THE BATTLE OF NEIGHBORHOODS

CORONA VIRUS OUTBREAK ANALYSYS

a Capstone Project Report by

MUJEEBUR RAHMAN

under the course

Applied Data Science Capstone

of the Professional Certification

IBM Data Science

Introduction:

Problem Background:

- In late 2019, a novel coronavirus, now designated SARS-CoV-2, was identified as the cause of an outbreak of acute respiratory illness in Wuhan, a city in China.
- In February 2020, the World Health Organization (WHO) designated the disease COVID-19, which stands for coronavirus disease 2019.
- Since the first reports of COVID-19, infection has spread to include more than 80,000 cases in China and increasing cases worldwide, prompting the WHO to declare a public health emergency in late January 2020 and characterize it as a pandemic in March 2020.
- The rate of new infections outside of China has surpassed that within China as epidemics have grown in other countries.
- The possibility of COVID-19 should be considered primarily in patients with fever and/or lower respiratory tract symptoms who have had recent close contact with a confirmed or suspected case of COVID-19, who reside in or have recently (within the prior 14 days) traveled to areas where community transmission has been reported (e.g., China, South Korea, most of Europe [including Italy], Iran, Japan) or who have had potential exposure from specific settings where COVID-19 cases have been reported.

Problem Description:

Suppose we travel to some country for some vacation or business purpose. In such a situation, we would not know whether that country is inflicted by this deadly virus. Even if that country or city is not affected, it might be affected soon due to the nearby countries or cities. As a summary, this problem can be asked in the form of questions as follows:

- How many countries are affected by the virus?
- Does the country where I want to go affected?
- Do the nearby countries are affected?
- Are the unaffected countries being followed with healthy and safety measures?

This model will give an analysis on the countries which are affected by this coronavirus.

Data:

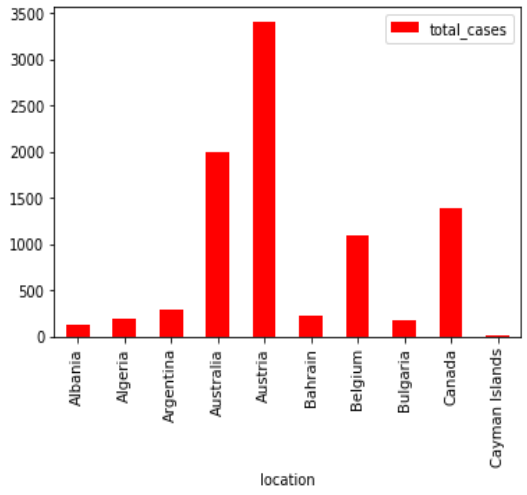
To understand and explore, we will need the source data for coronavirus and the latitude and longitude data for the countries. These data can be obtained from the following mentioned links.

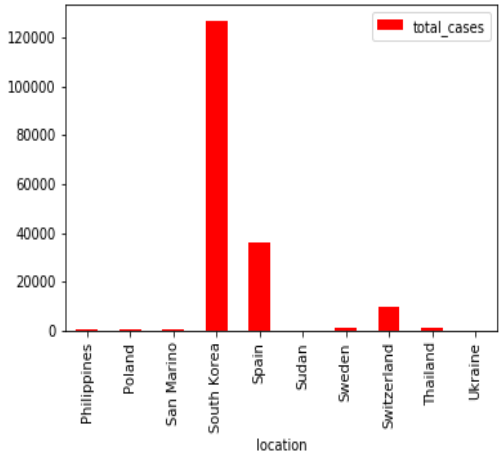
- **Coronavirus Source Data:** https://ourworldindata.org/coronavirus-source-data
- **Latitude and Longitude Data:** https://www.kaggle.com/parulpandey/world-coordinates

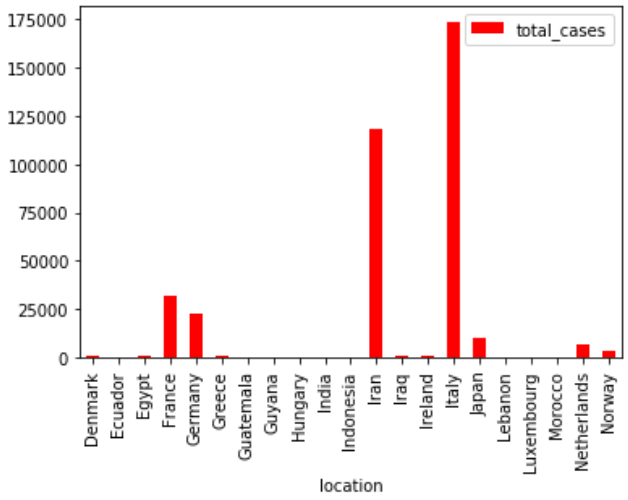
Methodology:

- First off, the required libraries are imported into the workspace for which analysis is to be done.
- Based on the above data, the source dataset about the Coronavirus is obtained from the website and is stored in a dataframe.
- This dataframe is processed and any unnecessary columns and null values are excluded and processed.
- Then the latitude and longitude information are also gathered and stored into another dataframe.
- Then these two dataframes are merged into a single dataframe so that the analysis can be done using a single dataframe itself.
- Now a Geojson file is created which will contain the information from the above created dataframe.
- Using this Geojson file, a choropleth map is made which contains all the locations which are affected by the corona virus.
- Then by using K-Means clustering, the locations are clustered together and about 3 clusters are formed.
- The remaining points which do not form these clusters are considered as outliers.
- These outliers indicate that these locations are the least affected areas, and there might be a great possibility that the surrounding areas might be affected soon enough, considering the rapid spread of this deadly virus.

The following plots show the number of people infected by the virus in each country:







Results:

- From the above map, it is obvious that European countries are the most affected.
- From the above plots, it can be seen that China, Iran, Italy and South Korea have the highest infected people so far.
- Therefore, migration from and to these countries should be avoided at any cost.
- It can't be said that the remaining countries are safe from this virus.
- So it is advisable that be in your own home and in your city and not to travel to any other place until this pandemic is completely dissolved.

Conclusion:

With the above discussions, it can be concluded that the areas which are mostly affected are most likely to affect the nearby areas. So migration to such areas should be avoided until the situation comes under control. It is advisable not to travel to any other city or country until this virus gets completely nullified.