Saimaa University of Applied Sciences
Faculty of Business Administration, Lappeenranta
Degree Programme in International Business

Daniil Tykheev

Big Data in marketing

Thesis 2018

Abstract

Daniil Tykheev
Big Data in marketing, 57 pages
Saimaa University of Applied Sciences
Business administration, Lappeenranta
International Business
Thesis 2018
Instructor: Principal lecturer Tuuli Mirola, Saimaa University of Applied Sciences

The objective of the study was to look into the concept of big data and find out its application in marketing of clothing retailers. The paper attempted to help in understanding the current state of big data in terms of the technological development and its popularity. The additional purpose was to enlarge the knowledge by reviewing several practical cases. The paper investigated the general aspects of the technology and as well as the more detailed features of big data employment in marketing.

Data for this study was collected solely by the author of the thesis. The information was gathered from literature and the Internet. The goal was to look for the most up-to-date sources, because of the high volatility of the subject. That is the reason why certain non-traditional sources, for example online courses, were used. The report also employs one literature source as the major foundation of the general theory part. The information for case study was collected from the Internet as well. The author of the thesis connected the theoretical concepts from the theoretic part of the paper with the practical cases observed in the empirical part.

The results of the report showed the vast potential of Big Data in marketing and further study is required to fully understand and take advantage of this tool. The paper provided a list of instructions for the competent use of big data as well several recommendations on the actions that need to be performed to start using the technology. The thesis contains a case study that has illustrated the opportunities of big data use in the clothing retail sector.

Keywords:

big data, marketing, analytics, data mining

Table of Contents

1 Introduction

The objective of the report is to find out the current state and capabilities as well as the future potential of big data, to gain a preliminary understanding of how this tool works and what is needed to use it properly and efficiently. First, the paper provides a general overview of big data, discussing its types, storage options and security issues. Then it concentrates on the technical side of the tool, examining the data mining techniques and the most popular open-source software Hadoop.

The report focuses on the use of big data technologies in marketing in general and the marketing of clothing retail companies. It discusses the reasons for big data employment in the field as well as the changes introduced by the emergence of the new concept. The paper investigates the variety of information sources and focuses on the customer data obtained from the social media. After covering the theoretical aspects, the paper provides an overview of the practical measures and suggestions regarding the use of big data.

After collection and analysis of the accessible knowledge from books, online courses and articles, a case study is conducted. It focuses on the use of the technology by the clothing retail companies Lesara and Nordstrom. The two chosen companies have a rather different approach to the application of big data, which highlights the versatility of the tool.

Big data is one of the most trendy concepts in the world nowadays. It has a vast potential and could be applied in many different ways. Some companies start exploring the new technology but there are still many of those who sit on a pile of valuable data without the necessary knowledge to reap benefits from it. Thus, big data needs to be studied in order to use a business' assets to the full potential.

2 Research

2.1 Research questions

The overall purpose of the study is to understand the concept of big data.

The other research questions are:

1. How do companies use big data in their marketing activities? In other words, the goal is to understand what is required for a competent use of data in marketing in technical, human and knowledge aspects.
2. What are the main steps of incorporating big data in the business? No organization could start using big data without planning and preparation for the innovation. The aim is to understand how to carry out the process of incorporating the technology in a business.
3. What are businesses' purposes for the utilization of big data? The opportunities for the application seem endless and their number increases with the development of the technology. The goal is to understand why companies use big data in the particular field of clothing retail marketing.
4. What are the costs of big data usage? Despite big data clearly having lots of advantages, it might seem rather expensive at first. The goal of this paper is to find out about the ways to cut the costs of big data use.
5. What kind of information do customers give to companies via social media? That is a very interesting issue for companies, because social media content, in its majority, is unstructured data, which is rather difficult to collect and process.

2.2 Research methods

The paper employs the case study method for the empirical part of the thesis process. The goal was to gain insight that could help in understanding how clothing retail companies use the big data technologies in their business. The cases reviewed were selected in such a way that they would provide different perspectives on the same subject to produce the conditions for comparison and discussion. The information about the case companies was gathered solely via the Internet. The Discussion and Comparison chapters were included after providing the necessary information about the case companies and supplementing it with related theory from the first part of the paper.

3 Limitations and delimitations

3.1 Limitation

High volatility of big data

Today's technological progress has an extremely fast pace and sometimes the knowledge one gains today may be worthless tomorrow, because a new technology can be discovered and introduced overnight. That is a slight exaggeration, but the truth is that the one who wants to keep up with the industry should always seek new ideas and upgrade their own knowledge.

3.2 Delimitations

1. Technical aspects

Most of the processes concerning data are done on a computer or a network of computers. Therefore, it is important to know and understand how it all works. However, this paper is not profound in terms of programming and computer engineering. Only the most important and popular technologies and utilities such as Hadoop and MapReduce are discussed.

2. Clothing retail companies

Clothing retail companies have been chosen as a sample population because this market is a subject of interest to the author of the thesis.

3. B2C marketing

The number of various possibilities for Big Data employment in B2C marketing is higher than in B2B marketing, which is a subject of interest to the author of the thesis.

4 Key concepts, theories

Big data refers to the dynamic, large and disparate volumes of data being created by people, tools and machines both inside and outside a company. These sets of data are so large and complex that processing requires innovative technology because traditional data management tools cannot perform the task as effective. (Big Data University 2016.)

The main purpose of data science is finding patterns based on data gathered, which will then lead to a decision. Pattern in general is *a regular and intelligible*

form or sequence discernible in the way in which something happens or is done. (Google dictionary n.d.) An example of a pattern can be a seasonal increase and decline in sales of a sunscreen.

Data mining is *a process used by companies to turn raw data into useful information* (Investopedia n.d.). There are many different techniques and approaches to data mining.

According to Maheshwari (2015), database *is a modeled collection of data that is accessible in many ways.* Secondly, data warehouse is *an organized store of data from all over the organization, specially designed to help make management decisions*. Finally, business intelligence (BI) *is an umbrella term that includes a variety of IT applications that are used to analyze an organization's data and communicate the information to relevant users.*

5 Big Data

5.1 Data

John Mashey, a US computer scientist is believed to be the inventor of the term big data, or he is at least responsible for the start of the buzz around the concept in the 1990s (Lohr 2013). However, it is impossible to say when big data actually appeared. As a matter of fact, mankind has been using different sorts of data throughout its history. One of the most remarkable examples is the Bills of Mortality, which were first published in 1603. Bills of Mortality are the accounts of births and deaths which have occurred in a certain district, during a definite space of time. (The free dictionary by Farlex n.d.) Although the initial goal of those lists were just to increase public knowledge, they were used almost 60 years later when the infamous London plague struck. A group of scientists used Bills of Mortality to find out how the disease had been spreading around the city. Then they matched this information with a map of urban water supply and sewage. Eventually, the source and peddlers of the plague were identified thanks to the data and city authorities published instructions on how to avoid being infected. (Cringely 2016.)

This report will mostly focus on the period of evolving of the modern concept of big data, starting from 2005, when O'Reilly Media introduced this term in its current meaning and form (van Rijmenam 2016). This very year is a truly important mark in the development of data, since the cloud technology was introduced alongside the huge reduction in calculation prices. These two factors meant firstly, a new innovative way of storage of gigantic amounts of data and secondly, relatively low costs of that storage. The ongoing process of digitalization of already existing and constant appearance of new information has raised many of new challenges and opened a world of opportunities, one of which being the search for application of this enormous volume of new knowledge. Aside from national security, medicine and military, one of the main purposes here is to get people buy more. (Cringely 2016.) This report will provide a closer look into that subject and find out how big data is changing marketing.

It is impossible to give a common definition of big data as there are plenty of those around the web. The one by Ernst and Young is general and it provides a comprehensive look on the phenomenon claiming that *big data refers to the dynamic, large and disparate volumes of data being created by people, tools and machines. It requires new, innovative, and scalable technology to collect, host and analytically process the vast amount of data gathered in order to derive real-time business insights that relate to consumers, risk, profit, performance, productivity management and enhanced shareholder value*. This definition captures all the essential points of the concept: how immense is the volume of information and necessary tools, and how broad can be the application of the knowledge. (Big Data University 2016.)

Looking further into the definition it is important to mention the 4+1 V's of big data. These are Velocity, Volume, Variety and Veracity. Those 4 V's lead to the fifth V, which every company seeks and wants to add to its customers' experience, and that is Value.

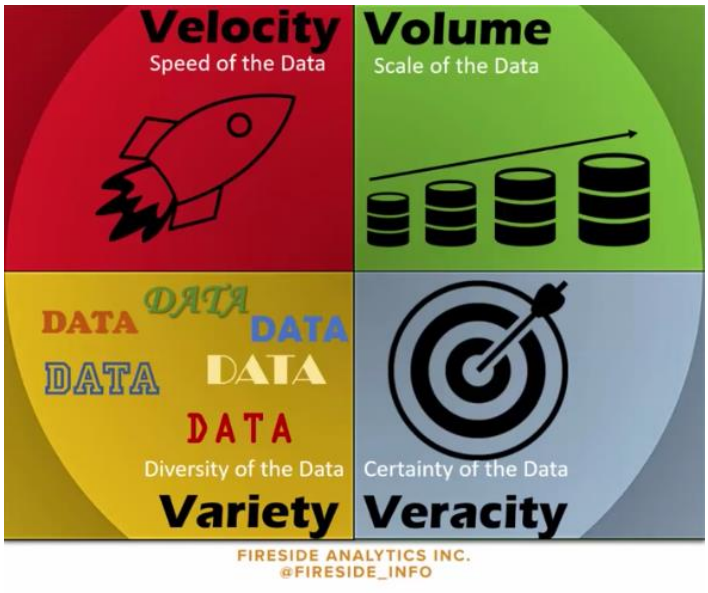

Figure 1. 4 V's of big data (Fireside Analytics Inc. n.d.)

Those four key indicators of big data allow to understand its features. In other words, there is a lot of various information, which grows constantly. Because of the speed, volume and diversity of data it is more difficult to find trustworthy information.

Variety of data is worth looking into in more detail, because it is important to note that there is structured and unstructured data. Nowadays one of the questions for marketers is how to process the unstructured data, such as posts in social networks, pictures and videos. (Big Data University 2016.)

Volume is the amount of data produced by millions of sources. The main challenge posed by the volume of big data is the storage, because the traditional means could no longer be used and new technologies and approaches have to be implemented to effectively manage the excessive amounts of data. (Big Data University 2016.) Databases and data warehouses will be discussed in more detail later on in this paper.

Veracity means trustworthiness of data. With great volume comes great uncertainty – that is the motto that everyone who wants to use big data must keep in mind. According to IBM, poor quality data costs the US economy around $3.1 trillion a year. When the data comes from everywhere, it cannot be blindly trusted, because it can be imprecise and full of biases and abnormalities. Data loses all

its value if the information is at least partially incorrect, because it can lead to wrong decisions. Therefore, those who look for the data must apply a critical approach to the process and always have two questions in mind: how accurate the information is and how reliable the source is. (Altintas n.d.)

