The Battle of the Neighborhoods - Week 1 and Week 2

Introduction & Business Problem:

Problem 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 is a dream place for gourmet to seek delicious cuisine. Its food culture includes an array of international cuisines influenced by the city's immigrant history.

As it is highly developed city so cost of doing business is also one of the highest. Thus, the market is highly competitive, any new business venture or expansion needs to be analysed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market.

Problem Description:

New York City attracts many to start their business in the food industry. It's food culture includes an array of international cuisines influenced by the city's immigrant history.

Central and Eastern European immigrants, especially Jewish immigrants introduced bagels, cheesecake, hot dogs to New York. Italian immigrants brought New York-style pizza and Italian cuisine. Jewish immigrants and Irish immigrants is famoused by making pastrami and corned beef. Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city. 4,000 license were issued to mobile food vendors by the city. Middle Eastern foods such as falafel and kebabs examples of modern New York street food. It is famous for not just Pizzerias, Cafe's but also for fine dining Michelin starred restaurants.The city is home to "nearly one thousand of the finest and most diverse haute cuisine restaurants in the world", according to Michelin. it is impractical to run an analysis for each type of restaurant. Based on the maximum total numbers among these restaurants, I choose Pizza Place for the following report. The study of other types of restaurants can be conducted with the same method.

Data Preparation

Scope of this project To recommend the correct location for Company A to locate and recommend to the management which neighborhood of New York city will be best choice to start a restaurant. The Management also expects to understand the rationale of the recommendations made.

Success Criteria: The success criteria of the project will be a good recommendation of borough/Neighborhood choice to Foodie Company based on Lack of such restaurants in that location and nearest suppliers of ingredients.

We will be using the below datasets to resturant business in New York city

Data 1 : Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segement the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs

and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood.
This dataset exists for free on the web. Link to the dataset is : https://geo.nyu.edu/catalog/nyu_2451_34572

Data 2 : Second data which will be used is the DOHMH Farmers Markets and Food Boxes dataset. In this we will be using the data of Farmers Markets.
https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2
Website-https://www.grownyc.org/greenmarketco/foodbox

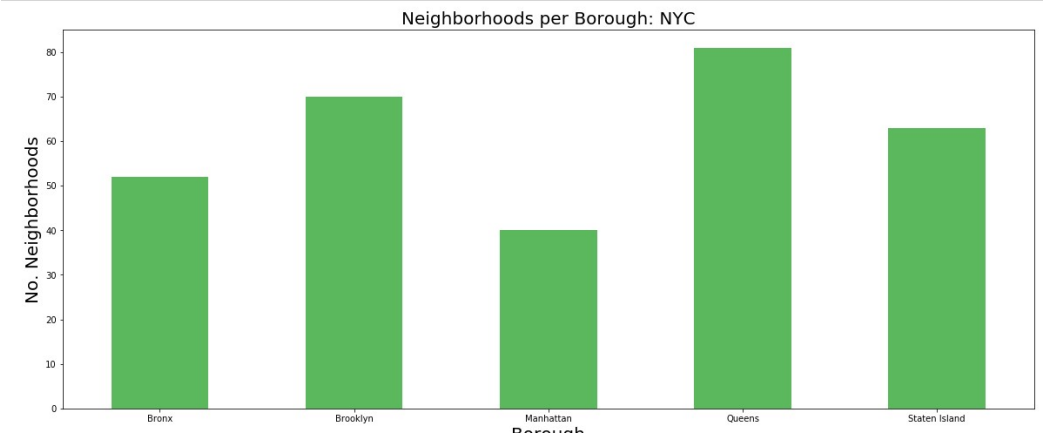
Data 3 : For the below analysis we will get data from wikipedia as given below :
New York Population New York City Demographics Cuisine of New York city https://en.wikipedia.org/wiki/New_York_City https://en.wikipedia.org/wiki/Economy_of_New_York_City https://en.wikipedia.org/wiki/Portal:New_York_City https://en.wikipedia.org/wiki/Cuisine_of_New_York_City https://en.wikipedia.org/wiki/List_of_Michelin_starred_restaurants_in_New_York_City

Data 4 : New York city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood.We will use the Foursquare API to explore neighborhoods in New York City. The below is image of the Foursquare API data.

ues' record of given venues of New York City.

