UNIVERSITÀ
DI PAVIA

Department of Economics & Management

Master's degree in International Business and Entrepreneurship

FROM DATA REPORTING TO PREDICTIVE INSIGHTS: THE INFLUENCE OF BIG DATA ANALYTICS ON STRATEGIC DECISONS

Advisor:

Chiar.mo Prof. Valerio Veglio

Master Thesis of:

Simone Berganti

Matr. n. 512802

Anno Accademico 2024-25

INDEX

1. INTRODUCTION

Data is the new oil, indeed the information economy is powered by data, just as the industrial economy was fuelled by oil (Hirsch, 2014). Data analysis has become essential for companies looking to enhance performance management, accelerate decision-making processes, and develop innovative business models. In the contemporary era, we find ourselves standing at the beginning of a new age, The Age of Analytics. This age is characterized by an astonishing data explosion, with the volume of information doubling every three years (Henke et al., 2016). This boom is driven by the proliferation of digital platforms, wireless sensors, virtual reality applications and billions of mobile phones. At the same time, the storage capacity of data has grown exponentially and the cost of data has decreased; as a result, data has become more accessible than ever for everyone. But this data-driven world is one of opportunities and challenges. On the one hand, it can cause the disruption of many traditional business models and the transformation of entire industries; in fact, leading companies are already leveraging their analytic capabilities to not only make improvements to their core businesses, but also to introduce entirely new business models. On the other hand, organizations are experiencing great difficulty in adapting to this new "paradigm". Many have heavily invested in technology, but they are struggling to change their organisational structures and cultures to take full advantage of these investments. The journey towards a data-driven world is full of challenges, but also enormous promises. As we navigate the Age of Analytics, it is critical to understand how we can effectively compete in this new landscape and leverage the power of data to drive innovation and growth. Today we live in an extremely dynamic world, where everything runs faster, nothing is sure or taken for granted and everything can change from one day to the next, especially in business; this makes it impossible and highly risky to base important decisions on mere intuition, only through data it is possible to truly understand, verify and quantify. This is why the adoption of data-driven decision making, i.e. using data to inform decision-making and validate a series of actions before taking it, becomes of paramount importance (Brynjolfsson & McElheran, 2016). A PWC's study of over 1,000 executives, revealed that organizations that are highly data-driven are three times more likely to achieve significant improvements in decision-making than those that don't (Stobierski, 2019). Despite these findings, which show clear evidence of the supremacy and competitive advantage of those companies that base their decisions and strategies on data analysis, many others still remain reluctant to adopt data-driven decision making. The PWC study also found that a consistent portion of the executives interviewed argue that important decisions are best made on the basis of experience and intuition, highlighting how, despite the enthusiasm for the potential use of data and analytics, there is still a certain underlying scepticism towards their application for important executive-level decisions (PWC, 2015). At the same time, other executives not only understand its potential, but proceed to apply data in decision-making, which turns out to be a determining factor in the company's growth path.

Furthermore in the last decades, the development of data has revolutionised the way companies operate and make decisions.

The analysis, by companies, of the data collected, for many years was mainly done for descriptive purposes, such as to monitor previous business performance, and to understand consumer behaviour through retrospective analysis (Davenport, 2016). However, with the continuous increase in data generated and at the disposal of companies, made possible by the spread of new technologies and digitization, data analysis has expanded its purpose. In fact, today data are not only used to describe the status quo, but have become a key tool for making predictions and anticipating the future. Through machine learning techniques and the application of artificial intelligence, companies can identify patterns and trends hidden within their data, thus they can, for example, anticipate consumer needs, making strategic decisions earlier and with greater accuracy (Brynjolfsson et al., 2021).

This shift from descriptive to predictive analysis represents one of the most significant transformations in modern business, deeply affecting business competitiveness and innovation. The ability of companies to collect, analyse, evaluate and make the best use of the vast amount of data at their disposal is therefore a factor that really determines their success, growth or decline. The aim of my research is to understand how the application and use of big data can influence the performance of companies in an environment where data is increasingly characterised by increased velocity, variety and volume.

2. RESEARCH OBJECTIVE

This thesis seeks to explore how a company can maximize the benefits from the vast amounts of information available to improve its decision-making strategies, optimize operational processes, and perform well in light of a changing business environment. Initially, I will focus on an overview of the use of data in business over time, particularly those times when it was used primarily for descriptive analysis. Indeed, in the past, companies have relied on historical data to assess past performance, identify trends, and create reports. However, this approach has been mostly retrospective; it focuses on what has already happened, allows you to react to events only after they have occurred, and therefore is more reactive than proactive. Although knowledge of the past is critical, in order to understand the mistakes that have been made, and the strategies that instead have been particularly proficient, this method has limitations, especially in the most dynamic markets, where the ability to anticipate change is often critical to achieving success.

With the growth of technologies and the exponential increase in the volume of data from a wide variety of sources, such as social media, IoT devices, and digital transactions, companies in today's world must not only manage huge volumes of data, but also translate and interpret useful information from them to predict future trends and guide their strategic decisions. For this reason, a winning approach today is not limited to descriptive analysis, but requires more advanced tools such as predictive and prescriptive analytics. New methodologies enable companies not only to see what has happened, but also to estimate what will happen in order to make targeted decisions in advance.

This study will examine ways in which big data analytics can be used by companies to move from reactivity to proactivity. Indeed, companies can analyse large volumes of data in real time, identify emerging patterns, and predict customer behaviour to optimize operations and respond more quickly to changes in their respective markets.

This study also aims to analyse how the adoption of Big Data Analytics improves business performance, ensuring competitive advantage and long-term business sustainability using examples from different industries.

In conclusion, the thesis aims to show how a company's ability to manage and integrate big data into its processes can be a key factor for innovation, efficiency and growth. To support these statements, examples will be presented of companies that have been able to leverage data strategically, turning it into a key factor for success. These illustrative examples will highlight examples of how big data has been exploited by companies that have completely revolutionized their respective industries, and that for this reason are considered disruptors, becoming an inspiration for all other companies that aspire to maximize the benefits of big data in the era of analytics.

This research turns out to be relevant since, the concept of big data and big data analytics, as much as it is a hot topic, is still quite recent, and in-depth investigations that focus on the various benefits generated by their application are not as widespread.

While there are some studies that specify how BDAs can be applied in business, for what activities, it is also true that many of them do not focus on the practical impacts of applying these techniques and how they influence strategic decisions; for this reason, the research question is: how can the application and use of big data in business processes influence relevant decisions and business performance?

Thesis structure:

After introducing the research topic and identifying the question on which it is based, the thesis evolves with the following structure: Chapter 3 provides the heoretical background and explores the existing literature on the topic of data use in the enterprise, going through its evolution, up to the explosion of big data and all the opportunities ad derived from it. Chapter 4 outlines the Research methodology. This chapter describes the methodology applied in gathering the findings presented in the next chapter, adopting a qualitative methodology based on case studies and illustrative examples. Chapter 5 presents the main findings that have emerged from the research, illustrating companies that have distinguished themselves in their sectors for their use of BDA and in-depth case studies on Toyota Motor Europe. Finally Chapter 6 draws conclusion by summarizing the main academic and managerial implications arising from this research

3. THEORETICAL BACKGROUND

3.1. Evolution Of Data In Business Practices

Over the last century, the use of data in business has been transformed from rudimentary data recording to sophisticated big data analysis; technology advancements and the digital revolution have exponentially increased the volume, variety and velocity of data available to companies, creating new opportunities for data-driven decision-making and strategic innovation. At the origins of data usage in business, companies began using data for manual recording and basic statistical analysis. At this stage, data were often collected in handwritten and paper records to track transactions, manage inventory and maintain financial accounts. These procedures required a lot of time and effort but were necessary in order to monitor business performance and ensure accuracy, however, the risk of making mistakes was particularly high as many steps were done manually. Pioneers such as Henry Ford have demonstrated how data-driven methodologies can and should be applied to optimize processes and improve the quality of operations; in fact, one of the earliest examples of data-driven optimizations involves the implementation of the Ford assembly line. (Velu, 2021).

The proliferation of the Internet and the digital revolution have undoubtedly contributed to transforming “how” companies use the data available to them; but not only the “how” also the “why” in that as said earlier, they no longer just analyse data to understand the past and react, but now want to have a reactive approach and predict future changes. The increasing number of digital technologies and easier and easier access to the Internet have brought to an explosion in the volume of data generated. At the same time, this period has seen the emergence of data warehousing and data mining techniques that allow companies to store much more data than in the past and extract from it highly valuable information for their strategies. Data analysis has therefore become a key factor in gaining competitive advantage, as companies can analyse consumer behaviour, market trends, and operational performance with unprecedented accuracy (Steve LaValle, 2010).

As entering the 21st century, the concept of big data has begun to take shape. Big data refers to extremely large, complex data sets that are difficult to process with traditional data management tools. The term ‘big data’ is often associated with the so-called three

Vs - Volume, Variety and Velocity - which highlight the immense scale, diverse types and rapid generation of data in the modern era. Advances in cloud computing, machine learning and artificial intelligence have enabled companies to harness more and more the power of big data. Predictive analytics, real-time data processing and advanced data visualisation techniques are now commonplace, enabling companies to make data-driven decisions with greater accuracy and speed.

Nowadays, Data is a too important strategic resource; it is a key tool in fact for generating innovation and growth. Companies, regardless of industry, use data to bring new products and services to market, optimize the supply chain, enhance and personalize the consumer experience, improve the accuracy of their operations.

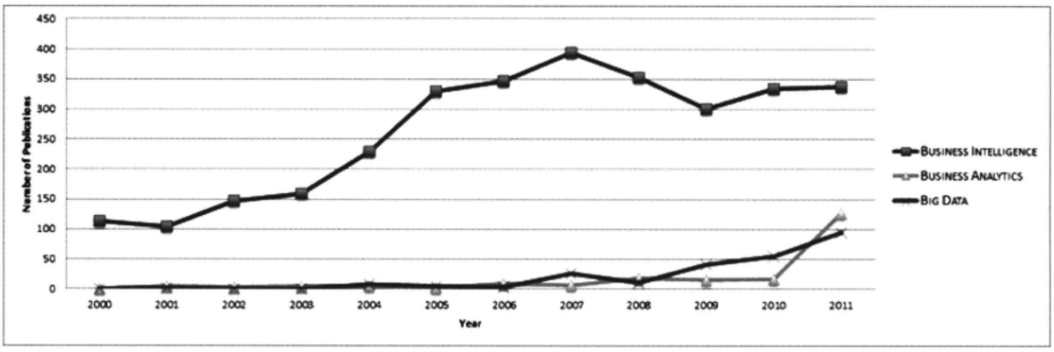
Thus, integration of data analysis into business strategies gave rise to such a phenomenon as data-driven organizations where data is put at the heart of every aspect comprising decision-making and strategic planning.

3.2. Data Concepts And Definitions

The necessity of adopting a decision-making process based on objective facts and not mere intuition, combined with the constant and accelerating technological progress over the last century, have made the ability to integrate data as a tool to support business strategies a key resource (Grover et al., 2018) . The relevance of this asset, has led to the development of numerous techniques and technologies to support the analysis of data and translate it into valuable information and actions; numerous terms have been used to describe the set of such techniques; in particular three of them have gained popularity over the years: Business Intelligence (BI), Business analytics (BA) and Big Data (BD) (Mashingaidze & Backhouse, 2017).

The growing importance of these topics was demonstrated by (Chen et al., 2012) who analysed the trend of publications between the early 2000s and 2011 containing the keywords ‘Business Intelligence’, ‘Business Analytics’ and ‘Big Data’ in the title or abstract.

Figure 1. BI,BA & BD publications trend from 2000 to 2011 (Chen et al., 2012)



Keyword	All Years	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Business Intelligence	**3,146**	113	104	146	159	229	330	346	394	352	201	334	338
Business Analytics	**213**	0	5	43	4	5	2	9	6	19	16	17	126
Big Data	**243**	0	1	0	0	7	4	3	26	11	41	44	95
Total	**3,602**	**113**	**110**	**149**	**163**	**241**	**336**	**358**	**426**	**382**	**358**	**356**	**560**

From the figure 1, it can be seen that ‘business intelligence’ has ‘older’ origins, as it is a term on which several articles were published even before the 2000s, but which saw a sharp rise at the beginning of the 21st century; ‘business analytics’ and ‘big data’, on the other hand, have a more recent origin, with the first articles appearing during 2001 but only reaching a certain level of importance a few years later. Although there is growing interest and research on Business Intelligence (BI), Business Analytics (BA) and Big Data (BD), these terms lack universally accepted definitions in academic and professional literature. Existing definitions have generated confusion as to their meaning and uncertainty as to how these terms are related. (Côrte-Real et al., 2014) consider Business Analytics and Business intelligence one singles thing, in fact they use the term BI&A (Business intelligence & analytics) “born from the success of Business Intelligence (BI) in the 1990s and the introduction of Business Analytics (BA) in the 2000 as a key data analysis element in BI”. Many other professional and academics tend to consider these three terms as interchangeable, without highlighting any particular differences (Bayrak, 2015).

In order to avoid confusion and delimit the boundaries and relationships between these three particularly popular and widely used terms, Mashingaidze & Backhouse, 2017, carried out a literature review, analysing different points of view from both the academic and professional worlds; their aim was to identify, by analysing the various sources and, by identifying the commonalities between the different articles, an univocal definition for each term and the relationships between them. As a result of this literature review Business Intelligence can be defined cohesive framework of strategies, applications, technologies, architectures, processes and methodologies designed to collect, store, retrieve and analyse data, thereby facilitating informed decision-making; the difference between Business Intelligence and Business Analytics would appear to be very thin, in fact the latter can be defined as a combination of skills, applications, technologies, architectures, processes and methodologies aimed at collecting, storing and retrieving data for analysis to support decision-making, strategic business planning and performance improvement. These analyses can be descriptive, predictive or prescriptive and use techniques from scientific fields such as mathematics and statistics. This definition closely resembles that of Business Intelligence (BI); however, Business Analytics (BA) is distinguished by the use of mathematical and statistical techniques and the classification of analysis into descriptive, predictive and prescriptive (Williams, 2016). Finally the definition of big data as data characterised by large volume, multiple sources and rapid generation and analysis. Its size and complexity exceed the capabilities of conventional technologies, necessitating advanced technologies and techniques for effective storage and analysis. After identifying "universal" definitions for each terms, Mashingaidze & Backhouse pointed out the relation among these three: "BD is a type of data that is used in advanced BI. BA is a component of BI. Thus, BD can also be used as the data source for BA."

From a historical perspective, the concept of big data is not entirely new. In the 1990s, the main motivation behind the creation of data warehouses was the need to store large amounts of information. At that time, reaching a terabyte of data was considered a significant milestone in data management (Watson, 2014).

But, once again, what exactly is meant by big data? One way to understand big data is to consider it as data that is too large and diverse to be effectively managed by traditional relational database management systems (RDBMSs). Some might define big data as any data set that exceeds 10 terabytes, but this numerical threshold is fluid and is likely to increase as organisations continue to collect, store and analyse increasing volumes of data (Russom, 2011).

Another useful perspective is to define big data according to its main characteristics: high volume, high velocity and high variety-often referred to as the 'three Vs.':

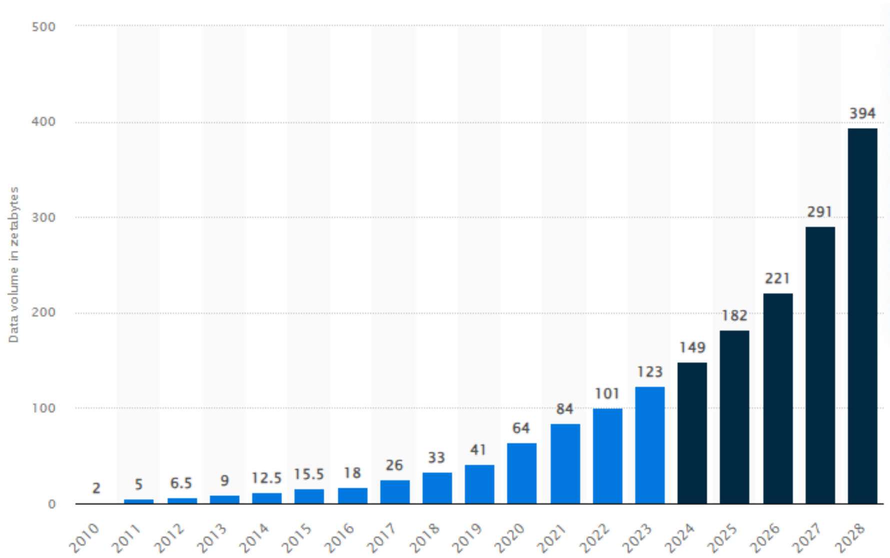
- High volume refers to the amount of data that is produced and stored.
- High speed describes the speed with which this data is generated and has to be processed.
- High variety indicates the wide range of data types, including structured, unstructured and semi-structured data.

In essence, ‘big data’ means that there is more data, it is generated at a faster pace and comes in a multitude of forms.

These perspectives are reflected in a widely accepted definition: big data refers to data characterised by high volume, high velocity and high variety, which require new technologies and techniques to be captured, stored and analysed. The goal of managing such data is to improve decision-making, foster insights and discoveries, and support and optimise various processes. (Glick, 2014)

In 2020, the digital universe is estimated to contain about 44 zettabytes of data. To put this number in perspective, a single zettabyte corresponds to about one trillion gigabytes. The amount and variety of available datasets has increased rapidly due to the collection of information from mobile devices, numerous sensors, aerial surveys (remote sensing), software logs, cameras, microphones, radio frequency identification (RFID) readers and wireless sensor networks. The global volume of data was predicted to grow exponentially from 4.4 zettabytes to 44 zettabytes between 2013 and 2020, underlining the vastness of Big Data and its continued growth in the digital age. (Taylor, 2023)

Figure 2. Amount of data created, consumed, and stored 2010-2023, forecast to 2028



3.3. Data Sources And Structures

As already presented, one of the key feature that characterizes big data and makes them particularly appealing is the Variety of sources these data come from.

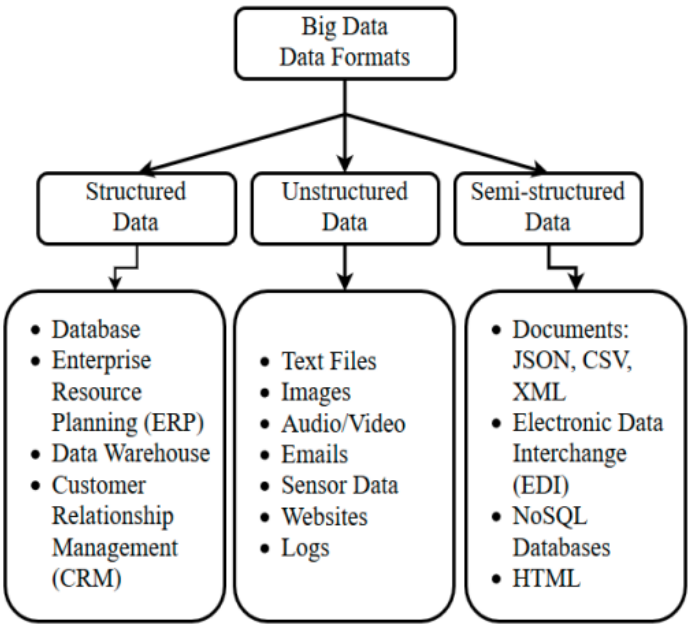
Big data contains relevant information coming from sources that go forward the traditional ones. Unlike traditional solutions, Big Data can handle large volumes of various types of data, especially unstructured data, at a much higher speed and with much lower resource consumption. One of the main advantages of this technology is the possibility of using advanced analytical techniques, such as predictive analysis, to operate on massive data and build high-quality predictive models to support decision-making. (Alonso et al., 2017)

For example, web log files record every mouse click, such data can be analysed to understand the customer buying habits and influence the customers' choices by presenting dynamic product recommendations. Social media platforms like Facebook and Twitter generate millions of comments and tweets that can be analysed to understand what people are thinking, for instance, about new-product launches (Mohankumar, 2013). In addition, devices such as smart meters produce continuous data on electricity, water or gas consumption. This data can be shared with customers and integrated with tariff plans to incentivise a change of energy consumption to off-peak times, such as postponing the washing of clothes.

Data can be divided into three main categories: structured, semi-structured and unstructured data. Structured data is highly organised and easily searchable, typically stored in relational databases or spreadsheets, where information is arranged in rows and columns with defined fields, such as customer names and transaction amounts. A sales database that stores customer records and their purchase history is one example of structured data. On the other hand, semi-structured data does not conform to any rigid schema but contains labels or markers that provide some level of organization to make them partially interpretable by machines. Common examples are the XML and JSON files, where data elements are labelled but not bound to a rigid format, thus allowing some flexibility in data representation. Finally, unstructured data do not have any predefined format, making analysis with traditional methods difficult. This type of data includes a wide range of formats, such as text documents, social media posts, images and videos.

For example, a collection of product reviews from different online platforms represents unstructured data, which requires advanced techniques such as Natural Language Processing (NLP) to extract meaningful information. (Blazquez & Domenech, 2018).
About 95% of big data is semi structured or unstructured, which underlines the importance of analytics in it and, hence, the importance of making efficient analytical methods fitted to unstructured data. Classical statistical techniques may be unable to meet the challenges because of heterogeneity and noise associated with unstructured datasets. The large amount of information coming from this kind of data makes crucial the necessity of focusing on unstructured data analytics to leverage the vast amounts of information available, which can lead to more informed decision-making and insights in various fields. (Gandomi & Haider, 2015)

Figure 3. Big data formats (Punn et al., 2019)



(Azad et al., 2019) Offer an in-depth analysis of the growing gap between structured and unstructured data, particularly in the context of the Internet of Things (IoT) (Wang et al., 2018).
Over the past decade, there has been a significant increase in the volume of unstructured data generated by numerous interconnected devices, which has become a distinctive feature of the IoT environment (Zhao et al., 2018). Unlike structured data, which are organized in predefined formats such as tables and can therefore be queried using a

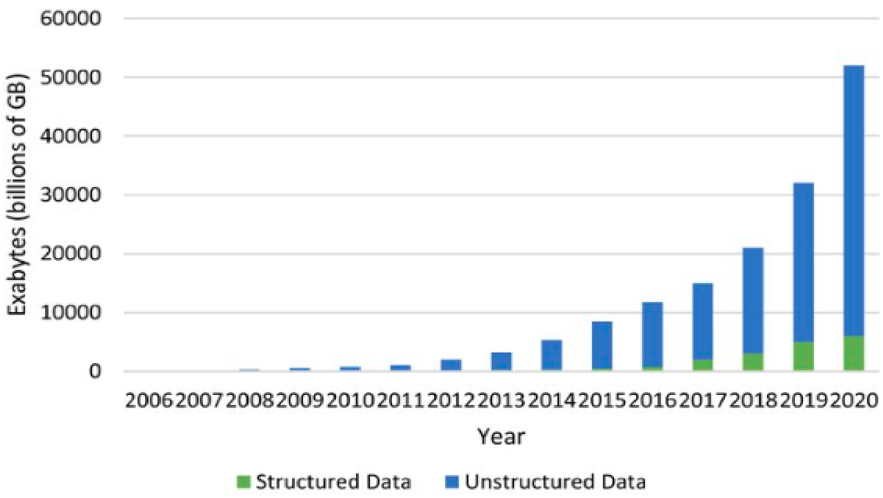
traditional database management system, unstructured data lacks a fixed schema and is thus intrinsically more difficult to manage and analyse (Jiang et al., 2015). This is because its nature is highly varied, made up of text, images, video, and sensor data, none of which adheres to a standardized structure (Gonizzi et al., 2015).

As the number of IoT devices expands, the proportion of unstructured data is expected to grow exponentially, leading to significant challenges in data management, storage and retrieval (Kim & Jeong, 2016). It is observed that the majority of data transmitted over the Internet today is unstructured, and this trend is expected to continue as the number of connected devices increases and large amounts of data are generated (Urbanczyk & Peter, 2016).

The implication of this growth is deep, because an organization has to adjust its strategies of data management to handle unstructured data effectively (Gu et al., 2017). This includes innovative storage solutions, analytical tools, and flexible data management frameworks that can accommodate the characteristics of unstructured data (Tao et al. 2017).

Finally (Azad et al., 2019) emphasises the urgency for researchers and practitioners to develop effective methodologies and technologies to address the challenges posed by the rise of unstructured data in the IoT ecosystem, as this becomes a crucial aspect of the big data challenges facing organisations (Sood et al., 2018).

Figure 4. The growth of structured versus unstructured data (Azad et al., 2019)



Information types not conforming to a predefined data model (unstructured data), such as text documents, emails, social media interactions, images, and videos, which, altogether, constitute a large bulk of the data generated in the modern digital era. Basically, much emphasis is placed on unstructured data because meaningful insights and hidden patterns, which perhaps would not emerge through traditional structured data analysis, can be revealed using this form of data, leading to enhanced decision-making processes in various sectors (Puri, 2012).

However, transforming unstructured data into valuable and usable information is full of challenges, including the embedded noise and uncertainty present in the data, as well as problems related to spelling errors, grammatical inconsistencies and semantic ambiguities (Shi et al., 2007).

These complexities are handled through sophisticated methodologies such as vector space models and unsupervised clustering algorithms, including k-means, which are applied together with thematic modelling techniques such as LDA. These techniques allow unstructured data to be structured and clustered to identify key concepts and establish relationships among the data. These sophisticated analytical techniques will then enable organizations to extract useful information from unstructured data, turning it into a strategic resource that enables innovation and evidence-based decision making (Chaudhry et al., 2023).

3.4. From Storing To Analysing – Different Analytics

Stored data, by itself, does not automatically generate value for the company, whether it lies in traditional databases, data warehouses or new technologies such as Hadoop, designed to handle large volumes of big data. The simple act of storing data is not enough to obtain meaningful results or provide usable insights.

However, once data are organized and arranged in an accessible format, their true power begins to be fully unlocked (Azeroual & Fabre, 2021). Companies can create substantial value from this data using numerous analytical techniques and tools. With deep analytics, companies are able to find trends, patterns, and relationships that would have otherwise remained hidden, thereby improving decision making, optimizing operations, and uncovering new opportunities for growth. (Ishwarappa & Anuradha, 2015).

It is through the transformation of raw and archived data into meaningful and actionable information that organizations can foster innovation and gain a competitive advantage in their respective industries. In the end, it is the analysis of archived data that then turns it into a valuable asset, which the business can use to support strategic decisions toward its success and sustainability for the long term (Azeroual & Fabre, 2021).

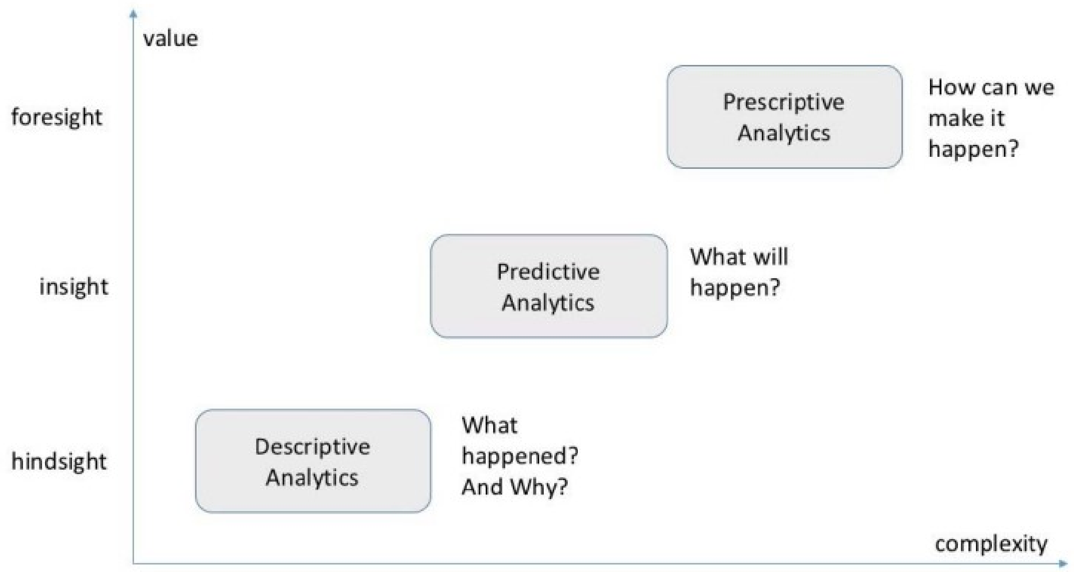
There are various forms of analytics that organisations embrace to make sense out of big data, which plays a major role in driving decision-making. The three key types are descriptive, predictive, and prescriptive analyses, each of which offers depth in understanding the insight that can be obtained from the data and its implication for future action. Understanding these analyses is crucial to organizations which aim to exploit big data and, hence, obtain a competitive advantage in their respective industries.

Descriptive analysis involves the study of historical data to realize what has happened and the trends involved. By summarizing and interpreting past occurrence data, it offers insights about 'what' had been happening within a given context. This type of analysis is crucial for organizations to discover patterns, measure performance, and get a basic understanding of their functionalities and operations (Chen et al., 2012). For instance, the companies could use descriptive analytics to evaluate their sales performance over time, analyse the behaviour of their customers, and then see where there is a need for improvement. It also allows organizations to make decisions based on facts rather than intuition through visualization in reports and dashboards of historic data.

On the other hand, predictive analytics relies on the employment of machines or other devices to make predictions about future results based on the analysis of existing data, employing statistical models mainly. Analysing historical data enables organizations to predict changes and assess the likelihood of risks or opportunities occurring (Gupta et al., 2024). This forward-looking approach enables rational selection of the best estimates in view of many decision outcomes and improves strategic planning and resource allocation.. For instance, a predictive analytic can be used by a retailing company to estimate inventory levels at certain peak seasons and manage supply chain costs due to excess or zero inventories (Souza, 2014).

Prescriptive analytics takes the analyses to the next level by suggesting specific actions based on insights gained from descriptive and predictive analytics. With it, it does not just identify possible outcomes, but it gives the best way for a firm to take that would provide the desired results for the business or organization (Bertsimas & Kallus, 2020). This kind of analysis becomes valuable in a complex decision-making environment where every other variable and uncertainties need consideration. In supply chain management, for instance, prescriptive analytics can be used to determine the most efficient delivery routes based on traffic patterns, weather conditions, and time windows for delivery (Hu et al., 2020). With actionable recommendations, prescriptive analytics therefore uses logic in enabling organisations to drive optimisation of operations for overall efficiency. The integration of descriptive, predictive and prescriptive analytics provides organisations with a robust framework to exploit available data. By understanding past trends, predicting future outcomes and recommending actionable strategies, organisations can improve their decision-making capabilities and achieve better business outcomes. As the (big) data landscape evolves, the importance of these analytics will grow further, making it essential for organisations to invest in the right tools and skills to exploit their full potential.

