IBM

IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

CAPSTONE PROJECT - THE BATTLE OF NEIGHBORHOODS

LOCATION OPTIMIZATION IN NEW YORK CITY

1. Introduction

OBJETIVE
Optimum location for new restaurant business.

BUSINESS PROBLEM
Choice of first neighborhood to start restaurant business.

- Easy to replicate.
- Low competition
- High demand
- Choice of menu

SUCCESS CRITERIA
The success of the business implementation depends on the location of the business and the ability to replicate it in new areas using the data. Therefore, the objective would be to locate and recommend the most optimal address to open a restaurant based on the casuistry of the area.

1. Introduction

The hospitality sector is very broad and culturally diverse in New York City. Its variety of international cuisines due to the high number of immigrants residing and visiting the city on business frequently, such as:

- Asian restaurants.
- Italian restaurants.
- Middle Eastern restaurants.
- Central and Eastern European restaurants.
- Jewish restaurants.
- Haute cuisine restaurants.
- Fast food restaurants.

Among many others.

2. Data

Data 1. 2014 New York City Neighbourhoods Names from NYU:
https://geo.nyu.edu/catalog/nyu_2451_34572

Data 2. DOHMH Farmers Markets from NYC:
https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2

Data 3. New York City from Wikipedia:
https://en.wikipedia.org/wiki/New_York_City

Data 4. Economy of New York City from Wikipedia:
https://en.wikipedia.org/wiki/Economy_of_New_York_City

Data 5. New York City Portal from Wikipedia:
https://en.wikipedia.org/wiki/Portal:New_York_City

Data 6. Cuisine of New York City from Wikipedia:
https://en.wikipedia.org/wiki/Cuisine_of_New_York_City

Data 7. List of Michelin starred restaurants in New York City from Wikipedia:
https://en.wikipedia.org/wiki/List_of_Michelin_starred_restaurants_in_New_York_City

Data 8. Foursquare API.

3. Methodology

BUSINESS UNDERSTANDING
Optimum location for new restaurant business.

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

3. Methodology

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.
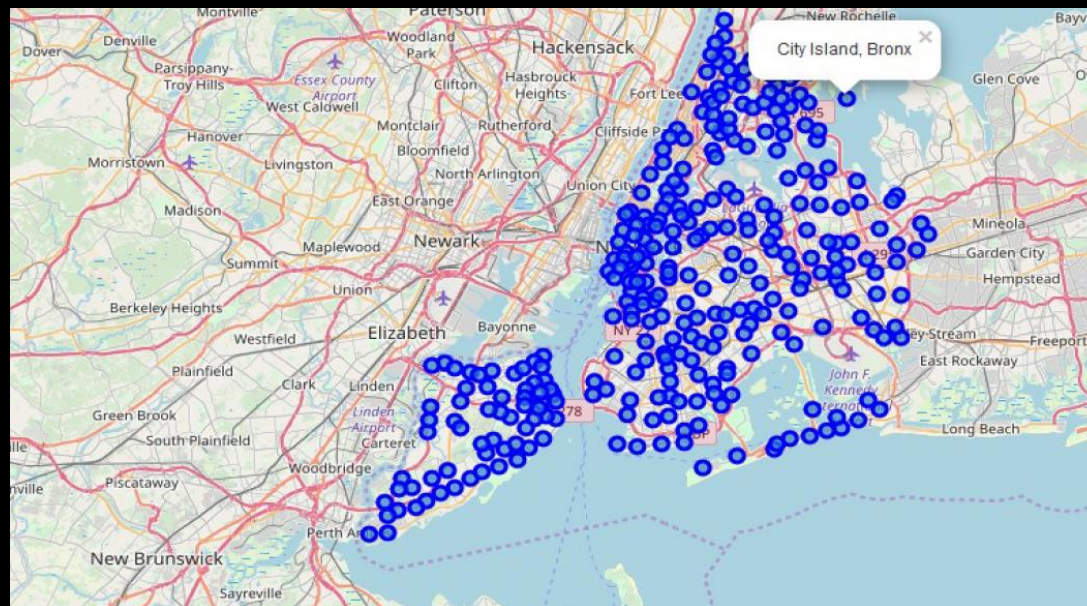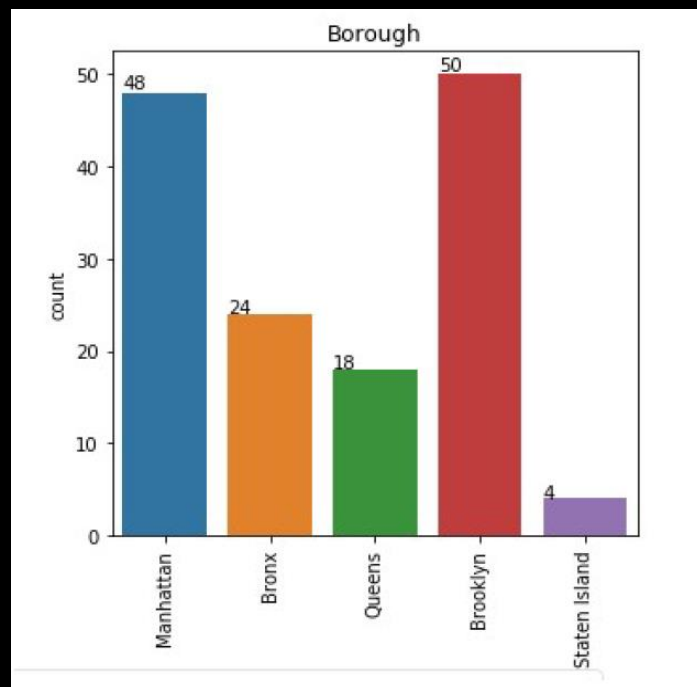
3. Methodology

