Post-COVID Data Analysis Report

Author: Baireddy Anjali

Email: baireddyanjali683@gmail.com

GitHub: https://github.com/anjalibaireddy/

LinkedIn: https://www.linkedin.com/in/baireddy-anjali-080405314/

---

1. Dataset Description

1.1 Source:

Publicly available COVID-19 dataset collected from multiple Indian health data portals, covering confirmed, recovered, and deceased case counts during the post-pandemic recovery phase (2020–2023).

1.2 Columns:

- **State:** Indian state or union territory
- **Date:** Reporting date
- **Confirmed:** Total number of confirmed COVID-19 cases
- **Recovered:** Number of recovered patients
- **Deceased:** Number of deaths reported
- **Active:** Currently active cases derived from totals

1.3 Data Quality:

- No missing or null values after preprocessing
- Uniform data types and consistent date formatting
- Verified accuracy by cross-referencing with government data sources
- Clean and structured dataset, ready for time-series and regional analysis

---

2. Operations Performed

2.1 Data Cleaning & Preparation

- Checked for null or inconsistent entries; standardized all date formats
- Computed new features such as active cases and recovery rate
- Removed duplicate records and validated cumulative totals

2.2 Exploratory Data Analysis (EDA)

- State-wise comparison of confirmed and recovered cases (bar charts)
- Trend visualization of national case progression (line charts)
- Correlation analysis among confirmed, recovered, and deceased cases
- Monthly and seasonal breakdown to observe recovery patterns

2.3 Statistical Analysis

- Calculated mean, median, and standard deviation for case counts
- Identified peaks and troughs in case trends by month
- Applied moving-average smoothing for time-series stabilization

---

3. Key Insights

3.1 National Recovery Trends

- Post-2022, recovery rates exceeded 95% across most states
- A steady decline in active cases was observed from mid-2022 onward

3.2 Regional Analysis

- Maharashtra, Kerala, and Karnataka recorded the highest total case counts
- Smaller northeastern states showed faster recovery percentages

3.3 Seasonal & Temporal Patterns

- Post-monsoon months (Oct–Dec) exhibited higher recovery surges
- Early 2021 showed residual case spikes before a steady decline

3.4 Correlation Insights

- Strong positive correlation ($r \approx 0.98$) between confirmed and recovered cases
- Mild correlation with deaths, indicating effective containment post-vaccination

---

4. Recommendations

- Encourage vaccination drives and continuous booster campaigns
- Strengthen healthcare infrastructure in rural and high-density urban regions
- Implement predictive modeling for early outbreak detection and response planning
- Create interactive dashboards for public health monitoring and awareness
- Integrate socio-economic recovery metrics with health data for holistic insights