IBM- COURSERA CAPSTONE PROJECT

The Battle of Neighbourhoods

ABSTRACT

This project aims to illustrate the application of machine learning and the Foursquare API services to help determine which would be the best Borough and Neighborhood in New York City (NYC) to open an Italian restaurant. The project will highlight a fictional scenario for demonstrating the benefits and intent of this project.

By: Tevin Richard

March 2021

Table of Contents

1. Introduction

Entrepreneurs are constantly attempting to start businesses for the purpose of deriving an income stream and establishing a successful business empire. One of the largest challenges entrepreneurs face when starting a new business is finding the optimal location for starting up.

Location is key when starting a business since an ideally situated business will receive high customer traffic which could potentially translate into profitability. Furthermore, it is important to start a business in a location where the product or service you intend to offer is not readily available. This implies strategically starting up in an area that is not rife with competition, thereby allowing you to meet the needs and/or wants of that population. By strategically opening a business in an area that has low levels of competition, it can be assumed that you will not have to keep up with competitive prices of the already established enterprises which would drastically increase your businesses' probability of success.

CNBC recently named New York City (NYC) as the "top city to open a small Business in America". NYC is regarded as one of the world's largest centers for finance therefore making it an ideal location to consider when opening a business. Furthermore NYC is also considered as one of the most diverse cities in the United States (Financial, 2019), thereby allowing your business to potentially reach a wide population with varying interests, habits as well as likes or dislikes. Opening a restaurant in NYC would increase your businesses' probability of success as it can be assumed that your businesses' products or services would at least appeal to a small/large faction of the diverse population.

The next problem is to determine which Borough and neighborhood in NYC would be the most ideal location to purchase/lease a property for the purpose of starting a business.

1.1 Scenario

For the purpose of this project, we will explore the fictional scenario of an Italian entrepreneur, Raymond who had recently migrated to New York with the ambition of starting up a successful Italian restaurant. Raymond has limited finances and must therefore carefully consider all risks when making business decisions.

1.2 Business Problem

Raymond is aware that there are numerous competitor Italian restaurants that are already established in New York City and therefore requires assistance in determining which areas/neighborhoods are least densely populated with competition.

The objective of this project is to answer the following problem questions:

1. Which borough in NYC would be the best borough to open an Italian restaurant? (Based on least competition)
2. Which location has the highest number of Italian restaurants?
3. Which Neighborhood would be the best Neighborhood to open an Italian restaurant? (Based on least competition)
4. Who would be the top 5 most competitive Italian restaurants in NYC that Raymond would have to compete with? (Based on customer ratings)
5. Which Italian restaurant would represent the lowest level of competition?
6. What would be the best locations (street names) to open an Italian restaurant?

1.3 Target Audience

This project was designed to assist the target audience (entrepreneurs like Raymond) in deciding upon which geographical location within a specified city (NYC for this project) would be ideal for starting up a new business.

2. Data Section:

This section describes the data that will be used for this project.

2.1 New York City data

This data set will be used to obtain information for the list of all Boroughs and Neighborhoods in NYC, together with their respective latitudes and longitudes. This dataset will be used as a starting point for which the latitudes and longitudes for each Borough and Neighborhood will be used within Foursquare to extract information for Italian Restaurants.

Source: This data can be obtained at the following link.

2.2 Italian Restaurants Data

The data for all NYC Italian restaurants, including their stats (likes, ratings and tips) will be extracted using the Foursquare API.

The first request will involve retrieving the latitudes, longitudes, unique venue id and venue names for each Italian restaurant for each Borough and it's respective neighborhoods. The second request will be to retrieve stats (likes, ratings and tips) for each Italian restaurant found.

Source: Foursquare API

3. Methodology

This section describes the methodology that was adopted to identify the best location(s) to for an entrepreneur like Raymond to open an Italian restaurant in NYC.

3.1 Import the required libraries:

The following python libraries were imported for use in the project:

```
import numpy as np
import pandas as pd
from pandas.io.json import json_normalize
import requests
import folium
import geocoder
import urllib.request
from urllib.request import urlopen
from geopy.geocoders import Nominatim
import os
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import matplotlib.colors as colors

# configure libraries settings
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
%matplotlib inline


print("All libraries have been successfully imported")
```

```
All libraries have been successfully imported
```

Figure 1: Libraries utilized in project

3.2 Retrieve the NYC location data

- The data containing Boroughs, Neighborhoods and their respective latitudes and longitudes were extracted and loaded into a panda's data frame for further analysis.
- This data frame formed the basis for further extracting venue data from the Foursquare API.
- A function was created to analyze the JSON file and extract the data in a pandas data frame. The function ran through each line of the "*features*" in the response query and appended the data to the data frame as a new row.
- The resulting data frame was of the following structure:

	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 2: NYC Location data frame

- Analyzing the shape of the dataset revealed that there are 306 Neighborhoods amongst all boroughs in NYC.
- The data was then grouped by Borough to determine the number of neighborhoods per Borough as represented in Figure 3 below.

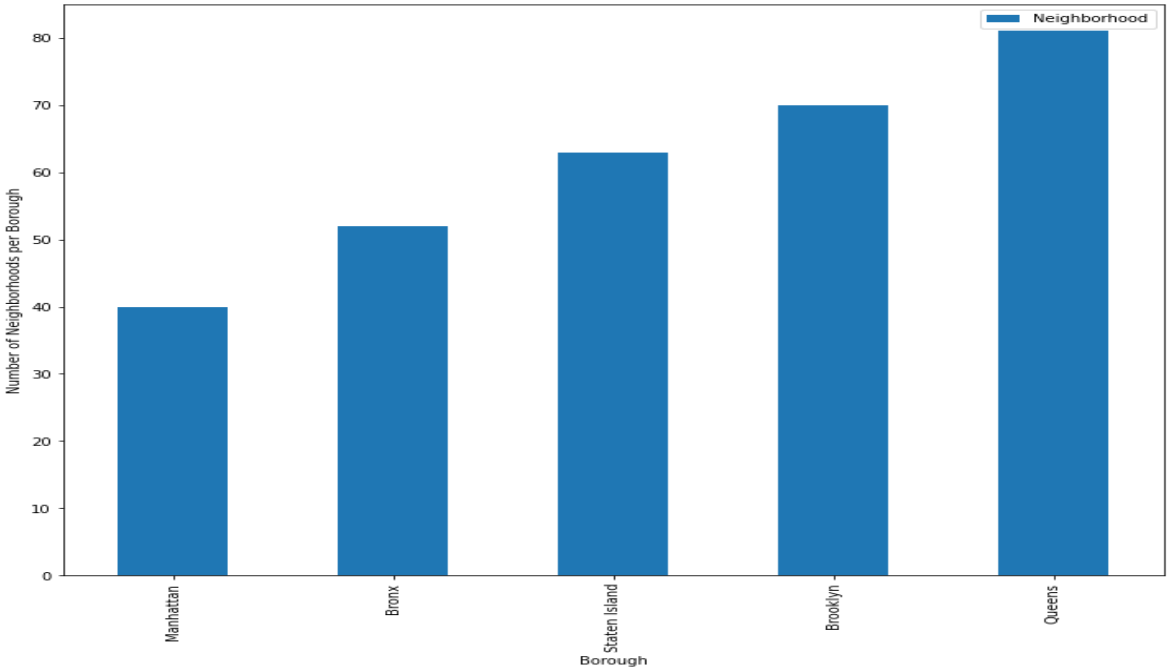

Figure 3: Grouping and count of Neighborhoods per Borough in NYC

- Manhattan was found to be the smallest Brough with 40 Neighborhoods and Queens was the largest borough with 81 neighborhoods.

