Capstone Project – REPORT

Contents:

1. Introduction Section:

    1.1 Discussion of the "background situation" leading to the problem at hand:
    1.2 Problem to be resolved
    1.3 Audience for this project.

2. Data Section:

    2.1 Data of Current Situation (current residence place)
    2.2 Data required to resolve the problem
    2.3 Data sources and data manipulation

3. Methodology section:

    3.1 Process steps and strategy to resolve the problem
    3.2 Data Science Methods, machine learning, mapping tools and exploratory data analysis.
4. Results Section:

    Discussion of the results and how they help to take a decision.

5. Discussion Section:

    Elaboration and discussion on any observations and/or recommendations for improvement.

6. Conclusion Section

    Decision taken and Report Conclusion.

1. Introduction Section:

Discussion of the business problem and the audience who would be interested in this project.

1.1 Scenario and Background

I have picked Gurgaon, Haryana in India as the current location for this Project

Within walking distance to many amenities and venues in the area, such as various international cuisine restaurants, cafes, gyms, food shops and entertainment.

1.2 Problem to be resolved:

The challenge to resolve is being able to find an apartment unit in Manhattan NY that offers similar characteristics and benefits to my current situation. Therefore, in order to set a basis for comparison, I want to place subject to the following conditions:

- Top ammenities in the selected neighborhood shall be similar to current residence (See item 2.1)
- Desirable to have venues such as coffee shops, restaurants, wine stores, gym and food shops
- As a reference, I have included a map of venues near current residence in Gurgaon Haryana.

1.3 Interested Audience

I believe this is a relevant project for a person or entity considering moving to a major city in Europe, US or Asia, since the approach and methodologies used here are applicable in all cases. The use of FourSquare data and mapping techniques combined with data analysis will help resolve the key questions arisen. Lastly, this project is a good practical case toward the development of Data Science skills.

2. Data Section:

Description of the data and its sources that will be used to solve the problem

2.1 Data of Current Situation

I Currently reside in the neighborhood of Gurgaon, Haryana, India.

I use Foursquare to identify the venues around the area of residence which are then shown in the Haryana, Delhi NCR map shown in methodology and execution in section 3.0 . It serves as a reference for comparison with the desired future location in Manhattan NY

2.2 Data Required to resolve the problem

In order to make a good choice of a similar apartment in Manhattan NY, the following data is required:

- List/Information on neighborhoods form Manhattan with their Geodata (latitude and longitude.
- List/Information about the subway metro stations in Manhattan with geodata.
- Listed apartments for rent in Manhattan area with descriptions (how many beds, price, location, address)
- Venues and ammenities in the Manhattan neighborhoods (e.g. top 10)

2.3 sources and manipulation

The list of Manhattan neighbourhoods is worked out during LAB exercise during the course. A csv file was created which will be read in order to create a dataframe and its mapping. The csv file 'mh_neigh_data.csv' has the following below data structure. The file will be directly read to the Jupiter Notebook for convenience and space savings. The clustering of neighbourhoods and mapping will be shown however. An algorithm was used to determine the geodata from Nominatim. The actual algorithm coding may be shown in 'markdown' mode because it takes time to run.

2.4 How the data will be used to solve the problem

The data will be used as follows:

- Foursquare and geopy data to map top 10 venues for all Manhattan neighbourhoods and clustered in groups (as per Course LAB)

2.5 Mapping of Data

The following maps were created to facilitate the analysis and the choice of the places to live.

- Manhattan map of Neighborhoods
- manhattan subway metro locations
- Manhattan map of places for rent
- Manhattan map of clustered venues and neighborhoods
- Combined maps of Manhattan rent places with subway locations
- Combined maps of Manhattan rent places with subway locations and venues clusters

3. Methodology section:

This section represents the main component of the report where the data is gathered, prepared for analysis. The tools described are used here and the Notebook cells indicates the execution of steps.

