

Searching a Location in New York City for a Chinese Restaurant

Jiang Li

1. Introduction

1.1 Background

Restaurants have always played an essential role in the business, social, intellectual and artistic life of a thriving society. The major events of life, personal and professional, are celebrated in restaurants. Acquaintances become friends around a table in the safe and controlled environment of a restaurant. Individuals become lovers across a restaurant table, sometimes. Beyond the basic purpose of restaurants to provide food and drink, restaurants have, historically, fulfilled a human need for connection and shaped social relations. Furthermore, there is a saying that “Fashion is in Europe, living is in America, but eating is in China.” The phrase is a testament to the popularity of Chinese food around the world. Unlike the fast food society of the U.S., the Chinese select live seafood, fresh meats and seasonal fruits and vegetables from the local market to ensure freshness. This means swimming fish, snappy crabs, and squawking chickens. Even prepared foods such as dim sum or BBQ duck for to go orders must gleam, glisten, and steam as if just taken out of the oven. This is why the Chinese-style restaurant is more and more popular with the people living in U.S. especially with the Asian people. New York City (NYC), often simply called New York, is the most populous city in the United States. With an estimated population of 8,253,213 distributed over about 302.6 square miles (784 km^2), NYC is also the most densely populated major city in the United States. Therefore, opening a Chinese restaurant in New York is very promising for business.

1.2 Problem

NYC is an international metropolis with dense population, but at the same time, prices and housing prices are very high. How to choose a suitable location to open a Chinese restaurant in New York is a very difficult problem. In addition, there are already many Chinese restaurants in New York, and opening a Chinese restaurant in New York also faces competition from many peers.

1.3 Interest

Many people who plan to start their own business will be interested in this topic, especially those who want to work in the catering industry. Also, those who are already living in New York or planning to live in New York, if they are looking for a career, will also be interested in this report. Not only that, for those people all over the world who are looking for business partners and those who are looking for investment opportunities, even though they may not be able to open a Chinese restaurant in NYC themselves, they are willing to participate in investment and cooperate with others to open a Chinese restaurant in NYC and therefore they are also interested in this topic.

2. Data acquisition and description

2.1 Data sources

The geographic data of NYC is from the IBM Developer Skills Network. This dataset contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude

coordinates of each neighborhood. The second dataset is about the demographics of NYC, which is introduced by Wikipedia at https://en.wikipedia.org/wiki/Demographics_of_New_York_City. Finally, the third data set about the property prices in NYC can be found in Kaggle datasets at <https://www.kaggle.com/new-york-city/nyc-property-sales>.

2.2 Description

The geographic dataset from the IBM Developer Skills Network is used for locations visualizing with Folium package, and also used for locations searching and exploring with Foursquare API. The demographic dataset from Wikipedia is used to analyze the population distribution and composition of NYC in order to choose a suitable location with a large number of potential customers to open a Chinese restaurant. In order to reduce the cost of opening a Chinese restaurant, we must consider the issue of housing rent. The Kaggle dataset shows the property sales in NYC and this dataset is used to estimate the housing rent of a location in NYC. These three datasets are used and analyzed together to search a proper location for opening a Chinese restaurant where is convenient and accessible for most people especially for the Asian people living in NYC and is also with acceptable housing rent.

3. Methodology

3.1 Map of neighborhoods in NYC

Using the geographic dataset from the IBM Developer Skills Network, we got the data contains 306 neighborhoods with latitude and longitude coordinates. With the coordinates data of the neighborhoods, I applied Folium package to generate the map of NYC with neighborhoods superimposed on top.

3.2 Population compositions and distribution in NYC

In order to get the demographic data of NYC, a web-scraping code with implementation of Beautiful Soup was used to get the table of race compositions in NYC from the Wikipedia webpage. The data was cleaned and was finally converted to a Pandas data frame. Subsequently, the population distribution in NYC was visualized with table, bar chart, and pie chart.

3.3 Exploring the housing price in NYC

The property sales dataset from Kaggle was cleaned. The distribution of housing price in NYC was plotted in a histogram. The average housing price in the five boroughs of NYC was shown in a bar chart. Especially, I explored the housing prices for the commercial units and employed the box plot to show the price distribution of commercial units in the five boroughs. In addition, relation between the number of commercial units and the sale price was explored and the results were visualized in a box plot.

