# THE BATTLE OF NEIGHBOURHOODS


## Abstract

This study will help business to understand New York untapped market to start grocery store



Charudatta Rahate

# CONTENTS

# 1. Introduction & Business Problem

## 1.1. Problem Background

The aim of this project is to find a safe and secure location for opening of Grocery shop in New York. The New York City often called New York (NY), is the most populous city in the United States. City is the centre of the New York metropolitan area, the largest metropolitan area in the world by urban landmass. It is diverse and is the financial capital of not only USA but also of the world. New York has emerged as a global node of creativity and entrepreneurship. It provides lot of business opportunities and business friendly environment. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world (Turin, n.d.). It has attracted many different players into the market. It is a global hub of business and commerce. New York is major capital for any business in US.
As a result, market in City is highly competitive. If you are looking to start a business, then cost will be one of the highest. Thus, market analysis is very important for any business starting new setup or expanding existing business. After the analyses, the insights getting from it will give good understanding to entrepreneur and will help them to derive their strategy to tap the market.

## 1.2. Problem Description

A grocery store or grocer's shop is a retail shop that primarily sells food, either fresh or preserved. Some groceries specialize in the foods of a certain nationality or culture, such as Indian, Chinese, Italian, Middle Eastern or Polish. These stores are known as ethnic markets.
There are different types of grocery store in New York, including supermarket, mega market, ethnic market. And you can get at variety of prices.
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 choose on the Location such as:

- New York Population
- New York City Demographics
- Are there any Super Markets, Wholesale markets etc nearby?
- Who are the competitors in that location?
- Segmentation of the Borough
- Untapped markets
- Saturated markets etc
- The list can go on...

Even though ABC Company Ltd. need to choose not only the correct location but also need to target the community to start its first venture for supermarket. If this is successful, they can replicate the same in other locations. First move is very important, thereby choice of location and type of product selection is very important.

## 1.3. Target Audience

To recommend the precise location along with choice of product, ABC Company Ltd has selected me to lead of the Data Science team. The objective is to locate and recommend to the management which neighbourhood of New York city will be best choice to start Supermarket. The Management also expects to understand the justification of the recommendations made.

This would interest anyone who wants to start a new Grocery store in New York city.

## 1.4. Success Criteria

The success criteria of the project will be a good recommendation of borough/Neighbourhood choice to ABC Company Ltd based on absence of such grocery store in that location.

# 2. DATA

Analysis done for: New York City.

We will be using the below datasets for analysing New York city

## 2.1. NEIGHBOURHOOD DATA

Neighbourhood 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 neighbourhood.

This dataset exists for free on the web. Link to the dataset is:

https://geo.nyu.edu/catalog/nyu_2451_34572

## 2.2. GROCERY STORE DATA

Second data which will be used is the grocery store present in each borough. In this we will be using the data of Recognized Shop Healthy Stores.

https://data.cityofnewyork.us/Health/Recognized-Shop-Healthy-Stores/ud4g-9x9z (2)

## 2.3. POPULATION DATA

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

1. New York Population
2. New York City Demographics

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

## 2.4. GEOGRAPHICAL DATA

New York city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighbourhood. We will use the Foursquare API to explore neighbourhoods in New York City.

# 3. Methodology

## 3.1 Business Understanding

The aim of this project is to find a safe and secure location for opening of Grocery store in New York for ABC Company Ltd. Where business need to understand population and demographic of city, ethnic of society and untapped location.

## 3.2 Analytic Approach

New York city neighbourhood has a total of 5 boroughs and 306 neighbourhoods. In this project first part is to segment the neighbourhoods and explore them, gets their location data. Second part is to understand the healthy grocery store present in the vicinity. Third part is to understand population information in the city and in last part we will analyse different borough to find untapped market location.

## 3.3 Exploratory Data Analysis

### 3.3.1 Data 1- New York city Geographical Coordinates Data.

In this we 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 neighbourhoods. This data will used to get venues data from Foursquare. We used geopy and folium libraries to create a map of New York city with neighbourhoods superimposed on top.

