PROJECT REPORT: BATTLE OF NEIGHBORHOODSṢ
REVANTH.MADAMALA

Content

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.
- 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

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.

Results section

- Discussion of the results and how they help to make a decision.

Discussion section

- Elaboration and discussion on any observations and/or recommendations for improvement.

Conclusion section

- Decision taken and Report Conclusion.

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 data scientist currently residing in INdia. I currently live In Hyderabad therefore I have access to good public transportation to work. Likewise, I enjoy many amenities in the neighborhood , such as international cuisine restaurants, cafes, 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.

1.2 Problem to be resolved:

The challenge to resolve is being able to find a rental apartment unit in Manhattan NY that offers similar characteristics and benefits to my current situation. Therefore, in order to set a basis for comparison, I want to find a rental unit subject to the following conditions:

- Apartment with min 2 bedrooms with monthly rent not to exceed US$7000/month
- Unit located within walking distance (<=1.0 mile, 1.6 km) from a subway metro station in Manhattan
- Area with amenities and venues similar to the ones described for current location (See item 2.1)

1.3 Interested Audience I believe this is a relevant project for a person or entity considering moving to a major city in Europe, US or Asia, since the approach and methodologies used here are applicable in all cases. The use of FourSquare data and mapping techniques combined with data analysis will help resolve the key questions arisen. Lastly, this project is a good practical case toward the development of Data Science skills.

2. Data Section:

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

2.1 Data of Current Situation

I Currently reside in the neighborhood of Srinagar colony,Hyderabad,India. I use Foursquare to identify the venues around the area of residence which are then shown in the Hyderabad 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 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 from 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 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.

	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
190 Street Subway Station Bennett Ave, New York, NY 10040, USA	40.858113	-73.932983
59 St-Lexington Av Station E 60th St, New York, NY 10065, USA	40.762259	-73.966271
57 Street Station New York, NY 10019, United States	40.764250	-73.954525
14 Street / 8 Ave 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-fGi0IxoCA7cQAvD_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 1.6 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 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 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 pop ups labels allow quick identification of location, price and features, 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 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 workplace (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.

|	name	categories	lat	lng
0	Naturals Ice Cream	Ice Cream Shop	17.428536	78.439620
1	Park Hyatt Hyderabad	Hotel	17.424764	78.429600
2	Big Byte	Bakery	17.430177	78.438352
3	Q-Mart	Department Store	17.426681	78.438616
4	Poorna Tiffin Center	Breakfast Spot	17.433254	78.426191
5	Testa Rossa Caffe	Coffee Shop	17.424747	78.438038
6	Creamstone	Dessert Shop	17.432899	78.429327
7	Harley Davidson	Motorcycle Shop	17.425133	78.433314
8	Tre Forni	Italian Restaurant	17.424806	78.429600
9	WOFL	Breakfast Spot	17.424944	78.438148
10	Deja Vu	Hookah Bar	17.426643	78.435349
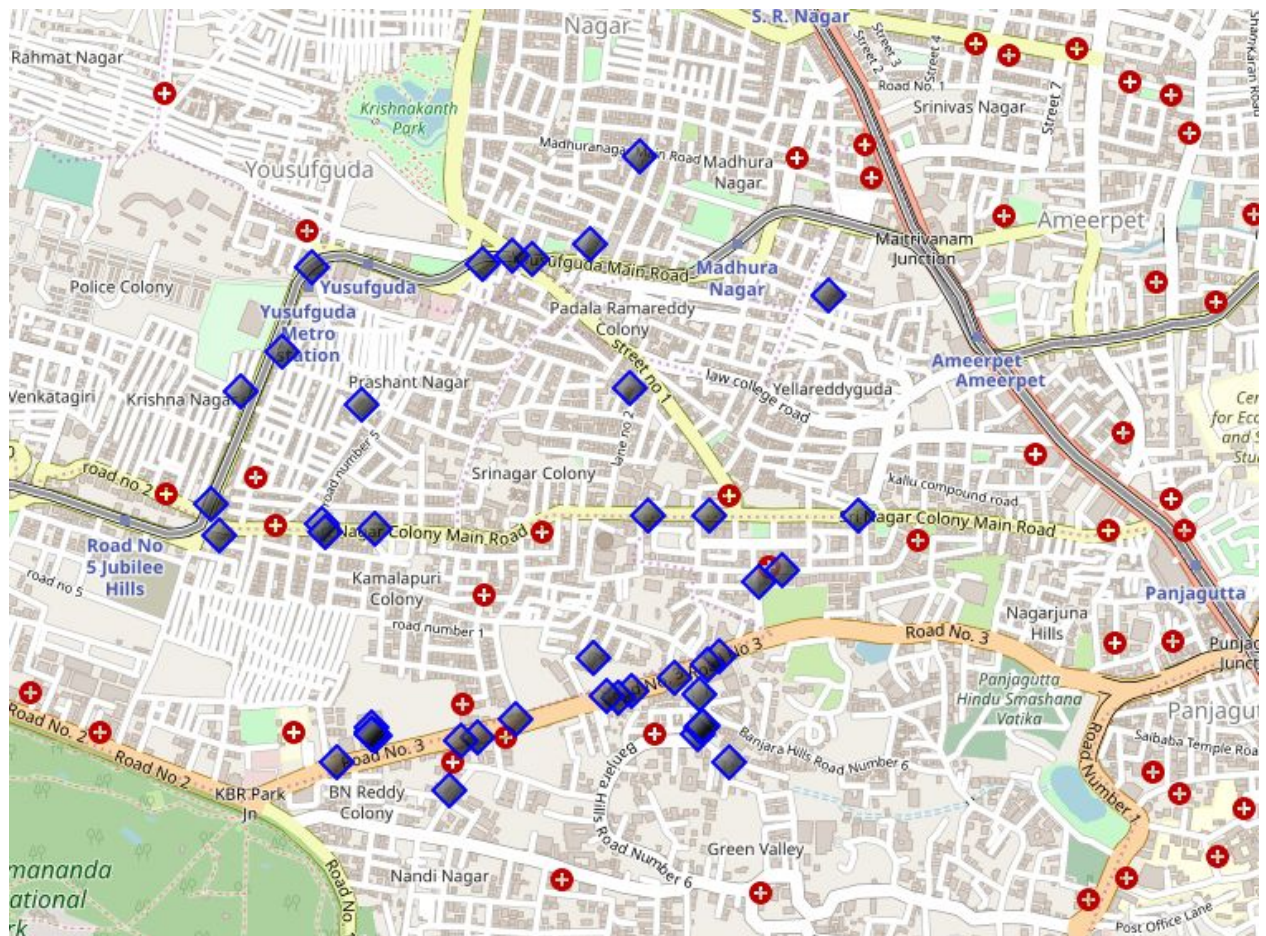
MANHATTAN NEIGHBORHOODS - DATA AND MAPPING Cluster neighborhood data was produced with Foursquare during course lab work. A csv file was produced containing the neighborhoods around the 40 Boroughs. Now, the csv file is just read for convenience and consolidation of report.