Methodology

I used **BeautifulSoup** to scrape boroughs from Wikipedia and organize a table containing Community Board, Area, Pop.Census, Neighborhoods information of New York City.And I used **Geopy** to get the geological location of each community board. The first result returned was the number of Italian restaurants in one area. As shows in below, Queens has most Italian restaurants. However, in later study, we discovered Queens is not our best option.



I utilized the **Foursquare API** to explore the boroughs and segment them. I designed the limit as 100 venues and the radius 500 meters for each borough from their given latitude and longitude information. Here is the header of the result, adding venue id, venue name, category, latitude, and longitude information from Foursquare API.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Rite Aid	40.896521	-73.844680	Pharmacy
2	Wakefield	40.894705	-73.847201	Cooler Runnings Jamaican Restaurant Inc	40.898283	-73.850478	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
4	Wakefield	40.894705	-73.847201	Dunkin Donuts	40.890631	-73.849027	Donut Shop

	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

I utilized the **Foursquare API** again by Italian restaurants places ID to explore the detailed record of these Italian restaurants places. Select out Rating, Price, Likes, Photos, Tips into a data frame. And drop those places without a rating.

	Neighborhood	Average Rating
86	Tribeca	9.200000
11	Boerum Hill	9.200000
12	Bushwick	9.200000
44	Hamilton Heights	9.000000
37	Fulton Ferry	8.900000
54	Lower East Side	8.900000
28	Dumbo	8.900000
64	Noho	8.766667
43	Greenwich Village	8.707143
80	Soho	8.700000

Then I tried to find correlations among these variables:

	Rating	Price	Likes	Photos	Tips
Rating	1.000000	0.468022	0.493203	0.450074	0.510075
Price	0.468022	1.000000	0.624174	0.528872	0.713245
Likes	0.493203	0.624174	1.000000	0.868937	0.892418
Photos	0.450074	0.528872	0.868937	1.000000	0.829932
Tips	0.510075	0.713245	0.892418	0.829932	1.000000

Showing from the correlation matrix, Likes, Photos, and Tips are highly correlated with each other. But Likes is not highly related to Rating. Customers who click Likes for some specific reasons but give lower ratings to the general performance might cause this low correlation. Therefore I choose Rating to represent the restaurant. Rating is somewhat correlated to Price, which indicates that the price might not affect impressions of customers in that place significantly.

I utilized the **Foursquare API** centering these Italian restaurants places to explore their neighborhoods with a 500-meter radius.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	id	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Linda's Pizzeria	40.823803	-73.909060	5956be26123a195de6701c2b	Porto Salvo	40.823887	-73.912910	Italian Restaurant
1	Linda's Pizzeria	40.823803	-73.909060	4fa534cee4b0fed4819dc7d4	Perry Coffee Shop	40.823433	-73.910940	Diner
2	Linda's Pizzeria	40.823803	-73.909060	4f0f4fe14fc6e3adf353a950	Blink Fitness	40.819543	-73.910554	Gym / Fitness Center
3	Linda's Pizzeria	40.823803	-73.909060	5591837d498ee4167d48bb73	Cinco de Mayo	40.822574	-73.911592	Mexican Restaurant
4	Linda's Pizzeria	40.823803	-73.909060	553980e0498e6809906f65c	Old Bronx Courthouse	40.822894	-73.909565	Art Gallery
5	Linda's Pizzeria	40.823803	-73.909060	542c31e1498e76106c273492	Blink Fitness St Ann's	40.8[illegible]	-73.9[illegible]	Gym
6	Linda's Pizzeria	40.823803	-73.909060	4d38c0949ca8236a1b12aee8	Popeyes Louisiana Kitchen	40.824505	-73.909819	Fried Chicken Joint

