Rental prices comparison (Jurong West vs Manhattan)

Than Lwin Aung

6th December 2019

Content

1. Introduction Section

1.1 Background

A person currently studying data science who was used to stay in Singapore for 8 years within walking distance to "Boon Lay MRT station as public transport to go for work. Now, considering to apply the skills and knowledge learnt in IBM Data Science courses to explore the rental price of Manhattan, New York compares to its of Downtown, Singapore.

1.2 Problem to be resolved

The following criteria are set to compare the rental price of Manhattan, NY and Downtown, SG: Apartment with min 2 bedrooms with monthly rent not to exceed US$4,500/month Unit located within walking distance (<=0.5 mile, 0.8 km) from a subway metro station in Manhattan

1.3 Interested Audience

This is aimed to help the migrant workers who are interested to work in Manhattan, NY

2. Data Section

2.1 Data of Current Situation

I currently reside in the Mandalay, Myanmar. However, the venues around Jurong West, SG will be fetched using Foursquare to show in the Singapore map 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 Data Required to resolve the problem

In order to make a good choice of a similar apartment in Manhattan NY, the following data is required: List/Information on neighborhoods 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 neighborhoods (e.g. top 10)

2.3 Sources and manipulation

The list of Manhattan neighborhoods 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 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 complied 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 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
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

A list of places for rent was collected by web-browsing real estate companies in Manhattan : http://www.rentmanhattan.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 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 0.8 km radius. Foursquare is used to find the avenues at Manhattan neighborhoods 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 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 0.5 mile (0.8 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 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 and prepared for analysis. The tools described are used here and the Notebook cells indicates the execution of steps.

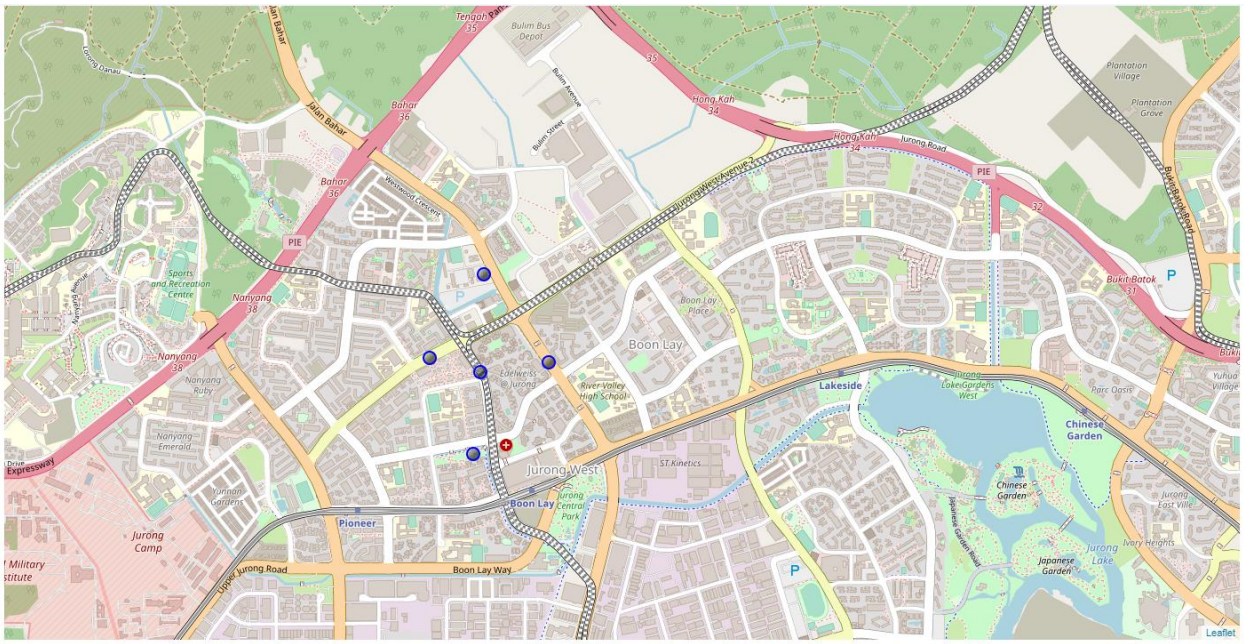
The analysis and the strategy:

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 to answer 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 half mile radius from each subway metro station? What is the area of Manhattan with best rental pricing that meets criteria established? 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.