	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
5	Manhattan	Manhattanville	40.816934	-73.957385	3
6	Manhattan	Central Harlem	40.815976	-73.943211	2
7	Manhattan	East Harlem	40.792249	-73.944182	4
8	Manhattan	Upper East Side	40.775639	-73.960508	0
9	Manhattan	Yorkville	40.775930	-73.947118	2

Manhattan Borough neighborhoods - 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
0	Manhattan	Marble Hill	40.876551	-73.910660	2	Coffee Shop	Discount Store	Yoga Studio	Steakhouse	Supplement Shop
1	Manhattan	Chinatown	40.715618	-73.994279	2	Chinese Restaurant	Cocktail Bar	Dim Sum Restaurant	American Restaurant	Vietnamese Restaurant
2	Manhattan	Washington Heights	40.851903	-73.936900	4	Café	Bakery	Mobile Phone Shop	Pizza Place	Sandwich Place
3	Manhattan	Inwood	40.867684	-73.921210	3	Mexican Restaurant	Lounge	Pizza Place	Café	Wine Bar
4	Manhattan	Hamilton Heights	40.823604	-73.949688	0	Mexican Restaurant	Coffee Shop	Café	Deli / Bodega	Pizza Place

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**
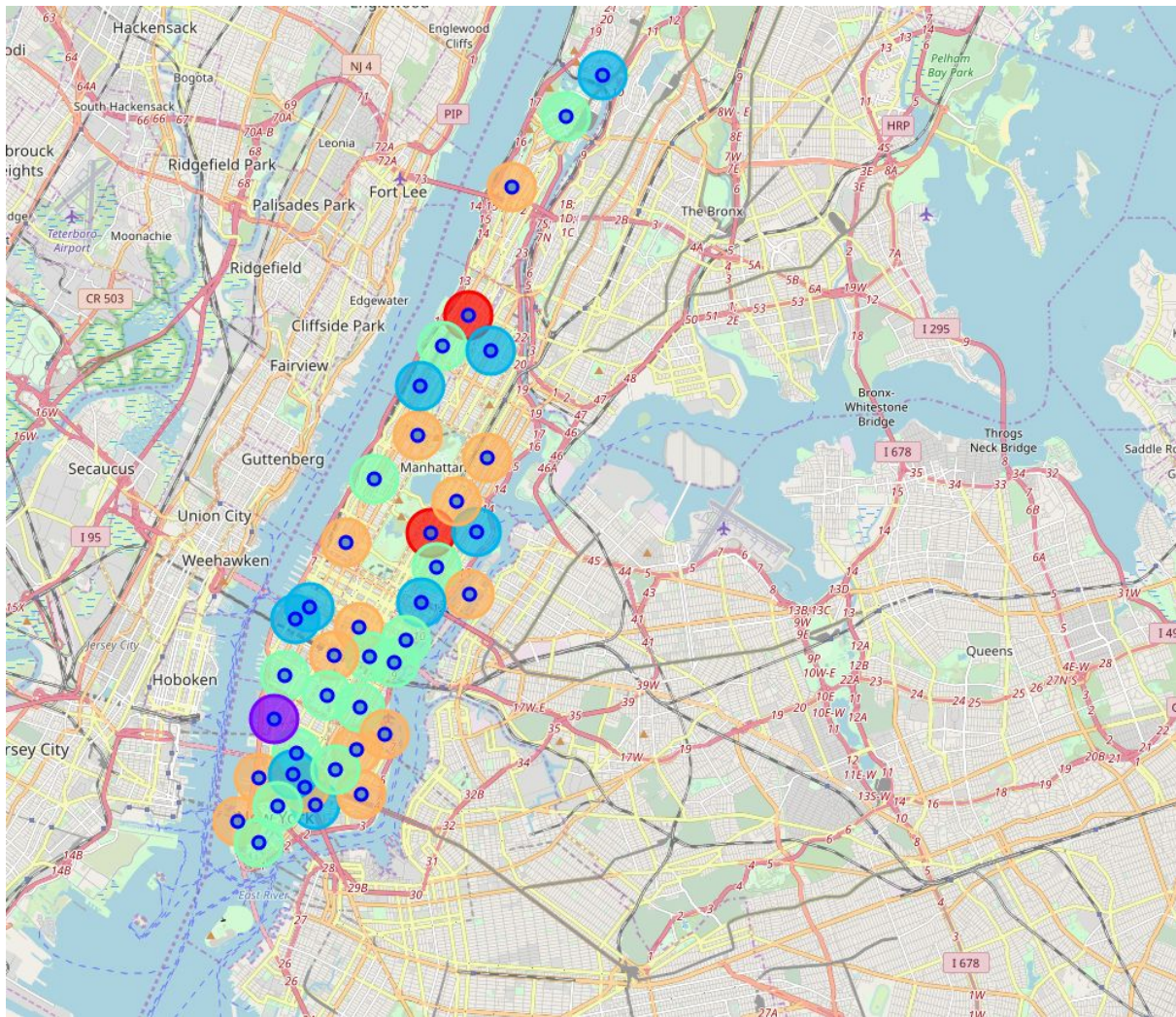
Examine a paticular Cluster - print venues

