Coursera Capstone pro

Coursera IBM Data Science Certificati

Report Content

1. Introduction Section :

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

Data Section:

- Describe Data requirements and Sources needed to solve the pro

Methodology section:

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

1.0 Introduction

1.1 Scenario and Background

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

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 subw
  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 r
or entity considering moving to a major city in US, Europe or Asia. E
Likewise, it can be helpful approach to explore the opening of a new
of FourSquare data and mapping techniques combined with data an
resolve the key questions arisen. Lastly, this project is a good practic
developing Data Science skills.

2.0 Data Section

2.1 Data Requirements

- Geodata for current residence in Singapore with venues established using Foursq
- List of Manhattan (MH) neighborhoods with clustered venues established via Four
Lab). https://en.wikipedia.org/wiki/List_of_Manhattan_neighborhoods#Midtown_nei
List of subway metro stations in Manhattan with addresses and geo data (lat,long):
en.wikipedia.org/wiki/List_of_New_York_City_Subway_stations_in_Manhattan) , (h
maps/search/manhattan+subway+metro+stations/@40.7837297,-74.1033043,11z/
apartments for rent in Manhattan area with information on neighborhood location, a
beds, area size, monthly rent price and complemented with geo data via Nominatim

www.rentmanhattan.com/index.cfm?page=search&state=results https://www.nestpi
city=new-
- 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
  Manhattan neighborhoods were obtained from Wikipedia and organized by Neighl
  via Nominatim for mapping with Folium.
- List of Subway stations was obtained via Wikipedia, NY Transit web site and Goo
- List of apartments for rent was consolidated from web-scraping real estate sites fo
  (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
  can visualize all details needed to make a selection of apartment

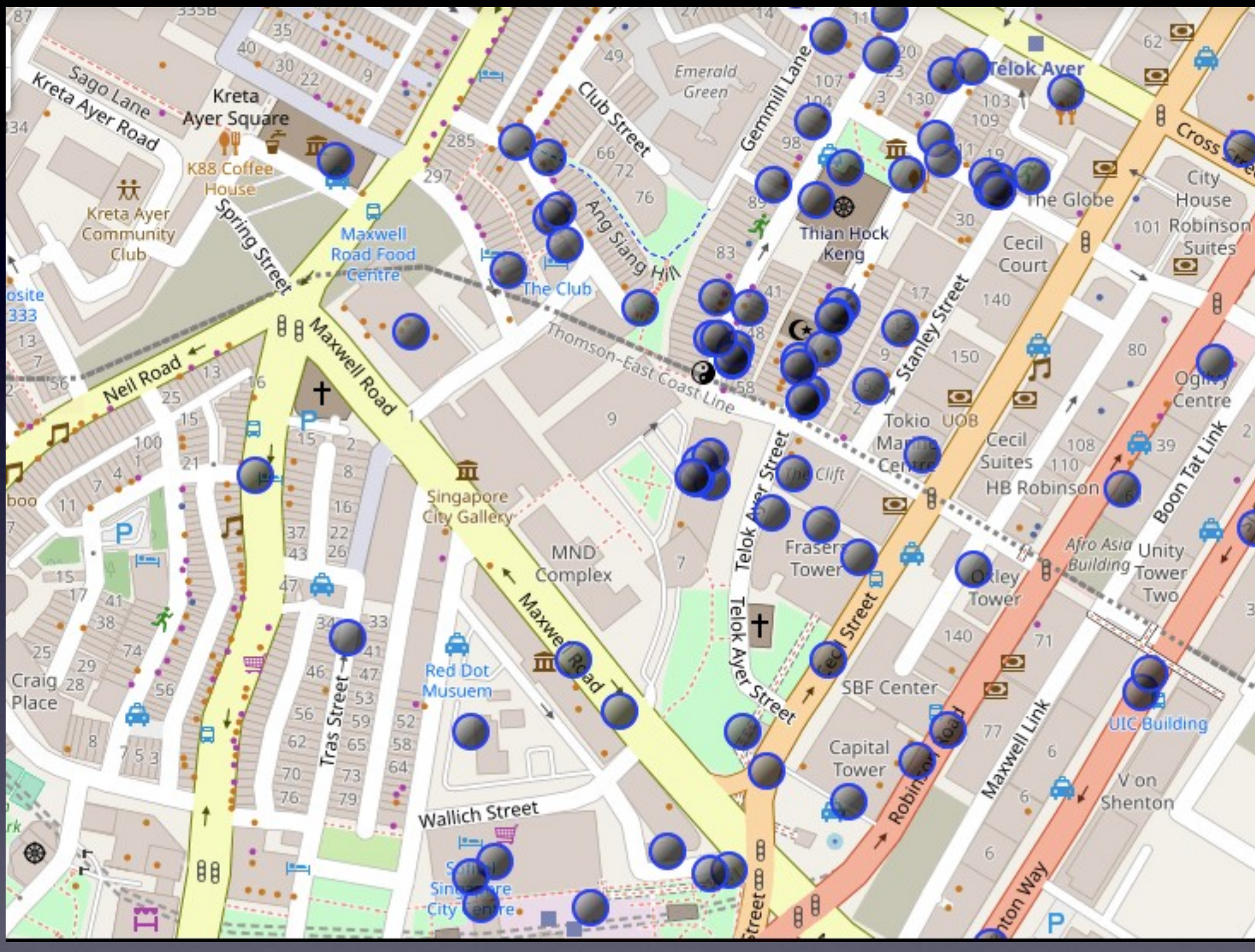
3.0 Methodology

The Strategy to find the answer:
The strategy is based on mapping the described data in section 2.0
facilitate the choice of at least two candidate places for rent. The inf
consolidated in ONE MAP where one can see the details of the apa
of venues in the neighborhood and the relative location from a subw
work place. A measurement tool icon will also be provided. The pop
items will display rent price, location and cluster of venues applicab

