Data Visualization in Data Science

[1]S M Nazmuz Sakib (sakibpedia@gmail.com)

S M Nazmuz Sakib is an eLearning expert and done more than 500 MOOCs or Massive Open Online Courses and experienced as an instructor in sites like Udemy. He has completed his BSc in Business Studies from School of Business And Trade, Switzerland with CGPA 4 in the scale of 4 and 97.06% grade marks on an average. He is also a certified Google IT Support Professional, Google Data Analytics Professional and IBM Customer Engagement Specialist Professional.

Abstract

Many of the important theoretical and practical issues that need to be addressed when developing data visualizations have been covered in the chapters that precede them. I also reviewed and evaluated a variety of data visualization examples, along with common mistakes and helpful approaches. Based on what we've learned, developing an efficient and ethical way to visualize data can be challenging. This chapter covers a wide range of topics, including the future of data visualization and new data visualization technologies. Finally, I have come to the conclusion of our study on data visualization. In the eleven chapters preceding this one, I have reviewed some of the most important theoretical and practical components of visualization methods and applications, and their applications. As I have seen, developing an efficient and effective approach to building a data visualization application is challenging. This procedure represents the relevant data. The data is then analyzed to derive the most important information about the problem, then transformed into a visualization, re-represented and finally integrated into a user-friendly solution. Consider a fantasy, scary image or something in between. The ecological plausibility of any form of representation facilitates its effectiveness: "On the one hand, the chart type and aesthetic components must be adapted to the data presented. display; otherwise, they must be visually appealing. However, the message you want to convey, as well as the audience's ability to perceive and understand the image, must be considered. Visualization is an art and a talent in its own right. R comes with a lot of tools, but learning them is mostly a matter of practice.

Introduction

For example, a map or graph may be used to visualize data in order to improve knowledge and insights of the human brain. It is the main goal of data presentation to make it easier for users to see patterns, trends, and outliers in large data sets via the use of graphics. For example, graphs, displays of information, and statistical graphs are often used with this word. There must be a way to display the findings once the data has been collected, analysed and modelled (Wenjun et al., 2008). Data visualisation is also a subset of the broader data processing and analysis (DPA) field,

[1]Graduate of BSc in Business Studies, School of Business And Trade; Pilatusstrasse 6003, 6003 Luzern, Switzerland
[1]Student of BSc in Civil Engineering, Faculty of Science and Engineering, Sonargaon University; 147/I, Green Road, Panthapath, Dhaka
[1]Student of LLB(Hon's), Faculty of Law, Dhaka International University; House # 4, Road # 1, Block - F, Dhaka 1213
[1]Student of BSc in Physiotherapy, Faculty of Medicine, University of Dhaka; Nilkhet Rd, Dhaka 1000

which aims to find, collect, analyse, prepare, and disseminate data as effectively as possible (Eckert et al., 2021).

Data visualisation is now essential in almost every field of endeavour. The use of teachers to display student exam results may benefit computer scientists researching artificial intelligence advancements, as well as managers seeking to convey information to stakeholders. It is also critical in large-scale data projects of any kind. Companies gathered enormous data collections throughout the early years of the big data movement, requiring the development of a technique for quickly and easily obtaining an overview of their data. The use of visualisation tools has shown to be a natural fit (Vila et al., 2018).

In advanced analytics, visualisation is important for the same reasons as previously mentioned. If a data scientist develops sophisticated predictive analyses or algorithms for machine learning, it is important to present the findings to ensure that the algorithms are properly monitored and ensured. This is because visual representations of complicated algorithms are frequently simpler to understand than numerical findings. Data visualisation is a fast and efficient method to transmit information visually to many individuals worldwide. It may also assist companies to identify which variables influence their client behaviour, highlight areas that need to be improved or cautioned, make data more remembered for players, decide when and where certain goods are placed and anticipate sales volumes (Tadse et al., 2021).

Literature Review

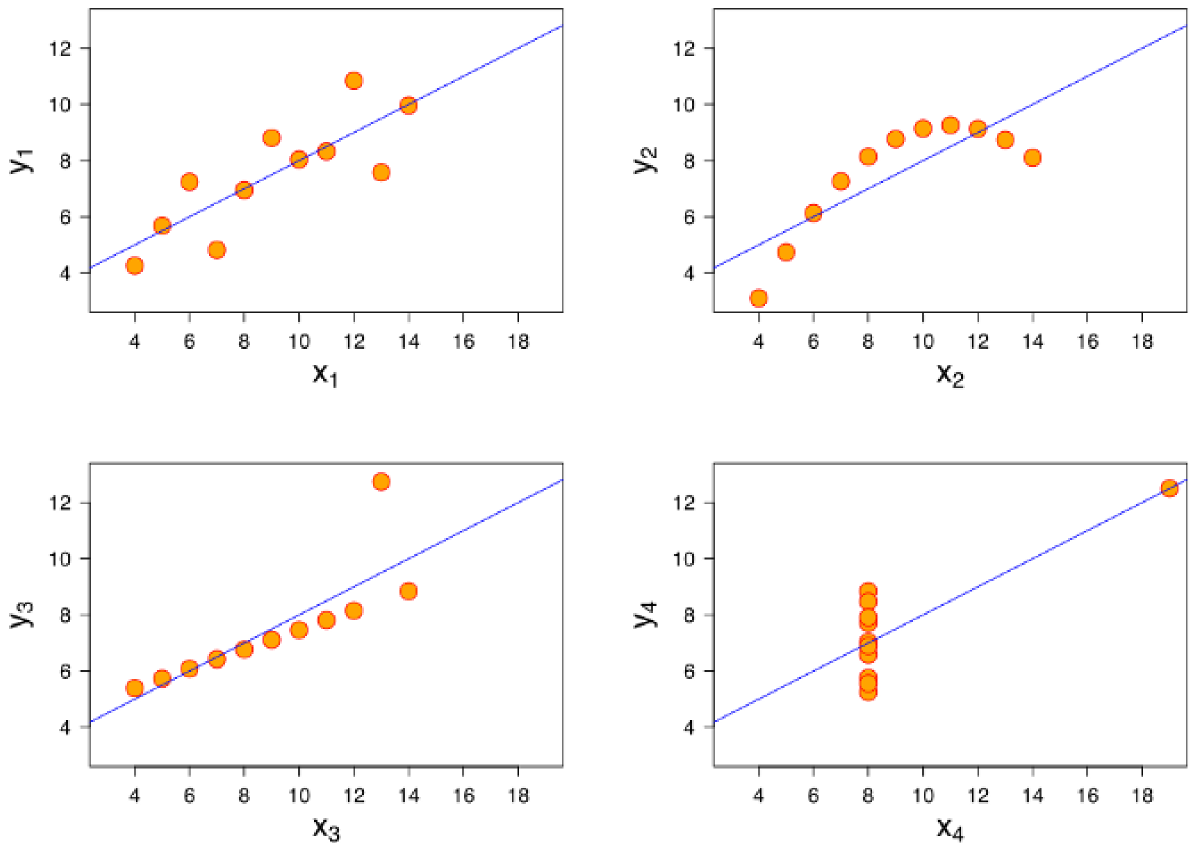
Ability to quickly absorb information, increase insight and make faster decisions, improve understanding of future actions to improve the organisation, and enhance the ability to maintain public interest through understanding of information; the ability to improve information; the simple dissemination of information which increases the ability to share ideas with all concerned; Capacity to improve Data scientists should be removed since data are more accessible and comprehensible than ever, and the ability to respond fast to findings and succeed more quickly, with less errors, should be enhanced (Starková 2020).

Because Big Data initiatives are becoming popular and data analyses are being analysed, visualisation is more essential than ever. Businesses depend more and more on machine learning to collect huge quantities of data that may be difficult to organise, analyse and explain. Visualization is a technology to speed up the process and provide information that company owners and stakeholders comprehend (Skaggs, 2017).

In many cases, high-quality data visualisation extends beyond traditional data visualisation techniques such as graphs, histograms, and organisational diagrams. Rather, it makes use of more complex visualisations such as heat maps and fever charts. For the collection, analysis, and translation of raw data into graphical representations that humans can utilise for quick understanding, sophisticated computer systems are needed. While big data visualisation offers advantages, it also has disadvantages for businesses. The following are the names of the individuals. To make the most of big data visualisation technologies, it is essential to hire a visualising specialist. This expert must be able to pick the most appropriate data sets and presentation styles to ensure that businesses get the most out of their data (Salazar-Cardona et al., 2018).

