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

Sinhgad Institutes

DEPARTMENT OF COMPUTER ENGINEERING

RMD SINHGAD SCHOOL OF ENGINEERING

WARJE, PUNE – 411058

2023 - 24

Sinhgad Institutes

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
Guide
Department of Computer Engineering

Dr. Deepali Newaskar
Head,
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

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

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 diverse 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 gender-wise breakdowns. The aim of this project is to analyze this data to identify patterns, disparities, 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 detail the methodology, algorithmic approach, results, and conclusions drawn from this analysis.

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 :

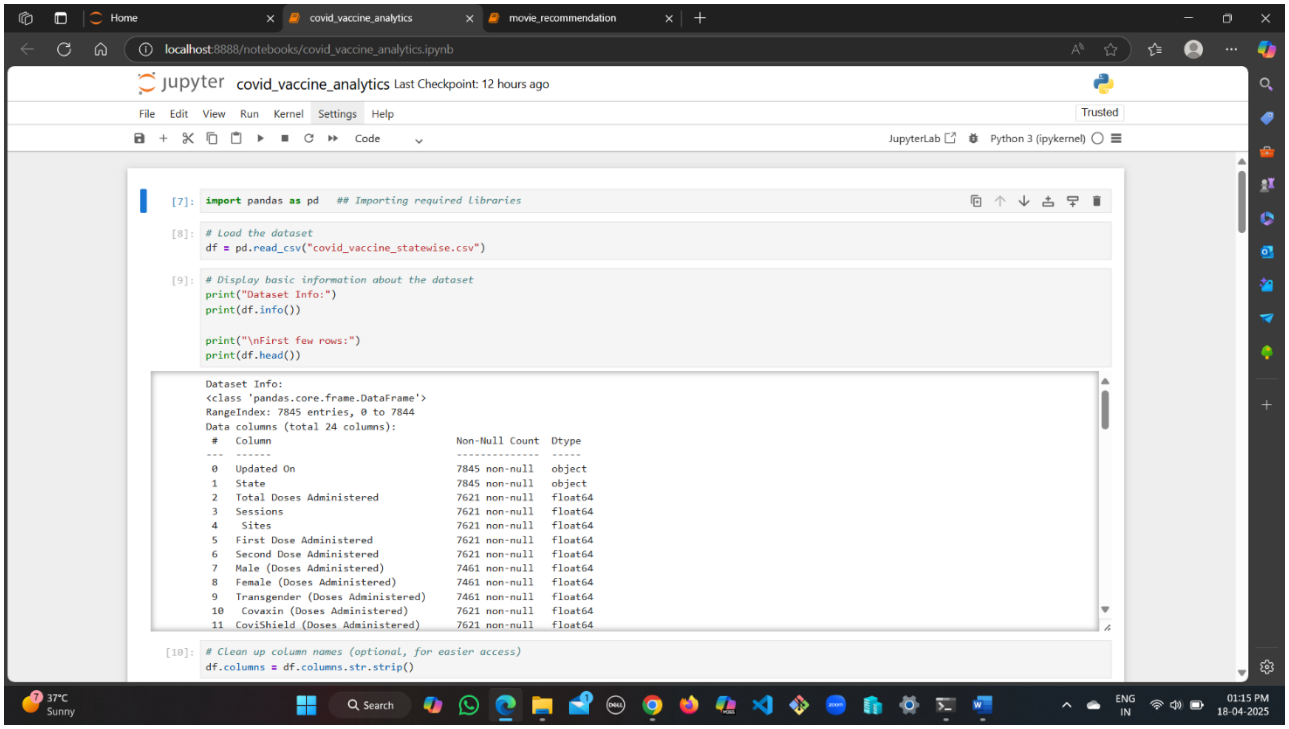
- Group by State, take max of 'Male(Individuals Vaccinated)', then sum across all states.

