**SAVITRIBAI PHULE PUNE UNIVERSITY**

****

**A MINI PROJECT REPORT ON**

## Covid-19 Vaccination Analytics

**Submitted by**

**Name: Pratik Vijay Manjarekar Roll no: 57**

**CLASS: TE DIV: A**

**Under the Guidance of Ms. Jyoti Rahtwan**

****

DEPARTMENT OF COMPUTER ENGINEERING

# RMD SINHGAD SCHOOL OF ENGINEERING

WARJE, PUNE – 411058

**2023 - 24**



DEPARTMENT OF COMPUTER ENGINEERING

# RMD SINHGAD SCHOOL OF ENGINEERING

WARJE, PUNE – 411058

**2023 – 24**

# CERTIFICATE

This is to certify that the project report entitled

## “Covid-19 Vaccination Analytics”

*Submitted by*

**Name: Pratik Vijay Manjarekar PRN No: 72262586D**

is a bona fide work carried out by them under the supervision of Ms. Jyoti Raghtwan and it is submitted towards the partial fulfilment of the requirement of Savitribai Phule Pune University for Third Year.

**Ms. Jyoti Raghtwan Dr. Deepali Newaskar**

Guide Head,

Department of Computer Engineering Department of Computer Engineering

**Dr. V. V. Dixit**

Principal

RMD Sinhgad School of Engineering Pune – 58

Place: PUNE Date:

# I

## Certificate by Guide

This is to certify that **Mr.Pratik Vijay Manjarekar** has completed the MINI Project work under my guidance and supervision and that, I have verified the work for its originality in documentation, problem statement, implementation and results presented in the Project. Any reproduction of other necessary work is with the prior permission and has given due ownership and included in the references.

Place: PUNE Date:

Signature of Guide

**Ms. Jyoti Raghtwan**

# II ACKNOWLEDGEMENT

It is my pleasure to acknowledge a sense of gratitude to all those who helped me in making this project.

I thank my Mini Project Guide **Ms. Jyoti Raghtwan** for helping me and providing all necessary information regarding this project.

I am also thankful to **Dr. Deepali Newaskar (Head - Department of Computer Engineering)** for providing me the required facilities and helping me while carrying out this project work.

Finally, I wish to thank all my teachers and friends for their constructive comments, suggestions and criticism and all those who directly or indirectly helped me in completing this project.

**Pratik Vijay Manjarekar**

|  |  |  |
| --- | --- | --- |
|  | **CONTENTS** |  |
| **Certificate** |  | **I** |
| **Certificate By Guide** |  | **II** |
| **Acknowledgement** |  | **III** |
| 1. **Problem Statement** |  | **1** |
| 2. **Introduction** |  | **2** |
| 3. **Requirements** |  | **3** |
| 4. **Algorithm Used** |  | **4** |
| 5. **Coding** |  | **5** |
| 6. **Conclusion** |  | **8** |
| 7. **References** |  | **9** |

# PROBLEM STATEMENT

Use the following covid\_vaccine\_statewise.csv dataset and perform following analytics on

the given dataset

https://www.kaggle.com/sudalairajkumar/covid19-inindia?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

e. Number of females vaccinated

# INTRODUCTION

The COVID-19 pandemic has profoundly impacted public health systems, economies, and societies around the world. In response, countries have mobilized massive vaccination drives to curb the spread of the virus and mitigate its devastating effects. India, with its vast and di- verse population, undertook one of the largest vaccination campaigns in history, aiming to ensure equitable access to vaccines across all states and union territories.

As part of my Data Science & Big Data Analytics Laboratory Mini Project, I have conducted an analytical study titled **"COVID-19 Vaccination Data Analytics Using Statewise Data"**. This project utilizes real-world vaccination data to gain meaningful insights into the progress and reach of India’s vaccination drive.