The veracity issue appeared together with big data. As the data volumes began to increase due to new sources, such as IoT and social media, the uncertainty started to grow as well, as illustrated in Figure 2. The level of uncertainty now is much higher than it was when the only data companies used was their own. (Altintas n.d.)

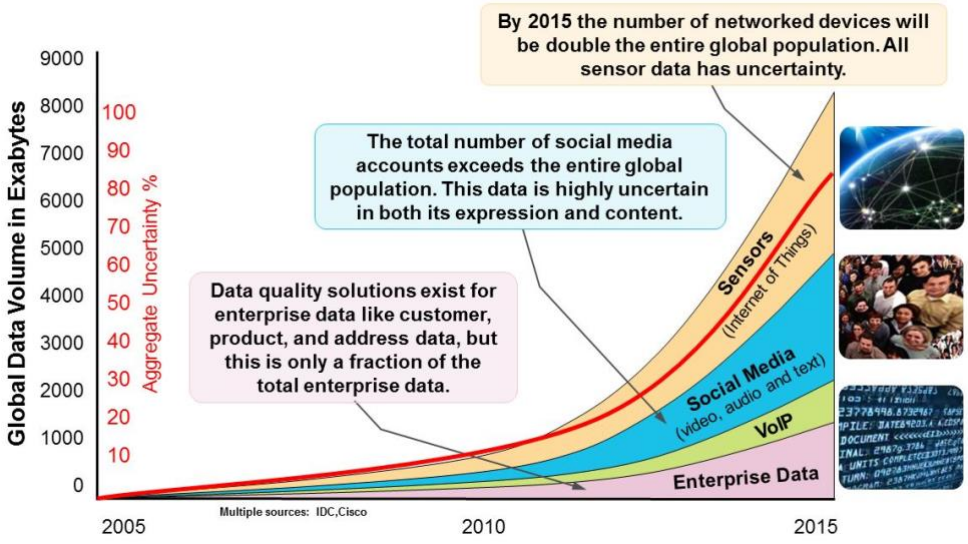

Figure 2. Global Data Volume (Altintas n.d.)

Velocity is the speed with which data is appearing. There are two main reasons for the increased velocity of data: new technologies and amplified amount of information sources. New hard- and software allow enormous data amounts to be generated in seconds and cloud technologies make instant data transfer possible. One of the examples that allows to get a grasp on the data velocity is the amount of pictures being uploaded to Facebook daily. According to Data Center Frontier (2015), in 2015 users were posting more than 900 million photos every day.

High frequency of data uploading brings a logical conclusion: the faster a company gets the data, the faster it needs to make a decision, based on that

data, while it is still fresh and up-to-date. Amazon CEO Jeff Bezos has recently emphasized the importance of quick decision making and suggested that *most decisions should probably be made with somewhere around 70% of the information you wish you had. If you wait for 90%, in most cases, you're probably being slow*. It is risky, but the risks are even higher if one waits too long for more data to analyze. (Dykes 2017.)

5.2 Types of data

Data is very diverse, as it comes from different sources and in different formats. There are many ways of categorizing the types of data, but one of the most fundamental and relevant difference is between structured and unstructured data.

Structured data is the information which is displayed in rows and columns usually in the form of numbers. One of the main characteristics of structured data is that the meaning of each item is defined. (Al Taie 2016.)

Experts suppose that structured data is somewhere from 10% to 20% of the whole amount of data in the world. Hurwitz, Nugent, Halper and Kaufman (2013) distinguish two main categories in which structured data is divided: machine- and human-generated. Both categories are further deconstructed into more particular types. Machine-generated data include:

1. Sensor data. This type of information comes from devices that detect and react to any form of physical impact. Sensors are being introduced throughout different industries. One of the most common uses is implementing sensors in a company's premises, for example warehouses. Sensors can detect motions, which information could be used for different purposes, such as security or inventory control. (Rouse n.d.)
2. Web log data is generated by servers, applications, websites and networks during their activity. This data is used for service-level agreements or to predict security breaches. (Hurwitz et al. 2013.)
3. Point-of-sale data is all the transactional data which is generated every time a purchase is made. (Hurwitz et al. 2013.)
4. Financial data. Even though in the past financial data was mostly produced by humans and some of it is still human-generated, most of the processes

have been automatized and they happen without any human intervention. Hurwitz et al. (2013) give an example of stock-trading, which contains structured data such as the company symbol and the dollar value.

Human-generated data include:

1. Input data. Humans enter different types of data into computers every day and some of this information is structured, for example names, age and e-mail addresses. One of the most useful types of data is qualitative survey data, which can be used to better understand customers. (Hurwitz et al. 2013.)
2. Click-stream data is another type of data which leads to a better understanding of consumer behaviour. Every click a person makes while surfing the web is recorded and used to find patterns. (Hurwitz et al. 2013.)
3. Gaming-related data. The same thing happens with every move customers make in a video game. With the increased popularity of video games, gaming-related data gains large enough volumes to be considered as a separate type. (Hurwitz et al. 2013.)

Structured data can be analyzed and processed using traditional analysis tools and software. For example, Microsoft Excel is a great and relatively simple tool to work with structured data. Millions of rows with numbers and titles can be operated with ease, knowing the right combinations of formulas and functions. Using this type of data, one can answer to such questions as ‘what product has been selling better in any particular month?’ or ‘which salesperson has been selling more additional products than others?’. It is much harder to know for sure why certain things happen or will happen in the future.

The answer to this kind of questions lays in unstructured data. The main difference between the two types is that unstructured data has no “identifiable internal structure”. (Sherpa Software n.d.) It corresponds to 80-90% of all the data in the world at the moment. Unstructured data can also be divided into machine- and human-generated.

Human-generated unstructured data includes:

1. Text files: Word documents, spreadsheets, presentations, email, logs.
2. Email: Email has some internal structure thanks to its metadata, and sometimes scientists refer to it as semi-structured. However, its message field is unstructured and traditional analytics tools cannot parse it.
3. Social Media: data from Facebook, Twitter, LinkedIn.
4. Website: YouTube, Instagram, photo sharing sites.
5. Mobile data: text messages, locations.
6. Communications: chat, IM, phone recordings, collaboration software.
7. Media: MP3, digital photos, audio and video files.
8. Business applications: MS Office documents, productivity applications. (Taylor 2017.)

Machine-generated unstructured data includes:

1. Satellite imagery: Weather data, land forms, military movements.
2. Scientific data: Oil and gas exploration, space exploration, seismic imagery, atmospheric data.
3. Digital surveillance: Surveillance photos and video.
4. Sensor data: Traffic, weather, oceanographic sensors. (Taylor 2017.)

Another big difference between structured and unstructured data is that the latter cannot be analysed via the traditional tools and services. The first obstacle is the volume of data. It is impossible to store such great amounts of data using the same storage systems as for the structured data. It is easy to get a grasp on the difference in sizes: an Excel document with 1000 rows and 8 columns full of information weighs 100 kilobytes, while a single image in JPEG format could easily be around 2-3 megabytes or more. The next problem is posed by the format of data, which is highly uncertain and is changing from one type of unstructured data to another. Excel works perfectly well with numbers, but it cannot process pictures, videos, Facebook profiles and long texts. Sometimes, it is necessary to process all these types of data simultaneously to derive a truly valuable insight.

5.3 Data storage

As it was mentioned before, the amount of data generated every day is enormous, therefore a very logical question comes up: how and where to keep all this data?

Nowadays, there are two main options for data storage: physical solutions and cloud solutions. They are rather similar in terms of appearance. In both cases there is a database of some sort, in which the data is stored. The main difference is the owner of the said database or databases. In the first case, a company owns the data facilities and has them located somewhere at the company's premises. In the second option, the company uses the services of third party organizations that provide the space for data storage. (Pursiainen 2016.)

A database is a modeled collection of data that is accessible in many ways (Maheshwari 2015). The main function of a traditional database is to provide transactional data for day-to-day operations. That means that the database is an OLTP (On-line transaction processing) based system. Most databases employ the relational data model. For example, a hospital's database contains different types of relationships, for instance between patients and doctors. One patient can be enrolled for an appointment to several doctors, as well as one doctor obviously has several patients. That is a many-to-many relationship. However, although one patient may have several appointments, a single appointment can be made by only one patient. That is a one-to-many relationship. One database can operate with different types of relationships. Maheshwari (2015) suggests that there are different types of database software systems, both commercial and open-source. Examples for the first batch are Oracle and Microsoft SQL Server, and the free-of-charge options are MySQL and PostgreSQL.

Databases are not designed for data analysis. It is crucial to have a data warehouse to effectively analyze the data stored in a database. Data warehouse is an additional layer on top of the database or several databases. (Maheshwari 2015.) It uses the data which is located in the related database to facilitate its analysis. This can be clearly seen by taking any company which has different departments: marketing, sales, logistics and human resources, for example. Each of these departments has its own database with all the data from the daily

operations. Several data warehouses could be created in order for each department to have its own warehouse for the analysis of the data. However, firstly it is often not cost-effective and secondly the decision making process usually requires a certain extent of integration between data from different departments. Thus, the data warehouse is created ‘on top’ of several databases.

Data warehouse is an OLAP (Online analytical processing) based system. The Figure 3 illustrates the difference between a database and data warehouse.

