**Q1. What is Data Analytics? Explain the difference between Data warehouse, Data Science, and Machine Learning. Write down various applications of Data Analytics.**

Data analytics is the science of analyzing raw data in order to make conclusions about that information. Many of the techniques and processes of data analytics have been automated into mechanical processes and [algorithms](https://www.investopedia.com/terms/a/algorithm.asp) that work over raw data for human consumption.

The process involved in data analysis involves several different steps:

1. The first step is to determine the data requirements or how the data is grouped. Data may be separated by age, demographic, income, or gender. Data values may be numerical or be divided by category.
2. The second step in data analytics is the process of collecting it. This can be done through a variety of sources such as computers, online sources, cameras, environmental sources, or through personnel.
3. Once the data is collected, it must be organized so it can be analyzed. Organization may take place on a spreadsheet or other form of software that can take statistical data.
4. The data is then cleaned up before analysis. This means it is scrubbed and checked to ensure there is no duplication or error, and that it is not incomplete. This step helps correct any errors before it goes on to a data analyst to be analyzed.

**Difference between Data Warehousing, Data Science and Machine Learning :-**

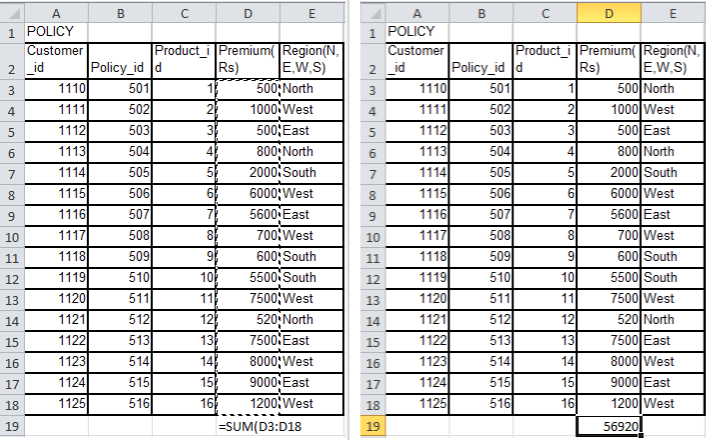
|  |  |  |
| --- | --- | --- |
| **Data Warehousing** | **Data Science** | **Machine Learning** |
| It is a process which is used to integrate data from multiple sources and then combine it into a single database. | Create insights from data dealing with all real-world complexities. This includes tasks like understand the requirement, extracting data etc. | Accurately Classify or predict the outcome for new data point by learning patterns from historical data, using mathematical models. |
| It provides the organization a mechanism to store huge amount of data. | Data science helps define new problems that can be solved using machine learning techniques and statistical analysis | The problem is already known and tools and techniques are used to find an intelligent solution |
| Knowledge of SQL is necessary to perform operations on data. | Knowledge of SQL is necessary to perform operations on data. | Knowledge of SQL is not necessary. Programs are written in languages like R, Python, Java, Lisp etc… |
| This process is solely carried out by engineers. | Data science is a complete process. | Machine learning is a single step in data science that uses the other steps of data science to create the best suitable algorithm for predictive analysis. |

**Applications of data analytics :-**

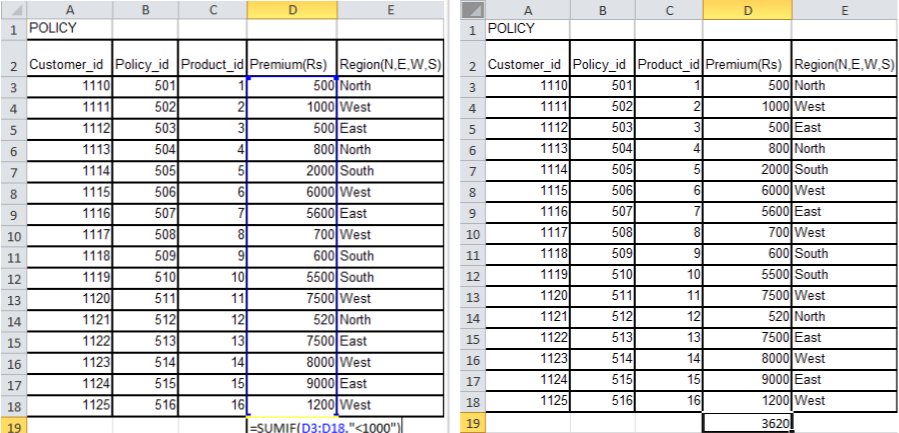
1. **Travel :-**Data analytics applications help in the optimization of traveler’s buying experience via social media and mobile/weblog data analysis. This is because customers’ preferences and desires can be obtained from this, therefore, making companies sell products from the correlation of the current sales to recent browse-to-buy conversion through customized offers and packages. Data analytics applications can also deliver personalized travel recommendations depending on the outcome from social media data.
2. **Digital Advertisement :**-Apart from web search, there is another area where data analytics and data science serves a very important purpose – digital advertisements. From the banners displayed on several websites to the digital billboards seen in the big cities; all are controlled by data algorithms.
3. **Internet/Web Search :**-When one mentions the word ‘search’, the first thing that comes to the mind is ‘Google’. In fact, Google to some point can be used in place of ‘search on the internet’ by saying ‘Google it’. Well, apart from Google, there are several other search engines such as Bing, Yahoo, Duckduckgo, AOL, Ask, etc. Each of these search engines is as a result of data science applications because they use algorithms to deliver the best results for any search query directed at them in just a split second. In respect to this, Google is known to process over 20 petabytes of data daily. Of course, without analytics and data science, this feat wouldn’t have been possible.
4. **Energy Management :**-Data analytics application here focuses mainly on monitoring and controlling of dispatch crew, network devices and make sure service outages are properly managed. Utilities get the ability to integrate as much as millions of data points within the performance of the network which allows the engineers make use of the analytics in monitoring the network.
5. **Healthcare :**-One challenge most hospitals face is coping with cost pressures in treating as many patients as possible, considering the quality of healthcare’s improvement. Machine and instrument data use has risen drastically so as to optimize and track treatment, patient flow as well as the use of equipment in hospitals. There is an estimation that a 1% efficiency gain will be achieved and would result to over $63 billion in worldwide health care services.
6. **Delivery Logistics :**-From data analytics applications, these companies have found the most suitable routes for shipping, the best delivery time, most suitable means of transport to select so as to gain cost efficiency and many others. Also, data generated by these companies through the use of GPS gives them enough opportunities to take advantage of data analytics and data science.

