

Opening an Asian restaurant in Queens, New York City

1.Introduction

New York City is the largest city in the United States with over 8 million residents. The city is the home to people from a diverse set of backgrounds and ethnicities. Asian Americans make up 11.8% of the population, more than a million¹. It provides many business opportunities that can serve the Asian population in many ways, including food industry.

New York City has five boroughs, including Bronx, Manhattan, Queens, Staten Islands, and Brooklyn. Among these five boroughs, Queens has the biggest Asian population with more than 550,000 residents, according to the 2010 US census².

1.1 problem statement

In this project, I identify best neighborhoods in Queens to open an Asian restaurant based on three criteria: The total number of restaurants, the number of Asian restaurants, and the Asian population within a neighborhood. The best location has the minimum number of restaurants and a reasonable Asian population as the potential customers.

1.2 Interest and target audience

The outcome of this project can be helpful for businessmen and businesswomen who want to open an Asian restaurant in Queens and are looking for the place for the restaurant. It can also provide insights to restaurants owners who want to invest in a chain store.

2.Data Selection

To address the problem statement in this project, data are gathered through three sources as follows.

A. New York City dataset: The NYC data contains all five Boroughs, their Neighborhoods, and the coordinates that define their location. The dataset provides the ability to explore all the Neighborhoods in Queens. This dataset is accessible through this [link](#).

Data sample:

¹ Data commons, Retrieved from <https://datacommons.org/>

² U.S. Census Bureau, Retrieved from https://www1.nyc.gov/assets/planning/download/pdf/planning-level/nyc-population/census2010/t_sf1_p9_nyc.pdf

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

B. Foursquare location dataset: The Foursquare dataset has the locations of different venues at each Neighborhood in Queens, NYC. The Foursquare API is used to gather restaurants and Asian restaurants of each neighborhood in order to identify the neighborhoods with minimum number of restaurants and Asian restaurants. This data is accessible [here](#).

Data sample:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	Elmhurst	40.744049	-73.881656	Little House Cafe 聚會園茶餐厅	40.741853	-73.874486	Asian Restaurant
2	Forest Hills	40.725264	-73.844475	MoCA Asian Bistro	40.720983	-73.845436	Asian Restaurant
3	Forest Hills	40.725264	-73.844475	Asian Bowl	40.726644	-73.851988	Asian Restaurant
4	Flushing	40.764454	-73.831773	Peking Duck Sandwich Stall	40.758883	-73.829913	Asian Restaurant
5	Flushing	40.764454	-73.831773	趙記港式雲吞粥粉麵飯	40.758444	-73.831640	Asian Restaurant

C. Asian population dataset: It provides the Asian population count and percentage in each Queens neighborhood. This dataset serves as the background information to assess whether a neighborhood has the potential customer base to support a new Asian Restaurant. It is accessible [here](#), through the statistical Atlas website.

Data sample:

	Neighborhood	Perctg_of_Asian_Population	Count	Rank
0	Astoria	15.90	26800	31
1	Auburndale	49.00	13000	6
2	Bay Terrace	24.00	3212	21
3	Bayside	43.50	21100	7
4	Bellerose	38.80	11600	11

3.Methodology

Based on the three datasets described in the Data Section, the final purpose is to obtain the neighborhoods information (latitude and longitude), venues (Asian restaurants), and Asian populations, in order to run a Classification analysis with K-mean Clustering.

