

THE INTELLIGENT USE OF DATA



IAPMEI



Powered by CAPACITAR PARA EMPREENDEDOR

TABLE OF CONTENTS

CHAPTER 1 | INTRODUCTION NOTE A quick introduction is made through a brief history of what intelligent use of data is in the field of Data Science and all of its sub-fields, as well as the benefits of appropriately implementing intelligent use of data. The value of examining data is also discussed, as are the issues of information vs knowledge.

4

CHAPTER 2 | Data SCIENCE Data science is explained in terms of topics and areas, as well as interdisciplinarity in economic sectors. The relevance of data analysis and visualization, as well as machine learning techniques, is mentioned. Also presented is the importance of Data Science in decision making.

10

CHAPTER 3 | DATA AS IMPORTANT INFORMATION FOR ORGANIZATIONS The science of data, its analysis, and visualization is handled as a key factor to consider for the business of organizations, highlighting of organizations, and highlighting of enterprises' business potential. Market examples are also provided, as well as an explanation of what a data scientist is.

22

CHAPTER 4 | TECHNOLOGICAL PLATFORMS The most often used technology platforms and technologies in Data Science are presented, as are the major supporting platforms.

33

CHAPTER 5 | OPPORTUNITIES, SOLUTIONS AND TRENDS The employment prospects for a data scientist are described, as well as the solutions available in the various areas/sectors. Data science trends are also discussed.

39

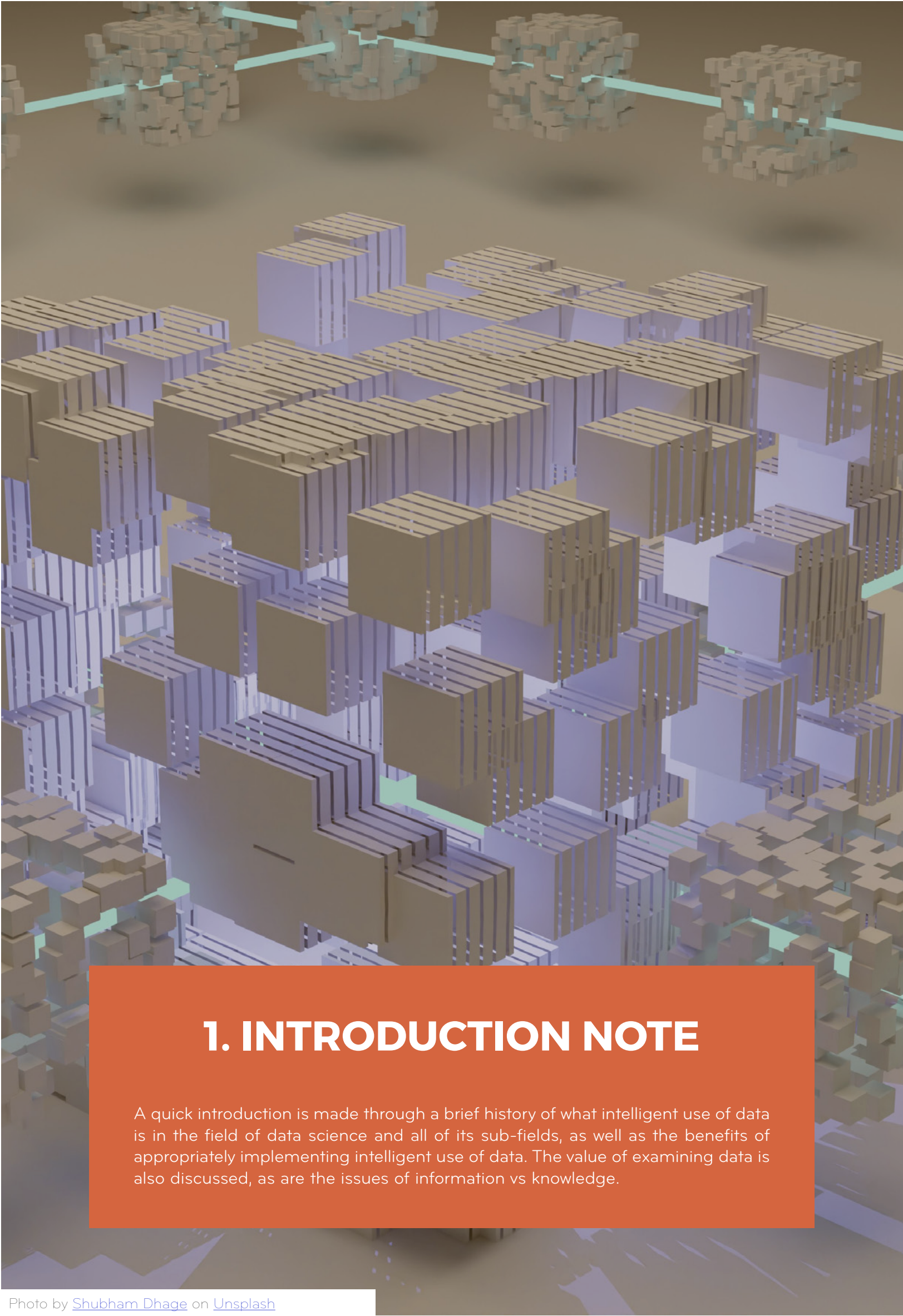
CHAPTER 6 | THE IMPACT OF THE INTELLIGENT DATA USE The importance of intelligent data exploitation, prediction as a result of effective data exploitation, and Data Science, as well as ensuring privacy and anonymity, are discussed.

44

CHAPTER 7 | CASE STUDIES The case studies refer to certain Data Science companies that are emerging in Portugal as well as international Data Science companies that will arise in 2022. Several funding opportunities for Data Science businesses are also mentioned.

50

TABLE OF CONTENTS	
GLOSSARY	61
REFERENCES	64
INDEX OF PICTURES, FIGURES AND TABLES	67



1. INTRODUCTION NOTE

A quick introduction is made through a brief history of what intelligent use of data is in the field of data science and all of its sub-fields, as well as the benefits of appropriately implementing intelligent use of data. The value of examining data is also discussed, as are the issues of information vs knowledge.

Photo by [Shubham Dhage](#) on [Unsplash](#)

1.1. Context

In recent years, the terms "Data Science" and "Big Data" have been discussed within academic and academic and business circles as themes that may help organizations in their strategic response to the rise of these new approaches. (Redman and Davenport (2021).

The truth is that these topics, when thoroughly researched and applied in current (or new) enterprise solutions, can make a significant difference in terms of strategic and technological positioning, as well as the identification of new opportunities, decision support, and the exploration of new market niches, among many other applications. (Avantika Monnappa, 2022)

This chapter is aimed to provide an overview of what intelligent use of data is in the context of Data Science and all of its subfields, as well as the benefits of using intelligent use of data correctly.

1.2. Information vs. Knowledge

There are Eras, Periods, and Epochs throughout history. Companies and organizations believe, for strategic purposes, that we are in the Information Age, despite the fact that no name has 'yet' been given to this period. Why? Since the rise of networks and the Internet in the 1990s, as well as all the Information Technology (IT) associated with this 'new' technology, there has been an abundance of information to be created, digitized, and changed. (Tucci, 2016)

However, there has always been a great deal of difficulty in understanding the difference between these two generally used terms: information and knowledge. What is the difference between information and knowledge? Is having information and having knowledge the same? No! Nevertheless, these two concepts are intricately interwoven, and it may be claimed, "There is no knowledge without information!"

It is a reality that both knowledge and information are founded on data. Data is the source of all information and knowledge used by businesses for any and all purposes. What does "Information" and "Knowledge" actually mean?

The primary distinction lies in the definitions of the steps in a process commonly referred to as "Knowledge Extraction," where "Knowledge" is the last step and "Information" is what is extracted after data processing. Figure 1.1 depicts a possible illustration of the "Knowledge Extraction" procedure. Extraction" method.



Figure 1.1: Phases of the Knowledge Extraction Process.
Source: Adapted from Guru99.com

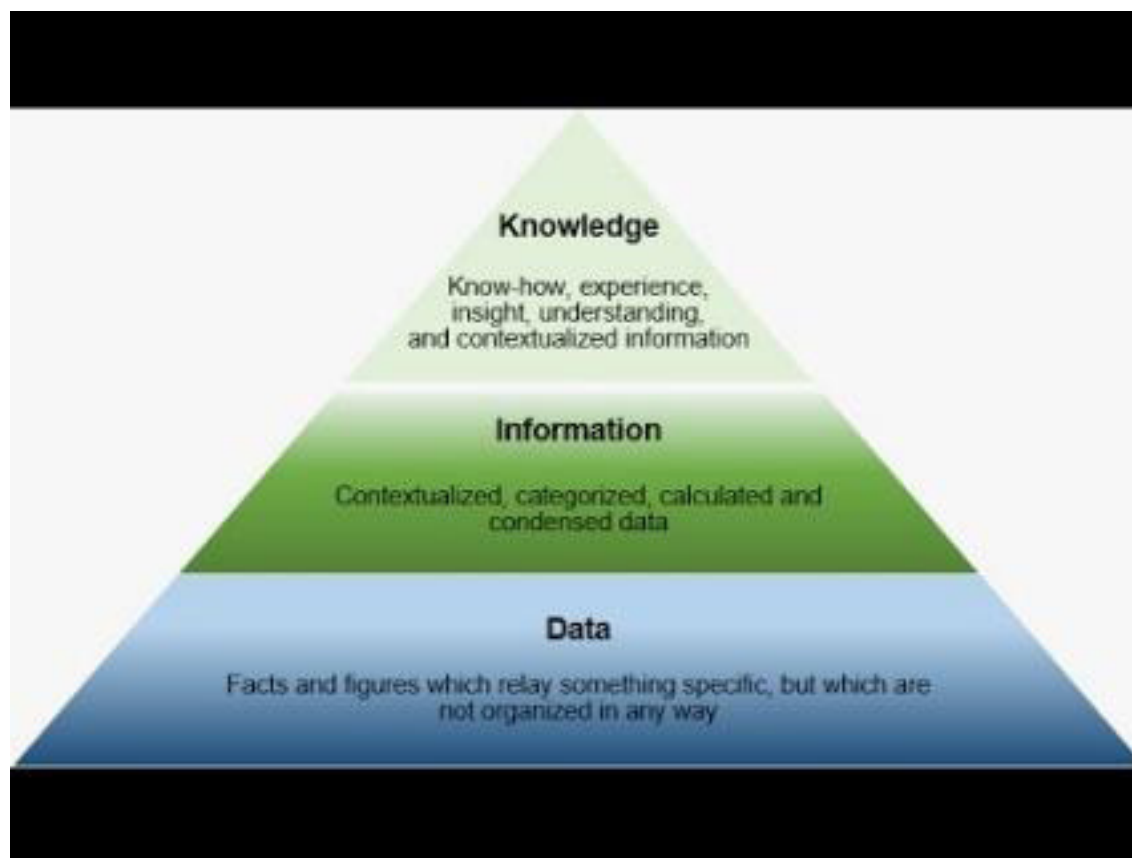
As the name implies, the goal of this technique is to extract knowledge. Starting with data, there is a requirement to process it in order to extract some type of information. These data can be gathered in a variety of ways, assuming that they are generally multimedia data, such as text, tables, graphs, photos, and videos, which can then be received / captured via documents, databases, cameras, sensors, and a variety of other methods. This is a step in the 'Information Extraction' process.

However, this knowledge requires a structure, a meaningful context, and a purpose. And it is through these aspects that knowledge is extracted. We can define knowledge as the mix of information processing, experience, and introspection about data, and it is directly tied to the know-how of individuals or the organization itself.

Knowledge is present in the organization if it is present in its personnel, knowing that it comes from experience and all of the information that has been processed.

When it comes to sharing and transfer, information and knowledge differ: information is easily shareable and transferable, however knowledge is not. To have Knowledge, it is important to learn, and this is the great bet of today's innovative solutions: knowing that Knowledge is required to be enhanced, reinvented, optimized, and changed. (Taylor, 2021)

We recommend watching the following video to complete this section interactively:



1.3. The importance of studying the data

In the business world, it is becoming increasingly important to understand what is an understanding and understanding about the data that is acquired for various purposes. And, as a result of the great need to collect information, a large amount of data is obtained every day in every day in companies as a result of their daily activity. But what is the 'additional' value of this information?

Previously, many companies did not explore data critically and strategically.They only used the data to store information about their customers, such as subscriptions to a particular service. However, due to the high level of competition in the business world, there is a need to analyze data more carefully and try to understand where they can improve, such as meeting consumer needs in a more optimized manner and establishing long-term, quality relationships (Wang, 2021).

