The Battle of Neighborhoods

1. Introduction And Business Problem

Problem statement : To help entrepreneur to find an ideal location for opening a new restaurant in New York

The city I will be analalysing for finding the suitable location will be **New York**

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.

City economic overview

New York City is a global hub of business and commerce, as a center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in the United States; while Silicon Alley, metonymous for New York's broad-spectrum high technology sphere, continues to expand.

In this project I will be analysing the neighbourhood of New York based on certain factors to find a location that would be suitable for opening a new restaurant. The different factors I will be considering are:

- Location of other restaurants in the area.
- Nearby colleges and school (so that they can start a section that could attract them)
- vicinity
- Distance from Airports
- Historically significant area in the neighbourhood that can attract toursits
- Menu of the competitors
- Availability of food materials (Any markets nearby)

2. Data description

The data includes :

1. A dataset to segement the neighborhoods and explore them ,the data set consist of 5 boroughs and 306 neighborhoods. https://geo.nyu.edu/catalog/nyu_2451_34572 (https://geo.nyu.edu/catalog/nyu_2451_34572)
2. The Foursquare API will be used to explore neighborhoods in New York City
3. Newyork city geographical coordinates data will be utilized as input for the Foursquare API
4. List of all the cities in United States with population density and coordinates: https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population (https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population)
5. The data about population, population density and per capita income of different boroughis analyzed. https://en.wikipedia.org/wiki/Boroughs_of_New_York_City (https://en.wikipedia.org/wiki/Boroughs_of_New_York_City)

Data about the neighbourhood with corresponding coordinates

	Borough	Neighborhood	Latitude	Longitude
0	Staten Island	St. George	40.644982	-74.079353
1	Staten Island	New Brighton	40.640615	-74.087017
2	Staten Island	Stapleton	40.626928	-74.077902
3	Staten Island	Rosebank	40.615305	-74.069805
4	Staten Island	West Brighton	40.631879	-74.107182

	Borough	Neighborhood	Latitude	Longitude
0	Staten Island	St. George	40.644982	-74.079353
1	Staten Island	New Brighton	40.640615	-74.087017
2	Staten Island	Stapleton	40.626928	-74.077902
3	Staten Island	Rosebank	40.615305	-74.069805
4	Staten Island	West Brighton	40.631879	-74.107182

	Borough	Neighborhood	Latitude	Longitude
0	Staten Island	St. George	40.644982	-74.079353
1	Staten Island	New Brighton	40.640615	-74.087017
2	Staten Island	Stapleton	40.626928	-74.077902
3	Staten Island	Rosebank	40.615305	-74.069805
4	Staten Island	West Brighton	40.631879	-74.107182

	Borough	Neighborhood	Latitude	Longitude
0	Staten Island	St. George	40.644982	-74.079353
1	Staten Island	New Brighton	40.640615	-74.087017
2	Staten Island	Stapleton	40.626928	-74.077902
3	Staten Island	Rosebank	40.615305	-74.069805
4	Staten Island	West Brighton	40.631879	-74.107182

	Borough	Neighborhood	Latitude	Longitude
0	Staten Island	St. George	40.644982	-74.079353
1	Staten Island	New Brighton	40.640615	-74.087017
2	Staten Island	Stapleton	40.626928	-74.077902
3	Staten Island	Rosebank	40.615305	-74.069805
4	Staten Island	West Brighton	40.631879	-74.107182

Data for comparing the five borough

	Borough	Population	per capita	Density
0	The Bronx	1471160	19570	34653
1	Brooklyn	2648771	23900	37137
2	Manhattan	1664727	378250	72033
3	Queens	2358582	31310	21460
4	Staten Island	479458	23460	8112

3. Methodology

Map will be used for evaluation of neighbourhood , a suitable location will be identified for opening a new restaurant based on the various factors above . Comparing opening and closing time of competitors,and closeness from different public transport systems.

My first task was choose the best among the five borough of New York. For that purpose I relied on data about population density ,population and per capita income of each borough. The data about them were accessed using pandas. After comparing the above three I came to the conclusion that Manhattan is the preferable one.

The next part was to retrieve the data about different restaurants in Manhattan with the help of Foursquare API. For getting the data ,I used the search query with restaurant as the keyword. The location of these restaurants were plotted in a map.

The nearby venues obtained where clustered into five clusters using k-means. The clusters obtained where analysed closely Each location where evaluated depending on the parameters

Neighbourhood along with the top 10 most common venues

:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Battery Park City	Park	Coffee Shop	Hotel	Memorial Site	Italian Restaurant	Wine Shop	Gym	Clothing Store	Women's Store	Men's Store
1	Carnegie Hill	Coffee Shop	Pizza Place	Café	Bar	Cosmetics Shop	Yoga Studio	Spa	French Restaurant	Japanese Restaurant	Grocery Store
2	Central Harlem	African Restaurant	Chinese Restaurant	French Restaurant	Bar	Public Art	American Restaurant	Gym / Fitness Center	Cosmetics Shop	Seafood Restaurant	Cycle Studio
3	Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Nightclub	Bakery	Hotel	Seafood Restaurant	Art Gallery	Theater	American Restaurant
4	Chinatown	Chinese Restaurant	Cocktail Bar	Salon / Barbershop	American Restaurant	Vietnamese Restaurant	Dumpling Restaurant	Spa	Bubble Tea Shop	Bakery	Ice Cream Shop

4. The target audience

The entrepreneur who wants to find the location to open a restaurant

5. Discussion

Eventhough I have considered the population , per capita income and population density of five borough's, I didn't considered these parameters in Manhattan alone. So the result can be improved by considering these factors. Also with further evaluation of interest of people we could determine which type of restaurant is suitable. I have done the analysis with limited number of data , so the competition caused by other restaurants and crime rates data can improve the location within the suggested location ie between East Harlem and Upper East Side

6.Result

At the end of the project I will be able to suggest a suitable place for opening of **restaurant**. After evaluating the clusters I have came to the conclusion that the suitable place for starting a new restaurant is between **East Harlem** and **Upper East Side.** This is because they are near two museum which is visited by a number of people and a hospital. Also East Harlem is a pilgrimage for foodies

The suggested location

7.Conclusion

In this project, I have gone through the process of identifying the business problem, specifying the data required, extracting , preparing and cleaning the data , performing the machine learning by utilizing k-means clustering and providing recommendation to the entrepreneur