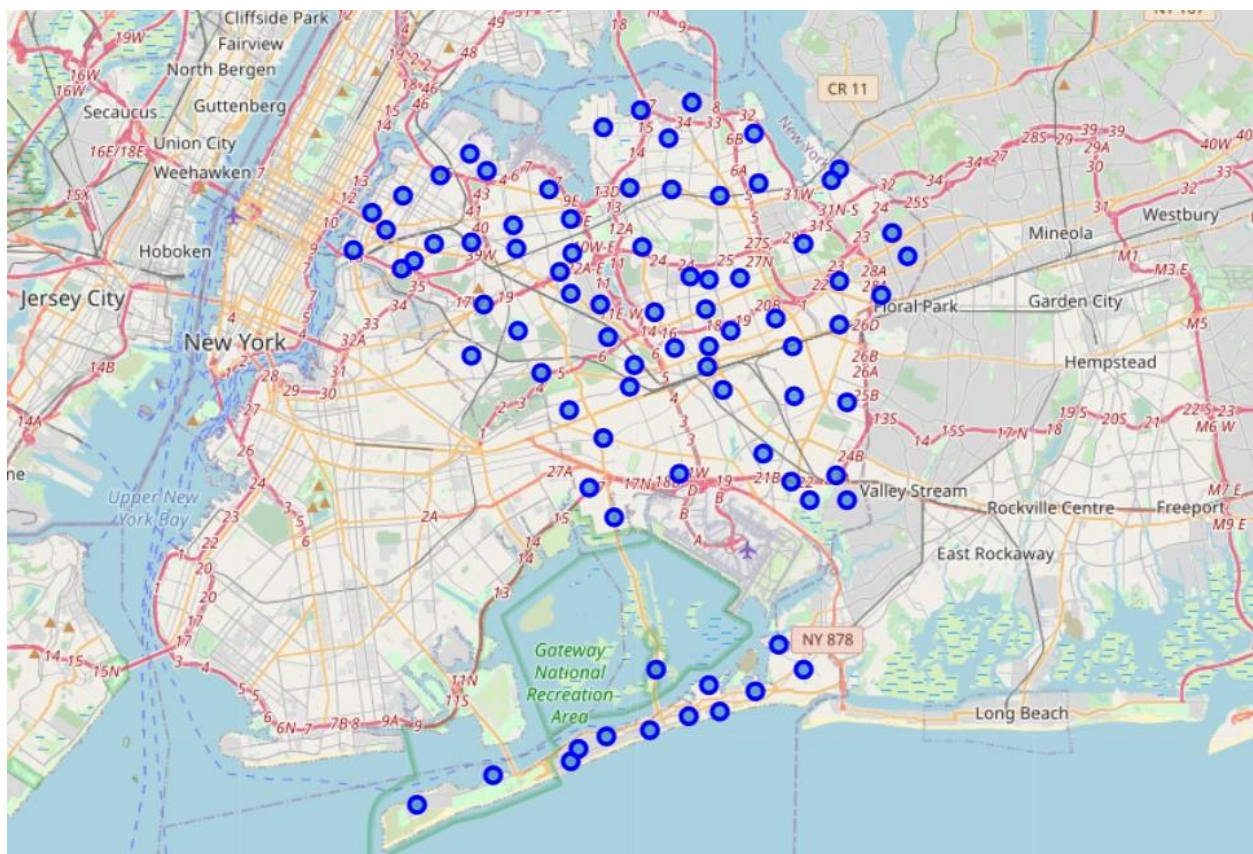
The "Elbow" method is used to determine the best number of k for the K-mean engine. Once k is determined, we will proceed to classify the neighborhoods and label them by cluster. The classified neighborhoods will be plotted on a Folium map for Analysis.

After analyzing each cluster, it turns out that one of them has neighborhoods with zero Asian restaurant and enough Asian residents. For further analysis, some neighborhoods in this cluster are chosen which have higher Asian population and lower number of restaurants. The data of these three top neighborhoods are captured in a data frame.

The use of these types of analysis and machine learning technique will help to find the neighborhood that lack of Asian restaurant and has a sizeable potential client's pool. Such location has a greater success rate to start a new business.

