

## ***Battle of the Neighborhood:***

### Fitness preferences explored in NYC

## 1. Introduction

### 1.1 Background

This final project explores which neighborhood in New York City has the BEST fitness options for enthusiasts. When people move to a major city like New York it might be overwhelming to find the best neighborhood/borough with the best options for fitness venues. This is where the genesis of this project came from. Hopefully this helps make your move to New York City easier by taking one decision off your plate if you're a fitness buff or trying to start your new years resolution on the right path.

### 1.2 Problem

Using your personal fitness interest (Yoga is used as an example) to find the most venue options and best location for that fitness interest in New York City. Also note, that we are focusing solely on the density of fitness venues in NYC neighborhoods and leaving any other factors up to you to decide personally.

### 1.3 Interest

I am a fitness enthusiast and I've always wondered what each borough or neighborhood in New York City would be known for if we focused on fitness venues. I think this project would be very helpful for any fan of fitness that lives in New York and wants to find what neighborhood to move to to get the most options for their particular interest of fitness.

## 2. Data acquisition and cleaning

### 2.1 Data sources

We collected data from 3 sources. The first 2 are used to get location data (lat, long), neighborhood and borough data for New York City. 1. [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset) 2. <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm>. The third data source is pulled from the **FOURSQUARE API**, which we use to pull all fitness related venues located in New York City.

## 2.2 Data cleaning

For the New York City location data json file we pull the relevant data from the *features* key, which is a list of New York City neighborhoods. Now we create a new variable called features and a new data frame called neighborhoods, (we will create columns 'Borough', 'Neighborhood', 'Latitude', 'Longitude') this is what we use to loop the data into the neighborhoods data frame. So our neighborhoods data frame now has 5 boroughs and 306 neighborhoods. We will use this data set to do our initial analysis when exploring the neighborhoods.

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Figure 1: Final Neighborhood Data frame.

Next we pull from the FourSquare API, the venue data we will need to do the next phase in our analysis. We create a function to call to the API the top 100 venues within a 500 mile radius and loop the data into a data frame we created called nyc\_venues. The final data frame shown below. Next we check all the different venue categories that exist in nyc\_venues and only pull the fitness venues from all the venues in the FourSquare API. The fitness venues chosen are: 'Gym', 'Yoga Studio', 'Sports Club', 'Cycle Studio', 'Gym / Fitness Center', 'Recreation Center', 'Outdoors & Recreation', 'Athletics & Sports', 'Pilates Studio', 'Gym Pool', 'Pool Hall', 'Gymnastics Gym', 'Martial Arts Dojo', 'Golf Course', 'Boxing Gym' and 'Indoor Play.' We name this data frame nyc\_fitness and we use it to do our fitness venue analysis. The final data frame shown below.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue_Category
62	40.890834	-73.912585	Hayden On Hudson Gym	40.889593	-73.917446	Gym
124	40.881687	-73.902818	East Coast United Brazilian Jiu Jitsu , MMA & ...	40.883792	-73.901573	Martial Arts Dojo
134	40.876551	-73.910660	Bikram Yoga	40.876844	-73.906204	Yoga Studio
142	40.876551	-73.910660	Astral Fitness & Wellness Center	40.876705	-73.906372	Gym
143	40.876551	-73.910660	Blink Fitness	40.877271	-73.905595	Gym

Figure 2: nyc\_fitness Data Frame.

Last, we create a data frame to just run our analysis for the Brooklyn borough (last part of our analysis of Yoga venues). We create a subset of the neighborhoods data frame and just pull the Brooklyn neighborhoods which is 70 total. We combine the new brooklyn\_data with the fitness\_venue data and after we cluster the data we have to remove the NA's as some of the neighborhoods in Brooklyn don't have fitness venues.