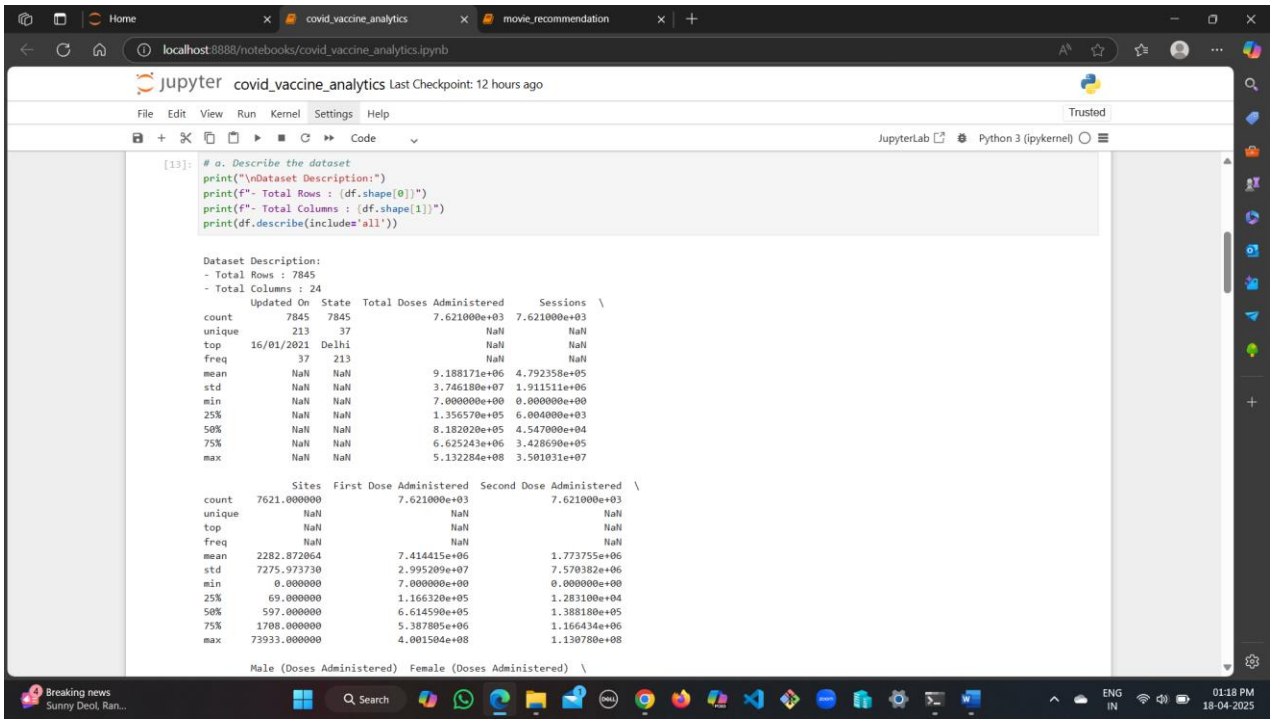
d. Total Females Vaccinated :

