

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

N. Freeman

August 9, 2021

Contents

1	Introduction	1
2	Data and Methodology	2
3	Results and Discussion	4
4	Conclusion	6

List of Figures

1	The Elbow Point Method using Inertia.	3
2	Two maps of New York city with B-J and Gym Clusters superimposed on top. . . .	4
3	Mean Frequencies of Burger Joints and Gyms according to each cluster.	5

List of Tables

1	The first five rows of the initial dataframe.	2
2	The first five rows of the main dataframe (f_m = mean frequency).	2
3	The first five rows of the main dataframe after clustering.	3
4	Number of neighborhoods and averages according to each cluster.	5
5	All common two-by-two neighborhoods between clusters.	6

1 Introduction

New York City is slowly starting to become the leading capital of the fast food nation, according to an article from [Crain's New York](#). Christian González-Rivera in his [report](#) found that the most numerous type of national retailer in the city in 2017 was fast food restaurant. Also, according to the Food Institute's analysis of data from the Bureau of Labor Statistics, the average American family spends half their food budget on restaurant food in 2018.

All of these findings show that Americans, and New Yorkers especially, love fast food so much that they cannot do without, although it tends to contain various substances that are generally unhealthful. It is high in sugar, salt, and saturated or trans fats, as well as many processed preservatives and ingredients. It is also low in beneficial nutrients. This means that people are consuming a large amount of unhealthy calories in the shape of fast food which leads to weight

gain and ultimately obesity. Of course, an occasional night of fast food will not hurt, but a habit of eating out could pose a serious danger to [human health](#).

Now obesity has become a public health problem in most nations. In New York City, more than half of adult New Yorkers have overweight (34%) or [obesity](#) (22%). People who are overweight are at increased risk for diabetes, heart disease, stroke, high blood pressure, arthritis and cancer. As a consequence, exercise is more important now than ever. It is becoming a necessity for maintaining a healthy weight or losing weight or, moreover, reducing the risk of chronic diseases. Simply walking or jogging on a daily basis can help people control their weight and improve their health, but it is not always enough. Currently, Gyms or health clubs are considered as a part of the solution for the health and wellbeing of future generations. They play an essential role in promoting physical, mental and emotional wellness.

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

2 Data and Methodology

New York City has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we needed, essentially, a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and longitude coordinates of each neighborhood. Luckily for us, all these required data can be found in a JSON file [here](#). Since the data is in a JSON format we first transformed it to 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.

Next, we used the Foursquare location data to create a new pandas dataframe that includes the name of neighborhoods within the five boroughs of the City of New York, the latitude and longitude coordinates of each neighborhood and the mean of the frequency of occurrence of both the Gyms and the Burger Joints that exist in each neighborhood.

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).

After creating our main dataframe, 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 Cluster
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 we used the Elbow Point Technique. 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.