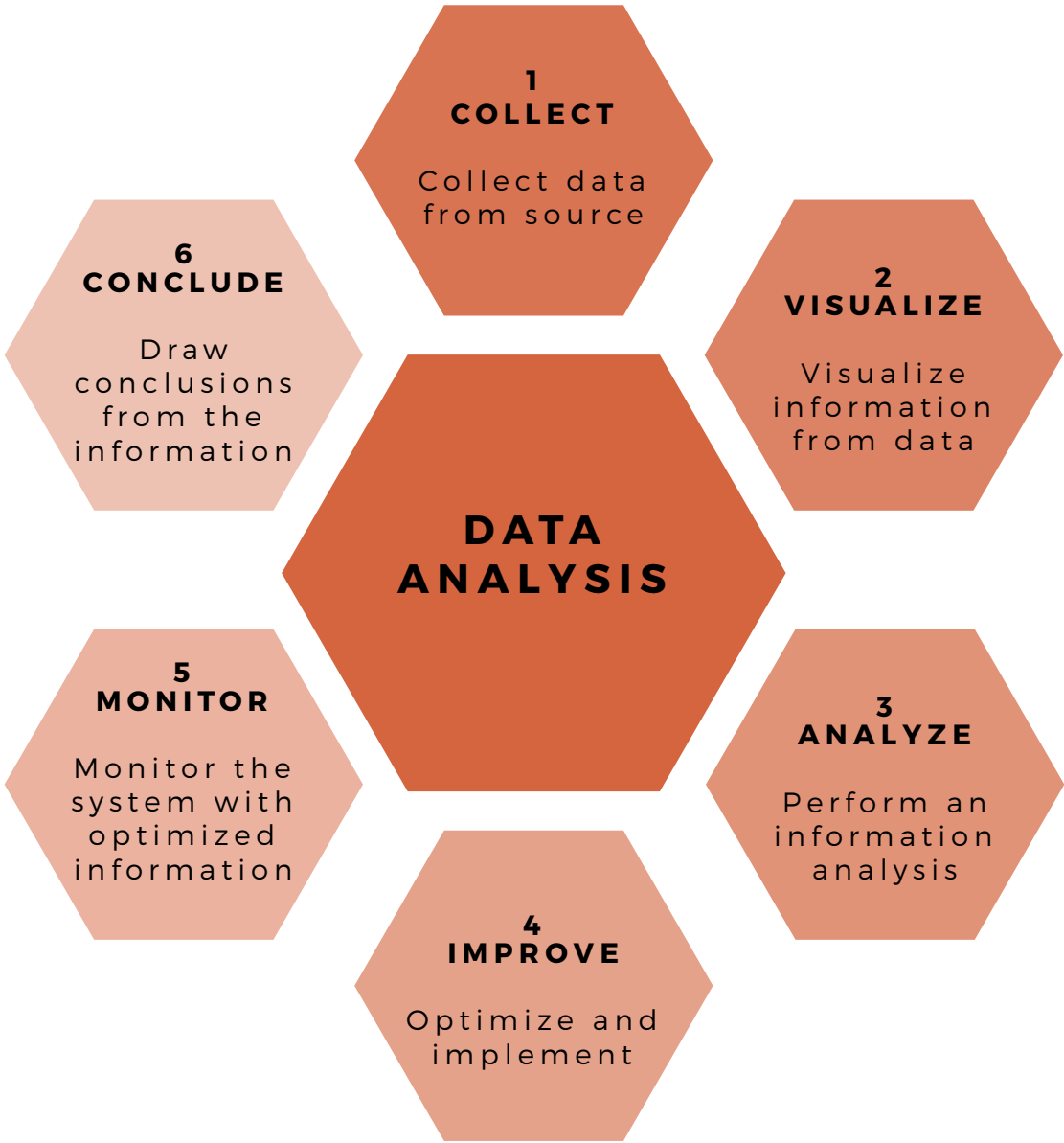


Figure 1.2: The basic concepts of data analysis
SOURCE: Adapted from HealthMinds

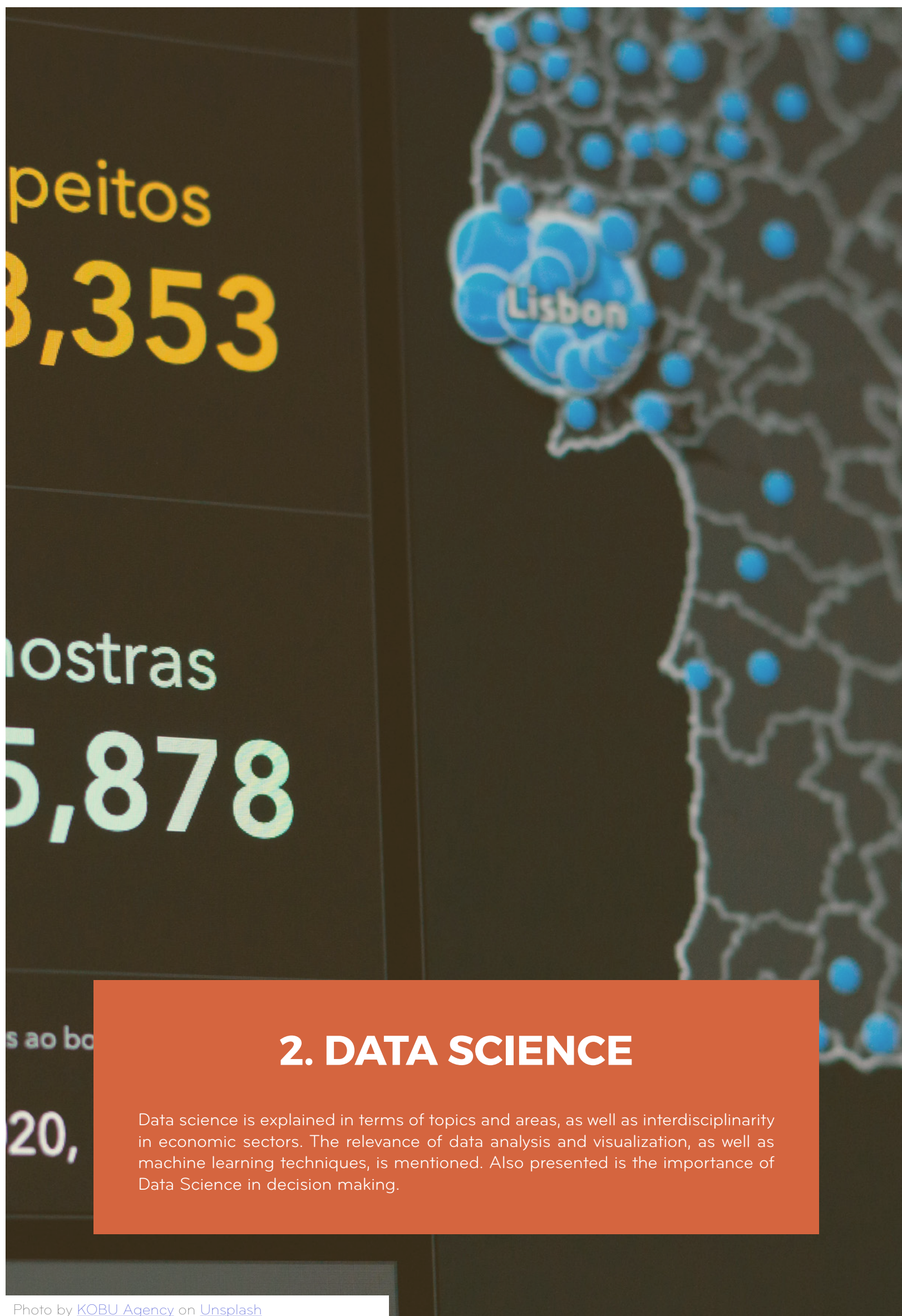
Companies have begun to strategically invest in new methodologies to study their clients or their products and/or services in response to this need. How? Using Data Science and all of its related methodologies and technologies.

The importance of data study, or data analysis, is based on several pillars of exploration, such as: the development of new products and/or services; market exploration, identifying what the customers' needs are and optimizing solutions for them (targeted marketing); optimization of internal and external processes, aimed at improving current techniques as well as updating them. techniques while also keeping them up to date.

And all of this is based on the common goals of all businesses: more and better, at a competitive cost, with quality, assertiveness, and on time. Ohri (2020)

We recommend watching the video below about data analysis and its importance in data analysis.





2. DATA SCIENCE

Data science is explained in terms of topics and areas, as well as interdisciplinarity in economic sectors. The relevance of data analysis and visualization, as well as machine learning techniques, is mentioned. Also presented is the importance of Data Science in decision making.

Photo by [KOBU Agency](#) on [Unsplash](#)

2.1. What is Data Science

Many of us have dealt with Data Science in one way or another. Today, just doing a search on the Internet or looking at dining deals on a mobile app for ordering food from a restaurant with a smartphone is an interaction with a product that uses Data Science.

Today, you could say that Data Science is involved in a lot of the things we do every day, and even if you don't know it, you're always hearing about ways to make those things better. (Igual & Seguí, 2020)

Data science is often considered of as the combination of several scientific fields, such as mathematics, statistics, computer science, engineering, programming, and Artificial Intelligence. When this combination is used with data, it gives rise to Data Science (Skiena, 2017)

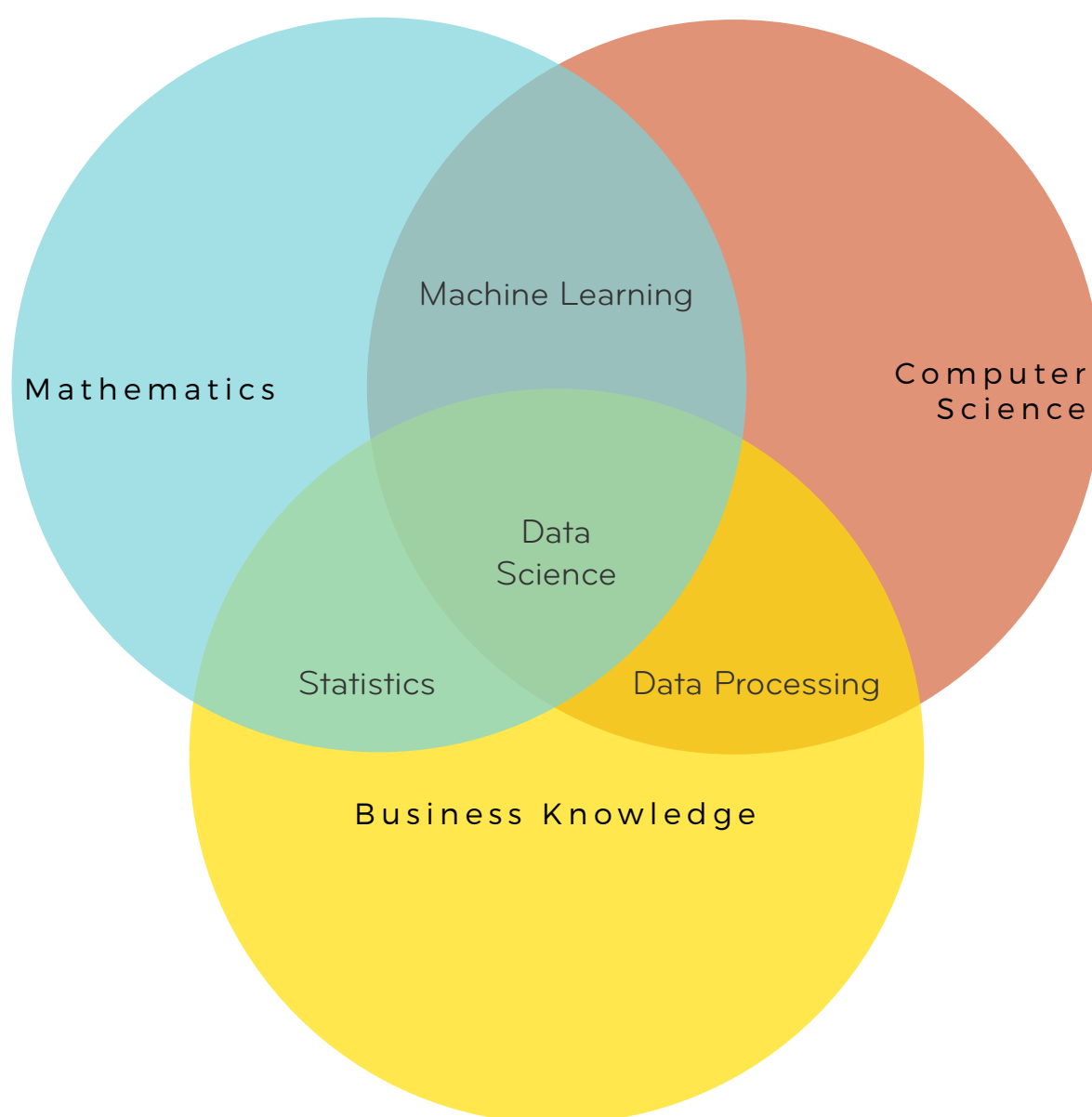


Figure 2.1: The domains and practice areas of Data Science.
Source: Adapted from: Analance

Some of these fields, such as mathematics and statistics, are quite old. Of course, computer science and machine learning are the most recent. However, it is also because of these areas that there has been a significant technological evolution in other areas. However, if these areas have been around for so long, why is there such a buzz about them now?

All of this hype is due to technological advances brought about by the dramatic increase in computing power, data power, and data available (Big Data). Previously, only large companies with economic power could perform what was known as data exploration with the techniques available on that period, which used traditional methods of information extraction.

Today, because of the massification of information and the emergence of new data exploration technologies, it is much more accessible and possible to extract knowledge from data. In a nutshell, Data Science is the study of data through the use of methodologies for processing, analyzing, visualizing, and comprehending data in order to draw conclusions that aid decision making. in making decisions Das (2016) and Schutt and O'Neil (2013)

We recommend watching the video below and learning about the demystification of Data Science.



2.3. The Interdisciplinary Nature of Data Science

Data Science has no context by itself. It is not specific. It does not have a specific field of application. It is completely inaccurate to believe that Data Science-based methodologies can only be applied in fields such as medicine or computer science.

Data Science is a versatile, adaptable field that can be applied to more common areas such as computer science and medicine, as well as the most diverse areas such as social sciences, engineering, economics, health sciences, and natural sciences, among others. (Anand, 2021)

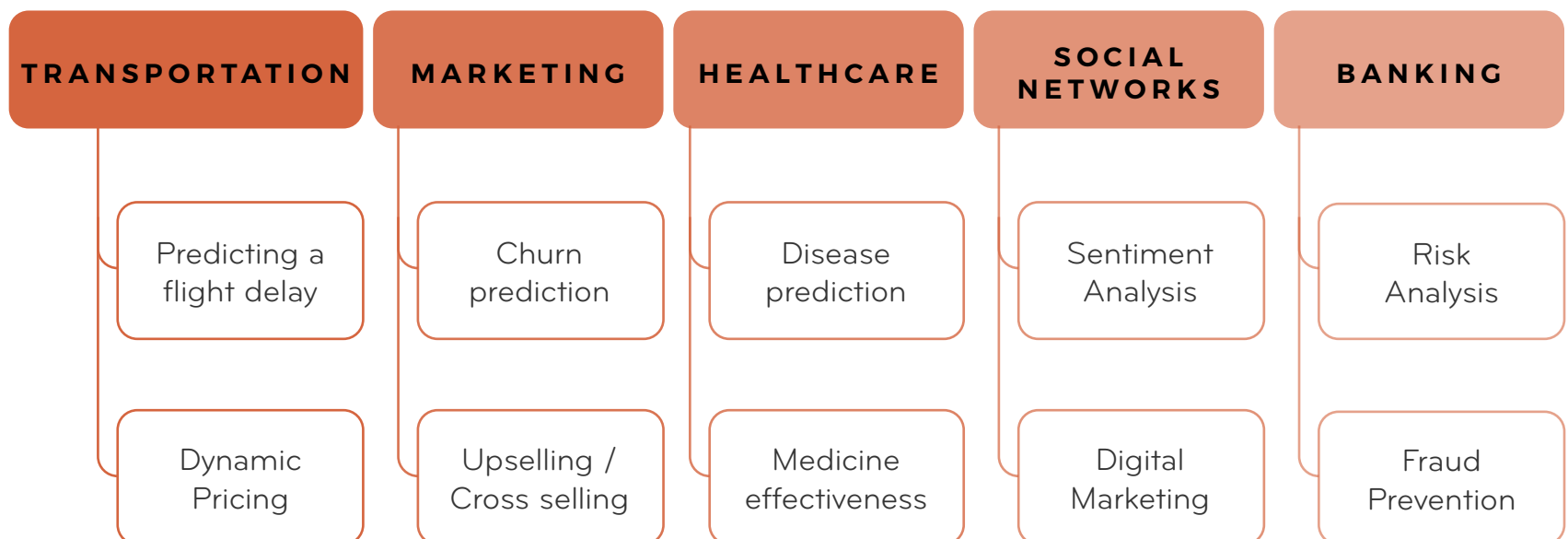


Figure 2.2: Examples of economic sectors where Data Science methodologies are used.
Source: Adapted from LucidInsights

When the goal is to extract knowledge from data in order to optimize processes, innovate solutions, and improve customer relationships, we can use Data Science. There is no difference between the limitations of Data Science and its actual application. We can use these methodologies if there is data.

Companies such as **Facebook**, for example, use Data Science to target advertising based on our interest profile, capturing what our preferences are, constantly analyzing our behavior within the platform, as well as the type of content we view or post.

Amazon, one of the largest online retailers, uses Data Science-supported methodologies to improve the user experience. They implement their recommendation system through these methodologies in order to perform predictive analyses on their products, making recommendations based on the customer's consumer profile, and improving customer satisfaction.

Uber, the company behind the same-named mobile application, uses these methodologies to optimize vehicle travel routes, taking into account current traffic conditions, as well as weather conditions, among other types of disturbances. It is also based on Data Science and provides the best price to its customers based on the current demand for vehicles and supply of drivers.

Airbnb, considered as one of the "largest accommodation companies, but which does not own any accommodation in its possession," employs Data Science in its strategy for presenting search results for an accommodation. Based on the preferences of the customer, it presents the best results for that customer first, in order to increase sales.

Airbnb discovered in 2014 that many users left the page after accessing accommodation results "in the vicinity of a specific destination." Also in 2014, **Airbnb** launched a new version of the platform that, instead of showing accommodations near the destination, shows the lodgings with the best reputation (top travel destinations), increasing total sales by 10%. (2021, Data Flair Tutorials)

To see more examples of the interdisciplinary application of Data Science, we recommend watching the video below, which shows the Top 20 applications of Data Science in the business world.



2.3. Analysis & visualization as a key driver in making conclusions

When dealing with large amounts of data, there is sometimes no simple way to determine whether the information is ready to be processed and whether it has been properly processed. To avoid misinterpretations and/or conclusions about data, any data scientist must check and validate all data before and after processing to ensure that the information it creates and/or transmits represents exactly what the data is.

Knowing that there is a great responsibility to ensure the reliability and veracity of our data's representation, what mechanisms exist to ensure that the processed information is as expected? What exactly does our data look like, and what conclusions can we draw from it? Skiena (2017)

It is only through visualization tools, in their various forms, that we can literally "look" at the shape of the data in order to perform Exploratory Data Analysis, which is the most common and widely used analysis to visualize information about our data.

In this analysis, we identify the variables of interest and conduct statistical analyses on them (averages, correlations) (averages, correlations, variances, relationships between variables, extraction of characteristics, etc.). The previously mentioned studies are also presented in the form of plots in this analysis. studies mentioned previously.

The detection of errors is one of the most significant benefits of using data visualization mechanisms. When we are already "watching" our data through visualization, we can find errors in the data that may have originated in the data source itself, or may be the result of its manipulation / transformation.

It is critical to ensure that our data, when processed, does not produce misleading information or errors that may cause problems in subsequent processes (for example, processing via Machine Learning algorithms).

When we want to convey our data, the information and knowledge extracted from it, visualization becomes important. When using visualization mechanisms, the process of 'communicating' the data, which is very common in the preparation of results, conclusions, and later reports, makes sharing information with others much more intuitive and understandable. (Grus, 2019)



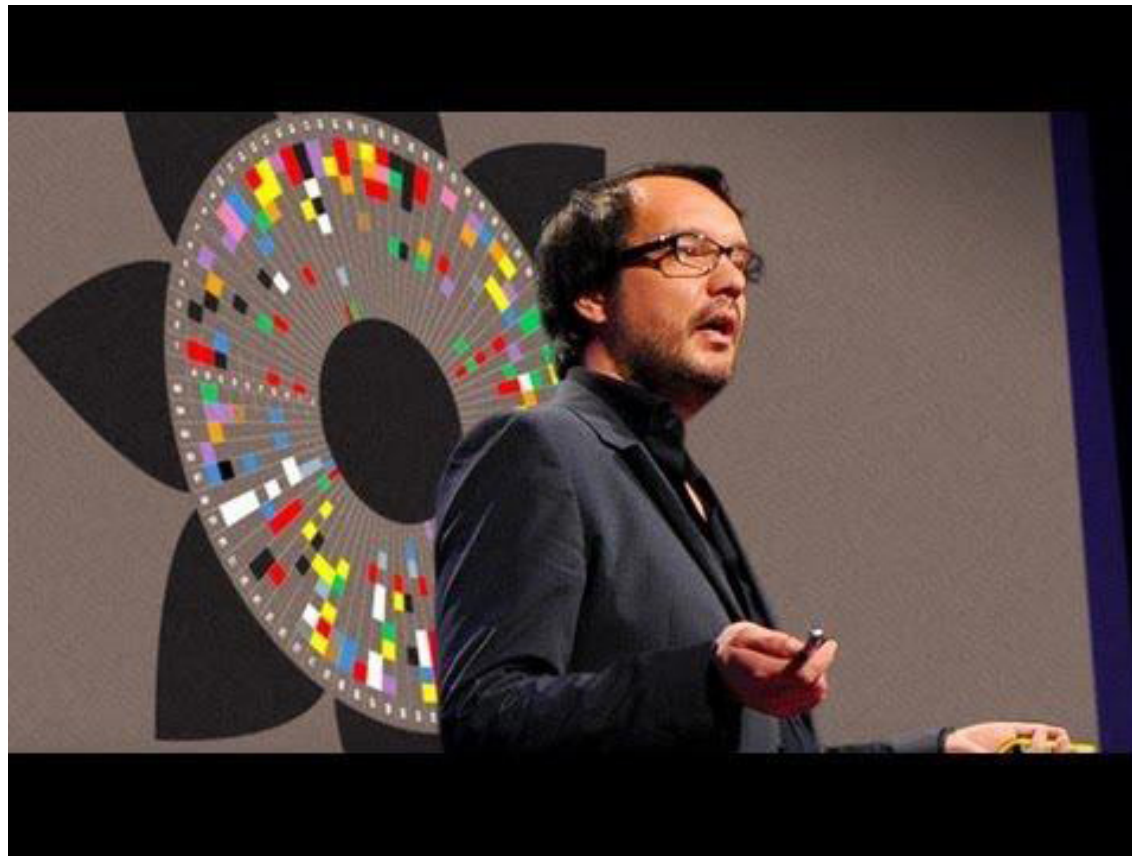
Figure 2.3: Land invasion points in Ukraine. Infographics: Jaime Figueiredo
Source: *Expresso.pt*, Date: 25/02/2022

We take Figure 5 as an example. Russia invaded Ukraine by air and ground this year. Figure 5 depicts the ground advance of Russian troops over eastern Ukraine. And, using a pink color scheme, the representation of regions under Russian control, with the circles (also in pink) representing Russian offensive attacks.

