ESTD. : 2013

GURU GOBIND SINGH FOUNDATION'S

Guru Gobind Singh College of Engineering & Research Centre, Nashik

Department of Computer Engineering

Academic Year 2024-25

Third Year SEM VI

DSBDAL (Mini Project)

Class: TE CO-A

Title of Project: State-wise COVID-19 Vaccination Analysis

Group Member:

Roll No.	Name
59	Sanket Vilas Jadhav

Abstract:

This project undertakes a comprehensive analysis of the COVID-19 pandemic posed significant health and logistical challenges across the globe. One of the key responses to combat the virus was the vaccination drive, which required rigorous planning and data monitoring. This mini-project focuses on analyzing state-wise COVID-19 vaccination data in India to derive insights regarding the number of doses administered, gender-wise distribution, and trends. Python, along with data analysis and visualization libraries such as Pandas, Matplotlib, and Seaborn, were used to perform the analysis and uncover patterns in the vaccination rollout.

INDEX

1.Introduction Of Project

The COVID-19 pandemic posed significant health and logistical challenges across the globe. One of the key responses to combat the virus was the vaccination drive, which required rigorous planning and data monitoring. This mini-project focuses on analyzing state-wise COVID19 vaccination data in India to derive insights regarding the number of doses administered, gender-wise distribution, and trends. Python, along with data analysis and visualization libraries such as Pandas, Matplotlib, and Seaborn, were used to perform the analysis and uncover patterns in the vaccination rollout.

2. Objective:

- To perform exploratory data analysis (EDA) on the vaccination dataset to understand trends and patterns.

- To derive meaningful insights from state-wise vaccination statistics for better understanding of regional differences.

- To visualize the data using graphs and charts for easy interpretation and comparison.

- To understand gender-based vaccination distribution and identify states with disparities.

3. Dataset Used:

- Dataset Name: StatewiseTestingDetails.csv
- Source: Extracted from archive.zip located in the student's Google Drive
- File Path: /content/drive/MyDrive/dataset1/archive.zip
- Description: The dataset contains daily state-wise data related to COVID-19 testing and vaccination, including the number of doses administered and gender-based distribution.

4. Tools and Technologies Used:

The analysis for this project would typically involve using the following tools and technologies:

- **Python:** A versatile programming language widely used for data analysis.
- **Pandas:** A powerful Python library for data manipulation and analysis.
- **NumPy:** A fundamental package for numerical computation in Python. • **Matplotlib and Seaborn:** Python libraries for creating visualizations and plots to understand trends and patterns in the data.
- **Jupyter Notebook/Google Colab:** Interactive environments for writing and executing code, and for creating reports with embedded code, visualizations, and text explanations.

5. Methodology:

1. Data Acquisition: Google Drive was mounted in Google Colab, and the ZIP file was extracted.
2. Data Loading: The StatewiseTestingDetails.csv file was read using Pandas.
3. Data Exploration: The structure and contents of the dataset were examined. Missing values and data types were identified.
4. Data Aggregation: State-wise grouping was performed for vaccine dose data and gender-wise counts.
5. Visualization: Bar charts and pie charts were generated to illustrate state-wise vaccinations and male-female ratios.