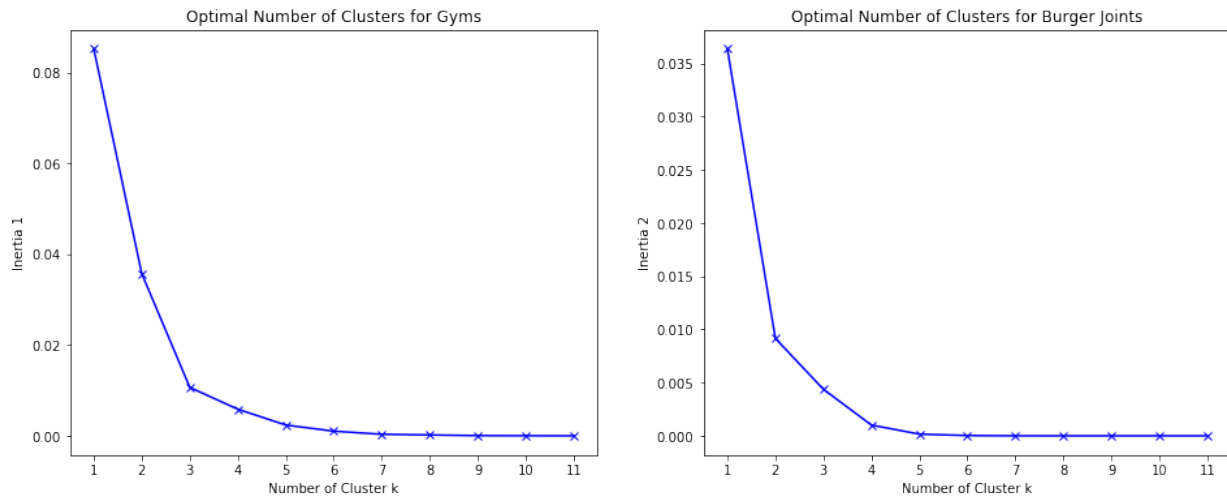
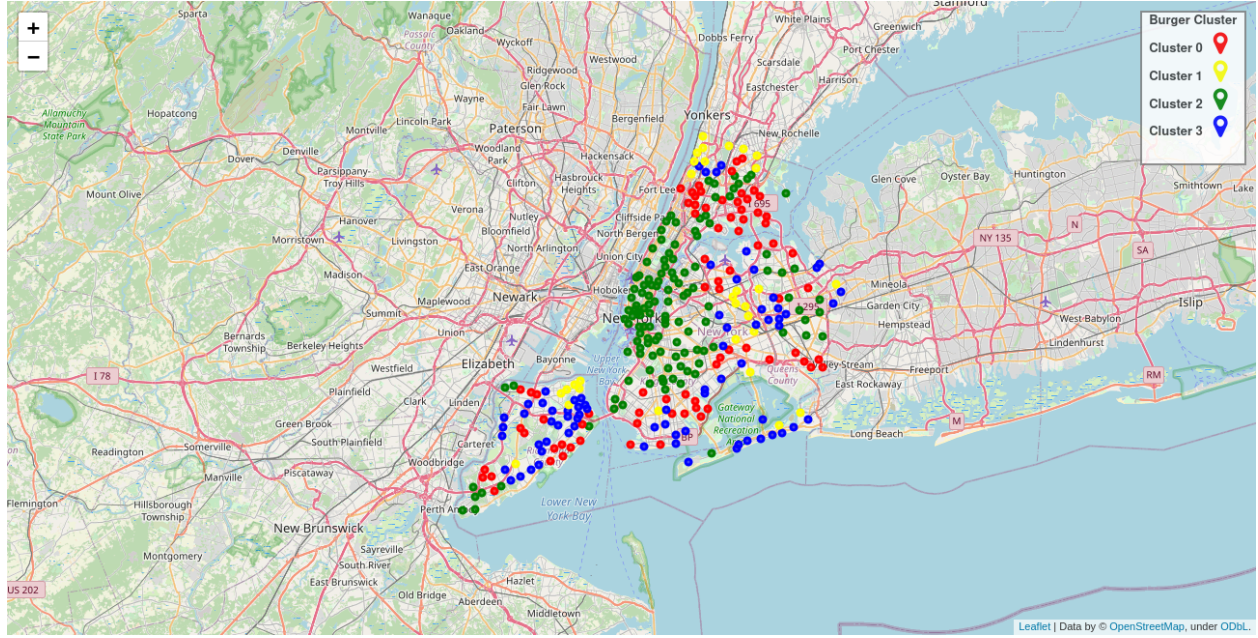
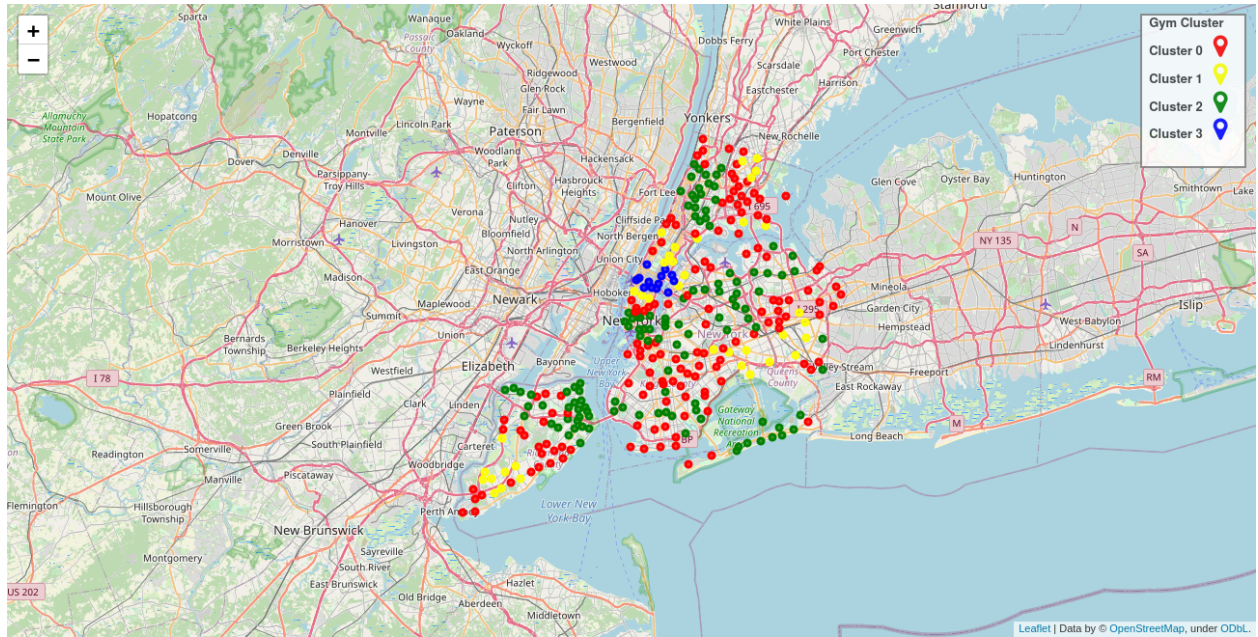


Figure 1: The Elbow Point Method using Inertia.

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.



(a) Burger Joint Clusters Map over New York city



(b) Gym Clusters Map over New York city

Figure 2: Two maps of New York city with B-J and Gym Clusters superimposed on top.

3 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.

The analysis of Figure (3), allowed us to draw the following preliminary conclusions:

1. The highest mean frequencies of occurrence of Burger Joints concern neighborhoods that are

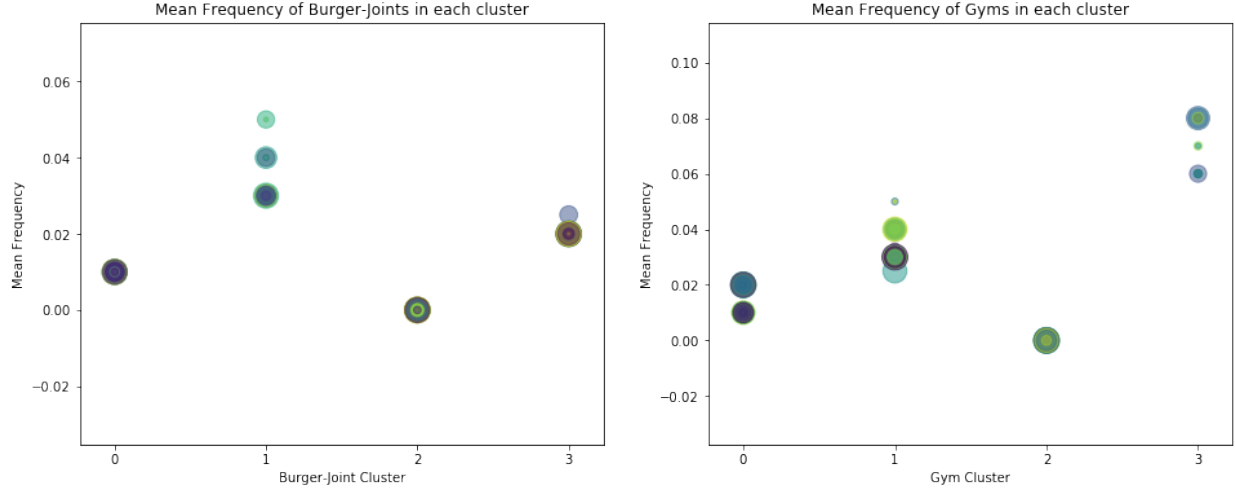


Figure 3: Mean Frequencies of Burger Joints and Gyms according to each cluster.

contained in cluster 1.

2. The highest mean frequencies of occurrence of Gyms concern neighborhoods that are contained in cluster 3.
3. All neighborhoods belonging to Burger Joint Cluster 2 do not contain any Burger Joint.
4. 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	Gym Cluster	Nbr of Neighborhoods	Average
0	77	0.01	0	140	0.015473
1	30	0.032666	1	37	0.033728
2	126	0	2	116	0
3	73	0.020088	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)

1. 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.
2. Most of the neighborhoods are grouped in B-J Cluster 2, however they do not contain any Burger Joint.
3. A large part of the neighborhoods are grouped in Gym Cluster 2, however they do not contain any Gym.
4. 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:

1. Gym Cluster 3 do not contains any Burger Joint.
2. 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.
3. Burger-Joint Cluster 2 and Gym Cluster 2 have 40 common neighborhoods that do not contain any Gym or Burger Joint.

4 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.