Capstone Project – The Battle of Neighborhoods

Part 1 – Problem Background & Description

I-   Problem Background

NYC is one of the most popular, cosmopolitan city in the United States. It is multicultural. There are many business opportunities and it is a business friendly environment especially for restaurants that are being developed everywhere as it is a global hub of business for the first world power. We are dealing with ABC Company, who wants to open a new restaurant in NYC but doesn't know where.

Indeed, the restauration market is much competitive in this city. As it is a highly developed city the cost of doing business is also one of the highest, with barriers to entry. So we have to think carefully and analyse things deeply before opening any new business. By conducting such an analysis (due diligence), it will help the investors or owners reducing the risks and make the best decisions possible.

II-   Problem Description

The City of New York is famous for its variety of restaurants. Its food culture includes an array of international cuisines influenced by the city's immigrant history.

Central, Eastern European, Jewish - bagels, cheesecake, hot dogs, knishes, delicatessens...
Italian - Pizza & Italian cuisine
Irish - pastrami and corned beef
Asian restaurants or sandwich joints, diners, and coffeehouses, sushis…
Truck food vendors (very trendy with gentrification)
Middle Eastern foods such as falafel and kebabs…

As a matter of fact, to survive in such competitive market, it is of paramount importance to plan carefully. Various factors need to be studied in order to decide on the location

Population, demographics, Farmers Markets, Wholesale markets nearby so that the ingredients can be purchased fresh to maintain quality and cost? venues like Gyms, Entertainment zones, Parks etc nearby where floating population is high etc Who are the competitors in that location? Cuisine served / Menu of the competitors Segmentation Untapped markets Saturated markets ...

We are working for ABC company that want to recommend to the management which neighborhood of NYC will be best to start a restaurant. The Management also expects to understand the rationale of the recommendations made.

The success criteria of the project will be a good recommendation of the Neighborhood choice to ABC Company based on lack of restaurants in that location and nearest suppliers of ingredients.

III-   Data

We will study the city of NYC. To do that, we will use the datasets below:

-Data 1 : NYC has a total of 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood.
Link to the dataset is : https://geo.nyu.edu/catalog/nyu_2451_34572

-Data 2 : Second data which will be used is the Farmers Markets and Food Boxes dataset.
https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2

-Data 3 : For the analysis we will get data from wikipedia as given below :

NYC Population, Demographics, types of cuisine in NYC
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 https://en.wikipedia.org/wiki/List_of_Michelin_starred_restaurants_in_New_York_City

-Data 4 : NYC geographical coordinates data will be utilized as input from Foursquare API, that will be leveraged to provision venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in NYC

IV-   Target audience

The objective is to locate for the management which neighborhood of NYC will be the best choice to start a restaurant. The Management also expects to understand the rationale of the recommendations made. As a matter of fact, This would interest anyone who wants to start a new restaurant in NYC.

Part 2: Analysis

I-   NYC Geographical Data

-load the data and explore data from newyork_data.json file.
- We transform the data of nested python dictionaries into a pandas DF
-This dataframe contains the geographical coordinates of NYC neighborhoods

-It will be used to get venues data from Fouresquare

-We also use geopy and folium libraries to create a map of NYC with neighborhoods superimposed
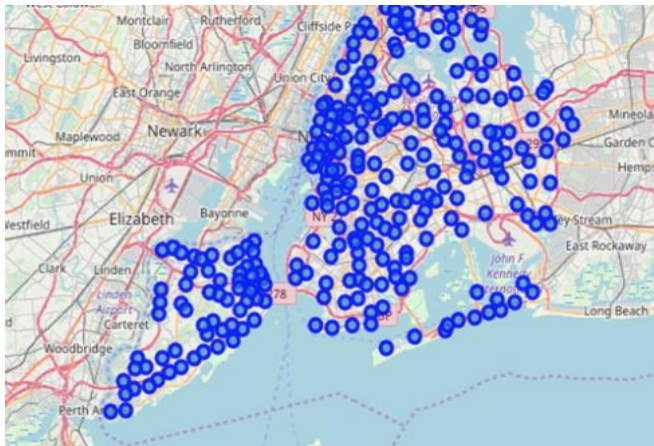
NYC neighborhood Visualization

II-   Farmers' markets

In this part we will use the Food Boxes Dataset as well as the DOHMH Farmers Market datasets. The highest numbers of farmers' markets are in Manhattan and Brooklyn, whereas the lowest are in Queens, Bronx and Staten Island. It can be seen on the graphs below, and we use folium libraries to create a map which highlights the presence of the farmer's markets
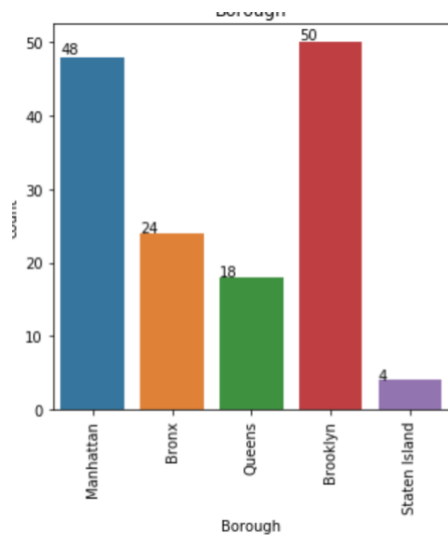
NYC Farmers' markets Visualization

III-   Wikipedia Data scrapped

To analyize NYC 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
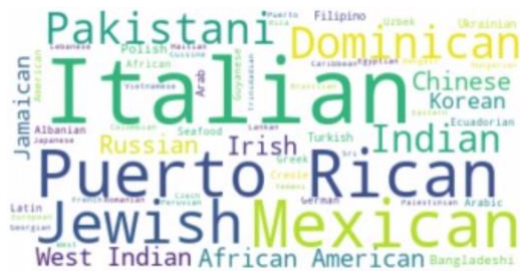
a- Population

→Manhattan is the smallest and most densely populated borough with a population density of 72,033 people per square mile (27,812/km²) in 2015
→Brooklyn, is the city's most populated borough
→Queens, is the largest borough

	Borough	County	Estimate_2017	square_miles	square_km	persons_sq_mi	persons_sq_km
0	Manhattan	New York	1,664,727	22.83	59.13	72,033	27,826
1	The Bronx	Bronx	1,471,160	42.10	109.04	34,653	13,231
2	Brooklyn	Kings	2,648,771	70.82	183.42	37,137	14,649
3	Queens	Queens	2,358,582	108.53	281.09	21,460	8,354
4	Staten Island	Richmond	479,458	58.37	151.18	8,112	3,132
5	City of New York		8,622,698	302.64	783.83	28,188	10,947	NaN
6	State of New York		19,849,399	47,214	122,284	416.4	159	NaN
7	Sources: [2] and see individual borough articles	NaN	NaN	NaN	NaN	NaN	NaN

b- NYC Cuisine

The city as a whole tends to like these cuisines:

BROOKLYN CUISINE -Most Preferred Food is Italian, Puerto Rican & Mexican

MANHATTAN CUISINE - Most Preferred Food is Italian, American & Puerto Rican

BRONX CUISINE - Most Preferred Food is – Italian,Puerto Rican & Albanian

QUEENS CUISINE - Most Preferred Food is Indian, Irish & Pakistani

IV-   Foursquare Data

NYC geographical coordinates data is used as an input for the Foursquare API, which has been leveraged to provision venues information for each neighborhood. We used the Foursquare API data to explore neighborhoods in New York City.

Brooklyn & Manhattan venue data

|	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Sam's Pizza	40.879435	-73.905859	Pizza Place
4	Marble Hill	40.876551	-73.91066	Loeser's Delicatessen	40.879242	-73.905471	Sandwich Place

Brooklyn & Manhattan venue map

Bronx & Queens venue data

|	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Ripe Kitchen & Bar	40.898152	-73.838875	Caribbean Restaurant
2	Wakefield	40.894705	-73.847201	Jackie's West Indian Bakery	40.889283	-73.843310	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Ali's Roti Shop	40.894036	-73.856935	Caribbean Restaurant
4	Wakefield	40.894705	-73.847201	Rite Aid	40.896521	-73.844680	Pharmacy

Bronx & Queens venue map
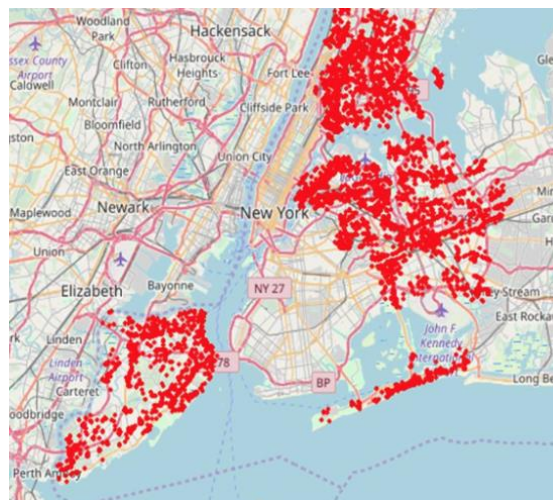
From this venues data we filtered and used the restaurant data for Brooklyn & Manhattan clustering and Bronx, Queens and Staten Island clustering. As a matter of fact, we focused only on restaurants business as we are opening a restaurant

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.

Brooklyn & Manhattan
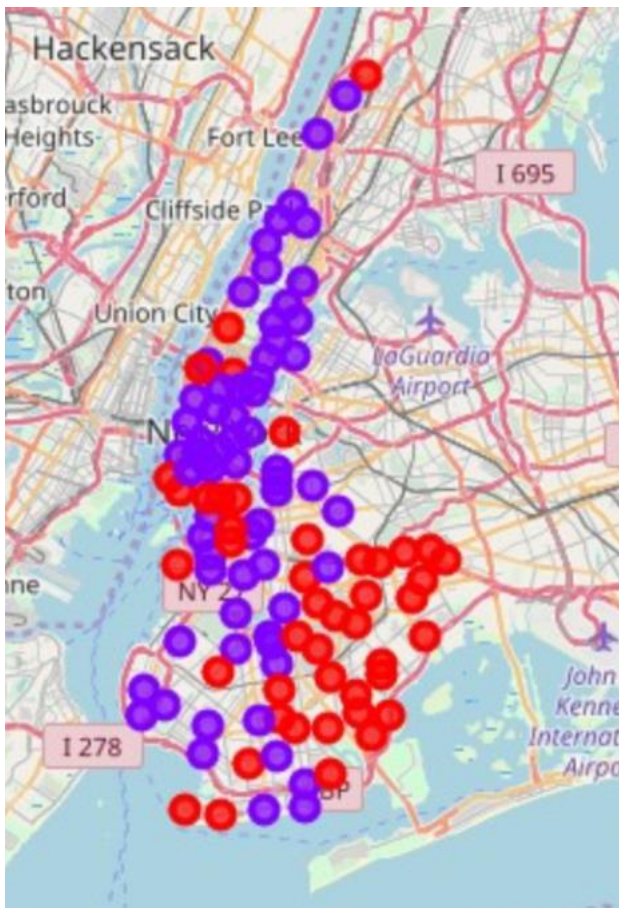
Cluster0 : The Total Sum of cluster0 has smallest value. It shows that the market is not saturated.

Cluster1 : The Total Sum of cluster1 has highest value. It shows that the markets are saturated. Number of restaurants are very high in this area

As we could have expect, there are no untapped neighborhoods in Brooklyn and Manhattan.

Bronx & Queens & Staten Island
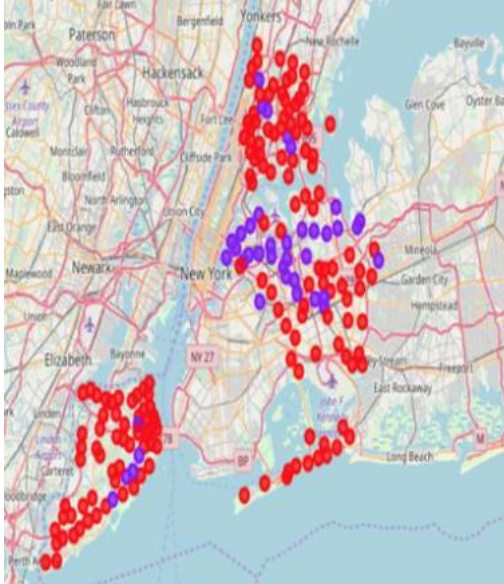
There are untapped markets in this area!

Part 3 – Conclusion Discussion

Discussion

1. There is an opportunity to increase Farmers markets in Bronx, Queens and Staten Island for ABC Company.
2. There is an opportunity to open restaurants with cuisines from various countries in Bronx, Queens and Staten Island, that are not present yet in these neighborhoods.
3. In Manhattan and Brooklyn restaurants of cuisines of many countries are available. So risks can be taken with tryning new things in these neighborhoods. It also shows people love eating cuisines of various countries.

Conclusion

- ABC Company can take many risks regarding cuisine: in Manhattan and Brooklyn, menus are widely multicultural (mostly Italian and Mexican though). Maybe avoiding opening another italian restaurant is a good idea, as this market seems saturated. An Indian restaurant might be a good idea, as many New Yorkers seem to appreciate this cuisine. Nonetheless, there are not that many Indian restaurant in the city, as compared to other types of cuisine.

- Manhattan and Brooklyn are saturated, so ABC should consider one of the three other Boroughs.
- In Queens, Bronx and Staten Island, there are not many Farmers markets. That can also be an opportunity for ABC Company, who could open a few to expand their business.
- Consequently, I would choose Staten Island. Then, Bloomfield as it is not saturated (yet) and ABC could have the first mover advantage in this neighborhood.

D