Places most Covid-19 infected cases and deaths in New York

STUDENT: RAFAEL MEJIA S.

INTRODUCTION/ BUSINESS PROBLEM

The pandemic Covid-19 is a tragedy that affected a millions of people around the world.

United States was not exception, especially New York City. It's a place where the illness growth exponentially during the last three months.

For this reason we need to know the places where the cases and deaths for Covid-19 were most affected. Which are the features of the neighborhood and how can improve or prevent for the future the cases will not raise quickly.
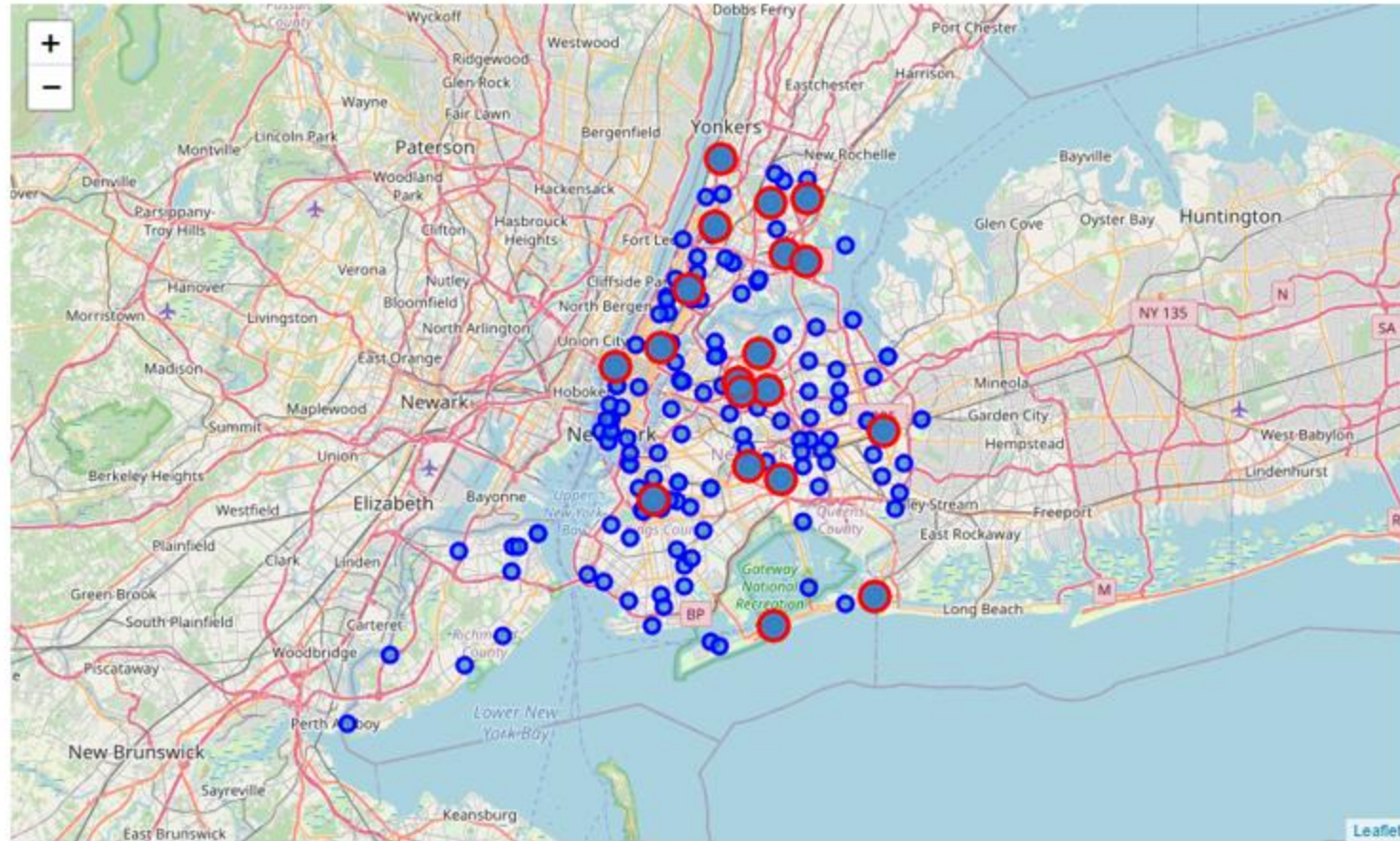
Data acquisition and cleaning

