COVID-19 Pandemic Analysis Report

Introduction

This project investigates the spread of the COVID-19 pandemic by analyzing data on cases, deaths, and recoveries across different regions and countries. The primary objectives are to understand the progression of the pandemic, visualize trends, and predict future outcomes using machine learning techniques.

Project Goals

1. **Understand Spread**: Examine how COVID-19 propagated over time globally and regionally.
2. **Data Analysis**: Assess cumulative data for confirmed cases, deaths, and recoveries.
3. **Data Visualization**: Highlight patterns in the distribution of cases, recoveries, and deaths.
4. **Prediction**: Employ machine-learning models to forecast future trends, including potential increases in cases and deaths.

Data Overview

The dataset includes the following features:

- **Lat (Latitude)**: Geographical location (north/south).
- **Long (Longitude)**: Geographical location (east/west).
- **Date**: Daily reporting date.
- **Confirmed**: Cumulative confirmed cases.
- **Deaths**: Cumulative deaths.
- **Recovered**: Cumulative recoveries.
- **Active**: Ongoing cases.
- **WHO Region**: Region categorized by the World Health Organization.

Sample Data Structure

1. **Geographic Data**:
    - Examples include Afghanistan, Albania, Algeria, Andorra, and Angola.
2. **Daily Confirmed Cases**:
    - Initial data: 555 confirmed cases on 2020-01-22.
    - Peak period (as of sample): Over 16 million cases by 2020-07-27.
3. **Country-level Snapshot**:
    - **US**: 4,290,259 confirmed cases.
    - **Brazil**: 2,442,375 confirmed cases.
    - **India**: 1,480,073 confirmed cases.

Analysis Approach

1. **Trend Analysis**:
    - Use cumulative data to plot progression over time.
    - Compare regions to identify patterns.
2. **Visualization**:
    - Employ graphs and maps to illustrate global spread.
    - Highlight regions with significant increases or decreases.
3. **Prediction**:
    - Apply machine learning to historical data for forecasting.
    - Evaluate the model's accuracy for predicting cases or deaths.