**Q2. Explain various functions in excel used for data analytics:**

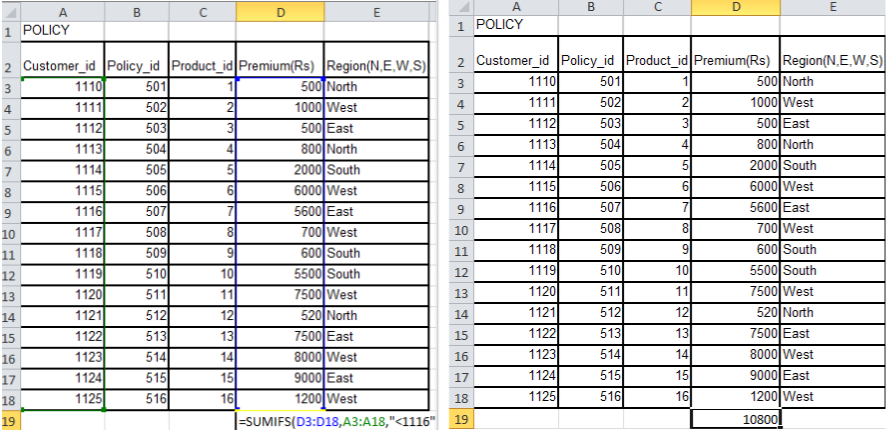
1. **Sum(), Sumif(), Sumifs(), SumProd()**



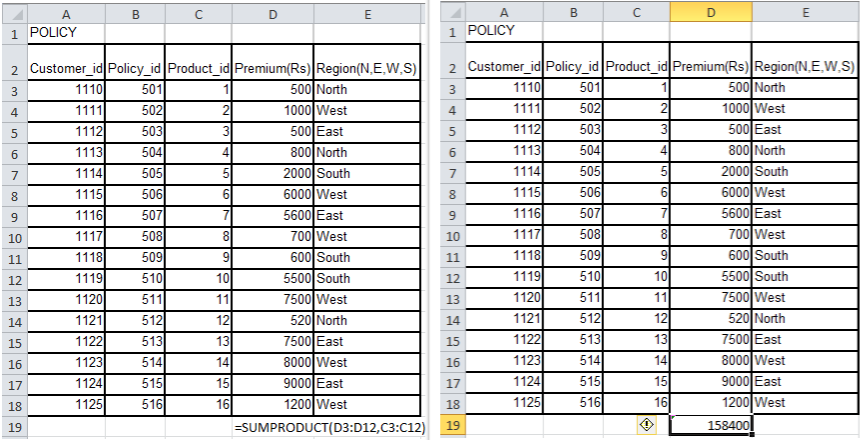
*Fig 1.1 Formula for SUM Operation Fig. 1.2 Result after applying Formula*