The dataset used in this study, **covid\_vaccine\_statewise.csv**, was sourced from Kaggle and contains detailed vaccination statistics across different Indian states. It includes information such as the number of individuals vaccinated with the first and second doses, along with gen- der-wise breakdowns. The aim of this project is to analyze this data to identify patterns, dis- parities, and trends that can inform public health decisions and awareness campaigns.

Using Python and powerful data analysis libraries such as **NumPy**, **pandas**, and **matplotlib**, this project performs an in-depth exploration of the vaccination dataset. It provides a clear state-wise and gender-wise distribution of vaccinated individuals, offering a comprehensive view of the vaccination progress in India.

Through this analytical study, we aim to not only summarize the vaccination efforts but also highlight areas that may require more attention and resources. The following sections will de- tail the methodology, algorithmic approach, results, and conclusions drawn from this analy- sis.

This project demonstrates how data-driven insights can support informed decision-making and contribute to the success of public health initiatives in times of global crisis.

# REQUIREMENTS

**Hardware Requirements:**

* Processor: Intel Core i3 or equivalent (later generations recommended)
* RAM: 4GB (8GB or more recommended for better performance)

**Software Requirements:**

* Operating System: Windows 10 (or macOS, Linux)
* Languages: Python (3.6.3 or later)
* Software: Anaconda, Jupyter Notebook
* Dataset: covid\_vaccine\_statewise.csv
* Libraries: numpy, pandas, scikit-learn

# ALGORITHM

**Covid-19 Vaccination Analytics Algorithm:**

**Step 1: Import Libraries**

* Use pandas for data handling.

**Step 2: Load Data**

* Read the dataset: "covid\_vaccine\_statewise.csv".

**Step 3: Initial Exploration**

* View dataset information (df.info()) and preview (df.head()).
* Clean column names using df.columns.str.strip().

**Step 4: Dataset Description**

* Print shape (rows and columns) and descriptive statistics using df.describe(include='all').

**Step 5: Analysis Steps:**

**a. First Dose by State :**

* **Group data by State, find the max of 'First Dose Administered'.**
* **Sort descending to see which states vaccinated the most.**

**b. Second Dose by State :**

* Same as above but for 'Second Dose Administered'.

