Battle of Neighborhoods in Manhattan

1. Introduction

New York City is one of the largest cities in the world with 8 million people. There are five boroughs: Manhattan, Brooklyn, Queens, Bronx, and Staten Island. As a resident in NYC, I decided to explore this city in my project. One interesting idea would be, if one wants to purchase an apartment in New York City, what are the best neighborhood to consider? The tradeoff is usually between price and convenience. The more centrally located neighborhoods with a lot of restaurants and grocery stores are often more expensive, while more affordable neighborhoods tend to be further from the city center, less convenient with less shops and restaurants.

With all the above in mind, I will create a ranking of neighborhoods according to their housing prices, distance to city center, number of shops, restaurants, and grocery stores. For simplicity, I will only consider the neighborhoods in Manhattan.

2. Data

The data to explore this question come from below sources:

- Housing price data

Housing price data is downloaded from StreetEasy Data Dashboard: https://streeteasy.com/blog/data-dashboard/. It provides downloadable csv files for price index, median sale price, median asking price for each neighborhood. Since median sale price data has a lot of missing values if there is no sales in a given month, I will just use the median asking price as a gauge of housing price. The data is a time-series data of monthly prices and I just use the latest available month, which is December 2019.

- Venue data

I use Foursquare API to get venue data such as the number of shops, restaurants and grocery stores in each neighborhood.

- Distance data

I use Google map to get distance to city center data for each neighborhood. The official New York City coordinates are those of the city hall, but most professional people work around midtown, and when we consider distance to center, we are

mostly concerned about commute for work, so I use midtown Manhattan coordinates 40.7549° N, 73.9840° W to calculate distance to center.

3. Methodology

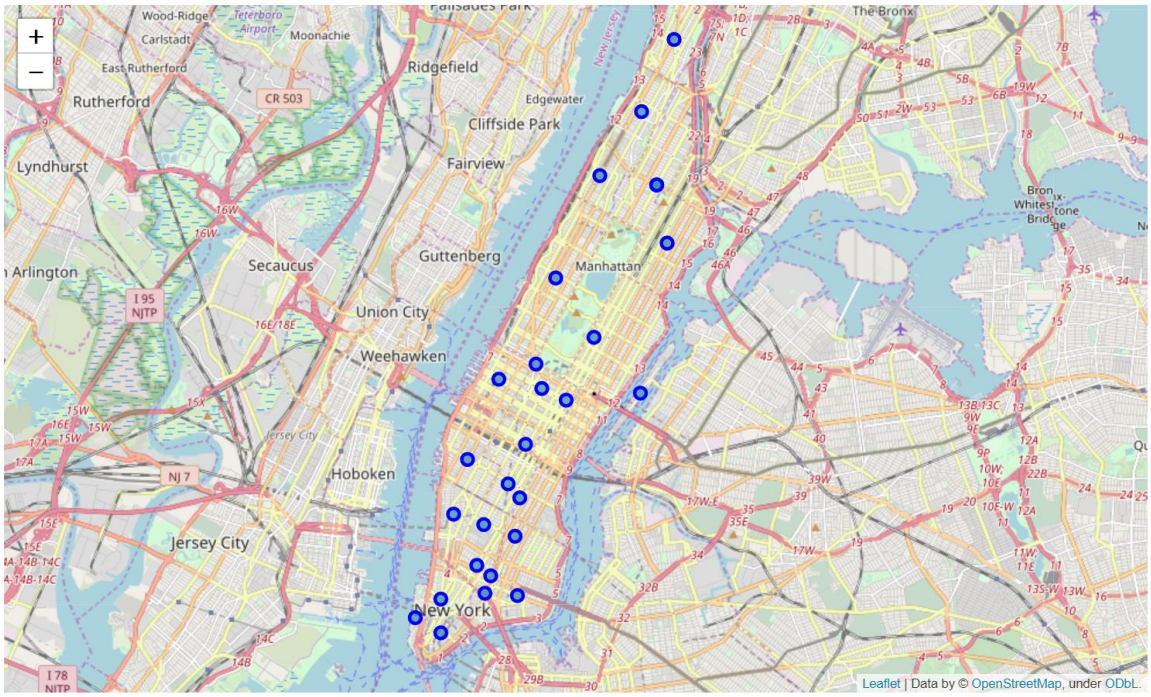
3.1 Import and clean data

I first download the Manhattan median asking price by neighborhood data from StreetEasy and combine it with distance to center data from Google Map. There are 32 neighborhoods in Manhattan. Then I clean the data by dropping the neighborhoods that have missing values. We now have 28 neighborhoods.