	Borough	Neighborhood	Latitude	Longitude	Neighborhood_Latitude	Neighborhood_Longitude	Venue	Venue_Latitude	Venue_Longitude	Venue_Category
0	Brooklyn	Bay Ridge	40.625801	-74.030621	40.625801	-74.030621	New York Sports Clubs	40.622364	-74.027163	Gym / Fitness Center
1	Brooklyn	Bay Ridge	40.625801	-74.030621	40.625801	-74.030621	Status Q	40.625906	-74.030192	Pool Hall
2	Brooklyn	Bay Ridge	40.625801	-74.030621	40.625801	-74.030621	Dahn Yoga	40.626273	-74.030208	Yoga Studio
3	Brooklyn	Bay Ridge	40.625801	-74.030621	40.625801	-74.030621	Bay Ridge Crossfit	40.624143	-74.030823	Gym / Fitness Center
4	Brooklyn	Sunset Park	40.645103	-74.010316	40.645103	-74.010316	Blink Fitness	40.645502	-74.013143	Gym

Figure 3: brooklyn\_merged Data Frame

### 2.3 Feature selection

We focus on the following features/variables in our analysis. For the first part of our analysis where we focused on all of New York City it is Borough and Venue. For the second part of our analysis where we focus on Brooklyn it is Borough, Venue\_Category and Neighborhood.

## 3. Methodology & Data Analysis

### 3.1 NYC Neighborhood Analysis

For the neighborhood analysis, we needed to see how each borough broke down and where each was in relation to New York City. So we graphed how many neighborhoods were in each borough (see below) and found that Queens had the most neighborhoods with more than 80.

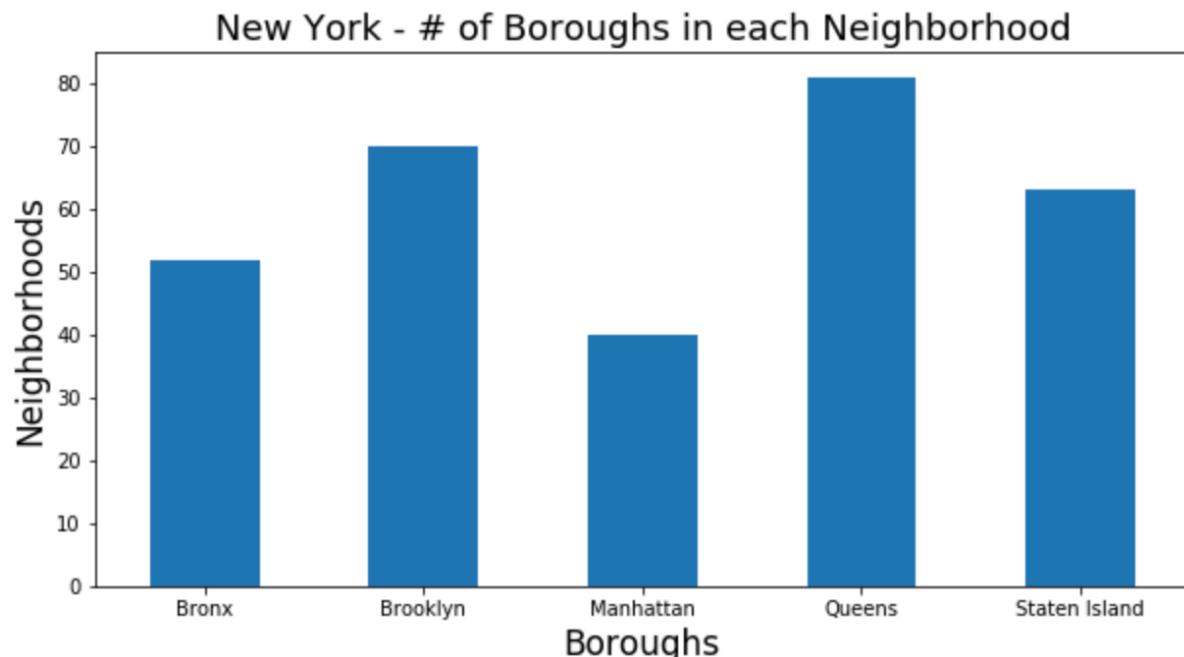


Figure 3: Bar Chart of Number of Neighborhoods in each New York City Borough

So we used folium (a python package) to graph all the neighborhoods in Queens. We use Queens as an example, since it has the most neighborhoods in New York, shown below.

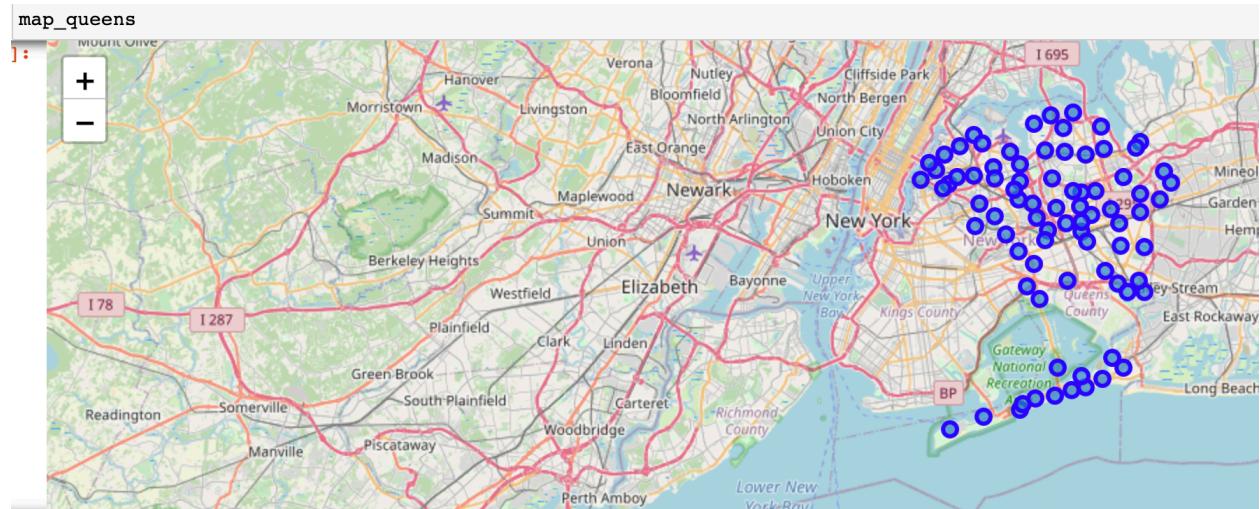


Figure 4: Map of Queens Neighborhoods

### 3.2 FourSquare Analysis of New York City Venues and Neighborhoods

Next we pull all venues from the FourSquare API and found that there were over 9000 venues and 423 were fitness venues. So that is about 4.25% of the total venues are fitness venues and on average about 14% of the venues in each neighborhood were fitness venues.

```
Total number of venues: 9953
Total number of fitness venues: 423
Percentage of fitness venues: 4.25%
Average number of fitness venues in each neighborhood: 14.714285714285714
```

Figure 5: List of Venue Totals and Percentages.

Next we checked to see how many fitness venues are in each borough. Manhattan was first with 193 fitness venues, second was Brooklyn with 111, third was Queens with 79, fourth was Bronx with 28 and last was Staten Island with 26. Shown below.

