

Article

# An Exploration of Big Data Practices in Retail Sector

Emel Aktas <sup>1,\*</sup>  and Yuwei Meng <sup>2</sup>

<sup>1</sup> Cranfield School of Management, Cranfield University, College Road, Cranfield MK43 0AL, UK

<sup>2</sup> Apple Computer Trading (Shanghai), No. 391 Yuanshen Road, Pudong, Shanghai 200135, China; yuwaymong@gmail.com

\* Correspondence: emel.aktas@cranfield.ac.uk; Tel.: +44-123-475-1122

Received: 29 September 2017; Accepted: 6 December 2017; Published: 12 December 2017

**Abstract:** Connected devices, sensors, and mobile apps make the retail sector a relevant testbed for big data tools and applications. We investigate how big data is, and can be used in retail operations. Based on our state-of-the-art literature review, we identify four themes for big data applications in retail logistics: availability, assortment, pricing, and layout planning. Our semi-structured interviews with retailers and academics suggest that historical sales data and loyalty schemes can be used to obtain customer insights for operational planning, but granular sales data can also benefit availability and assortment decisions. External data such as competitors' prices and weather conditions can be used for demand forecasting and pricing. However, the path to exploiting big data is not a bed of roses. Challenges include shortages of people with the right set of skills, the lack of support from suppliers, issues in IT integration, managerial concerns including information sharing and process integration, and physical capability of the supply chain to respond to real-time changes captured by big data. We propose a data maturity profile for retail businesses and highlight future research directions.

**Keywords:** big data; retail operations; maturity; availability; assortment; replenishment; pricing; layout; logistics

## 1. Introduction

Advances in information technologies and reductions in the cost of data collection have made individual and organizational data generation and recording an integral part of our lives. We leave information about our location via our phones and apps; reveal hints on our lifestyles to the supermarkets through every transaction; and share our sentiments in social media. The abundance of data encourages companies to use it to create a competitive advantage. Specific to the retail sector, McKinsey estimates the use of big data to increase the margins by up to 60% [1], highlighting the importance and the relevance of big data applications. In May 2013, Rangespan, a company using big data analysis to help retailers make decisions on the ranges to carry, is taken over by Google and stopped providing services to other retailers; now has become an in-house service for Google Shopping to expand its offerings [2]. Two months later, Boomerang Commerce, a company providing dynamic pricing services to retailers based on big data analysis, raised \$8.5 million [3].

By the time this study was carried out, most research on big data in retail has focused on how to obtain greater consumer insights to implement marketing activities more effectively. Although precise marketing can generate demand more effectively, without suitable operational support, these demands may not translate into sales [4]. Therefore, this study aims to unveil how big data currently is, and can be in the future used in retail operations. We formulate our research question as follows: *How can the use of big data benefit retailers in operations from a supply chain perspective?* To answer this research question, we first identify existing and potential applications of big data in retail operations through a state of the art literature review and expert interviews, and then we develop a roadmap from current to future

in the form of a maturity profile. From the literature review, we identify four areas for further research: availability, assortment, pricing, and layout planning; so, the research question can be crystallized as: “How big data is used, and can be used, in availability, assortment, pricing, and layout planning?”

The contribution of this paper is two-fold: First, we unveil how big data are applied in the retail business, without getting lost in the details of information technologies as well as hardware and software requirements. Second, we identify how big data can be used as a lever to increase operational efficiencies and help make better operational decisions, specifically in the retail sector. Retail operations can involve a wide spectrum of stakeholders from consumers to suppliers; however, the research is focused on the operations of retailers. To get a more complete picture, experts from academia and consultancies are interviewed. The applications described elaborate how data are collected, what insights can be provided, and how they link to the business. Benefits, risks, and enablers are also discussed.

This paper is organized as follows: Section 2 provides the literature review and the gaps in terms of big data applications in retail operations. Section 3 explains the research methods and materials, followed by findings in Section 4. Section 5 presents a discussion of the findings and concludes the work with academic and managerial implications.

## 2. Literature Review

We start with a background on the big data and the retail sector, identifying the current applications of big data and risks in implementation. Our review of big data focuses on its managerial applications rather than technical aspects. We provide the details of our literature review process in the Appendix A.

### 2.1. Big Data

In the 1960s, when computers first came into business, data were stored without structure and many efforts were required to create some value from them [5]. Relational databases were then introduced to add structure to data; enable companies to better organize them and obtain value [6]. When the volume of structured data grew faster than the computing power, data warehouses and data marts were introduced to divide the data into subsets and obtain better insights [6].

With the development of the Internet, however, the volume of unstructured data has started to grow rapidly, as the format of data is not limited to data tables [7,8]. Besides web content, data generated from devices and sensors have also become more accessible, such as data from cameras or GPS transceivers [2,9,10]. Although a vast amount of data has been collected and stored, the research is still ongoing in terms of how to analyze and gain insights from it [11,12]. It is reported that 23% of the data have value besides its original intent, but only 1% of them are analyzed [12,13]. The development of computing technologies such as distributed computing and cloud computing enabled these data to be analyzed in time [14,15]. As the costs for collecting, storing, and processing data reduced, including more data in the analyses became the trend [2,16].

‘Big data’ has different definitions that focus on size, range, and speed. While some incorporate a dynamic definition as *beyond the ability of typical software* [2,17], some use a static number as *100 terabytes* or *have a growth rate that is higher than 60% annually* [12]. Big data is also defined as using the population in the analysis rather than the sample, as this can change the mind-set of data analysis in three ways: accept errors as they can be smoothed by sheer quantity, accept correlation rather than causality, and discover value of data besides its original purpose [17]. Big data is also defined based on volume, variety, velocity, and veracity [11,12]. Volume represents to the amount of data, which can be affected by the scope, and is the primary attribute. Variety refers to the data included in the analysis which comes from various sources and are in many mediums. Velocity includes the frequency of both data generation and deliveries. Finally, veracity is to do with the uncertainty of the data. In this paper, the definition used is: *all population data that can be accessed timely*. This definition implicitly involves volume, variety, velocity, and veracity features of big data.

## 2.2. Retail Challenges and Potential Big Data Applications

Although retailers were distributing products passively in the past, with the information about the end customer demand, they are becoming more proactive in the supply chain [18]. The retail supply chain contains four major activities: assorting goods, breaking goods into smaller packages, holding inventory near the customer, and providing value-adding services such as gift wrapping and warranties [19]. Agrawal and Smith [20] mapped the supply chain of two major home furnishing retailers in the US and proposed a comprehensive set of planning processes: product design (private label merchandise) and assortment planning, sourcing and vendor selection, logistics planning, distribution planning and inventory management, clearance and markdown optimization, and cross-channel optimization.

Major activities of retailing are to decide what products to carry (assortment, product design, procurement), how to sell products to customers (marketing, including pricing), and how to complete these efficiently (supportive functions, such as logistic planning). Out of stock (OOS) situations and poor on shelf availability (OSA) lead to customer dissatisfaction and consequently loss of market share [21]. Assortment, pricing, and store layout are challenging processes in retail operations that are expected to benefit from big data analysis. Assortment and pricing involve a large amount of granular decisions and can be affected by various factors, such as customer preferences, store location, and demand elasticity. Without effective quantitative analysis, making these decisions can be resource-consuming and ineffective. Store layout is also worth exploring as it can affect customers' purchasing decisions.

### 2.2.1. Assortment, Pricing and Store Layout

Assortment is one of the most difficult tasks in the retail supply chain and has received high priority due to its impact on sales [22,23]. Big data is used to micro-segment customers and optimize assortment [2,24]. Common practices include analysis of the correlations between items purchased as well as time- and location- dependent purchase patterns [25,26]. The output helps retailers understand customer preferences and compositions, and improve forecast processes. A common practice in assortment planning is to group stock keeping units (SKUs) into categories [20]. With big data, it is possible to plan at SKU level, but the literature is yet to grow on this aspect.

Pricing is one of the most difficult issues facing retailers [26] due to the large amounts of SKUs whose price can vary in different locations and over time according to local demand and competition. In the case of markdowns and promotions, pricing is even more difficult because of the increased level of uncertainty and the lack of historical data with the same promotion conditions. As the promotions are a function of what else is also on the market, it is almost impossible to replicate a promotion, although some similarities can be captured for forecasting purposes. With extra computing and analysis capacity, big data analysis enables promotion decisions to be taken more effectively and efficiently [1]. Big data can also help retailers evaluate sources of sales lifts and plan future promotions more effectively.

A practice that can be adapted easily is the dynamic pricing commonly used in the airline industry. A dynamic pricing model can incorporate factors such as demand and production (available seats) to adjust the price based on the long-term strategic price [27] as prices from some airlines are responsive to rival prices [28]. An example from the UK is that major grocery retailers such as Sainsbury's, Tesco, ASDA, or Morrisons give price guarantees to the consumer on branded products, leveraging connectivity of information systems and big data flows.

Store layout is worth exploring in terms of how it can benefit from big data analysis as it has an impact on purchasing decisions. In the past, customer in-store behavior could be analyzed by observing sampled customers and this information could help retailers optimize store layout and shelf design [29]. However, this exercise requires a significant level of effort and reports have suggested that using videos, mobile services, WI-FI, and RFID tags attached to shopping carts to track customer movement in store can provide high volumes of information at low cost [2,24]. Although in-store

traffic alone cannot provide information on customer behavior, customers' in-store location data still provide useful correlations with sales.

### 2.2.2. E-Tailing and Multi-Channel Fulfillment

Besides potential traditional channels, the growing popularity of online shopping has also changed retail operations and revealed other opportunities to apply big data analysis [30]. As more people are gaining Internet access with the popularity of tablets and mobile devices, more people will shop online, especially the elderly [31]. Compared to traditional channels, online shopping has imposed new logistics requirements: increased volume of goods to be handled, a wider geographic range of customers to be served with pressing demands such as rapid and reliable delivery at convenient times [32].

Online shopping also influences food and non-food retail operations in different ways due to different product and demand characteristics that require different fulfillment models. For example, a typical online grocery order contains 60–80 items and customers expect the order to be completed in one delivery, necessitating an efficient order picking and consolidation process [18]. Regional distribution centers are used for picking; however, they are usually not set up for item-level picking, and their local deliveries may be inefficient due to the wide geographic area they serve [33]. Two types of fulfillment models are identified for online orders: picking in stores and picking in fulfillment centers [18]. While store-based picking minimizes the investment and local delivery transport costs, its substitution rate is significantly higher due to the on-shelf availability. Purpose-built fulfillment centers can provide real time inventory information to customers to avoid this problem; however, they have higher transport cost especially when serving a wide geographic area [21]. One solution that takes advantage of both models is to use a fulfillment center in densely populated areas and store-based picking in others. Tesco, a major British retailer, has adopted this strategy and built several fulfillment centers around London. Big data analysis can provide better demand forecasts and improve on-shelf availability.

Unlike online grocery orders, non-food orders only have one or two items per order [18]. In addition, since non-food retailing normally has many SKUs with a long tail of slow-moving items, it is common that these retailers have several distribution centers handling different product categories to minimize inventory investment. Compared to food retail, parcel carriers are preferred in non-food retail over own fleet for home delivery mainly because of their efficiency [33]. However, using parcel carriers means that the retailer has less control over the delivery and therefore requires earlier cut off times as packed items would go to the carrier first. This lead-time may dissuade customers from buying items from online merchants [34], which could be addressed by locating warehouses close to the carrier's hub.

