



# APPLIED DATA SCIENCE CAPSTONE

Battle of Neighborhoods

Muzothule Kubheka

# Report Content

## 1. Introduction Section :

- The “business problem” to be solved by this project and who may be interested

## 2. Data Section:

- Describe Data requirements and Sources needed to solve the problem

## 3. Methodology section:

- Main component of the report
- Execute data processing, describe/discuss any exploratory data analysis and/or inferential statistical testing performed, and/or machine learnings used.

## 4. Results section:

- Discussion of the results and finding of answer

## 5. Discussion section:

- Discussion of observations noted and any recommendations

## 6. Conclusion section:

- Answer chosen and conclusions.

# 1.0 Introduction

## Scenario and Background

I am currently living and working in Singapore, within walking distance to Town Center with great venues and attractions and restaurants, such as international cuisine, entertainment and shopping and kindergarten for my family. I am current considering a job offer that will require me to move to New York and I would like to explore if I can find a place to live that is similar with the one I currently live in.

## Problem to be resolved

How to find an apartment in Manhattan with the following conditions:

- Apartment with min 2 bedrooms
- Monthly rent not to exceed US\$7000/month
- Located within walking distance ( $\leq 1.0$  mile, 1.6 km) from a subway metro station in Manhattan
- Venues and amenities as in my current residence.

## 2.0 Data Section

### Data Requirements

- Geodata for current residence in Singapore with venues established using Foursquare.
- List of Manhattan (MH) neighborhoods with clustered venues established via Foursquare (as in Course Lab).  
[https://en.wikipedia.org/wiki/List\\_of\\_Manhattan\\_neighborhoods#Midtown\\_neighborhoods](https://en.wikipedia.org/wiki/List_of_Manhattan_neighborhoods#Midtown_neighborhoods)
- List of subway metro stations in Manhattan with addresses and geo data (lat,long):  
[https://en.wikipedia.org/wiki/List\\_of\\_New\\_York\\_City\\_Subway\\_stations\\_in\\_Manhattan](https://en.wikipedia.org/wiki/List_of_New_York_City_Subway_stations_in_Manhattan)  
(<https://www.google.com/maps/search/manhattan+subway+metro+stations/@40.7837297,-74.1033043,11z/data=!3m1!4b1>)
- List of apartments for rent in Manhattan area with information on neighborhood location, address, number of beds, area size, monthly rent price and complemented with geo data via Nominatim.  
<http://www.rentmanhattan.com/index.cfm?page=search&state=results>  
<https://www.nestpick.com/search?city=new-york>
- Place to work in Manhattan (Park Avenue and 53rd St) for reference
- 

### Data Sources, Data Processing and Tools used

- Singapore data and map is to be created with use of Nominatim , Foursquare and Folium mapping
- Manhattan neighborhoods were obtained from Wikipedia and organized by Neighborhoods with geodata via Nominatim for mapping with Folium.
- List of Subway stations was obtained via Wikipedia, NY Transit web site and Google map,
- List of apartments for rent was consolidated from web-scraping real estate sites for MH. The geolocation (lat,long) data was found with algorithm coding and using Nominatim.
- Folium map was the basis of mapping with various features to consolidate all data in ONE map where one can visualize all details needed to make a selection of apartment

## 3.0 Methodology

### Approach

The strategy is based on mapping the described data in section 2.0, in order to facilitate the choice of at least two candidate places for rent. The information will be consolidated in ONE MAP where one can see the details of the apartment, the cluster of venues in the neighborhood and the relative location from a subway station and from work place. A measurement tool icon will also be provided. The popups on the map items will display rent price, location and cluster of venues applicable.

### The Tools:

Web-scraping of sites is used to consolidate data-frame information which was saved as txt files for convenience and to simplify the report. Geodata was obtained by coding a program to use Nominatim to get latitude and longitude of subway stations and also for each of (144 units) the apartments for rent listed. Geopy\_distance and Nominatim were used to establish relative distances. Seaborn graphic was used for general statistics on rental data. Maps with popups labels allow quick identification of location, price and feature, thus making the selection very easy

