Coursera Capstone Project

The Battle of Neighborhoods -Final Report (Week 1 & week 2)

Contents

1.Introduction Section:

Discussion of the business problem and the audience who would be interested in this project.

1.1 Scenario and Background

I am a IT Professional residing in Leeds, uk. I currently live within walking distance to many amenities and venues in the area, such as parks, international restaurants, cafes, gyms, food shops and entertainment.

I have been offered a great opportunity to work in Manhattan, NY. Although, I am very excited about it, I am a bit stress toward the process to secure a comparable place to live in Manhattan. Therefore, I decided to apply the learned skills during the Coursera course to explore ways to make sure my decision is factual and rewarding. Of course, there are alternatives to achieve the answer using available Google and Social media tools, but it rewarding doing it myself with learned tools.

In order to make a comparison and evaluation of the rental options in Manhattan NY, I must set some basis, therefore the apartment in Manhattan must meet the following demands:

- 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 like current residence
- desirable to have venues such as coffee shops, restaurants Asian Thai, wine stores, gym and food shops
- as a reference, I have included a map of venues near current residence in Leeds.

1.2 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. The data required to resolve this challenge is described in the following section 2, below.

1.3 Interested Audience:

I believe this is a relevant challenge with valid questions for anyone moving to other large city in US, EU 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 Section:

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

Required Data:

2.1 Data of Current Situation:

I Currently reside in the neighbourhood of 'City centre' in Leeds. I use Foursquare to identify the venues around the area of residence which are then shown in the Singapore map shown in methodology and execution in section 3.0. It serves as a reference for comparison with the desired future location in Manhattan NY

2.2 In order to make a good choice of a similar apartment in Manhattan NY, the following data is required:

- List/Information on neighbourhood's form Manhattan with their Geodata (latitude and longitude.
- List/Information about the subway metro stations in Manhattan with geodata.
- Listed apartments for rent in Manhattan area with descriptions (how many beds, price, location, address)
- Venues and amenities in the Manhattan neighbourhoods (e.g. top 10)

2.3 sources and manipulation:

- The list of Manhattan neighbourhoods is worked out during LAb exercise during the course. A csv file was created which will be read in order to create a data frame 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 neighbourhoods and mapping will be shown, however. An algorythm was used to determine the geodata from Nominatim . The actual algorythm coding may be shown in 'markdown' mode because it takes time to run.

  mh_neigh_data.tail():

	Borough	Neighbourhood	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 complemented 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 algorythm 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
•	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
-

A list of places for rent was collected by web-browsing real estate companies in Manhattan :http://www.rntmanhattan.com/index.cfm?page=search&state=results 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 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 algorythm was used to create all the geodata using Nominatim, as shown in section 3.0. The actual algorythm 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 algorythms 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 neighbourhoods in general and a cluster is created to later be able to search for the venues depending of the location shown.

2.4 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 neighbourhoods 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 ammenities 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.

2.5 Mapping of Data:

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

How data will be used to solve the following problem:

- the cost of rent (per square ft) around a mile radius from each subway metro station.
- what is the area of Manhattan with best rental pricing that meets criteria established?
- distance from work place (Park Ave and 53 rd St) and the tentative future home?
- What are the venues of the two best places to live? How the prices compare?
- How venues distribute among Manhattan neighbourhoods and around metro stations?

