Coursera Capstone Project

IBM Data Science Certification

Ramakrishna Samudrala

Contents

1. Introduction
   Discuss the business problem and the audience who would be interested in this project.
2. Data
   Describe the data that will be used to solve the problem and the source of the data.
3. Methodology
   Represents the main component of the report where you discuss and describe all the exploratory data analysis, inferential statistical testing and the machine learning tools that were used and why.
4. Results
   Discussing the results
5. Discussion
   Discuss the observations noted and the recommendations can be made based on the results
6. Conclusion
   Concluding the report

1.0 Introduction

1.1 Background of the Problem

I am helping a friend who is residing in Chennai, India currently. He is living within the walking distance to Chennai Downtown and Railway Station with good public transportation access to work. He has been offered a great opportunity to work in Manhattan, NY and I am helping him to find a similar place to live with all the amenities using my skills that I learned in the Coursera IBM Data Science.

1.2 Problem to be resolved:

Finding a rental apartment unit in Manhattan NY that offer the below requirements

- Apartment with min 2 bedrooms with monthly rent not to exceed US$7000/month
- Unit located within walking distance (<=1.0 mile, 1.6 km) from a subway metro station in Manhattan
- Area with amenities and venues similar to the ones described for current location (as described in section 2.1)

1.3 The intended Audience for this Project

I believe this is a relevant project for any person or entity considering to move to a major city in US (Manhattan, New York), Europe, or Asia, since the approach and methodologies I am presenting here will be applicable in all cases.

2.0 Data

2.1 Data requirements for Present Situation (current residence place)

- Geodata for my friend's current residence 'Valachery' in Chennai, India was established using Foursquare.
- List of Manhattan neighborhoods with clustered venues established via Foursquare as in Course Lab: https://en.wikipedia.org/wiki/List_of_Manhattan_neighborhoods#Midtown_neighborhoods
- List of Subway metro sttions in Manhattan with addresses and geod ata (Latitude and Longitude): https://en.wikipedia.org/wiki/List_of_New_York_City_Subway_stations_in_Manhattan , https://www.google.com/maps/search/manhattan+subway+metro+stations/@40.7837297,-74.1033043,11z/data=!3m1!4b1
- List of rental apartments in Manhattan area with information on neighborhood location, address, number of beds, area size, rent – complemented with geo data vai nominatim. http://www.rentmanhattan.com/index.cfm?page=search&state=results https://www.nestpick.com/search?city=new-york&page=1&order=relevance&district=manhattan&gclid=CjwKCAiAjNjgBRAgEiwAGLlf2hkP3A-cPxjZYkURqQEswQK2jKQEpv_MvKcrIhRWRzNkc_r-fGiolxoCA7cQAvD_BwE&type=apartment&display=list https://www.realtor.com/apartments/Manhattan_NY
- Place to work in Manhattan (Park Avenue and 53rd St) for reference

2.2 Data Sources and Data Manipulation

- Chennai data and its map to be created using Nominatim, Foursquare and Foliup mapping
- Manhattan neighborhoods data was obtained using Wikipedia and the Neighborhoods data was organized with geodata via Nominatim and the mapping was done using Folium
- List of Subway stations was obtained from Wikipedia, NY Transit website and Google maps
- List of rental apartments was consolidated by wen-scraping the real estate sites for Manhattan. The geolocation (lat, long) data was obtained using Nominatim
- Folium map was the basis of mapping with various features to consolidate all data in one single map where we can visualized all details needed to select an apartment.

3.0 Methodology

Mapping:

