The Battle of Neighborhoods

Report

Problem Background:

- The New York city is the most populous city in the United States. It provides lot of business oppourtunities and business friendly environment.
- This also means that the market is highly competitive. Thus, any new business venture or expansion needs to be analysed careful. This will help in reduction of risk.

Problem Description

- 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.
- So it is evident that to survive in such competitive market it is very important to startegically plan.

Factors to be studied:

- New YorkPopulation
- New York CityDemographics
- Are there any Farmers Markets, Wholesale markets etc nearby so that the ingredients can be purchased fresh to maintain quality andcost?
- Are there any venues like Gyms, Entertainmnet zones, Parks etc nearby where floating population is highetc
- Who are the competitors in thatlocation?
- Cuisine served / Menu of thecompetitors
- Segmentation of theBorough
- Untappedmarkets
- Saturated markets and so on.

Objective and target audience

- The objective is to locate and recommend to the management which neighborhood of Newyork city will be best choice to start a restaurant.
- 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 Neighborhood based on Lack of such restaurants in that location and nearest suppliers of ingredients.

Data 1

- 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

Data 2

- 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

Data 3

- For the analysis we will get data from wikipedia as given below :
- New YorkPopulation
- New York CityDemographics
- Cuisine of New Yorkcity
- 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

- 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.

Methodology

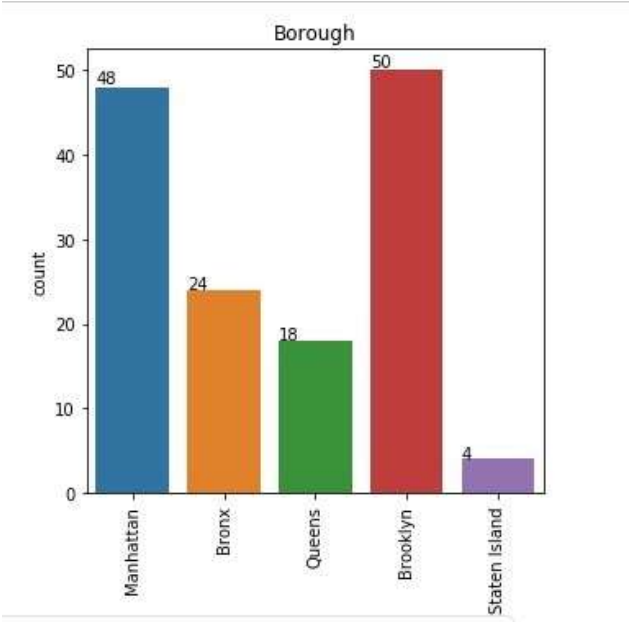
- Business Understanding :
- Our main goal is to get optimum location for new restaurant business in New York City for a company.
- Analytic Approach :
- New York city neighbourhood has a total of 5 boroughs and 306 neighborhoods. In this project first part is clustering of Manhattan and Brooklyn . And second part is clustering of Bronx, Queens and Staten Island. This is done due to the following Exploratory data analysis.

Exploratory data analysis - Data 1

- In this we load the data and explore data from newyork_data.jsonfile.
- Transform the data into a pandas dataframe.
- This dataframe contains the geographical coordinates of New York city neighborhoods.
- This data will used to get Venues data from Fouresquare.
- We used geopy, folium libraries to create a map of New York city with neighborhoods on top.

Data 2

- In this we will be using the data of Farmers Markets data.
- There are totally 144 Farmers Markets in New York city. Highest number are in Manhattan and Brooklyn.
- And lowest in Queens, Bronx and Staten Island.



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). Below are samples from New York Population, Demographics, and cuisine (word cloud)

	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

	Racialcomposition	2010	1990	1970	1940
0	White	44.0%	52.3%	76.6%	93.6%
1	—Non-Hispanic	33.3%	43.2%	62.9%	92.0%
2	Black or African American	25.5%	28.7%	21.1%	6.1%
3	Hispanic or Latino (of any race)	28.6%	24.4%	16.2%	1.6%
4	Asian	12.7%	7.0%	1.2%	–

Data 4

- NewYork city geographical coordinates data has be utilized as input for the Foursquare API. We used the Foursquare API data to explore neighborhoods in New York City.
- The sample of Brooklyn and Manhattan is shown here

Results

- Clustering is done. Two clusters are formed.

DISCUSSION

- There is scope to increase Farmers markets in Bronx, Queens and StatenIsland. There is scope to explore cuisines of various countries in Bronx, Queens and StatenIsland. In Manhattan and Brooklyn restaurants of cuisines of many countries are available. Since people love multi cuisines, risk can be taken.

Conclusion

- As the data is less, the insight from the analysis may be right or wrong. If a considerate amount of data is available, it would provide better results.
- As of now, a venue with less competition can be easily provided for any company seeking it.