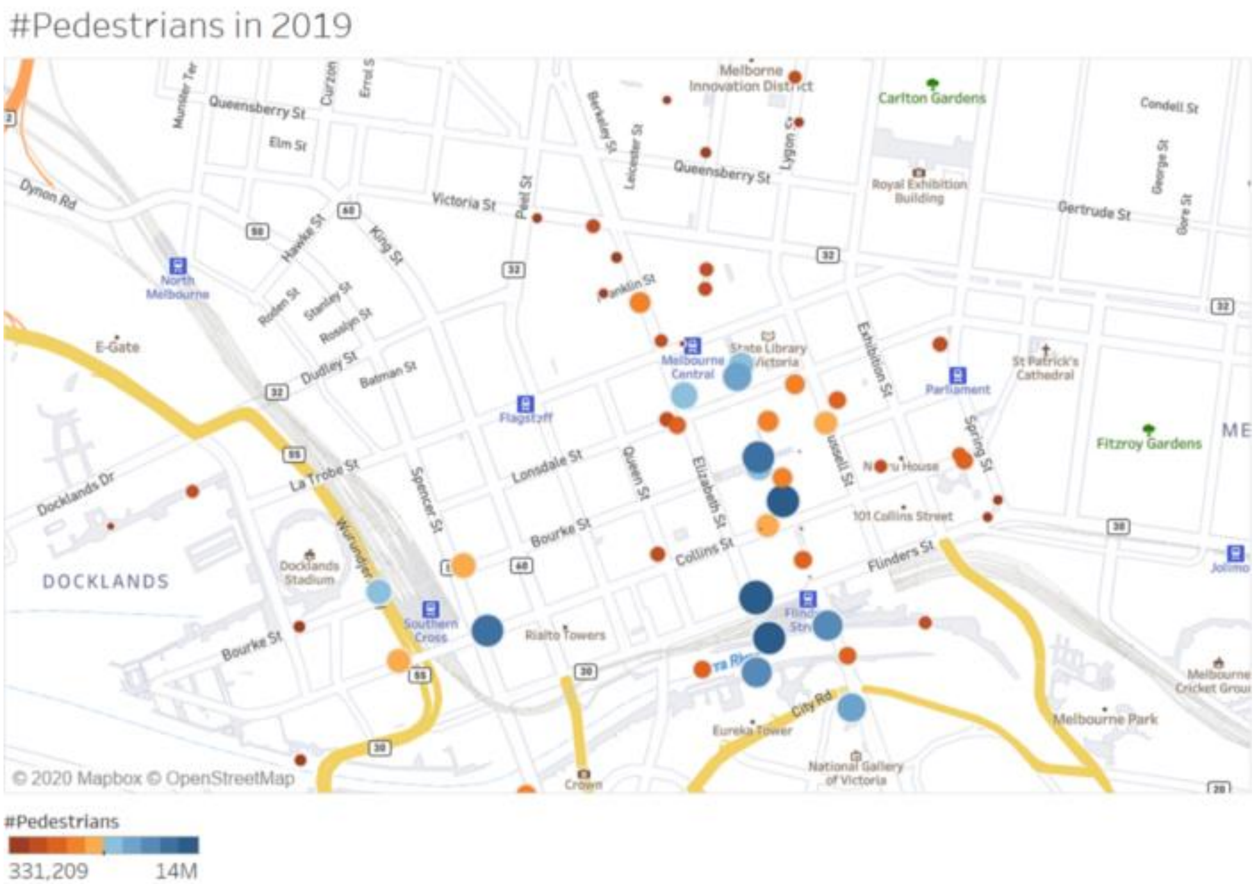
Because big data visualisation projects require the use of strong computer hardware, effective storage systems, and even a cloud migration, excellent data visualisation projects often necessitate the involvement and supervision of information technology professionals. The accuracy of big data visualisation insights is determined by the information that is displayed. As a result, it is necessary to create people and processes for monitoring and controlling the quality of corporate data, metadata, and data sources (Pérez et al., 2018).

Many departments use data visualisation tools to monitor their own progress. A marketing team, for example, might use the program to measure the effectiveness of an email campaign, tracking metrics such as open rate, click rate, and conversion rate, among others. Furthermore, as data visualisation providers improve the capabilities of their tools, they are becoming more popular as fronts for sophisticated Big Data sets. Depending on the conditions, data engineers and scientists may utilise data visualisation tools to monitor data sources and do basic exploratory study on data sets before to or after more sophisticated analysis. Many other businesses are well-known in the area of big data technology, including Microsoft, IBM, SAP, and SAS. Tableau, Qlik, and Tibco are among the most well-known firms that provide specialised large-data visualisation solutions (Pérez et al., 2018).

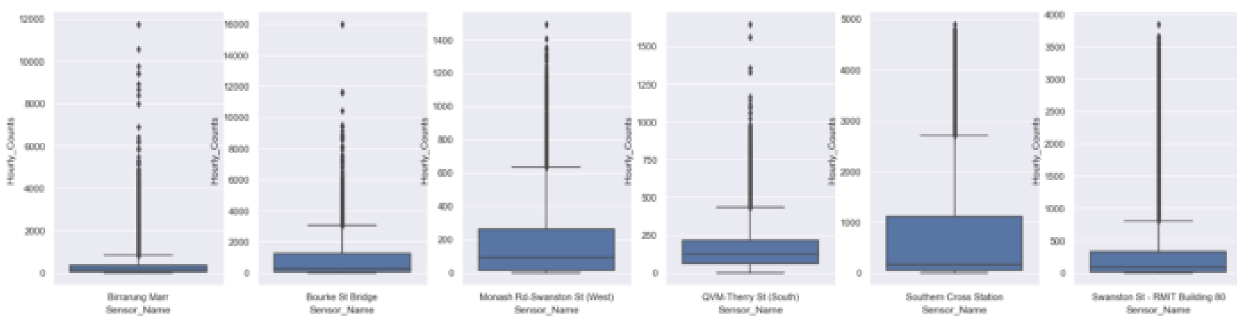


Data visualisation is called the technology of creating interactive visuals in order to analyse patterns, variations and derive valuable insights from data. Data visualisation is mainly used to verify and clean up data, to explore and to discover data and to communicate the results to business parties. Most data scientists overlook visuals and focus solely on numerical calculations, which may often be misleading (Pedroza et al., 2019).

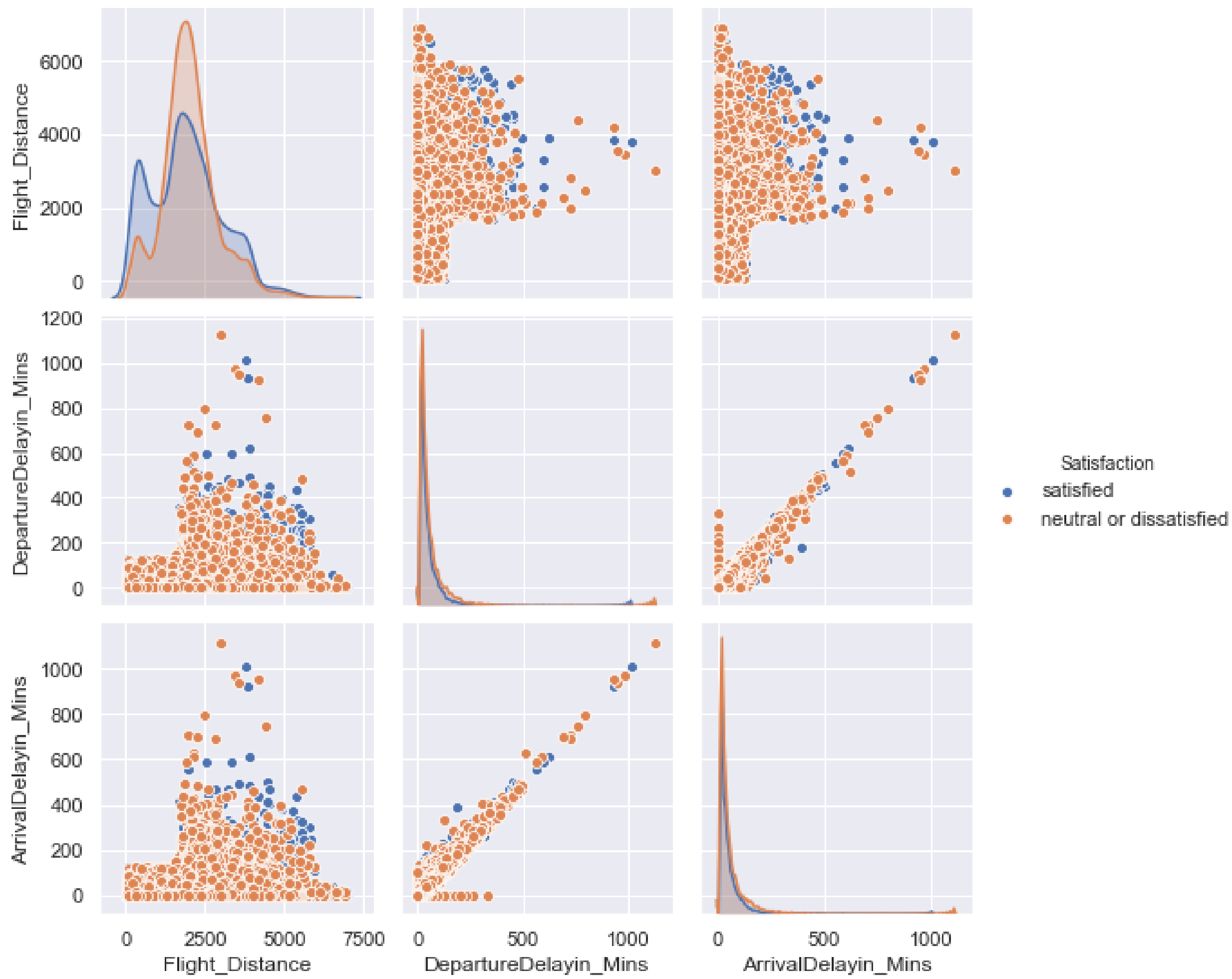
There are many businesses intelligence (BI) systems on the market today, each with its own set of benefits and drawbacks. Self-service dashboards were created to allow stakeholders with little or no knowledge of data science to work independently on data and derive specific conclusions that may help them make better business decisions on a daily basis. When dealing with evident data set issues such as zeros, random values, distinct records, date formats, regional data sensitivities, and string and character encoding, data visualisation may help pinpoint the underlying cause. It is critical to utilise data visualisation to comprehend data distribution, look for basic patterns (mean, median, and mode), detect outliers using a plot, monitor data skewing, and always understand the impact of data distribution fissurization. The assumption that the independent variables should be distributed, that there should be no relationship between the independent variables, that error terms should be homoscedastic, and a number of additional assumptions underlie linear regression and other classification models (Palacios et al., 2021). As a result, visualisations serve an essential role in verifying some of these assumptions. Individuals are often recruited by data scientists who work on loop analyses to examine the data, create a hypothesis, do enough analytics to validate the theory, and repeat the process until sufficient evidence is discovered. One such technique is the pair plot method, which is included in the popular Seaborn Python package. A link between the dependent and independent variables is very helpful in pair graphs. Using the visualisation, I hope to get a better understanding of whether any of the independent factors have an impact on the model results. Before deciding on a visualisation, it's essential to understand the many types of datasets that may be used. When dealing with tabular data, for example, a combination of bar and line charts may be helpful, while when dealing with spatial data, a map with a density track may be sufficient to convey the result effectively. Every day, organisations collect an increasing amount of data. In order to make use of this massive quantity of data, an increasing number of employees must deal with data analytics and data science. Data visualisation may help with data analysis, structural understanding, and the discovery of new insights (Olaya et al., 2020).