After examining several cluster data , I concluded that cluster # 2 resembles closer the
**Hyderabad place, therefore providing guidance as to where to look for the future**
apartment

	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
0	Marble Hill	Coffee Shop	Discount Store	Yoga Studio	Steakhouse	Supplement Shop	Tennis Stadium
1	Chinatown	Chinese Restaurant	Cocktail Bar	Dim Sum Restaurant	American Restaurant	Vietnamese Restaurant	Salon / Barbershop
6	Central Harlem	African Restaurant	Seafood Restaurant	French Restaurant	American Restaurant	Cosmetics Shop	Chinese Restaurant
9	Yorkville	Coffee Shop	Gym	Bar	Italian Restaurant	Sushi Restaurant	Pizza Place
14	Clinton	Theater	Italian Restaurant	Coffee Shop	American Restaurant	Gym / Fitness Center	Hotel
23	Soho	Clothing Store	Boutique	Women's Store	Shoe Store	Men's Store	Furniture / Home Store
26	Morningside Heights	Coffee Shop	American Restaurant	Park	Bookstore	Pizza Place	Sandwich Place
34	Sutton Place	Gym / Fitness Center	Italian Restaurant	Furniture / Home Store	Indian Restaurant	Dessert Shop	American Restaurant
39	Hudson Yards	Coffee Shop	Italian Restaurant	Hotel	Theater	American Restaurant	Café

Map of Manhattan places for rent Several Manhattan real estate webs were webscrapped to collect rental data, as mentioned in section 2.0 . The resut was summarized in a csv file for direct reading, in order to consolidate the proces.