3.4 Preliminary selection of the location

Based on the population distribution and the housing price information in NYC, the location for opening a Chinese restaurant was preliminarily determined. The approximate location is in Queens borough in NYC (more details are discussed in **Section 4**).

3.5 Decreasing the searching space

Once the preliminary location is determined, Four Square API was used to explore the neighborhoods around the selected borough. The top 20 most popular venues around the venues were analyzed and explored. K-means clustering algorithm was used to cluster the neighborhoods to see the neighborhoods distribution patterns.

3.6 Define a criterion of predicting the best place for opening a Chinese restaurant

If a place has a lot of restaurants, it indicates this place is suitable for opening a restaurant. However, if this place has a large number of Chinese restaurants, it means this place is very competitive to open a Chinese restaurant. Therefore, I defined parameter score. If a place has a restaurant but not a Chinese restaurant, this place will get 1 point; if this place has a Chinese restaurant already, it will get -1 point. By analyzing the top 20 most popular venues of the neighborhoods, we get an accumulated score for each location. For the top 20 most popular venues, the weights of points are different. If a restaurant (not a Chinese restaurant) is the 1st most popular venue in this location, the weight is 100%. As the rank of the most popular venues increases, the weight decreases. The weight of point is defined with the following equation,

$$w = \frac{100-k \times i}{100}$$

where k is a constant, i is the rank of a venue in the most popular venues. If k is a large number, the weight w decreases dramatically as i increases. In this project, we set k to be 0.5. With the weight of a point defined, we got a table of weighted points and finally got an integrated score for a location which is used as a criterion of evaluating the fitness of a place for opening a Chinese restaurant.

3.7 Visualize the best place for opening a Chinese restaurant

In order to visualize the final results, the score map was created with Folium with locations and corresponding scores labeled on the map. The larger radius of the cycle on the map, the better fitness of this place to open a Chinese restaurant.

4. Results and discussions

4.1 Map of NYC neighborhoods

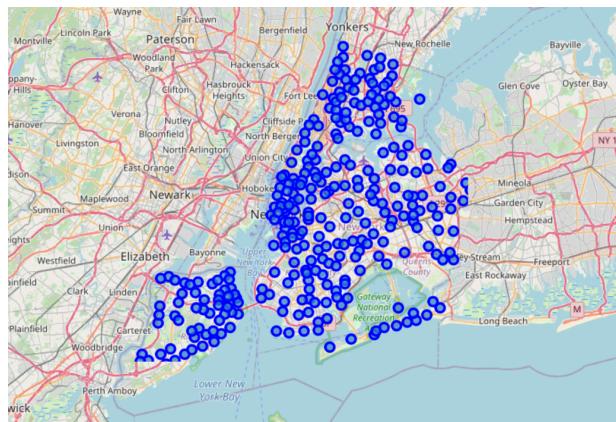


Figure 1 Map of NYC neighborhoods. 306 neighborhoods are shown in this map, distributed in five boroughs (Brooklyn, Queens, Manhattan, Bronx, and Staten Island) of NYC.

The map of NYC is shown in **Figure 1**. In this map, 306 neighborhoods are marked out. The neighborhoods are almost evenly distributed in five boroughs of NYC, i.e., Brooklyn, Queens, Manhattan, Bronx, and Staten Island.