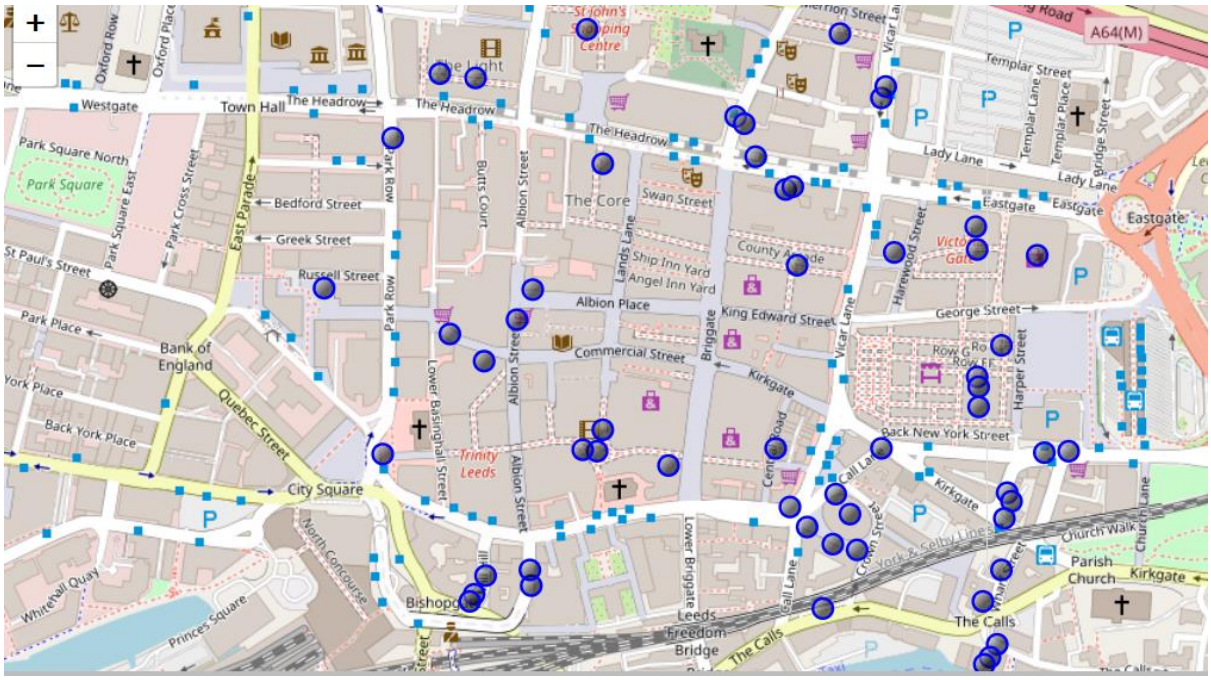
- Are there tradeoffs between size and price and location?
- Any other interesting statistical data findings of the real estate and overall data.

3. Explore Current residence neighbourhoods:

3.1Venues near current city Leeds,UK

	name	categories	lat	lng
0	Trinity Leeds	Shopping Mall	53.796525	-1.543937
1	Victoria Quarter	Shopping Mall	53.798170	-1.540943
2	Headrow House	Bar	53.798837	-1.541118
3	Mrs Athas	Coffee Shop	53.796542	-1.541268
4	Laynes Espresso	Coffee Shop	53.795323	-1.544939
5	Bundobust	Bar	53.795415	-1.545622
6	Friends of Ham	Bar	53.795470	-1.544944
7	The LEGO Store	Toy / Game Store	53.796535	-1.544166
8	200 Degrees Coffee	Coffee Shop	53.797560	-1.546153
9	Ox Club	Restaurant	53.798862	-1.541013

3.2 Map of LEEDS residence place with venues in Neighbourhood - for reference



MANHATTAN NEIGHBORHOODS - DATA AND MAPPING

Cluster neighbourhood data was produced with Foursquare during course lab work. A csv file was produced containing the neighbourhoods around the 40 Boroughs. Now, the csv file is just read for convenience and consolidation of report.

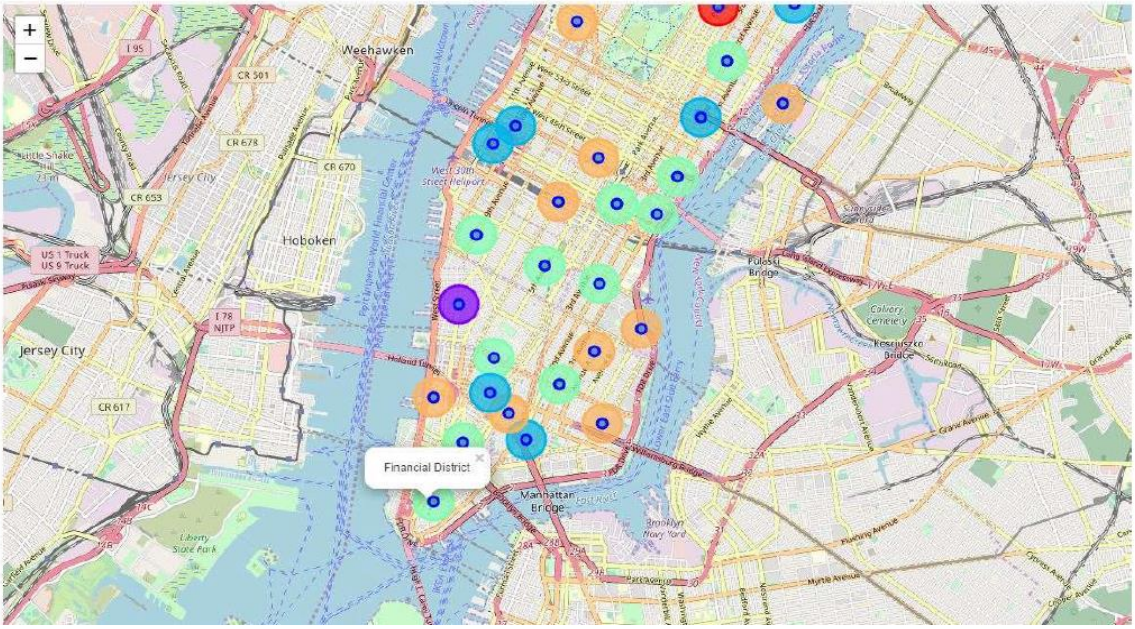
	Borough	Neighbourhood	Latitude	Longitude	Cluster Labels
0	Manhattan	Marble Hill	40.876551	-73.910660	2
1	Manhattan	Chinatown	40.715618	-73.994279	2
2	Manhattan	Washington Heights	40.851903	-73.936900	4
3	Manhattan	Inwood	40.867684	-73.921210	3
4	Manhattan	Hamilton Heights	40.823604	-73.949688	0

Manhattan Borough neighbourhoods - data with top 10 clustered venues

	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	2	Coffee Shop	Discount Store	Yoga Studio	Steakhouse	Supplement Shop	Tennis Stadium	Shoe Store	Gym	Bank	Seafood Restaurant
1	Manhattan	Chinatown	40.715618	-73.994279	2	Chinese Restaurant	Cocktail Bar	Dim Sum Restaurant	American Restaurant	Vietnamese Restaurant	Salon / Barbershop	Noodle House	Bakery	Bubble Tea Shop	Ice Cream Shop
2	Manhattan	Washington Heights	40.851903	-73.936900	4	Café	Bakery	Mobile Phone Shop	Pizza Place	Sandwich Place	Park	Gym	Latin American Restaurant	Tapas Restaurant	Mexican Restaurant
3	Manhattan	Inwood	40.867684	-73.921210	3	Mexican Restaurant	Lounge	Pizza Place	Café	Wine Bar	Bakery	American Restaurant	Park	Frozen Yogurt Shop	Spanish Restaurant
4	Manhattan	Hamilton Heights	40.823604	-73.949688	0	Mexican Restaurant	Coffee Shop	Café	Deli / Bodega	Pizza Place	Liquor Store	Indian Restaurant	Sushi Restaurant	Sandwich Place	Yoga Studio

Map of Manhattan neighbourhoods with top 10 clustered venues

popus allow to identify each neighbourhood and the cluster of venues around it in order to proceed to examine in more detail in the next cell:



4.Manhattan places for rent

Several Manhattan real estate webs were web scrapped to collect rental data, as mentioned in section 2.0. The result was summarized in a csv file for direct reading, in order to consolidate the process.

The initial data for 144 apartments did not have the latitude and longitude data (NaN) but the information was established in the following cell using an algorythm and Nominatim.

```
# csv files with rental places with basic data but still wihtout geodata ( latitude and longitude)
# pd.read_csv(' le.csv', header=None, nrows=5)
mh_rent=pd.read_csv('MH_flats_price.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	NaN	NaN
1	East 97th Street	Upper East Side	3.57	3.0	2100	7500	NaN	NaN
2	West 105th Street	Upper West Side	1.89	4.0	2800	5300	NaN	NaN
3	CARMINE ST.	West Village	3.03	2.0	1650	5000	NaN	NaN
4	171 W 23RD ST.	Chelsea	3.45	2.0	1450	5000	NaN	NaN

4.1 geodata (lat,long) for each rental place in Manhattan with Nominatim

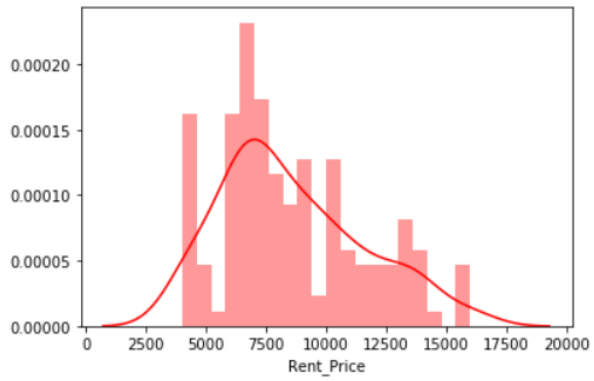
```
mh_rent=pd.read_csv('MH_rent_latlong.csv')
mh_rent.head()
```

	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

4.2 Manhattan apartment rent price statistics

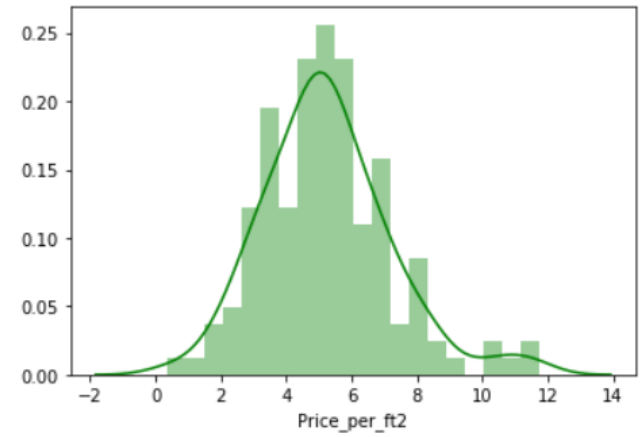
```
import seaborn as sns
sns.distplot(mh_rent['Rent_Price'],bins=20,color='red')
```

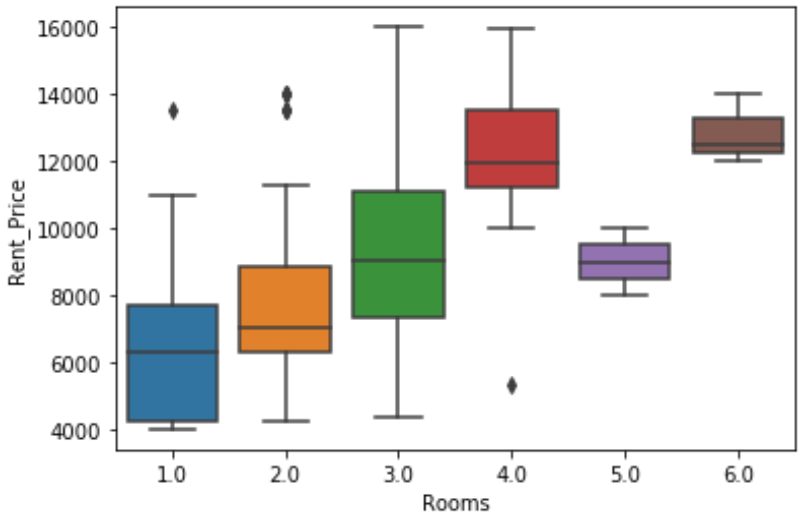
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f324780c0f0>
```




```
import seaborn as sns
sns.distplot(mh_rent['Price_per_ft2'],bins=20,color='green')
```

