

Selecting the Venue to set up Japanese Ramen Restaurant in Queens, New York

1. Introduction

Ramen is a noodle soup dish that was originally imported from China and has become one of the most popular dishes in Japan in recent decades. Ramen are inexpensive and widely available, two factors that also make them an ideal option for budget travelers. Ramen restaurants, or ramen-ya, can be found in virtually every corner of the country and produce countless regional variations of this common noodle dish [1].

Queens is the easternmost of the 5 boroughs in New York city. It is the largest borough geographically and is adjacent to the borough of Brooklyn at the southwestern end of Long Island. The borough of Queens is the second largest in population (after Brooklyn), with an estimated 2,358,582 residents in 2017, approximately 48% of them foreign-born [2].

With the large and diverse population at Queens, it provides good business opportunity for any established Japanese Ramen restaurants or chains to set up their business there. However, selecting the appropriate venue to set up the restaurant could be a big problem if no proper analysis is carried out to ensure maximum business potential.

2. Data Description

2.1. Data Source

The data utilized for the analysis can be separated into 2 categories, which are the geographical data of Queens and detail neighborhood data. Geographical data including all the neighborhoods and its respective coordinates data are extracted from NYU Spatial Repository. The detail neighborhood data including places of interested, categories and rating are extracted from Foursquare API.

2.2. Data Wrangling and Cleaning

Data downloaded from NYU Spatial Repository in the format of .json contains many information, and the geographical information which are borough, neighborhood, latitudes and longitudes coordinates for New York City. The data are extracted from the .json file into a *pandas* DataFrame followed by further filtering to obtain only the specific Data for Queens.

After getting the Queens specific data, the geographical data from each neighborhood in Queens are used as the input to obtain the common venues at each neighborhood through Foursquare API.

3. Exploratory Analysis

The raw data from .json file was first extracted into a *pandas* dataframe containing all the boroughs, neighborhoods and their geographical coordinates data. It was then further filtered to obtain the data for Queens only, as shown in Table 1.

Table 1

	Neighborhood	Borough	Latitude	Longitude
0	Astoria	Queens	40.768509	-73.915654
1	Woodside	Queens	40.746349	-73.901842
2	Jackson Heights	Queens	40.751981	-73.882821
3	Elmhurst	Queens	40.744049	-73.881656
4	Howard Beach	Queens	40.654225	-73.838138

Python **folium** library was then used to visualize geographic details of Queens area, with every neighbourhood indicated as one blue dot on the map, as shown in Figure 1 below.

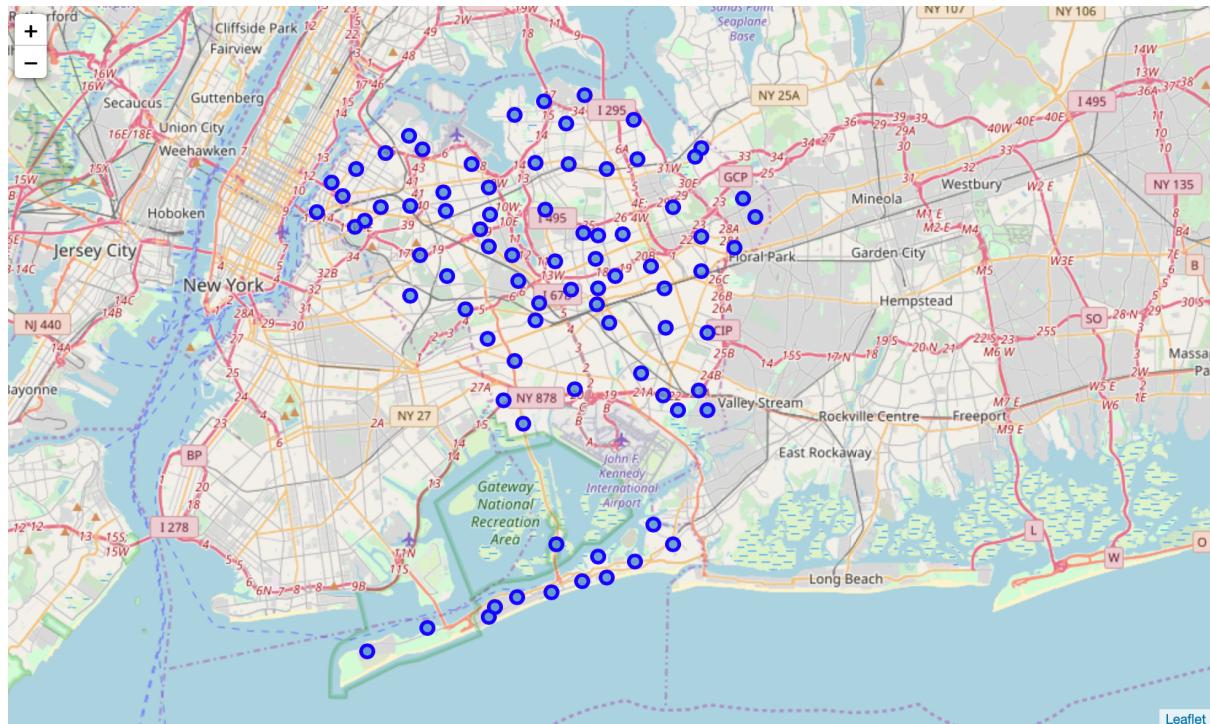


Figure 1: Neighborhoods in Queens.

Using Foursquare API, the top 100 most common venues within 3km radius of each neighborhood were extracted, joining the data with Table 1 and shown in Table 2 below.

