COVID-19 in India – Data Analysis Using Python

Prepared By: Harsh Belekar

Date: 30/12/2025

Tools Used: Python, Pandas, NumPy, Matplotlib, Seaborn, Jupyter Notebook

Objective:

To analyze COVID-19 case trends and vaccination progress in India using data-driven exploratory analysis, identify key patterns, and derive meaningful insights to support public health understanding and decision-making.

2. Executive Summary

This project presents an in-depth analysis of COVID-19 cases and vaccination data in India using Python-based data analysis techniques. The study focuses on understanding trends in confirmed cases, recoveries, deaths, and vaccination progress across different states, time periods, genders, and age groups. Through effective data cleaning, feature engineering, and exploratory data analysis (EDA), meaningful insights were extracted to highlight pandemic waves, seasonal patterns, and vaccination coverage. The results emphasize the importance of timely vaccination drives, state-level preparedness, and inclusive healthcare strategies.

3. Dataset Description

The project uses two datasets:

1. **COVID-19 India Dataset (covid_19_india.csv)**
    - Daily confirmed, recovered, and death cases
    - State/Union Territory-level data
    - Time-series data across multiple years
2. **COVID-19 Vaccination Dataset (covid_vaccine_statewise.csv)**
    - State-wise vaccination statistics
    - Gender-wise and age-group-wise vaccination data
    - Dose-level vaccination information

Both datasets were preprocessed and standardized to ensure consistency and accuracy for analysis.

4. Data Cleaning & Preprocessing

The following steps were performed to ensure data quality and consistency:

- Standardized column names for readability.
- Converted date columns to datetime format.
- Handled missing values and removed duplicate records.
- Corrected inconsistent state names and merged duplicate state entries.
- Removed non-state and invalid records.

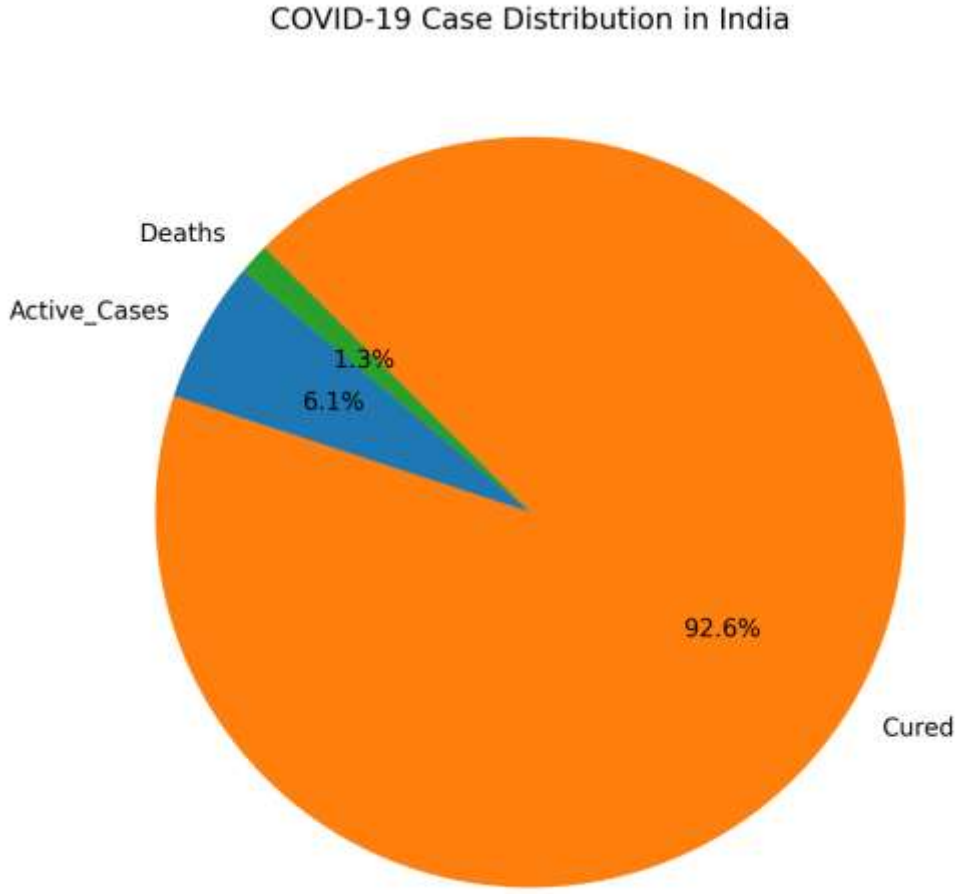
- Created additional time-based features:
  - Year
  - Month
  - Month Name
  - Day

Cleaned datasets were saved separately and reused for analysis to follow best data engineering practices.

