

The Battle of Neighborhoods - Finding The Right Location For a Gym

By
N. Freeman

August 9, 2021

Plan

1 Introduction

Plan

- 1 Introduction
- 2 Data and Methodology

Plan

- 1 Introduction
- 2 Data and Methodology
- 3 Results and Discussion

Plan

- 1 Introduction
- 2 Data and Methodology
- 3 Results and Discussion
- 4 Conclusion

Introduction

This project is aimed to help entrepreneurs or business owners, who want to open a Gym in New York City, better choose the right location using the Foursquare API and clustering. Neighborhoods are grouped based on the frequency of occurrence of Gyms and Burger Joints in each neighborhood.

Data and Methodology

The first step consisted in downloading the required data from [here](#) and transformed the obtained JSON file into a pandas dataframe.

Borough	Neighborhood	Latitude	Longitude
Staten Island	Tottenville	40.505334	-74.246569
Staten Island	Butler Manor	40.506082	-74.229504
Staten Island	Richmond Valley	40.519541	-74.229571
Staten Island	Pleasant Plains	40.524699	-74.219831
Staten Island	Prince's Bay	40.526264	-74.201526

Table 1: The first five rows of the initial dataframe.

Second, we used the Foursquare location data to create a new pandas dataframe

Borough	Neighborhood	Latitude	Longitude	B-J f_m	Gym f_m
Staten Island	Tottenville	40.505334	-74.246569	0	0.02
Staten Island	Butler Manor	40.506082	-74.229504	0	0.02
Staten Island	Richmond Valley	40.519541	-74.229571	0	0.02
Staten Island	Pleasant Plains	40.524699	-74.219831	0	0.02
Staten Island	Prince's Bay	40.526264	-74.201526	0.01	0.03

Table 2: The first five rows of the main dataframe (f_m = mean frequency).

Third, we used k-means clustering algorithm to group the neighborhoods into k clusters based on the neighborhoods that have similar f_m of Gyms and Burger Joints.

Neighborhood	Latitude	Longitude	B-J f_m	B-J Cluster	Gym f_m	Gym
Tottenville	40.505333	-74.246569	0	2	0.02	0
Butler Manor	40.506082	-74.229504	0	2	0.02	0
Richmond Valley	40.519541	-74.229571	0	2	0.02	0
Pleasant Plains	40.524699	-74.219831	0	2	0.02	0
Prince's Bay	40.526264	-74.201526	0.01	0	0.03	1

Table 3: The first five rows of the main dataframe after clustering.

To find the optimal k the Elbow Point Technique is used. Since the decreases in inertia 1 and inertia 2 begin to slow at $k = 4$, grouping Gyms and Burger Joints into 4 clusters is the right option.

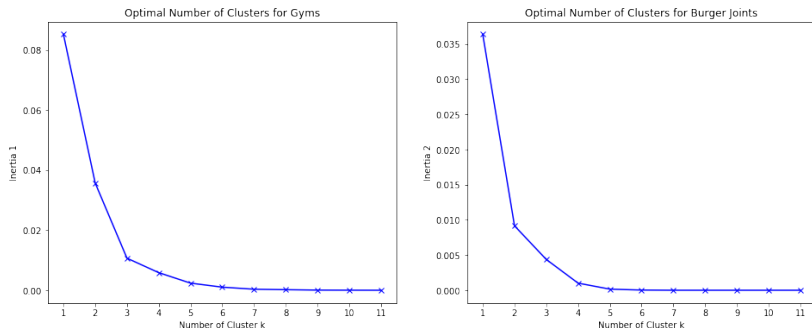


Figure 1: The Elbow Point Method using Inertia.

Finally, to visualize the neighborhoods in New York City and their emerging clusters the Folium library is used. The latitude and longitude values of New York City are obtained using geopy library.

