The Battle of Neighborhood- Week 1

Introduction and Business Problem

Introduction:

New York City (NYC), often called the City of New York or simply New York (NY), is the most populous city in the United States. With an estimated 2018 population of 8,398,748 distributed over about 302.6 square miles (784 km2), New York is also the most densely populated major city in the United States. Located at the southern tip of the U.S. state of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass. With almost 20 million people in its metropolitan statistical area and approximately 23 million in its combined statistical area, it is one of the world's most populous megacities. New York City has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. Home to the headquarters of the United Nations, New York is an important center for international diplomacy.

This also means that the market is highly competitive. As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analysed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

Problem Description:

As increasing numbers of consumers want to dine out or take prepared food home, the number of food-service operations has skyrocketed from 155,000 about 40 years ago to nearly 960,000 today. But there's still room in the market for your food-service business.

A restaurant is a business which prepares and serves food and drink to customers in return for money, either paid before the meal, after the meal, or with an open account. The City of New York is famous for its excelllent cuisine. It's food culture includes an array of international cuisines influenced by the city's immigrant history.

Central and Eastern European immigrants, especially Jewish immigrants - bagels, cheesecake, hot dogs, knishes, and delicatessens Italian immigrants - New York-style pizza and Italian cuisine Jewish immigrants and Irish immigrants - pastrami and corned beef Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city mobile food vendors - Some 4,000 licensed by the city Middle Eastern foods such as falafel and kebabs examples of modern New York street food It is famous for not just Pizzerias, Cafe's but also for fine dining Michelin starred restaurants.The city is home to "nearly one thousand of the finest and most diverse haute cuisine restaurants in the world", according to Michelin. So it is evident that to survive in such competitive market it is very important to startegically plan. Various factors need to be studied inorder to decide on the Location such as :

1. New York Population
2. New York City Demographics

3. Are there any Farmers Markets, Wholesale markets etc nearby so that the ingredients can be purchased fresh to maintain quality and cost?
4. Are there any venues like Gyms, Entertainmnet zones, Parks etc nearby where floating population is high etc
5. Who are the competitors in that location?
6. Cuisine served / Menu of the competitors
7. Segmentation of the Borough
8. Untapped markets
9. Saturated markets etc

Eventhough well funded ABC Company Ltd. need to choose the correct location to start its first venture.If this is successful they can replicate the same in other locations. First move is very important, thereby choice of location is very important.

Target Audience

To recommend the correct location, ABC Company Ltd has appointed me to lead of the Data Science team. The objective is to locate and recommend to the management which neighborhood of Newyork city will be best choice to start a restaurant. The Management also expects to understand the rationale of the recommendations made.

This would interest anyone who wants to start a new restaurant in Newyork city.

Success Criteria

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

Data

City choosed for analysis- New York

Data Set for New York City:- Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segement 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. This dataset exists for free on the web. Link to the dataset is : https://geo.nyu.edu/catalog/nyu_2451_34572

Second data which will be used is the DOHMH Farmers Markets and Food Boxes dataset. In this we will be using the data of Farmers Markets. https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2 Website- https://www.grownyc.org/greenmarketco/foodbox GrowNYC's 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.

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

New York Population New York City Demographics Cuisine of New York city

1. https://en.wikipedia.org/wiki/New_York_City
2. https://en.wikipedia.org/wiki/Economy_of_New_York_City
3. https://en.wikipedia.org/wiki/Portal:New_York_City
4. https://en.wikipedia.org/wiki/Cuisine_of_New_York_City

5. https://en.wikipedia.org/wiki/List_of_Michelin_starred_restaurants_in_New_York_City

Fourth Data: Newyork 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 below is image of the Foursquare API data.

3.Methodology :

Business Understanding :

Our main goal is to get optimum location for new restaurant business in New York City for ABC Company.

Analytic Approach :

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

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.**

Data 3 : To analyize New York city 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

1.New York Population : Insights from the data :

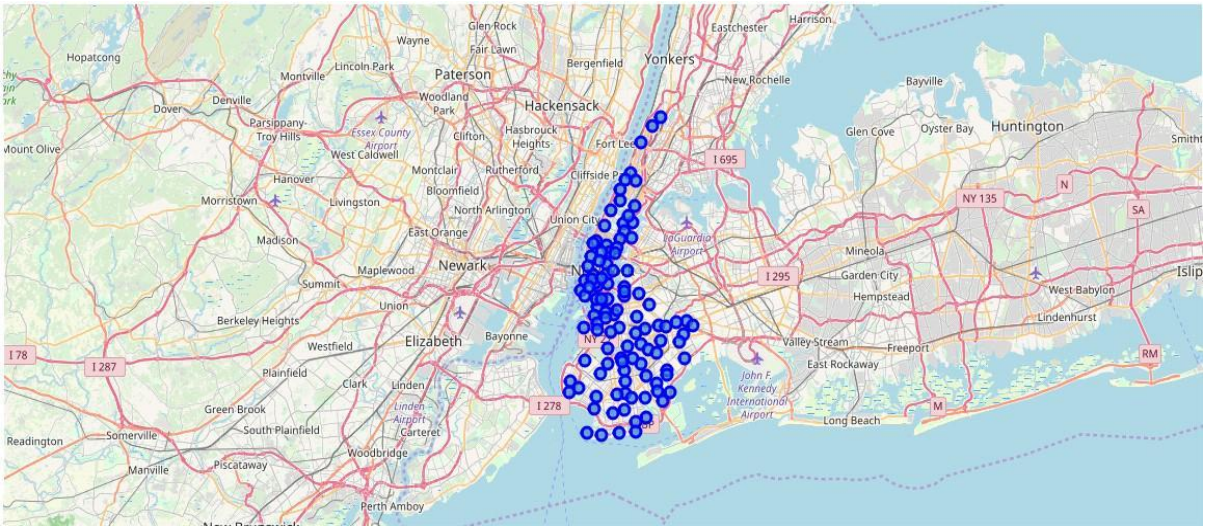
Manhattan (New York County) is the geographically smallest and most densely populated borough.
Manhattan's (New York County's) population density of 72,033 people per square mile (27,812/km²) in 2015 makes it the highest of any county in the United States and higher than the density of any individual 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.

	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
6		State of New York	19,849,399	47,214	122,284	416.4	159

Brooklyn and Manhattan :

Brooklyn and Manhattan Visualization :



Using the geographical coordinates of each neighbourhood foursquare API calls are made to get top 500 venues in a radius of 2000 meters. The venues data is as given below :

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.

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.

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.

.