

# Capstone Project – Battle of the Neighborhoods Report

## 1 Introduction and Business Problem

### 1.1 About New York

The City of New York, is the most populous city in the United States. It is diverse, and the financial capital of the United States. They're multicultural. It offers a lot of business opportunities and an atmosphere conducive to companies. It has brought into the market a number of different companies. It is a global hub of enterprise and trade. The city is a major center in the United States of America for banking and finance, retail, international trade, transportation, tourism, real estate, digital media, conventional media, legal services, accounting, insurance, theatre, music, and arts.

That also means a highly competitive marketplace. As it is a highly developed area, it is also one of the highest costs of doing business. Therefore any new business venture or expansion has to be carefully analyzed. The insights derived from the analysis will give a good understanding of the business environment that helps to target the market strategically. This will assist in considering the risk in opening up a restaurant.

### 1.2 Business Problem

New York is known for its diverse restaurant scene. One can find cuisines from all over the world. Being home to over millions of immigrants has helped New York cement its place as a city of multi-cultural foods and delicacies. Italian immigrants have a penchant for establishing fine pizzeria, Chinese and Asians bring in their cuisine for across the globe to New York. Jewish and Middle Eastern immigrants serve delicacies like Falafel, Hummus, etc. Indian immigrants bring their burst of spices and serve foods like Paneer Butter Masala, Masala Dosa, etc. Now, this just the tip of the iceberg; there are thousands of street vendors selling all sorts of things pretzels, bagels, burgers, hotdogs, etc.

From the above it's pretty obvious that to survive the food scene in New York is no small task. So, this brings us to the business problem which is how can one gain enough customers to make it in New York. This requires analyzing the location data of New York and look for some important insights like competitors, Boroughs, demographics, availability of suppliers, high footfall areas like parks, entertainment areas, etc.

By properly analyzing the above, one can decide the viability of setting up shop in a particular area.

### 1.3 Goal of this Data Science Project

The goal here is to provide a customer with enough insights into the restaurant scene in New York so that he/she can decide on setting up their establishment with some guarantee of success based on available data.

## 2 Data Requirements

The city we will be analyzing is New York.

The data sources for this project is as below:-

1.[https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)

New York has 5 Boroughs and this dataset has comprehensive data about latitudes and longitudes of all the neighborhoods in these 5 Boroughs.

2.<https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2>

This dataset contains information of Farmer's Markets. This gives access to locally and fresh produce available in New York.

3.[https://en.wikipedia.org/wiki/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/Economy_of_New_York_City)  
[https://en.wikipedia.org/wiki/Portal: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/Cuisine_of_New_York_City) [https://en.wikipedia.org/wiki/List\\_of\\_Michelin\\_starred\\_restaurants\\_in\\_New\\_York\\_City](https://en.wikipedia.org/wiki/List_of_Michelin_starred_restaurants_in_New_York_City)

The above links give information about New York's population, demographics and different cuisines.

4.Foursquare API data lets us explore location data of New York City.

	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
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	Loeser's Delicatessen	40.879242	-73.905471	Sandwich Place

A sample screenshot of the location data from the Foursquare API.

## 3 Methodology

The aim of this data science project is to give a customer the ideal location to open a restaurant in New York City.

