A Mini Project

ON

“Covid 19 Vaccination Analysis & Prediction”

OF

TE (Computer) Academic Year: 2023-2024

Submitted By

Kohraj Gajanan Kakade

Guide By

Prof. Bhosale S.S.

DEPARTMENT OF COMPUTER ENGINEERING

HSBPVT’S FACULTY OF ENGINEERING, KASHTI

SAVITRIBAI PHULE PUNE UNIVERSITY

2023-2024

CERTIFICATE

This is to certify that the Mini Project Report entitled,

"Covid 19 Vaccination Analysis & Prediction"

has been carried out by **Mr. Kohraj Gajanan Kakade** Under the guidance of **Prof. Bhosale S.S.** in partial fulfilment of the Bachelor of Computer Engineering at HSBPVT'S Faculty of Engineering, Kashti has completed the mini project satisfactorily in course **Data Science & Big Data Analytics(310251)** for the academic year 2023-2024 as prescribed in the curriculum.

(Prof. Bhosale S.S)
Guide

Date :-
Place :- Kashti

Acknowledgement

I would like to express my deepest appreciation to all those who provided me the possibility to complete this report. A special gratitude I will give to my mini project guide,

Prof. Bhosale S. S., who invested her full effort in guiding me in achieving the goal.

My great obligation would remain towards our Head of Department **Prof. Hiranwale S.B.**, whose contribution in stimulating suggestions and encouragement helped us for writing report. He provided with an opportunity to undertake the Mini Project at HSBPVT's Faculty of Engg. Kashti. I am thankful for the guidance given by other staff members of Computer Engineering Department for improving my presentation skills.

I sincerely thank to our respected Principal who has proved to be a constant motivation for the knowledge acquisition and moral support during our course curriculum.

Abstract

In India, a large country of about 1.3 billion people, the disease covid 19 was first detected on 30 January 2020, in a student returning from Wuhan. The total number of confirmed infections in India as of 3 May 2020, is more than 37,000 and is currently growing fast. Most of the prior research and media coverage focused on the number of infections in the entire country. However, given the size and diversity of India, it is important to look at the spread of the disease in each state separately, where in the situations are quite different. In this report, we aim to analyse data on the number of infected people in each Indian state using csv dataset and predict the number of vaccinations for that state. We hope that such state wise predictions would help the state governments better channelize their limited health care resources.

Keywords :-

Prediction, Analysis, Vaccination, administered, Data Visualization, Data Cleaning, Data Handling.

Index

Sr.No	Topic Name	Page No.
1	Introduction	6
2	Problem Statement	7
3	Hardware & Software Requirement	7
4	Implementation Details	8
5	Dataset Information	10
6	Program Code and Output	11
8	Conclusion	15
9	References	15

Introduction

The highly infectious coronavirus disease (COVID-19) was first detected in Wuhan,China in December 2019 and subsequently spread to 212 countries and territories around the world, infecting millions of people. In India, a large country of about 1.3 billion people, the disease (covid 19) was first detected on 30 January 2020, in a student returning from Wuhan. The total number of confirmed infections in India as of 9 August 2021, is more than 37,000 and is currently growing fast.

An effective rollout of vaccinations against COVID-19 offers the most promising prospect of bringing the pandemic to an end. We present the Our World in Data COVID19 vaccination dataset, a global public dataset that tracks the scale and rate of the vaccine rollout across the Country. This dataset is updated regularly and includes data on the total number of vaccinations administered, first and second doses administered, Male (Doses Administered), Female (Doses Administered) Transgender (Doses Administered) for which data are available (28 state as of 9 August 2021). It will be maintained as the global vaccination campaign continues to progress. Our intention is to maintain the database for the foreseeable future and include additional State as they implement their vaccination campaigns. This dataset tracks the total number of COVID-19 vaccinations administered, Number of persons state wise vaccinated for first dose in India Number of persons state wise vaccinated for second dose in India Number of persons state wise vaccinated for second dose in India, Number of Males vaccinated, Number of females vaccinated each State. In this project using python libraries doing various operation on state wise Covid 19 vaccination dataset and in this project use csvfile dataset.

In this project for analysing and predicting the data, we need some libraries, We have imported all the required libraries like pandas, NumPy, matplotlib, portly, seaborn, and word cloud that are required for data analysis

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

System Requirements

Software Requirements :-

- Jupiter Notebook Programming Languages -Python.

