The Battle of the Neighborhoods – Report

1. Introduction & Business Problem:

Problem Background:

The City of New York, usually called either New York City (NYC) 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 a land area of 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 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 and one of the world's most populous megacities, with an estimated 19,979,477 people in its 2018 Metropolitan Statistical Area and 22,679,948 residents in its Combined Statistical Area. A global power city, New York City has been described as the cultural, financial, and media capital of the world, and exerts a significant impact upon commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. The city's fast pace has inspired the term New York minute. Home to the headquarters of the United Nations, New York is an important center for international diplomacy.

Situated on one of the world's largest natural harbors, New York City consists of five boroughs, each of which is a separate county of the State of New York. The five boroughs – Brooklyn, Queens, Manhattan, The Bronx, and Staten Island – were consolidated into a single city in 1898. The city and its metropolitan area constitute the premier gateway for legal immigration to the United States. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world. New York City is home to more than 3.2 million residents born outside the United States, the largest foreign-born population of any city in the world. As of 2019, the New York metropolitan area is estimated to produce a gross metropolitan product (GMP) of US$1.9 trillion. If greater New York City were a sovereign state, it would have the 12th highest GDP in the world. New York is home to the highest number of billionaires of any city in the world.

So, the market is truly competitive in New York City and doing business is either a risk or an opportunity.

Problem description:

Restaurant is a public place. Provide Food and Beverage on a commercial basis. This is open to all to take refreshment, Food and beverage. Everybody can take food and Beverage against money. Restaurant offer service of Food and Beverage desires to satisfy the Guest. Actually, Guest take Rest in restaurant and pay Rent for the having refreshment, food and beverage. Restaurant comes from the word of “Rest and Rent”. “Rest & Rent” those words consisting the word restaurant. Where the guest/client/peoples take their Food & beverage. The city of New York is known for its famous restaurant such as French restaurants, Italian restaurants, Asian restaurants, African restaurants, and so on.

To survive in such competitive market it is very important to strategic plan.

Various factors need to be studied in order to decide on the location such as New York Population, New York City Demographics, Cuisine served / Menu of the competitors, and so on.

Here a company ABC Ltd needs to choose the exact location to start their business. If this is successful they can use the same approach to other locations.

Target Audience:

In order to get the right location for their business, ABC Company Ltd has hired a Data Science team. Their goal is to recommend to the management which neighborhood of New York City will be the best one to start a restaurant. The Data Science team should explain in details to the management their decision to choose the location.
The result of this would interest anyone who wants to start a new restaurant in New York 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 and farmers markets.

2. Data:

One city will be analyzed in this project: New York City.
We will be using the below datasets for analyzing New York City.

Data 1: Neighborhood has a total of 5 boroughs and 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 geographical coordinates of each neighborhood.

Here is the link of the dataset: https://geo.nyu.edu/catalog/nyu_2451_34572

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Data 2: We will use the data relative to the Farmers Markets and Food Boxes.

Dataset sources:

- https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2
- https://www.grownyc.org/greenmarketco/foodbox

Data 3 : We will use the dataset related to New York City Population and Cuisines (Restaurants).

Data sources:

- 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: We will use New York City geographical coordinates data and the Foursquare API to explore neighborhoods in New York City. Below is image of the Foursquare API data.

	Borough	County	Estimate_2017	GrossDomesticProduct\n	square_miles	square_km	persons_sq_mi	squarekm	persons/sq.mi	persons/sq.km\n
0	The Bronx	Bronx	1,471,160	28.787\n	19,570	42.10	109.04	NaN	NaN	NaN
1	Brooklyn	Kings	2,648,771	63.303\n	23,900	70.82	183.42	NaN	NaN	NaN
2	Manhattan	New York	1,664,727	629.682\n	378,250	22.83	59.13	NaN	NaN	NaN
3	Queens	Queens	2,358,582	73.842\n	31,310	108.53	281.09	NaN	NaN	NaN
4	Staten Island	Richmond	479,458	11.249\n	23,460	58.37	151.18	NaN	NaN	NaN
5	City of New York	8,622,698	806.863	93,574	302.64	783.83	28,188	NaN	NaN	NaN
6	State of New York	19,849,399	1,547.116	78,354	47,214	122,284	416.4	NaN	NaN	NaN
7	Sources:[14] and see individual borough articles	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