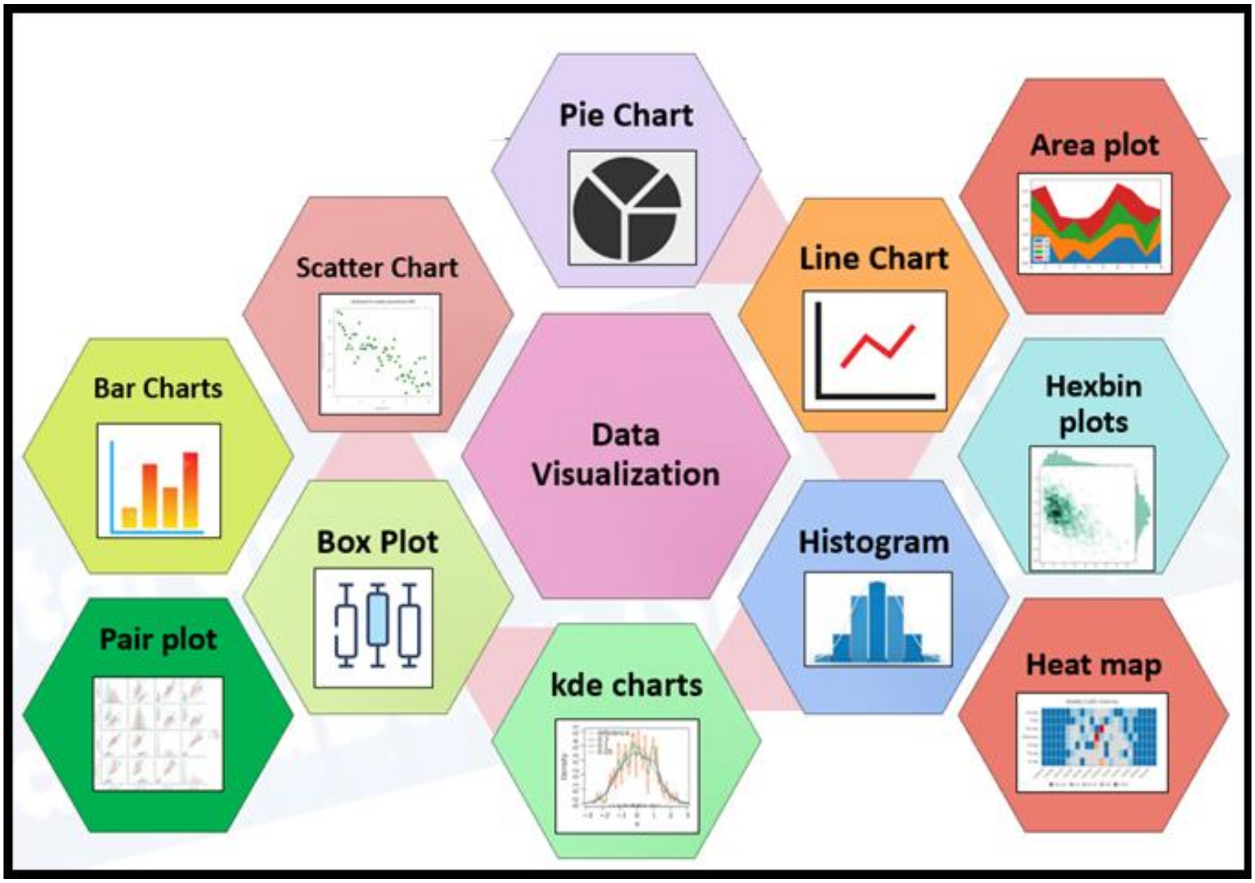
The current usage scenarios that customers face when interacting with data are driving forces behind the research into data visualisation and visualisation technologies. Without a doubt, the difficulties and duties that data scientists must complete are a potential source of future study in the area of data visualisation. Data scientists, on the other hand, are already using data visualisations to communicate their results on a daily basis, and this will continue. As a result, it is essential to explore how well-known visual analysis methods might be incorporated into contemporary data scientists' processes. It is generally accepted that data scientists' operations follow a similar workflow pattern, with various stages that may be identified along the way. When it comes to data processing, each step has its own set of challenges (Masini et al., 2020). For example, at the start of the workflow process, data wrangling is considered as an essential and time-consuming component. Data wrangling, for example, is a wide phrase that encompasses a variety of tasks such as data processing, cleaning, and combining. Data visualisation techniques may be used to identify data issues such as a lack of data, anomalies such as duplicates or outliers, and other inconsistencies. Following that, data scientists must analyse the data to evaluate its suitability for modelling. Understanding the data structure, locating correlations and clusters, and selecting modelling data components may all be aided by data visualisation methods. In recent years, the advancement of data science has given a large number of clients from a range of industries with data visualisation techniques. As a result, there has been an explosion of new data visualisation tools and frameworks (Masini et al., 2020). Furthermore, many of the libraries are open source and free, enabling them to be used in a variety of programming environments and languages. Because open-source software is widely used, data scientists may rely on a large community for guidance and support, as well as access to a wide range of libraries and plugins. A particular focus is placed on the regular expansion and maintenance of high-performance computing tools, numerical calculations, regression modelling, and visualisation, particularly for Python. However, in recent years, there has been a rise in the number of feature-rich, stand-alone applications for visual analysis that have been created (Malinverni et al., 2018).



Data science, as an interdisciplinary approach, includes input from a wide range of fields, including mathematics, statistics, computer science, and graphic design, among others. Many studies have been done in attempt to better understand the duties and needs of data scientists and to begin to categorise them due to the wide range of professions and skills accessible (Llerena et al., 2021).



Data visualisation techniques help us to detect patterns in business operations. Understand the statement and discover solutions with regard to design and use them to eliminate one or more of the underlying problems These techniques help us to detect market trends by collecting data on everyday business operations and generating trend reports, which helps to monitor the effect of the firm on the market. To understand our competitors and customers. This certainly helps to a long-term perspective. The knowledge of storytelling utilising available data is one of the most essential skills for business communication, especially in the field of data science (Lasaponara et al., 2010). This function may be greatly enhanced with the best visualisation and the objectives of business problems can be fulfilled. Businesses may get insight into KPIs, identify specific goals and strategic business planning and therefore optimize the data for present operations. The clarity of the KPIs indicating the productivity trends of the manufacturing unit and where the efficiency of the plant could be increased were enhanced using visualisation techniques. Data visualisation techniques are without doubt the most important component of data research. Visualization of data also plays an essential role in the field of data analysis. I will thoroughly use Python packages and how they assist with the flow of the data science process. This is an interesting topic for any data scientist (Kraak and sensing, 2003).

