MINI-PROJECT REPORT ON

"COVID VACCINE STATE ANALYTICS"

Submitted By

Mr. Thorve Avishkar Shrikrushna

Roll No: 63

Exam No: T400840394

Class: TE AIDS

UNDER THE GUIDANCE OF

Prof. Rokade N. G.

Department of Artificial Intelligence and Data Science Engineering

Jaihind College of Engineering, Kuran

A/p-Kuran, Tal-Junnar, Dist-Pune-410511, State Maharashtra, India

2024-2025

Department of Artificial Intelligence and Data Science Engineering

Jaihind College of Engineering, Kuran

A/p-Kuran, Tal-Junnar, Dist-Pune-410511, State Maharashtra, India

2024-2025

CERTIFICATE

This is to certify that the Internship Report Entitled

"COVID VACCINE STATE ANALYTICS"

SUBMITTED BY

Thorve Avishkar Shrikrushna

Is a Bonafide work carried out by student under the supervision of **Prof. Rokade N. G.** and it is submitted towards the partial fulfilment of the requirement of third year of Engineering in Artificial intelligence and data science under the Savitribai Phule Pune University during the academic year 2024-2025.

Prof. Rokade N. G.	**Prof. S. K. Said**
(Subject Teacher)	**(**HOD AI&DS)
JCOE, Kuran.	JCOE, Kuran.

ACKNOWLEDGEMENTS

With deep sense of gratitude we would like to thank all the people who have lit our path with their kind guidance. We are very grateful to these intellectuals who did their best to help during our project work. It is our proud privilege to express a deep sense of gratitude to Principal of Jaihind College Of Engineering, Kuran **Dr. D. J. Garkal** for his comments and kind permission to complete this project. We remain indebted to H.O.D. Artificial Intelligence And Data Science Engineering Department **Prof. S. K. Said,** for his timely suggestion and valuable guidance. The special gratitude goes to **Prof. Rokade N. G.** excellent and precious guidance in completion of this work .We thanks to all the colleagues for their appreciable help for our working project. With various industry owners or lab technicians to help, it has been our endeavor throughout our work to cover the entire project work. We are also thankful to our parents who provided their wishful support for our project completion successfully .And lastly we thank our all friends and the people who are directly or indirectly related to our project work.

Yours faithfully,

Mr. Thorve Avishkar Shrikrushna.

ABSTRACT

The COVID-19 pandemic has brought about an unprecedented global crisis, compelling rapid development and deployment of vaccines to curb its spread. This data science mini project aims to analyze the state of COVID-19 vaccination across different regions, focusing on various metrics such as vaccination rates, distribution disparities, and effectiveness.

Utilizing publicly available datasets sourced from reliable repositories, the project will employ data cleaning, preprocessing, and exploratory data analysis techniques to extract meaningful insights. This project seeks to provide valuable insights into the progress and challenges of COVID-19 vaccination efforts, informing policymakers, healthcare professionals, and the general public. Ultimately, the findings aim to contribute to evidence-based decision-making strategies for optimizing vaccination campaigns and mitigating the impact of the pandemic.

TABLE OF CONTENTS

CHAPTER 01
INTRODUCTION

The COVID-19 pandemic has spurred unprecedented global efforts to develop and administer vaccines to combat the spread of the virus. In India, a comprehensive vaccination campaign has been initiated to inoculate its vast population against COVID-19. Understanding the progress and distribution of vaccinations across different states is crucial for effectively managing the pandemic.

This study aims to analyze the COVID-19 vaccination data in India using the "covid_vaccine_statewise.csv" dataset available on Kaggle. The dataset provides valuable insights into the number of individuals vaccinated for the first and second doses, as well as the gender distribution of vaccinated individuals across various states in India.

Through this analysis, we seek to address the following key aspects:

a. Describe the Dataset: An overview of the dataset structure, including its columns and attributes, will be provided to understand the available information.

b. Number of Persons State-wise Vaccinated for First Dose: We will calculate and analyze the total number of individuals vaccinated for the first dose in each state across India.

c. Number of Persons State-wise Vaccinated for Second Dose: Similarly,. we will determine the total number of individuals vaccinated for the second dose in each state.

d. Number of Males and Females Vaccinated: This analysis requires summing up the "Males_Vaccinated" and "Females_Vaccinated" columns respectively to get the total number of males and females vaccinated.

1.1 MOTIVATION

The motivation behind the COVID vaccine state analytics mini project has had a profound impact on global health and economies. Understanding the distribution and effectiveness of COVID-19 vaccines is crucial for controlling the spread of the virus, minimizing its impact on public health, and facilitating the return to normalcy. Analyzing vaccination data allows policymakers and healthcare authorities to optimize vaccination strategies by identifying areas with low vaccination rates, addressing disparities in vaccine distribution, and implementing targeted interventions to improve coverage.

