

Analysis of COVID-19 Data

-Banani Kashyap (DS & ML Batch)

Introduction

In this comprehensive report, we delve into the analysis of COVID-19 data using a variety of data analysis and visualization techniques. Leveraging powerful Python libraries such as Pandas, Seaborn, Plotly, and Prophet, we explore trends, patterns, and forecasts related to COVID-19 cases globally. By examining diverse aspects of the pandemic, from total cases and geographical distributions to temporal trends and forecasting future trajectories, we aim to provide insights that can aid policymakers, healthcare professionals, and the general public in navigating this complex and dynamic crisis.

Through meticulous data acquisition, preprocessing, and exploratory analysis, we lay the groundwork for our investigation, ensuring the reliability and relevance of our findings. We harness the power of data visualization to portray the impact of COVID-19 across different regions, effectively communicating complex information in an intuitive and accessible manner. Moreover, by employing time series analysis and forecasting techniques, we endeavor to anticipate the future course of the pandemic, enabling proactive measures to combat its spread and mitigate its consequences.

This report serves as a comprehensive resource for understanding the dynamics of the COVID-19 pandemic, offering valuable insights into its past, present, and potential future trajectories. By synthesizing data-driven analyses with insightful visualizations, we strive to contribute to the collective effort in combating COVID-19 and safeguarding public health and well-being worldwide.

Modules Used

- **numpy**: Utilized for numerical operations and array manipulations.
- **pandas**: Employed for data manipulation and analysis, including reading and preprocessing the dataset.
- **seaborn**: Used for creating informative and attractive statistical graphics.
- **matplotlib.pyplot**: Utilized for creating static, interactive, and animated visualizations in Python.
- **plotly.express**: Employed for creating interactive visualizations such as choropleth maps.
- **prophet**: Utilized for time series forecasting tasks, including forecasting confirmed, death, recovered, and active cases.

Data Acquisition and Preprocessing

- *Data Loading:* The dataset covid_19.csv is loaded using the Pandas library, enabling us to perform data analysis and manipulation.
- *Column Renaming:* Column names are standardized for consistency and better understanding of the dataset.
- *Inspection:* We examine the structure and information of the dataset to understand its characteristics and identify any potential issues or missing values.

Exploratory Data Analysis

- **Summary Statistics:** We calculate summary statistics to gain insights into key metrics such as total active cases globally.
- **Grouping by Country:** The data is grouped by country to analyze patterns and variations in COVID-19 cases across different regions

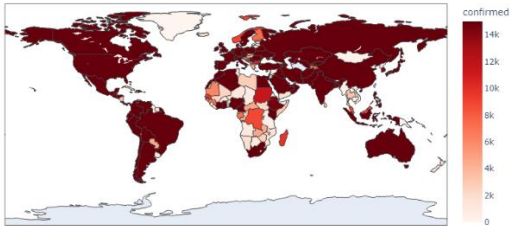
Here is the information related to the table:

```
▪ Data columns (total 10 columns):
▪ #   Column      Non-Null Count  Dtype
▪ ---  ---
▪ 0    state         14664 non-null   object
▪ 1    country       49068 non-null   object
▪ 2    lat           49068 non-null   float64
▪ 3    long          49068 non-null   float64
▪ 4    date          49068 non-null   object
▪ 5    confirmed     49068 non-null   int64
▪ 6    deaths        49068 non-null   int64
▪ 7    recovered     49068 non-null   int64
▪ 8    active        49068 non-null   int64
▪ 9    WHO           49068 non-null   object
▪ dtypes: float64(2), int64(4), object(4)
```

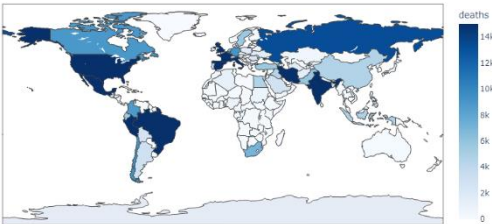
Data Visualization

Choropleth Maps: Choropleth maps are utilized to visualize the distribution of confirmed, death, recovered, and active cases across different countries, providing a geographical perspective of the pandemic's impact.