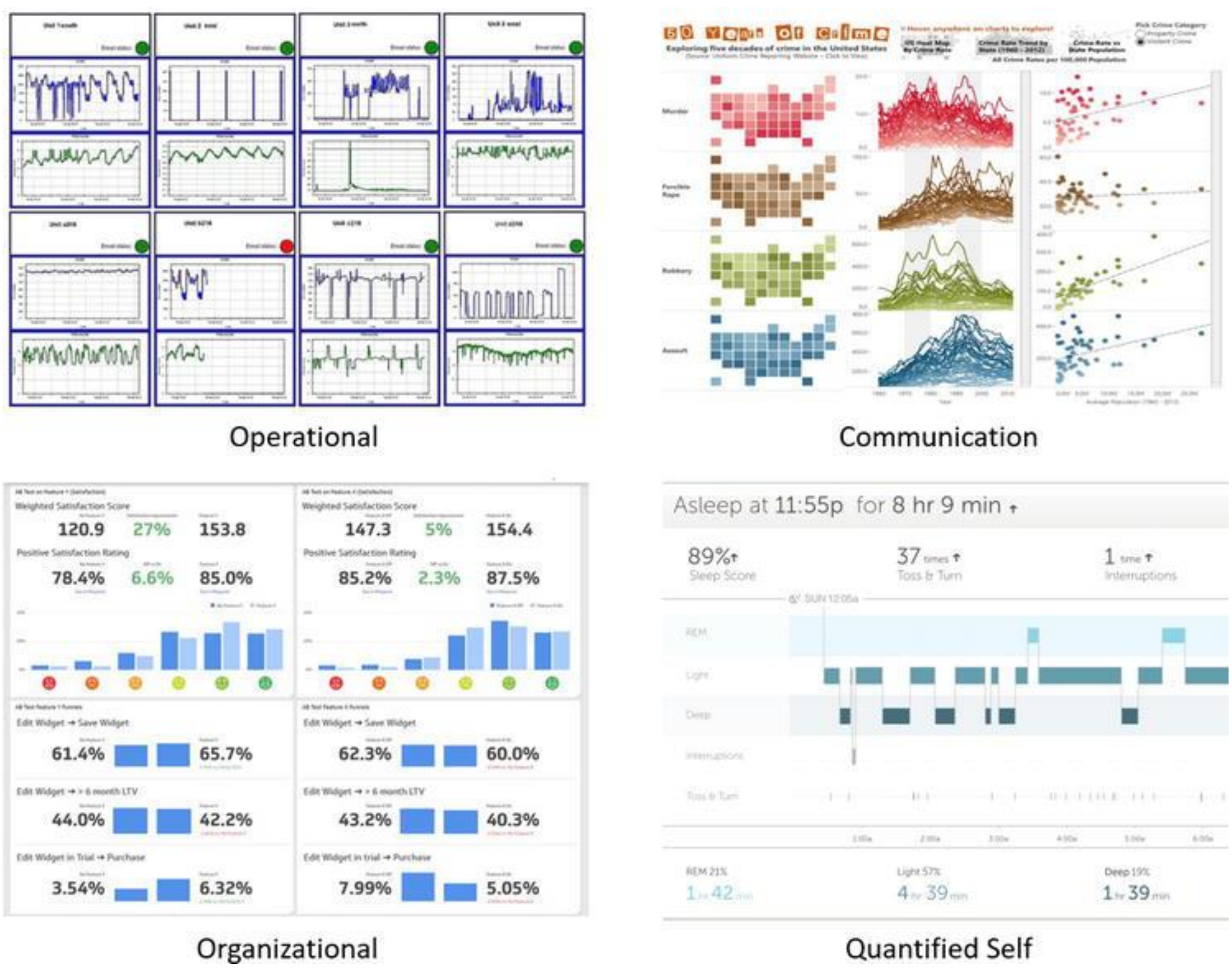


Data scientists make informed estimates about the kind of modifications needed for modelling. It also requires a knowledge of which data fields are most important in a particular analytical job. The data is utilised to construct the underlying phenomena models at this stage, which are subsequently used as input data for the model stage. While models are being created, the ability to test them against relevant real-world data is essential (Horák and Klír, 2017).

Simple models, such as regression models, are already supported by a variety of visualisation programmers. Furthermore, more sophisticated methods are being developed under the name in order to discover new, often visual ways for humans to explain the decision structure of AI models and to validate their choice based on their own fundamental understanding of what is true. When working with huge quantities of data, one issue that data scientists often raise is the scalability of model testing. Models are assessed in a way similar to that outlined in the Profile Stage, with the exception being the usage of EDA methods. According to projections, modern techniques for model output verification, especially with regard to input and output data, will be the focus of data visualisation research in the future. Because the findings of the data analysis stage must be conveyed to a larger audience at this level, visualisations at the report stage are often basic and straightforward. Simple charts, which are now accessible in the most current data science tools, may be enough for handling the bulk of use cases at this level (Holata et al., 2018).

In many instances, the outcomes are presented in the form of dashboards, with various degrees of interactivity. Dashboards are now accessible in a number of data science tools. This has only recently been acknowledged by the community of data visualisation practitioners. Dashboards are much more than a collection of graphs, and in order to be successful, they must be addressed as distinct study subjects in the context of data visualisation. They classified current dashboards into seven categories based on their study, the bulk of which were defined by the intended job (Gama et al., 2016).

Dashboards may differ in order to provide data results based on the intended user group and the activity being performed. This picture depicts dashboards that present your own (quantified) data for operational and strategic decision-making, communication, and research. The operational and organisational dashboards in the first column are designed for a small set of users with particular responsibilities inside the organisation. The dashboards in the second column (communication and quantified self) are intended for a broader audience than those in the first (Friendly and Statistics, 2002).



A critical component is selecting the most suitable visualisations for the data type under consideration. This is particularly essential when the exact structure of the data is unclear to the viewers and their previous knowledge of data visualisation is called into doubt. Academics started producing recommendations for data or task-oriented data visualisation concepts in the early 2000s, based on findings from human perception research and data visualisation possibilities. Draco uses predefined criteria to suggest several representations based on the data and to experiment with what the data is supposed to convey, all while keeping the data safe. Holtz and Healy explain several methods for visualising particular patterns on their website From Data to Viz, beginning with a single data type and continuing from there (Gama et al., 2016). The Data Visualization Catalog contains information on a range of visualisation techniques, as well as instructions on how to encode data using them. Overall, these approaches highlight the necessity for further data visualisation standards research in order to progress the field of data visualisation. As more people have started to engage in the area of data science in recent years, an increasing number of data analysis software programmers, many of which are open source, have been created. After completing all phases of the data science pipeline, data scientists must recharge their efforts and continue analytical activities. As a result of the extremely dynamic and undirected nature of the data science workflow, there are no applications that can cover the whole data science process. As a result, data scientists must continually use a variety of tool, script, and app combinations to achieve their goals. They are often focused on particular tasks such as data storage and access efficiency (as in Big Data applications), data wrangling, or automated analysis (e.g., machine learning). They may be created as fully functioning independent applications in a variety of computer languages (for example, Python, R, and JavaScript). This chapter will go over libraries and data visualisation tools in more detail (Friendly).

This notion is used when discussing libraries and data visualisation apps, among other things. To study the features of viewing libraries and applications, a specific chart was constructed using a variety of different instruments. The authors differentiate between visual libraries (also known as toolboxes) and apps in their study. Kandel et al. found that various kinds of data scientists like to utilise different technologies in their study. The typical hacker of a data scientist would be dissatisfied with the use of an autonomous programmed because he or she would be unable to access the most recent library in a scripting environment and therefore change the specific workflow. As a result, we'll stick with this distinction for the remainder of the chapter. Diagram libraries are any visualisation libraries that need the use of a certain programming environment to function properly (Filzwieser and Eichert, 2020). This is often a programming environment, as shown by the fact that many libraries are now based on Python or R. Because many data visualisation libraries are open source, they are often updated and improved. As a result, their popularity varies from year to year. Open-source data scientists have access to a large community of people who can provide advice and help, as well as a large number of libraries and plugins. This also implies that web-based visualisation methods are more likely to be presented or reported in a data science process than in the past. Furthermore, when addressing the client-server context for web-based displays, visualisation designers must carefully evaluate what data type to present, since huge data sets are unlikely to be transferred over the network and may create customer-side processing or rendering issues (e.g., smartphones). Most of the time, such a comprehensive design can only be done after the analysis is concluded (Filzweiser and Eichert, 2022).

Data visualisation, as explained in this article's "Brief Data Visualization History," is a quantitative information graphic representation. The way data is presented has the potential to turn apparently random facts into something comprehensible. Consider this data visualisation of the whole NFL's history, which takes a graph and turns it into a timeline for the entire sport. Data science candidates, whether online or in person, may be interested in understanding how visualisation art combines the tougher abilities of data analysis and organisation with a less technical skill set. Data may be shown in a variety of ways, including graphs. The lesson that follows will help you understand the basic ideas and methods for generating such visualisations, as well as the abilities required to do so (Filzweiser and Eichert, 2022). According to Friendly, data visualisations first emerged in geometric diagrams, tables of star locations, and the creation of maps to aid in navigation and exploration before the 17th century. Since then, data visualisation has advanced at a breakneck speed. Data scientists may be able to present data in previously unimaginable ways as a result of technological advancements. Interactive statistical computer systems, broad-based statistical and graphical software, and linear statistical modelling are only a few of the available technologies. In other words, "big data" refers to "bigger and more complicated data sets, particularly those coming from new sources... so enormous that conventional processing techniques are unable to handle them." The creation of large-scale and three-dimensional data representations may help data scientists better comprehend the results of their huge data studies, which could be challenging otherwise. There are many different kinds of data visualisation methods. On the other hand, some may be more useful than others when it comes to presenting various data information. It is important to do study before making a final choice to determine which approach will provide the best results for you (Espitia and Montilla, 2018).

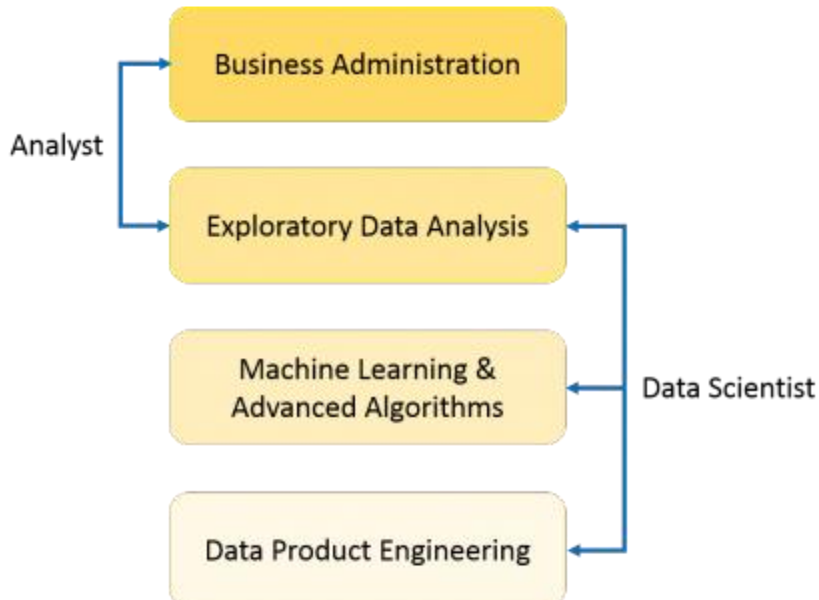
Although the term "big data" was very new, the roots of large sets were in the 1960s and 1970s when the world of information began to take shape when the first data centres were established and the first relationship database was developed. Around 2005, consumers started to understand how large the data produced by Facebook, YouTube and other Internet services was sold to other parties. Hadoop was launched the same year (an open-source framework especially designed to store and analyse large data volumes). During this time, NoSQL became more popular (Dehmer et al., 2020).

