Capstone Project - The Battle of Neighbourhoods against COVID19

June 2021

1. Introduction

Background

The ongoing Covid19 Pandemic caused a serious distress across globe. The statistics and data related to COVID19 is humongous and is being evolved day by day. The impact of COVID19 on each country is different based on population, social, health profile of that particular country. The response to COVID19 for each country is also different based on resources availability. The recovery from COVID19 pandemic is largely dependent on fighting it as one neighbourhood in quickly vaccinating all populations across the globe. This requires, first an evaluation of impact of COVID19 on all countries and their response so far. Mankind can win against COVID19 if we fight it as one neighbourhood and find vaccination clusters based on impact, need and urgency. Therefore, it is important to evaluate the COVID19 data for all countries and find how similar or dissimilar a particular country is to its nearest neighbour in terms of impact and response.

2. Business Problem

The objective of Capstone Project is to evaluate the impact of COVID19 on all countries and find how similar or dissimilar are the countries in terms of COVID19 impact.

Using Data Science, this project aims

1. To provide vaccination clusters based on clustering countries as per COVID19 impact for effective distribution
2. Based on Individual COVID19 response, find cluster of countries which require enhanced attention and resources

3. Data acquisition and cleaning

Data sources

Coronavirus Country profiles are available in Coronavirus Pandemic (COVID-19) - Statistics and Research - Our World in Data.

We built 207 country profiles which allow you to explore **the statistics on the coronavirus pandemic for every country in the world**.

In a fast-evolving pandemic it is not a simple matter to identify the countries that are most successful in making progress against it. For a comprehensive assessment, we track the impact of the pandemic across our publication and we built country profiles for 207 countries to study in depth **the statistics on the coronavirus pandemic for every country in the world**. Each profile includes **interactive visualizations**, **explanations** of the presented metrics, and the details on the **sources of the data**.

Every country profile is **updated daily**.

Data Preparation

The available Data extracted from above mentioned link is in JSON format. It includes all of our historical data on the pandemic up to the date of publication.

As of 3 June 2021, the columns are:

iso_code, continent, location, date, total_cases, new_cases, new_cases_smoothed, total_deaths, new_deaths, new_deaths_smoothed, total_cases_per_million, new_cases_per_million, new_cases_smoothed_per_million, total_deaths_per_million, new_deaths_per_million, new_deaths_smoothed_per_million, reproduction_rate, icu_patients, icu_patients_per_million, hosp_patients, hosp_patients_per_million, weekly_icu_admissions, weekly_icu_admissions_per_million, weekly_hosp_admissions, weekly_hosp_admissions_per_million, total_tests, new_tests, total_tests_per_thousand, new_tests_per_thousand, new_tests_smoothed, new_tests_smoothed_per_thousand, positive_rate, tests_per_case, tests_units, total_vaccinations, people_vaccinated, people_fully_vaccinated, new_vaccinations, new_vaccinations_smoothed, total_vaccinations_per_hundred, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, new_vaccinations_smoothed_per_million, stringency_index, population, population_density, median_age, aged_65_older, aged_70_older, gdp_per_capita, extreme_poverty, cardiovasc_death_rate, diabetes_prevalence, female_smokers, male_smokers, handwashing_facilities, hospital_beds_per_thousand, life_expectancy, human_development_index, excess_mortality

For the scope of the Business problem, it is aimed to study the covid19 spread across various countries in Data base and response in terms of tests, facilities and vaccinations

JSON data is further reduced to two separate Dataframes as below

1) **df_country_data** which consists each country's demographic and health profile ['location', 'population', 'population_density','median_age', 'aged_65_older', 'aged_70_older','cardiovasc_death_rate', 'diabetes_prevalence','life_expectancy','hospital_beds_per_thousand','human_development_index','extreme_poverty']
2) **df_covid** which consists each country's covid data till data publication

These dataframes were further cleaned by dropping rows with null values wherever applicable date wise data is aggregated with respect to each country. Geographic coordinates were extracted from geolocator

4. Exploratory Data Analysis

The data is explored thoroughly and statistics were reviewed.

