

Capstone Project Report



FEBRUARY 8

Authored by: Sneha Devadas

Identifying an Ideal Location for a Supplement Shop in Manhattan

Introduction

Manhattan, known regionally as **the City** and the urban core of the New York metropolitan area, is the most densely populated of the five boroughs of New York City. Manhattan has been described as the cultural, financial, media, and entertainment capital of the world, and the borough hosts the United Nations Headquarters. Anchored by Wall Street in the Financial District of Lower Manhattan, New York City has been called both the most economically powerful city and the leading financial center of the world. Manhattan real estate has since become among the most expensive in the world, with the value of Manhattan Island, including real estate, estimated to exceed US\$3 trillion in 2013; median residential property sale prices in Manhattan approximated US\$1,600 per square foot (\$17,000/m²) as of 2018, with Fifth Avenue in Midtown Manhattan commanding the highest retail rents in the world, at US\$3,000 per square foot (\$32,000/m²) per year in 2017.

To setup and run a business entity in such location requires a great deal of consideration and deliberation. Among important factors such as financial, operational and legal considerations, the location of the proposed business venture plays a very important role. It is so powerful a factor that it can make or break the business. We need to identify such a location that it effectively addresses the gap in demand. In other words, we need to identify an area which has a lot of market potential waiting to be tapped. One way to identify this is by analyzing the presence and popularity of such business ventures in the proposed area versus the demand.

Business Problem/Problem Statement

Identify a location for setting up a new Supplement Store in one of the neighborhoods in Manhattan, which may help to maximize the sales.

Stakeholder

An individual or a partnership interested to setup a Supplement Shop in the neighborhoods of Manhattan.

Data

To address the business problem defined in the previous section, we need the following data:

- Name of all the neighborhoods in Manhattan along with the geographical coordinates (latitude and longitude).
- List of popular venues in each of these neighborhoods.

“newyork_data.json” file available in the following link (https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json) contains the a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough in New York , where Manhattan is one borough, as well as the latitude and longitude coordinates of each neighborhood.

For example, below is snapshot of the data extracted from “newyork_data.json” file.

	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

This dataset is cleaned and processed to retrieve data of the neighborhoods in Manhattan alone.

List of popular venues in each of the neighborhood is obtained with the help of Foursquare API. The data is obtained in the form of a json file, which is then processed to retrieve the desired information i.e., the venue name, category to which it belongs, the latitude and longitude of the venue.

For example, the information extracted from the response.json file for the neighborhood Stuyvesant Town in Manhattan is as follows.

	name	categories	lat	lng
0	Stuyvesant Cove Park	Park	40.732513	-73.973876
1	Con Ed Field	Baseball Field	40.729206	-73.973732
2	Stuyvesant Oval	Park	40.731665	-73.977835
3	StuyFitness	Gym / Fitness Center	40.729013	-73.976676
4	Pouring Ribbons	Cocktail Bar	40.729093	-73.978033
5	Stuyvesant Oval Fountain	Fountain	40.731987	-73.977685
6	Stuyvesant Town Greenmarket	Farmers Market	40.731429	-73.978222
7	The Roost	Bar	40.729084	-73.978638
8	Unleashed Spa & Self Pet Wash	Pet Service	40.728984	-73.978380
9	B Cup Cafe	Coffee Shop	40.728783	-73.978565
10	Perk Kafe	Coffee Shop	40.729645	-73.978775
11	New York Skyports Seaplane Base	Heliport	40.734959	-73.972991
12	 le	Bar	40.728438	-73.978780
13	Rocks Off Cruises	Harbor / Marina	40.735434	-73.973098
14	Solar One	Park	40.734442	-73.974350
15	BP	Gas Station	40.735107	-73.974554

These datasets are then used for segmenting and clustering the neighborhoods in Manhattan. Further analysis of these clusters will help us to identify the neighborhood or list of neighborhoods which are promising.

