The Battle of the Neighborhoods - Week 1

Introduction & Business Problem

Imagine the company "Butterscotch Pancakes" from Russia. It is a chain of restaurants of Russian cuisine, specialized in Pancakes with different toppings. A CEO of a company has a dream - to enter a US market. In a cold winter of 2019, he decided to pursuit his dream! To open a restaurant, one need to choose a city, populated enough for a restaurant to succeed. Our CEO has chosen New York, because of big population of ex-soviet and Russian Immigrants.

The food market in NY is highly competitive. So, there has to be a thorough analysis of business environment to form a strategy. This will help to reduce the risk of restaurant failure.

Our goal is to find an optimal location for the Restaurant.
There is a huge variety of food on the streets of New York:

- Fast food (hot dogs, bagels, ice cream, burgers etc.)
- Italian restaurants
- Asian restaurants (Thai, Chinese, Indian etc.)
- Coffee shops
- Middle Eastern restaurants.

Various factors need to be studied in order to decide on the Location such as:

- New York Population
- New York City Demographics
- Sources of ingredients
- Popular attractions nearby
- Competitors
- Segmentation of neighborhoods (Boroughs) and so on...

The objective of this project is to deliver a recommendation of which neighborhood of NY will be the best choice to build the restaurant.

Success Criteria:
The success criteria of the project will be a good recommendation of Neighborhood or Borough choice for a new restaurant to CEO based on lack of such restaurants in that location and nearest suppliers of ingredients.

This project can be used by anyone who looking forward of opening a restaurant in any city.

Data

1 - New York City Neighborhood names (https://geo.nyu.edu/catalog/nyu_2451_34572)

2 - The list of farmers market of NY (https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets/8vwk-6iz2)

3 - Location data of Fresh food box. 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 (https://www.grownyc.org/greenmarketco/foodbox)

4 - Wikipedia (data on population, economy, demographics, cuisine etc.)

5 - Foursquare API.

Methodology

Business Understanding

Our goal is to find an optimal location for the Restaurant "Butterscotch Pancakes" in New York City.

Analytic Approach

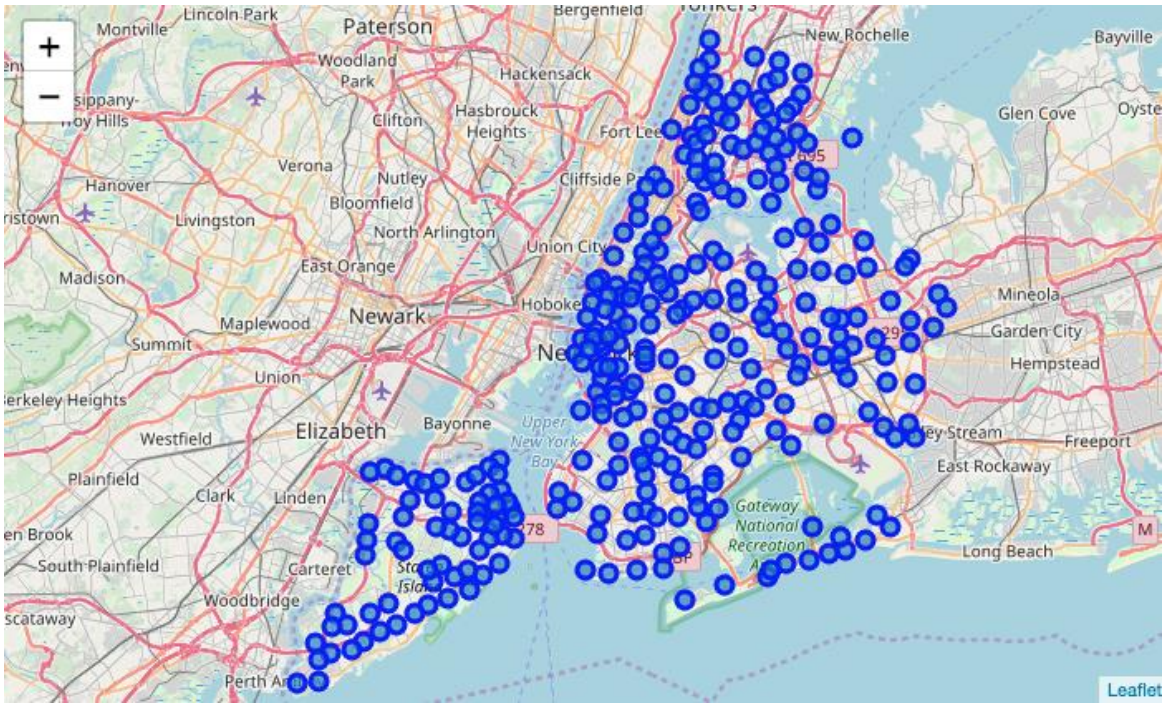
New York city neighborhood has a total of 5 boroughs and 306 neighborhoods. In this project we will cluster Manhattan and Brooklyn, Bronx, Queens and Staten Island. This is done because of the following Exploratory data analysis.