The analysis and the strategy:

The strategy is based on mapping the above described data in section 2.0, in order to facilitate the choice of at least two candidate places for rent. The choice is made based on the demands imposed : location near a subway, rental price and similar venues to Gurgaon. This visual approach and maps with popups labels allow quick identification of location, price and feature, thus making the selection very easy.

The procesing of these DATA and its mapping will allow to answer the key questions to make a decision:

- what is the cost of available rental places that meet the demands?
- what is the cost of rent around a mile radius from each subway metro station?
- what is the area of Manhattan with best rental pricing that meets criteria established?
- What is the distance from work place (Park Ave and 53 rd St) and the tentative future rental home?
- What are the venues of the two best places to live? How the prices compare?
- How venues distribute among Manhattan neighborhoods and around metro stations?
- Are there tradeoffs between size and price and location?
- - Any other interesting statistical data findings of the real estate and overall data.

4. Results

Current Residence in Gurgaon:
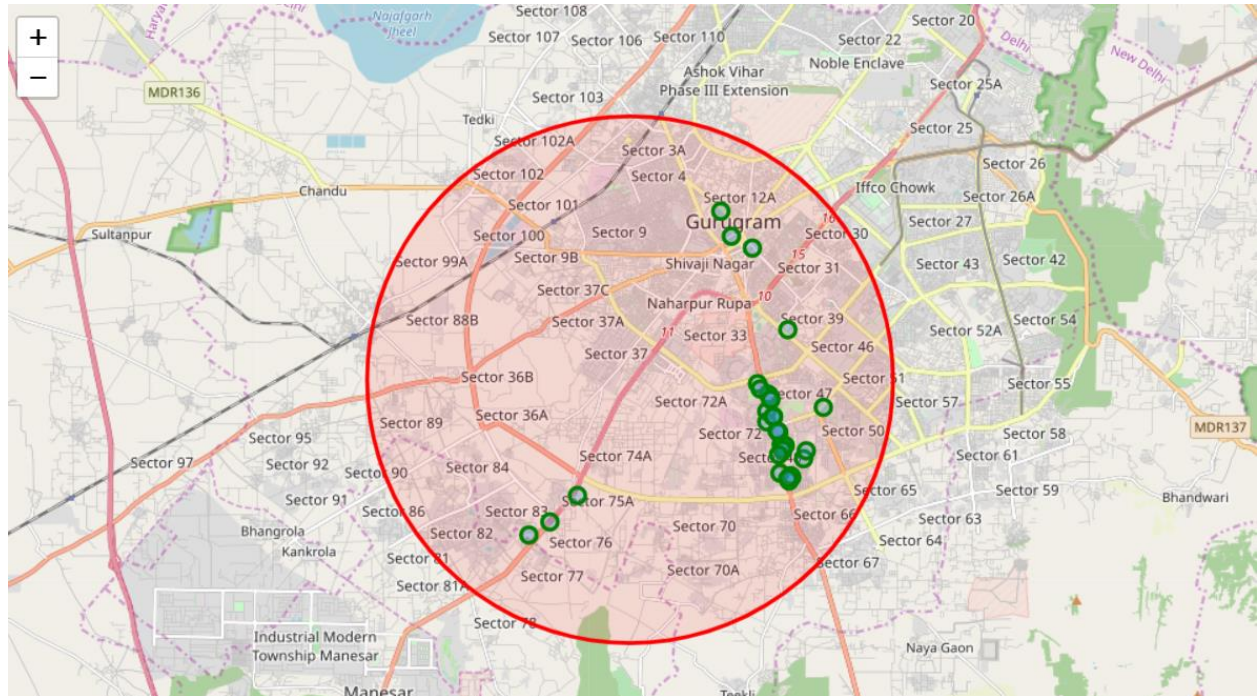
	name	categories	lat	lng
0	Decathlon Gurgaon	Sporting Goods Shop	28.427433	77.036258
1	Brewer Street	Brewery	28.424499	77.039134
2	Bikanervala	Indian Restaurant	28.419831	77.040156
3	Sagar Ratna	Indian Restaurant	28.421042	77.038509
4	Open Tap	Brewery	28.405130	77.044332
5	Good Earth City Centre	Shopping Mall	28.421766	77.053183
6	Starbucks	Coffee Shop	28.406085	77.044741
7	Hyatt Regency Gurgaon	Hotel	28.392552	76.976269
8	Vapour Bar Exchange	Brewery	28.406734	77.041820
9	Haldiram	Indian Restaurant	28.401791	76.989000
10	KFC	Fast Food Restaurant	28.424789	77.038722
11	Shopper's Stop	Clothing Store	28.413425	77.042197

