Frameworks for Data Science

- Author: Shadab Malik
- Mail ID: shadab.malik@iubh.de
- Matriculation Number: 32003135

- Module: Current Topics in Data Science (DLMDSSCTDS01)
- Date: 17-04-2024
- Place: Berlin, German

Table of Contents

1. Introduction 1
2. Background 2
3. What Is The Data Analysis Process? 3
4. Regression Analysis 4
5. Cluster Analysis 4
6. Time Series Analysis 6
7. Decision Tree 7
8. Data Mining 7
9. Conjoint Analysis 8
10. Multidimensional Scaling 8
11. Cohort Analysis 10
12. Sentiment Analysis 11
13. Conclusion 13
14. Literature 14

TABLE OF FIGURE

FIGURE 1 3

FIGURE 2 5

FIGURE 3 8

FIGURE 4..10

FIGURE 5 11

1. Introduction

The past 15 years have seen extensive investments in business infrastructure, which have improved the ability to collect data throughout the industries. Every aspect of the the business is now open to data collection and often even instrumented for data collection: operations, manufacturing, supply -chain management, customer behaviour, marketing campaign performance, workflow procedures and etc. At the same time, information is now widely available on external events such as market trends, industry news and competitors analysis. This broad availability of data has led to increasing interest in methods for extracting useful information and knowledge from data.

With vast amounts of data now available, companies in almost every industry are focused on exploiting data for competitive advantage. In the past, firms could employ teams of statisticians, modelers, and analysts to explore datasets manually, but the volume and variety of data have far outstripped the capacity of manual analysis. At the same time, computers have become far more powerful, networking has become ubiquitous, and algorithms have been developed that can connect datasets to enable broader and deeper analyses than previously possible. The convergence of these phenomena has given rise to the increasingly widespread business application of data science principles and data-mining techniques.[5]

Probably the widest applications of data-analysis techniques are in marketing for tasks such as targeted marketing, online advertising, and recommendations for cross-selling. Data analysis is used for general customer relationship management to analyze customer behaviour in order to manage attrition and maximize expected customer value. The finance industry uses data mining for credit scoring and trading, and in operations via fraud detection and workforce management. Major retailers from Walmart to Amazon apply data analysis throughout their businesses, from marketing to supply-chain management. Many firms have differentiated themselves strategically with data science, sometimes to the point of evolving into data mining companies.

2. Background

In today's digital age, the abundance of data generated from various sources such as social media, sensors, and transactional systems has transformed the way organizations operate and make decisions. The ability to extract meaningful insights from this vast amount of data has become a crucial competitive advantage for businesses and institutions across all sectors. This process of extracting insights from data, known as data analysis, encompasses a wide range of techniques and methodologies aimed at uncovering patterns, trends, and relationships within datasets.

The field of data analysis has a rich history that dates back centuries, with roots in disciplines such as statistics, mathematics, and computer science. Early practitioners relied on manual methods and basic statistical techniques to analyze data, often limited by the available computing power and data storage capabilities. However, with the advent of modern computing technologies and the proliferation of big data, data analysis has evolved into a sophisticated and multidisciplinary field encompassing statistical modeling, machine learning, data mining, and more.

Data analysis plays a pivotal role in enabling organizations to make data-driven decisions, optimize processes, and gain actionable insights into their operations and customers. From identifying market trends and forecasting demand to detecting fraud and improving healthcare outcomes, the applications of data analysis are diverse and far-reaching. Moreover, as industries continue to digitize and generate increasingly larger volumes of data, the demand for skilled data analysts and data scientists capable of interpreting and extracting value from this data is expected to grow exponentially.[12]

Against this backdrop, it is essential for individuals and organizations to develop a solid understanding of data analysis principles and techniques. By mastering the fundamentals of data analysis, individuals can unlock new opportunities for career advancement and contribute to driving innovation and growth within their respective fields. Thus, there is a pressing need for comprehensive educational resources that provide a structured introduction to data analysis, covering key concepts, methodologies, and real-world applications. This background sets the stage for the development of a comprehensive introduction to data analysis that equips readers with the knowledge and skills needed to navigate the complex landscape of data-driven decision-making.

