Name : Nidhi Sunil Patil
Class : TE-4
Batch : M4

Problem statement :

Use the following covid_vaccine_statewise.csv dataset and perform following analytics on the

given dataset- https://www.kaggle.com/sudalairajkumar/covid19-in-india?select=covid_vaccine_statewise.csv

a. Describe the dataset
b. Number of persons state wise vaccinated for first dose in India
c. Number of persons state wise vaccinated for second dose in India
d. Number of Males vaccinated d. Number of females vaccinated

Title : Data Visuallization and Analysis of data

Objective :

1) One will be able to understand the covid vaccination data and perform analytics on it.
2) One will be able to visuallize the covid vaccination data
3) Will be able to describe the covid database.

Outcome :

1) Able to perform analytics on covid dataset.
2) Able to perform a visuallization on a covid dataset.
3) Understood the various features of covid dataset.

Methodology :

Background:

Dataset- Covid_vaccine_statewise.csv

Covid-19 has affected our lives very much in very accepts it could be economical, mentally, etc. we have explored how the vaccination drive is going around the world. From the past 16 January 2021 to 15 August 2021. The number of new cases were increasing day by day around the world. This dataset has information from the states and union territories of India at daily level.

State level data comes from Ministry of Health & Family Welfare
Testing data and vaccination data comes from covid19india.

Implementation Details :

a. **Importing required libraries:**
For analyzing data, we need some libraries. In this section, we have imported all the required libraries like pandas, NumPy, matplotlib, plotly, seaborn that were required for data analysis.

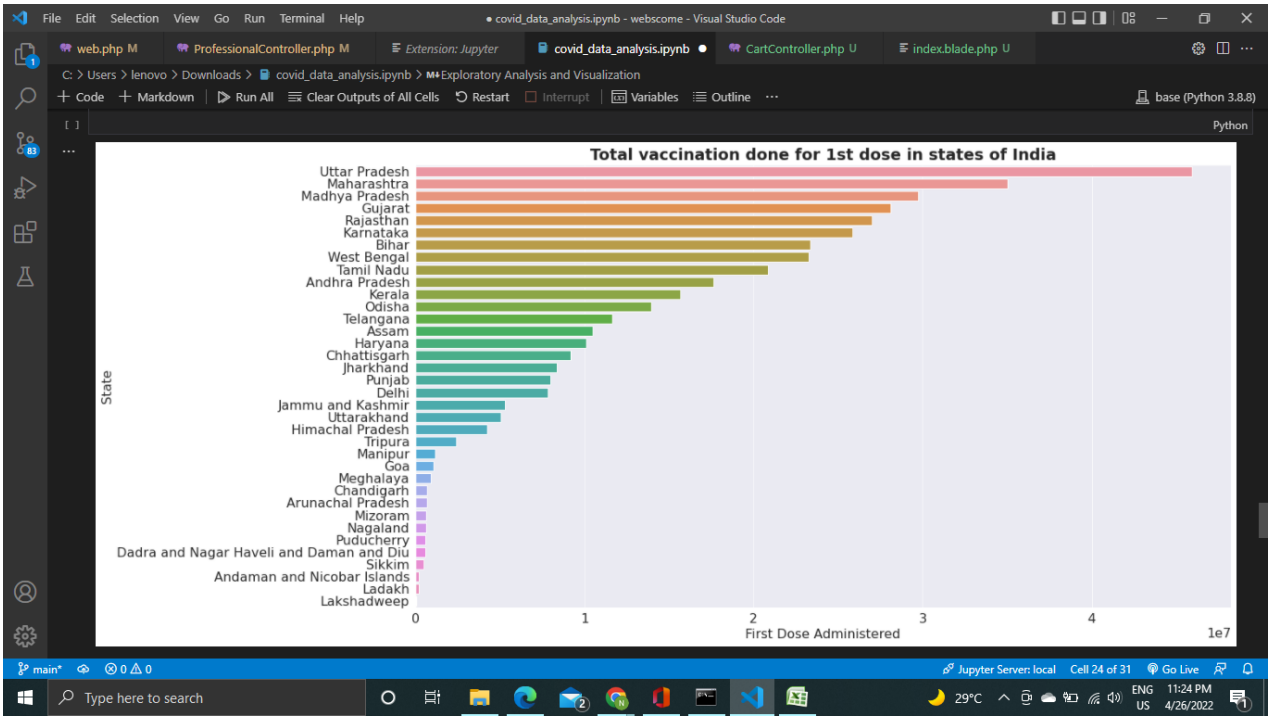
b. **Loading the dataset:**
Read the CSV file using pandas read_csv() function and show the output using head() function.

c. **Getting Basic Information of the dataset:**
Identify the dimensions of the dataset, identifying the columns in the dataset.