**c. Total Males Vaccinated :**

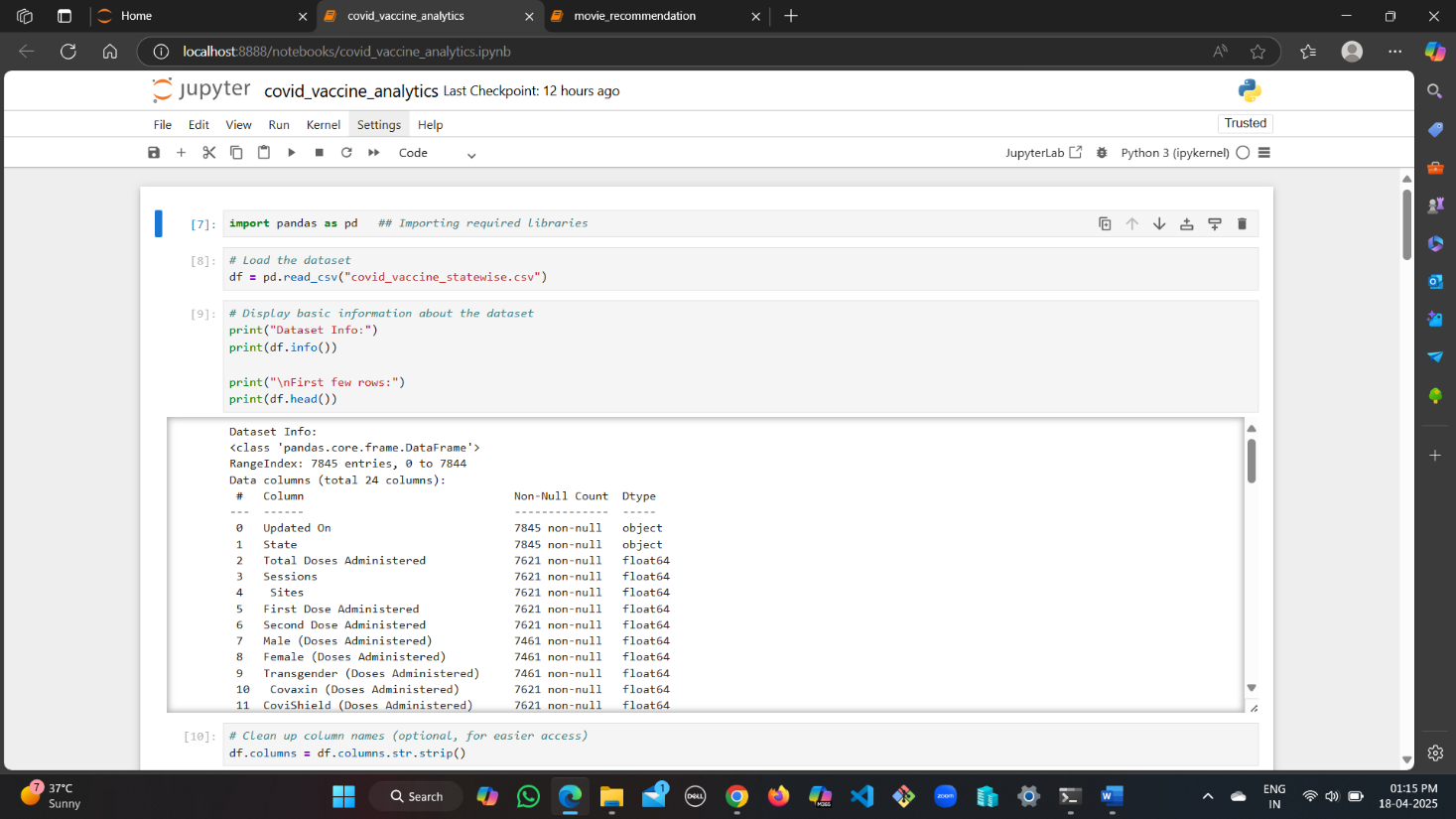
* Group by State, take max of 'Male(Individuals Vaccinated)', then sum across all states.