1.2 AIM

The aim of the COVID state analytics mini project is to delve into the intricate dynamics of the COVID-19 pandemic at the state level, leveraging data-driven insights to understand its spread, impact, and associated factors. Through comprehensive data collection from reliable sources such as government health departments and research organizations, the project aims to curate a dataset encompassing various metrics including confirmed cases, deaths, recoveries, testing rates, vaccination rates, and demographic information.

1.3 OBJECTIVE

1. Gather comprehensive data on COVID-19 cases, testing, vaccinations, hospitalizations, and demographic information at the state level.

2. Clean and preprocess the collected data to ensure accuracy and consistency, handling missing values and outliers appropriately.

3. Conduct exploratory data analysis (EDA) to identify trends, patterns, and correlations among different states regarding COVID-19 spread, impact, and associated factors.

4. Utilize statistical analysis techniques to quantify relationships, assess significance, and derive actionable insights from the data.

1.4 PROBLEM STATEMENT

Use the following covid_vaccine_statewise.csv dataset and perform following analytics on the given dataset 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, & Number of females vaccinated.

CHAPTER 02
SYSTEM REQUIREMENT SPECIFICATIONS

2.1 SOFTWARE REQUIREMENT :

1. Python:

Essential for data analysis, visualization, and statistical modeling.

Packages such as Pandas, NumPy, and Matplotlib/Seaborn for data manipulation and visualization.

2. Jupyter Notebook or JupyterLab:

Provides an interactive environment for data exploration, analysis, and documentation.

Facilitates iterative development and collaboration.

3. Statistical Analysis Tools:

SciPy and Statsmodels for statistical analysis and hypothesis testing.

Scikit-learn for machine learning tasks if predictive modeling is included.

4. Documentation Tools:

Markdown for writing documentation and reports within Jupyter notebooks.

2.2 HARDWARE REQUIREMENT :

1. Computer:

A desktop or laptop computer capable of running the required software.

2. Operating System:

Windows, macOS, or Linux distributions such as Ubuntu or CentOS are commonly used.

3. Memory (RAM):

Minimum of 8 GB RAM is recommended, but higher capacities (16 GB or more) may be beneficial for handling larger datasets and complex analyses.

4. Storage:

Sufficient storage space to store datasets, software installations, and project files.

CHAPTER 04
IMPLEMENTATION

To implement the analytics on the "covid_vaccine_statewise.csv" dataset, we'll follow these steps:

a. Describe the dataset: We'll start by loading the dataset using Pandas and then use descriptive functions like info() and describe() to provide an overview of the dataset, including column names, data types, and summary statistics.

b. Number of persons state wise vaccinated for first dose in India: We'll group the dataset by the "State" column and calculate the sum of "First Dose Administered" for each state to determine the total number of individuals vaccinated for the first dose in each state.

c. Number of persons state wise vaccinated for second dose in India: Similar to step (b), but this time we'll calculate the sum of "Second Dose Administered" for each state to determine the total number of individuals vaccinated for the second dose in each state.

d. Number of Males vaccinated & Number of females vaccinated: We'll sum up the values in the "Male(Individuals Vaccinated)" and "Female(Individuals Vaccinated)" columns respectively to determine the total number of males and females vaccinated.

3.1 ALOGORITHMIC STEPS

1. Load Dataset:

Use Pandas to read the dataset from the provided URL into a DataFrame.

2. Describe the Dataset:

Utilize functions like info() and describe() to provide an overview of the dataset, including column names, data types, and summary statistics.

3. Calculate Number of Persons Vaccinated for First Dose State-wise:

Group the DataFrame by the "State" column.

Aggregate the "First Dose Administered" column using the sum() function to calculate the total number of individuals vaccinated for the first dose in each state.

4. Calculate Number of Persons Vaccinated for Second Dose State-wise:

Group the DataFrame by the "State" column.

Aggregate the "Second Dose Administered" column using the sum() function to calculate the total number of individuals vaccinated for the second dose in each state.

5. Calculate Number of Males and Females Vaccinated:

Sum the values in the "Male(Individuals Vaccinated)" column to get the total number of males vaccinated.

Sum the values in the "Female(Individuals Vaccinated)" column to get the total number of females vaccinated.

7. Display Results:

Print or display the results obtained from steps 3, 4, and 5 to provide the desired analytics insights.