3.3 Determining the best Borough

The ideal location to open a new Italian restaurant would be one that is up-and-coming and is well known by food lovers. According to (Extra Space Storage, 2016), Manhattan is the best rated borough for dining and nightlife. From the plot above, Manhattan is also the smallest borough, with the least number of neighborhoods.

These suggest that **Manhattan** would be the **best location** for establishing an Italian restaurant due to it being the best rated by food lovers for dining and night life as well as it being the smallest brough (better chances of a new restaurant popping up in searches within that location).

The data set was then narrowed to include only neighborhoods within Manhattan.

3.4 Retrieve locations of Italian restaurants in Manhattan

- The Foursquare unique ID for Italian restaurants (“4bf58dd8d48988d110941735”) was used to make requests to the Foursquare API to retrieve the following data for each Italian restaurant:

Item	Description
1. Venue	Name of Italian restaurant
2. Venue ID	Unique restaurant ID assigned by Foursquare
3. Venue latitude	Latitude of Italian restaurant
4. Venue Longitude	Longitude of Italian restaurant
5. Venue Category	Category of venue

- A new function was then created to go through the data frame created in step 3.2 and retrieve the information for venue, venue ID, venue latitude, venue longitude and venue category for each Italian restaurant.
- The search was limited to restaurants within a 1000 m radius of the latitude and longitude of each neighborhood in the Manhattan borough.
- The extracted data obtained using the created function and Foursquare API had the following structure:

	Unnamed: 0	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category
0	0	Marble Hill	40.876551	-73.910660	Arturo's	4b4429abf964a52037f225e3	40.874412	-73.910271	Pizza Place
1	1	Marble Hill	40.876551	-73.910660	Broadway Pizza & Pasta	4be72770910020a16f1ad514	40.878822	-73.904494	Pizza Place
2	2	Marble Hill	40.876551	-73.910660	Patricia's of Tremont	4df272a918386ecb4e2cab2a	40.882543	-73.913501	Italian Restaurant
3	3	Marble Hill	40.876551	-73.910660	pinochio pizza	5302a84a11d26be8acd6bd89	40.877868	-73.918948	Italian Restaurant
4	4	Marble Hill	40.876551	-73.910660	Nonno Tony's	4b784cd3f964a5206ac32ee3	40.885288	-73.909588	Italian Restaurant
5	5	Chinatown	40.715618	-73.994279	Bacaro	472a027af964a520ea4b1fe3	40.714468	-73.991589	Italian Restaurant
6	6	Chinatown	40.715618	-73.994279	Forsythia	5f21f95204091a2439ed691b	40.722550	-73.992485	Italian Restaurant

Figure 4: NYC, Manhattan Italian restaurants data frame

- The latitude and longitude for each Italian restaurant was then plotted onto a map of NYC using the Folium library. The output was as follows:

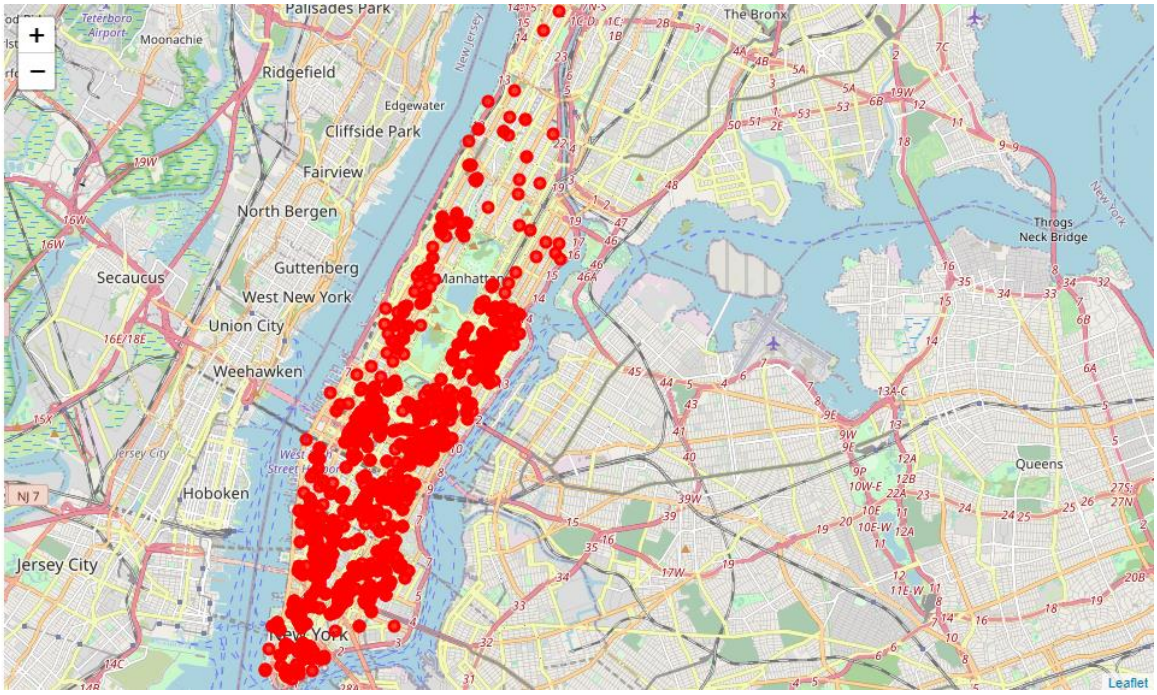

Figure 5:Map illustrating the distribution of Italian restaurants in NYC, Manhattan

3.5 Group the data set according to number of Italian restaurants per Manhattan Neighborhood

- The data set was grouped according to the number of Italian restaurants per neighborhood in the Manhattan borough.
- This allowed to visually identify the neighborhoods with the least and most number of Italian restaurant competitors.