	Neighborhood	Price	DistanceToCenter	Latitude	Longitude
0	Battery Park City	1340000.0	3.8	40.711017	-74.016937
1	Central Harlem	980000.0	4.5	40.807879	-73.945415
2	Central Park South	2345000.0	0.9	40.767750	-73.981318
3	Chelsea	2275000.0	1.4	40.746491	-74.001528
4	Chinatown	1995000.0	3.0	40.716491	-73.996250
6	East Harlem	950000.0	4.3	40.794722	-73.942500
7	East Village	1250000.0	2.4	40.729269	-73.987361
8	Financial District	1397000.0	3.7	40.707612	-74.009378
9	Flatiron	2695000.0	1.1	40.741059	-73.989642
10	Gramercy Park	1715000.0	1.5	40.737925	-73.985932
11	Greenwich Village	1975000.0	1.8	40.731980	-73.996566
12	Hamilton Heights	780000.0	5.6	40.824145	-73.950062
13	Inwood	487500.0	8.9	40.869258	-73.920495
15	Lower East Side	1397500.0	3.3	40.715936	-73.986806
17	Midtown	2225000.0	0.0	40.762268	-73.979544
18	Midtown East	950000.0	0.9	40.759822	-73.972471
19	Midtown South	1565550.0	0.3	40.749842	-73.984251
20	Midtown West	1492500.0	0.9	40.764423	-73.992392
21	Morningside Heights	799000.0	4.3	40.810000	-73.962500
22	Nolita	4750000.0	2.5	40.720370	-73.994582
23	Roosevelt Island	1089000.0	3.4	40.761418	-73.950228
24	Soho	3495000.0	2.5	40.722880	-73.998750
26	Tribeca	4274500.0	3.0	40.715380	-74.009306
27	Upper East Side	1825000.0	2.5	40.773702	-73.964120
28	Upper West Side	1972500.0	2.5	40.787045	-73.975416
29	Washington Heights	547000.0	6.8	40.840198	-73.940221
30	West Harlem	1155000.0	5.1	40.807879	-73.945415
31	West Village	2100000.0	1.9	40.734186	-74.005580

3.2 Get the latitude and longitude coordinates of each neighborhood and create map

I use geocoder to convert the neighborhoods into their latitude and longitude coordinates. Then I use folium to create a map of Manhattan neighborhoods:



3.3 Foursquare data

I use the Foursquare API to find venues of interest in each neighborhood. In this project, we are interested in two categories: restaurants, and grocery stores, as these are the proxy variables for convenience of daily life. I define the **limit=100** and **radius=500 meters**. The search returns 2233 restaurants and 310 grocery stores in Manhattan.

I then group them by neighborhood to calculate the number of restaurants and number of grocery stores in each neighborhood. Now we have the complete data for our analysis. Below is the summary statistics of our neighborhood data. We can see the average housing price is $1,779,359, average distance to city center is 2.96 miles, average number of restaurants is 80, and average number of grocery stores is 11 per neighborhood.

