The Battles of the Neighborhoods

1. Introduction:

1.1. 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 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 so 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 Description:

A restaurant is a business which prepares and serves food and drink to customers in return for money, either paid before the meal, after the meal, or with an open account. The City of New York is famous for its excellent cuisine. Its food culture includes an array of international cuisines influenced by the city's immigrant history.

Central and Eastern European immigrants, especially Jewish immigrants - bagels, cheesecake, hot dogs, knishes, and delicatessens
Italian immigrants - New York-style pizza and Italian cuisine
Jewish immigrants and Irish immigrants - pastrami and corned beef
Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city
mobile food vendors - Some 4,000 licensed 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.
So, it is evident that to survive in such competitive market it is very important to strategically plan. Various factors need to be studied in order to decide on the Location such as:

New York Population
New York City Demographics
Are there any Farmers Markets, Wholesale markets etc. nearby so that the ingredients can be purchased fresh to maintain quality and cost?
Are there any venues like Gyms, Entertainment zones, Parks? nearby where floating population is high etc.
Who are the competitors in that location?
Cuisine served / Menu of the competitors
Segmentation of the Borough
Untapped markets
Saturated markets
The list can go on...
Although well-funded XYZ Company Ltd. need to choose the correct location to start its first venture. If this is successful they can replicate the same in other locations. First move is very important, thereby choice of location is very important.

1.3. Objective:

The objective is 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.

This would interest anyone who wants to start a new restaurant in New York city.

1.4. Success Criteria:

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

2. Data

2.1. Data 1: Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment 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 latitude and longitude coordinates of each neighborhood.
Link : https://geo.nyu.edu/catalog/nyu_2451_34572

	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

2.2. Data 2: Second data which will be used is the DOHMH Farmers Markets dataset.
Link: https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2

	Borough	Market Name	Street Address	Latitude	Longitude	Days of Operation	Hours of Operations	Season Dates	Accepts EBT	Open Year-Round	Stellar Cooking Demonstrations	Food Activities for Kids	Location Point
0	Brooklyn	Woodhull Hospital Youthmarket	Broadway & Flushing Ave	40.700726	-73.941932	Wednesday	9 a.m. - 2 p.m.	07/10/2019-11/27/2019	Yes	No	No	No	(40.700726, -73.941932)
1	Manhattan	Mount Sinai Hospital Greenmarket	E 99th St bet Madison & Park Aves	40.789169	-73.952743	Wednesday	8 a.m. - 5 p.m.	06/12/19-11/27/19	Yes	No	No	No	(40.789169, -73.952743)
2	Bronx	170 Farm Stand	E 170th St & Townsend Ave	40.839882	-73.916783	Wednesday	2:30 - 6:30 p.m.	07/10/2019-11/27/2019	Yes	No	No	Yes	(40.839882, -73.916783)
3	Manhattan	Greenmarket at Oculus Plaza	Church & Fulton Sts, on Oculus Plaza	40.711535	-74.010464	Tuesday	7 a.m. - 7 p.m.	07/09/2019-11/30/19	Yes	Yes	No	No	(40.711535, -74.010464)
4	Queens	Ditmars Park Youthmarket	Steinway St bet Ditmars Blvd & 23rd Ave, at Di...	40.772854	-73.906061	Saturday	8 a.m. - 3 p.m.	07/13/2019-11/23/2019	Yes	No	No	No	(40.772854, -73.906061)

2.3. Data 3: We use the following data from Wikipedia:

New York Population

New York City Demographics

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

	Borough	County	Estimate 2020	square_miles	square_km	persons_sq_mi	persons_sq_km
0	The Bronx	Bronx	1,432,132	42.10	109.04	34,653	13,231
1	Brooklyn	Kings	2,582,830	70.82	183.42	37,137	14,649
2	Manhattan	New York	1,628,701	22.83	59.13	72,033	27,826
3	Queens	Queens	2,278,906	108.53	281.09	21,460	8,354
4	Staten Island	Richmond	476,179	58.37	151.18	8,112	3,132
5		City of New York	8,398,748	842.343	783.83	28,188	10,947
6		State of New York	19,745,289	1,701.399	122,284	416.4	159