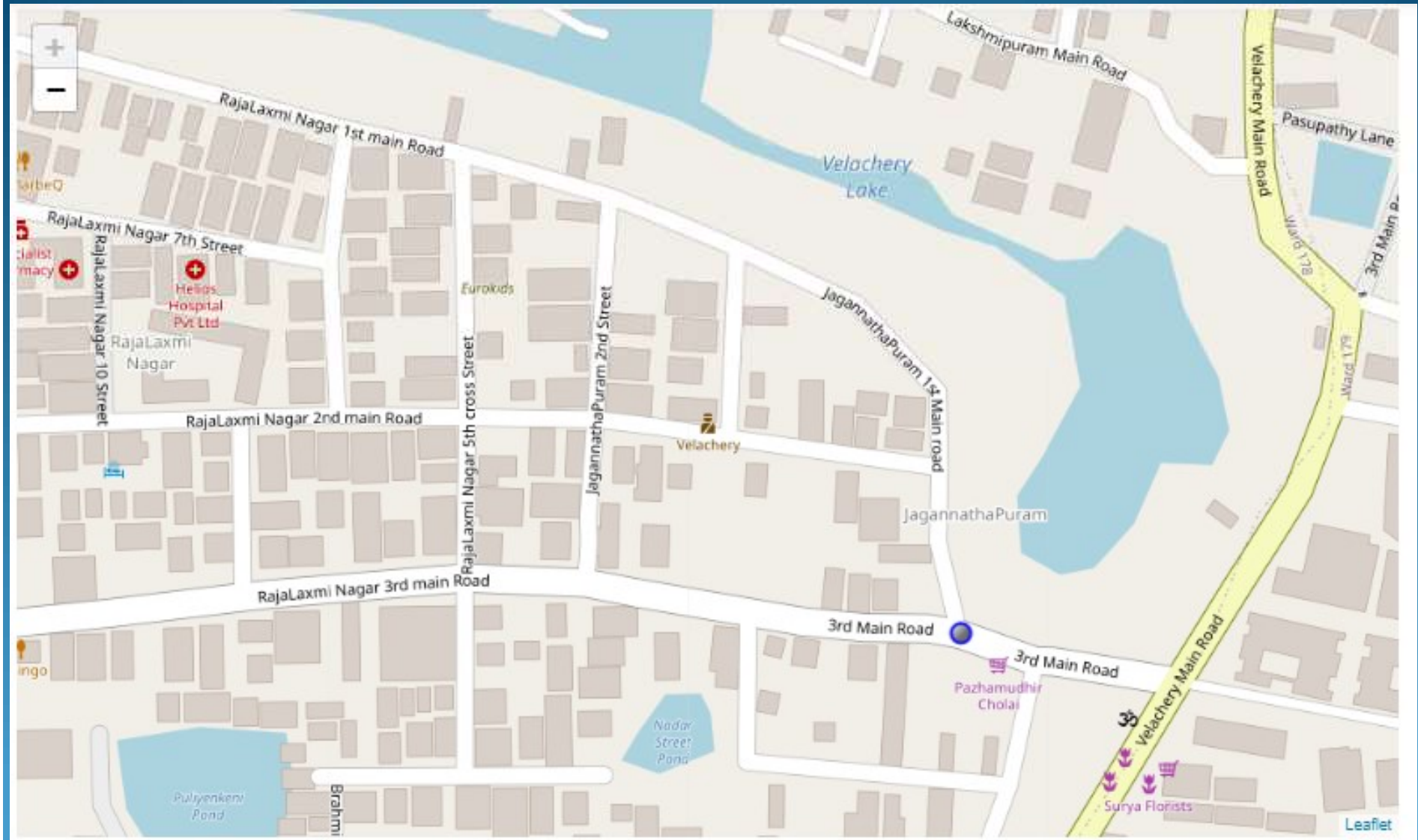
The data described in Data Section will be mapped in order to facilitate the choice of at least two candidate places for renting. Then the information will be consolidated in one single map where one can see the details of the apartment, the cluster of venues in the neighborhood and the relative location from a subway station and from work place. A measurement tool icon will also be provided. The popups on the map items will display rental price, location and cluster of venues applicable.

Tools used:

Web scraping of sites was used to consolidate data-frame information. Then the data was saved as .csv files for convenience and simplification of the report. Geodata was obtained by coding a program to use Nominatim and to get latitude and longitude of subway stations and also for each of the apartments for rent listed. Geopy_distance and Nominatim were used to establish relative distances. Seaborn graphic was used for general statistics on rental data. Maps with popups labels allow quick identification of location, price and other features, thus making the selection easier.

