# Capstone project

# where should I open a Chinese restaurant in New York city?

Jingya Zhang

05/11/2021

## Business Problem

New york city has the largest Chinese population outside of Asia and houses thousands of Chinese restaurants. Originally Chinese restaurants in new york city were almost exclusively Chinese-American and Cantonese. They were concentrated in Chinatown. Recently more immigrants began streaming into new york city from multiple regions of China, bringing all kinds of authentic Chinese dishes. These authentic Chinese food almost look nothing like American Chinese food. They vary tremendously by region: Sichuan and Hunan cuisine are spicy, Shandong and Dongbei cuisine are generally salty and crispy, and Shanghai cuisine is famous for “red-cooking”.

With an estimated Chinese population of 628,673 already, new york city’s demand for Chinese restaurant are continuously growing, especially for authentic Chinese restaurants. Here we explore which neighborhoods in new york city are suitable for opening a Chinese restaurant. Because American Chinese restaurants are so different from authentic Chinese restaurants, so we will research them separately.

The target audiences are Chinese immigrants or business owners who want to open Chinese restaurant or grow their current business.

## Data source and acquisition

### Source 1: New York Neighborhoods via coursera server

Like what we did in the coursera lab, we download a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood from coursera server.

A picture containing table

Description automatically generated

### Source 2: We use FourSquare API get all the venues in the neighborhoods of new York city, especially Chinese restaurants.

* Pull venue data for all neighborhoods from the Foursquare API.
* Select all the Chinese restaurant venues from new york venues data. Out of 10106 venues in 5 boroughs of new york city, 262 of them are Chinese restaurants.

Graphical user interface, table

Description automatically generated with medium confidence

## Methodology

### Data exploration

* Select representative American Chinese restaurants and authentic Chinese restaurants from the Chinese restaurant venues.
  + 1. American Chinese restaurant was filtered by the criteria that name of venue contains most common American Chinese word such as “Great Wall”, “Wok”, “Dragon”, “Golden”, “China”, “Panda” and “lucky”.
    2. Authentic Chinese restaurant was filtered by the criteria that the venues category is one of these: ‘Szechuan Restaurant’, ‘Hunan Restaurant’, ‘Hotpot Restaurant’ or ‘Dim Sum Restaurant’, ‘Cantoese Restaurant’.

* Create a map using Folium and map all the Chinese restaurants in new york city, representative American Chinese restaurants and authentic Chinese restaurants were color-coded as blue and red separately.

Map

Description automatically generated

### Machine learning – cluster 306 neighborhoods in new york city

* Use one hot encoding, individual venues for each neighborhood were turned from categorical data into numerical data 0 or 1, then the rows were grouped by boroughs and neighborhoods taking the average frequency of occurrence of each venue category. We also define a function return\_most\_common\_venues to get the top\_10 most common venues for each neighborhood.
* Then we use K-Means clustering to segment the neighborhoods based on average frequency of occurrence of each venue category in each neighborhood.
  + To get our optimum K value, we used the Yellowbrick package from KElbowVisualizer to accurately point out the optimum K value, which is 11.
  + Use K-Means clustering, neighborhoods that had a similar mean frequency of venue category were put into the same cluster. All the 306 neighborhoods were divided into 11 clusters.
* Using Folium and map all the neighborhoods, and each neighborhood is colored based on the cluster label.

Chart, map

Description automatically generated

* Calculate top 10 venues for each cluster and characterize each cluster.

### Incorporate Chinese restaurant venues with cluster labels. View the distribution of Chinese restaurants among different clusters and find the pattern.

## Results

### First we characterize the 11 clusters of neighborhoods in new york city.

Chart, bar chart

Description automatically generated

Cluster 1, 5,8,9 only has 1 neighborhood each;

Cluster 6 has 2 neighborhoods; cluster 2 has 4 neighborhoods; cluster 3 has 7 neighborhoods and all of them are from Staten Island.

So we can focus on 4 main clusters: 0,4,7,10.

Chart, bar chart

Description automatically generatedCluster 0 has 118 neighborhoods, the top 5 common venues are coffee shop, Italian restaurant, pizza place, bar and café.

Most of manhattan neighborhoods are in cluster 0. Some fancy Brooklyn downtown/waterfront neighborhoods and populated Queens neighborhoods in queens are also in cluster 0.

Cluster 0 neighborhoods are busy and popular area with large amount of foot traffic, lots of business, restaurants, stores and night life.