5. Exploratory Data Analysis (EDA)

1. COVID-19 Case Distribution in India

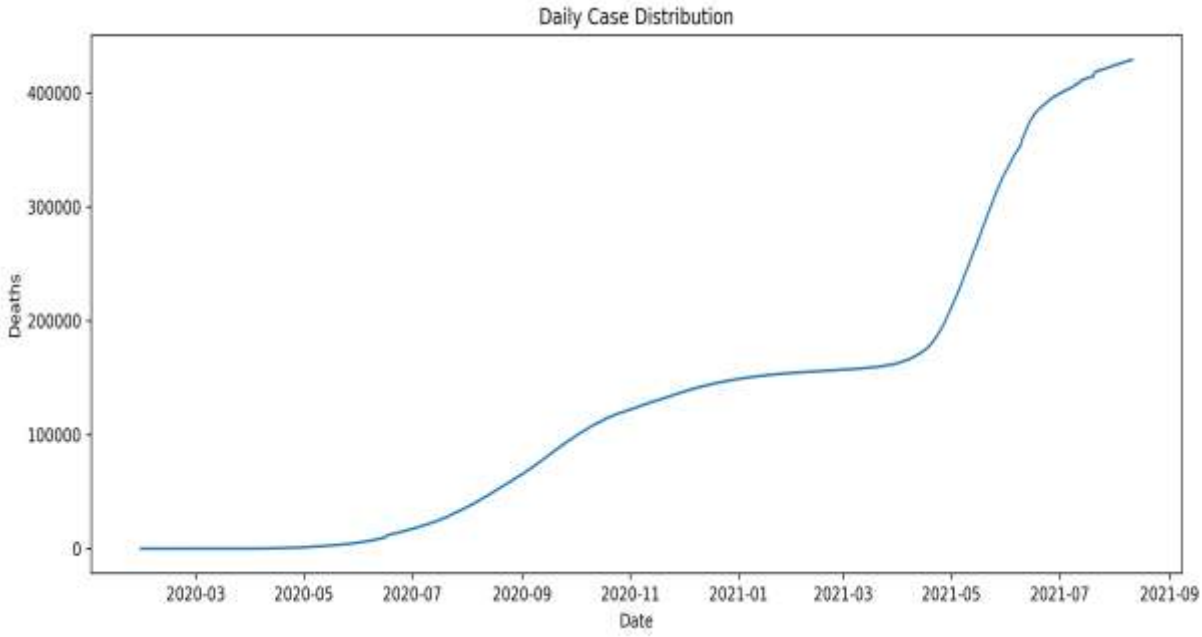
This pie chart shows the overall proportion of confirmed, cured, and death cases in India.



Insight: A significant majority of confirmed cases resulted in recovery, while deaths constituted a relatively small percentage, indicating improved treatment and healthcare response over time.

2. Daily Deaths Case Trend

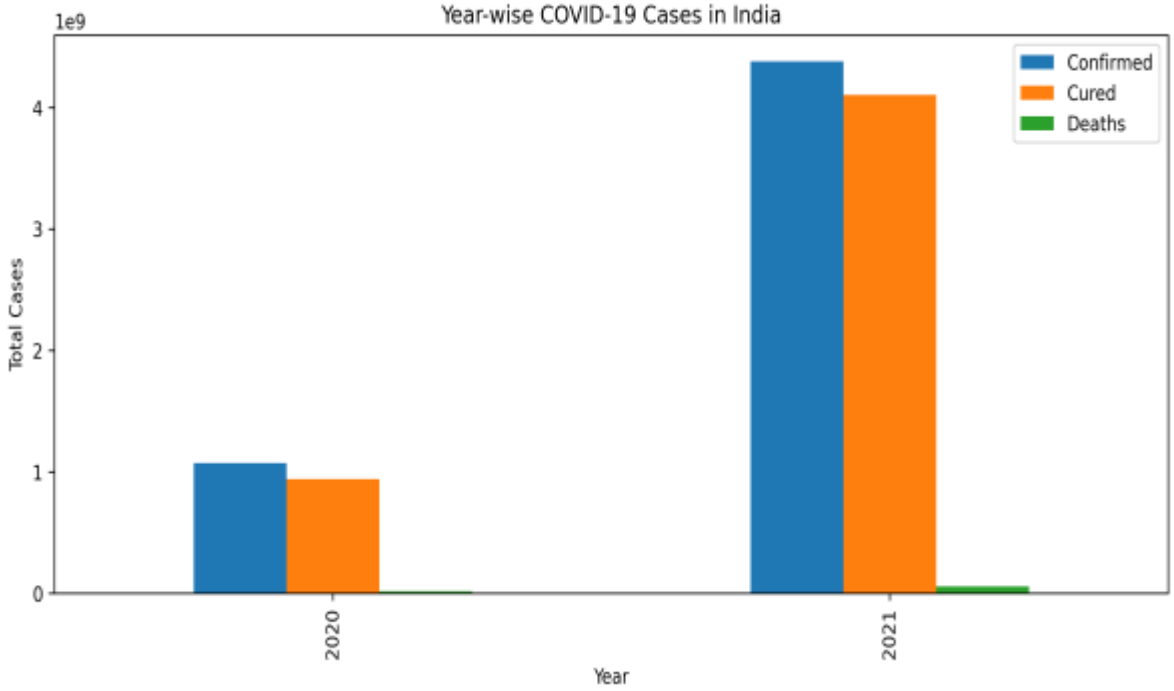
Represents daily reported deaths.



Insight: Death counts increased during peak infection periods but remained significantly lower than confirmed cases, suggesting improved clinical management over time.

3. Year-wise COVID-19 Cases in India

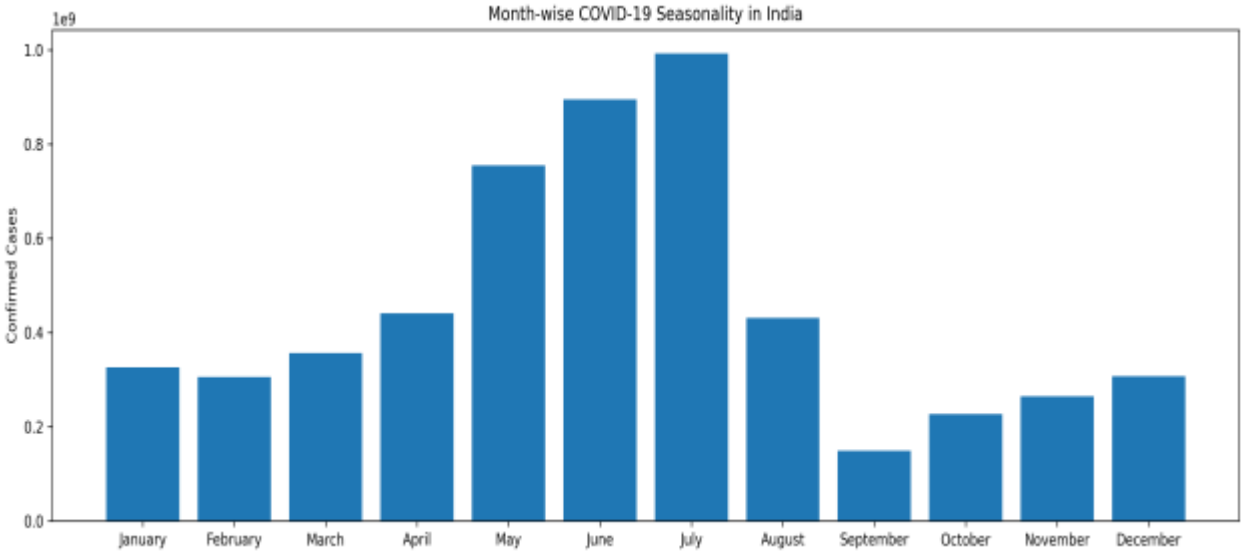
Compares total cases year by year.



