Capstone Project – Covid 19

Table of Contents:

1. 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.

2. Project Objective:

Objective of the project is to study the impact of covid 19 in the world. And also predicting the impact in the future week.

3. Data Description:

Feature Name	Description
Date	Week of Sales
Province / State	State or Province
Country / Region	Country or Region Name
Lat	Latitude value
Long	Longitude value
Confirmed	Number of Confirmed cases
Deaths	Number of Deaths
Recovered	Number of Recovered cases
Active	Number of Active cases
WHO Region	WHO Region

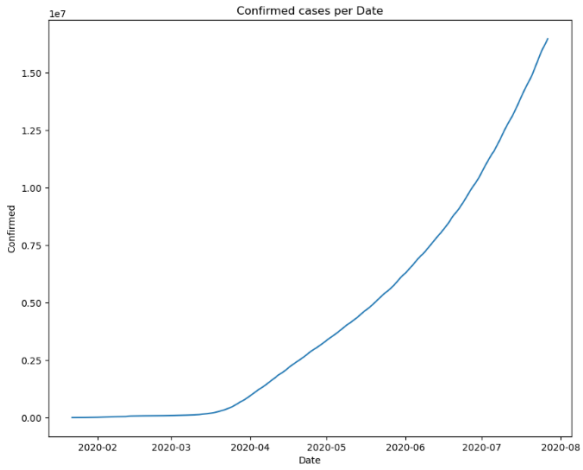
The dataset provided contains 49068 rows of data from different Region/Country. The 'Date 'column contains dates ranging from **"2020-01-22** to **2020-07-27".** Confirmed, Recovered, Active and Deaths columns contain data for number of cases observed on that particular date for the respective Country/Region.

4. Data pre-processing Steps and inspiration:

The preprocessing of the data included the following steps:

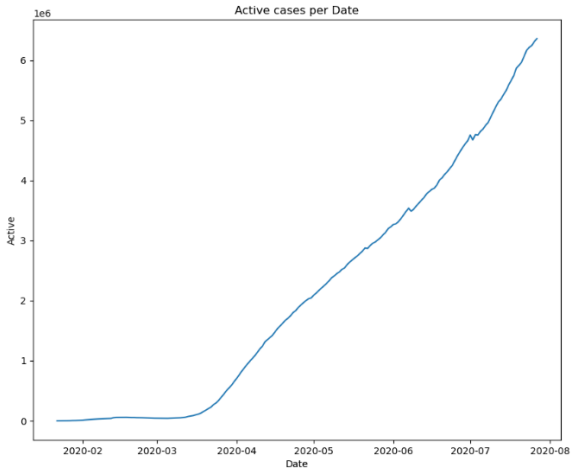
Removing of unwanted columns like "Lat", "Long" as we have Country column to make visualization of the data. Also, "WHO Region", "State" is not required for analysing the data and forecasting the data. For visualization, we will only take into consideration the data present for each Country/Region for the latest date in the dataset i.e. **2020-07-27**.

EDA I:



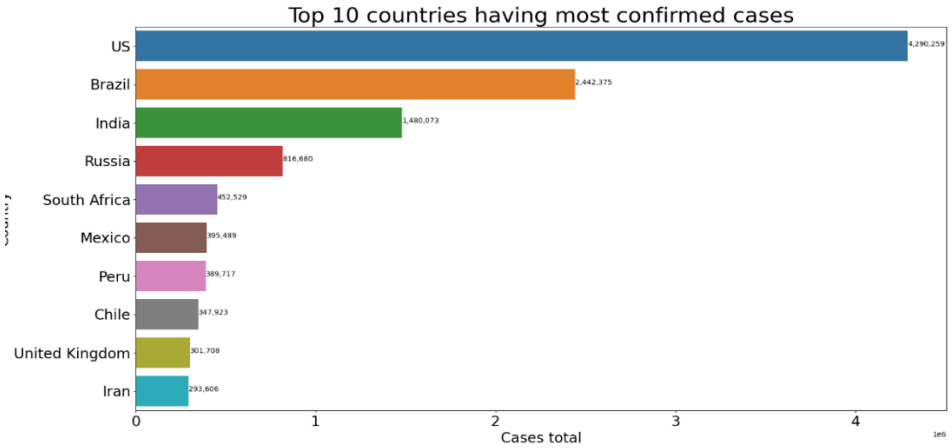
Above visualization shows the number of confirmed cases date wise from Jan-20 to Sep-20. It can be observed that till the mid of month of March there were slight increase in the number of cases but after that the number of confirmed cases started increasing gradually at higher rate.

EDA II:



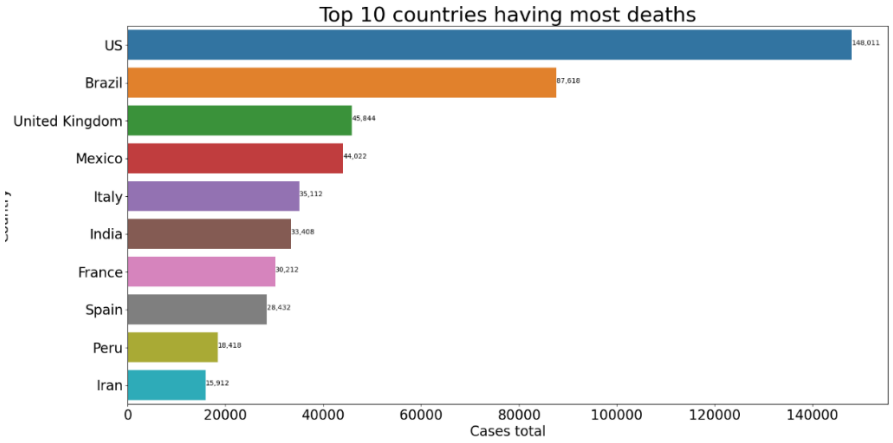
Above visualization shows the number of active cases date wise from Jan-20 to Sep-20. It can be observed that till the mid of month of March there were slight increase in the number of active cases but after that the number of active cases started increasing gradually at higher rate.

EDA III:



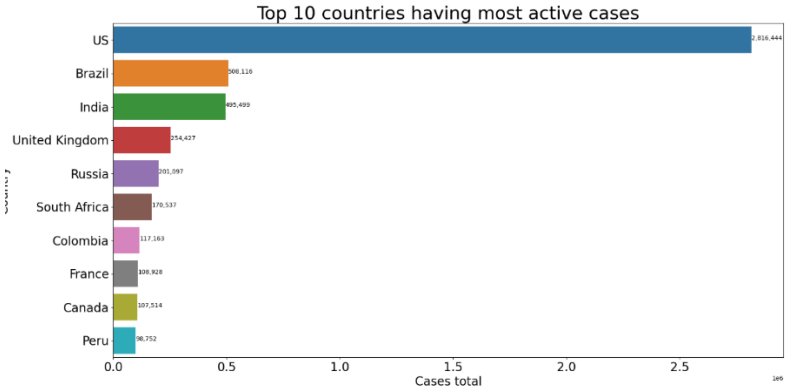
Above graph represents the top 10 countries having the highest number of confirmed cases. It can be observed from above graph that “US” has recorded highest number of confirmed cases. US (42,90,259), Brazil (24,42,375), India (14,80,073) and Russia (8,16,680) are the countries with relatively higher number of cases compared to other countries.

EDA IV:



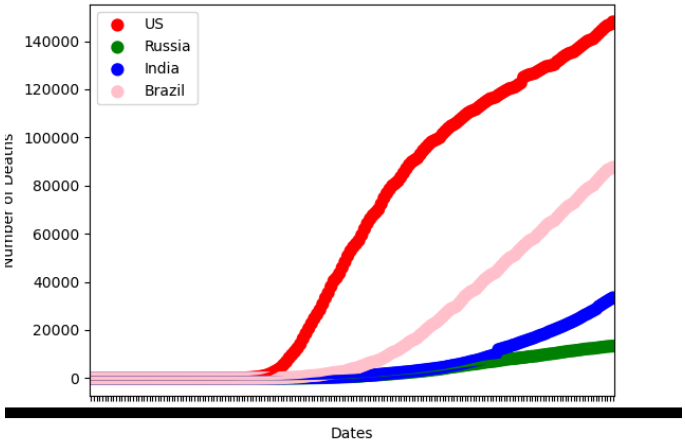
Above graph represents the top 10 countries having the highest number of death cases. It can be observed from above graph that “US” has recorded highest number of death cases of **1,48,011**. Brazil has second highest number of death cases recorded of **87,618**.

EDA V:



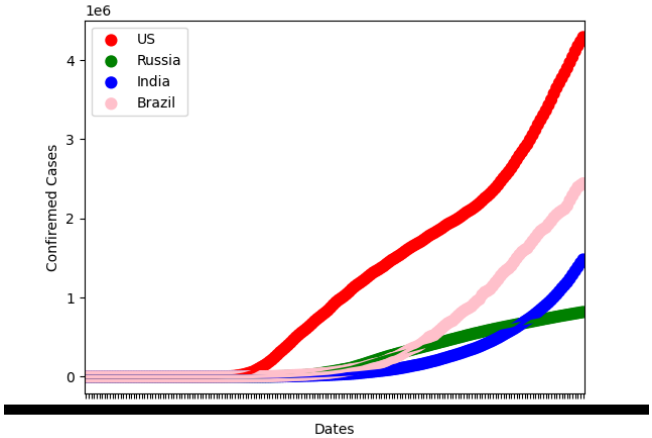
Above graph represents the top 10 countries having the highest number of active cases. It can be observed from above graph that “US” has recorded highest number of active cases compared to any other country. Total of 28,16,444 of active cases are recorded in US. Second highest country with highest number of active cases has only 18% of cases which are recorded in US.

