

Capstone: Battle of the Neighborhoods - Chinese Restaurants

Iain Drew

1. Introduction

A small group of restaurateurs are looking to open a Chinese restaurant somewhere in New York City. New York City comprises 5 boroughs with Manhattan as its epicenter, a densely populated borough and one of the world's major commercial, financial and cultural centers. There are numerous restaurants of similar genre, and this group of investors would like to understand what areas/neighborhoods would be advantageous to opening and sustaining a successful business. This report explores which neighborhoods and boroughs of New York City have the most and best Chinese restaurants and I will attempt to answer the question "Where should the restaurateurs open a Chinese restaurant?"

1.1 Problem

- What is the best area for a small group of investors to open a Chinese restaurant in New York City?
- What boroughs and neighborhoods have the most and highest ranked Chinese restaurants?

2. Data

Data for this project will be structured into data frames with the geographic data as the starting point and adding Foursquare metadata (Names, ID, rankings etc.) to additional data frames later in the workflow. These features will be used to evaluate the counts of restaurants per borough and neighborhood as well as identify the highest ranked restaurants and their locations. Throughout the workflow visualizations will be used to better understand the data as well.

2.1 Data sources

- New York City data containing neighborhoods, boroughs, latitudes and longitudes is obtained from: https://cocl.us/new_york_dataset
- All data related to locations and quality of Chinese restaurants will be obtained from the Foursquare API.

- New York City boundary data is from: <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm>

2.2 Data Preparation

Initial geographic data on NYC will be gathered and structured into a data frame with features for Borough, Neighborhood, Latitude, and Longitude.

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Figure 1 Data frame of NYC geographic data

From this df we can get some preliminary information about how many neighborhoods per borough exist. Here we see Queens has the greatest number of neighborhoods and Manhattan has the least.

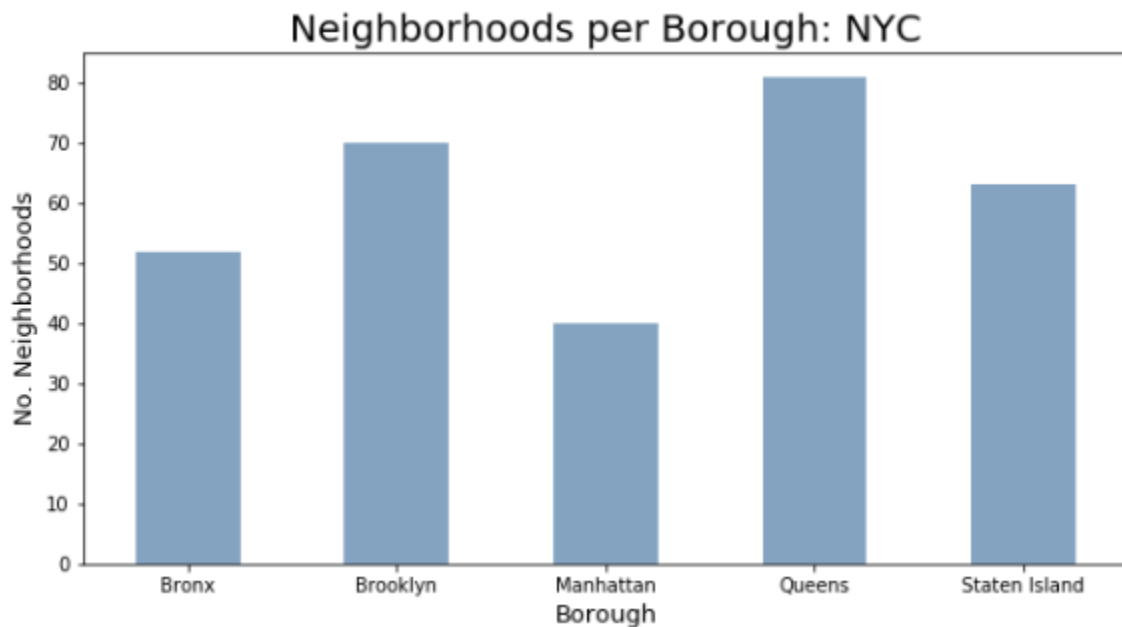


Figure 2 Bar plot of Neighborhoods per Borough in NYC

Next, a data frame was created to contain Borough, Neighborhood, Lat and Long as well as ID and Name (of Chinese restaurants) from Foursquare API to compare how many Chinese restaurants exist in each borough and neighborhood.

	Borough	Neighborhood	ID	Name
257	Staten Island	Prince's Bay	4bb7d282b35776b0b83dc801	Island Taste
258	Brooklyn	Madison	4be32bd7b02ec9b61fc34ec0	win hing
259	Bronx	Bronxdale	4c3f7f2eda3dc928b8f6c5b9	Peking Kitchen
260	Bronx	Allerton	4bc11181abf495219f7dc093	Li's Kitchen
261	Queens	Hammels	4c72ad2bad69b60cb31b84b9	301 Chinese Resturant

Figure 3 Data frame of Chinese restaurant Names, ID with geographic location data.

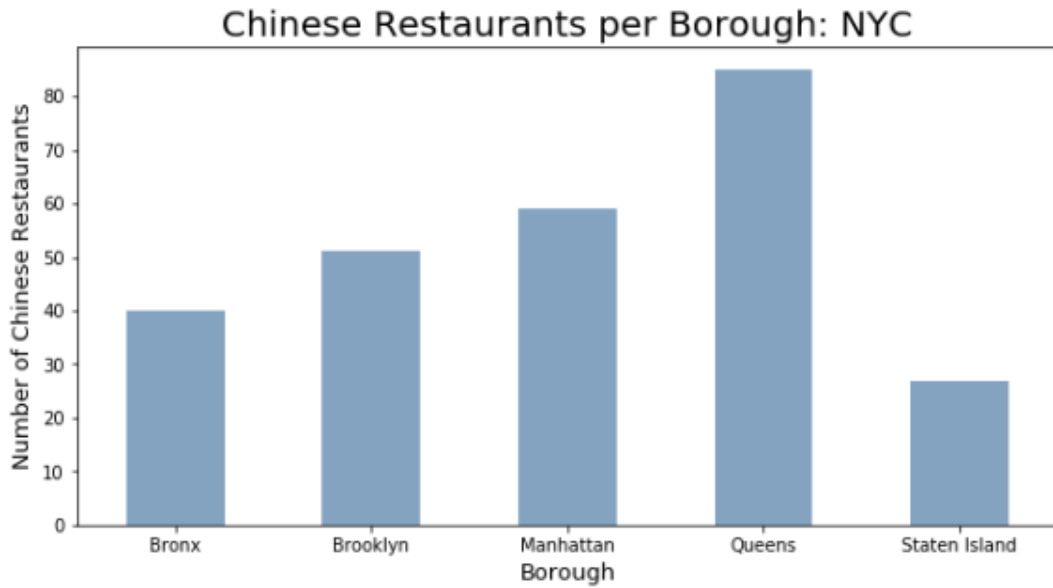


Figure 4 Bar plot of Chinese restaurants per Borough NYC

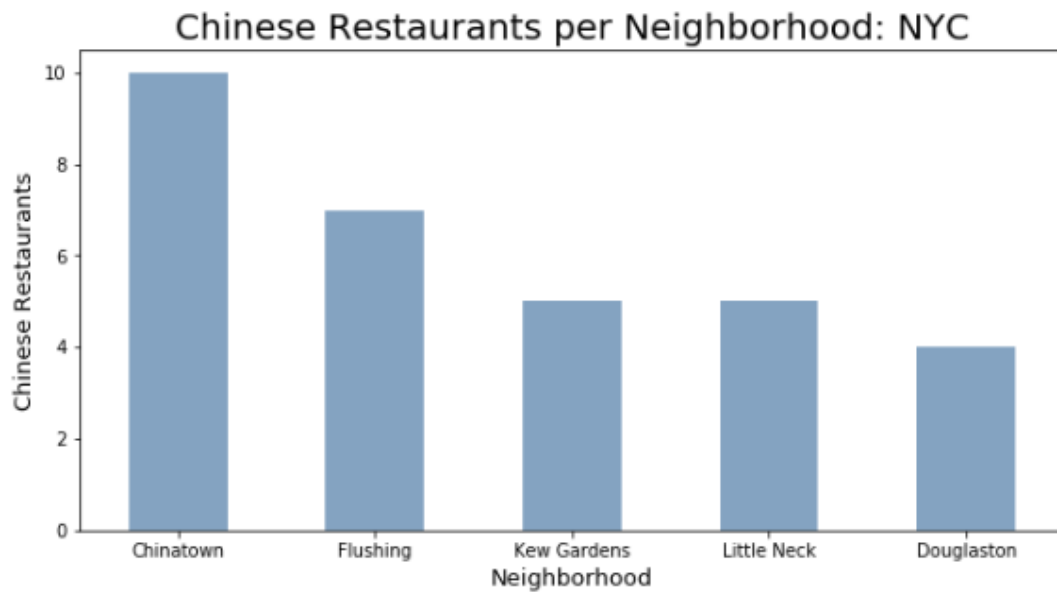


Figure 5 Chinese restaurants per Neighborhood NYC

Methodology

- Data will be collected from https://cocl.us/new_york_dataset, then cleaned and transformed into a dataframe.
- Foursquare data will be collected to locate all venues then filtered for Chinese restaurants. Metadata like ratings and tips by users will be counted and put into the dataframe.
- Data will get sorted by ranking.
- Data will be visually assessed via Python libraries.