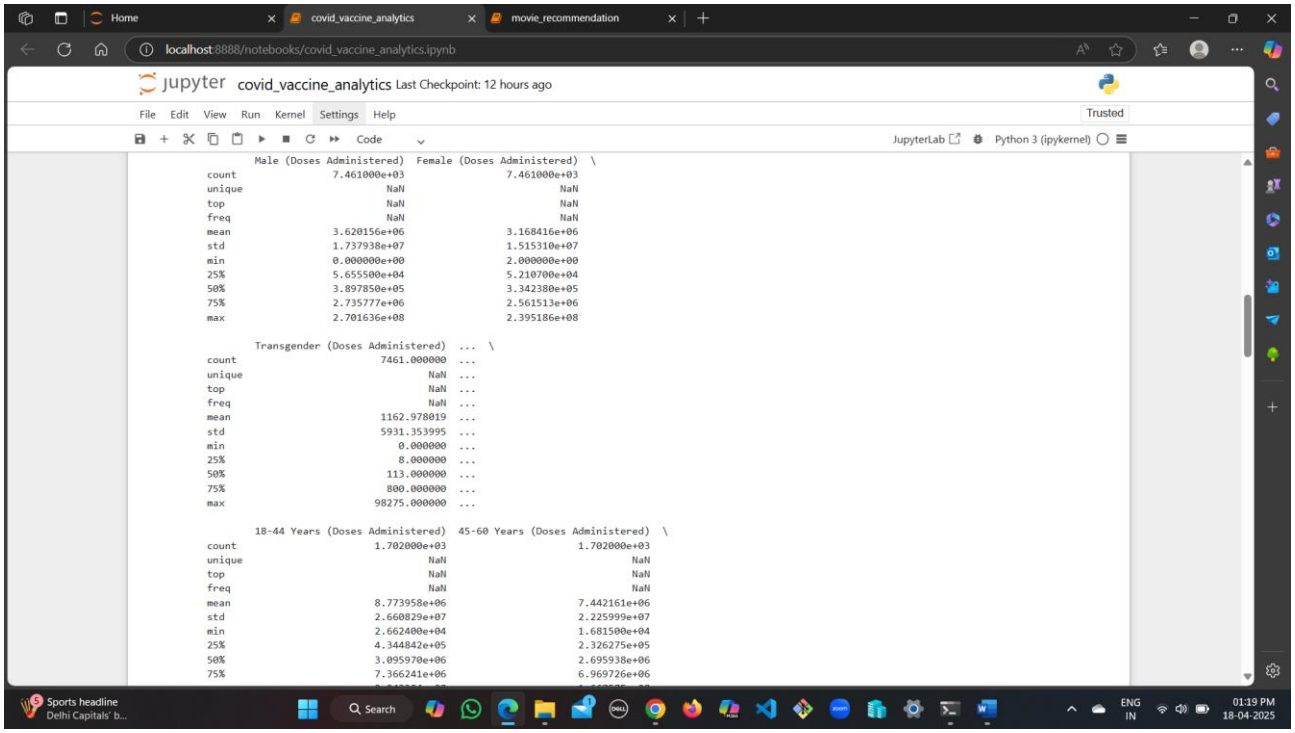
- Group by State, take max of 'Female(Individuals Vaccinated)', then sum across all states.

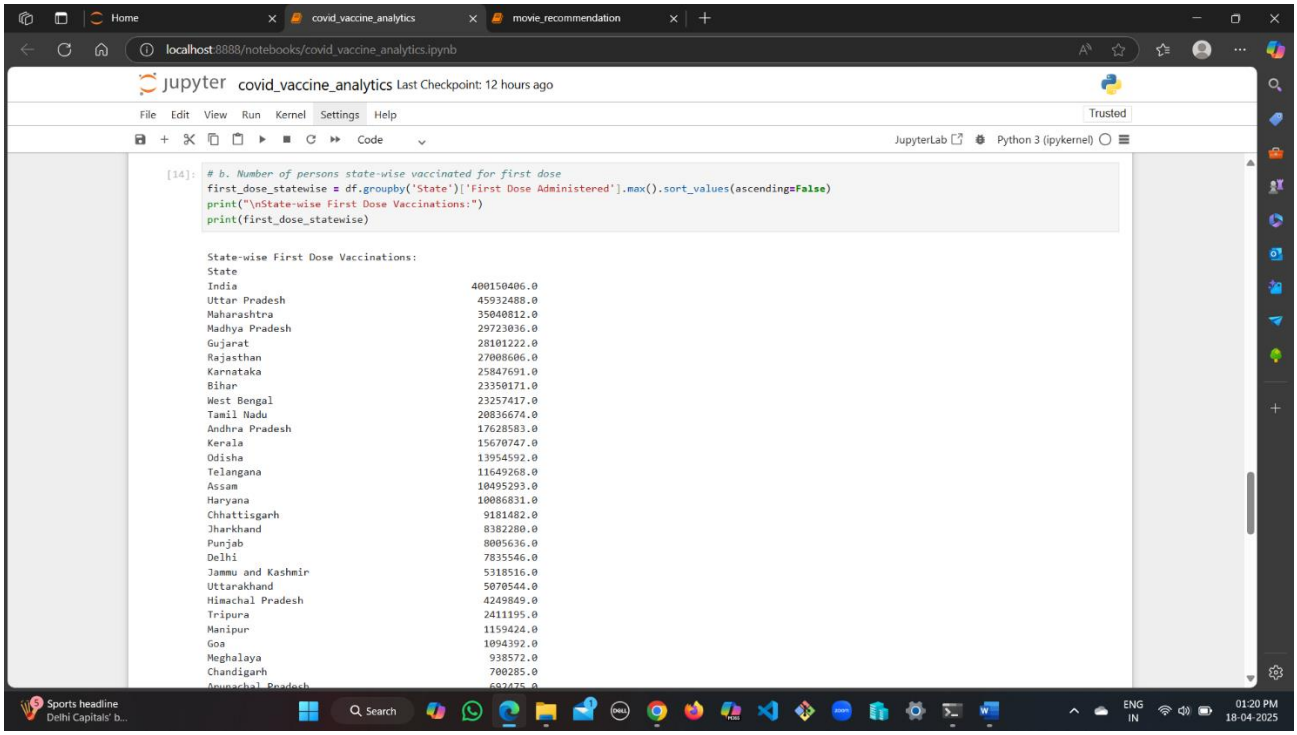
This structured approach ensures that the analysis is reproducible, scalable, and insightful. By methodically cleaning, processing, analyzing, and visualizing the data, we derive a comprehensive understanding of India's COVID-19 vaccination landscape

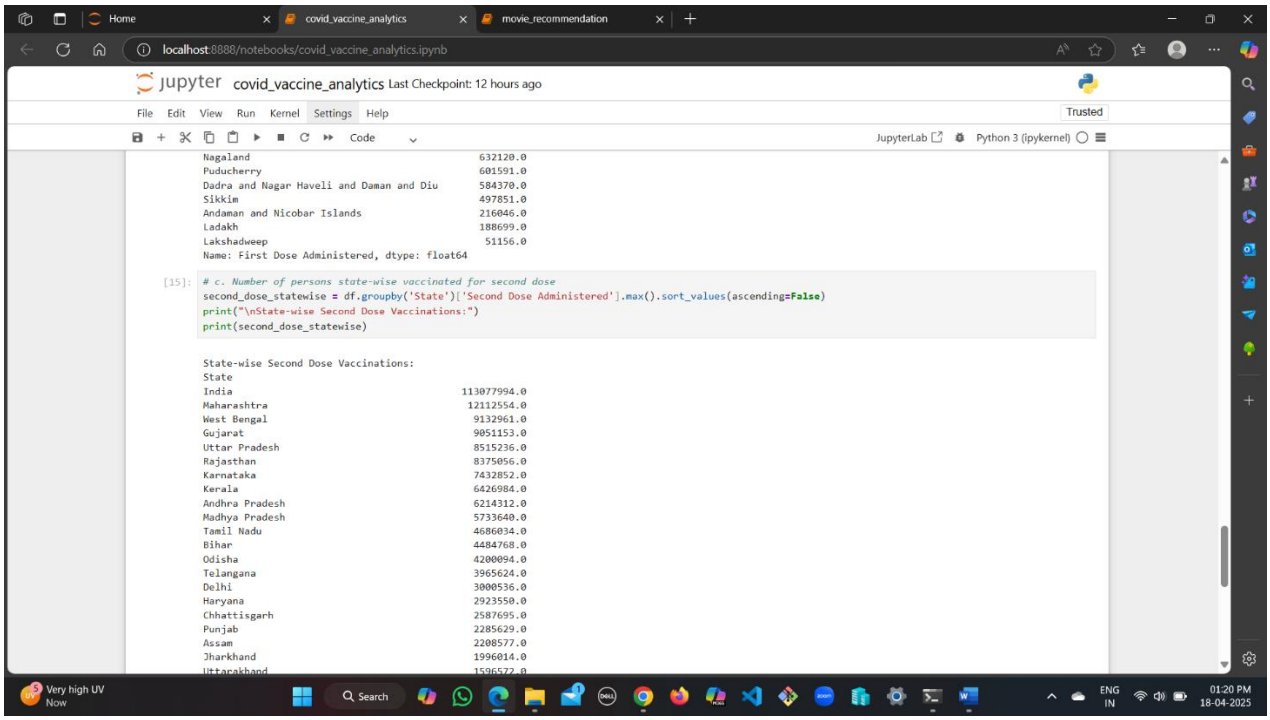
CODE

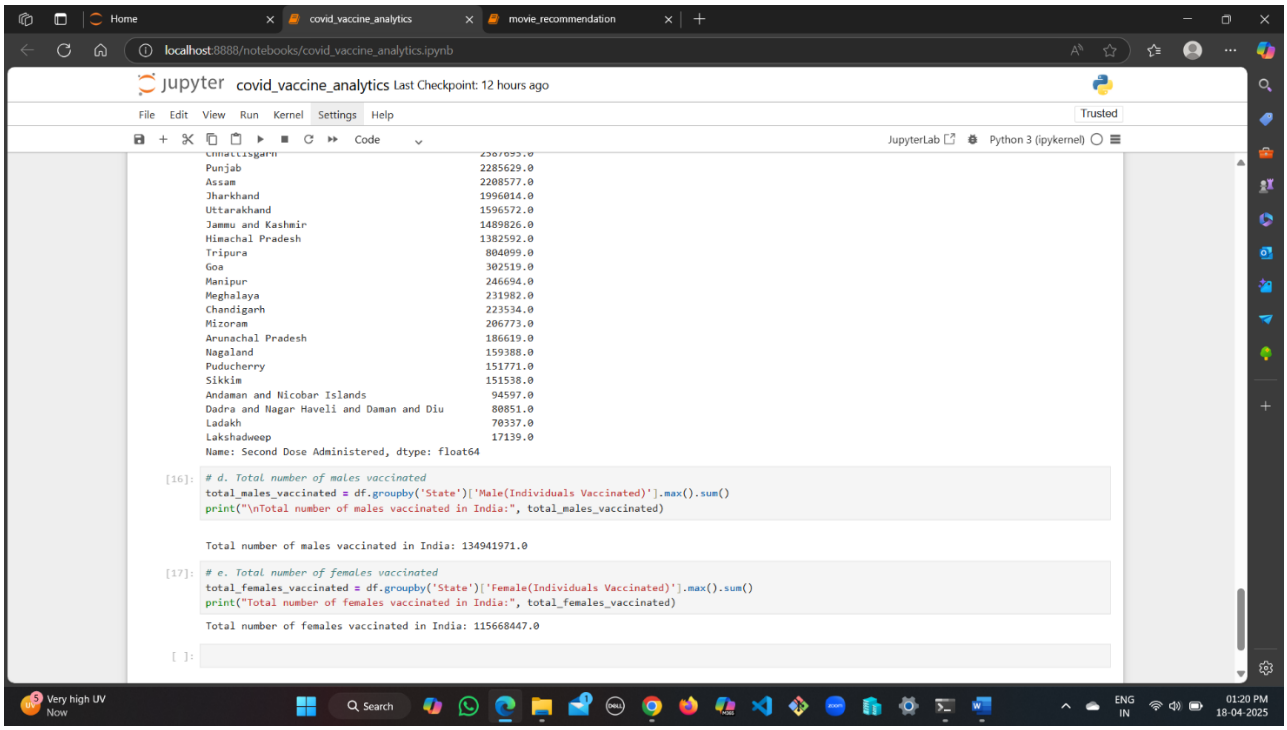

Home
covid_vaccine_analytics
movie_recommendation
localhost:8888/notebooks/covid_vaccine_analytics.ipynb
jupyter covid_vaccine_analytics Last Checkpoint: 12 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
[7]: import pandas as pd ## Importing required libraries
[8]: # Load the dataset
df = pd.read_csv("covid_vaccine_statewise.csv")
[9]: # Display basic information about the dataset
print("Dataset Info:")
print(df.info())
print("\nFirst few rows:")
print(df.head())
Dataset 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
[10]: # Clean up column names (optional, for easier access)
df.columns = df.columns.str.strip()
37°C Sunny
Search
ENG IN
01:15 PM 18-04-2025