Table 2

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Distance	Venue Category ID	Venue Category
0	Astoria	40.768509	-73.915654	Titan Foods Inc.	4a9c0105f964a520b03520e3	40.769198	-73.919253	312	4bf58dd8d48988d1f5941735	Gourmet Shop
1	Astoria	40.768509	-73.915654	Ample Hills Creamery	5ae115d7c84d18002393461a	40.765562	-73.919012	433	4bf58dd8d48988d1c9941735	Ice Cream Shop
2	Astoria	40.768509	-73.915654	CrossFit Queens	4c94d26d58d4b60c40fc2b29	40.769404	-73.918977	297	4bf58dd8d48988d176941735	Gym
3	Astoria	40.768509	-73.915654	Sitan Muay Thai	5208fb011d228eef0972843	40.766108	-73.913224	336	4bf58dd8d48988d101941735	Martial Arts Dojo
4	Astoria	40.768509	-73.915654	Al-sham Sweets and Pastries	4ba58abaf964a520d01039e3	40.768077	-73.911561	348	4bf58dd8d48988d115941735	Middle Eastern Restaurant

Some simple analysis was then done and discussed in the following graphs/tables. The most common venue categories in Queens were sorted and shown in bar plot in Figure 2.

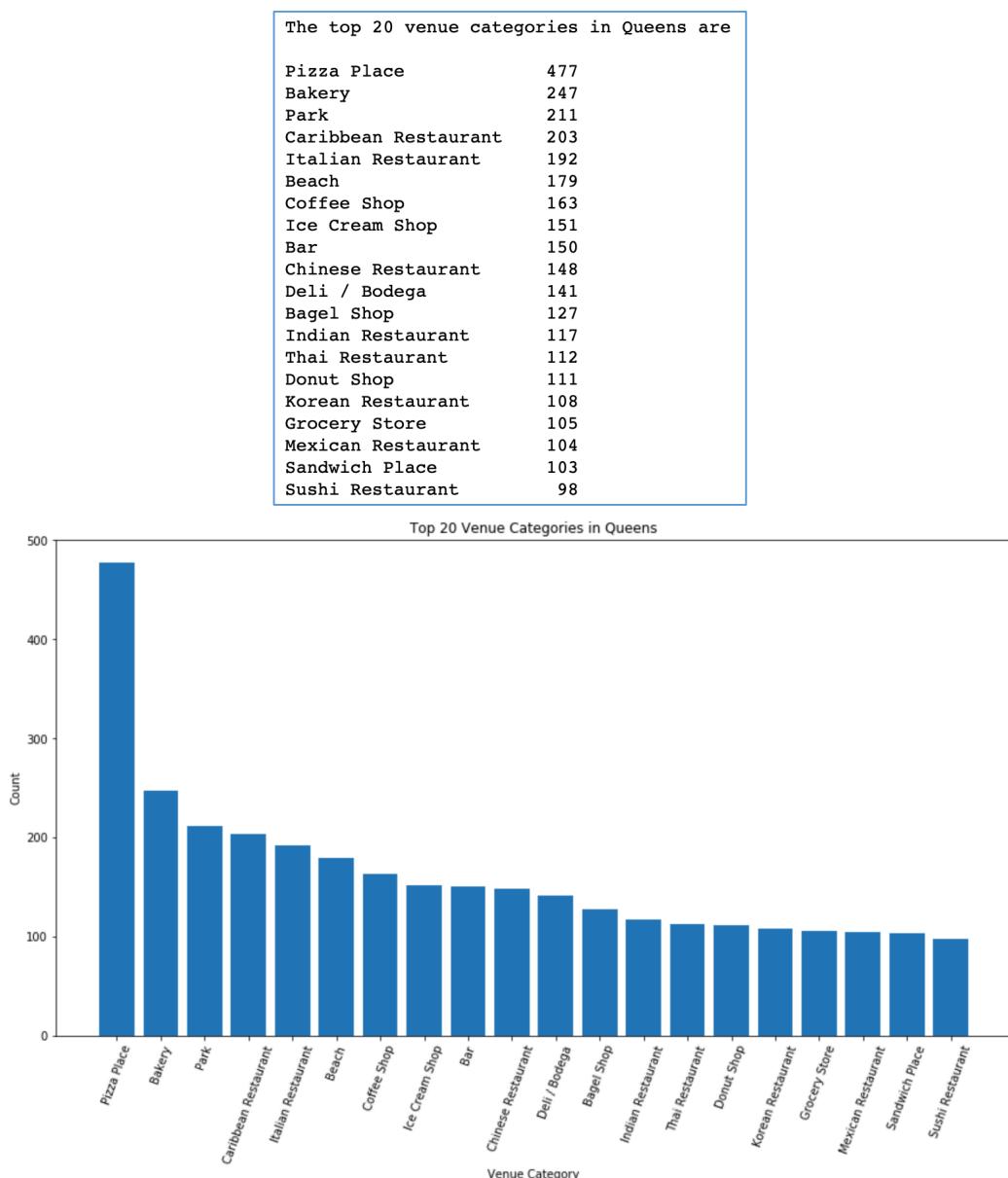


Figure 2: Top 20 venue categories in Queens.

It can be seen that the pizza place is the most common venue in Queens followed by bakery, park, Caribbean restaurant and Italian restaurant. Out of the top 20 venue categories, restaurant, bakery, café and bar take up 17 spots. Further drill down the data, the top 20 restaurant categories in Queens were sorted and shown in Figure 3 below.

