

Coursera Capstone proj

Coursera IBM Data Science Certification

MAR-EL OUARDI

1. Introduction Section :

- The “business problem” to be solved by this project and who may benefit.

2. Data Section:

- Describe Data requirements and Sources needed to solve the problem.

3. Methodology section:

- Main component of the report - Execute data processing, describe exploratory data analysis and/or inferential statistical testing performed and machine learnings used.

4. Results section:

- Discussion of the results and finding of answer

5. Discussion section:

- Discussion of observations noted and any recommendations

6. Conclusion section:

- Answer chosen and conclusions.

1.1 Scenario and Background

I am currently living in Singapore, within walking distance to Downtown MRT metro station". I also enjoy great venues and attractions, such as cuisine, entertainment and shopping. I have an offer to move to work to and I would like to move if I can find a place to live similar with similar v

1.2 Problem to be resolved:

How to find an apartment in Manhattan with the following conditions:

- Apartment with min 2 bedrooms
- Monthly rent not to exceed US\$7000/month
- Located within walking distance (≤ 1.0 mile, 1.6 km) from a subway Manhattan
- Venues and amenities as in my current residence.

1.3 Interested Audience

I believe the methodology, tools and strategy used in this project is relevant for any individual or entity considering moving to a major city in US, Europe or Asia. Furthermore, it can be a helpful approach to explore the opening of a new business or service in a specific location by utilizing the analysis of Foursquare data and mapping techniques combined with data analysis.

2.1 Data Requirements

- Geodata for current residence in Singapore with venues established using Foursquare
 - List of Manhattan (MH) neighborhoods with clustered venues established via Foursquare Lab). 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>)
 - List of apartments for rent in Manhattan area with information on neighborhood location, number of beds, area size, monthly rent price and complemented with geo data via Nestpick.com www.rentmanhattan.com/index.cfm?page=search&state=results <https://www.nestpick.com/city=new-york-city>
 - Place to work in Manhattan (Park Avenue and 53rd St) for reference

2.2 Data Sources, Data Processing and Tools used

- Singapore data and map is to be created with use of Nominatim , Foursquare and Folium.
 - Manhattan neighborhoods were obtained from Wikipedia and organized by Neighborhood via Nominatim for mapping with Folium.
 - List of Subway stations was obtained via Wikipedia, NY Transit web site and Google.
 - List of apartments for rent was consolidated from web-scraping real estate sites for Manhattan.
 - (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 place.

The Strategy to find the answer:

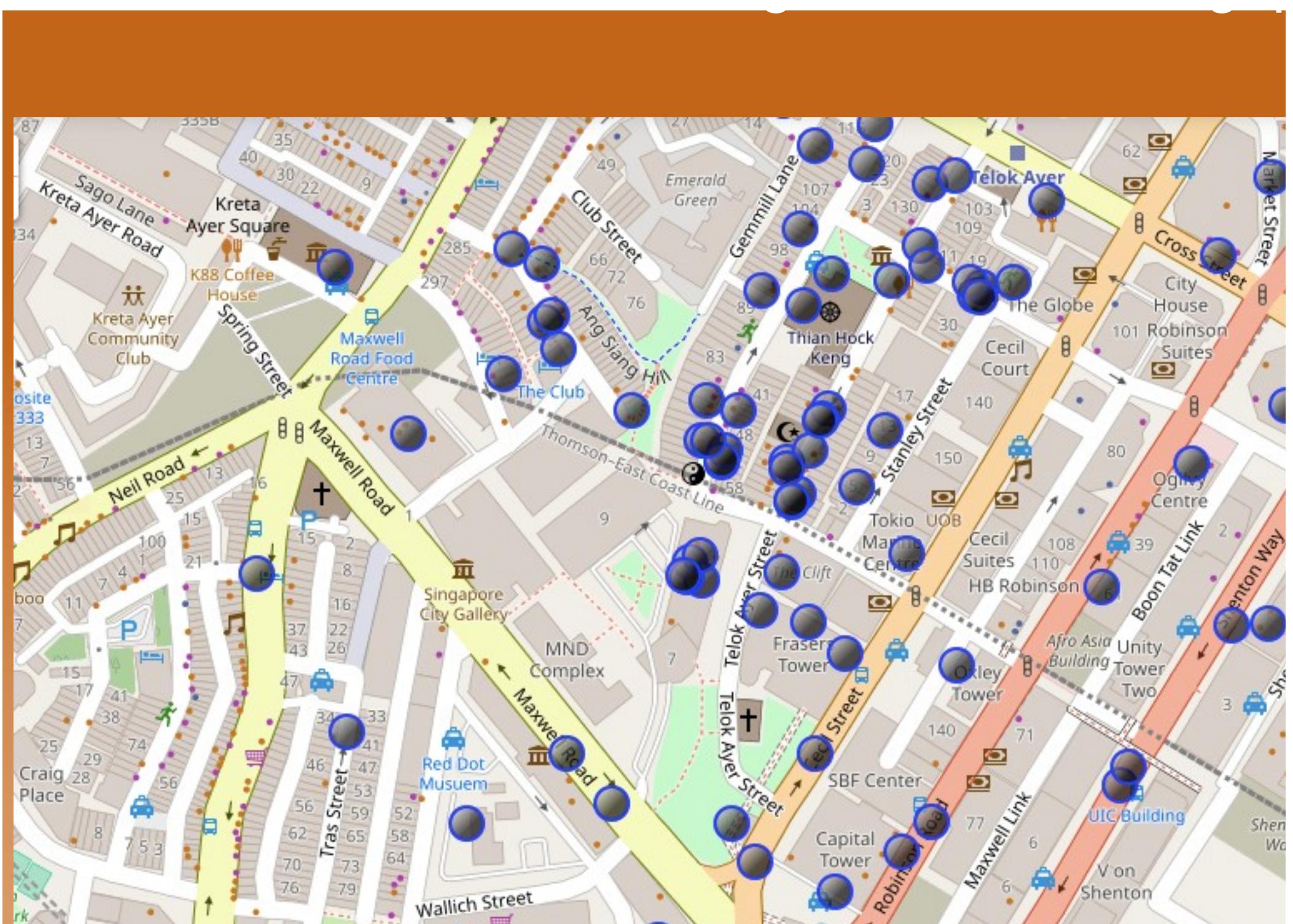
The strategy is based on mapping the described data in section 2.0, in facilitate the choice of at least two candidate places for rent. The inform consolidated in ONE MAP where one can see the details of the apartm of venues in the neighborhood and the relative location from a subway from work place. A measurement tool icon will also be provided. The p map items will display rent price, location and cluster of venues applic

The Tools:

Web-scraping of sites is used to consolidate data-frame information wh saved as csv files for convenience and to simply the report. Geodata w by coding a program to use Nominatim to get latitude and longitude of stations and also for each of (144 units) the apartments for rent listed. Geopy_distance and Nominatim were used to establish relative distance graphic was used for general statistics on rental data.

Maps with popups labels allow quick identification of location, price and making the selection very easy.

4.0 Execution and Results

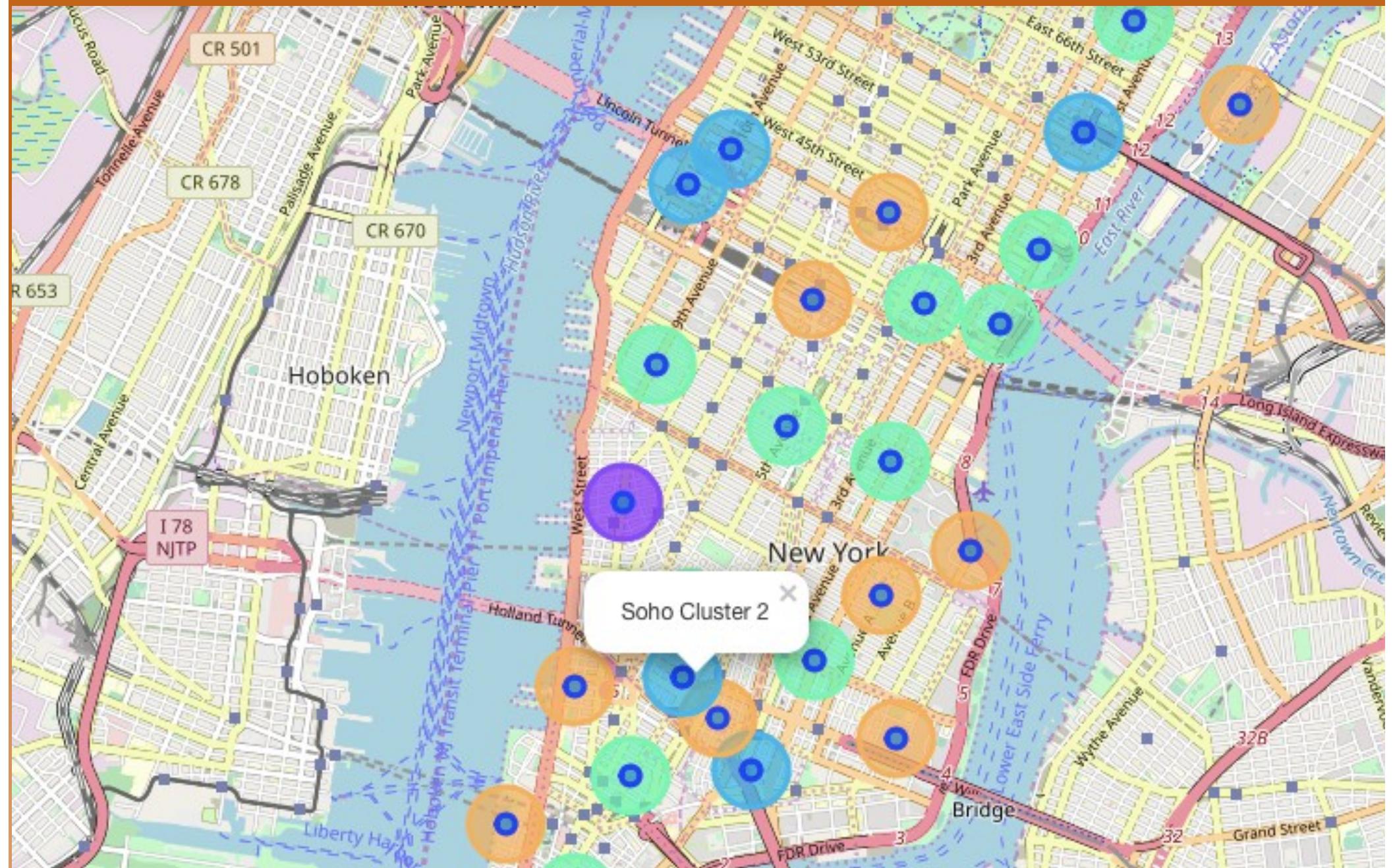


venues around neighbourhood

```
: # Venues near current Singapore residence place  
SGnearby_venues.head(10)
```

```
:
```

	name	categories	lat	lng
0	Napoleon Food & Wine Bar	Wine Bar	1.279925	103.847333
1	Park Bench Deli	Deli / Bodega	1.279872	103.847287
2	Native	Cocktail Bar	1.280135	103.846844
3	Muchachos	Burrito Place	1.279175	103.847082
4	Matt's The Chocolate Shop	Dessert Shop	1.280462	103.846950
5	Freehouse	Beer Garden	1.281254	103.848513
6	PS.Cafe	Café	1.280468	103.846264
7	왕대박 Wang Dae Bak Korean BBQ Restaurant	Korean Restaurant	1.281345	103.847551
8	Ancient Therapy	Massage Studio	1.280413	103.847481
9	Oven & Fried Chicken	Korean Restaurant	1.280479	103.847522



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 |


```

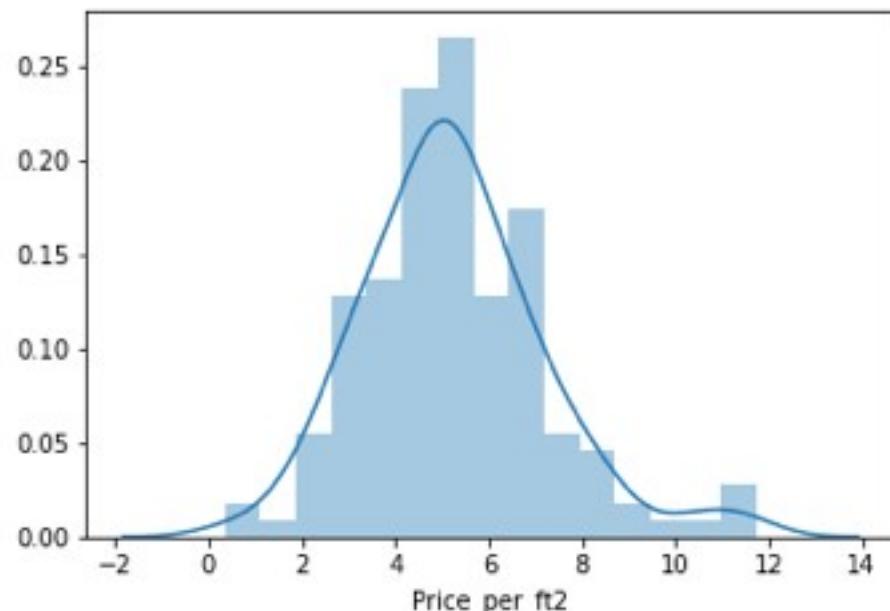
```
: mh_rent.tail()
```

```
:  

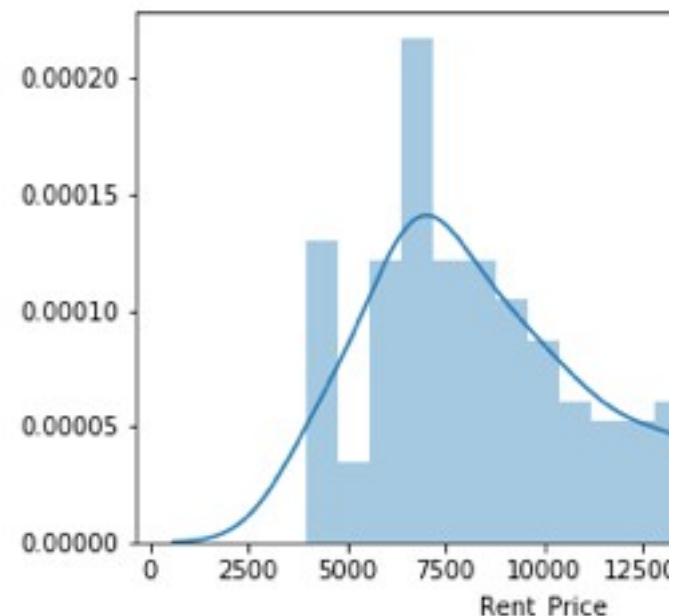

|     | Address              | Area                               | Price_per_ft2 | Rooms | Area-ft2 | Rent_Price |       |
|-----|----------------------|------------------------------------|---------------|-------|----------|------------|-------|
| 139 | 200 East 72nd Street | Rental in Lenox Hill               | 5.15          | 3.0   | 1700     | 8750       | 40.76 |
| 140 | 50 Murray Street     | No fee rental in Tribeca           | 7.11          | 2.0   | 1223     | 8700       | 40.71 |
| 141 | 300 East 56th Street | No fee rental in Midtown East      | 3.87          | 3.0   | 2100     | 8118       | 40.75 |
| 142 | 1930 Broadway        | No fee rental in Central Park West | 5.06          | 2.0   | 1600     | 8095       | 40.77 |


```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1a2415fc18>
```

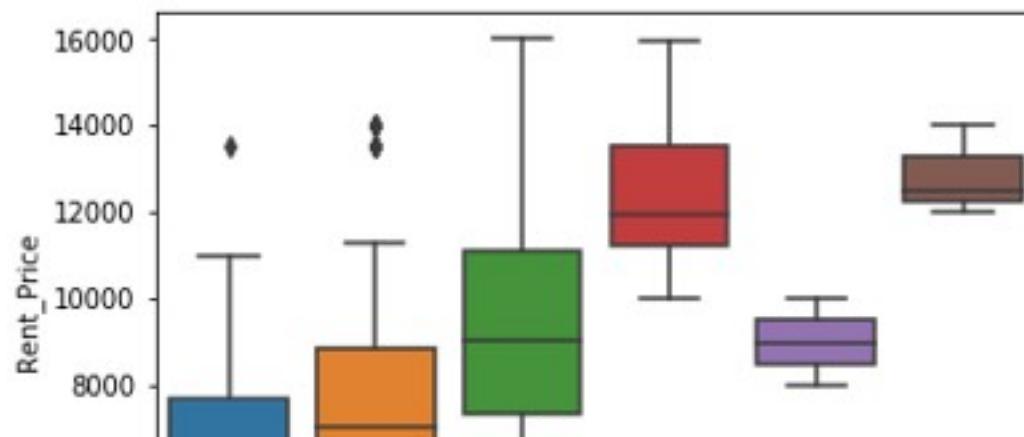


```
<matplotlib.axes._subplots.AxesSub
```



```
sns.boxplot(x='Rooms', y= 'Rent_Price', data=mh_rent)
```

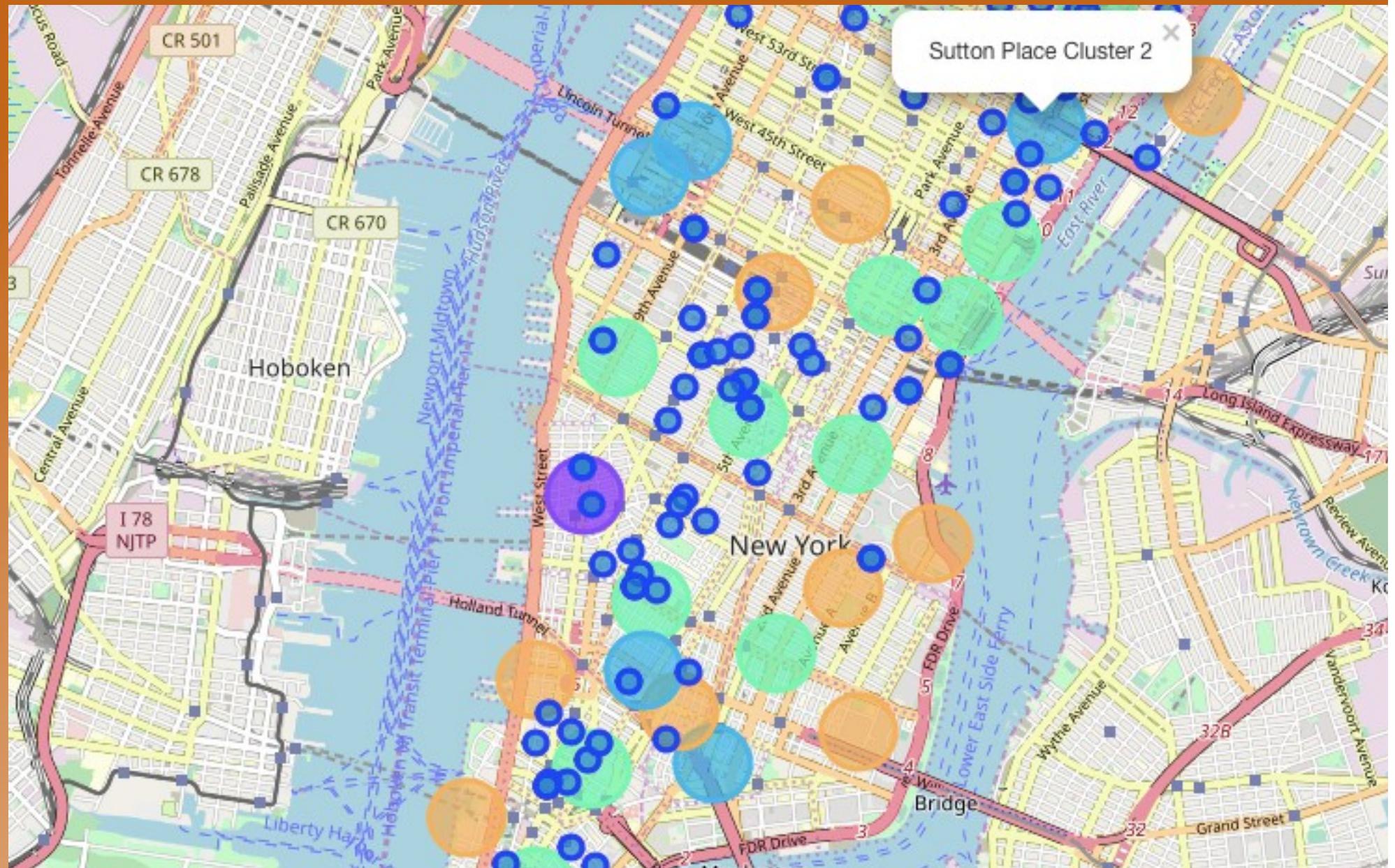
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a25f2a2b0>
```



Apartment for Rent in Mombasa



MH apts for rent with venue clus



Venn Diagram of Clusters

```
# k is the cluster number to explore
```

```
kk = 3
```

```
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == kk, manhattan_merged.columns[[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 N Comr Ve
3 Inwood	Mexican Restaurant	Lounge	Pizza Place	Café	Wine Bar	Bakery	American Restaurant	Italian Restaurant
5 Manhattanville	Deli / Bodega	Italian Restaurant	Seafood Restaurant	Mexican Restaurant	Sushi Restaurant	Beer Garden	Coffee Shop	Fast Food Restaurant
10 Lenox Hill	Sushi Restaurant	Italian Restaurant	Coffee Shop	Gym / Fitness Center	Pizza Place	Burger Joint	Deli / Bodega	Chinese Restaurant
12 Upper West Side	Italian Restaurant	Bar	Bakery	Vegetarian / Vegan Restaurant	Indian Restaurant	Coffee Shop	Cosmetics Shop	Wine Bar
16 Murray Hill	Sandwich Place	Hotel	Japanese Restaurant	Gym / Fitness Center	Coffee Shop	Salon / Barbershop	Burger Joint	French Restaurant
17 Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bakery	Nightclub	Theater	Art Gallery	Seafood Restaurant
18 Greenwich Village	Italian Restaurant	Sushi Restaurant	French Restaurant	Clothing Store	Chinese Restaurant	Café	Indian Restaurant	Baileys Bar
27 Gramercy	Italian Restaurant	Restaurant	Thrift / Vintage Store	Cocktail Bar	Bagel Shop	Coffee Shop	Pizza Place	Mexican Restaurant

Manhattan Subway Stations

click to scroll output; double click to hide

			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	

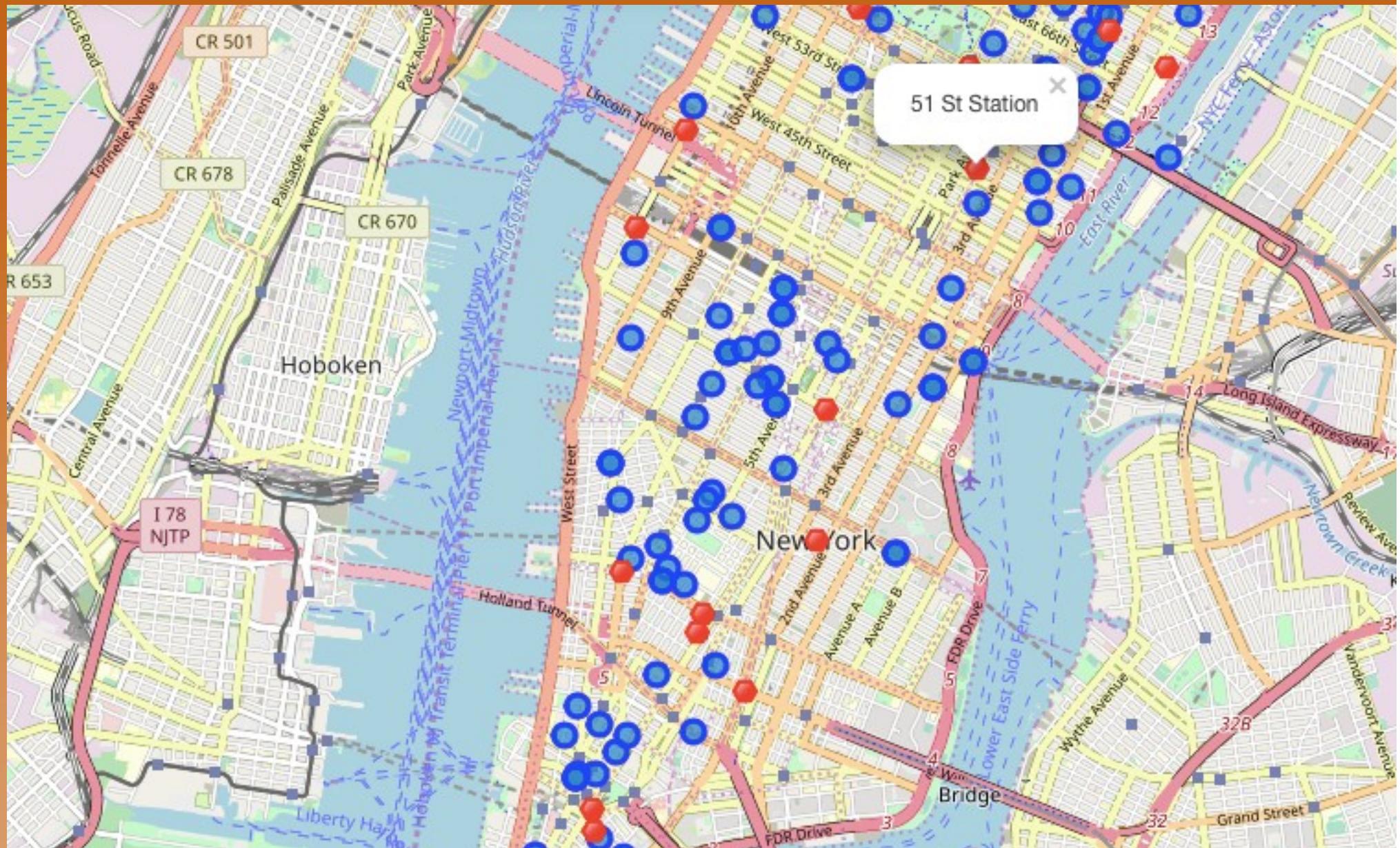
```
# removing duplicate rows and creating new set mhsub1
mhsub1=mh.drop_duplicates(subset=['lat','long'], keep="last").reset_index()
mhsub1.shape
```

```
(22, 4)
```

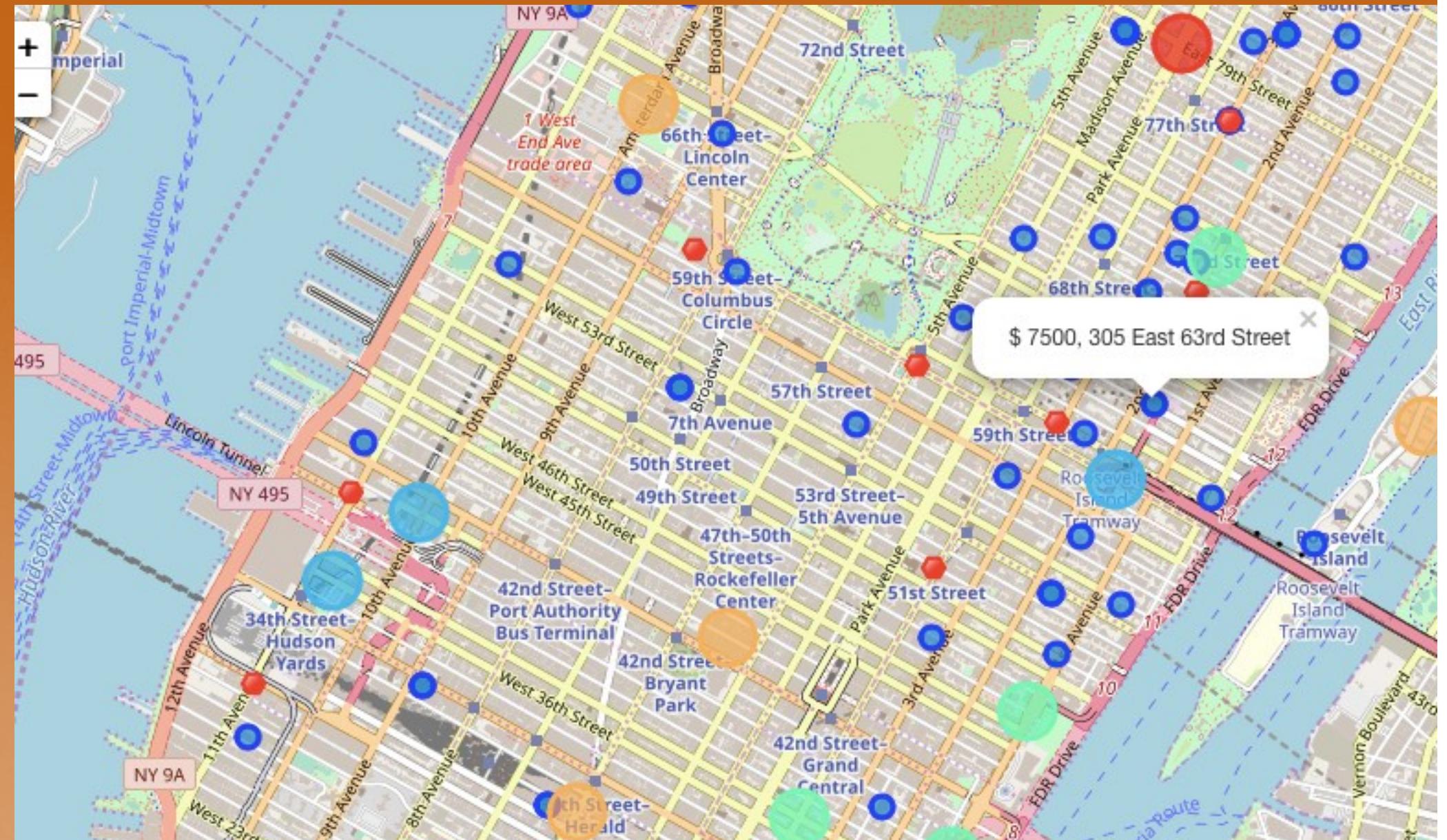
```
: mhsub1.tail()
```

	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	145th Street Station	145th Street, New York, NY 10032, USA	40.764250	-73.954525

Map for Port (Ports) and Gateway Strategies



The ONE consolidated map shows all information for decision making.
Apartments address, price, neighborhood, cluster of venues and subway stations.
Blue dots=apts , Red dots=Subway station, Bubbles=Cluster of venues.



Apartment Selection

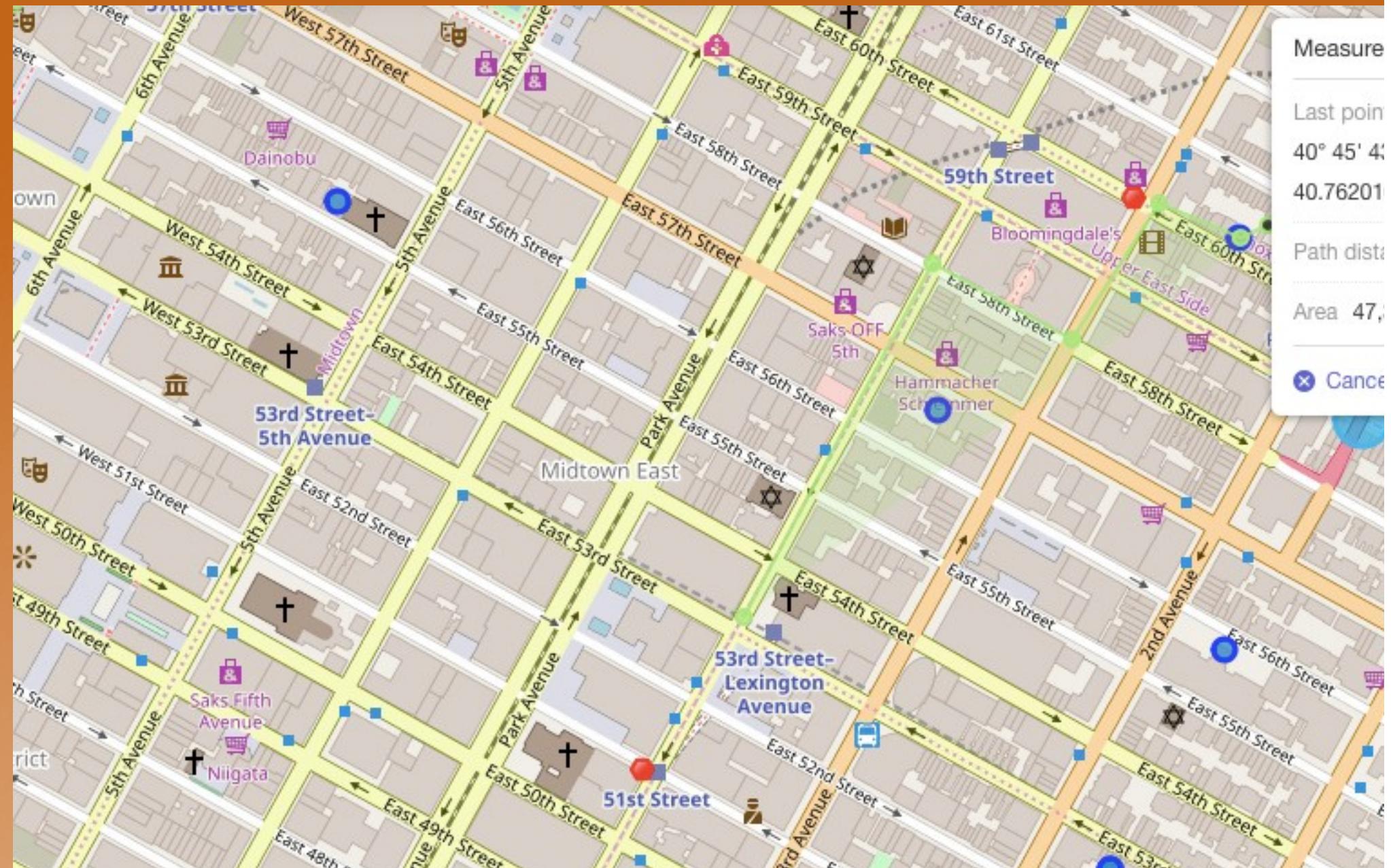
Using the "one map" above, I was able to explore all possibilities since it provides the information needed for a good decision.

Apartment 1 rent cost is US7500 slightly above the US7000 budget. Apartment is located 400 meters from subway station at 59th Street and work place (Park Avenue) is another 600 meters away. I can walk to work place and use subway for the rest of the way around. Venues for this apt are as of Cluster 2 and it is located in a first floor unit on the East side of Manhattan.

Apartment 2 rent cost is US6935, just under the US7000 budget. Apartment is located 400 meters from subway station at Fulton Street, but I will have to ride the subway to get to work , possibly 40-60 min ride. Venues for this apt are as of Cluster 2.

Based on current Singapore venues, I feel that Cluster 2 type of venue has more resemblance to my current place. That means that APARTMENT 1 is better choice.

Walk from home to work is less than 1 k



Venues in Cluster 2 Near Future

```
# kk is the cluster number to explore
kk = 2
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == kk, manhattan_merged.columns[[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 M Comm Ver
0 Marble Hill	Coffee Shop	Discount Store	Yoga Studio	Steakhouse	Supplement Shop	Tennis Stadium	Shoe Store	G
1 Chinatown	Chinese Restaurant	Cocktail Bar	Dim Sum Restaurant	American Restaurant	Vietnamese Restaurant	Salon / Barbershop	Noodle House	Bak
6 Central Harlem	African Restaurant	Seafood Restaurant	French Restaurant	American Restaurant	Cosmetics Shop	Chinese Restaurant	Event Space	Liquor St
9 Yorkville	Coffee Shop	Gym	Bar	Italian Restaurant	Sushi Restaurant	Pizza Place	Mexican Restaurant	Deli / Bode
14 Clinton	Theater	Italian Restaurant	Coffee Shop	American Restaurant	Gym / Fitness Center	Hotel	Wine Shop	\$
23 Soho	Clothing Store	Boutique	Women's Store	Shoe Store	Men's Store	Furniture / Home Store	Italian Restaurant	Mediterranean Restaur
26 Morningside Heights	Coffee Shop	American Restaurant	Park	Bookstore	Pizza Place	Sandwich Place	Burger Joint	C
34 Sutton Place	Gym / Fitness Center	Italian Restaurant	Furniture / Home Store	Indian Restaurant	Dessert Shop	American Restaurant	Bakery	Juice

- In general, I am positively impressed with the organization, content and lab works presented in the Coursera IBM Certification Course
- I feel this Capstone project presented me 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 create examples of predictive models.

0.0 CONCLUSIONS

- I feel rewarded with the efforts, time and money spent on this course with all the topics covered in it. I am very grateful and full of appreciation.
- This project has shown me a practical application of Data Science in a real situation that has impacting personal and professional life. It has shown how Data Science can have a positive impact using Data Science tools.
- The mapping with Folium is a very powerful technique to consolidate information and make the analysis a lot more thorough and with confidence. I would recommend its use in similar situations.