How Big Data Analytics can Contribute to the Marketing Performance of Supermarkets

Bachelor Thesis Business and Consumer Studies

Author:	Koen van der Schaaf
Student no.	921225728020
University:	Wageningen University
Chair Group:	Management Studies
Supervisor/Examiner:	dr. HB Kok
2nd Supervisor/Examiner:	dr. JLF Hagelaar

Course code:	YSS-81812
Study load:	12 credits
Date:	21th of August 2017, Wageningen

WAGENINGEN UR
For quality of life

Abstract

This thesis aims to create insight into how supermarkets can use Big Data Analytics in order to enhance their marketing performance. The methodology used in this thesis is a literature review. Papers from peer-reviewed journals are analysed, used, and put into context.

It was found that Big Data Analytics offers businesses, including supermarkets, new possibilities to collect business intelligence. With the use of Big Data Analytics supermarkets are able to gather information about their customers and products faster and from more sources than before. The insights Big Data Analytics can provide, offer possibilities for supporting decision making regarding supermarkets' marketing performance. Examples are given of current and future applications for using Big Data Analytics for marketing purposes in supermarkets. Challenges, such as privacy and data quality, regarding the use of Big Data in supermarkets are also discussed.

Table of Contents

Abstract ii

1 – Introduction 1

1.1 Motivation and Background: *Why Big Data Analytics could be used by supermarkets in order to enhance their marketing performance* 1

1.2 Relevance 2

1.3 Goals 3

1.4 Outline 3

2 – Method 4

3 – Which technical and BI conditions do datasets need to meet for usage in BDA? 5

3.1 Big Data 5

3.2 Technical Conditions 5

3.3 BI Conditions 6

3.4 Big Data Analytics 8

4 – What is a supermarket and which datasets are available to measure its marketing performance? 9

4.1 The retail industry: *What are the characteristics of supermarkets?d* 9

4.2 Aspects of marketing performance: *Which types of datasets are available for supermarkets to measure marketing performance?* 11

4.3 Empirical Confirmation of available datasets 17

5 – Can data that is used for measuring the marketing performance of supermarkets be utilized in Big Data Analytics? 18

5.1 Characterizing the marketing performance data of supermarkets 18

5.2 *Do supermarket datasets for determining marketing performance meet the conditions to be used in Big Data analytics?* 22

5.3 Current and Future Practices: *How can Big Data Analytics enhance the marketing performance of supermarkets?* 22

6 Discussion & Conclusion 24

References 27

1 – Introduction

This chapter elaborates on the motivation to write this thesis and how it is scientifically relevant. It also describes the goals that are set for this research and the structure of this thesis. Firstly this chapter explains how the subjects Big Data, supermarkets and marketing performance are linked to each other and why research on these linkages is necessary.

1.1 Motivation and Background: *Why Big Data Analytics could be used by supermarkets in order to enhance their marketing performance*

Why Big Data Analytics?

The current interest in Big Data from both businesses and academics is justifiable, as Big Data has the potential of enhancing operational and strategic processes of businesses (Fosso Wamba et al., 2015). Big Data is a term that is used to describe datasets whose size is so large that, until a few years ago, it was difficult to store and analyse them (Manyika et al., 2011). However the recent increase of storage capacity, computational power and advances in software development have made it possible to start analysing these large datasets (Emrouznejad & Marra, 2016). The usage of the term Big Data in scientific literature has increased rapidly in recent years, see figure 1. Leading technology companies, such as IBM and SAP, have contributed to this by promoting their analytics products (e.g. IBM's Watson Analytics and SAP's HANA) for Big Data (Gandomi & Haider, 2015). The sources of the datasets of which Big Data consists are diverse and can range from, for example: web and social media data, to machine data, to sensor data, to transaction data and even to the "Internet of Things" (Mohanty, 2015).

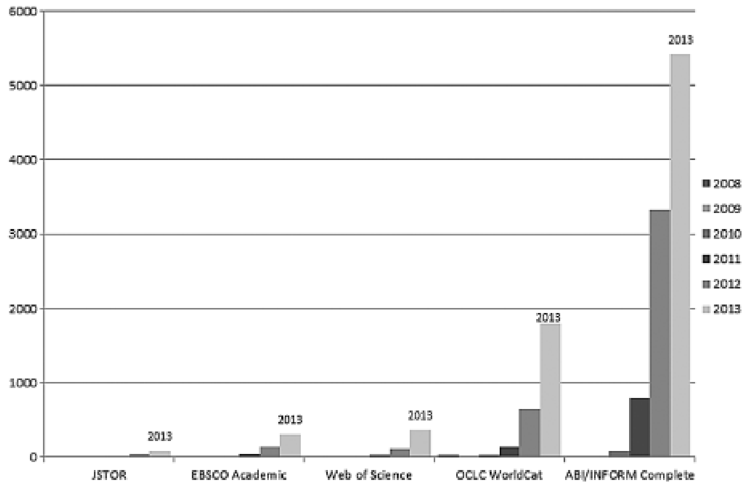

Figure 1: Number of publication with the phrase "Big Data" in five academic databases between 2008 and 2013
Source: Ekbia et al. (2015)

On its own Big Data can have little meaning, however when processed it can improve the decision making process for managers (McAfee & Brynjolfsson, 2012). In order to gain insights from Big Data one needs to process it, analyse it, and put it into context (Tien, 2013). This process is called Big Data Analytics (BDA) (LaValle et al., 2011). BDA is a tool that fits within a business intelligence system (Chen, Chiang & Storey, 2012). Business intelligence (BI) is all "relevant information and knowledge describing the business environment, the organization itself, and its situation in relation to its markets, customers, competitors, and economic issues." (Lönnqvist & Pirttimäki, 2006, P. 32). BI can "help an enterprise to better understand its business and market and make timely business decisions." (Chen et al., 2012, P. 1166). Analysing datasets in order to create BI is not a new concept, and the term business intelligence has been around since the 1990s (Ranjan, 2008; Chen et al., 2012) However BDA offers new tools and directions to obtain BI and can help businesses with creating insights for sustained value delivery, performance measurement, and establishing competitive advantages (Chen et al., 2012; Fosso Wamba et al., 2015; Tien, 2013).

Relevance to supermarkets

In industries with heavy competition, firms are required to continuously improve their activities in order to remain competitive (Porter, 1985). The supermarket industry can be classified as such an industry (Grewal, Krishnan, Levy & Munger, 2010; Sirohi, McLaughlin & Wittink, 1998). As explained above BDA offers possibilities for businesses to improve their activities and remain competitive. Another reason why BDA might offer interesting opportunities for supermarkets, is that they can collect large amounts of data on a daily basis. Wal-Mart, for example, collects 2.5 petabytes (1 petabyte = 1000 terabytes = 1.000.000 gigabytes) every hour from their customer transactions alone (McAfee & Brynjolfsson, 2012). Wal-Mart is one of the largest retailers in the world (Grewal et al., 2010), thus most supermarket chains will probably generate less data. But considering the speed at which these data are generated and storage volume they take up, BDA is an interesting option to process the datasets that are generated and collected by supermarkets. Also BDA can be used to determine consumer trends and preferences, which is relevant to supermarkets (Ridge, Johnston & O'Donovan, 2015; Grover & Kar, 2017)

Why marketing performance?

In order to answer what marketing performance is, it first must become clear what marketing exactly is. Marketing is "the task of finding and stimulating buyers for the firm's output. It involves product development, pricing, distribution, and communication" (Kotler & Levy, 1969, P. 10). This is an important activity for supermarkets, as stimulating their customers to buy more can have a positive impact on the sales of a retailer (Preuss, 2014; Berry, 2001; Zentes, Morschett & Schramm-Klein, 2007). However researching what influences consumers in their buying habits can be difficult to research, as it often requires interviews, surveys or monitoring by hand (Burke, 2010). However the recent technological advances in BDA offer possibilities to enhance that process for the marketing related aspects of BI (Fan, Lau & Zhao, 2015).

However despite the long history of research on marketing, the literature does not present a clear consensus of a definition on marketing performance (Clark, 2000; Gao, 2010). Gao (2010) states that "a review of the literature has failed to unearth a clear and explicit definition of the term marketing performance" (P. 31). In order to guide further research, Gao (2010) has defined 'marketing performance', based upon 13 previously published scientific papers, as follows: "A multidimensional process that includes the three dimensions of effectiveness, efficiency and adaptability; the effectiveness and efficiency of and organisation's marketing activities with regard to market-related goals, such as revenues, growth, and market share." (P. 35). Thus marketing performance is seen as a process within businesses, and BDA might contribute to this process as it is able to find correlations between various datasets and can create new/faster/more accurate insights that other types of analysis cannot (Fan et al., 2015; Fosso Wamba et al., 2015). However, a search in the literature on "marketing performance", "supermarkets", and "big data analytics", revealed no relevant scientific publications that explore what how BDA can be utilized by supermarkets for creating insights about their marketing performance. Thus, in this context, one can speak of a gap in the literature.

1.2 Relevance

Supermarkets are found in many neighbourhoods, and many consumers visit them on a regular basis (Sonneck & Sören Ott, 2010; Allaway et al., 2011). The market where supermarkets operate in is highly saturated, the competition is increasing and supermarkets face a major challenge in delivering greater value (in terms of experience, service or price) to customers than their competitors (Grewal et al., 2010). At the same time small margins on food products are standard practice in the supermarket industry (Zentes et al., 2007). In 2010 the average profit margin of European Retailers was only 0.7% (Simon, Von der Gathen & Daus, 2010). In order for supermarkets to survive, it is relevant to investigate opportunities on how they can gain better insights about the market they operate in and their customers (Sonneck & Sören Ott, 2010). Firms that are capable of implementing a BDA infrastructure, managing and coordinating the usage of this infrastructure and have personnel with expertise on this subject are proven to be able to enhance their performance (Fosso Wamba et al., 2016). E-commerce businesses have already proven the usability of BDA, as they are able to use BDA to better understand consumers and gain competitive advantages over their competitors (Akter & Fosso Wamba, 2016). The successes online businesses have in using BDA indicates that there are opportunities for traditional retailers to enhance their performance (Weitz & Whitfield, 2010). However, as stated in chapter 1.1, currently there is a gap in the scientific literature when it comes to the utilization of BDA for contributing to the marketing performance of supermarkets.

1.3 Goals

Based on the observed knowledge gap, the goal of this thesis is to create insight into how BDA can contribute to the marketing performance of supermarkets. This leads to the main research question: *How can Big Data analytics be used by supermarkets to enhance their marketing performance?* This will be done by answering the following sub-questions:

1. Which technical and BI conditions do datasets need to meet for usage in BDA?
2. What are the characteristics of supermarkets?
3. Which types of datasets are available for supermarkets to measure marketing performance?
4. Do supermarket datasets (on technical conditions and BI conditions) meet the conditions to be used in Big Data analytics for marketing purposes?

1.4 Outline

The underlying structure of this thesis is illustrated in figure 2. In chapter two the methodology is presented, then for chapters three till six the structure of the model is used. In chapter three, sub-question one is answered by researching the conditions which datasets used in BDA should meet. This is split up into technical and BI conditions, technical conditions referring to which conditions datasets have to meet in order to use them in BDA. BI conditions refer to what one wants to get out of the datasets for management purposes, e.g. to offer better insights for decision making purposes. The contents of this chapter are illustrated by the left part of the model in figure 2.

In chapter four it is researched what characterizes supermarkets and what datasets can be used to determine their marketing performance, thus answering sub question two and three. The found datasets will be characterized using the same technical and BI conditions as before, this will be done in chapter 5.1. This is illustrated by the right part of the model in figure 2.

In chapter 5.2 it will be researched if the supermarket datasets meet the required conditions in order to be used in BDA. This is where both parts of the model come together, hence it is in the middle of the model. This analysis will allow for a conclusion on whether or not the supermarket datasets can be used for BDA, thus answering sub question four.

In chapter 5.3 the main research question "*How can Big Data analytics be used by supermarkets to enhance their marketing performance?*" will be answered. This will be done through illustrating some current and future applications of BDA. In chapter six the findings, validity and reliability of this thesis will be discussed, and this thesis will be focused towards a conclusion.