4.2 Population compositions in NYC

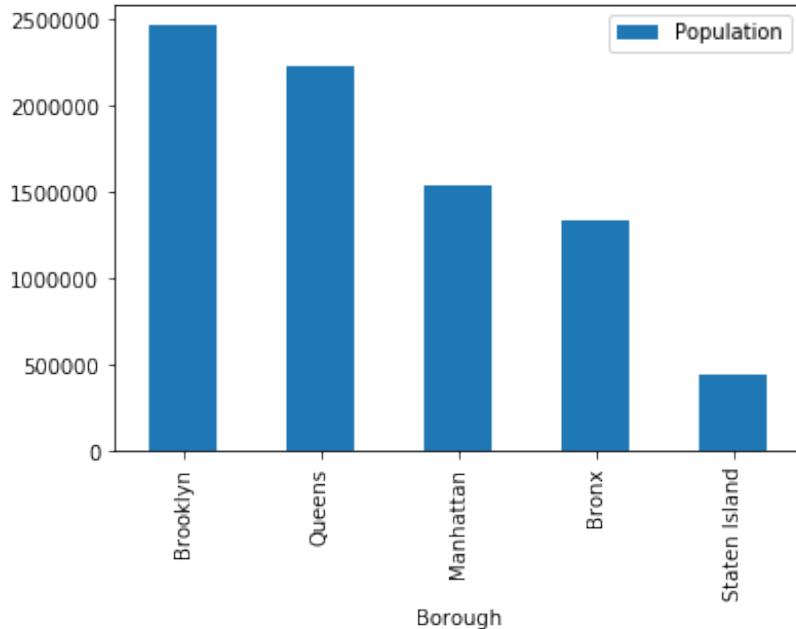


Figure 2 Population in five boroughs of NYC.

Figure 2 shows that Brooklyn has the highest population with number of 2500000, followed by Queens, Manhattan, Bronx, and Staten Island. High population means large number of potential customers. The boroughs with large population is potentially an ideal area for opening a restaurant.

Table 1 Race compositions in the five boroughs of NYC

| | Borough | White(%) | AfricanAmerican(%) | Asian(%) | Mixedrace(%) | Other(%) |
|---|---------------|----------|--------------------|----------|--------------|----------|
| 0 | Brooklyn | 41.2 | 36.4 | 7.5 | 4.3 | 10.6 |
| 1 | Queens | 44.1 | 20.0 | 17.6 | 6.1 | 12.3 |
| 2 | Manhattan | 54.4 | 17.4 | 9.4 | 4.1 | 14.7 |
| 3 | Bronx | 29.9 | 35.6 | 3.0 | 5.8 | 25.7 |
| 4 | Staten Island | 77.6 | 9.7 | 5.7 | 2.7 | 4.3 |
| 5 | NYC Total | 44.7 | 26.6 | 9.8 | 4.9 | 14.0 |

We are going to open a Chinese restaurant; therefore, we need to know the race compositions in NYC. Generally speaking, most of the frequent and stable customers of a Chinese restaurant are Asian people. From the above table, it shows that Queens has the highest Asian ratio among all the five boroughs, which can be visualized in the bar chart in **Figure 3** and pie charts in Figure 4.

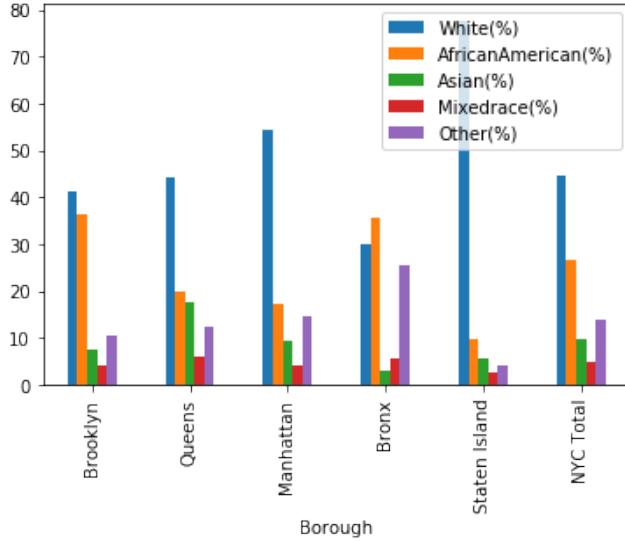


Figure 3 Bar chart showing the race compositions in the five boroughs in NYC.

