State wise Analysis of COVID-19 in India: Cases, Deaths and Recoveries

M. Ramya,
sahasrapramod@gmail.com,
GitHub link: Ramyamarripedda.

1) Project overview:

The project analyses COVID-19 data for each state in India, showing how many people got sick, recovered, died, and were vaccinated. It groups states by population size to see differences more clearly. The project uses easy-to-understand charts with interactive features to explore the data. This helps decision-makers understand how the pandemic affected different regions. Overall, it makes complex data simple and accessible for better planning and response.

2) Dataset Description:

Table 1: COVID Cases Data

Column	Description
Date	Report date
State_UT	Indian state or union territory name
Latitude	Geographical latitude of the state/UT
Longitude	Geographical longitude of the state/UT
TotalConfirmed	Total cumulative confirmed COVID-19 cases
Death	Total cumulative deaths due to COVID-19
CuredDischargedMigrated	Total cumulative recoveries, discharges, and migrations
NewCases	Number of new cases reported on the given date
NewDeaths	Number of new deaths reported on the given date
NewRecovered	Number of new recoveries reported on the given date

Table 2: Vaccination Data

Column	Description
State/UTs	Indian state or union territory name
Total Vaccination Doses	Total COVID-19 vaccine doses administered
Dose1	Number of people who received first dose
Dose2	Number of people who received second dose
Dose 1 15-18	First dose administered to age group 15-18
Dose 2 15-18	Second dose administered to age group 15-18
Dose 1 12-14	First dose administered to age group 12-14
Dose 2 12-14	Second dose administered to age group 12-14
Precaution 18-59	Precautionary (booster) dose given to age group 18-59
Population	Total population of the state/UT

3) Data Cleaning Process for COVID-19 & Vaccination Tables

i) Loading the Datasets

```
df_covid = pd.read_csv("F:/Ramya_course/Final_project/final_data/complete.csv")
df_vaccine = pd.read_csv("F:/Ramya_course/Final_project/final_data/Vaccine.csv")
```

ii) Cleaning Process for both tables

```
df_vaccine['State/UTs'] = df_vaccine['State/UTs'].replace({
    'Andaman And Nicobar': 'Andaman and Nicobar Islands',
    'Dadra And Nagar Haveli And Daman And Diu': 'Dadra and Nagar Haveli and Daman and Diu',
    'Jammu And Kashmir': 'Jammu and Kashmir'
})
```

```
df_covid.duplicated()
df_covid.duplicated().sum()
```

```
df_covid.isnull()
df_covid.isnull().sum()
```

```
df_covid.replace("Telangana***", "Telangana", inplace=True)

df_covid['Name of State / UT'] = df_covid['Name of State / UT'].str.replace(
    r'^Telengana$', 'Telangana', regex=True
)
```

```
df_covid['Death'] = df_covid['Death'].replace('0#', 0)
```

```
df_covid['Latitude'] = df_covid['Latitude'].replace(0, 18.1124)
df_covid['Longitude'] = df_covid['Longitude'].replace(0, 79.0193)
```

```
df_covid['Name of State / UT'] = df_covid['Name of State / UT'].replace({
    'Union Territory of Chandigarh': 'Chandigarh',
    'Union Territory of Jammu and Kashmir': 'Jammu and Kashmir',
    'Union Territory of Ladakh': 'Ladakh'
})
```

iii) Saving the Cleaned Datasets

```
df_covid.to_csv("F:/Ramya_course/Final_project/final_data/cleaned_covid_data.csv", index=False)
df_vaccine.to_csv("F:/Ramya_course/Final_project/final_data/cleaned_vaccine_data.csv", index=False)
```

