

The battle of the Neighborhoods

Derek W.

Section 1 Introduction

New York City, the center of the US, has always been the dream city for the international students and commuters. Manhattan as the heart of NYC has attracted peoples from all nations with various culture backgrounds. People who comes from Queens, Jersey City even Stamford struggle for a living on this amazing Island every single day. Students enjoyed their life in the world-famous universities like NYU and Colombia University beside the Hudson river. Most people choose to rent an apartment exactly in Manhattan in order to feel the heartbeat of the golden city, though the expenses are as extraordinary as the city himself.

So, suppose I am owning a consultant company for international students. They might have extreme curiosity about Manhattan and might face a lot of questions about the living conditions there. The fundamental problem in starting a new life in Manhattan must be which area should people choose to live. Students should acquire adequate information and keep a balance between personal living habits and financial budget. Therefore, it is my job to offer them reliable analysis based on the neighborhood information and provide appropriate suggestions for them. Such analysis will definitely save their time in searching repetitive information from the internet.

In this report we are focusing on the international students who wants to rent an apartment near some outdoor facilities. Thus, the need a home to at least brings them conveniences for their sporting habits. I want to help them choosing suitable apartments utilizing clustering analytics technique.

Section 2 Data Description

The data I will analyze basically consist of three parts.

The first one is the "2014 New York Neighborhood Names" dataset, which was provided by NYU Spatial Data Repository (<https://geo.nyu.edu/catalog/nyu-2451-34572>). By using such dataset, I can extract the Manhattan neighborhood data and visulize them to offer my clients a direct perception to the city area.

The second one is the Foursquare dataset (<https://developer.foursquare.com/docs/build-with-foursquare/>). By connecting the Foursquare API, I can filter the venues in the relevant neighborhood in order to fit the student's personalized requirements. I will extract all the venues in category of "Outdoors & Recreation" for the Manhattan borough.

The third one is the renting price data from RENTcafé website. I will scrape the average rent price data from the Manhattan rental market website (<https://www.rentcafe.com/average-rent-market->

[trends/us/ny/manhattan/](#)) and then match them into each neighborhood that I investigated. By utilizing these data, I will eventually give an integrated consideration on their living place selection.

Section 3 Methodology

First, we extract the neighborhood information including their coordinates from the NYU dataset mentioned in the previous section just as Table 1 shows.

Table 1

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Also, we can mark each neighborhood on the map of Manhattan using the python library Folium which shows in Fig 1.

