Cranfield University

Ian Di Tullio

Improving the direct marketing practices of FMCG retailers through better customer selection. An empirical study comparing the effectiveness of RFM (Recency, Frequency and Monetary), CHAID (Chi-squared Automatic Interaction Detection), stepwise logit (logistic regression) and ANN (Artificial Neural Networks) techniques using different data variable depths

Cranfield University, School of Management
Ph.D. Dissertation

Academic Year of Submission – 2014

Supervisor: Dr. Stan Maklan

September 1, 2014

This thesis is submitted in partial fulfilment of the requirements for the degree of Ph.D.

Abstract

The intent of this thesis is to understand Data Mining technique effectiveness in both shallow (RFM variable only) and expanded data environments. The thesis addresses two specific gaps in research: (1) the relationship between customer selection techniques and performance and (2) the effects of using different depths of data on performance. In shallow-data contexts stepwise logit and neural networks provided the greatest cumulative lift and outperformed both RFM and CHAID across all top deciles. However, RFM shows the second highest fit measure, illustrating its relative stability in predicting outcomes. In addition, the RFM technique performance was tested using both one-month and 12-month time series. The 12-month series performed better and showed a greater level of fit. The subsequent study comparing technique effectiveness under expanded variable sets demonstrated an even more significant and visible lift increase versus the RFM technique. Looking at logistic regression, CHAID and neural networks, the lifts and gains obtained at the first two deciles provide enough response lift to allow these techniques' cumulative performance to surpass RFM well past decile five into decile six. From a cumulative perspective, the strong performance of logit and ANN allow these techniques to outperform CHAID in deciles one and two, but as of decile three, cumulative performance of all three advanced techniques becomes virtually identical. Though CHAID remains the technique with the best fit performance, RFM fit value falls to last place once an expanded variable set is introduced. Furthermore, both logistic and ANN performance increases significantly, and though they remain very close from an overall Gini and PCC score perspective, the logistic regression outperforms ANN when using expanded data. In both studies, dimensionality reduction plays a role in optimising model response. In limited data sets, logit applications reduced data to achieve better response, whereas in extended data sets, all models applied reductions. These findings contribute to the growing literature on customer selection techniques and provide a specific contribution to data mining, RM, segmentation and marketing practice by demonstrating how these techniques can be used for better consumer selection for purposes of customer development in FMCG retail.

ACKNOWLEDGMENTS

The author wishes to express sincerest and deepest appreciation to Dr. Stan Maklan for his invaluable support, encouragement, guidance and friendship during the entire Ph.D. process. The author also wishes to express appreciation to Professors Hugh Wilson and Simon Knox for their assistance in the preparation of this manuscript and for their high degree of pragmatism through the entire process. Special thanks are also offered to Dr. Colin Pilbeam and Professor Allan Harrison, whose understanding of my aspirations and objectives motivated me to push through the difficult times and emerge a better and more balanced researcher.

Finally, the author wishes to thank his parents, Donato and France Di Tullio, for encouraging him to always push himself beyond his limits and aim for the stars.

DEDICATION

I wish to dedicate this work to my wife, Dr. Lisa Campisi. Her courage, dedication and outlook on life have made me a better person and have helped me find true fulfilment.

Contents

Table of Tables

Table of Figures

1 Introduction

Marketers live in a period of exponential data proliferation. Data available includes not only the standard set of customers' demographics, buying patterns, usage and attitudes, product attribute perceptions and advertisement awareness (Curry, 1993) but, with the increasingly dominant role of the web as a marketing channel and the emergence of radio-frequency identification (RFID) and near field communication (NFC) technologies, online and offline customer browsing habits have also become traceable (Naik, Wedel, Bacon, Bodapati, Bradlow, Kamakura, Kreulen, Lenk, Madigan and Montgomery, 2008). This phenomenon has been recently referred to as Big Data; the emergence of a windfall of information in the form of "public, proprietary, and purchased sources, as well as new information gathered from Web communities and newly deployed smart assets" (Bughin, Chui, and Manyika, 2010, p. 7). Big Data is increasingly issued from non-traditional sources such as interconnected real-time supply chain data transmitting information on market demand, weather patterns, unstructured digital channels and clickstream data from the Web, social media content (tweets, blogs, product reviews, Facebook wall posts and profiles, YouTube video posts and comments), smart phone applications and other Internet-based gadgets and widgets (LaValle, Lesser, Shockley, Hopkins, Kruschwitz, 2011). But Big Data also encompasses everything from "call-centre voice data to genomic and proteomic data from biological research and medicine" (Davenport, Barth, and Bean, 2012, p. 22); even customer complaints are being encoded and made available for analysis (Spangler and Kreulen, 2007).

This proliferation of data and the capacity to leverage it are widely considered to be significant strategic assets for the organisation particularly because of their relationship to marketing productivity and competitive advantage. Keller (1993, p. 2) states "a firm's most valuable asset for improving marketing productivity is the knowledge that has been created about the brand in consumers' minds from the firm's investment in previous marketing programs." If actionable, relevant and positive, this knowledge increases the value of an organisation's market offering (Hunt and Arnett, 2006) and ultimately leads to greater customer loyalty (Keller and Aaker, 1998), reduced vulnerability to competitive actions (Aaker, 1991; Kamakura and Russell, 1991; Keller, 1998), product differentiation (Bharadwaj, Varadarajan and Fahy, 1993) and better profit margins (Keller and Aaker, 1998; Yoo and Donthu, 2001).

However, the extent of data proliferation presents a unique challenge as "marketers are being challenged by a deluge of data that is well beyond the capacity of their organisations to comprehend and use" (Day, 2011, p. 183). Lilien and Rangaswamy (2004) estimate that the quantity of customer data available to managers increased a thousandfold over a five-year period. Other research indicates that the amount of data available would expand tenfold from 100-billion gigabytes in 2005 to 1,000-billion gigabytes by 2010 (Gantz, Reinsel, Chute, Schlichting, McArthur, Minton, and Manfrediz, 2007). Regardless of the exact figure or multiple, this suggests that the rate of growth of information per person is growing at more than 50 percent per year (Day, 2011, Davenport and Harris, 2007). The gap between the high growth of information and the moderate rate of data consumption of individuals is one of the major barriers to leveraging the growing data asset. Unlike the rate of information growth per person, the average growth of information consumption per person is only two percent a year (Gantz et al., 2007), illustrating a growing surplus of idle and unproductive information. Thus "absent any breakthroughs in human beings' ability to process data, unless new tools and approaches are adopted, the gap will continue to grow" (Day, 2011, p. 184).

Fortunately, this increased amount of data can be processed and actionned intelligently given commensurate advances in data-mining tools, processes and Database Marketing applications (Verhoef, Hoekstra and van Aalst, 2002). These advances have the potential to improve managers' access to the data asset, unlock its latent value, and, in the case of marketing applications, increase the relevance of communications and targeted offers to increasingly smaller segments and individuals (Jain, 1997).

From a marketing communication and promotions perspective, the combined application of data mining and Database Marketing provides marketers with the opportunity to generate a two-way exchange with the consumer. In such an exchange, the consumer "speaks" via his/her transactional purchases (that are encoded in relational databases) and the organisation "replies" by providing more relevant experiences in the form of varied communications (sales and promotions, triggered marketing programs and exclusive events) (Blattberg and Deighton, 1996). Deighton (1996) describes this "interactive" phenomenon as the strategic use of IT and information to address a single consumer's needs and preferences using collected historical response and transactional data in order to adequately adapt the next communication. However, even with the advent of such tools and advances in statistical techniques to manage the surge of data, scholarly research has yet to provide practitioners with a good understanding of which single, or combination of, marketing campaign design decision(s) works best in practice (Wedel and Kamakura, 2000).

From a Database Marketing perspective, the practice of selecting the right customers for targeting activities can be traced back to the first commercial transaction in ancient times (Grönroos, 1994); however, its modern appellation as segmentation can be traced back to its introduction by Smith (1956). Since then, "segmentation has become a dominant concept in marketing literature and practice. Besides being one of the major ways of operationalising the marketing concept, segmentation provides guidelines for a firm's marketing strategy and resource allocation among markets and products. Faced with heterogeneous markets, a firm following a market segmentation strategy usually can increase the expected profitability" (Wind, 1978). Since then, applications of segmentation have evolved greatly with its granularity extending to segments of one. Instead of the historical clusters of homogeneous customers, such customers can now be scored individually based on their specific propensity to belong to a segment, and as such, operational concepts of segmentation, targeting and selection have become somewhat interchangeable.

Though academically researched for a long time, practices of customer selection currently exhibited by marketers still favour uncomplicated customer selection techniques, such as simplistic Recency, Frequency, Monetary Value (RFM) segmentations, over more complex ones, such as predictive models (Verhoef et al., 2002). Simple techniques dominate because they are easy to use, rapidly implemented (Kahan, 1998), familiar and easier for managers to understand (Verhoef et al., 2002; Marcus, 1998). This is consistent with Day's observation (2011, p. 184) that "mastering exploitation of an existing activity often crowds out the necessary sensing, experimentation, and exploration that is the essence of a dynamic capability." Beyond aversion to change and exploitation mastery, other factors behind the managerial adoption gap arethe low degree of accessibility of academic research (Reisz, 2010), the challenges of co-producing research that is relevant to practice (Lunt, Shaw, and Fouche, 2010) and the lack of a systematic approach to academic research (particularly in Database Marketing technique research). On this last item, Bose and Chen (2009, p. 14) state that "no research paper provides detailed guidelines on which model should be used under what conditions for direct marketing."

Whilst the body of literature assessing the effectiveness of different techniques in Database Marketing is growing (Deichmann, Eshghi, Haughton, Sayek, and Teebagy, 2002; Levin and Zahavi, 2001; Malthouse, 1999; McCarty and Hastak, 2006), there has been limited work assessing techniques in relation to specific marketing or promotional objectives. Quinn and Dibb (2010, p. 1241) summarise this gap by noting that academic researchers are mainly interested with "the choice of variables and multivariate techniques available for the analysis and validation of that output", "while marketing managers have prioritised how segmentation outputs can be implemented in practice. Hence, I would characterise current research as heavily technique-centric rather than practice or decision-centric.

That this divide continues today suggests academe has failed to respond to requests for managerial relevance from practitioners, and it is exacerbated by the what Leigh Sparks calls the academic push "to think ever more narrowly...focusing...energies on 4 * journal articles" that are in some cases "heavily quantitative-oriented, often unreadable, collection of ever more narrow "research" (cited in Lee and Greenley, 2010, p. 9). Quinn and Dibb (2010) contrast recent research priorities in market segmentation research with those highlighted by Wind in 1978 and Wedel and Kamakura (2000), suggesting that priorities have remained fairly constant in the last three decades. However, in the midst of a very slow moving research agenda, one of the most salient observations when comparing priorities through time is the increasing importance of managerial relevance and implementation. The fragmented research priorities to date are confronted with an increasing focus on managerial; relevance; thus requiring research to bridge the theory-practice divide (Bailey, 2009). This attempted bridging of academically rigorous theories in use and the implementation of theories in practice (Grishikashvili and Dibb, 2014) is summarised by Sparks in Lee and Greenley's (2010, p. 8) editorial on the theory-practice divide in marketing scholarship: "The divide between practice and theory is a real one, but bridging the divide won't be possible until we understand where the divide, at least in part, originated and how it is maintained. Perhaps if we stopped and thought about what we do, how we present it, and what we are trying to achieve being ever more reductionist and ever more narrow, then practitioners could understand us and recognise what we have to offer".

1.1 Scope

From a scoping perspective, researching the Database Marketing challenge can be addressed using a variety of different research approaches. For instance, adopting a social interpretive philosophical stance could orient research towards the human information assimilation challenge and the related sources of opportunity that exist in expert systems (Jacob, Gaultney and Salvendy, 1986); an IT domain perspective could direct research towards assessing the benefits of software applications in simplifying data handling and increasing absorptive capacity (Chen and Cheng, 2004); and an action research orientation could lead down the path of understanding the managerial decision-making process in developing functional Database Marketing outputs (Quinn, Hines and Bennison, 2007).

For my part, I take a positivist view to the research challenge of leveraging growing information to improve marketing performance. Using a systematic research process, I first sought to understand Database Marketing objectives, the related Data Mining applications and then looked to compare different Data Mining techniques' performances in achieving objectives. Given the task of scoping out the relationships between Data Mining techniques and all objectives of Database Marketing is improbable in one piece of research, I focus my research on the objective of developing the relationship with existing customers. Given that this objective can also be achieved in a variety of different manners (mass marketing, telemarketing, personal sales), I narrowed the scope of my research to the specific application of Data Mining techniques for customer selection in Database Marketing campaigns. Much like other approaches to direct-to-consumer marketing, Database Marketing decisions are multi-faceted (Nash, 1984). I chose to focus my attention on the output decision of selecting the right customer (also referred to as targeting) because the list of consumers to be targeted is generally considered the most important and most researched decision dimension (Bult and Wansbeek, 1995; Reutterer et al., 2006; Piatetsky-Shapiro and Masand, 1999). Using a well-researched dimension whose current practices illustrate a striking simplicity and lack of actionability (Boulding, Staelin, Ehret, and Johnston, 2005; Ryals, 2005; Bailey, 2009) presents a unique opportunity for systematic empirical research.

The metric chosen to assess customer penetration is promotional response, the most widely used Database Marketing metric (Verhoef, Spring, Hoekstra, and Leeflang, 2003). Data is sourced from a large North American fast-moving consumer-goods retailer. This industry is selected because of the richness of the data available. Given the important role of customer variables in improving technique performance, particular attention is given to the incremental gains in effectiveness generated by appending the right data to the right techniques.

Research is conducted in a sequential fashion starting with a systematic review of literature. The systematic review provides an empirical review of Database Marketing techniques and variables. Using the output from the systematic review to define its design, the second study uses a large-scale, transactional sample of twelve months of consumer data to examine the gains in effectiveness issued from the application of different leading Data Mining techniques. The third and last study examines the impact of adding customer selection variables into each of the previously used techniques.

1.2 Research Rationale & Motivation

I have worked in the area of marketing analytics and strategy for more than fifteen years. During this period, I've seen first-hand the significant investments that organisations have made in customer databases, in enriching these databases, and in customer relationship management (CRM) to enable more effective customer interactions. I've also observed that many organisations fail to deliver on the expectations created by these investments. The selection of my thesis topic was greatly influenced by this gap in benefit realisation. In focusing on the very narrow topic of customer selection, I wanted to showcase the immediate value of creation potential that exists from leveraging existing database assets and accessible Data Mining techniques. I saw (and still see) this as being the first of many projects aimed at illustrating how improved understanding of the applications of Data Mining can allow organisations to systematically leverage the huge investments made in CRM systems and databases to achieve specific marketing objectives and subsequently generate significant increases in marketing and organisational performance.

1.3 Research Relevance

In this period of growing data richness, marketing channel multiplication, increasing customer brand interaction and growing consumer expectations, organisations have invested heavily in CRM processes and technologies that capture and analyse customer data in order to improve marketing and service decisions. Firth (2006, p. 21) estimates that "over 50 percent of organisations may be in the process of implementing some form of CRM technology." Precise estimates of annual CRM expenditures are hard to come by given the lack of openly published information and the overlap between CRM software and implementation expenditures; nevertheless, researchers estimate global spending in CRM to lie between US $8-10-billion (Firth, 2006) and US $50-billion per annum (Payne and Frow, 2006). Ang and Buttle (2006, p. 5) put the blended expenditures; "including software and implementation at circa US$13-billion; with "$3.2-billion … spent on CRM software worldwide in 2005, and more than three times this amount, or $9.8-billion … spent on software integration, administration and maintenance." Regardless of the exact figure, expenditures are substantial and return on such investment is highly debated and often limited by a lack of dynamic capabilities needed to leverage firms' CRM resources (Maklan, Knox and Peppard, 2011). Regardless of these challenges and increasing costs, Gartner's 2012 CEO Survey found that CEOs cited "CRM as their most important area of investment to improve their business over the next five years” (Gartner, 2012, online). Given the huge investments already made in building databases and CRM systems, improving the criteria used to select customers in the context of Database Marketing campaigns should not only generate immediate and substantial campaign performance improvements for most firms but could also act as the starting point for a more systematic application of statistical techniques across marketing objectives. This suggests a sizeable opportunity to advance research in CRM effectiveness exists in further understanding customer selection techniques regardless of whether they are simple segmentation, cross-tabulation or more complex predictive techniques (Wedel and Kamakura, 2000).

1.4 Desired Contributions

The research aims to generate pragmatic science that can be applied both in academic and organisational contexts. This is in line with the view that theoretical and methodological rigor combined with practical relevance effectively translates into actionable and relevant research results (Tranfield, Denyer and Smart, 2003). This thesis proposes contributions that address two specific gaps: (1) the relationship between customer selection techniques and performance and (2) the effects of using different depths of data on performance. The second gap, de facto, examines the often-implicit impact of dimensionality reduction on performance, and will shed some light on the predictive strength of specific variables and variable types. The first gap is described by Ryals (2005) as lack of a comprehensive assessment of Data Mining techniques, as well as by Bailey (2009) and Foedermayr (2008) who call for empirical evidence on the effectiveness of more data-intensive techniques on customer selection and segmentation. Kumar, Venkatesan and Reinartz (2006) identify the second gap by positing that there are significant gains in direct marketing effectiveness generated from deeper and more relevant customer data. The second gap addresses the dimensionality reduction problem (selecting the most appropriate variables from an extended set of variables) presented by Li (1991) and Neslin, Gupta, Kamakura Lu and Mason (2006) as the variable selection problem. This issue is also highlighted by Lee, Lee, Cho, Im and Kim (2011) as a highly correlated data-sets problem. This thesis addresses these gaps using the retail FMCG context and data.

These contributions address the 2010-2012 Marketing Science Institute (MSI) priority to “Develop conceptual frameworks and methods for understanding customer experience and behaviour in an increasingly complex landscape" (MSI, 2007, online). On the specific topic of marketing segmentation and customer selection or targeting, the MSI states: “Firms require new ways to leverage information about customer preferences and behaviour (including from “addressable” social and mobile media) to enhance or supplant conventional strategic planning, market segmentation, and targeting approaches" (MSI, 2007, online).

An overview of expected contributions is illustrated in Table 1.

Table 1: Expected Contributions

Theory	Practice
• Provide systematic approach on how to increase promotional effectiveness when pursuing the development objective in FMCG retail • Illustrate the differences in usage of advanced techniques in academic and practice • Demonstrate the strength of advanced techniques • Demonstrate the strength of extended data (loyalty, usage, channel socio-demo) • Identify the most salient data variables • Support the role of database marketing in achieving repeat transactional behaviours in RM (Eggert and Stieff, 1999; Maklan et al., 2011)	• Illustrate benefits of applying a systematic framework to database marketing • Support CEO's prioritisation of CRM for business performance improvement (Gartner, 2012) • Demonstrate the value of advanced technique in data-poor contexts • Demonstrate the value of advanced technique in (potentially) data-rich contexts • Identify the most salient data variables • Challenge research that suggests that "contacts from companies in low-involvement-purchase situations are not considered personal and, therefore, have little or no impact on relationship development" (Leahy, 2011, p. 651) • Investigate whether some segments of customers in FMCG contexts could be amenable to developing a repeat-transactional relationship

Contributions to Theory

My results will demonstrate a specific improvement in the measurable contribution of marketing campaigns in FMCG through the application of enhanced targeting facilitated by advanced Data Mining techniques and access to extended customer data. These results inform Eggert and Stieff's (1999) concept of behavioural Relationship Marketing (RM), whose goal it is "to achieve repeat transactions through a process of interaction with the buyer, typically driven by economic goals rather than including some of the wider aspects of the exchange such as customer satisfaction" (Palmer, Lindgreen and Vanhamme, 2005, p. 318). This contribution to behavioural Relationship Marketing is akin to Coviello, Brodie and Munro's (1997) and Coviello, Brodie, Danaher, and Johnston's (2002) Database Marketing stage of RM as it also leverages personal consumer data for targeting and personalisation of mainly tactical promotional offers. This view is supported by Richards (2008, p. 126): "as managers are able to use these types of sophisticated models to mine the information available in CRM technology, improved customer targeting will result." This improvement in the measurable economic contribution of enhanced targeting is supported by Rust, Zeithaml and Lemon's (2000) contention that individualised marketing messages, facilitated by direct and data marketing practices, contribute to generating firm value.

The research seeks to validate the specific effects of marketing variables on marketing performance and to contribute to enhancing Database Marketing (and related segmentation) theory around what customer selection variables (what Wedel and Kamakura (2000) call “segmentation bases”) are most effective for marketing promotions in FMCG retail.

From a process perspective, the portion of the research that aims to apply data reduction techniques illustrates the value of implementing Data Mining practices across the KDD (Knowledge Development in Databases) and Database Marketing processes, thus also informing literature on the relative strength of those conceptual frameworks (Fayyad, Piatetsky-Shapiro and Smyth, 1996; Blattberg, Kim and Neslin, 2008).

Contribution to Practice

From a practical perspective, the research output seeks to improve the understanding of how different customer selection techniques increase promotional response rates and thus the financial performance of marketing campaigns. Furthermore, the research aims to demonstrate the value of richer data and validate what specific variables increase marketing performance. The work will not only provide practitioners with a sense of the value of data and technique applications but will also illustrate how such applications perform in data-rich and data-poor contexts. This will be particularly important as most CEOs are prioritising CRM investments to enhance business performance (Gartner, 2012).

This contribution to practice in Database Marketing (in the context of FMCG retail) will also investigate the claim that “contacts from companies in low-involvement purchase situations are not considered personal and, therefore, have little or no impact on relationship development” (Leahy, 2011, p. 651). More specifically, the research investigates whether some segments of customers in such contexts could be amenable to developing a repeat-transactional relationship. By seeking to identify higher-probability promotional responders, this research will inform Leahy’s recommended future research suggestion that there may be a segment that values direct contact from the FMCG retailer.

Finally, should the application of the academic framework demonstrate a positive impact on performance; this would also provide practitioners with a valuable road map with which to conduct effective database marketing.

1.5 Thesis Structure

The thesis structure is detailed in Table 2 below. The structure allows readers to understand the progression from initial research scoping in the literature review section, to detailed scoping in study one (systematic review), and finally the contribution to knowledge and practice issued from empirical research studies two and three and the discussion section.

The structure starts with an introduction that discusses the general area of research; the research aims, questions and contribution; and the research motivations. The next section reviews the adopted philosophical research perspectives. It provides the philosophical lens by which the research is conducted along with a rationale behind the family of methods selected to research the topic area.

The literature review section that follows provides an overview of research domains informing the research area, identifies the main methodological bases for customer selection, and establishes some of the key questions and hypotheses to be taken and validated in the systematic review. The domain mapping will mainly examine the relevant research domains that inform the research area and provide a narrative overview by: (1) identifying the leading relevant concepts and definitions for each domain, (2) contrasting and critiquing concepts and definitions, (3) identifying the most-relevant definitions of terms, and (4) positioning the research within these domains. From this domain overview, the literature review subsequently identifies the main methodological bases set in the domain literature to conduct research on customer selection. These key bases include: (1) objective definition, (2) metric selection (dependent variable), (3) applicable techniques, (4) independent variables, and (5) performance measures. As the research topic is widely researched and wide-ranging, the literature review's aim is to provide a direction for future research to be conducted via the systematic review. As such, the literature review concludes with some key questions that need to be further clarified, and hypotheses that need to be validated in the context of a systematic review.

The systematic review that follows examines the key questions and hypotheses from the literature review in a much more organised, transparent and methodical manner. By framing the research against key research questions and search strings, the output is much more focused and allows for an effective framing of a research design that addresses the research topic of customer selection. The output of the systematic review frames the empirical research of the two subsequent empirical studies by identifying core objectives of customer selection, measures of effectiveness, the most used methods and variables applied with modelling techniques, and providing an initial assessment of relative performance. The systematic review output thus creates the foundation for the structure and content of subsequent quantitative research studies. As a result, studies two and three provide an in-depth quantitative analysis of the recommended research questions that emerge from the systematic review; namely the relative strength of highly effective techniques, the strength of different data variables and the moderating effects of data reduction. For parsimony and simplicity purposes, given study three is a follow-up to study two and results are highly relatable and comparable, the discussion section of study three will incorporate results from study two. This will allow for a comparison of results between shallow (RFM variables only) and extended variable applications, and will allow for a richer discussion of findings.

The empirical studies are followed by a review of implications for practice and theory, a discussion of limitations, and a conclusion section that also provides a short summary and identifies directions for future research.

Table 2: Thesis Structure

Section	Title	Content and Objectives
1	Introduction	• Roots and focus of the research • Research aims and questions • Research gaps and contribution • Discusses motivations behind the research and research rationale • Presents existing research gaps and location of the contribution • Reviews of the overall thesis structure
2	Philosophical Perspective	• Presents array of available research perspectives • Discusses the adopted philosophical research perspectives
3	Literature Review	• Identifies the phenomenon of interest • Identifies relevant research domains and concepts • Defines key terms • Identifies questions for systematic review • Identifies hypotheses for systematic review
4	Study 1: Systematic Literature Review	• Reviews systematic review methodology • Reviews literature on the tasks, techniques, applications and usage of Data Mining techniques • Reviews literature, reviews findings on objectives of customer selection, measures of effectiveness, selection techniques, their effectiveness and their moderators • Identifies research gaps and developing research propositions
5	Study 2: Effectiveness of Data Mining Techniques	• Reviews applied methodology • States dataset composition, origin, and variables • States set of techniques examined, and rationale for selection • States performance measures, and rationale for their selection • Specifies research proposition and research design • Analyses reliability of sampling • Validates/invalidates research propositions • Interprets emerging research outputs and parameters • Discusses the findings of the study
6	Study 3: Effectiveness of Data Mining Techniques (using richer data and imensionality reduction)	• Reviews applied methodology • Presents additional data variables • Examines techniques with dimensionality reduction function • Specifies research proposition and research design • Validates/invalidates research propositions • Interprets emerging research outputs and parameters • Discusses the findings of the study & contrasts them with study 2
7	Contribution	• Calls out overall research contributions
8	Implications for Practice & Theory	• Discusses research contributions in respect to practice and theory
9	Limitations	• Discusses research limitations
10	Conclusion & Future Research	• Suggests direction for future research • Summary and conclusion
11	Bibliography	
12	Appendices	

2 Philosophical Perspective

2.1 Introduction

This section discusses the ontological and epistemological perspective of this thesis. Adopting adequate philosophical stances is central to scholarly work because ontological and epistemological positions offer: (1) a varied and important foundation for coherent scholarly choices and their explanatory outcomes, (2) an often unrecognised source of disagreement about the relevance and trustworthiness of scholarly claims, and (3) a potential basis for novel contributions to scholarly conversation (Huff, 2007).

To look at the phenomenon of interest, I have adopted the realist ontological perspective and the positivist epistemological perspective, best defined by Huff (2007) as the logical positivist perspective. The following section provides an overview of the ontological and epistemological alternatives in social science and the rationale for the above choices for this thesis.

2.2 Ontology and Epistemology

Ontology is defined as the theory about 'what exists.' Orlikowski and Baroudi (1991, p. 7) state that ontological beliefs "have to do with the essence of phenomena under investigation; that is, whether the empirical world is assumed to be objective and hence independent of humans, or subjective and hence having existence only through the action of humans in creating and recreating it." Epistemology focuses on what humans can know about what exists. According to Orlikowski and Baroudi (1991, p. 8) "epistemological assumptions concern the criteria by which valid knowledge about a phenomenon may be constructed and evaluated." Burrell and Morgan's (1979) work, adapted by Anne Huff (2008), provides a simple overview of the ontological and epistemological alternatives available to researchers (Table 3).

Table 3: Ontological and Epistemological Alternatives in Social Science

Ontology	**Nominalist:** Labels are artificial creations. Social reality in particular is relative. **vs.** **Realist:** The real world, including the social world, exists independent of labels.
Epistemology	**Anti-positivist:** Scholars cannot create objective knowledge. They can only report on their own experience. **vs.** **Positivist:** Scholars can explain and predict by searching for patterns and testing hypotheses. Knowledge is cumulative.

Source: Adapted from (Huff, 2008)

The table shows that both ontology and epistemology exist across a wide spectrum. Namely, from an ontological perspective, alternatives span views that can be nominalist whereby reality is relative and socially derived, or realist, where the real world is said to exist free of social interpretations and independent of labels ascribed to it. From an epistemological perspective, a similar span of alternatives is defined. On one end of the spectrum, the phenomenological and generally anti-positivist perspective states that objective knowledge does not truly exist but rather that researchers can only report what they perceive to be their personal experience; thus, the researcher can never be separated from the sense-making process. Conversely, the positivist perspective takes the view that not only can objective knowledge be created and phenomenon explained but that this knowledge can also be cumulative through the aggregation of past research and further testing. Though Huff's spectrum illustrates the span of alternatives, there are numerous ontological perspectives that exist in between. A description of the different philosophical views is provided by Easterby-Smith, Thorpe and Lowe (2002) in Table **4**.

In their work, Easterby-Smith et al. (2002) identify three main ontologies of science: traditional realism, internal realism and relativism as well as three main ontologies for social sciences: representationalism, relativism and nominalism. The difference in ontologies emerges from the differing subject matters, social sciences being concerned with people whereas pure sciences are more concerned with physical objects. The topic of customer selection, though related to people, is mainly related to determining what mathematical derivations are most effective. Thus, it is more relevant to examine the ontologies of science. Nevertheless, it is important to note that this is only an overview of main ontological perspectives; there are numerous other intermediary perspectives that exist. For example, critical realism bridges the gap between realism and relativism as it recognises that social conditions have real consequences while also recognising that all concepts are issued from human interpretation.

Table 4: Ontologies in science and social science

Ontology of Science	**Traditional Realism**	**Internal Realism**	**Relativism**	
Ontology of social Science		**Representationalism**	**Relativism**	**Nominalism**
Truth	Is established by correspondence between observations and phenomena	Is determined through verification of predictions	Requires consensus between different viewpoints	Depends on who establishes it
Facts	Are concrete	Are concrete, but cannot be accessed directly	Depend on viewpoint of observer	Are all human creations
Epistemology of science	***Positivism***		***Relativism***	
Epistemology of social science		***Positivism***	***Relativism***	***Social Constructionism***

Source: Easterby-Smith, Thorpe and Lowe (2002)

Easterby-Smith et al. (2002) also highlight the main epistemological perspectives in science – positivism and relativism – and in social sciences – positivism, relativism and social constructionism (what Huff refers to as anti-positivism). The relativist perspective (the only one not discussed in Huff's continuum) refers to a view that generating knowledge is a relative concept. Knowledge can be generated in a realist fashion; however, it must exist within a social context and rely on consensus.

2.3 Adopted Research Philosophy

As in any inquiry, research topics may be investigated using a variety of lenses that vary with the type of inquiry conducted. For example, the research topic of this thesis may be examined via either a social constructivist (or positivist) perspective or a realist (or relativist) perspective. Indeed, many of the environmental, structural and human moderators of effectiveness discussed thoroughly by Dibb (2001) support the need for a more interpretative stance on Data Mining and customer selection. This view would be appropriate for considering the effect of such moderators on the implementation of customer selection techniques, the managerial capacity to handle complexity and execute it and other similar questions. However, the core focus of this thesis is less concerned with the interpretive perspectives about how people construe their environment, and more concerned with the explicit application of Database Marketing techniques, the understanding of which statistical techniques provide the highest response rate amongst customers targeted with a promotional offer.

Conducting research in the quantitative fields of Database Marketing and Data Mining is different from research conducted in other pure social sciences because there exist intrinsically tight relations between Database Marketing and the quantitatively oriented fields of mathematics and statistics. In fact, customer selection variables and empirical Database Marketing results are generally known and tangible. They emerge from an existing, measurable and tangible data set and, as such, indicate that the world of Database Marketing exists free and independent of socially ascribed labels (Huff, 2007). They also exist and have a specific form that can be measured in a quantitative fashion to provide empirical results for adoption in future Database Marketing applications. Though still a rapidly evolving area of practice, Database Marketing and Data Mining have been highly researched and can be qualified as being mature in a sense. These factors have a direct impact on the realist and positivist research stances adopted in this thesis, given that they indicate that the field of Data Mining can be mathematically reduced to elementary concepts and rules and occurs independent of “emotive” concepts. Furthermore, the intrinsic nature of scientific inquiry in fields highly related to mathematics and statistics invariably leads to authentic knowledge being sourced from the positive affirmation of theories informed by the application of a strict scientific method.

Within this context, cross-sectional or variance theories are helpful for identifying systemic relationship patterns (Langley, 2007). Granted, they don’t offer the temporally embedded processual accounts that facilitate the understanding of longer-term performance, variance theories suggest that a positivist view can effectively help tackle the tactical underpinnings of broader strategic philosophies. This thesis studies the tactical underpinnings of a much broader RM and CRM inquiry. Through the decomposition of factors and techniques that lead to increasing customer response, this thesis will help enhance knowledge on a portion of the broader inquiry on RM and CRM performance.

This viewpoint is certainly not unique. Nearly 83 percent of the studies included in the systematic review of specific Database Marketing are quantitative in nature and use analysis to mainly validate or falsify hypothetical outcomes through experimental design. The pursued outcome in most of these studies is usually the demonstration of causality. These studies ultimately end up being highly positivist with all the research design implications shown in Table 5.

Table 5: Overview of methodological implications of different epistemologies

Elements / Epistemology of Methods	**Positivism**	**Relativism**	**Social Constructionism**
Aims	Discovery	Exposure	Invention
Starting Points	Hypothesis	Suppositions	Meanings
Designs	Experiment	Triangulation	Reflexivity
Techniques	Measurement	Survey	Conversation
Analysis/Interpretation	Verification/falsification	Probability	Sense-making
Outcomes	Causality	Correlation	Understanding

Source: Easterby-Smith, Thorpe and Lowe (2002)

Though I could address ontology and epistemology distinctly and separately, and elaborate a view on where my research positions itself vis-à-vis both, I have chosen, rather, to look at them in tandem given their indivisible nature. This view is supported by Crotty (1998, p. 10): “to talk of the construction of meaning is to talk of the construction of meaningful reality.” Huff (2008) provides an overview of what she calls research worldviews and highlights how they line up with ontology, epistemology and scholarly activity outputs (Table 6). The research view that best espouses the principles of realism and positivism is logical positivism. Logical positivism reduces theory to the certainty of logic, mathematics and statistical probability coupled with propositions issued from sense experience (Cacioppo, Semin and Bernston, 2004). Though both positivism and realism are certainly the broader views that represent the philosophical grounding of this research, logical positivism brings them together. This choice of stance and its alignment between the positivist stance and quantitative experimental design is strongly supported by the work of Huff (2007), Cacioppo et al. (2004), and Morgan and Smircich (1980). Other than indicating the particular conversation one wishes to join, philosophical choices have some significant implications on what preferred research methods to adopt. Morgan and Smircich (1980) indicate that realist and positivist stances are usually best informed by conducting lab experiments or surveys. This is in line with Easterby-Smith et al.’s (2002) conclusions on research approaches to adopt with a positivist epistemology. This alignment is further validation that the selected philosophical stance is appropriate and adapted for the research topic at hand.

2.4 Conclusion

The research aims to generate pragmatic science that can be applied both in academic and organisational contexts. Much of the extant research in database marketing and data mining (and identified in the systematic review) adopts a logical positivist stance. The extent database marketing literature largely adapts a logical positivist stance as the evaluation of techniques' performance is grounded in logic, mathematics and statistical probability. As a result, the research design of this positivist research and the related articles quoted throughout the thesis are heavily rooted in quantitative analysis and experimental design (Morgan and Smircich, 1980).

This literature does not assess techniques in a systematic fashion, contrasting their relative effectiveness in controlled experiments. Nor does it necessarily provide practical guidance to marketing managers. This has been identified as the academic-practitioner gap (Quinn and Dibb, 2010). This thesis will make contributions to closing this gap by addressing the research question systematically, connecting the research to a shared practitioner agenda, and demonstrating the value of research outcomes in a context relevant to both academics and practitioners..

Table 6: World Views Influencing Scholarly Conversation

World View	Ontology	Epistemology	Focus of Scholarly Activity
Rationalist, Formalist Descartes Leibniz	The world is intelligible and orderly.	"Logic is a source of knowledge, independent of intuition or deduction."	Reasoned deductive description.
Empiricist Locke Hume Mill	"Knowledge of the physical world can be nothing more than a generalisation from particular instances and can never reach more than a high degree of probability."	"There is no source of knowledge other than sense experience." "For most empiricists, experience includes…reflection upon the mind and its operations."	Inductive generalisation.
Logical Positivist Comte Vienna Circle Wittgenstein Russell Carnap	The world is orderly and can be reduced to elementary concepts and rules. Its nature can be described without "emotive" concepts.	"The only authentic knowledge is scientific knowledge, and that can only come from positive affirmation of theories through strict scientific method."	Theories reduced "to the truths of logic and mathematics coupled with propositions referring to sense experience."
Pragmatist Pierce Dewey Rorty	"Truth is modified as discoveries are made and is relative to the time and place and purpose of inquiry."	"Thought is …simply an instrument for supporting the life aims of the human organism and has no real [epistemological] significance."	"Theories …ultimately justified by their instrumentality, or the extent which they help people meet their aims." (Rorty)
Social Constructivist Social Constructionist Hegel Durkheim Burger & Luckmann	"Individuals and groups participate in the creation of their perceived reality."	"All knowledge, including the most basic, taken-for-granted common sense knowledge of everyday reality, is derived from and maintained by social interactions." (Burger & Luckmann)	Theories that 'create' knowledge. Insight created from multiple accounts.
Historicist German Historicists Boas New Historicists	Change in historical objects (such as states) is the product of (local) causes external to those objects.	Since all knowledge is dependent at least to some extent on entities that are historically embedded, then all knowledge is limited by and relative to historical locations.	Theories that describe processes of change and development.
Critical Theorist Frankfurt School Marx	Historical materialism…shaped by class, ethnicity, gender, and socio-economic values.	Grand narratives dominate local knowledge. But there can be local resistance or grand knowledge narrative.	Theory that "aims to explain and transform all circumstances that enslave human beings."
Feminist Wollstonecraft Harding	"Social, political and economic inequality between the sexes which favours the male gender."	"The perspectives of marginalized individuals can help to create more objective accounts of the world." Harding: standpoint theory.	Theory that corrects androcentric bias. (However, debate over the purpose of feminist theory "defines the movement)."
Postmodernist Lyotard Latour	The world is transient, ephemeral, and emergent.	Knowledge and power are fragmented. "Local assemblages of 'organisings' …collectively make up social reality.	"Polyphonic, juxtaposed readings and writings [in a]…chorus of narratives. Play with varied perspectives.
Scientific Realist Einstein Putnam	"Unobservable things talked about by science are little different from ordinary observable things (such as tables and chairs)." "They exist independently of the minds or acts of scientists."	"Subject to the recognition that scientific methods are fallible and most scientific knowledge is approximate, we are justified in accepting the most secure findings of scientists 'at face value'."	Scientific success. Successively improved explanations.
Critical Realist Russell Sellars Bhaskar	Some sense data accurately reflect external entities, events, and so one, while other sense data do not.	"The objects of knowledge do not depend, for either their being or their nature, upon the knowledge of them." However, ideas about physical things are "molded by the conclusions of science."	Theory that produces generalised claims about causal mechanisms, often using multiple methods and abductive reasoning.

Source: Huff (2007)

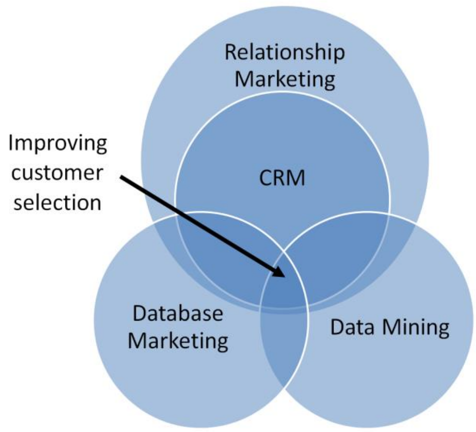
3 Literature Review

3.1 Introduction

By definition, a literature review is a selection of available texts and documents on a topic of interest which contains information, data, insight and evidence aimed at expressing a point of view on the topic, shedding light into how the topic is best investigated, and providing a critical assessment of the revised literature in relation to a research proposal. The ultimate aim of the literature review is to “demonstrate skills in library searching, to show command of the subject area and an understanding of the problem, to justify the research topic, design and methodology” (Hart, 2006, p. 13). Given the breadth and complexity of the field, the traditional literature review did not allow me to effectively reduce literature in order to identify: specific leading techniques, adapted methodological approaches and best metrics to utilise in the context of the research topic. To do so, I have elected to complement the literature review with a systematic review as my first study. The systematic review approach allows me to segue from a topical review subject, design and methodology to specific review of these areas. In addition, the systematic review also allows me to do so in a manner that is transparent and methodical.

This literature review first maps the field of research and defines which literature informs the phenomenon of interest. This mapping provides a narrative overview of: (1) the salient research domains, (2) the leading relevant concepts and definitions for each domain, (3) a contrast and critique of these concepts and definitions, (4) the selection of most relevant definitions of terms, and (5) the positioning of the research within these domains. Doing so allows me to identify the relevant texts that inform the topic, develop a point of view, and understand the designs and methodologies applied for the study of the research topic. The main research domains that inform this research on customer selection are relationship marketing, customer relationship management (CRM), database marketing, and Data Mining (Figure 1).

Figure 1: Phenomenon of interest: Key bodies of literature



Though each domain is described in subsequent sections, I summarise the relationships between each domain as follows:

- Relationship Marketing (RM) provides the theoretical grounding for customer-centred marketing;
- Customer relationship management (CRM) relates to the application of RM concepts with thorough customer information;
- Database Marketing "provides the a technological enabler, allowing vast quantities of customer-related data to be stored and accessed in ways that create strategic and tactical marketing opportunities" (Chaffey, Mayer, Johnston, and Ellis-Chadwick, 2000, p. 291); and
- Data Mining provides the techniques and processes by which data can be sourced and manipulated to create actionable insight from databases.

Looking at the sub-domains across the literature, there exists a high degree of overlap between them, and the interpretation of how the domains intersect (or indeed are sub-domains of one another) differs greatly. This presentation of CRM as a subset of RM is based on an interpretation of RM as the broader overriding concept. This view is supported by many authors, including Ryals and Payne (2001), Lindgreen (2001) and Das (2008). In the views of these authors, sub-domains are intrinsically related to RM and provide the tools, techniques and processes to effectively achieve its objectives. The classification of the different disciplines in RM provided by Das (2008) goes so far as defining CRM merely as direct marketing. Database Marketing and one-to-one marketing are thus the sole instruments of CRM. Though simplistic, it nevertheless underscores the somewhat different roles that domains can play in informing the application of Relationship Marketing strategies. Subsequent sections will delve deeper into roles, definitions and positions vis-à-vis one another. Though this research does somewhat align with the directional role of domains presented by Das (2008), I adopt a more strategic (yet complimentary) interpretation of CRM as a domain that emphasises the integration of processes and information across different functions "with particular emphasis on customer relationships – turned into practical application" (Gummeson, 2002, cited in Frow and Payne, 2009, p. 13). This interpretation underlines that, although both RM and CRM speak to the domain of managing relationships, they are different in scope.This interpretation of CRM is supported by Dibb (2004) and Frow and Payne (2009) as it includes the application of specific customer information assets, technology and processes.

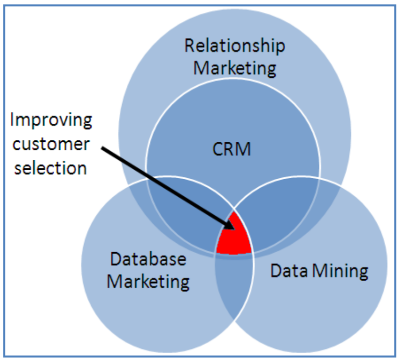
From a historical perspective, as discussed later, the first field to emerge was direct marketing (as a predecessor of Database Marketing). Soon thereafter, as utilisation of direct marketing expanded to other industries and data became more accessible, Database Marketing and Data Mining were popularised (Harker and Egan, 2006). RM as a formal domain (though its practices were applied in service marketing as early as Transactional Marketing) only officially emerged years later when the Transactional Marketing paradigm was challenged for its limited scope on customer acquisition (Berry, 1983). The changing scope and reframing of marketing as a customer-oriented field (versus the product orientation of Transaction Marketing) led RM to become one of the dominant marketing paradigms. As literature on RM increased, it was made operational by CRM, which overtook it in the literature as the focus of publication in early

2000s. Nevertheless, literature of all domains co-evolved to a large extent as CRM was (and still is) constantly affected by changes in Database Marketing and Data Mining, just as RM evolves and is affected by the B2B, services marketing, direct/Database Marketing and channel marketing literature (Moller and Halinen, 2000).

In the following sections, these terms are defined against recognised definitions and key concepts are identified and critiqued. More specifically, the section on RM grounds the research in marketing theory and examines what concepts it informs. The CRM section illustrates how the RM theory is applied using IT, data and processes as enablers. The bulk of the literature on the research topic itself lies in the section on Data Mining given that the main interest of this thesis lies in the application of Data Mining techniques. In addition to an overview of the field and its main concepts, the Data Mining section also provides an overview of: (1) the Data Mining tasks, (2) the phases of the Data Mining process, (3) the main Data Mining techniques, (4) the normative practices of Data Mining, and (5) the practical applications of techniques. These are all important to establish a strong common foundation for understanding subsequent sections. Finally, the section on Database Marketing, which in essence describes the application of Data Mining techniques using relational databases, is mainly focused on the explicit definition of the term, its objectives and its relationship to the other domains. Finally, in the summary section I summarise the previous sections and provide a holistic point of view that acts as the lead-in for a systematic literature review of the research topic.

Prior to exploring each of the literature domains, I first provide an overview of the research topic: customer selection. I then cover, in order: RM, CRM, Database Marketing and Data Mining.

3.2 Customer Selection



Customer selection, often referred to as customer targeting, has always been practiced by firms implicitly in their efforts to identify the best possible target audience for a product. Venkatesan, Kumar and Bohling (2007, p. 580) define its use by firms as: “Faced with a limited annual marketing budget, the firm can proactively contact only a fraction of its current customers and therefore must select customers to target with marketing communication.” In the context of direct and database marketing, the term becomes more precise and is best captured by one of four decision variables that define a successful Direct- or Database-Marketing effort: (1) the promotional offer, (2) communication elements, (3) timing or sequencing of communications, and (4) customer selection (Nash, 1984). Increased insight into individual customers has significant implications on each of these variables, and several researchers have suggested that customer profits can be considerably enhanced if marketing contacts were adapted to individual customer preferences (Ansari and Mela, 2003; Reinartz, Thomas, and Kumar, 2005; Venkatesan and Kumar, 2004).

However, decades of advances in statistical techniques and scholarly research have yet to provide practitioners with a good understanding of which single, or combination of, decision variable(s) works best in practice (Wedel and Kamakura, 2000). Venkatesan et al. (2004, p. 593) state that “practitioner implementation of this recommendation (customising customer contacts) is limited because the previous academic research does not show any causal link between using optimal contact levels and maximised customer profits.” Of these four variables for customising marketing contacts, customer selection is generally considered to be the most important and most researched (Bult and Wansbeek, 1995). As such, this thesis addresses gaps in the application of Data Mining techniques for purposes of customer selection optimisation in Database Marketing; more specifically, I examine the statistical and machine-learning techniques that allow customer selection to be most effective and best discriminate highly responsive consumers from non-responsive ones. The objective is to find a combination that increases response rates and promotional returns. Though used interchangeably with terms such as targeting (Kim, Street, Russell and Menczer, 2005) and operational segmentation (Wedel and Kamakura, 2000), customer selection is used as a term of reference given its prevalence and identification by Nash (1984). Authors refer to customer selection by describing it as contacting “customers through multiple channels, such as through a salesperson, by direct mail (including promotional catalogs and e-mail), and by telephone” (Venkatesan, Kumar and Bohling, 2004, p. 580); “marketing contacts across various channels” (Venkatesan and Kumar, 2004, p. 121); “consider which customers to target with the firm’s marketing program” (Winer, 2001, p. 95), target marketing (Anderson, Jolly and Fairhurst, 2007), and “the number of customers a firm should target in its market communications, and cross-buying models function to increase the number of products purchased by a customer”

(Reinartz and Venkatesan, 2008, p. 292).

Given the B2C focus of this research, in order to define the customer selection term, a direct-to-consumer angle must be adopted. This is in line with the definition of Database Marketing found later. Furthermore, given the importance of applying Data Mining techniques to the selection process, I also consider this to be of critical importance to crafting a definition of customer selection.' In light of these critical inputs from literature, I define the term customer selection as follows:

> *The selection of customers for the purpose of direct communication through the utilisation of patterns, statistical or predictive models issued from data to obtain a measurable response or transaction via one or multiple channels*.

As most customer selection in practice is informed by the application of statistical and machine learning techniques (Fayyad, Piatetsky-Shapiro and Smyth, 1996; Blattberg et al., 2008), it finds its roots in the fields of mathematics and statistics, computer sciences, economics, consumer behaviour and marketing literature. From mathematics and statistics, it adopts its mathematical formulations and statistical applications. From computer science/information technology (IT) it takes the computational power, relational database capabilities, artificial intelligence applications and software automation that allowsthe scalability of statistical applications to sizeable transactional and relational databases. From economics and microeconomics it sources the "assumption that the marketing mix is used to help a company 'optimise' [maximise] its profit function (Waterschoot and Van den Bulte, 1992; Grönroos, 1991). From consumer behaviour it sources the consumers' desire to reduce the complexity of the buying situation (Sheth and Parvatiyar, 1995). From marketing comes the consumer relational theories (Harker, 2006), and from marketing practice comes direct and Database Marketing applications.

Given the depth and breadth of literature in each domain, I examine the relevant core assumptions of each and look at the sub-domains of research that exist at their intersections in subsequent sections. However, prior to doing so, I provide an overview of Segmentation as a predecessor practice.

Segmentation as predecessor to Customer Selection

Market segmentation was first cited in the 1950s by Wendell Smith (1956) as "viewing a heterogeneous market as a number of smaller homogenous markets in response to differing product preferences amongst important segments." During the same period, marketing as a key driver of business strategy was advanced by Drucker (1954) and the specific importance of segmentation underscored shortly thereafter by Levitt (1960) as he stated: "if you are not talking segments, you're not talking marketing." Since then, alternative definitions have abounded and, though there is agreement on the need to identify homogeneous groups, differences persist. Much like the domains of RM and CRM, segmentation is defined strategically, as Analysing market attractiveness and auditing the company's critical success factors, or operationally, as a tool for resource allocation and planning (McDonald and Dunbar, 2000).

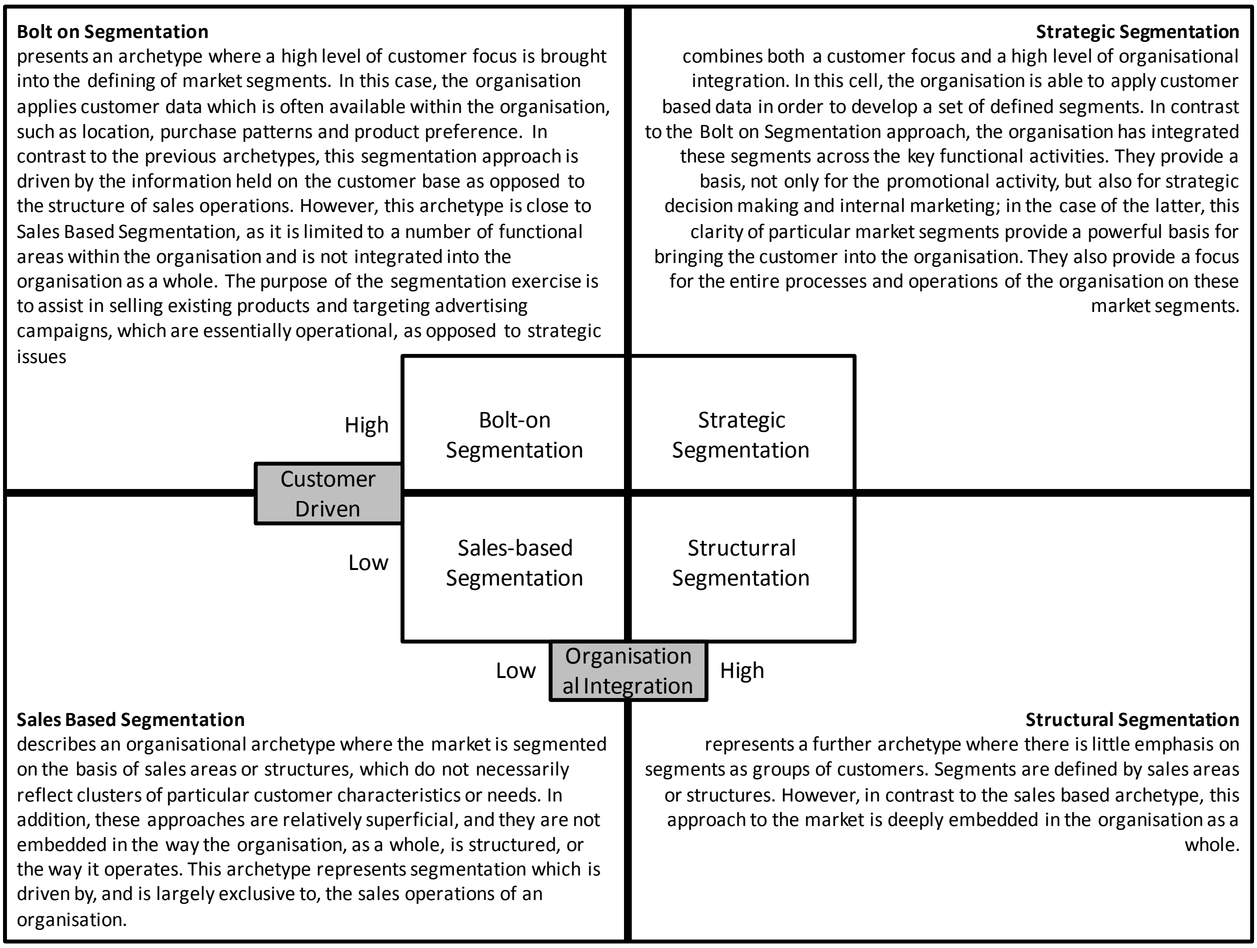
Types of segmentation

The strategic segmentation view addresses fundamental organisational questions, including who are our customers, what do they look like, what are their needs, and what are the potential value propositions that will meet those needs. The tactical view informs managerial decisions on the implementation of the market mix; for example, who are our most price-sensitive shoppers. Tactical segmentations address customer or organisational concerns generally individually and in absence of any constraints whereas strategic segmentations address both customer and organisational concerns simultaneously with all their related constraints. These views are conceptually captured in Jenkins and McDonald's (1997) Market Segmentation Archetypes, and are encompassed by four types of segmentation: Sales-Based, Structural, Bolt-on and Strategic. The type of segmentation approach adopted is largely a function of the degrees of customer centricity required (largely driven by firm strategy) and organisational integration. These archetypes and definitions are found in Figure 2..

In light of these different segmentation archetypes and segmentation orientations, some important differences appear in segmentation research. Blocker and Flint (2007) identifies this as the difference between research related to the pursuit of segmentation as a competitive advantage (Smith, 1956; Bonoma and Shapiro, 1983; Freytag and Clarke, 2001; Kotler and Bliemel, 2000) and research focusing more on developing it as a technique that employs sophisticated mathematical modelling (Brusco, Cradit, and Tashchian, 2003).

Segmentation can also be conducted in an a priori manner (intuitive based on specific, pre-determined variables) or post hoc (empirical based on segments and variables derived from data analysis).

Figure 2: Segmentation Archetypes and Definitions



Source: Jenkins and McDonald (1997)

Objectives of Segmentation

Though there is some disagreement on the definition of segmentation, the main managerial objectives of market segmentation are generally consistent. Doyle (1998) and Yankelovich (1964) identify the key reasons for segmenting markets as:

- Better matching of customer needs to the firm’s offer
- Enhancing profits (e.g. pricing discrimination based on individual price elasticity)
- Retaining customers
- Developing targeted communications to appropriate audiences

Sausen, Tomczak and Herrman’s (2005) empirical study identifies five managerial objectives of market segmentation:

- Acquisition of new customers
- Retention and penetration of existing customers
- Increasing customer profitability

- Achievement of higher efficiency of marketing activities
- Identification/exploitation of new sub-markets

These objectives are very similar to the stated objectives of Relationship and Database Marketing and are well in line with the the practical applications of bolt-on segmentation. However, Levin and Zahavi (2001, p. 3) point out that segmentation is “used to distinguish between customers and non-customers... to understand their composition and characteristics... and supports a whole array of decisions, ranging from targeting decisions to determining efficient and cost-effective marketing strategies, even evaluating market competition.” The profiling, competitive evaluation and new market identification dimension of segmentation, though certainly informing the domains of Relationship and Database Marketing, are broader in scope that the strict domains’ areas of interest and illustrate the importance of considering objectives based on the type of segmentation activities pursued by the firm, strategic or operational.

In addition to segmentation types, numerous approaches to conducting segmentation have been documented in literature. The normative segmentation process consists of a sequence of logical steps: market segmentation, targeting, and marketing mix development (Dibb and Simkin, 1994; Rudelius, Walton and Cross, 1987; Wind, 1978). Another process presented by Foedermayr and Diamantopoulos (2008) identifies six stages in the segmentation process: (1) segmentation variable selection, (2) segmentation method selection, (3) segment formation, (4) segment profiling, evaluation and selection of final target segments, (5) implementation, and (6) segmentation strategy evaluation.

Both processes can be conceptualised as an input to the Database Marketing process outlined by Blattberg et al. (2008) (similar to Data Mining-see next section). , Foedermayr and Diamantopoulos’ (2008) steps of the segmentation process can be included into the compilation, data analysis, design and evaluation steps of the Database Marketing process. The normative segmentation steps of market segmentation, targeting, and marketing mix development also align respectively with the steps of compiling and Analysing data and campaign design.

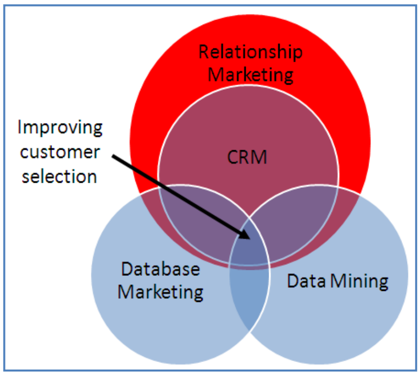
From an operational standpoint, Verhoef (2003, p. 471) states that segmentation in Database Marketing (also referred to as list segmentation in the context of marketing campaigns) “serves to group customers into clusters, which are internally homogenous and mutually heterogeneous, implying that the members of a segment react to (direct) marketing actions similarly, but differently than members of another segment.” This view of segmentation as an array of statistical techniques for purposes of profiling or response modelling has significant similarities with the Data Mining domain. This is supported by Verhoef’s (2003, pp. 471-472) view that “in the Database Marketing literature, a number of techniques, such as cross-tabulations, CHAID, probit analysis, and neural networks, have been examined that can be used in segmentation and/or predictive modelling” illustrating the interchangeability of the terms segmentation and predictive modelling as they pertain to outputs of Database Marketing. This is also supported by Danneels (1996, p. 36) as he observes that targeting is an inherent part of the segmentation process: “segmentation as a marketing tool thus involves dividing the market into homogeneous groups (segmentation), selecting one or more segments (targeting), and tailoring a marketing program to the target group (marketing mix development).”

The purpose of the bolt-on segmentation exercise is “to assist in selling existing products and targeting advertising campaigns, which are essentially operational, as opposed to strategic issues” (Jenkins and McDonald, p. 27). Such a definition is tightly related to all the domains pertaining to customer selection.

The relationship between bolt-on segmentation (that characterises this research) and the more operational facets of RM and CRM, the partial overlap between the segmentation process and the broader Database Marketing process, and the linkages between the segmentation techniques and Data Mining illustrate that literature on segmentation intersects with the multiple domains informing customer selection. As such, it will not be covered as a domain in itself but rather as a concept that is included in and informs the domains of customer selection.

Finally, as a point of clarity, this thesis researches the improvement of direct marketing practices using an existing customer database that contains historical transactional data; therefore, it uses a post hoc approach to segmentation or Data Mining.

3.3 Relationship Marketing - the theoretical grounding of customer-centric marketing



Although the concept of maintaining and building strong customer relationships has been around since the first commercial exchange in ancient times (Grönroos, 1994a), the roots of the term Relationship Marketing can be traced back to the late 1970s and 1980s. Through the years, RM has also been called customer-focused management (Gummesson, 1994b), relationship management (Payne, 1995), customer management (Frow and Payne, 2009) and a variety of other related terms. Early uses of the term can be related to Berry (1983) and Hammarkvist, Hakansson and Mattsson (1982). Berry (1983) defined RM as marketing with the purposes of attracting, maintaining, and enhancing customer relationships. For their part, Hammarkvist et al. (1982, p. 10) presented a similar definition: "all activities by the firm to build, maintain and develop customer relations." However, its usage by Berry (1983) was used to critique the service marketing literature by highlighting the disproportionate attention given to attracting new customers versus retaining them. Berry advocated a switch from a transactional approach (also known as traditional, conventional, or 4Ps marketing) focused on customer attraction to a relational approach to marketing (where customer attraction was only one of the steps in a broader marketing process) (Berry and Gresham, 1986).

Though originally issued from the services sector, the practice rapidly extended to the other marketing sectors. In a review of its broad marketing applications and historical categorisation, Palmer et al. (2005) identify three main RM schools of thought: the Nordic School, IMP Group, and the Anglo-Australian School. The Nordic School originates from services marketing (Gummesson et al., 1997) and first appeared in the late 1970s in response to the limitations of the dominant Transactional Marketing paradigm. Its research studies the concept of service as a way of improving relationship quality, customer loyalty, and maximising customer value throughout the life-cycle (Grönroos, 1990). The Nordic School takes the view that marketing is the process of cross-functional interactions with the customer and not the strict application of the marketing mix to these interactions (Grönroos and Gummesson, 1985). The IMP (Industrial Marketing and Purchasing) School of RM originates from organisational relationships in business-to-business (B2B) markets, which consider long-term inter-organisational interactions as the unit of analysis (Ford, 1990; Hakansson, 1982). Finally, the Anglo-Australian School originates from the perspective of marketing being built on quality and service with customer relationships being achieved by the delivery of enhanced levels of value to customers through cross-functional collaboration with stakeholder/market groups and markets (Christopher, Payne and Ballantyne, 1991). Although all schools agree that the intent of RM is to attract, maintain, and enhance customer relationships, the main difference is in the manner in which RM is practiced (i.e. B2B versus B2C sectors, product versus service context, the degree of differentiation of products, the definition of a product strictly or as an experience, the complexity of value chains and the number of involved partners).

Although criticized from its inception by the services marketing perspective (and later by other schools), the Transactional Marketing paradigm's product-orientated philosophy and growth reliant weaknesses were initially hidden by the buoyant economy and consumer growth of the post-WWII era (Harker and Egan, 2006) and came into severe question during the seventies and eighties. Global competition intensification via both the increase of domestic and foreign firms (Gummesson, 1987) signified the beginning of an era of maturity and hyper-competition for consumer goods (Hammarkvist et al., 1982; Kotler, 1991). In tandem, mass markets began to fragment as customers became more sophisticated and demanded more personalised products and services (Christopher et al., 1991). As a result, the passive consumer competing for mass-produced products was replaced by the active consumer seeking personalised products among highly competitive firms. This integral change directly exposed the Transactional Marketing theory to the criticism of being reliant on overall market growth and not suited to a period of stable population and growth through share gains (Grönroos, 1991; Gummesson, 1991). These changes challenged the product-orientated philosophy that underpinned Transactional Marketing and created a strong case for the adoption of a customer-oriented approach to business (Grönroos, 1994b; Gummesson, 1997a). Though the change indicated a shifting orientation to serving the end customer better, it didn't lead to an immediate customer orientation. Firms first organised around markets and segments, and only slowly (and due to continued increases in competition) started defining smaller niche segments. Sheth, Sisodia, and Sharma (2000, p. 55) characterise this shift as part of a greater evolution from "mass marketing to segmented marketing in the twentieth century... toward customer-centric marketing in the next century" where "the marketing function seeks to fulfil the needs and wants of each individual customer."

Another particularly important shift in modern RM is that firms no longer need to achieve a minimum size or market penetration to adopt a customer orientation. The Internet and its e-commerce applications now enable most companies to deal directly with end customers from inception. What more, "companies small and large are able to achieve a high level of accessibility and establish a two-way information flow directly with end users almost immediately and at low cost. Serving huge numbers of customers efficiently and effectively is made possible by the automation of numerous administrative tasks. Every company is potentially a global player from the first day of its existence (subject to supply availability and fulfilment capabilities)" (Sheth and Sisodia, 1999, p. 74).

This paradigmatic change stemmed from the intrinsic underlying assumptions behind Transactional Marketing: the optimisation of the firm's profit function solely through the marketing mix, the de facto perspective that marketing objectives were met at the moment of exchange (Harker, 2006), and its related product-oriented philosophy (Grönroos 1994b; Gummesson, 1997). Sheth, Sisodia, and Sharma (2000) also identify more recent antecedents for the change: pressure to improve marketing productivity, increasingly diverse consumer and business markets, and enhanced technology capabilities and applications. RM for its part focuses on developing a long-term relationship with customers (Grönroos, 1991; Gummesson, 1987b) where “placing significant emphasis on customer retention will yield dividends” (Harker, 2006, p. 221).

Another factor contributing to the ongoing reconceptualisation of marketing as RM, is the increasingly prevalent perspective that services are integrated along with the sale of products and become critical to the delivery of enhanced customer relationships. This movement is part of a continued evolution toward a “customer orientation” and away from the “pure product” orientation. In recent literature, this phenomenon of delivering an enhanced “product-service” (P-S) offering is referred to as “product-service system”, “servitization” and “integrated solutions” (Baines et al., 2007; Davies, Brady and Hobday, 2006; Johnstone, Dainty and Wilkinson, 2008; Neely, 2008).

Table 7 presents both the Transactional and Relational concepts and their differences as identified by Grönroos (1996) and Brodie (2002). It shows that Transactional Marketing is mainly focused on the discrete economic transaction using a marketing mix approach to sell products pre-positioned to specific segments of passive customers. This takes place in a centralised marketing department concerned with static and ad hoc metrics. RM, for its part, extends its interest to acquiring the full share of the customer throughout the life cycle by using personal information for targeting purposes and interacting with active customers. This takes place in a decentralised fashion across the organisation as marketing is perceived as a shared task; furthermore, it is a lot more dynamic as information (including previously ad hoc measures) circulates continuously from the customer.

Table 7: Transactional versus Relational Marketing

Concept	Transactional Marketing	Relationship Marketing
1. The marketing variables	• Focusing on activities that attract customers • Traditional marketing mix variables: i.e. 4 Ps	• Focusing on activities that establish, maintain, enhance and terminate relationships • Broader set of personal organisational activities involving interaction communication
2. The offering	• Predetermined ‘positioning’ of physical product or service	• Resources including personnel, technology, know-how, customer time as well as a management system that governs these resources so that a need-satisfying solution emerges
3. Where marketing takes place in the organisation	• Marketing department	• Marketing as part of the tasks of most of the organisation’s units • Decentralised organisation of marketing

Concept	Transactional Marketing	Relationship Marketing
4. Marketing planning	• Marketing plan	• Marketing is planned in all plans that have an impact on the development of relationships
5. Choosing customers	• Market segmentation • Individual customer not identified	• Choice of customers based on individual customer information
6. Performance measurement	• Market share, ad hoc customer satisfaction and perceived quality measures	• Continuous information from the customer interface supported by market share and other ad hoc measures

Source: Brodie (2002) adapted from Grönroos (1996)

3.3.1 Relationship Marketing Continuums

In a period of transition from Transactional Marketing, RM was seen by many as the new paradigm of choice for the marketing discipline (Berry, 1983; Grönroos, 1994b; Gummesson, 1991). However, the overly simplistic view that transaction and relational marketing are mutually exclusive was replaced by a more tempered and integrative view that marketing existed on a continuum (on which firms should place themselves and/or their products) made up of transactional-based relationships on one end and relationship-based interactions on the other (Grönroos, 1994b, 1997; Webster, 1992). This continuum view is certainly appealing and allows academics to better integrate a multitude of previously divorced concepts originally introduced under the guise of relational strategy (but were ultimately largely transactional) into a relational continuum, some of these concepts included Loyalty Marketing and CRM, (Harker and Egan, 2006).

This perspective led many academics to pursue their own iterations of a continuum. Mattson (1997) proposes a model where a narrow version of RM may indeed be well represented by Transactional Marketing, whereas the extended version is represented by the relationship/network perspective of the IMP Group. Eggert and Stieff (1999) introduced the concept of the relationship as being either behavioural or attitudinal. The actual nomenclature used by Eggert and Stieff to describe the behavioural or attitudinal relationship modes are respectively 'limited' and 'extended' relationship marketing. In their review, the limited RM mode is characterised by a "series of transactions on behalf of the seller designed to achieve repeat transactions through a process of interaction with the buyer, typically driven by economic goals rather than including some of the wider aspects of the exchange such as customer satisfaction while the extended RM mode is characterised by a "motivation to achieve a state of mutual acknowledgement that the relationship exists" (Bliemel and Eggert, 1998, cited in Palmer et al., 2005, p. 318).

Another example of such a relational continuum is Maklan, Knox and Peppard's (2011) marketing capabilities framework. Shown in Table 8, the framework showcases that marketing capabilities exist along three dimensions: transactional, one-to-one, and networked relationships. Transactional Relationship Marketing identifies "distinctive market segments (*creating marketing knowledge*) where their product brands (*building brands*) can command a leading market share (*demand management*) based on volume or value of sales without necessarily building one-to-one consumer relationships; …one-to-one Relationship Marketing identifies existing and potential high-value consumers (*creating marketing knowledge*) to maximise the lifetime value of those relationships (*demand management*) by creating customised offers (*CRM*) that are delivered by all the business processes of the company (*building brands*)…" and a networked relationship that "identifies key individuals of a community (*creating marketing knowledge*) and works with them collaboratively (*CRM*) to meet the community's needs (*demand management*). Participants use the total capabilities of the network (*building brands*) to create the solutions they really need (*demand management*). In other words, the company becomes the leader of an orchestra of network participants, shaping the offer in conjunction with its consumers" (Maklan et al., 2011, p. 81). This framework allows managers to position their firms' capabilities on the Relationship Marketing continuum and then develop plans on how to develop them in order to best execute their CRM strategy.

Table 8: Relationship Marketing Framework

	Marketing Relationships		
	Transactional Relationship	**One-to-one relationship**	**Networked relationship**
Demand management	Profitable transactions, product/service innovation and channel partnerships	Maximise lifetime value of consumers across all channels	Co-create value with a network of consumers
Creating marketing knowledge	Market trends, segmentation and competitive offers	Individual consumers (or segments): needs, purchasing styles, profitability	Key network participants and shapers
Building brands	Product and service brand management		Encourage consumers to access a network's capabilities through your company offer
Customer relationship Management	Standardised offer	Customised and/or Negotiated offer	Self-managed offer (consumer networks co-create the offer

Source: Maklan, Knox and Peppardl. (2011)

In an effort to more clearly define the continuum from a practical standpoint, Coviello et al. (1997) provide a conceptual breakdown of the relational continuum as four distinct types of marketing. The four types of marketing practices defined by Coviello et al. include: transaction marketing, database marketing, interaction marketing, and network marketing. Transactional Marketing stands as both its own class and practice while database marketing, interaction, marketing, and network marketing are all practices that are different applications of Relational Marketing. Transactional Marketing involves a firm creating discrete economic transactions by attracting new buyers through use of the marketing mix. The transactions, though sometimes ongoing, are generally treated independently as single transactions in a formal and impersonal process with a largely passive customer. Database Marketing involves marketers using information and technology tools to impact the market transaction, in a way very similar to Peppers and Rogers' (1995) "one-to-one" marketing where marketers rely on IT to build customer relationships. The major difference between this practice and Transactional Marketing is the movement away from mass marketing approaches and toward more personalised exchanges. Nevertheless, the nature of exchanges is discrete and individualised via technology (not via interpersonal interactions) and very much still "to" the customer, rather than "with" the customer (Pepper and Rogers, 1995). Contrary to Database Marketing, Interaction Marketing is based on social interaction and exchange and is grounded on principles of trust, mutual cooperation and interdependence, commitment, joint planning and continuous value creation (Grönroos, 1990; Wilson and Jantrania, 1994). Finally, Network Marketing is mainly focused on the delivery of an enhanced customer experience via concerted inter-firm connectivity and network of relationships (Anderson et al., 1994). Though it also implies a high degree of deep social interactions between firms and individuals, Network Marketing is ultimately heavily reliant on a high degree of coordination and exchange between firms of a network. As such, it has often been referred to as a form of strategic orientation of sorts (Coviello et al., 1997, 2002).

Contrasting these different views reinforces the perspective that transactional and relational practices are not mutually exclusive or independent. In fact, most of the continuums presented demonstrate that practices either become more complex and integrated as a shift between the purely transactional and relational occurs (Eggert and Stief, 1999) or that transactional and relational marketing applications can be practiced concurrently (Coviello et al., 1997, 2002).

In order to operationalise these continuums, Grönroos (1991) identifies where different categories of goods and services belong across the relational continuum. For instance, consumer packaged goods (or Fast Moving Consumer Goods), followed by consumer durables and industrial products, are identified as being most appropriate for completely transactional applications. Services, on the other hand, are best suited to relational applications. Harker and Egan (2006, from Baker 1995, p. 224) note that from the consumer perspective "most consumer goods are mass produced, mass distributed, convenience goods" and that, as such, customers "are not looking for a relationship with the seller and a mix-management approach may well satisfy them better."

Indeed, it is likely that, as Grönroos (1991) indicates, certain types of firms may, by nature of industry of operation or business model, gravitate more toward the transactional or relational dimensions. However, firms certainly do operate across the continuum and as such transactional and relational practices are not mutually exclusive and can co-exist within a same organisation's product portfolio (Pels, Coviello and Brodie, 1999; Brodie Brookes and Little, 1997). The views indicate that Relationship and Transactional Marketing are not only practised concurrently but that firms adopt positions according to the individual strategic and operational contexts (of firms' themselves and/or their product lines/brands).

Nevertheless, the way the different continuums are conceptualised is quite different. I would qualify the continuums as falling into one of two categories: aggregate level views of RM or specific categorical views of RM. Furthermore, while some continuums overtly call out individual RM schools of thought as end points (Mattson, 1997), others are more general (Eggert and Stief, 1999) or integrative (Coviello et al., 1997, 2002; Maklan et al., 2011). This can be visualised as a simple two-by-two matrix (Figure 3).

Figure 3: RM Continuum Categorisation

Specific School Focused

	Mattson (1997)		
Aggregate	Eggert and Stief (1999)	Coviello et al. (1997, 2002) Maklan et al (2011)	**Specific**

General or Integrative

Eggert and Stief (1999) present a general representation of the continuum that effectively presents the relational world as existing on a broad continuum without any description of its intrinsic operationalisation other than the mention that one mode of RM combines the other in its application. This point of view is effective for strategically positioning RM practices but provides little normative support for practical uses or categorisation. Mattson's (1997) view is slightly more specific as it presents that continuum from a school of thought perspective with one aspect of the relationship being purely transactional and the other mainly driven by a networked approach to RM. This continuum is likely the most restrictive as it limits its applications solely to the IMP school and the related B2B sectors. Coviello et al. (2002) and Maklan et al. (2011) present specific and integrative conceptualisations that provide both a high level of separation between the transactional and relational marketing and a normative view of how different marketing applications can fit and be effectively categorised within this continuum. The main difference between these conceptualisations lies in the way marketing applications are differentiated. For example, Coviello et al. (2002) represent that Transactional Marketing is mainly concerned with mass marketing, where the shallow RM application of Database Marketing initiates a direct contact with the customer with segmented or personalised communications. In contrast, Maklan et al. (2011) include segmented targeting in Transactional Marketing, and then create a separate category for individualised contact with customers under

One-to-One Marketing. In addition, Coviello et al. describe two additional RM schools of thought in their continuum (Interaction and Networks) whereas Maklan et al. only approach it from the Network perspective.

In reviewing these competing presentations of RM, other than differences in how some terms are defined or categorised, there is no inherent disagreement on the existence of an array of practices of RM that span the transactional, operational or analytical, all the way to the strategic, integrated or individualised. Rather than take a position on which one is best or most relevant, I present my work in relation to each continuum, leaving the task of resolving taxonomical issues to other researchers.

Nevertheless, in reviewing the different applications of RM, this work focuses on the more transactional section of the relational continuum and clearly lies in Mattson's (1997) Transactional Marketing category. Its use of data to target individual customers for Database Marketing purposes indicates that the focus is not only transactional but also relational (albeit in a more transaction-frequency-building manner). Subsequently, it also falls into Eggert and Stieff's (1999) behavioural RM category as well as in the Database Marketing phase of RM in Coviello et al.'s (2002) continuum (as it moves beyond simple mass marketing of products to segmented or targeted communications). The work can also be categorised as falling into the one-to-one relationship phase of Maklan et al.'s (2011) continuum as it targets individual consumers using data analytics.

3.3.2 Choice Reduction in Relationship Marketing

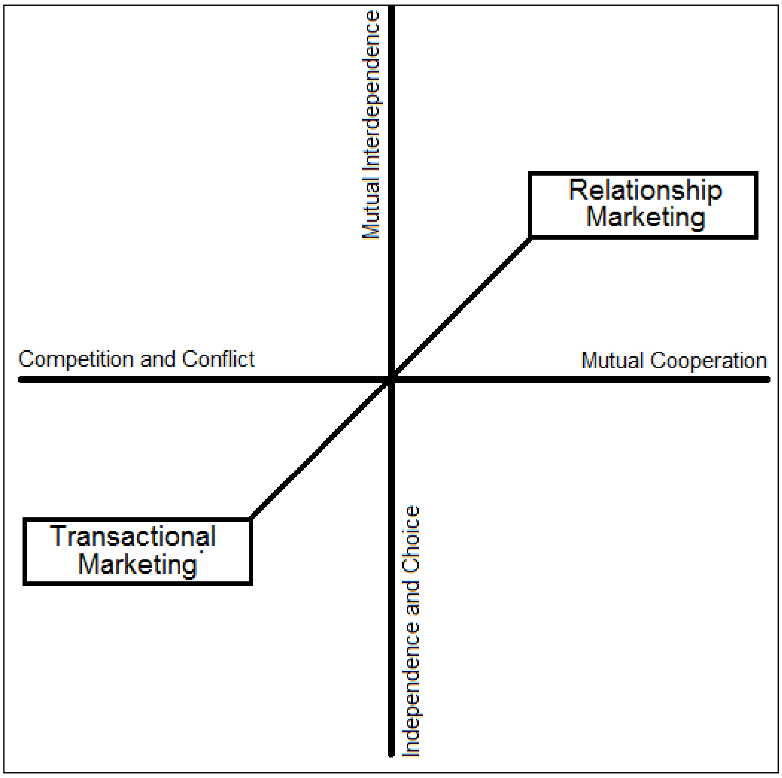
Sheth and Parvatiyar (1995) suggest that RM theory originates in consumer behaviour literature and that consumer choice reduction is the basic tenet of RM. Their contention is that consumers engage in relational behaviours because of personal, social, and institutional influences and that choice reduction and relational behaviours are pursued in order to "simplify their buying and consuming tasks, simplify information processing, reduce perceived risks, and maintain cognitive consistency and a state of psychological comfort" (p. 397). Other factors for engaging in relational behaviour also include family and/or social norms, peer group influences, religious beliefs, government directives, and marketing messaging policies. It is Sheth's contention that the willingness and ability of consumers and marketers to entertain mutual interdependent and cooperative relational exchanges leads to greater marketing efficiency.

A parallel, yet distinct, view is taken by proponents of Transaction Cost Economics. They assume that transactions have inherent costs related to researching purchase alternatives, negotiating these alternatives and enforcing contractual terms (if required). The existence of such transaction costs, though sometimes frictional, may result in choice set reduction as the bounded rationality of consumers leads to transactional outcomes that are not always optimal or assess alternatives completely (Williamson, 1985). When reviewing choice reduction research in grocery retail, several types of transaction costs have been identified:

(1) opportunity costs of time (travel and in-store shopping time), (2) transportation costs, (3) psychological costs (associated to undesirable aspects of the retail experience), (4) product unavailability adjustment costs, (5) search costs related to assortment differentials between retailers, (6) basket delivery costs, (7) physical (picking and packing) costs, and (8) other costs (waiting costs for order delivery, ability/inability to check product/produce prior to payment – mainly in the online or telephone order grocery environments) (Betancourt, 2005; Bell, Ho and Tang, 1998; Lewis, Singh and Fay, 2006).

Other relational aspects that can also lead to choice reduction include confidence and trust (Morgan and Hunt, 1994; Barnes, 1994; Bitner, 1995; Grönroos, 1990) and reciprocity and gratitude (Palmatier, 2008). Gwinner and Gremler and Bitner (1998, p. 102, citing Morgan and Hunt, 1994) effectively capture the trust dimension as follows: "trust, defined as confidence in the exchange partner's reliability and integrity, is a key mediating variable in relational exchanges. Although this sense of confidence and trust may be inextricably tied to the quality of the core service, it is likely perceived as an independent benefit of long-term relationships, particularly when customers perceive that there are comparable quality providers in the market."

Figure 4: Relational and Transactional Matrix



Source: Sheth and Parvatiyar (1995)

In addition to customers' tendency to want to reduce choice alternatives, the degree of customer involvement in the product decision is critically important. High-involvement products closely resemble service marketing relationships because of their greater relative value creation. Low involvement products, because of their inherent lower-value distribution, more closely resemble Transactional Marketing relationships (Sheth and Parvatiyar, 1995). This view is captured in Figure 4 showing a 2 x 2 matrix where Transactional Marketing is best suited to high competition and conflict and high independence and consumer choice contexts, whereas RM is best applied in contexts that are characterised by mutual cooperation within the value chain and mutual interdependence between the buyer and seller. Though the retail sale of FMCG products in the grocery sector may be a little more relationship oriented than the sales of individual FMCG products, it remains a highly competitive sector with a high degree of consumer independence and choice. This is supported by the increase in the number of options for offline and online purchasing of groceries (and most other consumer packaged goods categories, i.e. pharmacies) and the high intensity of retail competition characterised by informed consumers (Fox and Sethuraman, 2010). Most product categories can now be purchased across a wide variety of traditional (grocery, drug, and department stores), non-traditional (warehouse clubs, dollar store) retail formats, and online and offline retail options. A recent Food Marketing Institute study (2012, p. 23) recently reported that “primary stores lost share of weekly spending … with store trips going to other types of stores and formats. Warehouse clubs, drugstores, dollar stores, ethnic food stores and convenience stores all seeing increases in patronage while traditional supermarkets, supercenters, discount stores, and limited assortment stores saw decreases in visits.” In addition, the study also showed that although “46 percent of shoppers report that they never or rarely purchase groceries online…54 percent say they occasionally do” (FMI, 2012, p. 13).

This customer involvement view, supported by multiple authors (Leahy, 2011; Passingham, 1998; Grönroos, 1991) provides an alternative to the choice reduction literature that states that consumers reduce their available alternatives because of transaction costs (Williamson, 1985), confidence and trust (Gwinner, Gremler and Bitner, 1998; Morgan and Hunt, 1994; Barnes, 1994; Bitner, 1995; Grönroos, 1990), reciprocity and gratitude (Palmatier, 2008), and simplification of the consumption task (Sheth and Parvatiyar, 1995). This is not to say that low involvement decisions such as purchases in mass marketing channels may not indeed be well adapted to RM or contextualise this claim; however, Harker and Egan (2006) state that the notion of a strong relationship between a buyer and a seller in mass marketing contexts has yet to be established. They indicate that FMCG and mass marketing firms may very well operate mainly in the transactional side of the marketing continuum where not all customers want a relationship with suppliers. Nevertheless, Leahy (2011) does raise the possibility that there may exist segments of the market who may value some form of direct communication from FMCG organisations and that marketers should try to identify them instead of applying mass marketing strategies indiscriminately. Grönroos (2009, p. 351) further supports this view by stating: “In the case of Standardised consumer products, conventional approaches to marketing, such as the marketing mix management approach with its given set of marketing variables, can persuade customers to buy and also make them satisfied with the value created in their practices by what they have bought. The product variable, well managed and geared toward customer processes, may be enough to keep promises made." From an FMCG retail perspective, providing relevance in the form of targeted marketing offers may indeed provide consumers with the discounts or special offers they seek to affect their store patronage. In fact, promotions are stated as one of the top ten store selection criteria by consumers in the US (FMI, 2012).

The variety of conflicting views on the role that RM may play in lower engagement contexts is not inconsistent with the widely different engagement approaches adopted by retailers in the last decade. Frow and Payne (2009) show a slow migration of the FMCG and FMCG retail industry from a product-based selling to a customer-based form of RM. Though the product-based selling form was previously dominated by the FMCG industry, they demonstrate a renewed interest of this industry in engaging consumers and in the adoption of a customer orientation.

This thesis makes the assumption that even in low-involvement environments, customer choice can be reduced by one, or a combination, of the aforementioned factors. Practical examples of successful behavioural RM and CRM abound and include traditional retailers such as Tesco (Humby, Hunt and Phillips, 2003; Frow and Payne, 2009) and non-traditional ones such as Amazon.com (Lee-Kelley, Gilbert and Mannicom, 2003).

3.3.3 Research Positioning and Relationship Marketing Definition

Positioning the research requires selecting a definition of RM that accommodates its context without choosing a definition by convenience of fit with the topic, often referred to as single-issue definitions (Harker, 1999). Morgan and Hunt (1994, p. 21) identify this difficulty as they review definitions of the term and state that "extant definitions cover some kinds (of RM) but not others." For example, in the services marketing area, Berry (1983, p. 25) states, "RM is attracting, maintaining and -- in multi-service organisations -- enhancing customer relationships" and Berry and Parasuraman (1991, p. 133) propose that "RM concerns attracting, developing, and retaining customer relationships." In industrial marketing, Jackson (1985, p. 2) refers to RM as "marketing oriented toward strong, lasting relationships with individual accounts." Definitions similar to the preceding can be found in the areas of bank marketing, advertising, and business strategy (Beltramini and Pitta, 1991; Spekman and Johnston, 1986). This challenge emerges not only from the multiple schools of thought but also from the inherent multitude of RM forms identified by Coviello et al. (1997).

Table 9 below provides a list of compiled definitions. It should be reiterated that, as of the early 2000s, the CRM and RM literature started to blur, with CRM becoming the more dominant term of reference. As a result, I note that the evolution of RM definitions virtually stopped circa 2002 as academic interest shifted to RM's operationalisation as CRM (covered in the subsequent section). This is consistent with my earlier description of CRM as a more operational sub-domain of RM.