These conclusions were reached through the creation of a graphic of the geographic region type, which allowed for quick conclusions about the current state of the Russian invasion through the use of colors and legends. This clearly demonstrates the power of visualization in demonstrating outcomes and drawing quick conclusions.

There are numerous other mechanisms that show the power of visualization. Using the well-known adage that "a picture is worth a thousand words," these mechanisms are among the most important tools for visualizing information and extracting knowledge quickly and effectively in order to draw conclusions about data. knowledge that allows for effective data conclusion drawing 2020 (Analytiks)

Finally, we recommend watching the video below to learn about the "beauty" of data visualization.



2.4. Machine Learning as an Important Data Science Tool

Nowadays, business owners are more aware of market trends . One of them is the use of Machine Learning-enabled algorithms in their products, services, and internal processes, among other things. This is because ML algorithms are becoming more optimized, evolved, and accessible to all, making them valuable tools.

However, their application is heavily dependent on the use case and expected results, and when not used correctly, what could be a solution becomes a problem. Because this is an area that necessitates some specific technical knowledge, the algorithms supported by ML must be correctly applied in order to produce the expected results. And correct is dependent on a number of factors, one of which is the correct processing of the input data.(input).

By incorporating the domains mentioned in Figure 3 of this document, Data Science has the ability to explore and ensure that the data take the necessary form to produce the expected results. Even if the results are unexpected, they produce conclusions that are critical to decision making.

ML, a branch of Artificial Intelligence, is a field that allows for the generation of prediction models by using specific algorithms. These forecasting models are then used to develop decision-making tools. Companies today use these tools to optimize their products and/or services, as well as to develop new processes and solutions, all with the goal of innovating, staying current with market trends, and becoming more competitive. Dönmez (2013)

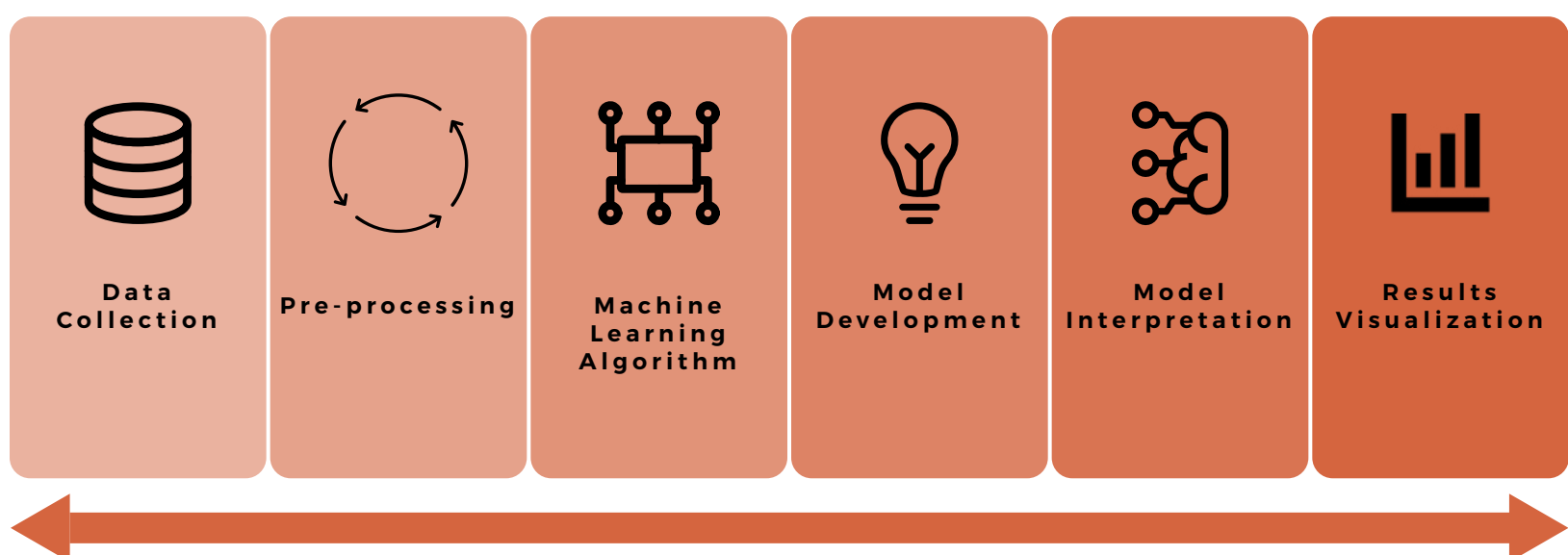


Figure 2.4: Processes of Machine Learning applied to data.
Source: Adapted from Medium

Data Science provides the language and techniques required for data comprehension and manipulation. It involves designing, collecting, analyzing, and interpreting numerical data in order to extract patterns and other useful information. Machine Learning, a subfield of Data Science, is concerned with the development of algorithms and computer resources that learn from data.

The benefits of using these mechanisms are realized through the use of these algorithms, which are required for data processing and allow for the automation of processes that will support decision making. (Sundarasan, 2011)

We recommend the following video to help with understanding of the ML process in Data Science:



2.5. Data Science and Decision Making: A Perfect Match

Decision making has always been a sensitive subject, because a decision can mean the difference between being one step ahead or not, creating a competitive advantage. Making a decision that has a negative impact on a company's mode of operation may have serious consequences for the company's future strategic and competitive positioning.

Data Science's ultimate goal is to improve decision making, which is critical for business, and it employs a methodology known as Data-Driven Decision Making (DDDM). DDDM is the practice of making data-driven decisions based on data analysis rather than just observation or intuition.

For example, a marketer could decide on his strategy based solely on his extensive experience and "eye." Alternatively, he could base his decision on an analysis of data on how consumers react to various ad campaigns.

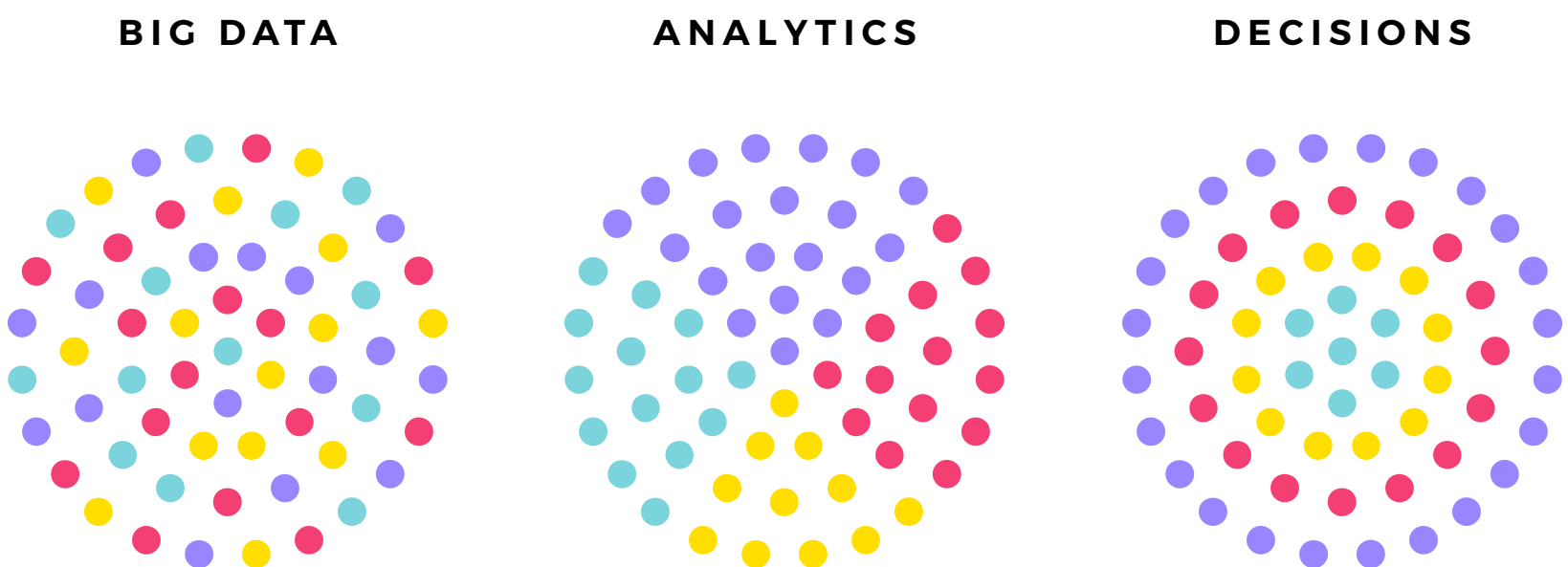


Figure 2.5: Data-Driven Decision Making (DDDM).
Source: *IdealMarketing*

When Hurricane Frances was on its way to the Caribbean and threatening a direct hit on the Atlantic coast of Florida, one **Walmart** store in the United States of America used data-driven decision making. The CEOs of Walmart stores in Bentonville decided that the situation presented a great opportunity to earn a competitive advantage using one of their newest weapons: predictive technology.

Linda Dillman, Walmart's chief information officer, pushed her staff a week before the storm to create forecasts based on what happened when Hurricane Charley hit several weeks earlier. Dillman believed that, backed by the "trillions of bytes" of purchase history stored in Walmart's databases, the company could "start predicting what's going to happen, rather than waiting for it to happen." (Provost and Fawcett, 2013)

Being 'one step ahead' is now an important mission in what companies' visions and business bets are. Knowing that there are methodologies that can be used to make better decisions is also a sign that there is a growing concern about fine-tuning what is fine-tuning of business decisions.

We recommend the following video to help you understand the importance of ML in decision making:





3. THE DATA AS IMPORTANT INFORMATION FOR ORGANIZATIONS

The science of data, its analysis, and visualization is handled as a key factor to consider for the business of organizations, highlighting of organizations, and highlighting of enterprises' business potential. Market examples are also provided, as well as an explanation of what a data scientist is.

Photo by [Nacional Cancer Institute](#) on [Unsplash](#)

With the help of **Data Science**, it is possible to transform, accelerate, and ensure the effectiveness and efficiency of existing business models. The basis for a careful analysis supported by algorithms that support the optimization of the optimization is the information that can be captured, the computing power computing capacity (local, distributed large data centers), and the efficiency of the communication systems (figure 3.1).



Figure 3.1: Data Science-Enabled Analysis

Big Data is a term that refers to the massive amounts of data – both structured and unstructured – that flood enterprises on a daily basis. This type of data can be difficult to handle due to its size and volume. But what companies do with the data is what really matters; it's not simply the variety or quantity of data that's significant; it's what they do with it. The analysis of large amounts of data can yield insights that help improve decision-making and give confidence when making important business actions. And when applied to analytics, **Big Data Analytics** is the process of analyzing massive volumes of data to discover hidden patterns in data, correlations, and other insights. With the technology that is available today, it is possible to analyze data and extract answers from it almost instantaneously. This is a process that would be longer and less efficient if it were done using more traditional business intelligence tools.

Companies that use **Data Science** to identify and create opportunities can gain a competitive advantage and, as a result, transform their business models and develop new ones for innovative growth. Companies can improve and optimize the supply of services, as well as support and customer relationships, using the data they collect (from various sources) and innovative aspects of **data analysis**.

Companies that use **Data Science** can discover new revenue, customer service, product, and service opportunities.

As Data Science continues to optimize **business processes**, there is a significant shift in focus toward discovering actual value with Data Science.

Companies today are accelerating their business growth by leveraging data's high **volume, value, variety, velocity and veracity (5Vs)** (Figure 3.2).

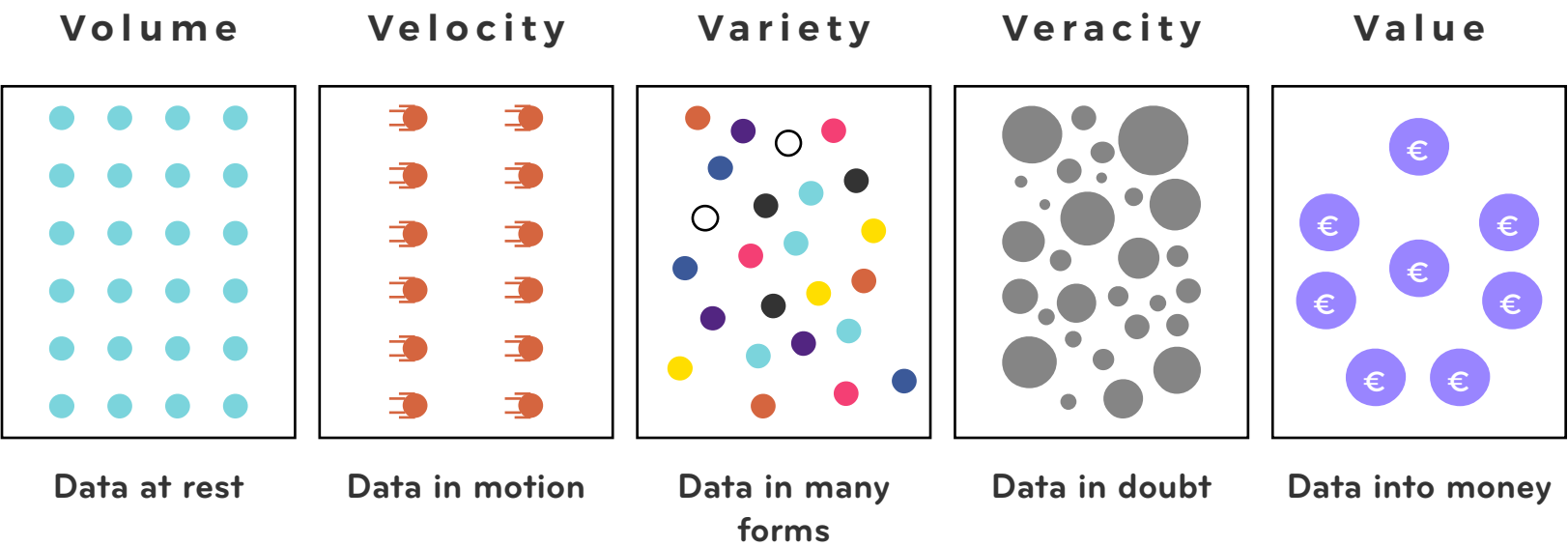


Figure 3.2: The 5Vs of Big Data – (adapted from *Information Catalyst* – informationcatalyst.com)

3.1. Data science opportunities for enterprise business

Data Science is not yet fully understood by businesses and the potential it represents. Data Science considers data from various sources and, as a result, in various formats. The data is mostly unstructured when it arrives, necessitating the use of algorithms to structure and analyze its useful information. There is now computing power with processing power to organize data and improve the quality of the results of its analysis. With current computing power (as a result of advances in electronics, hardware, and software), it is possible to obtain decisions that will help guide the business's path and strategy.

Data Science is evolving as an excellent tool to support business innovation and transformation through *Cloud Computing Technologies*, *Artificial Intelligence*, and the *Internet of Things*.

Data Science is a combination of data inference, algorithms, statistics, and technologies that, when combined, provide excellent business support through more complex analysis and more effective decisions. Data Science allows you to analyze customer behaviors, trends, and complex situations in order to make better business decisions. Google, Bing, Yahoo, and other search engines, for example, use Data Science algorithms to display the top search results (almost in real time).

3.2. The importance of data visualization in business

Data visualization is critical in business. Decision makers, CEOs, and departments in organizations can analyze data more effectively with the help of an effective data visualization tool. It would be impossible to analyze for a proper decision with the current volumes of information. Because of the complexity used in their presentation, visualization allows public and private organizations to have the appropriate information. It is possible to obtain results that have an impact on the prediction of consumer behavior or Key Performance Indicators (KPIs), that support business decisions using advanced analysis and visual modeling.

Figure 3.3 depicts some examples of specialized reference data visualization tools, both open source and custom developed.

