# IBM Applied Data Science Capstone Project

**Subject: Impact of Local Venues in COVID-19 Spread :**
**A Comparison Between London and New York**

**Semanur KAPUSIZOĞLU**
**semanurkps@gmail.com**

# Table of Contents

# Introduction

One of the biggest problems of humanity is COVID-19 plague 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 with the help of the course track, the author wanted to apply them in a real-world case and see the effect of data driven solutions.

# Business Problem

- Highly contagious virus
- People neglecting the effects of virus
- Governments being slow on taking precautions
- Public places remaining open and visited

The project aims to find the relation between the number of confirmed cases (till 05.04.2020) and commonly visited public venues. We will cluster the neighborhoods by taking the number of confirmed 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.

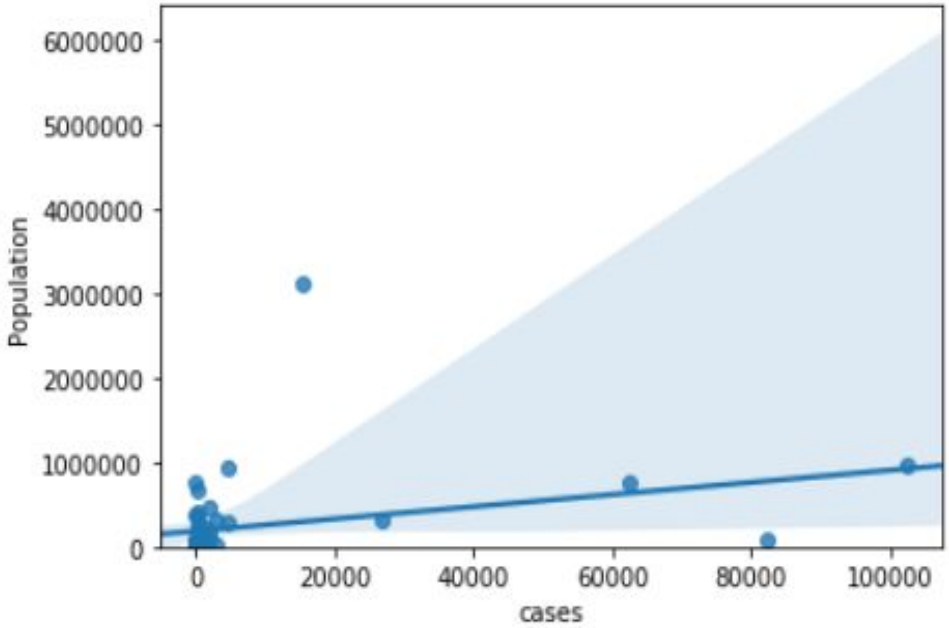
# Data- New York

| | Population | Latitude | Longitude | cases |
|---|---|---|---|---|
| count | 5.700000e+01 | 57.000000 | 57.000000 | 57.000000 |
| mean | 2.245278e+05 | 42.447361 | -75.532777 | 5608.403509 |
| std | 4.541461e+05 | 0.911117 | 1.743888 | 19083.955749 |
| min | 3.520000e+03 | 39.625500 | -79.466800 | 4.000000 |
| 25% | 3.004600e+04 | 42.008400 | -76.623800 | 70.000000 |
| 50% | 7.588000e+04 | 42.601200 | -75.165200 | 148.000000 |
| 75% | 2.237740e+05 | 43.025600 | -73.962600 | 1021.000000 |
| max | 3.116069e+06 | 44.447300 | -72.615100 | 102386.000000 |

```
Neighborhood     object
Population        int64
Latitude        float64
Longitude       float64
cases             int64
dtype: object
```

```
sns.regplot(x="cases", y="Population", data=ny_df)
plt.ylim(0,)
```

```
(0, 6395708.729889891)
```




```
ny_df[["Population","cases"]].corr()
```

| | Population | cases |
|---|---|---|
| Population | 1.000000 | 0.301913 |
| cases | 0.301913 | 1.000000 |