Recent big data research on retail logistics also use publicly available data such as customer inquiries on Twitter. Customers learn from experiences of others and understand the logistics-related service processes of the retailer better [35]. Interactions over Twitter also help the retailer connect with its customers and address queries much faster than other channels.

Besides, big data analysis can be used to shorten the time from receiving orders to delivery. Amazon is using historical order data to anticipate demand in certain geographic areas, so that items can be shipped before orders are placed, thus shortening the time customers have to wait [34].

One of the biggest challenges e-tailing brings, is on-shelf availability, especially for store-based picking fulfillment model [18]. Stock outs not only lead to higher substitution rates but also affect the sales in stores [36]. Companies try to produce better forecasts by exploring correlations not only between items bought but also other factors such as the weather [1] or the time of the year or specific events such as promotions or Christmas within the forecast horizon, and then make better inventory management decisions [24]. On the other hand, correlations discovered in this way may be incomprehensible and misleading [37], so care must be taken in arriving at conclusions.

Customers are not only using the Internet for placing orders but also for sharing their online shopping experiences within their own social network, and increasingly relying on peer sentiment and recommendations on these social networks to make purchasing decisions [38]. These data, if analyzed, can provide valuable insight for customer relationship management. Since retaining existing customers is easier than winning new ones, creating a better shopping experience is in line with this viewpoint [18]. By using information from social networks, retailers can better understand shoppers as individuals, and get insights in terms of their satisfaction towards services, preferences, and subsequently provide personalized offers and promotions to enhance the shopping experience [23]. Besides, customer location data can also be used in marketing. iBeacon, an extended location service provided by Apple, enables users' location to be tracked through Apple devices more precisely [39]. Taking advantage of these data, some apps are designed to send tailored promotions and offers to users when they pass certain stores or certain aisles in a store [40].

### 2.3. Risks in Implementing Big Data Solutions

One of the most significant challenges in implementing big data applications is the privacy concerns [41]. Data can be collected and traced to individuals in various forms without being known by the person: internet tracking, location data, video data etc. [42]. However, further concerns come from the extent to which data are integrated and analyzed to gain insight into the individual. When Google revised its privacy policy initially in 2012, which allowed it to integrate user information from different services to build more complete user profiles, data protection authorities investigated the changes, concluded that this policy did not comply with the EU legal framework and gave recommendations [43]. However, Google failed to meet the recommendations and was fined by the French Data Protection Authority [44]. In addition, supposedly anonymous data collected from social media and made public for research purposes can be de-anonymized [45].

Data credibility is another important issue when implementing big data solutions. Poor data quality and the noise in the data are common problems for all types of information within many organizations. The cost of validating data in terms of resources is enormous: studies show that it can take up to 50% of an employee's time to validate and correct the data [46]. In the UK retail industry, the product information data are inconsistent across the supply chain in over 80% of the instances, significantly impacting operations, coordination, and profits [47]. Insights from data can only be as good as the data they are based on [48]; therefore, statistical methods should be used to measure, monitor, and control data quality and credibility [49].

Another risk of applying big data analysis is the shortage of analytical talent to conduct the analysis. By 2018, the shortage of deep analytical talent could reach 190,000 in the US [1]. Although providing training and education is essential in addressing this problem, the analytical culture within organizations has a significant impact on talent retention in the long term [50].

One of the most fundamental obstacles in using big data for retail operations is the managerial culture [51]. Adopting technologies alone cannot guarantee improvements in productivity; changes are necessary in the decision-making culture and management practices for successful implementation of big data analysis [52].

The academic research carried out on this topic is yet to grow and most articles describing real-life practices are written from the marketing perspective: i.e., micro-segmenting customers for more effective marketing activities. Nonetheless, assortment planning, pricing, and store layout design have been identified in the literature review as areas that can benefit from greater quantitative analysis. Availability has emerged as a key issue of retail operations in the literature review, having been affected by online shopping and can be improved with use of big data.

## 3. Materials and Methods

The primary data in this paper are collected via interviews with experts in retail industry and big data applications (Figure 1). Considering the nature of this study, which aims to explore possibilities

with focus on pre-identified areas for big data applications in retail operations from the literature review, semi-structured interview is chosen as the method to obtain primary data as it allows the researchers to use open-ended questions and probe respondents to yield in-depth responses. Secondary data comprise academic literature, non-academic (grey) literature, and company websites as this field is yet to be established and still growing.

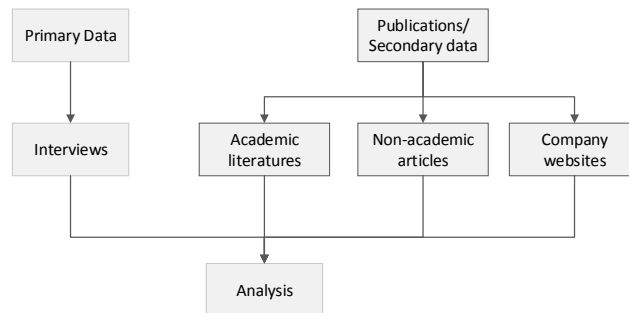


Figure 1. Data used in the study.

### 3.1. Interview Design

A semi-structured interview is designed to 1) identify existing applications of big data in retail operations; 2) discover potential applications of big data in retail operations; and 3) roadmap from current to future in the form of maturity. This includes a framework of how data are collected, what insights they can provide, and how they apply to the business. In addition, benefits, risks, and enablers are also explored. The basic structure of interviews is designed as shown in Table 1.

Table 1. Objectives and interview structure.

Objective	Data Types	Decisions	Analysis (Insight)	Challenges; Concerns	Next Steps
1	X	X	X	X	
2				X	X
3					X
	Framework	Framework; benefits	Framework,	Risks, enablers	

Assortment, pricing, layout planning, and availability are identified, as shown in Table 2 and discussed in literature review.

Table 2. A mapping of the literature with areas identified for big data applications.

Literature	Assortment	Pricing	Layout	Availability
[20]	X			
[22]	X			
[1]	X	X	X	X
[23]	X	X*	X	
[18]	X	X*		
[26]		X		
[29]			X	
[36]				X

\* in the form of personalizing offers.

A set of semi-structured questions and an invitation email template are developed. In the meantime, the interview is designed to be conducted via telephone as well, since it can be more feasible for participants sometimes. Where possible, face-to-face interviews are preferred. Prior to formal interview, a pilot is carried out to test if the questions are clear and if the interview can be finished within the designed length. In some cases, the respondent has expertise in certain areas but not the others; therefore, not all questions are answered. Although in these cases, the questions would

go deeper in the expertise area of the respondent. This adjustment enables interview quality to be improved as questions can be more focused, therefore getting a more detailed response in limited time.

### 3.2. Sample

Two main approaches to sampling are probability and non-probability sampling [53]. In line with the interview design, instead of selecting a sample from a large pool randomly, the sample is selected intentionally from people who hold a position related to retail operations or academics with relative retail operations experience [54] within the authors' networks. There are no exact rules for the sample size; the validity and the meaningfulness have more to do with the information richness and the quality of the analysis rather than sample size [55]. Due to the nature of this study, the sampling strategy is to have as many and as wide a range of respondents as possible to explore this emerging area. In total, 12 interviews are conducted and 13 respondents are involved in the primary data collection. Eight respondents work in retail industry, three work in consultancy or for a data solution provider, and four hold positions in academic institutes. Respondent 9 holds positions in both a service provider and an academic institution whereas Respondent 4 is employed by a retailer and is also an academic, as shown in Table 3.

**Table 3.** Respondents by type of organization.

Retailer	Consultancy/Services Provider	Academic
R1, R2, R3, R4, R5, R6, R10, R12	R9, R11, R13	R4, R7, R8, R9

### 3.3. Data Analysis

Two ways to organize and report qualitative data are storytelling or case study [55], to use in an analytical framework, which can be built around issues, processes, questions, or sensitizing concept. To keep in line with the interview design, the authors choose to use the interview structure as the thread for describing findings, and arranged them around themes as identified from the literature review. Challenges and plans are discussed separately.

To avoid ambiguity and inaccuracy in interpretation, transcripts are presented verbatim and key information is highlighted in bold with the relative code, where evidence is coded in order of appearance (see supplementary material). For example, I1E1 refers to interview one, evidence one. The codes are shown in brackets.

The major limitation of this research design is the small sample, risking generalizability. In addition, the time scale of the research overlapped with holiday season of retailers, making it difficult to reach them. Nevertheless, the authors believe the information collected and the insights achieved are useful theoretically and practically.

## 4. Findings

We first present descriptive information about interviews and then organize the findings into the following areas: availability, assortment/ranging, pricing, layout and space planning, and challenges. The retail businesses involved are clothing, grocery (foods), and others categorized as general merchandise. Among those who work in retail, most work in the general merchandise sector, and Respondents 3, 4, and 10 also have expertise in the grocery sector. Respondents 1 and 2 work in the clothing sector. Lengths of the interviews vary depending on the content; altogether they lasted around 6.5 h (391 min) with the average of 33 min per interview (Table 4).

**Table 4.** Interview information.

Interview No	Respondents	Type	Length (min)	Type of Organization	Retail Sector *
1	1,2	Telephone	25	Retailers	Clothing
2	3	Telephone	30	Retailers	Grocery/General
3	4	Face-to-face	66	Retailers/Academic	Grocery/General
4	5	Telephone	26	Retailers	General
5	6	Telephone	30	Retailers	General
6	7	Telephone	26	Academic	n/a
7	8	Face-to-face	25	Academic	n/a
8	9	Face-to-face	41	Provider/Academic	n/a
9	10	Telephone	26	Retailers	Grocery/General
10	11	Telephone	40	Provider	n/a
11	12	Telephone	22	Retailers	General
12	13	Telephone	34	Provider	n/a

\* Applies to those who work in retail sector; 'General' refers to general merchandise.

#### 4.1. Availability

As availability can involve multiple processes and activities before and after sales, findings are further grouped into demand sensing (primarily in terms of volume, to avoid overlaps with ranging), stock deployment, replenishment and inventory management, and substitution management.

In terms of *demand sensing*, using historical sales data to forecast the future demand is a common practice among retailers, as is mentioned by many respondents (R3, R4, R5, R9, R10, R11, R13). However, the granularity and richness of insights obtained from these data can be different from one retailer to another. To start with, historical sales data can be used to identify trends and therefore make forecasts about future demand (I2E3, I4E21, I9E1). Based on this, some models would also have elements such as seasonality (I4E23, I9E12, I12E33), weather (I9E13), consumer behavior (I2E4, I2E14, I3E86, I10E6), lifestyle (I3E4, I3E5, I3E90), demographic (I12E4), delivery time window (I3E47, I9E11), inventory (I3E80, I4E24, I5E25), and store location (I8E33) to enrich the analysis. Seasonality and weather are quite straightforward to incorporate in models built to exploit big data; holidays such as Christmas or Easter coupled with weather temperature would influence sales on certain products (I2E18, I9E13, I9E14). Some would use this forecast to determine the purchase order quantity quarterly (I4E24, I4E39), but in extreme cases, this sort of forecast can be used to inform replenishments within a day (I3E49). Besides the purpose of analysis, the *capability* and the *degree of sophistication in analysis* are also the reasons for differences across retailers (I9E14).