3. Methodology:

Business Understanding:

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

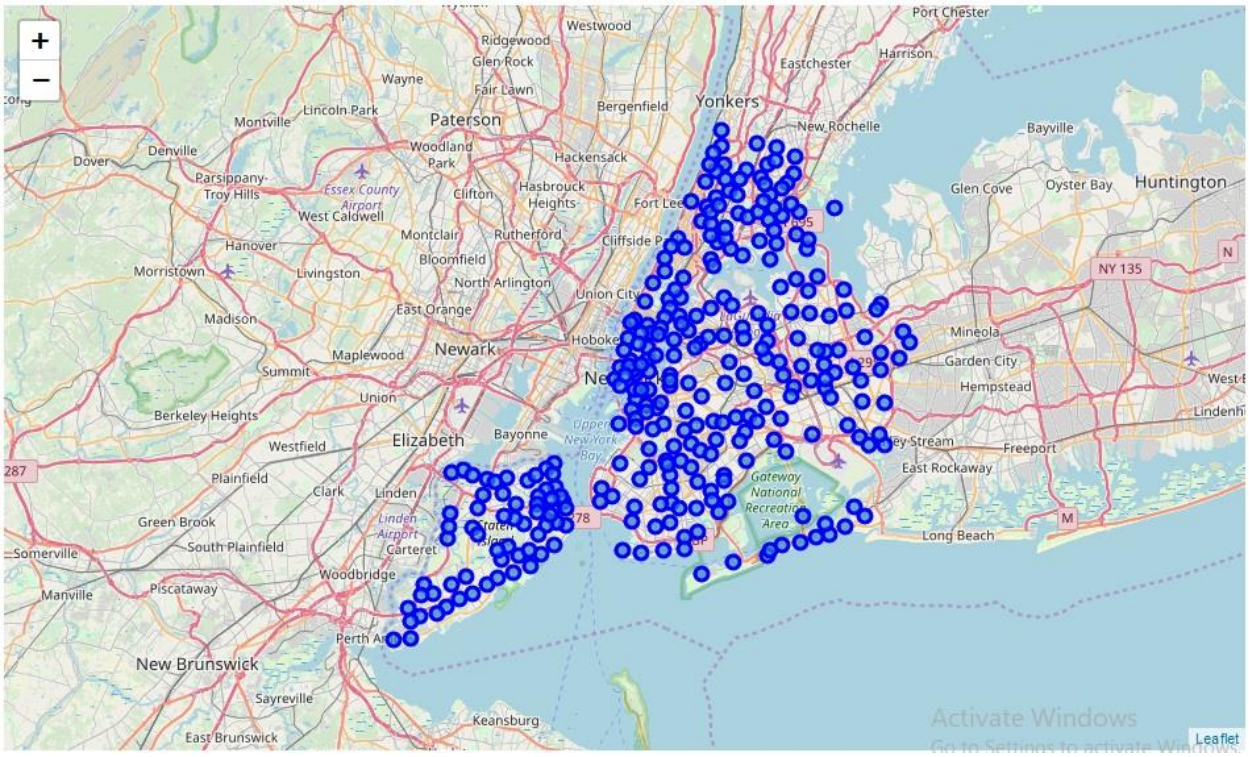
Analytic Approach:

New York City has a total of 5 boroughs and 306 neighborhoods. In this we will be clustering the city of New York and its neighborhoods.

Exploratory Data Analysis:

Data 1- New York City Geographical Coordinates Data.

1. First we load and explore data from newyork_data.json file.
2. Second we transform the data of nested python dictionaries into a pandas data frame.
3. This data frame contains the geographical coordinates of New York City neighborhoods.
4. We will use those data to get Venues data from Foursquare.
5. We used geopy and folium libraries to create a map of New York City with its neighborhoods.