## 4.0 Execution and Results

### Current residence neighborhood in Singapore

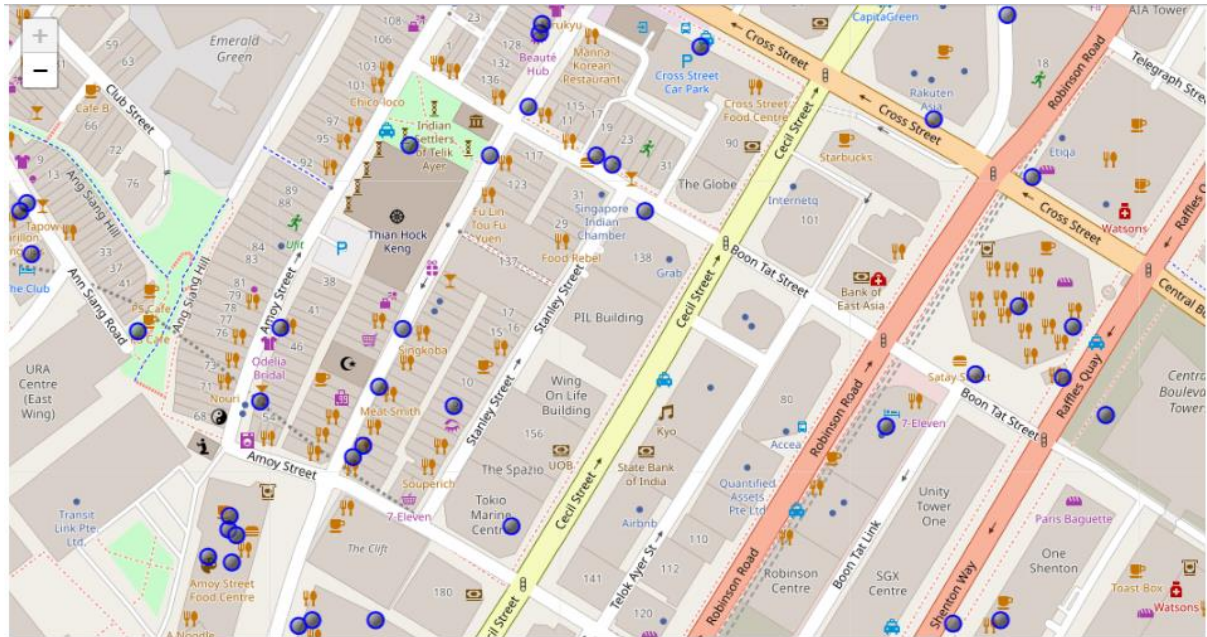


Figure 1: Map of current residence

### Venues around current neighborhood

```
# Venues near current Singapore residence place
SGnearby_venues.head(10)
```

	name	categories	lat	lng
0	The Westin Singapore	Hotel	1.278275	103.850772
1	Pure Fitness	Gym	1.278631	103.851487
2	Lau Pa Sat Satay Street	Street Food Gathering	1.280261	103.850235
3	Anglo Indian Cafe & Bar	Indian Restaurant	1.279084	103.850127
4	Westin Infinity Pool	Pool	1.278057	103.851077
5	Napoleon Food & Wine Bar	Wine Bar	1.279925	103.847333
6	Sofitel So Singapore	Hotel	1.280017	103.849813
7	Lobby Lounge Westin	Bar	1.277811	103.850966
8	Mellower Coffee	Café	1.277814	103.848188
9	Cook & Brew	Gastropub	1.277842	103.851103

Figure 2: Venues around current residence

## Manhattan Map - Neighborhoods and Cluster of Venues

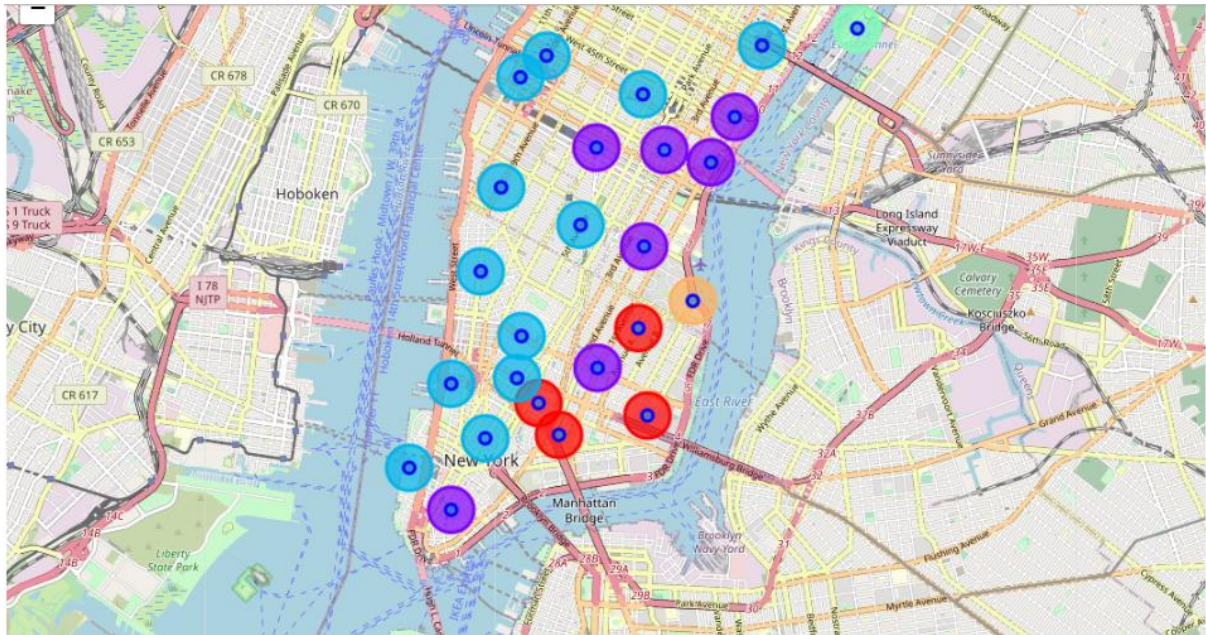


Figure 3: Manhattan Map with Venue Clusters

## GeoData Manhattan apartments for rent

```
mh_rent=pd.read_csv('MH_rent_latlong.txt')
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

## Rental Price Statistics MH Apartments

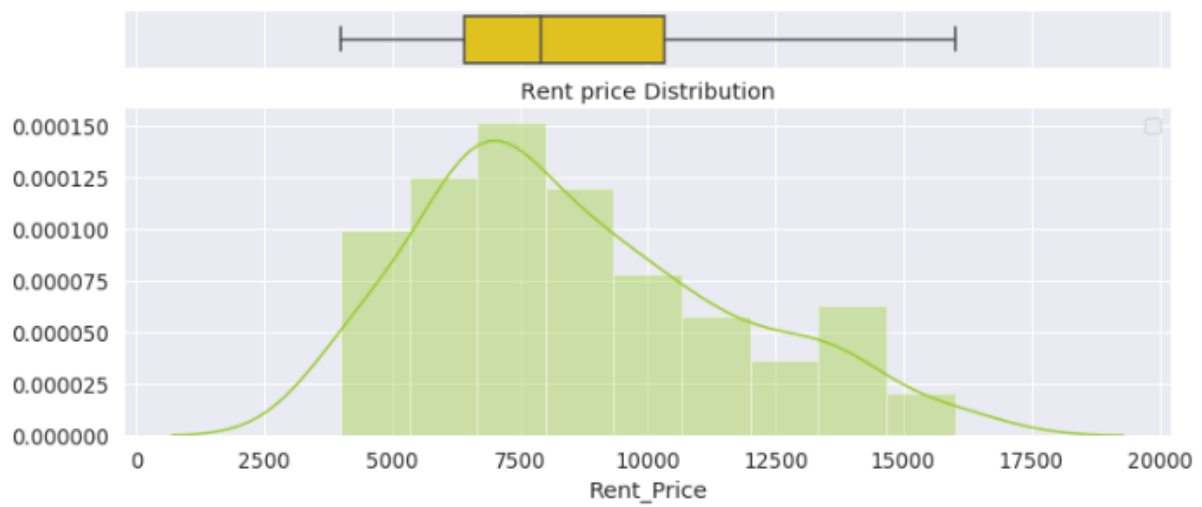


Figure 4: Rent price distribution



Figure 5: Price per sq.ft distribution

Budget US7000/month is around the mean



## Apartments for Rent in Manhattan

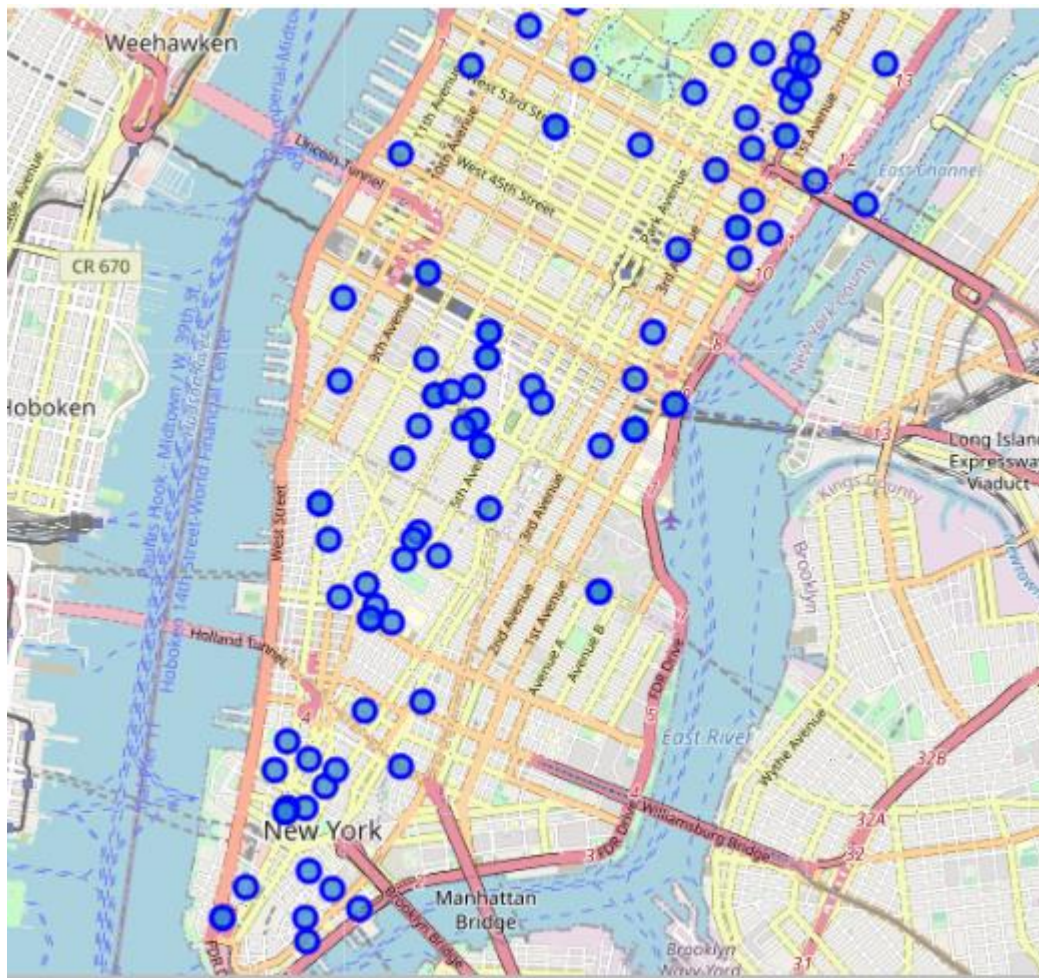


Figure 6: Map of apartments for rent in Manhattan

Manhattan apartments for rent with venue clusters

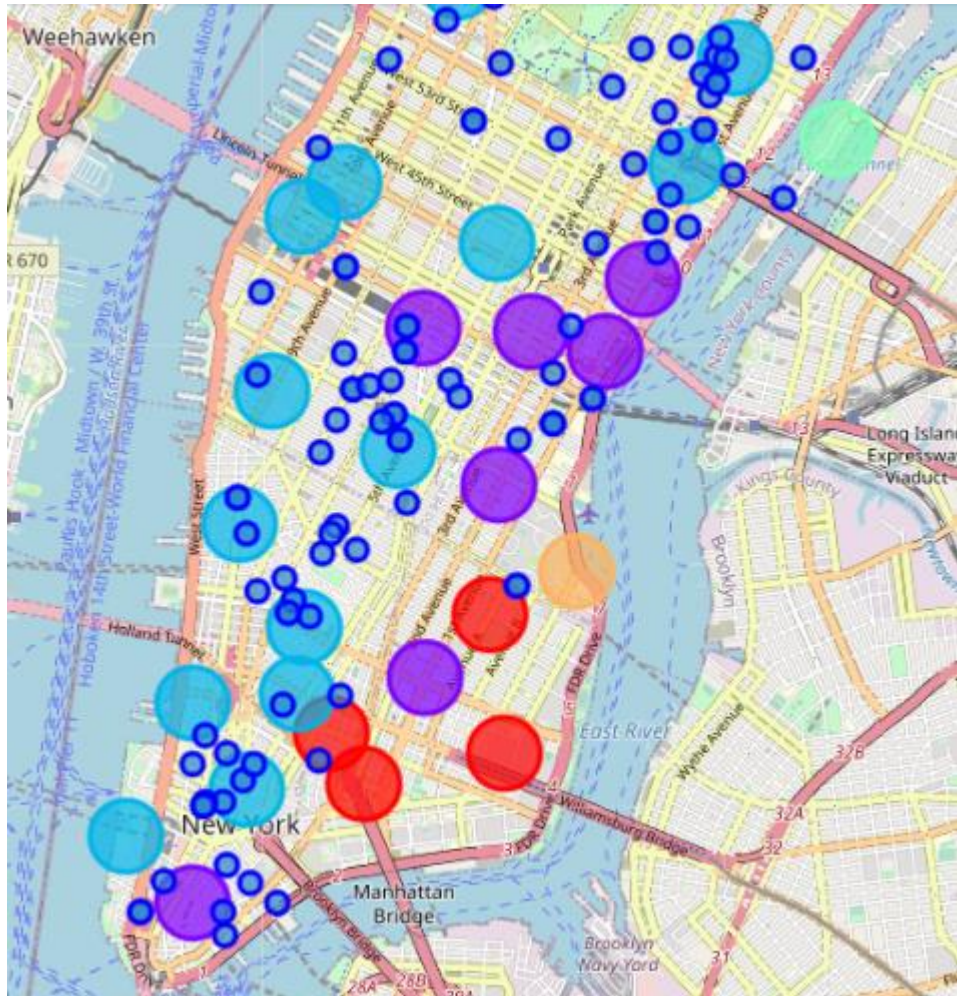


Figure 7: Apartments and Venues Clusters Map

## Venues of cluster 2

	Borough	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