*Fig 1.3 Formula for SUMIF Operation Fig. 1.4 Result after applying Formula*

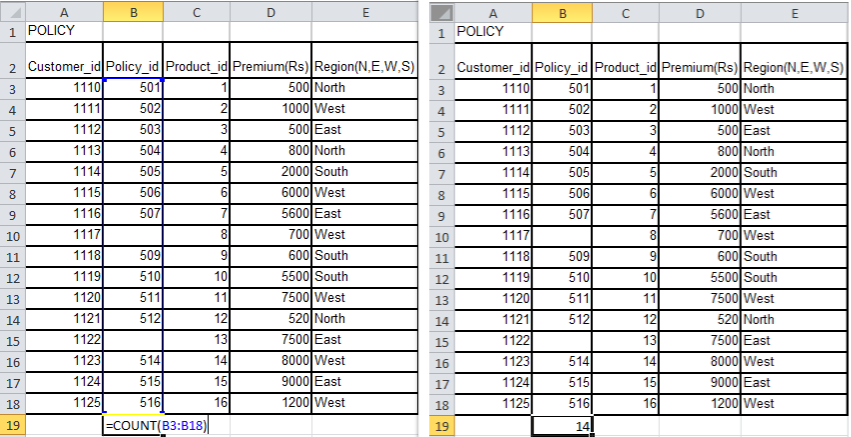


*Fig 1.5 Formula for SUMIFs Operation Fig. 1.6 Result after applying Formula*

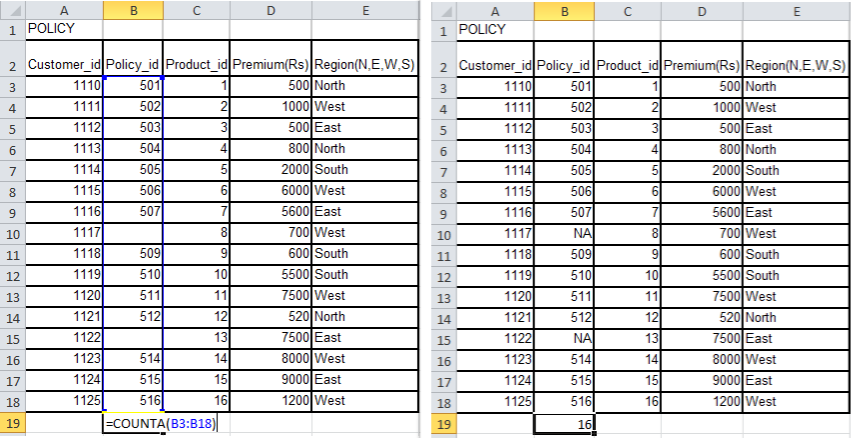
**

*Fig 1.7 Formula for SUMPRODUCT Operation Fig. 1.8 Result after applying Formula*

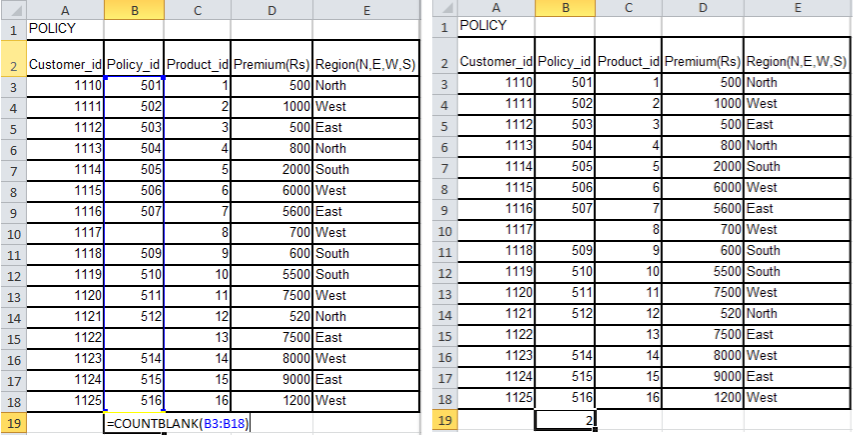
1. **Count(), CountA(), CountBlank(), Countif(), Countifs()**

****

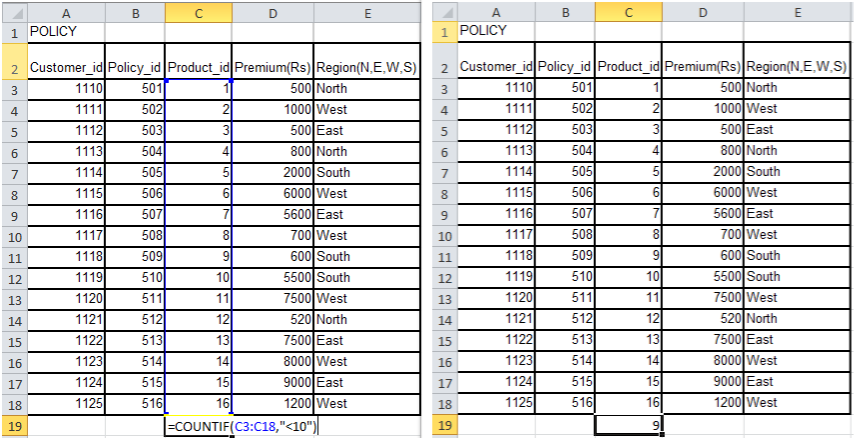
*Fig 2.1 Formula for COUNT Operation Fig. 2.2 Result after applying Formula*

****

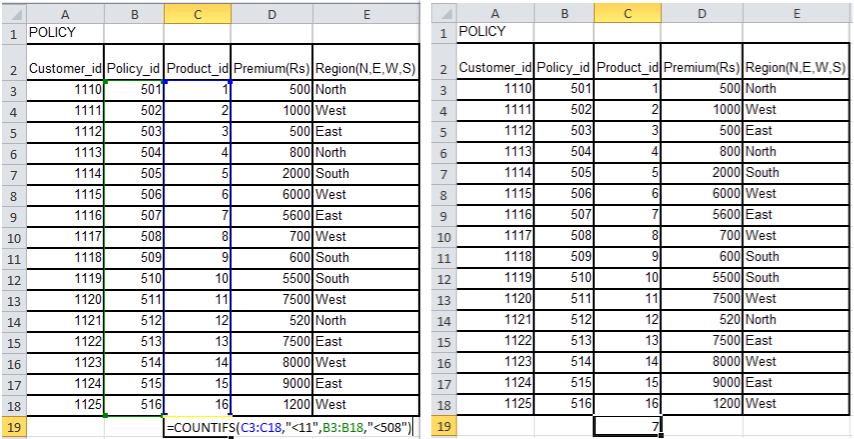
*Fig 2.3 Formula for COUNTA Operation Fig. 2.4 Result after applying Formula*