It is hard to envision a company that does not benefit from making data more understandable. In many areas of STEM, data understanding is important, including governance, finances, marketing, history, consumer goods, services, education, sports, etc. Although I often wax lyrical about data views (sometimes you are on the website of Tableau), practical uses of real-world views are obvious. In addition, since visualisations are so prevalent, they are considered to be one of the most important professionals to obtain (de Datos and Technology, 2021). The more efficient you utilise this information, the greater your capacity to transmit your thoughts visually in a dashboard or slide presentation. The concept of the citizen data scientist gains momentum. Ability changes in order to fulfil the requirements of a data-driven world. Increasingly, professionals are interested in two areas: utilising information for decision making and using graphics to explain the effects of data on who, what, when and where. The modern professional world honours persons who can connect the two professions using data visualisation via analyses and visual narratives, whereas traditional education frequently isolates creative storytelling from technical analysis. The data science field focuses on utilising advanced analytical and scientific procedures in order to get useful information from company decision-making, strategic planning and other applications. Businesses become more important: As a result of data science inspections, companies may improve operational efficiency while also discovering new business options and enhancing their marketing and sales strategy. It may lead to a competitive advantage over rivals in the long run (Chang et al., 2009).

You will not have to be restricted to your workstation in order to evaluate information. Your analytics may be accessible to you anywhere you go through your smartphone or tablet. Machine learning can identify the variables that drive your company's performance automatically, analyse data behaviour, and reveal hidden insights that may lead to better educated choices. You should be able to import and integrate data from a number of sources utilising preconfigured connections in your data visualisation tool, making it simple to mix various data sets and rapidly determining which data is really relevant. Furthermore, it should be built in such a manner that information can be accessible and shared throughout your organisation at any time and from any place. Many companies have an analytical ecosystem with multiple instruments: one for production reporting, another for management reporting, another for discovery, and so on. This may be costly, require the development of a wide range of skills, and result in instrument compatibility problems (Bravo et al., 2018).

In data visualisation, statistical visualisations, charts, information graphics, and other techniques are used to convey information simply and effectively. Numerical data may be represented graphically using points, lines, or bars to convey a quantitative statement visually. [6] When people utilise good visualisation, they can evaluate facts and evidence more effectively. This makes complex knowledge more accessible, comprehensible, and useful. The visual design concept (for example, displaying parallels or conveying causality) may be chosen based on the user's analytical goals (for example, drawing similarities or comprehending causation). When a user is searching for a particular measurement, data is usually presented in tables, while data is typically displayed in different kinds of diagrams when a user is looking for patterns or relationships between one or more unique variables in the data (Arevalo et al., 2019).

When I speak about data visualisation, I mean the technology used to communicate data or information by encoding it as graphics-based visual objects (e.g., points, lines or bars). The aim is to offer consumers with information that is clear and comprehensible. It is a stage in the gathering or analysis of data. Vitaly Friedman (2008) claims that the "The primary objective of data visualisation is to convey information in a clear and effective way via the use of graphic methods. This does not imply that the data display must be dull or too complex to be aesthetically attractive. In order to successfully communicate ideas, both aesthetic form and utility must collaborate, giving insight into a very restricted and complicated collection of data via a more natural transfer of its essential features. Designers, on the other hand, often fail to find a balance between form and function, resulting in visually appealing data visualisations that fall short of their primary objective of conveying information." A clerical mistake occurred. According to Fernanda Viejas and Martin M. Wattenberg, an ideal visualisation should not only effectively convey information, but should also evoke involvement and attention from audience members. To mention a few subjects, data visualisation is strongly related to graphs, information presentation, scientific visualisation, data analysis, and statistical graphics. Data visualisation has developed as an active field of study, education, and development in the twenty-first century. According to Post et al. (2002), it has combined scientific and information visualisation. Dashboards are a word used in the business sector to refer to data visualisation. Infographics are yet another popular way of graphically presenting information. It's difficult to envision a professional sector that doesn't seek to make data easier to comprehend for its clients. Data expertise is useful in all fields of STEM, as well as government, finance, marketing, history, consumer products, services, education, sports, and many other sectors. Despite the fact that I are always waxing poetic about data visualisation (after all, you can learn more about it on the Tableau website), there are practical uses that cannot be overlooked in real-world circumstances. Furthermore, since visualising is so common, it is one of the most useful skills to have. The more effectively you can visually convey your ideas, whether via a dashboard or a slide deck, the more successful you will be in utilising that knowledge. The idea of a citizen data scientist is gaining traction. To meet the needs of a data-driven world, skill sets are evolving (Arevalo et al., 2019). When information tells them who, what, when, where, and how, professionals are getting more proficient at utilising data to drive decision-making and using visuals to convey stories. However, although conventional training usually differentiates between creative storytelling and technical analysis, today's professional world rewards people who can bridge the gap between the two: data visualisation sits exactly in the middle of analysis and global visual history. According to the authors, as the "age of big data" advances, visualisation is becoming an increasingly important tool for understanding the billions of rows of data produced every day. By selecting data in such a manner that patterns and outliers are simpler to comprehend, data visualisation helps in the telling of stories. A good perspective, in addition to creating a story, removes noise from the data and brings attention to important information. However, it is not as simple as just dressing up a graph to make it seem nicer or putting the "data" portion of a computer image on top of another. A careful balance of design and function results in the most effective data display. The most visually attractive chart may be too dull to draw attention or tell a compelling story the most visually pleasing visualisation may be completely useless in delivering the appropriate message or fail to communicate successfully at all (Arango-López, 2018). Blending excellent analysis with captivating narrative is an art; both facts and images must work in tandem to be successful. Data visualisation is the visual presentation of graphical information and data. Data visualisation tools, which use visual elements like as charts, graphs, and maps, provide readily comprehensible representations of trends, contours, and patterns in data sets. In the era of big data, massive quantities of data and the capacity to visualize that data are essential. Data visualisation tools and technologies are critical for analyzing huge quantities of data and making choices based on that data (Aguilar et al., 2018).



Data analysts usually explain what occurs when data history is handled, as seen in the figure above. A Data Scientist, on the other hand, not only does exploratory research but also employs advanced machine learning techniques to predict the likelihood of a certain occurrence recurring in the future. A data scientist examines data from a variety of viewpoints, sometimes from previously unconsidered angles. As a result, in data science, predictive causal analytics, predictive analysis, and machine learning are often used to make decisions and predictions. You must apply predictive causal analysis to create a model that can anticipate the likelihood of a certain event occurring in the future. Consider the following scenario: you offer credit and are worried about your customers' capacity to make future credit payments on schedule. This section may be used to build a model that predicts whether or not future payments will be received on time based on a customer's payment history. Prescriptive analytics is required if you want a model that can make decisions on its own and modify them in real time as circumstances change. This relatively new profession is mostly focused with advising. In other words, it not only forecasts, but also suggests a broad range of necessary actions and their related consequences (Arango-López, 2018).

The most compelling example is the self-driving Google car, which I already mentioned. In the future, data gathered by cars may be used to educate automobiles. This information may be utilised to run algorithms that give intelligence to the user. As a result of this capability, you may select when to turn, slow down, or accelerate. If you have transactional data from a financial company and need to build a model to forecast future trends, machine learning algorithms are the most effective choice available. This is a component of the supervised learning paradigm, which was previously described. Monitoring is appropriate in this case since you already have the data on which your computers can be educated. A model for identifying fraud may be developed by looking at a historical record of fraudulent transactions, for example. Examine how the percentages of the techniques listed above vary to see how data analysis and data science are related. Data analysis, as shown in the picture below, incorporates both descriptive analytics and prediction to some degree. Data science, in contrast to other fields of study, places a greater emphasis on predictive causal analysis and machine learning techniques. There are a variety of definitions for data scientists to choose from. Simply said, a data scientist is someone who carries out the art of data science. The term "data scientist" was coined in recognition of the fact that a data scientist draws on a wide range of scientific knowledge and applications, including statistics and mathematics, to do his or her work. A common mistake in data science efforts is the collection and analysis of data without first thoroughly comprehending the needs or even properly identifying the business problem under consideration. As a result, it is very critical that you adhere to all phases of the data science life cycle in order to ensure that the project runs successfully (Arango-López, 2018).

Methodology

One of the most significant discoveries was that the variety and variance rendered reusing the materials impractical. In the first stage, data scientists are often on the search for appropriate data sets, which they may find in databases or on the internet, or they can ask their colleagues for suggestions. Finding and comprehending critical information is often seen as a major bottleneck in the work process, particularly in big organisations, and this is especially true due to access constraints. If a suitable format is available, it is essential to convert the datasets. Data conflict includes the processing of files, the administration of data layouts, and the integration of a large number of disparate data sources. There is a data discrepancy. Because data struggle is a time-consuming and mostly manual job, it occupies the bulk of the time spent on data analytics. After the data has been made accessible in the format of choice, its quality and appropriateness for analysis should be checked (Arevalo et al., 2019). A "dataset failure" occurs when data sets have major defects such as missing information, outliers, erroneous values, and other problems. Understanding data structure, on the other hand, is a critical job in data science. Finally, using data sets as training sets for prediction models is an important part of the data science process that should not be ignored. The models must be created and evaluated during this phase to verify that they operate as anticipated when compared to existing real-world data. Data scientists are in charge of determining which data will be utilised in their current research. This entails looking for information from both internal and external sources. At this time, the most important difficulties are the limited availability of data and the lack of documentation on data characteristics. This kind of capability is uncommon in data visualisation tools. Data visualisation methods, for example, are available to enhance the display of search engine results. During the Wrangle phase, the data war consists of focusing on data defects such as duplication and inconsistencies (for example, name) on one hand, and the process of profiling and converting data sets on the other. The main aim of data fighting is to make the data useful at later stages (Chang et al., 2009).