Methodology

In order to choose an ideal neighborhood to set up a Supplement Shop it is necessary to understand the characteristics of each neighborhood. The neighborhood chosen should be such that there is a considerable demand for the service we are offering and there are few market players or competitors. In this project, the approach used was to analyze the popular venues in a particular neighborhood, to identify potential demand as well as possible competitors. In simple words, look for a neighborhood which has no or few supplement shops among the popular venues. At the same time, it should also host venues related to health and fitness such as gym, playgrounds, hospital etc.

After extracting and preprocessing the required data, the neighborhoods were visualized on a map for better understanding.

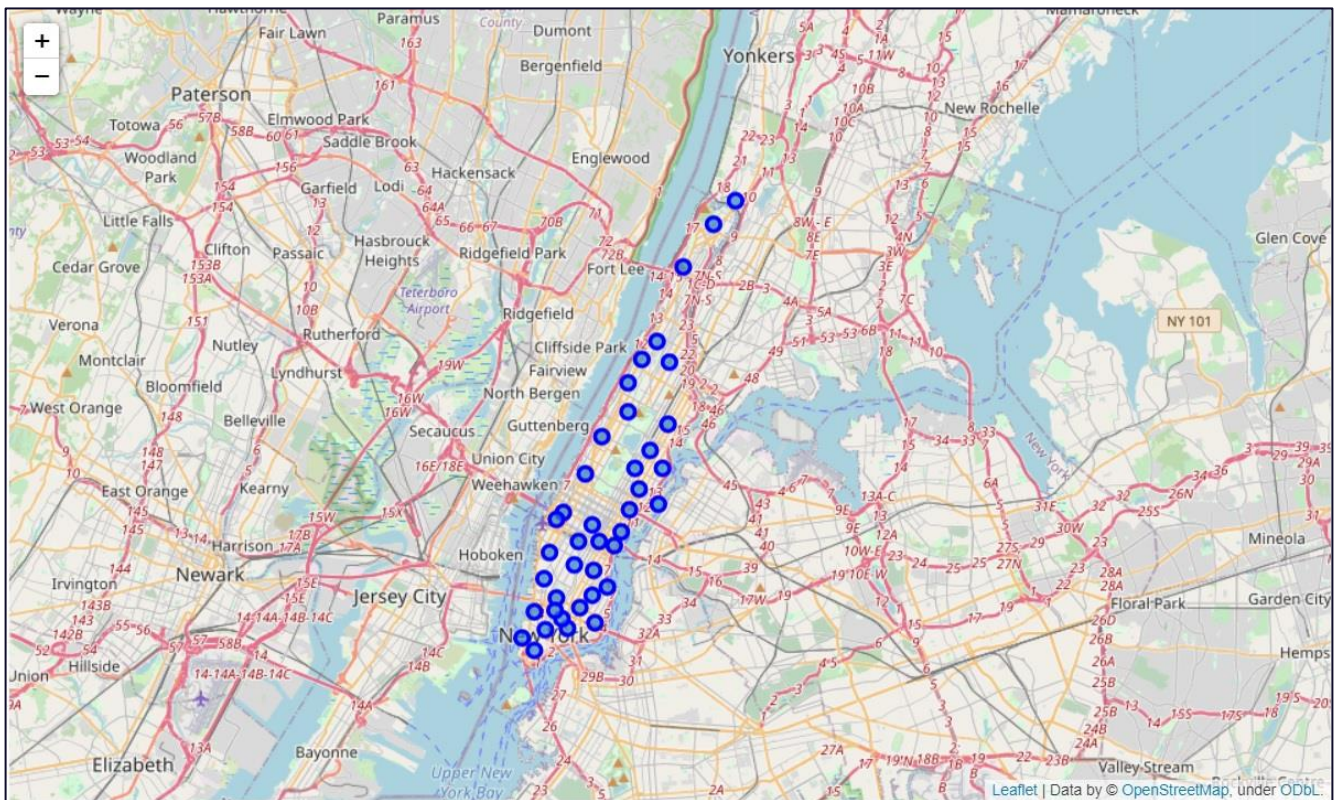


Figure 1 Neighborhoods of Manhattan

Next step is to identify the popular venues in each neighborhood. We used Foursquare API to provide the necessary data. The response was processed to obtain a dataset as shown below.