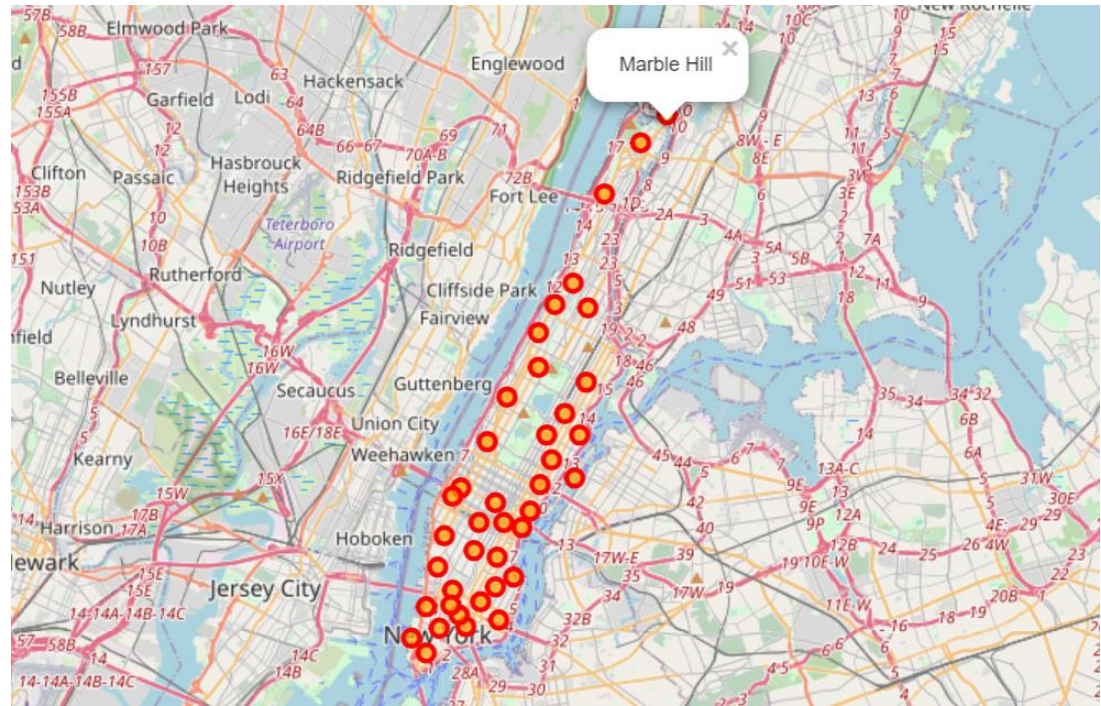


Fig. 1

We can see there are 40 neighborhoods distributed uniformly in the borough of Manhattan. Next, we scrape the rent price data from the RENTcafé website using the python library BeautifulSoup. The 10 cheapest neighborhoods are listed in Table 2.

Table 2

	Neighborhood	Average Rent
0	Marble Hill	\$1,708
1	Washington Heights	\$2,284
2	Inwood	\$2,375
3	Randalls and Wards Islands	\$2,394
4	Harlem	\$2,948
5	East Harlem	\$2,965
6	Roosevelt Island	\$3,416
7	Stuyvesant Town	\$3,611
8	Liberty Island	\$3,648
9	Governors Island	\$3,648
10	Ellis Island	\$3,648

In order to give our clients a direct perception of the rental, we using a bar chart in Fig. 2 to show all these prices in Manhattan neighborhoods.

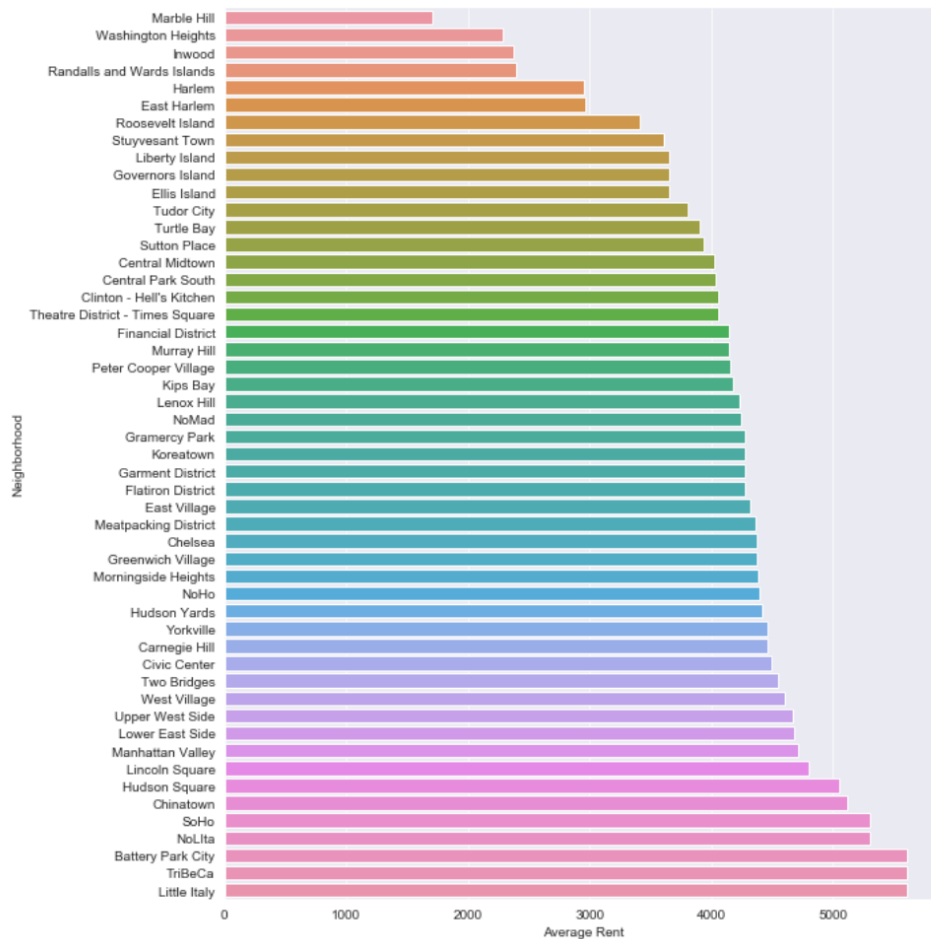


Fig. 2

Finally, we try to analyze each neighborhood using the K Means Clustering method and give our clients a reliable reference combining the average rental. Here we list 5 examples of the venues we required from through the Foursquare API in Table 3.

