**What is Data Analysis?**

**“Data is Everywhere”**, in sheets, in social media platforms, in product reviews and feedback, everywhere. In this latest information age it’s created at blinding speeds and, when data is analyzed correctly, can be a company’s most valuable asset. “**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 recognize 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 grow more.

All you need to do is analyze your business data and business processes. The process of studying the data to find out the answers to how and why things happened in the past. Usually, the result of data analysis is the final dataset, i.e. a pattern, or a detailed report that you can further use for [**Data Analytics**](https://www.geeksforgeeks.org/data-analytics-and-its-type/). In this article, we will explore ***What is Data Analysis?***, ***How it works***, the ***types of data analysis***, and the ***Tools required for data analysis***.

**What is Data Analysis?**

Before jumping into the term “**Data Analysis”,** let’s discuss the term **“Analysis”**. Analysis is a process of answering **“How?”** and **“Why?”**. For example, how was the growth of XYZ Company in the last quarter? Or why did the sales of XYZ Company drop last summer? So to answer those questions we take the data that we already have. Out of that, we filter out what we need. This filtered data is the final dataset of the larger chunk that we have already collected and that becomes the target of **data analysis**. Sometimes we take multiple **data sets** and **analyze** them to find a pattern. For example, take summer sales data for three consecutive years. I found out if that fall in sales last summer was because of any specific product that we were selling or if it was just a recurring problem. It’s all about looking for a pattern. We analyze things or events that have already happened in the past.

**Why Data Analysis is important?**

Let’s say you own a business and sell daily products. Your business model is pretty simple. You buy products from the supplier and sell them to the customer. Let’s assume the biggest challenge for your business is to find the right amount of stock at the given time. You can’t stock excess dairy products as they are perishable and if they go bad you can’t sell them, resulting in a direct loss for you. At the same time, you can not understock as it may result in the loss of potential customers. But data analytics can help you in predicting the strength of your customers at a given time. Using that result, you can sufficiently stock your supplies, in turn, minimizing the loss. In simple words, using data analysis, you can find out the time of the year when your store has the least or the most customers. Using this info, you can stock your supplies accordingly. So these are some reasons why analysis of data is important.

**Types of Data Analysis Methods**

The major Data Analysis methods are:

1. **Descriptive Analysis**
2. **Diagnostic Analysis**
3. **Predictive Analysis**
4. **Prescriptive Analysis**
5. **Statistical Analysis**

### ****1. Descriptive Analysis****

A [Descriptive Analysis](https://www.geeksforgeeks.org/descriptive-analysis-in-r-programming/) looks at data and analyzes past events for insight as to how to approach future events. It looks at the past performance and understands the performance by mining historical data to understand the cause of success or failure in the past. Almost all management reporting such as sales, marketing, operations, and finance uses this type of analysis.

**Example:** Let’s take the example of DMart, we can look at the product’s history and find out which products have been sold more or which products have large demand by looking at the product sold trends, and based on their analysis we can further make the decision of putting a stock of that item in large quantity for the coming year.

### ****2. Diagnostic Analysis****

Diagnostic analysis works hand in hand with [Descriptive Analysis](https://www.geeksforgeeks.org/descriptive-analysis-in-r-programming/). As descriptive Analysis finds out what happened in the past, diagnostic Analysis, on the other hand, finds out why did that happen or what measures were taken at that time, or how frequently it has happened. it basically gives a detailed explanation of a particular scenario by understanding behavior patterns.

**Example:** Let’s take the example of Dmart again. Now if we want to find out why a particular product has a lot of demand, is it because of their brand or is it because of quality. All this information can easily be identified using diagnostic Analysis.

### ****3. Predictive Analysis****

Information we have received from descriptive and diagnostic analysis, we can use that information to predict future data. it basically finds out what is likely to happen in the future. Now when future data doesn’t mean we have become fortune-tellers, by looking at the past trends and behavioral patterns we are forecasting that it might happen in the future.

**Example:** The best example would be **Amazon** and **Netflix** recommender systems. You might have noticed that whenever you buy any product from Amazon, on the payment side it shows you a recommendation saying the customer who purchased this has also purchased this product that recommendation is based on the customer purchase behavior in the past. By looking at customer past purchase behavior analyst creates an association between each product and that’s the reason it shows recommendation when you buy any product.