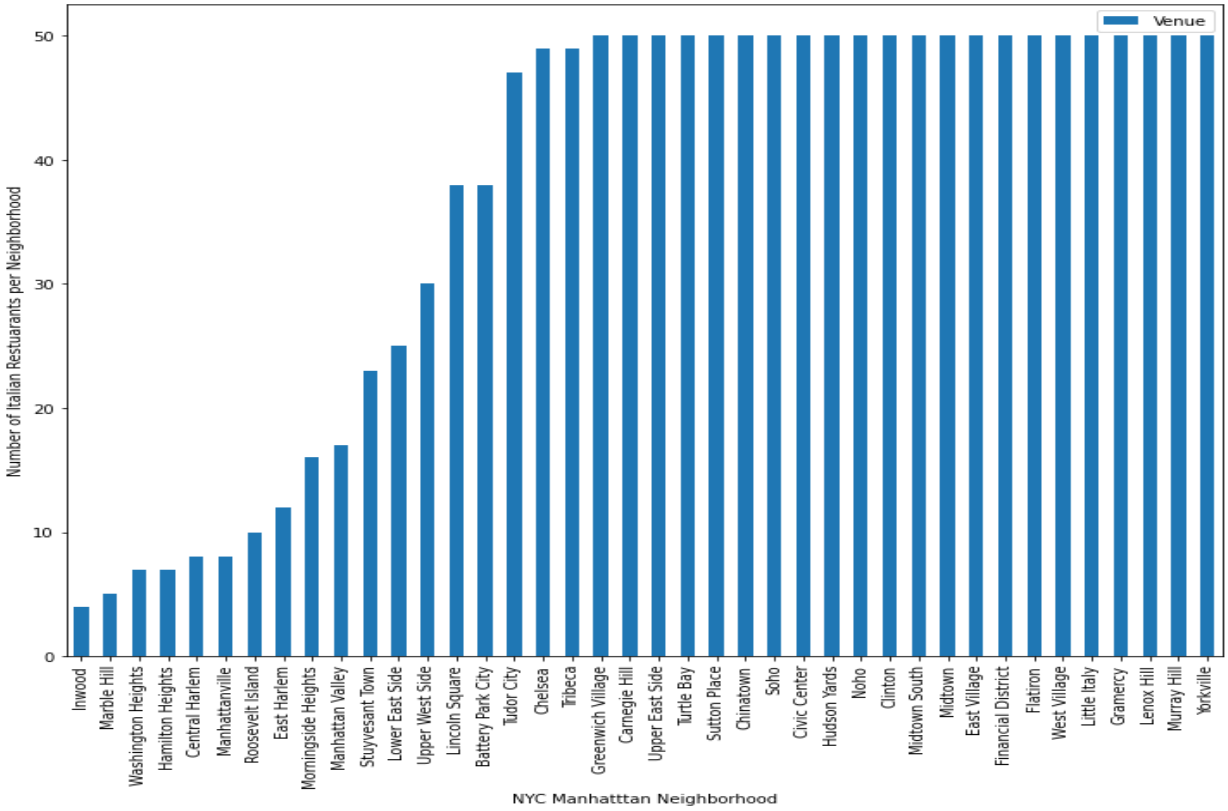

Figure 6: Grouping and count of Italian restaurants per NYC, Manhattan Neighborhood

3.6 Extract statistics for each of the Italian restaurants in the data set using the Foursquare API

- The Foursquare API was then used to extract statistics for each of the Italian restaurants. These stats included data for likes, ratings and tips for each Italian restaurant in the data frame.
- A function was created to loop through the entire data set and retrieve stats for each Italian restaurant based the restaurant's unique Foursquare ID as shown in Figure 7 below:

Unnamed: 0	Neighborhood	ID	Name
1110	Noho	4a790030f964a520cbe61fe3	Luzzo's
54	Chinatown	5ae016fc2faed54a007fbb89	Una Pizza Napoletana
1177	Midtown South	5b5b8ac6c21cb1002c540c6f	Cardoncello DiVino
653	Tribeca	4a149749f964a52055781fe3	Locanda Verde
1360	Tudor City	4afa15d8f964a520931622e3	Villa Berulia

Figure 7: Structure of data frame used to make Foursquare API requests for Italian restaurant's statistics

- The Foursquare API only allows for 500 premium calls for a free developer account. Therefore, the dataset was divided into three partitions (each containing < 500 rows) and each partition was analyzed individually. The results were then merged into a single dataset at the end.
- The extracted results had the following structure:

	Unnamed: 0	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude	Venue Category	ID	Name	Likes	Rating
0	1110	Noho	40.723259	-73.988434	40.730565	-73.983191	Pizza Place	4a790030f964a520cbe61fe3	Luzzo's	279	8.5
1	54	Chinatown	40.715618	-73.994279	40.721750	-73.988589	Pizza Place	5ae016fc2faed54a007fbb89	Una Pizza Napoletana	85	7.0
2	1177	Midtown South	40.748510	-73.988713	40.745090	-73.990090	Italian Restaurant	5b5b8ac6c21cb1002c540c6f	Cardoncello DiVino	10	7.5
3	653	Tribeca	40.721522	-74.010683	40.719981	-74.010002	Italian Restaurant	4a149749f964a52055781fe3	Locanda Verde	1434	9.1
4	1360	Tudor City	40.746917	-73.971219	40.746863	-73.980620	Italian Restaurant	4afa15d8f964a520931622e3	Villa Berulia	37	8.2

Figure 8: Structure of data frame containing restaurant stats

3.7 Apply the K-Means unsupervised learning algorithm to create clusters based on the likes and ratings of each Italian restaurant.

- The K-Means learning algorithm was then applied with a setting of 5 clusters.
- K-Means was utilized to group each Italian restaurant into a cluster so that we can categorize them and easily identify them as a strong or weak competitor.
- The cluster labels were then joined to the merged data.
- The cluster centroids in Figure 9 below were then analyzed to identify more information pertaining to each cluster:

Clustering Lables	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude	Likes	Rating
0	40.742723	-73.986828	40.743520	-73.987250	421.261682	8.529907
1	40.724783	-73.995446	40.728753	-73.993936	8129.000000	9.300000
2	40.746314	-73.984011	40.746860	-73.984290	71.165226	7.610855
3	40.741975	-73.988493	40.742978	-73.988784	979.592105	8.782895
4	40.731694	-73.990566	40.733081	-73.990482	2497.000000	9.300000

Figure 9: Clustering centroids obtained based on clustering of likes and rating data of all Italian Restaurants

- The NYC Manhattan restaurants were plotted on a map of NYC by their respective clusters as shown below:

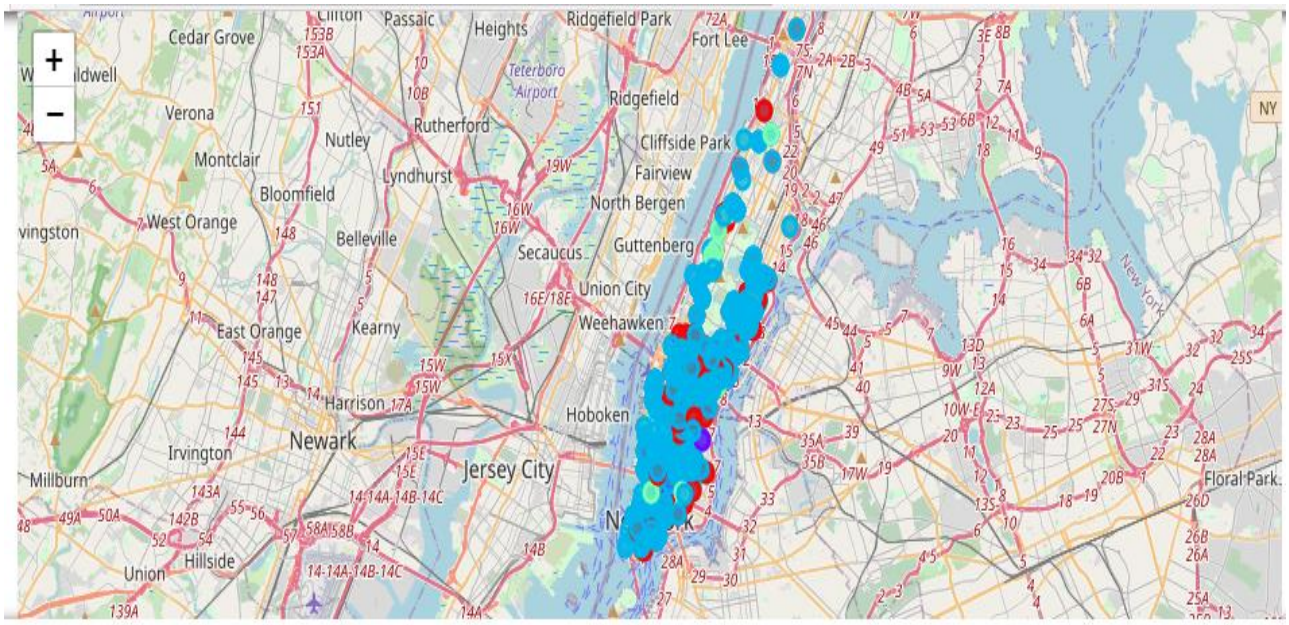

Figure 10: NYC, Manhattan map showing all Italian Restaurants as per their respective clusters

3.8 Group the Neighborhoods according to average Italian restaurant ratings

- The restaurant ratings data extracted in the previous step were then grouped by neighborhood to determine the neighborhoods with the highest and lowest average ratings.
- The reasoning is that a new restaurant should avoid opening in neighborhoods with high average ratings since these would represent high barriers for entry and high levels of competition. Rather the neighborhoods with low average ratings should be targeted since the levels of competition are low, thereby increasing the probability of success for a new restaurant. (The idea is that people would be keen to try a new Italian restaurant since the existing competition is not very good).

3.9 Choose the best neighborhood(s) to open a new Italian restaurant

- The next step in the methodology involved analyzing the data to find the neighborhoods with the lowest numbers of Italian restaurants and the lowest customer ratings. These neighborhoods would be chosen as the most ideal for a start up business.
- The chosen neighborhoods were then further analyzed by plotting existing competitors on a Folium map and identifying areas and street names that have little to no competition.
- These streets would represent the ideal locations for opening a new Italian restaurant.

4. Results

The results obtained from the methodology applied above are presented below:

4.1 Italian restaurants per Neighborhood

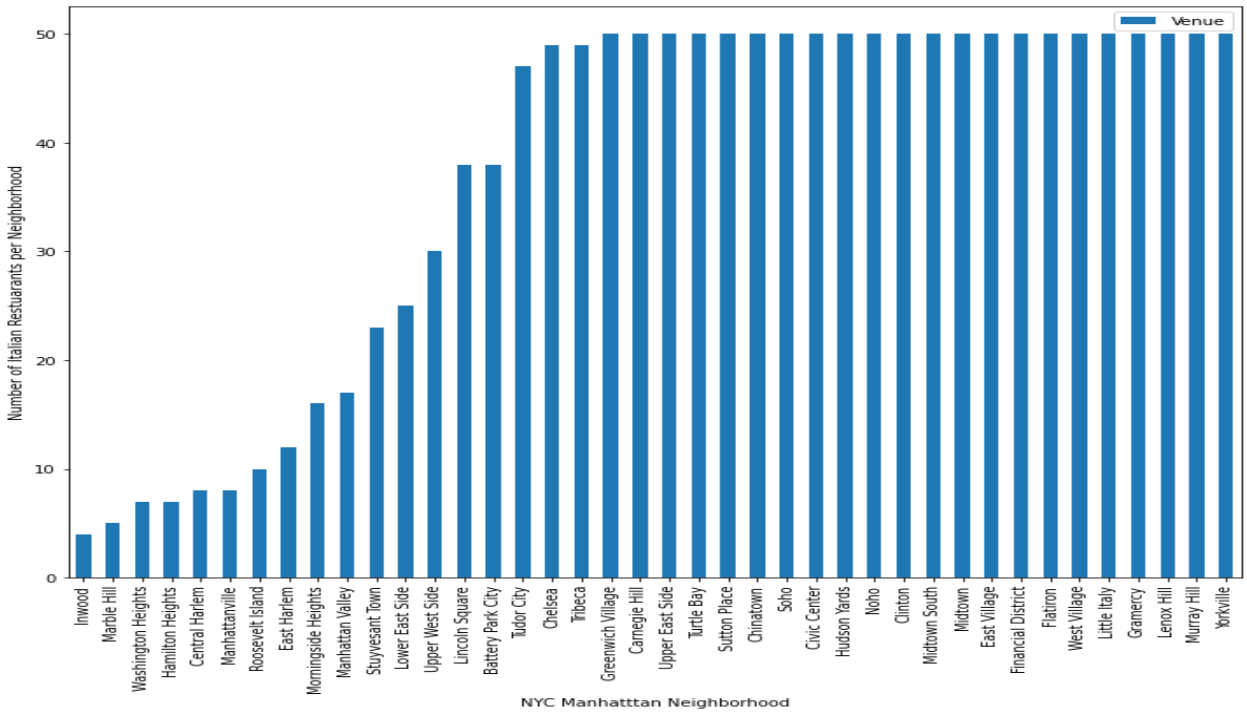

Figure 11: Number of Italian Restaurants per NYC Manhattan Neighborhood

Figure 11 reveals that the 5 Neighborhoods with the **least** number of Italian restaurants are:

1. Inwood
2. Marble Hill
3. Washington Heights
4. Hamilton Heights
5. Central Harlem

These represent the top 5 neighbours with the **least competition**.

