Data science Big Data Analytics Project on Covid Vaccine Statewise

in

Third Year Engineering - Computer Engineering

by

Hake Pratiksha
Sathe Shweta
Asalkar Sharayu
Gulme Aishwarya

Under the guidance of

Prof. Rupali Wagh

DEPARTMENT OF COMPUTER ENGINEERING

PARVATIBAI GENBA MOZE COLLEGE OF ENGINEERING - 412207

CERTIFICATE

This is to certify that seminar entitled "**Data science Big Data Analytics Project on Covid Vaccine Statewise** ", submitted by **Hake Pratiksha**, **Sathe Shweta**, **Asalkar Sharayu**, **Gulme Aishwarya**an undergraduate student of **T.Y. Engineering (2022-23)** in partial fulfillment for the award of degree of **Bachelor of Engineering** with specialization of **Computer Engineering**. To the best of my knowledge and belief this work has not been submitted elsewhere for the award of any other degree.

. . . .
Head of Department
Prof. Shrikant Dhamdhere

. . . .
Seminar Guide
Prof. Rupali Wagh

CERTIFICATE OF APPROVAL

The forgoing thesis report is hereby approved as a creditable study of "**Thesis Title**
"carried out and presented satisfactorily to warrant its acceptance as a pre-requisite
for the Degree of Master of Technology of University. It is understood that by this
approval the undersigned do not necessarily approve any statement made, opinion
expressed and conclusion drawn there in but approve the progress report only for the
purpose for which it is submitted.

Examiners:
1. . . . .
2. . . . .   **Place: Pune**
3. . . . .   **Date:. . . . .**

ACKNOWLEDGEMENTS

ABSTRACT

The given project involves analyzing the COVID-19 vaccine dataset for India as of 7th April 2021. The dataset contains information about the number of people vaccinated for COVID-19 in each state or union territory in India, the number of people vaccinated for the first and second doses, the percentage of the population vaccinated, and the number of doses administered per day.

The project involves performing various analytics on the dataset, including calculating the total number of people vaccinated for the first and second doses, the number of males and females vaccinated, and the number of people vaccinated in each state.

Overall, this project provides valuable insights into the COVID-19 vaccination efforts in India, which can be used to guide policy decisions and improve vaccination strategies to curb the spread of the disease.

Chapter 1

Introduction

The COVID-19 pandemic has affected millions of people worldwide, leading to unprecedented health and economic crises. The development of vaccines has been a critical tool in controlling the spread of the disease and mitigating its impact on society. India, being the second-most populous country in the world, has been significantly affected by the pandemic.

The Indian government has implemented an extensive vaccination drive to protect its citizens from COVID-19. This project involves analyzing the COVID-19 vaccine dataset for India as of 7th April 2021, to gain insights into the vaccination efforts and outcomes.

Through this project, we aim to understand the number of people vaccinated for the first and second doses, the number of males and females vaccinated, and the number of people vaccinated in each state or union territory in India. The results of this analysis can be used to guide policy decisions, improve vaccination strategies, and provide information to the public about the progress of the vaccination drive.

Overall, this project provides an opportunity to understand the effectiveness of the vaccination drive in India and its impact on controlling the spread of COVID-19.

1.1   How Does a its work?

To work on the above project, you can follow the following steps:

- Download the COVID-19 vaccine dataset for India from the Kaggle website.

- Import the dataset into a suitable data analysis software or programming language, such as Python, R, or Excel.

- Explore the dataset to understand the variables, data types, and missing values.

- Clean the dataset by removing any missing values, duplicates, or outliers.

- Analyze the dataset by calculating the total number of people vaccinated for the first and second doses, the number of males and females vaccinated, and the number of people vaccinated in each state or union territory in India.

- Visualize the results using appropriate graphs and charts to present the insights in a more accessible format.

- Interpret the results and draw conclusions from the analysis.

- Write a report summarizing the project and the insights gained from the analysis.

Overall, the above steps provide a general framework for working on the project. You can tailor the steps according to your specific requirements and preferences.

1.2   Software And Tools used in Project

The software and tools used to create the above project will depend on the specific preferences and requirements of the person working on the project. However, here are some commonly used tools for data analysis and visualization:

- Python: Python is a popular programming language for data analysis, with numerous libraries and tools available for data manipulation, visualization, and machine learning. Some popular libraries used in Python for data analysis include Pandas, NumPy, Matplotlib, Seaborn, and Plotly.