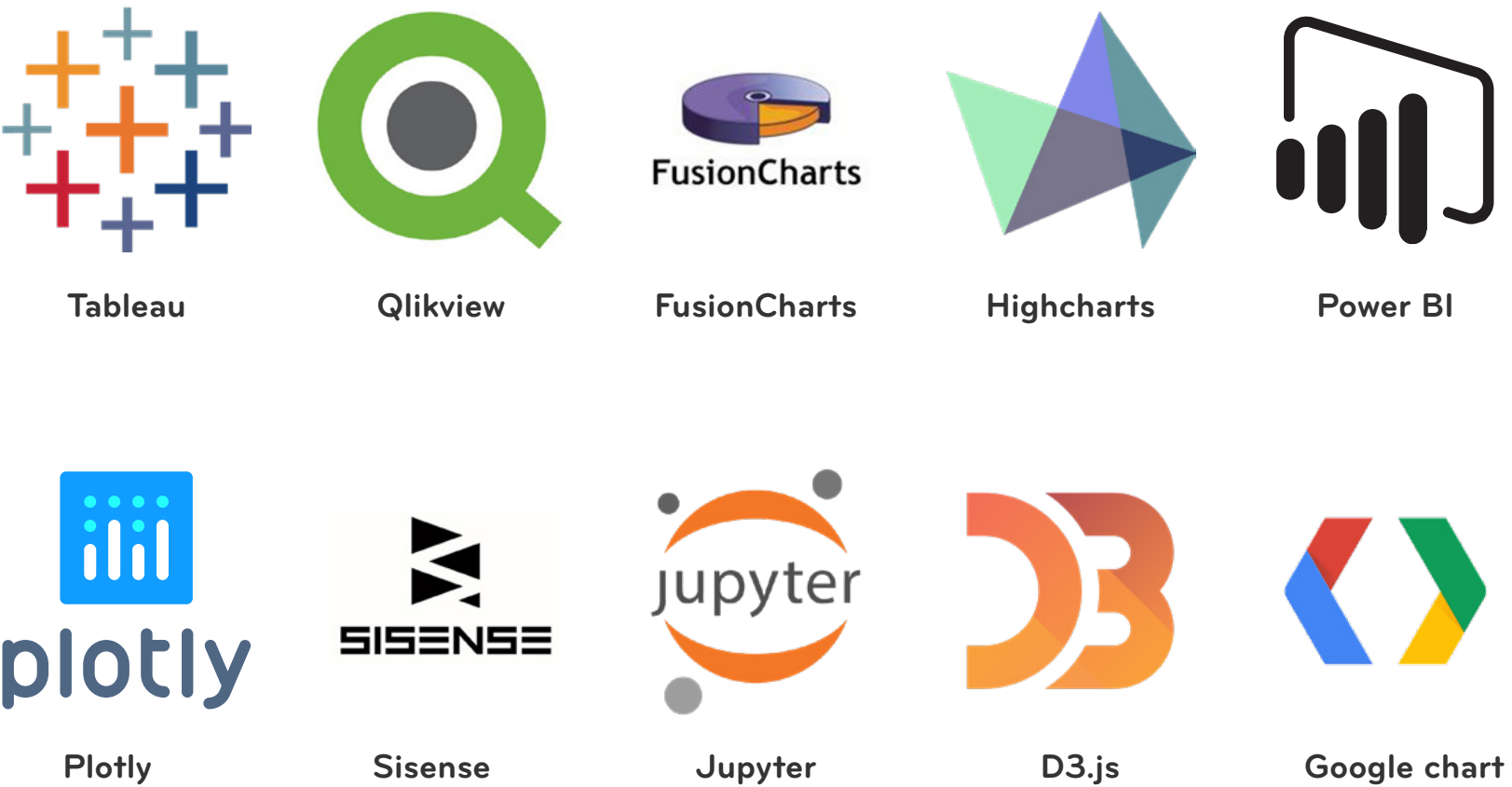


Figure 3.3: Data visualization software.

A Dashboard is a visual interface that contains information, metrics, and performance indicators about a company or organization. The purpose of the dashboard is to display numbers that are relevant to the business strategy and to the achievement of organizational goals.

Benefits of Data Visualization:

- i) Allows data (after processing) to be presented in tables and graphs and allows data to be viewed from different perspectives with a more complete dashboard and greater value in the information presented;
- ii) Allows data to be observed geographically (in space and time) in the form of geolocalized data, with important decision indicators;
- iii) Assists businesses in making intelligent decisions by allowing them to "get inside" the data and develop relevant analyses for the decision. Companies can observe patterns, trends, and correlations that support, for example, cost reduction, revenue increase, human resource needs, customers (type, location, etc.), production failures, commercial actions, and business growth and expansion

3.3. Data Analysis' Importance in Business

Data analytics can identify the path to a variety of business advantages, such as new revenue streams, efficient marketing practices, increased customer engagement, advanced operational efficiency, and competitive advantage. Data analytics enables data scientists to analyze an ever-increasing volume of data from various sources and formats that would otherwise go untapped by traditional Business Intelligence tools.

Data analytics, on the other hand, assists corporate bodies in deriving value from the data they have in order to make better decisions. Complex applications, such as predictive models, statistical algorithms, and hypothesis analysis, are combined with high-performance analytical systems in data analytics.

Data mining and Data Science enable the discovery of new opportunities, and extracting information from data enables better products, more efficient operations, higher profits, and higher customer satisfaction. Figure 3.4 shows how data analysis can be used to find value in a variety of ways.

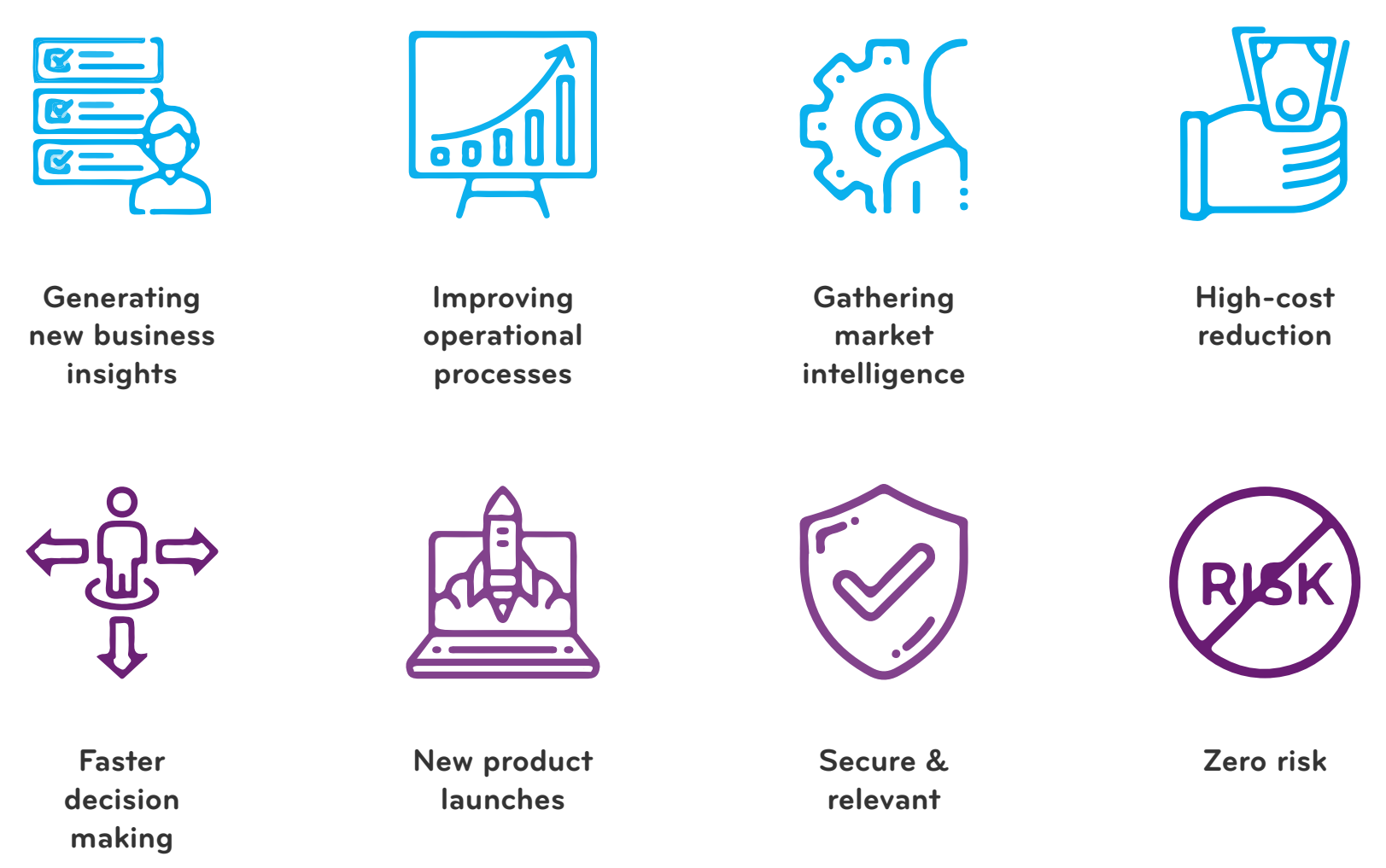


Figure 3.4: Different types of data value.

3.4. Markets that use and employ Data Science

Data science is used in various markets to improve management and procedures in the optimization of processes, products, and services, for example:

TELECOMMUNICATIONS

To help their customer support teams resolve customer queries more efficiently, telecom service providers use data analytics to analyze and process thousands of documents, images, manuals, customer data, and reports (in real time). This has an impact on their ability to solve customer problems in terms of saving time and money.

RETAIL

Customer service has evolved significantly in recent years, and customers now expect retailers to understand and respond to their needs in a timely manner. In this sense, retailers are leveraging Data Science to respond effectively, effectively, and efficiently to the needs of their customers. One example is the use of predictive analytics to forecast trends, future needs, and business strategies in order to increase profit for retailers and their customers.

HEALTHCARE

The healthcare industry is employing Data Science to conduct analysis across multiple dimensions, for example, healthcare providers are implementing data analytics to extract critical patient information. Data patterns and other information help physicians identify patients with chronic and other diseases so that they can provide timely and appropriate treatment to reduce risk.

MANUFACTURING

Manufacturing companies use data analytics to adjust and optimize production in order to avoid overstocking and balance the supply-demand relationship. Manufacturers have been using predictive analytics to create customized products that meet customer expectations while reducing inventory and warehousing infrastructure costs.

BANKING & INSURANCE

The banking and insurance sectors face numerous challenges, such as financial fraud alerts, tick analysis, card fraud detection, audit filing, credit risk reporting, trade visibility, and so on. Banks and other financial institutions can effectively monitor the financial market by using data analytics. To support trading, major banks and financial institutions are developing trade analytics using Data Science.

Many other industries, such as media, energy, technology, government organizations, fashion, and entertainment, use data analytics to improve business, customer satisfaction, and thus customer loyalty.

3.5. Data Scientists: The Key to Business Knowledge

Data scientists have emerged as critical human resources in almost every organization over the last decade. They are well-rounded individuals with advanced theoretical, technical, and practical knowledge who can build complex quantitative algorithms to organize and synthesize large amounts of data used to answer questions and drive strategy in their organization.

Data scientists must be curious and results-oriented, with exceptional knowledge and communication skills specific to the activity in which they perform their role, allowing them to explain highly technical results to non-technical coworkers. (Qian, 2021)



PROGRAMMING

Python and R are the two most common programming languages used in Data Science.



MATHEMATICS

Understand optimization and mathematical modeling concepts.



STATISTICS

Create statistical models and make inferences (predictions).



CREATIVITY

Being able to think outside the box and propose solutions that go beyond common sense.

Figure 3.5: Technical skills of a data scientist.
Source: *Sigmoidal.ai*

They have a strong quantitative background in statistics, mathematics, and algebra, as well as programming skills focused on data warehousing, data mining, and modeling in order to build and analyze algorithms.

Experienced data scientists are in charge of developing best practices for a company, from data cleaning to data processing and storage. They collaborate with other teams in your organization, such as marketing, operations, and customer relations. It is a highly sought-after profession, and the salary reflects the high level of responsibility they possess. (Berkeley University, 2018)

Data scientists are now considered as the key to 'first-hand' knowledge in businesses. They serve as a vehicle for strategic positioning excellence, allowing businesses to become more evolved in the niche markets they explore.

We recommend watching the video below on the most important data scientist skills.





4. TECHNOLOGICAL PLATFORMS

The most often used technology platforms and technologies in Data Science are presented, as are the major supporting platforms.

Photo by [ThisisEngineering RAEng](#) on [Unsplash](#)

4.1. The use of of technological platforms

A data scientist is someone who is in charge of extracting, manipulating, pre-processing, and making predictions from data. He or she will need a variety of statistical tools and programming languages to accomplish this. A data scientist typically employs one (or more) technological platforms to simplify his work throughout the processes.

These platforms are very useful because they allow you to analyze and process data in an accessible, fast, and intuitive manner, and some also allow you to present the results of the data study through the use of dashboards.

As previously stated, Data Science is interdisciplinary in nature and can be applied to any scientific field or economic sector. However, the platforms that allow data to be processed in order to apply all methodologies up to decision making reside in the technological domain. The following section discusses the technological applications that enable the implementation of Data Science methodologies.



Figure 4.1: Technological platforms for Data Science

Source: INSAID - International School of Artificial Intelligence & Data Science

4.2. Applications to help with the use of Data Science

In Data Science, there are data collection, analysis, and processing platforms. There are also 'all-in-one' platforms that allow you to perform all of the processes at once. There are platforms that are free, paid, and open source. Some require programming language experience, and some don't. The following are the top eight most popular Data Science applications (DataFlair, 2021; Mujthaba et al., 2020):

[RapidMiner](#)

RapidMiner is a tool that allows you to implement the entire predictive modeling life cycle. It has complete data preparation, model building, validation, and implementation capabilities.

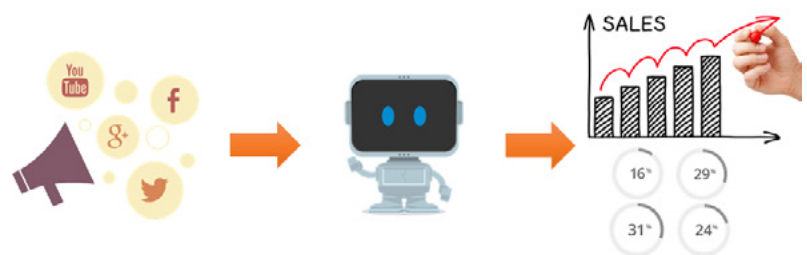
Figure 4.2: RapidMiner features
Source: Edureka



[Data Robot](#)

Data Robot is a platform for implementing machine learning-based algorithms. Appropriate for data scientists, CEOs, software engineers, and IT professionals.

Figure 4.3: Social media data collection and subsequent analysis using the Data Robot platform.
Source: Data Robot



[Alteryx](#)

Alteryx is a data discovery, preparation, and analysis platform. It also helps in the discovery of deeper insights by deploying and sharing analysis at scale.

Figure 4.4: Manipulating data from various sources in Alteryx.
Source: TrueCue



 [SAS](#)

SAS is a Data Science tool that was created specifically for statistical operations. SAS is a proprietary, closed-source data analysis software used by large corporations. SAS performs statistical modeling using the SAS programming language.

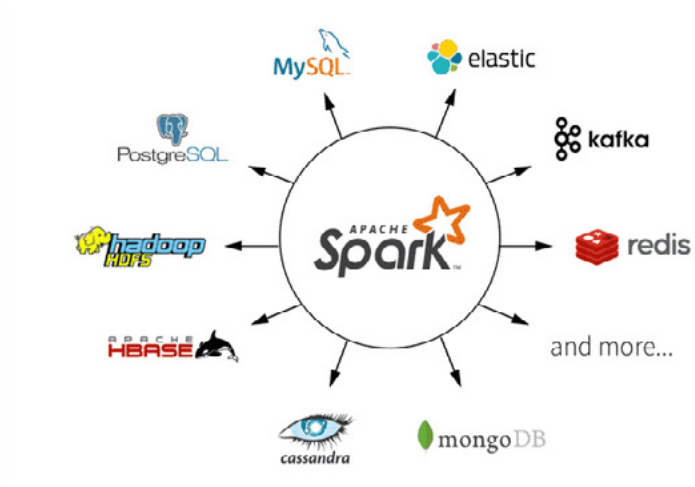
Figure 4.5: Features of the SAS platform.
Source: Edureka



 [Apache Spark](#)

Apache Spark, or simply Spark, is a fairly powerful open source analytics tool focused on speed, usability, and flow analysis. Apache is the company behind this tool.

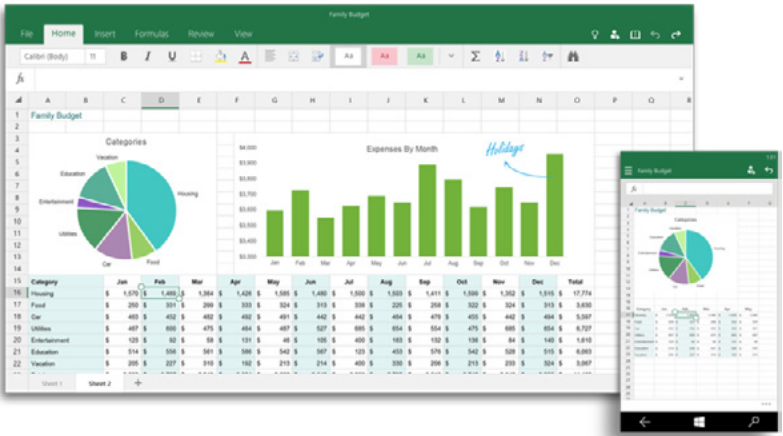
Figure 4.6: Supported tools for Apache Spark integration.
Source: DataBricks



 [Microsoft Excel](#)

Despite its popularity and age, it is still the most widely used Data Analysis tool. Excel is a powerful data analysis tool. Excel retains its prestige despite being the traditional tool for manipulating spreadsheets.