3. What Is The Data Analysis Process?

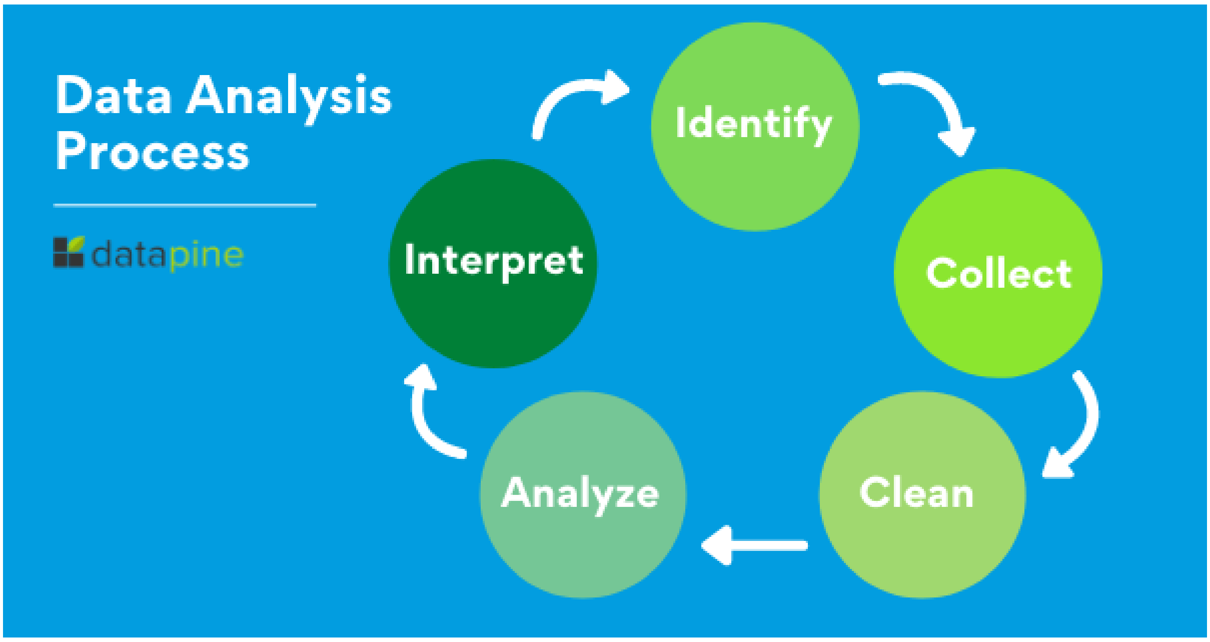

Figure 1

When we talk about analyzing data there is an order to follow in order to extract the needed conclusions. The analysis process consists of 5 key stages. We will cover each of them more in detail later in the post, but to start providing the needed context to understand what is coming next, here is a rundown of the 5 essential steps of data analysis.[2]

a) **Identify:** Before you get your hands dirty with data, you first need to identify why you need it in the first place. The identification is the stage in which you establish the questions you will need to answer. For example, what is the customer's perception of our brand? Or what type of packaging is more engaging to our potential customers? Once the questions are outlined you are ready for the next step.
b) **Collect:** As its name suggests, this is the stage where you start collecting the needed data. Here, you define which sources of data you will use and how you will use them. The collection of data can come in different forms such as internal or external sources, surveys, interviews, questionnaires, and focus groups, among others. An important note here is that the way you collect the data will be different in a quantitative and qualitative scenario.
c) **Clean:** Once you have the necessary data it is time to clean it and leave it ready for analysis. Not all the data you collect will be useful, when collecting big amounts of data in different formats it is very likely that you will find yourself with duplicate or badly formatted data. To avoid this, before you start working with your data you need to make sure to erase any white spaces, duplicate records, or formatting errors. This way you avoid hurting your analysis with bad-quality data.
d) **Analyze**: With the help of various techniques such as statistical analysis, regressions, neural networks, text analysis, and more, you can start analyzing and manipulating your data to extract relevant conclusions. At this stage, you find trends, correlations, variations, and patterns that can help you answer the questions you first thought of in the identify stage. Various technologies in the market assist researchers and average users with the management of their data. Some of them include business intelligence and visualization software, predictive analytics, and data mining, among others.
e) **Interpret:** Last but not least you have one of the most important steps: it is time to interpret your results. This stage is where the researcher comes up with courses of action based on the findings. For example, here you would understand if your clients prefer packaging that

is red or green, plastic or paper, etc. Additionally, at this stage, you can also find some limitations and work on them.

4. Regression Analysis

Regression ("value estimation") attempts to estimate or predict, for everyone, the numerical value of some variable for that individual. An example regression question would be: "How much will a given customer use the service?" The property (variable) to be predicted here is service usage, and a model could be generated by looking at other, similar individuals in the population and their historical usage. A regression procedure produces a model that, given an individual, estimates the value of the variable specific to that individual. Regression is related to classification, but the two are different. Informally, classification predicts whether something will happen, whereas regression predicts how much something will happen.[1]

Regression uses historical data to understand how a dependent variable's value is affected when one (linear regression) or more independent variables (multiple regression) change or stay the same. By understanding each variable's relationship and how it developed in the past, you can anticipate possible outcomes and make better decisions in the future.[3]

Let's bring it down with an example. Imagine you did a regression analysis of your sales in 2019 and discovered that variables like product quality, store design, customer service, marketing campaigns, and sales channels affected the overall result. Now you want to use regression to analyze which of these variables changed or if any new ones appeared during 2020. For example, you couldn't sell as much in your physical store due to COVID lockdowns. Therefore, your sales could've either dropped in general or increased in your online channels. Through this, you can understand which independent variables affected the overall performance of your dependent variable, annual sales.

