The Battle of the Neighborhoods - Report

1.Introduction & Business Problem :

**Problem Background**

City of New York, is the most crowded city in the United States. It is assorted and is the monetary capital of USA. It is multicultural. It gives part of business oppourtunities and business amicable condition. It has pulled in a wide range of players into the market. It is a worldwide center point of business and business. The city is a significant community for banking and fund, retailing, world exchange, transportation, the travel industry, land, new media, conventional media, publicizing, legitimate administrations, bookkeeping, protection, theater, style, and human expressions in the United States. This additionally implies the market is profoundly serious. As it is profoundly created city so cost of doing business is additionally one of the most elevated. In this manner, any new undertaking or development should be examined cautiously. The bits of knowledge got from examination will give great comprehension of the business condition which help in deliberately focusing available. This will help in decrease of hazard. What's more, the Return on Investment will be sensible

**Problem Description:**

An eatery is a business which plans and serves nourishment and drink to clients as a byproduct of cash, either paid before the dinner, after the feast, or with an open record. The City of New York is acclaimed for its excelllent cooking. It's nourishment culture incorporates a variety of global cooking styles affected by the city's worker history.

1. Focal and Eastern European workers, particularly Jewish foreigners - bagels, cheesecake, sausages, knishes, and shops

2. Italian workers - New York-style pizza and Italian cooking

3. Jewish workers and Irish foreigners - pastrami and corned meat

4. Chinese and other Asian eateries, sandwich joints, trattorias, cafes, and cafés are omnipresent all through the city

5. portable nourishment sellers - Some 4,000 authorized by the city

6. Center Eastern nourishments, for example, falafel and kebabs instances of current New York road nourishment

7. It is well known for Pizzerias, Cafe's yet additionally for high end food Michelin featured restaurants.The city is home to "about one thousand of the best and most various haute food eateries on the planet", as per Michelin.

So it is apparent that to make due in such serious market it is critical to startegically design. Different variables should be examined inorder to settle on the Location, for example, :

Eventhough well funded ABC 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.

**Success Criteria:**

success criteria of the project will be a good recommendation of borough/Neighborhood choice to

ABC companyLtd based on Lack of such restaurants in that location and nearest suppliers of

ingredients.

**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 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.



**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

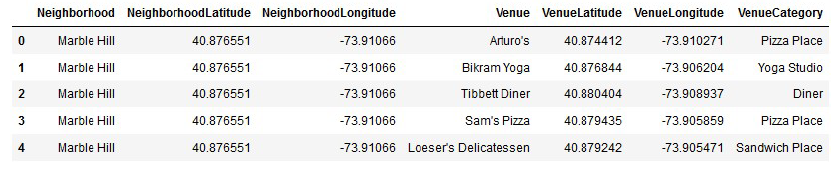
<https://en.wikipedia.org/wiki/Cuisine_of_New_York_City>

**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.



**3.Methodology :**

Business Understanding :

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

Company.

Analytic Approach :

New York city neighbourhood has a total of 5 boroughs and 306 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.

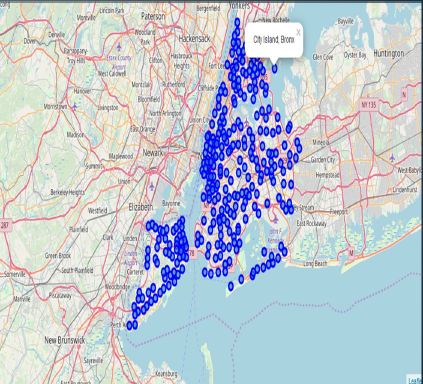
2. Transform the data of nested python dictionaries into a pandas dataframe.

3. This dataframe contains the geographical coordinates of New York city neighborhoods.

4. This data will used to get Venues data from Fouresquare.

5. We used geopy and folium libraries to create a map of New York city with neighborhoods

superimposed on top.

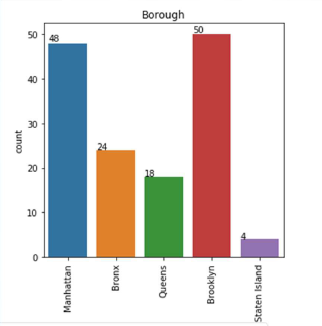


Data 2- Second data which is used is the DOHMH Farmers Markets and Food Boxes dataset. In this we will be using the data of Farmers Markets data.