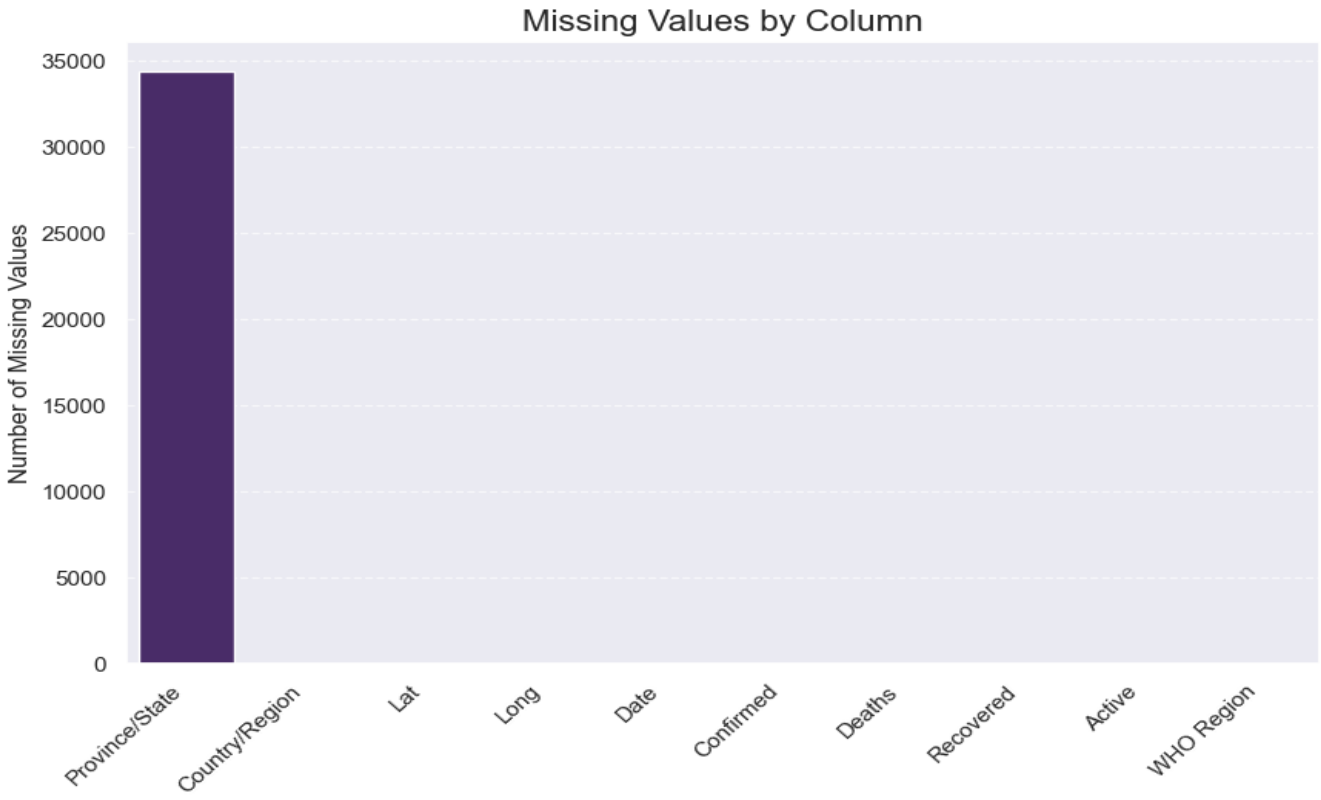
Next Steps

- **Data Cleaning**: Address missing or inconsistent values.
- **Model Training**: Train predictive models using a portion of the data.
- **Validation**: Test and refine models for better accuracy.
- **Interpretation**: Extract actionable insights from results.

This report outlines the groundwork for understanding and combating COVID-19, combining historical analysis with predictive modeling to aid public health decisions.

The first five row of the data

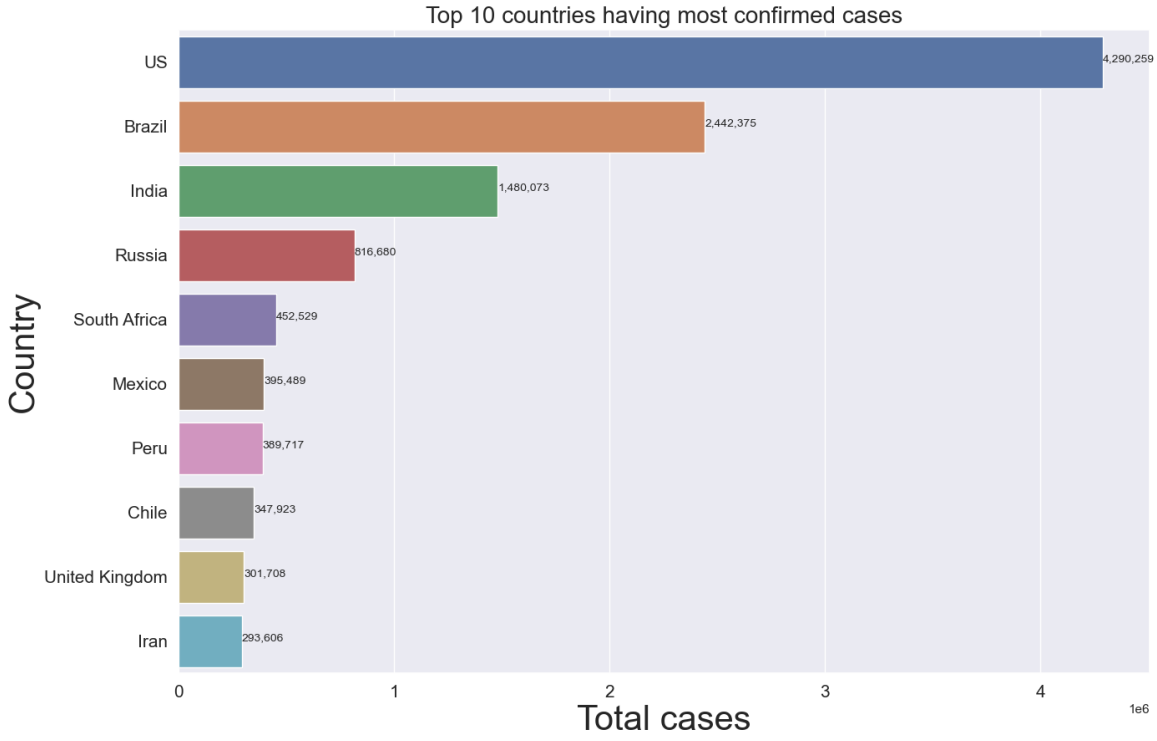
s/n	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths	Recovered	Active	WHO Region
0	NaN	Afghanistan	33.93911	67.709953	2020-01-22	0	0	0	0	Eastern Mediterranean
1	NaN	Albania	41.15330	20.168300	2020-01-22	0	0	0	0	Europe
2	NaN	Algeria	28.03390	1.659600	2020-01-22	0	0	0	0	Africa
3	NaN	Andorra	42.50630	1.521800	2020-01-22	0	0	0	0	Europe
4	NaN	Angola	-11.20270	17.873900	2020-01-22	0	0	0	0	Africa
…	…	…	…	…	…	…	…	…	…	…

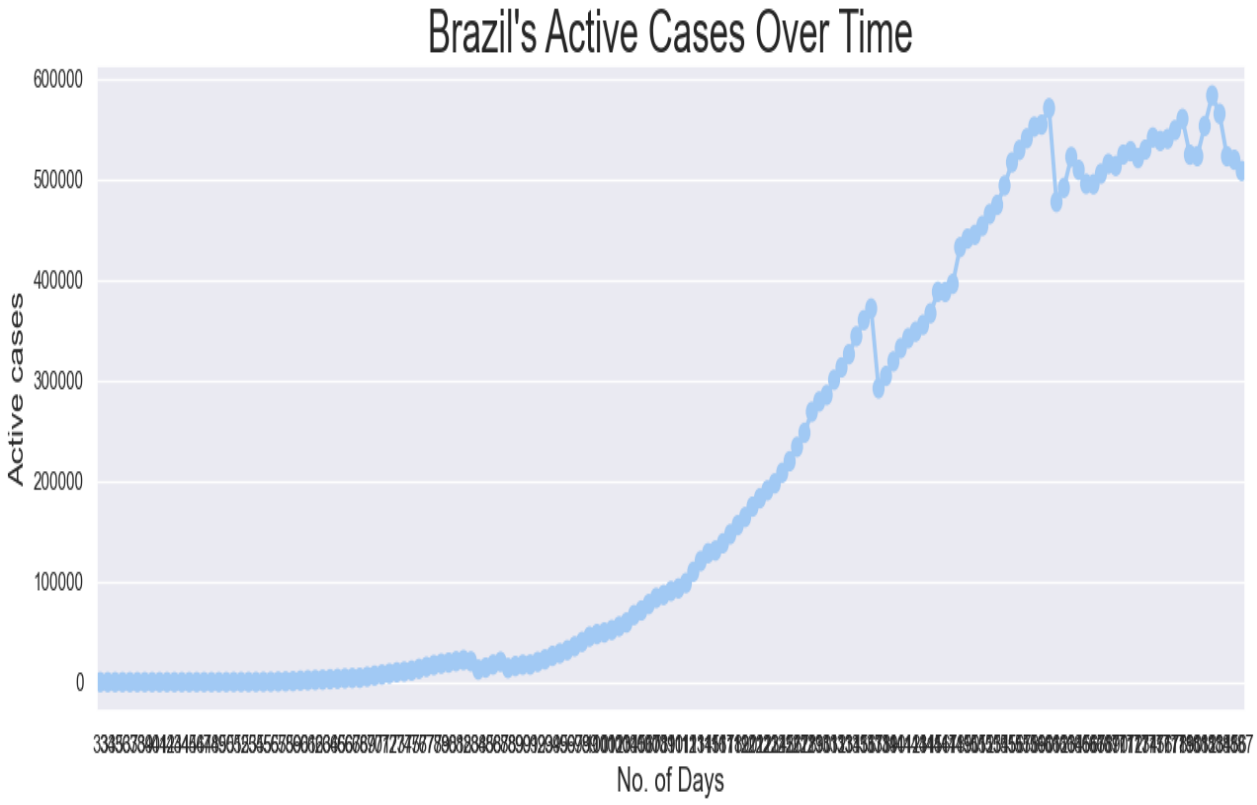


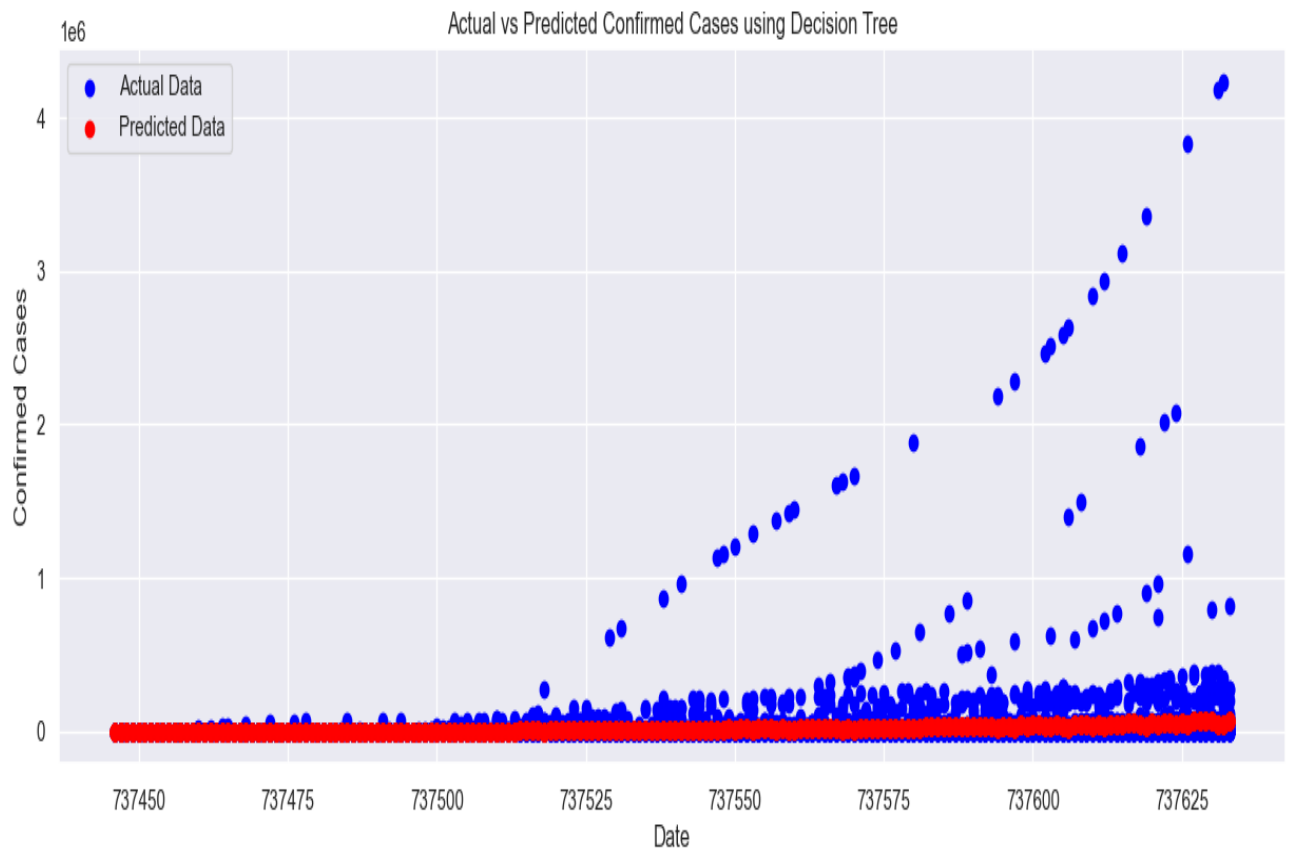
Confirmed Cases

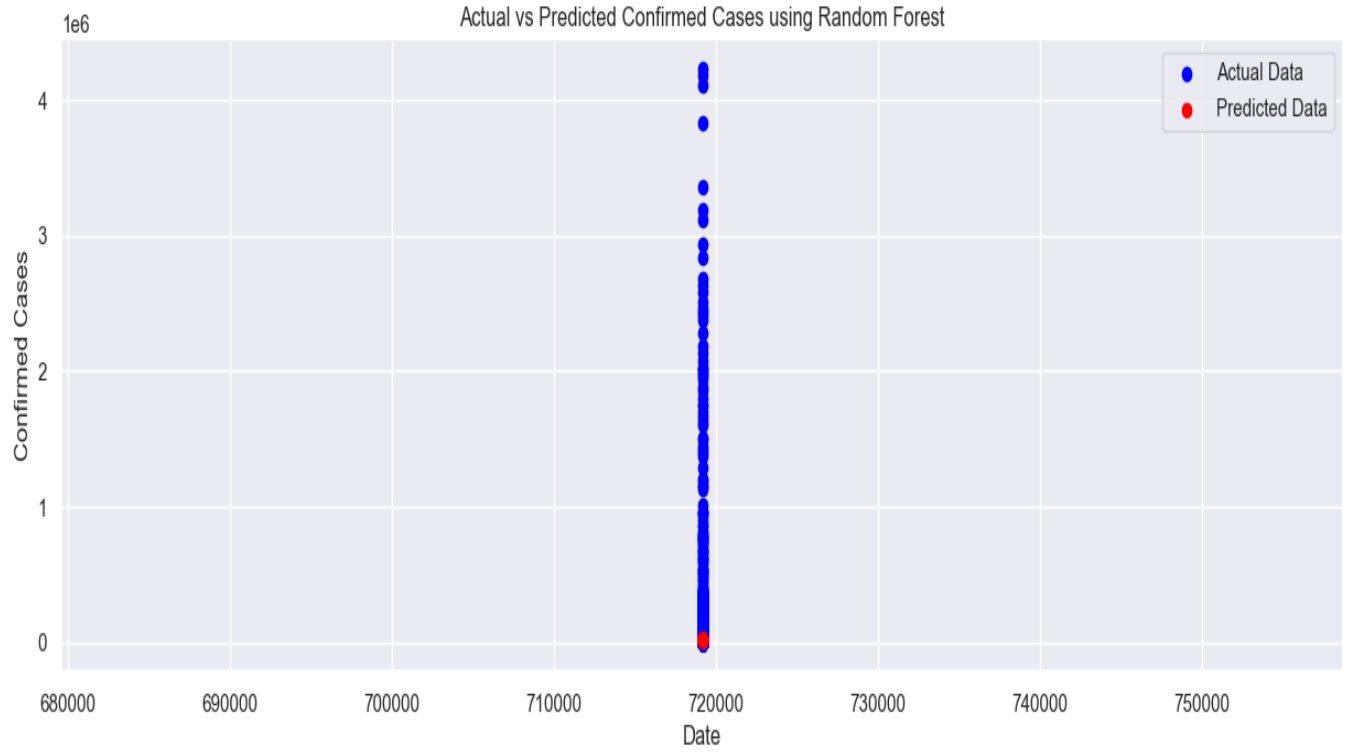
S/N	Date	Confirmed
0	2020-01-22	555
1	2020-01-23	654
2	2020-01-24	941
3	2020-01-25	1434
4	2020-01-26	2118
...	...	...
183	2020-07-23	15510481
184	2020-07-24	15791645
185	2020-07-25	16047190
186	2020-07-26	16251796
187	2020-07-27	16480485

S/N	Country/Region	Confirmed
0	US	4290259
1	Brazil	2442375
2	India	1480073
3	Russia	816680
4	South Africa	452529
5	Mexico	395489
6	Peru	389717
7	Chile	347923
8	United Kingdom	301708
9	Iran	293606


Top 10 countries having most confirmed cases
US
Brazil
India
Russia
South Africa
Mexico
Peru
Chile
United Kingdom
Iran
4,290,259
2,442,375
1,480,073
816,680
452,529
395,489
389,717
347,923
301,708
293,606
Country
0
1
2
3
4
1e6
Total cases


Brazil's Active Cases Over Time
600000
500000
400000
300000
200000
100000
0
Active cases
No. of Days


Actual vs Predicted Confirmed Cases using Decision Tree
1e6
Actual Data
Predicted Data
Confirmed Cases
0
1
2
3
4
737450
737475
737500
737525
737550
737575
737600
737625
Date


Actual vs Predicted Confirmed Cases using Random Forest
1e6
Actual Data
Predicted Data
Confirmed Cases
0
1
2
3
4
680000
690000
700000
710000
720000
730000
740000
750000
Date