

The Battle of the Neighborhoods

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 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 analysed 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.

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.

- ☐ Indian immigrants –Dhosa, Thali, Roti ,Biryani.
- ☐ Japanese –Athena.
- ☐ Central and Eastern European immigrants, - bagels, cheesecake, hot dogs, knishes, and delicatessens
- ☐ Italian immigrants - New York-style pizza and Italian cuisine
- ☐ Chinese and other Asian restaurants, sandwich, 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 :

1. New York Population
2. New York City Demographics
3. Are there any Farmers Markets, Wholesale markets etc nearby so that the ingredients can be purchased fresh to maintain quality and cost?
4. Are there any venues like College, Entertainment zones, Parks etc nearby where floating population is high increasing the visitors.
5. Who are the competitors in that location?
6. Cuisine served / Menu of the competitors .
7. Segmentation of the Borough

ABC Company wants to establish its first restaurant in the city of New York .To Be Successful in Business startup factors like Location, Demographics and their preference etc are very important for consideration.

To Recommend the Locations at which the restaurant could be started to have high profits we as a Data Science Team will locate and recommend the management of the company the best choices of Location.

2. Data :

One city will be analysed in this project : Newyork City.

We will be using the below datasets for analysing Newyork city

Data 1 : Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segement theneighborhoods 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

	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

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>

GrowNYC's Fresh Food Box Program is a food access initiative that enables under-served communities to purchase fresh, healthy, and primarily regionally grown produce well below traditional retail prices.

A farmers' market is often defined as a public site used by two or more local or regional producersfor the direct sale of farm products to consumers. In addition to fresh fruits and vegetables, markets may sell dairy products, fish, meat, baked goods, and other minimally processed foods.

	FacilityName	Service Category	Service_Type	Address	Address 2	Borough	ZipCode	Latitude	Longitude	AdditionalInfo	StartDate	EndDate	Monday	Tuesday	Wednesday	Thursday
0	Inwood Park Greenmarket	Farmers Markets and Food Boxes	Farmers Markets	isham St bet Seaman & Cooper		NaN	Manhattan	10034	40.869009	-73.920320	Open year-round	NaN	NaN	NaN	NaN	NaN
1	82nd Street Greenmarket	Farmers Markets and Food Boxes	Farmers Markets	82nd St bet 1st & York Aves		NaN	Manhattan	10028	40.773448	-73.948954	Open year-round	NaN	NaN	NaN	NaN	NaN
3	125th Street Farmers Market	Farmers Markets and Food Boxes	Farmers Markets	125th St & Adam Clayton Powell Jr Blvd		NaN	Manhattan	10027	40.808981	-73.948327	Market open dates: 6/13/2017 to 11/21/2017	06/13/2017	11/21/2017	NaN	10am-7pm	NaN
4	170 Farm Stand	Farmers Markets and Food Boxes	Farmers Markets	170th St & Townsend Ave		NaN	Bronx	10452	40.840095	-73.916827	Market open dates: 7/5/2017 to 11/22/2017	07/05/2017	11/22/2017	NaN	NaN	2:30pm-6:30pm
5	175th Street Greenmarket	Farmers Markets and Food Boxes	Farmers Markets	175th St bet Wadsworth Ave & Broadway		NaN	Manhattan	10033	40.845956	-73.937813	Market open dates: 6/29/2017 to 11/30/2017	06/29/2017	11/30/2017	NaN	NaN	8am-5pm

Data 3 : For the below analysis we will get data from wikipedia as given below :

1. New York Population
2. New York City Demographics
3. 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

	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

Data 4 : Newyork 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.

	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

3.Methodology :

Business Understanding :

Our main goal is to get optimum location for new restaurant business in New York City for XYZ Company.