4.0 Results
and
5.0 Discussion

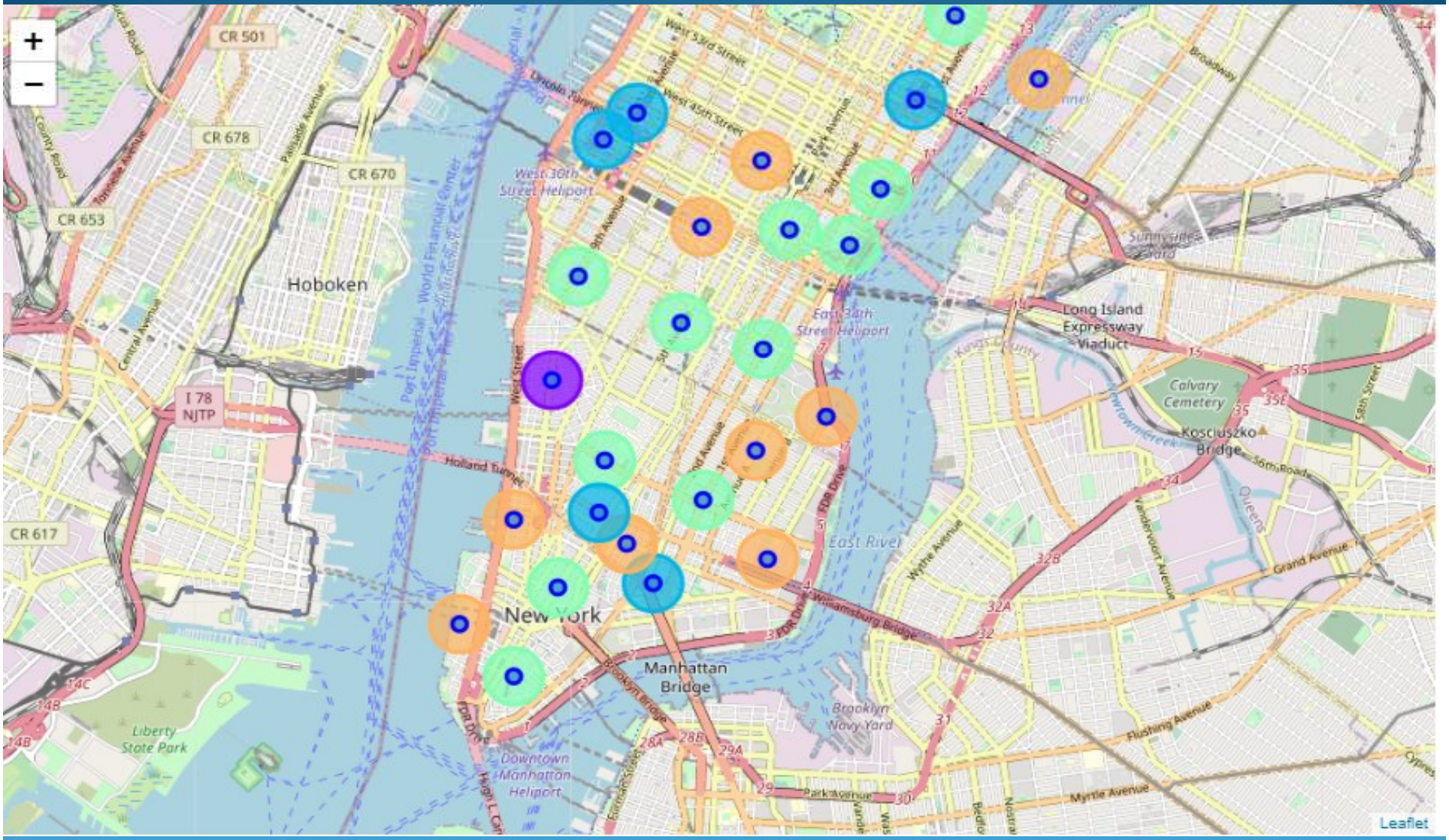
The Neighborhood map of current location - Chennai



Venues around neighborhood

	name	categories	lat	lng
0	The Westin	Hotel	12.989379	80.220885
1	Andhi Kadai	Indian Restaurant	12.982532	80.224662
2	Subway	Sandwich Place	12.987449	80.218362
3	Subway - Velachery	Sandwich Place	12.987412	80.218464
4	Peaches	Vegetarian / Vegan Restaurant	12.985401	80.218026
5	Pizza Empire	Pizza Place	12.984231	80.217984
6	Mast Kalandar	Indian Restaurant	12.984715	80.217821
7	Coal Barbecue	BBQ Joint	12.986020	80.218074
8	Pazhamudhir Cholai	Market	12.984232	80.221835
9	Dhandeeshwaram	Bus Stop	12.982931	80.222244