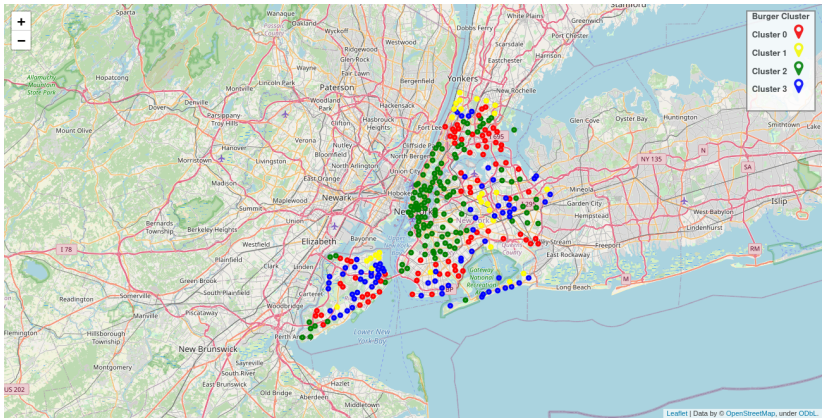


Figure 2: Burger Joint Clusters Map over New York city

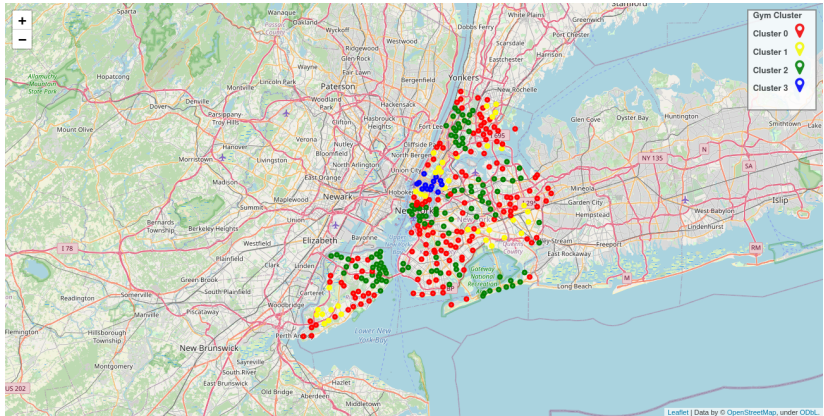


Figure 3: Gym Clusters Map over New York city

Results and Discussion

In order to have a better idea on the clusters nature, we analyzed them, step-by-step. So first, we plotted the mean frequency of each neighborhood according to each cluster containing it.

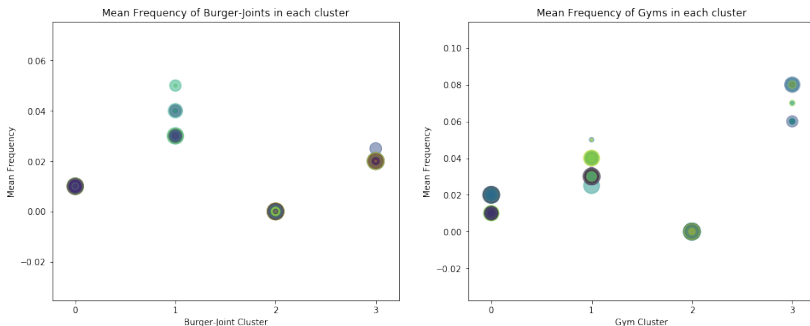


Figure 4: Mean Frequencies of Burger Joints and Gyms according to each cluster

The analysis of Figure (4), allowed us to draw the following preliminary conclusions:

- The highest mean frequencies of occurrence of Burger Joints concern neighborhoods that are contained in cluster 1.

The analysis of Figure (4), allowed us to draw the following preliminary conclusions:

- The highest mean frequencies of occurrence of Burger Joints concern neighborhoods that are contained in cluster 1.
- The highest mean frequencies of occurrence of Gyms concern neighborhoods that are contained in cluster 3.

The analysis of Figure (4), allowed us to draw the following preliminary conclusions:

- The highest mean frequencies of occurrence of Burger Joints concern neighborhoods that are contained in cluster 1.
- The highest mean frequencies of occurrence of Gyms concern neighborhoods that are contained in cluster 3.
- All neighborhoods belonging to Burger Joint Cluster 2 do not contain any Burger Joint.

