Universidad Zaragoza
1542

Undergraduate Dissertation

Big Data as a source of competitive advantages for companies

Data Analysis as the center of the firm's decision-making process

Author

Rafael Ruiz Amor

Supervisor

Eva María Pardos Martínez

Facultad de Economía y Empresa
Universidad Zaragoza

2021

ABSTRACT

The limit between the digital and the physical world is blurrier than ever. In this modern era, almost every individual action leaves a trace that represents an opportunity for companies to obtain information. In fact, a lot of the technology breakthroughs that are being created nowadays, are aimed at creating value from that mass of traces. Nevertheless, all of them base themselves in one wider concept: Big Data.

The proposition of Big Data is to study all those inputs in an aggregated way, in order to obtain knowledge and use it to improve different aspects of the society. Its applications are endless, however, one of the main beneficiaries are companies, who can find in this technology an ally to not only know their customers, but also themselves, enabling them to reach unprecedent heights.

The aim of this paper is to link that reality of Big Data, to the business world, by reviewing its notions and linking them to relevant economic concepts. Furthermore, the study will try to surpass the theoretical barrier by studying real companies' cases of Big Data usage.

Keywords: Big Data, digital world, Industry 4.0, competitive advantage, technology.

RESUMEN

Los límites entre el mundo físico y digital están más difuminados que nunca. En esta nueva era, prácticamente toda acción de un individuo representa una huella digital y una oportunidad para las empresas de obtener información. De hecho, muchos de los avances tecnológicos de hoy en día, tienen como objetivo crear valor de ese amasijo de huellas. Sin embargo, todos ellos se basan en un concepto primario: el Big Data o datos masivos.

Lo que el Big Data propone es estudiar todos estos inputs de una manera agregada, para obtener conocimiento y usarlo para mejorar distintos aspectos de la sociedad. Sus aplicaciones son incontables, sin embargo, uno de los principales beneficiados son las empresas, que pueden encontrar en él, el aliado perfecto para no solo conocer a sus clientes, sino también a sí mismos, para así poder llegar a límites inimaginables.

El objetivo de este trabajo es alinear los fundamentos teóricos del Big Data con conceptos empresariales relevantes. Además, el estudio va intentar ir más allá del plano teórico, mediante el análisis de casos reales de compañías que han adoptado esta herramienta.

Palabras clave: Big Data, mundo digital, Industria 4.0, ventaja competitiva, tecnología.

TABLE OF CONTENTS

LIST OF FIGURES

Figures:

Graphs:

Tables:

1. INTRODUCTION

The aim of this paper is to study one of the most important solutions applied to enterprises over the recent years: Big Data. This technology is still a stranger to the big public, who finds it, either difficult to understand, or inaccessible due to the lack of non-professional sources that talk about the topic.

In this way, this work will try to merge the potential of Big Data with the reality of the firms, by trying to prove why it is relevant for this area of the society and how it has affected companies that have already decided to take the chance that it offers.

In no way this paper tries to deepen into the knowledge of the technology itself, but what it tries to do is to contextualize it for the business world. Furthermore, the following document uses the concept and definition of Big Data as its base. It does not act as a guideline of implementation of Big Data in a company, but rather it tries to be informative about the potential that it holds.

In order to do so, firstly, a study of how the technology in the industrial world has evolved through the years will be performed, until reaching today's actual context and how these changes have brought new opportunities for companies. In fact, it will be noted how the Industry 4.0 is affecting to the current business world and why its characteristics make it the ideal candidate for a technology like Big Data.

Secondly, the foundations of Data Science and Big Data will be laid. In order to determine how they affect organizations; it is important to have a solid ground of what they represent and the general concepts that are linked with them. Afterwards, their implications for companies will be discussed.

In order to prove all the theoretical findings obtained, a real case analysis of three different companies which have integrated Big Data in their economic activity will be studied in section 5. This analysis will try to give answers to qualitative aspects of the applications and also, try to quantify what has been their impact in the general performance of these firms.

Finally, a series of conclusions will be drafted from these cases, which will try to bring these findings to a general level by pointing out common aspects, limitations and future lines of work.

But first, before analyzing Big Data in the actual business world, it is important to study how has technology evolved throughout the years, and what benefits have provided to companies in the different moments of the business history.

2. COMPETITIVENESS AND VALUE CREATION EVOLUTION THROUGH HISTORY

2.1 Evolution of the Industrial World

Across the history of the business world, decisive milestones have revolutionized the way that value and economic growth was being created at that moment. For example, in the case of the First Industrial Revolution (1784-1870), mechanization took a relevant role in all the aspects of human activity. Water power machines and steam engines made possible an unprecedent increase in productivity, which in turn made a great impact in the economic situation at the time (Pozdnyakova, U. A., et al., I. A., 2019).

The Second Industrial Revolution (1870-1969), was characterized by the ground-breaking conception of the assembly line, which limited the number of tasks performed per worker. Moreover, electrical power, enabled the constant functioning of all the machines needed to reach the point of mass production (Pozdnyakova, U. A., et al., I. A., 2019).

From 1969 up to some point in the early 2000s, the Third Industrial Revolution took place. The value creation process underwent a profound change due to automatization and electronics. Besides, Information Technologies provided organizations with the missing piece to have a total control over their activities, while at the same time offering a wide range of tools to reach new heights thanks to innovation, which was fostered by governments (Mowery, D. C., 2009).

All these periods are easily identifiable as they have similar characteristics and mainly concerned the production sector. However, there seems to be a current shift in the business world which is not as obvious and is being overlooked by many sectors, who find comfort on how things were previously done.

2.2 Industry 4.0

Lasi, H. et al, (2014) argued that we are undergoing our own industrial revolution at the moment. This paradigm-change is characterized by the digitalization of information, which has contributed to a rise in the intensity of the automatization of processes and increases in efficiency.

This new shift, fostered thanks to internet-technologies and future-oriented technologies, shows a predominant trend to connect physical and digital processes, breaking the important barrier that entailed working on the two dimensions at the same time, and increasing the ease to obtain information or control over the whole business, thanks to the information which can be obtained.

All previously mentioned changes are englobed under the term Fourth Industrial Revolution or Industry 4.0. This new situation is developing thanks to two drivers which are motivating the new reconfiguration of the processes behind economic activities (Lasi, H et al., 2014):

- Application-pull: The framework of how economic activity is performed is changing. Companies are drifting towards short-development periods, individual segmentations, flexibility, decentralization, resource efficiency…
- Technology-push: New technologies are pushing new business models forward, encouraging automatization, digitalization and miniaturization (being able to carry a business activity within small spaces).

Radical changes, like these ones, are always seen as threats to companies that are enjoying a privileged position in a previous established model. However, these actors do not correctly seize the opportunities that come with them. This new "industrial" picture is leading companies to an uncertain path as they have to adapt their current architecture to the one that the market is asking them in order to remain competitive.

That exact concept, competitiveness, is the main concern of any businessperson, as it is the key factor that will mark their situation in relation to their competitors, and consequently, the results of their activity.

However, these competitive advantages have not always been conceived to be the same. As the industrial history has been changing, the focus on different activities and areas has been shifting accordingly. In order to able to understand what drives a company to make certain decisions, a review of the evolution of this concept is needed.

2.3 Evolution of the concept of Sustained Competitive Advantage

Many scholars have tried to work out how companies work. Figuring out the functioning of the *black box* where you put inputs and you obtain a certain output has been one of

their main tasks. The processes which cannot be seen and that are key to understand how to manage a company. However, as the business world moves forward, previous conceptions become irrelevant, which makes this concept to be in constant study and evolution.

Barney (1991) proposed a Resource-Based View (RBV) of the firm. It defines the company as a set of processes and capabilities with given specific resources, which have to be managed to maximize their value. These resources have to be controlled to be completely efficient in the present.

However, they are not unlimited and they have to be managed by the firm to be able to assure the existence of them in the future. This concept could be summarized in: obtaining results in the short-run while guaranteeing a long-run continuity, which is the definition of Sustained Competitive Advantage (SCA).

In order to achieve that state, it is necessary that the company holds resources that are valuable, rare, inimitable and non-substitutable (VRIN), while at the same time having the capabilities (organization, O) necessary to manage and absorb them. This relation is represented by the following equation:

$$SCA = VRIN \; x \; O$$

This RBV of the firm later evolved into a Knowledge-Based View (KBV) of the firm. This theory proposed by Grant (1996), focused on superior knowledge as the central piece to maximize the value of the company and the key strategic resource through which companies can obtain SCA.

Kostopoulos, K. C., et al. (2002) highlighted the significance that intangible assets suppose for the companies, and how their relevance is increasing with time. However, knowledge is shaping to be the most important one of them, no matter if it is internal or external, as both are valuable for the firm.

Grant not only saw value in the company's knowledge, but also on the knowledge of the persons who integrated the organization. Being able to absorb and to manage it, is one of the critical activities that the company has to perform. Besides, what is wanted from the members is not to put together their knowledge into a common one, but rather to coordinate these different individual knowledges to provide strength to the company.

By this new definition of Sustained Competitive Advantage, it can be concluded that in order for a company to stay relevant in the short and long run, it has to be able to integrate new internal and external inputs in order to be able to use them to perform the changes necessary to obtain a competitive advantage from them.