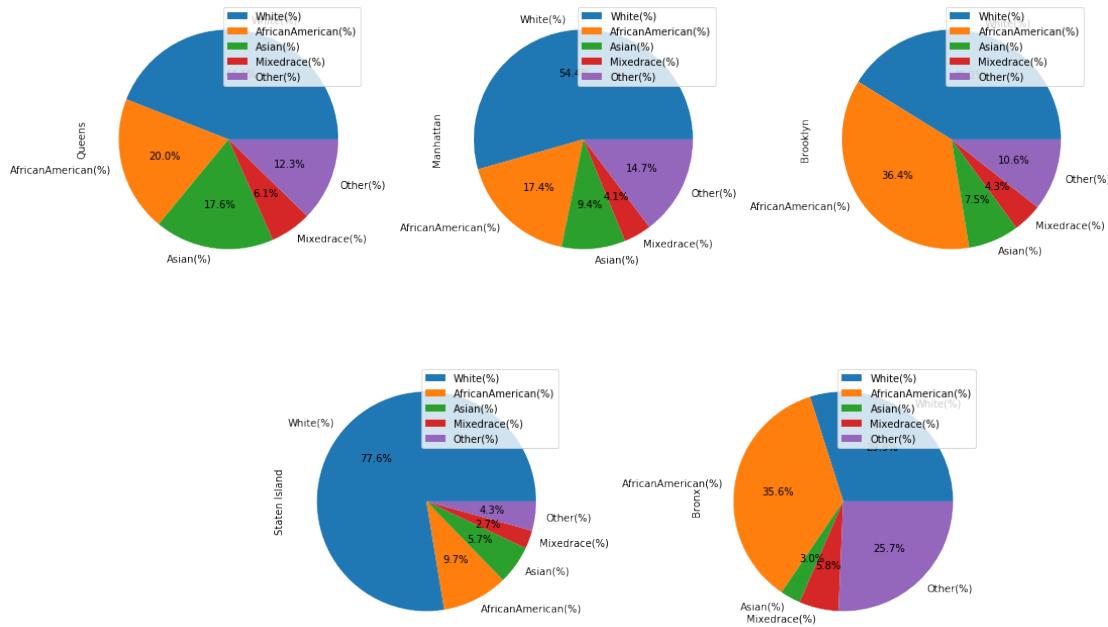


Figure 4 Pie charts of race compositions in Queens, Manhattan, Brooklyn, Staten Island, and Bronx. The Asian ratios are respectively 17.6%, 9.4%, 7.5%, 5.7%, and 3.0%.

Based on the demographic data shown above, Queens has the second largest population among the five boroughs of NYC. In addition, Asian people are majorly distributed in Queens. Therefore, we can approximately determine that Queens is a proper borough to open a Chinese restaurant. Of

course, from the perspective of customer sources, Queens is definitely an ideal borough for opening a Chinese restaurant, but how about the housing price in Queens? We should also consider the expense of opening a Chinese restaurant. In the next, we will see the housing prices in the five boroughs of NYC.

4.3 Housing price in NYC

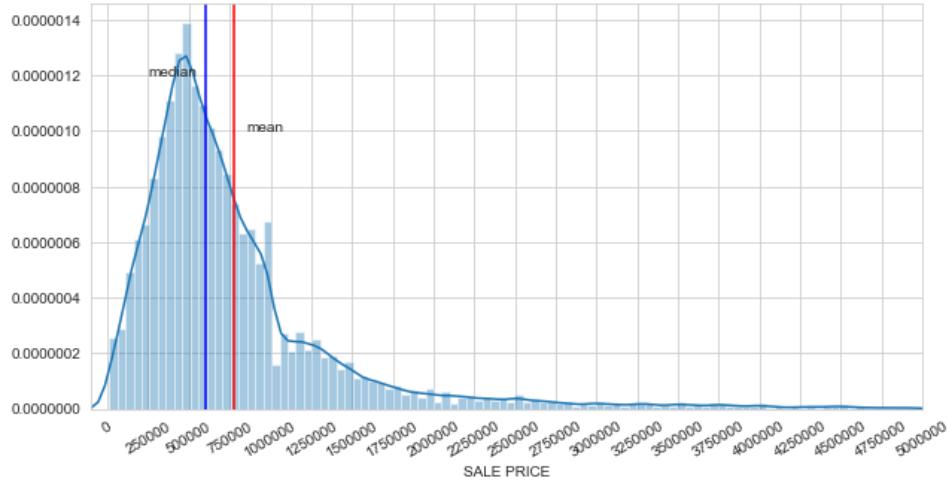


Figure 5 Distribution of housing price in NYC.