****

*Fig 2.5 Formula for COUNTBLANK OperationFig. 2.6 Result after applying Formula*

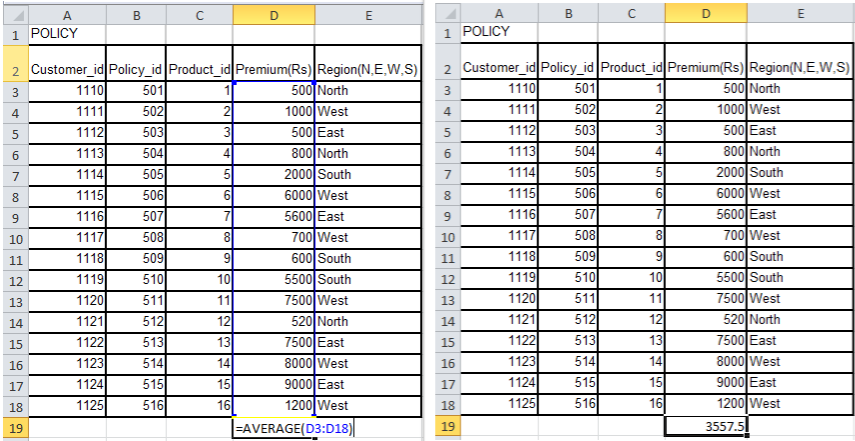
****

*Fig 2.7 Formula for COUNTIF OperationFig. 2.8 Result after applying Formula*

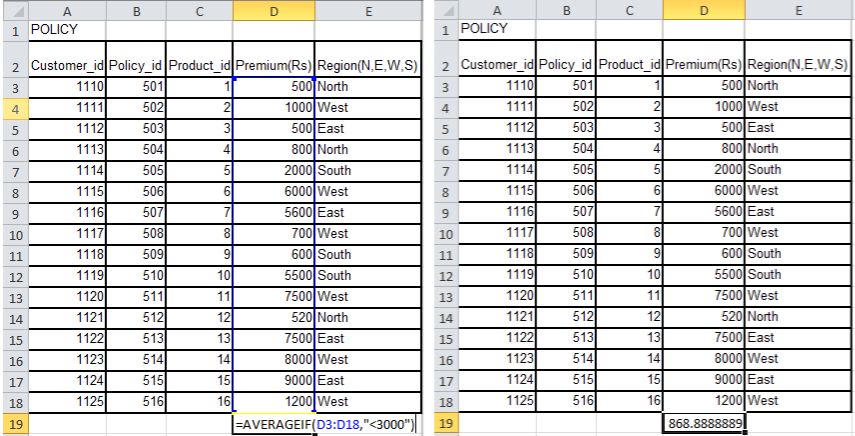


*Fig 2.9 Formula for COUNTIFs OperationFig. 2.10 Result after applying Formula*

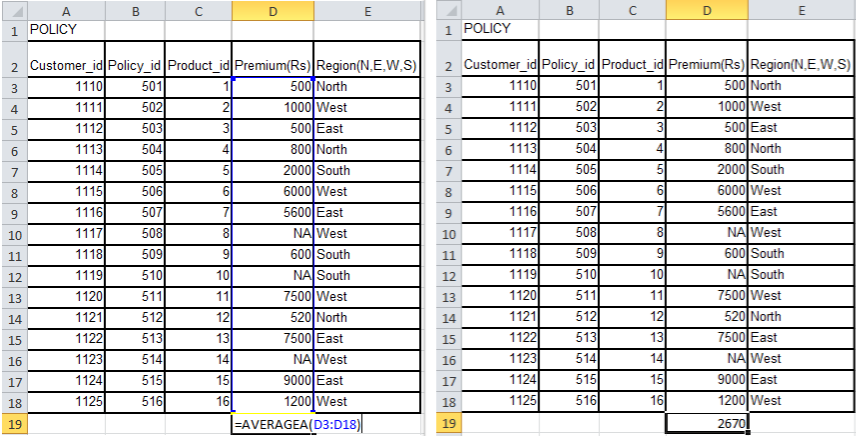
1. **Avg(), Avgif(), AvgA()**

****

*Fig 3.1 Formula for AVERAGE Operation Fig. 3.2 Result after applying Formula*

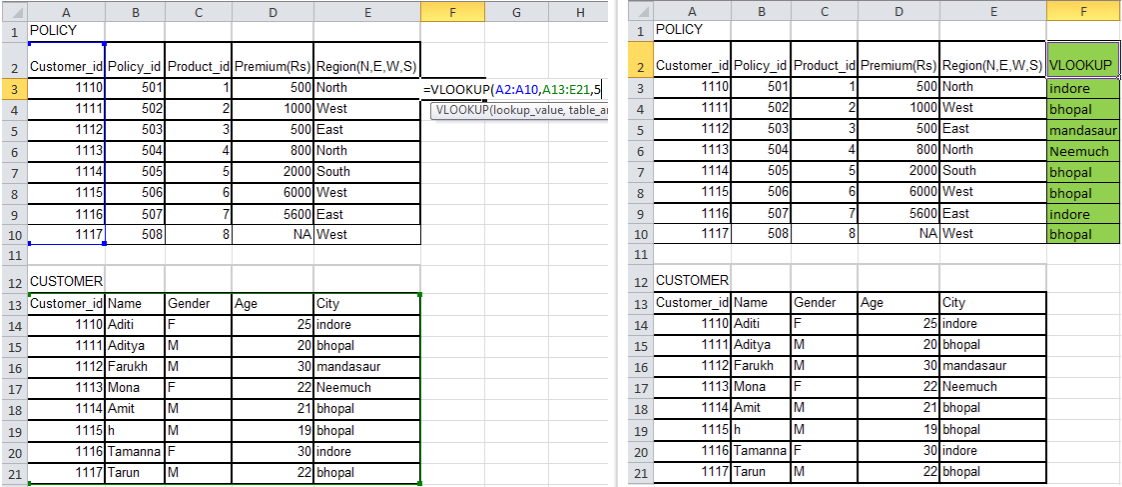


