Impact of COVID-19 Across Indian States

A Data-Driven Analytical Study

This project analyzes COVID-19 case and vaccination trends across Indian states using real datasets. Python was used for data cleaning and exploratory analysis, SQL was used for deriving state-wise insights, and Tableau was used to visualize patterns and comparisons.

Project Overview

Our Approach

This comprehensive study examines the pandemic's impact and healthcare response across India. We analyzed state-wise COVID-19 case data and vaccination records to identify the most affected regions, recovery and mortality trends, and vaccination progress.

The results provide crucial insights into understanding how different states responded to the pandemic and the effectiveness of healthcare interventions.

01

Data Collection

Gathered state-wise COVID-19 case data and vaccination records

02

Python Analysis

Performed data cleaning and exploratory analysis

03

SQL Insights

Derived state-wise analytical insights

04

Tableau Visualization

Created interactive dashboards for pattern analysis

Dataset Summary

The project uses state-wise COVID-19 case data and vaccination records collected during the pandemic period in India. The dataset is clean, structured, and suitable for time-series and comparative state-level analysis.

Date

Time-series tracking of pandemic progression

State

Geographic distribution across Indian states

Confirmed Cases

Total positive COVID-19 diagnoses

Recovered Cases

Patients who successfully recovered

Deaths

Mortality data for impact assessment

Vaccination Counts

Immunization progress tracking

Calculated Field: An additional field **Active Cases** was derived to analyze ongoing case pressure on healthcare systems using the formula: Active Cases = Confirmed – (Recovered + Deaths)

Exploratory Data Analysis Using Python

Exploratory Data Analysis was performed in Python to understand the spread of COVID-19 across different states and time periods. The dataset was first cleaned by removing unnecessary columns and converting the Date column into proper datetime format.

Analysis Process

State-wise summaries were generated to identify the top affected states, while recovery and mortality rates were calculated to measure health outcomes. Visualization libraries such as **Matplotlib** and **Seaborn** were used to create bar charts and line graphs for trend comparison.

```
covid_df['Active_Cases'] = covid_df['Confirmed'] - (covid_df['Cured'] +
covid_df['Deaths'])

top_10_active = covid_df.groupby("State/UnionTerritory").max()
[["Active_Cases"]].sort_values("Active_Cases", ascending=False).head(10)

sns.barplot(data=top_10_active, x=top_10_active.index, y="Active_Cases")
plt.xticks(rotation=45)
plt.title("Top 10 States with Highest Active Cases")
plt.show()
```

Key Libraries

- **Pandas:** Data manipulation and cleaning
- **Matplotlib:** Creating visualizations
- **Seaborn:** Statistical graphics
- **NumPy:** Numerical computations

Python Analysis Results

Top 10 States with Highest Active Cases

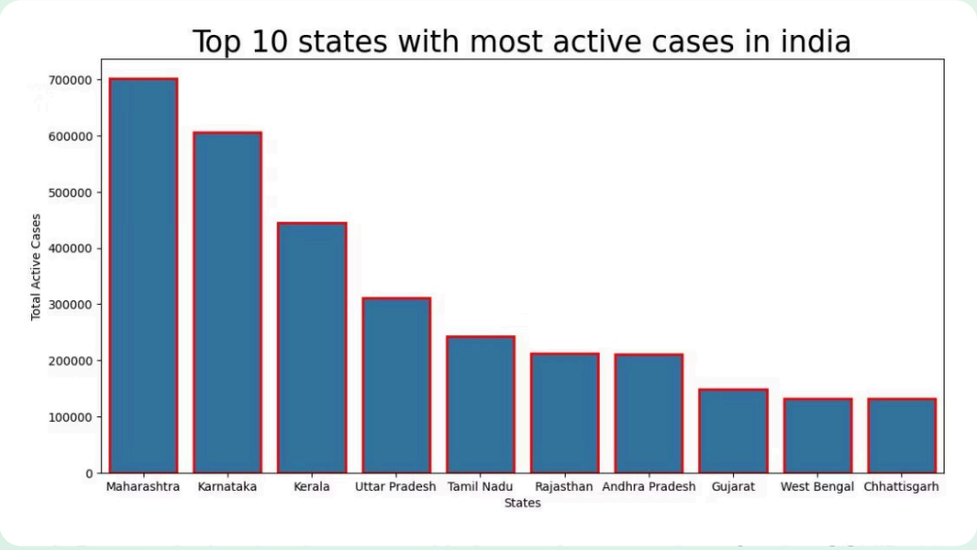


Figure 1: Bar graph showing the states with the maximum number of active COVID-19 cases

