Report :

The Battle of the Neighbourhoods

Introduction & Business Problem :

Background:

New York the financial capital of USA is the most populous city in the United States. It has become a global hub of business and commerce by attracting many different players into the market by providing business opportunities and business friendly environment. The market in New York city is highly competitive thus any new business venture or expansion needs to be analysed carefully. The insights derived from the current analysis will give good understanding of business environment which help in strategically targeting the market.

A restaurant or an eatery is a business that prepares and serves food and drinks to customers in return for money. Restaurants vary greatly in appearance and offerings including a wide variety of cuisines and service models ranging from inexpensive fast food restaurants and cafeterias to mid-priced family restaurants to high-priced luxury establishments. The City of New York is famous for its excellent cuisine. Its food culture includes an array of international cuisines influenced by the citys immigrant history. So it is evident that to survive in such competitive market it is very important to plan strategically. Various factors need to be studied in order to decide on the Location such as Population Cuisine served / Menu of the competitors and markets etc.

Objective:

The objective is to Identifying the best neighbourhood of New York city to start a restaurant business. This would interest to the people looking to start a new restaurant in New York city.

Data Sources:

- Latitude and longitude coordinates: https://geo.nyu.edu/catalog/nyu_2451_34572
- Farmers Markets: https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets/8vwk-6iz2
- GrowNYCs Fresh Food Box Program data: https://www.grownyc.org/freshfoodbox
- New York Population: https://en.wikipedia.org/wiki/New_York_City
- New York Economy: https://en.wikipedia.org/wiki/Economy_of_New_York_City
- New York Portal: https://en.wikipedia.org/wiki/Portal:New_York_City

Data Preparation

Data Tools/Libraries used in the analysis are listed below:

- IBM Watson® Studio: to prepare data and build models at scale for this project
- Pandas: To use for data manipulation and analysis
- Phython : Programming Language used for this project
- BeautifulSoup: to create parse trees that is helpful to extract the data
- GitHub: to share project notebooks using Git repository hosting service
- Geopy: to get geological location by address name
- folium: to visualize the distribution pattern
- Foursquare API: to get the most common venues of given Borough of New York City.
- Foursquare API: to get the venues record of given venues of New York City.
- etc.

Data Science Workflow

- **Outline the initial data that is required:**
  - Boroughs data for New York including names location data if available and any other details required.
- **Obtain the Data:**
  - Research and find suitable sources for the district data for New York.
  - Access and explore the data to determine if it can be manipulated for our purposes.
- **Initial Data Wrangling and Cleaning:**

- Clean the data and convert to a useable form as a dataframe.

Data Analysis and Location Data:

- Foursquare location data will be leveraged to explore or compare boroughs around New York.

- Data manipulation and analysis to derive subsets of the initial data.

- **Visualization:**

  - Analysis and plotting visualizations.

  - Data visualization using various mapping libraries.

- **Discussion and Conclusions:**

  - Recommendations and results based on the data analysis.

  - Discussion of any limitations and how the results can be used and any conclusions that can be drawn.

Methodology

Exploratory Data Analysis

New York has a total of 5 boroughs and 306 neighbourhoods. In order to segment the neighbourhoods and explore them we will essentially need a dataset that contains the 5 boroughs and the neighbourhoods that exist in each borough as well as the latitude and longitude coordinates of each neighbourhood. In this project first part is clustering of Manhattan and Brooklyn and second part is clustering of Bronx Queens and Staten Island.

New York city Geographical Coordinates

In this I have loaded the data from newyork_data.json file. and transformed the data of nested python dictionaries into pandas dataframe. This dataframe contains the geographical coordinates of New York city neighbourhoods which was used to get Venues data from Fouresquare. Map of New York city with neighbourhoods superimposed on top was created using geopy and folium libraries.

New York Population

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. Web scrapping of Population data from wikipedia page using BeautifulSoup was done and extracted the population details of New York.

Segmenting and Clustering Neighbourhoods - Brooklyn and Manhattan

New York city geographical coordinates data has be utilized as input for the Foursquare API, that has been leveraged to provision venues information for each neighbourhood. We used the Foursquare API data to explore neighbourhoods in New York City. Later the list of restaurants will be extracted and enlisted.