Figure 5. Descriptive, Predictive and Prescriptive Analytics (Baldassarre, 2016)



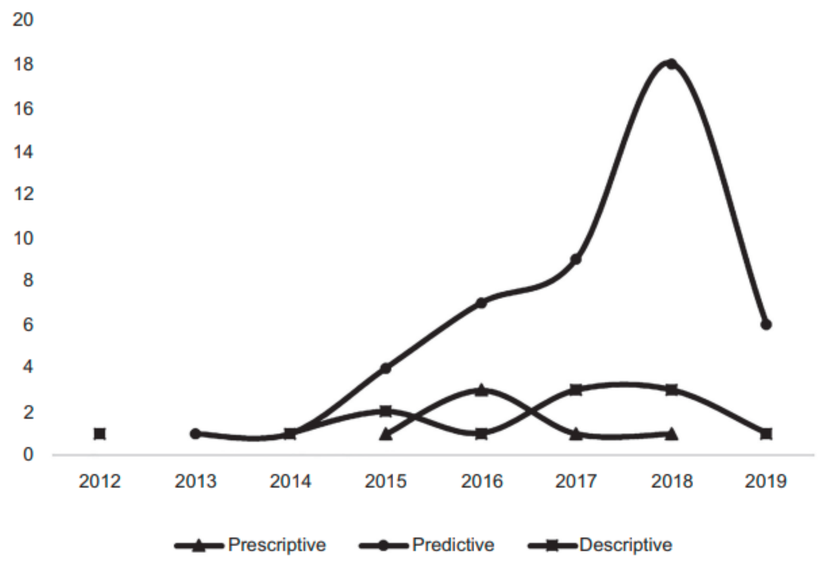
As highlighted by (Grander et al., 2021), the literature shows a significant focus on predictive analytics, which uses historical data to predict future outcomes. For example, studies in the healthcare sector have used predictive analytics for disease detection and treatment guidance, enabling proactive patient management. (Khansa et al., 2012)

In the beginning, the research landscape was dominated by descriptive analysis, which summarized historical data so that analysts could understand what had taken place. With greater complexity and availability of data, the trend has now shifted towards predictive analytics, which in logistics is applied in route optimization and operational cost reduction (Sathiaraj et al., 2018).

(Grander et al., 2021) predict increased use of prescriptive analytics, which not only forecasts outcomes but also recommends actions to maximize benefits. For example, in the context of autonomous vehicles, prescriptive analytics can guide route selection decision making based on real-time traffic data and historical patterns. This change is expected to improve decision-making capabilities in various fields, which include planning smart cities and identifying social problems through data analysis. (Semanjski et al., 2016).

Overall, (Grander et al., 2021) emphasise the growing importance of predictive and prescriptive analytics in addressing complex decision-making challenges, reflecting a broader trend in the use of big data for strategic advantage.

Figure 6. Types of analysis over time (Grander et al., 2021)



3.5. Decision Support System, Business Intelligence: Overview

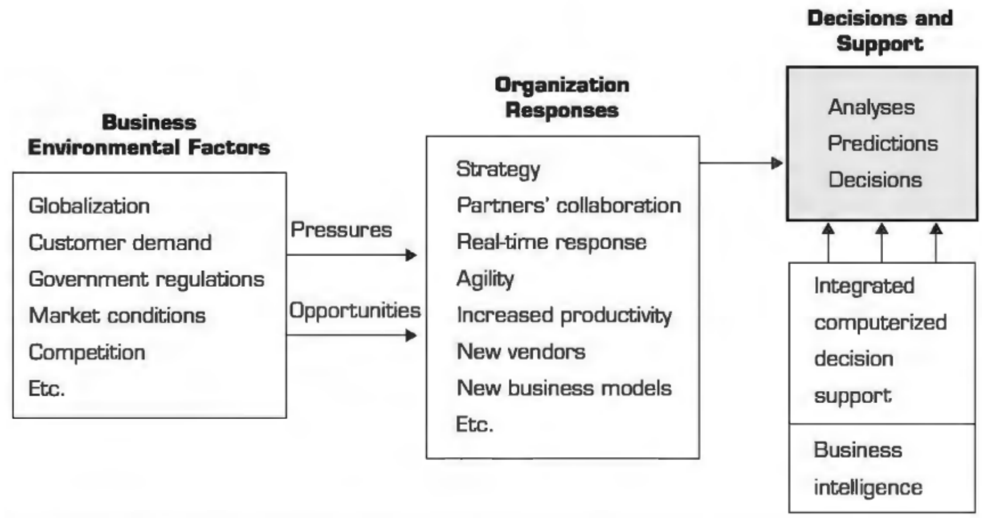
Over the years, companies increasingly adopted computerized solutions to support their operations. By regularly updating the tools and introducing new technologies such as automation and data analytics, companies can achieve efficiency and cost reductions, while being responsive to changes. In a digital era, this is one of the vital ways to keep yourself ahead in the race for long-term sustainability. Companies that prioritise innovation are better positioned to meet new challenges and seize emerging opportunities. To better identify the reasons why companies are adopting computerized support for their operations (Sharda et al., 2022) developed a new model called "Business Pressures-Responses-Support Model".

This model, as its name suggests, takes three aspects into account:

- Business pressure
- Response
- Computerised support

The first component (business pressure) refers precisely to the environment in which companies operate, constantly conditioned by changes and the need to implement new technologies in order to remain competitive.
The second component (responses) aims to show what actions are taken by companies to address and manage the pressure from the business environment.
Finally, the third component is computerised support, i.e. the digital and computerised tools that enable the company to monitor its environment and improve the effectiveness of the actions taken.

Figure 7. The Business Pressures-Responses-Support Model (Sharda et al., 2022)



The environment in which companies find themselves operating is increasingly unstable and constantly evolving. This dynamism means that there are events that can generate many opportunities, which companies can and must make the most of, but at the same time many problems and challenges that must be controlled, monitored and managed.
A typical example of an event that has changed and continues to change the factors with which companies have to operate is globalisation: indeed, while it is true that this phenomenon has brought with it numerous opportunities, it is also true that there are several challenges that companies are called upon to manage. In fact, globalisation has made the world a smaller place, making it easier to connect different areas; this certainly allows companies to expand their pool of potential customers, as their products can be sold and shipped even hundreds of thousands of kilometres away; at the same time, this gradual disappearance of territorial boundaries has increased market competition, once again underlining the existing trade-off between opportunity and risk.

As presented by (Sharda et al., 2022) the factors that populate and influence the business environment can be grouped into four main categories:

- Markets
- Consumer demands
- Technology
- Societal.

Figure 8. Business Environment Factors That Create Pressures on Organizations (Sharda et al., 2022)

Factor	Description
Markets	Strong competition Expanding global markets Booming electronic markets on the Internet Innovative marketing methods Opportunities for outsourcing with IT support Need for real-time, on-demand transactions
Consumer demands	Desire for customization Desire for quality, diversity of products, and speed of delivery Customers getting powerful and less loyal
Technology	More innovations, new products, and new services Increasing obsolescence rate Increasing information overload Social networking, Web 2.0 and beyond
Societal	Growing government regulations and deregulation Workforce more diversified, older, and composed of more women Prime concerns of homeland security and terrorist attacks Necessity of Sarbanes-Oxley Act and other reporting-related legislation Increasing social responsibility of companies Greater emphasis on sustainability

Changes and pressures can therefore arise from these four categories of factors, and companies must be able to respond in a timely manner with effective and innovative actions.
Possible examples of actions that can be taken to deal with changes and disruptions in the factors listed above may include the following:

- Leverage social media and mobile platforms for e-commerce and broader applications.
- Transition to make-to-order production and on-demand manufacturing and services.
- Implement new IT solutions to enhance communication, data access, and collaboration.
- React swiftly to competitors' moves, such as pricing, promotions, and product or service launches.
- Automate specific decision-making processes, particularly those related to customer interactions.
- Enhance decision-making through the use of analytics.

The majority, if not all, of these actions would benefit from some form of computerised support. Such actions are often supported by computerised decision-making systems (DSS).

A particularly recent example of an event that has profoundly changed the business environment, testing the ability of all companies to react quickly and adapt to new conditions is without doubt the pandemic.

The COVID-19 pandemic has profoundly disrupted the global business landscape, forcing organisations to face unprecedented challenges. However, many companies have managed to overcome these difficulties by accelerating digital transformation and increasing the use of data-driven strategies. The crisis has highlighted the importance of agility, with organisations adopting advanced technologies such as artificial intelligence (AI), cloud computing and big data analytics to improve operational efficiency and ensure business continuity (Brem et al., 2021).

The digitisation initiated by many companies due to COVID-19, however, was not limited to cope with the negative impacts of the pandemic; many digital processes and means introduced at that time are now considered standard: smart working, automated processes, use of online platforms for customer engagement... (O'Toole et al., 2020)

Many changes made during the pandemic era were in fact crucial for post-pandemic growth as well, and companies that embraced digital tools and increased the use of data in their strategic processes during the crisis proved to be stronger and more resilient, capitalising market opportunities that occurred during one of the darkest periods in modern history (Seetharaman, 2020).

The golden goal of computerized decision support is to improve the company performance and the effectiveness of the actions taken, by decreasing costs and time needed to take these actions.

3.5.1. Decision Support System (DSS)

In the early 1970s, Scott-Morton first articulated the key concepts that would come to form the basis of the decision support system (DSS) framework. He defined decision support systems (DSS) as "interactive computer-based systems which assist decision makers in utilising data and models to resolve unstructured problems" (Gorry & Scott Morton, 1971). However, it should be specified that Decision supports system (DSS) just like other terms from the IT field, e.g. management information system (MIS), do not have a universal definition. Nevertheless, in general, the various definitions that have been attributed to this term converge on a common point, i.e. the term DSS can be interpreted as any computerised system that supports decision-making.

(Gorry & Scott Morton, 1971) proposed a framework that is a 3-by-3 decision support matrix, which breaks down decision-making situations into nine different categories based on two dimensions:

- Degree of structuredness

- Type of control.

Figure 9. Information systems: a framework (Gorry & Scott Morton, 1971)

	Type of Control		
Type of Decision	**Operational Control**	**Managerial Control**	**Strategic Planning**
Structured	1 Accounts receivable Accounts payable Order entry	2 Budget analysis Short-term forecasting Personnel reports Make-or-buy	3 Financial management Investment portfolio Warehouse location Distribution systems
Semistructured	4 Production scheduling Inventory control	5 Credit evaluation Budget preparation Plant layout Project scheduling Reward system design Inventory categorization	6 Building a new plant Mergers & acquisitions New product planning Compensation planning Quality assurance HR policies Inventory planning
Unstructured	7 Buying software Approving loans Operating a help desk Selecting a cover for a magazine	8 Negotiating Recruiting an executive Buying hardware Lobbying	9 R & D planning New tech development Social responsibility planning

In total, the matrix consists of nine cells reflecting every possible combination of structured and unstructured decision-making scenarios. The first dimension, the degree of structuredness, describes the type of decisions: structured, semi-structured, and unstructured. Structured decisions represent repetitive decisions with well-defined alternatives, are usually supported by standard quantitative analysis methods, and are generally managed by lower-level managers. Semi-structured decisions require both standard procedures and human judgment and require more sophisticated DSS tools. Unstructured decisions are complex and unique, and require human judgement and innovative approaches, often supported by advanced DSS.

The second dimension, the type of control, is based on the division of managerial activities into three categories: operational control, managerial control, and strategic planning. Operational control is concerned with the efficient execution of specific tasks, usually handled by the lower-level managers. Managerial control focuses on the efficient acquisition and utilization of resources to achieve the organization's goals, whereas strategic planning includes the establishment of long-term goals and the determination of policies for resource allocation; it is normally the concern of the senior managers. Whereas cells 1, 2, and 4 in the matrix are individually concerned with structured operationally-oriented tasks, cells 6, 8, and 9 are reserved for the senior executives or highly qualified specialists dealing with unstructured strategic planning and decision making.

The use of DSS is critical for effective support, as traditional MIS and MS tools are insufficient for semi-structured and unstructured decisions. This matrix shall provide a framework for understanding how different types of decisions can be supported by various DSS tools and help organizations make informed and timely decisions in a complex business environment.

3.5.2. Business Intelligence (BI)

Business intelligence, among the various definitions that have been attributed to it, is said to have evolved from decision support systems, which began in the 1970s and developed through the mid-1980s. DSS originated in the computer-aided models created to assist decision making and planning. (Cebotarean & Maiorescu, 2011)

From DSS, over the years, the evolution of new tools, including OLAP (On-Line Analytical Processing), data warehousing, data mining and intelligent systems, has led to a significant improvement in the capabilities and accessibility of tools, models and data for computer-assisted decision-making.

In 1989 Howard Dresner (later a Gartner Group analyst) coined the term “business intelligence.” A typical definition is that “BI is a broad category of applications, technologies, and processes for collecting, storing, accessing, and analysing data to help business users make better decisions” (Watson, 2009). With this definition, BI can be seen as an umbrella term for all applications that support decision making.

(Negash, 2004) explains that "BI systems combine data gathering, data storage, and knowledge management with analytical tools to present complex internal and competitive information to decision makers"

From this definition it can be seen how, business intelligence system enables decision makers to have all the information they need, at the right time, in the right location and in the right format. The ultimate goal is to improve the timeliness and quality of inputs to the decision process, by facilitating managerial work.

BI allows to transform data into relevant information and then, thanks to analysis performed by the analysts, the information provided and transformed by BI, can generate knowledge that is key for the companies to driven crucial decisions.

Business Intelligence (BI) encompasses a variety of tasks designed to help organisations make data-driven decisions and gain a deeper understanding of their operations. Some of the key functions performed by BI include:

- Develop forecasts by analysing historical data, evaluating past and present performance and projecting future trends.
- Perform what-if analysis to assess the impact of change and explore alternative scenarios.
- Provide on-demand access to data to answer specific, non-standard queries.
- Provide strategic insight to support long-term planning and decision making.

Business Intelligence (BI) involves the use of structured and semi-structured data. Structured data is that information which is stored in a highly organized way in relational databases or flat files. In addition, semi-structured data is defined as all kinds of data that do not fit into those highly defined formats (Maimon & Rokach, 2010). Semi-structured data, including emails, social media content, and XML files, constitute more than 85 percent of all business information (Gandomi & Haider, 2015). Traditionally, BI tools were designed to manage structured data, such as data warehouses, enterprise resource planning (ERP), and customer relationship management systems. These systems are excellent at managing data from relational databases, enabling organizations to organize, retrieve, and analyse information effectively (Kimball & Ross, 2013). However, the vast volume of semi-structured data is often underutilized, despite its growing importance for decision making (Fan et al., 2015)

One of the critical features of a contemporary BI framework is that semi-structured data may be equally or, in fact, more valuable to planners and decision makers as structured data. The same processes of acquisition, cleaning, and integration applicable to structured data have to be extended for semi-structured sources as well (Borkar et al., 2012). This would ensure that the organizations derive meaningful insights from the entire spectrum of data base available with them.

To turn raw data into useful business information, analysts use tools such as extraction-transformation-loading (ETL) processes, data mining, and online analytical processing (OLAP) tools, which enable the integration, analysis, and delivery of structured and semi-structured data (Chaudhuri et al., 2011).

As just introduced, in the context of Business Intelligence (BI), data can be structured and semi-structured, and these two type of data can come from both internal and external data sources. These two dimensions:

- data type (structured vs semi-structured)
- data source (internal vs external)

offer a full picture of the variety of data the organizations can deal with.

The distinction between structured and semi-structured data, as well as between internal and external data sources, sometimes, is not immediately clear and intuitive. Indeed, semi-structured data, such as emails and contents from the web, can come from both internal and external sources.

Figure 10. BI Data Type/Source Matrix with Examples (Negash, 2004)

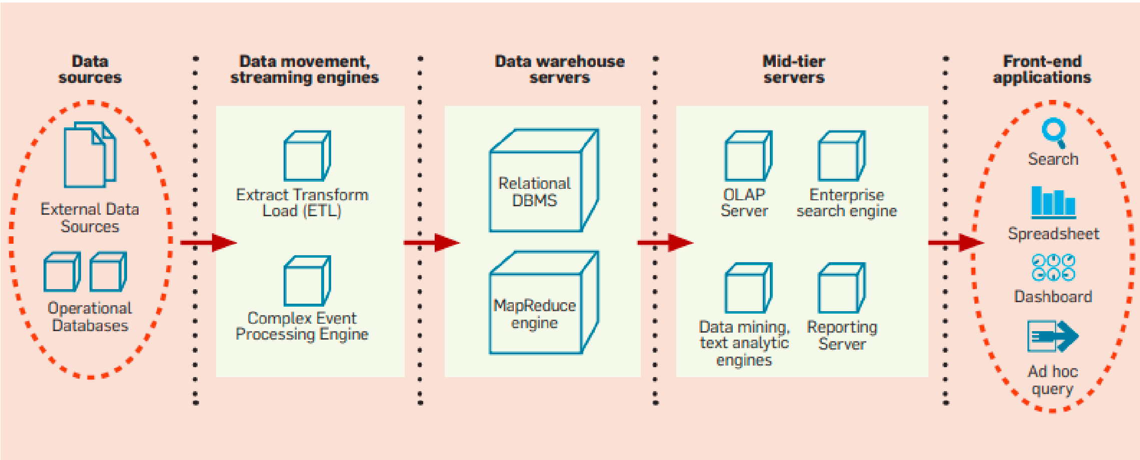
SOURCE / TYPE	INTERNAL	EXTERNAL
STRUCTURED	ERP	CRM
SEMI-STRUCTURED	BUSINESS PROCESSES	NEWS ITEMS

As shown in the matrix (Figure 10) proposed by (Negash, 2004) Enterprise Resource Planning (ERP) systems mainly deal with internal operational data presented in a structured way. On the other side, Customer Relationship Management (CRM) systems focus on external and customer related data, also organized in structured format.

Semi-structured data, on the other hand, is often captured from business processes, news and other non-standardised documents, which further increases the complexity of data analysis within organisations. This categorisation highlights the necessity of tailored BI tools that can efficiently process both structured and semi-structured data from a variety of internal and external sources.

Figure 11 illustrates a typical BI architecture within an organization proposed by (Chaudhuri et al., 2011).

Figure 11. Example of a typical BI architecture (Chaudhuri et al., 2011)



Since Business Intelligence (BI) rely on data from different sources, including multiple operational databases from various departments and external providers, this data are characterized by heterogeneous formats and quality standards, making the processes of data integration, cleaning and standardisation particularly complex and challenging (Inmon, 2005). As a consequence, efficient data loading is key for successful BI implementations, especially since these activities have often to be performed incrementally as new data arrives, for instance last month's sales data (Ranjan, 2005).

To address these requirements, organisations utilise extraction-transformation-loading (ETL) tools, that are vital to the preparation of data for BI applications. As today's business becomes increasingly fast-paced, the need for near real-time performance of BI tasks, to provide a firm with timely enterprise-wide insight decisions based on recent operational data have generated a high demand. In summary, such demand has produced

CEP (Complex Event Processing) engines that enable real-time analytics by allowing high volumes of data and stream processing with immediate, granular insight (Chaudhuri et al., 2011).

Typically, data utilised for business intelligence (BI) activities are loaded into a centralised repository, namely a data warehouse, which is managed by one or more data warehouse servers. Relational database management systems (RDBMSs) are frequently utilised for storing and querying data within such warehouses.

Over time, a multitude of data structures, optimisations and query processing techniques have been developed with the objective of facilitating the execution of complex SQL queries on large datasets, which is an essential prerequisite for BI (Elmasri & Navathe, 2011). For instance, an ad hoc SQL query might be structured to find customers who placed an order in the previous quarter that exceeded the average amount by at least 50%. In order to efficiently handle such queries on large quantities of data, big data warehouses often employ parallel RDBMS engines, thus reducing latency in query execution.

As digital data keeps growing, there is an increasing necessity to develop low cost data platforms with the capacity to manage volumes that exceed the capabilities of traditional RDBMSs. This is commonly referred to as the 'Big Data' challenge. In order to address this necessity, engines based on the MapReduce paradigm, which were initially developed for the analysis of web documents and search query logs, are now being adapted for the purpose of business analysis. Such engines are being adapted to support the execution of complex structured query language (SQL) queries, which are essential for traditional data warehousing scenarios (Dean & Ghemawat, 2008).

Data warehouse servers are generally integrated with mid-tier servers that provide specialized functionalities for various BI scenarios. Online analytical processing (OLAP) servers, for example, efficiently present multidimensional data visualizations to applications and or users, allowing standard BI operations such as filtering, aggregation, drill-down and pivoting (Plattner & Zeier, 2012).

Furthermore, the increasing relevance of unstructured data, including product reviews, emails, call centre transcripts and so on, presents a new challenge for business intelligence (BI) systems. The utility of enterprise search engines has increased considerably as a result of their capacity to facilitate keyword searches on textual and structured data within the warehouse.

In fact, it is important to point out that the conventional data warehouse had been designed to support structured data, but the rise of unstructured content has led to the incorporation of advanced technologies (Chen et al., 2012).

Data mining engines support more sophisticated analysis than OLAP or reporting servers, permitting the creation of predictive models. For example, an organization may use an analysis of historical data to determine which current customers are most likely to respond to future marketing.

Textual analysis engines are critical to process large volumes of unstructured textual data, allowing the extraction of significant information without the need for manual analysis. These engines are used, for example, for tasks such as identifying products or topics frequently mentioned in customer feedback, survey responses, and other textual sources (Feldman & Sanger, 2007).

Users interact with BI systems through diverse front-end applications, including spreadsheets, enterprise portals, and performance management tools that display KPIs in the form of visual dashboards. These enable decision makers to monitor business metrics in real time, making use of ad hoc data visualization to explore patterns, anomalies, and trends in a dynamic format (Few, 2006). The information is presented straight away because the capability of rapid, visual analysis allows users to arrive at decisions much quicker.

Business intelligence dashboards are essential tools for collecting, analysing and presenting data in a clear, visual and attractive way. The main purpose of a BI dashboard is to enable employees to create their own visualizations, facilitating sharing and quick access to data to generate rapid insights, ensuring a continuous flow of data within the company. Using various data visualization techniques, dashboards improve decision-making processes by enabling users to interact with data in real time and draw insights that can improve overall business efficiency and reactivity to internal and external events. The added value of using BI dashboards lies in their ability to simplify complex data, provide customized visualizations, and support executive decision making through key performance indicators (KPIs) and other critical metrics, fostering better analysis of business results (Khatuwal & Puri, 2022).

In times of economic crisis, optimising business processes to reduce costs and time is of paramount importance. Dashboards have proven to be a worthwhile investment in these times.

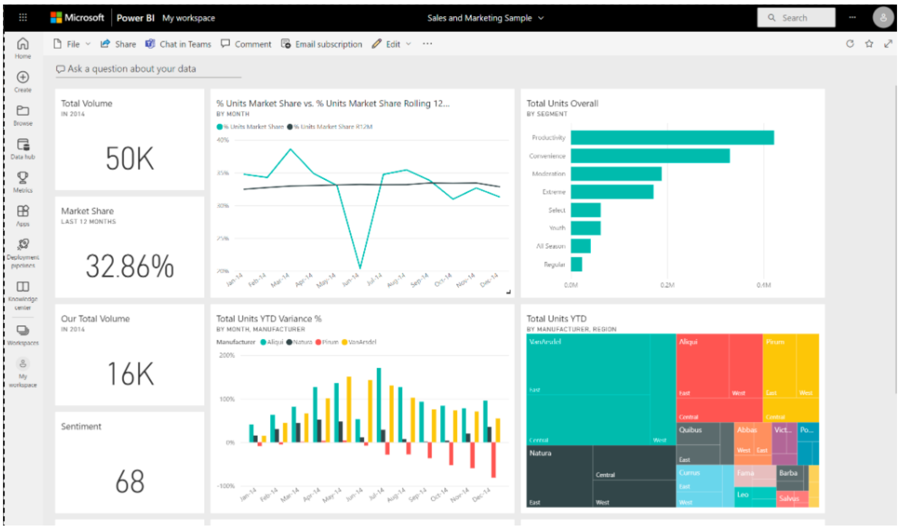
The main benefits of using a dashboard are:

- Represent performance information in a graphical format, allowing users to quickly recognise performance problems.
- Analyse performance information in a structured manner in the context of objectives.
- Help to facilitate decision-making and reduce costs incurred by manual information gathering.
- Help to reduce costs associated with administrative or manual work. (Hurwitz J. et al., 2005).

A variety of business intelligence tools are currently available on the market. One of the most widely used, and also one of the most utilised at company level, is Power BI.

Power BI has been developed with the objective of facilitating data analysis and reporting from a multitude of sources. It offers a range of features, including real-time visualisations, dashboards that consolidate various data representations, and the capacity to connect to Software as a Service (SaaS) applications without the necessity of transferring data to the cloud. This ensures secure and efficient data management. In conclusion, Power BI is an effective tool for organisations that require the ability to leverage data for informed decision-making and improved business performance.

Figure 12. Dashboard created with Power BI (Microsoft learn)



3.6. The advent of Big Data

Advancements in technologies for generating, processing, storing and networking data have significantly reduced the cost and complexity of acquiring, storing and sharing data. These improvements enable organizations to handle large volumes of data with great speed and diversity, commonly known as big data. When the challenges related with big data are properly tackled, several opportunities arise. Business intelligence (BI) is mainly concerned with transforming raw data into meaningful and actionable information for making informed decisions; it can be viewed as a form of data-driven decision support system. As the amount of data keeps growing exponentially, tools and technologies for storing, processing, and analysing it are becoming increasingly important for BI solutions. The emergence of big data has transformed BI concepts, architecture and capabilities. Unlike previous decades, BI today seeks to derive value from huge amounts of data using both big data tools and traditional methods. As a result, an intersection between big data and BI has emerged (Sirin, 2017).

With the increasing amount of data generated from mobile devices, social media, e-commerce and other sources, interest in Big Data has grown exponentially. Companies have begun to realize that there is value in turning this data into useful information through analytics (De Mauro, 2019). Because of this, Big Data analytics has become a critical topic, and most organizations have started heavily investing in Big Data analytics technology and talent (Akter et al., 2016). AI artificial intelligence and machine learning

have been integrated into BDA solutions, enabling advanced and predictive analytics (Dubey et al., 2020). This evolution and increased curiosity of organizations on Big Data Analytics to drive business strategies, has also brought to greater interest from the academic and research world. Many academics have dedicated their time to study the impact of BDAs on business performance in different industries. This interest is, in fact, dictated by the need to understand whether and how the use of BDA techniques can mark the fortune or defeat of a company.
Therefore, big data analytics from mobile devices, social media, e-commerce, and IoT give valuable insights to the firms to optimize operations, improve the experience of customers, and come up with better market strategies. Consequently, this growing interest in BDA has raised such a demand to go into more in-depth studies on the impact of BDA on business performance.

However, BDA is a relatively recent topic; as a result, research specifically addressing how BDA affects business performance is still limited.

As organizations increasingly use and analyse big data to gain competitive advantages and make strategic decisions, it becomes crucial to understand its effect on business performance. There are many reasons for the relevance of this topic:

- Impact on competitive advantage: Increasing volumes of data from a wide range of sources have made big data analysis key for organizations that intend to gain a competitive advantage. Companies that can better and more efficiently extract value from their data usually perform better, but there are many that still find it challenging to fully adopt and optimize BDA. (Davenport, 2014).
- Evolving field: Big data analytics is a relatively new field. Most of the development of tools, technologies, and methodologies has happened during the last decade. Therefore, how BDA's full potential is understood and its long-term impact on business performance remains an area that is relatively at the infancy stage of exploration (McAfee et al., 2012).
- Driver for decision making and innovation: BDA plays a key role in fostering innovation and enabling data-driven decisions. Increasingly, the ability to predict trends and make proactive data-driven decisions is becoming critical to compete in dynamic and changing markets (Günther et al., 2017).

In the last few years, some of the most emblematic technology players, such as Google, Meta, IBM, Microsoft, etc., have started investing billions of dollars in Big Data technologies, identifying BD (Big Data) as a key driver of innovation, productivity, competition and quality.

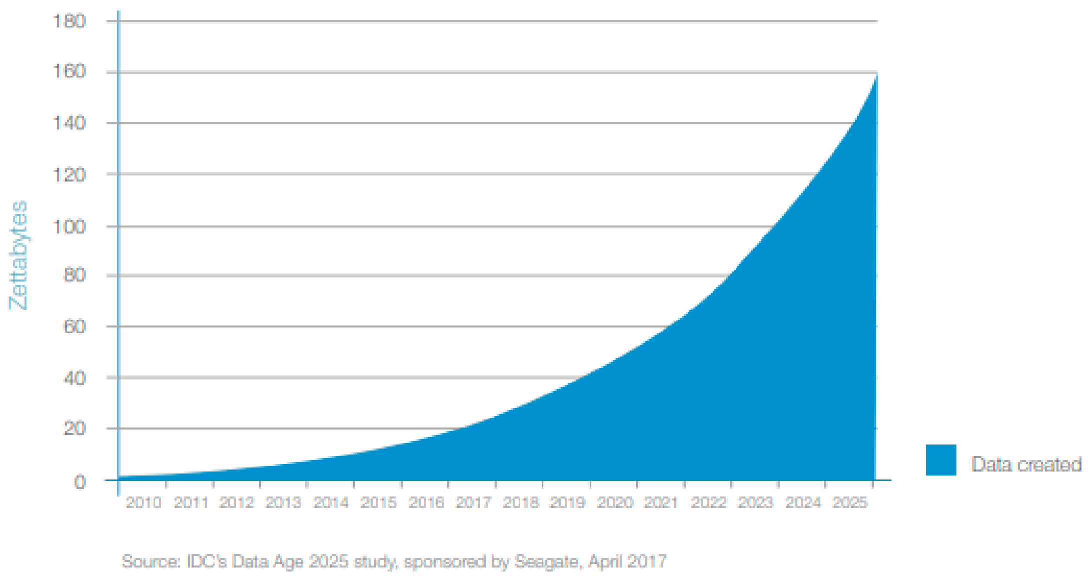
In 2011, McKinsey predicted that Big Data in five sectors (public sector, healthcare, retail, manufacturing, and personal location data) would generate about $223 billion.

Through a survey of these five major sectors that represent the global economy, McKinsey found that Big Data can play an essential role in the economic function, namely, improve the productivity, competitiveness of organizations and public sectors, and create enormous benefits for consumers (Manyika et al., 2011).

The term Big data was first coined by Cox and Ellsworth in 1997 (Cox & Ellsworth, 1997). According to the two authors, and according to the standards of the time, all those datasets characterized by a size of approximately 100 GB could be identified as "Big data." The capacity for data collection and storage has deeply evolved; what today can be easily stored in a USB flash drive, in the past was beyond the capabilities and traditional approaches of the time.