These changes can be defined as innovations that a company creates from the knowledge obtained. Kostopoulos, K. C., et al. (2002) confirmed a positive linkage between both concepts. If the company has a predisposition to integrate knowledge within the organization, it will be able to react accordingly and therefore create competitive advantages by evolving in relation to it.

However, thanks to its intangible nature, knowledge is difficult to identify and capture. This makes the creation of Sustainable Competitive Advantages a hard task for any company still running. Nonetheless, the new shift in the business world, the Industry 4.0, and all the changes that come with it, are providing companies with the tools necessary to handle it in a proper way.

3. THE ANALYSIS OF DATA AS THE KEY COMPETITIVE ADVANTAGE IN THE INDUSTRY 4.0

3.1 Data Science, the new discipline to interact with the external world

The Knowledge-Based View of the firm defines being competitive as the ability to respond to the external and internal world and to learn from them, to implement what has been observed into the organization and thus obtain an advantage over the rest of the players.

Nevertheless, in the current Industrial model, almost everything produces data and information due to the general digitalization of processes. It is impossible for firms to react to each and every single signal, and therefore the management of all of this data has been subject to a very exhaustive study over the recent course of history.

However, this new Industry 4.0 has come with new tools and innovations, which are able to manage such complicated matter, and therefore, are able to provide companies with the advantages that come with acquiring and absorbing knowledge.

The solution proposed is a new discipline that has been created to study how data is analyzed and studied. Its name is Data Science, which is expected to shape the way companies make decisions and react to the external and internal environment

Van Der Aalst, W. (2016) proposed the following definition for it:

> "Data science is an interdisciplinary field aiming to turn data into real value. (…) Value may be provided in the form of predictions, automated decisions, models learned from data, or any type of data visualization delivering insights. Data science includes data extraction, data preparation, data exploration, data transformation, storage and retrieval, computing infrastructures, various types of mining and learning, presentation of explanations and pre- dictions, and the exploitation of results taking into account ethical, social, legal, and business aspects..."

The treatment of data is not something new, however, the amount that was being dealt with some years ago is not the same as it is now. The innovation comes when companies are able to capture value from such huge amounts generated nowadays, which would have been impossible to work with before.

At the same time, companies do not only use this data to get to know their clients or even their own organization, but also to innovate by trying to predict these behaviors and to

stay ahead of possible changes in the market. That is, this innovation offers the firms a SCA through knowledge.

What we can also take away from the previous definition is that Data Science does not only involve the statistical and predictive management of data, it also entails the resolution of problems from a business perspective, therefore, it is an applied science. Numbers are not the only matter subject to study, as quantitative information is used to infer qualitative information that can be useful in the decision-making process.

It is obvious then, that in order to be able to handle such science, it is important to have a prior knowledge of the issues that are being dealt with. Data and information are meaningless without a proper contextualization of it (Waller, M. A., et al., 2013).

From a business point of view, Data Science presents itself as an incredibly powerful tool through which data is captured and studied, and also a way to quantify behaviors and even to try to predict them. For example, a company can anticipate the needs of its customers and adapt to them specifically, even creating an individualized segmentation where the activity is completely adjusted to the client. In the same way, the tool can give insights about the activity of the firm by recording processes and pointing out inefficiencies and possible changes to be made.

3.2 What is the process behind Data Science?

Inferring concepts is a process that must be followed step by step. Therefore, three hierarchies are set (Fávero, L. P., et al., 2019):

- In a first step, data is gathered. The company collects data from its customers, its market or its own firm and stores it. This data by itself does not mean anything, it is just numbers which do not give any prior value to the firm.
- Secondly, this data is treated and analyzed. Relation and indexes are found, and at the end, information is obtained. This information already gives us an insight on what the company decisions should be.

Figure 1 Process from Data to Knowledge

Source: Own elaboration

- Finally, this information is applied on the decision-making process and it provides the company with knowledge. There has been a contextualization of this information and it is useful without having to look at the raw data.

The discipline provides companies with a new way to decide on the different aspects of the organization. It leads to the creation of *data-driven decision making*. This new way to solve problems allows firms to not only work them out more efficiently, but also to recast these problems.

Prior to the handling of data, an issue can be seen as very simplistic and the solution to it very straightforward. However, once the data related to the problem is treated, the insights obtained can evidence that the efforts are not being directed towards the right direction. What was really evident at first sight, might be much more difficult than what it was thought. Thanks to this, organizations can recast themselves to a new much more accurate objective.

Nevertheless, we have to bear in mind that the treatment of data is not something exclusive to a particular company. The rest of competitors can invest in this discipline as well, so in order to keep that SCA is necessary to always stay a step ahead and to be constantly investing in Data Science (Fávero, L. P., & Belfiore, P., 2019).

In the case of the Industry 4.0, one of the main innovations that can offer new business opportunities to companies seems to be the study of data or Data Science. Therefore, firms should start to adopt these practices, not only to be able to survive this new revolution, but also to increase their competitive advantage.

In this way, we can say that Data Science is a disruptive innovation as expressed in Clayton's work of 1997 *The innovator's Dilemma.* He expressed that disruptive innovations give companies the competitive advantages necessary to surpass the rest of competitors in the market. As the Knowledge-Based View of the firm proposed by Grant in 1996 expresses, Sustained Competitive Advantage comes from absorbing this knowledge and Data Science could be considered as the missing piece to do so, providing competitive advantage and, therefore, being classified as a disruptive innovation.

However, Data Science is not as simple as treating and contextualizing data in a certain way. It holds much important and specific concepts and solutions that are shaping the future of the economic activity. The most important one, a word that it's one of the main concerns of companies nowadays: Big Data.

4. BIG DATA. THE NEW HORIZON

4.1 The Big Data concept

The term Big Data has been widely used with different meanings that do not give a clear representation of what it represents. A lot of people have applied it to different contexts, which has not helped to a homogenization of the concept and has led to it being a great stranger to the general public.

De Mauro, A., et al. (2016) were aware of this issue, and therefore tried to come up with a general definition to be used in the subsequent years, based on the research of previous works. According to them Big Data is defined as:

> "(…) the Information asset characterized by such High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value."

There are a few points that we can infer from this definition. First of all, Big Data by itself has no value. When we talk about this discipline, we are dealing with enormous databases that at first sight, do not give us any insight on the problem that we are trying to solve.

A second point would be that Big Data has its own tools to be analyzed with, so we can conclude that in order to capture value from it, we have to invest to obtain the right instruments or even technology to carry out proper studies. That is, it cannot be adopted as easily as it might seem.

A third point is the mention of a few concepts which are very relevant to the study of this phenomenon: Volume, Velocity and Variety. These are the so-called three V's of Big Data, which will be developed later on.

However, where does this data come from? The new Fourth Industrial Revolution paradigm has led to almost every single process to have a digital component. Almost everything that we do in the physical world, ends up, somehow, reflected in a digital trace. This happens thanks to a new concept that has gained popularity in recent years, the Internet of Things (IoT).

This term was coined by Kevin Ashton in 1999, in response to the dependence of many devices that we use on daily basis to the interconnection, thanks to networks, between them. This way, IoT refers to the connection to Internet of physical aspects of our

environment, that sometimes, even need this connection to function properly (Ashton, K., 2009).

In that way, most devices that are commonly used in the daily life of the population generate data that is later used by Big Data as its source to generate value.

Nevertheless, in order to be able to fully comprehend the meaning of this concept, it is necessary to study its characteristics in order to understand how can it be useful for companies all across the globe.

4.2 The three V's of Big Data and the two new ones

It is generally recognized that the concept of Big Data can be broken down into three main characteristics which have been previously mentioned: Volume, Velocity and Variety (Schroeck, M., et al., 2012).

- *Volume.* The data that is contained in these data sets are of such scale that cannot be humanly analyzed without some external tool. Schroeck, M., et al., 2012, pointed out that these amounts are not Terabytes or Petabytes, but Zettabytes that cannot be easily stored or used. Big Data works because of its volume. For example, to specifically know the preferences of customers, five samples of data would not return an accurate representation, as the value is obtained from insights that big amounts can provide. Furthermore, it could not be considered as Big Data.
- *Velocity.* This characteristic is referred to the speed at which the data is generated, what makes it impossible to process it with regular techniques. This seems obvious as Big Data is generated because it registers millions of transactions and all kinds of information per second. A perfect example would be a hospital in which the characteristics of the patients, pulse, electrocardiograms, oxygen in blood, and other metrics are measured every minute. If the hospital holds hundreds of patients, the amount of data that is being generated by the minute in order to keep up with the actual conditions of its guests is enormous to be able to offer a rapid response and detect any problem that might arise.
- *Variety.* Schroeck, M., et al., 2012 explained that the data generated comes from different types of sources which leads to have different kinds of formats. Three types can be distinguished, from easier to harder to analyze: *structured*, *semi structured* and *unstructured.* Taking the example of the hospital once again,

among the mentioned metrics, there are measurements that can be expressed as a number (oxygen in blood as a %), or measurements that can be considered as qualitative (cardiograms: sinus rhythm, tachycardia…).

These three characteristics explain the study of Big Data as a discipline of its own. The properties of these datasets represent such a difference from those which were generated before, that the techniques and challenges are completely different. Therefore, it forces this discipline to be looked at as an independent phenomenon.

