

Determining the location for a new gym in Brooklyn

Federico Frigola

July 20, 2021

Table of Contents

1. Introduction.....	3
2. Business Problem	3
3. Data	4
3.1 Data Cleaning	4
4. Methodologies used to conduct the analysis.....	6
4.1 Clustering	6
4.2 Competitors Analysis	7
5. Results.....	8
6. Conclusion	9
7. Future Directions.....	9

1. Introduction

During the Covid-19 pandemic, the gym sector has suffered a big drop in the revenues due to the restrictions put in place by governments to limit infections. Even though many fitness enthusiasts created their own home gym to keep on training during the lockdown phase, the gym represents in the heart of many, not only a place where you can train with other people, but also an environment where you can socialize and create meaningful relationships. Those relationships that the pandemic has taught us to be very important.

2. Business Problem

Having said that the gym sector has been hardly beaten by social restrictions in the last months, it's important to make good strategic decisions not to lose even more money. The purchase/rent of the building where the gym will be located, is definitely the biggest and the most important a gym owner has to make. Therefore, it is fundamental to take in consideration as many factors as possible in the choice of the location, as from it will depend a large part of the future of the business. PumpIt, is an international company that owns gyms in all the major European capitals; their goal for 2021 is target the US market starting from New York City and then expanding in the West Coast. More precisely, they have picked Brooklyn to be the starting point for their expansion project in the US. With a surface of 180 km², Brooklyn is the second largest borough in New York City after Queens; scanning such a big area to find the best location for a new gym, is not an easy task. This is why the company has hired as consultants, a team of data scientist to be supported during the decision making process with meaningful insights coming from data. What consultants have been asked to do, is not only to understand which is the best area

in Brooklyn where to open a new gym, but also to check the gym rating nearby the chosen area to understand a little bit more about the competition.

3. Data

The Data that will be used for this analysis are mainly of two types:

- 1) Data from the internet about New York City neighbourhoods and borough with their longitude and latitude to be able to create maps
- 2) Venues Data retrieved using the Foursquare API to explore and analyze neighborhoods in New York City, more specifically those located in the Brooklyn borough

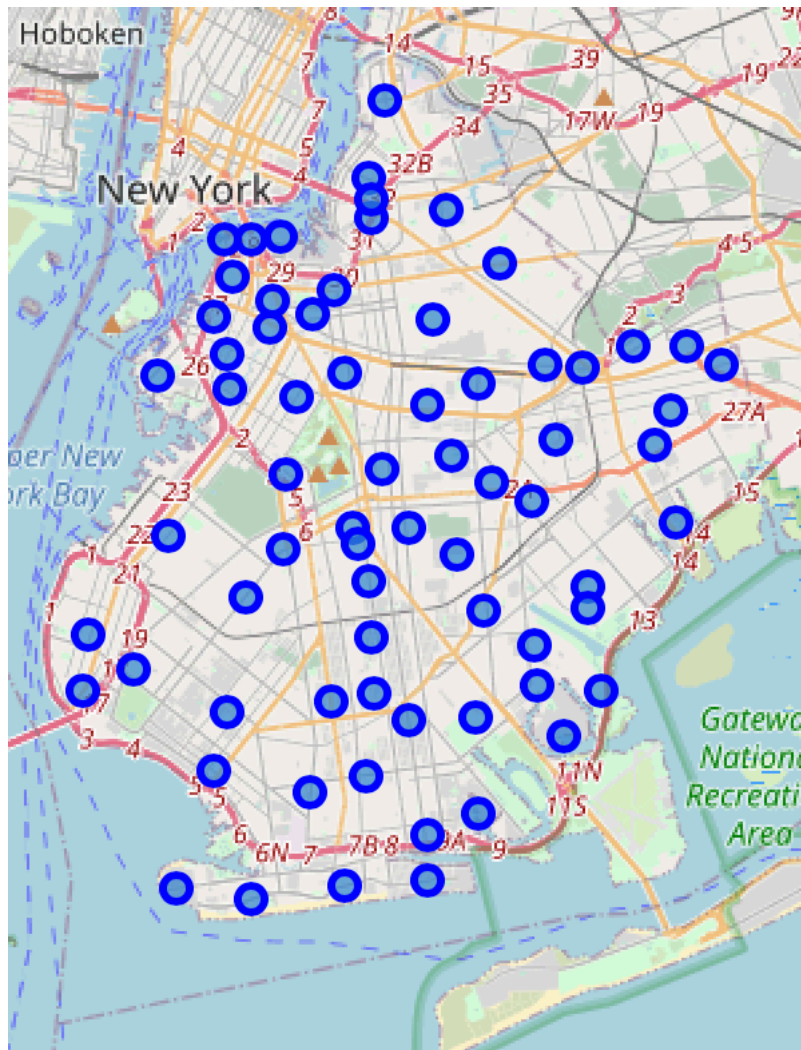
3.1 Data Cleaning

As mentioned, the first set of data has been downloaded from internet and cleaned to get to the final form presented in the figure below.