✓Use dataset of Covid-19 belong to New York City governance. This data contains the Zip, Neighborhood, Borough, Cases, Cases per 100.000, Deaths per 100.000 and the Percent of people tested who tested positive. The data contains information of 178 neighborhoods.

✓This data we will combine with the geolocation information of Foursquare to get the features of each neighborhood.

✓Only rename the columns of the dataset for better analysis.

✓I process the data to focus only in the top 20 most cases/death of Covid-19 because there are the current places more critical.
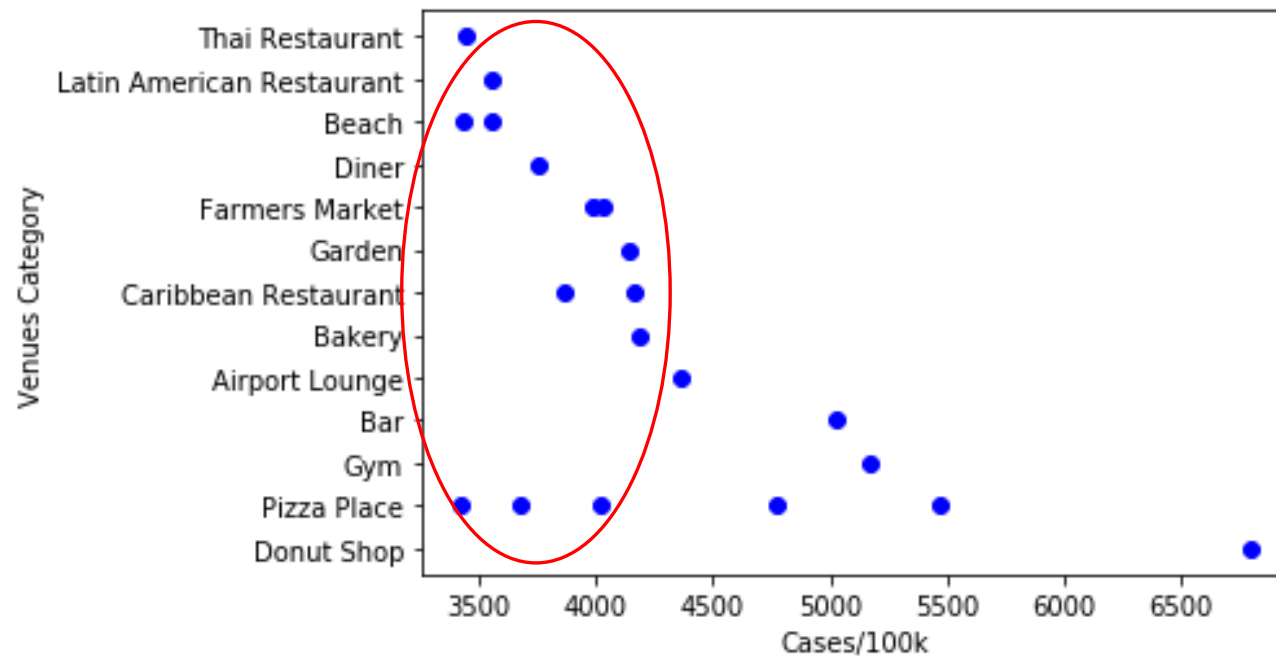
Data acquisition and cleaning

✓Use dataset of Covid-19 belong to [New York City governance](#). This data contains the Zip, Neighborhood, Borough, Cases, Cases per 100.000, Deaths per 100.000 and the Percent of people tested who tested positive. The data contains information of 178 neighborhoods.