Traditionally, it has been considered to have these three dimensions, but as the technology is further developed and studied, researchers have defined two more relevant aspects that have to be taken into account (Anuradha, J., 2015):

- *Value.* This characteristic is important at the time of selecting which data to analyze or store. It is useless to study data that is not meaningful to the problem that has to be solved. For example, taking into account the previously mentioned example of a hospital, obtaining data about the number of planes that take off in the airport of the city, is not relevant to the wellbeing of the patients.
- *Veracity.* This characteristic makes reference to the quality of the data. It is important that this characteristic is as present as possible, as it will have a direct impact on the outcomes of its study. Going back to the example of the hospital, if an ECG machine is not calibrated, the data that it will generate, will not be reliable, and therefore and it will not provide accurate information about the state of the patient.

Taking into account these last two V's, the general situation of the characteristics of Big Data could be presented as follows.

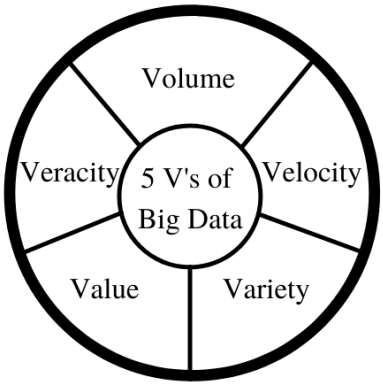


Figure 2 Summary of Big Data's 5 V's

Source: Own elaboration

4.3 Real Examples of Big Data

The concept Big Data can be too abstract. In order to understand it in a better way, some examples of applications in the real world will be presented.

Healthcare

Healthcare is now on the frontline more than ever. One of the core elements of the society is still stuck in the past from a technological perspective, as it could benefit from automatization and digitalization of processes. After all, from a data point of view, what a doctor does is to analyze information and variables, in order to give an answer, or in this case, a diagnosis. This makes healthcare a perfect candidate for the introduction of treatment of Big Data.

This is the case of Electronic Health Records (EHR's), as Dimitrov, D. V. detailed in his 2016 paper, *Medical Internet of Things and Big Data in Healthcare.* Something as simple as storing all the past information of a patient into a database can save many lives. As the written information is digitalized, it would be restructured to be easy to interpret and to be shared with everyone instantly.

Some years ago, the countless documents that had to be dealt with in a hospital, were often in written format. This made the job very difficult, given that for a simple procedure it was necessary to check up the personal file of a patient to know if that specific treatment could go ahead. However, with EHR's, it is as simple as accessing the platform and checking all the information in a matter of seconds.

Smart Cities

As the society goes forward, all its aspects start to get digitalized. All these data generated is stored in enormous databases that can be analyzed and give answers to radical problems of cities in terms of planning of resources, self-governance and environmental issues.

This is the case of Stockholm (Hashem, I. A. T., et al., 2016), which used Big Data as a solution for waste transportation. From vehicles, to traffic or weight of containers were recorded in a database that was used to draw more efficient routes of waste-collection. It not only helped in the managing of these resources, but it also contributed to traffic issues and, in turn, environmental issues as the routes were more efficient.

Helsinki, is often taken as a standard for smart cities (Hashem, I. A. T., et al., 2016), thanks to the open access to databases that the government granted. Anyone can work with them and propose solutions to many problems, which has led the city to transform into a smart city in a completely natural way.

Banking Sector

Big Data could be a game changer for the banking sector thanks to a tool called Blockchain **(**Hassani, H., et al., 2018). This complex technology could be defined as an easy and fast transactions system. This is it, thanks to nodes that record and check that the information transmitted matches with the copy of the register that each of them has. The data generated is stored in a block, and once this block is completed, another block is created. However, this new one not only contains its own data, but also the identification that links it with the previous block and its information.

We have to take into account that for a transaction to proceed, it has to be accepted by all the nodes. This way, if someone wanted to fake a bank transaction, it would take them an enormous amount of time to register it, as they would have to modify, not only every single block on the chain, but also, they would also have to modify all the exact copies of every single node to get accepted.

This Big Data application would not only revolutionize the banking sector in terms of privacy, it would also entail the elimination of bank intermediaries, who are in charge of checking the transactions, as they would be accepted by nodes automatically, resulting in a more secure and efficient way to manage monetary resources.

These are only a few of the applications that Big Data could have for the society, which could drastically change the current situation towards a more sustainable and efficient way of living. However, how can Big Data help companies to improve their processes and competitive advantage? What impact can this discipline have in organizations and how can it integrate in them?

4.4 Impact of Big Data in companies

Previously on this paper we talked about different views of the company with respect to how they obtain the competitive advantage needed to survive in the markets. The second one, KBV of the firm, sought SCA by acquiring and absorbing knowledge. Big Data could be used as the main engine to drive it and, therefore, as the main tool to obtain this competitive advantage. Bearing this in mind, we will take a more in detail look to how this discipline can affect companies.

The first and most important benefit of the Big Data Adoption (BDA) is the *information obtained from this tool*. It can not only provide insights about customers, but also is a great pool of knowledge of the market, which enables the company to act accordingly to what is happening on its external environment (Bartosik-Purgat, M., et al., 2018). In this way, the company can create its own database, or buy one, in order to be used by its workers, avoiding asymmetries in the information handled in every part of the organization. One database, one truth. (Lukić, J., 2017).

The second benefit is *time and cost reduction*. By obtaining information not only from their environment, but also form internal resources, the company can decide on which activities to prioritize. This will lead to a better management of the limited resources of the company and better financial performance, as its assets will be used in a more profitable way. Through Big Data, companies are able to detect which aspects of their activity (raw materials, factories, workers…) are incurring in more costs, giving them the chance to act over them (Barham, H., 2017). This results in a better supply chain management, among others.

The third benefit is a *higher probability of sales increase*. The term probability is used because demand and supply models have an unpredictable component that cannot be measured. However, as companies react accordingly to the preferences of their customers, they are able to position themselves in a privileged situation, as they are offering exactly what their public is asking for. In the same way, they are also ablet to target the right segments by using the proper strategy for them. (Bartosik-Purgat, M., et al., 2018).

The fourth benefit is *higher probability of getting new customers*. As with the previous case, nothing can be taken as certain, but the company can search new customers in segments that are right for the product that it offers. It can also redirect all its efforts of

promotion to these potential customers or, at least, reach users that are certainly interested in what it is offered (Lukić, J., 2017).

All these benefits can be summed up in a *better decision making* (Lukić, J., 2017). As companies have the right information, they are able to make the right decision at the right moment (Barham, H., 2017). These include new decisions that can be implemented, or decision on past actions that are now being monitored thanks to Big Data.

These benefits are, in fact, very general as we can see, however, their effects lead to *better financial results*. As the processes are optimized, the costs incurred decrease drastically by allocating the resources in a more effective way, and therefore, turning existing processes into more profitable activities.

Bartosik-Purgat, M., et al. (2018) summarized all these benefits into three main competitive advantages:

1. Product quality advantage
2. Customer relationship advantage
3. Risk reduction advantage

Which in other words could be summarized as: a personalized product for a specific targeted customer in a more profitable and secure way.

However, not all the companies are able to take advantage of this competitive advantages in the same way (Bartosik-Purgat, M. et al., 2018). The ability to capture value from Big Data depends on the nature of the company. For example, a public institution does not obtain benefit in terms of product quality advantage.

4.5 Managerial Implications and Challenges for Big Data

Big Data is a complex tool that cannot be implemented in every company. Most of them do not have the proper structure to capture value from the recording, monitoring and analysis of big amounts of information, either because their processes are not digitalized or because they do not have enough amounts of data to obtain relevant and conclusive results. Nevertheless, organizations that could benefit from it, might have to overcome challenges at its implementation, too.

The first issue is that teams in the company need to understand what success is for their organization. Using Big Data enables to quantify phenomena; therefore, workers have to

be able to ask the right questions so the value obtained corresponds to what was expected and a clear answer can be obtained. Moreover, the organization itself has to have a clear preference for data driven approaches to the decision-making in order to be able to capture all the value and advantages that this tool provides them (Grover, V., et al., 2018).

There are monetary implications as well. Of course, Big Data needs an infrastructure to function properly. For example, it needs a data warehouse where all the data can be stored or different tools or analytical solutions that might be used to manage this data, which of course have a price as well.

This cost is something that every company will have to face, but there is another bigger problematic monetary challenge at the time of using Big Data. Companies live on a budget and storing data, as well as processing it, has a clear cost. Therefore, companies will have to identify which data is important and can generate value to the organization, without making the mistake of leaving out parts that can be crucial for the analysis of the problem. The bigger the database, the higher the cost for the company (Almeida, F., et al., 2013).

On the other side, there is the issue of talent. The company needs to have the right employees to manage Big Data projects, these are the data scientists. However, this term might be too general and different specific roles should be identified (Stephenson, D., 2018). Such positions can be:

- Platform engineers
- Data engineers
- Algorithm specialists
- Business Analysts
- Web Analysts

Relating to the talent, there is also a clear need to hire and to appoint the right people for the projects. There will be times when the person who has the capabilities to manage the data is not the same person that knows how the economic activity of the company works. They might have to work very closely and there can be communication issues. However, it is crucial that they are able to collaborate, as real value can be obtained in that cooperation kind of approach (Almeida, F., et al., 2013).

These are some of the most important challenges from an implementation point of view. However, when dealing with technologies like this, there are bigger issues that are not as

clear as the previously mentioned and which can have more impact than any kind of monetary or communication issue.