4.2 NYC Manhattan Neighborhood average ratings

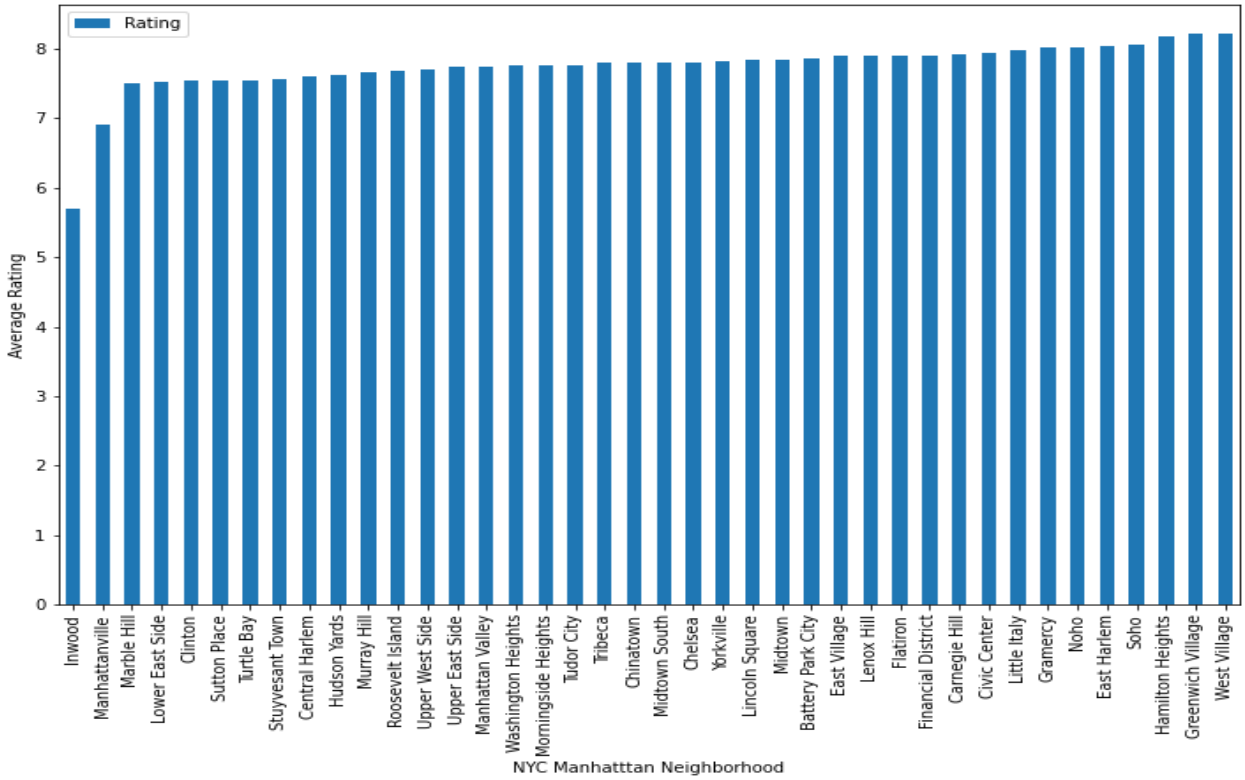

Figure 12: NYC, Manhattan average Italian Restaurant Neighborhood rating

Figure 12 reveals that the **5 lowest rated Neighborhoods** with respect to Italian restaurants are:

1. Inwood
2. Manhattanville
3. Marble Hill
4. Lower East Side
5. Clinton

4.3 Italian Restaurants clustering centroids

Clustering Lables	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude	Likes	Rating
0	40.742723	-73.986828	40.743520	-73.987250	421.261682	8.529907
1	40.724783	-73.995446	40.728753	-73.993936	8129.000000	9.300000
2	40.746314	-73.984011	40.746860	-73.984290	71.165226	7.610855
3	40.741975	-73.988493	40.742978	-73.988784	979.592105	8.782895
4	40.731694	-73.990566	40.733081	-73.990482	2497.000000	9.300000

Figure 13:Clustering centroids obtained based on clustering of likes and rating data of all Italian Restaurants

Analysis of the centroids in Figure 13 revealed the following labels/descriptors for each cluster:

1. Cluster 0 - Contains low likes & highly rated Italian restaurants
2. Cluster 1 - Contains highly liked & highly rated Italian restaurants
3. Cluster 2 - Contains low likes & lowly rated Italian restaurants
4. Cluster 3 - Contains medium liked and highly rated Italian restaurants
5. Cluster 4 - Contains highly liked and highly rated Italian restaurants

4.4 Highest and lowest rated Italian restaurants in NYC Manhattan

- Lowest rated Italian Restaurant in NYC, Manhattan:
  - Name: Pizza Shack
  - Rating: 4.8
  - Likes: 3.0
  - Neighborhood: Flatiron

	Clustering Lables	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude	Venue Category	ID	Name	Likes	Rating
611	2	Flatiron	40.739673	-73.990947	40.743401	-73.996019	Italian Restaurant	4aa5ae24f964a520254920e3	Pizza Shack	3.0	4.8
613	2	Clinton	40.759101	-73.996119	40.760100	-73.991378	Italian Restaurant	4d6e8e5d30d5b1f7a510c942	Abitino's Pizza & Cafe	1.0	5.3
169	2	Upper West Side	40.787658	-73.977059	40.783762	-73.977880	Italian Restaurant	4a4fec7df964a520d3af1fe3	Al Dente	10.0	5.3
1296	2	Midtown South	40.748510	-73.988713	40.742163	-73.989930	Italian Restaurant	4dee979db0fb43a756f7d7fe	Mezzogiorno	7.0	5.3
1302	2	Clinton	40.759101	-73.996119	40.755531	-73.994769	Italian Restaurant	4dee979db0fb43a756f7d7fe	Mezzogiorno	7.0	5.3

Figure 14: Five lowest rated Italian restaurants in NYC Manhattan (Low competition)

- Highest rated Italian Restaurant in NYC, Manhattan:
  - Name: L'Artusi
  - Rating: 9.4
  - Likes: 1172
  - Neighborhood: Chelsea, Little Italy

1306	1	Gramercy	40.737210	-73.981376	40.736031	-73.979160	Pizza Place	4c5ef77bfff99c74eda954d3	Eataly Flatiron	8129.0	9.3
1377	4	Carnegie Hill	40.782683	-73.953256	40.775019	-73.954046	Italian Restaurant	4cc6222106c25481d7a4a047	Rubirosa Ristorante	2497.0	9.3
994	4	Tribeca	40.721522	-74.010683	40.718935	-74.004940	Italian Restaurant	4cc6222106c25481d7a4a047	Rubirosa Ristorante	2497.0	9.3
929	3	Little Italy	40.719324	-73.997305	40.723117	-73.995791	Italian Restaurant	4a27289cf964a52062911fe3	L'Artusi	1172.0	9.4
1269	3	Chelsea	40.744035	-74.003116	40.741011	-74.004866	Italian Restaurant	4a27289cf964a52062911fe3	L'Artusi	1172.0	9.4

Figure 15: Five highest rated Italian Restaurants in NYC Manhattan (High Competition)

4.5 Best Neighbourhoods to open an Italian Restaurant

1. Inwood
2. Marble Hill

4.6 Results for best Neighborhoods

4.6.1 Inwood

Inwood Restaurants with ratings

	Clustering Lables	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude	Venue Category	ID	Name	Likes	Rating
953	2	Inwood	40.867684	-73.92121	40.86736	-73.917282	Italian Restaurant	4e473b97fa76a07fde65b0d8	La Trattoria	15.0	5.7

Figure 16: Inwood rated Italian Restaurants

Inwood Restaurants without ratings

	Unnamed: 0	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category
62	62	Inwood	40.867684	-73.92121	Il Sole	4310fe80f964a52050271fe3	40.866015	-73.927821	Italian Restaurant
63	63	Inwood	40.867684	-73.92121	Tenth Avenue NY | Rooftop Bar & Eatery	5cc451ce033693002ca2c39d	40.862050	-73.920470	Cocktail Bar
64	64	Inwood	40.867684	-73.92121	Viale NYC	5775f3e1498e58df24b9ee11	40.867360	-73.917282	Italian Restaurant
65	65	Inwood	40.867684	-73.92121	Trattoria Inwood	52cdfb4011d2f6cf0535f2d2	40.868416	-73.918647	Italian Restaurant

Figure 17: All Inwood Italian Restaurants (Including rated and non-rated)

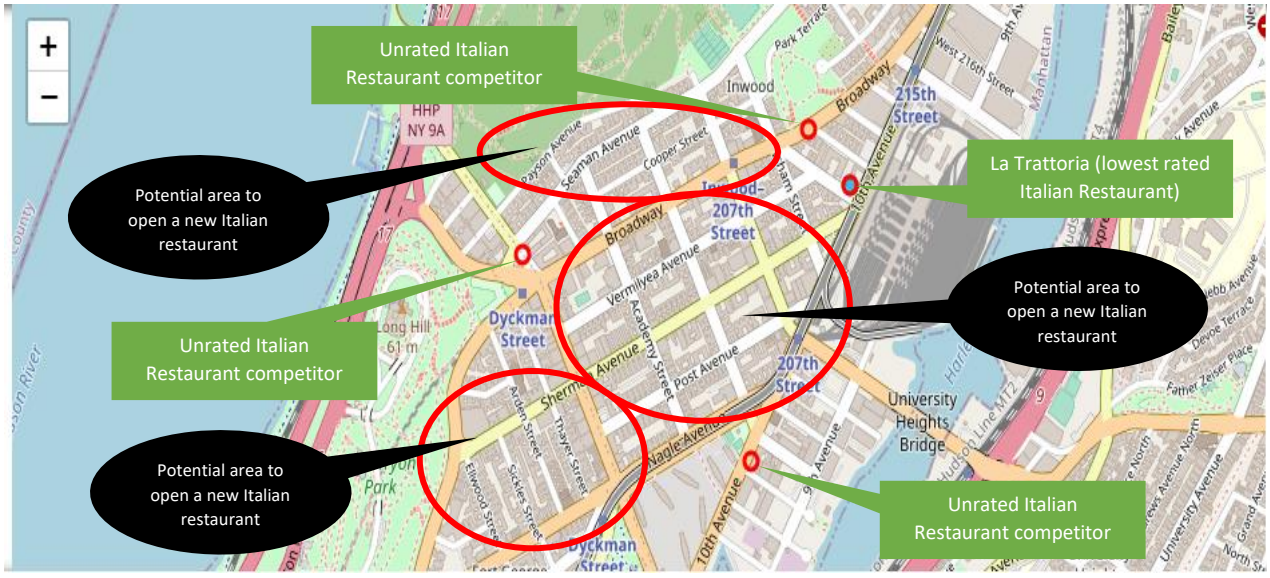

Figure 18: Map of NYC, Manhattan - Inwood showing existing Italian restaurants and potential locations for a new Italian restaurant

4.6.2 Marble Hill

Marble Hill Restaurants with ratings

	Clustering Lables	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue Latitude	Venue Longitude	Venue Category	ID	Name	Likes	Rating
86	2	Marble Hill	40.876551	-73.91066	40.878822	-73.904494	Pizza Place	4be72770910020a16f1ad514	Broadway Pizza & Pasta	12.0	7.4
584	2	Marble Hill	40.876551	-73.91066	40.878822	-73.904494	Pizza Place	49f1d030f964a520b9691fe3	Cipriani Dolci	114.0	6.7
763	2	Marble Hill	40.876551	-73.91066	40.885288	-73.909588	Italian Restaurant	57dad81f498e43ee1c4a75c4	Vago Restaurant	9.0	6.9
1082	2	Marble Hill	40.876551	-73.91066	40.878822	-73.904494	Pizza Place	4d840360d5fab60c7b25d99b	Bar Italia	108.0	8.3
1261	2	Marble Hill	40.876551	-73.91066	40.885288	-73.909588	Italian Restaurant	54ab15df498ee7d7723bca15	Adoro Lei	113.0	8.2

