

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

We now discuss the business problem and the audience who will be interested in this project.

1.1SCENARIO AND BACKGROUND

I have chosen CHANDIGARH, INDIA (my hometown) as the location to be used as current location for this project.

Being a growing town with many breweries, malls, gyms, clubs, restaurants, shops, cafes opening with a rapid pace, my hometown surely serves a great place to live

1.2PROBLEM TO BE RESOLVED

Many youngsters are migrating to foreign countries in search of jobs, better education and thus want to find an apartment at least similar to one they used to live.

Thus the problem is to find similar apartment in MANHATTAN, NEW YORK, USA with similar amenities which were around in Chandigarh

-as a reference I have included the map of venues in Chandigarh.

1.3AUDIENCE FOR THIS PROJECT

I think that this project will be useful to every person considering moving to foreign countries similar to USA, basically anyone migrating to developed nation.

Also I think this will help travel agencies partner with appropriate entities offering accommodation, homestays, lodging, etc.

Also this will surely develop my data science skills

The use of foursquare data and mapping techniques along with data analysis and visualization will help solve the core problem

2. DATA SECTION

We now describe the data and its sources used in this project

2.1 DATA OF CURRENT PLACE (CHANDIGARH HERE)

I currently reside in neighbourhood of Chandigarh in city Panchkula, Haryana, India.

It is directly connected to Chandigarh and shares its borders with it.

Data from FOURSQUARE will be used to identify the main and appropriate venues near Chandigarh.

This data will be superimposed on the map of Chandigarh, INDIA and thus will serve as reference to find desired location in MANHATTAN, NY.

2.2 DATA NEEDED TO SOLVE THE PROBLEM

To make the best prediction of the required apartment in Manhattan, data we need:

- Information of the neighbourhoods in Manhattan with their geo data (coordinates-latitude and longitude)
- Information of the apartments listed for renting out in Manhattan
- List of basic facilities provided at each neighbourhood (top 5)
- List of metro/bus stations near and in each neighbourhood.

2.3 DATA MANIPULATION AND DATA SOURCES

Data for Manhattan neighbourhoods will be taken directly from capstone week-3 lab since it was well cleaned. Also clustering of the data will be shown accordingly

Data of the venues around the current residence i. e Chandigarh will be gathered from the FourSquare API.

Data of other amenities, places of rent, of metro stations, etc will also be derived from foursquare API.

2.4 INFORMATION IN DATA

The data of venues around Chandigarh extracted from foursquare consists of information of venue and we thus clean it to get the NAME OF VENUE, CATEGORY OF VENUE, LATITUDE AND LONGITUDE OF VENUE

Coordinates will be used to map the data on the map of Chandigarh.

The data of manhattan neighborhoods consists of Borough name, neighbourhood name , postal code , and coordinates.

The data will be clustered based on the :

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

3. METHODOLOGY

3.1 STEPS AND STRATEGY USED TO SOLVE THE PROBLEM

We used FourSquare API to get the coordinates of the venues around the current residence (chandigarh) .

Then we compared those venues with clustered neighbourhood (which were clustered by K-Means clustering on the basis of venues around them and frequency of venues).

Then, following rules were kept in mind while filtering out apartments in manhattan from which we will finally choose apartments:

1. Must be close to 1 or more metro stations.
2. Must be in the neighbourhood containing similar venues (cluster=1 here in my personal case)
3. Rent price must be lower than our budget or atleast in our budget
4. No of rooms didn't matter in my personal case but could be an important factor if number of people are taken in account.

Visualization of data (Maps) gave us very easy way to identify the appropriate apartments since we could easily find the needed by checking if the apartment is near the required amenities or not.

