

Finding the best place of 3 bedroom apartment in Manhattan on Rent

Yogesh Balkrishna Salve

September 23, 2019

1. Introduction

1.1 Background

The easiest way to save money is to avoid paying a broker's fee. "Look for a building with a rental office directly on site," suggests architect Hayes Slade, whose Slade Architecture firm knows something about affordable housing "That way you can negotiate directly with the landlord and skip a real-estate agent fee." In order to ditch online searches, Slade recommends walking the streets of your desired neighborhood and looking for offices and for rent signs. When I was looking for an apartment," she says, "I would go to moving sales every weekend and ask, 'Has your landlord found someone to take your apartment yet? Can I see it?'"

"A good way to find an affordable apartment in Manhattan is to look in buildings that have between 15 and 30 apartments," says Joshua Price of the Price Law Firm LLC, which specializes in landlord-tenant representation. Odds are good that if the building has rent-stabilized apartments, they will stay that way. "In those buildings, it's less likely that, when an apartment becomes vacant, the landlord will expand the enormous amount of money necessary to deregulate an apartment. Generally an investor with a ten-unit building is trying to deregulate every unit and then sell the building. Someone with a 30-unit building is more likely to be thinking to hold it and not spend as much money on each vacancy."

Finally I decided to go with the help of data science and I started finding my own way for me using data science libraries. The problem is defined below.

1.2 Problem

The challenge is to find a suitable apartment for rent in Manhattan that complies with the demands on location, price and venues. This project aims to predict the best location for the apartment with which fulfills the demands on price and venues.

1.3 Interest

I believe this is a relevant challenge with valid questions for anyone moving to other large city in US or Asia The same methodology can be applied in accordance to demands as applicable. This case is also applicable for anyone interested in exploring starting or locating a new business in any city. Lastly, it can also serve as a good practical exercise to develop Data Science skills.

2. Data acquisition and cleaning

2.1 Data Sources

Description of the data and its sources that will be used to solve the problem

The following data is required to answer the issues of the problem:

1. List of Boroughs and neighborhoods of Manhattan with their geodata (latitude and longitude)
2. List of Subway metro stations in Manhattan with their address location
3. List of apartments for rent in Manhattan area with their addresses and price
4. Preferably, a list of apartment for rent with additional information, such as price, address, Area, no. of beds, etc
5. Venues for each Manhattan neighborhood (than can be clustered)
6. Venues for subway metro stations, as needed

The data will be used as follows:

Use Foursquare and geopy data to map top 10 venues for all Manhattan neighborhoods and clustered in groups (as per Course LAB) Use foursquare and geopy data to map the location of subway metro stations, separately and on top of the above clustered map in order to be able to identify the venues and amenities near each metro station, or explore each subway location separately Use Foursquare and geopy data to map the location of rental places, in some form, linked to the subway locations.

A .csv file was created which will be read in order to create a dataframe and its mapping. The csv file 'mh_neigh_data.csv' has the following below data structure. The file will be directly read to the Jupiter Notebook for convenience and space savings. The clustering of neighborhoods and mapping will be shown however. An algorithm was used to determine the geodata from Nominatim. The actual algorithm coding may be shown in 'markdown' mode because it takes time to run.

mh_neigh_data.tail ():

	Borough	Neighborhood	Latitude	Longitude
35	Manhattan	Turtle Bay	40.752042	-73.967708
36	Manhattan	Tudor City	40.746917	-73.971219
37	Manhattan	Stuyvesant Town	40.731000	-73.974052
38	Manhattan	Flatiron	40.739673	-73.990947
39	Manhattan	Hudson Yards	40.756658	-74.000111

A list of Manhattan subway metro stops was compiled in Numbers (Apple excel) and it was completed with Wikipedia data, https://en.wikipedia.org/wiki/List_of_New_York_City_Subway_stations_in_Manhattan) and information from NY Transit authority and Google Maps, (<https://www.google.com/maps/search/manhattan+subway+metro+stations/@40.7837297,-74.1033043,11z/data=!3m1!4b1>) for a final consolidated list of subway stops names and their address. The geolocation was obtained via an algorithm using Nominatim. Details will be shown in the execution of methodology in section 3.0.

