

PROJECT REPORT
THE
BATTELE OF NEIGHBORHOODS
NEW YORK

The Battle of Neighbourhoods - Project Report

1. Introduction & Business Problem:

Problem Background:

The City of New York, is the most populous city in the United States with an estimated 8,398,748 residents as of July 2018. It is a global hub of business and commerce. The city is a major centre for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theatre, fashion, and the arts in the United States. It is multicultural and offers a business-friendly environment.

New York's business market is highly competitive. New business ventures need lots of careful planning and Analysis. The insights derived from analysis of this project will summarise the current business environment, which help new business ventures to strategically target the market thus helping business venture to reduce of cost risk associated with setting a business.

Problem Description:

A well-established company wants to venture in to food business and are interested in opening a chain of restaurants in the city.

This company need to choose the correct location to start its first venture. If this is successful then they can replicate the same in other locations. First move is very important, thereby choice of location is very important for them.

Below data is taken from https://en.wikipedia.org/wiki/New_York_City

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.

1. Central and Eastern European immigrants, especially Jewish immigrants - bagels, cheesecake, hot dogs, knishes, and delicatessens
2. Italian immigrants - New York-style pizza and Italian cuisine
3. Jewish immigrants and Irish immigrants - pastrami and corned beef
4. Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city
5. mobile food vendors - Some 4,000 licensed by the city
6. Middle Eastern foods such as falafel and kebabs examples of modern New York street food
7. 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.

Based on the above stats, in order to survive in such competitive market, it is very important to strategically plan. Restaurant business depends on various factors as listed below, hence these needs to be studied in order to decide on the best possible location to start a first venture:

1. Population
2. Demographics
3. Location of the Whole sale markets (Ex: Meat, Vegetables etc.) so that ingredients can be purchased fresh to maintain quality and cost?
4. Nearby Venues in the Neighbourhoods of the borough
5. competitors in that location?
6. Cuisine served by the competitors
7. Segmentation of the Borough
8. Market information (Untapped / Tapped)

Target Audience:

The objective of this project is to identify, provide recommendation on which neighborhood of New York city will be best suitable to start a restaurant and what cuisine can be offered. Also provide the rationale behind the recommendations made. These recommendations can help company in discussion or anyone who has plans of starting a new restaurant venture in New York city.

2. Data:

As part of this project, we will be analysis the various factors that influence the decision on the Restaurant location. In order to conduct the analysis, we would require to analyse the following available datasets of New York City.

Dataset1: Neighborhood has a total of 5 boroughs and 306 neighbourhoods. In order to segment the neighbourhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighbourhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood. This dataset has been sourced from the following Link.

Link to the dataset: 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

Dataset2: Second data which will be used is the DOHMH Farmers Markets dataset. In this we will be using the data of Farmers Markets. <https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets/8vwk-6iz2>