The distribution of housing price in NYC is shown in **Figure 5**. It's apparent that the housing price distribution is roughly a right-skewed normal distribution with a median of \$ 500000 and mean of \$ 750000. In the following, let's see the average housing price in the five boroughs of NYC.

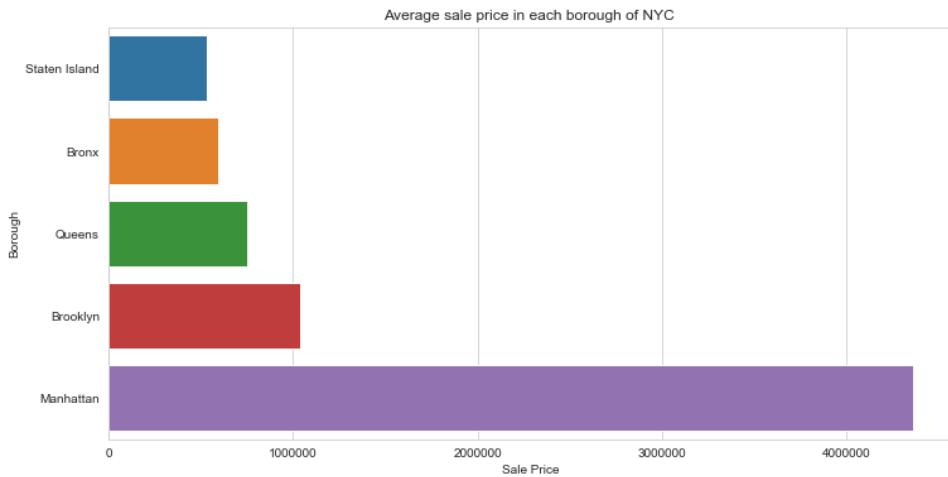


Figure 6 Bar chart showing the average housing price in the five boroughs of NYC.

As shown in **Figure 6**, the average housing price in Manhattan is the highest, which is above \$ 4000000, followed by Brooklyn, Queens, Bronx, and Staten Island. Because we are going to open a restaurant instead of a residential unit, we should not just focus on the average housing price. In the next, let's see the prices of commercial units in the five boroughs of NYC.

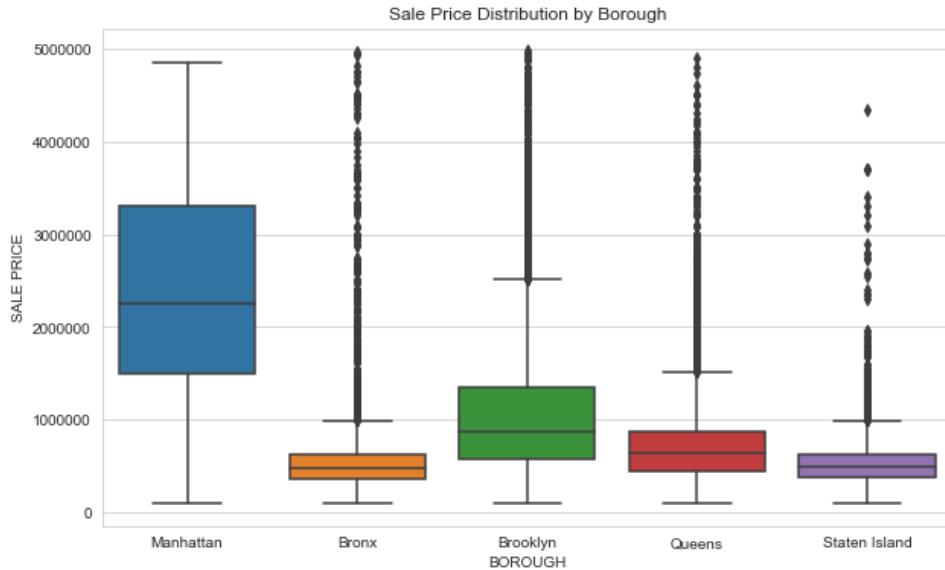


Figure 7 Prices of commercial units in the five boroughs of NYC.