5. Cluster Analysis

Cluster analysis or simply clustering is the process of partitioning a set of data objects (or observations) into subsets. Each subset is a cluster, such that objects in a cluster are similar to one another, yet dissimilar to objects in other clusters. The set of clusters resulting from a cluster analysis can be referred to as a clustering. In this context, different clustering methods may generate different clusterings on the same data set. The partitioning is performed, not by humans, but by the clustering algorithm. Hence, clustering is useful in that it can lead to the discovery of previously unknown groups within the data. Cluster analysis has been widely used in many applications, such as business intelligence, image pattern recognition, Web search, biology, and security. In business intelligence, clustering can be used to organize a large number of customers into groups, where customers within a group share strong similar characteristics. This facilitates the development of business strategies for enhanced customer relationship management. Moreover, consider a consultant company with a large number of projects. To improve project management, clustering can be applied to partition projects into categories based on similarity so that project auditing and diagnosis (to improve project delivery and outcomes) can be conducted effectively. In image recognition, clustering can be used to discover clusters or "subclasses" in handwritten character recognition systems. Suppose we have a data set of handwritten digits, where each digit is labeled as either 1, 2, or 3, and so on. Note that there can be a large variance in the way in which people write the same digit. Take the number "2", for example. Some people may write it with a small circle at the left bottom part, while some others may not. We can use clustering to determine subclasses for "2", each of which represents a variation on the way in which “2” can be written. Using multiple models based on the subclasses can improve the overall recognition accuracy. Clustering has also found many applications in Web search. For example, a keyword search may often return a very large number of hits (that is, pages relevant to the search) due to the extremely large number of Web pages. Clustering can be used to organize the search results into groups and present the results in a concise and easily accessible way. Moreover, clustering techniques have been developed to cluster documents into topics, which are commonly used in information retrieval practice.

A cluster analysis plot can look as follows:

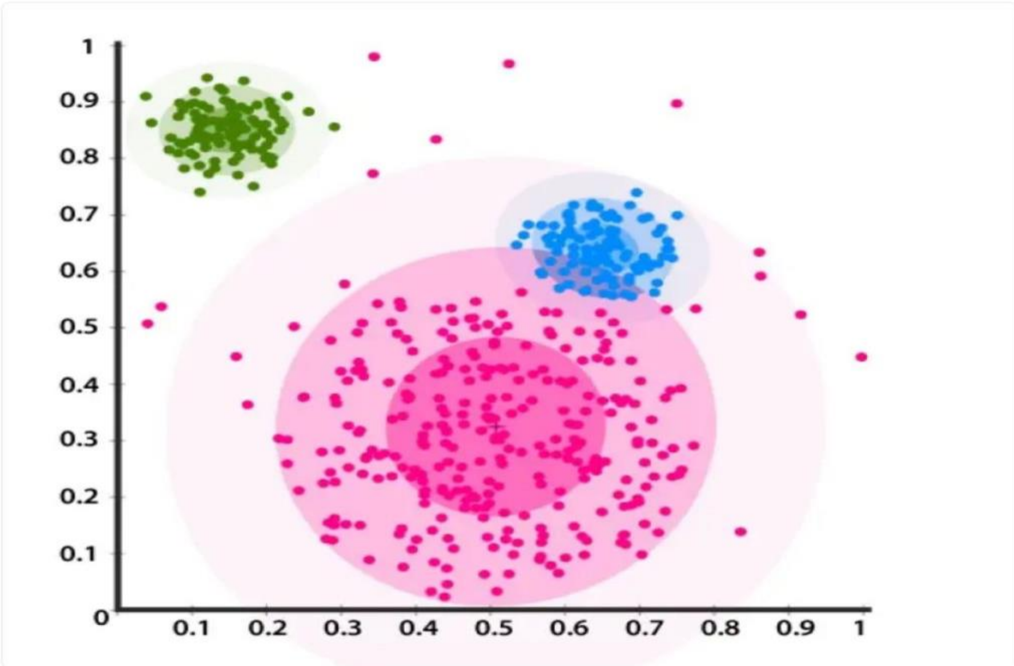

Image Source: https://byjus.com/maths/cluster-analysis/

Figure 2

As a data mining function, cluster analysis can be used as a stand-alone tool to gain insight into the distribution of data, to observe the characteristics of each cluster, and to focus on a particular set of clusters for further analysis. Alternatively, it may serve as a preprocessing step for other algorithms, such as characterization, attribute subset selection, and classification, which would then operate on the detected clusters and the selected attributes or features. Because a cluster is a collection of data objects that are similar to one another within the cluster and dissimilar to objects in other clusters, a cluster of data objects can be treated as an implicit class. In this sense, clustering is sometimes called automatic classification. Again, a critical difference here is that clustering can automatically find the groupings. This is a distinct advantage of cluster analysis. Clustering is also called data segmentation in some applications because clustering partitions large data sets into groups according to their similarity. Clustering can also be used for outlier detection, where outliers (values that are “far away” from any cluster) may be more interesting than common cases. Applications of outlier detection include the detection of credit card fraud and the monitoring of criminal activities in electronic commerce. For example, exceptional cases in credit card transactions, such as very expensive and infrequent purchases, may be of interest as possible fraudulent activities.[8]

6. Time Series Analysis

As its name suggests, time series analysis is used to analyze a set of data points collected over a specified period of time. Although analysts use this method to monitor the data points in a specific interval of time rather than just monitoring them intermittently, the time series analysis is not uniquely used for the purpose of collecting data over time. Instead, it allows researchers to understand if variables changed during the duration of the study, how the different variables are dependent, and how did it reach the end result.

In a business context, this method is used to understand the causes of different trends and patterns to extract valuable insights. Another way of using this method is with the help of time series forecasting. Powered by predictive technologies, businesses can analyze various data sets over a period of time and forecast different future events.[2]

A great use case to put time series analysis into perspective is seasonality effects on sales. By using time series forecasting to analyze sales data of a specific product over time, you can understand if sales rise over a specific period of time (e.g. swimwear during summertime, or candy during Halloween). These insights allow us to predict demand and prepare production accordingly. It is a part of predictive analysis, gathering data over consisten interval of time.

Let’s take an example of Coca Cola and look at a time series analysis example through the lens of the company’s sales. Two quarters from now, their expected sales will be anywhere between 250,000 and 300,000 units. Historical sales indicate a strong relationship between unit sales and weather. Based on that, it is likely that the numbers will be closer to 290,000 in the summer months. However, to achieve similar results in the winter quarter, the company will need some additional marketing investments.

The technique the Coca-Cola team can use to perform this type of future forecasting is precisely time series analysis. When applied, the model will provide a range of potential outcomes. In our example, the variable we are interested to predict is future sales volume. Therefore, the outcomes will vary depending on numerous factors, which may affect sales development throughout the year.

Let’s suppose the weather is 5% warmer than average, and Coca-Cola spend 5% more on marketing by investing in TV ads and promotional events. Then, based on historical data, we can reasonably expect that sales will be on the higher end of the range we indicated - 290,000 units. By changing the weather condition assumptions and running hypothesis testing on different marketing spend, the model would yield a separate time series analysis forecast. Typically, in practice, we will provide a range of estimates. For Coca-Cola, they might look something like this:

- 290,000 units in the best-case scenario
- 250,000 units in the worst-case scenario
- 270,000 in a base-case scenario

7. Decision Tree

Decision trees are a simple way to guide one's path to a decision. The decision may be a simple binary one, whether to approve a loan or not. Or it may be a complex multi-valued decision, as to what may be the diagnosis for a particular sickness. Decision trees are hierarchically branched structures that help one come to a decision based on asking certain questions in a particular sequence. Decision trees are one of the most widely used techniques for classification. A good decision tree should be short and ask only a few meaningful questions. They are very efficient to use, easy to explain, and their classification accuracy is competitive with other methods. Decision trees can generate knowledge from a few test instances that can then be applied to a broad population. Decision trees are used mostly to answer relatively simple binary decisions.[6]

Decision tree analysis is a helpful technique for strategic decision-making. The method displays potential costs, outcomes, and consequences via a tree-like model. The format makes assessing all the factors involved and choosing the best action easier. A decision tree works like a flowchart. It begins with the main decision or problem that branches out based on various potential consequences and outcomes for each decision. Each outcome outlines its costs, consequences, and gains. At the end of the analysis, you can compare and choose the best course of action based on each outcome. Businesses can use decision trees to determine the most cost-effective projects and those that provide long-term revenue. Let's say you want to decide between updating your existing app or building an entirely new one. You can do a decision tree analysis to assess and compare each option's total costs, time you need to invest, and potential revenue. It can give you a better idea about the more realistic, attainable, and profitable option for your company.[4]

8. Data Mining

A method of data analysis that is the umbrella term for engineering metrics and insights for additional value, direction, and context. By using exploratory statistical evaluation, data mining aims to identify dependencies, relations, patterns, and trends to generate advanced knowledge. When considering how to analyze data, adopting a data mining mindset is essential to success - as such, it's an area that is worth exploring in greater detail.[2]

An excellent use case of data mining is datapine intelligent data alerts. With the help of artificial intelligence and machine learning, they provide automated signals based on particular commands or occurrences within a dataset. For example, if you're monitoring supply chain KPIs, you could set an intelligent alarm to trigger when invalid or low-quality data appears. By doing so, you will be able to drill down deep into the issue and fix it swiftly and effectively.