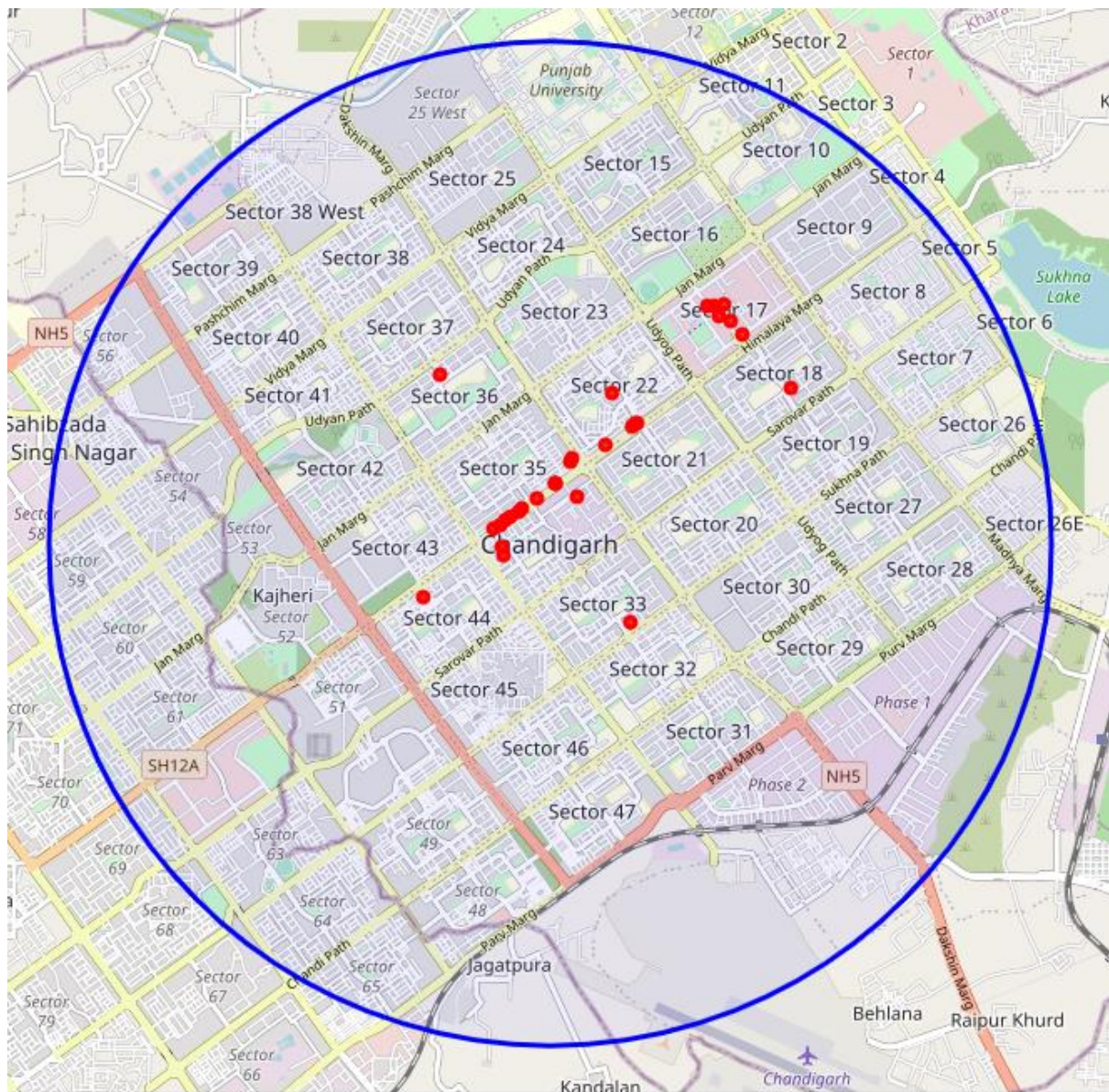
4. RESULTS

Venues around my hometown (chandigarh, india) extracted by making FourSqaure API calls.

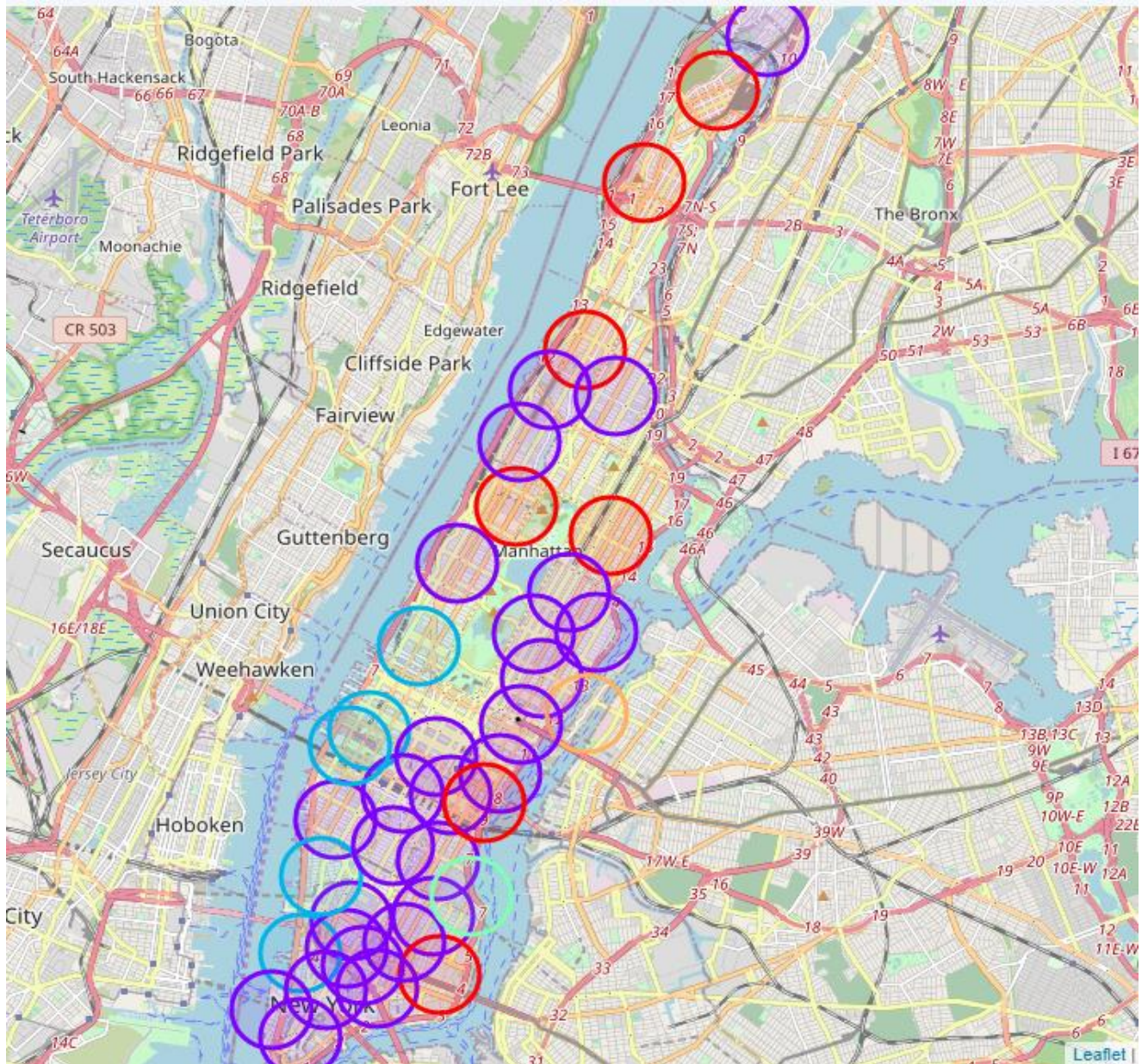
	name	categories	lat	lng
0	Nik Bakers	Bakery	30.721447	76.760176
1	JW Marriott Hotel Chandigarh	Hotel	30.726698	76.766827
2	The Coffee Bean & Tea Leaf	Coffee Shop	30.729740	76.773197
3	chef's gateway Jagat Singh & Sons Agencies	Organic Grocery	30.728041	76.770405
4	Chandigarh Mohali ਚੰਡੀਗੜ੍ਹ	City	30.712486	76.773030
5	Ovenfresh	American Restaurant	30.721800	76.760700
6	Aroma 22	Hotel Bar	30.730059	76.773615
7	Classic 44	Bakery	30.714746	76.751775
8	Subway	Sandwich Place	30.719116	76.759693
9	Barista	Café	30.722207	76.761461
10	Piccadily Square	Multiplex	30.723543	76.767475
11	Peddlers	American Restaurant	30.720781	76.758961
12	Sector 36	Plaza	30.734333	76.753392
13	Gopal's	Indian Restaurant	30.721003	76.759519
14	Basil and Bean Co.	Café	30.723363	76.763399
15	Baskin-Robbins	Ice Cream Shop	30.718387	76.760002
16	The Cafe @ JW	Café	30.726890	76.767007
17	Girl In The Café	Café	30.737845	76.784498
18	Giani's - 35/B	Ice Cream Shop	30.724726	76.765288
19	Nando's	Australian Restaurant	30.724701	76.765239
20	Ghazal	Indian Restaurant	30.739055	76.783358
21	Sector 17	Miscellaneous Shop	30.739541	76.782158
22	Softy Corner	Ice Cream Shop	30.740414	76.781619
23	Shangrila Plus	Chinese Restaurant	30.724732	76.765283
24	Cafe Coffee Day	Coffee Shop	30.732760	76.771070
25	Cafe Coffee Day	Café	30.722522	76.761861
26	McDonald's	Fast Food Restaurant	30.730100	76.773461

(*there were some more entries in the above table but I couldn't take screenshot of all.)

4.1 MAP OF CHANDIGARH WITH ABOVE VENUES



4.2 MAP OF MANHATTAN, NEW YORK WITH CLUSTERED NEIGHBORHOODS ALONG VENUES



This is the dataframe head containing clustered venues with neighborhoods

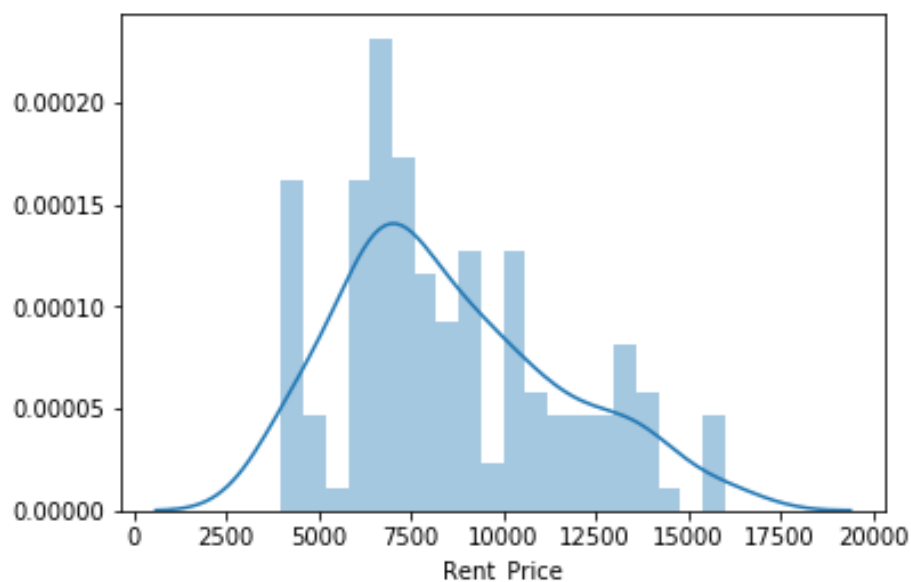
	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	1	Sandwich Place	Gym	Coffee Shop	Yoga Studio	Deli / Bodega	Steakhouse	Shopping Mall	Seafood Restaurant	Pizza Place	Department Store
1	Manhattan	Chinatown	40.715618	-73.994279	1	Chinese Restaurant	Cocktail Bar	American Restaurant	Spa	Bakery	Hotpot Restaurant	Optical Shop	Vietnamese Restaurant	Salon / Barbershop	Dessert Shop
2	Manhattan	Washington Heights	40.851903	-73.936900	0	Café	Bakery	Grocery Store	Chinese Restaurant	Mobile Phone Shop	Latin American Restaurant	Deli / Bodega	Tapas Restaurant	New American Restaurant	Park
3	Manhattan	Inwood	40.867684	-73.921210	0	Mexican Restaurant	Lounge	Restaurant	Café	Pizza Place	Wine Bar	American Restaurant	Bakery	Frozen Yogurt Shop	Deli / Bodega
4	Manhattan	Hamilton Heights	40.823604	-73.949688	0	Pizza Place	Coffee Shop	Café	Mexican Restaurant	Sandwich Place	Indian Restaurant	Sushi Restaurant	Park	Deli / Bodega	Yoga Studio

4.3GEODATA OF APARTMENTS ON RENT IN MANHATTAN USING nominatim()

	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	40.788517	-73.955118
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	40.730337	-74.002476
4	171 W 23RD ST.	Chelsea		3.45	2.0	1450	5000	40.744118	-73.995299

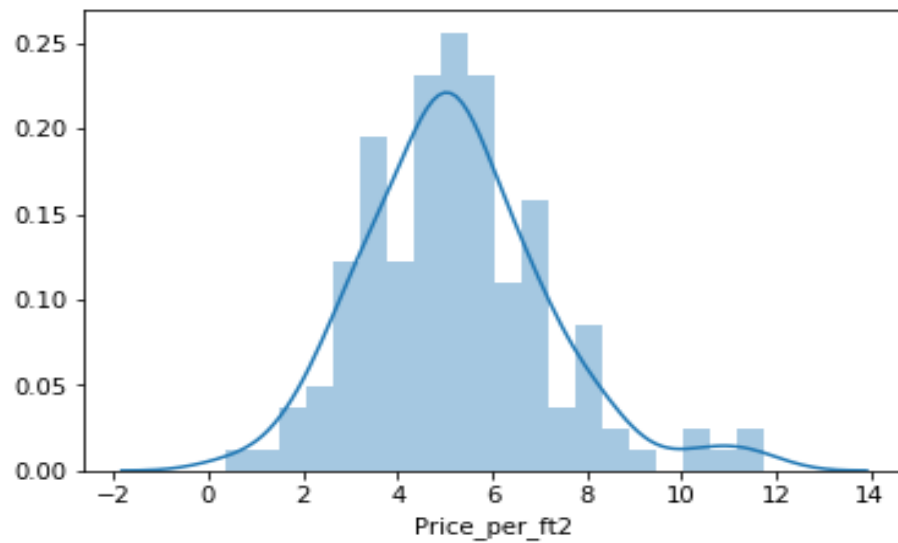
Analysis of above rental geodata

1. Plotting univariate distribution of Rent_price



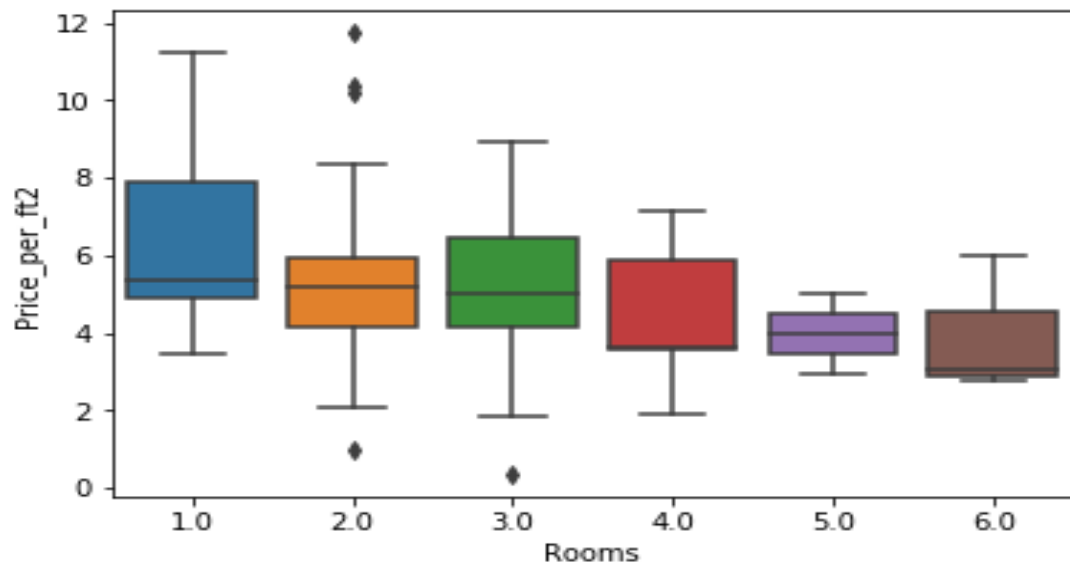
Observation from above plot- maximum rental apartments have rent close from 6000 to 7000 US dollar which is greater than my budget but being developed nation with better facilities they'll definitely charge more.

2. Plotting univariate distribution of Price per square foot



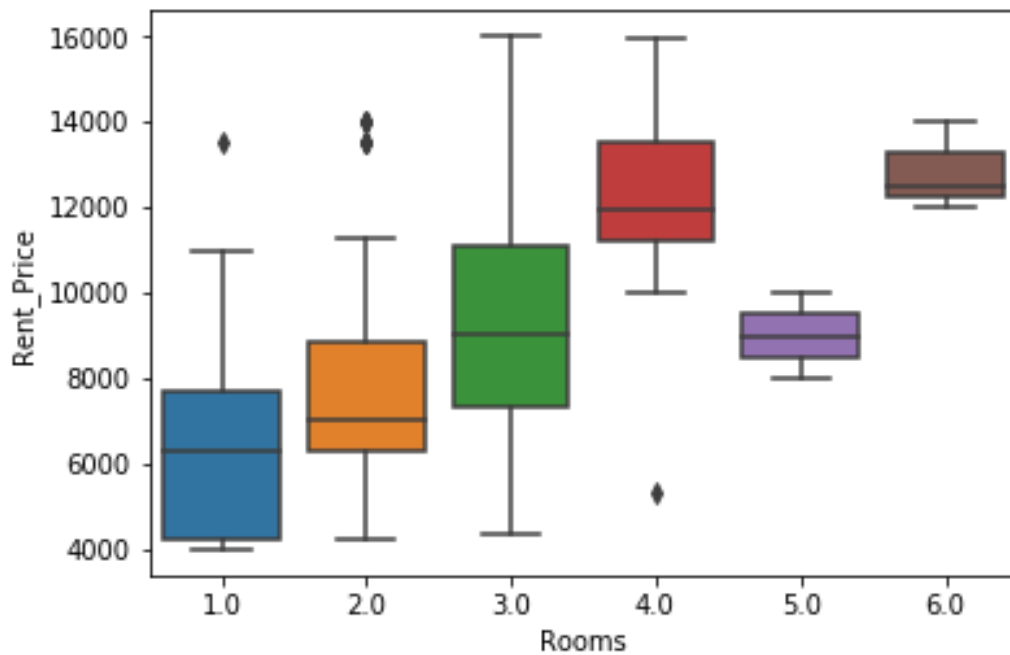
Observation from above plot- most are charging around 5.5 to 6 dollar per square feet

3. Boxplot of rooms vs price per square foot



Observation from above plot- if we consider 2 room apartment then median is around 5.6(price_per_ft2) i.e 50 percent of apartments on rent with 2 rooms are having per square foot price less than 5.5 US dollar which seems a good deal !

4. Boxplot of rooms vs rental price



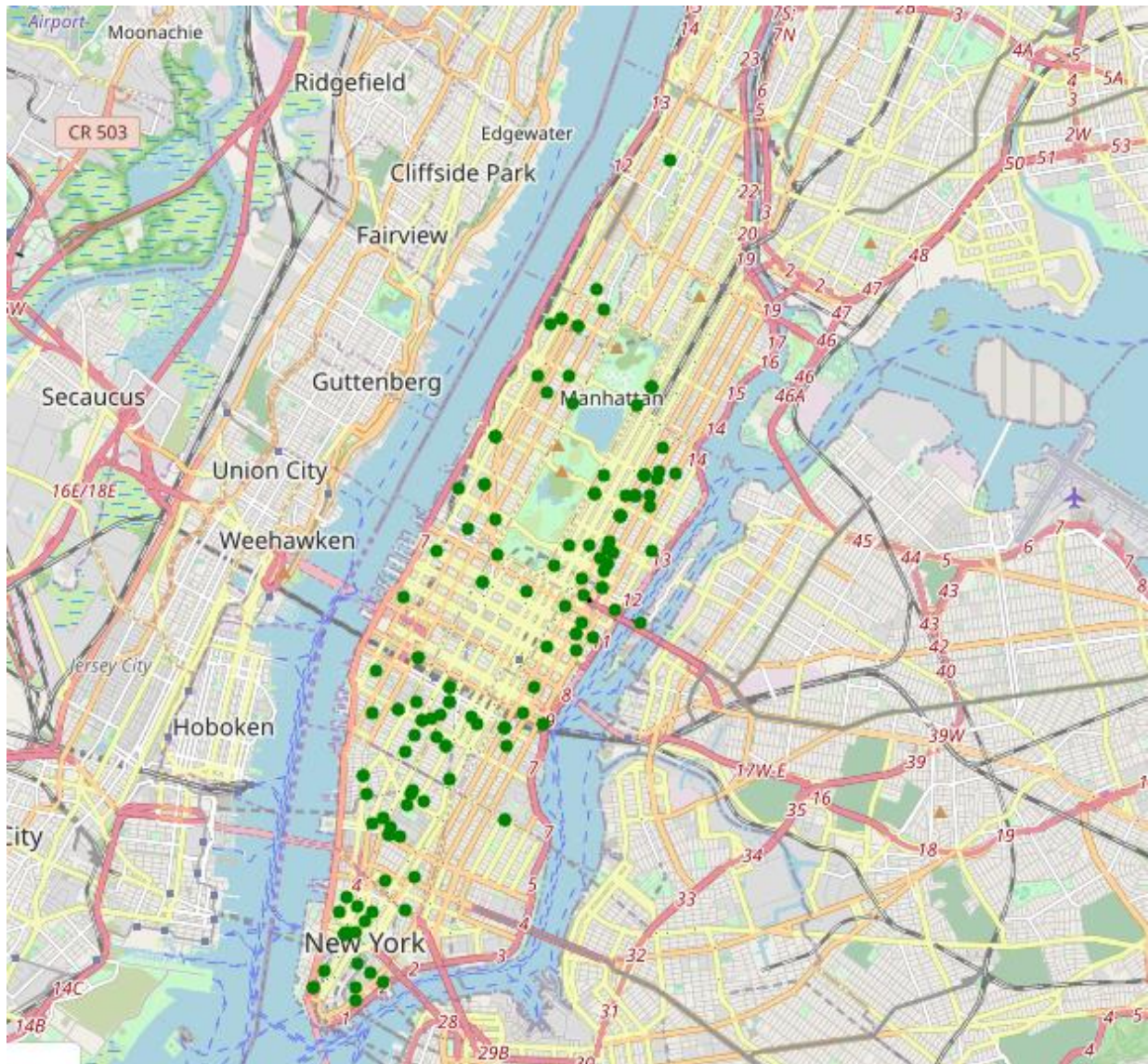
Observtion from above plot-

1 considering 1 room and 2 room apartments on rent, we can easily find apartment under 7500 US dollar.

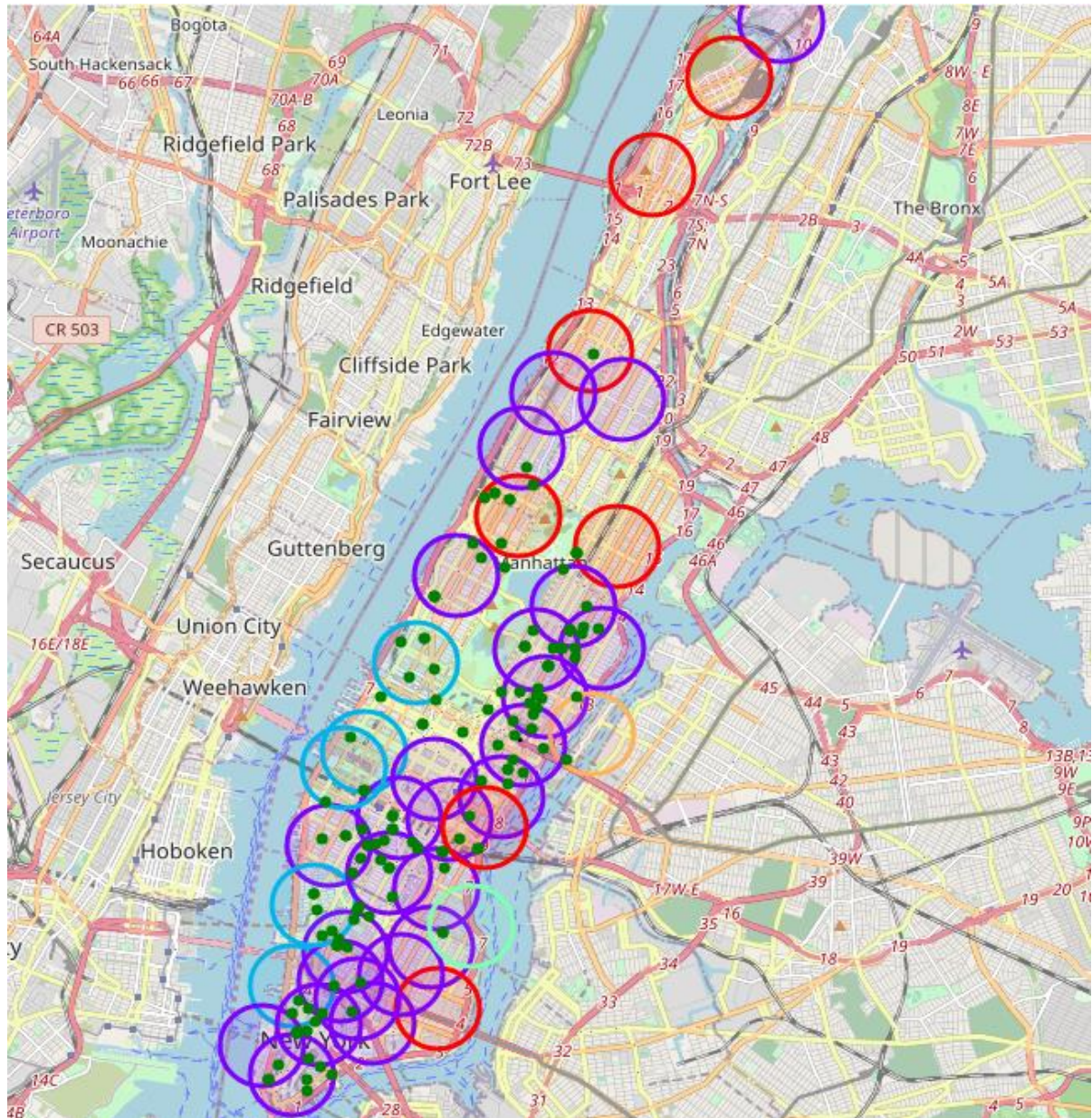
2 median of 1 room apartments is around 6500 rent price and median of 2 room apartments is above 7000 rent price

3 there is a outlier in 4 room apartments who i charging less than 6000 US dollar and thus also seems like a great deal.

