

# A Peek into New York's Fitness Preferences

- A data analysis report

<b><i>Introduction.....</i></b>	<b><i>2</i></b>
<b><i>Data.....</i></b>	<b><i>2</i></b>
<b><i>Methodology – Exploratory Analysis.....</i></b>	<b><i>3</i></b>
<b><i>Availability .....</i></b>	<b><i>7</i></b>
<b><i>Fitness vs Food Joints .....</i></b>	<b><i>9</i></b>
<b><i>Results .....</i></b>	<b><i>12</i></b>
<b><i>Clustering .....</i></b>	<b><i>12</i></b>
<b><i>Discussion .....</i></b>	<b><i>15</i></b>
<b><i>Conclusion .....</i></b>	<b><i>16</i></b>

## Introduction

The purpose of this project is to analyze the location data related to physical health and fitness related venues like gyms, fitness centers, sports arena and others. I would like to see how a model city like New York, which is the center of intense economic activity and thereby home to majority population following sedentary lifestyle, places its fitness industry. This report would give basic insights about the spread of different types of fitness venues across the five Boroughs of New York city. Also this can be a starting point for deeper analysis leading up to building and developing health & fitness as a viable business opportunity in any of the neighborhoods of New York.

## Data

I am using the data sources provided in the previous weeks of the Coursera Capstone Project course which includes,  
New York location data

Source URL - [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)  
File - newyork\_data.json

Also the location data of venues of all the neighborhoods in each of the 5 Boroughs, namely Manhattan, The Bronx, Brooklyn, Queens and Staten Island, is extracted using separate API calls to *Foursquare* database.

API call URL-  
[https://api.foursquare.com/v2/venues/explore?oauth\\_token=<my\\_access\\_token>&v=20180301&ll=<latitude>,<Longitude>&radius=500&limit=100](https://api.foursquare.com/v2/venues/explore?oauth_token=<my_access_token>&v=20180301&ll=<latitude>,<Longitude>&radius=500&limit=100)

Extracted data includes a venues' Name, Venue Category, Latitude and Longitude. The number of venues on which the above data points are extracted as below into separate datasets for each borough.

Manhattan – 2332  
Bronx – 2679  
Brooklyn - 3496  
Queens - 2138  
Staten Island – 3998

The venues that are related to the fitness and health activity are separated out from these datasets based on the venue categories - 'Gym', 'Gym / Fitness Center', 'Climbing Gym', 'Gymnastics Gym', 'Boxing Gym', 'Pilates Studio', 'Martial Arts Dojo', 'Physical Therapist', 'College Gym', 'Weight

Loss Center', 'Cycle Studio', 'Yoga Studio', 'Tennis Stadium', 'Sports Club', 'Athletics & Sports', 'Tennis Court', 'Golf Course', 'Volleyball Court', 'Mini Golf', 'Basketball Court', 'Soccer Field', 'Baseball Field', 'Soccer Field', 'Baseball Stadium', 'Golf Course', 'Stadium', 'Squash Court', 'Hockey Field', 'College Basketball Court', 'College Stadium', 'Pool', 'Gym Pool', 'Pool Hall', 'Bike Trail', 'Track', 'Trail', 'Skate Park', 'Skating Rink', 'Surf Spot', 'Ski Area', 'Roller Rink', 'Park', 'Playground', 'Dance Studio', 'Bowling Alley', 'Indoor Play Area', 'Outdoors & Recreation', 'Rock Climbing Spot', 'Other Great Outdoors'. This data is stored in separate datasets namely <Borough>\_fitness\_venues for further analysis.

Borough	Fitness Venues
Manhattan	344
The Bronx	141
Brooklyn	272
Queens	149
Staten Island	216

## Methodology – Exploratory Analysis

The data is grouped based on the ‘Venue Category’ to identify the most common location for each neighborhood. This is done for all the 5 boroughs,

### Manhattan:

1st Most Common Venue	Neighborhood
Park	14
Gym / Fitness Center	10
Yoga Studio	8
Gym	5
Boxing Gym	1
Cycle Studio	1
Dance Studio	1

### The Bronx:

1st Most Common Venue	Neighborhood
Park	9
Athletics & Sports	6

1st Most Common Venue	Neighborhood
Baseball Field	6
Gym	5
Bowling Alley	2
Pool	1
Trail	1
Track	1
Sports Club	1
Soccer Field	1
Outdoors & Recreation	1
Playground	1
Other Great Outdoors	1
Martial Arts Dojo	1
Gym / Fitness Center	1
Basketball Court	1
Weight Loss Center	1

*Brooklyn:*

1st Most Common Venue	Neighborhood
Park	8
Yoga Studio	7
Dance Studio	5
Gym / Fitness Center	5
Gym	4
Athletics & Sports	2
Baseball Field	2
Playground	2
Baseball Stadium	1
Golf Course	1
Martial Arts Dojo	1
Other Great Outdoors	1
Pilates Studio	1

*Queens:*

<b>1st Most Common Venue</b>	<b>Neighborhood</b>
Park	7
Gym	6
Gym / Fitness Center	6
Playground	4
Yoga Studio	4
Martial Arts Dojo	2
Trail	2
Athletics & Sports	1
Baseball Field	1
Basketball Court	1
Bowling Alley	1
College Basketball Court	1
Pool	1
Skate Park	1
Skating Rink	1
Surf Spot	1

*Staten Island:*

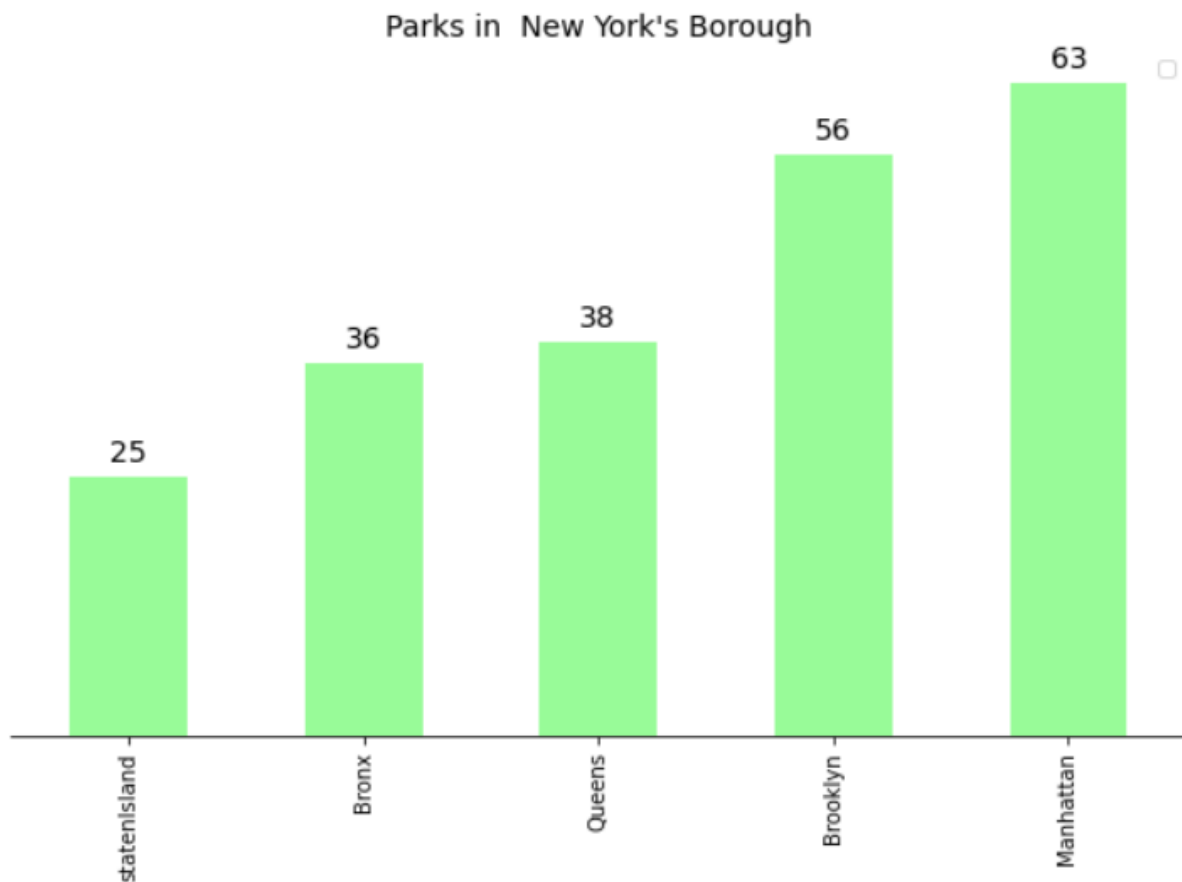
<b>1st Most Common Venue</b>	<b>Neighborhood</b>
Dance Studio	7
Park	6
Athletics & Sports	5
Martial Arts Dojo	4
Baseball Field	3
Gym / Fitness Center	3
Playground	2
Yoga Studio	2
Basketball Court	1
Gym	1
Mini Golf	1
Outdoors & Recreation	1
Physical Therapist	1

1st Most Common Venue	Neighborhood
Skating Rink	1
Soccer Field	1
Trail	1

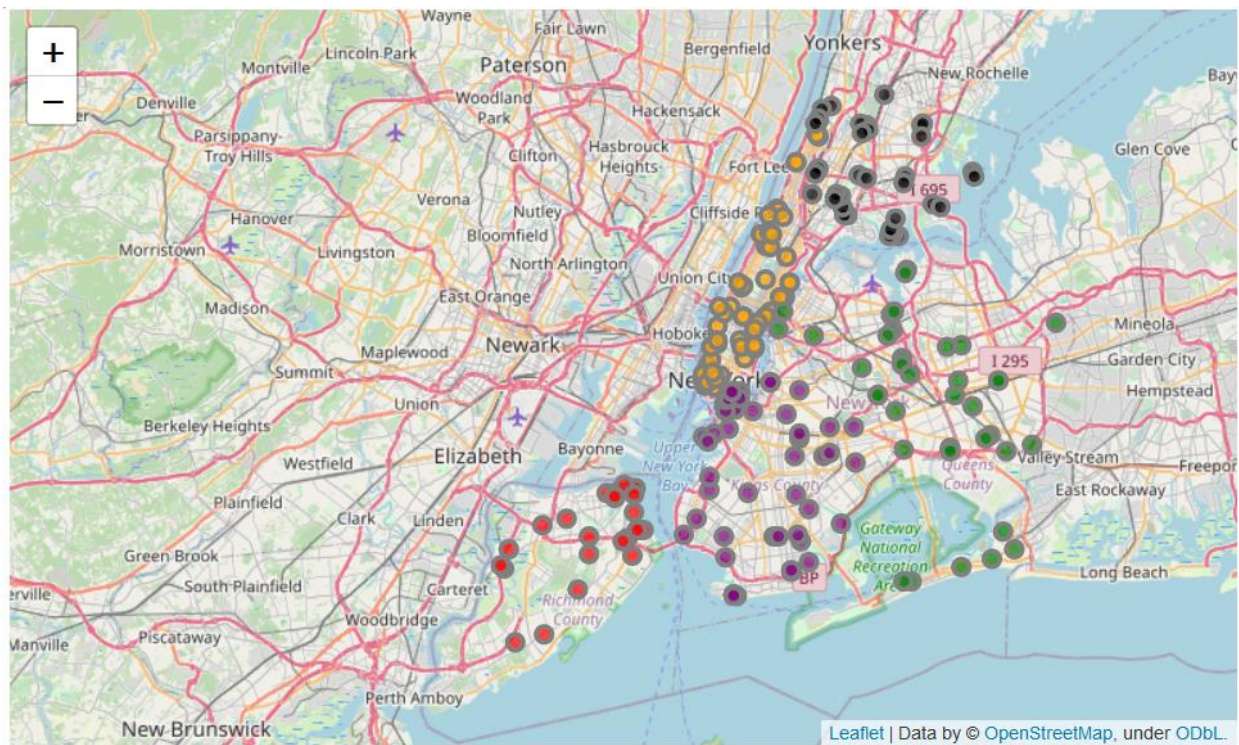
The above tables show the number of neighborhoods that favor each type of venues. It is apparent that most of Neighborhoods in all the 5 Boroughs have Parks as the most common location.

This is understandable as parks are more of a common location and is not specific to any or for only fitness activities. But these parks are included in the analysis since they play an important role for the population in any of the neighborhoods to maintain their physical as well as mental well-being.

In New York, there are around 208 parks spread across the 5 boroughs as below,



Now that we have the count of the available parks, their spread across New York is shown with the help of **Folium** library to plot the location data of these parks



Now that we have looked at the availability and importance of parks, let us get these out of our collection of fitness venues and proceed with further analysis.

## Availability

The venues are grouped as below into broader categories to help get a bird's eye view of their preferences

```
gym_cat = ['Gym', 'Gym / Fitness Center', 'Climbing Gym', 'Gymnastics Gym',  
'Boxing Gym', 'Pilates Studio', 'Martial Arts Dojo', 'Physical Therapist',  
'College Gym', 'Weight Loss Center', 'Cycle Studio']
```

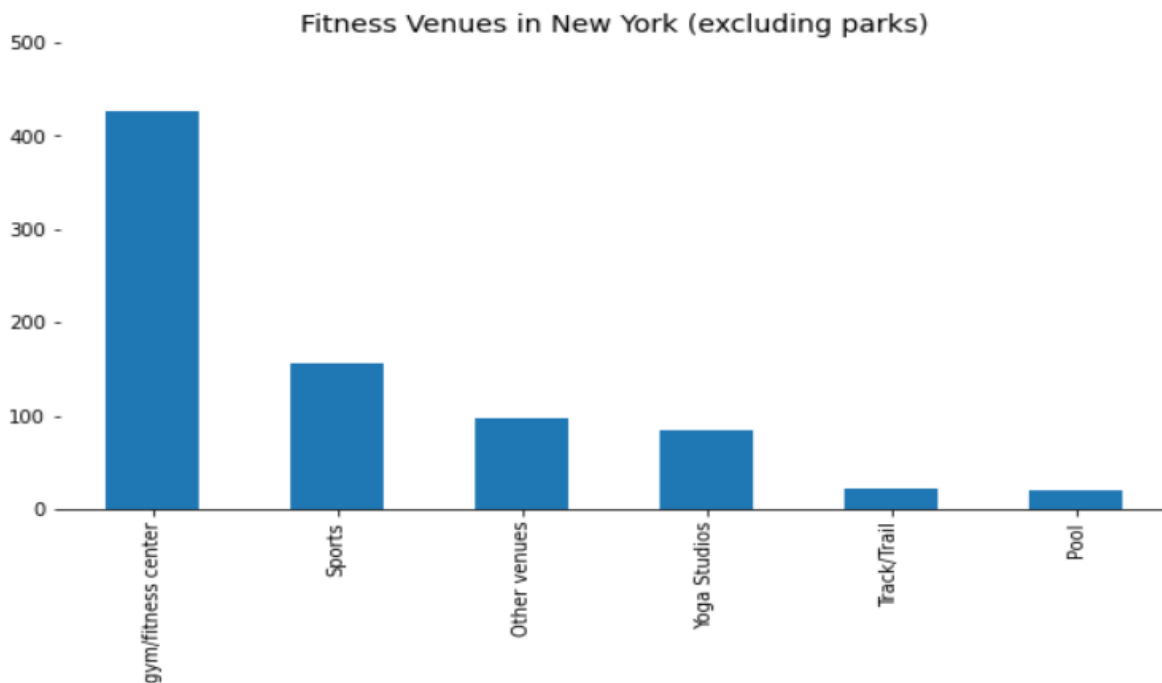
```
sports_cat = ['Tennis Stadium', 'Sports Club', 'Athletics & Sports', 'Tennis  
Court', 'Golf Course', 'Volleyball Court', 'Mini Golf', 'Basketball Court',  
'Soccer Field', 'Baseball Field', 'Soccer Field', 'Baseball Stadium', 'Golf  
Course', 'Stadium',  
              'Squash Court', 'Hockey Field', 'College Basketball Court',  
'College Stadium']
```

```
swim_cat = ['Pool', 'Gym Pool', 'Pool Hall']
```

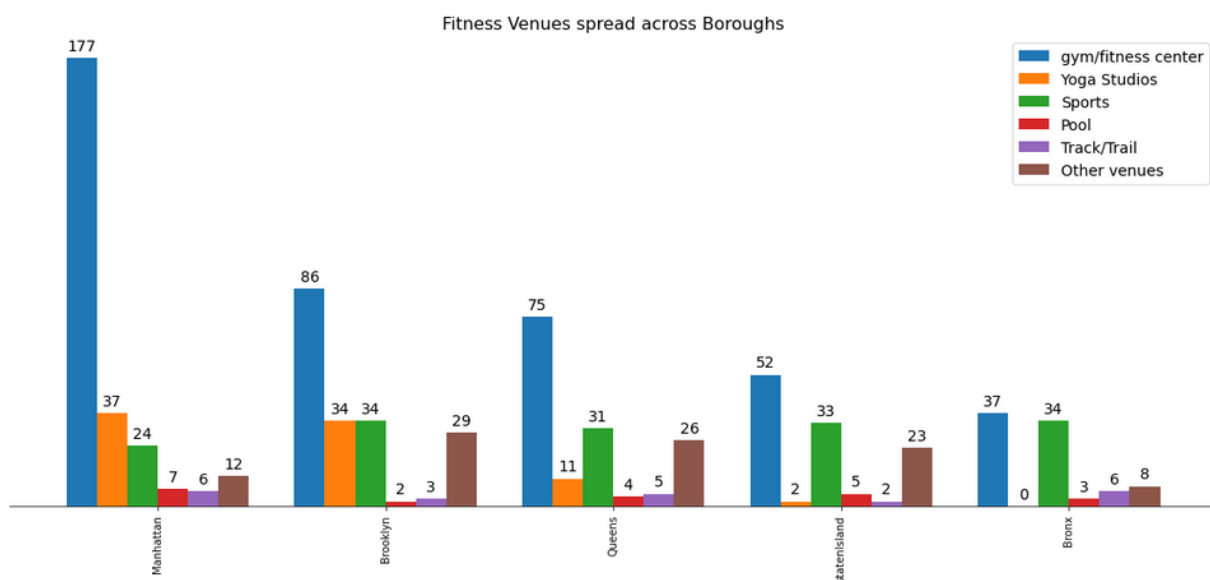
```
TT_cat = ['Bike Trail', 'Track', 'Trail']
```

```
Others = ['Skate Park', 'Skating Rink', 'Surf Spot', 'Ski Area', 'Roller Rink', 'Dance Studio', 'Bowling Alley', 'Indoor Play Area', 'Outdoors & Recreation', 'Rock Climbing Spot', 'Other Great Outdoors']
```