4.6 Legal and ethical issues when dealing with Big Data

As time goes by, the daily life of every person integrates more digital interaction points, which generate data, given that digitalization starts to play a relevant role in every aspect of the society. This digitalization comes with concerns regarding privacy, protection of interests and, of course, responsibilities for companies when handling this data. These responsibilities can be distinguished into two types: internal and external.

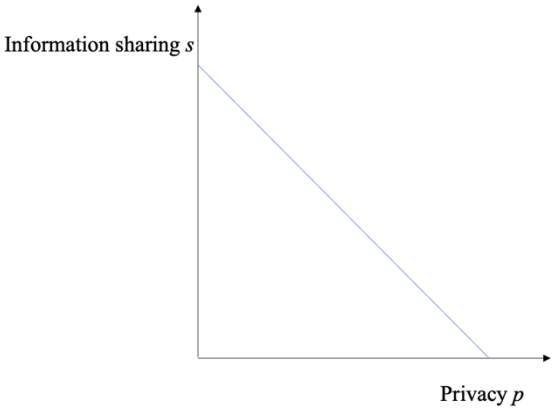
Internal responsibilities refer to the intentions and ethical behavior of a company (Zwitter, A., 2014). Privacy is one of the most general concerns in this field. Big Data needs those traces that digital transactions leave in order to work. Nevertheless, there is a bad perception of people regarding the sharing of their data. This problem comes with the fact that every piece of data is generated by a user, and the implications of a third party having someone's data are negative.

One of the main answers that are given to this matter is the anonymization of data. However, if this data is dehumanized and all the properties that identify it with the person that generated it are taken away, Big Data cannot capture value from it as the content that could be inferred would, in fact, give little to none insights on the issue. This is one of the main contradictions that this discipline faces.

In the same way, Big Data does not only generate value to the companies, but it also creates benefits to the customers. To obtain these benefits, users have to give consent to the treatment of their data. Even though there starts to be more trust in these practices, there is a significant part of the society who still find them unsettling and are still reluctant to do so.

As Puaschunder, J. M. (2019) pointed out, economics study the utility of economic factors, and this extended theory can be applied to the case of Big Data as well.

As it can be seen in the graph, assuming that the amount of data that someone can share is limited, an indifference curve is created with two variables that are substitutable between them, p (privacy) and s (information sharing). This means that the increase in one of them, produces a decrease in the other one. Which can be translated to the sharing of information entails the losing of privacy.



Graph 1 Relation between privacy and information sharing

Source: Puaschunder, J. M. (2019)

In this way, a utility function is created which represents the value that the user obtains from both variables:

$$U(s, p) = \alpha s + \beta p$$

Where α and β are the coefficient of information sharing and privacy respectively, which determine the utility that each variable adds to the whole function. The user can choose at which point of the indifference curve they want to situate themselves and what offers more value to them, their privacy or what they obtain from sharing their information.

As the years go by, users are more and more familiar with allowing the treatment of their data. This imposes a legal and ethical responsibility on companies to use it in a specific manner that should be stipulated and to follow a set of rules to protect the users.

There are multiple frameworks addressing these concerns of users when an external party deals with their data. These are the external responsibilities of the companies. The first and more important in this context, is the EU General Data Protection Regulation (GDPR). These regulation aims to harmonize all the regulation regarding data across the EU members states (Goddard, M., 2017).

It holds six principles which summarize its 100 articles. These are: fairness and lawfulness; purpose limitation; data minimisation; accuracy; storage limitation; and integrity and confidentiality.

This text is crucial for Big Data, as it deals with the raw input of the technology, data, in order to ensure an ethical and legal use of it.

Another international regulation regarding data is the ISO 20071. Its main objectives are to preserve confidentiality, integrity and availability of critical business information assets.

These two mentioned regulations are crucial and have to be taken into consideration when carrying Big Data projects, as companies will have to comply with them, not only to be considered as legal, but also to protect the users who generated the data that is being treated.

4.7 Big Data in Europe and Spain

Big Data is a questionably recent technology which is slowly making its way up to the top of companies' priorities. The reason is very simple, with the digitalization of the economy, the amount of data subject to be analyzed is extraordinary, and the bigger the database, the more accurate the results obtained.

In order to be able to analyze the current situation of the usage of Big Data in Europe, Pejić Bach, M. et al. performed a study in 2020 that concluded the following results:

33.5% of large European companies perform data analysis of any kind, whereas 19% of medium companies and only 10.5% of small ones use this technology. These figures show that this tool is still in development and it is far from being a common practice in the continent.

However, there are differences within the region that cannot be disregarded, in spite of the generalization of the analysis. Countries like Finland, Ireland or The Netherlands can be considered as the top Big Data countries in the continent, as their usage of the tool in their companies is the most frequent with figures of almost 50% of large companies and 30% and 20% of medium and small companies respectively.

Another set of countries like Germany or Norway, are following these leaders right behind with amounts that equal 35%, 20% and 12% of large, medium and small companies respectively.

Finally, countries like Greece, Italy or Spain represent the lowest figures in the whole European Union, with 27%, 14% and 7% of large, medium and small companies respectively. These countries share common characteristics like similar slow economic development and are some of the most damaged regions by the 2008 economic crisis.

The results show that even within the EU, there are inequalities regarding Big Data. However, the presence of this technology in the continent is obvious and deemed to keep growing, as the development of digitalization and the discipline continues to develop.

Throughout this whole section a deep analysis of Big Data has been carried out covering its meanings, requirements, challenges, problems, but more importantly, its advantages. However, there is not a better way to prove all these findings than looking and analyzing real cases where this technology has been implemented and the results obtained from it.

5. REAL COMPANIES CASES ANALYSIS

5.1 Methodology

As previously stated, the analysis of massive amounts of data can offer a lot of different competitive advantages to companies. However, as Grover et al. explained in their 2018 paper, *Creating strategic business value from big data analytics: A research framework,* the measurement of the impact of Big Data projects in the performance of companies is still something to be further studied, as there is not an indicator which offers a conclusive result.

For example, managers are always focused on reporting high levels of Return on Investment (ROI) on all of their projects, which can be an inconvenient given that Big Data projects need considerable amounts of money and money to start working properly. This can be one of the major reasons for companies to not see this kind of projects through, even when the long-term benefits are clearly evident. Linking that investment with actual results can be a difficult and inaccurate task.

This is the reason why, in order to be able to obtain a better understanding of the real value of integrating Big Data solutions into companies, I have performed an analysis of real cases of companies who have successfully introduced this technology into their organizations.

In order to avoid bias over the result, and understand the real influence of Big Data in the following companies, for each case an analysis on different indicators which measures the performance of the different areas affected by these projects has been carried out. These three applications cannot be measured in the same way as they affect different processes of the companies, with their respective characteristics.

These organizations have been chosen because of the variety of applications that they represent and because they are, in fact, success cases which can show the real potential of the tool. They are:

- *Walmart Inc.* and their use of real-time data analytics for a better supply-chain management.
- *Netflix Inc.* and the creation of algorithms of recommendations to increase the lifespan of its subscriptions.
- *Starbucks Corporation* and the improvement of its real state planning through Big Data tools as well as a better understanding of its customer preferences.

For each case a double analysis has been performed:

- In the first place, a qualitative analysis presents the case and the implications of the usage of Big Data in that area. This practical case has been taken from real uses and therefore, external sources.
- Then, in second place, an own quantitative analysis has been performed. A relevant amount of data has been gathered from each of the companies to be analyzed. These indicators obtained are the ones who could be affected by these Big Data projects, and therefore, the ones who can give true insights on the application of the technology.

5.2 Walmart Inc.

The first case to be analyzed is Walmart. Walmart is a large retailer store chain which focuses its activity in the United States. Its business model consists of two value propositions that drive customer to pick the brand over its direct competitors: a great diversity of products in each store combined with the lowest possible price.

At first, when identifying what part of the business could benefit from the use of Big Data, it can be difficult to link both realities as their nature is drastically different. After all, the case of Walmart is that of a physical model of retailing where the digital trace that its customers leave when they visit the stores, is only generated at the check-out machine when they leave.

The tracking of their customer behavior inside the store is not an easy task as it could be in a digital marketplace like Amazon or Ebay, where we can identify which path has the customer followed to reach the order point. However, diving into the business model, a key activity of the company which can benefit from the using of Big Data can be identified: supply chain management.

Walmart has a big network of stores all over the US, where different habits or ways of living are taking place. Trying to draft a common strategy for the supply of products for every store would be a mistake, as people do not behave identically in every place at the same time. For example, if a store is set up in a city where the temperature does not drop below 20ºC, the strategy regarding the supply of clothes would not be the same to that of a city where the temperature does not raise above the 10ºC.

The company was aware of this, and therefore, from 2004 they started to register in a database the transactions that were taking place in every store all over the country to analyze them in an aggregated way. As previously said, the point at where all this data was generated was the check-out machines, which were seen as a gate to gather information thanks to scanning of the bar codes of every product. In this way, Walmart obtained a real time data analytics of what items were being bought in any store at any time (Marr, B., 2016).

