

Capstone Project - The Battle of the Neighborhoods

Lei Guo
April 29, 2020

1. Introduction

1.1 Background

New York City is experiencing a Chinese food renaissance. Never before have the city's offerings been so diverse; not only are multiple regions represented, but price points range, too. There are fast-casual spots like Xi'an Famous Foods, and there's ambitious pan-regional fine dining like Atlas Kitchen. And where restaurateurs once needed to cater to Western tastes, many of today's New York Chinese restaurants don't have to so in order to survive.

1.2 Problem

In this project we will try to find an optimal location for a restaurant. Since there are lots of restaurants in New York we will try to detect the best location that are not already crowded with Chinese restaurants. Moreover, we are also particularly interested in areas with less competitor Chinese restaurants with good market reputation.

We will use our data science powers to generate a few most promising neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

1.3 Interest

Specifically, this report will be targeted to stakeholders interested in opening a Chinese restaurant in New York city, USA.

2. Data acquisition and cleaning

2.1 Data sources

Based on definition of our problem, factors that will influence our decision are:

- number of existing Chinese restaurants in the neighborhood.
- number of existing Chinese restaurants in the neighborhood with good market reputation (likes, tips etc.).

Following data sources will be needed to extract the required information:

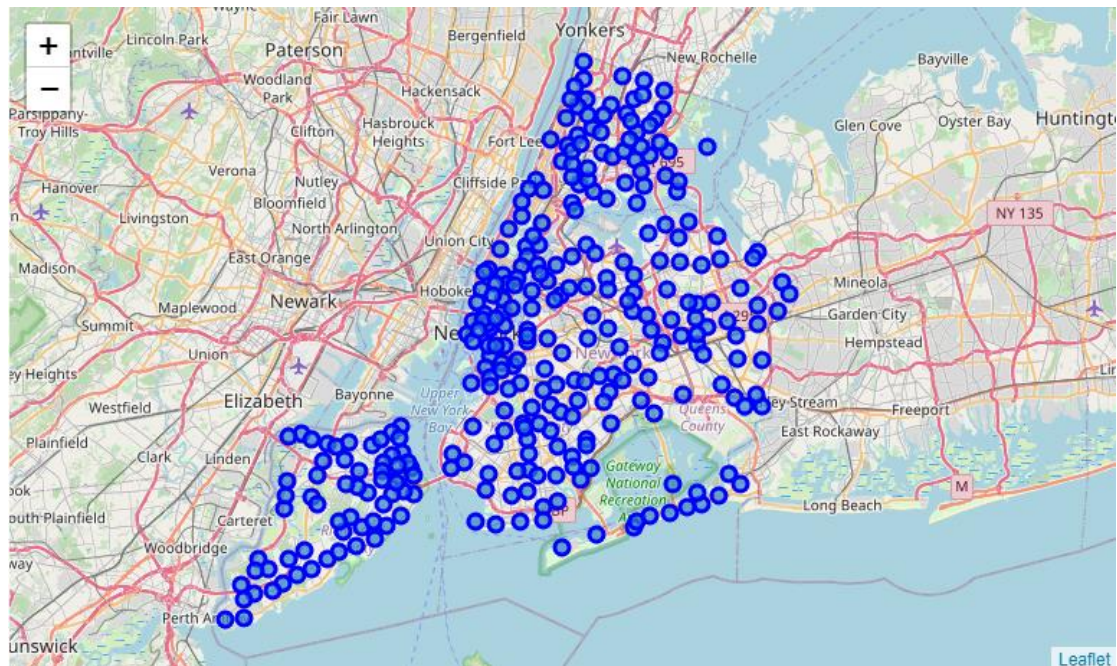
- New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.
Data source: https://cocl.us/new_york_dataset
- Chinese restaurants in each neighborhood of New York city.

Data source : Fousquare API

2.2 Data cleaning

New York has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

Create a map of New York with neighborhoods superimposed on top.



Next, we are going to start utilizing the Foursquare API to explore the neighborhoods and segment them. The details are in the below part.

3. Methodology

In first step we have collected the required data: location, id, name, category etc. of every Chinese restaurant within each neighborhood according to Foursquare.

Second step in our analysis will be calculation and exploration of 'Chinese restaurant quantity' across different neighborhoods of New York - we will use bar chart to identify a few promising neighborhoods with low number of Chinese restaurants(the first requirment of business) and focus our attention on those areas.

In third and final step we will focus on the promising areas and within those create clusters of locations that meet the second requirements established in discussion with stakeholders: we will take into consideration locations with less competitor Chinese restaurants with good market reputation. We will present map of all such locations using

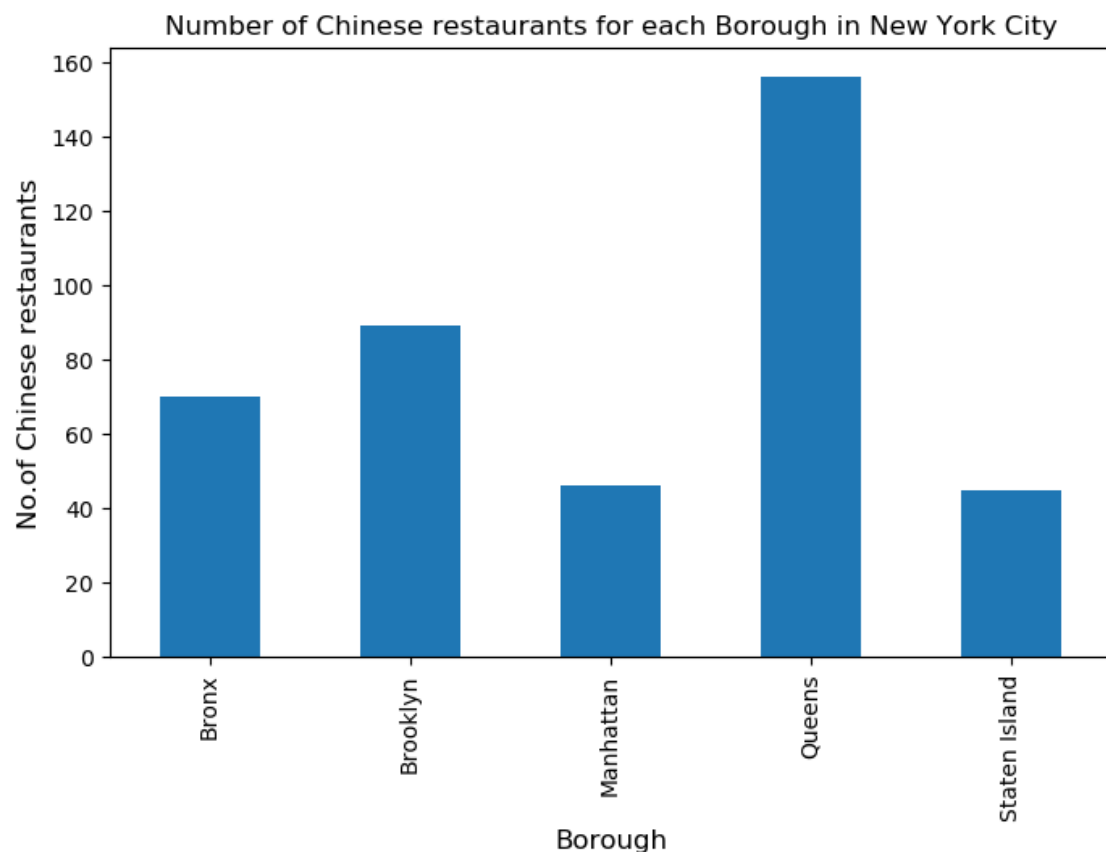
k-means clustering of those locations to identify general neighborhoods which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

4. Analysis

First let's count the number of Chinese restaurants in every neighborhood. Because there are many neighborhoods with only one Chinese restaurants, the promising areas could be a little bit confusing to choose from. Therefore, let's zoom out to the borough level and have a try.

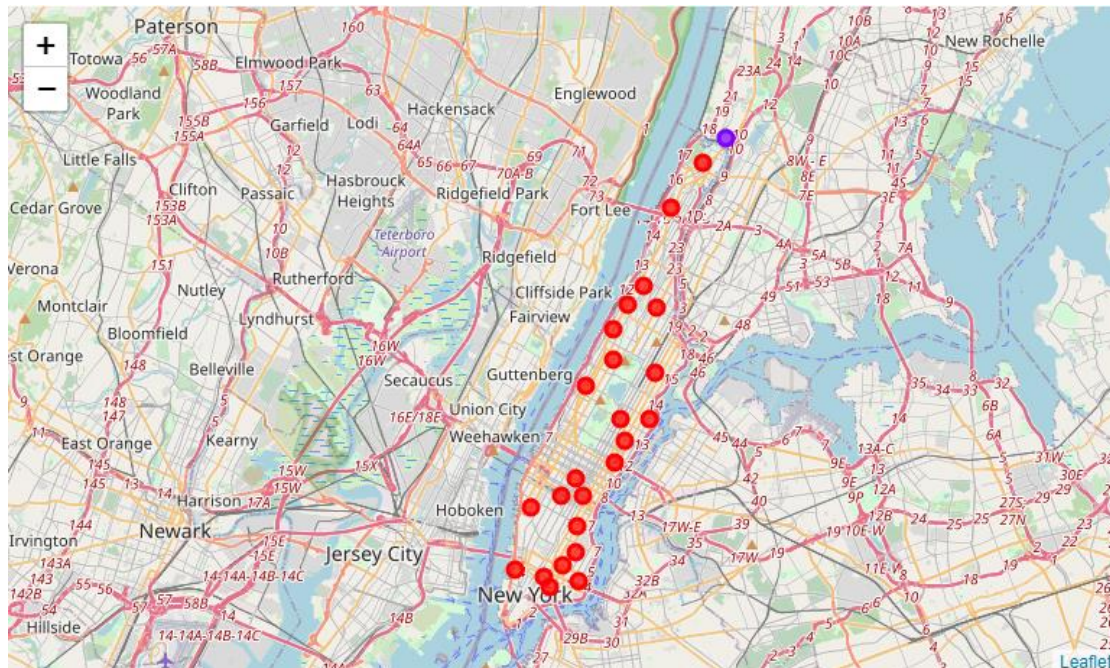
	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue ID	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bronx	70	70	70	70	70	70	70	70
1	Brooklyn	89	89	89	89	89	89	89	89
2	Manhattan	46	46	46	46	46	46	46	46
3	Queens	156	156	156	156	156	156	156	156
4	Staten Island	45	45	45	45	45	45	45	45

Let's create a bar chart to visualize the density of the Chinese restaurants in each borough.



In the second step, as we can see from the above bar chart that Manhattan and Staten Island have the least number of Chinese restaurants, we choose Manhattan as the promising area because it is the center of New York city.

We will get the ranking of each Chinese restaurant in Manhattan, then use k-means algorithm to cluster the neighborhood into 2 clusters.



5. Results and Discussion

Our analysis shows that although there is a great number of Chinese restaurants in New York city, there are boroughs of low Chinese restaurant density which are Manhattan and Staten Island. We never expect Manhattan has the least Chinese restaurants compared to other boroughs. Highest concentration of restaurants was detected in Bronx, Brooklyn and Queens, so we focused our attention to areas Manhattan and Staten Island. Furthermore, our attention was focused on Manhattan which offers a combination of popularity among tourists, closeness to city center, strong socio-economic dynamics and a number of pockets of low Chinese restaurant density.

After directing our attention to this more narrow area of interest, we first created a list of popularity ranking of Chinese restaurants using Foursquare data of likes, rating and tips; then we can tell the locations with good market reputation should not be considered as the promising areas for new Chinese restaurants location, because of the fierce competition.

Those neighborhoods in Manhattan were then clustered to create zones shown on the map above. The red circles indicate areas with potential market for Chinese restaurants,

the purple circles indicate areas with fierce market competition for Chinese restaurants. **Overall, we recommend the red circle areas on the map in Manhattan to open a new Chinese restaurant.**

The results may not be concise because of some limitations, one is the ranking depends purely on basis of rating of restaurants, the accuracy of data depends purely on the data provided by FourSquare. Another is the parameter set when we obtain the data from FourSquare, for example, the limitation set to 1000 and radius set to 1000 meters could be unreliable. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

6. Conclusion

Purpose of this project was to identify New York areas with low number of Chinese restaurants in order to aid stakeholders in narrowing down the search for optimal location for a new Chinese restaurant. By calculating Chinese restaurant number distribution from Foursquare data we have first identified general boroughs that justify further analysis (Manhattan), and then generated collection of locations which satisfy another requirement regarding existing nearby Chinese restaurants with good market reputation which are strong competitors. Clustering of those locations was then performed in order to create major zones of interest to be used as starting points for final exploration by stakeholders.

Final decision on optimal Chinese restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location, levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.