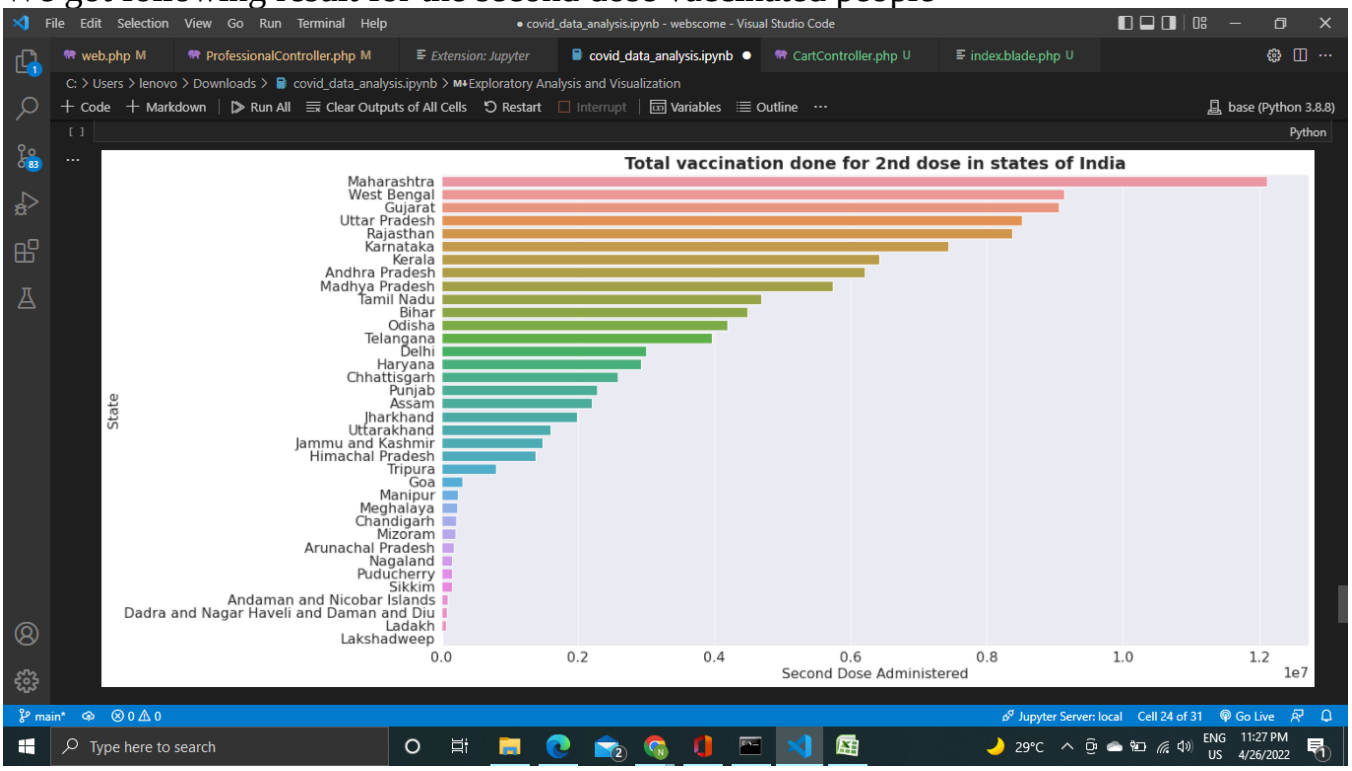
d. **Data Preperation and Cleaning:**
   1. Get an information about the dataset,
   2. Identify datatypes of the columns and found that the Updated on column is having the object datatype which we to change to the datetime.
   3. count the number of missing values in each column and found the technique to avoid the missing values that we filled them with the value 0.
   4. Found out all the unique values for all the states for which this dataset is created.
   5. In the listed states, India is also present which is not the state. May be it contains the sum of values in all states. But it will change the values in our analysis. So we will will drop the rows containing India as State by using `drop()` function.
      Here, after that our index has changed in the dataframe so resetted index using `reset_index` method

e. **Exploratory Analysis and Visuallization-**
In this section, we have explored relationships between columns by doing visualization using matplotlib and seaborn libraries of python.