Soon, this recording of data was applied not only to the products side of the business, but in every single aspect of the company. Registering the stock received by the suppliers, their delivery time, the relation between certain prices to certain outcomes, the behavior of its workers…

This data driven strategy allowed them to do several things (Schmarzo, B., 2013):

- *Predict sales.* Thanks to the data registered, they could anticipate which products would be bought in every set of circumstances, to supply that specific store with the stock that it needed to face the demand. For example, studying the products that are most commonly bought on Christmas or even in the occasion when there is threat of a hurricane.
- *Create specific promotion strategies for every product.* As they have the information needed to know how every product sell depending on the promotion techniques needed, Walmart could draft individualized strategies for every item of the store (Davenport, T. H., et al., 2007).
- *Choose the best suppliers.* By tracking different variables from the products delivered by its suppliers, the company could identify which supplier has a better cost efficiency and quality combination.
- *Enjoy dynamic pricing.* Thanks to real-time data analytics, Walmart is able to track the prices of all of its products and make instant changes to readjust them to maximize profit.
- *Exert influence over their suppliers.* As the company has information over the products and the suppliers themselves, they can use all that data in the negotiation of deals for the supply of products, imposing the conditions that the company desires.

Quantitative Analysis[1]

In order to carry out Walmart's quantitative analysis, the focus will be put into indicators that show the efficiency of the company's supply chain, which is the cycle in which Walmart is able to provide its stores and clients with new products.

However, due to the difficulty to access sources of relevant companies' data, the period from 2016 to 2020 will be studied, even though aggregated data analysis was implemented in 2004 at the firm.

Nevertheless, as previously stated, Big Data study is more efficient the bigger the datasets, thus, any subsequent period of years studied, should give some insights about the improvement generated, irrespectively of the date of introduction of the technology.

The first indicator to analyze is the *days in inventory*, which measures the number of days that the merchandise spends in the inventory before it is sold. The lower the indicator the more times the whole inventory can be sold. In Table 1 it can be clearly seen that this metric has been steadily decreasing over the studied period, which means that Walmart has improved its supply chain management.

From this indicator the *inventory turnover* can obtained, which indicates how many times the company's inventory is sold and replaced over a period. In 2020, the firm managed to sold its entire inventory more than 9 times, which is an outstanding figure due to the scale of the company (Table1).

Both indicators can be linked to the benefits of Big Data in this area. Walmart use of this tool gives the company a better understanding of the customer needs and preferences, which in turn allows them to better forecast the sales and the demand they will face for a better supply planning. In turn, this leads to a heavier rotation of inventory and a higher number of general sales.

Moreover, by taking a look at the figures of *total inventories* in Table 1, it can be seen that Walmart's inventories value has been increasing over the years, but its inventory turnover has not decreased as it would be expected (bigger inventory, more time to fully sell it). The figure has increased which means that, even with the growing trend of Walmart's inventory, the company is managing to replace it more times over a period.

[1] Raw data taken from Walmart's financial reports, as well as gurufocus.com, page that collect relevant indicators and figures about companies, and limits the data to the most recent 5 years for free users. Net Income CAGR and Net Income Nominal Growth Rate of own elaboration.

This better supply chain management, mixed with other relevant factors which could influence the result, should be, in part, translated into income for the company. Even though it is true that Walmart's figure has been decreasing from 2016 to 2019, last year's *net income* has returned the best result of this 5-year period, even with the current pandemic situation.

Looking deeper into those 2016-2019 bad results, it can be seen that they are explained by Other losses included in the Income Statement that the company has had over these periods and which are not clearly specified. Therefore, it cannot be certainly known if they are attributable to the Big Data usage in the company.

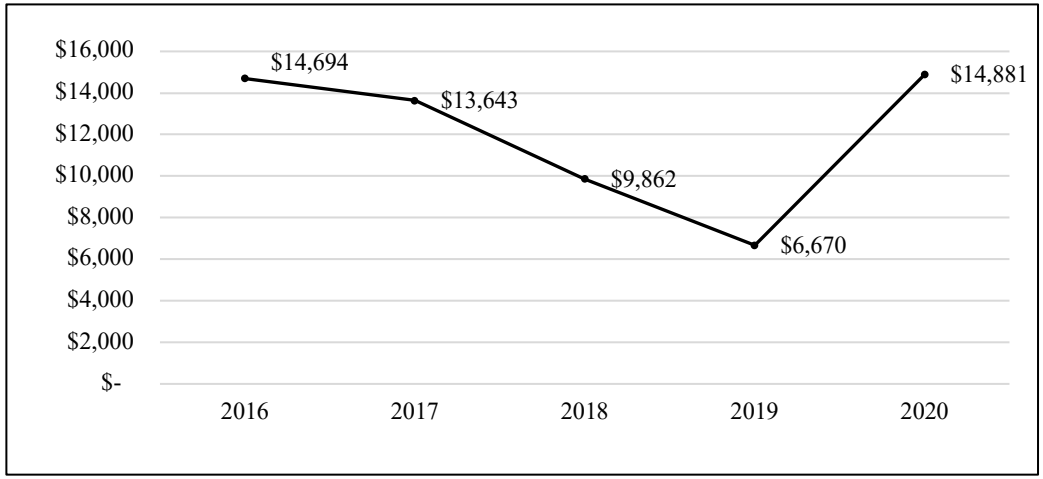
However, on the fiscal year 2020, these Other losses have turned into Gains, which explain the *nominal growth rate* from 2019 to 2020 of 123.1%. Nevertheless, as previously said, a direct proportional link between these findings and the reality of this technology cannot be defined, as more factors have to be taken into account.

Walmart is an example of a physical based company who was able to identify the advantages that data analysis could mean for them and turn it into one of its main sources of information for the decision-making of many parts of its company.

Table 1 Walmart's Data from 2016 to 2020[2]

	2016	2017	2018	2019	2020
Days in inventory	*44.21*	*42.44*	*41,71*	*41,02*	*38,81*
Inventory Turnover	*8.26*	*8.6*	*8.75*	*8.9*	*9.4*
Total Revenue (in millions of $)	*$ 482,130*	*$ 485,873*	*$ 500,343*	*$ 514,405*	*$ 523,964*
Cost of goods sold (in millions of $)	*$ 361,256*	*$ 373,396*	*$ 385,301*	*$ 394,605*	*$ 420,315*
Total Inventories (in millions of $)	*$ 43,046*	*$ 43,783*	*$ 44,269*	*$ 44,435*	*$ 44,949*
Net Income (in millions of $)	*$ 14,694*	*$ 13,643*	*$ 9,862*	*$ 6,670*	*$ 14,881*
Net Income CAGR		*-3.6%*	*-12.4%*	*-17.9%*	*0.3%*
Net income Nominal Growth Rate		*-7.2%*	*-27.7%*	*-32.4%*	*123.1%*

Graph 2 Net Income Walmart in thousands of $ (2016-2020) [2]



[2] Source: Walmart's financial reports and gurufocus.com. Tables and graphs of own elaboration.

5.3 Netflix Inc.

Netflix needs no introduction. The company is known worldwide for its on-demand model of series and films streaming platform and it is the perfect case of a business model who was able to adapt to the changes in the market before becoming obsolete.

The platform started as physical-based rental business where you could pick the movie you wanted to have sent home. This model was bound to be left behind with the convenience that technology could bring to users, yet not all the competitors seized the big shift that was coming their ways (Marr, B., 2016).

However, Netflix forecasted this change as an opportunity to grow and expand their company beyond the limitations that the traditional model was imposing on them. In this way, from early stages, the firm invested considerable amounts of money on innovation in order to develop the tools needed to launch the streaming platform service that we know them for today.

This forecast of the opportunities allowed Netflix to enjoy the so-called first mover advantage. Suarez, F. et al. defined the concept in their paper 2005 *The half-truth of first-mover advantage,* as the competitive advantage that a player enjoys when they enter into a whole new market where there are not any competitors. This anticipation enables the first mover to establish a foothold in this new niche which the followers cannot compete with.

The first mover advantage concept has been greatly questioned by scholars through the years, given the number of cases of first movers who got outperformed by second movers thanks to the knowledge they acquired from previous external experiences. However, it is obvious that since Netflix step into the streaming platforms market, the company has remained as the dominant player (Yahoo Finance, 2021, 5th of June, *Netflix: Recent Data Confirms Streaming Player's Dominance*).

Back to the Big Data perspective, Netflix proposed a model where the customers were not only able to pick what they wanted to watch, but also, they could get recommendations based on the titles they had already seen. This content recommendation is the key value proposition in streaming platforms. With the already preexisting information that Netflix had of its customers, which was useful even when it was a physical-based model, all that it was left to do is to set up an algorithm to profit from (Marr, B., 2016).

An algorithm is defined by the Cambridge dictionary as:

"… a set of mathematical instructions or rules, that, especially if given to a computer, will help to calculate an answer to a problem."

At the company they knew what this tool would could mean for the service. It would increase its product life cycle by bringing the focus of the user to new content when they finish what they were watching and it would keep them from unsubscribing from the platform.

This is why, in 2006 Netflix launched *the Netflix Prize,* a competition where programmers all over the world were challenged to create an algorithm for the company with a prize of 1 million $ for the winner. Up to this day this algorithm has remained unmatched by competitors (Marr, B., 2016).

Netflix enjoyed the competitive advantage that data could offer them in two ways:

- It granted them a differential value from its competitors, which at that time did not even have a streaming platform, let alone an algorithm of recommendation.
- It helped the service to retain customers and increase its product life cycle.