Hardware Requirements :-

- Laptop/Desktop
- 4 GB Ram
- Internet Connectivity

Implementation Details

1. Importing Python Libraries :-

i. import numpy as np

NumPy (Numerical Python) is **an open-source library for the Python programming language**. It is used for scientific computing and working with arrays. Apart from its multidimensional array object, it also provides high-level functioning tools for working with arrays. In this tutorial, you will learn how to install NumPy.

ii. import pandas as pd

pandas is a Python library used for working with data sets. It has functions for analysing, cleaning, exploring, and manipulating data. Pandas allows usto analyse big data and make conclusions based on statistical theories. Pandas can clean messy data sets and make them readable and relevant.

iii. import sequential

A Sequential model is appropriate for a plain stack of layers where each layer has exactly one input tensor and one output tensor.

iv. import matplotlib

Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible.

v. import seaborn

Seaborn is an open-source Python library built on top of matplotlib. It is used for data visualization and exploratory data analysis.

1. Read Data and Basic Information :-

- Read the CSV file **covid_vaccine_statewise.csv** using pandas read_csv()function and show the output using head() function.

2. Data Cleaning :-

- Dataset has many null values. To get rid of it we need to clean the data first, after cleaning we will perform our further analysis.

3. Data Visualization :-

- we are going to draw some visuals to get insights fromour dataset.
- **describe()** function in pandas used to get the statistics of each feature present in our dataset. we get include count, max, min, standard deviation, median, etc.

No of Rows:-

- There are 7840 rows used in the dataset.

Column Names Of Dataset:-

- Updated On
- State
- Total Doses Administered
- Sessions
- Sites
- First Dose Administered
- Second Dose Administered
- Male (Doses Administered)
- Female (Doses Administered)
- Transgender (Doses Administered)
- Covaxin (Doses Administered)
- CoviShield (Doses Administered)
- Sputnik V (Doses Administered)
- AEFI
- 18-44 Years (Doses Administered)
- 45-60 Years (Doses Administered)
- 60+ Years (Doses Administered)
- 18-44 Years(Individuals Vaccinated)
- 45-60 Years(Individuals Vaccinated)
- 60+ Years(Individuals Vaccinated)
- Male(Individuals Vaccinated)
- Female(Individuals Vaccinated)
- Transgender(Individuals Vaccinated)
- Total Individuals Vaccinated

Program Code & Output

#imorting required libraries import
pandasas pd import numpy as np import
matplotlib.pyplot as plt

#reading the dataset cov=pd.read_csv('covid_vaccine_statewise.csv')cov.head()

