Coursera Capstone project

PARAMESWARAN PERIASWAMY

MARCH,2019

Report Content

- **Introduction Section :**
  - The "business problem" to be solved by this project and who may be interested
- **Data Section:**
  - Describe Data requirements and Sources needed to solve the problem
- **Methodology section:**
  - Main component of the report - Execute data processing, describe/discuss any
  - exploratory data analysis and/or inferential statistical testing performed, and/or
  - machine learnings used.
- **Results section:**
  - Discussion of the results and finding of answer
- **Discussion section:**
  - Discussion of observations noted and any recommendations
- **Conclusion section:**
  - Answer chosen and conclusions

1. Introduction Section

- **1.1 Scenario and Background**

I am a data scientist residing in Bangalore India. I currently live within walking distance to MG Road metro station and I enjoy many amenities and venues in the area, such as various international cuisine restaurants, cafes, food shops and entertainment. I have been offered a great opportunity to work for a leader firm in Manhattan, NY. I am very excited and I want to use this opportunity to practice my learnings in Coursera in order to answer relevant questions arisen. The key question is : How can I find a convenient and enjoyable place similar to mine now in India? Certainly, I can use available real estate apps and Google but the idea is to use and apply myself the learned tools during the course.

1. Introduction Section

▶ **1.2 Problem to be resolved:**

▶ **Business Problem:**

    ▶ The challenge is to find a suitable apartment for rent in Manhattan NY that complies with the demands on location, price and venues. Therefore, in order to set a basis for comparison, I want to find a rental unit subject to the following conditions:

    ▶ apartment must be 2 or 3 bedrooms

    ▶ desired location is near a metro station in the Manhattan area and within 1.0 mile (1.6 km) radius

    ▶ price of rent not exceed $7,000 per month

    ▶ top amenities in the selected neighbourhood shall be similar to current residence

    ▶ desirable to have venues such as coffee shops, restaurants Indian, wine stores, gym and food shops

    ▶ as a reference, I have included a map of venues near current residence in India.

1. Introduction Section

- **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

▶ **2.1 Data Requirements**

- Geodata for current residence in Bangalore with venues established using Foursquare.

- List of Manhattan (MH) neighbourhoods with clustered venues established via https://en.wikipedia.org/wiki/List_of_Manhattan_neighborhoods#Midtown_neighborhoods

- List of subway metro stations in Manhattan with addresses and geo data (lat,long): 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 apartments for rent in Manhattan area with information on neighbourhood location, address, number of beds, area size, monthly rent price and complemented with geo data via Nominatim. http:/www.rentmanhattan.com/index.cfm?page=search&state=results https://www.nestpick.com/search?city=new-

- Place to work in Manhattan (Park Avenue and 53rd St) for reference

2 Data Section

▶ **2.2 Data Sources, Data Processing and Tools used**

- Bangalore data and map is to be created with use of Nominatim , Foursquare and Folium mapping

- Manhattan neighbourhoods were obtained from Wikipedia and organized by Neighbourhoods with geodatavia Nominatim for mapping with Folium.

- List of Subway stations was obtained via Wikipedia, NY Transit web site and Google map,

- List of apartments for rent was consolidated from web-scraping real estate sites for MH.

- The geolocation (lat,long) data was found with algorithm coding and using Nominatim.

- Folium map was the basis of mapping with various features to consolidate all data in ONE map where one can visualize all details needed to make a selection of apartment

3 Methodology

▶ The Strategy to find the answer:

The strategy is based on mapping the described data in section 2.0, in order to facilitate the choice of at least two candidate places for rent. The information will be consolidated in ONE MAP where one can see the details of the apartment, the cluster of venues in the neighbourhood 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 rent price, location and cluster of venues applicable.

3 Methodology

▶ The Tools:

Web-scraping of sites is used to consolidate data-frame information which was saved as csv files for convenience and to simply the report. Geodata was obtained by coding a program to use Nominatim to get latitude and longitude of subway stations and also for each of (144 units) 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 feature, thus making the selection very easy

4 Execution and Results

Current residence Neighbourhood in Bangalore
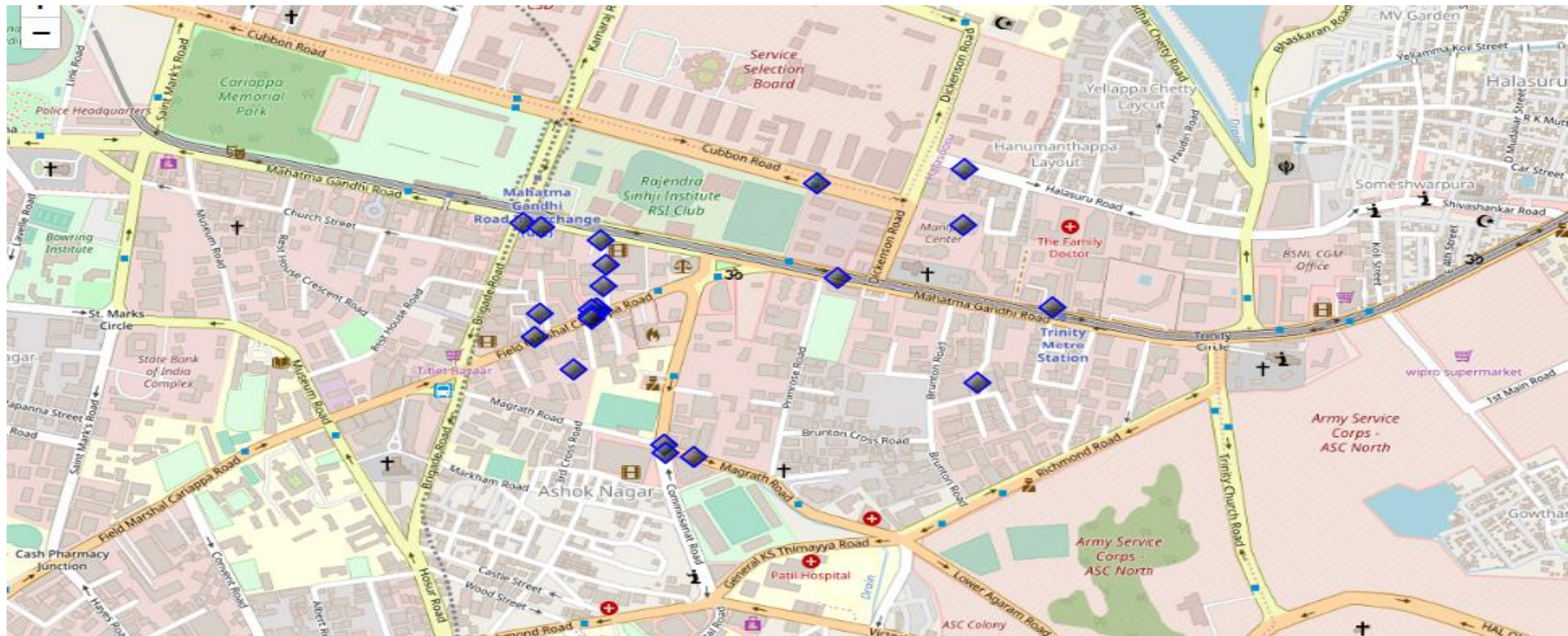
Venues around Neighbourhood

	name	categories	lat	lng
0	Nagarjuna	Andhra Restaurant	12.973145	77.609263
1	Chung Wah	Chinese Restaurant	12.974099	77.609386
2	Arbor Brewing Company	Brewery	12.970140	77.610857
3	Art of Delight	Ice Cream Shop	12.973198	77.609254
4	Corner House	Dessert Shop	12.973650	77.609362
5	Tandoor	Indian Restaurant	12.973819	77.613247
6	Karavalli	Indian Restaurant	12.971937	77.608869
7	The Permit Room	Pub	12.970388	77.610358
8	Green Onion Chinese Restaurant	Chinese Restaurant	12.973165	77.609159
9	Meghanas Biriyani	Indian Restaurant	12.973045	77.609163

Manhattan Map - Neighborhoods and Cluster of Venues
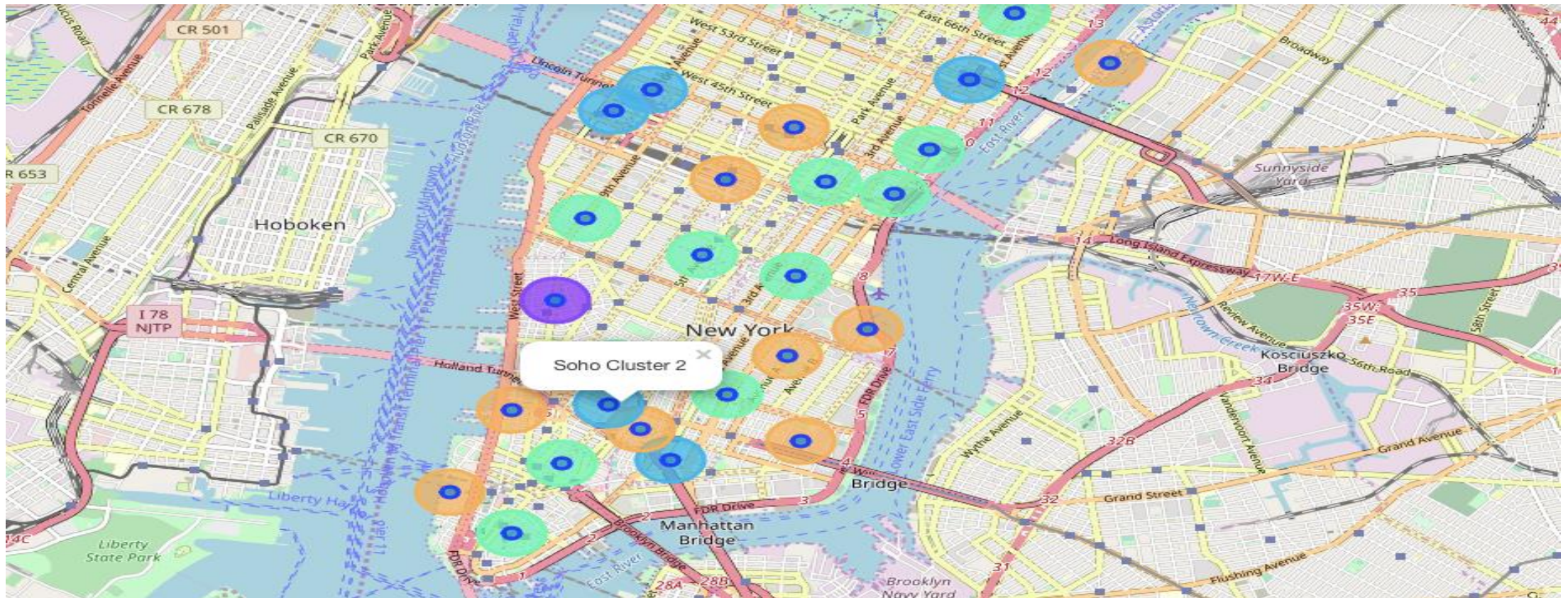
Geodata Manhattan Apts for rent

```
]: 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
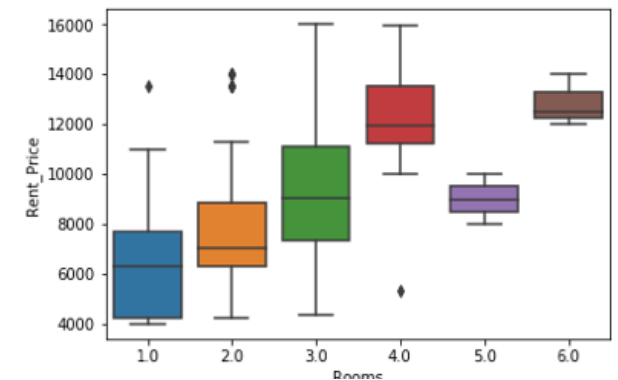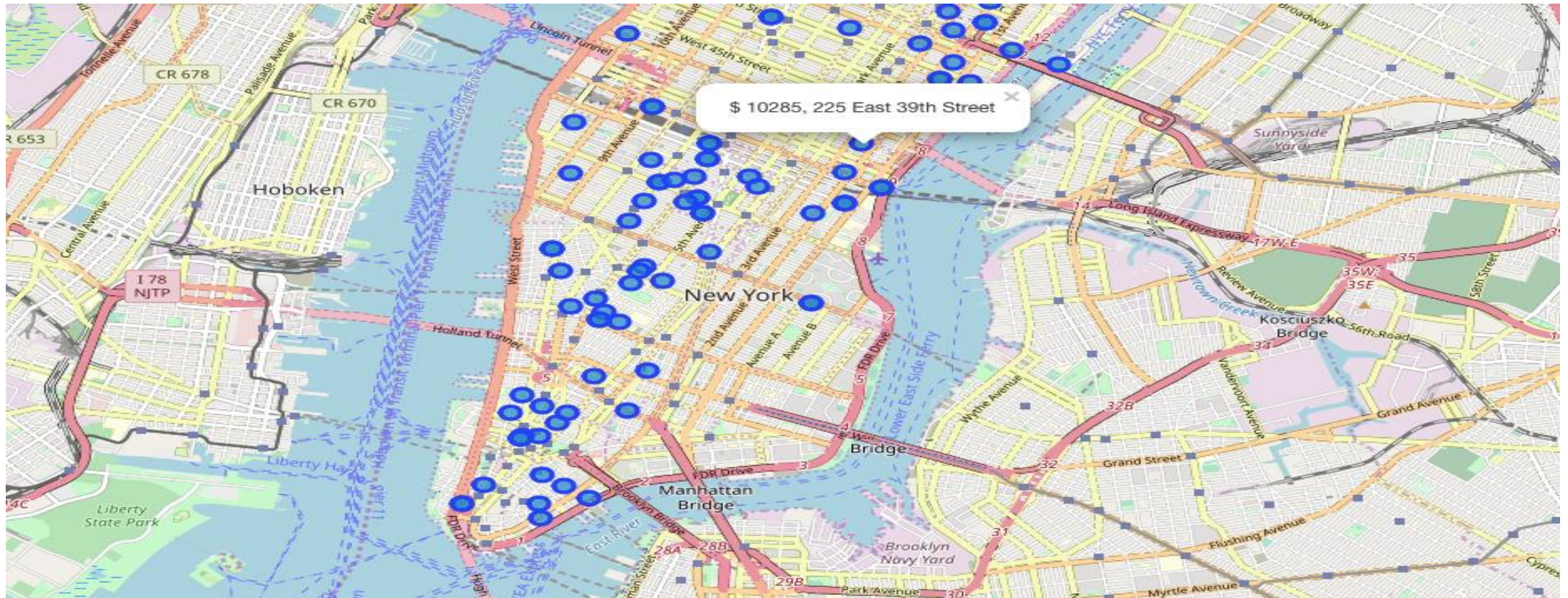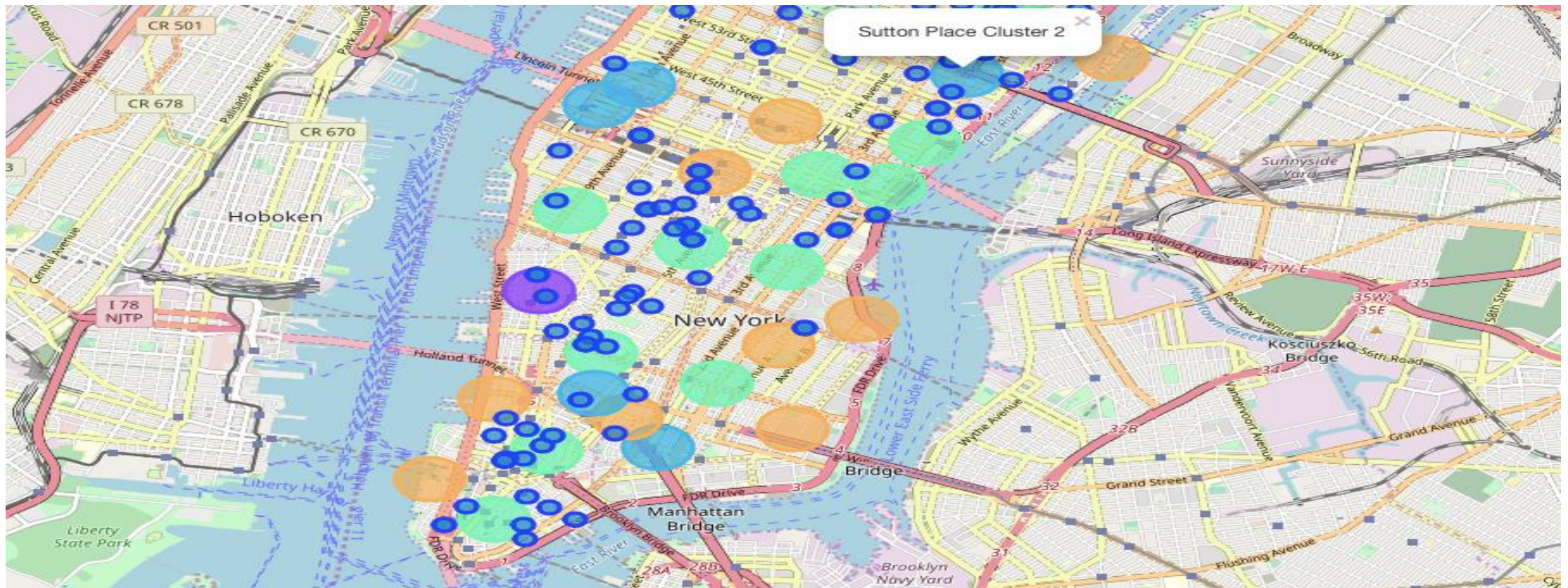
▶ Budget US7000/month is around the mean
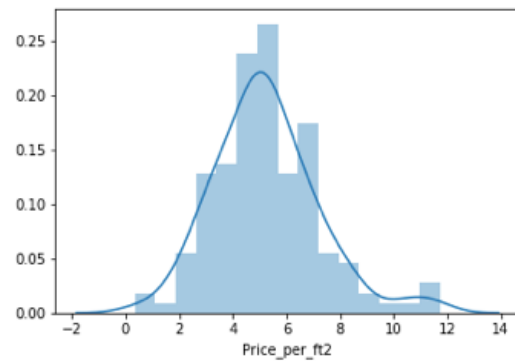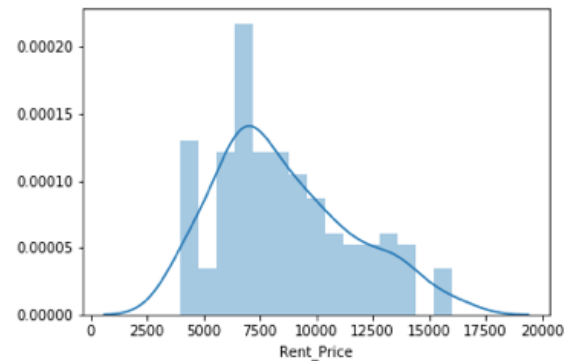
Apartments for Rent in MH

MH Apts for rent with venue clusters

Venues of cluster 2

```
## kk is the cluster number to explore
kk = 2
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == kk, 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
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

Manhattan subway stations geodata

|		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

```
# removing duplicate rows and creating new set mhsub1
mhsub1=mh.drop_duplicates(subset=['lat','long'], keep="last").reset_index(drop=True)
mhsub1.shape
```

```
(22, 4)
```

```
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

Apts for rent (blue) and subway stations (red)
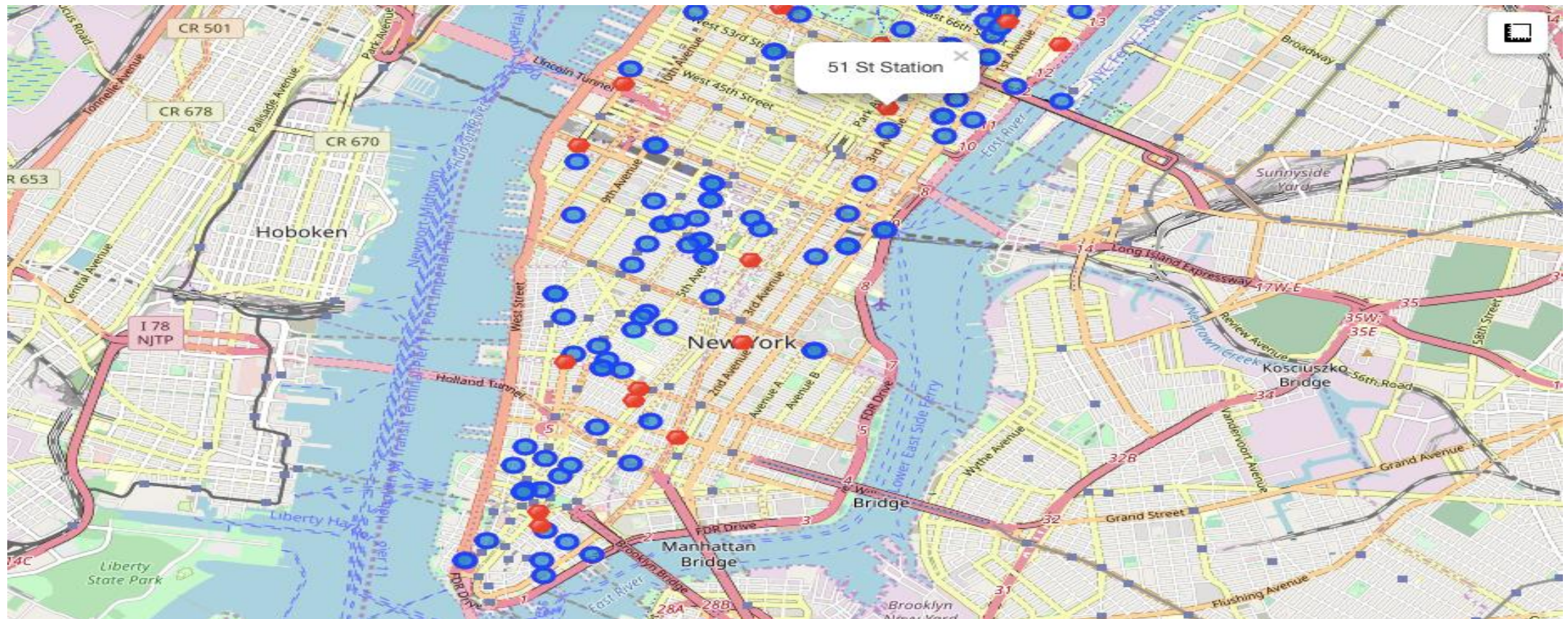
Selected Apartment!
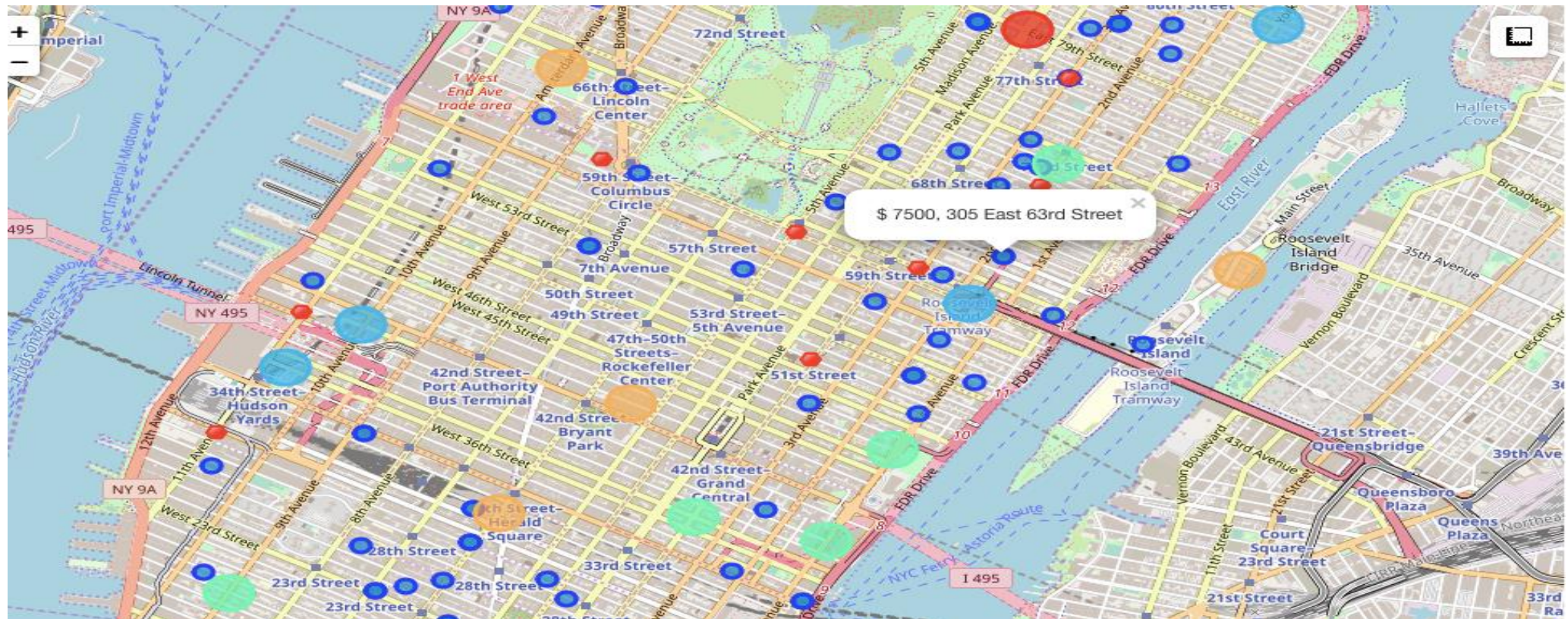
5 DISCUSSION

▶ Apartment Selection

  ▶ Apartment 1 rent cost is US7500 slightly above the US7000 budget. Apt 1 is located 400 meters from subway station at 59th Street and work place (Park Ave and 53rd) is another 600 meters way. 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 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.

6 CONCLUSIONS

Based on current Bangalore venues, I feel that Cluster 2 type of venues is a closer resemblance to my current place. And is within walkable distance.

That means that APARTMENT 1 is a better choice since the extra monthly rent is worth the conveniences it provides.