The subway csv file is "MH_subway.csv" and the data structure is: `mhsub.tail(): sub_station sub_address lat long`

A list of places for rent was collected by web-browsing real estate companies in Manhattan:

1. <http://www.rentmanhattan.com/index.cfm?page=search&state=results>
2. https://www.nestpick.com/search?city=new-york&page=1&order=relevance&district=manhattan&gclid=CjwKCAiAjNjgBRAGeiwAGLlf2hkP3A-cPxjZYkURqQEswQK2jKQEpv_MvKcrIhRWRzNkc_r-fGi0lxoCA7cQAvD_BwE&type=apartment&display=list
3. https://www.realtor.com/apartments/Manhattan_NY

A csv file was compiled with the rental place that indicated: areas of Manhattan, address, number of beds, area and monthly rental price. The csv file "nnnn.csv" had the following below structure. An algorithm was used to create all the geodata using Nominatim, as shown in section 3.0. The actual algorithm coding may be shown in 'markdown' mode because it takes time to run. With the use of `geolocator = Nominatim ()`, it was possible to determine the latitude and longitude for the subway metro locations as well as for the geodata for each rental place listed. The loop algorithms used are shown in the execution of data in section 3.0 "Great_circle" function from `geolocator` was used to calculate distances between two points, as in the case to calculate average rent price for units around each subway station and at 1.6 km radius. Foursquare is used to find the avenues at Manhattan neighborhoods in general and a cluster are created to later be able to search for the venues depending of the location shown.

2.2 How the data will be used to solve the problem

The data will be used as follows: Use Foursquare and geopy data to map top 10 venues for all Manhattan neighborhoods and clustered in groups (as per Course LAB) Use foursquare and geopy data to map the location of subway metro stations , separately and on top of the above clustered map in order to be able to identify the venues and amenities near each metro station, or explore each subway location separately Use Foursquare and geopy data to map the location of rental places, in some form, linked to the subway locations. Create a map that depicts, for instance, the average rental price per square ft, around a radius of 1.0 mile (1.6 km) around each subway station - or a similar metrics. I will be able to quickly point to the popups to know the

relative price per subway area. Addresses from rental locations will be converted to geodata (lat, long) using Geopy-distance and Nominatim. Data will be searched in open data sources if available, from real estate sites if open to reading, libraries or other government agencies such as Metro New York MTA, etc.

2.3 Mapping of Data

The following maps were created to facilitate the analysis and the choice of the palace to live. Manhattan map of Neighborhoods, Manhattan subway metro locations, Manhattan map of places for rent, Manhattan map of clustered venues and neighborhoods Combined maps of Manhattan rent places with subway locations, Combined maps of Manhattan rent places with subway locations and venues clusters.

3. Methodology section

This section represents the main component of the report where the data is gathered, prepared for analysis. The tools described are used here and the Notebook cells indicates the execution of steps.

The analysis and the strategy:

The strategy is based on mapping the above described data in section 2.0, in order to facilitate the choice of at least two candidate places for rent. The choice is made based on the demands imposed: location near a subway, rental price and similar venues to Singapore. This visual approach and maps with popups labels allow quick identification of location, price and feature, thus making the selection very easy.

The processing of these DATA and its mapping will allow answering the key questions to make a decision:

- What is the cost of available rental places that meet the demands?
- What is the cost of rent around a mile radius from each subway metro station?
- What is the area of Manhattan with best rental pricing that meets criteria established?
- What is the distance from work place (Park Ave and 53 rd St) and the tentative future rental home?
- What are the venues of the two best places to live? How the prices compare?
- How venues distribute among Manhattan neighborhoods and around metro stations?
- Are there tradeoffs between size and price and location?

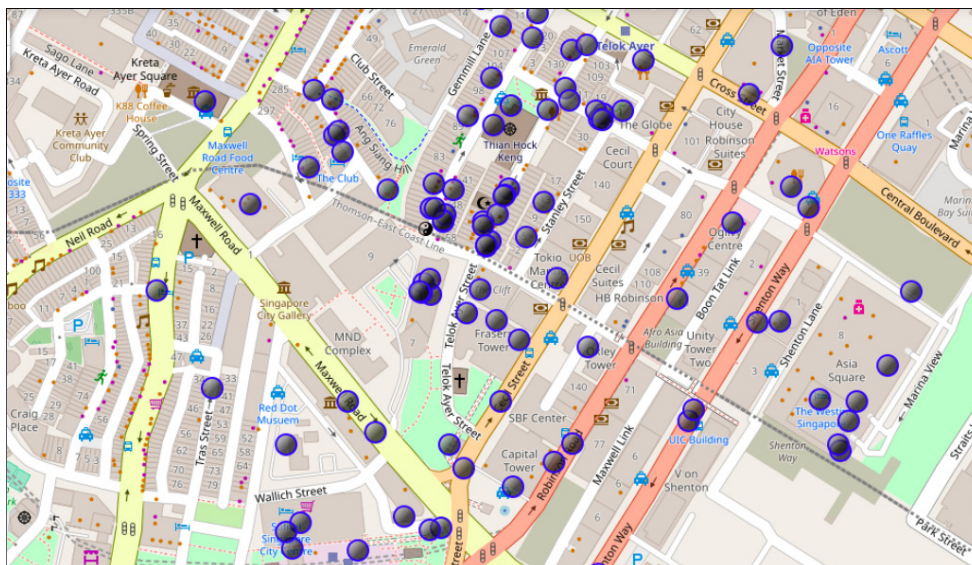
- Any other interesting statistical data findings of the real estate and overall data.

In order to do the analysis and suggest the best location, following are the steps we have to follow:

- The website Rentmanhattan page (<http://www.rentmanhattan.com/index.cfm?page=search&state=results>) was scraped using the BeautifulSoup library to build a pandas dataframe listing the cities, states, coordinates, area and population density. The dataframe was cleaned and processed appropriately.
- The Website page (https://www.nestpick.com/search?city=new-york&page=1&order=relevance&district=manhattan&gclid=CjwKCAiAjNjgBRAGeiwAGLlf2hkP3A-cPxjZYkURqQEswQK2jKQEpv_MvKcrIhRWRzNkc_r-fGi0lxoCA7cQAvD_BwE&type=apartment&display=list) was scraped using the BeautifulSoup library to build a pandas dataframe listing the cities, states and percapita income. The dataframe was cleaned and processed appropriately.
- The Foursquare API is then used to get the nearby venues in Manhattan.
- Based on the categories of each venue as decided above, I have assigned weights to each of them and got the city area that has the maximum weight.
- Once the area in Manhattan is finalized, I again use Four Square API to get the venues within that city and assign weights to each category.
- I will now use K means to cluster the venues based on the category and get the coordinates of the cluster that has maximum weight which is also our preferred location to for taking apartment on rent.

4. Execution and Result

Current Residence neighborhood in Singapore

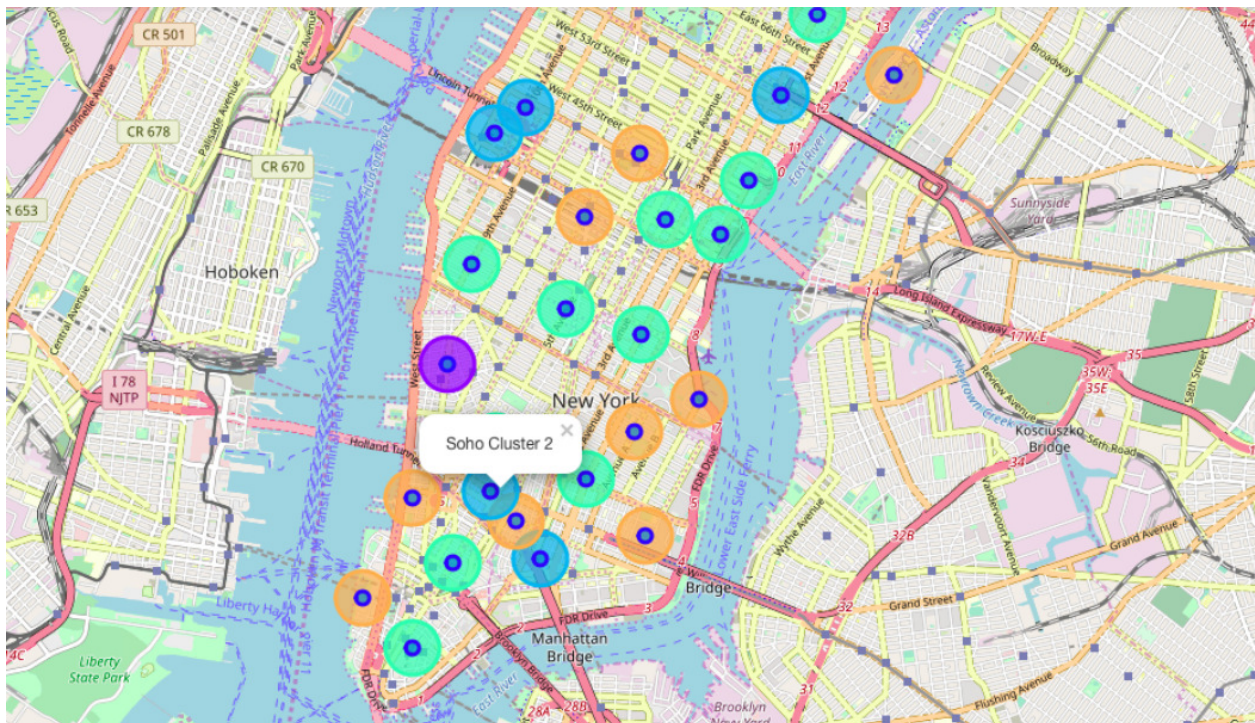


Venues around neighborhood in:

```
# 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

Manhattan Map – Neighborhoods and cluster of venues



Geodata: Manhattan Apartment 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

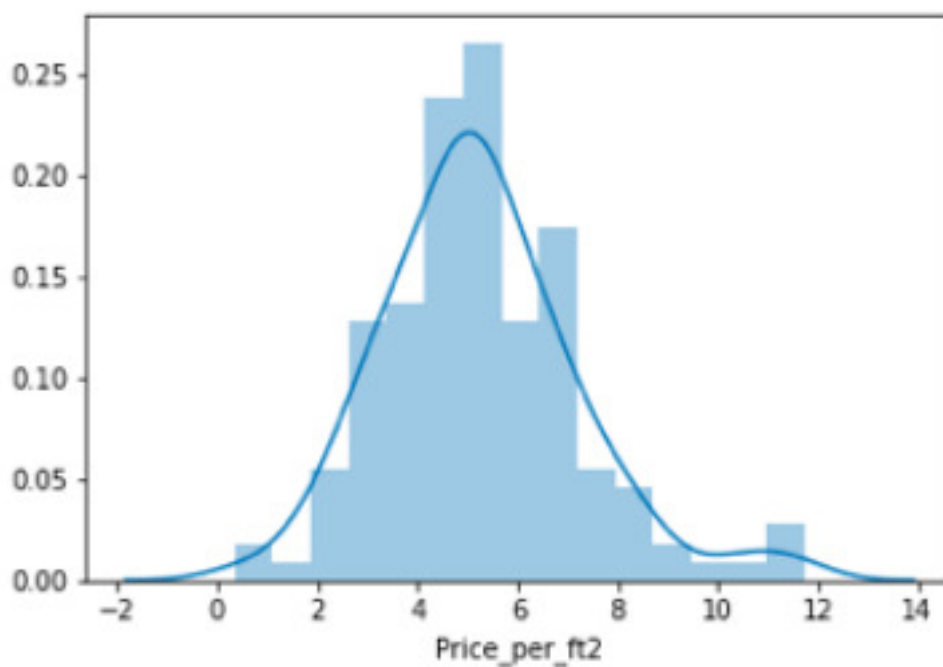
```
]: mh_rent.tail()
```