✓This data we will combine with the geolocation information of Foursquare to get the features of each neighborhood.

✓Only rename the columns of the dataset for better analysis.

✓I process the data to focus only in the top 20 most cases/death of Covid-19 because there are the current places more critical.

Map top 20 (red circle) places most cases/deaths Covid-19

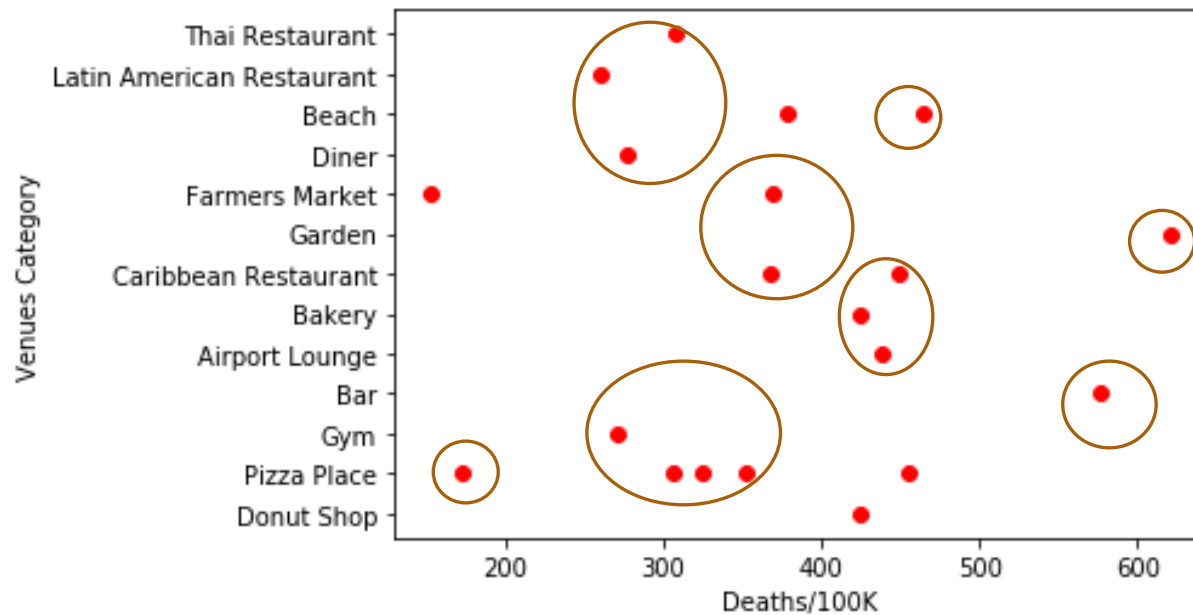# Relationship between Cases/100k and Venue Categories