3.1 Data Analysis

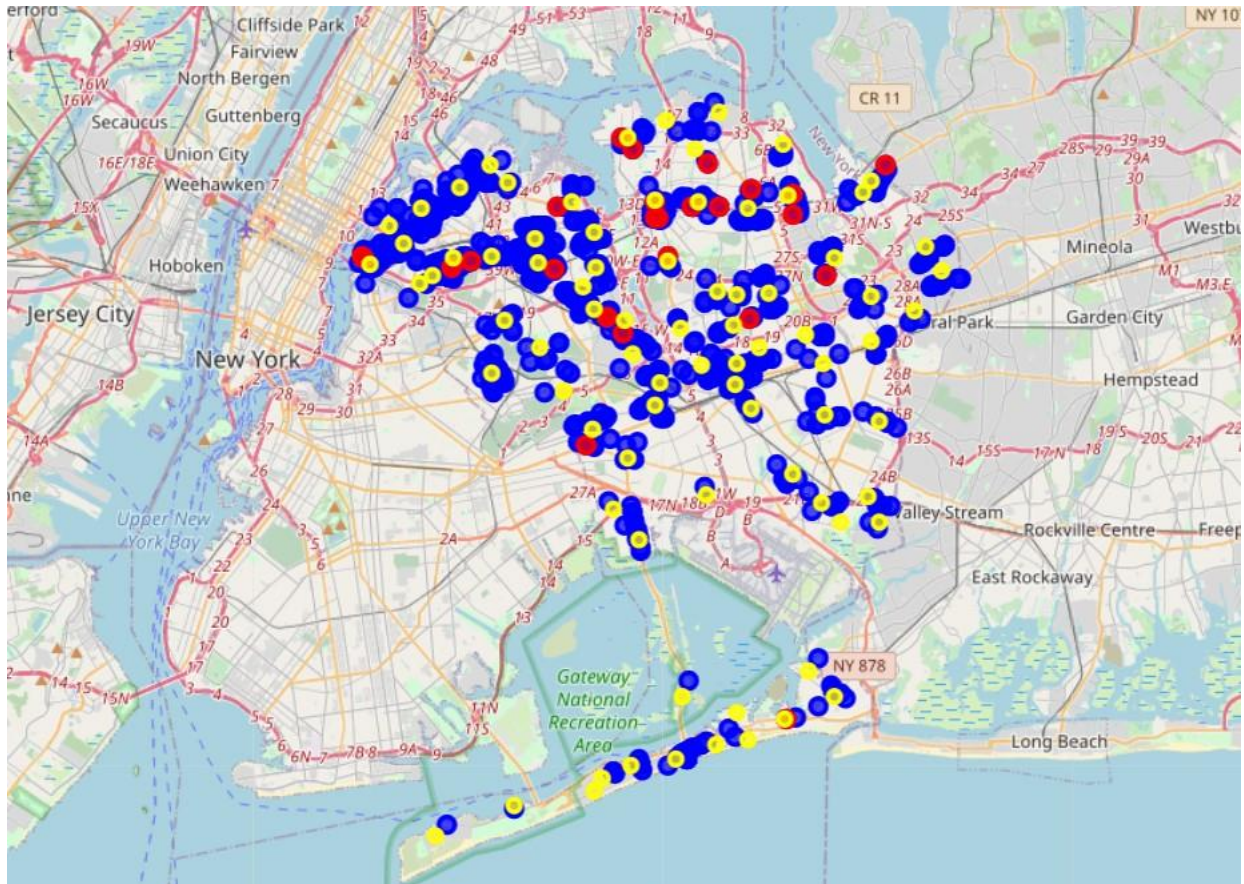
First, I used the New York City dataset to get the coordinates of each neighborhood in Queens. The raw data that was loaded was a Jason file with several features. It then was transformed to a panda data frame and borough, neighborhood, latitude, and longitude were extracted. This table was then filtered by borough and the final data frame returned 81 rows that each represented a neighborhood in Queens with their coordinates. This data was used to create a folium map that each neighborhood was illustrated by a blue dot.



Second, I used the Foursquare API to get the venues information (category, latitude, and longitude) in each neighborhood. After setting credential and connecting to the server, the call returned a JSON file

and it was put into a data-frame. It included 100 popular spots for each neighborhood were chosen with a radius of 750 meters. The one hot encoding for getting dummies of the venue category was ran, and rows were grouped by neighborhood to calculate the mean of the frequency of occurrence of each category.

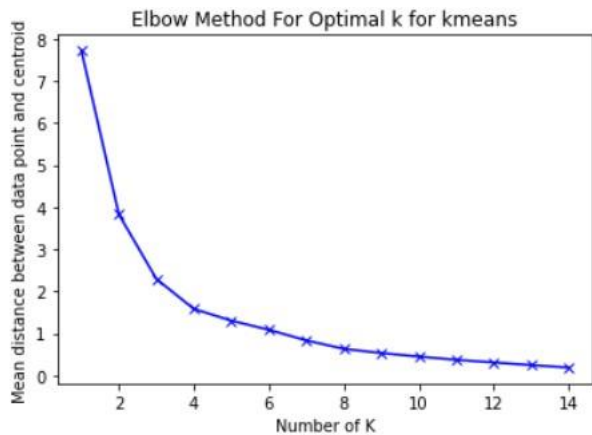
Next, the venues dataset was further filtered to return restaurants and Asian restaurants in each neighborhood. The below map, created by folium, show all restaurants in Queens in blue dots. Asian restaurants are illustrated with red dots and neighborhoods' centers are shown with yellow dots.



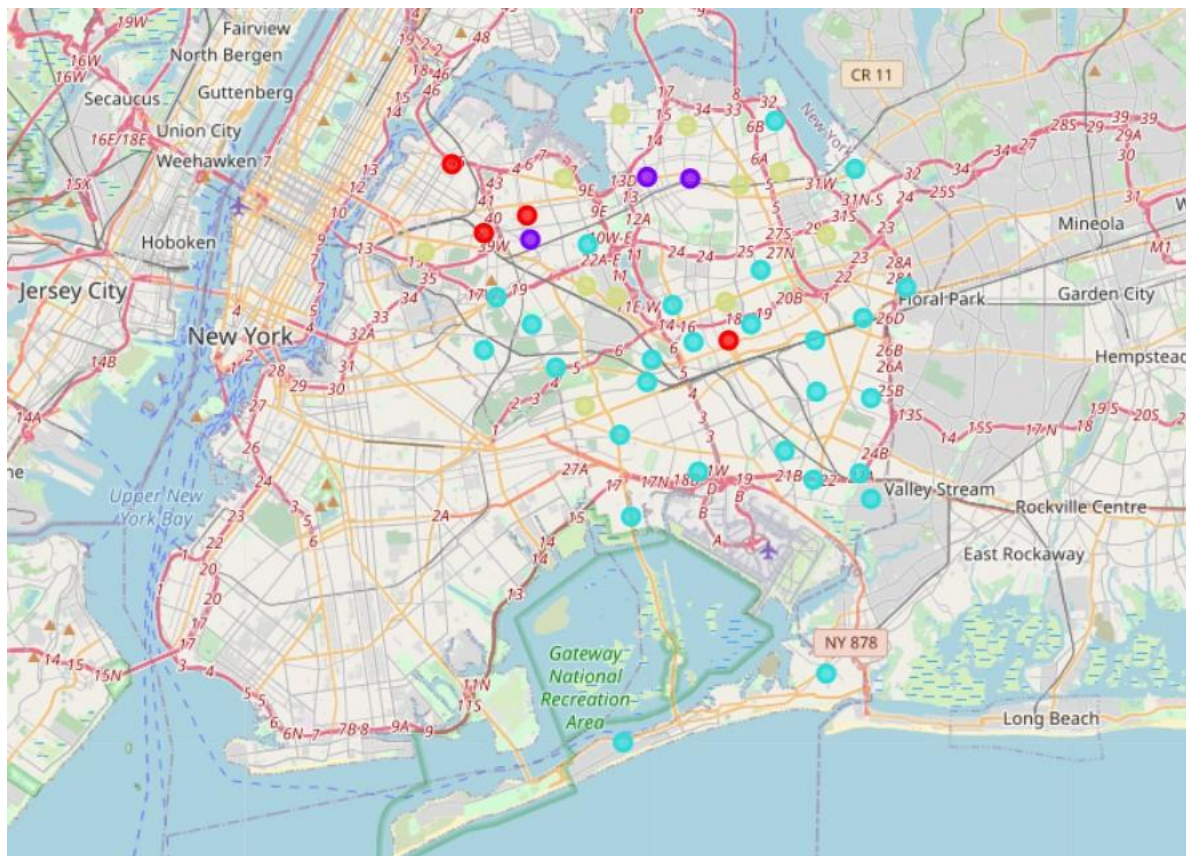
Third, Queens Asian population dataset was used to get the number of Asian residents in each neighborhood. Then, this data frame was merged with the previous tables to form a new data frame with these columns: neighborhood, total restaurant, Asian restaurant, Asian population, and total population.

	Neighborhood	Total Restaurants	Total Asian Restaurants	Asian_Population	Total_Population
0	Astoria	33.0	0.0	26800.0	168553.0
1	Auburndale	20.0	1.0	13000.0	26530.0
2	Bay Terrace	4.0	0.0	3212.0	13383.0
3	Bayside	30.0	2.0	21100.0	48505.0
4	Bellerose	3.0	0.0	11600.0	29896.0

Next, I used k-means clustering technique. The K-means clustering was performed based three criteria: total restaurants, Asian restaurants, and Asian population. Next, the "Elbow" method was used in order to determine the best number of k, the numbers of centroid. From the "Elbow Method for Optimal k for k-means" plot it was determined that the best k had to be 4. It is where the slope changes rapidly and more K does not considerably increase the accuracy of our model.