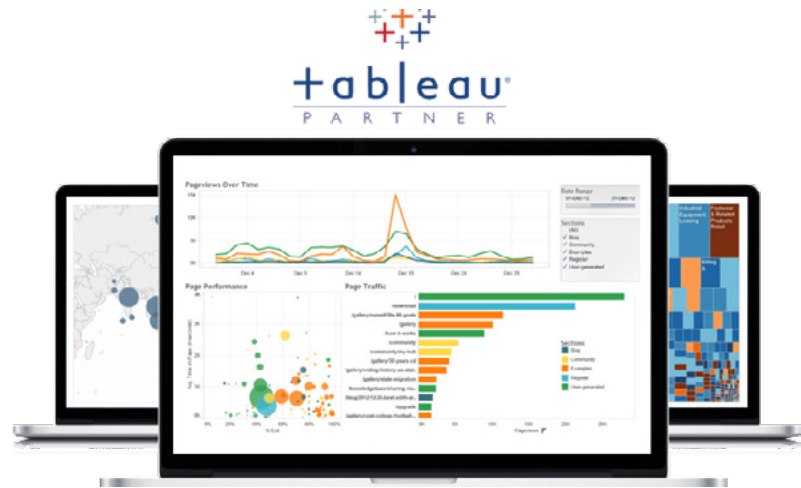
Figure 4.7: A dashboard created with Microsoft Excel.
Source: VHV.rs



[Tableau](#)

Tableau is a data visualization software that can create powerful graphs for interactive visualizations. It focuses on industries involved in business intelligence.

Figure 4.8: Example of a dashboard created using the Tableau platform.
Source: VHV.rs



[Jupyter](#)

Jupyter is a free and open source programming environment that allows developers to create open source software and experiment with interactive programming. Jupyter supports a number of programming languages, including Julia, Python, and R. It is a tool that allows you to present visualizations and interactive content by allowing you to write code on the fly. Jupyter is a well-known data science tool. It's completely free and open source

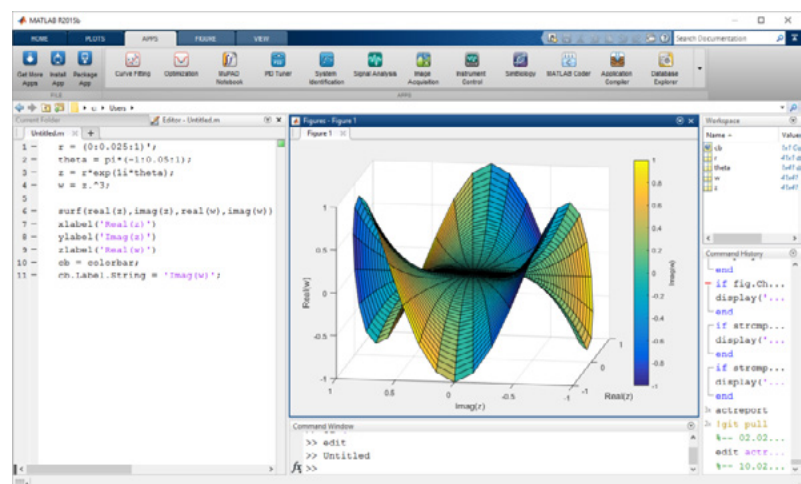


Figure 4.9: An example of Jupyter application in the Python programming language.
Source: DataCamp

[MATLAB](#)

MATLAB® is a platform for programming that was developed primarily for engineers and scientists to evaluate and design systems and products that transform the world. The MATLAB language is at the core of the MATLAB environment. It is a matrix-based language that enables the most natural expression of computational mathematics.

Figure 5: Source: Princeton Research Computing (researchcomputing.princeton.edu)



[RStudio \(Posit\)](#)

RStudio, which has now been rebranded as Posit, is an integrated development environment (IDE) that supports the programming languages R and Python. It comes with a console, an editor with syntax highlighting that allows for direct code execution, as well as tools for plotting, history, debugging, and workspace management. RStudio is a desktop application that is available in both open source and commercial editions (Windows, Mac, and Linux).

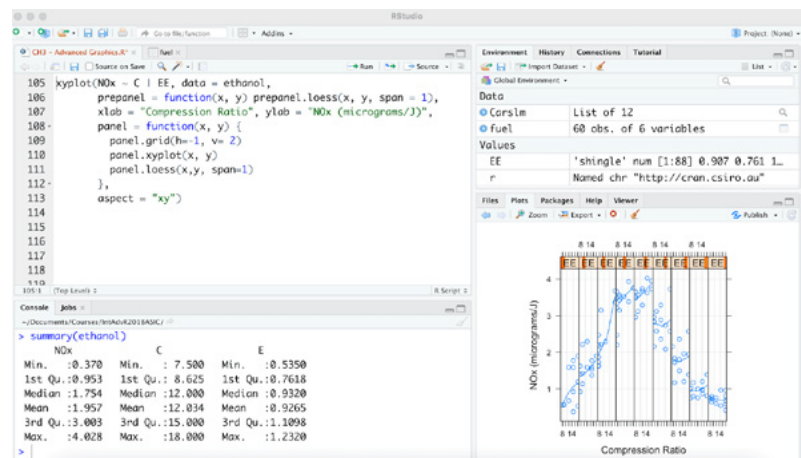


Figure 5.1: Source: *SolutionMetrics* (solutionmetrics.com.au)

[MySQL](#)

MySQL is a structured query language-based (SQL) relational database management system (RDBMS) created by Oracle. It is a well-known technology in the current big data environment. MySQL is sometimes referred to as the most popular database, and it is currently in broad, successful usage across industries. It is evident that everyone engaged with business data or general IT should seek for a fundamental understanding of MySQL.



Figure 5.2: Source: *Simplilearn* (simplilearn.com)

[TensorFlow](#)

TensorFlow is an open-source, comprehensive framework for machine learning that was created by Google. It includes a rich and flexible ecosystem of tools, libraries, and community resources, which not only allows developers to quickly design and deploy applications driven by machine learning, but also enables researchers to advance the state of the art in machine learning.

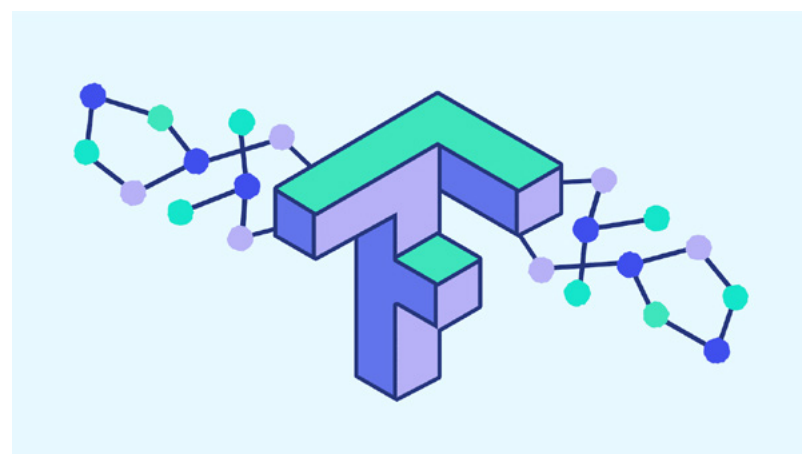


Figure 5.3: Source: *DataScientest* (datascientest.com)



5. OPPORTUNITIES, SOLUTIONS AND TRENDS

The employment prospects for a data scientist are described, as well as the solutions available in the various areas/sectors. Data science trends are also discussed.

Photo by [National Cancer Institute](#) on [Unsplash](#)

5.1. What are the main opportunities in Data Science?

Large amounts of data are generated every day, so it is relevant to address the topic of Data Science, i.e., **Data Scientist**, career opportunities.

According to a Harvard Business Review review, a Data Scientist "is a high-level professional with the skills and capacity to perform analysis in the world of Big Data." In this regard, Data Scientists are highly desirable professionals in the Big Data Analytics and IT industries (edureka, 2022).

Career opportunities in Data Science are evident due to a scarcity of skilled professionals who can "work" data to aid in decision making more effectively and efficiently.

It can be seen that there is a need for this type of professional to join the staff of large organizations (private and public), as well as startups, where their business is strongly linked to decision making supported by data analysis.

According to the Bureau of Labor Statistics (BLS), statistics in Data Science-related careers will grow 33.8 percent between 2016 and 2026, making it the fastest growing occupation in the mathematical industry.

A Data Science Career Guide can be found at [discoverdatascience](https://discoverdatascience.com/) (2022), which presents various types of careers in the Data Science field and summarizes the qualities of a Data Scientist.

5.2. What Data Science Solutions Exist?

Almost every industry has been transformed by data scientists. In healthcare, their algorithms help in the prediction of patient side effects. Their models and metrics have redefined "athletic potential" in sports. Data science has even been used to improve traffic, with route optimization models that take into account typical rush hours and weekend slowdowns.

DATA SCIENCE APPLICATIONS AND EXAMPLES

- Disease detection and prevention- Personalized medical advice
- Real-time shipping route optimization
- Identifying the next wave of world-class athletes
- Tax fraud eradication
- Automated digital ad placement
- Algorithms for finding love
- Predicting incarceration rates

Data Science should not be confused with data analysis. Both fields are ways of understanding Big Data, and both generally involve the analysis of massive databases using R and Python. Because of all these points of overlap, the fields are generally treated as one, but they differ significantly. For one thing, they have different relationships with time. Data analysts synthesize Big Data to answer specific questions about the past, such as "How did our subscriber base grow from 2016 to 2019?" In other words, they mine Big Data to gain insights into what has already occurred. Meanwhile, data scientists build on Big Data by developing models that can predict or analyze what will happen next. Of course, it is impossible to accurately simulate all of life's complexities. "All models are wrong, but some are useful," statistician George E.P. Box famously stated. Nonetheless, Data Science at its best can make informed recommendations about key areas of uncertainty (Rice, 2019).

Table 5.1 provides some examples of Data Science-based solutions. Many other startup-developed solutions emerge on a daily basis.

Table 5.1: Examples of data science-based solutions.

FIELDS/SOLUTION			
1	2	3	4
Healthcare			
Medical Image Analysis	Disease Phenotyping	Malignant tumor detection	Patient monitoring
Travel & Logistics			
Targeted marketing	Warehouse Management and Optimization	Route optimization	Travel and accommodation price forecasts
Sports			
Predicting Match Results	Health and physical performance monitoring	Scouting decision-making support	Viewing players heat maps
Government			
Fraud detection and tax evasion	Criminal activity monitoring	Cybersecurity protection	Rapid response from emergency services
Crime Prevention			
Intelligent Video Surveillance	Profile and background analysis	Natural Language Processing	Predictive policing
Finance			
Risk analysis	Real-time data analysis	Algorithms for trading in financial markets	Price forecasting in financial markets
E-Commerce and Marketing			
Price and Delivery Optimization	Recommendation systems	Product/service strategy optimization	Supply chain forecasting
Social Life			
Resource optimization in home care	Algorithms for adaptive learning	Remote and Virtual Assistance	Humanitarian Aid Resource Distribution Optimization

5.3. What are the current Data Science trends?

Companies use Data Science to optimize business processes and make the best decisions based on data analytics insights.

With the pandemic, SMEs and large enterprises were forced to adapt to changes in less time, necessitating increased investments in Data Science. We are seeing new trends emerge across industries as businesses rely on data analytics to overcome various challenges. One example of a development is Gartner's AI 2022 trends. The trends have been classified into three broad categories: accelerating change, operationalizing business value, and distributing everything (data and insights).

Companies, regardless of industry, place a high value on Data Science as a great business support tool. Figure 5.1 depicts the most significant Data Science trends.

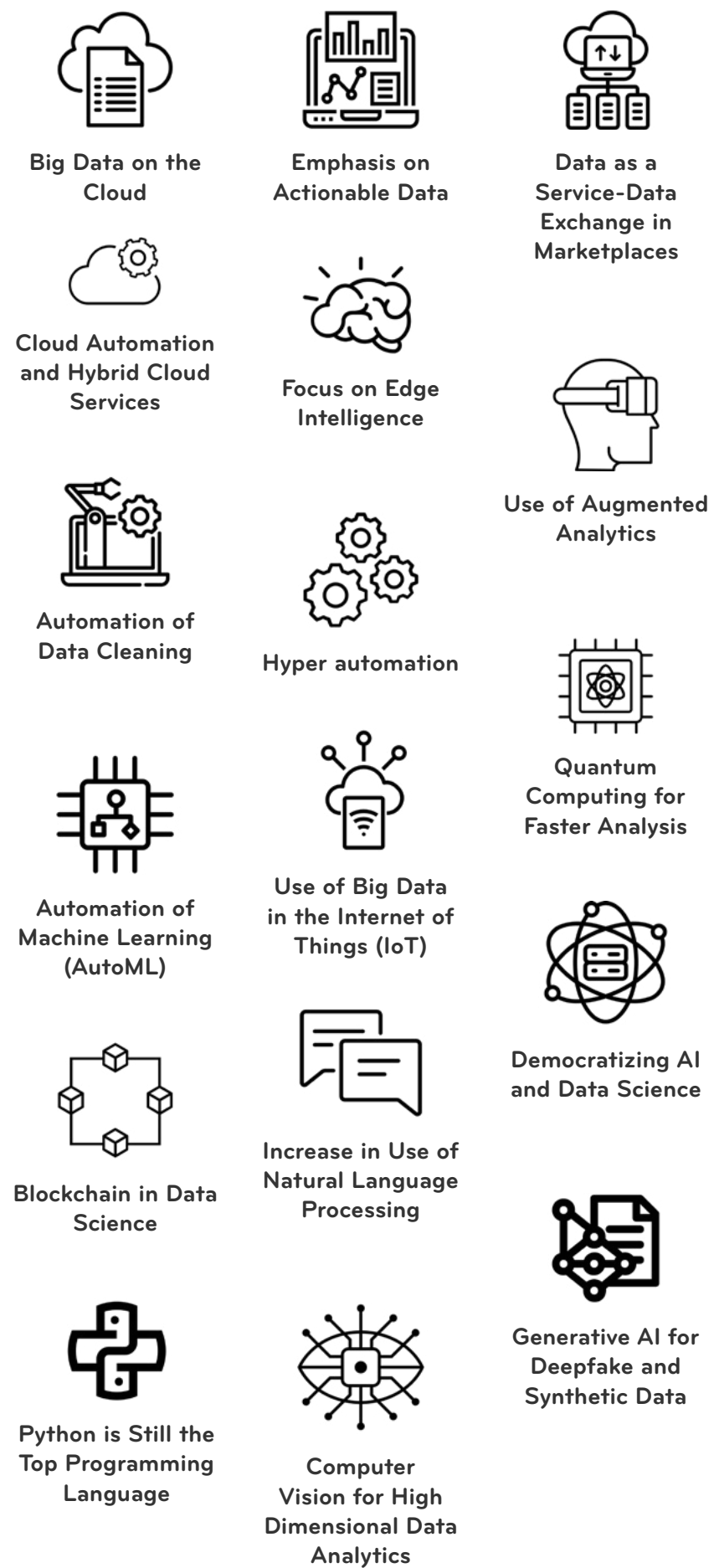


Figure 5.1: Key Data Science trends for 2022.



6. THE IMPACT OF THE INTELLIGENT DATA USE

The importance of intelligent data exploitation, prediction as a result of effective data exploitation, and Data Science, as well as ensuring privacy and anonymity, are discussed.

Photo by [Alexander Sinn](#) on [Unsplash](#)

6.1. The importance of intelligent data exploration

The goal of data exploration is to learn about a dataset's characteristics and potential problems without having to make assumptions about it beforehand. Data exploration is now a required component of all Data Science projects. But how can intelligent data exploration provide benefits? (Piringer, 2020)

Manual / traditional data exploration techniques can help in identifying specific areas of interest, which is feasible but falls short of more in-depth investigation. This is where machine learning can revolutionize data analysis.

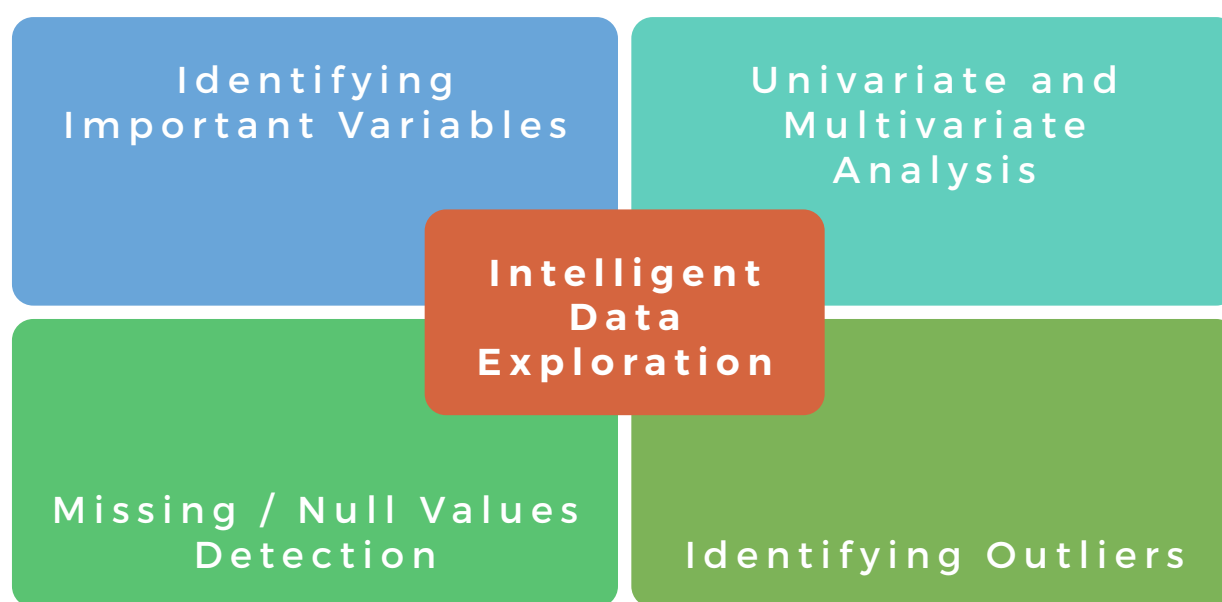


Figure 6.1: The steps for intelligent data exploration.
Source: Adapted from Sisudata

Both traditional data mining and data mining supported by ML algorithms can identify patterns and help draw conclusions from data. However, machine learning allows data scientists to extract information from large databases much faster, minimizing potential errors.

