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

CINDY DONG

APRIL 23, 2020

Introduction

- Manhattan has 40 neighborhoods
- Client Tina's problem:

  Find the best neighborhood to open her new pet service store

Client’s Requirement

- Fewer pet store competitor
- More parks
- Lower retail rental price
- Higher residential rental price

Data Gathering

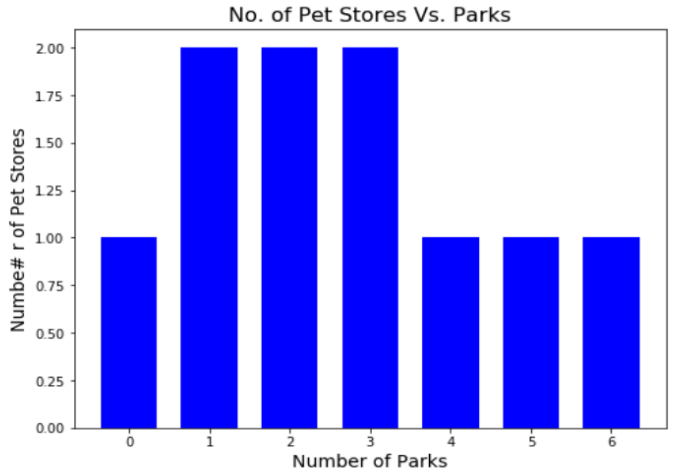
- Manhattan venue data from Foursquare
- Manhattan average retail rental price from rebny.com
- Manhattan average apartment rental price from rentcafe.com

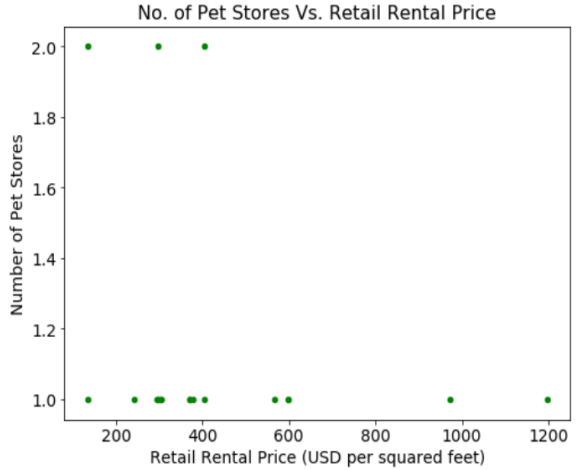
Data Cleaning and Sorting

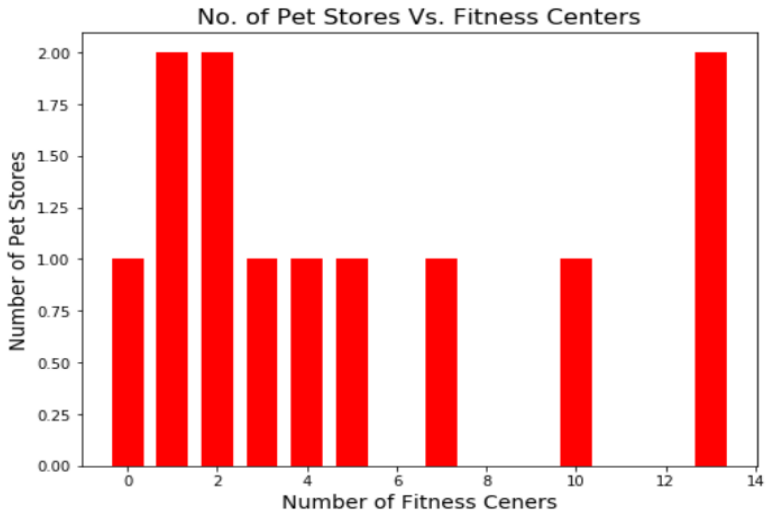
- Group pet store related venues in the same categories and add geographic coordinates
- Sort data by pet store “Competitor”(number of existing pet stores)
- First 18 neighborhoods have pet store >=1

	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

Relationship between Competitor and Other Features







In the first 18 neighborhoods which have pet stores, no simple linear relationship between number of pet stores and other related features.