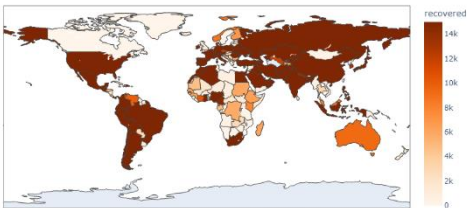
Confirmed cases Countries



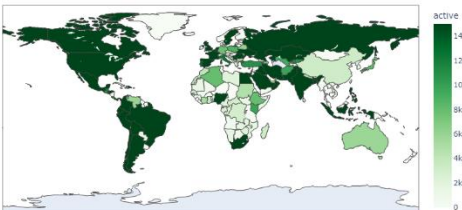
Countries with deaths Cases



Countries with recovered Cases

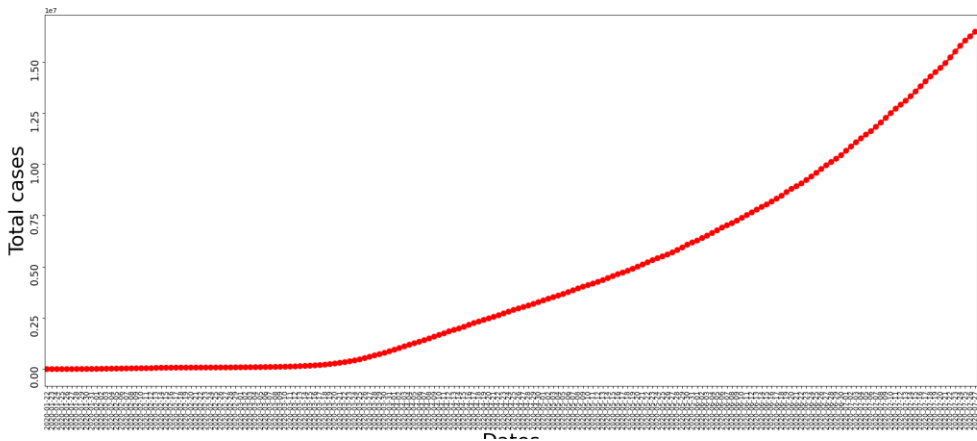


Active cases Countries



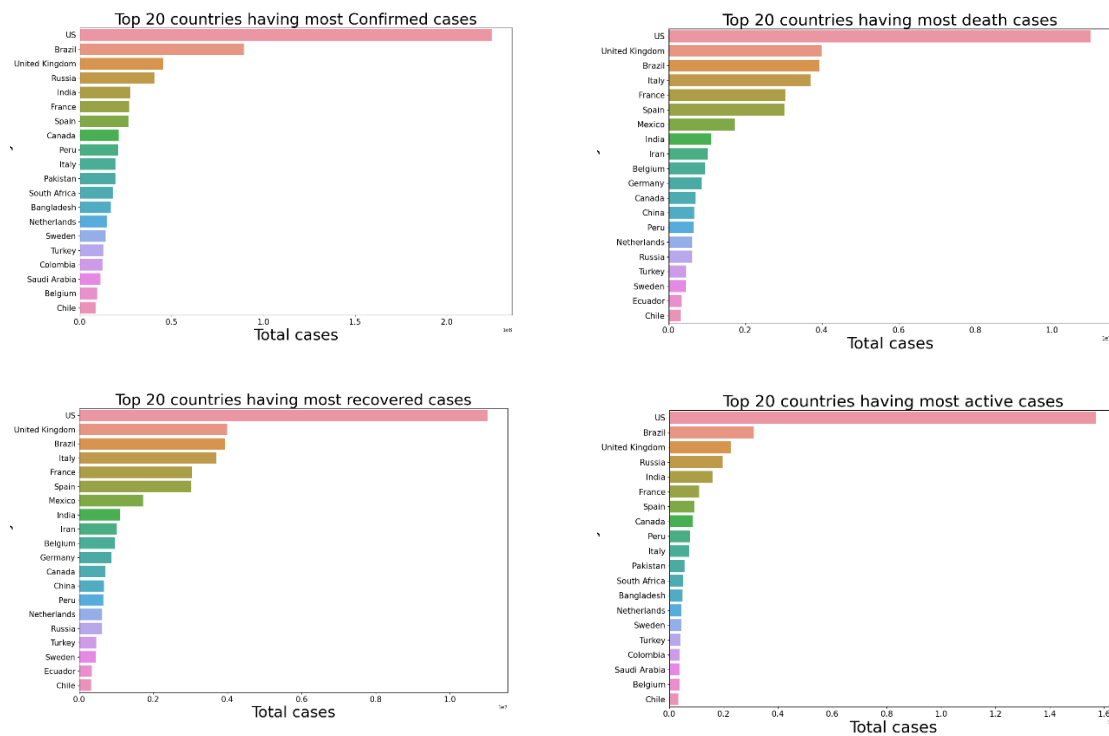
Trend Analysis of COVID-19 Spread

