

Capstone course project : Battle of Neighborhoods

Project Report : By Jigisha Desai

----- Table of Contents -----

1. Introduction Section

1.1 Scenario and Background

1.2 Problem to be solved

1.3 Interested Audience

2. Data Section

2.1 Data of Current city, the city of London

2.2 Data required to solve the problem

2.3 How the data will be used to solve the problem?

3. Methodology

3.1 Exploratory data analysis & Selecting Machine Learning
Algorithm

4. Result

5. Discussion

6. Conclusion

1. Introduction Section:

Define the business problem and the target audience

1.1 Scenario and Background:

Mr. X lives in the city of London with his wife. They love their neighborhood, mainly because of all the great amenities, transport facilities and other types of venues that exist in the neighborhood, such as gourmet fast food joints, multi cuisine restaurants, art gallery, pharmacies, parks, and so on. Mr. X has received a job offer from a great company in Manhattan, New York with great career prospects. However given the far distance from his current place he unfortunately must move if he decide to accept the offer. Wouldn't it be great if he is able to determine neighborhoods in the new city that are exactly the same as his current neighborhood, and if not perhaps similar neighborhoods that are at least closer to his new job?

1.2 Problem to be solved:

To develop a system/project/map that will help Mr. X to get an idea about the transport facilities, different venues and a rental apartment in Manhattan, New York that offers characteristics and benefits similar to his current neighborhood.

The system will provide the details about all the subway stations, rental places with the monthly rental between US\$2000 and US\$6000 and different venues in Manhattan. It will also provide a facility to check the distance between a subway station and a rental place or a distance between a rental place and a venue. This will help Mr. X to make a good choice of a similar apartment in Manhattan NY.

1.3 Interested Audience:

This system/project will be useful for a person who wants to relocate from a current city to a major city in Europe, US or Asia (since the approach and methodologies used here are applicable in all cases) and wants to get an idea about the new neighborhood. This project can also be helpful to those who want

to know about the amenities, transport facilities, venues in a major city in Europe, US or Asia just for knowledge or curiosity.

Although there are various websites available to check the transport facilities, rental apartments, different venues like coffee shop or restaurant etc. available in a major city, this project will bring all these details under one roof and will certainly help the interested audience by saving their time and effort.

2. Data Section:

Data section describes the sources of the data collected to solve the problem. It also indicates how the data will be used to solve the problem.

2.1 Data of Current city, the city of London:

Mr. X currently lives in the city of London. To get the details about the venues in London city, Foursquare credentials and APIs are used. These venues are then shown on the London map. The details about Foursquare APIs and how to show venues on London map are given in "Methodology section" of the report. It serves as a reference for comparison with the desired future location in Manhattan NY.

2.2 Data required to solve the problem:

In order to make a good choice of a similar apartment in Manhattan NY, we need data about subway stations, rental places and venues in Manhattan.

- **Subway Stations Data:**

To retrieve subway stations names and address (about 125 stations) in Manhattan following website is used.

https://en.wikipedia.org/wiki/Category:New_York_City_Subway_stations_in_Manhattan

To get the latitude and longitude of all the subway stations, “Nominatim” function from “geocoders” library is used.

List of the subway stations along with their latitude and longitude are saved in "**subway_stations.csv**" file.

	stations	Lat	Long
0	103rd Street (IND Eighth Avenue Line)	40.796040	-73.961420
1	103rd Street (IRT Broadway–Seventh Avenue Line)	40.799419	-73.968158
2	103rd Street (IRT Lexington Avenue Line)	40.790290	-73.947687
3	106th Street (Second Avenue Subway)	40.790600	-73.942500
4	10th Avenue (IRT Flushing Line)	40.759000	-73.996000
5	110th Street (IRT Lexington Avenue Line)	40.794773	-73.944426
6	116th Street (IND Eighth Avenue Line)	40.804389	-73.955412
7	116th Street (IRT Lenox Avenue Line)	40.802000	-73.950000
8	116th Street (IRT Lexington Avenue Line)	40.798607	-73.942022
9	116th Street (Second Avenue Subway)	40.797000	-73.938000
10	116th Street–Columbia University (IRT Broadway–Seventh Avenue Line)	40.808000	-73.964000
11	125th Street (IND Eighth Avenue Line)	40.810756	-73.952665

- **Rental Places Data:**

To retrieve the details of rental places with the monthly rental between US\$2000 and US\$6000 following websites are used. Details like rent, address, no of beds, bathrooms etc. are retrieved.