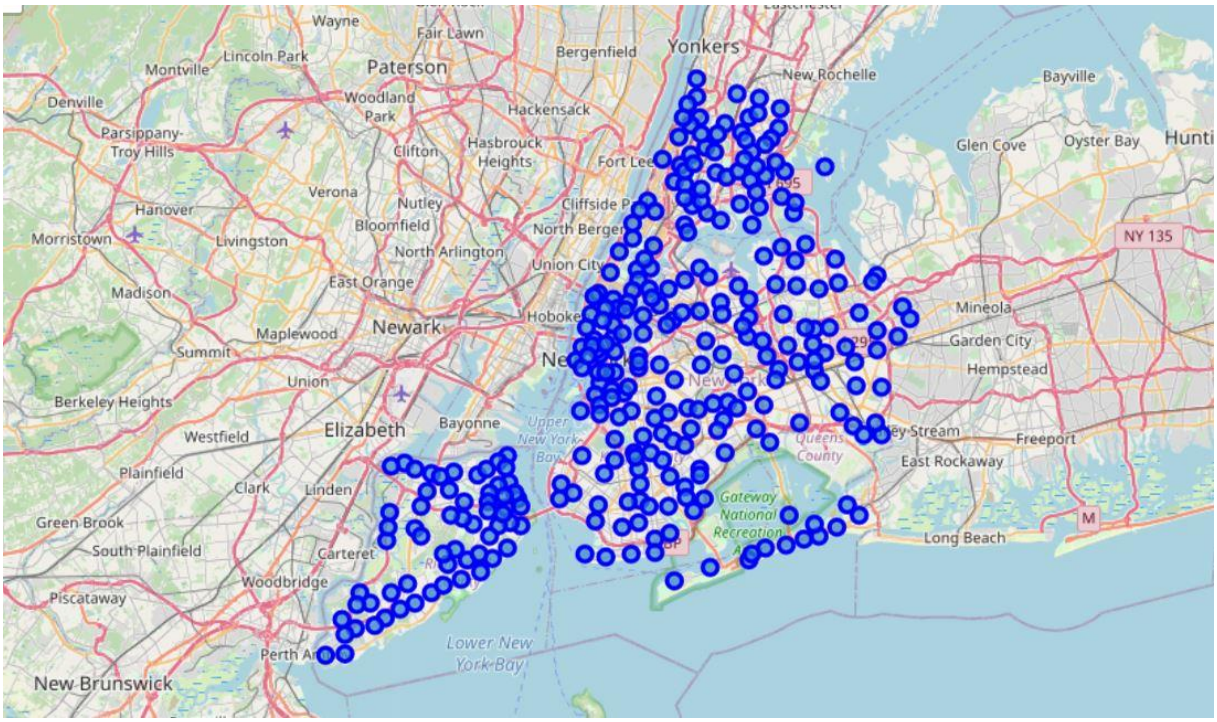


**FIGURE 1: NEW YORK NEIGHBOURHOOD VISUALIZATION**

### 3.3.2 Data 2: Healthy Grocery Store data

Second data which is used is the Healthy grocery store dataset. There are totally 368 grocery store marked as selling healthy food item to consumers. Highest number are in Bronx and Brooklyn. And lowest in rest of the Newy York city (i.e. Manhattan, Queens, and Staten Island). The proof of this is as given below.


**FIGURE 2: HEALTHY GROCERY STORE IN NEW YORK**

### 3.3.3 Data 3: New York city Population and Demographics Data

To analyse New York city Population and Demographics, scrapped the data from Wikipedia pages given above in the data section. We used BeautifulSoup python library. BeautifulSoup is a Python package for parsing HTML and XML documents (including having malformed mark-up, 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

**New York Population : Insights from the data**

- Queens (Queens County), on Long Island north and east of Brooklyn, is geographically the largest borough
- Manhattan (New York County) is the geographically smallest borough
- Brooklyn (Kings County) and Queens (Queens County), are city's most populous borough.
- Staten Island has lowest population as well as density amongst the other borough.

| | Borough | County | Estimate_2019 | square_mi | persons_sq_mi |
|---|---|---|---|---|---|
| **0** | The Bronx | Bronx | 1,418,207 | 42.10 | 109.04 |
| **1** | Brooklyn | Kings | 2,559,903 | 70.82 | 183.42 |
| **2** | Manhattan | New York | 1,628,706 | 22.83 | 59.13 |
| **3** | Queens | Queens | 2,253,858 | 108.53 | 281.09 |
| **4** | Staten Island | Richmond | 476,143 | 58.37 | 151.18 |
| **5** | | City of New York | 8,336,817 | 842.343 | 783.83 |
| **6** | | State of New York | 19,453,561 | 1,731.910 | 122,284 |

**FIGURE 3: NEW YORK POPULATION DISTRIBUTION**

**New York City Demographics**:

New York City is the most populous city in the United States, with an estimated record high of 8,336,817 residents as of 2019 incorporating more immigration into the city than outmigration since the 2010 United States Census. The racial composition is as given below.