In the graph that the infections are more in the beach, farmers market and food places where are people conglomerations.
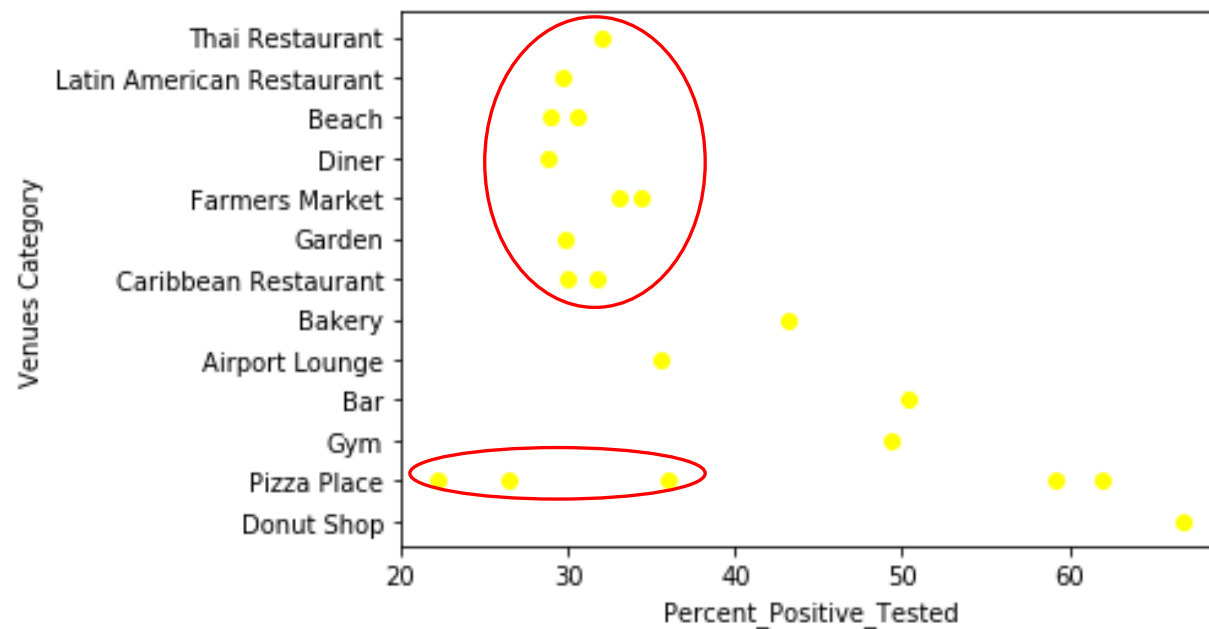
Relationship between Cases/100k and Venue Categories

In the graph not patterns in this case, in all categories are people death for the illness.
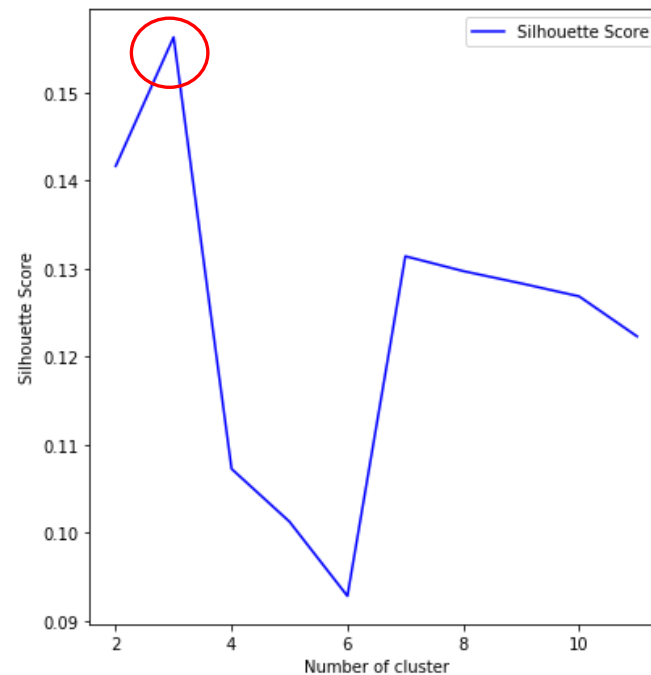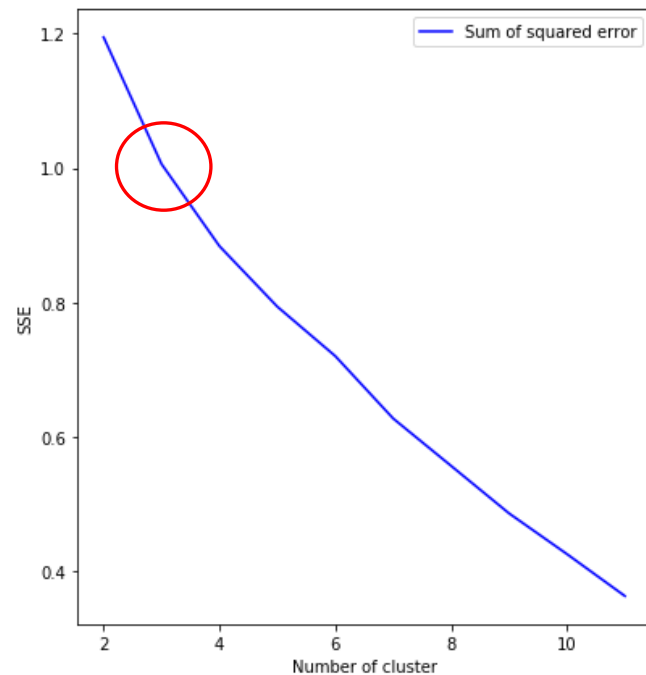
Relationship between Percent Positive Tested and Venue Categories