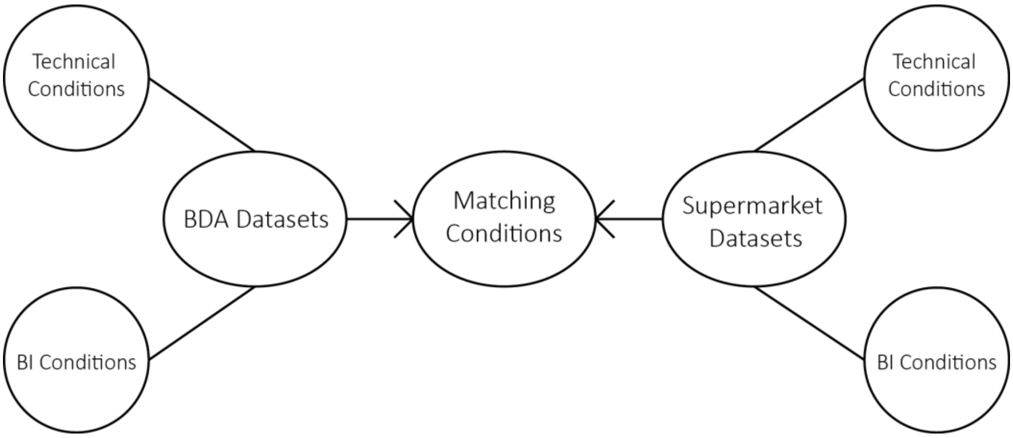

Figure 2: Underlying Structure of Thesis

2 – Method

In this thesis research was conducted by doing a literature review. The databases of the *WUR Library* and *Google Scholar* were used. The database of the WUR Library offers a global search tool, which searches various scientific databases at once (such as Scopus, Web of Science and JSTOR). Therefore mostly articles from peer-reviewed journals and chapters from academic books were used. Articles/chapters were first selected on their title, when the title seemed relevant to the research the abstract and introduction was read. After this it was decided whether they were relevant, if this was the case they would be included in the research. The entire article/chapter was then read and relevant information documented. While reading these articles their relevance to the sub-questions was constantly monitored, if it would turn out that an article was not relevant to the research it would be discarded. For sub-question four extra attention was paid to the generation and processing velocity of the datasets, their structure, origin and purpose.

Also the "snowball method" (i.e. using the references of one article to find new relevant articles (Jalali & Wohlin, 2012)) is used to find literature that could be relevant to this research. This resulted in the use of some grey literature, due to the lack of peer-reviewed alternatives. The articles found using the snowball method were also first selected on their title and then would undergo the same selection criteria as the articles found in the scientific databases. The following search terms were used to find relevant literature:

Sub-question one (*Which technical and BI conditions do datasets need to meet for usage in BDA?*):

- "big data" AND "analytics"
- "big data analytics" AND ("insight" OR "insights") AND "business intelligence"

Sub-question two (*What are the characteristics of supermarkets?):*

- "retail*" AND "industry" AND "characteristics"
- "grocery" AND "retail*"
- "supermarket" AND "retail*"

Sub-question three (*Which types of datasets are available for supermarkets to measure marketing performance?*):

- ("supermarket" OR "supermarkets") AND "performance" AND "measurement"
- "performance" AND "measurement" AND ("retail" OR "retailing")
- "marketing performance" AND "measurement"
- "marketing performance" AND ("supermarket" OR "supermarkets")

Sub-question four (*Do supermarket datasets, with regard to technical conditions and BI conditions, meet the conditions to be used in Big Data analytics for marketing purposes?):*

- "big data analytics" OR "BDA"AND ("retail" OR "retailing")
- "big data analytics" OR "BDA" AND "marketing"
- "big data analytics" OR "BDA" AND "data source*"
- "big data analytics" OR "BDA" AND ("supermarket" OR "supermarkets")

3 – Which technical and BI conditions do datasets need to meet for usage in BDA?

In the business environment Big Data Analytics is a new software tool that allows for the creation of business intelligence (Kimble & Milolidakis, 2015; Chen et al., 2012). To understand how this works, first an understanding of the concept “Big Data” is required, which will be covered in the first section of this chapter. This section will also cover the technical and BI conditions which datasets have to meet for usage in BDA. Then, in 3.2 and 3.3, this chapter will go into detail on what business intelligence and Big Data Analytics are and for what purpose businesses can use it.

3.1 Big Data

Big Data is a collection of “large” datasets from different sources, structured in different ways. Up until a few years ago, it was not possible to link or process these datasets, either due to lack of storage volume, processing power and/or software limitations (Tien, 2013). The usage of the term Big Data in the scientific literature is increasing, however a clear definition among scholars is lacking (Ekbia et al., 2015; Gandomi & Haider, 2015; Fosso Wamba et al., 2015). Fosso Wamba et al. (2015) performed a literature review and defined Big Data upon different definitions from 18 articles: “*a holistic approach to manage, process and analyse 5 Vs (i.e., volume, variety, velocity, veracity and value) in order to create actionable insights for sustained value delivery, measuring performance and establishing competitive advantages”* (P. 235). The five V’s refer to the conditions which datasets have to meet in order to be used in BDA, this will be elaborated in the upcoming paragraphs. This definition represents a common way of looking at Big Data in the management literature (e.g. McAfee & Brynjolfsson, 2012; Hofacker, Malthouse & Sultan, 2016; Erevelles, Fukawa & Swayne, 2016), and will be used in this thesis. In order to answer the sub question “*Which technical and BI conditions do datasets need to meet for usage in BDA?*” it is required to know what Big Data consists of and for what purposes the datasets can be used. Firstly the technical conditions are described using the three V’s: Volume, Variety, and Velocity. Secondly the BI conditions are described using the two V’s: Veracity and Value.

3.2 Technical Conditions

Volume indicates the magnitude of the datasets used in analysing Big Data. Currently datasets with a size larger than one terabyte are considered Big Data, however this is likely to change in the future with the continuous development of technology (Gandomi & Haider, 2015). Also the type of data has an influence on the size of a data set, e.g. millions of transaction details could fit into one terabyte but millions of hours of video would not. Therefore it is difficult to define a threshold of a certain volume for when a dataset can be called Big Data (Gandomi & Haider, 2015). The volume of datasets necessary for supermarkets to make marketing decisions can be considered large, as for proper analytics at least historical data of the last two years needs to be available (Spillecke & Umblijs, 2013). The transactional data alone can be huge for supermarkets, e.g. the retailer Wal-Mart collects 2.5 petabytes (one petabyte = 1000 terabytes) every hour from their customer transactions (McAfee & Brynjolfsson, 2012). This will be further discussed in chapter five.

Variety refers to the different types and sources of datasets that Big Data can contain, and can be classified into structured, semi-structured and unstructured data (Gandomi & Haider, 2015). Structured data refers to data that can be stored in a spreadsheet in an orderly manner (e.g. a balance sheet) (Erevelles et al., 2016; Gandomi & Haider, 2015). Semi-structured datasets are not stored in a tabular manner, but contain tags that allow software to more easily process it (e.g. hashtags on twitter make their data easier to process) (Erevelles et al., 2016; Gandomi & Haider, 2015). Unstructured data can be data from books, texts, photos or videos, these data can lack the structural organisation required by machines for analysis (Gandomi & Haider, 2015). Each type of dataset varies in the difficulty of processing, structured datasets are easier to process than unstructured datasets (Chen et al., 2012). Supermarket datasets for marketing decisions can be of all three types, e.g. transaction data (structured), customer satisfaction surveys (semi-structured), and data from video surveillance cameras (unstructured) (Japec et al., 2015; Marr, 2016). This will be further discussed in chapter five.

Velocity refers to the rate at which datasets are created and the speed at which they need to be processed (Gandomi & Haider, 2015). This can be split into four different categories, namely: batch, near real time, real time, and stream processing (Sagiroglu & Sinanc, 2013). Batch processing refers to periodical processing, e.g. at the end of each business day. Near real time processing happens almost instantly, that is, within a few minutes. An example would be waiting for results to be calculated in a statistics program. Real time processing has to happen instantly and requires processing within a hard deadline, e.g. ATM machines need to process the transaction details instantly in order to serve the customer. Stream processing requires continuous input and output of data, or in other words a 'stream' of data. Analysing the tweets that are currently placed on Twitter is an example of stream processing. (Gandomi & Haider, 2015; Ridge et al., 2015). Faster data analysation requires more processing power, and faces challenges in terms of data availability (Gandomi & Haider, 2015). Supermarket datasets can be generated at all different velocities, e.g. transactional data is generated in real-time (Marr, 2016), but customer surveys come in at batch velocity (Japec et al., 2015). This will be further discussed in chapter five.
In summary, BDA is required when one wants to process various datasets simultaneously which should meet at least one of the following technical conditions: large in volume, structured in different ways, and/or generated/processed at different velocities.

3.3 BI Conditions

For businesses, the goal of analysing Big Data datasets is ultimately to obtain business intelligence (BI) which can be used to aid management decisions (Kimble & Milolidakis, 2015; Chen et al., 2012). BI is a very broad concept, as it covers all relevant information and knowledge that describe the internal and external business environment (Lönnqvist & Pirttimäki, 2006). Nowadays BI is seen as an "umbrella term that includes the applications, infrastructure and tools, and best practices that enable access to and analysis of information to improve and optimize decisions and performance." (Gartner, 2013). BI can "help an enterprise to better understand its business and market and make timely business decisions." (Chen et al., 2012, P. 1166). The new technologies for analysing Big Data offer new opportunities for obtaining BI, as the usage of more and larger datasets allows for faster and more accurate predictions than before (Kimble & Milolidakis, 2015). Therefore the creation of BI is one of the conditions datasets must meet in order to be used in BDA.
The BI conditions that datasets for usage in BDA have to meet can be described using the two V's Veracity and Value. Veracity refers to the quality of the data, or in other words: not all data inside a dataset is necessarily accurate (Erevelles et al., 2016; Gandomi & Haider, 2015). In order for supermarkets to make marketing decisions based on their datasets, they need to assure that the datasets that are available are clean an trustworthy (Spillecke & Umblijs, 2013). Supermarkets can generate a lot of different datasets themselves, but are also reliant on, for example, market share data provided by third parties (Marr, 2016; "Retail Measurement", 2017). Thus this condition requires datasets to be trustworthy in order for managers to make reliable decisions (Fosso Wamba et al., 2015).
The second V from the BI conditions is Value; this refers to the extent to which the processing of datasets can generate insights for managerial decisions (Fosso Wamba et al., 2015). For supporting marketing decision making in supermarkets Big Data could, for example, allow for better customer segmentation in order to have customized actions for each segment (LaValle et al., 2011; McAfee & Brynjolfsson, 2012; Fosso Wamba et al., 2015).

The link between marketing performance and BI

Within the context of this thesis, supermarket datasets need to be able to give insights into marketing performance, in order to meet the BI conditions. However there is little consensus in the literature on which data should be used for measuring marketing performance (Katsikeas et al., 2016; Gao, 2010; Clark, 2000). Katsikeas et al. (2016) recognized that prior researchers have used many different performance measures and that there was little consistency across prior studies. Therefore the aim of their paper was to conceptualise and operationalise the performance outcomes in marketing. Katsikeas et al. (2016) state that marketing is an activity within the value-chain of a firm, this corresponds with the view of Gao (2010), which was presented in chapter 1.1. Figure 3 models marketing performance, and gives an overview of the "Marketing-Performance Outcome Chain" created by Katsikeas et al. (2016).

Literature on marketing within the retail sector supports the view of Katsikeas et al. (2016): Schramm-Klein & Morschett (2006) reviewed several papers and found that the result of successful marketing activities, within a retail perspective, can generate revenue through increasing sales volume and/or customer satisfaction. However, the broader and more nuanced view of Katsikeas et al. (2016) is one of the reasons their model will be used in this thesis. The other being that their model is based on an extensive literature review, as they reviewed 998 articles published in 15 different marketing journals.