### ****4. Prescriptive Analysis****

This is an advanced method of [Predictive Analysis](https://www.geeksforgeeks.org/step-by-step-predictive-analysis-machine-learning/). Now when you predict something or when you start thinking out of the box you will definitely have a lot of options, and then we get confused as to which option will actually work. Prescriptive Analysis helps to find which is the best option to make it happen or work. As predictive Analysis forecast future data, Prescriptive Analysis on the other hand helps to make it happen whatever we have forecasted. Prescriptive Analysis is the highest level of Analysis that is used for choosing the best optimal solution by looking at descriptive, diagnostic, and predictive data.

**Example:** The best example would be **Google’s self-driving ca**r, by looking at the past trends and forecasted data it identifies when to turn or when to slow down, which works much like a human driver.

**5. Statistical Analysis**

**Statistical Analysis** is a statistical approach or technique for analyzing data sets in order to summarize their important and main characteristics generally by using some visual aids. This approach can be used to gather knowledge about the following aspects of data:

1. Main characteristics or features of the data.
2. The variables and their relationships.
3. Finding out the important variables that can be used in our problem.

## What is the Data Analysis Process?

A [Data analysis](https://www.geeksforgeeks.org/data-analysis-with-python/) has the ability to transform raw available data into meaningful insights for your business and your decision-making. While there are several different ways of collecting and interpreting this data, most data-analysis processes follow the same six general steps.

1. **Specify Data Requirements**
2. **Collect Data**
3. **Clean and Process the Data**
4. **Analyse the Data**
5. **Interpretation**
6. **Report**

#### **1. Specify Data Requirements**

In step 1 of the data analysis process define what you want to answer through data. This typically stems from a business problem or questions, such as

* How can we reduce production costs without sacrificing quality?
* How do customers view our brand?
* How can we increase sales opportunities using our current resources?

#### **2. Collect Data**

* **Find Your Source**: Determine what information can be collected from existing sources, and what you need to find elsewhere.
* **Standardize Collection**: Create file storage and naming system ahead of time.
* **Keep Track**: Keep data organized in a log with dates and add any source notes as you go.

| **Internal Sources** | **External Sources** |
| --- | --- |
| Customer service data | Social media APIs |
| Marketing analytics | Google public data |
| Sales statistics | Public government data |
| Human resource data | Global finance data |
|  | Official research statistics |

#### **3. Clean and Process the Data**

Ensure your data is correct and usable by identifying and removing any errors or corruption.

* **Monitor Errors**: Keep a record and look at trends of where most errors are coming from.
* **Validate Accuracy**: Research and invest in data tools that allow you to clean your data in real-time.
* **Scrub for Duplicate Data**: Identify and remove duplicates so you save time during analysis.
* **Delete all Formatting**: Standardise the look of your data by removing any formatting styles.

#### **4. Analyse the Data**

Different data analysis techniques allow you to understand, interpret, and derive conclusions based on your business question or problem.

| **Descriptive Analysis** | **Inferential Analysis** |
| --- | --- |
| Analysis of data that helps show variables in a meaningful way and find patterns. | Exploring the relationship between multiple variables to make predictions. |
| **Measure of Tendency:** The central position of a frequency distribution for a group of data. | **Correlation:** Describe the relationship between two variables. |
| **Measure of Spread:** Summarising a group of data by describing how to spread out the scores are. | **Regression:** Shows or predicts the relationship between two variables. |

#### **5. Interpretation**

As you interpret the result of your data, ask yourself these key questions:

* Does the data answer your question? How?
* Does the data help you defend against any objections? How?
* Are there any limitations or angles you haven’t considered?

#### **6. Report**

Data Analysis can be used to report to different people:

* A primary collaborator or client
* Executive and business leaders
* A technical supervisor
* **Keep it Succinct:** Organize data in a way that makes it easy for different audiences to skim through it to find the information most relevant to them.
* **Make it Visual:** Use data visualizations techniques, such as tables and charts, to communicate the message clearly.
* **Include an Executive Summary**: This allows someone to analyze your findings upfront and harness your most important points to influence their decisions.