We watch in the graph similar pattern with Cases/100k chart. The concentration of people with percent positive teste are in the food places and beach.
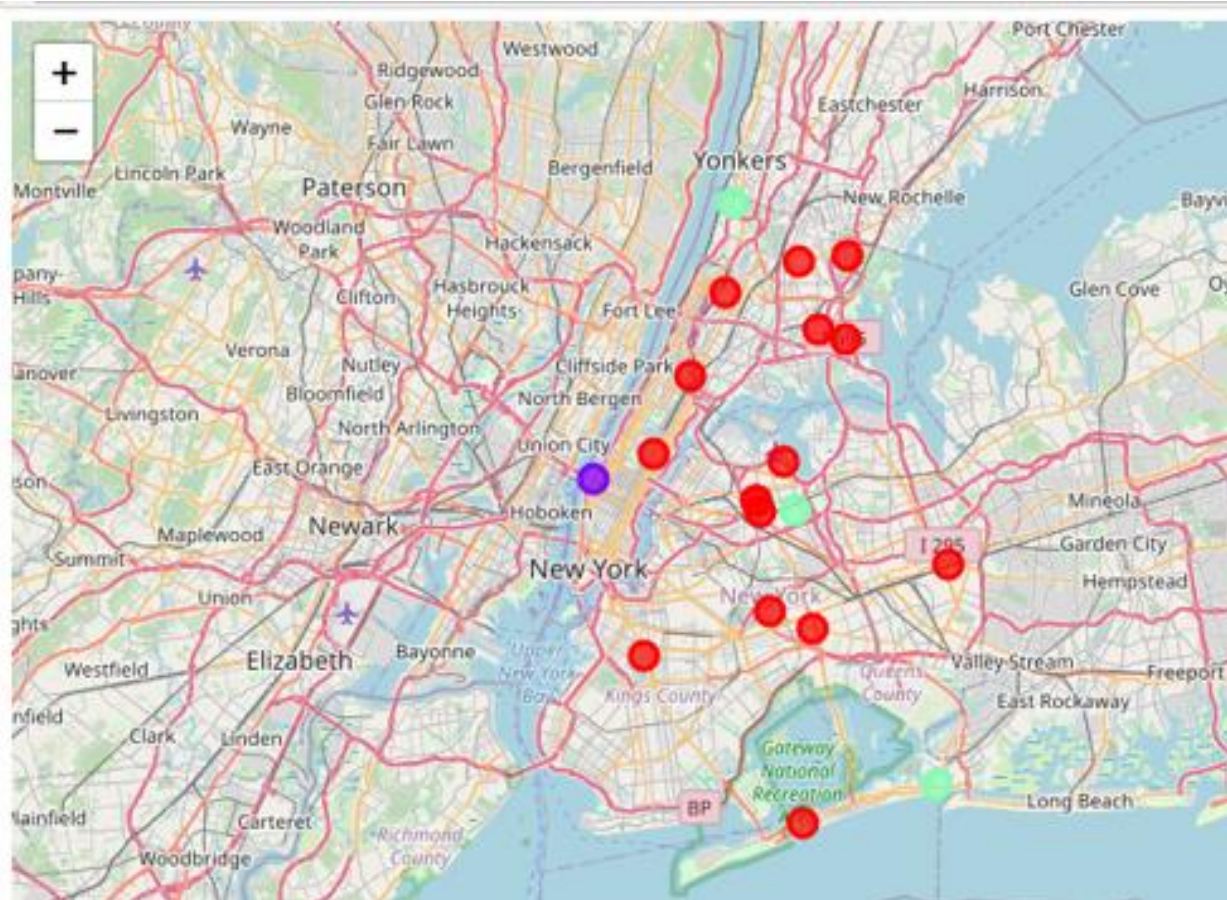
Best Optimal value K to clustering the data

