

A Comparison of Descriptive, Diagnostic, Predictive, and Prescriptive Analytics

23 June 2023

Contents

1. [Descriptive analysis: Walmart 2](#_bookmark0)

[Background on Walmart 2](#_bookmark1)

[What is descriptive analysis? 2](#_bookmark2)

[Data analytics approaches and methods. 3](#_bookmark3)

[How does Walmart use descriptive data to improve business operations? 3](#_bookmark4)

[Challenges 5](#_bookmark5)

[Legislation on ethics and data privacy 5](#_bookmark6)

[Conclusion 6](#_bookmark7)

1. [Diagnostic analysis: Cork company in Portugal 6](#_bookmark8)

[Background of Cork History and its challenge 6](#_bookmark9)

[What is diagnostic analysis? 6](#_bookmark10)

[How does the Cork Company use descriptive analytics? 7](#_bookmark11)

[Data analytics approaches and methods. 7](#_bookmark12)

[Challenges 8](#_bookmark13)

[Legislation on ethics and data privacy 8](#_bookmark14)

[Conclusion 9](#_bookmark15)

1. [Predictive analysis: Netflix 9](#_bookmark16)

[Background on Netflix 9](#_bookmark17)

[What is predictive analysis? 10](#_bookmark18)

[How does Netflix use predictive analysis? 10](#_bookmark19)

[Challenges 11](#_bookmark20)

[Legislation on ethics and data privacy 11](#_bookmark21)

[Conclusion 12](#_bookmark22)

1. [Prescriptive analysis: The Coca Cola Company 12](#_bookmark23)

[Background of The Coca Cola Company 12](#_bookmark24)

[What is prescriptive analysis? 13](#_bookmark25)

[How does Coca Cola use prescriptive analysis? 13](#_bookmark26)

[Challenges 14](#_bookmark27)

[Legislation on ethics and data privacy 14](#_bookmark28)

[Conclusion 15](#_bookmark29)

1. [References 15](#_bookmark30)

# Descriptive analysis: Walmart Background on Walmart

Walmart is a multinational retail corporation that was founded in 1962 in Arkansas, United States of America by Sam Walton (Walmart(a), n.d). It is currently one of the largest retailer in the world, with over 10 900 stores in 27 countries, and employs more than 2 million people. Walmart operates a variety of retail formats, including online stores and hypermarkets.(ProjectPro, 2023).

Walmart wants to continually understand their customer demands and optimal prices of products that their customers are willing to pay, so that Walmart can adjust their prices based on market prices and customer behaviour (ProjectPro, 2023). Walmart has always prioritized understanding its customers and their needs to satisfy them and increase product demand, which would lead to increased returns. Since Walmart has a several networks of stores located in different geographical locations it can be challenging to fully understand the needs of its customers(ProjectPro, 2023). Therefore, Walmart uses descriptive data analysis to analyze large amounts of data and gain valuable insights into its customer's behaviour (ProjectPro, 2023).

# What is descriptive analysis?

Descriptive data analysis is one of the methods used by Walmart to analyze data. The Harvard Business school describes descriptive analytics as the interpretation of historical data to better understand current or previous changes that occur in a business and answer the question “ What happened”. Descriptive data analysis involves summarizing and interpreting data to provide a clear picture of what happened in the past. Moreover, descriptive analytics describes the use of a range of historic data to draw comparisons, using sentiment analysis and measures of central tendency (Harper, 2019).

Furthermore, descriptive data analysis refers to the process of examining and summarizing a dataset using various statistical and visual representations to identify patterns that answer what happened (Whitney and Liu, 2017). It involves organizing, describing, and interpreting data in order to gain insights and understand the key characteristics of a dataset to answer the question “What happened”. Descriptive data analysis helps researchers, analysts, and decision-makers to gain an understanding of the dataset, identify potential issues or areas of interest, and generate hypotheses for further investigation.

The two main statistical analysis types used in data analysis is descriptive and inferential statistics. Descriptive statistics is concerned with summarizing and describing the features of a dataset, while inferential statistics focus on making inferences and generalizations about a population based on sample data. Both types of descriptive analytics provide valuable insights into data. Descriptive statistics involves calculating and interpreting measures of central tendency which are the mean- average of the data set, median-represents the middle value in a dataset, mode-most frequent values in dataset or datasets. The measures of central tendency are used to provide information about the central value or average value. The measures of variability (e.g.,

range, variance, standard deviation) are also used to understand the distribution and spread of the data points in a dataset (Chakrabarty, 2021).

Stevens (2023) explains that once the data has been analysed it is then visualized. Data visualization inlcudes creating graphs, charts, and plots to visually represent the data. Common visualizations include bar graphs, line graphs histograms, box plots and scatter plots. Visualization helps decision makers to easily see the patterns, trends, and relationships within the dataset.

To further analyze the data, exploratory data analysis is conducted. Exploratory data analysis provides deeper analysis by examining relationships between variables, identifying trends, and detecting outliers or anomalies. This may involve computing correlations, conducting regression analysis, or performing other statistical techniques to uncover insights. The data can then be used to accept or reject a hypothesis and to then write a summary to explain the findings, patterns and trends identified by the analysis.

# Data analytics approaches and methods.

Sing (2021) explained that in the descriptive analysis of Walmart, a variety of tools and techniques were employed to process and analyze the sales datasets. Apache Spark served as the primary framework for data processing, offering capabilities for both batch and stream processing. The Walmart sales datasets in CSV format were stored using the Hadoop Distributed File System (HDFS), which was integrated with Spark to ensure efficient data storage.

The CSV datasets were loaded into Spark DataFrames, which are distributed collections of data organized into named columns. Spark DataFrames simplify the processing and querying of big data. Spark SQL, a component of Spark, was utilized to query and manipulate the data stored in Spark DataFrames, enabling SQL-like operations on distributed data. Furthermore, MapReduce functions, applied through Spark SQL, allowed for operations such as grouping, sorting, and aggregating to be performed on the datasets. These operations were crucial in analyzing the data and extracting meaningful insights. After applying the necessary operations, the processed datasets were saved back to HDFS in CSV format.

To facilitate interactive and exploratory data analysis, iPython Notebook was employed as the Python shell for PySpark. This combination allowed for a interactive environment to conduct data analysis tasks. Additionally, the Pandas library was utilized for data visualization. In addition to descriptive analysis, a regression model utilizing machine learning techniques was employed. The model aimed to identify the relationship between variables such as fuel price, temperature, unemployment rate, holidays, and store sales. By understanding these relationships, Walmart could gain insights into the factors affecting sales and make informed decisions regarding resource allocation and competitive advantage.

# How does Walmart use descriptive data to improve business operations?

The analysis of historical data that is presented to Walmart, through descriptive analysis, is helpful to Walmart as it allows the company to understand its existing performance and identify areas where it can improve by answering questions such as; What is the distribution of sales across

different regions? or How does Walmart's pricing compare to competitors? (ProjectPro, 2023). Through the descriptive analysis of the customer data which contains information about; geographic location, brands purchased the most/least, consumer spending average, etc., Walmart gains insights into customer behaviour, trends, and preferences, which can inform the company's marketing and product development strategies. Furthermore, descriptive data analysis is helpful to Walmart as it enables the company to understand the factors -such as unemployment rate, fuel prices, temperature, and holidays- in the Walmart stores at different geographical locations - affecting sales (Singh, 2021). By analysing the abovementioned factors, Walmart manages to understand factors that are affecting the purchasing power and demands of their customers. Hence, resources of Walmart are efficiently managed and channelled to maximize returns and have a more precise target marketing strategy. For instance, Singh et al. 2021, who conducted a descriptive data analysis on Walmart’s big data found that Easter holiday induces a sharp increase in sales. Walmart can use the insights from (Singh et al., 2021) to better allocate more resources – increase the supply of goods that are sold the most around easter- around Easter holidays. This type of target marketing would increase returns around Easter holidays.

Walmart utilizes analytical tools such as Apache data science platforms, libraries, and tools to gain new insights into consumer behaviour and comprehend Walmart's marketing efforts and their data-driven strategies (Singh et al., 2021). Apache data science tools are software solutions that provide a range of functionalities for data analysis, manipulation, and visualization. They are often used in data science projects to help derive insights from data and enable data-driven decision making. The abovementioned tool is used to summarize and interpret data in a way that helps understand its characteristics, patterns, and trends. The tools facilitated the answering of questions such as what happened, when, where, and how. For example, the Apache Hadoop ecosystem provides a distributed computing framework for storing, processing, and analysing massive datasets (Singh et al., 2017). The Apache tool is used to perform exploratory data analysis, data cleaning, and data pre-processing tasks. On the other hand, the Apache Spark framework provides an interface for data analysis using distributed computing, which can be used to perform data transformations, filtering, aggregations, and statistical analysis (Singh et al., 2017). Lastly, the Apache Zeppelin notebook provides an interactive environment for data analysis and visualization (Cheng et.al 2018).

ProProject (2023), explained Walmart's big data processes multiple terabytes of new data and historical data every day by collecting 2.5 petabytes of unorganised data from 1 million customers per hour (ProProject, 2023). The company uses data analytics to optimize the shopping experience for customers in its stores, on its website, and on mobile devices. Walmart also used Hadoop and NOSQL technologies to provide real-time data analysis and centralized access to data. Furthermore, the company uses descriptive statistics to track the effectiveness of its e- commerce strategy, and its big data analysts have identified a 10% - 15% increase in online sales for USD 1 billion in incremental revenue. Walmart has developed applications, such as Savings Catcher and eReceipts, that leverage Hadoop data to improve customer experiences. The company also developed Mupd8, a real-time stream processing platform, to handle issues like performance and scalability (ProProject, 2023).

The company's big data team analyses every clickable action on the Walmart website, in-store and online purchases, social media posts, local events and more to access information on consumers (ProProject, 2023). Furthermore, Walmart’s descriptive analytics, which tracks and monitors individual and average consumer purchases, gathers exhaustive customer data. Subsequently, this information is used in predictive analysis by providing customized recommendations based on the consumer’s purchase history and trending goods which the consumer may be interested in, based on their purchase history (ProProject, 2023).

# Challenges

Webb (2020) explained that one of descriptive analysis’s challenges is biased sampling, which occurs when the data collected is not a representative of the population being studied, leading to inaccurate or misleading results. Another challenge is missing data, which can create incomplete or skewed results if not addressed using appropriate techniques like imputation. Outliers can also skew results if not identified and addressed appropriately. Confounding variables, which can influence the relationship between two variables being studied, must be accounted for to avoid inaccurate conclusions. Additionally, analysts must be careful when interpreting the results of descriptive analysis, ensuring that they have a clear understanding of the context and purpose of the analysis. Finally, there is a risk of overreliance on summary statistics, which may be misleading if not used appropriately or in combination with other techniques (Wang, 2013).

# Legislation on ethics and data privacy

Walmart follows its Privacy Notice which serves as a comprehensive guide for customers, ensuring transparency and informing them about their rights and choices regarding their personal information (Walmart(b), nd). Walmart collects various categories of personal information, which has been previously mentioned. This information is collected through various channels, such as direct input from customers, device-associated data, in-store technology, external third-party sources, and other sources of information. The company also shares personal information with other businesses, service providers and co-branded partners, in compliance with legal requirements.

In South Africa Walmart operates as Massmart shops (Walmart(c), nd). The first policy that Walmart has to adhere to is the Consumer Protect Act (CPA) Right to privacy: The CPA protects consumers' personal information from unauthorized use or disclosure by businesses. CPA obligates shops to handle customer data responsibly and obtain consent when collecting or using personal information. The main governing policy on data protection and privacy is the Protection of Personal Information Act (POPIA). The Act regulates the collection, processing, storage, and dissemination of personal information by entities including Walmart. POPIA establishes the conditions under which personal information can be lawfully processed, including for analysis purposes, and requires organizations to obtain consent from individuals before collecting and using their data. Furthermore, the South African government has developed a National Data Protection Framework that provides guidance on best practices for data protection and privacy. The framework outlines principles and standards for the responsible collection, processing, and use of personal information, emphasizing the need for transparency, purpose limitation, and

security measures. Online collection of data is further guided by the Electronic Communications and Transactions Act, 2002 (ECT Act), which addresses various aspects of electronic transactions and communications in South Africa. It includes provisions related to the integrity and accuracy of information, prohibiting the provision of false, misleading, or deceptive information.

Moreover, Sello et.al 2019, mentioned that to protect privacy during the data process, personal identifiers should be removed or de-identified from data ensuring that individuals cannot be readily identified from the data and sometimes there could regulations could require that only authorized individuals, have access to sensitive data.

# Conclusion

Walmart is a leader in the retail industry and operates in many countries worldwide. Walmart has to remain competitive and understand its business trends and customer needs in order to manage resources wisely. With the use of big data analytics, Walmart can analyze complex datasets and gain valuable insights into customer behaviour, trends, and preferences in line with the countries privacy policies and ethics. Descriptive data analysis is one of the methods used by Walmart to analyze data and provide a clear picture of what happened in the past. Through data analysis, Walmart is improving its marketing and product development strategies and manage its resources wisely to maximize returns.

# Diagnostic analysis: Cork company in Portugal Background of Cork History and its challenge

Silva et al., (2020), conducted a study on absenteeism in a Cork manufacturing factory in Portugal with 350 staff. This study took place at a cork stoppers producer. The cork company has a problem of high daily absenteeism rates which result in a drop in a company’s productivity and hence affecting revenue generation, costing the cork company EUR 1 200 000 per year. The objectives of the study were to find out what causes absenteeism of employers and identify the most sectors at the company which have the highest percentage of absenteeism, quantify its impact on related outcomes and identify issues of absenteeism causes.

# What is diagnostic analysis?

Diagnostic analytics seeks to provide answers to the question of "why did this happen?" Diagnostic analytics enables companies to gain a deeper understanding of the underlying causes behind patterns observed in their data such as high rate of absenteeism. Diagnostic analysis is made up of (1) data drilling – analysis data that is driving the observed trends, (2) data mining - process of uncovering hidden patterns, trends, and insights from large datasets using statistical and computational methods and (2) correctional analysis -examining relationships between or amongst different variables (Holiday, 2021). Diagnostic analytics helps companies such as the cork company to uncover the reasons behind events or patterns in their data, allowing them to take appropriate action (Silva et al., 2020). Diagnostic analytics is using data to determine the

causes of trends and relationships between variables. It is a logical next step after using descriptive analytics to identify trends. This type of analysis can be done using an algorithm, or a statistical software such as excel (Silva et al., 2020).

Furthermore, Cote (2021) explains that diagnostic analysis involves examining the assumptions and characteristics of the data to assess the quality of data and detect any potential sources of error or bias. Diagnostic analytics helps in identifying and addressing potential problems that may affect the accuracy and reliability of the results generated by the data or model. This type of analysis is a valuable approach that can build on descriptive analysis to gain insights into understanding the underlying causes of a specific outcome or result.

Diagnostic analysis also involves using various techniques and methods to assess the performance and accuracy of statistical models. This type of analysis is the process of using data to determine the causes of trends and correlations between variables. It can be used to understand why there is a relationship or no relationship between variables and to determine whether a hypothesis is true or not. Lastly, regression analysis is a form of diagnostic analysis that aims to identify and explain a relationship between variables (Cote, 2021).

# How does the Cork Company use descriptive analytics?

In a study conducted by (Silva et al., 2020), two types of absence metrics were used, which included: time lost not being at work and frequency of not being present at work (absence frequency). The time lost measurement shows absenteeism as a sum of units of time (e.g., hours or days) away from work, absence frequency, on the other hand, is the number of absences in a specific period irrespective of duration.

The study also used the absenteeism rate, which was calculated by summing up the total hours away from work and dividing the result by the sum of total hours available to work (total number of workers X number of each employee’s working hours). The first step was to analyze the company's overall absenteeism rate and compare it with the national average. The company’s absenteeism rate was 11% higher than the average national absenteeism rate in Portugal, in 2018. Then, the critical sectors – the ones with the highest absenteeism rate –, were identified along with the absenteeism causes for these sectors. The key cause of absenteeism was identified to be medical conditions caused by Musculo- skeletal (contributing to 47% absenteeism), breathing problems (contributing to 39% of the absenteeism rate and personal and family related issues (contributing to 13% of absenteeism). Finally, costs related to absenteeism

were calculated considering several factors, such as non-quality costs and productivity. The

company's annual cost incurred as a result of absenteeism was calculated to be EUR 1 200 000 per annum.

# Data analytics approaches and methods.

In the diagnostic study conducted by (Silva et al., 2020) the methodology which was employed was quantitative research, which involves a collaborative approach between an action researcher and a client to diagnose a problem and develop a solution based on the results from the diagnostic analysis.

The research was conducted within a large cork company comprising 350 employees in Portugal. Absenteeism was measured by the amount of time lost, representing the total number of hours away from work. Additionally, the study utilized the absenteeism rate, which was determined by summing up the total hours away from work and dividing it by the total hours available for work (obtained by multiplying the number of workers by the working hours of each employee).

The initial step involved analyzing the company's overall absenteeism rate and comparing it to the national average. Subsequently, the critical sectors characterized by the highest absenteeism rates, were identified, and the causes of absenteeism within these sectors were determined. Finally, the study computed the costs associated with absenteeism, considering various factors such as non-quality costs and productivity.

# Challenges

(Holiday, 2021), outlined that diagnostics analysis has limited analysis on cause and effect, the correlational identified through dialogistic analysis may need to be further investigated. This can lead to mistake correlation for causation, which can lead to addressing wrong root causes of the problem. Data analysts may need to rely on additional sources, such as third-party historical and real-time data, to supplement and make their analysis comprehensive.

# Legislation on ethics and data privacy

The study was conducted by the University of Aveiro, in Portugal. The university has an ethics council made up of the Ethics and Deontology Committee which is the he advisory and support body for the governing boards in matters of ethic which guides the collection and interpretation of data (University of Aveiro, nd). The committee’s aim is to promote reflection and contribute to the setting-up and implementation of appropriate directives for the establishment and consolidation of a policy of safeguarding ethical principles, by issuing guidelines or proposing its own codes of conduct. Furthermore, the policy states that researchers must obtain informed consent from individuals before collecting and using their data. Consent should be voluntary, specific, and based on adequate information provided to the individuals(University of Aveiro, nd).

The second partner university in this study, the University of Twente has adopted a university research ethics policy. This policy states that an ethical review of research will be conducted by a committee. For this study the Ethics Committee for Natural Sciences and Engineering Sciences (NES) played a crucial role in overseeing and promoting ethical standards in research. The NES also verified if informed consent was received during the study. The committee ensures that participants in a study are informed beforehand about all aspects of research to enable the participant to make an informed decision on their willingness to participate in the research and for their data to be collected (University of Twente, nd).

Since the universities involved in the study are based in Europe the General Data Protection Regulation (GDPR) applies. The GDPR emphasizes obtaining explicit consent, ensuring data security, and Informed Consent. Consent is one of the six legal grounds for lawful processing of

personal data, often consent is the most appropriate legal basis in research based on the GDPR. The consent requires action: it needs to be given by a written/oral statement or by clear affirmative action. Inactivity and pre-ticked boxes are not valid consent.

In addition, in South Africa, the POPIA applies as well as the Common Law. The South African common law principles related to fraud and misrepresentation apply to the misinterpretation of data. The application of the Common Law ensures that researchers handle data and insights derived from diagnostic analytics accurately. It emphasizes that misrepresenting data with the intention to deceive others can be deemed fraudulent, potentially resulting in legal repercussions in South Africa.

# Conclusion

The study on absenteeism in a Cork manufacturing company revealed a significant challenge of high absenteeism rates, resulting in decreased productivity and substantial financial losses. Through the application of diagnostic analytics, researchers were able to identify the underlying causes of absenteeism, including musculoskeletal and breathing problems, as well as personal and family-related issues. Diagnostics analytics provided a broader understanding of the overall absenteeism rate and enabled the calculation of the financial costs incurred by the company. However, it is important to note that diagnostic analysis has limitations in establishing causality and further investigation may be required.

In terms of ethics and data privacy, the study adhered to strict guidelines and regulations. Universities involved in the research maintained ethics committees to ensure informed consent was obtained from participants. Compliance with legal frameworks, such as the GDPR in Europe and the POPIA in South Africa, safeguarded data privacy and protected against fraudulent misrepresentation.

# Predictive analysis: Netflix Background on Netflix

Netflix was found in 1997 by Reed Hastings and Marc Randolph, it is now available in more than

180 countries. Netflix started by producing Digital Versatile disc, commonly known as DVDs. As technology and consumer demands shifted, Netflix started streaming movie service and one of the leading internet bandwidth consumers (Kariuki, 2023).

Netflix is one of the major players in the streaming business, which faces several challenges in the industry. One of the challenges Netflix is facing is the competition from other streaming services such as Amazon Prime video, Showmax and many others, which are seeking to differentiate themselves and attract more customers based on pricing, technological features and creation of original content. According to a study conducted by (Lengyel 2021), Netflix's shift in strategic course of action to personalized content has led to an increase in the use of the Netflix

streaming service, the characteristics of Netflix recommendation system led to increased usage and utility.

# What is predictive analysis?

Predictive analysis has helped companies to understand customers and is also used to identify risks and opportunities and guides in the decision-making process. Predictive analytics analyses current and historical data and use statistical, data mining and machine learning techniques to predict events of future occurrence. Furthermore, predictive modelling and classification analysis are one of the two key forms of predictive analysis (Mishra & Silakar, 2012). Mishra & Silakar (2012) breaks down predictive analysis into:

* Collection and cleaning of data: Collecting data from various sources, such as databases, files, or sensors and cleaning of data, which includes removing missing values, dealing with outliers, and normalizing the data.
* Transforming data: Selecting and creating features that are most relevant to the problem being addressed by the data.
* Creating the learning model: Training a learning model which could include regression, decision trees, neural networks, and support vector machines. The developed model is trained using part of the of the transformed data.
* Model testing: Testing model using some of the transformed data to improve the model’s accuracy.
* Using the model for predictions: After the model is trained, it can be used to make predictions on new data. E.g., the predictions are delivered as recommendations of movies that a Netflix user are most likely to watch and enjoy.

# How does Netflix use predictive analysis?

Netflix uses predictive analysis for multiple reasons, which includes marketing, maximising sales and returns, understanding customer behaviour and improve user experience which result in retaining higher number of users.

Netflix collects user data such as; user demographics, viewing history, search history, ratings, preferences, time spent watching movies, movie ratings, genres of movies watched and genres of movies watched half way. The collected data together with algorithms is used for predictive analysis, to curate a personal experience for each user on Netflix and understand consumer behaviour and consumer patterns based on the user’s historical monitored activities on Netflix (Team, 2013). The accuracy of Netflix predictive analysis is accurate for 75% of Netflix users. The movies recommended for users based on predictive analysis are 80% accurate on what the user might be interested in (Team 2013). Team 2013 also explains that Netflix collects data on their movies, the collected data includes; number of plays, number of ratings, number of searches, etc. This information helps to identify movie genres that are preferred the most by users.

Netflix uses several algorithms in their predictive analysis to suggest movies to users. Some of the algorithms include the personalized video ranking, trending now ranker, and aesthetic visual analysis tool (Team, 2013). Mishra & Silakar, (2012), highlighted one of the tools that Netflix uses is a A/B split testing, which is used to test the effectiveness of different images and screenshots for each user. For example, if a user has watched movies with certain actors/actresses or genres, Netflix recommender systems recommends more movies of those actors or movie genre. The A/B split testing also assesses emotional responses to movies by analysing data on how users have reacted to certain types of movie content in the past. For example, if a user has clicked on movies with romantic scenes, Netflix may show images of the romantic scenes to convey similar interest. On the other hand, if a user has watched more action scenes, Netflix may focus on images based on those action scenes. Figure 2 shows the different images of the “Good Will Hunting” movie that is shown to various users based on how they interacted with previous movies – showing interest through watching movies with certain cover photos -



Figure 1: Effects of the A/B split testing on the “Good Will Hunting” movie (Mishra & Silakar, 2012)

# Challenges

* Predictive analysis relies heavily on the quality and completeness of data. If the data used in the analysis is incomplete or inaccurate, the predictions generated will most likely not be accurate. Overfitting and biasness will occur when a predictive model is trained too closely to the data used to create it, making it less effective at making predictions on new data. Incorrect recommendations can reduce user’s usage of the Netflix platform (Lengyel, 2021).
* Insightsoftware states that predictive models can be complex and difficult to understand, deploy, use, scale and update.
* Predictive analysis can raise ethical concerns around privacy, discrimination, and accountability. For example, a model can be used to make decisions about individuals which are not fair and can discriminate (Mühlhoff, 2021).
* Predictive analytics tools provide recommendation information and insights, but they fail to provide the user with the next actions or follow up actions (Infosight, 2021).

# Legislation on ethics and data privacy

Hosanagar and Werbach (2019), explained that Netflix faced a lawsuit, in 2009, for sharing anonymized movie ratings data of around half a million subscribers as part of their Netflix Prize contest. Despite efforts to protect user privacy by Netflix, researchers from the University of Texas demonstrated that it was possible to identify individuals with as few as six movie ratings. However, Netflix does have a privacy policy they follow called Privacy Statement and Terms of Use. The Privacy Statement and Terms of Use promotes confidentiality of user information and ensures compliance with applicable laws and regulations and aligns Netflix with the South African POPIA.

Lengyel (2021), highlights the need for Netflix to consider user-centric key-performance indicators that should be inherent in recommendation system features. As users are becoming increasingly aware that companies are collecting data on them with or without their explicit consent, trust in systems develops into a central issue and source of customer loyalty towards trustworthy systems. System characteristics such as transparency and curation are aimed at building trust and user friendliness. The increasing customer awareness of private user data being used without the conscious consent of the user or by being collected from other sources also increases the scepticism and mistrust in Netflix. Therefore, if Netflix manages to establish trust- building mechanisms through increased transparency, the company stands to gain and keep users, potentially at the expense of competitors. In South Africa the same legislations mentioned in section 1 and 2 apply, which includes the (POPIA, to guide the collection of personal data and common law to ensure that Netflix does not mispresent that data. According to Netflix’s Privacy Statement and Terms of Use, Netflix's privacy and data protection policies prioritize user privacy, security, and control, ensuring a trustworthy and enjoyable streaming experience for its global user base.

# Conclusion

Netflix faces challenges in the industry due to the competition from other streaming services. Predictive analytics, specifically the recommendation system, improves user experience and provides users with personalized recommendations which could increase user usage of the platform that gives Netflix a competitive advantage in the streaming industry. To ensure that Netflix does not break the privacy laws when using the predictive analytics, Netflix applies its Privacy Statement and Terms of Use.

# Prescriptive analysis: The Coca Cola Company Background of The Coca Cola Company

The Coca Cola Company produces carbonated soft drinks and juices. Coca Cola was founded by Dr. John Pemberton in Atlanta United States of America in 1886 (The Coca Cola Company. n.d). It is now sold worldwide. One of the key challenges that Coca cola is experiencing is coming up with new soft drink favours that meets the market demands. As a result, Coca Cola has been using prescriptive analysis to understand the needs of the consumers and come up with new drink favours (Mixson, 2022).

# What is prescriptive analysis?

Prescriptive analysis is data analysis that uses optimization algorithms and machine learning models to provide recommendations on what actions should be taken based on the insights derived from the data analysed (Hansel, 2018). Prescriptive analysis goes beyond descriptive and predictive analysis by not only providing information on what happened (descriptive analysis) and what might happen (predictive analysis) but also suggesting what should be done. It answers the question “What should happen?” (Hansel, 2018). Prescriptive analysis is used in a wide range of applications such as supply chain management, finance, healthcare, and marketing. In the case of Coca Cola, it is used to optimize marketing campaigns and satisfy customer needs by recommending new drink flavours (Mixson, 2022).

Stevens (2021) further explains that prescriptive analysis is a data analysis process that goes step further from other data analysis processes by recommending the best actions to achieve desired outcomes. Prescriptive analysis uses machine learning together with mathematical models, algorithms, and optimization techniques to evaluates various potential actions and considers factors such as constraints, objectives, and preferences to draw valuable insights and practical recommendations for decision-makers, effectively allocating resources, streamlining operations, and enhancing overall performance. When applying prescriptive analysis it is important to understand the desired outcome, identifying key variables, and defining any constraints or limitations.

Stevens (2021) further adds that prescriptive analytics involves analysing historical data, real- time data, external data, and qualitative information potential future outcomes to determine and recommend the optimal actions to be taken. Its purpose is to guide decision-making by identifying strategies to avoid future problems, maximize benefits and leverage emerging trends. Prescriptive analysis utilizes algorithms, machine learning, statistical methods, predictive analysis and computational modelling to analyse data and make recommendations. It considers various decision patterns and potential scenarios that decision makers may adopt in future. The generated scenarios enable decision makers to measure the effects of specific decisions and select the most favourable course of action based on a comprehensive analysis of all possible scenarios and potential outcomes.

# How does Coca Cola use prescriptive analysis?

In the beverage industry prescriptive analytics analyses what is likely to happen and it also provides instructions or suggested solutions. Coca-Cola has been using prescriptive analytics in its business to leverage big data and artificial intelligence to better understand consumer trends and preferences. One of the problems that prescriptive data analysis has addressed is the need for Coca-Cola to understand and track the evolving taste of its customers in different regions and introduce a product offerings that meets the preference of the society ([Marr](https://www.forbes.com/sites/bernardmarr/), 2017). Coca Cola has invested extensively in research and development to better leverage the data it collects from customers around the world. Marr (2017) explains that Coca-Cola has used data analytics to inform product development and introduce new flavours based on customer preferences. For example, The Coca-Cola company launched Cherry Sprite after discovering that there was a high demand for the flavour through the insights gained from its freestyle vending machines – which

allowed consumers to combine flavour of drinks. Data on the cool drink flavours was collected from the cool drink fountain machine, while consumer were pouring and mixing them their drinks. Artificial Intelligence (Al) Algorithms were used to analysis the data from the fountain machine and develop a new drink flavour. Furthermore, Coca-Cola uses AI algorithms to gain insights into customer preferences regarding the time, place, and way the consumer prefers their products. This information us used to provide the consumers with their preferred products in sufficient quantities (AI of Things, 2021.)

Another example where Coca Cola and other beverage companies use prescriptive analytics to provide specific solutions, is the orange juice example explained by (Marr, 2017). Coca-Cola uses data such as weather, crop yields, pricing, and taste, combined with satellite imagery to inform an algorithm that matches Coca Cola’s products to customer tastes and to inform when and how they source oranges which are key input for juices.

# Challenges

1. Prescriptive analysis requires large amounts of data and advanced analytical techniques, which can make it difficult for organizations to develop and implement due to lack of skills and data being sourced from various sources. Furthermore, prescriptive analysis can be complex and difficult to understand, which can make it challenging for stakeholders to interpret and act on the results (Ransbotham, 2015).
2. Implementing the recommendations and suggestions generated by prescriptive analysis can be challenging, as it may require significant changes to existing business processes and systems. (Lepenioti, 2020).
3. Coca Cola is a company that operates based on hierarchy which can introduce duplicate innovative systems based on predictive and prescriptive analysis (Ransbotham, 2015). Prescriptive analysis can result in the implementation of duplicate and redundant systems within a hierarchical organization.

# Legislation on ethics and data privacy

Coca-Cola's privacy policy applies to their consumers, who are using their websites, mobile applications or providing personal information. The company collects various categories of personal information from different sources for different purpose indicating use for prescriptive analysis. As per the policy users must have the ability to permit Coca-Cola to use that information according to the privacy policy. However, for the vending machine data collection no consent is required because personal information is not collected.

Coca-Cola also collects information that is passively obtained when users interact with their websites. This includes information recorded by their system logs, such as web requests, IP addresses, device and browser information. If the consumer accepts cookies while they are on the Coca Cola website, the consumer gives consent for their information to be collected.

Coca-Cola acknowledges that in certain jurisdictions, including certain states in the USA and South Africa, the act of disclosing personal information to third parties, which may include

distribution partners, advertising agencies, and analytics providers, could be considered as either a sale or sharing of personal information. In South Africa the POPIA allows the selling of personal data. However as per the Electronic Communication and Transactions Act (ECT Act). Users have the right to opt-out of such disclosures and sales by enabling opt-out preferences on their browser or by opting out of cookies through Coca-Cola's cookie preference center. This is well aligned to the ECT Act. The privacy policy also addresses the use of Do Not Track (DNT) signals and Global Privacy Control (GPC). Coca-Cola handles user information in accordance with its privacy policy and does not specifically respond to DNT signals. GPC, on the other hand, is a setting users can enable in their web browser to communicate their privacy preferences. Coca-Cola’s privacy policy is well aligned with the POPIA.

# Conclusion

Coca-Cola uses prescriptive analysis to come up with new drink flavours that meets the consumer needs. By using prescriptive analytics, the company is better equipped to understand consumer trends and preferences, introduce new flavours, and remain relevant in the future. However, implementing prescriptive analysis can present challenges such as the need for advanced analytical techniques, changes to existing business processes, and the potential for duplicate systems within hierarchical organizations. Nonetheless, Coca Cola has demonstrated a commitment to privacy and data protection by adhering to legislation such as the POPIA and implementing measures to obtain user consent and provide opt-out preferences. Overall, prescriptive analysis has proven to be a valuable tool for Coca Cola in driving innovation and meeting customer needs in the highly competitive beverage industry.

# References

* 1. AI of Things. 2021. *Telefonica. Coca-Cola’s use of AI to stay at the top of the drinks market.* Blog. Available from: https://business.blogthinkbig.com/coca-colas-use-of-ai- to-stay-at-the-top-of-the-drinks-market/. [26 March 2023]
  2. Cheng, Y., Liu, F., Jing, S., Xu, W. & Chau, D. Building Big Data Processing and Visualization Pipeline through Apache Zeppelin. 2018. *ACM*, 1-7. <https://doi.org/10.1145/3219104.3229288>. [36 March 2023]
  3. Cote. C. 2021. 18 November 2021. *What is Diagnostic Analytics? 4 Examples.* Blog.

Available from: <https://online.hbs.edu/blog/post/diagnostic-analytics>. [24 May 2023].

* 1. Engati Team. 2013. Engati Team. April 2013. BLOG Netflix Predictive Analytics: Journey to 220Mn+ subscribers. Blog. Available: <https://www.engati.com/blog/predictiveanalytics>. [18 April 2023].
  2. Hansel, S. (2018). T Systems. February 2018. “*Big data and an orange juice, please!“: Prescriptive analytics is maturing into an effective, all-purpose weapon for better decisions*. https://[www.t-systems.com/de/en/newsroom/best-practice/02-2018-](http://www.t-systems.com/de/en/newsroom/best-practice/02-2018-) brilliantprospects/prescriptive-analytics-the-next-level-of-data-analytics. [26 April

2023]

* 1. Harper, J. 2019. AI Business. 26 November 2019. *The compelling case for descriptive analytics: sentiment analysis and natural language*: Blog. Available from https://aibusiness.com/nlp/the-compelling-case-for-descriptive-analytics- sentimentanalysis-and-natural-language. [20 April 2023].
  2. Harvard Business School. n.d. *A Beginner’s Guide to Data & Analytics*. Available from: https://online.hbs.edu/Documents/a-beginners-guide-to-data- andanalytics.pdf?\_ga=2.171073127.615557383.1681931258- 1643208751.1681931258. [17 April 2023].
  3. Holliday, M. (2021). Oracle Netsuite. 9 December 2019. *What Is Diagnostic Analytics?How It Works and Examples*. https://[www.netsuite.com/portal/resource/articles/datawarehouse/diagnostic-](http://www.netsuite.com/portal/resource/articles/datawarehouse/diagnostic-) analytics.shtml. [20 April 2023].
  4. Hosanagar, K and Werbach, K. 2019. 23 August 2019. *Knowledge at Warton. Is Data Privacy Real*? *Don’t Bet on It.* <https://knowledge.wharton.upenn.edu/article/data-privacy-real-dont-bet/>. [22 June

2023].

* 1. Insightsoftware. (2022). Insightsoftware. 17 January 2022. *The 4 Common Challenges of Predictive Analytics Solutions*. Blog. Available from: https://insightsoftware.com/blog/the-4-common-challenges-of-predictive- analyticssolutions/.[20 April 2023].
  2. KariukI, P. n.d. *How and When Did Netflix Start? A Brief History of the Company*. Available from: https://[www.makeuseof.com/how-when-netflix-start-brief-](http://www.makeuseof.com/how-when-netflix-start-brief-) companyhistory/:[19 April 2023].
  3. Lengyel, D. 2021. *Does the Netflix Recommender System Produce Customer Utility?: An Analysis of the Technology Acceptance of the Algorithmic-prediction- based Netflix* Recommender System and its Drivers. Master’s thesis, Universidade Católica Portuguesa. Available from: https://repositorio.ucp.pt/bitstream/10400.14/35694/1/202751350.pdf. [19 April

2023].

* 1. Lepenioti, K., Bousdekis, A., Apostolou, D. & Mentzas, G. 2020. Prescriptive analytics: Literature review and research challenges. *International Journal of Information Management*, 50:57-70. Available from: https://[www.researchgate.net/publication/338970341\_Prescriptive\_analytics\_Literatu](http://www.researchgate.net/publication/338970341_Prescriptive_analytics_Literatu) re\_review\_and\_research\_challenges. [29 April 2023].
  2. Marr, B. 2017. Forbes. 18 September 2017. *The Amazing Ways Coca Cola Uses Artificial Intelligence And Big Data To Drive Success:* Blog. Available from: https://[www.forbes.com/sites/bernardmarr/2017/09/18/the-amazing-ways-coca-](http://www.forbes.com/sites/bernardmarr/2017/09/18/the-amazing-ways-coca-) colauses-artificial-intelligence-ai-and-big-data-to-drive-success/?sh=40bcf28678d2. [25 April 2023].
  3. Mishra, N. and Silakari, S., 2012. Predictive analytics: a survey, trends, applications, opportunities & challenges. *International Journal of Computer Science and Information Technologies*, *3*(3), pp.4434-4438. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=5e6136dc033984b a2ad0222fa53eb69db11e60d2 [27 April 2023].
  4. Mixson, E. 2022. Data Analytics Network. 2 August 2022. *Coca-Cola vs. Pepsi: The Sweet Fight For Data-Driven Supremacy*: Blog. Available from https://[www.aidataanalytics.network/data-science-ai/articles/coca-cola-vs-pepsi-](http://www.aidataanalytics.network/data-science-ai/articles/coca-cola-vs-pepsi-) thesweet-fight-for-data- drivensupremacy#:~:text=Since%202018%2C%20Coca%2DCola%20has,consumer