Table 1 Processed Foursquare API output

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

This data was then consolidated on the basis of neighborhood. Further, frequency of occurrence of each category in each neighborhood was calculated. Based on the frequency of occurrence, top 10 venues were selected for each neighborhood.

Table 2 Top 10 venues of each neighborhood in Manhattan

	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	Coffee Shop	Hotel	Park	Memorial Site	Gym	Clothing Store	Burger Joint	Gourmet Shop	BBQ Joint	Plaza
1	Carnegie Hill	Coffee Shop	Café	Yoga Studio	Wine Shop	Bookstore	Pizza Place	Gym	Cosmetics Shop	Japanese Restaurant	Cocktail Bar
2	Central Harlem	African Restaurant	Seafood Restaurant	Art Gallery	Bar	French Restaurant	Chinese Restaurant	Cosmetics Shop	Gym / Fitness Center	American Restaurant	Park
3	Chelsea	Coffee Shop	Bakery	American Restaurant	French Restaurant	Art Gallery	Ice Cream Shop	Wine Shop	Seafood Restaurant	Cupcake Shop	Bookstore
4	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Hotpot Restaurant	Salon / Barbershop	Spa	American Restaurant	Optical Shop	Dessert Shop	Ice Cream Shop

The above dataset was used for segmenting and clustering the neighborhoods of Manhattan into three groups using k-means clustering algorithm. The value of k was determined by optimizing the value of Squared Error or Cost function as shown below.

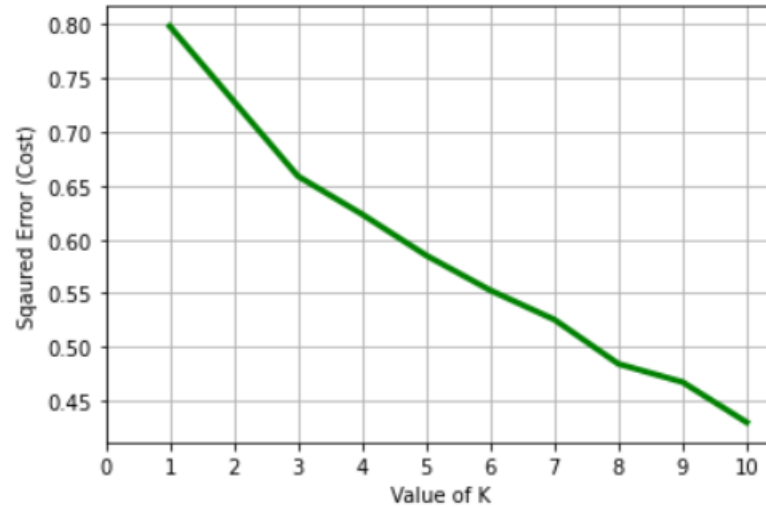


Figure 2 Squared Error (Cost) for each value of k

The clusters obtained were then analyzed to identify defining characteristics.

Results and Discussion

Upon running k-means clustering, with $k = 3$ on the neighborhood data, following clusters were obtained.