Insight: The year 2021 recorded the highest number of cases, corresponding to major pandemic waves.

4. Month-wise COVID-19 Seasonality in India

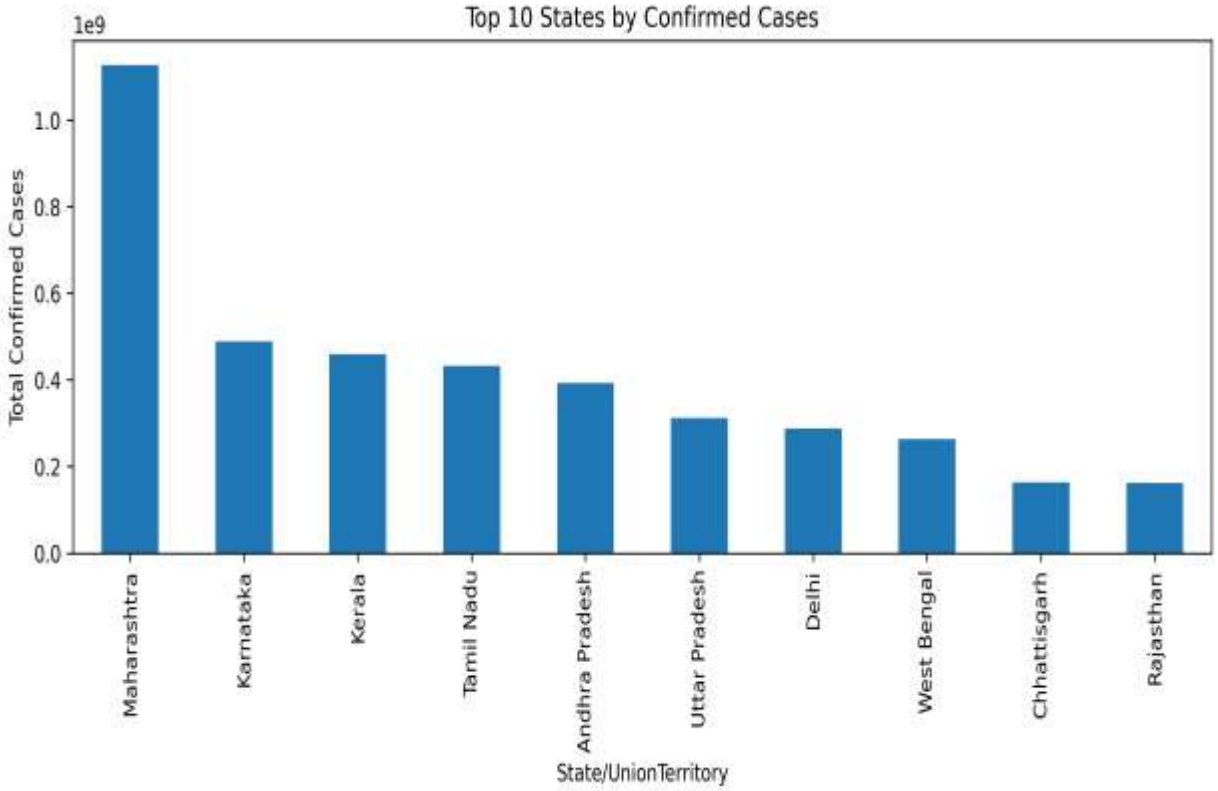
Shows seasonal patterns in case numbers.



Insight: Case surges were observed during specific months, indicating seasonal and behavioral influences on virus spread.

5. Top 10 States by Confirmed Cases

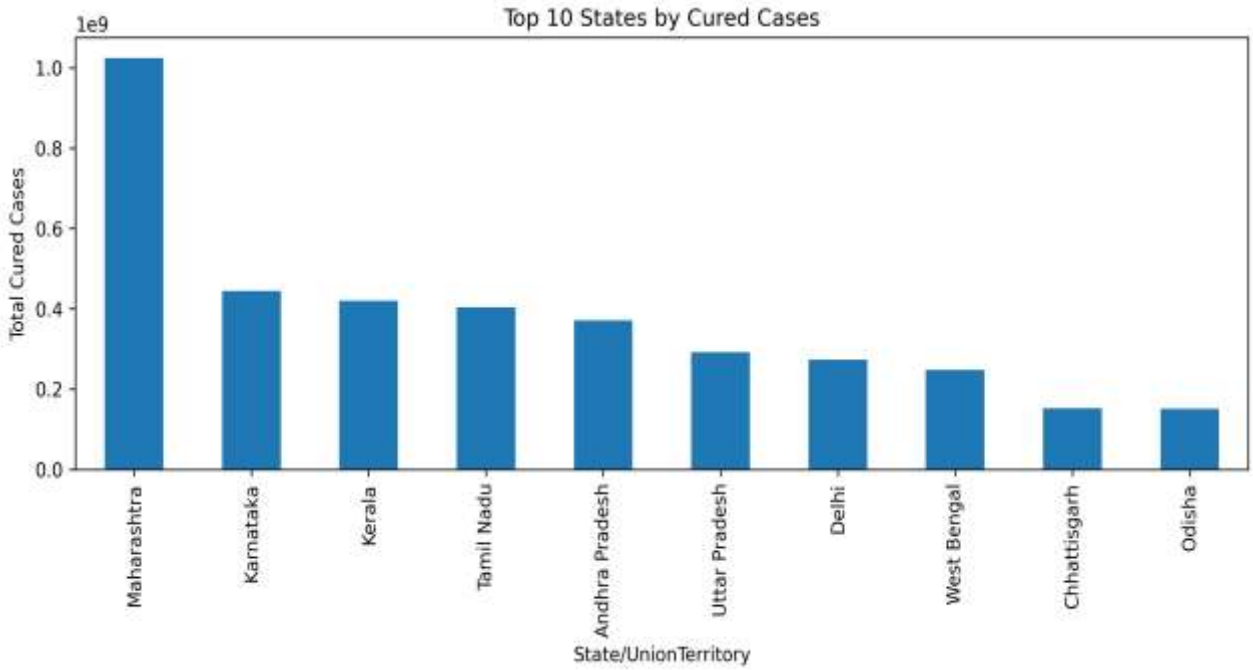
Shows states with the highest total cases.