It also shows that the price of commercial units in Manhattan is apparently higher than that in other boroughs. The prices of commercial units in Staten Island and in Bronx are comparable and are slightly lower than that in Queens. Overall, the price of commercial units in Queens is acceptable.

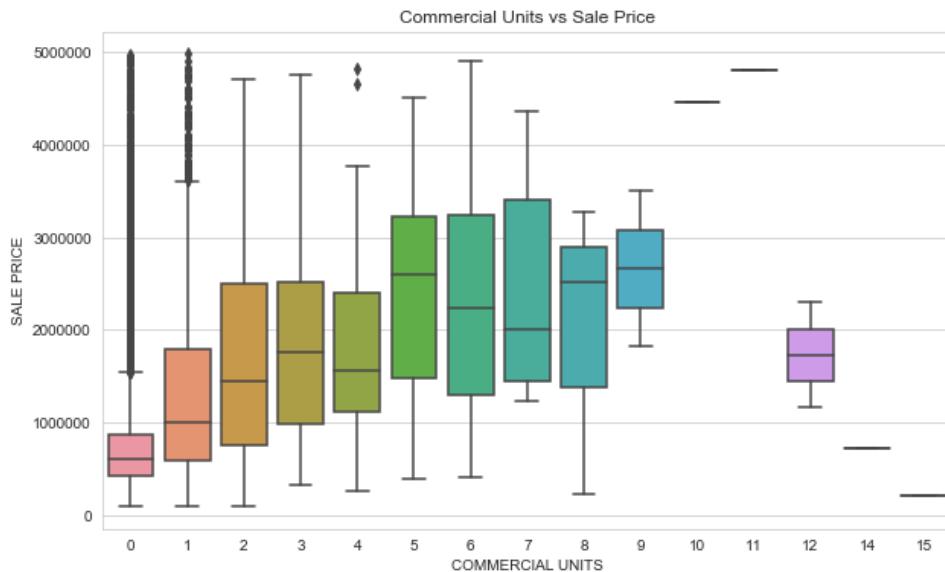


Figure 8 Prices of commercial units with relationship to the number of units.

As we know, the price of a commercial unit is dependent on the number of commercial units. The results are shown in **Figure 8**. It indicates that price goes up as the number of units increases and reach the highest at unit number of 5-8 and then the prices drops down as further increasing the number of units. According to the data shown above, 1 unit is too small and 5 or more is too expensive. Therefore, a proper number of units is 4, which is spacious enough and also has reasonable price.

4.4 Preliminary result and its neighborhoods

Based on population data analysis and the housing price analysis, we can roughly determine that Queens is the best borough to open a Chinese restaurant in NYC. In the next, let's explore the neighborhoods in Queens.