4) SQL Operations to Prepare Tables for Analysis

i) Create a New Database

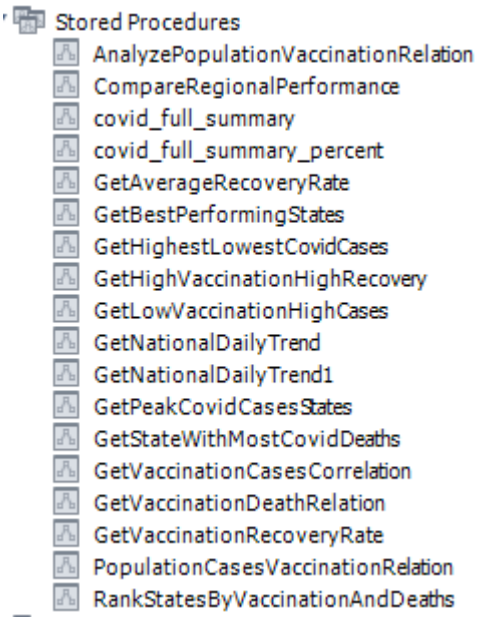
```
create database dummy;
use dummy;
```

ii) Create Tables for the Cleaned Data

```
CREATE TABLE vaccination_data (
    state_ut VARCHAR(100) PRIMARY KEY,
    total_vaccination_doses BIGINT,
    dose1 BIGINT,
    dose2 BIGINT,
    dose1_15_18 BIGINT,
    dose2_15_18 BIGINT,
    dose1_12_14 BIGINT,
    dose2_12_14 BIGINT,
    precaution_18_59 BIGINT,
    Population BIGINT
);
CREATE TABLE CovidCases (
    S_NO INT PRIMARY KEY,
    Date DATE,
    State_UT VARCHAR(100),
    Latitude FLOAT,
    Longitude FLOAT,
    TotalConfirmed INT,
    Death INT,
    CuredDischargedMigrated INT,
    NewCases INT,
    NewDeaths INT,
    NewRecovered INT,
    FOREIGN KEY (state_ut) REFERENCES vaccination_data(state_ut)
);
```

iii) Imported the Cleaned CSV Files into SQL Tables.

iv) Created Stored Procedures for COVID-19 and Vaccination Data



5) Connecting SQL to Python for Visualization and Dashboard Creation

a) Key Libraries Used:

- **mysql.connector** → Helps Python connect to your MySQL COVID-19 database.
- **pandas** → Helps load your COVID-19 tables into DataFrames and do cleaning.
- **matplotlib.pyplot & seaborn** → Used to draw COVID-19 charts (cases, deaths, vaccination, etc.).

b) Steps:

i) Set up database connection:

You first create a db_config dictionary that stores: (host, user, password, database name)
This allows Python to connect to your COVID-19 MySQL database.

- **Create one easy function to run stored procedures**

You write a function called **run_proc (proc_name, params).**
This function will:

- Call any stored procedure (e.g., total cases, vaccination summary, state-wise deaths, etc.)
- Receive the results using cursor.stored_results()
- Convert the output into a **pandas DataFrame**

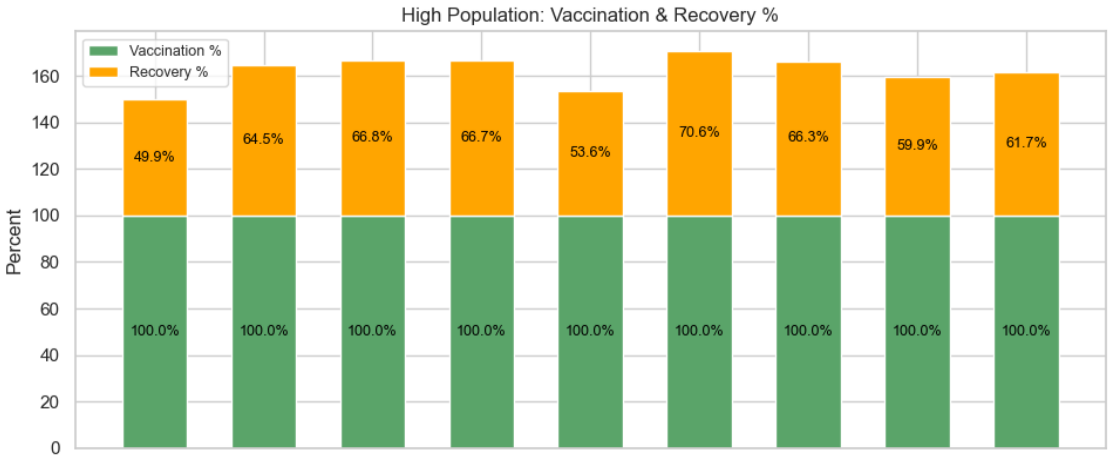
This makes it very simple to use your database results in Python.