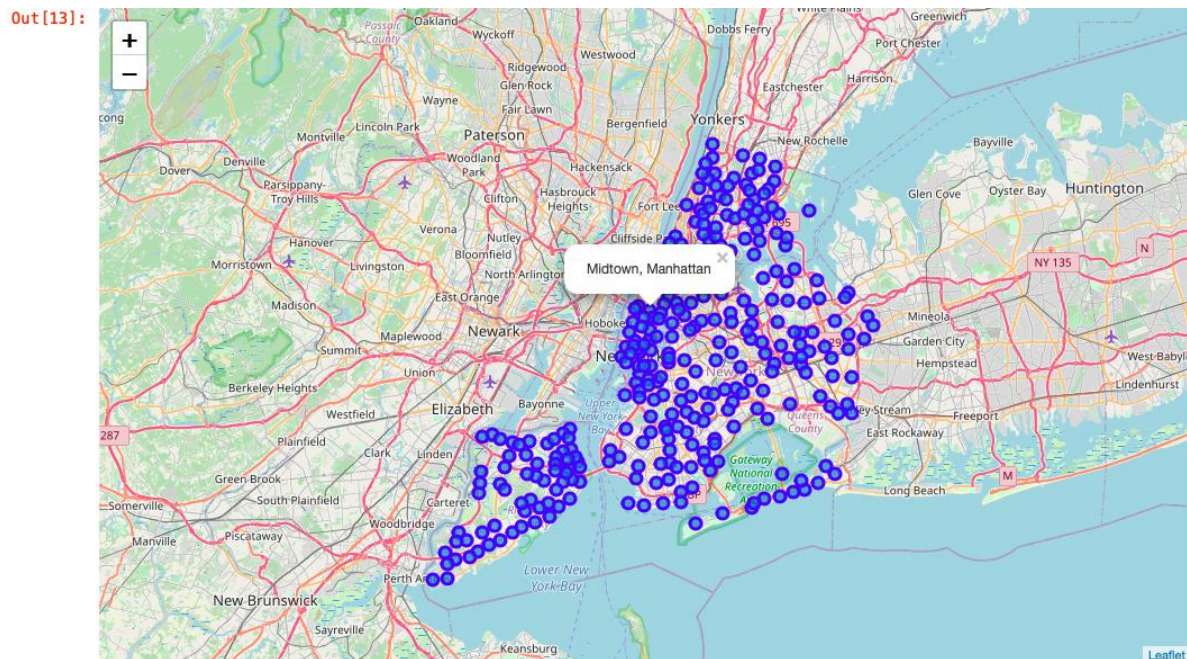
NYC has 5 boroughs – Manhattan, Brooklyn, Queens, Bronx and Staten Island. Manhattan and Brooklyn has been first subjected clustering and Queens ,Bronx and Staten Island are then clustered.

### 3.1 Exploratory Data Analysis

#### 3.1.1 New York City Geographical Data

- Load the data and explore data from newyork\_data.json file.
- Transform the data of nested python dictionaries into a pandas dataframe.

- This dataframe contains the geographical coordinates of New York city neighborhoods.
- This data will be used to get Venues data from Foursquare.
- The geopy and folium libraries are used to create a map of New York city with neighborhoods superimposed on top.



Map of all the neighborhoods in NYC.

### 3.1.2 NYC Demographics and Population

- Kings County in Brooklyn is the most populous in NYC.
- Richmond County is the least populous in Staten Island.

Out[34]:

	Borough	County	Estimate_2017	square_miles	square_km	persons_sq_mi	persons_sq_km
0	Manhattan	New York	1,664,727	22.83	59.13	72,033	27,826
1	The Bronx	Bronx	1,471,160	42.10	109.04	34,653	13,231
2	Brooklyn	Kings	2,648,771	70.82	183.42	37,137	14,649
3	Queens	Queens	2,358,582	108.53	281.09	21,460	8,354
4	Staten Island	Richmond	479,458	58.37	151.18	8,112	3,132
5		City of New York	8,622,698	302.64	783.83	28,188	10,947
6		State of New York	19,849,399	47,214	122,284	416.4	159

Dataframe of the population in NYC

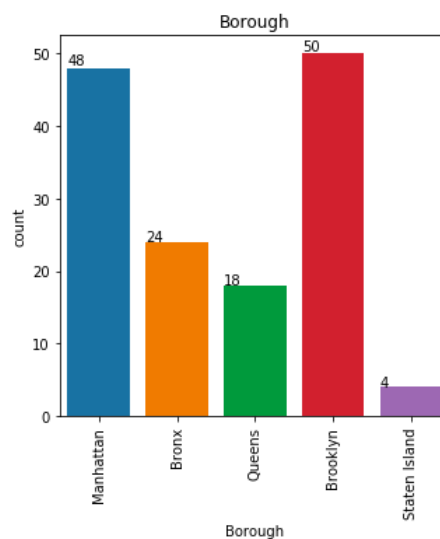
Out [45]:

	Racialcomposition	2010	1990	1970	1940
0	White	44.0%	52.3%	76.6%	93.6%
			43.2%	62.9%	92.0%
2	Black or African American	25.5%	28.7%	21.1%	6.1%
3	Hispanic or Latino (of any race)	28.6%	24.4%	16.2%	1.6%
4	Asian	12.7%	7.0%	1.2%	–

Demographics of NYC

### 3.1.3 Farmer's Market Dataset

- There are 144 Farmer's Market in NYC
- Manhattan and Brooklyn have the most
- Queens, Bronx and Staten Island have comparatively lower number of Farmer's Market.