Exploratory Data Analysis

Data 1- New york city Geographical Coordinates Data.

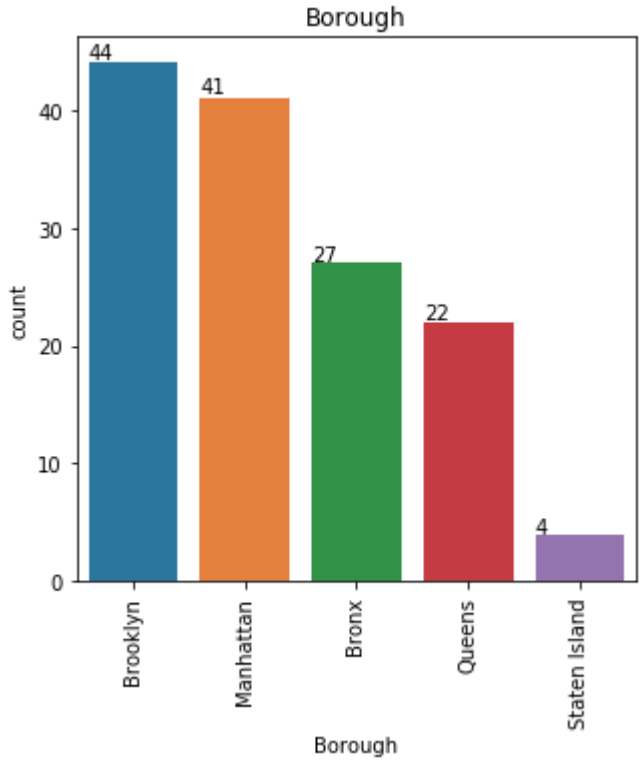
1. We load and explore the data from newyork_data.json file.
2. Transform the data into a pandas dataframe.
3. This dataframe contains the geographical coordinates of New York city neighborhoods.
4. This data will be used to get Venues data from Foursquare with its API.
5. We used Geopy and Folium libraries to create a map of New York city with neighborhoods.