The Tools:
Web-scraping of sites is used to consolidate data-frame information
as csv files for convenience and to simply the report. Geodata was
a program to use Nominatim to get latitude and longitude of subway
for each of (144 units) the apartments for rent listed.
Geopy_distance and Nominatim were used to establish relative dist
graphic was used for general statistics on rental data.
Maps with popups labels allow quick identification of location, price
making the selection very easy

4.0 Execution and Results

Current residence Neighborhood in Sing

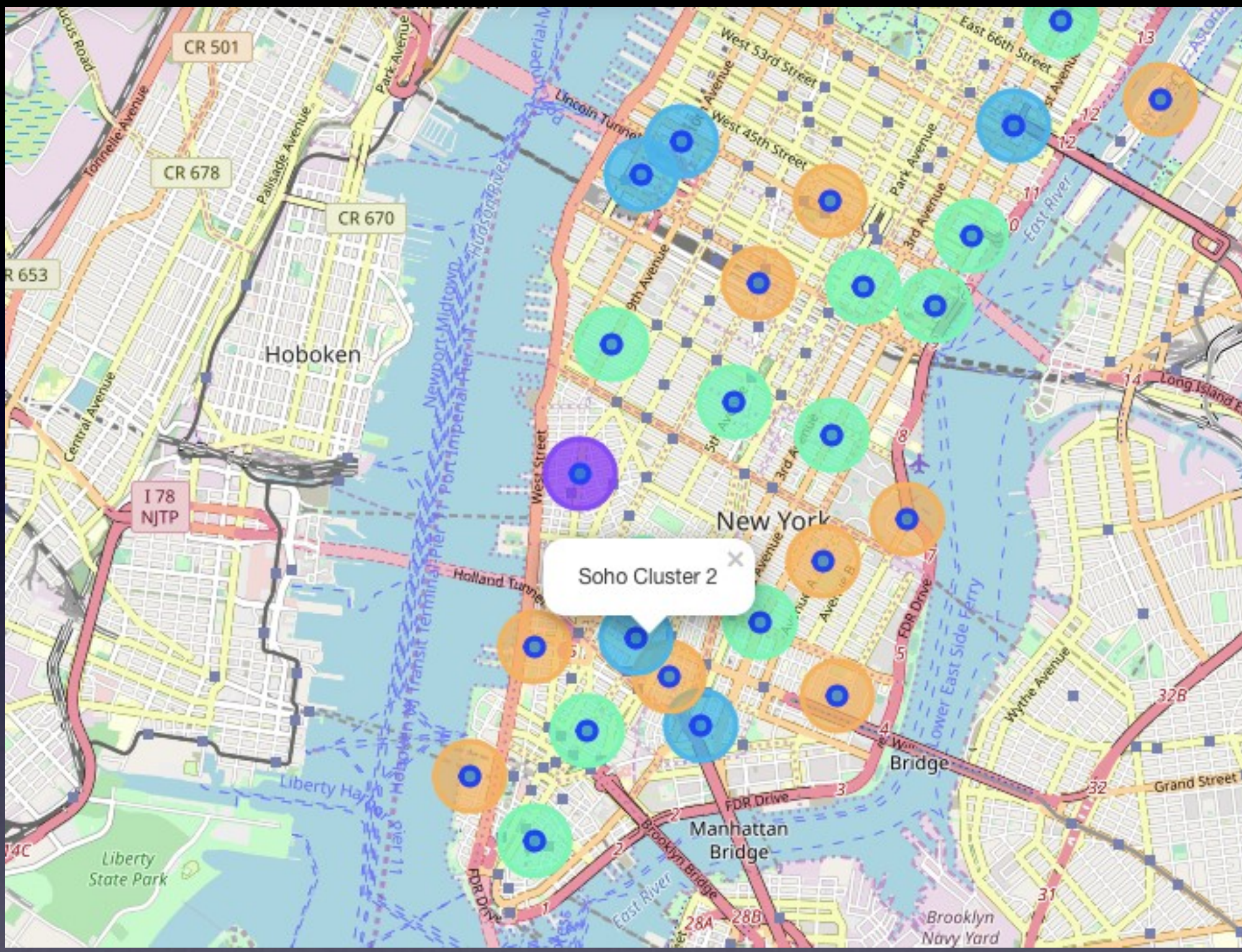


Venues around Neighborhoo

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

	name	categories	lat	
0	Napoleon Food & Wine Bar	Wine Bar	1.279925	103.847
1	Park Bench Deli	Deli / Bodega	1.279872	103.847
2	Native	Cocktail Bar	1.280135	103.846
3	Muchachos	Burrito Place	1.279175	103.847
4	Matt's | The Chocolate Shop	Dessert Shop	1.280462	103.846
5	Freehouse	Beer Garden	1.281254	103.848
6	PS.Cafe	Café	1.280468	103.846
7	왕대박 Wang Dae Bak Korean BBQ Restaurant	Korean Restaurant	1.281345	103.847
8	Ancient Therapy	Massage Studio	1.280413	103.847
9	Oven & Fried Chicken	Korean Restaurant	1.280479	103.847

