

COVID-19
Data Analysis &
Forecasting Report

Shreya Manohar

Introduction

The outbreak of COVID-19 caused widespread health, economic, and social challenges globally. Understanding the spread of the virus through data-driven analysis is essential for planning, decision-making, and forecasting future trends.

This project analyzes worldwide and India-specific COVID-19 data to identify infection and recovery trends, visualize patterns, and forecast the expected number of cases for the upcoming week.

Problem Statement

Given data about COVID-19 patients, the objective is to:

- Visualize the impact of COVID-19 across the world and in India.
- Analyze the trends in infection, recovery, and fatality rates.
- Predict the number of cases expected in the next week using time-series forecasting techniques.

Dataset Description

The dataset includes:

- **Confirmed cases**
- **Deaths**
- **Recovered cases**
- **Date-wise global records**
- **Country/Region data, specifically India**
- CSV/Excel files containing historical COVID-19 data

Tools and Libraries Used

Core Libraries

- **Pandas** – Data loading, cleaning, preprocessing
- **NumPy** – Numerical processing
- **Plotly** – Interactive visualizations
- **Facebook Prophet** – Time-series forecasting model

Methodology

Exploratory Data Analysis (EDA)

Global Trends

- Visualized cumulative confirmed, recovered, and death cases over time.
- Identified rising infection peaks.
- Observed improvement in recovery rates as time progressed.

New Trends

- Similar visualizations showed India's infection curve, recovery improvements, and stabilization phases.

Daily New Cases

- Bar charts combined with rolling averages highlighted short-term fluctuations and long-term trends.

Rates (Recovery & Fatality)

- Observed how recovery rate gradually increased.
- Fatality rate varied but generally stabilized over time
- Loaded datasets using Pandas.
- Converted the `Date` column into datetime format.
- Ensured numerical fields such as Confirmed, Deaths, and Recovered are properly typed.
- Aggregated data:
 - **Global daily data**
 - **India-specific daily data**
- Computed daily new cases using `.diff()`.
- Created additional metrics:
 - **7-day rolling averages**
 - **Case Fatality Rate (CFR)**
 - **Recovery Rate**

Visualizations (Plotly)

Interactive graphs created:

Cumulative Trends

- Line plots for Confirmed, Recovered, and Deaths.
- Separate visuals for Global and India data.

Daily New Cases

- Bar chart of new cases.
- Overlaid 7-day moving average line.

Recovery vs Fatality Rates

- Trend lines showing:
 - % of infected population recovering
 - % of infected population dying (CFR)

Data Preparation

Prophet requires:

- $ds \rightarrow$ date
- $y \rightarrow$ value to predict (confirmed cases)

Both global and India datasets were transformed accordingly.

Modeling

- Fitted Prophet models on confirmed cases.
- Generated future dataframes predicting **next 7 days**.
- Output included:
 - $yhat$ (predicted value)
 - $yhat_lower$
 - $yhat_upper$

Forecast Visualization

Prophet plots included:

- Trend components
- Seasonal patterns (if detected)
- Actual vs Predicted curves
- Confidence intervals (shaded region)

Step 1: Data Loading and Preprocessing (Fixed Date Parsing)

This section includes code and explanation for: **1: Data Loading and Preprocessing (Fixed Date Parsing)**.

Output:-

Global daily data:

	Confirmed	Deaths	Recovered
Date			
2020-01-22	555	17	28
2020-01-23	654	18	30
2020-01-24	941	26	36
2020-01-25	1434	42	39
2020-01-26	2118	56	52

India daily data:

	Confirmed	Deaths	Recovered
Date			
2020-01-22	0	0	0
2020-01-23	0	0	0
2020-01-24	0	0	0
2020-01-25	0	0	0
2020-01-26	0	0	0

```
/tmp/ipython-input-3140941454.py:9: UserWarning: Parsing dates in %Y-%m-%d
format when dayfirst=True was specified. Pass `dayfirst=False` or specify a
format to silence this warning.
```

```
df['Date'] = pd.to_datetime(df['Date'], dayfirst=True, errors='coerce')
```

Step 2: Daily new cases, rolling averages, and rates

This section includes code and explanation for: **2: Daily new cases, rolling averages, and rates**.

Global daily data (with new columns):

	Confirmed	Deaths	Recovered	NewConfirmed	NewRecovered	\
Date						
2020-07-23	15510481	633506	8710969	282756.0	169714.0	
2020-07-24	15791645	639650	8939705	281164.0	228736.0	
2020-07-25	16047190	644517	9158743	255545.0	219038.0	
2020-07-26	16251796	648621	9293464	204606.0	134721.0	
2020-07-27	16480485	654036	9468087	228689.0	174623.0	

	NewDeaths	NewConfirmed_7d	NewRecovered_7d	CFR	\
Date					
2020-07-23	9966.0	242565.142857	153818.285714	0.040844	
2020-07-24	6144.0	248154.571429	163706.428571	0.040506	
2020-07-25	4867.0	250713.142857	173456.142857	0.040164	
2020-07-26	4104.0	249278.714286	180175.571429	0.039911	
2020-07-27	5415.0	252408.857143	182472.857143	0.039685	

	RecoveryRate
Date	
2020-07-23	0.561618
2020-07-24	0.566103
2020-07-25	0.570738
2020-07-26	0.571842
2020-07-27	0.574503

India daily data (with new columns):

