1. Introduction

Data science is the art of manipulating data to gain insights and knowledge from it with the help of knowledge, tools. The scope of this course is to teach us how this process is handled in order to find data-driven solutions to problems we face.

One of the biggest problems of humanity is COVID-19 pleague nowadays. It can be thought as highly contagious flu-like virus. Many data scientists are working on different cases regarding the issue. After acquiring some knowledge withthe help of the course track, the author wanted to apply them in a real-world case and see the effect of data driven solutions.

2. Business Problem

Main problem of the new virus type is that it is so contagious. Scientists say it can contaminate you by touching, from air etc. Although some countries did not take this seriously at first, now many declared quarantine to slow down the spread. During he negligence time, public places took an important role. The project aims to find the relation between the number o confirmed cases (till 05.04.2020) and commony visited public venues. We will cluster the neighborhoods by taking the number of cconfirmed cases into consideration, then inside the clusters, we will find which type of venues are more likely to be visited by those neighborhoods. We will comment on them afterwards in the discussion section.

3. Data

usdata: Contains New York neighborhoods, their locations, population
    Data source: Kaggle (for the number of cases in each neighborhood)
    Google (for latitude ad longitude values)
ukdata: Contains Londan neighborhoods, their locations, population
    Data source: Public Health England (for the number of cases in each neighborhood)
    Wikipedia (for latitude and longitude values)
Foursquare Data: Venue information will be withdrawn from Foursuare

- usdata and ukdata will be used to cluster neighborhoods based on the number of confirmed cases, seperately.
  (https://github.com/semanurkps/IBM_capstone/blob/master/ukdata.xlsx)
  (https://github.com/semanurkps/IBM_capstone/blob/master/usdata.xlsx)
- After the clusters are formed, with the help of Foursquare location data, most commonly visited venues will be listed and venue types will be compared to see if a certain type has affect on the spread.
- London and New York will be evaluated seperately to see how different countries' big cities are affected by the virus.