With more data available than ever before, many businesses are faced with an abundance of data. However, they lack the necessary resources to analyze and process it. This is where machine learning algorithms come into play.

The software used to implement ML can also make the data much easier to 'digest.' Companies can extract meaningful information quickly and without spending time interpreting and questioning the results by exporting specific points in the data into data visualizations such as bar charts or scatter plots. (Henn, 2021)

6.2. The prediction as a result of proper data exploration

Forecasting is also a delicate matter, because a good forecast, such as tomorrow's weather, enables farmers to better adapt their work to the weather. This example is further evidence that if it is possible to make predictions, and if these predictions match reality, then it is certain that decisions made in advance will yield significantly better results.

Predictive Analysis is a type of advanced analysis that is used to forecast unknown future events. The use of data in ML algorithms and techniques to identify the likelihood of future outcomes based on historical data is known as predictive analytics. The goal is to investigate what has occurred in order to better predict what will occur in the future. (SAS, 2016)

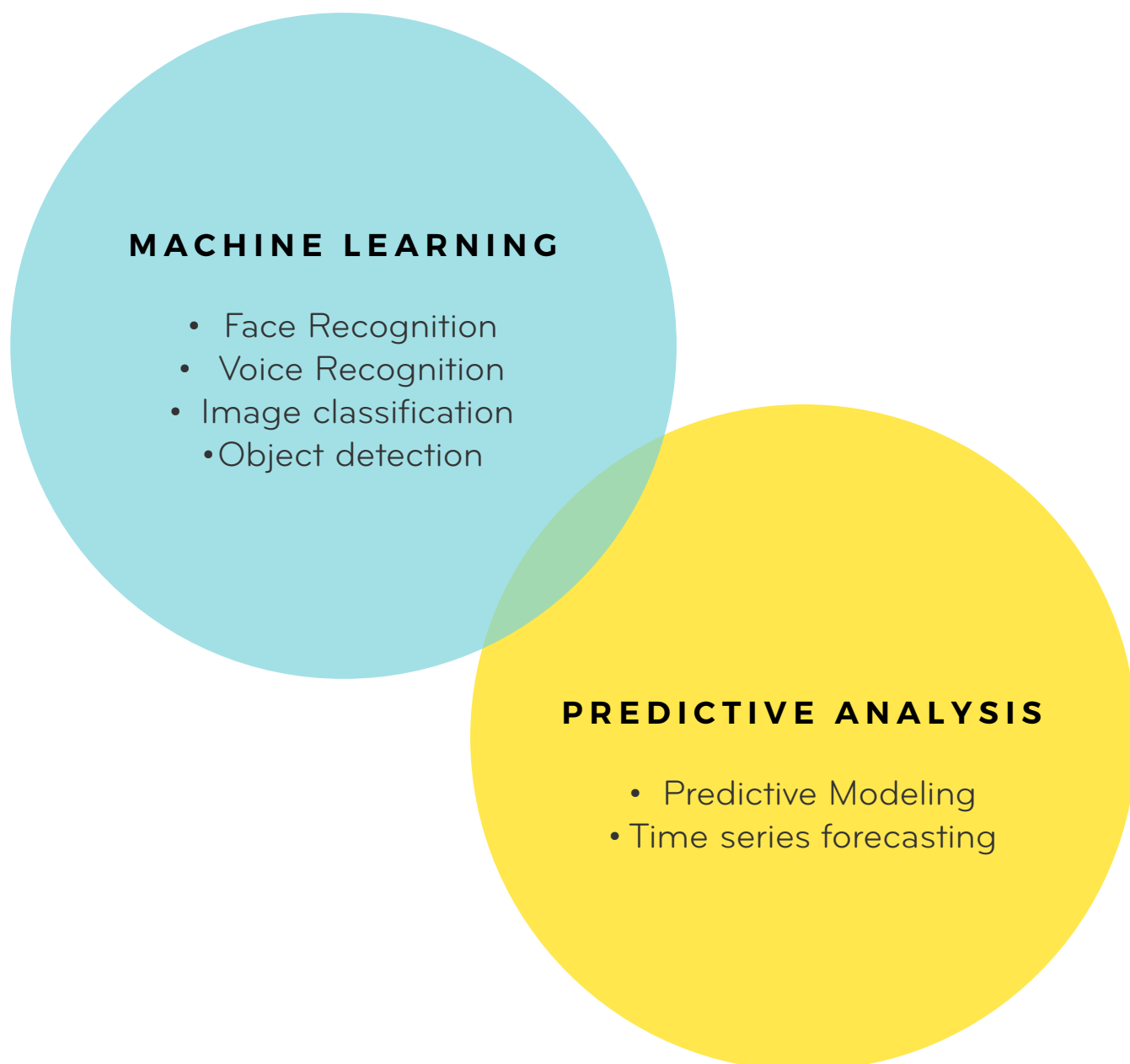


Figure 6.2: Machine Learning and Predictive Analytics Comparison
Source: Adapted from Sisudata

However, a common misunderstanding is that predictive analytics and machine learning have the same definition. Machine Learning and Predictive Analytics approach the problem in different ways. Predictive analytics can sometimes be integrated with different ML applications. The techniques supported by ML are more adaptive, newer, and have greater degrees of freedom, allowing it to be more flexible in its approach to a problem. Predictive analytics has been around longer and is more procedural in its application. (Johnson, 2020)

Predictions are extremely important for decision making because they use previous experience to make corrections and improvements to current processes.

We recommend watching the video below, which covers topics such as predictive analytics and ML-supported techniques.



6.3. Data Science: Privacy and anonymity assurance

Privacy is also one of the most controversial topics in Data Science, because when dealing with sensitive data, companies must guarantee its security in order to prevent leaks of sensitive information and becoming potential cybercrime targets.

Regarding data transparency and data protection, the application of Data Science methodologies is subject to considerable 'tension.' Without data collection and the free flow of information, Data Science would not exist. However, as more data is collected, data protection becomes more complicated.

Data scientists must find a way to ensure that data is both secure and accessible (for when legal disclosure is required), as well as shareable and editable.



Figure 6.3: GDPR is a European Union regulation that ensures the privacy of personal data..
Source: *Medium*

There are several ways to overcome these obstacles. One of them is data anonymization, which involves de-identifying the collected data and separating it from the actual people to whom it refers. In fact, many data privacy regulations, including the GDPR, have enforced the anonymization of data. Due to the irreversibility of data anonymization, this is not the most optimal business strategy. If you anonymize all personal information from data collected from prospective customers, it becomes very difficult to use that data in Data Science.

However, there are methods that guarantee both data integrity and privacy, and data generalization is one of them. Instead of storing personal information about customers, businesses "lump" data into fairly broad categories (age groups, geographic areas, etc.), ensuring that the data cannot be converted back into its granular form, thereby preserving the anonymity of the data and the privacy of their customers. (Wieringa et al., 2021)

We recommend watching the following video for a more comprehensive overview of privacy assurance and anonymity techniques in Data Science.



Randomized Response

Plausible Deniability


$$\hat{p} = \frac{E[y] - q}{1 - 2q}$$

p = True proportion
q = $\frac{1}{6}$ E[y] = results

© 2017 Microsoft AI. All Rights Reserved.



7. CASE STUDIES

The case studies refer to certain Data Science companies that are emerging in Portugal as well as international Data Science companies that will arise in 2022. Several funding opportunities for Data Science businesses are also mentioned.

Photo by [Walkator](#) on [Unsplash](#)

7.1. Some Data Science companies that are emerging in Portugal

Some examples of startups and other Portuguese companies that have already consolidated and are expanding their Data Science activity are presented as good examples of Portugal with dynamic teams and market value.

 [Unbabel](#) is a Portuguese startup that is revolutionizing the world of online translation by providing quality translations to businesses using a combination of AI technology and human quality regulators. It has offices in Lisbon and San Francisco (Figure 7.1).

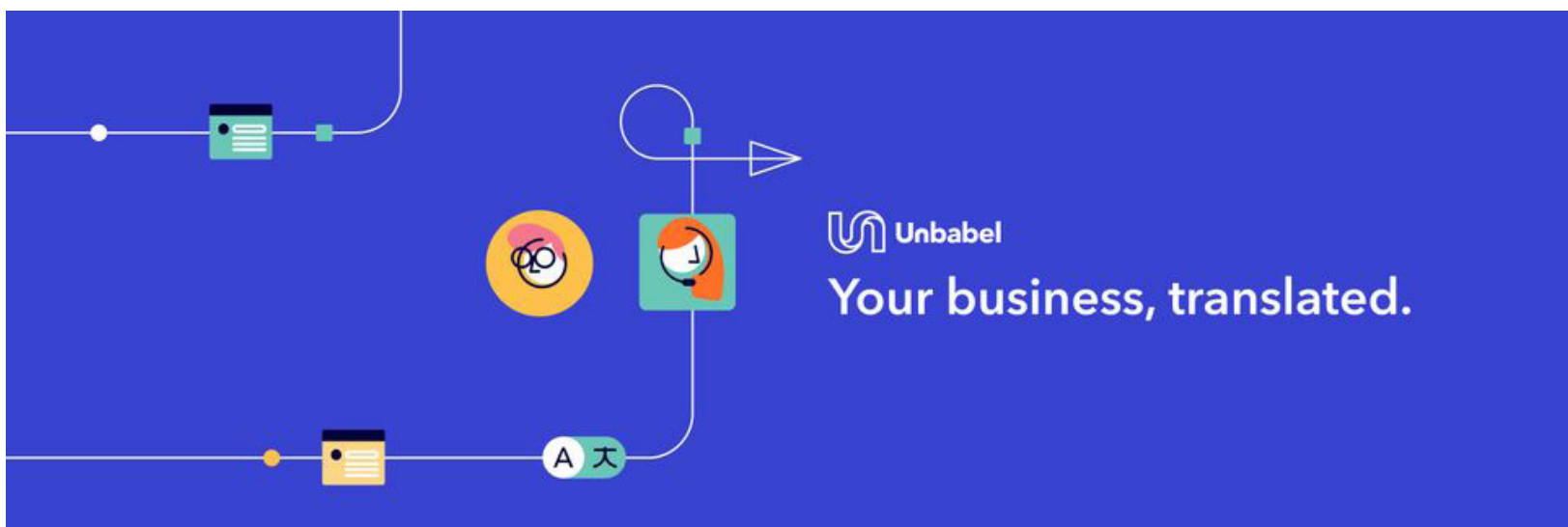


Figure 7.1: Unbabel in online translation with Data Science.

 [Uniplaces](#) is an online room rental platform that has quickly come to dominate the real estate market in several European capitals, including Lisbon. The brand's success is greatly supported by the simplicity of its website. It has offices in London, Madrid, and other major European cities, with its headquarters in Lisbon. Figure 7.2 depicts a platform example.

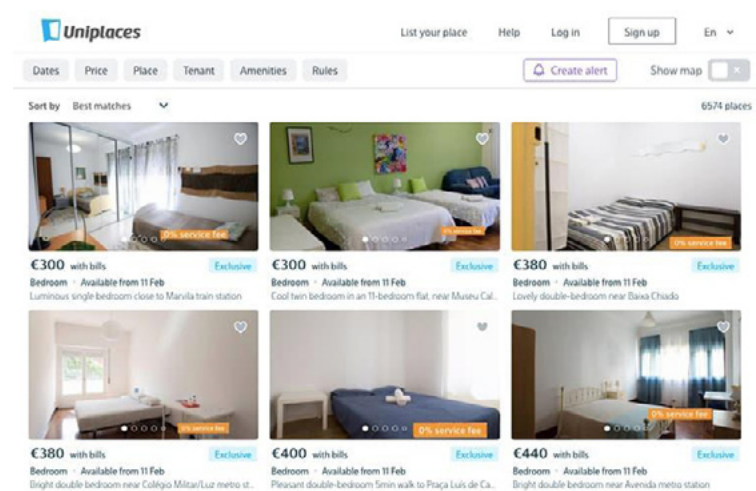


Figure 7.2: Uniplaces platform.

 [Prodsmart](#), a Portuguese company, received funding from other Portuguese startups that needed to help small businesses with their production processes. For production optimization, the application provides analysis and intelligent management. The production monitoring reports take into account information obtained in real time (Figure 7.3) via mobile devices used on the factory floor, allowing for real-time monitoring and analysis of the production flow, such as orders, workers, materials, machines, maintenance, waste, time, and costs.

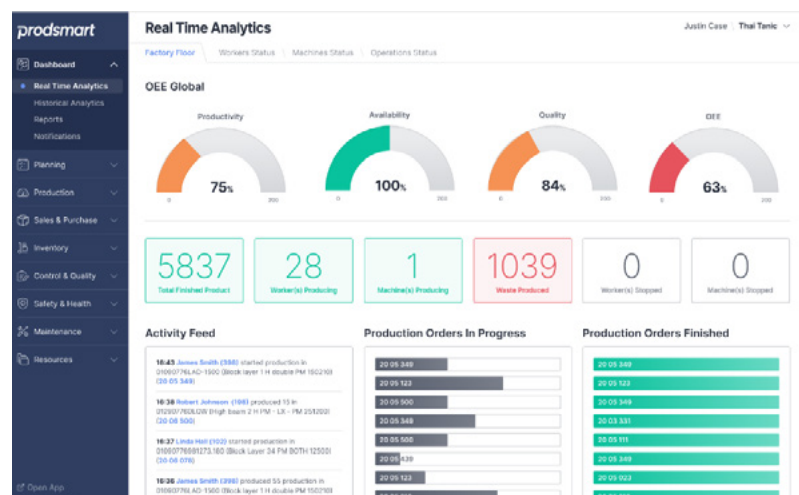


Figure 7.3: Prodsmart real-time analysis

✍ [James](#), formerly known as CrowdProcess, has created an Artificial Intelligence platform for credit risk. The system was created by and for credit experts with the goal of ensuring that risk teams have constant access to Data Science algorithms as well as up-to-date validation and monitoring requirements.

 [Talkdesk](#) was founded in 2011 after the founders identified a market gap for call center software. The company creates software that allows users to quickly and easily set up call center services. Cloud computing is used to power the various services.

 [Defined.ai](#) considers high performance models with training data from AI algorithms all over the world. Defined.ai offers an AI training platform. It has over 40 customers in the United States, including Mastercard, BMW, EDP, Grupo Mello Sade, Amazon, and Sony. They operate in the banking and insurance, retail, and telecommunications sectors. The company is currently valued at 35.2 million euros and has 140 employees across three continents.

 [Feedzai](#) was founded with 74 million euros and is now worth 521 million euros. Its primary goal is to secure banking activity through the use of anti-fraud Artificial Intelligence. It is one of the most promising financial technology startups in Europe. Its offices are currently located in Lisbon, Porto, Coimbra, Silicon Valley, New York, and the United Kingdom.

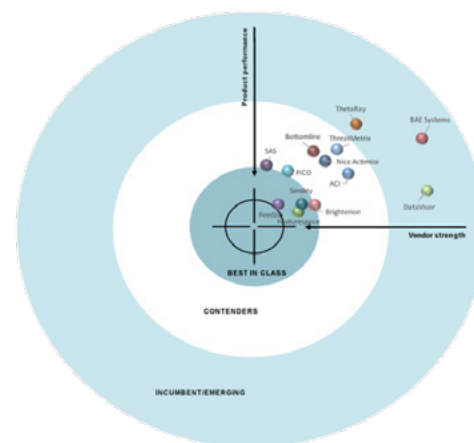
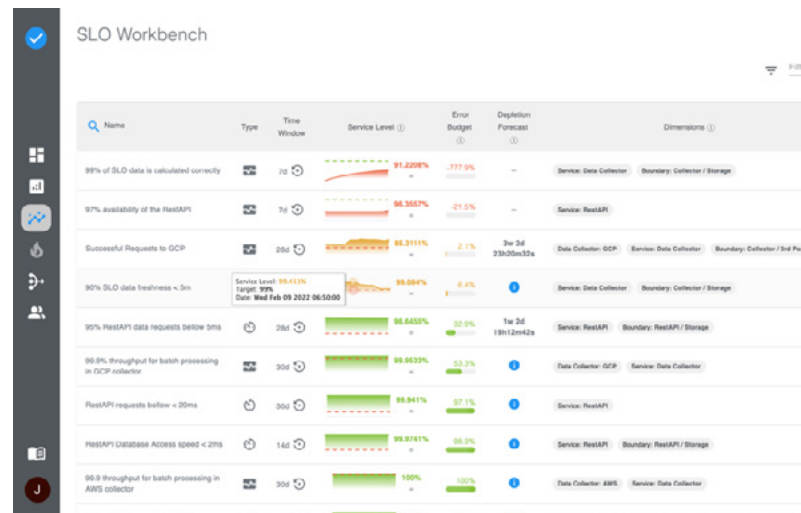


Figure 7.4: Feedzai in banking security.