Home
covid_vaccine_analytics
movie_recommendation
localhost:8888/notebooks/covid_vaccine_analytics.ipynb
jupyter covid_vaccine_analytics Last Checkpoint: 12 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
[13]: # a. Describe the dataset
print("\nDataset Description:")
print(f"- Total Rows : {df.shape[0]}")
print(f"- Total Columns : {df.shape[1]}")
print(df.describe(include='all'))
Dataset Description:
- Total Rows : 7845
- Total Columns : 24
Updated On State Total Doses Administered Sessions \
count 7845 7845 7.621000e+03 7.621000e+03
unique 213 37 NaN NaN
top 16/01/2021 Delhi NaN NaN
freq 37 213 NaN NaN
mean NaN NaN 9.188171e+06 4.792358e+05
std NaN NaN 3.746180e+07 1.911511e+06
min NaN NaN 7.000000e+00 0.000000e+00
25% NaN NaN 1.356570e+05 6.004000e+03
50% NaN NaN 8.182020e+05 4.547000e+04
75% NaN NaN 6.625243e+06 3.428690e+05
max NaN NaN 5.132284e+08 3.501031e+07
Sites First Dose Administered Second Dose Administered \
count 7621.000000 7.621000e+03 7.621000e+03
unique NaN NaN NaN
top NaN NaN NaN
freq NaN NaN NaN
mean 2282.872064 7.414415e+06 1.773755e+06
std 7275.973730 2.995209e+07 7.570382e+06
min 0.000000 7.000000e+00 0.000000e+00
25% 69.000000 1.166320e+05 1.283100e+04
50% 597.000000 6.614590e+05 1.388180e+05
75% 1708.000000 5.387805e+06 1.166434e+06
max 73933.000000 4.001504e+08 1.130780e+08
Male (Doses Administered) Female (Doses Administered) \
Breaking news Sunny Deol, Ran...
Search
ENG IN
01:18 PM 18-04-2025


Home
covid_vaccine_analytics
movie_recommendation
localhost:8888/notebooks/covid_vaccine_analytics.ipynb
jupyter covid_vaccine_analytics Last Checkpoint: 12 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
Male (Doses Administered) Female (Doses Administered) \
count 7.461000e+03 7.461000e+03
unique NaN NaN
top NaN NaN
freq NaN NaN
mean 3.620156e+06 3.168416e+06
std 1.737938e+07 1.515310e+07
min 0.000000e+00 2.000000e+00
25% 5.655500e+04 5.210700e+04
50% 3.897850e+05 3.342380e+05
75% 2.735777e+06 2.561513e+06
max 2.701636e+08 2.395186e+08
Transgender (Doses Administered) ... \
count 7461.000000 ...
unique NaN ...
top NaN ...
freq NaN ...
mean 1162.978019 ...
std 5931.353995 ...
min 0.000000 ...
25% 8.000000 ...
50% 113.000000 ...
75% 800.000000 ...
max 98275.000000 ...
18-44 Years (Doses Administered) 45-60 Years (Doses Administered) \
count 1.702000e+03 1.702000e+03
unique NaN NaN
top NaN NaN
freq NaN NaN
mean 8.773958e+06 7.442161e+06
std 2.660829e+07 2.225999e+07
min 2.662400e+04 1.681500e+04
25% 4.344842e+05 2.326275e+05
50% 3.095970e+06 2.695938e+06
75% 7.366241e+06 6.969726e+06
Sports headline
Delhi Capitals' b...
Search
ENG
IN
01:19 PM
18-04-2025


Home
covid_vaccine_analytics
movie_recommendation
localhost:8888/notebooks/covid_vaccine_analytics.ipynb
jupyter covid_vaccine_analytics Last Checkpoint: 12 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
[14]:
# b. Number of persons state-wise vaccinated for first dose
first_dose_statewise = df.groupby('State')['First Dose Administered'].max().sort_values(ascending=False)
print("\nState-wise First Dose Vaccinations:")
print(first_dose_statewise)
State-wise First Dose Vaccinations:
State
India 400150406.0
Uttar Pradesh 45932488.0
Maharashtra 35040812.0
Madhya Pradesh 29723036.0
Gujarat 28101222.0
Rajasthan 27008606.0
Karnataka 25847691.0
Bihar 23350171.0
West Bengal 23257417.0
Tamil Nadu 20836674.0
Andhra Pradesh 17628583.0
Kerala 15670747.0
Odisha 13954592.0
Telangana 11649268.0
Assam 10495293.0
Haryana 10086831.0
Chhattisgarh 9181482.0
Jharkhand 8382280.0
Punjab 8005636.0
Delhi 7835546.0
Jammu and Kashmir 5318516.0
Uttarakhand 5070544.0
Himachal Pradesh 4249849.0
Tripura 2411195.0
Manipur 1159424.0
Goa 1094392.0
Meghalaya 938572.0
Chandigarh 700285.0
Sports headline
Delhi Capitals' b...
Search
ENG
IN
01:20 PM
18-04-2025