	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

In total, our dataset is composed by 5 different types of Borough and 306 Neighborhoods.

As our client aims at investing specifically in Brooklyn, I've extracted from the dataset only those neighborhoods associated with "Brooklyn" in the "Borough" column. The map of Brooklyn with the related neighborhoods is presented below.



The second group of data that we required to conduct this analysis, can be retrieved using FootSquare API after creating a developer account. Using this method, I've downloaded info about at maximum 1000 venues per neighborhood in a range of 10km. The dataset comprising all the venues displays 2740 observations and the format is presented in the following figure.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bay Ridge	40.625801	-74.030621	Pilo Arts Day Spa and Salon	40.624748	-74.030591	Spa
1	Bay Ridge	40.625801	-74.030621	Bagel Boy	40.627896	-74.029335	Bagel Shop
2	Bay Ridge	40.625801	-74.030621	Georgian Dream Cafe and Bakery	40.625586	-74.030196	Caucasian Restaurant
3	Bay Ridge	40.625801	-74.030621	Leo's Casa Calamari	40.624200	-74.030931	Pizza Place
4	Bay Ridge	40.625801	-74.030621	Pegasus Cafe	40.623168	-74.031186	Breakfast Spot

We can notice that each row is a venue, and we were able to retrieve info about the Name, the location and the category. This represents the starting point of our analysis as it will help us in understanding which are the most common types of venues in each neighborhood.

4. Methodologies used to conduct the analysis

As the goal of this work is to find the best area in Brooklyn for a new gym opening, I've decided to use K – means clustering to divide the borough area according to their similarity. From a theoretical perspective, this will indicate us which is the most appropriate area to welcome a new gym.

Once the area has been defined, I've conducted an analysis of the competition by looking at how much the other gyms were appreciated by their clients.

4.1 Clustering

Once I've defined the top 15 most common venues per neighborhood, the next step is to cluster them and retrieve a summary table for each cluster such to identify the most prone to free-time and training activities.

I've decided to pick $k = 7$ as Brooklyn is the second biggest borough in New York City after the Queens. After having define the number of clusters into which I want to split my dataset, I applied K – means clustering. With the results I have created a dataset

3. Gott Gym
4. St Bernard Gymnasium
5. Harbor Fitness

Once competitors have been defined, the next step is to look for their rating, and if not available, look for the number of likes assigned to each gym. This can be done calling for “tips” again using FootSquare API that identifies each single gym with an “id”.

5. Results

Clustering results show that Cluster 2, Cluster 4 and Cluster 6 include a vast variety of venues and services: this may mean that a gym opening would attract many individuals with very different lifestyles and interest. However, it may result in an anonymous choice with little resonance due to the large set of venues available.

Cluster 1 seems to be mainly dominated by restaurants of all types; this indicates that is not the best environment to place a gym as people coming in this area are interested by other types of venues.

Cluster 3 includes mainly Parks and restaurants. It sounds attractive but at the same time we have to keep in mind that if there is the possibility to train outdoor for free, people may avoid paying a gym subscription when they can perform their workout at the fresh air without spending their money.

Cluster 5 is mainly focused on business, as it displays business services, childcare services and food in general. It may make sense to open a gym available 24h for workers nearby the place they work at. This would save them time to get to the gym therefore increasing the motivation.

Finally, Cluster 7 is mainly characterized by the presence of venues for free time and sports. For instance, we can find basketball field, playground, athletics & sports and parks.

From the competitors' side, unfortunately, only 1 of the 5 gyms present in the area of interest has been graded. The available grade is just over the sufficiency which means that Harbor Fitness' clients are not so much satisfied with services provided. This may indicate a low retention rate and the presence of clients willing to change if new players enter the market.

Moreover, it's interesting to note the gym with the highest number of likes by far, has a rating of only 6.9. We may infer that the competition is not so fierce as we expected and therefore there is room for growth for newcomers.

6. Conclusion

My final advice would be to pick Cluster 7, more precisely Bergen Beach, as the location for a new gym opening due to the big presence of free time/sport-related venues; the presence of factories may also attract those workers willing to train after/before work. In addition to that, the competitor analysis that has been conducted has highlighted a weak presence of popular gym, indicating a potentially low customer loyalty if a strong new player enters the market.

7. Future Directions

Because only 1 out of 5 competitors has been rated, a further analysis that may be conducted in the future could retrieve information about gyms appreciation using social media. For instance, using Twitter API one could easily get the Twitter audience sentiment by analyzing tweets containing gyms' name.