| | Year | White(includes White Hispanics)% | Non-Hispanic Whites % | Black % | Asian % | Other orMixed % | Hispanic/Latino % | %FB |
|---|---|---|---|---|---|---|---|---|
| **0** | 2000 | 44.66 | 34.98 | 26.59 | 9.90 | 18.85 | 26.98 | 35.85 |
| **1** | 2010 | 44.00 | 33.31 | 25.55 | 12.77 | 17.68 | 28.58 | 37.51 |

**FIGURE 4: DEMOGRAPHIC OF NEW YORK CITY**

### 3.3.4 Data 4: New York City Venue Information

New York city geographical coordinates data has be utilized as input for the Foursquare API, that has been leveraged to provision venues information for each neighbourhood. We used the Foursquare API data to explore neighbourhoods in New York City.

From above analysis we learned that grocery store market in Bronx and Brooklyn are already saturated, so we will analyse venue information for Manhattan, Queens and Staten Island borough.

**Manhattan, Queens and Staten Island Venues Visualization** : The "MQS_venues" dataframe has 10805 venues and 387 unique venue types.

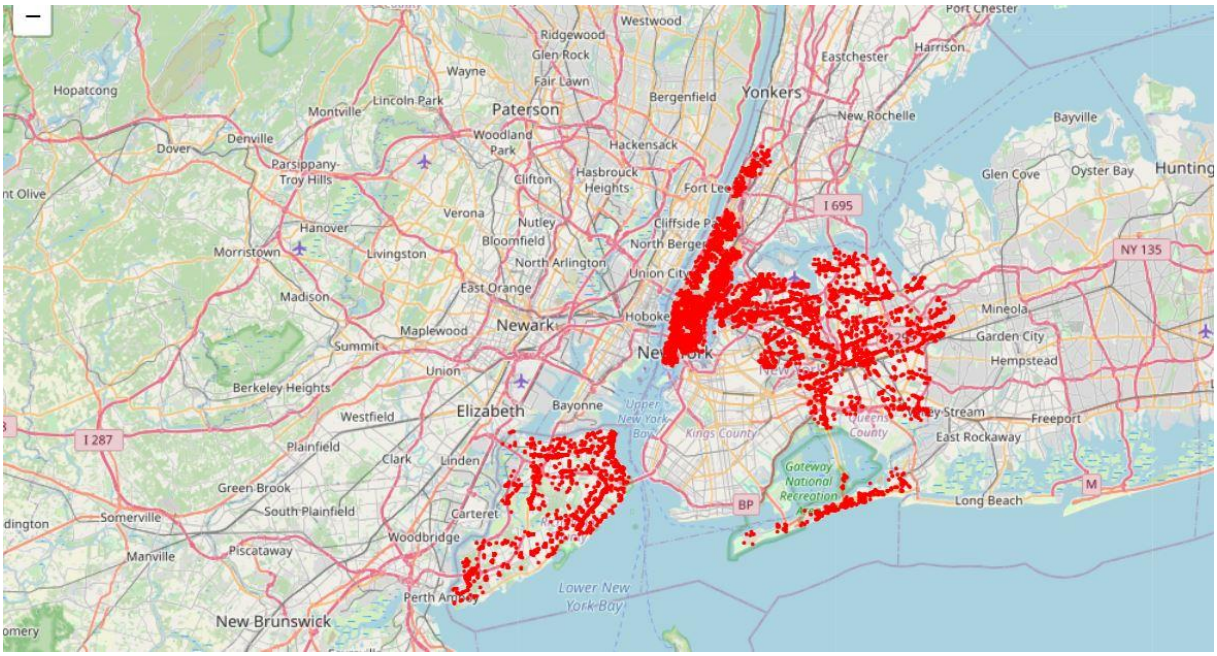


**FIGURE 5: MANHATTAN QUEENS AND STATEN ISLAND VENUES MAP VISUALIZATION**

# 4. RESULT

From this venues data we filtered and used only the grocery store and supermarket data for Manhattan, Queens and Staten Island clustering. As we focussed only on grocery store business.

**Neighbourhood K-Means clustering based on mean occurrence of venue category** :

To cluster the neighbourhoods into 9 clusters we used the K-Means clustering Algorithm. k-means clustering aims to partition and observations into k clusters in which each observation belongs to the cluster with the nearest mean. It uses iterative refinement approach.

