**COVID-19 DATA ANALYSIS**

The current situation is extremely overwhelming. The days are getting harder day by day. People are losing their jobs, dying out of covid and their families are getting shattered. Even after taking every little precautions, people are losing their ability to fight this virus.

Everyone is waiting for the day when this nasty killing virus would stop spreading and a vaccine would be provided to every individuals regardless of their race, gender, caste, creed etc.

The main motive is to flatten the curve. The faster the infection curve rises, the quicker the local health care system gets overloaded beyond its capacity to treat people. As we're seeing in Italy, more and more new patients may be forced to go without ICU beds, and more and more hospitals may run out of the basic supplies they need to respond to the outbreak.

A flatter curve, on the other hand, assumes the same number of people ultimately get infected, but over a longer period of time. A slower infection rate means a less stressed health care system, fewer hospital visits on any given day and fewer sick people being turned away.

In this project, I have analysed the whole dataset whose source is [Johns Hopkins University Center for Systems Science and Engineering](https://raw.githubusercontent.com/datasets/covid-19/master/data/time-series-19-covid-combined.csv).

All the codes have been written in Google Colab.

Since, this is a huge dataset. So, for reducing the data set for my own convenience, I had opted for Pivot Table which would take the necessary tables i.e, Confirmed cases, dates and country/region into consideration.

Several plots have also been plotted by using matplotlib.

The code for the maximum cases has been shown in the Google Colaboratory file.

* **Total Number of cases:** Everyday the cases are spiking up. This is the maximum value of total cases in the whole world is **14070273.0**
* **Country with maximum total no. of cases:** The country with maximum total number of cases are in **India.**
* **Worst affected country:**

It is becoming impossible for several countries to control the outbreak of this deadly virus even after taking every possible precautions. By evaluating the total number of cases, we can see that the worstly affected country is **India**.

* **Highest number of active cases** is **8106384**.
* **Country with maximum number of active cases:**  The country with maximum number of recovered cases is in the **United States**.
* **Maximum Recovery Rate** - However, the rate of recoveries is a poor, if not meaningless, measure of a country’s response to its COVID-19 epidemic. The maximum recovery rate is **0.500768049155146.**
* **Country with Maximum Recovery Rate:** This has been recorded in **Australia**. Recovery rate in a pandemic is not considered a scale to measure the success of containment strategy. The logic is simple that all those people who do not die during a pandemic, recover.
* **Maximum number of recovered people:-**

The only ray of hope which we can see is the number of recovery cases. This is a huge satisfaction in these crucial hours. The maximum number of recovered people are **38878**.

* **Country with maximum number of recovered people:**

The country with maximum number of recovered people are in **India**.

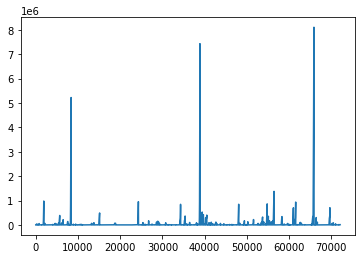
* **Maximum number of Deaths:** The maximum number of deaths has been recorded a figure of **65879** in the **United states**.

The graph plots are already there in the google colab file.