In the following picture, you can see how the intelligent alarms from datapine work. By setting up ranges on daily orders, sessions, and revenues, the alarms will notify you if the goal was not completed or if it exceeded expectations.

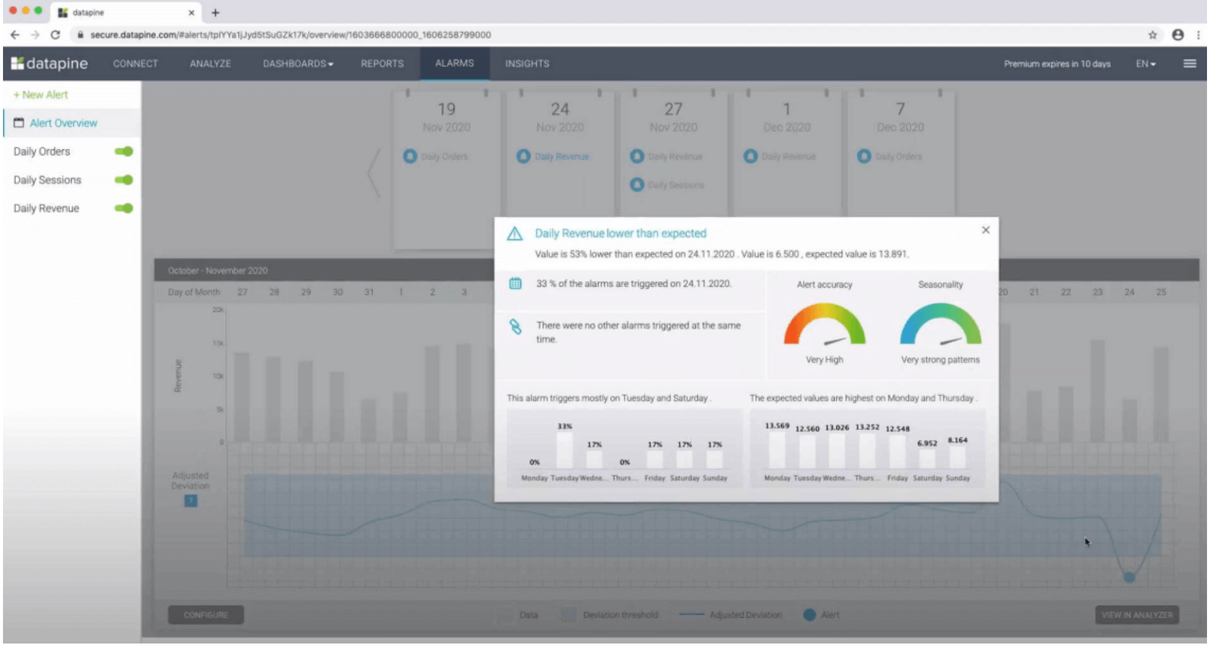

Figure 3

9. Conjoint Analysis

This approach is usually used in surveys to understand how individuals value different attributes of a product or service and it is one of the most effective methods to extract consumer preferences. When it comes to purchasing, some clients might be more price-focused, others more features-focused, and others might have a sustainable focus. Whatever your customer's preferences are, you can find them with conjoint analysis. Through this, companies can define pricing strategies, packaging options, subscription packages, and more.

A great example of conjoint analysis is in marketing and sales. For instance, a cupcake brand might use conjoint analysis and find that its clients prefer gluten-free options and cupcakes with healthier toppings over super sugary ones. Thus, the cupcake brand can turn these insights into advertisements and promotions to increase sales of this particular type of product. And not just that, conjoint analysis can also help businesses segment their customers based on their interests. This allows them to send different messaging that will bring value to each of the segments.

10. Multidimensional Scaling

Multidimensional scaling is one of several multivariate techniques that aims to reveal the structure of a dataset by plotting points in one or 2D. The basic idea can be motivated by a geographical example. Let's assume that the distance between pairs of cities are told and we are asked to reconstruct the 2D map from where those distances have been derived. We can do this by a process of simple trial and error method by moving points on a piece of paper until we get the distance right. An algorithm which does this automatically is known as multidimensional scaling (MDS)

MDS is a method used to observe the similarities or disparities between objects which can be colors, brands, people, geographical coordinates, and more. The objects are plotted using an "MDS map" that positions similar objects together and disparate ones far apart. The (dis) similarities between objects are represented using one or more dimensions that can be observed using a numerical scale. For example, if you want to know how people feel about the COVID-19 vaccine, you can use 1 for "don't believe in the vaccine at all" and 10 for "firmly believe in the vaccine" and a scale of 2 to 9 for in between responses. When analyzing an MDS map the only thing that matters is the distance between the objects, the orientation of the dimensions is arbitrary and has no meaning at all.