Nowadays, Big Data offers much more value to the company than the one that it offered back in its beginnings as a streaming platform. For example, when a user picks content to watch, the cover of that movie or series is not choose randomly. Each element has a set of different images that will show depending on the preferences and habits that the algorithm has detected from the customer. In this case Netflix uses Big Data as a marketing promotion strategy for its offering.

Another example of the way they use this tool is as a pool of ideas for content creation. The company has drifted from a distributor of visual media to a mix between a distributor and creator. Thanks to Big Data and the information they have regarding the preferences of their users, they can forecast what content the customers want to watch, producing that content from scratch (Marr, B., 2016).

Quantitative Analysis[3]

One of the key metrics to analyze the value that Big Data has offered to Netflix would be the Customer Lifetime Value, as the company's algorithm makes customers find endless constant content in its huge catalogue. Unfortunately, the company stopped offering detailed customer information, like cancelled subscriptions per year or new customers per year. Without this information it is impossible to calculate an accurate CLF.

However, a close general picture of the company's situation to see the improvement of the platform over the years could be obtained without this information. In the first place, the *number of users* will be analyzed. As it can be seen in Graph 4, this number has not stopped growing in the studied period. We can try to link this steady increase with the ability of the company to retain users thanks to the better targeting of preferences that the algorithm has been able to offer to them.

At the same time, this number of users can be linked with the *gross margin per user*. Thanks to the *revenue of the company*, the *cost of sales* and the *number of users* (Table 2) this metric can be obtained. As it can be seen in Graph 3, it presents an upwards tendency, except in 2012, which for reasons not related to Big Data, this margin sank and has been recovering ever since.

Even though a CLF cannot be obtained, performing a cross analysis of the evolution of customers, the gross margin per user over the years and the revenue per user every year, an accurate intuition of its uprising trend can be concluded. If the customers are increasing, the retention rate should be increasing accordingly, due to the nature of the service. At the same time as the margin per customer shows an upwards trend, then the Customer Lifetime Value should be increasing too.

Moreover, the *compound annual growth rate of the net income* between 2008 and 2020 (Table 2) has been of a 34%, which shows a positive image of the company as a whole throughout the years.

In this analysis one important milestone of the history of the company should be taken into consideration. In 2012, the company begin its international expansion to foreign markets. This expansion came with an aggressive marketing campaign which cause the

[3] Data taken from Netflix's financial reports, as well as macrotrends.net and businessofapps.com, pages that collect relevant indicators and figures about companies. Net Income CAGR, Cost per user, Revenue per user and Gross Margin per user, own elaboration.

company costs that had a relevant impact on the *operating income* and therefore on the *net income* of the company, explaining the dip from 2012 onwards.

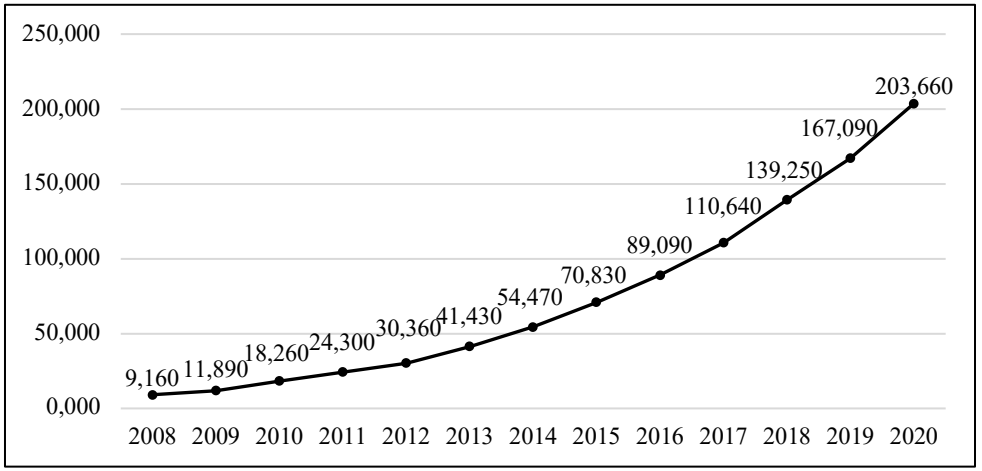
Even though a definitive quantitative analysis picture of Netflix usage of Big Data in this process cannot be obtained, it is obvious that the streaming platform is offering a value proposition that is attractive to its customers as it can be seen by the increase in the amount of them. This value proposition is, in a big proportion, the endless content that you will have at your disposition and how you will receive that content, and these areas are directly affected by their use of Big Data.

Table 2 Netflix's Data from 2008 to 2020[4]

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Revenues *(in million of $)*	$ 1,365	$ 1,670	$ 2,163	$ 3,205	$ 3,609	$ 4,375	$ 5,505	$ 6,780	$ 8,831	$ 11,693	$ 15,794	$ 20,156	$ 24,996
Nº of users *(in thousands)*	9,160	11,890	18,260	24,300	30,360	41,430	54,470	70,830	89,090	110,640	139,250	167,090	203,660
Cost of sales *(in million of $)*	$ 910	$ 1,079	$ 1,357	$ 2,040	$ 2,652	$ 3,117	$ 3,753	$ 4,591	$ 6,257	$ 8,033	$ 9,968	$ 12,440	$ 15,276
Cost per user	$ 99	$ 91	$ 74	$ 84	$ 87	$ 75	$ 69	$ 65	$ 70	$ 73	$ 72	$ 74	$ 75
Revenue per user	$ 149	$ 140	$ 118	$ 132	$ 119	$ 106	$ 101	$ 96	$ 99	$ 106	$ 113	$ 121	$ 123
Gross margin per user	$ 50	$ 50	$ 44	$ 48	$ 32	$ 30	$ 32	$ 31	$ 29	$ 33	$ 42	$ 46	$ 48
Net income *(in million of $)*	$ 83	$ 116	$ 161	$ 226	$ 17	$ 112	$ 267	$ 123	$ 187	$ 559	$ 1,211	$ 1,867	$ 2,761
Net Income CAGR		40%	39%	40%	-33%	6%	21%	6%	11%	24%	-34%	-29%	-25%

Graph 3 Evolution of Netflix's Gross Margin per user over the years[4]

Graph 4 Evolution of Netflix's users in thousands (2008-2020) [4]



[4] Source: Netflix's financial reports, macrotrends.net and businessofapps.com. Graphs and table of own elaboration.

5.4 Starbucks Corporation

The image of Starbucks is that of a coffee shop located in the center of the city, where people go with no other pretension but to enjoy a quiet time with its coffee or grab one to go.

The general perception is not of a very innovative company, apart from the way they prepare their products. This is why this company is the perfect example of a business that given the nature of its activity, we might suppose that it would not obtain value from data analytics, but finds in Big Data a great ally.

The limitations that Starbucks had to obtain data from its customers were obvious. There were no methods to register the transactions, apart from the cash machines located in every store, which would only give insights about the sales with no correlation to the customers themselves as they were not linked in any way (Marr, B., 2018).

However, this situation changed when the company launched its mobile app and its rewards program. Thanks to these new channels, the coffee giant was able put faces to a big part of the purchases that they were obtaining.

This data gathered from every customer was used to give multi-store joint experience. No matter to which Starbucks store you go, if you are subscribed to their reward program, the barista will know your order of choice (Marr, B., 2018).

Moreover, thanks to the mobile app, the company can offer brand new recommendations based on your previous orders. This history of orders is also used in targeted marketing, where Starbucks sends personalized offers and ads that are relevant to the products that the user usually consumes.

These applications of Big Data are not really different to what other companies do. Once the problem about how obtaining the data was solved, the strategies to follow regarding its customers were endless. However, Starbucks do not only use Big Data for making decisions about its users, it also uses this tool to plan its real estate strategy (Marr, B., 2018).

In 2008 the company found itself in a delicate position. There were a significant number of stores across the globe that were underperforming, which in turn caused millions of dollars in losses to the business. The economic crisis pushed customers to find more affordable coffee alternatives to the ones offered in the company.

As a result of this situation 600 coffee shops were closed (although this figure is not reflected on Graph 5, as there were stores that opened too), leaving thousands of workers unemployed (Herman, C., 2008). At that moment, Starbucks realized that if they wanted to recover from that crisis, they should be very careful about the location of new stores, in order to be able to obtain the maximum income from each of them.

To carry out this task, a solution was found on the platform Atlas, developed by the company Esri. This tool used thousands of different sources of data to inform about the conditions of a determined area in a map. Thanks to it, Starbucks was able to obtain dozens of variables about a potential store location (from traffic, demographics, preferences…) and also determine how the opening up of a new point of sales was going to affect its own stores in the area (Gordon, K., 2017).

This is the reason why in a given city, there can be 2 Starbucks stores and in other cities, in the area of 2 kms, 5 stores can be located close to each other. Atlas enabled them to know which exact points could suppose a potential source of clients without competing directly with its own business.

Quantitative Analysis[5]

There are a several Big Data's aspects which could be considered to study in this specific case. However, the focus will be to analyze the situation that took place in 2008, when a considerable number of stores that were underperforming, were closed down and how the company recover from that situation thanks to Big Data.

In Graph 5 it can be seen that, in fact, the growth of *number of Starbuck's stores* halted in 2008 to slightly descend in 2009, once the underperforming points were closed down. This decrease is not as intense as we could imagine, as there were stores that opened, which offsets that decrease, but coming from years when the growth had values around 10% - 20% this figure is significant.