Borough	Market Name	Street Address	Latitude	Longitude	Days of Operation	Hours of...	Season Dates	Accepts...	Open Year-Round	Stellar Cooking Demo
Brooklyn	Urban Oasis Farmers Market	681 Clarkson Ave	40.656255	-73.936608	Wednesday	2 - 5:30 p.m.	06/26/2019-11/06/2019	No	No	No
Staten Island	Staten Island Mall Greenmarket	Marsh Ave & Ring Rd	40.583804	-74.161245	Saturday	8 a.m. - 3 p.m.	Year-Round	Yes	Yes	No
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
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
Manhattan	Grass Roots Farmers Market	W 145th St bet Edgecombe & Bradhurst Aves, at Jackie Rob...	40.823647	-73.943844	Tuesday & Saturday	9 a.m. - 4 p.m.	07/11/2019-11/21/2019	Yes	No	No
Queens	Brooklyn Grange Farm Stand at Long Isla...	37-18 Northern Blvd (Roof-top)	40.752096	-73.925887	Saturday	11 a.m. - 4 p.m.	05/25/19-10/26/19	Yes	No	No
Bronx	CS211 GoHealthy Farm Stand	1919 Prospect Ave	40.843546	-73.889755	Wednesday	2:30 - 5:30 p.m.	07/08/2019 - 11/22/2019	Yes	No	No
Queens	Flushing Greenmarket	Kissena Blvd & Maple Ave, at Maple Playground	40.756357	-73.825773	Wednesday	8 a.m. - 3 p.m.	07/03/19-11/27/19	Yes	No	Yes
Bronx	Harvest Home Jacobi Hospital Farmers M...	1400 Pelham Pkwy S, at Eastchester Rd	40.857427	-73.847079	Tuesday & Friday	8 a.m. - 4 p.m.	06/11/2019-11/19/2019	Yes	No	No
Brooklyn	Marcus Garvey Apartments Farm Stand	300 Chester St	40.664477	-73.910384	Saturday	11 a.m. - 5 p.m.		Yes	No	No
Brooklyn	Fort Greene Park Greenmarket	Washington Park bet Dekalb Ave & Wiloughby St	40.691707	-73.973539	Saturday	8 a.m. - 3 p.m.	Year-Round	Yes	Yes	No
Bronx	Riverdale Y Sunday Farmers' Market	Independence Ave bet W 236th & 237th Sts, at MS/HS 141	40.8871	-73.913306	Sunday	9 a.m. - 2 p.m.	05/05/2019-11/24/2019	No	No	No
Brooklyn	Cypress Hills Youthmarket	Fulton & Richmond Sts	40.682391	-73.876616	Friday	10 a.m. - 5 p.m.	07/12/2019-11/22/2019	Yes	No	Yes
Bronx	South Bronx Farmers Market	E 138th St bet Willis & Alexander Aves	40.809107	-73.922887	Saturday	10 a.m. - 4 p.m.	06/15/2019-11/23/2019	Yes	No	Yes
Brooklyn	Isabellia Farm Stand	514 Rockaway Ave	40.667554	-73.910207	Friday & Saturday	(F) 9 a.m. - 4 p.m.	06/28/2019-10/25/2019 ; 06/29/2019-10/2...	Yes	No	No
Bronx	Kingsbridge-Riverdale Farmers Market	260 W 231st St	40.880038	-73.907005	Sunday	9 a.m. - 5 p.m.	05/26/19-11/24/19	Yes	No	No
Manhattan	Tompkins Square Greenmarket	F 7th St & Avenue A	40.726054	-73.983534	Sunday	9 a.m. - 6 p.m.	Year-Round	Yes	Yes	No

Dataset3: For the below analysis we will get data from Wikipedia as given below:

1. Population
2. Demographics
3. Cuisine of New York city
4. List of Michelin Stared Restaurants in the 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

Dataset4: A data set will be constructed by using Foursquare API. New York city geographical coordinates data will be used as an input to secure the venues information for each neighborhood of the Boroughs. This dataset then will be further segmented and clustered for analysis by using Machine learning algorithms.

Foursquare API URL = <https://api.foursquare.com/v2/venues/explore>

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
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

3. Methodology:

Business Requirement:

The Main aim of this project is to thoroughly study various factors of the New York city to identify a suitable location to start the first restaurant venture.