 [Detech.ai](#) is an Artificial Intelligence startup that provides a platform for converting monitoring data into information that will assist in business planning. The goal of Detech.ai is to improve customer experience through website reliability engineering. This startup was founded in 2021 with headquarters in Almada and a 400k€ investment from four investors.

Figure 7.5: Detech.ai on improving customer experience.



 [GoWithFlow](#) develops technology that helps companies and cities reduce their carbon footprint. The software tracks how company vehicles, scooters, and charging stations are used, allowing businesses to plan and manage mobility services. The company was established in 2019 and is headquartered in Leça Da Palmeira, Porto, Portugal. This startup has already made \$5.5 million in revenue.

 [Gorodata](#) is an Artificial Intelligence and software company that operates a cloud-based platform using Artificial Intelligence to classify data in order to identify relevant information to increase company sales. Gorodata provides businesses with customizable buying indicators to help them know when their services are required. The company was founded in 2020 and is based in Cascais, Portugal. It currently has between one and ten employees. The company received a 475k€ funding round in April 2021.



Figure 7.6: Gorodata sales analytics platform supported by Data Science.

 Other companies like [WeChangers](#), [Growhappy](#), [UNI2ALL](#), [Sensei](#), [Meight](#), [Tesselo](#), [Shimejito](#), [Virtuleap](#), [Visor.AI](#), [Criam](#), [Probely](#), [UpHill](#), etc... were recently created Portuguese startups that use Data Science in their business models.

7.2. International Data Science companies that are emerging in 2022

Global corporations are currently implementing Data Science. Using a variety of Data Science tools, businesses are preparing, organizing, developing, and employing data. **Emerging Data Science startups** offer cutting-edge, innovative tools for managing and analyzing massive amounts of data. A number of startups provide solutions to businesses in a variety of industries using a variety of strategies. According to Analytics Insight, the following startups are the top 10 startups that are emerging in 2022 (Sanyal, 2019):

- ✍ [Cropln](#) is a market leader in the Software as a service (SaaS) model for the agricultural industry (Figure 7.7). Cropln began its business by providing solutions that take into account financial analysis, weather forecasting, data interpretation, satellite monitoring, geotagging, Artificial Intelligence, and data analytics. Cropln's objective is to monitor the agricultural sector.



Figure 7.7: Cropln uses Data Science in an agricultural SaaS model.

 [Firebolt](#) is a startup that enhances cloud data storage platforms for more efficient and effective analytics. This company's SaaS solution provides quick query performance and combines the simplicity, elasticity, and low cost of the cloud with analytics-level innovation. It was designed specifically for Amazon AWS and has an efficient SQL query engine that separates storage and compute, enabling users to create multiple isolated resources within the same database (Figure 7.8).

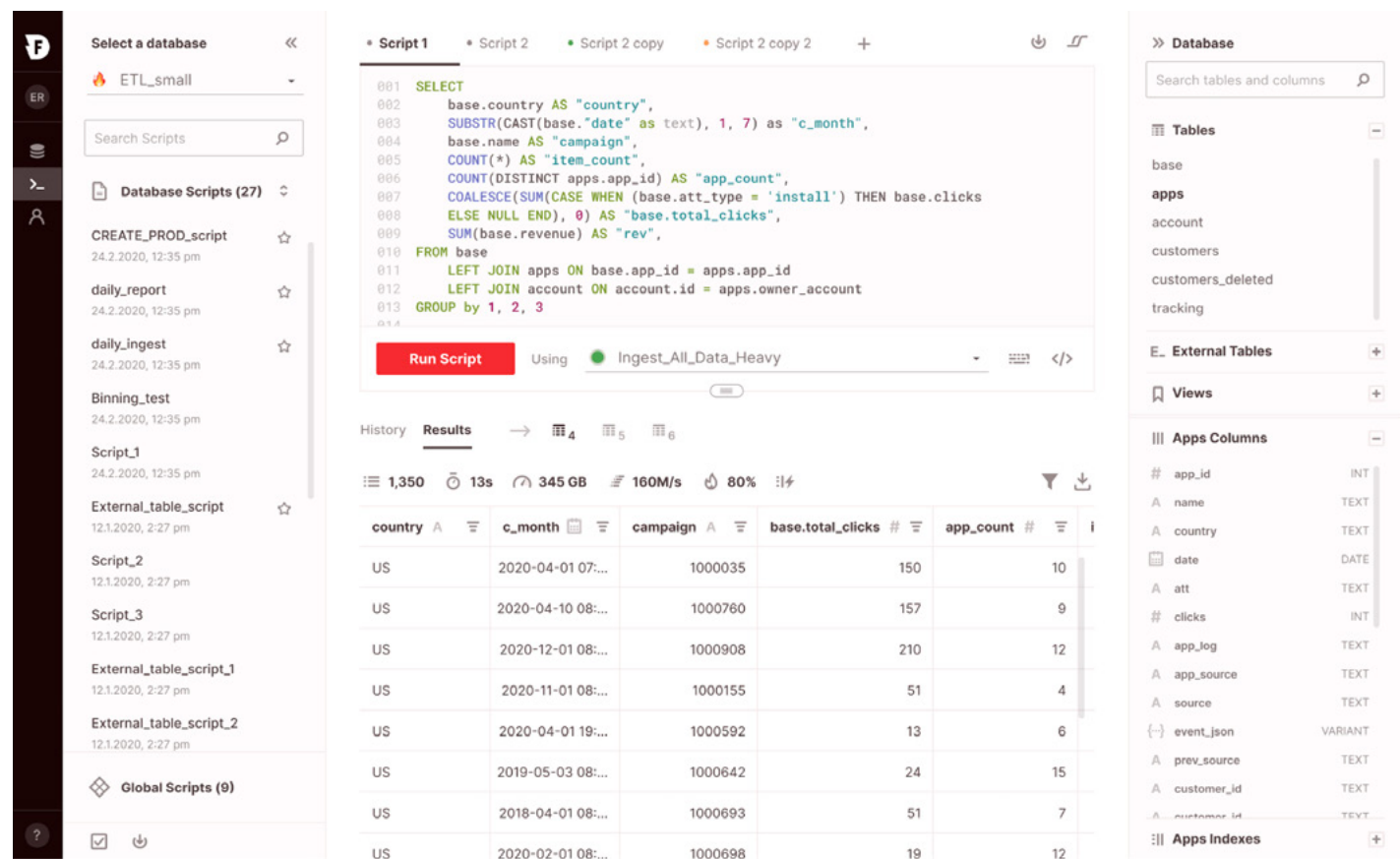


Figure 7.8: Firebolt with SaaS computing and data storage solutions.

 [Lymbyc](#) is a startup with a team of data scientists and analysts who have created robust analytical tools supported by cutting-edge technologies. The primary objective of this sophisticated Data Science infrastructure is to implement a practical and optimized decision-making system for business leaders, allowing them to structure business forecasting using business intelligence and natural language processing-driven data inputs (NLP).



Figure 7.9: Lymbyc with SaaS Data Analytics solution for business leaders.

 [InCountry](#) is a platform (as a service) for data storage, processing, and regulation that enables global data compliance for international businesses. The platform securely stores, processes, and regulates business data in the home country, assisting multinational corporations in scaling by allowing for quick integration. The company was founded in 2019 and has operations in 65 countries.

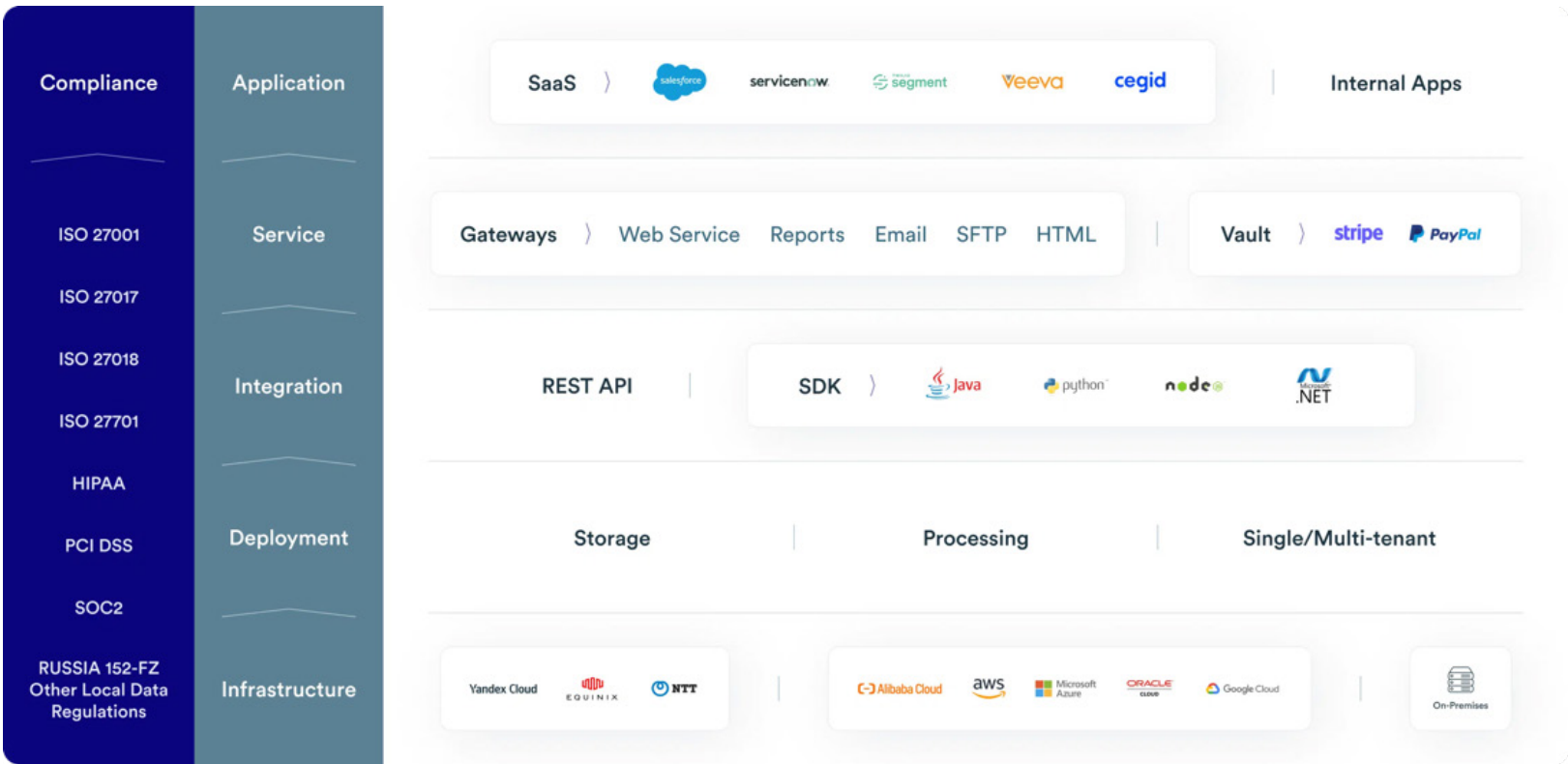
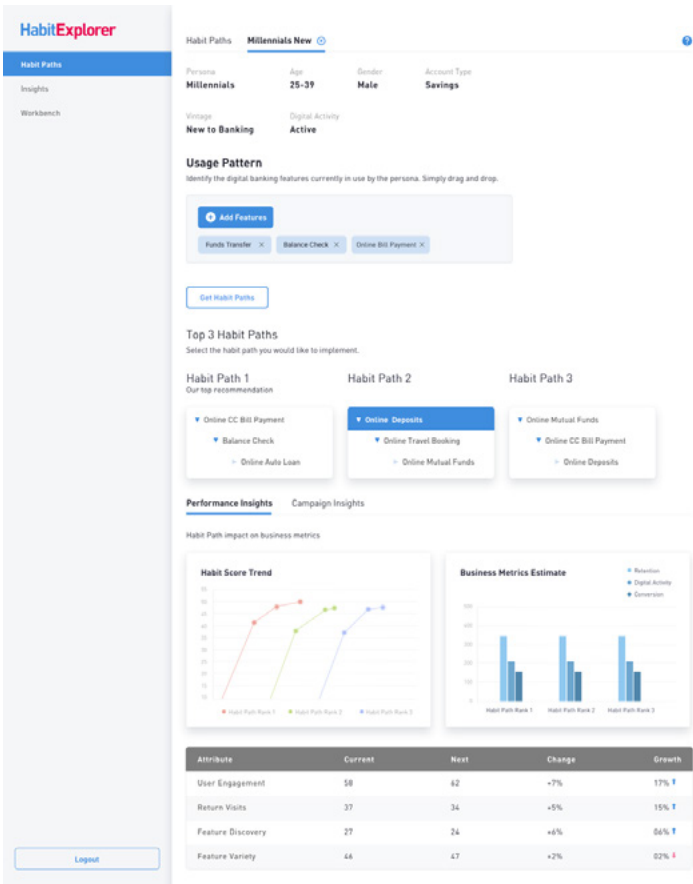


Figure 7.10: InCountry offers a data analytics platform for companies as a SaaS solution.

 [3LOQ Labs](#) uses advanced machine learning algorithms to identify individual customer behavior based on past actions and context. 3LOQ powers financial management systems (cash and assets) in terms of digital banking services, allowing end customers to receive simplified and personalized communication (Figure 7.11).

Figure 7.11: 3LOQ Labs with a smart financial management solution.



- ✍ [CloudNine AI](#) provides AI-driven (code-free) data management for healthcare organizations, relieving engineers of the burden of coding when dealing with interoperability issues. The startup was founded in 2019 and is focused on predictive analytics and cloud security. CloudNine AI offers a variety of Data Science services.



Figure 7.12: CloudNine AI with an intelligent financial management solution.

- ✍ [Razorthink](#) was founded as an emerging company that provides AI solutions. Individual investors provided funding for this project. The company aims to optimize AI computing, automation, dynamic models, and Data Science tools. Razorthink has a team of over 100 professionals who help small and large businesses optimize Artificial Intelligence-based customer service, risk management, fraud detection, and other services (figure 7.13).

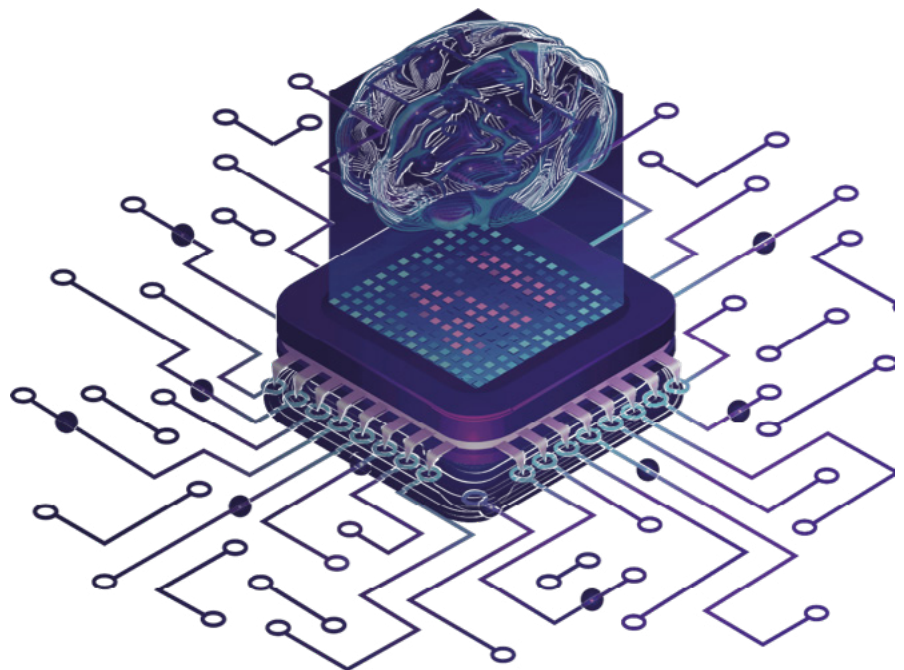


Figure 7.13: Razorthink with a forecasting solution to increase sales and profits.

- ✍ [SweetData.io](#) is a data marketplace where you can search, buy, sell, and download data sets for applications and machine learning. The company offers a simple way for people who have stored data to connect with other developers and industries (who need this data to create Artificial Intelligence applications and systems).



Figure 7.15: SweetData.io

- ✍ [Math Company](#) is a startup founded by data analytics experts as a collaborative venture. To support its customers, the company now employs over 50 people and is developing data analytics and machine learning, data visualization, design thinking, and domain knowledge.



Figure 7.16: The Math Company.

✎ [Zendrive](#) uses mobile sensor data to provide information that improves the safety of passengers and drivers worldwide. The global transportation industry is changing, and Zendrive is on the cutting edge of transportation automation and mobile service delivery.