Table 3

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.894705	-73.847201	The Block	40.891606	-73.844401	Other Great Outdoors
1	Marble Hill	40.894705	-73.847201	The Horseshoe	40.885253	-73.849457	Other Great Outdoors
2	Marble Hill	40.894705	-73.847201	my neighbor park	40.895864	-73.844522	Playground
3	Marble Hill	40.894705	-73.847201	P.S 87 community park	40.895973	-73.847461	Playground
4	Marble Hill	40.894705	-73.847201	East 224th Street And Laconia Ave	40.884043	-73.849180	Scenic Lookout

After acquiring all the venues data from the Foursquare, we can simply using the clustering methods to analyze the implicit relationship among those neighborhoods and provide possible classifications. The ultimate Clusters are showed in Fig. 3.

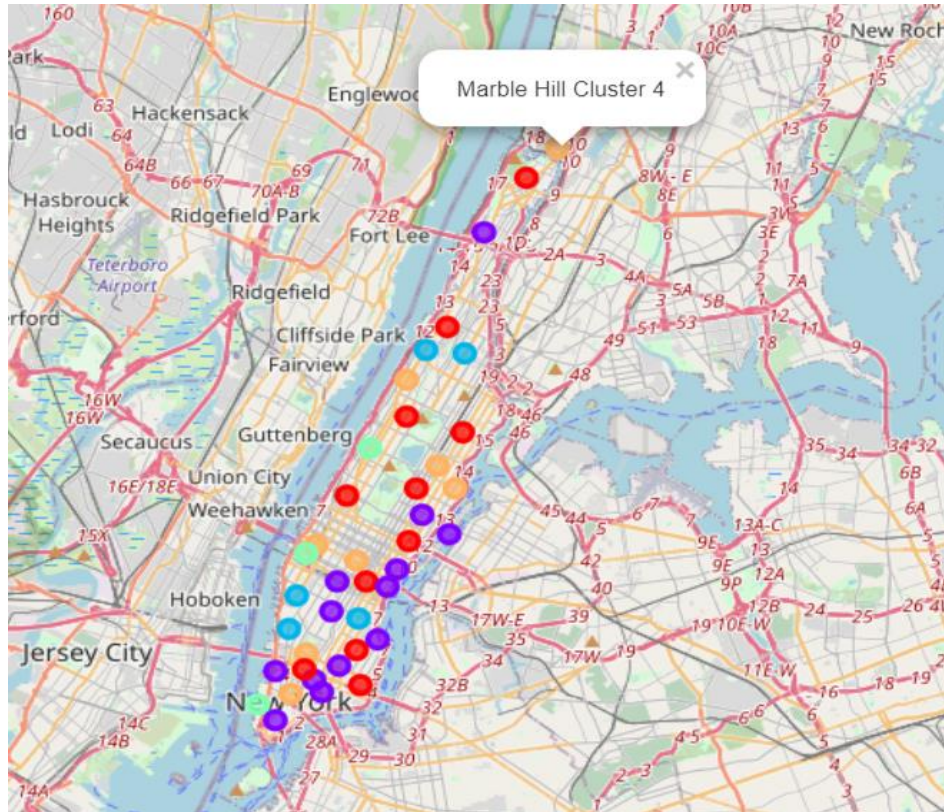


Fig. 3

We can see each color represent a single cluster. Those neighborhoods in the same cluster share some related common features.

Section 4 Results

In this section we list all the five clusters from Table 4 to Table 8 as follows. In the tables we can see 10 most common venues of each neighborhood and the average rent of the apartments. Our client can choose their future home according to the neighborhood feature with their acceptable expenditure in considerations.

Cluster 0:

Table 4

	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	Average Rent
3	Inwood	Park	Playground	Dog Run	Field	Other Great Outdoors	Plaza	Gym / Fitness Center	Athletics & Sports	Skating Rink	Pool	2394.0
4	Hamilton Heights	Park	Playground	Other Great Outdoors	Dog Run	Gym / Fitness Center	Gym	Plaza	Tennis Court	Pool	Athletics & Sports	2948.0
7	East Harlem	Garden	Playground	Park	Other Great Outdoors	Pool	Baseball Field	Field	Market	Beach	Bridge	3611.0
8	Upper East Side	Park	Dog Run	Gym	Garden	Playground	Cemetery	Plaza	Athletics & Sports	Gym / Fitness Center	Baseball Field	3648.0
13	Lincoln Square	Park	Garden	Baseball Field	Field	Gym / Fitness Center	Botanical Garden	Gym	Plaza	Cemetery	Dog Run	3941.0
16	Murray Hill	Park	Playground	Plaza	Gym / Fitness Center	Cemetery	Gym	Garden	Field	Athletics & Sports	Martial Arts Dojo	4053.0
19	East Village	Park	Playground	Other Great Outdoors	Bridge	Gym	Baseball Field	Plaza	Athletics & Sports	Skate Park	Historic Site	4143.0
20	Lower East Side	Playground	Park	Gym	Housing Development	Gym / Fitness Center	Residential Building (Apartment / Condo)	Other Great Outdoors	Garden	Baseball Field	Martial Arts Dojo	4154.0
23	Soho	Park	Playground	Baseball Field	Gym / Fitness Center	Plaza	Garden	Other Great Outdoors	Housing Development	Gym	Paintball Field	4243.0
25	Manhattan Valley	Park	Playground	Garden	Gym	Baseball Field	Residential Building (Apartment / Condo)	Martial Arts Dojo	Campground	Other Great Outdoors	Community Center	4276.0
34	Sutton Place	Park	Garden	Playground	Plaza	Gym / Fitness Center	Field	Athletics & Sports	Scenic Lookout	Botanical Garden	Other Great Outdoors	4416.0

Cluster 1:

Table 5

	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	Average Rent
1	Chinatown	Other Great Outdoors	Park	Baseball Field	Gym / Fitness Center	Playground	Pool	Plaza	Gym	Trail	Bridge	2284.0
2	Washington Heights	Park	Gym	Bridge	Baseball Field	Gym / Fitness Center	Athletics & Sports	Playground	Field	Cemetery	Bus Station	2375.0
10	Lenox Hill	Other Great Outdoors	Playground	Park	Gym / Fitness Center	Garden	Baseball Field	Pool	Plaza	Gym	Athletics & Sports	3648.0
11	Roosevelt Island	Gym	Playground	Other Great Outdoors	Pool	Gym / Fitness Center	Dog Run	Park	Outdoors & Recreation	Garden	Tennis Court	3808.0
21	Tribeca	Baseball Field	Gym	Playground	Park	Bridge	Garden	Other Great Outdoors	Housing Development	Gym / Fitness Center	Pool	4172.0
22	Little Italy	Baseball Field	Bridge	Playground	Park	Gym	Field	Other Great Outdoors	Dog Run	Pool	Garden	4231.0
29	Financial District	Playground	Gym / Fitness Center	Other Great Outdoors	Beach	Dog Run	Scenic Lookout	Bridge	Gym	Park	Pool	4363.0
31	Noho	Baseball Field	Gym / Fitness Center	Playground	Martial Arts Dojo	Other Great Outdoors	Cemetery	Gym	Park	Field	Outdoors & Recreation	4378.0
33	Midtown South	Gym	Playground	Gym / Fitness Center	Other Great Outdoors	Baseball Field	Field	Park	Pool	Cemetery	Outdoors & Recreation	4394.0
35	Turtle Bay	Other Great Outdoors	Scenic Lookout	Park	Gym / Fitness Center	Athletics & Sports	Tennis Court	Bridge	Gym	Dog Run	Playground	4458.0
36	Tudor City	Gym	Playground	Park	Other Great Outdoors	Baseball Field	Garden	Dog Run	Scenic Lookout	Pool	Field	4458.0
37	Stuyvesant Town	Gym / Fitness Center	Other Great Outdoors	Baseball Field	Gym	Playground	Tennis Court	Park	Dog Run	Weight Loss Center	Bridge	4489.0
38	Flatiron	Other Great Outdoors	Gym	Beach	Martial Arts Dojo	Park	Baseball Field	Vineyard	Cemetery	Farm	Building	4544.0

Cluster 2:

Table 6

	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	Average Rent
5	Manhattanville	Park	Gym / Fitness Center	Trail	Scenic Lookout	Dog Run	Plaza	Playground	Gym	Other Great Outdoors	Bridge	2965.0
6	Central Harlem	Park	Scenic Lookout	Other Great Outdoors	Gym / Fitness Center	Bridge	Basketball Court	Dog Run	Playground	Athletics & Sports	Gym	3416.0
17	Chelsea	Park	Other Great Outdoors	Playground	Gym	Basketball Court	Outdoors & Recreation	Trail	Dog Run	Tennis Court	Bridge	4059.0
24	West Village	Park	Other Great Outdoors	Dog Run	Baseball Field	Gym / Fitness Center	Gym	Pool	Plaza	Scenic Lookout	Paintball Field	4275.0
27	Gramercy	Park	Other Great Outdoors	Pool	Trail	Scenic Lookout	Playground	Track	River	Gym	Beach	4276.0