4.4 APARTMENTS ON RENT IN MANHATTAN,NY



4.5 APARTMENTS ON RENT IN MANHATTAN,NY WITH CLUSTERS



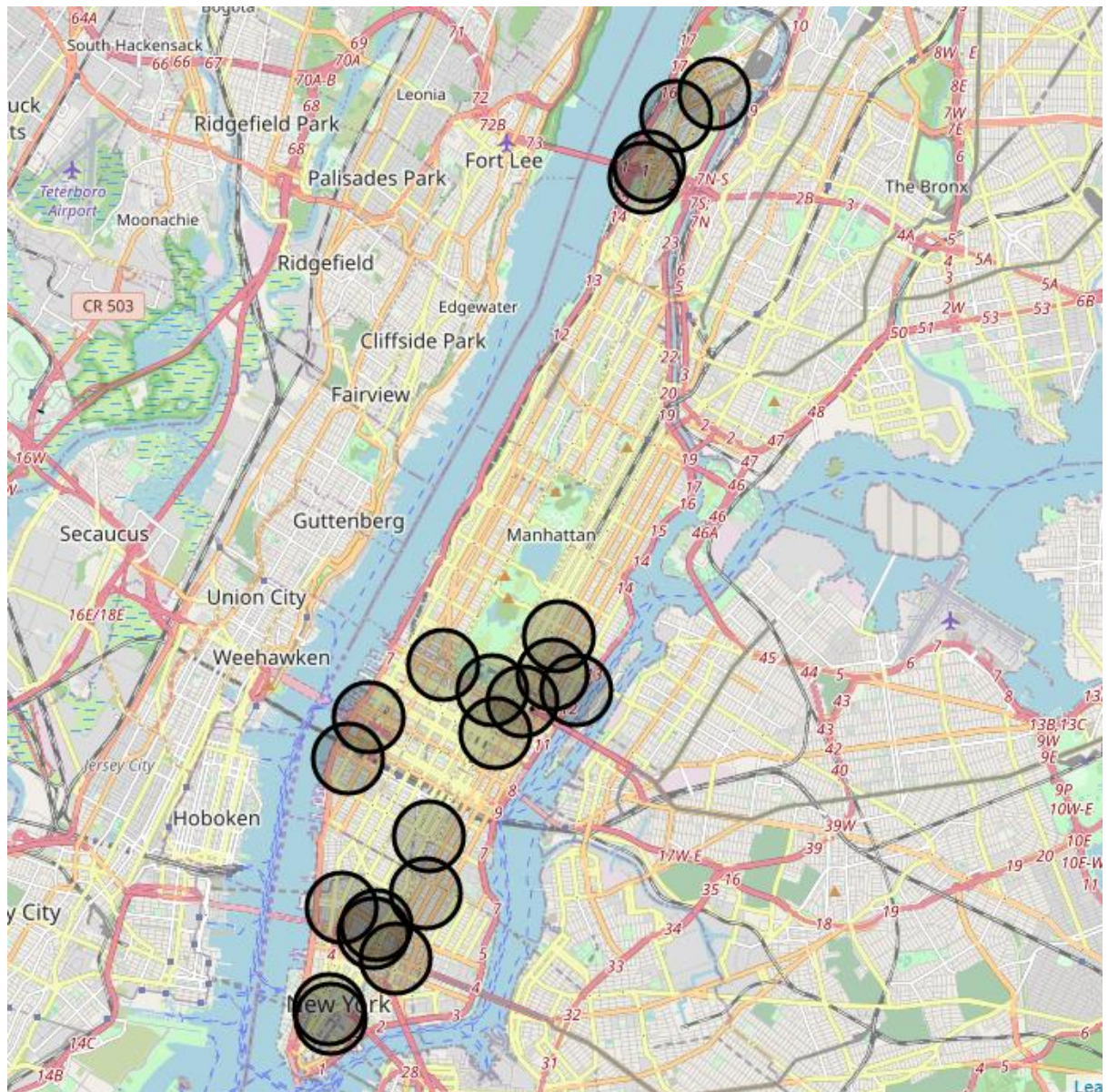
Here we can easily check If the apartments are in the preferred neighbourhood or not ...

Also some apartments on rent which are not in any clustered venue neighbourhood can be disqualified straight away.

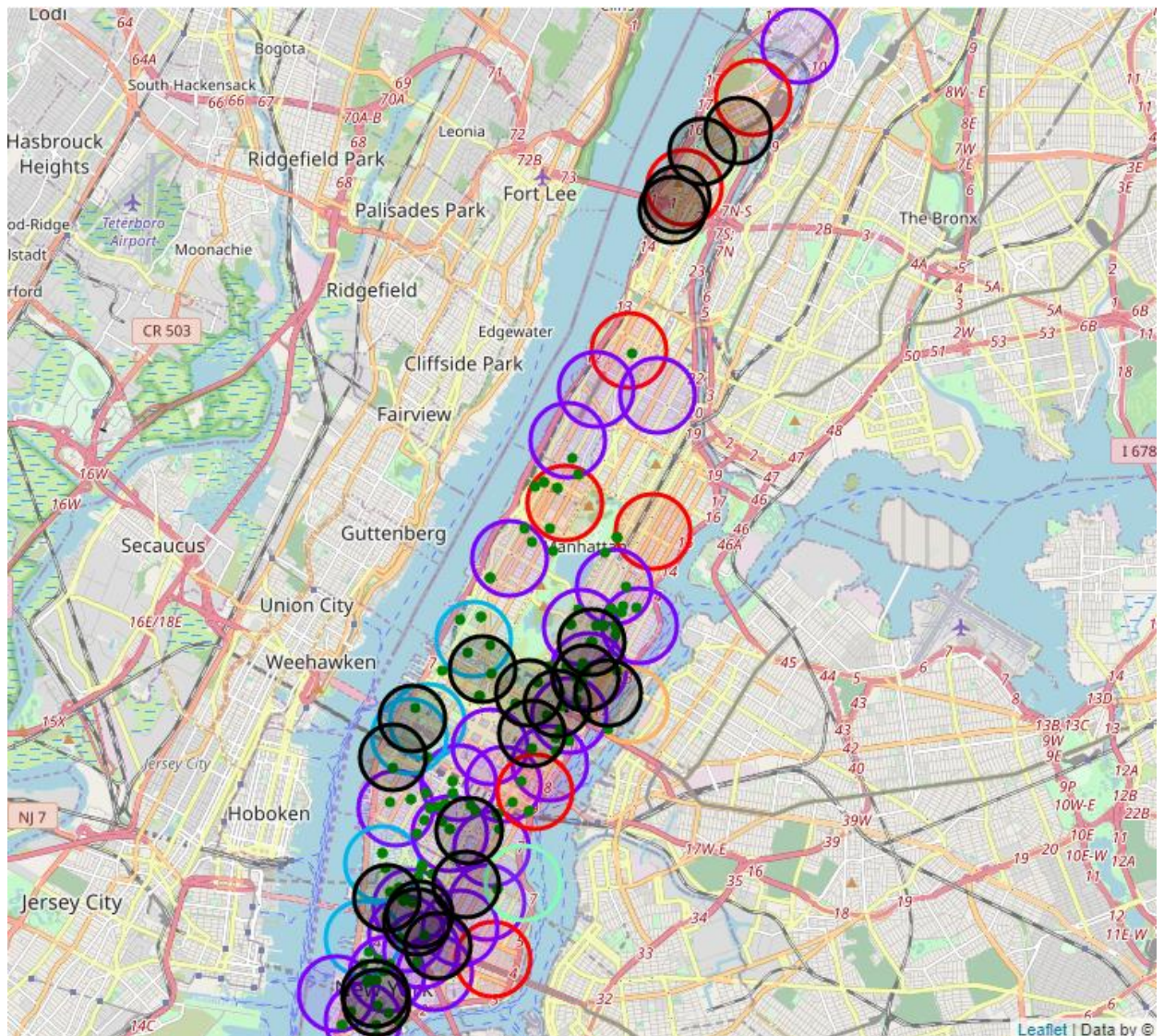
4.6 MAP OF MANHATTAN WITH METRO STATIONS MARKED

(Metro stations are in exactly middle of each black circle marker,

Circle markers are made bigger just to check if the apartment is near to the metro station or not !)



FINAL MAP OF MANHATTAN WITH APARTEMNTS ON RENT, METRO **STATIIONS AND CLUSTERS OF VENUES**



1. THE MAP ABOVE IS COMBINED MAP WITH ALL THE NECESSARY MARKERS

BLACK MARKERS TELL THE INFORMATION OF METRO STATIONS

GREEN MARKERS TELL THE INFORMATION OF APARTMENTS ON RENT

RED MARKERS TELL THE INFORMATION OF CLUSTER=0 NEIGHBORHOODS

PURPLE MARKERS TELL THE INFORMATION OF CLUSTER=1 NEIGHBORHOODS

SKY BLUE MARKERS TELL THE INFORMATION OF CLUSTER=2 NEIGHBORHOODS

2. we will pick the apartments if its in cluster=1 neighborhood and metro station needs to close

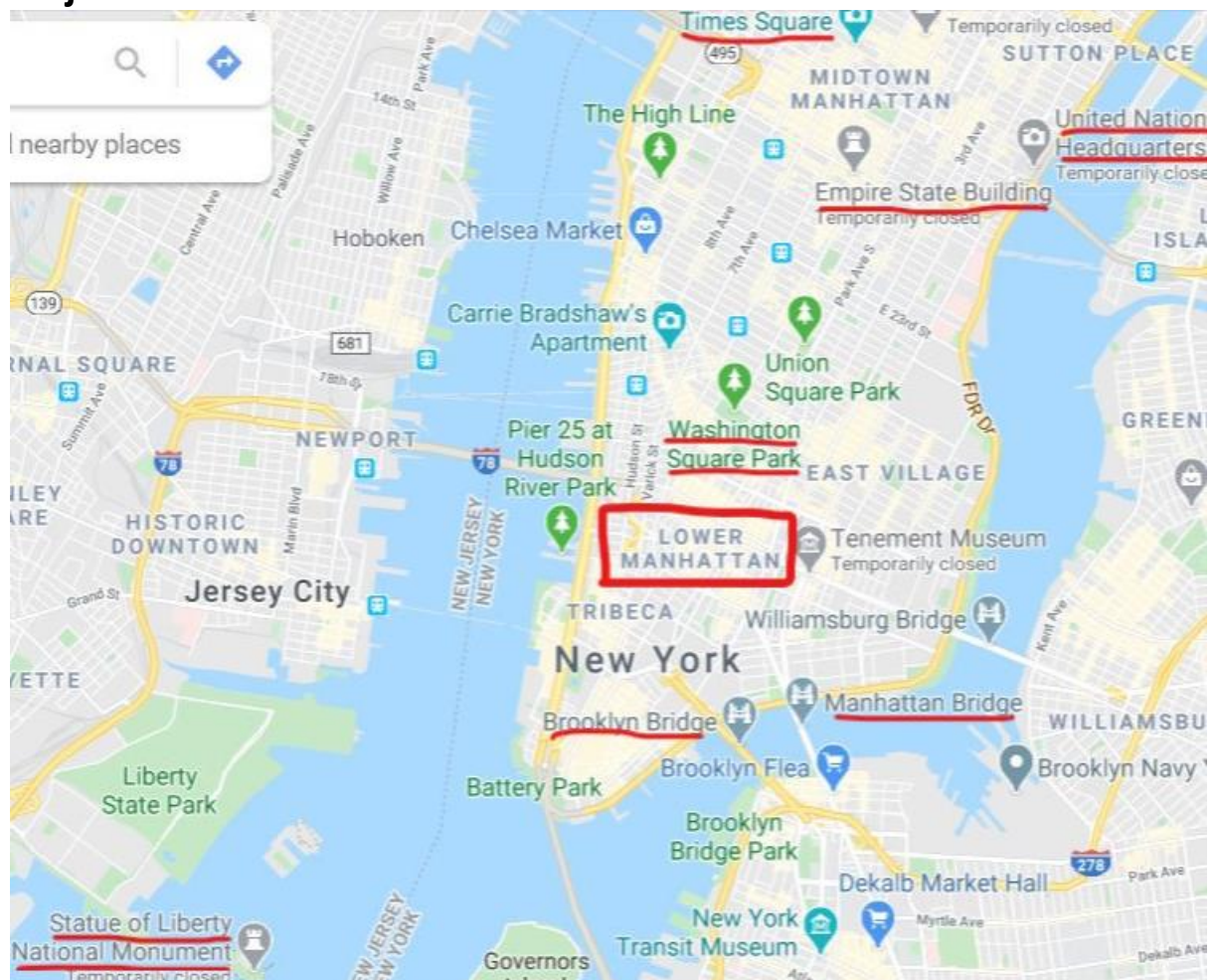
- if we have more stations near then the apartments is preferred
- also keeping in mind the budget is less than 6k -7k US dollars so it'll be great if we find cheap deal
- considering my personal case where i'll be moving alone , no. of rooms won't really affect the decision.

THESE ARE THE FOLLOWING PICKED APARTMENTS FROM FINAL MAP(along with addresses , i have added some comments to better understand what will help choosing final apartment appropriately) :-

- 70 Pine Street,6188\$ (OK CHOICE WITH CLOSE NEIGHBORHOOD BUT FAR METRO)
- 60 Fulton Street,6185 \$ (OK CHOICE WITH CLOSE MTERO BUT FAR NEIGH)
- 211 East 70th Street,4250\$ (CHEAP AND VERY CLOSE TO BOTH)
- 150 East 57th Street,5950\$ (GOOD CHOICE , PERFECT RENT, CLOSE NEIGH AND METRO)
- 52 Spring Street,6000\$ (VV GOOD SURR BY 2 METRO + 2 CLUS=1 NEIGHBORHOOD)
- 182 Bleecker Street,6000\$ (OK CHOICE)
- 160 East 48th Street,6431\$ (OK CHOICE)
- 265 East 66th Street,6425\$ (OK CHOICE ,CLOSE METRO BUT FAR GOOD NEIGHBORHOOD)
- 333 east 81st street ,4500\$ (CHEAP BUT FAR METRO AND FAR NEIGHBORHOOD)

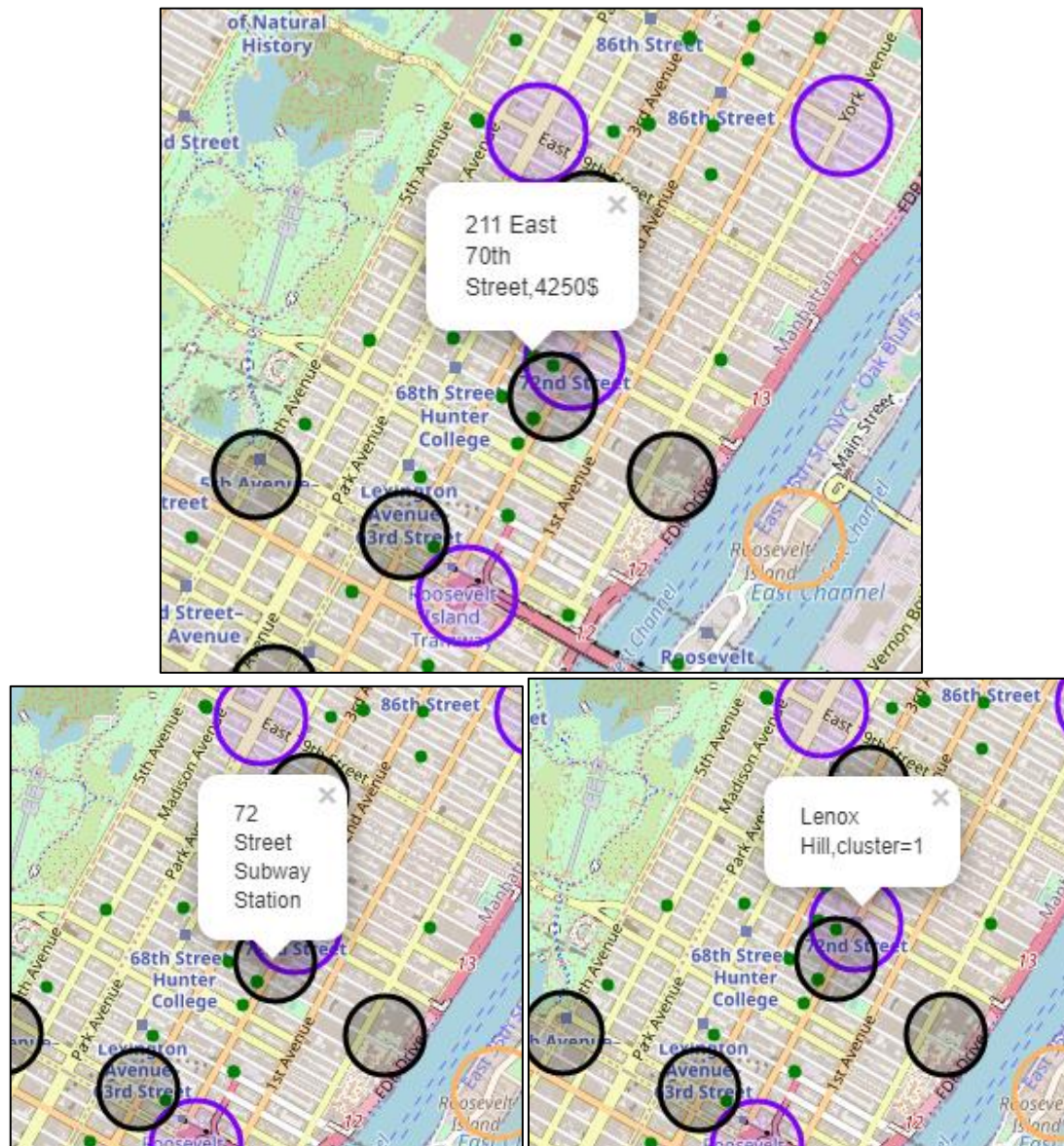
5. DISCUSSION

- The model can help someone moving to manhattan to choose an apartment but to generalize it more , we need to take into account more diverse data , from different countries and different kind of areas.
- Different person might require more facilities and bigger Apartment with more number of rooms and thus the model can help them also.
- One can observe that the apartments on rent are not available in north of manhattan which is isolated a little.
- Most apartments on rent plus metro stations as well as cluster venues can be found in southern part of manhattan which is close to the STATUE OF LIBERTY IN NY as well other major locations....



6. CONCLUSION

In this study , I found out how to find an apartment in New York, With atleast similar amenities as in chandigarh but better in ways such as metro stations nearby !



The above labelled apartment 211 East 70th street (4250\$ rent price) with nearest 72street Subway Station in LENOX HILL neighbourhood(cluster=1) is the one I have chosen personally by considering various personal as well as general factors and I can confidently say that the chosen apartment is the best deal at best price in my opinion.

THANK YOU ..