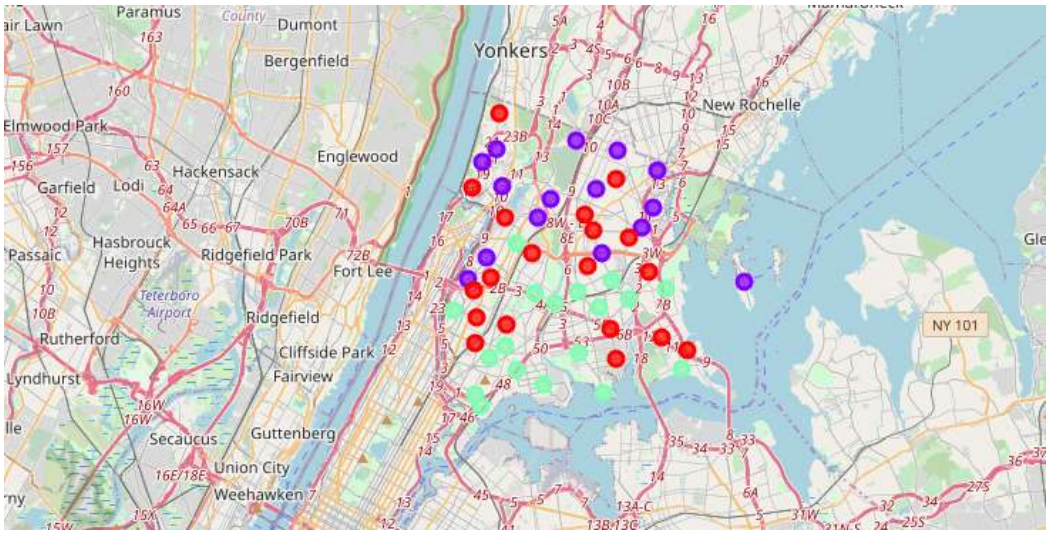
According to venue categories and numbers surrounding each Italian restaurants house, I use **k-means** to cluster Italian restaurants houses into several groups.

	Neighborhood	Latitude	Longitude	Count	Average Rating	High Rated Total	Mid Rated Total	Low Rated Total	Cluster Labels
0	Wakefield	40.894705	-73.847201	1257.0	4.22	331.0	375.0	551.0	1
1	Co-op City	40.874294	-73.829939	1260.0	4.60	393.0	386.0	481.0	1
2	Eastchester	40.887556	-73.827806	1261.0	4.51	374.0	376.0	511.0	1
3	Fieldston	40.895437	-73.905643	1270.0	4.57	407.0	350.0	513.0	1
4	Riverdale	40.890834	-73.912585	1281.0	4.49	387.0	371.0	523.0	1

I merged cluster labels of each Italian restaurants place with its geological location.

	PizzaPlace	ClusterLabels	id	Rating	Price	Likes	Photos	Tips	Venue Latitude	Venue Longitude
90	Sottocasa Pizzeria - Harlem	2	56e86a57cd1017cb53f3e8f9	9.4	1	87	39	29	40.805550	-73.947435
88	Saraghina	1	4a593de0f964a52015b91fe3	9.2	2	736	432	204	40.683590	-73.935340
63	Nick's Pizza	1	45ac11b0f964a5205b411fe3	9.1	2	191	143	94	40.718180	-73.840692
98	Zero Otto Nove	4	4aabd3e6f964a5204a5a20e3	9.1	3	231	144	76	40.854714	-73.888388
79	Posto	0	43c4bb46f964a5205a2d1fe3	9.0	2	346	146	121	40.734737	-73.983049
100	babbalucci	2	55a4456a498e29417f5c71a2	9.0	2	92	73	30	40.808875	-73.944796
47	La Villa Pizzeria	1	4b193722f964a52056d923e3	8.9	2	75	42	40	40.616867	-73.909858
73	Peppino's	1	4b58f8f2f964a5203f7628e3	8.8	1	32	24	16	40.629851	-74.028481
93	Vesuvio Pizzeria & Restaurant	0	4b3eaeff964a520bfa025e3	8.8	2	105	80	149	40.632580	-74.027096
23	F&F Pizzeria	0	5d56bcea6ac2400008de14d9	8.8	1	15	13	0	40.577314	-73.981112
27	Full Moon Pizzeria	4	4aee1220f964a520ced121e3	8.7	2	114	95	42	40.855506	-73.887557

Then I used **folium** to visualize the distribution of these Italian restaurants in NYC as below:



Conclusion

The Bronx has the least number of Italian restaurants per borough. However, of note, Belmont of The Bronx is the neighborhood in all of NYC with the most Italian Restaurants. After studying the Bronx's neighborhoods for Italian restaurants for our client.