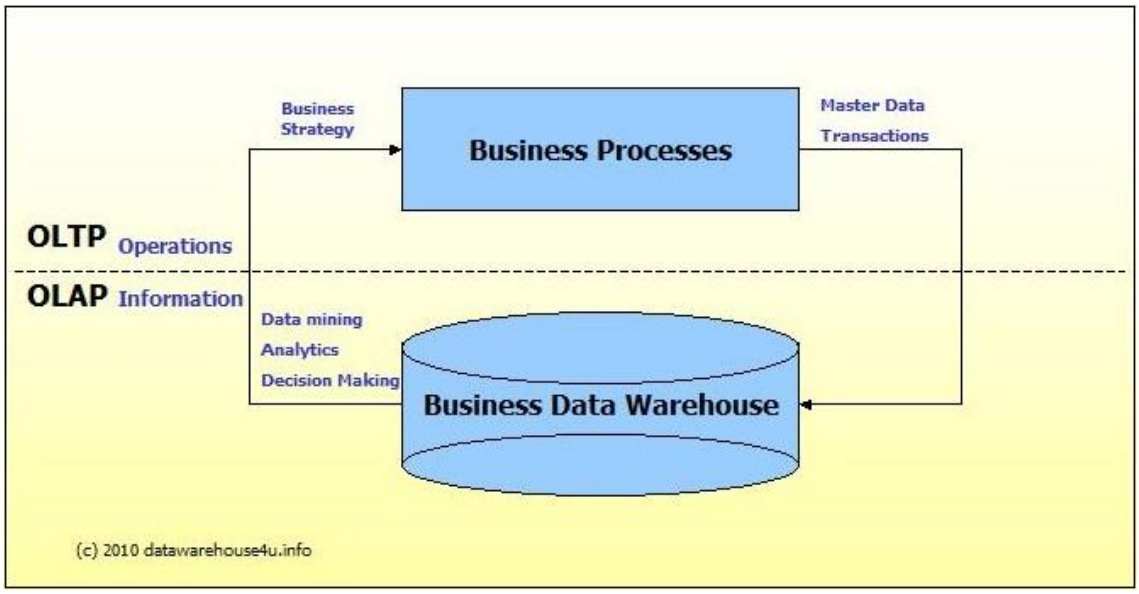

Figure 3. OLTP & OLAP (datawarehouse4u.info 2009)

Typically, the data in a data warehouse is much simpler and there are fewer attributes than in a database. This makes it easier to understand and manage the data. The designer of a data warehouse can regulate the amount of information attached to a piece of data to adjust it to different purposes. It is important to find the right balance of data granularity. The more information a data warehouse has, the more potential insights could be gathered through analysis. However, having too much data is hazardous, as this may lead to wrong conclusions because of too many potential relationships between the items. (Carden n.d.)

Maheshwari (2015) provides a comparison of databases with data warehouses. Table 1 illustrates the differences between the two systems.

Function	Database	Data warehouse
Purpose	Data can be used for many purposes including day-to-day operations	Reporting and analysis
Granularity	High	Low (only the needed features)
Complexity	Highly complex data with many interrelationships	Typically organized around a large fact table and many lookup tables
Size	Database size is moving in both directions: it is expanding when new data is uploaded, but the old data is deleted periodically to reduce size	Data warehouse size is growing with the uploading of new rows of data from the operational database. The old data is saved for future analysis.

Table 1. Comparison of Database and Data Warehouse (Maheshwari 2015)

5.4 Security

Big Data security is one of the permanent issues for every IT department. New threats for data confidentiality are emerging all the time as the value of information is growing exponentially. It is crucial to protect the insights and findings to stay ahead of the competition because otherwise, a company will not have an advantage over its rivals.

The function of big data security is clear enough: keep out the unauthorized users and intrusions with firewalls, strong user authentication, end-user training, and intrusion protection systems (IPS) and intrusion detection systems (IDS) (Taylor 2017).

One of the main challenges is to build an integrated process of data security. There are three main stages: data input, data storage and data output. A company's data has to be properly secured. The process of guarding may differ from one stage to another, as the data could be entered from one or many

sources into one or many databases and the outcome of the data analysis could be sent to one or many addressees. Another challenge is the newness of the analytical tools. Usually data scientists and programmers try to use the most innovative solutions and it is difficult to come up with security measures for the new software. The developers of the solutions usually provide their own security tools, but it is important not only to make sure they work but also to integrate them into the company's big data platform. There are five main technologies for data security: encryption, centralized key management, user access control, intrusion detection and prevention and physical security. (Taylor 2017.)

6 Data Mining

The following chapter focuses on data mining, its types, techniques and tools. Data mining is a very complex discipline that has features of different fields of science, such as statistics, computer science, databases area and business management (Maheshwari 2015).

Data mining is a process used by companies to turn raw data into useful information (Investopedia n.d.). Data mining is all about pattern recognition. The idea is that a pattern derived from a past data can help make better decisions in the future. There are several challenges for data scientists. Patterns have to be new, valid, understandable and most importantly they have to carry a certain value, which can benefit the end user of the information. Large data amounts contain billions of patterns, but only a few of those can actually prove to be useful for the decision making process. The main obstacles include poor data quality, data velocity and data variety. Poor quality data could include dirty data, missing values, inadequate data size, and poor representation in data sampling. Data velocity deals with the relevance of a pattern. New information has to be uploaded instantly to provide the full picture and opportunities to gain the proper insight without any missing parts. Data variety could become a problem if using only one tool for data mining, because data may come from various sources in different formats. Certain patterns can be proved useful only if several sources of data are taken into account. (Hinman 2013.)

6.1 Data mining preparations

Before the pattern recognition stage certain preparations are required to ensure that the quality of the data is appropriate and to provide the best conditions for the final stage of the data mining process. It is very difficult to get the perfect data set straight away, because most of data from public sources is usually incomplete, dirty, inconsistent or simply wrong. For example, one may want to study a store's productivity, but the sales report is missing the names of the salesperson related to sales. This data is incomplete and one has to look for more data to add the necessary information.

Three stages of preparation can be distinguished: cleansing, integration and selection. Sometimes the cleansing and integration stages can be confused, as their ultimate purpose is basically the same – to provide the best data for the mining stage. In the data selection process it is necessary to separate the sufficient amount of data from the large quantities of information. (WideSkills 2015.)

There are various actions involved in the cleansing and integration processes:

1. Removing duplicate data. Different sources may contain the same data and one has to remove the redundant information prior to the data mining process to ensure that the duplicate data will not affect the result of the process. (Maheshwari 2015.)
2. Filling in missing values. Average, modal or default values could be added to the spots where the values are missing. (Maheshwari 2015.)
3. Outlier detection. Outliers are the values which are considerably different from all the data in the distribution. In some cases software do not work properly if any outliers are included into the selection of data. (Körting 2014.)
4. Transforming data in a single format. Data needs to be comparable and a single format is required. The most simple example is currencies. If one has to work with monetary data containing values in dollars, euros and rubles, one has to choose the currency which is more useful to work with and then convert all the others to the format of the selected currency. (Maheshwari 2015.)

5. Ensure data objectivity. Data may contain biases which need to be eliminated. For example, a sample may include a much bigger amount of young people than the population it is derived from. Thus, the results of data analysis with the selected sample will not provide a fair and full picture. (Maheshwari 2015.)

The main process of data mining is very complex from the technical point of view. It deploys such methods as association, classification, prediction, clustering and time series analysis. (Maheshwari 2015.)

6.2 Data mining techniques

This paper reviews in details only a few of the most popular data mining techniques, one of which is the decision tree method. Decision tree is a graphical representation of specific decision situations that are used when complex branching occurs in a structured decision process. A decision tree is a predictive model that uses specific facts to make more generalized conclusions. (Techopedia n.d.) A decision tree has to be connected to a set of business rules to produce the most relevant and appropriate outcome. These business rules are usually presented as a number of if-then statements that show causality. (Maheshwari 2015.) The decision tree method could be best illustrated with the following example.

Using the past data can help predict which person is more likely to buy a certain product, judging by their demographic features. First a table which consists of past results is created (Table 2).

Age	Income level	Gender	Married	Buy?
28-40	Average	Male	False	Yes
28-40	Low	Female	False	Yes
18-27	High	Male	False	No
18-27	High	Male	True	No

41>	High	Male	False	Yes
28-40	Low	Female	True	No
41>	Low	Female	True	Yes
28-40	Average	Female	False	Yes
28-40	Average	Male	True	No
18-27	Average	Male	False	No
18-27	Low	Female	False	Yes
41>	Average	Male	True	Yes
18-27	Average	Female	True	Yes
41>	High	Female	False	Yes

Table 2. Customer data

Then, an actual decision tree is created on the basis of this table (Figure 4).

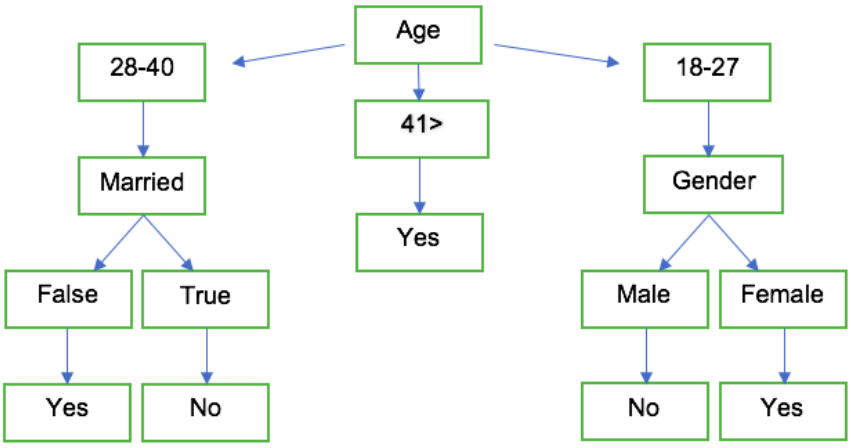

Figure 4. Decision tree

Now it is clear which demographic groups are more likely to buy a product. That will help in identifying a target audience. Usually the amount of data is much bigger than what is presented in the Table 2 above. Thus, a decision tree helps to speed up the decision making process and it also makes it more precise and reliable.

7 Big Data technologies

Processing modern data is impossible without innovative technologies. The problem of data storage and speed of data processing has eventually lead to creating one of the main tools for big data usage – Hadoop.

In the late 90's and early 2000's such companies as Google started to realize that physical storage of data will soon become irrelevant. Since it was impossible to keep all the data on a single physical storage device, it was decided to divide the vast amount of data into small portions and then map them to many devices with general access. That is a very basic explanation of the technology called MapReduce. Hadoop was developed on the base of this technology and introduced in 2005 by a group of creative software engineers of Apache Software Foundation. (Marr n.d.)

Marr (n.d.) distinguishes four modules of Hadoop : Hadoop Distributed File System (HDFS), Hadoop MapReduce, Hadoop YARN, and Hadoop Common. The first two are the most important modules which allow the application to work and perform the most necessary tasks.

HDFS deals with data storage. Data of any format can be stored in the HDFS, which makes the data easily accessible despite the format across a large number of linked storage devices. (Marr n.d.)

Hadoop MapReduce enables parallel processing of the data which is stored in the HDFS. The small portions of the information are directed to mappers that isolate targeted data and reducers collate meaning from it. (Denthumdas 2013.) During the process that module converts the data in a format suitable for analysis (Marr n.d.).

Hadoop YARN manages resources on the hosting storage cluster of devices. Hadoop Common provides the tools (in Java) needed for the user's computer systems to read data stored under the Hadoop file system. (Marr n.d.)

It is important to keep in mind that there are many more processes involved in the Hadoop functioning but the aforementioned four are the most principal of those. Hadoop is mostly used for four different purposes:

1. Search. Hadoop can help provide more relevant and more precise search results which are of greater value to the customer.
2. Flexible Analytics. The information which is uploaded in Hadoop can be very diverse and can have very high accuracy. It allows the users of data to have better flexibility in terms of data analysis. This data could have more than one application over a certain period of time and could be used in different stages and/or projects of data analysis.
3. Point-of-Sale Transaction Analysis. Retailers can use large quantities of recent and historical sales data from PoS systems combined with other data about the customers obtained in stores and online to forecast demand and increase sales.
4. Threat Analysis. Online businesses use Hadoop to detect threat and fraudulent activity. In order to do this, companies capture, store, and analyze content and pattern of messages that flow through the network to tell the difference between a legitimate and a fraudulent transaction. (Denthumdas 2013.)