Based on these broad categories, we get the below numbers for each borough

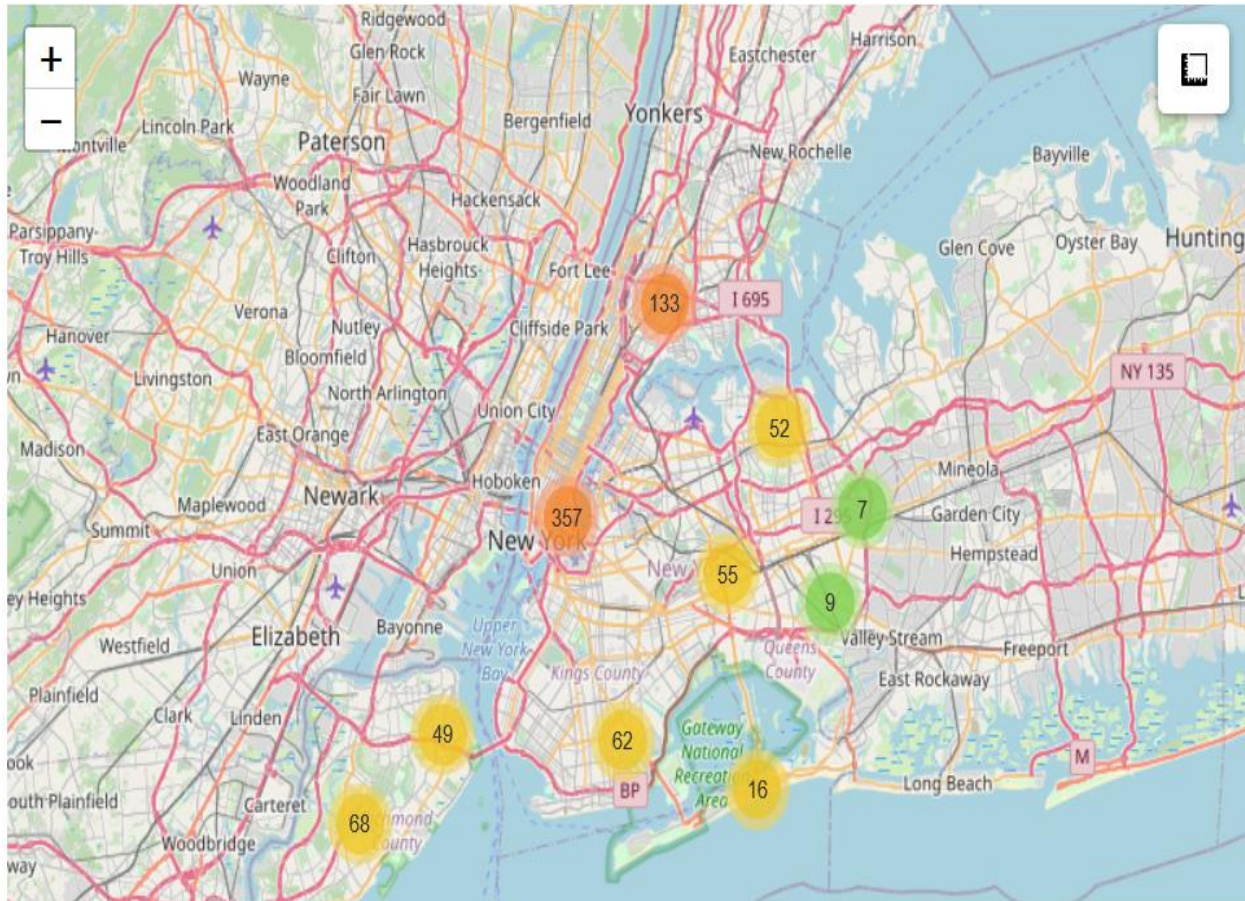


A further deep dive into this classification gives the below chart





Mapping these location would give us deeper insight as to whether these venues are spread evenly across all neighborhoods or if there any specific locations where they are in abundance/scarce.