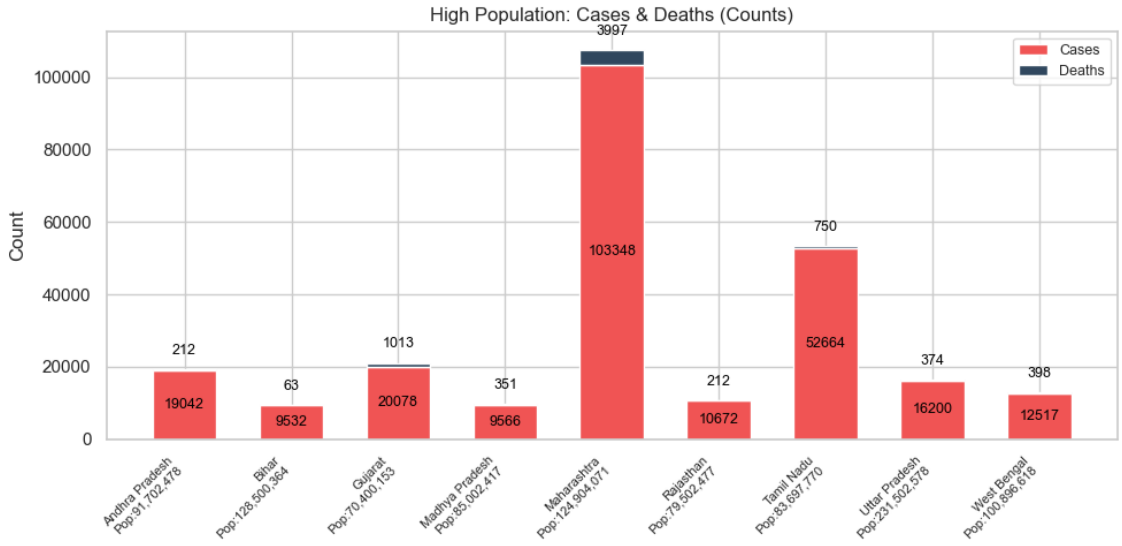
ii) Make separate functions for charts

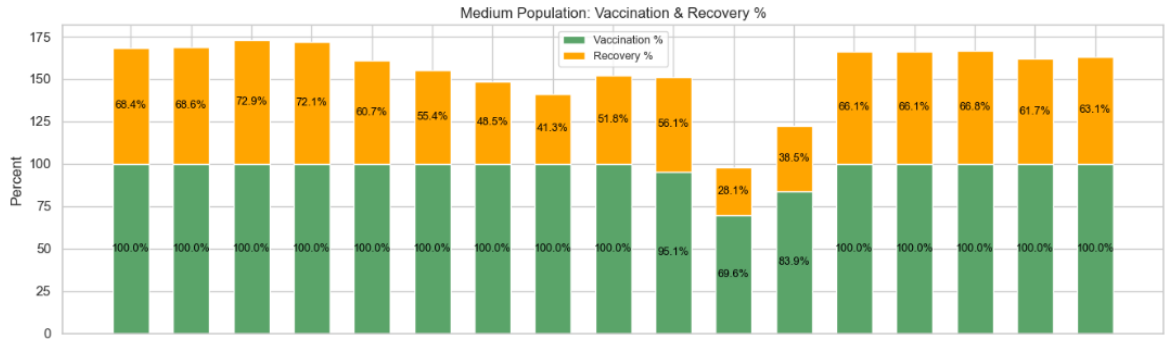
For each type of COVID-19 analysis, create one visualization function.
Examples:

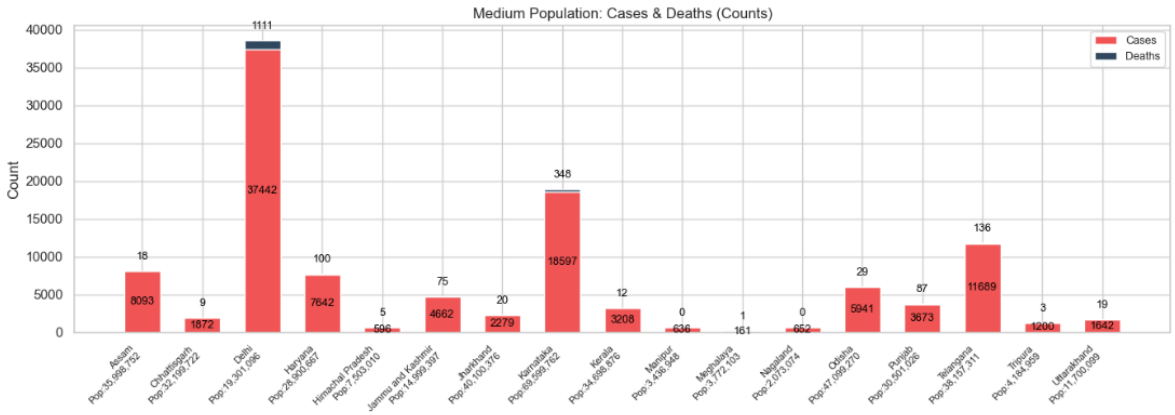
- Chart for **state-wise total cases**
- Chart for **vaccination comparison**
- Chart for **deaths vs recovery**
- Chart for **population categories (High, Medium, Low)**

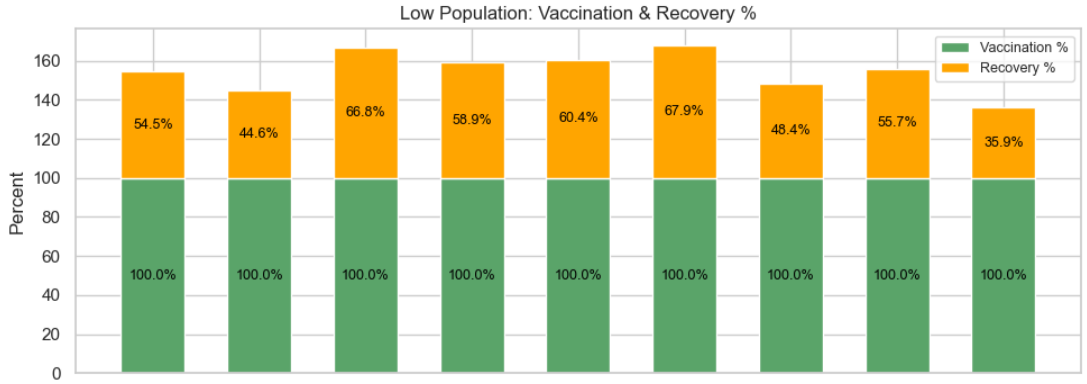
6) Dashboard Visualizations:

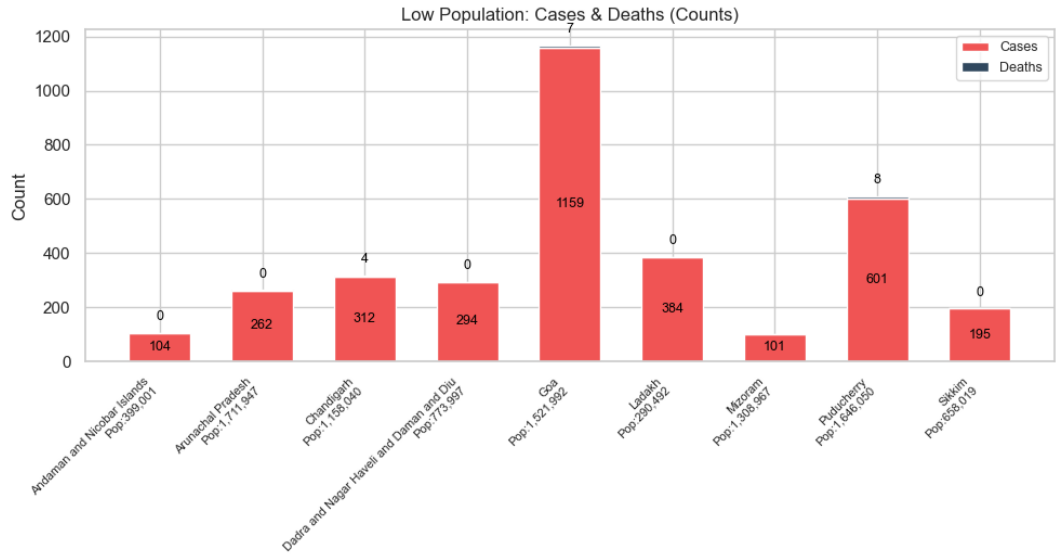












i) High Population State Analysis

States in the high population category have the largest number of people compared to other regions in India.