There are totally 144 Farmers Markets in New York city. Highest number are in Manhattan and

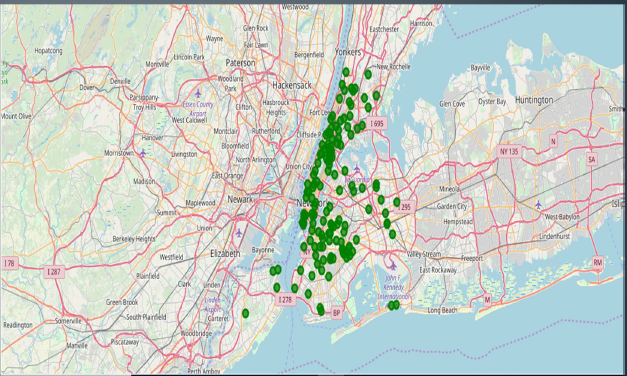
Brooklyn.And lowest in Queens, Bronx and Staten Island.

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 analyize New York city Population, Demographics and Cuisine , 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

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.



2.New York City Demographics : New York City is the most populous city in the United States,[9]

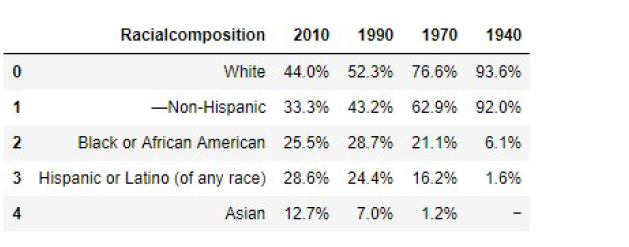
with an estimated record high of 8,622,698 residents as of 2017,[7] incorporating more immigration

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.



3.Cuisine of New York city : This data has been manually prepared. Data is taken from Wikipedia

page - https://en.wikipedia.org/wiki/Cuisine\_of\_New\_York\_City . Using this data we did word cloud.

NEW YORK CITY CUISINE : Most Preferred Food in New York City –Italian, Purto Rican,

Mexican, Jewish, Indian, Pakistani & Dominican.

BROOKLYN CUISINE -Most Preferred Food in Brooklyn is –Italian, Purto Rican & Mexican

MANHATTAN CUISINE - Most Preferred Food in Manhattan is – Italian, American, Puerto Rican and Indian.

QUEENS CUISINE - Most Preferred Food in Queens is – Indian, Irish, Pakistani and Mexican.

THE BRONX CUISINE - Most Preferred Food in The Bronx is – Italian,Puerto Rican, Albanian and Dominican.

There is very less data of cuisine relating to Staten Island. So could not develop word cloud with it.

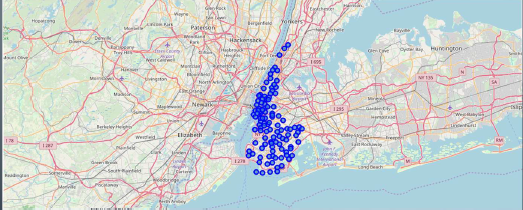
Data 4 : NewYork city geographical coordinates data has be utilized as input for the Foursquare

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

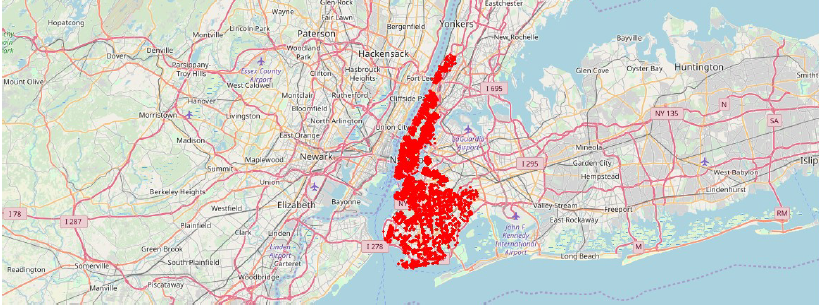


Brooklyn and Manhattan Venues

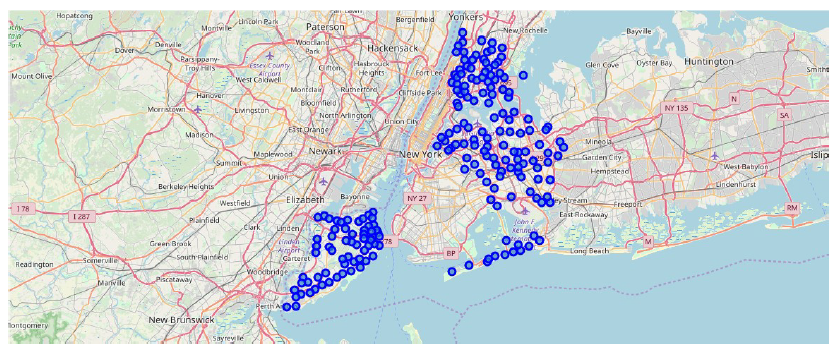


Brooklyn and Manhattan Venues Visualization : Generated the below Brooklyn and Manhattan

Venues Visualization. The "BM\_venues" dataframe has 9708 venues and 397 unique venue types.