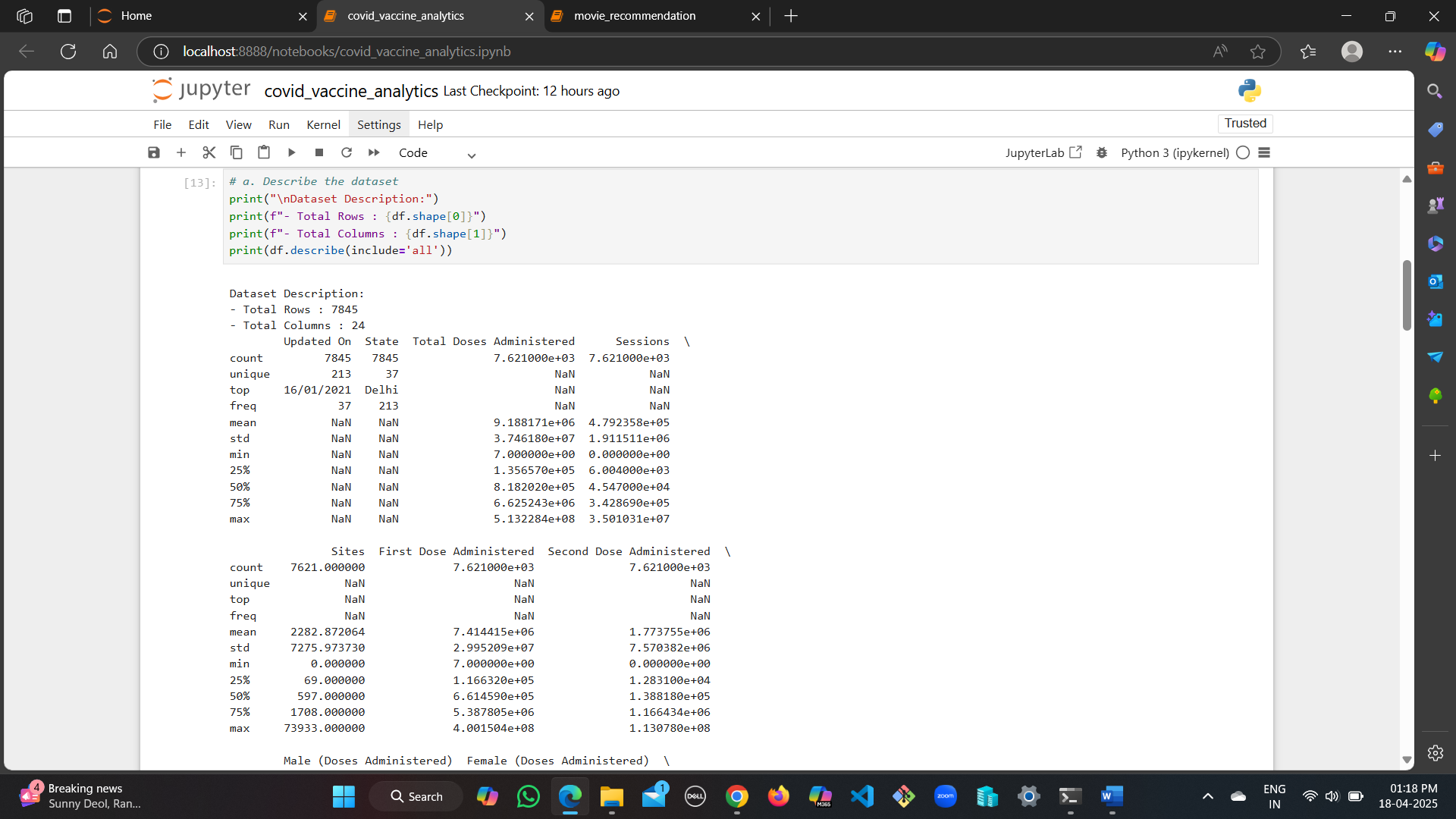
**d. Total Females Vaccinated :**

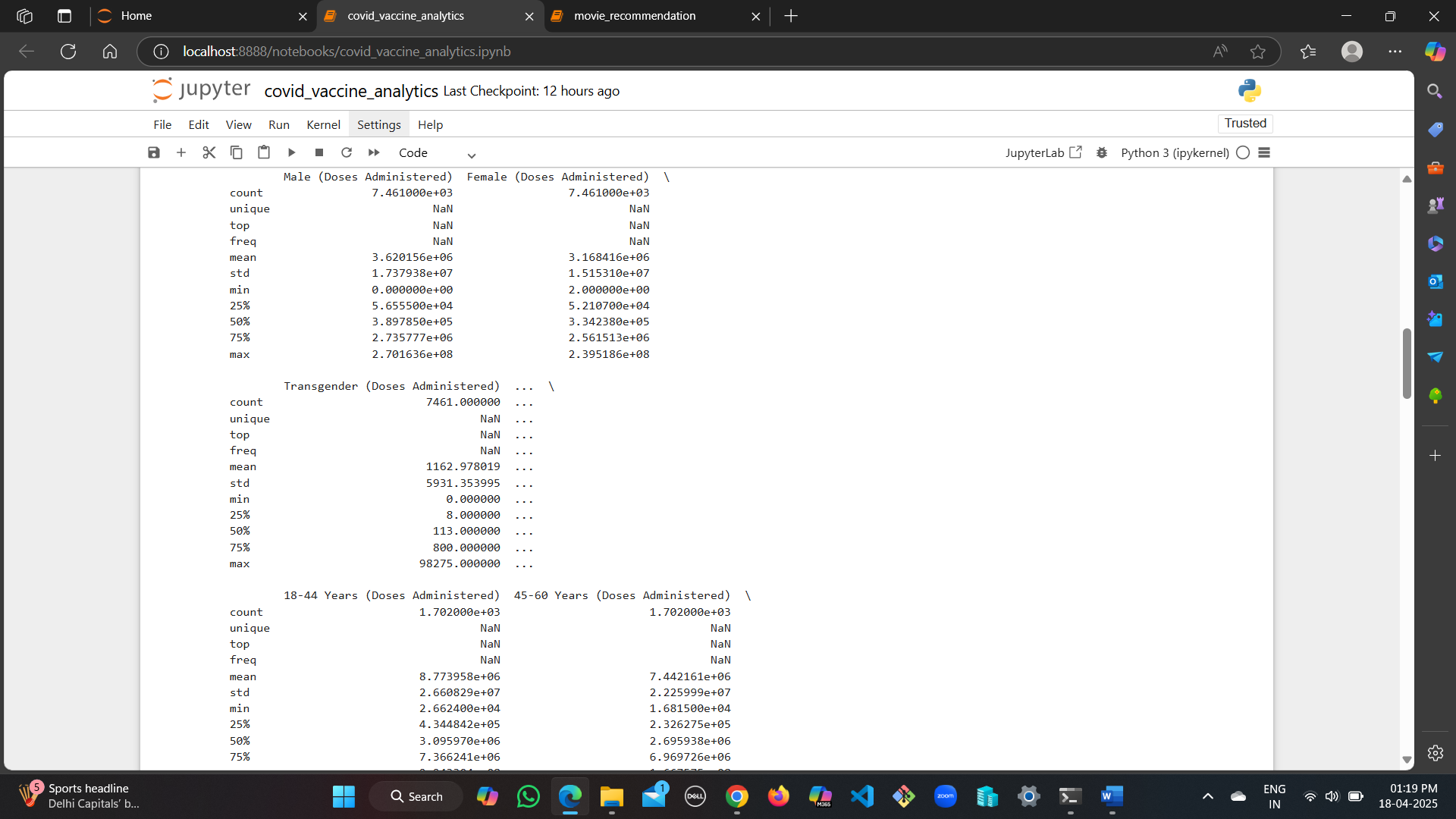
* Group by State, take max of 'Female(Individuals Vaccinated)', then sum across all states.