Code Highlights:

```
# Load and explore data
import pandas as pd import
matplotlib.pyplot as plt import
seaborn as sns import zipfile
import os

# Mount Google Drive from
google.colab import drive

# Extract and read CSV drive.mount('/content/drive')
zip_path = "/content/drive/MyDrive/dataset1/archive.zip"
extract_path = "/content/dataset_extracted" with
```

```python
zipfile.ZipFile(zip_path, 'r') as zip_ref:
zip_ref.extractall(extract_path)
csv_path = os.path.join(extract_path, "StatewiseTestingDetails.csv")
df = pd.read_csv(csv_path)

# Visualization example df.groupby('State')['First
Dose
Administered'].sum().sort_values().plot(kind='barh')
plt.title('First Dose Administered by State') plt.xlabel('Number
of Doses')
plt.ylabel('State')
plt.tight_layout() plt.show()
```

```
Extracted Files:
['StatewiseTestingDetails.csv', 'covid_vaccine_statewise.csv', 'covid_19_india.csv']
Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16336 entries, 0 to 16335
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Date          16336 non-null  object
 1   State         16336 non-null  object
 2   TotalSamples  16336 non-null  float64
 3   Negative      6969 non-null   object
 4   Positive      5662 non-null   float64
dtypes: float64(2), object(3)
memory usage: 638.3+ KB
None

Missing Values:
Date               0
State              0
TotalSamples       0
Negative        9367
Positive       10674
dtype: int64

Summary Stats:
        TotalSamples      Positive
count  1.633600e+04  5.662000e+03
mean   5.376466e+06  5.652654e+04
std    8.780338e+06  1.673108e+05
min    5.800000e+01  0.000000e+00
25%    1.729098e+05  5.362500e+02
50%    9.307580e+05  4.771000e+03
```
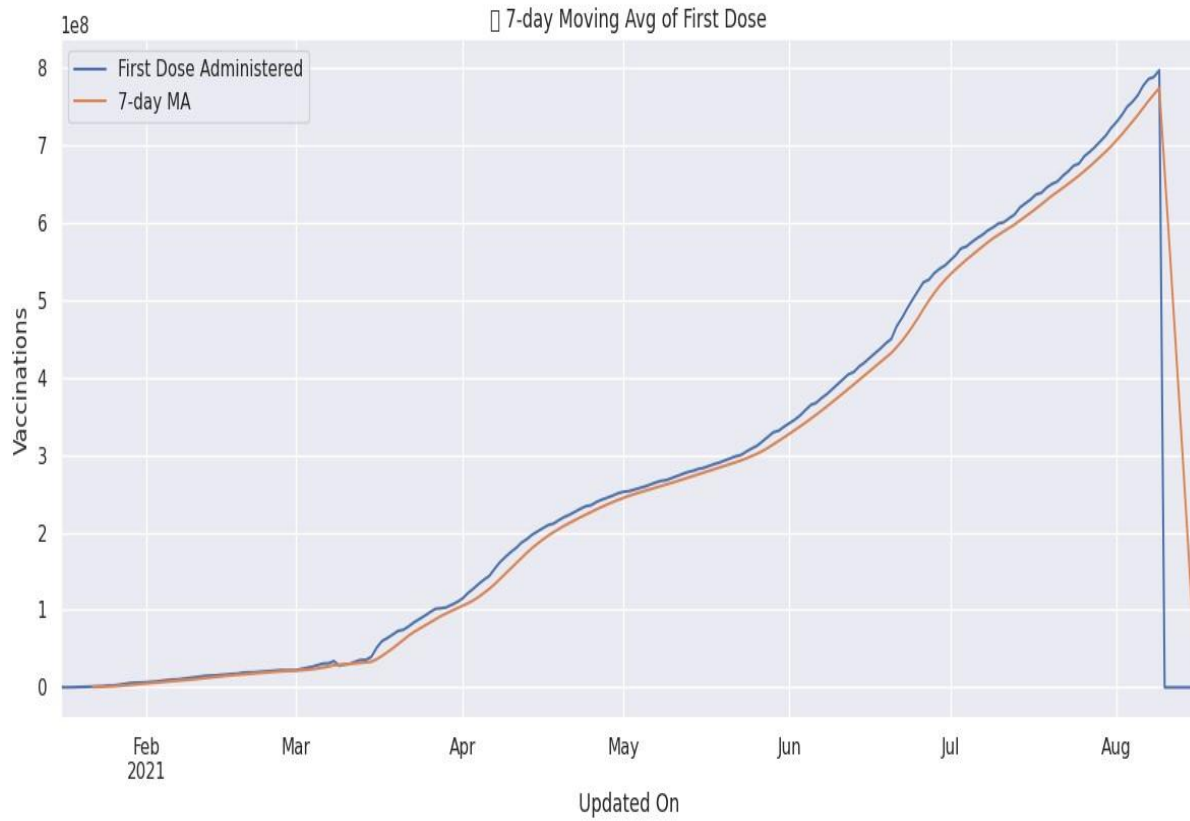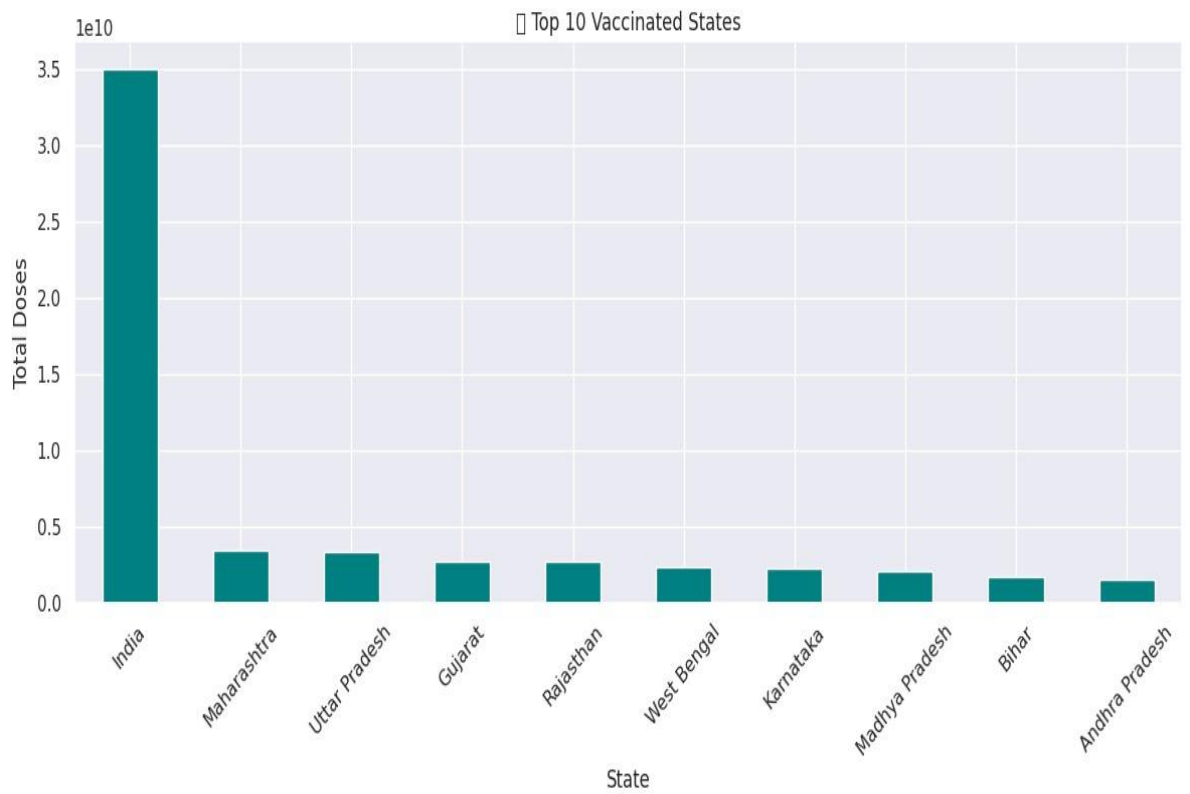
Top 10 Vaccinated States

7-day Moving Avg of First Dose

6. Conclusion:

This project highlighted the power of data science in public health analysis. By exploring vaccination trends across Indian states, it was possible to uncover patterns and disparities. Python's data analysis libraries proved to be highly effective for handling large datasets, and visualizations made it easy to communicate findings. Overall, the analysis promotes awareness and can support government efforts in managing health campaigns.

7. Future Work:

Further analysis could explore:

- **Real-time Data Integration:** Connecting to live government APIs for up-to-date vaccination data.
- **Predictive Analysis:** Using machine learning models to forecast future vaccination rates and identify lagging regions.
- **Interactive Dashboards:** Creating interactive dashboards using Plotly, Dash, or Tableau to allow dynamic data exploration.
- **Expanded Metrics:** Including more features such as age groups, urban-rural classification, and vaccine types for deeper analysis.

8. References:

- https://www.kaggle.com
- https://pandas.pydata.org
- https://matplotlib.org
- https://colab.research.google.com
- Government of India COVID-19 Portal