Insight: Highly populated and urbanized states reported the highest number of confirmed cases.

6. Top 10 States by Cured Cases

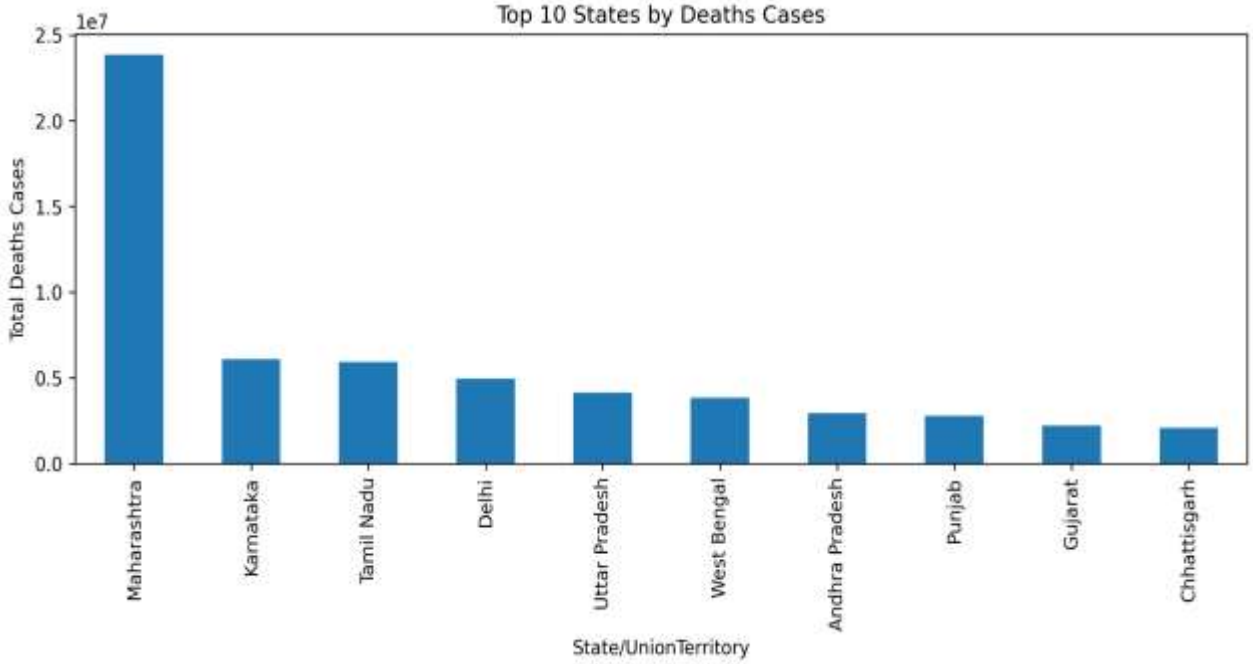
Displays states with the most recoveries.



Insight: States with better healthcare infrastructure showed higher recovery numbers.

7. Top 10 States by Deaths Cases

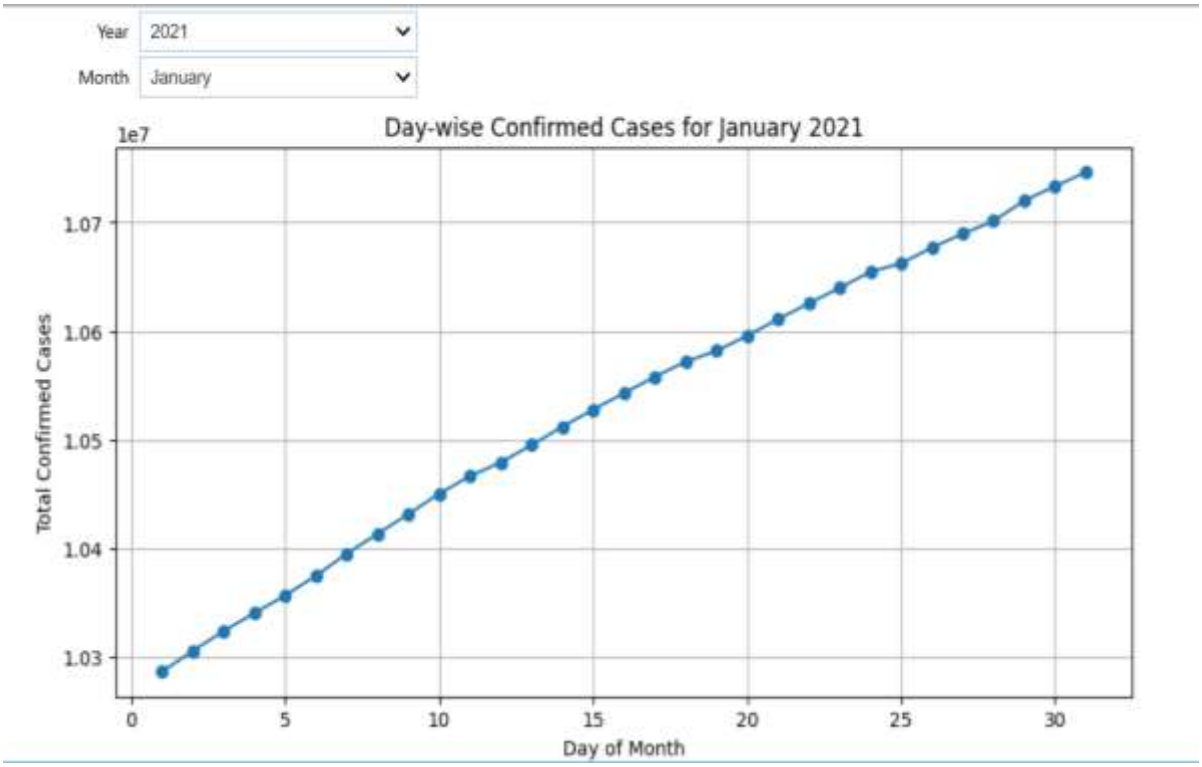
Highlights states with the highest deaths.