Multidimensional scaling is a valuable technique for market research, especially when it comes to evaluating product or brand positioning. For instance, if a cupcake brand wants to know how they are positioned compared to competitors, it can define 2-3 dimensions such as taste, ingredients, shopping experience, or more, and do a multidimensional scaling analysis to find improvement opportunities as well as areas in which competitors are currently leading.

Another business example is in procurement when deciding on different suppliers. Decision makers can generate an MDS map to see how the different prices, delivery times, technical services, and more of the different suppliers differ and pick the one that suits their needs the best.

A final example proposed by a research paper on "An Improved Study of Multilevel Semantic Network Visualization for Analyzing Sentiment Word of Movie Review Data". Researchers picked a two-dimensional MDS map to display the distances and relationships between different sentiments in movie reviews. They used 36 sentiment words and distributed them based on their emotional distance as we can see in the image below where the words "outraged" and "sweet" are on opposite sides of the map, marking the distance between the two emotions very clearly.[7]

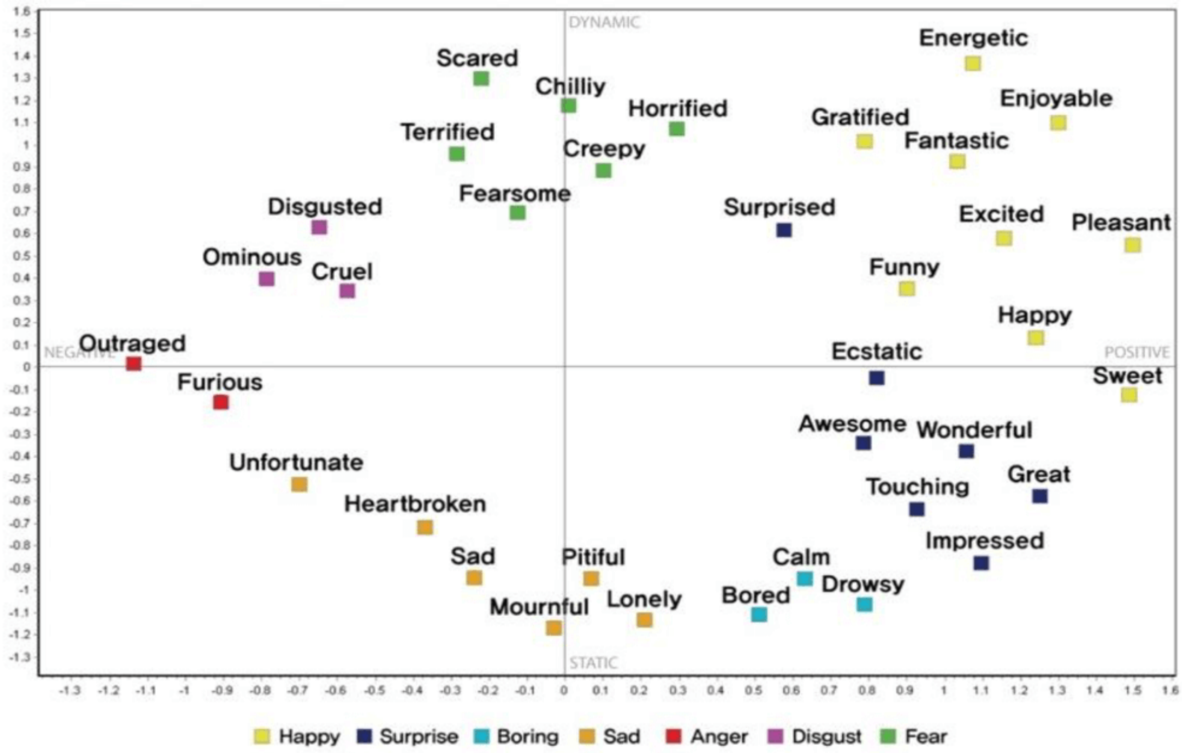

Figure 4

11. Cohort Analysis

Cohort analysis is a method of analyzing data that involves grouping users or customers into groups or "cohorts" based on shared characteristics or behaviors. This can be used to track changes in behavior or engagement over time, and to identify trends or patterns that may not be apparent when looking at data for the entire user or customer base.