In the map below, neighborhoods are divided into four clusters and depicted with different colors. Cluster one is represented by red dots, cluster 2 is purple, cluster 3 is blue, and cluster 4 is green.



Among these clusters, cluster 3 has no Asian restaurants and minimum number of restaurants. Some neighborhoods within this cluster have enough Asian population that makes them a good choice for our business. It is further discussed in the next section.

4.Results

As it is illustrated in the K-means clustering analysis, the neighborhoods in Queens are divided into 4 clusters based on 3 factors: Total restaurants, Asian restaurants, and Asian population.

The neighborhoods in cluster 1 have no Asian restaurants and medium to high Asian population, however, there are plenty of restaurants that makes the business environment very competitive. The cluster 2 includes neighborhoods with high Asian population and high number of restaurants. There are also 1-3 Asian restaurants in each neighborhood. It means that this cluster does not offer a good spot for our purpose. Cluster 4 has 1-2 Asian restaurants in each neighborhood with medium Asian population, which means that these neighborhoods are well served and offer low potential for a new Asian restaurant.

Cluster 3 contains neighborhoods with low number of restaurants. Most neighborhoods have less than 10 restaurants and there is no single Asian restaurant in any neighborhood within this cluster. While Asian population is low in some areas, in some other neighborhoods there are enough people as the potential customers of a new Asian restaurant. This cluster has some ideal neighborhoods for opening a new Asian restaurant. In the next section, we will discuss about the best neighborhood within the third cluster to open a new Asian restaurant.

5.Discussion

Among 27 neighborhoods in the third cluster, only 6 of them have more than 10,000 Asian population, which can be seen as the potential customers for our Asian restaurant. To further narrow the choices, we look at the total number of restaurants in these neighborhoods. In three neighborhoods, there are zero Asian restaurants and less than 5 restaurants in general, which means that there is a great potential for profitability with minimal number of competitors.

	Neighborhood	Total Restaurants	Total Asian Restaurants	Asian_Population	Total_Population
0	South Ozone Park	1.0	0.0	17100.0	74025.0
1	Bellerose	3.0	0.0	11600.0	29896.0
2	Ozone Park	5.0	0.0	16600.0	54426.0

South Ozone Park neighborhood is the top choice. It is located in the southern part of Queens, it has zero Asian population, there is only one restaurant in the neighborhood and around 17100 Asian live in the area. A new restaurant can target this population while there is almost no competitor in this neighborhood.

The second and third choices are Bellerose and Ozone Park neighborhood. Ozone Park is located next to the South Ozone Park, our top choice. It has no Asian restaurant and there are only five restaurants in

the area. It is a home to more than 16000 Asian residents which is a considerable base for our business. Bellerose neighborhood is located in the middle of Queens, it has no Asian restaurant and only 3 restaurants exist in the area. More than 11000 Asian people live in this neighborhood who can be the potential customers of our Asian restaurant.

6.Conclusion

In this project, I identified best neighborhoods in Queens to open an Asian restaurant based on three criteria: The total number of restaurants, the number of Asian restaurants, and the Asian population. To answer this problem statement, I gathered data from three sources: New York City dataset, Foursquare location dataset, and Asian population dataset. After cleaning data and combining them into new data frames, I used the K-means clustering technique, given $k=4$, to divide each neighborhood based on the three factors that was mentioned. It turned out that the third cluster offered neighborhoods with no Asian restaurant and most of them had less than 10 restaurants in total. I then transferred neighborhoods with more than 10,000 Asian population into a new data frame and chose top three neighborhoods which had minimum number of restaurants.

The top suggestion is South Ozone Park. There are no Asian restaurant and there exist only one restaurant in the neighborhood. With high number of Asian residents, there is a high demand for an Asian restaurant.

This choice is followed by two other neighborhoods: Ozone Park and Bellerose. In both neighborhoods, there is no Asian restaurant and less than 5 generic restaurants. Besides, there are enough Asian residents who can be the potential customers of a new Asian restaurant.