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

In [1]:
```
#imorting required libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [6]:
```
#reading the dataset
cov=pd.read_csv('covid_vaccine_statewise.csv')
cov.head()
```

Out[6]:

	Updated On	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Female (Doses Administered)	Transgender (Doses Administered)	...	18-44 Years (Doses Administered)	45-60 Year (Dose Administere
0	16/01/2021	India	48276.0	3455.0	2957.0	48276.0	0.0	NaN	NaN	NaN	...	NaN	Na
1	17/01/2021	India	58604.0	8532.0	4954.0	58604.0	0.0	NaN	NaN	NaN	...	NaN	Na
2	18/01/2021	India	99449.0	13611.0	6583.0	99449.0	0.0	NaN	NaN	NaN	...	NaN	Na
3	19/01/2021	India	195525.0	17855.0	7951.0	195525.0	0.0	NaN	NaN	NaN	...	NaN	Na
4	20/01/2021	India	251280.0	25472.0	10504.0	251280.0	0.0	NaN	NaN	NaN	...	NaN	Na

5 rows × 24 columns

In [3]:
```
#to check which null values
cov.isnull().sum()
```

Out[3]:
```
Updated On                          0
State                               0
Total Doses Administered          224
Sessions                          224
 Sites                            224
First Dose Administered           224
Second Dose Administered          224
Male (Doses Administered)         384
Female (Doses Administered)       384
Transgender (Doses Administered)  384
 Covaxin (Doses Administered)     224
CoviShield (Doses Administered)   224
Sputnik V (Doses Administered)   4850
AEFI                             2407
18-44 Years (Doses Administered) 6143
45-60 Years (Doses Administered) 6143
60+ Years (Doses Administered)   6143
18-44 Years(Individuals Vaccinated) 4112
45-60 Years(Individuals Vaccinated) 4111
60+ Years(Individuals Vaccinated)   4111
Male(Individuals Vaccinated)        7685
Female(Individuals Vaccinated)      7685
Transgender(Individuals Vaccinated) 7685
Total Individuals Vaccinated        1926
dtype: int64
```

In [13]:
```
#checking for null values
cov.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7845 entries, 0 to 7844
Data columns (total 24 columns):
 #   Column                             Non-Null Count  Dtype
---  ------                             --------------  -----
 0   Updated On                         7845 non-null   object
 1   State                              7845 non-null   object
```

Describe the dataset :-

A) #to describe the dataset, we can use describe method as follow:cov.describe()

A)Describe the dataset

In [4]:
```
#to describe the dataset we can use describe method as follow:
cov.describe()
```

Out[4]:

	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Female (Doses Administered)	Transgender (Doses Administered)	Covaxin (Doses Administered)	CoviShield (Doses Administered)	...	A
count	7.621000e+03	7.621000e+03	7621.000000	7.621000e+03	7.621000e+03	7.461000e+03	7.461000e+03	7461.000000	7.621000e+03	7.621000e+03	...	1
mean	9.188171e+06	4.792358e+05	2282.872064	7.414415e+06	1.773755e+06	3.620156e+06	3.168416e+06	1162.978019	1.044669e+06	8.126553e+06	...	8
std	3.746180e+07	1.911511e+06	7275.973730	2.995209e+07	7.570382e+06	1.737938e+07	1.515310e+07	5931.353995	4.452259e+06	3.298414e+07	...	2
min	7.000000e+00	0.000000e+00	0.000000	7.000000e+00	0.000000e+00	0.000000e+00	2.000000e+00	0.000000	0.000000e+00	7.000000e+00	...	2
25%	1.356570e+05	6.004000e+03	69.000000	1.166320e+05	1.283100e+04	5.655500e+04	5.210700e+04	8.000000	0.000000e+00	1.331340e+05	...	4
50%	8.182020e+05	4.547000e+04	597.000000	6.614590e+05	1.388180e+05	3.897850e+05	3.342380e+05	113.000000	1.185100e+04	7.567360e+05	...	3
75%	6.625243e+06	3.428690e+05	1708.000000	5.387805e+06	1.166434e+06	2.735777e+06	2.561513e+06	800.000000	7.579300e+05	6.007817e+06	...	7
max	5.132284e+08	3.501031e+07	73933.000000	4.001504e+08	1.130780e+08	2.701636e+08	2.395186e+08	98275.000000	6.236742e+07	4.468251e+08	...	2

8 rows × 22 columns

Number of persons state wise vaccinated for first dose in India:

cov_state_vacc1=cov.set_index('State').group by(level=0)['First Dose Administered'].sum()

cov_state_vacc1

B) #to confirm the sum

cov_delhi=cov.where(cov['State']= ='Delhi')['First Dose Administered'].sum() print(cov_delhi)

print(cov_state_vacc1['Delhi']) cov_state_vacc1['Delhi']= =cov_delhi

In [68]:
```
#to confirm the sum
cov_delhi=cov.where(cov['State']=='Delhi')['First Dose Administered'].sum()
print(cov_delhi)
print(cov_state_vacc1['Delhi'])
cov_state_vacc1['Delhi']==cov_delhi
```