Manhattan Map - Neighborhoods and Cluster of Venues
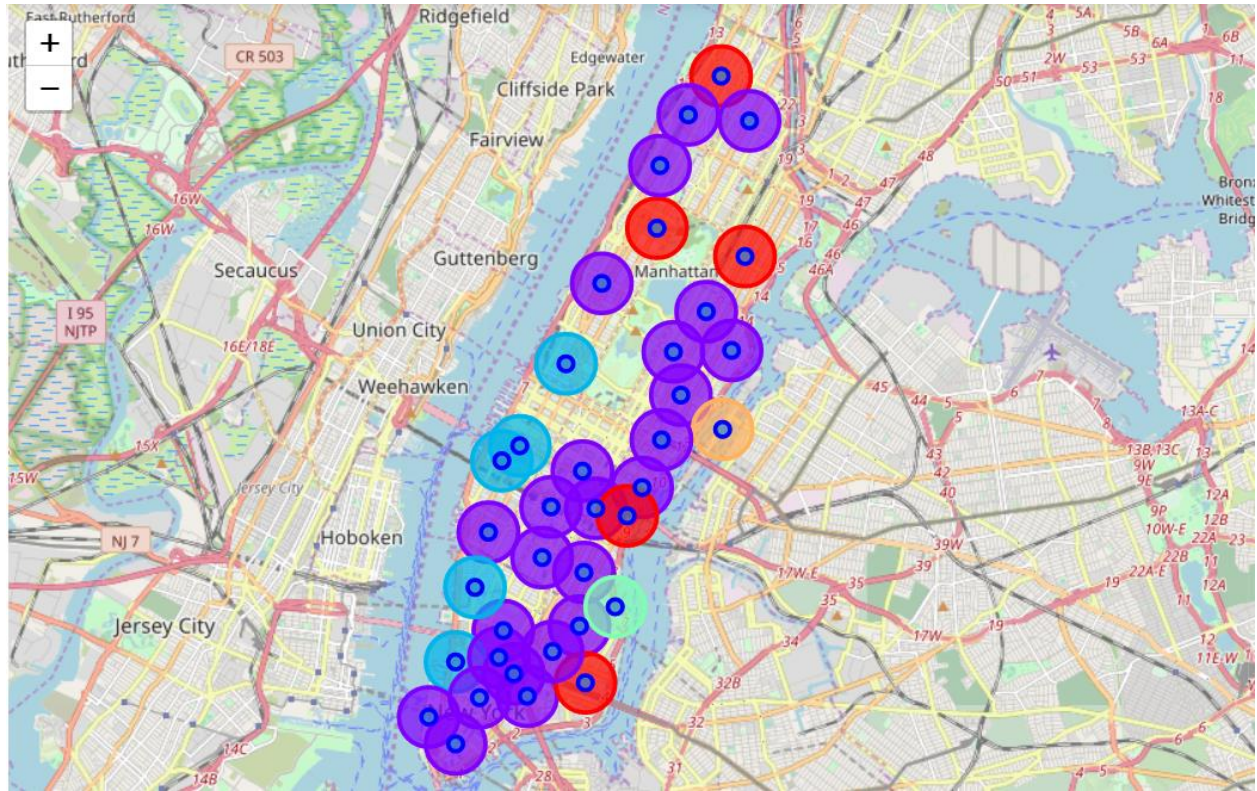
GeoData Manhattan Apartments - Rent

```
mh_rent=pd.read_csv('MH_rent_latlong.csv')
mh_rent.head()
```