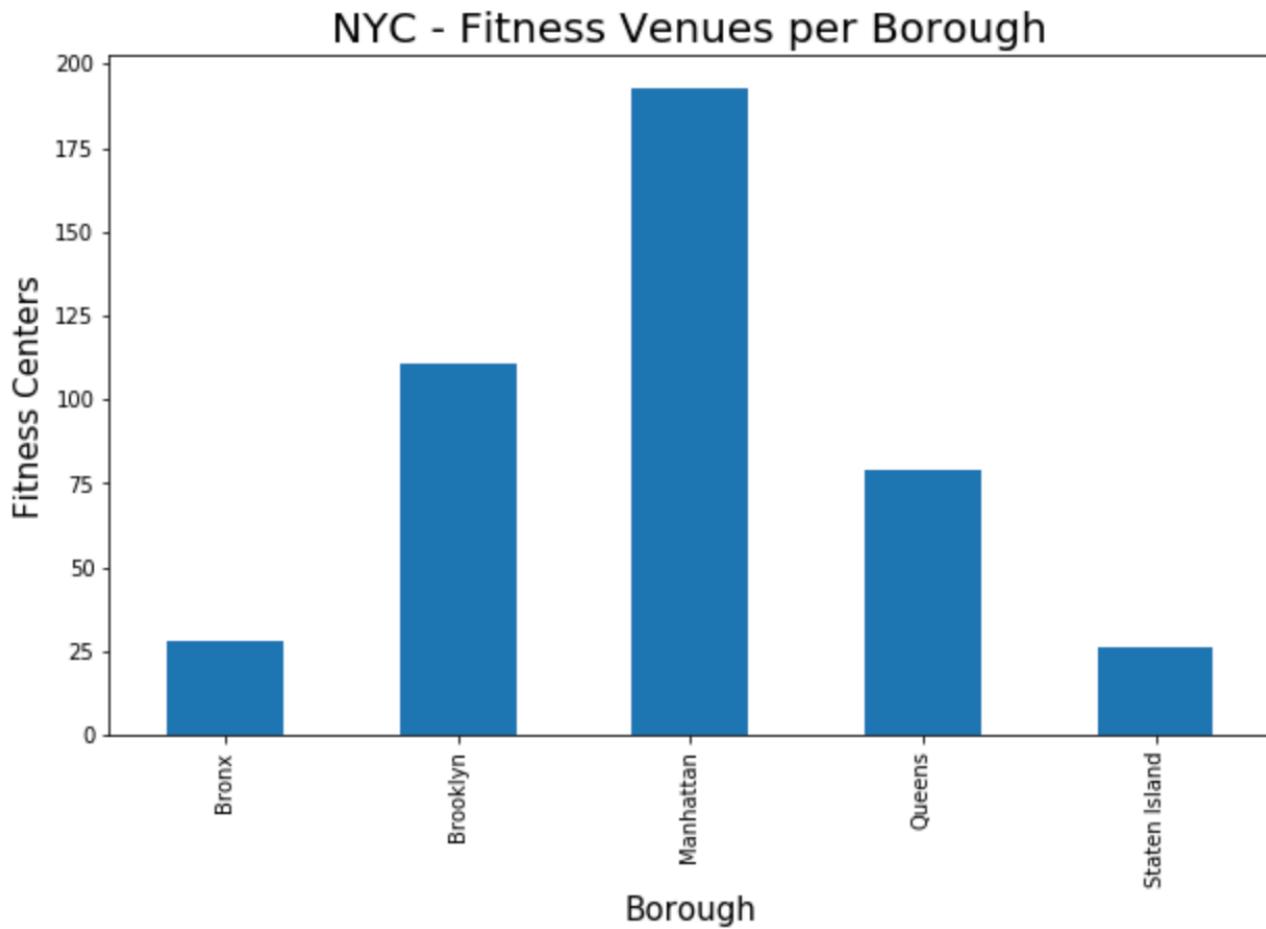


Figure 6: Bar Chart of Fitness Venues per Borough

Next we will analyze the frequency of fitness centers in each borough by creating dummy variables for each Venue\_Category as it pertains to each borough. There are a total of 19 fitness venue features we will analyzing going forward. The top 5 frequency of fitness venue is shown below. For the Bronx, Gym was the most frequent. For Brooklyn, Yoga was the most frequent. For Manhattan, Queens and Staten Island, Gym/Fitness Center was the most frequent.

----Bronx----		----Manhattan----		----Staten Island----	
	venue freq		venue freq		venue freq
0	Gym 0.50	0	Gym / Fitness Center 0.31	0	Gym / Fitness Center 0.23
1	Gym / Fitness Center 0.29	1	Gym 0.28	1	Gym 0.19
2	Martial Arts Dojo 0.14	2	Yoga Studio 0.17	2	Athletics & Sports 0.15
3	Athletics & Sports 0.07	3	Cycle Studio 0.08	3	Martial Arts Dojo 0.08
4	Golf Course 0.00	4	Boxing Gym 0.04	4	Cycle Studio 0.08

----Brooklyn----		----Queens----	
	venue freq		venue freq
0	Yoga Studio 0.31	0	Gym / Fitness Center 0.38
1	Gym / Fitness Center 0.23	1	Gym 0.29
2	Gym 0.16	2	Yoga Studio 0.10
3	Boxing Gym 0.06	3	Martial Arts Dojo 0.08
4	Cycle Studio 0.05	4	Athletics & Sports 0.05

Figure 7: Top 5 most frequent fitness venues per Borough

Please note, this clearly shows how we will cluster our boroughs (a total of 3 clusters) in the next step in our analysis outlined below.

Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bronx	Gym	Gym / Fitness Center	Martial Arts Dojo	Athletics & Sports
1	Brooklyn	Yoga Studio	Gym / Fitness Center	Gym	Boxing Gym
2	Manhattan	Gym / Fitness Center	Gym	Yoga Studio	Cycle Studio
3	Queens	Gym / Fitness Center	Gym	Yoga Studio	Martial Arts Dojo
4	Staten Island	Gym / Fitness Center	Gym	Athletics & Sports	Cycle Studio
					Yoga Studio

Figure 8: Top 5 most frequent fitness venue categories per Borough

Last, we will use both of our data frames (neighborhoods and nyc\_fitness), to create a clustering of all the fitness facilities in NYC. We will see first, which borough has the most fitness venues and second, we will choose a specific type of fitness venue and see how that breaks down on a neighborhood by neighborhood level. We chose a k of 3 because we can clearly see that there will be 3 clusters based on the frequency of venues in our breakdown above. We need, one for the Bronx, one for Brooklyn, and one for Manhattan, Staten Island and Queens. We used folium to map the 3 clusters (shown below). Cluster 0 is labeled as Bronx: RED & Gym/Martial Arts. Cluster 1 is labeled as Brooklyn: PURPLE & Yoga. Cluster 2 is labeled as Manhattan, Queens and Staten Island: LIGHT GREEN & Gym/Fitness centers.

Table 1: Breakdown of New York City Clusters

CLUSTERS	BOROUGH	CLUSTER COLOR	LABEL	NEIGHBORHOOD TOTALS
Cluster 0	Bronx	RED	Gym/Martial Arts	52
Cluster 1	Brooklyn	PURPLE	Yoga	70
Cluster 2	Manhattan, Queens, Staten Island	LIGHT GREEN	Gym/Fitness Centers	184

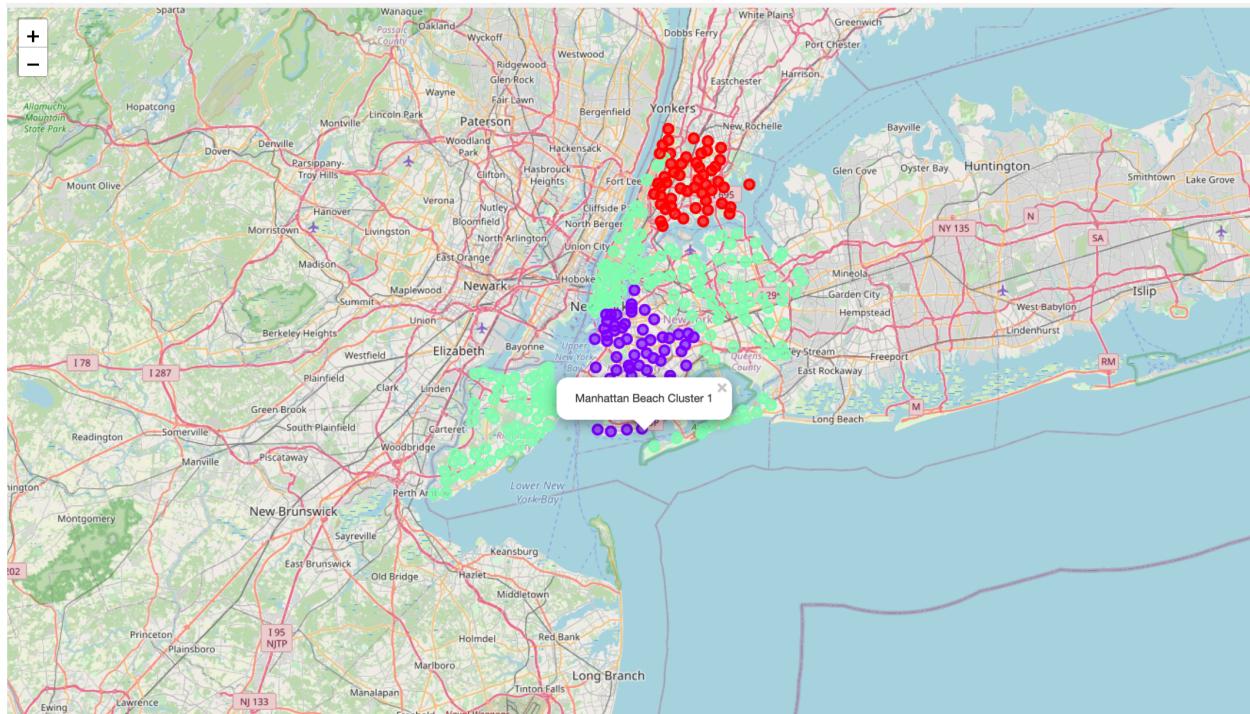


Figure 9: Cluster Map by Borough

### 3.3 FourSquare Analysis of Brooklyn Neighborhoods

The Brooklyn neighborhood with the most fitness venues is Brooklyn Heights with 12. The North Side has the third most with 9 which is the neighborhood we choose in our final analysis for Yoga and Cycling centric neighborhoods. Please note that there are a total of 70 neighborhoods in Brooklyn and some don't have fitness venues in them. Reducing the cluster neighborhoods in our analysis down to 37. The top 25 fitness venues per Brooklyn neighborhood is shown below.

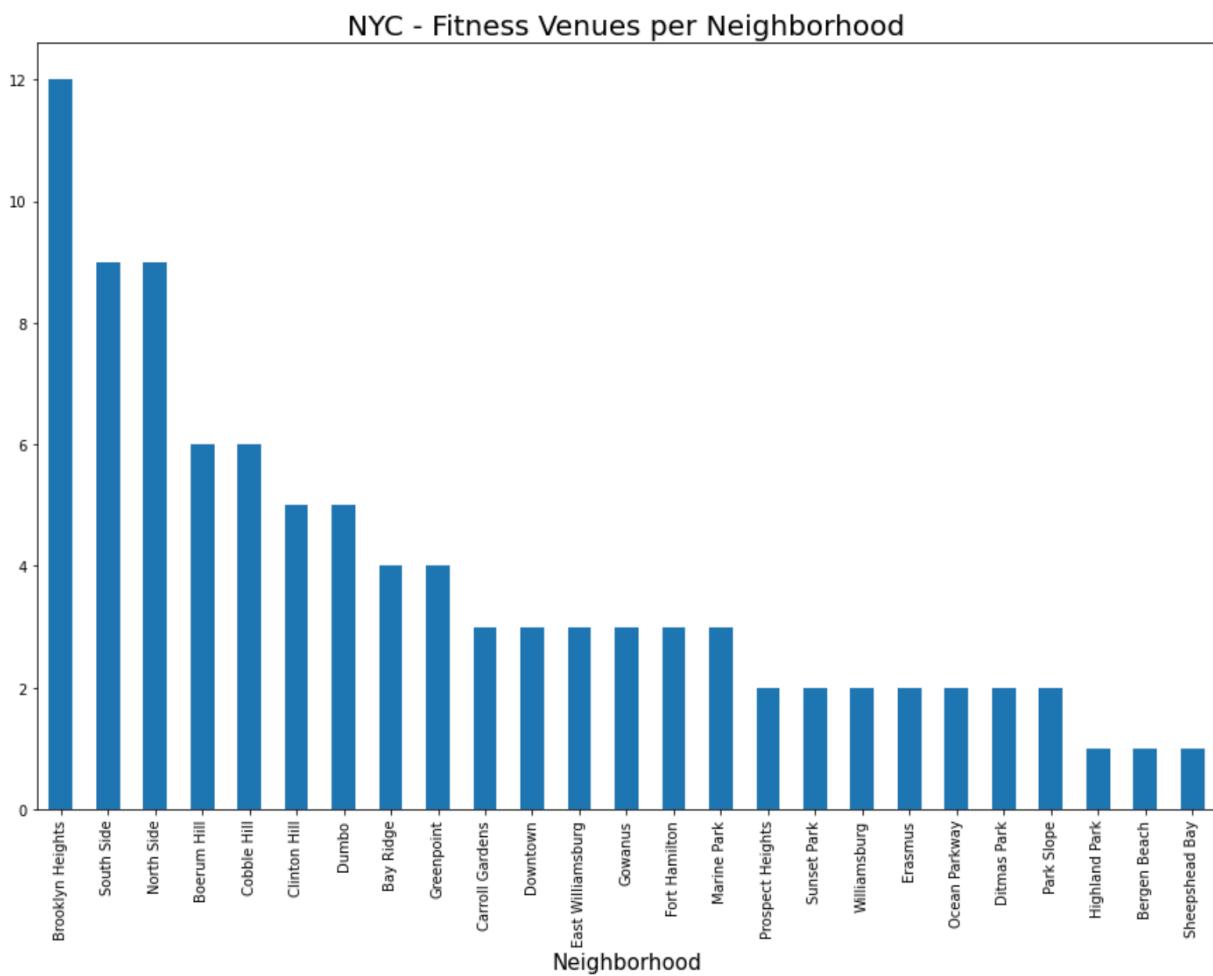


Figure 10: Bar Chart of Brooklyn Fitness Venues per Neighborhood

### 3.4 Clustering Analysis of Brooklyn Neighborhoods

Next we chose a specific type of fitness venue for our next set of analysis, which was Yoga. Yoga has the highest density in Brooklyn so we choose that borough for the last part of our analysis.

We chose 4 clusters for our analysis because 5 clusters gives us a blank cluster during our initial analysis. After we run our analysis ,we choose *CLUSTER 1* as the neighborhoods we would move to as it has the highest density of Yoga venues among the Brooklyn neighborhoods. There are a total of 11 choices to choose from but if we chose another fitness type in cluster 1 like cycling then we can narrow our search down to one neighborhood, The North Side (shown below in the map). Some of the neighborhoods were removed because they didn't have enough information for our analysis.

Table 2: Breakdown Brooklyn Clusters

CLUSTERS	BOROUGH	CLUSTER COLOR	LABEL	NEIGHBORHOOD TOTALS
Cluster 0	Brooklyn	RED	Balanced	11
Cluster 1	Brooklyn	PURPLE	Yoga	11
Cluster 2	Brooklyn	BLUE	Gym/Fitness Centers	8
Cluster 3	Brooklyn	LIGHT GREEN	Gym	7

