Capstone Project –
The Battle of Neighborhoods

Final Project: Recommending Location to open a new coffee shop based on K-Means Algorithm

Version 1.0

Author: Bharath Marimuthu

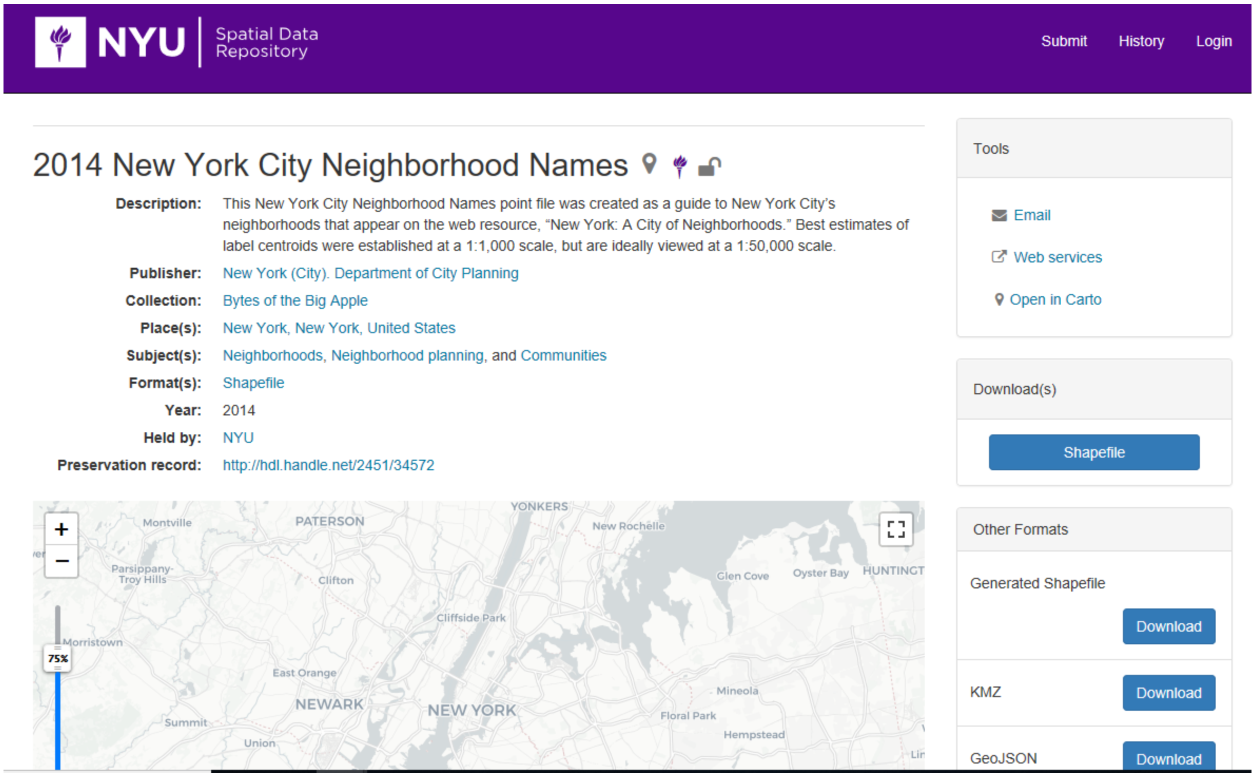
1. **Introduction**:

A coffee house may be one of the simpler businesses to start among businesses in the food and beverage industry. The ability to plan ahead and meet the needs of customers will help create a successful business that will endure time. The location should be energy efficient.

This project recommends location to an entrepreneur to open a coffee shop in New York City using data science. Whenever people want to open a new shop, they explore the place and try to fetch as much information as possible around it. It can be the neighborhood, venues, etc., This is can be termed as request for a search algorithm which usually returns the requested features such as population rate, schools/colleges/offices around, weather conditions, recreational facilities etc. It would be beneficial to have an application which could make easy by considering a comparative analysis between the neighborhood with provided factors.

2. **Data Section:**

New York City Neighborhood Names point file from https://geo.nyu.edu/catalog/nyu_2451_34572. It has a total of 5 boroughs and 306 neighborhoods.



2014 New York City Neighborhood Names

Description: This New York City Neighborhood Names point file was created as a guide to New York City's neighborhoods that appear on the web resource, "New York: A City of Neighborhoods." Best estimates of label centroids were established at a 1:1,000 scale, but are ideally viewed at a 1:50,000 scale.

Publisher: New York (City). Department of City Planning

Collection: Bytes of the Big Apple

Place(s): New York, New York, United States

Subject(s): Neighborhoods, Neighborhood planning, and Communities

Format(s): Shapefile

Year: 2014

Held by: NYU

Preservation record: http://hdl.handle.net/2451/34572

Figure 1 : Webpage where neighborhood of New York acquired.

2.1 Foursquare API:

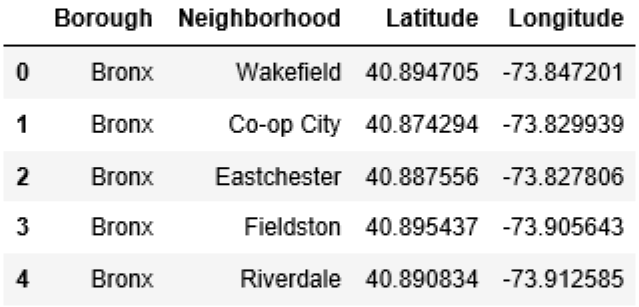
It has a database of more than 105 million places. This project would use Four-square API as its prime data gathering source.

2.2 Python Library Files:

- Pandas - Library for Data Analysis
- NumPy – Library to handle data in a vectorized manner
- JSON – Library to handle JSON files
- Folium – Map rendering Library
- Matplotlib – Python Plotting Module
- Geopy – To retrieve Location Data
- Requests – Library to handle http requests
- Sklearn – Python machine learning Library

2.3 Folium:

Python visualization library would be used to visualize the neighborhoods cluster distribution of Chicago city over an interactive leaflet map. Extensive comparative analysis of two randomly picked neighborhoods world be carried out to derive the desirable insights from the outcomes using python's scientific libraries Pandas, NumPy and Scikit-learn.



	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

Table 1 : Neighborhood and corresponding geo location.

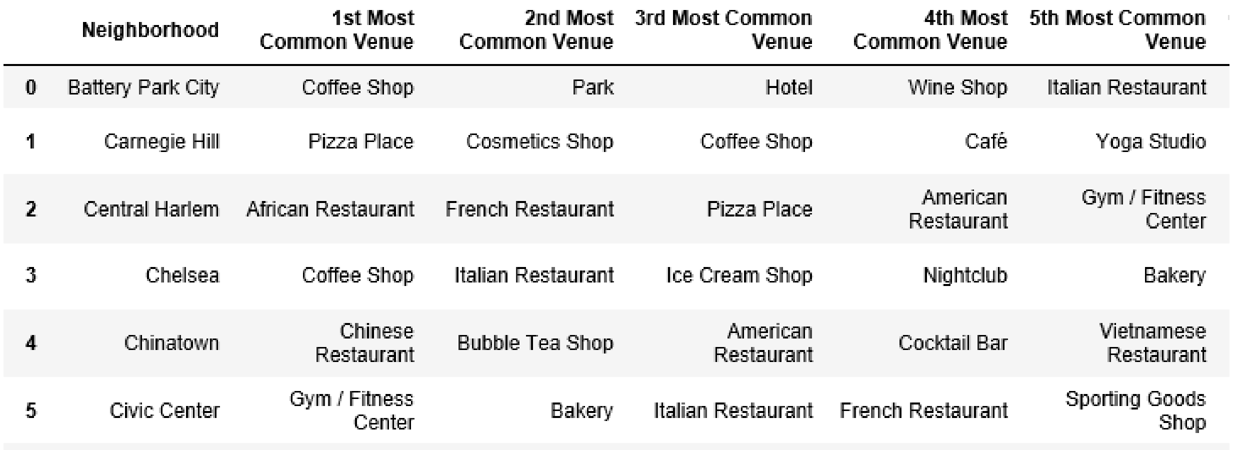
3. Methodology:

Once the neighborhood GPS data has been acquired for any given city the foursquare API call can be used to acquire the 10 most common 'Trending' venues around each neighborhood GPS point. The radius was set to 500m with a limit of 100 venues to be returned.

The returned venues are then grouped using a hot encoding method to display for top 5 venues for each neighborhood. Refer table 2.

3.1 Unsupervised machine learning algorithm:¶

K-mean clustering would be applied to form the clusters of different categories of places in and around the neighborhoods. Each of them would be analyzed individually and comparatively to derive the best location.

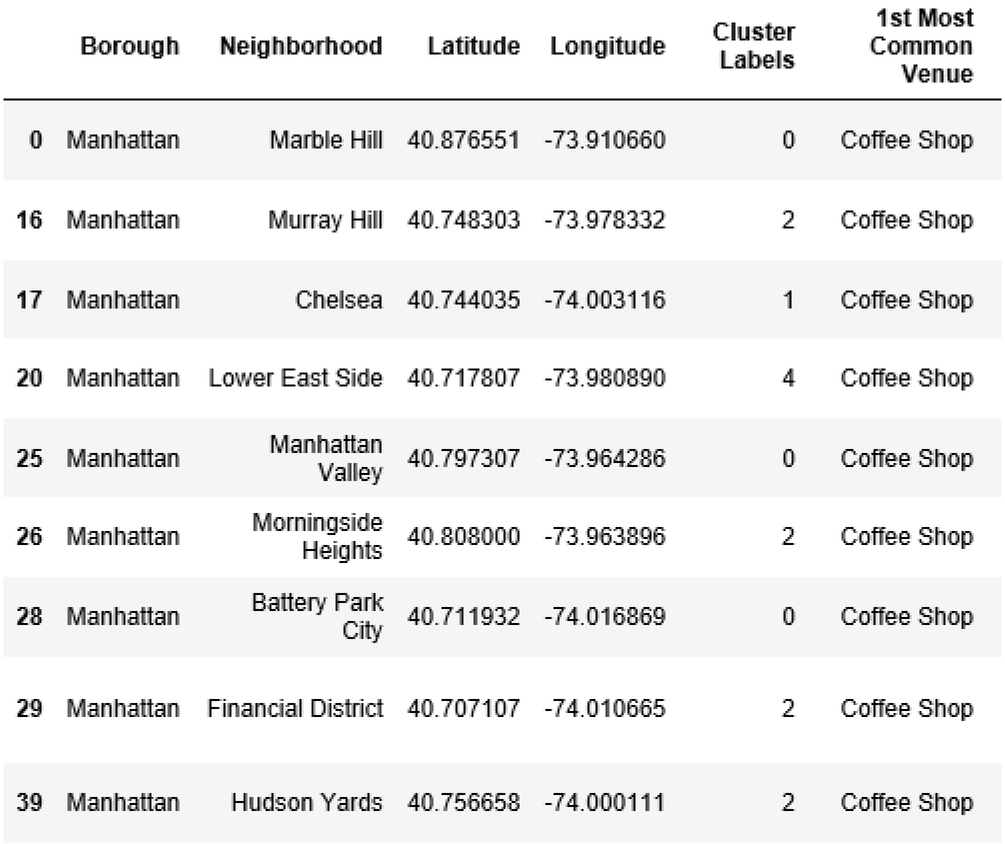


	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Battery Park City	Coffee Shop	Park	Hotel	Wine Shop	Italian Restaurant
1	Carnegie Hill	Pizza Place	Cosmetics Shop	Coffee Shop	Café	Yoga Studio
2	Central Harlem	African Restaurant	French Restaurant	Pizza Place	American Restaurant	Gym / Fitness Center
3	Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Nightclub	Bakery
4	Chinatown	Chinese Restaurant	Bubble Tea Shop	American Restaurant	Cocktail Bar	Vietnamese Restaurant
5	Civic Center	Gym / Fitness Center	Bakery	Italian Restaurant	French Restaurant	Sporting Goods Shop

Table 2 : Dataframe demonstrating top 5 venues of each neighborhood

4. **Results**:

The most visited/common venue is the best location for opening new shop. This model identified 9 best locations to open a new coffee shop based on input.



	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	0	Coffee Shop
16	Manhattan	Murray Hill	40.748303	-73.978332	2	Coffee Shop
17	Manhattan	Chelsea	40.744035	-74.003116	1	Coffee Shop
20	Manhattan	Lower East Side	40.717807	-73.980890	4	Coffee Shop
25	Manhattan	Manhattan Valley	40.797307	-73.964286	0	Coffee Shop
26	Manhattan	Morningside Heights	40.808000	-73.963896	2	Coffee Shop
28	Manhattan	Battery Park City	40.711932	-74.016869	0	Coffee Shop
29	Manhattan	Financial District	40.707107	-74.010665	2	Coffee Shop
39	Manhattan	Hudson Yards	40.756658	-74.000111	2	Coffee Shop

Table 3 : Data frame demonstrating top neighborhoods has top most common venue as Coffee shop.

5. **Discussion**:

From the results, an entrepreneur can apply this model to any city and produce a best location suggestion without any prior knowledge of the city. The disadvantage of the this system is location suggestion not considered population density and crime rate of the city. Using other end points may be a better solution.

6. **Conclusion**:

This model can be applied to any city where the GPS locations of a neighborhood are known. As it stands the model breaks the neighborhoods into 5 clusters of similar trending values. This model will cut down on manual research time and allow an entrepreneur to expand faster.