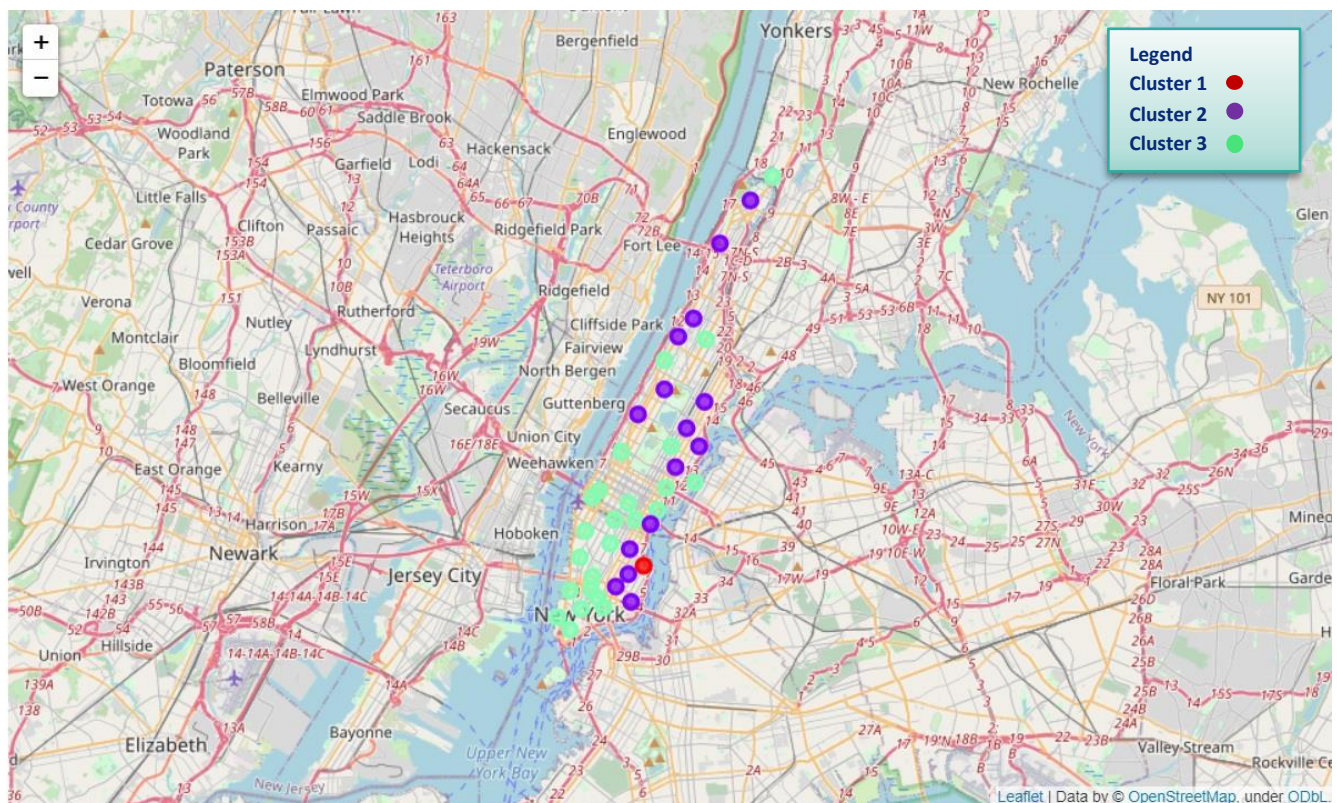


Figure 3 Clusters of Neighborhoods

Cluster 1

Neighborhoods and the top 10 popular venues of cluster 1 are given below. As can be seen clearly, it contains only one neighborhood namely Stuyvesant Town. Also, it has a lot of outdoor recreational and fitness related venues. This could possibly be a good fit for our supplement shop. Nonetheless let's analyze the remaining clusters as well.

Table 3 Neighborhood and popular venues of 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
37	Stuyvesant Town	Park	Bar	Coffee Shop	Pet Service	Baseball Field	Gym / Fitness Center	Cocktail Bar	Heliport	Farmers Market	Harbor / Marina

Cluster 2

Neighborhoods and the top 10 popular venues of cluster 2 are given below. Let's get the frequency distribution of this data to identify the most popular venues of this cluster.