In their model, marketing performance consists of four aspects, *Customer Mindset, Customer Behaviour, Customer-Level Performance,* and *Product-Market Performance*. Customer Mindset and Product-Market Performance are most frequently discussed in the articles researched by Katsikeas et al. (2016). Of the articles they researched, 14.6% discussed measures of Customer Mindset and 36.2% discussed measures of Product-Market Performance. Whereas measures of Customer Behaviour and Customer-Level Performance were only discussed in 12.2% and 4.9% of the articles. The types of data used for measuring Customer Mindset and Product-Market Performance lean respectively towards qualitative and quantitative data (Katsikeas et al., 2016), which might make an interesting combination for use in BDA (Venkatesh, Brown & Bala, 2013; Japec et al., 2015). Therefore the scope of this thesis will be limited to these two aspects only. They are defined in the following ways (Katsikeas et al., 2016):

- *Customer Mindset*: Customer perceptions of and attitude towards the firm and its value offering
- *Product-Market Performance*: Performance outcomes achieved (e.g. unit sales, penetration) in the marketplace in which the product is offered

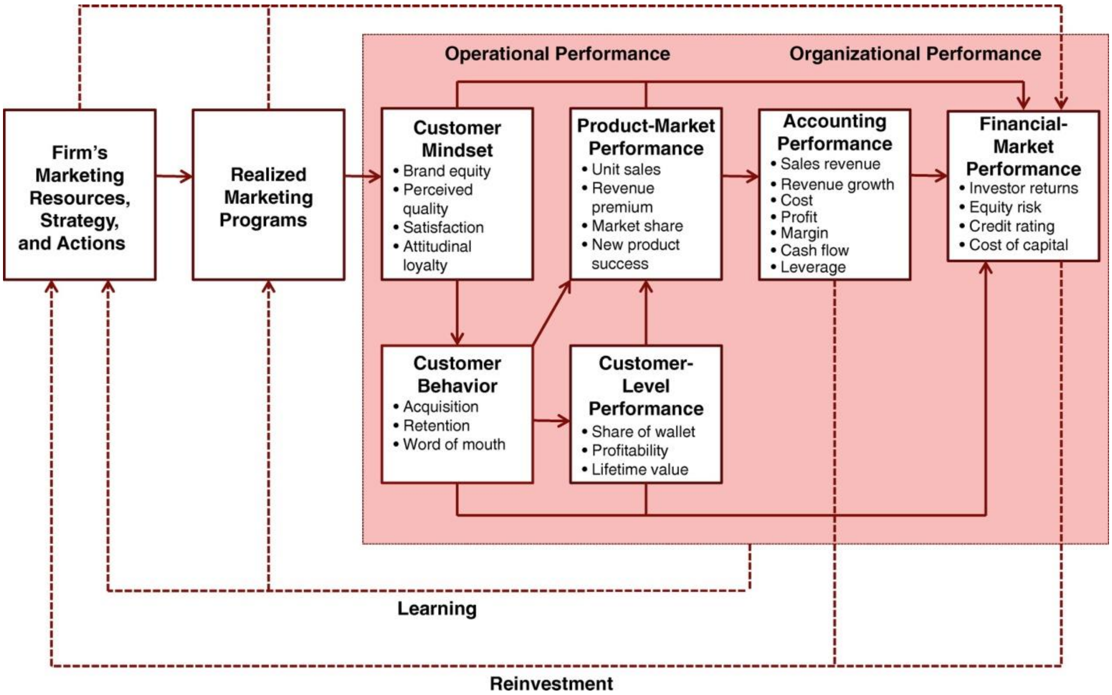

Figure 3: "The Marketing-Performance Outcome Chain"
Source: Katsikeas et al. (2016)

Marketing decisions are to be made within the measures the model presents, e.g. actions designed to improve customer mindset could focus on increasing brand equity, or one of the other measures from this aspect. Thus in order to meet the BI conditions within the context of marketing performance, the datasets should contain information corresponding with the measures of the model. In chapter 4.2 it is investigated which datasets are required to determine the marketing performance in supermarkets, by focussing on these measures.

An important note about Big Data: the datasets of which Big Data consists, can be anything and everything (Marr, 2016). Which makes it difficult to generalize the sources of the datasets (Walker, 2015). The datasets can come from conventional databases that already exist, such as customer databases, ERP systems, and/or customer relationship management systems, or from other, non-traditional, sources, such as audio, video, or social media (Sagiroglu & Sinanc, 2013; McAfee & Brynjolfsson, 2012). The current state of technology allows for the continuous analysis of these datasets from different sizes and sources.

3.4 Big Data Analytics

Regarding Big Data, "*analytics has been defined as complex procedures which run over large scale datasets in order to extract useful information*" (Ridge et al., 2015, P. 690). Thus basically Big Data analytics can be seen as the process of analysing Big Data in order to create BI (Kimble & Milolidakis, 2015). This process is necessary as the raw datasets of which Big Data consists have low value (in terms of money and/or information) if they are unprocessed (Gandomi & Haider, 2015; Tien, 2013). The value of these datasets only becomes clear after the data is processed, analysed, and put into context (LaValle et al. 2011). There are four different forms of analytics, namely: descriptive, diagnostic, predictive, and prescriptive (Ridge et al., 2015; Banerjee et al., 2013). They are defined in the following way (Ridge et al., 2015; Banerjee et al., 2013):

- Descriptive Analytics, describes and reports on the past. Answers the question: *What happened?*
- Diagnostic Analytics uses exploratory data analysis to find the root causes of a problem. Answers the question: *Why did it happen?*
- Predictive Analytics, uses statistical models and empirical methods to create predictions about the future. Answers the question: *What is likely to happen?*
- Prescriptive Analytics, uses mathematical algorithms in order to determine the best course of action. Answers the question: *What should I do about it?*

All four forms are relevant for BDA, however they vary in terms of difficulty. From descriptive analytics being the easiest form to realise, up to prescriptive analytics being the most difficult to design (Ridge et al., 2015). Technological and economic factors limited these forms of BDA, however since a few years businesses have started to utilize BDA (Emrouznejad & Marra, 2016). The possible value of BDA for the marketing performance of supermarkets lies in the new insights and BI it can obtain (Chen et al., 2012; Ridge et al., 2015).
In summary, BDA is a relatively new tool that is able to find correlations between multiple large datasets that other types of analysis cannot, therefore it can help business to acquire business intelligence and remain competitive (Fan et al., 2015; Fosso Wamba et al., 2015).

This chapter answered the sub-question "*Which technical and BI conditions do datasets need to meet for usage in BDA*?". It was found that when the technical an BI conditions are met, the analysis of Big Data can generate insights into what happened, why it happened, what is likely to happen next, and what one should do about this. Ultimately the supermarket datasets that are available for determining marketing performance will have to meet the technical and the management conditions.

4 – What is a supermarket and which datasets are available to measure its marketing performance?

This chapter will start with setting boundaries to the scope of this research, and determine within which context the supermarket datasets should fit. Therefore in the first part it is identified what a supermarket is and how the current market they operate in came to exist. And in the second part of this chapter it will be investigated which datasets are used in the literature to measure marketing performance within the context of supermarkets. The datasets that come forward from the literature are checked on their use in practice in the third part of this chapter.

4.1 The retail industry: *What are the characteristics of supermarkets?d*

Retailers are businesses that connect consumers to the suppliers (Newman & Cullen, 2002). Supermarkets are considered a format within the retailing industry (Zentes et al., 2007). A retail format is characterized by "the strategies that retailers employ in selling goods and services" (Zentes et al., 2007). The supermarket in its current form exists for about 90 years, but retailers have existed for a much longer period of time. The history of the retailing industry can be split up into three eras, namely the mercantile era, the modern era and the digital era (Niemeier, Zocchi & Catena, 2013). The first supermarkets opened up shop during the modern era. The following paragraph describes how Niemeier et al. (2013) characterize these eras.

The mercantile era ranges from the 13th century to the 18th century, and was characterized by regional stores selling food and non-food products from multiple suppliers. The modern era is marked by the start of the industrial revolution, and ranged from the mid-18th century up to the 20th century. It saw the emergence of several formats, such as department stores, supermarkets and category killers (these formats are explained in the follow-up paragraph). A store called "King Kullen", located in Queens, New York, is recognized as the first supermarket by the Smithsonian Institution. It opened its doors in 1930 and was the first to combine large scale assortment with low prices. The modern era saw a move to more centralized distribution systems and retailer-owned distribution centres, allowing stores to be stocked from a distribution centre rather than directly from suppliers. It also made retailers more powerful as a mediator, as they were the only ones who were able to match supply and demand on a large scale and sell at low prices. As the scale of the operations of retailers increased, so did their power as a mediator. This era also saw the emergence of vertical retailing, i.e. retailers owning the production and distribution of goods manufactured exclusively for them.

The start of the digital era was at the turn of this century and saw a change in focus towards an online retail environment. It marks the start of e-commerce and the rise of online retailers such as Amazon. However, next to advances in online shopping, it also saw an increase of internal productivity and retailers specifying their value propositions.

According to Zentes et al. (2007) supermarkets are a format within the food sector of the retailing industry. The food sector can be divided into the following formats: Supermarkets, Superstores, Large retail formats, Convenience stores, Hard discounters, Warehouse clubs and Non-Store formats. Even though the formats are categorized as food they often also sell non-food products. The percentage of non-food products depends on the format. Table 1 gives an overview of the characteristics of the various formats in the food sector.

One format that is not fully covered in the overview of Zentes et al. (2007) is the online retailing format. Online retailing is part of the e-commerce industry, "E-commerce refers to the online transactions: selling goods and services on the internet" (Akter & Fosso Wamba, 2016, P. 182). In the recent years sales made in online stores are increasing, and this trend is likely to continue (Weitz & Whitfield, 2010; Gorczynski & Kooijman 2015). Supermarkets are also starting to integrate online shopping into their activities, however in terms of money the online grocery market is still small but growing (Gorczynski & Kooijman, 2015). Currently both existing supermarkets and newcomers are trying to capitalize the online shopping market (Niemeier et al., 2013).

Table 1: Characteristics of different food sector formats *Source***:** Adapted from Zentes et al. (2007)

Retail Food Sector Formats	
Format	*Characteristics*
Supermarket	Extensive width and depth of assortment (many products and many different brands of a certain product); average product quality; sells manufacturer and store brands; 75-90% food; e.g. Edeka, Rewe
Superstore	Full assortment of conventional supermarket items, plus health and beauty aids and general merchandise; 60-80% food; e.g. Intermarché, Tesco
Large Retail Format	Full selection of supermarket and drugstore items, and general merchandise; extensive width and depth; 60-70% food; e.g. Carrefour, Wal-Mart, Target
Convenience Store	Medium width and low depth of assortment (less products than conventional supermarket, also of the products available less choice of brands is possible), average product quality; 90% food; e.g. Seven-Eleven, petrol station shops
Hard Discounter	Medium width and low depth of assortment; heavy use of store brands; 80-90% food; e.g. Aldi, Lidl
Warehouse Club	Medium width and medium depth of assortment; about 50% food; membership required; focus on business to business trade; e.g. Costco
Non-Store Format	Low width and low depth of assortment; online ordering or use of vending machines; e.g. internet shops, market stands, truck & van sales, vending machines

Supermarkets are characterized by the self-service element, i.e. customers are not dependent on a salesperson in order to make their purchases (Niemeier et al., 2013). Instead of asking an employee for a certain product in order to make a purchase, customers can serve themselves and take a product of the shelves. An individual supermarket store carries about 40,000 unique products (i.e. stock keeping units/SKUs) on average, and has average annual sales of $14 million. However this varies greatly among different regions, e.g. US supermarkets have about 1.5 times more SKUs and sales revenue than European supermarkets (Huddleston et al., 2009; Kumar, 2008). Another characteristic of supermarkets is the frequency and intensity of customer traffic, average customer traffic is about two times per person per week which is considered high in the retail industry (Gómez, McLaughlin & Wittink, 2004). Supermarkets are often organised as "chains", with one "supermarket brand" that is operated in many stores, e.g. Tesco (Kumar, 2008). These chains will be the focus of this thesis.

In summary, supermarkets are a format within the retail food-sector, and in the business of connecting manufacturers to the consumer. Delivering a wide range of products and brands, up to tens of thousands of SKUs. The assortment of supermarkets is focussed towards the food segment (70-90% of a supermarkets assortment consists of food products), but it also includes non-food products. Supermarkets are considered self-service stores, where the consumer has to collect his or her desired goods. They are typically located in neighbourhoods and meet the needs of consumers living in relative close proximity to the store (Sonneck & Sören Ott, 2010; Gorczynski & Kooijman 2015). In the current age supermarkets are able to sell their products through physical stores, online environments or a combination of both. The amount of products and customer traffic, in combination with ever changing trends and customer preferences (Gómez et al., 2004), make supermarkets interesting businesses for the application of BDA in a marketing context.

4.2 Aspects of marketing performance: *Which types of datasets are available for supermarkets to measure marketing performance?*

This section elaborates on marketing performance and how to determine marketing performance in a supermarket context. The first parts of this section, 4.2.1 and 4.2.2, will give an overview on what types of datasets are available for determining marketing performance in a supermarket context. The measures from the aspects *Customer Mindset* and *Product-Market Performance* of the "Marketing-Performance Outcome Chain" will be researched. The datasets that are found to be used in the scientific literature will be checked on their use in practice in chapter 4.3. The found "supermarket datasets" will then be characterised in chapter five.

4.2.1 Datasets for measuring Customer Mindset in the Literature

Within the Marketing-Performance outcome chain the aspect Customer Mindset consists of the measures *Brand Equity, Perceived Quality, Satisfaction,* and *Attitudinal Loyalty*. These will all be covered in the upcoming paragraphs.

Brand Equity

Brand equity in a supermarket environment is the result of the total "brand-building" efforts over time, "which involves the daily implementation of the marketing message through service, product, price, and promotion decisions which are experienced in the stores by consumers" (Allaway et al., 2011, P 202). The measure brand equity covers certain associations to a particular brand that is stored inside the mind of consumers (Keller, 2003). Building a brand is important for supermarkets, as it influences consumer's perceptions of supermarkets and their store of choice (Allaway et al., 2011). Allaway et al. (2011) researched consumers by using questionnaires about different supermarket brands in the USA. Questions about their expectations, experiences and preferences of different supermarket brands were asked. They found that the brand equity of supermarkets is mainly influenced by *service level* and *product quality*.