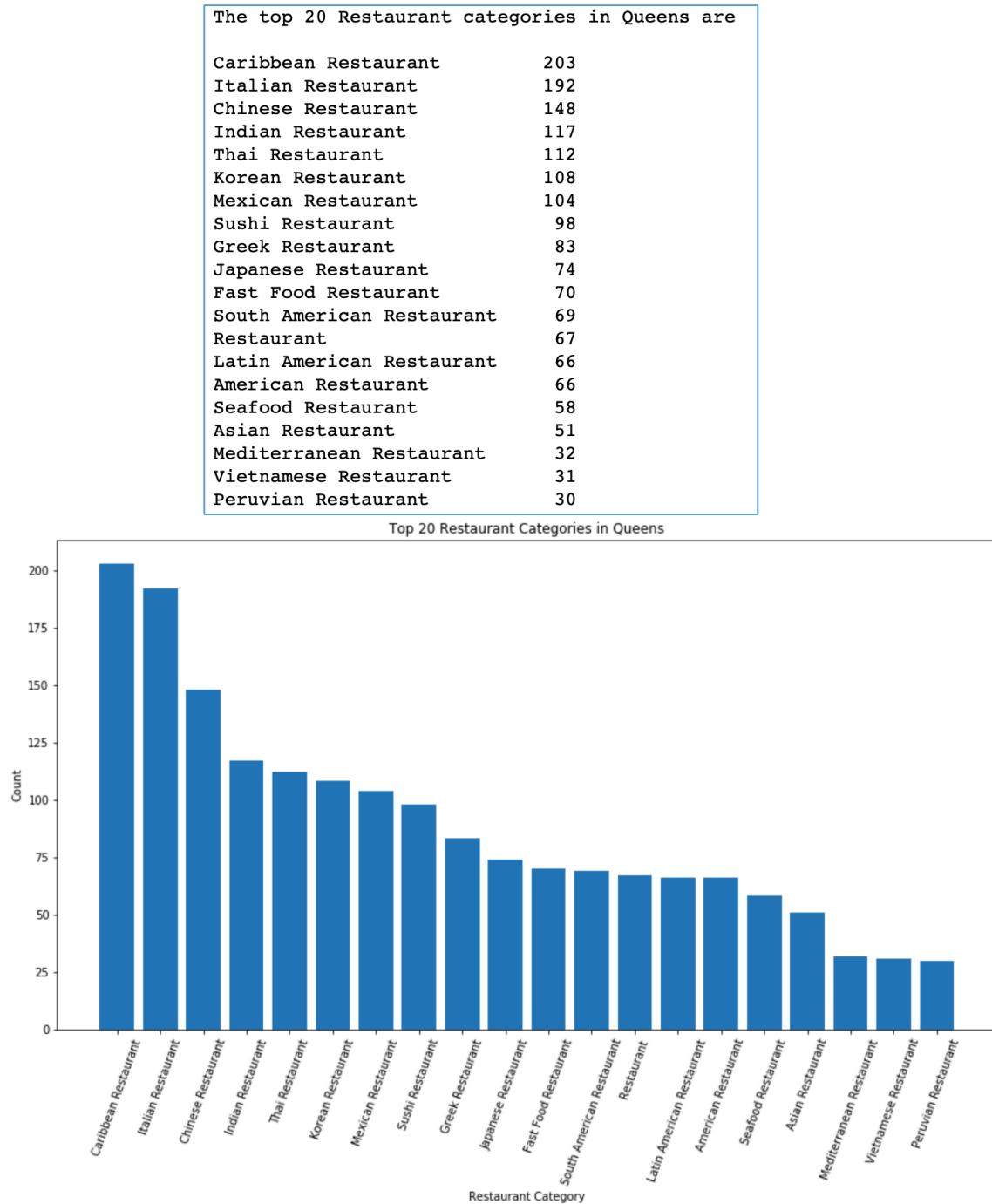


Figure 3: Top 20 restaurant categories in Queens.

From the graph, it can be seen that Caribbean restaurant is the most common type of restaurant in Queens, followed by Italian restaurant, Chinese restaurant, Indian restaurant and Thai restaurant. Sushi restaurant and Japanese restaurant are ranked 8th and 10th, and the combined sushi and Japanese restaurants quantity is 172, ranking it as 3rd most common.

Ramen restaurant is not ranked as top 20 venue categories, as the number of Ramen restaurants in Queens is relatively low at 11 outlets in 11 different neighborhoods, as shown in Figure 4 below.

There are 11 Ramen restaurants at Queens

The Ramen restaurants are at the neighborhood below
 ['Kew Gardens', 'Forest Hills Gardens', 'Sunnyside', 'Kew Gardens Hills', 'Long Island City', 'Sunnyside Gardens', 'Steinway', 'Woodside', 'Blissville', 'Forest Hills', 'Astoria Heights']

Figure 4: Ramen restaurants in Queens.

The top 10 neighborhoods with the most restaurants are sorted and shown in Figure 5 below.