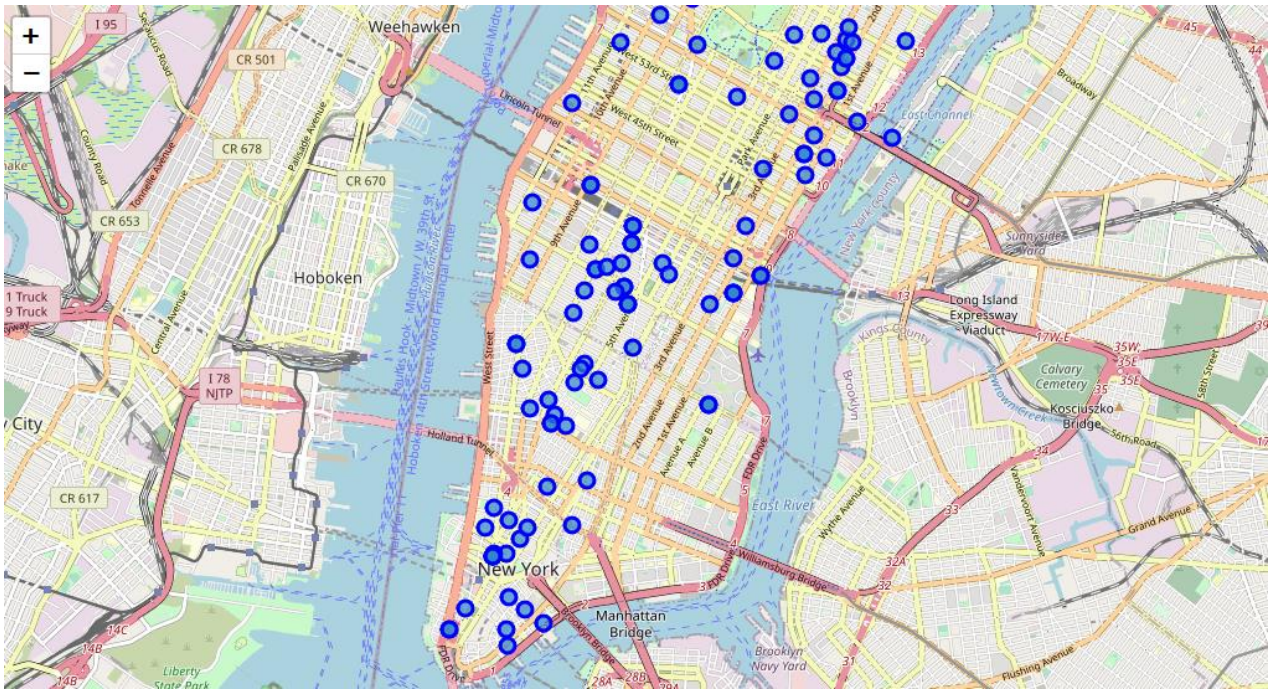
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f3246c83f60>
```



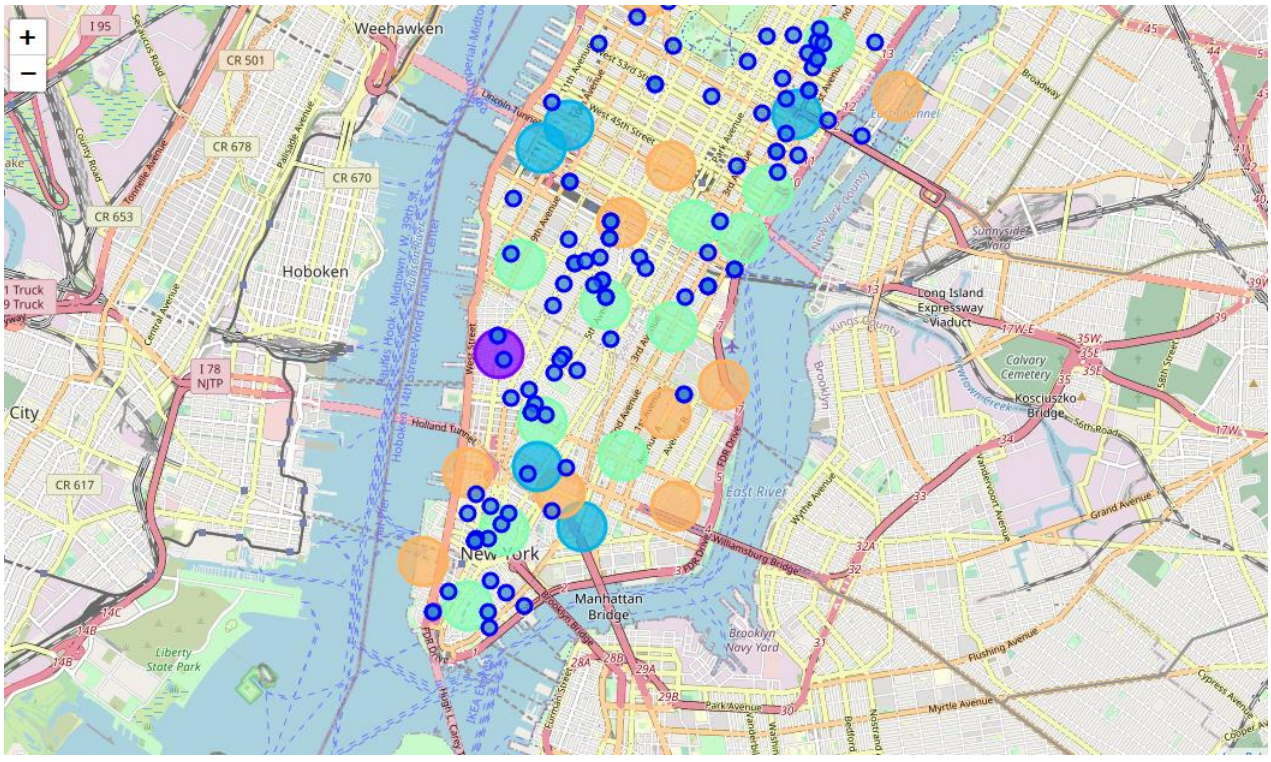


Map of Manhattan apartments for rent

The popups will indicate the address and the monthly price for rent thus making it convenient to select the target appartment with the price condition estipulated (max US7000)



Map of Manhattan showing the places for rent and the cluster of venues



Now one can explore a rental place and its venues in detail.

```
## kk is the cluster number to explore
kk = 3
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == kk, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.sha
```

	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
3	Inwood	Mexican Restaurant	Lounge	Pizza Place	Café	Wine Bar	Bakery	American Restaurant	Park	Frozen Yogurt Shop	Spanish Restaurant
5	Manhattanville	Deli / Bodega	Italian Restaurant	Seafood Restaurant	Mexican Restaurant	Sushi Restaurant	Beer Garden	Coffee Shop	Falafel Restaurant	Bike Trail	Other Nightlife
10	Lenox Hill	Sushi Restaurant	Italian Restaurant	Coffee Shop	Gym / Fitness Center	Pizza Place	Burger Joint	Deli / Bodega	Gym	Sporting Goods Shop	Thai Restaurant
12	Upper West Side	Italian Restaurant	Bar	Bakery	Vegetarian / Vegan Restaurant	Indian Restaurant	Coffee Shop	Cosmetics Shop	Wine Bar	Mexican Restaurant	Sushi Restaurant
16	Murray Hill	Sandwich Place	Hotel	Japanese Restaurant	Gym / Fitness Center	Coffee Shop	Salon / Barbershop	Burger Joint	French Restaurant	Bar	Italian Restaurant
17	Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bakery	Nightclub	Theater	Art Gallery	Seafood Restaurant	American Restaurant	Hotel
18	Greenwich Village	Italian Restaurant	Sushi Restaurant	French Restaurant	Clothing Store	Chinese Restaurant	Café	Indian Restaurant	Bakery	Seafood Restaurant	Electronics Store

We could find an apartment with at the right price and in a location with desirable venues. The next step is to see if it is located near a subway metro station.

5 Mapping Manhattan Subway locations

Manhattan subway metro locations (address) was obtained from web scrapping sites such as Wikipedia, Google and NY Metro Transit.

```
# A csv file summarized the subway station and the addresses for next step to determine geodata
mh=pd.read_csv('NYC_subway_list.csv')
mh.head()
```

	sub_station	sub_address
0	Dyckman Street Subway Station	170 Nagle Ave, New York, NY 10034, USA
1	57 Street Subway Station	New York, NY 10106, USA
2	Broad St	New York, NY 10005, USA
3	175 Street Station	807 W 177th St, New York, NY 10033, USA
4	5 Av and 53 St	New York, NY 10022, USA

	sub_station	sub_address
0	Dyckman Street Subway Station	170 Nagle Ave, New York, NY 10034, USA
1	57 Street Subway Station	New York, NY 10106, USA
2	Broad St	New York, NY 10005, USA
3	175 Street Station	807 W 177th St, New York, NY 10033, USA
4	5 Av and 53 St	New York, NY 10022, USA

Add colums labeled 'lat' and 'long' to be filled with geodata

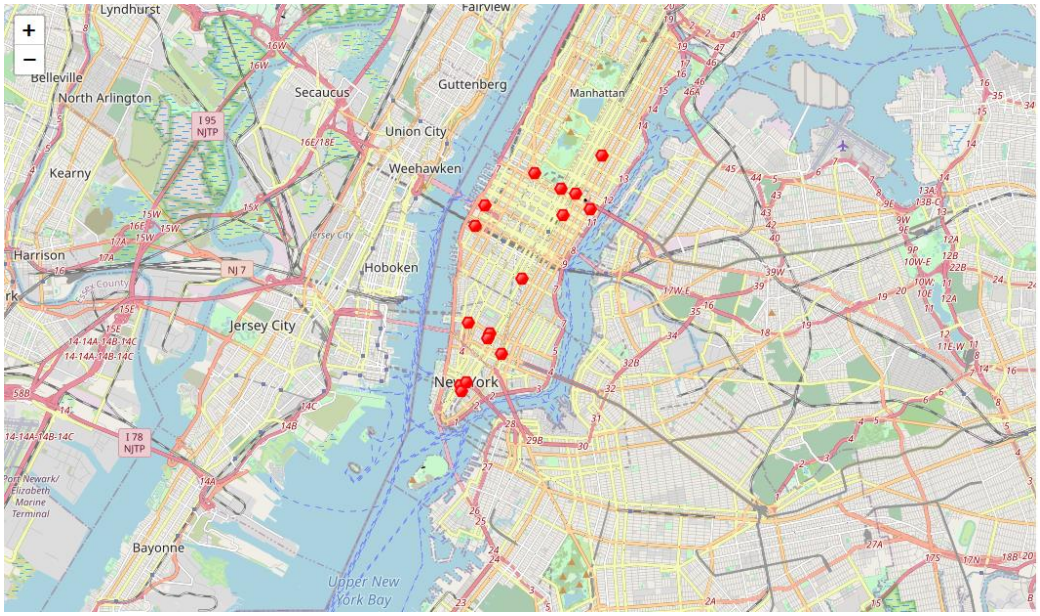
```
#  Add columns 'lat'  and 'long' to mh dataframe - with random temporary numbers to get started
sLength = len(mh['sub_station'])
lat = pd.Series(np.random.randn(sLength))
long=pd.Series(np.random.randn(sLength))
mh = mh.assign(lat=lat.values)
mh = mh.assign(long=long.values)
```

```
mh=pd.read_csv('MH_subway.csv')
print(mh.shape)
mh.head()
```

```
(76, 4)
```

	sub_station	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	40.758798	-73.962343
2	Broad St	New York, NY 10005, USA	40.712728	-74.006015
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	40.758798	-73.962343

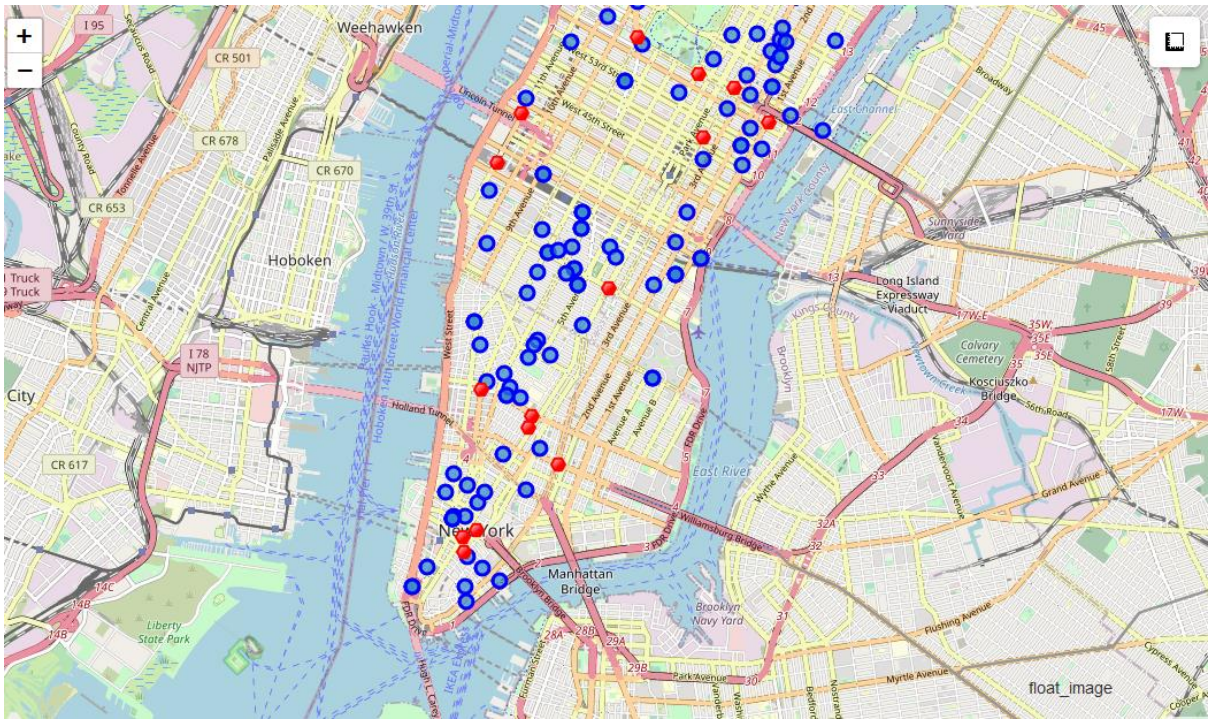
MAP of Manhattan showing the location of subway stations



Map of Manhattan showing places for rent and the subway locations nearby

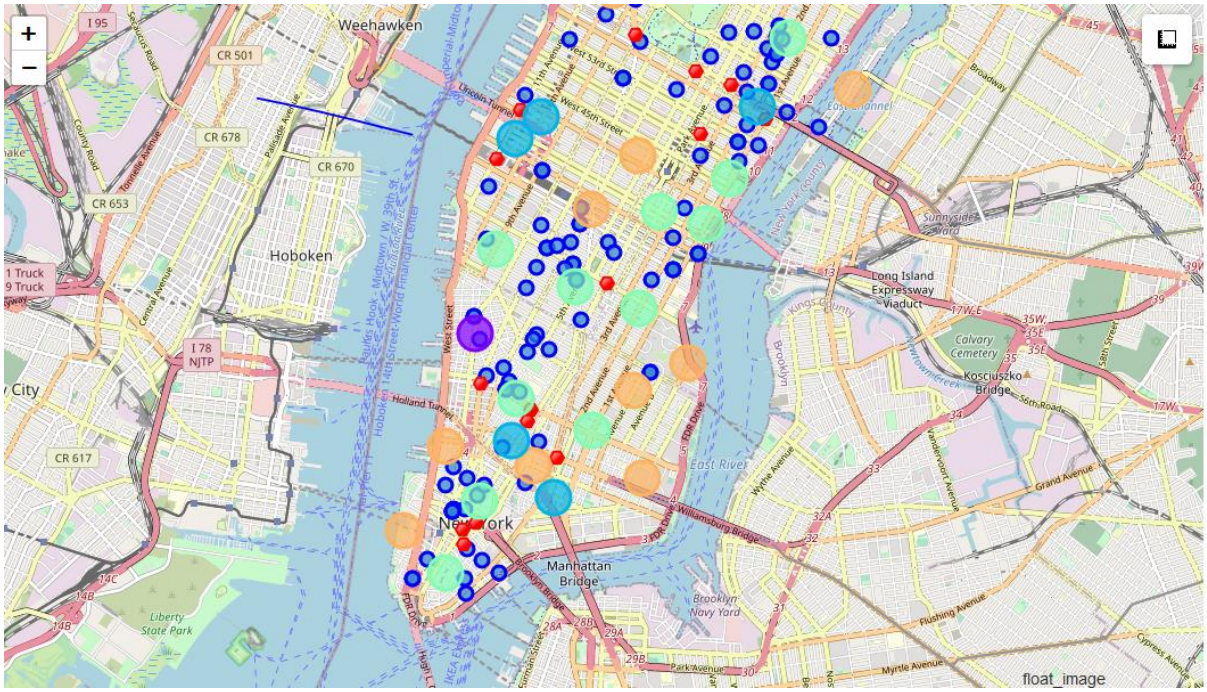
Now, we can visualize the desirable rental places and their nearest subway station. Popups display rental address and monthly rental price and the subway station name.

the icon in the top-right corner is a "ruler" that allows to measure the distance from a rental place to an specific subway station.



CONSOLIDATE MAP: Map of Manhattan with rental places, subway locations and cluster of venues

Red dots are Subway stations, Blue dots are apartments available for rent, Bubbles are the clusters of venues.



The above consolidate map was used to explore options.

After examining, I have chosen two locations that meet the requirements which will assess to make a choice.

- Apartment 1: 305 East 63rd Street in the Sutton Place Neighbourhood and near 'subway 59th Street' station, Cluster # 2 Monthly rent: 7500 Dollars
- Apartment 2: 19 Dutch Street in the Financial District Neighbourhood and near 'Fulton Street Subway' station, Cluster # 3 Monthly rent: 6935 Dollars

Venues for Apartment 1 - 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.sha
```

	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

Venues for Apartment 2 - Cluster 3

```
## kk is the cluster number to explore
kk = 3
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == kk, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.sha
```

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

Result:

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

Apartment 1 rent cost is US7500 slightly above the US7000 budget. Apt 1 is located 400 meters from subway station at 59th Street and workplace (Park Ave and 53rd) is another 600 meters way. I can walk to workplace and use subway for other places around. Venues for this apt are as of Cluster 2 and it is 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.

THANK YOU