<https://www.realtor.com/>

<http://www.rentmanhattan.com/>

<https://www.calibernyc.com/>

To get the latitude and longitude of the rental places, “**Nominatim**” function from “**geocoders**” library is used.

Details of the rental places along with their latitude and longitude are saved in "**rental_places.csv**" file.

	Address	Price	Details	Lat	Long
0	EAST 47TH STREET	\$2,300	Studio 1 BT	40.754613	-73.973734
1	265 EAST 78TH STREET	\$2,444	1 BR 1 BT	40.772875	-73.955833
2	1680 YORK AVENUE	\$3,295	2 B 1 BT	40.777483	-73.945756
3	304 EAST 62ND STREET	\$2,287	1 BR 1 BT	40.762117	-73.962541
4	516 EAST 78TH STREET	\$2,200	Studio 1 BT	40.769999	-73.949969

• Venues Details:

To retrieve different venues in Manhattan, first we need to get the dataset of neighborhoods in Manhattan. This can be retrieved from “**newyork_data.json**” file or from the following website.

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

Now we can get the details about the venues in Manhattan using neighborhood details, Foursquare credentials and APIs.

	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	Coffee Shop	Hotel	Gym	Wine Shop	Shopping Mall	Italian Restaurant	Food Court	Burger Joint	Fountain
1	Carnegie Hill	Coffee Shop	Pizza Place	Bar	Café	Yoga Studio	Spa	Bookstore	Cosmetics Shop	French Restaurant	Gym
2	Central Harlem	African Restaurant	American Restaurant	Public Art	Gym / Fitness Center	French Restaurant	Chinese Restaurant	Seafood Restaurant	Cycle Studio	Bookstore	Ethiopian Restaurant
3	Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bakery	American Restaurant	Nightclub	Hotel	Theater	Seafood Restaurant	Asian Restaurant
4	Chinatown	Chinese Restaurant	American Restaurant	Vietnamese Restaurant	Cocktail Bar	Dim Sum Restaurant	Noodle House	Salon / Barbershop	Bakery	Hotpot Restaurant	Ice Cream Shop

2.3 How the data will be used to solve the problem?

- Venue details collected about the current city, London will be displayed on the map of London. Categories of the five most available venues will also be displayed. This will be helpful to compare venues in Manhattan.
- Subway stations details stored in "**subway_stations.csv**" file will be retrieved and stations will be displayed on the map of Manhattan.
- Rental places details stored in "**rental_places.csv**" file will be retrieved and places will be displayed on the map of Manhattan.

- Venue details retrieved using Foursquare credentials and APIs will be clustered using the clustering algorithm and then clusters will be displayed on the map of Manhattan.
- To explore **subway stations** and **rental places**, the map will be created to show both the details together.
- To explore **rental places and clusters of venues**, the map will be created to show both the details together.
- The final map will be created that shows details about **rental places**, **subway stations** and **clusters of venues** together. It will also provide a facility to check the distance between a subway station and a rental place or a distance between a rental place and a venue. This will help Mr. X to make a good choice of a similar apartment in Manhattan NY.

