The Battle of the Neighbourhood – Report

1. **Introduction and business problem**

The city of New York being the most populated city in United States has a lot of business opportunities and also a business friendly environment.

The market here is very competitive, the multicultural environment many various cuisine restaurants being opened here

**Problem description:**

Opening a restaurant is challenging because of the competitive environment, the city of NY being famous for its excellent cuisine.

Its food culture includes an array of international cuisines influenced by the city’s immigrant history.

The following immigrant cultures are present here:

-Italian immigrants

-Jewish immigrants

-Central and Eastern Europe immigrants

-Chinese and Asian restaurants

-Middle Eastern foods

It is obvious that surviving in such a competitive market it is important to have a strategic plan.

In order to decide the location of the Italian restaurant many factors need to be taken into account:

-structure of the population

-city demographics

-the competitors

-segmentation of the borough

-untapped/saturated markets

Selecting the best neighbourhood to live, build a restaurant in New York

* Understanding the similarities and differences between the neighbourhoods using Unsupervised K-means clustering algorithm; K-means clustering is a type of unsupervised machine learning, the goal of the algorithm is to find unlabelled groups in the data; data points are clustered based on feature similarity

Choice of location is very important and is a prerequisite for the future success.

1. **Data**

Data regarding New York city was analysed.

The following datasets were used:

**Data 1: Neighborhoods from New York city**

* Neighborhood has a total of 5 boroughs and 306 neighbourhoods
* This dataset is already available on the web, the link is:

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

**Data 2: Farmers market**

* The data is available at the following link: <https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2>

**Data 3: Demographics and cuisine**

Data was taken using web scraping from Wikipedia as from the below links:

<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: Venues**

* Data was taken from foursquare.com
* 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

neighborhood. We will use the Foursquare API to explore neighborhoods in New

York City



The following techniques were used:

* Web scraping and data wrangling
* Top trending places extraction and clustering
* Decision making based on clustered neighbourhoods, population distribution, competition (other cuisine types opened in the same area),access to Farmers markets which are suppliers for the business

New York city neighborhood has a total of 5 boroughs and 306 neighborhoods

• PART 1 - Clustering of Manhattan and Brooklyn

• PART 2 - Clustering of Bronx, Queens and Staten Island.

• Only restaurant data is filtered from foursquare.com venues data and was further utilized for this project.

This is performed through the following Exploratory data analysis

1. **Methodology**

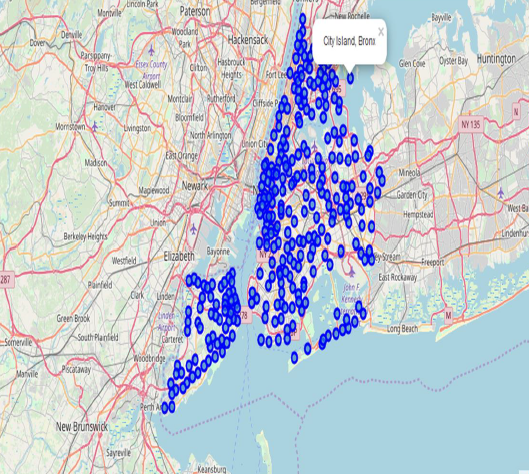
The main purpose was to find an optimum location for opening a new Italian business restaurant.

The following Python packages were used:

* Pandas-library for data analysis
* Numpy-library for handle data in a vectorised manner
* JSON-library to handle JSON files
* Geopy-to retrieve location data
* Requests-library to handle HTML requests
* Matplotlib-Python plotting module
* Sklearn-Python machine learning library
* Folium-map rendering library

Exploratory data analysis:

A.New York city Geographical Coordinates Data

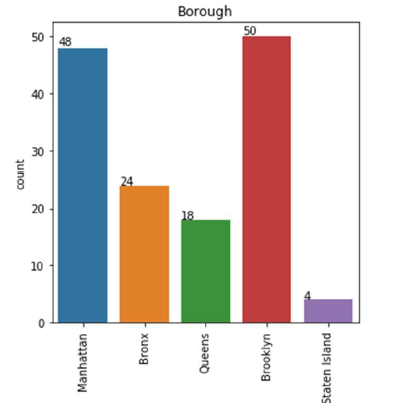


The following steps were performed:

* Load the data and explore-newyosk\_data.json
* Transform the data of nested python dictionaries into a panda dataframe
* Dataframe contains the geographical coordinates of NYC neighborhoods
* Data will use to get Venues data from Foursquare
* Geopy and folium libraries were used to create a map of New York

B. The farmers market

* There are totally 144 Farmers Markets in New York city
* Highest number are in Manhattan and Brooklyn
* Lowest are in Queens & State Island



C. NYC population, demographics and the data about cuisine

* New York is the most populated city in US, with an estimated record high of 8.7 mio residents as of 2017, incorporating more immigration into the city than outmigration🡺 cuisine is different and various

The highest density of the population is in Manhattan (New York County) which has the geographically smallest area but the population with highest density.

Brooklyn is the city’s most populated borough

Queens, on Long Island north and east of Brooklyn, is geographically the largest borough

New York City is the most populous city in the United States , with an

estimated record high of 8,7 mio residents as of 2017,incorporating

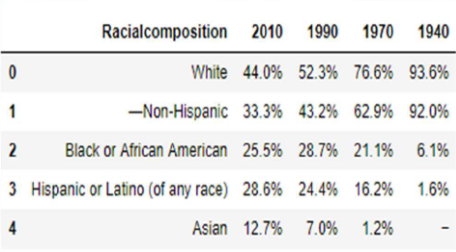
more immigration into the city than outmigration since the 2010 United States

Census.

The racial composition is as given below. This is the reason New York city has

restaurants serving cuisine from many countries such as India, Africa, Japan

etc. This also increases the scope for restaurants business in New York City.



1. **Results**

From all the venues restaurant data was filtered for the following areas:

-Brooklyn and Manhattan (Cluster0)

-Bronx, Queens and Staten Island (Cluster1)

Cluster0:Total and Total sum of cluster0 has smallest value🡺markets is not saturated

Cluster1:Total and total sum of cluster1 has highest value🡺markets are saturated, there’s very little room for opening a restaurant, which means there are no untapped neighbourhoods in Brooklyn and Manhattan.

1. **Discussion**

After the analysis it revealed that Manhattan and Brooklyn restaurants have cuisines from many countries so in order to have success a great menu should be considered to differentiate from the competition and take advantage position.

That’s also means that there is demand for eating cuisines of various countries.

Opening a restaurant in Bronx, Queens and State Island will benefit lower competition but in the same time the Farmers market are not so well represented so the suppliers of fresh vegetables are not so numbered with direct impact on the restaurant menu structure.

1. **Conclusion**

This analysis takes into account a limited amount of data.

It can be improved by including more data or any other variables that might be relevant for the analysis depending on their availability (e.g. income of the people from a specific neighbourhood, consumer behaviour)

The most suitable area for opening a new Italian cuisine restaurant is Bronx, Queens and Staten Island because Brooklyn and Manhattan have a high concentration of restaurant business from all cuisines.