

The Battle of the Neighborhoods - Report

1.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 oppourtunities 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 excelllent cuisine. It's 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. So it is evident that to survive in such competitive market it is very important to startegically plan. Various factors need to be studied inorder to decide on the Location such as :
 8. New York Population
 9. New York City Demographics
 10. Are there any Farmers Markets, Wholesale markets etc nearby so that the ingredients can be purchased fresh to maintain quality and cost?
 11. Are there any venues like Gyms, Entertainmnet zones, Parks etc nearby where floating population is high etc
 12. Who are the competitors in that location?
 13. Cuisine served / Menu of the competitors
 14. Segmentation of the Borough
 15. Untapped markets
 16. Saturated markets etc

In this case study two such factor - Restaurant availability and farmer markets accesibility will be studied.

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 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> 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 producers for 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.

Data 3 : Newyork city geographical coordinates data will be utilized as input for the Foursquare API,

It will be leveraged to provision venues information for each neighborhood.We will use the Foursquare API to explore neighborhoods in New York City.

3.Methodology :

Business Understanding :

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

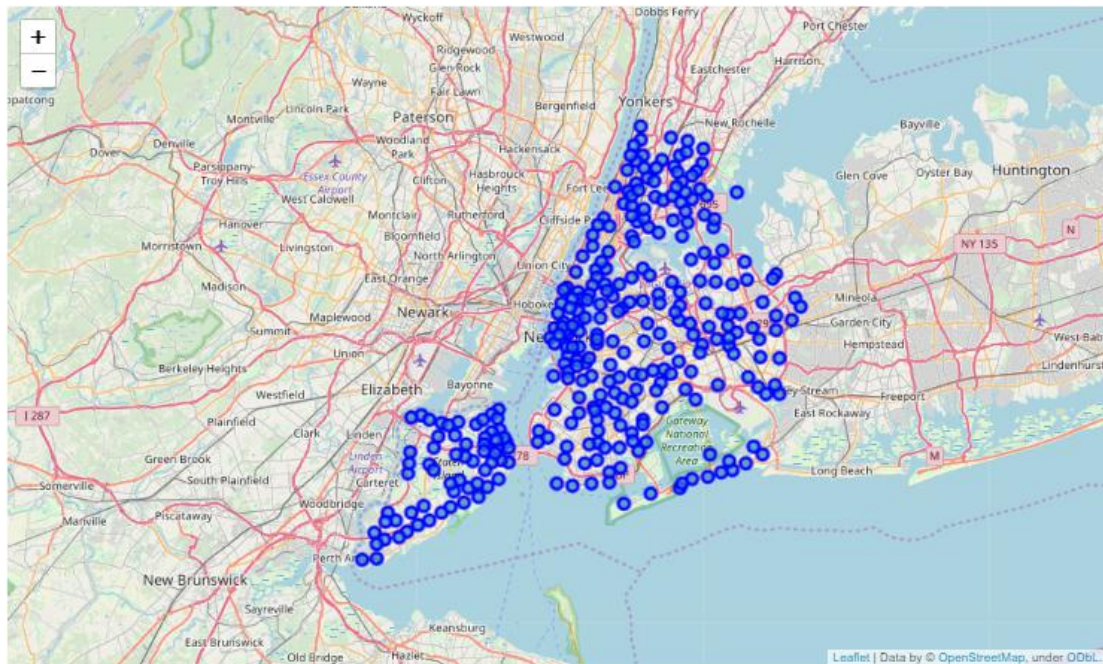
Analytic Approach :

New York city neighbourhood has a total of 5 boroughs and 306 neighborhoods. In this project first part is clustering of the first five neighborhoods . And second part is analyzing the farmer markets data in nyc.

Exploratory Data Analysis :

Data 1- New york city Geographical Coordinates Data.

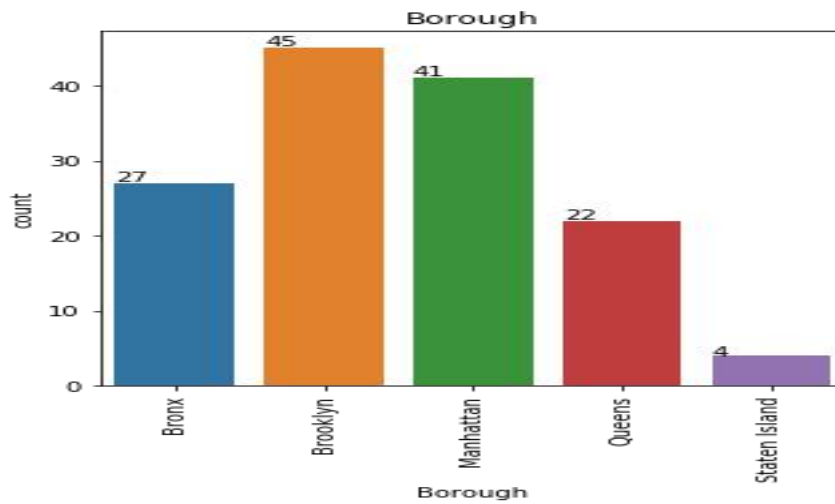
1. In this we load the data and explore data from newyork_data.json file.
2. Transform the data of nested python dictionaries into a pandas dataframe.
3. This dataframe contains the geographical coordinates of New York city neighborhoods.
4. This data will be used to get Venues data from Foursquare.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



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 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.