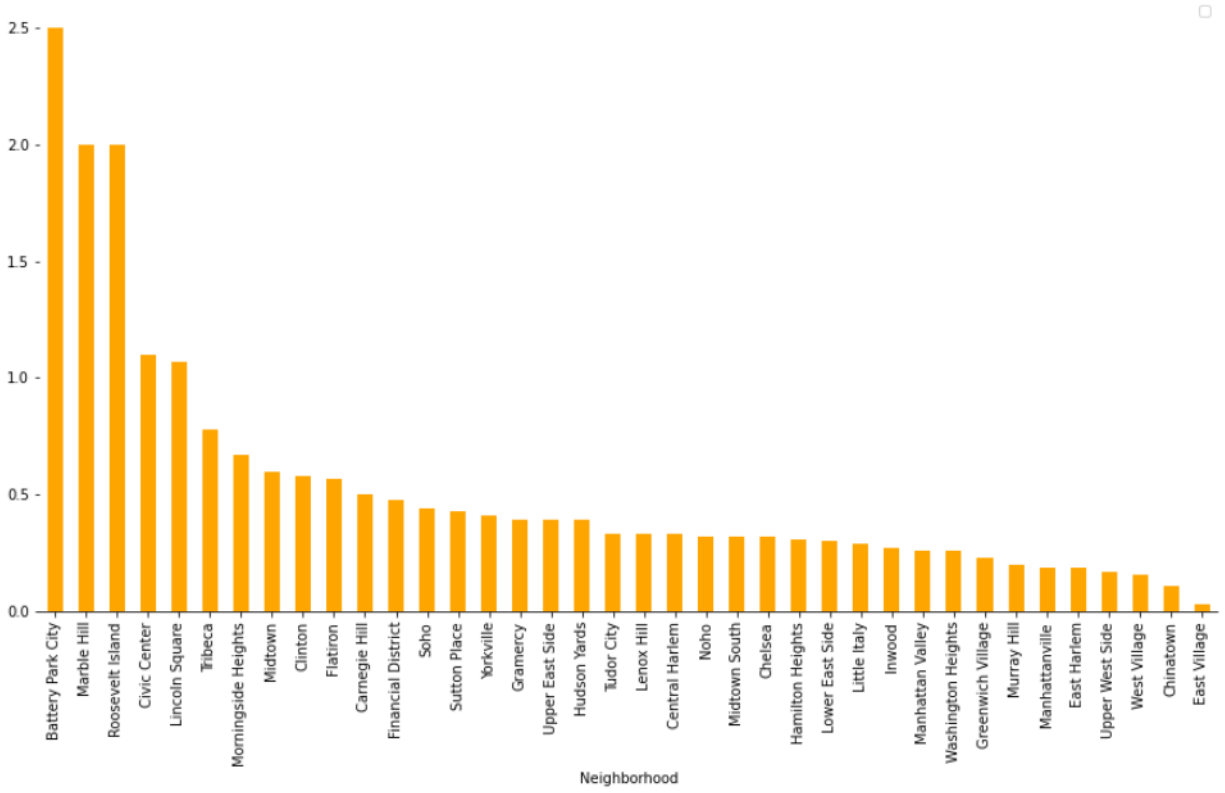


## Fitness vs Food Joints

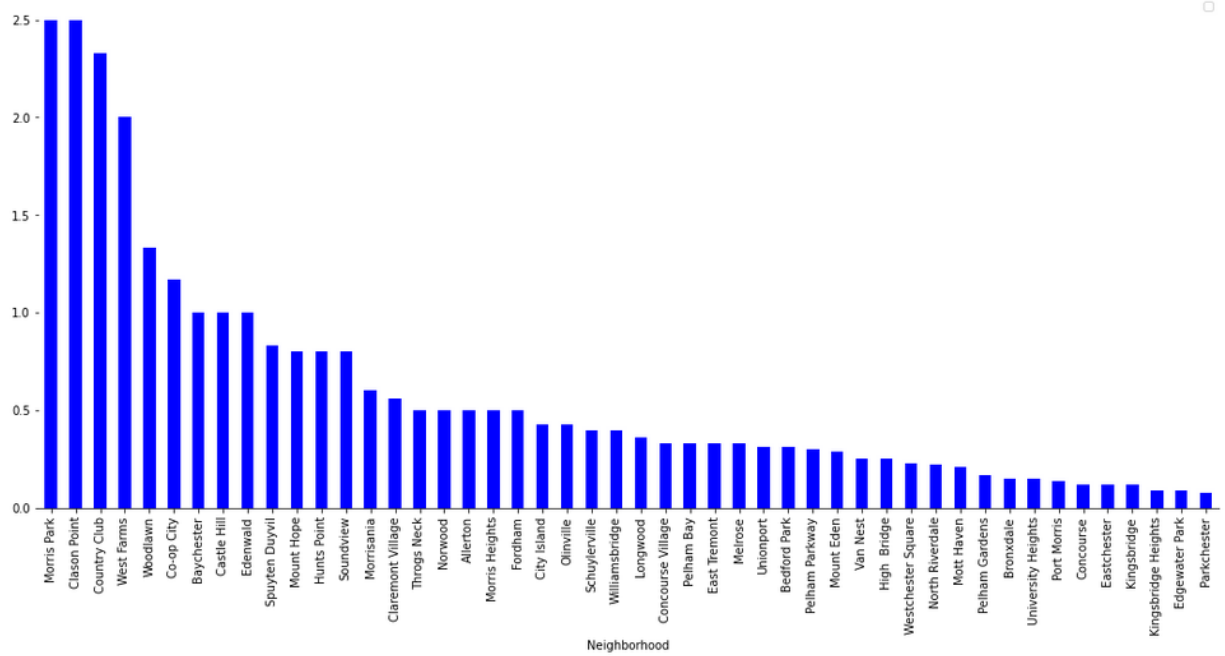
Now to get a perspective as to how fitness as an industry is present in New York, we compare with another industry responsible for our health and fitness i.e. the restaurants of the city. Here we look at the number of restaurants in each borough and get an idea about how these 2 industries go in terms of number of venues.

The below charts give a ratio of No. of fitness Venues/Restaurant in each of the neighborhoods

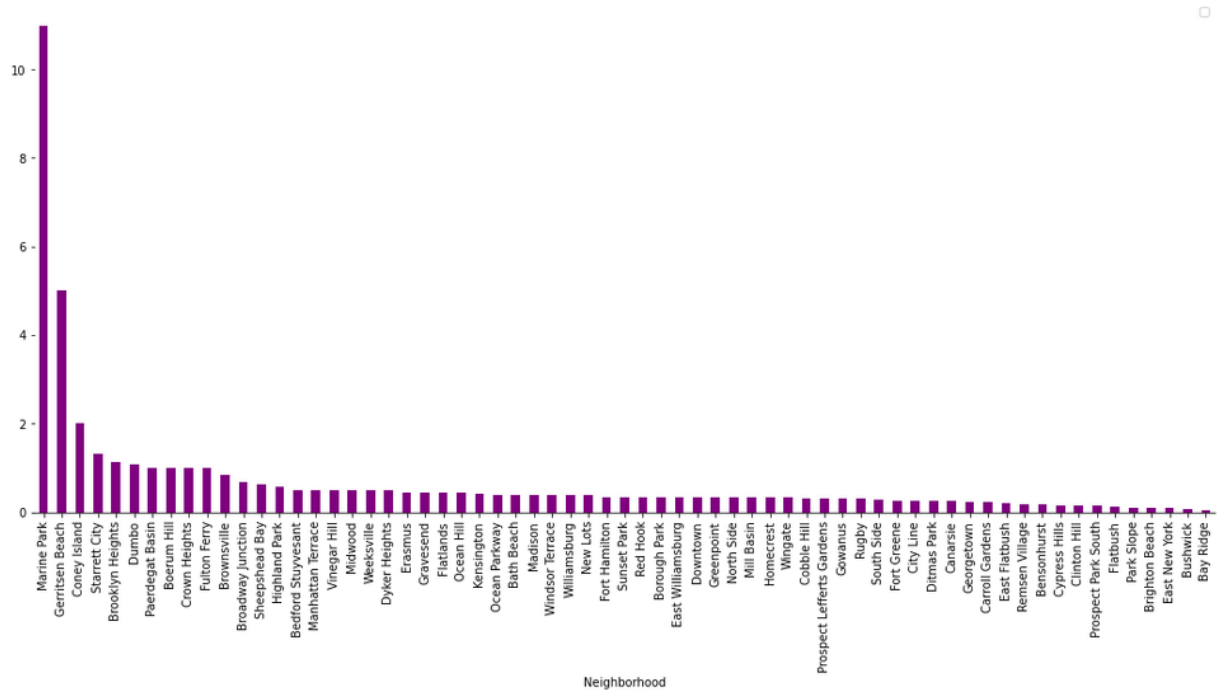
Manhattan - Ratio of Fitness Venues per Restaurant



