IBM Applied Data Science Capstone

Capstone Project

Comparison of New York Boroughs and the impact on outbreak of COVID 19

By: Usama Tariq

May, 2020

Introduction

The outbreak of COVID 19 pandemic has created a havoc around the global. Apart from infecting over 4 million people and causing almost 300 thousand deaths worldwide, the world economy is at a standstill. While researchers are still struggling to find a vaccine the world leaders have resorted to enforce social distancing through lockdowns. In the wake of these events, the need to break the transmission chain is of critical importance. Consequently, researchers need to identify the hotspots for the transmission to minimize the impact of COVID 19 on businesses and society at large.

Business Problem and Target Audience

The objective of my project is to compare the New York boroughs with highest and lowest number of COVID 19 cases and identify the neighborhood clusters which are existent in most affected borough only. This will help to highlight the hotspots in terms of the most common venues which might have contributed towards surge of COVID 19 cases in the respective boroughs. Once the hotspots are identified the analysis can be used effectively to control the outbreak and minimize the social and economic impact of COVID 19. The analysis can be used by governments to design public health policy by allocating medical resources in such neighborhoods which are at risk of the COVID 19 outbreak as well to lockdown the hotspots or to enforce strict prevention measures in these areas. Imagine if a location is at the risk of COVID 19 outbreak and the policy makers are already aware of the target areas in which the resources are to be deployed. Not only this will help the governments financially but will probably save a lot of lives too.

Data Requirements and Sources

In order to identify the boroughs with highest and lowest number of COVID 19 cases in New York, we will be using the data provided by Statista website. The data reference is attached as Annexure 1. Once the two boroughs are identified we will then use the New York Json file

available in week 3 of Coursera Capstone to acquire the geolocation of the neighborhoods in the New York boroughs. Since this data has already been tried and tested in Week 3 lab the data file is quite comprehensive and will successfully fetch the venue details from the foursquare API.

Further, Foursquare API calls will be used to identify the common venues in the neighborhoods, and they will be clustered using K-means. For visualization purpose, GEOJSON data available on GITHUB will be used to segregate the New York boroughs in the folium map.

Annexure 1