In the below Map Visualization, we can see the different types of clusters created by using K-Means for Manhattan, Queens and Staten Island.

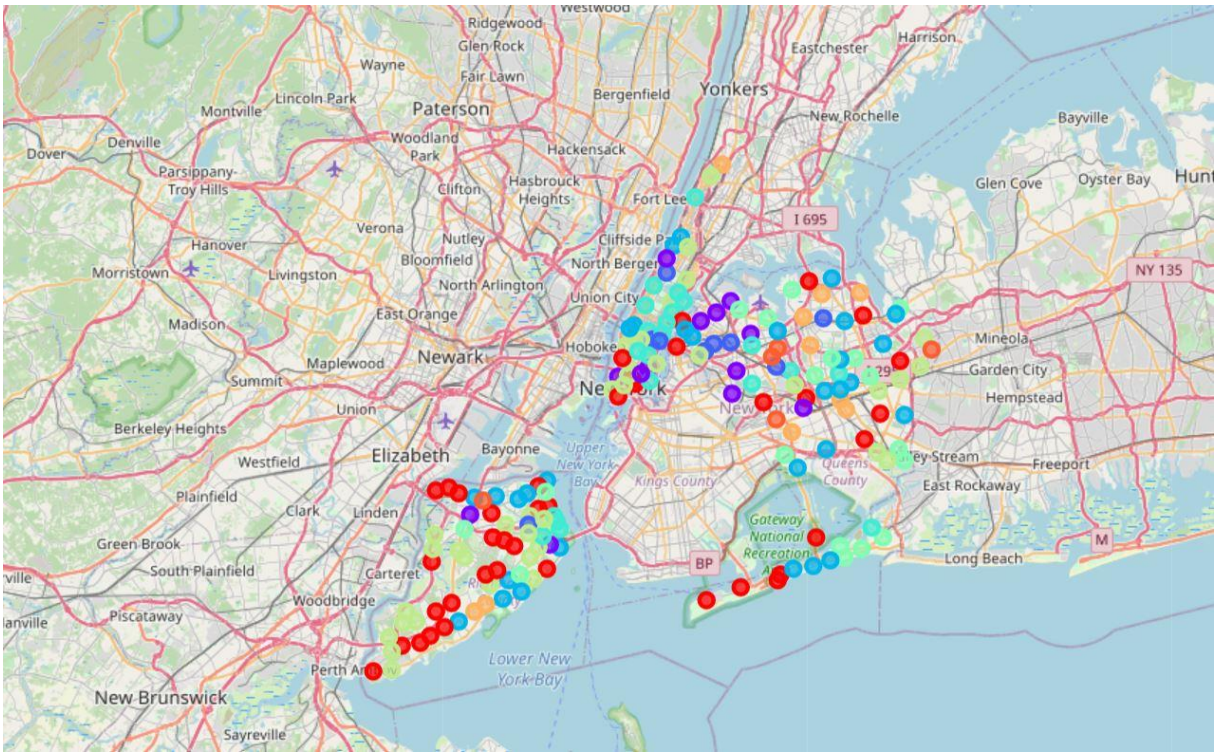


**FIGURE 6: MANHATTAN, QUEENS, STATEN ISLAND CLUSTER**

Out of 9 cluster created in Cluster0 the Total (0.051282) and Total Sum (0.102564 ) has smallest value. It shows that the market is not saturated.

