Introduction

Data Analysis is mostly about exploring data. In general people have this idea that it's all about building Machine learning models and predictive stuff. But actually it's really about the working with the data. The best thing that one can do with data is just understand what they have and start asking questions and just start exploring and from the results or observations one can make incredible like ,for businesses and organizations, help them drive improvements and address the problems are trying to solve.

The first step in the whole process of Data analysis is 'Getting the Data'. Data comes in all sorts of shapes and sizes and all different places. It could be a simple as files; some text files, or some documents or increasingly commonly if you working in a large organization, data is more likely to be in somewhere like a database. So one needs to go and get it out from there. Now-a-days we live in a connected world, where there are devices running, that are emitting data all the time, in real time. One have got sensors, and this so-called Internet of things is out there. So one have all this data from all these types of places.

Once the data is gathered, the next requirement is of some sort of tool to go and manipulate that data, work with that data, clean up and then start to analyse that data. And that tool might be something that is very commonly used like MS Excel, which is a really powerful tool that a lot of people use to Analyse data Or maybe it's a bit more specialized; like writing some R code or Python or something like that and then manipulating the data in that language. The crucial step is really understanding the kinds of questions one need to ask and how to think about data in hand.

Example: The Operations Head of a multinational Computer Hardware Production organization has a requirement to analyse production data and identify the defect patterns across all its manufacturing plants across the globe.

Using the proper Data Analysis technique and tool, the operations head can perform an analysis for every item produced on each machine. Also the data about the defects can be obtained and then analysed for questions like 'number of defects for each plant / machine / per day etc'. Thus through Data Analysis businesses can optimize their Supply-Chain Process and do a predication of Demand- Supply Chain.

1. Data Analysis

Now we understand that Data Analytics can be coined as a process of observing the data in order to draw meaningful and logical insight into the data and the information / trend contained in that data. With the trends and the information gained, the organizations can then make a better decisions. The process of Data Analytics can be classified into four types.

1.1 Types of Analytics

Different type of Data Analytics are as follows:

1.1.1 Descriptive Analytics

Descriptive Analytics explains what has already happened and it is the simplest form of Data Analytics. During the process of Descriptive Analytics data is sliced into smaller chunks and then appropriate information about, what has already happened, from those data chunks is extracted. For instance, for a Banking Client's data, using descriptive analytics one can find answers to following types of queries:

- Geographical dispersion of Client's
- What are different Segments of Client's (HNI etc.,)
- Clients expenditure's, income patterns etc.,

1.1.2 Diagnostic Analytics

In order to understand 'Why something' has happened from the data gathered one can use Diagnostic Analytics. This type of analytics digs deeper in to the data to identify the real cause.

1.1.3 Predictive Analytics

Predictive Analytics, as the name suggest is for making predications based on the trends identified in the Data. This type of Analytics uses different Statistical, Data modelling and Data mining techniques to study, observe and identify trends in the data gathered. Based on these trends, Business Analyst and Data scientist can make forecast for the future time. This type of analytics is of major use while defining the Business Strategies, Objectives and Processes for the specific time periods.

1.1.4 Prescriptive Analytics

Prescriptive Analytics helps to augment the decision making process of organizations by formulating the best solutions among the available options. While identifying the best possible solution among the available solutions, the prescriptive analysis process also considers the business limitations enforced.

1.2 Ares of Analytics

Following are the areas where Analytics is used extensively now-a-days

· 	Customer Analytics : Through Customer Analytics, the organizations gains the insight about customers and thereby help then make

decisions about the customers offers / discounts etc.,.

· Risk Analytics: Risk analysis makes attempts to predict the indecisions of the forthcoming business atmosphere. The results obtained through Risk Analytics that assists in evaluating the probability of project's success.

· Financial Analytics: is of greater use for Financial Executives. It provides various options to address financial related business queries and also to predict and device financial strategies.

. Performance Analytics: is conduct to optimize the day to day operations. It aids in planning daily operations, budgeting for short period, making strategies for meeting the SLA's and also to improve the overall operational performance.