Over the past decades, the amount of data generated has grown at an astonishing rate, it is enough to think that the size of a large dataset can vary between many Exabyte [EB, 10^18 byte] to many Zettabytes [ZB, 10^21 byte] (Manyika et al., 2011).

Figure 13. Annual Size of the Global Datasphere (Reinsel et al., 2018)

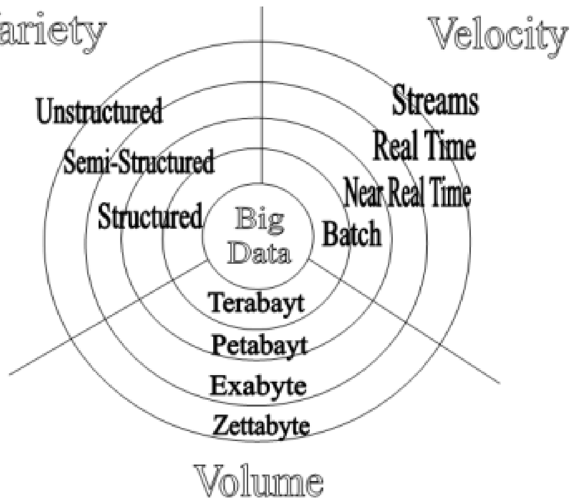


Being an ever-evolving phenomenon, as of today it is not yet possible to draw defined and delineated boundaries that can accurately quantify big data.
Nevertheless, (Dautov & Distefano, 2017) use the term big data to identify all those datasets that cannot be acquired, managed, and analysed with traditional IT tools (software and/or hardware) and in short time frames.
With the term traditional IT tools (software and/or hardware) refers, for example, to traditional relational database management systems (RDBMS), whose strength is the storage and query of data in a relational format, correlated, and contained in tables, consisting of rows and columns.
However, the relational format and structured nature is not always the case of Big Data, which is characterized by high heterogeneity and an unstructured or semi-structured nature. For this reason, what are the strengths of RDBMSs turn into limitations, which make them completely inapplicable to Big Data scenarios, or, minimize potential benefits, such as query speed or indexing, leading to situations where the resources invested to scale existing RDBMSs are disproportionate to the performance achieved.
The term "Big Data" is defined in a multitude of ways. There is no single, unified definition shared between the academic community, business industry, media, and the various stakeholders. The lack of a formal, systemic definition is one more point contributing to the mystery of the Big Data concept. The meaning of Big Data is context-specific and often varies across different industries: it depends on types of software tools and sizes of datasets commonly used in some industry. A well-defined concept of Big Data would help increase the awareness of the Big Data phenomenon among both practitioners and academics. This, in turn, would contribute to faster growth and more valued creation from Big Data, as observed by (Chen et al. in 2014). For the first time, a NASA team of scientists defined Big Data in 1997; they highlighted the need to address the challenges posed by the increasing volume of data, which required computer systems to manage larger amounts of main memory, local disk, and remote disk (Press, 2014).
In 2001, Doug Laney the analyst of the Meta Group (now Gartner), defined the challenges and opportunities associated with data growth as 3-dimensional (volume, velocity and variety). In 2013, the definition of Big Data was updated by Gartner, which defined the concept as follows: “High-volume, high-velocity and/or high-variety information

resources that require innovative and cost-effective forms of information processing to improve insight, decision making and process optimization" (Laney, 2001). In its definition, SAS defines Big Data as "a popular term used to describe the exponential growth, availability, and use of information, both structured and unstructured". IBM offers a further definition of Big Data: "Data from a multitude of sources, including sensors used to gather climate information, social media posts, digital images and videos, purchase transaction records, and GPS signals from cell phones, to name a few." (Al Nuaimi et al., 2015).

Based on the existing definitions, and their commonalities, it is possible to draw a kind of unified definition of big data, defined as a term that describes large volumes of high-speed (velocity), complex and variable data (variety) that require advanced techniques and technologies to enable the acquisition, storage, distribution, management and analysis of information.

Figure 14. The 3Vs of Big data (Sagiroglu & Sinanc, 2013)



The importance of efficient use of data was explained by (Brynjolfsson et al., 2021), among others, who stated how data-driven decision are better decisions, and that through the application of big data, managers can make rational decisions instead of decisions based on simple intuition.

This statement explains how the amount of data available is directly proportional to the quality of decisions made, as this allows to reduce uncertainty and have a complete picture of the situations on which companies are called to act. For this reason, the first step from which companies should start is the collection of data, but, of course, this is not enough; in fact, in parallel with the collection, data must be processed as quickly as possible, in real time, and from as many different sources as possible, so that relevant information and added value can be extracted from it to support the decision making process. (Manyika et al., 2011).

3.6.1. (V)olume

This feature refers to large amounts of data that can be collected and processed and, as a consequence, it is linked to companies' ability to successfully process large amounts of data.

In the corporate context, it is a measure of the amount of data available to an organisation, which do not have to be owned by the company, and therefore internal data, it is sufficient that these are accessible, and thus also from sources outside the organisation (Kaisler et al., 2013). Indeed, it is not only the dimension of data collected internally, such as information on inventory levels or sales history, that is relevant. External sources can also be integrated, bringing in new data sets. In this regard, the integration of external sources can increase the volume of data processed and stored. Consequently, a company must be able to handle large amounts of information from both internal tools and external sources. The volume or size of data is now larger than terabytes and petabytes. The scale and growth of data is challenging traditional storage and analysis.

Figures from 2012 said that, around 2.5 exabytes of data are created every day, and this figure will roughly double every 40 months. More data is flowing across the Internet every second than what was stored on the entire network just 20 years ago. This gives organisations the opportunity to work with many petabytes of data in a single record, and not just from the Internet. Walmart, for example, is estimated to gather more than 2.5 petabytes of data every hour from its shoppers' transactions. (Brynjolfsson et al., 2021). This significant increase of data is due to technological advancements in various areas of the IT industry, which have made hardware resources easily accessible, cheap and widespread. Today, at least one-third of the world's population owns a minimum of one mobile device, embedded microprocessors are now present in almost every aspect of daily life, and networking technologies together with the Internet enable the connection and remote control of smart devices. These are just some of the factors that have contributed

to the exponential growth of the data sets produced. From a data management perspective, volume is relatively easy to measure and quantify, as metrics can be expressed in traditional units such as megabytes, gigabytes, terabytes and so on.
More formally, as presented by (Dautov & Distefano, 2017) the volume of a Big Data process and/or activity "A" can be defined by the metric VolA:

VolA = Amount of data handled by "A"
which quantifies the total amount of data "A" must handle to process a single incoming activity or request.

3.6.2. (V)elocity

Velocity' refers to a company's ability to process data quickly and efficiently. As applications evolve, processing speed also increases. Computer systems do not just handle large volumes of data, but are able to process them at high speeds (Cecere, 2013). As stated by (Kaisler et al., 2013), data velocity assesses how quickly data is created, transmitted and aggregated'. Consequently, speed can influence different moments in the data management process, increasing the overall speed and allowing users to access the most up-to-date data. (Davenport et al., 2012) explain that companies that successfully harness big data will use real-time information from sensors, radio-frequency identification (RFID) and other tracking devices to address changes in patterns as soon as they happen. This will enable companies to react more quickly and effectively to events and develop data-driven strategies.
Data velocity is an indispensable factor for the utilisation of big data, not only in the context of large data sets, but also in relation to all processes. In the case of time-limited processes, the use of big data as it flows into the organisation is an optimal strategy for maximising its value.
Although the evolution of Big Data has its origins in the notion of volume, more and more applications focus on the speed at which data is created and processed, believing this to be crucial for making timely business decisions. Velocity is also related to volume, as it refers to the speed at which volumes of data are produced (or consumed) and processed. Consequently, (Dautov & Distefano, 2017) describe data velocity through two categories of metrics, which cover both the generation/arrival of data, as well as its processing, and which can be defined for generic activity "A" as follows.

$$VelG_A = \text{Velocity of data generation/arrival at } A = \frac{VolG_A}{TG_A}$$

$$VelP_A = \text{Velocity of data processing by } A = \frac{VolP_A}{TP_A}$$

VolGa and VolPa respectively represent the volume of data arriving at "A" during the time interval TGa and the volume of data processed by "A" during the time interval TPa. Simply speaking, these metrics are expressed as units of volume per unit of time, such as megabytes per second or gigabytes per minute.

3.6.3. (V)ariety

It refers to the ability of a company to effectively incorporate different types of data. This characteristic relates to the variety of sources from which the data originates. (Kaisler et al., 2013) define variety of data as a measure of diversity in its representation, including elements such as text, images, video, audio, and more. It is not just a matter of combining, for instance, inventory data with past demand information to plan production. New data sources can also be added, such as information from the web or RFID chips (Franks, 2012). Besides different sources, data can also vary in their form. In this context, (Baars & Kemper, 2008) analysed different strategies for handling structured and unstructured data in business decision support systems.

Big data come from a wide variety of sources and they are typically of three different types:

- Structured
- Semi-structured
- Unstructured

Structured data are entered into a data warehouse already labelled and easily sorted, while unstructured are random and difficult to analyse. Semi-structured data do not conform to fixed fields, but contain tags to separate data elements.

As datasets increase in size, the relational format is no longer the primary method of data storage. In the age of the Internet and social media, data comes in the form of videos, audio clips, images, and text. YouTube, for example, reports that users upload 400 hours of video every minute (Dwivedi et al., 2021), and this rate is steadily increasing. Similar trends are observed on platforms such as Facebook, Twitter, Instagram and other social networks. Another factor increasing the variety of data is the presence of many different vendors and manufacturers, as well as the lack of standardization in various areas (for example, sensory devices use different and often incompatible data formats). Even language differences used to convey simple textual information increase heterogeneity among different datasets.

As highlighted by Dautov and Distefano (2017), variety is more complex to measure in a quantitative way, as in most cases it is application-specific and depends on many external factors. However, it is still possible to classify data according to the degree of structuredness, i.e. 'the degree to which a system or component possesses a defined organisational pattern of its interdependent parts' (Boehm, et al., 1976), distinguishing between structured, semi-structured and unstructured data. Furthermore, different data sources can be identified, such as tables, files or data streams. Therefore, a possible metric to quantify the variety of data for a Big Data process or activity, A, is

$$Var_A = \text{Degree of structuredness of data managed by } A = \frac{Vol_A^{US}}{Vol_A}$$

in the formula VolUSA represents the volume of unstructured data with respect to the total volume VolA, resulting in a value between 0 and 1, i.e. 0 ≤ VarA ≤ 1. In this way, the variety of data can be represented as a one-dimensional array ranging from 0 to 1: 0 indicates fully structured data coming only from relational databases, while 1 represents the situation where the data is totally unstructured and comes from multiple different sources, making aggregation and processing more complex. It is important to note that the assignment of a variety coefficient could be based on an empirical process, due to the difficulty of precisely defining variety quantitatively during the design phase.

The three Vs just described are those that from the beginning, have been used to define big data, present in any study or research dealing with this topic; however, as time has progressed and studies have increased other "Vs" have been integrated:

- Veracity: it means the truthfulness of the data, how "certain" can we be about the data at hand? Can we trust what the data say? It mainly means the meaningfulness of the results obtained from the data for a given problem being analysed.

- Validity: it may seem similar to veracity. But in fact, it is not the same concept; data validity means the correctness and accuracy of the data with respect to its intended use. To exemplify, data may have no problem with veracity, but it may be invalid if not understood correctly.

- Volatility: it refers to the frequency with which the data change and remain useful to the business

- Value: this V represents the desired outcome of big data processing. Of course, the goal is always to extract the maximum value from any set of big data that needs to be analysed. (Khan et al., 2014).

3.6.4. The processing paradigms of big data

A processing paradigm describes a fundamental approach or model used in computer science to manage data processing activities. It includes methods such as batch processing and stream processing, each of which is tailored to specific data processing requirements (Liu et al., 2016).
Big Data processing, availability, and capturing fall into several categories, including batch processing, real-time processing, and hybrid processing. Batch processing is an efficient method to deal with large volumes of data that have been accumulated over time. In this type of approach, data are collected, stored into data sources and then processed, consequently batch results are produced. Nevertheless, several applications demand real-time data management (streams) from heterogeneous data sources. Real-time processing involves the continuous input, management and output of data. The main goal of this paradigm is low latency, i.e., data must be handled in a short or almost immediate period. Areas of application include smart cities, entertainment, and disaster management. Importantly, batch processing provides rigorous results because it allows more data to be used and predictive models to be better trained. However, it is not suitable for contexts that require fast response times. Real-time processing, on the other hand, generally ensures short response times, although this may involve less in-depth analysis of the data. Consequently, a hybrid approach is needed to enable the various application domains that use Big Data to benefit from both batch and real-time processing. With this approach, batch and real-time results are queried, merged, synchronized, or combined. Data acquisition and analysis then become more complex following this model (Casado & Younas, 2014).

The advent of Big Data has made it imperative to introduce innovative methods, tools and techniques to address the unique challenges that arise from its specific characteristics, namely: volume, variety and velocity. The typical processing solutions for Big Data, often summarized by the 3Vs, include batch processing, real-time processing, and hybrid models. Batch processing is an effective way to handle big data problems, while real-time processing is an efficient approach to meet the needs of high-speed data streams. Hybrid approaches offer solutions for applications that require analysis of both large static data sets and dynamic, streaming data, as they integrate the results of both batch and real-time processing. The characteristic of "variety" is generally relevant in all three processing paradigms.

The first known examples of batch processing for Big Data go back to 2003, when Google introduced two pioneering technologies: the Google File System and the MapReduce framework. However, at that time, companies and organizations were not widely confronted with the challenges associated with Big Data. Therefore, the first significant phase of Big Data is considered to have begun in 2006, marked by the development of Hadoop (White, 2009). Renowned for its reliability, Hadoop has become one of the most widely adopted technologies for batch processing. (Casado & Younas, 2014)'s research research indicates a lack of recent developments in batch processing technology, suggesting the conclusion of the first generation of Big Data processing solutions.

The second generation it is related to real-time processing. Indeed, companies such as Yahoo! and Twitter faced situations where they had to handle both large volumes of static data and large data in real time (e.g. streaming). One of the most well-known and powerful tools for processing Big Data in real time is the Scalable and Simple Streaming System (S4), developed by Yahoo! in 2010. S4 is a distributed computing platform that allows programmers to easily develop applications for processing continuous and unlimited streams of data (Mohammed et al., 2016). Other companies, more recently, have developed their own solutions to handle Big Data. For example, LinkedIn developed Samza and Google created Millwheel (Cumbane & Gidófalvi, 2019). Already this makes clear how real-time paradigm is constantly evolving since new technologies are emerging, and currently there are no established standard like Hadoop for batch processing.

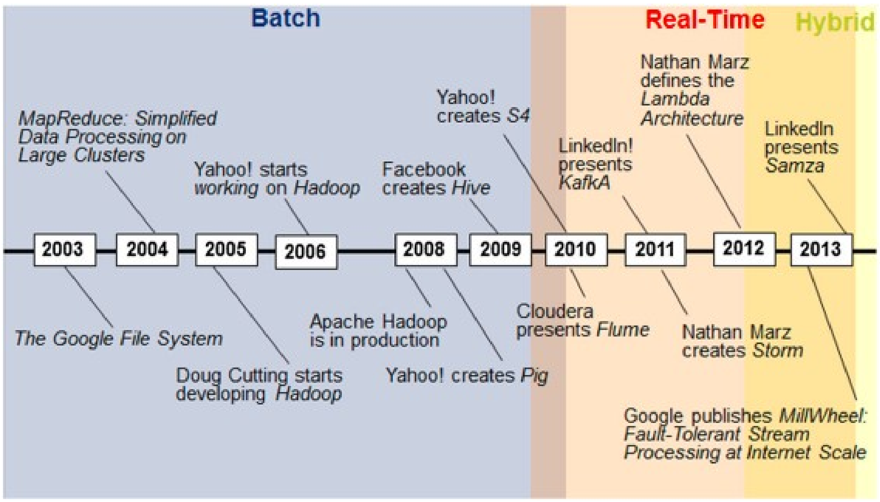
The hybrid paradigm combines both batch and streaming processing, and it is based on the Lambda Architecture. The Lambda Architecture is indeed, a data processing framework designed to handle large amounts of data by exploiting batch and streaming processing methods (Mohamed et al., 2019).

In the figure below, (Casado & Younas, 2014) illustrates the chronology and timeline of Big Data processing paradigms and technologies. In 2003 and 2004, Google published its papers on the Google File System and MapReduce. Inspired by these papers, Doug Cutting began developing Hadoop in 2005. Yahoo! supported and contributed to the development of Hadoop, releasing a stable version in 2008. Later, Facebook also began working on Hadoop, and Yahoo! developed abstract layers on top of MapReduce. In 2008, Yahoo! introduced Pig, and in 2009 Facebook introduced Hive.

In 2010, second-generation Big Data processing technologies were introduced when Yahoo! created S4, the first framework for real-time processing. In 2013, Google released a paper on MillWheel for real-time data processing. LinkedIn also released Samza, used for real-time data management.

The beginning of the third generation of Big Data processing technologies dates back to 2012, with the development of Lambda Architecture by Nathan Marz (Ounacer et al., 2017). However, it is still too early to say that the third generation has begun, although promising approaches exist.

Figure 15. Big Data processing paradigms, technologies, and timeline (Casado & Younas, 2014)



Let’s now drill down into the three paradigms.

Batch processing is a method used to process large amounts of already stored data. It is so called because it processes data present in memory and does not take into account any new data that arrives after the process has started.

The main strength of batch processing systems is their ability to scale, that is, to handle increasing workloads by expanding resources. To achieve this, batch processing often uses distributed and parallel frameworks such as MapReduce, the most commonly used technology for this type of task. MapReduce has several advantages: it offers a simple and uniform approach to working with data, it is easily scalable, and it simplifies the scheduling of distributed systems, which is complex to manage due to possible hardware failures, variations in network quality, and device diversity (Ghemawat et al., 2003).

However, MapReduce has limitations in some situations. For example, for tasks that must repeat or operate in real time it is difficult to use. Several recent implementations aim to overcome these difficulties, and further enhancements are being developed to support real-time computation, data access, and indexing.

In summary, batch processing is reliable and stable but takes longer to complete, and therefore is not suitable for tasks that require quick responses. Once started, a batch process cannot be interrupted or modified if new data arrive. An example of batch processing is the analysis of logs on a website to identify customers' buying habits. Today, batch processing is widely used for activities such as social network analysis, graph mining, scientific research, and more (Zikopoulos, 2012).

The purpose of real-time processing is to address the speed of Big Data, especially streaming data, and to ensure low latency. This processing approach is based on principles similar to batch processing, such as distribution and parallel processing. To achieve low latency, real-time processing analyses small portions of data that are stored in memory instead of on disks (Bonte & Tommasini, 2023).

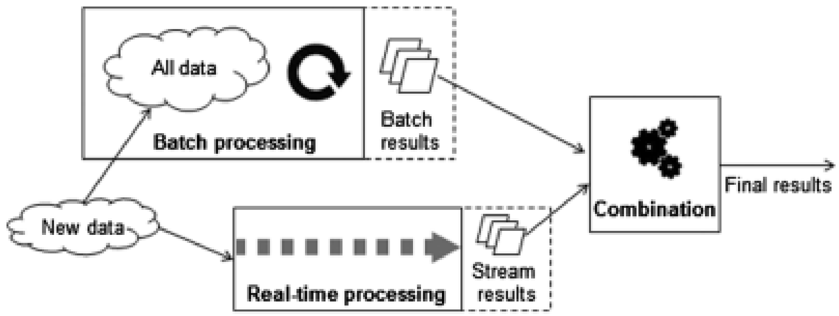
In other words, it is like continuously processing small batches of data directly in memory (diskless method). An example of real-time processing is the definition of current or trending topics on Twitter. Many applications require real-time processing of data streams from different sources. Some examples include: Smart Cities: Managing transportation, energy supply, and waste collection. Emergency Management: Leveraging data from emergency services, social networks and mobile devices to respond to critical situations. Manufacturing and Logistics: Using sensors in factories to control quality, optimize products and save resources. Entertainment: Analyzing streaming data from music, TV and gaming platforms for suggestions, user analytics and targeted advertising (Zhang et al., 2022).

Many application areas require the combination of batch processing and real-time processing. This is achieved through a hybrid model known as the Lambda Architecture. This architecture consists of three main parts:

- Batch Layer (Batch Processing): This part handles the main dataset, which is unchanged and stored in a distributed file system.
- Service Layer (Batch Results): This layer loads and makes accessible batch views within a datastore, allowing queries to be performed.
- Speed Layer (Real-Time Processing): This section deals exclusively with new data that needs fast processing.

To obtain complete results, it is essential to query both batch and real-time views and merge the results. At this stage, synchronization and result merging issues, along with other complex issues, must be addressed in what is called the Combination Layer.
The figure below proposed from (Marz & Warren, 2018) shows the high-level architecture of a hybrid processing model.

Figure 16. Hybrid processing model (Marz & Warren, 2018)



In the context of this hybrid model, new data is sent to both layers: the batch layer and the real-time layer. The batch layer continuously processes the entire dataset. However, because batch processes take time to complete, new information may arrive while processing is still in progress. This new information is not considered in the batch results. To compensate for this delay, the real-time layer deals only with the new data that has not been analysed by the batch layer. Each layer stores partial results in a database, which the Combination Layer uses to generate updated final results in real time.

3.6.5. Big Data(base)

With the growth of the Internet and cloud computing, the emergence of new applications have increased the demand for database technology. Key requirements include:

- ✓ *High read and write speed with minimal latency*: Databases need to support a very large number of read/write requests with good speed, so that applications can respond faster and create a better user experience.
- ✓ *Efficient storage and access to Big Data*: Large applications, such as social network systems and web search engines, need databases that can support very huge volumes of data-even up to petabytes-and that can handle millions of inquiries.
- ✓ *High scalability and availability*: With the increase in concurrent requests and growing volumes, databases should provide ease of scaling and updating while ensuring continuous service.
- ✓ *Reduced operating and management costs*: Increased data involves increased costs regarding databases, including hardware, software and operations. It is therefore essential to find more cost-effective solutions for storing large volumes of data.

Although relational databases have played a central role in data storage, they have some limitations in meeting these new demands, [see below the limitations]:

- ✘ *Low read and write speed*: Relational databases have a complex structure that can generate blocking problems and reduce read and write efficiency as data increases.
- ✘ *Limited capacity*: Traditional relational databases are not designed to support the big data needs of search engines, social networks, or large systems.
- ✘ *Scalability difficulties*: The multi-table structure of relational databases makes it complex to scale efficiently.

To meet these needs, various types of databases have been developed. These new databases are very different from traditional relational databases and are referred to as

“NoSQL.” The acronym “NoSQL” can also stand for “Not Only SQL,” emphasizing the advantages offered by these systems (Jing Han et al., 2011).
Of the various data models, the relational model has been the most widely used since the 1980s, with well-known systems such as Oracle, MySQL and Microsoft SQL Server all being part of the Relational Database Management Systems (RDBMS). However, recently, relational databases have begun to show some limitations, especially in terms of data modelling and horizontal scalability across multiple servers to handle large amounts of data.
There are two main trends that highlight these limitations of relational databases:

- The exponential growth of data generated by users, systems and sensors, often concentrated on large distributed platforms such as Amazon, Google and other cloud service providers.
- The increased complexity and interconnection among data, facilitated by the Internet, Web 2.0, social media and standardized access to multiple data sources. (Moniruzzaman & Akhter Hossain, 2013)

Many organizations are now contemplating the use of non-relational databases, commonly referred to as NoSQL databases, to cope with large volumes of unstructured data. NoSQL databases are specifically designed for large-scale analyses of large-size datasets and can provide greater scalability using conventional hardware. Big data analytics, business intelligence, and social networking applications that process massive volumes of data have stretched conventional SQL databases to their limits. Hence, distributed, horizontally scalable NoSQL databases have been developed such as Google's Bigtable and its open-source version HBase, and Cassandra from Facebook.
RDBMSs have significant scalability limitations, especially for data warehousing, Web 2.0 applications, grid computing, and cloud-based solutions. NoSQL databases other than RDBMSs have been developed to meet the needs of Web 2.0 applications that require more flexibility (Pokorny, 2011). For example, Web applications such as blogs contain text, comments, images and videos, source code, etc. that need to be distributed across a number of tables. Since these applications change very quickly, their databases must be designed to cope with structural changes without much effort (Hecht & Jablonski, 2011). While in traditional relational databases the addition/removal of a feature could lead to

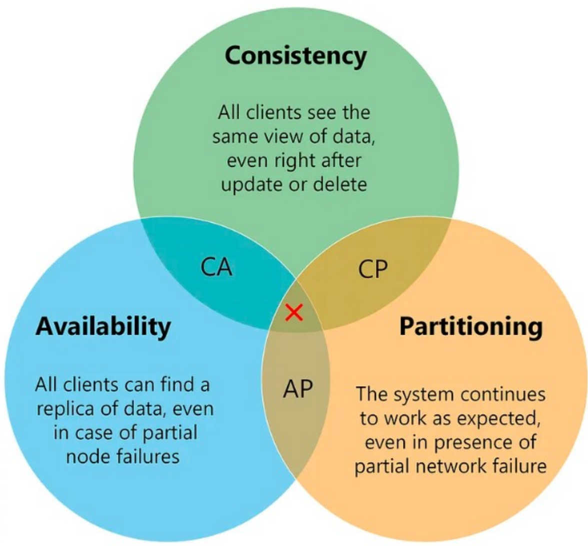
system disruption, NoSQL databases allow for seamless storage, indexing of large datasets, and handling a large number of simultaneous requests from users.

To ensure data accuracy, most traditional databases rely on transactions, which ensure data consistency in any situation. These properties, called ACID (Atomicity, Consistency, Isolation, and Durability), help make data stable and reliable. However, scaling these ACID systems presents challenges. There is a fundamental rule, so-called CAP theorem (Consistency-Avilability-Partitioning), in distributed systems that says it is impossible to achieve all three of the following characteristics simultaneously:

- Strong Consistency: Users view the same version of data, including after updates; this is possible in ACID or two-stage commit protocols.
- High Availability: Each user can always access the copy of data because if some servers are down, it will not affect.
- Partition Tolerance: The system operates correctly even if deployed on different servers, without users feeling any differences.

According to the CAP theorem, only two of these three aspects can be fully realized at the same time. For more details, see the figure below:

Figure 17. CAP theorem visualization (Litoiu et al., 2016)



NoSQL databases can be classified based on various classes that emerge from a proper way of structuring data. In other words, they are classified based on the actual data structure that can be used to store data in them.

- Key-Value: these databases store data in relation to a unique key. It follows architecture similar to the hash table, but it distributes keys and values across multiple physical servers so that they can efficiently manage large volumes of data.

- Wide Column (or Column Families): while relational databases store their data per row, these databases store it per column. Regarding this, some rows can only contain certain columns, making it more flexible.

- Document-oriented (Document-oriented): In this type of database, each document consists of fields with related values, such as first name="John" and last name="Smith," creating a document with two fields. Documents are stored in semi-structured formats, such as XML (eXtensible Markup Language), JSON (JavaScript Object Notation) or BSON (Binary JSON). Although similar to Key-Value databases, here each key represents a document ID and each value is a document (in JSON or XML format) with specific fields that can be queried.

Each category meets different needs based on structure and data access requirements. (Corbellini et al., 2017).

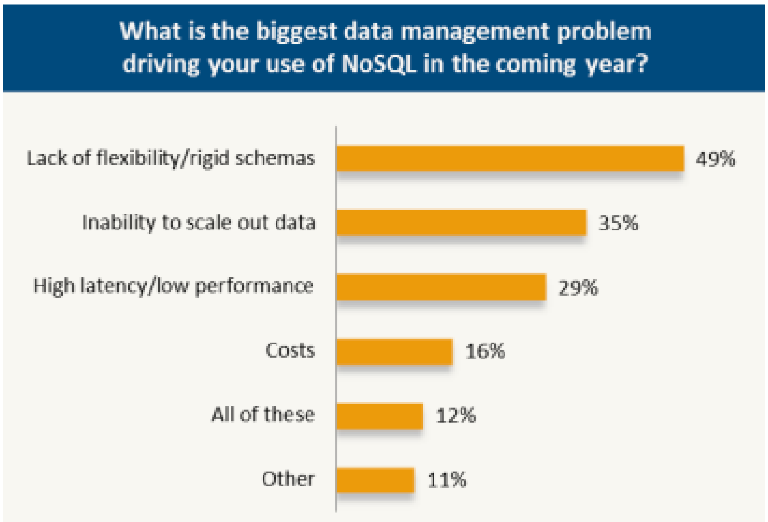
The term "NoSQL" contrary to what some may think, it is not intended as a criticism of SQL. In fact, it stands for "Not Only SQL," suggesting that both SQL and NoSQL technologies can be useful and coexist, each tackling specific needs. In recent times, the NoSQL approach has gained attention as leading Web 2.0 companies such as Facebook, Twitter, Digg, Amazon, LinkedIn and Google have integrated NoSQL databases in various ways.

A 2012 survey of NoSQL adoption conducted by Couchbase (Phillips, 2012) revealed some important trends:

- Almost half of over 1,300 respondents stated that they had funded NoSQL projects in the first half of that year. Among companies with over 250 developers, almost 70% were planning to invest in NoSQL by the end of 2012.
- 49% of those interviewed pointed to rigid schemas in traditional databases as a major reason for moving to NoSQL. Many also cited a lack of scalability and performance issues.
- Forty percent described NoSQL as important or very important to their work today, while another 37 percent told us NoSQL was growing in importance.

The figure 18 illustrates the main issues driving the move toward NoSQL databases.

Figure 18. Main issues in moving towards NoSQL databases (Phillips, 2012)

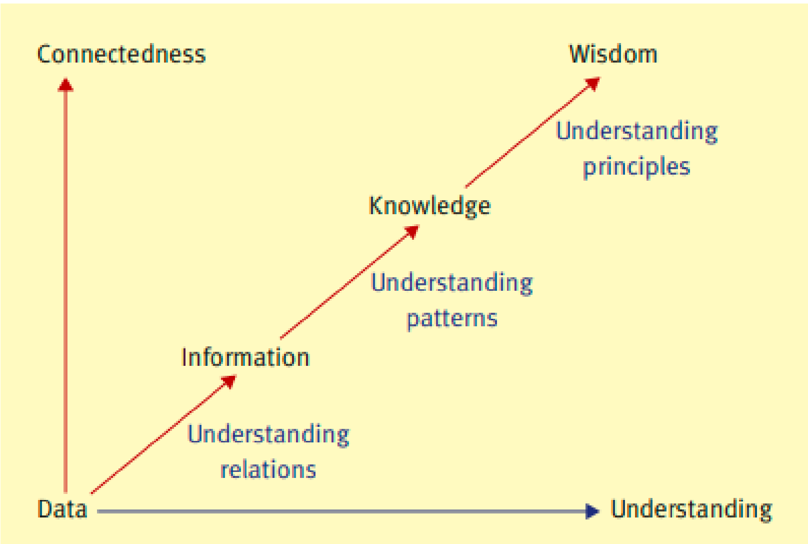


Many organizations with large data storage needs are now seriously considering NoSQL. As a result, NoSQL database experts are in high demand by many growing companies. Organizations that collect large amounts of unstructured data are increasingly turning to non-relational databases. NoSQL databases focus on analytical processing of large-scale datasets, offering greater scalability on commodity hardware. NoSQL systems are able to store and index arbitrarily large datasets while allowing a large amount of concurrent queries from users.