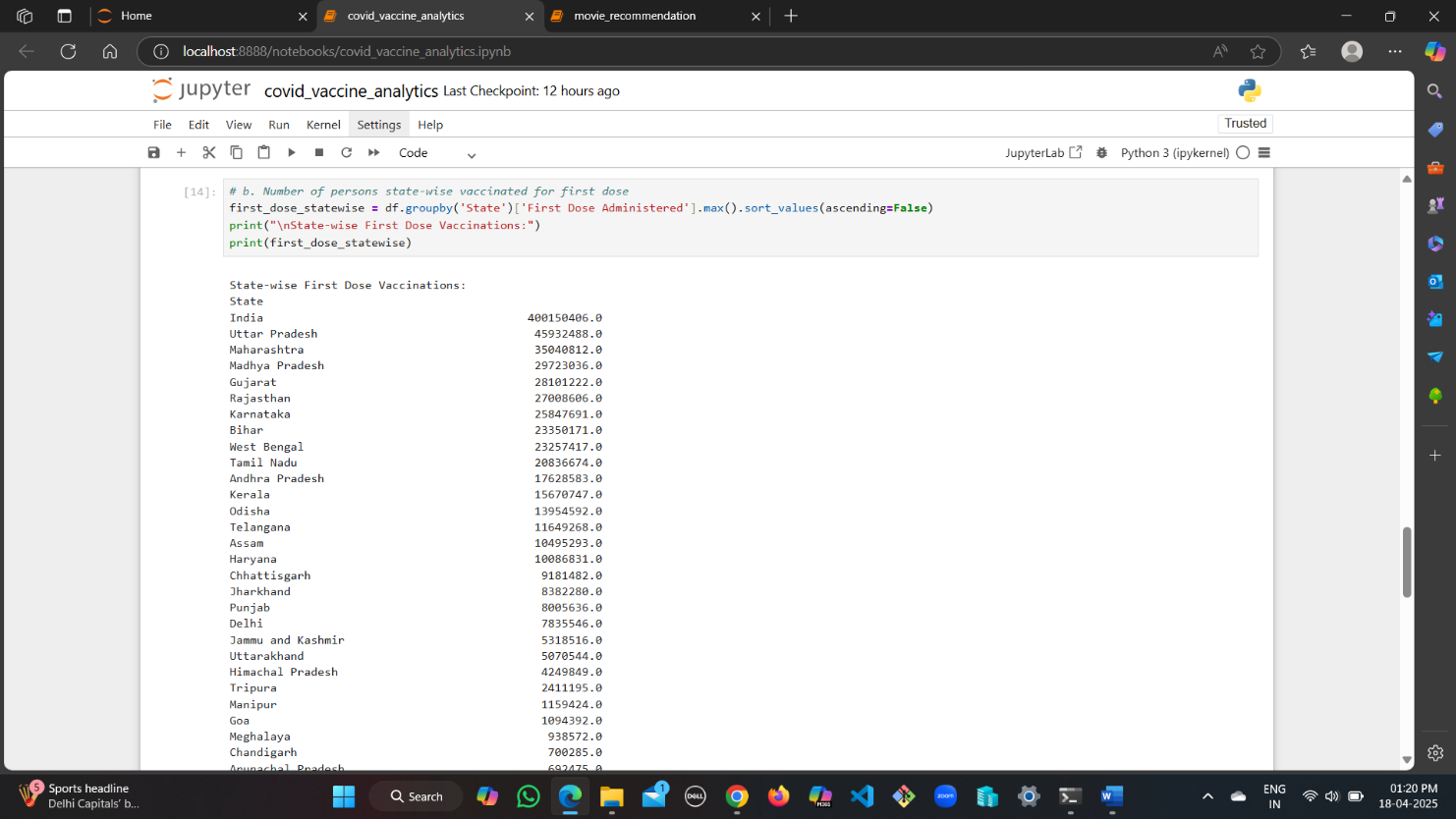
This structured approach ensures that the analysis is reproducible, scalable, and insightful. By methodically cleaning, processing, analyzing, and visualizing the data, we derive a compre- hensive understanding of India's COVID-19 vaccination landscape