Analytic Approach :

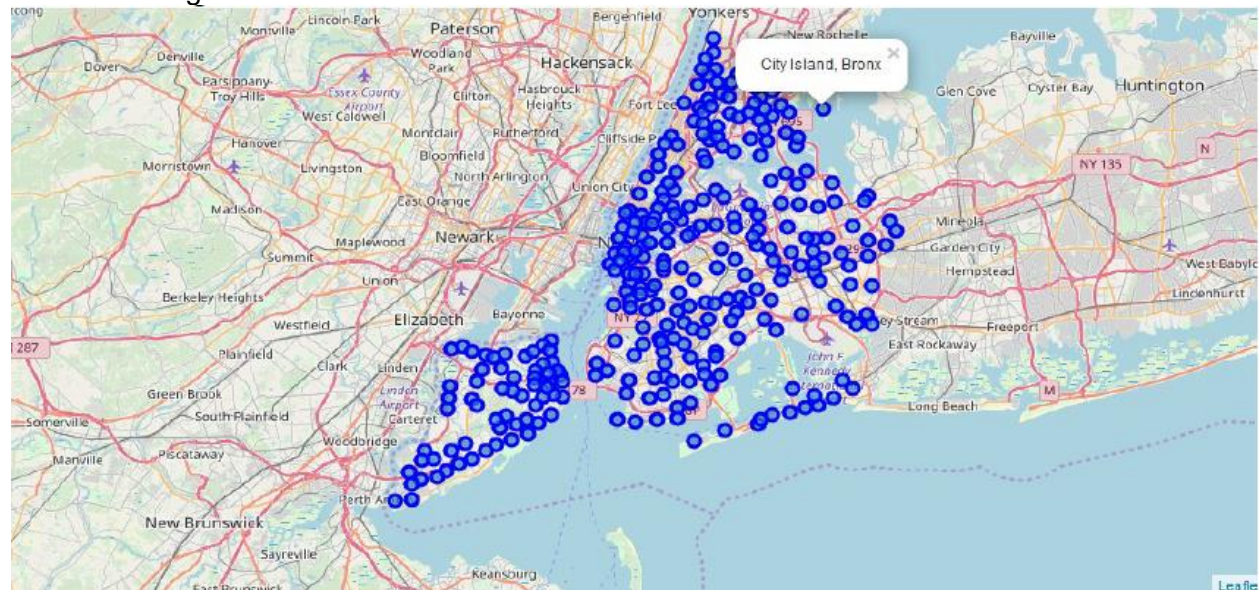
New York city neighbourhood has a total of 5 boroughs and 407 neighborhoods. In this project first part is clustering of Manhattan and Brooklyn . And second part is clustering of Bronx, Queens and Staten Island. This is done because of the following Exploratory data analysis.

Exploratory Data Analysis :

Data 1- New york city Geographical Coordinates Data.

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

New York neighbourhood visualization



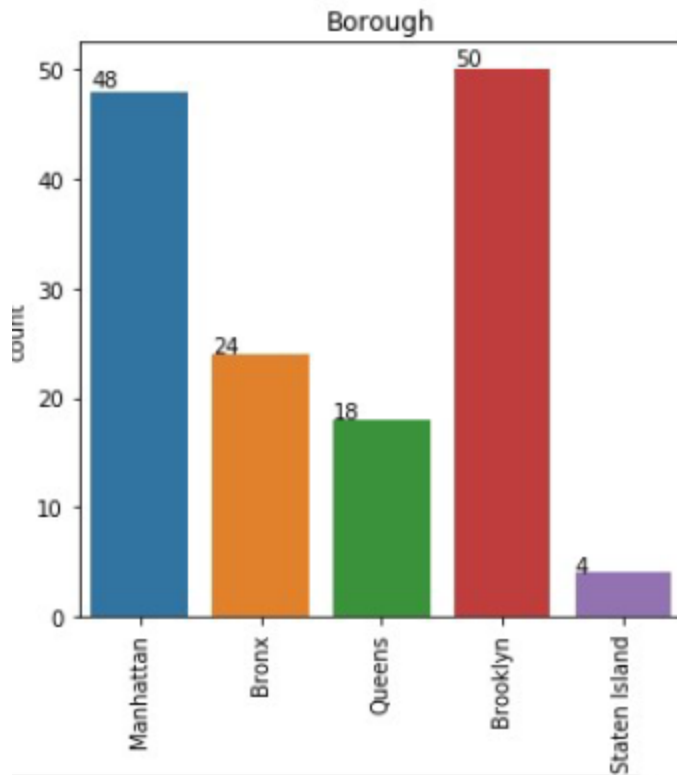
Data 2- Second data which is used is the DOHMH Farmers Markets and Food Boxes dataset. In this we will be using the data of Farmers Markets data.

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.

The proof of this is as given below.