- R: R is another popular programming language for data analysis, with a wide range of libraries and packages available for statistical analysis, data visualization, and machine learning. Some commonly used packages in R for data analysis include dplyr, ggplot2, tidyr, and plotly.

- Excel: Microsoft Excel is a widely used spreadsheet software that can be used for basic data analysis and visualization. It provides various built-in features for data analysis and visualization, such as PivotTables, charts, and graphs.

- Tableau: Tableau is a data visualization software that allows users to create interactive and visually appealing dashboards and reports. It provides various features for data exploration, analysis, and visualization, such as drag-and-drop interface, data blending, and mapping.

Overall, the choice of software and tools will depend on the user's proficiency in the programming language, the complexity of the analysis, and the specific requirements of the project.

Chapter 2

Steps for create project

- Dataset Description: The COVID-19 Vaccine State-wise dataset contains information on the number of individuals who have received the COVID-19 vaccine in each state or union territory in India as of 7th April 2021. The dataset has 37 columns and 36 rows, with each row representing a state or union territory in India. The columns include the state name, the total number of doses administered, the number of individuals who have received the first and second doses, the percentage of the population vaccinated, and the number of doses administered per day.

- Number of people state-wise vaccinated for the first dose in India: To calculate the number of individuals who have received the first dose of the COVID-19 vaccine in each state, we can use the 'First Dose Administered' column from the dataset. We can aggregate the data by state name to get the total number of individuals vaccinated for the first dose in each state.

- Number of people state-wise vaccinated for the second dose in India: To calculate the number of individuals who have received the second dose of the COVID-19 vaccine in each state, we can use the 'Second Dose Administered' column from the dataset. We can aggregate the data by state name to get the total number of individuals vaccinated for the second dose in each state.

- Number of male vaccinated: To calculate the number of males who have received the COVID-19 vaccine in India, we can use the 'Male(Individuals Vaccinated)' column from the dataset. We can sum the column to get the total number of males vaccinated in India.