CHAPTER 04

RESULT

1. DESCRIBE THE DATASET.

```
In [1]: # import requried Libraries

In [2]: import pandas as pd
        import numpy as np

In [3]: # Load the Dataset and Read this dataset

In [13]: df = pd.read_csv(r"C:\Users\saira\Downloads\archive\covid_vaccine_statewise.csv")

In [5]: # Describe the dataset

In [14]: df.describe()
```

Out[14]:

	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)
count	7.621000e+03	7.621000e+03	7621.000000	7.621000e+03	7.621000e+03	7.461000e+03
mean	9.188171e+06	4.792358e+05	2282.872064	7.414415e+06	1.773755e+06	3.620156e+06
std	3.746180e+07	1.911511e+06	7275.973730	2.995209e+07	7.570382e+06	1.737938e+07
min	7.000000e+00	0.000000e+00	0.000000	7.000000e+00	0.000000e+00	0.000000e+00
25%	1.356570e+05	6.004000e+03	69.000000	1.166320e+05	1.283100e+04	5.655500e+04
50%	8.182020e+05	4.547000e+04	597.000000	6.614590e+05	1.388180e+05	3.897850e+05
75%	6.625243e+06	3.428690e+05	1708.000000	5.387805e+06	1.166434e+06	2.735777e+06
max	5.132284e+08	3.501031e+07	73933.000000	4.001504e+08	1.130780e+08	2.701636e+08

8 rows × 22 columns

In [15]:
```
df.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 
 2   Total Doses Administered              7621 non-null   float64
 3   Sessions                              7621 non-null   float64
 4    Sites                                7621 non-null   float64
 5   First Dose Administered               7621 non-null   float64
 6   Second Dose Administered              7621 non-null   float64
 7   Male (Doses Administered)             7461 non-null   float64
 8   Female (Doses Administered)           7461 non-null   float64
 9   Transgender (Doses Administered)      7461 non-null   float64
 10   Covaxin (Doses Administered)         7621 non-null   float64
 11  CoviShield (Doses Administered)       7621 non-null   float64
 12  Sputnik V (Doses Administered)        2995 non-null   float64
 13  AEFI                                  5438 non-null   float64
 14  18-44 Years (Doses Administered)      1702 non-null   float64
 15  45-60 Years (Doses Administered)      1702 non-null   float64
 16  60+ Years (Doses Administered)        1702 non-null   float64
 17  18-44 Years(Individuals Vaccinated)   3733 non-null   float64
 18  45-60 Years(Individuals Vaccinated)   3734 non-null   float64
 19  60+ Years(Individuals Vaccinated)     3734 non-null   float64
 20  Male(Individuals Vaccinated)          160 non-null    float64
 21  Female(Individuals Vaccinated)        160 non-null    float64
 22  Transgender(Individuals Vaccinated)   160 non-null    float64
 23  Total Individuals Vaccinated          5919 non-null   float64
dtypes: float64(22), object(2)
memory usage: 1.4+ MB
```

In [52]:
```
print(df.columns)
```

```
Index(['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'],
      dtype='object')
```

2. NUMBER OF PERSONS STATE WISE VACCINATED FOR FIRST DOSE IN INDIA.

In [20]:
```
# Number of persons state wise vaccinated for first dose in India
```

In [21]:
```
first_dose_vaccinated = df.groupby('State')['First Dose Administered'].sum()
first_dose_vaccinated
```

Out[21]:

State	
Andaman and Nicobar Islands	1.642585e+07
Andhra Pradesh	1.232861e+09
Arunachal Pradesh	4.900498e+07
Assam	5.856002e+08
Bihar	1.470503e+09
Chandigarh	4.470310e+07
Chhattisgarh	7.960029e+08
Dadra and Nagar Haveli and Daman and Diu	3.359506e+07
Delhi	6.243395e+08
Goa	7.599137e+07
Gujarat	2.131646e+09
Haryana	7.557984e+08
Himachal Pradesh	3.162940e+08
India	2.826214e+10
Jammu and Kashmir	4.101018e+08
Jharkhand	6.036737e+08
Karnataka	1.873330e+09
Kerala	1.193845e+09
Ladakh	1.780925e+07
Lakshadweep	4.363655e+06
Madhya Pradesh	1.796605e+09
Maharashtra	2.784364e+09
Manipur	6.740957e+07
Meghalaya	6.261597e+07
Mizoram	4.787308e+07
Nagaland	4.241077e+07
Odisha	1.032633e+09
Puducherry	4.134686e+07
Punjab	5.843466e+08
Rajasthan	2.201044e+09
Sikkim	3.698093e+07
Tamil Nadu	1.288533e+09
Telangana	8.803206e+08
Tripura	1.926897e+08
Uttar Pradesh	2.788411e+09
Uttarakhand	3.631914e+08
West Bengal	1.796450e+09

Name: First Dose Administered, dtype: float64

3. NUMBER OF PERSONS STATE WISE VACCINATED FOR SECOND DOSE IN INDIA.

In [22]:
```
# Number of persons state wise vaccinated for second dose in India
```