The initial data for 144 apartment did not have the latitude and longitude data (NaN) but the information was established in the following cell using an algorythm and Nominatim.
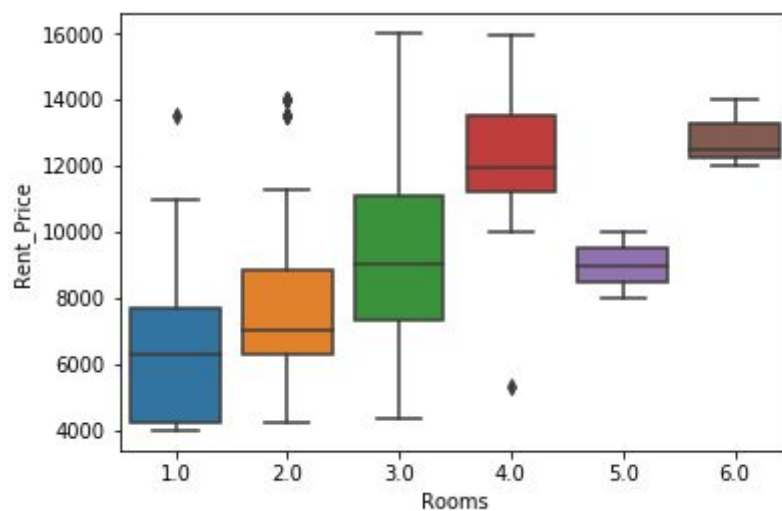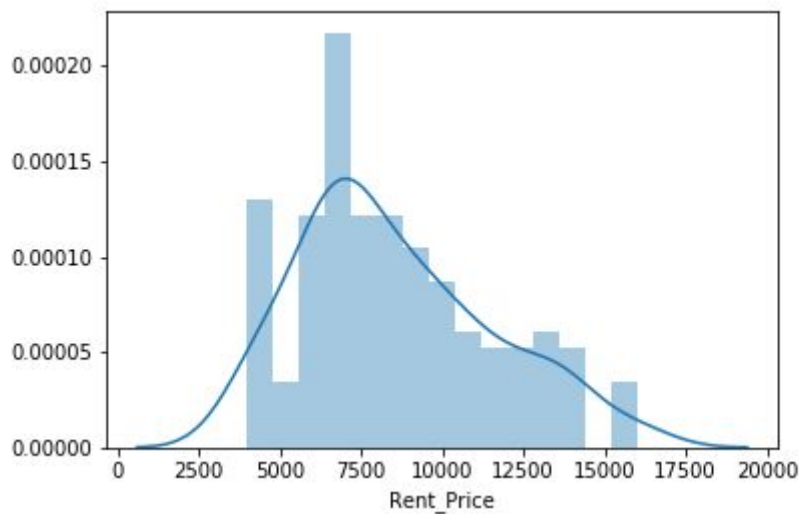
Map of Manhattan apartments for rent The popups will indicate the address and the monthly price for rent thus making it convenient to select the target appartment with the price condition estipulated (max US7000)
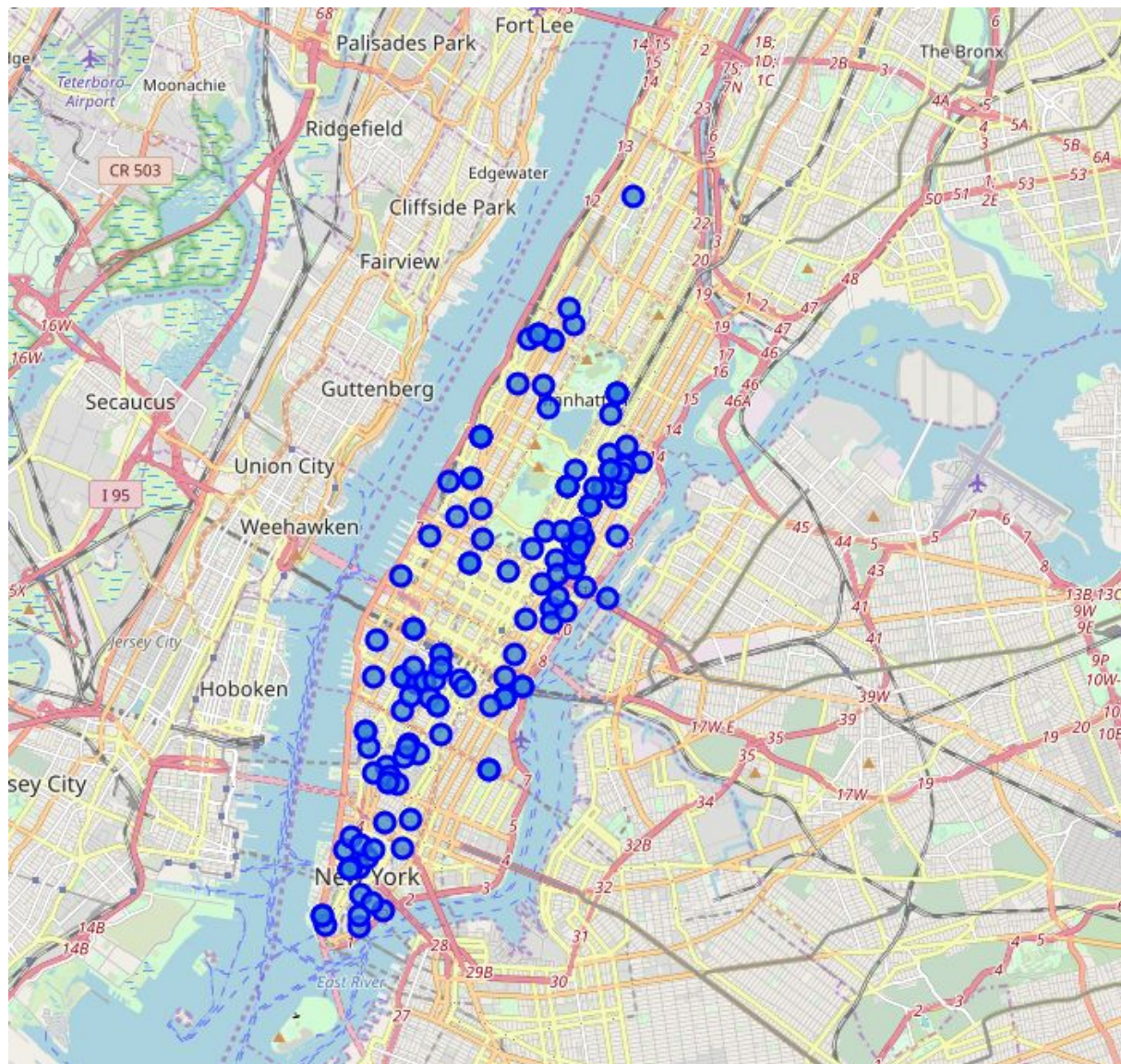
Map of Manhattan showing the places for rent and the cluster of venues