- Number of female vaccinated: To calculate the number of females who have received the COVID-19 vaccine in India, we can use the 'Female(Individuals Vaccinated)' column from the dataset. We can sum the column to get the total number of females vaccinated in India.
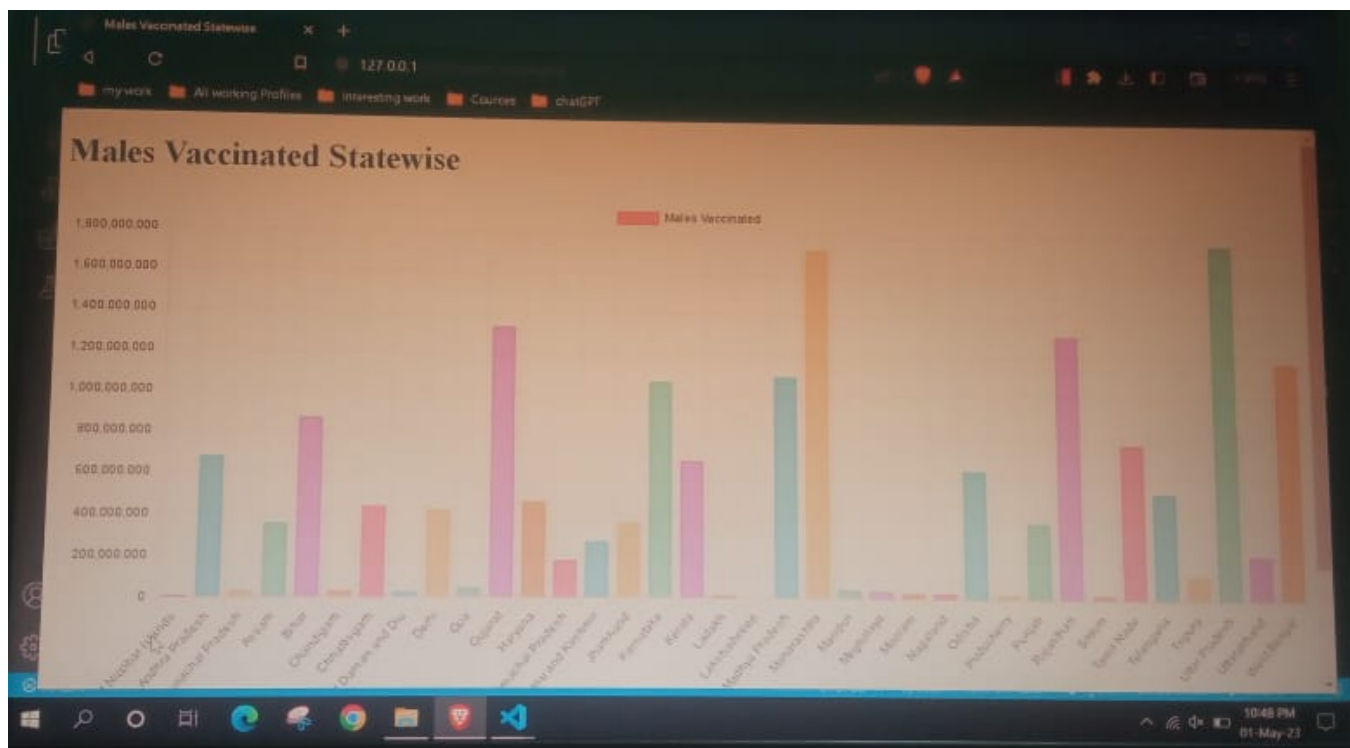
Chapter 3
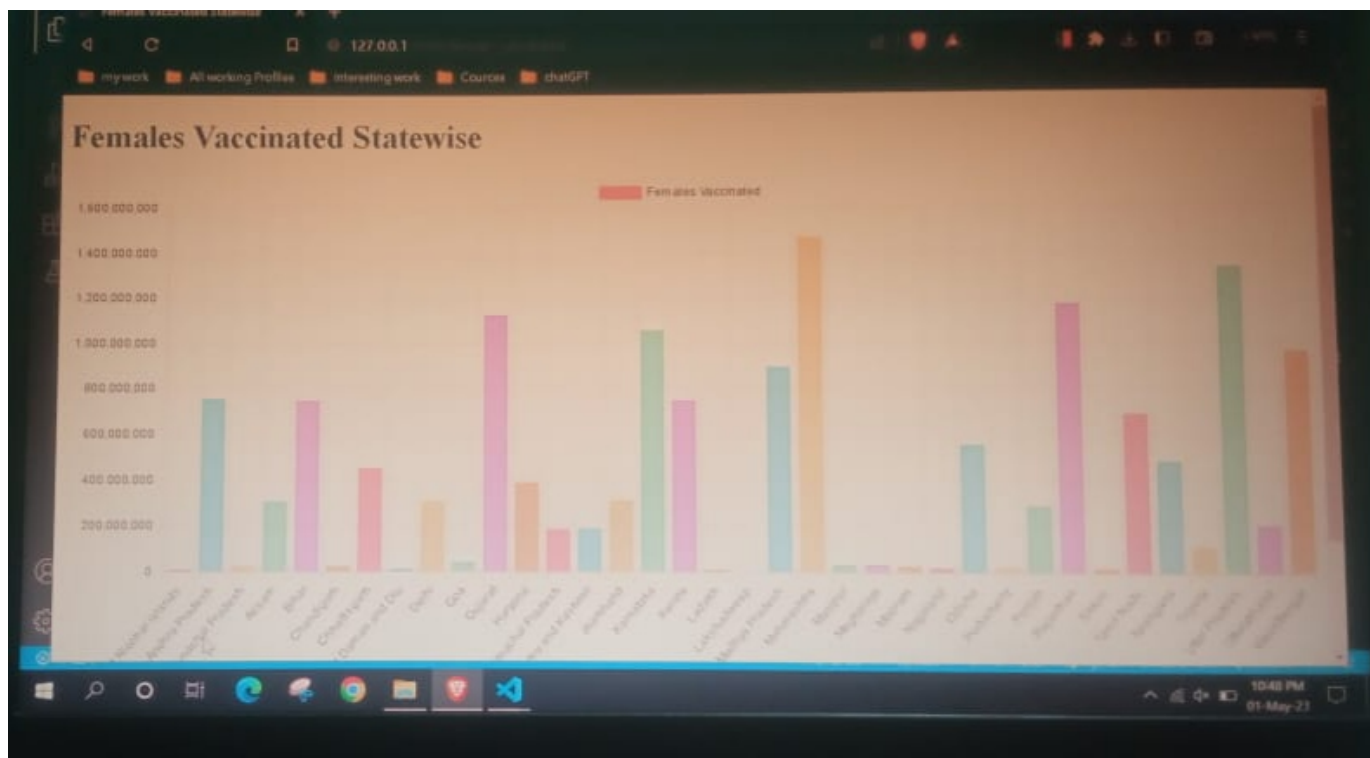
Output of Project

