

Find the Best Neighborhood for Opening a Pet Service Store in Manhattan

Cindy Dong

April 23, 2020

1 Introduction

Manhattan, New York is probably the most crowded place in US with over 1.6 million people living in a land area of 22.8 square miles. There are lots of opportunities of doing businesses here since so many people need all kinds of services. However, it is also very challenging to do business here because of the high rental price and many competitors. Considering the diversity of 40 neighborhoods in Manhattan, the first question is always “Where should I open my business?”

My client Tina plans to open her pet service store in Manhattan. The pet service store will provide grooming services on-site and at customer’s home. It may also expand the business to pets walking and boarding. I will use Manhattan location data to help her pick the right neighborhood for the new store, which has many potential customers nearby, fewer competitors and a reasonable rental price. Once the neighborhood is selected, Tina’s business risk and effort on searching the retail rental space can be greatly reduced.

2 Data Description

2.1 Manhattan location and venue data

The first data set is got from Foursquare API. The data set provides the list of 40 neighborhoods in Manhattan, their precise latitudes, longitudes, and venue information. For Tina’s pet service store, I don’t need all the venue information. The most relevant venues are pet related venues, such as pet service, pet store and pet café because they are potential competitors. Pet café may also be

potential customers if they don't groom their pets by themselves. But they are more likely to be competitors since they are able to offer pet boarding and other services. Park is also important because dog owners would like to walk their dogs there and very likely will visit the pet service stores near the park. Tina will also prefer parks nearby if she wants to expand her business to pets walking and boarding in the future.

Other related venues include gym, yoga studio and spa if the pet owners want to do some excises instead of going back home and waiting while their pets are in the store. Those venues also include supermarkets and shopping malls in case the pets owners would like to go shopping. Coffee shop, bakery, tea room and sandwich shop are other choices for pet owners to take a nibble. People more likely to choose a store or service if it is convenient for them.

The Manhattan venue data are cleaned up with only the related features left in the data frame as in the picture below.

	Neighborhood	Pet Service	Pet Store	Pet Café	Park	Yoga Studio	Gym / Fitness Center	Spa	Supermarket	Grocery Store	Shopping Mall	Farmers Market	Women's Store	Coffee Shop	Bubble Tea Shop	Tea Room	Bak
0	Battery Park City	0	1	0	6	0	0	0	0	1	2	0	0	1	0	0	
1	Carnegie Hill	0	1	0	0	3	3	2	1	2	0	0	0	8	0	0	
2	Central Harlem	0	0	0	1	0	2	1	0	0	0	0	0	0	0	0	
3	Chelsea	0	0	0	2	0	0	0	0	1	0	0	1	7	0	0	

2.2 Manhattan retail rental price

Another important data set is the retail rental price in Manhattan. Tina would like to open her store where retail rental price is reasonable. I got the information from rebny.com [1]. It has the most recent ground floor retail rental price per square feet in different Manhattan neighborhoods from spring 2018 to spring 2019. However those locations don't cover all the neighborhoods as listed in the Foursquare data. I manually map those prices to all neighborhoods by finding the nearest location in the data table to each neighborhood using google map search. The data is also sorted by neighborhood in alphabet order, which looks like the figure below.

1	Neighborhood	Spring 2019	Fall 2018	Spring 2018	Average Price in USD
2	Battery Park City	401	338	368	369
3	Carnegie Hill	1039	1160	1390	1196
4	Central Harlem	137	140	124	134
5	Chelsea	277	303	313	298
6	Chinatown	544	558	595	566
7	Civic Center	544	558	595	566
8	Clinton	1936	1860	1938	1911
9	East Harlem	137	140	124	134

2.3 Manhattan apartment rental price

The third data set is the apartment rental price in Manhattan I grabbed from rentcafe.com [2]. It provides the latest average apartment rental price in all Manhattan neighborhoods updated by February, 2020 for an average size of 703 square feet apartment. From those data, I will know about the income level of potential customers. People with higher income usually would like to spend more money on pet services. The data is sorted by neighborhood in alphabet order as the figure below.