Cohort analysis can be used in various contexts, including ecommerce digital marketing, and product development. Some common use cases for cohort analysis include:

1. Identifying trends in customer behavior: By analyzing data for different cohort groups, you can identify trends in customer behavior over time, such as changes in purchase frequency or average order value.
2. Identifying retention trends: Cohort analysis can be used to track retention rate over time, helping you understand how well you are retaining your customers and identifying opportunities to improve retention.
3. Identifying product or feature adoption: Cohort analysis can be used to track the adoption of new products or features, helping you understand how well your customers are receiving them and identify any issues that may be preventing adoption.
4. Identifying customer segments: By analyzing data for different cohort groups, you can identify distinct customer segmentation and understand their behaviors and preferences.

To perform cohort analysis, you will need to have a customer data platform or access to data on your customers or users, including information on their behaviors or actions over time. This data can be analyzed using various tools and techniques, such as spreadsheet software or specialized analytics platforms.[9]

This type of data analysis approach uses historical data to examine and compare a determined segment of users' behaviour, which can then be grouped with others with similar characteristics. By using this methodology, it's possible to gain a wealth of insight into consumer needs or a firm understanding of a broader target group.

Cohort analysis can be really useful for performing analysis in marketing as it will allow you to understand the impact of your campaigns on specific groups of customers. To exemplify, imagine you send an email campaign encouraging customers to sign up for your site. For this, you create two versions of the campaign with different designs, CTAs, and ad content. Later on, you can use cohort analysis to track the performance of the campaign for a longer period of time and understand which type of content is driving your customers to sign up, repurchase, or engage in other ways.

A useful tool to start performing cohort analysis method is Google Analytics. You can learn more about the benefits and limitations of using cohorts in GA in this useful guide. In the bottom image, you see an example of how you visualize a cohort in this tool. The segments (devices traffic) are divided into date cohorts (usage of devices) and then analyzed week by week to extract insights into performance.

	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
Tablet and Desktop Traffic 63,733 users	**99.99%**	**3.41%**	**1.99%**	**1.55%**	**1.05%**	**0.63%**	**0.16%**
Aug 13, 2017 - Aug 19, 2017 10,458 users	99.96%	3.35%	2.21%	1.67%	1.43%	1.08%	0.16%
Aug 20, 2017 - Aug 26, 2017 10,444 users	100.00%	3.54%	1.96%	1.77%	1.55%	0.18%	
Aug 27, 2017 - Sep 2, 2017 10,835 users	99.96%	3.78%	2.44%	2.41%	0.18%		
Sep 3, 2017 - Sep 9, 2017 9,735 users	100.00%	4.04%	2.96%	0.22%			
Sep 10, 2017 - Sep 16, 2017 9,876 users	100.00%	4.92%	0.32%				
Sep 17, 2017 - Sep 23, 2017 12,385 users	100.00%	1.32%					
Mobile Traffic 32,335 users	**99.97%**	**2.66%**	**0.90%**	**0.69%**	**0.58%**	**0.46%**	**0.00%**
Aug 13, 2017 - Aug 19, 2017 5,302 users	99.94%	2.81%	0.98%	0.64%	0.85%	0.81%	0.00%
Aug 20, 2017 - Aug 26, 2017 5,346 users	99.96%	2.49%	1.01%	1.12%	0.75%	0.11%	
Aug 27, 2017 - Sep 2, 2017 4,511 users	99.93%	2.33%	1.13%	0.84%	0.07%		
Sep 3, 2017 - Sep 9, 2017 4,541 users	99.98%	3.28%	0.99%	0.09%			
Sep 10, 2017 - Sep 16, 2017 4,915 users	99.98%	4.46%	0.39%				
Sep 17, 2017 - Sep 23, 2017 7,720 users	100.00%	1.36%					

Figure 5

12. Sentiment Analysis

It is the process of analyzing digital texts to determine if the emotional tone of the message is positive, negative or neutral. Today, companies have large volumes of text data like emails, customer support chat transcripts, social media comments, and reviews. Sentiment analysis tools can scan this text to automatically determine the author's attitude towards a topic.[10] Companies use the insights from sentiment analysis to improve customer service and increase brand reputation. Sentiment analysis, also known as opinion mining, is an important business intelligence tool that helps companies improve their products and services. We give some benefits of sentiment analysis below.

Businesses use sentiment analysis to derive intelligence and form actionable plans in different areas.

Improve customer service **:** Customer support teams use sentiment analysis tools to personalize responses based on the mood of the conversation. Matters with urgency are spotted by artificial intelligence (AI)–based chatbots with sentiment analysis capability and escalated to the support personnel.

Brand monitoring **:** Organizations constantly monitor mentions and chatter around their brands on social media, forums, blogs, news articles, and in other digital spaces. Sentiment analysis technologies allow the public relations team to be aware of related ongoing stories. The team can evaluate the underlying mood to address complaints or capitalize on positive trends.