Now, one can 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 possibilites
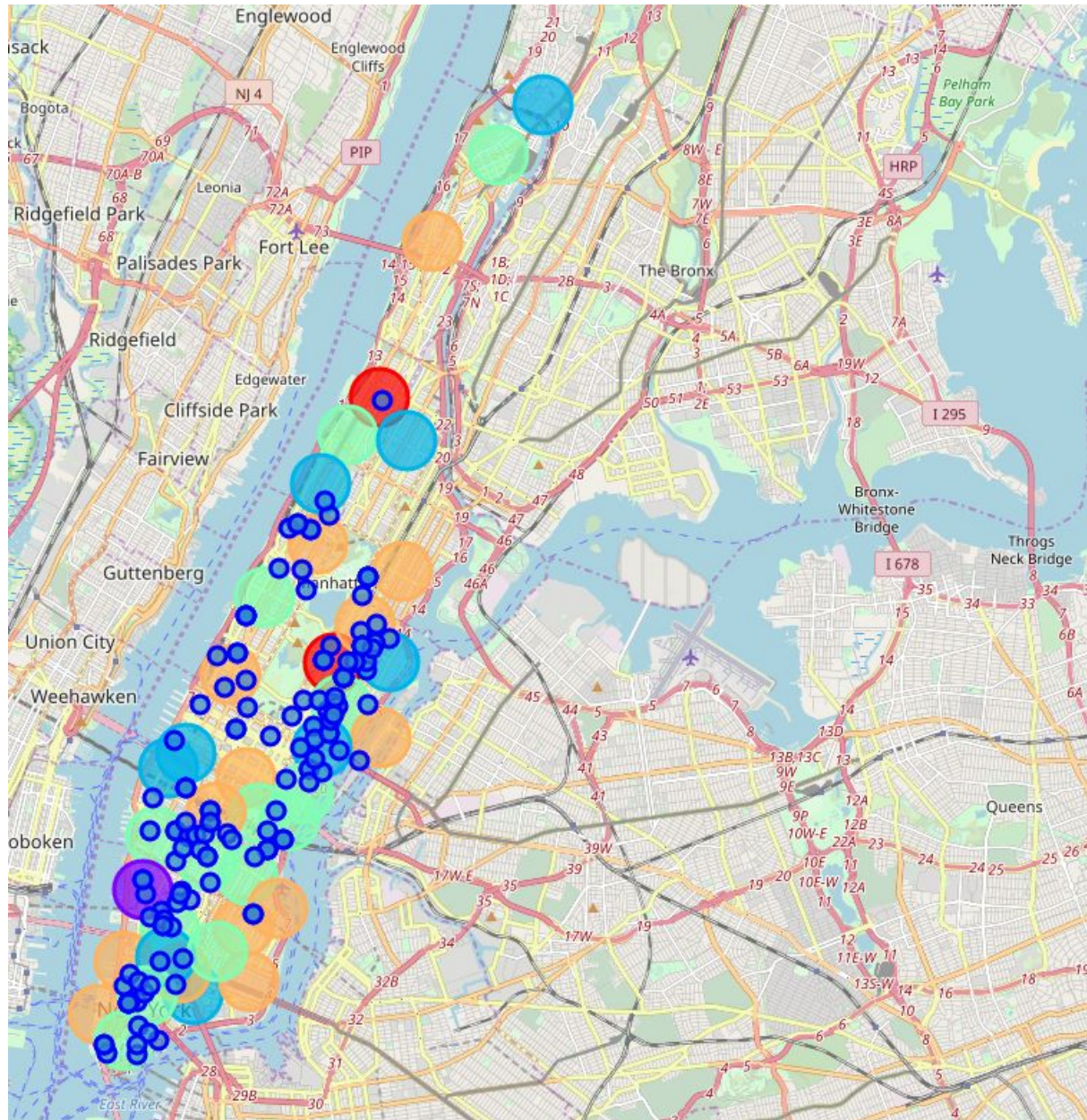
Now one can explore a particular rental place and its venues in detail In the map above, examination of appartments with rental place below 7000/month is straightforwad while knowing the venues around it. We could find an appartment with at the right price and in a location with desirable venues. The next step is to see if it is located near a subway metro station, in next cells work.

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

Mapping Manhattan Subway locations

Manhattan subway metro locations (address) was obtained from webscrapping 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.