Figure 19: Marble Hill rated Italian Restaurants

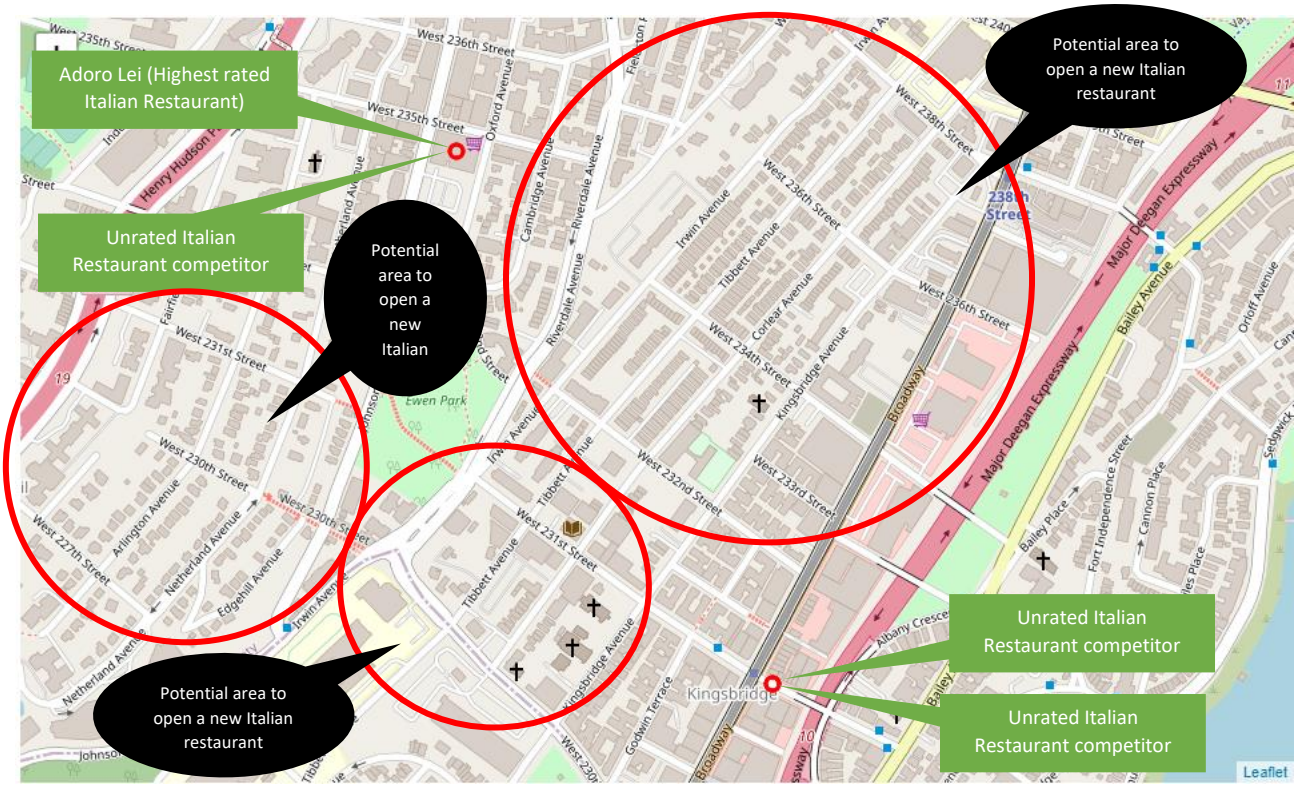

Figure 20: Map of NYC, Manhattan – Marble Hill showing existing Italian restaurants and potential locations for a new Italian restaurant

- Lowest rated Italian Restaurant in NYC, Manhattan – Marble Hill:
  - Name: Cipiriani Dolci
  - Rating: 6.7
  - Likes: 114
  - Neighborhood: Marble Hill
  -
- Highest rated Italian Restaurant in NYC, Manhattan – Marble Hill:
  - Name: Bar Italia
  - Rating: 8.3
  - Likes: 108
  - Neighborhood: Marble Hill

Answers to the problem questions:

1. Which borough in NYC would be the best borough to open an Italian restaurant? (Based on least competition)
   Answer : Manhattan

2. Which location has the highest number of Italian restaurants?
   Answer: The following Neighborhoods each have the highest number of Italian restaurants (50 restaurants):
   - *i. Greenwich Village*
   - *ii. Carnegie Hill*
   - *iii. Upper East Side*
   - *iv. Turtle Bay*
   - *v. Sutton Place*
   - *vi. Chinatown*
   - *vii. Soho*
   - *viii. Civic Center*
   - *ix. Hudson Yards*
   - *x. Noho*
   - *xi. Clinton*
   - *xii. Midtown South*
   - *xiii. Midtown*
   - *xiv. East Village*
   - *xv. Financial District*
   - *xvi. Flatiron*
   - *xvii. West Village*
   - *xviii. Little Italy*
   - *xix. Gramercy*
   - *xx. Lenox Hill*
   - *xxi. Murray Hill*
   - *xxii. Yorkville*

3. Which Neighborhood(s) would be the best Neighborhood to open an Italian restaurant? (Based on least competition)
   Answer: Inwood and Marble Hill

4. Who would be the top 5 most competitive Italian restaurants in NYC that Raymond would have to compete with? (Based on customer ratings)
   Answer: Top 5 rates Italian restaurants in NYC, Manhattan are (in descending order):
   - *i. L'Artusi (Rating = 9.4)*
   - *ii. Rubirosa Ristorante (Rating = 9.3)*
   - *iii. Eataly Flatiron (Rating = 9.3)*
   - *iv. Via Carota (Rating = 9.3)*
   - *v. Ristorante (Rating = 9.3)*

5. Which Italian restaurant would represent the lowest level of competition?
   Answer: The Italian Restaurant in NYC, Manhattan that represents the lowest level of competition (based on customer ratings is):

Pizza Shack (Rating = 4.8)

6. What would be the best locations (street names) to open an Italian restaurant?
   The following are the best proposed locations for opening an Italian restaurant within the best identified neighborhoods:

   6.1 Inwood best proposed locations:

   i. *Broadway*
   ii. *Vermilyea Avenue*
   iii. *Post Avenue*
   iv. *Academy Street*
   v. *Payson Avenue*
   vi. *Seaman Avenue*
   vii. *Cooper Street*
   viii. *Arden Street*
   ix. *Thayer Street*
   x. *Sickles Street*
   xi. *Ellwood Street*
   xii. *Sherman Avenue*

   6.2 Marble Hill best proposed locations:

   I. *West 236th Street*
   II. *West 238th Street*
   III. *Cambridge Avenue*
   IV. *Irwin Avenue*
   V. *Tibbett Avenue*
   VI. *Corlear Avenue*
   VII. *Kingsbridge Avenue*
   VIII. *West 234th Street*
   IX. *West 233rd Street*
   X. *Werst 232nd Street*
   XI. *West 231st Street*
   XII. *West 230th Street*
   XIII. *Arlington Avenue*
   XIV. *West 227th Street*
   XV. *Edgehill Avenue*
   XVI. *Netherland Avenue*

5 . Discussion

The results presented above provide guidance on the best location(s) for an entrepreneur like Raymond to open an Italian restaurant within New York City.

An analysis of the New York City location data revealed that Manhattan is the smallest Borough thereby suggesting that this would be the best brough to open a new restaurant. This is motivated by the fact that a restaurant opening in a smaller location would have a higher chance of being noticed/ recognised or being identified by search engines such as Google or Bing. Further research confirmed that Manhattan is the best borough due to it being highly rated for dining and nightlife (Extra Space Storage, 2016). For these reasons the project scope was limited to only evaluate Manhattan borough in New York City.