The top 10 neighborhoods with the most restaurants are

Elmhurst	65
Jackson Heights	64
Woodside	50
Maspeth	50
Murray Hill	47
Flushing	47
South Jamaica	44
Jamaica Center	43
Lefrak City	42
Pomonok	39

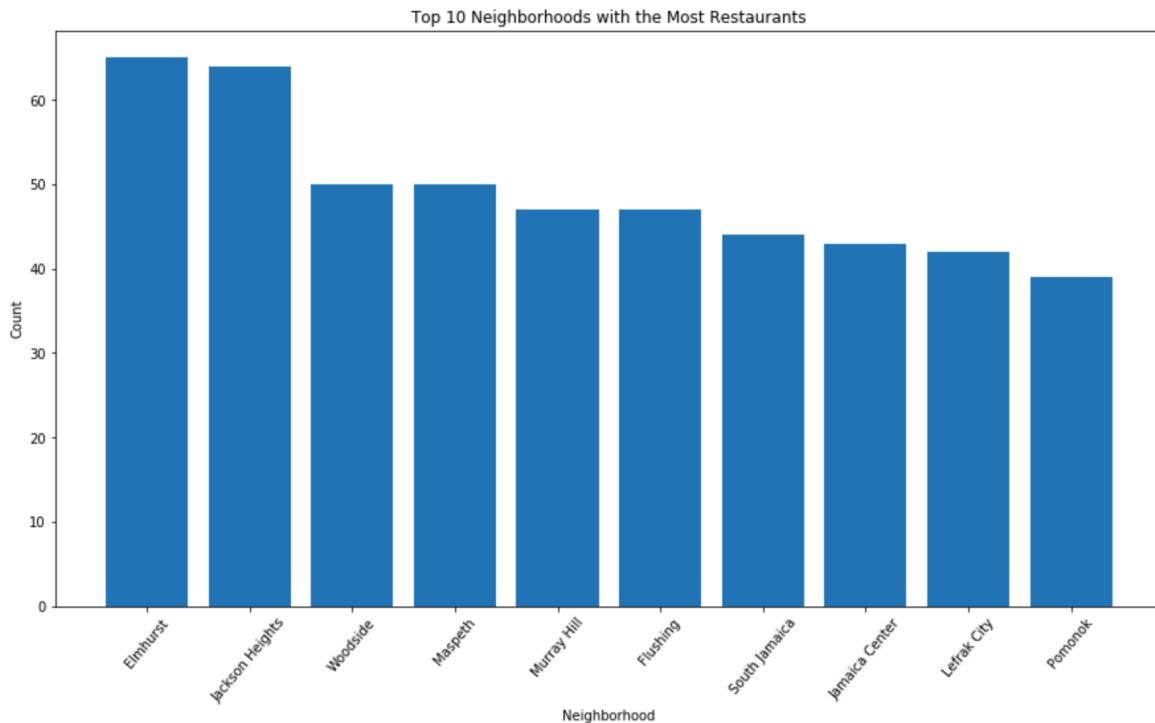


Figure 5: Top 10 Neighborhoods with the most restaurant quantity.

The top 5 venues in these top 10 neighborhoods are sorted and shown in Figure 6 below.

----Elmhurst----		
	venue	freq
0	Thai Restaurant	0.14
1	Bakery	0.07
2	Indian Restaurant	0.06
3	South American Restaurant	0.05
4	Argentinian Restaurant	0.04

----Woodside----		
	venue	freq
0	Thai Restaurant	0.09
1	Bar	0.05
2	Indian Restaurant	0.04
3	Bakery	0.04
4	Pizza Place	0.04

----Murray Hill----		
	venue	freq
0	Korean Restaurant	0.20
1	Chinese Restaurant	0.08
2	Bubble Tea Shop	0.07
3	Dumpling Restaurant	0.04
4	Coffee Shop	0.04

----Jackson Heights----		
	venue	freq
0	Thai Restaurant	0.13
1	Bakery	0.07
2	Indian Restaurant	0.05
3	South American Restaurant	0.05
4	Mexican Restaurant	0.05

----Maspeth----		
	venue	freq
0	Thai Restaurant	0.09
1	Pizza Place	0.08
2	Mexican Restaurant	0.07
3	Bar	0.07
4	Bakery	0.05

----Flushing----		
	venue	freq
0	Chinese Restaurant	0.15
1	Korean Restaurant	0.09
2	Bubble Tea Shop	0.08
3	Hotpot Restaurant	0.04
4	Dumpling Restaurant	0.04

----South Jamaica----		
	venue	freq
0	Caribbean Restaurant	0.11
1	Park	0.05
2	Pizza Place	0.05
3	Latin American Restaurant	0.04
4	BBQ Joint	0.04

----Jamaica Center----		
	venue	freq
0	Caribbean Restaurant	0.09
1	Indian Restaurant	0.05
2	Pizza Place	0.05
3	Coffee Shop	0.04
4	Fast Food Restaurant	0.03

----Lefrak City----		
	venue	freq
0	Tennis Stadium	0.10
1	Bakery	0.05
2	Pizza Place	0.04
3	Thai Restaurant	0.04
4	Chinese Restaurant	0.03

----Pomonok----		
	venue	freq
0	Chinese Restaurant	0.08
1	Pizza Place	0.08
2	Bakery	0.04
3	Grocery Store	0.03
4	Coffee Shop	0.03

Figure 6: Top 5 venues in the top 10 neighborhoods with most restaurants.

It can be seen that Elmhurst, Jackson Height and Woodside are having Thai restaurants, Indian restaurants and bakery as the top venues, while Murray Hill and Flushing are topping the Korean restaurants, Chinese restaurants and Bubble Tea shops

4. Machine Learning Modeling

Based on the data in Table 2, the 10 most common venues in each neighborhood, distance from the center of each neighborhood, and the restaurant percentage from the total venues within a neighborhood are summarized in a new dataframe to be used for the machine learning modeling.

kMeans clustering method is applied to check for any additional insight from the data available. In order to determine the suitable k to be used for the KMeans clustering, elbow method is carried out. The elbow method data can be seen in Figure 7 below. Y axis is the distortion score, while x axis is the number of cluster (k). k value of 4 is recommended based on the Figure 7, which a change in the drop in distortion score reduces significantly starting from k =5.