3.7. Big Data Analytics (BDA)

In recent years, the term "big data" has become extremely popular in fields such as economics, computer science, information science, information systems, statistics and many other areas. As technology advances, we are continually generating increasing amounts of data. This growth has no exception: it affects both people and businesses, involving the public and private sectors, from academic institutions to commercial activities. This increasing availability and dissemination of data actually represents a great opportunity, as there are more and more factors and variables that can and should be analysed to reduce the uncertainty and ambiguity of decisions (Maltby, 2011).
This need to make informed decisions that reduce risk and uncertainty is well represented by the so-called DIKW pyramid, which explains the relationship between data, information, knowledge, and wisdom. Each step of the pyramid answers several initial questions about the data and adds value to it.

Figure 19. Transition from data, to information, to knowledge, to wisdom (Cooper, 2014).



The data itself is a value, often a measure or simply a number, and in fact, when isolated, by itself, has no meaning. When data is placed in a specific context, it gains meaning and provides information.
Structured and organized information becomes knowledge, which can be divided into two groups: explicit, e.g., written guidelines, easily available and transferable to others, and implicit, i.e., acquired through experience and intuition.

Finally, there is wisdom, which can be seen as the application of knowledge.
Therefore, starting with data, itself a mere number without added value, it is essential to gather information and generate knowledge through understanding and analysing it, and then to acquire the maturity and wisdom that enable one to make thoughtful decisions and face the future with less uncertainty (Cooper, 2014).
Big Data analytics focuses on the structured processing and analysis of massive amounts of information and complex data sets, properly named Big Data, with the goal of deriving high-value insights. This analysis enables the detection of hidden trends, patterns and relationships within large-scale raw data, supporting analysts in making informed, data-driven decisions. Through this process, organizations can leverage exponentially increasing volumes of data from a variety of sources, such as Internet of Things (IoT) sensors, social platforms, economic transactions, and smart devices, to gain actionable operational insights through advanced analytical methodologies.
Since the beginning of the 2000s, advances in software and hardware capabilities have made it possible for organizations to collect and manage massive amounts of unstructured data. This unprecedented increase in useful data led the open-source communities to develop Big Data frameworks, tools designed to store and process this information in a distributed manner across computer networks. In combination with other tools and libraries, these frameworks enable a variety of core functions.
Among these functions, frameworks enable predictive modelling using sophisticated artificial intelligence and statistical algorithms, allowing organizations to predict future scenarios based on the available data. Another important use is statistical analysis, which enables deep exploration of data to detect hidden patterns and models. In addition, these frameworks enable "what-if" simulations, which are useful for exploring possible alternative scenarios and evaluating various potential outcomes. Finally, they can handle diverse datasets that include structured, semi-structured, and unstructured data from different sources, thus enabling flexible and versatile information management.

3.7.1. Differences Between Big Data And Traditional Analytics

(Mucci & Stryker, 2024) point that the main differences between Big Data analytics and traditional analytics concern the type of data processed, but more importantly the techniques and tools used for analysis. Traditional analytics is concerned with structured data, which is usually stored in tables within relational databases. These databases are designed in such a way that the data are clearly presented and easily understood by computer systems. This type of data is analysed using traditional statistical methods and machines, as a result of which a database can be queried accurately using SQL.

Big Data analysis is performed for a wide range of data, including not only structured data, but also semi-structured and unstructured data from sources as diverse as social media, IoT sensors, financial transactions, etc. Such complex data require advanced techniques that can handle its variety and size. The process of extracting useful information from complicated and often disorganized datasets requires machine learning and data mining as key techniques (Provost & Fawcett, 2013). These methods are useful when one is interested in discovering trends, patterns, and correlations that are difficult to obtain with standard techniques.

In addition, Big Data analysis requires the use of distributed processing technologies such as Hadoop or Spark. These enable the management and analysis of extremely large volumes of data, distributed across multiple servers, all in real time. These systems are highly scalable and enable the growing amount of data that characterizes Big Data. In summary, traditional analytics would suffice for well-structured and limited data, while Big Data analytics is a must for the comprehensive use of modern variable-source data to support business decisions and research in complex and interconnected domains.

The main differences between what is considered traditional data and big data is summarized in the table below (figure 20):

Figure 20. Points of difference between traditional data and big data (Kumari et al., 2016)

Parameters	Traditional Data	Big Data
Structure of data	Structures are defined	Mix of Structured, semi-structured and unstructured data
Data Volume	Based on business volumes and extent of digitization	Very high, in petabytes and even more
Variety of Data Sources	Data source from database systems	Besides data from business information systems, text (emails, documents), weblogs, sensors, RFID, etc.
Velocity	Low to moderate based on volume of business	High velocity
Flow	Fixed	Continuous round the clock accumulation of data
Structured Data	Structured Data	Structured, Semi-structured and Unstructured data
Sources of data	Organizational data, trading partners data	Organizational data, RFID, Sensor data, Google searches, Social media (Linked in, Facebook, Twitter, Whatsapp, etc.)
Analytics	Provide historical view, status reports	Real-time, direct feedback from the consumer, sentiment analysis, opinions

3.7.2. BDA: From Raw Data To Valuable Insights

The transformation of raw data into valuable information takes place through several essential steps, each of which builds the foundation for a structured analytic process. (Mucci & Stryker, 2024) identify four main steps, namely data collection, processing, cleaning and finally, analysing.

Data collection marks the beginning of the journey, when structured and unstructured data are captured from a wide variety of sources. At this stage, organizations develop strategies for collecting data in centralized repositories, such as data lakes, where metadata can be automatically assigned. This approach improves the information management and makes it easier to access.

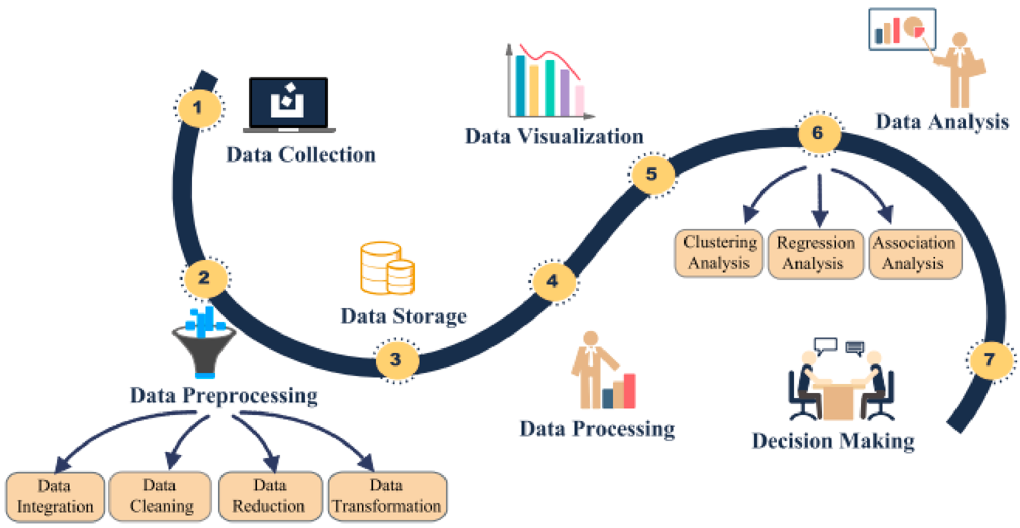
Collection is followed by processing. In this stage, data are organized, extracted, transformed, and loaded into a storage system to make them suitable for analysis. Processing is the stage where the raw data are transformed into a usable format; it may involve standardizing data types, or putting the information into structured formats. As volume grows at incredible rates, this can be a very challenging phase, typically requiring batch processing strategies suitable for large volumes over long periods of time, or streaming processing that handles small batches of data in real time.

After processing, the information must be cleaned to maintain high quality and relevance. This includes properly formatting the data, removing any duplicates, and deleting all irrelevant entries. Data cleaning is critical to eliminate errors in the results, as it makes the analysis reliable and accurate.

The final stage involves data analysis, which is the application of advanced analytical techniques to the already prepared data: data mining, predictive analytics, machine learning, and deep learning. Using the aforementioned methods, analysts can trace regularities, connections and trends in the operational set, based on which a qualified, data-driven decision can be made.

Also (Yu et al., 2021) illustrate the different stages that make up the big data processing (figure below) starting precisely with data collection, and then moving on to the pre-processing phase that requires, the integration, cleaning, reduction and transformation of the data, which, at this stage, are modified precisely from the point of view of their structure and/or format in such a way as to make them conform to their storage. Then the data are processed, visualized and finally analyzed; the last step in this process described by (Yu et al., 2021) concludes with decision making, which can indeed be supported by an accurate analytical process that can reduce its risk and uncertainty (Lytras et al., 2017).

Figure 21. Big data processing stages (Yu et al., 2021)



Under the umbrella of Analyse there is a broad range of technologies that work in together to bring out interesting and important insights from the data. Data mining identifies patterns and relationships in large datasets, while predictive analytics predict future trends and opportunities (Larose, 2015) Deep learning, in turn, emulates human learning processes to reveal much more complex and abstract ideas. Deep learning, unlike traditional machine learning that most of the time needs manual intervention, is based on layered artificial neural networks to recognize complex patterns in data and thus learns directly from images, sounds, and text. This makes it particularly effective in big data analysis; it cannot be disturbed by large volumes of complex data. (Najafabadi et al 2015). Another important technology is NLP, natural language processing, which enables computers to understand and interpret and also generate human language. In the context of big data analytics, this technology is considered of great value because it extracts information from unstructured textual data generated inside and outside an organization, making this information valuable insight (Jain et al., 2018).

The four main approaches are descriptive, diagnostic, predictive, and prescriptive analysis, which are generally applied to analyse an organization's data and extract meaningful information from it.

These approaches help to gain a broader and deeper view not only of changes in the market, but also of its composition, customers, and other relevant information that is critical to making informed strategic decisions and improving the effectiveness of the company as a whole (Mucci & Stryker, 2024). Diagnostic analysis allows companies to know the real reasons behind certain events or behaviour by extracting certain patterns and relations from the available data. Among the main benefits, this analysis helps understand the reason for changes or anomalies (such as sudden changes in sales), facilitating more informed decisions.

Descriptive analysis, on the other hand, is concerned with synthesizing available data to highlight trends and performance. It is useful for giving meaning to large volumes of data, providing insight into the current situation and allowing comparisons with past data, which is critical for predicting future trends.

Prescriptive analytics utilizes insights from all the above-mentioned forms of analytics and prescribes actions to be taken for business process optimization. It helps companies enhance their strategies, processes, customer service, or whatever, to bring things towards optimal levels by suggesting the most effective paths forward, which would increase financial performance.

Predictive analytics focuses on the forecast of future trends using data-driven models and algorithms. It thus helps companies to make an accurate forecast concerning customer needs, inventory management, and market trends. This improves their planning and resource management, and in refining market strategies.

In conclusion, these four forms of big data analytics help to draw meaningful conclusions from the raw data. Each one of them has a very distinct objective - ranging from comprehension of trends of the past, forecasting trends of the future, improving efficiency, to decision making. Thus, it is evident that the adoption of these analysis methodologies within organizations would enable them to transform big data as an operational strategy to enhance growth and improve business performance (Bay Atlantic University, 2023).

3.7.3. How Companies Can Benefit From BDA

Organizations face several challenges, for instance ensure data quality and integrity, integrate disparate data sources, ensure data privacy and security, and find the right personnel for analysis and interpretation. However, organizations that succeed in applying big data analytics gain key advantages.

One of the most valuable advantages for the company, is the capability to generate real-time intelligence through the analysis of large amounts of data as it is produced from different sources and in different forms. Real-time intelligence has the effect of promoting rapid decision making; there are timely responses to market changes and very rapid action on emerging opportunities (Althati et al., 2024).

In addition, big data analysis helps the organization reveal hidden trends, patterns, and correlations that were not previously known; as a result, leaders are more informed when making strategic decisions. Better understanding improves decision making in heterogeneous fields.

Other benefits include cost efficiency, where big data analysis identifies ways through which processes and resources can be optimized. Big data processing enables organizations to identify areas where unnecessary expenses are being incurred and, in the process, optimize operations to ensure productivity (Patil & Bhosale, 2018). In addition, predictive analytics extends this benefit to a higher level by enabling the prediction of future trends; as a result, organizations ensure economical use of available resources and are able to help companies avoid costly mistakes.

Big data analysis provides the organization with tools for deeper insight into customer needs and behaviours, preferences and engagement. Knowledge of consumer sentiment and preferences enables the company to tailor marketing strategies and improve the customer relationship process to increase engagement.

Big data analytics enhance the risk management capability of an organization. It helps organizations in threat identification, assessment, and response along with real-time development of preventive strategies that may help reduce risks before they actually occur. Predictive analytics, in this regard, supports this proactive approach by enabling an organization to anticipate such potential threats and protect itself.

(Farmer, 2024) stating that big data analytics is a great resource to make smart business decisions and to initiate changes, lists eight benefits generated by using big data in business and applying big data analytics. among the benefits and potential use of BDA highlighted, (Farmer, 2024) says that with purchase, social media, and survey data, companies can personalize offers and the overall customer experience.

Data collected from social media and competitors will provide greater insight into market trends and consumer preferences that will shape an appropriate marketing strategy, influencing product development. Aggregated customer demand data, along with supplier and shipping data, will optimize inventory, pricing and staffing to minimize supply chain disruptions. Finally, big data recommendation engines enable companies to offer highly personalized products or content, further enhancing customer interaction. In a nutshell, big data is the key to better manage businesses, sustain competitiveness, and respond quickly to customer needs.

(Mills, 2022) explains how innovations in software systems and analytics give companies the ability to make quick decisions that enable them to increase revenues, reduce costs, and stimulate growth; among the benefits listed in investing in Big Data analytics, also (Mills, 2022) include the opportunity to better understand their customers and their preferences; in fact, big data plays a crucial role in customer acquisition and retention. Furthermore, consumption patterns allow companies to reinforce brand loyalty and create extremely personalized experiences, such as those created by Amazon. In fact, big data allows, as (Farmer, 2024) and (Mills, 2022) also explain, for more targeted marketing campaigns to ensure that messages actually reach people they would be relevant to, thus improving customer satisfaction and loyalty. Big data also helps identify potential risks, improve risk management strategies, and offer innovative products based on customer feedback and market trends. Big data optimizes the efficiency of complex supplier networks, enabling the collaboration and accuracy needed for supply chain optimization.

3.7.4. BDA In Different Industries

Big Data has become very important in many fields in recent times. Today it has involved wide applications in enterprises, organizations, businesses and many other domains. The main areas that are likely to be involved in Big Data are banking, agriculture, chemistry, data mining, cloud computing, finance, marketing, stock trading, health care, and many others. All these sectors have integrated Big Data in different ways and for their specific purposes. Big Data finds its most effective application in the government and corporate sectors, helping to improve decision making and operations (Sravanthi & Subba Reddy, 2015).

Big Data made its appearance in the early 21st century, with online companies and start-ups being among the earliest adopters. By their very nature, companies such as Google, LinkedIn, eBay, and Facebook were created from scratch with Big Data in mind. Over time, for example, many organizations began adopting Hadoop software directly from Apache or from one of several third-party vendors such as Cloudera, Hortonworks, EMC and IBM. For developers, this was a convenient way to process huge volumes of data.

Today, companies use Hadoop to process, store and analyse huge volumes of web log data, helping them better understand their customers' browsing and purchasing behaviours. In the past, many organizations outsourced the analysis of click stream data or neglected it completely because they did not have the tools to process it efficiently and economically. However, with the advent of technologies such as Hadoop, they can now manage this data in a more timely and cost-effective manner (Nandimath et al., 2013).

(Sravanthi & Subba Reddy, 2015) analyse the impact of big data analytics over different industries and fields, specifying, as did the previous subchapter (3.7.3), the potential benefits that the application of these technologies may generate.

Big data shapes industries through insights to improve customer engagement, operational efficiency and risk control. In customer acquisition and retention, big data enables companies to analyse consumer behaviour so they can develop their products and services to be closer to customer loyalty and needs. This can be explained by Amazon, which applies big data to create a fully personalized shopping experience, suggesting items based on previous purchases and search history. (Akter & Wamba, 2016).

Big data in marketing helps companies create targeted campaigns by tracking transactions and online behaviours, which identify a model to drive effective advertising strategies. Automatically, with a highly satisfied customer, brand loyalty will be strong and proactive. Big data improves risk management, especially in high-risk environments. Advanced analytics enable companies such as banks to increase fraud detection rates and independently assess credit risks, relying minimally on third-party assessments.

Big data acts as an enabler for product innovation. The data collected will allow existing products to be refined and new ones to be created, enabling companies to become competitive through their responses to changing customer needs and preferences (Zhan et al., 2017). Big data enables better management of supply chains: contextual intelligence enables the supplier network to predict demand, minimize disruptions, and improve efficiency with better resource allocation.

Although challenges related to data integration and privacy are not specifically mentioned, so far big data insights have only just begun to benefit healthcare. Now, big data insights can be leveraged to improve patient outcomes by analysing huge volumes of data across various providers (Wang et al., 2018). Treatment efficiency and related costs can also be estimated. Biotech companies working in agriculture are using sensor data to optimize environmental conditions and increase crop yields (Misra et al., 2022).

Big data has shown real-time insights from the telecommunications sector to the financial sector. A telecom operator will use predictive analytics to ensure network load optimization and improve customer satisfaction, while financial institutions will analyse transaction data to detect fraud and provide secure customer interactions.

Big data is thus helping organizations in all vertical industries find critical insights for innovation, efficiency and accuracy that meet customer needs for sustainable success.

(Manyika et al., 2011) performed a study on the use of big data in different industries, noting and affirming that its application, is able to generate substantial productivity growth, demonstrating that this phenomenon, is not limited to being useful in restricted and specific industries. The opportunities resulting from the use of big data analytics have the potential to improve efficiency and effectiveness, allowing companies to use fewer inputs but produce more and higher quality outputs.

Also (Manyika et al., 2011), exemplify some potential applications of big data that can have the greatest impact on the companies' performance; again, among the various examples illustrated, they refer to the possibility of using BDAs as a tool to support product innovations and to better understand customer needs and desires.

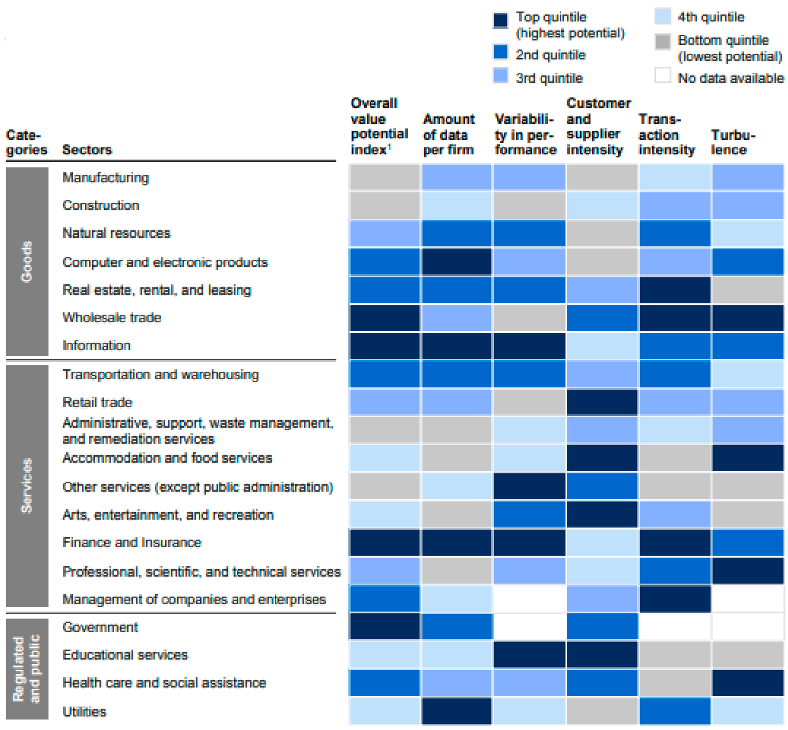
The figure below shows the potential financial impact of big data in the industries studied by (Manyika et al., 2011), placing an important focus on the ability to influence productivity and economic value

Figure 21. Big data: value generated across industries (Manyika et al., 2011)

Big data in healthcare has the potential to generate $300 billion in value per year in the United States and increase productivity by 0.7 percent. Along the same lines, the European public sector can capture 250 billion euros of value per year, with productivity growth of about 0.5 percent. The global personal location data market represents a large revenue opportunity: $100 billion or more for service providers and up to $700 billion in value for end users. U.S. retailers can increase net margins by more than 60 percent, while productivity could increase by up to 1.0 percent per year. Finally, the manufacturing industry could see up to a 50% reduction in product development and assembly costs and up to a 7% decrease in working capital. This array of information demonstrates the power of big data in enabling efficiency and improved financial performance across industries. (Manyika et al., 2011) additionally explore the potential value and challenges associated with the use of big data across five different sectors, identifying in this key characteristics that indicate which sectors are most likely to benefit from big data and which are likely to face significant obstacles. They developed two indices to systematically assess each sector's position.

The first index, named the Value Potential Index, is about the value that big data can generate in different sectors. Sectors with high scores on this index are those from which considerable dividends can be derived based on successful uses of big data.

Figure 22. Potential of using Big data across industries (Manyika et al., 2011)



This index has been computed using five criteria:

1. Volume of data per company: The more significant the volume of data, the better one stands to gain transparency in application.

2. Performance variability: High performance variability means there is greater potential to use the data for experimentation and performance improvement.

3. Customer-supplier intensity: Companies with a number of customers and suppliers stand to gain most by segmentation and targeted strategies.

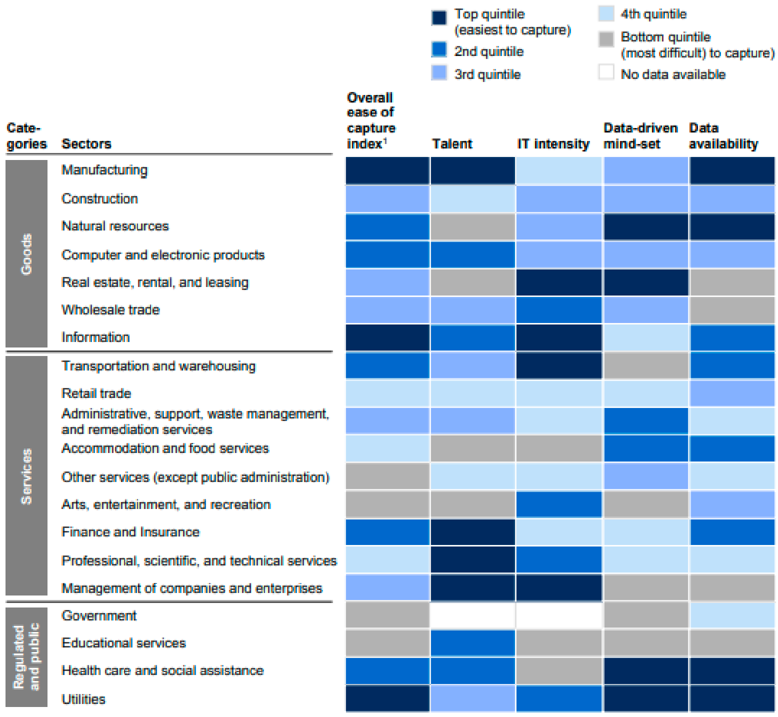
4. Transaction intensity: High transaction volume suggests the industry's propensity to benefit from algorithm-based decision making.

5. Turbulence: Frequent entry and exit of industry leaders indicates a high potential for innovation and discontinuity.

All these factors, determine whether or not an industry is ready to derive value from big data.

The second index is the Ease of Capture Index, which shows the ease with which each sector could realize this value in practice. High scores in the index mean fewer barriers to exploiting the potential of big data, while low scores mean there are significant challenges or obstacles.

Figure 23. Accessibility to Big data across industries (Manyika et al., 2011)



It is based on four criteria, namely:

1. Talent: more analytic talent can increasingly be able to leverage big data; it is measured by talent per firm size.
2. IT intensity: The greater the IT stock, the lower the technology barrier.

3. Data-driven mindset: Reflects the openness of companies to the use data. It is based on a McKinsey survey on decision making.
4. Data availability: Refers to the availability of industry data and is measured in terms of the number of databases.

These criteria could assess the various barriers that prevent the effective exploitation of big data in different sectors.

Together, these indices allow the researchers to compare sectors in terms of their potential gains from big data and the relative difficulty each sector faces in realizing these gains. The indices are built on multiple criteria, which together give a balanced view of which sectors might be well-positioned to capitalize on big data and which might encounter considerable hurdles.

4. RESEARCH METHODOLOGY

4.1. Introduction to the approach

This study aims to take a qualitative approach to examine how Big Data and real-time analytics affect the performance of companies, and business models, with a focus on decision making and outcomes in the automotive industry. The overall research query that will guide this work will be understanding how organizations leverage predictive and prescriptive analytics in the course of making strategic decisions that optimize operations and competitive advantage. Focusing primarily on the practical applications of Big Data, this research went in depth to gain a contextualized understanding of the growing influence of data-driven methodologies especially in a complex and competitive industry such as the automotive industry.

In the present study, a qualitative method design was chosen, combining illustrative examples and case study based on in depth interviews with industry professionals. The method is appropriate for the research objective, as it is able to carry out an in-depth multivariate investigation of Big Data applications from external benchmarks, i.e., the illustrative examples themselves, and from internal, hands-on insights, i.e., case studies carried out through in depth interviews. This research focuses on established, data-driven companies, named hereafter as NATU (Netflix, Airbnb, Tesla, Uber) that have been the subject of high-profile examples on innovative ways of using Big Data and real-time analytics (Tarnaud, 2020). These organizations are often cited as leaders in digital disruption and have used Big Data to reimagine entire industries. These companies provide key lessons about the potential and challenges of data-driven decision making.

In parallel, semi-structured interviews have been conducted with employees of a major automotive company, a company where I currently work. This will serve to add an inside perspective to the research on trends in the incorporation of Big Data applications into the strategic processes of a major automaker. In addition, both previously, during my internship period, and now as a full-time employee of the company, I have had the opportunity to be directly involved in some of the analytics discussed in the interviews; this direct involvement also allows me to bring to this research my direct experience and perspective on the effect of these analytics as a tool to optimize performance.

The explanatory nature of the research is also in line with the qualitative approach, as understanding Big Data in real life from a business perspective seeks to be as detailed as possible rather than testing some predefined hypotheses or arriving at numerical generalizations.

Illustrative examples will explore best practices and lessons learned from leading data companies. They offer a broad and contextualized view of how Big Data shapes the strategic decisions of modern companies. On the other hand, the interviews will seek rich and detailed testimonies from key automotive company decision makers to gain insight into the internal processes and challenges of this specific industry.

This will allow triangulation of data sources and provide a much deeper understanding of the problem studied. Illustrative examples will add an outside perspective and show evidence from other industries, while case studies with in depth interviews will present the inside perspective on how Big Data affects decision making and operational strategy in the automotive industry. The research will uncover not only what works in practice to align these findings from different data sources, but also the challenges and limitations companies face in effectively leveraging Big Data for predictive and prescriptive analytics.

4.2. Research Type: Qualitative research

This research is of a Qualitative nature, seeing as the interest of the research is in the "how" and "why" of Big Data application, rather than in the measurement or quantification of variables. Qualitative approach is, by its very nature, particularly suited to understand those complex and context-dependent phenomena ranging from how companies make decisions using big data to how they apply real-time analytics to improve performance. Quantitative research sets cause-and-effect relationships using numerical data and statistical tests, while qualitative research relies more on in-depth descriptions to delineate the nuances of human experience, organisational behaviour and strategic decision-making (Gelo et al., 2008). For this purpose, the qualitative nature of this research will allow the in-depth analysis needed to assess how Big Data is being integrated into business processes at both strategic and operational levels. It allows to investigate how data-driven business models evolve and, in general, real-time analytics

inform decision-making, living and practising something within the organisation instead of through theories. Different qualitative research methods, such as interviews or illustrative examples, can be used to flexibly explore all dimensions of Big Data usage.
The method design means that the research project will be based on an integrated approach, which in this case, will provide two complementary but distinct methods, illustrative examples, and single case studies with in depth interviews, able to grasp the multidimensionality of the research question related Big Data applications in decision-making.

1. NATU illustrative examples: The research work will analyse the practices of four of the most well-known digital disruptors, namely Netflix, Airbnb, Tesla and Uber. All these organisations are highly recognised for their use of Big Data and real-time analytics to drive decision-making and define business strategy. These companies have adopted advanced data-driven models that help them make informed and accurate decisions on aspects such as predicting customer behaviour, optimising operations and allocating resources in real time. The illustrative examples will explore a comparison between the use of predictive and prescriptive analytics to solve problems, improve efficiency and maintain an edge over competitors. The subjects of the illustrative examples were chosen because the selected companies represent benchmarks in the light of digital disruption and their experiences provide insight into the most important role of Big Data in business transformation.
2. The case studies through in depth interviews within the automotive company will be the second part of the research and will include semi-structured interviews with five key employees working within a large automotive company where I am employed. These professionals are involved in processes that involve Big Data directly in business decisions, from forecasting sales and parts demand to optimising production volumes and analysing vehicle journeys. These interviews will provide an in-depth understanding of the internal process, challenges and opportunities associated with Big Data in the automotive industry. The combination of these two approaches allows the study to build a comprehensive view of how Big Data analysis influences business practices in disruptive global companies and in a rather traditional but innovation-driven industry: the

automotive industry. While the illustrative examples will show best practices in the industries in general, the interviews offer an inside view of the implementation of Big Data applications in a specific organisational context.

Therefore, the research is qualitative in nature and the combination of illustrative examples and interviews was used to investigate the role Big Data and real-time analytics in strategic decision-making. It is based on case of leading digital companies, including Netflix, Airbnb, Tesla and Uber, supplemented by semi-structured interviews within the automotive industry. The methodology will therefore provide an in-depth understanding of Big Data applications from both a theoretical and industrial perspective. It is a study that seeks to bridge the gap between what academics have theorised and put into practice, thus providing insight into how data-driven business models and real-time analytics will remake business processes, operational efficiency and competitive differentiators.