Bar graph showing the number of Farmer's Markets in the Boroughs.

### 3.1.3 Cuisine of NYC Dataset

- New York has a diverse collection of restaurants offering various cuisines.
- These are scattered throughout the Boroughs of NYC.
- These have been visualized using wordclouds.



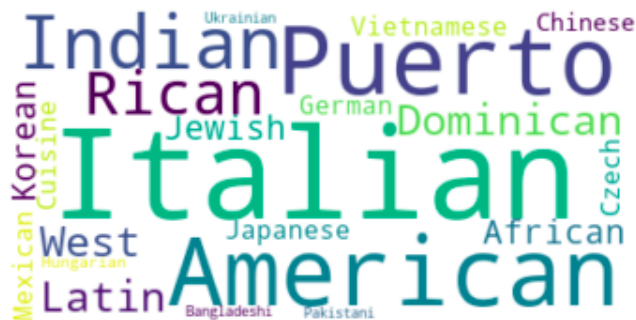
NYC Wordcloud



Brooklyn Wordcloud



Queens Wordcloud



Manhattan Wordcloud





Bronx Wordcloud

## 4 Results

- Brooklyn and Manhattan Venues using Foursquare API.

	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Marble Hill	40.876551	-73.91056	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91056	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91056	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91056	Sam's Pizza	40.879435	-73.905859	Pizza Place
4	Marble Hill	40.876551	-73.91056	Loeser's Delicatessen	40.879242	-73.905471	Sandwich Place

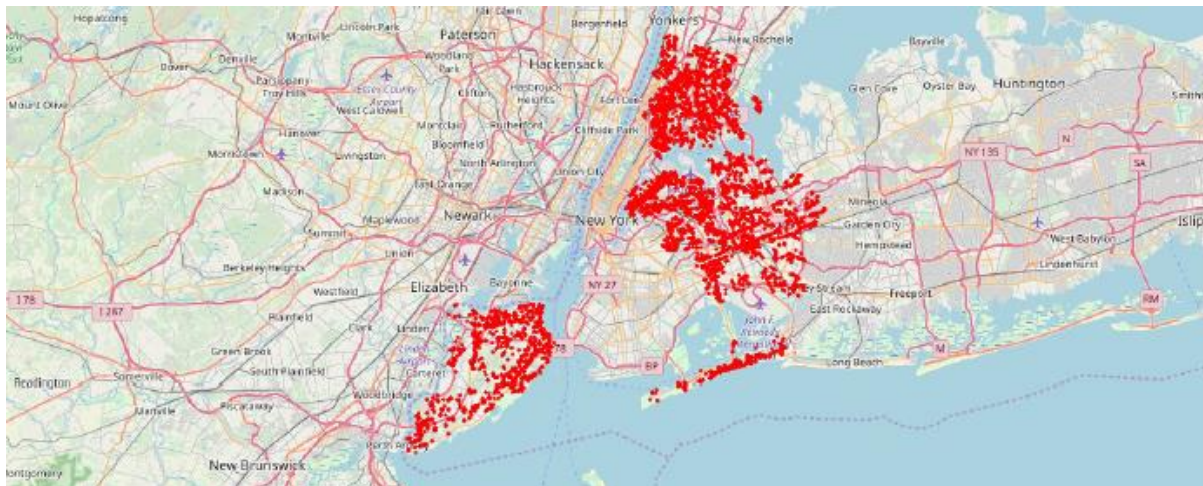
- Brooklyn and Manhattan Venues Visualization.