1	Neighborhood	Average Rent
2	Battery Park City	5605
3	Carnegie Hill	4458
4	Central Harlem	2965
5	Chelsea	4370
6	Chinatown	5116
7	Civic Center	4489
8	Clinton	4053

3 Methodology

3.1 Gather venue information in Manhattan neighborhoods related to pet store

From file “newyork_data.json” which is available from previous course lab, I was able to extract New York neighborhood list and their latitudes and longitudes information can be got using geopy library.

Once I have the list of 40 neighborhoods of Manhattan and their geographical coordinates, I can use Foursquare API to explore nearby top 100 venues within a

radius of 500 meters to all the neighborhoods in Manhattan. There are 322 different venue categories. For our goal: find the best neighborhood to open Tina's new pet service store, I only need those venues related to pet service as described in the data section.

To further simplify the data, I group venues in the same categories and sum their store number in each Manhattan neighborhood. After adding geographical coordinates, retail rental price and apartment rental price, the cleaned-up dataframe looks like the picture below, where "Competitor" column records the number of pet stores in each neighborhood.

	Neighborhood	Competitor	Park	Fitness	Shopping	Cafe	Retail Rental	Apartment Rental	Latitude	Longitude
0	Battery Park City	1	6	0	2	2	369	5605	40.876551	-73.910660
1	Carnegie Hill	1	0	7	3	12	1196	4458	40.715618	-73.994279
2	Central Harlem	0	1	3	0	1	134	2965	40.851903	-73.936900
3	Chelsea	0	2	0	2	14	298	4370	40.867684	-73.921210
4	Chinatown	0	0	4	3	18	566	5116	40.823604	-73.949688

3.2 Analyze relationship between pet stores and related neighborhood features

I sorted the data by "Competitor" and found the first 18 neighborhoods have pet stores and the rest of 22 neighborhoods have no pet stores as the picture below.

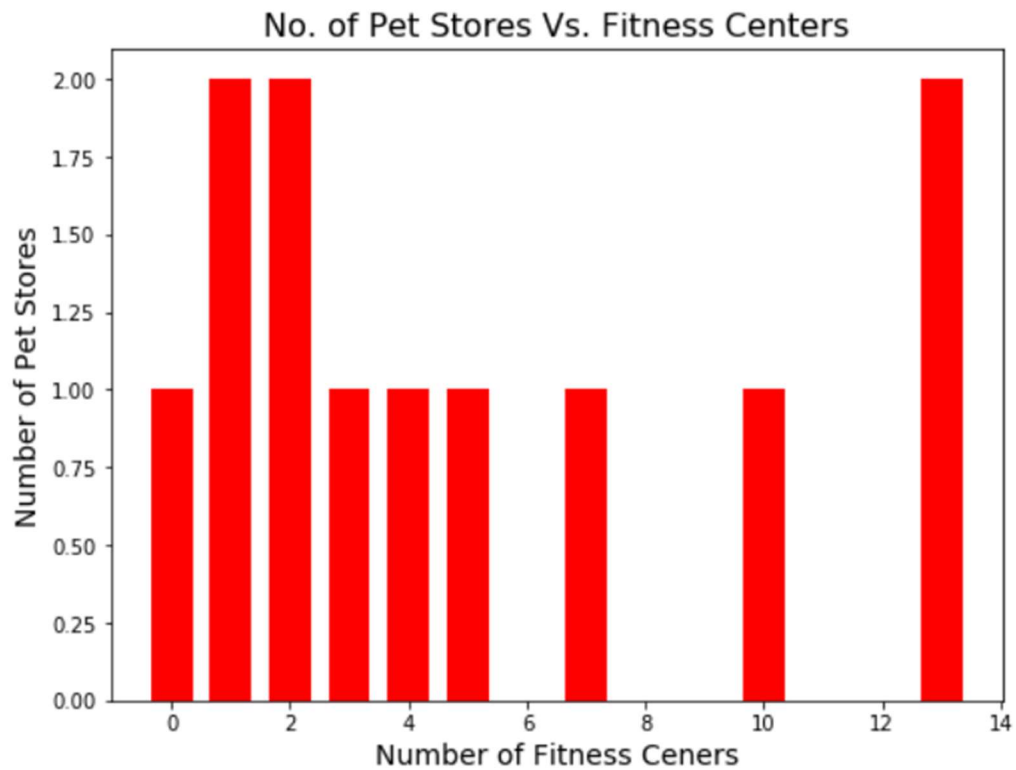
	Neighborhood	Competitor	Park	Fitness	Shopping	Cafe	Retail Rental	Apartment Rental	Latitude	Longitude
0	Washington Heights	2	2	2	7	11	134	2284	40.731000	-73.974052
1	East Village	2	1	1	2	11	298	4320	40.775639	-73.960508
2	Flatiron	2	3	13	3	7	405	4276	40.768113	-73.958860
3	Battery Park City	1	6	0	2	2	369	5605	40.876551	-73.910660
4	Inwood	1	2	1	3	7	134	2375	40.754691	-73.981669
5	Murray Hill	1	0	4	2	9	597	4143	40.808000	-73.963896
6	Noho	1	0	0	4	14	566	4394	40.737210	-73.981376
7	Carnegie Hill	1	0	7	3	12	1196	4458	40.715618	-73.994279
8	Lower East Side	1	1	1	1	6	298	4676	40.727847	-73.982226
9	Little Italy	1	0	4	2	15	369	5607	40.726933	-73.999914