Table 4 Neighborhoods and popular venues of cluster 2

	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	Café	Bakery	Bank	Mobile Phone Shop	Grocery Store	Chinese Restaurant	Shoe Store	Italian Restaurant	Latin American Restaurant	Tapas Restaurant
3	Inwood	Mexican Restaurant	Restaurant	Café	Lounge	Bakery	Pizza Place	Park	Chinese Restaurant	Caribbean Restaurant	Frozen Yogurt Shop
4	Hamilton Heights	Pizza Place	Café	Coffee Shop	Mexican Restaurant	Deli / Bodega	Yoga Studio	Latin American Restaurant	Park	Liquor Store	Cocktail Bar
5	Manhattanville	Coffee Shop	Seafood Restaurant	Sushi Restaurant	Mexican Restaurant	Italian Restaurant	Boutique	Spanish Restaurant	Burger Joint	Bus Station	Cuban Restaurant
7	East Harlem	Mexican Restaurant	Deli / Bodega	Thai Restaurant	Bakery	Sandwich Place	Latin American Restaurant	Dance Studio	Taco Place	Gas Station	Beer Bar

We notice that Mexican Restaurant, Coffee Shop, Café, Pizza Place, Italian Restaurant, Bar, Sushi Restaurant, Bakery, Deli / Bodega and Latin American Restaurant are the most popular venues. We also notice that fitness related or outdoor recreational venues are not that popular in this neighborhood. Different kinds of restaurants and coffee shops are predominant in this neighborhood.

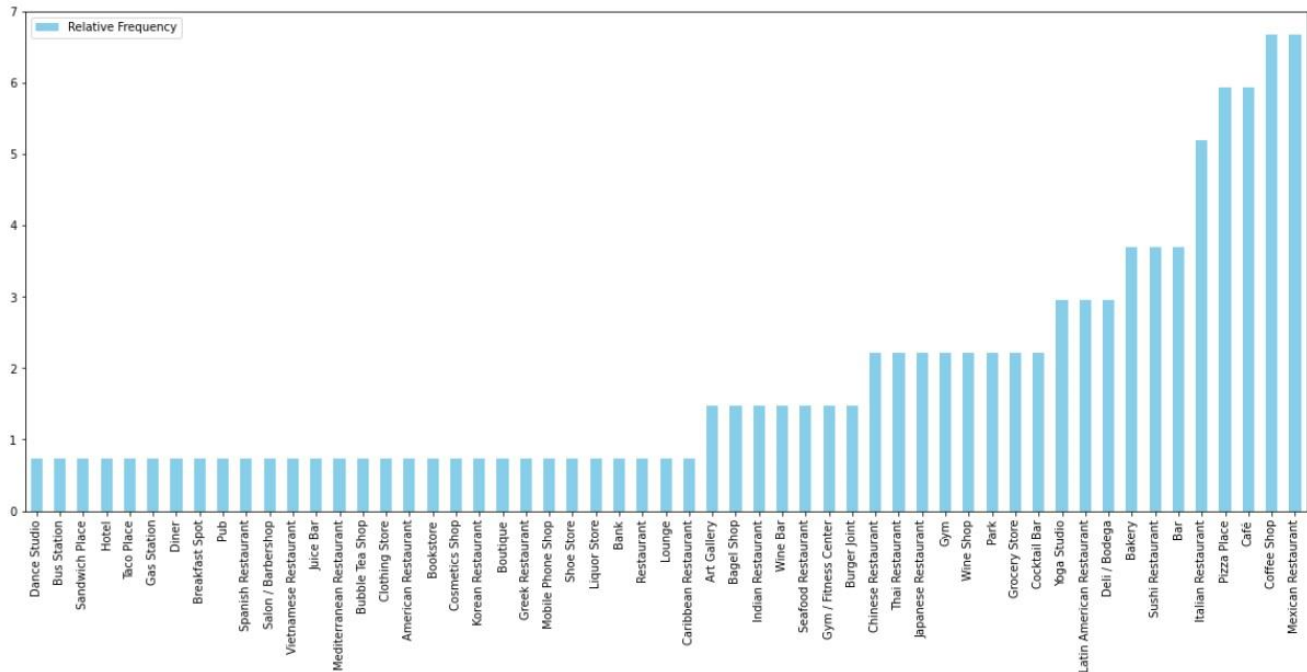


Figure 4 Frequency Distribution of popular venues in cluster 2

Cluster 3

Neighborhoods and the top 10 popular venues of cluster 3 are given below. Let's get the frequency distribution of this data to identify the most popular venues of this cluster.