*Fig 3.3 Formula for AVERAGEIF Operation Fig. 3.4 Result after applying Formula*

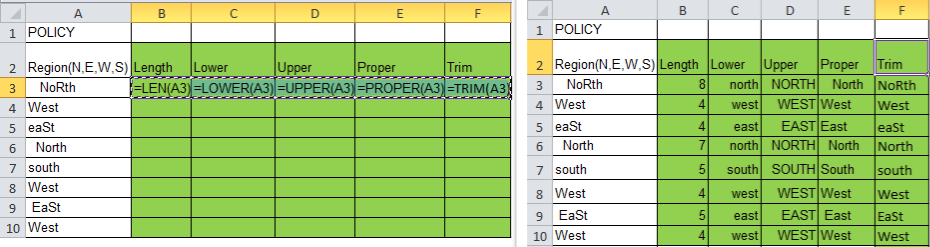


*Fig 3.5 Formula for AVERAGEA Operation Fig. 3.6 Result after applying Formula*

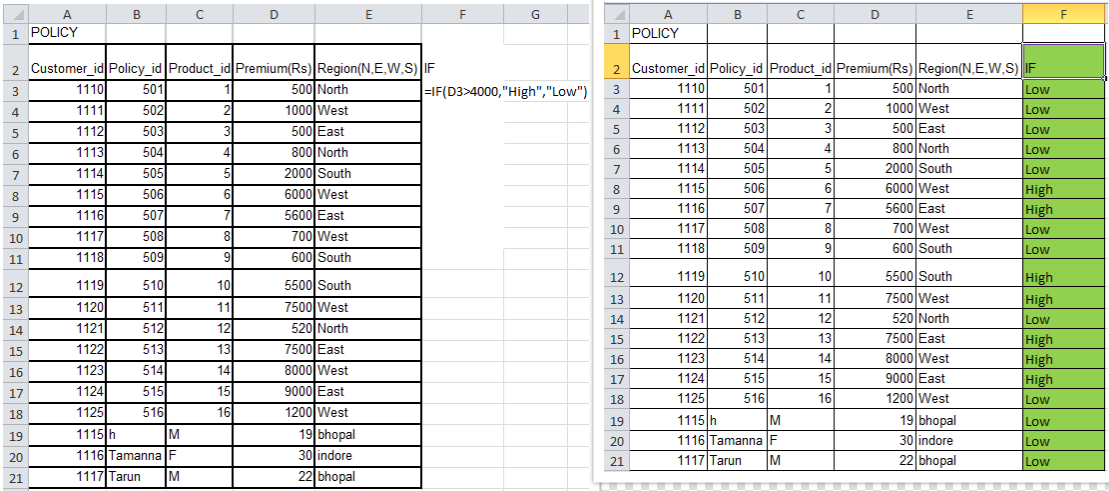
1. **Vlookup in excel, Concatenate(), Len(), Lower(), Upper(), Proper(), Trim(), If()**

****

*Fig 4.1 Formula for VLOOKUP Operation Fig. 4.2 Result after applying Formula*

**

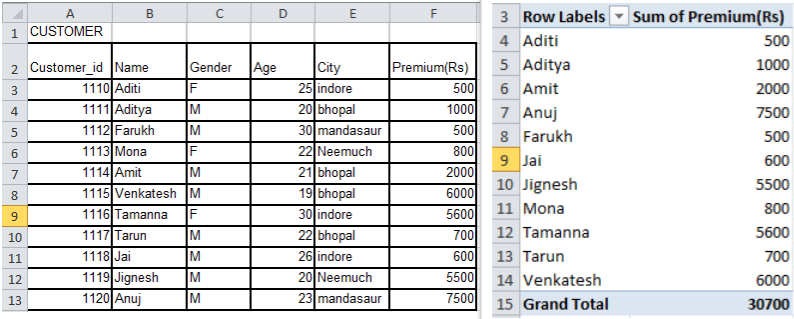
*Fig 4.3 Formula for LEN,Lower,Upper,Proper and Trim Operation Fig. 4.4 Result after applying Formula*