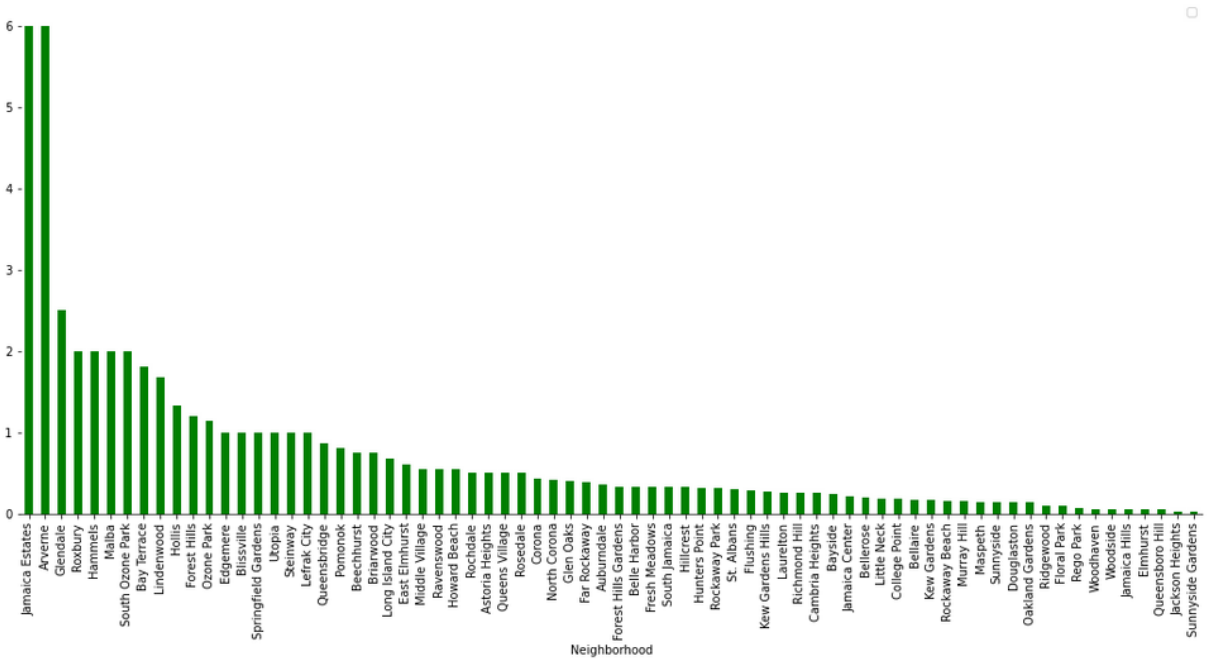
The Bronx - Ratio of Fitness Venues per Restaurant

