Analysis NYC Neighborhoods for New Chinese Restaurant Business

Yizheng Li 03/22/2020

1. Introduction

1.1 Background

United Sates of America is an immigration country, which create ethnic diversity. New York as the economic center of U.S., because of this unsurpassed ethnic diversity, the restaurant goer will feel that he has died and gone to heaven while in New York City. The infinite numbers of Chinese dishes, flavors, and methods of cooking which make eating Chinese food joyful and exciting. So, when someone want to open a Chinese restaurant business in NYC, which location and neighborhood is the best for business? Location is very important for business, a bustling block can help restaurant attract more people, meanwhile, the restaurant trend are also very important for a new restaurant.

1.2 Problem

For a restaurant owner, the tradition way is field investigation by ask information from people and check out each location personally. However, this is very inefficient way, people waste money and time during this process. In addition, their information may not be accurate or complete, this will cause huge economic losses when the new restaurant performance does not meet owner's expectation. The worst thing is the owner doesn't even realize the problem of location and waste more money on propaganda and other expense. Using data science to analysis and cluster Chinese food to find the best location is best way for owner, this will save both money and time. We will use data to make the decision on the place that is the most convenient for owner's new business.

2.Data acquisition and cleaning

2.1 Data source

For data and location analysis, we need geographic data, borough and neighborhoods Data, Chinese restaurant data from data sources.

- NYC borough and neighborhoods from use Wikipedia information and NYC data Json file from previous Lab: https://cocl.us/new_york_dataset
- NYC geophysics data from Foursquare location data
- Chinese Restaurants data in NYC from NYC Open Data
- NYC population from NYC Department of Health

2.2 Data Cleaning

The initial Chinese restaurant data frame contained redundant information such as street, zip codes, phone, etc. So, a new data frame was created with just the columns containing boroughs and Chinese Restaurant Name. The borough and neighborhoods data frame were cleaned and leave Borough, Neighborhoods, coordinate to give us better observation.

3. Exploratory Data Analysis

3.1 Explore the Chinese Restaurant in Different in Borough

In New York City, different borough has their own culture and food preference. If the number of Chinese restaurants in a borough is high, then it is reasonable to think people live here prefer Chinese food. New York City contains total 42450 Chinese Restaurant, the borough has high proportion reveals the taste of residence. By count the number of Chinese Restaurant in each borough, and visualized with pie chart, it is clearly that Queens and Brooklyn have most Chinese

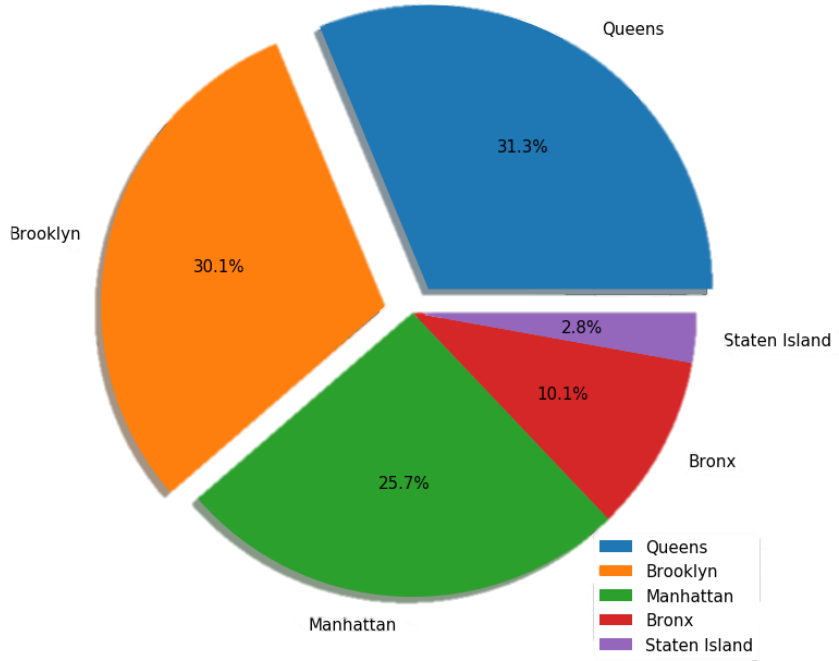

Fig.1 Distribution of 42450 Chinese Restaurant in each borough of NYC

Restaurant in NYC. So, it is worth to investigate both borough and their neighborhoods.

3.2: Exploring New York City data for K-means clustering

After two boroughs was found with highest Chinese restaurant number, the json data of New York City was imported to a data frame. The data frame helps explore the features of the neighborhoods. To start with, all neighborhoods in New York City were visualized on a map using Folium as shown in Fig.1.

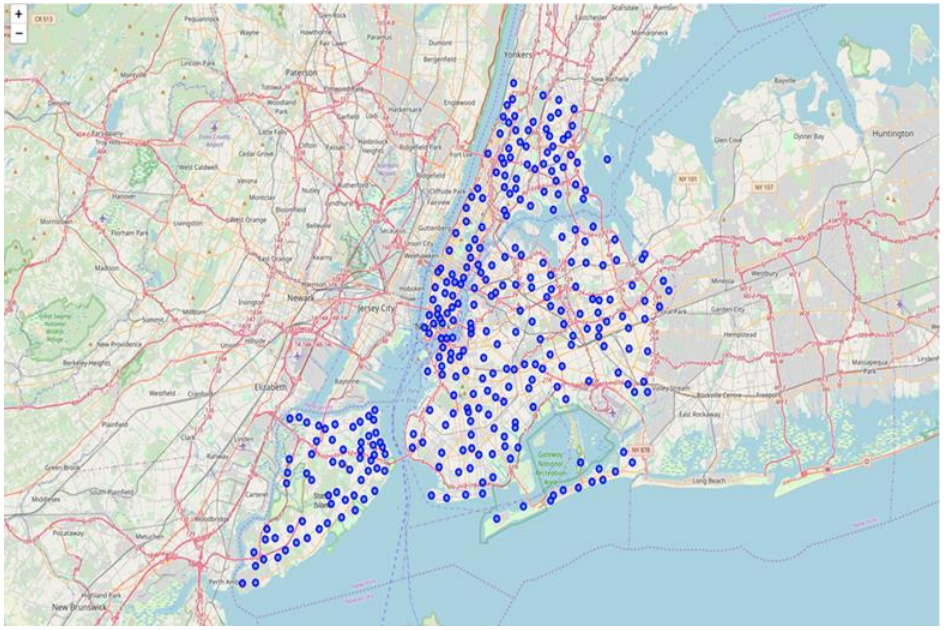

Fig.2 All neighborhoods in NYC

Fig.3 Neighborhoods of Queens and Brooklyn in NYC

Because we only need analysis Queens and Brooklyn area, all other borough was then removed as shown in Fig.3 To acquire the detailed information of venues in these neighborhoods, the Foursquare API was imported to create a data frame contains borough, neighborhoods, and coordinate information. By merge this neighborhood information with nearby venue data information, we then have complete data frame to find top 10 venues in these neighborhoods. One-hot encoding was the applied to this data frame to help us find top 10 most common venue in each neighborhood. K-means clustering then employed with K=5 to cluster all venues in these neighborhoods of Queens and Brooklyn. By observe cluster data frame, we can then find out the signature of each cluster:

- Cluster 0 (red) has high frequencies for restaurant category. This appear to be residence dining area
- Cluster 1 (purple) has low frequencies for all venue categories. This appears to be small affluent neighborhood. This area geographically close to beach, people likely to come here for outdoor activity. Drinks or fast food would be more popular here.
- Cluster 2 (blue) has consistently high frequencies for all venue categories. This is the most diversely developed part of city, most likely to be commercial district.
- Cluster 3 (green) has low frequencies. These appear to be bodega and farm area.

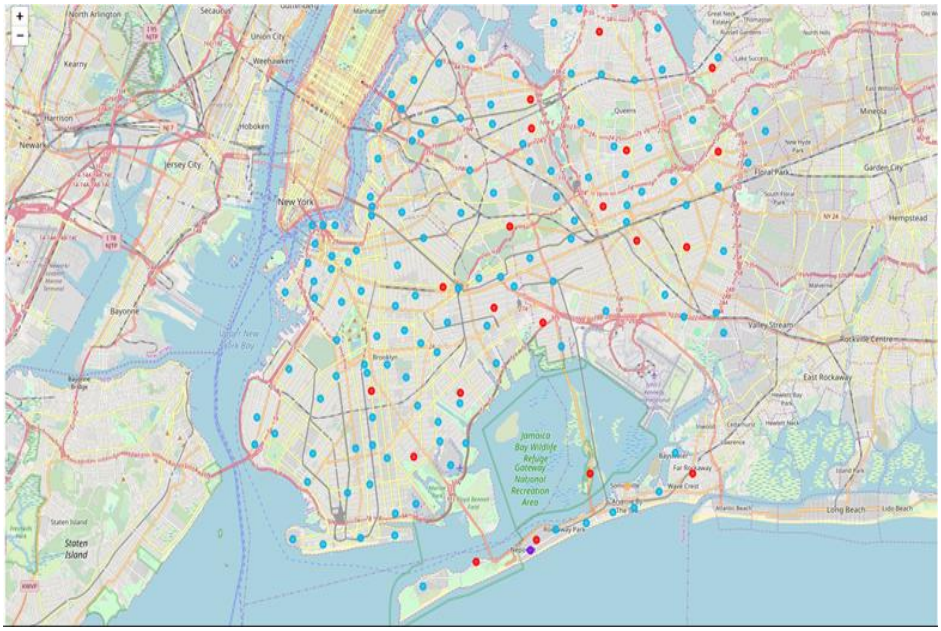

Fig.4 Neighborhoods Cluster in Queens and Brooklyn

- Cluster 4 (orange) has low frequencies but with more park and services. This is park and outdoor district

From Fig.4 we know most of the restaurants could be found in cluster 0 and cluster 2. Although cluster 2 neighborhoods have more Chinese restaurant in numbers, however, the ratio of Chinese restaurant is relatively lower. The ratio of venues appeared in top 10 category are equally distributed, which proves this is the most diversely developed part of city, most likely to be commercial shopping distinct. On the other hand, cluster 0 neighborhoods have most restaurant or bar venue in this area, it seems appears to be prosperous residential area, open a new Restaurant would be a good idea.

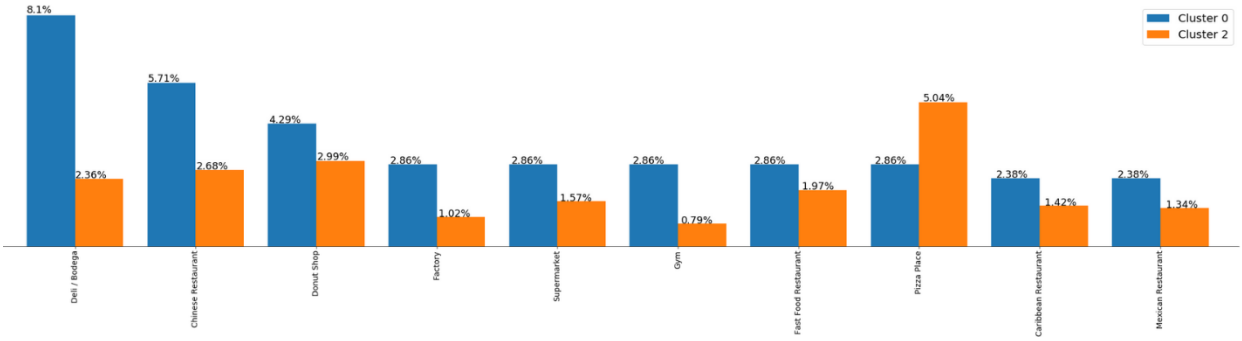

Fig.5 The ratio of each venues appeared in top 10 common venue

4.Result

From Fig.5 we know cluster 0 neighborhoods are dinning areas, however, there are more Chinese restaurant as most common venue in cluster 2 as shown in Fig.6. So, these 10 neighborhoods should be all satisfy new Chinese restaurant requirement. (Table 1)

	Borough	Neighborhood
0	Brooklyn	Bensonhurst
1	Brooklyn	Bath Beach
2	Queens	Corona
3	Queens	Kew Gardens
4	Queens	Beechhurst
5	Queens	Queensboro Hill
6	Queens	Ravenswood
7	Queens	Lindenwood
8	Brooklyn	New Lots
9	Brooklyn	Mill Basin

Table.1 Neighborhoods with Chinese Restaurant as most common venue

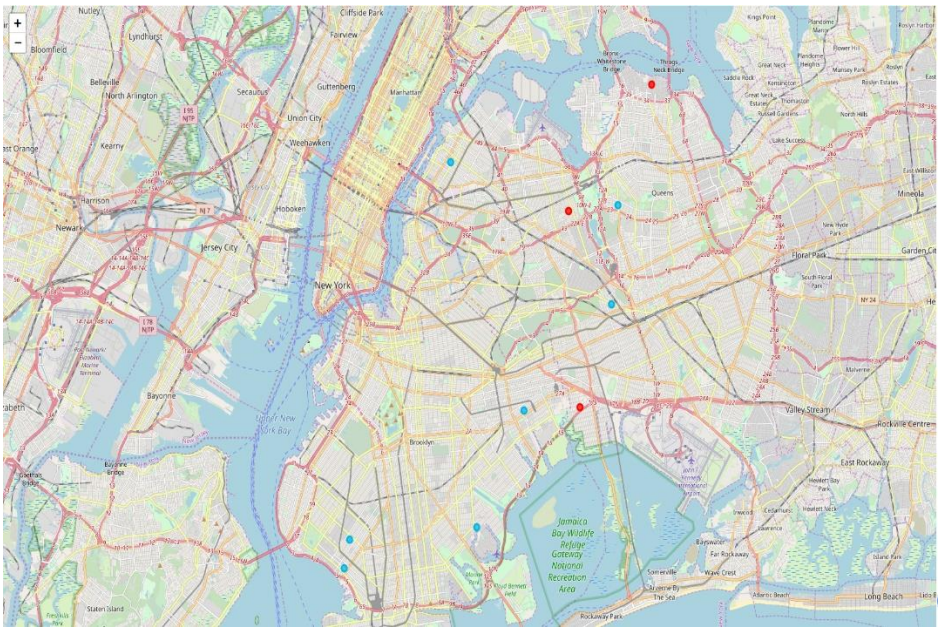

Fig.6 Neighborhoods with Chinese Restaurant as most common venue

5. Conclusion

Although the data analysis provides us result that these 10 neighborhoods are benefit Chinese Restaurant business. But there are many other factors didn't involve. In the additional data should be considered to improve accuracy, such as transportation data like accessibility, parking lot space, traffic condition, and population distribution and food taste.