Of course, according to the decrease in the number of stores, the decrease in *the number of employees* is sharper as Graph 6 proves. However, the most worrying figure in this

[5] Data taken from Starbucks' financial reports, as well as marcotrends.com, stock-analysis-on.net and knoema.com, pages that collect relevant indicators and figures about companies. Income per stores and increase in stores, own elaboration.

case is the *income*, which fell dramatically from 673 million of $ in 2007 to 316 million of $ in 2008 which is the moment where the number of stores peaked, and the moment that made evident the company's problem.

We have no reliable source of information of those underperforming stores, but from the data a general picture of the situation at that time can be obtained.

In order to do this, a relation of the income of the company, as a whole, to the number of stores was performed. As a result, the variable obtained, *income per store*, shows concerning results in 2008, with a dip of a 58%, which the company took time to recover from.

However, as previously said, this event made the company realize the need to implement a better decision-making tool for its real estate planning, and they found that ally in Big Data. From this point onwards, the number of stores increases simultaneously to the amount of income. In fact, the income increases at a faster pace than the number stores as the income per store gradually increases year by year.

It is important to remark two points in the timeline of this analysis:

- In 2013 Starbucks had to pay 2.75 billion $ to Kraft due to a lawsuit over the distribution of Starbucks's coffee products in grocery stores. (Strom, S., 2013)
- In 2018, due to the acquisition of their East China joint venture (which entailed a gain of 1.4 billion $ not subject to taxes), the income skyrocketed instead of growing with the trend, which explains the decrease from 2018 to 2019 as it was an extraordinary earning (Starbucks's financial statements).
- In 2020 the upwards trendline is broken due to the COVID-19 world pandemic, which obviously affected a physical-based company.

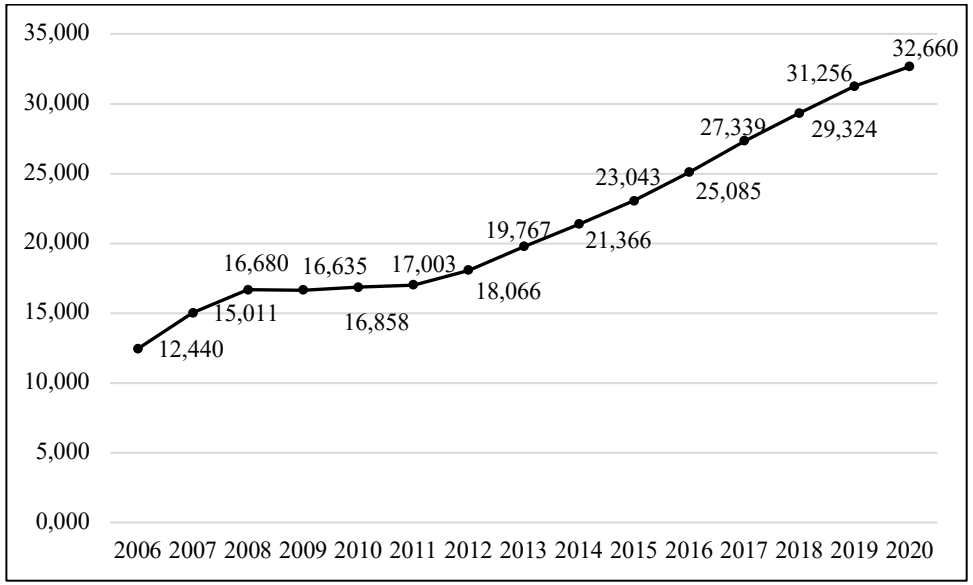
Taking those considerations into account, it is evident that the new strategy that Starbucks took regarding its stores has been effective by the good health of the income per store indicator.

Furthermore, as previously said early in the case, Starbucks does not only use Big Data to address its store's location, it also launched customers program's which might have had a positive effect on the revenue and thus, in the net income. However, the key matter in these results is that, as it can be seen, the inefficiency presented was solved thanks to the introduction of a data analysis tool in the decision-making of the company.

Table 3 Starbucks' Data from 2006 to 2020[6]

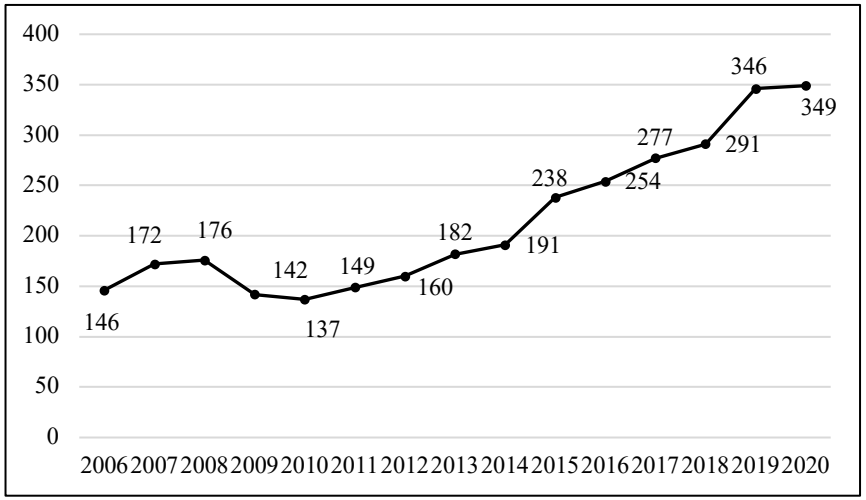
	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Number of stores	*12,440*	*15,011*	*16,680*	*16,635*	*16,858*	*17,003*	*18,066*	*19,767*	*21,366*	*23,043*	*25,085*	*27,339*	*29,324*	*31,256*	*32,660*
Return on Assets (ROA)	*12.74%*	*12.59%*	*5.56%*	*7.01%*	*14.81%*	*16.92%*	*16.84%*	*0.07%*	*19.23%*	*22.15%*	*19.66%*	*20.08%*	*18.7%*	*18.73%*	*3.16%*
Nº of employees *(in thousands)*	*146*	*172*	*176*	*142*	*137*	*149*	*160*	*182*	*191*	*238*	*254*	*277*	*291*	*346*	*349*
Income *(in millions of $)*	*$ 564*	*$ 673*	*$ 316*	*$ 391*	*$ 946*	*$ 1,246*	*$ 1,384*	*$ 8*	*$ 2,068*	*$ 2,757*	*$ 2,818*	*$ 2,885*	*$ 4,518*	*$ 3,599*	*$ 928*
Income per store	*$ 45,338*	*$ 44,834*	*$ 18,945*	*$ 23,505*	*$ 56,116*	*$ 73,281*	*$ 76,608*	*$ 405*	*$ 96,789*	*$ 119,646*	*$ 112,338*	*$ 105,527*	*$ 154,072*	*$ 115,146*	*$ 28,414*
Increase stores		*21%*	*11%*	*0%*	*1%*	*1%*	*6%*	*9%*	*8%*	*8%*	*9%*	*9%*	*7%*	*7%*	*4%*

Graph 5 Evolution of nº of Starbucks' stores over the years[6]

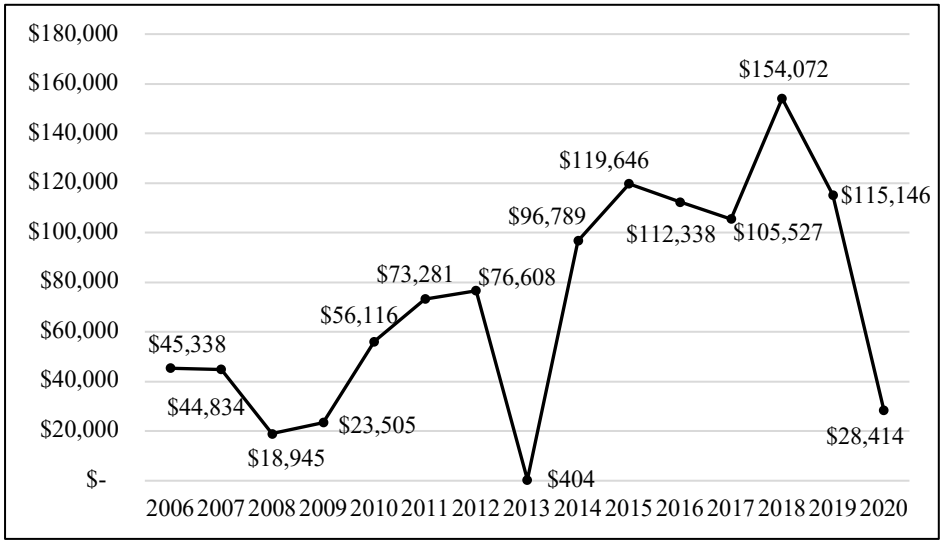


[6] Source: Starbuck's financial reports, macrotends.com, stock-analysis-on.net, knoema.com and graphs and table of own elaboration.

Graph 7 Evolution of n° of Starbucks' employees over the years (in thousands) [6]



Graph 6 Evolution of Starbucks' income per store over the years[6]



6. CONCLUSIONS

These previous analyses of cases are of own elaboration, taken as a starting point literature from other scholars, who have studied from a theoretical perspective the cases referenced. In this way, taking into account their findings, relevant parameters have been identified and the information necessary has been collected and studied in order to reach these conclusions.