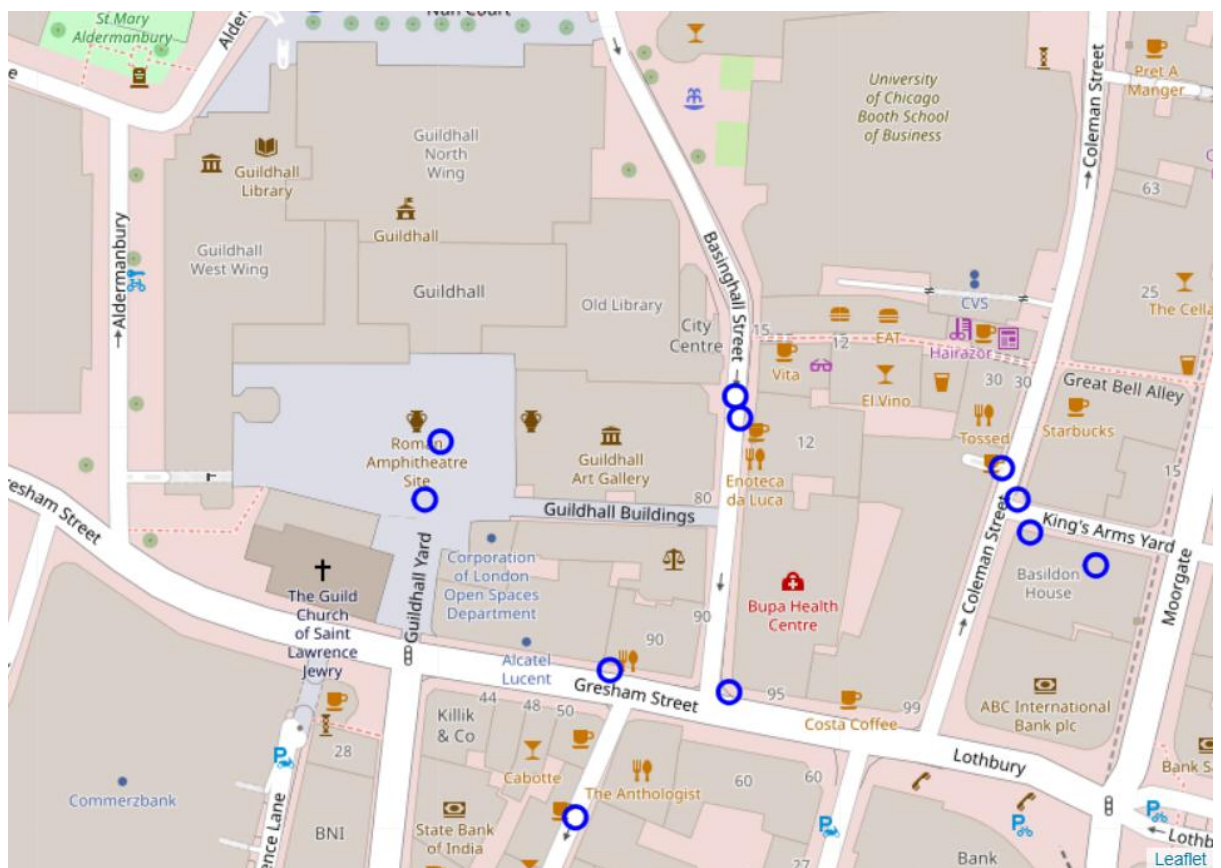
3. Methodology:

Methodology section which represents the main component of the report where one discusses and describes any exploratory data analysis that was done, any inferential statistical testing that was performed, and what machine learning were used and why.

- As a database, I used rental places details and subway stations details in Manhattan in my study. These details were retrieved using web scrapping. Rental places details has the main components Address, Monthly Rental, Details, Latitude and Longitude information of the city. Subway stations details has the main components Station name, Latitude and Longitude.
- I utilized the Foursquare API to explore the venues in London. I designed the limit as **100 venues** and the radius **500 meter**. Here is a head of the list Venues name, category, latitude and longitude information from Forsquare API.

	name	categories	lat	lng
0	Goodman Steak House Restaurant	Steakhouse	51.514398	-0.090745
1	Hawksmoor Guildhall	Steakhouse	51.515647	-0.090997
2	City Càphê	Vietnamese Restaurant	51.514750	-0.091545
3	Daunt Books	Bookstore	51.513982	-0.092995
4	The Ned Hotel	Hotel	51.513565	-0.090166
5	The Merchant House	Cocktail Bar	51.513264	-0.093039

- I used python **folium** library to visualize geographic details of venues in London. I used latitude and longitude values to get the visual as below:



- To retrieve different venues in Manhattan, first we need to get the dataset of neighborhoods in Manhattan. This can be retrieved from “newyork_data.json” file.

Exploratory data analysis & Selecting Machine Learning Algorithm:

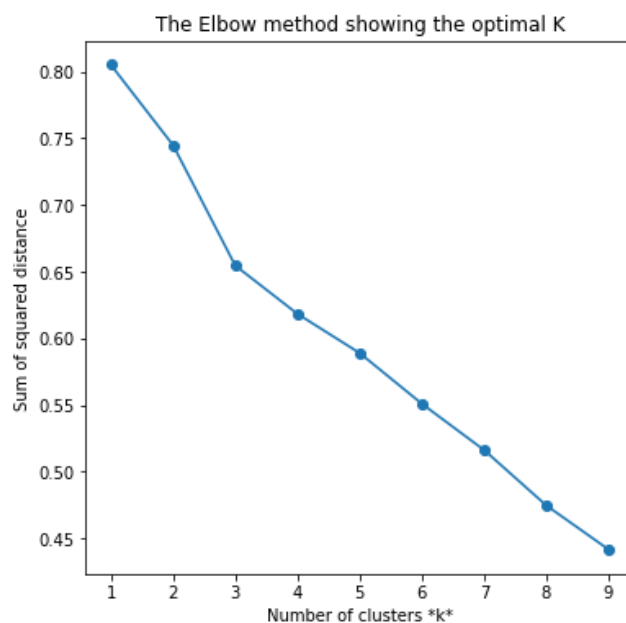
- Again I utilized the Foursquare API to explore the venues in Manhattan. I designed the limit as **100 venues** and the radius **500 meter**. Then I

created a table which shows list of top 10 venue category for each borough in below table.

	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	Coffee Shop	Hotel	Gym	Wine Shop	Shopping Mall	Italian Restaurant	Food Court	Burger Joint	Fountain
1	Carnegie Hill	Coffee Shop	Pizza Place	Bar	Café	Yoga Studio	Spa	Bookstore	Cosmetics Shop	French Restaurant	Gym
2	Central Harlem	African Restaurant	American Restaurant	Public Art	Gym / Fitness Center	French Restaurant	Chinese Restaurant	Seafood Restaurant	Cycle Studio	Bookstore	Ethiopian Restaurant
3	Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bakery	American Restaurant	Nightclub	Hotel	Theater	Seafood Restaurant	Asian Restaurant
4	Chinatown	Chinese Restaurant	American Restaurant	Vietnamese Restaurant	Cocktail Bar	Dim Sum Restaurant	Noodle House	Salon / Barbershop	Bakery	Hotpot Restaurant	Ice Cream Shop

- We have some common venue categories in boroughs. In this reason I used unsupervised learning K-means algorithm to cluster the boroughs. K-Means algorithm is one of the most common cluster methods of unsupervised learning.

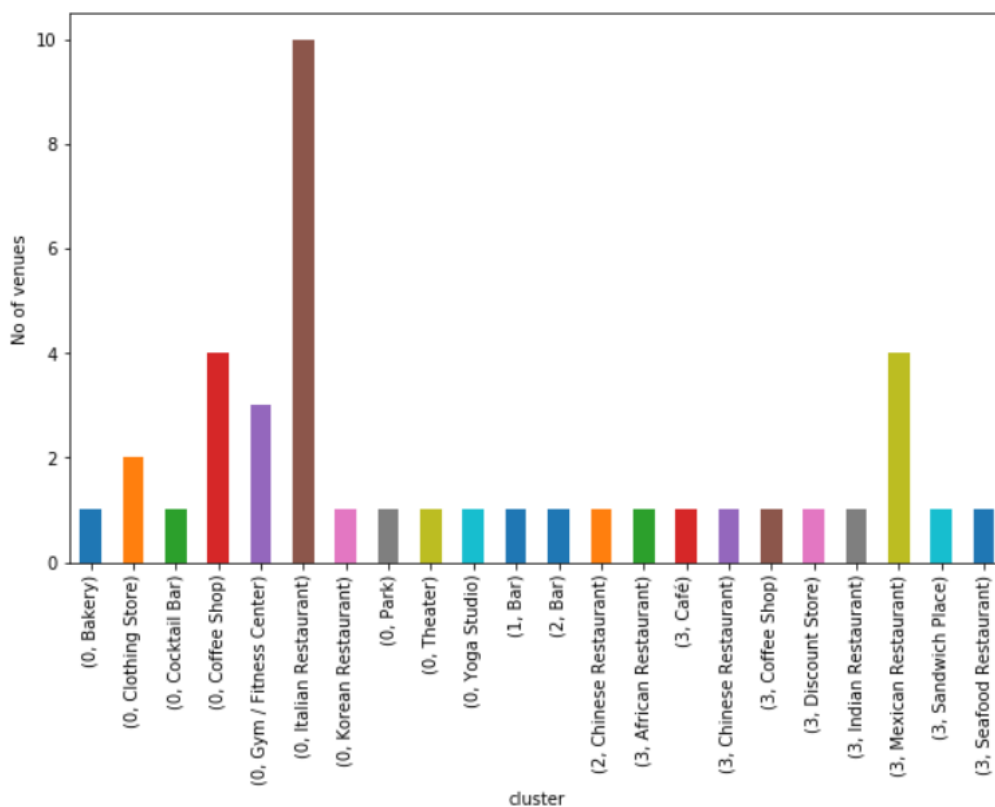
First, I will run K-Means to cluster the boroughs into 4 clusters because when I analyze the K-Means with elbow method it ensured me that optimum k of the K-Means can be 3 or 4.