Initially, data visualisation did not take data conflict into account; instead, it began to work only after the data was made accessible in the appropriate format. Because data confrontation has become a critical and difficult job in data science, and because it consumes a considerable amount of time throughout the workflow (up to 50-80 hours), academics have begun to investigate methods to assist in this process. Data visualisation Wrangler, an interactive data transformation system, is the most well-known application in this category. Data scientists may investigate potential actions after being provided with information regarding data changes. Trifecta, a firm that specializes in data pipeline development, embraced the Wrangler concept and included it into their product (de Datos and Technology, 2021).

To summaries, data conflict is an intriguing example that may get greater attention in the future as a consequence of data visualisation research. Currently, the bulk of data scientists must depend on scripting tools and manual procedures to convert their data into the appropriate format for their needs. The most difficult step is the visualisation phase, which needs data scientists to analyse the data in order to comprehend its structure. This is a very circular and undirected process with no clear aim. A fundamental knowledge of the relevant issue area is required. The aim is to recognize and understand the patterns in the data. This encompasses, but is not limited to, the distribution of values, linkages, outlines, and clusters, among other things (Eckert and Britos, 2019).

Datasets often include a range of quality problems, such as missing data, outliers or high values, and inconsistencies, to name a few. In many instances, the loss of data is caused by a total lack of observations inside a data collection, which may be recognized by empty cells or null values in the data collection in question. Integers indicating missing information (for example, 0 or 1), as well as characters are also conceivable and must be considered throughout the analysis. Both 0 and 1 are instances of missing data. Humans often contribute inconsistencies and heterogeneous information into the system, particularly in the case of names and conditions, and because information was overloaded in certain parts of the system. When dealing with a specific dataset, data scientists must be aware of any faults in the data collecting process. To guarantee that a dataset is of good quality, several visualisation methods have been employed in the past. Profiler's aim was to allow users to assess the visual resemblance of datasets for time series data in order to establish their quality. Quality measures for multidimensional datasets have been established and are extensively used in general. Quality controls may now be accessed as distinct visualisation components that can be utilised independently. When used with Python, the module automatically provides quality controls (Eckert et al., 2021).

Discussion & Conclusion

Open-source framework frameworks like as Hadoop (and more recently Spark) have proven essential to big data development, simplifying and reducing the costs of handling and storing huge volumes of data. Since then, the amount of big data has increased. Users continue to generate huge quantities of data, not only from themselves. Due to the Internet of Things (IoT), more products and gadgets are linked to the internet to collect data on patterns of consumer usage and product performance. Machine learning progress led to a huge flow of data. While big data is still in its early phases, it has only recently started practical application. The introduction of cloud computing has further increased the potential of big data. Due to the very elastic scalability of the cloud, developers may build ad hoc clusters rapidly and inexpensively to test a portion of data. Graphical databases also become popular because to the capacity to show large quantities of data so that they may be analysed quickly and fully. Over the past several decades, visualisation researchers have had great success, developing a plethora of new methods for the visual presentation of data (Pérez et al., 2018). Methods of data representation (e.g., parallel coordinates) and suggested methods for user assistance and involvement, among other things, are among these approaches (e.g., overview-first, details-on-demand). Current surveys show that there is a wide range of presentation methods. In a previous evaluation of McNabb and Laramie information visualisation surveys over 80 survey articles detailing pertinent state-of-the-art methods were categorized, and a more recent analysis of information visualisation books showed a comparable number and diversity. Unfortunately, at present moment, the most recent advances in visualisation research are only compatible with data visualisation capabilities of charting libraries and applications. More complicated visualisation techniques (such as scatter plots and bar charts) can be found in the majority of tools and apps, while more sophisticated visualisation methods cannot be found in the majority of tools and apps (e.g. chord charts, horizon graphs) (Skaggs, 2017).

Many research on the integration of standard libraries and the use of visualisation methods have shown the validity of this. Herger and Crosson carried research to see how feature-rich open-source visual analysis software is. They assessed the research toolkits based on the basic graph kinds (for example, bar graphs and line graphs), graph visualisation types (for example, circular or force-directed layouts), and geo-spatial visualisation techniques they had access to (e.g. Choropleth maps, cartograms). Certain toolkits are more analytically focused, while others are more aesthetically oriented, according to their results performed a research similar to this one in which he looked at how people utilise tools and apps and rated their visualisation methods. The emphasis of this research was on the visualisation methods themselves, rather than the features that were obtained from them (such as feature richness). It also took into account recent advances in visual research, open-source tools, and commercial applications to provide a complete picture of the usage and efficacy of visualisation techniques. Apart from that, they focused on two-dimensional information visualisation methods since they are more important for data science and analysis, whereas spatial techniques such as three-dimensional volume rendering were overlooked (Wenjun et al., 2008).

It should come as no surprise that the most basic kinds of graphs, such as scatter plots and bar charts, have gotten substantial support in all of the studies thus far by all of the tools and applications investigated. A number of multi-dimensional methods, such as parallel coordinates and radar charts, are already extensively used and well-known from sophisticated visualisation techniques, and therefore appear in a broad range of products. The same may be true for heat maps and dispersion plot matrices, among other things. Hierarchical data techniques, particularly open-source tools, are very popular and well-supported. In the overwhelming majority of tools and mobile apps, there are no visualisation methods for temporal data. This is because temporal data (for example, time series data) is a specialised kind of data that is only utilised for certain purposes. Users are therefore more inclined to utilise their own resources as a consequence. Temporary data methods have not been incorporated in popular tools and apps since they are intended to appeal to a wider variety of data scientists and analysts. Despite their popularity, certain visualisation methods such as time networks, data vases, and people gardens have yet to be incorporated in a tool or application (Masini et al., 2020).

Polly and D3 have the most functionalities of any open-source tools that have been looked at from the perspective of tools and applications. Several additional tools, such as scientific plot digraphs, aim for outstanding qualities while only providing a limited set of visualisation methods. Chart.js and Google Charts, for example, are additional libraries designed for usage in web-based apps, but they don't provide any methods to display web-based applications. Open-source tools, particularly ggplot2, are very helpful since they enable the development of a variety of sophisticated visualisation methods that are only accessible via extensions. Tableau, Microsoft Power BI, and High charts all include the bulk of the visualisation methods that have been covered here when it comes to business tools (Olaya et al., 2020).

Data scientists may benefit from the usage of visual tools at any step of their process. It's worth noting that visualisation methods, which were formerly utilised only at the conclusion of the data science process, are now mostly used at the reporting stage. Visualisation research, interestingly, significantly promotes interactive data scanning methods, which is in direct contradiction to reality. In addition, there is currently limited support for sophisticated visualisation methods, especially for interactive data exploitation, which is a significant restriction. The "Interactive Visualization Gap," as defined by Batch and Alquist is a "gap in interactive visualisation." A desirable and essential objective is to increase the amount of interaction between the visualisation community and data scientists. According to previous data science study, there is still a considerable gap between new advances in visualisation research and their usage "in the field," and hopefully in the future. Data visualisation methods may be beneficial to those who deal with data on a regular basis. Data visualisation methods have been proven to be useful at many stages of the data analysis process. The methods vary in terms of how they interact and how complicated they are. A variety of visualisation methods have been effectively incorporated into libraries and data visualisation applications. A flurry of new possibilities has emerged in recent years, particularly in the open-source community. Libraries targeted towards programming environments like Python, R, and Scala have grown in popularity as a result of the extensive use of these languages. Two examples of such tools are Polly for Python and ggplot2 for R. Web-based applications are likewise growing more popular. This is why the JavaScript programming language is used in libraries like D3 and Chart. JavaScript has some of the most commonly used database visualisation libraries. Commercial intelligence solutions are likewise a fast-growing industry, although there are three major players: Tableau, Microsoft Power BI, and Qlik, all of which are headquartered in the United States. A wide range of libraries and tools are required by data scientists of different colors (Dehmer et al., 2020).

As a consequence, applications are more narrowly focused on a single objective and intended to do particular activities. In contrast, the Interactive Visualization Gap for exploratory data analysis still exists when looking at the data visualisation methods employed by the most popular libraries and apps. Despite their significance, many new findings and implementations in data visualisation research do not make their way into existing libraries and applications. As a result, continuing to share data science and visualisation is highly encouraged, as both sides may learn from one another's experiences and collaborate to improve the use of data visualisation for data analytics. Furthermore, many organisations, which may include professional business intelligence (BI) specialists, business analysts, data-savvy business users, data engineers, and other workers who lack formal data science expertise, are becoming more dependent on citizen data scientists, to varying degrees (Kraak and sensing, 2003).

This thorough data science book delves further into why data science is so essential to companies, how it works, the economic advantages it offers, and the difficulties it presents. An introduction to data science applications, tools, and methods is available, as well as information on the activities done by data scientists and the skills needed by data scientists. The guide contains links to relevant TechTarget articles that examine the subjects covered in this guide and provide insight and expert guidance on data science initiatives, as well as additional resources (Espitia and Montilla, 2018).