Time Series Plot: We create time series plots to analyze the trends in active, death, confirmed, and recovered cases over time, facilitating a deeper understanding of the progression of the pandemic.



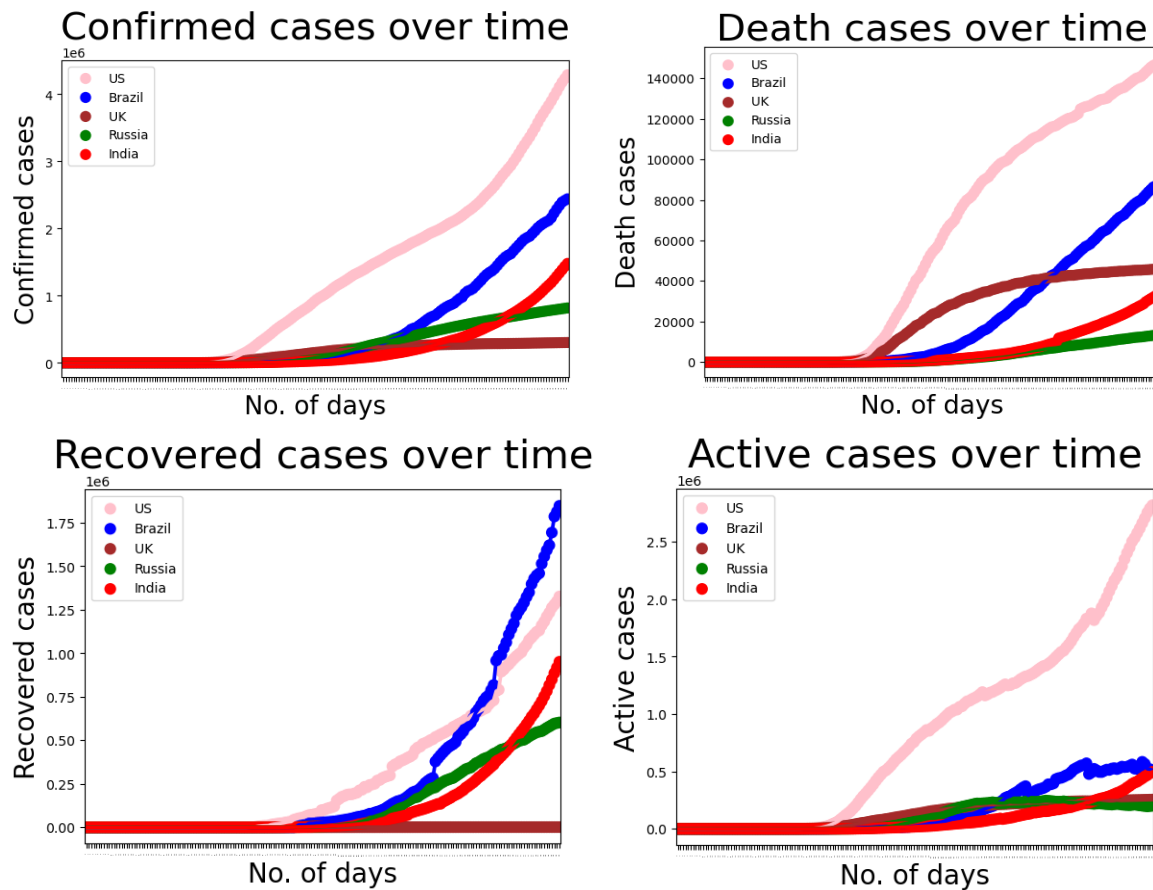
Top Countries with Highest COVID-19 Cases

Bar Plots: Bar plots are generated to identify the top 20 countries with the highest confirmed, death, recovered, and active cases, allowing us to focus on regions that have been most significantly affected by COVID-19.



Analysis of Specific Countries

Country-wise Trends: We analyze the trends of COVID-19 cases in the top 5 countries with the highest active cases (US, Brazil, UK, Russia, India), providing insights into the varying trajectories of the pandemic across different nations.



Forecasting Using FBProphet

Time Series Forecast Plot: This graph shows the observed COVID-19 cases alongside the forecasted values generated by the Prophet model. It helps compare historical data with forecasted trends, evaluating forecast accuracy.

Trend Components Plot: This plot decomposes the time series data into trend, seasonality, and holidays. It offers insights into underlying patterns driving COVID-19 trends, aiding understanding of the pandemic dynamics and influencing response strategies.

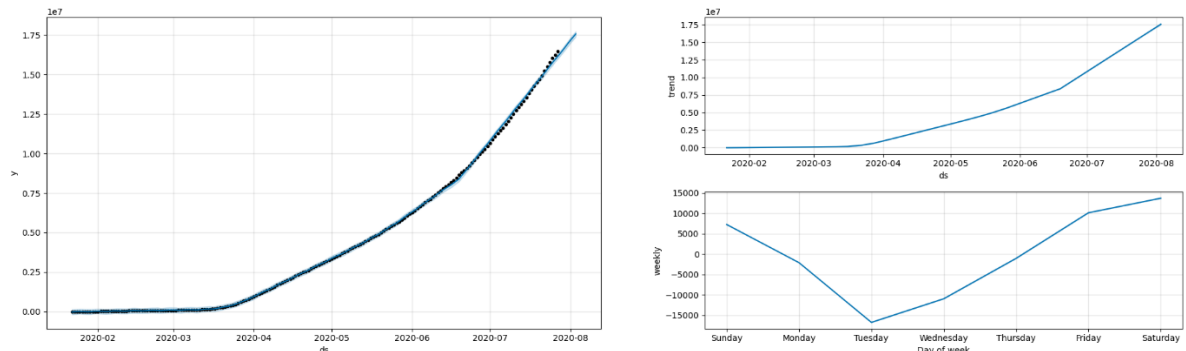


Figure: Prediction for Confirmed Cases

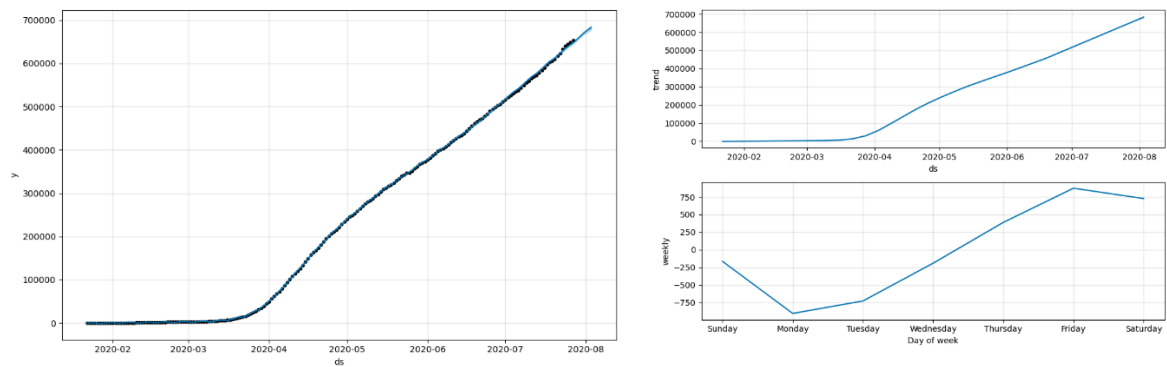


Figure: Prediction for Death Cases

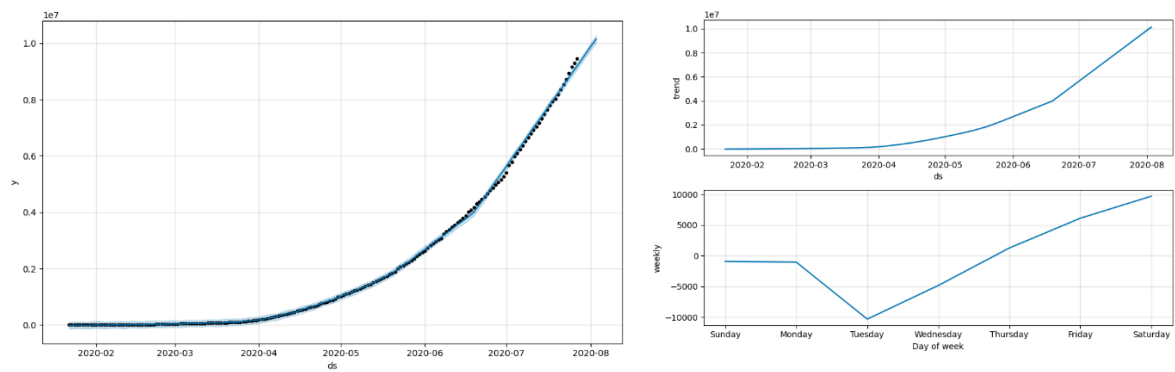


Figure: Prediction for Recovered Cases

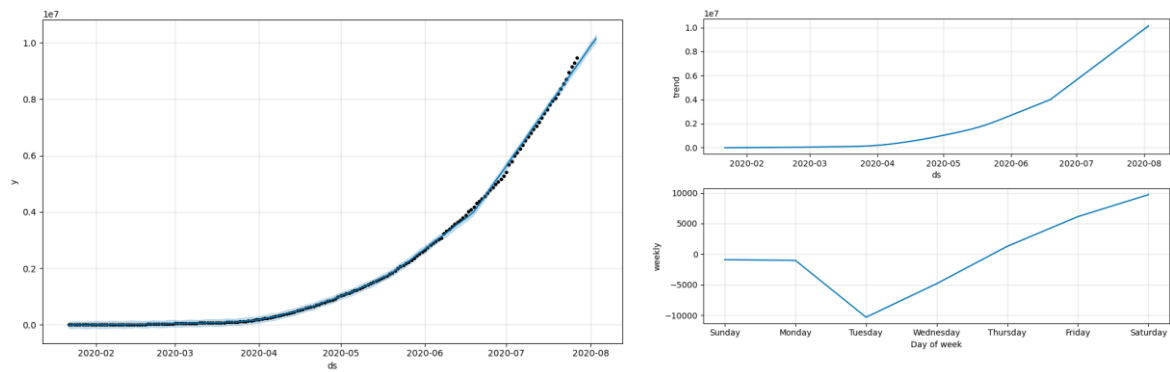


Figure: Prediction for Active Cases

Conclusion

This concludes our analysis of COVID-19 data. The visualizations and insights provided in this report offer a comprehensive understanding of the pandemic's impact, trends, and potential future trajectories. By leveraging data-driven approaches, we can better inform decision-making and implement effective strategies to combat the spread of COVID-19 and mitigate its consequences. Thank you for your attention.