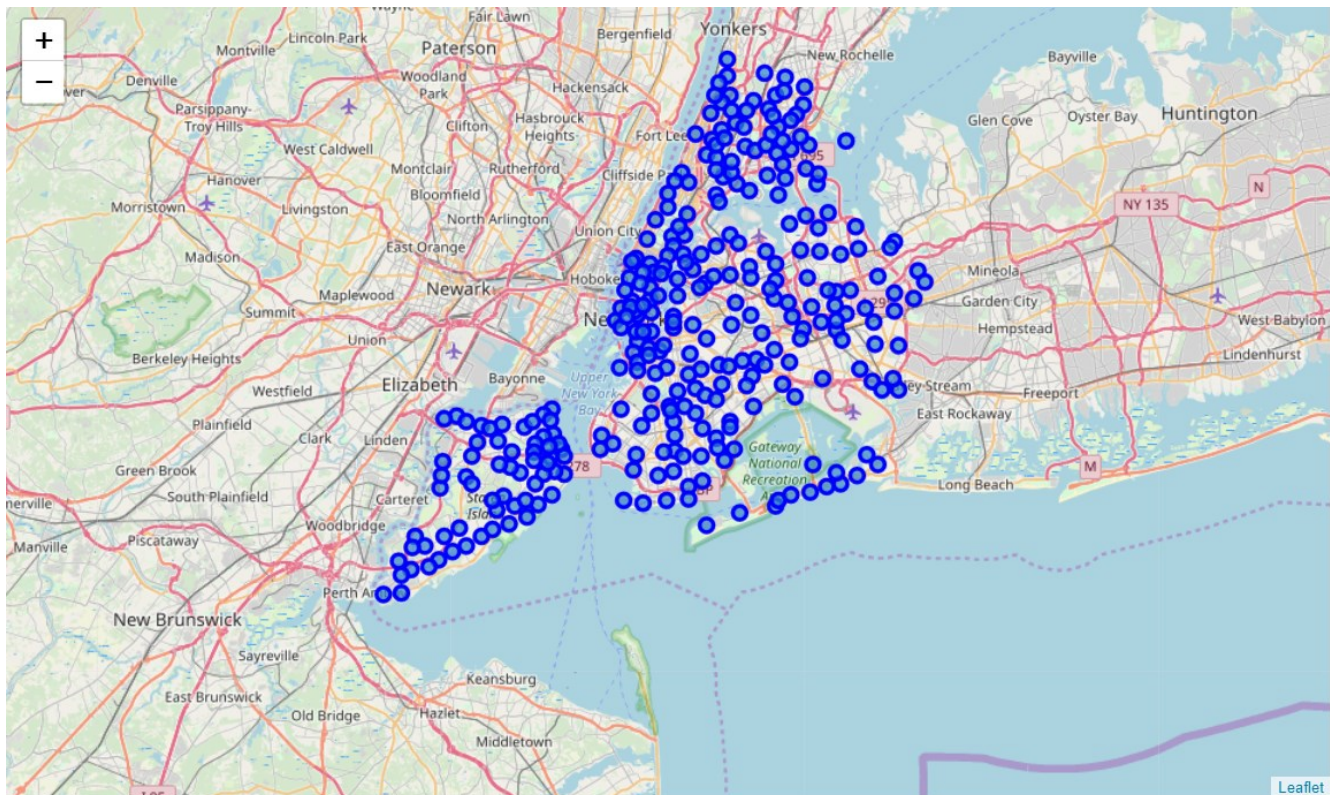
Exploratory Data Analysis:

The following exploratory analysis will provide an overview of the analysis conducted against each dataset mentioned in the data section and also visual representation is provided in the form of either Folium maps or Bar charts.

Dataset1: Data of Geographical co-ordinates of New York city



1. Data for the data set has been sourced from https://geo.nyu.edu/catalog/nyu_2451_34572 to form a **newyork_data.json** file.
2. The next task is essentially transforming this data of nested Python dictionaries into a *pandas* data frame(neighbourhoods).
3. The data frame has 5 boroughs and 306 neighbourhoods and contains geographical coordinates of all the neighbourhoods of New York city boroughs
4. Once this data is ready, with the help of folium library a map has been created by super imposing the neighbourhoods on the New York city map



Dataset2: Data of Farmers Markets

DOHMH Farmers Markets dataset. In this we will be using the data of Farmers Markets.

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

After analysis the farmers market, it was observed that both **Manhattan and Brooklyn** have highest concentration of these markets.