Apply two methods: silhouette score and elbow (sum of squared error) algorithm.

```
For n_clusters = 2 The average silhouette_score is : 0.1416357207815753
For n_clusters = 3 The average silhouette_score is : 0.15626182241596792
For n_clusters = 4 The average silhouette_score is : 0.10727002617160457
For n_clusters = 5 The average silhouette_score is : 0.10127641846402476
For n_clusters = 6 The average silhouette_score is : 0.09283780680471007
For n_clusters = 7 The average silhouette_score is : 0.13140143641138
For n_clusters = 8 The average silhouette_score is : 0.12971286714015517
For n_clusters = 9 The average silhouette_score is : 0.12831980036515797
For n_clusters = 10 The average silhouette_score is : 0.1268479329030705
For n_clusters = 11 The average silhouette_score is : 0.12230592815501277
```

Watching the graphs and according to the results, the best value for K is n_cluster = 3 because the average silhouette scores the max value is 0.156.
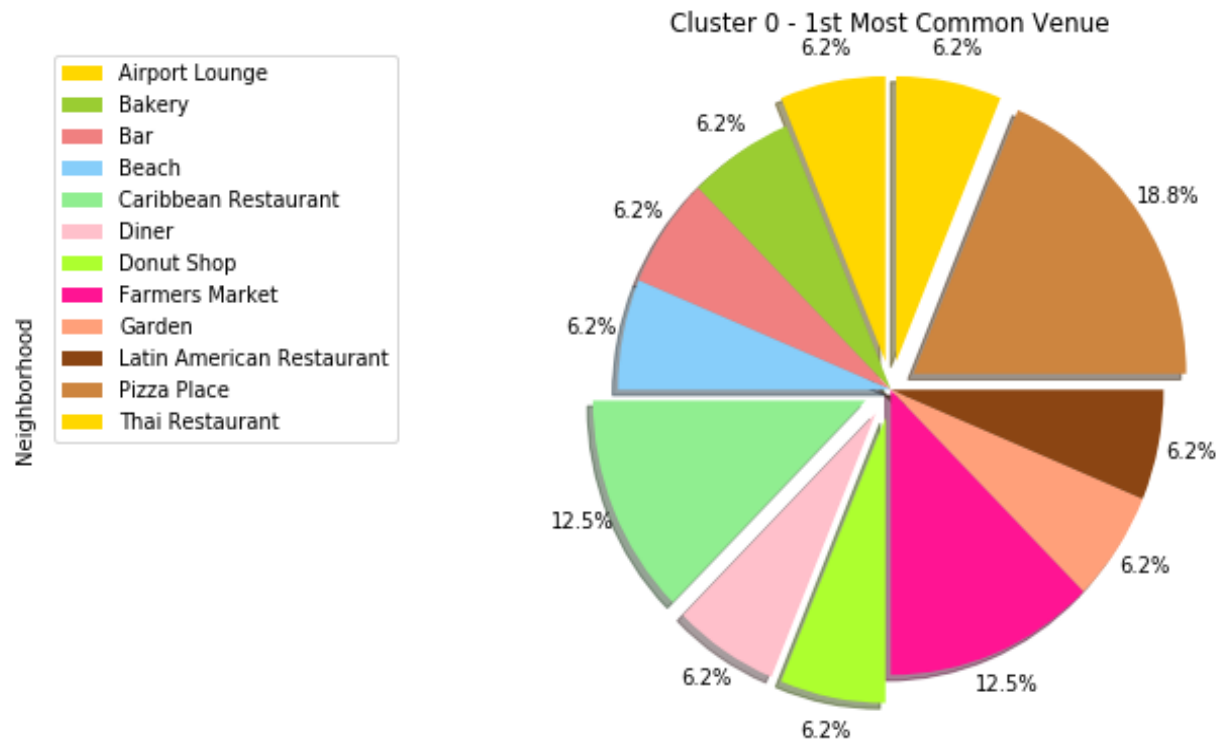
Map New York clustering Covid-19 cases/deaths