	Price	DistanceToCenter	Latitude	Longitude	NumRestaurant	NumGrocery
count	2.800000e+01	28.000000	28.000000	28.000000	28.000000	28.000000
mean	1.779359e+06	2.957143	40.761105	-73.978035	79.750000	11.071429
std	1.031887e+06	2.012264	0.042684	0.025049	25.383904	13.213149
min	4.875000e+05	0.000000	40.707612	-74.016937	14.000000	1.000000
25%	1.061750e+06	1.475000	40.727672	-73.996329	64.000000	5.000000
50%	1.529025e+06	2.500000	40.754832	-73.985092	100.000000	7.000000
75%	2.131250e+06	3.925000	40.788965	-73.959432	100.000000	11.250000
max	4.750000e+06	8.900000	40.869258	-73.920495	100.000000	65.000000

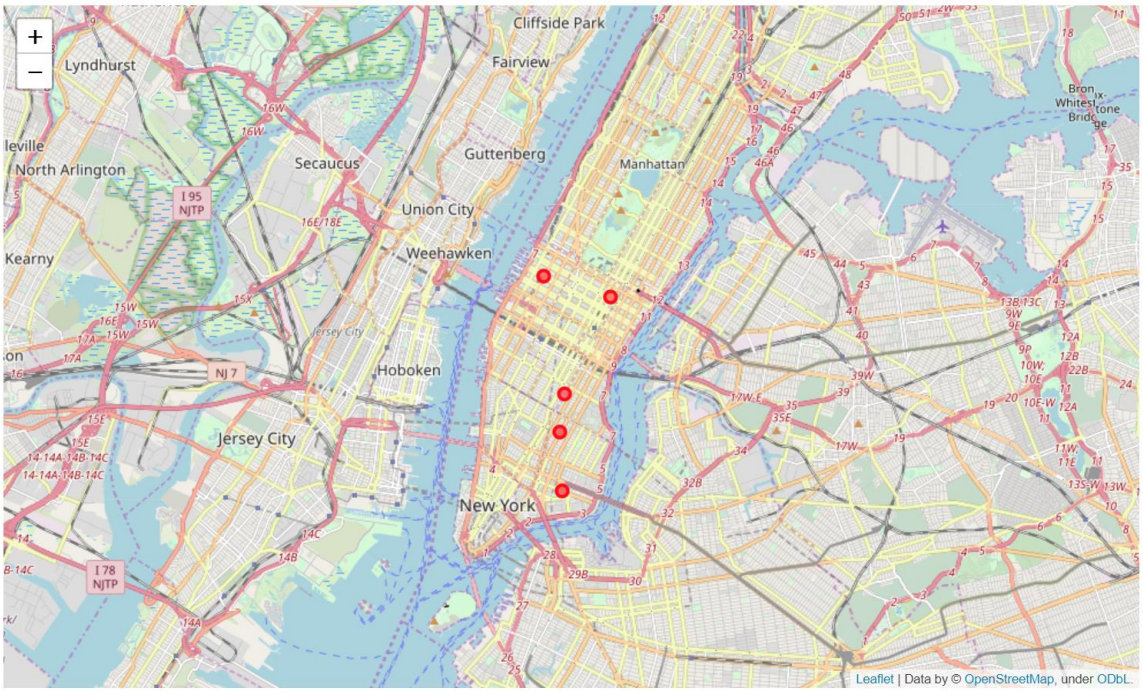
3.4 Rank neighborhoods

Now we have the complete data for each neighborhood, including housing price, distance to center, number of restaurants and number of grocery stores. Next, we will rank the neighborhoods by these four metrics. I do this by first ranking each neighborhood according to each individual metric. Then assuming equal weight for each metric, I will sum up all four ranks to get an aggregate score. The smaller the aggregate score, the higher the rank of a particular neighborhood.

The results show that the top 5 neighborhoods to consider to purchase an apartment are: **East Village, Midtown West, Lower East Side, Gramercy Park, and Midtown East.** They represent the overall most affordable, closest to city center and most convenient neighborhoods in Manhattan.