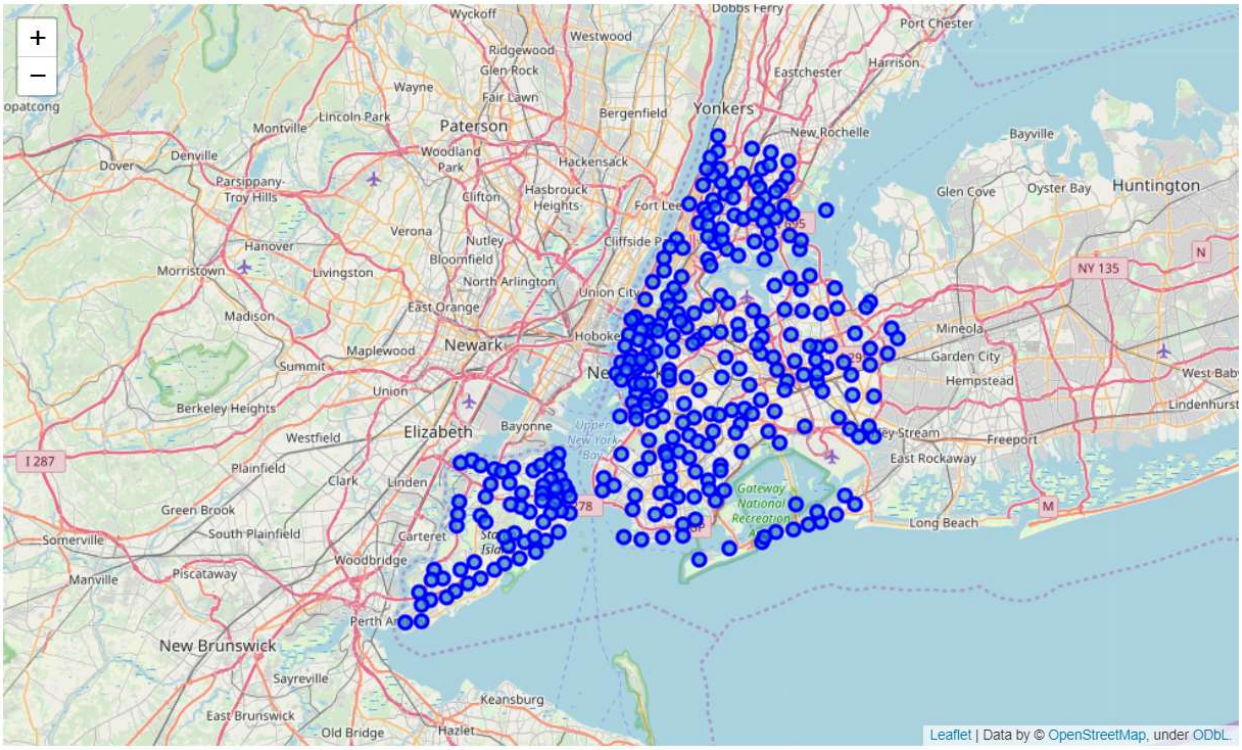
2.4. Data 4: New York city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in New York City.

	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Sam's Pizza	40.879435	-73.905859	Pizza Place
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop

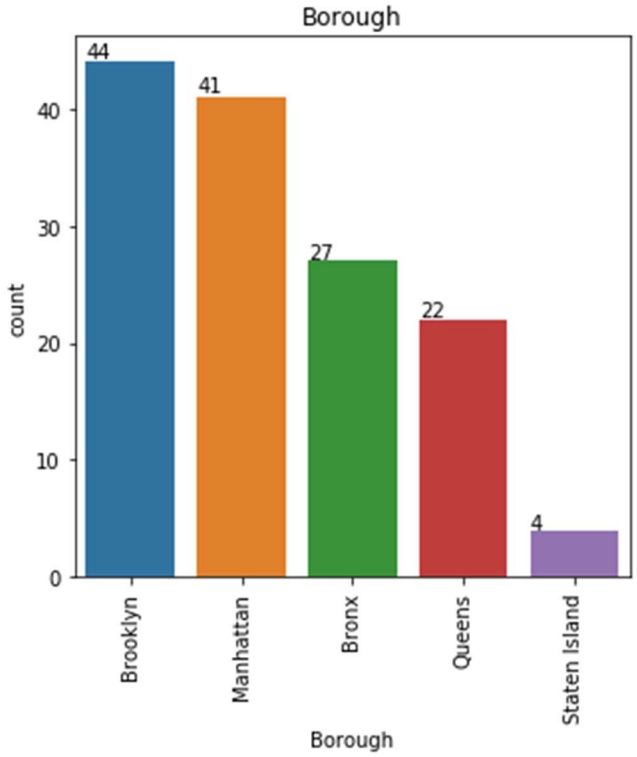
3. Data Analysis

3.1. Data 1: New York city Geographical Coordinates Data.

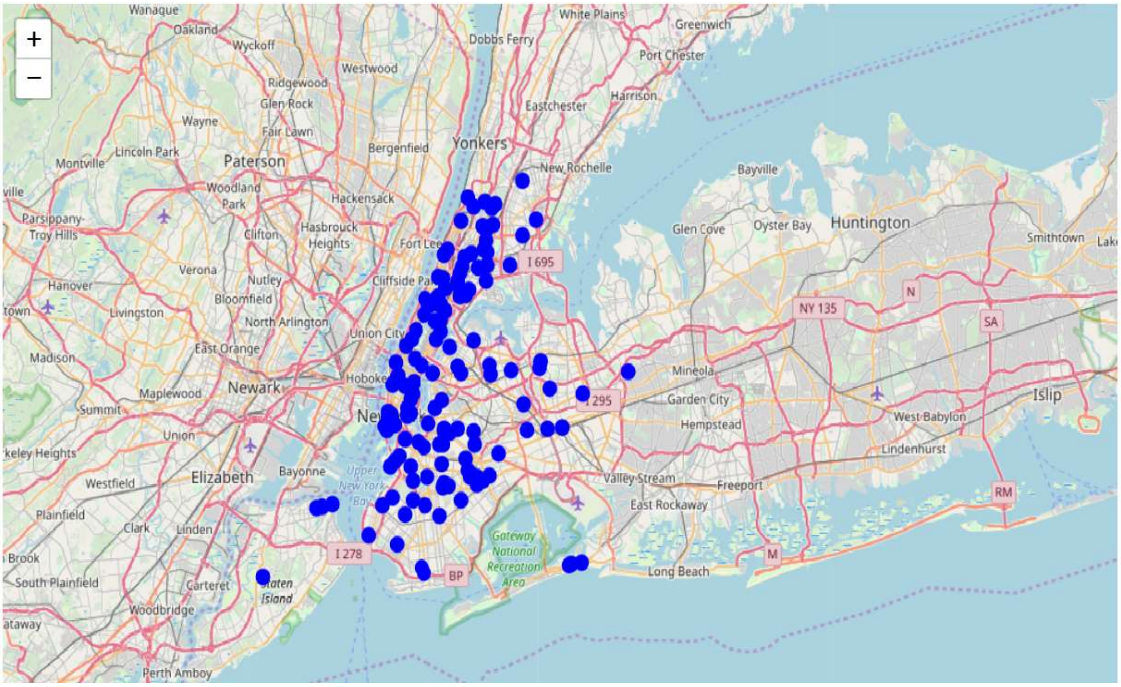
1. In this we load the data and explore data from New York data json file.
2. Transform the data of nested python dictionaries into a pandas' data frame.
3. This data frame contains the geographical coordinates of New York city neighborhoods.
4. This data will used to get Venues data from Four square.
5. We used geopy and folium libraries to create a map of New York city with neighborhoods superimposed on top.



3.2. Data 2: Second data which is used is the DOHMH Farmers Markets and Food Boxes dataset. In this we will be using the data of Farmers Markets data. There are totally 144 Farmers Markets in New York city. Highest number are in Manhattan and Brooklyn. And lowest in Queens, Bronx and Staten Island.



We used geopy and folium libraries to create a map to visualise farmers markets of New York city.



3.3. Data 3: To analyze New York city Population, scrapped the data from Wikipedia pages given above in the data section. We used BeautifulSoup python library. Beautiful Soup is a Python package for parsing HTML and XML documents (including having malformed markup, i.e. non-closed tags, so named after tag soup). It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping.