Insight: Higher deaths correlate with higher infection volumes rather than fatality rate alone.

8. Day-wise Confirmed Cases by Year & Month

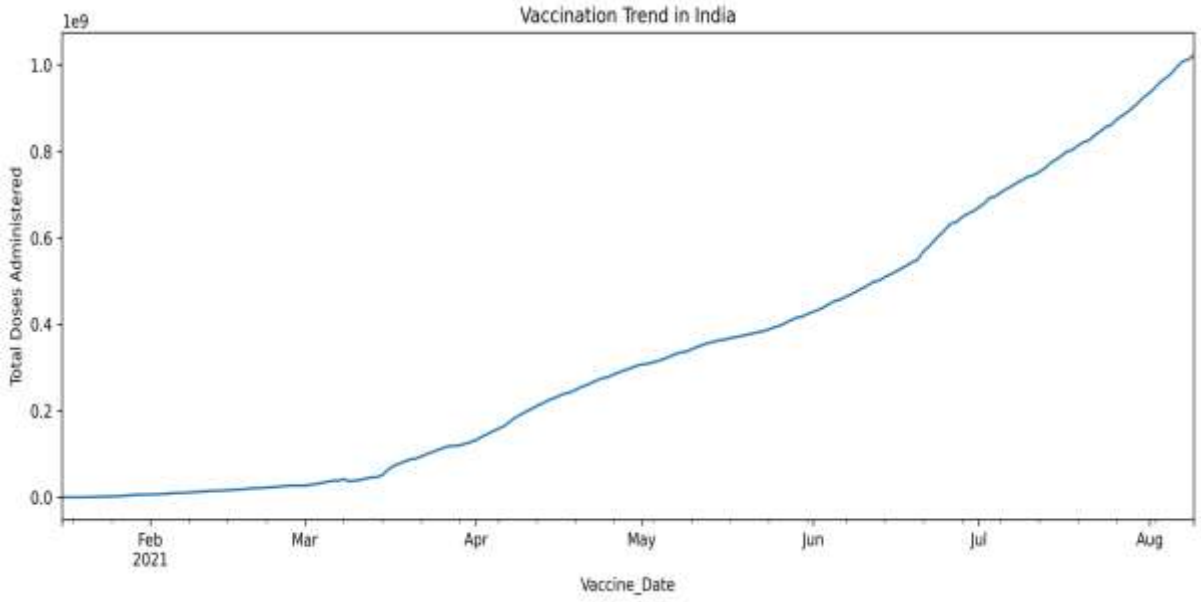
Analyzes daily patterns across months and years.



Insight: Sharp spikes indicate sudden outbreak waves, emphasizing the importance of early intervention.

9. Vaccination Trend in India

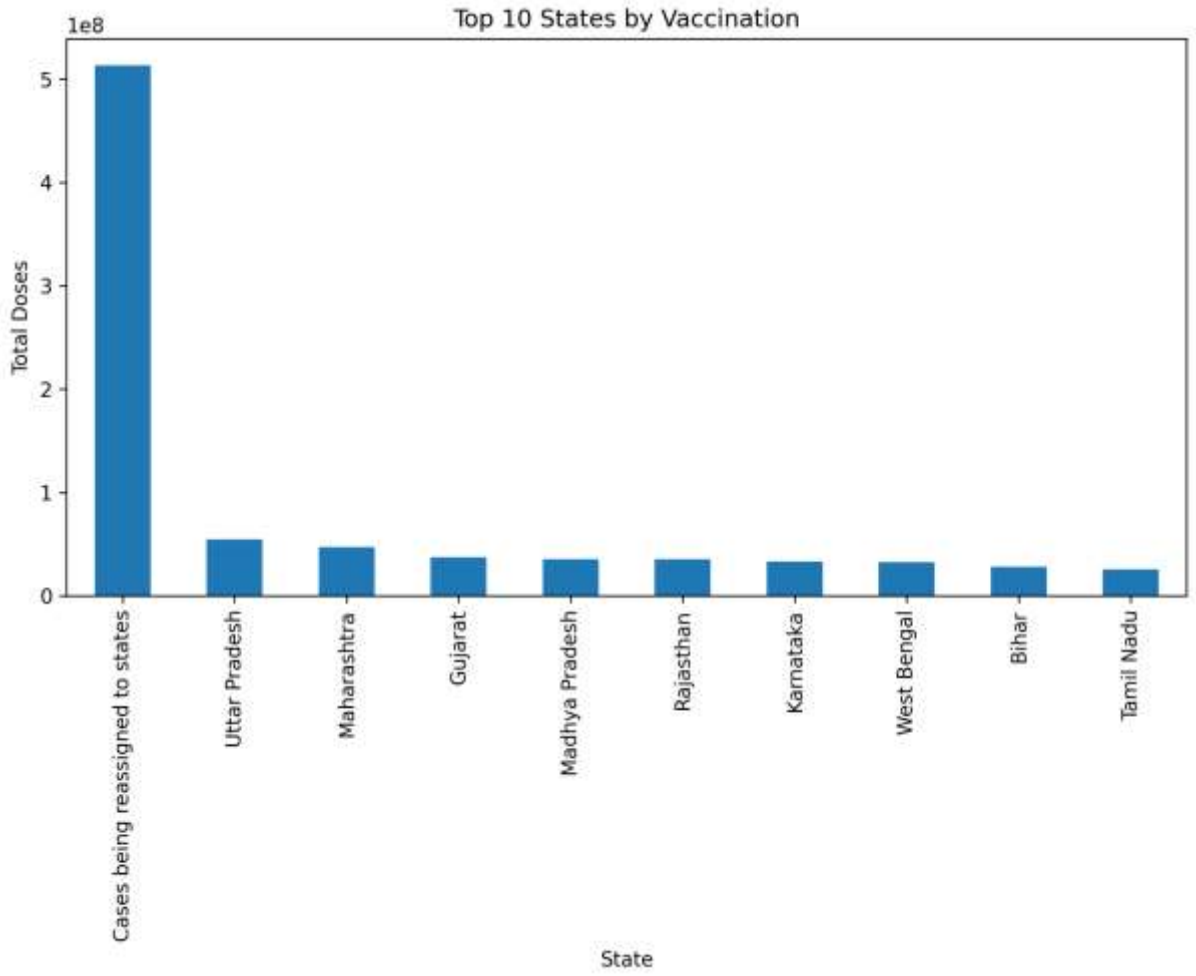
Displays cumulative vaccination progress over time.