Data Splitting

Training data
First 18 rows
Pet store>=1

Test data
Last 22 rows
Pet store=0

	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
:		Y_train			:		X_train			:
16	Upper West Side	1	0	0	1	10	301	4668	40.746917	-73.971219
17	West Village	1	5	0	1	11	306	4598	40.739673	-73.990947
18	Soho	0	0	3	2	9	369	5301	40.707107	-74.010665
19	Roosevelt Island	0	2	1	2	2	377	3416	40.711932	-74.016869
:					:		X_test			:
38	Central Harlem	0	1	3	0	1	134	2965	40.851903	-73.936900
39	Yorkville	0	2	3	1	15	377	4458	40.756658	-74.000111

Y_predict
=?

Data Normalization

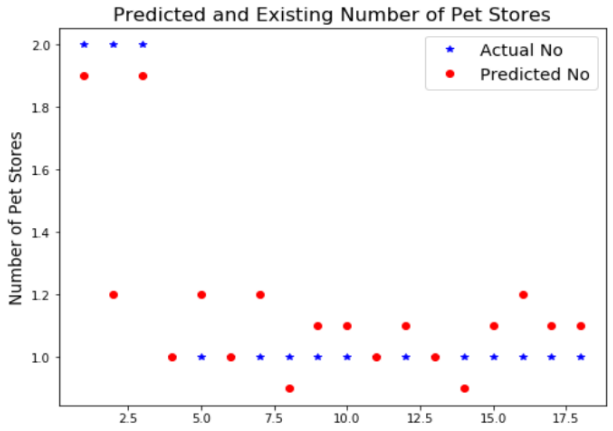
Normalize x_train and x_test by column max so each feature has similar weight in the model

	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

Machine learning

	R^2	Mean Square Error
Multiple Linear Regression	0.462	0.074
Ridge Regression	0.680	0.044

Ridge Regression model fits training data well



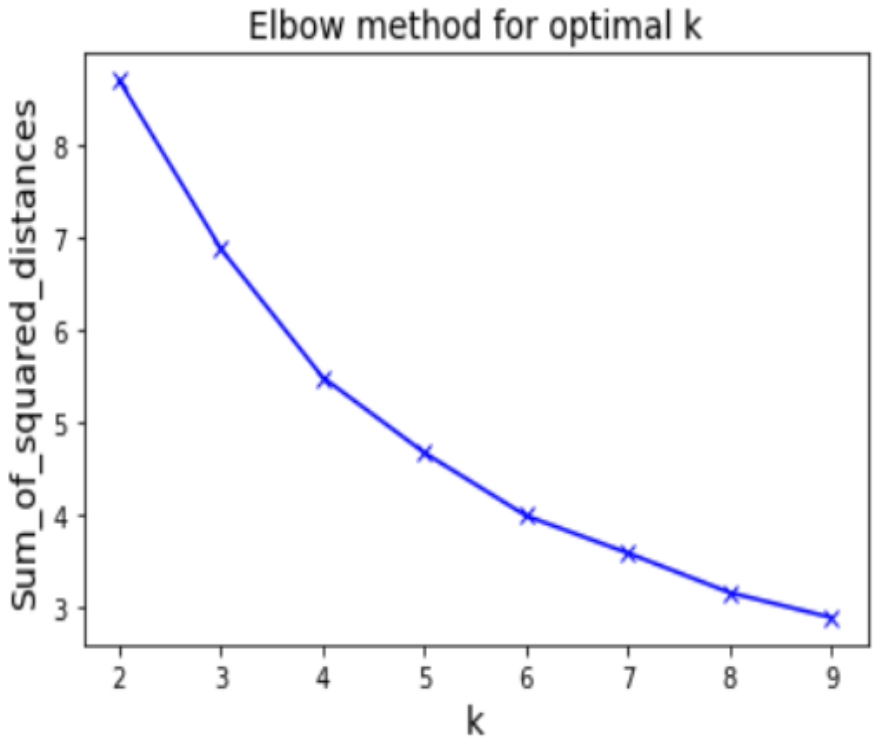
Predict Potential Pet Store Number

- Predict number of pet stores using ridge regression on test data
- Add predicted number back to data as “Potential Pet Store”
- Sort data by “Potential Pet Store” in descending order
- Choose top 5 neighborhoods as potential candidates:

  Sutton Place, Civic center, Yorkville, Chelsea and Morningside Heights

	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

K-means Clustering



6 is the optimal number of clustering

K-means Clustering Results

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

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

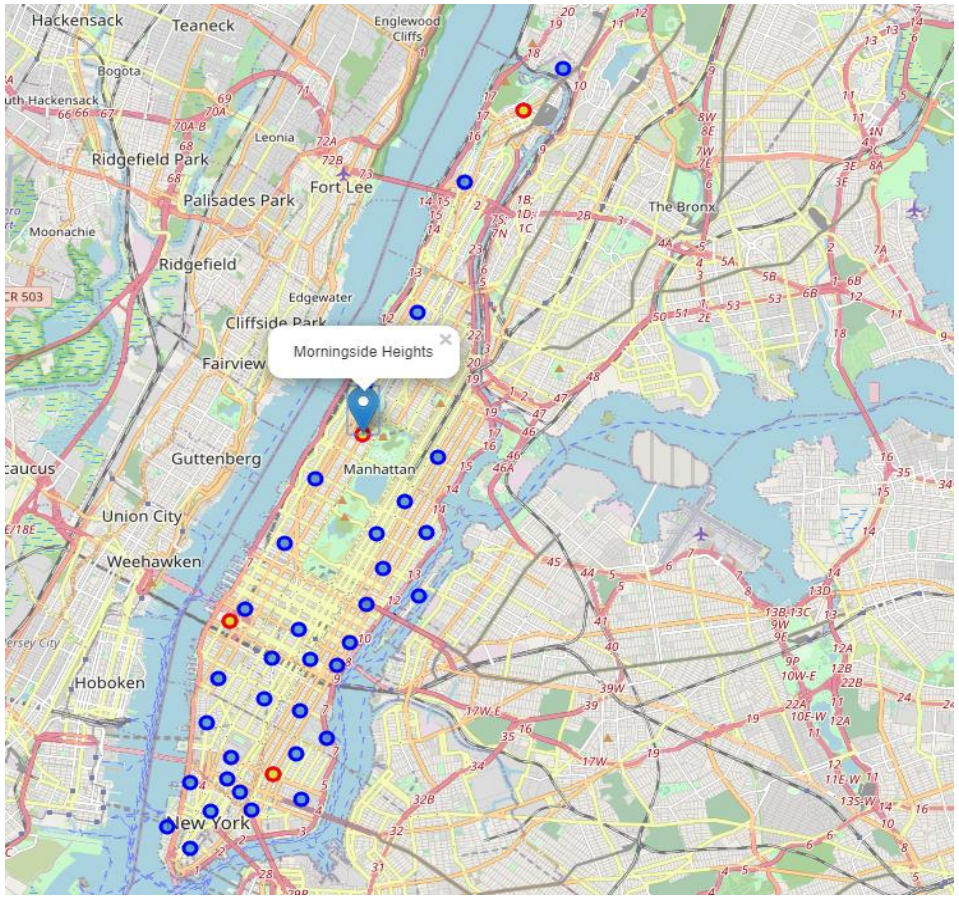
Cluster 6

	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	-73.96428

Morningside Heights is the final choice:

- 4 parks
- Low retail rental price
- Apartment rental price higher than median

Results Visualization



Morningside Heights is the best neighborhood for opening new pet service store.

Discussion

- Mathematic tools are useful but don’t only reply on them
- Pick the right features before running K-means clustering
- Focus on features which clients really care for decision making

Conclusion

- Introduce Tina's problem: pick best neighborhood for her new pet service store
- Gathering data related to problem
- Analyze data using machine learning and K-means clustering
- Make decision and visualize the result