****

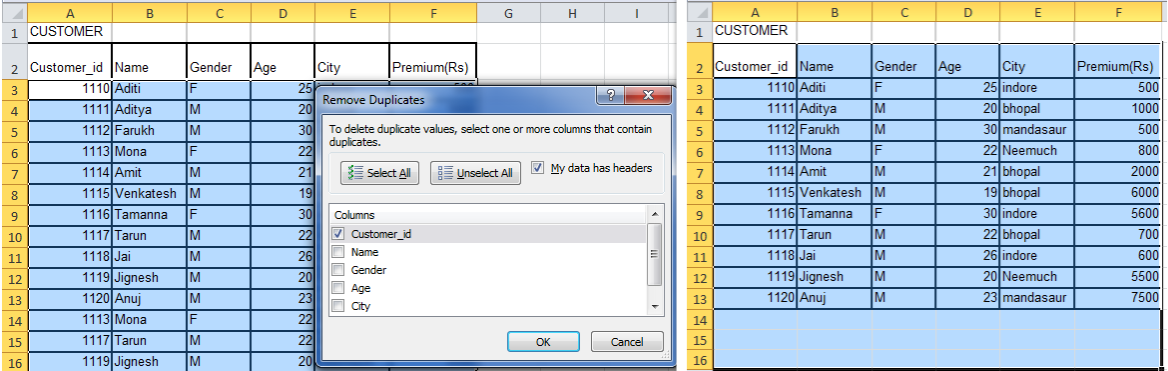
*Fig 4.5 Formula for IF Operation Fig. 4.6 Result after applying Formula*

**Q3. Explain Pivot table with proper example. Perform Data Cleaning Operations such as Removing Duplicate Values, Text to Column Conversion.**

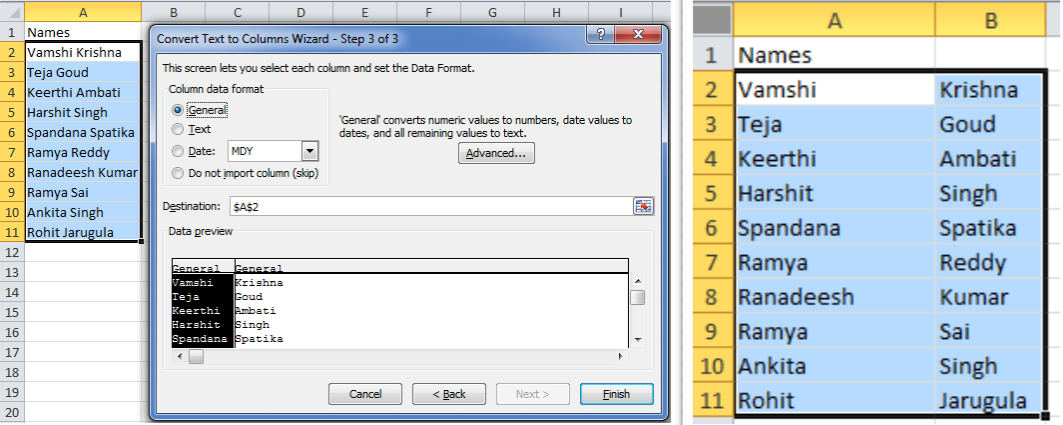
**Pivot Table :-Pivot tables** are one of **Excel**'s most powerful features. A pivot table allowsto extract the significance from a large, detailed data set.

****

*Fig 5.1 Getting the data of selected users’ Premium and their total*

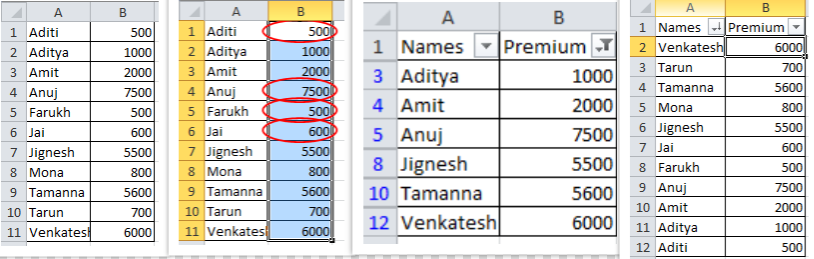
**

*Fig 5.2 Removing the duplicate customer\_id’s from the table*

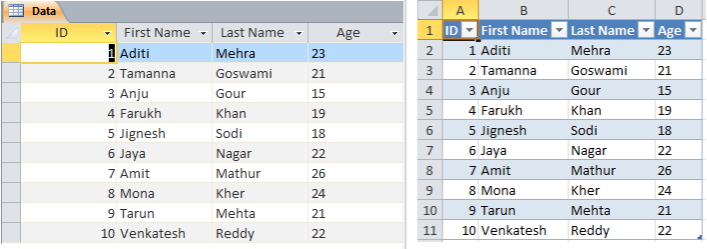
**

*Fig. 5.3 Seperating the fisrst name and last name in the table using Text-to-Column Conversion*

