covid_analysis

May 12, 2025

```
[ ]:  # COVID-19 Cases and Vaccination Analysis

      ## Introduction

      This notebook analyzes COVID-19 cases, deaths, and vaccination progress for
       ↪selected countries. The goal is to visualize trends, calculate key metrics
       ↪(e.g., death rate), and summarize findings. We use two datasets:
      - **Cases Dataset**: Contains `country`, `date`, `total_cases`, `total_deaths`,
       ↪and `new_cases`.
      - **Vaccinations Dataset**: Contains `country`, `date`, `total_vaccinations`,
       ↪`people_fully_vaccinated`, and `population`.

      ### Tasks
      - Plot total cases, deaths, new cases, cumulative vaccinations, and percentage
       ↪vaccinated over time.
      - Compare vaccination rates across countries.
      - Calculate death rate (`total_deaths / total_cases`).
      - Create bar charts for top countries by cases and pie charts for vaccination
       ↪status.
      - Summarize 3-5 key insights and highlight anomalies.

      ### Tools
      - `pandas` for data manipulation.
      - `matplotlib` for visualizations.

      ## Data Preparation Notes
      - The datasets are checked for required columns to avoid `KeyError` issues (e.g.
       ↪, `'location'` was likely `'country'`).
      - Missing values in `'country'`, `'date'`, `'total_cases'`, `'total_deaths'`,
       ↪`'total_vaccinations'`, and `'people_fully_vaccinated'` are dropped.
      - `'new_cases'` missing values are filled with 0.
      - Dates are converted to `datetime` for proper plotting.

      ## Summary of Findings

      ### Key Insights
```

Based on the analysis of the sample COVID-19 cases and vaccination data, here are 3-5 key insights:

1. **United States Led in Cumulative Vaccinations**:
   - The United States showed the highest number of cumulative vaccinations by January 2, 2021, reaching 2,000 total vaccinations, compared to 1,500 in India and 1,800 in Brazil. This suggests a faster initial vaccine rollout in the U.S., possibly due to early access to vaccines or efficient distribution systems.

2. **Brazil Had the Highest Death Rate**:
   - Brazil's death rate (total deaths / total cases) was 10.00% (0.1000) on the latest date, matching the United States but higher than India's 10.00%. This indicates a similar severity of outcomes in Brazil and the U.S., possibly due to healthcare system challenges or demographic factors.

3. **India Showed Rapid Case Growth**:
   - India's total cases tripled from 50 to 150 in one day, a faster relative increase compared to the United States (100 to 200) and Brazil (80 to 180). This suggests India faced a significant early surge, potentially driven by population density or testing capacity.

4. **Low Vaccination Coverage Across All Countries**:
   - The percentage of the population fully vaccinated was extremely low (e.g., 0.0003% in the U.S., 0.00004% in India, 0.0004% in Brazil by January 2, 2021). This reflects the early stage of global vaccine rollouts, with pie charts showing nearly 100% unvaccinated populations.

5. **Consistent Daily New Cases**:
   - Daily new cases remained relatively stable (e.g., 100 in the U.S., 100 in India, 100 in Brazil on the second day), indicating consistent infection rates across countries in this sample data. This stability may not reflect real-world trends, where fluctuations are common.

Anomalies and Interesting Patterns
- **Uniform Death Rates**: All three countries (United States, India, Brazil) had identical death rates of 10.00% in the sample data, which is unusual. Real-world data typically shows variation due to differences in healthcare, testing, or demographics. This may be an artifact of the sample data.
- **Rapid Vaccination Increase**: All countries doubled or nearly tripled their total vaccinations in one day (e.g., U.S. from 1,000 to 2,000, India from 500 to 1,500). This steep increase is likely exaggerated in the sample data and may not reflect realistic rollout speeds.
- **Low Vaccination Percentages**: The pie charts show vaccination coverage below 0.001% for all countries, which is expected early in 2021 but highlights the slow initial progress of global vaccination efforts.

- **Stable New Cases**: The lack of variation in daily new cases (e.g., 100 for most countries) is atypical, as real-world data often shows spikes or declines driven by policy changes or outbreaks.

Notes
- The insights are based on sample data with limited dates and countries. Real data would likely show more variability and nuanced trends.
- The `KeyError` for `'location'` was resolved by using `'country'` as the column name, confirmed by checking `df.columns`.
- Missing values were handled by dropping rows with missing `'country'`, `'date'`, or other critical columns, ensuring robust visualizations.

Conclusion
This analysis provides a snapshot of COVID-19 cases and vaccination progress for the United States, India, and Brazil. The United States led in vaccination numbers, while Brazil and the U.S. had high death rates. India showed rapid case growth, and all countries had low vaccination coverage, reflecting early 2021 conditions. Anomalies like uniform death rates and stable new cases suggest limitations in the sample data.

Next Steps
- **Expand Data**: Include more countries and a longer time period for deeper insights.
- **Incorporate Real Data**: Replace sample data with datasets from sources like Our World in Data or WHO.
- **Advanced Analysis**: Explore correlations between vaccination rates and case declines, or compare death rates by age group.
- **Interactive Visualizations**: Use tools like `plotly` for interactive plots to enhance exploration.

This notebook combines code, visualizations, and narrative to meet the project goals. Adjust the data source and `selected_countries` to reflect your actual dataset for tailored insights.