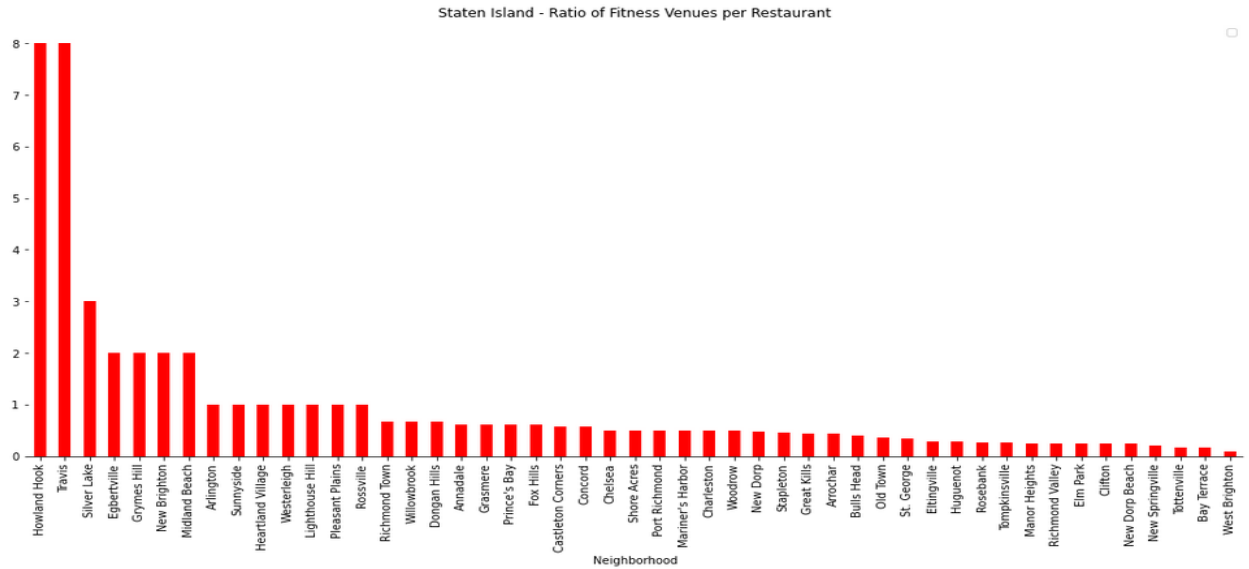


Brooklyn - Ratio of Fitness Venues per Restaurant



Queens - Ratio of Fitness Venues per Restaurant



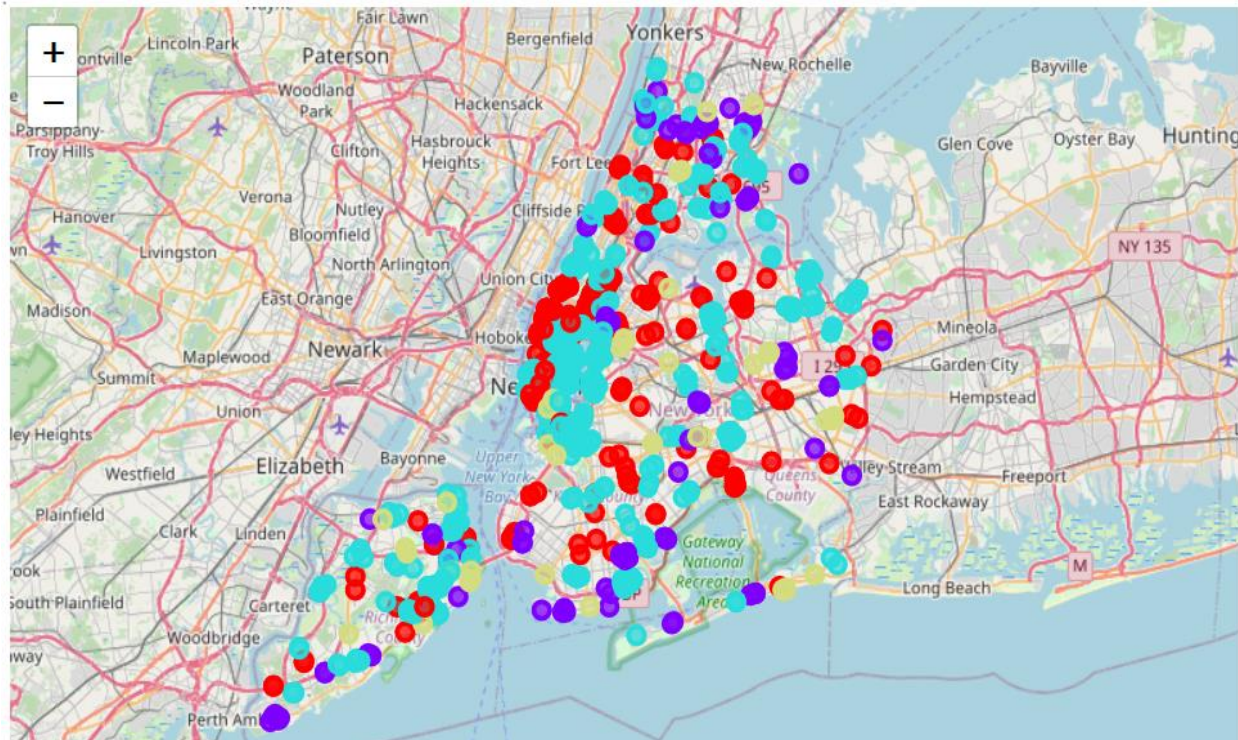


We see that there is clear parity in the distribution of the venues.

## Results

### Clustering

Now that we have understood about the availability of these venues, let us cluster them using KMeans clustering algorithm and see if they have any innate pattern in the way they are distributed and preferred. We got 4 clusters spread across New York as below,





Sample data from each of the clusters show the most common location for each cluster

### Cluster 1

Neighborhood	Venue Latitude	Venue Longitude	Venue Category	Category	color	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Washington Heights	40.848562	-73.936941	Gym	gym/fitness center	Orange	0	gym/fitness center	Pool	Yoga Studios
Central Harlem	40.812608	-73.939699	Gym	gym/fitness center	Orange	0	gym/fitness center	Yoga Studios	Track/Trail
Yorkville	40.775625	-73.947402	Gym	gym/fitness center	Orange	0	gym/fitness center	Sports	Pool
Yorkville	40.777947	-73.949378	Gym	gym/fitness center	Orange	0	gym/fitness center	Sports	Pool
Yorkville	40.779062	-73.944095	Gym	gym/fitness center	Orange	0	gym/fitness center	Sports	Pool
...	...	...	...	...	...	...	...	...	...
Lindenwood	40.662833	-73.852720	Weight Loss Center	gym/fitness center	Orange	0	gym/fitness center	Other venues	Yoga Studios
Howard Beach	40.651928	-73.837653	Sports Club	Sports	purple	0	gym/fitness center	Sports	Other venues
Flushing	40.766154	-73.831431	Pool Hall	Pool	red	0	gym/fitness center	Pool	Yoga Studios
Lindenwood	40.666577	-73.851563	Dance Studio	Other venues	yellow	0	gym/fitness center	Other venues	Yoga Studios
Howard Beach	40.651032	-73.837298	Indoor Play Area	Other venues	yellow	0	gym/fitness center	Sports	Other venues

There are 252 venues under this cluster. This cluster seems to contain locations that have gym as their most preferred location. From the map, it is observed that the cluster 1(red) is mostly in Manhattan region with higher concentration the Bay Area.

### Cluster 2

Neighborhood	Venue Latitude	Venue Longitude	Venue Category	Category	color	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Morningside Heights	40.811358	-73.965748	Tennis Court	Sports	purple	1	Sports	gym/fitness center	Yoga Studios
Morningside Heights	40.811324	-73.965724	Tennis Court	Sports	purple	1	Sports	gym/fitness center	Yoga Studios
Bedford Park	40.873893	-73.888768	Gym	gym/fitness center	Orange	1	Sports	gym/fitness center	Yoga Studios
Spuyten Duyvil	40.878628	-73.914568	Tennis Stadium	Sports	purple	1	Sports	gym/fitness center	Yoga Studios
Fieldston	40.894779	-73.905598	Athletics & Sports	Sports	purple	1	Sports	gym/fitness center	Yoga Studios
...	...	...	...	...	...	...	...	...	...
Rockaway Beach	40.584960	-73.821430	Golf Course	Sports	purple	1	Sports	gym/fitness center	Yoga Studios
Hillcrest	40.724029	-73.794458	College Basketball Court	Sports	purple	1	Sports	gym/fitness center	Yoga Studios
Hillcrest	40.723030	-73.796101	College Stadium	Sports	purple	1	Sports	gym/fitness center	Yoga Studios
Roxbury	40.567476	-73.891433	Trail	Track/Trail	green	1	Track/Trail	Sports	gym/fitness center
Glendale	40.702862	-73.869528	Dance Studio	Other venues	yellow	1	Sports	Other venues	gym/fitness center

Cluster 2 prefers Sports venues like baseball field, tennis courts, basketball courts etc. Majority of the venues in this cluster are present in The Bronx region followed by interior parts of Queens and Brooklyn away from the bay area.