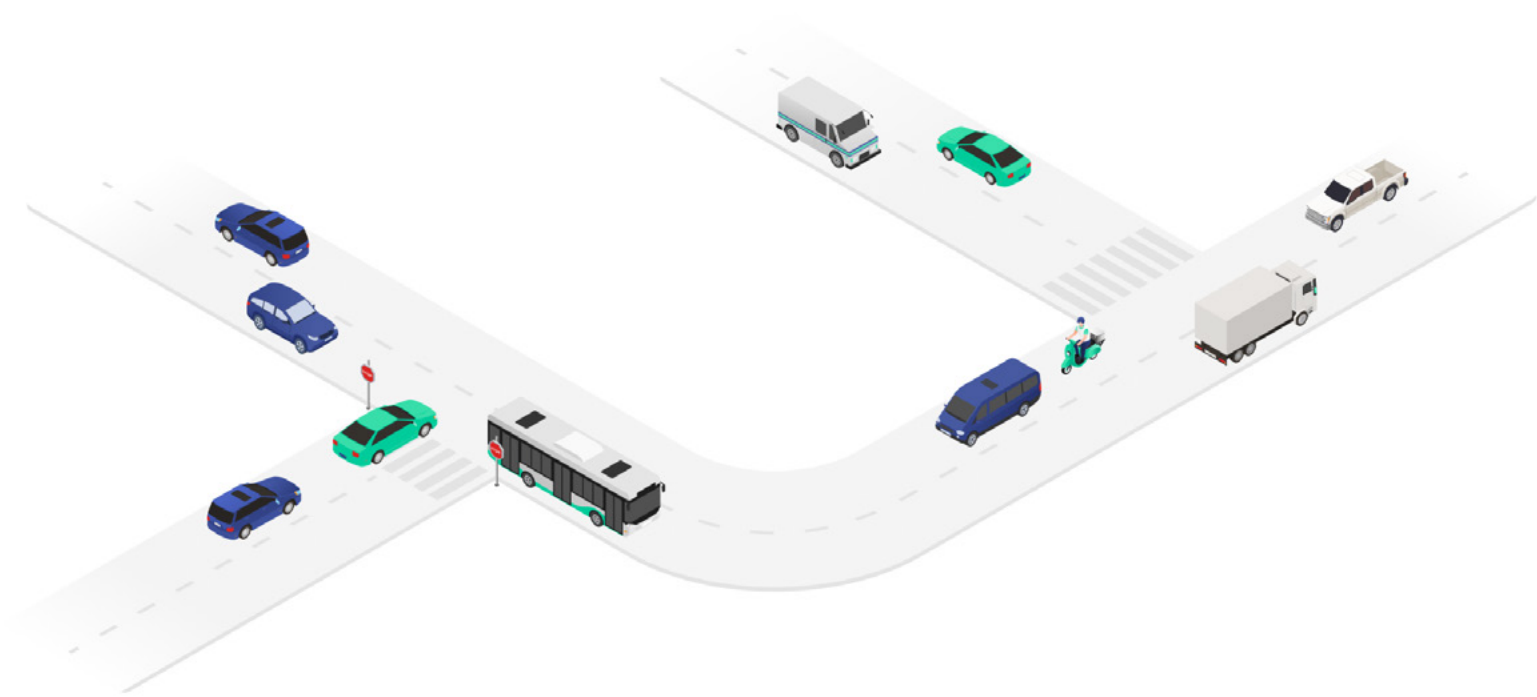


Figure 7.17: Risk mitigation with Zendrive's Intelligent Mobility Risk Platform.

✎ Other promising Data Science startups include [Avo](#), [Bigeye](#), [DataKitchen](#), [Deepnote](#), [Dhiva.ai](#) and [Hex Technologies](#) (Avidon, 2021).

7.3. Sources of Funding for Data Science Startups

✎ In addition to [IAPMEI](#), [Portugal Ventures](#), [Caixa Capital](#), [RAIZE](#) and [Previsional](#), which invest in startups in various ways and stages of development, [Faber Venture Capital](#) launched Faber Tech II to invest in startups that aim to accelerate the digital transformation of companies that use Data Science in its various aspects. This fund expects raising at least 30 million euros in funding (Computerworld, 2020).

This fund invests in early stage companies (pre-series A) that operate and lead areas of Data Science, with the main goal of **global digital transformation of companies and industries**.

Geographically, the fund invests in startups that use Portugal and Spain as a launching pad for the global market, exploring Europe's potential as a rising technology center.

✎ Faber has already made investments, including a €15.6 million Series A funding round by [SWORD Health](#) (a Portuguese startup that created the first digital physiotherapy system and brings together Data Science and clinical teams), the leadership of a pre-seed round by [YData](#) (a platform that helps data scientists improve data quality and ensure data privacy), and participation in a pre-seed round by Emotai (which developed a neuroscience-based pedicure). ✎

The fund continues to draw investors who want to bet on assets in a more specialized manner. Faber Tech II's main investors, in addition to several private (private and institutional) investors, are the ✎ [European Investment Fund](#) (EIF) - Europe's largest institutional investor in this asset class - and the ✎ [Development Finance Institution](#) (DFI) which have committed 15 million euros under the Portugal Tech program.

Glossary

- **Algorithm**

Any program (usually software) designed to be implemented on machines to perform finite processes, such as problem solving, by following a defined set of operations.

- **Predictive Analysis**

It involves predicting future outcomes by analyzing trends in current and historical data. It can use machine learning-supported algorithms to extract relevant information from data and future behavior patterns.

- **Data Analysis / Analytics**

It involves extracting, cleansing, transforming, modeling, and visualizing data in order to extract useful information for drawing conclusions and making decisions.

- **Machine Learning**

Artificial Intelligence subfield that implements algorithms that perform tasks and autonomously learn data-related information. It is applied to produce predictive models from data.

- **Data Science**

The inter-disciplinary study of extracting knowledge from massive amounts of data. It includes the analysis and preparation of data, as well as the presentation of results, for the purpose of informing high-level organizational decisions. It combines computer science, mathematics, statistics, and domain expertise.

- **Competitiveness**

Capability of an organization to accomplish its mission more effectively than other organizations. It is based on the organization's capacity to meet the needs and expectations of the customers or citizens it serves in its target market, in accordance with the mission for which it was established.

- **Knowledge**

Knowledge or familiarity with someone or something, such as facts, abilities, or objects that contribute to its comprehension. Acquired through a variety of methods and sources, including perception, logic, memory, scientific research, education, and experience.

- **Data**

Observations that are recorded as a result of measurements, collections, captures, or extractions. Typically, data must be processed before information and knowledge can be extracted. They describe entities on which algorithms can operate.

- **Data Mining**

Extraction of information, patterns, and trends from unprocessed data. Data mining is a technique used to investigate, analyze, and display data in order to discover useful information. It is a subfield of Data Analytics.

- **Strategy**

A collection of competitive moves and actions that a business employs to attract customers, compete successfully, enhance performance, and achieve organizational goals. It describes how the organization must operate to achieve its objectives.

- **Data Exploration**

Data analysts use data visualization and statistical techniques to describe characterizations of data sets, such as precision, quantity, and accuracy, in order to get a deeper understanding of the nature of the data.

- **Information**

Analyst-relevant data that has been processed, organized, and/or classified. Information is the collection of processed data that integrates the Knowledge Extraction process and upon which decisions and actions are based.

- **Interdisciplinarity**

A method or idea that merges concepts and consists of implementing an integration of traditionally considered singular and/or separate scientific fields and/or areas.

- **Jupyter Notebook**

A free, interactive, open-source application that enables the implementation of notepads with computational functions. It enables the combination of programming languages, text, multimedia resources, data visualization, computational operations, and machine learning (ML) algorithms within a single document.

- **Market**

A strategy used by a company to sell products or services to a particular group of customers. Usually, trade markets facilitate sales from one company to another when one company is planning to reuse or resell another company's products or services.

- **Niche**

A subset of a larger market with distinct needs and preferences. It is defined as a concentration of strategic efforts directed at a clearly defined group of target customers.

- **Python**

High-level, interpreted, object-oriented, high-level programming language with dynamic semantics, used for programming general application software. It is commonly used to develop websites, software, automate tasks, implement algorithms, and conduct data analysis.

- **Data Processing**

Transformation and manipulation of data to generate relevant, meaningful information. It is considered a sub-process of Knowledge Extraction and enables the transformation of raw data into content that can be used by machine learning algorithms.

- **Systema**

Physical and/or virtual platform able to execute stored instructions that interpret stored data and/or data read from an input, perform operations, and generate an output.

- **Task**

Action or collection of actions to be carried out by a machine using one or more algorithms. It is the execution of specific actions resulting from the application of models generated by machine learning-based algorithms.

- **Decision Making**

Method of gathering information and evaluating between two or more alternatives in order to achieve a desired result. It is also a process of reasoning that can be based on data, assumptions, preferences, and the decision maker's beliefs.

- **Data Visualization**

Visualization of information and data in a pictorial or graphical format (Example: digrams, charts and maps). The use of data visualization tools enables the identification and comprehension of trends, patterns, and outliers in data.

References

Analytiks. (2020). Why Data Visualization Is Important - Analytiks. Analytiks.Co. <https://analytiks.co/importance-of-data-visualization/>

Anand, A. (2021). 10 Data Science Applications in Real Life. AnalyticSteps. <https://www.analyticssteps.com/blogs/10-data-science-applications-real-life>

Avantika Monnappa. (2022). Why Data Science Matters And How It Powers Business Value? Simplilearn, 1–8. <https://www.simplilearn.com/why-and-how-data-science-matters-to-business-article>

Avidon, E. (2021). 8 analytics startups to watch over the next year. TechTarget. <https://www.techtarget.com/searchbusinessanalytics/feature/8-analytics-startups-to-watch-over-the-next-year>

Berkeley University. (2018). What is Data Science? | The Data Science Career Path. Berkeley School of Information. <https://ischoolonline.berkeley.edu/data-science/what-is-data-science/>

Computerworld. (2020). Faber investe 30 milhões de euros em startups ibéricas de "Data Science" | Computerworld. <https://www.computerworld.com.pt/2020/11/13/faber-investe-30-milhoes-de-euros-em-startups-ibericas-de-data-science/>

Das, S. R. (2016). Data Science: Theories, Models, Algorithms, and Analytics. In S. R. Das.

Data Flair Tutorials. (2021). Top 6 Data Science Use Cases that are Changing the World. Data Flair Tutorials. <https://data-flair.training/blogs/data-science-use-cases/>

DataFlair. (2021). 14 Most Used Data Science Tools for 2021 – Essential Data Science. <https://data-flair.training/blogs/data-science-tools/>

Dönmez, P. (2013). Introduction to Machine Learning. Natural Language Engineering, 19(2), 285–288. <https://doi.org/10.1017/s1351324912000290>

discoverdatascience, (2022) Data Science Career Guide Overview, <https://www.discoverdatascience.org/career-information/>

Edureka, (2022). Data Science Career Opportunities: Your Guide To Unlocking Top Data Scientist Jobs, <https://www.edureka.co/blog/data-science-career-opportunities-your-guide-to-top-data-scientist-jobs>

Grus, J. (Software engineer). (2019). Data Science from Scratch: First Principles with Python.

Henn, B. (2021). What is Data Exploration? Sisudata. <https://sisudata.com/blog/what-is-data-exploration-in-machine-learning>

Igual, L., & Seguí, S. (2020). Introduction to Data Science. <https://doi.org/10.4018/978-1-7998-3053-5.ch001>

Johnson, J. (2020). Predictive Analytics vs Machine Learning: What's The Difference? BMC Blogs. <https://www.bmc.com/blogs/machine-learning-vs-predictive-analytics/>

Mujthaba, G. M., Ameen, A. Al, Kolhar, M., & Rahmath, M. (2020). Data Science Techniques, Tools and Predictions. International Journal of Recent Technology and Engineering, 8(6), 5661–5668. <https://doi.org/10.35940/ijrte.f9887.038620>

Ohri, A. (2020). Importance Of Data Analytics in 2021 | Jigsaw Academy. Importance Of Data Analytics in 2021. <https://www.jigsawacademy.com/blogs/business-analytics/importance-of-data-analytics/>

Piringer, H. (2020). Benefits of Data Exploration. Visplore.Com. <https://visplore.com/benefits-of-data-exploration/>

Provost, F., & Fawcett, T. (2013). Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. O'Reilly Media, Inc.

Qian, L. (2021). What is a Data Scientist? March. <https://www.mastersindatascience.org/careers/data-scientist/>

Redman, T. C., & Davenport, T. H. (2021). 4 Ways to Democratize Data Science in Your Organization. 1–7. <https://hbr.org/2021/03/4-ways-to-democratize-data-science-in-your-organization>

Rice, M. (2019). 17 Data Science Applications & Examples You Should Know. Built-In. <https://builtin.com/data-science/data-science-applications-examples>

Sanyal, S. (2019). Top 10 Emerging Data Science Startups To Keep an Eye on 2022. Analytics Insight. <https://www.analyticsinsight.net/top-10-emerging-data-science-startups-to-keep-an-eye-on-2022/>

SAS. (2016). Predictive Analytics: What it is and why it matters | SAS. Predictive Analytics Overview, 1–12. https://www.sas.com/pt_pt/insights/analytics/predictive-analytics.html

Schutt, R., & O'Neil, C. (2013). Doing Data Science: Straight Talk.

Skiena, S. S. (2017). The Data Science Design Manual. In Springer. <http://link.springer.com/10.1007/978-3-319-55444-0>

Sundaresan, N. (2011). Data science and machine learning at scale. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6911 LNAI(November), 10. https://doi.org/10.1007/978-3-642-23780-5_9

Taylor, D. (2021). Information vs Knowledge: Key Differences. Guru99. <https://www.guru99.com/information-vs-knowledge-difference.html>

Tucci, L. (2016). What is information age? In TechTarget. <https://www.techtarget.com/searchcio/definition/Information-Age>

Wang, J. (2021). The Importance of Data Analytics in Today's Business World. IMC. <https://www.imc.edu.au/news-archive/the-importance-of-data-analytics-in-today-s-business-world>

Wieringa, J., Kannan, P. K., Ma, X., Reutterer, T., Risselada, H., & Skiera, B. (2021). Data analytics in a privacy-concerned world. Journal of Business Research, 122(July 2018), 915–925. <https://doi.org/10.1016/j.jbusres.2019.05.005>

Pictures

Photo by [Shubham Dhage](#) on [Unsplash](#)
Photo by [KOBU Agency](#) on [Unsplash](#)
Photo by [Nacional Cancer Institute](#) on [Unsplash](#)
Photo by [ThisisEngineering RAEng](#) on [Unsplash](#)
Photo by [National Cancer Institute](#) on [Unsplash](#)
Photo by [Alexander Sinn](#) on [Unsplash](#)
Photo by [Walkator](#) on [Unsplash](#)

Figures

[Figure 1.1: Phases of the Knowledge Extraction Process.](#)
[Figure 1.2: The basic concepts of data analysis](#)
[Figure 2.1: The domains and practice areas of Data Science.](#)
[Figure 2.2: Examples of economic sectors where Data Science methodologies are used.](#)
[Figure 2.3: Land invasion points in Ukraine. Infographics: Jaime Figueiredo](#)
[Figure 2.4: Processes of Machine Learning applied to data.](#)
[Figure 2.5: Data-Driven Decision Making \(DDDM\).](#)
[Figure 3.1: Data Science-Enabled Analysis](#)
[Figure 3.2: The 3Vs of data supporting business growth.](#)
[Figure 3.3: Data visualization software.](#)
[Figure 3.4: Different types of data value.](#)
[Figure 3.5: Technical skills of a data scientist.](#)
[Figure 4.1: Technological platforms for Data Science](#)
[Figure 4.2: RapidMiner features](#)
[Figure 4.3: Social media data collection and subsequent analysis using the Data Robot platform.](#)
[Figure 4.4: Manipulating data from various sources in Alteryx.](#)
[Figure 4.5: Features of the SAS platform.](#)
[Figure 4.6: Supported tools for Apache Spark integration.](#)
[Figure 4.7: A dashboard created with Microsoft Excel.](#)
[Figure 4.8: Example of a dashboard created using the Tableau platform.](#)
[Figure 4.9: An example of Jupyter application in the Python programming language.](#)
[Figure 5.1: Key data science trends for 2022.](#)
[Figure 6.1: The steps for intelligent data exploration.](#)
[Figure 6.2: Machine Learning and Predictive Analytics Comparison](#)
[Figure 6.3: GDPR is a European Union regulation that ensures the privacy of personal data..](#)
[Figure 7.1: Unbabel in online translation with data science.](#)
[Figure 7.2: Uniplaces platform.](#)
[Figure 7.3: Prodsmart real-time analysis](#)
[Figure 7.4: Feedzai in banking security.](#)
[Figure 7.5: Detech.ai on improving customer experience.](#)
[Figure 7.6: Gorodata sales analytics platform supported by data science.](#)
[Figure 7.7: CropIn uses data science in an agricultural SaaS model.](#)
[Figure 7.8: Firebolt with SaaS computing and data storage solutions.](#)
[Figure 7.9: Lymbyc with SaaS Data Analytics solution for business leaders.](#)
[Figure 7.10: InCountry offers a data analytics platform for companies as a SaaS solution.](#)
[Figure 7.11: 3LOQ Labs with a smart financial management solution.](#)
[Figure 7.12: CloudNine AI with an intelligent financial management solution.](#)
[Figure 7.13: Razorthink with a forecasting solution to increase sales and profits.](#)
[Figure 7.15: SweetData.io](#)
[Figure 7.15: The Math Company.](#)

[Figure 7.16: Risk mitigation with Zendrive's Intelligent Mobility Risk Platform.](#)

Tables

[Table 5.1: Examples of data science-based solutions.](#)

AUTHORS

António Raimundo and
Pedro Sebastião

**PROMOTING
ENTITY**

IAPMEI, Agência para a Competitividade e Inovação, I.P.
Departamento de Valorização e Capacitação Empresarial
Departamento de Empreendedorismo e Financiamento

**COORDINATION &
REVISION**

AUDAX – Centro de Inovação e Empreendedorismo do ISCTE-IUL
Augusto Albuquerque

GRAPHIC DESIGN

I AM - The Creative House

DATE OF ISSUE

October 2021

COPYRIGHT

2021, IAPMEI

PRODUCTION

audax _iscte



ISBN: 978-972-8191-69-6

Co-financed by



UNIÃO EUROPEIA

Fundo Social Europeu