METHODOLOY EXECUTION - Mapping Data

Singapore Map - venues and residence in neighborhood around Boon Lay MRT station, Jurong West, SG for comparison to future Manhattan renting place.

Map of Singapore residence place with venues in Neighborhood - for reference



MANHATTAN NEIGHBORHOODS - DATA AND MAPPING

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

```
manhattan_data = pd.read_csv(body)
manhattan_data.head()
```

Out[11]:

	Borough	Neighborhood	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

In [12]:
```
manhattan_data.tail()
```

Out[12]:

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels
35	Manhattan	Turtle Bay	40.752042	-73.967708	3
36	Manhattan	Tudor City	40.746917	-73.971219	3
37	Manhattan	Stuyvesant Town	40.731000	-73.974052	4
38	Manhattan	Flatiron	40.739673	-73.990947	3
39	Manhattan	Hudson Yards	40.756658	-74.000111	2

Manhattan Borough neighborhoods - data with top 10 clustered venues

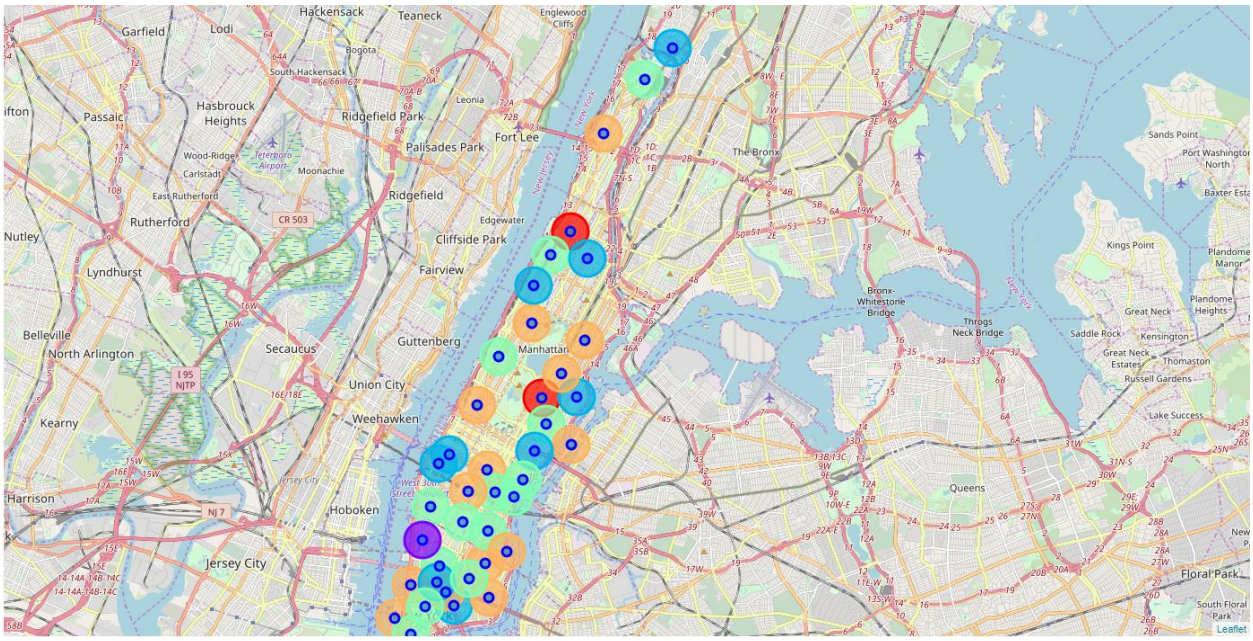
```
manhattan_merged = pd.read_csv(body)
manhattan_merged.head()
```

Out[13]:

	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 neighborhoods with top 10 clustered venues

popups allow to identify each neighborhood and the cluster of venues around it in order to proceed to examine in more detail in the next cell.



Examine a particular Cluster - print venues

After examining several cluster data, I concluded that cluster # 2 resembles closer the Singapore place, therefore providing guidance as to where to look for the future apartment. Assign a value to 'kk' to explore a given cluster.

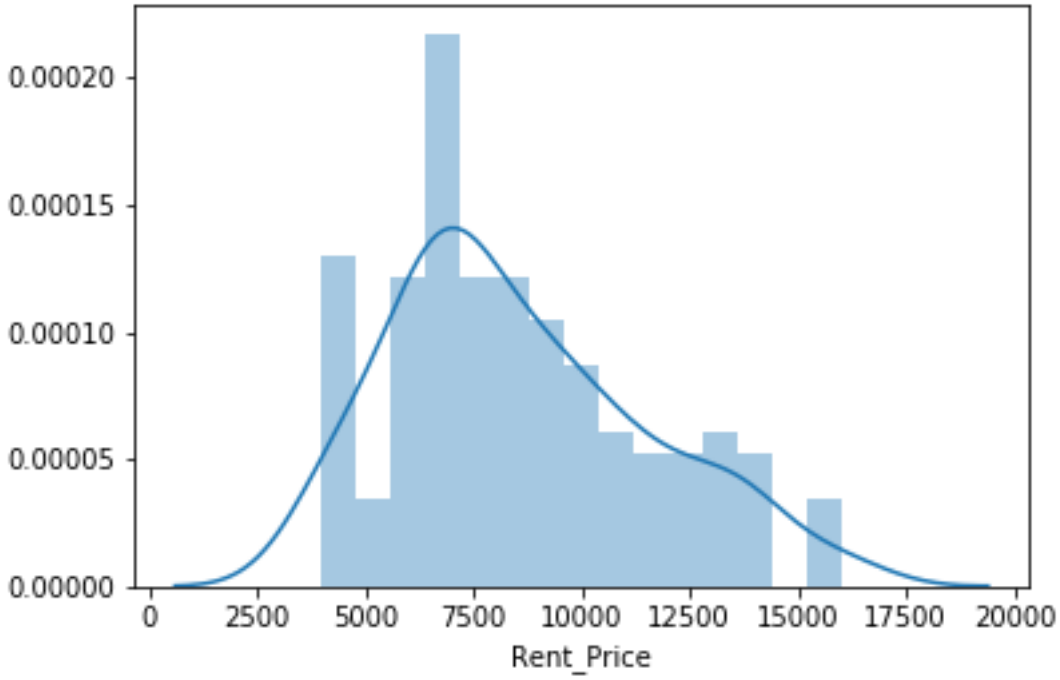
Map of Manhattan places for rent

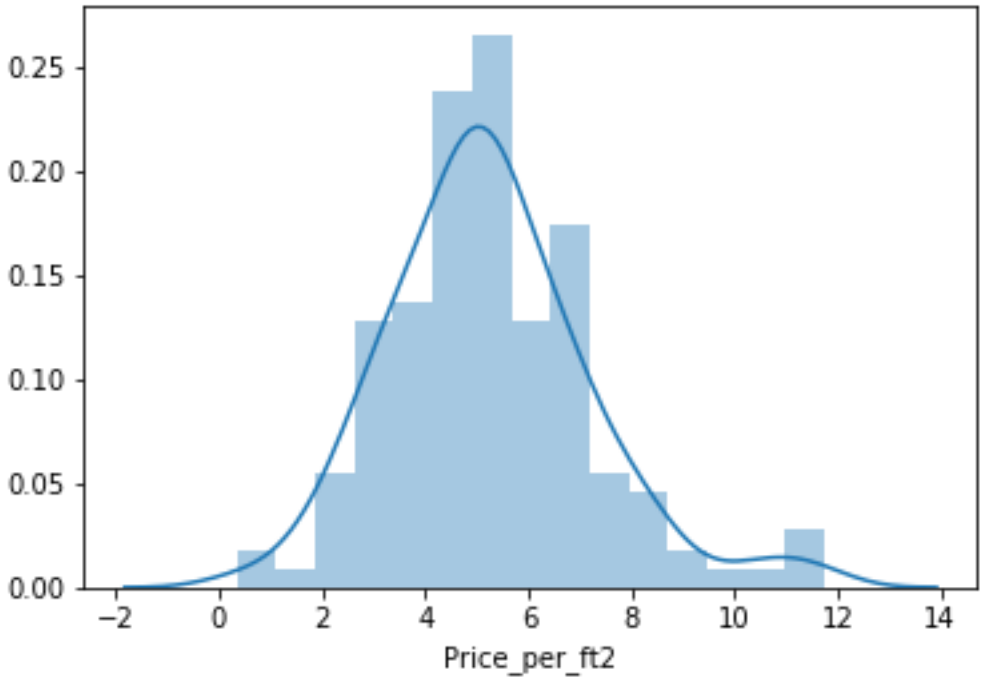
Several Manhattan real estate webs were web scrapped to collect rental data, as mentioned in section 2.0 . The result was summarized in a csv file for direct reading, in order to consolidate the process. The initial data for 144 apartments did not have the latitude and longitude data (NaN) but the information was established in the following cell using an algorithm and Nominatim.