The Vaccination & Recovery % chart shows that almost all people have received the COVID-19 vaccine, and most have recovered from the disease, though recovery rates are slightly lower than vaccination rates in some states.

The Cases & Deaths chart displays total COVID-19 cases and deaths for each high-population state. For example, Maharashtra is the most affected state, reporting the highest number of cases (103,348) and deaths (3,997), which is much greater than other states in this

group.
Population size directly influences the numbers of cases, deaths, and recoveries; the largest states often see the biggest impacts, making them critical for understanding the overall pandemic in India.

ii) Medium Population States

Medium population states have a moderate number of people, so their COVID-19 numbers are not as high as the largest states but still important for analysis.
The vaccination chart shows that almost everyone is vaccinated, and recovery rates are different for each state.
Case and death numbers vary by state, but Delhi stands out as the most affected among medium-sized states, with the highest case count (37,442) and deaths (1,111), much more than any other state in this group.
By comparing these states, we can understand how COVID-19 affected people where populations are neither very large nor very small.

iii) Low Population States

Low population states have fewer people, so their COVID-19 cases and deaths are much lower compared to bigger states.
The vaccination chart shows almost all people have been vaccinated, and recovery rates vary from one state to another.
Most states in this group have small numbers of cases and deaths, but Goa is the most affected, reporting 1,159 cases and 7 deaths, which is higher than other low population states.

7) Key Insights

COVID-19 affected regions differently based on population. States with higher populations—such as Maharashtra—reported the largest numbers of cases and deaths, making them vital for understanding the national impact.

Among medium population states, Delhi experienced a much heavier impact compared to others, with significantly higher case and death counts, highlighting regional differences even within similar population groups.

In low population states, most regions had only a few cases and deaths, but Goa was an outlier, reporting higher numbers than its peers, underscoring that smaller populations are not always shielded from severe outbreaks.

Vaccination coverage across all population categories was very high, suggesting effective rollout and strong government/public response nation-wide.

Recovery rates showed more variation between states, indicating that factors beyond vaccination—such as healthcare infrastructure and public behaviour—also played a major role in pandemic outcomes.

8) COVID-19 Awareness and Prevention Education

To effectively educate people about COVID-19 and how to protect themselves, awareness programs should be conducted regularly in schools, colleges, companies, and offices. These programs can teach individuals about the importance of hygiene practices such as frequent hand washing, using hand sanitizer, wearing masks properly, and maintaining physical distancing. Educational sessions can demonstrate how keeping personal and workplace environments clean reduces the spread of the virus. Schools and colleges can share age-appropriate information with students, while companies and offices can organize training and reminders for employees on safety protocols. Such widespread, organized efforts help create a culture of health consciousness and collective responsibility to prevent COVID-19 transmission in all settings.

In addition to urban and institutional awareness, it is equally important to educate villagers through local ward members, health volunteers, and ASHA workers. They can conduct door-to-door campaigns, community meetings, and small group sessions to explain the symptoms of COVID-19, how it spreads, and the steps villagers can take to protect themselves. ASHA workers can demonstrate simple practices such as proper mask usage, handwashing techniques, and maintaining distance during social gatherings. They can also guide villagers on what to do if someone develops symptoms—such as isolating the individual, seeking early medical help, monitoring oxygen levels, and avoiding panic.

Moreover, educating villagers on the importance of good nutrition is crucial. Encouraging them to consume healthy food such as green leafy vegetables, seasonal fruits, turmeric, ginger, and vitamin-rich diets helps improve immunity. Information on maintaining mental well-being, staying hydrated, and following doctor-recommended medications during illness assists villagers in overcoming COVID-19 more effectively. With the collective effort of ward members, ASHA workers, and community leaders, rural areas can become better prepared to prevent and manage COVID-19 situations.