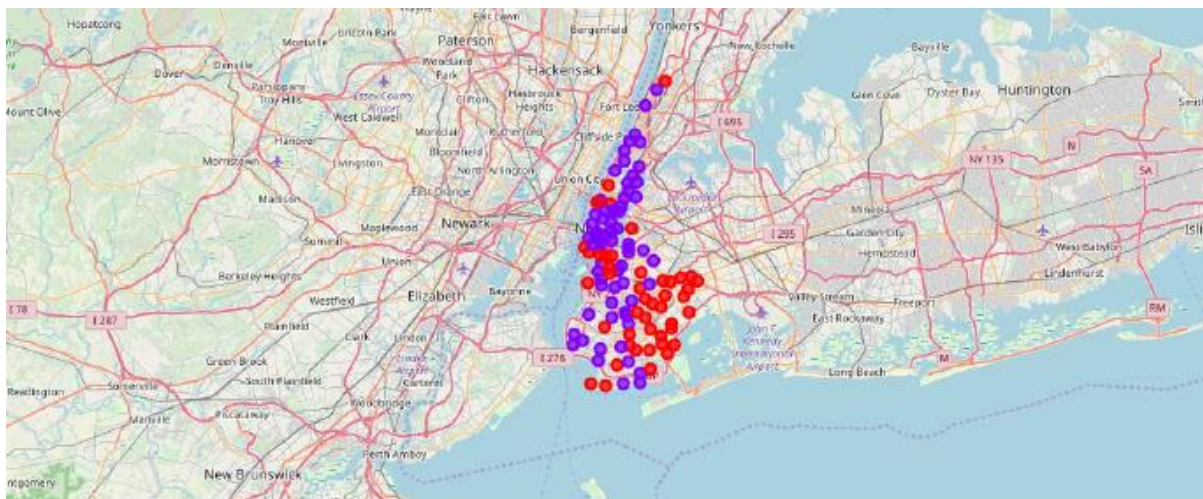
- Queens, Bronx and Staten Island Venues using Foursquare API.

	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Ripe Kitchen & Bar	40.898152	-73.838875	Caribbean Restaurant
2	Wakefield	40.894705	-73.847201	Jackie's West Indian Bakery	40.889283	-73.843310	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Ali's Roti Shop	40.894036	-73.850935	Caribbean Restaurant
4	Wakefield	40.894705	-73.847201	Rite Aid	40.896521	-73.844680	Pharmacy

- Queens, Bronx and Staten Island Venues Visualization.



- From this the venues data is filtered and is used for only the restaurant data for Brooklyn & Manhattan clustering.
- Bronx, Queens and Staten Island clustering is done later.
- Neighborhoods 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.
- The below map visualization shows the different clusters formed in Brooklyn and Manhattan.



- The below dataframe show 2 clusters:-
  1. Cluster 1 - This indicates the saturated markets in those 2 boroughs.
  2. Cluster 2 - This indicates that there no untapped markets available in those 2 boroughs
- This means that setting up a restaurant in Brooklyn or Manhattan isn't a viable option.