In the current age brand equity of supermarkets is usually measured using "pen and pencil" instruments, i.e. using surveys (Christodoulides & De Chernatony, 2010; Çifci et al. 2016). These surveys ask questions about brand awareness, trust, social image, etc. However these questions can vary greatly among industry or even products (Christodoulides & De Chernatony, 2010). To specifically measure the brand equity of supermarkets/retailers, questions about service, product quality, brand awareness, brand satisfaction, and loyalty can be asked (Christodoulides & De Chernatony, 2010; Çifci et al. 2016; Pappu & Quester, 2006).

Thus to measure brand equity in supermarkets, datasets are available that contain information from customers visiting different supermarkets combined with their preferences about different supermarket brands. Usually, in the literature, these type of datasets are collected using surveys.

Perceived Quality

Consumers often have imperfect information about products, therefore their perception of a product may differ from reality (Steenkamp, 1986). Their perceived quality is based on these perceptions (Steenkamp, 1986). The perceived quality of supermarkets can be defined as "the perception of quality of the retailer according to the consumer" (Pappu & Quester, 2006, P. 320). A search on the perceived quality of supermarkets reveals that there are two dimensions: perceived product quality and perceived service quality.

There is minimal literature available on the perceived product quality of supermarkets, therefore papers that focused on other retail sectors or specific product categories have been included. In these papers perceived product quality was measured using surveys (Beneke et al., 2013; Das, 2014; Monroe & Dodds, 1988; Rao & Monroe, 1989; Steenkamp, 1986).

A commonly agreed definition of perceived service quality in a retailing environment is: "a global judgement, or attitude, relating to the superiority of the service" (Spreng & Mackoy, 1996, P. 202). Dabholkar, Thorpe and Rentz (1996) developed the "Retail Service Quality Scale" (RSQS), this model

is considered to be a proper tool for measuring service quality in a supermarket environment and is used in multiple papers that researched service quality (Martinelli & Balboni, 2012). In the RSQS data is collected using surveys.
In summary, perceived quality from a supermarket perspective can be split up into two dimensions, perceived product- and perceived service quality. Both are generally measured in the literature using survey datasets. The dataset requires data from customers visiting a certain supermarket's stores, and should consist of customer's experiences with this specific supermarket brand or even specific stores of this brand.

Satisfaction
(Customer)Satisfaction is the third measure of the aspect customer mindset, and can be seen as a customer's emotional reaction to his or her evaluation of the total set of experiences realized from visiting a particular store (Zielke, 2008). Customer satisfaction is a multidimensional construct (Grace & O'Cass, 2005), and therefore can be measured using many different types of datasets.
According to Gómez et al. (2004) customer satisfaction in supermarkets is determined by three factors: Customer service, Quality, and Value. Customer service can be divided into two different attributes: service provided by employees, and speed and accuracy of checkout. Quality can be split up into three different attributes: perceived product quality of perishable departments, availability of everyday grocery items, and store ambiance. Value can be divided into two attributes: value for money and the pricing of products. This measure has some attributes that overlap with the measure perceived quality. A reason for this is that perceived quality is seen as an enabler of customer satisfaction (Seth, Deshmukh & Vrat 2005). In the following paragraph it is elaborated how the aforementioned factors can be measured.

Customer Service
Service provided by employees: Corresponds with the perceived service quality explained in the paragraph "perceived quality", which can be found above. This attribute is measured using the survey as a research instrument.

Speed and accuracy of checkout: Measuring the (average) waiting time of customers in supermarkets is done using an observer, a stopwatch, and logging the waiting time of individual customer (Tom & Lucey, 1997).
The accuracy of the checkout can be measured by physically checking whether the prices on the shelves correspond with the prices at the checkout (Clodfelter, 1998). Studies on the accuracy of checkout scanners generally scanned thousands of products, and checked whether customers were over- or undercharged (Clodfelter, 1998). Thus this attribute of customer satisfaction can be measured using quantitative data on waiting times and the accuracy of checkout scanners at supermarkets.

Quality
Perceived product quality of perishable departments: Improving perceived product quality does not necessary lead to an increase in customer satisfaction, on the contrary a decrease in perceived quality can reduce customer satisfaction (Gómez et al., 2004). This attribute is explained in the paragraph "perceived quality", there it was concluded that the perceived product quality is generally measured using customer surveys.

Availability of everyday grocery items: There are two ways this attribute can influence customer satisfaction. Diminishing or preventing the amount of products that are out-of-stock can increase customer satisfaction (Matsa, 2011). And increasing the amount of different products customers can choose from, can also increase customer satisfaction (Briesch, Chintagun & Fox, 2009). Measuring the amount of out-of-stock items in a supermarket is usually done by hand by a surveyor, who logs each item that is out-of-stock (Matsa, 2011).

Briesch at al. (2009) and Borle et al. (2005) have aimed to measure the preferred amount of grocery items by consumers, they retrieved the required data for measuring this out of so called "market-basket" datasets. Nielsen and IRI are companies that provide these type of data, see Bronnenberg, Kruger & Mela (2008) for an example of a market-basket dataset. A Market-basket dataset gives information on the demographics of consumers, how often they buy, what they buy, how loyal they are, and what their brand preferences are. The brand preferences of consumers are based on the quantities they bought of a certain brand. All data in a market-basket dataset is stored in a tabular manner.
Thus the availability of everyday grocery items can be measured using datasets containing information on the amount of items that are out-of-stock, and market-basket data.

Store ambiance: The store ambiance can be measured using surveys, or through observational research (Burke, 2005). "A survey might ask shoppers to rate the store on specific features, including the breadth and depth of product assortments, the perception of product quality and value, the attractiveness of displays and merchandising, the ease of navigation, the level of shopping convenience, the availability of product information, the frequency of out-of-stocks, the quality of service, and the speed of checkout" (Burke, 2005, P. 217). Observational research is usually done through the use of one or more video cameras, which record the customer's activity. It can give insights into drivers or obstacles for purchase, as what consumers say and do can be two different things (Burke, 2005). Video cameras are not the only option for tracking customers, supermarkets can also make use of data from electronic traffic counters, infrared sensing devices, handheld shopping systems, or RFID tags (Burke, 2005).
Thus store ambiance can be measured using survey or video datasets, which can be combined with datasets generated by one or more tracking devices.

Value
Value for money is the perceived value of products relative to the price (Gómez et al., 2004). Gómez et al. (2004) measured this asking whether customers of particular stores thought they received value for their money. Grace and O'Cass (2005) also used a survey and implemented value for money as "prices are reasonable", "offers good value for money", "good service for price", and "is economical". Thus both used a survey for measuring this attribute of customer satisfaction.

Pricing: several scholars aimed to measure the impact of price on the satisfaction of customers in a supermarket (Gómez et al., 2004; Martínez-Ruiz, Jiménez-Zarco & Izquierdo-Yusta, 2010; Zielke, 2008). Price is a difficult attribute of customer satisfaction, as "the marketing literature does not establish a clear position in relation to the perception that the client has on price" (Martínez-Ruiz, 2010, P. 279). However all of the mentioned scholars made use of surveys to measure the impact of price on customer satisfaction.

Concluding these paragraphs on customer satisfaction: customer satisfaction can be split into three factors, each of which can have an influence on the total satisfaction of a customer. Most of the datasets available for measuring customer satisfaction in supermarkets are survey datasets. However datasets containing information on waiting times, the accuracy of checkout-scanners, market-basket data of households, video data and/or data from other tracking devices are also found in the literature. The datasets require data from customers visiting supermarkets, and their perceptions on the different attributes mentioned above.

Attitudinal Loyalty
Attitudinal loyalty is defined as "a deeply held commitment to rebuy or repatronize a preferred product or service consistently in the future, despite situational influences and marketing efforts having the potential to cause switching behavior" (Oliver, 1997, P. 392). There are two common ways to measure attitudinal loyalty, either using 'brand-specific' or 'individual' measures (Bennett &

Rundle-Thiele, 2002). Both views measure attitudinal loyalty using surveys, which is also the common method from a supermarket/retail perspective (Bennett & Rundle-Thiele, 2002; Jensen, 2011; Liu-Thompkins & Tam, 2013; Yi & Jeon, 2003).
In summary, datasets are found that contain information on consumers' preferences of/ behaviour towards supermarkets that they visit. This is qualitative data that is obtained using surveys.

Available datasets for the aspect Customer Mindset from a supermarket perspective
The measures of the aspect customer mindset are almost all measured using qualitative data, and collected using surveys. See table 2 for an overview of which type of datasets are available to measure customer mindset in supermarkets, this is the first part of the answer to sub-question three "*Which types of datasets are available for supermarkets to measure marketing performance?*". The second part will be answered in chapter 4.3.2.

Table 2: Overview of datasets for customer mindset in supermarkets

Available datasets for measuring Customer Mindset in Supermarkets			
Measures	*Attributes*	*Available type of dataset*	*References*
Brand Equity	Service level, product quality, brand awareness, brand satisfaction, and loyalty	Survey dataset	(Pappu & Quester, 2006; Christodoulides & De Chernatony, 2010; Allaway et al., 2011; Çifci et al. 2016)
Perceived Quality	Perceived product quality, and perceived service quality	Survey dataset	(Steenkamp, 1986; Monroe & Dodds, 1988; Rao & Monroe, 1989; Martinelli & Balboni, 2012; Beneke et al., 2013; Das, 2014)
Satisfaction	Service provided by employees, speed and accuracy of checkout, perceived product quality of perishable departments, availability of everyday grocery items, store ambiance, value for money and the pricing of products	Survey dataset Waiting time dataset Accuracy of checkout-scanners dataset Out-of-stock dataset Market-basket dataset Video data and/or data from other tracking devices	(Clodfelter, 1998; Tom & Lucey, 1997; Gómez et al., 2004; Borle et al., 2005; Burke, 2005; Grace & O'Cass, 2005; Zielke, 2008; Briesch et al., 2009; Martínez-Ruiz et al., 2010; Matsa, 2011)
Attitudinal Loyalty	Purchase intention, brand commitment, tendency to be brand loyal, product category loyalty	Survey dataset	(Bennett & Rundle-Thiele, 2002; Yi and Jeon, 2003; Jensen, 2011; Liu-Thompkins & Tam, 2013)

4.2.2 Datasets for measuring Product-Market Performance in the literature
The aspect Product-Market performance is defined as: performance outcomes achieved in the marketplace in which the product is offered (Katsikeas et al., 2016). It consists of the measures *Unit Sales, Revenue Premium, Market Share,* and *New Product Success*. These will all be discussed in the upcoming paragraphs.

Unit Sales
This measure refers to the amount of products a firm sells or is planning to sell in a particular market (Katsikeas et al., 2016). For supermarkets this would mean the amount of products they sell to consumers (Gedenk, Neslin & Ailawadi, 2010). The stimulation of the amount of units sold is crucial

for supermarkets, as this has a direct impact on revenue (Schramm-Klein & Morschett, 2006). Most of the studies that research the sales of supermarkets use scanner data (Gedenk et al., 2010). "Scanner data are electronic records of transactions that establishments collect as part of the operation of their businesses. The most familiar and now ubiquitous form of scanner data is the scanning of bar codes at checkout lines of retail stores." (Feenstra & Shapiro, 2003, P. 1). The information of each product with a barcode that is scanned at a store will be saved in the cash register system of that specific shop. The data, which is scanner data, can then be transferred to the head office of the store chain (Sammar, Norberg & Tongur, 2012). Scanner data can contain information on quantities sold and other product attributes, such as, brand, packaging size, and product category (Feenstra & Shapiro, 2003; Sammar et al., 2012).
Thus, in order to measure unit sales in supermarkets, scanner datasets from the stores are required. At least information on the quantities sold should be stored inside the dataset in order to measure unit sales. This is quantitative data that can be collected from the stores of the supermarket chain itself.

Revenue Premium
In general this measure refers to "the ability of a brand to charge a higher price than an unbranded equivalent charges" (Ailawadi, Lehman & Neslin, 2003, P. 2). However from a supermarket perspective this is difficult to measure, as there are no unbranded supermarkets (Ailawadi & Keller, 2004). Also some of the most successful retailers today, e.g. Aldi & Wal-Mart, actually charge lower prices than their competitors (Ailawadi & Keller, 2004). Hence, revenue premium is not a valid measure for supermarkets (Rashmi & Dangi, 2016). Ailawadi & Keller (2004) coined the term 'resources premium' as a valid alternative, i.e. the extra resources that consumers are willing to invest in order to shop at a certain retailer. Unfortunately there is no more research done on the topic of resources premium. Since revenue premium is considered an invalid measure and there is no literature on resources premium, it is impossible to determine which type of datasets could be used for this measure concerning supermarkets.