%20behavior%20in%20incredible%20detail. [20 April 2023].

* 1. Projectpro. 2023. Projetpro. 24 April 2023. *How Big Data Analysis helped increase Walmarts Sales turnover*?. Blog. Available from: https://[www.projectpro.io/article/howbig-data-analysis-helped-increase-walmarts-](http://www.projectpro.io/article/howbig-data-analysis-helped-increase-walmarts-) sales-turnover/109 [14 April 2023].
  2. Ransbotham, S. 2015. *Coca-Cola’s Unique Challenge: Turning 250 Datasets Into One.* MIT Sloan Management Review. Available from: https://calipgca.org/images/Coca\_Cola\_Unique\_Challenge.pdf. [26 April 2023].
  3. Sello, P., Bagula, A., Ajayi, O. 2020. Laws and Regulations on Big Data Management: The Case of South Africa. In: Zitouni, R., Agueh, M., Houngue, P., Soude, H. (eds) e-Infrastructure and e-Services for Developing Countries. AFRICOMM 2019. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 311. *Springer*, Cham. <https://doi.org/10.1007/978-3-030-41593-8_12>. [22 June 2023].
  4. Silva, A., Ramos, A.L., Brito, M. and Ramos, A. 2020. A Diagnostic Analysis of Absenteeism—A Case Study in a Portuguese Cork Industry. *Occupational and Environmental Safety and Health II*. 793-799. Available: https://[www.researchgate.net/profile/MarleneBrito/publication/339395338\_A\_Diagno](http://www.researchgate.net/profile/MarleneBrito/publication/339395338_A_Diagno) stic\_Analysis\_of\_AbsenteeismA\_Case\_Study\_in\_a\_Portuguese\_Cork\_Industry/links/ 5ea7f04c458515ca49f29aae/ADiagnostic-Analysis-of-Absenteeism-A-Case-Study-in- a-Portuguese-Cork-Industry.pdf.[27 April 2023].
  5. Singh, M., Ghutla, B., Jnr, R., Mohammed, A. & Rashid, M. 2017. Walmart's Sales Data Analysis - A Big Data Analytics Perspective. Available: [http://repository.usp.ac.fj/10798/1/Article\_93.pdf.](http://repository.usp.ac.fj/10798/1/Article_93.pdf) [20 April 2023].
  6. Stevens E, 2023. Career Foundry. 11 May 2023. *The 4 Types of Data Analysis [Ultimate Guide].* Blog. Available from: [The 4 Types of Data Analysis [Ultimate Guide]](https://careerfoundry.com/en/blog/data-analytics/different-types-of-data-analysis/) [(careerfoundry.com)](https://careerfoundry.com/en/blog/data-analytics/different-types-of-data-analysis/). [11 May 2023]
  7. The Coca Cola Company. N.d. History. Available : https://[www.cocacolacompany.com/company/history.](http://www.cocacolacompany.com/company/history) [18 April 2023].
  8. University of Aveiro. nb. Ethics Council. Available: [https://www.ua.pt/en/ethics-](https://www.ua.pt/en/ethics-council) [council](https://www.ua.pt/en/ethics-council). [22 June 2023].
  9. University of Twente. nb. ET RESEARCH SUPPORT SERVICES. Available: [https://www.utwente.nl/en/service-portal/research-support/additional-information-per-](https://www.utwente.nl/en/service-portal/research-support/additional-information-per-faculty/engineering-technology-et/et-research-support-services) [faculty/engineering-technology-et/et-research-support-services](https://www.utwente.nl/en/service-portal/research-support/additional-information-per-faculty/engineering-technology-et/et-research-support-services). [23 June 2023].
  10. Walmart(a). n.d. *Walmart History* . Available: https://careers.walmart.com/history#:~:text=Walmart%20started%20with%20one%20 man,help%20more%20customers%20than%20ever. [3 April 2023].
  11. Walmart(b). n.d. Walmart Privacy Notice. Available: <https://corporate.walmart.com/privacy-security/walmart-privacy-notice>. [23 June

2023].

* 1. Walmart(c). n.d. Africa. Available: [https://corporate.walmart.com/our-story/our-](https://corporate.walmart.com/our-story/our-business/international/africa) [business/international/africa](https://corporate.walmart.com/our-story/our-business/international/africa) [23 June 2023].
  2. Wang, R. & Ware, J. H. 2013. Detecting Moderator Effects Using Subgroup Analyses. Prevention Science, 14: 111-120. Available: https://0-link- springercom.oasis.unisa.ac.za/article/10.1007/s11121-011-0221-x. [15 April 2023].
  3. Whitney, C. R. & Liu, J. 2017. What we’re missing: A descriptive analysis of part-day absenteeism in secondary school. *AERA Open,* 3**,** 2332858417703660.