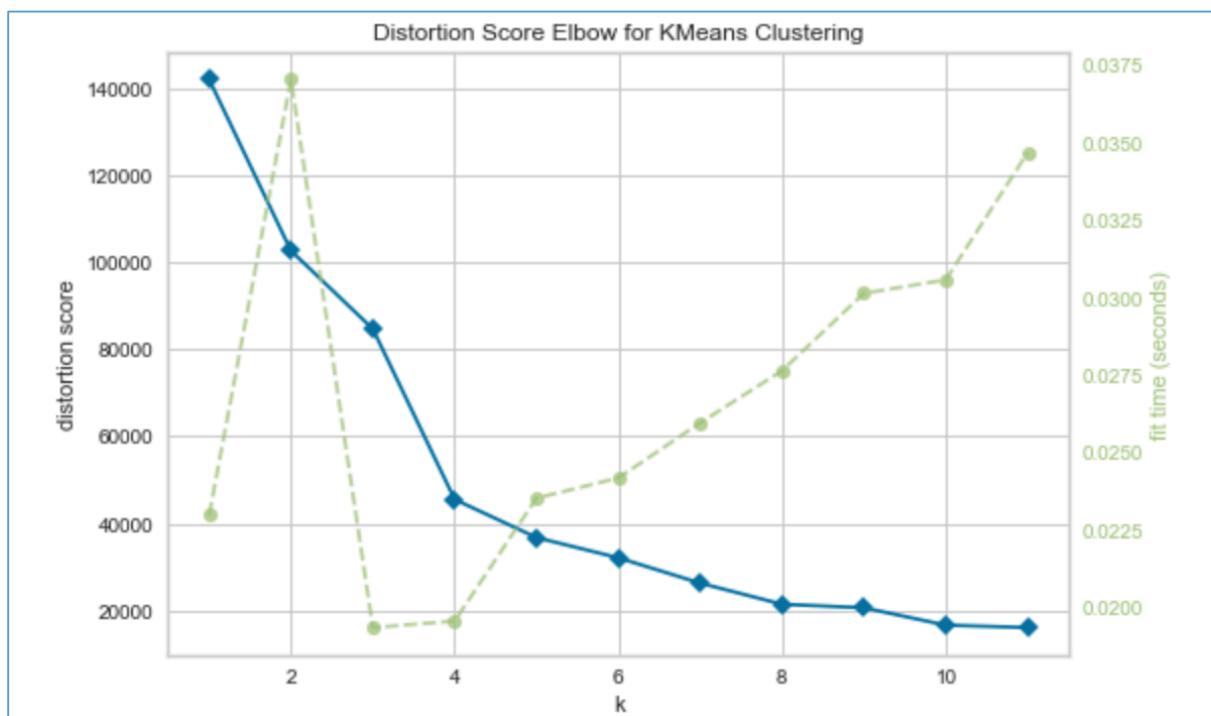


Figure 7: Elbow method plot

After getting the k value, KMeans clustering analysis is carried out and the neighborhoods are clustered into 4 different clusters. The neighborhood clusters are illustrated in the map below.

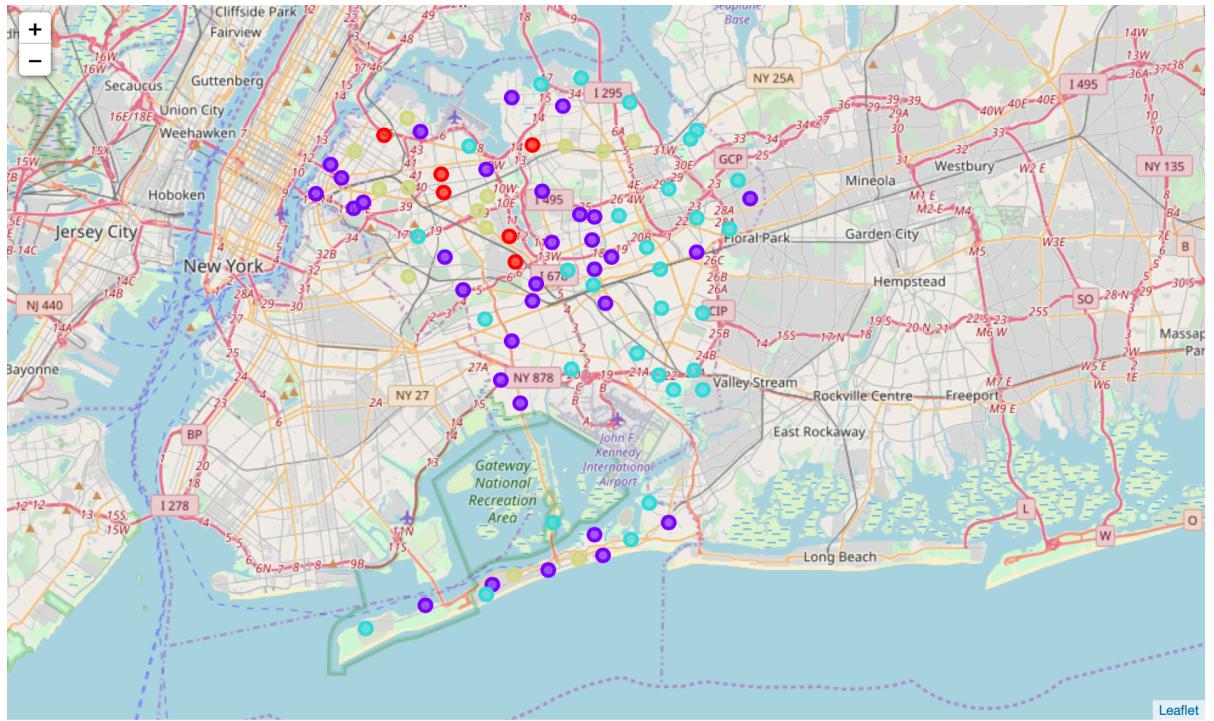


Figure 8: Neighborhood in Queens in accordance to clusters.