	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

**MAP of Manhattan showing the location of subway stations**
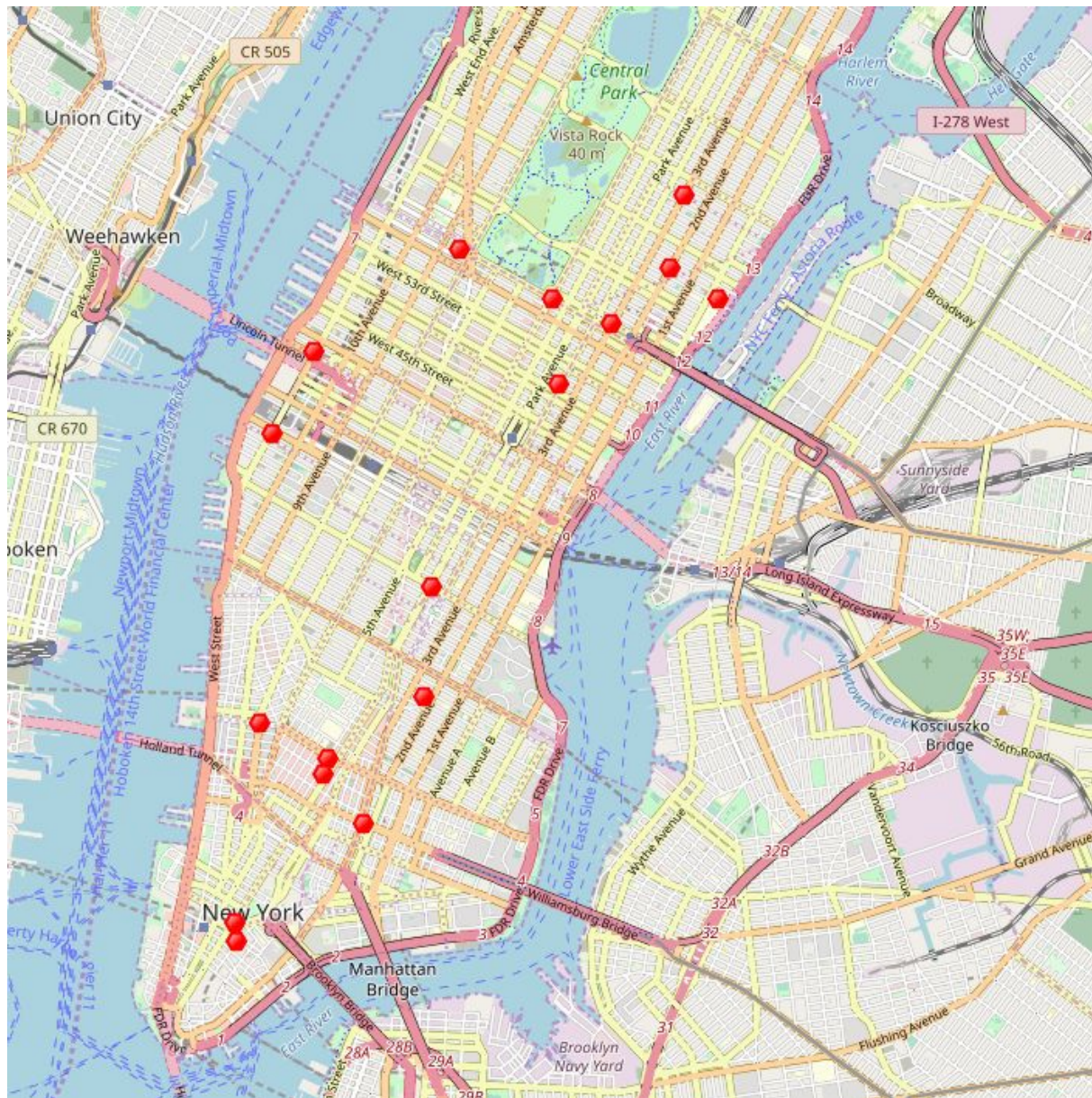
Map of Manhattan showing places for rent and the subway locations nearby

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.

