Capstone Project — The Battle of Neighborhoods | Manhattan, New York.

Introduction:

Manhattan serves as New York city's economic and administrative center, cultural and historical birthplace. Manhattan's most important economic sector lies in its role as the headquarters for the U.S. financial industry, also known as Wall Street. New York City is home to the most corporate headquarters of any city in the United States, the overwhelming majority based in Manhattan.

The purpose of this Capstone Project is to help people who move there in exploring better facilities around their neighborhood. It will help people making smart and efficient decision on selecting great neighborhood out of numbers of other neighborhoods in Manhattan, New York. Ideally it needs lots of research for good housing prices, restaurants, reputed schools for their children. I personally was there a couple of years back for work, I wish I had something like this prior I visited Manhattan.

This Capstone Project aims to create an analysis of features to search a best neighborhood as a comparative analysis between neighborhoods. The features include median housing price and better restaurants according to ratings.

Data Section:

Data Link: https://geo.nyu.edu/catalog/nyu_2451_34572

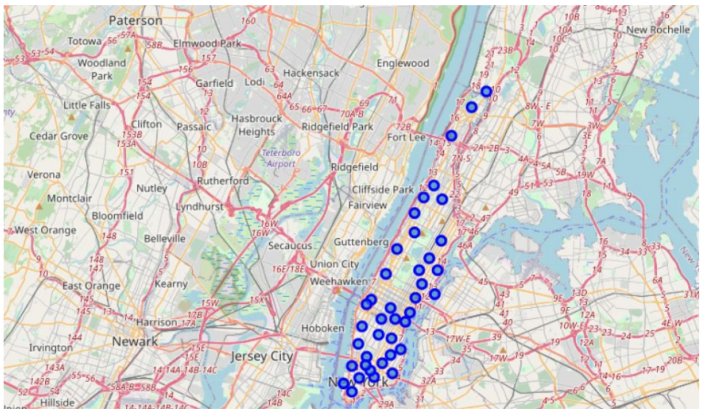
Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a data-set that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and longitude coordinates of each neighborhood.

Foursquare API Data:

We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" location information (Source : https://foursquare.com/). Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus, tips and more. As such, it will be used as the sole data source since all the stated required information can be obtained through the

API.

Map of Manhattan:



Methodology Section:

Clustering Approach:
To compare the similarities, I decided to explore neighborhoods, segment them, and group them into clusters to find informative data. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

Most Common Venues near Neighborhood:

[33]:

	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	Hotel	Coffee Shop	Memorial Site	Gym	Plaza	Boat or Ferry	Food Court	Gourmet Shop	Mexican Restaurant
1	Carnegie Hill	Coffee Shop	Italian Restaurant	Café	Yoga Studio	Gym / Fitness Center	Wine Shop	Gym	Bookstore	Bakery	Grocery Store
2	Central Harlem	African Restaurant	Seafood Restaurant	American Restaurant	Bar	Gym / Fitness Center	French Restaurant	Chinese Restaurant	Café	Bookstore	Boutique
3	Chelsea	Coffee Shop	Art Gallery	American Restaurant	Ice Cream Shop	Italian Restaurant	Café	French Restaurant	Theater	Cocktail Bar	Pizza Place
4	Chinatown	Chinese Restaurant	Bakery	Dessert Shop	Bubble Tea Shop	Ice Cream Shop	Hotpot Restaurant	Optical Shop	Cocktail Bar	Spa	Vietnamese Restaurant

Using K-Means Clustering Approach:

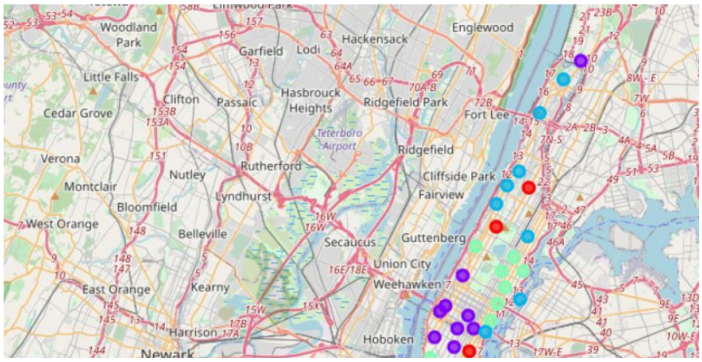
[37]:
```
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

manhattan_merged = manhattan_data

# merge toronto_grouped with toronto_data to add latitude/longitude for each neighborhood
manhattan_merged = manhattan_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

manhattan_merged.head() # check the last columns!
```

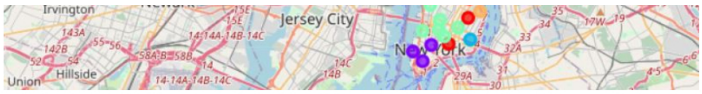
[37]:

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	Manhattan	Marble Hill	40.876551	-73.910660	1	Sandwich Place	Gym	Coffee Shop	Yoga Studio	Pharmacy	Deli / Bodega	Department Store	Diner	Discount Store	Donut Shop
1	Manhattan	Chinatown	40.715618	-73.994279	0	Chinese Restaurant	Bakery	Dessert Shop	American Restaurant	Ice Cream Shop	Spa	Bar	Hotpot Restaurant	Vietnamese Restaurant	Bubble Tea Shop
2	Manhattan	Washington Heights	40.851903	-73.936900	2	Café	Bakery	Mobile Phone Shop	Deli / Bodega	Mexican Restaurant	Latin American Restaurant	Coffee Shop	Tapas Restaurant	Park	Italian Restaurant
3	Manhattan	Inwood	40.867684	-73.921210	2	Lounge	Mexican Restaurant	Restaurant	Café	Deli / Bodega	Bakery	Chinese Restaurant	Caribbean Restaurant	Park	Frozen Yogurt Shop
4	Manhattan	Hamilton Heights	40.823604	-73.949688	2	Pizza Place	Coffee Shop	Deli / Bodega	Café	Mexican Restaurant	Sushi Restaurant	Caribbean Restaurant	School	Bakery	Chinese Restaurant

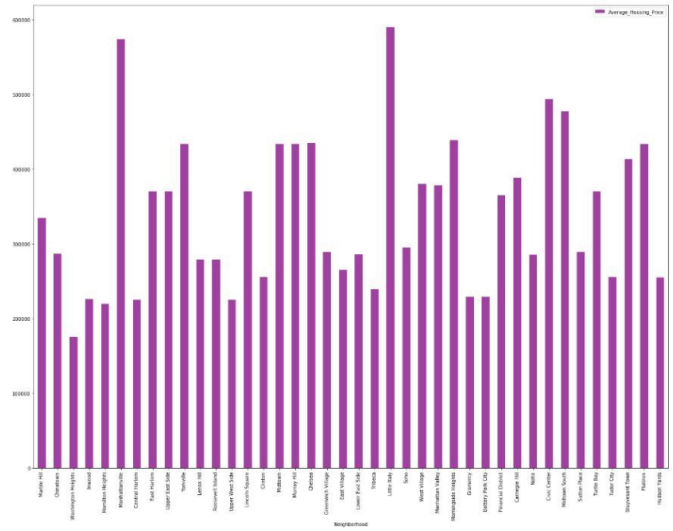
Results Section:

Map of Clusters in Manhattan:

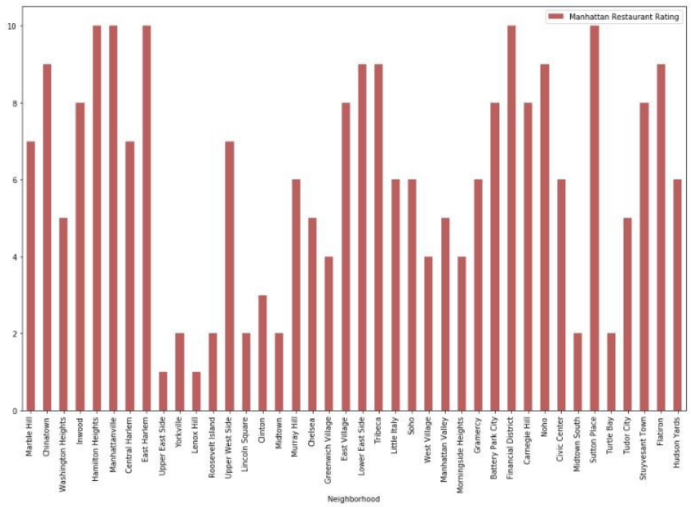




Average Housing Price:



Restaurant Ratings:



Discussion Section:

Solution to the problem: The major purpose of this project is to suggest a better neighborhood in a new city for people who are shifting there.
Here, I did the below:
- Sorted list of house in terms of housing prices in an ascending order.
- Sorted restaurant ratings in terms of location, rating and reviews.

The project can be extended for the below or may be more scenarios, to name a few below:
- Making it more precise in terms to find best house/restaurant/school in Manhattan.
- Connectivity to the airport, railway station, city center, markets and other daily needs things nearby.

Conclusion Section:

Have used k-means cluster algorithm. I separated the neighborhood into 5(five) different clusters and for 40 different latitude and longitude from data-set, which have very similar neighborhoods around them. Using the

data-set, which have very-similar neighborhoods around them. Using the charts above results presented to a particular neighborhood based on average house prices and restaurant rating have been made.

I think I have acquired a lot through this course and especially with this project. It has shown me a practical application to resolve a real life situation which directly impacts personal and financial aspects of a person who moves to a new locality. The same can be used by tweaking a different data-set for a another location using Data Analysis, Data Visualization, Machine Learning. Learning Folium, Geocoder, Beautiful Soup have just been a feather on the cap.

This has been my first self made Data Science project. Any feedback/suggestions are welcome :)
Will continue learning and will keep getting better and better.
LIFE LONG LEARNER!!

Github Link :
https://github.com/kurrysamir/Coursera_Capstone/tree/master/Capstone%20Week%205