2.Data Analysis and Excel

In the market, there are numerous Software and packages available today for performing Data analysis. Now the question arises **Why use excel for analysing the data?**

2.1Why Use Excel

Excel is extensively used for Data Analysis tool, primarily for following reasons:

1. Cost
2. Ease of Use
3. Learning
   MS excel is installed and accessible on almost all the Windows based computers so one need not to put in or invest in any other software for their basic Data analysis requirements. With minimum or almost negligible learning period, a novice in Data analysis can utilize MS excel for analytical purpose as easily as any other Windows based program.

2.2 Formatting and Conditional Formatting

In order to make Data ready and accessible for analysis, Formatting and conditional Formatting features of excel are used.

2.2.1 Formatting

With appropriate formatting techniques in place user is able to present the data more effectively and efficiently to the audience. Using Formatting users can add patterns or trends to the data. Patterns or trends can be implemented either by colors or Icons or formulas etc. Cells of Microsoft Excel spreadsheet can be formatted manually by selecting fonts, font color, font size, background colors, and borders. In addition, excel provides several predefined table styles for formatting the tables automatically.

2.2.2 Conditional Formatting

Conditional formatting enables users to apply formatting patterns on particular cells of excel spreadsheet, satisfying the specified condition/s.

Through Conditional formatting, data can be visualized in such a manner which makes complex data easy to understand and relevant information can be highlighted, using colors, Bars, Icons etc.,

There are few pre-sets styles defined in excel. These are:

* **Data Bars:** The horizontal bars which are added to cells, making it look like a Bar Graph
* **Color Scales:** Changes the color of cells based on the value of cell. Each color scale uses a two- or three-color gradient.
* **Icon Sets:** Icon are added to the cells based on its value.

2.3 Excel Functions and Formulas

MS excel provides several type of functions which can be of enormous help while analysing big chunks of data. Major type of Excel functions are:

2.3.1 Logical Functions

Logical functions provided by Excel, assesses specific cell or cells for the specified criteria and then return a Boolean Value either True or False. Various Logical functions supported by Excel are

Boolean Operator Functions: AND, OR, XOR, NOT

Functions Returning Constant Value: TRUE and FALSE.

Conditional Functions: IF, IFERROR, IFNA, IFS, SWITCH

2.3.2 Lookup and Reference Functions

Excel LOOKUP functions are used when it is required to look into a single row or column and find a value from the same position in a second row or column i.e., these functions returns the value from a specified range of cells , either a Colum or a Row OR from an Array. There are two forms of Excel Lookup Functions. These are:

Vector Form: searches one column or a row. It 'looks up' for a value in a data vector which is a 1-dimensional list of data and then

returns corresponding value from a second data vector. In case the function is not able to find an exact lookup value, it returns the closest value below the lookup value.

Array Form: A collection of values in the form of rows and columns is defined as an Array. This functions 'looks up' for a supplied value in the first column or row of a supplied data array which is a 2-dimensional table of data. It returns the corresponding value from the last column or row of the array. One Prerequisite for using the Array Form is that Data must be sorted.

Microsoft recommends using the VLOOKUP and HLOOKUP functions instead of using Array form of the Lookup functions.

VLOOKUP: performs a vertical Lookup and searches for the supplied data in the current worksheet. It is used to look for data in a table or a range by row.

HLOOKUP: performs a Horizontal Look up and searches for the value in first or top row of a table or an array of values. It returns a value in the same column from the specified row in the table or array.

Following Reference Functions are provided by Excel

INDEX: function utilizes and index to pick a value from a reference or array. This function returns a reference to a cell (or cells) for requested rows and columns.

OFFSET: This function is used to return a reference offset from a given reference range.

MATCH: This function looks for specified value in an array of cells and then returns the relative location of that element.

