# A Research on Data Analytics

**Akhil Nashaboina**

Student of ACE Engineering College, Jawaharlal Nehru Technological University
Department of Electrical and Electronics Engineering, Hyderabad, Telangana, India

*Abstract*: In this contemporary world, with the presence of enormous data around us it is necessary to analyse the data for our benefits either for gathering hidden insights of a generating reports. These data analytics benefits the company by performing proper market analysis and improving the business requirements. So, in today's market this field has gained a lot of popularity because it lets us to gather generate reports perform market analysis and also improve business requirement. Data analytics refer to the techniques to analyse the data to strengthen the productivity and business gain.

## I. INTRODUCTION

Data analytics is a method of applying qualitative and quantitative techniques to analyse data, aiming for valuable insights. With the help of data analytics, we can explore data and we can even draw various conclusions about our data. This process can also include data collection, organization, pre-processing, transforming, modelling and interpretation of the data with the goal of discovering useful information suggesting conclusions and supporting decision making. Now the techniques and tools used to perform data analytics vary from organisation to organization.

## II. TYPES OF DATA ANALYTICS

The four main types of data analytics are **descriptive, diagnostic, predictive**, and **prescriptive.**

**1. Descriptive analytics:**

Descriptive analytics is a simple, entry-level type of analysis that looks at what has happened in the past. The two main techniques used in descriptive analytics are data aggregation and data mining. So, the data analyst first gathers the data and presents it in a shorter format (that's the aggregation part) and then 'mines' the data to explore patterns. Then the data is presented in a way that it can be easily understood by numerous audiences. It's significant is to note that descriptive analytics doesn't try to explain the historical data it's simply determining and describing the 'what'.

**2. Diagnostic analytics:**

While descriptive analytics looks at the 'what', diagnostic analytics inspect the 'why'. When running diagnostic analytics, data analysts will first seek to identify anomalies within the data that is, anything that cannot be explained by the data in front of them.

**3. Predictive analytics:**

Just as the name suggests, predictive analytics tries to predict what is likely to happen in the future. This is where data analysts start to come up with actionable, data-driven insights that the company can use to inform their next steps. Predictive analytics estimates the probability of a future outcome based on historical data, and while it can never be completely accurate, it does eliminate much of the conclusion from key business decisions.

**4. Prescriptive analytics:**

Prescriptive analytics shows us how we can take advantage of the outcomes that have been predicted. When conducting prescriptive analysis, data analysts will consider a range of possible sketch and guess the different actions the company might take.



**Fig -1:** Types of Data Analytics

## III. THE DATA ANALYSIS PROCESS

→ In the modern world, the power of data is a big game-changer for many businesses and organisations as it helps them to gain better insights make fast and better decisions and improve their products and services but the data by itself is just an information source. Unless we analyse and understand it fully, we will not be able to use it effectively.

→ Data analytics combines different processes to extract information from the data. The different steps involved in data analytics process are Defining our objective, collecting the data, clean the data, analyse the data, visualize and share your findings.
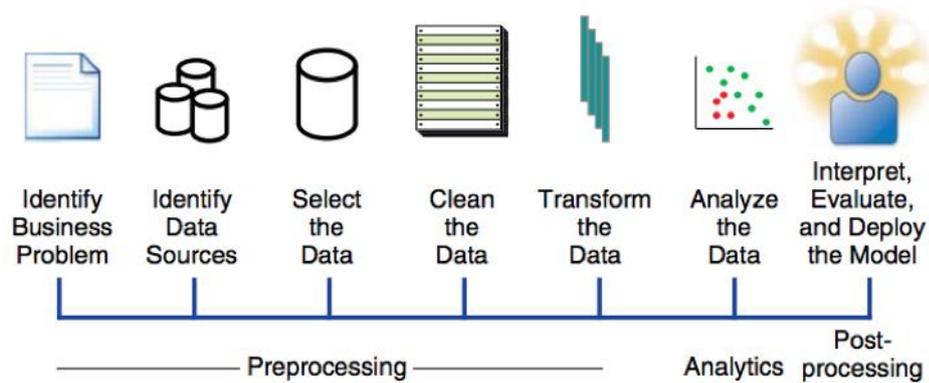
**Fig -2:** Overview of the Analytics Process Model

## STEP-1 DEFINING OUR OBJECTIVE:

first step in our data analysis process or any data analysis process is to define our objective in data analytics terms this is called the problem statement. Defining our objective means coming up with an assumption and figuring out how exactly to test it we can start by asking what business problem am I trying to solve now. For instance, an organisation's manager might pose a question such as why the last month business was quite low? Now as a data analyst we need to understand the business and the business's goals, and we need to work according to it.

## STEP-2 COLLECTING THE DATA:

Once we set up our objective, we'll need to create a strategy for collecting the relevant data a key part of this is determining which data we need this might be qualitative data or quantitative data. Qualitative data is nothing but datatype that consists of descriptive statements. And quantitative data is the one that can be measured and expressed numerically. Data collection should be done in three categories. First party data, second party data, third party data.

### 1. First party data:

First party data is data that we or our company has directly collected from customers. For example, it can be transactional tracking data or information from our customer relationship management system your CRM system whatever it sources first party data is usually collected in a clear and structured way other sources of first party data might include surveys, interviews or direct observation.

### 2. Second party data:

Second party data is simply the first party data of other organizations this might be available directly from the company or from private marketplace the main benefit of second party data is that it's usually structured and although it's less relevant than first party data it tends to be reliable examples of second party data include website app or social media activity like online purchase history or shipping data.

### 3. Third party data:

Third-party data involves that the data collected or distributed by a third-party source and has no direct connection to the consumer. Some of them uses a DMP (data management platform) to do this. DMP nothing but it is a tool where we can do first party data collection or organize third party data from any data source. In other words, buying the data from the third-party sources.
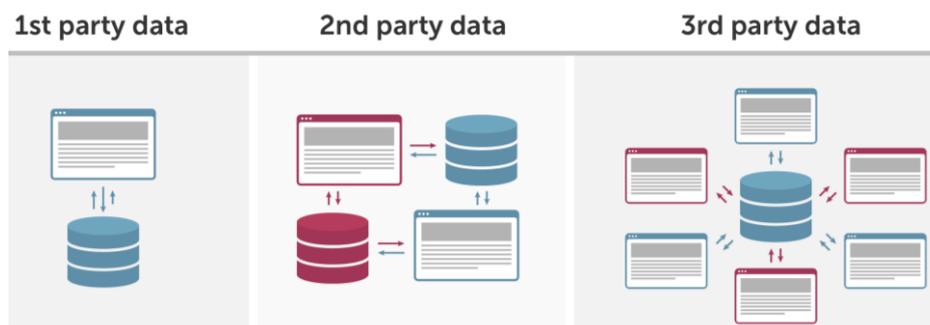


**Fig -3:** Data collection

## STEP-3 CLEANING THE DATA:

The third step is getting data ready for analysis this means cleaning or scrubbing the data and this is crucial to make sure that we're working with high quality data. Data cleaning tasks include removing major errors duplicates or outliers all of which are problems when we aggregate data from numerous sources. This brings structure to our data so for example fixing typos which will help us manage our data more easily and finally it helps filling in major gaps. A good analyst will spend about 70% to 90% for cleaning their data. If we get any wrong data or analysing error in this data it will severely impact our results. There exists some of the tools that we can use to clean our data such as excel, SQL, open refine are excellent for basic data cleaning and python libraries such as pandas and some r packages are better suited to heavy data scrubbing.



**Fig -4:** Data Cleaning

## STEP-4 ANALYSING THE DATA:

Once we cleaned our data now, we should analyse it. The type of data analysis we conduct largely depends on what our goal is but there are many techniques available univariate or bivariate analysis time series analysis and regression analysis. We should apply them depending on what type of insights we are hoping to gain. Data analysis fit into the four categories descriptive analysis which is analysis which identifies what has already happened this is a common first step that companies do before proceeding with deeper explorations. Diagnostic analysis where the focus is on understanding why something has happened. Predictive analysis which is where we identify future trends by the analysis of historical data. Predictive analysis is commonly used by businesses to forecast future growth and lastly prescriptive analysis which allows us to make recommendations for the future.

## STEP-5 VISUALIZING THE DATA:

Data visualization is a technique which is used to present the data in a pictorial or graphical format. It enables stakeholders to analyse data visually the data in graphical formats and allow them to identify new trends and patterns easily. There are many tools to visualize the data most popular tools include Tableau and Power BI. Tableau is a powerful data visualization tool which is used in bi industry. By using this we can simplify raw data into an easily understandable format for professionals at any level. Tableau can handle large volumes of data quickly. It is faster and provides extensive features for visualizing the data. Power BI is easy to use. It is faster and performs better when the volume of data is limited. Power BI tends to drag slowly while handling bulk data.
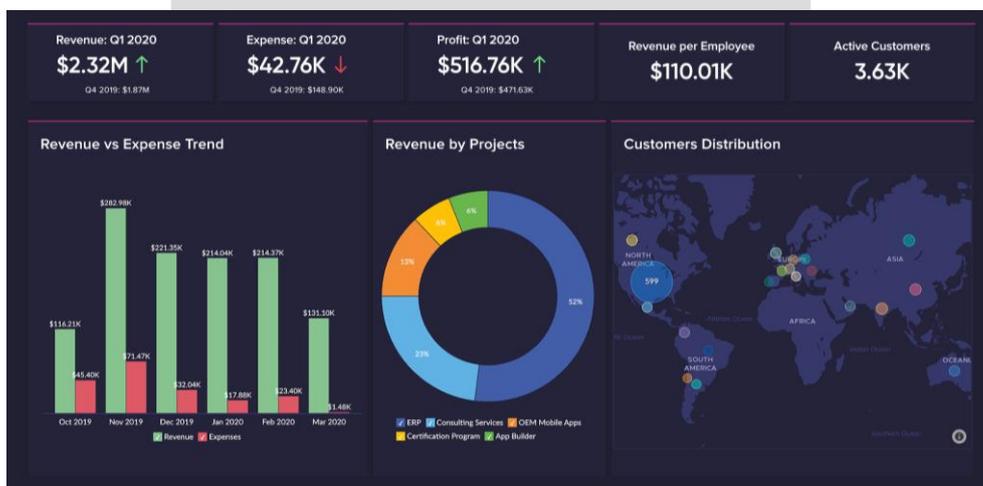


**Fig -5:** Data Visualization

**CONCLUSION**

Data analysis is process where we can draw conclusions by means of graphs and charts with the help of data, which is either qualitative or quantitative which uses many tools such as excel, SQL, python, and tableau for data visualization. This helps the stakeholders to understand what is going to happen in future or what happened in the past.

**REFERENCES**

[1] Francesco Corea, "An Introduction to Data", Springer Nature Switzerland AG publishers, 2019
[2] Wiley, "Data Science & Big data analytics", John Wiley & sons Inc. publishers, 2015
[3] Murtaza Haider, "Getting Started with Data Science", Pearson plc publishers, 2015
[4] Sanjiv Ranjan Das, "Data Science", published by S. R. Das, 2016
[5] John. W. Tukey, "Exploratory data analysis", Addison-Wesley publishers, 1997
[6] Albright, Winston, "Business Analytics", Cengage learning publishers, 2013