Cluster 3:

Table 7

	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	Average Rent
12	Upper West Side	Harbor / Marina	Beach	Park	Other Great Outdoors	Playground	Surf Spot	Pool	Scenic Lookout	Cemetery	Sports Club	3906.0
28	Battery Park City	Other Great Outdoors	Harbor / Marina	Park	Beach	Building	Residential Building (Apartment / Condo)	Tennis Court	Playground	Trail	Martial Arts Dojo	4320.0
39	Hudson Yards	Other Great Outdoors	Beach	Park	Harbor / Marina	Gym	Tennis Court	Farm	Scenic Lookout	Pedestrian Plaza	Building	4598.0

Cluster 4:

Table 8

	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	Average Rent
0	Marble Hill	Playground	Other Great Outdoors	Athletics & Sports	Gym / Fitness Center	Pool	Garden	Track	Scenic Lookout	Park	Financial or Legal Service	1708.0
9	Yorkville	Playground	Park	Other Great Outdoors	Dog Run	Gym	Scenic Lookout	Field	Cosmetics Shop	Plaza	Ski Area	3648.0
14	Clinton	Playground	Park	Garden	Plaza	Gym / Fitness Center	Basketball Court	Gym	Other Great Outdoors	Athletics & Sports	Campground	4022.0
15	Midtown	Playground	Park	Bridge	Other Great Outdoors	Gym / Fitness Center	Basketball Court	Plaza	Garden	State / Provincial Park	River	4038.0
18	Greenwich Village	Park	Other Great Outdoors	Playground	Bridge	Scenic Lookout	Gym	Garden	Dog Run	Residential Building (Apartment / Condo)	Community Center	4142.0
26	Morningside Heights	Park	Playground	Basketball Court	Other Great Outdoors	Dog Run	Gym	Athletics & Sports	Scenic Lookout	Field	Residential Building (Apartment / Condo)	4276.0
30	Carnegie Hill	Playground	Other Great Outdoors	Park	Gym	Plaza	Basketball Court	Dog Run	Martial Arts Dojo	Residential Building (Apartment / Condo)	Campground	4370.0
32	Civic Center	Playground	Park	Other Great Outdoors	Gym	Garden	Zoo	College Gym	Gym / Fitness Center	Housing Development	Office	4388.0

As we can see in the tables, there are 11 neighborhoods in the Cluster 0 many of their venues are public outdoor park places, and the rental are varied from \$2394 to \$4416. There are a lot of options in these neighborhoods.

The Cluster 1 includes 13 neighborhoods, and the venues there are varied, almost have all kinds of sports places. It will be a good choice for students who want to try different sports every day. The rentals are varied from \$2284 to \$4544.

All the 5 neighborhoods in Cluster 2 have lots of parks. It will be definitely welcomed by those who like go jogging and enjoy the sunshine. The rentals are good.

The 3 neighborhoods in Cluster 3 are special. They have beach and harbor views in common. Folks who love to swim shall consider such places. The rentals are higher than the other clusters, varying from \$3906 to \$4598.

The last 8 neighborhoods are in Cluster 4, which has lots of playground. Young folks might love these places. And the rental in Marble Hill is the lowest which is only \$1708.

Section 5 Discussions

For the advantages of the analysis:

1. We extract the data from the internet so that we can use almost whatever factors we want to consider. For example, we can even take the crime data from NYPD website of traffic convenience into consideration.
2. The K Means Clustering methods as a unsupervised learning method can generate the result automatically without any subjective preference.
3. The Foursquare API does offers us large amount of geographic information that can help our client for some personalized purposes.

For the limitations of the analysis:

1. The data we use are not the recent updated ones, for example the 2016 New York neighborhood data, since the data sources are limited.
2. Using the K Means Clustering in such contents might lead some issues in cluster explanatory. Sometimes the differences between clusters are not that distinct. So, the method is still a rough one that cannot reach a high precision analytics need.

Section 6 Conclusion

In conclusion, by utilizing such method, we can easily extract the useful information from the internet to fit out clients' need. And then, use a simply clustering method to shrink the scope of their choice so that they can simply find their ideal position in a relatively small clusters according to the venue features. With the reference of the rental from professional website, our clients can simply being targeted to focus on some neighborhoods they like to start their new life. Such analytics do save their time on finding a suitable home, especially for those who have limit understanding for the borough of Manhattan. Hope our Client can start an amazing life in the amazing New York City!