The analysis of Figure (4), allowed us to draw the following preliminary conclusions:

- The highest mean frequencies of occurrence of Burger Joints concern neighborhoods that are contained in cluster 1.
- The highest mean frequencies of occurrence of Gyms concern neighborhoods that are contained in cluster 3.
- All neighborhoods belonging to Burger Joint Cluster 2 do not contain any Burger Joint.
- All neighborhoods belonging to Gym Cluster 2 do not contain any Gym.

Then, we created a dataframe including the number of neighborhoods that exist in each cluster and the average of each cluster.

B-J Cluster	Nbr of Neighborhoods	Average
0	77	0.01
1	30	0.032666
2	126	0
3	73	0.020088
Gym Cluster	Nbr of Neighborhoods	Average
0	140	0.015473
1	37	0.033728
2	116	0
3	13	0.070769

Table 4: Number of neighborhoods and averages according to each cluster.

Similarly, the following interesting conclusions are made according to the above Table (4)

- With only a minimum of neighborhoods, most of Burger Joints and most of Gyms are grouped in B-J Cluster 1 and Gym Cluster 3, respectively.

Similarly, the following interesting conclusions are made according to the above Table (4)

- With only a minimum of neighborhoods, most of Burger Joints and most of Gyms are grouped in B-J Cluster 1 and Gym Cluster 3, respectively.
- Most of the neighborhoods are grouped in B-J Cluster 2, however they do not contain any Burger Joint.

Similarly, the following interesting conclusions are made according to the above Table (4)

- With only a minimum of neighborhoods, most of Burger Joints and most of Gyms are grouped in B-J Cluster 1 and Gym Cluster 3, respectively.
- Most of the neighborhoods are grouped in B-J Cluster 2, however they do not contain any Burger Joint.
- A large part of the neighborhoods are grouped in Gym Cluster 2, however they do not contain any Gym.

Similarly, the following interesting conclusions are made according to the above Table (4)

- With only a minimum of neighborhoods, most of Burger Joints and most of Gyms are grouped in B-J Cluster 1 and Gym Cluster 3, respectively.
- Most of the neighborhoods are grouped in B-J Cluster 2, however they do not contain any Burger Joint.
- A large part of the neighborhoods are grouped in Gym Cluster 2, however they do not contain any Gym.
- Gyms are more common than Burger Joints in New York City.

Finally, we created the last dataframe that contains all common two-by-two neighborhoods between clusters.

	Gym Cluster 0	Gym Cluster 1	Gym Cluster 2	Gym Cluster 3
Burger Cluster 0	40	11	26	0
Burger Cluster 1	10	4	16	0
Burger Cluster 2	55	18	40	13
Burger Cluster 3	35	4	34	0

Table 5: All common two-by-two neighborhoods between clusters.

After analyzing Table (5), we concluded that the higher the mean of the frequency of occurrence of Burger Joints, the lower is that of Gyms, and vice versa. For example, we found that:

- Gym Cluster 3 do not contains any Burger Joint.

After analyzing Table (5), we concluded that the higher the mean of the frequency of occurrence of Burger Joints, the lower is that of Gyms, and vice versa. For example, we found that:

- Gym Cluster 3 do not contains any Burger Joint.
- Burger-Joint Cluster 1 contains only 7.14% of neighborhoods that are in Gym Cluster 0, and 10.81% of neighborhoods that are in Gym Cluster 1.

After analyzing Table (5), we concluded that the higher the mean of the frequency of occurrence of Burger Joints, the lower is that of Gyms, and vice versa. For example, we found that:

- Gym Cluster 3 do not contains any Burger Joint.
- Burger-Joint Cluster 1 contains only 7.14% of neighborhoods that are in Gym Cluster 0, and 10.81% of neighborhoods that are in Gym Cluster 1.
- Burger-Joint Cluster 2 and Gym Cluster 2 have 40 common neighborhoods that do not contain any Gym or Burger Joint.

Conclusion

In view of the previous results, we note that to have a very high success rate concerning the opening of a Gym in New York City, the 40 common neighborhoods that are in Burger-Joint Cluster 2 and Gym Cluster 2 will be a better choice as a location. This is also valid for entrepreneurs or business owners who want to open a Burger Joint in New York City. It deserves to be mentioned that our perspectives are limited by the Foursquare data availability.