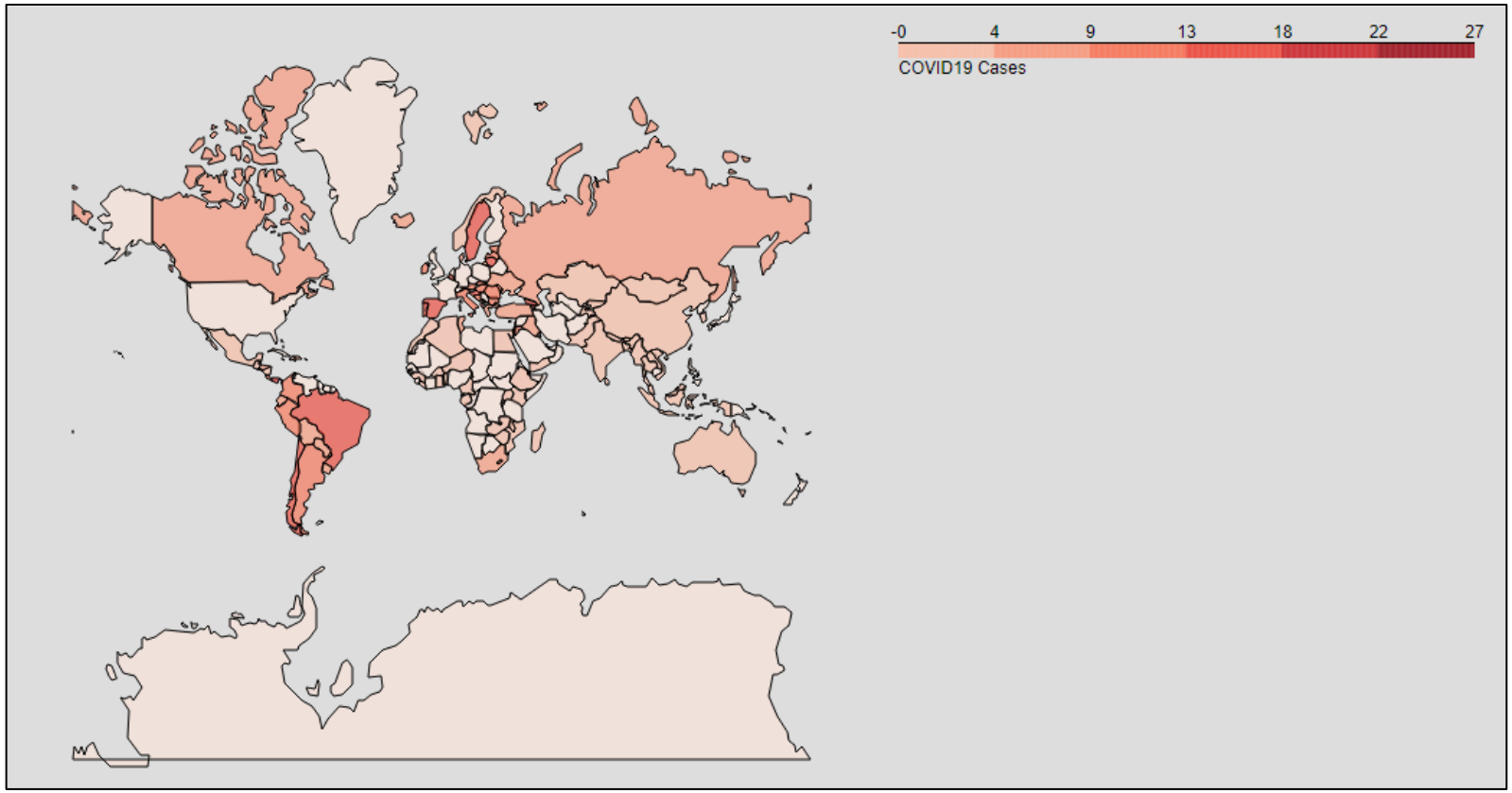

Figure 1 World map to show covid cases relative to population

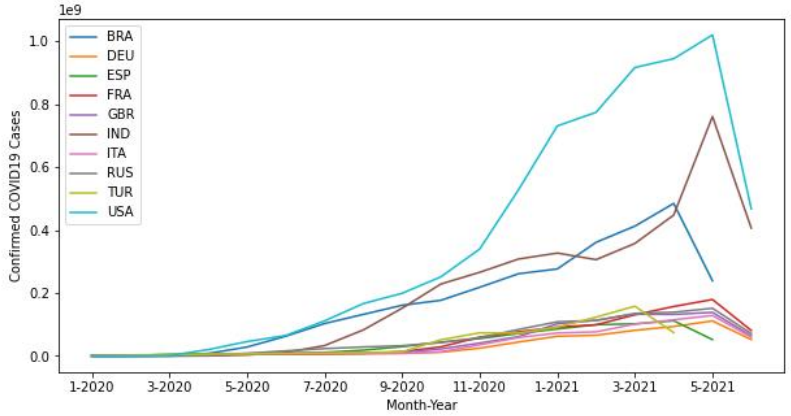

Figure 2 Plot Cumulative Confirmed Cases for Top 10 Countries

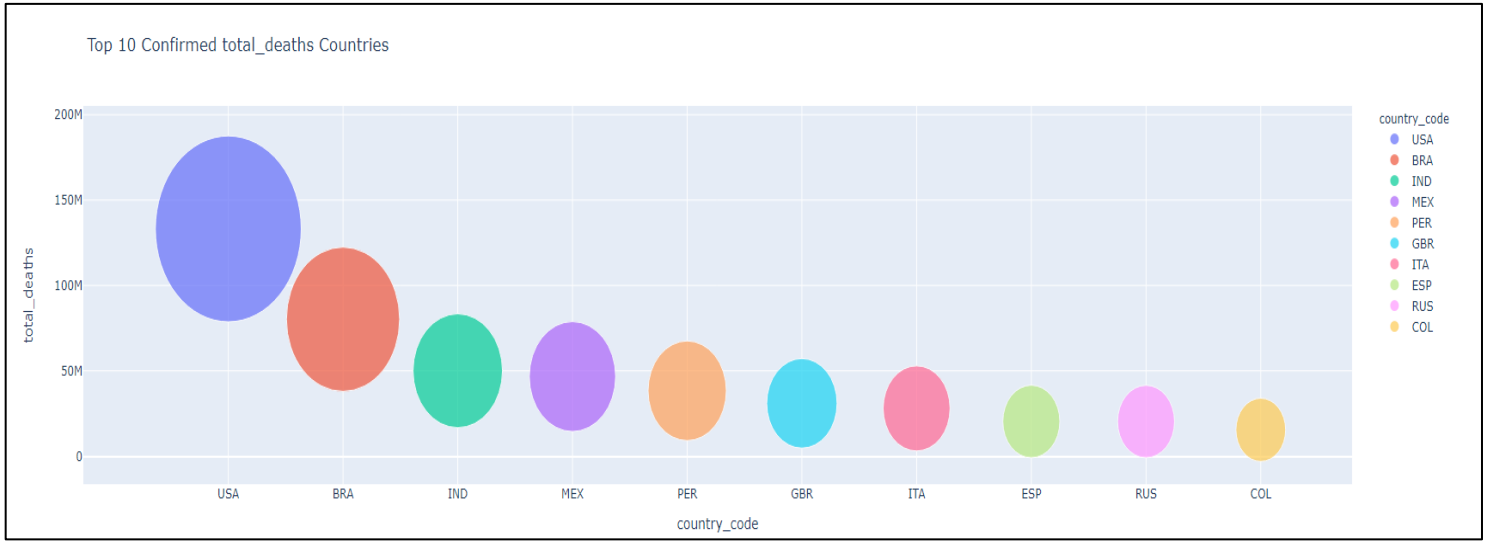

Figure 3 Top 10 Countries confirmed total_deaths countries

5. Feature selection

By study various Data visualisation and correlation techniques, below features are selected to study the COVID19 impact on each country.

Data_final[['location','population','population_density','median_age','aged_65_older','aged_70_older','cardiovasc_death_rate','diabetes_prevalence','life_expectancy']]

Similarly, to understand the response of each country against COVID19, these features as below are selected.

Data_final[['location','hospital_beds_per_thousand','human_development_index','extreme_poverty','total_cases','total_deaths','total_tests','people_fully_vaccinated','total_cases_per_million','total_deaths_per_million','people_fully_vaccinated_per_hundred']]

6. Classification

Kmeans classification is performed on the finalized data separately one for COVID impact and secondly for COVID response. Elbow method is used for selecting optimum clusters.

Figure 4 World Map with Covid data for clustering countries

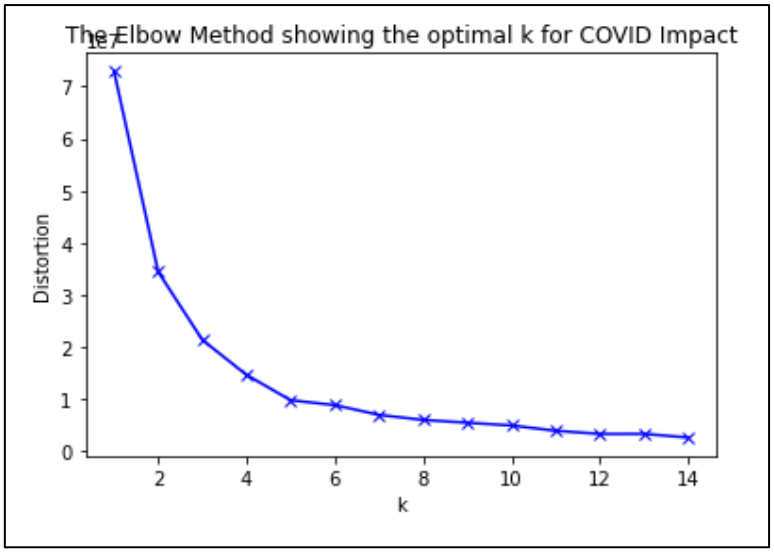

Figure 5 Kmeans Clustering of countries as per COVID19 Impact: Fit Kmean Clustering

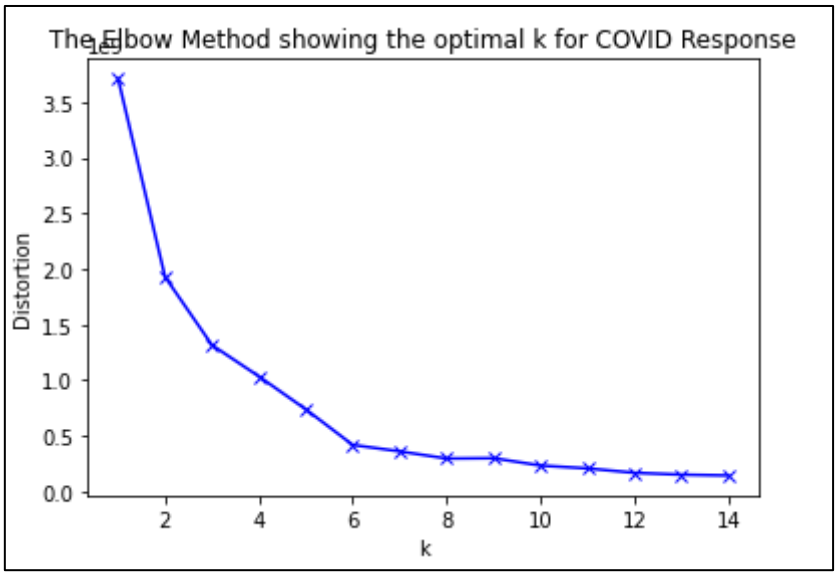

Figure 6 Kmeans Clustering of countries as per COVID19 Response: Determine optimal k

7. Conclusions

8 clusters were found based on COVID19 impact as shown in Figure 7. Similarly, 8 clusters were formed as per COVID19 response