The choice of this combination reflects the complexity of the following research question: *how does Big Data affect strategic decision making and performance in data-driven organizations?* These two "tools" can cover different needs (Eisenhardt, 1989) because with the illustrative examples, this research brings practical examples of BDA applications from companies that have harnessed data to become leaders in their respective industries. In parallel, case studies with in depth interviews within the automotive company provide a focused, firsthand look at how Big Data is being used; and by having the opportunity to speak with people involved in these types of analyses on a daily basis, it is possible to go deeper to highlight what benefits are generated, but also challenges and obstacles in their application. Together, these methods support a comprehensive exploration of the different applications of Big Data and their relevance to business performance

These two tools, applied for the purpose of answering the research question are not independent: the illustrative examples and thus the history of NATU companies and their exploitation of big data, has been useful to identify the questions and points to be touched upon during the case studies with in-depth interviews. Although these are in some cases companies operating in different sectors, and therefore companies that offer different products and/or services in the market, the topic of big data is common among them, and its application can be related to the same activity even if carried out for different products;

for example, sales volume forecasting is a crucial activity for any company, regardless of whether it belongs to the automobile or food and beverage sector (Sarker et al., 2023).

4.2.1. Illustrative examples Best Practitioners

The selection of so-called "NATU" companies is justified by the fact that they have earned the label of data-driven industry disruptors over the years. In fact, all of these four companies (Netflix, Tesla, Airbnb, Uber) are known for having data-driven business models and making heavy and effective use of big data so that their decisions and strategies are driven by predictive and prescriptive analytics. In fact, choosing to bring these four companies as illustrative examples allows to highlight so-called best practices in the efficient and effective use of big data, showing how adopting a driven model, can indeed be a key condition for becoming a leader in one's industry (David Fayon, 2020). All this can finally illustrate how big data can be an engine of competitiveness and innovation.

The NATU illustrative example is included in the research to answer, through concrete examples, the question, "How can BDAs influence business performance?" Their success stories highlight the incredible power of Big Data; illustrative benchmarks that could inform and inspire organizations in other industries, such as automotive. In addition, these cases also seek to illustrate how appropriate and effective use of Big Data can not only optimize business performance, but also influence the direction of long-term strategies.

4.2.2. Case studies: In Depth Interviews

While the illustrative examples provided insight on how some data-driven strategies have been successfully used by top practitioners, the semi-structured interviews with automotive employees provide the internal and pragmatic insight needed for the research. The interviews aim to investigate how Big Data is currently being used within an established automaker. This aspect of the research is particularly relevant given my role within the company, where I am involved in some of the analytics in scope. As such, I am able to contribute to the discussion with an insider's view, offering objective insights from interview participants and first-hand observations of BDA applications.

Through the interviews, the study wants capture the day-to-day realities and challenges of implementing Big Data in this rapidly evolving industry. More specifically, these interviews with employees can offer a very clear explanation of how Big Data empowers the daily processes and operations of vehicle trip analysis, accessory forecasting, and production planning, among concrete examples of their contribution to decision optimization and operational effectiveness.

Interviews also perform the unique task of connecting theory to practice. In some cases, interview results may agree with the literature on the application of Big Data, in which predictive and prescriptive analytics develop decision making by predicting trends, demand, and optimal resource utilization. For example, Big Data used for accessory sales forecasting identifies how BDA can provide predictive capabilities for inventory and supply chain management.

At the same time, the interviews could highlight differences between the literature and practice. While the literature talks a lot about Big Data applied to product development based on customer needs, an automotive company's use of data may be less focused on this aspect, or by the way, this process of identifying new products to maximize customer desires may not be pursued through the application of BDA. The interviews will allow investigation of these gaps and possible reasons why some BDA applications, such as the use of customer data in product design, are less meaningful or less utilized. These insights add depth to the study and present the factors that shape an organization's approach to Big Data beyond theoretical expectations.

This research is expected to achieve, by combining these two approaches, illustrative examples and case studies through in depth interviews, a broad understanding of the role of Big Data in improving performance using theoretical and practical best practices within an organization. Therefore, the mixed-method design supports a balanced analysis in which findings can be validated across sources, while allowing for a fuzzy exploration of the factors that influence data-driven decision making in different contexts.

4.3. Data Collection Procedures

In this research, data collection is carried out systematically, considering secondary sources from illustrative examples and primary data collected through case studies carried out with in depth interviews at a leading automotive company. This combination provides wide opportunities for a well-founded study of Big Data analysis applied to business performance evaluation.

In the illustrative examples section, the focus is on the companies already mentioned, Netflix, Airbnb, Tesla and Uber, as they enjoy the status of being pioneers in adopting business models based on extensive use of big data and real-time analytics. The purpose of this part is to illustrate the benefits of applying BDA by bringing as examples companies among the leading exponents of their implementation in decision-making and strategic processes.

The Big Data strategies of each of these companies are analysed in light of information from industry reports, academic research, and executive interviews to fully present data-driven initiatives. Industry reports provided insights into each company's data strategies, technology investments, and the impact on business performance from data analytics. Theoretical perspectives of Big Data as a driver of innovation and competitive advantage emerge from academic publications. In addition, public interviews with executives reveal leadership's view of the strategic importance of Big Data and the alignment of these initiatives with each company's most important goals. The set of sources presents a holistic view of the impact of Big Data on strategies.

To integrate the broader perspective provided by the illustrative examples, this research also includes case studies with semi-structured interviews focused on professionals actively involved in Big Data processes; this part of the study captures real-world insights into how Big Data analytics affects decision making, operational efficiency, and overall business performance.

Interviewees were selected based on their involvement in analytics, with different levels of experience, from specialists and senior specialists to project managers and managers.

The interviews were designed to be semi-structured, following a predefined set of open-ended questions that allow for consistency and at the same time flexibility to explore individual insights in depth. Although each interview has some independence, some key points will be common, in particular I intend to explore, with the interviewers the following aspects:

Interviewees will be asked to describe in detail the type of analysis they work with, such as vehicle trip analysis, monitoring of accessory performance, production volume allocation, and accessory forecasting.

Participants will explain the nature of the analytics, which can be descriptive, predictive or prescriptive, and discuss the main objectives of each type of analysis.

Specific questions will be asked about the value and performance impact of each analytical application: What kind of results are associated with decision making, efficiency, customer insight and overall company performance?

In addition, interviewees will be asked to further discuss problems they have encountered in using analytics and how they see further improvement or expansion of the use of Big Data for strategic decision-making.

Each interview will be scheduled at a time convenient for participants and with a duration of 45 minutes. Informed consent will be obtained prior to each interview, and all participants will be informed of the research objectives, confidentiality measures, and data use.

The interviews will all be audio-recorded, with prior agreement, to obtain as complete and accurate responses as possible. Recordings will be transcribed in writing for the purpose of detailed analysis and summary of findings. Recordings and transcripts will be stored securely, and data will be anonymized before use for the purpose of analysis and presentation of results.

5. FINDINGS

5.1. Illustrative examples

NATU companies stand for Netflix, Airbnb, Tesla, and Uber. These are emerging technology giants that represent great scale-up and innovation in all lines of business.
NATU companies are the ones that have truly disrupted traditional industries. Indeed each company has created a new, innovative, data-driven business model to transform their respective industries.
These companies are well positioned to play a powerful role in disrupting industries, characterized by rapid growth, huge data usage, and a strong focus on consumer behaviour. Their model is based on an extensive use of big data to gain a competitive advantage, optimizing service and predicting future market demand. The more NATU companies grow, the more they reshape the economy, make traditional industries adapt, and thus can be the potential giants of the future. (Tarnaud, 2020).
NATU valuable representatives in the digital economy because their business models are based on big data, which continues to transform and disrupt each of these industries. NATU companies, also referred to as "GAFAM infants," have an innovation and growth orientation that comes from the largest technology companies, including Google, Apple, Facebook, Amazon, and Microsoft. However, each of them has developed specific business models by leveraging data.
Let's now analyse the main points of these four companies:
Disruptive effect, NATU companies have disrupted traditional industries in new ways. Netflix made media a tailored streaming business, Airbnb disrupted hospitality without property ownership, Tesla rewired the automotive industry with electric and autonomous driving, and Uber redefined urban transportation by introducing demand-side flexibility.
Data-driven decisions, every business strategy involves extensive use of big data interpretation and customer anticipation, with the dual purpose of optimizing the delivery of better services and expanding the market reach. Big data will enable the creation of experiences at scale, address emerging trends, and keep users engaged.

Continuous improvement with Big Data and AI, Through the use of AI and analytics, NATU companies are refining their services and operations. For example, Netflix uses data on users to create content, while Uber uses ride data to predict demand and manage drivers' schedules. Their reliance on data is a means to innovate and compete rapidly (David Fayon, 2020).

Netflix, Airbnb, Tesla, and Uber in addition fall into that category called sharing economies, which have experienced great growth over the years, driven primarily by changing consumer preferences and habits.

The sharing economies have completely disrupted entire industries by harnessing changing consumer behaviour. These companies, through rapid and creative disruption, have managed not only to reshape the traditional business landscapes of a few years ago, but also to position themselves as key drivers of one of the most important global trends- and success stories-of recent times and probably of the future.

Many of these sharing economy pioneers have experienced phenomenal growth. In July 2015, the Wall Street Journal reported that Uber was valued at over $50 billion, placing it above more than 80 percent of the companies in the S&P 500, while Airbnb was valued at $24 billion (MacMillan & Demos, 2015) A related study by PwC puts into perspective the possible extent of this industry's growth: it estimates that the revenues of sharing economy companies, in the five sectors most important to this model, will grow from $15 billion in 2013 to $335 billion by 2025. By that date, sharing-based business models will account for half of all revenues generated in these sectors (Osztovits & Kőszegi, 2015).

Netflix evolved from a DVD-by-mail rental service to a worldwide streaming platform, leveraging cloud computing, data analytics and recommendation algorithms in order to deliver on-demand content at scale globally while enabling personalized recommendations. This data oriented strategy lowered operational costs and increased customer satisfaction by giving easy access to a big library. This strategy led to rapid growth in subscribers, revenue, and customer retention — proof that a digital transformation can streamline processes while also forging an organization around the most critical element of success: the customer.

By building a peer-to-peer accommodation network that directly connects hosts and guests, also Airbnb created a disruption, in the hospitality industry characterized by issues of trust and competition from hotels. What really improved the customer experience was the way Airbnb automated its operations, identified unique and affordable accommodations through verification processes, developed mobile applications through which reservations could be made seamlessly, and used data analytics to drill down into user preferences. The results were rapid globalization, high market valuation, strong revenue growth, and the disruption of traditional hotel models.

Tesla set out to disrupt the automotive industry by sitting on electric vehicle sales, which was facing numerous problems: high battery costs, limited charging facilities, and generally sceptical consumers. key points at the heart of its digital transformation strategy include vertical integration, in which most battery and vehicle manufacturing is controlled by Tesla itself; over-the-air software upgrades; the development of Autopilot autonomous driving technology; and the adoption of direct-to-consumer sales, bypassing the traditional dealership model.

Tesla has also leveraged big data analytics to improve vehicle performance, optimize manufacturing processes, and ultimately satisfy customers. All of this is accomplished efficiently through the use of data to improve operations and the customer experience. In this way, Tesla has become one of the most highly regarded automakers globally; it has experienced record sales growth, has been well positioned to be a leader in the electric vehicle industry, and has enhanced its brand image through a focus on sustainability and technological advancement.

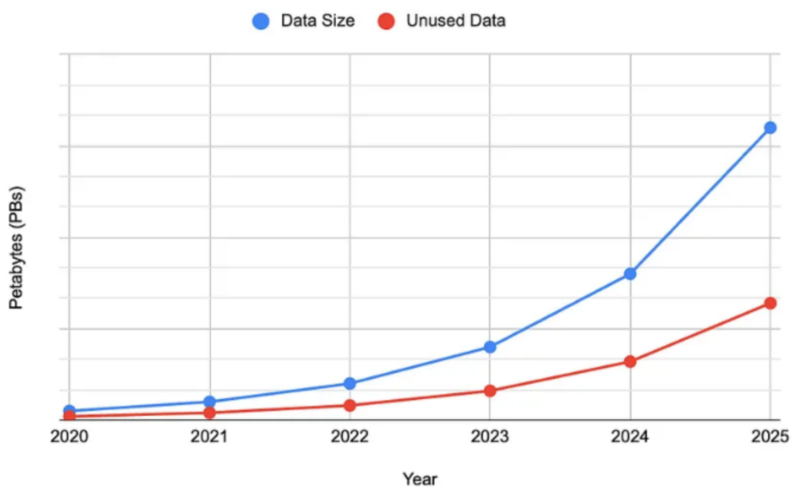
Finally, Uber disrupted traditional transportation using a ride-sharing platform that responded to problems created by inconsistent services, high fares, and dependence on cabs. Using digital means, Uber designed a mobile application to connect passengers and drivers, introduced geolocation to enable real-time matching, and developed in-app payments and surge pricing mechanisms that adjust fares based on demand.

Uber has leveraged real-time data and GPS, optimized resource allocation through effective matching of drivers and riders, and streamlined operations without taxicab stations or physical hubs. This fact-based approach has improved the customer experience: reliable, convenient, and transparent service. The service has been able to achieve astonishing growth in geographic coverage in almost every major city in the world, achieving a high market valuation, deriving more revenue from commissions earned per transaction, and offering an adaptable urban transportation option (Stankovska, 2024).

5.1.1.Netflix

With more than 65 million members in over 50 countries, Netflix streams 100 million hours of content daily and has come to account for about a third of peak-time Internet traffic in the U.S., making it a powerhouse in the streaming business. This huge engagement creates a lot of data and Netflix not only collects this data but analyses this using some great Big Data and analytics techniques. Indeed, every week, hundreds of Netflix studios around the world generate data, we are talking about 2 Petabytes of data per week. This huge amount of data comes from a variety of sources, including text, images, image sequences and so on (Netflix Technology Blog, 2024).

Figure 24. Unused data size growth (Netflix Technology Blog, 2024)



By continuously using this data, Netflix has transformed the goals of its core business model and approached many of the complexities of streaming, including viewer retention, personalized content recommendations, and content distribution optimization. Even before the emergence of Big Data and nowadays, knowing the right solution to produce successful content was particularly difficult. To tackle this challenge, Netflix decided to do things differently, namely, to use a data driven approach and analytics to find out what people would like to watch and would potentially enjoy, generating an higher customer satisfaction and stimulating subscriptions increase and renewal rates.

Since then, data has become the cornerstone of the company influencing every decision they make, from content recommendations to original programming decisions.

The recommendation engine is at the heart of Netflix's data-driven operations. Based on algorithms and built to analyse user behaviour, it tries to accurately predict what customers will like, based on their previous viewing history. Netflix's prediction efforts began as early as 2006, before streaming began, with the Netflix Prize, which puts up a million dollars for the best algorithm that can accurately predict how customers will rate movies. Although the company was streaming almost nothing at that point, its early investments in data analytics laid the foundations for the extremely sophisticated recommendation systems that would be developed in the years that followed. With Netflix's transition from DVDs to streaming, the company's menu of customer data has expanded exponentially. Now the company is able not only to track what customers watch, but also when and how offering a much more sophisticated and accurate information on customer behaviour. With that, Netflix is enabled to produce predictive models capable of correctly anticipating the taste of each subscriber and recommending certain content in order to satisfy their needs, building greater engagement and satisfaction.

The information available to Netflix not only helps to understand the main tastes of viewers, but also offers a guideline for content creation. Data on viewing habits and preferences were important when the company began producing original content. For example, through Netflix's analysis, it became evident that its subscribers were interested in content directed by David Fincher starring Kevin Spacey. This data insight convinced Netflix to bid for and win the rights to develop House of Cards, beating industry tradition for a pilot episode. Data informed the entire production, from the storyline to minor aspects such as the colour scheme used in promotional images. This data-driven approach to content creation has proven incredibly successful, driving subscriber growth and engagement and leading Netflix to produce other series such as Orange is the New Black (Maddodi & K, 2019).

The success of Netflix's Big Data strategy is reflected in its impact on key performance metrics. In the first quarter of 2015, Netflix reported 4.9 million new subscribers, up from 4 million in the same quarter a year earlier, largely attributed to popular original content. Not only Netflix's subscriber base grew, but hours spent by users watching the content provided also increased significantly; in the first quarter of 2015, users streamed a total of 10 billion hours of content. This is an important metric, as increased viewing hours indicate that subscribers feel they are getting value from their subscription and will therefore be less likely to disconnect from the service. This reduces churn and increases subscriber retention over time (Marr & Blaker, 2024).

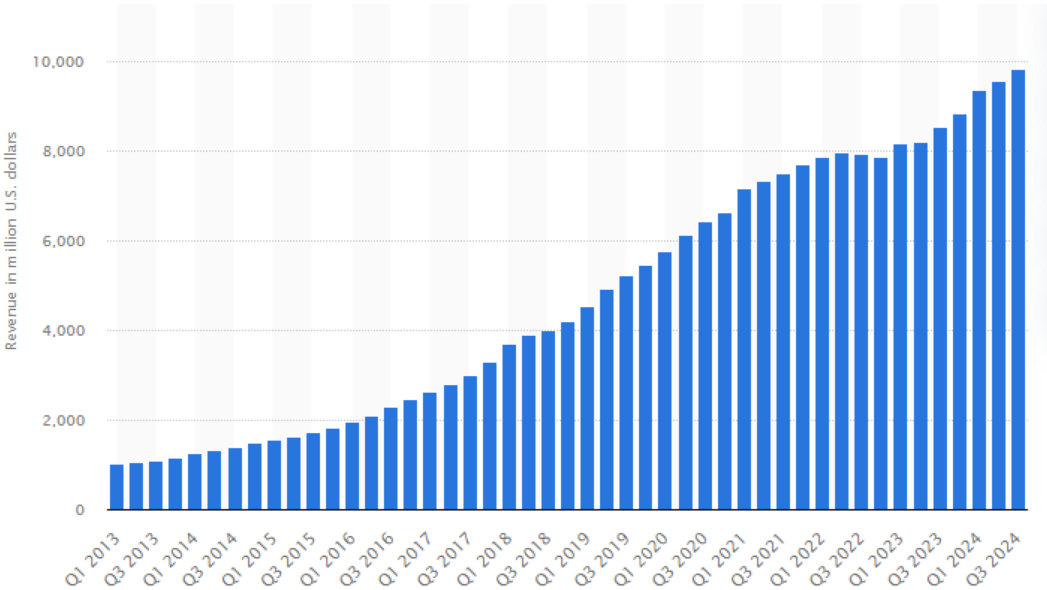
Netflix is also approaching machine learning tools, including Spark, to extend its ability to analyse streaming data that could support future developments in personalized content delivery (Kwiatkowska, 2023). One of the main challenges Netflix faced was extracting useful information from unstructured data, such as video and audio content. At first, Netflix addressed this issue by hiring live taggers to categorize content based on themes and subject matter.

Recently, however, Netflix has begun automating this process by using image recognition and colour analysis algorithms to visually categorize content. The result is automation that speeds up the tagging process, taking it a step further in understanding viewer preferences by adding much more detail about the types of scenes and visual styles that connect to segments in different ways.

This makes Netflix the clear leader in this emerging area of personalized television, thanks to the predictive power of its data. Netflix is targeting a future where people can enjoy a much more personalized viewing experience, by analysing what users have watched down to the smallest details, where personal taste shapes each user's content offer. To better explain, Netflix has leveraged data as one of the key tools to reinvent itself from a pure DVD rental service to a leading streaming platform and creator of engaging content. The above approach, through Big Data analytics, has enabled Netflix to wonderfully deliver personalized recommendations, develop original content, and optimize its content delivery infrastructure to thrive with subscriber growth rates, higher engagement rates, and a satisfied customer base. Indeed, their analytics have been instrumental in understanding and anticipating viewer behaviour, which has positioned Netflix as a leader in personalized content.

Since Netflix implemented data-driven strategies, the company has experienced astonishing growth in its revenues: from $3.1 billion in 2011 to $33.7 billion in 2023. The first quarterly decline was in the fourth quarter of 2022 (Vasupradha, 2024).

Figure 25. Revenue generated by Netflix from 1st quarter 2013 to 3rd quarter 2024 (Stoll, 2024)



5.1.2. Airbnb

Airbnb, as of today is, undoubtedly, the most important global platform capable of connecting hundreds of millions of people looking for accommodation; this platform is not only known for being an excellence in its field, but has become a true case study in the transformative power of big data and analytics. Since its foundation, Airbnb has accumulated an enormous amount of data, gaining relevant insights into the vacation habits and preferences of its users. This extensive collection of data, both structured and unstructured, has enabled and allows the company to fulfil its ultimate goal, which is to efficiently connect a large group of guests, looking for accommodations that meet their requests, and hosts who instead represent the supply, all while optimizing property usage across its global network.

The key point of Airbnb's strategy, as explained by the company's head of data science, Riley Newman, lies in deeply understanding user behaviour; in fact, any interaction by users with the platform, from booking, to looking for accommodations or reviews left by customers, generates data, and this data is valuable input for business strategies (Roelofsen & Minca, 2018). Airbnb uses the information from data collection to drive fundamental operational decisions, such as signing up landlords in destinations with particularly high demand, setting prices efficiently, and generally improving the experience of both guests and owners. Newman emphasizes how the ability to interpret information derived from customer voices, enables the generation of actionable insights. One of the biggest use cases of data analytics in Airbnb involves price optimization. The platform incorporates a sophisticated algorithm within a machine learning system known as Aerosolve to help hosts set competitive but appropriate prices. Aerosolve analyses a number of factors, including location, seasonality, and accommodation characteristics. It also takes into account and analyses photos uploaded by hosts to link the appeal of accommodations with booking success rates (Hill, 2015). For example, a cozy bedroom might generate more interest than an elegant living room, thus prioritizing listings. In addition, the algorithm segments cities into micro-neighbourhoods to better define location-based pricing. In this way, Aerosolve captures and operationalizes customer preferences into more accurate pricing recommendations.

Another key tool is dynamic pricing, taken from models used in hotels and airlines. Factors such as local events, like the Edinburgh Festival, are incorporated to ensure that prices reflect fluctuations in demand.

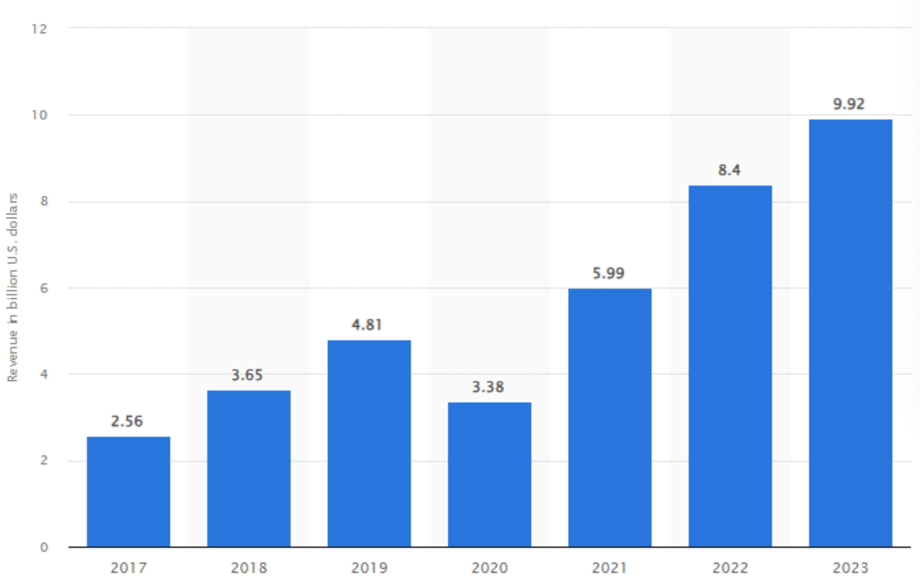
The addition of guest reviews further informs pricing, as properties with more positive feedback can get higher rates. All of this information is condensed into easy-to-use dashboards to give hosts a greater ability to set their own prices. The result is an analysis that is fully integrated into the user experience, which has a direct impact on booking rates and revenue generation (Marr & Blaker, 2024).

Airbnb's exploitation of the available data is not limited to price-related strategies: through a machine learning algorithm, in fact, it is possible to identify in advance potential frauds, suspicious and fraudulent transactions, therefore building a relationship of trust with both guests and hosts, improving in this way the transparency of each transaction, an absolutely necessary and fundamental condition for this type of business that is based on personal relationships.

The impact of these data-driven initiatives on Airbnb's performance, while not easy to measure, is extremely profound. In fact, in just a few short years (founded in 2007) it has experienced incredible growth rates, going from being a modest start-up to becoming a global leader in its industry and beyond, with more than 2 million listings in more than 191 countries in 2017 (Zhang et al., 2018), demonstrating how efficient use of data, and adoption of data driven strategies, enables strong scalability and market domination.

As a demonstration of the company's steady growth, the graph below (Statista, 2024) shows Airbnb's revenues trending from 2017 to 2023, with a positive trend except for 2020 which was obviously a year impacted by the pandemic that affected all sectors but particularly travel and tourism. The company, however, recovered abundantly after the shock caused by the coronavirus, not only returning to pre-pandemic levels but going on to more than double revenues.

Figure 26. Revenue of Airbnb worldwide from 2017 to 2023 (Statista, 2024)



Additionally, after a 40 percent drop in bookings due to the pandemic in 2020, Airbnb quickly recovered and doubled the gross value of bookings by 2022 due to increased international travel after Covid-19. In 2023, the company reported its most profitable quarter ever and surpassed its pre-pandemic profitability levels. Thanks to remote work trends, long-term stays have increased for Airbnb, but flexible travel has driven its continued growth (Armstrong & Richter, 2023).

Figure 27. Number of experiences booked on Airbnb since 2015 (Armstrong & Richter, 2023)

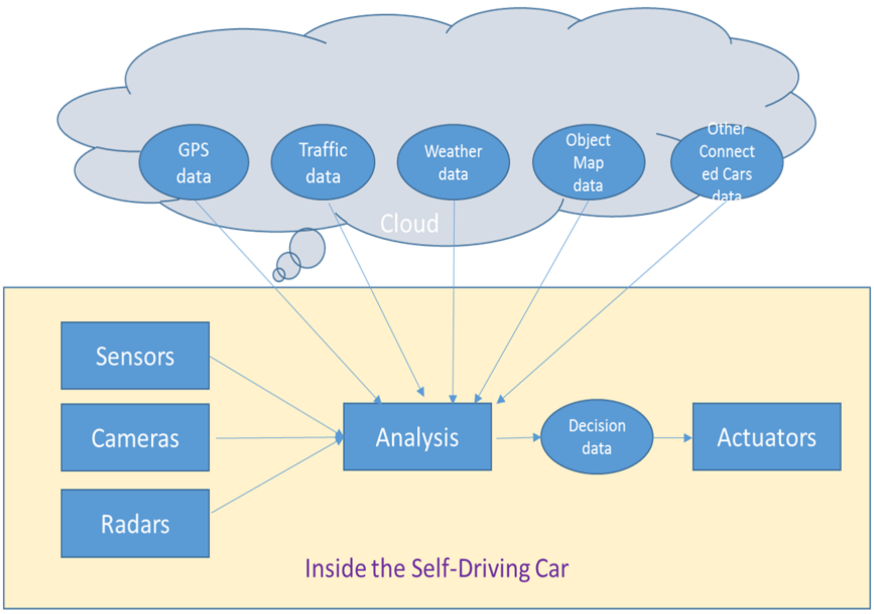
As a conclusion, Airbnb represents an important example of how the application of big data analytics can transform and make business more efficient if adopted strategically. Price optimization, fraud and scam listing detection, proper, efficient, and automated matching of supply and demand-all that and much more testifies the strong commitment Airbnb has in using data for improvement in business performance. This company continuously refines its technology infrastructure while fostering a culture of data-driven decision making, placing Airbnb at the forefront of its industry and showing ways in which analytics can drive growth, innovation, and competitive advantage.

5.1.3.Tesla

Tesla represents another significant example of a company that has demonstrated great application of BDAs to drive its business; in fact, tesla, besides having revolutionized the automotive industry and the electric cars market, has made innovative use of big data over the years. BDAs are being applied in a variety of domains and for a variety of purposes, including self-driving technology, safety enhancements, and personalized driving experiences, thus increasingly reinforcing Tesla's reputation as a technology leader and customer satisfaction.

A key point of Tesla's success is the use of big data to develop and implement self-driving capabilities and increasingly improve the driving experience (Kumari & Bhat, 2021). The company is able to collect a huge amount of data from its vehicles which are equipped with sensors, cameras, radars ... these therefore continuously provide information either about the vehicle and the driver's driving habits, or about the surrounding environment such as road conditions, weather, traffic situation. In addition, Tesla cars interact with cloud servers that aggregate real-time GPS information, object maps, and connectivity data from other Tesla vehicles (Cheruvu, 2015) This complete data pipeline supports Tesla's autonomous driving features, which are offered as optional enhancements and represent a significant part of the global market for advanced driver assistance systems.

Figure 27. Pipeline for Tesla data collection (Wang, 2023).



Tesla's self-driving model fuels a virtuous circle of data collection and improvement. Since the vehicles would be running on self-driving mode, they continue gathering new data flowing back into the system, with algorithms continuously refined. That is further augmented by Tesla's "fleet learning" system, whereby all of Tesla's cars together provide data to improve the performance of each single vehicle. Regular "over-the-air" software updates further ensure that improvements can be quickly implemented on a mass scale and within a short period of time, enabling iterative advancement in self-driving technology (Wang, 2023).

However, managing the huge amount of data collected is not at all straightforward, and includes numerous challenges, for example, all that unstructured data from video and/or image captured by the sensors and cameras in the vehicles obviously needs to be managed by substantial computational resources as it requires cleaning and transformation before it can then be analysed. In addition, the accuracy of autonomous driving systems can be hampered by noisy or faulty sensor data and subjective biases in user feedback.

Tesla, however, is not just using Big data to implement a self-driving experience; it is also being used to improve safety through advanced predictive and preventive systems.

All Tesla models have safety systems that make use of algorithms, machine learning, and artificial intelligence in order to identify in advance potential hazards that could endanger people's safety, such as icy roads or objects on the ground. These safety systems alert the driver to the potential danger and can autonomously implement countermeasures to prevent any damage from occurring. Features such as "Automatic Emergency Braking" and "Lane Departure Avoidance" are examples of Tesla's ability to integrate real-time data with high-accuracy algorithms in order to minimize risk; in fact, Tesla collects data from millions of cars to analyse collisions patterns so that it can redefine its safety measures to anticipate and thus prevent any road accidents from occurring (Tesla's website, 2024)

Despite the many advantages, the predictive models underlying Tesla's safety features face the challenge of balancing conservatism and functionality. Models that are too complex may sacrifice usability, while models that are too liberal may fail to prevent critical incidents. This point is one of the major trade-offs related to the use of big data in the highly risky context of vehicle safety.