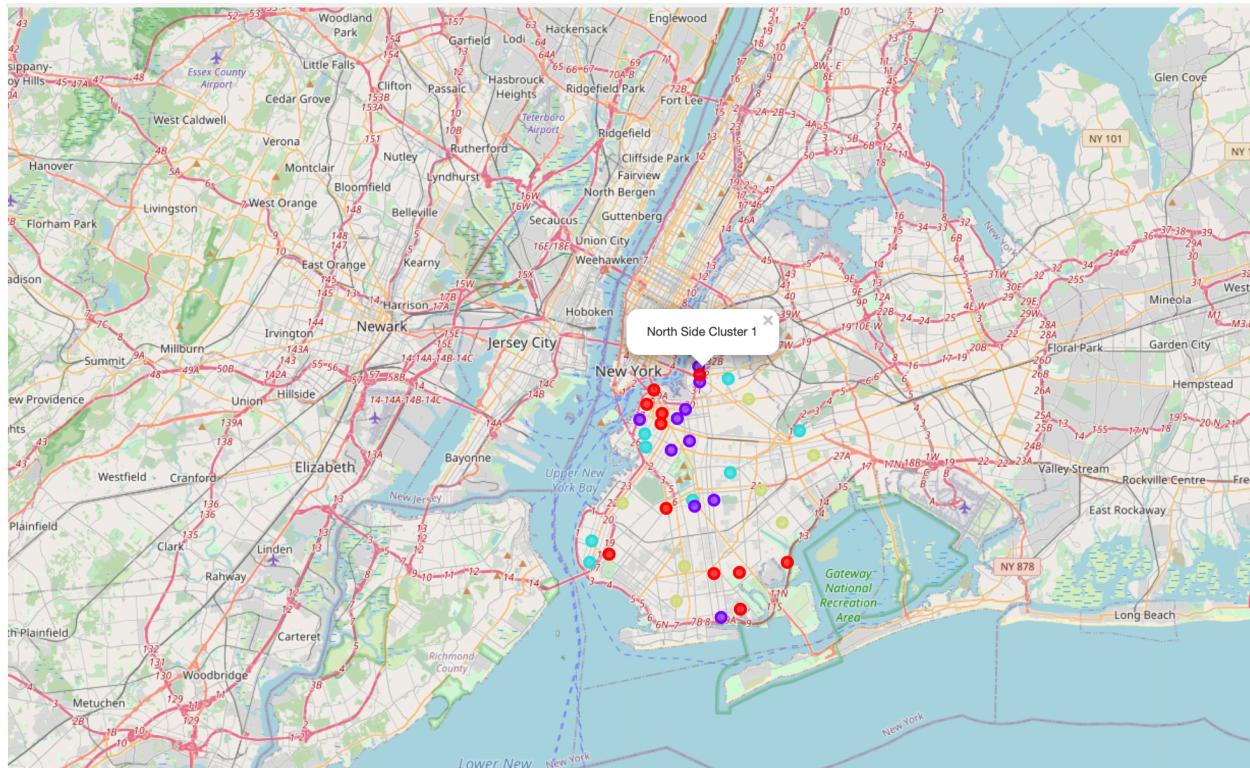


Figure 11: Cluster Map of Brooklyn Neighborhoods

#### 4. Conclusions & Results

In this study, our decision comes down to first, which Borough you want to live in based on your fitness preferences. Then second, which neighborhood in that borough you want to live in based on the number of fitness venue options. For the Borough analysis, the highest density of fitness venues from most to least were: Queens, Brooklyn, Staten Island, Bronx and Manhattan. The clusters broke down as follows, Cluster 0 had 184 total fitness venues (composed of mostly Gym & Fitness) and included Boroughs: Queens, Staten Island and Manhattan. Cluster 1 had 70 total fitness venues (composed of mostly Yoga) and included Brooklyn. Cluster 2 had 52 total fitness venues (composed of mostly Gym & Martial Arts) and included the Bronx. Please note that of the 70 neighborhoods in Brooklyn only 37 of 70 had fitness venues in them so we can eliminate 33 of the neighborhoods right away. So based on those results you have to decide what type of fitness centers you are mostly looking for and make your decision/pivot on that decision. Next we analyzed Brooklyn, to see how Yoga studios density continue to breakdown by neighborhood. We created 4 clusters and the one with the most yoga studios was cluster 1, which included 11 neighborhoods to chose from. If you have a second choice, you can further narrow down your specific neighborhood choice (like shown above with yoga and cycling, with the best choice being the North Side). So for a quick analysis this is an good way to see where you should live based on fitness venue density, but if you want a more comprehensive analysis on where to move, you will need to look into other factors like where you work and your other interest like nightlife etc.

As you can see this is a great start for finding out which neighborhood to move to in New York City based on your fitness preferences. Also note, all of the above analysis is dependent on the accuracy of the FourSquare data.