- Here is my merged table with cluster labels for each borough.

	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
0	Manhattan	Marble Hill	40.876551	-73.910660	3	Discount Store	Coffee Shop	Yoga Studio	Clothing Store	Tennis Stadium	Big Box Store	Supplement Shop	Spa	
1	Manhattan	Chinatown	40.715618	-73.994279	2	Chinese Restaurant	American Restaurant	Vietnamese Restaurant	Cocktail Bar	Dim Sum Restaurant	Noodle House	Salon / Barbershop	Bakery	Rest
2	Manhattan	Washington Heights	40.851903	-73.936900	3	Café	Bakery	Mobile Phone Shop	Deli / Bodega	Gym / Fitness Center	Supermarket	Spanish Restaurant	Latin American Restaurant	Rest
3	Manhattan	Inwood	40.867684	-73.921210	3	Mexican Restaurant	Café	Deli / Bodega	Lounge	Pizza Place	Restaurant	Chinese Restaurant	Frozen Yogurt Shop	
4	Manhattan	Hamilton Heights	40.823604	-73.949688	3	Mexican Restaurant	Coffee Shop	Café	Pizza Place	Deli / Bodega	Liquor Store	Indian Restaurant	Sushi Restaurant	Sar

- We can also estimate the number of **1st Most Common Venue** in each cluster. Thus, we can create a bar chart which may help us to find proper labels for each cluster.



- When we examine above graph we can label each cluster as follows:

Cluster 0: Coffee shop venues & Intensive Italian restaurants

Cluster 1: Bar venues

Cluster 2: Chinese restaurants

Cluster 3: Multi-cuisine restaurants

4. Result:

- The result of cluster analysis is as follows:

Cluster 0:

```
# Cluster 0
#Cluster Label:Coffee shop venues & Intensive Italian restaurants
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 0, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

	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
8	Upper East Side	Italian Restaurant	Coffee Shop	Exhibit	Art Gallery	Bakery	Juice Bar	Gym / Fitness Center	French Restaurant	Boutique	Hotel
9	Yorkville	Italian Restaurant	Coffee Shop	Gym	Bar	Pizza Place	Diner	Deli / Bodega	Mexican Restaurant	Thai Restaurant	Sushi Restaurant
10	Lenox Hill	Italian Restaurant	Sushi Restaurant	Coffee Shop	Gym / Fitness Center	Gym	Deli / Bodega	Burger Joint	Pizza Place	Sporting Goods Shop	Café
12	Upper West Side	Italian Restaurant	Bar	Vegetarian / Vegan Restaurant	Indian Restaurant	Wine Bar	Mediterranean Restaurant	Bakery	Coffee Shop	Burger Joint	Ice Cream Shop
13	Lincoln Square	Gym / Fitness Center	Theater	Café	Plaza	Concert Hall	Italian Restaurant	French Restaurant	Performing Arts Venue	Park	Opera House
14	Clinton	Theater	American Restaurant	Hotel	Gym / Fitness Center	Indie Theater	Wine Shop	Spa	Coffee Shop	Gym	Italian Restaurant
15	Midtown	Clothing Store	Hotel	Cocktail Bar	Steakhouse	Theater	Bakery	Coffee Shop	Spa	American Restaurant	Bookstore
16	Murray Hill	Coffee Shop	Hotel	Spa	French Restaurant	American Restaurant	Italian Restaurant	Bar	Sandwich Shop	Gym	Japanese Restaurant

Cluster 1:

```
# Cluster 1
#Cluster Label: Bar venues
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 1, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

	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
37	Stuyvesant Town	Bar	Park	Pet Service	Baseball Field	Harbor / Marina	Farmers Market	Beer Garden	Cocktail Bar	Coffee Shop	Heliport