Top 10 States with Highest Death Cases

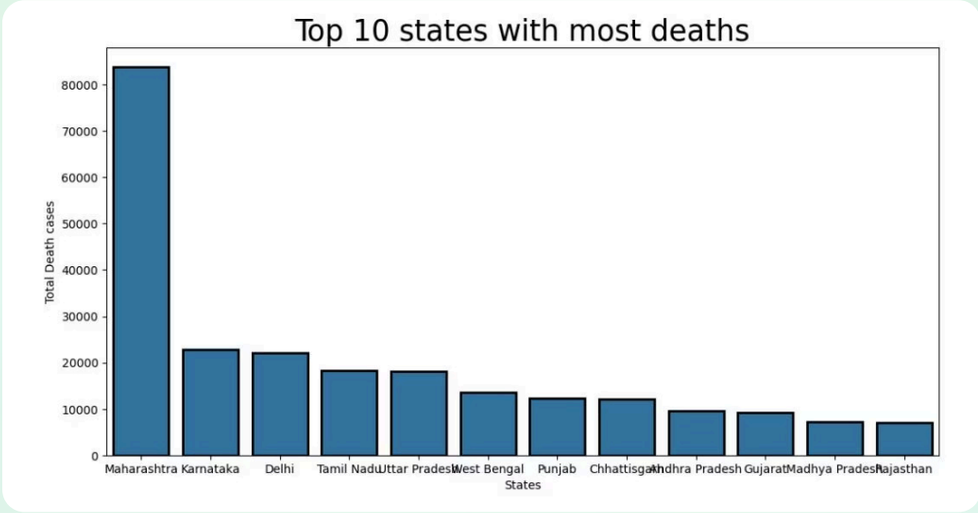


Figure 2: Bar graph showing the states with the maximum number of death COVID-19 cases

Most Impacted States

Maharashtra, Kerala, Karnataka, and Tamil Nadu were the most heavily impacted states

Recovery Improvement

Recovery rates increased over time due to improved medical response and vaccination

Regional Mortality Variation

Mortality varied by region, influenced by healthcare infrastructure and population density

Wave Patterns

Multiple waves of increasing and decreasing case trends were observed throughout the timeline

Data Analysis Using SQL

SQL was used to perform analytical queries on the cleaned COVID-19 dataset stored in PostgreSQL. The purpose was to identify the most affected states, calculate national recovery and mortality rates, and study vaccination distribution across regions.

Data Storage

PostgreSQL database for structured COVID-19 data

Query Analysis

Aggregation functions: SUM, MAX, GROUP BY

Trend Identification

Ranking states by case severity and outcomes

Insights

Deeper interpretation of spread and health outcomes

Key SQL Queries and Insights

Total National Impact

```
SELECT SUM(confirmed) AS total_confirmed,
SUM(cured) AS total_recovered,
SUM(deaths) AS total_deaths
FROM covid_cases;
```

Insight: Shows the overall impact of COVID-19 across the country

Top 10 States by Active Cases

```
SELECT state, MAX(active_cases) AS highest_active
FROM covid_cases
GROUP BY state
ORDER BY highest_active DESC
LIMIT 10;
```

Insight: Identified Maharashtra, Kerala, Karnataka, and Tamil Nadu as the most affected states

Recovery and Mortality Rates

```
SELECT
(SUM(cured)::float / SUM(confirmed)) * 100 AS recovery_rate,
(SUM(deaths)::float / SUM(confirmed)) * 100 AS mortality_rate
FROM covid_cases;
```

Insight: Recovery rate improved over time, while mortality remained comparatively low