Table 9: Relationship Marketing Definitions

Definition	Year	Source
All activities by the firm to build, maintain and develop customer relationships	1982	Hammarkvist, Hakansson and Mattsson
... attracting, maintaining and - in multi-service organisations enhancing customer relationships	1983	Berry
... they want to build and maintain lasting - and profitable-relationships with their customers	1985	Jackson, in Harker 1999
... relates marketing to the development of long-term relationships with customers and other parties …	1990	Grönroos, in Harker 1999
… marketing can be viewed as the building, maintenance, and liquidation of networks and interactive relationships between the supplier and the customer, often with long-term implications. As a consequence marketing becomes first and foremost RM	1990	Gummesson, in Harker 1999
Establishing a relationship involves giving promises, maintaining a relationship is based on fulfilment of promises; and, finally, enhancing a relationship means that a new set of promises is given with the fulfilment of earlier promises as a prerequisite.	1991	Gummesson
Integrated efforts to identify, maintain, and build up a network with individual customers and to continuously strengthen the network for the mutual benefits of both the sides, through interactive, individualised and value-added contacts over a long period of time	1992	Shani and Chalasani
The process whereby the seller and the buyer join in a strong personal, professional, and mutually profitable relationship over time	1993	Pathmarajah, in Harker 1999
RM aims to identify and establish, maintain and enhance and, where necessarily terminate, relationships with customers and other stakeholders, at a profit, so that the objectives of all parties involved are met; and this is done by mutual exchange and fulfilment of promises	1994	Grönroos
All marketing activities directed toward establishing, developing, and maintaining successful relational exchanges". This is achieved by a "mutual exchange and fulfilment of promises	1994	Morgan and Hunt
RM is to identify and establish, maintain and enhance and when necessary also to terminate relationships with customers and other stakeholders, at a profit, so that the objectives of all parties are met, and that this is done by a mutual exchange and fulfilment of promises	1994	Grönroos, in Harker 1999
RM emphasises a long-term interactive relationship between the provider and the customer, and long-term profitability	1994	Gummesson in Harker 1999
… the understanding, explanation and management of the ongoing collaborative business relationship between suppliers and customers	1994	Cravens and Piercy, Sheth, in Harker 1999
...the process whereby a firm builds long term alliances with both prospective and current customers so that both buyer and seller work towards a common set of specified goals	1994	Evans and Laskin, in Harker 1999
RM (RM) is concerned with the development of long-term 'relationships with customers and other stakeholders, at a profit, so that the objectives of all parties are met	1996	Grönroos
Consumer RM seeks to establish long-term, committed, trusting and co-operative relationships with customers, characterised by openness, genuine concern for the delivery of high-quality goods and services, responsiveness to customer suggestions, fair dealing, and (crucially) the willingness to sacrifice short-term advantage for long-term gain. Suppliers attempt to create and strengthen lasting bonds with their customers; they shift from attempting to maximise profits on each individual transaction towards the establishment of solid, dependable and, above all, permanent relationships with the people they serve	1996A	Bennett, in Harker 1999
Consumer RM is the organisational development and maintenance of mutually rewarding relationships with customers achieved via the total integration of information and quality management systems, service support, business strategy and organisational mission in order to delight the customer and secure profitable lasting business	1996B	Bennett A, in Harker 1999
… fundamentally, RM involves the total fulfilment of all the promises given by the supplying organisation, the development of commitment and trust … and the establishment (where possible) of personal contacts and bonds between the customer and the firm's representatives; the eventual emergence of feelings within each party of mutual obligation, of having common goals, and of involvement with and empathy for the other side	1996C	Bennett B, in Harker 1999

Definition	Year	Source
RM is to establish, nurture and enhance ... relationships with customers and other partners, at a profit, so that the objective of the partners involved are met. This is achieved by a mutual exchange and fulfilment of promises	1996	Grönroos in Harker 1999
Marketing is to establish, maintain and enhance relationships with customers and other parties at a profit so that the objectives of the parties involved are met. This is done by a mutual exchange and fulfilment of promises	1996	Bennett C in Harker 1999
... is not directly aimed at immediate transactions but is based on building, supporting and extending customer relationships	1996	Matthyssens and Van den Bulte, in Harker 1999
RM refers to all marketing activities directed towards establishing, developing and maintaining successful relational exchanges	1996	Bennett, in Harker 1999
RM is the process of co-operating with customers to improve marketing productivity through efficiency and effectiveness	1996	Paravatiyar, in Harker 1999
The core of RM is relations, a maintenance of relations between the company and the actors in its micro-environment … The idea is first and foremost to create customer loyalty so that a stable, mutually profitable and long-term relationship is enhanced	1996	Ravald and Grönroos, in Harker 1999
Establishing, strengthening, and developing customer relations was stressed. The focus was on the profitable commercialisation of customer relationships, and the pursuit of individual and organisational objectives. In particular, long-term and enduring relationships with customers were stressed	1996	Takala and Uusitalo, in Harker 1999
RM is the process of planning, developing and nurturing a Relationship climate that will promote a dialogue between a firm and its customers which aims to imbue an understanding, confidence and respect of each others' capabilities and concerns when enacting their role in the market place and in society	1996	Tzokas and Saren, in Harker 1999
RM is concerned with the development and maintenance of mutually beneficial relationships with strategically significant markets	1996	Buttle, in Harker 1999
RM is an emergent disciplinary framework for creating, developing and sustaining exchanges of value between the parties involved, whereby exchange relationships evolve to provide continuous and stable links in the supply chain,	1997	Ballantyne, in Harker 1999
RM concerns attracting, developing, and retaining customer relations	1997	Berry and Parasuraman, in Harker 1999
RM has as its concern the dual focus of getting and keeping customers	1997	Christopher et al., 1991 in Harker 1999
RM is marketing seen as relationships, networks and interaction	1994, 1997	Gummesson, in Harker 1999
RM involves the identification, specification, initiation, maintenance and (where appropriate) dissolution of long-term relationships with key customers and other parties, through mutual exchange, fulfilment of promises and adherence to relationship norms in order to satisfy the objectives and enhance the experience of the parties concerned	1997	O'Malley et al., in Harker 1999
RM is designed to achieve repeat transactions through a process of interaction with the buyer, typically driven by economic goals rather than including some of the wider aspects of the exchange such as customer satisfaction	1999	Eggert and Stieff
RM theory provides the conceptual underpinning of one-to-one marketing since it emphasises enhanced customer service through knowledge of the customer, and deals with markets segmented to the level of the individual.	2000	Chaffey et al
The aim is to convert buyer behaviour and status from fleeting casual encounter, through marketing interventions, to committed relationships	2002	Varey
Relationship Marketing is the strategic management of relationships with all relevant stakeholders in order to achieve long term shareholder value. Critical tasks include the identification of relevant relational forms for different stakeholders and the segments and sub-groups within them and the optimal management of interactions within these stakeholder networks.	2009	Frow and Payne

Table 10: Count of Aims of Relationship Marketing

Aims of RM	Count (out of 35 definitions)
Building/attracting/acquiring/establishing/creating/initiating or getting customers	16
Retaining/maintaining/supporting or keeping customers	13
Developing/enhancing/extending or nurturing customers	16
Terminating or dissolving customer relationships	2

A count of the aims of RM stated in the aforementioned definitions (Table 10) indicates that aims are increasingly aligned among authors and include the: building, retaining, development, and potential termination of customer interactions, with the objective of building long-term relationships. This degree of consensus on the objectives of RM is not, however, translated into a consensus of the term's definition and deeper shared understanding in itself. Though definitions may seem similar, there are different interpretations of these definitions between RM schools of thought. In light of this disconnect, Harker (1999) attempted to apply a more empirical research methodology to define RM and bridge this gap. He did so by categorising definitions into seven conceptual categories shown in Table 11. He then scored definitions on a scale of one to seven depending on how many of the categories were addressed.

Table 11: Categorised Definitions of RM

Primary Construct	Other Common Constructs
Creation	Attracting, establish, getting
Development	Enhancing, strengthening, enhance
Maintenance	Sustaining, stable, keeping
Interactive	Exchange, mutually, co-operative
Long Term	Lasting, permanent, retaining
Emotional Content	Commitment, trust, promises
Output	Profitable, rewarding, efficiency

Source: Harker, 1999

This scoring system immediately excludes any single-issue or tactical definitions of the term and provides a good baseline with which to assess definitions since a conceptually "comprehensive" definition would address all seven categories. As identified by Harker (1999), narrow definitions of the term, while valid, are not comprehensive. They include Berry (1983) "who emphasises the 'beginnings' of marketing relationships, Christopher et al. (1991) who stress the importance of RM's 'customer keeping' orientation, and Paravatiyar (1996, in Harker 1999) who highlights the potential benefits of an RM strategy….while…other 'single-issue' definitions include Grönroos (1990), Gummesson (1994a, 1997), Cravens and Piercy (1994 in Harker 1999) and Sheth (1994 in Harker 1999)". In Harker's review, no definition achieved a perfect score; however, several definitions did score six out of a possible seven. These are listed next:

Consumer RM seeks to establish long-term, committed, trusting and co-operative relationships with customers, characterised by openness, genuine concern for the delivery of high-quality goods and services, responsiveness to customer suggestions, fair dealing, and (crucially) the willingness to sacrifice short-term advantage for long-term gain. Suppliers attempt to create and strengthen lasting bonds with their customers; they shift from attempting to maximise profits on each individual transaction towards the establishment of solid, dependable and, above all, permanent relationships with the people they serve. (Bennett, 1996A, cited in Harker, 1999)

Consumer RM is the organisational development and maintenance of mutually rewarding relationships with customers achieved via the total integration of information and quality management systems, service support, business strategy and organisational mission in order to delight the customer and secure profitable lasting business. (Bennett, 1996B, cited in Harker, 1999)

RM is to identify and establish, maintain and enhance and when necessary also to terminate relationships with customers and other stakeholders, at a profit, so that the objectives of all parties are met, and that this is done by a mutual exchange and fulfilment of promises. (Grönroos, 1994, cited in Harker, 1999)

RM is to establish, nurture and enhance ... relationships with customers and other partners, at a profit, so that the objective of the partners involved are met. This is achieved by a mutual exchange and fulfilment of promises. (Grönroos, 1996, cited in Harker, 1999)

The core of RM is relations, a maintenance of relations between the company and the actors in its micro-environment … The idea is first and foremost to create customer loyalty so that a stable, mutually profitable and long-term relationship is enhanced. (Ravald and Grönroos, 1996, cited in Harker, 1999)

RM involves the identification, specification, initiation, maintenance and (where appropriate) dissolution of long-term relationships with key customers and other parties, through mutual exchange, fulfilment of promises and adherence to relationship norms in order to satisfy the objectives and enhance the experience of the parties concerned. (O'Malley et al., 1997, cited in Harker, 1999)

"Bennett (1996) achieves this twice (A+B), although interestingly his most explicit attempt to define RM (C) dwells on its 'emotional' component, rather than being a general definition. Grönroos has sole authorship of two subtly different definitions and a third in partnership with Ravald (compare Grönroos 1989; 1990; 1994; 1997; and Ravald and Grönroos, 1996). The last of the half-dozen 'near-perfect' definitions collected here is from O'Malley et al. (1997)" (Harker, 1999, p.15).

Of the remaining definitions, four are rejected outright as they do not meet Harker's stringent criteria; Hammarkvist, Hakansson and Mattsson's (1982) definition, similar to Berry's (1983), is too narrow and restrictive; Eggert and Stieff's(1999) definition is too restrictive as it only focuses on the repeat transaction and the economic achievement of the relationship. Similarly Vary (2002) focuses mainly on customer conversion while dodging any reference to interactivity, emotional content or output. Finally, Chaffey et al. (2000) do not specify any of the customer lifecycle's objectives, and instead focus on the single dimension of service.

The one definition that remains is Shani and Chalasani's (1992). It effectively covers the creation, development, maintenance, interaction, and long-term definition aspects. It also touches on the emotional content and output facets by introducing the notions of mutual benefits and value-added contacts. One weakness in the definition lies in the lack of a description of "network"; this leaves the definition to interpretation and adds confusion given its potential linkages with the networking approach to RM espoused by the IMP school. It also touches on the significant concepts of integrated and interactive marketing, and value co-creation, whereby on one hand marketers "support customers' value creation by facilitating it with products that fit their processes in a value-creating way," while on the other hand marketers support "customers' value creation to achieve value fulfilment" (Grönroos, 2009, p. 354).

> *Integrated efforts to identify, maintain, and build up a network with individual customers and to continuously strengthen the network for the mutual benefits of both the sides, through interactive, individualised and value-added contacts over a long period of time. (Shani and Chalasani, 1992)*

An acceptance of Harker's methodology implies that one or more of the identified definitions could be selected as inclusive and comprehensive enough to be acceptable. However, some definitions remain different by virtue of underlining certain conceptual "aspects" of RM instead of taking a comprehensive view. "Those from Bennett (1996), while certainly being as broad as any of the others, must be considered descriptive rather than as explicit.…Grönroos (1997) seems superior to Grönroos (1989; 1990; 1994; 1997) and Ravald and Grönroos (1996) because of its developed awareness of the potential necessity for relationship "termination". The substantive terms culled from O'Malley et al. (1997, in Harker 1999) and Grönroos (1994; 1995) are very similar, with identical terms in four categories and the same category omitted from both ("temporal"). Of the two, the definition presented explicitly by Grönroos (1994; 1995) seems both more elegant and more succinct (Harker, 1999, p. 15).

Given the strength of Grönroos' definition and its high complementarity with Shani and Chalasani (1992), I have opted to take the strongest components of both definitions and craft a definition that contains all of the concepts required by Harkers' (1999) categorisation while also highlighting the integrated and coordinated marketing capabilities core to supporting the deployment of CRM strategy (Payne and Frow, 2005) defined in the next section. The definition follows:

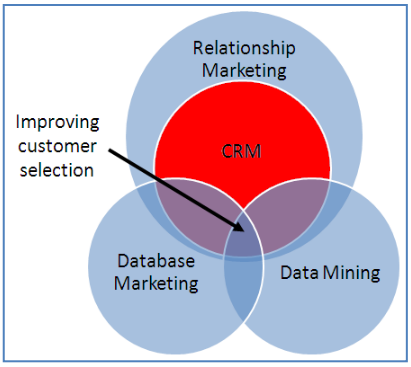
> *Integrated efforts to identify and establish, maintain and enhance and when necessary also terminate relationships with customers and other stakeholders, at a profit, and to continuously strengthen the relationship through mutual exchange and fulfilment of promises, through interactive, individualised and value-added contacts over a long*

period of time.

This definition, while being comprehensive enough to accommodate the field as a whole, also accommodates the context of the research:

- the focus on maintaining a customer relationship aligns with the promotional objective of achieving a response to the promotion;
- the use of a database for targeting align with the definition's individualised and value-added aspects; and
- the context of the promotional effort is part of a loyalty program and part of an ongoing set of (albeit discrete) communications aligns with the longitudinal nature of the definition.

3.4 CRM – the application of RM concepts with superior customer information



Richards and Jones (2008, p. 121) states that "research in marketing has been focused on relationships and building partnerships for some time (Crosby, Evans, & Cowles, 1990; Dwyer, Schurr and Oh, 1987; Morgan and Hunt, 1994), but it was not until technology became available to support managers in building relationships that CRM became an important part of this research (Chen and Popovich, 2003)." The acronym CRM originated in the IT community to describe the emerging technology solutions of managing and automating managing customers as individuals; the "information-enablement of RM" (Ryals and Payne 2001, p. 3). In the academic community, the terms CRM and RM are often stated as being interchangeable (Parvatiyar and Sheth, 2001). Dibb (2004, p. 114) states "despite the new terminology which is widely adopted across different sectors, CRM is simply a fresh perspective on RM ideas. Like RM, the costs of customer acquisition and desirability of customer retention are key features of CRM."

Research Positioning and CRM Definition

Early in the development of the CRM concept, Winer (2001, p. 91) stated that it means different things to different people: "CRM means direct e-mails. For others, it is mass customisation or developing products that fit individual customer's needs. For IT consultants, CRM translates into complicated technical jargon related to terms such as OLAP (on-line analytical processing) and CICs (customer interaction centres)." Payne and Frow's (2005, p. 167) research on the meaning of the term also highlights the lack of a common understanding of the term: "to some, it meant direct mail, a loyalty card scheme, or a database, whereas others envisioned it as a help desk or a call centre. Some said that it was about populating a data warehouse or undertaking data mining; others considered CRM an e-commerce solution, such as the use of a personalisation engine on the Internet or a relational database for SFA (sales force automation)." This lack of a common understanding is highlighted by the different definitions of CRM captured in Table 12.

Table 12: Definitions of CRM

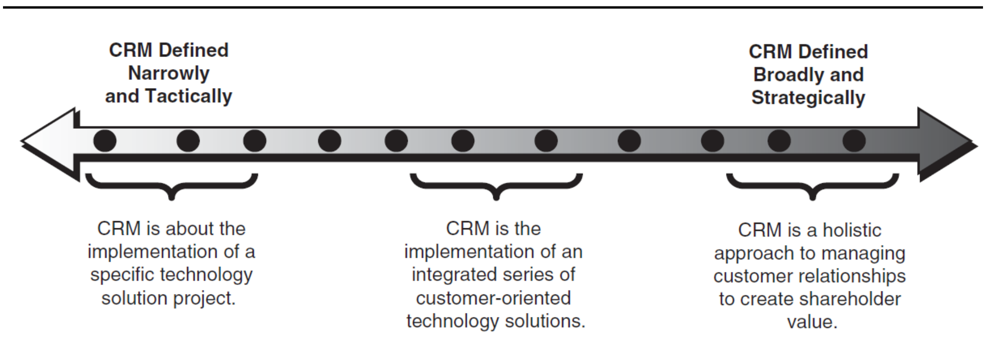
Definition	Year	Source
CRM is the marketing field aimed at establishing, developing, maintaining and enhancing relationships with customers	1990	Grönroos
CRM attempts to provide a strategic bridge between information technology and marketing strategies aimed at building long-term relationships and profitability. This requires "information-intensive strategies	1997	Glazer
CRM is data-driven marketing	1997	Kutner and Cripps
CRM can be viewed as an application of one-to-one marketing and RM, responding to an individual customer on the basis of what the customer says and what else is known about that customer	1999	Peppers, Rogers, and Dorf
CRM is a management approach that enables organisations to identify, attract, and increase retention of profitable customers by managing relationships with them	1999	Hobby
CRM involves using existing customer information to improve company profitability and customer service	1999	Couldwell
CRM is a macro level process that subsumes numerous subprocesses, such as prospect identification and customer knowledge creation.	1999	Srevastava, Shervani and Fahey
CRM is an enterprise approach to understanding and influencing customer behaviour through meaningful communications in order to improve customer acquisition, customer retention, customer loyalty, and customer profitability	2000	Swift
CRM includes numerous aspects, but the basic theme is for the company to become more customer-centric. Methods are primarily Web-based tools and Internet presence	2000	Gosney and Boehm
CRM is a relationship orientation, customer retention and superior customer value created through process management	2001	Ryals & Knox
CRM is a comprehensive strategy and process of acquiring, retaining, and partnering with selective customers to create superior value for the company and the customer. It involves the integration of marketing, sales, customer service, and the supply-chain functions of the organisation to achieve greater efficiencies and effectiveness in delivering customer value	2001	Parvatiyar & Sheth
CRM is an e-commerce application	2001	Khanna
CRM is a term for methodologies, technologies, and ecommerce capabilities used by companies to manage customer relationships	2001	Stone and Woodcock
CRM is a comprehensive strategy and process of acquiring, retaining, and partnering with selective customers to create superior value for the company and the customer	2001	Parvitiyar and Sheth
CRM is about the development and maintenance of long-term, mutually beneficial relationships with strategically significant customers	2001	Buttle
CRM is the core business strategy that integrates internal process and functions, and external networks, to create and deliver value to targeted customers, at a profit. It is grounded on high quality customer data and enabled by IT	2002	Knox, Maklan, Payne, Peppard and Ryals
CRM is an enterprise-wide integration of technologies working together , such as data warehouse, web site, intranet/extranet, phone support system accounting, sales, marketing, and production	2002	Bose
CRM aligns business processes with customer strategies to build customer loyalty and increase profits over time	2002	Rigby, Reichheld, & Schefter
CRM allows companies to gather customer data swiftly, identify the most valuable customers over time, and increase customer loyalty by providing customised products and services	2002	Rigby, Reichheld and Schefter
CRM is a philosophy of doing business that will affect the entire enterprise	2003	Newell
CRM is a customer-focused business strategy that aims to increase customer satisfaction and customer loyalty by offering a more responsive and customised services to each customer	2003	Croteau & Li

Definition (continued)	Year	Source
CRM is a strategy used to learn more about customer's needs and behaviours in order to develop stronger relationship with them	2003	Gupta & Lehmann
CRM is the strategic use of information, processes, technology, and people to manage the customer's relationship with a company across the whole customer life cycle	2003	Kincaid
A process to compile information that increases understanding of how to manage an organisation's relationships with its customers	2003	Zikmund et al.
Resources destined for relationship building and maintenance efforts should be allocated based on customers' lifetime value to the firm	2003	Ryals
CRM is leveraging technology to engage individual customers in a meaningful dialogue so that firms can customise their products and services to attract, develop, and retain customers	2003	Campbell
CRM is an enterprise-wide initiative that belongs in all areas of an organisation	2003	Singh and Agrawal
CRM is a business strategy designed to optimise profitability, revenue and customer satisfaction by organising the enterprise around customer segments, fostering customer-centric behaviours and implementing customer-centric processes	2003	Gartner Inc.
CRM is the strategic use of information, processes, technology, and people to manage the customer's relationship with a company across the whole customer life cycle.	2003	Kincaid
CRM is the practice of leveraging technology to engage individual customers in a meaningful dialogue so that firms can customise their products and services to attract, develop, and retain customers".	2003	Campbell
CRM refers to the idea that the most effective way to achieve loyalty is by proactively seeking to build and maintain long term relationships with customers	2004	Zablah Bellenger and Johnston, 2004
CRM is a technology or software solution that helps track data and information about customer to enable better customer service	2004	Peppers & Rogers
CRM is the business process in which customer equity (i.e., aggregate lifetime value of a firm's existing and potential customers) is continuously created, enhanced, and managed by interacting with customers through multiple channels	2004	Kumar and Reinartz
Management of mutually beneficial relationship(s) from the seller's perspective	2004	LaPlaca
A systematic process to manage customer relationship initiation, maintenance, and termination across all customer contact points in order to maximise the value of the relationship portfolio	2004	Reinartz, Krafft, & Hoyer
CRM is the outcome of the continuing evolution and integration of marketing ideas and newly available data, technologies, and organisational forms.	2005	Boulding, Staelin, Ehret, & Johnston
CRM is a strategic approach for systematically targeting, tracking, communicating, and transforming relevant customer data into actionable information on which strategic decision-making is based	2005	Karakostas, Kardaras and Papathanassion
CRM is a strategic approach that is concerned with creating improved shareholder value through the development of appropriate relationships with key customers and customer segments. CRM unites the potential of RM strategies and IT to create profitable, long-term relationships with customers and other key stakeholders. CRM provides enhanced opportunities to use data and information to both understand customers and to create value with them. This requires a cross functional integration of processes, people, operations, and marketing capabilities that is enabled through information, technology, and applications	2005	Payne & Frow
The process that identifies customers, creates customer knowledge, builds customer relationships, and shapes customers' perceptions of the firm and its products/ solutions	2006	The Sales Educators
CRM technology is a suite of information technology-based solutions designed to support the customer relationship management process	2005	Jayachandran, Sharma, Kaufman, ,Raman
CRM is a strategy used to learn more about customers' needs and behaviours in order to develop stronger relationships with them.	2007	Tarokh & Ghahremanloo
CRM is the complex of software and technologies, automating and performing business processes in the following areas: sales, marketing, service, and customer support	2008	Urbanskienė, Žostautienė & Chreptavičienė
CRM is the philosophy, policy and coordinating strategy mediated by a set of information technologies, which focuses on creating two way communications with customers so that firms have an intimate knowledge of their needs, wants, and buying patterns	2008	Lun et al.

Definition (continued)	Year	Source
CRM is a customer-cantered enterprise management mode, which discovers the customers' value and satisfies their requirements to realise the interaction between enterprise management and customers	2009	Huang & Wang
CRM is a key business strategy in which a firm needs to stay focused on the needs of its customers and must integrate a customer-oriented approach throughout the organisation	2009	Liou
CRM is an enabling technology for organisations to foster closer relationships with their customers	2009	Hsieh
CRM is a cross-functional strategic approach concerned with creating improved shareholder value through the development of appropriate relationships with key customers and customer segments. It typically involves identifying appropriate business and customer strategies, the acquisition and diffusion of customer knowledge, deciding appropriate segment granularity, managing the co-creation of customer value, developing integrated channel strategies and the intelligent use of data and technology solutions to create superior customer experiences	2009	Frow and Payne
A more expansive and holistic approach in developing sound and productive relationships with customers, while CRM technology, one of major components of CRM, has been defined as the information technology that is deployed for the specific purpose of managing customer relationships	2010	Chang, Park and Chaiy
CRM is a managerial strategy that helps organisations collect, Analyse, and manage customer related information through the use of information technology tools and techniques in order to satisfy customer needs and establish a long term and mutually beneficial relationship	2010	Hung et al.,
CRM involved the evolution and integration of marketing ideas and newly available data, technologies and organisational forms, and it concentrates on establishing, developing, and maintaining successful long-term relationships with well-chosen customers	2010	Ernst, Hoyer, Krafft, and Krieger

Payne and Frow (2006) posit that definitions of CRM abound and differ because CRM is not firmly entrenched in any one field but rather in many fields. Payne and Frow's (2004) review suggests that CRM can be viewed from three perspectives: "narrowly and tactically as a particular technology solution, wide-ranging technology, and customer centric" (Payne and Frow, 2005, p. 168). This CRM definition continuum spanning a narrow view of technology solutions on one end and a holistic approach to managing customer relationships on the other is portrayed in Figure 5.

Figure 5: CRM Continuum



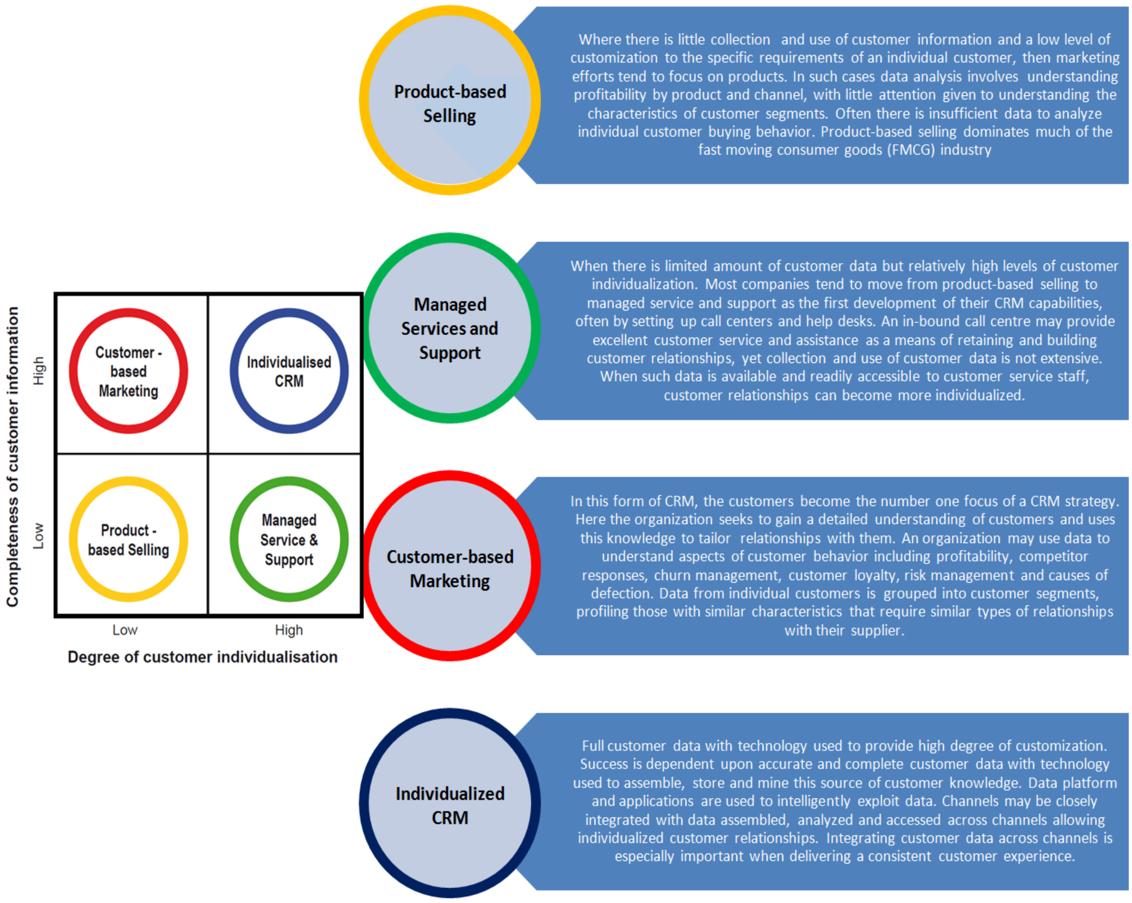
Source: Payne and Frow, 2005

Within this continuum, the technology view is characterised as being too narrow, limited in its applications and failing to embrace a sufficient customer or human perspective. Payne and Frow define CRM as "a strategic approach that is concerned with creating improved shareholder value through the development of appropriate relationships with key customers and customer segments." They continue to say "CRM unites the potential of RM strategies and

IT to create profitable, long-term relationships with customers and other key stakeholders. CRM provides enhanced opportunities to use data and information to both understand customers and co-create value with them. This requires a cross-functional integration of processes, people, operations, and marketing capabilities that is enabled through information, technology, and applications" (Payne and Frow, 2005, p. 407).

To operationalise their continuum, Frow and Payne (2009) present a CRM strategy matrix that considers both the strategic context of organisations and its implication for the evolution of CRM strategies. In their matrix, they present four alternatives to building customer relationships and identify migration paths that organisations can use to move between them. The CRM Strategic Matrix framework (Figure 6) shows 'Completeness of customer information' (how much information is held on customers) on the Y axis, and 'Degree of customer individualisation' (the extent to which the organisation can use whatever information it has on customers to give them individualised or customised service) on the X axis. The outcome is the four broad strategic forms of CRM that organisations can adopt: Product-based selling, Managed Services and support, Customer-based marketing, and Individualised CRM. Definitions of each form can be found in Figure 6 as well.

Figure 6: CRM Strategy Matrix & Definitions



Source: Frow and Payne, 2009

Compared to the RM continuums, Frow and Payne (2009) reframe the discussion into a more operational data asset and customer-centric strategy conversation. In this view, the transactional dimension is largely encompassed in the product-based selling and, similar to Mattson (1997), the RM dimension is imbedded, albeit to a different degree in all forms within the matrix. Other views classify CRM into many different categories. Richard and Jones (2008) identify two categories of CRM: strategic or operational. This definition categorisation resembles that of Leigh and Tanner (2004) who suggest that CRM is either analytical or operational. Zablah, Bellenger, and Johnston (2004) offer five perspectives of CRM: (1) a process, (2) a strategy, (3) a philosophy, (4) a capability, and (5) a technology. These are detailed in Table 13.

Table 13: Perspectives on CRM

Perspective	Description	Implications for CRM success	Representative conceptualisation
Process	Buyer– seller relationships develop over time and must evolve to persist	CRM success is contingent upon a firm's ability to detect and respond to evolving customer needs and preferences.	[CRM is concerned with] the creation and leveraging of linkages and relationships with external marketplace entities, especially channels and end users (Srivastava et al., 1999, p. 169).
Strategy	A customer's lifetime value determines the amount and kinds of resources that a firm invests in a particular relationship.	CRM success requires that firms continually assess and prioritise customer relationships based on their relative lifetime profitability	[CRM] enables companies to] invest in the customers that are (potentially) valuable for the company, but also minimise their investments in nonvaluable customers (Verhoef & Donkers, 2001, p. 189).
Philosophy	Customer retention (and hence profitability) is best achieved through a focus on relationship building and maintenance.	CRM success requires that firms be customer-centric and driven by an understanding of customers' changing needs.	CRM is not a discrete project—it is a business philosophy aimed at achieving customer centricity for the company (Hasan, 2003, p. 16).
Capability	Long-term, profitable relationships result only when firms are able to continuously adapt their behaviour towards individual customers.	CRM success is contingent upon a firm's possession of a set of tangible and intangible resources that afford it the flexibility to change its behaviour towards individual customers on an ongoing basis.	[CRM] means being willing and able to change your behaviour toward an individual customer based on what the customer tells you and what else you know about that customer (Peppers et al., 1999, p. 101).
Technology	Knowledge and interaction management technologies represent the key resources firms need to build long-term, profitable customer relationships.	CRM success is primarily driven by the functionality and user acceptance of the technology firms implement in an attempt to build customer knowledge and manage interactions.	CRM is the technology used to blend sales, marketing, and service information systems to build partnerships with customers (Shoemaker, 2001, p. 178).

Source: Zablah, Bellenger and Johnston, 2004

Jayachandran, Sharma, Kaufman and Raman (2005) provide a distinct separation between CRM technology (the suite of IT solutions) and the CRM process they support. Jayachandran et al.'s CRM process showcases how CRM technology acts as an enabler to the customer-oriented relational information process that helps bridge the gap between the CRM vision and organisation performance. This process is very similar to the KDD process (covered later in section 3.5.4). In fact, CRM has a lot of commonalities with Data Mining, with the KDD process potentially being a substitute for Jayachandran et al.'s (2005) relational information process.

More recent treatments from Chang, Park and Chaiy (2010) have depicted CRM as a more expansive and holistic approach in developing sound and productive relationships with customers, while CRM technology, one of major components of CRM, has been defined as the information technology that is deployed for the specific purpose of managing customer relationships (Chen and Popovich, 2003; Sin et al., 2005). In this context, CRM technology use equates to "the degree to which firms use supporting information technology" to manage customer relationships effectively (Reinartz et al., 2004, p. 296).

In order to better understand how definitions are framed in literature, I use the taxonomies of Payne and Frow (2005), Richard and Jones (2008) and Zablah, Bellenger and Johnston (2004), and based on judgement, classify each definition against taxonomical categories. The results are shown in Table 14 (with detailed results in Appendix 1). Results illustrate that the strategic view clearly dominates the IT, operational and analytical views. In Payne and Frow's (2005) and Richard and Jones' (2008) categorisations, the strategic view garners more than 60 percent of definitions' scope overall. This is not surprising given that the technology-centred views have suffered increasing criticism of inadequacy and lack of user-centricity, whereas the holistic approach has flourished (Maklan et al., 2011) given its strategic actionnability. This shift to a holistic approach to defining CRM was tested and accepted with practitioners by Payne and Frow (2005) and subsequently supported by Boulding et al. (2005). Observing Zablah, Bellenger and Johnston's (2004) five perspectives of CRM, strategy also emerges as the dominant perspective with a 44-percent representation of definitions; and the perspectives of process, philosophy and technology following with nearly 20 percent each.

Table 14: CRM Definition Taxonomical Categorisation

	Payne and Frow (2005)			Richard and Jones (2008)		Zablah, Bellenger, and Johnston (2004)				
Definition Type	**Narrow & Tactical Technology Solution**	**Wide-Ranging Technology**	**Customer Centric Strategy**	**Strategic**	**Operational/ Analytical**	**Process**	**Strategy**	**Philosophy**	**Capability**	**Technology**
COUNT 1990-2000	1	1	9	6	4	2	5	2	3	0
COUNT 2001-2005	2	5	20	18	8	5	10	5	2	6
COUNT 2006-2013	0	4	9	9	3	2	5	2	1	4
TOTAL COUNT	2	10	36	31	14	10	18	9	4	10

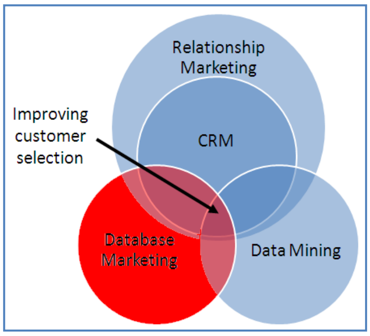
	Payne and Frow (2005)			Richard and Jones (2008) Tanner (2004)		Zablah, Bellenger, and Johnston (2004)				
% 1990-2000	9%	9%	82%	60%	40%	20%	50%	20%	30%	0%
% 2001-2005	7%	19%	74%	69%	31%	22%	43%	22%	9%	26%
% 2006-2013	0%	31%	69%	75%	25%	17%	42%	17%	8%	33%
TOTAL %	4%	21%	75%	69%	31%	24%	44%	22%	10%	24%

In a way, this outcome is very similar to the RM definition categorisation outcome as perspectives advanced by the authors each conceptualise CRM in a somewhat unique way and sometimes single-usage manner. The strategic/macroprocess perspective of CRM provides the best conceptual foundation for the CRM phenomenon as it “offers the most comprehensive, inclusive view of CRM (i.e., subsumes highly related sub-processes, such as interaction management) and, more importantly, explicitly acknowledges the process aspects of relationship development and maintenance (i.e., buyer–seller relationships develop over time and must evolve to perdure)” (Zablah et al., 2004, p. 479). I also contend that given the significantly high representation of Wide-Ranging Technology and Technology Applications across the table, the role of IT as an enabler of CRM must have a place in the definition of choice. The same can be said about the process dimension given the instrumentality in connecting IT with customer data in an organised and systematic manner. Definitions that position CRM as a philosophy were discarded given RM and CRM cannot both be the philosophy behind highly overlapping concepts. This perspective is supported by Ryals and Payne’s (2001) and Gummesson’s (2002b) views of RM as the broader and more dominant concept. The only definition that focuses on the strategic perspective while highlighting both the process and technology dimensions is the one provided by Payne and Frow (2005):

> *CRM is a strategic approach that is concerned with creating improved shareholder value through the development of appropriate relationships with key customers and customer segments. CRM unites the potential of RM strategies and IT to create profitable, long-term relationships with customers and other key stakeholders. CRM provides enhanced opportunities to use data and information to both understand customers and to create value with them. This requires a cross functional integration of processes, people, operations, and marketing capabilities that is enabled through information, technology, and applications.*

This is the definition that will be used for the thesis. This definition not only covers many of the central CRM facets but also supports the type of CRM form addressed by the empirical research: Customer-based marketing (Frow and Payne, 2009). In this form of CRM there is a formal use of advanced analytics and data to seek a detailed understanding of customer behaviour, namely response propensity of the loyal high value customer segments. A particularly relevant example of a grocer in the customer-based marketing quadrant quoted by Frow and Payne (2009, p. 19) is Tesco: “The company owes much of its success to the intelligent use of customer data. Tesco has gathered vast amounts of data on individual customers, largely through the Tesco Club Card. This loyalty card, which accumulates transaction data on customer purchases, allows the supermarket to refine aspects of the store layout, product positioning and presentation to enhance the customer buying experience." Not unlike the case in this research, Tesco also uses its detailed knowledge of customers and customer segments to increase the effectiveness of promotions and increase customer share and patronage (Humby, Hunt and Phillips, 2003). Other examples of increasingly analytical firms include Anheuser-Busch, Google, Wal-Mart, FedEx, and Harrah's Entertainment (Verhoef and Lemon, 2012).

3.5 Database Marketing - the technological enabler



Though sometimes used synonymously today with “direct mail”, the term was coined in the 1960s by Lester Wonderman and encompassed the concepts of consumer targeting and long-term value. Early applications of Database Marketing first occurred in the areas of direct marketing, then referred to as “mail-order”, “direct mail” or the “catalogue business”, and can be traced back to Aaron Montgomery Ward’s mail-order catalogue in 1872, with Sears, Roebuck following in 1887.

Prior to the 1950s, given the limited marketing channels available and limited applications from the computer science domain, Database Marketing was mainly referred to as direct marketing and the practice existed mainly in the mail-order sector. At the time, statistical applications such as the "12-month prune rule”, whereby inactive consumers (in the last twelve months) were excluded from a mailing list selection (Ross, 1992), and basic segmentations and cross-tabulations, were applied using cardboard stencils. Therefore, though the practice of direct marketing was much enhanced and informed by the advances in computing, it was born and actively practiced years before the computational and relational database booms. We can also glean from these early applications that the early adoption of segmentation practices occurred firstly in direct marketing before being adopted in the broader Database Marketing applications (this is discussed later).

The foundation for the revolutionary changes in direct marketing came in the 1960s with the application of computer data processing technology to customer files. Petrison, Blattberg and Wang (1993) describe the impact of computing on direct marketing as follows: “The integration of computers into the direct marketing industry is a watershed moment because nearly all of the developments in the business since then have been based on computer technology. In addition to replacing the cumbersome address plates and stencils previously used to record customer information, computer technology also led to the development of statistical modelling, merge/purge, personalised communications, credit cards, scanner technology, and dozens of other innovations” (Petrison et al., 1993, p. 30).

Though the practice of Database Marketing was much enhanced and informed by the advances in computing, it was not until the 1980s, the era of integrated and organised mainstream direct marketing programs, that non-direct-mail companies also began applying the concepts used in direct marketing to their own core businesses, thus extending the practice beyond the concept of mail order. This point is the clear break-off point between what became perceived as the channel-specific tactic of direct marketing and the data and insight practice and application of database marketing. Petrison et al. (1993, p. 119) capture this in their work and state: “Advances in computer and database technology quickly led to a new age of what eventually became known as ‘database marketing’. While expressing some of the same concepts that the term ‘direct marketing’ had been designed to address a decade earlier, the term ‘database marketing’ also suggested that the use of individualised consumer information did not have to be confined to the direct mail industry, but instead was also relevant to manufacturers of packaged and durable goods and to business-to-business companies.” It is these changes that were the driving force behind modern Database Marketing today and its incorporation of both traditional and online information into its practice (Zahay, Mason and Schibeowski, 2009).

3.5.1 Database Marketing Positioning and Definition

An early definition of Database Marketing states: "Database Marketing is an interactive approach to marketing communication, which uses the individually addressable marketing media and channels (such as mail, telephone and the sales force); to extend help to a company's target audience; to stimulate their demand; and to stay close to them by recording and keeping an electronic database memory of the customer, prospect and all commercial contacts, to help improve all future contacts” (Shaw and Stone, 1988, cited in Fletcher and Wright, 1995, p. 118).

The identification of sales as an addressable marketing channel illustrates that definitions of direct and Database Marketing may be influenced by the RM tradition and context in which the definition is used. Nevertheless, though digital or direct marketing channels aren’t overtly identified, the key concept in this definition is the use of addressable media and database assets to help define the targeting of the appropriate audience. This focus on targeting is reiterated by Dibb (2004, p. 113): given its concern with “improving efficiency by targeting marketing activities more efficiently” with the objectives of increasing “customer retention and loyalty, providing customer value and maximising customer lifetime value”.

Figure 7: Database Marketing Activities

Source: Blattberg, Kim and Neslin (2008)

More recent definitions of the term qualify the use of relational databases in the facilitation of database marketing. Hughes (1996a, p. 4, from Blattberg et al., 2008) quotes the National Centre for Database Marketing to define the domain as: “Managing a computerised relational database, in real-time, of comprehensive, up-to-date, relevant data on customers, inquiries, prospects and suspects, to identify our most responsive customers for the purpose of developing a high-quality, long-standing relationship of repeat business by developing predictive models which enable us to send desired messages at the right time, in the right form to the right people – all with the result of pleasing our customers, increasing our response rate per marketing dollar, lowering our cost per order, building our business and increasing our profits." Given its length, Blattberg et al. (2008, p. 4) also provide a shortened version of a definition: "the use of customer databases to enhance marketing productivity through more effective acquisition, retention and development of customers." In both these definitions, relational databases are the key triggers of the Database Marketing activity with the intent of improving a customer relationship and increasing marketing productivity. This practice is also described as "one-to-one" marketing since it uses deep transactional information processed through intelligent IT solutions to provide firms the ability to develop personal relationships with individual customers using "individually addressable and interactive media" (Pepper and Rogers, 1995, p. 48). The focus on deepening the customer relationship is also expressly supported and extended by Coviello et al. (1997) as they describe Database Marketing as a form of Relational Marketing. This view maintains that "the marketer relies on information technology (possibly in the form of a database or the Internet) to form a type of relationship, thus allowing firms to compete in a manner different from "mass marketing" (Coviello et al., 1997, p. 512). Coviello et al.'s continuum of RM placing Database Marketing as the closest "relative" of Transactional Marketing is supported by Iacobucci's (1994) view that characterises Database Marketing as a deeper and closer form of Transaction Marketing.

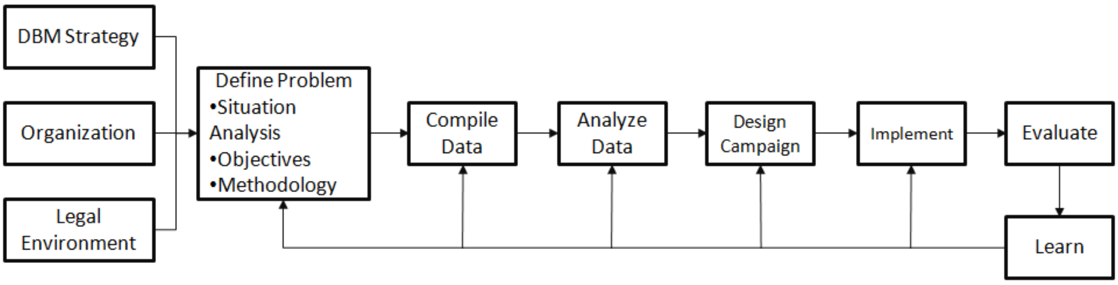
In Coviello et al.'s view, Database Marketing is a tool or technique used by practitioners to attract, develop and manage long-term relationships with a target group of customers by means of technology that normally does not involve interpersonal communications. This tactical and operational definition of Database Marketing is supported by Perrien and Ricard (1995) as they identify Database Marketing as a communication process aimed at improving a seller's product or service offering.

It is my view that, although Database Marketing is the cornerstone of RM, its applications and uses extend to the entire field of marketing starting with Transactional and Relationship Marketing and extending into other more loosely related areas, such as product assortment management, pricing and price sensitivity analysis, sales force enablement, and communication and product personalisation (Blattberg et al., 2008) detailed in part in Figure 7. Nevertheless, this research's interest is on the applications of Database Marketing in the context of RM, via the development of deeper and recurrent exchanges with existing customers.

3.5.2 The Database Marketing (DBM) Process

It is also important to note that Database Marketing is often supported by a process (Figure 8). This process hinges on the use of a database to contribute data that is converted into insights through the application of analytical techniques. To be consistent with the previous section's overview of Database Marketing activities, the process I've used as a reference is also from Blattberg et al. (2008) and illustrates how opportunities/problems that stem from the Database Marketing strategy and the organisation of the legal environment can be scoped, defined, and systematically addressed through a sequence of a data compilation, analysis, design, implementation and feedback loop.

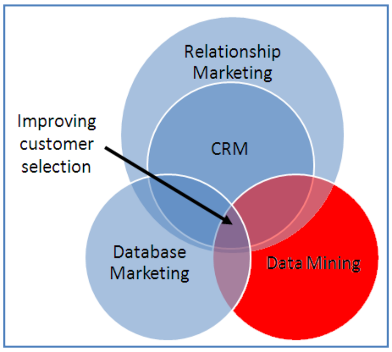
Figure 8: Database Marketing Process



Source: Blattberg, Kim and Neslin (2008)

Though many of the techniques used in the process are the realm of Data Mining (discussed later), it is important to point out that they are generally applied at the data analysis phase of the Database Marketing process. Some of the most tangible outputs from the data analysis phase of the Database Marketing process are customer segments, clusters, ranks and scores that are issued from the application of statistical techniques such as RFM analysis, cluster analysis, decision trees, predictive regression models and machine learning techniques. These methods will be discussed in detail in later sections.

3.6 Data Mining – the techniques and processes to create actionable insights



If Database Marketing is the usage of database assets to form a type of relationship by using derived insights, then Data Mining comprises the tools and process by which these insights are extracted. This is particularly important as the sizes of database assets increase, and is a critical contributor to the success in generating value for what academics are now openly referring to as Big Data.

3.6.1 Big Data

First used in an academic context by Weiss and Indurkhya (1998, p. xi), Big Data was defined and qualified as "very large collections of data ... compiled into centralised data warehouses, allowing analysts to make use of powerful methods to examine data more comprehensively." In theory, Big Data can lead to much stronger conclusions for data-mining applications, but in practice many difficulties arise. More recent definitions of the term qualify Big Data as being issued from non-traditional sources such as: interconnected real-time supply chain data transmitting information on market demand, weather patterns, unstructured digital channels and clickstream data from the Web, social media content (tweets, blogs, product reviews, Facebook wall posts and profiles, YouTube video posts and comments), smart phone applications and other Internet-based gadgets and widgets (LaValle, Lesser, Shockley, Hopkins, Kruschwitz, 2011). But Big Data also encompasses everything from "call-centre voice data to genomic and proteomic data from biological research and medicine" (Davenport, Barth, and Bean, 2012, p. 22). This growing flow of data sources has led the SAS institute to define Big Data as "when volume, velocity and variety of data exceed an organisation's storage or compute capacity for accurate and timely decision-making" (SAS, 2011).

Much like CRM, the term Big Data is used both tactically and strategically. The use of the term in IT refers to software and hardware applications that can manage the increasing flows of information. Strategically, Big Data describes the phenomenon of the growing data asset and the potential value it implies. This is increasingly supported in practitioner literature which describes Big Data as a business concept evolving in parallel with the technologies underlying its enablement, a "meme and a marketing term ... but also shorthand for advancing trends in technology that open the door to a new approach to understanding the world and making decisions" (Lohr, 2012, p. SR1). Brown, Chui and Manyika (2011) identify the key elements of Big Data as: (1) the collection of data across business units, from partners and customers, (2) the IT infrastructure that allows for the integration, scaling and scalability of information, and (3) the experiments, algorithms, and analytics that allow for the information to be transformed into actionable insight. Davenport et al. (2012) state that though the term Big Data often refers to "smarter, more insightful data analysis...Big Data is really much more than that. Indeed, companies that learn to take advantage of Big Data will use real-time information from sensors, radio frequency identification and other identifying devices to understand their business environments at a more granular level, to create new products and services, and to respond to changes in usage patterns as they occur. A more conceptually friendly way of distinguishing Big Data from standard data and data analysis is via three unique characteristics it embodies: Volume, Velocity and Variety (McAfee and Brynjolfsson, 2012). Volume is characterised with the capacity of organisations to work with many petabytes (one-quadrillion bytes) of data in a dataset. Velocity has to do with the speed of data creation. Unlike typical data management, Big Data is real-time or near real-time and allows marketers to make inferences on data before any transactional data is even recorded (i.e. sales forecasting based on traffic patterns). Finally, Variety refers to the ever-increasing array of non-traditional data that is sourced and collected on consumer behaviour.

Big Data, whether restricted to the applications of advanced analytics or the extended process of generating real-time actionable information from data, requires the tools and processes of Data Mining to generate meaning and value. The next section delves into some of these ideas. However, it should be noted that the full mining of Big Data may indeed require a specialised set of tools and/or hardware (e.g. advanced Cloud-based management). These will not be covered specifically in this thesis. The Data Mining techniques referred in the subsequent section are certainly applicable (and widely applied) to Big Data today; however, it should be noted that most of the articles used as references are more specific to the world of traditional Database Marketing and data analytics.

3.6.2 Data Mining Positioning and Definition

In the last twenty years, definitions of Data Mining have been relatively closely knit, albeit they have existed within a continuum defined by any form of data analysis on one end and tighter statistical approaches on the other (Kamber and Han, 2001). In addition, definitions of the term range from being a process of searching and Analysing large data stores (Frawley, 1992) to the pure application of statistical techniques on such data stores (Berry and Linoff, 1997; Han and Kamber, 2001; Fayyad and Uthurusamy, 2002).

Shankar and Winer (2006, p. 2) define Data Mining as “the process of automatically searching large volumes of data for patterns such as association rules and robust relationships. It also is defined as the non-trivial extraction of implicit, previously unknown, and potentially useful information from data (Frawley, Piatetsky-Shapiro and Matheus, 1992). Data Mining involves computational techniques drawn from disciplines such as computer science and statistics.” For their part, Han and Kamber (2001, p. 5) define the Data Mining founding disciplines as follows: “Data Mining involves an integration of techniques from multiple disciplines such as database technology, statistics, machine learning, high-performance computing, pattern recognition, neural networks, data visualisation, information retrieval, image and signal processing, and spatial data analysis.” Though many other definitions abound as will be shown later, all definitions albeit slightly different, embody the need for a combined use of computational power and mathematical and statistical applications for the derivation of meaningful patterns and rules in raw data sets. This domain is of particular interest given most of the statistical approaches and machine-learning techniques applied for the selection of particular customers for means of Database Marketing take their roots in this literature.

Berry and Linoff's (1997, p. 5) definition became the more widely accepted (Thanakorn, 2003): "the exploration and analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns and rules." This definition is also supported by Rahman (2008). A later definition by Fayyad and Uthurusamy (2002, p. 28) extends Berry and Linoff's definition to "the identification of interesting structures in data. Structures designate patterns, statistical or predictive models of the data, and relationships among parts of the data."

Given the thesis' interest in customer selection techniques in a data-driven interactive process of direct marketing, Fayyad and Uthurusamy's (2002) definition will be used for reference as it explicitly extends the practice of Data Mining to include not only general patterns but "statistical or predictive models," and other types of "relationships among parts of the data."

Payne and Frow (2005) acknowledge that a data-driven approach to integrating all customer contact channels is at the very centre of CRM strategy. As a result, the role of Data Mining extends to informing both RM, and enabling database marketing. It is important to reiterate that whether used as a process or a technique, data mining's unit of interest is either the techniques applied or the process under which these techniques are applied. Thus, Database Marketing involves transforming a database and its information into business actions using Data Mining applications.

3.6.3 Data Mining Tasks

The two primary tasks that Data Mining aims to achieve are prediction and description (Wedel and Kamakura, 2000). "Prediction involves using some variables or fields in the database to predict unknown or future values of other variables of interest, and description focuses on finding human-interpretable patterns describing the data" (Fayyad et al., 1996a, p. 44). Although the boundary between prediction and description is grey, it remains useful in understanding the goals of the Data Mining application. To make the constructs applicable, a number of researchers subdivide them into specific tasks or functions. Three examples of task subdivisions sourced from frequently cited articles are illustrated in Table 15. To facilitate interpretation, tasks from different sources that are related to one another are positioned side-by side.

The data presented in Table 15 illustrates the close alignment between definitions of Data Mining task categories between scholars. However, Hand, Mannila and Smyth (2001) include a task category specifically related to text and image analysis while Han and Kamber (2006) explicitly address the dimensionality of time with the addition of the evolution analysis task.

Given that Hand, Mannila and Smyth's (2001) definition of retrieval by content is mainly driven by the identification of patterns, it could be argued that it aligns with the dependency of modelling and mining frequent patterns, association and correlation tasks presented by Fayyad, Piatetsky-Shapiro and Smyth (1996) and Han and Kamber (2006). The task of evolution analysis, though it may include elements of "characterisation, discrimination, association and correlation analysis, classification, prediction or clustering, is distinct and includes time-series analysis, sequence or periodicity pattern matching, similarity based analysis" (Han and Kamber, 2006, p. 27) as well as elements of survival analysis (Tabachnick et al., 2001). Given the holistic nature and the addition of the time dimension, the Data Mining task categorisation of Han and Kamber (2006) is the definition adopted for this thesis. These tasks are used to categorise literature according to the tasks met by research.

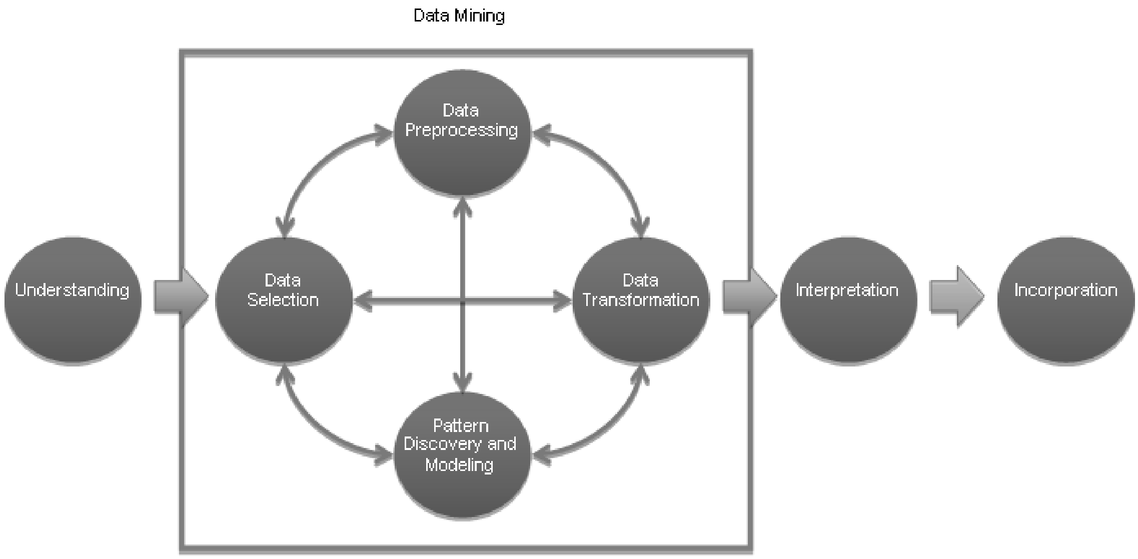
Table 15: Data Mining Tasks

Fayyad, Piatetsky-Shapiro and Smyth (1996) Kantardzic (2002)	Hand, Mannila and Smyth (2001)	Han and Kamber (2006)
Summarisation Methods for finding compact descriptions for a subset of data	**Exploratory Data Analysis** Interactive and visual techniques allowing for the exploration of data	**Concept/Class Description: Characterisation and Discrimination** Association of data with classes or concepts
Classification Learning function that maps (classifies) a data item into one of several predefined classes	**Predictive Modelling: Classification and Regression** Models that permit the value of one variable to be predicted from the known values of another variable	**Classification and Prediction** Process of finding a model (or function) that describes and distinguishes data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown based on a training set whose class label is known.
Regression Discovery of a predictive learning function that maps a data item to a real-value prediction variable		
Clustering Common descriptive task where one seeks to identify a finite set of categories or clusters to describe the data	**Descriptive Modelling** Models allowing for the description of subsets within data and the relationship between variables	**Cluster Analysis** Analysis of data objects without consulting a known class label
Dependency modelling Finding a model that describes significant dependencies between variables.	**Rule and Pattern Discovery** Task of finding combination of items that occur frequently or that emerge as outliers	**Mining Frequent Patterns, Associations & Correlations** Identification of patterns that occur frequently in data
Change and deviation detection discovering the most significant changes in the data from previously measured or normative values		**Outlier Analysis** Identification of data objects that do not comply with the general behaviour or model of the data
	Retrieval by Content Discovery of patterns of interest mostly in texts and image data sets	
		Evolution Analysis Description and modelling of regularities or trends for objects whose behaviour changes over time

3.6.4 Phases of Data Mining

Fayyad et al. (1996b) identify five phases of Data Mining that they refer to as Knowledge Discovery in Databases (KDD). These five phases consist of: (1) Selection, (2) Preprocessing, (3) Transformation, (4) Data Mining, and (5) Interpretation/Evaluation. Though seemingly static, the process is interactive and iterative, with many user-driven decisions along the way (Brachman and Anand, 1996; Kantardzic, 2002). Fayyad et al. (1996b) also identify two additional phases to the process given that the standard KDD phases are typically preceded by the understanding of the application domain and by the evaluation of prior knowledge and goals of end-users, and usually followed with the incorporation of knowledge into the system via decisions or actions. The explicit incorporation of these two additional phases thus increases the total number of phases to seven. In addition, I have renamed the data-mining phase to Pattern Discovery and Modelling because the definition of Data Mining utilised in this review includes steps two through five of the KDD process (Figure 9).

Figure 9: KDD Process



Source: Adapted from Fayyad, Piatetsky-Shapiro and Smyth (1996).

These additions and the renaming of the data-mining phase to pattern discovery and modelling are aligned with Kantardzic's (2002) data-mining process while the continued segregation of the data preprocessing and transformation phases is consistent with Han's (2006) interpretation. Definitions of phases are presented in Table 15. These phases are used to categorise literature in the systematic review. Distinguishing these phases is important because different techniques are applied at different phases and only by methodically identifying individual research by phase can researchers improve research and practice outcomes in an unbiased and systematic fashion.

Table 16: Expanded KDD Process

Phase	Definition
1: Understanding	Understanding of the application domain and prior knowledge and goals of end-users
2: Selection	Creating a target data set or focusing on a subset of variables, on which discovery is performed
3: Pre-processing	Data cleaning and pre processing in order to obtain consistent and significant data and/or variables. This phase includes outlier detection and scaling, encoding and selection features
4: Transformation	Transforming data using "dimensionality reduction or transformation methods"
5: Pattern Discovery & Modelling	Search for patterns of interest linking the data to the discovery objectives;
6: Interpretation / Evaluation	Interpretation and evaluation of identified patterns to assess actionnability
7: Incorporation	Knowledge incorporation into the system via decisions or actions

Each phase uses somewhat specific techniques and, in some cases, more than one phase is covered in research. Looking back at the Database Marketing process, it is easy to visualise how the data analysis step is fulfilled by the KDD or Data Mining process. Much like the segmentation process can be attached to part of the Database Marketing process, the KDD process matches up in a similar fashion. More specifically, the understanding step can be related to problem definition, the broader Data Mining steps (data selection, preprocessing, transformation and pattern discovery and modelling) and interpretation steps can all be related to data compilation and analysis and, finally, the incorporation step can take place in the campaign design and implementation of the Database Marketing process.

In terms of the intensity of research conducted on each of the Data Mining phases, most research has focused on the process of pattern discovery and modelling (Shaw, Tan, Subramaniam and Welge, 2001).

3.6.4.1 Feature Reduction

According to Kantardzic (2003), data reduction/transformation is applied mainly for the reduction or transformation of three elements of Data Mining models: (1) features, (2) cases, or (3) values. In short, a researcher can eliminate or change an entire independent variable (or a feature), a case (or line item, observation, or, in this study, a customer's transactional record), and/or a value (the contents of an observation or one of the transactional record's individual values).

Kantardzic (2002) states that two feature reduction approaches exist: (1) a bottom-up empirical approach that starts with the data itself and, based on a full set of variables, assesses which ones to remove or transform and (2) the top-down heuristic approach that starts with the researcher making a decision of which variables to include in the modelling exercise based on some heuristic criteria informed by literature. Han and Kamber (2001, pp. 119-120) support this view and further expand the heuristic approaches into three types:

1. "**Stepwise forward selection**: The procedure starts with an empty set of attributes. The best of the original attributes is determined and added to the set. At each subsequent iteration or step, the best of the remaining original attributes is added to the set.
2. **Stepwise backward elimination**: The procedure starts with the full set of attributes. At each step, it removes the worst attributes remaining in the set.
3. **Combination of forward selection and backward elimination**: The stepwise forward selection and backward elimination methods can be combined so that, at each step, the procedure selects the best attribute and removes the worst from among remaining attributes."

To a certain extent, both approaches can be applied simultaneously and it is often hard to determine when authors have applied one or the other. Though the actual distribution of all variable sets is unknown, this approach assumes that a normal distribution exists in every case and that variables are independent from one another. The most popular statistical methods for empirical dimensionality reduction are PCA and PLS (Lee et al., 2011). Limited studies in the Database Marketing literature were found to compare PLS and PCA; however, literature in the field of natural sciences has been shown that PLS in high dimensionality environments produces improvements in class separation and discrimination, and requires less transformation and variable compression to provide the same level of predictive accuracy versus PCA (Kemsley, 1996). Other commonly used methods for purposes of data reduction and transformation include: stepwise logistic regression, decision trees (CART, CHAID, C4.5, ID3), neural networks and other machine learning techniques (Weiss and Kulikowski, 1991). Decision tree algorithms, though originally intended for classification, have also been found to be very successful in data reduction given the approach's simplicity (all attributes that don't show up on a tree's output are classified as irrelevant) and its higher degree of accuracy (given it optimises for both a modelled output and the removal of noisy or insignificant attributes or variables (Han and Kamber, 2001). Finally, sliced inversed regression is a less used but not uncommon method to reduce variable complexity (Li, 2010); however, it is more prevalent in the natural science fields. Definitions of these data reduction methods can be found in Kantardzic (2002), Hand and Manila (1998) and Li (2010).

Other ways of reducing features consist of compressing existing features into a more limited yet behaviourally significant subset. One of the most famous and simplest partitional algorithms to solving the principal point problem is the iterative K-means clustering algorithm (Reutterer, Mild, Natter and Taudes, 2006; Jain, 2010). As a centre-based clustering algorithm, K-means solution can sometimes converge to a terminal minimum that is far from the global minimum (Hamerly and Elkan, 2002). Other alternatives to K-means that have been suggested recently include: K-medians, K-midranges, K-modes, transfer algorithms, K-harmonic means, symmetry-based K-means, Fuzzy K-means, Global K-means and hybrid methods (Steinley, 2006). Other than simply eliminating predictors, alternative methods of dealing with the issue of multicollinearity problems include: ridge regression and principal component regression (PCR) (Abdi, 2010).

3.6.4.2 Value Reduction

Value reduction and transformation, for its part, aims to decrease or modify the values of a specific variable or feature. The benefits are dataset simplification, ease of understanding, and interpretability of data-mining results. Similar to feature reduction and transformation, this exercise can be conducted manually in an a-priori and heuristic fashion. For example, cut-off levels for middle-aged individuals can be defined based on a researcher's knowledge of a specific population distribution and/or previous knowledge of how middle-aged is interpreted by responder according to the goals of a study. However, empirical applications to value discretization do exist with an abundance of value reduction techniques found in literature including unsupervised and supervised algorithms. Unsupervised algorithms include the equal-width interval, equal-frequency interval k-mean clustering or unsupervised MCC algorithms (Dougherty, Kohavi and Sahami, 1995). K-means is a popular method of cluster analysis that divides n observations into k clusters according to the observations' mean proximity. Unsupervised MCC algorithms use an unsupervised clustering approach to dividing borders or boundaries that provide the highest degree of contrast based on a given contrast function (Dougherty et al., 1995). Supervised algorithms include CAIM (Kurgan and Cios, 2004), ChiSplit (Bertier and Bouroche, 1981), ChiMerge (Kerber, 1992; Mejia-Lavalle, Arroyo-Figueroa and Morales, 2009; Hu and Cercone, 2001), Chi2 (Liu and Setiono, 1997; Arroyo-Figueroa and Morales, 2009; Tay and Shen, 2002), Khiops (Boullé, 2004), CADD (Ching, Wong and Chan, 1995), maximum entropy (Kumar and Zhang, 2007; Le and Satoh, 2007) or 1R (Holte, 1993) and most recently Amiva (Gonzales, Cuberos, Velasco and Ortega, 2009). A complete list of techniques can be found in Dougherty et al. (1995). Most of these supervised algorithms are automated and evaluate intervals within a given variable and use X^2 statistics to determine whether adjacent intervals exhibit similarities and should be merged, for example. The Chi2, CAIM and Amiva algorithms are essentially sequential modifications made to the ChiMerge algorithms or its followers (Chi2) that further automate the reduction process by introducing different approaches to evaluating the inconsistency rate and stopping criterion either manually or automatically (Shen, 2002).

3.6.4.3 Case Reduction

Finally, case reduction concerns itself with the correction of data that is usually sourced in a secondary fashion by researchers (and is thus of more or less questionable quality). Examples of standard forms of case reduction include the exclusion of outliers and the elimination of cases with missing data. Some of the most utilised methods for case reduction related to outliers, commonly referred to as exceptional cases, include: the standard deviation method, the Z-Score or Modified Z-Score, the visual use of Boxplot or Adjusted Boxplot, and the Median Absolute Deviation (MAD). Choice of preferred technique depends on a variety of factors including the presence of extreme values in the data and preference for visual versus numerical representation (Seo, 2002).

It is important to note that case reduction techniques take their importance when the extended or complete samples (i.e. a full population) of data cannot be managed by a data-mining technique or due to computing limitations related to file size. In such cases, samples must be managed to best represent a population since a sampling process inherently generates sampling errors. However, if all extended samples can be managed, there is no technical or theoretical reason for case reduction since sampling errors are eliminated when a full population is selected (Kantardzic, 2002). Definitions of additional data transformation methods can be found in Kantardzic (2002), Hand and Manila (1998) and Li (2010).

3.6.4.4 Impact on Promotional Effectiveness

From a promotional effectiveness moderation perspective, all steps in the KDD process have been shown to provide improvement. Crone, Lessmann and Stahlbock (2006), for example, show the direct impact of data preprocessing on predictive power and computational efficiency. Abbott, Stone and Buttle (2001), for their part, showcase the gains from pre-processing are essential to successful CRM performance as they enable CRM strategies to be more effective and efficient – regardless of the campaign tactics employed.

For its part, data reduction is the most-often stated application in pre-processing (Deichmann et al., 2002). It specifically addresses the issue that adding too many predictor variables increases the variance of the predicted values. “Models having too many parameters can overfit the data, where the stochastic variation in the data is modelled as well as the underlying relationship between the dependent and predictor variables. Conversely, having too few parameters will produce biased estimates of the dependent variable” (Malthouse, 1999, p. 14). Thus, by reducing the number of independent predictive variables in a model, the variance of the dependent predicted values and the chance of overfitting decrease.

Moderators in the transformation phase of the KDD process include Shih and Liu’s (2003) application of an AHP technique to appropriately weigh variables for modelling, and Rao and Ali’s (2002) use of genetic algorithms to select neural network inputs. Though both these studies effectively transformed traditional variables, the transformation phase will be particularly pressed for advances in the near future as progress in computational and storage capacity enables the accumulation of non-traditional data such as text, voice and other rich datasets with heterogeneous scales. This non-traditional data creates significant challenges in transforming different scales into statistically usable and computationally friendly formats. Both show that this phase still holds much to discover with significant upside in performance enhancement opportunities along the way.

3.6.5 Data Mining Techniques

A wide array of statistical techniques is currently mentioned in literature to deliver on the aforementioned applications. Cui, Wong and Lui (2006) categorises these techniques and other data-mining techniques in two classes: statistical analysis techniques and machine learning techniques.

Statistical techniques are applied across the range of Data Mining applications. From associations to prediction and time course of events, statistical techniques involve a high degree of human interaction and judgment in their application.

Machine learning techniques “refers to computer-based techniques that can extract patterns or knowledge from data and perform optimisation tasks with minimum human intervention. Most of these techniques have their roots in artificial intelligence and dynamic programming” (Cui et al., 2006, p. 598). These techniques are dubbed learning techniques because they use algorithms that recognise complex patterns and allow systems to make intelligent decisions. Kantardzic (2002) defines the two common learning algorithm types as supervised and unsupervised. Supervised learning estimates an unknown dependency based on input-output samples. Supervision denotes that outputs from a training sample are made available. Unsupervised learning uses samples with input value only with no notion of the output during the learning process. As a result, the system discovers, or learns from, the natural structure it derives from the input data. Some common machine learning and statistical techniques are detailed in Tables 17 and 18.

Table 17: Classes of Machine Learning Techniques

Technique Classes	Description
Neural Networks (NN)	A class of systems modelled after the human brain. As the human brain consists of millions of neurons that are inter-connected by synapses, NN are formed from large numbers of simulated neurons, connected to each other in a manner similar to brain neurons. As in the human brain, the strength of neuron inter-connections may change (or be changed by the learning algorithm) in response to a presented stimulus or an obtained output, which enables the network to "learn".
Case-Based Reasoning (CBR)	Technology that tries to solve a given problem by making direct use of past experiences and solutions. A case is usually a specific problem that was encountered and solved previously. Given a particular new problem, CBR examines the set of stored cases and finds similar ones. If similar cases exist, their solution is applied to the new problem, and the problem is added to the case base for future reference.
Genetic Algorithms (GA)	Procedures modelled upon the evolutionary biological processes of selection, reproduction, mutation, and survival of the fittest to search for very good solutions to prediction and classification problems. GA are used in Data Mining to formulate hypotheses about dependencies between variables in the form of association rules or some other internal formalism.
Decision Trees (DT)	Similar to those used in decision analysis, an individual non-terminal node represents a test or decision on the data item considered. Depending on the outcome of the test, one chooses a certain branch. To classify a particular data item, one would start at the root node and follow the assertions down until a terminal node (or leaf) is reached; at that point, a decision is made. DT can also be interpreted as a special form of a rule set, characterised by their hierarchical organisation of rules.
Association Rules (AR)	Statements about relationships between the attributes of a known group of entities and one or more aspects of those entities that enable predictions to be made about aspects of other entities who are not in the group, but who possess the same attributes. More generally, AR state a statistical correlation between the occurrences of certain attributes in a data item, or between certain data items in a data set. The general form of an AR is $X1…Xn \Rightarrow Y[C,S]$ which means that the attributes $X1,…,Xn$ predict Y with a confidence C and a significance S.

Source: Johnson and Wichern (1998)

Table 18: Classes of Statistical Techniques

Technique	Description
Descriptive and Visualisation Techniques	Techniques that include simple descriptive statistics such as: • averages and measures of variation, • counts and percentages, and • cross-tabs and simple correlations They are useful for understanding the structure of the data. Visualisation is primarily a discovery technique and is useful for interpreting large amounts of data; visualisation tools include histograms, box plots, scatter diagrams, and multi-dimensional surface plots
Cluster Analysis	Techniques that seek to organise information about variables so that relatively homogeneous groups, or "clusters," can be formed. The clusters formed with this family of methods should be highly internally homogenous (members are similar to one another) and highly externally heterogeneous (members are *not* like members of other clusters).
Correlation Analysis	Techniques that measure the relationship between two variables. The resulting correlation coefficient shows if changes in one variable will result in changes in the other. When comparing the correlation between two variables, the goal is to see if a change in the independent variable will result in a change in the dependent variable. This information helps in understanding an independent variable's predictive abilities. Correlation findings, just as regression findings, can be useful in Analysing causal relationships, but they do not by themselves establish causal patterns.
Discriminant Analysis	Techniques used to predict membership in two or more mutually exclusive groups from a set of predictors, when there is no natural ordering on the groups. Discriminant analysis can be seen as the inverse of a one-way multivariate analysis of variance (MANOVA) in that the levels of the independent variable (or factor) for MANOVA become the categories of the dependent variable for discriminant analysis, and the dependent variables of the MANOVA become the predictors for discriminant analysis.
Factor Analysis	Techniques useful for understanding the underlying reasons for the correlations among a group of variables. The main applications of factor analytic techniques are to reduce the number of variables and to detect structure in the relationships among variables; that is to classify variables. Therefore, factor analysis can be applied as a data reduction or structure detection method. In an exploratory factor analysis, the goal is to explore or search for a factor structure. Confirmatory factor analysis, on the other hand, assumes the factor structure is known a priori and the objective is to empirically verify or confirm that the assumed factor structure is correct.
Regression Analysis	Statistical tools that use the relation between two or more quantitative variables so that one variable (dependent variable) can be predicted from the other(s) (independent variables). But no matter how strong the statistical relations are between the variables, no cause-and-effect pattern is necessarily implied by the regression model. Regression analysis comes in many flavours, including simple linear, multiple linear, curvilinear, and multiple curvilinear regression models, as well as logistic regression, which is discussed next.
Logistic Regression	Techniques used when the response variable is a binary or qualitative outcome. Although logistic regression finds a "best fitting" equation just as linear regression does, the principles on which it does so are rather different. Instead of using a least-squared deviations criterion4 for the best fit, it uses a maximum likelihood method, that is, it maximises the probability of obtaining the observed results given the fitted regression coefficients. Because logistic regression does not make any assumptions about the distribution for the independent variables, it is more robust to violations of the normality assumption. Some of the more common flavours that logistic regression comes in include simple, multiple, polytomous and Poisson logistic regression models.

Source: Johnson and Wichern (1998)

Full and detailed definitions of Data Mining techniques can be found in Appendix 2.

3.6.6 Normative Technique Applications

The individual techniques reviewed can be classified with the Data Mining tasks reviewed previously in section 3.6.3. To simplify presentation and ensure relevance, only techniques that were applied within the context of Pattern Discovery and Modelling were retained. With the exception of outlier analysis, all of the Data Mining tasks can be used for customer selection (e.g. concept and/or class related customers, clusters of customers, customers with high product association patterns, and modelled customers). Outlier analysis is excluded, as it is applied after the customer selection phase and is explicitly part of the subsequent phase 6 of Interpretation/Evaluation. This classification is illustrated in Table 19.

Table 19: Normative Technique Applications by Class

	Concept/ Class Description: Characterisa-tion & Discrimina-tion	***Cluster Analysis/ Segmentation***	***Mining Frequent Patterns, Associations & Correlations***	***Classification & Prediction***	***Evolution Analysis***
Classes of Machine Learning Techniques					
Descriptive & Visualisation Techniques	●				
Cluster Analysis		●			
Correlation Analysis			●		
Discriminant Analysis				●	●
Factor Analysis		●			
Regression Analysis	●		●	●	●
Logistic Regression				●	●
Classes of Machine Learning Techniques					
Neural Networks		●	●	●	
Case-Based Reasoning (CBR)		●	●	●	
Genetic Algorithms			●		
Decision Trees				●	
Association Rules			●		

The most recent and relevant study of technique application and usage among academics is provided by Ngai (2009) in his systematic review of Data Mining application in CRM. Ngai reviews nine hundred academic articles in search of a relationship between Data Mining techniques and CRM. He identifies eighty-seven articles of substance that he subsequently categorises into four CRM dimensions (Customer Identification, Customer Attraction, Customer Retention and Customer Development), seven Data Mining functions (Association, Classification, Clustering, Forecasting, Regression, Sequence Discovery and Visualisation) and nine sub-categories of CRM. From my review of the literature, Ngai's study is the most comprehensive review to date on Data Mining application conducted in CRM and connected fields of Database Marketing. Of particular interest is the technique usage rate among academics. Out of 87 studies and 125 technique applications, Ngai identifies 34 separate Data Mining techniques. Of those techniques, neural networks are the most frequently used by academics with a usage in 34.5 percent (30 out of 87) of assessed studies. Following neural networks, techniques with the most usage included decision trees, association rules and regression with respectively 26.4 percent, 20.7 percent and 11.5 percent of studies.

The majority of the reviewed articles relate to customer retention. Of these, 51.9 percent (28 articles) and 44.4 percent (24 articles) are related to one-to-one marketing and loyalty programs, respectively. These articles could provide insight to organisation policy makers on the common Data Mining practices used in retaining customers.

Table 20: Distribution of articles by Data Mining techniques

Data Mining Techniques	**Mentions/Usage**
Neural Networks	30
Decision Tree	23
Association Rules	18
Regression	10
Genetic Algorithm	4
Markov Chain	4
Survival Analysis	4
K means	3
K nearest neighbour	3
Bayesian network classifier	2
If-the-else	1
Set Theory	1
Support Vector Machine	1
Attribute Oriented Induction	1
Constructive Assignment	1
Customer Map	1
Data Envelopment Analysis	1
Data Mining by Evolutionary Learning	1
Expectation Max	1
Expectation Max Mod	1
Farthest First	1
Goal Oriented Sequential Pattern	1
Latent Class Model	1
MARFS1/S2	1
Mixture Transition Distribution	1
Multi-Classifier Class Combiner	1
Multivariate Adaptive Regression Splines (MARS)	1
Online Analytical Mining	1
Outlier Detection	1
Pattern Based Cluster	1
Rule Based RIPPER	1
S-Means	1
S-Means Mod	1
Total	**125**

Source: Ngai (2009)

As mentioned in the introduction section, much of the literature on techniques is conducted with the express objective of providing performance comparisons. Of these comparison studies, a number provide compelling, but often contradictory, evidence on comparative technique performance. As a full technique performance comparison will be conducted in the systematic review, I will focus the attention of this literature review on the comparative performance of some leading techniques identified by Ngai (2009).

Some of the most recent work by Olson and Chae (2012) compare traditional RFM performance against more modern RFM-based technique variants, and the top three techniques identified by Ngai: logistic regression, decision trees, and neural networks. Results of their work illustrated that, though the RFM variant performed only slightly better, the Data Mining techniques of logistic regression, neural networks and decision trees all generated better predictive and performance results. Olson and Chae also provide an interesting comparison of methods that adds some colour to performance measures (Table 21).

Table 21: Comparison of Methods

Model	Relative Advantages	Relative Disadvantages	Inferences
Basic RFM	• Widely used • Data readily available • Software obtainable	• Predictive accuracy consistently weak	• Can do better using conventional Data Mining (RFM implicitly a special case)
RFM with Balanced Data	• Better statistical practice	• May not actually improve accuracy	• Not worth the trouble
Value function	• Eay to apply (uses the 2 of the 3 RFM variables, so data reasily available) • Focuses on uncorrelated variables	• May not actually improve accuracy • Not necessarily more accurate	• Value function is superior to RFM
Logistic regression	• Can get better fit • Can include many variables • Model statistically interpretable	• Logistic output harder to interpret than OLS for managers	• Decision trees easier to interpret
Neural network	• Can get better fit • Can include many variables	• Output not conducive to interpretation • Can't apply model outside of software used to build model	• Decision trees easier to interpret
Decision trees	• Can get better fit • Can include many variables • Output easily understandable by managers	• Model may involve an excessive number of rules	• Best option, if can control the number of rules obtained (through minimum required response parameter)

Source: Olson and Chae (2012)

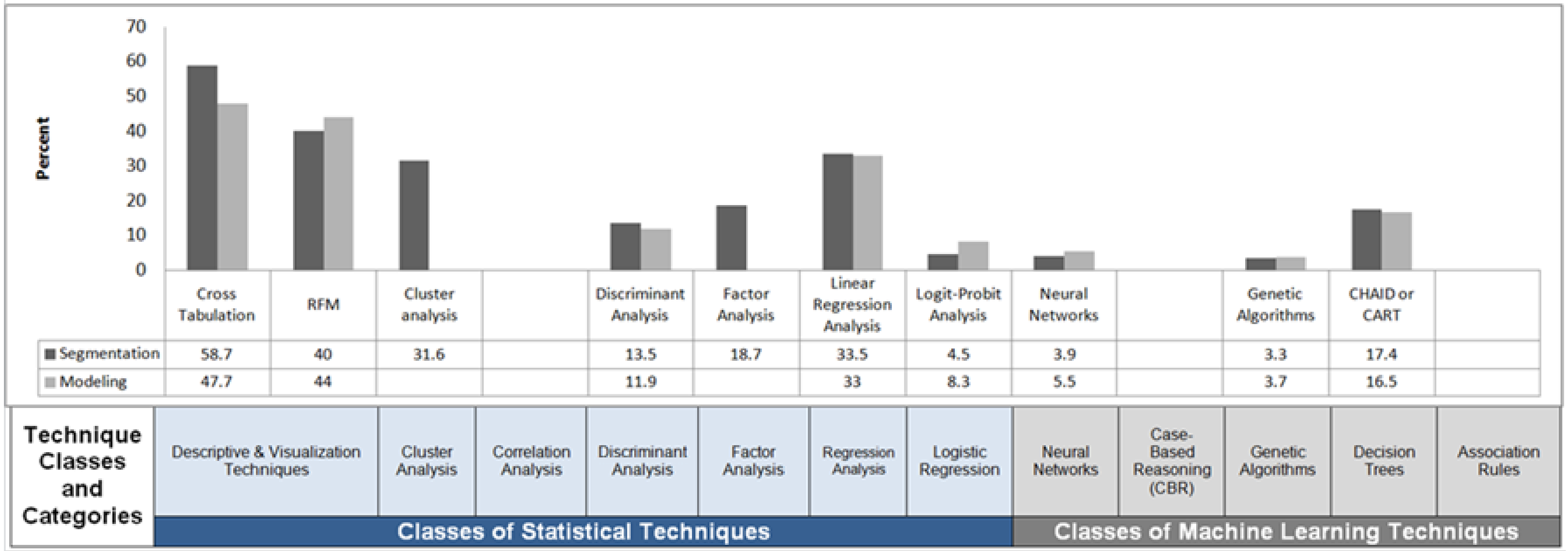
On the traditional RFM and value function techniques, as most authors before them, Olson and Chae note that data and software availability are its main advantages, while predictive ability is consistently weak. Moving into more comprehensive methods, the authors state that decision trees are certainly the easiest for managers to use given their 'interpretability'; however, all three techniques (decision tree, neural network and logistic regression) provide a greater predictive ability and can all manage extended variable sets. Unsurprisingly, neural networks remain the most difficult to interpret given its black box design.

3.6.7 Practical Technique Applications

Regardless of technique classification or application, practitioner usage of these techniques varies significantly. Verhoef, Spring, Hoekstra, and Leeflang's (2003, p. 475) investigation on the usage of statistical techniques among practitioners for the purposes of segmentation and modelling shows that "despite the fact that literature advocates more sophisticated techniques, cross-tabulation is the most widely utilised technique followed by RFM analysis". Linear regression is the third most popular technique followed by CHAID or CART, which are employed by approximately 16 percent of the respondents, while logit or probit models are used less. Verhoef et al. (2003) suggest that neural networks and genetic algorithms are ranked last because of the disappointing performance of these solutions in practice and a black-box design that is hard to understand. Performance issues are tied to these solutions not delivering against the objectives that are stated later in the review in section 4.5. This black box design has also lead to neural networks performance uncertainty due to 'overlearning or underlearning' issues (Thieme, Song and Calantone, 2000) – the process by which the network's training either 'overcomplexifies' a problem or doesn't include the full complexity of a problem. Finally, with the exception of cluster analysis and factor analysis, the rankings are consistent even when responders are asked to evaluate these techniques against the differing objectives of segmentation and modelling.

What emerges as the key finding when comparing this survey of practitioner practices (Figure 10) with normative applications of techniques (Table 19) is that techniques are rarely used appropriately. For example, the cross-tabulation technique is the most popular modelling technique in practice even through it is not designed for predictive purposes. The same can be said of all other techniques with the exception of neural networks that seem to be appropriately used by practitioners.

Figure 10: Statistical Technique Usage among Practitioners



Source: Adapted from Verhoef, Spring, Hoekstra, and Leeflang (2003)

3.6.8 Data Mining Variables

In direct marketing, the RFM variables are the most likely candidates as bases for segmentation (Bult and Wittink, 1996; Van den Poel, 2003). The RFM variables measure consumer response behaviour in three dimensions. The first dimension is Recency, which indicates how long it has been since the customer has last responded. The second is Frequency, which provides a measure of how often the customer has responded to mailings received. And finally, Monetary Value measures the amount of money or the number of products that the customer has spent in response to the mailings (Jonker, Piersma, and Van den Poel, 2004).

However, the dominance of RFM variables as Data Mining bases is eroding rapidly as other bases are emerging and becoming more accessible. Mama (2007) provides a good overview of these bases in Table 22.

Table 22: Typology of Customer Selection/Segmentation Bases

Basis	Description	Authors
Geographic	Dividing markets into different geographical units including countries, regions, cities, towns and population density.	Kotler 1991, Beane and Ennis 1987, Haley, 1995
Demographic	Dividing a market based on demographic variables such as age, gender, family size, income, occupation, education, religion or nationality	Haley 1995,Kotler 2002, Blattberg, 1976, Greenberg and McDonald (1989), Beane and Ennis (1987), Tynan and Drayton
Psychographic/ lifestyle	Dividing markets based on consumer values, attitudes, interests, opinions.	Alpert, 1972, Frank et al. 1972, Pessemier et al.1967, Lazer 1963, Plummer 1974, Yankelovich 1964 and 2004
Benefits	Dividing the market into groups according to different benefits that consumers seek from the product or service.	Haley 1968, Myers 1976, Greenberg & McDonald 1989, Beane and Ennis, 1987, Kotler 2002
Usage	Dividing markets based on usage patterns such as non-user, ex-user, potential user, first time user, regular user, high volume user.	Twedt, 1964, Young et. Al, 1978
Loyalty	Dividing consumers based on brand loyalty: loyals, habituals, variety seekers and switchers	Knox, Simon 1998
Situation	Related to usage segmentation, situation segmentation divides markets on the basis of the consumption or purchase situation of consumers.	Dickson, 1982
Behavioural	Dividing markets based on consumer's knowledge of attitude toward uses for and responses to a product.	Kotler et al, 2002

Source: Mama (2007)

Geographic selection/segmentation is based on the view that consumers' needs and wants may vary geographically (Beane and Ennis, 1987). This type of segmentation divides markets into different geographical entities that may be represented by continents, countries, cities, regions, sub-regions, and other similar dimensions. Examples of geographical segmentations at postal-code/neighbourhood levels include: Mosaic and PRIZM in the US, Generation 5 in Canada and ACORN in the UK. Demographic selection/segmentation divides individuals according to variables such as gender, age, income, occupation, race/ethnic background and language (Kotler, 2000). Psychographic selection/segmentations divide individuals into groups based on indicators of motivations that may include: personality characteristics, values, beliefs, lifestyles and social class (Greenberg and McDonald, 1989). Needs/benefit selection/segmentation groups persons according to the unobserved motivations and perceptions behind category choice making (Greenberg and McDonald, 1989). Situation selection/segmentation, for its part, breaks down individuals into groups according to consumer usage situations (Dickson, 1982). Behavioural selection/segmentation aims to separate individuals into groups based on variables such as: transactional details (including active/inactive, frequency, recency, monetary value, product consumption and usage) and loyalty and engagement with the brand that can usually be tracked via transactional and relational databases supported by targeted couponing and/or loyalty programs (Knox, 1998). Loyalty segmentation can be characterised as behavioural segmentation or as a separate category. Finally, occasion selection/segmentation looks to divide groups based on specific events that coincide with consumer purchases. These can be birthdays, holidays, sporting events and the like (Dubow, 1992).