	Address	Area	Price_per_ft2	Rooms	Area-ft2	Rent_Price	Lat	Long
0	West 105th Street	Upper West Side	2.94	5.0	3400	10000	40.799771	-73.966213
1	East 97th Street	Upper East Side	3.57	3.0	2100	7500	40.788585	-73.955277
2	West 105th Street	Upper West Side	1.89	4.0	2800	5300	40.799771	-73.966213
3	CARMINE ST.	West Village	3.03	2.0	1650	5000	40.730523	-74.001873
4	171 W 23RD ST.	Chelsea	3.45	2.0	1450	5000	40.744118	-73.995299

```
mh_rent.tail()
```

|	Address	Area	Price_per_ft2	Rooms	Area-ft2	Rent_Price	Lat	Long
139	200 East 72nd Street	Rental in Lenox Hill	5.15	3.0	1700	8750	40.769465	-73.960339
140	50 Murray Street	No fee rental in Tribeca	7.11	2.0	1223	8700	40.714051	-74.009608
141	300 East 56th Street	No fee rental in Midtown East	3.87	3.0	2100	8118	40.758216	-73.965190
142	1930 Broadway	No fee rental in Central Park West	5.06	2.0	1600	8095	40.772474	-73.981901
143	33 West 9th Street	Rental in Greenwich Village	6.67	2.0	1500	10000	40.733691	-73.997323