The details of each cluster are analyzed and summarized in Figure 9 to Figure 12 respectively.

Thai Restaurant is ranked the most in 1st most common venue in the cluster 1 neighborhood with total 2 counts

Bakery is ranked the most in 2nd most common venue in the cluster 1 neighborhood with total 3 counts

The average distance from the neighborhood center is 812.0950000000001 m
Average restaurant probability is 0.451666666666667

	Neighborhood	Cluster Labels	Distance	Restaurant	Ramen	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Astoria	0	677.22	0.33	0.00	Greek Restaurant	Bagel Shop	Bakery	Bar	Italian Restaurant
22	Elmhurst	0	777.37	0.65	0.00	Thai Restaurant	Bakery	Indian Restaurant	South American Restaurant	Mexican Restaurant
25	Flushing	0	789.80	0.46	0.00	Chinese Restaurant	Korean Restaurant	Bubble Tea Shop	Tennis Stadium	Dumpling Restaurant
26	Forest Hills	0	975.46	0.35	0.01	Bakery	Pizza Place	Italian Restaurant	Sushi Restaurant	Bagel Shop
27	Forest Hills Gardens	0	821.60	0.28	0.01	Pizza Place	Bakery	Park	Italian Restaurant	Bagel Shop
37	Jackson Heights	0	831.12	0.64	0.00	Thai Restaurant	Bakery	Indian Restaurant	Mexican Restaurant	South American Restaurant

Figure 9: Summary of Cluster 1.

Pizza Place is ranked the most in 1st most common venue in the cluster 2 neighborhood with total 14 counts

Italian Restaurant is ranked the most in 2nd most common venue in the cluster 2 neighborhood with total 4 counts

The average distance from the neighborhood center is 1710.0442900181743 m
Average restaurant probability is 0.27363500416308967

Neighborhood	Cluster Labels	Distance	Restaurant	Ramen	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Arverne	1 1739.722222	0.129630	0.00	Beach	Surf Spot	Beach Bar	Wine Bar	Pizza Place
2	Astoria Heights	1 1640.640000	0.310000	0.01	Bar	Bakery	Greek Restaurant	Pizza Place	Bagel Shop
9	Belle Harbor	1 1657.266667	0.160000	0.00	Beach	Ice Cream Shop	Pizza Place	Pharmacy	Deli / Bodega
11	Blissville	1 1828.450000	0.300000	0.01	Coffee Shop	Mexican Restaurant	Café	Yoga Studio	Bar
17	College Point	1 1793.170000	0.270000	0.00	Chinese Restaurant	Korean Restaurant	Deli / Bodega	Gym	Park
23	Far Rockaway	1 1749.535714	0.119048	0.00	Beach	Pizza Place	Supermarket	Chinese Restaurant	Golf Course
24	Floral Park	1 1818.810000	0.210000	0.00	Pizza Place	Indian Restaurant	Italian Restaurant	Park	Deli / Bodega
30	Glendale	1 1750.840000	0.230000	0.00	Pizza Place	Italian Restaurant	Bakery	Deli / Bodega	German Restaurant
32	Hillcrest	1 1633.500000	0.360000	0.00	Pizza Place	Coffee Shop	Park	Mexican Restaurant	Chinese Restaurant
35	Howard Beach	1 1603.920000	0.190000	0.00	Pizza Place	Italian Restaurant	Discount Store	Deli / Bodega	Diner
36	Hunters Point	1 1871.040000	0.200000	0.00	Park	Gym	Cocktail Bar	Japanese Restaurant	American Restaurant
39	Jamaica Estates	1 1576.330000	0.380000	0.00	Pizza Place	Indian Restaurant	Mexican Restaurant	Sandwich Place	Chinese Restaurant
40	Jamaica Hills	1 1825.890000	0.390000	0.00	Pizza Place	Caribbean Restaurant	Indian Restaurant	Coffee Shop	Bakery

Figure 10: Summary of Cluster 2.

Pizza Place is ranked the most in 1st most common venue in the cluster 3 neighborhood with total 14 counts

Pizza Place is ranked the most in 2nd most common venue in the cluster 3 neighborhood with total 7 counts

The average distance from the neighborhood center is 2048.7786062330188 m
Average restaurant probability is 0.25776414174359386