Manhattan Map – Neighborhoods and Cluster of Venues



Geodata for Manhattan Apts for Rent

```
# csv files with rental places with basic data but still wihtout geodata ( latitude and longitude)
# pd.read_csv(' le.csv', header=None, nrows=5)
mh_rent=pd.read_csv('MH_flats_price.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	3400	10000	NaN	NaN
1	East 97th Street	Upper East Side	3.57	3	2100	7500	NaN	NaN
2	West 105th Street	Upper West Side	1.89	4	2800	5300	NaN	NaN
3	CARMINE ST.	West Village	3.03	2	1650	5000	NaN	NaN
4	171 W 23RD ST.	Chelsea	3.45	2	1450	5000	NaN	NaN

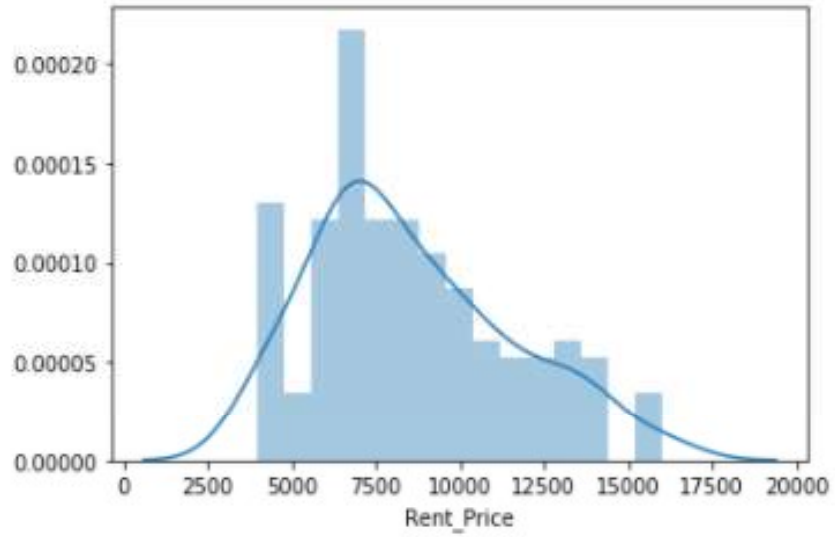
```
mh_rent.tail()
```

	Address	Area	Price_per_ft2	Rooms	Area-ft2	Rent_Price	Lat	Long
139	200 East 72nd Street	Rental in Lenox Hill	5.15	3	1700	8750	NaN	NaN
140	50 Murray Street	No fee rental in Tribeca	7.11	2	1223	8700	NaN	NaN
141	300 East 56th Street	No fee rental in Midtown East	3.87	3	2100	8118	NaN	NaN
142	1930 Broadway	No fee rental in Central Park West	5.06	2	1600	8095	NaN	NaN
143	33 West 9th Street	Rental in Greenwich Village	6.67	2	1500	10000	NaN	NaN

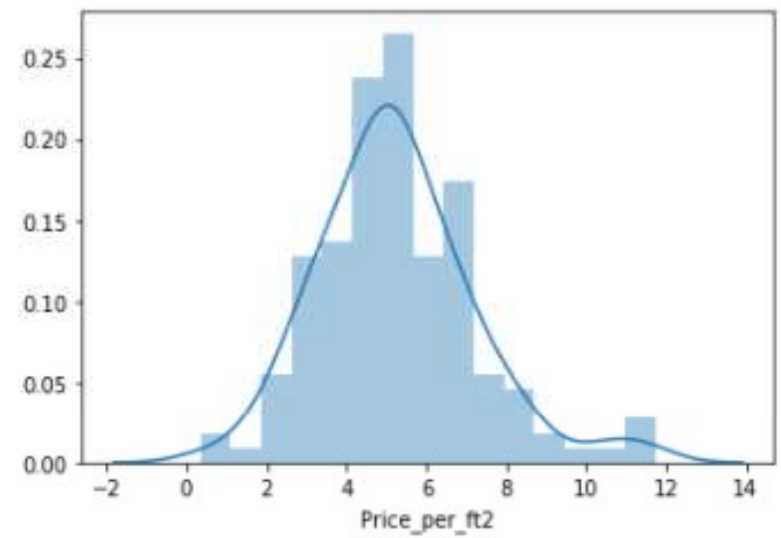
Rental Price Statistics – Manhattan Apartments

Budget $7000/month around is around the mean.

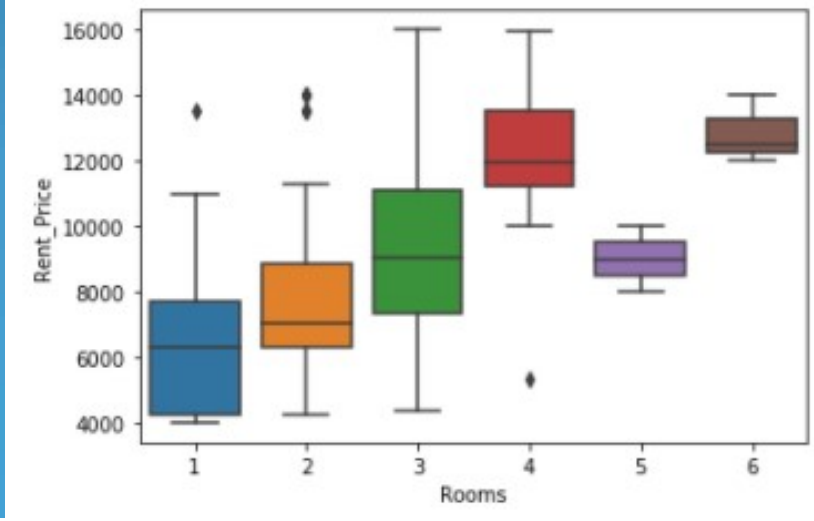
<matplotlib.axes._subplots.AxesSubplot at 0x2030f338e88>



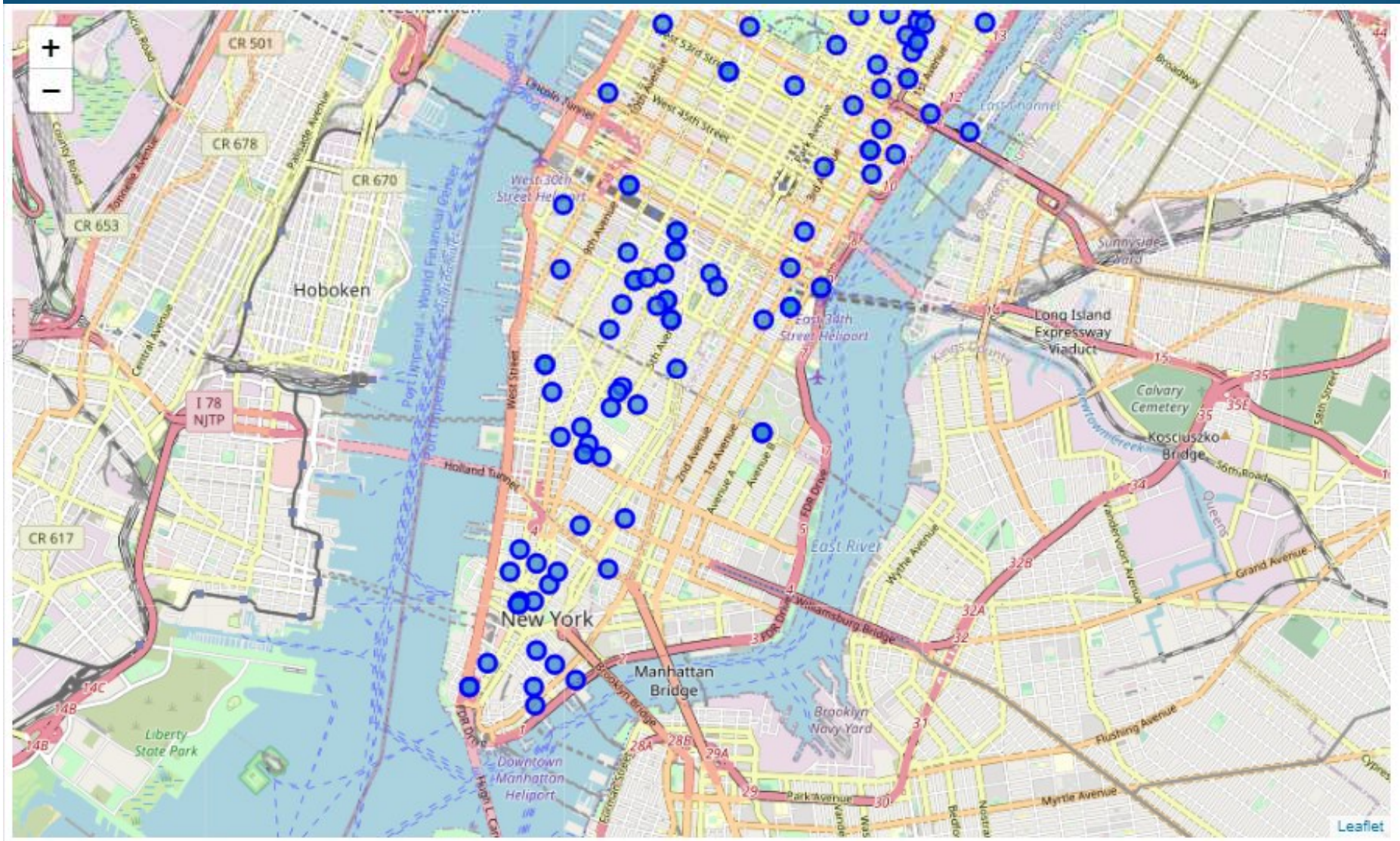
<matplotlib.axes._subplots.AxesSubplot at 0x2030fafb3c8>



<matplotlib.axes._subplots.AxesSubplot at 0x2030fb7e588>



Manhattan Map with Rental Apartments



Manhattan Rental Apts with Venue Clusteres

Venues of Cluster 3

	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
3	Inwood	Mexican Restaurant	Lounge	Pizza Place	Café	Wine Bar	Bakery	American Restaurant	Park	Frozen Yogurt Shop	Spanish Restaurant
5	Manhattanville	Deli / Bodega	Italian Restaurant	Seafood Restaurant	Mexican Restaurant	Sushi Restaurant	Beer Garden	Coffee Shop	Falafel Restaurant	Bike Trail	Other Nightlife