Lowest number of Farmers market are in **Staten Island**

Dataset3: Analysing New York city population data, its Demographics and cuisines

Source: https://en.wikipedia.org/wiki/New_York_City#Demographics

Population Density

New York City is the most populous city in the United States, with an estimated 8,398,748 residents as of July 2018. In 2017, the city had an estimated population density of 28,491 inhabitants per square mile (11,000/km²)

Manhattan (New York County) is the geographically smallest and most densely populated borough.

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, the most ethnically diverse county in the United States

New York City's five boroughs V · T · E								
Jurisdiction		Population	Gross Domestic Product		Land area		Density	
<i>Borough</i>	<i>County</i>	<i>Estimate (2017)^[191]</i>	<i>billions (US\$)^[192]</i>	<i>per capita (US\$)</i>	<i>square miles</i>	<i>square km</i>	<i>persons / sq. mi</i>	<i>persons / sq. km</i>
The Bronx	Bronx	1,471,160	28.787	19,570	42.10	109.04	34,653	13,231
Brooklyn	Kings	2,648,771	63.303	23,900	70.82	183.42	37,137	14,649
Manhattan	New York	1,664,727	629.682	378,250	22.83	59.13	72,033	27,826
Queens	Queens	2,358,582	73.842	31,310	108.53	281.09	21,460	8,354
Staten Island	Richmond	479,458	11.249	23,460	58.37	151.18	8,112	3,132
City of New York		8,622,698	806.863	93,574	302.64	783.83	28,188	10,947
State of New York		19,849,399	1,547.116	78,354	47,214	122,284	416.4	159

New York City is the most populous city in the United States, with an estimated 8,398,748 residents as of July 2018, incorporating more immigration into the city than outmigration since the 2010 United States Census. More than twice as many people live in New York City as compared to Los Angeles, the second-most populous U.S. city, and within a smaller area.

New York City gained more residents between April 2010 and July 2014 (316,000) than any other U.S. city. New York City's population is about 43% of New York State's population and about 36% of the population of the New York metropolitan area.

Racial composition [hide]	2010^[270]	1990^[272]	1970^[272]	1940^[272]
White	44.0%	52.3%	76.6%	93.6%
—Non-Hispanic	33.3%	43.2%	62.9% ^[273]	92.0%
Black or African American	25.5%	28.7%	21.1%	6.1%
Hispanic or Latino (of any race)	28.6%	24.4%	16.2% ^[273]	1.6%
Asian	12.7%	7.0%	1.2%	–

Cuisine of New York City

The **cuisine of New York City** comprises many cuisines belonging to various ethnic groups that have entered the United States through the city. Almost all ethnic cuisines are well represented in New York City, both within and outside the various ethnic neighborhoods

Data for the analysis has been sourced from https://en.wikipedia.org/wiki/Cuisine_of_New_York_City

Below word cloud images provide a clear picture of the preferred cuisine in each New York borough



<matplotlib.figure.Figure at 0x7f562c289160>

Most Preferred Food in New York City -

1. Italian
2. Puerto Rican
3. Mexican
4. Jewish
5. Indian
6. Pakistani
7. Dominican



<matplotlib.figure.Figure at 0x7f562c149b00>

Most Preferred Food in Brooklyn is -

1. Italian
2. Puerto Rican
3. Mexican



<matplotlib.figure.Figure at 0x7f562c122160>

Most Preferred Food in Queens is -

1. Indian
2. Irish
3. Pakistani
4. Mexican



<matplotlib.figure.Figure at 0x7f562c178d68>

Most Preferred Food in Manhattan is -

1. Italian
2. American
3. Puerto Rican
4. Indian



```
<matplotlib.figure.Figure at 0x7f562c126978>
```

Most Preferred Food in The Bronx is -

1. Italian
2. Puerto Rican
3. Albanian
4. Dominican

Most Preferred Food in Staten Island is -

1. Italian
2. Indian
3. Russian
4. Arab and Polish

Dataset4: Analysing New York city Boroughs segmented and Clustered in two groups

Foursquare API has been used to explore neighbourhoods of the New York city. Geographical data analysed and gathered in Dataset1 has been used as an input to the API to source the Venues information for each borough.