## Top 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. Below is the list of some popular tools explain briefly:

* **SAS** :[SAS](https://www.geeksforgeeks.org/introduction-to-sas-programming/) was a programming language developed by the SAS Institute for performed advanced analytics, multivariate analyses, business intelligence, data management, and predictive analytics. , SAS was developed for very specific uses and powerful tools are not added every day to the extensive already existing collection thus making it less scalable for certain applications.
* **Microsoft Excel :**It is an important spreadsheet application that can be useful for recording expenses, charting data, and performing easy manipulation and lookup and or generating pivot tables to provide the desired summarized reports of large datasets that contain significant data findings. It is written in [C#](https://www.geeksforgeeks.org/csharp-programming-language/), [C++](https://www.geeksforgeeks.org/c-plus-plus/), and [.NET Framework](https://www.geeksforgeeks.org/introduction-to-net-framework/), and its stable version was released in 2016.
*  **R**:It is one of the leading programming languages for performing complex statistical computations and graphics. It is a free and open-source language that can be run on various ***UNIX platforms, Windows, and macOS***. It also has a command-line interface that is easy to use. However, it is tough to learn especially for people who do not have prior knowledge about programming.
*  **Python**:It is a powerful high-level programming language that is used for general-purpose programming[. Python](https://www.geeksforgeeks.org/python-programming-language/) supports both structured and functional programming methods. Its extensive collection of libraries make it very useful in [data analysis](https://www.geeksforgeeks.org/data-analysis-with-python/). Knowledge of [Tensorflow](https://www.geeksforgeeks.org/introduction-to-tensorflow/), [Theano](https://www.geeksforgeeks.org/theano-in-python/), [Keras](https://www.geeksforgeeks.org/difference-between-tensorflow-and-keras/), [Matplotlib](https://www.geeksforgeeks.org/python-introduction-matplotlib/), [Scikit-learn](https://www.geeksforgeeks.org/learning-model-building-scikit-learn-python-machine-learning-library/), and Keras can get you a lot closer to your dream of becoming a machine learning engineer.
*  **Tableau Public**:Tableau Public is free software developed by the public company “[Tableau Software](https://www.geeksforgeeks.org/tableau-installation-guide/)” that allows users to connect to any spreadsheet or file and create interactive data visualizations. It can also be used to create maps, dashboards along with real-time updation for easy presentation on the web. The results can be shared through social media sites or directly with the client making it very convenient to use.

 **RapidMiner**: RapidMiner is an extremely versatile data science platform developed by “RapidMiner Inc”. The software emphasizes lightning-fast data science capabilities and provides an integrated environment for the preparation of data and application of machine learning, deep learning, text mining, and predictive analytical techniques. It can also work with many data source types including Access, [SQL](https://www.geeksforgeeks.org/sql-tutorial/), Excel, Tera data, [Sybase](NULL), Oracle, [MySQL](https://www.geeksforgeeks.org/sql-tutorial/), and Dbase.

*  **Knime**:Knime, the Konstanz Information Miner is a free and open-source data analytics software. It is also used as a reporting and integration platform. It involves the integration of various components for [Machine Learning](https://www.geeksforgeeks.org/machine-learning/) and [data mining](https://www.geeksforgeeks.org/data-mining/) through the modular data-pipe lining. It is written in Java and developed by **KNIME.com AG**. It can be operated in various operating systems such as Linux, OS X, and Windows.

## How to Become Data Analyst?

To become a ***data analyst*** you must require least a bachelor’s degree. To those who are at higher level , you may require a master’s degree. You also need to developed skills such as : [Statistical Analysis](https://www.geeksforgeeks.org/what-is-exploratory-data-analysis/), [Data Visualization](https://www.geeksforgeeks.org/what-is-data-visualization-and-why-is-it-important/), [Data Cleaning](https://www.geeksforgeeks.org/data-cleaning-in-r/), [Database Mnagement](https://www.geeksforgeeks.org/introduction-of-dbms-database-management-system-set-1/), and [MS-Excel.](https://www.geeksforgeeks.org/excel-tutorial/) Start with internships to gain experience and make projects that will demonstrate your skills. The files of ***Data Analytics*** is changing rapidly So, you need to keep yourself updated as according to the time by taking online sessions, attending workshops, or reading related books and articles published. As you grow in the field of data science you might find the specific industries to work with and you can explore more in-depth about Data Analysis.