

Battle of the Neighborhoods

Aditya Hariharan

25/03/2020

1.Introduction

1.1 Background

The City of New York, is the most populous city in the United States. It is diverse and is the financial capital of USA. It is multicultural. It provides lot of business opportunities and business friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce. The city is a major center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in the United States.

This also means that the market is highly competitive. As it is highly developed city the cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analyzed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

1.2 Problem

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. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world. English remains the most widely spoken language, although there are areas in the outer boroughs in which up to 25% of people speak English as an alternate language, and/or have limited or no English language fluency. English is least spoken in neighborhoods such as Flushing, Sunset Park, and Corona.

With New York's diverse culture, comes diverse food items. There are many restaurants in New York City, each belonging to different categories like Chinese, Italian, Indian, French etc. So as part of this project, we will list and visualize all major parts of New York City that has great Italian restaurants and determine the best place to build a new restaurant.

1.3 Interest

To recommend the correct location, AH Food Company Ltd has appointed me to lead of the Data Science team. The objective is to locate and recommend to the management which neighborhood of New York City will be best choice to start an Italian Restaurant.

2. Data Acquisition and Cleaning

2.1 Data Sources

We get the list of boroughs and neighborhoods along with their coordinates using the [New-York-Dataset](#). We also use the Demographic data for the boroughs found [here](#). We then use the foursquare API which can be found [here](#), to get the venues in NYC.

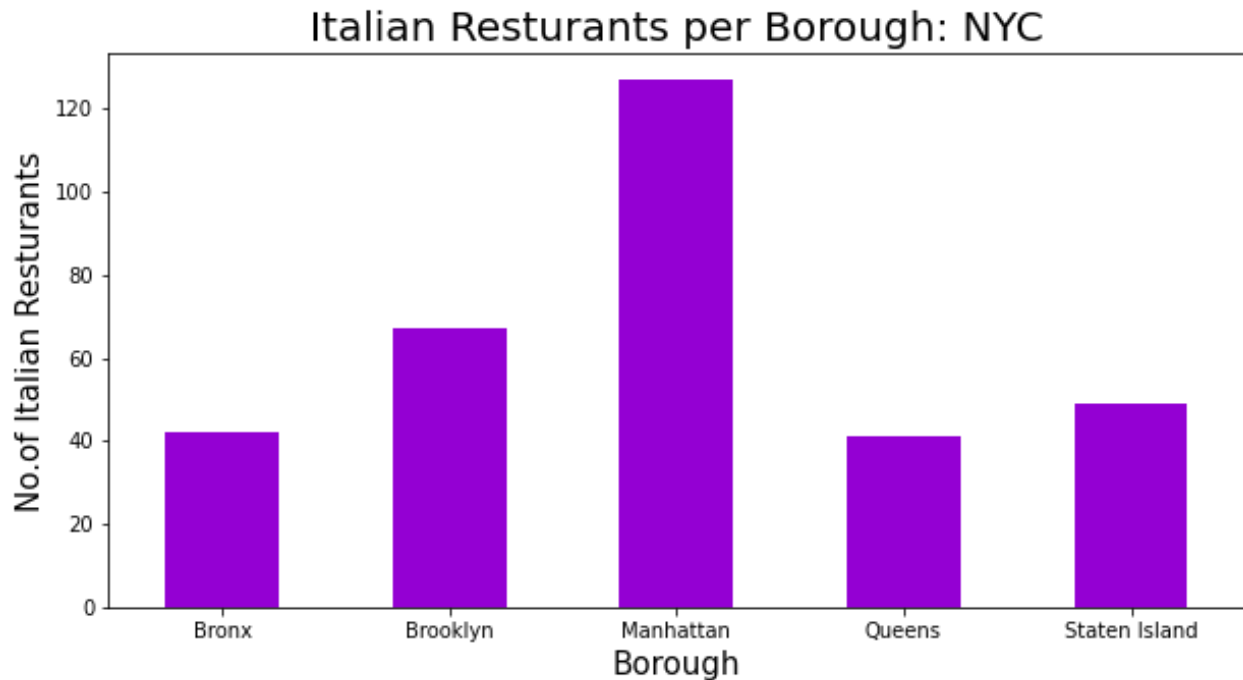
2.2 Data Cleaning

The data from the New York dataset is already in clean form and all that is done is to request it and put the data in a data frame. The Venues in NYC were then obtained using the FourSquare API and then filtered so that a new data frame contained only the Italian Restaurants in NYC. The New York data and the venues data were then merged to make a final data frame. The demographic data was found in a wiki html page, which was first extracted using beautifulsoup4 and then put into a data frame. I then dropped unnecessary columns from the demographic data frame.

3. Analysis

3.1 Analyzing the Demographic Data

The demographic data is mainly used to see the population in each borough and the per capita income which suggests the number of people that would go out and spend their money in a Restaurant. On analyzing the number of restaurants in each borough and the demographic of each borough, we can choose a Borough to build a restaurant in.



We can see that Queens has the lowest number of Italian restaurants from the above graph.

	Borough	Population	per capita
0	The Bronx	1,432,132	29,200
1	Brooklyn	2,582,830	34,600
2	Manhattan	1,628,701	360,900
3	Queens	2,278,906	39,600
4	Staten Island	476,179	30,300

Looking at the demographic data, we can see that Queens has the 2nd highest number of people and the 2nd highest per capita.