We have used **explore** function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters.

k -means clustering algorithm has been used to complete the analysis

Finally, Folium library has been used to visualize the neighborhoods

Segmenting and Clustering:

Since New York city has a total of 5 boroughs and 306 neighbourhoods, activity is divided into 2 parts.

1. Clustering of Manhattan Brooklyn.
2. Clustering of Bronx, Queens and Staten Island

Clustering of Manhattan Brooklyn:

For the analysis, Venues have been limited to 200 and have collected nearby venues within a radius of 1000 meters.

A Total of 397 unique venue types have been sourced. Below is the snapshot of the top 5 rows of the DF

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
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

Clustering of Bronx, Queens and Staten Island:

For the analysis, Venues have been limited to 200 and have collected nearby venues within a radius of 1000 meters.

A Total of 387 unique venue types have been sourced. Below is the snapshot of the top 5 rows of the DF

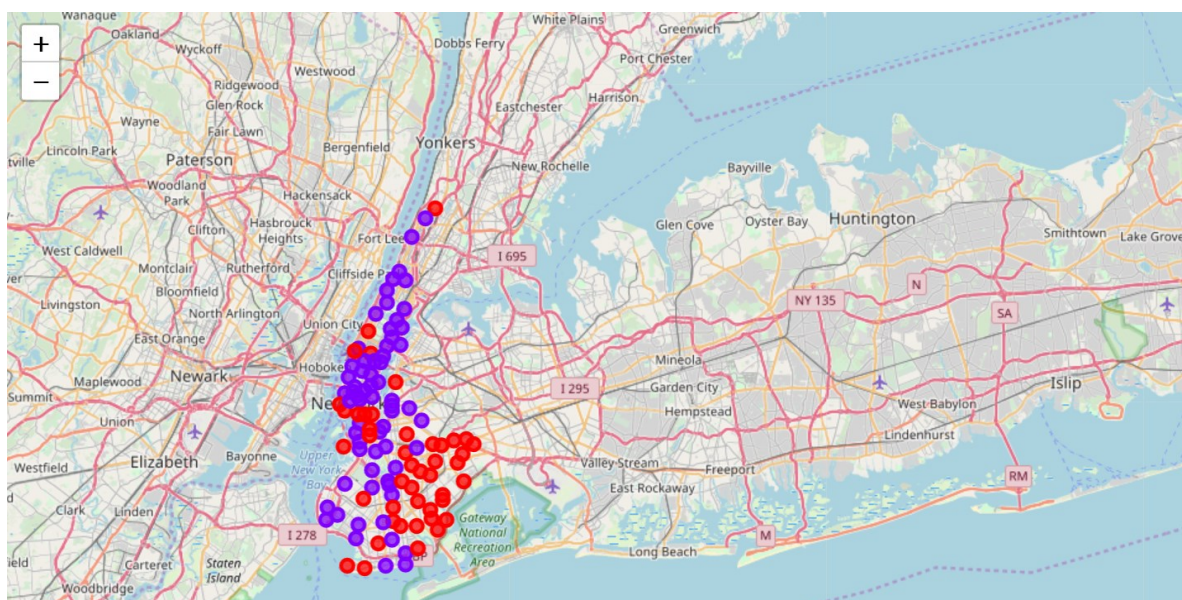
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Ripe Kitchen & Bar	40.898152	-73.838875	Caribbean Restaurant
2	Wakefield	40.894705	-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	-73.856935	Caribbean Restaurant
4	Wakefield	40.894705	-73.847201	Rite Aid	40.896521	-73.844680	Pharmacy

4. Results and Discussion:

From the Foursquare API only venues that are restaurants have been sourced for both the clusters and super imposed the venues on the Neighbourhoods.