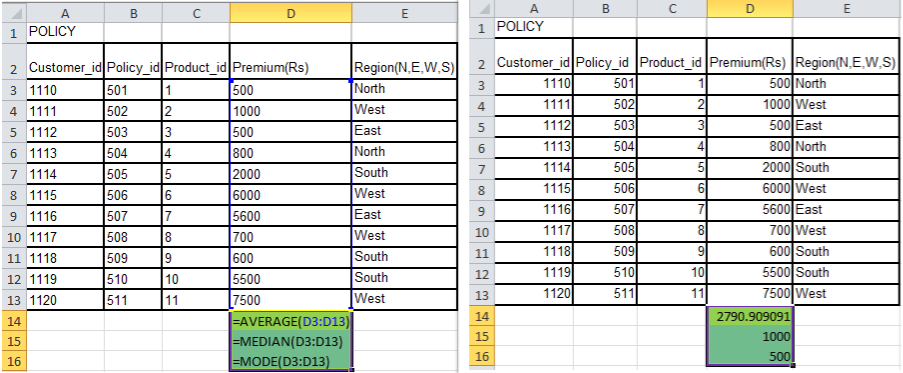
**Q4. Perform Data Validation, Filter function, sorting function and importing an external file into Excel. Perform Mean, Median, and Mode operation on dataset for data analytics operation.**

****

*Fig 6.1 Finding Invalid data, Filtering the users who have premium less 1000 and Sorting the table by ordering the names from z to a*

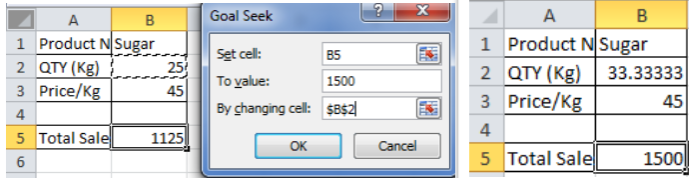
**

*Fig. 6.2 Importing the table from Microsoft Access to Excel*

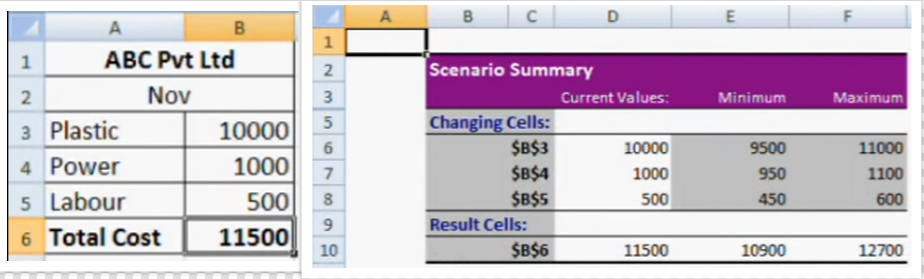
**

*Fig. 6.3 Finding Mean, Median and Mode of Premium Column*

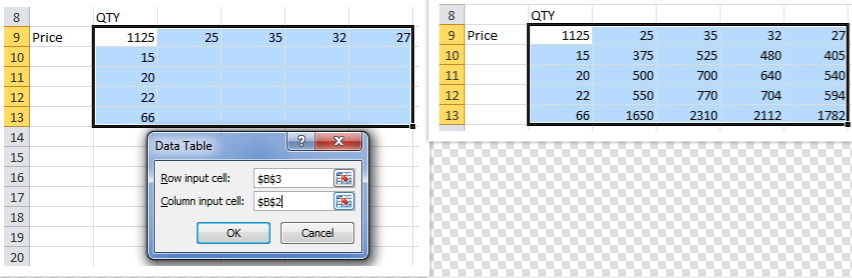
**Q5. Perform What-If-Analysis in Excel for Goal Seek, Scenario manager, Data Table.**

****

*Fig. 7.1 Performing Goal Seek Operation*

****

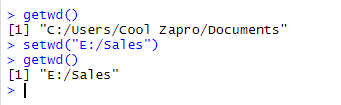
*Fig. 7.2 Finding Scenario for Maximum and Minimum Cost using Scenario Manager*

**

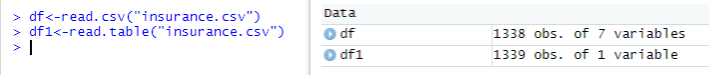
*Fig 7.3 Performing Goal Seek using Data table Operation*