Market research **:** A sentiment analysis system helps businesses improve their product offerings by learning what works and what doesn't. Marketers can analyze comments on online review sites, survey responses, and social media posts to gain deeper insights into specific product features. They convey the findings to the product engineers who innovate accordingly.

Track campaign performance **:** Marketers use sentiment analysis tools to ensure that their advertising campaign generates the expected response. They track conversations on social media platforms and ensure that the overall sentiment is encouraging. If the net sentiment falls short of expectation, marketers tweak the campaign based on real-time data analytics.

Despite advancements in natural language processing (NLP) technologies, understanding human language is challenging for machines. They may misinterpret finer nuances of human communication such as those given below.

Sarcasm **:** It is extremely difficult for a computer to analyze sentiment in sentences that comprise sarcasm. Consider the following sentence, *Yeah, great. It took three weeks for my order to arrive*. Unless the computer analyzes the sentence with a complete understanding of the scenario, it will label the experience as positive based on the word *great*.

Negation **:** Negation is the use of negative words to convey a reversal of meaning in the sentence. For example, *I wouldn't say the subscription was expensive.* Sentiment analysis algorithms might have difficulty interpreting such sentences correctly, particularly if the negation happens across two sentences, such as, *I thought the subscription was cheap. It wasn't.*

Multipolarity **:** Multipolarity occurs when a sentence contains more than one sentiment. For example, a product review reads, *I'm happy with the sturdy build but not impressed with the color.* It becomes difficult for the software to interpret the underlying sentiment. You'll need to use aspect-based sentiment analysis to extract each entity and its corresponding emotion.

13. Conclusion

The field of data science offers diverse array of methods and techniques to extract valuable insights from the data, enabling informed decision-making and driving innovation across various industries. From regression analysis and data mining to cluster analysis, data analysts have wide range of algorithms and tools at their disposal, to analyse, interpret and derive insights.

The importance of data analysis cannot be overstated in today's data-driven world, where organizations rely on data-driven insights to gain a competitive edge, optimize processes, and enhance customer experiences. By leveraging techniques such as inferential statistics, predictive modelling, and association analysis, businesses can uncover hidden patterns, identify trends, and anticipate future outcomes with a high degree of accuracy.

Furthermore, advancements in technology, such as big data analytics, machine learning, and artificial intelligence, have expanded the horizons of data analysis, enabling analysts to tackle larger and more complex datasets with unprecedented speed and efficiency. Techniques such as natural language processing and spatial analysis have also opened up new avenues for analyzing unstructured and spatially distributed data, unlocking valuable insights from text documents, social media feeds, and geographic information systems.

However, with these opportunities come challenges, including data privacy concerns, ethical considerations, and the need for robust data governance frameworks. As data analysis continues to evolve, it is essential for practitioners to uphold ethical standards, ensure data privacy, and use data responsibly to avoid bias and discrimination.

In conclusion, data analysis methods and techniques play a vital role in shaping our understanding of the world around us, driving innovation, and informing decision-making processes. By mastering the fundamentals of data analysis and staying abreast of emerging trends and technologies, individuals and organizations can harness the power of data to drive positive change and create a brighter future for all.

14. Literature

[1] https://careerfoundry.com/en/blog/data-analytics/data-analysis-techniques/

[2] https://www.datapine.com/blog/data-analysis-methods-and-techniques/

[3] https://economictimes.indiatimes.com/jobs/c-suite/7-highly-effective-data-analysis-methods-and-techniques/articleshow/103310016.cms?from=mdr

[4] https://www.polymersearch.com/data-analysis-guide/15-data-analysis-techniques-methods

[5] Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking" by Foster Provost and Tom Fawcett pp. 39, 170-180, 2018

[6] Data Analytics made accessible by Dr. Anil Maheshwari pp. 112-124

[7] Hyoji Ha, Hyunwoo Han, Seongmin mun, Sungyun Bae, Jihye Lee and Kyungwon Lee, 2019 " An improved study of multilevel semantic network visualization for analyzing sentiment word of Movie review data"
https://doi.org/10.3390/app9122419

[8] https://hanj.cs.illinois.edu/cs412/bk3/10.pdf pp 4-6

[9] https://sarasanalytics.com/blog/cohort-analysis/

[10] https://aws.amazon.com/what-is/sentiment-analysis/

[11] Saif, S. & Wazir, S. (2018). Performance Analysis of Big Data and Cloud Computing Techniques: A Survey, Procedia Computer Science, vol 132, pp. 118-127.

[12] Nelli, F. (2018). Python Data Analytics With Pandas, NumPy, and Matplotlib. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-3913-1