EXPLORATORY DATA ANALYSIS

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

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.
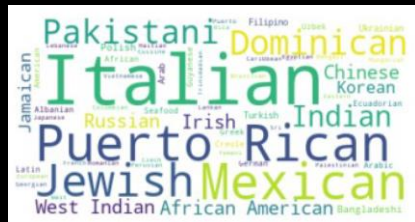
3. Methodology

EXPLORATORY DATA ANALYSIS

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

**NEW YORK CITY CUISINE**   **BROOKLYN CUISINE**   **MANHATTAN CUISINE**   **QUEENS CUISINE**   **THE BRONX CUISINE**

3. Methodology

EXPLORATORY DATA ANALYSIS

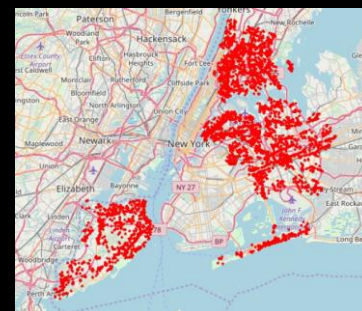**Data 4: New York city geographical coordinates data**

It has been 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 neighbourhoods in New York City.
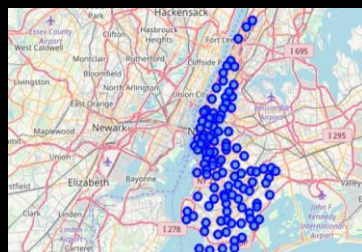
BROOKLYN AND MANHATTAN

BRONX, QUEENS AND STATEN ISLAND

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.

4. Results

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.

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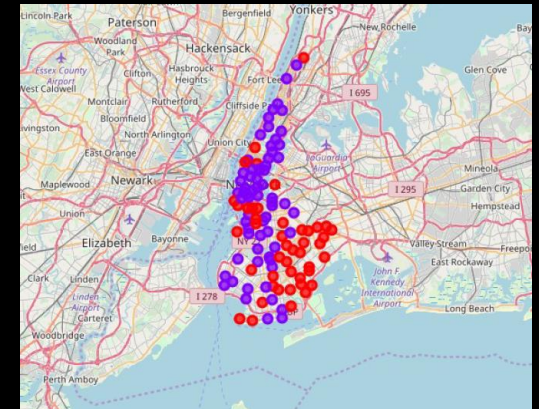
5. Discussion

1. There is scope to increase Farmers markets in Bronx, Queens and Staten Island.

2. There is scope to explore cuisines of various countries in Bronx, Queens and Staten Island.

3. In Manhattan and Brooklyn restaurants of cuisines of many countries are available. So if risk can be taken with great menu on board. It also shows people love eating cuisines of various countries.

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.
As per the neighbourhood or restaurant type mentioned like Indian Restaurant analysis can be checked. A venue with lowest risk and competition can be identified.

THANK YOU & BON APPÉTIT