Opening a Traditional Chinese Restaurant in New York City

August 2020

by V Badri

Coursera Capstone

IBM Applied Data Science Capstone Project

The Battle of the Neighborhoods

Introduction & Business Problem

The New York City, is one of the most populous cities in the United States, which is also ethnically diverse and is the financial capital of USA. It is the largest city in USA with long history of international migration. Being business friendly it has become the centre for banking & finance, tourism, food industry, retail, real estate, news media, legal services, insurance, transportation, fashion and arts in the USA. It also provides homes for 8.7 million people (2020) accounting close to 38% of total population of New York state.

It is evident that any business is highly competetive, establishing a new business is also requires huge capital as it is highly developed city. The project we are looking into explores all the possible options available for opening a Traditional Chinese restaurant by our company ABC restaurant chain, as this will be their first chinese restaurant for them and they have many questions that need be answered. The insight provided by this analysis will help in understanding the competetion they currently have, where they can target and eventually have best ROI.

Business Problem Description

Whole point of restautant is to serve food and drink to consumers in return for money, as it is evident that City of New York is already famous for excellent food culture and wide range of international cuisines influenced by its immigration history.

The problem **ABC** are facing is they don't have all the information/insights and data about existing restaurants available in hand to choose the best place to start their inagural Chinese restaurant and also want to satisfy the investors by showing the insights and have their full backing to proceed further.

The main objective of this project is to analyse and understand the data of existing restaurants, explore to find the best place to open the restautant, the insights provided by our analysis should help the investor to make the decision easily.

Target Audience

Investors of ABC Restaurant Chain, management of the chain has recruited a Data Science Team in order to locate and recommend the best neighborhood of New York city to start their inagural Chinese Restaurant and in turn they can put it to the investors, make them understand how quick they can turn the ROI.

Data

In order to get a complete view we will be using the below data sets and initiate our analysis

Neighborhood data - This covers all the details of 5 boroughs and 306 neighborhoods
Geographical data - This covers all the coordinates for the respective neighborhoods, eventually used to plot the maps
Venue Data - This covers all the existing venues/restaurants established in the same business sector, we will use to provide more insights on how they impact our decision

Data Sources, Extract and merging plan

The open data source (https://cocl.us/new_york_dataset) will be used to read borough and neighborhoods data, then geographical coordinates will be added to the respective neighborhoods using Pythin Geocoder Package.

Foursquare API will be used to get the venue data for the neighborhoods and will eventually help us to provide insights to our business problem.

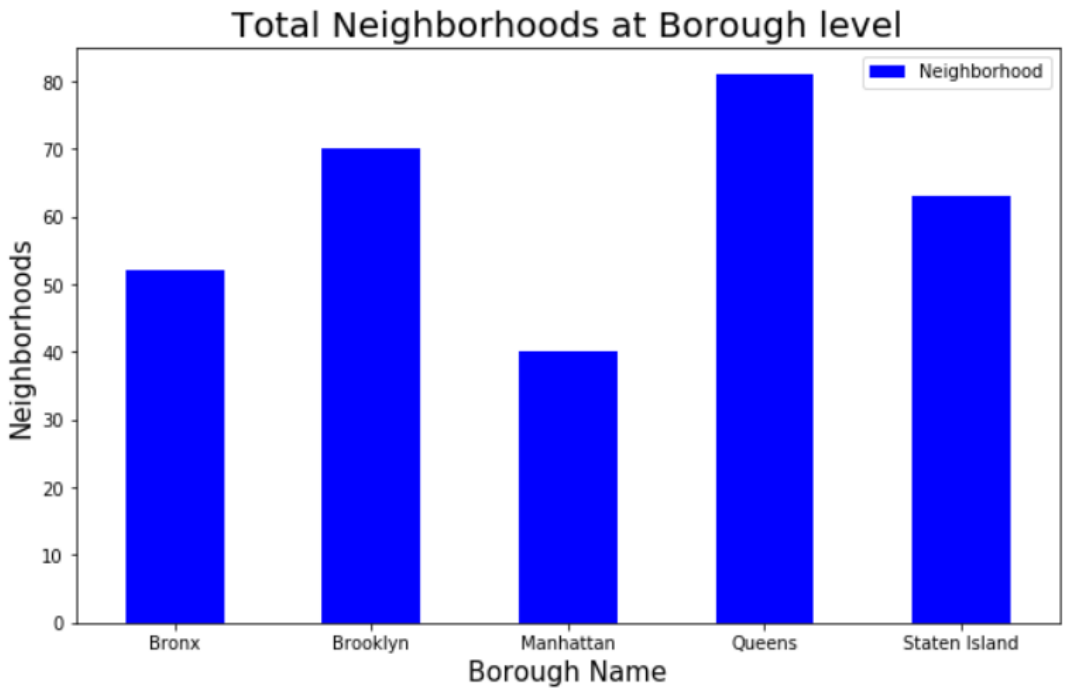
Data Science Techniques used

Web Scrapping
Fetching Data from API(json)
Data Cleaning
Grouping
Word Cloud
Data Wrangling
Map Plotting

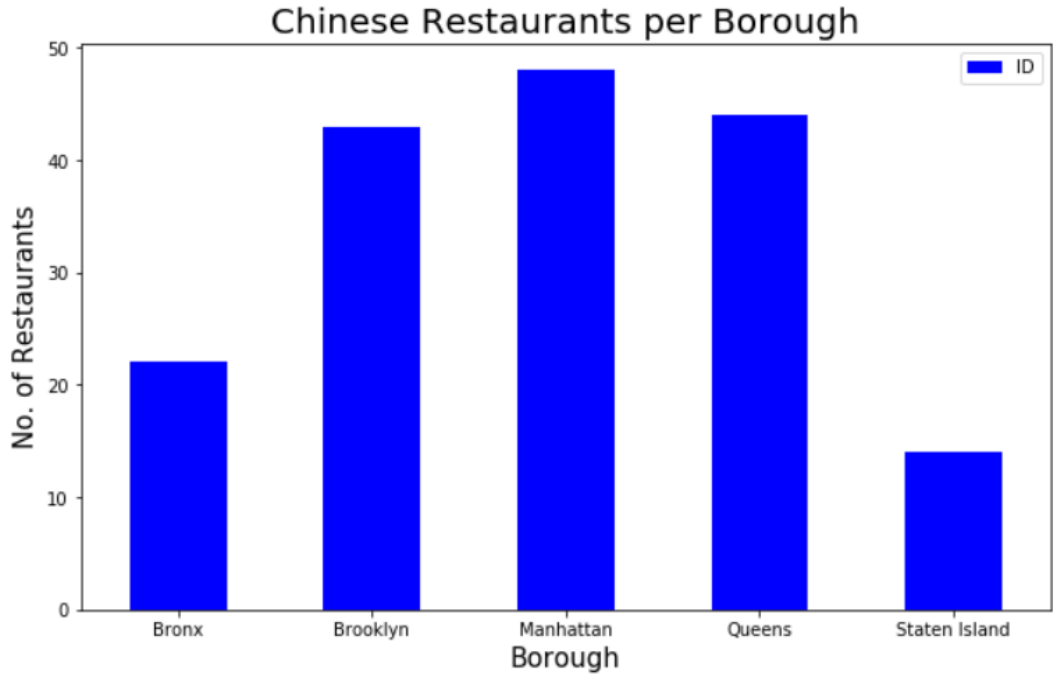
Methodologies

• Extracted data from https://cocl.us/new_york_dataset will be cleaned before storing into a dataframe
• FourSquare API will be used to locate all the NY City venues and then Chinese restaurants will be filtered out into the dataframe
• Rankings will be provided by sorting out the data
• The extracted data will be graphically represented, maps will be plotted using the given co-ordinates

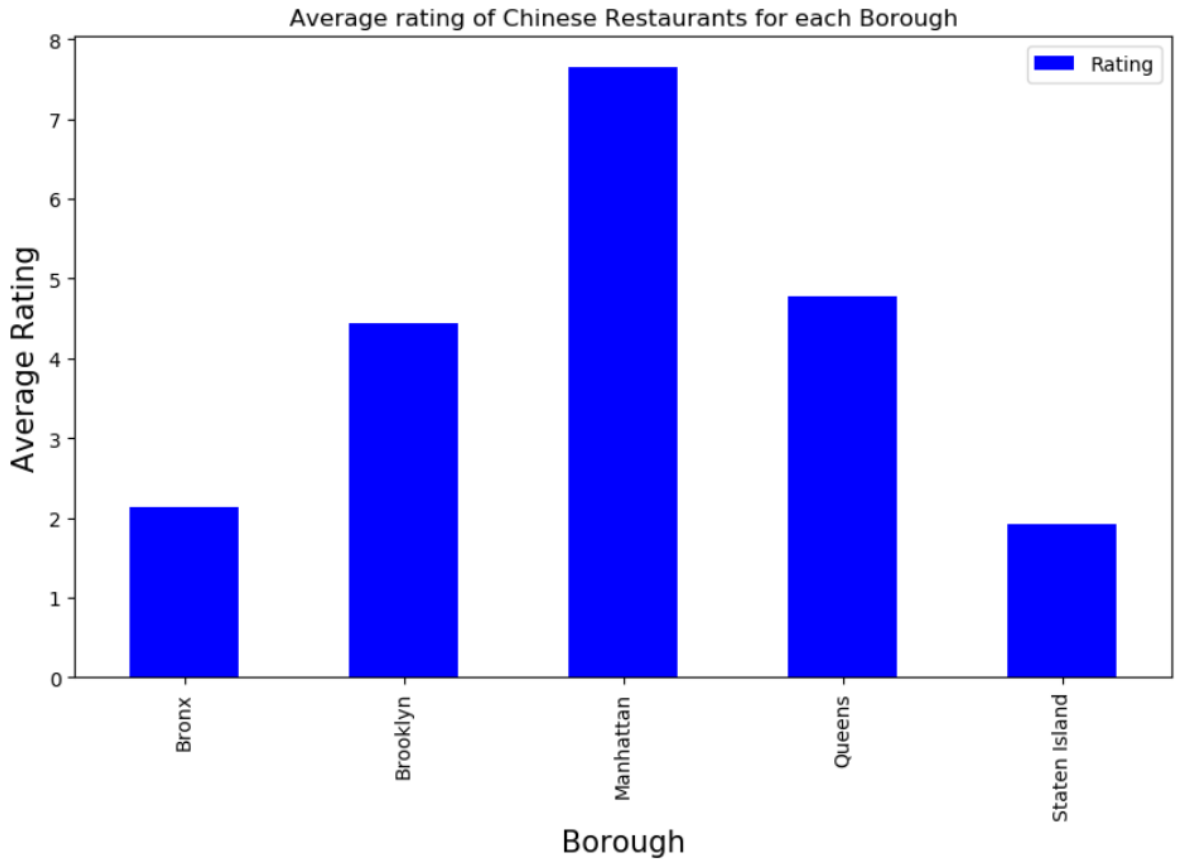
Results



It is clearly visible that Queens leads the chart with highest number of Neighborhoods across other borough's

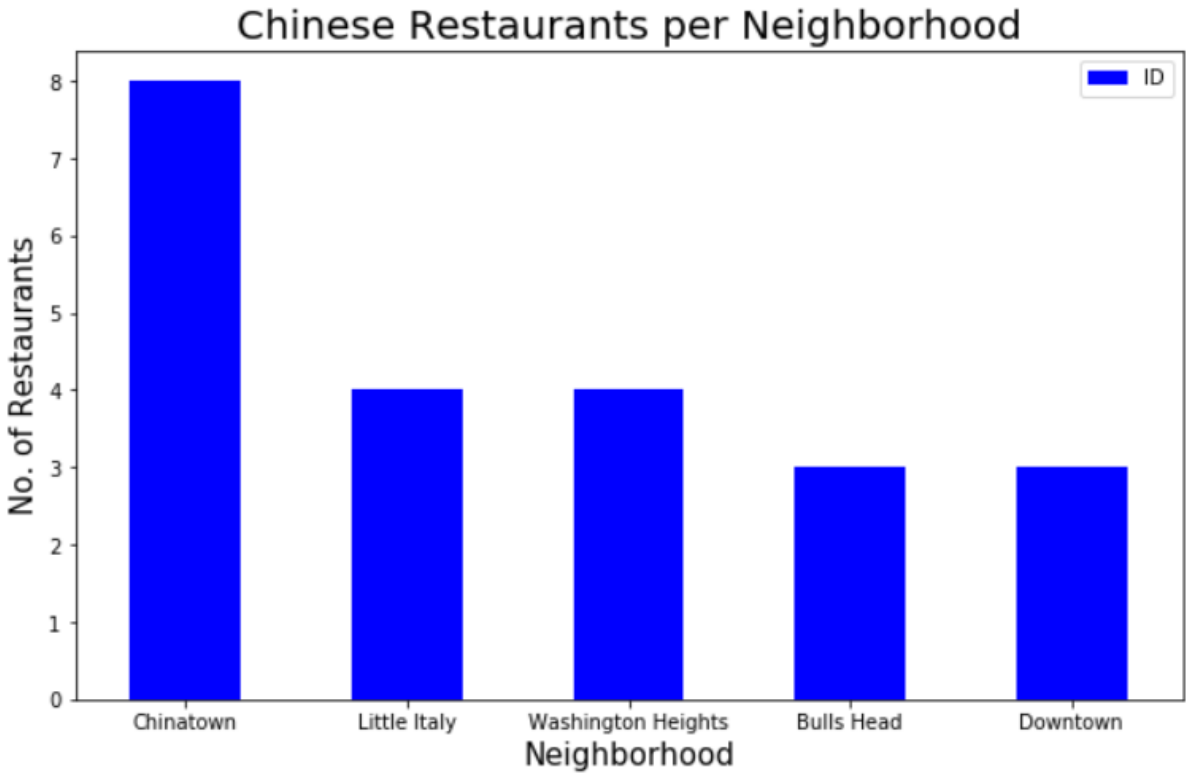


Though Manhattan has only 40 neighborhoods, it races to the top with 48 restaurants



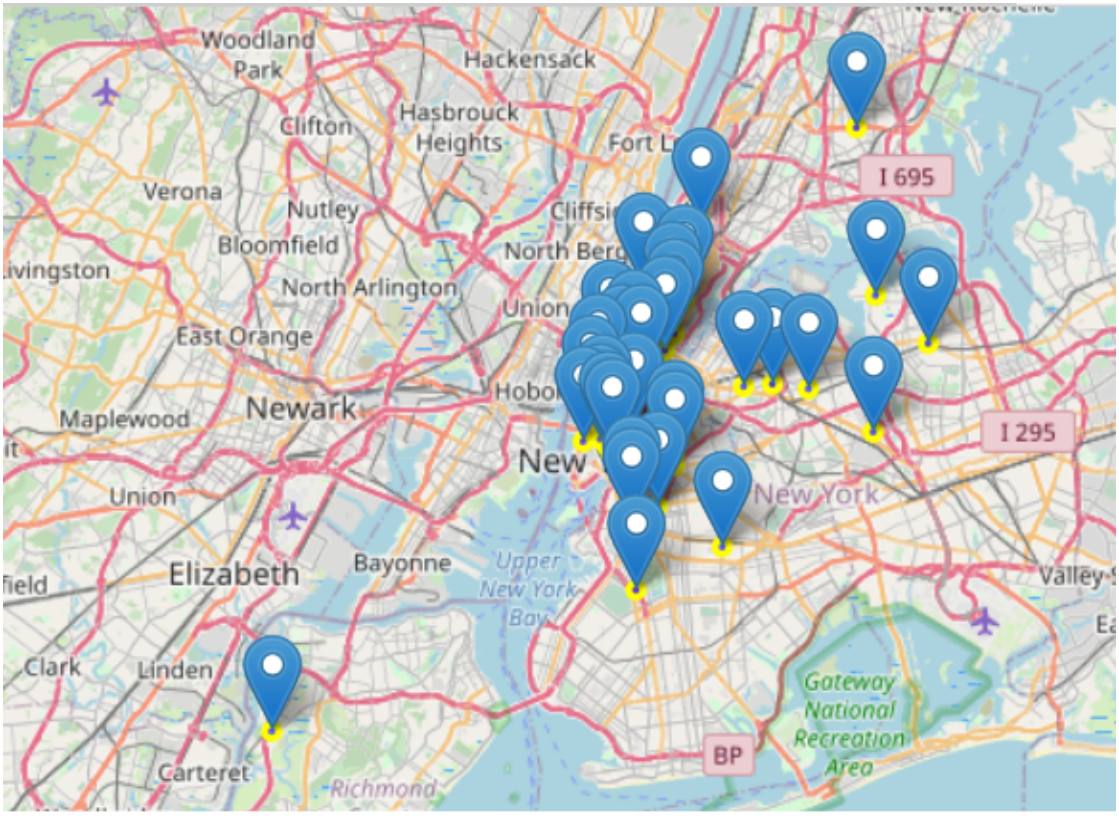
	Borough	Average Rating
2	Manhattan	7.647917
3	Queens	4.770455
1	Brooklyn	4.439535
0	Bronx	2.127273
4	Staten Island	1.921429

Manhattan, again, stood top with highest average ratings for the respective chinese restaurants



	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
47	Manhattan	Chinatown	4db3374590a0843f295fb69b	Spicy Village	501.0	8.7	179.0
48	Manhattan	Chinatown	4a96bf8ff964a520ce2620e3	Wah Fung Number 1 Fast Food 華豐快餐店	194.0	8.5	96.0
49	Manhattan	Chinatown	5894c9a15e56b417cf79e553	Xi'an Famous Foods	104.0	8.9	29.0
50	Manhattan	Chinatown	3fd66200f964a520b1ea1ee3	Great N.Y. Noodletown	547.0	7.9	287.0
51	Manhattan	Chinatown	5c965dad5455b2002c058659	Yi Ji Shi Mo Noodle Corp	21.0	8.7	6.0
52	Manhattan	Chinatown	59d828f0916bc1155fde2c04	Hwa Yuan	69.0	8.0	19.0
53	Manhattan	Chinatown	59ad625a2619ee5cdded881f	House Special 甘來飯店	22.0	8.3	13.0
54	Manhattan	Chinatown	3fd66200f964a520ede41ee3	Big Wong King 大旺	422.0	8.4	185.0

With in Manhattan, Chinatown stood upto the name by having highest number of Chinese restaurants across rest of neighborhoods and intrestingly worth noting that all the restaurants within China Town are having ratings minimum of 8 (except one 7.9)



Discussion

Based on the churned numbers and stats, I would say we may have two options to pick a place, we can pick one of the top 3 (Manhattan, Queens, Brooklyn) but the investment would relatively linear to the average ratings of our boroughs as we can see Manhattan is very popular place and the rents would be high and very competetive, especially in chinatown it is very competetive.

On the other hand, Queens and Brooklyn don't have the best restaurants and it is worth a shot in establishing a restaurant with relatively little low investment in comparison to Manhattan and maintaining it to the par of Manhattan restaurants would also be another option explore further.

Conclusion

I think we have covered everything, assuming this would help ABC to reach a conclusion. Before concluding this, I would like to say as these results are completely data driven, the results we have are completely subjective to the credibility of the data we currently have, if we have some more data available via different sources, with more FourSquare calls (premium), we may be able to do further more analysis and may impact the results in future.