Where the red color is the cluster 0, red cyan is belong to cluster 1 and the color purple if for cluster 3.

Analysis Cluster 0 – Places Foods

Cluster 0 - 1st Most Common Venue

Neighborhood
- Airport Lounge
- Bakery
- Bar
- Beach
- Caribbean Restaurant
- Diner
- Donut Shop
- Farmers Market
- Garden
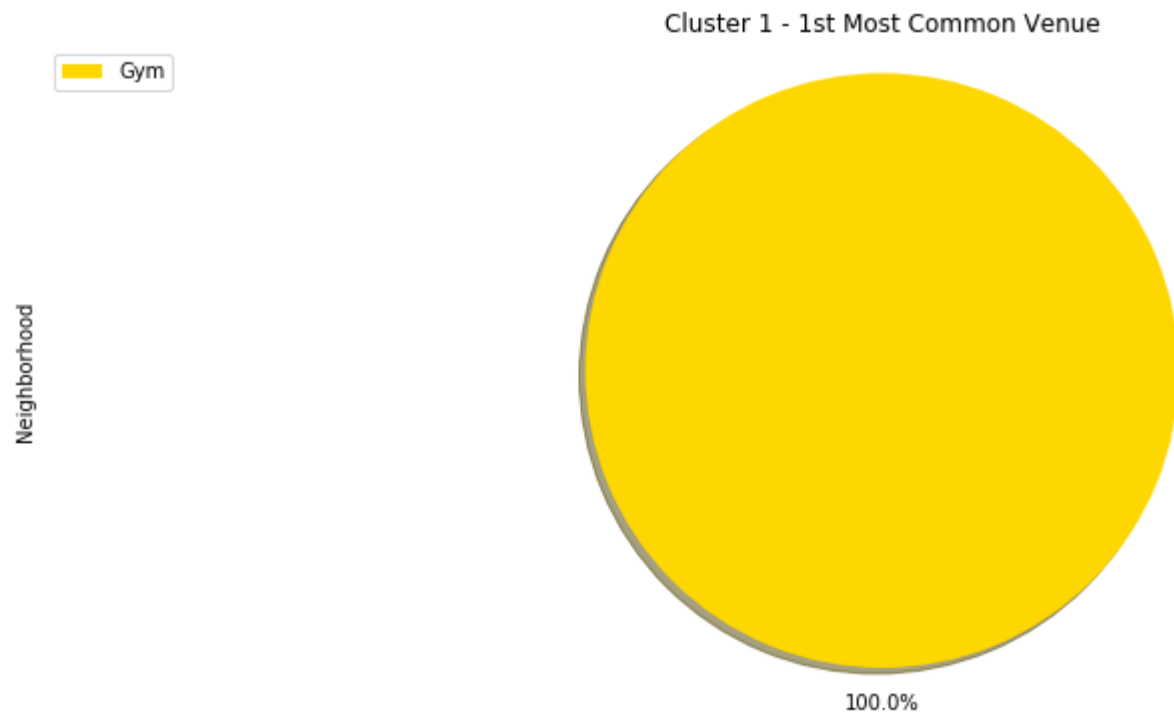- Latin American Restaurant
- Pizza Place
- Thai Restaurant