Neighborhood	Cluster Labels	Distance	Restaurant	Ramen	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Bay Terrace	2 1901.840000	0.290000	0.0	Pizza Place	Italian Restaurant	Park	Bakery	Greek Restaurant
6	Bayswater	2 2183.203704	0.055556	0.0	Pizza Place	Beach	Donut Shop	Sandwich Place	Pharmacy
7	Beechhurst	2 1951.350000	0.180000	0.0	Pizza Place	Deli / Bodega	Italian Restaurant	Park	Donut Shop
8	Bellaire	2 2090.650000	0.230000	0.0	Pizza Place	Park	Indian Restaurant	Bank	Discount Store
10	Bellerose	2 2095.050000	0.200000	0.0	Pizza Place	Park	Indian Restaurant	Bakery	Italian Restaurant
12	Breezy Point	2 2065.875000	0.041667	0.0	Beach	Bar	Bus Stop	Café	Scenic Lookout
13	Briarwood	2 1908.110000	0.340000	0.0	Pizza Place	Caribbean Restaurant	Coffee Shop	Bar	Vegetarian / Vegan Restaurant
14	Broad Channel	2 2167.111111	0.111111	0.0	Beach	Surf Spot	Bagel Shop	Beach Bar	Wine Bar
15	Brookville	2 2183.620000	0.200000	0.0	Caribbean Restaurant	Clothing Store	Donut Shop	Park	Pizza Place
16	Cambria Heights	2 2131.880000	0.280000	0.0	Caribbean Restaurant	Pizza Place	Pharmacy	Donut Shop	Discount Store
19	Douglaston	2 1994.280000	0.340000	0.0	Pizza Place	Bakery	Grocery Store	Burger Joint	Italian Restaurant
20	East Elmhurst	2 1968.060000	0.390000	0.0	Tennis Stadium	South American Restaurant	Bakery	Pizza Place	Latin American Restaurant
21	Edgemere	2 1901.356164	0.136986	0.0	Beach	Pizza Place	Donut Shop	Supermarket	Surf Spot
28	Fresh Meadows	2 1891.610000	0.360000	0.0	Pizza Place	Park	Korean Restaurant	Bakery	Greek Restaurant
29	Glen Oaks	2 2095.530000	0.250000	0.0	Bakery	Pizza Place	Italian Restaurant	Indian Restaurant	Ice Cream Shop

Figure 11: Summary of cluster 3.

Bakery is ranked the most in 1st most common venue in the cluster 4 neighborhood with total 2 counts
 Bagel Shop is ranked the most in 2nd most common venue in the cluster 4 neighborhood with total 2 counts
 The average distance from the neighborhood center is 1296.2743672456575 m
 Average restaurant probability is 0.32560794044665015

Neighborhood	Cluster Labels	Distance	Restaurant	Ramen	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	
3	Auburndale	3	1281.590000	0.340000	0.00	Korean Restaurant	Bagel Shop	Greek Restaurant	Pizza Place	Coffee Shop
5	Bayside	3	1086.790000	0.290000	0.00	Bakery	Pizza Place	Cosmetics Shop	Korean Restaurant	Greek Restaurant
18	Corona	3	1391.490000	0.350000	0.00	Tennis Stadium	South American Restaurant	Pizza Place	Bakery	Park
31	Hammels	3	1376.096774	0.112903	0.00	Beach	Surf Spot	Bagel Shop	Pizza Place	Beach Bar
44	Lefrak City	3	1359.640000	0.430000	0.00	Tennis Stadium	Thai Restaurant	Bakery	Pizza Place	Park
51	Murray Hill	3	1209.640000	0.460000	0.00	Korean Restaurant	Chinese Restaurant	Bubble Tea Shop	Dumpling Restaurant	Coffee Shop
60	Ravenswood	3	1322.380000	0.260000	0.00	Bar	Bagel Shop	Park	Italian Restaurant	Coffee Shop
61	Rego Park	3	1135.650000	0.350000	0.00	Bakery	Italian Restaurant	Cosmetics Shop	American Restaurant	Pizza Place
63	Ridgewood	3	1412.590000	0.280000	0.00	Bar	Mexican Restaurant	Bakery	Coffee Shop	Nightclub
66	Rockaway Park	3	1447.440000	0.150000	0.00	Beach	Pizza Place	Ice Cream Shop	Bagel Shop	Bar
74	Steinway	3	1209.470000	0.330000	0.01	Greek Restaurant	Bar	Bakery	Pizza Place	Italian Restaurant
76	Sunnyside Gardens	3	1250.120000	0.370000	0.01	Pizza Place	Mexican Restaurant	Sushi Restaurant	Coffee Shop	Bar
80	Woodside	3	1368.670000	0.510000	0.01	Thai Restaurant	Bar	Indian Restaurant	South American Restaurant	Pizza Place

Figure 12: Summary of Cluster 4.

5. Results and Discussion

As shown in the exploratory analysis, Queens has 17 out of 20 top venue category to be either restaurant, bakery or café. Further analysis shows that Caribbean restaurants, Italian restaurants, Chinese restaurant, Indian restaurant and Thai restaurant are the top 5 restaurant category in Queens. Sushi restaurant and Japanese restaurant are ranked 8th and 10th respectively, and the combined sushi and Japanese restaurants quantity is 172, ranking it as 3rd most common. Ramen restaurant is not ranked as top 20 venue categories, as the number of Ramen restaurants in Queens is relatively low at 11 outlets in 11 different neighborhoods, as shown in Figure 4. With such a low quantity of ramen restaurant in Queens, it is a good choice to set up a new Ramen restaurant there as the competition between Ramen restaurant will be relatively lower compared to setting up a new Caribbean, Italian, Chinese or Indian restaurants. Even though the combined sushi and Japanese restaurants are ranked 3rd, it will not affect the Ramen restaurant business significantly as the types of food sold are different.

After confirming the presence and competition of Ramen restaurant in Queens, the next step is to find the suitable location to set up one. Elmhurst has the most restaurant quantity, followed by Jackson Heights, Woodside, Maspeth and Murray Hill. There is no Ramen restaurant in any of these top 10 neighborhoods, thus eliminating the competition concern. Based on research, ramen restaurants are normally available in areas with a large demand for Asian cuisine, probably linked to the higher Ramen popularity by Asians, particularly Chinese, Korean, Hongkie and Taiwanese [3]. With this fact, the top 5 venues in these top 10 neighborhoods are analyzed, Elmhurst, Jackson Height and Woodside are having Thai restaurants, Indian restaurants and bakery as the top venues, while Murray Hill and Flushing are topping the Korean restaurants, Chinese restaurants and Bubble Tea shops. As our strategy is to find neighborhoods with high concentration of Asians, Elmhurst, Jackson Height, Woodside, Murray Hill and Flushing are shortlisted based on the type of restaurants listed. Checking on the demographic population in Queens, Flushing, Elmhurst and Woodside are topping the list with highest concentration of Asian [4]. Thus, based on the exploratory analysis, the top 5 recommended venues for setting up new Ramen restaurant in Queens are **Flushing, Murray Hill, Elmhurst, Woodside and Jackson Height**.

