The Battle of the Neighbourhoods

Project Report

Applied Data Science Capstone

IBM Data Science Professional Certificate

Coursera

Coursera

Background and Objective

- New York, is the most populous city in the United States.
- Global hub of business and commerce by business opportunities
- It's food culture includes an array of international cuisines
- Highly competitive, thus, any new business venture needs to be analysed carefully
- Very important to plan strategically

- Objective: To facilitate in identifying the neighbourhood of New York city to start a restaurant business

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
- GrowNYC's 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 Tools used in the analysis

- IBM Watson® Studio: to prepare data and build models at scale for this project
- Pandas: To use for data manipulation and analysis
- 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

- **New York city Geographical Coordinates**
  - Map of New York city with neighbourhoods superimposed on top was created using geopy and folium libraries
- **New York Population**
  - Web scrapping of Population data from wikipedia page using BeautifulSoup to extract population numbers
- **Segmenting and Clustering Neighbourhoods - Brooklyn and Manhattan**
  - Foursquare API data to explore neighbourhoods in New York City.
  - Later the list of restaurants extracted and enlisted.

Results: Geographical Coordinates

- Neighbourhood has a total of 5 boroughs and 306 neighbourhoods
- Their latitude and longitude coordinates as shown below

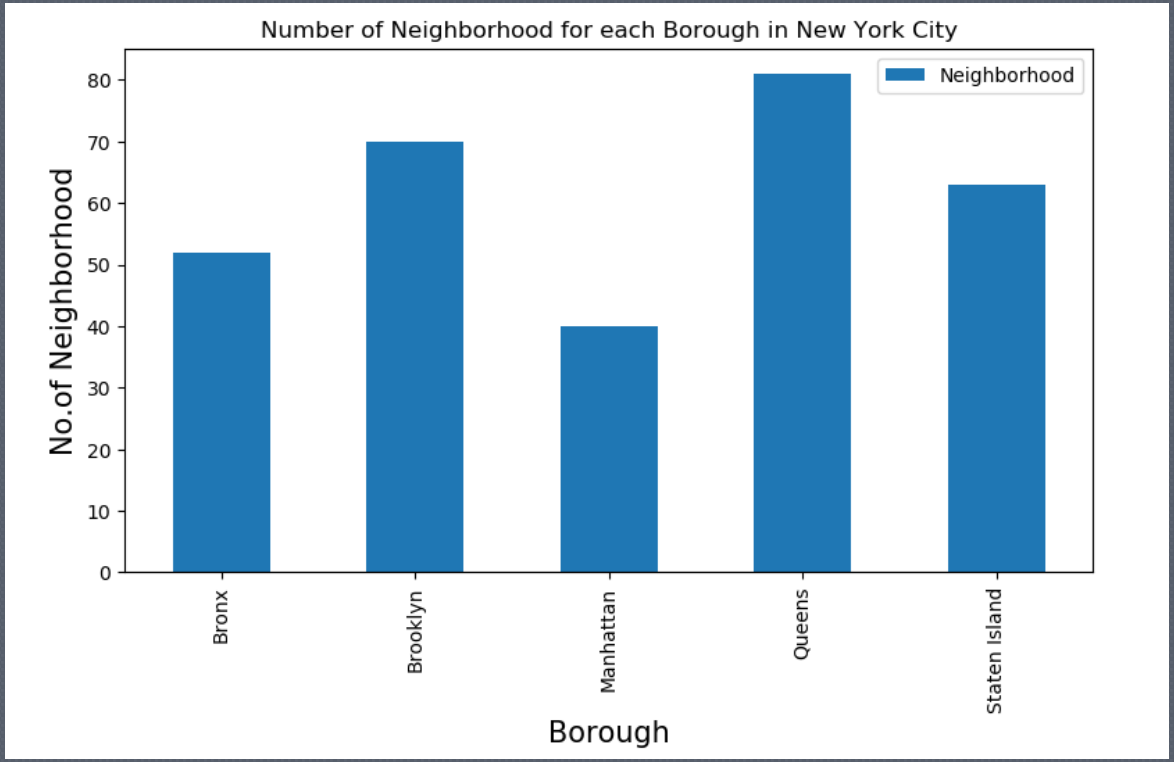
	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Results: Geographical Coordinates