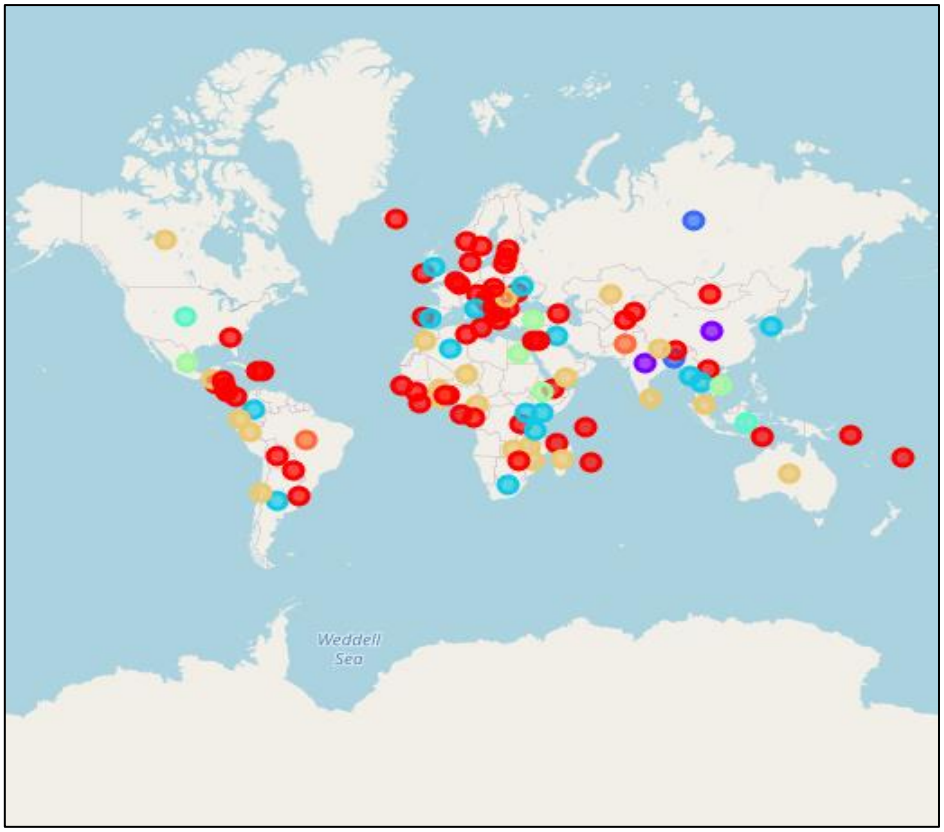

Figure 7 Clustering Results: Create a map of countries with similar COVID19 impact in terms of cases and deaths

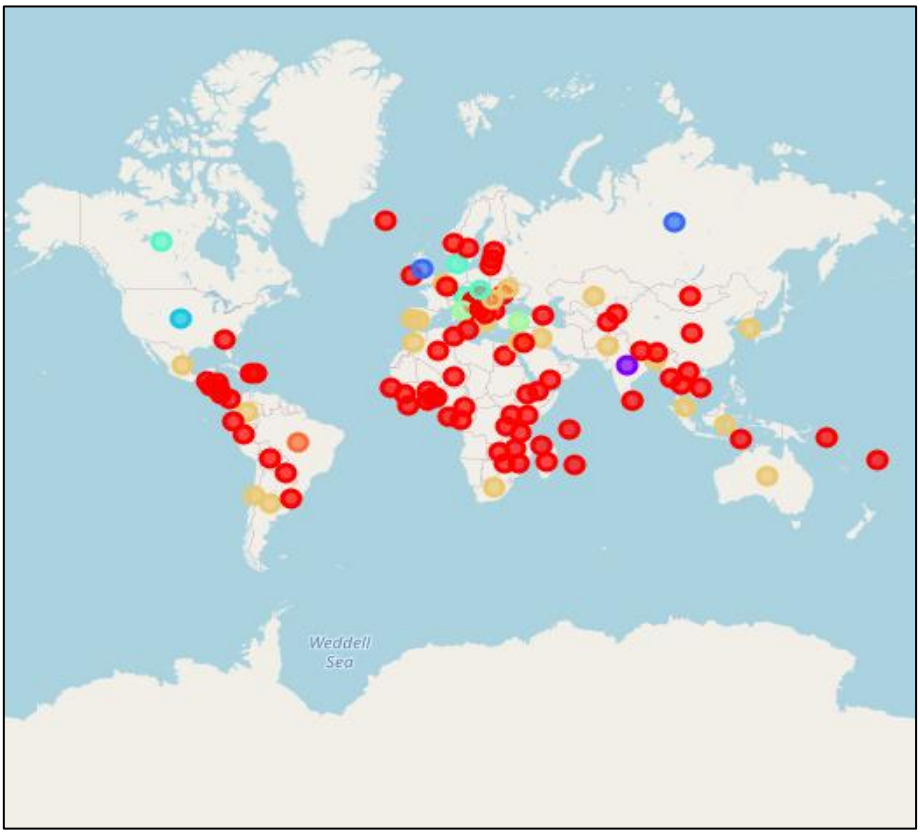

Figure 8 Clustering Results: Create a map of countries with similar COVID19 response in terms tests and Vaccinations

8. Model Evaluation and Future Scope

The model has to be evaluated with evolving data as the pandemic is spreading with multiple variants. Clustering approach and Resulting clusters will help the best use of resources for strong recovery from COVID19