Cluster 4 has 30 neighborhoods, the top 5 common venues are deli/bodega, pizza place, Italian restaurant, bus stop and supermarket.

Cluster 7 has 104 neighborhoods. only 2 of them are located in Manhattan, all other 102 are from other 4 boroughs of new york city.

Cluster 7 neighborhoods are relatively less busy residential neighborhoods compared to cluster 0. The top 5 most common venues are pizza place, sandwich place, bank, pharmacy and donut shop.

Cluster 10 has 39 neighborhoods, and the top 5 most common venues are bus station, Caribbean restaurant, playground, Chinese restaurant and park.

### The distribution of Chinese restaurants among different clusters

Chart, bar chart

Description automatically generated Chart

Description automatically generated

Chinese restaurants are only located in the 4 main clusters. Cluster 0 has 139 Chinese restaurants, the highest number of all clusters. But cluster 0 has the highest number of neighborhoods, wo we divide the number of Chinese restaurants by neighborhoods, so we can compare average number of Chinese restaurants per neighborhood among clusters.

The distribution of Chinese restaurants by cluster labels on map.

Map

Description automatically generated

### The distribution of authentic Chinese restaurants among different clusters

Chart, bar chart

Description automatically generated Chart, histogram

Description automatically generated

Authentic Chinese restaurants are only located in cluster 0 and cluster 7. Cluster 0 has 46 authentic Chinese restaurants; cluster 7 has 7 authentic Chinese restaurants.

From the map, we can see that authentic Chinese restaurants are only in high foot traffic area cluster 0. The less busy neighborhoods in cluster 7 that has authentic Chinese restaurants have high density of Chinese population.

Map

Description automatically generated

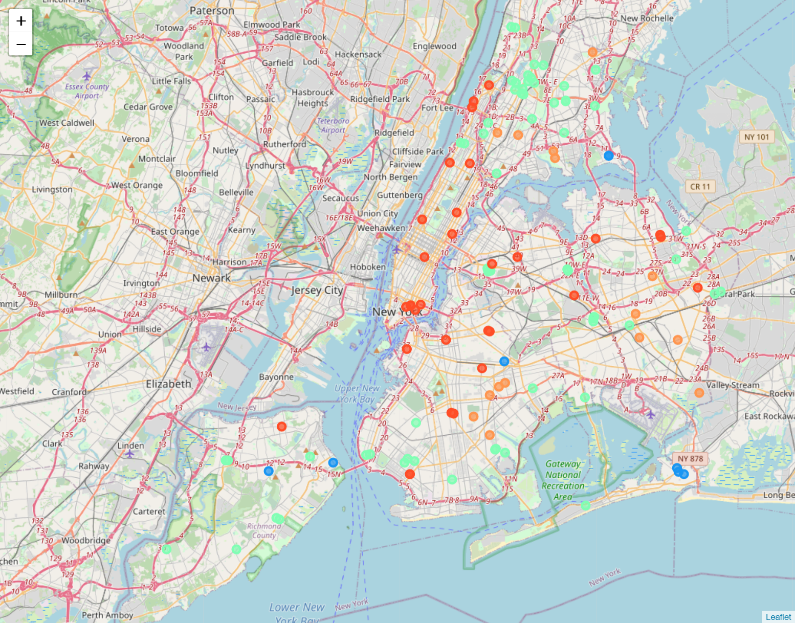
### The distribution of American Chinese restaurants among different clusters

Chart, bar chart

Description automatically generatedChart, bar chart

Description automatically generated

By number cluster7 has the most American Chinese restaurants, cluster 4 has the least. By per neighborhood, American Chinese restaurants are relatively evenly distributed among different clusters (cluster 7 and 10 has a little bit higher number).



We can get the same conclusion by distribution map.

## Discussion

### The difference between authentic Chinese restaurants and American Chinese restaurants

If you want to open a Chinese restaurant, first you have to know what kind of Chinese restaurant do you want to open. American Chinese food is very different from authentic Chinese food.

American Chinese food have adapted to American people for centuries by adding American vegetables like celery, broccoli, and bell peppers; replacing pork with beef beef and shrimp. Recipes were also adapted to sweet and sour American tastes, avoiding strong flavors like fish sauce and spicy ingridients. The result was such signatures as chop suey, chow mein, egg foo yung, beef with broccoli, and mGeneral Tsao’s chicken.

But before they came to USA, most Chinese people have never heard of General Tsao’s chicken and they hardly can believe this overly sauced fried chicken is Chinese dish. Compared to General Tsao’c chicken, they much more prefer lazi ji (chicken with chillies) – spicy Sichuan dish that consists of fried chunks of chicken breast with peppercorn, toasted sesame and chili. Authentic Chinese restaurants serve Chinese people mainly and they provide authentic Chinese dish instead of the American version, like scallion pancake instead of egg roll, yuxiangrousi (shredded pork in hot garlic sauce) instead of beef with broccoli, zhajiang mian instead of lo mein.

So the neighborhoods is suitable for authentic Chinese restaurants may be not a fit for American Chinese restaurants, and vice versa.

### Where should I open an authentic Chinese restaurant?

From our results, we can see authentic chinese restaurants are mainly(46) in cluster 0. Seven authentic Chinese restaurants are located in cluster 7. Cluster 7 is not as busy and popular as cluster 0, but neighborhoods inside cluster 0 that authentic Chinese restaurants are located always have lots of Chinese immigrants. So if you want to open an authentic Chinese restaurants, there are 2 main factors you need to take into consideration: Chinese population and foot traffic.

The 2 biggest chinatown, chinatown(Manhattan) and flusing(Queens) each has 10 authentic chinese restaurants. They both have huge Chinese population and large amount foot traffic. But these areas are too competitive and expensive.

On the distribution map, we can see other areas that would be more cost-effective for authentic chinese restaurant:

* 1. Bayside-Douglaston-Little Neck area, which represents one of the least urban communities in new york city, with many areas having a suburban feel. New Chinese immigrants have been flooding the area because it’s near Flushing and have good schools. Authentic Chinese restaurants are following. From the map, we can see Bayside and Douglaston have 1, and little neck has 2 authentic Chinese restaurant venues. Given its growing Chinese population and relatively low rent, this area is good choice for opening authentic Chinese restaurants.
  2. Bensonhurst-bath beach in Brooklyn. There are 4 authentic Chinese restaurants in this area. Recently Bensonhurst has become home to New York City’s largest Chinese population. 86th Street — where the D line rumbles above ground — has become a mini Chinatown in itself. The Chinese immigrants live here are newly from mainland China and they miss authentic Chinese food, which makes Bensonhurst and its adjacent neighborhood bath beach a perfect location for opening a Chinese restaurant.
  3. Brooklyn downtown and waterfront area and long island city area, which are the two fastest growing area of New York city. From the distribution map, we can tell that besides Manhattan (cluster 0), flushing, Bensonhurst and little neck area, authentic Chinese restaurants are mainly located in these 2 areas. They grow so fast and they are in cluster 0 with large amounts of foot traffic. Their fast growing attract Chinese people too, although not as many as other Chinatowns. High foot traffic and certain amount of Chinese population is a good combination for opening an authentic Chinese restaurant too.

### Where should I open an American Chinese restaurant?

From our results, we can see that American Chinese restaurants are more equally distributed among different cluster compared to authentic Chinese restaurants. They don’t need to be in cluster 0 with high foot traffic, they don’t need to be in area with high density Chinese population either. American Chinese food has a much board range of customers, the price is very reasonable, and American Chinese restaurants do a lot of take-outs, so American Chinese restaurants are suitable for most neighborhoods in new york city. If you want to open an American Chinese restaurant, you should probably choose an area with no American Chinese restaurants yet.

For example, cluster 7 and 10 which has more American Chinese restaurants per neighborhood, which means that they better environmental for American Chinese restaurants. They we choose the neighborhoods in cluster 7 ,10 that don't have American Chinese restaurants right now, which might be good choices. For example, the neighborhoods in cluster 10 listed a below could be a good start.

Table

Description automatically generated

## Conclusion

For opening an authentic Chinese restaurant, the best choice will be in the areas going through transformation from cluster 7 to cluster 0 and with high density with Chinese population. So Bayside-Douglaston-Little Neck area in Queens and Bensonhurst-bath beach in Brooklyn will be the top choices. Brooklyn downtown, waterfront area and long island city area will be secondary choices, because they are already in cluster 0, has already become expensive and Chinese population are not as high as the top choices.

For opening an American Chinese restaurant, the good choice will be the neighborhoods in cluster 7 and 10 that don’t have American Chinese restaurants yet, like Kew Gardens Hills in queens, coney island in Brooklyn.