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

|	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	40.785255	-73.947350
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
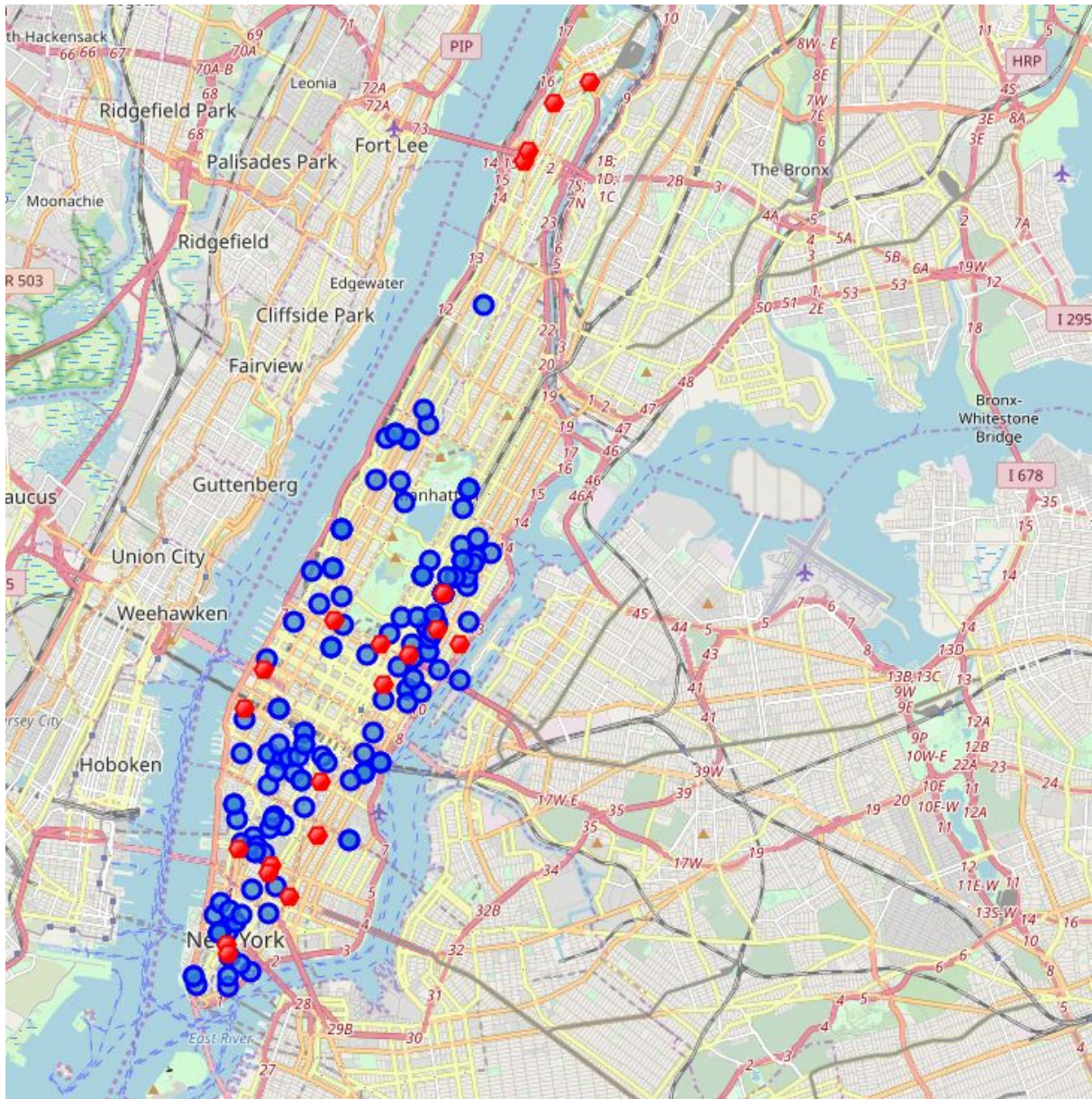
4.0 Results

ONE CONSOLIDATE MAP

Let's consolidate all the required inforamtion 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
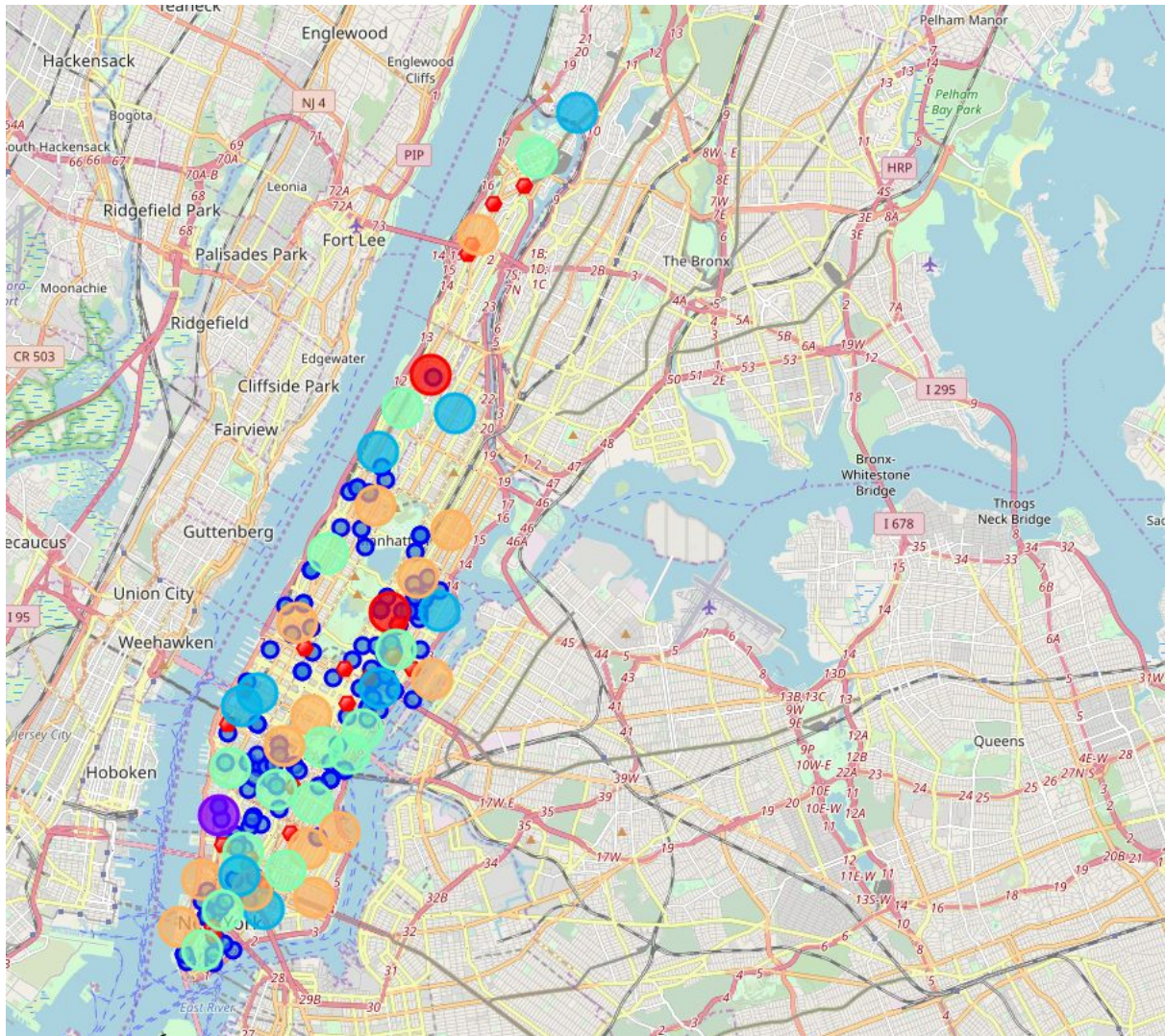
Problem Resolution