10	Lenox Hill	Sushi Restaurant	Italian Restaurant	Coffee Shop	Gym / Fitness Center	Pizza Place	Burger Joint	Deli / Bodega	Gym	Sporting Goods Shop	Thai Restaurant
12	Upper West Side	Italian Restaurant	Bar	Bakery	Vegetarian / Vegan Restaurant	Indian Restaurant	Coffee Shop	Cosmetics Shop	Wine Bar	Mexican Restaurant	Sushi Restaurant
16	Murray Hill	Sandwich Place	Hotel	Japanese Restaurant	Gym / Fitness Center	Coffee Shop	Salon / Barbershop	Burger Joint	French Restaurant	Bar	Italian Restaurant
17	Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bakery	Nightclub	Theater	Art Gallery	Seafood Restaurant	American Restaurant	Hotel
18	Greenwich Village	Italian Restaurant	Sushi Restaurant	French Restaurant	Clothing Store	Chinese Restaurant	Café	Indian Restaurant	Bakery	Seafood Restaurant	Electronics Store
27	Gramercy	Italian Restaurant	Restaurant	Thrift / Vintage Store	Cocktail Bar	Bagel Shop	Coffee Shop	Pizza Place	Mexican Restaurant	Grocery Store	Wine Shop
29	Financial District	Coffee Shop	Hotel	Gym	Wine Shop	Steakhouse	Bar	Italian Restaurant	Pizza Place	Park	Gym / Fitness Center
31	Noho	Italian Restaurant	French Restaurant	Cocktail Bar	Gift Shop	Bookstore	Grocery Store	Mexican Restaurant	Hotel	Sushi Restaurant	Coffee Shop
32	Civic Center	Gym / Fitness Center	Bakery	Italian Restaurant	Cocktail Bar	French Restaurant	Sandwich Place	Coffee Shop	Gym	Yoga Studio	Park
35	Turtle Bay	Italian Restaurant	Coffee Shop	Steakhouse	Wine Bar	Sushi Restaurant	Hotel	Noodle House	Indian Restaurant	Japanese Restaurant	French Restaurant
36	Tudor City	Café	Park	Pizza Place	Mexican Restaurant	Greek Restaurant	Sushi Restaurant	Hotel	Deli / Bodega	Diner	Dog Run
38	Flatiron	Italian Restaurant	American Restaurant	Gym	Gym / Fitness Center	Yoga Studio	Vegetarian / Vegan Restaurant	Bakery	Clothing Store	Cosmetics Shop	Cycle Studio

Geodata for Manhattan Subway Stations