Brooklyn and Manhattan Clustering

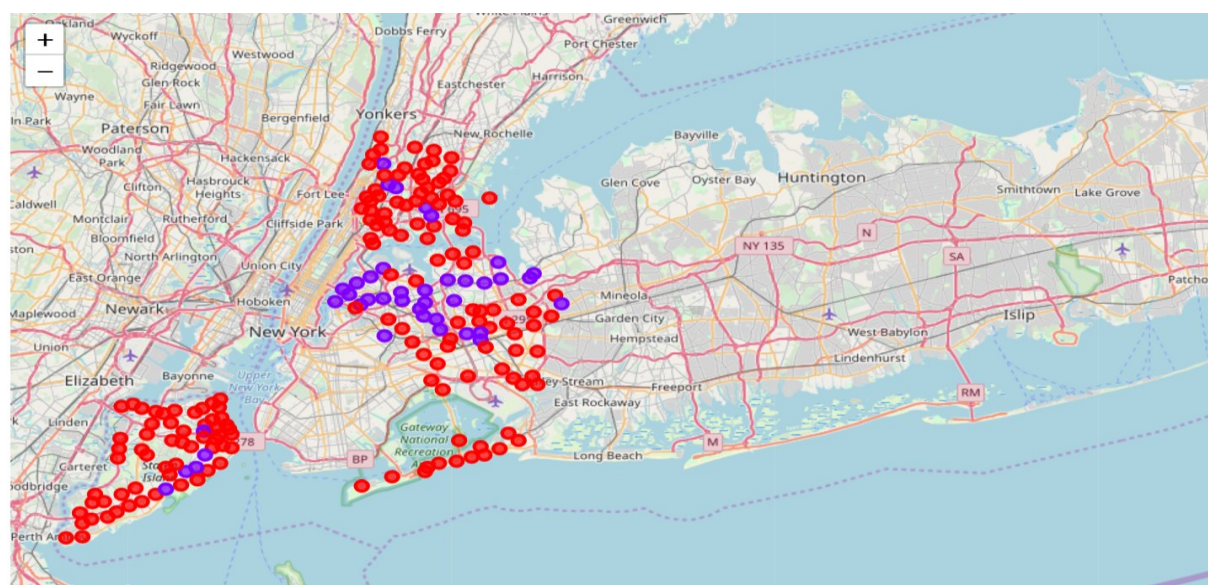
Based on the cluster analysis, it is identified that the Market is very saturated in both these boroughs and there aren't any neighbourhoods that were not tapped so far.



Bronx, Queens and Staten Island Clustering

Based on the cluster analysis, it is identified that even though the Market of most of the neighbourhoods is saturated in these boroughs, there are still few neighbourhoods that were not tapped so far

	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



Based on the above, as the competition is high in boroughs like Brooklyn and Manhattan with varieties of cuisines offered, a new restaurant will need to offer unique Menu in order to survive the competition. Based on the type of Cuisine that the restaurant is planning an optimal location needs to be identified taking the analysis of dataset3 into consideration.

Even though some neighbourhoods in Staten Island lack the competition, it is entirely possible that there is a very good reason for small number of restaurants in any of those neighbourhoods, reasons which would make them unsuitable for a new restaurant regardless of lack of competition in these neighbourhoods.

5. Conclusion:

The objective of this project is to identify, provide recommendation on which neighborhood of New York city will be best suitable to start a restaurant and what cuisine can be offered. Also provide the rationale behind the recommendations made. These recommendations can help company in discussion or anyone who has plans of starting a new restaurant venture in New York city

Data gathered mostly sourced from Wikipedia and also only a limited number of venues information has been sourced from Foursquare API. Clustering, segmenting and the Analysis was performed on a limited set of data.

Brooklyn and Manhattan have maximum number of restaurants, hence there might be a scope for a new restaurant if the location is carefully chosen as reason for maximum concentration may be due to huge demand for restaurant food.

Recommended Neighbourhoods should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met