Analyzing the Trends of COVID-19

Sekhar Chandra Padhi

Index :-

Introduction:-

Problem Statement:

Given data about COVID-19 patients, write code to visualize the impact and analyze the trend of rate of infection and recovery as well as make predictions about the number of cases expected a week in future based on the current trends.

Guidelines:

- Use pandas to accumulate data from multiple data files.
- Use plotly (visualization library) to create interactive visualizations.
- Use Facebook prophet library to make time series models.
- Visualize the prediction by combining these technologies.

Data Information :-

Dataset:

CSV and Excel files containing data about the number of COVID-19 confirmed deaths and recovered patients both around the world and in India.

Source:

Intellipaat.

Description:

The dataset likely contains information about confirmed cases, deaths, and recoveries from COVID-19 for different countries/regions and dates.

Rows/Index (49068 entries, 0 to 49067)
Data columns (total 10 columns)

Key Variables:

- Date: The date of the record.
- Country/Region: The country or region where the data was collected.
- Province/State: The province or state within the country/region (if applicable).
- Confirmed: The number of confirmed COVID-19 cases.
- Deaths: The number of deaths due to COVID-19.
- Recovered: The number of recovered COVID-19 cases.

Data Loading and Cleaning :-

The project uses the covid_19_clean_complete.csv dataset. We load it using the pandas library and perform the following cleaning steps:

Data Type Conversion:

Convert the 'Date' column to datetime format for proper analysis.

Column Renaming:

Rename columns 'Province/State' and 'Country/Region' to 'state' and 'country' for better readability.

Handling Missing Values:

The 'state' column contains null values. However, removing these null values could result in data loss since some data collection is based on country-wise information. Given that our project also focuses on country-wise analysis, we will skip removing the null values here.

Duplicate Check:

Checked duplicate entries in the dataset.

Exploratory Data Analysis (EDA) :-

I conduct EDA to gain insights into the data

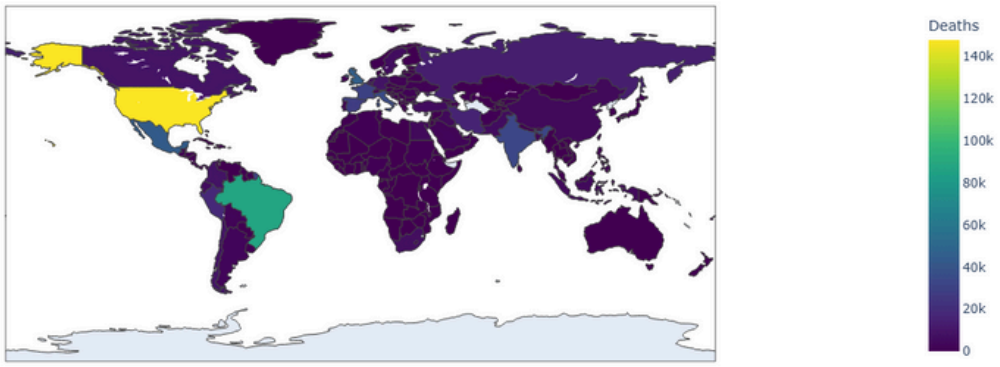
- Analyze confirmed, active, recovered, and death cases globally.
- Identify countries with the highest and lowest numbers of cases.
- Visualize the distribution of cases using choropleth maps to see the global impact.

Global Spread and Impact :-

I visualize the global spread of COVID-19 using choropleth maps, showing the distribution of:

- Deaths
- Active cases
- Recovered cases
- Confirmed cases These visualizations help understand the geographical impact of the pandemic.

country wise death cases



country wise active cases

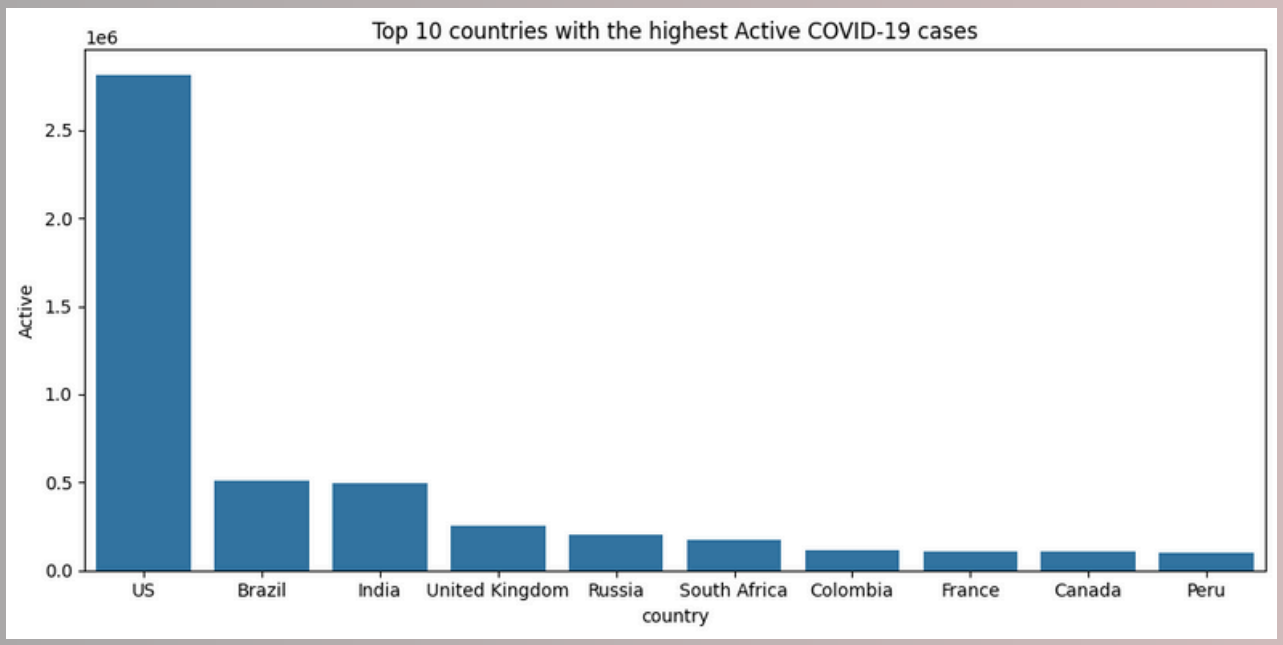
country wise recovered cases

country wise recovered cases

Country-wise Analysis :-

I analyze the data on a country-by-country basis, focusing on:

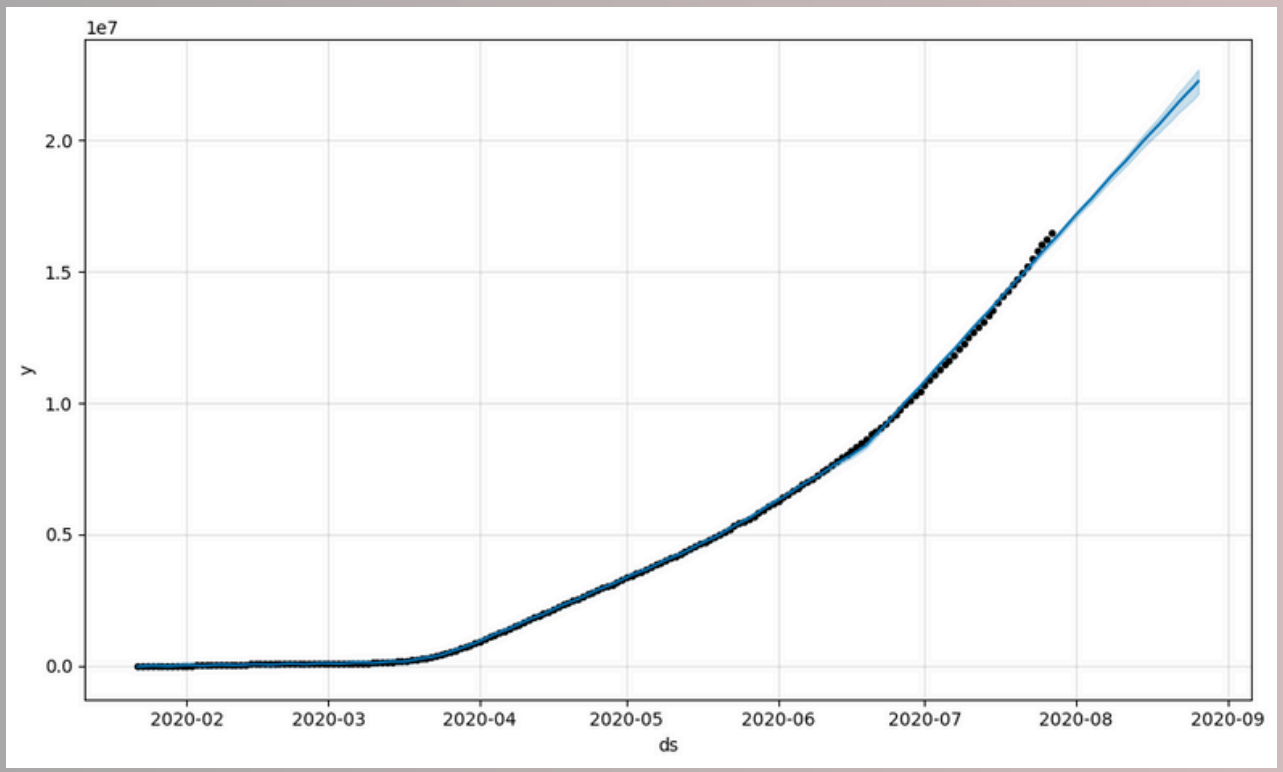
- Top 10 countries with the highest active cases. This analysis reveals the countries most affected by the pandemic. You can potentially expand this to include other metrics or comparisons as per your modifications.

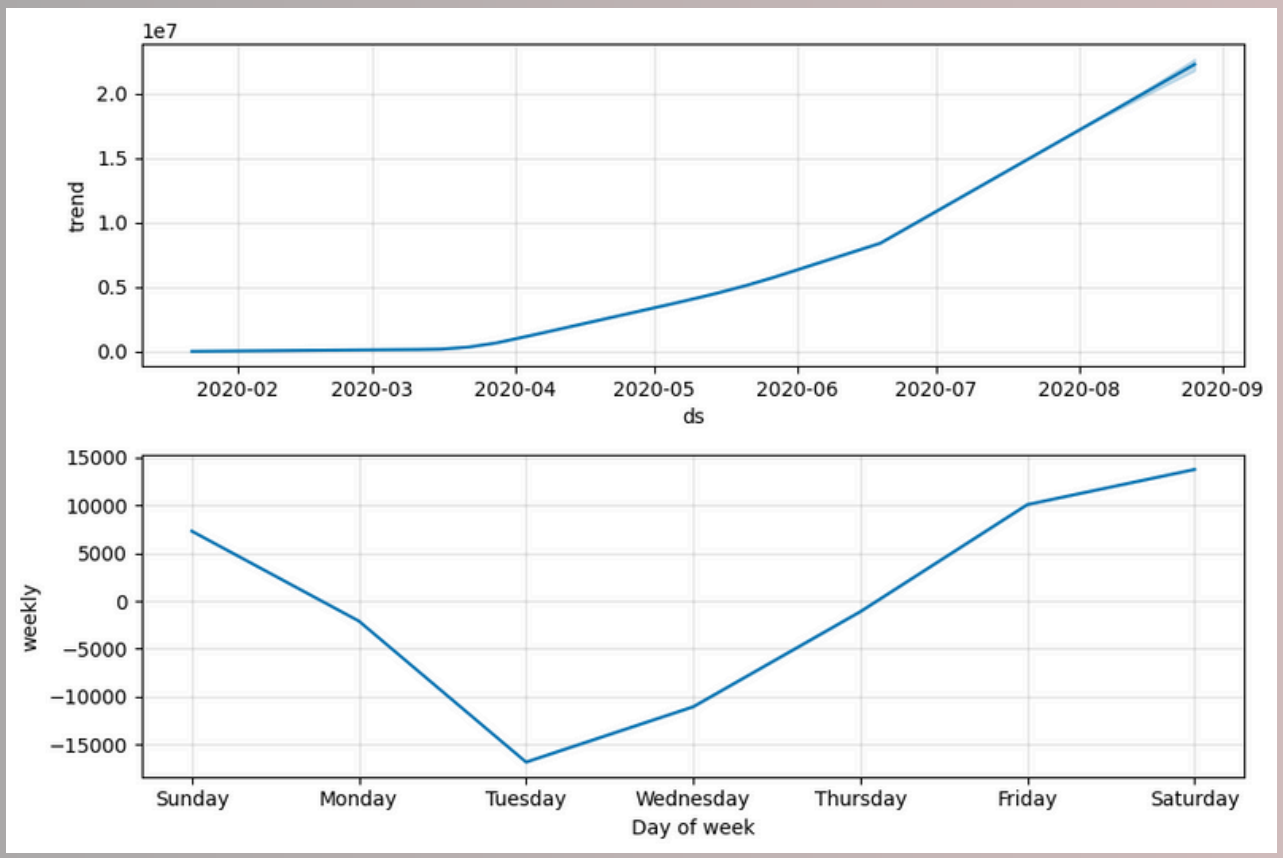


Time Series Analysis and Forecasting with Prophet :-

I use the Facebook Prophet library to perform time series analysis and forecasting:

- Analyze the trend of confirmed cases over time.
- Build a Prophet model to predict future confirmed cases.
- Visualize the forecast with uncertainty intervals.
- Analyze the components of the forecast, such as trend and seasonality.





Conclusion :-

This project provides a comprehensive analysis of COVID-19 trends, highlighting the global impact and country-specific patterns. Data cleaning and preprocessing steps were implemented to enhance the accuracy of the analysis. The use of the Prophet library allows for forecasting future trends, which can be valuable for planning and resource allocation. By addressing the null values in the 'state/province' column, the analysis provides a more complete and reliable picture of the pandemic's impact.