Tina would prefer to open her new business in one of those 22 neighborhoods, which have no competitor now. Before that, I want to know the relationship between number of pet stores and its related features in the first 18 neighborhoods where people have already run their pet store businesses.

- First, plot number of pet stores versus number of parks in each neighborhood as the picture below. I don't see any obvious linear relationship between them.



- Secondly, plot number of pet stores versus number of fitness centers in each neighborhood as the picture below. I still don't see any obvious linear relationship between them.



- Thirdly, plot number of pet stores versus retail rental price in each neighborhood as the picture below. I still don't see any obvious linear relationship between them.



Now I know the relationship between number of pet stores and all the related features in the data is complicated. I need more advanced tool such as machine learning to model it.

3.3 Machine learning to predict potential pet store number

- Clean and Normalize data for machine learning

Before doing machine learning, I cleaned the data by dropping column “Neighborhood”, “Competitor”, “Latitude” and “Longitude”. Then I normalized the data by the maximum number of each column so each feature has similar weight in the model. Finally, the data is split into training data and test data for model input. The “Competitor” column in the original data is split in the same way as training and testing data as data label for prediction.

The training data includes first 18 rows which neighborhoods have pet stores. The training data will be used to train the model. The test data has the rest of 22 row which neighborhood has no pet stores currently and I will use the trained model to predict number of potential pet store in these neighborhoods. The training data looks like the picture below.

	Park	Fitness	Shopping	Cafe	Retail Rental	Apartment Rental
0	0.333333	0.153846	1.000000	0.611111	0.070120	0.407348
1	0.166667	0.076923	0.285714	0.611111	0.155939	0.770465
2	0.500000	1.000000	0.428571	0.388889	0.211931	0.762618
3	1.000000	0.000000	0.285714	0.111111	0.193093	0.999643
4	0.333333	0.076923	0.428571	0.388889	0.070120	0.423578

- Multiple Linear Regression

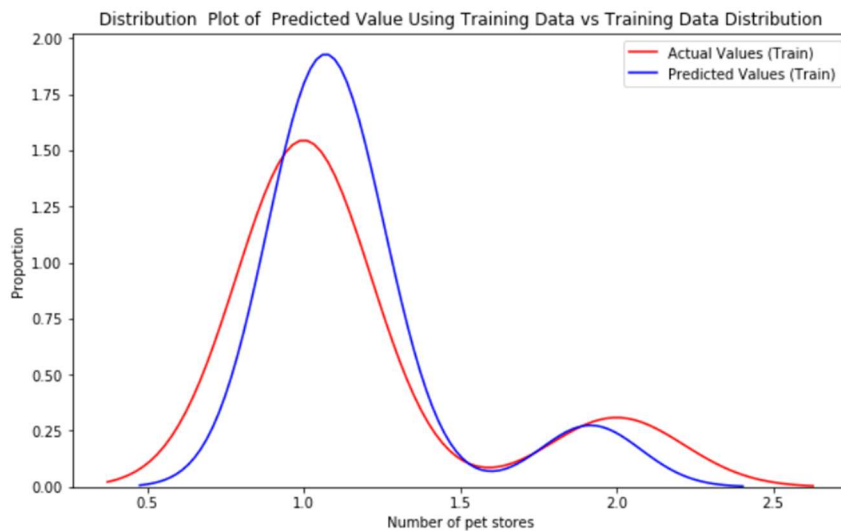
First, I use multiple linear regression to train the model. It's R squared is 0.462 and mean square error is 0.074.