The Manhattan borough was then further analysed by extracting details for all Italian restaurants within the location using the Foursquare API. This resulted in a pandas data frame with 1493 rows indicating that there are at least 1493 Italian restaurants in Manhattan. The next step involved analysing the data extracted from Foursquare to identify the top 5 NYC Manhattan neighborhoods with the **least** number of Italian restaurants. Figure 11 revealed that the Inwood, Marble Hill, Washinton Heights, Hamilton Heights and Central Harlem neighborhoods have the least number of Italian restaurants. These represent the neighborhoods with the least number of competitors, thereby suggesting possible locations for opening a new Italian restaurant. This is based on the premise that a new business owner would have a high chance of succeeding in an area that has fewer competitors.

Once the top 5 areas with the least number of competitors were identified, it was required to determine which areas have Italian restaurants that are highly and poorly rated by customers.

Figure 12 depicts the average ratings per neighborhood in ascending order. From this it was determined that the five neighborhoods with the **highest rated** Italian restaurants were West Village, Greenwich Village, Hamilton Heights, Soho and East Harlem. These neighborhoods represent the highest levels of competition and it is advised to avoid opening a new Italian restaurant in these neighborhoods. Figure 12 also revealed that the five neighborhoods with the **lowest rated** Italian restaurants were Inwood, Manhattanville, Marble Hill, Lower East Side and Clinton with Inwood having the lowest average customer ratings. These neighborhoods represent the areas with the lowest levels of competition and it would therefore be advised to open a new restaurant in one of these neighborhoods.

It was determined that it would also be beneficial to identify the highest and lowest rated Italian restaurants in the entire Manhattan borough. This information can be useful since an entrepreneur could visit the highest and lowest rated Italian restaurants and get and idea of what to do and what not to do respectively. By analysing the dataset containing the restaurants statistics obtained from the Foursquare API, it was found that L'Artusi in Chelsea and Little Italy have the highest customer ratings of 9,4. Raymond could potentially visit this franchise and evaluate his largest competitor to better understand what makes L'Artusi so successful. Raymond could also visit Pizza Shack in Flatiron which was the lowest rated Italian restaurant in Manhattan. Further analysing this weak competitor would provide Raymond with an idea of what to avoid in his restaurant that is ultimately resulting in low customer ratings. This information does not necessarily assist in identifying the best location; however it would assist an entrepreneur in determining the factors of success and failure for existing competitors.

The next step in the analysis involved using the K-Means clustering algorithm to group the NYC, Manhattan Italian restaurants according to the number of likes and customer ratings. The algorithm was run with a setting of 5 clusters. The centroids as per Figure 13 for each cluster was then further analysed to draw conclusions about descriptions for each cluster. Cluster 0 was found to contain all Italian restaurants with a low number of likes, but high customer ratings. Cluster 1 contained Italian restaurants that have high numbers of likes and high customer ratings . Cluster 2 represented Italian restaurants with a low number of likes and relatively low customer ratings. Cluster 3 was found to contain Italian restaurants with a medium number of likes and high customer ratings. Finally, cluster 4 was found to contain Italian restaurants with a high number of likes and high customer ratings. The clustering will assist in further determining areas or locations to avoid. From the K-Means clustering, it is obvious that an entrepreneur must avoid opening a new Italian restaurant in areas that contains many restaurants belonging to Cluster 0, Cluster 1, Cluster 3 and Cluster 4. This leaves only Cluster 2 which contains the lowly liked and low rated Italian restaurants. Therefore, it would be advised to open a new Italian restaurant in an area that contains a high number of Cluster 2 restaurants since this would increase the probability of success since competition is not very good.

Based on the results discussed above, a deeper analysis revealed that Inwood and Marble Hills would be the best neighborhoods to open a new Italian Restaurant. This is because both Inwood and Marble Hill are neighborhoods with the lowest numbers of Italian restaurants that also have the lowest average customer ratings. These indicate that there are low levels of competition in these neighborhoods and that customers will be more inclined to trying a new restaurant as opposed to the existing poorly rated restaurants.

Figures 16 and 19 revealed that Inwood and Marble Hill both contain Italian restaurants that are within Cluster 2. This confirms that these neighborhoods have low levels of competition with respect to Italian cuisine.

Figure 18 geographically represents the distribution of all Italian restaurants in the Inwood neighborhood. Only one out of the four Italian restaurants had statistics obtained from the Foursquare API. The four markers indicate the locations of existing Italian restaurants that should be avoided since one would not want to open a new restaurant too close to a competitor (even if they are poorly rated). Therefore, the areas within the red circles represent potential locations for opening a new Italian restaurant that contains low levels of competition. From this analysis there are approximately twelve (12) streets within Inwood that can be used for opening a new Italian restaurant. These include Broadway, Post Avenue, Academy Street, Payson Avenue, Seaman Avenue, Cooper Street, Arden Street, Thayer Street, Sickles Street, Ellwood Street and Sherman Avenue.

Figure 20 similarly represents the distribution of all Italian restaurants in the Marble Hill neighborhood. Analysis of the restaurant statistics obtained from the Foursquare API revealed that the five rated Italian restaurants were very close to one another. Therefore, an entrepreneur like Raymond should avoid opening a new Italian restaurant that is in close proximity to these established competitors. The red circled areas on Figure 20 represents the potential locations in which Raymond can open a new Italian restaurant and avoid having high levels of competition. This revealed sixteen (16) potential streets in which Raymond can open his business. This includes West 236th Street, West 238th Street, Cambridge Avenue, Irwin Avenue, Tibbett Avenue, Corlear Avenue, Kingsbridge Avenue, West 234th Street, West 233rd Street, West 232nd Street, West 231st Street, West 230th Street, Arlington Avenue and Arlington Avenue.

6. Conclusion

From the analysis above it can be concluded that Raymond or any other entrepreneur wishing to open an Italian restaurant in NYC would have the highest chances of success by opening a restaurant in the Manhattan Borough. Further analysis revealed that the Inwood and Marble Hill neighborhoods are the best choices since these neighborhoods have the lowest levels of competition, thereby increasing the probability of success for a start up business.

It was determined that there are 12 and 16 possible streets within Inwood and Marble Hill respectively that Raymond can use for opening his restaurant. This affords Raymond the opportunity of choosing a neighborhood and subsequent street that is most affordable in terms of rent, the most safe and perhaps the one that is most closely situated to his place of residence (shorter commutes to work).

And thus, the battle of neighborhoods was concluded with Inwood and Marble Hill emerging as the victors.

References

Extra Space Storage. (2016, December 14). *Find The Best Borough in New York City For You*. Retrieved from ExtraSpace Storage: https://www.extraspace.com/blog/moving/city-guides/best-boroughs-nyc/

Financial, F. (2019, March 11). *5 Reasons to Open a Business in New York*. Retrieved from Fora Financial: https://www.forafinancial.com/blog/small-business/open-business-in-new-york/#:~:text=CNBC%20recently%20named%20New%20York,industries%20like%20IT%20and%20entertainment.