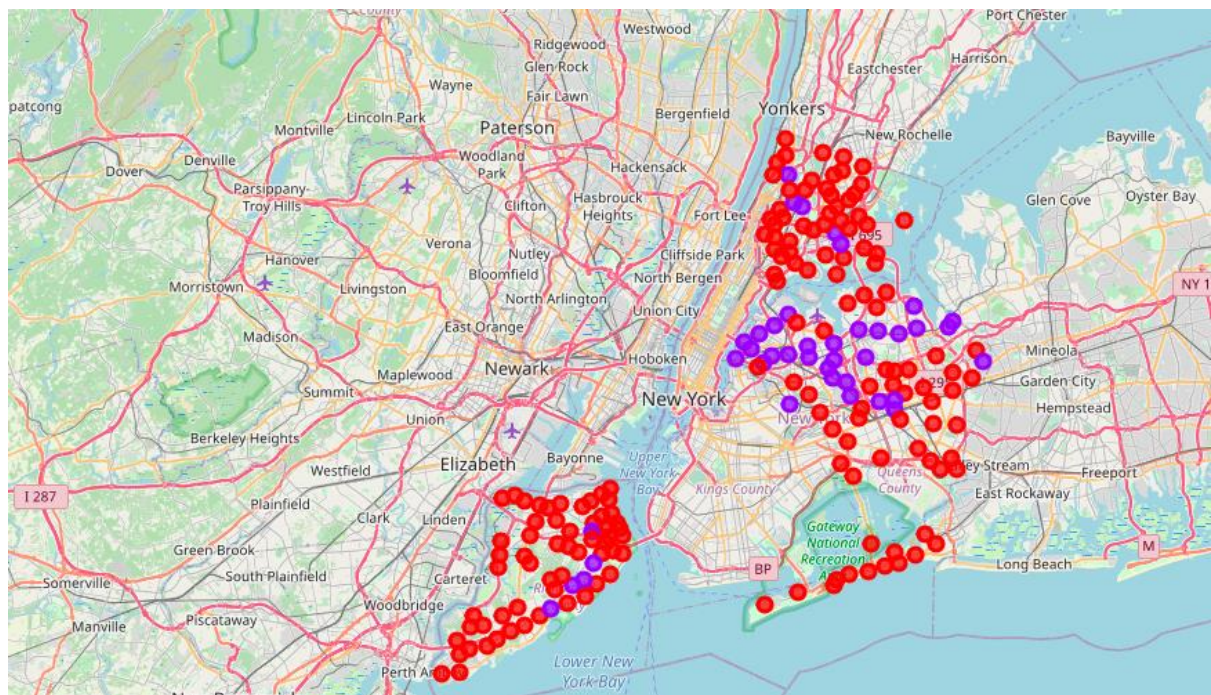


Cluster 1 - Saturated Markets						
In [93]: BM_merged[BM_merged['Cluster_Labels'] == 1].reset_index(drop=True)						
Out[93]:						
	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Brooklyn	Bay Ridge	40.625801	-74.030621	33	1
1	Brooklyn	Bensonhurst	40.611009	-73.995180	27	1
2	Brooklyn	Sunset Park	40.645103	-74.010316	40	1
3	Brooklyn	Brighton Beach	40.576825	-73.965094	22	1
4	Brooklyn	Sheepshead Bay	40.586890	-73.943186	34	1
5	Brooklyn	Flatbush	40.636326	-73.958401	35	1
6	Brooklyn	Kensington	40.642382	-73.980421	23	1
7	Brooklyn	Windsor Terrace	40.656946	-73.980073	22	1
8	Brooklyn	Prospect Heights	40.676822	-73.964859	30	1
9	Brooklyn	Williamsburg	40.707144	-73.958115	32	1
10	Brooklyn	Bushwick	40.698116	-73.925258	23	1

Cluster 0 - Untapped Markets						
In [94]: BM_merged[BM_merged['Total'] == 0].reset_index(drop=True)						
Out[94]:						
	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels

- The below map visualization shows the different clusters formed in Queens, Bronx and Staten Island.



- The below dataframe show 2 clusters:-
  1. Cluster 1 - This indicates the saturated markets in those 3 boroughs.
  2. Cluster 2 - This indicates that there no untapped markets available in those 3 boroughs
- We can see that there is a chance that a restaurant might be able to survive the NYC restaurant by operating in Staten Island.



#### Cluster 1 - Saturated Markets

```
In [119]: BQS_merged[BQS_merged['Cluster_Labels'] == 1].reset_index(drop=True)
```

```
Out[119]:
```

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Bronx	Kingsbridge	40.881687	-73.902818	22	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	23	1
5	Queens	Astoria	40.768509	-73.915654	31	1
6	Queens	Woodside	40.746349	-73.901842	46	1
7	Queens	Jackson Heights	40.751981	-73.882821	58	1
8	Queens	Elmhurst	40.744049	-73.881656	60	1
9	Queens	Corona	40.742382	-73.856825	21	1
10	Queens	Forest Hills	40.725264	-73.844475	33	1

#### Cluster 0 - Untapped Markets

```
In [120]: BQS_merged[BQS_merged['Total'] == 0].reset_index(drop=True)
```

```
Out[120]:
```

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Staten Island	Todt Hill	40.597069	-74.111329	0	0
1	Staten Island	Port Ivory	40.639683	-74.174645	0	0
2	Staten Island	Bloomfield	40.605779	-74.187256	0	0

## 5 Conclusions

1. The number of Farmer's Markets should be increased in Queens, Bronx and Staten Island.
2. Manhattan and Brooklyn pretty much have restaurants of cuisines from many countries.  
So, it could be tough to operate a successful restaurant in these 2 Boroughs unless something different is offered on the menu.
3. Individual analysis of a particular type of cuisine might yield better insights as there is a constraint on the datasets here.
4. Also another way to view these results is to consider a saturated market and assume that the demand is more and hence could encourage more competition.