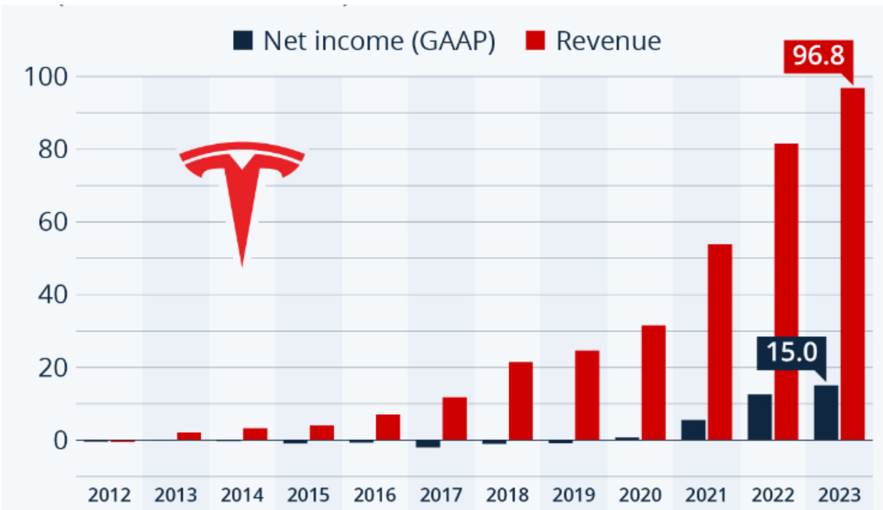
Tesla can also provide, based on the available data, a range of personalized experiences to drivers so that they can dynamically modify the vehicle configuration according to their subjective driving preferences, thereby maximizing satisfaction. By installing a series of sensors, in fact, Tesla cars can tailor a wide range of driving dynamics to the subjective preferences of different drivers. These include adjustments to music, mirror placement, air conditioning and navigation routes, which change as the driver's habits and needs evolve. Cars are even able to distinguish one driver from another and adjust settings appropriately, a high degree of personalization (Morgan, 2024)

Tesla also shows its commitment to maximizing customer satisfaction by actively using third-party data from online forums, for example, to discover and address customer needs. These kinds of feedback loops enable Tesla to prioritize features and design changes based on user satisfaction; hence, it is most likely that one will observe higher scores of customer satisfaction. Combining technical innovation with a customer-oriented approach empowers Tesla to build a community of loyal users and maintain the brand image.

Summarizing, Tesla represents big data transformation in the automotive industry. From data collection to machine learning and analytics, Tesla has reinvented what was thought possible for an automotive company. Autonomous driving, safety, and personal experiences rely on robust data strategies designed to move quick with emerging challenges and opportunities. Tesla reaches this point within the industry, where so many technical and operational challenges are yet to be overcome, due to big data being utilized effectively, so that it leads the industry both in technology as well as in consumer satisfaction.

Tesla, with its focus on using big data to drive its business has seen continued growth rates in its revenues and profits as shown in the chart below, the company ended Q4 of 2023 with $7.9 billion in net profit and $15 billion in the full calendar year (Zandt & Richter, 2024)

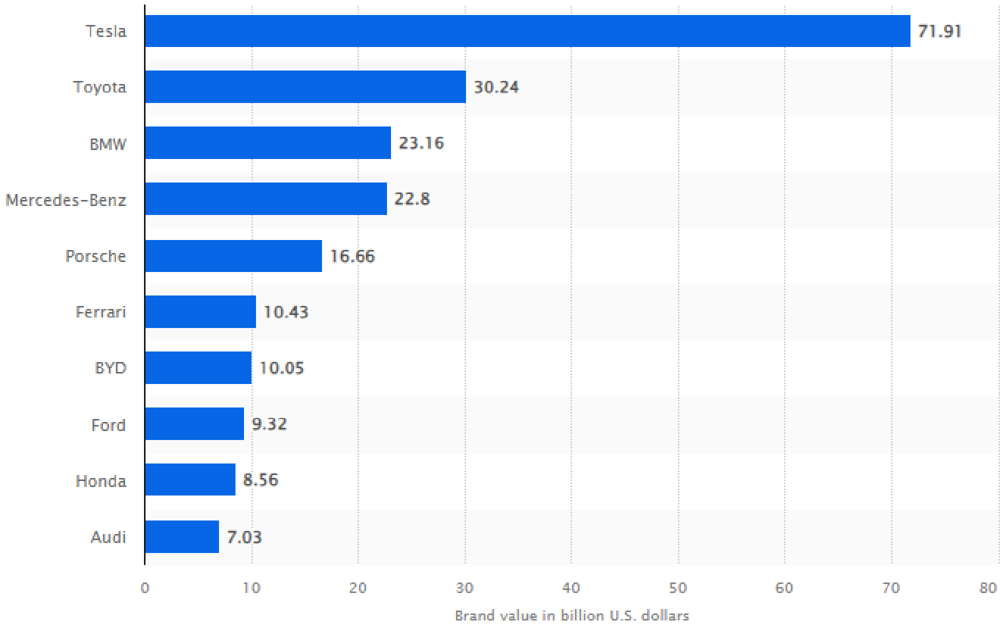
Figure 28. Yearly net income and revenues of Tesla (in billion U.S. $) (Zandt & Richter, 2024)



In addition, Tesla, since its founding in 2003 as Tesla Motors, has consistently climbed the rankings of the world's most valuable automotive brands, reaching the top in 2021.

The Tesla brand has been ranked as the world's most valuable automotive brand in 2024, with a brand value of about $71.9 billion. Toyota, the 2020 leader, is now in second place, followed by BMW (Carlier, 2024).

Figure 29. Brands value within the automotive sector worldwide as of 2024 U.S billion $ (Carlier, 2024).



5.1.4.Uber

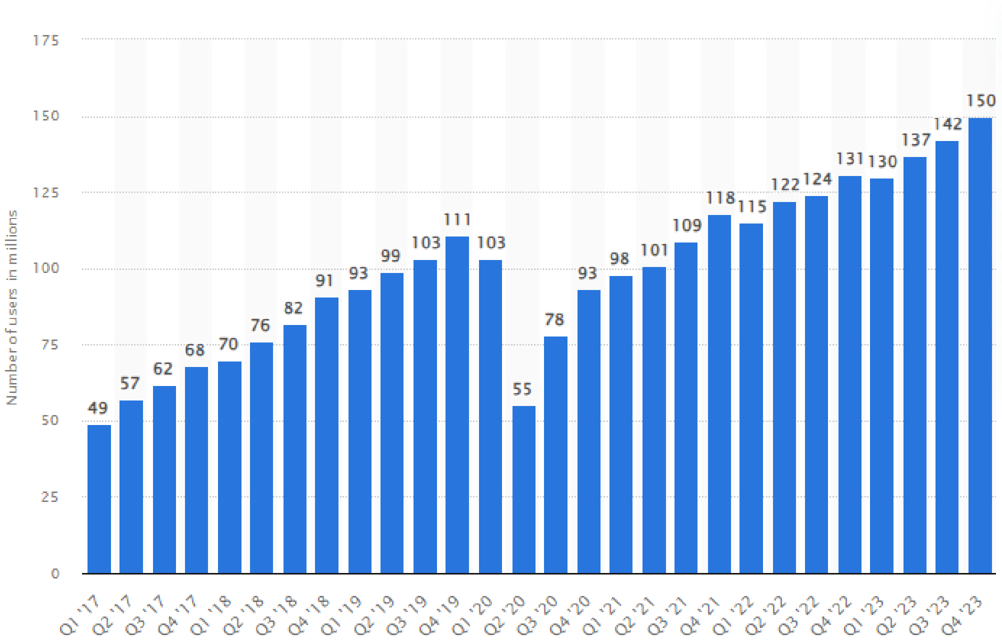
Uber, founded in 2009, as of today establishes its strategies based on continuous and massive data collection in order to generate competitive advantage and maximize its performance in an increasingly dynamic and competitive global market.

Uber's ultimate goal is to provide safe, high-quality transportation around the world. To do this, the company relies strongly on data-driven strategies and decisions, from predicting rider demand during periods of heavily congested traffic, to identifying powerful bottlenecks in their driver partner sign-up process. Over the years, the need for more and more information has led to the generation of more than 100 petabytes of analytical data that require cleaning, storing, and making available with minimal latency (Uber's Website, 2024).

Within just a few years, the company has completely revolutionized the transportation industry, becoming a brand known worldwide and operating on every continent.

Demonstrating the skyrocketing growth, as reported by (Statista, 2024) the sharing economy services provider has experienced a steady increase in its monthly users, 150 million worldwide in the fourth quarter in 2023, exceeding pre-pandemic volumes.

Figure 30. Number of Uber's active users worldwide from 2017 to 2023, by quarter (Statista, 2024)



The success of the company based in San Francisco, like the other illustrative examples mentioned above, is closely linked to its innovative and strategic use of big data, a cornerstone of its operations. In fact, since its inception, Uber has revolutionized the cab and transportation industry through targeted and efficient management of the available data to offer its customers a higher quality service, through a business model that surpasses the traditional one applied by its competitors. The impact, and the role that big data plays within Uber's operations and its impact on business growth cannot and should not be underestimated, as it allows to solve the challenges and critical issues that characterize this industry and offer a user-friendly experience.

The business model of Uber is based on the concept of crowdsourcing, made possible thanks to the use of big data. The application of this concept allows any person who owns a car to become an Uber driver, thus representing the supply side, and to connect with millions of people who require a ride who instead, represent the demand side. By using a platform that efficiently connects supply and demand, Uber offers a service that has the ambition to maximize the satisfaction of its customers by targeting potential market gaps and covering areas where public transportation is scarce or limited (Heiskala et al., 2016). Uber additionally pursues environmentally sustainable goals by encouraging people to share rides and pool journeys, limiting the number of vehicles on the roads to reduce traffic and emissions. This crowdsourcing model, fuelled by data, has been the distinguishing feature of Uber vs. the tractional cab system.

Data collection and analysis are the steps that precede and are at the heart of any operation carried out by Uber (Torre-Bastida et al., 2018). In fact, the company collects and stores information obtained from each trip, and analyses this information to anticipate the market and the demand, allocating in advance the necessary resources but not only that, through the collected data, Uber calculates the best fares to apply depending on the demand for that specific route. Through predictive analytics Uber is able to make predictions in real time, thus allowing drivers to be located in areas where there is a high concentration of demand so that people requesting the service, do not have to attend long waiting times. This demand density information is also integrated with real-time traffic status information, so trip durations are better estimated and fares are adjusted accordingly (Marr & Blaker, 2024).

In fact, one of the most relevant uses that the company makes of the available data concerns precisely the surge pricing mechanism, which is a dynamic pricing strategy that adjusts fares based on traffic conditions and demand. Through an algorithm that takes into account the demand for trips and the availability of drivers, Uber wants to make sure that drivers have an incentive to offer their service, to avoid demand disruption. To achieve this goal, i.e., the availability of drivers at times of peak demand, and/or when traffic on the roads is particularly high, Prices increase so that drivers are stimulated to offer their service. This mechanism is indeed called “surge pricing,” and it is a strategy particularly used in the aviation and hotel industries (Uber's website, 2024).

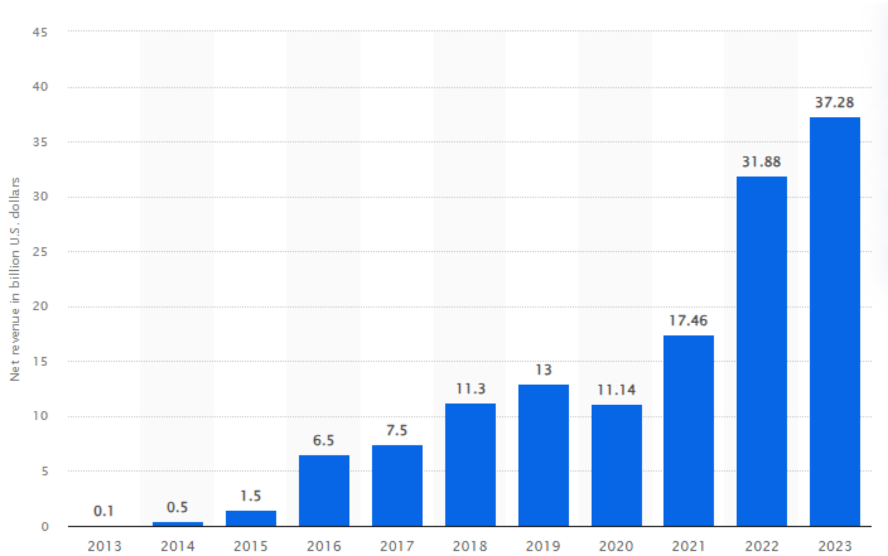
In addition to price-related strategies, Uber also leverages Big Data to optimize routes, which benefits both passengers and drivers. This route optimization is done through GPS data and real-time data on traffic conditions, so the platform will suggest to drivers the shortest route, both to reach the customer and to take them to their destination, thus going to reduce travel time and fuel consumption. These positive effects generated by route optimization make it possible to both increase customer satisfaction but also to pursue sustainability goals resulting in less pollution. This, however, is not the only sustainable initiative undertaken by Uber through the use of data; in fact, the UberPoll service connects riders traveling to similar destinations who can potentially share part or the entire ride. This is made possible by analysing multiple data on trips that have similar starting points and similar destination points. It is effectively a car-sharing service that positively impacts passengers, as they share not only the trip but also the cost, but also environmental sustainability (Uber's Website, 2024).

Trust and accountability are essential in Uber's peer-to-peer business model, and data plays a crucial role in fostering these qualities; uber in fact has an extensive Review System, where both drivers and passengers can leave their comments about their counterpart and rate their respective experience. Through this system, which actually generates unstructured data all the time, Uber can maintain high quality standards because, drivers with low ratings risk not getting high demand; but it doesn't end there, because drivers are also rated based on their "acceptance rate" i.e., the number of accepted rides once requested; in fact, Uber requires its drivers to have a high score on this metric to ensure that demand is always met (Hyun Kim et al., 2021).

The data at Uber's disposal come from both internal and external sources. For example, among the external data, as already mentioned, the company collects information on road traffic and the public transportation schedule, so potential gaps can be identified and holes in the supply can be covered. Internal data obviously include all information related to business operations, so real time data from GPS, user interactions, and all data from trips.

The impact of the data driven strategies adopted by Uber on its performance are, again, deep. Its ability to handle huge amounts of data from multiple sources has accelerated its expansion around the world and to continually increase its revenues and reach the point of generating more than $37 billion worldwide

Figure 31. Revenue of Uber worldwide from 2013 to 2023 (Statista, 2024)



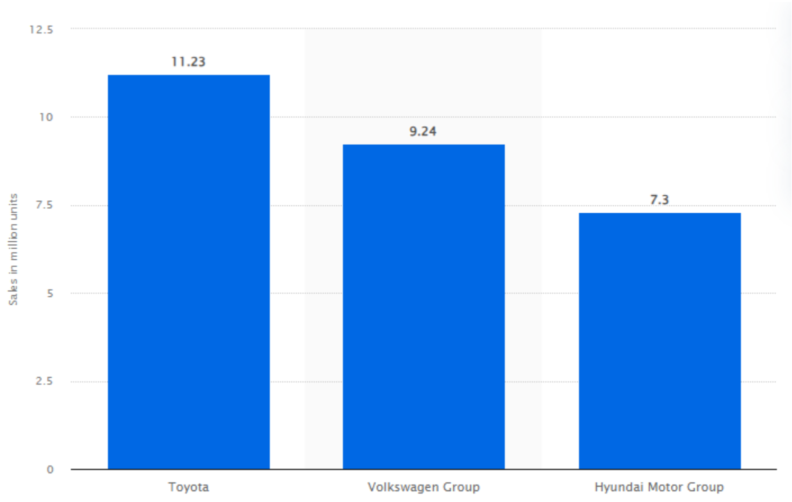
Beyond Uber's success in just a few years, its strong application of big data for many is destined to radically change the transportation industry, but not only that. In fact, former CEO and Uber co-founder Travis Kalanick said that services like Uberpool will help to significantly reduce the number of cars on the roads, especially in major cities, and that in the future "people will not own cars, they'll have a service that takes them where they want to go, when they want to go there. And that's what Uber is" (Rosamond, 2016).

5.2. Case Studies - In Depth Interviews – Big data in Toyota

In order to further study big data and its impact on business performance, I performed a series of interviews at Toyota's European headquarter.

The Japanese car company represents as of today the largest car manufacturer in the world, in fact in 2023 Toyota was the company that sold the greatest number of cars worldwide, more than 11.2 million, ranking first among all car manufacturers, followed by Volkswagen Group (9.2 million) and Hyundai (7.3 million).

Figure 32. Ranking of vehicle manufacturers worldwide by global sales 2023 (Carlier, 2024)



Toyota was founded in 1937 by Kiichiro Toyoda: the headquarters of the Japanese company, Toyota Motor Corporation (TMC), is precisely located in Japan, in the city that is indeed renamed Toyota City; but the company has sales operations in more than 170 countries and regions around the world, and for that it has a number of regional headquarters that in turn manage Toyota's operations in different countries. For example, the European headquarters, Toyota Motor Europe, is located in Brussels.

Figure 33. Toyota’s regional headquarters (Toyota Motor Europe’s website)

Toyota's cars and services are sold throughout Europe, but the way they are marketed and the relationships with customers are different in each country. The so-called NMSCs (national marketing and sales companies) and dealers approach customers in the right way for each market.

European operations extend to every corner of the continent and include manufacturing plants, logistics hubs, sales and marketing operations, research and development facilities, training and design centres, thousands of local dealers, KINTO mobility services, and Toyota Connected.

An important entity with increasing relevance and weight within Toyota in Europe is Toyota connected Europe whose mission is to transform mobility experiences through leveraging data from connected vehicles to generate value. "Toyota Connected Europe (TCEU) is changing the way people experience mobility. Leveraging the power of data, we create value by delivering inclusive, sustainable, and safety-focused technology solutions with a positive impact on society" (Toyota Motor Europe's website, 2024).

The automotive sector is an excellent example and object of analysis when it comes to big data and its impact on the performance of companies, this is because it is an industry that has changed a lot in recent years, mainly due to technology and technological progress applied on cars.

Car manufacturers use big data in different areas and for different purposes, for example with the aim of optimizing production, making logistics more efficient, personalizing the customer experience.

On the other hand, connected vehicles, represent a huge generator of data, we have seen this in the illustrative examples chapter with the examples of both Tesla and Uber; through the data collected from the vehicles themselves, companies can obtain and/or generate crucial information, for example, they can develop predictive maintenance models, which alert the driver about future maintenance to be performed; they can improve road safety, for example by reporting recurring problems on a particular section of road; they can provide drivers with information about their driving style, giving advice on how to improve or adapt it to reduce fuel consumption or the risk of breakdowns and malfunctions.

Through the interviews conducted, I tried to understand how big data is used during standard, regularly performed processes within Toyota Motor Europe and how it contributes to make key decisions that impact the company's performance. Based on what the literature says, and explained in chapter 3 of this research ("theoretical background"), big data and big data analytics are a very powerful and crucial tool for driving business growth, allowing, for example, to better predict the future and take necessary actions in advance, avoiding falling victim to potential disruptions, in the case of negative events, or failing to capitalize on future opportunities arising from the market in the case of positive events. Therefore, big data are often used to improve the accuracy of sales forecasts in order to avoid disruptions within the supply chain; they can be used to predict economic performance in order to ensure that targets set in advance are achieved and to suggest possible corrective actions, allowing the company to move in advance and tackle any problems before they become critical and difficult to solve.

Another example that emerged from the study of existing literature on the topic of big data concerns the identification of new products to be launched in the market that better meet the needs and preferences of consumers, through, for example, the analysis of surveys, trends, reviews, and comments written by users.

The examples just mentioned concern applications of big data that may be common to different industries, since, regardless of the products and/or services offered in the market, activities such as sales forecasting, economic performance analysis, and the study of new products and innovations must be carried out to ensure the proper and efficient running of the company; big data analysis is what enables these activities to be determinants of long-term success. Through the interviews conducted in Toyota, I tried to understand if and how big data analysis is employed to carry out these activities and identify what are the main benefits, the impact on business performance from different points of view (e.g. cost reduction, revenue increase, worker workload reduction, company reputation...) in fact, the existing literatures focus a lot on how Big Data is employed, for what activities, but has some gaps on its practical impact.

In the course of the interviews, I also had the opportunity to investigate analytics typical of the automotive industry, as the data that are collected come directly from connected vehicles; both in the existing literature and in the illustrative examples mentioned in Chapter 5. 1 of this research, it has been mentioned how cars, are a huge generator of data and thus information, the analysis of these, after cleaning and transforming them into structured and usable data, allows companies to have additional and crucial insights, and determine the best actions and strategy based on them; in this regard, we have seen how Uber and Tesla are excellent examples in the use of unstructured data coming from vehicles and not only that, but also Toyota on this topic carries out very important activities that allow it to provide its customers with relevant information about their driving experience and at the same time allow the company to collect data about drivers' behaviour and use accordingly to make decisions and adopt strategies.

5.2.1. Big data as a tool for forecasting

During the interviews, as also pointed out in existing literature, the importance of big data for making forecasts of future sales volumes often emerged (Seyedan & Mafakheri, 2020). As explained by M.M. senior specialist and project manager in the accessories department at Toyota Motor Europe, in order to efficiently satisfy the customer, it is essential to forecast, as accurately as possible, sales and the number of accessories that will need to be installed in vehicles. To better understand the complexity of this activity M.M. explained how the accessories that can be ordered by customers are “infinite” this group includes, for example, floor mats, towing hitches, Electric Vehicle charging cables, scuff plates, deck protection, trunk liner, alloy wheel, complete winter wheel ... and many others; for each vehicle in Toyota's line-up there are several dozen versions of these accessories, all of which are characterized by a code (i.e. part number) and forecasts have to be made at the part number level, so we are talking about tens of thousands of accessories, for each European market under TME's responsibility (roughly 29 NMSC); this serves to give a measure of the size of this activity.

To forecast future sales and installations and thus predict the number of pieces of each accessory, for each vehicle, for each market, massive analysis is done. In order to meet future customer demand, it is necessary to analyse both existing data (past data) and other forecasts and estimations. The data used to be able to carry out accessory forecasts include data on past sales of both accessories and vehicles and therefore on the installation rate for each part number, as well as current data that is captured in real time as soon as the sales are recorded; data on installation locations (hub or dealer) since the installation capacity varies depending on where it occurs; but in this forecasting methodology, future data, and so other estimations, regarding future vehicle sales volumes are of course also considered, for each market; in addition, at its discretion, each market may decide to launch promotional campaigns on certain accessories, or change their strategy stop selling a particular accessory or part number thereby affecting future volumes. All of this data and information, is aggregated and used as input to forecast sales for each part number over the next 6 months, so that it is possible to have transparent communication with suppliers and ensure the supply of accessories.

In order to focus on critical products with a significant impact on the accessories business, a committee, SSIC (supply sales installation committee), was created with the very aim of fostering business growth through a more accurate estimation of future volumes, avoiding in occurrence of situations of inability to meet customer orders, i.e. backorders, but also situations in which many more products are produced than actually required by the market, generating therefore overstock; a correct estimation of such volume allows to maximize customer satisfaction and storage capacity.

This process of forecasting accessories requires the participation of many stakeholders, both internal and external to the company, and it obviously impacts them; M.M explains how "Accessories and parts is really something which is generating profit, and because we are all profit driven as we can imagine, a lot of parties are involved to ensure this profit." (Simone - M.M. interview) among the many teams involved obviously there is the logistics team, which has to ensure that there is the necessary space to store the parts that then have to be shipped to the customers, TPCE (Toyota parts centre Europe) is in charge of these logistics activities; the Purchasing department is heavily involved, as it is in charge of managing the relationships and connections with the suppliers, determining the purchase prices based on the expected volumes, e.g. how many parts we will buy and how often we will buy them. If we talk about the purchase price, we can't not also talk about the selling price, so the pricing team that precisely has to determine the price at which parts will be sold, often involved in sell out campaigns where you want to stimulate demand for certain products in stock through temporary price reductions.

The sales department is a very important stakeholder, as it is responsible for providing data on future vehicle production and sales volumes, based on which accessory sales forecasts are then calculated.

Accessories must then comply with precise regulations, which is why the homologation team is a key part of ensuring that components comply with regulations, and this requires attention to both their design and development. At the same time, the styling team is responsible for making the accessories aesthetically appealing, because we know how important it is for customers to buy something that is not only useful, but also attractive.

Among the stakeholders who are directly impacted by this analysis, of course, are all those people who are in charge of installing the accessories, whether in the production plants, in the Hubs or directly at the dealers, a correct estimation and consequently a correct supply of the necessary accessories and parts allows them to complete the set up of the vehicle, which can then be sold; in fact, if, for example, the Hubs do not have all the accessories that need to be fitted on the vehicle, depending on what the end customer requires, the vehicle cannot leave the plant and is therefore blocked,

In conclusion, as is obvious, perhaps the most important stakeholder directly impacted by this forecasting activity is precisely the end customer, as explained earlier, in order to ensure that the vehicle is sold and equipped as per the customer's request, all the parts and accessories must be in the right place at the right time, and this is possible only and exclusively through careful forecasting activity, which involves a great many stakeholders.

This type of analysis is defined by M.M. as a combination of descriptive, predictive, and prescriptive analysis, in that the starting point is certainly what has happened in the past, the trends that have generated previous sales of different part numbers in different countries, but based on past and current data, combined with other estimates and forecasts, the prediction of future accessory sales volumes is made. But these forecasts actually provide input on what actions need to be taken in order to ensure an accurate and smooth supply.

The data analysis performed in this process of forecasting accessory volumes, carried out in SSIC, starts by consolidating information in different databases, which are managed through queries to then create pivot tables and dashboards useful to better understand the data.

Mainly Microsoft tools are used, and some of the vehicle data is processed in TM1, which is part of the same Microsoft ecosystem.

A significant portion of the analysis is done through external tools for more advanced statistical analysis. For example, SO99, a system that helps make forecasts based on statistical data, is used so that we can improve our planning capacity.

This type of analysis has a very strong impact on company performance, in fact the focus of this activity is on enhancing efficiency, minimizing errors, and speeding up processes to ensure greater accuracy and effectiveness. By doing this, TME can save money, manage business growth effectively with the current workforce, and maintain robustness in its operations. This continuous research to improve existing processes, and to maximize the efficiency of its operations, as explained by M.M. during the interview, is part of Toyota's DNA, in fact in Japanese culture there is a term, "Kaizen" which means indeed "continuous improvement" which is at the heart of Toyota's philosophy as it is always trying to do things better, more efficient, quicker, and more profitable. *"Kaizen (Continuous improvement): A philosophy that helps to ensure maximum quality, the elimination of waste, and improvements in efficiency, both in terms of equipment and work procedures. Kaizen improvements in standardised work help maximise productivity at every worksite"* (Toyota Motor Europe's website, 2024).

Drilling down into the specifics of how and why this type of analysis impacts business performance and in what areas, the following practical impacts were identified in the interview with M.M:

- *Costs reduction and revenues optimization*

There are two main ways that can generate a loss of money.

The first occurs when you are unable to deliver a part, leaving the customer dissatisfied and potentially driving them away; this demand dissatisfaction can occur when the expected quantity is less than what was actually demanded. As a customer, you expect to receive the part you ordered; if this does not happen, you are likely to be disappointed and choose not to return to the company, that in the case of Toyota mean that the customer may decide to buy non-genuine accessories. Whenever an accessory goes on backorder it creates a negative perception of the company, giving the impression that you are unable to meet customer expectations, but more importantly this situation leads to a lost opportunity, as a secure revenue was not registered because there was not an accurate forecast in advance

To address this problem, TME through the use of big data to make accessory forecasts, works to prevent backorders by improving forecasts and maintaining adequate stock levels. This ensures customer satisfaction, and the ability to fully meet demand.

The second challenge concerns the overstocking. If, on the other hand, much fewer accessories and parts are ordered than planned, they still have to be stored somewhere,

but storage space is limited. Continually expanding storage space involves significant costs, including rent, electricity, and labour, in addition to the physical land required.
In a nutshell, a balance must be established: avoiding losses due to dissatisfied customers because of stock shortages while reducing the costs associated with excessive inventory. This balance is essential to maintaining efficiency and profitability, so an accurate and structured forecasting process is critical to reducing inventory costs and lost opportunities.

- *Reduction of workers workload*

Correct estimation of future sales makes it possible to have all the accessories in the right place and at the right time, this ensures that the people who are in charge of installing the accessories on the vehicles do not have to spend their time in waiting for the procurement of these parts delaying all the operations and delivery to the final customer, at the same time, the storage of accessories takes time as well as costs, so the more accurate the forecasts are, the less time needs to be spent in storing excess parts.

- *Strengthening corporate reputation*

Having a correct forecast of future sales, with all the consequences it implies, also supports the company's position in the market.
M.M. says that "Good reputation Pays. We know that good reputation and connections is 60 to 70% of the success; and whatever activity we do is all linked to revenue and to the reputation that we create as a Toyota".
Therefore, reducing backorders, and so enhancing customer satisfaction, also contributes to strengthen Toyota's presence in the European market. Customer satisfaction plays a key role in promoting the company's growth. That said, it is not only SSIC that determines the company's position in the market. There are many factors at play, but in the end it all comes back to the goal of ensuring customer satisfaction. Obviously, an important step in any process, analysis and beyond, involves measuring results, validation of what has been done and is being done to measure its effects and act accordingly through improvements if needed.
To measure the effectiveness of accessory forecasts, M.M. explains specific KPIs are defined and accessed, such as the number of backorders, inventory overstock levels, and

overall forecast accuracy through "forecast vs. actuals" comparisons. In addition, vehicles stuck at installation locations are monitored to see if the number of blocked vehicles is gradually decreasing.
In addition to this, KPIs on economic performance, for example sales turnover vs. target, are also considered; these are particularly important since, as mentioned, the ultimate goal is always profit generation. Moreover, in logistics, it's important to establish KPIs such as installation location capacity, the number of vehicles processed for installation each month. Regular follow-ups are critical to stay on track and achieve the goals.

A careful and accurate planning of future volumes through analysis of the big data available to the company, however, emerged during the interviews as a key activity not only for the accessories business, but also for the vehicles themselves. In fact, L.C., a graduate in the Supply Planning division at TME, highlighted the importance of big data in determining the volumes of vehicles to be allocated to the various European markets.
The team in charge of allocations must ensure, each month, a fair and correct distribution of vehicles to the various NMSCs for the following months, and to do this they make extensive use of the big data available, from many sources and in different formats. L.C. in the interview highlights the crucial importance of analytics within this process, as allocations must be extremely precise in order to satisfy end customers and maximize sales "because every unit that is sent to every NMSC counts of course" (Simone - L.C. interview).