Farmer Markets by Boroughs



Farmer markets superimposed on the map of new york



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.

The dataframe consisting of the venue data

	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	Ali's Roti Shop	40.894036	-73.856935	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Rite Aid	40.898649	-73.844846	Pharmacy
4	Wakefield	40.894705	-73.847201	Jackie's West Indian Bakery	40.889283	-73.843310	Caribbean Restaurant

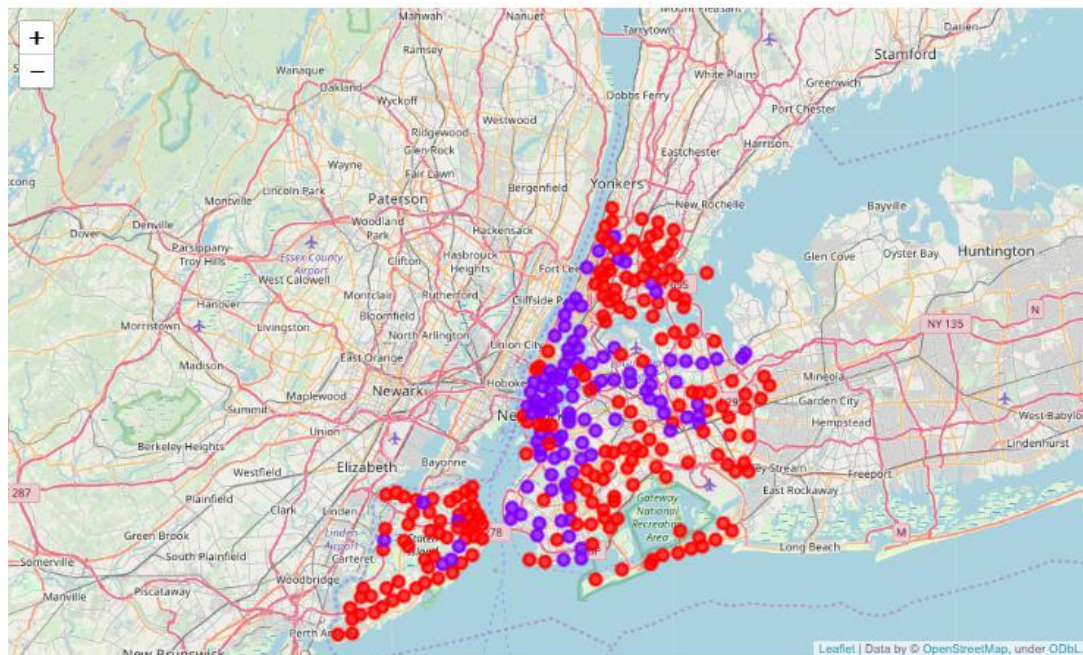
4.RESULTS :

From this venues data we filtered and used only the restaurant data for Brooklyn, Manhattan ,Bronx, Queens and Staten Island 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 app

In the below Map Visualization, we can see the different types of clusters created by using K-Means



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.

There are a few untapped markets i.e in cluster 0 for the following markets

2 *

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Bronx	Clason Point	40.806551	-73.854144	0	0
1	Staten Island	Todt Hill	40.597069	-74.111329	0	0
2	Staten Island	Port Ivory	40.639683	-74.174845	0	0
3	Staten Island	Rossville	40.549404	-74.215729	0	0
4	Staten Island	Bloomfield	40.605779	-74.187256	0	0

6.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 has high concentration of restaurant business. Very competitive market. Bronx, Queens and Staten Island also has good number of restaurants but not as many as required. So this can be explored.

Clason Point, Todt Hill, Port Ivory, Rossville, Bloomfield could be possible neighborhoods to setup a restaurant as there is no competition. Further analysis is required on this front by considering the demographics of these neighborhoods.

The neighborhoods in the cluster 1 are saturated markets, i.e, the competition is high and also the demand for restaurants seems to be high. The neighborhoods in the cluster 0 market seem to be less saturated.

The farmer markets seem to be closer to the neighborhoods in the Boroughs Brooklyn and Manhattan.