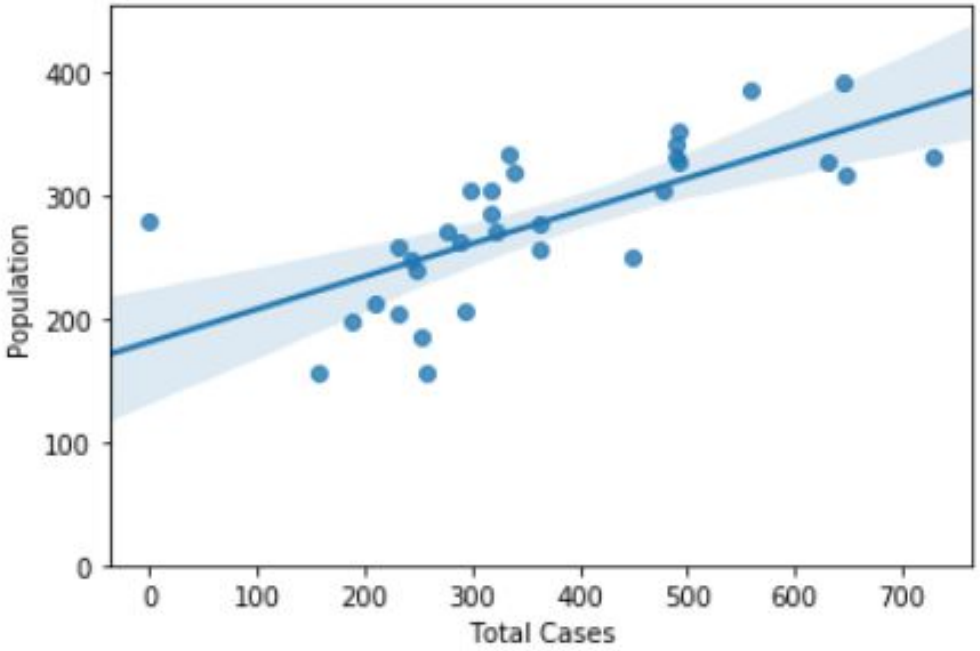
# Data-London

| | Population | Latitude | Longitude | Total Cases |
|---|---|---|---|---|
| count | 32.000000 | 32.000000 | 32.000000 | 32.000000 |
| mean | 277.503188 | 51.505666 | -0.119197 | 363.531250 |
| std | 61.422621 | 0.072662 | 0.161904 | 163.914018 |
| min | 156.197000 | 51.361800 | -0.476000 | 0.000000 |
| 25% | 245.229000 | 51.456250 | -0.205325 | 251.000000 |
| 50% | 278.182500 | 51.505600 | -0.114100 | 320.500000 |
| 75% | 326.056250 | 51.558850 | -0.011525 | 489.000000 |
| max | 392.140000 | 51.653800 | 0.183700 | 728.000000 |

```
Neighborhood      object
Population       float64
Latitude         float64
Longitude        float64
Total Cases        int64
dtype: object
```

```
sns.regplot(x="Total Cases", y="Population", data=lon_df)
plt.ylim(0,)
```

```
(0, 453.2948313607337)
```




```
lon_df[["Population","Total Cases"]].corr()
```

| | Population | Total Cases |
|---|---|---|
| Population | 1.000000 | 0.708661 |
| Total Cases | 0.708661 | 1.000000 |

# Methodology

- Clustering
- Foursquare Venue Data
  - get_category_type function
  - getNearbyVenues function
  - return_most_common_ones function
- Folium
  - mapping

# Results

| **New York** | **1st Common Venue** | **2nd Common Venue** | **3rd Common Venue** |
|---|---|---|---|
| Cluster 0 | Pizza Place | Restaurants | Cafe-Coffee Places |
| Cluster 1 | Deli / Bodega | Yoga Studio | Dessert Shop |
| Cluster 2 | Yoga Studio | Cafe-Coffee Places | Beach |
| Cluster 3 | Gym | Cafe-Coffee Places | Shops |
| **London** | | | |
| Cluster 0 | Cafe-Coffee Places | Clothing store | Pub |
| Cluster 1 | Pub | Cafe | Clothing Store |
| Cluster 2 | Cafe-Coffee Places | Pub | Restaurants |
| Cluster 3 | Cafe-Coffee Places | Clothing Store | Restaurants |

# Conclusion and Future Directions

Before diving into the details, the plan was to compare how the preventive actions taken by those big cities, what is the effect of public places in this plague. I was hoping to emphasize the importance of taking effective measures earlier and slowing the spread. However, when I further think about it, I should have selected cities with similar timelines. For example first case seen on the same day or consecutive day to compare the two cities better, without bias.