Among the various data that are considered, to establish vehicle allocations, based on the model and specifications required by the market, there are for example customer contracts, which are precisely agreements signed with private individuals, which include a series of information, (in text format) regarding the configuration of the car, such as engine type, transmission, interior trims, exterior features; it is essential to follow these specifications to ensure customer satisfaction. As mentioned precisely all this information is captured in text form; the contracts are uploaded regularly by the respective NMSCs, and then structured with a code that indicates the exact configuration of the vehicle through an internal tool called A2P that then makes the data included in the contracts, standard and usable.
Other data that are analysed are local stock levels (LSL); this reflects how many vehicles are stored in the yards or hubs of an NMSC; It is very important to keep stock levels

within specific thresholds to ensure availability and improve 'cost efficiency. To give a practical example: on some models, the goal is to keep the LSL between 1.2 and 1.6 months. If the stock level is 1.6 months, it means that even in case of a production shutdown, the units available in depots or at retailers could sustain sales for 1.6 months before running out. Exceeding this threshold leads to overstocking and thus costs, while falling below it leads to the risk of supply exhaustion and thus demand dissatisfaction.
Another key KPI is the sales plan, which predicts the number of sales expected in the next 6 or 3 months, depending on the type of plan. For example, if the forecast shows 1,000 vehicles sold in a given NMSC, we need to ensure that these vehicles are allocated accordingly. Failure to do so would result in missed targets, reduced performance, and financial losses.
This process allows to do predictive analysis, the data available already allows future decisions to be made, for example, if the sales plan of a certain country has 3000 vehicles sold in a particular month, this already suggests to allocate 3000 units to it.
there is a trend suggesting that in July Ireland and the United Kingdom will sell many more cars than in June, because in July the license plate number changes, and if the license plate number is the new one, people are more willing to buy a car because it looks new because of the new license plate, and therefore the sales plan increases by four times. Therefore, it is necessary to make sure that there is enough production in the previous months to support sales.

L.C. provides specific examples of how these analytics have improved the company's performance:

- *Cost reduction and revenues optimization*

In Toyota, the concept of "just-in-time," coined precisely for the first time in the Japanese company, has a certain relevance; in fact, it states that production is based on actual demand, with the aim of minimizing inventory and producing only what is strictly necessary: "if we make a mistake in calculating what the customer will want then of course we have to scrap those units, there are costs associated with the fact that we are producing units that are not sellable" (Simone - L.C. interview).
For this reason, accurate forecasts are absolutely crucial, especially during periods of change for a vehicle, such as a minor change, a new generation, or any significant

improvement. Whenever a model is subjected to changes, such as new colours, engines, or particular features, forecasting demand for the previous version must be extremely accurate to avoid overproduction and waste, reducing costs and ensuring a smooth transition to the new model year.
On the other side of the coin, a forecast that reflects demand allows you not to miss sales opportunities, thereby satisfying the customer and maximizing revenue.

- *Customer satisfaction*

The ability to predict more precisely customer demand in the future allows cars to be produced more accurately. This means that the customer will have to wait less time to receive their car. “If the customer receives the car in a shorter period of time than expected he will be a happy customer, and a happy customer is a returning customer” (interview with Simone - L.C.), so accurate forecasts and timely deliveries increase customer satisfaction.

- *Employees workload*

“Every month we receive an aggregate number of orders from the NMSCS imagine if we had two single handedly look at every single customer contract for every single NMSC for every single car. That would be absolutely impossible. That would be huge workload, basically impossible for anybody” (Simone – L.C. interview).

The availability of effective tools ensures that the necessary data are provided accurately and in detail, enabling efficient management of vehicle allocation. This access to aggregated and accurate data significantly reduces the workload and improves the efficiency of the allocation process.
if we extend the view beyond Toyota Motor Europe employees to include those employed in the hubs and dealers, the impact of this analytics on people's workloads becomes even more evident. If these people are informed in advance about the number of vehicles they will receive, they can better plan their workloads. Therefore this proactive approach helps streamline operations and reduce the overall workload of employees involved in the later stages of the automotive supply chain.

- *Corporate reputation*

L.C. points out that Toyota is recognized for being one of the automakers with the shortest waiting time between order and delivery, with an important (positive) gap against others. This in fact has a positive impact on the company's reputation. “Toyota is capable of delivering a vehicle in approximately two to three months, depending on the model and market conditions” (Simone – L.C. interview).
This efficiency is attributed to the use of extremely accurate data, allowing for predictions of future customer demand that are close to perfection. This ability, as a consequence, facilitates coordination with suppliers and factories regarding the feasibility of customer requests.
In contrast, other manufacturers may take five to eight months to deliver their vehicles, because of issues like improper specifications, supply chain disruptions, factory shutdowns, and other challenges. Long waiting periods can lead to customer dissatisfaction and negatively impact brand reputation.

These examples, in which big data are applied to predict and anticipate demand, both of the accessories and the vehicles themselves, with the associated positive consequences, demonstrate how important it is to “play it by ear,” and this is only possible through predictive analytics such as those explained by L.C. and M.M..
However, these analytics also have limitations: while it is true that many things can be anticipated, it is also true that not everything can be predicted, even through precise and advanced tools. Unfortunately, there are external factors and events, such as wars, pandemics, and others, that generate a huge impact on demand and the market, but these are outside of any kind of prediction, so rather than trying to predict them, it is critical to figure out how to mitigate them.

5.2.2. Vehicles as generators of big data

As seen in the illustrative examples of both Tesla and Uber, the data that companies can use as input to their strategic decisions can also and mostly come from sources that are not exactly "traditional". Cars, and in these cases vehicles, are exceptional generators of data and thus information, which are collected by companies and received very often, in an unstructured form. This data is used for a wide variety of purposes, allowing automakers, for example, to learn more about their drivers and to provide them with valuable statistics about their driving styles and overall experience.

Toyota also collects a lot of data from its connected vehicles, in order to better understand the use and potential of this data I carried out an interview with A.U. manager in Toyota Motor Europe.

A.U. is responsible for B2C connected services that focus on improving the “out-of-car” customer experience, which is primarily concerned with providing services to customers when they are not inside their vehicles. Although some aspects of the in-car experience are considered, the emphasis remains on out-of-car functionality.

These services are accessed primarily through apps such as MyToyota and Lexus Link+. These apps serve as the main channels through which customers can interact and use connected services.

The process involves defining the features to be offered, overseeing their development and implementation, and launching them effectively through these apps. The goal is aimed at meeting customer needs while improving the overall experience.

A.U. explains that through the data that is collected from vehicles, Toyota provides its drivers with a number of features, accessible through the apps mentioned earlier, that enable them to improve their experience with their cars and beyond.

A typical feature is “Find My Car,” which uses geolocation data to help customers identify their parked vehicle. The system shows the location of the vehicle on a map, simplifying the process of finding it.

Another feature is “Trip History,” which provides information on previous trips, driving behaviour, and calculates a driving score. For example, the system can compare scores between trips, highlighting differences such as excessive acceleration that can negatively affect the score. In addition, the app offers recommendations for improving driving habits, contributing to greater efficiency, reduced cost of ownership, lower fuel consumption, and lower CO2 emissions.

For trip planning, particularly for electric vehicle (EV) users, there are features that provide detailed information on charging stations. These include availability, prices, and whether the station is in a public or private area, supporting more informed and efficient trip planning. Depending on the data on previous trips and related driving styles, the customer will also receive information and advice on how many kilometres they will be able to travel before they need to recharge their vehicle and where it is convenient to proceed with recharging, thus providing a detailed plan on how often to stop and recharge the vehicle, where, and what driving style is best to adopt in order to save money and not deviate from the plan provided.

A.U. highlights how through data from connected vehicles, predictions can be made about future events related to the condition of the car, which allows for more predictive analysis, such as predictive maintenance. Typically, customers should be asked to do maintenance once a year or when they reach 10,000 kilometres. Through some logic and mechanisms to predict this based on the number of trips and average daily or weekly mileage. The system estimates when the vehicle will reach 10,000 kilometres and communicates that date to the customer. For example, it might communicate to the customer, “In 25 days it will reach 10,000 kilometres. Please book maintenance.”

A.U. explains that the data received from the car is very descriptive and unstructured. However, algorithms are applied provide prescriptive and predictive analyses such as those related to vehicle maintenance.

Through this data as already mentioned Toyota calculates for each driver, a score, according to his or her driving style, which considers how the driver drove on previous trips, whether he or she made accelerations, swerves, sudden braking, whether he or she maintained a constant speed and within limits, and much more, all of which makes it possible to calculate a proper score, offering the driver advice on how to improve it in order to reduce the vehicle's usura time (e.g., tires in case of sudden acceleration) to reduce fuel consumption and prevent possible damage to himself, the vehicle, and the surrounding environment.

Currently, says A.U., studies are being made on how to predict when brake pads need to be replaced or when it is time to change the oil based on the data received from the car. For example, if it is observed that the customer drives very aggressively, the time for an oil change might come earlier than usual. Similarly, if the customer brakes very hard, this data can be used to recommend a preventive brake condition check at a repair shop.

A further implementation that is really a “buzzword” is the so-called V2X technology, Vehicle to infrastructure; which would enable reporting, for example to the police or appropriate entities, sections of road where accidents or damage to vehicles occur on a recurring basis, so the aim is to "sell" the data collected from vehicles. currently there is low demand from governments as it is really at an early stage, but it is an implementation that is on Toyota's target map.

Another goal of Toyota concerns the analysis of customer satisfaction through machine learning techniques; in fact, the App Store contains numerous customer comments that can be leveraged for this analysis; a country-by-country review is carried out in order to identify customer sentiments, strengths, and possible areas for improvement. Different categories are applied to the feedback, such as issues related to remote control functionality or app design. This approach aims to identify the specific areas in which customers express dissatisfaction, allowing necessary improvements to be applied.

The benefits generated by these analytics, which use data from the machines, are significant, impacting both Toyota itself, the dealers and workshops, but most importantly the end customers; in fact, A.U. states that: "When we develop connected services, our goal is to provide benefits, whether to the customer, the company, the retailer, or all of them together." (Simone – A.U. interview). One practical example, where both customers and Toyota retailers are happy, concerns precisely maintenance reminders: "we have this prescriptive mechanism so customer has less anxiety. He always knows that he will be informed."

This mechanism turns out to be very convenient for both the customer but at the same time for the retailer, as they can better schedule capacity in the workshop but, most importantly, they make sure that the customer comes back to them and not to someone else.

Cerchiamo di strutturare anche in questo caso, i benefici generati dai big data provenienti dai Vecoli connessi:

- *Revenues & cost optimization and customer satisfaction*

Maintenance remainders obviously belong to this group, as they provide Toyota dealers with income from returning customers.

Another good example of how vehicle data are used is "driving coaching." This feature enables to build a score (from 1 to 100) on driving behaviour, which is used in a special insurance program offered to customers. If a customer drives safely, his or her driving score improves, giving him or her access to higher insurance discounts.

This approach generates new revenue streams and attracts customers who prefer Toyota insurance over other insurance programs. It is a win-win solution: the customer benefits from lower insurance premiums, the retailer increases its business, and Toyota strengthens customer loyalty and satisfaction.

The driving behaviour score is based on four main criteria: speed, how aggressively you accelerate; braking, measures the aggressiveness of braking; curving, measures how aggressively the driver takes curves, such as at roundabouts considering angular acceleration; and finally constant speed, the system encourages customers to keep a steady speed, as frequent stops and acceleration can indicate risky behaviour.

Drivers who show a low score are classified as high-risk and thus risk not being able to obtain insurance discounts.

Currently, the R&D team is developing a proof of concept to improve these ratings by adding new parameters. This is particularly relevant since the main business of insurance is risk prediction, i.e., classifying drivers as low, medium or high risk.

With the help of Toyota's safety systems, the team is working to evaluate driver behaviour more accurately. For example, efforts are being made to include in the analysis parameters such as the average distance the driver maintains from the vehicle ahead and how frequently the brake assist system is activated. This data makes it possible to identify high-risk drivers whose safety systems activate frequently, suggesting potential distraction. This increases the likelihood of accidents or vehicle damage. The goal is to incorporate these additional parameters into the overall score formula, further improving the risk prediction model. Thus, this feature allows Toyota to optimize its costs, as it identifies particularly hectic and "risky" drivers in advance, but at the same time encourages Toyota and Lexus drivers to join Toyota's insurance program, generating new annuities, and if they adopt a safe driving style they can get discounts that increase their satisfaction.

- Employees workload

"There are different way we apply the data so our data, business team for example they recently introduced a solution thanks to the data we have that they can reduce the workload of people in the port and hubs"

Historically, staff at Hubs such as Zeebrugge had to manually check thousands of vehicles as they arrived. Workers would do a "walk around check" around the vehicles in all weather conditions to see various parameters, such as any damage or scratches. This process was time-consuming and very hard work.

Now sensor data from the vehicles are used to determine their condition, identifying whether they are damaged, scratched or in good condition. In addition, you have detailed information about the exact time each vehicle arrives in Europe. Upon arrival, the vehicle data communication module activates, connects to the network and identifies itself, "I am the module with this SIM card, installed in the vehicle with this VIN number" (vehicle identification number).

This system eliminates the need for a physical inspection just to confirm the vehicle's arrival.

A.U. says that through this digitization, the workload has been significantly reduced. Instead of manually inspection of each vehicle, staff can now refer to a back-end system. The team developed an algorithm and dashboard that provide a real-time overview of vehicle arrivals, listing details such as VIN numbers and vehicle condition. If the system flags a vehicle that needs verification, only then is it manually inspected.

This approach has streamlined operations, saving time and effort while maintaining high quality standards.

- Corporate reputation

In the past, much data from vehicles has been used to understand how Toyota's hybrid cars performed. In particular, data were collected on the distance and time customers spent driving using their hybrid powertrains compared to petrol engines. This data played a crucial role in establishing Toyota's reputation as the number one hybrid car brand in the world.

The marketing team used this information to communicate a strong message: our Toyota hybrid drivers drive practically as if they were driving electric vehicles (EVs) because their fuel consumption is so low. In fact, they cover about 60 percent of their distance using the hybrid powertrain. This data helped Toyota to build a powerful marketing narrative and provide factual evidence to support it.

In addition, this data was also used to promote hybrid vehicles at points of sale. A program called “Test Drive the Hybrid.” was implemented: When customers went to the dealer to look for a car, they could take a test drive, and information was immediately shared about the efficiency of their trip compared to a normal trip in the same model equipped with a petrol engine.

Direct comparisons were then made, highlighting the differences in fuel consumption and CO2 emissions between hybrid and non-hybrid models. This approach not only informed customers but also encouraged them to consider the benefits of hybrid driving.

5.2.3. Big data to steer the business performance

During the interview performed with A.S. specialist in the accessory planning and management team in TME, it emerged how important it is to have the most accurate and detailed data possible in order to analyse economic performance, foresee any problems or bottlenecks, and consequently, take the necessary actions and countermeasures to solve them and achieve the planned targets.

A.S. also explains how it is not enough to have the data at hand, it is necessary to visualize them in the best way and use the right tools that allow a detailed and complete analysis, of each and every aspect.

A.S. throughout his career at Toyota Motor Europe has been responsible for the “styling” category of accessories, i.e., those accessories whose function is mainly related to aesthetics (e.g., scuff plates, side sills, alloy wheels); her main responsibility was to ensure that the styling category for accessories was set up correctly. This included making sure that NMSCs associated with Toyota Motor Europe were able to sell and order these accessories effectively.

A significant challenge during that period, A.S. explains, was dealing with delays in the supply chain while developing a strategic approach to the styling category. This involved analysing the specific needs and goals of each of the 30 markets in which TME operates and defining a tailored strategy for each product within the category.

To ensure smooth operation of the accessories business, the various responsible people require data from different sources, both internal to the company such as historical data on sales of each part number, for each model, for each NMSc, data on vehicle sales

volumes, annual targets, and sales plans, but also external sources such as data from the various partners with whom Toyota cooperates.
Two particularly relevant KPIs for the accessories business are the EUR/NCS, which represents the revenue represented by accessories for each new vehicle sold, and the installation ratio (IR), which is the number of pieces of a specific accessory (part number) divided by the number of vehicles sold that are compatible with that accessory (i.e., a part number with an IR = 100 percent means that for each vehicle X an accessory Y was sold, so if 100 vehicles X are sold, at the same time 100 accessories Y are sold).
In order to better access all the data needed to monitor economic performance, a series of reports were created within the APM team on Power BI through a connection to databases containing these data through SQL analysis services, "This database doesn't just include accessories data but also encompasses sales data, pricing data, ordering data, and other information. To address this, we developed the Power BI dashboard to streamline our work" (Simone – A.S. interview).
the use of software such as power BI in fact allows the creation of automated and customizable reports, so all people have clear and standardized access and visualization of data.
This analysis tool plays a key role in the management of accessory categories or specific vehicle models, it is in fact a tool used daily for all operations as it is updated in real time, as soon as new data are recorded, and all views represent the latest version of the accessories business. Its main purpose is to provide a clear view of trends for each product, and to enable constant and detailed monitoring.
This makes it possible to identify patterns and understand the reasons behind them. While some results are in line with targets, others may deviate from them, indicating potential problems that require investigation and solutions.
Users can access the system at any time of day, be informed of any changes and react proactively if necessary.
This capability makes the analysis tool a crucial resource to maintain control, promptly address discrepancies, and support informed decisions.
This approach and tool enables several key activities to ensure a profitable performance of the accessories business; a key aspect of this process is to understand market trends.
Effective category management requires a clear and detailed view of the current situation

in all markets. This makes tools like Power BI a must, since they provide automated and accurate visualizations for each market. As they eliminate the need to manually create separate visualizations, such tools simplify the process of monitoring and analysing trends, enabling more efficient management.

Another significant activity involves projections on the evolution of economic performance in the following months. through the inclusion of targets directly in dashboards, the tool ensures that teams can constantly assess whether they are on track to meet year-end targets. This enables timely action if projections fall short of targets. The information accessible through power BI is critical to prioritize focus areas and actions to be taken; for example, the identification of underperforming products or categories becomes significantly easier with clear visualizations of country-specific performance.

A.S. also specifies what are the improvements, that are currently being investigated, that can further improve this analytics; for example, the creation of flows on power automate can allow to send, for example by email, with a specific recurrence (weekly/daily/monthly ...) the data stored in the power BI, pre-filtered according to the recipient, making the analysis process even faster and more immediate. Another improvement, however, refers to the enhancement of AI within the power BI; the goal is to interview the AI and then interact directly with it and no longer with the report itself.

The impact of this analytics process on operations can be evaluated in a number of ways, both direct and indirect. Although it is difficult to directly attribute increased revenues or reduced costs to these tools, they have undoubtedly brought significant time savings, reduced waste, and reduced workloads. For example, the time that previously was spent to prepare reports or to manually verify information, has been drastically reduced (i.e. an estimated 50 percent from the daily workload). This freed-up time allows teams to focus on other crucial activities, like addressing problems visible by consulting Power BI and implementing strategies to improve economic performance.

This lower workload, therefore indirectly generates a positive impact on cost reduction, because category management and daily data reviews, are simplified; thus teams can identify problems and alarm bells, before they become critical. For example, supply chain issues can now be detected early, thereby taking preventive rather than reactive action. In the past, with less efficient tools such as Excel, the high workload often led to delays in

the identification of problems, sometimes up to a month after they arose. These delays negatively affected the company's revenues and even brand reputation.
In addition, the consistency and accuracy of the data presented strengthened communication both internally and with customers. Standardization of visualizations ensures that all stakeholders, from management to customers, have a common understanding of performance metrics. Negotiations and discussions now begin with clear and shared data, improving service quality and fostering trust.
In summary, even though the financial impact of these tools is likely to be less visible, their role in operational efficiency, proactivity, communication, and service quality is unquestionable.
These changes not only make internal processes smoother, but also help position the organization as an innovative leader in data-driven management.

5.2.4. New products development

Existing literature has often shown how big data analytics can be used as a tool to support the design of new products to be launched in the market (Zhan et al., 2016). Indeed, BDAs can make it possible to collect and analyse consumer feedback, and translate it into information that enables the identification of new products that better meet customer desires. Often this information comes from unstructured data, in text format, which then needs a transformation aimed at analysis; this text data can be reviews left on the Internet, comments on social media, reviews even in video format.
However, this type of approach is not easy to implement, and obviously cannot totally cover the process of bringing a new product to market, which obviously includes several stages, from market research to product ideation, and actual development.
To explore this type of process and understand how it is carried out in the business, and in this case in Toyota Motor Europe, I interviewed K.K. project manager responsible for the entire "newness" process in the accessories department.
This process encompasses all those steps that have as their ultimate goal to the identification of new accessories to be sold with the car and that can obviously differ depending on the customer, in this case the target market, so a new accessory can be

launched for example in Italy but not in the Nordic markets (Sweden, Norway, Denmark ...) depending on the demand and therefore the interest of consumers.

K.K. explains that this process is divided into 4 distinct phases:

- Ideation: this phase consists of identifying potential new products to be introduced to the market, based on proposals that can come from both internal stakeholders (purchasing team, design team, and external stakeholders, the NMSCs representing the voices of consumers in the different countries involved)
- NPRS: this is a framework, a tool to evaluate whether or not to proceed with a given idea. It is overseen by an L3 Committee, which analyses the feasibility and priority of ideas. This assessment is essential because of the presence of numerous ideas in the pipeline, making it necessary to determine whether to pursue a specific idea or allocate resources elsewhere.
- Feasibility study: at this stage, the feasibility of the investment is checked in order to better assess the quality and priority of the ideas and decide whether to proceed with development, a score is given to each potential new product. This scoring system is based on several variables. One of the key parameters is the reason for developing the accessory; for example, if the development is motivated by a legal requirement, the idea receives a high initial score. Another parameter is the method of installation; there are three options: in-line installation (done at the factory), installation at a hub, and installation by the dealer. Also considered is the development effort required to design the accessory, classified into four levels: very easy, easy, medium or difficult, based on the time and resources required for development. Potential volumes and prices are additional parameters considered in the evaluation. These variables collectively determine the final score for each proposal.
- Development stage: after the prioritization of ideas on the basis is carried out, the development of the identified accessories proceeds.

In the past, Toyota has involved external companies or partners to identify markets for accessories through more automated processes based on big data analytics where online comments and feedback were considered. However, the standard process involves investigating customer desires through input from NMSCs who propose their ideas directly. A web-based SharePoint site is used to distribute the questionnaires, which include approximately 10 questions. Customers are invited to share their ideas, provide feedback on the proposals, and indicate suggested volumes, installation rates, and prices. The data are downloaded from the website and evaluated to proceed with the evaluation of the ideas, determining which of them deserve a feasibility study and which should be rejected. Feedback from NMSCs is used to prepare a high-level business case.

The process then begins with a comprehensive list of potential new accessories, from which decisions are made regarding which ideas will proceed to the next stage.

K.K. explains that the duration of the complete process, may vary depending on the product and its complexity, but for example, for a product with a non-high complexity rate the minimum duration is one year.

K.K. also specifies how the possibility of automating phases of the Newness process is being investigated at this time, pointing out how this could have a strong positive impact on the workload of all engaged stakeholders. The stage that could benefit most from automation and introduction of BDA is the ideation and evaluation of ideas. It can help reduce business costs because, by identifying the right products, the company's efforts and expenditures can be focused more effectively; in addition, the correct identification of the right products and thus "capturing" customer desires enables additional revenue generation. Thus, identifying the best ideas through automated and structured processes increases their potential success.

From a workforce perspective, reducing workloads can lead to improved efficiency and, as a result, a positive impact on corporate reputation. Simplifying processes can facilitate better handling of customer requirements, which are sometimes obstructed by complexity. An automated system will help meet customer requirements more effectively.

Ultimately, improving customer satisfaction should enhance the company's reputation. This example from Toyota on how new accessories are determined to be developed shows how, the application of BDA is not so straightforward, its potential is often recognized and companies aim to maximize their potential, but at the same time, their implementation is not so simple and quick, clear and defined strategies and requirements are needed, as well as technical expert

6. CONCLUSIONS

"Data is the new oil." This statement, which introduced this analysis, is fully reflected in the conclusions of this thesis. After extensive research on the use of data, and in particular big data, it was possible to identify a clearer and more complete picture regarding its impact on the strategic decisions that companies face on a daily basis.
The results confirm what had been hypothesized: the ever-increasing volume of data, from increasingly heterogeneous sources and in less structured forms, represents an invaluable potential for companies. However, the value of this data is non-existent unless accompanied by adequate analytical capacity. Information must be extracted from the data and processed in order to drive operational and strategic decisions that positively impact business performance.

Big data analytics is not the solution but the means, an essential means to achieve goals and drive business economic performance. As highlighted in the illustrative examples, companies such as Netflix, Tesla, Uber, and Airbnb are clear examples of how the extensive use of big data can create disruption in their respective industries, making them almost untouched market leaders. These success stories, by the way, do not come only from the availability of big data, but more from the ability to implement advanced tools and strategies to pull maximum value from the data.

Interviews carried out during the research highlighted how big data is used for a number of business activities. Among these, sales volume forecasting is a crucial example: by analysing historical data and market trends, companies can optimize supply capacity and anticipate any critical issues in the supply chain.

In parallel, big data is applied in performance monitoring and in stimulating the achievement of set targets. This approach not only improves operational efficiency, but also enables timely reaction to changes in the market.

Particularly interesting was the example of data coming directly from connected vehicles, this showed how data also comes from "unconventional" sources but allows for the collection of information in real time and use it for as input to strategic decisions that allow, for example, to expand business and increase sales.

A key finding of the study concerns the practical consequences of using big data analytics in different business activities. Among the most significant and recurring impacts are reduced costs and increased revenues, achieved through improved resource allocation and identification of new business opportunities. In addition, employee workload can be significantly reduced through automation of repetitive processes. These benefits are also reflected in customer satisfaction, improving the overall experience and enhancing corporate image and reputation.

However, the adoption of big data is not without its challenges. Among the recurring problems is the development of appropriate skills within the organization. In addition, effective data governance that ensures that information is not only secure and private, but also of good quality and reliable, is critical to adoption.

These include the need for a strategic vision to guide the adoption and use of big data. Success certainly does not come only from technology, but also from the ability to incorporate it into business processes and, therefore, to generate a data-driven culture that involves every level of the organization.

No less relevant is the impact of big data on innovation. The ability to analyse large volumes of data in real time enables companies to identify new and recurring trends, research the launch of new products or services, and meet evolving market needs more quickly. In an increasingly competitive and dynamic environment, this can be a major strategic advantage.

The conclusion of this research shows how big data and related analysis techniques have become the key tools for business success in today's competitive world.

However, its actual value depends on the ability of companies to integrate it effectively into decision-making and operational processes. The examples analysed show how big data can be a key differentiator only if supported by clear strategies, and appropriate skills.

Looking ahead, it is normal to expect an increasingly central role for big data analytics within business operations.

Artificial intelligence, machine learning, and emerging technologies will make analytic capabilities ever greater, thus generating new opportunities for companies that show themselves able to take advantage of these innovations. In a world where data is now an indispensable resource, and used for every activity, the main challenge for companies will be not only to collect and analyse data, but to translate it into concrete actions that generate maximum value.

As anticipated in the introduction chapter, the existing literature focuses heavily on the theoretical aspects of BDAs, also pointing out possible applications of big data within the enterprise. As far as the academic implications are concerned, this study aims to contribute to academia by bringing a more practical view of the application of big data in business processes, highlighting real cases of how big data can be integrated into business processes and decisions, but also the obstacles and complications that may preclude their application. This was done both from a more distant point of view through the illustrative examples, as well as through a real zoom inside an automotive big player such as TME.
In terms of managerial implications, it is clear how companies can, through this study, find references, of "extreme" companies (NATU) that have made big data the key to their success, which has allowed them to become leaders in their sectors; at the same time, this study through the case studies carried out with in-depth interviews, takes a closer look at the activities that can be supported and improved through the use of big data by highlighting the benefits generated.
Regarding the limitations of this research, the interviews were all carried out within the same company and therefore within the same sector but as anticipated, the topic of big data is common in many if not all sectors, and the activities that can be improved through their application in many cases are the same between different sectors (forecasting, performance monitoring and steering, new products development) that thus offer different products and/or services. In any case, possible future developments of this research may include interviewing players from other sectors as well to confirm precisely the commonalities among them when it comes to Big data.

7. BIBLIOGRAPHY

10minutes. PWC. (2015, May). https://www.pwc.com/ee/et/publications/pub/pwc-making-strategic-decisions.pdf

Akter, S., & Wamba, S. F. (2016). Big Data Analytics in e-commerce: A systematic review and Agenda for Future Research. *Electronic Markets*, *26*(2), 173–194. https://doi.org/10.1007/s12525-016-0219-0

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using Big Data Analytics capability and business strategy alignment? *International Journal of Production Economics*, *182*, 113–131.

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, *6*(1). https://doi.org/10.1186/s13174-015-0041-5

Alonso, S. G., de la Torre Díez, I., Rodrigues, J. J., Hamrioui, S., & López-Coronado, M. (2017). A systematic review of techniques and sources of big data in the healthcare sector. *Journal of Medical Systems*, *41*(11).

Althati, C., Arasu Malaiyappan, J. N., & Shanmugam, L. (2024). Ai-Driven Analytics: Transforming data platforms for real-time decision making. *Journal of Artificial Intelligence General Science (JAIGS) ISSN:3006-4023*, *3*(1), 392–402. https://doi.org/10.60087/jaigs.v3i1.155

Armstrong, M., & Richter, F. (2023, November 2). *Airbnb bookings already past pre-pandemic high in 2023*. Statista Daily Data. https://www.statista.com/chart/23520/number-of-nights-and-experiences-booked-on-airbnb/

Arun, V. (2021). INFLUENCE OF BUSINESS INTELLIGENCE AND ANALYTICS ON BUSINESS VALUE . *International Engineering Journal For Research & Development*, *6*(1).

Azad, P., Navimipour, N. J., Rahmani, A. M., & Sharifi, A. (2019). The role of structured and unstructured data managing mechanisms in the internet of things. *Cluster Computing*, *23*(2), 1185–1198.

Azeroual, O., & Fabre, R. (2021). Processing big data with Apache Hadoop in the current challenging era of covid-19. *Big Data and Cognitive Computing*, *5*(1), 12.

Baars, H., & Kemper, H.-G. (2008). Management support with structured and unstructured data—an integrated business intelligence framework. *Information Systems Management*, *25*(2), 132–148. https://doi.org/10.1080/10580530801941058

Baldassarre, M. (2016). Think big: Learning contexts, algorithms and data science. *Research on Education and Media*, *8*(2), 69–83. https://doi.org/10.1515/rem-2016-0020

Bay Atlantic University. (2023, November 7). *The four types of big data analytics and their benefits*. Bay Atlantic University - Washington, D.C. https://bau.edu/blog/big-data-analytics-types/#:~:text=There%20are%20four%20main%20types,%2C%20prescriptive%2C%20and%20predictive%20analytics.