2.3.3 Statistical Functions

MS Excel offers a wide range of Statistical Functions that can accomplish majority of Statistical calculations ranging from basic mean, median & mode to complex statistical distribution and probability tests like Anova, CHITEST etc., Major Statistical functions of Excel are listed below:

MIN: returns the minimum value from a list of provided data readings

MAX: returns the maximum value from a list of provided data readings

AVERAGE: returns the average of the list of data readings provided

SUMIFS: calculates the sum of the range of selected cells based on the defined criteria

COUNTIFS: returns the number of cells among the range of cells provided meeting the defined criteria.

PERCENTILE: calculates and returns the K'th percentile of data readings provided, where K is in the range 0 - 1 (both inclusive)

QUARTILE: calculates and returns identified quartile of a set of data readings provided based on percentile value 0 - 1

STDEV: calculates the standard deviation for the given dataset or values representing population sample

MEDIAN: calculates the Median for the given dataset or values

MODE: finds the Mode or most frequently occurring value in the given dataset.

COVAR: Finds the Covariance for the given set of values or dataset

GEOMEAN: Calculates the Geometric mean of the given dataset or values.

2.4 Excel Pivot Tables

MS excel provides another commanding tool called Pivot Table which helps to tabulate and summarize large datasets for easy analysis. Using Pivot Tool, a lot of effort and time can be saved by summarizing the data in few clicks instead of writing thousands of formulas to achieve the same results.

2.4.1 Pivot Tables

Pivot tables were earlier referred as Crosstab. It's believed that building a Pivot table is very complex and time consuming task but it a myth. Rather, if the data is well organized then it's a minute job to create a pivot table. Characteristics of Pivot tables are listed below:

- Summarizes, analyses, explores and present the data efficiently
- Instant calculation and summarization of data
- Quickly Pivots or reorganizes the data
- Provides valuable insight (trends and patterns) about a large dataset that is otherwise difficult to identify
- Saves Effort and Time
- Provides Drill Down Capabilities

2.4.2 Analysing data using Pivot Tables

To analyse a dataset using Pivot table, user needs to simply drag and drop the relevant data in the appropriate cells. The tool itself re arrange the data enabling the user to play around with the summarized or reorganized data to discover and identify data trends and patterns. During data analysis following features offered by excel pivot tables are of great help:

Grouping: groups the data according to the Header values. Any field which is added in to the Pivot table either as row or column is can be grouped using the Grouping option. By default, the numeric values, including date and time fields, are grouped by pivot table

Filters: While doing an analysis on a large data it is often required to narrow down the data based on field values to bring out trends in the data. This can be effectively achieved using the Filters feature of the Pivot table.

Slicers: Slicers are just another form of filtering Slicers provided set of buttons based on data values of the field selected for slicer. There are no drop downs in this case. Slicers can be defined for any type of field in the pivot table

Custom Calculation: Very often there is requirement to change the way field values are displayed. For instance, instead of showing the sum or count of the numeric value it is required to display the numeric data in the form of percentages. This can be achieved using the Custom Calculation feature offered by Pivot table.

Calculated Field: In case there is a requirement to do some runtime calculation using the Pivot tables fields then calculated fields are used. These type of fields allows to write custom formulas in the Pivot table using already present fields in Pivot table.

2.5 Analytical Tools of Excel

Excel not only helps in performing basic calculations but is capable of doing complex to very complex data analysis that can be create sophisticated Data Models that can be utilized for financial Modelling, Business planning and Risk evaluations. In fact Excel has now become a first step for all those people who are new in the world of Data Analytics.

2.5.1 What-If Analysis Tools

While analysing large data sets there is often a situation where some pieces data is missing. In such cases, to continue with the data analysis process, it becomes mandatory to fill in the values for the missing data. The What-If analysis tools of Excel comes handy in these situations. The process of changing the values in excel worksheet cells to observe the changes in the formulas outcome is termed as What-if analysis. With this feature it is possible to play around and experiment with the incomplete datasets also.