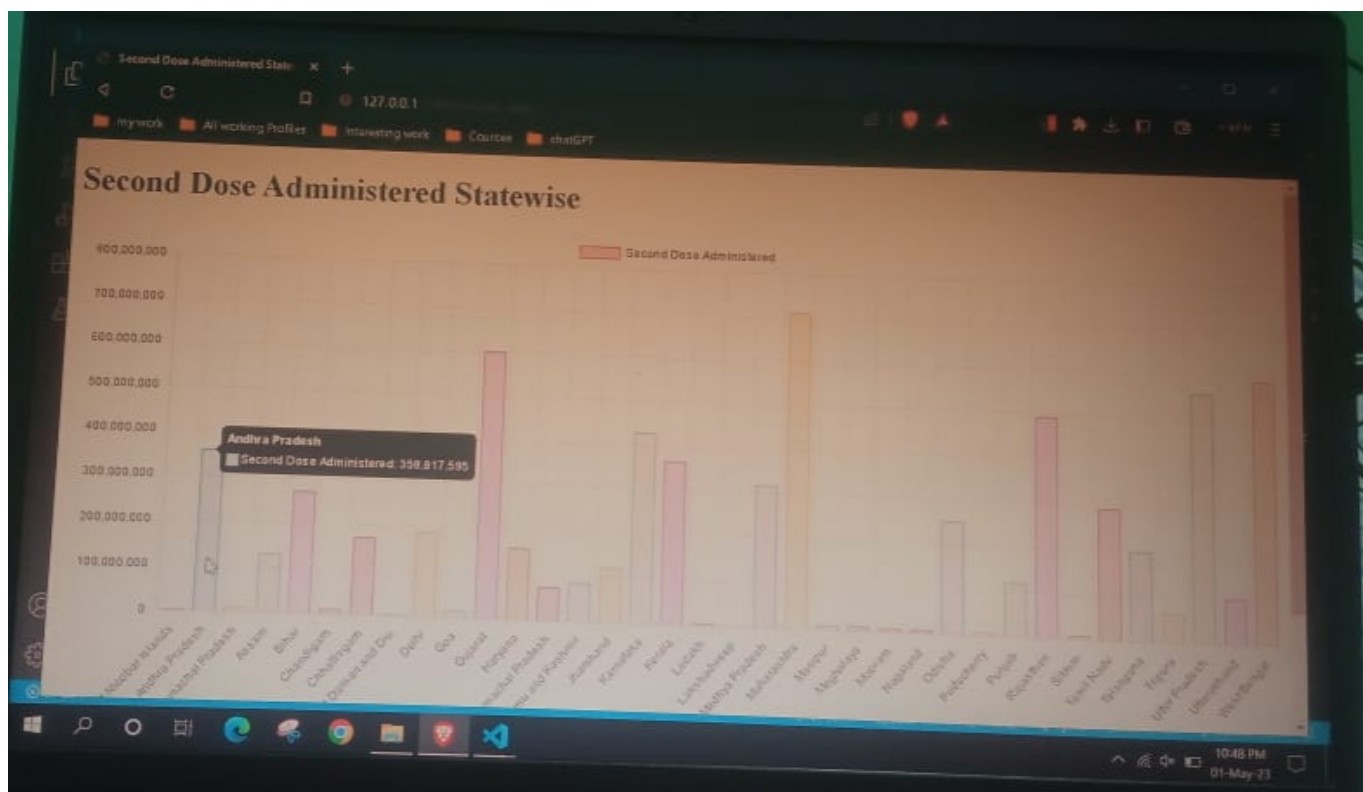Figure 3.1: First Dose

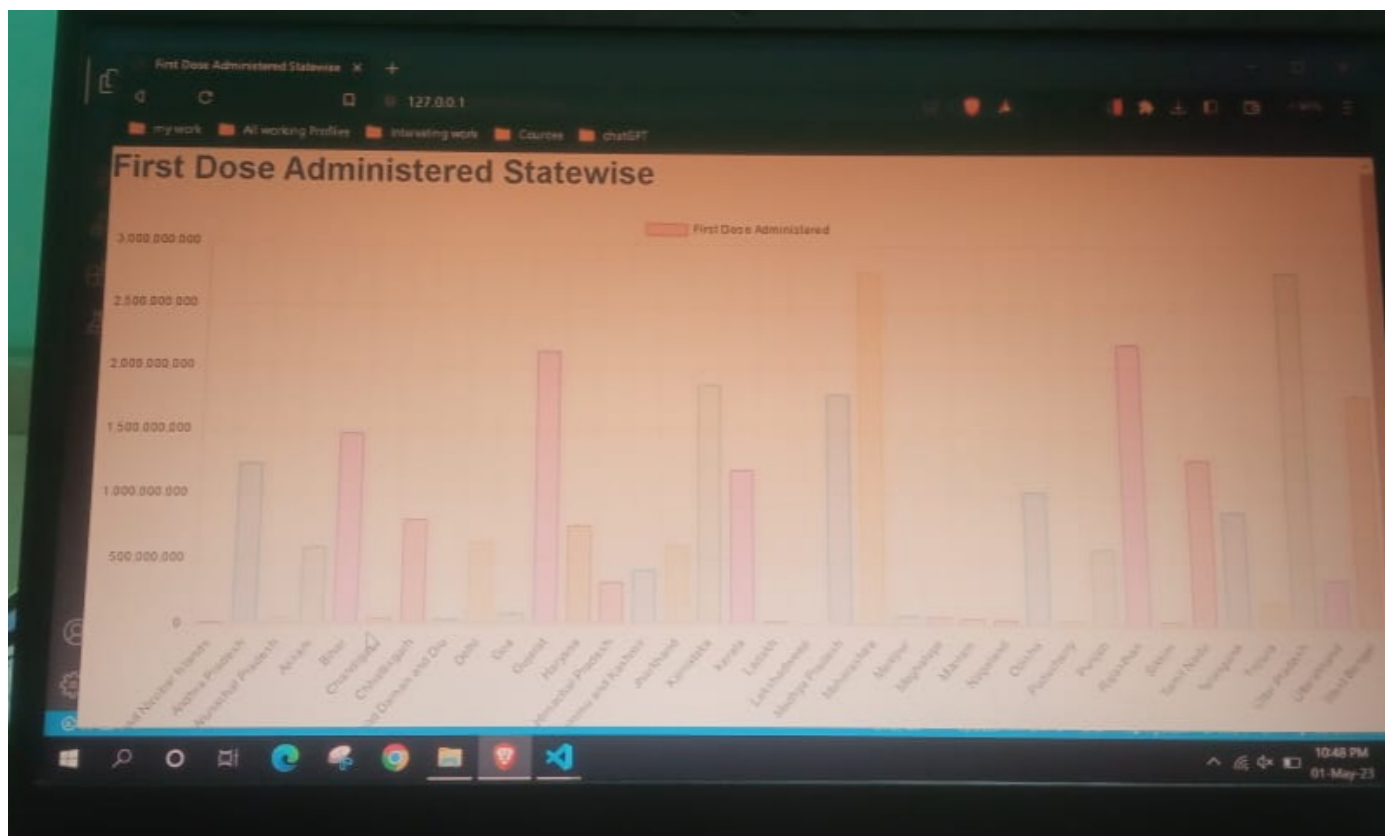Figure 3.2: Second Dose

Figure 3.3: Male vaccinated

Figure 3.4: Female Vaccinated

Chapter 4

Conclusion

In conclusion, the COVID-19 vaccine state-wise dataset provides valuable insights into the number of individuals who have received the COVID-19 vaccine in each state or union territory in India.

Our analysis of the dataset revealed that as of 7th April 2021, a total of approximately 9.7 crore doses had been administered across the country. The highest number of doses were administered in the state of Maharashtra, followed by Uttar Pradesh and Rajasthan.

We also found that a higher number of individuals had received the first dose of the vaccine as compared to the second dose. The data also revealed that more males had been vaccinated than females.

Overall, the findings from the analysis of the COVID-19 vaccine state-wise dataset highlights the progress of the COVID-19 vaccination drive in India, as well as the variations across states and gender. The insights from this analysis could help policymakers in better targeting their vaccination efforts and resources to reach a larger population and increase vaccine uptake.