```
mh=pd.read_csv('MH_subway.csv')
print(mh.shape)
mh.head()
```

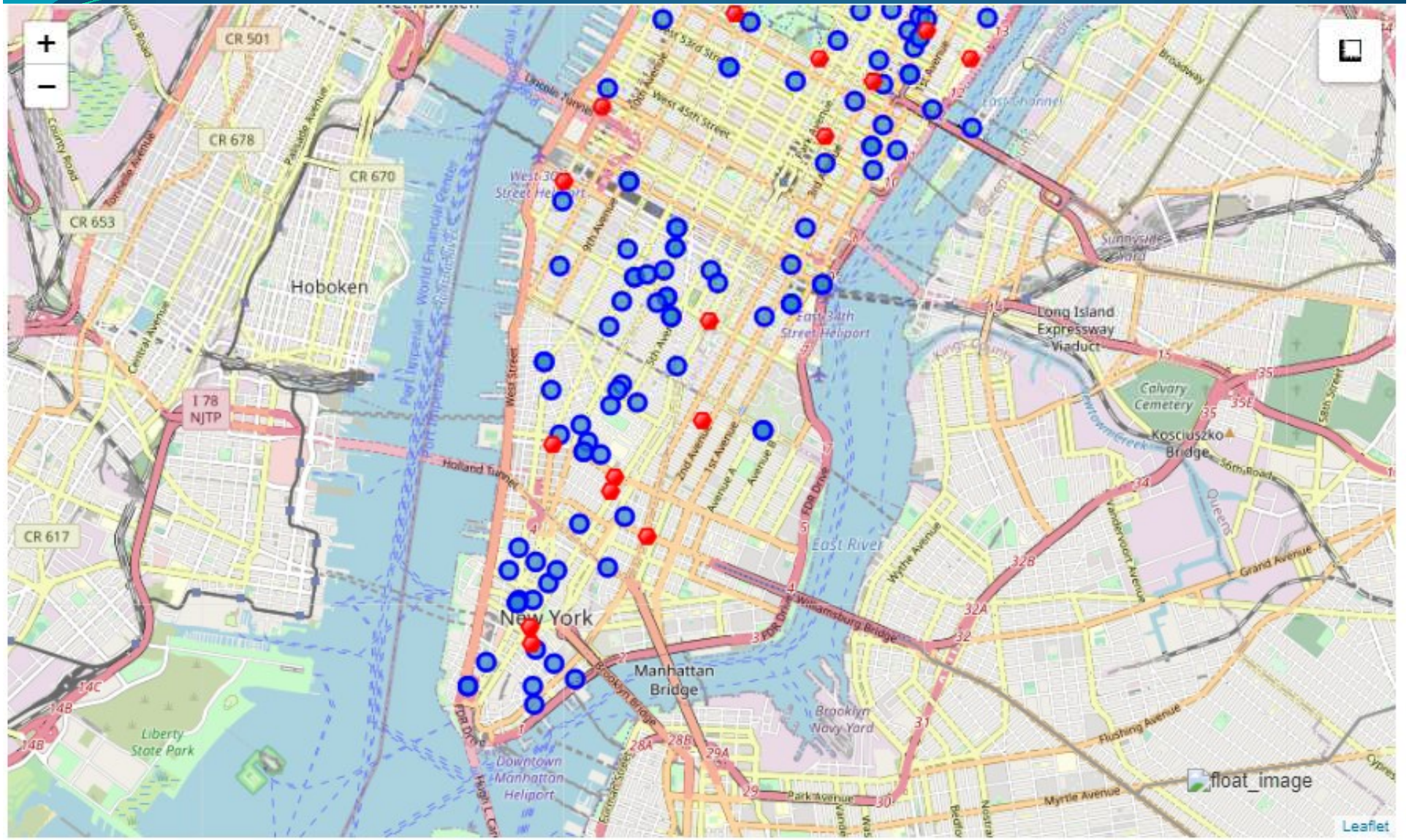
```
(76, 4)
```

	sub_station	sub_address	lat	long
0	Dyckman Street Subway Station	170 Nagle Ave, New York, NY 10034, USA	40.861857	-73.924509
1	57 Street Subway Station	New York, NY 10106, USA	40.764250	-73.954525
2	Broad St	New York, NY 10005, USA	40.730862	-73.987156
3	175 Street Station	807 W 177th St, New York, NY 10033, USA	40.847991	-73.939785
4	5 Av and 53 St	New York, NY 10022, USA	40.764250	-73.954525

