Exploring New York Neighborhoods - to open a Chinese restaurant

Qingling Shen

January 14, 2020

1 Introduction

1.1 Background

New York is the most populous city in the United States. It is the largest metropolitan area in the world by urban landmass and one of the world's most populous megacities. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world. New York City is home to 3.2 million residents born outside the United States. Also, New York is home to the largest ethnic Chinese population outside of Asia, with multiple distinct Chinatowns across the city. **(**Taken from Wiki**)** Thus, it is one of the best places to start a Chinese restaurant.

1.2 Problem

In this project we will go through process to make a decision whether it is a good idea to open a Chinese restaurant. We analyze the neighborhoods in NY to identify the most profitable area since the success of the restaurant depends on the number of persons who might be shopping. Since we already know that it is a good idea to start the supermarket here, but we just need to make sure where we can place it, so it yields more profit to the owner.

1.3 Target audience

Who would be more interested in this project? And what type of customers or group of people would benefit from this?

1. Businessman who want to invest or open a Chinese restaurant in New York. The analysis will be a comprehensive guide for opening or expanding Chinese restaurants.

2. Chinese gathering area with demand for Chinese restaurants.

3. Freelancers who like to have their own Chinese restaurant as a sideline. The analysis will give an idea of how good it is to open a Chinese restaurant, and what are the pros and cons of the business.

I hope to use exploratory data analysis and other statistical and machine learning techniques to analyze the situation around New York for opening a Chinese restaurant to get all the necessary data, perform some operations on it and draw corresponding conclusions

2 Data acquisition and cleaning

2.1 Data source

We will use the data mentioned in the last week's lab study. Neighborhood 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. This dataset exists for free on the web. Here is the link to the dataset: (https://geo.nyu.edu/catalog/nyu_2451_34572).

From Foursquare API (https://developer.foursquare.com/docs), I retrieved the following for each venue: Name: The name of the venue; Category: The category type as defined by the API; Latitude: The latitude value of the venue; Longitude: The longitude value of the venue.

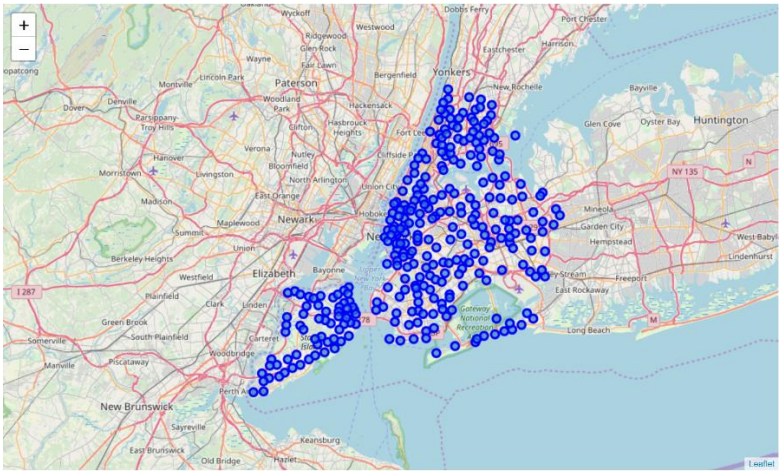
2.2 Data cleaning

We will get location data using Foursquare. Foursquare API is very useful online application used my many developers & other applications like Uber etc. In this project I have used it to retrieve information about the places present in the neighborhoods of NY. The API returns a JSON file and we need to turn that into a data-frame. For example, we could choose 100 popular spots for each neighborhood within a radius of 1km. Then, we could try to find and analyze relationship between neighborhood and Chinese restaurant.

3 Exploratory Data Analysis

3.1 Folium Library and Leaflet Map

Folium is a python library, and I’m using it to draw an interactive leaflet map using coordinate data. The map of New York with neighborhoods superimposed on top is shown below:

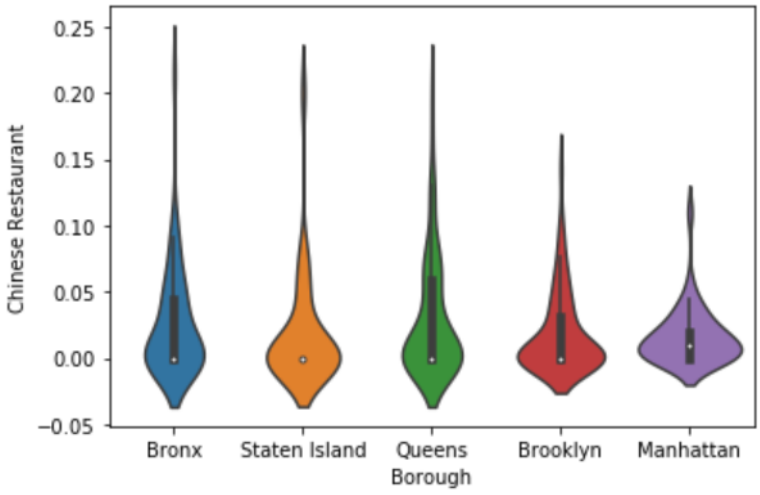


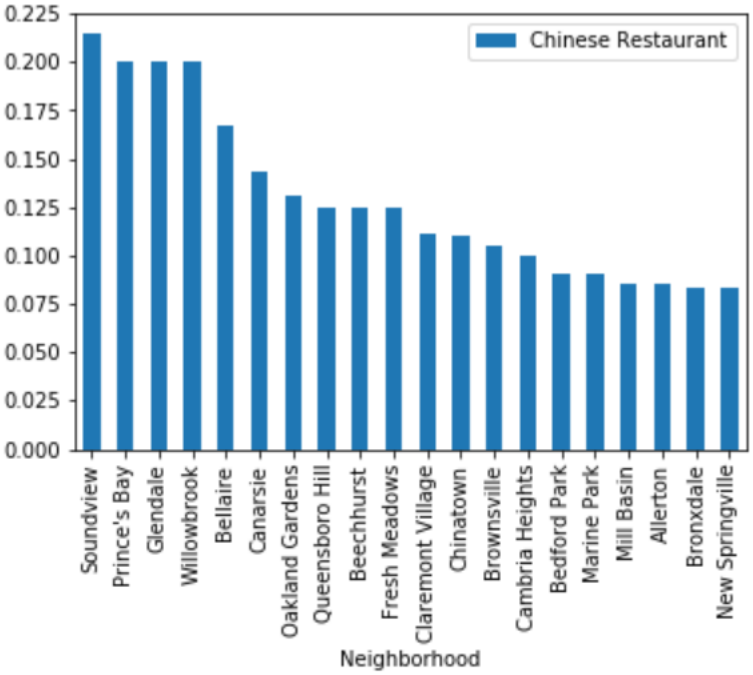
3.2 Relationship between neighborhood and Chinese Restaurant

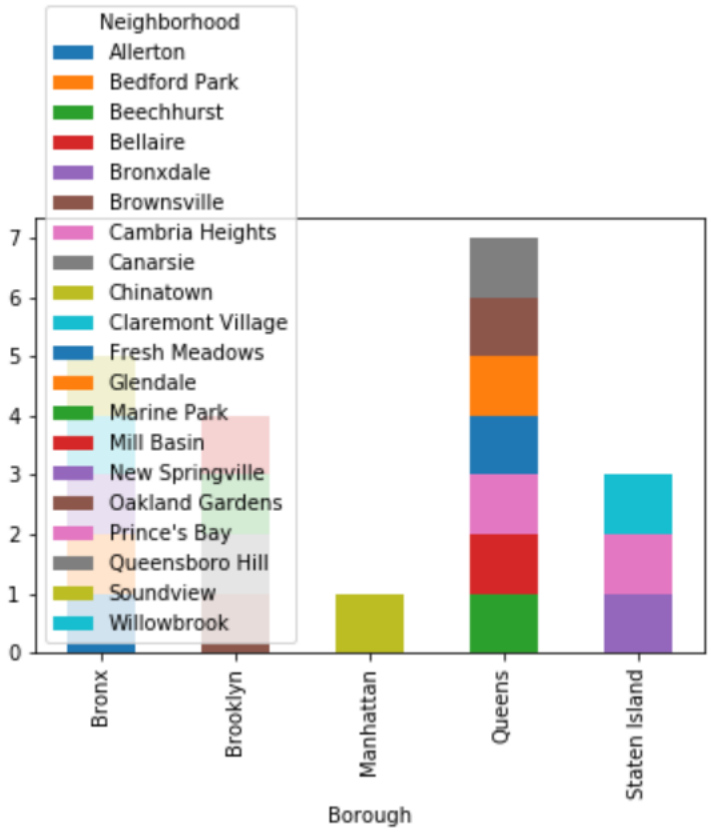
First, we will extract the Neighborhood and Chinese Restaurant column from the above data frame for further analysis:

	Neighborhood	Borough	Chinese Restaurant
252	Soundview	Bronx	0.214286
218	Prince's Bay	Staten Island	0.200000
114	Glendale	Queens	0.200000
296	Willowbrook	Staten Island	0.200000
20	Bellaire	Queens	0.166667
43	Canarsie	Brooklyn	0.142857
200	Oakland Gardens	Queens	0.130435
223	Queensboro Hill	Queens	0.125000
19	Beechhurst	Queens	0.125000
109	Fresh Meadows	Queens	0.125000

Also, we draw some plots to show the relationship more clearly.





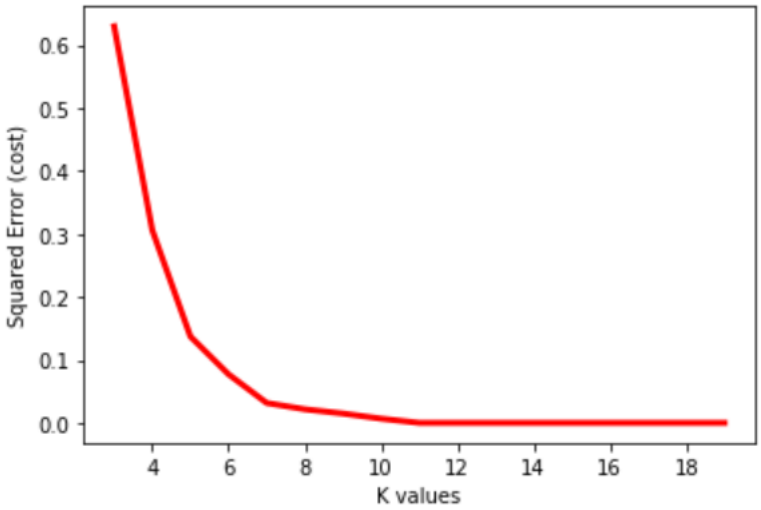


From the plots above we can find that Queens has the largest number of Chinese restaurants. Manhattan and Brooklyn have relatively few Chinese restaurants. Due to the big size of the whole dataset, we only pay attention to Bronx in a follow-up study.

4 Predictive modelling

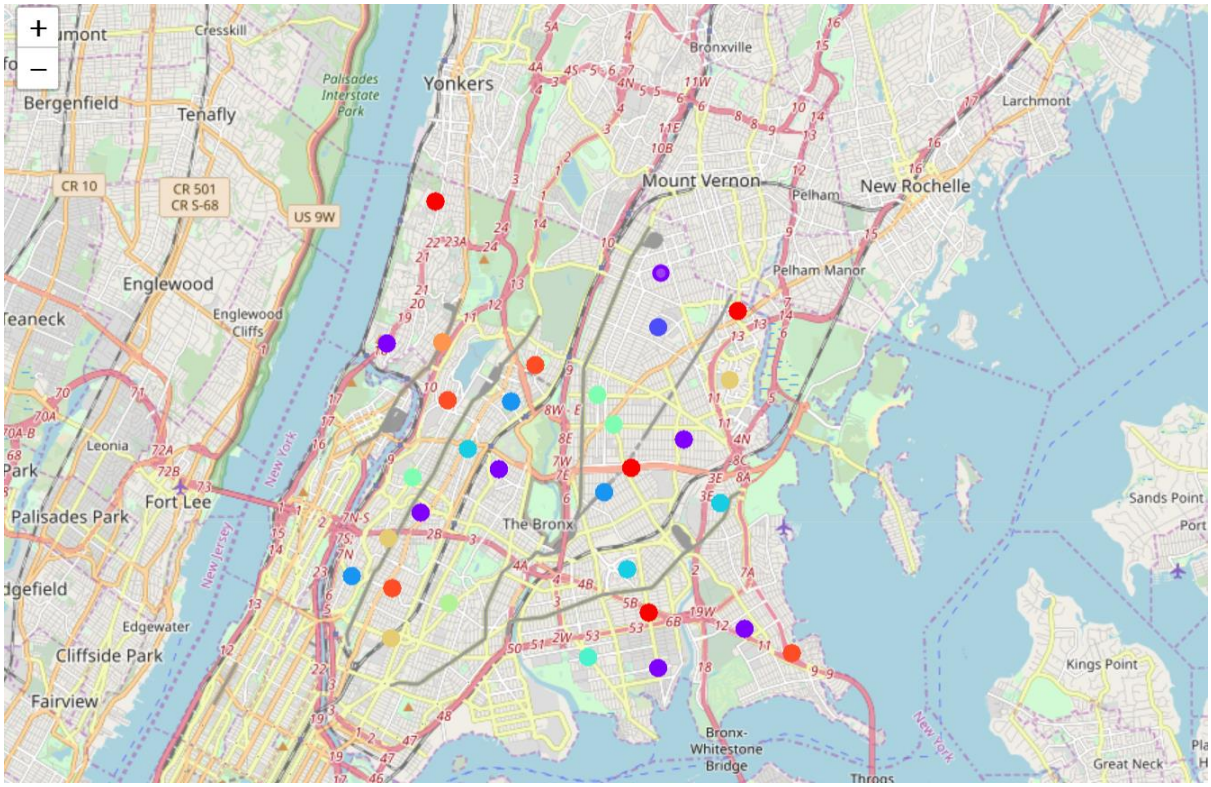
4.1 Clustering Neighborhoods of Queens:

First step in K-means clustering is to identify best K value meaning the number of clusters in a given dataset. To do so we are going to use the elbow method on the Bronx borough dataset with Chinese restaurant percentage.



It looks like K = 11 is the best value.

The Folium map for the clusters of different neighborhoods is shown below: (we just extract the first 800 points with sorted dataset)



4.2 Examine the clusters

Cluster 0 contains all the neighborhoods which has least number of Chinese restaurants. It is shown in red color in the map:

```
NY_merged.loc[NY_merged['Cluster Labels'] == 0, NY_merged.columns[[1] + list(range(5, NY_merged.shape[1]))]]
```

	borough	Venue Latitude	Venue Longitude	Venue Category	Cluster Labels	Borough	Chinese Restaurant
26	Bronx	40.885656	-73.829197	Caribbean Restaurant	0.0	Bronx	0.052632
27	Bronx	40.886332	-73.827616	Diner	0.0	Bronx	0.052632
28	Bronx	40.888628	-73.831260	Pizza Place	0.0	Bronx	0.052632
29	Bronx	40.889318	-73.831453	Seafood Restaurant	0.0	Bronx	0.052632
30	Bronx	40.885384	-73.828099	Donut Shop	0.0	Bronx	0.052632
31	Bronx	40.888235	-73.831282	Deli / Bodega	0.0	Bronx	0.052632
32	Bronx	40.888488	-73.831083	Caribbean Restaurant	0.0	Bronx	0.052632
33	Bronx	40.886020	-73.823207	Bowling Alley	0.0	Bronx	0.052632
34	Bronx	40.888249	-73.831661	Caribbean Restaurant	0.0	Bronx	0.052632
35	Bronx	40.887969	-73.831413	Diner	0.0	Bronx	0.052632
36	Bronx	40.885636	-73.826840	Fast Food Restaurant	0.0	Bronx	0.052632
37	Bronx	40.885940	-73.827870	Bus Stop	0.0	Bronx	0.052632
38	Bronx	40.889147	-73.829297	Food & Drink Shop	0.0	Bronx	0.052632

Cluster 14 contains all the neighborhoods which is densely populated with Chinese restaurants. It is shown in orange color in the map:

```
NY_merged.loc[NY_merged['Cluster Labels'] == 10, NY_merged.columns[[1] + list(range(5, NY_merged.shape[1]))]]
```

	borough	Venue Latitude	Venue Longitude	Venue Category	Cluster Labels	Borough	Chinese Restaurant
153	Bronx	40.875269	-73.879563	Pizza Place	10.0	Bronx	0.034483
154	Bronx	40.877033	-73.877331	Park	10.0	Bronx	0.034483
155	Bronx	40.874933	-73.879404	Coffee Shop	10.0	Bronx	0.034483
156	Bronx	40.880766	-73.877808	Pizza Place	10.0	Bronx	0.034483
157	Bronx	40.878234	-73.883164	Park	10.0	Bronx	0.034483
158	Bronx	40.880200	-73.883434	Mexican Restaurant	10.0	Bronx	0.034483
159	Bronx	40.881665	-73.879484	Deli / Bodega	10.0	Bronx	0.034483
160	Bronx	40.874727	-73.879660	Bank	10.0	Bronx	0.034483
161	Bronx	40.874499	-73.879515	Pharmacy	10.0	Bronx	0.034483
162	Bronx	40.881566	-73.879299	Restaurant	10.0	Bronx	0.034483
163	Bronx	40.875392	-73.879734	Bank	10.0	Bronx	0.034483
164	Bronx	40.879039	-73.874222	Pharmacy	10.0	Bronx	0.034483
165	Bronx	40.872818	-73.878312	Sandwich Place	10.0	Bronx	0.034483
166	Bronx	40.881672	-73.878996	Caribbean Restaurant	10.0	Bronx	0.034483
167	Bronx	40.872880	-73.878151	Pizza Place	10.0	Bronx	0.034483
168	Bronx	40.873496	-73.878825	Supermarket	10.0	Bronx	0.034483
169	Bronx	40.880516	-73.883233	Spanish Restaurant	10.0	Bronx	0.034483
170	Bronx	40.872894	-73.878677	Fast Food Restaurant	10.0	Bronx	0.034483

5 Results and discussion

5.1 Results

Among these 5 boroughs and hundreds of neighborhoods, we found that only Queens Bronx and Staten Island have relatively high numbers of Chinese restaurants. With clusters examining and violin plots, Queens seems to be packed with Indian restaurants. Therefore, it's best to exclude this borough and just consider Bronx as the location for the new restaurant. After careful consideration, it is a good idea to open a new Chinese restaurant in Bronx, because the Chinese restaurant there is relatively moderate and the competition is lower than Queens. It has more customer possibilities.

5.2 Discussion

According to analysis, Bronx will provide relatively large passenger flow and relatively little competition for new Chinese restaurants that are about to open. Therefore, there is no doubt that the area may be the ideal place to open a new restaurant. However, the disadvantage of this analysis is that the clustering based only on data obtained from the Foursquare API.

6 Future direction

In the last analysis, we divided Bronx borough into 11 clusters. And we can further analyze the density distribution of restaurants in the Bronx community. From the folium map we can easily find the neighborhood which is relatively suitable for opening a new Chinese restaurant.

Some disadvantages or areas for improvement show that this analysis can be further improved with the help of more data and different machine learning techniques. Similarly, we can use the project to analyze any situation, such as opening other gourmet restaurants or opening a new gym and so on.