In those neighborhoods there are many venues to conglomeration people, this in a concentration of food places like restaurants, bars, pubs and beach with more than 68.000 cases confirmed and more than 5.900 deaths per 100.000 persons.
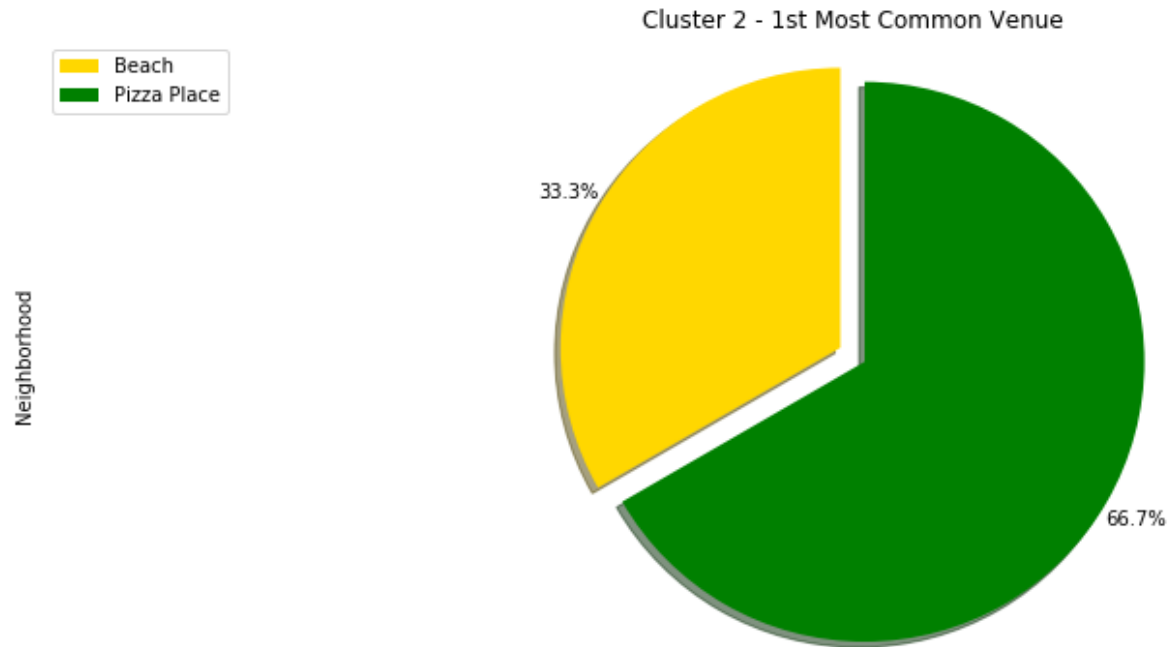
Analysis Cluster 1 – Gym/Bike Rental

Cluster 1 - 1st Most Common Venue

Gym

Neighborhood

100.0%

Only in this neighborhood are more 5.000 cases per 100.000 persons. A gym is a cabin where the people produce perspiration and the manipulations of equipment is frequently.

Analysis Cluster 2 – Beach/Food place

For the last cluster, we can identify the beach as a focal venue of infection and also the restaurants. In this cluster we have more than 11.000 cases per 100.000 persons and more than 1.100 deaths.

Conclusions and Future Directions

✓ The most important for this analysis is first to the New York State Government to do the best campaigns in the focal points infected places to social distancing, sanitizations and disinfection particularly for manipulations of any equipment.

✓ The second is for the people, these persons who live in the neighborhoods can make a prevent measures to not increment the contagious

✓ Other important point is a strong campaign of informative in these venues to prevent more infected and deaths with covid-19.

✓ I think the best vacuum for Covid-19 is the education and the prevention.