- Neighbourhoods using geopy and folium libraries

Results: Geographical Coordinates

- Neighbourhoods in for each borough in New York City



Results: New York Population

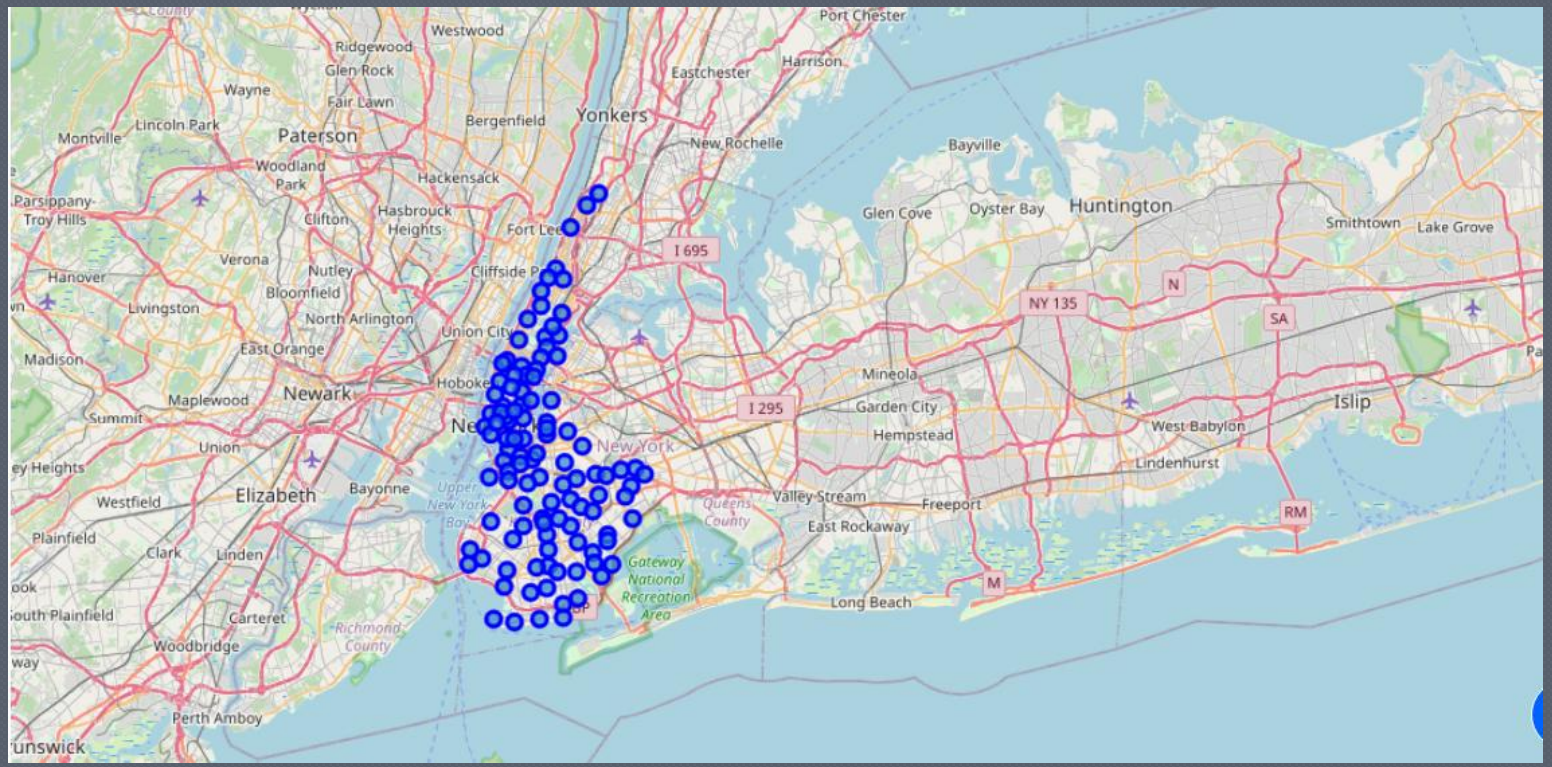
- Web scrapping of Population data from wikipedia page using BeautifulSoup

:

	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...						

Results: Segmenting and Clustering Neighbourhoods - Brooklyn and Manhattan

- Foursquare API, that has been leveraged to provision venues in Brooklyn and Manhattan Visualization



Results: Segmenting and Clustering Neighbourhoods - Brooklyn and Manhattan

- Using the geographical coordinates of each neighbourhood foursquare API calls are made to get top 200 venues in a radius of 1000 meters

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)

- List of restaurants available at: https://github.com/KNRaju1984/github-Example/blob/master/The%20Battle%20of%20Neighborhoods%20Final%20Report.pdf

Discussion

- 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 offers an opportunity analysis but lacks risk analysis like the cost of the location and competition in that area

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. At very competitive market
- Could able to identify coordinates of neighbourhoods and the list of restaurants
- Can be utilized cautiously to identify the location to initiate restaurant business after validation of current report

The Battle of the Neighbourhoods

Project Report

Thank You

Coursera