Machine learning modeling using KMeans method is carried out to check for additional insights from the data obtained. Before performing the KMeans analysis, the optimum k (number of cluster) is calculated based on Elbow Method and 4 clusters is determined as the optimum k value. Based on the 4 clusters data, we can see that distance from the neighborhood center affects the concentration of restaurants. There are more restaurants within 1km from the neighborhood center, and the restaurant decreases further from the neighborhood center, as shown in Figure 13. Thus, it is better to set up restaurant at neighborhoods in cluster 1 and 4 based on table 3. Besides, based on the cluster analysis from Figure 9 to Figure 12, Pizza place is the most common venues in neighborhoods in cluster 2 and cluster 3, further supporting the exclusion of these 2 clusters as the prefer locations to set up Ramen restaurant is at Asian restaurants centric area, which are lacking in cluster 2 and 3.

Table 3

Distance from Center	Restaurant Percentage	Cluster
0	812.095000	0.451667 1
1	1296.274367	0.325608 4
2	1710.044290	0.273635 2
3	2048.778606	0.257764 3

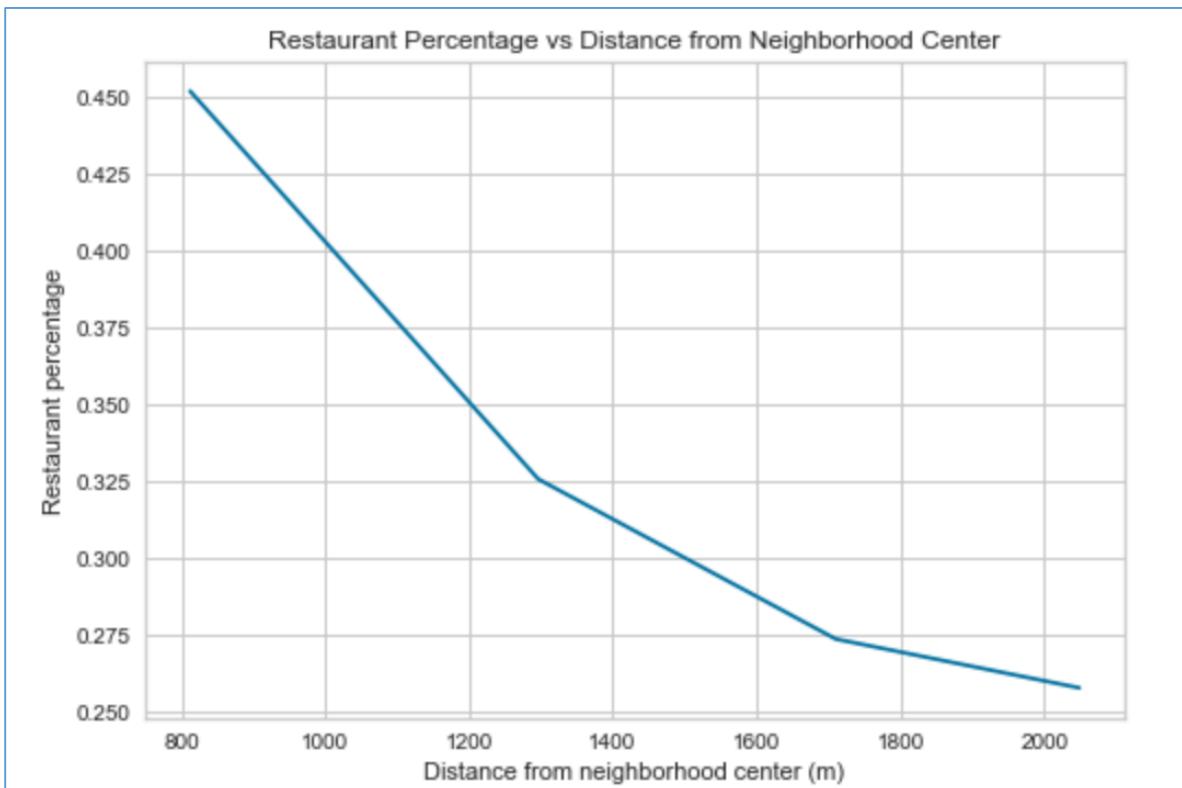


Figure 13: Restaurant concentration by the distance from the neighborhood center.

6. Conclusion

Based on the exploratory analysis, the top 5 recommended venues for setting up new Ramen restaurant in Queens are Flushing, Murray Hill, Elmhurst, Woodside and Jackson Height. Based on KMeans clustering modeling, neighborhoods in cluster 1 and cluster 4 are recommended, with the preference to be neighborhoods in cluster 1 due to its location closer to neighborhood center and higher restaurant concentration, providing higher foot traffic. Combining exploratory analysis and KMeans modeling results, the 3 proposed neighborhoods to set up a new Ramen restaurant are **Flushing, Elmhurst and Jackson Heights**.

7. References

- [1] <https://www.japan-guide.com/e/e2042.html>
- [2] <https://en.wikipedia.org/wiki/Queens>
- [3] <https://en.wikipedia.org/wiki/Ramen>
- [4] https://en.wikipedia.org/wiki/Demographics_of_Queens