Data science, which is becoming more essential, influences almost all areas of company operations and strategy. For example, it gives businesses with consumer information, allowing them to create more successful marketing campaigns and targeted advertising to boost product sales. It helps with the control of financial risks, the detection of fraudulent transactions, and the prevention of equipment failures in manufacturing facilities and other industries. It aids in the prevention of cyber assaults on information technology systems as well as other security concerns (Pérez et al., 2018).

Data science efforts, from an operational perspective, may assist improve the management of supply chains, product inventories, distribution networks, and customer service by detecting patterns and trends. They lay out the path to better efficiency and lower costs on a more basic level.

Furthermore, data science allows businesses to develop business plans and goals based on an in-depth examination of consumer behaviour, industry trends, and competition. Companies that do not have it may lose out on opportunities and make bad choices (Starková, 2020).

Data science is equally essential in domains other than regular business operations. Some of the uses of artificial intelligence in healthcare include medical diagnosis, image analysis, treatment planning, and medical research. Data science is used by academic institutions to assess student performance and enhance the marketing of future students to potential students. Sports teams utilise data science to assess their players' performance and create game tactics. The service is also heavily used by government agencies and public policy organisations. When Jessica Steth, Data Science Manager at Fidelity Investments, spoke during an October 2020 webinar hosted by the Institute for Applied Computational Science at Harvard University, she claimed the Institute for Applied Computational Science demonstrated the "very obvious connection" between data science effort and business outcomes. This study highlighted possible business benefits such as improved ROI, revenue growth, more effective operations, quicker time to market, and increased customer engagement and satisfaction (Llerena et al., 2021).

One of the most important advantages of data science in general is that it empowers and enables improved decision-making in organisations. It has the ability to allow businesses to provide quantitative, data-driven proof for their business choices. In an ideal environment, data-driven choices improve company performance, save money, and simplify business and workflow processes. The particular economic advantages of data science differ depending on the organisation and sector. Data science, for example, may help businesses that are focused on their consumers in identifying and refining target audiences. Marketing and sales teams may mine consumer data to create more focused marketing and promotional campaigns, resulting in higher sales conversion rates. Others will involve, among other things, reducing fraud rates, enhancing risk management, boosting financial transaction profitability, increasing manufacturing uptime, improving supply chain performance, improving cybersecurity, and improving patient performance. Data science also allows for real-time data analysis as it is produced – see another Farmer article for more information on the advantages of real-time analysis, which include quicker decision-making and greater business agility. Data scientists' main job is to analyse data, often in enormous numbers, in order to discover valuable information that may be communicated with managers, managers and workers, government officials, doctors, researchers, and a variety of other people and organisations. Data scientists are also working on artificial intelligence tools and deployment techniques for a wide range of applications. In both cases, data is collected, analytical models are developed, and models are trained, tested, and run against data (Tadse et al., 2021).

Consequently, data scientists must have a broad variety of abilities, including data preparedness, data mining, predictive modelling, machine education, statistical analysis and mathematical skills, and algorithms and coding expertise, such as Python, R and SQL programming skills. Many people also create data displays, dashboards and reports that illustrate the results of analytical research. Business intelligence and reporting foundations, such as data science, help drive operational decisions and strategic planning. Business intelligence focuses primarily on descriptive analytics: what occurred or is occurring now, which requires the reaction of an organisation? Intelligence analysts and self-service users Business intelligence (BI) depends largely on the collecting, purification, transformation and storage of structured transaction data in a data warehouse or analysis mart. Corporate performance, procedures and trends are actual instances of how BI is used (Vila et al., 2018).

Data science involves more advanced analytical applications not confined to statistics. It also includes predictive analytics, which anticipate future conduct, and prescriptive analysis, which seeks to identify the optimal answer to the question under consideration. Predictive analytics is a kind of predictive analysis that predicts future behaviour and occurrences. In data science applications, unstructured or semi-structured data types, such as log files, sensor data or text, are frequently used along with structured data. For the evaluation of the whole dataset or filter and for the preparation of specific analytical applications, data scientists frequently require raw data before cleaning and aggregation. In Hadoop data lake, a cloud object storage facility, a NoSQL database or any big data platform, raw data may be stored as described above. Machine learning and implementation methods are widely used in the field of data science. Machine Learning is a sophisticated analysis in which computers discover patterns, anomalies, and insights into data sets and search them for them. It uses supervised, uncontrolled, semi-supervised and reinforced methods for learning and algorithms which are trained and overseen by data scientists (Pedroza et al., 2019).

The display of data or information in the form of graphs, charts, or other visual representations is known as data visualisation. It uses pictures to convey data connections. This is important since it helps identifying trends and patterns easier. With the advent of Big Data, I will need to be able to understand ever-increasing amounts of data. Machine learning simplifies analyses such as predictive analysis, which may then be utilised to generate aesthetically attractive visualisations. However, understanding data visualisation is essential not just for data researchers and analysts; it is also important for individuals in all other professions. You must be able to visualize information regardless of whether you work in finance, marketing, technology, design, or anything else. This finding emphasizes the significance of data visualisation. I need data visualisation instead of staring at hundreds of elements in a table since it makes it simpler to identify patterns and trends when looking at a visual information summary. This is how the human brain functions and thinks. Because data analysis is supposed to provide insight, data that is graphically displayed is much more useful. It is possible to get insights from data without utilising visualisation, however communicating this knowledge without using visualisation is more challenging. Even while patterns may be discovered without the use of graphs and charts, they are more readily conveyed when they are. Undergraduate students are often taught the importance of using visual representation to communicate data results. Without a visual depiction of the insights, viewers may struggle to understand the full significance of the results. For example, reeling off your boss's statistics will not tell you why they are worried about the information. But you'll almost certainly show them a graph showing how much money your insights might save or earn them in the future. Many data points gathered and processed by companies have a geographic component that may be easily shown on a map. An example of this is a map depiction of the number of purchases made by consumers in each state in the United States. In this scenario, each nation is represented by a distinct colour, with states with less shopping having lighter hues and states with more shopping having deeper colors. Understanding location information may also be advantageous for company executives, making this essential data visualisation tool. A heat map is a color-coded matrix in its most basic form. Each matrix cell is colored in the shade that corresponds to the relative value or risk connected with that cell using a formula. The colors of a heat map usually run from green to red, with green suggesting a better outcome and red indicating a poorer one. This kind of representation is beneficial since colors are more easily understood than numerical numbers (Molano et al., 2019).

"A picture is worth a thousand words," as the saying goes. And, in the age of Big Data, when companies are bombarded with data from a number of sources, both on-premises and in the cloud, the old adage has never been more applicable. It is getting increasingly difficult to sort through information to decide what is important and what is not. Analytical images make it simpler to recognize what is significant at a glance and to comprehend what isn't. Furthermore, most individuals react considerably more favorably to pictures than they do to text—90 percent of information supplied to the brain is visual, and images are processed in the brain at a pace 60,000 times quicker than text1. Considerations like these offer strong arguments in favor of using data visualisation for information analysis and dissemination. Data visualisation is an essential component of many technologies, particularly advanced analytics. It assists individuals in making sense of the enormous quantity of information or data that is currently accessible (Aguilar et al., 2018).

The presentation of information in graphical form, such as a pie chart, a graph, or another kind of visual display, is known as data visualisation. Effective data visualisation is essential to success in data analysis and decision-making based on data. A table or raw data table may not be able to help customers identify patterns, connections, and developing trends as fast and readily as a graph or chart. In most instances, there is no requirement for specialised expertise to comprehend what is presented in the pictures, allowing universal understanding. A well-designed graph may not only offer information, but it can also enhance the effect of that information by drawing attention and creating interest, which is especially helpful when no table or table can be used to display the information. The vast majority of data visualisation tools are capable of connecting to data sources such as relational databases (Holata et al., 2018).

For analytical reasons, this data may be kept on-premises or in the cloud. Users may then choose the best method to present data from a number of choices. Some technologies create display suggestions automatically depending on the data that is given to them. A graph should constantly consider the data type as well as the purpose of the data. Some data is more suited to one kind of graph than another: a bar graph rather than a pie chart, for example, is better suited to some data.

Nonetheless, most systems provide a broad variety of visual analytics choices, including simple diagrams and bar graphs, as well as timelines, maps, schematics, histograms, and bespoke designs. Standard diagrams and bar diagrams are among the various choices. A tool with this capacity should be able to assist you with information analysis and transmission at all stages of the process, beginning with data preparation. Because they needed significant data gathering and processing, traditional data preparation methods were time-consuming, complex, and prone to error (Horák and Klír, 2017).

Consider technology that may automate data preparation by gathering and combining information from many sources. These speeds up the procedure and lowers the chances of making a mistake. Furthermore, the tool should be able to help you improve your study by suggesting fresh data packages that will provide more accurate findings that may be included in the review. You want an interactive data visualisation tool that lets you ask questions and receive answers quickly and simply, find what you're searching for, and get to the data straight away. This may be done by using natural language interfaces, which enable you to interact with your data sources using natural language. The APIs may also be used to alter the requests and parameters used in the data collection process. Included should be a tool that allows you to choose the best graph for presentation or make a suggestion based on data findings. Also essential is the fact that consumers should be able to get predictive and predictive analytics with a single click, even if they lack specialised skills such as coding expertise, in order to identify patterns, forecast future events, and anticipate trends. Consider the potential for proactive, personalised analysis offered by a data visualisation mobile application. This feature is accessible to the user in a machine learning tool. You may have a personal assistant who knows precisely what you need and when you want it. For example, you may need to decide on the report and visuals for your business meeting in Manhattan. When you make a mobile voice query, it may convert your words into text and alert you of fresh information that can be analysed while you're on the road (Olaya et al., 2020).

Result

After our brief Conclusion of data visualisation, it is clear that this topic has many potential applications in a number of areas, but I must also be mindful of the practical and ethical problems that it raises. Many important theoretical and practical issues that must be addressed during the development of a data visualisation were presented in the chapters that preceded them. I also examined and assessed a range of data visualisation examples, as well as common errors and helpful approaches. Based on what we've learnt, developing an efficient and ethical way of viewing data may be challenging. This chapter discusses a range of topics, including the future of data visualisation and new data visualisation technologies. Finally, I came to the conclusion of our study into data visualisation. In the eleven chapters before this one, I examined some of the most important theoretical and practical components of visualisation methods and applications, as well as their applications. As we've seen, developing an efficient and effective approach for building a data visualisation application is challenging. This procedure represents relevant data. The data is then analysed in order to extract the most important information about the problem, which is then translated into a visual representation, represented again, and finally integrated into a user-friendly solution. Consider an image that is either fantastical, terrifying, or something in between. The ecological rationality of any representation determines its effectiveness: "On the one hand, the graph type and aesthetic components must be appropriate for the data being presented; on the other hand, they must be visually appealing." However, the message to be conveyed, as well as the audience's ability to perceive and comprehend the graph, must be taken into account. Visualization is an art and a talent in its own right. R comes with a plethora of tools, but learning them is mostly a matter of practice (Vila et al., 2018).

Even if you aren't seeing anything, keep in mind that any portrayal may be excellent or bad for a variety of reasons. You should not mistake this with a logical puzzle interface since the majority of people are not as interested in information as you are. If at all possible, your graphs should include information other than statistics, such as skill or knowledge. Provide your users with clear conclusions and useful recommendations based on the study's results. Looking at a Matisse or Kandinsky painting provides me with knowledge and inspiration, while it gives my wife a headache (Aguilar et al., 2018).

References

1. AGUILAR, J., CORDERO, J. & BUENDÍA, O. J. J. O. E. C. R. 2018. Specification of the autonomic cycles of learning analytic tasks for a smart classroom. 56**,** 866-891.
2. ARANGO-LÓPEZ, J. Learning Analytics as a Tool for Visual Analysis in an Open Data Environment: A Higher Education Case. Advances in Computing: 13th Colombian

Conference, CCC 2018, Cartagena, Colombia, September 26–28, 2018, Proceedings, 2018. Springer, 55.

3. AREVALO, M. L. S., PORRAS, A. A., HERRERA, J. D. G. & ESCOBAR, R. F. Proposal for the Implementation of a Business Intelligence Tool to Detect Cases of Student Desertion at the Francisco Jose de Caldas District University. International Conference on Advanced Engineering Theory and Applications, 2019. Springer, 482-489.
4. BRAVO, L. E. C., BERMUDEZ, G. M. T. & MOLANO, J. I. R. Big data: An exploration toward the improve of the academic performance in higher education. International Conference on Data Mining and Big Data, 2018. Springer, 627-637.
5. CHANG, R., ZIEMKIEWICZ, C., GREEN, T. M., RIBARSKY, W. J. I. C. G. & APPLICATIONS 2009. Defining insight for visual analytics. 29**,** 14-17.
6. DE DATOS, M. D. C. J. J. O. C. S. & TECHNOLOGY 2021. Proposed Extended Analytic Hierarchy Process for Selecting Data Science Methodologies. 21.
7. DEHMER, M., MOUTARI, S. & EMMERT-STREIB, F. 2020. 10. Mathematics as a language for science. *Mathematical Foundations of Data Science Using R.* De Gruyter Oldenbourg.
8. ECKERT, K. & BRITOS, P. V. Data science methodologies selection with hierarchical analytical process and personal construction theory. XXV Congreso Argentino de Ciencias de la Computación (CACIC)(Universidad Nacional de Río Cuarto, Córdoba, 14 al 18 de octubre de 2019), 2019.
9. ECKERT, K., BRITOS, P. V. J. J. O. C. S. & TECHNOLOGY 2021. Proposed extended analytic hierarchical process for selecting data science methodologies. 21.
10. ESPITIA, E. & MONTILLA, A. F. Applying CRISP-DM in a KDD process for the analysis of student attrition. Colombian Conference on Computing, 2018. Springer, 386-401.
11. FILZWEISER, R. & EICHERT, S. J. R. P. 2022. Towards an Online Database for Archaeological Landscapes. 2.
12. FILZWIESER, R. & EICHERT, S. J. H. 2020. Towards an Online Database for Archaeological Landscapes. Using the Web Based, Open Source Software OpenAtlas for the Acquisition, Analysis and Dissemination of Archaeological and Historical Data on a Landscape Basis. 3**,** 1385-1401.
13. FRIENDLY, M. Visualizing Multivariate Uncertainty: Some Graphical Methods for Multivariable Spatial Data.
14. FRIENDLY, M. J. J. O. E. & STATISTICS, B. 2002. Visions and re-visions of Charles Joseph Minard. 27**,** 31-51.
15. GAMA, J. A. P., CARO, R., HERNAN, C., ALVARADO, D. B., GOMEZ, C. L. C., GOMEZ, G. H. & MENA, A. M. Work in progress—New education model based on competencies of higher education and iMIS with architectures. 2016 IEEE Global Engineering Education Conference (EDUCON), 2016. IEEE, 1065-1070.
16. HOLATA, L., PLZÁK, J., SVĚTLÍK, R. & FONTE, J. J. R. S. 2018. Integration of low-resolution ALS and ground-based SfM photogrammetry data. A cost-effective approach

providing an 'Enhanced 3D Model'of the Hound Tor archaeological landscapes (Dartmoor, South-West England). 10**,** 1357.

17. HORÁK, J. & KLÍR, T. J. I. A. N. S. I. A. 2017. Pedogenesis, pedochemistry and the functional structure of the waldhufendorf field system of the deserted medieval village spindelbach, the Czech Republic. 8**,** 2017.
18. KRAAK, M.-J. J. I. J. O. P. & SENSING, R. 2003. Geovisualization illustrated. 57**,** 390-399.
19. LASAPONARA, R., COLUZZI, R., GIZZI, F. T., MASINI, N. J. J. O. G. & ENGINEERING 2010. On the LiDAR contribution for the archaeological and geomorphological study of a deserted medieval village in Southern Italy. 7**,** 155-163.
20. LLERENA, J., ÁLAVA-MORÁN, N. & ZAMORA-GALINDO, J. Learning analytics for student academic tracking, a comparison between Analytics Graphs and Edwiser Reports. 2021 Second International Conference on Information Systems and Software Technologies (ICI2ST), 2021. IEEE, 101-107.
21. MALINVERNI, E. S., GIULIANO, A. A. & MARIANO, F. 3D information management system for the conservation of an old deserted military site. 2018 Metrology for Archaeology and Cultural Heritage (MetroArchaeo), 2018. IEEE, 188-192.
22. MASINI, N., LASAPONARA, R. J. J. O. A. M. & THEORY 2020. On the Reuse of Multiscale LiDAR Data to Investigate the Resilience in the Late Medieval Time: The Case Study of Basilicata in South of Italy. 1-28.
23. MOLANO, J. I. R., ZEA, L. D. F. & REINA, Y. F. P. J. I. S. 2019. Proposal of Architecture And Application of Machine Learning (Ml) as A Strategy for the Reduction of University Desertion Levels Due to Academic Factors. 15**,** 1-23.
24. OLAYA, D., VÁSQUEZ, J., MALDONADO, S., MIRANDA, J. & VERBEKE, W. J. D. S. S. 2020. Uplift Modeling for preventing student dropout in higher education. 134**,** 113320.
25. PALACIOS, C. A., REYES-SUÁREZ, J. A., BEARZOTTI, L. A., LEIVA, V. & MARCHANT, C. J. E. 2021. Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in Chile. 23**,** 485.
26. PEDROZA, K. D., CHASOY, C. & GÓMEZ, A. R. Review of techniques, tools, algorithms and attributes for data mining used in student desertion. Journal of Physics: Conference Series, 2019. IOP Publishing, 012003.
27. PÉREZ, B., CASTELLANOS, C. & CORREAL, D. Predicting student drop-out rates using data mining techniques: A case study. IEEE Colombian Conference on Applications in Computational Intelligence, 2018. Springer, 111-125.
28. SALAZAR-CARDONA, J., ANGARITA-GARCIA, D. & ARANGO-LÓPEZ, J. Learning Analytics as a Tool for Visual Analysis in an Open Data Environment: A Higher Education Case. Colombian Conference on Computing, 2018. Springer, 55-69.
29. SKAGGS, S. J. I. D. J. 2017. The semiotics of non-virtuous data visualization: Why information design can never be pure. 23.

30. STARKOVÁ, L. J. G. 2020. Toward a High-Definition Remote Sensing Approach to the Study of Deserted Medieval Cities in the Near East. 10**,** 369.
31. TADSE, S., JAIN, M. & CHANDANKHEDE, P. Parkinson's Detection Using Machine Learning. 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021. IEEE, 1081-1085.
32. VILA, D., CISNEROS, S., GRANDA, P., ORTEGA, C., POSSO-YÉPEZ, M. & GARCÍA-SANTILLÁN, I. Detection of desertion patterns in university students using data mining techniques: A case study. International Conference on Technology Trends, 2018. Springer, 420-429.
33. WENJUN, D., HAI, R., SHENGLEI, F., JUN, W., LONG, Y. & ZHANG, J. J. J. O. E. S. 2008. Natural recovery of different areas of a deserted quarry in South China. 20**,** 476-481.