Market Share
Market share is defined as "the percentage of a market (defined in terms of either units or revenue) accounted for by a specific entity" (Farris, Bendle, Pfeifer & Reibstein, 2010, P. 32). It can be calculated by dividing the sales of a firm by the total sales of a market (Cooper & Nakanishi, 1989).
Data used for measuring market share can be obtained from multiple sources, such as: Point of Sale data (POS), Market-Basket data, and changes in inventory data of wholesale warehouses (Cooper & Nakanishi, 1989). It is most common for retailers to acquire market share data through POS sources (Ing & Mitchell, 1994). There are retail analytic services that collect these data and process them for market and industry overviews, e.g. A.C. Nielsen, SAS and GFK are companies that provide those services. A retailer supplies these companies with sales data from their stores, this data is nothing more than a list of Universal Product Codes (UPCs, i.e. barcodes). The dataset consists of the number of units sold and their price for each individual UPC ("Retail Measurement", 2017). This type of data is called sales data (Garber, Goldenberg, Libai & Muller, 2004). Thus measuring the market share of supermarkets is often done by analytics businesses that use datasets containing sales information, these datasets are supplied to them by various supermarkets.

New Product Success
The measure 'new product success' is a multidimensional measure, which is difficult to capture into one definition (Hultink & Robben, 1995). There can be many different reasons for product failure or success, and there is no clear definition on what success exactly is (Hultink & Robben, 1995). For supermarkets the success of products is relevant, as the selection of the right products can improve their success in the market (Montgomery, 1975). The success of new products in supermarkets is generally measured using sales data, i.e. how many units are sold (Garber et al., 2004; Hultink, Thölke & Robben, 1999). However as stated previously, success can also be something different than unit

sales, such as: customer satisfaction with the new product, customer acceptance of the new product and speed to market (Hultink & Robben, 1995).
The stores of a supermarket chain can be seen as their 'products', therefore the brand of a supermarket is connected to its physical stores (Ailawadi & Keller, 2004; Zentes et al., 2007). Thus it is also relevant for them to measure the success of new stores and/or new store designs (Burke, 2010). This is can be done by looking at sales per square meter, store traffic, dwell time and purchase conversion rate. Video datasets can be used to determine the dwell time and traffic through the store, also customer interviews can be used to measure how they like a new store/new store design (Burke, 2010).
In summary, new product success in a supermarket context can be measured using scanner, sales, survey, video, or tracking device datasets.

Measuring the aspect Product-Market Performance from a supermarket perspective
The measures of the aspect Product-Market Performance consist mostly of quantitative data, such as scanner, market-basket, and sales data. Table 3 gives an overview of which type of datasets are required to measure this aspect in supermarkets, and is part two of the answer sub-question answered in this chapter.

Table 3: Overview of available datasets for Product-Market Performance in supermarkets

Available datasets for measuring Product-Market Performance in Supermarkets			
Measures	*Attributes*	*Required type of data*	*References*
Unit Sales	Amount of units sold	Scanner dataset	(Gedenk et al., 2010)
Revenue Premium	Inapplicable to supermarkets	Inapplicable to supermarkets	Inapplicable to supermarkets
Market Share	Sales of the firm, and total sales of the industry	Sales dataset Market-basket dataset	(Ing & Mitchell, 1994; Garber et al., 2004)
New Product Success (for supermarkets defined as new stores or store designs)	Amount of units sold, customer satisfaction, customer acceptance, speed to market, sales per square meter, store traffic, dwell time, and purchase conversion rate	Scanner dataset Sales dataset Survey dataset Video data and/or data from other tracking devices	(Hultink & Robben, 1995; Garber et al., 2004; Burke, 2010)

4.3 Empirical Confirmation of available datasets

Chapters 4.1 and 4.2 set the scope wherein supermarket datasets must fit in order to be measure marketing performance of supermarkets. The datasets found in 4.2 are checked on their use in practice. The datasets that are found to be used by supermarkets will be researched further on whether or not they can be used in BDA. These datasets are called the “supermarket datasets” in the model presented in chapter 1.1. Table 4 gives an overview of the found results. The papers included in this table received datasets from supermarkets in order to conduct their research. Which indicates that supermarkets are able to collect and/or obtain these datasets, hence they are called “supermarket datasets” in this thesis. In chapter 5 it is elaborated for what purposes supermarkets use these datasets.

Table 4: Overview of datasets used in practice by supermarkets

Type of Dataset	Used in practice?	References
Survey	Yes	(Gómez et al., 2004; Hultink & Robben, 1995; Liu-Thompkins & Tam, 2013)
Waiting time	Yes	(Grewal, Roggeveen & Nordfält, 2017; Inman & Nikolova, 2017)
Accuracy of checkout-scanners	No	-
Out-of-stock	Yes	(Inman & Nikolova, 2017)
Market-basket	Yes	(Corstjens & Lal, 2002; Casabayo et al., 2004; Gómez et al., 2004; Gedenk et al., 2010; Lal & Bell; 2003)
Video	Yes	(Burke, 2005; Inman & Nikolova, 2017)
Tracking device(s)	Yes	(Burke, 2005; Inman & Nikolova, 2017)
Scanner	Yes	(Barsky et al., 2003; Casabayo et al., 2004; Gedenk et al., 2010; Inman & Nikolova, 2017; Lal & Bell; 2003; Liu-Thompkins & Tam, 2013; Mulhern & Leone, 1991)
Sales	Yes	(Burke, 2005; Casabayo et al., 2004; Garber et al., 2004; Gómez et al., 2004; Hultink & Robben, 1995; Hultink et al., 1999; Ing & Mitchell, 1994)

Datasets on the accuracy of checkout-scanners were not found to be used by supermarkets. These datasets were used for investigating the fairness of stores towards their customers concerning the prices on the shelves and the prices at the checkout. One might say that this is apparently a subject where supermarkets are less interested in, as no empirical evidence was found that suggested the existence of these datasets.
The use of video datasets is mainly to track the customers’ emotional responses throughout the store, this is done through the use of facial recognition software. However these techniques are still in the start-up phase, and not commonly used (Inman & Nikolova, 2017).

In summary, in this chapter it was presented what characterises a supermarket, and which datasets are used in scientific papers for measuring the different aspects of marketing performance in a supermarket context. These datasets were checked on their use in practice, which resulted in the elimination of one dataset (accuracy of checkout scanners). The other eight datasets were confirmed on their use in practice by supermarkets. These will be further discussed in chapter 5.

5 – Can data that is used for measuring the marketing performance of supermarkets be utilized in Big Data Analytics?

This chapter will start off with characterizing the eight supermarket datasets for measuring marketing performance that were presented in the previous chapter. This is done in order to determine whether the supermarket datasets meet the technical and BI conditions required for use in BDA. Then in 5.2 it is answered whether or not these datasets can be used for BDA. In 5.3 it will be explained how these datasets can be used in BDA, thus answering the main research question.

5.1 Characterizing the marketing performance data of supermarkets

In order to answer sub question four: "*Do supermarket datasets for determining marketing performance meet the conditions to be used in Big Data analytics?*", the supermarket datasets will be characterized using the conditions explained in chapter 3. Table 5 gives an overview of the characteristics of the datasets.

Table 5: Characteristics of marketing performance data of supermarkets

Type of Dataset	Volume	Structure	Data generation velocity	Data processing velocity	Type of Analytics
Survey	Varies	Structured or Semi-structured	Batch	Batch	Descriptive & Diagnostic
Waiting time	Varies	Structured	Stream	Batch, real time, or stream	Descriptive & Predictive
Out-of-stock	Varies	Structured	Batch or stream	Batch or stream	Descriptive
Market-basket	Varies	Structured	Batch	Batch	Descriptive & predictive
Video	Varies	Unstructured	Stream	Batch or stream	Descriptive, diagnostic, predictive
Tracking device(s)	Varies	Structured, Unstructured, or semi-structured	Stream	Batch or stream	Descriptive, predictive
Scanner	Varies	Structured	Stream	Batch or stream	Descriptive & predictive
Sales	Varies	Structured	Stream	Batch or stream	Descriptive & predictive

The table should be read in the following way: for example, survey datasets can vary in their volume, are semi-structured, generated in batches, processed in batches, and can be used for descriptive and diagnostics analytics. Thus each column indicates how the different datasets are characterized.
Each dataset has the value '*Varies*' in the column '*Volume*'. This is because it can't be generalized for all supermarkets. For example survey data that is generated from clients visiting a single specific store would nowadays not be considered large in volume. However if you have thousands of supermarkets under management (e.g. the Dutch retailer SPAR has over 9000 supermarkets in Europe alone (SPAR International, 2017)), the volume would increase drastically and could be considered large. As Gandomi and Haider (2015) already stated, it is "*impractical to define a specific threshold for Big Data volumes.*" (P. 138). Also for BDA volume is no longer considered a real issue or challenge, as storage has become larger, faster, and cheaper (Tien, 2013). Therefore all datasets are characterized as '*Varies*' regarding volume.

Survey datasets are considered structured or semi-structured, as surveys can contain both closed and open questions. When a survey only contains closed multiple choice questions it is considered structured, as it can be easily stored in a tabular manner and the context is clear for each question. If a survey contains open and closed question, or only open questions, it is considered semi-structured. As the questions present context about the content of the answer (like hashtags on Twitter), but the answer itself can be completely different for each respondent (Japec et al., 2015). The found literature states that supermarkets send surveys to customers that participate in their loyalty programs (Gómez et al., 2004; Liu-Thompkins & Tam, 2013). Surveys usually have a specific start and ending dates, and are therefore collected in batches. This is done either through "pen and pencil" tools or online forms (Christodoulides & De Chernatony, 2010; Çifci et al. 2016). Surveys can be processed once the required amount of respondents is reached, or after the collection deadline has passed. This means that they are also processed in batches. These characteristics mean that the technical conditions are met.
The veracity of survey datasets, or in other words the quality of those datasets, is dependent on several factors, errors in survey datasets can arise from "the survey frame deficiencies, the sampling process, interviewing and interviewers, respondents, missing data, and coding, keying, and editing processes" (Biemer, 2010, P. 818). There are frameworks which provide guidelines on collection error-free survey data (Japec et al., 2015). Thus depending on those factors a manager has to decide for himself whether or not a survey dataset has sufficient quality for supporting marketing decisions.
The value of survey data lies in the insights they can create. Surveys conducted by supermarkets usually ask questions to shoppers about past events, such as experienced service (Gómez et al., 2004) or about their preferences (Liu-Thompkins & Tam, 2013). Thus they can give information on what happened, and in case of open questions also why it happened (Japec et al., 2015). Therefore it can be used for descriptive and diagnostic analytics, which can add value to the marketing decision making process. Thus also the BI conditions are met.

Waiting time datasets are structured, as it is a log with the waiting time of each individual customer (Tom & Lucey, 1997). Supermarkets use tracking devices to log the waiting time of customers at the checkout (Inman & Nikolova, 2017), this means the data will be generated in a stream (Kitchin & McArdle, 2016). The processing can happen in batch, real-time or stream form, this depends on the purpose. If the purpose of collecting this data is to be able to immediately react when waiting times reach unacceptable proportions, then real-time or stream processing is required (Inman & Nikolova, 2017; Grewal, Roggeveen & Nordfält, 2017). If the data is merely used to monitor the waiting times, then batch processing will suffice. These characteristics mean that the technical conditions are met.
The quality of this dataset depends on the technology that is in place for logging the waiting times. Managers will have to decide whether or not the data inside the dataset is of high enough quality to use in decision making.
Supermarkets use this dataset descriptive and predictive analytics (Inman & Nikolova, 2017). Analysing it answers how long the waiting times were, and it can provide predictions about how long the waiting time is going to be at certain times of the day (Inman & Nikolova, 2017). Depending on how much value managers attach to controlling or monitoring waiting times, the analysis of this dataset can add insights to his/her decision making process (Grewal et al., 2017). Meaning that the BI conditions are also met.

Out-of-Stock datasets are structured, as it is a list of all SKUs that are out of stock in a particular supermarket (Matsa, 2011). Supermarkets commonly generate this dataset by hand, a surveyor logs each item that is out of stock (Gruen & Corsten, 2007; Inman & Nikolova, 2017). But it is also possible to generate this dataset using POS data (Gruen & Corsten, 2007), or through hardware specifically installed for this purpose on the shelves (Inman & Nikolova, 2017). Therefore this data is generated in batches or as a stream (Gruen & Corsten, 2007; Kitchin & McArdle, 2016). Depending on the purpose of gathering this data, either batch or stream processing is applicable. Batch processing will only allow for evaluation of a certain store cluster, individual store, or department, which is common

in for supermarkets (Gruen & Corsten, 2007). Whereas stream processing will also allow for immediate responses on the operation side. For example, if an item is logged as out-of-stock, supermarkets automatically schedule new items for delivery to the store as soon as possible (Gruen & Corsten, 2007; Inman & Nikolova, 2017). These characteristics indicate that the technical conditions for using this dataset in BDA are met.
The veracity of this dataset is determined by the accuracy of the surveyor or the technology in place. It is again up to managers whether or not the dataset is ought to be reliable enough for marketing decision purposes.
This dataset can be used for descriptive analytics, as it will only indicate what happened and not why it happened, what is likely to happen or what should be done. The value of it lies in possible shorter out-of-stock times, as past trends allow for better stocking in the future (Gruen & Corsten, 2007). Therefore the BI conditions are also met.