**Q6. Explain Various R commands used in Data Analytics:**

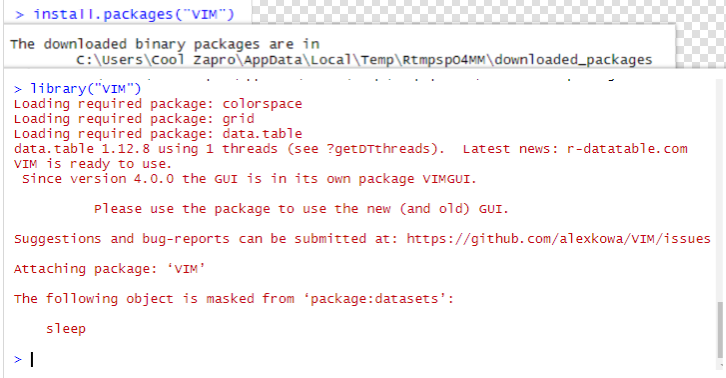
1. **getwd(), setwd() :-**

****

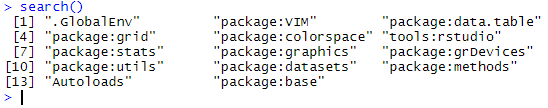
1. **Importing the data files in R environment with the help of read.table () and read.csv**

****

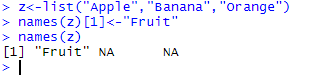
1. **Installing and including packages in R environment.**

****

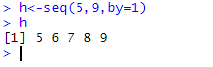
1. **Search()**

****

1. **Names()**

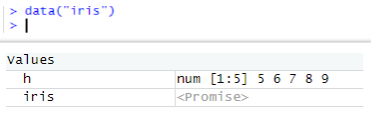
****

1. **Use of sequence operator**

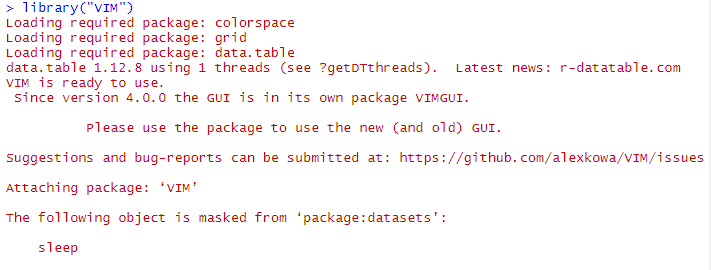
****

**Q7. Perform the following on R (use the IRIS dataset to perform the below operations :**

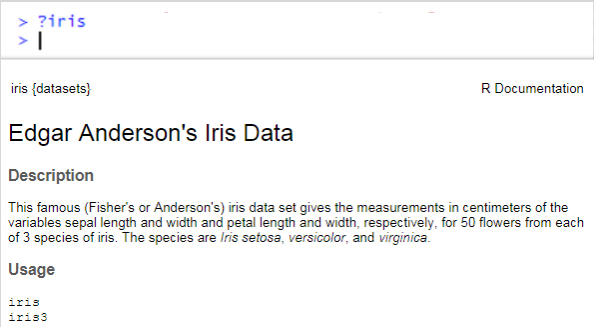
1. **Data() :-**

****

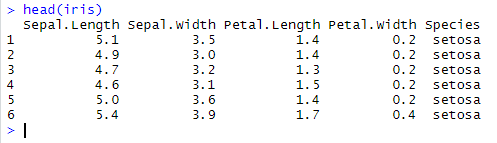
1. **library(datasets) :-**

****

1. **? iris :-**

****

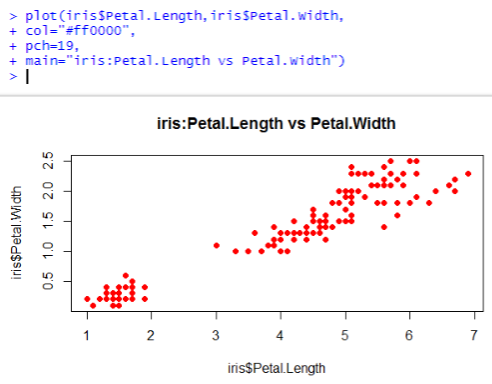
1. **How to fetch the first 6 records from datasets :-**



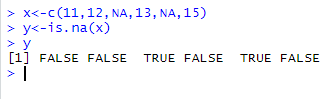
1. **Plot the graph for knowing the distribution of different species in iris dataset :-**

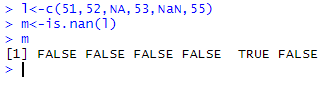
****

1. **Plot the graph for petal length and petal width. Give X axis as label “Petal length”, Y axis as label “Petal Width”, graph name as “IRIS : Petal Length vs Petal Width”, color as red and point character (pc) as 19.**

****

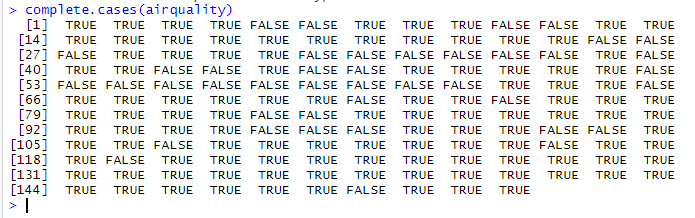
**Q8. a. How to Determine the missing data in dataset by using na() and nan(). Show it making your own dataset.**

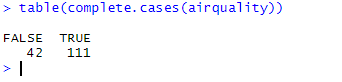
****

****

**b. Used Air Quality dataset in R (inbuilt dataset) and determine the following:**

1. **Missing values in complete dataset.**
2. **Use of table() and complete.case() in data analytics.**

****

****