New York Neighborhood visualization

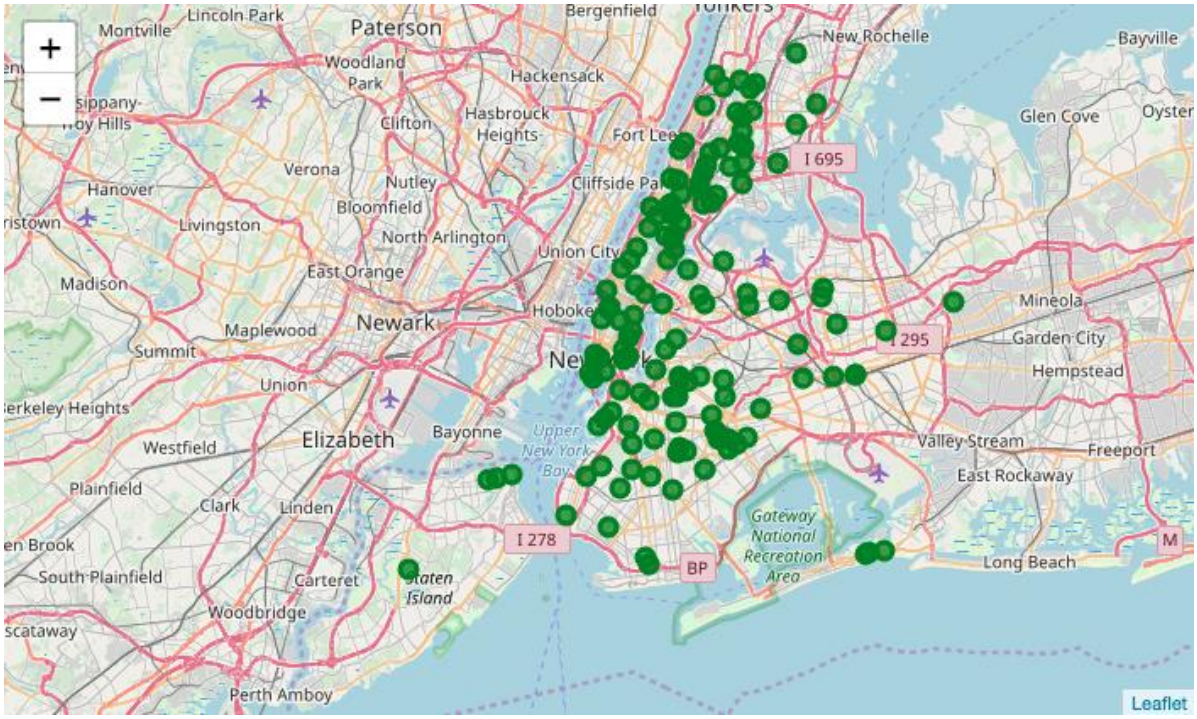


Data 2- Second data which is used is the DOHMH Farmers Markets dataset.

There are totally 138 Farmers Markets in New York city. Highest number are in Manhattan and Brooklyn. And lowest in Queens, Bronx and Staten Island.



We used Geopy and Folium libraries to create a map of farmers markets in 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 Beautiful Soup 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

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.

New York City Demographics:
New York City is the most populous city in the United States, with an estimated record high of 8,622,698 residents as of 2017, incorporating more immigration into the city than outmigration since the 2010 United States Census.

Cuisine of New York city

This data has been manually sourced from Wikipedia page - https://en.wikipedia.org/wiki/Cuisine_of_New_York_City . Using this data we created word cloud.

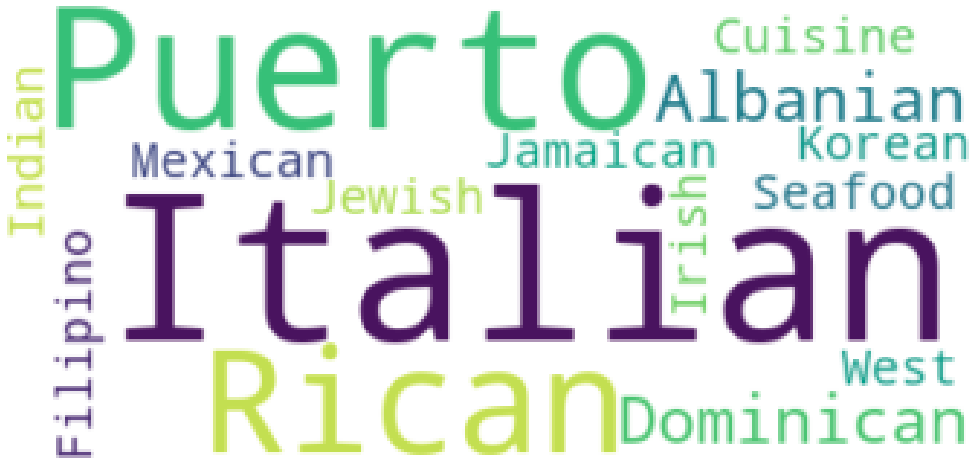
NEW YORK CITY CUISINE: Most Preferred Food in New York City – Italian, Mexican, Puerto Rican, Jewish, Chinese.

BROOKLYN CUISINE - Most Preferred Food in Brooklyn is – Italian, Puerto Rican, Mexican, Chinese, Jamaican and Russian.

MANHATTAN CUISINE - Most Preferred Food in Manhattan is – Indian, Pakistani, Italian.

QUEENS CUISINE - Most Preferred Food in Queens is – Italian, Chinese, and Puerto Rican.