Market-basket datasets that supermarkets use contain information on the demographics of consumers, how often they buy, what they buy, how loyal they are, and what their brand preferences are (Corstjens & Lal, 2002; Casabayo et al., 2004; Gómez et al., 2004; Gedenk et al., 2010; Lal & Bell; 2003). The brand preferences of households are based on how often consumers buy a certain brand, and not by asking them what brands they like, i.e. they are based on qualitative data (Bronnenberg et al., 2008). All data in a market-basket dataset is stored in a tabular manner, and is considered to be a structured dataset (Akter & Fosso Wamba, 2016). This data is generated by scanning the products households bought, this means that it is generated as a batch (Borle et al., 2005 & Briesch et al., 2009). This dataset comes from third party companies, such as Nielsen, therefore batch processing is the only option (Bronnenberg et al., 2008). These characteristics indicate that the technical conditions are met for usage in BDA.
The quality of this dataset depends on the methods and processes used to collect the data. This dataset only contains information of past events, i.e. when and what consumers bought, therefore it is suited for descriptive and predictive analytics. Supermarkets use this dataset to gather on what happened, and based on the past purchases of households, predict what is likely to happen (Casabayo et al., 2004; Lal & Bell, 2003). This adds value in terms of creating insights into consumer behaviour, e.g. Casabayo et al. (2004) used market-basket dataset to predict which customers were likely to switch stores. It is up to supermarket managers to determine how much value they believe analysis of this dataset will add. The BI conditions are, however, met.

A Video dataset is considered unstructured data (Gandomi & Haider, 2015). The generation of video data is in a stream, and has to be saved immediately to memory in order to process it (Kitchin & McArdle, 2016). The processing can happen instantly, which falls under stream processing, or it can be processed in batches, this depends on the purpose of collecting the data. Supermarkets use video datasets for tracking customers throughout the store and customers' emotional responses during their shopping trip (Inman & Nikolova, 2017). For observational research, batch processing would suffice (Burke, 2005). However if managers want to be able to instantly react to the incoming video data, then stream processing is required. For example, retailers are attempting to reach customers with personalized offers in real time through the analysis of continuous streams of sensor data (including video) (Inman & Nikolova, 2017). These characteristics indicate that this dataset can be used in BDA.
The veracity of video datasets depends on the quality of the software analysing the dataset (Burke, 2010). Managers have to decide whether or not the quality of the software analysing the video dataset is sufficient enough for usage in marketing decisions.
Video datasets are used for descriptive, diagnostic and predictive analytics. And is able to track what customers did, why they showed certain behaviours, and with this information it is possible to predict what behaviour is likely to be expected (Burke, 2010; Inman & Nikolova, 2017). This can offer several insights for managers to make marketing decisions, hence the BI conditions are met.

Datasets from tracking devices can give information on how many customers visited a store and which isles they visited (Burke, 2005). The data can come structured, unstructured, or semi structured, and is generated as a stream (Hayden, 2013; Kitchin & McArdle, 2016). After the data is generated and stored, it can undergo either stream or batch processing. The reason for this is similar to that of video datasets, supermarkets use tracking devices' datasets for offering real-time promotions to their customers, but also for identifying problems with store layout (Burke, 2010; Inman & Nikolova, 2017). Regarding veracity, the same characteristics apply as explained in the paragraph video datasets. However tracking devices datasets have one more aspect where the veracity might be compromised, e.g. if supermarkets make use of tracking devices implemented in their shopping carts, results are compromised if customers leave their carts to shop in the store aisles itself (Burke, 2010).
Tracking devices' datasets are used for descriptive and predictive analytics. As analysing it can give information on what happened, and what is likely to happen (Inman & Nikolova, 2017. But not on why something happened or what one should do about it (Burke, 2010). The analysis of this dataset can offer managers insights into the behaviour of their customers, which adds value to the marketing decision making process. Hence the BI conditions are met.

Scanner datasets contain information on quantities sold and other product attributes, such as, brand, packaging size, and product category (Feenstra & Shapiro, 2003; Sammar et al., 2012). This is quantitative data which is stored in a table, and therefore structured (Erevelles et al., 2016). The data is generated as a stream, as customers check out constantly during the day, e.g. Wal-Mart generates 2.5 petabytes of data relating one million customer transactions per hour (Kitchin, 2014; Kitchin & McArdle, 2016). The data can be processed in a batch or as stream, depending on the desired use.
Supermarkets use this dataset to monitor the sales of products or product categories (Inman & Nikolova, 2017; REF), for this type of analysis batch processing will suffice. However supermarkets also use this dataset to have live updates about the stock of each store (Inman & Nikolova, 2017; REF), which means stream processing is required. These characteristics indicate that scanner datasets meet the technical conditions for use in BDA.
In the literature scanner data is deemed as accurate and reliable (Klee, 2008), but the veracity of scanner datasets can be compromised by the cashier at the checkout (or customer, if self-checkout is available). If all products that customers take out of the store are scanned, then the veracity is not compromised. Unfortunately this is not always the case in reality, this is something managers should take into account when assessing the veracity of this dataset.
This dataset is used by supermarkets for descriptive and predictive analytics. As, for example, analysing the past 12 months of scanner data can give information on what happened to the sales, and through the use of statistics predict what is likely to happen with them. But not on why something happened or what one should do about it (Spillecke & Umblijk, 2013). This means that this dataset can add value in terms of insights into past and future trends. Thus the BI conditions are also met for usage in BDA.

A *sales dataset* is a list of UPCs, that shows the quantities sold and price for each UPC (Garber et al., 2004). This can be classified under a structured dataset. The data is generated at the Point of Sale, which means the data is generated as a stream (Ing & Mitchell, 1994; Kitchin & McArdle, 2016). The processing happens usually in batches, as it is will be send over to analytic companies at the end of each day (Sammar et al., 2012; "Retail Measurement", 2017). The data of this dataset is generated at the same point as the scanner data. And is basically a stripped down version of the scanner dataset, as it does not include product attributes. Therefore the same characteristics apply as to the scanner dataset, thus indicating that both the technical and BI conditions are met.

The datasets characterized above cannot be used for prescriptive analytics. This is not surprising, as prescriptive analytics requires a combination of several datasets, algorithms and models in order to produce results (Grover & Kar, 2017; Banerjee et al., 2013). However the datasets can all be used in BDA, thus combining them for prescriptive analytics should be possible.

5.2 *Do supermarket datasets for determining marketing performance meet the conditions to be used in Big Data analytics?*

The answer to the question "*Do supermarket datasets, with regard to technical conditions and BI conditions, meet the conditions to be used in Big Data analytics for marketing purposes?*", is yes they do. The presented supermarket datasets all meet the technical and BI conditions in order to be used in BDA. Six out of the eight datasets are structured, which are easier to process than unstructured datasets (Chen et al., 2012). The video and tracking devices datasets are unstructured, and therefore harder to implement in a BDA system but it is possible to combine them with other structured datasets (Chen et al., 2012; Gandomi & Haider, 2015; Japec et al., 2015).
Also the frequency at which the dataset are generated show possibilities for BDA. Nowadays supermarkets are open most days of the week, and average customer traffic is about two times per person per week (Gómez et al., 2004), which means that there are many customers and many product sales that can generate data. To give an indication of the volume of supermarket datasets: the real-time transactional database of Wal-Mart contains 40 petabytes of data, and covers only the most recent weeks (Marr, 2016). Thus the high store traffic and SKUs combined with hundreds or even thousands of stores, means that a lot of data will be generated at a high velocity. The next chapter will go into detail on how BDA can be used in order to enhance the marketing performance of supermarkets.

5.3 Current and Future Practices: *How can Big Data Analytics enhance the marketing performance of supermarkets?*

In this chapter the main research question "*How can Big Data Analytics enhance the marketing performance of supermarkets?"* will be answered. This will be done by illustrating current practices, as well as highlighting what future possibilities BDA has to offer.
The first measure of marketing performance that is covered in this thesis, *brand equity,* can already be measured utilizing BDA instead of survey datasets. This is done through analysing "user generated content", i.e. opinions customers share online and on social media (Verhoef, Kooge & Walk, 2016). This is done through collecting these data, and then through the usage of *text analytics* the negativity or positivity of the words is learned (Verhoef et al., 2016; Erevelles et al., 2016). This process can be done completely automatically, whereas this used to be done manually, and is called opinion or sentiment mining (Grover & Kar, 2017; Verhoef et al., 2016). Survey data is more difficult to obtain than user generated content, as it requires the sampling and participation of respondents which is very labour intensive (Burke, 2010). The measure *perceived quality* is also commonly measured using surveys, but this can also be done using opinion mining. For example, Gensler et al. (2015) used opinion mining and found that one of the associations consumers have with McDonalds is "greasy" (Verhoef et al., 2016). Survey datasets can be used in BDA, but also have added value if they are used complementary to BDA (Japec et al., 2015; Ozgur et al., 2017). This is further elaborated under "findings" in the next chapter.

The measure *customer satisfaction* can be measured using different types of datasets, but surveys were mainly used. Xiang, Schwartz, Gerdes, and Uysal (2015) and Liu et al. (2017) showed that it is possible to use Big Data text analytics on user generated content, in order to determine attributes of satisfaction in the hotel industry. This shows possibilities for analysing and increasing customer's satisfaction of supermarkets. Ramanathan, Subramanian and Parrott (2017) already made some steps with this, as they looked at the impact of social media on customer satisfaction in the UK retail sector. *Attitudinal loyalty* is also usually measured using surveys, and according to the literature BDA should offer opportunities to measure this in other ways (e.g. Hofacker et al., 2016). However using the research methods described in chapter two, the author was not able to find more literature on this topic at the time of writing.

Measuring *unit sales* is done through the use of scanner data, this type of data is already widely spread for the usage in analytics since the 1980s (Chintagunta, Hanssens & Hauser, 2016). Using it for BDA however can offer new opportunities. For example, the Big Data team of Wal-Mart discovered that the declining sales of a particular produce was due to a pricing error. Due to BDA, finding and resolving the issue was a matter of days instead of weeks (Marr, 2016). The same opportunities are possible for the usage of sales data.

Another opportunity for where BDA can contribute to the marketing performance of supermarkets, lies in the utilization of video and tracking device data. As BDA allows for the combination of these datasets other datasets, such as scanner data for example (Burke, 2010). For example, one of the attributes of customer satisfaction was the speed of checkout, in chapter 5 it was indicated that real time analytics could be used in order to reduce checkout times. The American retailer Kroger (operates several store formats, including supermarkets) has developed QueVision (Irisys, 2017). Which “gives retailers insights into how many registers are needed and the expected wait times, using data garnered from infrared sensors over store doors and cash registers, predictive analytics, and real-time data feeds from point-of-sale systems” (Grewal et al., 2017, P. 2). This technology has allowed the average waiting time to drop from over 4 minutes to less than 30 seconds (Grewal et al., 2017). Also new technologies are available that allow for the automatic and real-time counting of out-of-stock items (Inman & Nikolova, 2017), which could also be implemented into a BDA system for supermarkets. Amazon is already using a combination of these real-time tracking devices in their convenience stores. They have developed a technology that automatically detects which shoppers take products from or return to the shelves, these products are then automatically added to their account. They can walk out of the store without going through checkout, and the total amount is charged to their account (Amazon, 2017). This shows possibilities for the supermarket format.

Also for improving new product success BDA offers opportunities, for example, it allows to better predict the success of a movie far before it is released (Xu, Frankwick & Ramirez, 2016). This shows opportunities for supermarkets and the products they stock. There are more opportunities for BDA in supermarkets, such as variety and price optimization, product placement design, labour inputs optimization, distribution and logistics optimization, and web based markets (Manyika et al., 2011). However these are not related to the datasets described in this thesis.

There is no single answer to the question “how can Big Data Analytics improve the marketing performance of supermarkets?”. However the examples given above offer some indication as to how supermarkets can utilize BDA for improving their marketing performance. They can use opinion mining to measure their brand equity and find which attributes are important to satisfy customers, reduce the time it takes to spot and solve problems drastically, and use tracking technologies to shorten the waiting time of customers, track unit sales and out-of-stocks. These new technologies allow supermarkets to use more data sources than before, and BDA is able to combine and process these datasets for improved marketing insights. All in all big data analytics offers plenty of opportunities for supermarkets to better understand and serve their customers.

6 Discussion & Conclusion

Findings

The goal of this thesis is to create insight into how BDA can be used for enhancing the marketing performance of supermarkets. It was found that, for businesses, BDA is a tool that allows for new methods for obtaining business intelligence. These new methods consists of analysing large datasets, that can be structured, unstructured, or semi-structured in nature, and can be generated at high velocity. BDA has become possible in the recent years, as technological advances in storage and processing power have made the fast analysis of different structured datasets possible. For datasets to be used in BDA they have to meet both technical and BI conditions.

It was found that supermarkets are a format of food retailing, and that this format came to light about 90 years ago. Today supermarket stores carry around 40.000 SKUs on average, which consumers can grab of the shelves themselves. Resulting in an average annual sales of $14 million sales per store. Average customer traffic is two times per person, per week, which is considered high in the retailing industry. It was also found that retailers, like Wal-Mart, generate a lot of data in a very short time frame. Thus offering interesting possibilities of BDA.

There is little consensus in the marketing literature about the definition and measurement of marketing performance. Recognising that this was troublesome, Katsikeas et al. (2016) developed a framework to guide future research. From this framework the two most common aspects, *Customer Mindset* and *Product-Market Performance,* were further researched in this thesis. It was found that all but one performance measures from these aspects are relevant for supermarkets. Only the measure revenue premium was found to be inapplicable to supermarkets, due to the fact there exist no unbranded supermarkets.

In the previous chapter it was shown that the two researched aspects of marketing performance, can be measured using BDA. The supermarket datasets that were found to be used in practice all met the required technical and BI conditions. Six out of the eight datasets were considered to be structured, which can easily be incorporated into a BDA system. The literature presented offers insights into how supermarkets can utilize BDA in order to enhance their marketing performance. They can, for example, use tracking devices to streamline the checkout procedure. This reduces waiting times for customers, which increases their satisfaction. Customer satisfaction is one of the measures that has an impact on the overall marketing performance of businesses. Thus, in this example, supermarkets can enhance their marketing performance using BDA. Many measures of marketing performance were found to rely on survey datasets, or other forms of qualitative data. The sampling techniques for these datasets are often time consuming and labour intensive. BDA is able to use other sampling techniques for these measures, and these techniques can be automated. This results in less time needed for problems to be spotted and for them to be solved. Thus time and cost reduction is also an area where BDA can enhance the marketing performance of supermarkets.

The results of this thesis are in line with the general consensus of the literature, which is that BDA offers businesses new opportunities for creating value for customers and enhancing their performance. The results presented add to a better understanding of how BDA can be utilized by supermarkets in order to improve their marketing performance.

One aspect where BDA seems to lack, is in delivering diagnostic analytics. Using a combination of Big Data, statistics, models, and algorithms can result into insights on what happened, what is likely to happen, and even what one should do about it. But answering why certain trends happen or customers behave the way they do seems problematic (Japec et al., 2015). However combining BDA with "traditional" research methods, offers chances to improve research results (Japec et al., 2015). For example, imagine a Big Data team at a large supermarket has access to real-time store traffic and sales volume data. They can spot trends and irregularities by combining these datasets, and conduct in-store surveys which go further into detail about the discovered patterns. Which would help them better understand why certain trends are happening.

Validity

The validity of this research is partly compromised due to the fact that little literature was found for classifying the datasets. Therefore the data generation/processing velocity, and type of analytics sometimes had to be characterized using the author's own insight. However the results are still in accordance with existing bodies of literature, thus they are probably not too compromised.

Also, in hindsight, the scope of this research was probably too wide, as marketing performance in itself is a vague construct. Many more uses for BDA exist, and if this research had focused on just one aspect of marketing performance, or even just one measure, it might had resulted in more detailed results. For instance which datasets supermarkets should combine in order to instantly gain insights about the behaviour of their customers.

On the other hand one could argue that the scope was too limited, as only eight different datasets were found to be used in practice by supermarkets. To give an indication, the Big Data team at Wal-Mart uses 200 different datasets (Marr, 2016). Some examples of the datasets they use are: meteorological data, economic data, telecoms data, social media data, gas prices data, and data of events taking place in the neighbourhood of Wal-Mart stores (Marr, 2016). Both arguments show that there are many more possibilities for supermarkets to use BDA.

Challenges for the usage of Big Data

There are some issues when one makes use of Big Datasets, which might temper the expectations. Firstly, it would be a mistake to think that online data and social media data represent all consumers. It shows only the sentiment of those people actively posting it online, which means that the overall sentiment retrieved using opinion mining does not automatically translate to all consumers. Secondly, the quality of those datasets cannot be assumed, thus data used for analytics would still have to "cleaned" before it can be used in BDA.

Thirdly, correlation does not mean causation. For example, using BDA one could find a perfect correlation between dropping homicide rates and the dropping market share of Microsoft's Internet Explorer in the US. However it is very unlikely that the one causes the other. One could argue that causation is not important, and that only the numbers matter. However then one has the risk that the data is interpreted in the wrong context, or made to fit within the models. An understanding of causal relationships thus remains critical when using BDA.

Fourthly privacy is an issue, consumers are often not aware that the data they generate are used or even shared with third parties. And sometimes they are not even aware that they are being tracked (e.g. Wi-Fi tracking in stores). One could question whether the data collection methods for Big Data can be seen as ethical. Especially when the Internet of Things starts to take off, this will allow for the collection of a lot of privacy sensitive information (e.g. from fridges, or thermostats). Businesses and researchers should be aware that consumers can experience the analysis of these data as an intrusion of their privacy.

Future research

As stated previously, the scope of this thesis was probably too wide. Future research on the marketing performance of supermarkets could focus more on the data inside the datasets, and how these datasets should be combined for optimal results. Another aspect that could be researched, is how the datasets are being generated and structured, in order to find optimal dataset generation techniques for usage in BDA. For example, what the optimal technique would be for logging sensor data from inside the stores in order to use in BDA.

This thesis only focussed on two aspects of marketing performance, investigating the other two could produce results that show other usable datasets for measuring marketing performance in supermarkets.

Is BDA here to last?

The author's view on BDA is that it is, or at least was, a bit overhyped, which does not mean he thinks it adds no value or that the technology is not here to stay. When one reads or talks about Big Data, not necessary in the literature, but more in general, the author has the feeling that sometimes it is spoken of as some "magic box" where one puts a bunch of data in and then spontaneously results will come out. While in reality it is just a tool that is still reliant on the quality of the data and the algorithms that are written to produce results. Or in other words: garbage in still means garbage out. The author thinks that in a few years from now Big Data Analytics will no longer be called 'big', as the storage and volume of datasets is becoming less and less of a problem. And that it will become a normal management tool that many businesses use to support their decision making. But only time will tell what the future holds.

Conclusion

Overall it was found that Big data Analytics is a tool that businesses and scholars are starting to utilize more and more. And that it can prove valuable for many different disciplines and industries. The supermarket sector is one of those industries that can benefit from the usage of BDA, since supermarkets tend to generate a lot of data, at a very high rate. For improving the marketing performance of supermarkets it was shown which datasets could be used in BDA. Examples were given of current and possible future practices for using BDA in supermarkets. This thesis offers new insights into how BDA can be used for improving marketing performance of supermarkets. Previous studies have not zoomed in on the supermarket industry this much. It can be concluded that using the found datasets in BDA offers new possibilities for supermarkets to track, measure, and understand their customers.

References

Ailawadi, K. & Keller, K. (2004). Understanding retail branding: Conceptual insights and research priorities. *Journal of Retailing*, 80(4), 331–342.

Akter, S. & Fosso Wamba, S. (2016). Big data analytics in E-commerce: a systematic review and agenda for future research. *Electronic Markets,* 26(2), 173-194.

Allaway, A., Hiddleston, P., Whipple, J. & Ellinger, A. (2011). Customer-Based brand equity, equity drivers, and customer loyalty in the supermarket industry. *Journal of Product & Brand Management,* 20(3), 190-204.

Amazon (2017). *Amazon Go: Frequently Asked Questions*. Retrieved on July 27, 2017 from https://www.amazon.com/b?node=16008589011

Banerjee, A., Bandyopadhyay, T. & Acharya, P. (2013). Data Analytics: Hyped Up Aspirations or True Potential? *Vikalpa*, 38(4), 1–12.

Barsky, R. B., Bergen, M., Dutta, S., & Levy, D. (2003). What Can the Price Gap between Branded and Private-Label Products Tell Us about Markups? *In R. C. Feenstra & M. D. Shapiro (Eds.), Scanner Data and Price Indexes* (Vol. 1, pp. 165–228). University of Chicago Press.

Beneke, J., Flynn, R., Greig, T. & Mukaiwa, M. (2013). The influence of perceived product quality, relative price and risk on customer value and willingness to buy: A study of private label merchandise. *Journal of Product & Brand Management,* 22(3), 218-228.

Bennett, R. & Rundle-Thiele, S. (2002). A comparison of attitudinal loyalty measurement approaches. *Journal of Brand Management*, 9(3), 193-209.

Berry, L. (2001). The Old Pillars of New Retailing. *Harvard Business Review,* 79(4), 131-137.

Biemer, P. (2010). Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly* ,74(5), 817–848.

Borle, S., Boatwright, P., Kadane, J., Nunes, J. & Shmueli, G. (2005). The Effect of Product Assortment Changes on Customer Retention. *Marketing Science,* 24(4), 616-622.

Briesch, R., Chintagunta, P. & Fox, E. (2009). How Does Assortment Affect Grocery Store Choice? *Journal of Marketing Research,* 46(2), 176-189.

Bronnenberg, B., Kruger, M. & Mela, C. (2008). Database Paper —The IRI Marketing Data Set. *Marketing Science*, 27(4), 745–748.

Burke, R. (2005). Retail Shoppability: A Measure of the World's Best Stores. In *Future Retail Now: 40 of the World's Best Stores*, 206-219. Washington, DC: The Retail Industry Leaders Association.

Burke, R. (2010). The Third Wave of Marketing Intelligence. In Krafft, M. & Mantrala, M. *Retailing in the 21st Century: Current and Future Trends (2)*, 221-238. Berlin: Springer.

Casabayó, M., Agell, N., & Aguado, J. C. (2004). Using AI Techniques in the Grocery Industry: Identifying the Customers Most Likely to Defect. *The International Review of Retail, Distribution and Consumer Research*, 14(3), 295–308.

Chen, H., Chian, R. & Storey, V. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *Management Information Systems Quarterly,* 36(4), 1165-1188.

Chintagunta, P., Hanssens, D. M., & Hauser, J. R. (2016). Editorial—Marketing Science and Big Data. *Marketing Science*, 35(3), 341–342.

Christodoulides, G. & De Chernatony, L. (2010). Consumer Based Brand Equity Conceptualization & Measurement: A Literature Review. *International Journal of Market Research*, 52(1), 43-66.

Çifci, S., Ekinci, Y., Whyatt, G., Japutra, A., Molinillo, S. & Siala, H. (2016). A cross validation of Consumer-Based Brand Equity models: Driving customer equity in retail brands. *Journal of Business Research,* 69, 3740-3747.

Clark, B. (2000). Managerial perceptions of marketing performance: efficiency, adaptability, effectiveness and satisfaction. *Journal of Strategic Marketing*, 8(1), 3-25.

Clodfelter, G. (1998). Pricing accuracy at grocery stores and other retail stores using scanners. *International Journal of Retail and Distribution Management*, 26(11), 412-420.

Cooper, L., & Nakanishi, M. (1989). Market-share analysis: Evaluating competitive marketing effectiveness (1). *Springer Science & Business Media*.

Corstjens, M., & Lal, R. (2000). Building Store Loyalty Through Store Brands. *Journal of Marketing Research*, 37, 281–291.

Dabholkar, P., Thorpe, D. & Rentz, J. (1996). A Measure of Service Quality for Retail Stores: Scale Development and Validation. *Journal of the Academy of Marketing Science*, 24(1), 3-16

Das, G. (2014). Linkages of retailer awareness, retailer association, retailer perceived quality and retailer loyalty with purchase intention: A study of Indian food retail brands. *Journal of Retailing and Consumer Services*, 21 (3), 284-292.

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., & Bowman, T. et al. (2015). Big data, bigger dilemmas: A critical review. *Journal Of The Association For Information Science And Technology*, *66*(8), 1523-1545.

Emrouznejad, A. & Marra, M. (2016). Big Data: Who, What and Where? Social, Cognitive and Journals Map of Big Data Publications with Focus on Optimization. *Studies In Big Data*, 18, 1-16.

Erevelles, S., Fukawa, N. & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897–904.

Fan, S., Lau, R. & Zhao, J. (2015). Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix. *Big Data Research*, 2, 28-32.

Farris, P., Bendle, N., Pfeifer, P., & Reibstein, D. (2010). Marketing Metrics: The Definitive Guide to Measuring Marketing Performance (2). Upper Saddle River, New Jersey: Pearson Education, Inc.

Feenstra, R. & Shapiro, M. (2003). Introduction to “Scanner Data and Price Indexes”. In *Scanner Data and Price Indexes (1),* 1-14.

Gruen, T. W., & Corsten, D. (2007). A Comprehensive Guide To Retail Out-of-Stock Reduction In the Fast-Moving Consumer Goods Industry. Retrieved from https://www.nacds.org/wp-content/uploads/2017/02/Retail-Out-of-Stock-Reduction.pdf

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G. & Gnanzou, D. (2015). How Big Data can make big impact: findings from a systematic review and a longitudinal case study. *International Journal of Production Economics,* 165, 234-246.

