

DATA ANALICTICS

¹S.Vigneshwaran - B.com CA

²G.Gokul - B.com CA

³J.Antony moses – B.com CA

Sri Krishna Adithya college of arts and science

Abstract:

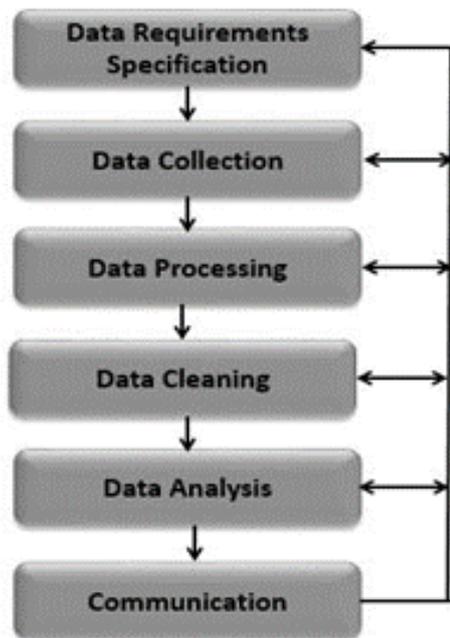
Data Analysis is a process of collecting, transforming, cleaning, and modeling data with the goal of discovering the required information. Data analysis is evaluating data using analytical and statistical tools to discover useful information and aid in business decision making. Data analysis is a process of inspecting, cleansing, transforming and modeling data with the goal of discovering useful information, informing conclusions and supporting decision-making. Data analysis is a part of a larger process of deriving business intelligence. Data Analysis Process is nothing but gathering information by using proper application or tool which allows you to explore the data and find a pattern in it. Once the data is collected, cleaned, and processed, it is ready for Analysis.

1. Introduction:

Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data. Data Analysis is the process of systematically applying statistical and logical techniques to describe and illustrate, condense and recap, and evaluate data. He process of evaluating data using analytical and logical reasoning to examine each component of the data provided. There are a variety of specific data analysis method, some of which include data mining, text analytics, business intelligence, and data visualizations. The purpose of Data Analysis is to extract useful information from data and taking the decision based upon the data analysis. The process of Inspecting, cleaning, transforming, modelling data with the objective of discovering useful information. Models and process for data analysis.

The terms Data Modeling and Data Analysis mean the same. Data visualization is at times used to portray the data for the ease of discovering the useful patterns in the data. The results so obtained are communicated, suggesting conclusions, and supporting decision-making.

2. The Process consists of the following phases that are iterative in nature :



2.1 Data Requirements Specification:

- The data required for analysis is based on a question or an experiment.
- Based on the requirements of those directing the analysis, the data necessary as inputs to the analysis is identified (e.g., Population of people).
- Data may be numerical or categorical.
- Specific variables regarding a population (e.g., Age and Income) may be specified and obtained.

2.2 Data Collection:

- Data is collected from various sources ranging from organizational databases to the information in web pages.
- Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes.
- Data Collection is the process of gathering information on targeted variables identified as data requirements.
- Data Collection ensures that data gathered is accurate such that the related decisions are valid.
- The emphasis is on ensuring accurate and honest collection of data.
- Data Collection provides both a baseline to measure and a target to improve.
- The data thus obtained, may not be structured and may contain irrelevant information.
- The collected data is required to be subjected to Data Processing and Data Cleaning.

2.3 Data Processing:

- The data that is collected must be processed or organized for analysis.
- Data processing occurs when data is collected and translated into usable information.
- Usually performed by a data scientist or team of data scientists, it is important for data processing to be done correctly as not to negatively affect the end product, or data output.
- Data processing starts with data in its raw form and converts it into a more readable format (graphs, documents, etc.), giving it the form and context necessary to be interpreted by computers and utilized by employees throughout an organization.
- This includes structuring the data as required for the relevant Analysis Tools.
- A Data Model might have to be created.

For example, the data might have to be placed into rows and columns in a table within a Statistical Application or Spreadsheet .

2.4 Data Cleaning:

- The processed and organized data may be incomplete, contain duplicates, or contain errors.
- Data cleansing or data cleaning is the process of detecting and correcting corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.
- There are several types of Data Cleaning that depend on the type of data.
- Data Cleaning is the process of preventing and correcting these errors.

For example, while cleaning the financial data, certain totals might be compared against reliable published numbers or defined thresholds. Likewise, quantitative data methods can be used for outlier detection that would be subsequently excluded in analysis.

2.5 Data Analysis:

- Data that is processed, organized and cleaned would be ready for the analysis.
- Data Visualization may also be used to examine the data in graphical format, to obtain additional insight regarding the messages within the data.
- Various data analysis techniques are available to understand, interpret, and derive conclusions based on the requirements.
- The process might require additional Data Cleaning or additional Data Collection, and hence these activities are iterative in nature.
- Statistical Data Models such as Correlation, Regression Analysis can be used to identify the relations among the data variables.
- These models that are descriptive of the data are helpful in simplifying analysis and communicate results.

2.6 Communication:

- The results of the data analysis are to be reported in a format as required by the users to support their decisions and further action.
- The data analysts can choose data visualization techniques, such as tables and charts, which help in communicating the message clearly and efficiently to the users.
- The feedback from the users might result in additional analysis.
- The analysis tools provide facility to highlight the required information with color codes and formatting in tables and charts.

3. Purpose Data Analysis :

To grow your business even to grow in your life, sometimes all you need to do is Analysis. If your business is not growing, then you have to look back and acknowledge your mistakes and make a plan again without repeating those mistakes. And even if your business is growing, then you have to look forward to making the business to grow more. All you need to do is analyze your business data and business processes.

4. Data Analysis Tools:



Data analysis tools make it easier for users to process and manipulate data, analyze the relationships and correlations between data sets, and it also helps to identify patterns and trends for interpretation. Here is a complete list of tools.

5. Types , Techniques and Methods of Data Analysis Tools :

There are several types of data analysis techniques that exist based on business and technology. The major types of data analysis are:

- Text Analysis
- Statistical Analysis
- Diagnostic Analysis
- Predictive Analysis
- Prescriptive Analysis

5.1 Text Analysis:

- Text analysis is the automated process of understanding unstructured text data and making it easier to manage.
- Text analysis is a classic example of machine learning, and is widely used to gain valuable insights from social media comments, survey responses, and product reviews.
- Text Analysis is also referred to as Data Mining.
- It is a method to discover a pattern in large data sets using databases or data mining tools.
- It used to transform raw data into business information.
- Business Intelligence tools are present in the market which is used to take strategic business decisions.
- Overall it offers a way to extract and examine data and deriving patterns and finally interpretation of the data.

5.2 Statistical Analysis:

- Statistical Analysis shows "What happen?" by using past data in the form of dashboards.
- Statistical analysis is a component of data analytics.
- In the context of business intelligence (BI), statistical analysis involves collecting and scrutinizing every data sample in a set of items from which samples can be drawn. A sample, in statistics, is a representative selection drawn from a total population.
- Statistical Analysis includes collection, Analysis, interpretation, presentation, and modeling of data.
- It analyses a set of data or a sample of data. There are two categories of this type of Analysis - Descriptive Analysis and Inferential Analysis.

Descriptive Analysis: Analyses complete data or a sample of summarized numerical data.

- It shows mean and deviation for continuous data whereas percentage and frequency for categorical data.

Inferential Analysis: Analyses sample from complete data. In this type of Analysis, you can find different conclusions from the same data by selecting different samples.

5.3 Diagnostic Analysis:

- Diagnostic Analysis shows "Why did it happen?" by finding the cause from the insight found in Statistical Analysis.
- This Analysis is useful to identify behavior patterns of data.
- If a new problem arrives in your business process, then you can look into this Analysis to find similar patterns of that problem.
- And it may have chances to use similar prescriptions for the new problems.

5.4 Predictive Analysis :

- Predictive Analysis shows "what is likely to happen" by using previous data.
- Predictive analytics is the branch of the advanced analytics which is used to make predictions about unknown future events.
- Predictive analytics uses many techniques from data mining, statistics, modeling, machine learning, and artificial intelligence to analyze current data to make predictions about future.
- The simplest example is like if last year I bought two dresses based on my savings and if this year my salary is increasing double then I can buy four dresses.
- But of course it's not easy like this because you have to think about other circumstances like chances of prices of clothes is increased this year or maybe instead of dresses you want to buy a new bike, or you need to buy a house!

So here, this Analysis makes predictions about future outcomes based on current or past data. Forecasting is just an estimate. Its accuracy is based on how much detailed information you have and how much you dig in it.

5.5 Prescriptive Analysis :

- Prescriptive Analysis combines the insight from all previous Analysis to determine which action to take in a current problem or decision.
- Prescriptive Analytics is the area of data analytics that focuses on finding the best course of action in a scenario given the available data.
- It's related to both descriptive analytics and predictive analytics but emphasizes actionable insights instead of data monitoring.
- Most data-driven companies are utilizing Prescriptive Analysis because predictive and descriptive Analysis are not enough to improve data performance.
- Based on current situations and problems, they analyze the data and make decisions.

6. Data Analysis Process:

Data Analysis Process is nothing but gathering information by using proper application or tool which allows you to explore the data and find a pattern in it. Based on that, you can take decisions, or you can get ultimate conclusions. Data Analysis consists of the following phases:

- Data Requirement Gathering
- Data Collection
- Data Cleaning
- Data Analysis

- Data Interpretation
- Data Visualization

6.1 Data Requirement Gathering:

- First of all, you have to think about why do you want to do this data analysis? .
- All you need to find out the purpose or aim of doing the Analysis.
- You have to decide which type of data analysis you wanted to do!.
- In this phase, you have to decide what to analyze and how to measure it, you have to understand why you are investigating and what measures you have to use to do this Analysis.

6.2 Data Collection:

- After requirement gathering, you will get a clear idea about what things you have to measure and what should be your findings.
- Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes.
- Now it's time to collect your data based on requirements.
- Once you collect your data, remember that the collected data must be processed or organized for Analysis.
- As you collected data from various sources, you must have to keep a log with a collection date and source of the data.

6.3 Data Cleaning:

- Now whatever data is collected may not be useful or irrelevant to your aim of Analysis, hence it should be cleaned.
- Data cleaning is the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.
- The data which is collected may contain duplicate records, white spaces or errors.
- The data should be cleaned and error free.
- This phase must be done before Analysis because based on data cleaning, your output of Analysis will be closer to your expected outcome.

6.4 Data Analysis:

- As you manipulate data, you may find you have the exact information you need, or you might need to collect more data.
- During this phase, you can use data analysis tools and software which will help you to understand, interpret, and derive conclusions based on the requirements.

6.5 Data Interpretation:

- After analyzing your data, it's finally time to interpret your results.
- Data Interpretation is the process of making sense out of a collection of data that has been processed.
- This collection may be present in various forms like bar graphs, line charts and tabular forms and other similar forms and hence needs an interpretation of some kind.
- You can choose the way to express or communicate your data analysis either you can use simply in words or maybe a table or chart.
- Then use the results of your data analysis process to decide your best course of action.

6.6 Data Visualization

- Data visualization is very common in your day to day life; they often appear in the form of charts and graphs.
- Data visualization is the graphical representation of information and data.
- By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.
- In other words, data shown graphically so that it will be easier for the human brain to understand and process it.
- Data visualization often used to discover unknown facts and trends.
- By observing relationships and comparing datasets, you can find a way to find out meaningful information.

7. Perfomance of data analysis :

- Data analysis is a part of a larger process of deriving business intelligence. The process includes one or more of the following steps:

Defining Objectives: Any study must begin with a set of clearly defined business objectives. Much of the decisions made in the rest of the process depends on how clearly the objectives of the study have been stated.

- **Posing Questions:** An attempt is made to ask a question in the problem domain. For example, do red sports cars get into accidents more often than others?
- **Data Collection:** Data relevant to the question must be collected from the appropriate sources. In the example above, data might be collected from a variety of sources including: DMV or police accident reports, insurance claims and hospitalization details. When data is being collected using surveys, a questionnaire to be presented to the subjects is needed. The questions should be appropriately modeled for the statistical method being used.
- **Data Wrangling:** Raw data may be collected in several different formats. The collected data must be cleaned and converted so that data analysis tools can import it. For our example, we may receive DMV accident reports as text files, insurance claims from a relational database and hospitalization details as an API. The data analyst must aggregate these different forms of data and convert it into a form suitable for the analysis tools.
- **Data Analysis:** This is the step where the cleaned and aggregated data is imported into analysis tools. These tools allow you to explore the data, find patterns in it, and ask and answer what-if questions. This is the process by which sense is made of data gathered in research by proper application of statistical methods.
- **Drawing Conclusions and Making Predictions:** This is step where, after sufficient analysis, conclusions can drawn from the data and appropriate predictions can be made.

8.Conclusion:

Data analytics can breach customer privacy as information such as online transactions, purchases, or subscriptions, can be viewed by the parent companies. There are chances that the companies will exchange these databases for mutual benefits. Data analysis means a process of cleaning, transforming and modeling data to discover useful information for business decision-making. Now same thing analyst does for business purposes, is called Data Analysis. Types of Data Analysis are Text, Statistical, Diagnostic, Predictive, Prescriptive Analysis. Data Analysis consists of Data Requirement Gathering, Data Collection, Data Cleaning, Data Analysis, Data Interpretation, Data Visualization. The price of the tools normally depends on the features and applications that they can support. Moreover, some tools are complex and require proper training. Whenever we take any decision in our day-to-day life is by thinking about what happened last time or what will happen by choosing that particular decision. The price of the tools normally depends on the features and applications that they can support. Moreover, some tools are complex and require proper training. This is nothing but analyzing our past or future and making decisions based on it. The information that is obtained by making use

of data analytics can be misused. For that, we gather memories of our past or dreams of our future. One of the toughest jobs is to select the correct analytics tool. So that is nothing but data analysis. It is clear that when businesses bring data together, it is easier for them to get real-time insights about sales and finance, marketing, product development, and much more. Data allows the teams within a business to better collaborate, to achieve better results, and outsell rival companies.

9. Reference:

1. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=2ahUKEwjwsZf3s_nmAhXn63MBHdNCBIIQFjAAegQIBRAB&url=https%3A%2F%2Fen.wikipedia.org%2Fwiki%2FData_analysis&usg=AOvVaw0PtrPRyofPLTrDjKs32CtX
2. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=10&cad=rja&uact=8&ved=2ahUKEwjwsZf3s_nmAhXn63MBHdNCBIIQFjAJegQIBBAB&url=https%3A%2F%2Fwww.guru99.com%2Fwhat-is-data-analysis.html&usg=AOvVaw23jGSdIOQQZkXo9xU60RIO
3. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=11&cad=rja&uact=8&ved=2ahUKEwjwsZf3s_nmAhXn63MBHdNCBIIQFjAKegQIAhAB&url=https%3A%2F%2Fwww.predictiveanalyticstoday.com%2Fdata-analysis%2F&usg=AOvVaw3VVmosZsh21fvtjsEVROpK
4. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=13&cad=rja&uact=8&ved=2ahUKEwjwsZf3s_nmAhXn63MBHdNCBIIQFjAMegQIBxAB&url=https%3A%2F%2Fwww.makeuseof.com%2Ftag%2Fwhat-is-data-analysis%2F&usg=AOvVaw0CLDondwH-0ZgG0Qw6yIw2
5. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=12&cad=rja&uact=8&ved=2ahUKEwjwsZf3s_nmAhXn63MBHdNCBIIQFjALegQIBhAB&url=https%3A%2F%2Fwww.tutorialspoint.com%2Fexcel_data_analysis%2Fdata_analysis_process.htm&usg=AOvVaw2CBVd68G5SBkge4eagqui2
6. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=16&cad=rja&uact=8&ved=2ahUKEwjwsZf3s_nmAhXn63MBHdNCBIIQFjAPegQIChAB&url=https%3A%2F%2Fchartio.com%2Flearn%2Fdata-analytics%2Ftypes-of-data-analysis%2F&usg=AOvVaw22zzFS5P6k2SFv37W5hFls