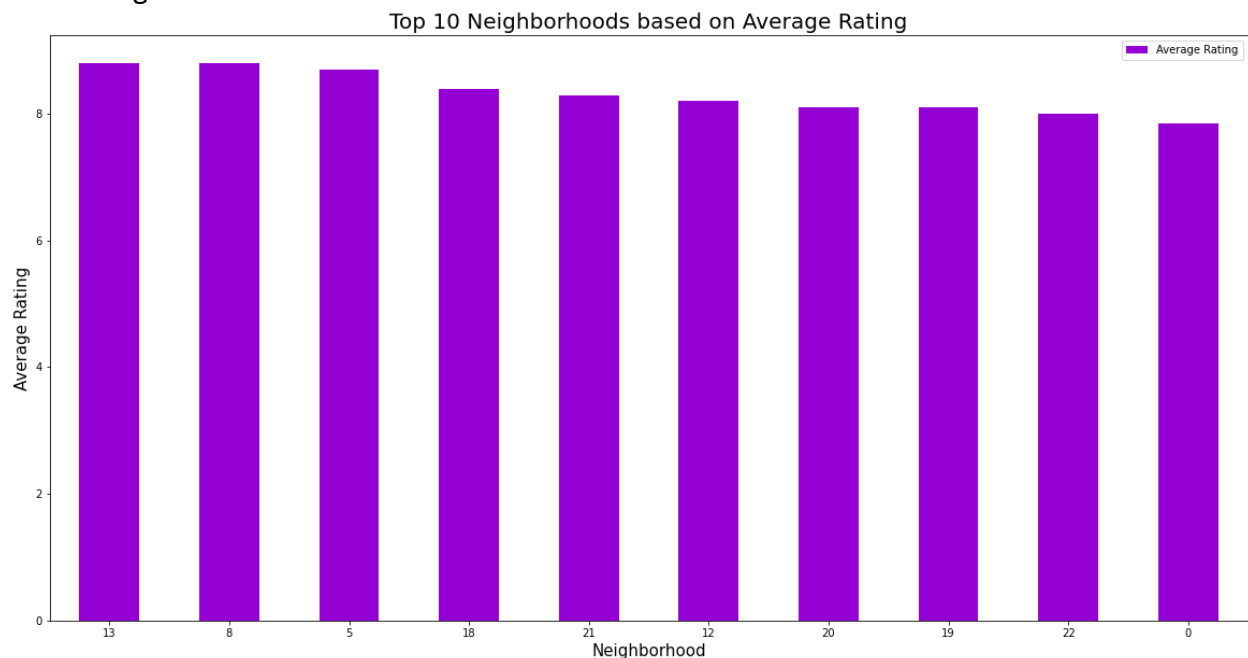
Therefore, we can choose Queens as the Borough to Build the new Italian Restaurant.

3.2 Analyzing the ratings of existing restaurants to determine neighborhoods

Using the FourSquare API, we get the rating, tips and likes of each of the existing restaurants into a data frame. Below is the `df.head()` of said data frame.

	Neighborhood	ID	Likes	Name	Rating	Tips
0	Murray Hill	4afa15d8f964a520931622e3	38.0	Villa Berulia	8.5	13.0
1	Murray Hill	4b3c0273f964a520488025e3	54.0	Rossinis	7.9	21.0
2	Murray Hill	4d99cbdc7cfc145660ac49f6	250.0	Piccolo Cafe	7.9	115.0
3	Astoria	5010b169e4b05edd9983ec7d	72.0	Antika Restaurant & Pizzeria	8.2	31.0
4	Astoria	4ebf0e380aaf6450ba608920	57.0	Via Trenta	7.5	28.0

We then average the rating based on the neighborhood to see which neighborhoods have the best ratings.



Based on the graph, we then filter out all the neighborhoods that have an average rating of less than 8.

	Neighborhood	Average Rating
22	Ridgewood	8.000000
0	Astoria	7.850000
11	Howard Beach	7.466667
3	Bayside	7.450000
15	Little Neck	7.233333
14	Kew Gardens	7.200000
24	Sunnyside	7.166667
17	Maspeth	7.100000
9	Douglaston	7.000000
1	Astoria Heights	6.900000
16	Long Island City	6.900000
2	Auburndale	6.900000
23	Steinway	6.700000
10	Forest Hills	6.500000
4	Beechhurst	6.400000
7	Bellerose	6.100000
6	Belle Harbor	5.900000

We then check the number of restaurants in each of these Neighborhoods and make a list of neighborhoods that have only 1 restaurant.

4. Results

Our Data shows that there a huge number of Italian Restaurants spread across all five boroughs in NYC, around 315 in total. Looking at the distribution, we found out that Manhattan had the highest number of Italian Restaurants, 127, and Queens had the lowest number of restaurants, 41. This statistic made me realize Queens as the Borough to pick for building the new restaurant.

As only one statistic is not enough, we looked at the Demographics of NYC to determine the population and per capita income per borough as both relate to how many customers the restaurant gets. We found out that Queens had the second highest population in NYC, 2,278,906, and the second highest per capita income, \$39,600 which makes it an ideal spot to open a new restaurant.

We then found out which Neighborhood in Queens have Italian restaurants and we also looked at the average rating of these restaurants based on their neighborhood. Based on this and the number of Italian Restaurants in each neighborhood, there are a number of ideal locations where the stakeholders can choose to build a restaurant. They are:

Ridgewood, Kew Gardens, Maspeth, Douglaston, Astoria Heights, Long Island City, Steinway, Forest Hills, Beechhurst and Belle Harbor.

5. Conclusion and Future Directions

Purpose of this project was to identify NYC areas with low number of restaurants (particularly Italian restaurants) in order to aid stakeholders in narrowing down the search for optimal location for a new Italian restaurant. By calculating restaurant density distribution from Foursquare data, we have first identified general boroughs that justify further analysis and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby restaurants.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.