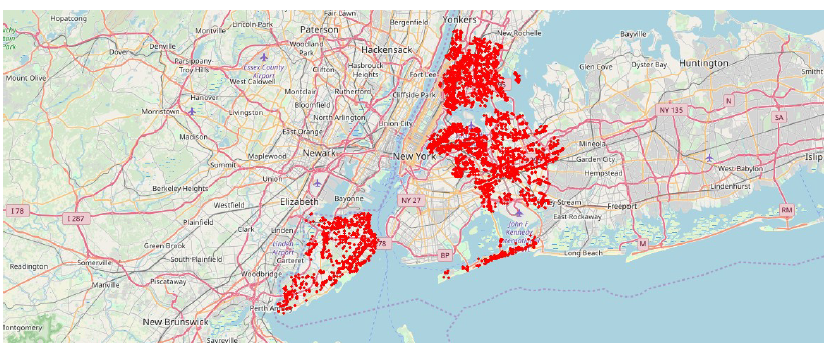
Bronx, Queens and Staten Island :

Bronx, Queens and Staten Island Neighborhoods Visualization :****

Bronx, Queens and Staten Island Venues Visualization : The "BQS\_venues" dataframe has 10805

venues and 387 unique venue types

****

****

**4.RESULTS :**

From this venues data we filtered and used only the restaurant data for Brooklyn & Manhattan

clustering and 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

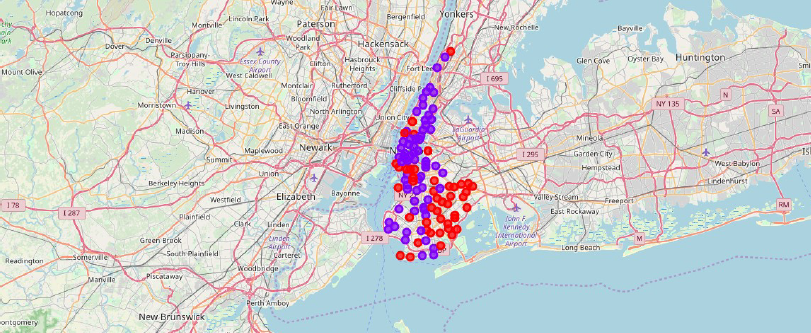
clustering aims to partition n observations into k clusters in which each observation belongs to the

cluster with the nearest mean. It uses iterative refinement approach.

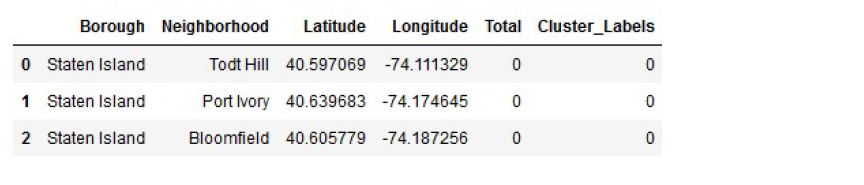
Brooklyn & Manhattan :

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.

There are no untapped neighborhoods in Brooklyn and Manhattan.

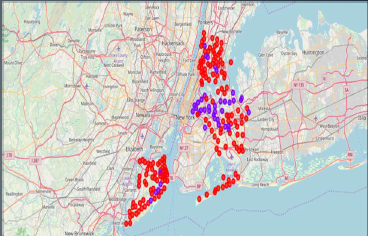
Bronx, Queens and Staten Island :

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

for Bronx, Queens and Staten Island.

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

****

**5.DISCUSSION**

**1**. There is extension to build Farmers advertises in Bronx, Queens and Staten Island.

2. There is degree to investigate cooking styles of different nations in Bronx, Queens and Staten Island.

3. In Manhattan and Brooklyn cafés of cooking styles of numerous nations are accessible. So if chance

can be accepted with incredible menu. It likewise shows individuals love eating cooking styles of different nations.

**6.CONCLUSION**

This investigation is performed on restricted information. This might be correct or might not be right. Be that as it may, if great measure of information is accessible there is degree to think of better outcomes. In the event that there are parcel of cafés presumably there is part of interest. Brooklyn and Manhattan has high centralization of resturants business. Very serious market. Bronx, Queens and Staten Island likewise has great number of Resturants however not as numerous as required. So this can be investigated. According to the area or restaurants type referenced like Indian Restaurant investigation can be checked. A setting with most reduced hazard and rivalry can be distinguished.