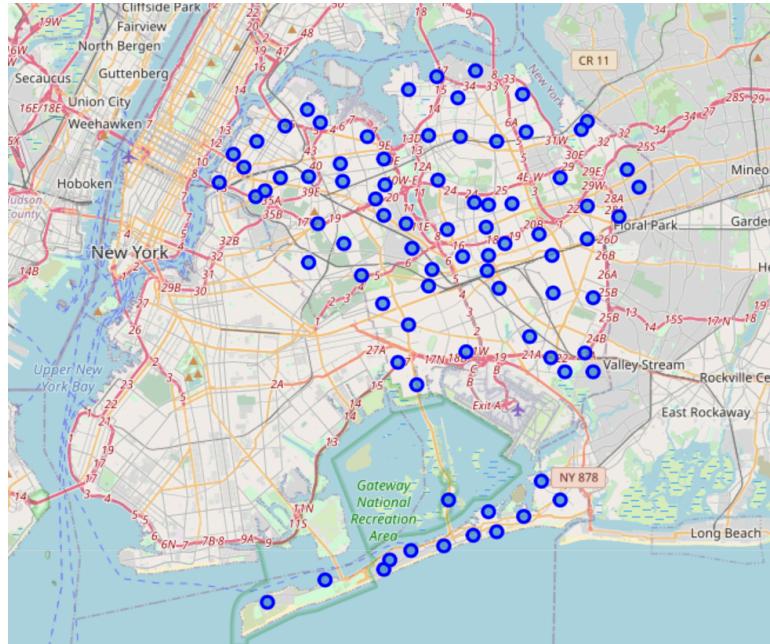


Figure 9 Neighborhoods distribution in Queens

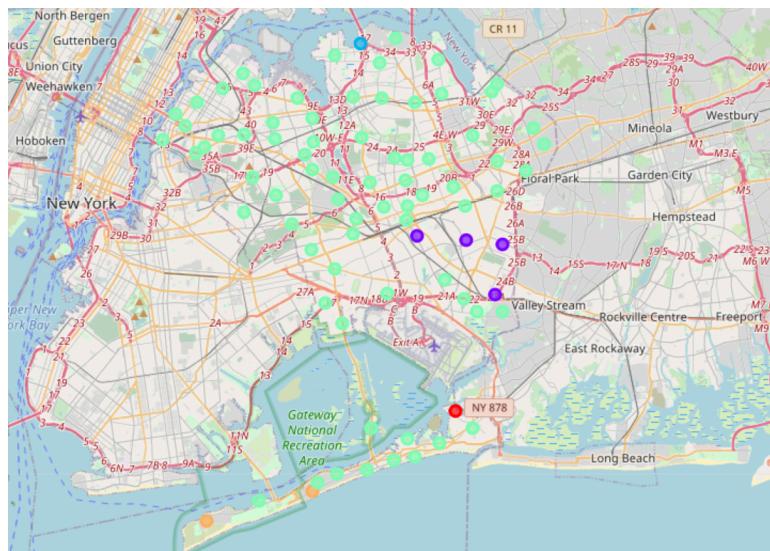


Figure 10 Five clusters of neighborhoods in Queens analyzed with K-means cluster algorithm.

There are 81 neighborhoods in Queens and are evenly distributed as shown in **Figure 9**, except the airport area with very sparse neighborhoods. With K-means clustering algorithm, the neighborhoods in Queens are assigned into five different clusters shown in **Figure 10**.

4.5 Most common venues around the neighborhoods

The top 20 most common venues in each neighborhood in Queens are explored with the FourSquare API.

Table 2 (part) Most common venues around neighborhoods in Queens

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|-----------------|---------------------------|-----------------------|---------------------------|---------------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------------|------------------------|
| 0 | Astoria | Middle Eastern Restaurant | Bar | Greek Restaurant | Café | Hookah Bar | Seafood Restaurant | Bakery | Indian Restaurant | Mediterranean Restaurant | Deli / Bodega |
| 1 | Woodside | Grocery Store | Bakery | Thai Restaurant | Latin American Restaurant | Filipino Restaurant | American Restaurant | Donut Shop | Bar | Pub | Café |
| 2 | Jackson Heights | Latin American Restaurant | Peruvian Restaurant | South American Restaurant | Bakery | Mobile Phone Shop | Supermarket | Thai Restaurant | Diner | Empanada Restaurant | Spanish Restaurant |

From the data in **Table 2**, we see that restaurants are one of the most common venues in most of the neighborhoods. If a neighborhood has lots of restaurants, it indicates that neighborhood is proper to open a restaurant but the competition there may be very severe at the same time. Based on the score parameter defined in the Methodology section (see **Section 3.6**), the fitness score of opening a Chinese restaurant for each neighborhood is shown in **Table 3a and 3b**.

Table 3a Top 5 best places for opening a Chinese restaurant.

| | Neighborhood | normalized_score | Latitude | Longitude |
|---|-------------------|------------------|-----------|------------|
| 0 | Jackson Heights | 100.00 | 40.751981 | -73.882821 |
| 1 | Elmhurst | 89.30 | 40.744049 | -73.881656 |
| 2 | Sunnyside Gardens | 87.55 | 40.745652 | -73.918193 |
| 3 | Bayside | 77.02 | 40.766041 | -73.774274 |
| 4 | Corona | 75.63 | 40.742382 | -73.856825 |

Table 3b Top 5 worst places for opening a Chinese restaurant.

| | Neighborhood | normalized_score | Latitude | Longitude |
|----|----------------|------------------|-----------|------------|
| 76 | Jamaica Center | 11.63 | 40.704657 | -73.796902 |
| 77 | Blissville | 10.53 | 40.737251 | -73.932442 |
| 78 | Rosedale | 10.41 | 40.659816 | -73.735261 |
| 79 | Arverne | 10.30 | 40.589144 | -73.791992 |
| 80 | Forest Hills | 0.00 | 40.725264 | -73.844475 |

Table 3a shows that the top 5 best places in NYC for opening a Chinese restaurant are Jackson Heights, Elmhurst, Sunnyside Gardens, Bayside, and Corona and the fitness scores are 100, 89.3, 87.55, 77.02, and 75.63, respectively. **Table 3b** shows that the top 5 worst places in NYC for opening a Chinese restaurant are Forest Hills, Arverne, Rosedale, Blissville, and Jamaica Center

with fitness scores below 20. The worst 5 places for opening a Chinese restaurant are either no restaurants in those places or a lot of Chinese restaurants have been already in those places. In order to visualize the best places in NYC for opening a Chinese restaurant, the locations are marked as cycles on the score map. The larger the cycle is, the better the place is for opening a Chinese restaurant.

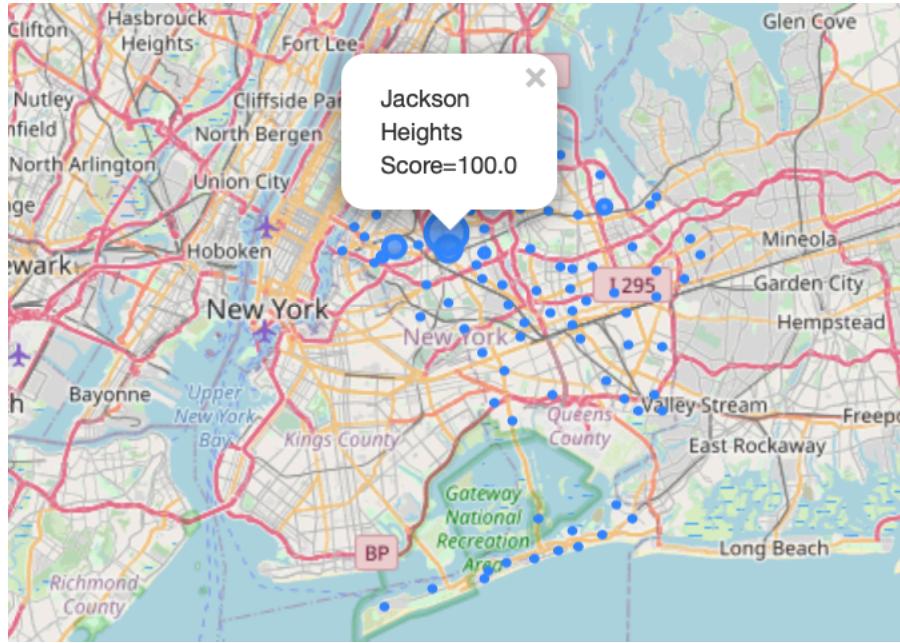


Figure 11 Score map showing the fitness scores for opening a Chinese restaurant in the neighborhoods in Queens of NYC.

According to the score map shown in **Figure 11**, Jackson Heights has the largest radius, which indicates that Jackson Heights is the best place to open a Chinese restaurant. All the other 4 places that are proper to open a Chinese restaurant are very close to Jackson Heights and all these places are located in the northern Queens in NYC.

5. Conclusions

In this project, I analyzed the population distribution in NYC and found that Queens borough has the second largest population and the highest Asian people percentage. In addition, the housing price in Queens is relatively lower among all the five boroughs of NYC. Therefore, I approximately determined that the Chinese restaurant should be opened in Queens. Furthermore, I explored the most common venues around each neighborhood in Queens and found that some neighborhoods have lots of restaurants (non-Chinese style) and some neighborhoods have large number of restaurants but most of them are Chinese-style. I defined a fitness score to evaluate whether a place is proper or not to open a Chinese restaurant. The best place is Jackson Heights with fitness score as high as 100, followed by Elmhurst (89.3), Sunnyside Gardens (87.55), Bayside (77.02), and Corona (75.63). In addition, the housing price analysis reveals that buying or renting a commercial housing with 4 units is the best choice to open the restaurant.