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

Data was stored in a csv file for simplification report purposes and saving code processing time in future. This coding section was 'markdown' for the report because its execution takes few minutes. Therefore, the csv saved will be just read directly in the following cell. for n in range(len(mh_rent)): address= mh_rent['Address'][n] address=(mh_rent['Address'][n]+ ' , '+' Manhattan NY ') geolocator = Nominatim() location = geolocator.geocode(address) latitude = location.latitude longitude = location.longitude mh_rent['Lat'][n]=latitude mh_rent['Long'][n]=longitude

Manhattan apartment rent price statistics





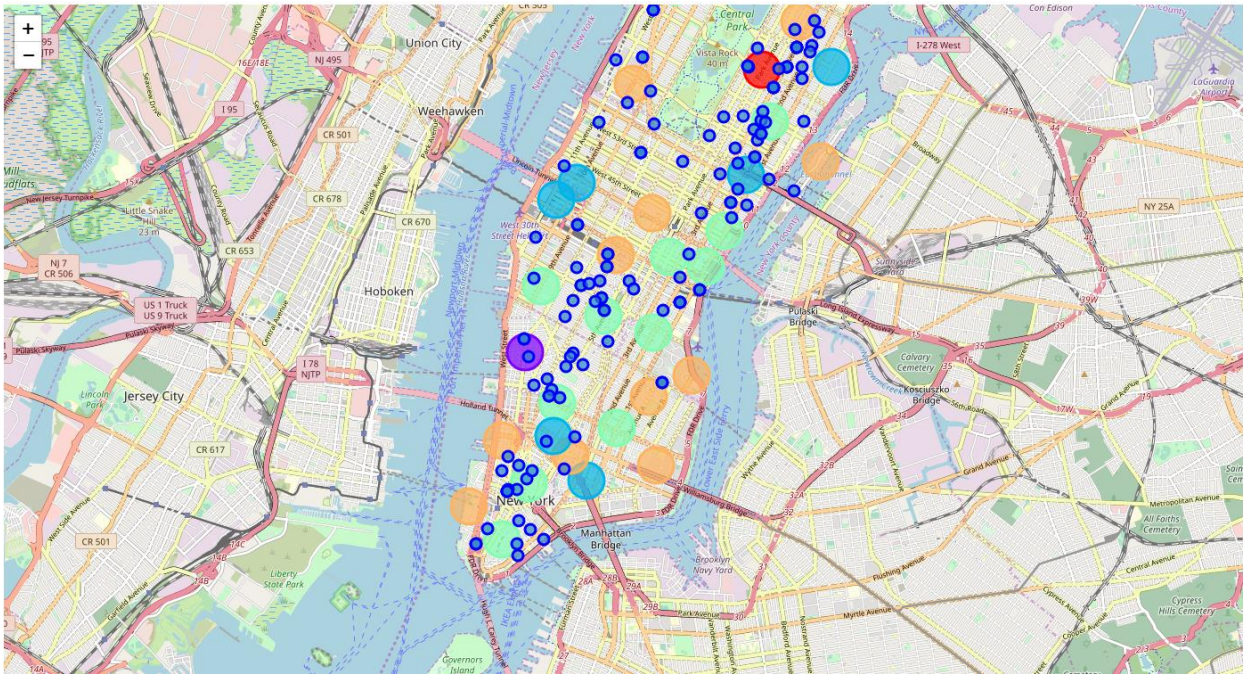
Luckily, there are apartment with monthly rental of USD4500 available with 2 bedrooms according to data points.

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 apartment with the price condition stipulated (max US4500).

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

Now, one can point to a rental place for price and address location information while knowing the cluster venues around it. This is an insightful way to explore rental possibilities.



Now can explore a particular rental place and its venues in detail

In the map above, examination of apartments with rental place around USD4500/month is straightforward while knowing the venues around it. 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, in next cells work.

Manhattan subway metro locations (address) was obtained from web scrapping sites such as Wikipedia, Google and NY Metro Transit. For simplification, a csv file was produced from the 'numbers' (Apple excel) so that the reading of this file is the starting point here. The geodata will be obtain via Nominatim using the algorithm below.

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. Notice that 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.

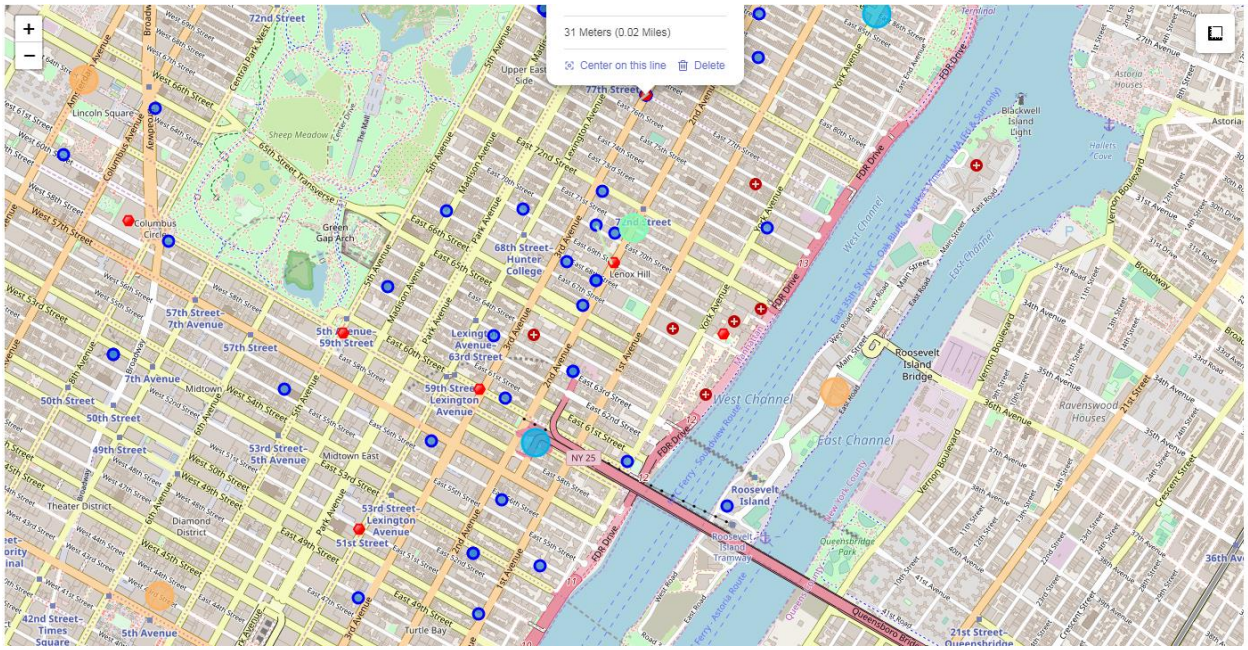
4. Results

ONE CONSOLIDATE MAP

Let's consolidate all the required information to make the apartment selection in one 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.



Problem Resolution - Select the apartment for rent

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: 211 East 70th Street in the Lenox Hill Neighborhood and near '72th Street Subway' station, Cluster # 3 Monthly rent: USD4,250

Apartment 2: East 77th Street in the Upper East side Neighborhood and near 'subway 77th Street' station, Cluster # 0 Monthly rent: USD4,000

Apartment Selection

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 (Apt 1) - rent cost is US4250 lower than the USD4,500 budget. Apt 1 is located 159 meters from subway station at 72nd Street. I can walk to use subway for other places around. Venues for this apt are as of Cluster 3 and it is located with various restaurant and near the college.

Apartment 2 (Apt 2) - rent cost is US4000 much lower than the USD4,500 budget. Apt 2 is located 31 meters from subway station at 77th Street. Venues for this apt are as of Cluster 0.

Based on experience in Singapore venues, I feel that Cluster 3 type of venues is a closer resemblance to my taste. That means that APARTMENT 1 is a better choice because it is lower than the budget despite slightly higher than Apt 2.

5. Discussion

In general, I am positively impressed with the overall organization, content and lab works presented during the Coursera IBM Certification Course despite I am struggling on coding. I feel this Capstone project presented a great opportunity to practice and apply the Data Science tools and methodologies learned. I have tried my best to create a good project that can be presented as an example to show my learning and hardworking. I feel I have acquired a good starting point to become a professional Data Scientist and I will continue exploring to creating examples of practical cases.

6. Conclusions

The topics covered in this course are very well worthy of appreciation. I struggled to finish this final report due to weakness in coding, however, finally made up mind to finish it off. The project really motivates me to understand more on Data Science not necessary to know more coding and what it brings to through consolidation of information to analyse and make a decision. I am full of confident that Data Science will be applied in several business fields.