We got following result for the 1st dose vaccinated people in this period

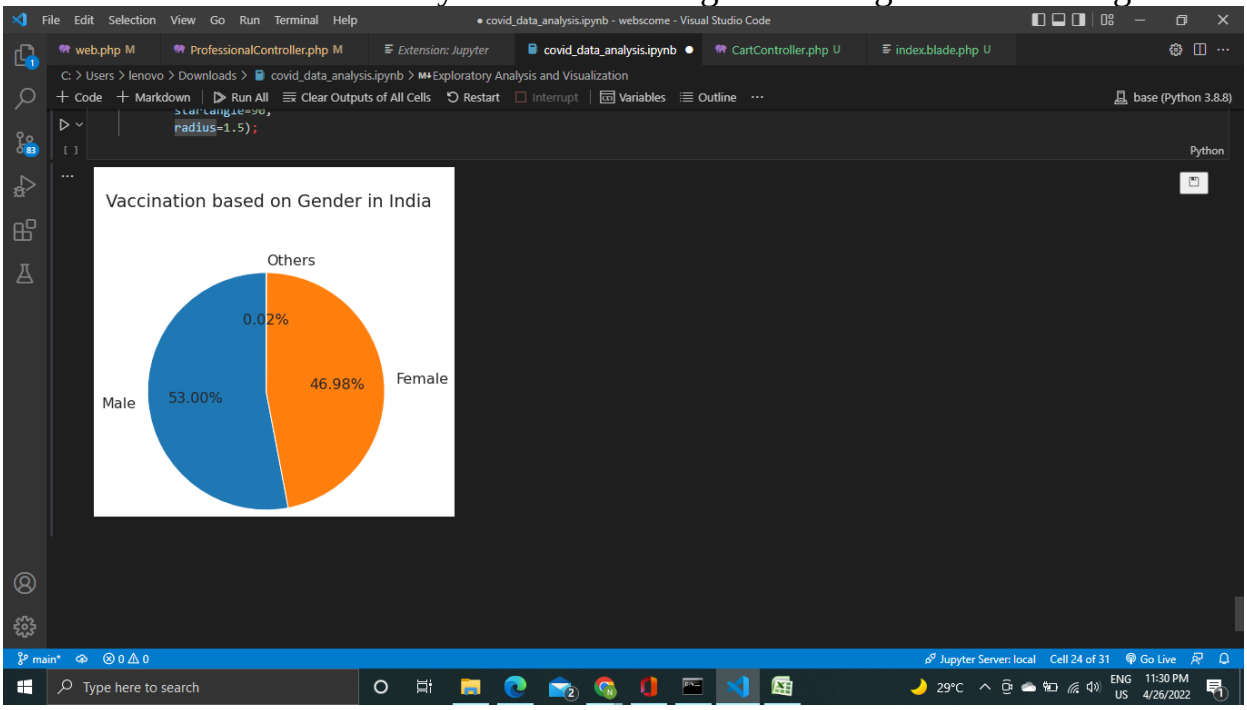


We got following result for the second dose vaccinated people



This is done using the barpot in the seaborn library.

After that found out the Analysis based on the gender and got the following results.



Observations-

1. Most of the people from Uttar Pradesh were vaccinated for the first dose of vaccine in this time interval.
2. Most of the people from Maharashtra were vaccinated for the second dose of the vaccine in this time interval.
3. Least Vaccinated People were from the Lakshadweep, Ladakh and Andaman and Nicobar.
4. Among the vaccinated people 46.98% were females and 53% were males and other contains 0.02%.

Conclusion

Hence we understood the Data Visualization and EDA of covid data generated daily and also done some data cleaning which was required.

References

1) https://www.tableau.com/learn/articles/data-visualization

2) https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15

3) https://medium.com/analytics-vidhya/data-cleaning-and-preprocessing-a4b751f4066f#:~:text=Data%20preprocessing%20involves%20the%20transformation,outcomes%20of%20any%20analytic%20algorithm.