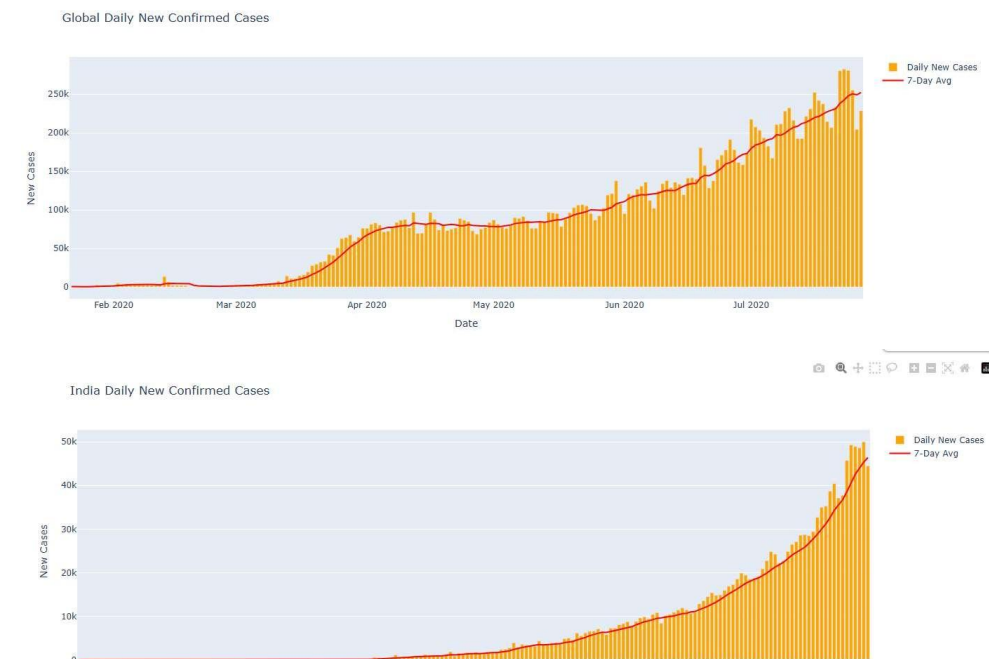
Date	Confirmed	Deaths	Recovered	NewConfirmed	NewRecovered	\
2020-07-23	1288108	30601	817209	49310.0	34602.0	
2020-07-24	1337024	31358	849432	48916.0	32223.0	
2020-07-25	1385635	32060	885573	48611.0	36141.0	
2020-07-26	1435616	32771	917568	49981.0	31995.0	
2020-07-27	1480073	33408	951166	44457.0	33598.0	

Date	NewDeaths	NewConfirmed_7d	NewRecovered_7d	CFR	\
2020-07-23	740.0	40610.857143	25921.714286	0.023757	
2020-07-24	757.0	42562.857143	27954.428571	0.023454	
2020-07-25	702.0	43979.142857	29735.714286	0.023137	
2020-07-26	711.0	45344.285714	31068.714286	0.022827	
2020-07-27	637.0	46390.714286	32369.714286	0.022572	

Date	RecoveryRate
2020-07-23	0.634426
2020-07-24	0.635315
2020-07-25	0.639110
2020-07-26	0.639146
2020-07-27	0.642648

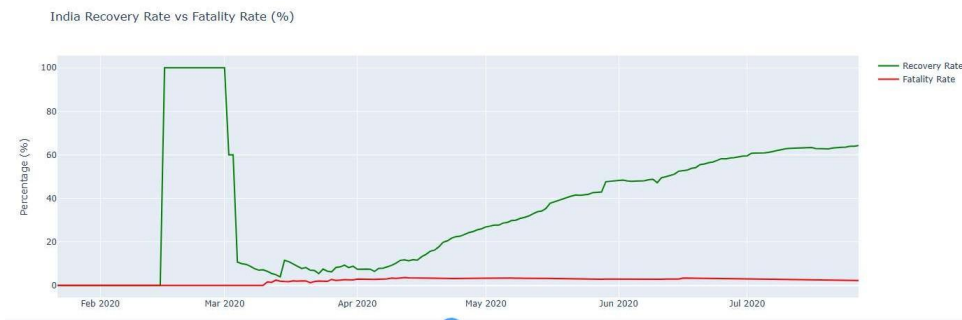
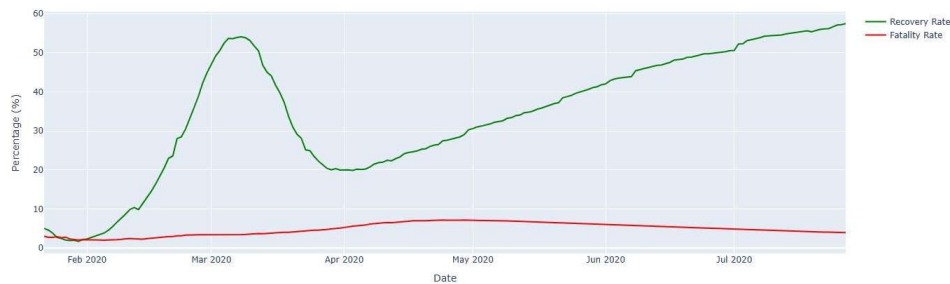
Step 3 (Plotly Version): Visualization of cumulative and daily trends

This section includes code and explanation for: **3 (Plotly Version): Visualization of cumulative and daily trends.**



Step 4 (Plotly Version): Infection vs Recovery Trend Analysis

This section includes code and explanation for: **4 (Plotly Version): Infection vs Recovery Trend Analysis.**



Step 5: Time Series Forecasting using Facebook Prophet

This section includes code and explanation for: **5: Time Series Forecasting using Facebook Prophet.**

Step 6: Visualize Prophet Predictions with Plotly (Combined with Historical Data)

This section includes code and explanation for: **6: Visualize Prophet Predictions with Plotly (Combined with Historical Data).**