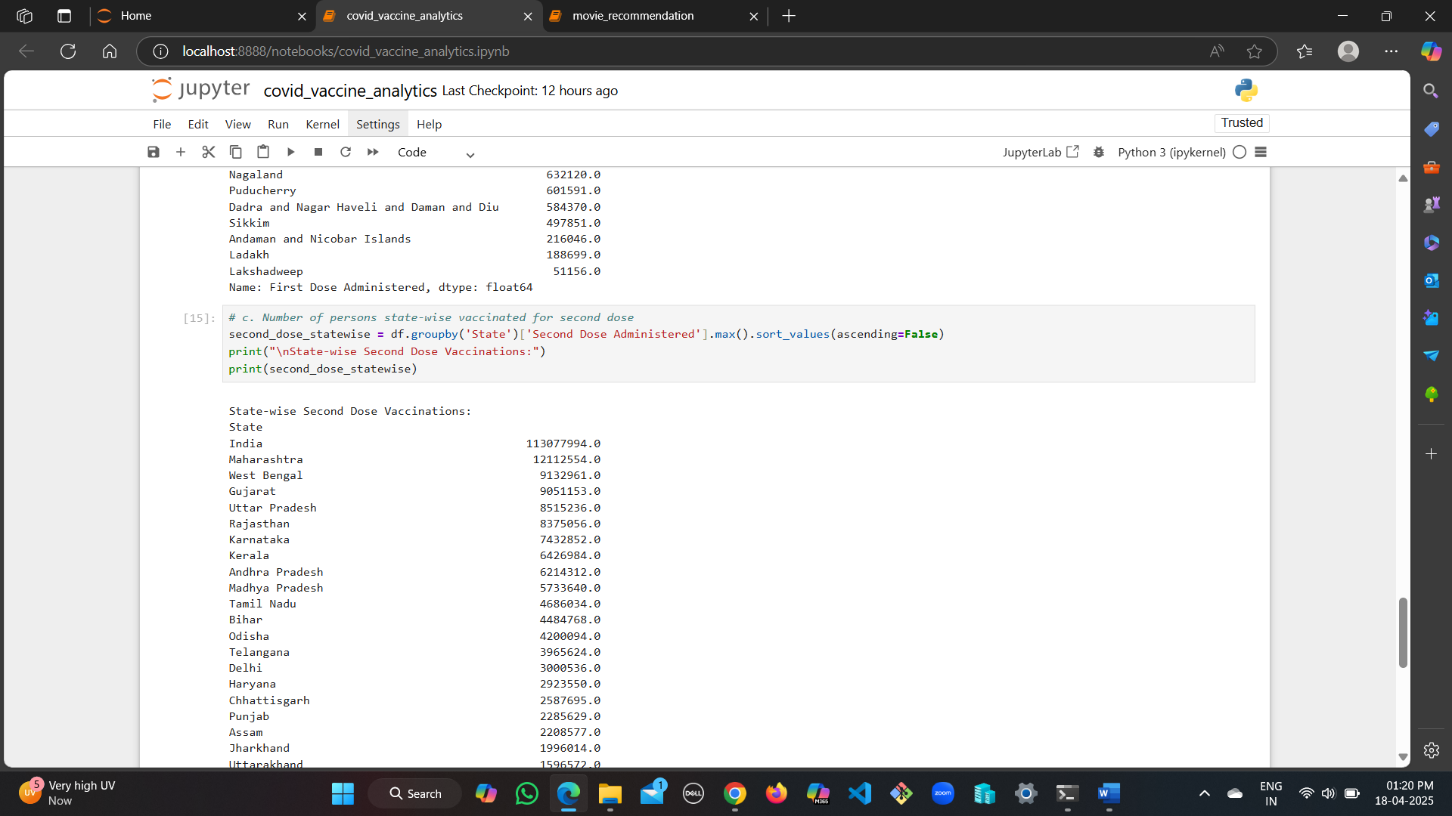
# CODE

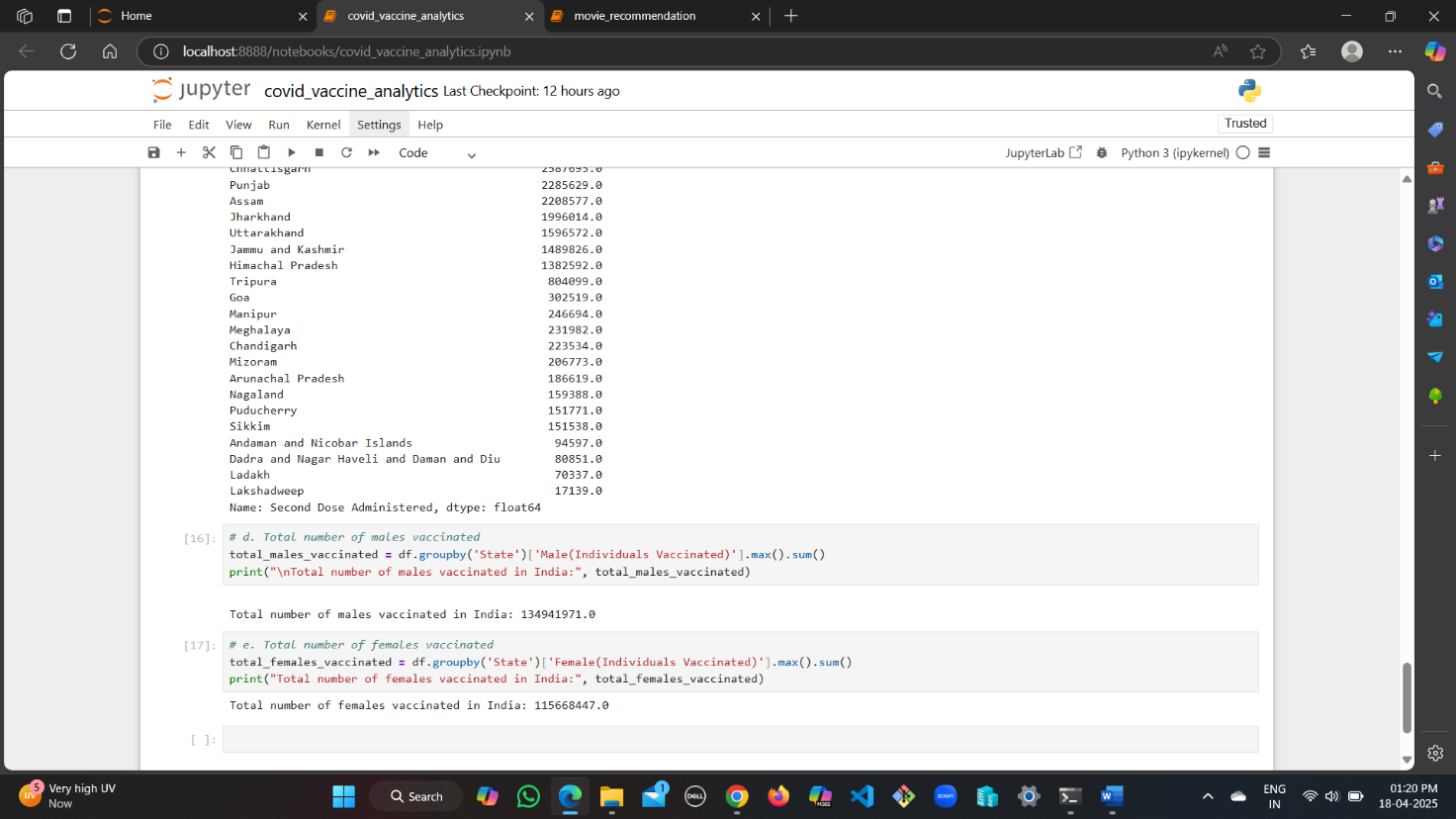
****

****

****

****





**CONCLUSION**

In conclusion, this COVID-19 Vaccination Data Analytics project has successfully demon- strated the power of data-driven approaches in understanding and evaluating public health in- itiatives. By leveraging Python libraries such as **NumPy**, **pandas**, and **matplotlib**, we were able to perform a detailed analysis of the covid\_vaccine\_statewise.csv dataset, uncovering in- sightful patterns in vaccination distribution across Indian states.

The project provided valuable insights into **statewise vaccination trends**, highlighting both the progress and disparities in vaccine coverage for **first and second doses**, as well as **gen- der-wise distribution**. Such analysis is crucial in identifying regions that may require in- creased focus, policy adjustments, or awareness campaigns to improve outreach and equity.

This project underscores the important role that **data science** plays in **public health moni- toring and strategic decision-making**. It not only aids in visualizing complex datasets but also empowers stakeholders with actionable insights.

Looking ahead, this analytical model can be further enhanced by integrating additional da- tasets such as **population demographics**, **vaccination centers**, or **district-level data**. Incor- porating **temporal analysis** or **predictive modeling** could provide forecasts on vaccination progress and help in optimizing future distribution strategies.

With continuous development and richer datasets, such analytical systems can become vital tools for **governments, health officials, and researchers** to make informed, impactful deci- sions in managing public health crises.

# REFERENCES

* Dataset Source: [Kaggle - COVID-19 India](https://www.kaggle.com/sudalairajkumar/covid19india?select=covid_vaccine_statewise.csv)
* Python Libraries Used: Pandas, [numpy](https://numpy.org/)
* Government of India COVID-19 Vaccine Data (for comparison/validation): [CoWIN](https://www.cowin.gov.in/) [Portal](https://www.cowin.gov.in/)