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

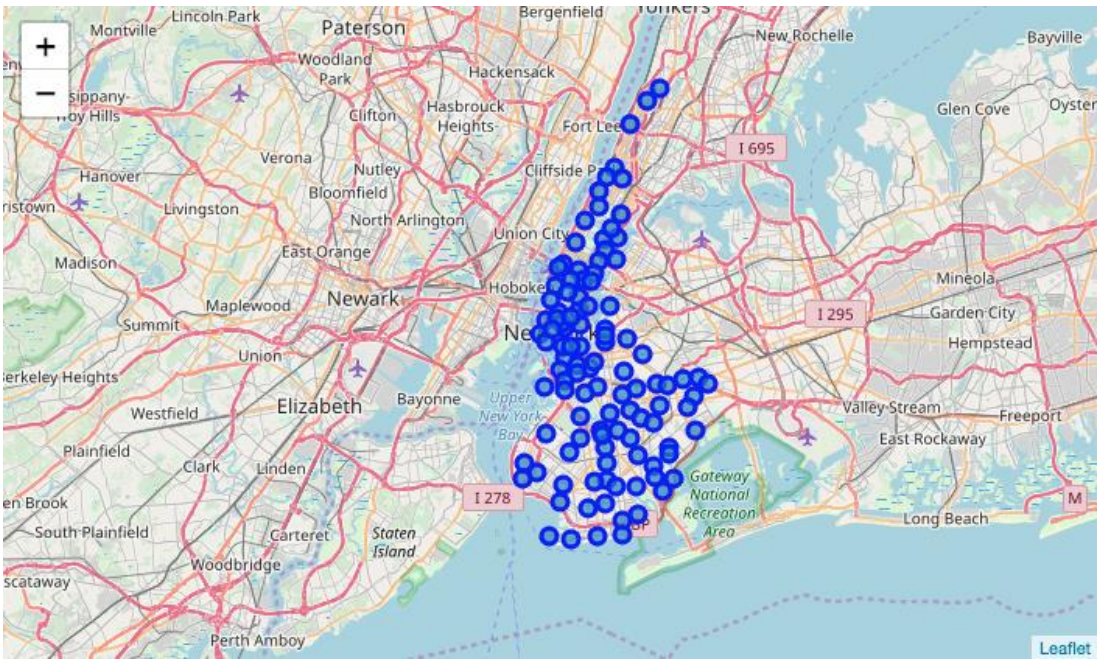


Data 4: New York city geographical coordinates data must be utilized as input for the Foursquare API, that has been leveraged to provision venues information for each neighborhood. We used the Foursquare API data to explore neighborhoods in New York City.

Brooklyn and Manhattan:

Brooklyn and Manhattan Visualization:

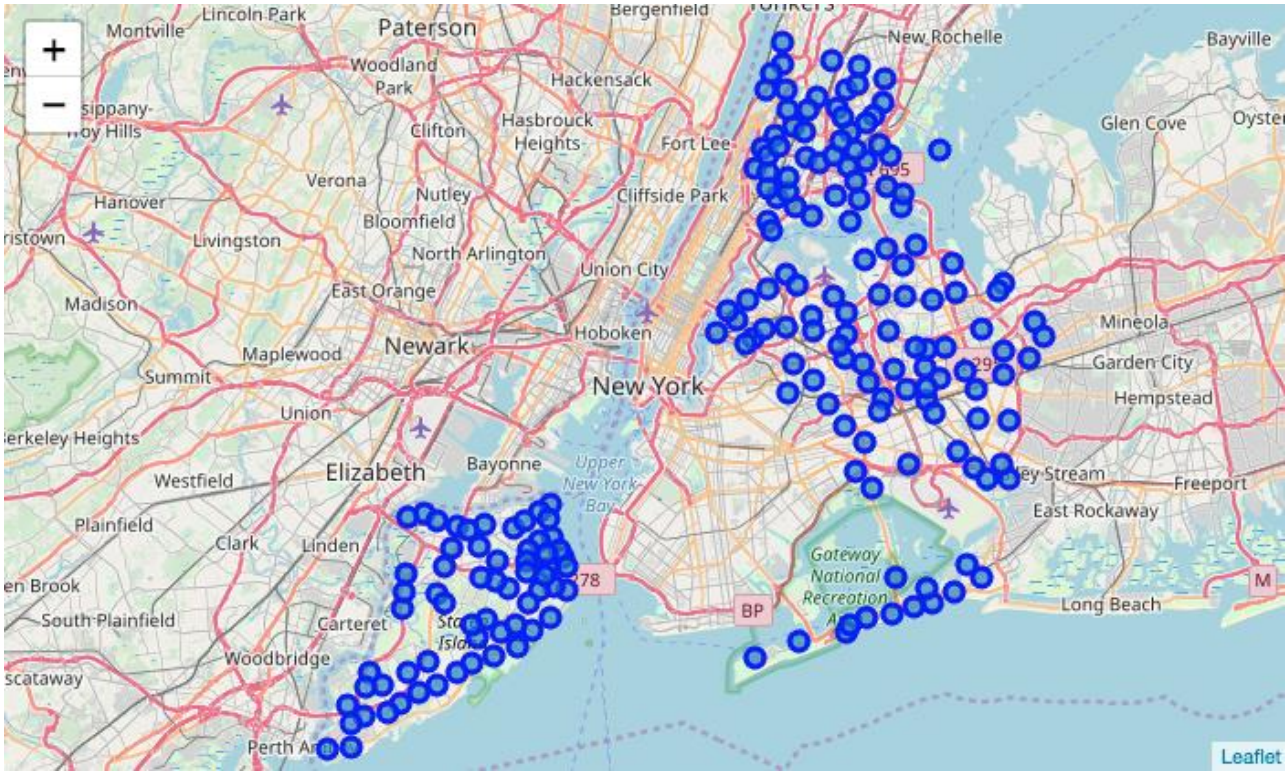
Brooklyn and Manhattan Venues Visualization: Generated the below Brooklyn and Manhattan Venues Visualization. The "BM_venues" dataframe has 9670 venues and 407 unique venue types.



Using the geographical coordinates of each neighborhood foursquare API calls are made to get top 200 venues in a radius of 1000 meters.

Bronx, Queens and Staten Island:

Bronx, Queens and Staten Island Neighborhoods Visualization:



The "BQS_venues" dataframe has 10911 venues and 395 unique venue types.

Results

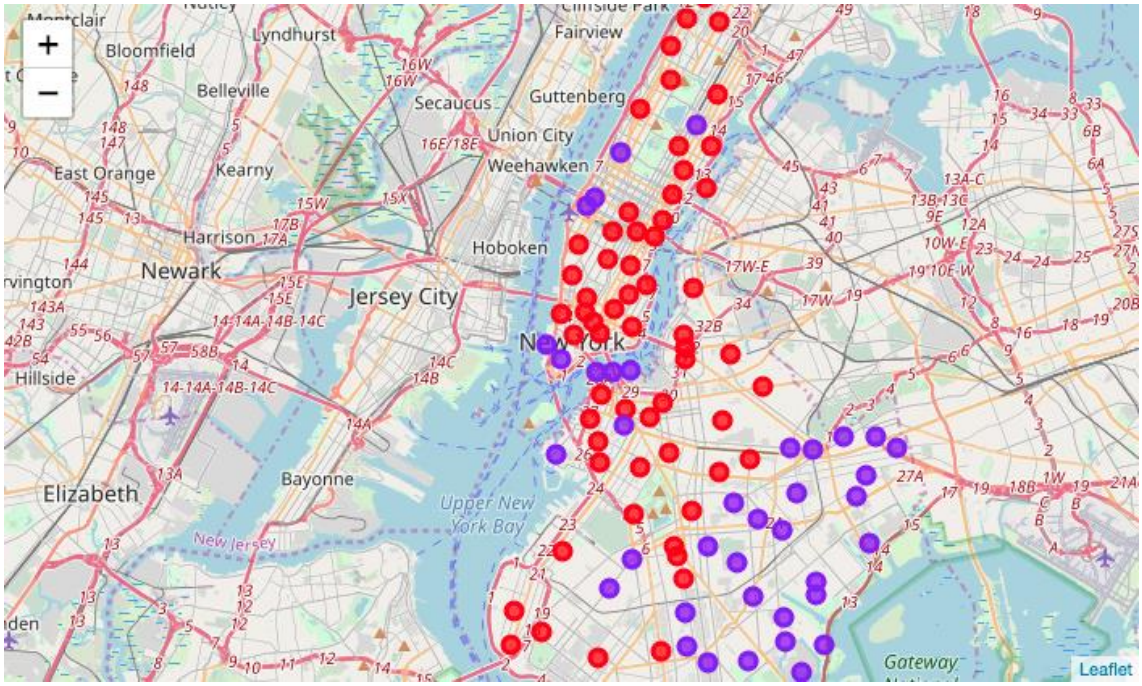
From this venues data we filtered and used only the restaurant data for Brooklyn & Manhattan clustering and Bronx, Queens and Staten Island clustering. As we focused only on restaurants business.