Monthly Trend Analysis

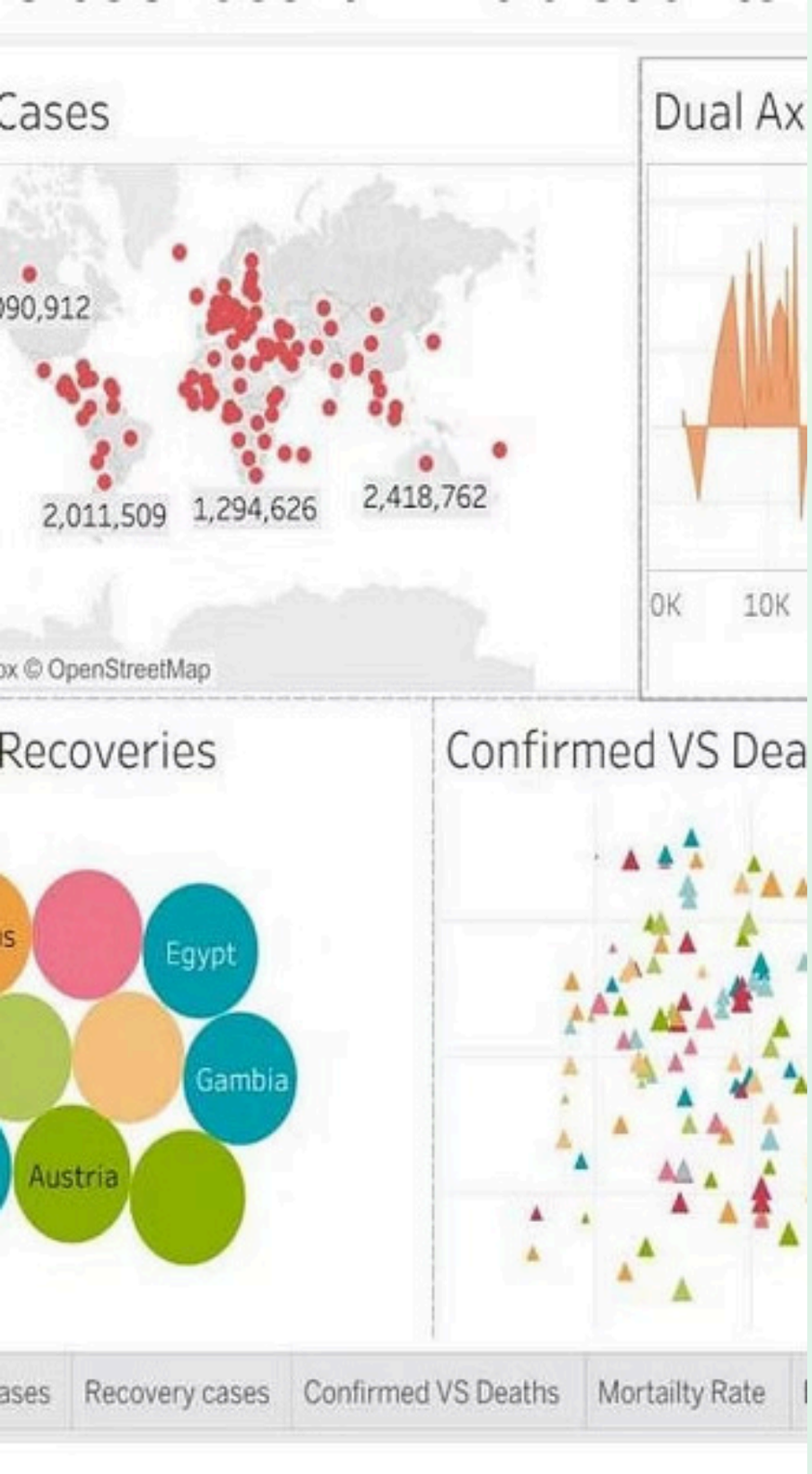
```
SELECT DATE_TRUNC('month', date) AS month,
    SUM(confirmed) AS monthly_cases
FROM covid_cases
GROUP BY month
ORDER BY month;
```

Insight: Clearly shows wave patterns and rise/fall cycles during pandemic periods

Vaccination Gender Distribution

```
SELECT SUM(male_vaccinated) AS total_male,
SUM(female_vaccinated) AS total_female
FROM covid_vaccination;
```

Insight: Vaccination among males was slightly higher, but gap narrowed over time



Interactive Dashboard in Tableau

An interactive dashboard was built in **Tableau** to present insights visually and enable dynamic exploration of COVID-19 trends across Indian states.

Figure 3: Tableau dashboard displaying overall COVID-19 case statistics

Consistent Top States

Maharashtra, Kerala, Karnataka, and Tamil Nadu were consistently the most affected states

Recovery Improvement

Recovery rate improved significantly across several states after vaccination rollout

Wave Patterns

Case trends show multiple waves, indicating periodic outbreak patterns

State Comparison

Visual comparison helps understand how different states responded to the pandemic

Key Findings

Finding	Interpretation
Maharashtra, Kerala, Karnataka and Tamil Nadu had the highest number of confirmed and active cases	These states have high population density and major urban centers, increasing transmission risk
The recovery rate improved steadily over time across most states	Indicates better medical support, awareness, and vaccination effect
Mortality rates varied between states	Differences in healthcare infrastructure and response strategies influenced outcomes
Multiple waves of infection were observed during the timeline	Shows that the virus spread was periodic and influenced by public movement and safety measures
Vaccination rates increased significantly, reducing active cases in later stages	Vaccination played a key role in controlling the pandemic impact

Conclusion and Future Scope

Project Conclusion

This project successfully analyzed the spread and impact of COVID-19 across India using real case and vaccination data. Python enabled efficient data cleaning and visualization, SQL provided deeper analytical insights, and Tableau helped represent the results interactively.

The findings highlight the most affected states, improvements in recovery, and the role of vaccination in reducing case severity. Overall, the project provides meaningful understanding of pandemic trends and supports decision-making in public health analysis.

Future Scope

Machine Learning Forecasting

Apply ML models to forecast upcoming case trends

Live Data Integration

Automate data updates using live API integration

Web-Based Dashboard

Create a web platform for real-time monitoring