- Ridge Regression

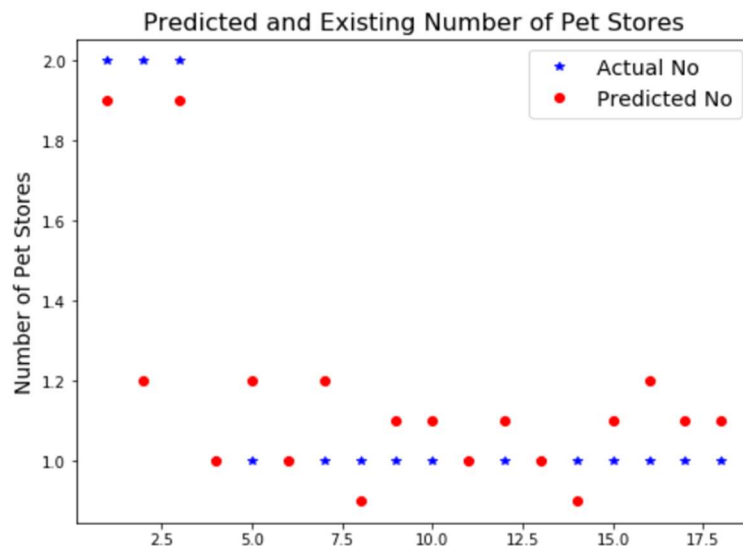
Then, I use ridge regression to train the model. Its R squared is 0.680 and mean square error is 0.044.

With higher R squared and lower mean square error, ridge regression is better than multiple linear regression to train my model.

I used ridge regression trained model to predict the number of pet stores for the training data, the distribution figure is as below. The model seems to be doing well in learning from the training dataset.



And Let's plot the predicted number on training data against the actual number of pet stores. The model seems to fit the training data reasonably well.



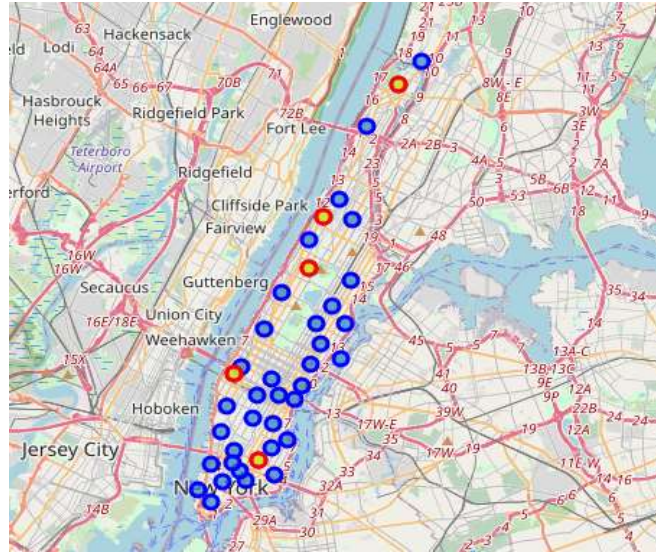
Since I didn't have a large data set, I skipped validating the model on validation data.

- Predict number of pet stores using ridge regression on test data

I predicted the number of potential pet stores on the test data and add the those numbers as column “Potential Pet Store” back to last 22 row of the original data set which neighborhood has no pet store currently. The new 22-rows data set is sorted by column “Potential Pet Store” in descending order to get the top 5 neighborhood which have highest number of potential pet stores as below. The number of potential pet store in the 5 neighborhoods are very close between 1.3 to 1.5. Note that I don’t round the number into integer because the number just reflects the potential of that neighborhood to have a new pet store.

	Neighborhood	Competitor	Park	Fitness	Shopping	Cafe	Retail Rental	Apartment Rental	Latitude	Longitude	Potential Pet Store
0	Sutton Place	0	4	8	2	7	241	3941	40.723259	-73.988434	1.5
1	Civic Center	0	3	10	1	6	566	4489	40.816934	-73.957385	1.4
2	Yorkville	0	2	3	1	15	377	4458	40.756658	-74.000111	1.3
3	Chelsea	0	2	0	2	14	298	4370	40.867684	-73.921210	1.3
4	Morningside Heights	0	4	0	3	5	134	4388	40.797307	-73.964286	1.3

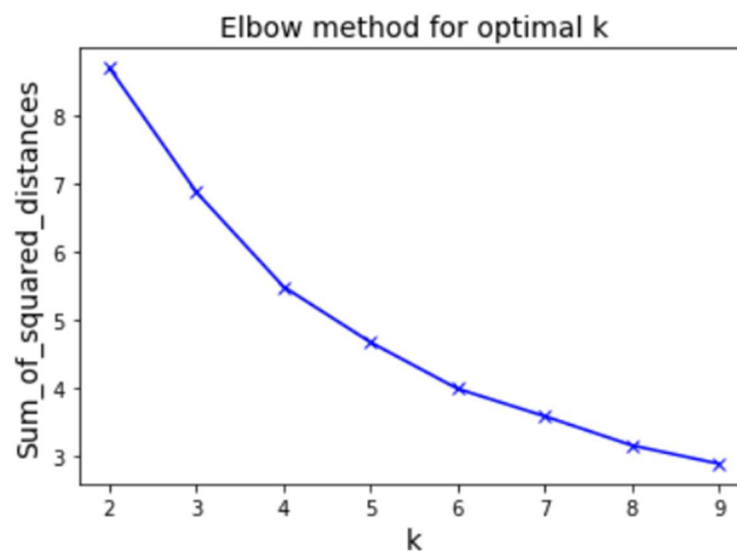
We can visualize the centers of these top 5 neighborhood on top of all Manhattan neighborhoods (marked in blue circle) on Manhattan map. Sutton Place, Civic center, Yorkville, Chelsea and Morningside Heights are marked in red circle. They spread out over Manhattan. Which one could be the best choice for Tina to open her new pet service store? I will analyze these neighborhoods with more details.



3.4 Cluster neighborhood to validate results

In order to further investigate the top 5 neighborhoods, I group all Manhattan neighborhoods based on their normalized features related to pet stores.

First, I run k-means clustering for 2 to 10 clusters and plot their sum of squared distances into elbow plot to choose a reasonable number of clusters. From the elbow plot below, we can see the drop in the sum of squared distance starts to slow down after $k=6$. Hence 6 is the optimal number of clusters for my analysis.

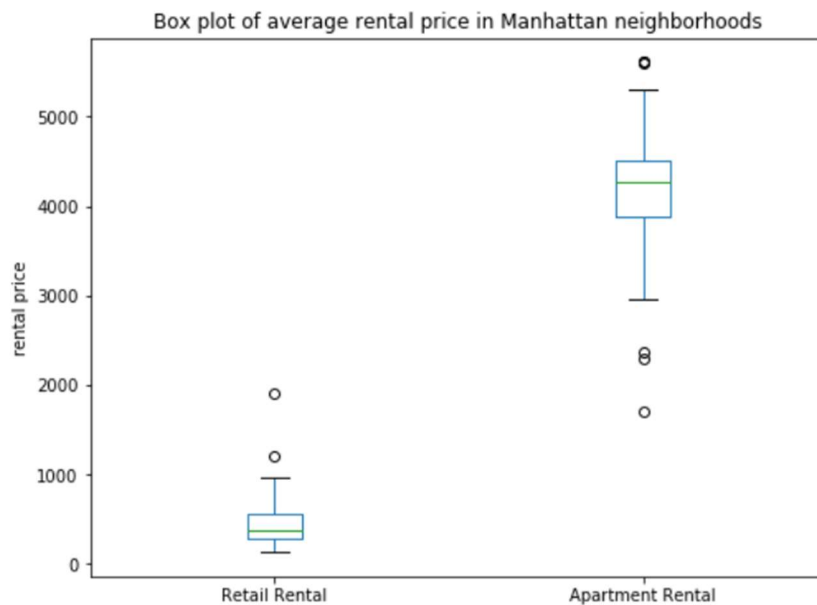


I rerun k-means clustering for 6 clusters and get 6 group of neighborhoods. I will discuss the clustering results in the results section.

4 Results

Using machine learning method, I got 5 neighborhood candidates for Tina to open her new pet service store. I also grouped all Manhattan neighborhoods into 6 clusters to analyze more details of those neighborhoods about their pet store related features.

First, let's review the range of retail and apartment rental price for all Manhattan neighborhoods. It is critical for my decision-making. Tina would like to open the new pet service store where the retail rental price is relatively low because it reduces the cost of business. Meanwhile she prefers the nearby apartment rental price is relatively high because rental price reflects the income level of potential customers. Customer with higher income tend to spend more money on their pets.



We can see from the box chart that the median retail rental price is 369 USD per square feet and median apartment rental price is 4276 USD.

Now let's check the details of each cluster.

- Cluster 1

	Cluster Labels	Neighborhood	Competitor	Park	Fitness	Shopping	Cafe	Retail Rental	Apartment Rental	Latitude	Longitude
7	0	Carnegie Hill	1	0	7	3	12	1196	4458	40.715618	-73.994279
25	0	Midtown	0	1	5	1	10	971	4022	40.722184	-74.000657
34	0	Clinton	0	1	8	1	11	1911	4053	40.815976	-73.943211

None of top 5 potential neighborhood is in this cluster.

- Cluster 2

	Cluster Labels	Neighborhood	Competitor	Park	Fitness	Shopping	Cafe	Retail Rental	Apartment Rental	Latitude	Longitude
1	1	East Village	2	1	1	2	11	298	4320	40.775639	-73.960508
5	1	Murray Hill	1	0	4	2	9	597	4143	40.808000	-73.963896
6	1	Noho	1	0	0	4	14	566	4394	40.737210	-73.981376
9	1	Little Italy	1	0	4	2	15	369	5607	40.726933	-73.999914
11	1	Lenox Hill	1	0	3	2	10	241	4231	40.748303	-73.978332
16	1	Upper West Side	1	0	0	1	10	301	4668	40.746917	-73.971219
18	1	Soho	0	0	3	2	9	369	5301	40.707107	-74.010665
30	1	Greenwich Village	0	2	3	1	15	566	4378	40.787658	-73.977059
31	1	Gramercy	0	2	3	3	9	405	4275	40.762160	-73.949168
36	1	Chinatown	0	0	4	3	18	566	5116	40.823604	-73.949688
37	1	Chelsea	0	2	0	2	14	298	4370	40.867684	-73.921210
39	1	Yorkville	0	2	3	1	15	377	4458	40.756658	-74.000111

Neighborhood Chelsea and Yorkville are in the same cluster with East Village which already has 2 pet stores. That means they have similar features for running successful pet store businesses.

Chelsea has 2 parks. Its retail rental price is 298 USD, which is lower than the median price and its apartment rental price is 4370 USD, which is a little higher than the median price.

Yorkville also has 2 parks. Its retail rental price is 377 USD, which is around the median price and its apartment rental price is 4458 USD, which is a little higher than the median price.

- Cluster 3

	Cluster Labels	Neighborhood	Competitor	Park	Fitness	Shopping	Cafe	Retail Rental	Apartment Rental	Latitude	Longitude
4	2	Inwood	1	2	1	3	7	134	2375	40.754691	-73.981669
8	2	Lower East Side	1	1	1	1	6	298	4676	40.727847	-73.982226
10	2	Lincoln Square	1	1	5	2	3	294	4795	40.744035	-74.003116
12	2	Stuyvesant Town	1	2	1	1	1	405	3611	40.782683	-73.953256
19	2	Roosevelt Island	0	2	1	2	2	377	3416	40.711932	-74.016869
20	2	Midtown South	0	0	6	2	4	597	4276	40.734434	-74.006180
24	2	Manhattan Valley	0	1	3	2	4	294	4711	40.717807	-73.980890
26	2	Marble Hill	0	0	1	0	6	134	1708	40.719324	-73.997305
27	2	Manhattanville	0	2	0	1	6	134	2965	40.721522	-74.010683
28	2	Hudson Yards	0	2	4	1	2	948	4416	40.759101	-73.996119
29	2	Hamilton Heights	0	1	2	0	8	134	2965	40.773529	-73.985338
33	2	East Harlem	0	1	1	1	5	134	2965	40.792249	-73.944182
38	2	Central Harlem	0	1	3	0	1	134	2965	40.851903	-73.936900

None of top 5 potential neighborhood is in this cluster.

- Cluster 4

	Cluster Labels	Neighborhood	Competitor	Park	Fitness	Shopping	Cafe	Retail Rental	Apartment Rental	Latitude	Longitude
2	3	Flatiron	2	3	13	3	7	405	4276	40.768113	-73.958860
15	3	Upper East Side	1	1	10	2	8	377	4038	40.752042	-73.967708
21	3	Sutton Place	0	4	8	2	7	241	3941	40.723259	-73.988434
32	3	Financial District	0	2	7	0	11	369	4142	40.775930	-73.947118
35	3	Civic Center	0	3	10	1	6	566	4489	40.816934	-73.957385

Neighborhood Sutton Place and Civic Center are in the same cluster with Flatiron which already has 2 pet stores. That means they have similar features for running successful pet store businesses.

Sutton Place has 4 parks. Its retail rental price 241 USD and apartment rental price 3941 USD are below the median prices.

Civic Center has 3 parks. Its retail rental price 566 USD and apartment rental price 4489 USD are higher than the medium prices.

- Cluster 5

	Cluster Labels	Neighborhood	Competitor	Park	Fitness	Shopping	Cafe	Retail Rental	Apartment Rental	Latitude	Longitude
0	4	Washington Heights	2	2	2	7	11	134	2284	40.731	-73.974052

Neighborhood Washington Heights is different from all other neighborhoods.

- Cluster 6

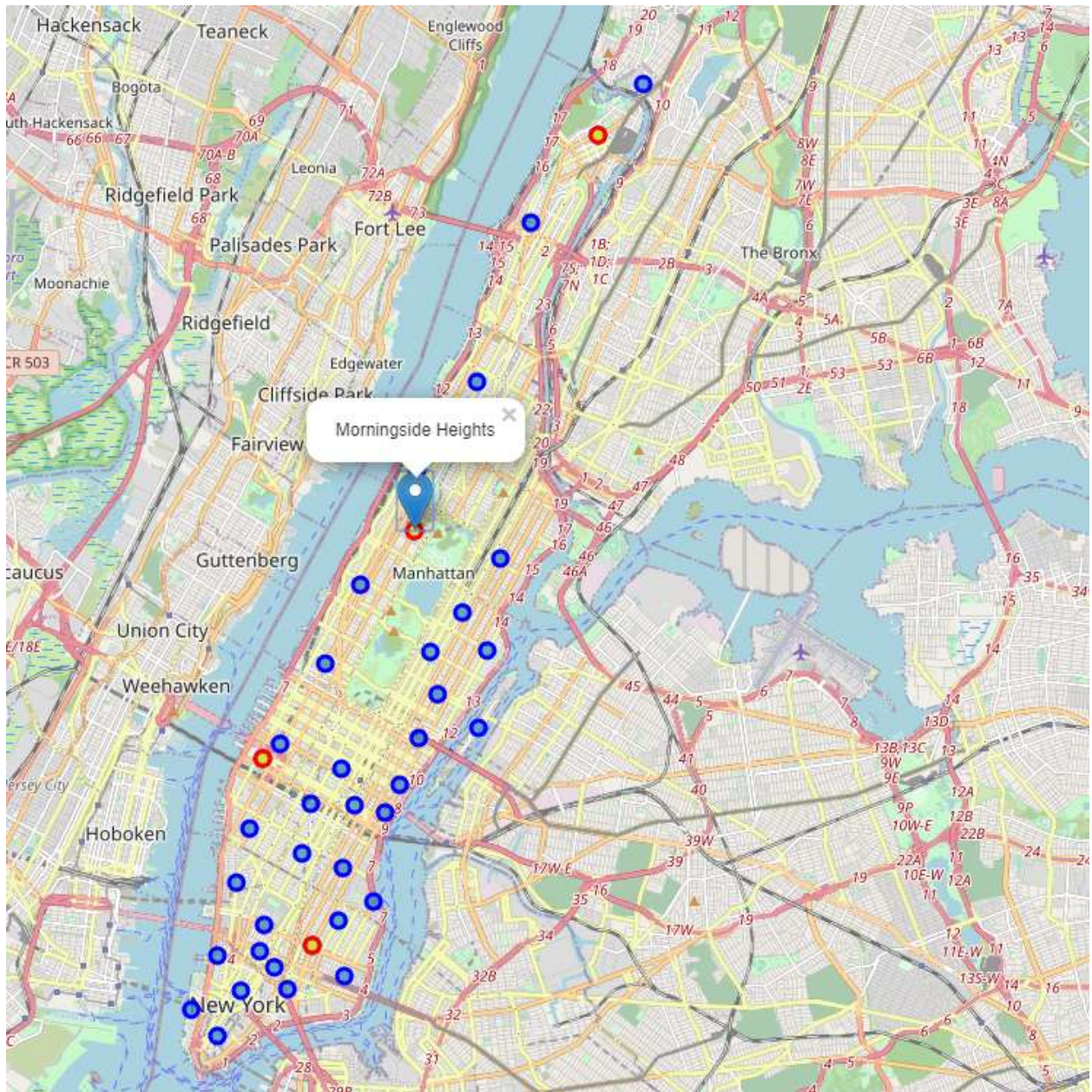
	Cluster Labels	Neighborhood	Competitor	Park	Fitness	Shopping	Cafe	Retail Rental	Apartment Rental	Latitude	Longitude
3	5	Battery Park City	1	6	0	2	2	369	5605	40.876551	-73.910660
13	5	Tudor City	1	5	4	0	1	597	3808	40.748510	-73.988713
14	5	Turtle Bay	1	4	3	2	9	971	3906	40.760280	-73.963556
17	5	West Village	1	5	0	1	11	306	4598	40.739673	-73.990947
22	5	Tribeca	0	5	5	0	4	369	5607	40.715229	-74.005415
23	5	Morningside Heights	0	4	0	3	5	134	4388	40.797307	-73.964286

Neighborhood Morningside Heights is in the same cluster with four other neighborhoods which already have 1 pet stores. That means they have similar features for running successful pet store businesses.

Morningside Heights has 4 parks. Its retail rental price is 134 USD, much lower than median price and its apartment rental price is 4388 USD which is a little higher than median price.

With 4 parks, low retail rental price and potential customers with higher than medium income, Morningside Heights sounds like the best choice for opening a new pet store.

Let's mark our choice of neighborhood on Manhattan map



5 Discussion

Mathematical tools such as machine learning are always useful for data analysis. But we cannot only rely on mathematical tools to make decision for our clients. In previous section, I got 5 neighborhood candidates with similar scores on potential pet stores using machine learning method. To pick the best choice for my client

Tina, I need more analysis on the neighborhood details related on her pet service store business.

K-means clustering is also a useful tool to group data for comparing similarity between them. To cluster the data into meaningful groups, we need to focus on those features that are important for clients, such as rental price, parks and shopping centers.

To make business decision for my client Tina, I also focused more on the features she really cares, such as retail rental price and parks. Since Tina wants to have the potential to expand her businesses to pet walking and boarding, number of parks in the neighborhood is important. Because nearby parks make pet walking more convenient and enjoyable, more parks not only bring more pet owners to visit the pet service store, but also make them more likely to choose Tina's pet walking and boarding services. If Tina would like to focus on providing more luxury grooming services at customer's home, we will focus more on higher apartment rental price because we will have more potential customers who would like to pay more for luxury pet services.

6 Conclusion

In this report, I introduced my client Tina's problem on choosing the best neighborhood in Manhattan to open her pet service store. To help her make decision, I gathered Manhattan neighborhood venue data from Foursquare, retail rental price and apartment rental price online. I analyzed these data using machine learning and k-means clustering method to pick the best neighborhood for Tina to open her new business. The final choice is visualized on Manhattan map. Once the neighborhood is selected, Tina will reduce her business risk and save a lot of effort on retail rental searching.

7 References

[1] THE REAL ESTATE BOARD OF NEW YORK
https://www.rebny.com/content/dam/rebny/Documents/PDF/News/Research/Retail%20Reports/2019_Spring_ManhattanRetailReport.pdf

[2] Average Rent in Manhattan, NY

<https://www.rentcafe.com/average-rent-market-trends/us/ny/manhattan/>