Choosing a Location for a Chinese Restaurant in New York City, USA

Sam Thompson

11th May, 2020

1 Introduction

1.1 Background

New York City is a major world financial hub, a leading tourist destination, and the largest city in the United States. According to the US Census Bureau (2019), New York had a total population of over 8.3 million, comprising of the following racial demographics: White (32.1%), Hispanic or Latino (29.1%), Black or African American (24.3%), Asian (13.9%), American Indian and Alaska Native (0.4%), Native Hawaiian and Other Pacific Islander (0.1%), and Two or More Races (3.5%). New York is clearly a racially diverse city with a significant Asian population. As a resident of New York and an enthusiast of Chinese food, I would like to conduct a feasibility study to determine an ideal location for a new Chinese restaurant.

According to New York State's Office of State Comptroller (2014), the five counties with the highest concentrations of Asian New Yorkers were Queens (27%), Manhattan (14%), Brooklyn (13%), Tompkins (12%), and Nassau (10%). In searching for an ideal location within New York City, I will pay attention to Queens, Manhattan, and Brooklyn out of the five boroughs owing to the larger Asian population concentration. As recommended by Entrepreneur Media (2020), however, other factors must be taken into consideration, including target market, customer base, restaurant service style, and food concept before carving a niche for a restaurant business. These factors are explored more in the following sections.

1.2 Problem Statement

Foursquare is a leading location-based service provider with millions of venues from multiple geographical regions around the world. This project aims to utilize location-based characteristics to determine an ideal location for a new Chinese restaurant in New York City.

1.3 Interest

This project could benefit restaurant entrepreneurs in making data-driven decisions before choosing a restaurant location. A number of restaurants fail in their first year. Making the right decisions at the onset could go a long way in averting poor outcomes.

2 Data Acquisition and Cleansing

2.1 Data Sources

The New York City neighborhood data used in this project was originally sourced from the NYU Spatial Data Repository (https://geo.nyu.edu/catalog/nyu_2451_34572). For convenience, the dataset was accessed from the *Cognitive Class* course link https://cocl.us/new_york_dataset. Cognitive Class is an IBM-led initiative to democratize data knowledge and skills. New York City

has a total of 5 boroughs and 306 neighborhoods. The boroughs are Bronx, Brooklyn, Manhattan, Queens, and Staten Island. The dataset was used to obtain neighborhood details, including borough, neighborhood name, latitude, and longitude before segmenting and exploring the neighborhoods. For ease of reference, this dataset is referred to as the *New York Geospatial Dataset*.

The second dataset was scraped from the New York City Government website https://www1.nyc.gov/assets/planning/download/pdf/data-maps/nyc-population/census2010/t_pl_p3a_nta.pdf. The site contains 2010 population census details tabulated by neighborhood. For ease of reference, this dataset is referred to as the *New York Population Dataset*.

For a given neighborhood in the New York geospatial dataset, I explored and segmented nearby venues based on Foursquare location-based data and radius. Attributes such as venue name, category, latitude, and longitude were explored. I subsequently joined the New York geospatial, Foursquare, and New York population datasets. I then identified neighborhoods that have a higher Asian population density and lower Asian restaurant density. In the end, I presented a list of top 10 neighborhoods and made a final choice for a suitable neighborhood for a new Chinese restaurant.

2.2 Data Cleansing

The following steps were taken to cleanse the data:

1. Removed records with no neighborhood names.
2. Removed unwanted columns and borough names that were inadvertently scraped from the pdf source containing the New York population data.
3. Dropped records of neighborhoods that did not have Asian population.
4. Dropped records where Foursquare categories were blank.

3 Methodology

3.1 Overview

This project focused on utilizing New York geospatial, population, and Foursquare venue data to determine an ideal location for opening a new Chinese restaurant in New York City. The main steps were as follows:

1. Identify New York neighborhoods with higher Asian population densities
2. Identify Foursquare venue categories that are deemed to be Asian restaurants
3. Identify Foursquare neighborhoods and their counts of Asian restaurants
4. Join New York Asian population and restaurant data
5. For each neighborhood in the unified dataset, determine the Asian restaurant-to-population ratio
6. Identify the neighborhood with the lowest Asian restaurant-to-population ratio as the ideal location for a new Chinese restaurant.

3.2 Data Science / Machine Learning Tools Used

The following tools were utilized in this project:

1. Scraping data from web pages
2. Extracting data from location data providers
3. Creating Pandas dataframes and manipulating data

4. Geospatial mapping using folium
5. Graphing using Matplotlib

4 Data Exploration and Manipulation

4.1 New York Geospatial Data

The data elements obtained from the New York geospatial dataset include borough, neighborhood, latitude, and longitude. Sample representation is provided below:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.89471	-73.8472
1	Bronx	Co-op City	40.87429	-73.8299
2	Bronx	Eastchester	40.88756	-73.8278
3	Bronx	Fieldston	40.89544	-73.9056
4	Bronx	Riverdale	40.89083	-73.9126

Since the dataset contained details from all five boroughs, records for Bronx and Staten Island were removed from the analysis since these were determined to have low percentages of Asian population.

4.2 New York Population Data

The New York population dataset originally contained 13 data elements, namely: Borough, Census FIPS County Code, NTA Code, Neighborhood, Total Population, White, Black/African American, American Indian and Alaska Native, Asian, Native Hawaiian and Other Pacific Islander, Some Other Race, Total, and Hispanic Origin. It was determined that not all columns in the original dataset were relevant in the current project. After a series of manipulations and transformations, four data elements (Borough, NTA Code, Neighborhood, and Asian Population Percentage) were retained for use in this project. A sample representation is provided below:

	Borough	NTA Code	Neighborhood	Asian Pop. Percentage
0	Brooklyn	BK40	Windsor Terrace	9.89%
1	Staten Island	SI37	Stapleton	9.70%
2	Staten Island	SI37	Rosebank	9.70%
3	Brooklyn	BK25	Homecrest	9.51%
4	Manhattan	MN32	Yorkville	9.27%

Since the dataset contained details from all five boroughs, records for Bronx and Staten Island were removed from the analysis since these were determined to have low percentages of Asian population.

4.3 Joining New York Geospatial with Population Dataframes

The New York geospatial and population dataframes were joined and records without Asian population dropped. A sample representation is provided below:

	Borough	Neighborhood	Latitude	Longitude	NTA Code	Asian Population Percentage
0	Brooklyn	Windsor Terrace	40.65695	-73.9801	BK40	9.89%
1	Brooklyn	Homecrest	40.59853	-73.9592	BK25	9.51%
2	Manhattan	Yorkville	40.77593	-73.9471	MN32	9.27%
3	Brooklyn	Flatbush	40.63633	-73.9584	BK42	9.18%
4	Brooklyn	Brooklyn Heights	40.69586	-73.9938	BK09	8.75%

For ease of analysis, a dataframe was created for each of the three New York City boroughs of interest, namely Manhattan, Brooklyn, and Queens.

4.4 Foursquare Search Definition

After defining my Foursquare credentials, the search definition was set to include a limit of 200 (number of venues returned by Foursquare API) and radius of 15000 (meant to accommodate 15 kilometers from a venue). The Foursquare dataframe was defined to include venue name, category, latitude, and longitude. A sample representation is provided below.

	Name	Categories	Latitude	Longitude
0	Carl Schurz Park	Park	40.775118	-73.943763
1	City Swiggers	Beer Store	40.777515	-73.95082
2	Levain Bakery	Bakery	40.777354	-73.955284
3	Equinox East 92nd Street	Gym	40.7825	-73.95058
4	The Noguchi Museum	Art Museum	40.766735	-73.938445

A function was created to obtain Foursquare venues from all the neighborhoods in Manhattan. This was repeated for Brooklyn and Queens boroughs.

4.5 Joining New York Geospatial, Population, and Foursquare Venue Data

For each borough (Manhattan, Brooklyn, and Queens), a three-way join was made among geospatial, population, and Foursquare venue data. For ease of reference, this dataset will be referred to as the *3-way unified dataset*. A sample representation is provided below.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Asian Pop. %	Borough
0	Yorkville	40.77593	-73.947118	Peng's Noodle Folk	40.77726	-73.94911	Asian Restaurant	9.27%	Manhattan
1	Yorkville	40.77593	-73.947118	Mansion Restaurant	40.77596	-73.946777	Diner	9.27%	Manhattan
2	Yorkville	40.77593	-73.947118	Nica Trattoria	40.77569	-73.95057	Italian Restaurant	9.27%	Manhattan
3	Yorkville	40.77593	-73.947118	PuTawn Local Thai Kitchen	40.7746	-73.951042	Thai Restaurant	9.27%	Manhattan
4	Yorkville	40.77593	-73.947118	Poke Restaurant	40.77651	-73.950548	Sushi Restaurant	9.27%	Manhattan

With the foregoing manipulations and transformations, the data was set for further analysis and geospatial mapping.

5 Results

5.1 Identify Asian Restaurants in the Unified Dataset

Foursquare venue categories were examined with the aim of identifying Asian restaurants in the three boroughs. This is not a straight forward exercise since venue categories are not definitive. For instance, the following could all represent Asian restaurants: Afghan Restaurant, Asian Restaurant, Chinese Restaurant, and Japanese Restaurant. Based on venue categories that were available across the three boroughs, I visually inspected them and defined wider criteria for identifying Asian restaurants. A sample representation of the Asian restaurants is provided below.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Asian Pop. %	Borough
2	Yorkville	40.77593	-73.947118	Peng's Noodle Folk	40.777258	-73.94911	Asian Restaurant	9.27	Manhattan
20	Yorkville	40.77593	-73.947118	PuTawn Local Thai Kitchen	40.774599	-73.951042	Thai Restaurant	9.27	Manhattan
21	Yorkville	40.77593	-73.947118	Poke Restaurant	40.776508	-73.950548	Sushi Restaurant	9.27	Manhattan
22	Yorkville	40.77593	-73.947118	Pho Shop	40.778669	-73.948111	Vietnamese Restaurant	9.27	Manhattan
40	Yorkville	40.77593	-73.947118	Charley Mom	40.774154	-73.948486	Chinese Restaurant	9.27	Manhattan

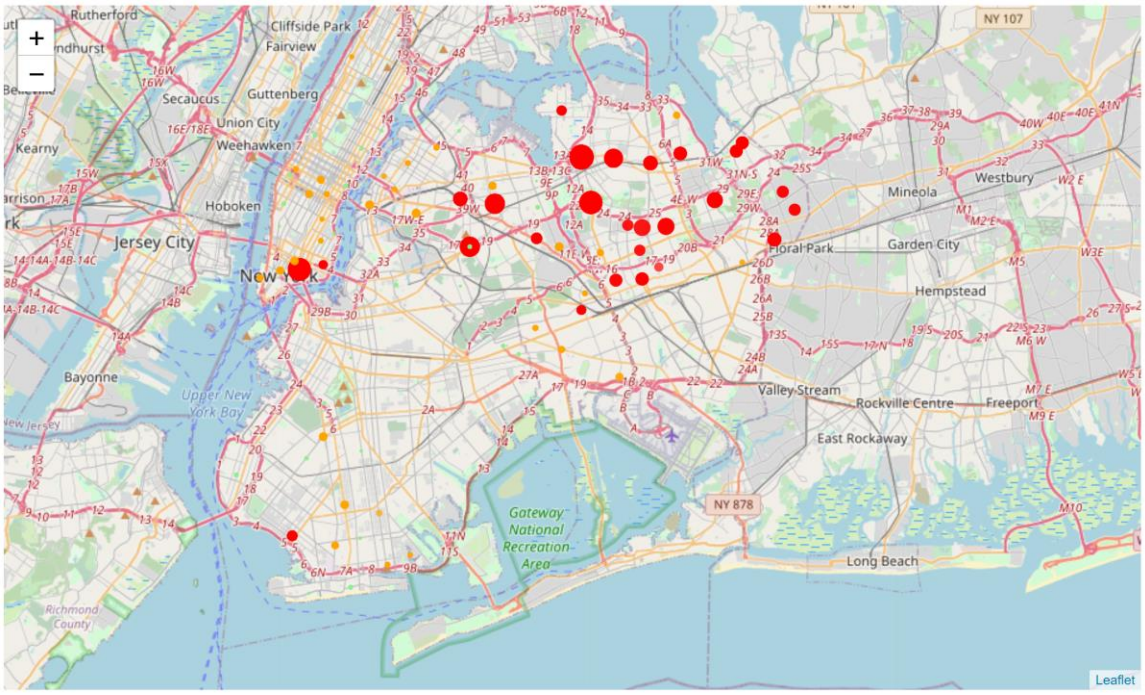
5.2 Asian Population Percentage and Restaurant Count by Neighborhood

Further analysis was conducted in order to obtain the Asian population percentage and restaurant counts by neighborhood across all three boroughs. For ease of reference, the abbreviation QUN is used in place of Queens, BKN for Brooklyn, and MHT for Manhattan. A sample representation is provided below.

Neighborhood	Asian Population %	Asian Restaurant Count
Flushing (QUN)	69.20	14
Queensboro Hill (QUN)	65.03	6
Chinatown (MHT)	63.87	23
Elmhurst (QUN)	56.07	24
Maspeth (QUN)	56.07	8
Murray Hill (QUN)	52.54	25
Utopia (QUN)	47.05	1

5.3 Mapping of Asian Population Density Across the Three Boroughs

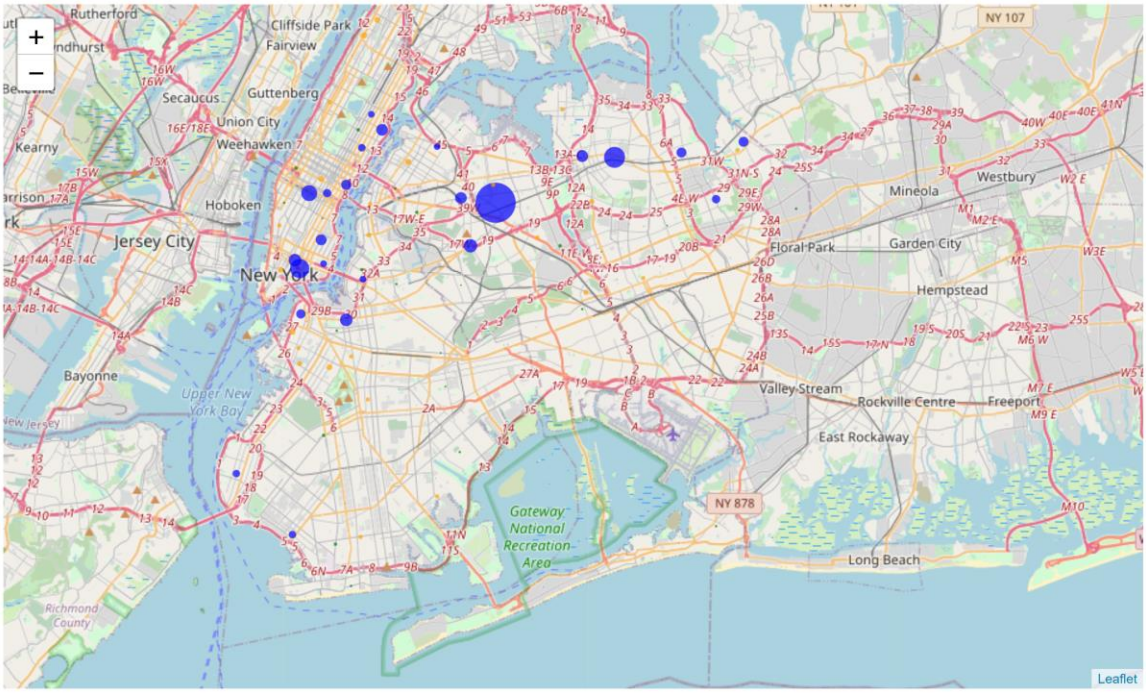
Further analysis was conducted so as to facilitate the mapping of the Asian population percentages across the three boroughs.



The size of the bubble marker is based on the Asian population percentage by neighborhood. The color of the bubble markers represents quartiles of the Asian population percentage (light blue for 1st quartile, light green for 2nd, orange for 3rd, and red for 4th quartile respectively).

5.4 Mapping of Asian Restaurant Counts Across the Three Boroughs

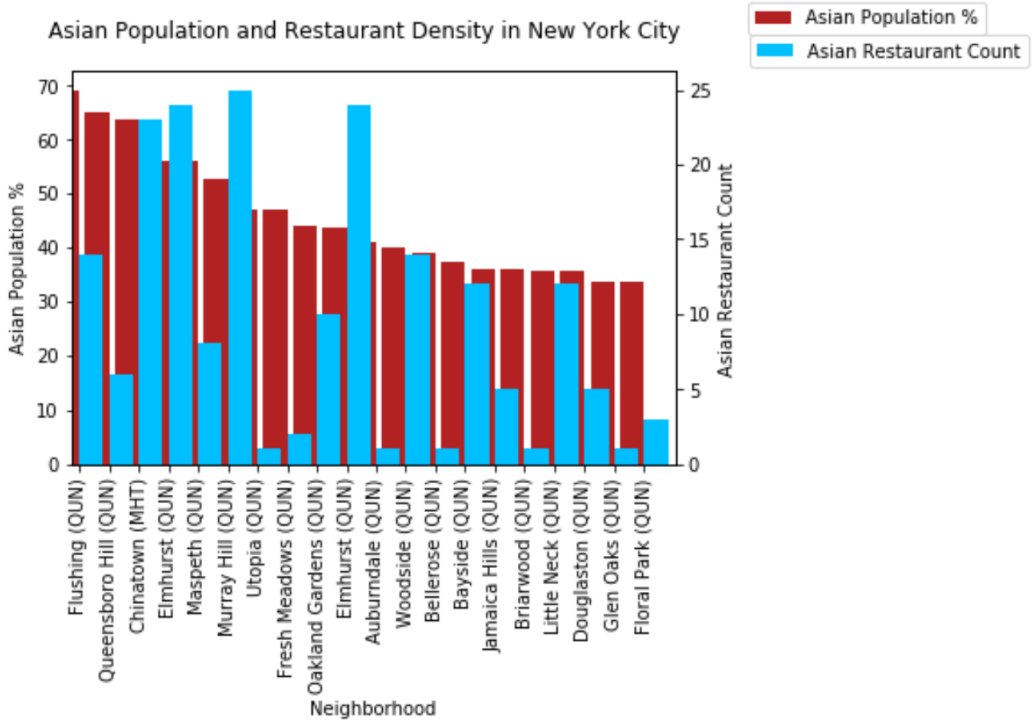
Further analysis was conducted so as to facilitate the mapping of the Asian restaurant counts across the three boroughs.



The size of the bubble marker is based on the Asian restaurant counts by neighborhood. The color of the bubble markers represents quartiles of the Asian restaurant counts (green for 1st quartile, orange for 2nd, blue for 3rd, and purple for 4th quartile respectively).

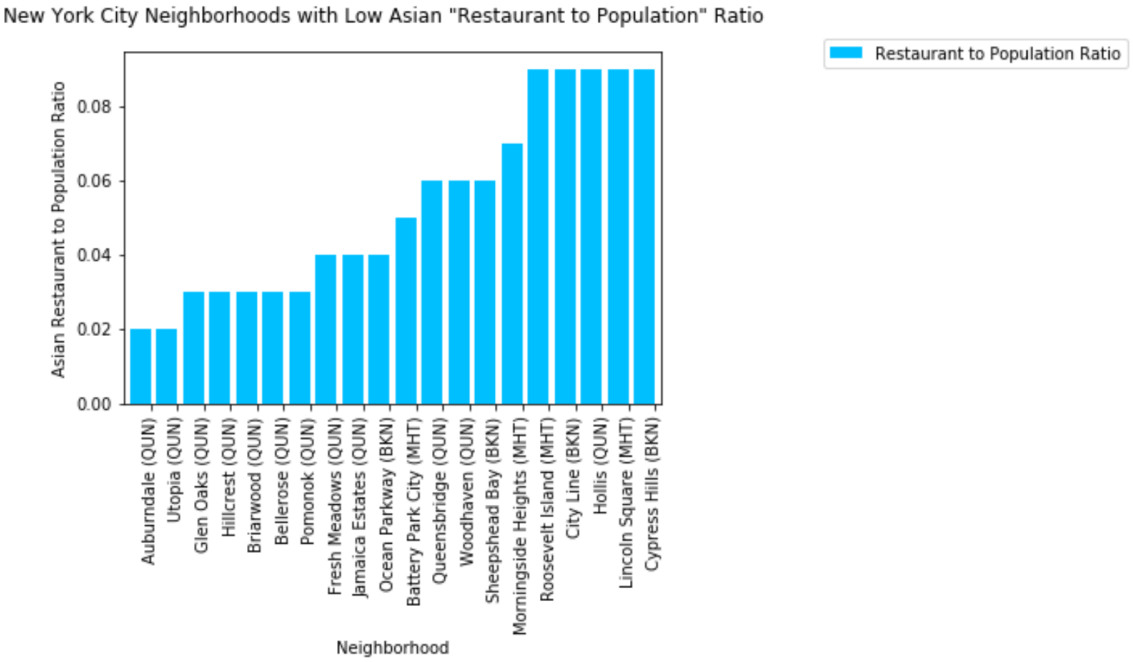
5.5 Combined Graph of Asian Population and Restaurant Density

Further analysis was conducted so as to facilitate the combined graphing of Asian population and restaurant densities (by neighborhood) across the three boroughs. Results are shown below (based on top 20 neighborhoods by population)



5.6 Graph of Restaurant-to-Population-Ratio

Further analysis was conducted so as to facilitate the graphing of restaurant-to-population-ratio (by neighborhood) across the three boroughs. This was meant to identify neighborhoods with higher Asian populations but low Asian restaurants. The top 20 neighborhoods are shown in the graph below.



6 Discussion of Results

Based on lowest restaurant-to-population-ratio, the best neighborhood for the next Chinese restaurant in New York City is **Auburndale** (located in the Borough of Queens). The top 10 neighborhoods for opening a Chinese restaurant in New York City are as follows:

1. Auburndale (Queens)
2. Utopia (Queens)
3. Glen Oaks (Queens)
4. Hillcrest (Queens)
5. Briarwood (Queens)
6. Bellerose (Queens)
7. Pomonok (Queens)
8. Fresh Meadows (Queens)
9. Jamaica Estates (Queens)
10. Ocean Parkway (Brooklyn)

Besides low restaurant-to-population-ratio, however, investors should consider other factors such as neighborhood security, infrastructure, business model (such as eat-in vs. delivery), and neighborhood income which are not accounted for in this project. Opportunities for further research are presented in the next section.

7 Conclusion and Opportunities for Further Research

This project focused on utilizing New York geospatial, population, and Foursquare venue data to determine an ideal location for opening a new Chinese restaurant in New York City. In the end, the neighborhood of Auburndale (Queens) was identified as the most ideal based on its lowest restaurant-to-population ratio. This project demonstrates the power of data science in driving

business decision-making such as choosing a location for a business enterprise. Opportunities for further research include the following:

1. It is possible that not all Chinese restaurants are listed on Foursquare. This was evident in Manhattan's Chinatown which has a very high Asian population but no fewer Asian restaurants on Foursquare. The research could be improved by reviewing other location-based service providers.
2. It was not easy to distinguish among Asian and Chinese restaurant categories, thereby making the assumption that higher Asian populations include Chinese people. Moreover, it was assumed that non-Chinese Asian people will dine at Chinese restaurants if there is no other Asian restaurant in the neighborhood. The results could be improved by researching the dining tendencies of Chinese people and Asians in general in New York City.

8 References

Entrepreneur Media Inc (2020). How to Start a Restaurant. Retrieved from https://www.entrepreneur.com/article/73384.

New York City Department of City Planning (2011). Total Population by Mutually Exclusive Race and Hispanic Origin: New York City Neighborhood Tabulation Areas, 2010. Retrieved from https://www1.nyc.gov/assets/planning/download/pdf/data-maps/nyc-population/census2010/t_pl_p3a_nta.pdf.

Office of the New York State Comptroller (2016). The Asian Community in New York State. Retrieved from https://www.osc.state.ny.us/osdc/rpt9-2016_eng.pdf.

United States Census Bureau (2019). QuickFacts New York City, New York. Retrieved from https://www.census.gov/quickfacts/newyorkcitynewyork.