```
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

Rental Apts (blue) and Subway Stations (red)



Solving the Problem - Select the apartment for rent

The above consolidate map was used to explore options.

After examining the above map, I have chosen two locations that meet all the requirements for my friend - which will be further assessed to make a choice.

- Apartment 1: 305 East 63rd Street in the Sutton Place Neighborhood and near 'subway 59th Street' station, Cluster # 2 Monthly rent : 7500 Dollars
- Apartment 2: 19 Dutch Street in the Financial District Neighborhood and near 'Fulton Street Subway' station, Cluster # 3 Monthly rent : 6935 Dollars

Venues for Apartment 1 - Cluster 2

	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	Coffee Shop	Discount Store	Yoga Studio	Steakhouse	Supplement Shop	Tennis Stadium	Shoe Store	Gym	Bank	Seafood Restaurant
1	Chinatown	Chinese Restaurant	Cocktail Bar	Dim Sum Restaurant	American Restaurant	Vietnamese Restaurant	Salon / Barbershop	Noodle House	Bakery	Bubble Tea Shop	Ice Cream Shop
6	Central Harlem	African Restaurant	Seafood Restaurant	French Restaurant	American Restaurant	Cosmetics Shop	Chinese Restaurant	Event Space	Liquor Store	Beer Bar	Gym / Fitness Center
9	Yorkville	Coffee Shop	Gym	Bar	Italian Restaurant	Sushi Restaurant	Pizza Place	Mexican Restaurant	Deli / Bodega	Japanese Restaurant	Pub
14	Clinton	Theater	Italian Restaurant	Coffee Shop	American Restaurant	Gym / Fitness Center	Hotel	Wine Shop	Spa	Gym	Indie Theater
23	Soho	Clothing Store	Boutique	Women's Store	Shoe Store	Men's Store	Furniture / Home Store	Italian Restaurant	Mediterranean Restaurant	Art Gallery	Design Studio
26	Morningside Heights	Coffee Shop	American Restaurant	Park	Bookstore	Pizza Place	Sandwich Place	Burger Joint	Café	Deli / Bodega	Tennis Court
34	Sutton Place	Gym / Fitness Center	Italian Restaurant	Furniture / Home Store	Indian Restaurant	Dessert Shop	American Restaurant	Bakery	Juice Bar	Boutique	Sushi Restaurant
39	Hudson Yards	Coffee Shop	Italian Restaurant	Hotel	Theater	American Restaurant	Café	Gym / Fitness Center	Thai Restaurant	Restaurant	Gym

Venues for Apartment 2 – Cluster 3

	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
3	Inwood	Mexican Restaurant	Lounge	Pizza Place	Café	Wine Bar	Bakery	American Restaurant	Park	Frozen Yogurt Shop	Spanish Restaurant
5	Manhattanville	Deli / Bodega	Italian Restaurant	Seafood Restaurant	Mexican Restaurant	Sushi Restaurant	Beer Garden	Coffee Shop	Falafel Restaurant	Bike Trail	Other Nightlife
10	Lenox Hill	Sushi Restaurant	Italian Restaurant	Coffee Shop	Gym / Fitness Center	Pizza Place	Burger Joint	Deli / Bodega	Gym	Sporting Goods Shop	Thai Restaurant
12	Upper West Side	Italian Restaurant	Bar	Bakery	Vegetarian / Vegan Restaurant	Indian Restaurant	Coffee Shop	Cosmetics Shop	Wine Bar	Mexican Restaurant	Sushi Restaurant
16	Murray Hill	Sandwich Place	Hotel	Japanese Restaurant	Gym / Fitness Center	Coffee Shop	Salon / Barbershop	Burger Joint	French Restaurant	Bar	Italian Restaurant
17	Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bakery	Nightclub	Theater	Art Gallery	Seafood Restaurant	American Restaurant	Hotel
18	Greenwich Village	Italian Restaurant	Sushi Restaurant	French Restaurant	Clothing Store	Chinese Restaurant	Café	Indian Restaurant	Bakery	Seafood Restaurant	Electronics Store
27	Gramercy	Italian Restaurant	Restaurant	Thrift / Vintage Store	Cocktail Bar	Bagel Shop	Coffee Shop	Pizza Place	Mexican Restaurant	Grocery Store	Wine Shop
29	Financial District	Coffee Shop	Hotel	Gym	Wine Shop	Steakhouse	Bar	Italian Restaurant	Pizza Place	Park	Gym / Fitness Center
31	Noho	Italian Restaurant	French Restaurant	Cocktail Bar	Gift Shop	Bookstore	Grocery Store	Mexican Restaurant	Hotel	Sushi Restaurant	Coffee Shop
32	Civic Center	Gym / Fitness Center	Bakery	Italian Restaurant	Cocktail Bar	French Restaurant	Sandwich Place	Coffee Shop	Gym	Yoga Studio	Park
35	Turtle Bay	Italian Restaurant	Coffee Shop	Steakhouse	Wine Bar	Sushi Restaurant	Hotel	Noodle House	Indian Restaurant	Japanese Restaurant	French Restaurant
36	Tudor City	Café	Park	Pizza Place	Mexican Restaurant	Greek Restaurant	Sushi Restaurant	Hotel	Deli / Bodega	Diner	Dog Run

5.0 Discussion

I was able to explore all the possibilities using the above map and using popups I was able to make a right decision in selecting an apartment.

Apartment 1 rent costs US7500 slightly above the US7000 budget. Apt 1 is located 400 meters away from subway station at 59th Street and it is 600 meters away from work place (Park Ave and 53rd). I can walk to work place and use subway for other places around. Venues for this apt are as of Cluster 2 and it is located in a fine district in the East side of Manhattan.

Apartment 2 rent cost is US6935, just under the US7000 budget. Apt 2 is located 60 meters away 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 3.

Based on current residence - Velachery, Chennai venues, I feel that Cluster 2 type of venues is a closer resemblance to my current place. That means that APARTMENT 1 is a better choice since the extra monthly rent is worth the conveniences (especially reduces the subway journey to work place) it provides.

6.0 Conclusion

I built a model here that is very useful to predict a possible rental place for newly coming people to Manhattan neighborhood.

This model is also useful for other people who want to set up new business ventures in any new places like Manhattan or some other countries.

And on from my personal perspective, I just wanted to speak a little bit about how this course work has been -

I am very impressed with the content, lab work and overall organization of the Coursera IBM Data Science Certification Course

Also, I feel this Capstone project helped me to find a great opportunity to practice and apply the Data Science skills I learned.