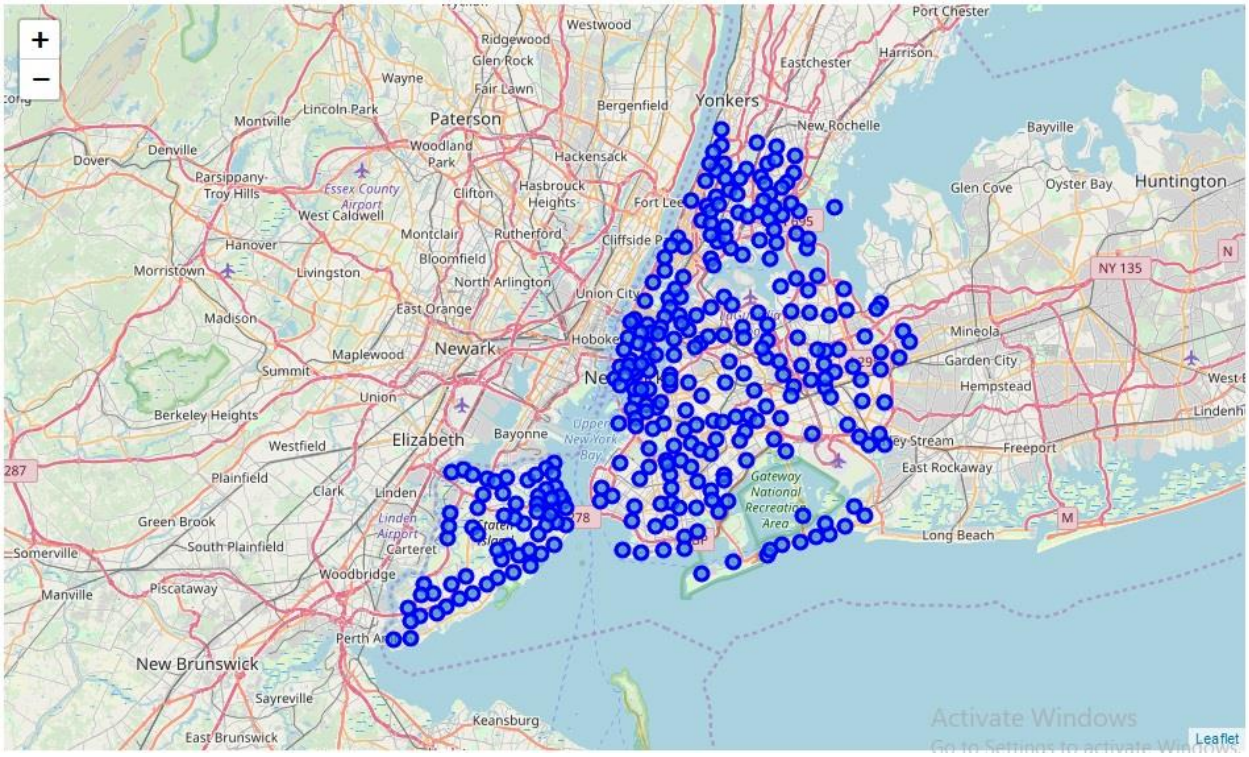


Data 2- Second data which is used is the Farmers Markets dataset. In this There are a total of 144 Farmers Markets in New York City.

Data 3: Here we analyze New York City Population, Demographics and neighborhoods. To do this, we scrapped the data from Wikipedia pages given in the above data section. We used BeautifulSoup python library which is a Python package for parsing HTML and XML documents (including having malformed markup, i.e. non-closed tags, so named after tag soup).

	Borough	County	Estimate_2017	GrossDomesticProduct\n	square_miles	square_km	persons_sq_mi
0	The Bronx	Bronx	1,471,160	28.787\n	19,570	42.10	109.04
1	Brooklyn	Kings	2,648,771	63.303\n	23,900	70.82	183.42
2	Manhattan	New York	1,664,727	629.682\n	378,250	22.83	59.13
3	Queens	Queens	2,358,582	73.842\n	31,310	108.53	281.09
4	Staten Island	Richmond	479,458	11.249\n	23,460	58.37	151.18
5		City of New York	8,622,698	93,574	806.863	302.64	783.83
6		State of New York	19,849,399	78,354	1,547.116	47,214	122,284

Data 4: We used New York City geographical coordinates data as input for the Foursquare API. The Foursquare API data is used to explore neighborhoods in New York City.



4. RESULTS:

From these data we filtered and used only the data from clustering New York City and its neighborhoods.

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.

5. 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 because the market is very Competitive. In this way the location with the lowest risk and competition can be identified.