Cluster 2:

```
# Cluster 2
#Cluster Label: Chinese Restaurants
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 2, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

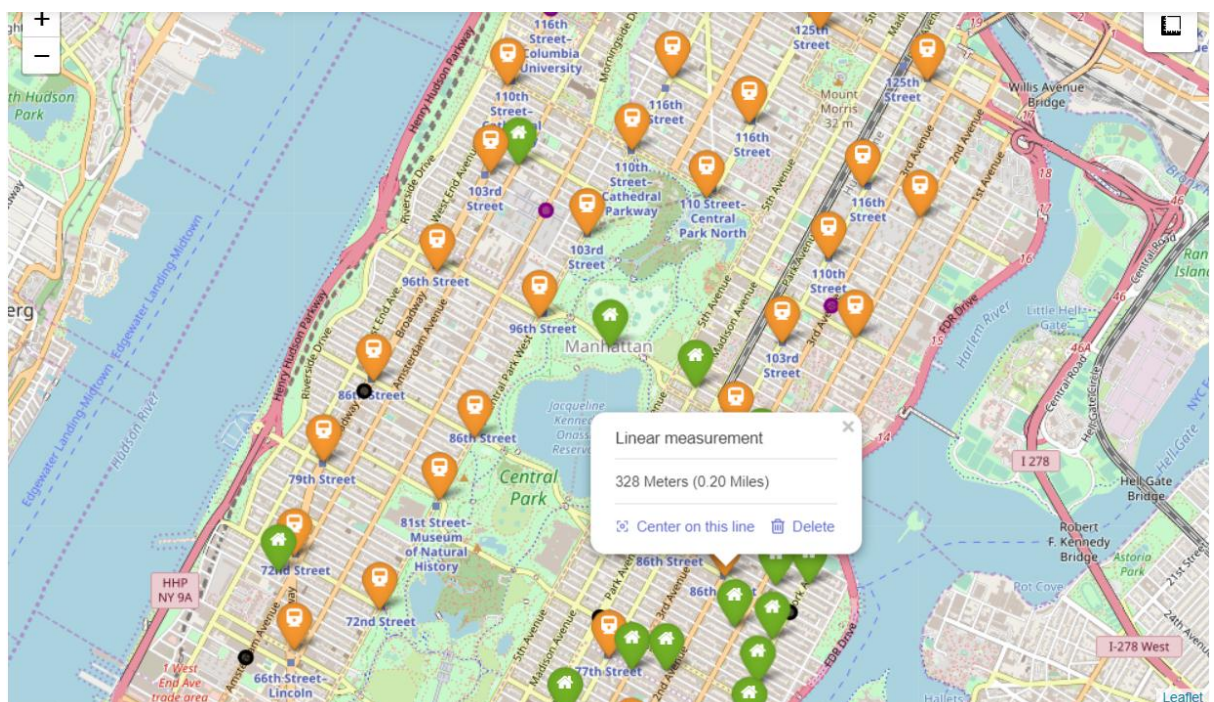
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Chinatown	Chinese Restaurant	American Restaurant	Vietnamese Restaurant	Cocktail Bar	Dim Sum Restaurant	Noodle House	Salon / Barbershop	Bakery	Hotpot Restaurant	Ice Cream Shop
19	East Village	Bar	Wine Bar	Ice Cream Shop	Mexican Restaurant	Chinese Restaurant	Pizza Place	Ramen Restaurant	Cocktail Bar	Japanese Restaurant	Vegetarian / Vegan Restaurant

Cluster 3:

```
# Cluster 3
#Cluster Label: Multi-cuisine restaurants
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 3, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape
```

	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	Marble Hill	Discount Store	Coffee Shop	Yoga Studio	Clothing Store	Tennis Stadium	Big Box Store	Supplement Shop	Spa	Shoe Store	Seafood Restaurant
2	Washington Heights	Café	Bakery	Mobile Phone Shop	Deli / Bodega	Gym / Fitness Center	Supermarket	Spanish Restaurant	Latin American Restaurant	Mexican Restaurant	Tapas Restaurant
3	Inwood	Mexican Restaurant	Café	Deli / Bodega	Lounge	Pizza Place	Restaurant	Chinese Restaurant	Frozen Yogurt Shop	Coffee Shop	Bakery
4	Hamilton Heights	Mexican Restaurant	Coffee Shop	Café	Pizza Place	Deli / Bodega	Liquor Store	Indian Restaurant	Sushi Restaurant	Sandwich Place	Yoga Studio
5	Manhattanville	Seafood Restaurant	Mexican Restaurant	Park	Italian Restaurant	Coffee Shop	Chinese Restaurant	Food & Drink Shop	Bike Trail	Sushi Restaurant	Supermarket
6	Central Harlem	African Restaurant	American Restaurant	Public Art	Gym / Fitness Center	French Restaurant	Chinese Restaurant	Seafood Restaurant	Cycle Studio	Bookstore	Ethiopian Restaurant
7	East Harlem	Mexican Restaurant	Bakery	Deli / Bodega	Latin American Restaurant	Thai Restaurant	Pet Store	Convenience Store	Dance Studio	Beer Bar	Pizza Place

- Using Folium library in Python, The final map will be created that shows details about **rental places**, **subway stations** and **clusters of venues** together. It will also provide a facility to check the distance between a subway station and a rental place or a distance between a rental place and a venue. This will help Mr. X to make a good choice of a similar apartment in Manhattan NY.



- Using Folium library in Python, maps are also created to show details about **rental places**, **subway stations** together and details about **rental places**, **cluster venues** together.



- By comparing the top 5 categories of the venues in London with these 4 clusters, we can say that **Cluster 0** resembles more to the venues in neighborhood in London.
- By examining the map with clusters, subway stations and rental places, we can say that neighborhoods like "Yorkville", "Lenox Hill", "Carnegie Hill", "Turtle Bay" can be ideal for selecting rental place as they have many rental places with the monthly rental between US\$2000 and US\$6000, there are many subway stations situated near by these rental places and the venues around these places resemble more to the venues in the current neighborhood in London.
- The best choice among the above mentioned neighborhoods is **Yorkville**. Few options for the rental places in Yorkville with the distance to these places from the nearby subway stations is as mentioned below.

Subway Stations	Rental Places	Distance
86th Street (Second Avenue Subway)	East 83rd Street \$3695 2 Beds	279 mt(0.17 miles)
86th Street (Second Avenue Subway)	York Avenue \$2395 2 Beds	457 mt(0.28 miles)
86th Street (Second Avenue Subway)	EAST 88TH STREET \$4500 2 Beds	191 mt(0.12 miles)
86th Street (Second Avenue Subway)	EAST 87TH STREET \$3000 1 Bed	328 mt(0.20 miles)
96th Street (Second Avenue Subway)	EAST 95TH STREET \$2963 1 BR	174 mt(0.11 miles)
96th Street (Second Avenue Subway)	East 91st street \$4650 2 Beds	458 mt(0.28 miles)
96th Street (Second Avenue Subway)	East 91st Street \$4800 3 Beds	531 mt(0.33 miles)
96th Street (Second Avenue Subway)	EAST 88TH STREET \$4500 2 Beds	614 mt(0.38 miles)
96th Street (Second Avenue Subway)	EAST 87TH STREET \$3000 1 Bed	772 mt(0.48 miles)

5. **Discussion:**

- New York City comprises 5 boroughs sitting where the Hudson River meets the Atlantic Ocean. At its core is Manhattan, a densely populated borough that's among the world's major commercial, financial and cultural centers.

- London is the capital and largest city of both England and the United Kingdom. London is considered to be one of the world's most important global cities and has been termed the world's most powerful, most desirable, most influential, most visited, most expensive, innovative, sustainable, most investment friendly, most popular for work, and the most vegetarian friendly city in the world. London exerts a considerable impact upon the arts, commerce, education, entertainment, fashion, finance, healthcare, media, professional services, research and development, tourism and transportation.
- Both Manhattan and London are the big cities. However, in this project when the venues are explored using the Foursquare API the limit is set to **100 venues** and the radius **500 meter**. For more detailed and accurate result from K-means, the data set can be expanded by increasing these limits.
- Also for getting the rental places details only limited number of websites are explored. By exploring more number of the websites we can provide more options to the audience of this project for rental place selection.

6. Conclusion:

People are turning to the new cities/countries for a job. For this reason, before actually moving to the new city, people can determine neighborhoods in the new city that are exactly the same as their current neighborhood, and if not perhaps similar neighborhoods that are at least closer to their new job through their access to the platforms where the detailed information like transportation, rental places, venues are mentioned together. This will indeed make people's life easier in the new city.

Thank You,

Jigiha Desai