Consumer behavior is another area for retailers to benefit from big data applications. With loyalty schemes, they can monitor customers' purchase history, and gain some insights into customer preferences and demand patterns for each individual shopper, hence potentially better anticipating what a customer is going to buy in a shopping trip (I3E86, I3E90, I3E91, I3E92, I12E47, I12E51). However, this is not yet mature (I12E51). With more advanced location based technologies such as iBeacon, retailers would be able to update their forecasts if they detect a certain loyal customer stepping into a store (I3E88). If a retailer operates in more than one channel or even more than one business, they can collect more information about their customers' lifestyles and build a clearer picture of the individual (I3E5, I12E4). This, coupled with demographic information can benefit retailers in demand sensing in a way similar to the loyalty scheme, but the demographic information is used at a more aggregated level.

Operational data can also be used to enrich the analysis of historical sales. Delivery time windows and inventory levels can help determine purchasing and replenishment quantities at a more granular level. Store locations enable sales to be analyzed per store or per region to inform stock deployment decisions. In addition, with more advanced technologies, demand information can be more visible further up the pipeline, and especially for general merchandise, items may be labeled to specific customers rather than to stores (I8E4, I10E28, I10E29).

Suppliers would also have their own analysis and sometimes share that with retailers, especially in the case of new product launches (I9E10, I9E33, I10E39, I10E40, I10E42, I10E44, I10E46, I11E11, I11E19). This is because historically, suppliers, especially big-brand manufacturers have the knowledge



and the capability to analyze sales of their products. In addition, forecasting sales for new products would have more elements of an art than forecasting of existing products as there is no statistical data to inform the new product forecasts (I3E70). Sometimes it is possible for retailers to consider product characteristics from historical sales and compare it with the new product features (I4E8, I4E9, I9E10).

Although big data can help retailers to have better anticipation of future demand, the forecasts it supports may contain errors and should be treated with caution (I8E29, I8E31). Retailers can now personalize promotions, but to predict the demand it generates is still difficult in general, but possible in theory (I6E21, I7E6, I12E51). In addition, this information may not be capitalized on if actions it suggested are beyond the physical capability of the supply chain (I8E26, I12E29). Therefore, how the supply chain operates in terms of stock deployment and replenishment are indispensable components in delivering availability.

By analyzing the historical demand of stores or regions, retailers can *deploy their stock* more effectively by segmenting stores and understanding differences in demand patterns among them (I8E33, I8E35, I8E38, I8E41). This decision can be regarded as assortment in terms of volume and may include not only stores but also warehouses. Detailed transaction data enable retailers to analyze how each individual SKU performs in different stores, even at different times within a day (I8E38, I10E35). Therefore, retailers can make stock deployment decisions based on this. However, some retailers would also consider performance in clusters, as sales between stores with geographic proximity can be interrelated (I5E34, I12E27). In terms of sophistication in analysis, some retailers do this using an algorithm or a mathematical model (I8E39) and some rely on guesswork (I9E29, I9E31). More data with better quality can help improve sales and stock investment (I9E14, I9E17). Retailers are segmenting stores and this would also be reflected in the stock deployment decisions (I8E38, I12E22). For example, some stores may function more like a showroom or collection point (I5E6); therefore, do not necessarily have to maintain a high stock level. This is particularly the case for general merchandise.

In terms of the *replenishment and inventory management*, measuring availability is a crucial activity, as it triggers and drives replenishment decisions. Although measuring availability is not a new concept, with the advancement of technology and customers who are more demanding, what it means can be different. In a store environment, normally the customer would only interact with what is available on the shelf; therefore, on-shelf availability is what the retailers are interested in. However, what stock data retailers have on hand may only indicate the stock level in the stores, but not necessarily on the shelf, therefore, a discrepancy may exist between the physical stock and the stock records in the business systems (I3E29, I3E30, I8E1, I8E3, I9E16). One direct way to solve this problem is to conduct a stock check to make sure the data are up-to-date (I5E12, I9E19). These stock checks would also record gaps on the shelf and whether the customer chooses a substitute, especially in the environment of general merchandise, as staff can interact with customers and suggest substitutions (I2E10).

Indirectly, but maybe more cost effectively, some retailers also measure availability using sales data. They match historical demand patterns with actual demand at the SKU level, and if the actual demand is significantly lower than what is indicated in the pre-identified demand pattern, there might be gaps on the shelf (I3E45, I3E46, I3E52, I3E53, I3E56, I9E18). Similar to the case of demand sensing, differences in velocity are also identified. Some retailers would monitor this based on minutes (I3E53, I3E55), whereas some would do that in days (I9E18). To make sure the variances can reflect on-shelf availability, the retailer who monitors the variance in minutes would also monitor their rate of check out, making sure the differences come from stock outs, not from long queues at the till (I3E54).

Availability has evolved to reflect customer satisfaction level (I3E6, I10E55, I10E56). In the past, the indicator may refer to stock level at one point in time on a specific day; nowadays, it can refer to the stock level at any one time (I10E55). Replenishment process efficiency can be monitored at the store or shelf level. Replenishing the stores in the past would be done by generating a mass replenishment order at a certain time in a day, and that would capture all demand within a day (I3E48). A more advanced practice would be setting a reorder level, and when the stock is consumed to that level,

placing an order (I9E3). To sum, these practices are similar in nature, which is to replenish what is sold and it is difficult to detect the lost sales.

Some retailers allow the store managers to place orders based on their anticipations (I9E5), in this way they can be more proactive in fulfilling future demand. However, since this would rely on the store manager's own judgment, it might be used less frequently in the future (I9E7, I9E8). With advanced technologies, it is possible to integrate sales forecast, minimum presentation level of stock depending on space planning, delivery time windows, and sales in real time to manage the process more systematically. To be more specific, from sales forecast and delivery time windows, the retailer can work out what the demand would be between the next two deliveries; therefore, would be able to decide to which level the stock should be replenished to meet the demand (I3E48, I9E11). Taking the minimum presentation stock level into consideration could help adjust the quantity and make the order more economic (I3E80). Real time sales information could feed into the forecast and update it, but more importantly, it is used to calculate stock levels within the store (I3E92).

The abovementioned solution considers detailed demand forecasts; however, as forecasts contain error, retailers have developed another application to only respond to actual demand, the continuous replenishment method. In this case, demand would be accumulated as items go through the tills. Once what is sold (retailer's order to its supplier) reaches the point where it can fill a vehicle, the order is released. In this way, replenishment transport is optimized (I3E50).

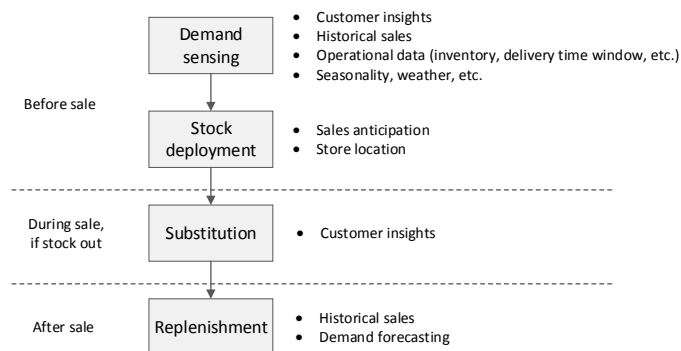
Although the focus of methods is different, both require the supply chain to be able to handle more items in smaller quantities and on a more frequent basis. It places new pressures on the supply chain in terms of responsiveness and reliability, and raises the standard for information technology (I10E23, I10E27, I10E58). Traditionally, supply chain activities end at the point when stocks are replenished to the stores (I10E24). However, to realize the benefit from the above-mentioned applications, replenishing activities within stores should also be part of the solution, as it can influence the overall responsiveness as well as discrepancy between stock data and actual stock level on the shelf. This notion starts when some retailers start fulfilling online orders by picking inside the stores and finding their actual on-shelf availability is below what they expect (I3E28, I3E31). The discrepancy, as a shocking fact for retailers, drives them to make the process of moving stock from the back of the receiving area to the shelf more smoothly and efficiently (I3E32), considering also package size and shelf design (I3E79). Ideally, all goods within one case could be filled into the shelf in one 'touch' (I3E80), which necessitates shelf and space planning.

Although retailers have been trying to improve availability, stock-outs and substitutions are inevitable. Managing the stock-out situations effectively can affect customer experience, especially for grocery home delivery, as customers do not have the opportunity to choose substitutions themselves, whereas in a store environment, especially for general merchandise, staff can suggest substitutions (I2E10) or consumers themselves can find similar products to the items that have stocked-out. In the past, retailers tended to substitute items with lower value, however they found customers are not quite satisfied with this (I3E35). Today the retailers provide substitute products that are more valuable than the customer's original order, increasing the value of the basket size (I3E38). Overtime, retailers will accumulate a pool of information on acceptance rates for substitutions and hence will be able to make better decisions on substitute products. In addition, with loyalty scheme data, the acceptance rate can be based on individual customers and therefore can suggest substitutions that are more likely to be accepted for a specific item in an order placed by a specific consumer (I3E39).

The availability issues encompass several activities before and after the sale. Big data have potential to be used in all these aspects, as shown in Figure 2.

Before sales, demand sensing and forecasting is the starting point for the rest of the activities. Sophisticated methods and algorithms that involve several different factors, such as customer insights, historical data, weather forecasting, etc., are used. Then, stock is deployed to individual stores based on anticipated demand. In case of stock-outs, effective substitution management prevents poor customer experience. Customer insights can be used to propose alternatives that are more likely to be accepted.

Once stock level goes down, the replenishment process should be sufficiently effective to support further sales. Velocity in transmitting sales data can help detect gaps on the shelf and forecasted demand to be more proactive.



**Figure 2.** Processes involved in availability that can benefit from big data applications.

#### 4.2. Assortment Planning

Historical sales data is the major source for making assortment decisions in most retailers. Typically, retailers examine what sells well in a specific fashion or category (I3E67, I4E1, I4E6, I9E10, I9E31), and then use this information to reach an assortment decision. In addition to the sales, retailers also consider the margin performance, although it acts more like an incentive and target (I4E12, I6E14). Some retailers consider forecasted sales of specific products in certain stores together with the margin to decide if it is economic to range the item (I3E64, I8E39), whereas others carry out assortment trials and review the results (I3E63, I3E65). Companies that do not analyze assortment decisions in detail mainly lack the capabilities to approach assortment planning scientifically (I9E34).

Whether to conduct trials or not would also depend on the nature of the business. Some businesses are rigorous in controlling the range of their products to get true cost advantages (I3E60, I3E66). In these cases, the number of SKUs they carry is stable, so items need to demonstrate the ability to have better performance in terms of the range offered in stores (I3E64). In contrast, other businesses choose to grow in variety and give items the opportunity to be tested in the store environment as part of the assortment decision (I3E65). Due to the space constraints of physical stores, it might not be possible to range all products in the store (I9E35), which is tightly linked to space planning. The retailers generally decide on the SKUs to present in the store, but some retailers also consider store performance in clusters as part of planning the stock deployment (I5E34).

*Consumer insight* data have the potential to enrich sales data and help localize ranging decisions for stores; however, this idea has not matured sufficiently yet to be of any real use in operations (I10E5, I12E32). Consumer insights are more often used to identify trends in the long term, and the marketing team is using this to segment shoppers for marketing purposes (I2E3, I6E5). Historical sales data merged with loyalty schemes can help the retailer understand not only what has been bought but also who bought it (I6E15, I12E10, I12E11). If this information can be linked to customer lifestyles, assortment decisions can be made more effectively (I10E11, I12E12). Tuning product range to a different time within a day for each individual store is possible, so the same store could carry different products in the same day to meet customer demand more effectively (I10E8, I10E34). For example, the data may be able to tell that most customers of a specific store are people who work nearby; therefore, during lunchtime, sandwiches and hot meals sell better, whereas in the afternoon, evening ready meals are more popular. Some retailers are learning to adapt their assortment accordingly, but this idea is yet to mature (I10E9, I10E10, I10E35).