Insight: A rapid acceleration phase is visible after vaccine rollout, followed by stabilization.

10. Top 10 States by Vaccination

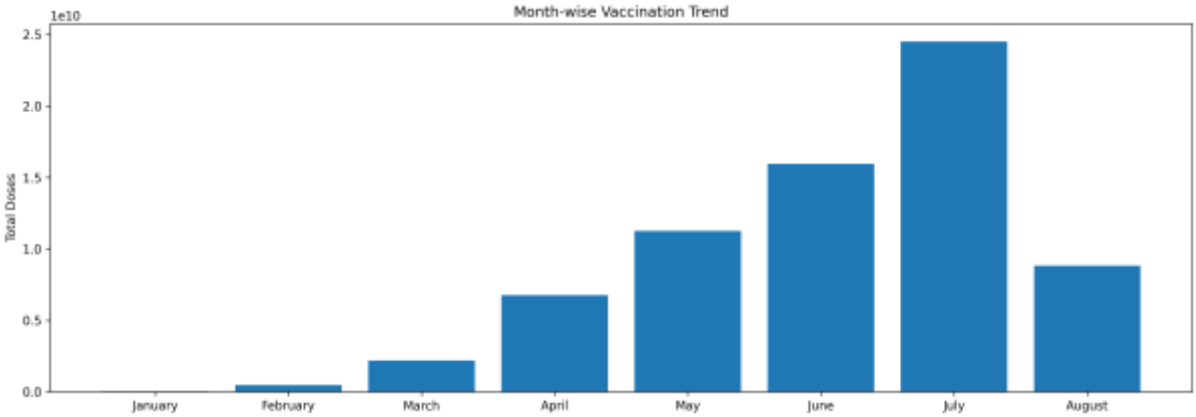
Shows states with the highest vaccination coverage.



Insight: States with strong public health infrastructure achieved faster vaccine rollout.

11. Month-wise Vaccination Trend

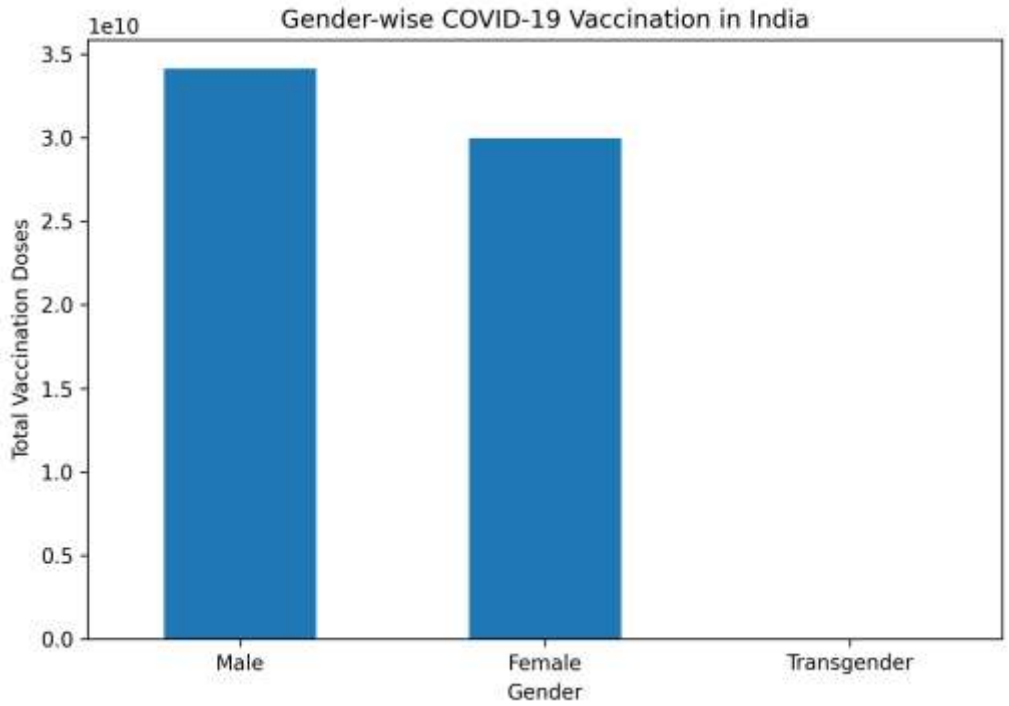
Analyzes monthly vaccination activity.



Insight: Peak vaccination occurred during high-risk periods, reflecting responsive policy actions.