	Neighborhood	Price	DistanceToCenter	Latitude	Longitude	NumRestaurant	NumGrocery	OverallRank
6	East Village	1250000.0	2.4	40.729269	-73.987361	100	15	33.0
17	Midtown West	1492500.0	0.9	40.764423	-73.992392	100	7	41.5
13	Lower East Side	1397500.0	3.3	40.715936	-73.986806	100	16	42.0
9	Gramercy Park	1715000.0	1.5	40.737925	-73.985932	100	10	42.5
15	Midtown East	950000.0	0.9	40.759822	-73.972471	100	3	42.5
16	Midtown South	1565550.0	0.3	40.749842	-73.984251	100	6	44.0
4	Chinatown	1995000.0	3.0	40.716491	-73.996250	100	65	45.5
7	Financial District	1397000.0	3.7	40.707612	-74.009378	100	12	47.0
8	Flatiron	2695000.0	1.1	40.741059	-73.989642	100	11	47.5
10	Greenwich Village	1975000.0	1.8	40.731980	-73.996566	100	9	48.0
14	Midtown	2225000.0	0.0	40.762268	-73.979544	100	6	50.0
2	Central Park South	2345000.0	0.9	40.767750	-73.981318	100	7	51.5
19	Nolita	4750000.0	2.5	40.720370	-73.994582	100	44	51.5
21	Soho	3495000.0	2.5	40.722880	-73.998750	100	13	53.5
27	West Village	2100000.0	1.9	40.734186	-74.005580	100	7	54.5
5	East Harlem	950000.0	4.3	40.794722	-73.942500	49	14	57.5
3	Chelsea	2275000.0	1.4	40.746491	-74.001528	66	10	60.5
24	Upper West Side	1972500.0	2.5	40.787045	-73.975416	64	11	61.5
0	Battery Park City	1340000.0	3.8	40.711017	-74.016937	70	7	64.5
25	Washington Heights	547000.0	6.8	40.840198	-73.940221	49	8	66.5
1	Central Harlem	980000.0	4.5	40.807879	-73.945415	68	5	71.5
26	West Harlem	1155000.0	5.1	40.807879	-73.945415	68	5	74.5
23	Upper East Side	1825000.0	2.5	40.773702	-73.964120	54	5	75.5
11	Hamilton Heights	780000.0	5.6	40.824145	-73.950062	64	1	78.0
18	Morningside Heights	799000.0	4.3	40.810000	-73.962500	38	3	78.5
22	Tribeca	4274500.0	3.0	40.715380	-74.009306	88	6	78.5
12	Inwood	487500.0	8.9	40.869258	-73.920495	41	3	80.0
20	Roosevelt Island	1089000.0	3.4	40.761418	-73.950228	14	1	82.5

We can visualize the results on the map:



4. Results and Discussion

In the above analysis, we collect and create data for each of the 32 neighborhoods in Manhattan, New York their median housing price, distance to city center, number of restaurants and number of grocery stores. We then rank the neighborhoods by each of the four metrics that we think are most relevant in making decision on purchasing an apartment. Finally, we combine the four ranks by equal weight to get an overall rank. We found that the top 5 neighborhoods in Manhattan are East Village, Midtown West, Lower East Side, Gramercy Park, and Midtown East. These candidates overall represent the neighborhoods that are most affordable, centrally located, and abundant in restaurants and grocery stores.

Of course, we can add more metrics to evaluate the neighborhoods, depending on the target client's preferences. For example, we can add the number of subway stations nearby to represent the convenience of commute, number of movie theaters for movie goers, number of gyms for athletic people, number of bars to represent the abundance of nightlife, or school zone rating for families with school-aged children.

In addition, we can also apply different weightings for the selected metrics. One can easily adjust up or down the individual weighting of each metric according to their own preferences. For example, if the target client is more sensitive to housing price, then we can apply a higher weight on the price metric etc. The model is very flexible.

5. Conclusion

In this project, we try to answer the question: what's the best neighborhood to consider if we want to purchase an apartment in Manhattan? We present a simple approach by collecting data of four different metrics: housing price, distance to city center, number of restaurants, and number of grocery stores. By leveraging Foursquare API data, we are able to calculate the number of restaurants and grocery stores in each neighborhood. We then rank the neighborhood by each metric and sum them together to get the overall score. We find that the best neighborhood options are East Village, Midtown West, Lower East Side, Gramercy Park, and Midtown East.

The model is flexible and can be easily modified according the stakeholder's preferences. Using the same framework, we can collect data on different metrics and apply different weights on each metric. We can also cluster the neighborhoods to find commonalities among the neighborhoods as an extension of the project.