Table 5 Neighborhoods and popular venues of cluster 3

	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	Sandwich Place	Gym	Coffee Shop	Discount Store	Pharmacy	Steakhouse	Supplement Shop	Seafood Restaurant	Tennis Stadium	Donut Shop
1	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Hotpot Restaurant	Salon / Barbershop	Spa	American Restaurant	Optical Shop	Dessert Shop	Ice Cream Shop
6	Central Harlem	African Restaurant	Seafood Restaurant	Art Gallery	Bar	French Restaurant	Chinese Restaurant	Cosmetics Shop	Gym / Fitness Center	American Restaurant	Park
8	Upper East Side	Coffee Shop	Italian Restaurant	Exhibit	Bakery	Gym / Fitness Center	American Restaurant	Yoga Studio	Juice Bar	Cosmetics Shop	Spa
11	Roosevelt Island	Park	Monument / Landmark	Kosher Restaurant	Food & Drink Shop	Soccer Field	Farmers Market	Supermarket	Metro Station	School	Outdoors & Recreation

We notice that Coffee Shop, Italian Restaurant, American Restaurant, Gym/Fitness Center, Hotel, Park, Gym, Café, Cocktail Bar and Bakery are the most popular venues. We also notice that fitness related or outdoor recreational venues are quite popular in

this neighborhood. This cluster is a balanced mix of restaurants, fitness and recreational amenities as well commercial establishments.

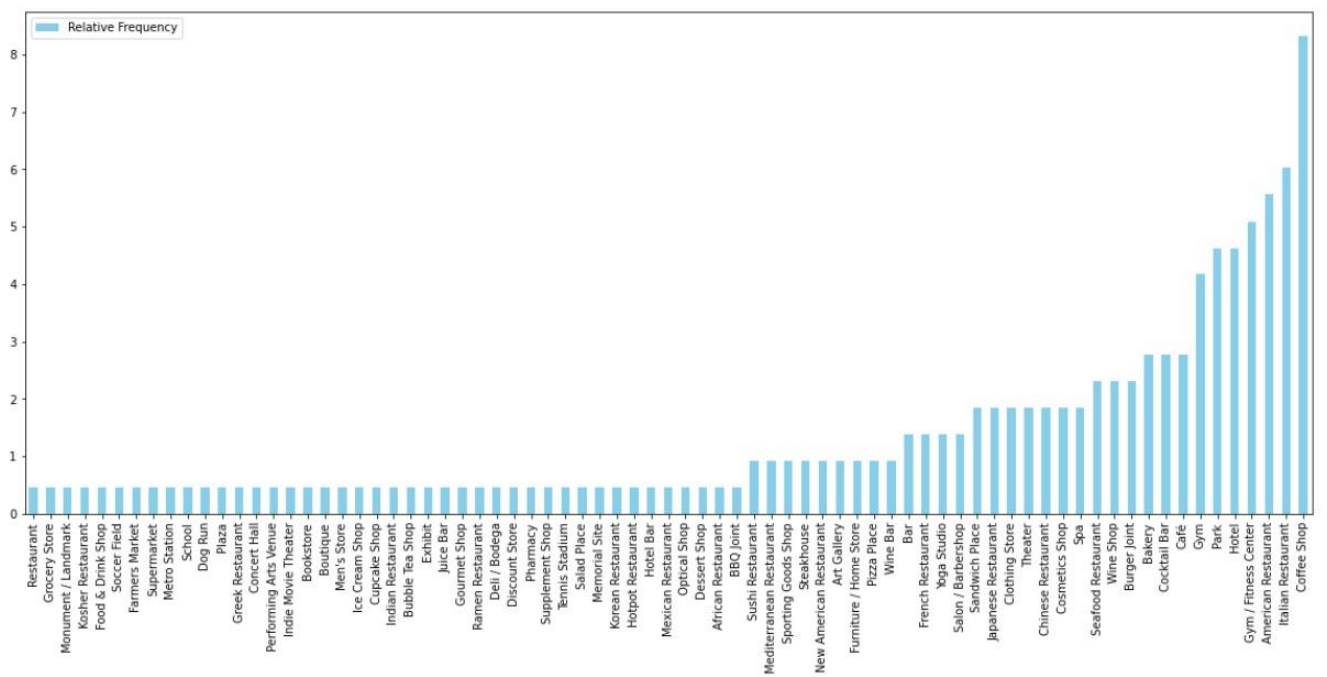
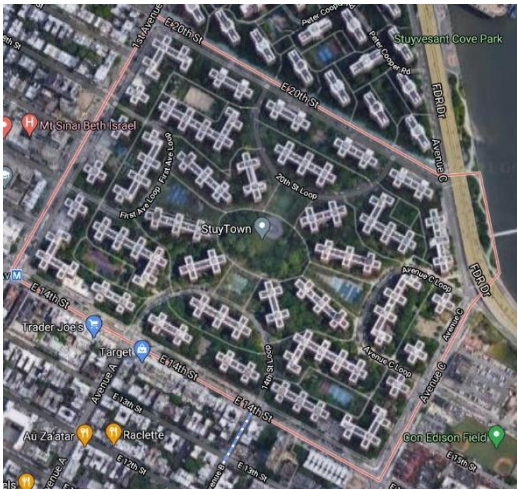


Figure 5 Frequency Distribution of popular venues in cluster 3

Since fitness related venues are quite popular, we can consider the neighborhoods in this cluster as well as a promising location. Please note that the neighborhood of Marble Hill already has a Supplement Store. However, the remaining neighborhoods may prove to be promising, but further analysis is required to pin point the neighborhoods.

Focusing more on cluster 1, we find that Stuyvesant Town is a very unique neighborhood. **Stuyvesant Town–Peter Cooper Village** is a large, post-World War II private residential development on the east side of the New York City borough of Manhattan. The complex consists of 110 red brick apartment buildings on an 80-acre (32 ha) tract. This neighborhood has almost all sports grounds such as baseball field, tennis field etc. It also has a lot of recreational venues like parks, health and fitness



related establishments such as gyms, hospitals etc. A list of 16 most popular venues of Stuyvesant Town is given below.

Table 6 List of popular venues in Stuyvesant Town

	name	categories	lat	lng
0	Stuyvesant Cove Park	Park	40.732513	-73.973876
1	Con Ed Field	Baseball Field	40.729206	-73.973732
2	Stuyvesant Oval	Park	40.731665	-73.977835
3	StuyFitness	Gym / Fitness Center	40.729013	-73.976676
4	Pouring Ribbons	Cocktail Bar	40.729093	-73.978033
5	Stuyvesant Oval Fountain	Fountain	40.731987	-73.977685
6	Stuyvesant Town Greenmarket	Farmers Market	40.731429	-73.978222
7	The Roost	Bar	40.729084	-73.978638
8	Unleashed Spa & Self Pet Wash	Pet Service	40.728984	-73.978380
9	B Cup Cafe	Coffee Shop	40.728783	-73.978565
10	Perk Kafe	Coffee Shop	40.729645	-73.978775
11	New York Skyports Seaplane Base	Heliprot	40.734959	-73.972991
12	B-Side	Bar	40.728438	-73.978780
13	Rocks Off Cruises	Harbor / Marina	40.735434	-73.973098
14	Solar One	Park	40.734442	-73.974350
15	BP	Gas Station	40.735107	-73.974554

Apart from Supplement Store, other promising business ventures in this neighborhood are Juice Bar, Dance Studio, Yoga Studio and Salad Place.

However, it should be noted that other factors like the operational cost, taxes and other legal requirements should also be analyzed for this neighborhood.

Conclusion

The neighborhoods of Manhattan were studied to identify a promising location for a Supplement Store. The defining feature data obtained from the Foursquare API was used

for segmenting and clustering the neighborhoods into three groups using k-means clustering. Further analysis of the clusters suggests cluster 1 and cluster 3 as promising clusters. Among these neighborhoods Stuyvesant **Town** appears to be the best fit.