**Project Report**

The World Health Organization (WHO) first declared COVID-19 a world health emergency in January 2020.COVID-19 pandemic is having a noticeable impact on global economic growth. Estimates so far indicate the virus could trim global economic growth by as much as 2.0% per month. Our project is to estimate the impact of COVID-19 on the economy and as well clustering and classifying the countries based on the metrics from the COVID-19 data. Collected data from Our World in Data. Our complete COVID-19 dataset is a collection of the COVID-19 data maintained by Our World in Data. It is updated daily and includes data on confirmed cases, deaths, and testing. Economic data is gathered from OECD and also Gap minder. After retrieving the data needed to clean it and check for sanity and make sure that all metrics are in proper data type and also find out it there are any missing values and replace them with average or delete the row if it was not useful. keeping location, date, total\_ cases, total\_ deaths, case per million, death per million as useful columns for the prediction. After cleaning and doing the sanity checks, I did some basic analysis on the data to find some trends and also find the country with total number of cases and deaths. Also Scatter plot of the five countries with the total number of cases and total death. Now we move on to more complex features such as Principal component analysis(PCA).I applied PCA on total\_cases' , 'total\_deaths' ,'total\_cases\_per\_million and took two components as PCA1 and PCA2 which will help in the regression and prediction. Based on this we used PCA to find the Kmeans clustering for the countries and divided into 3 clusters based on the countries . Inflation rate (CPI): Inflation measured by consumer price index (CPI) is defined as the change in the prices of a basket of goods and services that are typically purchased by specific groups of households. Inflation is measured in terms of the annual growth rate and in index, 2015 base year with a breakdown for food, energy and total excluding food and energy. Share price: Share price indices are calculated from the prices of common shares of companies traded on national or foreign stock exchanges. They are usually determined by the stock exchange, using the closing daily values for the monthly data, and normally expressed as simple arithmetic averages of the daily data.We now do the logistic regression on X= new case rate and death rate and y = inflation rate using linear regression and we got the score as -2.9 which does not fit the model . so we set the binary and try to fit the model we get score as 0.8 hence the model fits and has high accuracy. ROC-AUC curve can be used to judge the classification effect, above we can see two AUC number, 0.85 for “share price value”, 0.375 for “inflation rate”, so the accuracy of classification for “inflation rate” is not good. Logistic regression model fits this case well. From the above results of two index: “new\_case\_rate” and “death\_rate” , we conclude that Covid19 affects the economy by affecting share prices and inflation.

A close up of a person

Description automatically generated

A close up of a device

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A screenshot of a social media post

Description automatically generated

A screenshot of a map

Description automatically generated

A close up of a map

Description automatically generated