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

Farmers Market visualisation-New York City



Data 3 : To analyze New York city Population, Demographics and Cuisine , scrapped the data from

Wikipedia pages given above in the data section. We used BeautifulSoup python library. BeautifulSoup

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

1. 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_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

2 .New York City Demographics : New York City is the most populous city in the United States,^[9] with an estimated record high of 8,622,698 residents as of 2017,^[7] incorporating more immigration into the city than outmigration since the 2010 United States Census. The racial composition is as given below. This is the reason New York city has restaurants serving cuisine from many countries such as Indian, African, Japan etc. This also increases the scope for restaurants business in New York City.

	Racialcomposition	2010	1990	1970	1940
0	White	44.0%	52.3%	76.6%	93.6%
1	—Non-Hispanic	33.3%	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%	–

Data 3 : NewYork city geographical coordinates data has be 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.

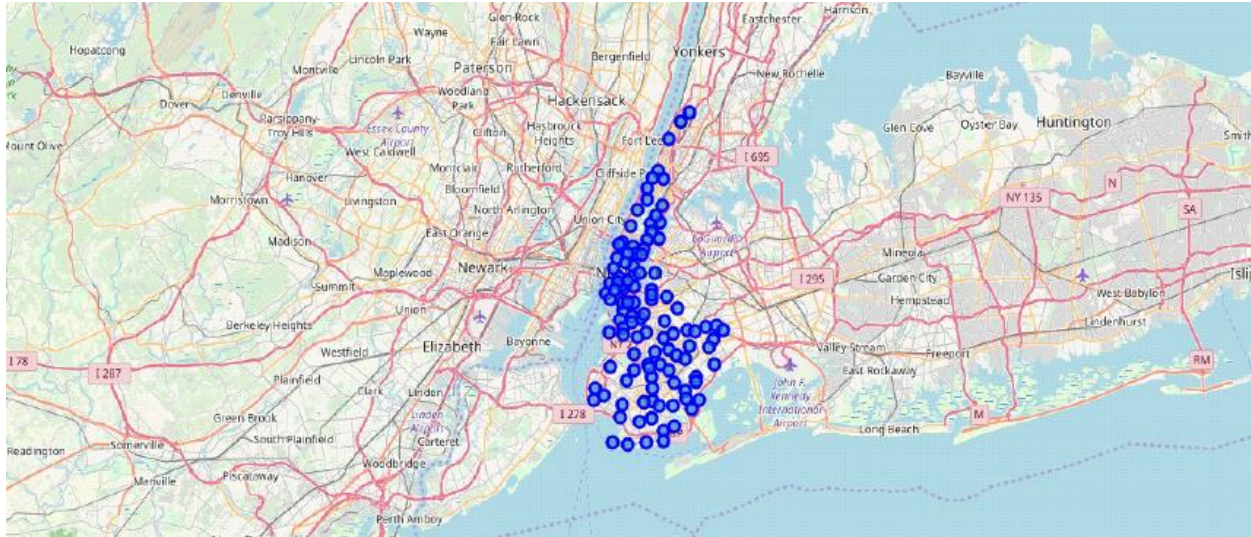
Brooklyn and Manhattan :

Brooklyn and Manhattan Visualization :

Using the geographical coordinates of each neighbourhood foursquare API calls are made to get top

200 venues in a radius of 1000 meters. The venues data is as given below :

Brooklyn and Manhattan Venues :



Brooklyn and Manhattan Venues Visualization : Generated the below Brooklyn and Manhattan Venues Visualization. The "BM_venues" dataframe has 9708 venues and 397 unique venue types.

4.RESULTS :

From this venues data we filtered and used only the restaurant data for Brooklyn & Manhattan clustering. As we focussed 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.

Cluster0 : The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated.

Cluster1 : The Total and Total Sum of cluster1 has highest value. It shows that the markets are saturated. Number of restaurants are very high.

Cluster0 : 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	Brooklyn	Sea Gate	40.576375	-74.007873	0	0

Cluster1 : The Total and Total Sum of cluster1 has highest value. It shows that the markets are saturated. Number of restaurants are very high.

CONCLUSION:

Manhattan has its restaurant market saturated whereas Brooklyn has a neighborhood Sea gate where restaurant could be opened .Also in Manhattan we can raise above the competitors by providing Quality and Tasty foods of the people's desire varity. Also there are many other places in New York which can be explored other that Brooklyn & Manhattan .

.