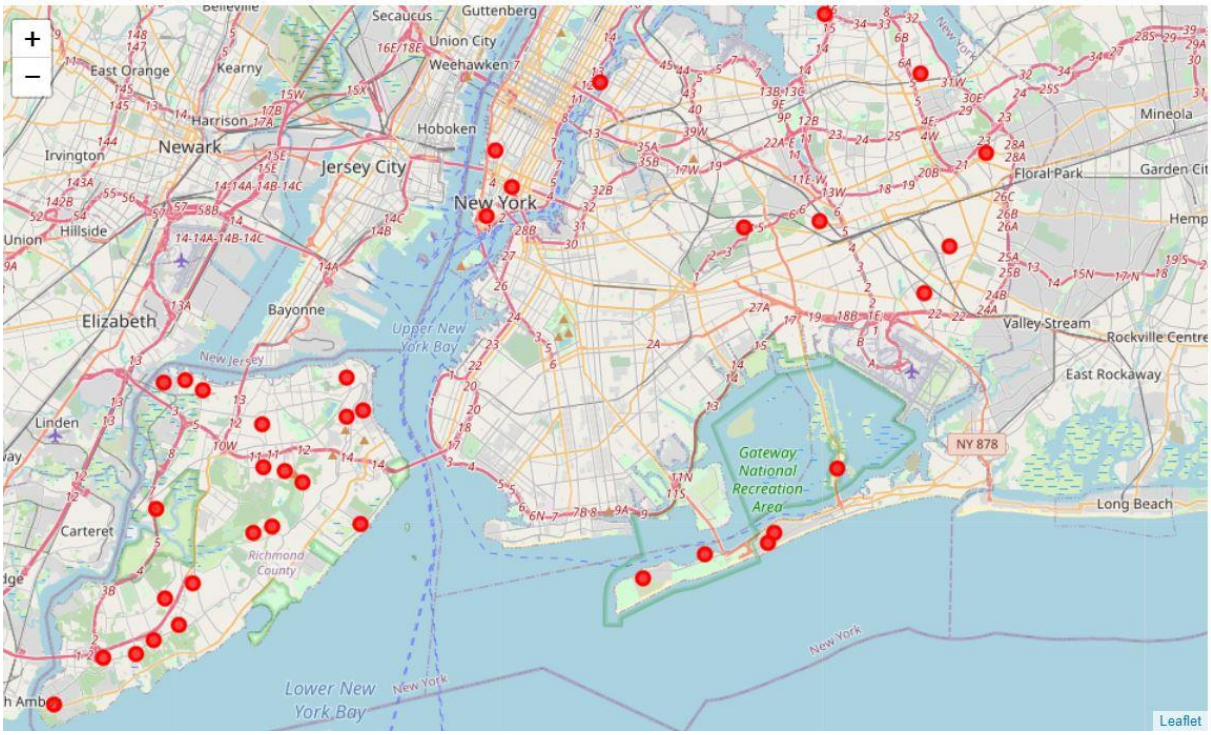


**FIGURE 7: CLUSTER0 VISUALISATION ON MAP**

Based on this result location in Queens borough will be most suited for upcoming grocery store

| Borough | Neighborhood | Latitude | Longitude |
|---|---|---|---|
| Queens | Kew Gardens | 40.705179 | -73.829819 |
| Queens | Glendale | 40.702762 | -73.870742 |
| Queens | Bayside | 40.766041 | -73.774274 |
| Queens | St. Albans | 40.694445 | -73.758676 |
| Queens | Rochdale | 40.675211 | -73.772588 |
| Queens | Broad Channel | 40.603027 | -73.820055 |
| Queens | Breezy Point | 40.557401 | -73.925512 |
| Queens | Neponsit | 40.572037 | -73.857547 |
| Queens | Belle Harbor | 40.576156 | -73.854018 |
| Queens | Bellaire | 40.733014 | -73.738892 |
| Queens | Roxbury | 40.567376 | -73.892138 |
| Queens | Malba | 40.790602 | -73.826678 |

**FIGURE 8: SUGGESTED PLACES FOR GROCERY STORE**

# 5. Discussion

In this study, I analysed the population of New York, how it is distributed and how different types of ethnic present with in the city.

For grocery store it is important to understand population density on which they can decide size of the store. Ethnic background will give them types of products they can keep.

From study we can see that there is lack of Grocery store selling Healthy food item in Manhattan, Queens and Staten Island. Out of these 3 borough Queens borough has highest population, so Queens borough can be right place to start grocery store where we can sell healthy food items.

One more interesting fact need to be check here is that, population from Asian country is increasing rapidly and as grocery store business they need to tapped this opportunity by adding more Asian products in there selling list.

# 6. Conclusion

Model in study is based on very limited available data, But if good amount of data is available there is scope to come up with better results.

We concluded the lack of Healthy food item availability in area but need more data to understand the reason. Is that area is more industrial one or any restriction present in this area. Also regarding ethnic back ground there is not enough data to understand in which area or neighbourhood particular culture has more influence.

This study will help ABC company Ltd. To narrow down there location on ‘Queens’ borough and try to get more data mention above to lower the risk of failure.

# References

(n.d.). Retrieved from https://www.businessinsider.com/queens-languages-map-2017-2?r=US&IR=T

(n.d.). Retrieved from https://data.cityofnewyork.us/Health/Recognized-Shop-Healthy-Stores/ud4g-9x9z

Turin, M. (n.d.). *The World's Most Linguistically Diverse Location? New York City*. Retrieved 6 9, 2020, from PopAnth: http://popanth.com/article/the-worlds-most-linguistically-diverse-location-new-york-city