



Applied Data Science 🏋️ Capstone Project

Opening a GYM in New York City

Where?

New York City is famous for all the gyms and fitness centers that are open 24/7 all around the city.

Therefore, it is a challenging choice to find the right location when entering the market.

Manhattan

My client would like start in Manhattan; thus,
we will be focusing on this borough
throughout the analysis.

Data Selection

01

Category

Let's find the category using Foursquare API

02

Location

We are using Google Maps API to find geographic coordinates

03

Quite a lot

There are 1696 gyms in New York

Neighborhood	Venue Latitude	Venue Longitude	Venue Category
-73.910986	40.877271	-73.905995	Gym
-73.910986	40.876705	-73.905372	Gym
-73.910986	40.882748	-73.907625	Gym
-73.910986	40.879628	-73.914668	Tennis Stadium
-73.910986	40.885669	-73.912437	Gym



In [25]: newyork_venues_gym.shape

Out[25]: (1696, 7)



Methodology

- Converting addresses to latitude and longitude
- Applying Foursquare API to explore neighborhoods in Manhattan
- Creating a map of all gyms in Manhattan

Gyms in Manhattan



Create columns according to the most common venues

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0 Battery Park City	Gym	Gym / Fitness Center	Yoga Studio	Corporate Amenity	Gymnastics Gym	Building	Clothing Store	Club House	Community Center	Cultural Center
1 Carnegie Hill	Gym	Gym / Fitness Center	Yoga Studio	Community Center	Pilates Studio	Gym Pool	Gymnastics Gym	Building	Clothing Store	Club House
2 Central Harlem	Gym	Gym / Fitness Center	Yoga Studio	Gymnastics Gym	Building	Clothing Store	Club House	Community Center	Corporate Amenity	Cultural Center
3 Chelsea	Gym	Gym / Fitness Center	Yoga Studio	Residential Building (Apartment / Condo)	Martial Arts School	School	Recreation Center	Building	Clothing Store	Club House
4 Chinatown	Gym	Gym / Fitness Center	Martial Arts School	Clothing Store	Pool	Office	Gym Pool	Building	Club House	Community Center

Using K-means to cluster data with less number of GYMs

Cluster 0

In [50]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 0, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]											
Out[50]:	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
13	Lincoln Square	Gym	Gym / Fitness Center	Yoga Studio	Residential Building (Apartment / Condo)	Cultural Center	Pilates Studio	Gym Pool	Building	Clothing Store	Club House
16	Midtown	Gym	Gym / Fitness Center	Yoga Studio	Building	Residential Building (Apartment / Condo)	Gymnastics Gym	Clothing Store	Club House	Community Center	Corporate Amenity
19	East Village	Gym	Gym / Fitness Center	Yoga Studio	Spa	Pool	Outdoor Gym	Gym Pool	Building	Clothing Store	Club House
21	Tribeca	Gym	Gym / Fitness Center	Martial Arts School	Office	Corporate Amenity	Yoga Studio	Recreation Center	Spa	Building	Clothing Store
23	Soho	Gym	Gym / Fitness Center	Yoga Studio	Martial Arts School	Office	School	Spa	Building	Clothing Store	Club House
27	Gramercy	Gym	Gym / Fitness Center	Yoga Studio	Pool	Outdoor Gym	Martial Arts School	Spa	Residential Building (Apartment / Condo)	Tennis Court	Building
32	Civic Center	Gym	Gym / Fitness Center	Martial Arts School	Office	Corporate Amenity	Yoga Studio	Recreation Center	Spa	Building	Clothing Store

Using K-means to cluster data with less number of GYMs

Cluster 1

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1 Inwood	Gym	Gym / Fitness Center		Yoga Studio	Gymnastics Gym	Building	Clothing Store	Club House	Community Center	Corporate Amenity
4 Hamilton Heights	Gym	Basketball Court	Gym / Fitness Center	Gymnastics Gym	Building	Clothing Store	Club House	Community Center	Corporate Amenity	Cultural Center
7 East Harlem	Gym	Gym / Fitness Center		Yoga Studio	Gymnastics Gym	Building	Clothing Store	Club House	Community Center	Corporate Amenity
8 Upper East Side	Gym	Gym / Fitness Center		Yoga Studio	Gym Pool	Club House	Recreation Center	Plate Studio	Building	Clothing Store
9 Yorkville	Gym	Gym / Fitness Center		Yoga Studio	Community Center	Plate Studio	Gym Pool	Gymnastics Gym	Building	Clothing Store
10 Upper West Side	Gym	Gym / Fitness Center		Yoga Studio	Cultural Center	Playground	Plate Studio	Gym Pool	Building	Clothing Store
24 West Village	Gym	Gym / Fitness Center		Rose	Yoga Studio	Gymnastics Gym	Building	Clothing Store	Club House	Community Center
25 Manhattan Valley	Gym	Gym / Fitness Center		Playground	Yoga Studio	Gymnastics Gym	Building	Clothing Store	Club House	Community Center
26 Battery Park City	Gym	Gym / Fitness Center		Yoga Studio	Corporate Amenity	Gymnastics Gym	Building	Clothing Store	Club House	Community Center
27 Financial District	Gym	Gym / Fitness Center		Corporate Amenity	Yoga Studio	Gymnastics Gym	Building	Clothing Store	Club House	Cultural Center
30 Carnegie Hill	Gym	Gym / Fitness Center		Yoga Studio	Community Center	Plate Studio	Gym Pool	Gymnastics Gym	Building	Clothing Store
31 Yonkers Bay	Gym	Gym / Fitness Center	Residential Building Apartment / Condo	Yoga Studio	Gymnastics Gym	Building	Clothing Store	Club House	Community Center	Corporate Amenity
33 Tudor City	Gym	Gym / Fitness Center	Residential Building Apartment / Condo	Yoga Studio	Gymnastics Gym	Building	Clothing Store	Club House	Community Center	Corporate Amenity

Using K-means to cluster data with less number of GYMs

Cluster 2

```
In [52]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 2, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

Out[52]:

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2 Washington Heights	Gym	Yoga Studio	Gymnastics Gym	Building	Clothing Store	Club House	Community Center	Corporate Amenity	Cultural Center	Dance Studio
5 Manhattanville	Gym	Basketball Court	Gymnastics Gym	Building	Clothing Store	Club House	Community Center	Corporate Amenity	Cultural Center	Dance Studio
6 Central Harlem	Gym	Gym / Fitness Center	Yoga Studio	Gymnastics Gym	Building	Clothing Store	Club House	Community Center	Corporate Amenity	Cultural Center
20 Lower East Side	Gym	Gym / Fitness Center	Clothing Store	Pool	Yoga Studio	Gymnastics Gym	Building	Club House	Community Center	Corporate Amenity
26 Morningside Heights	Gym	Gym / Fitness Center	Yoga Studio	Gymnastics Gym	Building	Clothing Store	Club House	Community Center	Corporate Amenity	Cultural Center
37 Stuyvesant Town	Gym	Martial Arts School	Outdoor Gym	Gymnastics Gym	Building	Clothing Store	Club House	Community Center	Corporate Amenity	Cultural Center

Using K-means to cluster data with less number of GYMs

Cluster 3

In [53]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 3, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]										
Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
16 Murray Hill	Gym	Gym / Fitness Center		Yoga Studio	Building	Marital Arts School	Residential Building (Apartment / Condo)	Recreation Center	Tennis Court	Clothing Store
18 Greenwich Village	Gym	Gym / Fitness Center	Martial Arts School		Office	Yoga Studio		Spa	Recreation Center	Building
22 Little Italy	Gym	Gym / Fitness Center		Yoga Studio	Martial Arts School	Office		School	Spa	Building
33 Midtown South	Gym	Gym / Fitness Center	Residential Building (Apartment / Condo)		Yoga Studio	Building	Marital Arts School	Recreation Center	Spa	Tennis Court
36 Flatiron	Gym	Gym / Fitness Center		Yoga Studio	Residential Building (Apartment / Condo)	Pool	Marital Arts School		Tennis Court	Building

Using K-means to cluster data with less number of GYMs

Cluster 4

In [54]:	manhattan_merged.loc[manhattan_merged['Cluster_Labels']] == 4, manhattan_merged.columns[[1]] * 1.15,color[5, manhattan_merged.shape[1]]])									
Out[54]:										
Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0 Marble Hill	Gym	Gym / Fitness Center	Tennis Stadium	Yoga Studio	Gymnastics Gym	Building	Clothing Store	Club House	Community Center	Corporate Amenity
1 Chinatown	Gym	Gym / Fitness Center	Martial Arts School	Clothing Store	Pool	Office	Gym Pool	Building	Club House	Community Center
10 Lenox Hill	Gym	Gym / Fitness Center	Tennis Court	Club House	Recreation Center	Yoga Studio	Gym Pool	Building	Clothing Store	Community Center
11 Roosevelt Island	Gym	Gym / Fitness Center	Tennis Court	School	Recreation Center	Club House	Dance Studio	Yoga Studio	Gym Pool	Building
14 Clinton	Gym	Gym / Fitness Center	Yoga Studio	Building	Residential Building (Apartment / Condo)	Gymnastics Gym	Clothing Store	Club House	Community Center	Corporate Amenity
17 Chelsea	Gym	Gym / Fitness Center	Yoga Studio	Residential Building (Apartment / Condo)	Martial Arts School	School	Recreation Center	Building	Clothing Store	Club House
31 Noho	Gym	Gym / Fitness Center	Yoga Studio	Pool	Clothing Store	Outdoor Gym	Martial Arts School	Recreation Center	Playground	Building
34 Sutton Place	Gym	Gym / Fitness Center	Tennis Court	Residential Building (Apartment / Condo)	Club House	Yoga Studio	Gym Pool	Building	Clothing Store	Community Center
39 Hudson Yards	Gym	Gym / Fitness Center	Building	Residential Building (Apartment / Condo)	Yoga Studio	Gymnastics Gym	Clothing Store	Club House	Community Center	Corporate Amenity

Resulting Map

To select an ideal location for a gym, we need to take other factors into account as well. However, the below clustered map can help us in identifying key areas.

Cluster 2 would be an ideal area

Discussion

- This analysis is based on limited data, however it is sufficient for the Applied Data Science Capstone
- With a more precise business plan, the potential is there to create a comprehensive and live report
- Other elements, such as public transport, schools, workplace, restaurants nearby would be crucial to determine an ideal location
- The aim of the project is to showcase some of the elements covered in the IBM Professional Data Science Certificate

Conclusion

- This report has been designed for a client who would like to open a gym in Manhattan, New York City
- We have used Google Map's API, Foursquare API and K-Means Clustering to analyze the location of current gyms & neighborhoods
- We could look at several possible locations in Manhattan, however Cluster 2 – Central Harlem may be a good start.
- The report could use more data for data-driven decision making; however, it duly covers the current Capstone Project.



THANK YOU.

OLIVER KISSZABO.