Results

Neighbourhood has a total of 5 boroughs and 306 neighbourhoods. In order to segment the neighbourhoods and explore them New York city geographical coordinates data has be utilized as input for the Foursquare API that has been leveraged to provision venues information for each neighbourhood identified their latitude and longitude coordinates as shown below

	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

Using the geographical coordinates of each neighbourhood foursquare API could able to depict New York city map with neighbourhoods using geopy and folium libraries

Number of Neighborhood for each Borough in New York City
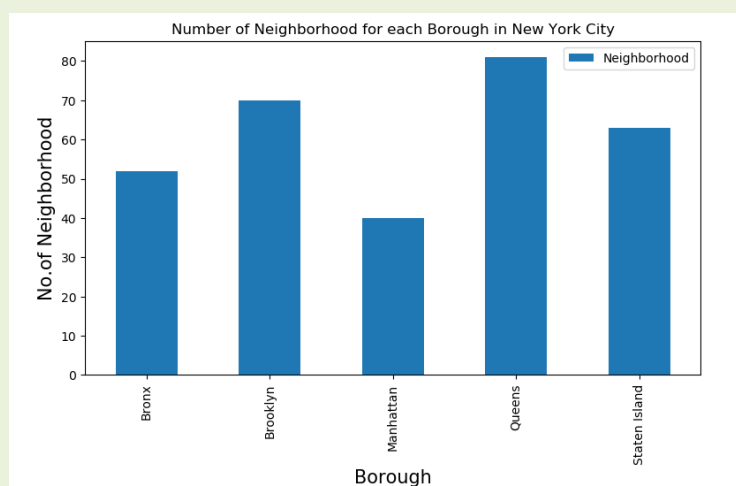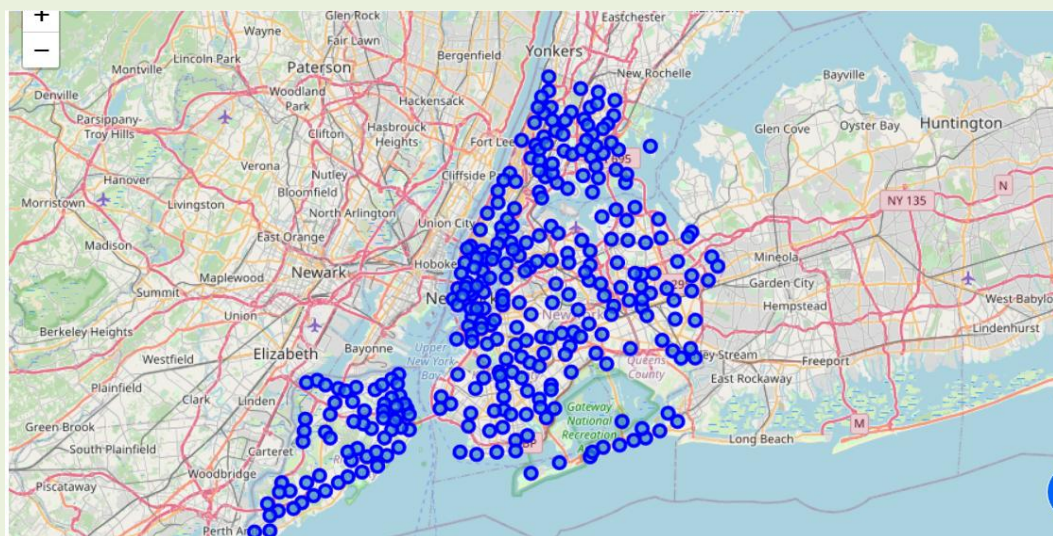
Manhattan is the geographically smallest and most densely populated borough. Manhattans (New York Counties) population density in 2015 makes it the highest of any

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 citys most populous - borough. Queens (Queens County) on Long Island north and east of Brooklyn is geographically the - largest borough

	New York City's five boroughsvte	Jurisdiction	Population	Gross Domestic Product	Land area	Density	Borough
0	The Bronx\n	\n Bronx\n	1,432,132\n	42.695\n	29,200\n	42.10\n	109.04\n
1	Brooklyn\n	\n Kings\n	2,582,830\n	91.559\n	34,600\n	70.82\n	183.42\n
2	Manhattan\n	\n New York\n	1,628,701\n	600.244\n	360,900\n	22.83\n	59.13\n
3	Queens\n	\n Queens\n	2,278,906\n	93.310\n	39,600\n	108.53\n	281.09\n
4	Staten Island\n	\n Richmond\n	476,179\n	14.514\n	30,300\n	58.37\n	151.18\n
5	City of New York	8,398,748	842.343	97,700	302.64	783.83	28,188
6	State of New York	19,745,289	1,701.399	85,700	47,214	122,284	416.4
7	Sources:[14] and see individual borough articl...						

Segmenting and Clustering Neighbourhoods - Brooklyn and Manhattan: New York city geographical coordinates data has be utilized as input for the Foursquare API, that has been leveraged to provision venues information for each neighbourhood. We used the Foursquare API data to explore neighbourhoods in New York City. For example Brooklyn and Manhattan Visualization :
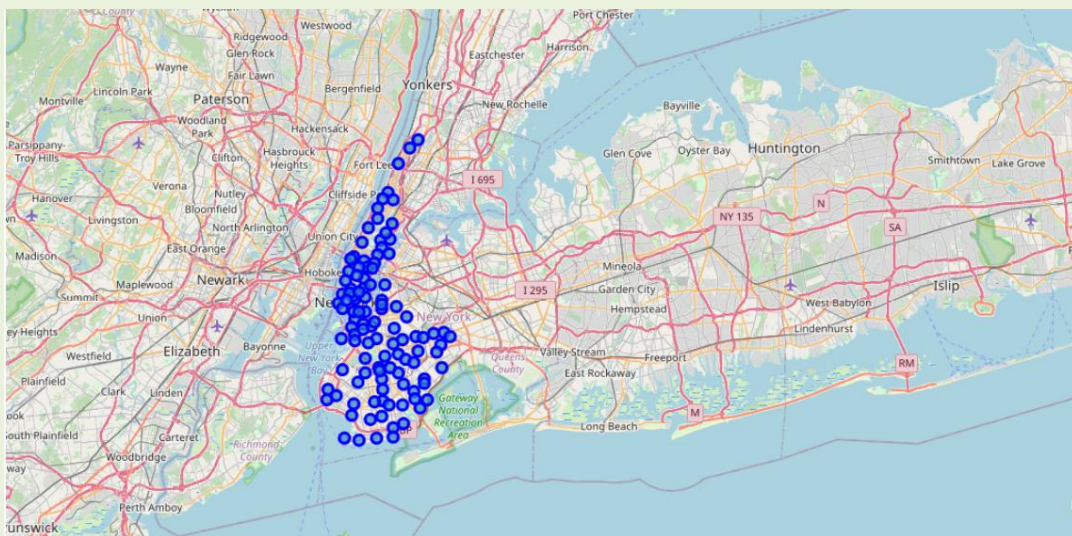
Using the geographical coordinates of each neighbourhood foursquare API calls are made to get top 200 venues in a radius of 1000 meters. The venues data is as given below

Out[76]:

	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
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	Starbucks	40.877531	-73.905582	Coffee Shop

In [77]: BM_venues.shape

Out[77]: (9669, 7)

Later the list of restaurants in were extracted and enlisted below

African Restaurant	Eastern European Restaurant	Korean Restaurant	Shanghai Restaurant
American Restaurant	Empanada Restaurant	Kosher Restaurant	South American Restaurant
Arepa Restaurant	English Restaurant	Latin American Restaurant	Southern / Soul Food Restaurant
Argentinian Restaurant	Ethiopian Restaurant	Lebanese Restaurant	Spanish Restaurant
Asian Restaurant	Falafel Restaurant	Malay Restaurant	Sushi Restaurant
Australian Restaurant	Fast Food Restaurant	Mediterranean Restaurant	Swiss Restaurant
Austrian Restaurant	Filipino Restaurant	Mexican Restaurant	Szechuan Restaurant
Burmese Restaurant	French Restaurant	Middle Eastern Restaurant	Taiwanese Restaurant
Cajun / Creole Restaurant	German Restaurant	Molecular Gastronomy Restaurant	Tapas Restaurant
Cambodian Restaurant	Greek Restaurant	Moroccan Restaurant	Thai Restaurant
Cantonese Restaurant	Halal Restaurant	New American Restaurant	Theme Restaurant
Caribbean Restaurant	Hawaiian Restaurant	North Indian Restaurant	Tibetan Restaurant
Caucasian Restaurant	Hotpot Restaurant	Paella Restaurant	Turkish Restaurant
Chinese Restaurant	Indian Restaurant	Pakistani Restaurant	Udon Restaurant
Colombian Restaurant	Israeli Restaurant	Persian Restaurant	Ukrainian Restaurant
Comfort Food Restaurant	Italian Restaurant	Restaurant	Vegetarian / Vegan Restaurant
Cuban Restaurant	Japanese Curry Restaurant	Russian Restaurant	Venezuelan Restaurant
Czech Restaurant	Japanese Restaurant	Scandinavian Restaurant	Vietnamese Restaurant
Dim Sum Restaurant	Jewish Restaurant	Seafood Restaurant	Ukrainian Restaurant
Dumpling Restaurant	Kebab Restaurant	Shabu-Shabu Restaurant	

Discussion

As a recommendation to those who plan to operate a restaurant location selection is one of the fundamental problem to think over. The analysis of this report gives some idea about the location population and etc. to decide about the restaurant business. It cannot solve the problem of whether a type of restaurant is the most popular type and how many customers will visit every day. And as for location suggestion it offers an opportunity analysis but lacks risk analysis like the cost of the location and competition in that area.

Although in this report there is scope to find the population co-ordinates and the list of several business firms in New York city. this project was done for training purpose so please be cautious while considering the output from each parameter.

Conclusion

This analysis is performed on limited data hence there is further need of analysis to get the detailed idea about the identification of particular location and type of business to be established. Brooklyn and Manhattan has high concentration of restaurant business. Very competitive market. finally based on Segmenting Neighbourhoods - Brooklyn and Manhattan. This project could able to identify coordinates of neighbourhoods and the list of restaurants, which can be utilized cautiously to identify the location to initiate restaurant business. . So this can be explored after validation of current report.

Thank you