A rather mature application is to examine repeat purchases as a reliable source to judge the success of a product (I5E20). If a consumer declines to buy a product twice it is likely that the demand will decline when there is a lack of first time purchasers. Although historically retailers would

segment stores according to demographic factors (I12E25, I12E26), consumer insights data give them new evidence on this aspect (I10E12, I12E27). This type of segmentation enables retailers to make judgements on what is more likely to be sold in which store. The analysis technology enables product ranges to be changed more frequently to meet the demand, but it is unknown if the supply chain has the physical capability to carry that out (I6E36, I8E32, I10E13, I12E29), and consumers may not get used to frequent changes on ranging; therefore, it may even have an opposite effect (I10E13).

*Collaboration with suppliers* can benefit the assortment planning, although suppliers may have their own anticipation and that may not necessarily be the same as those shared by retailers (I3E72, I9E9, I9E33, I10E39, I10E43, I10E45, I11E10). This is because suppliers tend to be more focused when analyzing how a specific item performs and what element drives the sales up for that item, whereas retailers are more interested in performances across retail estates (I10E39, I10E40). Therefore, especially in the case of launching a new product, the supplier would have a better idea of how new products share similarities with existing products and have a better anticipation of demand consequently (I9E9). It is possible to exploit big data by integrating supplier and retailer data when launching new products collaboratively (I9E9, I11E10).

On the other hand, if the retailer could share certain sales information of products with suppliers, suppliers could make better forecasts as the input data quality is improved (I10E41). Using repeat purchases to determine if products are successful is an example of this (R7E16). However, not all suppliers have the capability to analyze such data and therefore, not all suppliers are interested in such information (I3E19, I3E20, I6E26). What retailers would do consequently is to tier suppliers, and have a different degree of information exchange and monetize that exchange at different degrees (I3E21, I4E40). While some retailers would have a portal for suppliers to access such data (I3E18), some retailers would exchange information by communication documents (I9E9).

Besides anticipating sales collaboratively, another reason for retailers to work with suppliers in coordinating retail logistics is that suppliers, especially big-brand manufactures, would put certain marketing budgets behind their product, and some of these budgets would go to retailers as a listing fee (I3E69). There would still be room for better collaboration between the two parties to improve the decision-making (I10E41, I11E11, I11E12) where the trust between the two sides may pose an issue (I3E71).

#### 4.3. Pricing

Pricing is an essential element in the retail business, and differences in pricing strategies are identified across companies from the interviews. Some companies would have a pre-defined price band for their product range (I4E10, I9-This retailer has only pre-set price for their products) where dynamic pricing would not be applicable in the selling season, but relevant for regular price reviews across the year. Alternatively, these companies focus on working out the purchase cost and push that back to customers (I4E18). Some retailers have an overall pricing strategy, i.e., 'so much higher than the competitors' for *everyday prices*, but are flexible at the SKU level (I2E41, I5E28, I10E17). In these cases, retailers would include competition and margins to work out prices (I2E43, I5E30). Some retailers would also monitor sell-through rate, which is the number of products that can be sold before they are discounted, as an indicator to judge the success of pricing decisions. If the rate is not as expected, they would revise the price (I2E42). In some cases, the retailer would build these elements into a pricing algorithm so that they are able to respond to competition more swiftly and update the price more regularly, even daily (I5E28). However, stock information must be built into such an algorithm that takes availability into consideration (I5E29).

Although the analysis can be run daily, retailers may not necessarily change the prices that frequently, especially those operate in more than one channel. Changing the prices on the website is relatively easy compared to the off-line channel. The retailers mention that they have put effort into price consistency, but it is difficult to keep it stable and their customers are now aware of the differences between channels (I5E31). If the dynamic pricing strategy is extended to the off-line channel, technologies such as digital price tags for display prices will be indispensable (I10E58).

*Promotions* are a special case in pricing and a traditional marketing activity; customer insights data is used more heavily in this area (I10E52). With loyalty scheme data, retailers can build a more complete picture of how people react to price changes and their preferences; therefore, they can personalize offers for individual customers according to purchase history (I3E7, I12E16, I12E17). Over time, as retailers accumulate more data about how customers shop, the models they use in discounting products become more refined. For example, retailers can trace customer browsing history in the online channel and identify when and after how much time a customer drops out (I3E23). In the past, retailers may have sent a customer a discount offer about the item, which the customer would have viewed, straight away. Today the retailer realizes the shopping journey: customers compare the product with similar items or in different retailers, ask people's view on it, visit the store to try the product off-line, etc.; consequently, retailers would not send out the offer so soon (I3E11).

Big data and the insights from it can be used to improve *space utilization* and to maximize the range within a limited space as well as to improve the attractiveness of the store and maximize the sales on a limited range. Big data has recently been incorporated to the processes that inform stock location decisions (I10E52). Retailers would segment stores according to demand patterns and demography, besides stock deployment and ranging, this segmentation would also influence layout and space planning. Retailers would have a standard design based on store types, although they can still use different sizes (I3E74). When considering utilization, the challenge is to fit more ranges into a limited space. Some retailers can only get 40% of their range into one store (I9E35).

Retailers would use computer modelling to optimize space based on the forecasted sales (I3E78, I6E9). With advanced technologies, it is possible to develop 3D models for the store layout design (I3E84). To do this, product information such as volume, weight, and dimensions as well as shelf specifications such as height, width, and depth would need to be included in the model (I3E78). Combined with the forecasted sales rate, the model can work out how many facings are needed for a specific product (I3E81, I12E15).

As the planning is on product or packaging size, it is not surprising that this has always been a collaborative activity with suppliers (I10E48). To make the replenishment process more efficient, the packaging size should allow fitting a complete incoming case into the shelf so that the retailer does not need to replenish piece by piece (I3E80). This would also influence the minimum presentation stock level. Because the retailer would prefer to replenish a complete case at a time, it is acceptable to have fewer items on the shelf which can be easily replenished, keeping the display looking good (I3E80). This practice is already adopted by some retailers and there are also other retailers who have a plan to implement this soon (I9E37). Although the packaging sizes are decreasing to use the shelf space more efficiently, this is not a sustainable way for the future as this may introduce more variances in packaging size and stresses the supply chain (I10E53). Therefore, the feasible way may be dynamic ranging and dynamic shelf space to use the shelves more effectively (I10E51, I10E54).

Whereas the utilization aims to increase sales by increasing the range, attractiveness deals with how to increase sales of limited range. This can be achieved by making the shopping journey more pleasant for customers (I2E36, I12E28). One application is a heat-mapping planogram based on historical sales data (I2E34, I3E87). Fast moving items would be presented as 'hot spots' and slow moving items as 'cold spots'. If hot spots are moved around cold spots, slow moving items could be exposed to more customers (I2E35, I3E93). Consequently, sales may be increased. In addition, this 'hot' and 'cold' difference can go down to different levels on a shelf (I3E81).

Analyzing correlations among the items purchased in one basket can inform which products should be placed near each other (I3E94, I10E50). However, this is still a conceptual application although there are 'stories' (I3E94, I10E52), such as placing nappies next to beer for fathers doing the grocery shopping. Similarly, another way for the layout planning to prompt more purchases is to understand the 'mission' customers have when they are shopping, and use that information to plan layouts that are convenient to them (I12E28).

Effort that has been put into reflecting the trends on store design to appeal customers and therefore the layout may change significantly if the range of products to be displayed or trend is changed (I2E36, I3E63). Big data can help in terms of identifying those trends. Unlike utilization, attractiveness is more difficult to measure, at least in the design phase. Therefore, retailers could set up pilot stores to see if the new design is working before it is rolled out (I2E39, I3E83). There are also retailers who do not have fixed layouts and are very flexible in placing stocks (I5E35). In any case, some techniques in warehouse design, such as placing certain items near checkouts are applied to enable customers flow through the store more quickly (I5E36).

#### 4.4. Challenges

There is great potential to use big data and analysis methods to improve retail operations; however, some challenges are also identified in interviews. These challenges are grouped into managerial issues, people and capabilities, technical issues, and customer concerns around privacy.

##### 4.4.1. Managerial Issues

Managerial issues refer to the culture within an organization and the process integration that is required for successful big data applications and implementation of learnings. Main issues are being open to changes even if there is a possibility of failure, and information sharing. Implementing big data applications can mean the traditional processes need to be changed and people within the organization can be resistant to these changes (I6E24). The organization should be curious and open to innovative ways of running the business, although the new process cannot guarantee success (I8E42). In more advance retailing, there is a culture that can be called 'fast fail' (I3E100) and it has two parts to its meaning: The first part is to be a little more adventurous and open to new opportunities when success is not guaranteed (I3E102), which does not tend to sit well in traditional retailing (I3E101). The second part is that once the change has proven to be faulty, they should be able to get out of it quickly before further resources are invested (I3E100). This is validated as respondents mentioned trials and prototypes having been carried out and tested for improving big data applications (I2E40, I5E29, I5E38, I9E36).

Another managerial issue is the willingness to integrate processes and share information between functions. Currently more organizations are starting to exploit the value of data; however, historically these data are handled in isolation (I10E38). In some cases, people in different positions would make their own analysis and manipulate data for their own interests (I5E17); therefore, effort would be spent in reaching an agreement rather than solving problems. If data cannot be integrated for further analysis or solutions carried out without impact are fully assessed, it would have the opposite result once implemented (I8E24). A typical example may be promotions. If the customer-facing team wants to implement promotions within a time frame that is not aligned with the supply chain capability, it is possible that the availability cannot be assured and the customer would be let down (I8E24). Even within the same function, if information is not shared, there is a risk to obtain different results from different data sets (I12E7). In short, faster and more robust analyses are only possible when people work in alignment (I3E103).

##### 4.4.2. People and Capabilities

Before implementing big data analysis to enhance current retail operations the following issues should be addressed: people, timeliness, and support from suppliers. One of the barriers for implementing more advanced data analysis is the lack of human resources with the right skills set, which may be different for each organization. Sometimes there are not enough people to use the system effectively; especially when they try to consolidate functions and integrate processes, some experts are lost along the way (I4E30). Training the people to use the system is relatively easy (I1E22), and it is not too difficult to find people who can produce and analyze complex data (I5E15). However, there are no right people to interpret the analysis results and develop insights, especially when it

comes to integrating data across organizations (I5E15). A third party would be used to help retailers in this (I5E39, I10E42), but a third party can only be useful for specific analyses, if the organization has a clear idea of what is going to be analyzed and in what way (I8). When it comes to interpreting data, it is important and more reliable to have in-house or fully controlled data analysis capability (I12—The company did not provide services to competitors of its main client).

Besides interpreting the data, organizations should also be able to have foresight and understand how things can be improved with knock-on effects (I3E96). To do this, organizations need people who are motivated by data and who would be able to find solutions to the problems that might arise from the data, understand patterns, implications, and then develop and test hypotheses (I3E96, I3E98). These people are difficult to find and are expensive to keep (I3E97, I3E98).

Timeliness is a challenge for either implementing solutions or further enhancing current retail operations practices. Main issues include data processing and reporting, physical capability as well as managerial issues. Firstly, data need to be processed in a more timely manner, and this could mean that improvements in both hardware and software are needed. The IT infrastructure within the organization might be old and need to be updated to enable a faster and more sophisticated analysis (I6E36, I6E37). In addition, in terms of software, most retailers have a degree of batch processing built in, which hinders the system from handling information in real time (I10E59). Besides, in many cases data are available, but not reported in a format that can be worked on (I1E20, I2E22, I8E23); therefore, the overall process of analyzing and interpreting data is not prompt enough.

Other issues, such as not having the right skills set and managerial issues also have an influence on the overall responsiveness. Lastly, some companies do not have the physical capability with sufficient agility to leverage the data (I8E32, I12E29). For example, if retailers want to implement dynamic assortment in their stores, the supply chain would need to be able to handle more items in smaller quantities and on a more frequent basis (I10E19, I10E20), which may not be easily achieved.

When leveraging the big data for retail operations, it is important that suppliers provide the necessary support (I2E15, I5E41, I5E48). This is straightforward for a collaborative process such as space planning. In addition, supplier performance can also influence other activities. For example, if the supplier cannot deliver products as required by the retailer, no matter how advanced the replenishment process, availability cannot be ensured (I5E44). Retailers who performed better in availability mentioned that they have had good relationships with suppliers, and the supply chain is geared up to enable them to be responsive (I2E15). A retailer who realizes the problem has initiatives to interact with suppliers more frequently (I5E48).

#### 4.4.3. Technical and Data Issues

IT integration impacts data quality and data availability. Companies have many different systems managed by different functions historically, and linking these together is a challenge (I1E1, I11E25). While some companies have overcome this challenge (I1E9, I2E23, I12E8), some are just starting to tackle it (I4E5, I8E23, I11E26). The integration process, the lack of clarification in data generating processes, and different format of data lead to poor data quality (I4E29, I8E22). Manually created data are likely to have many errors, especially when there are several different systems in use (I5E54, I11E16, I11E17, I11E18).

The data available still have value to be discovered (I3E95, I6E23, I8E28, I10E36). Transactional data are available at a detailed level and storing the data is no longer an issue (I1E25, I8E28). The barrier now is how these data are reported and accessed (I10E38). This is partly a historical issue because systems have originally been set up for different purposes (I8E44, I8E46, I10E38). These data are handled in isolation by different functions resulting in different kinds of reporting schemes. Retailers try to maintain the integrity of these reports (I8E23) and develop reporting tools for further analysis (I1E26, I11E29). However, reporting requirements from other functions can change frequently (I1E18) and be challenging depending on the requirements (I1E19).

#### 4.4.4. Customer Concerns over Privacy

Lastly, customer concerns over privacy may hinder the retailer from implementing big data applications because it can prevent retailers from using customer data for specific operational decisions. When the customers realize the value of sharing their data to get for example personalized offers, they might not have an objection for their data to be used (I3E105). In the past, consumers were quite nervous about retailers having insight into their life (I3E8). More recently, they would expect personalized offers from retailers (I3E9). Customers' attitude towards big data solutions around their data have changed, because of the benefit they get from sharing their data. Although some people do not prefer sharing their data, the customer pool is sufficiently large to allow retailers to make inferences about shoppers who are not included in their loyalty schemes (I3E107).

Customer insights, as a most-recognized big data benefit, play a role in all these activities. In some cases, data do not necessarily need to be analyzed to yield insights. The sheer granularity and volume can help businesses improve the process, for example, in the case of optimizing replenishment deliveries according to customer demand. In addition, insights yielded from proposed applications, for example, customer demand, is not new in concept, but with better granularity and velocity, and if used more integrally it can lead to significant improvements. "The principle of supply chain has not really changed, what has changed is the degree of granularity and the acuteness of the need for all activities to be working optimally" (I10E30), and advanced technologies and big data applications enable this.

## 5. Discussion

### 5.1. Demand Sensing and Segmentation

In terms of demand sensing, some retailers do not have a systematic forecasting method and rely on personal judgements (I9E5, I9E15). Although this might not necessarily be less accurate, it is less sophisticated, more time-taking compared to using algorithms, more difficult to scale up and manage. Among retailers who have demand-forecasting algorithms, the most common input is historical sales data to identify the trend, and this could be based on the store level to make decisions more tailored to local demand (I2E3, I4E21, I9E1). Based on this, some retailers would build in elements that have an immediate impact in forecasting volume, such as delivery time windows, inventories, and weather forecasting (I3E47, I9E11, I3E80, I4E24, I5E25, I9E13). Customer insights, such as preferences and behavior could be included in forecast models to improve accuracy (I2E4, I2E14, I3E4, I3E5, I3E86, I3E90, I10E6) as a potential big data application [1]. However, currently, customer insights data are not yet implemented for short-term demand forecasting (I2E3, I2E4). Based on these findings, four levels of sophistication in demand sensing are summarized in Table 5.

**Table 5.** Different levels of sophistication in demand sensing.

Demand Sensing	Level 1	Level 2	Level 3	Level 4
Sophistication	Intuition or manual anticipation	Analyzing historical sales	Adding in other short-term factors, i.e., delivery time window, inventories, weather forecasting	Adding in customer behavior and preference

Segmenting stores according to the demand pattern is a common practice [56] and influences how stocks are deployed. Depending on the granularity, three levels of sophistication are identified in deploying stock according to segmentation (Table 6). Level 1 is planning on store type, this can be based on store roles such as destination stores or show-rooms, therefore they would have different requirements for stock levels (I5E6, I8E38, I8E41). Level 2 is planning on individual stores using the more granular transactional data (I2E31, I8E28). In addition, sales between stores with geographic proximity can be interrelated (I5E34, I12E27), therefore clusters of stores can be considered, for example, stores located in shopping malls or stores on high streets. Apart from deploying stock based on stores, another dimension of granularity, product, can also introduce differences. As the transactional data



are available, the decision could theoretically be made at the SKU level. In the case of launching new products, since no previous data are available, retailers would use data of products with similar features (I4E8, I4E9, I9E10). Planning at category level is quite common [20].

**Table 6.** Different levels of stock deployment.

Stock Deployment	Level 1	Level 2	Level 3
Granularity in terms of store	Per store type	Per store	Per store, considering clusters
Granularity in terms of product	Per category	Considering similar features	SKU

Replenishment is probably the most critical link to deliver high availability and it can comprise several components from measuring the gap on the shelf (availability), to ordering and replenishing. When retailers try to measure on-shelf availability, one way is to check stock manually (I9E19); therefore, in this case the availability standards would be on-shelf availability when the check is conducted. However, now with more demanding customers, the standard can be raised to on-shelf availability at any one time (I2E6, I10E55). One evidence indicates that when the grocery retailer starts fulfilling online orders from stores, the ‘real availability’ is revealed and therefore standards are changed from ‘in store’ to ‘any time’ [18]. Another way to measure availability is to compare actual sales and the pre-identified demand patterns and use variances as an indicator of stock-outs. Some retailers would compare this in days and some could compare it in minutes with more advanced technologies (I3E45, I3E46, I3E52, I3E53, I3E56, I9E18). Therefore, different levels of availability standards are summarized as in Table 7.

**Table 7.** Different levels of measuring availability.

Measuring Availability	Level 1	Level 2	Level 3	Level 4
Standard	In-store availability	On-shelf availability at certain point of time	On-shelf availability at any time	n/a
Measuring	Rely on stock data	Manual counting	Matching sales and demand pattern by days	Matching sales and demand pattern by minutes

Availability is a signal to inform replenishment and differences in how it is measured would also lead to different ways of how stores place their replenishment orders. A traditional practice would be to create a mass replenishment order for all items that have been sold on that day (I3E48). In this case, the ordering frequency would be once a day and the quantity is what is sold (base stock policy). Some retailers would use a reorder point (I9E3) which fixes the delivery quantity. A similar approach is to fix the delivery time slot (I3E47) where the quantity would be what is sold until the cut-off time. The differences are conceptualized in ordering frequency and quantity, as in Table 8.

**Table 8.** Different levels of store ordering.

Ordering From Stores	Level 1	Level 2	Level 3	Level 4
Frequency	Once a day	Depend on ordering points	Multiple times with fixed slots within a day	Multiple times with flexible deliveries depending on actual demand
Quantity	What is sold until ordering point (certain time or certain quantity)	Manual anticipation	Based on forecasting algorithm	What is sold until quantity reach a vehicle load

If the retailer is responsive enough, they would use the continuous replenishment scheme (I3E40), which is an innovative method to improve the efficiency of inventory management throughout the supply chain [57]. Although the approach proposed, which is once a day, is not as dynamic as it can be nowadays (several times a day, I3E50); it serves the point that response to actual demand is more efficient than anticipation, if the supply chain is responsive enough, as retailers do not need to worry about errors in forecast. In the ‘quantity’ row in Table 8, Level 1 and Level 4 are similar in nature

as they respond to actual demand, but the Level 4 practice requires a responsive supply chain and relatively more advanced technology.

When stocks arrive at stores, the next step in the replenishment process is to place the product on the shelves. As much as 25% of stock out situations are caused by issues in shelf replenishment [58]. Retailers try to make this process smoother to decrease the discrepancy between the stock data and what is on the shelf (I3E32). Replenishment efficiency will increase if products are moved directly from the receiving area to the shelf without extra handling (e.g., milk). Therefore, three levels for how the shelves can be replenished is presented in Table 9.

The backroom is the reason why stock data cannot reflect on-shelf availability and retailers have put efforts into reducing their backroom to have a smoother replenishment process from receiving area to the shelf. However, the backroom is more of an issue of overall supply chain responsiveness rather than the specific activity of filling the shelves. Retailers may have even more stock-outs if they reduce the backroom without increasing delivery frequency or shelf space.

**Table 9.** Different levels of shelf replenishment.

Filling the Shelves	Level 1	Level 2	Level 3
Efficiency	Taking items out from cases piece by piece	'One-touch' replenishment, without replenishing piece by piece	Without extra handling from receiving to shelves

Although improving availability is prioritized over managing substitutions, stock-out situations are not totally avoidable. When it does occur, what retailers choose to do may be different. These different levels of substitutions suggestions are presented in Table 10.

**Table 10.** Different levels of substitution management.

Substitution	Level 1	Level 2	Level 3	Level 4
Suggest substitution	Staff suggestion	Based on population statistics on accept rate	Based on individual statistics on accept rate	Personalized with other customer insights, i.e., brand or package size preferences

Ranging is an important decision with direct impact on sales [21,23]. Most suggestions on potential big data applications focus on using customer data to micro-segment customers and to optimize assortment; useful insights can be correlations between items purchased, time-dependent purchase patterns, and location-dependent demand patterns [2,25,26]. Although this is how customer insights are intended to be used, due to poor process integration or data privacy concerns, the idea is still immature compared to how it has been used in marketing activities (I10E5, I10E9, I10E10). Besides customer data, research involving suppliers into the planning process can be beneficial, but currently not all suppliers are involved, as some suppliers do not have the capability (I3E20, I9E32, I10E41). Assortment is normally planned at category level [20]. In interviews, almost all respondents suggest planning at SKU level where possible.

The main difference between retailers in range planning is the variety of input data for the analysis. Some retailers would use historical sales data only (I4E1, I4E2, I4E37), some would include product margin (I3E64, I6E14, I8E12), and some would include customer data from loyalty schemes (I10E11, I12E14, I12E19). In some cases, if suppliers have the capability, retailers would also share forecasts with them (I3E62, I9E33). Based on this, different levels of sophistication are summarized in Table 11.

The Level 2 practice in Table 11 excludes space planning because although product margins are quite often used as incentives for buyers, the space element usually acts as a pre-set constraint and buyers would not necessarily be involved in planning for maximizing space utilization (I6E9). Another difference could be how often retailers review and change their range. Retailers would review their range periodically and make changes, for example, every three or four months (I3E61, I4E23, I6E7). With big data, it is possible to make range changes within a day (I10E35). At this stage, we only identify current common practice (review and change every three to four months) and propose potential

applications (change within a day). Due to this reason, further research is needed to quantify and update the frequency of changes at different levels of practices.

**Table 11.** Different levels of sophistication in range planning.

Ranging	Level 1	Level 2	Level 3	Level 4
Sophistication in analyses	Based on past sales only	Add in margins (excluding space planning)	Add in consumer insights	Collaborative planning with suppliers

## 5.2. Pricing

Retailers would have two scenarios for making pricing decisions: for ‘normal’ price and for ‘promotions’, although in certain markets, a large portion of sales are driven by promotions and customers would expect a relatively large number of products to be on promotion at any time (I2E45). When this is the case, the ‘promotion’ prices are considered to be ‘normal’ prices.

### 5.2.1. ‘Normal’ Price

Airlines would take demand, supply, and competition into consideration for pricing decisions [28,29]. Similarly, dynamic pricing is a potential big data application for setting the ‘normal’ price in a retail environment [1]. Some retailers already implement dynamic pricing and test big data applications for this purpose (I5E28). However, retailers are more sensitive to competitors’ prices (I10E17) than anything else, so the elements built into the algorithm tend to include margins and competitors’ prices more than consumers’ actions (I2E41, I2E43, I5E28, I5E30). Nonetheless, stock data and consumer insights can be included in such analyses (I5E29, I10E52). Customer insights can help retailers obtain a more accurate picture of price elasticity of individual shoppers and make more refined pricing decisions (I12E15, I12E16, I12E17). Apart from using algorithms to update prices dynamically, some retailers would monitor their own pricing using sell-through rate as an indicator (I2E42), and some would have a pre-set price bands, so there is little analysis (I4E10, I9). These different practices are summarized in Table 12. Dynamic pricing analysis can be done on a daily basis, but due to the responsiveness in other processes, retailers suggest they would not change prices that often especially in a physical channel (I5E38, I5E29).

**Table 12.** Different levels of making ‘normal’ price decisions.

Pricing	Level 1	Level 2	Level 3	Level 4
‘normal’ pricing	Fix prices or price bands	Monitoring sell-through to inform changes	Dynamic pricing with considering margin and competitions	Dynamic pricing with considering margin, competitions, stock and customer insights

### 5.2.2. Planning for ‘Promotion’

The most naïve method of promotion would be to offer discounts to all customers for a certain period. Some retailers can now not only personalize what is to be discounted, but also decide on the best timing to offer discounts (I3E11). Consequently, four levels of promotion planning activities are identified in Table 13.

**Table 13.** Different levels of promotion planning.

Pricing	Level 1	Level 2	Level 3	Level 4
Promotion planning	Mass promotions	Personalized offers	Personalized offers and timings	Understand the source of sales lifts and the consequences when planning promotions

Mass-promotions stress the supply chain and may not increase the sales but motivate forward buying [4]. Big data can be used to analyze demand elasticity and evaluate sources of sales lifts to

decide promotion parameters [1]. However, even with big data, to forecast sales lifts is still not easy, and how customer insights data can be used to determine whether promotions pay for themselves is still a challenge, although possible in theory (I6E21, I7E6, I8E27, I8E28, I12E17, I12E51).

Not all retailers have a fixed layout design and some are flexible in their daily practices or have little planning on it (I5E35, I11E23). Some retailers use more advanced technologies for modelling shelf space, for which inputs include product dimensions, weight, shelf specifications, forecasted sales, and margin to maximize return (I3E78, I3E81, I3E84, I3E78, I6E9, I12E15). Advanced technologies may simplify the process of analyzing customer in-store behavior and therefore enable shelving to be designed in a more attractive manner [2,24,30]. Retailers can also consider the replenishment process when they design shelving to make the process more efficient (I3E80, I9E37). In terms of space utilization, a respondent suggests changing the range more frequently, maybe daily (I10E54). Table 14 presents the identified levels of sophistication for space planning.

**Table 14.** Different levels of space planning.

Store Layout Design	Level 1	Level 2	Level 3	Level 4
Utilization	Flexible	Model with product and shelf specification, and forecasted sales	Consider replenishment process	Plan frequent range changes

One of the purposes of studying customer in-store behavior is to make the design layout more attractive. Customer insights can be used to analyze correlations between items purchased in one basket and inform adjacencies of products, although this is still conceptual (I3E94, I10E50, I10E52, I12E28). However, retailers study the trends in the long-term (I2E36, I3E63). One mature application is to produce heat-maps of in-store traffic and place less popular items around hot items (I2E34, I2E35, I3E87, I3E93). Again, some retailers are quite flexible in their design, as reflected in Table 15.

**Table 15.** Different levels of designing for attractiveness.

Store Layout Design	Level 1	Level 2	Level 3	Level 4
Attractiveness	Flexible	Heat-mapping based on in-store traffic	Using customer insights to identify long-term trends	Using customer insight to inform adjacencies

Recent research on three online retailers which have massive data sets compared to brick and mortar stores improves forecasting, learning, and price optimization by exploiting these massive data sets on sell-through distributions of products to cope with high demand uncertainty, short product life cycle, and limited inventory [59]. These are the key issues for not only emerging online-only retailers but also omni-channel retailers who frequently face the issue of stocking out in one channel (e.g., online store) and having too much stock and facing excess inventory costs in the other channel (e.g., physical store).

### 5.3. Challenges

Main challenges associated with the abovementioned applications are people, supplier management, and IT integration. In some cases, challenges identified under these headings could have different meanings. In other cases, the challenges are prerequisites for successful implementation, and the issue is whether the organization has the problem rather than to what extent they have the problem.

There are shortages in terms of people who can conduct deep analysis and managers who know how to leverage insights form data, so that the two parties can work together and apply big data solutions to business [1]. In fact, respondents think what is really in utmost shortages are people who have both analytical skills and the business awareness that can work with data in an integrated way to draw foresight and understanding of how things can be improved (I3E96, I3E97, I3E98). However, the retail sector also needs people who can interpret data, especially when information is integrated (I5E15), and sometimes there are also shortages of people who can use the system effectively (I4E30). Based on these views, four levels of capability requirements on people are summarized in Table 16.

**Table 16.** Different levels of capability requirements on people.

People	Level 1	Level 2	Level 3	Level 4
Capability requirements on people	People who can use the system	People with deep analytical skills and managers who can use insights	People who can interpret data and apply to business	People who can interpret data, apply to business and develop hypotheses for improvement

As some of the applications above also include collaboration with suppliers, it is not surprising that supplier management is a challenge and require different levels of capability, as summarized in Table 17. Some retailers cannot deliver availability because suppliers are unable to supply on-time-in-full (I5E44); therefore, in their big data implementation plan, ensuring that suppliers can improve their capability is a critical link (I5E41, I5E48). Some retailers have good relationships with suppliers historically; therefore, they can implement new changes to improve their responsiveness with less difficulty (I4E40, I4E41, I9E32). To support decisions driven by big data, the supply chain needs to be coordinated accordingly (I2E15).

**Table 17.** Different levels of supplier capability.

Supplier Management	Level 1	Level 2	Level 3	Level 4
Supplier capability	Poor performance	Improve performance with retailers	Have good performance and relationship with retailers	The whole supply chain coordinated

The degree of IT integration has a significant impact on implementing big data solutions, and organizations are at different stages of IT integration, which also depends on the degree of process integration. For example, in an organization, sales data and inventory data may not be integrated and people who do range planning would not have access to inventory data, because historically these data are not used outside merchandising or the supply chain (I4E37). Since different reporting schemes and tools are used in the past, systems integration has been a challenge (I1E1, I11E25). An additional problem is the inconsistency in reporting or even inputting the data which leads to poor data quality (I4E29, I8E22, I11E16, I11E17, I11E18). In some cases, although the IT system is integrated, the users of data do not understand it in a systematic way and therefore change their requirements frequently (I1E18, I1E19). The best practices found are having an integrated information system driven by processes, which allows the user to understand how data flows and how it can support the process (I3E58, I12E34). The different levels of IT integration are presented in Table 18.

**Table 18.** Different levels of IT integration.

IT Integration	Level 1	Level 2	Level 3	Level 4
IT integration	Handles data in isolation	Intends to integrate system, but reporting is inconsistent	IT integrated, but user cannot understand it and changes requirements frequently	IT system driven by process

Besides the above-mentioned challenges, managerial issues (open to new possibilities and process integration, timeliness, and privacy of customer data) act as prerequisites, but could not be shown at progressive levels. The decision-making culture and management practices are the critical and fundamental issues in implementing big data analysis [52,60]. Processes need to be integrated and information needs to be shared. The technology is available, but needs it to be applied (I8E28, I10E38, I12E7); again, the barrier is the old-fashioned way of running the business (I5E17, I10E60). If the process is not integrated, people may interpret data for their own interest and arrive at different results (I5E17). Timeliness refers to both the data processing and the physical capability. In short, most of the current or potential practices can be compared as different maturity levels under the same theme, as shown in Table 19.

Table 19. Big data maturity profile.

Category	Indicator	Level 1	Level 2	Level 3	Level 4
Demand sensing	Sophistication	Intuition or manual anticipation	Analyzing historical sales	Adding other short-term factors, i.e., delivery time windows, inventories, weather forecasting	Adding in customer behavior and preference
Stock deployment	Granularity in terms of store	Per store type	Per store	Per store, considering clusters	n/a
	Granularity in terms of product	Per category	Considering similar features	Per SKU	n/a
Measuring availability	Standard	In-store availability	On-shelf availability at certain point of time	On-shelf availability at any time	n/a
	Measuring	Rely on stock data	Manual counting	Matching sales and demand pattern by days	Matching sales and demand pattern by minutes
Store order process	Frequency	Once a day	Depends on order points	Multiple times with fixed slots within a day	Multiple times with flexible delivery depending on actual demand
	Quantity	What is sold until reorder point (certain time or certain quantity)	Manual anticipation	Based on forecasting algorithm	What is sold until quantity reaches a vehicle load
Shelf replenishment	Efficiency	Taking items out from cases piece by piece	'One-touch' replenishment, without replenishing piece by piece	Without extra handling from receiving to shelves	n/a
Substitution management	Suggest substitution	Staff suggestion	Based on population statistics	Based on individual statistics on accept rate	Personalized with other customer insights, i.e., brand or package size preferences
Ranging	Sophistication in analyses	Based on past sales only	Add margins (excluding space planning)	Add consumer insights	Collaborative planning with suppliers
Pricing	'Normal' pricing	Fixed prices or price bands	Monitoring sell-through to inform changes	Dynamic pricing considering margin and competitors	Dynamic pricing considering margin, competitors, stock and customer insights
	Promotion planning	Mass promotions	Personalized offer	Personalized offer and timing	Understand sources of sales lifts and the consequences
Store layout design	Utilization	Flexible	Modeling with product and shelf specification, forecasted sales	Consider replenishment process	Plan with frequent range changes within a day
	Attractiveness	Flexible	Heat-mapping based on in-store traffic	Using customer insights to identify long-term trends	Using customer insight to inform adjacencies
Challenges	Capability requirements on people	People who can use the system	People with deep analytical skills (and with managers who can use insights)	People who can interpret data and apply to business	People who can interpret data, apply to business, and develop hypotheses for improvement
	Supplier capability	Poor performance	Improve performance with retailers	Have good performance and relationship with retailers	The whole supply chain coordinated
	IT integration	Data handled in isolation	Intends to integrate system, but reporting is inconsistent	IT integrated, but user cannot understand it and changes requirements frequently	IT system driven by process

## 6. Conclusions

This paper set out to identify big data applications for the retail sector. Decisions around availability, assortment, pricing, and layout planning are identified as key retail operations that can benefit from more advanced data processing and analysis. Although there are several definitions of big data, we propose a comprehensive definition of all population data that can be accessed and processed in a timely fashion to improve operations and processes. Existing applications of big data in retail operations concern sophisticated analyses and access to high volumes of data in a short period of time. *Existing applications are summarized as follows:*

- Using loyalty scheme data to identify consumer preferences and lifestyles, which is referred to as consumer insights, and using these insights in marketing activities.
- Using transaction data together with delivery time windows, inventories, and weather forecasts to produce more accurate demand forecasts and inform replenishment and procurement accordingly.
- Analyzing transactions at in-store SKU level to inform stock deployment.
- Matching pre-identified demand patterns of each SKU with actual sales in minutes to identify gaps on the shelves.
- Transmitting sales data in real time to optimize replenishment transport around consumer demand.
- Using consumer insights to personalize substitutions; therefore, increasing acceptance rates.
- Making pricing decisions based on margins and competitors' actions.
- Using loyalty scheme data to personalize promotional offers and timing.
- Integrating data on product specifications, shelf specifications, and forecasted demand to plan layout and subsequently to maximize margins in a limited space.
- Heat-mapping in-store traffic to inform layout planning and increase exposure of less popular items to boost their sales.
- Using consumer insights to identify long-term trends and inform layout planning, to make the store more attractive.
- Using consumer insights to understand 'missions' of shopping trips for each consumer and to use this information in layout planning.

During the interviews, respondents also mentioned that some applications are either still conceptual or planned. *Potential applications of big data in retail operations are summarized as follows:*

- Using transactional data, delivery time windows, inventories, weather forecasts, and consumer insights to improve demand forecast accuracy and inform replenishment.
- Analyzing transactional data at store level, while considering cluster performance to inform stock deployment.
- Using consumer insights to localize ranging on stores, increasing sales and improving space utilization.
- Using consumer insights and demand patterns within a day to adjust the range within a day, increase sales, and improve availability and space utilization.
- Regularly updating prices based on margin, competitions, inventories, and consumer insights to maximize profit.
- Using consumer insights to understand the sources of sales lifts during promotions and better control promotions.
- Using consumer insights to inform adjacencies in layout planning, to make the store more attractive and generate more sales.

### 6.1. Roadmap from Current to the Future

Some applications aim to answer the same question, underpinning different levels of maturity for the same activities. The complete set of maturity profiles is shown in Table 19. In short, the higher

levels of practices would use data in a more granular and timely manner, and therefore can be more responsive.

To implement big data solutions, there are also some prerequisites and challenges that should be overcome. For example, companies need to be adventurous and processes need to be integrated first. In addition, depending on the level of sophistication, capability requirements would also vary, as already summarized in the Discussion section. In general, the responsiveness of processes need to be able to support the decision-making cycle, and this could be affected by issues related to the right skills set, support from suppliers, and suitable IT systems. The outcomes of this research are based on the literature review that informed the interviews. Although the research comprises a small sample, the findings are still generalizable to a certain extent and useful to retailers and other interested parties.

This research provides a general overview of how big data is used in the retail operations, identifies some future possibilities, and challenges for implementation in the areas of availability, assortment, pricing, and layout planning. Further research can be carried out in one of those areas and explore some of the identified solutions to quantify the benefits. We propose a maturity profile that contains gaps which may be filled and tested in the future.

## 6.2. Managerial Implications

The findings can be used to inform the business in terms of how retailers can leverage big data to be more effective in assortment, pricing, layout planning, and availability. We compare findings from different respondents with the literature and summarize them in the form of a maturity profile, which can be used for benchmarking. Retail logistics, especially internet-enabled and ICT-supported real time logistics have emerging and progressive applications such as RFID-based product tracking, assisted driving via on-board vehicle systems integrated with intelligent traffic applications, monitoring of environmental parameters for perishable goods and augmented mapping [61]. Challenges highlighted for big data applications extend to this domain in terms of managerial issues, people and skills shortages, and technical integration. However, it is impossible to decouple big data applications in retail operations from the logistics as it is a key aspect of fulfilment.

**Supplementary Materials:** Transcripts of interviews with interview and evidence numbers can be requested from the corresponding author.

**Acknowledgments:** This paper is produced out of an MSc thesis completed as part of the MSc Logistics and Supply Chain Management Programme at Cranfield School of Management. LCP Consulting made connections and covered the costs of the second author in accessing the retailers who were interviewed to explore the big data applications in the retail sector. The authors agreed to submit this paper to an open access journal on the condition that article processing charges are waived.

**Author Contributions:** E.A. supervised the work undertaken as part of this research paper and Y.M. collected the primary data and wrote an MSc thesis. E.A. transformed the thesis material into an academic paper and rewrote the paper in the style expected in an academic paper.

**Conflicts of Interest:** The authors declare no conflict of interest. Thesis sponsor LCP Consulting was not involved in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## Appendix A. Literature Review Process

To perform the review, we identified the search strings in Table A1 and present the number of hits from each within the four databases used. The character “\*” in Table A1 is a wild card, which allows the database to search everything that starts with the word before the “\*”. For example, “collect\*” will force the database to return “collection”, “collectible”, “collectable”, “collected”, “collective”, “collector”, “collecting”. We limited the results to title, abstract, and keywords. We screened the papers returned from the search based on the title and abstract.



**Table A1.** Search strings and number of result.

Search Strings	ABI/Inform	EBSCO	Science Direct	Emerald
ti("big data") OR ab("big data")	12,315 (275)	2450 (318)	324	10
(ti("big data") OR ab("big data")) AND all(collect *)	505 (45)	-	239	-
(ti("big data") OR ab("big data")) AND all(tech *)	5246 (169)	-	308	-
(ti(retail *) OR ab(retail *) AND (ti(challeng *) OR ab(challeng *)))	282,724 (8477)	8380 (1053)	339	313
ti(retail *) AND (ti(challeng *) OR ab(challeng *))	4339 (364)	1453 (453)	74	293
ti(retail *) AND ti(challeng *)	1026 (60)	300 (70)	-	-
ti(retail *) AND ti(challeng *) NOT all(bank *)	857 (52)	245 (56)	-	-
(ti("big data") OR ab("big data")) AND (ti(retail *) OR ab(retail *))	525 (5)	100 (9)	3	0
(ti("big data") OR ab("big data")) AND (all(retail *))	609 (6)	514 (44)	52	4
(ti("big data") OR ab("big data")) AND (all(supply chains))	156 (3)	198 (21)	42	1

In Table A1, the number given in brackets refers to the number of academic articles. Search strings presented in the table are written in the format used by ABI/Inform. The justifications for the search strings are provided in Table A2.

**Table A2.** Justifications for search strings.

Search Strings	Purpose
ti("big data") OR ab("big data")	To get a general understanding of big data.
(ti("big data") OR ab("big data")) AND all(collect *)	To get information on what data are currently collected, and corresponding applications.
(ti("big data") OR ab("big data")) AND all(tech *)	To get some ideas about technical enablers of big data.
(ti(retail *) OR ab(retail *) AND (ti(challeng *) OR ab(challeng *)))	To find trends and challenges currently discussed.
ti(retail *) AND (ti(challeng *) OR ab(challeng *))	To find trends and challenges in retail operations.
ti(retail *) AND ti(challeng *) NOT all(bank *)	To exclude results on banking.
(ti("big data") OR ab("big data")) AND (ti(retail *) OR ab(retail *))	To find researches on big data applications in retailing.
(ti("big data") OR ab("big data")) AND (all(retail*))	Since last search strings returned too few result, the search field for key word 'retail*' are change to 'all.
(ti("big data") OR ab("big data")) AND (all(supply chain))	To find big data applications in supply chain in general, as these could apply to retail supply chains.

After screening by title, abstracts of returned results are read first, if relevant, the article is then skimmed and decided to be included in the review if suitable. We used cross-referencing to identify other relevant articles. For example, where the conclusions from an article are cited in another article that is already in the review, the cited articles are found and reviewed for inclusion. The literature review process is illustrated in Figure A1.

Managerial articles are preferred over technical ones, such as those categorized under computer science and those research on algorithms. Besides the main review process, other topics researched include Amazon and anticipatory shipping, Apple and iBeacon, Google and its privacy policies, Dunnhumby and its customer data analysis practices, Tesco and its multi-channel fulfillment operation, and dynamic pricing in the airline industry.

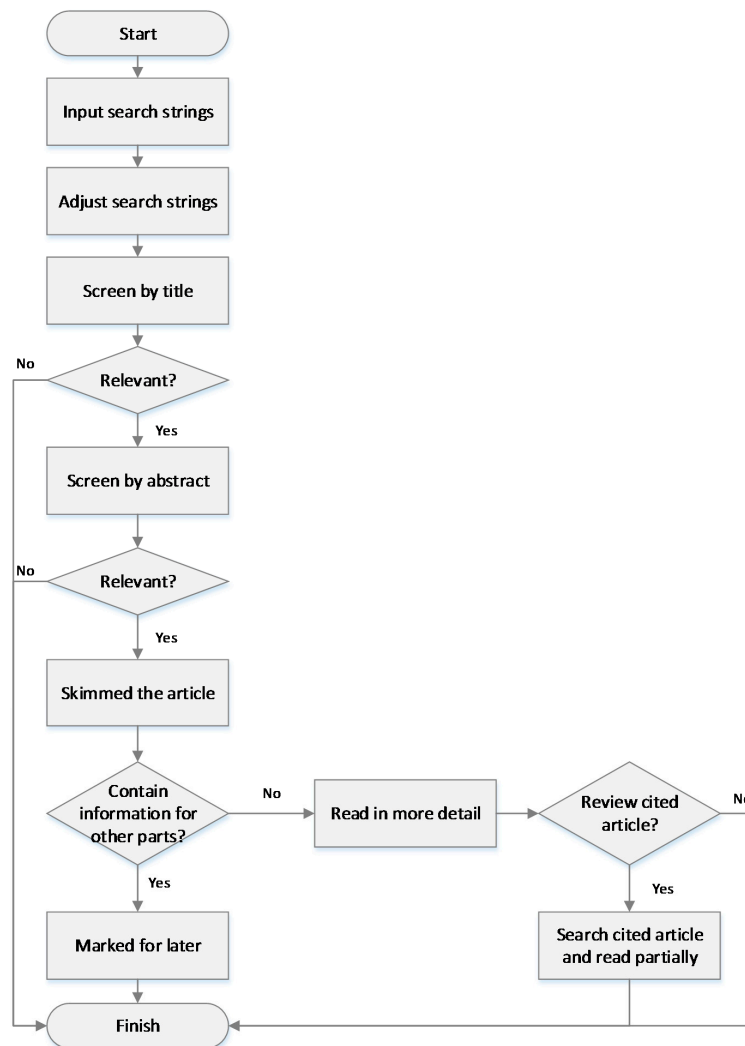


Figure A1. Literature review processes.