Goal Seek: It is used to obtain the results when the starting values are not known. As the name suggest, it tries to find out the goal from the end results. For instance, if the sum of two numbers are provided as 10 and one number is 6 then what the second number is? This can be solved using Goal Seek. In the large volume of data where the data readings are missing then Goal Seek is used to fill up the values with more complex logic

Scenario Manager: This tool enable users to create different circumstances with different set of data values and then switch between them. It allows to change 32 variables concurrently thus enabling to examine the outcomes for different input values at the same time.

Data Tables: Data Table in excel is a tool for doing a comparative analysis of the available values. Using data tables possible outcomes are determined by taking a set of input values. Instead of creating different situations as with Scenario manager, different output values for a formulae for different inputs is determined by making use of data tables.

Solver: Solver is used to find out the optimal solution for a formula according to the defined constraints and specifications. Constraints in Solver tools plays a crucial role. They confine the boundaries for the resources that helps solver to determine a feasible and optimal solution.

2.5.2 Statistical Analysis

One of the component of data analytics is Statistical Analysis which includes collection and examination of data sample of a population. It's a discipline of science in which large data is collected, studied, explored and manipulated to identify the hidden trends and patterns. In the end results are presented in such a manner that the end user can easily understand the findings about the data.

Statistical Analysis consists of following steps:

- Data collection
- Data Exploration
- Data Modelling
- Study of the Data Model
- Deductions from the data Model

  The objective of Statistical Analysis is to identify and discover the Trends and patterns in the Data.
  Excel is a widely used software package to implement the statistical concepts and perform the calculations accordingly. It can:

- Hold the data to analyse
- Provide various features and functions to implement statistical Concepts on the data
- Perform statistical calculations according to the functions chosen
- Display the output in various form like table or Chats

Statistical Analysis in excel can be performed using the Data Analysis command. In case data Analysis command is not available then it needs to be enabled by loading the Analysis Toolpack addin in excel.

Major types of Statistical operations or calculations that can be performed using the Data analysis command in excel are as follows:

Moving Average: is used for forecasting purposes by creating a series of Averages for different subsets of the dataset provided.

Hypothesis Testing: determines the characteristics of population by analysing a random sample drawn from the population itself. Two contrary hypothesis are formulated to establish the characteristics. The First Hypothesis called the Null Hypothesis, referred as H0, is assumed to be true until strong evidences are discovered against it by performing the Hypothesis tests. Second one, called Alternate Hypothesis and referred as H1, is assumed to be true when the Null Hypothesis is false.

ANOVA: is a statistical method which observes and reports the variance by comparing the Means of different groups of data. ANOVA stands for Analysis Of Variance. This method allows to test the Null Hypothesis for more than two groups also.

COVARIANCE: is a method which determines the relationship or behaviour between two random variables say X and Y. If Y increases with X then covariance is positive and both variables show similar behaviour. If Y decreases with X then covariance is negative and both variables show opposite behaviour.

CORRELATION: is a measure of extent to which two or more variables vary together. Correlation coefficient vary between -1 to +1. Correlation Coefficient of value +1 indicates perfect correlation and a value of -1 indicates perfect negative correlation. A value of 0 indicates no correlation between the variables

Regression: is most widely used statistical method and is used to define the relationship strength between on dependent variable (denoted as Y) and series of series of changing variables knows as independent variables.

3.   Conclusion

Undoubtedly MS excel is a tool which can cover and serve almost all the perspectives of Data analysis and due to this it has become a by default choice for a first step towards the world of Data Analysis. Though excel has almost everything to offer for Data crunching and analysis still there are some limitations of this package like:

- ·   Handling huge data volumes
- ·   Handling streaming data from disparate sources
- ·   Excel's limitations for scientific applications.
- ·   Many useful functions or tests are still not available in Excel and its Analysis Tools,

Even, with the above mentioned shortcomings, excel is the preferred tool for Data analysis for many users. Also if someone is an experience user in the field of Data Analytics still MS Excel offers something new to learn.

From <https://www.codeproject.com/Articles/1173811/Data-Analysis-with-MS-Excel>