In [23]:
```
second_dose_vaccinated = df.groupby('State')['Second Dose Administered'].sum()
second_dose_vaccinated
```

Out[23]:

State	
Andaman and Nicobar Islands	4.118554e+06
Andhra Pradesh	3.588176e+08
Arunachal Pradesh	1.193232e+07
Assam	1.307888e+08
Bihar	2.707906e+08
Chandigarh	1.159374e+07
Chhattisgarh	1.721204e+08
Dadra and Nagar Haveli and Daman and Diu	4.594416e+06
Delhi	1.882189e+08
Goa	1.619817e+07
Gujarat	6.004184e+08
Haryana	1.586561e+08
Himachal Pradesh	7.383858e+07
India	6.759621e+09
Jammu and Kashmir	8.595165e+07
Jharkhand	1.221211e+08
Karnataka	4.271872e+08
Kerala	3.640488e+08
Ladakh	5.453762e+06
Lakshadweep	1.056446e+06
Madhya Pradesh	3.169330e+08
Maharashtra	7.128811e+08
Manipur	1.185815e+07
Meghalaya	1.216663e+07
Mizoram	9.998418e+06
Nagaland	9.204637e+06
Odisha	2.513028e+08
Puducherry	8.608859e+06
Punjab	1.211210e+08
Rajasthan	4.917030e+08
Sikkim	9.723640e+06
Tamil Nadu	2.906706e+08
Telangana	1.981529e+08
Tripura	6.527014e+07
Uttar Pradesh	5.544351e+08
Uttarakhand	1.000850e+08
West Bengal	5.861469e+08

Name: Second Dose Administered, dtype: float64

4. NUMBER OF MALES VACCINATED & NUMBER OF FEMALES VACCINATED.

In [44]:
```
# Number of Males vaccinated
```

In [8]:
```
print("Number of Males vaccinated")
males_vaccinated = df['Male(Individuals Vaccinated)'].sum()
print(males_vaccinated)
```

```
Number of Males vaccinated
7138698858.0
```

In [9]:
```
print("Number of Females vaccinated")
Females_vaccinated = df['Female(Individuals Vaccinated)'].sum()
print(Females_vaccinated)
```

```
Number of Females vaccinated
6321628736.0
```

CHAPTER 05
CONCLUSION

After obtaining the above results, you can draw conclusions based on the analysis. For example, you can compare vaccination rates between states, assess gender distribution in vaccination, identify any trends over time if the dataset includes timestamps, and so on. Overall, by leveraging data-driven insights from state-wise COVID vaccine analytics, stakeholders can make informed decisions to enhanOverall, the COVID vaccine state analytics mini-project serves as a valuable tool for understanding the dynamics of vaccination efforts in India and informing evidence-based decision-making to combat the COVID-19 pandemic effectively.ce vaccination efforts, mitigate the spread of the virus, and safeguard public health in India.

CHAPTER 06

FUTURE SCOPE

The COVID vaccine state analytics mini-project offers numerous opportunities for future development and expansion. Here are some potential avenues for further exploration:

1.Temporal Analysis:

Incorporate time-series analysis to track vaccination rates over time and identify trends, seasonality, and patterns in vaccine uptake. This could involve analyzing daily, weekly, or monthly vaccination data to monitor progress and assess the impact of vaccination campaigns and policy changes.

2.Predictive Modeling:

Develop predictive models to forecast future vaccination rates based on historical data, demographic factors, disease prevalence, and other relevant variables. These models can help anticipate future vaccination needs, optimize resource allocation, and inform decision-making.

3.Geospatial Analysis:

Utilize geospatial analysis techniques to map vaccination coverage and identify geographic areas with low vaccination rates or underserved populations. This can aid in targeting interventions, allocating resources, and implementing localized vaccination strategies.

4.Demographic Insights:

Explore the demographic characteristics of vaccinated populations, including age, gender, occupation, and socio-economic status. Analyzing demographic trends can help identify vulnerable populations, assess equity in vaccine distribution, and tailor outreach efforts to address specific needs.

REFERENCE

[1] Ministry of Health and Family Welfare, Government of India, "COVID-19 Vaccine Statewise Data," Jan. 2024. [Online]. Available: [Dataset Link].

[2] World Health Organization (WHO), "COVID-19 Vaccines," [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/covid-19-vaccines.

[3] Centers for Disease Control and Prevention (CDC), "COVID-19 Vaccines," [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/vaccines/index.html.

[4] N. I. Khachfe, A. A. Chahrour, and M. S. Mikati, "Characteristics of COVID-19 Vaccines and Their Efficacy," Molecular Therapy - Methods & Clinical Development, vol. 20, pp. 1-10, Mar. 2021. DOI: 10.1016/j.omtm.2021.01.015.