12. Gender-wise COVID-19 Vaccination in India

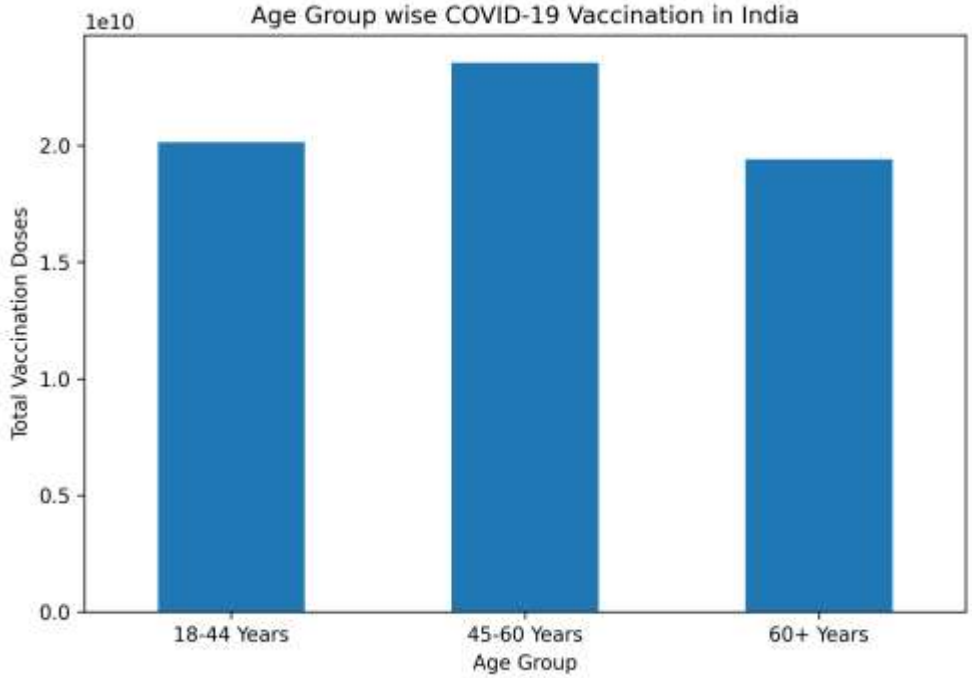
Compares male, female, and transgender vaccination.



Insight: Male and female vaccination counts are relatively comparable, while transgender vaccination remains significantly lower, highlighting the need for inclusive healthcare initiatives.

13. Age-Group-wise COVID-19 Vaccination in India

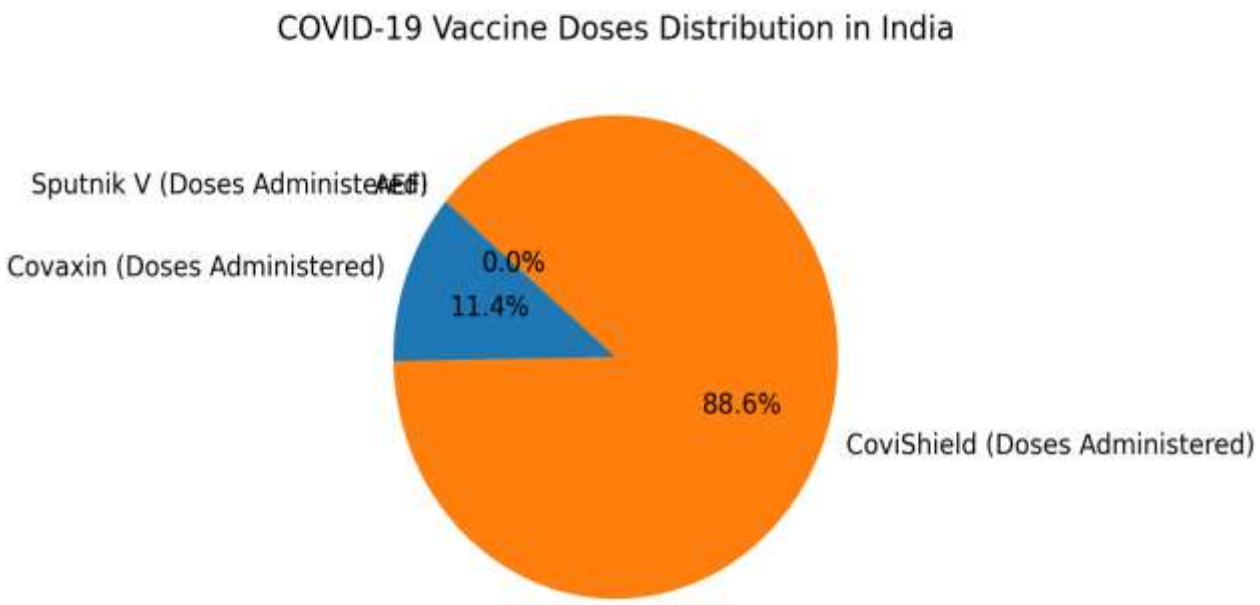
Shows vaccination across age groups.



Insight: Higher vaccination coverage in working-age and senior populations reflects prioritization strategies.

14. COVID-19 Vaccine Doses Distribution in India

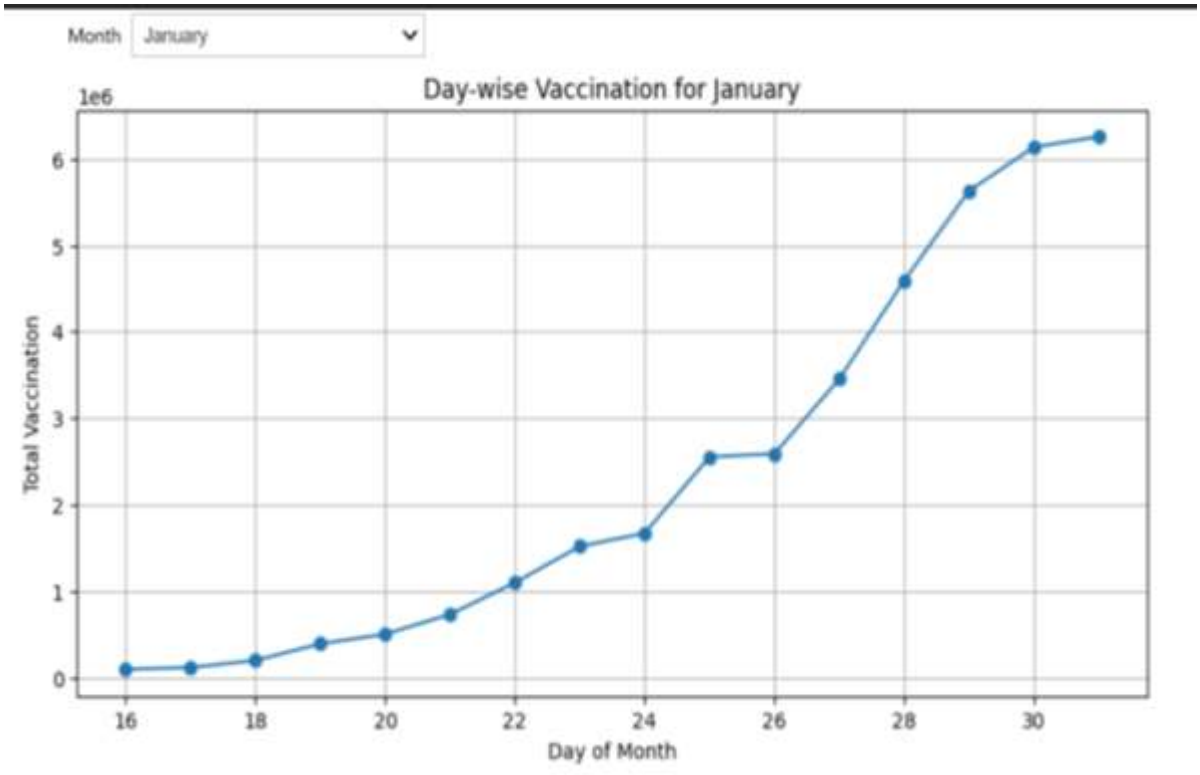
Displays overall dose distribution.



Insight: First-dose coverage dominates early stages, followed by a steady rise in second doses.

15. Day-wise Vaccination by Month

Shows daily vaccination trends within months.



Insight: Sudden spikes indicate vaccination drives and government-led campaigns.

6. Key Insights Summary

- The COVID-19 outbreak in India followed **distinct multi-wave patterns**, with a significant surge observed during the year 2021, highlighting the impact of highly transmissible variants.
- **Recovery rates consistently remained high** relative to confirmed cases, indicating improvements in clinical treatment, healthcare capacity, and disease management over time.
- **Death rates, although comparatively low**, showed noticeable spikes during peak infection periods, emphasizing the importance of early intervention and healthcare readiness during surge phases.
- **Seasonal trends** revealed higher infection rates during specific months, suggesting that climatic conditions, population movement, and public behavior influenced virus transmission.
- **State-wise analysis** showed that highly populated and urbanized states contributed the largest share of confirmed cases, recoveries, and deaths, indicating a strong correlation between population density and virus spread.
- The **vaccination rollout in India accelerated rapidly** after its introduction, with clear month-wise growth reflecting large-scale government vaccination drives.
- States with stronger healthcare infrastructure and higher population sizes achieved **higher vaccination coverage**, demonstrating the role of logistics and accessibility in public health campaigns.
- **Gender-wise vaccination analysis** showed relatively balanced coverage between males and females, while vaccination among the transgender population remained significantly lower, highlighting the need for more inclusive outreach initiatives.
- **Age-group-wise vaccination trends** reflected policy-driven prioritization, with higher coverage among older age groups during early rollout phases to reduce severe outcomes and mortality.
- **Vaccine dose distribution analysis** indicated a structured, phased approach, beginning with primary doses and later expanding to booster doses to enhance long-term immunity.

7. Recommendations

- **Strengthen early-warning systems** by continuously monitoring daily case, recovery, and death trends to enable faster detection of future outbreak surges.
- **Improve healthcare surge preparedness** in high-population and high-case states by increasing hospital capacity, oxygen supply, and critical care infrastructure during peak periods.
- **Adopt season-based preventive strategies**, such as targeted awareness campaigns and mobility restrictions during months historically associated with higher infection rates.
- **Enhance state-level resource allocation** by prioritizing regions with consistently high confirmed cases and lower recovery-to-case ratios to reduce mortality risk.
- **Expand inclusive vaccination outreach**, with special focus on underrepresented groups, including the transgender population, to ensure equitable healthcare access.
- **Promote age-specific vaccination strategies**, ensuring timely booster dose administration for elderly and high-risk groups to minimize severe outcomes.

- **Leverage data-driven vaccination planning** by using month-wise and state-wise vaccination trends to optimize supply chain management and distribution efficiency.
- **Encourage public-private collaboration** to strengthen healthcare delivery, vaccination coverage, and emergency response capabilities during health crises.
- **Invest in digital health infrastructure** to improve real-time data reporting, transparency, and decision-making at national and state levels.
- **Maintain long-term pandemic preparedness plans**, including periodic vaccination drives, public education programs, and continuous epidemiological monitoring.

8. Conclusion

This project presents a comprehensive analysis of **COVID-19 case trends and vaccination progress in India** using structured data exploration and visualization techniques. By analyzing confirmed, cured, and death cases alongside vaccination data, the study highlights the evolving nature of the pandemic and the impact of large-scale immunization efforts.

The **Exploratory Data Analysis (EDA)** revealed clear temporal patterns, including distinct daily, monthly, and yearly trends in COVID-19 cases. Seasonal surges emphasized the importance of timely interventions, while state-wise analysis exposed regional disparities in infection rates, recoveries, and fatalities. These findings underscore the need for localized and data-driven public health strategies.

Vaccination analysis demonstrated a **steady increase in coverage over time**, with noticeable variations across states, gender groups, age categories, and dose types. The insights confirm the critical role of vaccination in reducing severe outcomes and improving recovery rates, while also highlighting areas where equitable access can be further strengthened.

Overall, this project showcases how **data analytics can support informed decision-making in public health**. The insights derived from the analysis can assist policymakers, healthcare administrators, and organizations in improving preparedness, optimizing resource allocation, and strengthening future pandemic response strategies. The methodology and findings reinforce the value of **data-driven approaches in managing large-scale health crises**.