Even though these three companies show positive results on the areas affected by the Big Data applications, it has to be noted that the limitations of measuring the impact of this technology on enterprises are more than evident. After all, these solutions are applied to areas of the businesses which are influenced by other factors, and the amount of importance of each of them is difficult to establish.

In fact, this is one of the main lines of work of study for scholars in the area of analytical solutions applied to enterprises. Obtaining a relevant framework in order to be able to fully comprehend the possibilities that Big Data can offer, not only in a theoretical way, but also in a practical way while taking into account all the factors that can influence the results of the company. That is exactly the aim of this paper, trying to add and propose a different approach to the growing body of works on this area.

Of course, this analysis is made from an external perspective of the enterprise, and thus, is subject to the lack of relevant private information that would have been useful, as well as internal indicators which would show a more accurate picture of each situation.

It is important to note, that the contribution of this work is an initial approach, which could be considered as the first steps towards a more complete analysis. It opens up new future lines of work on this field, to improve the methods used and evolve into a more accurate study of the impact of Big Data in companies.

Nevertheless, what is clear is that, overall, the perception of many scholars is that technologies are shaping to be one of the main keys in order to remain competitive in the market and to offer companies a continuity in their activities. This seems obvious, because the business world follows the same trendline as the society in itself, and both are drifting towards a general digitalization of processes.

This digitalization of processes offers new possibilities to firms, who are constant generators of data, to be able to look at their situation in a more aggregated way, instead of using an individualized approach. In this way, trends, patterns and habits can be

identified, and actions can be implemented to improve inefficiencies, which is exactly the use of Big Data for enterprises.

Furthermore, even though it is an advanced technology, its applications are not only limited to digital based companies. As it has been seen, companies which operate in the physical world, can also benefit from all the advantages that it offers. Nonetheless, it is true that part of its activity has to have connecting points with the digital world in order to generate the inputs for this tool.

However, there is still the question of whether these projects are cost efficient or not, given that the amount of monetary resources in order to make them work are really high in comparison with other regular projects. But as the technology settles in, this relation will become clearer and more obvious, as the number of cases grow and the results will be more conclusive.

What it is clear is that technology is moving to the center of the everyday life of the society. This new opening of opportunities, is offering a lot of possibilities to companies, who are still reluctant of leaving their comfort zone as they are used to their usual ways of doing things and scared of taking on new directions. Their attitude towards these new changes will be the determining factor which will decide, whether they keep growing and moving forward with the pass of the time, or they stagnate in a past, that is bound to be left behind.

In this way, Big Data becomes the perfect ally for organizations who are ready to take the challenge that the Industry 4.0 suppose. Thanks to this technology, firms will be able to obtain the most valuable resource they could have: knowledge, which will give them a more accurate perspective of their situation and the situation of the external environment of the company, as well as a better control over their operations, enabling them to trace appropriate strategies.

In a world where everyone lives in two different realities at the same time, the physical and the digital one, ignoring one of them would suppose losing valuable information and opportunities for the company. The only way to stay relevant, is to stay present in both levels, and by doing so not only the firm is able to survive, but also, it is able to enhance.

BIBLIOGRAPHY

Almeida, F., & Calistru, C. (2013). The main challenges and issues of big data management. *International Journal of Research Studies in Computing*, *2*(1), 11-20.

Anuradha, J. (2015). A brief introduction on Big Data 5Vs characteristics and Hadoop technology. *Procedia computer science*, 48, 319-324.

Ashton, K. (2009). That 'internet of things' thing. *RFID journal,* 22(7), 97-114.

Barham, H. (2017, July). Achieving competitive advantage through Big Data: a literature review. In 2017 *Portland international conference on management of engineering and technology (PICMET)* (pp. 1-7). IEEE.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of management,* 17(1), 99-120.

Bartosik-Purgat, M., & Ratajczak-Mrożek, M. (2018). Big data analysis as a source of companies' competitive advantage: A review. *Entrepreneurial Business and Economics Review*, *6*(4), 197.

Christensen, C. and Raynor, M. (2003). *The Innovator's Solution: Creating and Sustaining Successful Growth*. Boston, MA: Harvard Business School Press.

Christensen, C. M. (1997). *The innovator's dilemma: when new technologies cause great firms to fail.* Harvard Business Review Press.

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Boston, Mass: Harvard Business School Press.

De Mauro, A., Greco, M., & Grimaldi, M. (2016). *A formal definition of Big Data based on its essential features. Library Review.*

Dimitrov, D. V. (2016). Medical internet of things and big data in healthcare. *Healthcare informatics research*, *22*(3), 156.

Fávero, L. P., & Belfiore, P. (2019). *Data science for business and decision making*. Academic Press.

Goddard, M. (2017). The EU General Data Protection Regulation (GDPR): European regulation that has a global impact. *International Journal of Market Research*, *59*(6), 703-705.

Gordon, K. (2017). Business site selection, location analysis, and gis.

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic management journal*, 17(S2), 109-122.

Grover, V., Chiang, R. H., Liang, T. P., & Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems*, *35*(2), 388-423.

Guru Focus. (2021). Guru Focus. *https://www.gurufocus.com/*

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., ... & Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management*, *36*(5), 748-758.

Hassani, H., Huang, X., & Silva, E. (2018). Banking with blockchain-ed big data. *Journal of Management Analytics*, 5(4), 256-275.

Herman, C. (2008, 2 julio). *Coffee Crisis? Starbucks Closing 600 Stores*. ABC News.

Knoema. (2021). Knoema. *https://knoema.com*

Kostopoulos, K. C., Spanos, Y. E., & Prastacos, G. P. (2002, May). The resource-based view of the firm and innovation: identification of critical linkages. *In The 2nd European Academy of Management Conference* (pp. 1-19). Stockholm, Sweden: EURAM.

Lasi, H., Fettke, P., Kemper, H. G., Feld, T., & Hoffmann, M. (2014). Industry 4.0. *Business & information systems engineering*, 6(4), 239-242.

Lukić, J. (2017). The impact of big data technologies on competitive advantage of companies. *FACTA UNIVERSITATIS-Economics and Organization*, *14*(3), 255-264.

Macrotrends. (2021). Macrotrends. *https://www.macrotrends.net*

Marr, B. (2016). *Big data in practice: how 45 successful companies used big data analytics to deliver extraordinary results*. John Wiley & Sons.

Marr, B. (2018). *Starbucks: Using Big Data, Analytics And Artificial Intelligence To Boost Performance.* Bernard Marr & Co.

Mowery, D. C. (2009). Plus ca change: Industrial R&D in the "third industrial revolution". *Industrial and corporate change*, *18*(1), 1-50.

Netflix Revenue and Usage Statistics (2021). (2021, 9 marzo). Business of Apps. *https://www.businessofapps.com/data/netflix-statistics/*

Netflix: Recent Data Confirms Streaming Player's Dominance. (2021, 5 junio). Yahoo Finance.

Netflix. (2020). *FY20 Netflix's Annual Report.*

Pejić Bach, M., Bertoncel, T., Meško, M., Suša Vugec, D., & Ivančić, L. (2020). *Big data usage in European countries: cluster analysis approach.* Data, 5(1), 25.

Pozdnyakova, U. A., Golikov, V. V., Peters, I. A., & Morozova, I. A. (2019). Genesis of the revolutionary transition to industry 4.0 in the 21st century and overview of previous industrial revolutions. In *Industry 4.0: Industrial Revolution of the 21st Century* (pp. 11-19). Springer, Cham

Puaschunder, J. M. (2019). Big data ethics. *Journal of Applied Research in the Digital Economy*, *1*, 55-75.

Sanders, N. R. (2016). How to use big data to drive your supply chain. *California Management Review*, *58*(3), 26-48.

Schmarzo, B. (2013). *Big Data: Understanding how data powers big business*. John Wiley & Sons.

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). Analytics: the real-world use of big data: How innovative enterprises extract value from uncertain data, Executive Report. *IBM Institute for Business Value and Said Business School at the University of Oxford*.

Starbucks Corporation. (2020). *FY20 Starbuck's Annual Report*.

Stephenson, D. (2018). Big Data Demystified: *How to use big data, data science and AI to make better business decisions and gain competitive advantage.* Pearson UK.

Stock Analysis on Net. (2021). Stock Analysis on Net. *https://www.stock-analysis-on.net*

Strom, S. (2013, 14 noviembre). *Starbucks to Pay Kraft $2.75 Billion, Ending Broken-Deal Dispute*. The New York Times.

Suarez, F., & Lanzolla, G. (2005). The half-truth of first-mover advantage. *Harvard business review.*

Van Der Aalst, W. (2016). Data science in action. In *Process mining* (pp. 3-23). Springer, Berlin, Heidelberg.

Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management.

Walmart. (2020). *FY20 Walmart's Annual Report.*

Zwitter, A. (2014). Big data ethics. *Big Data & Society*, *1*(2), 2053951714559253.

ANNEX

Table 4 Usage of Big Data in Europe by country and size of the company. Pejić Bach, et al. (2020)

	Large Companies	Medium Companies	Small Companies
Europe	33.5%	19%	10.5%
Group 1 (Finland, Ireland, The Netherlands)	50%	30%	20%
Group 2 (Germany, Norway)	35%	20%	12%
Group 3 (Greece, Italy, Spain)	27%	14%	7%