Home
covid_vaccine_analytics
movie_recommendation
localhost:8888/notebooks/covid_vaccine_analytics.ipynb
jupyter covid_vaccine_analytics Last Checkpoint: 12 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
Nagaland 632120.0
Puducherry 601591.0
Dadra and Nagar Haveli and Daman and Diu 584370.0
Sikkim 497851.0
Andaman and Nicobar Islands 216046.0
Ladakh 188699.0
Lakshadweep 51156.0
Name: First Dose Administered, dtype: float64
[15]:
# c. Number of persons state-wise vaccinated for second dose
second_dose_statewise = df.groupby('State')['Second Dose Administered'].max().sort_values(ascending=False)
print("\nState-wise Second Dose Vaccinations:")
print(second_dose_statewise)
State-wise Second Dose Vaccinations:
State
India 113077994.0
Maharashtra 12112554.0
West Bengal 9132961.0
Gujarat 9051153.0
Uttar Pradesh 8515236.0
Rajasthan 8375056.0
Karnataka 7432852.0
Kerala 6426984.0
Andhra Pradesh 6214312.0
Madhya Pradesh 5733640.0
Tamil Nadu 4686034.0
Bihar 4484768.0
Odisha 4200094.0
Telangana 3965624.0
Delhi 3000536.0
Haryana 2923550.0
Chhattisgarh 2587695.0
Punjab 2285629.0
Assam 2208577.0
Jharkhand 1996014.0
Uttarakhand 1596572.0
Very high UV Now
Search
ENG IN
01:20 PM 18-04-2025


Home
covid_vaccine_analytics
movie_recommendation
localhost:8888/notebooks/covid_vaccine_analytics.ipynb
jupyter covid_vaccine_analytics Last Checkpoint: 12 hours ago
File Edit View Run Kernel Settings Help
Trusted
Code
JupyterLab
Python 3 (ipykernel)
Punjab 2285629.0
Assam 2208577.0
Jharkhand 1996014.0
Uttarakhand 1596572.0
Jammu and Kashmir 1489826.0
Himachal Pradesh 1382592.0
Tripura 804099.0
Goa 302519.0
Manipur 246694.0
Meghalaya 231982.0
Chandigarh 223534.0
Mizoram 206773.0
Arunachal Pradesh 186619.0
Nagaland 159388.0
Puducherry 151771.0
Sikkim 151538.0
Andaman and Nicobar Islands 94597.0
Dadra and Nagar Haveli and Daman and Diu 80851.0
Ladakh 70337.0
Lakshadweep 17139.0
Name: Second Dose Administered, dtype: float64
[16]:
# d. Total number of males vaccinated
total_males_vaccinated = df.groupby('State')['Male(Individuals Vaccinated)'].max().sum()
print("\nTotal number of males vaccinated in India:", total_males_vaccinated)
Total number of males vaccinated in India: 134941971.0
[17]:
# e. Total number of females vaccinated
total_females_vaccinated = df.groupby('State')['Female(Individuals Vaccinated)'].max().sum()
print("Total number of females vaccinated in India:", total_females_vaccinated)
Total number of females vaccinated in India: 115668447.0
[ ]:
Very high UV Now
Search
ENG IN
01:20 PM 18-04-2025

CONCLUSION

In conclusion, this COVID-19 Vaccination Data Analytics project has successfully demonstrated the power of data-driven approaches in understanding and evaluating public health initiatives. 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 insightful 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 **gender-wise distribution**. Such analysis is crucial in identifying regions that may require increased focus, policy adjustments, or awareness campaigns to improve outreach and equity.

This project underscores the important role that **data science** plays in **public health monitoring 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 datasets such as **population demographics**, **vaccination centers**, or **district-level data**. Incorporating **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 decisions in managing public health crises.

REFERENCES

- Dataset Source: Kaggle - COVID-19 India

- Python Libraries Used:

  Pandas, numpy

- Government of India COVID-19 Vaccine Data (for comparison/validation): CoWIN Portal