```
]:
```

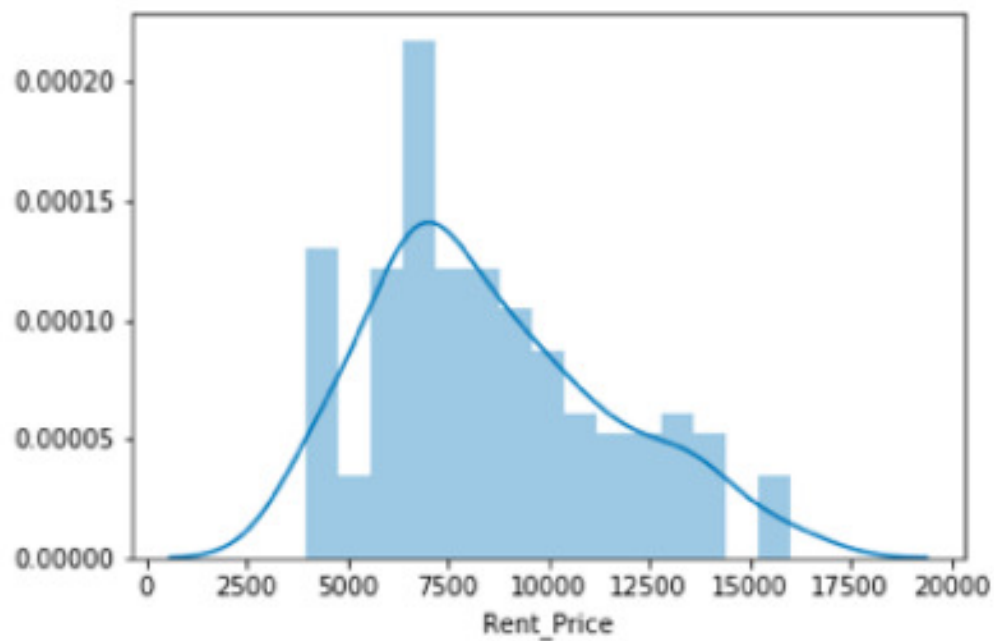
	Address	Area	Price_per_ft2	Rooms	Area-ft2	Rent_Price	Lat	Long
139	200 East 72nd Street	Rental in Lenox Hill	5.15	3.0	1700	8750	40.769465	-73.960339
140	50 Murray Street	No fee rental in Tribeca	7.11	2.0	1223	8700	40.714051	-74.009608
141	300 East 56th Street	No fee rental in Midtown East	3.87	3.0	2100	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 Manhattan Apartments

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

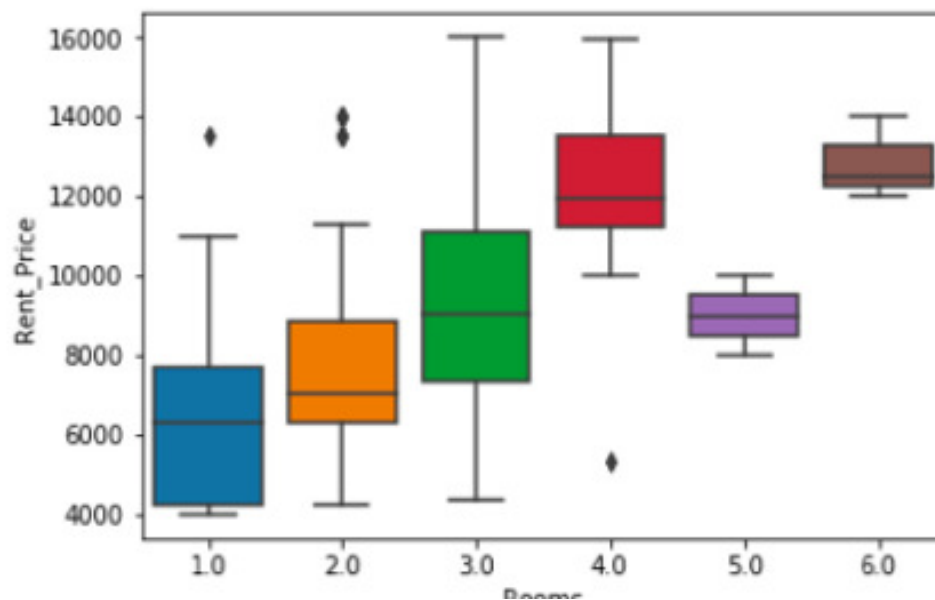


```
<matplotlib.axes._subplots.AxesSubplot at 0x1a25dd8400>
```

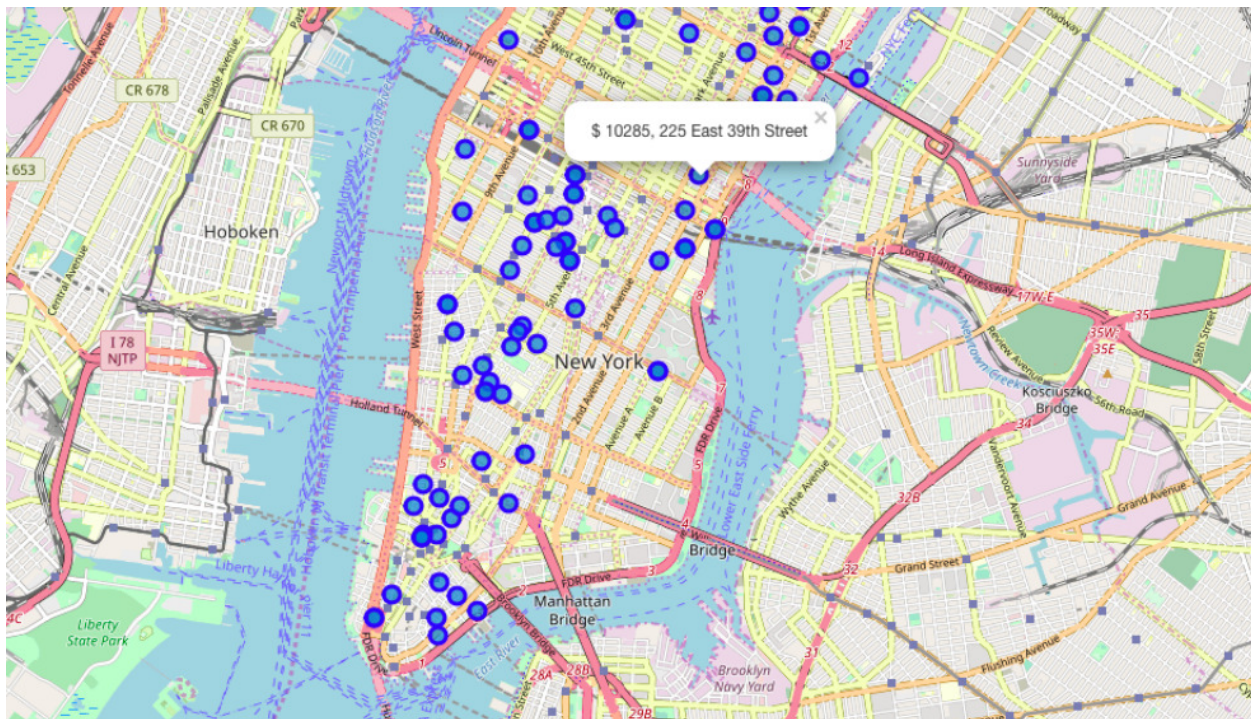


```
sns.boxplot(x='Rooms', y='Rent_Price', data=mh_rent)
```

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



Apartments for rent in Manhattan:



Manhattan Subway station geodata:

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(drop=True)
mhsub1.shape

(22, 4)

```

```

: mhsub1.tail()

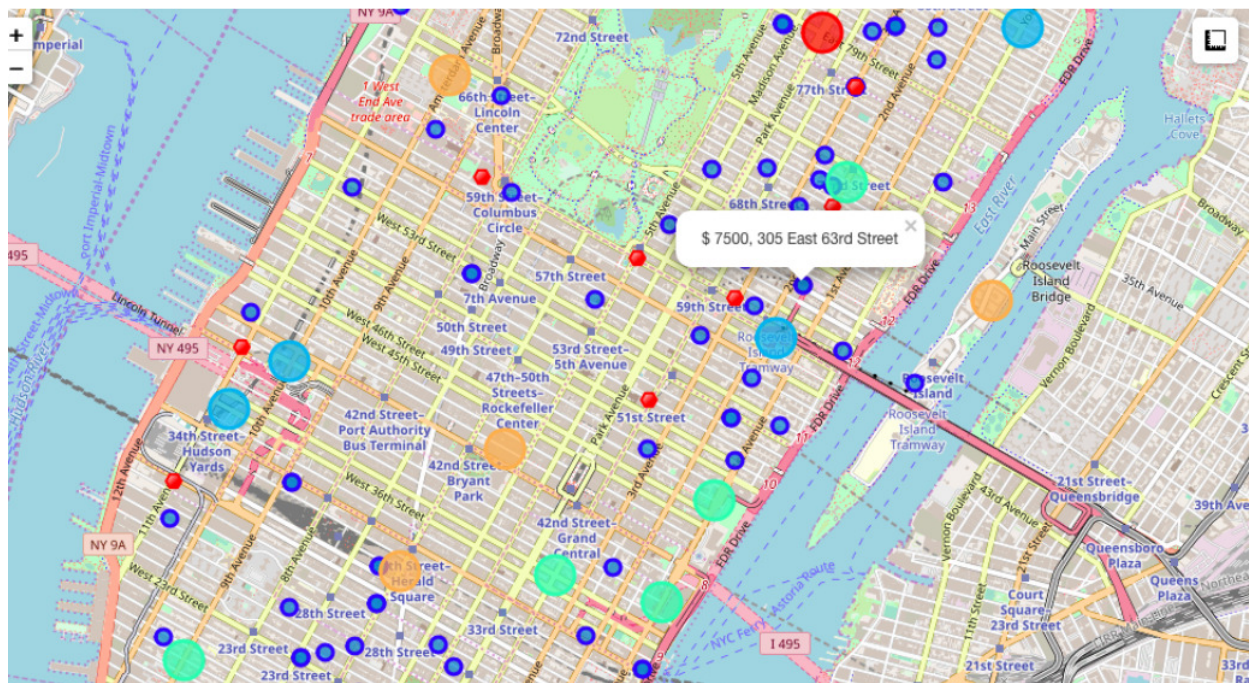
```

	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

**Apartments for Rent Shown in Blue and
Subway Stations are shown in Red**



**Apartment is to be selected in a way that,
The rent is near about equal to \$7000 in my budget.**



Based on the analysis we have done, the following is the result we have got:

Plot to show the location of final arcade that we are suggesting

Map

5. Discussions and further Improvements

- In the Four Square API, I have queried the Venues of a locality by specifying the LIMIT and Radius of our choice. We have chosen less LIMIT as the number of API calls that can be done using a free account in Four Square are less. I can increase the limit for more accurate results.
- In the venue categories we are choosing only few out of 2000 that are available to give weights and identify the best cluster. Hence, assigning weights must be done relatively for each category and then considering more number of venue categories would actually yield better output.

6. Conclusions

Based on the given constraints, a apartment on rent in Manhattan can be taken in a place closer to Murray Hills , which will take your rent less than \$7000 and all the basic needs can be fulfilled due to near metro station.

Resources:

- List of all the cities in New York with apartment rent per apartment
: <http://www.rentmanhattan.com/index.cfm?page=search&state=results>
- List of all the apartments in Manhattan with rent
: https://www.nestpick.com/search?city=new-york&page=1&order=relevance&district=manhattan&gclid=CjwKCAiAjNjgBRAGeiwAGLI2hkP3A-cPxjZYkURqQEswQK2jKQEpv_MvKcrIhRWRzNkc_r-fGi0lxoCA7cQAvD_BwE&type=apartment&display=list
- Four Square API : <https://foursquare.com>