EDA VI:



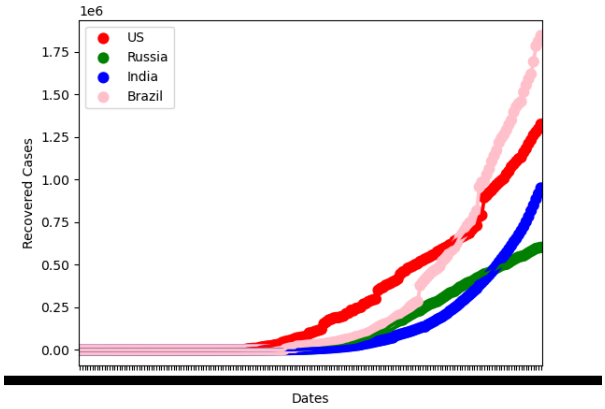
Above visualization shows comparison on number of deaths cases from feb-20 to sep-20 of 4 countries US, Russia, India, Brazil. It is observed that for US the number of deaths cases started increasing early compared to other 3 countries. For Russia There is not much increase in death cases, the increase is at lower rate compared to other 3 countries.

EDA VII:



Above visualization shows comparison on number of confirmed cases from feb-20 to sep-20 of 4 countries US, Russia, India, Brazil. It can be seen that number of confirmed cases increased to greater extent for US. It can also be seen from the visualization that Russia had higher number of confirmed cases than India and Brazil but after a certain time both for India and Brazil there was increase in confirmed cases which went higher than the confirmed cases of Russia.

EDA VIII:



Above visualization shows comparison on number of recovered cases from feb-20 to sep-20 of 4 countries US, Russia, India, Brazil. From above visualization it can be observed that recovery rate of US was higher but after some time the recovery cases of Brazil increased at a very higher rate. Also, it can be seen that at first the recovery cases of Russia were higher than India, but after certain period of time the recovered cases of India gradually increased and the recovery cases of India were higher than that of Russia.

5. Choosing the algorithm for the project:

The main purpose of creating models is to predict the cases for next upcoming week. This prediction was done using the FB Prophet library. In this study we have forecasted the number of confirmed cases, death cases and recovered cases for the next upcoming week.

6. Motivation and Reason for choosing the Algorithm:

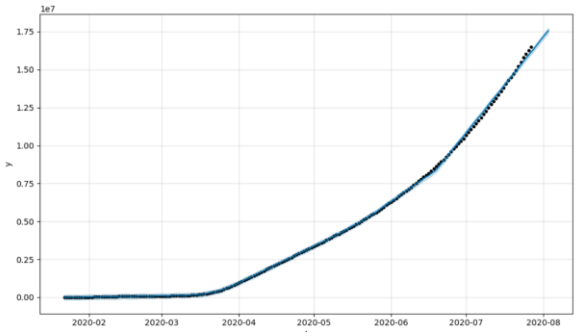
Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality.

The input to Prophet is always a dataframe with two columns: ds and y. The ds (datestamp) column should be of a format expected by pandas. The y column must be numeric, and represents the measurement we wish to forecast.

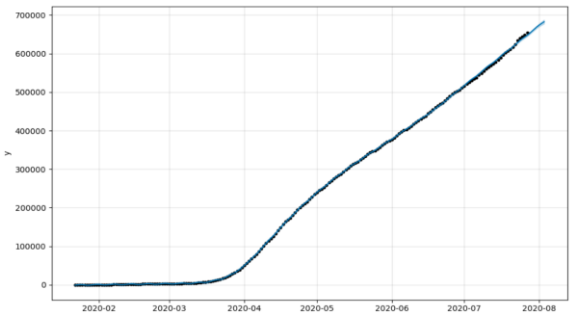
7. Model Evaluation and Techniques:

Below is the observation after forecasting the data for upcoming week for number of confirmed cases, death cases and recovered cases.

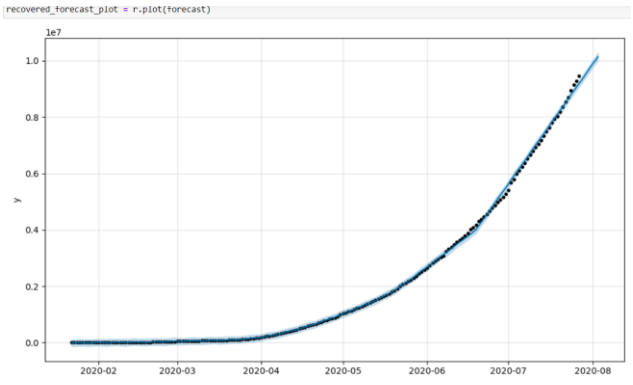
Forecast of Confirmed Cases:



Forecast of Deaths cases:



Forecast of Recovered Cases:



8. Inferences from the same:

From the above forecasted visuals for confirmed, death and recovered cases for a week. It can be seen that there will be rise in number of cases for all the three forecasted parameters (Confirmed, Death, and Recovered cases).

9. Conclusions:

From the visuals it can be seen that the highly affected country is United Sates (US). And the highest recovered cases are that of Brazil compared to other countries, recovery rate is higher compared to others. From the forecasted visual it can be observed that there will be increase in recovered cases in a week which is a good thing but on other hand there is also increased in confirmed cases and death cases which is not a good thing.