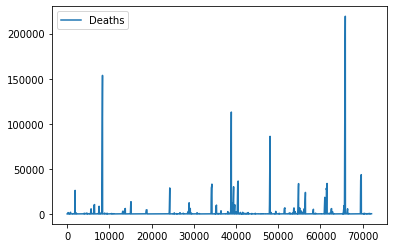
**Visualisation:**

**Plots:**

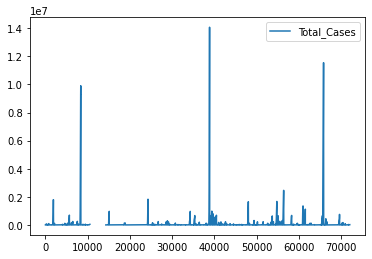
**Confirmed Cases Line Graph:**



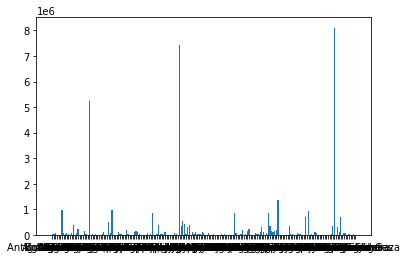
**Death Cases Line Graph:**



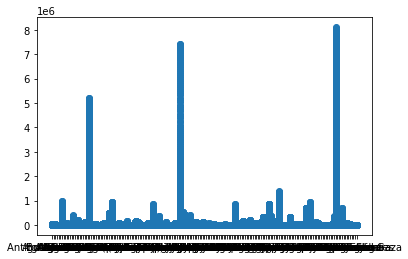
**Total Cases Line Graph:**



**Bar Graph:**

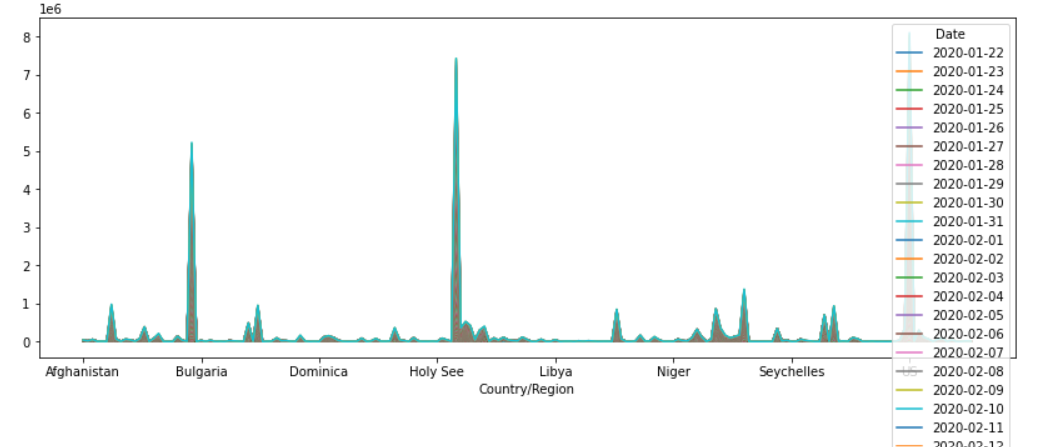


**Scatter Plot:**



Since, this is a huge set and most of the values are same so they are overlapping. And, we are not getting a proper scatter plot.

**Pivot Table Graph:** Here, columns are filled with **Dates**, Rows are **Country/Region** and the values are filled with **Confirmed Cases**.

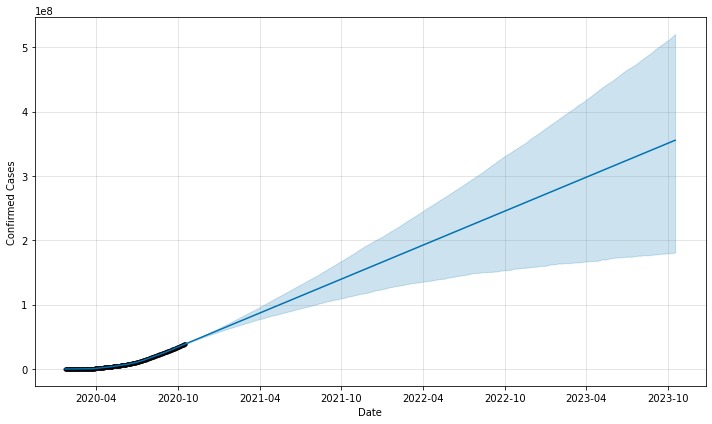
****

**Forecasting**

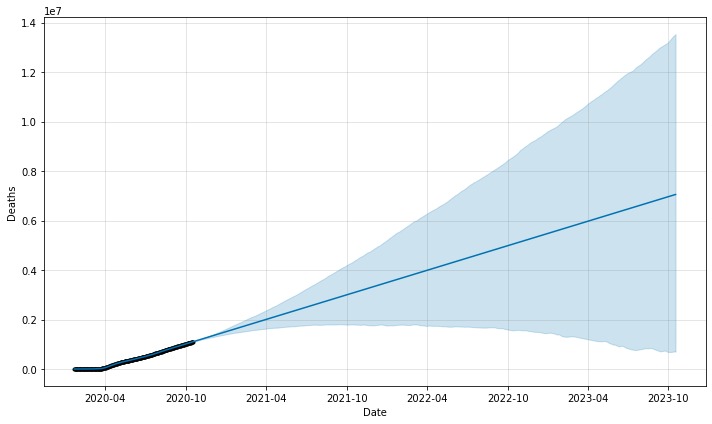
After **Forecasting**, these plots can be predicted within a span of three years. And, it can be seen from the graphs that it would take 2023 to completely flatten the curve.

The **predicted plots** for confirmed, recovered and death cases are:-

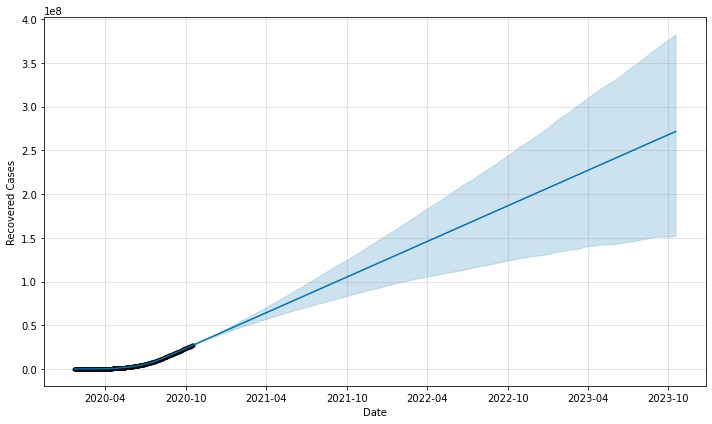
**Predicted Confirmed Cases’ Plot:**



**Predicted Death Cases’ Plot:**

****

**Predicted Recovered Cases’ Plot:**

****