```
624339473.0
624339473.0
```

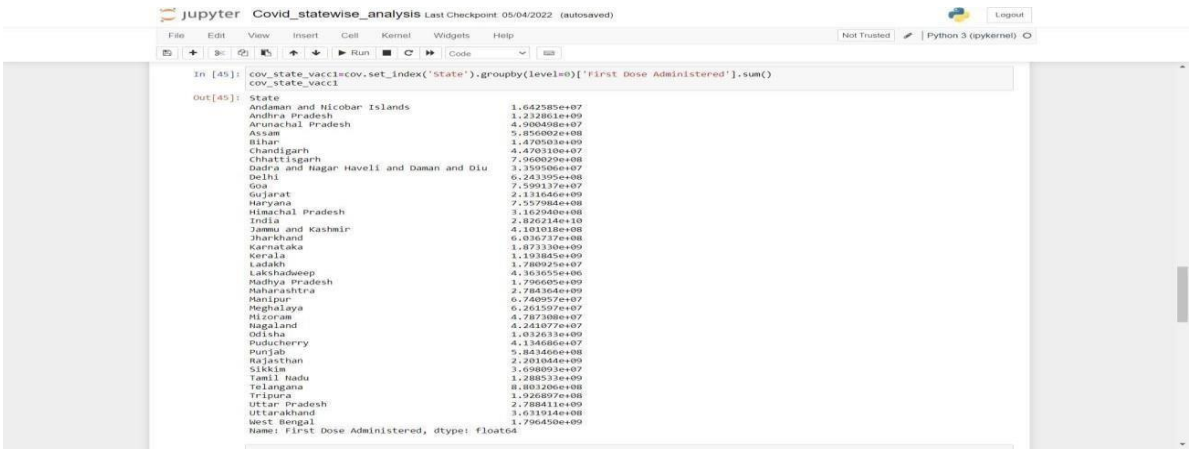
Out[68]: True

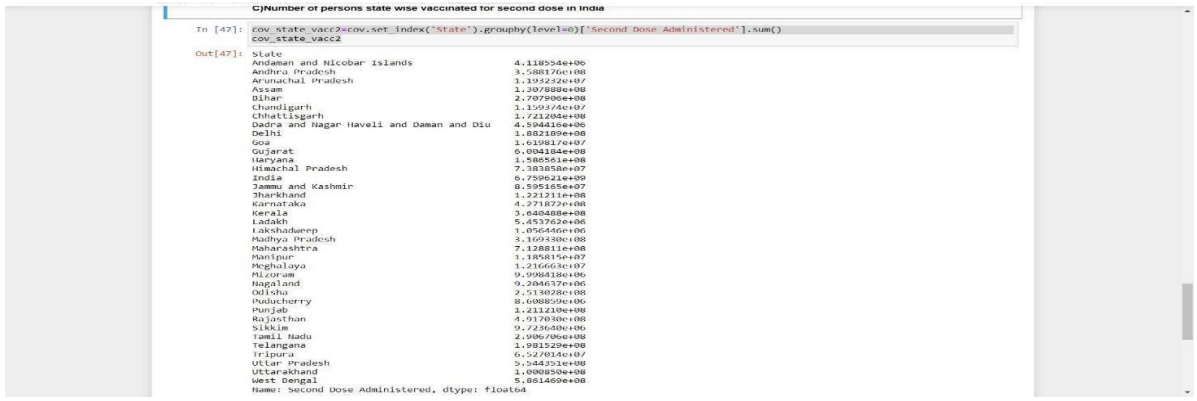
Number of persons state wise vaccinated for second dose in India:

cov_state_vacc2=cov.set_index('State').groupby(level=0)['Second Dose Administered'].sum()cov_state_vacc2

C) #to confirm :-

cov_MH=cov.where(cov['State']= ='Maharashtra')['Second Dose Administered'].sum() print(cov_MH) print(cov_state_vacc2['Maharashtra']) cov_state_vacc2['Maharashtra']= =cov_MH






```
In [71]: cov_MH=cov.where(cov['State']=='Maharashtra')['Second Dose Administered'].sum()
         print(cov_MH)
         print(cov_state_vacc2['Maharashtra'])
         cov_state_vacc2['Maharashtra']==cov_MH

         712881086.0
         712881086.0

Out[71]: True
```

D) Number of Males vaccinated:

cov['Male (Doses Administered)'].sum()

E) Number of females vaccinated:

cov['Female (Doses Administered)'].sum()

D) Number of Males vaccinated

```
In [74]: cov['Male (Doses Administered)'].sum()
Out[74]: 27009983996.0
```

E)Number of females vaccinated

```
In [75]: cov['Female (Doses Administered)'].sum()
Out[75]: 23639554465.0
```

Conclusion

This covid vaccination dataset provides vaccination status of covid 19 all over India. It shows result of state wise number of individual males and females vaccination. It shows the vaccination status state wise. Using visualization, its visuals to get insights from our dataset. Some of the information we get include Min, Max, count, Standard Deviation and Mean Etc. the DIR slowly become zero or negative for a consecutive 14 days to be able to declare the end of the pandemic. It gives the maximum correct accuracy of data.

References

1) Coronavirus disease: What you need to know. World Health Organization.

2) URL:https://www.afro.who.int/news/ coronavirus-disease-what-you-need-know.

3) URL:https://www.ecdc.europa.eu/sites/default/files/ documents/RRA-sixth-update Outbreak-of-novel-coronavirus- disease-2019-COVID- 19.pdf [accessed 2020-04- 05].

4) Novel coronavirus disease 2019 (COVID-19) pandemic: increased transmission in theEU/EEA and the UK – sixth update. European Centre for Disease Control and Prevention.