Manhattan Map - Neighborhoods and Cluste


CR 501
CR 678
CR 670
Hoboken
Soho Cluster 2
New York
Manhattan Bridge
Liberty State Park
Brooklyn Navy Yard
West 53rd Street
West 45th Street
East 66th Street
Park Avenue
3rd Avenue
9th Avenue
West Street
FDR Drive
East River
Brooklyn Bridge
Grand Street
Wythe Avenue
Central Avenue
Palisade Avenue
Tonnelle Avenue
Lincoln Tunnel
Holland Tunnel
Bridge
I 78
NJTP

GeoData Manhattan apts for

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

]:

	Address	Area	Price_per_ft2	Rooms	Area-ft2	Rent_Price	Lat	Lo
0	West 105th Street	Upper West Side	2.94	5.0	3400	10000	40.799771	-73.9662
1	East 97th Street	Upper East Side	3.57	3.0	2100	7500	40.788585	-73.9552
2	West 105th Street	Upper West Side	1.89	4.0	2800	5300	40.799771	-73.9662
3	CARMINE ST.	West Village	3.03	2.0	1650	5000	40.730523	-74.0018
4	171 W 23RD ST.	Chelsea	3.45	2.0	1450	5000	40.744118	-73.9952

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

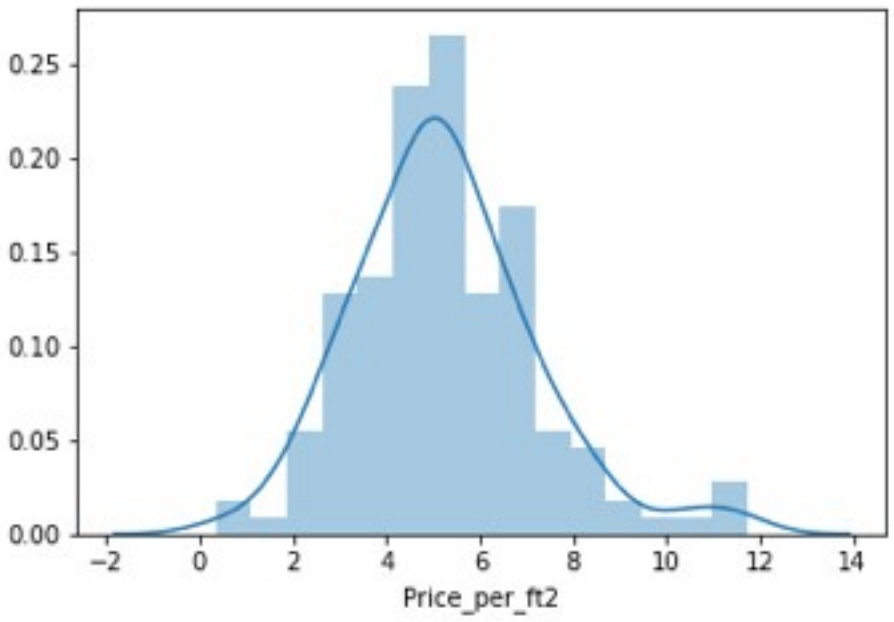
]:

	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
140	50 Murray Street	No fee rental in Tribeca	7.11	2.0	1223	8700
141	300 East 56th Street	No fee rental in Midtown East	3.87	3.0	2100	8118
142	1930 Broadway	No fee rental in Central Park West	5.06	2.0	1600	8095
143	33 West 9th Street	Rental in Greenwich Village	6.67	2.0	1500	10000

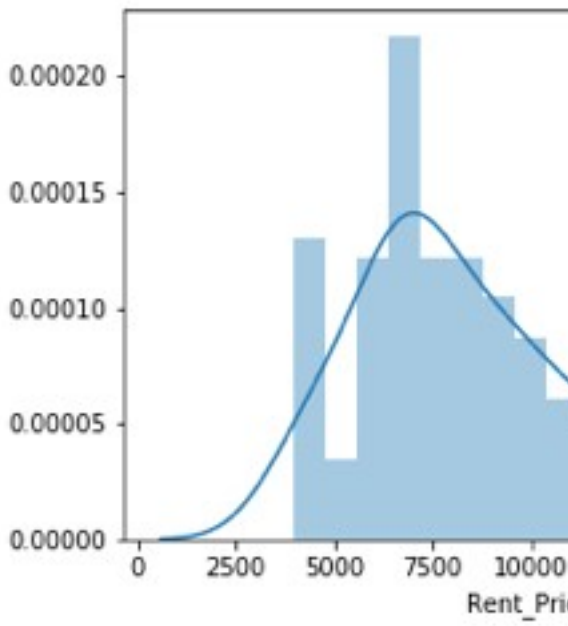
Rental Price Statistics MH Apartme

Budget US7000/month is around the mean

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

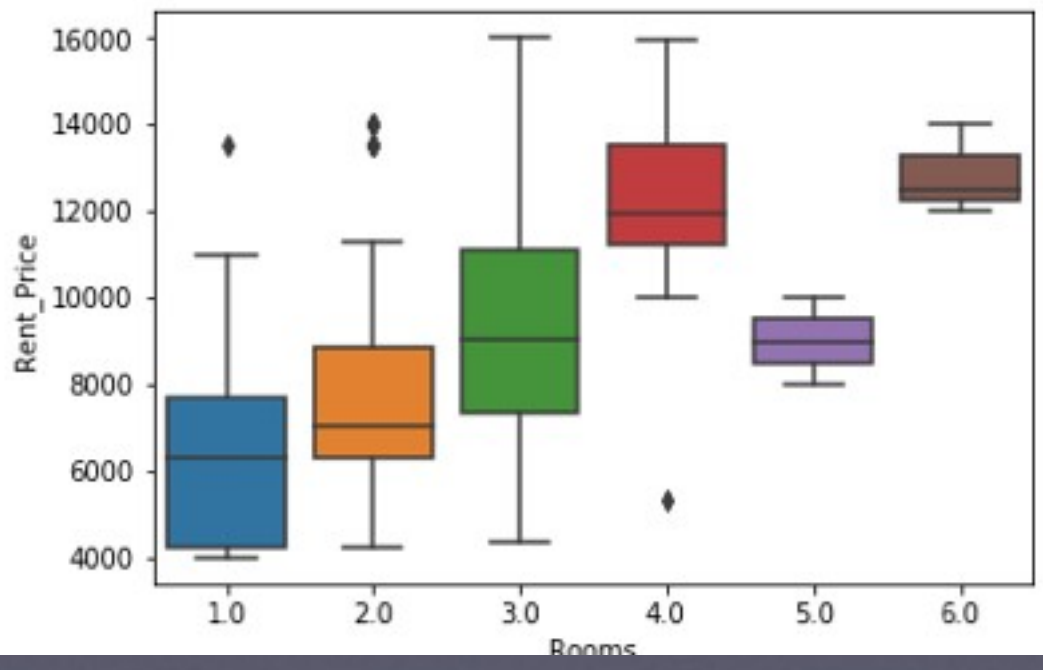



```
<matplotlib.axes._subplots.Axe
```

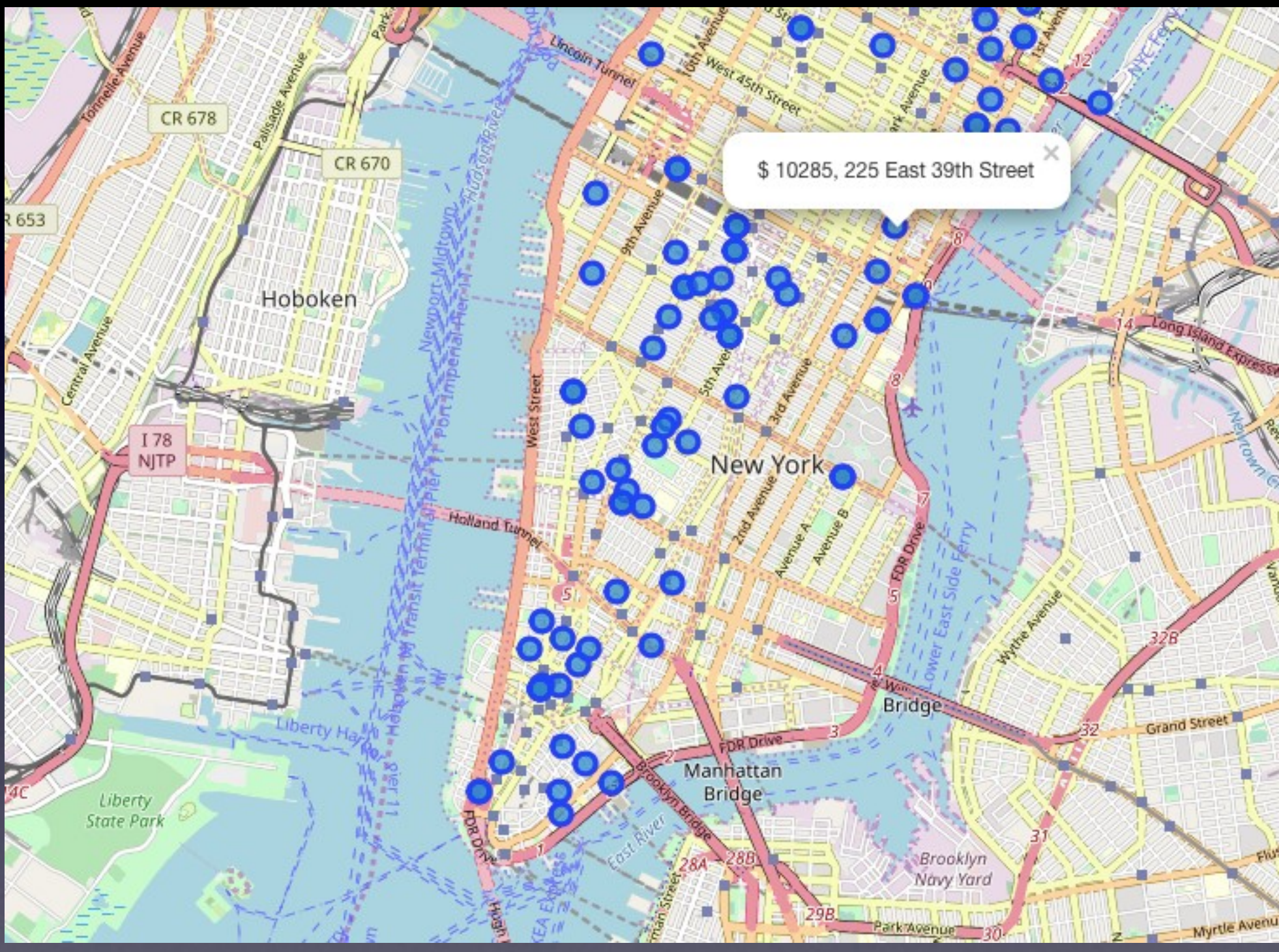



```
sns.boxplot(x='Rooms', y= 'Rent_Price', data=mh_rent)
```

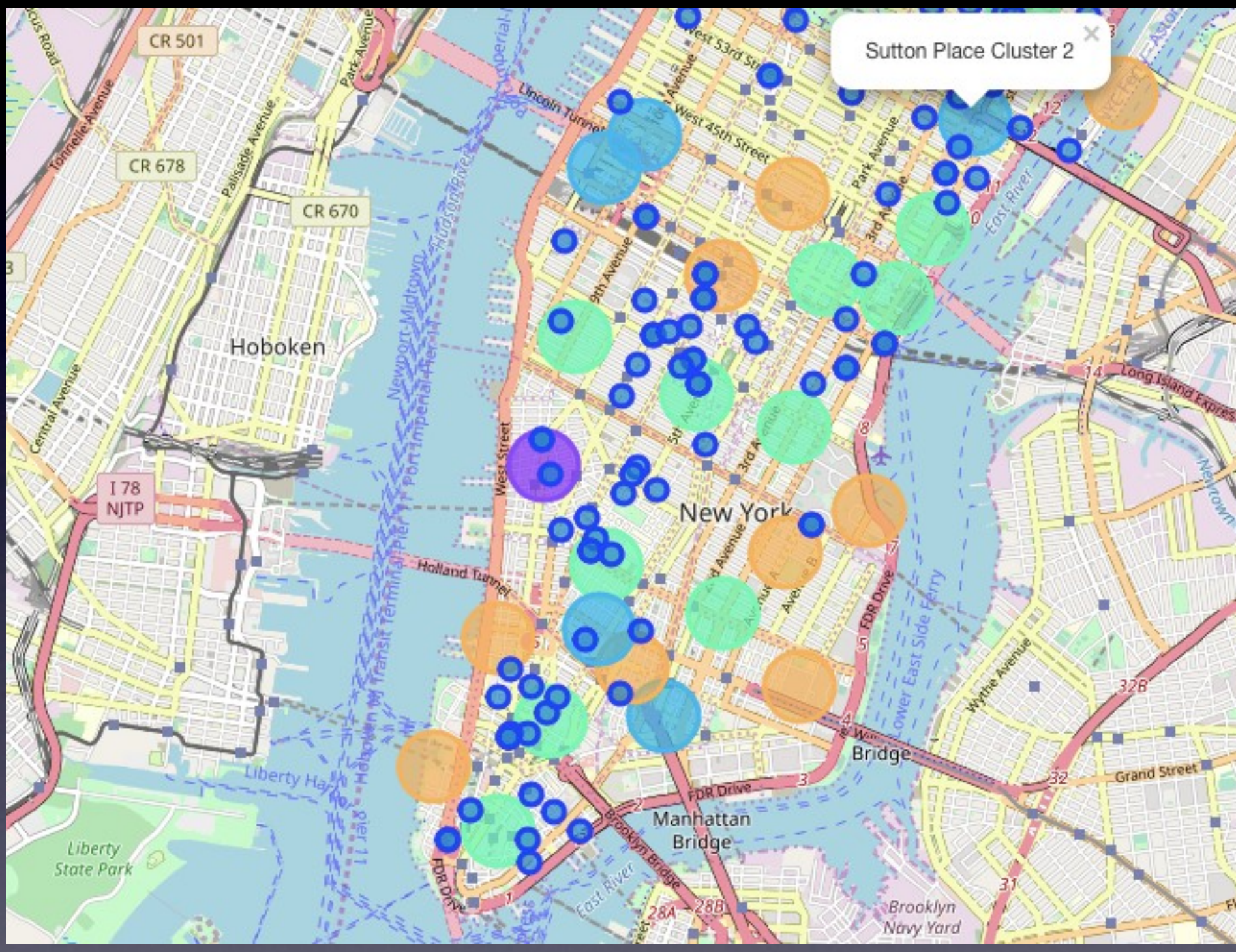
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a25f2a2b0>
```



Apartments for Rent in MH



MH apts for rent with venue clu



Venues of cluster 3

