

## COVID 19 PANDEMIC ANALYSIS

Advanced analytics & Dashboard design

TOOL : JUPYTER & TABLEAU PUBLIC

LANGUAGE : PYTHON



Covid 19 is an ongoing global pandemic and has become one of the deadliest pandemic in history, here we will uncover how the different countries across the globe are dealing with this.

**Project Goal** :To get the insights about Covid 19 Pandemic throughout the World in terms of mortality and vaccination drives of different countries using different analytical approaches

**Skills** : Data cleaning and Wrangling, Exploratory analysis, Machine Learning Techniques – regression and clustering ,Geospatial analysis, Analysing time series data.

**Data Source.** : open-sourced from [ourworldindata.org](https://ourworldindata.org). This data describes the country wise details for the important variables like cases, deaths, vaccination and many more from beginning of the Pandemic till 10<sup>th</sup> Dec 2021. The original data set has 138727 Rows and 28 Columns

For drawing insights, I first started with exploring the relationships between different variables in data, then I worked on the strongest relationships. While analyzing Number of People vaccinated & Mortality rate, I formed the hypothesis – “ Countries with high vaccination rates have low mortality “

Later I tested that hypothesis using regression analysis in python and found out that the relationship between the variables was not entirely linear and there were many outliers

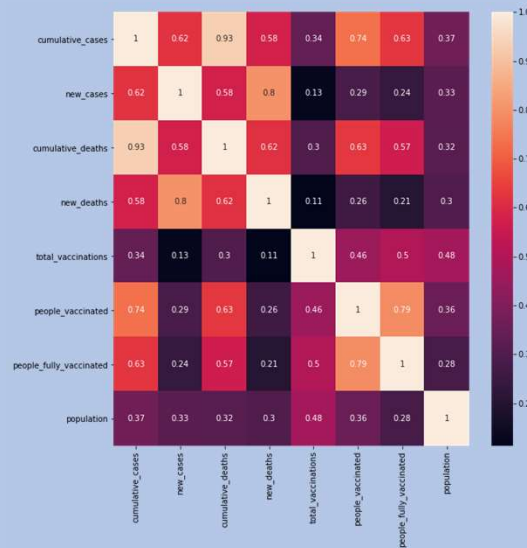


Fig 1 : Correlation Matrix

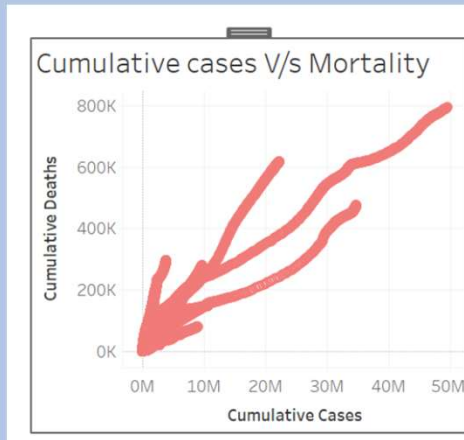


Fig 2 : Exploring relationship between cumulative cases & Mortality using scatter plot

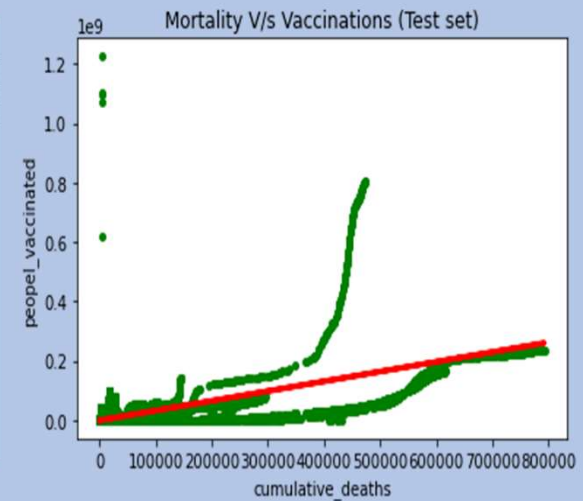


Fig 3 : Linear Regression Analysis

As Linear regression wasn't enough to fully explain the data, so I used another non linear approach to analyze the data – Cluster Analysis where I was able to group the data into three different clusters and by further analyzing the differences in clusters , I was able to identify the where are these clusters located and how differently all the countries are coping with the Pandemic

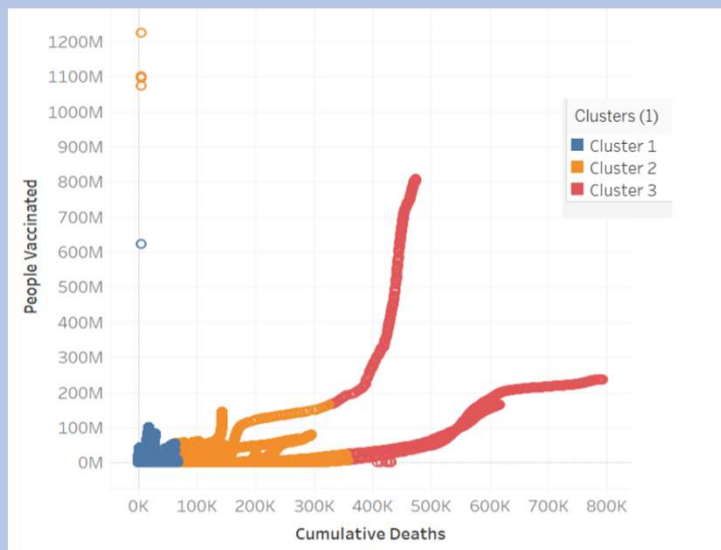
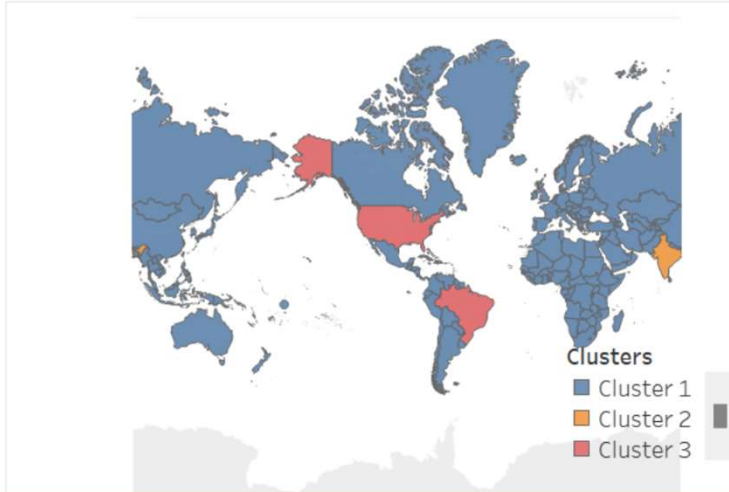


Fig 1 : Cluster Analysis



Fig 2 : Differences between Clusters

Where are these clusters located



[Click to see Tableau Story](#)

[Click to see project on Github](#)

### Key findings :

with most cases have most fatality, and the data points we have so far for these variables shows that vaccination is not making much difference in controlling mortality rate.

I was able to identify three clusters – countries with high fatality and majority population Vaccinated, countries with high fatality but under vaccinated, countries with moderate fatality and under vaccinated

### Limitations :

There weren't enough datapoints to yield a highly significant result.  
Data Collection bias

### Next Steps :

- Gather more datapoints on these variables and run the cluster analysis again.
- Analyze the impact of additional variables