Rental Price Statistics MH Apartments

Budget US7000/month is around the mean
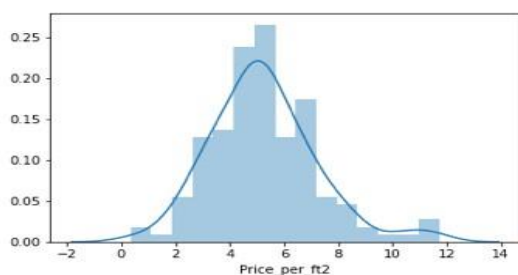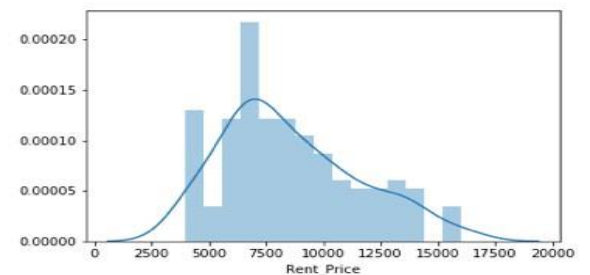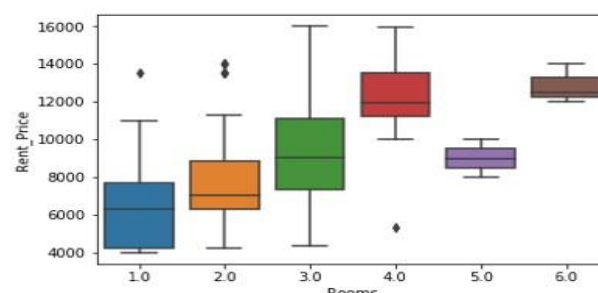