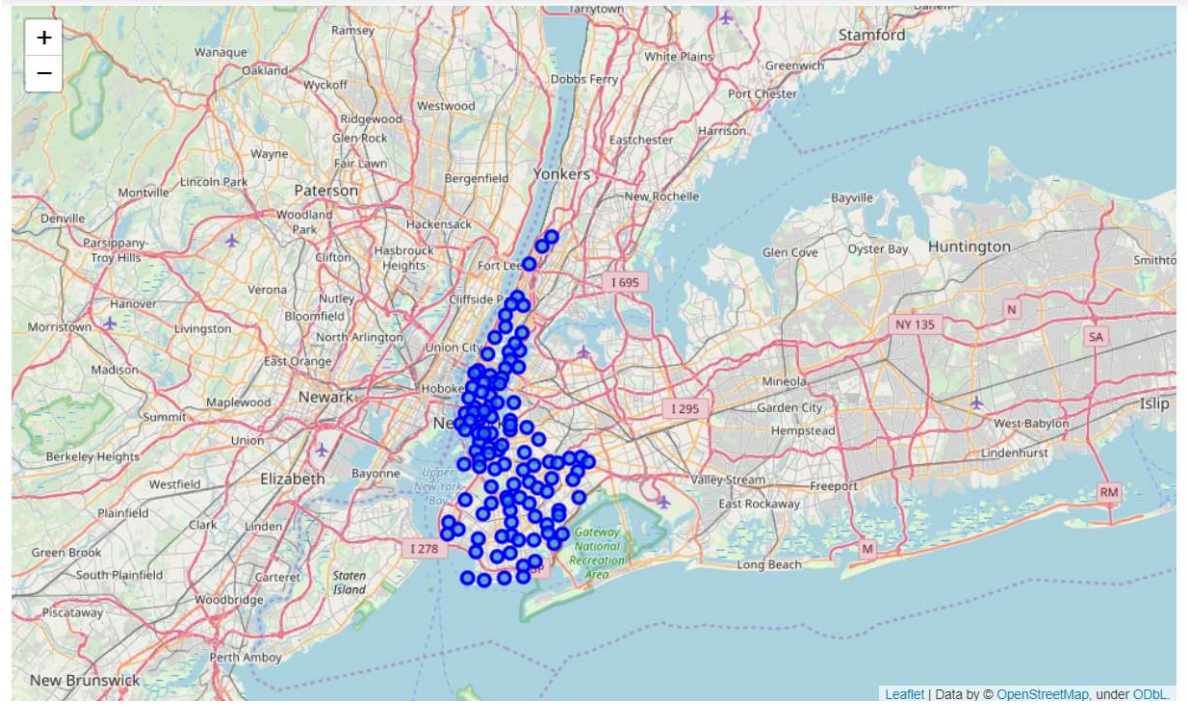
New York Population: Insights from the data :

- Manhattan (New York County) is the geographically smallest and most densely populated borough.
- Manhattan's (New York County's) population density of 72,033 people per square mile (27,812/km²) in 2015 makes it the highest of any county in the United States and higher than the density of any individual American city.
- Brooklyn (Kings County), on the western tip of Long Island, is the city's most populous borough.
- Queens (Queens County), on Long Island north and east of Brooklyn, is geographically the largest borough.

	Borough	County	Estimate 2020	square_miles	square_km	persons_sq_mi	persons_sq_km
0	The Bronx	Bronx	1,432,132	42.10	109.04	34,653	13,231
1	Brooklyn	Kings	2,582,830	70.82	183.42	37,137	14,649
2	Manhattan	New York	1,628,701	22.83	59.13	72,033	27,826
3	Queens	Queens	2,278,906	108.53	281.09	21,460	8,354
4	Staten Island	Richmond	476,179	58.37	151.18	8,112	3,132
5		City of New York	8,398,748	842.343	783.83	28,188	10,947
6		State of New York	19,745,289	1,701.399	122,284	416.4	159

3.4. Data 4: New York city geographical coordinates data has been utilized as input for the Foursquare API, that has been leveraged to provision venues information for each neighborhood. We used the Foursquare API data to explore neighborhoods in New York City.

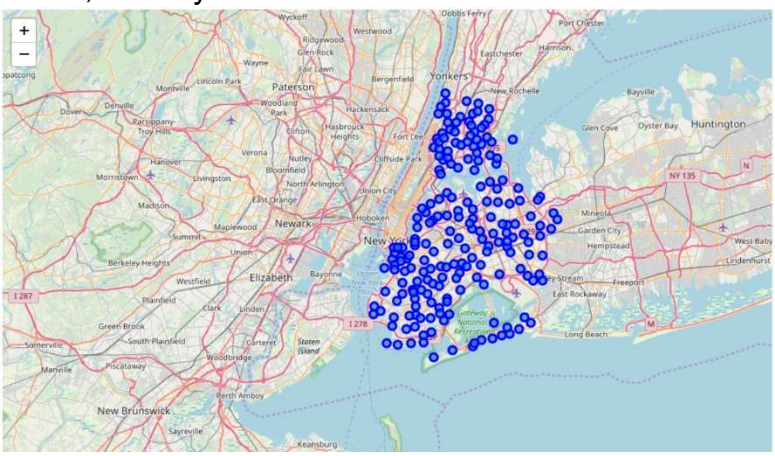
Segmenting and Clustering Neighborhoods - Brooklyn and Manhattan Visualization:



Brooklyn and Manhattan Venues:

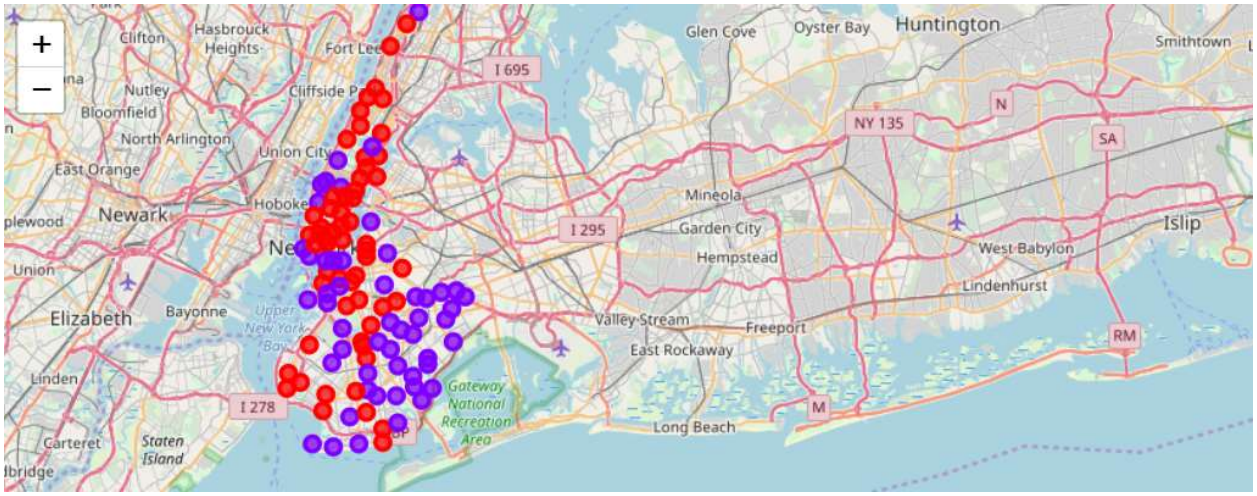
	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Sam's Pizza	40.879435	-73.905859	Pizza Place
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop

Bronx, Brooklyn and Queens Venues Visualization:



4. Results

From this venues data we filtered and used only the restaurant data for Brooklyn & Manhattan clustering and Bronx, Brooklyn and Queens clustering. As we focused only on restaurants business. Neighborhood K-Means clustering based on mean occurrence of venue category: To cluster the neighborhoods into two clusters we used the K-Means clustering Algorithm. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. It uses iterative refinement approach. Brooklyn & Manhattan. In the below Map Visualization, we can see the different types of clusters created by using K-Means for Brooklyn & Manhattan.



Cluster 0: The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated.

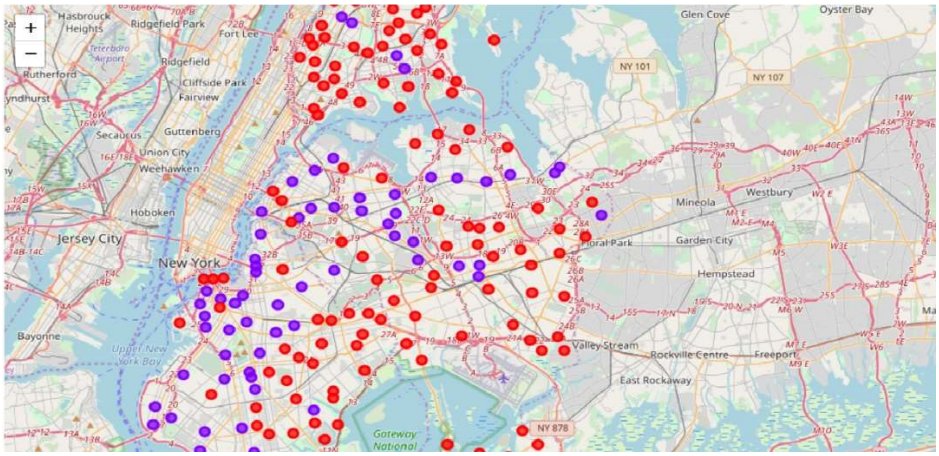
	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Brooklyn	Bay Ridge	40.625801	-74.030621	32	0
1	Brooklyn	Bensonhurst	40.611009	-73.995180	29	0
2	Brooklyn	Sunset Park	40.645103	-74.010316	39	0
3	Brooklyn	Sheepshead Bay	40.586890	-73.943186	32	0
4	Brooklyn	Flatbush	40.636326	-73.958401	34	0
5	Brooklyn	Crown Heights	40.670829	-73.943291	22	0
6	Brooklyn	Prospect Heights	40.676822	-73.964859	28	0
7	Brooklyn	Williamsburg	40.707144	-73.958115	25	0
8	Brooklyn	Bushwick	40.698116	-73.925258	25	0
9	Brooklyn	Brooklyn Heights	40.695864	-73.993782	23	0
10	Brooklyn	Cobble Hill	40.687920	-73.998561	24	0
11	Brooklyn	Fort Greene	40.688527	-73.972906	29	0
12	Brooklyn	Park Slope	40.672321	-73.977050	23	0
13	Brooklyn	Manhattan Beach	40.577914	-73.943537	22	0
14	Brooklyn	Bath Beach	40.599519	-73.998752	33	0

Cluster 1: The Total and Total Sum of cluster1 has highest value. It shows that the markets are saturated. Number of restaurants are very high.

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Manhattan	Marble Hill	40.876551	-73.910660	13	1
1	Brooklyn	Greenpoint	40.730201	-73.954241	21	1
2	Brooklyn	Gravesend	40.595260	-73.973471	8	1
3	Brooklyn	Brighton Beach	40.576825	-73.965094	20	1
4	Brooklyn	Manhattan Terrace	40.614433	-73.957438	17	1
5	Brooklyn	East Flatbush	40.641718	-73.936103	5	1
6	Brooklyn	Kensington	40.642382	-73.980421	21	1
7	Brooklyn	Windsor Terrace	40.656946	-73.980073	20	1
8	Brooklyn	Brownsville	40.663950	-73.910235	9	1
9	Brooklyn	Bedford Stuyvesant	40.687232	-73.941785	21	1
10	Brooklyn	Carroll Gardens	40.680540	-73.994654	21	1
11	Brooklyn	Red Hook	40.676253	-74.012759	11	1
12	Brooklyn	Gowanus	40.673931	-73.994441	21	1
13	Brooklyn	Cypress Hills	40.682391	-73.876616	11	1
14	Brooklyn	East New York	40.669926	-73.880699	4	1
15	Brooklyn	Starrett City	40.647589	-73.879370	8	1

There are no untapped neighborhoods in Brooklyn and Manhattan.

Bronx, Brooklyn and Queens: In the below Map Visualization, we can see the different types of clusters created by using K-Means for Bronx, Brooklyn and Queens.



Cluster 0: The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated. There are untapped neighborhoods. List is as given below.

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Bronx	Wakefield	40.894705	-73.847201	10	0
1	Bronx	Co-op City	40.874294	-73.829939	12	0
2	Bronx	Eastchester	40.887556	-73.827806	13	0
3	Bronx	Fieldston	40.895437	-73.905643	8	0
4	Bronx	Riverdale	40.890834	-73.912585	9	0

Cluster 1: The Total and Total Sum of cluster1 has highest value. It shows that the markets are saturated. Number of restaurants are very high.

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Bronx	Kingsbridge	40.881687	-73.902818	20	1
1	Bronx	Fordham	40.860997	-73.896427	26	1
2	Bronx	Parkchester	40.837938	-73.856003	27	1
3	Bronx	Belmont	40.857277	-73.888452	25	1
4	Bronx	Unionport	40.829774	-73.850535	20	1

There is one value with Total = 0 (Untapped Market

5. Conclusion

This analysis is performed on limited data. This may be right or may be wrong. But if good amount of data is available there is scope to come up with better results. If there are lot of restaurants probably there is lot of demand. Brooklyn and Manhattan have high concentration of restaurant business. Very competitive market. Bronx, Brooklyn and Queens also have good number of restaurants but not as many as required. So, this can be explored. As per the neighborhood or restaurant type mentioned like Indian Restaurant analysis can be checked. A venue with lowest risk and competition can be identified.