```
## kk is the cluster number to explore
kk = 3
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == kk, manhattan_merged.columns[[
```

	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
3	Inwood	Mexican Restaurant	Lounge	Pizza Place	Café	Wine Bar	Bakery	American Restaurant
5	Manhattanville	Deli / Bodega	Italian Restaurant	Seafood Restaurant	Mexican Restaurant	Sushi Restaurant	Beer Garden	Coffee Shop
10	Lenox Hill	Sushi Restaurant	Italian Restaurant	Coffee Shop	Gym / Fitness Center	Pizza Place	Burger Joint	Deli / Bodega
12	Upper West Side	Italian Restaurant	Bar	Bakery	Vegetarian / Vegan Restaurant	Indian Restaurant	Coffee Shop	Cosmetics Shop
16	Murray Hill	Sandwich Place	Hotel	Japanese Restaurant	Gym / Fitness Center	Coffee Shop	Salon / Barbershop	Burger Joint
17	Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bakery	Nightclub	Theater	Art Gallery
18	Greenwich Village	Italian Restaurant	Sushi Restaurant	French Restaurant	Clothing Store	Chinese Restaurant	Café	Indian Restaurant
27	Gramercy	Italian Restaurant	Restaurant	Thrift / Vintage Store	Cocktail Bar	Bagel Shop	Coffee Shop	Pizza Place
29	Financial District	Coffee Shop	Hotel	Gym	Wine Shop	Steakhouse	Bar	Italian Restaurant
31	Noho	Italian Restaurant	French Restaurant	Cocktail Bar	Gift Shop	Bookstore	Grocery Store	Mexican Restaurant

Manhattan subway stations ge

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

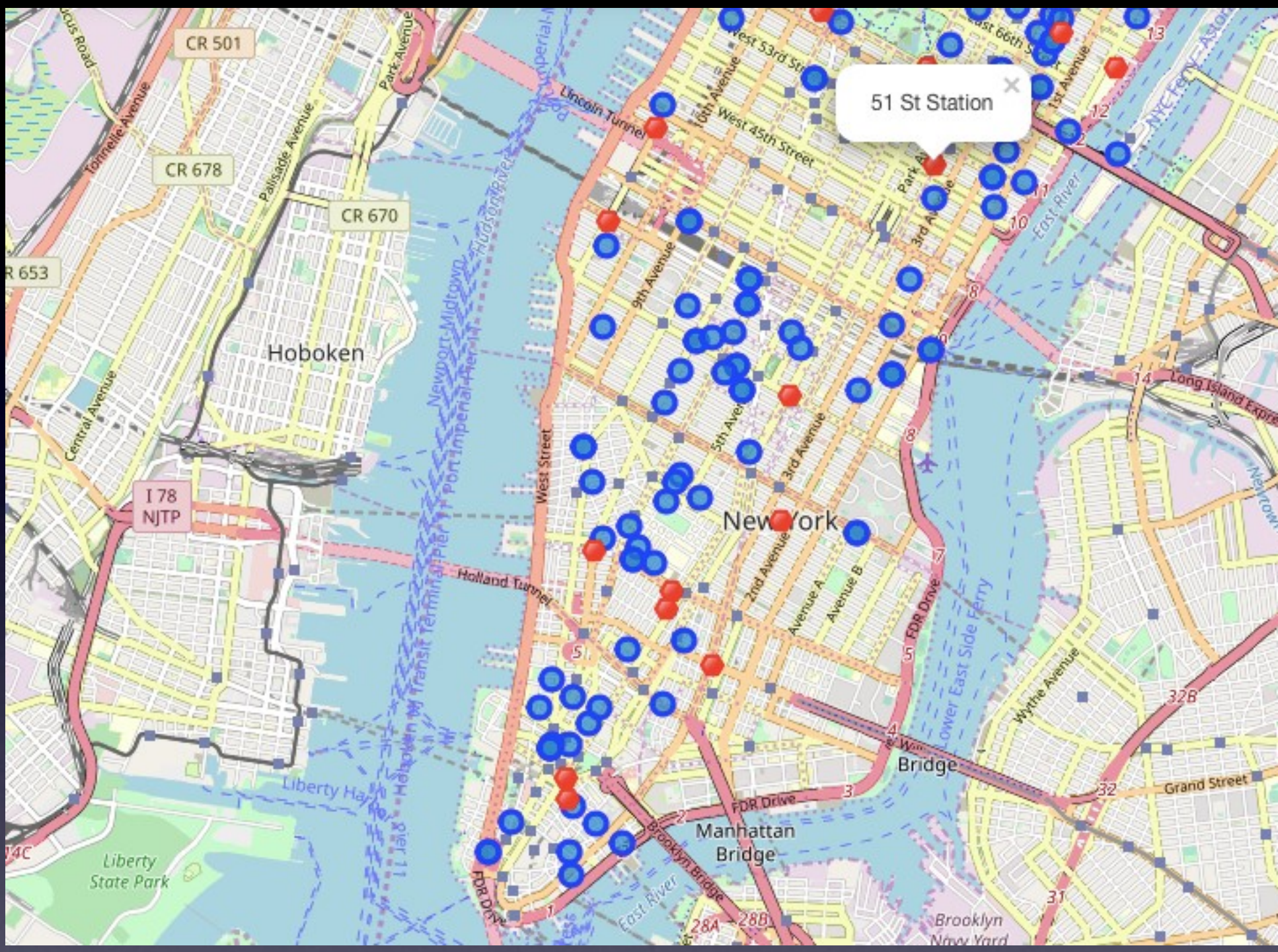
```
# removing duplicate rows and creating new set mhsub1
mhsub1=mh.drop_duplicates(subset=['lat','long'], keep="last").reset_index
mhsub1.shape
```

```
(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	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 station

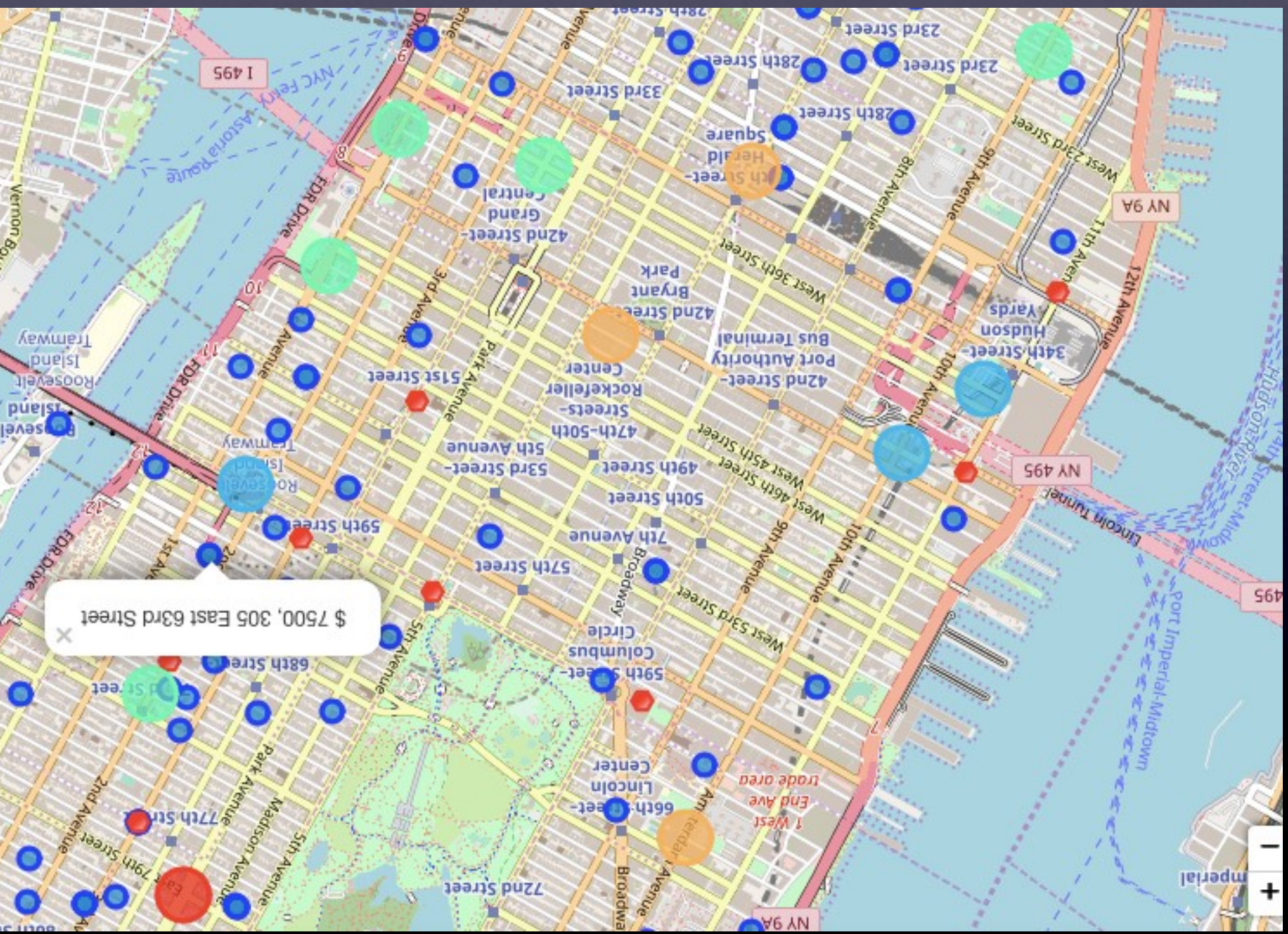


Selected Apartment!

The ONE consolidated map shows all information for de

Apartments address, price, neighborhood, cluster of venues and subw

Blue dots=apts , Red dots=Subway station, Bubbles=Cluster


$ 7500, 305 East 63rd Street
42nd Street-Grand Central
42nd Street Bryant Park
42nd Street-Port Authority Bus Terminal
34th Street-Hudson Yards
33rd Street
28th Street
23rd Street
Herald Square
8th Avenue
6th Avenue
10th Avenue
11th Avenue
West 23rd Street
West 36th Street
West 45th Street
West 46th Street
West 53rd Street
51st Street
49th Street
50th Street
47th-50th Streets-Rockefeller Center
53rd Street-5th Avenue
57th Street
59th Street
59th Street-Columbus Circle
66th Street-Lincoln Center
72nd Street
68th Street
77th Street
79th Street
7th Avenue
Broadway
Park Avenue
Madison Avenue
5th Avenue
2nd Avenue
3rd Avenue
1st Avenue
Lexington Avenue
Amsterdam Avenue
FDR Drive
NY 9A
NY 495
I 495
Lincoln Tunnel
Port Imperial-Midtown
Hudson River Greenway
NYC Ferry - Astoria Route
Roosevelt Island Tramway
1 West End Ave trade area

Apartment Selection

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

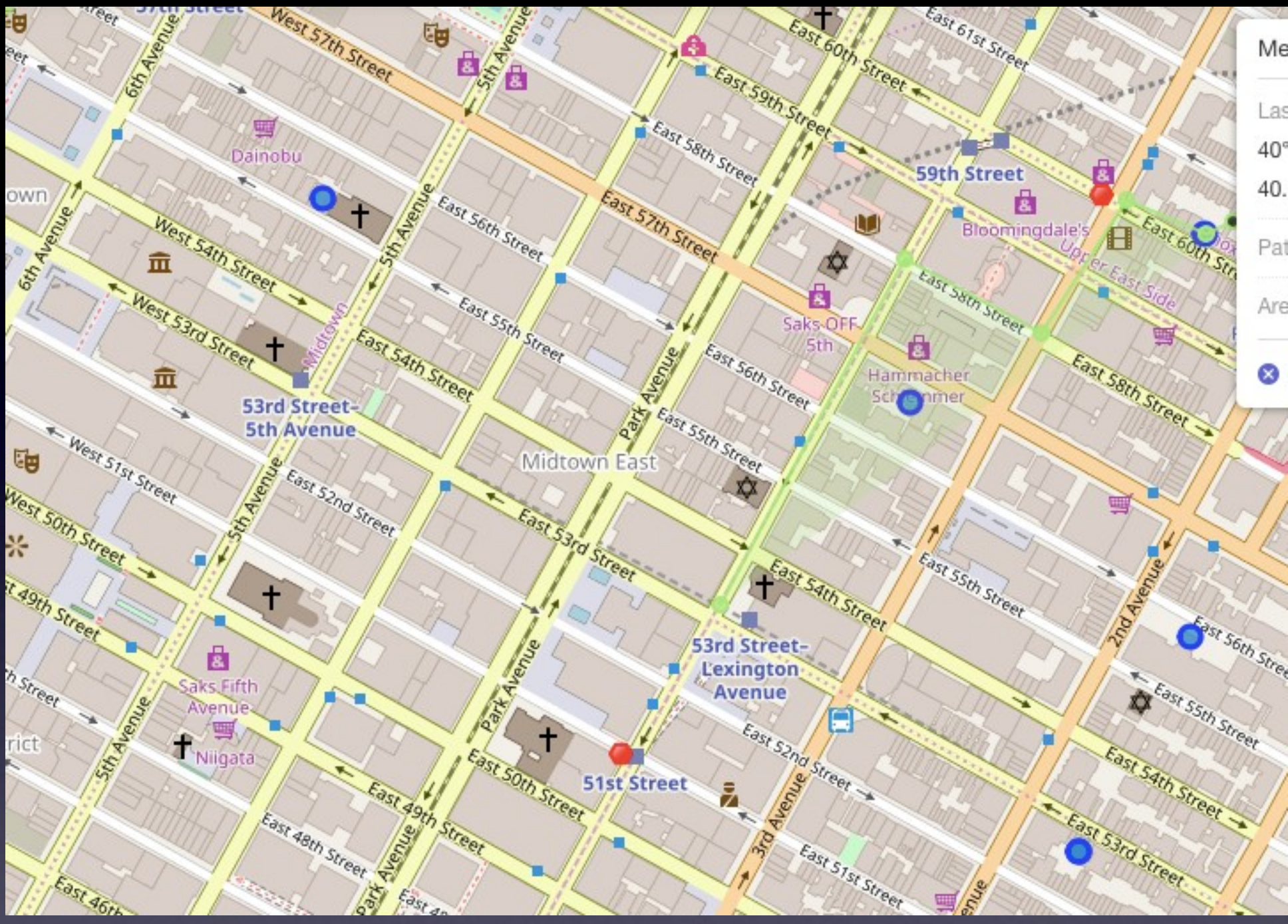
Apartment 1 rent cost is US7500 slightly above the US7000 budge
400 meters from subway station at 59th Street and work place (Pa
another 600 meters way. I can walk to work place and use subway
around. Venues for this apt are as of Cluster 2 and it is located in a
East side of Manhattan.

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

Based on current Singapore venues, I feel that Cluster 2 type of ve
resemblance to my current place. That means that APARTMENT 1
since the extra monthly rent is worth the conveniences it provides.

I will walk to work Walk

from home to work is less than 1 km!


West 57th Street
East 60th Street
East 61st Street
East 59th Street
East 58th Street
East 57th Street
East 56th Street
West 54th Street
East 55th Street
West 53rd Street
East 54th Street
East 56th Street
East 55th Street
East 58th Street
West 51st Street
East 52nd Street
East 53rd Street
East 54th Street
East 55th Street
West 50th Street
East 49th Street
East 50th Street
East 48th Street
East 51st Street
East 52nd Street
East 53rd Street
East 54th Street
East 55th Street
East 56th Street
6th Avenue
5th Avenue
Park Avenue
3rd Avenue
2nd Avenue
Dainobu
59th Street
Bloomingdale's
Upper East Side
Saks OFF 5th
Hammacher Schlemmer
Midtown
53rd Street–5th Avenue
Midtown East
Saks Fifth Avenue
Niigata
51st Street
53rd Street–Lexington Avenue

Venus 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	
0	Marble Hill	Coffee Shop	Discount Store	Yoga Studio	Steakhouse	Supplement Shop	Tennis Stadium	Shoe Store	
1	Chinatown	Chinese Restaurant	Cocktail Bar	Dim Sum Restaurant	American Restaurant	Vietnamese Restaurant	Salon / Barbershop	Noodle House	
6	Central Harlem	African Restaurant	Seafood Restaurant	French Restaurant	American Restaurant	Cosmetics Shop	Chinese Restaurant	Event Space	Liq
9	Yorkville	Coffee Shop	Gym	Bar	Italian Restaurant	Sushi Restaurant	Pizza Place	Mexican Restaurant	Deli
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	Medi R
26	Morningside Heights	Coffee Shop	American Restaurant	Park	Bookstore	Pizza Place	Sandwich Place	Burger Joint	
34	Sutton Place	Gym / Fitness Center	Italian Restaurant	Furniture / Home Store	Indian Restaurant	Dessert Shop	American Restaurant	Bakery	
39	Hudson Yards	Coffee Shop	Italian Restaurant	Hotel	Theater	American Restaurant	Café	Gym / Fitness Center	R

5.0 Discussion

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

- I feel this Capstone project presented m
  opportunity to practice and apply the Da
  tools and methodologies learned. • I hav
  good project that I can present as an exa
  my potential.

6.0 Conclusions

- I feel I have acquired a good starting poin
  a professional Data Scientist and I will co
  exploring to creating examples of practic

- I feel rewarded with the efforts, time and mo
  believe this course with all the topics covere
  worthy of appreciation.

- This project has shown me a practical applica
  a real situation that has impacting personal an
  impact using Data Science tools.

- The mapping with Folium is a very powerful t
consolidate information and make the analysi
thoroughly and with confidence. I would reco
use in similar situations.

- One must keep abreast of new tools for DS t
to appear for application in several business f