### Cluster 3

Neighborhood	Venue Latitude	Venue Longitude	Venue Category	Category	color	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Marble Hill	40.876705	-73.906372	Gym	gym/fitness center	Orange	2	gym/fitness center	Yoga Studios	Sports
Marble Hill	40.877271	-73.905595	Gym	gym/fitness center	Orange	2	gym/fitness center	Yoga Studios	Sports
Chinatown	40.717812	-73.992624	Gym	gym/fitness center	Orange	2	gym/fitness center	Yoga Studios	Sports
Chinatown	40.718700	-73.992028	Gym	gym/fitness center	Orange	2	gym/fitness center	Yoga Studios	Sports
Hamilton Heights	40.826827	-73.949984	Gym	gym/fitness center	Orange	2	Yoga Studios	gym/fitness center	Track/T
...	...	...	...	...	...	...	...	...	...
Ozone Park	40.679797	-73.840457	Bowling Alley	Other venues	yellow	2	gym/fitness center	Other venues	Yoga Studios
Hunters Point	40.746423	-73.951848	Bowling Alley	Other venues	yellow	2	gym/fitness center	Yoga Studios	Track/T
Little Neck	40.771499	-73.734334	Indoor Play Area	Other venues	yellow	2	gym/fitness center	Other venues	Sports
Little Neck	40.772553	-73.733803	Outdoors & Recreation	Other venues	yellow	2	gym/fitness center	Other venues	Sports
Far Rockaway	40.606464	-73.756246	Other Great Outdoors	Other venues	yellow	2	gym/fitness center	Other venues	Yoga Studios

This cluster consists of neighborhoods that have mostly gyms and yoga centers as their most common venue and that have mainly Yoga studios as their 2<sup>nd</sup> most common venue.

#### Cluster 4

Neighborhood	Venue Latitude	Venue Longitude	Venue Category	Category	color	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
Williamsbridge	40.882686	-73.862978	Dance Studio	Other venues	yellow	3	Other venues	gym/fitness center	Yoga Studios
East Tremont	40.844950	-73.887340	Dance Studio	Other venues	yellow	3	Other venues	gym/fitness center	Yoga Studios
Eastchester	40.886020	-73.823207	Bowling Alley	Other venues	yellow	3	Other venues	gym/fitness center	Yoga Studios
East Tremont	40.845818	-73.884549	Outdoors & Recreation	Other venues	yellow	3	Other venues	gym/fitness center	Yoga Studios
Gowanus	40.676820	-73.998652	Yoga Studio	Yoga Studios	blue	3	Yoga Studios	Other venues	gym/fitness center
Fulton Ferry	40.704328	-73.994547	Trail	Track/Trail	green	3	Track/Trail	Other venues	gym/fitness center
Bath Beach	40.595941	-74.000917	Surf Spot	Other venues	yellow	3	Other venues	gym/fitness center	Yoga Studios
City Line	40.681956	-73.866227	Ski Area	Other venues	yellow	3	Other venues	gym/fitness center	Yoga Studios
Fulton Ferry	40.699822	-73.998349	Roller Rink	Other venues	yellow	3	Track/Trail	Other venues	gym/fitness center
Gowanus	40.670342	-73.991967	Dance Studio	Other venues	yellow	3	Yoga Studios	Other venues	gym/fitness center

Cluster 4 contains majority neighborhoods that uniquely has other venues like Bowling Alleys, Roller Rinks, Ski Area etc. as the most common venues.

#### Discussion

From the above analysis we get to know which neighborhood is ideal for a particular kind of fitness activity. Some of the valuable suggestions/directions that someone from the fitness industry can take away are listed below,

- Neighborhoods in the Bronx region mostly has sports arena like Baseball, Tennis or Basketball courts. Apparently, this might be due to the fact that there are almost equal number of gyms and sports venues in Bronx and zero yoga studios. So it would be a good idea to start a Yoga Studio here and see how the response turns up. Else, opening a new multi sports arena with basketball, tennis and volleyball courts would attract a good enough response.
- Brooklyn is the borough where the ratio of fitness venues/restaurant is less than 1 in most of the neighborhoods. This can be noted and taken advantage of especially for developers of gyms and yoga studios since most of the neighborhoods in Brooklyn fall under cluster 3.
- It is important to note that 83% (approx.) of the neighborhoods all over New York have Fitness Venues/Restaurants ratios less than 1. Though these 2 industries and their business models works very different, this should be looked at from a personal hygiene perspective. If a person can have these may places come up as suggestions for his search on restaurants, there should

also be a good number of places around his neighborhood where he could get his physical fitness right.

These are some of the observations made from the analysis. The results are however restricted to the limitations with ForuSquare API calls and number of records that can be extracted per neighborhood. But this can be a good starting point for further analysis with demographic variables like population's age, income, population density and health information like diabetics for each neighborhood which could give a more accurate outcomes and insights.

## Conclusion

Health & Fitness as industry could gain from analyzing these preferences and thereby understanding the expectations of the population of a locality before starting an establishment like a gym or sports arena. This is going to be crucial with increasing emphasis on physical fitness and health amidst the current health crisis since general immunity plays a crucial factor. Fitness industry can solve this problem by strategically bringing up establishments according to the needs of the people of a neighborhood so that the latter doesn't have to compromise on the way they would like to maintain their physical health. This in turn would be profitable for businesses because of improved customer acquisition and retention – the two important metrics for fitness industry.