Bayrak, T. (2015). A review of Business Analytics: A business enabler or another passing fad. *Procedia - Social and Behavioral Sciences*, *195*, 230–239. https://doi.org/10.1016/j.sbspro.2015.06.354

Belhadi, A., Kamble, S. S., Zkik, K., Cherrafi, A., & Touriki, F. E. (2020). The integrated effect of Big Data Analytics, lean six sigma and green manufacturing on the environmental performance of manufacturing companies: The case of North Africa. *Journal of Cleaner Production*, *252*, 119903.

Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, *66*(3), 1025–1044.

Blazquez, D., & Domenech, J. (2018). Big data sources and methods for social and economic analyses. *Technological Forecasting and Social Change*, *130*, 99–113. https://doi.org/10.1016/j.techfore.2017.07.027

Boehm, B.W., Brown, J.R., Lipow, M.: *Quantitative Evaluation of Software Quality*. In: Proceedings of the 2nd International Conference on Software Engineering. pp. 592–605. IEEE Computer Society Press (1976)

Bonte, P., & Tommasini, R. (2023). Streaming linked data: A survey on life cycle compliance. *Journal of Web Semantics*, *77*, 100785. https://doi.org/10.1016/j.websem.2023.100785

Borkar, V., Carey, M. J., & Li, C. (2012). Inside “big data management.” *Proceedings of the 15th International Conference on Extending Database Technology*. https://doi.org/10.1145/2247596.2247598

Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017). The role of Big Data and predictive analytics in retailing. *Journal of Retailing*, *93*(1), 79–95.

Brem, A., Viardot, E., & Nylund, P. A. (2021). Implications of the coronavirus (COVID-19) outbreak for Innovation: Which Technologies will improve our lives? *Technological Forecasting and Social Change*, *163*, 120451.

Brynjolfsson, E., Jin, W., & McElheran, K. S. (2021a). The power of prediction: Predictive analytics, workplace complements, and Business Performance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3849716

Brynjolfsson, E., Jin, W., & McElheran, K. S. (2021b). The power of prediction: Predictive analytics, workplace complements, and Business Performance. *SSRN Electronic Journal*.

Carlier, M. (2024, June 14). *Most valuable car brands worldwide 2024*. Statista. https://www.statista.com/statistics/267830/brand-values-of-the-top-10-most-valuable-car-brands/

Casado, R., & Younas, M. (2014). Emerging trends and technologies in Big Data Processing. *Concurrency and Computation: Practice and Experience*, *27*(8), 2078–2091. https://doi.org/10.1002/cpe.3398

Cebotarean, E., & Maiorescu, T. (2011). Business Intelligence. Journal of Knowledge Management, Economics and Information Technology, 1(2).

Cecere, L. (2013). Big Data Handbook How to Unleash the Big Data Opportunity. *Supply Chain Insights LLC*.

Chaudhry, M., Shafi, I., Mahnoor, M., Vargas, D. L., Thompson, E. B., & Ashraf, I. (2023). A systematic literature review on identifying patterns using unsupervised clustering algorithms: A data mining perspective. *Symmetry*, *15*(9), 1679. https://doi.org/10.3390/sym15091679

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of Business Intelligence Technology. *Communications of the ACM*, *54*(8), 88–98. https://doi.org/10.1145/1978542.1978562

Chen, Chiang, & Storey. (2012a). Business Intelligence and Analytics: From Big Data to big impact. *MIS Quarterly*, *36*(4), 1165.

Chen, H., Chiang, R. H., & Storey, V. C. (2012b). Business Intelligence and Analytics: From Big Data to big impact. *MIS Quarterly*, *36*(4), 1165.

Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A survey. *Mobile Networks and Applications*, *19*(2), 171–209. https://doi.org/10.1007/s11036-013-0489-0

Cheruvu, R. (2015). *Big Data Applications in Self-Driving Cars*. https://doi.org/10.13140/RG.2.2.12302.05445

Cooper, P. (2014). Data, information, knowledge and wisdom. *Anaesthesia & Intensive Care Medicine*, *15*(1), 44–45. https://doi.org/10.1016/j.mpaic.2013.11.009

Corbellini, A., Mateos, C., Zunino, A., Godoy, D., & Schiaffino, S. (2017). Persisting big-data: The nosql landscape. *Information Systems*, *63*, 1–23. https://doi.org/10.1016/j.is.2016.07.009

Côrte-Real, N., Ruivo, P., & Oliveira, T. (2014). The diffusion stages of business intelligence & analytics (BI&A): A systematic mapping study. *Procedia Technology*, *16*, 172–179. https://doi.org/10.1016/j.protcy.2014.10.080

Cox, M., & Ellsworth, D. (1997). Application-controlled demand paging for out-of-core visualization. *Proceedings. Visualization '97 (Cat. No. 97CB36155)*, *ii*. https://doi.org/10.1109/visual.1997.663888

Cumbane, S. P., & Gidófalvi, G. (2019). Review of Big Data and processing frameworks for disaster response applications. *ISPRS International Journal of Geo-Information*, *8*(9), 387. https://doi.org/10.3390/ijgi8090387

Dautov, R., & Distefano, S. (2017). Quantifying volume, velocity, and variety to support (BIG) data-intensive application development. *2017 IEEE International Conference on Big Data (Big Data)*, *1142*, 2843–2852. https://doi.org/10.1109/bigdata.2017.8258252

Davenport, T. (2016, June 21). *A predictive analytics primer*. Harvard Business Review. https://hbr.org/2014/09/a-predictive-analytics-primer

Davenport, T. H. (2014). *Big data work: Dispelling the myths, uncovering the opportunities*. Harvard Business Pr.

Davenport, T., Barth , P., & Bean, R. (2012). How 'Big Data' is Different. *MIT Sloan Management Review*, *54*.

David Fayon. (2020, April 30). *Analyse de la transformation digitale des natu (netflix airbnb Tesla Uber)*. https://davidfayon.fr/2020/04/analyse-transformation-digitale-natu-netflix-airbnb-tesla-uber/comment-page-1/

De Mauro, A. (2019). *Big Data Analytics: Analizzare E interpretare dati con il machine learning*. Apogeo.

Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified Data Processing on Large Clusters. *Communications of the ACM*, *51*(1), 107–113. https://doi.org/10.1145/1327452.1327492

Dubey, R., Gunasekaran, A., Childe, S. J., Bryde, D. J., Giannakis, M., Foropon, C., Roubaud, D., & Hazen, B. T. (2020). Big Data Analytics and artificial intelligence pathway to operational performance under the effects of entrepreneurial orientation and environmental dynamism: A study of manufacturing organisations. *International Journal of Production Economics*, *226*, 107599.

Dwivedi, Y. K., Ismagilova, E., Hughes, D. L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A. S., Kumar, V., Rahman, M. M., Raman, R., Rauschnabel, P. A., Rowley, J., Salo, J., Tran, G. A., & Wang, Y. (2021). Setting the future of digital and social media marketing research: Perspectives and Research Propositions. *International Journal of Information Management*, *59*. https://doi.org/10.1016/j.ijinfomgt.2020.102168

Dylan Maltby. Big data analytics. In 74th Annual Meeting of the Association for Information Science and Technology (ASIST), pages 1–6, 2011

Elmasri, R., & Navathe, S. (2011). Fundamentals of Database Systems. Addison-Wesley.

Fan, S., Lau, R. Y. K., & Zhao, J. L. (2015). Demystifying big data analytics for business intelligence through the lens of Marketing Mix. *Big Data Research*, *2*(1), 28–32. https://doi.org/10.1016/j.bdr.2015.02.006

Farmer, D. (2024, May 30). *8 benefits of using big data for businesses*. TechTarget. https://www.techtarget.com/searchbusinessanalytics/feature/6-big-data-benefits-for-businesses

Feldman, R., & Sanger, J. (2007). *The text Mining Handbook: Advanced Approaches in analyzing unstructured data*. University Press.

Few, S. (2006). *Information dashboard design: Displaying data for at-a-glance Monitoring*. O'Reilly.

Franks, B. (2012). *Taming the big data tidal wave: Finding opportunities in huge data streams with advanced analytics*. John Wiley & Sons, Inc.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big Data Concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gelo, O., Braakmann, D., & Benetka, G. (2008). Quantitative and qualitative research: Beyond the debate. *Integrative Psychological and Behavioral Science*, *42*(3), 266–290. https://doi.org/10.1007/s12124-008-9078-3

Ghemawat, S., Gobioff, H., & Leung, S.-T. (2003). The google file system. *Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles*, 29–43. https://doi.org/10.1145/945445.945450

Glick, D. (2014). *Demystifying big data: A practical guide to transforming the business of government*. Breaking Government. https://breakinggov.com/documents/demystifying-big-data-a-practical-guide-to-transforming-the-bus/

Glick, D. (n.d.). *Demystifying big data: A practical guide to transforming the business of government*. Breaking Government.

Gonizzi, P., Ferrari, G., Gay, V., & Leguay, J. (2015). Data dissemination scheme for distributed storage for IOT observation systems at large scale. *Information Fusion*, *22*, 16–25.

Gorry, G. A., & Scott Morton, M. S. (1971). *A framework for Management Information Systems, by G. Anthony Gorry and Michael S.S. Morton*. M.I.T.

Grander, G., da Silva, L. F., & Santibañez Gonzalez, E. D. (2021). Big Data as a value generator in decision support systems: A literature review. *Revista de Gestão*, *28*(3), 205–222.

Grover, V., Chiang, R. H. L., 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.

Gu, F., Ma, B., Guo, J., Summers, P. A., & Hall, P. (2017). Internet of things and Big Data as potential solutions to the problems in Waste Electrical and Electronic Equipment Management: An exploratory study. *Waste Management*, *68*, 434–448.

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from Big Data. *The Journal of Strategic Information Systems*, *26*(3), 191–209. https://doi.org/10.1016/j.jsis.2017.07.003

Gupta, R., Sharma, A., & Alam, T. (2024). Building predictive models with machine learning. *Studies in Big Data*, 39–59. https://doi.org/10.1007/978-981-97-0448-4_3

Hecht, R., & Jablonski, S. (2011). NoSQL evaluation: A use case oriented survey. *2011 International Conference on Cloud and Service Computing*, 336–341. https://doi.org/10.1109/csc.2011.6138544

Heiskala, M., Jokinen, J.-P., & Tinnilä, M. (2016). Crowdsensing-based transportation services — an analysis from business model and sustainability viewpoints. *Research in Transportation Business & Management*, *18*, 38–48. https://doi.org/10.1016/j.rtbm.2016.03.006

Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B., & Sethupathy, G. (2016, December 7). *The age of analytics: Competing in a data-driven world*. McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-age-of-analytics-competing-in-a-data-driven-world

Hill, D. (2015). How much is your spare room worth? *IEEE Spectrum*, *52*(9), 32–58. https://doi.org/10.1109/mspec.2015.7226609

Hirsch, D. D. (2014). *The Glass House Effect Big Data, the new oil, and the power of analogy*. SSRN.

Hu, Q., Cai, M., Mohabbati-Kalejahi, N., Mehdizadeh, A., Alamdar Yazdi, M. A., Vinel, A., Rigdon, S. E., Davis, K. C., & Megahed, F. M. (2020). A review of data analytic applications in road traffic safety. part 2: Prescriptive modeling. *Sensors*, *20*(4), 1096.

Hyun Kim, J., Nan, D., Kim, Y., & Park Min, H. (2021). Computing the user experience via Big Data Analysis: A case of uber services. Computers, Materials & Continua, 67(3), 2819–2829. https://doi.org/10.32604/cmc.2021.014922

Inmon, W. H. (2005). *Building the data warehouse, Fourth Edition W.H. inmon*. Wiley.

Ishwarappa, & Anuradha, J. (2015). A brief introduction on Big Data 5vs characteristics and Hadoop Technology. *Procedia Computer Science*, *48*, 319–324.

Jain, A., Kulkarni, G., & Shah, V. (2018). Natural language processing. *International Journal of Computer Sciences and Engineering*, *6*(1), 161–167. https://doi.org/10.26438/ijcse/v6i1.161167

Jiang, H., Shen, F., Chen, S., Li, K.-C., & Jeong, Y.-S. (2015). A secure and scalable storage system for aggregate data in IOT. *Future Generation Computer Systems*, *49*, 133–141.

Jing Han, Haihong E, Guan Le, & Jian Du. (2011). Survey on NoSQL database. *2011 6th International Conference on Pervasive Computing and Applications*, 363–366. https://doi.org/10.1109/icpca.2011.6106531

Judith Hurwitz, Fern Halper and Marcia Kaufman, "Dashboards — Enabling Insight and Action", Hurwitz Research, 2005, Hurwitz & Associates, Massachusetts, USA. (Publication).

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big data: Issues and challenges moving forward. *2013 46th Hawaii International Conference on System Sciences*. https://doi.org/10.1109/hicss.2013.645

Khansa, L., Forcade, J., Nambari, G., Parasuraman, S., & Cox, P. (2012). Proposing an intelligent cloud-based Electronic Health Record System. *International Journal of Business Data Communications and Networking*, *8*(3), 57–71.

Khan, M. A., Uddin, M. F., & Gupta, N. (2014). Seven V's of big data understanding big data to extract value. *Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education*, 1–5. https://doi.org/10.1109/aseezone1.2014.6820689

Khatuwal, V. S., & Puri, D. (2022). Business intelligence tools for dashboard development. *2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)*, *4*, 128–131. https://doi.org/10.1109/iciem54221.2022.9853086

Kim, H.-W., & Jeong, Y.-S. (2016). Efficient auto-scaling scheme for rapid storage service using many-core of desktop storage virtualization based on IOT. *Neurocomputing*, *209*, 67–74.

Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The definitive guide to dimensional modeling*. Wiley.

Kumari, D., & Bhat, S. (2021). Application of artificial intelligence in tesla- A case study. *International Journal of Applied Engineering and Management Letters*, 205–218. https://doi.org/10.47992/ijaeml.2581.7000.0113

Kwiatkowska, M. (2023). Big Data Analytics - Tools and Technologies: Studying tools and technologies for big data analytics, including distributed computing frameworks like Hadoop and Spark. *AUSTRALIAN JOURNAL OF MACHINE LEARNING RESEARCH & APPLICATIONS* , *3*(1).

Kumari, S., Patil, Y., & Jeble, S. (2016). Role of big data and predictive analytics. *International Journal of Automation and Logistics*, *2*(4), 307. https://doi.org/10.1504/ijal.2016.10001272

Laney, D. (2001) *3D Data Management: Controlling Data Volume, Velocity and Variety*. META Group Research Note, 6.

Larose, D. T. (2015). *Data Mining and Predictive Analytics*. John Wiley & Sons, Incorporated.

LaValle, S., & Lesser, E. (2010, December 21). *Big Data, analytics and the path from insights to value*. MIT Sloan Management Review. https://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value/

Litoiu, M., Gaikwad, P., Shtern, M., Ramprasad, B., Beigi-Mohammadi, N., Zareian, S., Fokaefs, M., & Khazaei, H. (2016). How do I choose the right nosql solution? A comprehensive theoretical and experimental survey. *Big Data and Information Analytics*, *1*(2/3), 185–216. https://doi.org/10.3934/bdia.2016004

Liu, X., Dastjerdi, A. V., & Buyya, R. (2016). Stream processing in IOT: Foundations, state-of-the-art, and Future Directions. *Internet of Things*, 145–161. https://doi.org/10.1016/b978-0-12-805395-9.00008-3

Lytras, M. D., Raghavan, V., & Damiani, E. (2017). Big Data and Data Analytics Research. *International Journal on Semantic Web and Information Systems*, *13*(1), 1–10. https://doi.org/10.4018/ijswis.2017010101

MacMillan , D., & Demos, T. (2015, July 31). *Uber valued at more than $50 billion* . The Wall Street Journal. https://www.wsj.com/articles/uber-valued-at-more-than-50-billion-1438367457

Maddodi, S., & K, K. P. (2019). Netflix bigdata analytics- the emergence of data driven recommendation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3473148

Maimon, O. Z., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook*. Springer.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). *Big Data: The Next Frontier for Innovation, competition, and Productivity*. McKinsey Global Institute.

Marr, B., & Blaker, L. (2024). *Big Data in practice: How 45 successful companies used big data analytics to deliver extraordinary results*. Ascent Audio.

Marz, N., & Warren, J. (2018). *Big data principles and best practices of Scalable Realtime Data Systems*. Manning Publications : distributed by Skillsoft Books.

Mashingaidze, K., & Backhouse, J. (2017). The relationships between definitions of big data, business intelligence and Business Analytics: A literature review. *International Journal of Business Information Systems*, *26*(4), 488. https://doi.org/10.1504/ijbis.2017.087749

McAfee, A.P., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard business review*, 90 10, 60-6, 68, 128 .

Mills, T. (2022, April 14). *Council post: Five benefits of Big Data Analytics and how companies can get started*. Forbes. https://www.forbes.com/councils/forbestechcouncil/2019/11/06/five-benefits-of-big-data-analytics-and-how-companies-can-get-started/

Mir, S. A., & Quadri, S. M. (2009). Decision Support Systems: Concepts, progress and issues – A Review. *Climate Change, Intercropping, Pest Control and Beneficial Microorganisms*, 373–399.

Misra, N. N., Dixit, Y., Al-Mallahi, A., Bhullar, M. S., Upadhyay, R., & Martynenko, A. (2022). IOT, Big Data, and artificial intelligence in agriculture and Food Industry. *IEEE Internet of Things Journal*, *9*(9), 6305–6324. https://doi.org/10.1109/jiot.2020.2998584

Model 3 earns 5-star safety rating from Euro NCAP. Tesla. (n.d.). https://www.tesla.com/blog/model-3-earns-5-star-safety-rating-euro-ncap

Mohamed, A., Najafabadi, M. K., Wah, Y. B., Zaman, E. A., & Maskat, R. (2019). The state of the art and Taxonomy of Big Data Analytics: View from new Big Data Framework. *Artificial Intelligence Review*, *53*(2), 989–1037. https://doi.org/10.1007/s10462-019-09685-9

Mohammed, A. F., Humbe, V. T., & Chowhan, S. S. (2016). A review of Big Data Environment and its related technologies. *2016 International Conference on Information Communication and Embedded Systems (ICICES)*, 1–5. https://doi.org/10.1109/icices.2016.7518904

Mohankumar, P. (2013). BIG Data Analytics: A Framework for Unstructured Data Analysis. *International Journal of Engineering and Technology*.

Moniruzzaman, A. B. M., & Akhter Hossain, S. (2013). NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison. *International Journal of Database Theory and Application*, *6*(4). https://doi.org/ https://doi.org/10.48550/arXiv.1307.0191

Morgan, B. (2024, February 20). *3 ways tesla creates a personalized customer experience*. Forbes. https://www.forbes.com/sites/blakemorgan/2021/05/10/3-ways-tesla-creates-a-personalized-customer-experience/

Mucci, T., & Stryker, C. (2024, April 5). *What is Big Data Analytics?*. IBM. https://www.ibm.com/topics/big-data-analytics

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in Big Data Analytics. *Journal of Big Data*, *2*(1). https://doi.org/10.1186/s40537-014-0007-7

Nandimath, J., Banerjee, E., Patil, A., Kakade, P., Vaidya, S., & Chaturvedi, D. (2013). Big Data Analysis using Apache Hadoop. *2013 IEEE 14th International Conference on Information Reuse & Integration (IRI)*. https://doi.org/10.1109/iri.2013.6642536

Negash, S. (2004). Business intelligence. Communications of the Association for Information Systems, 13. https://doi.org/10.17705/1cais.01315

Netflix Technology Blog. (2024, March 26). *Navigating the Netflix Data Deluge: The imperative of effective data management*. Medium.

https://netflixtechblog.medium.com/navigating-the-netflix-data-deluge-the-imperative-of-effective-data-management-e39af70f81f7

O'Toole, C., Schneider, J., Smaje, K., & LaBerge, L. (2020, October 5). *How COVID-19 has pushed companies over the technology tipping point-and transformed business forever*. McKinsey & Company. https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/how-covid-19-has-pushed-companies-over-the-technology-tipping-point-and-transformed-business-forever

Osztovits, Á., & Kőszegi, Á. (2015). *Sharing or paring? growth of the sharing economy*. PWC. https://www.pwc.com/hu/en/kiadvanyok/assets/pdf/sharing-economy-en.pdf

OUNACER, S., TALHAOUI, M. A., & ARDCHIR, S. (2017). Real-time data stream processing - challenges and perspectives. *International Journal of Computer Science Issues*, *14*(5), 6–12. https://doi.org/10.20943/01201705.612

Patil, P. C., & Bhosale, A. (2018). Big Data Analytics. *Open Access Journal of Science*, *2*(5). https://doi.org/10.15406/oajs.2018.02.00095

Phillips, J. (2012, February 8). *Couchbase Survey shows accelerated adoption of NoSQL in 2012*. Couchbase. https://www.couchbase.com/press-releases/couchbase-survey-shows-accelerated-adoption-of-nosql-in-2012/

Phillips-Wren, G., Daly, M., & Burstein, F. (2021). Reconciling Business Intelligence, analytics and decision support systems: More Data, deeper insight. *Decision Support Systems*, *146*, 113560.

Plattner, H., & Zeier, A. (2012). *In-memory data management: An inflection point for Enterpise applications*. Springer.

Pokorny, J. (2011). NoSQL databases. *Proceedings of the 13th International Conference on Information Integration and Web-Based Applications and Services*, 278–283. https://doi.org/10.1145/2095536.2095583

Press, G. (2014). 12 Big Data Definitions: What's Yours?," . *Forbes*.

Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. O'Reilly.

Punn, N. S., Agarwal, S., Syafrullah, M., & Adiyarta, K. (2019). Testing big data application. *2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 159–162. https://doi.org/10.23919/eecsi48112.2019.8976972

Puri, S. (2012). A technical study and analysis on fuzzy similarity based models for text classification. *International Journal of Data Mining & Knowledge Management Process*, *2*(2), 1–15.

Ranjan, J. (2005). Business Intelligence: Concepts, Components, Techniques And Benefits. *Journal of Theoretical and Applied Information Technology*.

Reinsel, D., Gantz, J. & Rydning, J. (2018) Data age 2025: the evolution of data to life-critical. An IDC whitepaper. Seagate https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf (2018).

Roelofsen, M., & Minca, C. (2018). The superhost. Biopolitics, home and community in the airbnb dream-world of Global Hospitality. *Geoforum*, *91*, 170–181. https://doi.org/10.1016/j.geoforum.2018.02.021

Rosamond, H. (2016, June 28). *Uber CEO Travis Kalanick: soon, nobody will own a car*. World Economic Forum. https://www.weforum.org/stories/2016/06/uber-travis-kalanick-driverless-cars/

Russom, P. (2011). Best practices for big data analytics. *TDWI Best Practices Report*.

Sagiroglu, S., & Sinanc, D. (2013). Big Data: A Review. *2013 International Conference on Collaboration Technologies and Systems (CTS)*. https://doi.org/10.1109/cts.2013.6567202

Sarker, S., Arefin, M. S., Kowsher, M., Bhuiyan, T., Dhar, P. K., & Kwon, O.-J. (2023). A comprehensive review on big data for industries: Challenges and opportunities. *IEEE Access*, *11*, 744–769. https://doi.org/10.1109/access.2022.3232526

Sathiaraj, D., Punkasem, T., Wang, F., & Seedah, D. P. K. (2018). Data-driven analysis on the effects of extreme weather elements on traffic volume in Atlanta, GA, USA. *Computers, Environment and Urban Systems*, *72*, 212–220. https://doi.org/10.1016/j.compenvurbsys.2018.06.012

Seetharaman, P. (2020). Business models shifts: Impact of covid-19. *International Journal of Information Management*, *54*, 102173. https://doi.org/10.1016/j.ijinfomgt.2020.102173

Semanjski, I., Bellens, R., Gautama, S., & Witlox, F. (2016). Integrating big data into a sustainable mobility policy 2.0 planning support system. *Sustainability*, *8*(11), 1142.

Seyedan, M., & Mafakheri, F. (2020). Predictive big data analytics for supply chain demand forecasting: Methods, applications, and research opportunities. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00329-2

Sharda, R., Delen, D., Turban, E., Aronson, J. E., Liang, T.-P., & King, D. (2022). *Business Intelligence and Analytics: Systems for Decision Support*. Pearson India Education Services Pvt. Ltd.

Shi, L., Zhang, J., Liu, E., & He, P. (2007). Text classification based on nonlinear dimensionality reduction techniques and support vector machines. *Third International Conference on Natural Computation (ICNC 2007)*, *34*, 674–677. https://doi.org/10.1109/icnc.2007.706

Sirin, E. (2017). A review on business intelligence and Big Data. *International Journal of Intelligent Systems and Applications in Engineering*, *4*(5), 206–215. https://doi.org/10.18201/ijisae.2017533861

Song, H., Li, M., & Yu, K. (2021). Big Data Analytics in digital platforms: How do financial service providers customise supply chain finance? *International Journal of Operations & Production Management*, *41*(4), 410–435.

Sood, S. K., Sandhu, R., Singla, K., & Chang, V. (2018). IOT, Big Data and HPC based Smart Flood Management Framework. *Sustainable Computing: Informatics and Systems*, *20*, 102–117. https://doi.org/10.1016/j.suscom.2017.12.001

Souza, G. C. (2014). Supply Chain Analytics. *Business Horizons*, *57*(5), 595–605.

Sravanthi, K., & Subba Reddy, T. (2015). Applications of Big data in Various Fields . *International Journal of Computer Science and Information Technologies*, *6*(5).

Stankovska, B. (2024, July 31). *10 examples of digital transformation - snapstack - transforming businesses*. Snapstack. https://snapstack.cz/examples-of-digital-transformation/

Statista. (2024, May 22). Airbnb revenue worldwide 2023. https://www.statista.com/statistics/1193134/airbnb-revenue-worldwide/

Statista. (2024a, April 8). *Uber's users of ride-sharing services worldwide 2017-2023*. https://www.statista.com/statistics/833743/us-users-ride-sharing-services/

Stobierski, T. (2019, August 26). *The advantages of data-driven decision-making: HBS Online*. Business Insights Blog. https://online.hbs.edu/blog/post/data-driven-decision-making

Stoll, J. (2024, October 18). *Netflix: Quarterly revenue 2024*. Statista. https://www.statista.com/statistics/273883/netflixs-quarterly-revenue/

Tao, M., Ota, K., & Dong, M. (2017). Ontology-based data semantic management and application in IOT- and cloud-enabled Smart Homes. *Future Generation Computer Systems*, *76*, 528–539.

Tarnaud, N. (2020, February 11). *Les Natu Vont-Elles rejoindre Les Gafam ?*. Forbes France. https://www.forbes.fr/technologie/les-natu-vont-elles-rejoindre-les-gafam/

Taylor, P. (2023, November 16). *Data Growth Worldwide 2010-2025*. Statista. https://www.statista.com/statistics/871513/worldwide-data-created/

Torre-Bastida, A. I., Del Ser, J., Laña, I., Ilardia, M., Bilbao, M. N., & Campos-Cordobés, S. (2018). Big Data for transportation and mobility: Recent advances, trends and challenges. *IET Intelligent Transport Systems*, *12*(8), 742–755. https://doi.org/10.1049/iet-its.2018.5188

Uber's website. (2024). *How surge pricing works*. Uber. https://www.uber.com/us/en/drive/driver-app/how-surge-works/

Uber's Website. (2024). *Uber's big data platform: 100+ petabytes with minute latency - uber engineering blog*. Uber. https://eng.uber.com/uber-big-data-platform/

Uber's Website. (2024a). *Taking uberpool trips | A driver's guide to uberpool*. Uber. https://www.uber.com/au/en/drive/services/shared-rides/

Urbanczyk, T., & Peter, L. (2016). Database development for the Urgent Department of hospital based on Tagged entity storage following the IOT Concept. *IFAC-PapersOnLine*, *49*(25), 278–283.

Vasupradha. (2024, February 11). *Netflix-through the years with Business Analytics*. Medium. https://medium.com/@btsvasupradha/netflix-through-the-years-with-business-analytics-92b34d57c0de

Wang, S. (2023, October 17). *Big Data in Tesla Inc.*. Digital Innovation and Transformation. https://d3.harvard.edu/platform-digit/submission/big-data-in-tesla-inc/

Wang, W., Xu, P., & Yang, L. T. (2018). Secure data collection, storage and access in cloud-assisted IOT. *IEEE Cloud Computing*, *5*(4), 77–88.

Wang, Y., Kung, L., & Byrd, T. A. (2018). Big Data Analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, *126*, 3–13. https://doi.org/10.1016/j.techfore.2015.12.019

Wang, Y., Kung, L., Gupta, S., & Ozdemir, S. (2019). Leveraging Big Data Analytics to improve quality of care in healthcare organizations: A configurational perspective. *British Journal of Management*, *30*(2), 362–388. https://doi.org/10.1111/1467-8551.12332

Watson, H. J. (2009). Tutorial: Business intelligence – past, present, and future. Communications of the Association for Information Systems, 25. https://doi.org/10.17705/1cais.02539

Watson, H. J. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, *34*.

White, T. (2009). *Hadoop the definitive guide.* . O'Reilly Media, Inc.

Williams, S. (2016). Business intelligence in the era of Big Data and cognitive business. *Business Intelligence Strategy and Big Data Analytics*, 27–50. https://doi.org/10.1016/b978-0-12-809198-2.00002-6

Yu, S., Qing, Q., Zhang, C., Shehzad, A., Oatley, G., & Xia, F. (2021). Data-driven decision-making in COVID-19 response: A survey. *IEEE Transactions on Computational Social Systems*, *8*(4), 1016–1029. https://doi.org/10.1109/tcss.2021.3075955

Zandt, F., & Richter, F. (2024, January 25). *Infographic: How successful is Tesla?*. Statista Daily Data. https://www.statista.com/chart/26705/yearly-net-income-and-revenue-of-tesla/

Zhan, Y., Tan, K. H., Ji, G., Chung, L., & Tseng, M. (2017). A big data framework for facilitating product innovation processes. *Business Process Management Journal*, *23*(3), 518–536. https://doi.org/10.1108/bpmj-11-2015-0157

Zhan, Y., Tan, K. H., Li, Y., & Tse, Y. K. (2016). Unlocking the power of Big Data in new product development. *Annals of Operations Research*, *270*(1–2), 577–595. https://doi.org/10.1007/s10479-016-2379-x

Zhang, C., Kolte, Prajakta, & Kettinger. (2018). Established Companies' Strategic Responses to Sharing Economy Threats. *MIS Quarterly Executive*, *17*(1).

Zhang, D., Pee, L. G., Pan, S. L., & Cui, L. (2022). Big Data Analytics, resource orchestration, and Digital Sustainability: A case study of smart city development. *Government Information Quarterly*, *39*(1), 101626. https://doi.org/10.1016/j.giq.2021.101626

Zhao, X., Lucani, D. E., Shen, X., & Wang, H. (2018). Reliable IOT storage: Minimizing bandwidth use in storage without newcomer nodes. *IEEE Communications Letters*, *22*(7), 1462–1465.

Zikopoulos, P. (2012). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill.