

The Battle of the Neighborhoods Capstone Project

Searching for Chinese Food in New York City

Vince Spinnato

September, 2019

1. Introduction

1.1 Background

New York City (“NYC”) is the most densely populated major city in the U.S. with an estimated 8.4 million residents as of July 2018. NYC’s population is approximately 43% of New York State’s population. NYC is often referred to as the five boroughs, and in turn, there are hundreds of distinct neighborhoods throughout the boroughs, many with a definable history and character to call their own. If the boroughs were each independent cities, four of the five boroughs (Brooklyn, Queens, Manhattan, and the Bronx) would be among the ten most populous cities in the United States.

Throughout its history, New York City has been a major point of entry for immigrants. The term “melting pot” was coined to describe densely populated immigrant neighborhoods on the Lower East Side of Manhattan. Approximately 37% of the city’s population is foreign born where no single country or region of origin dominates. The ten largest sources of foreign-born individuals in the city as of 2011 were the Dominican Republic, China, Mexico, Guyana, Jamaica, Ecuador, Haiti, India, Russia, and Trinidad and Tobago. New contains the highest total Asian population of any U.S. proper, exceeding more than one million. The Chinese population constitutes the fastest-growing nationality in New York State.

Given its diverse culture, NYC’s food scene includes an array of international cuisines influenced by its immigrant history. They range from Central and Eastern Middle Eastern, Indian, Italian, and Chinese to name just a few.

1.2 Objective

This project will explore the many neighborhoods of NYC to visualize and identify the ones with the best Chinese restaurants.

1.3 Interest

There are several people who would be interested in the results of this project. The obvious ones are folks who just enjoy good Chinese food. Others might be, Asian American's visiting or looking to relocate to NYC who are interested in knowing which neighborhoods have the most and/or best Chinese restaurants. Another interested group could be Restaurateur's or investors who are looking to invest in or open a restaurant.

2. Data Acquisition and Processing Approach

2.1 Data sources

This project will require the following data:

2.1.1 New York City data that contains a list of Boroughs and Neighborhoods along with their latitude and longitude coordinates

- a. Data Source: https://cocl.us/new_york_dataset.com
- b. Description: This data set will be used to explore the various boroughs and neighborhoods of New York City

2.1.2 List of Chinese restaurants in New York City neighborhoods

- a. Data Source: Foursquare API
- b. Description: The Foursquare API will be used to identify venues in each neighborhood and will filter the venues for Chinese restaurants to identify location and ratings.

2.1.3 GeoSpace Data

- a. Data Source: <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm>

b. Description: Use the geo space data to identify New York Borough boundaries to visualize locations on a Map

2.2 Data Processing Approach

We start the data processing by collecting the necessary New York City data sets from the internet as outlined in section 2.1.1(a) above and examine the data to ensure it has the information needed for the analysis which includes the latitude and longitude coordinates for the boroughs and neighborhoods in NYC. Then we will access the Foursquare API to find all venues for each neighborhood in NYC and filter that data for Chinese restaurants. Once we have identified the restaurants, we will use the Foursquare API to acquire the ratings for each Chinese restaurant by utilizing the venue details. Then we will filter and sort the ratings data to identify the restaurants with the highest ratings. We will then merge that data with the New York City data set we acquired for the internet to match the venue details with the geospatial coordinates for each respective restaurant so that we can visualize the rankings on a map of New York City by using Python's Folium library.

3. Data Analysis

3.1 Import the necessary libraries for Python

Import pandas and numpy for handling and scraping the data.

Request module for using Foursquare API

Geopy to get coordinates of the City of New York

Folium to visualize the results on a map of NYC

Matplotlib to plot the data in charts

3.2 Define functions to extract and process our data

3.2.1 Define a function to get the latitude and longitude geocodes of a given location using geopy. The resulting coordinates for NYC were Latitude: 40.693943 and Longitude: -73.985880.

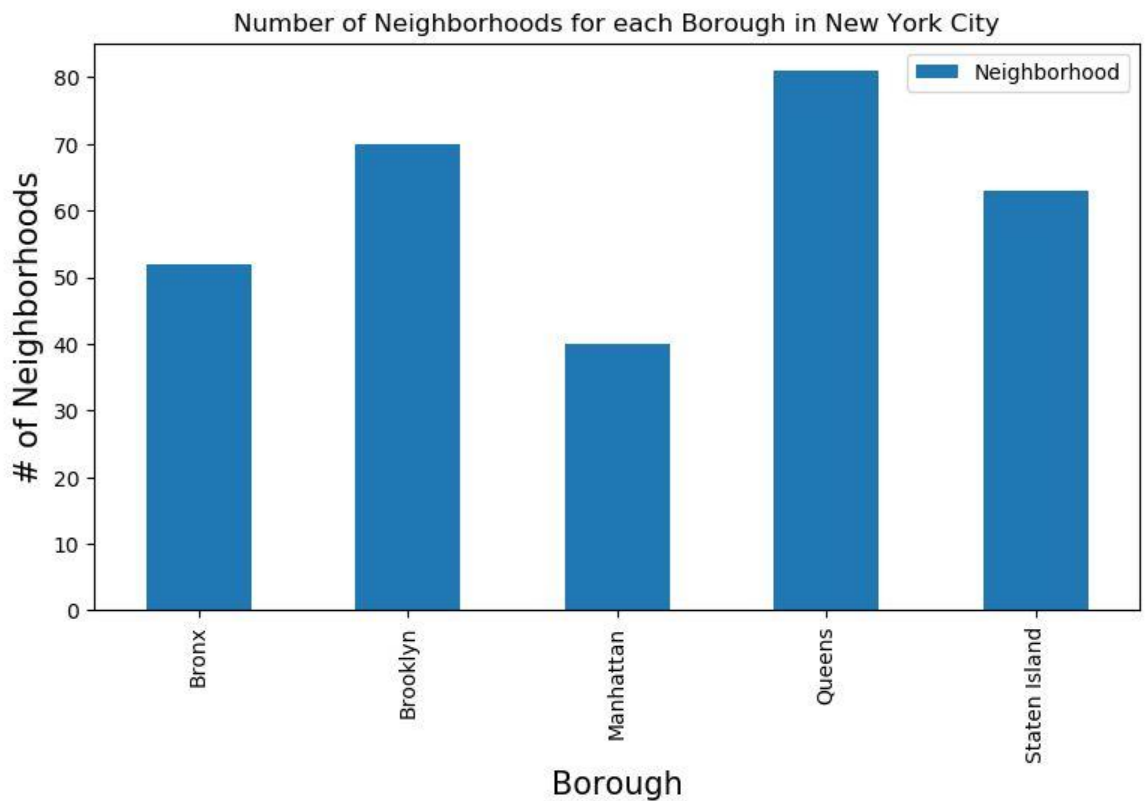
3.2.2 Define a function to interact with the Foursquare API to get a 100 venues within a radius of 1000 meters for a given latitude and longitude. This function returns the venue id, venue name and category

3.2.3 Define a function to get the venue details such as Rating for a given venue id. This info will be used to rank our results.

3.2.4 Define a function to get the New York City data such as Boroughs and Neighborhoods along with their respective latitude and longitude coordinates. This data will be merged with the venue data to match venue ids with their respective geo-coordinates to map the results.

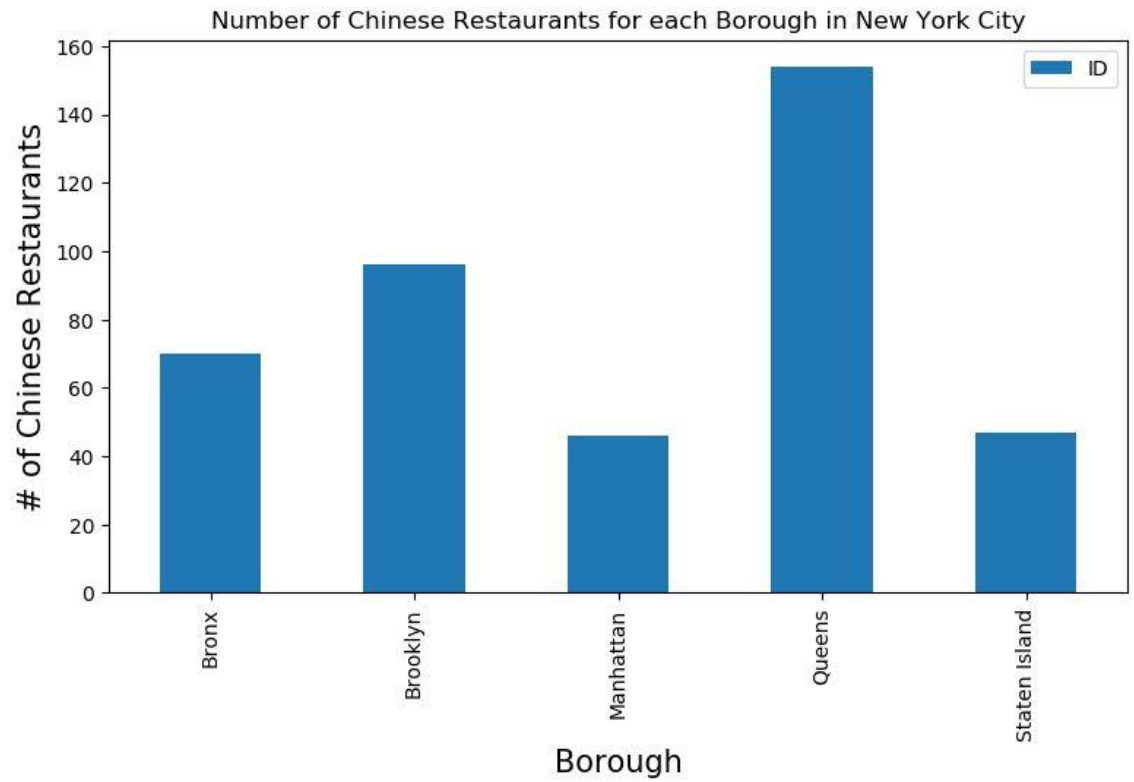
3.3 Results from processing the functions

The New York City data identifies that there are 306 different neighborhoods in NYC and the number of neighborhoods for each Borough in NYC is:



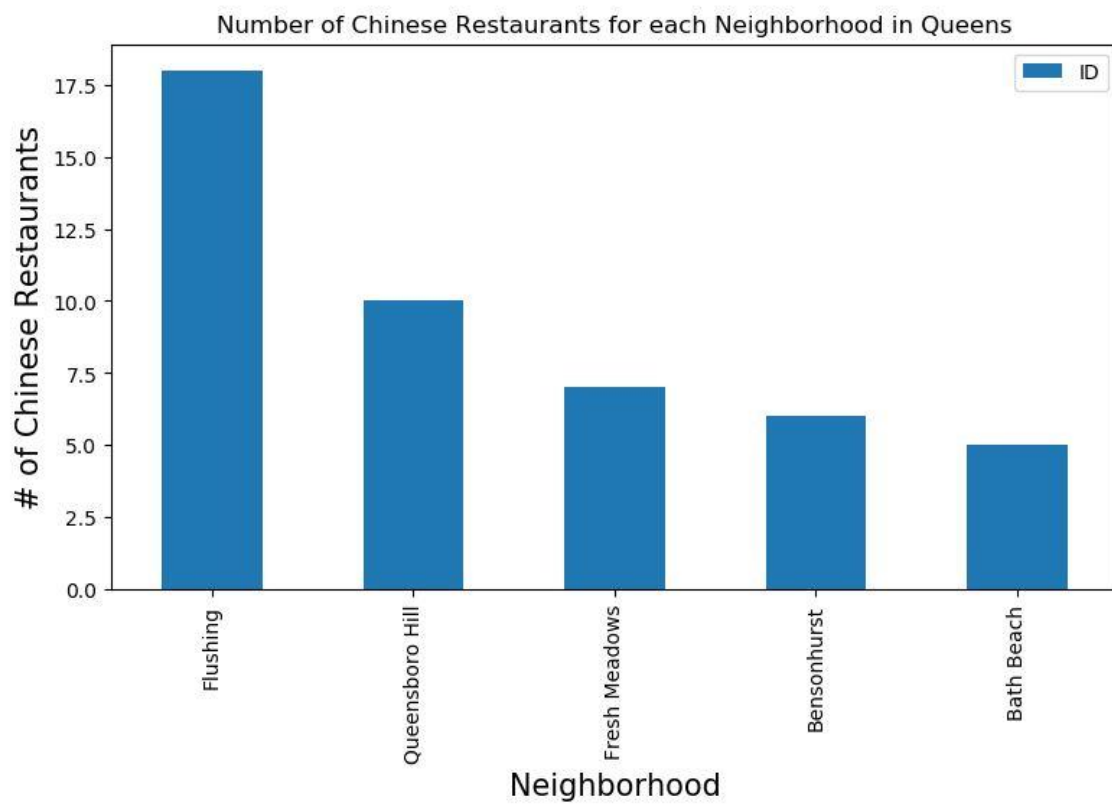
From the chart above, we see that Queens has the greatest number of neighborhoods.

From the data sets we will prepare a list of neighborhoods that contain Chinese restaurants. The result of this analysis identified 413 Chinese restaurants in NYC. We will graph this data to identify the number of restaurants by Borough.



From this chart we can see that Queens, NY has the largest number of Chinese restaurants.

Now we will process this data further to identify the neighborhood in Queens with the largest number of Chinese restaurants.



Flushing is the neighborhood in Queens with the largest number of Chinese restaurants (18).

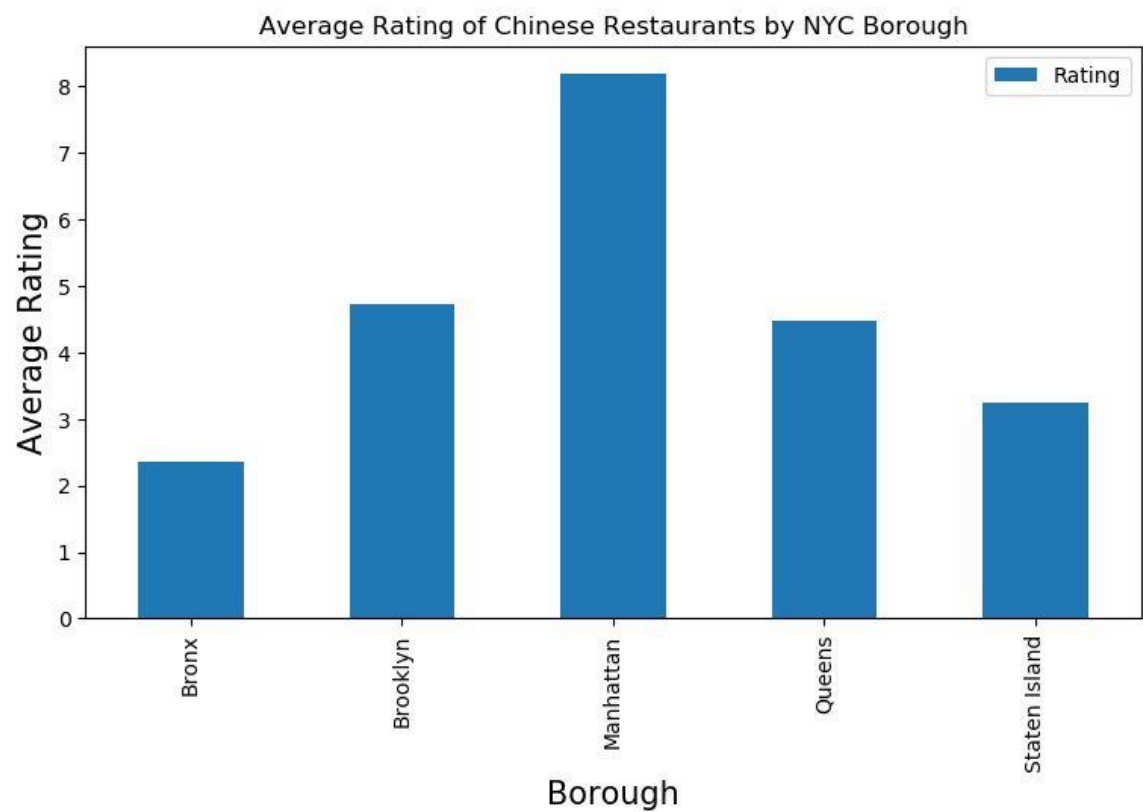
Now we will utilize the Foursquare API to retrieve the ranking for all 413 Chinese restaurants in NYC and scrape the data for different metrics. By sorting for maximum rating, we identified the highest rated Chinese restaurant in NYC is Buddakan located in the Chelsea neighborhood of Manhattan with a rating of 9.2. Now let's sort the data to identify the Neighborhoods with the highest average rating. The results from that analysis was:

Neighborhood	Average Rating
Flatiron	9.2
Chelsea	9.2
Long Island City	9.1
Lower East Side	9.0
East Village	8.9
West Village	8.9
Gramercy	8.9

Now let's identify the average ratings by Borough, the results were:

Borough	Average Rating
Manhattan	8.2
Brooklyn	4.7
Queens	4.5
Staten Island	3.2
Bronx	2.4

We used this data to create a chart of the average rating of Chinese restaurants by Borough:



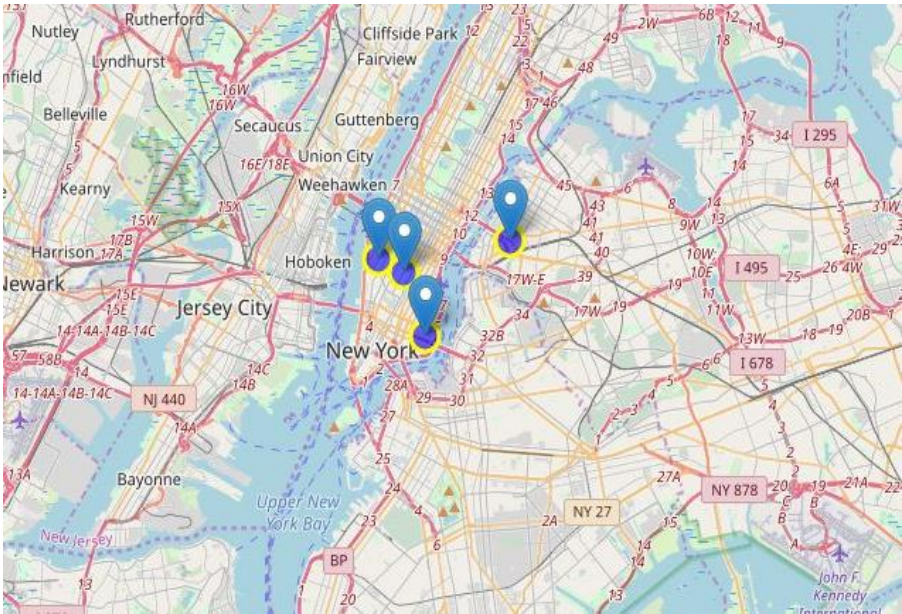
We processed the data further to sort the neighborhoods with an average rating equal to and greater than 9.0 and the results were:

Neighborhood	Average Rating
Chelsea	9.2
Flat Iron	9.2
Long Island City	9.1
Lower East Side	9.0

Then we merged this data with the New York City data set to get the latitudes and longitude coordinates to map the neighborhoods with the highest average rating and the results were as follows:

Borough	Neighborhood	Latitude	Longitude	Avg Rating
Manhattan	Chelsea	40.744035	-74.003116	9.2
Manhattan	Flatiron	40.739673	-73.990947	9.2
Queens	Long Island City	40.750217	-73.939202	9.1
Manhattan	Lower East Side	40.717807	-73.980890	9.0

Here is the resulting map:



4. Conclusions

4.1 What we learned from the data:

- a. New York City has **306** different Neighborhoods across the five Boroughs with **413** different Chinese restaurants.
- b. **Queens** is the Borough with the greatest number of Neighborhoods (80) and Chinese restaurants (154).
- c. **Flushing, Queens** is the Neighborhood with the most Chinese restaurants (18).
- d. The highest rated Chinese restaurant is **Buddakan** in **Chelsea, Manhattan** with a **9.2** rating.
- e. The Borough with the highest average rating is **Manhattan** with **8.2** and the Borough with the lowest rating is the Bronx with 2.4.

4.2 How to apply the results of the data.

- a. If you are in NYC and looking for good Chinese food, then you will want to seek out Buddakan in Chelsea, Manhattan.
- b. If you are moving to the NYC area and having easy access to multiple Chinese restaurants is an important criterion then you might consider Flushing, Queens as your future homestead.
- c. Restauranteurs may consider Queens as a potential location for their next Chinese restaurant due to its large neighborhood count or from a competitive standpoint, they may want to explore the Bronx due to its low average rating resulting in less competition.