```
sns.boxplot(x='Rooms', y= 'Rent_Price', data=mh_rent)
```
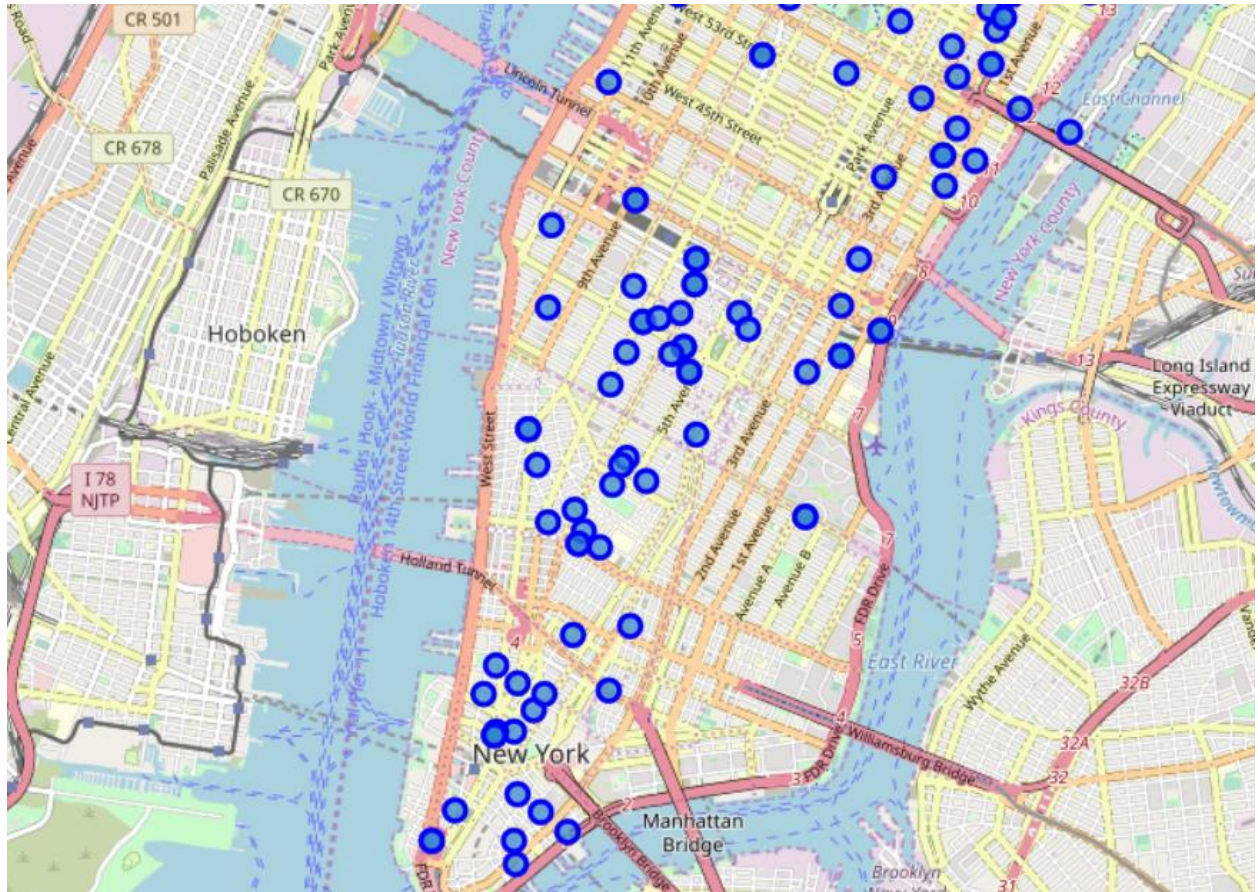
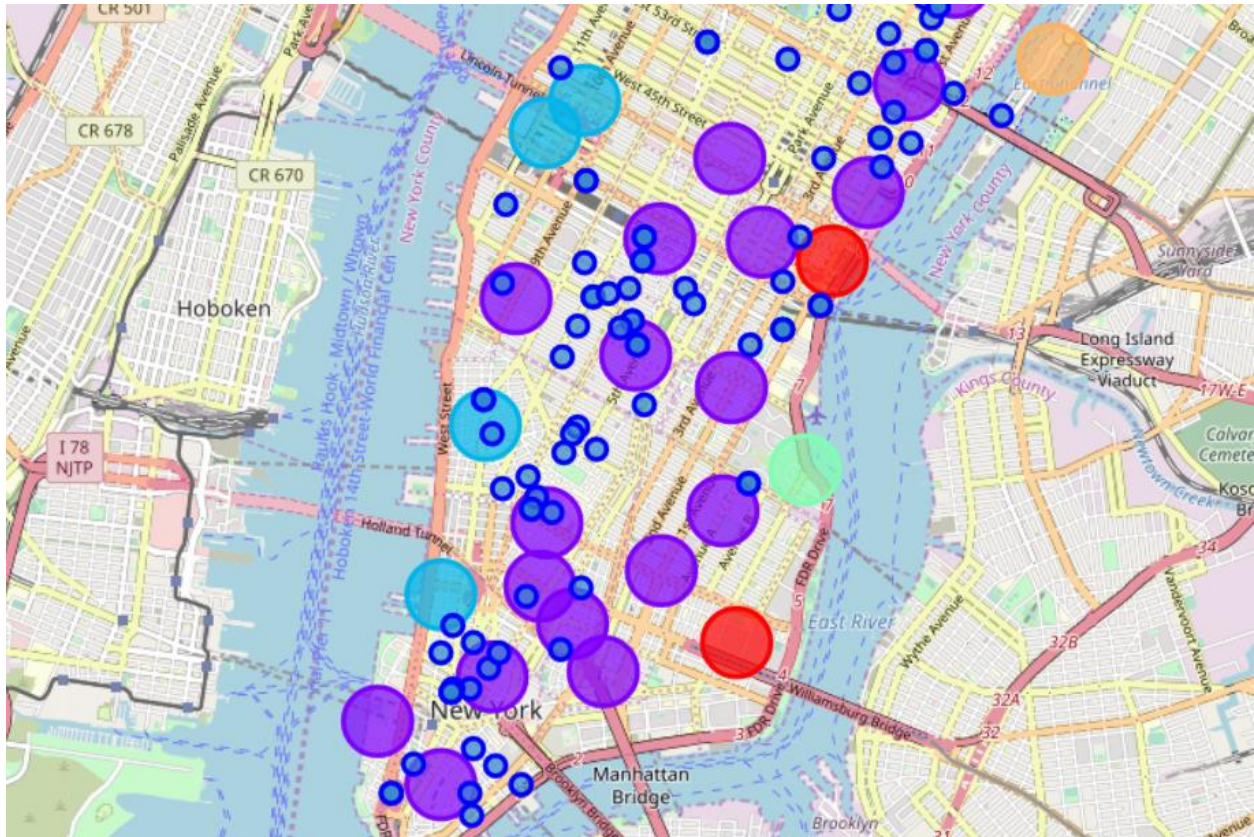<matplotlib.axes._subplots.AxesSubplot at 0x1a25f2a2b0>