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.

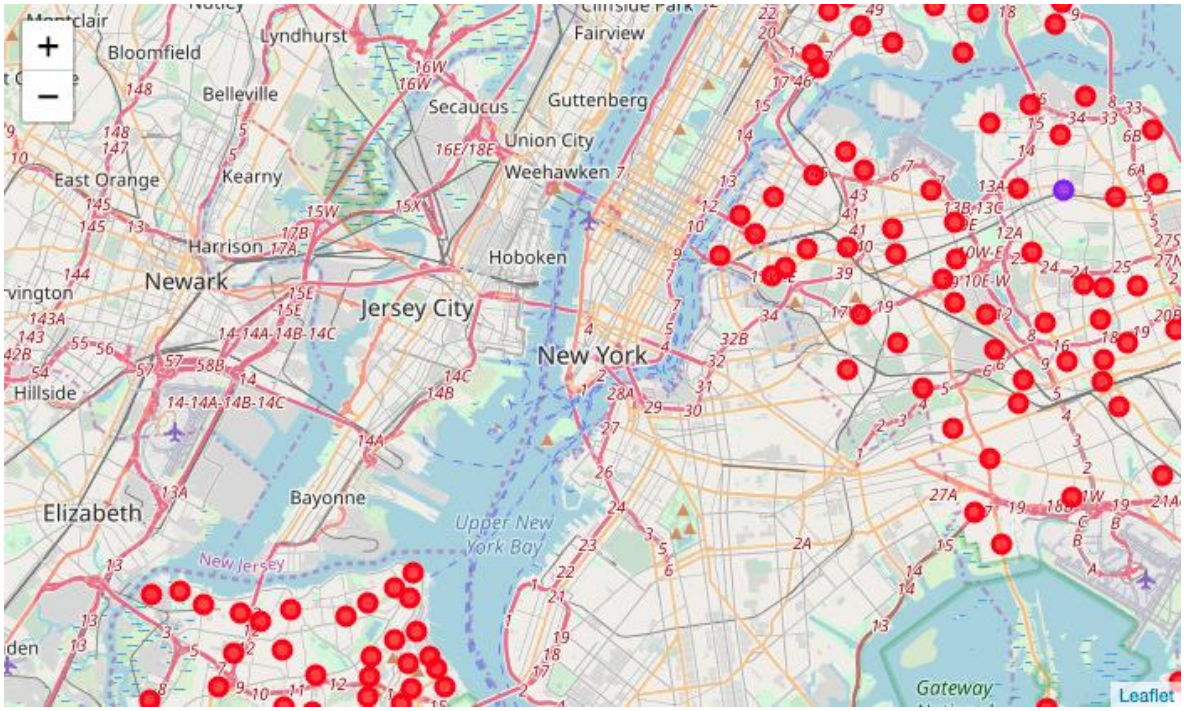


Cluster0 : The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated.
Cluster1 : The Total and Total Sum of cluster1 has highest value. It shows that the markets are saturated. Number of restaurants are very high.

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.

Discussion

1. There is scope to increase Farmers markets in Bronx, Queens and Staten Island.
2. In Manhattan and Brooklyn restaurants of cuisines of many countries are available. It shows that people love eating cuisines of various countries. Or there is very high density of people with different nationalities.

Conclusion

To conclude it should be mentioned that the analysis performed on limited data. Therefore, there may be incorrect clustering results. Brooklyn and Manhattan have high concentration of restaurant business and very
competitive market.

Regarding our objective, based on data and analysis it may be stated that a perfect place to start "Butterscotch Pancakes" restaurant is Brooklyn (probably due to a high concentration of ex-soviet people living there).