COVID vs NEW YORK

**INTRODUCTION/BUSINESS PROBLEM**

New York Has been one of the most affected cities in the world during Corona Virus pandemic. It is one of the most influential and populous cities of the world which makes it even more worrying as such cities can have disastrous end effects with COVID 19. But I hail from the second most populous country in the world - India. India has done well in controlling the virus to some extent. It is a fact that it is only due to the lockdowns and restrictions imposed on every state. However India has now come up with a new Categorization of each district which puts every district into one of the three zones. 1)Red Zone (High alert for COVID 19 cases) 2)Orange Zone (Medium alert) 3)Green Zone ( Relatively very less or no cases of COVID 19)

This is a very good system to get the country back into normal ways as green zones dont have to be locked down and just needs to take containment measures. And also breaking the country down into such units can ensure that the green zones will help the country move forward instead of being stuck at home and deadstill. The target audience is the whole world. The Government of USA and the people of New York.

**DATA SECTION**

I will be trying to implement the same system on New York and try and cluster the Neighborhoods into such zones. Data used will include the Boroughs and Neighborhood data of New York City. The most populous places in New York can be found(Using Foursquare API) which will help to cluster them together. Data about the number of cases in New York and worldwide will help us create visualizations to understand the situation better.

**METHODOLOGY**

1)Data Requirements and Data Collection:

I started out with by scraping out a table from Wikipedia about the number of death cases worldwide as it is a good place to start and observe which countries are most affected by the virus and how bad America is affected compared to other countries.

Data Understanding: The data collected is a representative of the worldwide cases compared to America.

Data Preparation : The table had to be cleaned up , certain unrequired column were dropped, the columns that were integers needed to be changed into int type so as to perform functions on them

Data Visualization : The folium library was used to plot a chloropleth map and we can see from it that America is the most affected country in terms of deaths

Next I decided to get data for the number of cases reported in each Borough of New York City so we can better understand the situation in the country alone. Plotting a chloropleth map of the following data shows us the most affected Borough is Queens .

Using matplotlib I further was able to map the neighborhoods of Queens into another figure. The question from here to be answered from our business understanding is , Which of these neighborhoods are to be under serious lockdown ?(Only considering the Queens Borough as it is the most affected by the pandemic).

From here , I used the Foursquare API to get all venues of the main neighborhoods in Queens and was able to group them according to number of venues in a neighborhood. It is logically sound that a neighborhood with a lot of venues is much more capable of spreading the corona virus and so I used K Means Clustering to group the neighborhoods of Queens into 3 parts

**RESULT :**

2- Neighborhoods that require a strict lockdown

1. Neighborhoods that require a moderate lockdown
2. Neighborhoods that need to take precautions.

**DISCUSSION**:

The model only uses the number of venues in Queens as a source but this can be immensely improved using all sorts of other information such as the age and population demographic in Queens. (I was not able to get those data tables and due to a lack of time had to use only the venues data). If we are able to add in more variables as such and cluster the neighborhoods again for a second time we might get a much better segregation of the zones required.

**CONCLUSION**:

We were able to classify the Neighborhoods of Queens into 3 zones against the COVID 19 virus. Certain improvements to make are left free to make by the reader.