Denthumdas (2013) also suggests that Hadoop has four main features:

1. High availability. The whole system has a great load balance thanks to the MapReduce technology. It means that the tasks are performed independently and failure of one part of the program does not affect the whole process. Hadoop also restarts failed processes automatically.
2. Scalability of storage/compute. Once again, thanks to MapReduce, no matter how many nodes are involved in the process, applications do not have to re-architect the system.

3. Controlling cost. One does not have to know the required storage or processing time beforehand. Usually the process starts with the minimum amount of nodes, which can then be multiplied, if necessary, based on evolution of storage. That means that a company will not lose money because of inaccurate preliminary calculations or unexpected progress of a process.
4. Agility and innovation. Hadoop is a very flexible software which keeps data in its original format. This makes it easy to apply new techniques of data analysis.

Hadoop has been attracting extensive attention and it is used by many corporate giants, such as Adobe, EBay, Facebook and LinkedIn. Adobe use the technology in *several areas from social services to structured data storage and processing for internal use*. LinkedIn use Apache Hadoop alongside some other technologies for their People You May Know feature and other similar models on their web platform (Hadoop Wiki n.d.).

8 Big Data in Marketing

Considering the first steps of big data in marketing, it is impossible not to mention Jeff Bezos, the CEO of Amazon and his contribution to the field. He is basically one of the fathers of a popular marketing tool which is known as contextual marketing. His goal was to have a non-stop interaction with an individual customer. For example, his intention was to welcome a customer back on Amazon with such a phrase as “are you still looking for a pan/shirt/book?”. That concept of remembering the customer’s last step (and later all the others) was a huge leap forward for big data development. (Cringely 2016.)

Big data has been introducing new channels of getting information from customers and markets, analyzing and using it in forms and scope which were not possible before.

8.1 Reasons for use

The main reason for companies to use big data is to facilitate and/or improve the decision making process. As Maheshwari (2015) suggests, there are two main

types of decisions: strategic and operational. Wise use of data can have a positive impact on both of the types.

Strategic decisions are difficult to make for different reasons. Sometimes a company's purposes and the direction of changes are not clear, and the outcome of the decisions may be noticed only after a certain period of time. The use of past data helps to reduce the doubts by taking into account similar cases and analyzing many potential scenarios. New ideas can also be generated through pattern recognition. All four main categories of marketing strategic decisions (product, price, place and promotion) can be facilitated by using both internal and external data. For example, a clothing company can plan the time of its winter collection promotional campaign launch based on the effectiveness of previous similar campaigns, climate conditions and predicted demand for the product.

Operational decision mostly uses past data. Maheshwari (2015) claims that a classification model can be created and modelled using the data of past instances to develop a good model of the domain. This model will then be used to improve and often automatize operational decisions in the future. One of the daily tasks of a marketing department is the evaluation of prospects. Those prospects have certain parameters and features, which can help group and distinguish said prospects. These patterns can be later used to predict the probability of successful interaction and thus increase the effectiveness of lead generation.

According to SAS (n.d.), there are three key areas in which the competent use of big data and its combination with a wise marketing strategy could make the most significant impact on a company's business:

1. Customer engagement. Big data brings the opportunity to considerably enlarge the base of knowledge about a customer. The amount of data that companies were able to collect mere ten years ago has been multiplied several times. The main concern of the past was to find out who are the people buying a product and who may be turned into customers. Nowadays, one can get access to information that will lead to knowledge of customers' location, tastes or preferences in terms of when and where to be contacted, for example.

2. Customer retention and loyalty. Big data also comes in handy when the relationship with a customer has already been established. The more interaction with customers a company has, the more information it can gather and the better it can understand what impacts customers' decisions.
3. Marketing optimization and performance. Continuous testing and analysis can help not only in cutting the costs of marketing operations, by calibrating the spend between channels, but also to optimize the marketing programs.

8.2 Changes brought by Big Data

The world of today has been significantly transformed with the evolvement of big data already and new applications of that tool are being introduced daily. Marketing is changing as well and nowadays companies have a much broader arsenal to understand customer behaviour and provide better targeting. In this paper the period of 2010 and later is compared to the previous years.

The first development that is impossible to miss is the difference in the volume of data. The Worldwide Web allows Japanese companies to successfully sell in Argentina, studying the local market from the other side of the world. Another example could be found in every big online clothing store. While browsing for new products a customer can always set specific requirements for the search results they want to get. These could be size, price range, color or brand. All this information is stored and analyzed to understand what exactly the customer was looking for and what are their shopping habits, patterns and expectations. On the basis of this information, the customer will later get a personalized promotion. (Franks 2014.)

The next aspect which exists in close connection with the volume of data is its accessibility. Once again with the help of the Internet companies are able to get different sorts of information almost in no time. For example government archives, academic publications and customers' feedback can easily be found online. Thus more and more information can contribute to the decision making process.

Another element is customer' feedback, which is produced mainly by social media. There is a large amount of unstructured data ready to be analyzed. Blogposts, tweets, Instagram photos, YouTube videos and billions of comments can contain useful information about brands, products, customer service, anything and everything. Basically, there is direct feedback and everything else. Direct feedback involves posts which mention brand or product names and express a person's attitude towards their customer experience. Companies have never been able to get this close to that many customers by simply opening a computer. One of the main benefits that companies get is an opportunity to understand their customers better and then to create personalized ads and present them to the customers.

Another part of marketing that has been transformed by big data has to deal with planning, estimations and expectations from a marketing campaign and the product itself. Now it is much easier to forecast results of certain marketing actions. (Reynolds 2016.) One of the most famous examples of wise preparatory work is the Netflix show "House of Cards". It is a common belief that Netflix executives were sure that the show is going to be successful a while before they started filming the first episode. They have achieved that as follows: Netflix had been collecting data and using already existing one to predict what its customer were going to like. They paid the closest attention to so-called "user actions", for instance how many times had a user re-watched his favorite comedy special, which categories was he searching for and which he was flipping through, or what actor or director had been involved in a customer's favorite shows and movies. There were three main facts, which Netflix took into account. Firstly, a significant part of their audience liked David Fincher's movies. Secondly, Kevin Spacey was one of the favorites of the same people who liked Fincher. Thirdly, the British version of House of Cards had proved to be successful. Those three factors brought Netflix to the decision on buying the series, spending $100 million for the first two seasons. What is also noticeable here is that Fincher directed only the first two episodes, leading the show into the right direction, while being more cost-effective. (Carr 2013.)

That is a perfect example of predicting customer behavior and forecasting the success of a future product. However, some were dissatisfied with such a "cold"

and pragmatic approach that Netflix used. For example, Andrew Leonard (2013) claimed that Netflix has turned its customers into puppets and killed creativity. An issue that should be discussed is how much privacy is left with big data being everywhere and whether it is a healthy practice when companies know any customer's next step. This report does not cover a discussion on that topic and leaves it to the related authorities to define the legislative frameworks for big data usage. However, it is necessary to note that most of the common people are seemingly not preoccupied with data being collected by companies, as long as their activity is not interfering customers' personal space. Furthermore, some see the introduction of big data as a virtue. They claim that both customers and companies may benefit from the data algorithms, for instance, used in online shopping. Thus a company can better identify the need and a customer gets a better and wider choice.

There is one indisputably positive change that was introduced by big data which is bringing small businesses closer to the top tier companies. Even though budgets are different, small enterprises still have equal access to zettabytes of information. Therefore there is an increased competition and customers benefit from it as always. (Reynolds 2016.) The main reason is the introduction of many free-of-charge software, such as Google Analytics or Hadoop. Now, a company does not have to pay for a license every month to continue using the product, and the only investment that is required is the one in knowledge.

Many can have mixed feelings about thinning privacy and rapidly growing information of different sorts. However, from the business point of view it can be clearly seen that there is an unprecedented opportunity that was brought about by big data, and it is important to try to seize that opportunity as fast and as effectively as possible.

8.3 New challenges

Every innovation poses certain challenges to those who are eager to use this innovation in their favour and big data is no different. This report discusses several of these challenges, which are mainly related to marketing. The following list has no particular order.

1. How to select the right information (veracity of information)

Big data refers to an extremely large amount of information. Marketers wonder how to separate all the junk, irrelevant and unreliable information from the piece which can bring value. Sometimes it can be similar to looking for a needle in a haystack. Therefore companies should carefully select appropriate data sources. (Spiegel 2014.)

2. Fast pace of changes

Big data is that kind of an innovation which cannot be put in a framework and studied in class for years. It requires constant adjustment and learning from those who want to reap the biggest benefit. As new methods of data processing are being introduced periodically, and new sources and types of information can be discovered daily, one just cannot keep up with everything. That is why many companies use services of independent data scientists for certain projects. However, since using big data can bring benefits on the daily basis it may be wise to have a big data expert on constant payroll. This thought brings up the next challenge

3. Lack of experts

Since big data is a comprehensive term, it is needless to say that one person cannot manage it alone. Therefore, usually a team of data scientists, analysts, developers and managers is required. Big data is a relatively new concept as well, which means there is not enough properly educated and skilful experts. The next problem is that even if a company has managed to gather suitable people and made a team, all of them need to possess specific knowledge of the company's strategy and sphere of business, for example. There are two main obstacles here: need of education and security of confidential information. (Spiegel 2014; Qubole 2017.) Obviously some of the big data aspects are easy to comprehend, but if the goal is to reach maximum efficiency, an organization will need the best people, who are harder to find.

4. Integration and connectivity throughout the company

When introducing big data to the employees, company executives need to find new ways of integration between departments, since the proper use of big data requires different sorts of corporate information. While selecting the needed information one needs to know about the product design, pricing strategy, marketing, manufacturing and storage capabilities. To sum up, a great deal of different data can impact the decisions on big data usage.

5. How to get value from unstructured data

Unstructured data, as mentioned above, is the sort of data which is not organized in any pre-defined way or does not fit into any data model. Such data usually involves plentiful text information, which can contain any sort of information also with features of structured data, such as dates, numbers and facts. Examples of unstructured data are academic publications, documents, blogposts, videos, photos and audio samples. Since that information cannot be put into any framework, it is much more difficult to use any automatic methods of data mining and analysis. Obviously, there is software which can perform such tasks, but manual tagging is still involved. (Sherpa Software n.d.)

6. Accessibility

Big data is an enormous amount of information, some parts of which can be covered and hidden from a common man by such features as firewalls or subscription fees. Therefore one needs to be sure (or at least try to do so) that a certain closed source contains the desired information before trying accessing it.

7. Security

The security topic has already been slightly touched upon while discussing some of the previous challenges. Companies are anxious about the security of confidential information, which keeps them from using their data in its full potential. This anxiety is undoubtedly reasonable, but one has to take risks if he wants to use all of the data he possesses. (Spiegel 2014; Qubole 2017.)

8. Software

Hadoop is one of the most widely used software solutions for big data processing, and it is indeed a very convenient tool for data mining and analysis. However, because of the aspects such as the newness of the concept and lack of talent, it is hard to master it. In the survey by Qubole, 73% of the respondents claimed to understand the Big Data platform was the most significant challenge of a big data project (Qubole 2017.)

8.4 Information sources

Nowadays businesses should take into account the fact that everything people do leaves a digital trace. Andreasen (2013) has designed a graphic that perfectly illustrates what information sources are included in big data and explains all of them individually.

Figure 5. Intelligence by variety. (Andreasen 2013.)

As Figure 5 shows, there are nine main data sources with different levels of velocity, variety, volume, structure and accessibility. Some of them are quite new, including those various new devices (sensor data) that are being introduced to the market every now and then and that are able to provide new kinds of information. For example, during the last five years there has been the rise of wearable technologies. Smart-watches like Apple Watch or Samsung Gear are the trailblazers, which will lead the way for smart-glasses, shoes and clothes, for example. All those devices will collect and transmit unique information, which was impossible to get in such a quantity and with such a quality only a decade ago. Alongside some physical indicators like pulse or walking pace, even more interesting data can be gathered, for instance where and when people use the device, what are they using it for, what are the most performed tasks and so on. This sort of data is used to predict customers' behavior not only in store or online but in real life. and make decisions on the essential questions: when, where and how to reach customers. (Andreasen 2013.)

Another relatively new source of information is the social media. Here are some statistics to consider: 79% of online adults use Facebook (Greenwood, Perrin, Duggan 2016), and 38% of the millennial age group have made a post about a brand. (Ypulse 2016). The first fact gives a perspective on how many people are active online, meaning they all provide information of different sorts, which can be used to analyze and predict their behavior. The second statistics means that roughly one out of four millennials has made a post about a certain brand. That kind of posts are a combination of indirect feedback and a peer's opinion, which is known to be one of the most important information sources for a customer. That sort of feedback may contain more valuable information than a traditional one, because customers tend to be more honest and straight-forward online. By organizing such posts using certain keywords via the big data software, a company can provide better customer service and get a more detailed insight about customers.

8.5 Customer data from social media

Social media is undoubtedly one of the major sources of data and interaction platforms between companies and customers. The growing popularity of social networks is a huge advantage for marketers, as they have never before been able to get that close to the end user of their products and services. Social networks are also rich in unstructured data – one of the most valuable data source nowadays. York (2016) suggests to consider social media as a raw source of data. The information one gets from social networks is later used for the analytics to derive value.

Every social network has its own metrics. It is vital to understand the differences between the networks and these metrics, because the audience's behaviour can vary depending on the platform. Today with the increasing importance of data analysis, major social networks provide native tools for such purposes (York 2016.)

Facebook's most essential metrics, according to York (2016), are:

1. Engagement shows the total amount of post clicks, likes, comments and shares on a time scale of one week.
2. Impressions are the number of times when the Facebook page is displayed.
3. Likes:
   • Organic likes are the number of likes from users that do not come from the paid advertisements.
   • Paid likes are the number of likes from users that come from paid advertisements.
   • Page likes are the total number of likes.
4. Unlikes are the number of people who unliked the page.
5. Post reach is the total number of people who have seen any content or advertisement of the page.

Instagram's most essential metrics, according to Jackson (2017), include:

1. Comments received. Jackson argues that Instagram comments are more important than likes, because a comment means a much higher level of engagement from a user. It takes much more time to come up with a comment than to leave a like on a picture. It shows that a customer received a certain emotional impulse from the brand's content.
2. Most engaged hashtags. Hashtags work as keywords for the suggestions mechanism. It is important to know which ones work in accordance with the goals of a business.
3. Engagements per follower. Businesses should not solely concentrate on the size of their audience. This metric shows how many likes and comments per follower there are on the page.
4. Followers gained. The overall amount of followers a page attracted over certain period of time.
5. Referral traffic. This metric shows the relation between an Instagram account and a website, i.e. how many people clicked through to the website of a company after visiting the Instagram page.
6. Instagram stories metrics. This metric shows how popular the stories of a page are.

An important issue for a business is to remember is that this information is just raw pieces of data that need to be analyzed to be relevant in the decision making process. Combining the patterns derived from this data with the actual comments and messages from customers allows companies to have a better understanding of their audience and adjust their marketing activities.

9 How to use Big Data

9.1 How to start using Big Data

The stage of preparation before taking any action is vital in terms of big data use. It is easy to be overwhelmed by the vast amount of data if starting using it without careful planning and strategy. Businesses must first realize their goal of implementing solutions, meaning what kind of real problems will they solve using big data.

Nowadays big data is used in many different areas from healthcare to military operations. Maheshwari (2013, 354) suggests that in business, regardless of the industry, the main application of big data is the protection and enhancement of customer relationships and customer experience. Therefore, a company has to identify what parts of their interaction with customers could be improved and try to answer the most important questions and solve customers' problems with data solutions.

Glass & Callahan (2014) emphasize that by better understanding current customers, a company will be able to better understand its prospects. They also quote Ruth P. Stevens, CEO of eMarketing Strategy, who claims that *it's about analyzing the buying process of your target audience. The first step is to figure out what it is that you need to measure, and in order to do that you need to understand what are your goals. Are you going to measure based on leads generated? Are you going to measure based on sales or revenue?* Thus, a company needs to have a goal and then carefully choose the means by which this goal could be achieved.

Understanding what precise sections of a business could be improved leads to another important step, which is understanding what precise parts of data should be looked at. The problem that many companies come across, especially in the beginning of their big data journey, is the already mentioned staggering volume of accessible data. It is difficult to act when it is unclear where to start. Therefore Glass & Callahan (2014) suggest to "start small". They claim that the process of incorporating big data in the daily routine of a company is not an easy one. Thus, one has to start with a small scale of data use and gradually evolve. That kind of approach has an advantage in terms of adaptation: it is easier to integrate a new technology starting with small steps, rather than a full-on campaign. Besides, it is not clear for everybody that big data is worth investing in and some small wins with low costs can convince the management board to invest more actively.

The next step big data amateurs should take after determining their current issues is to look for those companies who have already succeeded in deploying big data. These examples could be use as cases in the learning process and as benchmarks in the future.

Efficient use of all the data a company possesses is impossible without a competent team of specialists. There are several ways of solving this issue. First of all, an organization may try to use the existing human resources, enhancing their capabilities by training and educating the personnel. Specialists from both business and IT sides of the company are required. (Pramanick 2013.) As big data technologies are in the state of constant development, a mindset of permanent learning has to be adapted to stay relevant and use the most innovative tools. A second option is to hire specialists from the outside. There is a growing awareness of the importance of data analysis skills, and more professionals are shifting their preference towards learning more about big data technologies. A third option is the complete or partial outsourcing of the data analysis. It may prove wise to get a strategic big data implementation partner, for example IBM or Oracle, which will help the company to implement the big data practices into company's operations and facilitate the process of changing and adapting corporate activities. (Pramanick 2013.) This method is especially useful for smaller businesses which do not have enough funds and/or have no necessity of a full-scale data analysis department. Using a third-party provider of such services might significantly lower the costs of big data use. Together with the team and/or the implementation partner a company is to decide on the infrastructure, tools and architecture for the big data department.

Maheshwari (2013, 354) suggests that in the preliminary testing phase, an organization should deploy the company's internal data sources. The point is that it is much easier to understand all the big data capabilities when one is using the data which one knows is reliable and has full control of.

Finally, the company as a whole must evolve to use all the opportunities that data provide. The level of interaction between departments inside the company must be as high as possible. Thus, it is important to invest in devices and software that facilitate communication inside the organization. (Delgado 2017.) Besides Maheshwari (2013, 355) suggests that, even though data storage is not infinite, employees should not delete data, because even if it cannot be used to derive immediate profit, it can add perspective and bring unexpected value in future projects. Past data is also used in machine learning.

9.2 How to use Big Data

After a company has gotten comfortable with the new routine of daily data analysis, it is important to remember and follow several guidelines to continue a successful big data journey.

1. People and data should be together. Maheshwari (2013, 354) claims that the most valuable insight is gathered when combining the best features of machine-based data analysis and human intuition and experience.
2. The team has to be carefully assembled. It comes as no surprise that using big data requires a very competent team with excellent technical skills. Even when applying big data solutions in marketing, a company will need data scientists, architects, mathematicians and technologists. However, it is important to find a balance between the analytical and creative people. The latter come up with ideas of how to use the knowledge obtained by using the technology. (Glass & Callahan 2014.) The demand for senior data scientists is growing and according to Glassdoor, in 2016 data science was the highest paid field to get into. That is why it may pose a challenge for a company to find a professional without spending half of its salary budget on one person. (Valchanov 2017.)
3. Greater data diversity is more important than its mere amount. Thus more perspectives appear and better quality insights could be produced. (Maheshwari 2015.)
4. Data value depreciates with time. The speed of data analysis is of utmost importance. Sometimes, if the information has not been analyzed in 15 minutes, the insight loses part of its value. (Maheshwari 2015.)
5. One copy of data is enough. Storing many copies of one dataset has a couple of drawbacks. Most obviously, the storage capacity is not infinite. Secondly, when dealing with large amounts of data, its duplicates might create confusion. (Maheshwari 2015.)
6. Exponential growth should be planned. Experts predict data quantities to continue growing with storage and applications getting more accessible. Soon, every progressive company will have to use big data technology in certain extent. (Maheshwari 2015.)

7. Flexible processing environment. With the perpetually changing environment, companies need to establish a resilient and efficient system, which will adapt to these changes as quickly as possible. (Maheshwari 2015.)
8. Cooperation is the key. It is evident that CMO and CIO should work together to achieve the best results using the big data technology. However, other departments, such as sales or human resources also need to be integrated into the process. Thus, the sales department could provide valuable statistics, which should be taken into account while analyzing different data. Vice versa, based on the knowledge derived from the data, sales specialists could adjust their approach to customers, for example by offering certain products to a specific group of people. The human resource department should know the requirements of the job that involves active use of big data to hire the right people. (Maheshwari 2015.)

9.3 Big data use in marketing

Big data generally brings the same improvements to marketers as it does to many other professionals of different fields. The quality of decision making along with the detailing of processes significantly increase with a competent application of the technology.

Basically, the main advantage that big data introduces is the enhanced knowledge of customers. While in the past marketers were making decisions based on intuition and experience, nowadays their guesses could be confirmed by using the data. Besides, the number of customers' characteristics has increased as well. In addition to the standard demographic information, such as age, gender and marital status, companies are now able to literally track customers' online behavior. (Spencer 2014.)

The different sorts of data need to be added to a customer's profile. This data could be gathered, for example, via cookie files. Cookies collect information about a customer's activity on a certain website. The website owners have access to the cookie files, created by their site. A clothing retailer's website cookie file could contain such information as customers' clothing and shoe size, his preferred colors, styles, brands and his estimated income level, based on the price range

choice. This data is gathered while the customer is searching for a new product to purchase, and the more data is collected, the more accurate decision marketers could make. Cookies are also the primary method of collecting data from social media. The owner of the social media platform can sell the first-party cookies to companies or place the third-party cookies directly in the platform. (All AboutCookies.org n.d.)

All the features mentioned above allow marketers to create personalized promotions for individual customers. In the past, a shoe company would design a single advertisement for all the 20-year-old female university students living in Helsinki. Now, each of these students would get an individual ad, because one of them might prefer white Adidas sneakers, like the majority of the members of this group, while the minority would like black high Dr. Martens boots. Not a single customer would be left unnoticed, thanks to big data, and therefore the effectiveness of promotional campaign will increase.

Continuing the story of Amazon, the pioneers of personalized advertisement, a relevant example from a customer point of view will be discussed. Amanda Zantal-Wiener (2017), a Marketing Blog staff writer, shared her experience with Amazon, while discussing the brands which use personalized marketing in the best way possible. She describes herself as a person with a ‘borderline obsession with hip hop’ and attaches a screenshot of the Amazon main page, offering her different products related to her interest (Figure 6). She then notes that personalized ads also help companies provoke unplanned purchases.

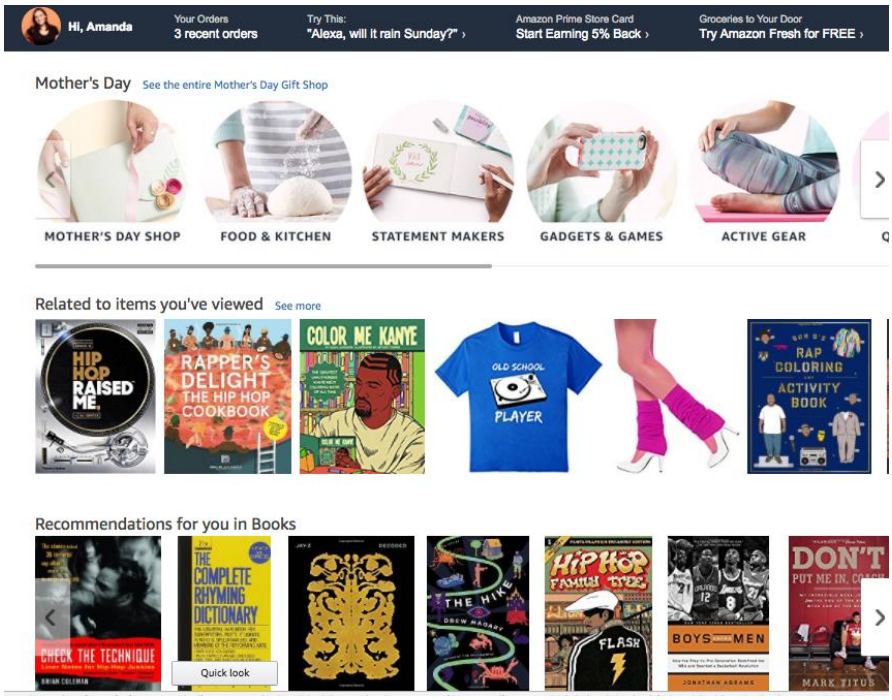

Figure 6. Amazon main page. (Zantal-Wiener 2017)

A similar example could be found very near. The author of the paper himself was recently looking for a certain model of Nike's black sneakers. After that he is constantly receiving adverts on websites that allow contextual ads, containing several similar products:

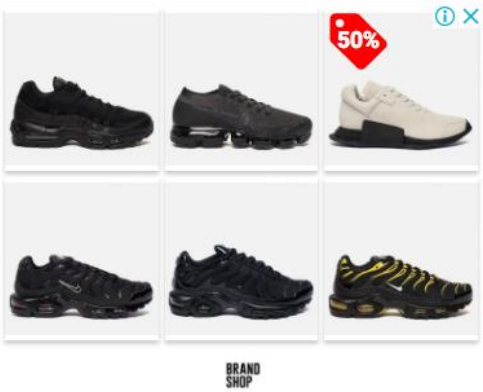

Figure 7. Contextual advertisement (Sports.ru 2017)

Another substantial improvement presented by big data is the broad abilities of analyzing the success of advertising campaigns. Again, the level of details is impressive. Apart from the native analyzing tools, that all popular social media

platforms have, there is a content scoring method that allows marketers to measure the effect of a single post. (Murdock 2014.) Thus, companies could improve and adjust their strategies in real time when some actions prove efficient and others do not.

Big data is not only a virtue for those concerned with online marketing. It helps with the most common and basic decisions as well, for example pricing. Companies are used to determining the price considering the cost of the product, competitors' prices, and value of the product to the customer, for example. Sometimes, when the business is not going well, marketers adapt the seemingly easiest way to boost sales – a 10% discount. (Baker, Kiewell, Winkler 2014.)

Baker et al. (2014) suggest that with big data capabilities it becomes possible to use many more factors to make a better decision. Those could be data from individual deals, decision-escalation points, incentives and performance scoring data, for instance. They stress the importance of approaching any price decision as individual, especially in the B2B sector, as circumstances may vary from one deal to another.

Baker et al. (2014) point out two main guidelines to remember when using big data for setting the price:

1. Listening to data. Companies usually have plentiful unused data at their disposal already, so such a common process as price setting is not a complex task in terms of data searching. The challenge is to derive the most valuable insight from the information that a company has in its database. The often-overlooked data include customer preferences, sales reps negotiations information, and general economic situation.
2. Automatization. Apart from saving time, automated price setting based on data helps in recognizing unfamiliar patterns.

According to Ric Kostick, a board member of the cosmetic company 100% Pure, they managed to increase their revenue by 13.52% after employing a machine learning algorithm. It can determine the amount of discount for a specific group of customers, based on their interest in a product. (Tanner 2014.)

9.4 Costs of Big Data

The costs of using the big data technology are usually divided into three categories. These are hardware, software and human resources. The big data solutions provider Cooldata (2014) claim in their blog that building an analytical database is one of the two major investments as 1TB per month will require roughly $45 000. However, Mark Johnson (2016) of Fusion Alliance suggests that this cost could be considerably cut down if using an open-source analytical tool like Hadoop. Comparing the traditional data warehousing provider Oracle to the open-source software, Johnson (2016) states that with Hadoop the cost of 1TB will fall down to the quite affordable $400. Another solution is the use of cloud services for data storage. Amazon, Microsoft and Google offer cloud data storage for the price of $355 to $413 per month (Bauer 2017). Cooldata (2014) also suggest that the cost could be reduced by purchasing an end-to end-solution, which will not only be cheaper than building the whole system from scratch but will also facilitate the integration process. Big data outsourcing is the best solution for small businesses, which are not capable of building a complex infrastructure by themselves but still want to deploy the technology.

The costliest investment according to Cooldata (2014) would be the team of specialists. A web source claims that at least eight professionals are required: ETL developer ($85k/year), cloud infrastructure engineer ($84.5k/year), Python developer ($92k/year), database administrator ($71.5k/year), data analyst ($65k/year), dashboard developer ($80k/year) and at least two data engineers ($91k/year). Thus, a small big data team costs $660 000 per year. (Glassdoor 2013; Glassdoor 2018a-d; Payscale 2018a-b.)

All in all, for most of the companies which see big data as an auxiliary tool it is an obviously wise decision to use a third-party provider of the service, because it would significantly cut the costs. The setbacks are privacy issues, partial control over operations and data access. That is why larger corporations opt for creating their own data department.

10 Future of Big Data

Big Data is already in use, but according to most of the experts, it is only on the way to its prime. Here are several developments that are most likely to happen in the nearest future which will affect big data and everyone who wants to use it.

1. The volume of data will increase even more. Alongside volume, the amount of sources from which this data will be derived will increase as well.
2. Software for big data will improve and more diverse. Free-of-charge, easily accessible and adjustable software like Hadoop makes a huge contribution to the progress of big data technology.
3. The privacy issue will become more important. As it was mentioned already, more companies and common people start to worry about the privacy of their data. Thus more obstacles for big data development are arising.
4. Big Data volume and complexity requires extreme speed of analyzing, that is why new technologies such as machine learning will become more necessary. (Marr 2016.)

11 Case study

11.1 Lesara

11.1.1 Company description

Berlin-based company Lesara, founded in 2013, is a leading player in the European agile fashion market. In 2017, they raised $40m in funding for strategic development and expansion. (Finsmes 2017.) The majority of the company's achievements were reached with the help of big data.

Agile fashion is a new concept that focuses on predicting current and future demand, quantities needed and style trends via big data analytical tools, for example the most popular open-source software Hadoop (Mitchell 2016).

It is important to mention that Lesara's recipe for success involves not only the competent use of big data; the company has exceptionally effective

manufacturing and distribution systems, which allow them to substantially cut costs. Another decision that let Lesara save a large amount of money was to abandon the traditional brick and mortar stores, in favor of online selling. (Mitchell 2016.)

11.1.2 Purpose of big data use

However, quick distribution and flexible production cycles would be pointless without big data. It allows Lesara to make accurate predictions of fashion trends and therefore customers' future demands. That is a significant turnaround in the approach to fashion. Such large and successful retailers as H&M and Inditex are still mostly oriented on trends, set by world-famous designers, and revealed at fashion shows twice a year. With smart analytics, such companies as Lesara are always several steps ahead of traditional fashion retailers. (Mitchell 2016.)

11.1.3. Lesara's benefit from using big data

Roman Kirsch, the CEO of Lesara, describes the advantage his company has over competitors: *By relying on data, we are able to offer the customers the exact trends they seek and want to buy, utilizing a much more personalized shopping experience than any retailer relying on traditional, subjective trend forecasting would ever be able to.* (Mitchell 2016.) Trend forecasting could be done in many ways, one of the most common ones being Google Trends – a free tool that analyzes different aspects of popularity in a certain search term. (Matias 2012.) Google Trends is a rather simple tool, but its data could be combined with information from other sources, contributing to the final decision.

Kirsch also predicts that in ten years technology will rule the fashion market and the trends are not going to be created inside designers' workshops, but in tech companies' databases (Mitchell 2016).

11.1.4 Three problems Lesara solves via big data

Lesara primarily uses big data for three things: to minimize the excessive production, to take mass production on a whole new level and to decrease the time between the start of design and the sale (Costa 2016).

Usually fashion retailers produce more items than they need, because of the fear of not satisfying demand. That leads to many products being transferred to discount retailers, which results in loss of money. Big data allows companies to anticipate the demand much more precisely, by analyzing previous seasons and combining the results with current customer data. Thus, there is minimum to no unsold items, which means lower production, inventory and transportation costs. (Costa 2016.)

According to Costa (2016), Lesara produces from 50,000 to 100,000 new pieces of clothing annually, which is more than five times more than the numbers shown by H&M and Zara. Data helps to quickly understand which items are growing in demand and then adjust the production cycle to satisfy this demand.

The fashion world still lives from autumn/winter to spring/summer collection, but this scheme will soon become a thing of the past. Agile fashion with the assistance of data drastically decreases the time between the initial idea and shipping of the finished product to a customer. Roman Kirsch claims that by using data from different sources, they *are able to design, manufacture, and start shipping to customers in 10 days.* (Costa 2016.)

11.1.5 Lesara's sources

Speaking of sources, Kirsch stresses the importance of extending the data pool. Lesara uses Google trends, social media, fashion bloggers, its own web analytics and inventory management systems, for example. The company tries to find new sources to stay up-to-date and discover new insights (Bartlett n.d.).

According to the Chapter 9.3, most companies use cookie files to collect and store information from websites and social media platforms. Lesara definitely allocates a significant part of its spending to buying the first-party cookies from many websites, connected with the fashion world.

Kirsch explains that "the more data and scale [Lesara] have, the more sophisticated and accurate the algorithms become across all relevant fields of application – from understanding and predicting consumer demands in real-time, to improving marketing efficiency." (Mitchell 2016.)

However, as mentioned in Chapter 5 , more data could pose the problem of data volume. To be able to handle all the data, Lesara would need a large and reliable storage system.

11.1.6 Summary

All in all, Lesara could well be responsible for revolutionizing the clothing retail sector, by introducing a big data-driven approach to fashion. The company not only deploys such multi-industry practices as precise demand estimation to eliminate over-production but also successfully predicts fashion trends ahead of the competition. As the pace of life increases, agile fashion with its constant new designs and low prices is likely to soon replace the traditional seasonal fashion.

11.2 Nordstrom

11.2.1 Company description

Unlike Lesara, Nordstrom is a rather old company, with an age-long history. It was founded in 1901 in the USA. Nordstrom operates in the fashion retail sector in three countries with the majority of the stores located in the USA. The company has always strived for innovation and exceptional customer service. To reach the latter, they started exploring big data.

11.2.2 Focus on traditional stores

Another difference to Lesara's case lies in Nordstrom's approach to the technology, as they wanted to implement it in the traditional stores; more specifically, they wanted to collect data about customers when they are in a Nordstrom's store. The company has employed a software that is an analog of RetailNext, a tool that tracks customers in physical stores, allowing to track customers' behavior in a shop in terms of where they walk, how long they stand in front of which products and how much time they spend in the store. There was no personal identification, but customers were upset about their privacy being disrupted and Nordstrom had to shut down the experiment. (Van Rijmenam n.d.)

RetailNext is a tool that uses data from different sources to collect information about customers. The company's website lists the sources: video camera feeds,

point of sale systems, Wi-Fi- and Bluetooth-enabled mobile devices, guest Wi-Fi-connected devices, workforce management systems, staffing schedules, promotional calendars, weather systems and third-party API applications. Then the data is brought to a single place to be analyzed and visualized. RetailNext suggests that by using the software companies are able to answer such questions as who are the customers, how many people enter the store, how they behave in the store and what the employees are doing. (RetailNext n.d.)

After a while, Nordstrom introduced another project involving the use of data. Some customers are willing to provide personal information to companies when acquiring a loyalty card. Such information as previous purchases or favorite brands is collected by Nordstrom and used when a particular customer enters the store. The system identifies the customer and sends his or her profile to a Nordstrom's salesperson. Thus, with a better knowledge of a customer, the company is able to provide a better customer experience and increase the potential profit from a single interaction, as they become more personalized. (Van Rijmenam n.d.)

With more than 300 stores and thousands of customers, Nordstrom would want to invest in a physical data storage. As the Chapter 5.3 suggests, in that case Nordstrom would own the facility, which will ensure the confidentiality of the information customers give to the company. Nordstrom would not have to keep the customer profiles on the third-party servers.

11.2.3 Connecting online and offline

Nordstrom has found another application for big data in creating a cross-channel inventory system that allows customers to see which products are or will be available at a certain location. Thus, Nordstrom combines online and offline capabilities to provide improve customer experience. (Van Rijmenam n.d.)

Apart from that, Nordstrom analyze great amounts of posts from social networks to determine which items are popular among the target audience. Then they put corresponding tags on products in their stores, thereby providing another reference to customers and connecting online resources with brick and mortar stores. (dturenshine 2017.)

11.2.4 Nordstrom's sources

Nordstrom uses a number of different sources to collect data from, including Point-of-Sales data, their own website, Facebook, Pinterest, Twitter and loyalty cards. (Van Rijmenam n.d.)

According to the Chapter 9.3, Nordstrom, just like Lesara collects information from online sources by plugging their own cookie files into different websites and by buying third-party cookies. All in all, Nordstrom is using a combination of structured data of two types (machine-generated, for example the Point-of-Sales data, and human-generated, for example the input and click-stream data) and unstructured human-generated data (social media posts).

11.2.5 Summary

On the whole, Nordstrom is a perfect example of successful use of big data to improve customer experience both in stores and online. The company understands its main purpose and designs all the new features in accordance with the main customer-centric strategy.

11.3 Comparison

Lesara and Nordstrom have quite different approaches to big data, but both companies succeed in reaching their goals, which shows the diversity of big data practices. While the Germans mostly look forward and use data to predict the future, the Americans focus on the past and present.

The difference is also shown in the purpose for what the companies are using big data. Lesara, as a company that does not have any physical interferences with customers, are mostly concerned with improving customer experience through extending their assortment and decreasing prices and shipping time. Nordstrom, on the other hand, invest in thorough studying of their customers to provide the best service at the points of contact.

11.4 Discussion

While considering the two cases of successful big data use, this paper will discuss the following questions: What are the key assets needed in both cases? Is it possible to combine the achievements of the two companies?

Lesara needs a perfect interaction between the analytical and logistics departments, as one of their main targets is shrinking the time between the initial concept and the finished product. Flawless communication, information flow and accessibility throughout the company are an absolute necessity. This helps Lesara in dealing with the problem of data velocity. Every part of the company needs to be ready to adjust to changes in the strategy, depending on the appearance of new insights, derived from fresh information. At the same time, Lesara should strictly define the data privacy frameworks. Their office and production units are situated in different locations, for example China. (Mitchell 2016.) Thus, it is getting more difficult to control the accessibility of the data.

Nordstrom situation requires a close connection between the customer experience manager and the IT department. Judging by the first experiment of the company, big data technology needs to be introduced appropriately so that customers will not get upset over innovations. Nordstrom's audience is used to a more traditional shopping style than Lesara's customers, and any change should be properly explained and justified. Apart from that, a large and powerful data storage is a necessity for Nordstrom, because they need to store and process thousands of customer profiles and millions of data units.

Both companies need a large pool of reliable information sources. For Lesara it is a bit more vital, as their goal is to determine the future trends and this information could come from unexpected sources. Nordstrom also need to keep pace with the rapid evolvement of social networks, where most of the customers leave their reviews on products.

12 Summary

The present thesis has investigated the topic of big data. It provided a general overview of the concept with a discussion of subjects such as data storage,

security, data mining and Hadoop that was followed by a thorough examining of the technology's application in marketing.

The paper has provided answers to the research questions. It has discovered what a company needs in order to use big data from the technical, human and knowledge point of view. First, it needs a data storage facility, either physical or cloud-based. Then, a team with a variety of both technical and creative skills must be assembled. The company also needs software and the related know-how along with the pool of data sources.

The questions of how to start and how to use big data were answered in the in the Chapter 9, which has provided an understanding of the necessary actions and assets that a company must do and have in order to successfully operate the big data capabilities. The process of incorporating big data in a company could be divided into six main steps: setting goals, selecting a small portion of data to start with, deciding on the human resource component (assembling a team or outsourcing), selecting the infrastructure, tools and architecture for the big data department, preliminary testing and increasing interaction between departments.

The study has determined that in general companies use big data to increase the accuracy of different types of decisions. Competent employment of big data in marketing results in enhanced customer knowledge. Companies then use this knowledge to provide exceptionally personalized promotional content. As for the clothing retail market, another significant advantage that big data brings is the ability to predict trends and therefore the demand.

The thesis has discovered that cutting the costs of big data is possible via using third-party services. The method works well for a smaller business, with privacy concerns and partial control over operations being the main downsides.

The research has established that the data that a company gathers from social media is usually raw and requires additional processing. The data is unstructured, which makes it harder to analyze but it often brings more value. Companies are able to get the most honest and direct feedback from the online interactions with customers. Social media data is a great indicator of customer engagement as well, which contributes to the overall understanding of an audience.

The theoretical part of the paper was followed by a case study featuring two clothing retail companies Lesara and Nordstrom. The case study has proved the versatility of potential big data employment: the first company operates solely online while the second one mostly focuses on the traditional stores. Both organizations were able to find a way to use big data and benefit from it.

Overall the report has aided in understanding the basics of big data concept as well as the more detailed aspects of its use in marketing.

List of figures

List of tables

List of references

All AboutCookies.org n.d. All AboutCookies.org - FAQ Section. http://www.allaboutcookies.org/faqs/cookie-file.html. Accessed on 20 March 2017.

Altintas, I. n.d. Characteristics of Big Data – Veracity. https://www.coursera.org/learn/big-data-introduction/lecture/Ln0ll/characteristics-of-big-data-veracity. Accessed on 6 December 2017.

Andreasen, S. 2013. Intelligence by variety in Big Data World. https://www.columnfivemedia.com/work-items/infographic-intelligence-by-variety. Accessed on 21 February 2018.

Baker, B. Kiewell, D. and Winkler, G. 2014. Using big data to make better pricing decisions. https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/using-big-data-to-make-better-pricing-decisions. Accessed on 23 March 2018.

Bartlett, B. n.d. What Happens When Big Data Meets High Fashion. http://www.adweek.com/sponsored/what-happens-when-big-data-meets-high-fashion-174112/. Accessed on 23 March 2018

Bauer, R. 2017. Transparency in Cloud Storage Costs. https://www.backblaze.com/blog/transparency-in-cloud-storage-costs. Accessed on 21 February 2018.

Big Data University 2016. What is Big Data online course. https://bigdatauniversity.com/courses/what-is-big-data/. Accessed on 21 February 2018.

Carden, D. n.d. Database vs Data Warehouse: A Comparative Review. https://www.healthcatalyst.com/database-vs-data-warehouse-a-comparative-review. Accessed on 28 February 2018.

Carr, D. 2013. Giving viewers what they want. http://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html?_r=0. Accessed on 21 February 2018.

Cooldata. 2014. The True Cost of Building a Big Data Solution. https://www.cooladata.com/blog/true-cost. Accessed on 21 February 2018.

Costa, D. 2016. Lesara: Big Data Is the Hot New Trend in Fashion. https://tech.co/lesara-big-data-fashion-startup-2016-04. Accessed on 23 March 2018.

Cringely , R. X. 2016. Thinking about Big Data – Part One. http://www.cringely.com/2016/07/05/thinking-big-data-part-one/. Accessed on 21 February 2018.

Data Center Frontier 2015. Inside Facebook's Blu-Ray cold storage data center. https://datacenterfrontier.com/inside-facebooks-blu-ray-cold-storage-data-center. Accessed on 23 April 2018.

Datawarehouse4u 2009. OLTP vs. OLAP. http://datawarehouse4u.info/OLTP-vs-OLAP.html. Accessed on 28 February 2018.

Delgado, R. 2017. 5 Tips to Prepare Yourself for Big Data Analytics. https://insidebigdata.com/2017/06/15/5-tips-prepare-big-data-analytics/. Accessed on 19 January 18.

Denthumdas, S. 2013. Apache Hadoop Explained in 5 Minutes or Less. https://www.credera.com/blog/technology-insights/open-source-technology-insights/apache-hadoop-explained-5-minutes-less/. Accessed on 10 January 2018.

Dturenshine 2017. Nordstrom: Integrating Digital with Retail. https://digit.hbs.org/submission/nordstrom-integrating-digital-with-retail. Accessed on 23 March 2018.

Dykes, B. 2017. Big Data: Forget Volume and Variety, Focus On Velocity. https://www.forbes.com/sites/brentdykes/2017/06/28/big-data-forget-volume-and-variety-focus-on-velocity/#65341bda6f7d. Accessed on 6 December 2017.

Finsmes 2017. Lesara Raises $40M in Funding. http://www.finsmes.com/2017/09/lesara-raises-40m-in-funding.html. Accessed on 23 March 2018.

Fireside Analytics Inc. n.d. 4 V's of Big Data. https://bigdatauniversity.com/courses/what-is-big-data/. Accessed on 21 February 2018.

Franks, B. 2014. The analytics revolution: How to improve your business by making analytics operational in the Big Data era. Hoboken, NJ: Wiley.

Glass, R. Callahan, S. 2014. The Big Data-Driven Business: How to Use Big Data to Win Customers, Beat Competitors, and Boost Profits. San Francisco, CA: Wiley.

Glassdoor 2013. BI Dashboard Software Developer Salaries. https://www.glassdoor.com/Salaries/bi-dashboard-software-developer-salary-SRCH_KO0,31.htm. Accessed on 21 February 2018.

Glassdoor 2018a. Cloud Infrastructure Engineer Salaries. https://www.glassdoor.com/Salaries/cloud-infrastructure-engineer-salary-SRCH_KO0,29.htm. Accessed on 21 February 2018.

Glassdoor 2018b. Data Analyst Salaries. https://www.glassdoor.com/Salaries/data-analyst-salary-SRCH_KO0,12.htm. Accessed on 21 February 2018.

Glassdoor 2018c. ETL Developer Salaries. https://www.glassdoor.com/Salaries/etl-developer-salary-SRCH_KO0,13.htm. Accessed on 21 February 2018.

Glassdoor 2018d. Python Developer Salaries. https://www.glassdoor.com/Salaries/python-developer-salary-SRCH_KO0,16.htm. Accessed on 21 February 2018.

Greenwood, S. Perrin, A. Duggan, M. 2016. Social Media Update 2016. http://www.pewinternet.org/2016/11/11/social-media-update-2016/. Accessed on 21 February 2018.

Hadoop Wiki n.d. Powered by Apache Hadoop. https://wiki.apache.org/hadoop/PoweredBy#A. Accessed on 10 January 2018.

Hinman, H. 2013. 9 Data Mining Challenges From Data Scientists Like You. https://info.salford-systems.com/blog/bid/305673/9-Data-Mining-Challenges-From-Data-Scientists-Like-You. Accessed on 24 December 2017.

Hurwitz, J. Nugent, A. Halper, F. Kaufman, M. Unstructured data in a big data environment. http://www.dummies.com/programming/big-data/engineering/unstructured-data-in-a-big-data-environment. Accessed on 23 April 2018.

Investopedia n.d. Data Mining. https://www.investopedia.com/terms/d/datamining.asp. Accessed on 23 April 2018.

Jackson, D. 2017. 6 Instagram Metrics That Truly Measure Your Efforts. https://sproutsocial.com/insights/instagram-metrics. Accessed on 18 February 2018.

Johnson, M. 2016. 5 factors for calculating the cost of a big data strategy. https://fusionalliance.com/calculating-cost-data-strategy. Accessed on 21 February 2018.

Körting, T.S. 2014. How data mining works. https://www.youtube.com/watch?v=W44q6qszdqY. Accessed on 24 December 2017.

Leonard, A. 2013. How Netflix is turning viewers into puppets. http://www.salon.com/2013/02/01/how_netflix_is_turning_viewers_into_puppets/. Accessed on 21 February 2018.

Lohr, S. 2013. The Origins of ‘Big Data': An Etymological Detective Story. https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/?_r=0 . Accessed on 21 February 2018.

Maheshwari, A. 2013. Data Analytics Made Accessible. Fairfield, IA: Wiley.

Marr, B. 2016. 17 Predictions About The Future Of Big Data Everyone Should Read. https://www.forbes.com/sites/bernardmarr/2016/03/15/17-predictions-about-the-future-of-big-data-everyone-should-read/#225409e11a32. Accessed on 21 February 2018.

Marr, B. n.d. Big Data: What Is Hadoop - An Easy Explanation For Absolutely Anyone. https://www.bernardmarr.com/default.asp?contentID=1080. Accessed on 10 January 2018.

Matias, Y. 2012. Insights into what the world is searching for -- the new Google Trends. https://search.googleblog.com/2012/09/insights-into-what-world-is-searching.html. Accessed on 5 April 2018.

Mitchell, J. 2016. This 27-Year-Old Founder Is Transforming High-Fashion Into A Billion-Dollar Tech Industry. https://www.forbes.com/sites/julianmitchell/2016/07/11/this-27-year-old-founder-is-transforming-high-fashion-into-a-billion-dollar-tech-industry/#281577f34efd. Accessed on 23 March 2018.

Murdock, T. 2014. What Is Content Scoring? https://marketeer.kapost.com/what-is-content-scoring. Accessed on 20 March 2018.

PayScale 2018a. Data Engineer Salary. https://www.payscale.com/research/US/Job=Data_Engineer/Salary. Accessed on 21 February 2018.

PayScale 2018b. Database Administrator (DBA) Salary. https://www.payscale.com/research/US/Job=Database_Administrator_(DBA)/Salary. Accessed on 21 February 2018.

Pramanick, S. 2013. 6 steps to start your big data journey. http://www.ibmbigdatahub.com/blog/6-steps-start-your-big-data-journey. Accessed on 19 January 2018.

Pursiainen, S. 2016. Big Data in the marketing of Finnish B2C companies. Haaga-Helia University of Applied Sciences. Degree Programme in Business Information Technology. Bachelor's thesis.

Qubole. 2017. Big Data challenges and opportunity https://www.qubole.com/resources/solution/big-data-challenges/. Accessed on 30 April 2017.

RetailNext n.d. How it works. https://retailnext.net/en/how-it-works. Accessed on 5 April 2018.

Reynolds, C. 2016. 5 Important Ways Big Data is Changing Marketing. https://datafloq.com/read/5-important-ways-big-data-is-changing-marketing/1675. Accessed on 21 February 2018.

Rouse, M. n.d. sensor data. http://internetofthingsagenda.techtarget.com/definition/sensor-data. Accessed on 28 December 2017.

SAS n.d. Big Data, Big Marketing. https://www.sas.com/en_us/insights/big-data/big-data-marketing.html. Accessed on 26 December 2017.

Sherpa Software n.d. Structured and Unstructured Data: What is It? https://sherpasoftware.com/blog/structured-and-unstructured-data-what-is-it/. Accessed on 28 December 2017.

Spencer, J. 2014. 5 Ways Marketers Can Actually Use Big Data. https://www.salesforce.com/blog/2014/11/5-ways-marketers-can-actually-use-big-data-gp.html. Accessed on 20 March 2018.

Spiegel, E. 2014. Six challenges of Big Data. https://blogs.wsj.com/experts/2014/03/26/six-challenges-of-big-data/. Accessed on 21 February 2018.

Tanner, A. 2014. Different Customers, Different Prices, Thanks To Big Data. https://www.forbes.com/sites/adamtanner/2014/03/26/different-customers-different-prices-thanks-to-big-data/#51a542595730. Accessed on 23 March 2018.

Taylor, C. 2017. Big Data Security. https://www.datamation.com/big-data/big-data-security.html. Accessed on 6 December 2017.

Techopedia n.d. Decision Tree. https://www.techopedia.com/definition/28634/decision-tree. Accessed on 24 December 2017.

The free dictionary by Farlex n.d. Bills of Mortality. https://legal-dictionary.thefreedictionary.com/Bills+of+mortality. Accessed on 21 February 2018.

Valchanov, I. 2017. Is data science really a rising career? https://www.quora.com/Is-data-science-really-a-rising-career/answer/Iliya-Valchanov. Accessed on 23 March 2018.

Van Rijmenam, M. 2016. A short history of Big Data. https://datafloq.com/read/big-data-history/239. Accessed on 21 February 2018.

Van Rijmenam, M. n.d. How Fashion Retailer Nordstrom Drives Innovation With Big Data Experiments. https://datafloq.com/read/how-fashion-retailer-nordstrom-drives-with-innovat/398. Accessed on 23 March 2018.

Wide Skills n.d. Data Mining Processes. http://www.wideskills.com/data-mining-tutorial/data-mining-processes. Accessed on 24 December 2017.

York, A. 2016. How to Successfully Mine Your Social Media Data. https://sproutsocial.com/insights/social-media-data/. Accessed on 18 February 2018.

Ypulse 2016. 5 Stats on millennials, teens & social media. https://www.ypulse.com/post/view/5-stats-on-millennials-teens-social-media1. Accessed on 21 February 2018.

Zantal-Wiener, A. 2017. These 7 Brands Take Personalized Marketing to a New Level. https://blog.hubspot.com/marketing/marketing-personalization-examples. Accessed on 20 March 2018.