Apartments for Rent in Manhattan

Manhattan Apartments For Rent (With Clusters)

On Careful Examination, Soho in Cluster #1 resembles in amenities to our Current Residence

	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	Sandwich Place	Gym	Coffee Shop	Yoga Studio	Deli / Bodega	Steakhouse	Shopping Mall	Seafood Restaurant	Pizza Place	Department Store
1	Chinatown	Chinese Restaurant	Cocktail Bar	American Restaurant	Spa	Bakery	Hotpot Restaurant	Optical Shop	Vietnamese Restaurant	Salon / Barbershop	Dessert Shop
5	Manhattanville	Coffee Shop	Seafood Restaurant	Italian Restaurant	Mexican Restaurant	Chinese Restaurant	Park	Deli / Bodega	Supermarket	Boutique	Spanish Restaurant
6	Central Harlem	African Restaurant	Bar	French Restaurant	American Restaurant	Seafood Restaurant	Chinese Restaurant	Caribbean Restaurant	Spa	Dessert Shop	Beer Bar
8	Upper East Side	Italian Restaurant	Coffee Shop	Exhibit	Art Gallery	Bakery	Gym / Fitness Center	Juice Bar	Pizza Place	French Restaurant	Yoga Studio
9	Yorkville	Italian Restaurant	Gym	Coffee Shop	Bar	Deli / Bodega	Pizza Place	Diner	Mexican Restaurant	Sushi Restaurant	Japanese Restaurant
10	Lenox Hill	Italian Restaurant	Coffee Shop	Sushi Restaurant	Pizza Place	Cocktail Bar	Burger Joint	Café	Gym	Gym / Fitness Center	Mexican Restaurant
12	Upper West Side	Italian Restaurant	Wine Bar	Bar	Coffee Shop	Indian Restaurant	Bakery	Café	Mediterranean Restaurant	Restaurant	Pub
15	Midtown	Hotel	Coffee Shop	Clothing Store	Bookstore	American Restaurant	Theater	Bakery	Sporting Goods Shop	Café	French Restaurant
16	Murray Hill	Sandwich Place	Coffee Shop	American Restaurant	Japanese Restaurant	Italian Restaurant	Gym / Fitness Center	Mediterranean Restaurant	Hotel	Gym	Chinese Restaurant
17	Chelsea	Coffee Shop	Bakery	Italian Restaurant	American Restaurant	Ice Cream Shop	Hotel	Wine Shop	Breakfast Spot	Tapas Restaurant	Cycle Studio
18	Greenwich Village	Italian Restaurant	Clothing Store	Sushi Restaurant	Café	Indian Restaurant	French Restaurant	Gym	Chinese Restaurant	Gourmet Shop	Dessert Shop

Manhattan Subway Stations-GeoData

```
mhsub1.tail()
```

	sub_station	sub_address	lat	long
17	190 Street Subway Station	Bennett Ave, New York, NY 10040, USA	40.858113	-73.932983
18	59 St-Lexington Av Station	E 60th St, New York, NY 10065, USA	40.762259	-73.966271
19	57 Street Station	New York, NY 10019, United States	40.764250	-73.954525
20	14 Street / 8 Av	New York, NY 10014, United States	40.730862	-73.987156
21	MTA New York City	525 11th Ave, New York, NY 10018, USA	40.759809	-73.999282

Venue Selection

- I was able to explore all possibilities since the popups provide the information needed for a good decision.

- Apartment 1 rent cost is US6900. It is located approximately 600 meters from subway station at Prince Street. I can walk to work place and use subway for other places aroung. Venues for this apt are as of Cluster 1 and it is located in a Soho Neighborhood of Manhattan.

- Apartment 2 rent cost is US6935, just under the US7000 budget. Apt 2 is located 60 meters from subway station at Fulton Street, but I will have to ride the subway daily to work possibly 40-60 min ride. Venues for this apt are as of Cluster 1.

- Based on current location venues, venues in cluster 1 resemble the places in my current location. I would pick APARTMENT 1 as it is more in the middle of the town and hence provides more convenience.

5.0 DISCUSSION

- In general, I am positively impressed with the overall organization, content and lab works presented during the Coursera IBM Certification Course

- I feel this Capstone project presented me a great opportunity to practice and apply the Data Science tools and methodologies learned.

- I have created a good project that I can present as an example to show my potential.

- I feel I have acquired a good starting point to become a professional Data Scientist and I will continue exploring to creating examples of practical cases

6.0 CONCLUSION

- I feel rewarded with the efforts, time and money spent. I believe this course with all the topics covered is well worthy of appreciation.

- This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools.

- The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision thoroughly and with confidence. I would recommend for use in similar situations.

- One must keep abreast of new tools for Data Science that continue to appear for application in several business fields.