Table 23: Decisional Depth and Links to Data Mining

	Issues the business wants to address	**Consumers' decisions**	**What the segmentation should try to find out**
Shallowest Decisions	• Whether to make small improvements to existing products • How to select targets of a media campaign • Whether to change prices	• How relevant and believable new product claims are • How to evaluate a given product • Whether to switch products	• Buying and usage behaviour • Willingness to pay a small premium for higher quality • Degree of brand loyalty
Middle-of-the spectrum Decisions	• How to position the brand • Which segments to pursue • Whether to change the product fundamentally • Whether to develop an entirely new product	• Whether to visit a clinic about a medical condition • Whether to switch one's brand of car • Whether to replace an enterprise software	• Whether the consumer being studied are do-it yourself or do-it-for-me types • Consumers' needs (better service, convenience, functionality) • Their social status, self-image, and lifestyle

(continued)	Issues the business wants to address	Consumers' decisions	What the segmentation should try to find out
Deepest Decisions	• Whether to revise the business model in response to powerful social forces changing how people live their lives	• Choosing a course of medical treatment • Deciding where to live	• Core values and beliefs related to the buying decision

Source: Yankelovich (2006)

As shown in Table 22, data variables/bases to be used in Data Mining and segmentation have been discussed by many authors. To add some actionnability to these data variables/bases, some authors, such as Yankelovich (2006), have designed frameworks to define the bases to be selected and in what circumstances (Table 23). In this depth of decision framework, the more observable bases, the shallower the decisions to be taken are enabled and vice-versa. Other authors, such as Wedel and Kamakura (2000), outline these bases grounded by how generic and observable these bases are, and subsequently assess them based on their relative contributions to identifiability, substantiality, accessibility, stability, actionnability and responsiveness (Table 24). Identifiability represents the degree to which distinct segments can be recognised as different from a behaviour perspective. Substantiality signifies the level to which the segment is sizeable enough to meet objectives such as profitability. Accessibility is defined as the segments' capacity to be reached via promotional campaigns and initiatives. Actionability is achieved when a segment is identifiable and promotional targeting results in customer behavioural changes. This dimension presumes an inherent degree of segment member behavioural consistency as it relates to marketing mix stimuli. Stability is defined as a segment's consistency of results in terms of its composition steadiness following repeated clustering and grouping applications. And finally, responsiveness relates to a segment's distinct response to marketing stimuli versus other segments.

Of particular interest to the thesis are the criteria of identifiability, responsiveness and stability. These criteria directly address the fit and performance measures of effectiveness. More particularly, the bases most likely to have a positive impact on these criteria are the bases of purchase, usage and benefits. Rust (2005) and Ortmeyer, Lattin and Montgomery (1991) support this claim as well.

Other representations of selection/segmentation attractiveness include Kotler (1997), Evans and Berman (1997) and Morrit (2007). Kotler's (1997) five criteria for segment attractiveness include: measurability, accessibility, substantiality, differentiation, and actionability. Evans and Berman (1997) propose five slightly different criteria: differentiation, targeting homogeneity, measurability, substantiality, and accessibility. Morrit (2007) includes Kotler's criteria and adds five more: defensibility from competition, stability over time, inter-segment homogeneity, segment competitive differentiation, and compatibility with other segments. For this literature review, I will refer to Wedel and Kamakura's approach as it generally encompasses most of the dimensions proposed by Kotler (1997), Evans and Berman (1997), and Morrit (2007). Furthermore, this approach is also complemented by an evaluation of segmentation bases that is particularly helpful in assessing selection/segmentation variables.

Though both Yankelovich's (2001) and Wedel and Kamakura's (2000) approaches show merit, they are difficult to reconcile. They do, however, indicate that different bases are useful for different objectives and both converge on the importance of incremental data variables and bases to deliver results and, more specifically in the shallow dimensions quadrant, to the importance of purchase (buying and usage) behaviour in helping the customer selection decision. Rossi (1996) echoes this final finding by stating "the tremendous potential for improving profitability of direct marketing lies in utilising household purchase histories" and empirically demonstrating that even short purchase histories can yield revenue gains of 150 percent or more (versus a random sample).

Table 24: Evaluation of Segmentation Bases

Bases/Criteria	Identifia-bility	Substan-tiality	Accessa-bility	Stability	Actionna-bility	Respon-siveness
1. General, observable	++	++	++	++	-	-
2. Specific, observable						
Purchase	+	++	-	+	-	+
Usage	+	++	-	+	-	+
3. General, unobservable						
Personality	+/-	-	+/-	+/-	-	-
Psychographics	+/-	-	+/-	+/-	-	-
4. Specific, unobservable		-	+/-	+/-	-	-
Psychographics	+/-	+	-	-	++	+/-
Perceptions	+/-	+	-	-	+	-
Benefits	+	+	-	+	++	++
Intentions	+	+	-	+/-	-	++

Source: Wedel and Kamakura, 2000

Studies demonstrating that customer transactional data variables are generally the most effective at predicting response behaviour in CRM are numerous (Rud, 2001). Van de Poel (2003) illustrates that all three traditional variables for RFM (Recency, Frequency and Monetary value) are very significant in predicting repeat purchase. In his study, these variables explain 50 percent of performance effectiveness (assessed with AUC). Though often the combination of recent purchases, high frequency of transactions and high monetary expenditures correlate well with customer value, the RFM combinations that best predict promotional behaviours can be quite different than those that align with customer value. For example, the work of Lal and Bell (2003) indicates that in the context of a US grocery chain's loyalty program, it was the incremental sales from casual shoppers, also referred to as cherry pickers, that generated the greatest return for retailers. In RFM terms, this would mean that RFM segments with lower Frequencies, might be prioritised over higher ones. This insight is often very useful in promotional design. Along with RFM variables, other strong predictive variables include credit usage, length of relationship, and historical order buying behaviour.

McCrary (2009) conducted a study of an American multi-channel, Fortune 500 speciality retailer and considered more than 500 independent variables as predictors of customer sales and profit. This list consisted of (among other dimensions) a good representation of Wedel and Kamakura's (2000) segmentation dimensions: past shopping behaviour, payment information, products purchased, seasonality or time period shopped, and household demographics such as marital status, age and home value. Following the application of a stepwise procedure and the application of Schwarz's Bayesian Criterion (SBC) statistic to identify the best model, significant variables shrank to the following subset in his first of two studies: customer tenure, number of departments they shopped, use of proprietary charge card, number of transactions, recency since customers had shopped with other retailers and the vehicle composition of the customer's neighbourhood. The second study results were similar and the short list of variables included: customer's number of trips, number of products purchased, average spent per trip, number of transactions made in past three months, total spent, recency, and indicator variable for single transaction customer and sales shoppers.

Evidence from non-transactional variables has been less consistent in its direct relationship with technique effectiveness. Haley (1985) argues that geographic and demographic segmentations are meagre indicators of predictive behaviour while Beane and Ennis (1987) also argue that demographic segmentation is better at being a profile descriptor than a causal predictor variable. When it comes to psychographic segmentations, Yankelovich (2006) argues against their use for CRM use acquisition.

Particular variable effectiveness can be debated at considerable length. However, no matter how powerful a model is, the true risk remains that irrelevant independent variables lead to poor accuracy (Ho, Cho and MacLachlan, 2005). It is thus important to try to limit input predictor variables to the most significant ones early in the process. However, this is not to say that parsimony is always required in model design and technique application. In fact, numerous authors do argue that adding relevant variables to advanced technique does have a direct and positive impact on mode effectiveness. Work conducted by Kumar (2006) identifies significant gains in the direct marketing effectiveness of firms utilising deeper and more relevant data on customers. Similar work by McCarty and Hastak (2006) and Yang (2004), where the model performance using basic recency, frequency and monetary value variables is tested against expanded data sets mainly composed of additional behavioural variables, also illustrates the increasingly predictive nature of deeper, more relevant data variable sets.

3.7 Summary

Customer selection is defined as the selection of customers or prospects for the purpose of direct communication through the utilisation of patterns, statistical or predictive models issued from data to obtain a measurable response or transaction via one or multiple channels. It is interested in customer selection in low involvement contexts. The main research domains that inform this research on customer selection are relationship marketing, customer relationship management (CRM), database marketing, and data mining.

Review of each of these sub-domains individually illustrates that RM is complex and is enabled by: CRM processes and technologies, the advancement of Data Mining techniques, and the applications of database marketing. It also shows that the domains of RM, CRM and Database Marketing overlap and aim to achieve the same objectives of building customer relationships through the extended use of relevant information and insight. RM, CRM, Database Marketing and segmentation all support the development of long-term customer relationships via building, retaining, developing and potentially terminating unprofitable customer relationships. The added CRM, Database Marketing and Data Mining lenses do so by leveraging the customer information and technology assets and processes in a targeted, one-to-one marketing fashion (Blattberg et al., 2008; Sausen et al., 2005; Grönroos, 1990; Hobby, 1999; Swift, 2000; Parvatiyar and Sheth, 2001; Campbell, 2003; Reinartz et al., 2004; Peppers et al., 1999; Payne and Frow, 2005). Therefore, I suggest that the objectives of customer selection should be: building, retaining, developing and potentially terminating customer relationships, using data and technology to build long-term relationships. Ngai (2009) also adds the additional objective of Customer Identification, which involves Analysing customers for purposes of profiling and segmentation.

By nature of their common roots and objectives, these sub-domains have high degrees of overlap but are all mainly concerned with the acquisition, retention and development of customers. However, through this research I have tried to distinguish them and characterise the relationships between each sub-domain as follows:

- Relationship Marketing (RM) provides the theoretical grounding for customer-centred (also referred to as one-to-one) marketing;
- Customer relationship management (CRM) relates to the application of RM concepts with superior customer information;
- Database Marketing “provides the technological enabler, allowing vast quantities of customer-related data to be stored and accessed in ways that create strategic and tactical marketing opportunities” (Chaffey et al., 2000, pp. 290–91); and
- Data Mining provides the techniques and processes by which data can be sourced and manipulated to create actionable insight from databases.

Key insights from the RM literature include the increasingly reframed view of RM as a concept that spans both Transaction and Relational Marketing and includes applications that vary between schools of thought. Some of the interesting concepts related to the study of transactional Database Marketing campaigns include the notion of transactionally oriented RM strategies. Such strategies are differentiated from purely Transactional Marketing applications by their: formal use of advanced analytics and data to seek a detailed understanding of customer behaviour (Frow and Payne, 2009; Coviello et al., 2002), the shift away from mass marketing practices and in favour of targeted marketing of customised offers (Maklan et al., 2011) and the aim of achieving repeat transactions through some degree of customer interactivity (behavioural CRM). Nevertheless, even with this relational shift, the application of RM in low involvement contexts remains generally directed "to" the customer, with relationships being arms length, tactical and discrete in nature and facilitated through the use of technology. As a result, it is safe to assume that low involvement retail applications of RM typically inform the more operational and behavioural aspects of RM. The areas informed from the literature include:

- Mattson’s (1997) transactional practice within the relational continuum
- Coviello et al.’s (2002) Database Marketing practice of RM
- Eggert and Stieff’s (1999) limited RM mode and behavioural RM
- Frow and Payne’s (2009) customer-based marketing form
- Maklan et al.’s (2011) one-to-one type of RM

Across various definitions and forms of RM, I noted that the aims of RM are increasingly aligned among authors and include the building, retaining, development, and potential termination of customer interactions, with the objective of building long-term relationships. This degree of consensus on the objectives of RM is not, however, translated into a consensus of the term’s definition and deeper shared understanding in itself. Using Harker’s (1999) seven categories scoring system allowed me to exclude any single-issue or tactical definitions of the term, and provides a good baseline with which to select a conceptually “comprehensive” definition that would address all seven categories. Given the strength of Grönroos' definition and the highly complementary nature and Shani and Chalasani (1992), I opted to take the strongest components of both definitions and craft a definition that contains all of the concepts required by Harker's (1999) categorisation while also highlighting the integrated and coordinated marketing capabilities core to supporting the deployment of CRM strategy (Payne and Frow, 2005). The definition follows:

> *Integrated efforts to identify and establish, maintain and enhance and, when necessary, also terminate relationships with customers and other stakeholders at a profit, and to continuously strengthen the relationship through mutual exchange and fulfilment of promises, through interactive, individualised and value-added contacts over a long period of time.*

The review of CRM literature suggests that CRM can be viewed from many perspectives: "as a particular technology solution, wide-ranging technology, and customer centric" (Payne and Frow, 2005, p. 168); strategically or operationally (Richard and Jones, 2008); analytically or operationally (Leigh and Tanner, 2004); as a technology or a process (Jayachandran, Sharma, Kaufman, and Raman, 2005); or as a process, a strategy, a philosophy, a capability, and a technology (Zablah, Bellenger and Johnston, 2004). I contend that the strategic view/ macroprocess perspective of CRM provides the best conceptual foundation for the CRM phenomenon as it presents the most comprehensive view of CRM (Zablah et al., 2004). I also support the critical role of IT as an enabler of CRM and the view of CRM as a process given the instrumentality of connecting IT with customer data in an organised and systematic manner. I discard the view of CRM as a philosophy in favour of its role as an enabler of RM through the use of technology and processes. From a process perspective, CRM has a lot of commonalities with data mining, with the KDD process potentially being a substitute for Jayachandran et al.'s (2005) relational information process. These observations lead me to conclude that the most appropriate definition is provided by Payne and Frow (2005):

> *CRM is a strategic approach that is concerned with creating improved shareholder value through the development of appropriate relationships with key customers and customer segments. CRM unites the potential of RM strategies and IT to create profitable, long-term relationships with customers and other key stakeholders. CRM provides enhanced opportunities to use data and information to both understand customers and to create value with them. This requires a cross-functional integration of processes, people, operations, and marketing capabilities that is enabled through information, technology, and applications.*

One particular area of opportunity for research emerges from the disagreement in literature on the applicability of RM and CRM to the areas of low involvement decisions such as retail FMCG. Although most RM conceptualisations inherently support the existence of some form of customer relationship development potential within low involvement contexts, Harker and Egan (2006) and Sheth and Parvatiyar (2005) indicate that the notion of a strong relationship has yet to be established between a buyer and a seller in mass marketing or Standardised consumer product contexts. However, Leahy (2011) and Grönroos (2009) raise the possibility that there may exist segments and individuals that may value some form of transactional relationship and that some forms of personalised marketing mix approaches may be sufficient to persuade customers to buy, increase satisfaction with the value created and ultimately support the development of a transactional relationship.

In reviewing Database Marketing literature, relational databases were identified as the key triggers of the Database Marketing activity with the intent of improving a customer relationship and increasing marketing productivity. Beyond the significant opportunities offered by traditional data analytics and relational database, I identify that the advent of Big Data provides practitioners and academics with ever-increased incentives in understand the links between data and customer behaviour. As such, the view of Database Marketing as the cornerstone of RM is increasingly relevant, with applications that extend to the entire field of marketing (and indeed throughout the customer experience). I also relate research on segmentation to research on Database Marketing, as segmentation is "used to distinguish between customers and non-customers... to understand their composition and characteristics... and supports a whole array of decisions, ranging from targeting decisions to determining efficient and cost effective marketing strategies, even evaluating market competition" (Levin and Zahavi, 2001, p. 3). This view indicates that segmentation may indeed also be an output of the Database Marketing process. For purposes of this study, I adopt the position that segmentation is indeed an output of database marketing. As in other sections, Database Marketing can be viewed in multiple ways and, in this case, as a tool or a process. Unlike other sections where the research needed to be positioned clearly within specific domains, the process and utilitarian tool perspectives of Database Marketing are complementary. This research applies Database Marketing as a process to attract, develop and manage long-term relationships with a target group of customers by means of using Data Mining for targeting and CRM tactics. Some statistics from the Food Marketing Institute in the US (2012) do add weight to the argument that frequency-based transactional relationships may indeed exist as promotions are stated as one of the top ten store selection criteria by consumers in the US (FMI, 2012). The addition of an additional level of relevance and targeting to such promotions certainly aligns with Tesco's strategy of Customer-based Marketing. This customer-based marketing approach is certainly in line with recent choice reduction research in grocery retail where several types of transaction costs have been identified and indeed can contribute to increasing the barriers to switching and encourage transactional loyalty and relationships (Betancourt, 2005; Bell, Ho and Tang, 1998; Lewis, Singh and Fay, 2006; Chintagunta, Chu, and Cebollada, 2012). Such barriers can certainly be raised when firms aim to provide an increasingly relevant dialogue with consumers that not only simplifies information processing and reduces alternative research, but ultimately also influences pertinent buying and consumption behaviours, all of which result in a higher degree of consumer psychological comfort.

Finally, the Data Mining literature is identified as the cornerstone of the research study given its intrinsic role in applying adequate techniques to marketing practice. The two primary aims of Data Mining are characterised by Wedel and Kamakura (2000) as prediction and description. From a categorisation tool perspective, literature on Data Mining provides two frameworks, the first to categorise tasks and the second to categorise the phases of data mining. In order to effectively categorise systematic review literature, the Data Mining task categorisation of Han and Kamber (2006) is adopted. In order to categorise phases of data mining, I identify Fayyad et al.'s (1996b) five phases of Data Mining (Knowledge Discovery in Databases-KDD). Outside these process observations and in comparing academic versus practitioner literature, I note that academic studies have a greater breadth of applications whereas practitioner studies seem to illustrate greater simplicity and limited span of applications. Furthermore, review of normative and practical applications illustrates that techniques are rarely used appropriately according to normative recommendation. Nevertheless, from Ngai's (2009) systematic review of Data Mining applications in CRM, most frequently used techniques include: neural networks, decision trees, association rules, and regression. From an independent variable perspective, numerous types of variables exist to inform data mining. These include: geographic, demographic, psychographic/lifestyle, benefits, usage, loyalty, situation, behavioural. Yankelovich (2006) provides an interesting overview on how types of data may inform different types of decision. This framework, although interesting, is prescriptive and well complemented by the conceptual work of Wedel and Kamakura (2000) that illustrates what types of variables impact different dimensions of segmentation. Variables best aligned with the intent of customer selection for promotional campaigns include purchase and usage variables (Rud, 2001) and the generally specific, unobservable variables of psychographics, benefits and intentions (Wedel and Kamakura, 2000). Beyond the individual data variables, literature also indicates that increasing the amount of data variables in appropriate categories has a direct effect on increasing technique effectiveness (Van de Poel, 2003; Kumar, 2006; McCarty and Hastak, 2006; Yang, 2004).

3.8 Conclusion

To identify simple or complex preferences or predictors, Data Mining techniques are particularly helpful. However, decades of advances in statistical techniques and scholarly research have yet to provide practitioners with a good understanding of which single, or combination of, decision variable(s) works best in practice (Wedel and Kamakura, 2000). Nevertheless, the literature review supports that increased insight into individual customers has significant implications on each of these sub-domains and, therefore, on customer selection. Indeed, several researchers have suggested that customer profits can be considerably enhanced if marketing contacts were adapted to individual customer preferences (Ansari and Mela, 2003; Reinartz, et al., 2005; Venkatesan and Kumar, 2004) across RM objectives, throughout CRM technologies and Database Marketing applications.

As stated earlier, the benefit realisation gap between data availability and Database Marketing effectiveness may partly be due to the unsystematic fashion in which research in Database Marketing and Data Mining has been conducted. With little attention given to contexts and objectives and more attention to comparative outputs, most of the research can be qualified as technique centric. Given the significant amount of unsystematic, technique-centric research conducted throughout the years, moving directly from a literature review to empirical research risks falling prey to the same technique-centric limitations of previous research. This literature review provides the foundations for a subsequent systematic review of literature by identifying the critical concepts and questions to be assessed in the quantitative phase of my research. The systematic review itself delves much deeper into key questions informing the thesis topic in addition to establishing a strong baseline of information required to adequately design the empirical research studies that follow. In order to inform the conceptual model, the main question that is asked of the systematic review is:

What is the empirical evidence for the effectiveness of Data Mining techniques for customer selection in database marketing?

In order to fully answer the systematic review question and adequately categorise information, six sub-questions are asked:

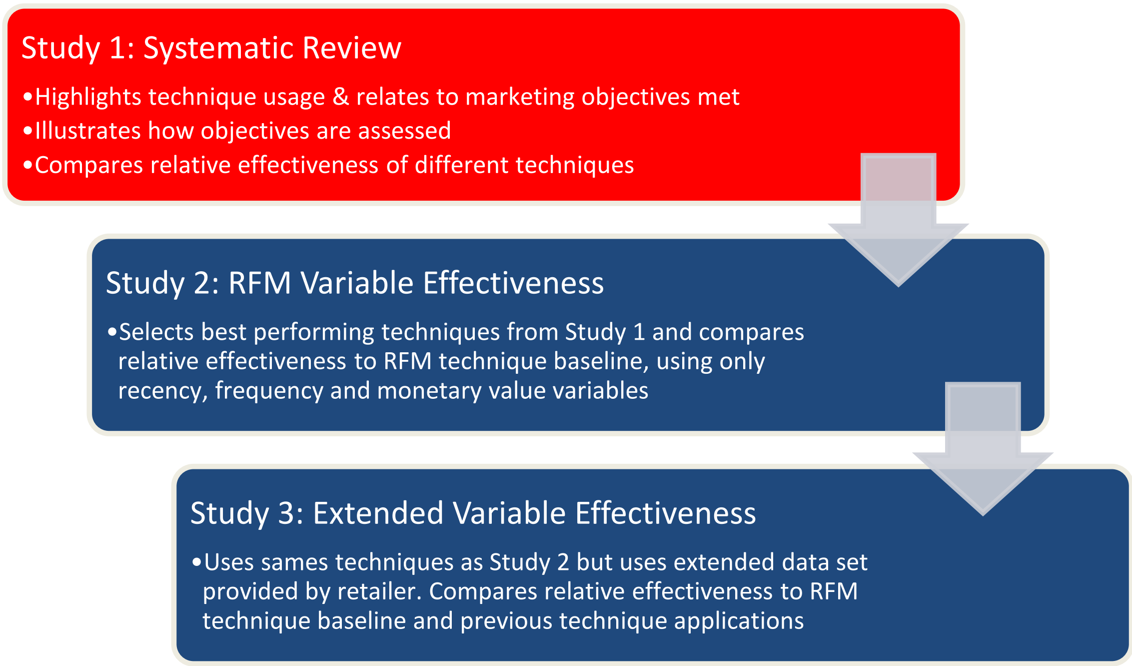
1. *What are the objectives pursued by customer selection?*
2. *In what phases of Data Mining has most of the research on customer selection been conducted?*
3. *What selection techniques are most often researched empirically and used by practitioners?*
4. *How do selection techniques relate to specific marketing objectives?*
5. *How is the attainment of these objectives assessed by researchers?*
6. *What is the empirical evidence comparing the effectiveness of customer selection techniques?*

4 Study 1: Systematic Literature Review

4.1 Introduction

Given the amount of available literature on the topic, the first study, in the form of a systematic literature review, researches some key questions informing the research topic in addition to establishing a strong baseline of information required to adequately design the empirical research studies that follow. Studies two and three subsequently take the answers to the systematic review questions and empirically test relevant techniques using validated measures in the context of FMCG retail. Figure 11 illustrates how each study informs the next.

Figure 11: Thesis Research Progression Overview



4.2 Methodology

Hart (2006, p. 13) defines a literature review as “the selection of available documents on the topic, which contains information, ideas, and evidence written from a particular standpoint to fulfil certain aims or express certain views on the nature of the topic and how it is to be investigated, and the effective evaluation of these documents in relation to the research being proposed.” There are different types of reviews available to social scientists: (1) traditional or narrative, (2) systematic, (3) meta-analysis and (4) meta-synthesis. Cronin et al. (2008) define each type as follows:

- **Traditional or narrative literature review** “critiques and summarises a body of literature and draws conclusions about the topic in question. The body of literature is made up of the relevant studies and knowledge that address the subject area. It is typically selective in the material it uses, although the criteria for selecting specific sources for review are not always apparent to the reader.” (Cronin et al., 2008, pp. 38-39)

- **Systematic review** "provides as complete a list as possible of all the published studies relating to a particular subject area. While traditional reviews attempt to summarise results of a number of studies, systematic reviews use explicit and rigorous criteria to identify, evaluate critically, and synthesise all the literature on a particular topic." (Cronin et al., 2008, pp. 38-39)
- **Meta-analysis** "is the process of taking a large body of quantitative findings and conducting statistical analysis in order to integrate those findings and enhance understanding. Meta-analysis is seen as a form of systematic review, which is largely a statistical technique. It involves taking the findings from several studies on the same subject and analysing them using Standardised statistical procedures." (Cronin et al., 2008, pp. 38-39)
- **Meta-synthesis** "is a non-statistical technique used to integrate, evaluate and interpret the findings of multiple qualitative research studies. Such studies may be combined to identify their common core elements and themes. Findings from phenomenological, grounded theory or ethnographic studies may be integrated and used. Unlike meta-analysis, where the ultimate intention is to reduce findings, meta-synthesis involves analysing and synthesising key elements in each study, with the aim of transforming individual findings into new conceptualisations and interpretations." (Cronin et al.,, 2008, pp. 38-39)

Whilst many social sciences researchers adopt the traditional literature review approach as the mainstay of their thesis work, I elected to conduct both a traditional and a systematic review because of the inherent size of the field of literature and because of the inherent lack of systematic research in the field. The literature review allowed me to scope out the broad literature while the systematic review allowed me to focus on specific questions and topics informed by the literature review. Typically the systematic review is adopted for the following reasons:

- The systematic review identifies relevant literature in a replicable and transparent manner thereby avoiding the selection bias inherent in traditional reviews
- A meta-analysis approach is impractical if not impossible to apply, and though some research is quantitative in nature, datasets for each piece of work are not available for comparison, different types of selection approaches exist in each article, contexts differ dramatically, and/or approaches to measurement vary.
- Unlike the meta-synthesis, the objective of the research is not to achieve a conceptualisation and interpretation of widespread concepts but rather to achieve an empirical understanding of the specific performance of a specific phenomenon of interest.

A systematic literature review "adheres closely to a set of scientific methods that explicitly aim to limit systematic error, mainly by attempting to identify, appraise and synthesise all relevant studies in order to answer a particular question or set of questions" (Petticrew and Roberts, 2006, p. 9). Given the extensive and disorganised state of research on the topic, a systematic review approach allowed me to answer the research question and sub-questions in a transparent, comprehensive and organised fashion. It also adheres to high standards and meets scholarly requirements for peer reviewing academic studies.

4.2.1 Steps of the Systematic Literature Review

Petticrew and Roberts (2006) outline the steps of the systematic review process as: (1) defining the question, (2) drawing up a steering or advisory group, (3) writing a protocol and having it reviewed, (4) carrying out the literature search, (5) screening the references, (6) assessing the remaining studies against the inclusion/exclusion criteria, (7) data extraction, (8) critical appraisal, (9) synthesis of the primary studies, (10) considering the effects of publication bias, and other internal and external biases, (11) writing up the report, and (12) wider dissemination. The methodology for the relevant step is detailed in the following sections, and findings are shown in the Study 1 section.

4.2.2 Defining the Question(s)

A systematic review begins with the identification of specific questions that can be answered from the literature (Pettigrew and Roberts, 2006). The answers establish the broad state of knowledge from extant research and subsequently direct the development of the empirical research design aimed at addressing the identified gaps. This review asks a specific question of the literature in order to establish a systematic presentation of the findings and appropriately scope out the research topic. This systematic review question is the following:

What is the empirical evidence for the effectiveness of Data Mining techniques for customer selection in Database Marketing?

This review question is itself subdivided into six sub-questions:

1. *What are the objectives pursued by customer selection?*
2. *In what phases of Data Mining has most of the research on customer selection been conducted?*
3. *What selection techniques are most often researched empirically and used by practitioners?*
4. *How do selection techniques relate to specific marketing objectives?*
5. *How is the attainment of these objectives assessed by researchers?*
6. *What is the empirical evidence comparing the effectiveness of customer selection techniques?*

4.2.3 Advisory Panel and Protocol

A panel of experts was selected to audit and advise the review process, and to provide expertise in the following areas: (1) RM, CRM Database Marketing and Data Mining, (2) the systematic review process and (3) academic database searching. The experts in the fields of marketing, CRM and segmentation consisted of:

Dr. Stan Maklan, Reader in Strategic Marketing at Cranfield University and thesis supervisor; Professor Hugh Wilson, Professor of Strategic Marketing at Cranfield University; and Professor Simon Knox, Emeritus Professor of Strategic Marketing at Cranfield University.

The expert selected for the process of conducting the systematic review was Professor Richard Wilding, subject-matter expert in systematic reviews. The expert in literature research was Ms. Heather Woodfield, Information Specialist at the Cranfield University King's Norton Library.

A systematic research protocol essentially outlining the methodology stated herein was presented to the expert panel on June 30, 2009, and was approved with minor corrections that have been integrated.

4.2.4 Search Strategy

From an initial scoping review, keywords and search terms relating to the review objectives were identified, compared and cross-referenced to extant literature for completion. Search strings were then created and applied to the following sources: peer-reviewed academic journals, books, and conference proceedings. Thirdly, studies judged useful by the researcher were added using selected academic work references. Finally, panel recommendations were added to the search output. All additional papers that hadn't been captured by the original search were added to the full research output and assessed as if they had been captured originally. Each of these steps is defined below.

Key Words and Search Strings Auditability

A research protocol detailing the search strategy was reviewed with the panel, at which point search terms were reviewed and refined. The keywords and associated cross-referenced search strings are summarised in Table 25. They include truncations and wild card characters to ensure maximum inclusion of relevant terminologies. Keyword searches were restricted to abstracts only, as searching entire texts extended the search beyond the immediate relevance of the search question. Furthermore, searches beyond the abstract were ineffective as bibliographic and full-text electronic resources were not capable of tracking the large number of research results.

Table 25: Construct Keywords

Construct	Keywords
Data Mining Techniques	Data min*, quantitative analys* , quantitative Method* , quantitative Technique* , quantitative modelling, quantitative modeling statistical analy*, statistical method*, statistical technique* , statistical model*, predictive analy*, predictive method*, predictive technique* , predictive model*, database analy*, database method*, database technique* , database model*, choice model*, customer base analy*, customer base model*, machine learning analy*, machine learning method*, machine learning technique* , machine learning model*, segmentation analy*, segmentation method*, segmentation technique*, segmentation model*, clustering analys*, clustering method*, clustering technique*
Direct/ Database Marketing	Database Marketing, direct marketing, one-to-one marketing, target marketing, segment marketing, direct mail, e-mail marketing, functional CRM
Customer Selection	Customer selection, consumer selection, customer segmentation, consumer segmentation, list selection
Effectiveness	effective* OR effectual* OR efficacious* OR efficien* OR productive*

Since the output of a single combined search string generated too many results for bibliographic and full-text electronic resources, the most popular construct, data mining, was decomposed into four manageable combinations based on feedback from the research panel. The final four search strings are detailed in Table 26.

Table 26: Search Strings

Construct Strings	Data Mining Techniques	Direct and Database Marketing	Effectiveness
1	Data-min* OR quantitative analys* OR quantitative Method* OR quantitative Technique* OR quantitative model* OR database analy* OR database method* OR database technique* OR database model*	Database Marketing OR direct marketing OR one-to-one marketing OR target marketing OR segment marketing	effective* OR effectual* OR efficacious* OR efficien* OR productive*
2	statistical analy* OR statistical method* OR statistical technique* OR statistical model* OR predictive analy* OR predictive method* OR predictive technique* OR predictive model* OR choice model*	Database Marketing OR direct marketing OR one-to-one marketing OR target marketing OR segment marketing	effective* OR effectual* OR efficacious* OR efficien* OR productive*
3	machine learning analy* OR machine learning method* OR machine learning technique* OR machine learning model*	Database Marketing OR direct marketing OR one-to-one marketing OR target marketing OR segment marketing	effective* OR effectual* OR efficacious* OR efficien* OR productive*
4	segmentation OR cluster*	Database Marketing OR direct marketing OR one-to-one marketing OR target marketing OR segment marketing	effective* OR effectual* OR efficacious* OR efficien* OR productive*

Databases and Electronic Journals

Once key terms and search strings were accepted by the panel, a search of the key databases and electronic journals, highlighted in Table 27, was conducted.

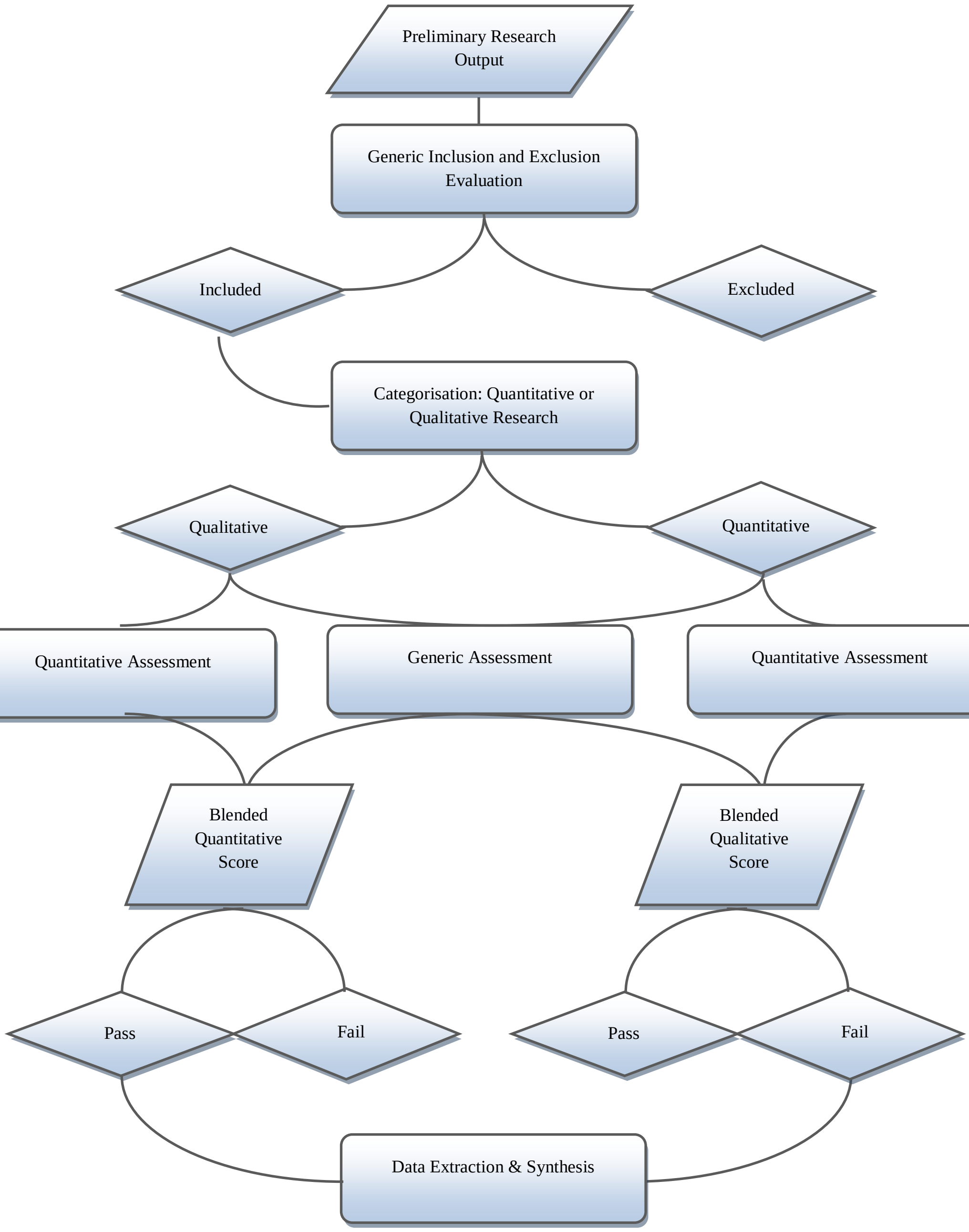
Table 27: Databases and electronic journals

Database	Description & Explanation
Google Scholar	Unlike a conventional Google search, GS searches only academic papers, many of which may be held in institutional repositories, and so would not be included in conventional databases. Google Scholar is particularly useful for articles which are difficult to track down using other sources, or for non-journal formats such as working papers of conference proceedings.
ABI/INFORM (ProQuest)	One of the most comprehensive sources for articles related to business and management, ABI can be searched by keyword, either in the full text of the article, the title and abstract, or in a subject indexing field, if you need a more precise search. The database goes back to 1970, but one can narrow or broaden your results by changing the time period, or by selecting scholarly journals only.
Business Source Premier (EBSCO)	This major source for business and management will be used alongside ProQuest for any management-related search. There is a degree of overlap between the two, but EBSCO sometimes contains complementary journals or articles not accessible through ProQuest
IEEE	IEEE Xplore is a more technical database that will complement the management focused databases
Scopus	Scopus is another technical database that will complement the management focused databases. In addition, Scopus will be used to track forward and backward through time literature from seminal articles.

Selection Criteria

Following the selection of literature based on the construct of keywords and search strings, the subsequent step involved screening results for relevance. Search results were first screened against inclusion and exclusion criteria in order to provide a first level of filtering. Subsequently, remaining articles were assessed based on the quality of the content. Figure 12 illustrates this process.

Figure 12: Research Selection Process



Inclusion and Exclusion Criteria

Once the full set of references was identified, references were summarily evaluated for relevance. The criteria for inclusion in this first step of the selection process were: topic relevance, context relevance, language, and time of publication. Given the manageable list issued from the application of the search strings, and given the large number of low impact journals and practitioner journals that emerged from the search, no exclusions were applied for the impact factor of journals. Rather, this assessment was deferred to the quality appraisal phase to ensure that relevant data in lower-impact publications was not omitted. Table 28 illustrates the inclusion/exclusion criteria as well as a detailed description. To counter the criticism that the systematic review process can be overly mechanistic, articles of relevance that did not qualify for the review list from the start were added and a similar quality assessment was conducted (Dixon-Woods, Bonas, Booth, Jones, Miller, Sutton and Young, 2006). These articles were mainly referred by my advisory panel.

Table 28: Inclusion and Exclusion Criteria

Criteria	Description
Topic Relevance	Only research dealing specifically with customer selection methods in direct marketing was retained. This allowed for the elimination of all literature around mass marketing and other non direct methods of communications that are out of scope for this systematic review. More specifically, papers that contained references to customer selection: • Methods • Moderators (including bases of customer selection) • Measures of effectiveness
Context relevance	Only research relating to business to consumer direct marketing promotions was retained in order to provide as much contextual relevance as possible. This exclusion criterion more explicitly ensures the elimination of completely unrelated contexts such mass market promotions, other types of non-targeted promotions and business to business promotions.
Language	Searches were limited to material published in English.
Time of publication	Customer selection, segmentation, and modelling have a long research tradition. However, given advancements in IS, statistical methods and CRM, with limited exceptions, only studies published after 1990 were retained.

Quality Appraisal Criteria for Full/Empirical Papers

Engel and Kuzel (1992, p. 505) state that “systematic reviews, due to their positivistic origins, sit comfortably with studies that use quantitative methods… therefore establishing criteria for ascertaining what is relevant or good quality.” However, as mentioned earlier, though many are time based on quantitative research, tracing quality back to the original data source is frequently difficult and assessment of quality is often based on rather subjective conclusions (Tranfield, Denyer and Smart, 2003).

Nevertheless, in an attempt to remain as transparent as possible, selected studies were classified into quantitative studies and qualitative studies. As both classes of selected studies retained are related to the research question, I used Huff's (2007) assessment grid as a base to evaluate all studies (Table 29).

The content of quantitative and qualitative studies should be assessed using different quality criteria (Huff, 2007). These are identified in Table 29 as well and were marked using a five-point Likert scale.

Each set of quality criteria is judged equally important to the evaluation of quality; each set was therefore weighted 50 percent. Cut-off levels for articles to be excluded were established after scoring the entire literature set instead of using a predetermined cut-off level a priori. This recommendation was made by the research panel.

Table 29: Systematic Review Research Questions

General Research Questions
Degree of relevance to field of enquiry
Research methodology – clarity of research objectives, assumptions, findings, limitations
Quality of discussion
Recommendations for future research.
Number of citations (secondary measure)
Does the research link techniques to outcomes?
Does the study provide the reader with a sense of the relative effectiveness of techniques discussed?
Context specificity (How context specific are the research findings?)
How robust does the data-set that is referred to seem?
What is the degree of expertise of the authors conducting the study?
What is the degree of managerial relevance of findings?
To what degree do the findings allow managers to determine which selection techniques they should use? Which is most effective?

1=Not at all. 2=Only to a limited extent. 3=At an acceptable level. 4=To a significant level. 5=Completely

Source: Huff (2007, p.158)

Detailed Questions for Quantitative Papers
Questions for Quantitative Papers
Theoretical framework and development of hypotheses (if appropriate)
Are the study's propositions and hypotheses clearly articulated?
Are the basic arguments of the paper important and interesting?
Are important premises & assumptions identified?
Is there a graphic depiction of the relationship between key variables in the paper?
Are the key terms identified?
Description and evaluation of the methods if appropriate
Description & evaluation of methods (if appropriate)
Is the methodology of the paper clearly identified?
Are data collection methods described adequately?
Are the sampling strategy and sample explained?
Is the operationalisation of the variables and constructs plausible (content validity)?
Are dependent variables identified and described?
Are independent variables identified and described?
Are control variables identified and described?
Are questionnaire or other measurement items identified and described?
Was the discussion of the interview or questionnaire construction and response rates clear and comprehensive?
Have steps been taken to avoid data collection errors?
Is there evidence of reliability or internal consistency in the study?
Results
Are the findings adequately and accurately described?
Are results clearly related back to original propositions, hypotheses, research questions, and data analysis?
Do tables provide sufficient and accurate data to allow the reader to reach independent conclusions?
Are figures and appendices used effectively?
Is implied causality justified?
Has the author adequately considered alternative explanations for the results found?

1=Not at all. 2=Only to a limited extent. 3=At an acceptable level. 4=To a significant level. 5=Completely

Source: Huff (2007, p.158)

Questions for Qualitative Papers
Is the purpose of the research adequately established?
Are the duration and intensity of observation clear?
Are the nature of the sit, and key players, adequately discussed?
Are methods and collecting and Analysing of data adequately described?
Does the writer convince the reader that he or she was able to gather information about key events from appropriate sources?
Is there evidence that informants trusted the researcher and were likely to honestly share information with the researcher?
Has the author adequately considered alternative interpretations of the data presented?
Is there evidence of systematically considering evidence that contradicts the author's interpretations/
Has the author adequately considered alternative interpretations of the data presented?
Is there evidence of systematically considering evidence that contradicts the author's interpretations?

1=Not at all. 2=Only to a limited extent. 3=At an acceptable level. 4=To a significant level. 5=Completely

Source: Huff (2007, p.158)

4.2.5 Data Extraction

Referencing of literature was done using Refworks, but the summarisation, cataloguing and critiques were captured via Microsoft Access using a bespoke format designed for this thesis. All evaluation variables were treated as separate fields in the database to provide full flexibility and comparability as required. The data extraction form in Table 30 was used for organisation of systematic review information.

Table 30: Data Extraction Table

Category	Author(s)
Citation Information	Title
	Periodical Full
	Publication Year
	Authors, Primary
	Abstract
	Access Key #
	Refworks ID
	Reasons for Inclusion
	Journal Type (Academic, Practitioner)
	Peer Reviewed
Study Background	Type of literature (theoretical, practice, research, policy)
	Type of Intellectual Project (knowledge for understanding, critical evaluation, action, instrumentalism, reflexive action)
	Study Methodology (empirical, theoretical)
	Research paradigm (positivist, realist, interpretivist)
Thematic information	Key concepts, ideas, theories, approaches
Methodology	
Study Methodology	
Data Collection Methodology	Sample selection process
	Sample size
	Composition and potential bias
	Methods applied: Experiment, cross-sectional analysis, simulation
Context	
Industry Context	B2B FMCG/Grocery Retail, B2C General Retail, B2B, Other Other (manual entry)
Evidential Contribution	
Research Question	What is the research question?
	What is the customer selection issue addressed by the research?
Research Hypothesis	What is the research hypothesis?
	What is the underlying customer selection hypothesis?
Methodology	Data source
	Sample size
	Sample composition
	Data analysis conducted
Objectives	Unit of analysis
	Objectives pursued by application of techniques
Measures	Metrics used to measure performance effectiveness
	Metrics used to measure fit
Limitations	What are the stated limitations as they relate to the topic of customer selection?
	What are the unstated limitations?
Techniques	Techniques evaluated by authors
	Technique class
	Technique type (according to Wedel and Kamakura, 2000)
	Effectiveness ranking
Bases/Data Variables of customer selection	Bases used
	Base type (according to Mama, 2007)
	Base type (according to Wedel and Kamakura, 2000)
Moderating Factors	What are the organisational moderating factors? How can they potentially impact the research findings?
	What are the consumer moderating factors? How can they potentially impact the research findings
	Are there any other moderating factors?
Main Findings	What are the main findings?
	What are the main findings that related to customer selection?
	How can I use this research?
Recommendations for Future Research	
Limitations	Stated limitations
	Unstated limitations

4.2.6 Synthesis

Given the review questions cover the topic of effectiveness of selection techniques in database marketing, the synthesis was conducted using cross-tabulations of the key information contained in the extraction tables and was related back to the literature for secondary validation. More specifically, by sub-question, a number of actions were taken:

1. What objectives are pursued? Overall distribution of objectives was assessed to provide a sense of research intensity. This is detailed in section 4.5.

2. What selection techniques are used? Selection techniques, categories and Data Mining phases of application were extracted and distribution tracked to provide a sense of research intensity. This is detailed in sections 4.4 and 4.6.

3. How do selection techniques relate to objectives? Objectives were cross-tabulated with different selection techniques, categories and phases of application to illustrate their relationship. This is detailed in section 4.7.

4. How is effectiveness assessed? Each study's measures (both for predictive accuracy and performance) were tracked and cross-tabulated to relate back to objectives, selection techniques, categories and phase of application. This is detailed in section 4.8.

5. What is the comparative effectiveness of selection techniques? To generate a transferable program theory of what works, for whom and in what contexts, selection techniques and categories were first categorised within the phase of application and second ranked by objectives. This is detailed in section 4.9.

The accumulation of understanding across a range of studies provided by the synthesis tables allowed the completion of the literature review using a narrative synthesis approach to allow for a greater socialisation of findings. This synthesis acted as the base with which to:

(1) select the techniques to be compared in empirical studies;
(2) ensure techniques met the stated research objective(s);
(3) validate the metric to be used and independent variables for effectiveness measurement;
(4) understand how to measure and compare effectiveness.

4.3 Systematic Review Search Output

The overall search using the aforementioned search strings generated a significant output of related content. A total of 558 references were identified. These references can be categorised as follows:

- 353 practitioner journal articles and white papers
- 205 academic references
- 26 non-peer-reviewed references such as conference proceedings, unpublished articles, and theses
- 179 academic peer-reviewed journal articles

In this first phase of assessment, of the 558 retained papers, 418 were excluded and 140 were retained. Since English was set as the only search language, none of the 558 papers had to be eliminated because of language. However, of the excluded papers, 140 were eliminated because of topic irrelevance, 147 because of context irrelevance, and 131 because they were written prior to 1990.

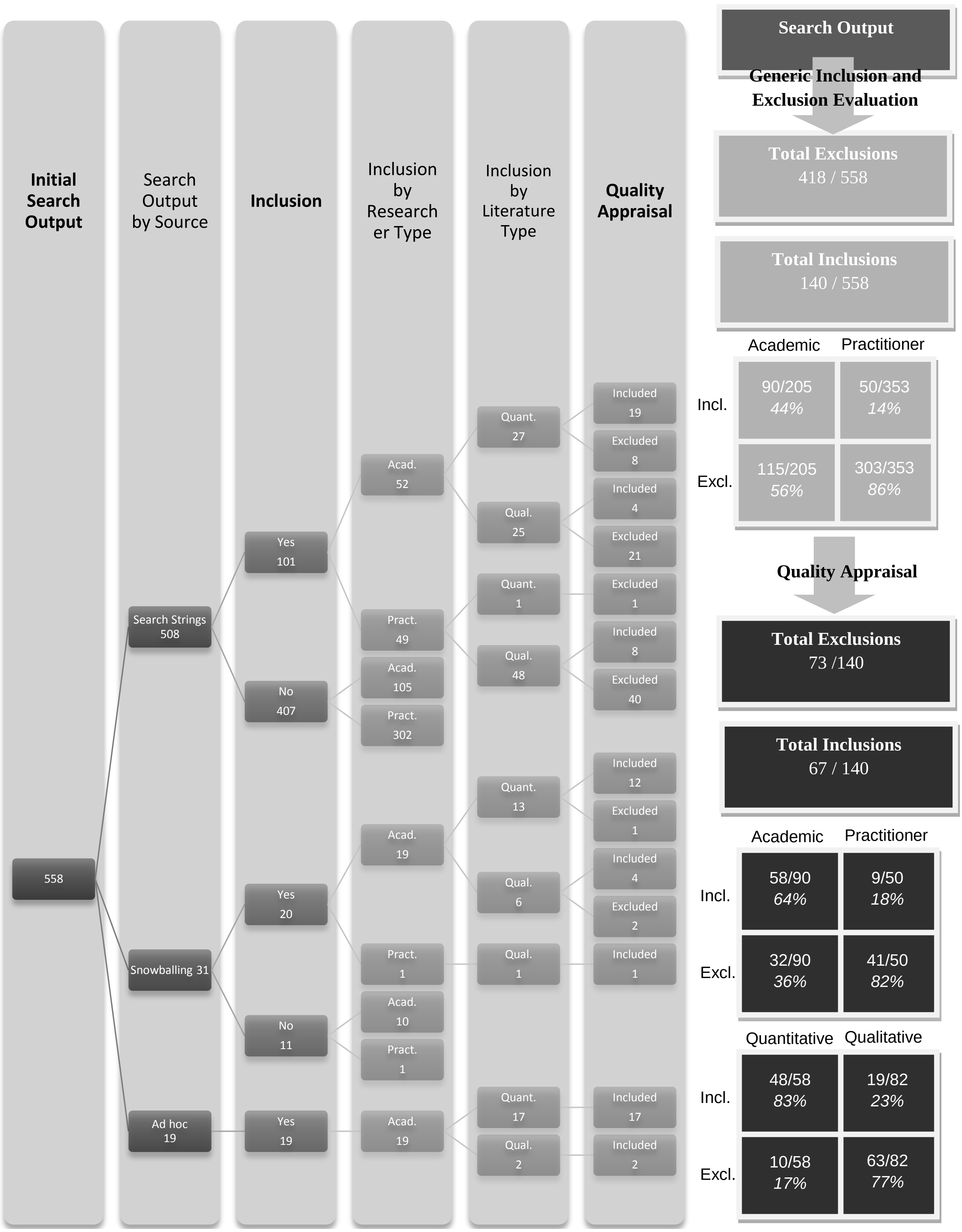
Given exclusions eliminated all topic- and time-irrelevant papers; of the retained papers, all had to reference techniques, moderators or measures of effectiveness. Given these were not mutually exclusive, the sum of their reasons for inclusion is greater than 140. As a result, 75 covered specific customer selection techniques instead of dealing with the topic on aggregate, 76 discussed some moderating factors, and 47 included a presentation of some form of measure of effectiveness.

Once scored, a distinct difference existed between high- and low-scoring papers. Of reviewed studies, 51 percent scored between zero and two, 13 percent between two and three and 36 percent scored between three and five. Given this outcome, only articles scoring three and more were retained. A high proportion (61 percent) of excluded content was practitioner literature. Of the practitioner literature, only 2 percent were assessed as relevant to the review. As a result, this literature was eliminated from the review because of its non-empirically supported positions. Finally, quantitative literature contributed 83 percent of content included in the systematic review, whereas the more conceptual, qualitative literature contributed 23 percent.

Details of this search and its outcome can be found in Figure 13.

Subsequent sections will take this reduced output from the systematic review and categorise literature by Data Mining phases applied, pursued objectives, applied techniques, measures of effectiveness and performance used, and data variables utilised. I will also examine the comparative effectiveness of different Data Mining techniques and of different data variables and depths of data.

Figure 13: Systematic Review Literature Inclusion Flowchart



4.4 Data Mining Phases

Section 3.6.4 identifies seven phases of data mining: understanding, selection, pre-processing, transformation, pattern discovery and modelling, interpretation/evaluation and incorporation. Though most of this research is concerned with the fifth phase of pattern discovery and modelling, it remains that there is evidence to indicate that researchers integrate other phases of the process as Data Mining inputs (Abbott, Stone and Buttle, 2001; Deichmann et al., 2002; Shih and Liu, 2003).

Categorising studies was conducted using a simple interpretive assessment of phase applications within each individual study. This was captured in the same Microsoft Access forms and tables designed for the systematic review information logging. Given the systematic review's main objective is the identification of customer selection techniques for purposes of database marketing, it was expected that a disproportionate number of studies would fall into the pattern discovery and modelling phase. This outcome was seen as being acceptable given the pursued objective. However, in a minority of studies, the pattern discovery and modelling phase was also complemented by the application of selection, pre-processing and transformation phases.

Of the 33 retained studies, 25 fell squarely into the pattern discovery and modelling phase, while a minority of studies focused solely on pre-processing and transformation. Five studies conducted the research in a more holistic fashion by applying both the Transformation and Pattern Discovery and Modelling phases. Table 31 provides a breakdown of how many studies within the list of studies retained from the systematic review fall into each phase of data mining. Studies that include multiple phases of Data Mining are presented using italics. Albeit biased because of the modelling phase orientation of the research objective, this outcome is nevertheless consistent with Shaw et al.'s (2001) observation that most research in Data Mining has been found to be mainly focused on the process of pattern discovery.

Table 31: Data Mining Phase Applications

Data Mining Phases	# of Studies
Selection	0
Pre-Processing	1
Transformation	2
Pattern Discovery & Modelling	25
Transformation AND Pattern Discovery & Modelling	*4*

Source: From 33 studies explicitly addressing data mining techniques

Given the focus of the thesis lies in the techniques applied in the pattern discovery and modelling phase, these are discussed in detail in section 4.6. However, in light of the important role of transformation issued from the review, the remainder of this section will discuss the transformation phase's key outcomes vis-à-vis current literature.

From the review of relevant literature, I identify seven techniques used in the transformation phase to overtly weight, modify or reduce variables (also known as data or dimensionality reduction). These techniques include: analytical hierarchical programming (AHP), multiple adaptive regression splines (MARS), partial least squares (PLS) regression, principle component analysis (PCA), sliced inverse regression, vector quantization (VQ) algorithms, and factor analysis. AHP was used to appropriately weight the individual RFM variables. MARS was used to transform independent variables into basis functions that consider non-linearity and variable interactions that facilitate the modelling phase. Principle component analysis (PCA), partial least squares (PLS), sliced inverse regression and factor analysis were used mainly for dimensionality reduction. Finally, ridge regression was used as an alternative to standard data reduction methods to manage the issue of multicollinearity. As discussed below, in these studies, appropriate weighting of variables, variable transformations and dimensionality reduction allowed models to perform better given that selected variables are less correlated and thus provide a better level of predictive accuracy.

There are also cases where the transformation and data-mining phases are not only applied in isolation or in sequence but rather where techniques are applied simultaneously to achieve the objectives both transformation and data mining. In the systematic review, techniques that simultaneously reduced dimensionality and created optimal models include stepwise logistic regression and stepwise linear regression. In both cases, the aim is to reduce dimensionality and deliver enhanced predictive accuracy given the reduction in model noise via the reduction of variable multicollinearity. Deichmann et al. (2002) and Malthouse (2008) both applied the stepwise logistic regression technique for dimensionality reduction purposes. These outputs are shown in Table 32.

Other techniques that also simultaneously reduce dimensionality and optimise model outputs include decision trees (CART, CHAID, C4.5, ID3) and neural networks (Weiss and Kulikowski, 1991). Decision tree outputs and design ultimately lead to an inherent reduction in variables used to optimally categorise groups by response rates. Nowhere is this clearer than in decision tree outputs, where only retained variable nodes are shown along with their respective response rates. Neural network black box design makes such a visualisation impossible; however, just like decision trees and stepwise regression, they meet the dual objectives. All neural networks and decision tree models (CART, CHAID) issued from the systematic review could thus also have been incorporated into the Transformation phase.

For parsimony purposes and because the intent was often unstated, the application of logistic regression, decision tree and neural networks was restricted to the Pattern Discovery and modelling phase. Nevertheless, it remains that they still achieve the dual objectives of both transformation and data mining.

Table 32: Data Transformation Studies

	Year	Transformation Techniques	Data Mining Techniques
Transformation			
Malthouse	1999	Ridge Regression Stepwise Linear Regression	n/a
Naik, Hagerty and Tsai	2000	Partial Least Squares Principal Components Sliced Inverse Regression	n/a
Transformation & Data Mining			
Deichmann, Eshghi, Haughton, Sayek and Teebagy	2002	Multiple Adaptive Regression Splines (MARS)	Stepwise Logistic Regression Stepwise Linear Regression
Shih and Liu	2003	Analytic hierarchy process (AHP)	Clustered RFM
Heilman, Kaefer and Ramenofsky	2003	n/a	Multinomial Logit model and an Artificial Neural Network model
Yang	2004	Factor analysis	RFM, CHAID, Logit
Reutterer, Mild, Natter and Taudes	2006	Vector quantization (VQ) algorithms	Assignment rule

4.5 Objectives of Customer Selection

Though this research is focused on the development objective, I did not limit output from the systematic review search to this objective alone in an a priori fashion. Rather, conducting the review without the limitation to development allowed me to assess the intensity in which objectives are pursued and post-hoc reduced the search set to refocus on development objectives only. This allowed me to validate the work by Ngai (2009) on Data Mining research intensity.

As shown in section 3, RM and CRM each support the development of long-term customer relationships via acquiring, retaining, developing and potentially terminating unprofitable customer relationships. These objectives are entirely supported by the combined objective of CRM, RM, Database Marketing and segmentation stated in the literature review.

The first slight difference between the systematic review categorisation and RM objectives categories was the expanded scope provided to the acquisition category in the systematic review. Given no RM objective captured customer reactivation (the process of re-acquiring a customer after a prolonged period of inactivity), this was categorised into the acquisition (or re-acquisition) objective of RM. The second difference with the RM categorisation of section 3 is the expanded definition of retention used in the systematic review categorisation. The systematic review definition of retention included loyalty and churn management objective, given they were measured using metrics similar to retention and used interchangeably.

Looking at the distribution of objectives pursued within the studies, 28 out of 35 studies were interested in one of the objectives of acquiring, retaining, and developing customers. More specifically, three studies were interested in the acquisition of new customers. Twenty-three studies examined customer development (otherwise referred to as customer penetration). Only one study examined customer retention. The final classification and counts of customer selection objectives is detailed in Table 33.

Three studies specifically examined Customer Lifetime Value (CLV) as an overt objective whereas four studies aimed solely at examining how techniques allowed for the creation of most heterogeneous groups. Segmentation studies used various clustering and segmentation techniques such as Structural Equation Modelling (SEM), logit, and ANN to improve the classification accuracy and heterogeneity of segments. CLV-focused studies used hazard function models (Gönül, Kim and Shi, 2000) and genetic algorithms (Chan, 2008) for CLV maximisation of estimation.

Table 33: Customer Selection Objectives

Customer Selection Objectives	Link to RM & CRM Objectives	# of Studies
New customer acquisition *or reactivation*	Building/attracting/acquiring/establishing/ creating/initiating or getting customers	3
Customer penetration (up-sell, cross-sell)	Developing/enhancing/extending or nurturing customers	23
Retention/Churn & Loyalty	Retaining/maintaining/sup-porting or keeping customers	1
Lifetime value Maximisation	Building long-term relationships	3
Pure Segmentation / Classification/ Analytical profiling	N/A	5
	Terminating or dissolving customer relationships	0

In terms of importance, the topic with the highest degree of interest was customer penetration, or the development of customer relationships. With twenty-three of the total studies examining this topic it is by far the one that garnered the most interest followed by a lagging interest in segmentation, new customer acquisition, CLV maximisation and retention.

4.6 Technique Usage

Only the 33 studies that explicitly addressed the comparative effectiveness of Data Mining techniques were retained to generate the most objective possible findings. Of the retained studies, the majority were quantitative in nature. Table 35 provides an overview of these 33 studies along with the techniques used, the objectives pursued, the phases of Data Mining where techniques are applied, and the relative performance of the techniques.

The results show a greater use of statistical techniques over machine learning ones by a factor of almost two to one. Within the statistical techniques, RFM (or its iterations), cluster analysis and logistic regression are the most often used, though the class of regression analysis techniques also shows varied usage. Within the machine learning techniques, neural networks and AID/CHAID are most often used. Table 34 illustrates the count and distribution of techniques as a percentage of studies in the systematic review.

A similar exercise is conducted in Table 36; however, instead of assessing the percentage of studies where each technique is applied, I examine the distribution of technique applications as a percentage of total technique applications. Results are fairly similar to the above but also provide a split of techniques between statistical and machine learning categories. The results show that 60 percent of applied techniques fall into the statistical category while 40 percent fall into the machine learning category. Of statistical techniques, as referenced above, regression analysis represents that bulk of applications with nearly 40 percent of techniques being represented by either logistic regression or multiple linear regression applications, and a small percentage being represented by cluster analysis. Within the regression sub-category, logistic regression applications are dominant with more than one-third of all regression usage. Machine learning applications are split more evenly between neural networks and decision trees. In fact, decision-tree applications represent 16 percent of applications whereas neural networks represent 15 percent. Of decision-tree applications, it is noteworthy to highlight the dominant representation of the AID/CHAID technique with 11-percent representation.

Table 34: Distribution of Data Mining Techniques

Technique	Technique Count	% of Studies
Logistic/Logit Regression	15	45%
AID/CHAID	11	33%
Multiple/Linear regression	8	24%
Neural Networks/ANN	8	24%
CART (Classification and Regression Trees)	5	15%
RFM	5	15%
Discriminant analysis	3	9%
Log-linear	3	9%
Latent Class Models	3	9%
Bayesian Networks	2	6%
Cluster Analysis	2	6%
finite mixture structural equation models	1	3%
VQ (K-means)	2	6%
Tobit	1	3%
Two-stage models	1	3%
Ridge Regression	1	3%
Partial Least Squares Regression (PLSR)	1	3%
Principal Components Regression (PCR)	1	3%
Sliced Inverse Regression (SIR)	1	3%
FRAC-Frequency, Recency, Amount, Category	1	3%
Genetic Algorithms	1	3%
Cross-tabulation	1	3%
Multiple Adaptive Regression Splines-MARS)	1	3%
AHP	1	3%
Markov Chains/Markov Chain Monte Carlo	1	3%
Automatic Relevance Detection (ARD)	1	3%
Random Forests	1	3%
Bagging Neural Networks	1	3%
ANN/GA	1	3%
PCA/Logit	1	3%
Support Vector Machines	1	3%
Crossed-basis sub-segmentation	1	3%
Sum	**87**	

Table 35: Study Overview

Authors, Primary	Year	*Main objectives pursued by methods*	*Phase of Data Mining*	*Techniques evaluated*	*Method Performance Ranking based on cumulative lift (or similar metrics) of top 5 deciles*				
					Best Method	**2nd Best**	**3rd Best**	**4th Best**	**5th Best**
Magidson	**1988**	Customer penetration (upsell, cross-sell)	Data Mining	AID/CHAID ; Discriminant analysis; Logistic/Logit Regression; Log-linear; Multiple regression	AID/CHAID; Logistic/Logit Regression; Log-linear	AID/CHAID (Classification Tree); Discriminant analysis; Multiple regression			
Thrasher	**1991**	Customer penetration (upsell, cross-sell)	Data Mining	CHAID CART	CHAID	CART			
Bult and Wansbeek	**1995**	Customer penetration (upsell, cross-sell)	Data Mining	AID/CHAID ; Gains chart analysis; PM approach	PM approach	Gains chart analysis	AID/CHAID (Classification Tree)		
Lix, Berger and Magliozzi	**1995**	New customer acquisition or reactivation	Data Mining	Log-linear; Multiple regression; Stepwise Linear Regression	Multiple regression; Stepwise Linear Regression	Log-linear			
Haughton and Oulabi	**1997**	Customer penetration (upsell, cross-sell)	Data Mining	CART CHAID	CART CHAID				
Jedidi, Jagpal and DeSarbo	**1997**	Pure Segmentation	Data Mining	finite mixture structural equation models; cluster analysis	Finite mixture structural equation models	Cluster analysis			
West, Brockett and Golden	**1997**	Customer penetration (upsell, cross-sell)	Data Mining	Discriminant analysis; Logistic/Logit Regression; Neural Networks/ANN	Neural Networks/ANN	Discriminant analysis; Logistic/Logit Regression			

Authors, Primary	Year	*Main objectives pursued by methods*	*Phase of Data Mining*	*Techniques evaluated*	*Method Performance Ranking based on cumulative lift (or similar metrics) of top 5 deciles*				
					Best Method	**2nd Best**	**3rd Best**	**4th Best**	**5th Best**
Zahavi and Levin	**1997**	Customer penetration (upsell, cross-sell)	Data Mining	Logistic/Logit Regression; Neural Networks/ANN	Logistic/Logit Regression; Neural Networks/ANN				
Levin and Zahavi	**1998**	Customer penetration (upsell, cross-sell)	Data Mining	Tobit Logit Linear Regression Two-stage models	Tobit Two-stage models	Logit	Linear Regression		
Malthouse	**1999**	Customer penetration (upsell, cross-sell)	Transformation	Ridge Regression (adapted multiple linear regression); Stepwise Linear Regression	Ridge Regression (adapted multiple linear regression)	Stepwise Linear Regression			
Naik, Hagerty and Tsai	**2000**	Pure Segmentation	Transformation	Discriminant analysis; Multiple regression; Partial Least Squares Regression (PLSR); Principal Components Regression (PCR); Sliced Inverse Regression (SIR)	Sliced Inverse Regression (SIR)	Partial Least Squares Regression (PLSR)	Principal Components Regression (PCR)	Multiple regression	Latent Root Regression
Levin and Zahavi	**2001**	Customer penetration (upsell, cross-sell)	Data Mining	AID/CHAID (Classification Tree); FRAC-Frequency, Recency, Amount, Category; Genetic Algorithms; Logistic/Logit Regression; RFM	Logistic/Logit Regression	AID/CHAID (Classification Tree); Genetic Algorithms	FRAC-Frequency, Recency, Amount, Category	RFM	

Authors, Primary	Year	*Main objectives pursued by methods*	*Phase of Data Mining*	*Techniques evaluated*	*Method Performance Ranking based on cumulative lift (or similar metrics) of top 5 deciles*				
					Best Method	**2nd Best**	**3rd Best**	**4th Best**	**5th Best**
Ratner	**2001**	Customer penetration (upsell, cross-sell)	Data Mining	Cross-tabulation CHAID	CHAID	Cross-Tabulation			
Deichmann, Eshghi, Haughton, Sayek and Teebagy	**2002**	Customer penetration (upsell, cross-sell)	Transformation; Data Mining	Multiple Adaptive Regression Splines-MARS); Stepwise Linear Regression; Stepwise Logistic Regression	Multiple Adaptive Regression Splines-MARS)	Multiple Adaptive Regression Splines-MARS); Stepwise Linear Regression	Logistic/Logit Regression; Multiple Adaptive Regression Splines-MARS)	Logistic/Logit Regression	
Shih and Liu	**2003**	Lifetime value maximisation	Transformation; Data Mining	AHP Weighted Clustered RFM; RFM	AHP Weighted Clustered RFM	RFM			
Linder, Geier and Kolliker	**2004**	Customer penetration (upsell, cross-sell)	Data Mining	AID/CHAID; Logistic/Logit Regression; Multiple regression; Neural Networks/ANN	Model combination; Multiple regression	Neural Networks/ANN	Logistic/Logit Regression	AID/CHAID (Classification Tree); Neural Networks/ANN	
Cui and Wong	**2004**	Customer penetration (upsell, cross-sell)	Data Mining	Bayesian Networks; CART (Classification and Regression Trees); Latent Class Models; Neural Networks/ANN	Bayesian Networks; Latent Class Regression	Latent Class Models	Neural Networks/ANN	CART (Classification and Regression Trees)	
Jonker, Piersma and Van Den Poel	**2004**	New customer acquisition or reactivation	Transformation; Data Mining	AID/CHAID (Classification Tree); Markov Chains/Markov Chain Monte Carlo (MCMC); RFM	Markov Chains/Markov Chain Monte Carlo (MCMC); RFM	AID/CHAID (Classification Tree)			

Author, Primary	Year	*Main objectives pursued by methods*	*Phase of Data Mining*	*Techniques evaluated*	*Method Performance Ranking based on cumulative lift (or similar metrics) of top 5 deciles*				
					Best Method	**2nd Best**	**3rd Best**	**4th Best**	**5th Best**
Yang	**2004**	Customer penetration (upsell, cross-sell)	Data Mining	RFM CHAID Logit	CHAID Logit	RFM			
Buckinx and Van den Poel	**2005**	Churn Management	Data Mining	Automatic relevance determination (ARD); Logistic/Logit Regression; Random Forests	Automatic relevance determination (ARD); Logistic/Logit Regression; Random Forests				
Jonker, Piersma and Potharst	**2005**	Lifetime value maximisation	Data Mining	Markov Chains/Markov Chain Monte Carlo (MCMC)	NA				
Ho, Cho and MacLachlan	**2005**	Customer penetration (upsell, cross-sell)	Data Mining	Bagging Neural Networks ANN Logit	Neural Networks using bagging technique (BMLP)	Logit ANN			
Kim, Street, Russell and Menczer	**2005**	Customer penetration (upsell, cross-sell)	Data Mining	ANN/GA PCA/Logit	ANN/GA	PCA/Logit			
McCarty and Hastak	**2006**	Customer penetration (upsell, cross-sell)	Data Mining	AID/CHAID (Classification Tree); Logistic/Logit Regression; RFM	AID/CHAID (Classification Tree)	RFM	Logistic/Logit Regression		
Crone, Lessmann, and Stahlbock	**2006**	Customer penetration (upsell, cross-sell)	Preprocessing	AID/CHAID; Neural Networks/ANN; Support Vector Machines	Decision Tree Algorithm; Neural Networks/ANN; Support Vector Machines				

Authors, Primary	Year	*Main objectives pursued by methods*	*Phase of Data Mining*	*Techniques evaluated*	*Method Performance Ranking based on cumulative lift (or similar metrics) of top 5 deciles*				
					Best Method	**2nd Best**	**3rd Best**	**4th Best**	**5th Best**
Cui, Wong and Lui	**2006**	Customer penetration (upsell, cross-sell)	Data Mining	Bayesian Neural Networks CART Latent class model ANN	ANN	Bayesian NN	CART	Latent Class	
Malthouse and Elsner	**2006**	Customer penetration (upsell, cross-sell); New customer acquisition or reactivation	Data Mining	Crossed-basis sub-segmentation	Crossed-basis sub-segmentation	NA			
Reutterer, Mild, Natter, and Taudes	**2006**	Pure Segmentation; Customer penetration (upsell, cross-sell)	Transformation; Data Mining	Vector quantization (VQ) algorithms (K-means);	VQ + Assignment rule	NA			
Greene and Greene	**2008**	Pure Segmentation	Data Mining	Cluster analysis; Clustering with additional variables	Clustering with additional variables	Cluster analysis			
Chan	**2008**	Lifetime value maximisation	Data Mining	Genetic Algorithms	Genetic Algorithms				
Cui, Wong, Zhang and Li	**2008**	Customer penetration (upsell, cross-sell)	Data Mining	Bayesian Neural Networks CART Latent class model Logit	Bayesian NN	Logit	CART	Latent Class	

Authors, Primary	Year	*Main objectives pursued by methods*	*Phase of Data Mining*	*Techniques evaluated*	*Method Performance Ranking based on cumulative lift (or similar metrics) of top 5 deciles*				
					Best Method	**2nd Best**	**3rd Best**	**4th Best**	**5th Best**
Malthouse and Derenthal	**2008**	Customer penetration (upsell, cross-sell)	Data Mining	Logit Aggregated models logit Aggregated models w/ addtl data logit	Aggregated models with addtl data logit	Aggregated models logit	Logit		
Guido, Prete, Miraglia and De Mare	**2011**	Customer penetration (upsell, cross-sell)	Data Mining	ANN Logit Multiple regression	ANN	Logit	Multiple Regression		

4.7 Data Mining Technique Usage and Objectives Met

Having detailed what techniques have been utilised for purposes of Customer Selection, the next logical step is to assess how techniques have been applied to specific objectives. Linking usage to specific objectives was one of the main gaps that was identified earlier and that made advancements in the field of Database Marketing difficult, particularly for practitioners. Table 36 provides a detailed overview of how the different techniques from the systematic review relate to objectives. The vertical axis illustrates all the techniques that emerged from the systematic review while the horizontal line illustrates the pursued objectives. Each technique is accompanied by three critical pieces of information: references, count of studies, and the representation of this count as a percentage of total techniques applied across studies. (Using a percent of studies representation proved cumbersome given the count of technique applications is nearly threefold the number of studies.) The numbers that are presented on the reference line for each technique indicate specific studies that relate the techniques to objectives. A reference table identifying the individual study authors follows in Table 37. For example, for the MARS technique, reference number 14 relates its application to the customer penetration objective. The row below the reference illustrates that this is the only study that ties MARS and customer penetration together and that it represents about 1 percent of the techniques (1 out of 91) identified across the review. Reference 14 is identified in Table 37 as the study by Deichmann et al. (2002).

When crossing statistical techniques with RM and CRM marketing objectives, cluster analysis is tightly tied to pure segmentation, whereas RFM and logistic regression are mainly used for customer penetration. Regression techniques were mostly used in pure segmentation though some also addressed the objectives of acquisition and penetration. The machine learning techniques are almost fully dedicated to customer penetration with more than 60 percent of overall applications used against this objective and almost all of AID/CHAID and neural networks applications.

As mentioned in the previous sections, a disproportionate amount of studies is concerned with customer penetration. Table 36 illustrates that 76 percent of techniques are applied in the context of customer penetration, 12 percent in the context of segmentation and 5 percent for acquisition. Looking at the distribution by category, statistical technique applications were applied nearly 70 percent of the time to customer penetration while machine-learning applications were applied nearly 90 percent of the time. The other significant objective that statistical techniques were concerned with was segmentation, with 20 percent of techniques applied in that concern.

Table 36: Direct Marketing Objectives and Relationship to Data Mining Technique

Techniques \ Objectives		New customer acquisition/ reactivation		Customer penetration		Retention/Churn & Loyalty		Lifetime value Maximisation		Pure Segmentation		# Techniques	% total
TOTAL STUDIES:	33												
Statistical Techniques		4	4%	37	41%	1	1%	2	2%	11	12%	55	60%
Descriptive & Visualisation Techniques		0	0%	5	5%	0	0%	1	1%	0	0%	6	7%
Descriptive Cross-Tabulation	*References*			*13*									
	# studies / %total			1	1%							1	1%
RFM	*References*			*12, 19, 24*				*15*					
	# studies / %total			3	3%			1	1%			4	4%
FRAC	*References*			*12*									
	# studies / %total			1	1%							1	1%
Cluster Analysis		1	1%	3	3%	0	0%	1	1%	5	5%	10	11%
Cluster Analysis	*References*			*1*						*6, 28, 29*			
	# studies / %total			1	1%					3	3%	4	4%
Cluster Analysis w/ additional variables	*References*			*1*						*28, 29*			
	# studies / %total			1	1%					2	2%	3	3%
AHP Weighted Clustered RFM	*References*	*18*		*1*				*15*					
	# studies / %total	1	1%	1	1%			1	1%			3	3%
Discriminant Analysis		0	0%	1	1%	0	0%	0	0%	1	1%	2	2%
Discriminant Analysis	*References*			*7*						*11*			
	# studies / %total			1	1%					1	1%	2	2%
Factor Analysis		0	0%	0	0%	0	0%	0	0%	1	1%	1	1%
Structural Equation Model	*References*									*6*			
	# studies / %total									1	1%	1	1%
Regression Analysis		3	3%	12	13%	0	0%	0	0%	4	4%	19	21%
Multiple Regression	*References*	*4*		*1, 16, 33*						*11*			
	# studies / %total	1	1%	3	3%					1	1%	5	5%
Linear Regression	*References*			*9*									
	# studies / %total			1	1%							1	1%
Stepwise Linear Regression	*References*	*4*		*10, 14*									
	# studies / %total	1	1%	2	2%							3	3%
Log-Linear Model	*References*	*4*		*1*									
	# studies / %total	1	1%	1	1%							2	2%
Latent Class Regression	*References*			*17, 26, 31,*									
	# studies / %total			3	3%							3	3%
Ridge Regression	*References*			*10*									
	# studies / %total			1	1%							1	1%
Sliced Inversed Regression	*References*									*11*			
	# studies / %total									1	1%	1	1%
Partial Least Squares Regression (PLSR)	*References*									*11*			
	# studies / %total									1	1%	1	1%
Principal Components Regression (PCR)	*References*									*11*			
	# studies / %total									1	1%	1	1%
Multiple Adaptive Regression Splines (MARS)	*References*			*14*									
	# studies / %total			1	1%							1	1%
Logistic Regression		0	0%	16	18%	1	1%	0	0%	0	0%	17	19%
Tobit	*References*			*9*									
	# studies / %total			1	1%							1	1%
Two stage model	*References*			*9*									
	# studies / %total			1	1%							1	1%
Stepwise Logistic Regression	*References*			*14*									
	# studies / %total			1	1%							1	1%
Logistic/Logit Regression	*References*			*1, 7, 8, 9, 12, 16, 19, 22, 24, 31, 32, 33*		*20*							
	# studies / %total			12	13%	1	1%					13	14%

Techniques			New customer acquisition/ reactivation		Customer penetration		Retention/Churn & Loyalty		Lifetime value Maximisation		Pure Segmentation		# Studies	% total
Machine Learning Techniques			1	1%	32	35%	2	2%	1	1%	0	0%	36	40%
Neural Networks			0	0%	14	15%	0	0%	0	0%	0	0%	14	15%
Neural Network	*References*				*7, 8, 16, 17, 22, 25, 26, 33*									
	# studies	%total			8	9%							**8**	**9%**
Bayesian Neural Networks	*References*				*17, 26, 31.*									
	# studies	%total			3	3%							**3**	**3%**
Neural Networks using bagging techniqe (BMLP)	*References*				*22*									
	# studies	%total			1	1%							**1**	**1%**
Neural Networks using Genetic Algorithm	*References*				*12, 23*									
	# studies	%total			1	1%							**1**	**1%**
Support Vector Machines	*References*				*25*									
	# studies	%total			1	1%							**1**	**1%**
Genetic Algorithm			0	0%	1	1%	0	0%	1	1%	0	0%	2	2%
Genetic Algorithm	*References*				*12*				*30*					
	# studies	%total			1	1%			1	1%			**2**	**2%**
Decision Trees			1	1%	15	16%	2	2%	0	0%	0	0%	18	20%
AID/CHAID	*References*		*18*		*1, 2, 3, 5, 12, 13, 16, 19, 24, 25*									
	# studies	%total	1	1%	10	11%							**11**	**12%**
CART	*References*				*2,5, 17, 26, 31*									
	# studies	%total			5	5%							**5**	**5%**
Random Forests	*References*						*20*							
	# studies	%total					1	1%					**1**	**1%**
Automatic Relevance Determination (ARD)	*References*						*20*							
	# studies	%total					1	1%					**1**	**1%**
Mixed Models			0	0%	2	2%	0	0%	0	0%	0	0%	2	2%
Mixed models	*References*				*32*									
	# studies	%total			1	1%							**1**	**1%**
Mixed models w/ additional data	*References*				*32*									
	# studies	%total			1	1%							**1**	**1%**
# Total Techniques Used	**# studies**	**%total**	**5**	**5%**	**69**	**76%**	**3**	**3%**	**3**	**3%**	**11**	**12%**	**91**	**100%**

Table 37: Author References for Table 36

Study	Main Author	Year
1	Magidson	1988
2	Thrasher	1991
3	Bult and Wansbeek	1995
4	Lix, Berger and Magliozzi	1995
5	Haughton and Oulabi	1997
6	Jedidi, Jagpal and DeSarbo	1997
7	West, Brockett and Golden	1997
8	Zahavi and Levin	1997
9	Levin and Zahavi	1998
10	Malthouse	1999
11	Naik, Hagerty and Tsai	2000
12	Levin and Zahavi	2001
13	Ratner	2001
14	Deichmann, Eshghi, Haughton, Sayek and Teebagy	2002
15	Shih and Liu	2003
16	Linder, Geier, and Kolliker	2004
17	Cui and Wong	2004
18	Jonker, Piersma and Van den Poel	2004
19	Yang	2004
20	Buckinx and Van den Poel	2005
21	Jonker, Piersma and Potharst	2006
22	Ha, Cho and MacLachlan	2005
23	Kim, Street, Russell and Menczer	2005
24	McCarty and Hastak	2007
25	Crone, Lessman and Stahlbock	2006
26	Cui, Wong and Lui	2006
27	Malthouse and Elsner	2006
28	Reutterer	2006
29	Greene and Greene	2008
30	Chan	2008
31	Cui, Wong and Zhang	2008
32	Malthouse and Derenthal	2008
33	Guido, Prete, Miraglia and De Mare	2011

Source: From 33 studies explicitly addressing Data Mining techniques

4.8 Measures of Effectiveness

Calder and Malthouse (2002) and Malthouse (1999) suggest that the main criteria that should be used to appraise a predictive model are measures of fit and measures of performance. Fit measures determine how well a model describes the dependent variable across all data observations. Though useful in theory, these measures are not useful for evaluating financial or behavioural performance since it has been repeatedly demonstrated that models that do not fit particularly well still may perform well (Roberts and Berger, 1993; Malthouse, 2002; Magliozzi and Berger, 1993).

In practice, performance measures are what marketers are most interested in, oftentimes at the expense of predictive accuracy (Jonker et al., 2004; Kim et al., 2005). In most cases, database marketers elect to use descriptive statistics to conduct financial and behavioural performance analysis and assess business objectives and achievement. However, though practitioners may prefer performance measures, both types of measures are related and thus cannot, without high risk of error, be considered in isolation. A model that yields a strong performance but a poor fit may not be stable or even desirable, whereas a high-fit but low-performance model does not meet managerial expectations. Furthermore, effectiveness is not a linear function and, consequently, models also need to be assessed at intervals to determine whether these are stronger within certain percentiles, quintiles, and/or other levels or segment types.

4.8.1 Effectiveness Measures of Objectives

The effectiveness measures of objectives assess what output is expected from an exercise where an objective is attempting to be achieved. In cases where models are applied, the effectiveness measures are often referred to as the dependent variable. These measures are necessary to produce the data points by which models are being compared.

From the review, objectives are assessed using five key performance measures: response rate, sales increments generated by direct marketing activities, profitability, defection rate, and attitudinal measures (such as customer satisfaction). The use of measures across studies presented in Table 38 highlights that response rate is disproportionately the preferred measure of researchers. All other measures only emerge from a narrow subset of studies. In some cases, more than one metric was used which explains why the sum of measures is greater than the number of studies.

Table 38: Performance Measure Usage

Performance Measure	# of Studies
Response Rate	24
Sales Increment	4
Profitability Increments / Return on sales / Return on Investment	4
Retention / Defection Rate	2
Customer Lifetime Value (CLV)	2
Customer Satisfaction OR Other Attitudinal Measures	2

Source: From 33 studies explicitly addressing Data Mining techniques

Table 39: Performance Measures Used per Retained Articles

		Metric						
Primary Author	Year	Response Rate	Sales Increment	Profitability Increments /	Retention /	Customer Lifetime	Customer Satisfaction	Sum
Magidson	1988	1						1
Thrasher	1991	1						1
Bult and Wansbeek	1995			1				1
Lix, Berger and	1995	1						1
Haughton and Oulabi	1997	1						1
Jedidi, Jagpal and DeSarbo	1997						1	1
West, Brockett and Golden	1997						1	1
Zahavi and Levin	1997	1						1
Levin and Zahavi	1998			1				1
Malthouse	1999		1					1
Naik, Hagerty and Tsai	2000	1						1
Levin and Zahavi	2001	1						1
Ratner	2001	1						1

		Metric						
Primary Author	Year	Response Rate	Sales Increment	Profitability Increments /	Retention /	Customer Lifetime	Customer Satisfaction	Sum
Deichmann, Eshghi, Haughton, Sayek and Teebagy	2002	1						1
Shih and Liu	2003					1		1
Linder, Geier, and Kolliker	2004	1						1
Cui and Wong	2004	1						1
Jonker, Piersma and Van den Poel	2004		1					1
Yang	2004	1						1
Buckinx and Van den	2004				1			1
Jonker, Piersma and Potharst	2005	1		1				2
Ha, Cho and MacLachlan	2005	1						1
Kim, Street, Russell and Menczer	2005	1						1
McCarty and Hastak	2006	1						1
Crone, Lessman	2006	1						1
Cui, Wong and Lui	2006	1						1
Malthouse and Elsner	2006	1						1
Reutterer	2006			1				1
Greene and Greene	2008	1	1					2
Chan	2008		1		1	1		3
Cui, Wong and Zhang	2008	1						1
Malthouse and Derenthal	2008	1						1

Primary Author	Year	Metric Response Rate	Sales Increment	Profitability Increments /	Retention /	Customer Lifetime	Customer Satisfaction	Sum
Guido, Prete, Miraglia and De Mare	2011	1						1
Sum		**23**	**4**	**4**	**2**	**2**	**2**	**37**

Source: From 33 studies explicitly addressing Data Mining techniques

Nevertheless, though the response rate may dominate from a measures' usage perspective, it remains that 14 other measures were used either in combination or in isolation in the studies' subset. Other direct campaign profitability measures such as sales or profit increases, return on sales (ROS) and return on investment were used eight times while other measures (directly tied to campaign profitability) such as retention/defection rate, customer lifetime value (CLV) and customer satisfaction were used six other times.

Of the entire subset of studies, three studies used multiple measures. Jonker et al. (2005) used both response rate and profitability measures, Greene and Greene (2008), for their part, used response rate and average sales per customer, and Chan used CLV, retention and sales. The combination of measures allowed for a more balanced and holistic view of performance extending beyond pure campaign profit (in some cases) to customer value generation. Table 40 illustrates how these measures are related to customer selection objectives by examining the number of times an objective has been evaluated by a given performance measure. Of note are the high usage of response rate versus all others and the limited use of sales and profit measures across all objectives. Retention and defection measures were limited to measuring the objectives of retention and lifetime value maximisation. Evidently, the customer lifetime value measure was also used to evaluate the lifetime value maximisation objective. This does indicate that academics separate the tactical objectives of acquisition, penetration and retention from the more longitudinal and strategic objective of lifetime value maximisation. Finally, another finding that emerges from the table is the use of all measures for the assessment of lifetime value maximisation. This illustrates an interest by academics in demonstrating performance impacts on both the aggregate measure of CLV and its more specific submeasures. This focus on the sources of CLV is consistent with Ryals' (2007) three elements of the customer relationship: duration of relationship, revenues, and costs. These three elements are all intrinsically linked to all the measures shown in Table 40.

Table 40: Objectives and Performance Measurement

Measures / Objectives/ Performance	Response Rate	Sales / Sales incr.	CLV	Retention/ Defection	ROS/ROI/ Profitability	Other
Acquisition	2	1	0	0	0	0
Customer Penetration	19	1	0	0	1	1*
Retention/Churn/Loyalty	0	0	0	1	0	0
Lifetime value Maximisation	1	1	2	1	1	0
Pure Segmentation	2	1	0	0	1	1**

** Other metric for customer penetration includes: consumers' attitudes and perceptions and behavioural response measure of consumer patronage*

*** Other metrics for pure segmentation includes: a) customer satisfaction and b) consumers' attitudes and perceptions and behavioural response measure of consumer patronage*

Source: From 33 studies explicitly addressing data mining techniques

4.8.2 Comparative Model Performance Methods

As mentioned in the section's introduction, marketers are very heavily focused on the development of models that meet business needs and performance objectives (sometimes at the expense of model fit). Thus, the goal is to ultimately differentiate customers based on their likelihood to respond to various marketing stimuli and ultimately increase the responsiveness of customers targeted by a model's application. Given they act as the dependent variables, objective measures are important, but equally important are the methods used for comparative model performance and model performance at different levels of file depth (as specific models also aim to separate the better customers from the rest).

In the systematic review, methods found to be most commonly used to differentiate segment (decile) performances are: lift charts, gains charts and decile charts. Lift and gain charts represented nearly one-third of all utilised methods with 30-percent and 27-percent usage, while response rate charts represented 24 percent of measures used. Together these three methods account for more than 80 percent of all methods used to assess performance. It is important to look at these methods in a holistic manner since they are all very intimately related. It is not uncommon to see methods presented in combinations of two in research. The most frequent combination, because of the alignment of groupings (deciles), are decile and gains charts. This is consistent with a survey conducted among members of the DMA Research Council that illustrated that the methods of choice for assessing response models were first and foremost gains and cumulative gains charts, followed by lift and cumulative lift (Greene and Milne, 2010).

A lift chart is a visual aid for measuring model performance. It measures a predictive model's effectiveness as the ratio between results obtained with and without the application of the model. A lift chart can be Analysed by decile and break down performance in 10-percent increments from highest performing to lowest performing sub-groups. Gains charts are commonly created after lift charts and offer a common index representing the ratio of the decile or segment to the average response rate of the total sample (which is then multiplied by 100). This provides the marketers with a way of assessing what deciles or segments performed better or worse than the average.

Gains and lift charts are invaluable for assessing the decile level response and profit comparisons of marketing campaigns; however, they are not particularly useful for evaluating or comparing overall model fit. It is a known weakness of these measures that full model performance or model reliability can be ambiguous given there are no inferential tests that address reliability or model equality with these statistics (Greene and Greene, 2005; Greene and Milne, 2010).

For parsimony purposes, I renamed deciled response rate charts, response rate charts in the output of the systematic review because not all studies organised information according to deciles (though the objective and presentation are the same). Other examples or groupings that substituted deciles include:

- Magidson (1988): specific depths of file of 18%, 25% and 43%
- Thrasher (1991): 5% increments
- Zahavi and Levin (1997): specifically 50% and 80% depths of file
- Kim et al. (2005): specifically 20% and 50% depths of file
- McCarty and Hastak (2006): specifically 20%, 30%, 40%, and 50% depths of file

The mean Pareto curve measure could have also been included in the response rate metrics as it essentially provides technique performance values at 20 percent depth of file for each propensity measure (Lix et al., 1995). However, given the different denomination, it was treated separately. It should also be noted that the profit and cumulative profit by depth of file measures used by Levin and Zahavi (1998) was used in combination with the other leading measures of cumulative response rate and lift/cumulative lift charts.

Table 41: Performance Measures Count and Distribution

Performance Measures	Count	% total
Gains Chart/Cumulative Gains Chart	10	30%
Lift Chart/Cumulative Lift Chart	9	27%
Response Rate/Cumulative Response Rate by depth of file	8	24%
Profit/Cumulative Profit by depth of file	1	3%
Mean Pareto Curve	1	3%
Multi-year response rate comparison	1	3%
Significance Test vs. holdout sample	1	3%
Lift index	1	3%
Fitness Value	1	3%
Sum	**34**	**100%**

Source: From 33 studies explicitly addressing Data Mining techniques

Though other methods were used very infrequently, before continuing, I will highlight the multi-year response rate comparison by Jonker et al. (2005). This specific method was used because the aim of the research was to develop the optimal frequency of contacts that would result in profitable long-term relationships with target customers. I highlight this study and method specifically among the minority set because it illustrates that, although performance method usage may seem to converge, all cases are unique and must use adapted methods as required by each case and objective.

For clarity, though I will detail the second type of effectiveness measures (fit measures) in the next section, it should be noted that for the purposes of this thesis (unless stated otherwise) reference to effectiveness moving forward will be related to measures of performance (sometimes at an aggregate level and other times at a decile level). This allows me to better tie the concept of effectiveness to the review question, as it is mainly concerned with the capacity of Data Mining techniques to deliver results against direct marketing objectives and less the techniques' capacity to deliver statistical fit. Nevertheless, fit will be assessed and discussed, though overtly described when doing so.

4.8.3 Fit Measures – Predictive Accuracy

Measures of fit (also referred to as measures of predictive accuracy) are critical to determine whether models that emerge from the predictive exercise are plausible and that predictions will effectively carry over to data not used in the exercise of model fitting (Hawkins, Bask and Mills, 2003). Fit measures depend heavily upon the objective pursued and the technique used.

For example, techniques related to building clusters, segments or groupings of customers are typically assessed using within and inter-group heterogeneity measures (Wedel and Kamakura, 2000). The performance of the binary models is evaluated by the area under the receiver operating characteristic curve (AUC), which is a widely accepted criterion since it evaluates the ranking for different thresholds (Ha et al., 2004). Model generalisation can also be captured via the percentage correctly classified (PCC) measure, confusion matrices, and the receiver operating curve (ROC) (to which AUC is related) (Stahlbock, Crone, Lessmann, 2010). Techniques for predictive modelling, when continuous covariates are present, use deviance and Pearson chi-square tests, R^2, the adjusted R^2 and the Root of the Mean Squared Error (RMSE), whereas the Hosmer-Lemeshow, Cox and Snell's R Square, Nagelkerke's R Square, Wald statistic, and adaptations of the Pearson chi-square tests have been used for logistic regression and Chi-squared 'goodness of fit' tests for classification trees (Archer, Lemeshow, Hosmer, 2007; Field, 2005; Olson and Chae, 2012).

Greene and Milne (2010) state that traditional goodness of fit statistics such as R^2, the F statistic and the Chi Square statistic are not adequate for evaluating model performance "when the objective is to create models that maximise differentiation between population segments in terms of response rates. The traditional statistical measures are appropriate for assessing how well individual response values fit a given model (for example, minimise least square errors between response data points and predicted values); however, they are not appropriate for effectively selecting market segments or individual customers for targeting and meeting business objectives. The descriptive measures used by practitioners, although visually appealing, do not assess overall model performance with statistical certainty" (Greene and Milme, 2010, p. 36). Greene's research subsequently recommends the use of the Gini index as a useful and valuable substitute measure for assessing a response model's performance given it can also be used as a measure to the consistency (reliability). Calculated using the AUC score (see Study 1, measures section), the Gini coefficient is a measure often used to assess income inequality but has gained acceptance in its use for overall model performance assessment. The Gini coefficient evaluates the ability of different models to rank promotional responders correctly. Risselada, Verhoef and Bijmolt (2010) refer to this as degree acceptance and use it to compare the output of a random selection of customers with that of a model-based selection. Kamakura, Wedel, De Rosa and Mazzon (2003) and many other authors (Neslin, Gupta and Kamakura, 2006; Lemmens and Croux, 2006; Kim, 2006; Antipov and Pokryshevskaya, 2010) also support the use of Gini coefficients to summarise the predictive power of individual output curves and assess 'discriminability'. Finlay (2012, p. 218) captures this in his latest book on reponse modelling: "The most popular measures used to assess the performance of binary classification models are the percentage correctly classified (PCC), the Gini coefficient and the KS-statistic. Measures such as Gini coefficient and KS-statistic are also widely applied during model development to provide a Standardised benchmark of model performance as different versions of a model are developed. So, when several competing models are being evaluated, the one with the best KS or Gini statistic is deemed to be the best."

The review of top measures of fit that emerged from more than one study in Table 42 illustrates a high level of consistency with measures issued from the aforementioned literature. Common measures that emerged from systematic review include PCC, AUC, R-square (and adjusted R-square), significance test, confusion matrix, F statistic, ROC and Gini coefficient. There was no mention of the KS statistic. The two measures with the most significant use were PCC and AUC, with respectively 28-percent and 19-percent representation on the total metric use in the studies using fit measures. Secondary measures netting between 6-percent and 9-percent representation included R-square, prediction error, and confusion matrices. However, 6 percent to 9 percent represents two to three studies on a total of thirty-one, thus indicating that characterising them as secondary may even be a bit of an overstatement. Nevertheless, these measures were more used than the rest that only garnered one use.

It is also important to note that many of the studies that used tertiary type metrics such as significance tests, correlation and the like also used many of the primary measures. For example, the use of significance tests by Deichmann et al. (2002) relates mainly to the comparison of individual variables' p-values between training and test samples. However, Deichmann et al. also complement the use of a significance test with an R-square measure to assess model fit. Ho, Cho and MacLachlan (2005) use the ROC statistic as a substitute for AUC, but also accompany it with the use of a confusion matrix.

Table 42: Measure of Fit Count and Distribution

Measures of Fit	Count	% total
Percentage correctly classified (PCC)	9	28%
Area under the receiver operating characteristic curve (AUC)	6	19%
R-square/Adjusted R2	3	9%
Prediction error	2	6%
Confusion Matrix/table	2	6%
Gini Coefficient	1	3%
Significance Test	1	3%
Goodness-of-fit (GFI) index	1	3%
Root mean residual (RMR)	1	3%
Receiver Operating Characteristics (ROC)	1	3%
Percentage of buyers captured relative to the total	1	3%
Within sample accuracy and Out-of Sample accuracy	1	3%
Arithmetic Mean performance	1	3%
Geometric mean performance	1	3%
Correlation coefficient	1	3%
Sum	**32**	**100%**

Note: Not all 33 studies used Fit measures and some studies used more than one

4.9 Effectiveness of Data Mining Techniques

More than three quarters of the thirty-three studies in the systematic review centred on the Data

Mining phase and a similarly high proportion centred on customer penetration as an objective. Of these, a disproportionate number of studies used response rate as a performance metric. This generated a subset of seventeen (of the original thirty three) studies that are highly comparable and from which I could effectively compare performance. These studies are detailed in Table 43.

Table 43: Comparative Studies

Study	Primary Author	Year	Title Primary	Periodical Full	Industry Context
1	Magidson	1988	Improved Statistical Techniques for Response Modelling	Journal of Direct Marketing	Petrol Retail Credit Card
2	Thrasher	1991	CART: A Recent Advance in Tree-Structured List Segmentation Methodology	Journal of Direct Marketing	Direct Mail
3	Haughton and Oulabi	1997	Direct Marketing Modelling with CART and CHAID	Journal of Direct Marketing	Not specified
4	Zahavi and Levin	1997	Applying neural computing to target marketing	Journal of Direct Marketing	unknown
5	West, Brockett and Golden	1997	A comparative analysis of neural networks and statistical methods for predicting consumer choice	Marketing Science	Simulated data, Mass-merchandise retailers
6	Levin and Zahavi	1998	Continuous predictive modelling; a comparative analysis	Journal of Interactive Marketing	Home equity loans (Banking)
7	Ratner	2001	Identifying the best customers: Descriptive, predictive and look-alike profiling	Journal of Targeting, Measurement and Analysis for Marketing	Cellular Phone Carrier
8	Levin and Zahavi	2001	Segmentation analysis with managerial judgment	Journal of Direct Marketing	Collectibles
9	Linder, Geier, and Kolliker	2004	Artificial neural networks, classification trees and regression: Which method for which customer base?	Journal of Database Marketing & Customer Strategy Management	Simulated data
10	Yang	2004	How to develop new approaches to RFM segmentation	Journal of Targeting, Measurement and Analysis for Marketing	Not specified
11	Ha, Cho and MacLachlan	2005	Response Models based on bagging neural networks	Journal of Interactive Marketing	Catalogue
12	Kim, Street, Russell and Menczer	2005	Customer Targeting: A Neural Network Approach Guided by Genetic Algorithms	Management Science	Property & Casualty Insurance
13	Cui, Wong and Lui	2006	Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming	Management Science	Multi-division mail order
14	McCarty and Hastak	2006	Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression	Journal of Business Research	Multi-division mail order; Fundraising

Study	Primary Author	Year	Title Primary	Periodical Full	Industry Context
15	Cui, Wong and Zhang	2008	Model selection for direct marketing: performance criteria and validation methods	Marketing Intelligence & Planning	Catalogue
16	Malthouse and Derenthal	2008	Improving predictive scoring models through model aggregation	Journal of Interactive Marketing	Catalogue, Fundraising
17	Guido, Prete, Miraglia and De Mare	2011	Targeting direct marketing campaigns by neural networks	Journal of Marketing Management	Book mail order

To compare the effectiveness of individual techniques, a relative ranking system was designed and ranked each technique application in relation to another. When one technique performed better than another, it received a score of 1 on the “Performs Better” total column. The technique that scored worse received the score on the “Performs Worse” row. In cases where a specific technique was used with different variables (for example, when the category variable was added to an RFM) and comparative performance was available, this was captured with a footnote at the intersection of this specific technique. The scoring can be found in Table 44; for transparency purposes, each point of comparison in the table is tied to a specific article denoted by its number. Each number relates to references at the end of the table.

The key findings in this table are located in the extreme values. The three techniques with the most significant overperformance scores include: logistic regression, neural networks and CHAID. Logistic regression techniques were shown to perform better in nine instances including against the following techniques: FRAC (Levin and Zahavi, 2001; Yang, 2004), neural networks (Zahavi and Levin, 1997; Ho, Cho and MacLachlan, 2005), and CHAID (Levin and Zahavi, 2001; Linder et al., 2004) in two instances each. It performed better than multiple linear regression (Guido, 2011), latent class models, and CART (Cui et al., 2008) in one instance each. Neural networks followed, with eight instances of better performance including two instances with logistic regression (Linder et al., 2004; Guido, 2011), two with CHAID (West et al., 1997; Linder et al., 2004) and one instance for discriminant analysis (West et al., 1997), multiple regression (Guido, 2011), Bayesian neural networks and CART (Cui et al., 2006). CHAID decision tree ranked third with seven instances including three instances with RFM (Magidson, 1988; Yang, 2004; McCarty and Hastak, 2006), and unique instances against descriptive cross-tabulation (Ratner, 2001), FRAC (Levin and Zahavi, 2001), logistic regression (McCarty and Hastak, 2006) and CART (Thrasher, 1991).

Secondary techniques that also seemed to provide good overperformance scores included Bayesian neural networks and mixed models. Shortlisted Bayesian neural networks performed better against CART applications (Cui et al., 2008), latent class regression and logistic regression (Cui et al., 2008), whereas mixed models did better against logistic regression (Linder et al., 2004; Malthouse and Elsner, 2006), Bayesian neural networks (Linder et al., 2004), and CHAID (Linder et al., 2004).

Given that many of the leading and most prevalent techniques are used in many studies, it is not surprising to often find the same technique showing up in the highest underperforming scores tally as well. Logistic regression shows the most mixed results with 11 instances where it did not perform as well as other techniques. The main techniques that outperformed logistic regression were neural networks (Linder et al., 2004; Guido, 2011), mixed models (Linder et al., 2004; Malthouse, 2008) and Bayesian neural networks (Ho et al., 2005; Cui et al., 2008), while single-instance cases included RFM (McCarty and Hastak, 2006), discriminant analysis (West et al., 1997), Tobit (Levin and Zahavi, 1998), Two-stage models (Levin and Zahazi, 1998), and CHAID (McCarty and Hastak, 2006). CHAID was, for its part, outperformed by logistic regression in the studies by Levin and Zahavi (2001) and Linder et al. (2004), by neural networks in the studies by West et al. (1997) and Linder et al. (2004), and in single instances by CART and mixed models in work by Haughton and Oulabi (1997) and Linder et al. (2004). CART results mainly underperformed vis-à-vis Bayesian neural networks (Cui et al., 2006; Guido, 2011) and against smaller instances of logistic regression (Cui et al., 2008), neural networks (Cui et al., 2006) and CHAID (Haughton and Oulabi, 1997).

Performance was also positively impacted when studies considered both the pattern discovery and modelling, and the transformation phases. For example, Deichmann et al. (2002) applied the MARS and stepwise logistic regression technique to reduce the number of data dimensions required for modelling. Malthouse and Elsner (2006), though already counted in Table 43 because of the predictive strength of ridge regression, also used stepwise logistic regression for dimensionality reduction. Lix et al. (1995) similarly applied a stepwise linear regression for the same reasons. Shih and Liu (2003) applied an AHP technique to appropriately better weigh variables and enhance model outcomes. Some additional literature from Rao and Ali (2002) shows how genetic algorithms have been used with neural networks as a hybrid system to select neural network inputs, evolve their topology and evolve their learning rules.

Table 44: Data Mining Techniques' Relative Effectiveness

Performs worse than → / Performs better than ↓	Descriptive cross-tabulation	RFM	FRAC	Discriminant Analysis	Multiple Regression	Latent Class Regression	Tobit	Two stage model	Linear Regression	Logistic/Logit Regression	Logistic/Logit Regression w/ PCA	Neural Network	Bayesian Neural Networks	Neural Networks using bagging technique (BMLP)	Neural Networks using Genetic Algorithm	CHAID	CART	Mixed models	TOTAL: Overperforming techniques
Descriptive cross-tabulation																			0
RFM										14									1
FRAC																			0
Discriminant Analysis										5									1
Multiple Regression																			0
Latent Class Regression																			0
Tobit									6	6									2
Two stage model									6	6									2
Linear Regression																			0
Logistic/Logit Regression			8, 10		17	15			6			4, 11				8, 9	15		9
Logistic/Logit Regression w/ PCA																			0
Neural Network				5	17					9, 17			13			5, 9	13		8
Bayesian Neural Networks						15				15							13, 15		4
Neural Networks using bagging techniqe (BMLP)										11		11							2
Neural Networks using Genetic Algorithm											12								1
CHAID	7	1, 10, 14	8							14							2		7
CART						15										3			2
Mixed models										9, 16			9			9			4
TOTAL Underperforming techniques	1	3	3	1	2	3	0	0	3	11	1	3	2	0	0	6	5	0	

NOTES

Study 6:

Tobit and Two-stage models performed equally well

Study 10:

CHAID and Logit performed equally well

Study 16:

Study contained two mixed models, one model with additional data variables, and another without. The model with additional data performed better that the one with limited data. However, the model with limited data still outperformed the logit model.

4.10 Data Variables

The segmentation base taxonomy developed by Mama (2007) shows that marketers have numerous variable categories to choose from when looking to build statistical models. Variables can be geographic, demographic psychographic/lifestyle, benefits based, usage-based, loyalty-based, situation-based, and/or behavioural. Wedel and Kamakura (2000) argue that not all these bases are created equal and that some based are better than others at supporting different aspects of the segmentation and selection decision.

Table 45 illustrates what variables have been shown to be significant in the development of multiple technique applications in the 17 retained studies focusing on customer development. Each study's variables are then classified into Mama's (2007) taxonomy. Results highlight that the four types of significant variables that emerge from technique applications include demographic, geographic, usage and behavioural. Of these four, usage-based variables are the most significant followed by demographic variables.

It should be noted that though all significant variables from studies were identified in the table, different variables may emerge as significant in different models; for example, Magidson (1988) shows the most significant variables for decision-tree applications include date account opened, proprietary credit card holder, recency, and marketing region. However, the variables that fit best for regression are the following: activity (month since last purchase), credit card balance, date account opened, number of times moderately delinquent in the last 12 months, brand affinity score, marketing region, number of times current on balance in the past 12 months, and country size. This is to be expected as this phenomenon occurs naturally between different techniques given their different discrimination logic. Similar cases are documented across most studies with substantial data variable predictors applied to different techniques (Haughton and Oulabi, 1997).

Table 45: Variable / Segmentation Base used by Study

			Variable / Segmentation Base Category							
Authors, Primary	**Year**	**Significant Variables**	**Geographic**	**Demographic**	**Psychographic/ lifestyle**	**Benefits**	**Usage**	**Loyalty**	**Situation**	**Behavioural**
Magidson	1988	Date account opened Proprietary credit card holder Recency Marketing Region Activity (month since last purchase) Credit Card balance Number of times moderately delinquent in the last 12 months Brand Affinity score Number of times current on balance in the past 12 months County Size	X				X			
Thrasher	1991	undefined list more than 6OO demographic and financial lifestyle variables								
Haughton and Oulabi	1997	age mailing history (6 months, total) geographical census area income variable from Census # target list consumer on gender geodemographic data	X	X			X			
Zahavi and Levin	1997	gender leading product code previous category purchase history previous transaction spend total yearly spend specific SKU purchase history recency price levels consumer # categories shopped direct mail purchase transaction history (24 months) historical response rate by category average time between purchases		X			X			

			Variable / Segmentation Base Category							
Authors, Primary	Year	Significant Variables	Geographic	Demographic	Psychographic/ lifestyle	Benefits	Usage	Loyalty	Situation	Behavioural
West, Brockett and Golden	1997	Consumer survey response to competitive store imaging: Low/High Price Hard/Easy Credit Cluttered/Spacious Low/High Quality Merchandise Bad/Good Selection Low/High Calibre Hard/Easy to Exchange				X				
Levin and Zahavi	1998	not specified								
Ratner	2001	gender current cellular phone ownership income have children		X			X			
Levin and Zahavi	2001	not specified								
Linder, Geier, and Kolliker	2004	time of first purchase monetary turnover purchase frequency recency of last purchase age gender education mobility housing type family status size of community industrialization	X	X			X			
Yang	2004	Recency of purchase frequency of purchase monetary expenditures					X			

Authors, Primary	Year	Significant Variables	Geographic	Demographic	Psychographic/ lifestyle	Benefits	Usage	Loyalty	Situation	Behavioural
			Variable / Segmentation Base Category							
Ha, Cho and MacLachlan	2005	not specified								
Kim, Street, Russell and Menczer	2005	"Average Family" psychographic segment Amount of contribution to third-party policy, car policy, moped policy, and fire policy, and number of households holding third-party, policies and social security policies		X			X			
Cui, Wong and Lui	2006	Recency of purchase frequency of purchase monetary expenditures					X			
McCarty and Hastak	2006	Recency of purchase frequency of purchase monetary expenditures					X			
Cui, Wong and Zhang	2008	not specified								
Malthouse and Derenthal	2008	recency of purchase frequency of purchase monetary expenditures day of first purchase/ donation average order/donation amount, frequency and monetary value (FM) during the most recent year FM from 1 year ago FM from more than 2 years ago three interaction terms measuring frequency across years.					X			
Prete, Miraglia and De Mare	2011	past purchase of guide/handbooks age group 19–23 years gender		X			X			
COUNT			**3**	**6**	**0**	**0**	**11**	**0**	**0**	**1**

4.11 Discussion

The systematic review on Data Mining applications provides a detailed view of the objectives, techniques, variables and measures of performance issued from Database Marketing literature. This section will break down the discussion into these objectives.

4.11.1 Objectives

Looking at the results obtained from the objective categorisation of studies, three of the objectives of RM and CRM identified in the literature review were validated: acquisition, retention and development. Out of 35 studies, 24 examined customer development while only a small remainder focused on acquisition and retention. The two major differences in terms of objectives of Customer Selection vis-à-vis objectives of RM were the addition of Customer Lifetime Value and Pure Segmentation objectives.

The only RM objective not validated by the Customer Selection objectives in the systematic review was the termination of customer relationships. This is not surprising given that in non-contractual settings, terminating relationships does not hold the same meaning as in contractual (and usually B2B) settings. In non-contractual settings customer selection practices applied to Database Marketing activities aim at acquiring, increasing or maintaining customer share, not specifically terminating relationships (as in relationships where terms are contractually dictated). In such contexts, customers that are unprofitable, unresponsive or undesirable generally do not meet the criteria for campaign selection or for continued relationship investment. Given customers cannot simply be “terminated” or “fired” in retail settings, marketers choose instead to not communicate with them. Kumar and Reinartz (2012, p. 303) allude to this in their study of applications of CRM in B2B and B2C as they denote that “differential treatment of customers is therefore being accepted as a way of life by both firms and customers.” In short, in non-contractual settings, relationship termination is generally an output of the exhibited customer behaviour profile instead of an actual practiced behaviour.

When comparing this with the objectives distribution by Ngai (2009), the distribution of objectives is not consistent. Ngai’s systematic review of Data Mining applications in CRM illustrates that the most common objective supported by Data Mining was retention. In his study, 62 percent of studies focused on the retention dimension while 15 percent of studies concerned customer identification (acquisition) and another 15 percent concern customer development. Ngai also highlights customer segmentation as a sub-objective; however, it is categorised under the broader grouping of customer identification (with 8-percent representation). The customer identification category also included the conduct of customer analysis and profiling; however, this review did not yield any research of the sort. Nevertheless, customer analysis is routinely applied directly to achieving the outputs of other objectives (i.e. segmentation and clusters such as RFM are routinely used for targeting purposes). As a result, it is not particularly surprising to not see it emerge as an overt objective of the review.

The importance of customer penetration as an objective was, however, validated by Verhoef et al. (2002), albeit in practice. However, this is, to my knowledge, the first study to provide a detailed assessment on how the customer penetration objective is pursued from a Data Mining perspective. Ngai (2009) does relate Data Mining techniques to CRM objectives; however, the majority of the articles reviewed in his research are classified as aiming to improve customer retention. Only 10 articles are identified for focusing on the customer development objective and of those 10, six are directed to the CLV and customer analysis objectives. As a result, Ngai's work may provide policy makers insight on the frequent Data Mining techniques used for retaining customers, but the output of the systematic review provides policy makers with insight on what Data Mining techniques to use for targeting customers in the context of the customer development objective.

4.11.2 Technique Usage

When observing the usage of techniques, I note that general technique usage aligns with Verhoef et al. (2002). The main difference lies in the ranking of methods between Statistical and the Machine Learning categories. The output from this systematic review shows a much stronger representation of non-descriptive techniques, such as regression analysis and clustering techniques, and a much less dominant role of RFM and cross tabulation. In addition, ANN and decision-tree techniques are much more present in the systematic review than in Verhoef's sample. This is likely explained by two factors: (1) the year of publication would have limited researchers' access to computing power and probably know-how of more advanced techniques and (2) the target audience was not academics of their research but rather practitioners.

Comparing this to the more academic sample extracted by Ngai (2009) presents a different perspective. Ngai identifies 34 separate Data Mining techniques. Of those techniques, neural network is the most frequently used with a presence in 34 percent of assessed studies. Following neural networks, techniques with the most usage included decision trees, association rules and regression with, respectivel,y 26 percent, 21 percent and 11 percent of studies. Table 46 illustrates not only the count and distribution of techniques as a percentage of studies in the systematic review; it also compares the output with Ngai systematic review output. The central column provides a very simple heat map illustrating the areas of most significant difference between studies. Green areas represent significant differences in favour of the current study, whereas Red areas represent differences in favour of Ngai's study. Yellow areas represent marginal differences. Finally, the only other difference of note (greater than 10-percent difference) includes the higher proportion of RFM technique usage represented in the current study. In fact, RFM as a technique does not show up at all in Ngai output (though it is possible that it is conducted under the guise of a clustering approach).

Table 46: Comparison of Data Mining Technique Count versus Ngai (2009)

Technique	Technique Count	% of Studies	Diff	% of Studies	Technique Count	Technique by Ngai
Logistic/Logit Regression	15	45%	34%	11%	10	Regression
AID/CHAID	11	33%	7%	26%	23	Decision tree
Multiple/Linear regression	8	24%	24%			
Neural Networks/ANN	8	24%	-10%	34%	30	Neural network
CART (Classification and Regression Trees)	5	15%	15%			
RFM	5	15%	15%			
Discriminant analysis	3	9%	9%			
Log-linear	3	9%	9%			
Latent Class Models	3	9%	8%	1%	1	Latent class model
Bayesian Networks	2	6%	4%	2%	2	Bayesian network classifier
Cluster Analysis	2	6%	6%			
finite mixture structural equation models	1	3%	3%			
VQ (K-means)	2	6%	3%	3%	3	K means
Tobit	1	3%	3%			
Two-stage models	1	3%	3%			
Ridge Regression	1	3%	3%			
Partial Least Squares Regression (PLSR)	1	3%	3%			
Principal Components Regression (PCR)	1	3%	3%			
Sliced Inverse Regression (SIR)	1	3%	3%			
FRAC-Frequency, Recency, Amount, Category	1	3%	3%			
Genetic Algorithms	1	3%	-2%	5%	4	Genetic algorithm
Cross-tabulation	1	3%	3%			
Multiple Adaptive Regression Splines-MARS)	1	3%	2%	1%	1	Multivariate adaptive regression splines
AHP	1	3%	3%			
Markov Chains/Markov Chain Monte Carlo	1	3%	-2%	5%	4	Markov chain
Automatic Relevance Detection (ARD)	1	3%	3%			
Random Forests	1	3%	3%			
Bagging Neural Networks	1	3%	3%			
ANN/GA	1	3%	3%			
PCA/Logit	1	3%	3%			
Support Vector Machines	1	3%	2%	1%	1	Support vector machine
Crossed-basis sub-segmentation	1	3%	3%			
Sum	**87**		-21%	21%	18	Association rules
			-5%	5%	4	Survival analysis
			-3%	3%	3	K nearest neighbour
			-1%	1%	1	If-then-else
			-1%	1%	1	Set theory
			-1%	1%	1	Attribute oriented induction
			-1%	1%	1	Constructive assignment
			-1%	1%	1	Customer map
			-1%	1%	1	Data envelopment analysis
			-1%	1%	1	Data mining by evolutionary learning
			-1%	1%	1	Expectation Max
			-1%	1%	1	Expectation Max Mod
			-1%	1%	1	Farthest first
			-1%	1%	1	Goal oriented sequential pattern
			-1%	1%	1	Logical analysis of data
			-1%	1%	1	MARFS1/S2
			-1%	1%	1	Mixture transition distribution
			-1%	1%	1	Multi-classifier class combiner
			-1%	1%	1	Online analytical mining
			-1%	1%	1	Outlier detection
			-1%	1%	1	Pattern based cluster
			-1%	1%	1	Rule-based RIPPER
			-1%	1%	1	S-means
			-1%	1%	1	S-means Mod
					125	**Sum**

Remark: Each article may have used more that one data mining techniques.

4.11.3 Data Mining Phases and Transformation

In the systematic review of findings, principle component analysis (PCA), partial least squares (PLS), sliced inversed regression and ridge regression emerged as data reduction methods. All these methods are applied prior to predictive techniques and had a positive impact on performance. Performance was also positively impacted when studies considered both the pattern discovery and modelling and the transformation phases either in sequence or in combination. For example, Deichmann et al. (2002) applied the MARS and stepwise logistic regression technique to reduce the number of dimensions required for modelling. Lix et al. (1995) and Malthouse and Derenthal (2008) respectively used stepwise linear and logistic regression for dimensionality reduction. Shih and Liu (2003) applied an AHP technique to appropriately better weigh variables and enhance model outcomes. In fact, literature from Rao and Ali (2002) shows that other less common model combinations have also been used to increase performance. In their study, genetic algorithms were used with neural networks as a hybrid system to select neural networks input variables and evolve their learning rules. These examples do not even include the potential inherent value of neural networks or decision trees in reducing variable subsets; however, it is recognised that just like stepwise regression, the impact is significant. Much like in studies comparing RFM to more advanced techniques, in all cases where dimensionality reduction is applied, enhanced performance follows.

Feature Reduction and Transformation

Though Kantardzic (2003) states that two feature reduction approaches exist, empirical bottom-up and heuristic top-down. Approaches that emerged from research were all bottom-up and fully empirical approaches, meaning that they started with the data itself and, based on a full set of variables, assessed which ones to remove or transform. None of the approaches where issued from the top-down school. As a reminder, this approach starts with the researcher making an informed decision of the variables to include in the modelling exercise based on some heuristic criteria (a priori). To a certain extent, both approaches can be applied simultaneously and it is often hard to determine when authors have applied one or the other.

The techniques that emerged from the systematic review include: PCA, PLS, sliced inverse regression, MARS, AHP, stepwise regression and decision-tree techniques (CART, AID, CHAID). These techniques all address Kantardzic's (2003) first objective of data reduction/transformation. PCA, PLS, sliced inverse regression, stepwise regression and decision trees achieve this by reducing dimensionality. MARS transforms features/variables to make them easier to model. And AHP weighs the individual features/variables to better affect model outcomes.

The most popular statistical methods for strict empirical dimensionality reduction are PCA and

PLS (Lee et al., 2011). Both did emerge in the systematic review, as did Sliced Inversed Regression. All three techniques were compared in Naik et al.'s (2000) dimension reduction research in data-rich marketing environments. Findings showed that Sliced Inverse regression provided the greatest reduction effectiveness followed respectively by PLS, PCA and Sliced Regression. In addition, Naik's comparative performance of PLS and PCA are consistent with Kemsley (1996) that illustrates that PLS produces improvements in class separation and discrimination, and requires less transformation and variable compression versus PCA in high dimensionality environments.

A less utilised method for variable subset selection that emerged from the study was ridge regression. In its application by Malthouse (1999), ridge regression is used as an alternative to standard data reduction methods as well as stepwise regression to manage the collinearity bias. Malthouse compares the ridge regression estimates with stepwise regression and suggests that ridge regression provides more model stability than simply dropping variables. The application of ridge regression is consistent with Abdi (2010) and its objective of addressing issues of multicollinearity.

The Multiple Adaptive Regression Splines (MARS) technique also appeared in the review. MARS is a technique that is mainly used to transform features/variables to make them easier to model. Deichmann et al. (2002) investigated its use in combination with logistic regression in the context of modelling direct response. The output from his work is that the usage of MARS technique to convert variables allowed prior to the application of logistic and linear regressions allowed for a lift in performance versus a pure application of logistic or linear regressions. It was not compared to any other transformation technique.

From a data weighting transformation perspective, the analytic hierarchy process (AHP) was the only method to emerge from the systematic review. Its application by Shih and Liu (2003) consists of combining the analytic hierarchy process and clustering analysis to maximise customer lifetime value ranking. The cluster analysis is conducted on RFM variables, so the result of the aim of the research is to compare lifetime value ranking of a clustered RFM output and the weighted clustered RFM output weighted with AHP. Results show a significant predictability of the enhanced AHP clustered RFM approach versus the standard RFM clustering.

An atypical study that emerged from review was Heilman, Kaefer and Ramenofsky (2003) that examined the sensitivity of classification accuracy to each additional purchase. Unlike other combined approaches, Heilman et al. aim to use results from the Pattern Discovery phase applications of Logit and neural networks to direct the degree of historical customer data utilised a priori for future Data Mining exercises. In essence, Heilman et al. use logit and neural networks as a simulation or sensitivity analysis to assess the right type of data reduction to apply (from a data history and frequency standpoint. Findings suggest that an optimal number of transactional observations exists and allows for an optimal classification and/or targeting purposes.

Yang (2004) illustrates the inefficiency of using RFM clusters because intuition-coded clusters are incapable of generating true segments. He argues that many options exist to address the limited reliability of the RFM variables. Firstly, the use of result-based segments identified by a statistical ‘z-score’ is demonstrated to generate increased returns by adequately recognising the relative strength of variables. Other approaches advanced by Yang to improve performance include using a CHAID technique, where more refined breakdowns can be achieved, instead of clusters, and the transition to an individual score-based focus based on the conversion of the RFM variables into a customer value variable.

Finally, Reutterer et al. (2006) uses vector quantization (VQ) algorithms (an online version of K-means type clustering) to compress unwieldy customer shopping basket transactional information database into a manageable set of simplified classification prototypes for the observed shopping baskets. A simple assignment rule logic based on behavioural persistency is then used to segment customers for selection purposes. Empirical findings demonstrate significant impacts on profitability and sales for the targeted segment-specific campaigns. As observed by Jain (2010), K-means (and its related applications such as VQ) is by far the most popular and simplest partitional algorithm. It is thus not surprising to find VQ as one of the systematic review outputs for data transformation.

Value and Case Reduction

Case or value reduction/transformation did not emerge in the search of customer selection. This is certainly not to say that they are not important. Any Data Mining exercise requires an ongoing application of case or value reduction/transformation (Kantardzic, 2003; Fayyad, Piatetsky-Shapiro and Smyth, 1996; Hand, Mannila and Smyth, 2001; Han and Kamber, 2006). However, the focus of the search on customer selection naturally reduced the intensity of search outputs focused on reduction and transformation techniques in favour of techniques applied in the context of pattern discovery.

This research is conducted using the full population of customers targeted for purposes of database marketing. No sample of targeted customers or responders was taken. In addition, though data is secondary, it is sourced from a transactional database with a very high degree of reliability and little or no requirements for case reduction or transformation, as data fields are already pre-transformed to optimise data analysis.

4.11.4 Measures of Effectiveness

The finding that the response rate is the most important metric of performance is consistent with findings from Cao and Gruca (2005), which indicate that some of the most utilised direct and Database Marketing metrics are response rates, as direct marketers usually seek the optimal response likelihood in the selection of prospects. One of the leading metrics that Cao and Gruca highlighted in their work that does not emerge directly from the output is the use of cost per response. However, cost per responder is an integral component of the composite profitability metric (Piatetsky-Shapiro and Masand, 1999). Given, the profitability of a Database Marketing campaign is a function of the lift of the model, the cost per mailing, and the marginal revenue per responder. It can be argued that this increased focus toward profitability and away from the cost dimension indicates a shift for research away from the pure cost and response view toward a more holistic profitability view. This evolution is consistent with requirements of customer orientation and RM to re-evaluate traditional marketing metrics in favour of customer-centric measures, such as customer profitability, that can provide managers with better indications of how CRM policies and programs are delivering results (Verhoef and Lemon, 2012; Gupta and Aggarwal, 2012).

Though practitioner based, the strong usage of response rate, the usage of sales (sales increment) and profitability metrics was also validated by work of Verhoef, Spring, Hoekstra and Leeflang (2003). In this practitioner overview, 90 percent of companies employed segmentation for purposes of target-customer selection. Other reasons include its use for the identification of specific prices or products to promote to the customer, the adequate timing of offer delivery or for building predictive models. But more importantly, of the same practitioner responders, the metrics they were trying to affect with their segmentation activity were response rate, purchase amount, profitability, and credit worthiness. Particularly telling was the wide margin by which the response rate measure outpaced all others (Table 47).

Table 47: Segmentation Criterion Variables

Criterion Variables (n=160)	Percent
Response Rate	87.4
Purchase Amount	61.6
Profitability	42.8
Creditworthiness	6.9

Source: Verhoef, Spring, Hoekstra and Leeflang (2003)

The main difference between the results for the systematic review output and the Verhoef et al.

(2003) study is the lower relative usage of response rate as a criterion in the systematic review output. Unlike the academic research distribution shown here, practitioner usage exhibits stronger use of response rates, sales and profitability metrics. Though the practitioner study was conducted in the limited context of the Netherlands, it is the only study of its kind to have been conducted. It illustrates the lack of alignment between academic and practitioner interests in metrics. Note should, however, be taken that effect sizes may exist when comparing both sets of statistics as only a limited number of studies were compared against Verhoef's sizeable managerial population sample (Figure 10).

The measures of comparative technique performance, performance and fit that emerged from the systematic review were also very consistent with the literature review. Performance measures found to be most commonly used are: lift charts, gains charts and deciled response rate charts (Magliozzi and Berger, 1993; Shepard, 1995; Hughes, 1996; David Shepard Associates, 1999; Roberts and Berger, 1999). Common measures of fit that emerged from both the systematic review and the literature review included: PCC, AUC, R-square (and adjusted R-square), significance test, confusion matrix, F statistic, ROC, and Gini co-efficient. The studies that used alternate methods usually did so with specific objectives in mind (i.e. comparing the significance of different variables between test and training samples) and usually accompanied these measures with standard measures of fit. Nevertheless, most utilised measures were PCC and AUC.

The two measures with the most significant use were PCC and AUC, with respectively 28-percent and 19-percent representation. Secondary measures included R-square, prediction error (6 percent) and confusion matrices (9 percent).

4.11.5 Effectiveness of Data Mining Techniques

When observing their relative performance and given that many of the leading and most prevalent techniques are used in many studies, it is not surprising to find the same technique showing up in the highest and lowest underperforming score tallies. This is the case for logistic regression and CHAID decision trees as they both came up as most and least effective techniques. Nevertheless, in many cases logistic regression and CHAID performed equally as well (often exchanging lead positions) and also compared effectively with Neural Networks. The CART decision tree did not perform as well as CHAID or logistic regression. Neural Networks performed generally much better without suffering from the same wide-ranking issues shown by CHAID and logistic regression. Bayesian Neural Networks and Mixed Models also showed mixed results. Though many techniques didn't make the list of best or worst, the weaknesses of the RFM/FRAC, log-linear models and CART techniques and the strength of the AID/CHAID techniques are well documented by Wedel and Kamakura (2000) and in light of the disproportionate number of studies comparing such techniques, they should thus be evaluated on the basis of relative strength versus relative weakness.

Another interesting observation is the relation between performance and (1) incremental data and (2) the application of techniques to the broader process of data mining. Firstly, studies by Greene and Greene (2008) and Reutterer et al. (2006) both show the positive effects of adding customer selection variables to clustering techniques, whereas Malthouse and Elsner (2006) show the benefits of complementing the standard RFM technique with the additional variable of category. Secondly, performance was positively influenced when studies considered both the pattern discovery and modelling and the transformation phases. For example, Deichmann et al. (2002) applied the MARS technique and Malthouse and Derenthal (2008) used stepwise logistic regression to reduce the number of data dimensions required for modelling. Shih and Liu (2003), for their part, applied an AHP technique to appropriately weigh variables for modelling. Some additional literature from Rao and Ali (2002) shows how genetic algorithms have been used with neural networks as a hybrid system to select neural network inputs and evolve their topology and their learning rules.

4.11.6 Data Variables

The significance of usage and geographic variables is somewhat consistent with literature from Wedel and Kamakura (2000), McCrary (2009), Rud (2001) and Van de Poel (2003) highlighting the predictive strength of usage, also referred to as transactional, variables. However, the significance of demographic variables is not consistent with the findings of the above-mentioned authors nor with the guidance provided by Beane and Ennis (1987) and Haley (1985). Beane and Ennis and Haley all argue against the discriminatory potential of demographic variables, rather favouring their use for profiling purposes instead.

Given variables are seldom applied strictly in unique category groups (i.e. only behavioural variables or only benefit variables), it is hard to prove or disprove Wedel and Kamakura’s criteria validity on a pure category level since degrees of significance, collinearity and model contributions need to be considered. However, single category studies do provide clues on the relationship between variables/bases and performance. For example, West et al.’s (1997) study uses survey responses providing benefit level information. Yang (2004), Cui et al. (2006, 2008) and Malthouse and Derenthal (2008) all use strictly usage-based variables. These cases provide some supporting evidence on the linkage between these categories as a whole and response.

Another interesting observation is the relation between performance and (1) incremental data and (2) the application of techniques to the broader process of data mining. Studies by both Greene and Greene (2008) and Reutterer et al. (2006) show the positive effects of adding customer selection variables to clustering techniques, whereas Malthouse (2006) shows the benefits of complementing the standard RFM technique with the additional variable of category. This said, when adding variables it is of critical importance to remain aware that more is not always better and that it is important to try to limit input of incremental predictor variables to the most significant ones early in the process. The addition of irrelevant independent variables (without prior validation of prediction or collinearity) may lead to poor accuracy (Ho, et al., 2005).

In reviewing the differences, some significant gaps appear. Firstly, if I add both logistic and multiple linear regression counts together, the current systematic review counts twice as many regression technique applications versus Ngai's study. In percentage terms, this represents 69 percent of studies versus 34 percent for the blended regression figure in Ngai. The next significant difference is shown with the decision-tree output. Adding both CHAI and CART decision-tree outputs together brings their representation in the current systematic review to 48 percent (33 percent and 15 percent, respectively) while Ngai's study shows 26 percent. Ngai's shows a significantly higher proportion of studies conducted using Neural Networks with 34 percent of studies represented versus 24 percent in the current study. The other significant difference that emerges from the Ngai study is the strong presence of Association Rules. The systematic review yielded none.

4.12 Limitations

The first set of limitations of this systematic review is inherently related to the process of the review itself. The review is limited by the output of literature that emerges from the research strings and the additions made post-hoc. The research is also a function of the researcher's scoring of literature. Without a double-blind scoring, this is a strict researcher bias limitation.
The second set of limitations is related to the generalisability of findings. Though a substantial amount of literature emerged for evaluation, the specific literature on Data Mining techniques was limited. As a result, care should be taken when interpreting these results as the overall sample size is moderate and there is a high level of inconsistency in the findings. Furthermore, the results were not weighted according to any rule or level of confidence, each study receiving equal weight. In the context of a systematic review leading to the definition of a research question, this is appropriate.

4.13 Summary

In summary, I conducted a systematic review of literature given the extensive and disorganised state of research on the topic. The systematic review approach allowed me to answer research questions and sub-questions in a transparent, comprehensive and organised fashion. It also adheres to high standards and meets scholarly requirements for peer reviewing academic studies.

This systematic review question is guided by the thesis question and is the following: "*What is the empirical evidence for the effectiveness of different Data Mining techniques in database marketing?"* This review question is subdivided into six sub-questions:

1. *What are the objectives pursued by customer selection?*
2. *In what phases of Data Mining has most of the research on customer selection been conducted?*
3. *What selection techniques are most often researched empirically and used by practitioners?*
4. *How do selection techniques relate to specific marketing objectives?*
5. *How is the attainment of these objectives assessed by researchers?*
6. *What is the empirical evidence comparing the effectiveness of customer selection techniques?*

In line with these questions, the findings from the systematic review are the following:

1. Objectives pursued by customer selection studies are: new customer acquisition and activation, customer penetration, retention/churn and loyalty, lifetime value maximisation and pure segmentation
2. Data Mining techniques are not used to address appropriate marketing objectives. This was true for academic research as well as in practice. However, in the case of this study the most researched objective was customer penetration.
3. The pattern discovery and modelling phase is the most pursued in the systematic review subset, though this was expected given the Data Mining orientation of the research objective.
4. Objectives are assessed using multiple types of performance measures. The most utilised performance measure issued from the review was disproportionately response rate. In the case of technique performance, there are two types of performance measures: comparative model performance measures and fit/predictive accuracy measures. Model performance measures that emerged as being most popular in this study include: gains charts, lift charts and response rates by depths of file. Most utilised Fit measures included: PCC and AUC (and to a lesser extent, R-square).
5. The most often utilised techniques were RFM, cluster analysis, logistic regression, neural networks and AID/CHAID; with statistical techniques being used twice as much as machine learning techniques.
6. The techniques showing the least promise were RFM/FRAC, log-linear models and CART, whereas the techniques showing the most promise were AID/CHAID and Neural networks (more specifically Bayesian Neural Networks). All studies that examined the impact of using incremental data variables showed positive performance effects related

to the addition of variables to models. In addition, studies that applied additional phases of the Data Mining process, such as data transformation to reduce dimensionality, showed strong incremental improvement potential.

This review appears to be one of the very few systematic comparisons of technique effectiveness and, to my knowledge, the first to dissect technique application by tasks and phases of data mining. It makes a contribution by explicitly identifying how each technique informs the achievement of specific marketing objectives and to what degree the Data Mining process is applied in objective achievement.

4.14 Conclusion

The literature review allowed me to express a point of view on a topic, shedding light onto how the topic is best investigated and providing a critical assessment of the revised literature in relation to a research proposal. The systematic review approach allowed me to segue from the topical review of the topic, design and methodology to a specific review of these areas. In addition, the systematic review allowed me to do so in a manner that is transparent and methodical, which is critically important given the criticism of research as being too technique centric.

Given the amount of available literature on the topic, the systematic literature review provided answers to key questions informing the research topic in addition to establishing a strong baseline of information required to adequately design the empirical research studies that follow. The systematic review outputs allowed me to define the metrics used to assess success, identify techniques with the highest prevalence and performance, and understand what variable might be most effective in increasing technique prediction. These outputs allowed me to effectively select techniques, data variables and design test propositions for the two subsequent empirical studies.

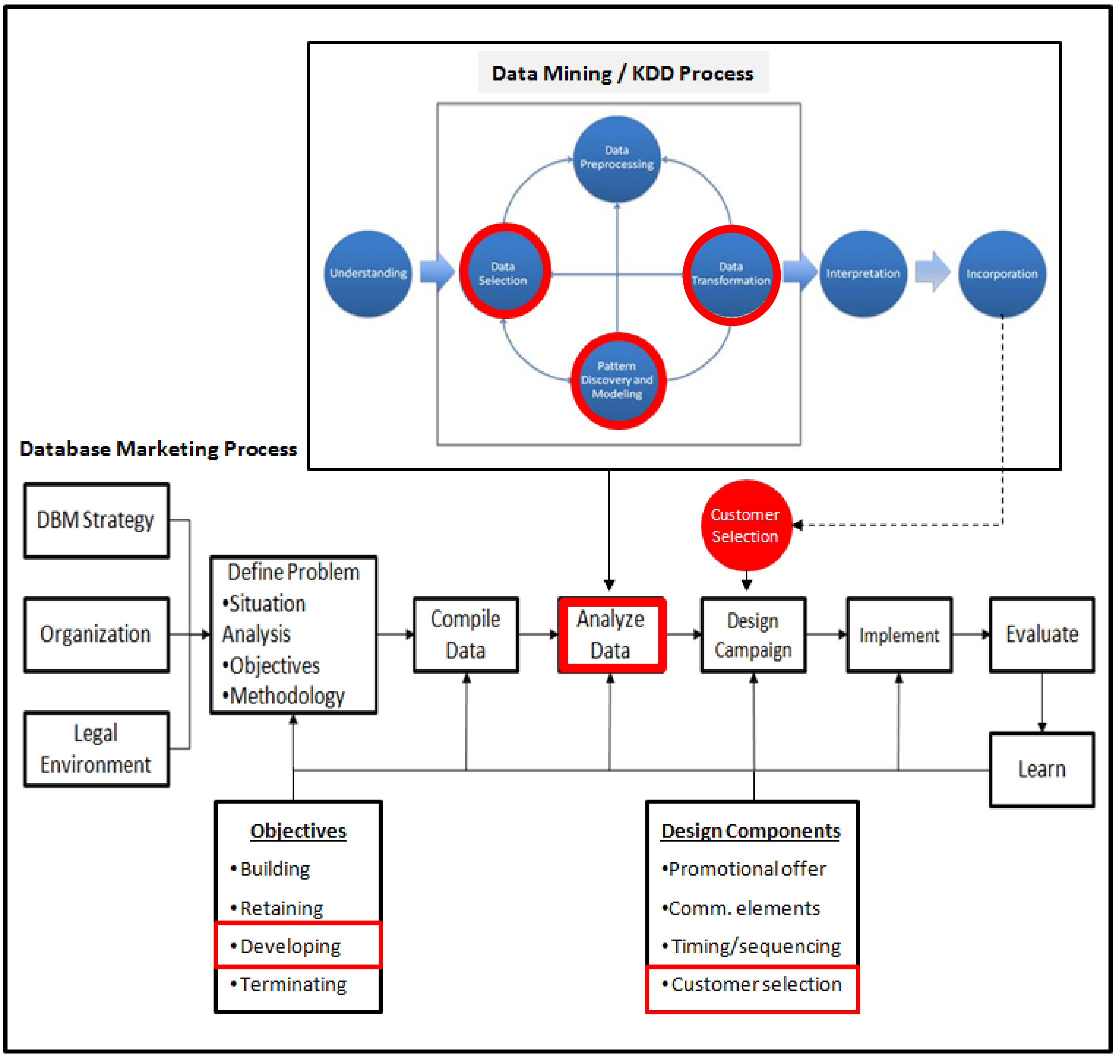
More specifically, high prevalence and performance techniques were identified for comparative testing. These included: RFM, stepwise logistic regression, CHAID, and Neural networks. As one of the most applied techniques, RFM will be mainly used to establish a baseline level of performance against which other techniques will be compared. Given the growing body of literature on the significant impact of incremental data on model performance, two depths of data are tested. Shallow depth of data techniques, composed of the generally available transactional variables of recency, frequency and monetary expenditures, are tested against deeper depth of data techniques, composed of extended transactional and non-transactional variables obtained from the extensive customer database. The four techniques are tested against one another. In Study 2, only RFM variables are tested, while in Study 3, extended data variables are added and their impact on performance examined. In addition, given the important significance of dimensionality reduction to performance enhancement, rather than apply PLS or PCA techniques as a separate step to the Data Mining process, improvements will be imbedded into the utilised techniques and, as a result, a stepwise logistic regression (that optimises variable integration into the final model run) will be utilised versus a simple logistic application (that would result in necessarily sub-optimal results).

FMCG retail was chosen not only because of the value of such a contribution to its context but also because of the availability and widespread use of customer data. Literature review findings also directly affected research design decisions. Given their prevalence in the broader academic and practitioner literature, the RFM technique is used as a baseline for comparison. Techniques against which it is compared are CHAID, Stepwise Logistic Regression and Neural Networks.

Given all the salient elements that emerged from the discussion section and the critical importance of applying a systematic process to the practice of customer seleciotn, I have detailed in Figure 14 the areas from the Database Marketing and the Data Mining processes that will be informed by studies two and three. The figure illustrates that: (1) objectives for the research will be restricted to customer development, (2) design components will aim to select appropriate customers for campaign targeting purposes, and (3) applied stages of the Data Mining process will include Data Selection (informed by the variable literature), Pattern Discovery and Modelling (informed by the technique literature) and the dimensionality reduction applications in Data Transformation. Given data is sourced from a FMCG retailer's transactional and relational databases, input variables were already transformed (not including dimensionality reduction) and were preprocessed to optimise response models. This is very similar to Ho et al.'s (2005) and many similar studies that utilise predetermined datasets provided by third parties and prequalified based on internal conceptions of high predictability. As a result, the study does not cover transformation and preprocessing phases of data mining.

This conceptualisation allows for a direct linkage between the extraction of customer knowledge from databases with the organisation and application of knowledge for the purposes of making effective marketing decisions in low-involvement contexts. This will enable what Shaw et al. (2001, p. 128) relate to as true customer relationship management, that integrates "the knowledge discovery process with the management and use of the knowledge for marketing strategies (that will help) marketers address customer needs based on what the marketers know about their customers, rather than a mass generalisation of the characteristics of customers".

Figure 14: Database Marketing and its Related Processes and Inputs

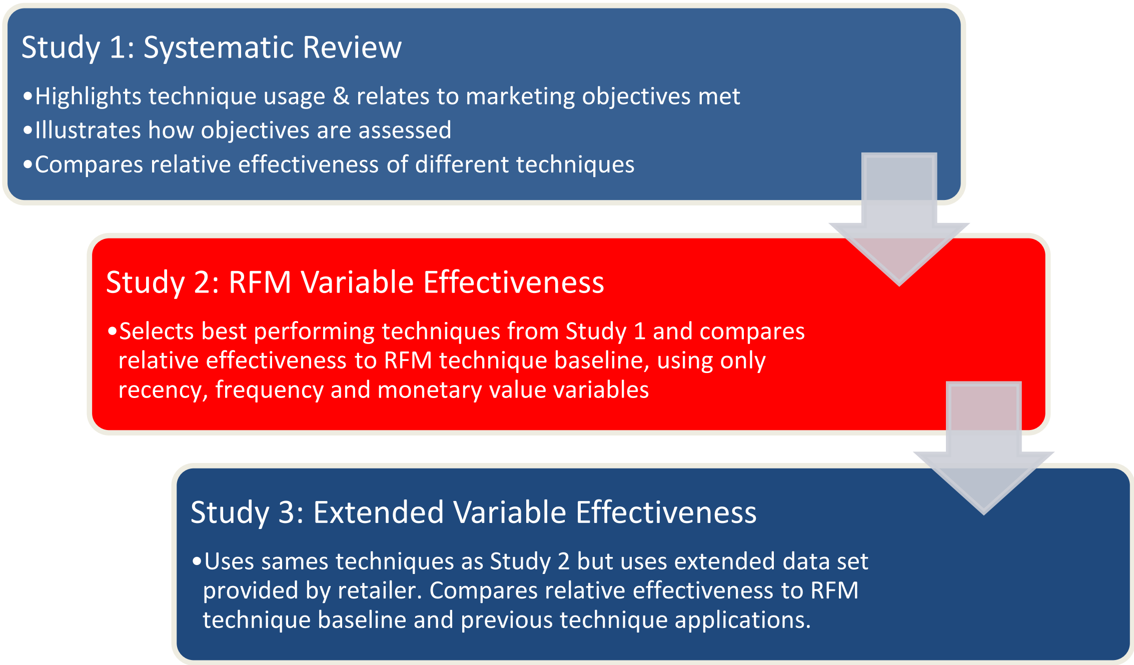


5 Study 2: Effectiveness of Data Mining techniques with RFM Variables

5.1 Introduction

Study 1 researched some key questions informing the research topic in addition to establishing a strong baseline of information required to adequately design the empirical research studies that follow. This baseline includes some of the most used metrics of success, most effective data modelling techniques and variables that enhance predictive strength. Study 2 takes these findings and incorporates them into a quantitative study aimed at understanding technique effectiveness in shallow (RFM variable only) data environments in the context of FMCG retail. Figure 15 illustrates how each study informs the next and the specific objective of Study 2.

Figure 15: Thesis Research Progression Overview



Given their prevalence and high overperformance rating from the systematic review, and the capacity to integrate dimensionality reduction, techniques tested against one another include: RFM, stepwise logistic regression, CHAID, and neural networks. RFM is used to establish a baseline level of performance against which other techniques are compared. Data variables used to operationalise techniques include the generally available transactional variables of recency, frequency and monetary expenditures. Performance is assessed using response rate.

The next section discusses the methodology followed to perform empirical testing, the statistical modelling outputs, a presentation of empirical results and analysis, and a discussion of findings. The methodology section explains the steps taken to select customers most likely to respond to a promotion, more specifically: (1) the dataset and variables used as inputs for statistical techniques, (2) an in-depth review of the techniques applied in the study, (3) the measures of performance used to compare technique effectiveness, (4) the procedure used to achieve a representative and significant output from the statistical exercise, and (5) the final overarching research design used to answer the research question. The statistical modelling and output generation section details how statistical modelling was conducted in collaboration with a private third-party firm. The empirical results and analysis section specifically describes the degree of reliability of the sampling methodology, the comparative performance of techniques, and the relative performance of parameters (variables). Finally, the discussion section examines the findings of the study and provides an assessment of results vis-à-vis literature. For parsimony purposes, contributions are only discussed at the end of the thesis using outcomes from all studies.

5.2 Methodology

5.2.1 Dataset

The dataset is provided by a large North American FMCG retailer with 30-percent market share and 90-percent market coverage in the region in which it operates. The promotion from which data is sourced was held in January 2009 and was communicated to consumers via direct mail. The promotion period lasted one month, meaning that customers had one month to take advantage of the offer. The retailer's loyalty programme rewards customers with points for dollars spent, and also allows the retailer to offer customers point incentives for specific purchase behaviours. Within the context of this specific promotion, customers earned double the loyalty programme points on their purchase.

The dataset contains both the data on customers targeted for the promotion as well as results of the targeted direct marketing promotion. Results include indicators for all consumers that were targeted and responded to the promotion. The dataset comprises 152,157 loyalty card members that received the direct marketing promotion from the retailer. The average response rate of customers targeted for the promotion was 15.5 percent. The promotion recipients were consumers the retailer identified as its top 30 percent of loyalty programme members based on 12 months of expenditures. This is consistent with common direct marketing practices where, due to budgetary and other considerations, mainly the top deciles within a customer base are selected to receive promotions (Zahavi and Levin, 1997). Since a marketing campaign intends to contact only customers in top deciles, the performance of a model over the entire dataset is less relevant. In this case, the goal of a model is to select a specific target group for analysis and interpretation; its performance on the rest of the population is of little or no material consequence (Cui and Wong, 2004).

5.2.2 Variables

The variables used for this study include some of the more prevalent customer data variables available to marketers: recency of last purchase, frequency of purchases, and monetary value of purchases. The use of RFM variables and RFM techniques first emerged from the recognition by catalogue marketers that these three variables were particularly well suited to predicting the likelihood that specific groups of customers out of the mass of customers on total order files would respond to particular offers (McCarty and Hastak, 2006).

In a world of emerging data richness, such limited data variables may seem less relevant. However, the use and importance of recency, frequency, and monetary value variables has been recognised by direct marketing professionals and academics for decades (Baier et al., 2002). In fact, the use of RFM variables and segmentation is perhaps "the most widely recognised behavioural analysis technique...and the easiest and fastest methodology to implement" (Kahan, 1998, p. 491). Given the suboptimal use of techniques by practitioners highlighted in the systematic review and the still very prevalent use of RFM as a basis of segmentation and selection, I felt it critical to establish a performance baseline that was restricted to these three variables using the traditional and recognised RFM technique.

Definition of these terms is quite intuitive and self-explanatory. Recency is a measure of the time elapsed since a customer's last purchase in a specific period. Frequency indicates the amount of transactions the customer has conducted in a specified period. And monetary value is the expenditures of the customer (pre-taxes) in the same period. The examined time periods were one month and one year prior to the selection of these customers for the promotion. These time periods were selected given their immediate availability with the retailer's database and are consistent with other research periods used in retail contexts. The monthly time period is particularly relevant given that, in non-contractual settings such as FMCG, retail customers can continuously change their purchase behaviour without having to inform the retailer. Research shows that more than 70 percent of all customers shop in several supermarkets in any given month (Buckinx and Van den Poel, 2005). This indicates a high volatility of purchases with a need for retailers to not only understand the full value of a customer, but the monthly value as well in order to deliver offers that are as relevant as possible close into their current purchase period. The 12-month period addresses the full customer value perspective and is particularly relevant given it is the standard period used to calculate brand loyalty and retention (Rust et al., 2004).

In order for these variables to be effectively integrated into the technique applications, a summary of each customer's transactional history was created, based on the following: (1) date of the last or most recent purchase, (2) total number or frequency of purchases, and (3) total amount spent per order. Where required (techniques not capable of handling date formats), to calculate the actual recency, the date of the last purchase was translated into the number of days since the last purchase. The frequency of purchases field is a summation of all the individual number of transactions from the customer during the period. The total monetary value is the summation of the total amount spent for all purchases in the period.

5.2.3 Techniques

RFM (Recency, Frequency and Monetary Value)

As noted by Olson and Chae (2012) and McCarty and Hastak (2006), RFM can be applied in a number of different ways. The empirically based RFM technique was advocated by Arthur Hughes (2000). "The first step in the approach is for the marketer to sort the customer file according to how recently customers have purchased from the firm. The database is then divided into equal quintiles and these quintiles are assigned the numbers 5 to 1. The next step involves sorting the customers within each recency quintile by how frequently they purchase from the marketer. For each of these sorts, the customers are divided into equal quintiles and assigned a number of 5 to 1 for frequency. Each of these groups (25 groups) is sorted according to how much money the customers have spent with the company. These sorts are divided into quintiles and assigned numbers 5 to 1. Therefore, the database is divided into 125 roughly equal groups (cells) according to recency, frequency, and monetary value" (McCarty and Hastak, 2006, p. 657). After responses from a promotion are received, responder proportions are calculated for each cell, and cells are then ordered by highest to lowest response rate (usually by deciles or percentile). Marketers typically target the customers contained in percentiles above a promotion's estimated break-even point for subsequent marketing campaigns.

Another common approach to RFM is known as hard coding (Drozdenko and Drake, 2002), where each variable is assigned a weight and an RFM score is calculated for each customer in the database. Weights can be derived from researcher judgement or calculated empirically based on historical responses to campaigns.

For this research, the Hughes approach is adopted given its simplicity and widespread application (Kohavi and Parekh, 2004). The notion of simplicity for selecting this type of RFM approach is important given the objective of the research is to illustrate the gains provided by applying advanced Data Mining techniques vis-à-vis simple, intuitive and widespread applications. The Hughes approach meets these criteria and is indeed the simplest yet well understood application of RFM in practice (McCarty and Hastak, 2006; Kohavi and Parekh, 2004). Though it is simpler than other techniques it is tested against, some researchers have gone so far as to claim that segmentation and customer selection based on RFM variables explain between 75 and 85 percent of direct marketing success (Retka, 2001, Hughes, 2000).

CHAID (Chi-squared Automatic Interaction Detection)

CHAID is a decision-tree technique application. Decision trees are “intuitive methods for classifying a pattern through a sequence of rules or questions, in which the next question depends on the answer on a current question. They are particularly useful for categorical data, as rules do not require any notion of metric” (Crone et al., 2006, p. 784).

“Basically, all automatic tree classifiers share the same structure. Starting from a ‘root’ node (the entire sample), tree classifiers employ a systematic approach to grow a tree into ‘branches’ and ‘leaves.’ In each stage, the algorithm looks for the ‘best’ way to split a ‘father’ node into several ‘children’ nodes, based on some splitting criteria. Then, using a set of predefined termination rules, some nodes are declared as ‘undetermined’ and become the father nodes in the next stages of the tree development process, some others are declared as ‘terminal’ nodes. The process proceeds in this way until no node is worth splitting any further as ‘terminal’ nodes define the resulting segments” (Levin and Zahavi, 2001, p. 5). What makes CHAID different from other decision tree techniques is that its output is based upon adjusted significance testing (Bonferroni testing). Gelman, Hill and Yajima (2012, p. 191) define the Bonferroni correction as follows: “The Bonferroni correction adjusts the p value at which a test is evaluated for significance based on the total number of tests being performed. Specifically, the working p value is calculated as the original p value divided by the number of tests being performed.” This means that tree nodes are split differently than other decision tree techniques in that they use an adjusted p-value calculation (Bonferroni adjusted p-value) (Hill, Thomas, and Lewicki, 2006). One of the main advantages of this correction is that it protects against Type I errors.

CHAID is similar to RFM in the sense that it creates groupings of customers; however, these groupings (called nodes) are not determined a priori nor are quantities predetermined. Nodes are generated based on individuals’ response rates from historical data. As mentioned in the previous paragraph, the process starts with a root node that includes the entire sample. Then, based on the independent RFM variable that best discriminates customer response, CHAID splits the root node into as many children nodes as are significant. The process continues until further splitting of nodes becomes insignificant. As mentioned earlier in the systematic review, this all-in-one process for discriminating variables (dimensionality reduction) based on their significance to the overall model predictive power inherently means that decision trees and CHAID applications apply both transformation and pattern discovery and modelling phases of Data Mining in an integrated manner.

The output can be evaluated in much the same way as RFM with terminal nodes being prioritised according to revenue per responder, cost per responder, profit per responder, break even, and profit margin. It is also similar to RFM in that it can accommodate relationships between dependent and independent variables that are non-monotonic. However, the main difference between both approaches is the number of variables that can be accommodated by both techniques. CHAID can accommodate an unlimited variable set (including recency, frequency, and monetary value), whereas RFM is by definition restricted to its constituent variables. In this study, only RFM variables are incorporated in the independent variable set.

Logistic Regression

"Logistic regression is a modelling procedure where a set of independent variables are used to model a dichotomous criterion variable. Therefore, it is appropriate for direct marketers who would like to model the dichotomous variable of respond/don't respond to a mailing. Logistic regression is particularly useful in these circumstances in that the actual criterion variable is dichotomous; however, the predicted variable is the response probability, which varies from zero to one. Therefore, the model can provide a probability of response for everyone in the file, given the estimated parameters for a set of predictor variables" (McCarty and Hastak, 2006, p. 658).

Although logistic regression, also referred to as logit, finds a 'best fitting' equation just as linear regression does, its principles are different. "Instead of using a least-squared deviations criterion for the best fit, it uses a maximum likelihood technique, that is, it maximises the probability of obtaining the observed results given the fitted regression coefficients" (Jackson, 2002, p. 275). Because logistic regression does not make any assumptions about the distribution for the independent variables, it is more robust to violations of the normality assumption.

Comparing RFM and CHAID to logistic regression, McCarty and Hastak (2006, p. 658) identifies two major differences: "first, logistic regression provides a response probability for individual members of the dataset rather than creating discrete groups of people. Therefore, in theory, each person in the dataset may have a different response probability. Second, for continuous predictor variables, logistic regression model relationships of the independent variables with the dichotomous dependent variable that are monotonic, both RFM and CHAID are distribution free. This has implications for the performance of logistic regression in instances where the relationship between a predictor variable and the response variable is neither continuously increasing nor decreasing. For example, when the relationship between recency of previous purchase and purchase on the test mailing is curvilinear, logistic regression may not be able to capture the relationship in ways similar to that for RFM or CHAID."

The logistic regression can be applied in its purest form with all variables being included in the model. This approach necessarily leads to an increase in variable collinearity and model volatility. Though it is the simplest application and sometimes effective at increasing a model's explanatory power (R^2), it is often applied in a manner that is referred to as stepwise. A stepwise logistic regression involves a stepwise selection of variables, meaning that variables are added to the logistic equation one-by-one and then tested to see whether the model performs better with the variable included or excluded.

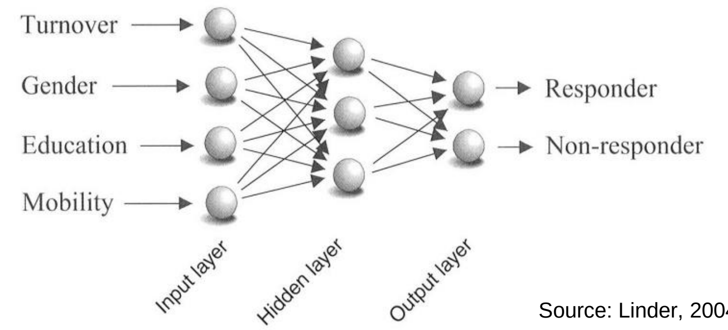
The stepwise logistic regression is used in this thesis as it allows for the inclusion of large variable subsets into a logit equation and is intended to find the most parsimonious set of predictor variables. In a stepwise regression output, one should not be surprised to only observe significant variables in the model output as all other variables are already pre-excluded for non-significance. The addition of the stepwise logistic regression technique is consistent with the studies of Lix et al. (1995), Malthouse (1999) and Deichman et al. (2002) whereby the logistic approach is used to increase technique effectiveness through dimensionality reduction. It would be a lost opportunity to strictly apply a standard logistic regression technique given that CHAID and neural networks also inherently decrease dimensionality. In a case such as FMCG retail, where predictive variables for promotions have not been thoroughly studied, stepwise methods are particularly relevant given they have been proven to be valuable in situations where little or no previous research exists on which to base hypothesis for testing or in situations where model fitting is the primary concern (Menard, 1995, Agresti and Finlay, 1986).

The specific stepwise logistic regression applied in the context of this research is the forward stepwise logistic regression. The forward method starts with a basic model that contains only a constant and then adds predictors to the model. Variables are added or excluded based on their capacity to increase the score statistic (the contribution to prediction of the dependent variable). The addition of variables continues until none have any additional contribution to the score statistic. The model also constantly re-evaluates already retained variables to determine whether they should be removed in light of overall contribution of the newly added variable and the rest of the model variables to the score statistic. The model does so using the likelihood ratio statistic, also referred to as the Forward: LR method (Field, 2005). The forward stepwise regression method was selected over the backward stepwise method because of the very large set of potentially independent variables used as inputs and the generally exploratory nature of the variable reduction exercise. I say exploratory because there are little or no studies in the context of FMCG to refer to for variable significance and, while studies on other domains do exist, I wanted to adopt an exploratory approach to identifying significant variables. This necessarily introduces the risk of 'casting too wide a net' and identifying independent variables only accidentally (random error) related to the dependent variable. However, given the logit model is cross-validated ten times, results allow for the identification of variables that occur only accidentally.

Neural Networks

Crone (2006, p. 783) defines artificial neural networks (ANN) as "a class of statistical techniques capable of universal function approximation, learning non-linear relationships between independent and dependent variables directly from the data without previous assumptions about the statistical distributions.” Neural networks are modelled after the human brain in that they accept a number of inputs and process these inputs in order to forecast a foreseeable output. Just like a human brain, some of the association between these inputs and outputs are transparent and visible; but many of these associations are hidden. Input layers contain the visible inputs while a hidden layer is situated between input and output layers. Both layers are critical to the approximation of a more robust output function. Each of these inputs and hidden layers are called nodes. Given some of these inputs can be linked to a common hidden phenomenon or node, nodes take one or more input value and combine them into a single value. This facilitates the transformation of this value into an output value. This is illustrated in the Figure 16.

Figure 16: Neural Network Representation



Source: Linder, 2004

Neural networks can be either supervised or unsupervised. The more common of the two techniques is supervised neural networks, where samples of data containing both inputs and outputs are entered for analysis purposes. Each sample item outcome is compared to a desired output to determine item importance. This is repeated until performance is acceptable. Unsupervised networks work in much the same way except no output data is fed in. Rather, the network is trained to detect salient features that are then used to group inputs into different classes. As a result, neural networks with supervised learning algorithms can be used in lieu of regression and discriminant analysis (Venugopal and Baets, 1994). This is what was done in this thesis where inputs included the independent variables of RFM and response was the dependent variable.

The specific neural network application used in this study is the multilayer perceptron. The multilayer perceptron (MLP) is one of the most popular neural network applications. It is a supervised network because it requires an output (dependent variable) to learn from. The MLP and most other neural network applications use a back propagation algorithm to generate models. In short, this means that input data is repeatedly introduced to the network and each introduction generates an output that is compared to the desired output (with an error being subsequently calculated). This error is reintroduced to the network and used to adjust the weights of each input in order to decrease the error with each subsequent iteration. This process continues until the optimal output is reached. This process is often referred to as training (Silva, 2008).

As it concerns the calculation of the MLP error function, though the mean square error (MSE) function is commonly used, in cases of logistic outcomes SPSS uses the cross entropy (CE) error function for data classification and prediction problems. Ultimately, both measures can be used to calculate the error function, so this choice is one of convenience. As Plunkett and Elman (1997, p. 166) state, "the cross-entropy measure has been used as an alternative to squared error. Cross-entropy can be used as an error measure when a network's outputs can be thought of as representing independent hypotheses (e.g. each node stands for a different concept), and the node activations can be understood as representing the probability (or confidence) that each hypothesis might be true. In that case, the output vector represents a probability distribution, and our error measure – cross-entropy – indicates the distance between what the network believes this distribution should be (training) and what the teacher says it should be (testing). For cross-entropy measures, smaller values indicate that the training results in a network that classifies cases with a lower error rate (Hopfield, 1987; Bishop, 1995).

The use of cross-entropty is in line with work from Santos, Alexandre and Marques de Sá (2004) and Silva, Marques de Sá and Alexandre (2005).

5.2.4 Performance Measures

Performance measures include goodness of fit and performance measures. Traditional goodness of fit measures include R^2, F and Chi Square statistics and other various classification indices highlighted in the previous literature and systematic reviews. However, practitioners are typically concerned with strict measures of performance such as response rates and performance measurement comparison methods such as lift charts, gains tables and decile response rate comparisons. Though all of these measures are often still used in research today, they have limitations and must be used in combination and not in isolation.

Goodness of Fit

Greene (2010) states that traditional goodness of fit statistics such as R^2, the F statistic and the Chi Square statistic are not adequate for evaluating model performance "when the objective is to create models that maximise differentiation between population segments in terms of response rates. The traditional statistical measures are appropriate for assessing how well individual response values fit a given model (for example, minimise least square errors between response data points and predicted values); however, they are not appropriate for effectively selecting market segments or individual customers for targeting and meeting business objectives. The descriptive measures used by practitioners, although visually appealing, do not assess overall model performance with statistical certainty” (Greene, 2010, p. 36). Greene’s research subsequently recommends the use of the Gini index as a useful and valuable substitute measure for assessing a response model's predictive performance given it can also be used as a measure for consistency (reliability). Risselada, Verhoef and Bijmolt (2010), among others, also use it to compare the output of a random selection of customers with that of model-based selections.

The Gini coefficient is calculated by dividing the area between the cumulative lift curve and the 45-degree line (how much a model contributes to increase cumulative response versus a random selection) by the area under the 45-degree line. The area under the cumulative lift curve (AUC) is also a widely used measure of performance as illustrated in the output of the literature review. The relation between both Gini and AUC measures (and their interchangeability) is evidenced by their relationship in the following formula:

Gini coefficient = 2 x AUC–1

Gini coefficients are normalised so that a value of 1.0 corresponds to the best-fitted lift curve. Therefore, higher values imply better overall model fit. Given the sample size of the dataset and the high registered response rates, the Gini measure is extremely appropriate given its inherent stability and insignificant standard errors when data files are larger than 60,000 customer records (Greene, 2010).

Because of its widespread usage, the other measure used to assess model fit is the percentage correctly classified (PCC) measure. The PCC measure is a central output of many models including neural network and CHAID but not from the logistic regression output. (It is, however, calculated separately for purposes of this study.) It is not a measure that is possible in the case of the pure RFM method. Nevertheless, given its prevalence, I will report it, along with the Gini coefficient for relevant techniques.

It is important to note that, although both measures are referred to as measures of fit, they measure somewhat different dimensions of fit. PCC compares the actual yes/no response behaviour of a customer with the predicted behaviour and, thus, measures divergence. The Gini coefficient for its part (and the AUC score and ROC curves AUC it is derived from) reflects how well the two groups are separated from one another (Thomas, 2009).

Performance Measures

Though the Gini coefficient is an excellent measure of fit, it must be complemented by adequate measures of comparative performance. Methods retained for this purpose are lift and gains charts (Shepard, 1995) (also referred to as top-decile lift measures by Malthouse (1999), and Neslin et al. (2006)). As shown in the systematic review, these methods are very prevalent in quantitative model research and provide more granular model performance information at a decile level. As a reminder, lift and gain charts are a visual aid that complements decile analysis and measures a predictive model's effectiveness as the ratio between results measures obtained with and without the application of the model.

The decile breakdown of lift and gain charts provides a breakdown of performance between techniques at the decile level. This is presented from the decile most likely to respond (on the left of the X axis on charts) to the decile least likely to respond (on the right of the X axis on charts). When authors refer to lift and gains charts, it is important to mention that they usually calculate these for individual deciles and cumulative deciles. Decile level representation provides a view on the individual decile performance, whereas cumulative representations allow a researcher to compare aggregate top performing deciles and establish cut-off levels where declining marginal cumulative performance rates make inclusion of incremental deciles prohibitive. A visual representation of these charts can be found in Figure 17 below.

Figure 17: Sample of Lift Charts

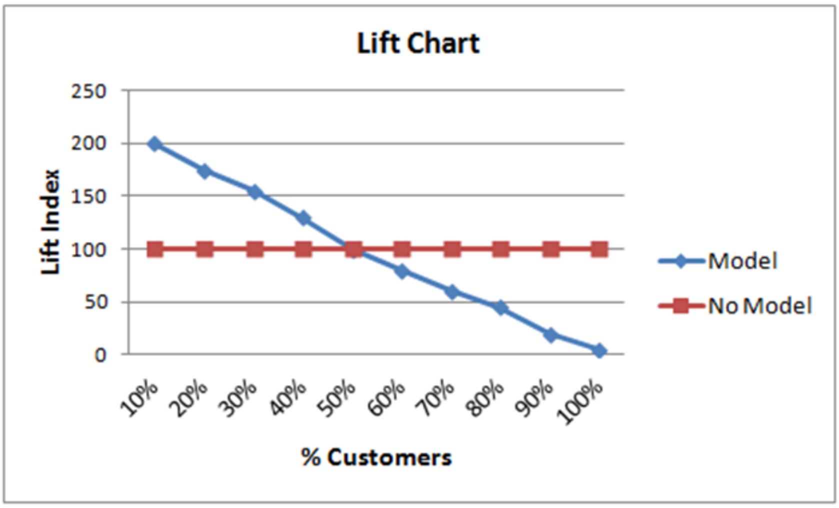
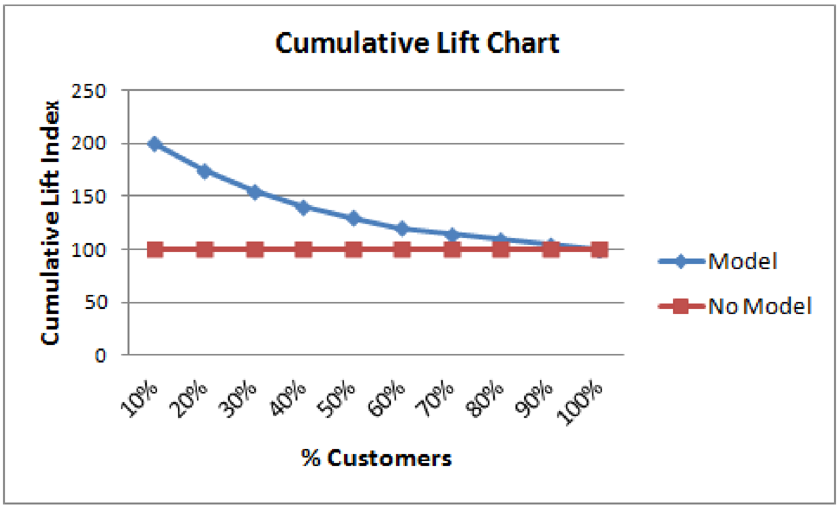


Figure 17 (continued): Sample of Lift Charts



More specifically, the lift score of each decile is calculated as follows:

Lift = response rate for each decile / overall response rate ×100

The accompanying cumulative lift score builds on the decile lift score and provides a view of how the top-scoring deciles perform as a cumulative group between different techniques. Cumulative response rate is calculated as follows:

Cumulative Lift: cumulative response rate / overall response rate ×100

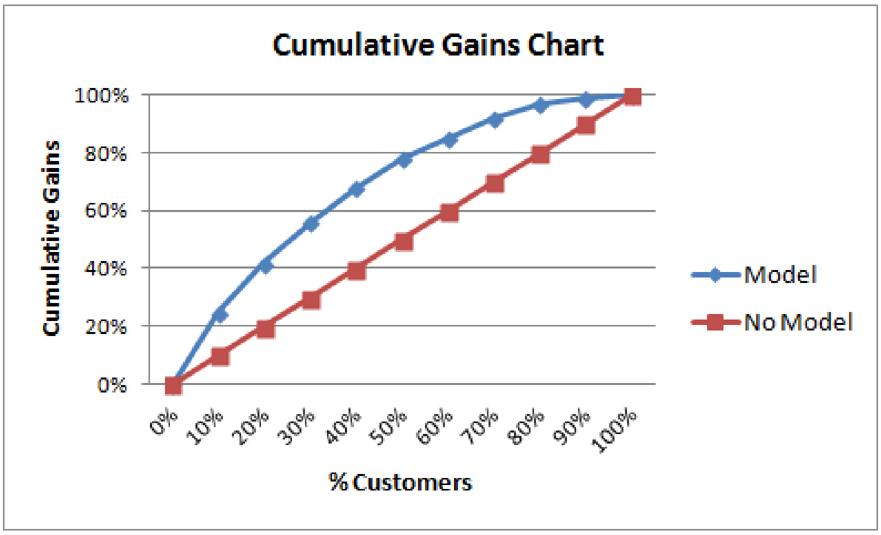
Gains charts are commonly created after lift charts and offer a common index representing the ratio of the decile or segment to the total sample (which is then multiplied by 100). Calculations for both gains and cumulative gains charts are as follows:

Gains = responders per decile / overall responders x 100

Cumulative Gains = cumulative responders / overall responders ×100

Given deciled lift and gains charts do provide similar information, for parsimony purposes only the cumulative gains chart will be used (in addition to both lift charts). A visual representation of this chart can be found in Figure 18.

Figure 18: Sample of Cumulative Gains Chart



Cui et al. (2008) is particularly interested in the continued use of these methods of comparative performance, given that no method is universally accepted for all performance criteria and validation methods. He highlights the importance of selecting the right performance criterion and validation in predictive modelling. He concludes, stating that given limited resources, "the effectiveness of target marketing efforts requires models that render consistently superior performance on the top deciles of the file. As opposed to overall performance in correct classification, cumulative lift remains the golden rule for assessing the true performance of direct marketing models" (Cui et al., 2008, p. 289).

It is important to note that, although response rates by decile could also have been used as a direct measure of performance, it was not used in this research. Unlike lift and gains charts that normalise outputs, response rates at a decile level are issued from different cross-validation applications (see next section) that can vary and generate somewhat different response rates. Thus, comparing response rates by decile can sometimes be misleading and lead to faulty interpretations. Nevertheless, response rates by decile constitute the backbone of decile and gains chart outputs and as such they remain a critical input.

In short, lift measures assess the relative effectiveness of a predictive model. Lift charts reveal the greater likelihood of identifying responders if specific models are used versus a random sample. Gains charts show the cumulative percentage of total responders acquired by targeting at different decile depths. Cumulative lift and gains charts are thus critical measures for comparing different models' performances. To do so, both include a baseline (referred to as the no-model scenario) that allows for the visual comparison between the lift or gains achieved by the application of different models and the baseline "no model application" (shown by the 100 index).

5.2.5 Sampling Procedure

When building and validating the performance of models based on historical data, researchers take a subset of customer data, utilise it to derive required insights and then test these insights on the remaining data for validation. If data from the sample used to derive insights is biased, so are the results of its downstream applications. Therefore, it is critical for data sampling procedures to be as accurate as possible in order to ensure that validity of derived insights.

Random sampling is, typically, the starting point of any robust sampling exercise. A population is sampled in order to make inferences about the entire population. Random sampling is critical since it helps eliminate the biasing influences of confounding variables. However, even with significant sample sizes, one important consequence of random sampling is sampling variability: where different samples provide different results. For example, luck of the draw may have produced a highly fitted model that does not generalise to the rest of the population and thus performs poorly on unseen cases.

Theoretically speaking, an unlimited number of samples will lead the error rate of the samples to equal the true error rate of the full sample. However, the issue isn't so much sampling as it is the often-limited data available for sampling. In limited data sets, systematic sampling procedures are necessary to ensure proper sampling outputs. The idea behind sample validation (also referred to as re-sampling) is to attempt to reduce sampling variability by adding more rigour to the sampling process by adding more samples to the exercise. Validation procedures are critical for evaluating the performance of a model that is derived from training data onto testing data as different derivations may have different degrees of sturdiness when dealing with sampling variations (Berardi and Zhang, 1999). Samples can be exclusive or overlap depending on methods used or the amount of data available. Since validating model performance using future data is often impractical, in many cases samples are picked from the very data set from which inferences will be made. Typically researchers use one of several sample validation procedures, including data partition, bootstrap, and k-fold cross-validation (Kohavi, 1995; Malthouse, 2001) to generate effective training and test sets and add statistical rigor to individual decile computations.

Data Partition

Data partition involved splitting the data into two groups of equal size, with one group being used for training and the other for testing. Given the limited number of iterations, typically error rates from this approach lead to a poor estimation of future performance. Cui et al. (2008) state that training and testing a model using simple data partitioning is not sufficient to effectively ascertain a model's performance. They suggest multi-partition approaches such as cross-validation to provide a less biased estimate of model error rates (Berardi and Zhang, 1999;

Cooil et al., 1987). Some of these cross validation approaches include leave-one-out cross-validation and repeated random subsampling. Leave-one-out cross-validation expressly reserves a testing sample up front and trains a model using the rest of the data. However, this method shows a higher than average variance when training sets resemble the full data set too much (Efron and Tibshirani, 1993). The repeated random subsampling method carries out the sampling exercises n times and doesn't exclude the testing data from the training data. As a result, when datasets are medium or small, it is possible for some of the subsamples to be biased and unrepresentative of the full data set.

Bootstrap

Cui et al. (2008) identifies two bootstrap approaches that seem to generate best classification results: e0 and the 0.632 bootstrap. For e0 boostrapping, training samples are drawn from a broader sample and returned after use (repeated use is allowed). The e0 estimator is approximated by repeating the sampling twice (two folds), using a split of 50/50 of train-and-test cases. The 0.632 bootstrapping, for its part, goes beyond two folds and continues to draw training samples until the number of cases in the total samples is equal to the total cases in the full sample set. The key benefit of 0.632 bootstrapping is that replacements allow the researcher to have enough cases to achieve stable estimates of parameters in a situation of smaller data sets. In bootstrapping, the error rate is calculated as the average of the error rates over the number of samples extracted.

K-fold cross-validation

K-fold cross-validation randomly selects cases and divides them into k training groups (folds) of equal size. The last group that is held out is the test group. This is done k times. The cross-validated error rate is calculated as the average error of all k groupings. K-fold cross-validation is significantly different from bootstrapping. In k-fold cross-validation, all available data sets are used for training and all cases are also used for testing. In addition, k-fold cross-validation does not use reuse cases in different folds (each fold is unique and cannot have duplicate cases with other folds). As a result, there is never any overlap between training and test data sets. These differences reduce overfitting. However, one of the main issues of the cross validation approach is the variance in estimated error rates of test samples, particularly with small datasets.

One of the most commonly used and highly discriminating choice of methods applied to a sizeable dataset is a form of k-fold cross-validation called tenfold cross-validation (Yang, 2007). Experimental results from Kohavi (1995) and Cui and Wong (2004) comparing the effectiveness of cross-validation and bootstrapping in the selection of predictive models have shown that cross-validation, and more specifically ten-fold cross-validation, was more effective in yielding accurate results.

In fact, empirical research on ten-fold cross-validation with large datasets conducted by Kohavi (1995) and Cui and Wong (2004) illustrate that that ten-fold cross-validation yields more accurate results than bootstrap validation with lower error rates and variances. Mitchell (1997) and Kohavi (1995) both show that ten-fold cross-validation seems accurate and very appropriate for uses with large samples. More folds are not shown to add much incremental value.

For more clarity, ten-fold cross-validation involves partitioning the original data sample into ten randomly split sub-samples (these are also commonly referred to as folds). Of the ten sub-samples, one sub-sample is retained as the validation set for model testing, while the remaining nine sub-samples are utilised as the training set. The process is repeated ten times, each of the ten sub-samples being used once as the validation set. The ten result outputs from each repetition are then averaged to produce a single combined output. The major benefit of using this method over others that use repeated random sub-sampling is that cross-validation uses all available observations for both training and validation.

5.2.6 Dimensionality Reduction

The inherent complexity and opportunity presented by dimensionality reduction in the dynamic programming of CRM campaigns is not only captured in the author output from the systematic review, but is also identified in the work of Kamakura, Mela, Ansari, Bodapati, Fader, Iyengar, Naik, Neslin, Sun, Verhoef, Wedel and Wilcox (2005, p. 286) where they state that “as CRM data typically contains many variables, the states and controls ‘explode,’ leading to the appropriately called ‘curse of dimensionality,’ which hinders estimation.”

As stated in the systematic review portion of this thesis, Kantardzic (2003) indicates that two feature reduction approaches exist: the empirical bottom-up and heuristic top-down. Approaches that emerged from the research were all bottom-up empirical approaches. The most popular statistical method for empirical dimensionality reduction are PCA and PLS (Lee et al., 2011). In addition, other commonly used methods for purposes of data reduction include stepwise logistic regression, decision trees (CART, CHAID, C4.5, ID3), neural networks and other machine-learning techniques (Weiss and Kulikowski, 1991). Finally, sliced inversed regression is a less used but not uncommon method to reduce variable complexity (Li, 2010); however, it is more prevalent in the natural science fields.

Indeed, in the systematic review of findings, principle component analysis (PCA), partial least squares (PLS) and sliced inversed regression did emerge as data reduction methods. In addition, ridge regression was used as an alternative to standard data reduction methods to manage the issue of multicollinearity. All these methods are applied prior to predictive techniques. Other techniques that reduced dimensionality, albeit in integrated and all-in-one application manners, include stepwise logistic regression, CHAID and neural networks. In all cases, dimensionality reduction did allow for enhanced performance given the subset of utilised variables shown less multicollinearity and thus provide a better level of predictive accuracy.

As a result, one of the additional aims of this study is to demonstrate the benefit of extending the customer selection process deeper into the KDD process by adding dimensionality reduction to the already effective use of data and advanced techniques. Lee et al. (2011) and many authors apply dimensionality reduction in a sequenced fashion to build an accurate and concise predictive model. Lee's experiment demonstrates that successful initiation and management of marketing programs (in his case, churn management programs), both predictive accuracy and model comprehensiveness, are critical. In sequenced applications, the first step includes the application of PLS and PCA methods for reduction, while subsequent steps aim to achieve prediction.

The approaches used in this study are not sequential but rather integrated applications that reduce dimensionality as part of an overall technique application. In sequenced and integrated applications, the result is to increase the strength of models applied in high variable contexts with potentially high correlations among one another. The selection of integrated approaches such as CHAID, ANN and stepwise logistic regression is convenient because integrated approaches were identified as also being the most effective individual techniques that emerged in the systematic review. Demonstration of dimensionality reduction improvements will be illustrated by the exclusion of specific variables that decrease individual models' contribution to predicting response.

I recognise that by applying a stepwise logistic regression I am not applying one of the dimensionality reduction techniques preferred by Lee et al. (2011) (PLS or PCA). However, I am applying a recognised approach that, in the context of this research, has the added benefit of demonstrating not only a measure of improvement versus other techniques but also a measure of improvement in the context of the specific regression technique application. In short, the usage of integrated approaches will determine whether all variables truly allow a model to increase prediction or whether some variables simply do not. This will also be contrasted with the dimensionality reductions of other similar methods.

5.2.7 Study Design

The thesis aims to address two gaps: (1) the relationship between customer selection techniques and performance and (2) the impact of dimensionality reduction on performance. Two sets of propositions are tested to address these gaps and assess the effectiveness of the aforementioned choices. The first set of 12 propositions contained in this study tests the effectiveness of customer selection techniques using recency, frequency and monetary value of purchases only, as these three variables represent the lowest common denominator for all techniques. The set of 12 propositions is detailed in Table 48.

Table 48: Technique Effectiveness Propositions

RECENCY, FREQUENCY & MONETARY VALUE VARIABLES ONLY					
		More Effective than			
		RFM	CHAID	Logistic Regression	Neural Networks
Less Effective than	RFM	NA	P4	P7	P10
	CHAID	P1	NA	P8	P11
	Logistic Regression	P2	P5	NA	P12
	Neural Networks	P3	P6	P9	NA

These propositions can be read as follows:

In the context of FMCG retail promotions aiming to maximise the response rate of direct to consumer promotions using only recency, frequency and monetary value of purchases variables, the technique of:

P1: RFM is more effective than CHAID

P2: RFM is more effective than Logistic Regression

P3: RFM is more effective than NN

P4: CHAID is more effective than RFM

P5: CHAID is more effective than Logistic Regression

P6: CHAID is more effective than NN

P7: Logistic Regression is more effective than RFM

P8: Logistic Regression is more effective than CHAID

P9: Logistic Regression is more effective than NN

P10: NN is more effective than RFM

P11: Neural Networks is more effective than CHAID

P12: NN is more effective than Logistic Regression

Given the availability of data in two timeframes for RFM variables, one for activity one month prior to the promotion and the second for activity 12 months prior, I test proposition P0 to establish the adequate RFM baseline model to be used for comparative purposes throughout the rest of the research. Proposition P0 reads as follows:

P0: 12-month period variables are more effective than one-month period variables when applying the RFM technique

This proposition is structured in favour of the 12- month period because of an expected lower volatility due to the aggregation of more data and the well-established relationship between long term profitability and future behaviour. Using one month's worth of data versus 12 months inherently implies a smaller number of transactions per customer and allows for abnormal transactional patterns to be much more significant in ranking extremes than in the case of 12 months' worth of data. In 12-month data sets, monthly extremes can be offset by activity in subsequent months and thus customers can be more adequately ranked. It is also possible that 12-month activity might also over-represent customer activity. For example, a customer that started frequenting a competing commerce during his last month might still be highly scored. It is for this reason that I must also refer to studies showcasing the relationship between long-term profitability and future behaviour. In fact, numerous conceptual studies (Sheth and Parvatiyar, 1995; Morgan and Hunt, 1994; Bendapudi and Berry, 1997) and empirical studies support the benefits of extended customer relationships to profitability (Reichheld and Teal, 1996).

However, it is important to note that, regardless of the outcome for proposition P0, only the RFM technique application will be restricted to the winning timeframe outcome. Other techniques, because of their inherent capacity to accommodate extended data variables, will incorporate all variables under both timeframes.

5.3 Statistical Modelling and Output Generation

As detailed in the previous section, I conducted the study design. However, because of the extreme size of the dataset (individual targeted customers X sum of cumulative transactions X categories by transaction) and the advanced set of techniques applied in the context of the analysis, statistical modelling was conducted in collaboration with a private third-party firm, Precision Consulting from New York, NY. It is important to note that modelling was conducted and outputs provided to the specifications I had specifically set forth and all key decisions made during the conduct of the analysis were made by me, sometimes in consultation with the third-party firm. Said otherwise, based on the KDD process followed in the course of this research, I personally conducted the understanding and selection phases and provided specific instructions on the pre-processing and pattern discovery and modelling phases. Pre-processing instructions included the application of listwise deletion for customer records with missing or incomplete data and the rescaling of recency from a date to the number of days since the last transaction (when required). Details of each party's involvement are listed in Table 49 below in the KDD process followed for this research. Pattern discovery and modelling instructions included the type of validation to conduct on the dataset, the measures used to assess cross-validation success, the required data variables to apply to each study iteration, the measures of fit and performance to use, the structuring of the data into deciles, and the provision of data allowing for the creation of gains and lift charts. All the interpretation of data was conducted by me.

Table 49: Roles of Third Party

Phase	Definition	Lead
1: Understanding	Understanding of the application domain and prior knowledge and goals of end-users	Researcher (me)
2: Selection	Creating a target data set or focusing on a subset of variables, on which discovery is performed	Researcher
3: Pre-processing	Data cleaning and preprocessing in order to obtain consistent and significant data and/or variables. This phase includes outlier detection and scaling, encoding and selection features	Precision Consulting conducted this step with researcher validation on listwise deletions, rescaling and/or recoding of variables
4: Transformation	Transforming data using dimensionality reduction or transformation methods	Given the transformation is an inherent component of applied techniques; it was conducted by Precision Consulting.
5: Pattern Discovery & Modelling	Search for patterns of interest linking the data to the discovery objectives;	Precision Consulting conducted this step and provided the researcher with requested model output
6: Interpretation / Evaluation	Interpretation and evaluation of identified patterns to assess actionnability	Researcher
7: Incorporation	Knowledge incorporation into the system via decisions or actions	Not applicable

5.4 Empirical Results and Analysis

This section provides the results of the empirical analysis. It starts by confirming that the sampling for cross-validation was indeed random (given that any issues at this stage would indicate an erroneous sample and the need to restart the randomised cross-validation exercise). This is followed by a detailed review of lifts at a decile level for each technique. To understand how techniques perform from fit perspective, the Gini coefficient and PCC section provides a review of overall technique performance. Finally, to allow readers to understand the potential sources of higher performance within the technique, the parameter assessment section identifies which variables provide the greatest predictive power for each technique.

5.4.1 Reliability of the cross-validation folds

To ensure sampling for cross-validation was random, Pearson's Chi-Square tests were used to assess the independence of the 10 sub-samples or folds issued from the ten-fold cross-validation. Table 50 illustrates the Pearson Chi-Square results and shows that folds are independent and show no bias. Actual and expected counts in the folds are fairly close to one another and, more importantly, the significance value of 0.175 is commonly interpreted as justification for accepting the null hypothesis that the row variable only randomly relates to the column variable (Han and Kamber, 2006). Said otherwise, the Pearson Chi-Square tests confirm that the test and training sets are only randomly related. These test and training samples are, therefore, used across all studies and statistical and machine-learning applications in this thesis.

Table 50: Validation Test Results

Validation Group * PROMO_RESPONDER Cross tabulation

Count

		PROMO_RESPONDER		
		No Response	Response	Total
Validation Group	.00	2	0	2
	1.00	15503	2035	17538
	2.00	15349	2190	17539
	3.00	15392	2147	17539
	4.00	15403	2136	17539
	5.00	15452	2087	17539
	6.00	15450	2089	17539
	7.00	15454	2085	17539
	8.00	15453	2086	17539
	9.00	15335	2204	17539
	10.00	15388	2151	17539
Total		154181	21210	175391

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	13.967[a]	10	.175
Likelihood Ratio	14.198	10	.164
Linear-by-Linear Association	.768	1	.381
N of Valid Cases	175391		

a. 2 cells (9.1%) have expected count less than 5. The minimum expected count is .24.

5.4.2 Technique performance

5.4.2.1 Assessing Proposition P0

In order to establish a robust and consistent baseline for comparison, Proposition P0 is tested by evaluating the performance of the RFM technique using both one- and 12-month timeframes.

Performance results of the tests are contained in Figures 19, 20 and 21. Figures respectively illustrate RFM technique lift charts, cumulative lift charts and cumulative gain charts. For clarity, RFM-12 indicates the application of the RFM technique on the 12-month timeframe, whereas RFM-1 indicates the application on a one-month timeframe. Results are contrasted with a no-model scenario. As with all other techniques in the research, results are aggregated and averaged from the estimates of all 10 cross-validation folds.

Results from the lift chart in Figure 19 show that RFM-12 has a greater discriminatory power for the top two deciles with a lift index of nearly 200 for RFM-12 at decile 1 versus 186 for RFM-1.

Lift indices for deciles three and four slightly favour RFM-1 before becoming fairly similar for deciles five to seven. The significant difference in lift between RFM-12 and RFM-1 shown at top deciles reverses its trend for decile eight, where RFM-1 captures a greater proportion of responders before trading places in deciles nine and 10.

Figure 20 illustrates the aggregate performance of deciles by showing the performance of deciles in a cumulative fashion. The cumulative lift index by decile indicates that the significant lifts of RFM-12 in the top two deciles allows the technique to surpass RFM-1 (in terms of cumulative response differential) until deciles eight to 10, where both techniques become fairly similar in terms of cumulative lift indices. This is also shown in the cumulative gains chart in Figure 21. The figure presents the results in a different fashion and shows the percentage of responders captured in a cumulative fashion from deciles one to 10. As in the cumulative lift chart, the relative strength of RFM-12's deciles one and two response rate differential versus the baseline allows for the cumulative gains to be maintained until decile eight. Said otherwise, if the targeting issued from RFM-12 was used, the total response rate of the marketing campaign would be higher than RFM-1 until 80 percent of the top performers are targeted. According to Cui et al. (2008), the performance of RFM-12 at top deciles should make it the preferred model and selected baseline for the comparison in this thesis.

However, a measure of predictive fit must also be applied prior to deciding upon the selection of one technique over another. As mentioned earlier, fit is assessed using the Gini coefficient given its strong capacity to measure dispersion. Table 51 contains the Gini coefficients of both RFM-1 and RFM-12. The slightly stronger coefficient of RFM-12 versus RFM-1 indicates that 12-month RFM variables provide more predictive reliability than similar one-month variables.

Table 51: Model Gini Coefficients

Technique & Rank	**Gini Coeff**
1. RFM 12 months	0.403871
2. RFM 1 month	0.384962

The stronger measures of performance and fit showcased by the 12-month variables in the context of the RFM technique provide a confirmation of Proposition P0 that a 12-month timeframe indeed performs better than a one-month timeframe.

Given the confirmation of a stronger performance of the 12-month timeframe, going forward, references to RFM technique outputs will always refer to results issued from the 12-month timeframe. Detailed information on the indices and gains used to generate the aforementioned conclusions can be found in Appendix 3.

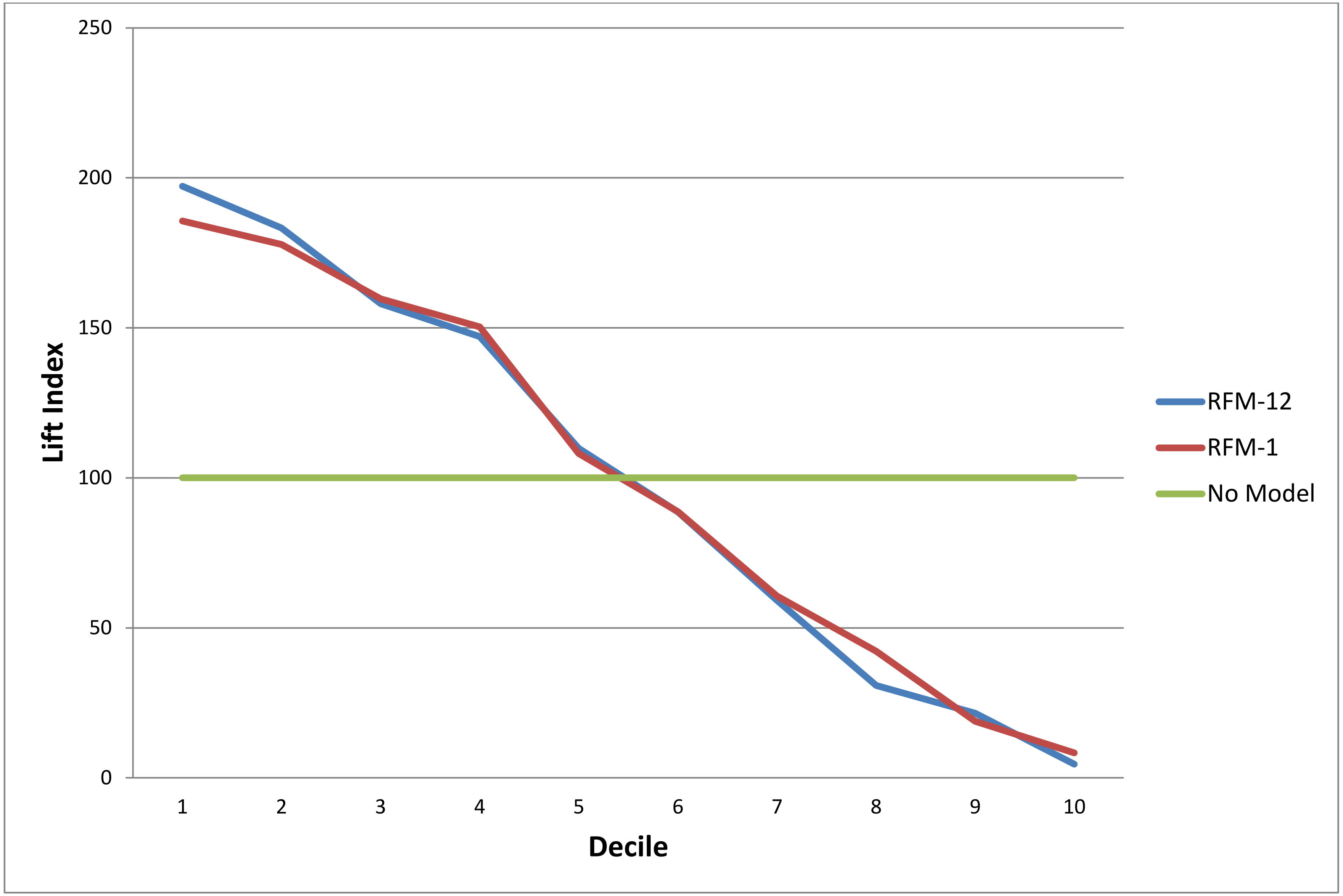


Figure 19: Lift Chart - 1 vs. 12 month RFM Variables

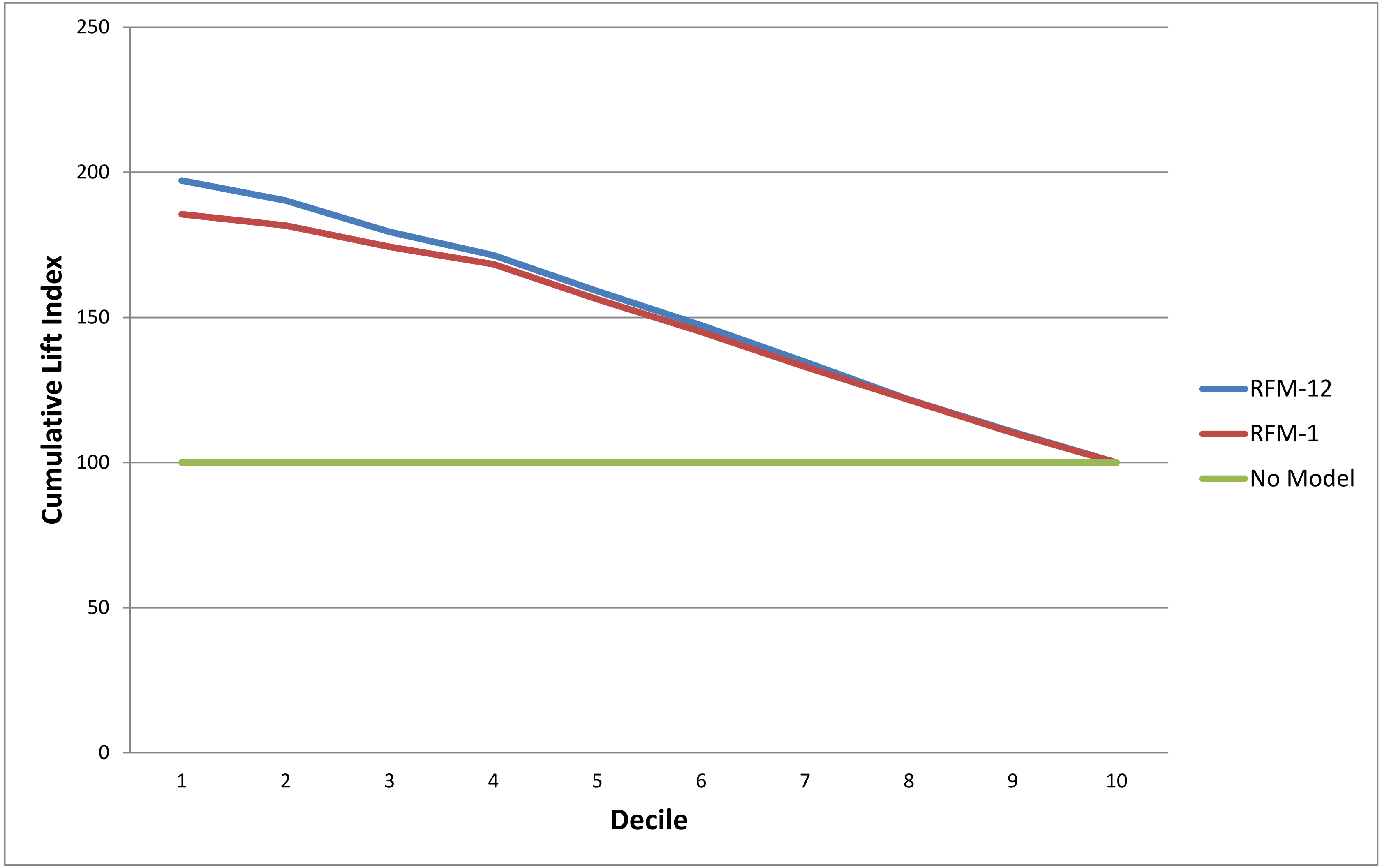


Figure 20: Cumulative Lift Chart – 1 vs. 12 month RFM Variables

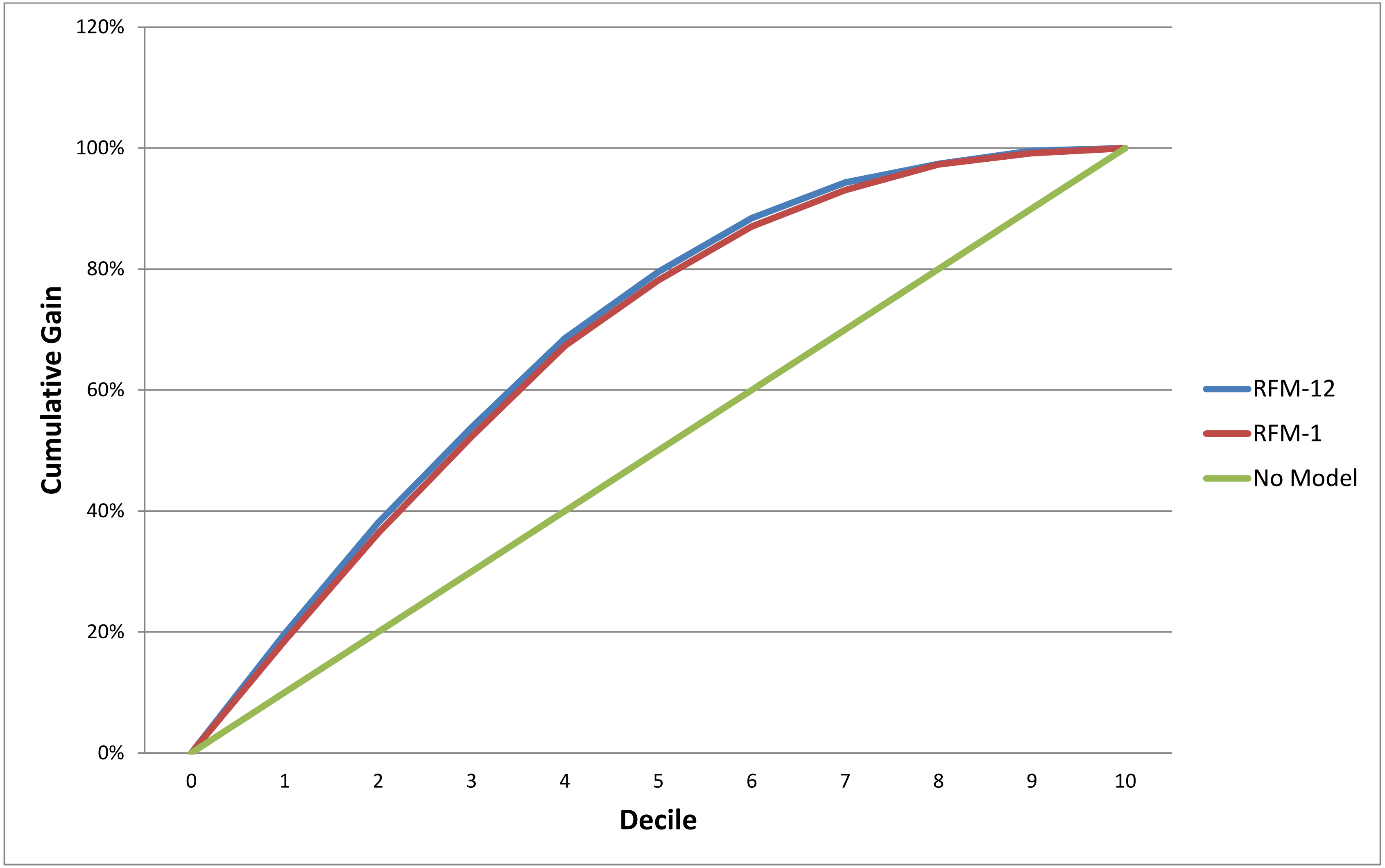


Figure 21: Cumulative Gains Chart - 1 vs. 12 month RFM Variables

5.4.2.2 Assessing Propositions P1 to P12

Following the assessment of the RFM technique timeframe; this section presents the relative performance results of applying advanced techniques (CHAID, ANN, and stepwise logistic regression) against the winning 12-month RFM model (RFM-12). Figures 21 to 23 show the lift and gains by decile of the four different techniques applied using RFM variables only. Results shown are averages of all ten-fold cross validation estimates.

From a top decile performance perspective, the lift chart (Figure 22) visually shows that all techniques seem to provide a fairly similar lift at each decile level versus the average response rate. Given the high number of techniques compared and the proximity of data points, to assess the actual relative performance of techniques I refer to the raw input data of the lift charts (Table 52) to validate this visual observation. From this table, differences do emerge and one can observe that Neural Networks provide the highest increase in lift (response rate of decile divided by average response rate) at the top decile, followed by logistic regression, CHAID and lastly RFM. In observing deciles two through five, with the exception of decile four, the lift provided by the application of logistic regression is consistently highest, while neural networks provide the second highest lift. CHAID and RFM perform very similarly between deciles two and three, before RFM lift overtakes CHAID in deciles four and five. RFM interestingly has the strongest lift versus other techniques at decile four. With an index of 147, RFM provide 47-percent more lift vis-à-vis the average response rate while the next closest technique, logistic regression, only provides 37 percent. Deciles six to 10 are less relevant as techniques with less predictive power naturally outperform in lower deciles.

Table 52: Input for Lift Charts - RFM Variables

Deciles	**RFM**	**CHAID**	**Stepwise LR**	**NN**	**No Model**
Top	197	203	204	206	100
2	183	183	189	187	100
3	158	160	167	164	100
4	147	135	137	136	100
5	110	106	114	110	100
6	89	88	84	86	100
7	59	61	56	57	100
8	31	36	31	33	100
9	22	21	15	17	100
Bottom	5	7	3	3	100

Cumulatively, results observed across lifts and gains charts (Figures 22 and 23) naturally present a different perspective. Cumulative lift and gains charts provide the running tally of how techniques perform as deciles are aggregated from best to worst. This approach is particularly useful when promotional cut-off levels (level at which a marketer chooses to stop targeting customers due to break-even or target-hurdle rates being reached) need to be established.

Table 53: Input for Cumulative Lift Charts - RFM Variables

Deciles	RFM	CHAID	Stepwise LR	NN	No Model
Top	197	203	204	206	100
2	190	193	197	197	100
3	179	182	187	186	100
4	171	170	174	173	100
5	159	157	162	161	100
6	147	146	149	148	100
7	135	134	136	135	100
8	122	121	123	122	100
9	111	110	111	111	100
Bottom	100	100	100	100	100

Table 54: Input for Cumulative Gains Charts - RFM Variables

Deciles	RFM	CHAID	Stepwise LR	NN	No Model
Top	19.72%	20.26%	20.37%	20.60%	10.00%
2	38.04%	38.55%	39.30%	39.34%	20.00%
3	53.84%	54.56%	55.99%	55.70%	30.00%
4	68.55%	68.09%	69.66%	69.28%	40.00%
5	79.53%	78.66%	81.09%	80.30%	50.00%
6	88.39%	87.44%	89.52%	88.92%	60.00%
7	94.31%	93.57%	95.14%	94.65%	70.00%
8	97.39%	97.14%	98.25%	97.92%	80.00%
9	99.54%	99.25%	99.70%	99.65%	90.00%
Bottom	100.00%	100.00%	100.00%	100.00%	100.00%

Looking at the cumulative lift and gain indices provided as input for the construction of cumulative lift charts (Table 53 and Table 54), stepwise logit and neural networks provide a very similar level of performance at top deciles one to five. In fact, both techniques cumulatively outperform both RFM and CHAID across all five top deciles. CHAID for its part, mainly due to the stronger lift in decile one provides a greater cumulative lift for deciles one through three versus RFM. At decile four, due to the strong single decile lift of RFM highlighted earlier, both techniques are practically equal before RFM outpaces CHAID at decile five.

In addition to model performance, model fit must also be examined. The Gini coefficient is used for assessing a response model's fit and consistency (reliability).

5.4.2.3 Gini Coefficient and Percentage Correctly Classified (PCC)

The Gini coefficient evaluates the ability of different models to rank promotional responders correctly. Risselada et al. (2010) refer to this as degree acceptance and use it to compare the output of a random selection of customers with that of model-based selections. As mentioned in the systematic review, other output measures could have been used to assess model fit, including the generally used percentage correctly classified (PCC) measure (among others). The challenge is that many measures are limited to specific methods. As demonstrated in the model summaries and classification outputs of the previous sections, the PCC measure is a central output of neural network and CHAID outputs. In the case of logistic regression, it needs to be calculated separately. In addition, it is not a measure that is possible in the case of the pure RFM method. As a result, I present the Gini coefficient for all techniques and the PCC for logistic regression, CHAID and ANN.

The results from the calculation of the average Gini coefficients of techniques (averaged from Gini coefficients from all cross-validation folds) are shown in rank order Table 55. The output of the Gini coefficient calculation shows that CHAID provides the highest level of predictive power and fit. RFM ranks just below CHAID, with logistic regression and neural networks performing third and fourth best when RFM variables are used.

Table 55: Average Technique Gini Coefficients and PCC – RFM

Technique & Rank	**Gini Coeff**	**PCC**
1. CHAID	0.411492	87.2%
2. RFM	0.403871	NA
4. LR	0.397962	84.0%
5. ANN	0.392185	84.5%

The table also shows the related PCC scores for all relevant techniques. Had PCC been used as a ranking criterion, CHAID would have remained the strongest performing technique but ANN would have ranked higher than logistic regression. It is important to remember that this can be explained by the somewhat different fit measurement focus of each measure. PCC compares the actual yes/no response behaviour of a customer with the predicted behaviour and measures the divergence. The Gini coefficient for its part reflects how well the two groups are separated from one another.

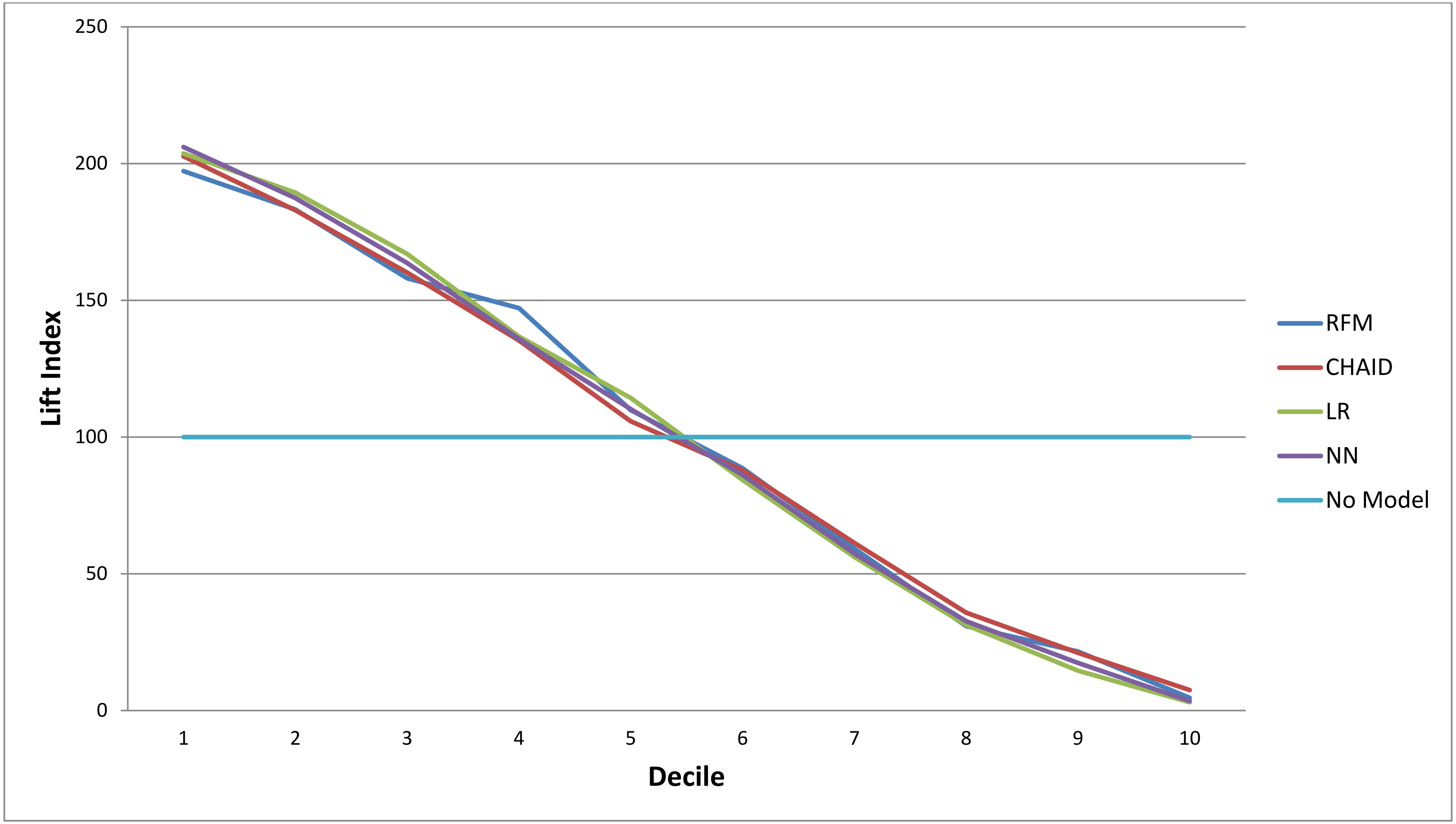


Figure 22: Lift Chart - RFM variables

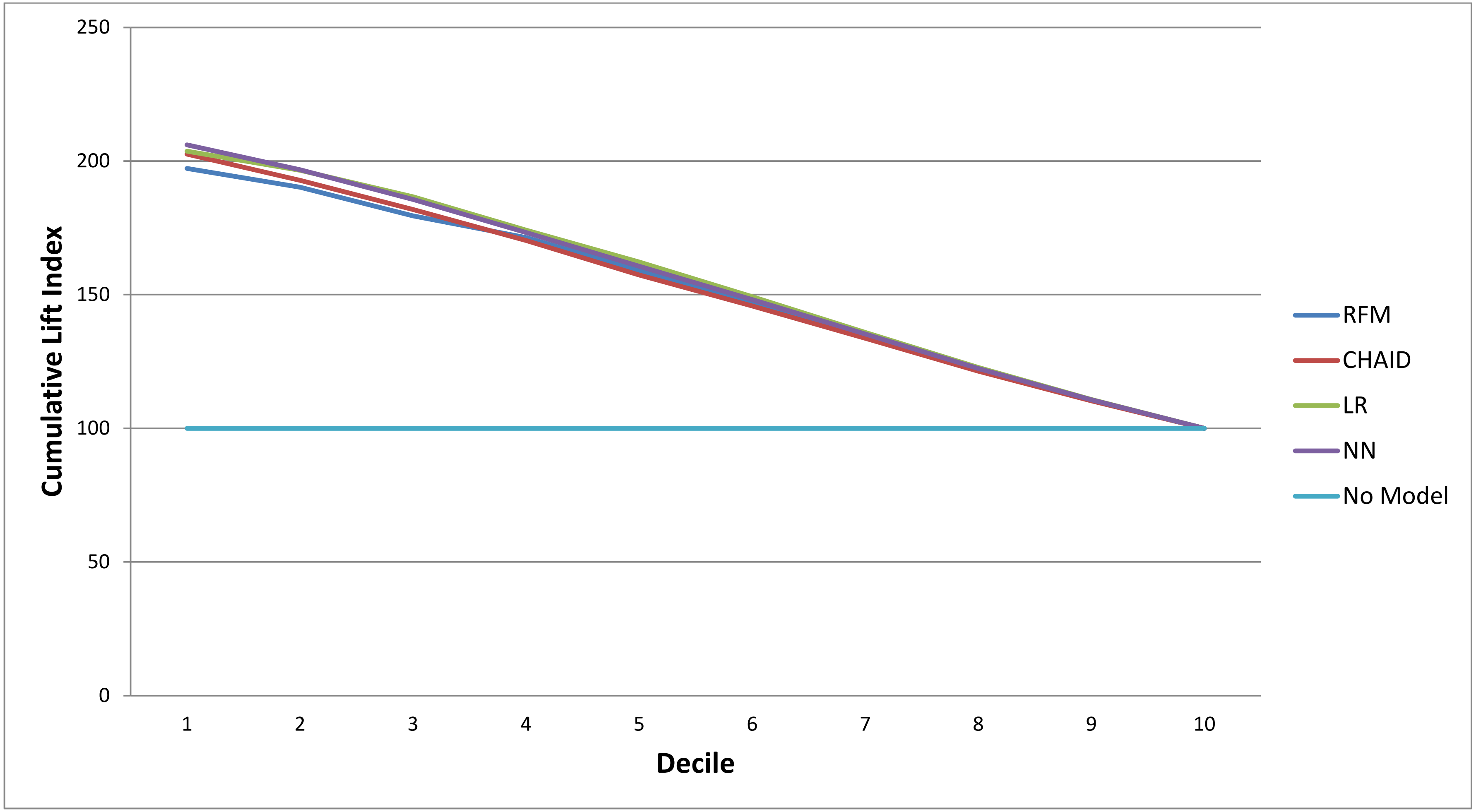


Figure 23: Cumulative Lift Chart - RFM Variables

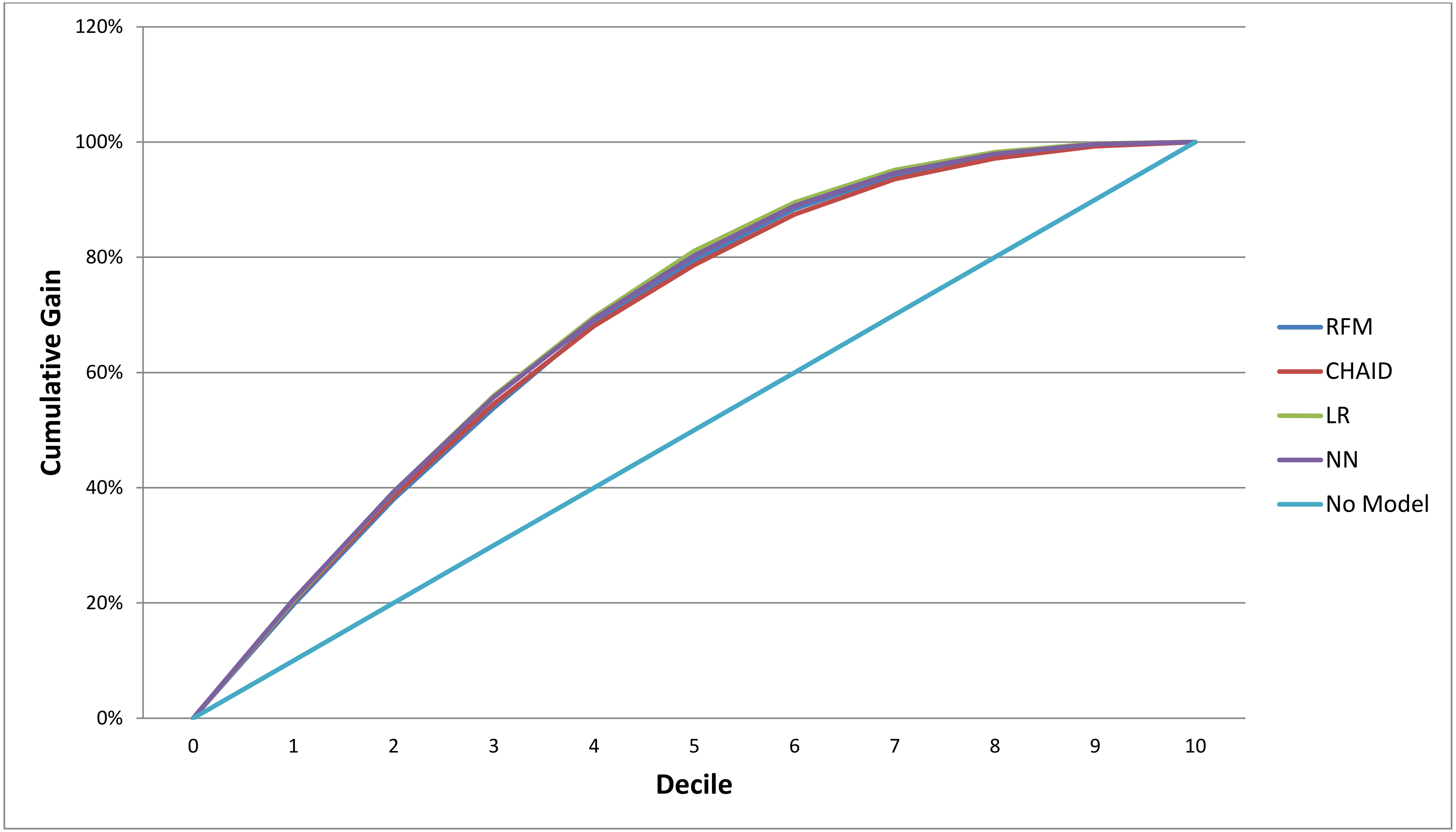


Figure 24: Cumulative Gains Chart - RFM Variables

5.4.3 Evaluation of Techniques outputs

5.4.3.1 Stepwise Logistic Regression

The first relevant information in the stepwise logistic regression output is the validation of whether input variables contribute to the predictive power of the model. This overall assessment is possible (as long as the degree of inter-variable multicollinearity is modest) by observing the residual chi-square statistic for variables that are not included in step 0 of a stepwise regression process. As a reminder, in step 0, only the constant is included in the model. Only in subsequent steps are the additional variables added. As a result, comparing the residual chi-square statistic for these variables provides an overview of whether it is worth moving forward or whether independent variables are only randomly related to the dependent variable. The residual chi-square statistic for these variables is 8029.397, which is significant at $p < .001$, indicating that the addition of these variables to a predictive model will have significant predictive impact. If the probability of the residual chi-square had been higher then .05, then none of the variables in step 0 (all the input variables) would have made a significant contribution to model prediction (Table 56).

Table 56: Step 0 Overall Statistics

Variables not in the Equation		
Fold	**Score**	**Sig.**
1	8029.397	.000
2	8040.553	.000
3	8304.768	.000
4	8016.676	.000
5	8065.552	.000
6	8028.637	.000
7	8081.638	.000
8	8095.147	.000
9	8017.658	.000
10	7935.591	.000

The next step consists of observing how many steps are conducted to achieve an optimal outcome and examine the variables that emerge as significant. The outputs from all ten folds are represented in Table 57, with a summary contained at its end.

Table 57: Variables in the Equation

Variables in the Equation

			B	S.E.	Wald	df	Sig.	Exp(B)
Fold 1	Step 3	FREQUENCY_full_Year	-.001	.000	78.190	1	.000	.999
		MONETARY_full_year	.000	.000	2829.851	1	.000	1.000
		MONETARY_prepromo	-.001	.000	155.684	1	.000	.999
		Constant	-2.533	.015	26822.603	1	.000	.079
Fold 2	Step 3	FREQUENCY_full_Year	-.001	.000	79.513	1	.000	.999
		MONETARY_full_year	.000	.000	2781.785	1	.000	1.000
		MONETARY_prepromo	-.001	.000	129.556	1	.000	.999
		Constant	-2.547	.016	26945.081	1	.000	.078
Fold 3	Step 3	FREQUENCY_full_Year	-.001	.000	79.227	1	.000	.999
		MONETARY_full_year	.000	.000	2860.932	1	.000	1.000
		MONETARY_prepromo	-.001	.000	148.687	1	.000	.999
		Constant	-2.554	.016	26891.442	1	.000	.078
Fold 4	Step 3	FREQUENCY_full_Year	-.001	.000	75.585	1	.000	.999
		MONETARY_full_year	.000	.000	2794.536	1	.000	1.000
		MONETARY_prepromo	-.001	.000	145.482	1	.000	.999
		Constant	-2.543	.016	26900.360	1	.000	.079
Fold 5	Step 3	FREQUENCY_full_Year	-.001	.000	77.475	1	.000	.999
		MONETARY_full_year	.000	.000	2798.005	1	.000	1.000
		MONETARY_prepromo	-.001	.000	134.843	1	.000	.999
		Constant	-2.541	.015	26889.116	1	.000	.079
Fold 6	Step 3	FREQUENCY_full_Year	-.001	.000	77.724	1	.000	.999
		MONETARY_full_year	.000	.000	2769.674	1	.000	1.000
		MONETARY_prepromo	-.001	.000	134.993	1	.000	.999
		Constant	-2.536	.015	26872.588	1	.000	.079
Fold 7	Step 3	FREQUENCY_full_Year	-.001	.000	77.399	1	.000	.999
		MONETARY_full_year	.000	.000	2819.811	1	.000	1.000
		MONETARY_prepromo	-.001	.000	143.481	1	.000	.999
		Constant	-2.540	.016	26853.766	1	.000	.079
Fold 8	Step 3	FREQUENCY_full_Year	-.001	.000	75.376	1	.000	.999
		MONETARY_full_year	.000	.000	2812.420	1	.000	1.000
		MONETARY_prepromo	-.001	.000	147.985	1	.000	.999
		Constant	-2.541	.015	26918.440	1	.000	.079
Fold 9	Step 3	FREQUENCY_full_Year	-.001	.000	71.176	1	.000	.999
		MONETARY_full_year	.000	.000	2774.236	1	.000	1.000
		MONETARY_prepromo	-.001	.000	143.932	1	.000	.999
		Constant	-2.550	.016	26944.145	1	.000	.078
Fold 10	Step 3	FREQUENCY_full_Year	-.001	.000	82.667	1	.000	.999
		MONETARY_full_year	.000	.000	2756.859	1	.000	1.000
		MONETARY_prepromo	-.001	.000	139.570	1	.000	.999
		Constant	-2.534	.015	26786.763	1	.000	.079
Summary								
FREQUENCY_full_Year		min	-.001	.000	71.176	1	.000	.999
		max	-.001	.000	82.667	1	.000	.999
MONETARY_full_year		min	.000	.000	2756.859	1	.000	1.000
		max	.000	.000	2860.932	1	.000	1.000
MONETARY_prepromo		min	-.001	.000	129.556	1	.000	.999
		max	-.001	.000	155.684	1	.000	.999
Constant		min	-2.554	.015	26786.763	1	.000	.078
		max	-2.533	.016	26945.081	1	.000	.079

This table contains similar information to the coefficient table in standard OLS (Ordinary Least Square) regression output. For example, the logistic coefficient for the constant is analogous to the intercept in a standard regression while the Wald statistic serves a similar function to the chi-square statistic (testing the significance of variables). As such, the Wald statistic tests the significance of individual logistic regression coefficients (Menard, 1995), with greater Wald statistics indicating a greater significance of independent variables. Finally, Exp(B) explains the change in the odds ratio that can be attributed to a specific variable. The output from the table illustrates that, in all fold iterations, there were three steps applied by the stepwise regression in order to achieve the optimal outcome. In all three steps, variables that significantly contributed to the prediction power of the model included the monetary value of expenditures (both for a full year and for the pre-promotional period) and the frequency of transactions during the full year. These variables are determined as being significant because they enhance the model's predictive power; however, the Wald statistic presented for each variable in Table 57 indicates whether the b-coefficient for the predictor variable is significantly different than 0 (as it also has a chi-square distribution).

In the case of all the variables included in the last step of the folds, the significance scores indicate that values are significantly different than 0 at a .05 level and so the variables do significantly predict response. Variables that were excluded from the model because they did not add any additional predictive power included the recency of transaction variables in both timeframes and the frequency of transactions during the pre-promotional period.

The final step consists of observing the outputs that are generated by the multiple steps of stepwise regression. The overall fit of the model is generally evaluated by using the log-likelihood statistic. The log-likelihood reported by SPSS is a modified version of the actual log-likelihood statistic, as its value is actually multiplied by -2 because this transformation has an approximate chi-square distribution and thus allows for the comparison against what we might expect to get by chance. This log-likelihood is often referred to as -2LL; the lower the -2LL, the better fitting the statistical models (Field, 2005).

Table 58 shows these improvements in the incremental improvements from steps one to three. Table 58 also answers the question of how much better each step is at improving prediction. It does so by assessing the chi-square value of each step, subtracting the value of the -2LL statistic of a previous step by that of the evaluated step. The outputs demonstrate a significant chi-square improvement at each subsequent step, with step one illustrating the improvement of prediction versus only the model constant. This value has a chi-square distribution and so a significance score is also provided. The significance score displayed for each step (Table 59) indicates that the value is significant at a .05 level, and so the model does indeed predict response significantly better at each additional model step.

There are no perfect substitutes for the R^2 value derived in linear regression in the context of logistic regression. However, in terms of interpretation, three measures can be used fairly interchangeably as proxies of logistic model significance: Hosmer and Lemeshow R^2, Cox and Snell R Square and Nagelkerke R Square adjusted value (Field, 2005). The measures of explanatory power provided by the SPSS output includes the Cox and Snell R Square and Nagelkerke R Square adjusted value. As a result, they will be used for comparison. The scores of both values indicate that each additional step in the logistic regression contributes to increasing the explanatory power of the model.

Table 58: Model Summary

Step	-2 Log likelihood		Cox & Snell R Square		Nagelkerke R Square	
	min	max	min	max	min	max
1	95974.857	96524.507	.057	.059	.098	.102
2	95786.587	96344.409	.058	.060	.101	.104
3	95704.145	96263.063	.059	.061	.102	.105

Table 59: Omnibus Tests of Model Coefficients

Step	Chi-square		Sig.
	min	max	
1	6991.710	7239.080	.000
2	7165.676	7427.350	.000
3	7251.851	7509.792	.000

5.4.3.2 CHAID

CHAID SPSS outputs consist of the model summary, the tree diagrams, and risk statistics and classification results. The model summary details the decision tree growing method, the dependent variable, the total variables used in the model, the final variables retained, and the number of nodes used to construct the tree. The tree diagrams showcases how different variables are positioned across the tree design. Risk statistics provide the estimated risk (and the related standard error) of misclassification. In the case of a logistic dependent variable, risk is calculated as "the proportion of cases in the sample incorrectly classified by the tree. A table is also displayed indicating the numbers of cases corresponding to specific prediction errors" (SPSS, 1999, p. 68).

To obtain segments large enough for analysis, the minimum size of parent nodes was set to 100 observations and child nodes to 50 observations.

Table 60: CHAID Model Summary

Specifications	Growing Method	CHAID	
	Dependent Variable	PROMO_RESPONDER	
	Independent Variables	RECENCY2_Full_Year, FREQUENCY_full_Year, MONETARY_full_year, RECENCY2_prepromo, FREQUENCY_prepromo, MONETARY_prepromo	
	Validation	Split Sample	
	Maximum Tree Depth	3	
	Minimum Cases in Parent Node	100	
	Minimum Cases in Child Node	50	
Results	Independent Variables Included	MONETARY_full_year, MONETARY_prepromo, FREQUENCY_full_Year, RECENCY2_Full_Year, FREQUENCY_prepromo, RECENCY2_prepromo	
		min	max
	Number of Nodes	80	97
	Number of Terminal Nodes	55	67
	Depth	3	

By observing the fold is one iteration of the decision tree (Appendix 4) as a representative example, there is a significant difference between the predictive power of each key node. The first breakout node for the tree is the yearly monetary value. The response rate by node indicates that customers who spend more in a given year also have a higher response rate. In fact, the response rate varies between 0.2 percent for lower spenders and 29.3 percent for the highest spending group within this layer. From the second layer, a couple of other observations are also possible. Reponse rates are inversely related to the frequency of customer patronage. This may seem counterintuitive; however, higher-frequency customers also typically exhibit higher share of wallet and thus have less need to purchase additional goods during promotions.

It is possible that such customers are loyal to the store regardless of promotions. This is true for nodes five to 10. Nodes four and five indicate an inconclusive relationship. This may be explained by the initially lower monetary expenditure that may change the outcome of the previously described frequency behaviour exhibited by customers.

Subsequent nodes are made up of numerous other variables including a combination of frequency of transactions for the year, in the pre-promo period, monetary value in the pre-promo period, and recency. It is difficult to interpret the outcomes of every single additional layer, given there are numerous factors that can impact response; and generally, beyond general observations, it is difficult to come to any conclusion at a terminal level other than to use this information to generate the appropriate ranking for group prioritisation in gains and lift charts. Nevertheless, the discrimination power of CHAID is shown by the wide gap between the highest responding terminal node and the lowest. The highest terminal node's response rate is 35.5 percent while the lowest is 0.0 percent.

Looking at the risk statistics and classification results showcases how well the model classifies predicted cases. The risk estimate of between .128 and .139 for the test sample (Table 61) indicates that the category predicted by the model (response) is wrong for 12.8 percent to 13.9 percent of the cases. This is the risk of misclassifying customers. Results in the classification table (Table 62) are naturally consistent with the risk estimate with data indicating that the model classifies approximately between 86.1 percent and 87.2 percent of customers correctly. This table showcases that the overall percentage of cases that are correctly classified by the CHAID model is relatively high both for the test and training set.

Table 61: Risk Statistics

Sample	**Estimate**		**Std. Error**	
	min	max	min	max
Training	.133	.134	.001	.001
Test	.128	.139	.003	.003

Growing Method: CHAID

Dependent Variable: PROMO_RESPONDER

Table 62: Classification

	Overall Percentage Correctly Classified	
	min	max
Training	86.6%	86.7%
Test	86.1%	87.2%

Growing Method: CHAID

Dependent Variable: PROMO_RESPONDER

5.4.3.3 Neural Networks

As illustrated in Table 63, the structure of the neural network was composed of six units in the input layer (all the RFM variables in both timeframes) and one hidden layer comprised of between two to seven 'hidden' units in the hidden layer (depending on the fold iteration used). The one hidden layer is typical of multi-layer perceptron applications while the hidden layers are a function of the model's variable interconnections. As is often the case with neural networks, the model obtained is not understandable in terms of physical parameters given its inherently black-box design. This black-box constraint is echoed by Ha et al. (2005, pp. 29-30) as they describe neural networks: "the input variables are combined in a complicated, nonlinear way to produce an output...Those marketers who would like to understand how individual predictor variables influence the target and how they interact might be baffled by the neural network model's inability to provide any insight in that regard." As a result, no significant amount of insight on parameter assessment can be derived from the neural network output. This difficulty is circumvented by using alternative approaches with easily interpretable results; i.e., decision tree or regression techniques (Ho et al., 2008). In this specific case, there is no need to apply alternative approaches given that the required techniques were already included in the analysis. Given that all techniques were applied independently, findings remain just as valid as if they had been applied in a two-step fashion. As such, I refer the reader to the previous section on logistic regression and CHAID decision tree for assessment of different variable estimates signs, sizes, and significance.

Table 63: ANN MLP Network Information

Network Information				
Input Layer	Covariates	1		RECENCY2_Full_Year
		2		FREQUENCY_full_Year
		3		MONETARY_full_year
		4		RECENCY2_prepromo
		5		FREQUENCY_prepromo
		6		MONETARY_prepromo
		Number of Units[a]		6
		Rescaling Method for Covariates		Standardised
Hidden Layer(s)		Number of Hidden Layers		1
		Number of Units in Hidden Layer 1[a]	min	2
			max	7
		Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1		PROMO_RESPONDER
	Number of Units			2
	Activation Function			Softmax
	Error Function			Cross-entropy

Table 64: ANN MLP Model Summary

Model Summary		min	max
Training	Cross Entropy Error	13538	13966.831
	Percent Incorrect Predictions	15.3%	15.9%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error	
	Training Time	00:00:05.811	
Testing	Cross Entropy Error	31654	31930.965
	Percent Incorrect Predictions	15.4%	15.6%
Holdout	Percent Incorrect Predictions	14.8%	16.1%

Dependent Variable: PROMO_RESPONDER

a. Error computations are based on the testing sample.

However, it is possible to observe the error and misclassification (1-PCC) statistics of the neural network output. Looking at the cross-entropy error, though results in Table 64 may look high in this output, they are not necessarily indicative of performance, given they are not compared to any other output. The PCC score, for its part, does indicate a similar classification/ misclassification rate versus previous techniques.

5.5 Discussion

This first quantitative study sheds some light on the relative performance of techniques sourced from the systematic review using the simple data variables of recency, frequency and monetary value. As mentioned earlier, the thesis aims to address two gaps: (1) the relationship between customer selection techniques and performance and (2) the impact of dimensionality reduction on performance. The second gap de facto examines the often implicit impact of dimensionality reduction on performance and will shed some light on the predictive strength of specific variables and variable types. This section will discuss findings as they relate to these gaps in order, with the first portion covering technique performance, while the second will address variable set and dimensionality reduction.

5.5.1 Overview of Technique Performance

From a technique performance perspective, the study provides a number of interesting theoretical and empirical validations, namely that a 12-month timeframe outperforms a one-month timeframe for the RFM application, and that advanced technique provides a slightly better performance than the simple RFM technique application.

The finding that a 12-month variable timeframe generally outperforms a one-month timeframe validates the works of Sheth and Parvatiyar (1995), Morgan and Hunt (1994), Bendapudi and Berry (1997) and Reichheld and Teal (1996). The stronger response rates exhibited by top deciles validate the relationship between long-term customer relationships and future behaviour (versus shorter term relationships). In addition, the stronger Gini coefficient demonstrated by the 12-month dataset validates that models with longer time series have a tendency to be more stable and thus show a lower rate of dispersion versus shorter term datasets. In other words, this indicates that using shorter periods for modelling RFM response is riskier from a model reliability perspective than using longer periods. Though previous research supports this finding, it is also indirectly supported by research by the Food Marketing Institute that indicates that "primary stores lost share of weekly spend … with store trips going to other types of stores and formats such as warehouse clubs, drugstores, dollar stores, ethnic foods stores and convenience stores" (FMI, 2012, p. 13). Primary stores are defined as the store of choice of a particular survey responder; as a result, loss of weekly spend to other channels indicates that the short-term volatility of purchases is indeed increasing. The higher Gini coefficient of RFM-12 seems to indicate that (from a modelling perspective) this can be partly offset by selecting longer reference periods. As a result, Proposition P0 can be answered as follows:

RFM variables using 12 months of historical data provide greater RFM technique fit and performance

Moving to the comparison of performance between techniques, a detailed review of technique effectiveness accompanied by the validation/invalidation of propositions is presented in Table 65.

Table 65: Technique Effectiveness by Proposition for RFM Variable Set

Proposition	Validation	Description
P0: RFM 12 month variables are more effective than 1 month when applying the RFM technique	Validated	• RFM variables using 12 months of historical data provide greater RFM technique fit and performance.
P1: RFM is more effective than CHAID	Invalidated	• CHAID performs better than RFM at an individual level for decile 1. Deciles 2 and 3 are fairly similar while RFM lift is better for deciles 4 and 5. Lower deciles are generally similar in terms of lift with some small variance in decile 7 and 8. • Cumulatively, the lift in decile 1 allows CHAID to have a cumulative lift/gain greater than RFM until deciles 3. As of decile 4, cumulative lift is similar. • The Gini coefficient of CHAID is slightly greater than RFM, illustrating a slightly better performance at top deciles but a slightly better fit overall.

Proposition	Validation	Description
P2: RFM is more effective than Stepwise Logistic Regression	Invalidated	• Logistic Regression performs better than RFM at an individual level for deciles 1, 2 and 3. RFM lift is better for deciles 4, while Logistic Regression performs better at decile 5. Lower deciles are a mixed lot with RFM showing a greater lift for decile 6, 7, 9 and 10 and Logit doing better for decile 8. • Cumulatively, the lift in deciles 1 to 3 allows Logistic Regression to have a cumulative lift/gain greater than RFM until deciles 6. As of decile 7, cumulative lift is similar with only slight differences. • The Gini coefficient of RFM is slightly greater than Logistic Regression, illustrating disconnect between performance and fit.
P3: RFM is more effective than NN	Invalidated	• NN performs better than RFM at an individual level for deciles 1, 2 and 3. RFM lift is better for deciles 4, while NN performs better at decile 5. Lower deciles are again a mixed lot with RFM showing a greater lift for decile 6, 7, 9 and 10 and NN performing better for decile 8. • Cumulatively, the lift in deciles 1 to 3 allows NN to have a cumulative lift/gain greater than RFM until decile 5. As of decile 6, cumulative list is generally similar with only slight differences. • The Gini coefficient of RFM is slightly greater than NN, illustrating disconnect between performance and fit.
P4: CHAID is more effective than RFM	Validated	• CHAID performs better than RFM at an individual level for decile 1. Deciles 2 and 3 are fairly similar while RFM lift is better for deciles 4 and 5. Lower deciles are generally similar in terms of lift with some small variance in decile 7 and 8. • Cumulatively, the lift in decile 1 allows CHAID to have a cumulative lift/gain greater than RFM until deciles 3. As of decile 4, cumulative list is generally similar. • The Gini coefficient of CHAID is slightly greater than RFM, illustrating a slightly better performance at top deciles but a slightly better fit overall.
P5: CHAID is more effective than Stepwise Logistic Regression	Validated	• Stepwise logit performs better than CHAID for all five top deciles. • Cumulatively, the lift in top deciles allows stepwise logit to have a cumulative lift/gain greater than CHAID until decile 9 • The Gini coefficient of CHAID is slightly greater than stepwise logit, illustrating a better overall relative fit.
P6: CHAID is more effective than NN	Invalidated	• ANN performs better than CHAID for all five top deciles. • Cumulatively, the lift in top deciles allows ANN to have a cumulative lift/gain greater than CHAID until decile 9. • The Gini coefficient of CHAID is greater than ANN, illustrating a better overall relative fit.

Proposition	Validation	Description
P7: Stepwise Logistic Regression is more effective than RFM	Validated	• Logistic Regression performs better than RFM at an individual level for deciles 1, 2 and 3. RFM lift is better for deciles 4, while Logistic Regression performs better at decile 5. Lower deciles are a mixed lot with RFM showing a greater lift for decile 6, 7, 9 and 10 and Logistic Regression doing better for decile 8. • Cumulatively, the lift in deciles 1 to 3 allows Logistic Regression to have a cumulative lift/gain greater than RFM until deciles 6. As of decile 7, cumulative list is generally similar with only slight differences. • The Gini coefficient of RFM is slightly greater than Logistic Regression, illustrating disconnect between performance and fit.
P8: Stepwise Logistic Regression is more effective than CHAID	Invalidated	• Stepwise logit performs better than CHAID for all five top deciles. • Cumulatively, the lift in top deciles allows stepwise logit to have a cumulative lift/gain greater than CHAID until decile 9 • The Gini coefficient of CHAID is slightly greater than stepwise logit, illustrating a better overall relative fit.
P9: Stepwise Logistic Regression is more effective than NN	Partly validated	• ANN performs better than stepwise logit for decile 1, while stepwise logit performs better for deciles 2 to 5. • Cumulatively, in line with the lift chart, only decile 1 allows ANN to perform better than stepwise logit. For all subsequent deciles, performance is generally aligned. • The Gini coefficient of stepwise logit is greater than ANN.
P0: NN is more effective than RFM	Validated	• NN performs better than RFM at an individual level for deciles 1, 2 and 3. RFM lift is better for deciles 4, while NN performs better at decile 5. Lower deciles are again a mixed lot with RFM showing a greater lift for decile 6, 7, 9 and 10 and NN performing better for decile 8. • Cumulatively, the lift in deciles 1 to 3 allows NN to have a cumulative lift/gain greater than RFM until decile 5. As of decile 6, cumulative list is generally similar with only slight differences. • The Gini coefficient of RFM is slightly greater than NN, illustrating disconnect between performance and fit.
P11: NN is more effective than CHAID	Validated	• CHAID performs better than NN models at deciles one, two and three before performance degrades to meet similar performance levels for deciles four to ten.
P12: NN is more effective than Stepwise Logistic Regression	Partly validated	• ANN performs better than stepwise logit for decile 1, while stepwise logit performs better for deciles 2 to 5. • Cumulatively, in line with the lift chart, only decile 1 allows ANN to perform better than stepwise logit. For all subsequent deciles, performance is generally aligned. • The Gini coefficient of stepwise logit is greater than ANN.

Looking at the comparative technique performance (hypotheses one to 12), the generally stronger performance of the advanced statistical and machine learning technique (for top deciles one to three) vis-à-vis the classical RFM technique supports the observations made in the systematic review that advanced techniques (specifically CHAID, ANN and logistic regression) outperform RFM. More specifically, this supports findings from empirical research by Magidson (1988), Levin (2001), Yang (2004) and McCarty and Hastak (2006). This is also very much in line with a recent study from Olson and Chae (2012) where RFM variable-based predictive models were similarly built and compared to logistic regression, decision trees, and neural networks. As in Olson and Chae's study (2012, p. 75), the "balancing expected cell densities and compressing RFM variables into a value function" issued from the CHAID, logistic regression and ANN models provided greater cumulative gains especially at cut-off level deciles.

Beyond the RFM comparison, from an advanced technique-specific perspective, both logistic regression and ANN outperform CHAID for all top deciles both at an individual and cumulative lift level. Albeit from different contexts, for ANN, this validates findings from West et al. (1997) and Linder et al. (2004). The validation of West et al. (1997) and Linder et al.'s (2004) studies must be made with the appropriate limitation that these studies included more input variables sourced from surveys or simulated data that could limit the generalisation of findings to a simple RFM variable set (such as the one used in this study). No studies from the examined literature illustrated CHAID outperforming ANN. As it pertains to logistic regression, results also validate the findings of Linder et al. (2004), as well as Levin and Zahavi (2001), but do not support McCarty and Hastak's (2004) findings. McCarty and Hastak's results indicated that CHAID produced better results that logistic regression. The different ranking with McCarty and Hastak's (2006) work may be related to the significantly different contexts of both studies. In studies by Levin and Zahavi (2001), Shih and Liu (2003), Yang (2004) and in McCarty and Hastak's (2006) first of two case studies, I posit that the response rate of the targeted industries (collectibles, hardware retail, automotive) are much lower than McCarty and Hastak's (2006) second case study (non-profit), which generated response rates between 20 percent and 50 percent. Therefore, empirical findings from this study of RFM performance in FMCG retail are consistent with the ranking obtained in similar contexts and support McCarty and Hastak's (2006) view that RFM may be less effective in low response rate contexts. Given that only 30 percent of the total customer database was mailed, this also validates McCarty and Hastak's (2006) finding that RFM performance is diminished when targeting is restricted to relatively small portions of an entire file (30 percent or less).

However, unlike Olson and Chae's (2012) findings, the predictive accuracy and fit of advanced models were not all better than RFM. In fact, the RFM application was second only to CHAID in terms of its Gini coefficient. For its part, CHAID showed a better capacity to predict outcomes

(PCC) and discriminate between groups (Gini) versus all other techniques. This indicates that in cases where limited RFM variables are used for predictive modelling (at least in the context of FMCG retail), the inherently assumed accuracy of advanced models is not a certainty. In fact, this study demonstrates that the risk of misalignment between a model's performance and fit, often highlighted by academics, exists in this study. This further adds credibility to the findings of Leahy (1992), Roberts and Berger (1993), Magliozzi and Berger (1993), and Malthouse (2002) that do indicate that in many cases models that don't fit particularly well still may perform well. From a practitioner perspective, given performance measures are what marketers are most interested in oftentimes at the expense of predictive accuracy (Jonker et al., 2004; Kim et al., 2005), such models may nevertheless still be adopted for targeting purposes even in light of their lower predictive accuracy. Such an action, though potentially financially lucrative in the short term, presents a higher risk of instability given a model that yields a strong performance but a poor fit may not be stable through time or iterations. Said otherwise, the performance advantage of lower Gini techniques may indeed disintegrate, as higher risk makes the model outputs less certain.

5.5.2 Data Variables and Dimensionality Reduction

Across most techniques, all RFM variables for both one- and 12-month timeframes seem to significantly contribute to increasing model strength with the exception of logistic regression where the recency variable is not always a significant predictor.

In the case of the logistic regression application, the recency variable (in both one- and 12-month timeframes) did not emerge as being significant. The frequency of transactions during the pre-promotional period was also excluded from the logit variable predictors. However, these variables (under all timeframes) were consistently applied in both ANN and CHAID applications. The exclusion of RFM variables contrasted with all studies from the systematic review.

Nevertheless, it is possible that a group bias exists given that the list extracted for the promotion only targeted the top 30 percent of retail customers. This targeting may have led to selecting a group of customers with a high and relatively consistent recency of transactions, thus making the variable potentially less significant in the logistic application. Concerning the exclusion of the one-month frequency variable, this may be explained by the inherently higher volatility of shorter period transactional behaviours in non-contractual settings where more than 70 percent of all customers shop in several supermarkets in any given month (Buckinx and Van den Poel, 2005). This is in sharp contrast with the significance of monetary expenditures for both timeframes. This incongruence between frequency and monetary value variables indicates a potential need for retailers to not only understand the short-term customer value but also their behavioural triggers and changes, as well as if they are to effectively improve responsiveness to promotions.

The exclusion of the recency and frequency variables within the logistic application also demonstrates the benefit (at least in the case of logit) of using a technique that has a built-in dimensionality reduction step. This certainly supports the fact that even in cases of limited RFM data variables, there is a dimensionality reduction benefit in using the stepwise regression approach adopted by Lix et al. (1995) instead of a standard logistic regression approach. The benefits of this integrated approach (that applied the pattern discovery and modelling and the transformation phases together) provides support to the claims of Deichmann et al. (2002), Rao and Ali (2002), Shih and Liu (2003), and Malthouse (2008) that either an integrated application of dimensionality reduction or a two-step dimensionality reduction application has a greater potential to increase performance.

Other interesting findings on variables issued from the CHAID decision tree indicate customers that spend more in a given year also have a higher response rate. Furthermore, the higher expenditure customer segments from this study that also shop less frequently were more likely to respond to the promotion than more regular customers. This may seem counterintuitive; however, high monetary value customers with a higher frequency may, by default, have a higher share of wallet and thus have less need to purchase additional goods during promotions. It is possible that such customers are loyal to the store regardless of promotions. As such, a sub-segment of high spenders that has a lower frequency of visits may indeed be more likely to respond to a promotion because their relative share is lower (and as a result, they are more likely to be able to take better advantage of the promotion). This finding is very much aligned with the work of Lal and Bell (2003) highlighted in the literature review which indicates that, in the context of a US grocery chain’s loyalty program, it was the incremental sales from casual shoppers, also referred to as cherry pickers, that often generated the greatest return for retailers. Much like the conclusion of Lal and Bell, this finding is inconsistent with existing theories on frequent shopper program impacts.

It is thus possible that a specific strategy targeting this sub-segment of customers specifically could be viable from a promotional perspective. It is important to mention that such a segment would indeed be limited to customers that fall within nodes five to 10 for both full-year and pre-promo frequency as other nodes indicate an inconclusive relationship. This may be explained by the initially lower monetary expenditure changing the type of frequency behaviour exhibited by customers.

5.6 Summary

The intent of this second study is to understand technique effectiveness in shallow (RFM variable only) data environments in the context of FMCG retail. Given their prevalence, high performance ratings from the systematic review, and the capacity to integrate dimensionality reduction, techniques tested against one another include: RFM, stepwise logistic regression, CHAID, and neural networks. RFM is used to establish a baseline level of performance against which other techniques are compared. Given the suboptimal use of techniques by practitioners highlighted in the systematic review and the still very prevalent use of RFM as a basis of segmentation and selection, I felt it critical to establish a performance baseline that was restricted to these three variables and the strict RFM technique. Furthermore, in order to provide the analysis with the most complete and impactful outcomes, all techniques used included a dimensionality reduction facet. While inherent to ANN and CHAID, the logistic regression approach selected to enable this was a forward stepwise logistic regression. Data variables used to operationalise techniques include the generally available transactional variables of recency, frequency and monetary expenditures.

The thesis aims to address two gaps: (1) the relationship between customer selection techniques and performance and (2) the impact of dimensionality reduction on performance. In order to address these gaps, the main objectives of this study were the following:

1) Establish the optimal timeframe to be applied to the RFM technique.
2) Evaluate the relative performance and fit of different techniques in predicting response.
3) Determine the impact or lack thereof of different variables on performance and whether dimensionality reduction was required.

Performance was assessed using response rate while comparative technique performance was assessed using lift charts, cumulative lift charts and cumulative gains charts. Fit and reliability measures used consisted of the Gini coefficient of dispersion and the PCC (when applicable). To ensure that sampling was random, the ten-fold cross validation sampling procedure was applied.

The dataset is provided by a large North American FMCG retailer with 30-percent market share and 90-percent market coverage in markets it serves. The promotion from which data is sourced was held in January 2009 and was communicated to consumers via direct mail. The promotion period lasted one month, meaning that customers had one month to take advantage of the offer.

Results of the study were the following:

- In the evaluation of the timeframe for the RFM technique, the stronger measures of performance and fit were exhibited by the 12-month variables. As a result, the RFM technique used as a baseline throughout Study 2 (and 3) is solely based on the 12-month timeframe for all RFM variables. The validation that a 12-month variable timeframe generally outperforms a one-month timeframe further validates the works of Sheth and Parvatiyar (1995), Morgan and Hunt (1994), Bendapudi and Berry (1997) and Reichheld and Teal (1996). Indeed, the stronger response rates exhibited by top deciles validated the relationship between long-term customer relationships and future behaviour.

- Looking at the cumulative lift and gain indices provided as input for the construction of cumulative lift charts, stepwise logit and neural networks provide a very similar level of performance at top deciles one to five. In fact, both techniques cumulatively outperform both RFM and CHAID across all five top deciles. CHAID, for its part, mainly due to the stronger lift in decile one, provides a greater cumulative lift for deciles one through three versus RFM. At decile four, due to the strong single decile lift of RFM highlighted earlier, both techniques perform equally well before RFM outpaces CHAID at decile five.
- The generally stronger performance of the advanced statistical and machine-learning technique (for top deciles one to three) vis-à-vis the classical RFM technique supports the observations made in the systematic review that advanced techniques (specifically CHAID, ANN and logistic regression) outperform RFM. More specifically, this supports findings from empirical research by Magidson (1988), Levin and Zahavi (2001), Yang (2004) and McCarty (2006).
- The output of the Gini coefficient calculation shows that CHAID provides the highest level of predictive power and fit. RFM ranks just below CHAID, with logistic regression and neural networks performing third and fourth best when RFM variables are used. Had PCC been used as a ranking criterion, CHAID would have remained the strongest performing technique but ANN would have ranked higher than logistic regression. It is important to remember that this can be explained by the somewhat different fit measurement focus of each measure. PCC compares the actual yes/no response behaviour of a customer with the predicted behaviour and measures the divergence. The Gini coefficient for its part reflects how well the two groups are separated from one another.
- All techniques, except logistic regression, used all RFM variables under both one-month and 12-month timeframes as significant inputs. Logistic regression output included the monetary value of expenditures (both for a full year and for the pre-promotional period), and the frequency of transactions during the full year. Variables that were excluded from the model because they did not add any additional predictive power included the recency of transaction variables in both timeframes and the frequency of transactions during the pre-promotional period. The exclusion of the recency and frequency variables within the logistic application also demonstrates the benefit (at least in the case of logit) of using a technique that has a built-in dimensionality reduction step. This certainly supports that, even in cases of limited RFM data variables, there is a dimensionality reduction benefit in using the stepwise regression approach. Nevertheless, it is possible that a group bias exists given that the list extracted for the promotion only targeted the top 30 percent of retail customers.
- Other interesting findings on variables issued from the CHAID decision tree indicate that

customers who spend more in a given year also have a higher response rate. Furthermore, the higher-expenditure customer segments from this study that also shop less frequently were more likely to respond to the promotion than more regular customers. This may seem counterintuitive; however, high monetary value customers with a higher frequency may, by default, have a higher share of wallet and thus have less need to purchase additional goods during promotions. This finding is very much aligned with the work of Lal and Bell (2003) that indicates that in the context of a US grocery chain's loyalty program, it was the incremental sales from casual shoppers, also referred to as cherry pickers, that often generated the greatest return for retailers.

5.7 Conclusion

Study 2 incorporated findings from the systematic review into a first quantitative study aimed at understanding technique effectiveness in limited variable (RFM) environments in the context of FMCG retail. Findings illustrated a moderate improvement in performance of advanced techniques and some slight differences on the impact of different variables on performance depending on the techniques utilised. These results provide a strong baseline from which to examine the performance of these same techniques with extended data.

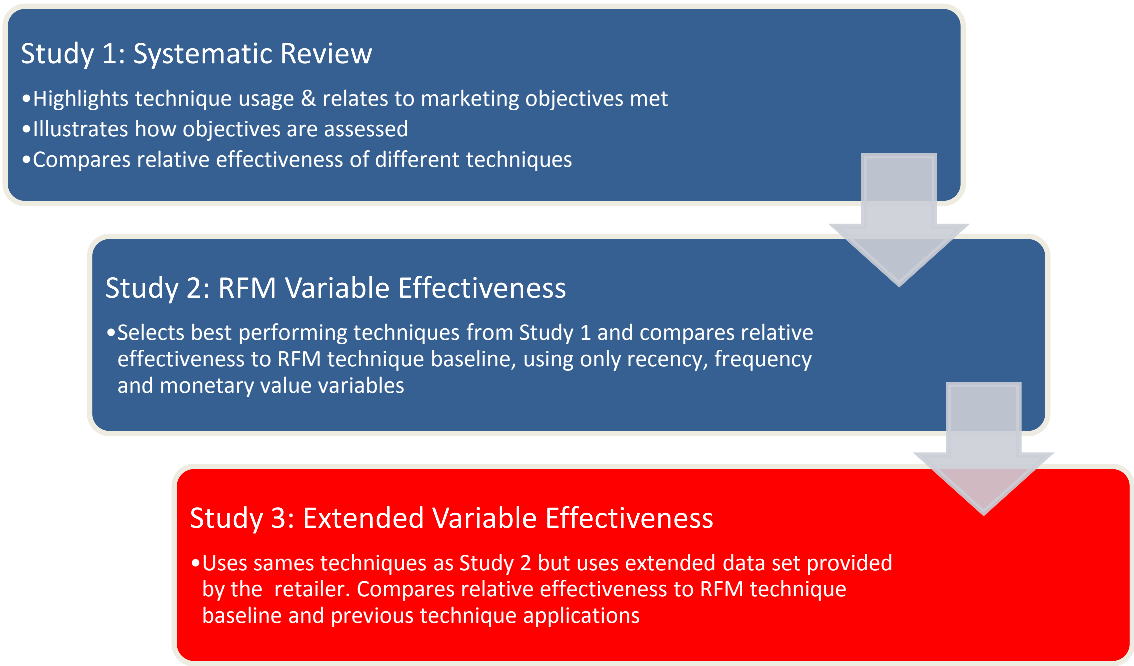
Study 3 repeats the research approach and techniques (RFM, stepwise logistic regression, CHAID, and Neural networks) of the second study but examines the performance impact of using expanded data variables. Expanded variables include the previously used variables of recency, frequency and monetary value, and all other variables provided by the retailer, including numerous detailed transactional, demographic and psychographic variables. The use of an extended data set will not only generate additional insights on the type of data that impacts response but will also illustrate the true power of applying the systematic processes of database marketing and data mining.

6 Study 3: Effectiveness of Data Mining techniques using richer data

6.1 Introduction

Given that the methodology of this study is very similar to that of Study 2, only additional methodological aspects will be covered in the next sections. These aspects include the description of the expanded variables added to the study and the related revised study design. The empirical results and analysis section reviews the comparative performance of variable enhanced techniques, and the relative performance of parameters (variables). As before, the discussion section examines the findings of the study, provides an assessment of results vis-à-vis literature, and contrasts these with the findings from Study 2. Neverthless, for consistency, Figure 25 illustrates how each study informs the next and the specific objective of Study 3.

Figure 25: Study Progression Overview



6.2 Methodology

The methodology used to conduct this study is very similar to the one used in Study 2. The dataset remains unchanged and remains that of a large North American FMCG retailer. Techniques applied are the same, as are the measures of performance and fit. In addition, the cross-validation folds created and validated in Study 2 are reutilised in Study 3. In essence, the main dimension that changes significantly between both studies is the expansion of variables used in the application of statistical techniques. With the addition of data variables also comes the increased benefit derived from enhancing technique inputs along with the additional insights provided by the dimensionality reduction outcomes (what variables matter most).

6.2.1 Variables

The variables used for this study include some of the more prevalent customer data variables available to marketers (recency of last purchase, frequency of purchases, and monetary value of purchases) and some frequently used variables in the context of FMCG retail. Such variables cut across a number of Wedel and Kamakura's (2000) bases and include variables that are issued from usage, geographic, loyalty, situation and behavioural categories. Variables retained at the start of the selection process are shown in

Table **66**. Other than the typological differences, variables are also differentiated by the time horizons that they each span. Some variables are generally static and/or have decision rules that impart them with values that are hard-coded into the retailer's database: age, gender and life stage, for example. They are identified in the table with the value NA (Not Applicable) in the timeframe column. Other variables are time constrained: frequency of last transaction, monetary expenditures, loyalty programme promotional usage, period over period retention rate and channel usage, for example. As in Study 2, time horizons include a one-month (prior to the promotion) timeframe and 12 months. In most cases, variables are sourced from both time horizons in order to provide the most flexibility in model building. Variables that are constrained to a specific timeframe are identified accordingly.

In exceptional cases, variables with a very low incidence rate within the sample were removed. These variables included the bonus reward points earned on spend-related promotions one month prior to the promotion (i.e. spend $20 and receive 100 reward points), the bonus reward product points earned on specific products purchased one month prior to the promotion (i.e. buy 750ml San Pellegrino Sparkling water and receive 10 reward points), customer’s date of birth and customer’s first redemption date (date when a customer may have used points to redeem a reward of some sort). The incidence rate of these variables was significantly lower than other variables. In the case of the promotional, product-related reward miles and redemption date, this is related to the short period sampled and/or the lack of demonstrated behaviours within the sample. In the case of date of birth, it was not compulsory as a requirement for joining the loyalty programme, and such data was largely absent from the customer profile. Though these variables were eliminated, it should be noted that their 12-month counterparts were retained given a much higher incidence rate for the 12-month period.

Table 66: Variable Selection Table

Bases	Description	Variables Examined	Timeframe		Type of Variable
			1M	12 M	
General, observable					
Demographic	Dividing a market based on demographic variables such as age, gender, family size, income, occupation, education, religion or nationality	-Date of birth Age			Ratio
		-Gender	NA	NA	Nominal
		-Income	NA	NA	Ordinal
		-Lifestage			
		-Boomer	NA	NA	Nominal-Binomial
		-Family	NA	NA	Nominal-Binomial
		-Gen X	NA	NA	Nominal-Binomial
		-Senior	NA	NA	Nominal-Binomial
		-HH income	NA	NA	Ordinal
		-HH Size	NA	NA	Ordinal
		-Education			
		-College	NA	NA	Nominal-Binomial
		-No grade	NA	NA	Nominal-Binomial
		-Grade 9-13	NA	NA	Nominal-Binomial
		-Trade	NA	NA	Nominal-Binomial
		-University	NA	NA	Nominal-Binomial
Specific, observable					
Usage	Dividing markets based on usage patterns such as non-user, ex-user, potential user, first time user, regular user, high volume user.	-Recency of last transaction	●	●	Temporal
		-Frequency	●	●	Ratio
		-Monetary expenditures	●	●	Ratio
		-Loyalty program			Time
		promotional usage	●	●	Nominal-Binomial
		-Base points earned	●	●	Ratio
		-Bonus points earned		●	Ratio
		-Vendor points earned		●	Ratio
		-First issuance date	NA	NA	Time
		-First redemption date			Time
Loyalty	Dividing consumers based on brand loyalty: loyals, habituals, variety seekers and switchers	-Loyalty program avidity			
		-Avidity score	NA	NA	Interval
		-Program Tier	NA	NA	Nominal
		-Retention rate	●	●	Ratio
Situation	Related to usage segmentation, situation segmentation divides markets on the basis of the consumption or purchase situation of consumers.	-Marketing channel usage in previous periods			
		-email activity	●	●	Nominal-Binomial
		-web activity	●	●	Nominal-Binomial
Behavioural	Dividing markets based on consumer's knowledge of attitude toward uses for and responses to a product.	-Dept level consumption			
		-Dept used	●	●	Nominal-Binomial
		-Dept frequency	●	●	Ratio
		- # products purchased	●	●	Ratio
		- # of products shopped	●	●	Ratio

6.2.2 Study Design

As mentioned earlier, the thesis aims to address two gaps: (1) the relationship between customer selection techniques and performance and (2) the impact of dimensionality reduction on performance. In Study 2, the timeframe for RFM that had the most impact was tested and then used as a baseline to compare the relative effectiveness of customer selection techniques using the winning recency, frequency and monetary value variables.

In this study, a revised yet similarly constructed set of propositions is tested to further address these gaps using extended data. This new set of 12 propositions tests the effectiveness of the same techniques by removing the constraint of using solely the recency, frequency and monetary value variables. Techniques are applied using the expanded set of FCMG retail variables including the recency, frequency and monetary value variables. Propositions aim to compare the relative effectiveness of stepwise logistic regression, CHAID and neural network techniques to the RFM technique (that uses only the 12-month RFM variables).

This design choice is inspired by similar work by McCarty (2006) and Yang (2004), where the model performance using basic recency, frequency and monetary value variables is tested against expanded data sets mainly composed of additional behavioural variables. The set of propositions is detailed in Table 67.

Table 67: Technique Effectiveness Propositions

ALL SIGNIFICANT SELECTION VARIABLES					
		More Effective than			
		RFM	CHAID	Logistic Regression	Neural Networks
Less Effective than	RFM	NA	P16	P19	P22
	CHAID	P13	NA	P20	P23
	Logistic Regression	P14	P17	NA	P24
	Neural Networks	P15	P18	P21	NA

These propositions can be read as follows:

In the context of FMCG retail promotions aiming to maximise the response rate of direct to consumer using only recency, frequency and monetary value of purchases variables, the technique of:

P13: RFM is more effective than CHAID-DD

P14: RFM is more effective than Logistic Regression- DD

P15: RFM is more effective than NN- DD

P16: CHAID- DD is more effective than RFM

P17: CHAID- DD is more effective than Logistic Regression-AD

P18: CHAID- DD is more effective than NN- DD

P19: Logistic Regression- DD is more effective than RFM

P20: Logistic Regression-AD is more effective than CHAID- DD

P21: Logistic Regression-AD is more effective than NN- DD

P22: NN-AD is more effective than RFM-DD

P23: NN-AD is more effective than CHAID- DD

P24: NN- DD is more effective than Logistic Regression- DD

Where DD represents applications using the extended variable set

6.3 Empirical Results and Analysis

6.3.1 Technique Performance

6.3.1.1 Assessing Propositions P13 to P24

Following the assessment of the optimal RFM technique timeframe and the comparison of the relative performance of techniques in Study 2, this section presents the relative performance results of advanced techniques (CHAID, ANN, and logistic regression) using an expanded set of variables issued from FMCG retail. Figures 26, 27 and 28 show the lift and gains by decile of the four different techniques applied using the extended data variable set. As in the previous study, results shown are averages of all 10 cross-validation estimates issued from applying the variable enhanced techniques on the original cross-validation folds.

Figure 26 presents a visual representation of top decile performance in the form of average lift per decile. The chart shows that all advanced statistical and machine-learning techniques seem to provide a fairly similar lift at each decile level versus the overall average response rate. Unlike Study 2, advanced technique results demonstrate a significant and visible lift increase versus the RFM technique. The raw input data of the lift charts (Table 68) allows for a more quantitative assessment of comparative performance. From this table, it can be concluded that there is an important difference in the lift index between RFM and statistical and machine-learning techniques. RFM results show that response rates at the top decile are nearly 100-percent higher than the average response rate, while comparable results from CHAID, neural networks and logistic regression show increases of more than 200 percent. Within the subset of advanced techniques, results indicate that neural networks and logistic regression provide the highest increase in lift (response rate of decile divided by average response rate) at the top decile, followed by CHAID and RFM.

A similar phenomenon can be observed at decile two, although the relative position of techniques changes. The best performing techniques at the second decile include CHAID and neural networks, with lift indices of 215 and 214, or otherwise said an increase of response of nearly 115 percent versus the average response rate. They are followed by logistic regression and by RFM, respectively. Deciles three, four and five show a reversal of the trend, with RFM lift surpassing all other techniques. This is to be expected considering the high discrimination power shown by the advanced statistical and machine-learning techniques in deciles one and two. More specifically, in deciles three and four, the lift index of CHAID is the second highest followed by logistic regression and neural networks, while in decile five (and most of subsequent deciles) all advanced techniques provide a similar lift index.

Table 68: Input for Lift Charts – Expanded FMCG Retail Variables

Deciles	RFM	CHAID	Stepwise LR	NN	No Model
Top	197	308	319	318	100
2	183	215	209	214	100
3	158	153	147	143	100
4	147	114	107	106	100
5	110	77	76	78	100
6	89	51	55	54	100
7	59	42	38	42	100
8	31	26	25	25	100
9	22	11	17	13	100
Bottom	5	4	9	7	100

Table 69: Input for Cumulative Lift Charts - Expanded FMCG Retail Variables

Deciles	RFM	CHAID	Stepwise LR	NN	No Model
Top	197	308	319	318	100
2	190	261	264	266	100
3	179	225	225	225	100
4	171	197	195	195	100
5	159	173	171	172	100
6	147	153	152	152	100
7	135	137	136	136	100
8	122	123	122	122	100
9	111	111	110	110	100
Bottom	100	100	100	100	100

Table 70: Input for Cumulative Gains Charts - Expanded FMCG Retail Variables

Deciles	RFM	CHAID	Stepwise LR	NN	No Model
Top	19.72%	30.80%	31.88%	31.77%	10.00%
2	38.04%	52.27%	52.74%	53.21%	20.00%
3	53.84%	67.52%	67.46%	67.54%	30.00%
4	68.55%	78.90%	78.17%	78.11%	40.00%
5	79.53%	86.65%	85.74%	85.89%	50.00%
6	88.39%	91.74%	91.20%	91.28%	60.00%
7	94.31%	95.99%	94.99%	95.47%	70.00%
8	97.39%	98.57%	97.46%	97.99%	80.00%
9	99.54%	99.63%	99.12%	99.34%	90.00%
Bottom	100.00%	100.00%	100.00%	100.00%	100.00%

Examining results cumulatively across lifts charts (Figure 27), gains charts (Figure 28) and input data

(Table 69 and Table 70) provides a more holistic perspective and showcases the overall impact of the high discrimination of advanced techniques (at top two deciles) on the individuals techniques' performance across all five top deciles. Looking at logistic regression, CHAID and neural networks, the lifts and gains obtained at the first two deciles provide enough response lift to allow these techniques' cumulative performance to surpass RFM well past decile five into decile six. Of course, the cumulative lift and gain changes vary somewhat by technique and are greatly affected by the individual decile lifts and gains. For example, the reason that CHAID can achieve a cumulative lift that is similar to neural networks and logistic regression is the individual CHAID lift from deciles two, three and four offsets the lower lift from decile one. In contrast, logistic regression and neural networks' relative lift decrease in deciles two and three. However, neural networks' relative lift is stronger in decile two, whereas logistic regression is stronger in decile three.

As mentioned earlier, in addition to model performance, model fit must also be examined. With the addition of additional variables, Gini coefficients change. New coefficients are described in the next section.

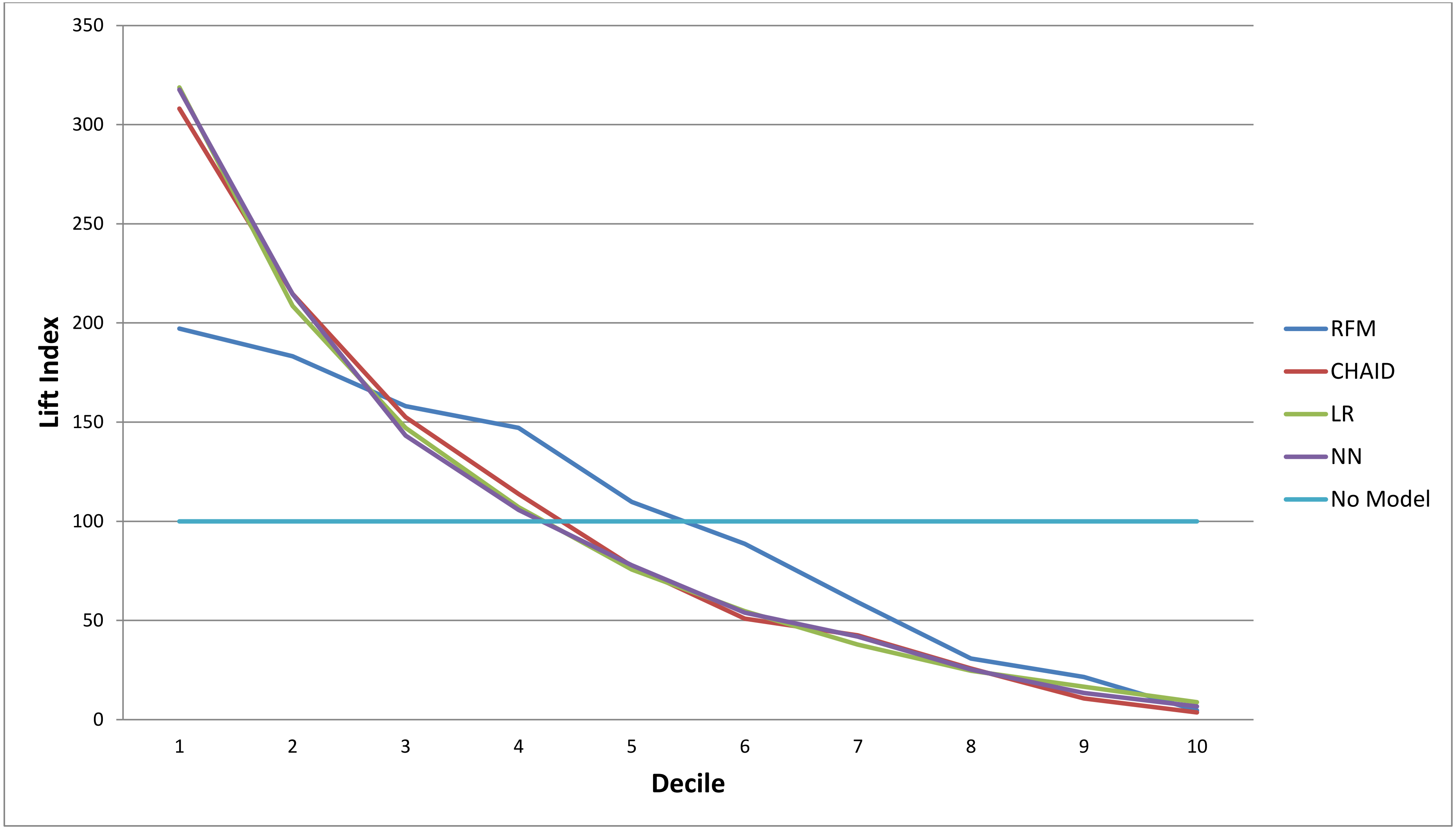


Figure 26: Lift Chart - Expanded FMCG Retail Variables

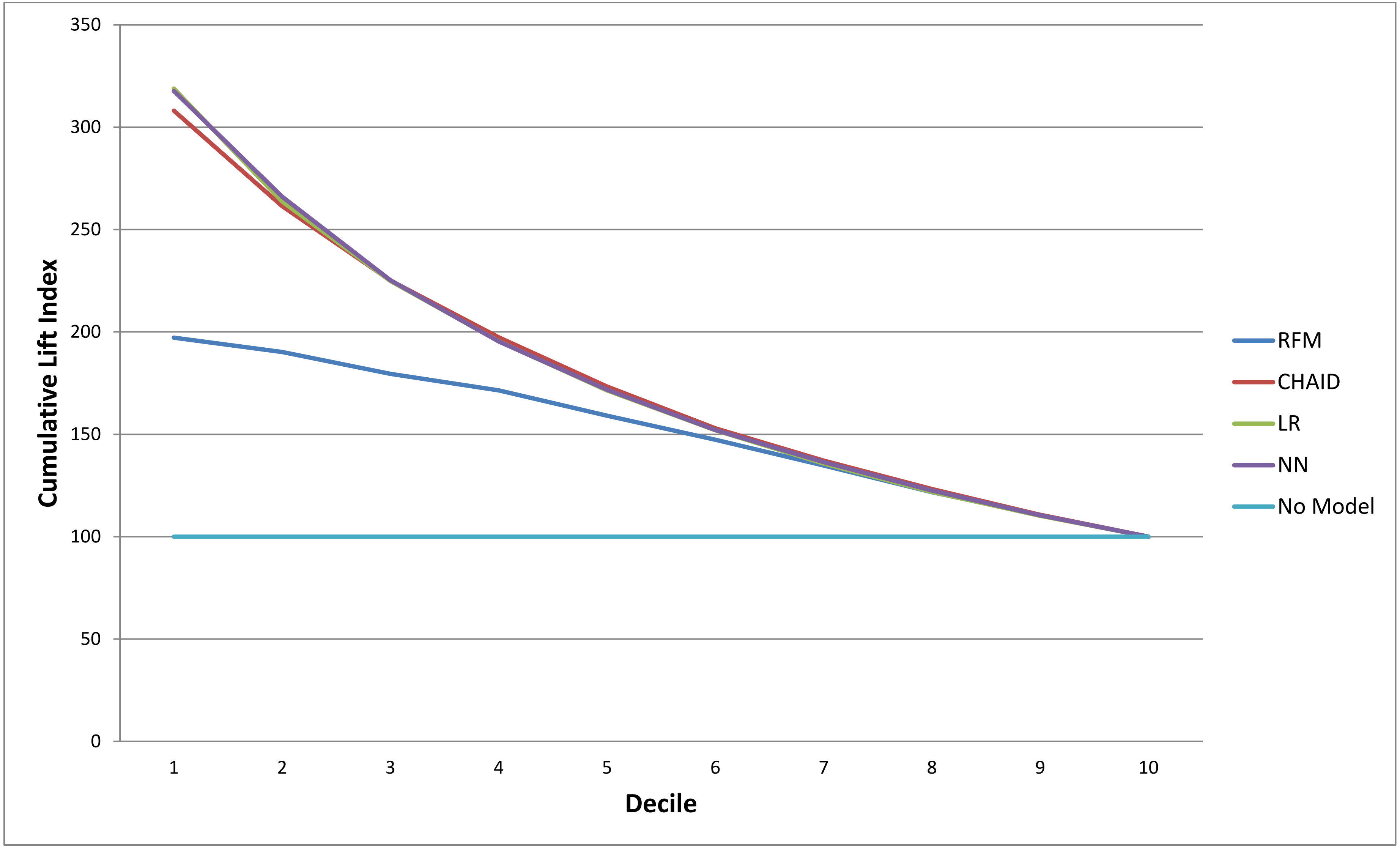


Figure 27: Cumulative Lift Chart - Expanded FMCG Retail Variables

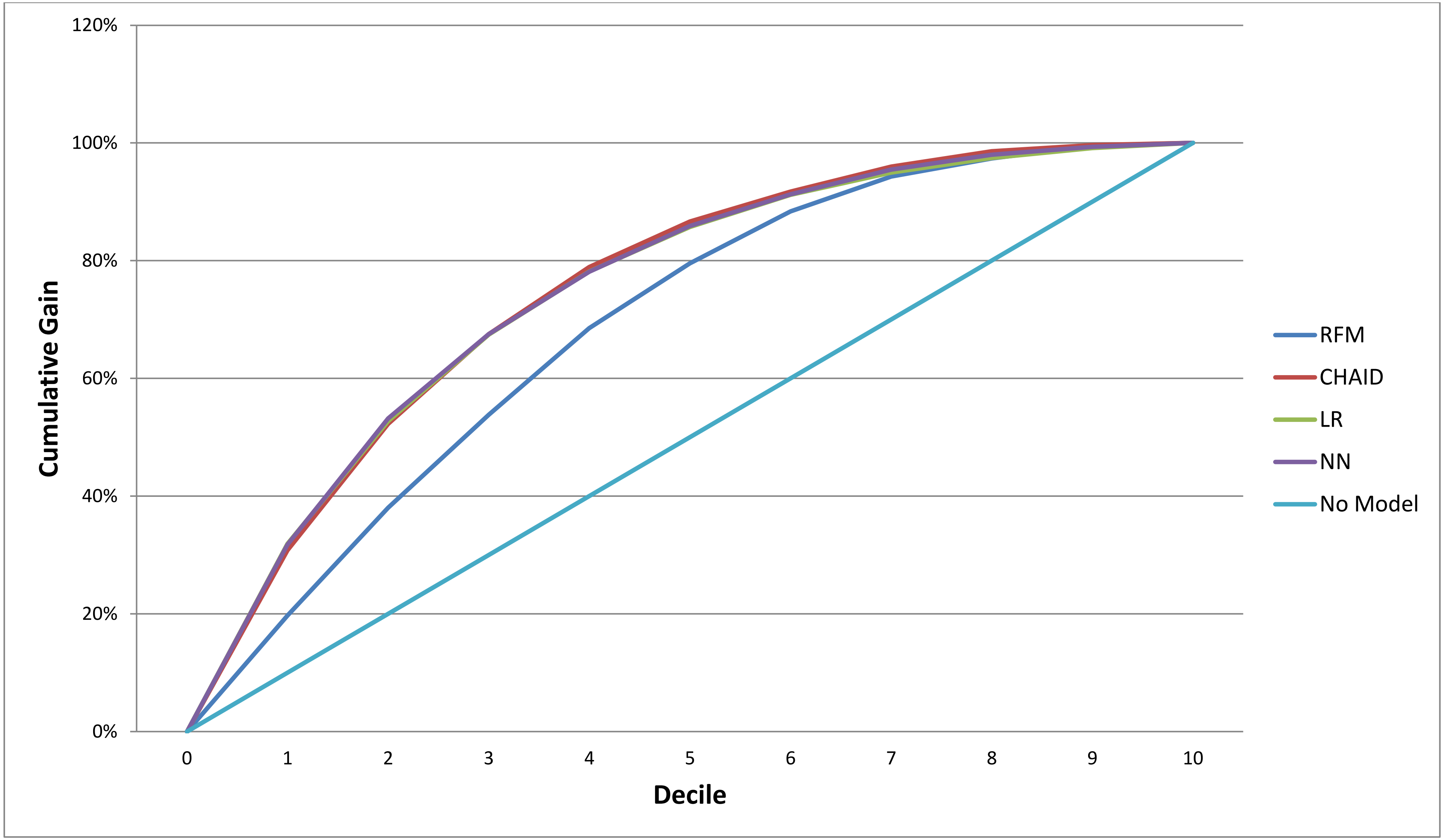


Figure 28: Cumulative Gain Chart - Expanded FMCG Retail Variables

6.3.1.2 *Gini Coefficient and Percentage Correctly Classified (PCC)*

The Gini coefficients of techniques (averaged from Gini coefficients from all cross-validation folds) are shown in Table 71. The output of the Gini coefficient calculations shows that, among the extended variable models, CHAID provides the highest level of fit. The second-best fitting technique is neural networks followed very closely by logistic regression. Finally, and quite far behind, is RFM with a Gini coefficient of 0.403871. This difference is not unexpected, as additional variables allow the individual techniques' fit to be much more discriminatory vis-à-vis the RFM technique which (by its very design) still utilises the limited set of RFM variables.

The table also shows the related PCC scores for all relevant techniques. Using PCC as ranking criteria, CHAID remains the strongest performing technique but logistic regression would have ranked higher than ANN. As mentioned in the previous study, this can be explained by the somewhat different fit measurement dimension of each measure; PCC comparing actual yes/no responses between actual response behaviour and modelled behaviour and Gini indicating how well the two groups show separation from one another.

Table 71: Average Technique Gini Coefficients and PCC – Expanded FMCG Variables

Model & Rank	**Gini Coeff**	**PCC**
1. CHAID	0.521057	88.2%
2. ANN	0.500625	84.7%
3. LR	0.49701	84.8%
5. RFM	0.403871	NA

6.3.2 Evaluation of Techniques Outputs

6.3.2.1 *Stepwise Logistic Regression*

As mentioned in Study 2, the first relevant information in the stepwise logistic regression output is the validation of whether input variables contribute to the predictive power of the model. This is not possible at step 0 of the full dataset model, given the high degree of inter-variable multicollinearity. This redundancy results in SPSS not being able to adequately report the residual chi-square statistic. Nevertheless, observing the outputs from subsequent steps illustrates the predictive contributions of the model and its variables (versus the constant).

As a result, the first step consists of observing how many steps are conducted to achieve an optimal outcome and examining the variables that emerge as significant. Given the difficulty in visually representing the number of folds in addition to representing the significant variables, I report the minimum and maximum number of steps from the logistic regression and then proceed to visually represent the variable outcomes in Table 72 (a variable description table follows).

The minimum number of steps conducted within one fold model totalled 36 while the maximum number of steps totalled 44. From a variable perspective, not all folds produced the same significant variable output. Across all folds, 52variables emerged as being significant in one fold or another. Seventeen of these variables were socio-demographic in nature, 10 were loyalty program specific, three indicated channel preferences and utilisation, and 22 were transactional (meaning they covered all retailer patronage and available product transaction details).

From a variable representation perspective, most variables emerged in six or more model iterations while a small amount of variables emerged in one or two models. Though some variables were different, it remains that in all 36 to 44 steps of each model iteration, variables from socio-demographic, loyalty, channel and transaction dimensions emerged as contributing to the prediction power of the model. Similar to Study 2, this also included the monetary value of expenditures both for a full year and for the pre-promotional period, and the frequency of transactions during the full year. The Wald statistic presented for each variable in Table 72 indicates whether the b-coefficient for the predictor variable is significantly different than 0. In the case of all the variables included in the last step of the folds, the significance scores indicate that the values are significantly different than 0 at a .05 level and so all of the variables that emerge from all of the fold models do predict responses with the exception of one variable: Household_Size_Cat(1). This indicates that this variable’s individual significance is not different than 0, but its addition to other model variables increases prediction.

Table 72: Overview of Variables in the Equation

Category	Variables	Fold Count	B		S.E.		Wald		df	Sig.	Exp(B)	
			min	max	min	max	min	max			min	max
Socio-Demographic	Household_Size_Cat(1)	10	-.035	.054	.047	.050	.001	1.154	1	0.472-.0979	0.965	1.055
	Household_Size_Cat(2)	10	.179	.201	.026	.026	48.261	60.840	1	0	1.196	1.223
	Household_Size_Cat(3)	10	.153	.179	.030	.030	25.756	35.453	1	0	1.165	1.196
	Household_Size_Cat(4)	10	.105	.128	.028	.028	14.262	21.086	1	0	1.111	1.136
	Household_Size_Cat(5)	10	-.144	-.097	.034	.034	8.163	17.836	1	0	0.866	0.908
	Gender_Rec(1)	10	.036	.058	.027	.027	1.753	4.540	1	0	1.037	1.060
	Gender_Rec(2)	10	.101	.130	.021	.021	23.187	37.951	1	0	1.106	1.139
	FLAG_LIFESTAGE_GENX	1	.161	.709	.022	.424	2.787	60.492	1	0	1.175	2.031
	FLAG_LIFESTAGE_FAMILY	1	.248	.769	.032	.421	3.342	76.025	1	0	1.281	2.157
	FLAG_LIFESTAGE_BOOMER	1	.931	.931	.421	.421	4.895	4.895	0	0	2.536	2.536
	FLAG_LIFESTAGE_SENIOR	1	1.026	1.026	.421	.421	5.936	5.936	0	0	2.791	2.791
	NOT_CL	2	.455	.525	.226	.226	4.047	5.330	0	0	1.576	1.691
	MOVABLE_DWLG	9	.229	.290	.098	.101	5.323	8.816	1	0	1.258	1.337
	NOGRADE	9	-.465	-.176	.076	.107	5.098	19.475	1	0	0.628	0.839
	GRADE913	10	.348	.542	.112	.126	7.696	23.321	1	0	1.416	1.719
	UNIVERSITY	6	-.249	-.205	.097	.097	4.444	6.621	0	0	0.779	0.815
Loyalty	FirstIssuanceDate	10	.000	.000	.000	.000	30.615	41.101	1	0	1.000	1.000
	AVID_08	10	-.175	-.171	.009	.009	337.932	355.164	1	0	0.839	0.843
	FLAG_AMFB	10	-.087	-.065	.023	.024	7.595	13.366	1	0	0.917	0.937
	FLAG08_ENTERPRISE_ MAINTAIN	10	-.407	-.373	.061	.062	37.318	43.488	1	0	0.666	0.689
	FLAG08_ENTERPRISE_LOW	1	-.551	-.551	.255	.255	4.669	4.669	1	0	0.576	0.576
	FLAG_SPON_RETAINED_12M	10	.346	.416	.100	.103	12.009	16.412	1	0	1.413	1.516
	MILESB1_12M	10	-.002	-.001	.000	.000	43.083	119.533	1	0	0.998	0.999
	MILESBONUS_12M	8	-.001	.013	.000	.000	4.298	4329.615	1	0	0.999	1.013
	MILESVENDOR_12M	8	.001	.013	.000	.000	160.190	4309.758	1	0	1.001	1.013
	MILESOTHER_12M	6	.000	.001	.000	.000	3.682	181.079	0	0	1.000	1.001
	MILESOTHER_1M	2	.000	.000	.000	.000	5.511	5.560	0	0	1.000	1.000

Category	Variables	Fold Count	B		S.E.		Wald		df	Sig.	Exp(B)	
			min	max	min	max	min	max			min	max
Channel	FLAG_EMAILABLE	10	.094	.107	.021	.021	20.593	26.480	1	0.000	1.099	1.113
	FLAG_MAILABLE	10	.481	.534	.095	.097	25.553	30.585	1	0.000	1.618	1.706
	WEB_12MTHS	10	.070	.095	.022	.022	10.550	18.958	1	0.000	1.073	1.099
Trasactional Information	FREQUENCY_full_Year	10	-.002	-.001	.000	.000	11.978	105.246	1	0.000	0.998	0.999
	MONETARY_full_year	10	.000	.000	.000	.000	71.060	112.051	1	0.000	1.000	1.000
	MONETARY_prepromo	10	-.001	-.001	.000	.000	104.646	124.344	1	0.000	0.999	0.999
	NB_DEPT_FULL_YEAR	9	.049	.058	.005	.005	87.917	123.838	1	0.000	1.050	1.059
	NB_SKU_FULL_YEAR	9	.001	.050	.000	.005	96.784	282.230	1	0.000	1.001	1.052
	DEPT_FREQUENCY.2	9	.001	.020	.000	.004	15.229	289.905	1	0.000	1.001	1.020
	DEPT_FREQUENCY.4	9	-.046	.020	.004	.004	21.726	117.538	1	0.000	0.955	1.020
	DEPT_FREQUENCY.5	8	-.073	-.041	.004	.024	5.854	101.122	1	0.000	0.930	0.960
	DEPT_FREQUENCY.9	9	-.068	-.009	.005	.024	4.123	8.857	1	0.000	0.935	0.991
	DEPT_FREQUENCY.10	10	.027	.032	.005	.005	28.383	41.617	1	0.000	1.028	1.033
	DEPT_FREQUENCY.11	9	.013	.019	.005	.006	5.868	11.497	1	0.000	1.013	1.019
	DEPT_FREQUENCY.16	10	.019	.024	.007	.007	6.529	10.896	1	0.000	1.019	1.024
	DEPT_FREQUENCY.20	1	-19.719	-19.719	12511.967	2511.967	.000	.000	0	0.000	0.000	0.000
	DEPT_FREQUENCY.21	10	.040	.048	.005	.005	61.897	91.000	1	0.000	1.041	1.049
	DEPT_FREQUENCY.22	10	.174	.190	.033	.034	25.770	32.839	1	0.000	1.189	1.209
	DEPT_FREQUENCY.23	10	.038	.046	.009	.009	16.838	25.344	1	0.000	1.038	1.047
	DEPT_FREQUENCY.25	6	.057	.410	.021	.150	4.752	7.434	1	0.000	1.059	1.507
	DEPT_FREQUENCY.31	1	-.072	.066	.012	.021	6.428	35.319	1	0.000	0.930	1.068
	DEPT_FREQUENCY.32	9	-.173	-.069	.012	.042	17.180	42.636	1	0.000	0.841	0.933
	DEPT_FREQUENCY.64	1	-.526	.050	.025	.251	3.932	14.894	1	0.000	0.591	1.051
	DEPT_FREQUENCY.34	7	-.191	-.161	.042	.043	14.442	19.332	0	0.000	0.826	0.851
	DEPT_FREQUENCY.60	2	2.071	2.071	.693	.693	8.921	8.921	0	0.000	7.929	7.929
	Constant	10	.574	2.518	1.103	1.208	.225	5.215	1	0.000	1.775	12.404

Table 72 Addendum: Variable Descriptions

Variable	Description
APARTMENT_FLAG	A flag used to indicate if the customer address is an Apartment.
AVID_08	Avid decile 2008
AVID_FLAG08	Avid flag 2008
BIRTH_DATE	Date of birth
FIRST_ISSUANCE_DATE	Date of the customer's first loyalty accumulation transaction
FIRST_REDEMPTION_DATE	Date of the customer's first redemption transaction.
FLAG08_ENTERPRISE_BEST	Enterprise Segment of 2008 = Best
FLAG08_ENTERPRISE_LOW	Enterprise Segment of 2008 = Low
FLAG08_ENTERPRISE_MAINTAIN	Enterprise Segment of 2008 = Maintain
FLAG08_ENTERPRISE_NEXTBEST	Enterprise Segment of 2008 = Next Best
FLAG_AMFB	Derived yes/no flag indicating that this account has been determined to be a small business account.
FLAG_EMAILABLE	Yes/no flag indicating whether the customer may be contacted by LMG via Email.
FLAG_LIFESTAGE_BOOMER	Current Lifestage Segment = Boomer
FLAG_LIFESTAGE_FAMILY	Current Lifestage Segment = Family
FLAG_LIFESTAGE_GENX	Current Lifestage Segment = Gen X
FLAG_LIFESTAGE_SENIOR	Current Lifestage Segment = Senior
FLAG_MAILABLE	A flag to indicate whether can contact the customer via Mail.
FLAG_SPON_INACTIVE_12M	Migration 12 years = Program Inactive customer (inactive during the current year)
FLAG_SPON_INACTIVE_1M	Migration 1 month = Program Inactive customer (inactive during the current month)
FLAG_SPON_NEW_12M	Migration 12 years = Program New customer (active for the first time in the current year)
FLAG_SPON_NEW_1M	Migration 1 month = Program New customer (active for the first time in the current month)
FLAG_SPON_REACTIVATED_12M	Migration 12 years = Program Reactivated customer (active in the current year but not new and inactive during the previous year)
FLAG_SPON_REACTIVATED_1M	Migration 1 month = Program Reactivated customer (active in the current month but not new and inactive during the previous month)
FLAG_SPON_RETAINED_12M	Migration 12 years = Program Retained customer (active during the current and the previous years)
FLAG_SPON_RETAINED_1M	Migration 1 month = Program Retained customer (active during the current and the previous months)
GENDER	Represents the customer's gender.
GOLD_FLAG08	customer with a gold status in 2008
HHLD_INCOME_CATEGORY	Household income category of the customer
HHLD_SIZE_CATEGORY	Household size of the customer
MILESB1_12M	Base 1 points earned 12 months
MILESB1_1M	Base 1 points earned 1 month
MILESBONUS_12M	Bonus points earned 12 months
MILESBONUS_1M	Bonus points earned 1 month
MILESOTHER_12M	Other points earned 12 months (other than base 1, bonus or vendor)
MILESOTHER_1M	Other points earned 1 month (other than base 1, bonus or vendor)
MILESVENDOR_12M	Vendor points earned 12 months
MILESVENDOR_1M	Vendor points earned 1 month
PROMO_BONUS_MILES	Bonus points earned during the January 2009 promotion
PROMO_NB_COUPONS	#Coupons used during the January 2009 promotion
REP_ADDRESS_POSTAL_CODE	Customer's Reporting address Postal code.
SBYA_ACTIVE_12M	Active flag 12 months
SBYA_ACTIVE_1M	Active flag 1 month
TRSB1_12M	Number of base 1 transactions done 12 months
TRSB1_1M	Number of base 1 transactions done 1 month
WEB_12MTHS	Web active 12 months
WEB_1MTH	Web active 1 month
AGE017	%of persons with age less than 17
AGE1824	%of persons with age between 18-24

Variable	Description
AGE2534	%of persons with age between 25-34
AGE3544	%of persons with age between 35-44
AGE4554	%of persons with age between 45-54
AGE5564	%of persons with age between 55-64
AGE65P	%of persons with age 65+
APARTMENT	Dwelling type: Living in an apartment
COLLEGE	Education: College degree
COMMON_LAW	%Living in common-law
G5_FOOD_FROM_STORES	$spent annually on food from stores
NOGRADE	Education: No certificate, diploma or degree
GRADE913	Education: High school certificate
HHLD_SIZE_1	Household size = 1
HHLD_SIZE_2	Household size = 2
HHLD_SIZE_3	Household size = 3
HHLD_SIZE_4or5	Household size = 4 or 5
HHLD_SIZE_6P	Household size = 6+
HOUSE	Dwelling type: Living in a house
INCOME_AVG	Average household income
INCOME_MED	Median household income
MARRIED	Marital Status: Married
MOVABLE_DWLG	Dwelling type: Living in a movable dwelling
NBCHILD	Average number of children at home
NOT_CL	% living in common-law
OWNED	%Owned dwelling
RENTED	%Rented dwelling
SEPARATED	Marital Status: Separated
SINGLE	Marital Status: Single
TRADE	Education: Apprenticeship, trades certificate or diploma
UNIVERSITY	Education: University
WIDOWED	Marital Status: Windowed
FREQUENCY	Frequency of customer 1 mth prior to Promotion
MONETARY	Monetary value of active customer 1 mth prior to Promotion
NB_DEPT	Nb of dept shopped by active customer 1 mth prior to during Promotion
NB_SKU	Nb of skus shopped by active customer during 1 mth prior to Promotion
RECENCY	Recency of active customer 1 mth prior to l Promotion
POS_DEPT_NUMBER	Dept shopped by active customer during Promotion
DEPT_FREQUENCY	Frequency of customer within dept during Promotion

As mentioned in the overview of techniques, one of the weaknesses of forward selection stepwise logistic regression is the risk of casting too wide a net and identifying independent variables only accidentally related to the dependent variable. However, the ten-fold cross-validation of models allows for the detection of such accidentally significant variables as these variable inclusion counts from Table 72 are very low. Variables with the lowest inclusion counts in this table included customer indicators of Generation X, Family, Baby Boomer or Senior life stages; the indicator for customer living in common law; the indicator for loyalty program members being low on the program involvement scale; the points earning indicator at other loyalty program partners; and the frequency of transactions in departments 20, 31, 64 and 60. Variables that were excluded from the model because they did not add any additional predictive power are not included in the list. However, it is important to note that the RFM variables of recency (in both one- and 12-month timeframes) and the frequency of transactions (one-month timeframe) did not emerge as significant. This is consistent with the outcome of Study 2.

Looking at the top 25 retained variables in Table 73, two loyalty program related variables are clearly more significant contributors to the model than others: the number of bonus points accumulated by the customer in the last 12 months and the number of points accumulated on product-specific offers (provided by FMCG vendors) over the same period. These variables are at least three to 10 times more significant than other variables based on the differential between Average Wald statistic scores. Two subsequent variables of importance that are two to three times more significant than other variables include the level of engagement of the customer in the loyalty program (avid) followed by the number of unique products purchased by the customers with the retailer specifically. With the exception of household size, which is the next-strongest individual variable, the top 10 variables are all either loyalty or transactional variables. As it pertains to socio-demographic variables, the most significant variables among the retained groups all consist of household size, life stage or gender variables.

Table 73: Logit Variable Significance Ranking by Wald Statistic

Rank	Variable	Present in Models Count	Wald Statistic Average
1	MILESBONUS_12M	8	3,387.329
2	MILESVENDOR_12M	8	1,001.341
3	AVID_08	10	345.669
4	NB_SKU_FULL_YEAR	9	248.565
5	Household_Size_Cat	10	128.709
6	MONETARY_prepromo	10	112.237
7	NB_DEPT_FULL_YEAR	9	106.750
8	MONETARY_full_year	10	101.396
9	MILESB1_12M	10	99.565
10	DEPT_FREQUENCY.4	9	98.464
11	DEPT_FREQUENCY.21	10	74.817
12	FREQUENCY_full_Year	10	68.084
13	FLAG_LIFESTAGE_FAMILY	1	63.272
14	Household_Size_Cat(2)	10	53.928
15	FLAG_LIFESTAGE_GENX	1	51.200
16	MILESOTHER_12M	6	46.905
17	DEPT_FREQUENCY.2	9	45.306
18	FLAG08_ENTERPRISE_MAINTAIN	10	40.998
19	DEPT_FREQUENCY.31	9	34.424
20	FirstIssuanceDate	10	34.002
21	DEPT_FREQUENCY.10	10	33.574
22	Gender_Rec	10	33.247
23	Gender_Rec(2)	10	32.148
24	Household_Size_Cat(3)	10	30.168
25	DEPT_FREQUENCY.22	10	29.698

The final step of the logit analysis process is to observe the outputs that are generated by the multiple steps of stepwise regression. As mentioned in Study 2, the overall fit of the model is generally evaluated by using the -2LL log-likelihood statistic. Given the high number of steps and the difficulty in representing the -2LL improvements in a table, Figure 29 below illustrates the improvement in -2LL from folds one to 44. Across all fold iterations, the decreasing -2LL scores indicate an improvement in prediction.

Figure 29: -2 Log Likelihood Statistic by Fold

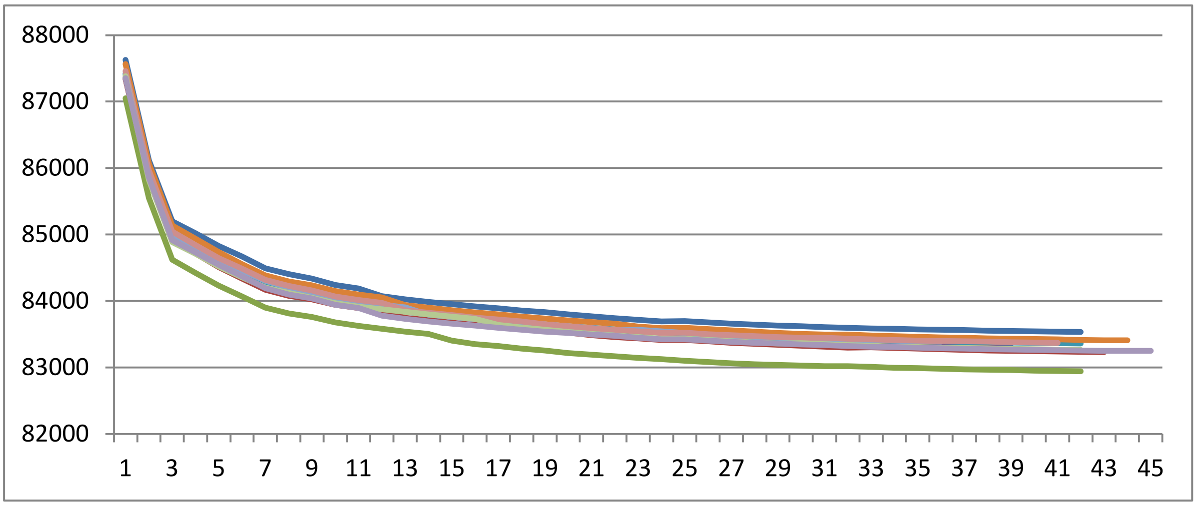
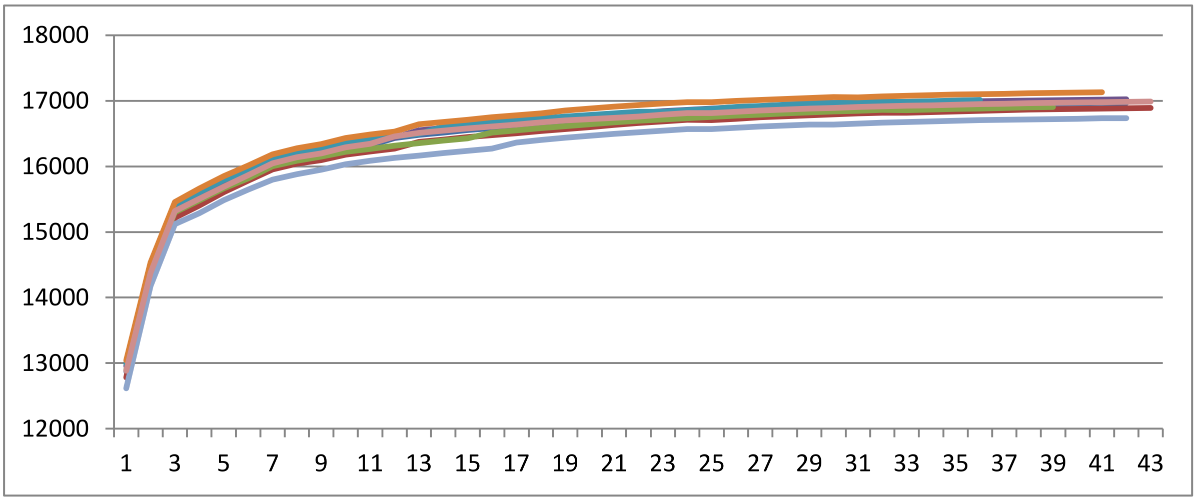


Figure 30 answers the question of how much better each step is at prediction. It does so by assessing the chi-square value of each step by subtracting the value of the -2LL statistic of a previous step by that of the evaluated step. The visual outputs demonstrate a significant chi-square improvement at each subsequent step for each of the 10 represented folds. The significance score for each step and fold is significant at a .05 level.

Figure 30: Chi-Square Value by Logistic Regression Step and Fold



Other measures of explanatory power include the Cox and Snell R Square and Nagelkerke R Square adjusted value. The increasing statistics by step and fold of both R2 values shown in Figure 31 and Figure 32 also consistently indicate that each additional step in the logistic regression contributes to increasing the explanatory power of the model.

Figure 31: Cox and Snell R Square

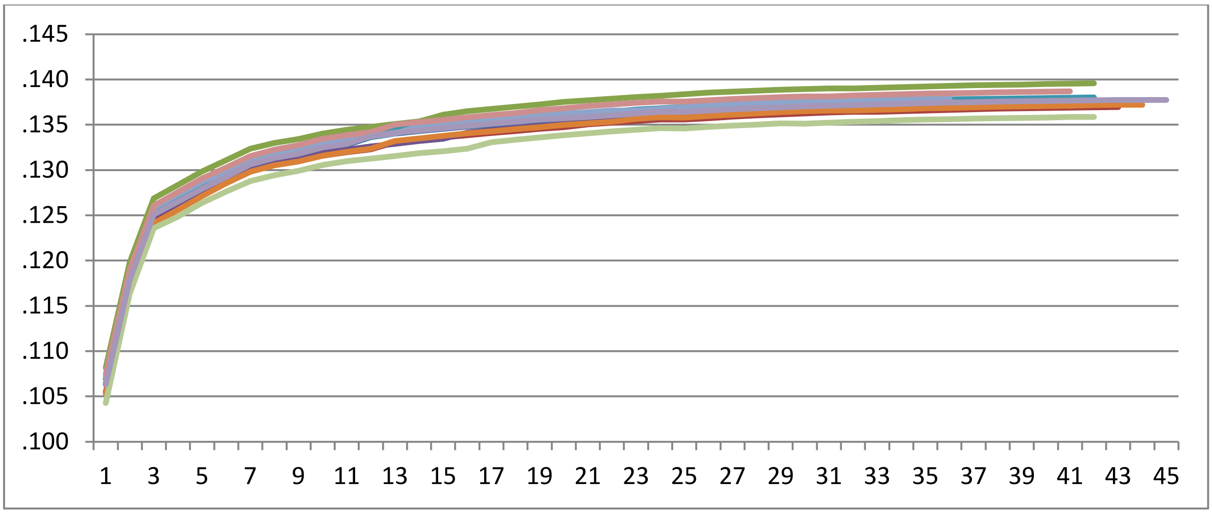
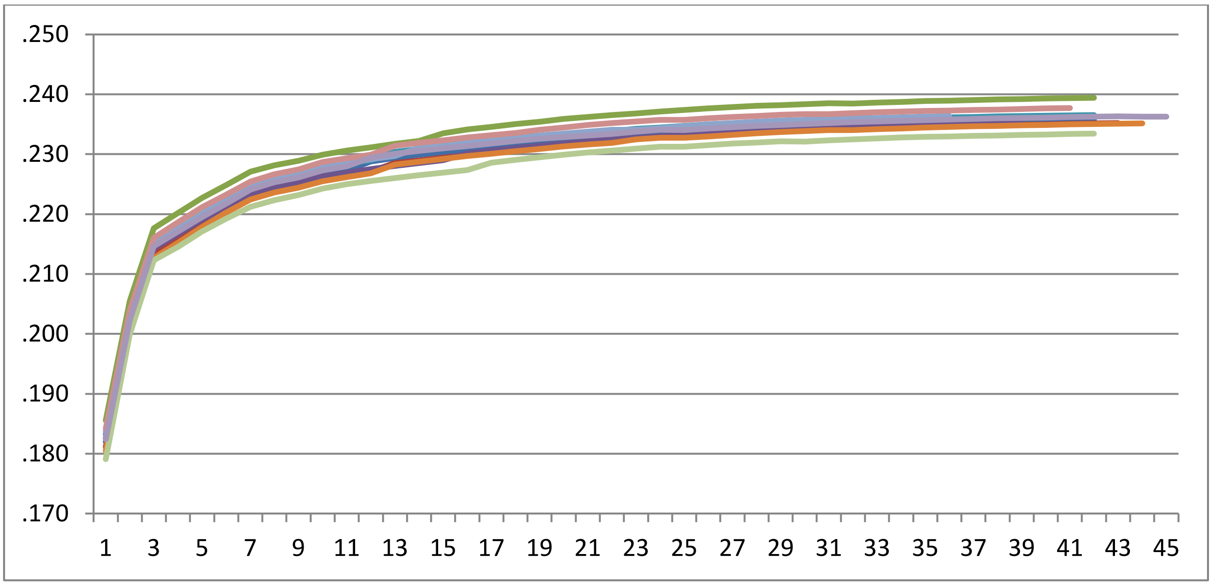


Figure 32: Nagelkerke R Square



In terms of pure final explanatory power statistic scores, the different fold iterations provided a Cox and Snell R Square between 0.136 and 0.140 and a Nagelkerke R Square between 0.233 and 0.239.

6.3.2.2 CHAID

CHAID model SPSS outputs consist of the model summary, the tree diagrams demonstrating how different variables are positioned across the tree design, and risk statistics and classification results.

As with the RFM-based model, to obtain segments large enough for analysis, the minimum size of parent nodes was set to 100 observations and child nodes to 50 observations. All variables presented in Table 74 were included in the model. However, only a select number were retained. As with the logistic output, significant variables all fell within the categories of socio-demographic, loyalty, channel usage and transactional information.

Table 74: CHAID Model Summary

Specifications	Growing Method	CHAID
	Dependent Variable	PROMO_RESPONDER
	Independent Variables	All variables
	Maximum Tree Depth	3
	Minimum Cases in Parent Node	100
	Minimum Cases in Child Node	50
INDEPENDENT VARIABLES INCLUDED		
Category	**Variable**	**Fold Count**
Socio-Demographic	Gender_Rec	9
	AGE017	3
	AGE4554	3
	AGE2534,	1
	AGE3544,	1
	HHLD_SIZE_1	2
	HHLD_SIZE_2	9
	HHLD_SIZE_3	2
	FLAG_LIFESTAGE_BOOMER	1
	FLAG_LIFESTAGE_GENX	2
	FLAG_LIFESTAGE_FAMILY	5
	WIDOWED	4
	FLAG_LIFESTAGE_SENIOR	4
	SINGLE	1
	Household_Income_Cat	6
	Household_Size_Cat	1
	APARTMENT_FLAG	1
	TRADE	1
	UNIVERSITY	5
	NOT_CL	2
	NBCHILD	1
	INCOME_AVG	1
	INCOME_MED	1
Loyalty	MILESBONUS_12M	10
	MILESB1_12M	7
	MILESVENDOR_12M	10
	AVID_08	10
	AVID_FLAG08	7
	GOLD_FLAG08,	1
	FLAG_AMFB	9
	FLAG08_ENTERPRISE_BEST	3
	FLAG08_ENTERPRISE_NEXTBEST	7
	FLAG08_ENTERPRISE_LOW	2
	TRSB1_12M	10
	MILESOTHER_12M	10
	FirstIssuanceDate	8

INDEPENDENT VARIABLES INCLUDED (continued)		
Category	**Variable**	**Fold Count**
Channel	FLAG_EMAILABLE	8
	WEB_12MTHS	7
Transactional Information	RECENCY2_Full_Year	8
	FREQUENCY_full_Year	5
	MONETARY_prepromo	2
	RECENCY2_prepromo	3
	MONETARY_full_year	10
	NB_SKU_FULL_YEAR	1
	NB_DEPT_FULL_YEAR	3
	DEPT_FREQUENCY.12	7
	DEPT_FREQUENCY.23	1
	DEPT_FREQUENCY.2	4
	DEPT_FREQUENCY.4	2
	DEPT_FREQUENCY.6	1
	DEPT_FREQUENCY.7	6
	DEPT_FREQUENCY.10	1
	DEPT_FREQUENCY.13	4
	DEPT_FREQUENCY.16	5
	DEPT_FREQUENCY.21	1
	DEPT_FREQUENCY.22	1
	DEPT_FREQUENCY.31	4
	DEPT_FREQUENCY.32	1
	G5_FOOD_FROM_STORES	2
	min	Max
Number of Nodes	80	97
Number of Terminal Nodes	55	67
Depth	3	

Variables with the lowest inclusion counts in Table 74 (and as a result, the highest level of consistency in predicting model outputs) included age group indicators (0-17, 45-54, 25-34, 35-44), household size indicators (single person, three persons) and category, life stage (boomer and gen x), marital status (single and common law), number of recorded children, indicators of income (average), living in an apartment indicator, indicators for loyalty program engagement (best and low), trade education, monetary value and recency of last transaction (both for one month timeframe), number of total individual products purchased in a year, number of total departments shopped in a year, and frequencies of transactions with specific departments (4, 6, 10, 21, 22, 23, 31, 32). This list is substantially longer than the list of potential accidental logit variables but still contains many similarities including the limited predictive consistency of life stage variables, spousal status, loyalty program engagement levels and category frequency. It also contains some discrepancies where variables in the logit output seemed more reliable; most notably these include the monetary value of the one-month timeframe as well as specific department purchase frequency variables (4, 6, 10, 21, 22, 23 and 32) and the annual aggregate spend in all food stores.

Variables that were excluded from the model because they did not add any additional predictive power are not included in the list. However, it is important to note that, unlike in Study 2, the RFM variables of frequency of transaction for the one-month timeframe did not emerge as significant in this iteration. Furthermore, the variables of recency and monetary value (both in a one-month timeframe) were inconsistently present across cross-validation fold outputs.

Given that no growth limit was imposed on the CHAID decision tree, the ultimate size of the final decision tree makes it very difficult to represent in the limited confines of this thesis. As a result, I reorganised the data in a spreadsheet format to highlight the key findings related to individual variables/parameters based on one representative decision-tree fold iteration. Results are presented by the response rates by branch level from top to bottom (Table 75, Table 76, Table 77). Much like the results from Study 2, results from this representation also show what variables (and variables values and/or combinations) specifically contribute to explaining response.

Table 75 shows the first branch level and illustrates that the more bonus points a customer had earned, the higher the actual response rate during the promotion. This is not surprising given that bonus points are a promotional currency used throughout the year to encourage minimum spend thresholds. This indicates that the more customers had developed a habit of taking advantage of promotions, such as the one being analysed, the more likely they would be to respond to similar promotions in the future.

The second branch level (Table 76) reiterates the previous finding on bonus points but further decomposes the response drivers to include the monetary value spent by customer and the number of products they purchased throughout the year. The main finding (tracing down one branch level) is that consumers who are higher bonus-mile earners and/or that spend higher amounts throughout the year are more likely to be promotional responders. The number of products purchased did allow the model to discriminate consumers but did not provide a branch level response rate above the average of total responders.

Finally, at leaf or terminating node level (Table 77), vendor points are an additional key differentiator for response, given it holds the top-10 spot (if we also include bonus transactions as a proxy) for highest response rate driver within the terminal node ranking. Other variables' relationships to response are harder to assess, given the variety of nodes and the much smaller size of each node. Other than vendor points, it is difficult to come to any conclusion at a terminal level other than to use this information to generate the appropriate ranking for group prioritisation in gains and lift charts. Nevertheless, the discrimination power of CHAID is significant and is demonstrated by the wide gap between the highest responding terminal node and the lowest. The highest terminal node's response rate is 63.2 percent while the lowest is 0.0 percent.

Table 75: CHAID Output at Branch Level 1

Branch Level 1	Branch Level 2	Leaf OR Terminating Node	Response Rate	Level
Total Responders			**11.6%**	**1**
1: miles bonus 12M (0)			1.8%	2
2: miles bonus 12M (0-2)			3.9%	2
3: miles bonus 12M (2-5)			5.9%	2
4: miles bonus 12M (5-12)			9.3%	2
5: miles bonus 12M (12-24)			14.6%	2
6: miles bonus 12M (24-42)			19.2%	2
7: miles bonus 12M (42-79)			27.6%	2
8: miles bonus 12M (80+)			46.9%	2
9: miles bonus 12M (missing)			0.0%	2

Table 76: CHAID Output at Branch Level 2

Branch Level 1	Branch Level 2	Leaf OR Terminating Node	Response Rate	Level
8: miles bonus 12M (80+)	57: Monetary full yr (7532$+)		52.7%	3
8: miles bonus 12M (80+)	56: Monetary full yr (5662-7532$)		49.9%	3
8: miles bonus 12M (80+)	55: Monetary full yr (4349-5662$)		47.9%	3
8: miles bonus 12M (80+)	54: Monetary full yr (3291-4349$)		43.4%	3
8: miles bonus 12M (80+)	53: Monetary full yr (2398-3291$)		35.8%	3
7: miles bonus 12M (42-79)	48: Monetary full yr (3291-7532$)		31.9%	3
8: miles bonus 12M (80+)	52: Monetary full yr (1669-2398$)		28.3%	3
7: miles bonus 12M (42-79)	47: Monetary full yr (2398-3291$)		27.3%	3
7: miles bonus 12M (42-79)	49: Monetary full yr (7532$+)		25.0%	3
6: miles bonus 12M (24-42)	41: Monetary full yr (2398-4349$)		24.7%	3
7: miles bonus 12M (42-79)	46: Monetary full yr (1669-2398$)		21.2%	3
6: miles bonus 12M (24-42)	42: Monetary full yr (4349-7532$)		20.8%	3
8: miles bonus 12M (80+)	51: Monetary full yr (1072-1669$)		20.8%	3
6: miles bonus 12M (24-42)	43: Monetary full yr (7532$+)		19.1%	3
5: miles bonus 12M (12-24)	33: Monetary full yr (2398-3291$)		18.0%	3
5: miles bonus 12M (12-24)	34: Monetary full yr (3291-4349$)		17.2%	3
5: miles bonus 12M (12-24)	35: Monetary full yr (4349-7532$)		15.6%	3
8: miles bonus 12M (80+)	50: Monetary full yr (0-1072) OR Missing		14.7%	3
5: miles bonus 12M (12-24)	32: Monetary full yr (1072-2398$)		14.3%	3
6: miles bonus 12M (24-42)	40: Monetary full yr (1669-2398$)		13.9%	3
7: miles bonus 12M (42-79)	45: Monetary full yr (584-1669$)		13.8%	3
5: miles bonus 12M (12-24)	36: Monetary full yr (7532$+)		13.3%	3
Total Responders			11.6%	1
4: miles bonus 12M (5-12)	29: Nb SKUs full yr (461+)		11.6%	3
4: miles bonus 12M (5-12)	28: Nb SKUs full yr (279-461)		10.6%	3
3: miles bonus 12M (2-5)	25: Nb SKUs full yr (368+)		9.8%	3
6: miles bonus 12M (24-42)	39: Monetary full yr (1072-1669$		9.4%	3
6: miles bonus 12M (24-42)	37: Monetary full yr (0-584$) OR Missing		8.0%	3
2: miles bonus 12M (0-2)	20: Monetary full yr (+2398$)		7.1%	3
4: miles bonus 12M (5-12)	27: Nb SKUs full yr (115-279)		6.2%	3
1: miles bonus 12M (0)	16: Monetary full yr (+4349$)		5.3%	3
6: miles bonus 12M (24-42)	38: Monetary full yr (584-1072$)		5.3%	3
2: miles bonus 12M (0-2)	19: Monetary full yr (1072-2398$)		4.8%	3
1: miles bonus 12M (0)	15: Monetary full yr (2398-4349$)		4.5%	3
5: miles bonus 12M (12-24)	31: Monetary full yr (584-1072$)		4.2%	3
1: miles bonus 12M (0)	14: Monetary full yr (1669-2398$)		3.9%	3
7: miles bonus 12M (42-79)	44: Monetary full yr (0-584$) OR Missing		3.7%	3
2: miles bonus 12M (0-2)	18: Monetary full yr (584-1072$)		2.9%	3
3: miles bonus 12M (2-5)	22: Nb SKUs full yr (46-195)		2.8%	3
3: miles bonus 12M (2-5)	24: Nb SKUs full yr (279-368)		2.8%	3
3: miles bonus 12M (2-5)	23: Nb SKUs full yr (195-279)		2.6%	3
1: miles bonus 12M (0)	13: Monetary full yr (1072-1669$)		1.3%	3
1: miles bonus 12M (0)	12: Monetary full yr (584-1072$)		0.9%	3
1: miles bonus 12M (0)	11: Monetary full yr (212-584$)		0.3%	3
1: miles bonus 12M (0)	10: Monetary full yr (0-212$) OR Missing		0.2%	3
2: miles bonus 12M (0-2)	17: Monetary full yr (0-584$) OR Missing		0.0%	3
3: miles bonus 12M (2-5)	21: Nb SKUs full yr (0-46) OR Missing		0.0%	3
4: miles bonus 12M (5-12)	26: Nb SKUs full yr (0-115) OR Missing		0.0%	3
5: miles bonus 12M (12-24)	30: Monetary full yr (0-584$) OR Missing		0.0%	3
9: miles bonus 12M (missing)	58: Monetary full yr (0-212$) OR Missing		0.0%	3
9: miles bonus 12M (missing)	59: Monetary full yr (212-584$)		0.0%	3
9: miles bonus 12M (missing)	60: Monetary full yr (584$+)		0.0%	3

Table 77: CHAID Output at Leaf or Terminating Node Level

Branch Level 1	Branch Level 2	Leaf OR Terminating Node	Response Rate
8: miles bonus 12M (80+)	57: Monetary full yr (7532$+)	164: Vendor Miles 12mth (161+)	62.3%
8: miles bonus 12M (80+)	56: Monetary full yr (5662-7532$)	161: Vendor Miles 12mth (161+)	56.5%
8: miles bonus 12M (80+)	55: Monetary full yr (4349-5662$)	156: Vendor Miles 12mth (0-47)	52.1%
8: miles bonus 12M (80+)	55: Monetary full yr (4349-5662$)	158: Vendor Miles 12mth (100+)	49.7%
8: miles bonus 12M (80+)	53: Monetary full yr (2398-3291$)	152: Vendor Miles 12mth (100+)	48.7%
8: miles bonus 12M (80+)	54: Monetary full yr (3291-4349$)	153: Bonus Transactions 12mth (0-78)	48.4%
8: miles bonus 12M (80+)	56: Monetary full yr (5662-7532$)	160: Vendor Miles 12mth (68-161)	47.9%
8: miles bonus 12M (80+)	57: Monetary full yr (7532$+)	163: Vendor Miles 12mth (100-161)	44.4%
8: miles bonus 12M (80+)	57: Monetary full yr (7532$+)	162: Vendor Miles 12mth (0-100)	43.7%
8: miles bonus 12M (80+)	55: Monetary full yr (4349-5662$)	157: Vendor Miles 12mth (47-100)	42.7%
8: miles bonus 12M (80+)	56: Monetary full yr (5662-7532$)	159: Vendor Miles 12mth (0-68)	42.1%
6: miles bonus 12M (24-42)	43: Monetary full yr (7532$+)	132: Other Miles (0+)	39.1%
8: miles bonus 12M (80+)	54: Monetary full yr (3291-4349$)	154: Bonus Transactions 12mth (78-143)	38.7%
8: miles bonus 12M (80+)	52: Monetary full yr (1669-2398$)	150: Vendor Miles 12mth (0-19)	36.4%
7: miles bonus 12M (42-79)	48: Monetary full yr (3291-7532$)	140: Loyalty Prog Avid (High)	35.5%
6: miles bonus 12M (24-42)	41: Monetary full yr (2398-4349$)	128: Vendor Miles 12mth (68+)	34.8%
7: miles bonus 12M (42-79)	47: Monetary full yr (2398-3291$)	138: Vendor Miles 12mth (100+)	34.7%
8: miles bonus 12M (80+)	54: Monetary full yr (3291-4349$)	155: Bonus Transactions 12mth (143+)	33.3%
8: miles bonus 12M (80+)	53: Monetary full yr (2398-3291$)	151: Vendor Miles 12mth (19-100)	29.7%
4: miles bonus 12M (5-12)	28: Nb SKUs full yr (279-461)	104: Vendor Miles 12mth (161+)	27.8%
8: miles bonus 12M (80+)	51: Monetary full yr (1072-1669$)	147: Loyalty Progr. Segment Next Best (Y)	27.8%
5: miles bonus 12M (12-24)	33: Monetary full yr (2398-3291$)	112: Vendor Miles 12mth (68+)	27.1%
7: miles bonus 12M (42-79)	48: Monetary full yr (3291-7532$)	139: Loyalty Prog Avid (Low)	26.0%
7: miles bonus 12M (42-79)	49: Monetary full yr (7532$+)	142: Other Miles (0)	25.1%
5: miles bonus 12M (12-24)	34: Monetary full yr (3291-4349$)	114: Dept 2 Frequency (4-6)	24.7%
7: miles bonus 12M (42-79)	49: Monetary full yr (7532$+)	143: Other Miles (0+)	24.2%
6: miles bonus 12M (24-42)	42: Monetary full yr (4349-7532$)	130: Family Lifestage (N)	23.8%
7: miles bonus 12M (42-79)	45: Monetary full yr (584-1669$)	134: Widowed (0.07+)	23.8%
7: miles bonus 12M (42-79)	47: Monetary full yr (2398-3291$)	137: Vendor Miles 12mth (0-100)	23.3%
7: miles bonus 12M (42-79)	48: Monetary full yr (3291-7532$)	141: Loyalty Prog Avid (Med)	23.1%
7: miles bonus 12M (42-79)	46: Monetary full yr (1669-2398$)	135: Loyalty Prog Avid (High)	22.6%
8: miles bonus 12M (80+)	52: Monetary full yr (1669-2398$)	151: Vendor Miles 12mth (19-100)	22.6%
5: miles bonus 12M (12-24)	35: Monetary full yr (4349-7532$)	118: Vendor Miles 12mth (161+)	21.9%
8: miles bonus 12M (80+)	50: Monetary full yr (0-1072)	145: Gender (Female)	21.4%
5: miles bonus 12M (12-24)	34: Monetary full yr (3291-4349$)	115: Dept 2 Frequency (6+)	19.7%
6: miles bonus 12M (24-42)	41: Monetary full yr (2398-4349$)	127: Vendor Miles 12mth (0-68)	19.3%
5: miles bonus 12M (12-24)	36: Monetary full yr (7532$+)	119: Transaction 12mth (0-143)	18.1%
5: miles bonus 12M (12-24)	35: Monetary full yr (4349-7532$)	117: Vendor Miles 12mth (100-161)	17.8%
6: miles bonus 12M (24-42)	42: Monetary full yr (4349-7532$)	129: Family Lifestage (Y)	17.4%
6: miles bonus 12M (24-42)	43: Monetary full yr (7532$+)	131: Other Miles (0)	17.2%
8: miles bonus 12M (80+)	53: Monetary full yr (2398-3291$)	150: Vendor Miles 12mth (0-19)	16.7%
7: miles bonus 12M (42-79)	46: Monetary full yr (1669-2398$)	136: Loyalty Prog Avid (Low)	16.0%
6: miles bonus 12M (24-42)	40: Monetary full yr (1669-2398$)	125: Monetary Value Pre Promo (0-262) OR Missing	15.0%
5: miles bonus 12M (12-24)	33: Monetary full yr (2398-3291$)	111: Vendor Miles 12mth (0-68)	14.8%
5: miles bonus 12M (12-24)	32: Monetary full yr (1072-2398$)	109: Gold Loyalty Status (N)	14.4%
5: miles bonus 12M (12-24)	32: Monetary full yr (1072-2398$)	110: Gold Loyalty Status (Y)	14.0%
5: miles bonus 12M (12-24)	35: Monetary full yr (4349-7532$)	116: Vendor Miles 12mth (0-100)	13.7%
4: miles bonus 12M (5-12)	29: Nb SKUs full yr (461+)	105: Transactions 12mth (0-143)	12.2%
Total Responders			11.6%
3: miles bonus 12M (2-5)	25: Nb SKUs full yr (368+)	93: Other Miles (+0)	11.4%
4: miles bonus 12M (5-12)	28: Nb SKUs full yr (279-461)	103: Vendor Miles 12mth (68-161)	11.4%
7: miles bonus 12M (42-79)	45: Monetary full yr (584-1669$)	133: Widowed (0-0.07) OR Missing	10.6%
6: miles bonus 12M (24-42)	39: Monetary full yr (1072-1669$	123: Other Miles (0)	10.2%
5: miles bonus 12M (12-24)	34: Monetary full yr (3291-4349$)	113: Dept 2 Frequency (0-4)	10.0%
8: miles bonus 12M (80+)	50: Monetary full yr (0-1072)	144: Gender (Male or unknown)	10.0%
3: miles bonus 12M (2-5)	25: Nb SKUs full yr (368+)	92: Other Miles (0)	9.7%
4: miles bonus 12M (5-12)	28: Nb SKUs full yr (279-461)	102: Vendor Miles 12mth (0-68)	9.6%
4: miles bonus 12M (5-12)	29: Nb SKUs full yr (461+)	106: Transactions 12mth (143+)	9.3%
2: miles bonus 12M (0-2)	20: Monetary full yr (+2398$)	83: Gold Loyalty Status (N)	8.5%
4: miles bonus 12M (5-12)	28: Nb SKUs full yr (279-461)	101: Vendor Miles 12mth (0)	8.3%
6: miles bonus 12M (24-42)	40: Monetary full yr (1669-2398$)	126: Monetary Value Pre Promo (262+)	8.3%
4: miles bonus 12M (5-12)	27: Nb SKUs full yr (115-279)	100: Dept Frequency (3+)	7.6%
5: miles bonus 12M (12-24)	36: Monetary full yr (7532$+)	120: Transactions 12mth (143+)	6.7%
3: miles bonus 12M (2-5)	23: Nb SKUs full yr (195-279)	88: Dept 7 Frequency (0-1)	6.5%
4: miles bonus 12M (5-12)	27: Nb SKUs full yr (115-279)	98: Dept 4 Frequency (0-2)	6.4%
6: miles bonus 12M (24-42)	38: Monetary full yr (584-1072$)	121: HH income (under 35K OR 75K+)	6.1%
2: miles bonus 12M (0-2)	19: Monetary full yr (1072-2398$)	81: Vendor Miles 12mth (0-47)	5.6%
1: miles bonus 12M (0)	16: Monetary full yr (+4349$)	76: Age 0-17 (0.19+) OR Missing	5.3%
1: miles bonus 12M (0)	16: Monetary full yr (+4349$)	75: Age 0-17 (0-0.19)	5.1%
1: miles bonus 12M (0)	14: Monetary full yr (1669-2398$)	72: Gold Loyalty Status (Y)	4.8%
1: miles bonus 12M (0)	15: Monetary full yr (2398-4349$)	74: Emailable (Y)	4.7%

Branch Level 1	Branch Level 2	Leaf OR Terminating Node	Response Rate
5: miles bonus 12M (12-24)	31: Monetary full yr (584-1072$)	108: Age 45-54 (0.12+)	4.5%
1: miles bonus 12M (0)	15: Monetary full yr (2398-4349$)	73: Emailable (N)	4.4%
1: miles bonus 12M (0)	14: Monetary full yr (1669-2398$)	71: Gold Loyalty Status (N)	3.6%
2: miles bonus 12M (0-2)	18: Monetary full yr (584-1072$)	79: Dept 23 Frequency (0)	3.0%
2: miles bonus 12M (0-2)	20: Monetary full yr (+2398$)	84: Gold Loyalty Status (Y)	3.0%
3: miles bonus 12M (2-5)	24: Nb SKUs full yr (279-368)	90: Web active 12mth (N)	3.0%
3: miles bonus 12M (2-5)	22: Nb SKUs full yr (46-195)	85: Dept 12 Frequency (0)	2.8%
3: miles bonus 12M (2-5)	24: Nb SKUs full yr (279-368)	91: Web active 12mth (Y)	2.7%
4: miles bonus 12M (5-12)	27: Nb SKUs full yr (115-279)	99: Dept 4 Frequency (2-3)	2.1%
1: miles bonus 12M (0)	12: Monetary full yr (584-1072$)	66: Bonus Miles 12mth (20-64)	1.9%
1: miles bonus 12M (0)	13: Monetary full yr (1072-1669$)	69: Loyalty Prog Avid (High)	1.9%
1: miles bonus 12M (0)	13: Monetary full yr (1072-1669$)	70: Loyalty Prog Avid (Low)	1.9%
3: miles bonus 12M (2-5)	23: Nb SKUs full yr (195-279)	87: Dept 7 Frequency (0)	1.6%
1: miles bonus 12M (0)	11: Monetary full yr (212-584$)	64: Web active 12mth (Y)	0.6%
1: miles bonus 12M (0)	10: Monetary full yr (0-212$)	61: Dept freq (0)	0.2%
1: miles bonus 12M (0)	12: Monetary full yr (584-1072$)	65: Bonus Miles 12mth (0-20)	0.2%
1: miles bonus 12M (0)	10: Monetary full yr (0-212$)	62: Dept freq (+0)	0.0%
1: miles bonus 12M (0)	11: Monetary full yr (212-584$)	63: Web active 12mth (N)	0.0%
1: miles bonus 12M (0)	12: Monetary full yr (584-1072$)	67: Bonus Miles 12mth (64-96)	0.0%
1: miles bonus 12M (0)	12: Monetary full yr (584-1072$)	68: Bonus Miles 12mth (96+)	0.0%
2: miles bonus 12M (0-2)	17: Monetary full yr (0-584$)	77: Business customer (N)	0.0%
2: miles bonus 12M (0-2)	17: Monetary full yr (0-584$)	78: Business customer (Y)	0.0%
2: miles bonus 12M (0-2)	18: Monetary full yr (584-1072$)	80: Dept 23 Frequency (0+)	0.0%
2: miles bonus 12M (0-2)	19: Monetary full yr (1072-2398$)	82: Vendor Miles 12mth (47+)	0.0%
3: miles bonus 12M (2-5)	22: Nb SKUs full yr (46-195)	86: Dept 12 Frequency (+0)	0.0%
3: miles bonus 12M (2-5)	23: Nb SKUs full yr (195-279)	89: Dept 7 Frequency (+1)	0.0%
4: miles bonus 12M (5-12)	26: Nb SKUs full yr (0-115)	94: Vendor Miles 12mth (0-10)	0.0%
4: miles bonus 12M (5-12)	26: Nb SKUs full yr (0-115)	95: Vendor Miles 12mth (10-19)	0.0%
4: miles bonus 12M (5-12)	26: Nb SKUs full yr (0-115)	96: Vendor Miles 12mth (19-47)	0.0%
4: miles bonus 12M (5-12)	26: Nb SKUs full yr (0-115)	97: Vendor Miles 12mth (47+)	0.0%
5: miles bonus 12M (12-24)	31: Monetary full yr (584-1072$)	107: Age 45-54 (0-0.12)	0.0%
6: miles bonus 12M (24-42)	38: Monetary full yr (584-1072$)	122: HH income (35K-75K)	0.0%
6: miles bonus 12M (24-42)	39: Monetary full yr (1072-1669$	124: Other Miles (0+)	0.0%
8: miles bonus 12M (80+)	51: Monetary full yr (1072-1669$)	146: Loyalty Progr. Segment Next Best (N)	0.0%
9: miles bonus 12M (missing)	59: Monetary full yr (212-584$)	165: HH Size 2 (0-0.41) OR Missing	0.0%
9: miles bonus 12M (missing)	59: Monetary full yr (212-584$)	166: HH Size 2 (0.41+)	0.0%

The risk statistics and classification results illustrate how well the model classifies predicted cases. Table 78 indicates that the category predicted by the model (response) is wrong in 11.2 percent to 12.2 percent of cases. Results in the classification table (Table 79) are consistent with the risk estimate, with data indicating that the model classifies approximately between 87.8 percent and 88.8 percent of customers correctly. This table showcases that the overall percentage of cases that are correctly classified by the CHAID model is relatively high both for the test and training set.

Table 78: Risk Statistics

Sample	Estimate		Std. Error	
	min	max	min	max
Training	.117	.118	.001	.001
Test	.112	.122	.002	.002

Growing Method:
Dependent Variable: PROMO_RESPONDER

Table 79: Classification

	Overall Percentage Correctly Classified	
	min	max
Training	88.2%	88.3%
Test	87.8%	88.8%

Growing Method: CHAID
Dependent Variable: PROMO_RESPONDER

6.3.2.3 Neural Networks

As shown in Table 80, the final structure of the neural network was composed of 76 units in the input layer (all available variables) and one hidden layer comprised of six to 12 'hidden' units in the hidden layer (depending on the fold iteration used). The one hidden layer is typical of multi-layer perceptron applications while the hidden layers are a function of the model’s variable interconnections. Naturally, given the greater number of covariates (variables) in the input layer (76) as compared to the RFM model (3), the number of hidden layers increases in this case.

Table 80: ANN MLP Network Information

Network Information				
Input Layer	Covariates	1		all variables
		2		
		3		
		...		
		...		
		76		
		Number of Units[a]		134
		Rescaling Method for Covariates		Standardised
Hidden Layer(s)		Number of Hidden Layers		1
		Number of Units in Hidden Layer 1[a]	min	6
			max	12
		Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables			PROMO_RESPONDER
	Number of Units			2
	Activation Function			Softmax
	Error Function			Cross-entropy

Table 81: ANN MLP Model Summary

Model Summary		min	max
Training	Cross Entropy Error	11810	12406.750
	Percent Incorrect Predictions	14.8%	15.5%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error	
	Training Time	00:00:05.811	
Testing	Cross Entropy Error	28171	28764.772
	Percent Incorrect Predictions	15.1%	15.4%
Holdout	Percent Incorrect Predictions	14.7%	15.7%

Dependent Variable: PROMO_RESPONDER

a.Error computations are based on the testing sample.

In observing the error and misclassification (1-PCC) statistics of the neural network output and comparing it to the RFM variable iteration, the extended data seems to moderately increase the PCC rate versus the RFM variable set. The cross-entropy error provides a similar observation and seems to indicate a stronger reliability of the extended variable model.

6.4 Discussion

This second quantitative analysis sheds some light on the relative performance of techniques sourced from the systematic review using extended FMCG retail variables. As mentioned earlier, the thesis aims to address two gaps: (1) the relationship between customer selection techniques and performance and (2) the impact of dimensionality reduction on performance. This section will discuss findings as they related to these gaps in order with the first portion covering technique performance, while the second will address variable set and dimensionality reduction.

6.4.1 Overview of Technique Performance

Advanced statistical and machine-learning techniques (for top deciles one and two) perform much better than RFM for identifying high-response segments. This significant positive variation in top-two decile performance allows all techniques to outperform RFM well into the fifth decile. This result supports observations made in the systematic review, and provides evidence that not only do advanced statistical and machine-learning techniques outperform RFM, but they also do so with a much greater margin when extended variables are applied. Results support findings from empirical research by Magidson (1988), Levin and Zahavi (2001), Yang (2004) and McCarty (2006).

Unlike the results in Study 2, Study 3 predictive accuracy and fit results do support Olson and Chae's (2012) observations that RFM methods are less accurate and that Data Mining algorithms yield better prediction accuracy and cumulative gains. However, the outputs from advanced models also illustrate that both the accuracy and performance of models can be significantly improved across top deciles, not only "slightly improved" by increasing the cut-off level as stated by Olson and Chae (2012, p. 75).

Moving to the comparison of performance between techniques, a detailed review of technique effectiveness accompanied by the validation/invalidation of propositions is presented in Table 82.

Table 82: Technique Effectiveness by Proposition for Extended Variable Sets

Proposition	Validation	Description
P13: RFM is more effective than CHAID-DD	Invalidated	• CHAID-DD performs better than RFM at an individual level for deciles 1 and 2. From deciles 3 on, RFM lift is greater than CHAID-DD. • Cumulatively, the lift in top deciles 1 and 2 allows CHAID-DD to have a cumulative lift/gain greater than RFM until deciles 7 and on, where performance becomes similar • The Gini coefficient of CHAID-DD is much greater than RFM, illustrating not only a greater performance at top deciles but a greater fit overall as well.
P14: RFM is more effective than Stepwise Logistic Regression-DD	Invalidated	• Logistic Regression-DD performs better than RFM at an individual level for deciles 1 and 2. From deciles 3 on, RFM lift is greater than Logistic Regression-DD. • Cumulatively, the lift in top deciles 1 and 2 allows Logistic Regression-DD to have a cumulative lift/gain greater than RFM until deciles 7 and on, where performance becomes similar • The Gini coefficient of Logistic Regression-DD is much greater than RFM, illustrating not only a greater performance at top deciles but a greater fit overall
P15: RFM is more effective than NN-DD	Invalidated	• NN-RDD performs better than RFM at an individual level for deciles 1 and 2. From deciles 3 on, RFM lift is greater than NN-RDD. • Cumulatively, the lift in top deciles 1 and 2 allows NN-RDD to have a cumulative lift/gain greater than RFM until deciles 7 and on, where performance is similar • The Gini coefficient of NN-RDD is much greater than RFM, illustrating not only a greater performance at top deciles but a greater fit overall as well.
P16: CHAID-DD is more effective than RFM	Validated	• CHAID-DD performs better than RFM at an individual level for deciles 1 and 2. From deciles 3 on, RFM lift is greater than CHAID-DD. • Cumulatively, the lift in top deciles 1 and 2 allows CHAID-DD to have a cumulative lift/gain greater than RFM until deciles 7 and on, where performance becomes similar • The Gini coefficient of CHAID-DD is much greater than RFM, illustrating not only a greater performance at top deciles but a greater fit overall as well.
P17: CHAID-DD is more effective than Stepwise Logistic Regression-DD	Invalidated	• Logistic Regression-DD performs better than CHAID-DD for decile 1. CHAID lift is stronger for deciles 2, 3 and 4. Both technique perform equally well for decile 5 and performance varies for subsequent deciles. • Cumulatively, the lift in decile 1 allows Logistic Regression-DD to have a cumulative lift/gain greater than CHAID-DD until decile 3, at which point performance becomes similar • The Gini coefficient of CHAID-DD is greater than the coefficient for logistic Regression-DD, thus demonstrating a slightly better fit

Proposition	Validation	Description
P18: CHAID-DD is more effective than NN-DD	Invalidated	• NN-DD performs better than CHAID-DD for decile 1. Both techniques perform similarly for decile 2. CHAID lift is stronger for deciles 3 and 4. Both technique perform equally well for decile 5 and are fairly similar for subsequent deciles • Cumulatively, the lift in decile 1 allows NN-DD to have a cumulative lift/gain greater than CHAID-DD until decile 3, at which point performance becomes similar • The Gini coefficient of CHAID-DD is greater than the coefficient for NN-DD, thus demonstrating a slightly better fit
P19 Stepwise Logistic Regression-DD is more effective than RFM-DD	Validated	• Logistic Regression-DD performs better than RFM at an individual level for deciles 1 and 2. From deciles 3 on, RFM lift is greater than Logistic Regression-DD. • Cumulatively, the lift in top deciles 1 and 2 allows Logistic Regression-DD to have a cumulative lift/gain greater than RFM until deciles 7 and on, where performance becomes similar • The Gini coefficient of Logistic Regression-DD is much greater than RFM, illustrating not only a greater performance at top deciles but a greater fit
P20: Stepwise Logistic Regression-DD is more effective than CHAID-DD	Validated	• Logistic Regression-DD performs better than CHAID-DD for decile 1. CHAID lift is stronger for deciles 2, 3 and 4. Both technique perform equally well for decile 5 and performance varies for subsequent deciles. • Cumulatively, the lift in decile 1 allows Logistic Regression-DD to have a cumulative lift/gain greater than CHAID-DD until decile 3, at which point performance becomes similar • The Gini coefficient of CHAID-DD is greater than the coefficient for Logistic Regression-DD, thus demonstrating a slightly better fit
P21: Stepwise Logistic Regression-DD is more effective than NN-DD	Invalidated	• Both techniques demonstrate very similar lift indices across top deciles with the slight exception of decile 2 where NN-DD performs slightly better and decile 3. All other deciles can be qualified as being fairly similar with a little exception in favour of NN-DD in decile 7. • Cumulatively, the lift proximity of top deciles 1 and 2, necessarily lead both techniques to perform equally well on all deciles (with never more than 1-2 points in cumulative lift difference). • The Gini coefficient of NN-DD is only slightly greater than that of Logistic regression-DD
P22: NN-DD is more effective than RFM-DD	Validated	• Logistic Regression-DD performs better than RFM at an individual level for deciles 1 and 2. From deciles 3 on, RFM lift is greater than Logistic Regression-DD. • Cumulatively, the lift in top deciles 1 and 2 allows Logistic Regression-DD to have a cumulative lift/gain greater than RFM until deciles 7 and on, where performance becomes similar • The Gini coefficient of Logistic Regression-DD is much greater than RFM, illustrating not only a greater performance at top deciles but a greater fit overall as well.

Proposition	Validation	Description
P23: NN-DD is more effective than CHAID-DD	Validated	• NN-DD performs better than CHAID-DD for decile 1. Both techniques perform similarly for decile 2. CHAID lift is stronger for deciles 3 and 4. Both technique perform equally well for decile 5 and are fairly similar for subsequent deciles • Cumulatively, the lift in decile 1 allows NN-DD to have a cumulative lift/gain greater than CHAID-DD until decile 3, at which point performance becomes similar • The Gini coefficient of CHAID-DD is greater than the coefficient for NN-DD, thus demonstrating a slightly better fit
P24: NN-DD is more effective than Stepwise Logistic Regression-DD	Invalidated	• Both techniques demonstrate very similar lift indices across top deciles with the slight exception of decile 2 where NN-DD performs slightly better and decile 3. All other deciles can be qualified as being fairly similar with a little exception in favour of NN-DD in decile 7. • Cumulatively, the lift proximity of top deciles 1 and 2, necessarily lead both techniques to perform equally well on all deciles (with never more than 1-2 points in cumulative lift difference). • The Gini coefficient of NN-DD is only slightly greater than that of Logistic regression-DD

From an advanced technique perspective, both logistic regression and ANN outperform CHAID at the top decile while, at an individual decile level, CHAID and ANN perform better at decile two, and CHAID continues outperforming for remaining top deciles three to five. From a cumulative perspective, the strong performance of logit and ANN allow these techniques to outperform in deciles one and two, but as of decile three cumulative performance of all three advanced techniques becomes virtually identical.

Comparing these results with the lift and cumulative lift and gains charts of Study 2 (Figures 33, 34 and 35), it is immediately noticeable that the combination of expanded data and advanced technique usage generates significant individual decile gains for deciles one and two. This indicates a very strong capacity of CHAID, ANN and logit models to classify top responders. The power of these models is indeed so strong at the top-two individual decile levels that the cumulative lifts and gains persist well into the bottom-half deciles. Given the controlled dataset and reuse of cross-validation folds, improvements in performance and fit are solely due to the combined effects of applying advanced techniques to an extended data variable set. This further strengthens the relationship between performance and incremental data highlighted by Reutterer et al. (2006), Malthouse (2006) and Greene (2008).

The somewhat similar performance of logistic regression and ANN is corroborated in work by Zahavi and Levin (1998). Zahavi and Levin's study showed that both techniques achieve similar performance and fit levels but that technique interpretation was simpler for logistic regression. However, the finding is contrary to numerous other studies that found ANN to perform better across top deciles than logistic regression. Studies by West et al. (1997), Cui et al. (2008), Guido (2011) and Ho et al. (2008) all find ANN provides a higher hit-ratio (versus a control group) than logistic regression. Linder et al. (2004)

for their part found that ANN performed best when the sample size was small, while logistic regression performed best when sample size was large.

As it pertains to ANN overall and consistent with previous literature, neural networks provide a stronger performance over tree-based techniques such as CART (Cui and Wong, 2004; Cui et al., 2006; Cui et al., 2008) as well as CHAID (West et al., 1997; Linder et al., 2004). Linder et al.'s study, in which CHAID performed well, actually shows that not only CHAID's performance but the "performance of all methods increased with the size of the customer base" (Linder et al., 2004, p. 344), but that "this relation was less strong for Neural Networks than for CTs (classification trees) and LR (logistic regression), especially when data complexity was high. As a consequence ANNs outperformed the other methods when sample size was small, but CTs and LR yielded better results when sample size was large — with LR being generally superior to CTs" (Linder et al., 2004, p. 344). Considering the sizable dataset utilised for this study, and the ANN performance, this challenges the stronger performance of CHAID vis-à-vis Neural Networks. However, it does support Linder et al.'s claim that, in such a context, logistic regression yields generally superior results. When discussing relative performance of ANNs to traditional statistical methods, Blattberg et al. (2008) indicate that performance depends on the types of data available and their applications. In the case of ANN, they indicate that the technique has an intrinsic edge in modelling highly non-linear relationships and when significant interactions among independent variables exist. In cases where relationships are more linear, linear regression performs better because "the linearity assumption (of logit) functions like additional prior information. On the other hand, when the true relationship is complex, ANNs may outperform linear (or logistic) regression because the wrong assumptions made in linear (logistic) regression will bias its results" (Blattberg et al., 2008, pp. 446-447). The increase in relative performance of ANN between Study 2 and Study 3 (specifically for decile one) may indeed indicate that this is the case.

Some studies found that both CHAID and logit performed equally well (Magidson, 1988; Yang, 2004), findings from Study 3 illustrate that logit outperforms CHAID at the top decile and at the two cumulative deciles (Levin and Zahavi, 2001; Linder et al., 2004) the findings do not support the view by McCarty (2006) whereby CHAID performs better than logit. One of the reasons cited by Levin and Zahavi (2001, p. 9) on the potentially lower performance of CHAID versus logistic regression is that "increasing the minimum size constraint usually increases the variance of the fit results across all segments of a tree. This is related to the fact that larger segments are less 'homogeneous' and thus exhibit larger variations." This finding is not supported as both studies use a sizeable dataset with a consistently higher Gini coefficient score for CHAID versus ANN and logit. However, Levin and Zahavi's study was conducted in a very different contextual setting: the collectible industry, which has unique characteristics that link consumer purchases to variables such as product affinity scores, recency and frequency. Levin and Zahavi note that their study may not extend easily to other industries. In fact, Levin and Zahavi's study was the only one that showed that adding predictors to the process did not improve model fit and contradicts most studies including those by Greene (2008), Reutterer et al. (2006) and Malthouse (2006).

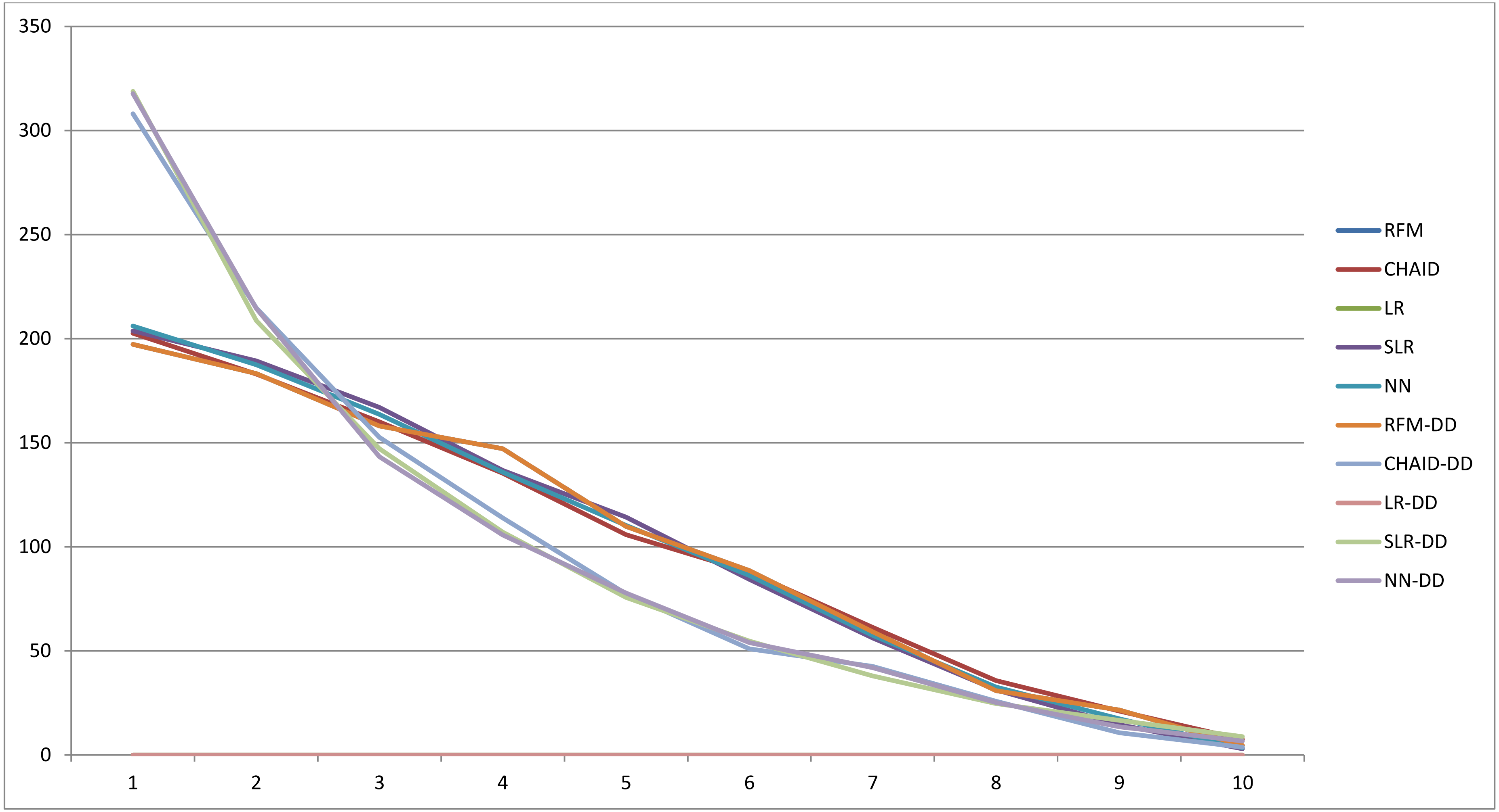


Figure 33: Lift Chart - All Techniques

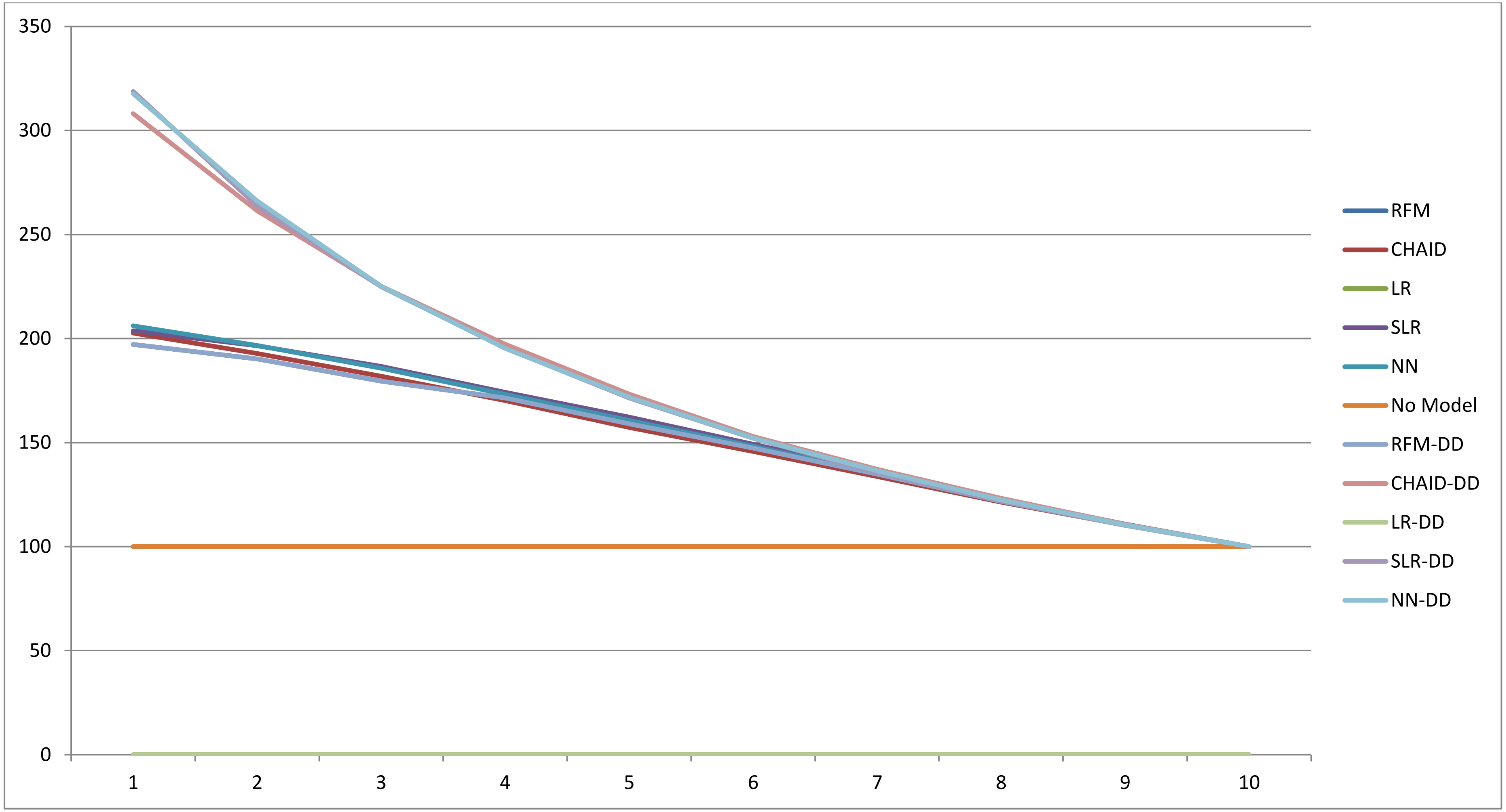


Figure 34: Cumulative Lift Chart - All Techniques

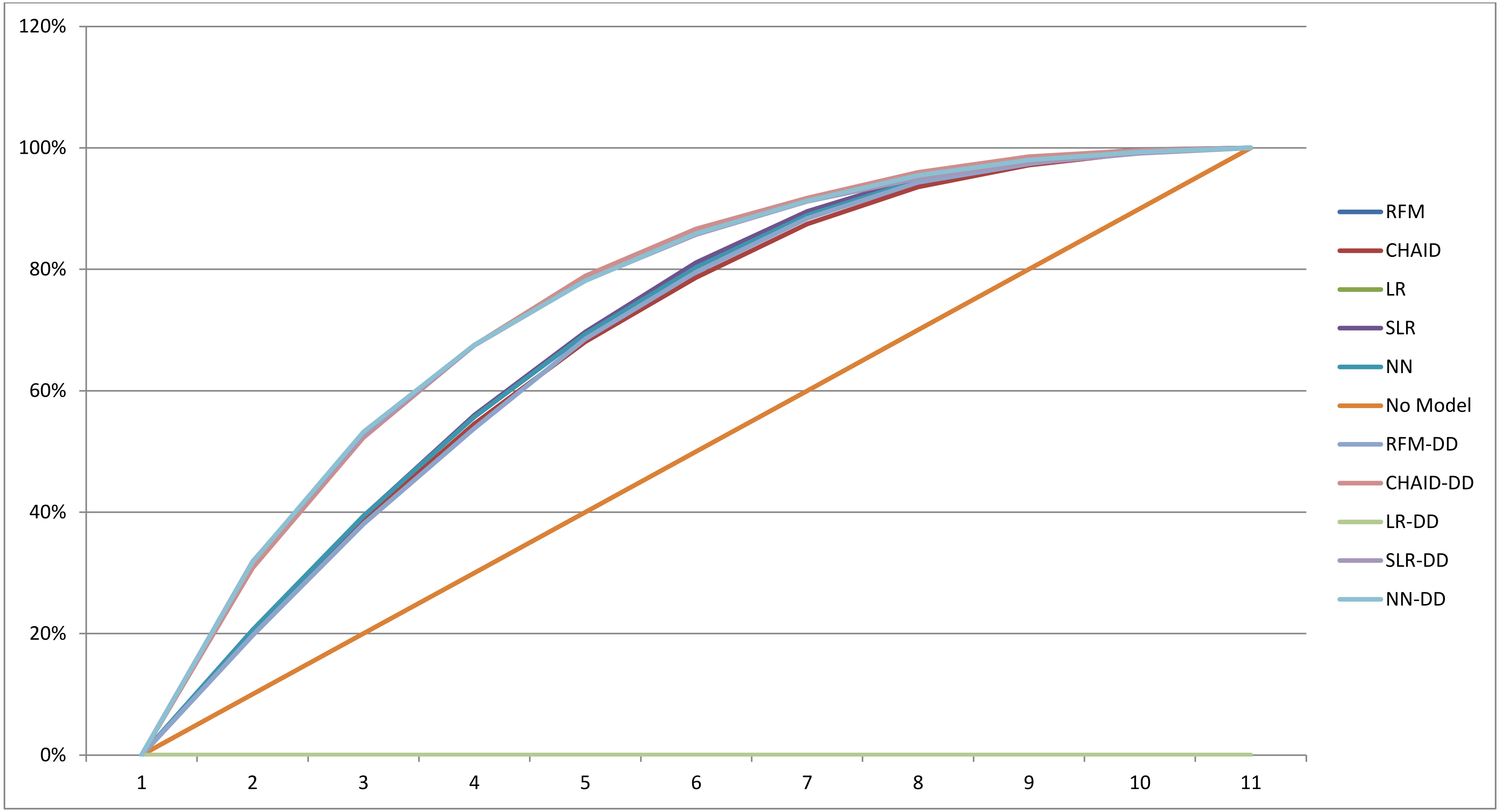


Figure 35: Cumulative Gains Chart - All Techniques

In addition to measures of performance, measures of fit are also compared in Table 83, which shows the comparison between results from Study 2 and Study 3. In both studies, CHAID is the technique that provides the most significant model fit both from a model discrimination and fit perspective (Gini coefficient) as well as predictive accuracy (PCC). However, all techniques (with the exception of the baseline RFM) show a significant increase in the Gini performance, indicating the important contribution of an expanded dataset to the reliability and fit of models. However, though discrimination is significantly increased, the percentage correctly classified statistic shows only moderate improvement: a 0.3-percent point improvement for ANN, 0.8-percent point for logistic regression and 1-percent point for CHAID.

Another significant finding is the changing order of performance ranking of the individual techniques. Though CHAID remains the technique with the best fit performance, RFM falls to last once an expanded variable set is introduced. Furthermore, both logistic and ANN performance increases significantly, and though they remain very close from an overall Gini and PCC score perspective, logistic regression outperforms ANN when using expanded data. Unlike Study 2 where accuracy of RFM was second only to CHAID, results from Study 3 indicate that additional variables lead advanced techniques to provide the expected increase of advanced model accuracy. Furthermore, unlike Study 2, Study 3 shows a much greater degree of alignment between techniques' predictive accuracy and performance, thus aligning the practitioner and academic interests of generating performance gains that are not realised at the expense of model stability.

Table 83: Comparative Model Gini Coefficients and PCC

RFM Variables				**Expanded FMCG Retail Variables**		
Model & Rank	**Gini Coeff**	**PCC**		**Model & Rank**	**Gini Coeff**	**PCC**
1. CHAID	0.411492	87.2%		1. CHAID	0.521057	88.2%
2. RFM	0.403871	NA		2. ANN	0.500625	84.7%
3. LR	0.397962	84.0%		3. LR	0.49701	84.8%
4. ANN	0.392185	84.5%		4. RFM	0.403871	NA

The Gini and PCC score differences between Studies 2 and 3 are also consistent with the additional measures of risk specific to the ANN and logit techniques. In the case of ANN, as mentioned in Study 2, cross-entropy measures the rate of error classification with smaller value indicating a lower error rate (Hopfield, 1987; Bishop, 1995). Comparing the cross-entropy results of the extended variable set with those of the RFM variables of Study 2 shows a reduced error rate in Study 3. This is consistent with the lower percentage of incorrect predictions.

Table 84 shows the estimates parameter of the logit models for both extended and RFM variable sets. Under the extended data depth logit model, both the Cox and Snell R Square and Nagelkerke R Square confirm that the enhanced data models explain a greater proportion of responses versus the

RFM variable-only applications.

Table 84: R^2 Measure

R^2 Measure	RFM		Extended Data		Average	
	Min	Max	Min	Max	RFM	Extended
Cox & Snell R Square	.059	.061	0.136	0.140	0.059	0.138
Nagelkerke R Square	.102	.105	0.233	0.239	0.103	0.236

6.4.2 Data Variables and Dimensionality Reduction

From a variable perspective, the addition of variables resulted in an increased performance and fit of all advanced techniques versus the simple RFM. This is evidenced by the fact that numerous variables from the total that were originally input into all models were deemed to be significant across techniques and thus contributed to increase the predictive strength of techniques.

In Study 2, frequency and monetary value were strong predictors of response across techniques, while recency was used (and significant) in RFM, CHAID and ANN applications. In the case of more extended data sets, a much greater number of variables emerged. In fact, 53 (out of the total of 85) retained variables emerged as significant in six or more model iterations (folds) of the logit application while a smaller amount emerged in one or two models. These latter variables are those that Field (2005) qualifies as accidental and are one of the weaknesses of applying an exploratory forward stepwise regression. This means that, although a number of variables emerged 'accidentally,' the elimination (or reduction) of 32 or more variables in each iteration of the logit application resulted in increasing model strength. This demonstrates the critical importance of dimensionality reduction in complementing technique applications in order to optimise model performance. Variables that were the most impactful included the number of bonus points accumulated by the customer in the last 12 months and the number of points accumulated on product-specific offers (provided by CPG vendors) over the same period. These variables are at least three to 10 times more significant than other variables based on the differential between average Wald statistic scores. These were followed by a clustering of other loyalty and transactional variables to complete the top-10 list of most significant variables. As it pertains to socio-demographic variables, the most significant variables among the retained groups all consist of household size, life stage or gender variables. The list of salient variables was completed by an assortment of other loyalty and transactional variables.

CHAID-DD output also shows that dimensionality reduction is important with a minimum of 25 of the 85 variables not being included in model iterations. Furthermore, as with the logit output, the technique outputs show that the number of bonus points that customers had earned in the past year was one of the stronger predictors followed by monetary spend and vendor product point accumulation. Across techniques, variables from socio-demographic, loyalty, channel and transactional dimensions also emerged as contributing to the prediction power of the model.

In comparing these categories of significant variables to Wedel and Kamakura's (2000) evaluation of segmentation bases responsiveness criteria, a number of interesting findings emerge. Firstly, the high discrimination power of transactional, channel and loyalty variables supports the power of what Wedel and Kamakura (2000) call purchase and usage variables. This is also consistent with the findings of Zahay, Peltier, Schultz, and Griffin (2004) and Yankelovich (2006). However, the significant number of socio-demographic variables, such as household size, gender, income category, among others, that did add to the model strength was unexpected and is not supported by Wedel and Kamakura's prescriptive guidance on variable categories showing increased responsiveness (though it is mentioned that contextual effects may exist). Although these variables were present across techniques, it remains that their usage count (and subsequent reliability) was lower than variables under the transactional categories. This indicates that these variables may be less reliable and often be due to sheer random error (what Field (2005) refers to as accidental inclusion into models due to the expansiveness of the dataset). Examples of low inclusion variables include life stages, marital status, loyalty program engagement, and frequency of consumption of specific departments.

Some variables in low responsiveness variable segments showcased by Wedel and Kamakura (2000) (general, observable and life style segments) did have a consistent and recurring role in increased response. Among variables that emerged both in the logit and CHAID outputs are: gender, household sizes of two persons, university level of education (six counts in logit and five in CHAID). It is difficult given the inherently different methods of variable prioritisation within each technique application to unequivocally state that these variables will constantly predict response across model applications. However, this research does provide evidence that (in addition to all the other variables that are issued in high counts from these less predictive categories) there are sub-segments of variable categories that contribute to predicting response.

Without trying to reconcile the findings with the Wedel and Kamakura (2000) segments, based on the ranking, count of significant variables and presence in model count of individual significant variables from the logit output, variable category strength for predicting promotional response in this research can be represented as follows in Table 85.

Table 85: Variable Category Strength

Variable Category	**Variable Category Strength**
Socio-Demographic	+/-
Loyalty	++
Channel	-
Transactional/Usage	++

++ very good + good +/- moderate - poor -- very poor

Another finding that emerges from the dimensionality reduction process is the fact that categories and variables are not predictive solely in a vacuum. In many cases (whether accidental random occurrences or not), variables and variable combinations contribute to increase the predictive power of models (-2LL, pseudo R^2 measures and cross-entropy) while variable collinearity between two individually strong variables explaining the same phenomenon may lead to exclusion. Said otherwise, some variables may be close substitutes of one another and, as a result, they may not both need to be included (even though they are both strong candidates for inclusion). The same can be said for variable combinations where two less explanatory variables effectively substitute a very strong explanatory variable.

This is evident in the cases of logistic regression and CHAID given the techniques' overt optimisation, but the existence of the multiple hidden layers in ANN also provides evidence that combinatorial effects exist. As a result, though variable categories may be an effective way to illustrate what variable could be included in a prescriptive fashion, it is not a good indicator for researchers of what combination of variables may result in the best a priori outcome (since ultimately, model strength is driven by the comprehensive contribution of all variables and not only individually strong variables). This indicates that an opportunity exists to further expand research to include the relationships that may exist between variable categories (versus looking solely at individual variables and variable categories in isolation). This could help provide a stronger degree of prescriptive direction to both academics and practitioners.

The findings also address the literature gap highlighted by Kumar (2006) on the lack of documented effects of using extended data and variable sets on performance and supports the emerging literature on the topic by Greene and Greene (2008), Reutterer et al. (2006) and Malthouse (2006) in which a relationship exists between performance and the use of extended data variables (hereafter referred to as depth of data for contrast purposes with shallow data sets).

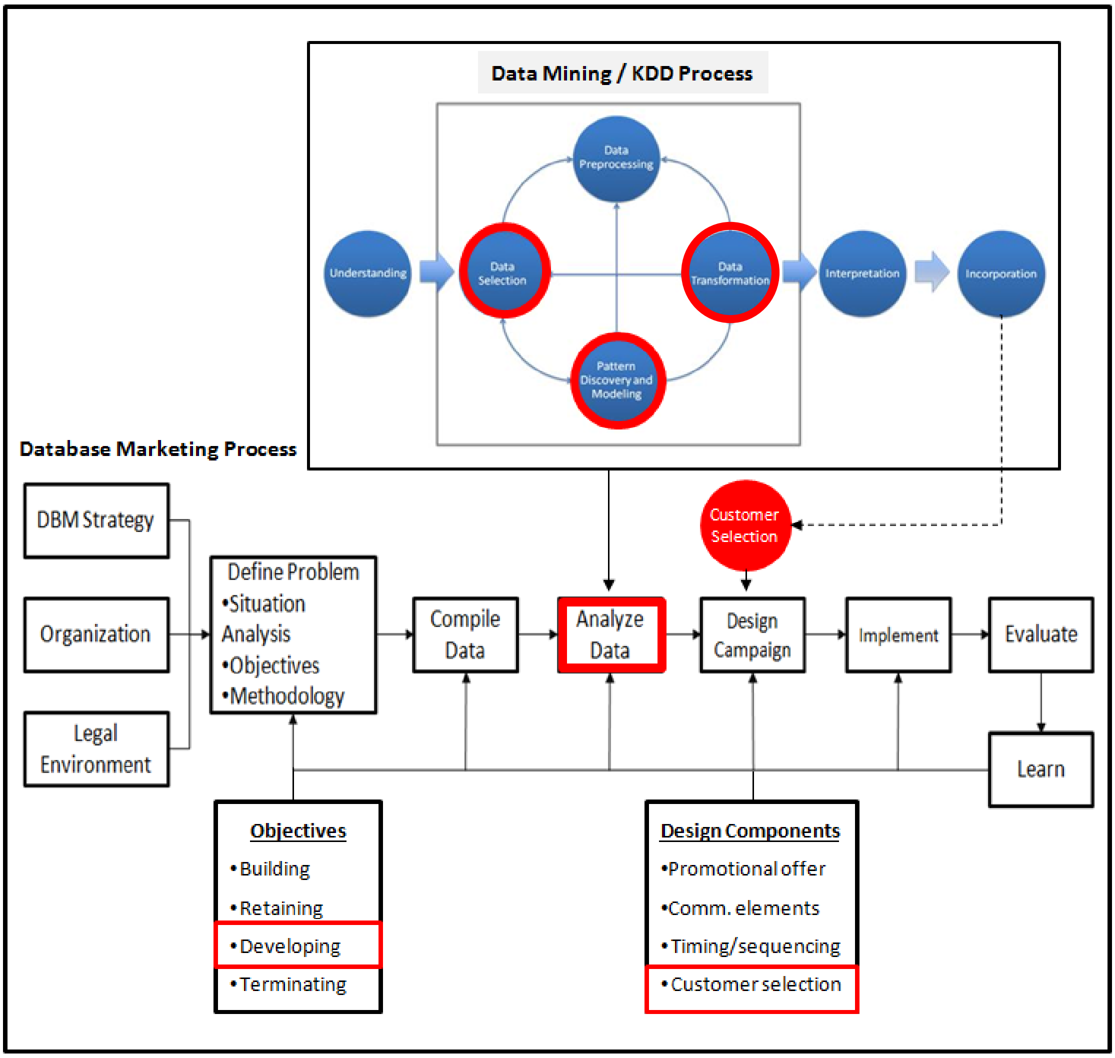
6.4.3 Database Marketing and Related Processes and Inputs

The increase in campaign performance and fit demonstrated by the sequential addition of variables to a set of advanced statistical and machine-learning techniques is more than a simple contribution to literature in the areas of statistics, machine learning or segmentation. It sheds some light on the significant gains possible by systematically applying all the steps in the KDD process. As mentioned in Figure 14, Studies 2 and 3 were focused on enhancing knowledge on the RM objective of customer development by increasing direct to consumer campaign response rates via a more effective customer selection approach. This process applied three of the four stages of the Data Mining process that included Data Selection, Pattern Discovery and the dimensionality reduction applications in Data Transformation. In applying the CHAID, logit and ANN techniques, the Pattern Discovery and Modelling and Data Transformation phases were conducted simultaneously as variable combination optimisation was embedded in model applications.

Consequently, in addition to demonstrating the contribution of variables and techniques to prediction and performance, this study illustrates that there is a direct relationship between the application of the broader process of KDD and performance. In addition, it has the added benefit of specifically calling out the contextual applications as well as the performance metrics it aims to pursue. As such, it supports the works of researchers such as Deichmann et al. (2002), which illustrate that Pattern Discovery can be well complemented by Dimensionality reduction.

This demonstrates the direct linkage between the extraction of customer knowledge from databases with the organisation and application of knowledge for the purposes of making effective marketing decisions in low-involvement contexts. Furthermore, this demonstration further supports Shaw et al.'s (2001, p. 128) observation that true customer relationship management integrates "the knowledge discovery process with the management and use of the knowledge for marketing strategies (that will help) marketers address customer needs based on what the marketers know about their customers, rather than a mass generalisation of the characteristics of customers."

Figure 36: Database Marketing and its Related Processes and Inputs



6.5 Summary

The intent of this third study is to understand technique effectiveness using expanded data variables in the context of FMCG retail. Expanded variables include the previously used variable of recency, frequency and monetary value, and all other variables provided by the retailer, including numerous detailed transactional, demographic and psychographic variables. All applied techniques are the same as those in Study 2.

As in Study 2, performance was assessed using response rate while comparative technique performance was assessed using lift charts, cumulative lift charts and cumulative gains charts. Fit and reliability measures used consisted of the Gini coefficient of dispersion and the PCC (when applicable). To ensure that sampling was random, the same validated cross validation folds from Study 2 were used. Results of the study were the following:

- All advanced statistical and machine-learning techniques seem to provide a fairly similar lift at each decile level versus the overall average response rate. Unlike Study 2, advanced technique results demonstrate a significant and visible lift increase versus the RFM technique. Looking at logistic regression, CHAID and neural networks, the lifts and gains obtained at the first two deciles provide enough response lift to allow these techniques' cumulative performance to surpass RFM well past decile five into decile six.
- From the raw input data, it can be firmly concluded that there is an important difference in the lift index between RFM and statistical and machine-learning techniques. RFM results show that response rates at the top decile are nearly 100-percent higher than the average response rate, while comparable results from CHAID, neural networks and logistic regression show increases of more than 200 percent. This result supports observations made in the systematic review, and provides evidence that not only do advanced statistical and machine-learning techniques outperform RFM, but they also do so with a much greater margin when extended variables are applied.
- From an advanced technique perspective, both logistic regression and ANN outperform CHAID at the top decile, while at an individual decile level, CHAID and ANN perform better at decile two, and CHAID continues outperforming for remaining top deciles three to five. From a cumulative perspective, the strong performance of logit and ANN allow these techniques to outperform in deciles one and two, but as of decile three, cumulative performance of all three advanced techniques becomes virtually identical. Though they remain very close from an overall Gini and PCC score perspective, the logistic regression outperforms ANN when using expanded data.
- As in Study 2, CHAID remains the technique with the best fit performance, while RFM has the lowest degree of fit (a significant fall from its Study 2 ranking). Fit for both logistic and ANN performance increases significantly versus Study 2.
- Unlike Study 2 where accuracy of RFM was second only to CHAID, results from Study 3 indicate that additional variables lead advanced techniques to exhibit higher model accuracy

than simple RFM. Furthermore, unlike Study 2, Study 3 shows a much greater degree of alignment between the techniques' predictive accuracy and performance, thus aligning the practitioner and academic interests of generating performance gains that are not realised at the expense of model stability.

- From a variable perspective, many emerged as being significant across technique and variable sets. While many variables emerged only 'accidentally,' across techniques, variables that emerged as contributing to the prediction power of the model could be classified into the following categories: socio-demographic, loyalty, channel and transactional. The high discrimination power of transactional, channel and loyalty variables supported the power of what Wedel and Kamakura (2000) call purchase and usage variables. This is also consistent with the findings of Zahay (2004), Yankelovich (2006), and Vogel (2008). However, the significant number of socio-demographic variables (such as household size, gender, income category, among others) that did add to the model strength was unexpected and is not supported by prescriptive guidance on variable categories showing increased responsiveness.
- The emergence of significance across numerous variable sets is necessarily accompanied by necessity to eliminate many variables from model inputs. The in-model elimination of 20-32 variables in each respective application of CHAID and logit demonstrates the critical importance of dimensionality reduction in complementing technique applications in order to optimise model performance.
- Finally, as the increase in campaign performance and fit is generated by the sequential addition of variables to a set of advanced statistical and machine-learning techniques, it is more than a simple contribution to literature in the areas of statistics, machine learning or segmentation. It sheds some light on the significant gains possible by systematically applying all the steps in the KDD process. Consequently, in addition to demonstrating the contribution of variables and techniques to prediction and performance, this study illustrates that there is a direct relationship between the application of the broader process of KDD and performance.

6.6 Conclusion

Study 3 repeated the research approach and techniques (RFM, stepwise logistic regression, CHAID, and Neural networks) of the second study but examined the performance impact of using expanded data variables. Expanded variables include the previously used variable of recency, frequency and monetary value, and all other variables provided by the retailer, including numerous detailed transactional, demographic and psychographic variables. The use of an extended data set not only generated additional insights on the type of data that impacted response but illustrated the true power of applying the systematic processes of database marketing and data mining.

The next sections will look at all studies and their individual findings and will provide an overview of contributions and implications to theory and practice, limitations of the current research, and guidance on future research opportunities.

7 General Conclusion

The thesis addresses two gaps in knowledge by presenting empirical evidence as to the relationship between customer selection techniques and performance and the effects of different data depths and specific data variables (and variable permutations) on performance. Table 86 illustrates the academic and practitioner contributions of the research. It categorises academic contributions as theoretical and methodical. Table 86 also categorises contributions based on the research priorities defined by Quinn and Dibb (2010): managerial relevance and implementation, value of segmentation, segmentation variables, methods and conceptualisations, stability and change, segmentation strategies; plus the additional priority of critical evaluation of current knowledge. Though not an overt research priority in their research output, systematic codification of knowledge was supported by some respondents in Quinn and Dibb's (2010, p. 1248) research as it provided a much needed "critical evaluation of knowledge so that weak links in the process" could be identified and addressed through potentially novel technique applications.

For clarity, the research priority of managerial relevance and implementation is not treated as a priority in the table but rather as a practitioner relevant contribution. Contributions are categorised based on the aforementioned research priorities and are also labelled as contributing new knowledge, supporting existing knowledge or challenging existing knowledge.

Table 86: Contributions

	Research Priority	Academic Contributions		Practioner Contributions
		Contributions to Theory	Contributions to Method	
New	**Critical Evaluation of Current Knowledge**		• Provides systematic approach on how to increase promotional effectiveness when pursuing the development objective in FMCG retail (Boulding et al., 2005; Ryals, 2005; Bailey, 2009)	
Supports	**Segmentation Variables**	• Strength extended data strength (loyalty, usage, channel, socio-demo) (Reutterer et al., 2006; Malthouse, 2006; Greene and Greene, 2008)	• In data rich contexts, data investments may contribute significant improvements, specifically with loyalty, usage, and channel variables	• In data rich contexts, data investments may contribute significant improvements, specifically with loyalty, usage, and channel variables • Validates that a cherry picker segment is amenable to developing a repeat-transactional relationship effectiveness (Lal and Bell, 2003)
	Methods and Conceptualisations	• In data rich contexts, advanced techniques offer significant performance improvements (Levin and Zahavi, 2001; Yang. 2004; McCarty, 2006)		
	Value of Segmentation	• Individualised marketing messages, facilitated by direct/data marketing practices, contribute to generating firm value (Rust et al., 2000) • Role of database marketing in achieving repeat transactional behaviours in RM (Eggert and Stieff, 1999; Maklan et al, 2011)		

	Research Priority	Academic Contributions		Practioner Contributions
		Contributions to Theory	Contributions to Method	
Challenges	Segmentation Variables	• Lack of impact of socio-demo variables (Wedel and Kamakura, 2000) • Contrary to loyalty theory that loyal and frequent customers are most profitable, identifies cherry pickers' contribution to loyalty returns (Lal and Bell, 2003)		• Lack of impact of socio-demo variables (Wedel and Kamakura, 2000)
	Methods and Conceptualisations	• In data poor contexts, investments in advanced techniques may not provide significant performance improvement as intuitive techniques such as RFM perform well	•	• In data poor contexts, investments in advanced techniques may not provide significant performance improvement as intuitive techniques such as RFM perform well

7.1 Contributions

This research contributes to theory in three of the key research areas discussed by Quinn and Dibb (2010):

1. Segmentation variables,

2. Methods and conceptualisations

3. Determining the value of segmentation

In addition, this thesis also contributes a critical evaluation of current knowledge through a systematic review of the literature.

7.1.1 Critical Evaluation of Current Knowledge

Findings illustrate that customer selection research has been overly technique-centric, focused on comparing various techniques in an unsystematic manner. Studies comparing the effectiveness of segmentation techniques identified in the systematic review do not thoroughly justify the selection of techniques investigated. This is partly illustrated by the wide difference in technique usage between academic research (Ngai, 2011) and practice (Verhoef et al., 2003). The justification, where presented, tends to be influenced by the techniques prevalence at the time research is conducted.

This thesis uses a systematic review to provide a transparent and evidence based process for technique selection and is less interested in the time relevance of techniques.

Unlike previous studies, the systematic review approach taken for this study examined techniques and data variables in tandem using a transparent, evidence-based process. This process was not only systematic and transparent, but it was also accompanied the use of widely recognised typologies of objectives, phases of data mining and data mining processes.

This represents an original contribution to method. The framework and methodical process utilised allows researchers to align their work to specific marketing objectives, identify the phases of data mining impacted and the steps followed within the data mining process. Whereas the extant literature does not consider specific marketing objectives, rather it assumes a universal comparison of techniques across all market contexts. Following the thesis process, executing as many of its steps as possible and applying the appropriate techniques to objectives is demonstrated in this study to provide enhanced campaign response. This outcome directly addresses the research priority of more critically evaluating current knowledge (Quinn and Dibb, 2010) by providing a systematic approach to linking objectives and techniques (Boulding et al, 2005, Ryals, 2005, Bailey, 2009).

7.1.2 Segmentation Variables

The findings of this thesis support research that demonstrates the positive relationship between model performance and the use of extended data variables. Results support the empirical findings of Magidson (1988), Levin and Zahavi (2001), Yang (2004), McCarty (2006), Reutterer et al. (2006), Malthouse (2006) and Greene (2008). The thesis findings also support research on the explanatory power of specific variables, namely: loyalty, usage, channel and socio-demographic variables. The high discrimination power of transactional, channel and loyalty variables supports research by Wedel and Kamakura (2000) on purchase and usage variables and is also consistent with the findings of Zahay, Peltier, Schultz, and Griffin (2004) and Yankelovich (2006). However, the significant contribution of socio-demographic variables to technique strength was unexpected and is not supported by research.

Findings also support the conclusions of McCarty and Hastak (2006), i.e., that data variables cannot be considered in absence of contextual applications. Hastak’s work with the collectible industry illustrates that many variables are significant in specific contexts only. The same can be said about the variables in both studies on FMCG retail. In Study 2, for example, given the relatively high frequency of visits of the top customer segment selected for analysis and the high commensurate expenditures, it was counterintuitive to see response rates being inversely related to monetary expenditures and frequency of visits. In light of the heavily engaged segment selected for this analysis, it is possible that share and frequency are indeed so high, that promotional response is more moderate and that such customers are loyal to the brand regardless of promotions. As such, a sub-segment of high spenders that has a lower frequency of visits may indeed be more likely to respond to a promotion and, as a result, the promotion might increase their visits. Though potentially contextual, it remains that such results contribute to enhancing segmentation theory by indicating that sub-segments within historically underperforming bases may indeed perform better (Reinartz and Kumar, 2000). This indicates that variable performance is not easily separated from the industry and customer segment contexts they are issued from and therefore questions the academic and practitioner efficacy of prescriptive models of variable effectiveness; such as those used by Wedel and Kamakura (2000).

These research findings address calls both from academic literature on the continued need to better understand what variables and techniques provide best results and from marketing managers on how segmentation outputs can be implemented in practice (Quinn and Dibb, 2010). While academics investigate complex techniques and the impacts of covariances on model outputs; the contribution of the thesis findings to loyalty, usage, channel and select socio-demographic variables provides practical guidance on what data investments may be immediately profitable for retailers. This would certainly allow CEOs to better prioritise CRM investments to enhance business performance (Gartner, 2012).

From a sub-segment- perspective, the significant responsiveness of the casual shopper, often called "cherry-pickers" suggests that practitioners could target specific groups of customers using shallow data to realise short-term improvements in marketing campaign performance. This finding is very much aligned with the work of Lal and Bell (2003) that indicates that, in the context of a US grocery chain's loyalty programme, it was the incremental sales from casual shoppers that often generated the greatest return for retailers. This is, nevertheless, a contrast with RM and loyalty theory that suggests that loyal and frequent customers are typically the most profitable (Reicheld and Teal, 2001). This also challenges research that suggests that "contacts from companies in low involvement purchase situations are not considered personal and, therefore, have little or no impact on relationship development" (Leahy, 2011, p.651). However, the ROI calculations of such investments should not assume that there is any sustained post promotion loyalty as cherry-picking behaviour is defined by promiscuous, promotion induced, shopping (Lal and Bell,2003).

7.1.3 Methods and Conceptualisations

Research findings provide a specific contribution to data mining in retail FMCG and further support that advanced statistical and machine-learning techniques perform much better than RFM for identifying high-response segments. More specifically, this provides evidence that not only do advanced statistical and machine-learning techniques outperform RFM, but they also do so with a much greater margin when extended variables are applied. This is consistent with the conclusions of empirical research by Magidson (1988), Levin and Zahavi (2001), Yang (2004) and McCarty (2006). The more "advanced" techniques, logistic regression and ANN, outperform CHAID at the top decile. The similar performance of logistic regression and ANN is corroborated in work by Zahavi and Levin (1998) while their stronger relative performance versus CHAID is corroborated by West et al. (1997)

and Linder et al. (2004)

Of interest to managers, this research indicates that in environments where managers have limited access to data, there is little difference between individual customer selection techniques. Where managers can access broader bases of data, there is a substantial difference between techniques. Considering the research by McCarty and Hastak (2006) on the robust performance of RFM in high-response contexts; practitioners should not only consider the depth of data available to them but the individual context in which techniques are used. This implies that whilst practitioners should adopt advanced techniques when extended variables are available, adopting lower complexity techniques such as RFM may be almost as effective in data constrained contexts.

Given the important positive effects of extended customer data on technique performance, practitioners should invest in gathering significant and relevant consumer data variables of loyalty, usage, channel and socio-demographics. However, the advanced techniques require well trained personnel (Drozdenko and Drake, 2002), therefore the datamarketing strategy of investing in extended customer data is linked to a commensurate investment in analytical capabilities. My conclusions build on the views Richards (2008, p. 126): “as managers are able to use these types of sophisticated models to mine the information available in CRM technology, improved customer targeting will result.”

7.1.4 Value of Segmentation

The combination of customer selection techniques and extended data variable set in the context of this thesis, provides some of the first empirical evidence (in FMCG retail) of combined performance effects for direct-to-consumer campaigns aimed at customer development. This contribution directly addresses the MSI's priority of "leveraging information about customer preferences and behaviour to enhance or supplant conventional strategic planning, market segmentation, and targeting approaches" (MSI, 2010, online). However, it is important to note that the information leverage cannot be separated by the information process as the research has demonstrated the importance of looking at data benefits across the database and data mining processes.

These results contribute to informing Eggert and Stieff’s (1999) concept of behavioural Relationship Marketing (RM), whose goal it is “to achieve repeat transactions through a process of interaction with the buyer, typically driven by economic goals rather than including some of the wider aspects of the exchange such as customer satisfaction” (Palmer, Lindgreen and Vanhamme, 2005, p. 318). This contribution to behavioural Relationship Marketing is akin to Coviello et al.’s (1997, 2002) Database Marketing stage of RM as it also leverages personal consumer data for targeting and personalisation of mainly tactical promotional offers. This improvement in the measurable contribution of enhanced targeting further supports Rust et al.’s (2000) contention that individualised marketing messages, facilitated by direct and data marketing practices, may indeed contribute to generating firm value.

Finally, beyond contributions to database marketing, segmentation and Data Mining processes, the research provides evidence that sophisticated marketing analytics practice improves marketing effectiveness. In fact, this research specifically illustrates that collecting extended data, integrating and transforming it effectively, and applying the right set of techniques and algorithms can produce improvements in marketing performance. These are the same elements that constitute the components of Big Data (Brown, Chui and Manyika, 2011) with the difference that Big Data has greater Volume, Velocity and Variety (McAfee and Brynjolfsson, 2012). By deduction, if a subset of extended data with relatively manageable volume, velocity and variety can generate significant improvement vis-à-vis a shallow dataset, a dataset that would include Big Data should not only allow for an even greater improvement in performance but also respond to changes in usage patterns in real time (Davenport et al., 2012).

Given the huge investments already made in building databases and CRM systems, investments in the required dynamic capabilities needed to leverage firms' CRM resources (Maklan et al., 2011) seem necessary support the priority that CEOs made of CRM "as their most important area of investment to improve business over the next five years" (Gartner, 2012, online).

7.2 Limitations

Whilst one endeavours to apply best the most stringent significance tests and the highest degree of objectivity in examining outputs; , trade-offs and decisions are inevitable in any research and findings must be considered in light of limitations.

Firstly, though findings are compelling, they remain focused on the objective of customer development and may not extend to other objectives of RM, CRM of database marketing.

Secondly, research was conducted on the context of FMCG retail. It is unclear whether results are generalisable to other industries, different promotional objectives, different time periods or different customer segments. Different organisations within the same industry may have their own particular characteristics, unique direct marketing objectives and different offer designs to address these objectives. Conclusions made with respect to the effectiveness of the four customer selection techniques are made within the context of a North American FMCG retailer. The promotion construct was designed to increase average customer sales with a spend-threshold promotional offer. Therefore, it is important to note that different promotional types and/or constructs may elicit different customer reactions and/or favour other techniques and data combinations. Techniques applied considered a static snapshot of response as measured by the Gini coefficient. Response rates may change over time and seasons and, as a result, this study should be replicated using a different yet comparable dataset to ensure consistency of findings. Furthermore, the promotional sample targeted only the top 30 percent of the retailer's most valuable loyalty-programme customers. Generalisability may be limited to the top customer segments as well as the industry. This speaks to the issue of endogenous 'observability' whereby not every customer in a database has an equal probability of being selected for a mailing, as organisations generally only mail customers considered most profitable (Donkers, 2006). Recent studies by Donkers (2006) showed that alternative techniques can lead to improved out-of-sample performances in that they could inevitably increase performance beyond the standard selection approaches.

Thirdly, data variables used as inputs for the research included variables that were available to the retailer and, although the list is comprehensive, it is still limited and specific to the retailer and the FMCG retail context. In addition, the analysis of recency, frequency, monetary value, or any other customer level variable – whether by an RFM model or by statistical techniques such as CHAID, ANN or logistic regression – focuses entirely on customers' past behaviour. Though powerful in its predictive ability, historical data likely limits marketers' in their ability to understand their customers. In their review of transactional and relational data, Zahay et al. (2004) argue that overemphasis on transactional information leads to a highly sales-oriented approach to customers that may support sales growth in the short term, but does not enhance or grow the long-term customer relationship. McCarty and Hastak (2006, p. 662) adds that increased consideration of relational data (i.e. motivations, attitudes, values, and lifestyles), "although these variables may be less useful than transaction information in their ability to predict a response to an immediate marketing activity...they may be enormously useful in understanding the underlying tendencies in customers. This consideration would favour analytical techniques such as CHAID and logistic regression that can accommodate a variety of personality and individual difference information." A valuable extension of this study would, therefore, be to analyse more time periods in conjunction with motivation and needs based variables.

Fourthly, for reasons of parsimony, only the four most researched and significant (as shown in the systematic review) selection techniques were chosen for testing and comparison purposes. Techniques that were not tested show promise in other contexts and/or when applied in combination. The thesis cannot determine if techniques that were not tested may perform better than the four highlighted in this study.

Lastly, it should also be noted that the marketing campaign was delivered via direct marketing. As such, conclusions may not be generalisable to other channels of communications such as email, banner advertising and retargeting among others.

7.3 Future Research

The future research direction that emerges from this research can be categorised into five topics: 1) the systematic codification of research, 2) replication against other marketing objectives and/or contexts, 3) the application of multiple and mixed techniques across the KDD process, 4) the examination of the impact of using extended data variables and Big Data in existing technique applications and 5) the evaluation of selection/segmentation success.

7.3.1 Systematic Codification of Research

Whilst this research represents a systematic approach to comparing the effectiveness of techniques against different data depths, there is a need to extend this to future research endeavours. Currently, research is characterised by a rather ad hoc approach to investigating customer selection techniques; each researcher using his or her own variable sets, preferred comparison methods and therefore the body of research is building very slowly towards an understanding of the most effective techniques. The systematic approach proposed in this thesis, if adopted by other researchers, would permit individual research activities to complement each other more and build a more accessible and scalable body of evidence as to the effectiveness of different techniques and variables in different contexts.

The expanded typology of objectives developed as part of the systematic review could be used by future researchers to design their own systematic reviews and thus start organising the research in the field of customer selection in a much more systematic and evidence-based manner. This could make research more accessible to practitioners by focusing their attention on research mainly relevant to practical questions related to the applicability to their marketing strategies (Cheron and Kleinschmidt, 1985). The same can be said about the need to systematically tag research into contextual or similar meta-contextual categories as this would also allow academics to start engaging more effectively with the practice agenda.

In addition, there is also significant opportunity for researchers to develop prescriptive applications of techniques against marketing objectives prior to starting a research endeavour by considering the techniques' normative applications. Systematically researching, tagging and testing based on objectives and phases of database marketing and data mining will allow for the growing body of knowledge on customer selection to become a lot more robust and scientific while allowing future research to effectively scale. Such a systematic approach would also allow for the researchers to better understand the relative contribution of each stage of the database marketing and data-mining phases.

7.3.2 Replication against other contexts and marketing objectives

The research focused solely on the context of retail FMCG. Though data availability in this context is vast and applications extensive, there have been very few studies beyond Guido's (2011) that have looked to examine the comparative effective of statistical techniques in data mining in this context. As detailed in the literature review, studies in charity, catalogue and mail order retail as well as financial services abound but there is a dearth of application in FMCG and even broader retail contexts. Future research could validate the research findings in different contexts using both comparable and context-specific data variables. This would allow researchers to assess the generalisability of findings on techniques, broad variable category (referred to by Wedel (2000) as segmentation bases) effectiveness, as well as to identify additional contextually sensitive contributions.

Similarly, using this dataset or another, replication of the current research approach against other marketing objectives aside from promotion response, could further contribute to future direction. Other research objectives can be sourced from the expanded typology of objectives developed as part of the systematic review. This would further contribute to organising research in the field of customer selection in a much more systematic and evidence-based manner.

7.3.3 Application of Multiple and Mixed Techniques

Although I selected the techniques that showed the most merit based on their systematic comparison in literature, future research should also focus on testing new emerging techniques (or mixed technique applications) on existing data-sets or further enhancing the outputs of applied techniques. A good example of technique improvement could be the further training ANNs to increase predictive response and further reduce cross-entropy error.

Furthermore, rather than limiting research to technique versus technique type comparisons, significant opportunities also exist for the integration of next-generation models to be evaluated against existing techniques. Approaches such as Linder et al.'s (2004) and Malthouse's (2008) use of multiple model outputs as inputs into a logistic regression mixed model regression approach show promise and illustrate the significant predictive power of the mixed model logit approach.

7.3.4 Extended and Big Data

Given the general evidence provided by the research on the importance of data, a promising avenue for future research is the identification of other key data variables that may allow for further performance improvements. In this research, for pragmatic reasons, existing data variables were selected as model inputs; however, the impact of alternative data variables on customer purchase intentions was not considered. Wedel and Kamakura (2000) suggest that additional data variables (that they refer to as specific/unobservable), such as benefit and intention type data, may provide even greater discrimination potential. Vogel (2008) suggests that customer perceptions of value, brand, and relationship (customer equity drivers) directly impact loyalty and future sales. Though applied in the context of a large European DIY (do-it-yourself) retailer, his study suggests that customer equity drivers can predict future customer actions and sales. In the era of Big Data, the availability of increasingly rich and complex information combined with systematic research in the area of predictive variables, variable combinations and technique impact is critical and will certainly allow researchers to continue establishing the tacit link between marketing and customer and firm equity.

In parallel, given the growing structured and unstructured datasets held by organisations, future research must continue to be pursued in the area of dimensionality reduction. Growing datasets will only further complicate the predictive process and as such studies similar to Deichmann et al. (2002) and Lee et al. (2011) could provide alternative avenues for testing model input variables and whether unbundled approaches could improve response further.

7.3.5 Evaluation of Selection/Segmentation Success

Given the growing interest of the marketing community on return on marketing investment and the dearth of academic studies using financial valuation metrics such as ROS, ROI, CLV, and other similar measures, research should move beyond a single-minded focus on response rate as a dependent variable to consider the aforementioned financial-based measures to facilitate cost/benefit analysis outcomes of applying techniques at different depths of data. At a basic level, researchers should start including financial returns and ROI as additional dimensions of customer selection decile analysis.

The integration of financial metrics would create immediate managerial relevance and would provide practitioners with further incentives to engage with academic literature. This managerial relevance could take the form of estimating the incremental financial value of: technique, variable or process improvements. Incremental value would necessarily need to be measured contextually as different contexts enjoy unique margins, operating models and particular competitive responses to name but a few particularities. In this, researchers in database marketing might follow the studies that analyse customer satisfaction: satisfaction has been linked to cash flow (Gruca, 2005), stock prices (Fornell, 2006), market share (Rego, 2103).

7.4 Conclusion

This research provides supporting evidence on the relationship between customer selection techniques and performance and the effects of using different depths of data. Furthermore, this research starts codifying literature into specific objectives, contexts and data depths. Findings that emerge from the thesis contribute the advancing knowledge in Database Marketing and segmentation, Data Mining techniques applications, and Data Mining processes and behavioural relationship marketing. It also provided a very clear contribution to technique applications in both data rich and poor retail FMCG context.

The codification mentioned above, albeit a very shallow start can provide the roadmap for future research organisation and also act as a bridge to reduce the practitioner-academic research gap.

Understandably, marketing science academics will always want to push the boundaries of knowledge both from a data input and model innovation perspective. Not allowing them to do so would certainly slow the rate of potential innovations, however, organising these efforts and attempting to bring more managerial relevance (either through more transparent contextual examinations or dimensions of marketing strategy to be evaluated) would certainly be a significant step in growing the intrinsic synergies that exist between practical and academic research interests.

8 References

1. Aaker, D. A. (1991), *Managing brand equity: Capitalizing on the value of a brand name,* Free press New York.
2. Abbott, J., Stone, M. and Buttle, F. (2001), "Integrating customer data into customer relationship management strategy: An empirical study", *Journal of Database Marketing & Customer Strategy Management*, vol. 8, no. 4, pp. 289-300.
3. Abdi, H. (2010). “Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdisciplinary Reviews: Computational Statistics*”, vol. 2, no. 1, pp 97-106.
4. Agresti, A, Finlay, (1986), *Statistical Methods for the Social Sciences (2nd edition),* Dellen Publishers, San Francisco, CA
5. Anbalagan, C. (2011), “Prospects from Different Channels of Direct Marketing on Indian Markets”, *International Journal of Research in Management and Technology*, Vol. 1, No.1, October
6. Anderson, J. L., Jolly, L. D., and Fairhurst, A. E., (2007), “Customer relationship management in retailing: A content analysis of retail trade journals”, *Journal of Retailing and Consumer Services*, vol. 14, no. 6, pp. 394-399.
7. Andersson, P. and Soderlund, M., (1988) "The Network Approach to Marketing", *Irish Marketing Review*, Vol. 1 pp. 63-67
8. Ang, L. and Buttle, F. (2006), "CRM software applications and business performance", *Journal of Database Marketing & Customer Strategy Management,* vol. 14, no. 1, pp. 4-16.
9. Ansari, A., & Mela, C. F. (2003), “E-customisation”, *Journal of Marketing Research*, vol. *40, no.* 2, pp. 131-145.
10. Antipov, E., & Pokryshevskaya, E., (2010), “Applying CHAID for logistic regression diagnostics and classification accuracy improvement”, *Journal of Targeting, Measurement and Analysis for Marketing*, vol. 18, no. 2, pp. 109-117.
11. Archer, K. J., Lemeshow, S., & Hosmer, D. W. (2007), “Goodness-of-fit tests for logistic regression models when data are collected using a complex sampling design”, *Computational Statistics & Data Analysis*, vol. 51, no. 9, pp. 4450-4464
12. Baier M, Ruf K.M., Chakraborty, G., (2002) *Contemporary database marketing: concepts and applications*, Racom Communications, Evanston, IL
13. Bailey, C., Baines, P. R., Wilson, H. and Clark, M. (2009), "Segmentation and customer insight in contemporary services marketing practice: why grouping customers is no longer enough", *Journal of Marketing Management,* vol. 25, no. 3, pp. 227-252.
14. Baines, T.S., (2007), “State-Of-The-Art in Product-Service Systems”, *Proceedings of the Institution of Mechanical Engineers - Part B,* Engineering vol. 221, no.10, pp.1543–1552.
15. Barnes, James G. (1994) “*The Issue of Establishing Relationships With Customers in Service Companies: When Are Relationships Feasible and What Form Should They Take*?” Working paper, Memorial University of Newfoundland
16. Beane T.P. and Ennis, D.M., (1987), “Market Segmentation: a review”, Eu*ropean Journal of*

Marketig, vol. 21, October, pp. 20-42

17. Bell, D., T.H. Ho, C. Tang, (1998), "Determining where to shop: fixed and variable costs of shopping", Journal of Marketing Research, vol.35, no. 3, pp. 352-369.
18. Beltramini, R. F., & Pitta, D. A. (1991), "Underlying dimensions and communications strategies of the advertising agency-client relationship", *International Journal of Advertising*, vol. 10, no. 2, pp. 151-59.
19. Bendapudi, N., & Berry, L. L., (1997), "Customers' motivations for maintaining relationships with service providers", *Journal of retailing*, vol. *73, no. 1*, pp. 15-37.
20. Berardi, V.L. and Zhang, G.P., (1999), "The effect of misclassification costs on neural network classifiers", *Decision Sciences*, Vol. 30, no. 3, pp. 659-82.
21. Berry, L.L., (1983), *Relationship marketing. Emerging Perspectives on Services*, AMA Press, Chicago, IL
22. Berry, L.L. & Gresham, L.G.,(1986), "Relationship Retailing; Transforming Customers into Clients" *Business Horizons,* November/December, pp 43-47
23. Berry, L.L. & Parasuraman, A.,(1991), *Marketing Services*, The Free Press, New York, NY
24. Berry, L. L., & Parasuraman, A. (2004). Marke*ting services: Competing through quality*. Simon and Schuster.
25. Bertier, P., & Bouroche, J. (1981). *Analyse des données multidimensionnelles*, Presses Universitaires de France, Paris, France
26. Betancourt, R. (2005). *The Economics of Retailing and Distribution*. Edward Elgar Publishing.
27. Bharadwaj, S. G., Varadarajan, P. R., & Fahy, J. (1993). Sustainable competitive advantage in service industries: A conceptual model and research, *Journal of Marketing*, vol. *57, no.*4
28. Bishop, C. M. (1995) *Neural Networks for Pattern Recognition*, Oxford University, Oxford
29. Bitner, Mary Jo. (1995). "Building Service Relationships: It's All about Promises." *Journal of the Academy of Marketing Science*, vol. 23, Fall, pp. 246-251.
30. Blattberg, R. C. and Deighton, J. (1996), "Manage marketing by the customer equity test", *Harvard Business Review,* vol. 74, pp. 136-145.
31. Blattberg, R. C., Kim, B.D., Neslin, N.A., 2008, *Database Marketing–Analysing and Managing Customer*, Springer, New York, NY
32. Blocker, C. P. and Flint, D. J. (2007), "Customer segments as moving targets: Integrating customer value dynamism into segment instability logic", *Industrial Marketing Management,* vol. 36, no. 6, pp. 810-822.
33. Bonoma, T. V., & Shapiro, B. P. (1983), "*Segmenting the industrial market*", Lexington Books, Lexington, MA:
34. Bose, R. (2002), "Customer relationship management: key components for IT success", *Industrial Management & Data Systems*, vol.102, 1/2, p 89.
35. Bose, I. and Chen, X. (2009), "Quantitative models for direct marketing: A review from systems perspective", *European Journal of Operational Research,* vol. 195, no. 1, pp. 1-16.
36. Boulding, W., Staelin, R., Ehret, M., & Johnston, W. J. (2005). "A customer relationship management roadmap: what is known, potential pitfalls, and where to go", *Journal of Marketing*,

pp. 155-166

37. Boullé, M. (2004). “Khiops a statistical discretization method of continuous attributes”, *Machine Learning*, vol. 55, pp. 53–69.
38. Boyer, K. K., & Hult, G. T. M. (2005),”Extending the supply chain: integrating operations and marketing in the online grocery industry”, *Journal of Operations Management*, vol. *23, no.* 6, pp. 642-661.
39. Brachman, R. and Anand, T. (1996), *The Process of Knowledge Discovery in Databases. Advances in Knowledge Discovery and data Mining,* The MIT Press, Boston.
40. Brodie, R.J., Coviello N.E, Brookes, R.W., Little, V., (1997), “Towards a Paradigm Shift in Marketing? An Examination of Current Marketing Practices”, *Journal of Marketing Management*, vol. 13, pp. 383-406
41. Brodie, R. J. (2002), “The challenge to include relational concepts”, M*arketing Theory*, vol. *2*, no. 4, pp. 339-343.
42. Brodie, R.J. and De Kluyer, C.A., (1987), “A Comparison of the Short Term Forecasting Accuracy of Econometric and Naive Extrapolation Models of Market Share”, *International Journal of Forecasting*, vol. 3, pp. 423-437.
43. Brown, B., Chui, M., & Manyika, J. (2011), “Are you ready for the era of ‘Big Data’?”, *McKinsey Quarterly*, vol. 4
44. Brusco, M. J., Cradit, J. D. and Tashchian, A. (2003), "Multicriterion Clusterwise Regression for Joint Segmentation Settings: An Application to Customer Value", *Journal of Marketing Research,* vol. 40, no. 2, pp. 225-234
45. Buckinx, W. and Van den Poel, P. (2005), "Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting", *European Journal of Operational Research,* vol. 164, no. 1, pp. 252-268.
46. Bughin, J., Chui, M., Manyika, J., (2010) “Clouds, Big Data, and smart assets: Ten tech-enabled business trends to watch”, *McKinsey Quarterly*, vol. 3
47. Bult, J. R. and Wansbeek, T. (1995), "Optimal Selection for Direct Mail", *Marketing Science,* vol. 14, no. 4, pp. 378.
48. Bult, J.R., Wittink, D.R., (1996), “Estimating and validating asymmetric heterogeneous loss functions applied to health care fund raising”, *International Journal of Research in Marketing*, vol. 13, pp. 215-226
49. Burrell, G. and Morgan, G. (1979), *Sociological paradigms and organisational analysis,* Heinemann, London.
50. Buttle, Francis A. (2001), “The CRM Value Chain,” *Marketing Business*, (February), pp. 52–55.
51. Cacioppo, J. T., Semin, G. R. and Berntson, G. G. (2004), "Realism, Instrumentalism, and Scientific Symbiosis: Psychological Theory as a Search for Truth and the Discovery of Solutions", *American Psychologist,* vol. 59, no. 4, pp. 214.
52. Calder, B.J., & Malthouse, E.C. (2002). *What Is Integrated Marketing?, Kellogg on Integrated Marketing*, John Wiley and Sons, Evanston, IL
53. Campbell, A. J. (2003), “Creating customer knowledge competence: managing customer

relationship management programs strategically", *Industrial Marketing Management*, vol. 32, no. 5, 375.

54. Chaffey, D., Mayer, R., Johnston, K. And Ellis-Chadwick, F. (2000), *Internet Marketing*, Harlow, Pearson Education,
55. Chan, C. C. H. (2008), "Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer", *Expert Systems with Applications*, vol. 34, no. 4, pp. 2754-2762.
56. Chandon, P., Wansink, B., & Laurent, G. (2000), "A benefit congruency framework of sales promotion effectiveness", *Journal of Marketing*, pp. 65-81.
57. Chang, W., Park, J. E., & Chaiy, S. (2010). "How does CRM technology transform into organisational performance? A mediating role of marketing capability", *Journal of Business Research*, vol. *63, no.* 8, pp. 849-855.
58. Chen, C. J. (2004). The effects of knowledge attribute, alliance characteristics, and absorptive capacity on knowledge transfer performance. *R&D Management*,*34*(3), 311-321.
59. Chen, J. S., & Ching, R. K. (2004), "An empirical study of the relationship of IT intensity and organisational absorptive capacity on CRM performance", *Journal of Global Information Management*, vol. 12, no. 1, pp. 1-17.
60. Chen, I. J., & Popovich, K. (2003), "Understanding customer relationship management (CRM): People, process, and technology", *Business Process Management Journal*, vol. 9, no. 5, pp. 672−688.
61. Cheron, E.J., and Kleinschmidt, E.J., (1985), "A review of industrial market segmentation research and a proposal for an integrated segmentation framework", *International Journal of Research in Marketing*, vol. 2, no. 2, pp. 101–115.
62. Ching, J., Wong, A., & Chan, K. (1995), "Class-dependent discretization for inductive learning for continuous and mixed-mode data", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol.17, no. 7, pp. 641–651.
63. Chintagunta, P. K., Chu, J., & Cebollada, J. (2012), "Quantifying transaction costs in online/off-line grocery channel choice", *Marketing Science*, vol. 31, no. 1, pp. 96-114.
64. Christopher, M., Payne, A. and Ballantyne, D. (1991), *Relationship Marketing*, Butterworth-Heinemann, Oxford
65. Cooil, B., Winer, R.S. and Rados, D.L. (1987), "Cross-validation for prediction", *Journal of Marketing Research*, Vol. 24 No. 3, pp. 271-9.
66. Cook, R. Dennis and Hakbae Lee (1999), "Dimension Reduction in Binary Response Regression," *Journal of the American Statistical Association*, vol. 94, December, pp.1187–1200.
67. Coviello, N. E., Brodie, R. J., & Munro, H. J. (1997), "Understanding contemporary marketing: development of a classification scheme", *Journal of Marketing Management*, vol. 13, no.6, pp. 501-522.
68. Coviello, N. E., Brodie, R. J., Danaher, P. J. and Johnston, W. J. (2002), "How Firms Relate to Their Markets: An Empirical Examination of Contemporary Marketing Practices", *Journal of Marketing,* vol. 66, no. 3, pp. 33-46.

69. Couldwell, Clive (1999), “Loyalty Bonuses,” *Marketing Week*, (February 18), p. 14.
70. Crone, S. F., Lessmann, S. and Stahlbock, R. (2006), "The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing", *European Journal of Operational Research*, vol. 173, no. 3, pp. 781.
71. Cronin, P., Ryan, F., & Coughlan, M. (2008), ”Undertaking a literature review: a step-by-step approach”, *British Journal of Nursing*, vol. *17, no.* 1, p. 38.
72. Crosby, L. A., Evans, K. R., & Cowles, D. (1990, “Relationship quality in services selling: An interpersonal influence perspective”, *Journal of Marketing*, vol. 54, July, pp. 68−81.
73. Croteau, A., & Li, P. (2003), “Critical success factors of CRM technological initiatives”, *Canadian Journal of Administrative Sciences*, vol. 20, no. 1, pp. 21-34.
74. Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*, SAGE Publications Limited.
75. Cui, G. and Wong, M. L. (2004), "Implementing neural networks for decision support in direct marketing", *International Journal of Market Research,* vol. 46, pp. 235.
76. Cui, G., Wong, M. L. and Lui, H. (2006), "Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming", *Management Science,* vol. 52, no. 4, pp. 597.
77. Cui, G., Wong, M. L., Zhang, G., & Li, L. (2008), ”Model selection for direct marketing: performance criteria and validation methods”, *Marketing Intelligence & Planning*, vol. *26, no.* 3, pp. 275-292
78. Curry, D. J. (1993), *The new marketing research systems: How to use strategic database information for better marketing decisions*, Wiley, New York.
79. Danneels, E. (1996). “Market segmentation: normative model versus business reality: an exploratory study of apparel retailing in Belgium”, European *Journal of Marketing*, vol. 30, no. 6, pp. 36-51.
80. Das, K., (2008), “Relationship Marketing research (1994-2006): An academic literature review and classification”, *Marketing Intelligence & Planning*, Vol. 27 No. 3, pp. 326-363
81. Davenport, T. H., Barth, P., & Bean, R. (2012). How ‘big data’is different. *MIT Sloan Management Review*, vol. *54, no.* 1, pp. 22-24.
82. Davenport, T. H. and Harris, J. G. (2007), *Competing on analytics: the new science of winning,* Harvard Business Press, Boston, MA.
83. David Shepard and Associates, (1999), *The New Direct Marketing, 3rd edition*, McGraw Hill, New York, NY
84. Davies, A., Brady, T. and Hobday, M. (2006), “Charting a path towards integrated solutions”, *Sloan Management Review*, Spring, pp. 38–49
85. Day, G. S. (2011), "Closing the marketing capabilities gap", *Journal of Marketing,* vol. 75, no. 4, pp. 183-195.
86. Deichmann, J., Eshghi, A., Haughton, D., Sayek, S. and Teebagy, N. (2002), "Application of Multiple Adaptive Regression Splines (Mars) in Direct Response Modeling", *Journal of Interactive Marketing,* vol. 16, no. 4, pp. 15-27.

87. Dickson, P. R. (1982), "Person-Situation: Segmentation's Missing Link", *Journal of Marketing,* vol. 46, no. 4, pp. 56–64.
88. Dibb, S., Meadows, M., (2004) “Relationship Marketing and CRM: a financial services case study”, *Journal of Strategic Marketing*, Vol. 12 Issue 2, pp.111-125
89. Dibb, S. and Simkin, L. (1994), “Implementation problems in industrial market segmentation,” *Industrial Marketing Management*, Vol. 23, February, pp. 55-63.
90. Dibb, S. and Simkin, L. (2001), "Market segmentation: Diagnosing and treating the barriers", *Industrial Marketing Management,* vol. 30, no. 8, pp. 609.
91. Dixon-Woods, M., Bonas, S., Booth, A., Jones, D. R., Miller, T., Sutton, A. J., & Young, B. (2006), “How can systematic reviews incorporate qualitative research? A critical perspective”, *Qualitative research*, vol. *6, no.*1, pp. 27-44.
92. Dolnicar, S. & Lazarevski, K., 2009. Methodological reasons for the theory/practice divide in market segmentation. *Journal of Marketing Management*, 25(3/4), pp.357–373.
93. Dougherty, J., Kohavi, R., & Sahami, M. (1995). “Supervised and unsupervised discretization of continuous features”, *Machine learning: Proceedings of the twelfth international conference*. Morgan Kaufmann Publishers, San Francisco
94. Doyle, D. P. (1998), *Adding Value to Marketing: The Role of Marketing in Today's Profit-Driven Organisation,* Kogan Page.
95. Drozdenko, R. G., and Drake, P. D. *Optimal Database Marketing: Strategy, Development, and Data Mining*. Sage Publications, 2002.
96. Dubow, J. S. (1992), “Occasion-based vs. user-based benefit segmentation: A case study”, *Journal of Advertising Research*.
97. Dwyer, R. R., Schurr, P. H., & Oh, S. (1987), “Developing buyer–seller relations”, *Journal of Marketing*, vol. 51,pp. 11−28.
98. Easterby-Smith, M. P. V. Thorpe. R., & Lowe, A.(2002). *Management research: An Introduction,* Sage, London
99. Efron, B., & Tibshirani, R. J. (1993), An introduction to the bootstrap, Chapman and Hall , New York, NY
100. Eggert, A. and Stieff, J. (1999), “What constitutes a relationship? Towards a conceptualisation of relationship marketing’s central construct”, *Proceedings of the Industrial Marketing and Purchasing Group Conference*, University College Dublin, Dublin, 2-4.
101. Evans, J.R. & Berman, B. (1997). *Marketing (7th edition),* Prentice Hall, Upper Saddle River.
102. Ernst H., Hoyer W. D., Krafft, M., Krieger,K., (2010). “Customer Relationship Management and Company Performance—the Mediating Role of New Product Performance”, *Journal of the Academy of Marketing Science*, vol. 39, no. 2, pp.290–306.
103. Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996a), "From Data Mining to knowledge discovery in databases", *Communications of the ACM*, vol. 39, no. 11, pp. 24-26.
104. Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996b), "Knowledge discovery and data mining: Towards a unifying framework", P*roc. 2nd Int. Conf. on Knowledge Discovery and Data Mining*, Portland, OR, pp. 82.

105. Fayyad, U. and Uthurusamy, R. (2002), "Evolving Data Mining into Solutions for Insights-Introduction", *Communications of the ACM*, vol. 45, no. 8, pp. 28-32.

106. Field, A. (2005). *Discovering statistics using SPSS,* Sage publications, London, UK

107. Finlay, S. (2012). *Credit Scoring, Response Modeling, and Insurance Rating: A Practical Guide to Forecasting Consumer Behavior*. Palgrave Macmillan

108. Firth, D.R. and Lawrence, C. (2006), "An Institutional Perspective on Customer Relationship Management", *Journal of Information Technology Theory and Application,* vol. 8, no. 1, pp. 21-31.

109. Fletcher, K., & Wright, G. (1995), "Organisational, strategic and technical barriers to successful implementation of database marketing", *International Journal of Information Management*, vol. 1*5, no.* 2, pp.115-126.

110. Foedermayr, E. K. (2008), "Market Segmentation in Practice: Review of Empirical Studies, Methodological Assessment, and Agenda for Future Research", *Journal of Strategic Marketing,* vol. 16, no. 3, pp. 223-265.

111. Foedermayr, E. K., & Diamantopoulos, A. (2008), "Market segmentation in practice: Review of empirical studies, methodological assessment, and agenda for future research", *Journal of Strategic Marketing*, vol. *16, no.* 3, pp. 223-265.

112. Food Marketing Institute (2012), *U.S. Grocery Shopper Trends 2012 Executive Summary*, available at http://www.icn-net.com/docs/12086_FMIN_Trends2012_v5.pdf, (accessed January 12, 2013)

113. Ford, D. (1990), *Understanding Business Markets: Interaction, Relationships and Networks*, Academic Press, London.

114. Fornell, C. et al., (2006), "Customer Satisfaction and Stock Prices: High Returns, Low Risk", *Journal of Marketing*, vol. 70 no. 1, pp.3–14.

115. Fox, E. J., & Sethuraman, R. (2010) *Retail competition, Retailing in the 21st Century* , Springer, Berlin Heidelberg, Chicago, IL

116. Frawley, W., Piatetsky-Shapiro, G., & Matheus, C. (1992), "Knowledge Discovery in Databases: An Overview". *AI Magazine*, Fall, pp. 213–228.

117. Freytag, P. V. and Clarke, A. H. (2001), "Business to Business Market Segmentation", *Industrial Marketing Management,* vol. 30, no. 6, pp. 473-486.

118. Frow, P., Payne, A. (2009), "Customer relationship management: a strategic perspective", *Journal of Business Marketing Management*, Vol. 3 No.1, pp.7-27.

119. Gantz, J. F., Reinsel, D., Chute, C., Schlichting, W., McArthur, J., Minton, S., Manfrediz, A. (2007). The *Expanding Digital Universe: A Forecast of Worldwide Information Growth Through 2010*. Framingham, MA: IDC.

120. Gartner Group (2003), "CRM Economics: Figuring Out the ROI on Customer Initiatives," white paper, Stamford, CT.

121. Gartner, (2012), 2012 CEO survey, available at http://www.gartner.com/newsroom/id/1993515, (accessed April 30, 2013)

122. Gelman, A., Hill, J., & Yajima, M. (2012), "Why we (usually) don't have to worry about multiple comparisons", *Journal of Research on Educational Effectiveness*,vol. 5, no. 2, pp.189-211.

123. Glazer, R. (1997), “Strategy and structure in information-intensive markets: the relationship between marketing and IT”, *Journal of Market Focused Management, vol. 2*, no. 1, pp. 65-81

124. Gönül, F., Kim, B.D., Shi, M. (2000), "Mailing smarter to catalog customer", *Journal of Interactive Marketing*, Vol. 14 No.2, pp.2-16.

125. Gonzalez-Abril, L., Cuberos, F. J., Velasco, F. and Ortega, J. A. (2009), “Ameva: An autonomous discretization algorithm”, *Expert Systems with Applications*, vol. 36, no. 3, pp.5327-5332.

126. Gosney, John and Thomas Boehm (2000), *Customer Relationship Management Essentials*, Prima Publishing, Indianapolis, IN

127. Greenberg, M., & McDonald, S. S. (1989), “Successful needs/benefits segmentation: A user's guide”, *Journal of Consumer Marketing*, vol. 6, no. 3, pp. 29-36.

128. Greene, H. J.,(2005),"A statistic for measuring response model performance: Determining the significance of the gains chart", *Electronic Doctoral Dissertations* , University of Massachusetts, Amherst.

129. Greene, H. and Greene, S. (2008), "Enhancing segmentation systems", *Journal of Targeting, Measurement & Analysis for Marketing,* vol. 16, no. 4, pp. 298-311.

130. Greene, H. J ., Milne, G. R., (2010), “Assessing model performance: The Gini statistic and its standard error”, *Database Marketing & Customer Strategy Management* Vol. 17,no.1, pp. 36–48

131. Grishikashvili, K., Dibb, S., and Meadows, M., (2014) “Investigation into Big Data Impact on Digital Marketing”, I*nternational Conference on Communication, Media, Technology and Design* 24 - 26 April 2014, Istanbul – Turkey.

132. Grönroos, C., (1990). “Relationship Approach to Marketing in Service Contexts: The Marketing and Organisational Behavior Interface.” *Journal of Business Research*, vol. 20, January, pp. 3-11.

133. Grönroos, C. (1991) "The Marketing Strategy Continuum: Towards a Marketing Concept for the 1990s", *Management Decision,* Vol. 29, no 1 pp. 7-13

134. Grönroos, C. (1994a) "From Marketing Mix to Relationship Marketing: Towards a Paradigm Shift in Marketing", *Management Decision,* Vol. 32, no. 2 pp. 4-20

135. Grönroos, C. (1994b) "Quo Vadis, Marketing? Toward a Relationship Marketing Paradigm" *Journal of Marketing Management*, vol. 10, pp. 347-360

136. Grönroos, C. (1996), “The Rise and Fall of Modern Marketing – And Its Rebirth”, in S. A. Shaw and N. Hood (eds) *Marketing in Evolution*: Essays in Honour of Michael J. Baker, pp. 14–35. Basingstoke: Macmillan

137. Grönroos, C. (1997), “Interaction, dialogue and value processes of relationship marketing”, *Proceedings of the 5th International Colloquium on Relationship Marketing*, Cranfield School of Management, Cranfield, November.

138. Grönroos, C., (2009), “Marketing as promise management: regaining customer management”, *Journal of Business & Industrial Marketing*,vol. 24, no. 5/6, pp. 351–359

139. Gruca, T.S, Rego, L.L., (2005), “Customer Satisfaction, Cash Flow, and Shareholder Value”, *Journal of Marketing*, vol. 69 no. 3, pp.115–130.

140. Guido, G., Prete, M. I., Miraglia, S. and De Mare, I. (2011), "Targeting direct marketing campaigns by neural networks", *Journal of Marketing Management,* vol. 27, no. 9-10, pp. 992-1006.

141. Gummesson, E. (1987), "The New Marketing - Developing Long-Term Interactive Relationships", *Long Range Planning,v*ol. 20, no 4, pp. 10-20

142. Gummesson, E. (1991) "Marketing Orientation Revisited: The Crucial Role of the Part-Time Marketer", *European Journal of Marketing*, vol. 25, no. 2, pp.60-75

143. Gummesson, E. (1997) "In Search of Marketing Equilibrium: Relationship Marketing Versus Hypercompetition", *Journal of Marketing Management* , vol. 13, pp. 421-430

144. Gummesson, E. (1994a) "Is Relationship Marketing operational?" *23rd EMAC Conference*, Maastricht

145. Gummesson, E. (1994b) "Making Relationship Marketing Operational", *International Journal of Service Industry Management, v*ol. 5, no. 5 pp. 5-20

146. Gupta, G., & Aggarwal, H., (2012)," Improving Customer Relationship Management Using Data Mining", *International Journal of Machine Learning and Computing*, vol. 2, no.6, pp.874-877.

147. Gupta, S., & Lehmann, D. (2003), "Customers as assets", *Journal of Interactive Marketing*, vol. 17, no. 1, pp. 9-24.

148. Gwinner, K. P., Gremler, D. D., & Bitner, M. J. (1998), "Relational benefits in services industries: the customer's perspective", *Journal of the Academy of Marketing Science*, vol. 26, no. 2, pp.101-114.

149. Ha, K., Cho, S., & MacLachlan, D. (2005), "Response models based on bagging neural networks", *Journal of Interactive Marketing*, vol. *19, no.* 1, pp. 17-30.

150. Hakansson, H. (1982), *International Marketing and Purchasing of Industrial Goods*, Wiley & Sons, New York, NY.

151. Haley, R. (1985), D*eveloping Effective Communication Strategy: A benefit Segmentation Approach*, John Wiley & Sons, New York, NY.

152. Hamerly, G., & Elkan, C. (2002), "Alternatives to the k-means algorithm that find better clusterings", *Proceedings of the eleventh international conference on Information and knowledge management,* pp. 600-607.

153. Hammarkvist, K.O., Hakansson, H. and Mattsson, L. (1982) *Marketing for Competitiveness*, Lund, Liber

154. Han, J. and Kamber, M. (2006), *Data mining: concepts and techniques,* Morgan Kaufmann, San Francisco.

155. Hand, D. J., Mannila, H. and Smyth, P. (2001), *Principles of data mining,* The MIT Press, Boston.

156. Harker, M. J. (1999), "Relationship marketing defined? An examination of current relationship marketing definitions", *Marketing Intelligence & Planning*, vol. *17, no.* 1, pp.13-20.

157. Harker, M. J., Egan, J. (2006), "The Past, Present and Future of Relationship Marketing", *Journal of Marketing Management*, Vol. 22, pp. 215-242

158. Hart, C. (2006). *Doing a literature review: Releasing the social science research imagination*.

Sage, London

159. Hasan, M. (2003), “Ensure success of CRM with a change in mindset”, *Marketing Management*,vol. 37, no. 8, p. 16.

160. Haughton, D. and Oulabi, S., (1997), *Direct marketing modeling with CART and CHAID*, John Wiley & Sons, Inc.

161. Hawkins, D. M., Bask, S. C., & Mills, D. (2003), “Assessing model fit by cross-validation”, *Journal of chemical information and computer sciences*, vol. 43, no. 2, pp. 579-586.

162. Heilman, C.M., Kaefer, F., and Ramenofsky, S.D., 2003, “Determining the appropriate amount of data for classifying consumers for direct marketing purposes”, *Journal of Interactive Marketing*, vol. 17, no. 3

163. Hill, T., Thomas, M. and Lewicki, P., (2006). *Statistics: methods and applications: a comprehensive reference for science, industry, and data mining*. StatSoft, Inc.

164. Hobby, J. (1999), “Looking After the One Who Matters,” *Accountancy Age*, (October 28), pp. 28-30

165. Holte, R. C. (1993), “Very simple classification rules perform well on most commonly used datasets”, *Machine Learning*, vol. 11, pp.63–90.

166. Hopfield, J. J. (1987), “Learning Algorithms and Probability Distributions in Feed-Forward and Feed-Back Networks”, *Proceedings of the National Academy of Sciences*, vol. 84, pp. 8429-8433.

167. Hsieh, M. (2009), “A case of managing customer relationship management systems: Empirical insights and lessons learned”, *International Journal of Information Management*, vol. 29, no. 5, pp/ 416-419.

168. Hu, X., & Cercone, N. (2001), “Discovering maximal generalised decision rules through horizontal and vertical data reduction”, Computational *Intelligence*, vol. 17, no. 4,pp. 685-702.

169. Huang, Y., & Wang, J. (2009), “Services-Oriented CRM System and Enabling Technologies for Insurance Enterprises”, *Proceedings of the Second International Workshop on Knowledge Discovery and Data Mining*.

170. Huff, A.S. (2007), *Designing Scholarly Research,* Sage, London

171. Hughes, A. (1996) *The Complete Database Marketer*, McGraw Hill, New York, NY

172. Hughes, A., (2000), *Strategic Database Marketing, 2nd edition*, McGraw Hill, New York, NY

173. Humby, C., Hunt, T., & Phillips, T. (2003). *Scoring Points: How Tesco is winning customer loyalty*. Kogan Page Limited.

174. Hung, S.-Y., Hung, W.-H., Tsai, C.-A., & Jiang, S.-C. (2010), “Critical factors of hospital adoption on CRM system: Organisational and information system perspectives”, *Decision Support Systems*, vol. 48, no. 4, pp. 592-603.

175. Hunt, S. D. and Arnett, D. B. (2006), "Does marketing success lead to market success?", *Journal of Business Research,* vol. 59, no. 7, pp. 820.

176. Iacobucci, D. (1994), ‘Towards defining relationship marketing’’, in Sheth, J.N. and Parvatiyar, A. (Eds), *1994 Research Conference Proceedings. Relationship Marketing: Relationship Marketing: Theory, Methods, and Applications*, Centre for Relationship Marketing, Emory University, Atlanta, GA.

177. Jackson, J. (2002), “Data mining: A conceptual overview’, *Communications of the Association for Information Systems*, vol. 8, no. 19, pp. 267-296.

178. Jacob, V. S., Gaultney, L. D., and Salvendy, G. (1986), “Strategies and biases in human decision-making and their implications for expert systems”, *Behaviour & Information Technology*, vol. 5, no. 2, pp. 119-140.

179. Jain, S. C. (1997), *Marketing Planning and Strategy,* 5th edition ed, Southwestern Publishing, Dallas, TX.

180. Jain, A. K. (2010), “Data clustering: 50 years beyond K-means”, *Pattern Recognition Letters*, 31(8), pp. 651-666.

181. Jayachandran, S., Sharma, S., Kaufman, P., Raman, P., (2005), “The Role of Relational Information Processes and Technology Use in Customer Relationship Management”, *Journal of Marketing*, vol. 69, no.4, pp.177–192.

182. Jedidi, K., Jagpal, H. S. and DeSarbo, W. S. (1997), "Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity", *Marketing Science,* vol. 16, no. 1, pp. 39.

183. Jenkins, M., & McDonald, M. (1997), “Market segmentation: organisational archetypes and research agendas”, *European Journal of Marketing*, vol. 31, no.1, pp.17-32.

184. Johnstone, S., Dainty, A. and Wilkinson, A. (2008), “In search of ‘product-service’: evidence from aerospace, construction and engineering”, *The Service Industries Journal*, Vol. 28 No. 6, pp. 861–75

185. Jonker, J. J., Piersma, N. and Van den Poel, D. (2004), "Joint optimisation of customer segmentation and marketing policy to maximise long-term profitability", *Expert Systems with Applications,* vol. 27, no. 2, pp. 159-168.

186. Jonker, J., Piersma, N. and Potharst, R. (2006), "A decision support system for direct mailing decisions", *Decision Support Systems,* vol. 42, no. 2, pp. 915-925.

187. Kahan, R. (1998), “Using Database Marketing techniques to enhance your one-to-one marketing initiatives”, *Journal of Consumer Marketing*, vol. 15, no. 5, pp. 491-493.

188. Kamakura, W., Mela, C. F., Ansari, A., Bodapati, A., Fader, P., Iyengar, R., Naik, P., Neslin, S., Sun, B., Verhoef, P.C., Wedel. M., Wilcox, R. (2005), “Choice models and customer relationship management”, *Marketing Letters*, vol.16, no. 3-4, pp. 279-291.

189. Kamakura, W. A. and Russell, G. J. (1991), "Measuring consumer perceptions of brand quality with scanner data: implications for brand equity", .

190. Kamakura, W. A., Wedel, M., De Rosa, F., & Mazzon, J. A. (2003), “Cross-selling through Database Marketing: a mixed data factor Analyser for data augmentation and prediction”, *International Journal of Research in Marketing*, vol. 20, no. 1, pp. 45-65.

191. Kantardzic, M. (2002), *Data mining: concepts, models, methods and algorithms,* Wiley-Interscience, New York.

192. Karakostas, B., Kardaras, D., & Papathanassiou, E. (2005). “The state of CRM adoption by the financial services in the UK: an empirical investigation”, *Information & Management*, vol. 42, no.6, pp.853-863.

193. Keller, K. L. (1993), "Conceptualising, measuring, and managing customer-based brand equity", *The Journal of Marketing,* vol. 57, no. 1, pp. 1-22.

194. Keller, K. and Aaker, D. (1998), "The impact of corporate marketing on a company's brand extensions", *Corporate Reputation Review,* vol. 1, no. 4, pp. 356-378.

195. Kemsley, E. K. (1996), “Discriminant analysis of high-dimensional data: a comparison of principal components analysis and partial least squares data reduction methods”, *Chemometrics and intelligent laboratory systems*, vol. 33. No.1, pp. 47-61.

196. Kerber, R. (1992). Chimerge: Discretization of numeric attributes. In: Proceedings of the AAAI-9, 9th international conference on artificial intelligence. MIT Press

197. Khanna, S. (2001), “Measuring the CRM ROI: Show Them Benefits,” (accessed November 20, 2002), [available at http://www.crm-forum.com].

198. Kim, Y. (2006), “Toward a successful CRM: variable selection, sampling, and ensemble”, Decision *Support Systems*, vol. 41, no. 2, pp.542-553.

199. Kim, Y. S., Street, W. N., Russell, G. J. and Menczer, F. (2005), "Customer targeting: A neural network approach guided by genetic algorithms", *Management Science,* vol. 51, no. 2, pp. 264-276.

200. Kincaid, J. (2002). *Customer relationship management: getting it right!* : Prentice Hall PTR.

201. Knox, S. (1998), “Loyalty-based segmentation and the customer development process”, *European Management Journal*, *vol. 16, no. 6, pp. 729-737.*

202. Knox S, Maklan S, Payne A, Peppard J, Ryals L (2003). *Customer relationship management: perspectives from the marketplace*. Butterworth-Heinemann, Burlington, MA

203. Kohavi, R. (1995), “A study of cross-validation and bootstrap for accuracy estimation and model selection”, *Proceedings of the Fourteenth International Joint Conference on AI*, Montreal, August.

204. Kohavi, R., & Parekh, R. (2004), “Visualising RFM segmentation”, *Proceedings of the Fourth SIAM International Conference on Data Mining,* Society for Industrial Mathematics.

205. Kotler, P. (1991), *Marketing Management: Analysis, Planning, Implementation and Control,* Prentice Hall, New York.

206. Kotler (1997), *Marketing Management: Analysis, Planning, Implementation and Control (9th edition),* Prentice Hall, Upper Saddle River

207. Kotler, P. and Bliemel, F. (2000), "Marketing Management", *Englewood Cliffs*

208. Kumar, A. (1998), “New techniques for data reduction in a database system for knowledge discovery applications”, *Journal of Intelligent Information Systems*, vol.10, no.1, pp. 31-48.

209. Kumar, V., & Reinartz, W. (2012), “Applications of CRM in B2B and B2C Scenarios (Part I)”, *Customer Relationship Management,* Springer Berlin Heidelberg, (pp. 303-333).

210. Kumar, V., Venkatesan, R. and Reinartz, W. (2006), "Knowing what to sell, when, and to whom.", *Harvard business review,* vol. 84, no. 3, pp. 131.

211. Kumar, A., & Zhang, D. (2007), “Hand-geometry recognition using entropy-based discretization”, *IEEE Transactions on Information Forensics and Security*,vol. 2, no.2, pp. 181–187.

212. Kumar, V., & Reinartz, W. (2012), *Applications of CRM in B2B and B2C Scenarios (Part I). In Customer Relationship Management,* Springer Berlin Heidelberg, Chicago, IL

213. Kurgan, L., & Cios, K. (2004), "CAIM discretization algorithm", *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 2,pp. 145–153.

214. Kutner, S., & Cripps, J. (1997), "Managing the customer portfolio of healthcare enterprises", *The Healthcare Forum Journal,* vol. 4 (Sept-Oct), pp. 52-54.

215. Lal, R., & Bell, D. E. (2003), "The impact of frequent shopper programs in grocery retailing", *Quantitative Marketing and Economics*, vol. 1, no. 2, pp. 179-202.

216. Langley, A. (2007), "Process thinking in strategic organisation", *Strategic Organisation,* vol. 5, no. 3, pp. 271.

217. LaPlaca, P. J. (2004), "Letter from the editor: Special issue on customer relationship management", Industrial *Marketing Management*, vol. 33, no.6, pp. 463−464.

218. LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). "Big Data, analytics and the path from insights to value". *MIT Sloan Management Review*, vol.52, no.2, pp. 21-32.

219. Le, D.-D., & Satoh, S. (2007), "Ent-boost: Boosting using entropy measures for robust object detection", *Pattern Recognition Letters*, vol. 28, pp. 419–425.

220. Leahy, R. (2011). Relationships in fast moving consumer goods markets: The consumers' perspective. *European Journal of Marketing*, vol. *45, no.* 4, pp. 651-672.

221. Lee, N. & Greenley, G., 2010. The theory-practice divide: thoughts from the Editors and Senior Advisory Board of EJM. *European Journal of Marketing*, 44(1/2), pp.5–20.

222. Lee, H., Lee, Y., Cho, H., Im, K., & Kim, Y. S. (2011), "Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model", *Decision Support Systems*, vol. 52, no.1, pp. 207-216.

223. Lee-Kelley, L., Gilbert, D., & Mannicom, R. (2003), "How e-CRM can enhance customer loyalty", *Marketing Intelligence & Planning*, vol. 2, no. 4, pp. 239-248.

224. Leigh, T. W., & Tanner Jr., J. F. (2004), "Introduction: JPSSM special issue on customer relationship management", *Journal of Personal Selling and Sales Management*, vol. 24, no. 4, pp. 259−262.

225. Lemmens, A., & Croux, C. (2006), "Bagging and boosting classification trees to predict churn", *Journal of Marketing Research*, pp. 276-286.

226. Levin, N. and Zahavi, J. (2001), "Predictive modeling using segmentation", *Journal of Interactive Marketing,* vol. 15, no. 2, pp. 2-22.

227. Levin, N. and Zahavi, J. (1998), "Continuous Predictive Modeling--a Comparative Analysis", *Journal of Interactive Marketing (John Wiley & Sons),* vol. 12, no. 2, pp. 5-22.

228. Levin, N. and Zahavi, J. (2001), "Predictive modeling using segmentation", *Journal of Interactive Marketing,* vol. 15, no. 2, pp. 2-22.

229. Lewis, M., V. Singh, S. Fay. (2006), "An empirical study of the impact of nonlinear shipping & handling fees on purchase incidence and expenditure decisions", *Marketing Science*, vol.25, no., pp. 51-64.

230. Li, Ker-Chau (1991), "Sliced Inverse Regression for Dimension Reduction," *Journal of the American Statistical Association*, vol. 86, June, pp. 316–42.

231. Li, L. (2010), "Dimension reduction for high-dimensional data", *Statistical Methods in Molecular Biology*, Humana Press. pp. 417-434.

232. Lilien, G.L., 2011. Bridging the Academic-Practitioner Divide in Marketing Decision Models. *Journal of Marketing*, 75(4), pp.196–210.

233. Lilien, G. L. and Rangaswamy, A. (2004), *Marketing engineering: computer-assisted marketing analysis and planning,* Trafford Publishing, Victoria, BC.

234. Linder, R., Geier, J. and Kölliker, M. (2004), "Artificial neural networks, classification trees and regression: Which method for which customer base?", *Journal of Database Marketing & Customer Strategy Management,* vol. 11, no. 4, pp. 344-356.

235. Lindgreen, A. (2001), "A framework for studying Relationship Marketing dyads", *Qualitative Market Research: An International Journal*, Vol. 4 No. 2, pp. 75-87, 88.

236. Liou, J. J. H. (2009), "A novel decision rules approach for customer relationship management of the airline market", *Expert Systems with Applications*, vol. 36, no.3, Part 1, pp.4374-4381.

237. Liu, H., & Setiono, R. (1997), "Feature selection via discretization", *IEEE Transactions on Knowledge and Data Engineering*, vol. 9, no. 4, pp. 642–645.

238. Lix, T. S., Berger, P. D. and Magliozzi, T. L. (1995), "New customer acquisition: Prospecting models and the use of commercially available external data", *Journal of Direct Marketing,* vol. 9, no. 4, pp. 8.

239. Lohr, S., (2012)."The age of Big Data", *New York Times*, February 11, 2012, p. SR1

240. Lun, Z., Jinlin, L., & Yingying, W. (2008), "Customer relationship management system framework design of Beijing Rural Commercial Bank", *Proceedings of IEEE International Conference on Service Operations and Logistics, and Informatics*, 2008. (IEEE/SOLI 2008).

241. Lunt, N., Shaw, I., and Fouche, C., (2010)., "Practitioner research: Collaboration and knowledge production", *Public Money and Management*, vol 30, no , 4, pp. 235–242.

242. Magidson, J. (1988), "Improved Statistical Techniques for Response Modeling Progression Beyond Regression", *Journal of Direct Marketing,* vol. 2, no. 4, pp. 6-18.

243. Magliozzi, T. L. and Berger, P. (1993), "List segmentation strategies in direct marketing", *The International Journal of Management Science*, vol. 2, no. 71.

244. Malthouse, E. C. (1999), "Ridge regression and direct marketing scoring models", *Journal of interactive marketing,* vol. 13, no. 4, pp. 10-23.

245. Malthouse, J.C., (2001), "Assessing the performance of direct marketing scoring models", *Journal of Interactive Marketing*, vol. 15, no. 1, pp. 49-62.

246. Malthouse, E. C. and Elsner, R. (2006), "Customisation with crossed-basis sub-segmentation", *Journal of Database Marketing & Customer Strategy Management,* vol. 14, no. 1, pp. 40.

247. Maklan, S., Knox, S., Peppard, J. (2011), "Why CRM Fails – and How to Fix It", *Sloan Management Review*, vol. 52, no. 4, pp. 77-85

248. Mama, F., (2007), "Actioning market segmentation in FMCGs-a multiple case study

approach", MsC Thesis, Cranfield University, Cranfield, UK

249. Marcus, C. (1998), "A practical yet meaningful approach to customer segmentation", *Journal of Consumer Marketing,* vol. 15, no. 5, pp. 494-504.

250. Marketing Science Institute (MSI)(2007), *2010-2012 Research Priorities*, available at: http://www.msi.org/research/index.cfm?id=271 (accessed 12th September 2011).

251. Malthouse, E. C. (1999), "Ridge regression and direct marketing scoring models", *Journal of interactive marketing,* vol. 13, no. 4, pp. 10-23.

252. Malthouse, E. C. and Derenthal, K. M. (2008), "Improving predictive scoring models through model aggregation", *Journal of Interactive Marketing (John Wiley & Sons),* vol. 22, no. 3, pp. 51-68.

253. Malthouse, E. C. and Elsner, R. (2006), "Customisation with crossed-basis sub-segmentation", *Journal of Database Marketing & Customer Strategy Management,* vol. 14, no. 1,

254. Marshall, A. et al., (1991), *On Markets; Markets, Hierarchies and Networks*, SagePublishing, London

255. Mattsson, L.-G. (1997). "Relationship marketing and the Markets-as-Networks approach - A comparative analysis of two evolving streams of research", *Journal of Marketing Management,* vol. 13, pp. 447–461.

256. McAfee, A., & Brynjolfsson, E. (2012), "Big Data: the management revolution", *Harvard business review*, vol. 90, no. 10, pp. 60-66

257. McCarty, J. A. and Hastak, M. (2007), "Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression", *Journal of Business Research,* vol. 60, no. 6, pp. 656.

258. McCrary, M. (2009), "Enhanced customer targeting with multi-stage models: Predicting customer sales and profit in the retail industry", Jou*rnal of Targeting, Measurement and Analysis for Marketing*, vol. 17, no. 4, pp. 273-295.

259. McDonald, M. and Dunbar, I. (2000), "Using structured processes and systems to help managers develop strategic segmentation", *Journal of Targeting, Measurement and Analysis for Marketing,* , no. 2, pp. 109.

260. Mejia-Lavalle, M., Arroyo-Figueroa, G., & Morales, E. F. (2009), "Innovative applications of diagnosis, forecasting, pattern recognition and knowledge discovery in power systems", *Power & Energy Society General Meeting*, PES'09. IEEE, pp. 1-9)

261. Menard, S. (1995). *Applied Logistic Regression Analysis*, Sage Publications, Thousand Oaks, CA

262. Mitchell, T. (1997), *Machine Learning*, McGraw-Hill, New York, NY.

263. Moller, K. and Halinen, A. (2000), "Relationship Marketing theory: its roots and direction", Journal *of Marketing Management* , vol. 16, no. 1,pp. 29–54.

264. Morgan, R.M., Hunt, S. D., (1994) "The commitment-trust theory of relationship marketing", *Journal of Marketing*; vol. 58, no. 3; p. 20

265. Morgan, G., & Smircich, L. (1980), "The case for qualitative research", *Academy of management review*, vol. *5* no. 4, pp. 491-500.

266. Morrit, R. (2007), *Segmentation Strategies for Hospitality Managers: Target Marketing for Competitive Advantage,* Haworth Press, Binghamton

267. Naik, P. A., Hagerty, M. R. and Tsai, C. (2000), "A New Dimension Reduction Approach for Data-Rich Marketing Environments: Sliced Inverse Regression", *Journal of Marketing Research (JMR),* vol. 37, no. 1, pp. 88-101.

268. Naik, P., Wedel, M., Bacon, L., Bodapati, A., Bradlow, E., Kamakura, W., Kreulen, J., Lenk, P., Madigan, D. and Montgomery, A. (2008), "Challenges and opportunities in high-dimensional choice data analyses", *Marketing Letters,* vol. 19, no. 3, pp. 201-213.

269. Nash, E. L. (1984), *The direct marketing handbook,* McGraw-Hill, New York.

270. Neely, A., (2008), “Exploring the financial consequences of the servitization of manufacturing”, *Operations Management Research*, vol. 1, no. 2, pp.103–118.

271. Neslin, S. A., Gupta, S., Kamakura, W., Lu, J., & Mason, C. H. (2006), “Defection detection: Measuring and understanding the predictive accuracy of customer churn models”, *Journal of Marketing Research*,pp. 204-211.

272. Neslin, Scott A., (2006): "Defection detection: Measuring and understanding the predictive accuracy of customer churn models." *Journal of Marketing Research*, pp. 204-211.

273. Newell, F. (2003), *Why CRM doesn't work: how to win by letting customers manage the relationship*, Kogan Page.

274. Ngai, E.W., Xiu, L., Chau, D.C.,(2009) “Application of data mining techniques in customer relationship management: A literature review and classification”, *Expert Systems and Applications*, vol 36, pp. 2592–2602

275. Olson, D. L., & Chae, B. K. (2012), “Direct marketing decision support through predictive customer response modelling”, *Decision Support Systems*.

276. Orlikowski, W. J., & Baroudi, J. J. (1991), “Studying information technology in organisations: Research approaches and assumptions”, *Information systems research*, vol. *2, no.* 1, pp. 1-28.

277. Ortmeyer, G., Lattin, J. M., & Montgomery, D. B. (1991), “Individual differences in response to consumer promotions”, *International Journal of Research in Marketing*, vol. 8, no. 3, pp. 169-186.

278. Palmatier, R. W. (2008). *Relationship marketing*, Marketing Science Institute Cambridge, MA

279. Palmer, R., Lindgreen, A., & Vanhamme, J. (2005). Relationship marketing: schools of thought and future research directions. *Marketing Intelligence & Planning*, vol. *23, no.*3, pp 313-330.

280. Parvatiyar, A., & Sheth, J. N. (2000), “The domain and conceptual foundations of Relationship Marketing Handbook of relationship marketing, 3, 38.

281. Parvatiyar, A., & Sheth, J. N. (2001), “Customer Relationship Management: Emerging Practice, Process, and Discipline”, *Journal of Economic & Social Research*, vol. 3, no. 2

282. Passingham, J. (1998) “Grocery Retailing and the Loyalty Card”, *Journal of the Market Research Society*, January, pp.55-63

283. Payne, A (1995b) "Relationship Marketing: A Broadened View of Marketing" In: Payne, A. (Ed) (1995a) Advances in Relationship Marketing The Cranfield Management Series, Kogan Page.

284. Payne, A. (Ed) (1995a), *Advances in Relationship Marketing* The Cranfield Management Series, Kogan Page.

285. Payne, A., & Frow, P. (2005), “A strategic framework for customer relationship management”, *Journal of Marketing*, vol. 69, no. 4, pp. 167-176.

286. Payne, A. and Frow, P. (2006), "Customer Relationship Management: from Strategy to Implementation", *Journal of Marketing Management,* vol. 22, no. 1, pp. 135-168.

287. Pels, J., Coviello, N. and Brodie, R. (1999), “Transactions versus relationships? The risk of missing the real issues”, *Proceedings of the Industrial Marketing and Purchasing Group Conference*, University College Dublin, Dublin, 2-4 September.

288. Peppers, D., & Rogers, M. (1995), “A new marketing paradigm: share of customer, not market share”, *Managing Service Quality*, vol. 5, no. 3, pp. 48-51.

289. Peppers, D., & Rogers, M. (2004), *Managing customer relationships: A strategic framework*, Wiley, New York, NY

290. Peppers, D., & Rogers, M. and Dorf, B. (1999), “Is Your Company Ready for One-to-One Marketing?” *Harvard Business Review*, vol. 77 (January–February), pp. 151–60.

291. Perrien, J., & Ricard, L. (1995), “The meaning of a marketing relationship: a pilot study”, *Industrial Marketing Management*, vol. *24, no.* 1, pp. 37-43

292. Petrison, L.A., Blattgerg, R.C., Wang, P., (1993), “Database Marketing: Past, Present, and Future”, journal of Interactive Marketing, Vol. 7, No. 3

293. Petticrew, M. and Roberts, H. (2006), *Systematic Reviews in the Social Sciences: A Practical Guide,* Blackwell Pub.

294. Piatetsky-Shapiro G., Masand, B., (1999), “Estimating Campaign Benefits and Modeling Lift”, *Proceedings of 5th International Conference on Knowledge Discovery & Data Mining,* pp. 185–193.

295. Plunkett, K., & Elman, J. L. (1997), *Exercises in rethinking innateness: A handbook for connectionist simulations*, The MIT Press, Boston, MA

296. Quinn, L. (2009), “Market segmentation in managerial practice: a qualitative examination”, *Journal of Marketing Management*, vol. 25, no. 3-4, pp. 253-272.

297. Quinn, L. And Dibb, S. (2010), “Evaluating market-segmentation research priorities”, *Journal of Marketing Management*, vol. 26, no. 13-14, December 2010, pp. 1239-1255

298. Quinn, L., Hines, T., & Bennison, D. (2007), “Making sense of market segmentation: a fashion retailing case”, European *Journal of Marketing*, vol. 41, no. 5/6, pp. 439-465.

299. Rao, C. P., & Ali, J. (2002), “Neural network model for database marketing in the new global economy”, *Marketing Intelligence & Planning*, vol. *20, no.* 1, pp. 35-43.

300. Ratner, B. (2001), "Finding the best variables for direct marketing models", *Journal of Targeting, Measurement & Analysis for Marketing,* vol. 9, no. 3, pp. 270.

301. Rego, L.L., Morgan, N.A. & Fornell, C., (2013), “Reexamining the Market Share- Customer Satisfaction Relationship”, *Journal of Marketing*, vol. 77 (September), pp.1–20.

302. Reichheld, F. F. (2001), T*he loyalty effect: The hidden force behind growth, profits, and lasting value,* Harvard Business Press, Boston, MA.

303. Reichheld, F. F., Teal, T. (2001), "*The loyalty effect: The hidden force behind growth, profits, and lasting value",* Harvard Business Press, Boston, MA.

304. Reichheld, F. F., & Teal, T, (1996) "The loyalty effect", *Harvard Business School Press, Boston*, MA

305. Reinartz, W., Krafft, M., & Hoyer, W. D. (2004), "The customer relationship management process: Its measurement and impact on performance", *Journal of Marketing Research*, vol. XLI, pp. 293−305

306. Reinartz, W., Thomas, J., Kumar, V., (2005), "Balancing Acquisition and Retention Resources to Maximise Profitability," *Journal of Marketing*, vol. 69, January, pp. 63–79.

307. Reinartz, W. J., & Kumar, V. (2000), "On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing", *Journal of Marketing*, pp. 17-35.

308. Reinartz, W. J., & Venkatesan, R. (2008), "Decision models for customer relationship management (CRM)", *Handbook of marketing decision models,* Springer, US.

309. Reisz, M., (2010, July 15), "Style points" *Times Higher Education*, pp. 34–39.

310. Retka, R., "Beyond Traditional Segmentation", *Direct Marketing News*, Nov 2001

311. Reutterer, T., Mild, A., Natter, M. and Taudes, A. (2006), "A dynamic segmentation approach for targeting and customising direct marketing campaigns", *Journal of Interactive Marketing,* vol. 20, no. 3-4, pp. 43-57.

312. Richards, K.A., Jones, E., (2008), "Customer relationship management: Finding value drivers", *Industrial Marketing Management* , vol. 37, pp. 120–130

313. Richards, K. A., & Jones, E. (2008), "Customer relationship management: Finding value drivers", *Industrial marketing management*, vol. 37, no. 2, pp.120-130.

314. Rigby, D. K., Reichheld, F. F., & Schefter, P. (2002), "Avoid the four perils of CRM", *Harvard Business Review*, vol. 80, no. 2, pp. 101−109

315. Risselada, H., Verhoef, P. C., & Bijmolt, T. H. (2010), "Staying power of churn prediction models", *Journal of Interactive Marketing*, vol. *24, no.* 3, pp. 198-208

316. Roberts, M. L. and Berger, P. (1999), *Direct Marketing Management, 2nd ed.,* Prentice Hall Upper Saddle River, NJ

317. Ross, N. (1992), "A History of Direct Marketing," Unpublished Paper, Direct Marketing Association, NY

318. Rossi, P. E., McCulloch, R. E. and Allenby, G. M. (1996), "The value of purchase history data in target marketing", *Marketing Science,* , pp. 321-340.

319. Rud, O. P. (2001). *Data Mining Cookbook*. Wiley, New York, NY

320. Rudelius, W., Walton, J.R. and Cross, J.C. (1987), "Improving the managerial relevance of market segmentation studies," in Houston, M.J. (Ed.), *Review of Marketing*, American Marketing Association, Chicago, IL, pp. 385-404.

321. Rust, R.T., Zeithaml, V.A. and Lemon, K.N., (2000), *Driving Customer Equity: How Customer Lifetime Value Is Reshaping Corporate Strategy*, Free Press, New York, NY

322. Rust, R. T. and Verhoef, P. C. (2005), "Optimising the Marketing Interventions Mix in

Intermediate-Term CRM", *Marketing Science,* vol. 24, no. 3, pp. 477-489.

323. Ryals, L., & Knox, S. (2001), "Cross-Functional Issues in the Implementation of Relationship Marketing Through Customer Relationship Management", *European Management Journal*, vol. 19, no. 5, pp. 534-567.

324. Ryals, L. (2003), "Making customers pay: measuring and managing customer risk and returns", *Journal of Strategic Marketing*, vol.11, no. 3, pp. 165-175.

325. Ryals, L. (2005), "Making Customer Relationship Management Work: The Measurement and Profitable Management of Customer Relationships", *Journal of Marketing,* vol. 69, no. October, pp. 252-61.

326. Ryals, L., & Payne, A. (2001). Customer relationship management in financial services: towards information-enabled relationship marketing. *Journal of strategic marketing*, vol. *9*, no. 1, pp. 3-27.

327. Santos, J., Alexandre, L., & Marques de Sá, J. (2004), "The error entropy minimisation algorithm for neural network classification". *Int. conf. on recent advances in soft computing*.

328. SAS, (2011), *Big Data and the Analytic Race*, available at source from the web: http://regions.cmg.org/regions/mspcmg/Agenda_Summer_July2012_files/Big%20data%20and%20the%20Analytic%20Race_CMG2012-07-26.pdf, (accessed May 23, 2012)

329. Sausen, K., Tomczak, T. and Herrmann, A. (2005), "Development of a taxonomy of strategic market segmentation: a framework for bridging the implementation gap between normative segmentation and business practice", *Journal of Strategic Marketing,* vol. 13, no. 3, pp. 151-173.

330. Seo, S. (2002), "A review and comparison of methods for detecting outliers in univariate data sets", Doctoral dissertation, University of Pittsburgh

331. Shani, D. and Chalasani, S. (1992), "Exploiting niches using relationship marketing", *Journal of Consumer Marketing,* vol. 9, no. 3, pp. 33-42.

332. Shaw, R. and Stone, M., (1988), *Database Marketing*, Wiley, New York, NY

333. Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001),"Knowledge management and data mining for marketing", *Decision Support Systems*, vol. *31, no.* 1, pp.127-137.

334. Sheth, J. N., Sisodia, R. S., & Sharma, A. (2000), "The antecedents and consequences of customer-centric marketing", *Journal of the Academy of Marketing Science*, vol. 28, no. 1, 1988 55-66.

335. Sheth, J. N., & Sisodia, R. S. (1999), "Revisiting marketing's lawlike generalisations", *Journal of the Academy of Marketing Science*, vol. 27, no.1, 1988 71-87.

336. Sheth, J. N., Parvatiyar, A., (1995), "Relationship Marketing in Consumer Markets: Antecedents and Consequences", *Journal of the Academy of Marketing Science*, vol. 23, no. 4, pp. 255-271

337. Shen, L. (2002),"A Modified Chi2 Algorithm for Discretization", *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 3

338. Shoemaker, M. E. (2001), "A framework for examining IT-enabled market relationships", *Journal of Personal Selling and Sales Management*,vol. 21, no. 2, pp. 177–185.

339. Shih, Y. Y. and Liu, C. Y. (2003), "A method for customer lifetime value ranking Combining

the analytic hierarchy process and clustering analysis", *The Journal of Database Marketing &# 38; Customer Strategy Management,* vol. 11, no. 2, pp. 159-172.

340. Silva, L., Alexandre, L., & Marques de Sá, J. (2005), “Neural network classification: Maximising zero-error density”. *LNCS*, vol. 3686.

341. Silva, L. M., Marques de Sá, J., & Alexandre, L. A. (2008), “Data classification with multilayer perceptrons using a generalised error function”, *Neural Networks*, vol. 21no.9, 1pp.1302-1310

342. Singh, D. and Agrawal, D.P. (2003), “CRM Practices in Indian Industries,” *International Journal of Customer Relationship Management*, vol. 5 (December–January), pp. 241–57.

343. Smith, W. R. (1956), "Product Differentiation and Market Segmentation as Alternative Marketing Strategies", *Journal of Marketing,* vol. 21, no. 1, pp. 3-8.

344. Spangler, S. and Kreulen, J. (2007). *Mining the talk: Unlocking the business value in unstructured information*. Pearson Education.

345. Spekman, R. E., & Johnston, W. J. (1986), “Relationship management: Managing the selling and the buying interface”, *Journal of Business Research*,vol.14. no. 6, pp. 519-531.

346. SPSS, (1999), *AnswerTree 2.0, User’s Guide*, available at source from the web: http://www.uic.edu/classes/idsc/ids422/trees.pdf, (accessed June, 3, 2012)

347. Srivastava, R. K., Shervani, T.A. and Fahey, L. (1999), “Marketing, Business Processes, and Shareholder Value: An Organisationally Embedded View of Marketing Activities and the Discipline of Marketing,” *Journal of Marketing*, vol. 63 (Special Issue), pp. 168–79.

348. Stahlbock, R., Crone, S. F., & Lessmann, S. (2010). *Data Mining: Special Issue in Annals of Information Systems* (Vol. 8). Springer.

349. Steinley, D. (2006), ”K-means clustering: A half-century synthesis”, *British Journal of Mathematical and Statistical Psychology*,vol. 59, no. 1,pp. 1-34.

350. Stevens, M. S. (1966), "Direct Mail Marketing: What Is It?," *The Reporter of Direct Mail Advertising,* July.

351. Stone, M. and Woodcock N.(2001), “Defining CRM and Assessing its Quality,” in *Successful Customer Relationship Marketing*, Brian Foss and Merlin Stone Kogan, London

352. Swift, R. (2001), *Accelerating customer relationships: Using CRM and relationship technologies,* Prentice Hall, Upper Saddle River, NJ.

353. Tarokh, M. J., & Ghahremanloo, H. (2007), “Intelligence CRM: A Contact Center Model”, *Proceedings of IEEE International Conference on Service Operations and Logistics, and Informatics*

354. Tay, F., & Shen, L. (2002), “A modified Chi2 algorithm for discretization”, IEEE *Transactions on Knowledge and Data Engineering*, vol.14, no. 3, pp. 666–670.

355. Thieme, R. J., Song, M. and Calantone, R. J. (2000), "Artificial neural network decision support systems for new product development project selection", *Journal of Marketing Research,* vol. 37, no. 4, pp. 499-507.

356. The Sales Educators (2006). Stra*tegic sales leadership: BREAKthrough thinking for BREAKthrough results,* Thomson, Mason, OH:

357. Thomas, L. C. (2009). Consumer Credit Models: Pricing, Profit and Portfolios: Pricing, Profit

and Portfolios , Oxford, UK, Oxford University Press

358. Thompson, G. et al., 1991. *Markets, Hierarchies and Networks*, Sage, London

359. Thrasher, R. P. (1991), "CART: A Recent Advance in Tree-Structured List Segmentation Methodology", *Journal of Direct Marketing,* vol. 5, no. 1, pp. 35-47.

360. Tranfield, D., Denyer, D. and Smart, P. (2003), "Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review", *British Journal of Management,* vol. 14, no. 3, pp. 207-222.

361. Urbanskienė, R., Žostautienė, D., & Chreptavičienė, V. (2008), “The Model of Creation of Customer Relationship Management (CRM) System.”, *Ryšių su klientas valdymo (crm) sistemos kūrimo modelis.*, vol. 58, no.3,pp 51-57

362. Van den Poel, D. (2003), “Predicting mail-order repeat buying: which variables matter?”, *Tijdschrift voor economie en management*, vol. 48, no. 3, pp.371-404.

363. Varey, R. (2002), *Relationship Marketing, Dialogue and Networks in the E-commerce Era*, Wiley, London.

364. Venkatesan, R., Krishnan, T.V., Kumar, V., (2004), “Evolutionary Estimation of Macro-Level Diffusion Models Using Genetic Algorithms: An Alternative to Nonlinear Least Squares,” *Marketing Science*,vol. 23, no. 3, pp. 451–64.

365. Venkatesan, R., Kumar, V., & Bohling, T. (2007), “Optimal CRM using Bayesian decision theory: an application for customer selection”, *Journal of Marketing Research*, pp.579-594.

366. Venkatesan, R. and Kumar, V. (2004), "A customer lifetime value framework for customer selection and resource allocation strategy", *Journal of Marketing,* vol. 68, no. 4, pp. 106-125.

367. Venkatesan, R., Kumar, V., & Bohling, T. (2007), “Optimal customer relationship management using Bayesian decision theory: An application for customer selection”, *Journal of Marketing Research*, vol. *44, no.* (4), pp. 579-594.

368. Verhoef, P. C., & Donkers, B. (2001), “Predicting customer potential value: An application in the insurance industry”, Decision Support Systems, vol. 32, no. 2, pp.189–199.

369. Verhoef, P. C., Hoekstra, J. C. and van Aalst, M. (2002), "The effectiveness of direct response radio commercials Results of a field experiment in The Netherlands", *European Journal of Marketing,* vol. 34, no. 1/2, pp. 143.

370. Verhoef, P. C., Spring, P. N., Hoekstra, J. C. and Leeflang, P. S. H. (2003), "The commercial use of segmentation and predictive modeling techniques for database marketing in the Netherlands", *Decision Support Systems,* vol. 34, no. 4, pp. 471-481.

371. Verhoef, P. C., & Lemon, K. N. (2012), “Successful customer value management: Key lessons and emerging trends”, *European Management Journal*.

372. Vogel, V., Evanschitzky, H., & Ramaseshan, B. (2008), ”Customer equity drivers and future sales”, *Journal of Marketing*, vol. 72, no.6, pp. 98-108.

373. Waterschoot, W. and Van Den Bulte (1992) "The 4P Classification of the Marketing Mix Revisited" , *Journal of Marketing*: Vol. 56, October, pp.83-93

374. Webster, F.E. (1992), “The changing role of marketing in the corporation”, *Journal of Marketing*, vol. 56, no. 4, pp. 1-17.

375. Webster, F. E. (2005), “A Perspective on the Evolution of Marketing Management”, *Journal of Public Policy & Marketing*,vol. 24, Spring, pp121-126

376. Wedel, M. and Kamakura, W. (2000), *Market Segmentation: Conceptual and Methodological Foundations,* Kluwer Academic Publishing, Norwell.

377. Weiss, S. and Kulikowski, C. (1991). *Computer Systems that Learn*, Morgan Kaufman Publishers.

378. Weiss, S. M., & Indurkhya, N. (1998). *Predictive data mining: a practical guide*. Morgan Kaufmann.

379. Winer, R. S. (2001), “A framework for customer relationship management”, *California Management Review*, vol. 43, no. 4, pp. 89−105.

380. Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001),”Knowledge management and Data Mining for marketing”, *Decision Support Systems*, vol. 31, no. 1,pp. 127-137.

381. Wensley, R. (1995), “A critical review of research in marketing”, *British Journal of Management*, vol. 6, December, pp. S63-S82.

382. West, P. M., Brockett, P. L. and Golden, L. L. (1997), "A comparative analysis of neural networks and statistical methods for predicting consumer choice", *Marketing Science,* vol. 16, no. 4, pp. 370-391.

383. Wilkie, W. L. and Moore E.S. (2003), “Scholarly Research in Marketing: Exploring the ‘4 Eras’ of Thought Development,” *Journal of Public Policy & Marketing*, vol. 22, Fall, pp.116–46.

384. Williamson, Oliver E. (1985), *The Economic Institution of Capitalism*. The Free Press, New York, NY

385. Wilson, D. T., & Jantrania, S. (1994), “Understanding the value of a relationship”, *Asia-Australia Marketing Journal*, vol. *2,* no.1, pp. 55-66.

386. Wind, Y. (1978), “Issues and advances in segmentation research,” *Journal of Marketing Research*, Vol. XV, August, pp. 317-37.

387. Winer, R. S. (2001), “A framework for customer relationship management”, *California management review*, vol. *43, no.* 4.

388. Yang, A. X. (2004), "How to develop new approaches to RFM segmentation", *Journal of Targeting, Measurement and Analysis for Marketing,* vol. 13, pp. 50-60.

389. Yankelovich, D. (1964), "New Criteria for Market Segmentation", *Harvard business review,* vol. 42, no. 2, pp. 83-90.

390. Yankelovich, D. and Meer, D. (2006), "Rediscovering Market Segmentation", *Harvard business review,* vol. 84, no. 2, pp. 122-131.

391. Yoo, B. and Donthu, N. (2001), "Developing and validating a multidimensional consumer-based brand equity scale", *Journal of Business Research,* vol. 52, no. 1, pp. 1-14.

392. Zablah, A. R., Bellenger, D. N., & Johnston, W. J. (2004), “An evaluation of divergent perspectives on customer relationship management: Towards a common understanding of an emerging phenomenon”, *Industrial Marketing Management*, vol. 33, no. 6, pp. 475−489.

393. Zahavi, J. and Levin, N. (1997), "Applying neural computing to target marketing", *Journal of Direct Marketing,* vol. 11, no. 1, pp. 76-93.

394. Zahay, D., Peltier, J., Schultz, D. E., & Griffin, A. (2004), “The role of transactional versus relational data in IMC programs: Bringing customer data together”, *Journal of advertising research*, vol. *44, no.* 1, pp. 3-18.

395. Zahay, D., Mason, C. H., Schibrowsky J. A., (2009)“The Present and Future of IMC and Database Marketing”, *International Journal of Integrated Marketing Communications*, Fall

396. Zikmund, W., McLeod, R., & Gilbert, F. (2003), *Customer relationship management: integrating marketing strategy and information technology*: Wiley, New York, NY

9 Appendices

Appendix 1: Categorisation of CRM Definitions

			Payne and Frow (2005)			Richard and Jones (2008) Tanner (2004)		Zablah, Bellenger, and Johnston (2004)				
Definition	**Year**	**Source**	**Narrow & Tactical Technology Solution**	**Wide-Ranging Technology**	**Customer Centric Strategy**	**Strategic**	**Operational/ Analytical**	**Process**	**Strategy**	**Philosophy**	**Capability**	**Technology**
CRM is the marketing field aimed at establishing, developing, maintaining and enhancing relationships with customers	1990	Grönroos			X	X			X			
CRM attempts to provide a strategic bridge between information technology and marketing strategies aimed at building long-term relationships and profitability. This requires “information-intensive strategies”	1997	Glazer		X	X	X	X		X		X	
CRM is data-driven marketing	1997	Kutner and Cripps			X	X			X			
CRM can be viewed as an application of one-to-one marketing and RM, responding to an individual customer on the basis of what the customer says and what else is known about that customer	1999	Peppers, Rogers, and Dorf			X		X				X	
CRM is a management approach that enables organisations to identify, attract, and increase retention of profitable customers by managing relationships with them	1999	Hobby			X		X	X				
CRM involves using existing customer information to improve company profitability and customer service	1999	Couldwell			X	X			X			
CRM is a macro level process that subsumes numerous subprocesses, such as prospect identification and customer knowledge creation.	1999	Srevastava, Shervani and Fahey			X		X	X				

			Payne and Frow (2005)			**Richard and Jones (2008) Tanner (2004)**		**Zablah, Bellenger, and Johnston (2004)**				
Definition	**Year**	**Source**	**Narrow & Tactical Technology Solution**	**Wide-Ranging Technology**	**Customer Centric Strategy**	**Strategic**	**Operational/ Analytical**	**Process**	**Strategy**	**Philosophy**	**Capability**	**Technology**
CRM is an enterprise approach to understanding and influencing customer behaviour through meaningful communications in order to improve customer acquisition, customer retention, customer loyalty, and customer profitability	2000	Swift			X	X			X	X		
CRM includes numerous aspects, but the basic theme is for the company to become more customer-centric. Methods are primarily Web-based tools and Internet presence	2000	Gosney and Boehm	X		X	X				X	X	
CRM is a relationship orientation, customer retention and superior customer value created through process management	2001	Ryals & Knox			X	X		X				
CRM is a comprehensive strategy and process of acquiring, retaining, and partnering with selective customers to create superior value for the company and the customer. It involves the integration of marketing, sales, customer service, and the supply-chain functions of the organisation to achieve greater efficiencies and effectiveness in delivering customer value	2001	Parvatiyar & Sheth			X	X			X			
CRM is an e-commerce application	2001	Khanna	X				X					X
CRM is a term for methodologies, technologies, and ecommerce capabilities used by companies to manage customer relationships	2001	Stone and Woodcock		X			X					X
CRM is a comprehensive strategy and process of acquiring, retaining, and partnering with selective customers to create superior value for the company and the customer	2001	Parvitiyar and Sheth			X	X			X			

			Payne and Frow (2005)			Richard and Jones (2008) Tanner (2004)		Zablah, Bellenger, and Johnston (2004)				
Definition	**Year**	**Source**	**Narrow & Tactical Technology Solution**	**Wide-Ranging Technology**	**Customer Centric Strategy**	**Strategic**	**Operational/ Analytical**	**Process**	**Strategy**	**Philosophy**	**Capability**	**Technology**
CRM is about the development and maintenance of long-term, mutually beneficial relationships with strategically significant customers	2001	Buttle			X	X			X	X		
CRM is an enterprise-wide integration of technologies working together , such as data warehouse, web site, intranet/extranet, phone support system accounting, sales, marketing, and production	2002	Bose		X			X					X
CRM aligns business processes with customer strategies to build customer loyalty and increase profits over time	2002	Rigby, Reichheld, & Schefter			X	X		X				
CRM allows companies to gather customer data swiftly, identify the most valuable customers over time, and increase customer loyalty by providing customised products and services	2002	Rigby, Reichheld, and Schefter.			X	X			X			
CRM is a philosophy of doing business that will affect the entire enterprise	2003	Newell			X	X				X		
CRM is a customer-focused business strategy that aims to increase customer satisfaction and customer loyalty by offering a more responsive and customised services to each customer	2003	Croteau & Li			X	X			X			
CRM is a strategy used to learn more about customer's needs and behaviours in order to develop stronger relationship with them	2003	Gupta & Lehmann			X	X			X			

			Payne and Frow (2005)			Richard and Jones (2008) Tanner (2004)		Zablah, Bellenger, and Johnston (2004)				
Definition	**Year**	**Source**	**Narrow & Tactical Technology Solution**	**Wide-Ranging Technology**	**Customer Centric Strategy**	**Strategic**	**Operational/ Analytical**	**Process**	**Strategy**	**Philosophy**	**Capability**	**Technology**
CRM is the strategic use of information, processes, technology, and people to manage the customer's relationship with a company across the whole customer life cycle	2003	Kincaid			X	X			X			
A process to compile information that increases understanding of how to manage an organisation's relationships with its customers	2003	Zikmund et al.			X	X		X				
Resources destined for relationship building and maintenance efforts should be allocated based on customers' lifetime value to the firm	2003	Ryals			X					X		
CRM is leveraging technology to engage individual customers in a meaningful dialogue so that firms can customise their products and services to attract, develop, and retain customers	2003	Campbell		X			X					X
CRM is an enterprise-wide initiative that belongs in all areas of an organisation	2003	Singh and Agrawal			X		X			X		
CRM is a business strategy designed to optimise profitability, revenue and customer satisfaction by organising the enterprise around customer segments, fostering customer-centric behaviours and implementing customer-centric processes	2003	Gartner Inc.			X	X			X			
CRM is the strategic use of information, processes, technology, and people to manage the customer's relationship with a company across the whole customer life cycle.	2003	Kincaid			X	X			X			

			Payne and Frow (2005)			Richard and Jones (2008) Tanner (2004)		Zablah, Bellenger, and Johnston (2004)				
Definition	**Year**	**Source**	**Narrow & Tactical Technology Solution**	**Wide-Ranging Technology**	**Customer Centric Strategy**	**Strategic**	**Operational/ Analytical**	**Process**	**Strategy**	**Philosophy**	**Capability**	**Technology**
CRM is the practice of leveraging technology to engage individual customers in a meaningful dialogue so that firms can customise their products and services to attract, develop, and retain customers".	2003	Campbell		X			X					X
CRM refers to the idea that the most effective way to achieve loyalty is by proactively seeking to build and maintain long term relationships with customers	2004	Zablah et al., 2004			X	X				X		
CRM is a technology or software solution that helps track data and information about customer to enable better customer service	2004	Peppers & Rogers		X			X					X
CRM is the business process in which customer equity (i.e., aggregate lifetime value of a firm's existing and potential customers) is continuously created, enhanced, and managed by interacting with customers through multiple channels	2004	Kumar and Reinartz			X	X		X				
Management of mutually beneficial relationship(s) from the seller's perspective	2004	LaPlaca			X	X			X			
A systematic process to manage customer relationship initiation, maintenance, and termination across all customer contact points in order to maximise the value of the relationship portfolio	2004	Reinartz, Krafft, and Hoyer			X	X		X				
CRM is the outcome of the continuing evolution and integration of marketing ideas and newly available data, technologies, and organisational forms.	2005	Boulding			X	X					X	
CRM is a strategic approach for systematically targeting, tracking, communicating, and transforming relevant customer data into actionable information on which strategic decision-making is based	2005	Karakostas et al.			X	X		X				

			Payne and Frow (2005)			Richard and Jones (2008) Tanner (2004)		Zablah, Bellenger, and Johnston (2004)				
Definition	**Year**	**Source**	**Narrow & Tactical Technology Solution**	**Wide-Ranging Technology**	**Customer Centric Strategy**	**Strategic**	**Operational/ Analytical**	**Process**	**Strategy**	**Philosophy**	**Capability**	**Technology**
CRM is a strategic approach that is concerned with creating improved shareholder value through the development of appropriate relationships with key customers and customer segments. CRM unites the potential of RM strategies and IT to create profitable, long-term relationships with customers and other key stakeholders. CRM provides enhanced opportunities to use data and information to both understand customers and to create value with them. This requires a cross functional integration of processes, people, operations, and marketing capabilities that is enabled through information, technology, and applications	2005	Payne & Frow			X	X		X				
CRM technology is a suite of information technology-based solutions designed to support the customer relationship management process	2005	Jayachandran, Sharma, Kaufman, ,Raman	X				X				X	
The process that identifies customers, creates customer knowledge, builds customer relationships, and shapes customers' perceptions of the firm and its products/ solutions	2006	The Sales Educators			X	X		X				
CRM is a strategy used to learn more about customers' needs and behaviours in order to develop stronger relationships with them.	2007	Tarokh & Ghahremanloo			X	X			X			
CRM is the complex of software and technologies, automating and performing business processes in the following areas: sales, marketing, service, and customer support	2008	Urbanskiene et al.		X			X					X
CRM is the philosophy, policy and coordinating strategy mediated by a set of information technologies, which focuses on creating two way communications with customers so that firms have an intimate knowledge of their needs, wants, and buying patterns	2008	Lun et al.		X	X	X	X			X		X

			Payne and Frow (2005)			Richard and Jones (2008) Tanner (2004)		Zablah, Bellenger, and Johnston (2004)				
Definition	**Year**	**Source**	**Narrow & Tactical Technology Solution**	**Wide-Ranging Technology**	**Customer Centric Strategy**	**Strategic**	**Operational/ Analytical**	**Process**	**Strategy**	**Philosophy**	**Capability**	**Technology**
CRM is a customer-centred enterprise management mode, which discovers the customers' value and satisfies their requirements to realise the interaction between enterprise management and customers	2009	Huang & Wang			X	X				X		
CRM is a key business strategy in which a firm needs to stay focused on the needs of its customers and must integrate a customer-oriented approach throughout the organisation	2009	Liou			X	X			X			
CRM is an enabling technology for organisations to foster closer relationships with their customers	2009	Hsieh		X			X					X
CRM is a cross-functional strategic approach concerned with creating improved shareholder value through the development of appropriate relationships with key customers and customer segments. It typically involves identifying appropriate business and customer strategies, the acquisition and diffusion of customer knowledge, deciding appropriate segment granularity, managing the co-creation of customer value, developing integrated channel strategies and the intelligent use of data and technology solutions to create superior customer experiences	2009	Frow and Payne			X	X			X			
A more expansive and holistic approach in developing sound and productive relationships with customers, while CRM technology, one of major components of CRM, has been defined as the information technology that is deployed for the specific purpose of managing customer relationships	2010	Chang, Park and Chaiy			X	X		X				
CRM is a managerial strategy that helps organisations collect, Analyse, and manage customer related information through the use of information technology tools and techniques in order to satisfy customer needs and establish a long term and mutually beneficial relationship	2010	Hung et al.,		X	X	X			X		X	X

			Payne and Frow (2005)			Richard and Jones (2008) Tanner (2004)		Zablah, Bellenger, and Johnston (2004)				
Definition	**Year**	**Source**	**Narrow & Tactical Technology Solution**	**Wide-Ranging Technology**	**Customer Centric Strategy**	**Strategic**	**Operational/ Analytical**	**Process**	**Strategy**	**Philosophy**	**Capability**	**Technology**
CRM involved the evolution and integration of marketing ideas and newly available data, technologies and organisational forms, and it concentrates on establishing, developing, and maintaining successful long-term relationships with well-chosen customers	2010	Ernst, Hoyer, Krafft, and Krieger			X	X			X			

	Payne and Frow (2005)			Richard and Jones (2008) Tanner (2004)		Zablah, Bellenger, and Johnston (2004)				
Definition / **Year** / **Source**	**Narrow & Tactical Technology Solution**	**Wide-Ranging Technology**	**Customer Centric Strategy**	**Strategic**	**Operational/ Analytical**	**Process**	**Strategy**	**Philosophy**	**Capability**	**Technology**
TOTAL COUNT 1990-2000	1	1	9	6	4	2	5	2	3	0
TOTAL COUNT 2001-2005	2	5	20	18	8	5	10	5	2	6
TOTAL COUNT 2006-2013	0	4	9	9	3	2	5	2	1	4
TOTAL COUNT	2	10	36	31	14	10	18	9	4	10
TOTAL % 1990-2000	9%	9%	82%	60%	40%	20%	50%	20%	30%	0%
TOTAL % 2001-2005	7%	19%	74%	69%	31%	22%	43%	22%	9%	26%
TOTAL % 2006-2013	0%	31%	69%	75%	25%	17%	42%	17%	8%	33%
TOTAL %	4%	21%	75%	69%	31%	24%	44%	22%	10%	24%

Appendix 2: Definitions of Data Mining Techniques

Descriptive and Visualisation Techniques

These techniques include “simple descriptive statistics such as averages and measures of variation, counts and percentages, and cross-tabs and simple correlations. They are useful for understanding the structure of the data. Visualisation is primarily a discovery technique and is useful for interpreting large amounts of data; visualisation tools include histograms, box plots, scatter diagrams, and multi-dimensional surface plots” (Jackson, 2002, p. 274).

RFM

“A well known, empirically based RFM technique is a procedure advocated by Arthur Hughes (2000). Hughes' approach is applicable in instances when a marketer intends to send a mailing to customers in its database and would like to find those in the database who are the most likely to respond to the specific mailing. Hughes recommends a test mailing to a sample of customers in the file; then the selection of the members of the rest of the file is made as a function of the results of the test …. The first step in the approach is for the marketer to sort the customer file according to how recently customers have purchased from the firm. The database is then divided into equal quintiles and these quintiles are assigned the numbers 5 to 1. Therefore, the 20% of the customers who most recently purchased from the company are assigned the number 5; the next 20% are assigned the number 4, and so on. The next step involves sorting the customers within each recency quintile by how frequently they purchase from the marketer. For each of these sorts, the customers are divided into equal quintiles and assigned a number of 5 to 1 for frequency. Each of these groups (25 groups) is sorted according to how much money the customers have spent with the company. These sorts are divided into quintiles and assigned numbers 5 to 1. Therefore, the database is divided into 125 roughly equal groups (cells) according to recency, frequency, and monetary value. Hughes recommends conducting a test mailing to a randomly sampled subset of each cell (e.g., 10%). After the responses of the test mailing are received, the proportion of respondents in each cell can be calculated. The cells can then be ordered as a function of response percent. The marketer can then elect to mail to a certain portion of the remaining file (e.g., the top 20% of the cells). Alternatively, the marketer can elect to mail to the cells that are above a breakeven percent, given the cost of the mailing and the expected revenue for each return.” (McCarty and Hastak, 2006, p.657)

Cluster analysis

Cluster analysis assigns data points or individuals into groups (otherwise referred to as clusters) so that those same data points or individuals from the same cluster are much closer to one another than data points or individuals from different clusters. Such clusters can be organised to be non-overlapping, overlapping or fuzzy.

Correlation Analysis

This technique "measures the relationship between two variables. The resulting correlation coefficient shows if changes in one variable will result in changes in the other. When comparing the correlation between two variables, the goal is to see if a change in the independent variable will result in a change in the dependent variable. This information helps in understanding an independent variable's predictive abilities. Correlation findings, just as regression findings, can be useful in Analysing causal relationships, but they do not by themselves establish causal patterns" (Jackson, 2002, p. 274).

Regression Analysis

"Statistical tool that uses the relation between two or more quantitative variables so that one dependent variable can be predicted from other independent variables. Regardless of the statistical relations that exist between the variables, no cause-and-effect pattern is implied by the regression model. Regression analysis comes in many forms including: simple linear regression, multiple linear regression, curvilinear regression, and multiple curvilinear regression, as well as logistic regression" (Jackson, 2002, p. 274).

Ridge Regression

"Variant of ordinary Multiple Linear Regression whose goal is to circumvent the problem of predictor colinearity. It gives-up the Least Squares (LS) as a method for estimating the parameters of the model, and focuses instead of the X'X matrix. This matrix will be artificially modified so as to make its determinant appreciably different from 0. By doing so, it makes the new model parameters somewhat biased (whereas the parameters as calculated by the LS method are unbiased estimators of the true parameters). Moreover, the predictions errors of the Ridge Model will also turn out to be more accurate than that of the LS regression model when predictors exhibit near co linearity. Therefore, the idea behind of Ridge Regression is at the heart of the "bias-variance trade-off" issue."

Source: http://www.aiaccess.net/English/Glossaries/GlosMod/e_gm_ridge.htm, accessed April 15, 2010

Logistic Regression

Technique that is "used when the response variable is a binary variable or has a qualitative outcome. Although logistic regression "finds a "best fitting" equation just as linear regression does, the principles on which it does so are rather different. Instead of using a least-squared deviations criterion for the best fit, it uses a maximum likelihood method, that is, it maximises the probability of obtaining the observed results given the fitted regression coefficients. Because logistic regression does not make any assumptions about the distribution for the independent variables, it is more robust to violations of the normality assumption" (Jackson, 2002, p 275). Some of the common types of logistic regression include: simple, multiple, polytomous and Poisson logistic regression models.

Factor Analysis

Technique "useful for understanding the underlying reasons for the correlations among a group of variables. The main applications of factor analytic techniques are to reduce the number of variables and to detect structure in the relationships among variables; that is to classify variables. Therefore, factor analysis can be applied as a data reduction or structure detection method. In an exploratory factor analysis, the goal is to explore or search for a factor structure. Confirmatory factor analysis, on the other hand, assumes the factor structure is known a priori and the objective is to empirically verify or confirm that the assumed factor structure is correct" (Jackson, 2002, p. 274)..

Discriminant Analysis

Technique "used to predict membership in two or more mutually exclusive groups from a set of predictors, when there is no natural ordering on the groups. Discriminant analysis can be seen as the inverse of a one-way multivariate analysis of variance (MANOVA) in that the levels of the independent variable (or factor) for MANOVA become the categories of the dependent variable for discriminant analysis, and the dependent variables of the MANOVA become the predictors for discriminant analysis" (Jackson, 2002, p. 274).

Principal components analysis

"Principal component analysis is a variable reduction procedure. It is useful when you have obtained data on a number of variables (possibly a large number of variables), and believe that there is some redundancy in those variables. In this case, redundancy means that some of the variables are correlated with one another, possibly because they are measuring the same construct. Because of this redundancy, you believe that it should be possible to reduce the observed variables into a smaller number of principal components (artificial variables) that will account for most of the variance in the observed variables.

Because it is a variable reduction procedure, principal component analysis is similar in many respects to exploratory factor analysis. In fact, the steps followed when conducting a principal component analysis are virtually identical to those followed when conducting an exploratory factor analysis. However, there are significant conceptual differences between the two procedures, and it is important that you do not mistakenly claim that you are performing factor analysis when you are actually performing principal component analysis" (Hatcher, 1994, p 2-3)

Machine Learning Techniques

Case-Based Learning

"Case-Based Reasoning (CBR) is a technology that tries to solve a given problem by making direct use of past experiences and solutions. A case is usually a specific problem that was encountered and solved previously. Given a particular new problem, CBR examines the set of stored cases and finds similar ones. If similar cases exist, their solution is applied to the new problem, and the problem is added to the case base for future reference" (Jackson, 2002, p. 273).

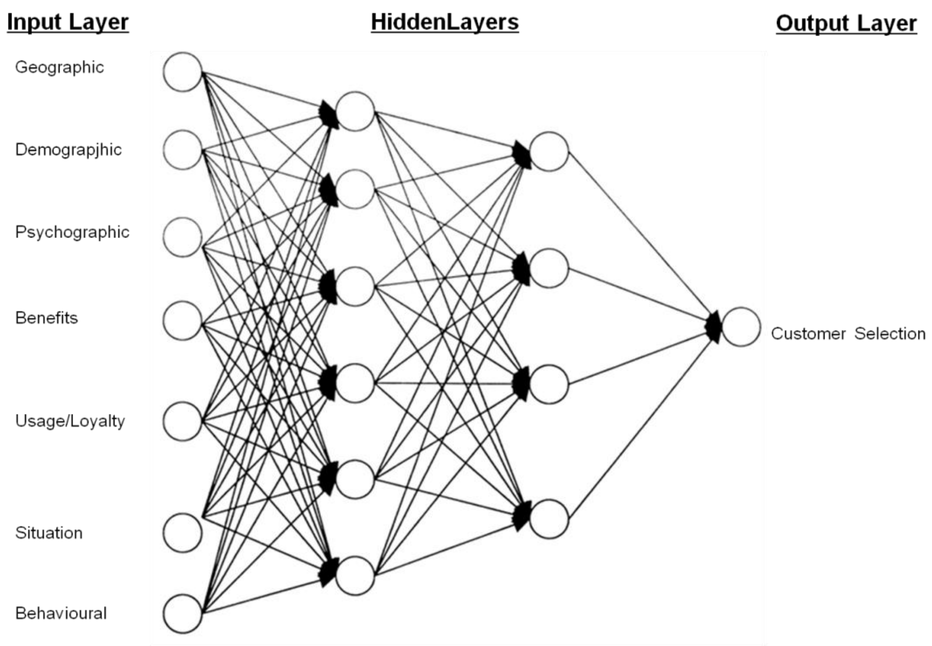
Genetic Algorithms

"Genetic Algorithms (GA) operate through procedures modelled upon the evolutionary biological processes of selection, reproduction, mutation, and survival of the fittest to search for very good solutions to prediction and classification problems. GA are used in Data Mining to formulate hypotheses about dependencies between variables in the form of association rules or some other internal formalism" (Jackson, 2002, p. 273).

Neural Networks

Crone (2006, p. 783) defines neural networks as "a class of statistical methods capable of universal function approximation, learning non-linear relationships between independent and dependent variables directly from the data without previous assumptions about the statistical distributions. Neural networks are modelled after the human brain. As such, they accept a number of inputs and process these inputs in order to forecast a foreseeable output. Just like a human brain, some of the association between these inputs and outputs are transparent but many of these associations are hidden. As such, the neural networks develop hidden layers between input and output layers in order to best approximate the output function. Each of these inputs and hidden layers are called nodes. Given some of these inputs can be linked to a common hidden phenomenon or node, nodes take one or more input values and combine them into a single value. This facilitates the transformation of this value into an output value". This is illustrated in Figure 37.

Figure 37: Neural Network for Customer Selection Decisions Schematic



Neural networks can be either supervised or unsupervised. The more common of the two techniques is supervised neural networks, where samples of data containing both inputs and outputs are entered for analysis purposes. Each sample item outcome is compared to a desired output to determine item importance. This is repeated until performance is acceptable. Unsupervised networks work in much the same way except no output data is fed in. Rather the network is trained to detect salient features that are then used to group inputs into different classes. As a result, it is not surprising to note that in general neural networks with supervised learning algorithms can be used in place of regression and discriminant analysis…whereas, neural networks with unsupervised learning can be used for grouping purposes (Venugopal and Baets, 1994).

Decision Trees

Decision trees are "intuitive methods for classifying a pattern through a sequence of rules or questions, in which the next question depends on the answer on a current question. They are particularly useful for categorical data, as rules do not require any notion of metric" (Crone et al., 2006, p. 784).

"Basically, all automatic tree classifiers share the same structure. Starting from a "root" node (the whole population), tree classifiers employ a systematic approach to grow a tree into "branches" and "leaves." In each stage, the algorithm looks for the "best" way to split a "father" node into several "children" nodes, based on some splitting criteria. Then, using a set of predefined termination rules, some nodes are declared as “undetermined” and become the father nodes in the next stages of the tree development process, some others are declared as “terminal” nodes. The process proceeds in this way until no more node left in the tree is worth splitting any further. The terminal nodes define the resulting segments” (Levin and Zahavi, 1996, p. 1272).

In data mining, trees fall into a number of different categories:

- Classification And Regression Trees (CART)
- Chi-squared Automatic Interaction Detector (CHAID
- Random Forest

Association Rules

“Association Rules (AR) are statements about relationships between the attributes of a known group of entities and one or more aspects of those entities that enable predictions to be made about aspects of other entities who are not in the group, but who possess the same attributes. More generally, AR state a statistical correlation between the occurrences of certain attributes in a data item, or between certain data items in a data set” (Jackson, 2002, p. 273).

Appendix 3: Technique Application Outputs

RFM Variables 1 MTH vs 12 MTH

Average Response Rate from 10 folds

	12 mths	1 mth
Deciles	RFM-12	RFM-1
Top	31.9%	29.8%
2	29.6%	28.5%
3	25.5%	25.6%
4	23.8%	24.1%
5	17.7%	17.4%
6	14.3%	14.2%
7	9.6%	9.7%
8	5.0%	6.8%
9	3.5%	3.0%
Bottom	0.7%	1.3%

Lift Charts

Lift Chart

12 mths	1 mth	
RFM-12	RFM-1	No Model
197	186	100
183	178	100
158	160	100
147	150	100
110	108	100
89	89	100
59	61	100
31	42	100
22	19	100
5	8	100

Cumulative Lift Chart

12 mths	1 mth	
RFM-12	RFM-1	No Model
197	186	100
190	182	100
179	174	100
171	168	100
159	156	100
147	145	100
135	133	100
122	122	100
111	110	100
100	100	100

Gains Charts

Cumulative Gains Chart

12 mths	1 mth	
RFM-12	RFM-1	No Model
19.7%	18.6%	10.0%
38.0%	36.3%	20.0%
53.8%	52.3%	30.0%
68.6%	67.3%	40.0%
79.5%	78.1%	50.0%
88.4%	87.0%	60.0%
94.3%	93.1%	70.0%
97.4%	97.3%	80.0%
99.5%	99.2%	90.0%
100.0%	100.0%	100.0%

RFM Variables

Average Response Rate from 10 folds

Deciles	RFM	CHAID	LR	NN
Top	31.9%	33.1%	31.6%	32.7%
2	29.6%	29.9%	29.4%	29.8%
3	25.5%	26.1%	25.9%	26.0%
4	23.8%	22.1%	21.2%	21.6%
5	17.7%	17.3%	17.7%	17.5%
6	14.3%	14.3%	13.1%	13.7%
7	9.6%	10.0%	8.7%	9.1%
8	5.0%	5.8%	4.8%	5.2%
9	3.5%	3.4%	2.2%	2.8%
Bottom	0.7%	1.2%	0.5%	0.5%

Lift Charts

Lift Chart

RFM	CHAID	LR	NN	No Model
197	203	204	206	100
183	183	189	187	100
158	160	167	164	100
147	135	137	136	100
110	106	114	110	100
89	88	84	86	100
59	61	56	57	100
31	36	31	33	100
22	21	15	17	100
5	7	3	3	100

Cumulative Lift Chart

RFM	CHAID	LR	NN	No Model
197	203	204	206	100
190	193	197	197	100
179	182	187	186	100
171	170	174	173	100
159	157	162	161	100
147	146	149	148	100
135	134	136	135	100
122	121	123	122	100
111	110	111	111	100
100	100	100	100	100

Gains Charts

Cumulative Gains Chart

RFM	CHAID	LR	NN	No Model
0.0%	0.0%	0.0%	0.0%	0.0%
19.7%	20.3%	20.4%	20.6%	10.0%
38.0%	38.6%	39.3%	39.3%	20.0%
53.8%	54.6%	56.0%	55.7%	30.0%
68.6%	68.1%	69.7%	69.3%	40.0%
79.5%	78.7%	81.1%	80.3%	50.0%
88.4%	87.4%	89.5%	88.9%	60.0%
94.3%	93.6%	95.1%	94.7%	70.0%
97.4%	97.1%	98.3%	97.9%	80.0%
99.5%	99.3%	99.7%	99.7%	90.0%
100.0%	100.0%	100.0%	100.0%	100.0%

All Data Variables

Average Response Rate from 10 folds

Deciles	RFM	CHAID	LR	NN
Top	31.9%	50.1%	50.6%	50.4%
2	29.6%	34.9%	33.1%	34.0%
3	25.5%	24.8%	23.4%	22.7%
4	23.8%	18.5%	17.0%	16.8%
5	17.7%	12.6%	12.0%	12.3%
6	14.3%	8.3%	8.7%	8.6%
7	9.6%	6.9%	6.0%	6.6%
8	5.0%	4.2%	3.9%	4.0%
9	3.5%	1.7%	2.6%	2.1%
Bottom	0.7%	0.6%	1.4%	1.1%

Lift Charts

Lift Chart

RFM	CHAID	LR	NN	No Model
197	308	319	318	100
183	215	209	214	100
158	153	147	143	100
147	114	107	106	100
110	77	76	78	100
89	51	55	54	100
59	42	38	42	100
31	26	25	25	100
22	11	17	13	100
5	4	9	7	100

Cumulative Lift Chart

RFM	CHAID	LR	NN	No Model
197	308	319	318	100
190	261	264	266	100
179	225	225	225	100
171	197	195	195	100
159	173	171	172	100
147	153	152	152	100
135	137	136	136	100
122	123	122	122	100
111	111	110	110	100
100	100	100	100	100

Gains Charts

Cumulative Gains Chart

RFM	CHAID	LR	NN	No Model
0.0%	0.0%	0.0%	0.0%	0.0%
19.7%	30.8%	31.9%	31.8%	10.0%
38.0%	52.3%	52.7%	53.2%	20.0%
53.8%	67.5%	67.5%	67.5%	30.0%
68.6%	78.9%	78.2%	78.1%	40.0%
79.5%	86.6%	85.7%	85.9%	50.0%
88.4%	91.7%	91.2%	91.3%	60.0%
94.3%	96.0%	95.0%	95.5%	70.0%
97.4%	98.6%	97.5%	98.0%	80.0%
99.5%	99.6%	99.1%	99.3%	90.0%
100.0%	100.0%	100.0%	100.0%	100.0%

Reduced Dataset

Average Response Rate from 10 folds

Deciles	RFM	CHAID	LR	NN
Top	31.85%	50.14%	65.29%	51.12%
2	29.60%	35.21%	37.83%	35.44%
3	25.53%	25.14%	22.39%	21.98%
4	23.76%	18.64%	13.97%	16.01%
5	17.74%	12.88%	8.82%	12.25%
6	14.31%	8.79%	4.90%	9.01%
7	9.56%	7.29%	2.54%	6.36%
8	4.98%	4.58%	1.50%	4.32%
9	3.47%	2.26%	0.00%	1.99%
Bottom	0.74%	0.98%	0.00%	0.97%

Lift Charts

Lift Chart

RFM	CHAID	LR	NN	No Model
197	302	415	321	100
183	212	241	222	100
158	152	142	138	100
147	112	89	100	100
110	78	56	77	100
89	53	31	57	100
59	44	16	40	100
31	28	10	27	100
22	14	0	12	100
5	6	0	6	100

Cumulative Lift Chart

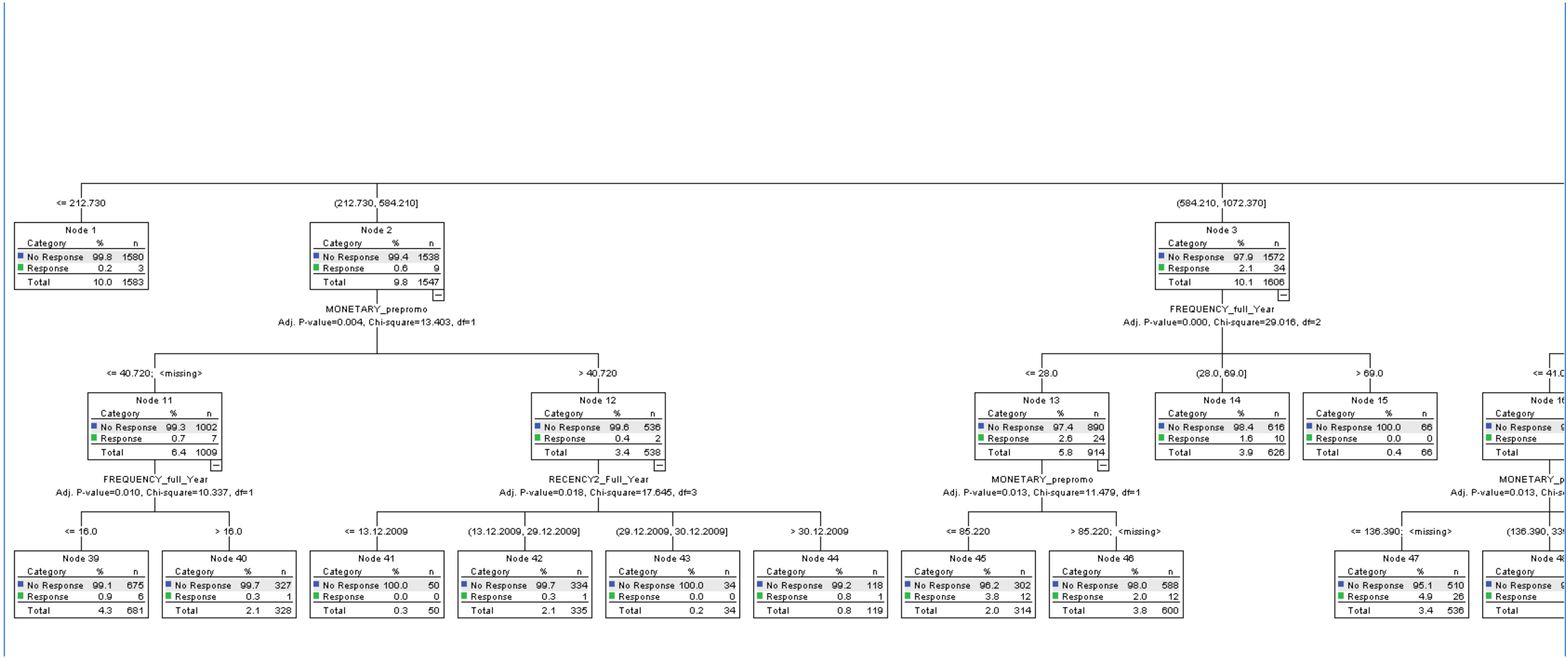
RFM	CHAID	LR	NN	No Model
197	302	415	321	100
190	257	328	271	100
179	222	266	227	100
171	195	222	195	100
159	171	189	172	100
147	151	162	152	100
135	136	141	136	100
122	123	125	123	100
111	110	111	110	100
100	100	100	100	100

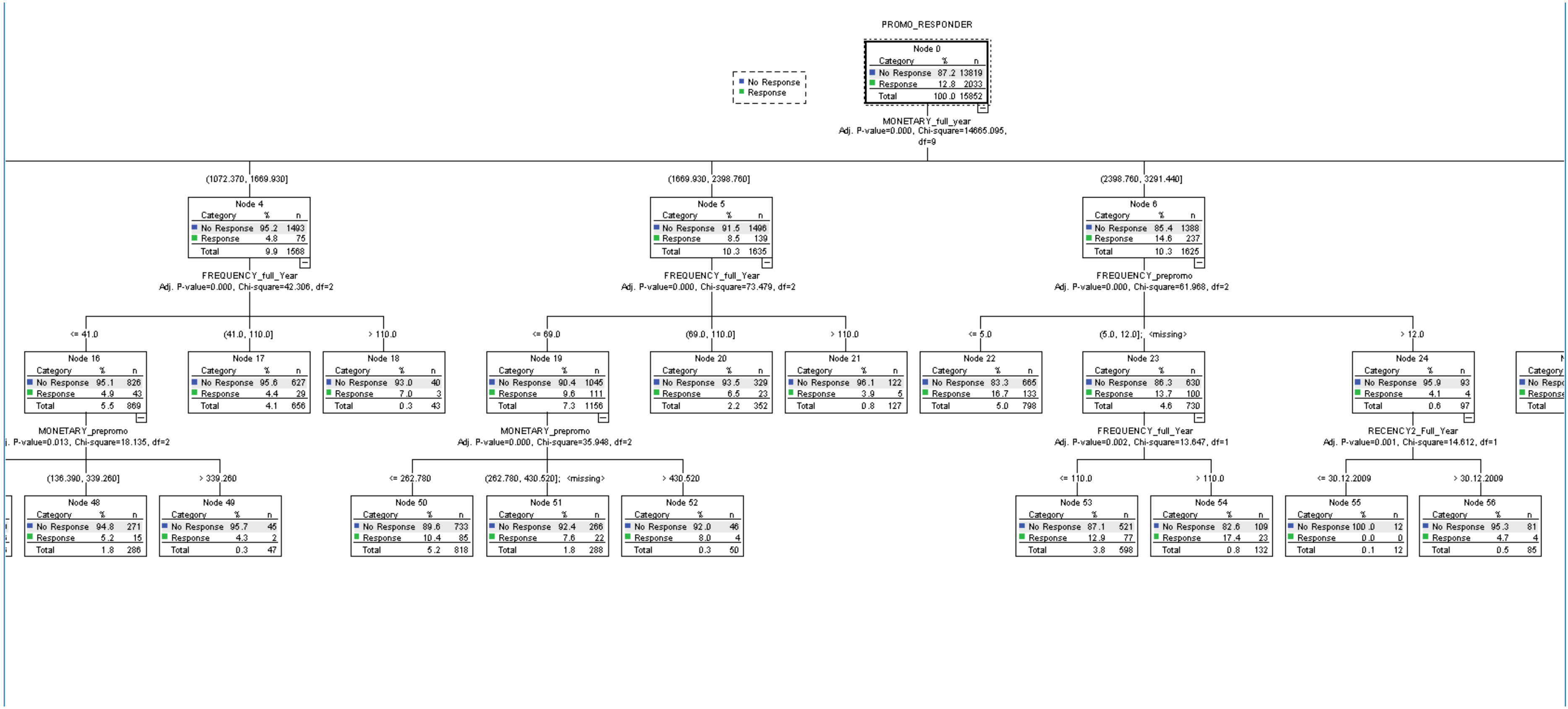
Gains Charts

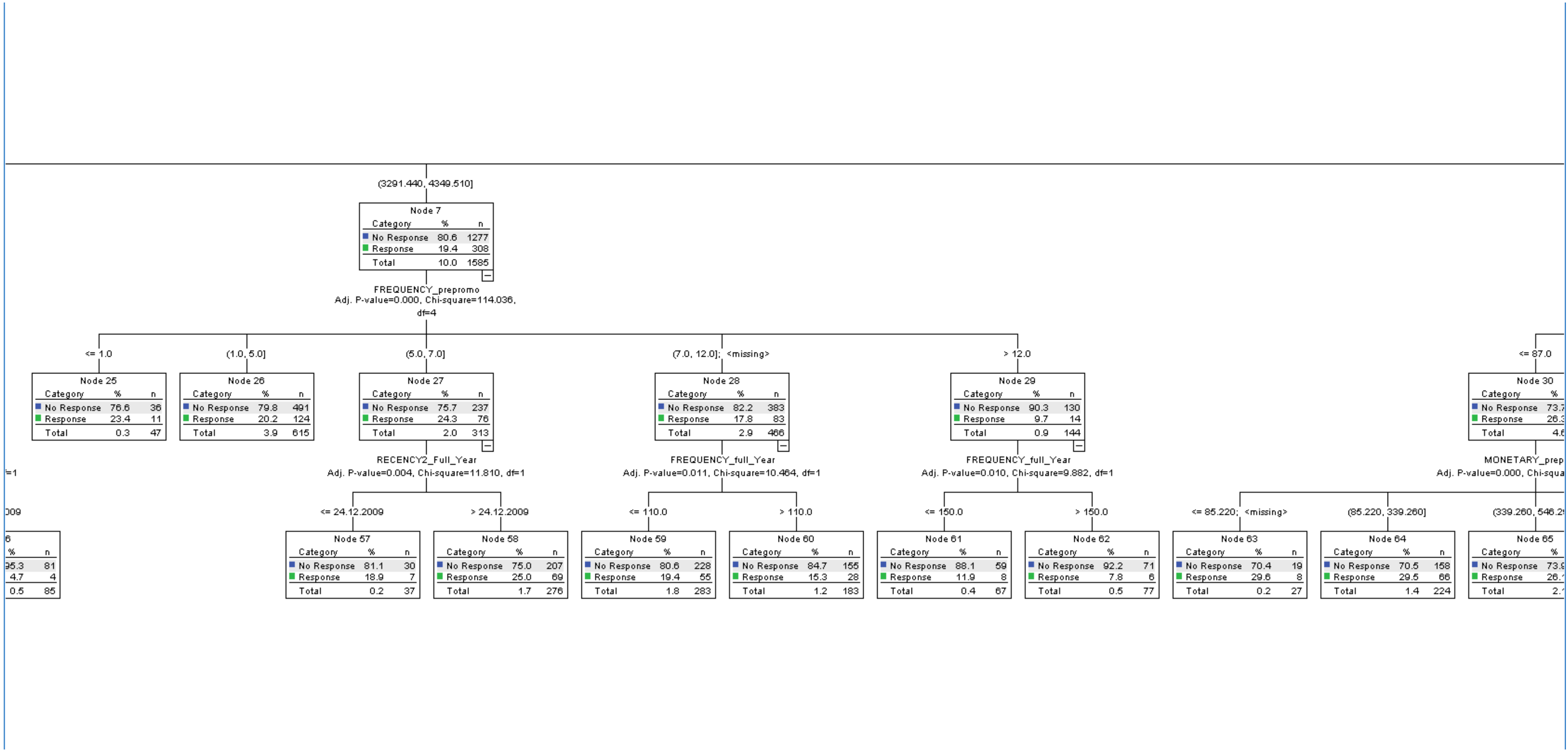
Cumulative Gains Chart

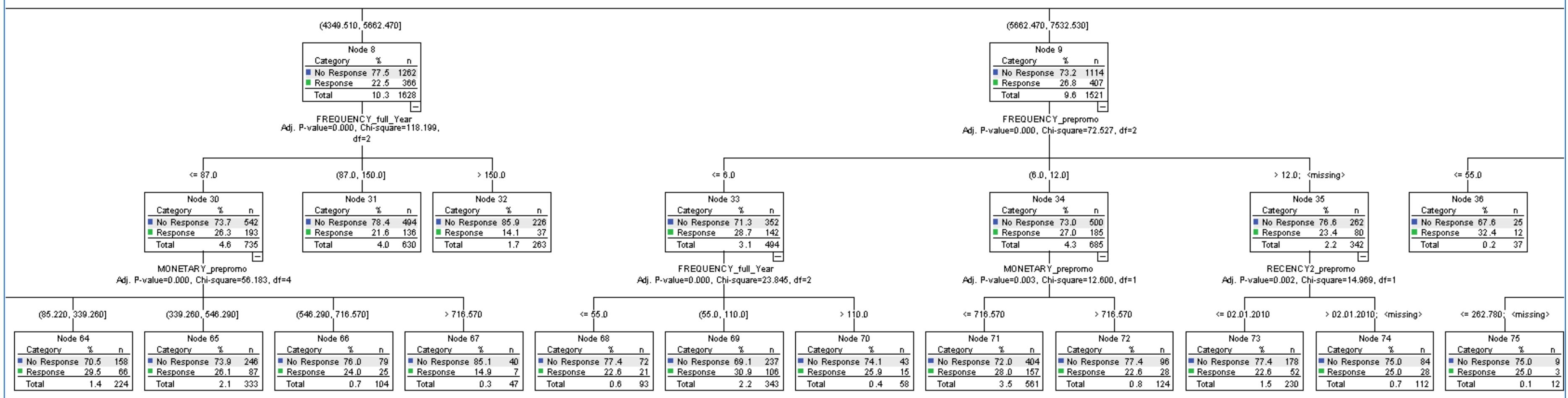
RFM	CHAID	LR	NN	No Model
0.0%	0.0%	0.0%	0.0%	0.0%
19.7%	30.2%	41.5%	32.1%	10.0%
38.0%	51.4%	65.6%	54.3%	20.0%
53.8%	66.6%	79.8%	68.1%	30.0%
68.6%	77.8%	88.7%	78.1%	40.0%
79.5%	85.6%	94.3%	85.8%	50.0%
88.4%	90.9%	97.4%	91.4%	60.0%
94.3%	95.3%	99.0%	95.4%	70.0%
97.4%	98.1%	100.0%	98.1%	80.0%
99.5%	99.4%	100.0%	99.4%	90.0%
100.0%	100.0%	100.0%	100.0%	100.0%

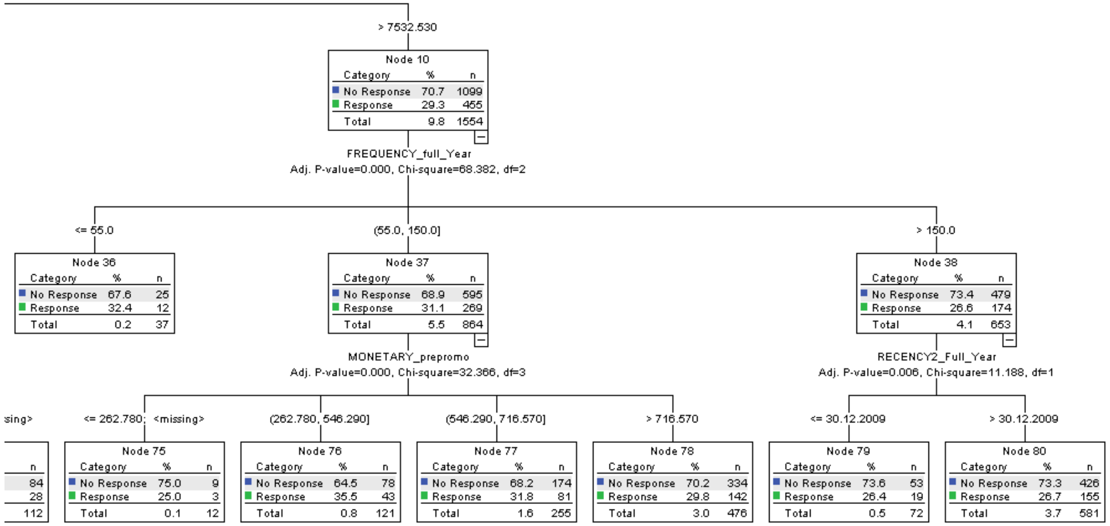
Appendix 4: Editor Output for RFM CHAID Fold 1 Output




PROMO_RESPONDER
No Response
Response
Node 0
Category % n
No Response 87.2 13819
Response 12.8 2033
Total 100.0 15852
MONETARY_full_year
Adj. P-value=0.000, Chi-square=14665.095, df=9
(1072.370, 1669.930]
(1669.930, 2398.760]
(2398.760, 3291.440]
Node 4
No Response 95.2 1493
Response 4.8 75
Total 9.9 1568
FREQUENCY_full_Year
Adj. P-value=0.000, Chi-square=42.306, df=2
Node 5
No Response 91.5 1496
Response 8.5 139
Total 10.3 1635
FREQUENCY_full_Year
Adj. P-value=0.000, Chi-square=73.479, df=2
Node 6
No Response 85.4 1388
Response 14.6 237
Total 10.3 1625
FREQUENCY_prepromo
Adj. P-value=0.000, Chi-square=61.968, df=2
<= 41.0
(41.0, 110.0]
> 110.0
<= 69.0
(69.0, 110.0]
> 110.0
<= 5.0
(5.0, 12.0]; <missing>
> 12.0
Node 16
No Response 95.1 826
Response 4.9 43
Total 5.5 869
Node 17
No Response 95.6 627
Response 4.4 29
Total 4.1 656
Node 18
No Response 93.0 40
Response 7.0 3
Total 0.3 43
Node 19
No Response 90.4 1045
Response 9.6 111
Total 7.3 1156
Node 20
No Response 93.5 329
Response 6.5 23
Total 2.2 352
Node 21
No Response 96.1 122
Response 3.9 5
Total 0.8 127
Node 22
No Response 83.3 665
Response 16.7 133
Total 5.0 798
Node 23
No Response 86.3 630
Response 13.7 100
Total 4.6 730
Node 24
No Response 95.9 93
Response 4.1 4
Total 0.6 97
MONETARY_prepromo
j. P-value=0.013, Chi-square=18.135, df=2
MONETARY_prepromo
Adj. P-value=0.000, Chi-square=35.948, df=2
FREQUENCY_full_Year
Adj. P-value=0.002, Chi-square=13.647, df=1
RECENCY2_Full_Year
Adj. P-value=0.001, Chi-square=14.612, df=1
(136.390, 339.260]
> 339.260
<= 262.780
(262.780, 430.520]; <missing>
> 430.520
<= 110.0
> 110.0
<= 30.12.2009
> 30.12.2009
Node 48
No Response 94.8 271
Response 5.2 15
Total 1.8 286
Node 49
No Response 95.7 45
Response 4.3 2
Total 0.3 47
Node 50
No Response 89.6 733
Response 10.4 85
Total 5.2 818
Node 51
No Response 92.4 266
Response 7.6 22
Total 1.8 288
Node 52
No Response 92.0 46
Response 8.0 4
Total 0.3 50
Node 53
No Response 87.1 521
Response 12.9 77
Total 3.8 598
Node 54
No Response 82.6 109
Response 17.4 23
Total 0.8 132
Node 55
No Response 100.0 12
Response 0.0 0
Total 0.1 12
Node 56
No Response 95.3 81
Response 4.7 4
Total 0.5 85


(3291.440, 4349.510]
Node 7
Category % n
No Response 80.6 1277
Response 19.4 308
Total 10.0 1585
FREQUENCY_prepromo
Adj. P-value=0.000, Chi-square=114.036, df=4
<= 1.0
(1.0, 5.0]
(5.0, 7.0]
(7.0, 12.0]; <missing>
> 12.0
<= 87.0
Node 25
Category % n
No Response 76.6 36
Response 23.4 11
Total 0.3 47
Node 26
Category % n
No Response 79.8 491
Response 20.2 124
Total 3.9 615
Node 27
Category % n
No Response 75.7 237
Response 24.3 76
Total 2.0 313
Node 28
Category % n
No Response 82.2 383
Response 17.8 83
Total 2.9 466
Node 29
Category % n
No Response 90.3 130
Response 9.7 14
Total 0.9 144
Node 30
RECENCY2_Full_Year
Adj. P-value=0.004, Chi-square=11.810, df=1
FREQUENCY_full_Year
Adj. P-value=0.011, Chi-square=10.464, df=1
FREQUENCY_full_Year
Adj. P-value=0.010, Chi-square=9.882, df=1
<= 24.12.2009
> 24.12.2009
<= 110.0
> 110.0
<= 150.0
> 150.0
<= 85.220; <missing>
(85.220, 339.260]
Node 57
Category % n
No Response 81.1 30
Response 18.9 7
Total 0.2 37
Node 58
Category % n
No Response 75.0 207
Response 25.0 69
Total 1.7 276
Node 59
Category % n
No Response 80.6 228
Response 19.4 55
Total 1.8 283
Node 60
Category % n
No Response 84.7 155
Response 15.3 28
Total 1.2 183
Node 61
Category % n
No Response 88.1 59
Response 11.9 8
Total 0.4 67
Node 62
Category % n
No Response 92.2 71
Response 7.8 6
Total 0.5 77
Node 63
Category % n
No Response 70.4 19
Response 29.6 8
Total 0.2 27
Node 64
Category % n
No Response 70.5 158
Response 29.5 66
Total 1.4 224
Node 65


(4349.510, 5662.470]
Node 8
Category % n
No Response 77.5 1262
Response 22.5 366
Total 10.3 1628
FREQUENCY_full_Year
Adj. P-value=0.000, Chi-square=118.199, df=2
(5662.470, 7532.530]
Node 9
Category % n
No Response 73.2 1114
Response 26.8 407
Total 9.6 1521
FREQUENCY_prepromo
Adj. P-value=0.000, Chi-square=72.527, df=2
<= 87.0
(87.0, 150.0]
> 150.0
<= 6.0
(6.0, 12.0]
> 12.0; <missing>
<= 55.0
Node 30
Category % n
No Response 73.7 542
Response 26.3 193
Total 4.6 735
Node 31
Category % n
No Response 78.4 494
Response 21.6 136
Total 4.0 630
Node 32
Category % n
No Response 85.9 226
Response 14.1 37
Total 1.7 263
Node 33
Category % n
No Response 71.3 352
Response 28.7 142
Total 3.1 494
Node 34
Category % n
No Response 73.0 500
Response 27.0 185
Total 4.3 685
Node 35
Category % n
No Response 76.6 262
Response 23.4 80
Total 2.2 342
Node 36
Category % n
No Response 67.6 25
Response 32.4 12
Total 0.2 37
MONETARY_prepromo
Adj. P-value=0.000, Chi-square=56.183, df=4
FREQUENCY_full_Year
Adj. P-value=0.000, Chi-square=23.845, df=2
MONETARY_prepromo
Adj. P-value=0.003, Chi-square=12.600, df=1
RECENCY2_prepromo
Adj. P-value=0.002, Chi-square=14.969, df=1
(85.220, 339.260]
(339.260, 546.290]
(546.290, 716.570]
> 716.570
<= 55.0
(55.0, 110.0]
> 110.0
<= 716.570
> 716.570
<= 02.01.2010
> 02.01.2010; <missing>
<= 262.780; <missing>
Node 64
Category % n
No Response 70.5 158
Response 29.5 66
Total 1.4 224
Node 65
Category % n
No Response 73.9 246
Response 26.1 87
Total 2.1 333
Node 66
Category % n
No Response 76.0 79
Response 24.0 25
Total 0.7 104
Node 67
Category % n
No Response 85.1 40
Response 14.9 7
Total 0.3 47
Node 68
Category % n
No Response 77.4 72
Response 22.6 21
Total 0.6 93
Node 69
Category % n
No Response 69.1 237
Response 30.9 106
Total 2.2 343
Node 70
Category % n
No Response 74.1 43
Response 25.9 15
Total 0.4 58
Node 71
Category % n
No Response 72.0 404
Response 28.0 157
Total 3.5 561
Node 72
Category % n
No Response 77.4 96
Response 22.6 28
Total 0.8 124
Node 73
Category % n
No Response 77.4 178
Response 22.6 52
Total 1.5 230
Node 74
Category % n
No Response 75.0 84
Response 25.0 28
Total 0.7 112
Node 75
Category % n
No Response 75.0 9
Response 25.0 3
Total 0.1 12



> 7532.530
Node 10
Category % n
No Response 70.7 1099
Response 29.3 455
Total 9.8 1554
FREQUENCY_full_Year
Adj. P-value=0.000, Chi-square=68.382, df=2
<= 55.0
(55.0, 150.0]
> 150.0
Node 36
Category % n
No Response 67.6 25
Response 32.4 12
Total 0.2 37
Node 37
Category % n
No Response 68.9 595
Response 31.1 269
Total 5.5 864
Node 38
Category % n
No Response 73.4 479
Response 26.6 174
Total 4.1 653
MONETARY_prepromo
Adj. P-value=0.000, Chi-square=32.366, df=3
RECENCY2_Full_Year
Adj. P-value=0.006, Chi-square=11.188, df=1
<= 262.780; <missing>
(262.780, 546.290]
(546.290, 716.570]
> 716.570
<= 30.12.2009
> 30.12.2009
n
84
28
112
Node 75
Category % n
No Response 75.0 9
Response 25.0 3
Total 0.1 12
Node 76
Category % n
No Response 64.5 78
Response 35.5 43
Total 0.8 121
Node 77
Category % n
No Response 68.2 174
Response 31.8 81
Total 1.6 255
Node 78
Category % n
No Response 70.2 334
Response 29.8 142
Total 3.0 476
Node 79
Category % n
No Response 73.6 53
Response 26.4 19
Total 0.5 72
Node 80
Category % n
No Response 73.3 426
Response 26.7 155
Total 3.7 581