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 Neighborhood and near 'subway 59th Street' station, Cluster # 2 Monthly rent : 7500 Dollars

Apartment 2: 19 Dutch Street in the Financial District Neighborhood and near 'Fulton Street Subway' station, Cluster # 3 Monthly rent : 6935 Dollars

Venues for Apartment 1 - Cluster 2

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Marble Hill	Coffee Shop	Discount Store	Yoga Studio	Steakhouse	Supplement Shop
1	Chinatown	Chinese Restaurant	Cocktail Bar	Dim Sum Restaurant	American Restaurant	Vietnamese Restaurant
6	Central Harlem	African Restaurant	Seafood Restaurant	French Restaurant	American Restaurant	Cosmetics Shop
9	Yorkville	Coffee Shop	Gym	Bar	Italian Restaurant	Sushi Restaurant
14	Clinton	Theater	Italian Restaurant	Coffee Shop	American Restaurant	Gym / Fitness Center
23	Soho	Clothing Store	Boutique	Women's Store	Shoe Store	Men's Store
26	Morningside Heights	Coffee Shop	American Restaurant	Park	Bookstore	Pizza Place
34	Sutton Place	Gym / Fitness Center	Italian Restaurant	Furniture / Home Store	Indian Restaurant	Dessert Shop
39	Hudson Yards	Coffee Shop	Italian Restaurant	Hotel	Theater	American Restaurant

cluster-3

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
3	Inwood	Mexican Restaurant	Lounge	Pizza Place	Café	Wine Bar
5	Manhattanville	Deli / Bodega	Italian Restaurant	Seafood Restaurant	Mexican Restaurant	Sushi Restaurant
10	Lenox Hill	Sushi Restaurant	Italian Restaurant	Coffee Shop	Gym / Fitness Center	Pizza Place
12	Upper West Side	Italian Restaurant	Bar	Bakery	Vegetarian / Vegan Restaurant	Indian Restaurant
16	Murray Hill	Sandwich Place	Hotel	Japanese Restaurant	Gym / Fitness Center	Coffee Shop
17	Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bakery	Nightclub
18	Greenwich Village	Italian Restaurant	Sushi Restaurant	French Restaurant	Clothing Store	Chinese Restaurant
27	Gramercy	Italian Restaurant	Restaurant	Thrift / Vintage Store	Cocktail Bar	Bagel Shop
29	Financial District	Coffee Shop	Hotel	Gym	Wine Shop	Steakhouse
31	Noho	Italian Restaurant	French Restaurant	Cocktail Bar	Gift Shop	Bookstore
32	Civic Center	Gym / Fitness Center	Bakery	Italian Restaurant	Cocktail Bar	French Restaurant
35	Turtle Bay	Italian Restaurant	Coffee Shop	Steakhouse	Wine Bar	Sushi Restaurant
36	Tudor City	Café	Park	Pizza Place	Mexican Restaurant	Greek Restaurant
38	Flatiron	Italian Restaurant	American Restaurant	Gym	Gym / Fitness Center	Yoga Studio

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 rent cost is US7500 slightly above the US7000 budget. Apt 1 is located 400 meters from subway station at 59th Street and work place (Park Ave and 53rd) is another 600 meters way. I can walk to work place and use subway for other places aroung. Venues for this apt are as of Cluster 2 and it is located 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.¶ Based on current Singapore venues, I feel that Cluster 2 type of venues is a closer resemblance to my current place. That means that APARTMENT 1 is a better choice since the extra monthly rent is worth the conveniences it provides.

5.0 DISCUSSION In general, I am positively impressed with the overall organization, content and lab works presented during the Coursera IBM Certification Course I feel this Capstone project presented me a great opportunity to practice and apply the Data Science tools and methodologies learned. I have created a good project that I can present as an example to show my potential. 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.

6.0 CONCLUSIONS I feel rewarded with the efforts, time and money spent. I believe this course with all the topics covered is well worthy of appreciation. This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools. The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision thoroughly and with confidence. I would recommend for use in similar situations.

In [ ]: