DATA ANALYSIS AND VISUALIZATION REPORT

DATASET: RAINFALL IN PAKISTAN

GROUP MEMBERS:

SYEDA ALIZA

RUBA NIZAM

NARMEEN



DEPARTMENT OF MATHEMATICS

(BS FINANCIAL MATHEMATICS)

(SECOND YEAR 4TH SEMESTER)

PROJECT SUBMITTED TO:

(MR SYED UMAID AHMED)

CONTENTS

1) Introduction to data sciences

2) Uses of data sciences in daily life

3) Introduction to libraries used in a project dataset

4) Introduction to dataset used in a project

5) INPUT NO: 1

6) INPUT NO: 2

7) INPUT NO: 3

8) INPUT NO: 4

8) INPUT NO: 5

9) INPUT NO: 6

10) INPUT NO: 7

11) INPUT NO: 8

12) INPUT NO: 9

13) INPUT NO: 1O

14) INPUT NO: 11

15) INPUT NO: 12

INTRODUCTION TO DATA SCIENCES:

Data sciences is a field that uses scientific methods, algorithms and system to extract an insight and knowledge from the unstructured data and apply knowledge and actionable insights from data across a broad range of application domain. Data sciences encompasses preparing for data including cleaning, aggregating and manipulating the data. Data sciences is one of the most exciting fields out here today in modern world.

USES OF DATA SCIENCES IN DAILY LIFE:

Data sciences is one of the most common and famous field in today's world. It is a career field that stems from multiple disciplines following are some main and common applications of the data sciences given below

1) Banking

2) E-Commerce

3) Finance

4) Manufacturing

5) Transport

6) Healthcare

7) Predictive models for diagnosis

LIBRARIES USED IN PROJECT DATASET:

Following are the libraries that we have used in our data analysis project:

1) NumPy

2) Pandas

3) Matplotlib

4) Seaborn

1) NUMPY:

NumPy is a python library used for working with arrays. It also has a function of working in domain of linear algebra and matrices. It was created in 2005 by Travis Oliphant. It is an open-source project which we can use freely. NumPy stands for numerical python. NumPy facilitate advanced mathematical operations on a large scale of data.

2) MATPLOTLIB:

Matplotlib is a plotting library in python programming language and its numerical extension NumPy. A python Matplotlib script is structured so that a few lines of code all are that is required in most instances to generate a visual data plot. Matplotlib is developed by Michael Droettboom in 2003.

3) PANDAS:

Pandas is a python library providing fast, flexible and expressive data structure designed to make working with relational or labelled data easily. Is is used to perform machine learning task efficiently. Pandas makes it simple to do many of the time-consuming tasks efficiently which includes data cleansing, data fill, data normalization, merges and joins, data visualization, loading and saving data, data inspection, statistical analysis and many more.

4) SEABORN:

Seaborn is a data visualization library built on the top of Matplotlib. Visualization is the central part of Seaborn which helps in exploration and understanding of a data. This library is used to make statistical graphs in python. Its plotting function operate on data frames and arrays containing whole dataset and perform the necessary mapping to produce informative plots.

INRTODUCTION TO DATASET USED IN A PROJECT:

The selected dataset contains rainfall data of Pakistan. The parameter considered for the evaluation of the performance and the efficiency of the given rainfall prediction model. Following are the steps which we are going to discuss in our project dataset

1) Data Cleaning

2) Data Analysis

3) Forecast and prediction using the dataset

4) Graph plotting of given dataset

INPUT NO:1

IMPORTING LIBRARIES:

First of all, we have to import the libraries as show below.



INPUT NO: 2

READ THE FILE:

Now we are reading the file on jupyter notebook on which we have to complete our data analysis.

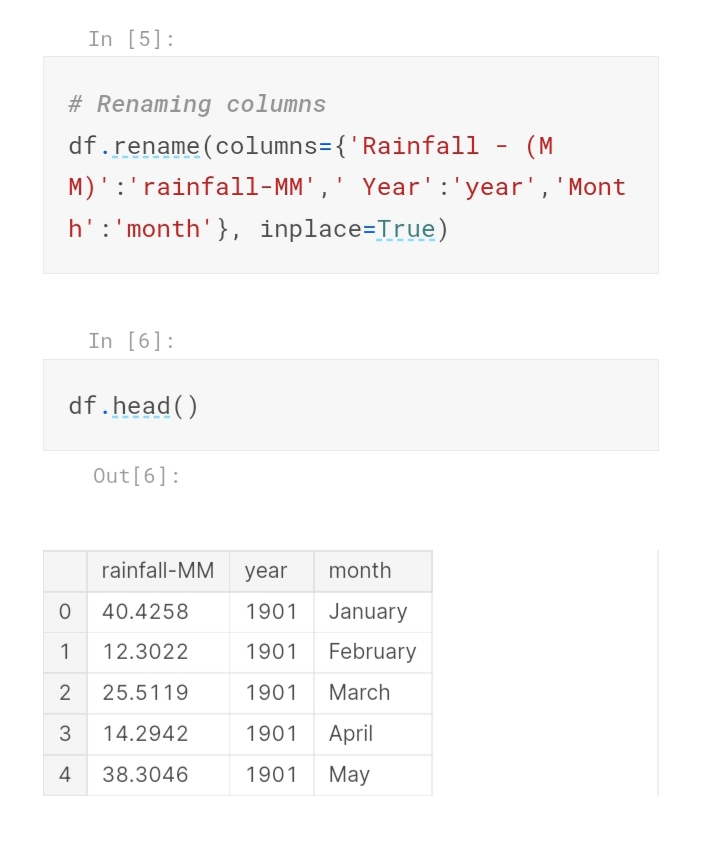


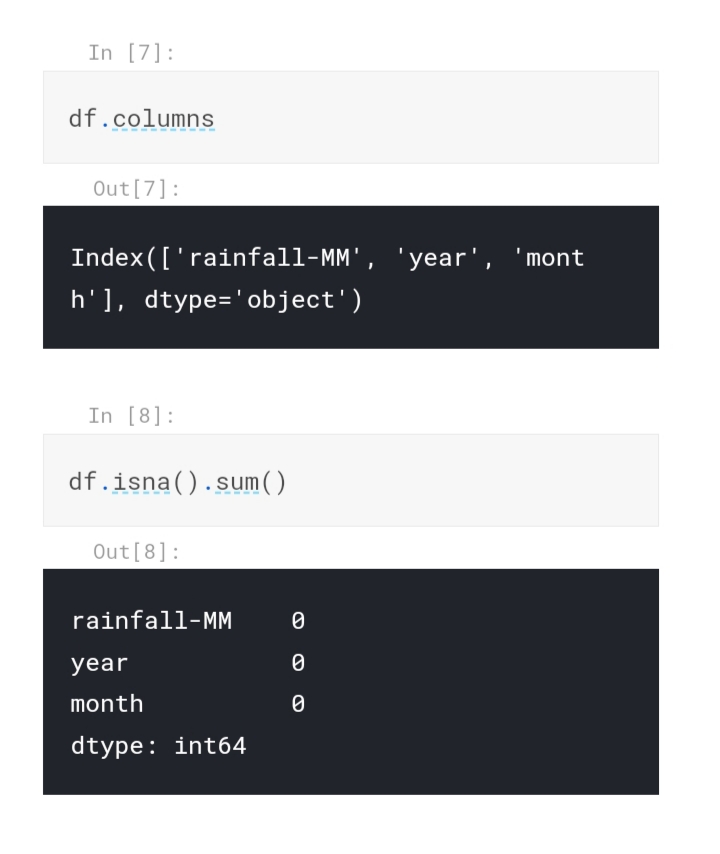
INPUT NO; 3

CHECKING THE COLUMNS AND HEADS:

Now we will check the head and columns of the data using the following inputs.



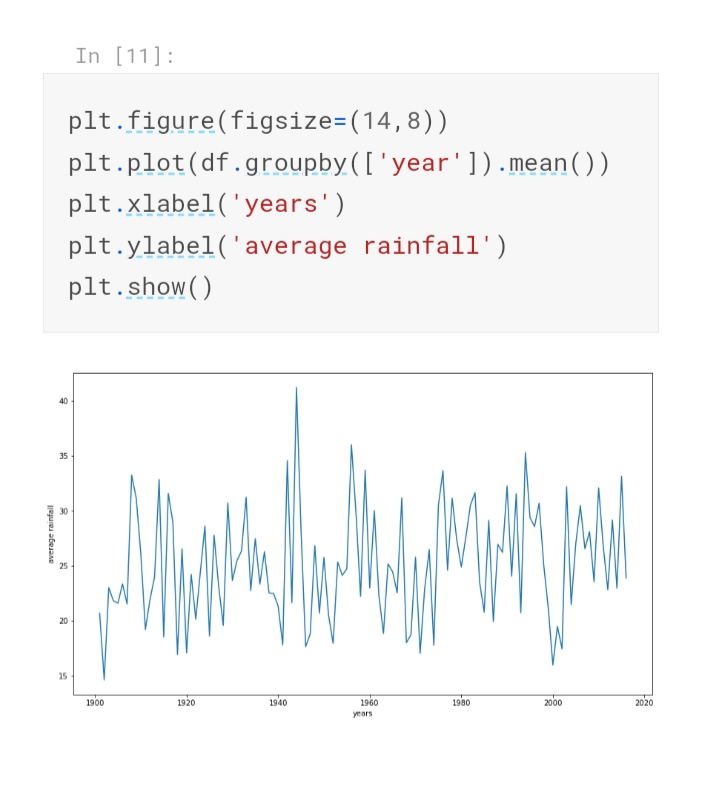




INPUT NO: 4:

PLOTTING OF GRAPH:

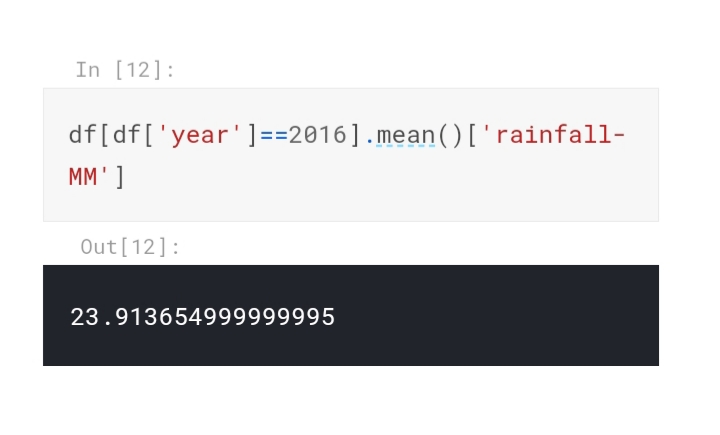
Now we will plot the graph of average rainfall throughout the data.



INPUT NO: 5

FINDING MEAN:

Now we are going to find the average of rainfall from our dataset.



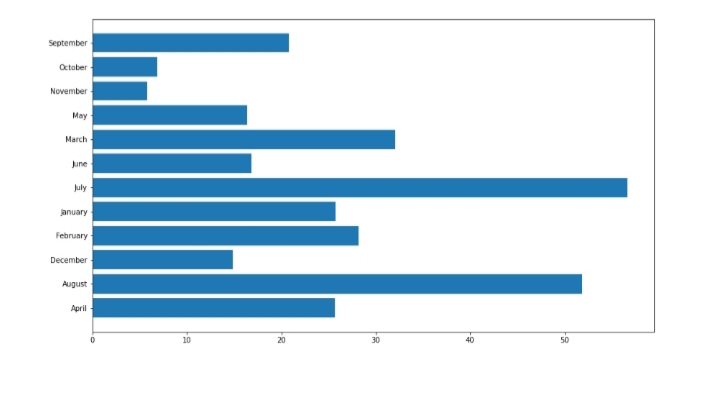
INPUT NO: 6

FINDING MONTHLY AVERAGE:

Now we are going to find monthly average of rainfall.







INPUT NO: 7

MONTHLY AVERAGE:

In the given input we will find the monthly average of our rainfall dataset.

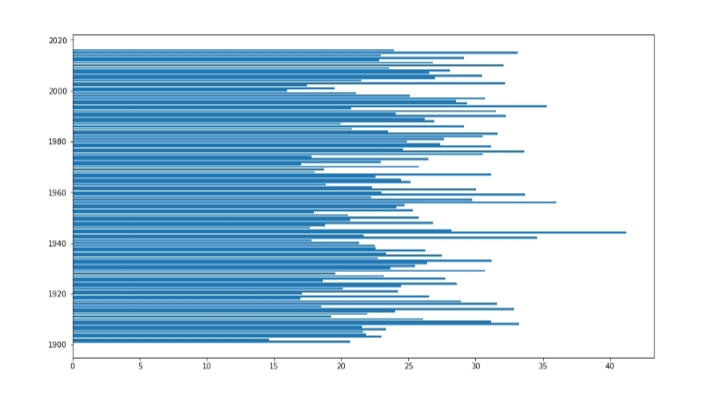


INPUT NO: 7

GRAPH PLOTTING:

For now, we will plot the graph of our monthly average of rainfall obtained above.





INPUT NO: 8

DATA SORTING:

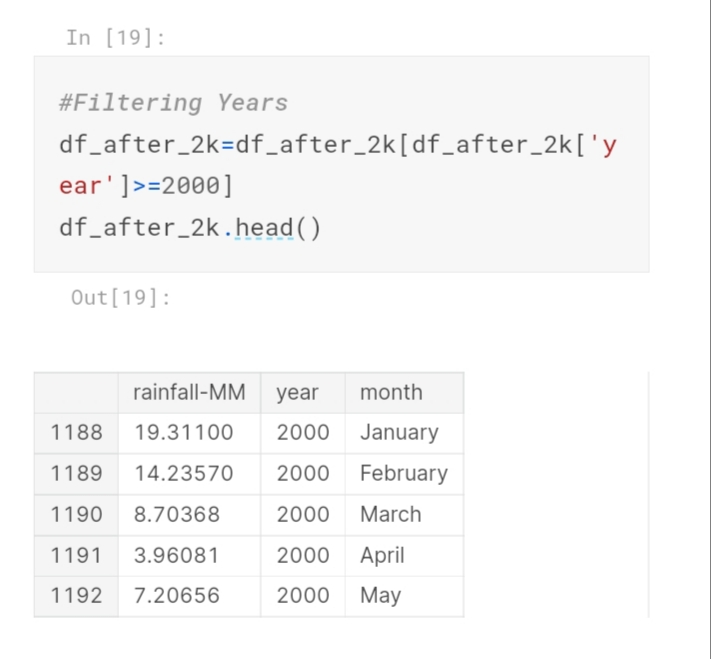
Now we are going to sort the data so that we can easily predict the top most rainfall year.



INPUT NO: 9

FILTERING THE DATA:

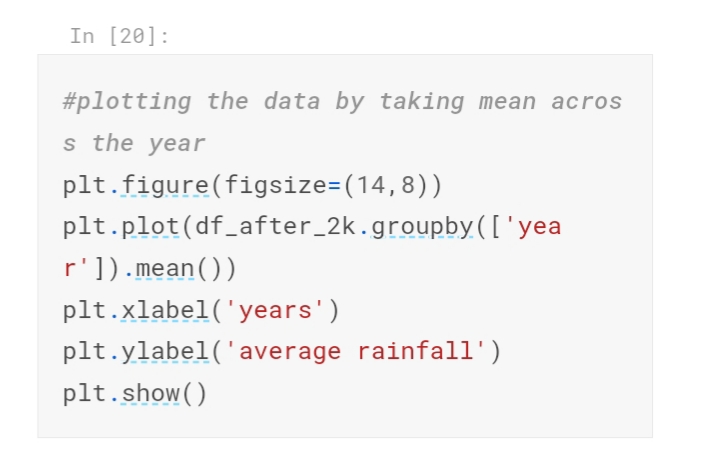


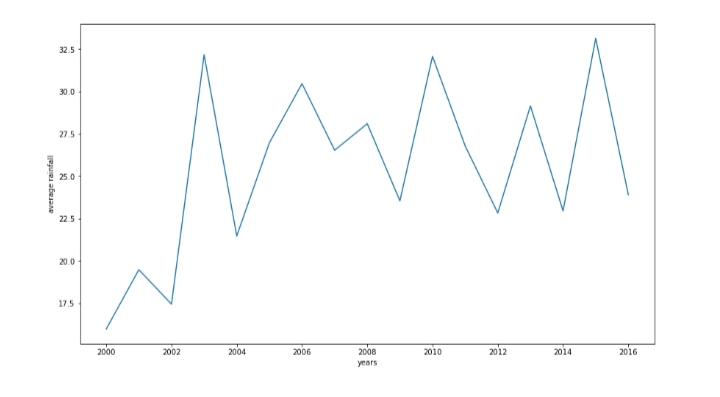


INPUT NO: 10

PLOTTING OF A GRAPH:

Now going to plot the graph of rainfall data on yearly basis.

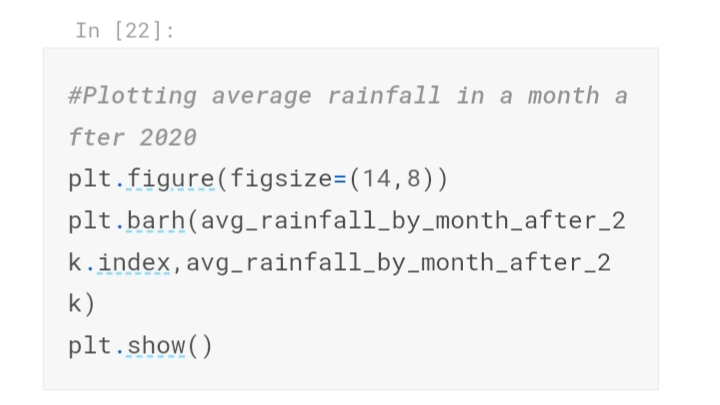


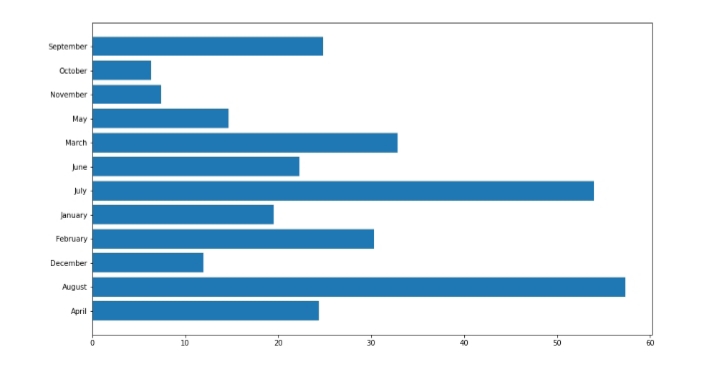


INPUT NO: 11

GRAPH PLOTTING:

Plotting the graph again of average rainfall occurred in a month after a year of 2020.



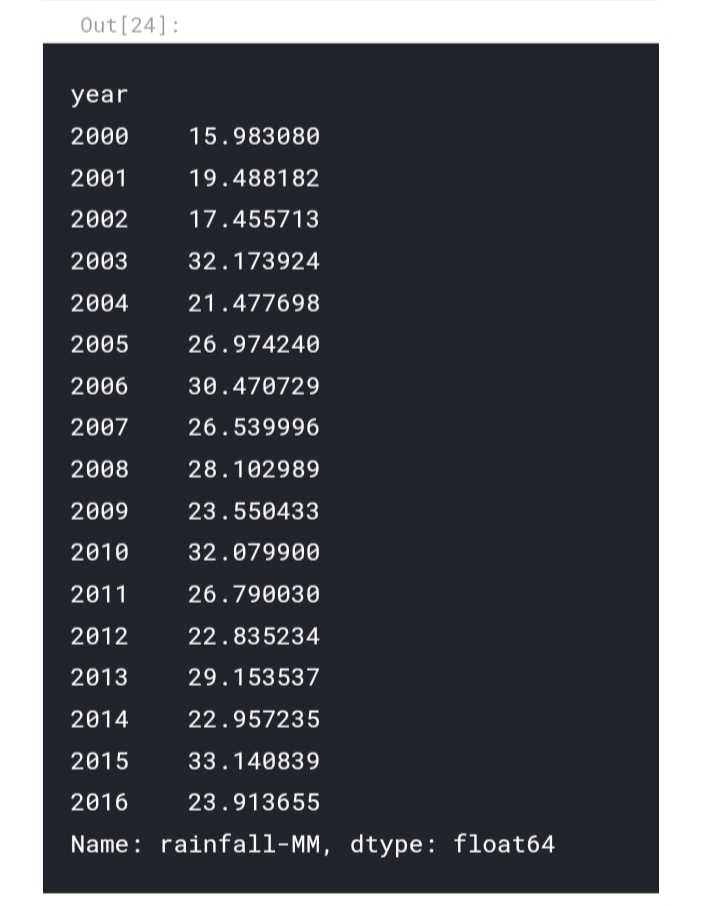


INPUT NO: 11

NUMERAL VALUE OF AVERAGE RAINFALL:

We will find the numeral values of average rainfall occurred from year 2000 to 2016.





INPUT NO: 12

SEASON WISE GRAPH PLOTTING:

Now we will plot the graph season wise which will show that which season receives highest rainfall.