## References

1. Brown, B.; Bughin, J.; Byers, A.H.; Chui, M.; Dobbs, R.; Manyika, J.; Roxburgh, C. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*; Technical Report for Mckinsey&Company: Washington, DC, USA, 2011.
2. Google Buys UK Retail Forecasting Firm Rangespan. Available online: <http://www.bbc.co.uk/news/technology-27261504> (accessed on 24 September 2017).
3. Boomerang Commerce Closes \$8.5 mln Series A. Available online: <https://www.pehub.com/2014/07/boomerang-commerce-closes-8-5-mln-series-a/> (accessed on 24 September 2017).
4. Fisher, M. What Is the Right Supply Chain for Your Product? *Harv. Bus. Rev.* **1997**, *75*, 105–117.
5. O'Regan, G. *A Brief History of Computing*, 1st ed.; Springer: Berlin, Germany, 2008; pp. 53–62. ISBN 978-1-4471-2359-0.
6. Hurwitz, J.; Nugent, A.; Halper, F.; Kaufman, M. *Big Data for Dummies*, 1st ed.; John Wiley & Sons: Hoboken, NJ, USA, 2013; pp. 9–24. ISBN 978-1-118-50422-2.
7. Davenport, T.H. Analytics 3.0. *Harv. Bus. Rev.* **2013**, *91*, 64–72.
8. The Internet of Things. Available online: <http://www.mckinsey.com/industries/high-tech/our-insights/the-internet-of-things> (accessed on 24 September 2017).
9. Big Data Analytics. Available online: <https://www.ibm.com/analytics/us/en/big-data/> (accessed on 24 September 2017).

10. TDWI Best Practices Report | Big Data Analytics. Available online: <https://tdwi.org/research/2011/09/best-practices-report-q4-big-data-analytics.aspx> (accessed on 24 September 2017).
11. McAfee, A.; Brynjolfsson, E. Big Data: The Management Revolution. *Harv. Bus. Rev.* **2012**, *90*, 60–68. [PubMed]
12. Balfour, G. Big Data's Big Promise. *Backbone* **2013**, *4*, 12–15.
13. Chen, H.; Chiang, R.H.L.; Lindner, C.H.; Storey, V.C.; Robinson, J.M. Business intelligence and analytics: From big data to big impact. *MIS Q.* **2012**, *36*, 1165–1188.
14. Jacobs, A. The pathologies of big data. *Commun. ACM* **2009**, *52*, 36–44. [CrossRef]
15. Clouds, Big Data, and Smart Assets: Ten Tech-Enabled Business Trends to Watch. Available online: <http://www.mckinsey.com/industries/high-tech/our-insights/clouds-big-data-and-smart-assets-ten-tech-enabled-business-trends-to-watch> (accessed on 24 September 2017).
16. Purcell, B. The emergence of “big data” technology and analytics. *J. Technol. Res.* **2013**, *4*, 1–7.
17. Mayer-Schönberger, V.; Cukier, K. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, 1st ed.; Houghton Mifflin Harcourt: Boston, USA, 2013; pp. 100–121, ISBN 978-1848547926.
18. Fernie, J.; Sparks, L. *Logistics and Retail Management*, 4th ed.; Kogan Page: London, UK, 2014; pp. 1–28, ISBN 9780749468231.
19. Levy, M.; Weitz, B.A. *Retailing Management*, 9th ed.; McGraw-Hill Irwin: New York, NY, USA, 2013; pp. 514–539, ISBN 978-1259060663.
20. Agrawal, N.; Smith, S.A. Supply Chain Planning Processes for Two Major Retailers. In *Retail Supply Chain Management*, 1st ed.; Agrawal, N., Smith, S.A., Eds.; Springer: Berlin, Germany, 2008; pp. 11–23, ISBN 978-1-4899-7562-1.
21. Fernie, J.; Sparks, L.; McKinnon, A.C. Retail logistics in the UK: Past, present and future. *Int. J. Retail Distrib. Manag.* **2010**, *38*, 894–914. [CrossRef]
22. Ayers, J.B.; Odegaard, M.A. *Retail Supply Chain Management*, 2nd ed.; Auerbach Publications: Boca Raton, FL, USA, 2008; pp. 3–13, ISBN 9781498739146.
23. Customer Science: Growing Business with Customer Data. Available online: <https://www.dunnhumby.com/customer-science-growing-business-customer-data> (accessed on 24 September 2017).
24. Bala, P.K. *Data Mining for Retail Inventory Management*, 1st ed.; Springer: Dordrecht, The Netherlands, 2009; pp. 587–598, ISBN 978-90-481-2311-7.
25. Videla-Cavieres, I.F.; Rios, S.A. Extending market basket analysis with graph mining techniques: A real case. *Expert Syst. Appl.* **2014**, *41*, 1928–1936. [CrossRef]
26. Grewal, D.; Levy, M. Retailing research: Past, present, and future. *J. Retail.* **2007**, *83*, 447–464. [CrossRef]
27. Narangajavana, Y.; Garrigos-Simon, F.J.; García, J.S.; Forgas-Coll, S. Prices, prices and prices: A study in the airline sector. *Tourism Manag.* **2014**, *41*, 28–42. [CrossRef]
28. Clark, R.; Vincent, N. Capacity-contingent pricing and competition in the airline industry. *J. Air Transp. Manag.* **2012**, *24*, 7–11. [CrossRef]
29. Underhill, P. *Why We Buy: The Science of Shopping*, 1st ed.; Simon & Schuster Pbks: New York, NY, USA, 2009; pp. 75–85, ISBN 978-1587990441.
30. Williams, D.E. The evolution of e-tailing. *Int. Rev. Retail. Distrib. Consum. Res.* **2009**, *19*, 219–249. [CrossRef]
31. Verdict in Association with SAS, How the UK Will Shop: 2013. Available online: <http://fliphtml5.com/gxqh/qzzy/basic> (accessed on 24 September 2017).
32. Xing, Y.; Grant, D.B. Developing a framework for measuring physical distribution service quality of multi-channel and ‘pure player’ internet retailers. *Int. J. Retail Distrib. Manag.* **2006**, *34*, 278–289. [CrossRef]
33. Rushton, A.; Croucher, P.; Baker, P. *The Handbook of Logistics and Distribution Management*, 5th ed.; Kogan Page: London, UK, 2014; pp. 83–85. ISBN 9780749466275.
34. Spiegel, J.R.; McKenna, M.T.; Lakshman, G.S.; Nordstrom, P.G. Method and System for Anticipatory Package Shipping. U.S. Patent 8,271,398, 18 September 2012.
35. Bhattacharjya, J.; Ellison, A.; Tripathi, S. An exploration of logistics-related customer service provision on Twitter. *Int. J. Phys. Distrib. Logist. Manag.* **2016**, *46*, 659–680. [CrossRef]
36. Pramatar, K.; Evgeniou, T.; Doukidis, G. Implementation of collaborative e-supply-chain initiatives: An initial challenging and final success case from grocery retailing. *J. Inf. Technol.* **2009**, *24*, 269–281. [CrossRef]
37. Hayashi, A.M. Thriving in a Big Data World. *MIT Sloan Manag. Rev.* **2014**, *55*, 35–39.

38. Harris, L.; Dennis, C. Engaging customers on Facebook: Challenges for e-retailers. *J. Consum. Behav.* **2011**, *10*, 338–346. [CrossRef]
39. About iBeacon on Your iPhone, iPad, and iPod touch. Available online: <https://support.apple.com/en-gb/HT202880> (accessed on 24 September 2017).
40. Fashion Retailer Tries on Apple's iBeacon for Size. Available online: <https://blogs.wsj.com/cio/2014/02/13/fashion-retailer-tries-on-apples-ibeacon-for-size/> (accessed on 24 September 2017).
41. Cumbley, R.; Church, P. Is 'Big Data' creepy? *Comput. Law Secur. Rev.* **2013**, *29*, 601–609. [CrossRef]
42. Church, P.; Kon, G. Google at the heart of a data protection storm. *Comput. Law Secur. Rev.* **2007**, *23*, 461–465. [CrossRef]
43. Time for Action on Google's Privacy Policy. Available online: <https://bigbrotherwatch.org.uk/2013/04/time-for-action-on-googles-privacy-policy/comment-page-1/> (accessed on 24 September 2017).
44. France Fines Google 150,000 Euro after It Refuses to Comply with Data Protection Act. Available online: <https://techcrunch.com/2014/01/08/france-fines-google-150000-euro-after-it-refuses-to-comply-with-data-protection-act/> (accessed on 24 September 2017).
45. Boyd, D.; Crawford, K. Critical Questions for Big Data. *Inf. Commun. Soc.* **2012**, *15*, 662–679. [CrossRef]
46. Redman, T.R. Data's Credibility Problem. *Harv. Bus. Rev.* **2013**, *91*, 84–88.
47. GS1 UK; IBM; The Institute of Grocery Distribution; Cranfield School of Management, (KTP project); Value Chain Vision. *Data Crunch Report: The Impact of Bad Data on Profits and Customer Service in the UK Grocery Industry*; Technical Report for GS1 UK and Cranfield University School of Management: Bedford, UK, 2009.
48. Hazen, B.T.; Boone, C.A.; Ezell, J.D.; Jones-Farmer, L.A. Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *Int. J. Prod. Econ.* **2014**, *154*, 72–80. [CrossRef]
49. Jones-Farmer, L.A.; Ezell, J.D.; Hazen, B.T. Applying Control Chart Methods to Enhance Data Quality. *Technometrics* **2014**, *56*, 29–41. [CrossRef]
50. Santaferraro, J. Filling the Demand for Data Scientists: A Five-Point Plan. *Bus. Intell. J.* **2013**, *18*, 13–18.
51. Lavalley, S.; Lesser, E.; Shockley, R.; Hopkins, M.S.; Kruschwitz, N. Big Data, Analytics and the Path from Insights to Value. *MIT Sloan Manag. Rev.* **2011**, *52*, 21–32.
52. Ferguson, R.B. How Analytics Can Transform Business Models. *MIT Sloan Manag. Rev.* **2013**, *54*, 1–8.
53. Maylor, H.; Blackmon, K.L.; Huemann, M. *Researching Business and Management*, 2nd ed.; Palgrave: Basingstoke, UK, 2017; pp. 101–127, ISBN 978-0333964071.
54. Marshall, M.N. Sampling for qualitative research. *Fam. Pract.* **1996**, *13*, 522–526. [CrossRef] [PubMed]
55. Patton, M.Q. *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*, 4th ed.; SAGE: London, UK, 2015; pp. 311–315, ISBN 9781412972123.
56. Moye, L.N.; Kincade, D.H. Shopping orientation segments: Exploring differences in store patronage and attitudes toward retail store environments among female apparel consumers. *Int. J. Consum. Stud.* **2003**, *27*, 58–71. [CrossRef]
57. Cachon, G.; Fisher, M. Campbell soup's continuous replenishment program: Evaluation and enhanced inventory decision rules. *Prod. Oper. Manag.* **1997**, *6*, 266–276. [CrossRef]
58. Corsten, D.; Gruen, T. Desperately seeking shelf availability: An examination of the extent, the causes, and the efforts to address retail out-of-stocks. *Int. J. Retail Distrib. Manag.* **2003**, *31*, 605–617. [CrossRef]
59. The New Frontier in Price Optimization. Available online: <http://cdn.executive.mit.edu/86/ce/f0157f5a418c8b670a5f2295d5f1/mit-sloan-executive-education-webinar-david-simchi-levi-the-new-frontier-in-price-optimization.pdf> (accessed on 1 September 2017).
60. Brown, B.; Chui, M.; Manyika, J. *Are you Ready for the Era of 'Big Data'?* Technical Report for McKinsey & Company: Washington, DC, USA, 2011.
61. Wang, Y. *E-Logistics*, 1st ed.; Kogan page: London, UK, 2016; pp. 14–17, ISBN 9780749472665.