Fosso Wamba, S., Gunasekaran, A., Akter, S., Ren, S., Dubey, R., & Childe, S. (2016). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal Of Business Research.*

Gandomi, A. & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal Of Information Management*, *35*(2), 137-144.

Gao, Y. (2010). Measuring marketing performance: a review and a framework. *The Marketing Review*, 10(1), 25–40.

Garber, T., Goldenberg, J., Libai, B. & Muller, E. (2004). From Density to Destiny: Using Spatial Dimension of Sales Data for Early Prediction of New Product Success. *Marketing Science,* 23(3), 419-428.

Gartner (2013). Business intelligence. Gartner IT Glossary, Retrieved from http://www.gartner.com/it-glossary/business-intelligence-bi

Gedenk, K., Neslin, S. & Ailawadi, K. (2010). Sales Promotion. In Krafft, M. & Mantrala, M. *Retailing in the 21st Century: Current and Future Trends (2)*, 393-407. Berlin: Springer

Gómez, M., McLaughlin, E. & Wittink, D. (2004). Customer satisfaction and retail sales performance: an empirical investigation. *Journal of Retailing,* 80, 265-278.

Gorczynski, T. & Kooijman, D. (2015). The real estate effects of e-commerce for supermarkets in the Netherlands*. The International Review of Retail, Distribution and Consumer Research*, 25(4), 379-406.

Grace, D. & O'Cass, A. (2005). An examination of the antecedents of repatronage intentions across different retail store formats. *Journal of Retailing and Consumer Services*, 12, 227–243.

Grewal, D., Krishnan, R., Levy, M. & Munger, J. (2010). Retail Success and Key Drivers. In Krafft, M. & Mantrala, M. *Retailing in the 21st Century: Current and Future Trends (2)*, 15-30. Berlin: Springer

Grewal, D., Roggeveen, A. L., & Nordfält, J. (2017). The Future of Retailing. *Journal of Retailing*, 93(1), 1–6.

Grover, P., & Kar, A. K. (2017). Big Data Analytics: A Review on Theoretical Contributions and Tools Used in Literature. *Global Journal of Flexible Systems Management*, 18(3), 203–229.

Hofacker, C. F., Malthouse, E. C., & Sultan, F. (2016). Big Data and consumer behavior: imminent opportunities. *Journal of Consumer Marketing*, 33(2), 89–97.

Huddleston, P., Whipple, J., Nye Mattick, R., & Jung Lee, S. (2009). Customer satisfaction in food retailing: comparing specialty and conventional grocery stores. *International Journal of Retail & Distribution Management*, 37(1), 63-80.

Hultink, E., & Robben, H. (1995). Measuring New Product Success: The Difference that Time Perspective Makes. *Journal of Product Innovation Management*, 12(5), 392–405.

Hultink, E., Thölke, J. & Robben, H. (1999). Retailers' Adoption Decision of New Consumer Durables. *Journal of Product Innovation Management,* 16(5), 483-490.

Inman, J. J., & Nikolova, H. (2017). Shopper-Facing Retail Technology: A Retailer Adoption Decision Framework Incorporating Shopper Attitudes and Privacy Concerns. *Journal of Retailing*, 93(1), 7–28.

Irisys (2017). *Faster checkout at Kroger creates customer loyalty and increased sales.* Retrieved on July 27 2017 from: http://www.irisys.net/kroger-queuevision

Jalali, S. & Wohlin, C. (2012). Systematic literature studies: database searches vs. backward snowballing. *Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement*, 29-38.

Japec, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., … Usher, A. (2015). Big Data in Survey Research. *Public Opinion Quarterly*, 79(4), 839–880.

Jensen, J. (2011). Consumer loyalty on the grocery product market: an empirical application of Dick and Basu's framework. *Journal of Consumer Marketing*, 28(5), 333-343.

Katsikeas, C., Morgan, N., Leonidou, L. & Hult, T. (2016). Assessing Performance Outcomes in Marketing. *Journal of Marketing,* 80, 1-20.

Keller, K. (2003). *Strategic Brand Management: Building, Measuring, and Managing Brand Equity, (2),* Prentice Hall, Upper Saddle River, NJ.

Kimble, C. & Milolidakis, G. (2015). Big Data and Business Intelligence: Debunking the Myths. *Global Business and Organizational Excellence*, 35(1), 23–34.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 1-12.

Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. Big Data & Society, 3(1), 1-10.

Klee, E. (2008). How people pay: Evidence from grocery store data. *Journal of Monetary Economics*, 55, 526–541.

Kotler, P. & Levy, S. J. (1969). Broadening the concept of marketing. *The Journal of Marketing*, 33(1), 10-15.

Kumar, S. (2008). A study of the supermarket industry and its growing logistics capabilities. *International Journal of Retail & Distribution Management*, 36(3), 192-211.

Lal, R., & Bell, D. E. (2003). The Impact of Frequent Shopper Programs in Grocery Retailing. *Quantitative Marketing and Economics*, 1, 179–202.

Laney, D. (2001, February 6). 3D Data Management: Controlling Data Volume, Velocity, and Variety. Retrieved from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. & Kruschwitz, N. (2011). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review*. 52 (2). 21-31.

Liu-Thompkins, Y. & Tam, L. (2013). Not All Repeat Customers Are the Same: Designing Effective Cross-Selling Promotion on the Basis of Attitudinal Loyalty and Habit. *Journal of Marketing*, 77(5), 21-36.

Liu, Y., Teichert, T., Rossi, M., Li, H., & Hu, F. (2017). Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tourism Management*, 59, 554–563.

Lönnqvist, A. & Pirttimäki. (2006). The measurement of business intelligence. *Information Systems Management,* 23(1), 32-40.

Matsa, D. (2011). Competition and product quality in the supermarket industry. *Quarterly Journal of Economics*, 126(3), 1539–1591.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Hung Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity.* Retrieved from McKinsey Global Institute: http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

Marr, B. (2016). *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results.* (1st ed.). Wiley.

Martinelli, E. & Balboni, B. (2012) Retail service quality as a key activator of grocery store loyalty. *The Service Industries Journal,* 32(14), 2233-2247.

Martínez-Ruiz, M., Jiménez-Zarco, A. & Izquierdo-Yusta, A. (2010) Customer satisfaction's key factors in Spanish grocery stores: Evidence from hypermarkets and supermarkets. *Journal of Retailing and Consumer Services*, 17, 278-285.

McAfee, A. & Brynjolfsson, E. (2012). Big data: The Management Revolution. *Harvard Business Review*, 90(10), 61-67.

Mohanty, H. (2015). Big Data: An Introduction. *Studies In Big Data*, 11, 1-28.

Monroe, K. & Dodds, W. (1988). A Research Program for Establishing the Validity of the Price-Quality Relationship. *Journal of the Academy of Marketing Science*, 16(1), 151-168.

Montgomery, D. (1975). New product distribution: An analysis of supermarket buyer decisions. *Journal of Marketing Research,* (12), 255-264.

Mulhern, F. J., & Leone, R. P. (1991). Implicit Price Bundling of Retail Products: A Multiproduct Approach to Maximizing Store Profitability. *Journal of Marketing*, 55(4), 63–76.

Newman, A. & Cullen, P. (2002). *Retailing: Environment & Operations*. London: Cengage Learning EMEA.

Niemeier, S., Zocchi, A. & Catena, M. (2013). *Reshaping retail: why technology is transforming the industry and how to win in the new consumer driven world.* John Wiley & Sons.

Oliver, R. (1997). *Satisfaction: A Behavioral Perspective on the Consumer*. New York: McGraw-Hill.

Ozgur, C., Colliau, T., Rogers, G., Hughes, Z. & Myer-Tyson, E. (2017). Matlab vs. Python vs. R. *Journal of Data Science,* 15, 355-372.

Pappu, R. & Quester, P. (2006). A consumer-based method for retailer equity measurement: Results of an empirical study. *Journal of Retailing and Consumer Services*, 13, 317-329.

Porter, M. (1985). *Competitive Advantage: Creating and sustaining superior performance*. New York: The Free Press.

Preuss, C. (2014). Retail marketing and sales performance: a definitive guide to optimizing service quality and sales effectiveness. *Springer Science & Business Media.*

Ramanathan, U., Subramanian, N., & Parrott, G. (2017). Role of social media in retail network operations and marketing to enhance customer satisfaction. *International Journal of Operations & Production Management*, 37(1), 105–123.

Ranjan, J. (2008). Business justification with business intelligence. *VINE Journal of Information and Knowledge Management Systems,* 38(4), 461-475.

Rao, A. & Monroe, K. (1989) The Effect of Price, Brand Name, and Store Name on Buyers' Perceptions of Product Quality: An Integrative Review. *Journal of Marketing Research,* 26(3), 351-357.

Rashmi. & Dangi, H. (2016). Act Like a Retailer, Think Like a Brand: An Overview of Retailer Brand Equity and Agenda for Future Research in Indian Context. *Asia-Pacific Journal of Management Research and Innovation*, 12(1), 67–84.

Retail Measurement. (2017). Retrieved 29 June 2017, from http://www.nielsen.com/us/en/solutions/measurement/retail-measurement.html

Ridge, M., Johnston, K., & O'Donovan, B. (2015). The use of big data analytics in the retail industries in South Africa. *African Journal of Business Management*, 9(19), 688–703.

Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. In *2013 International Conference on Collaboration Technologies and Systems (CTS)* (pp. 42–47). IEEE.

Sammar, M., Norberg, A. & Tongur, C. (2012). Scanner Data – A Collection Method for the Future. Retrieved from United Nations Economic Commission for Europe on the 8 June 2017, available at: https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2012/Scanner_data_-_A_Collection_Method_for_the_Future.pdf

Schramm-Klein, H. & Morschett, D. (2006). The Relationship between marketing performance, logistics performance and company performance for retail companies. *The International Review of Retail, Distribution and Consumer Research,* 16(2), 277-296.

Seth, N., Deshmukh, S. & Vrat, P. (2005). Service Quality Models: A review. *International Journal of Quality & Reliability Management,* 22(9), 913-949.

Simon, H., Von der Gathen, A. & Daus, P. (2010). Retail Pricing - Higher Profits Throug Improved Pricing. In Krafft, M. & Mantrala, M. *Retailing in the 21st Century: Current and Future Trends (2)*, 319-336. Berlin: Springer

Sirohi, N., McLaughlin, E. & Wittink, D. (1998). A Model of Consumer Perceptions and Store Loyalty Intentions for a Supermarket Retailer. *Journal of Retailing,* 74(2), 223-245.

Sonneck, P. & Sören Ott, C. (2010). Future Trends in Multi-channel Retailing. In Krafft, M. & Mantrala, M. *Retailing in the 21st Century: Current and Future Trends (2)*, 221-238. Berlin: Springer.

SPAR International (2017) *SPAR International Annual Review 2016*. Retrieved from http://spar-international.com/wp-content/uploads/2017/05/SPAR-International-2016-Annual-Review.pdf

Spillecke, D. & Umblijs, A. (2013). *Making data-driven marketing decisions. In Retail Marketing and Branding: A Definitive Guide to Maximizing ROI* (2nd ed., pp. 10–13). West Sussex, United Kingdom: John Wiley & Sons.

Steenkamp, J. (1986). Perceived Quality of Food Products and its Relationship to Consumer Preferences: Theory and Measurement. *Journal of Food Quality,* 9, 373-386.

Tien*, J. (2013).* Big Data: Unleashing information*. Journal of Systems Science and Systems Engineering, 22(2), 127–151.*

Tom, G., & Lucey, S. (1997). A Field Study Investigating the Effect of Waiting Time on Customer Satisfaction. *The Journal of Psychology*, 131(6), 655–660.

Venkatesh, V., Brown, S. & Hillol, B. (2013). Bridging the Qualitative-Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. *Management Information Systems Quarterly,* 37(1), 21-54.

Verhoef, P., Kooge, E. & Walk, N. (2016). *Creating Value with Big Data Analytics: Making Smarter Marketing Decisions* (1st ed.). Abingdon: Routledge.

Walker, R. (2015). What is “Big Data”? In *From Big Data to Big Profits* (pp. 3–34). Oxford University Press.

Weitz, B. & Witfield, M. (2010). Trends in U.S. retailing. In Krafft, M. & Mantrala, M. *Retailing in the 21st Century: Current and Future Trends (2)*, 15-30. Berlin: Springer

Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, 44, 120–130.

Xu, Z., Frankwick, G. L., & Ramirez, E. (2016). Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research*, 69(5), 1562–1566.

Zeithaml, V. (1988). Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence. *Journal of Marketing,* 52(3), 2-22.

Zentes, J., Morschett, D. & Schramm-Klein, H. (2007). Strategic retail management. Wiesbaden: Gabler.

Zielke, S. (2008). Exploring asymmetric effects in the formation of retail price satisfaction. *Journal of Retailing and Consumer Services*, 15, 335-347.