8	Manhattan	2	Italian Restaurant	Bakery	Gym / Fitness Center	Spa	Coffee Shop	Exhibit	Hotel	Juice Bar	Yoga Studio	Wine Shop
10	Manhattan	2	Coffee Shop	Italian Restaurant	Pizza Place	Cocktail Bar	Café	Sushi Restaurant	Gym / Fitness Center	Gym	Burger Joint	Thai Restaurant
13	Manhattan	2	Italian Restaurant	Plaza	Café	Concert Hall	Theater	Performing Arts Venue	American Restaurant	Wine Shop	French Restaurant	Gym / Fitness Center
14	Manhattan	2	Theater	Gym / Fitness Center	Coffee Shop	Hotel	Gym	Italian Restaurant	Sandwich Place	Cocktail Bar	Pizza Place	Spa
15	Manhattan	2	Coffee Shop	Bakery	Hotel	Theater	Japanese Restaurant	Clothing Store	Cuban Restaurant	Pizza Place	Sandwich Place	Gym
17	Manhattan	2	Art Gallery	Coffee Shop	Café	American Restaurant	Ice Cream Shop	Bakery	Market	Seafood Restaurant	Park	Cupcake Shop
18	Manhattan	2	Italian Restaurant	Café	Gym	Sushi Restaurant	Pizza Place	Coffee Shop	Indian Restaurant	Bakery	Vietnamese Restaurant	Comedy Club
21	Manhattan	2	Italian Restaurant	Park	American Restaurant	Bakery	Café	Spa	Wine Bar	Steakhouse	Skate Park	Scenic Lookout
23	Manhattan	2	Italian Restaurant	Mediterranean Restaurant	Sandwich Place	Gym	Coffee Shop	Clothing Store	Pizza Place	Sushi Restaurant	Ice Cream Shop	Hotel
24	Manhattan	2	Wine Bar	Italian Restaurant	American Restaurant	Park	Bakery	Jazz Club	Pizza Place	New American Restaurant	Cocktail Bar	Coffee Shop
28	Manhattan	2	Park	Coffee Shop	Hotel	Wine Shop	Gym	Memorial Site	Shopping Mall	Food Court	Gourmet Shop	Playground

Figure 8: Some of the venues within Cluster 2



## Manhattan subway stations geodata

	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	40.764250	-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

Figure 9: Dataframe head for Manhattan Subway Stations

Map with apartments for rent in blue and subway stations in red

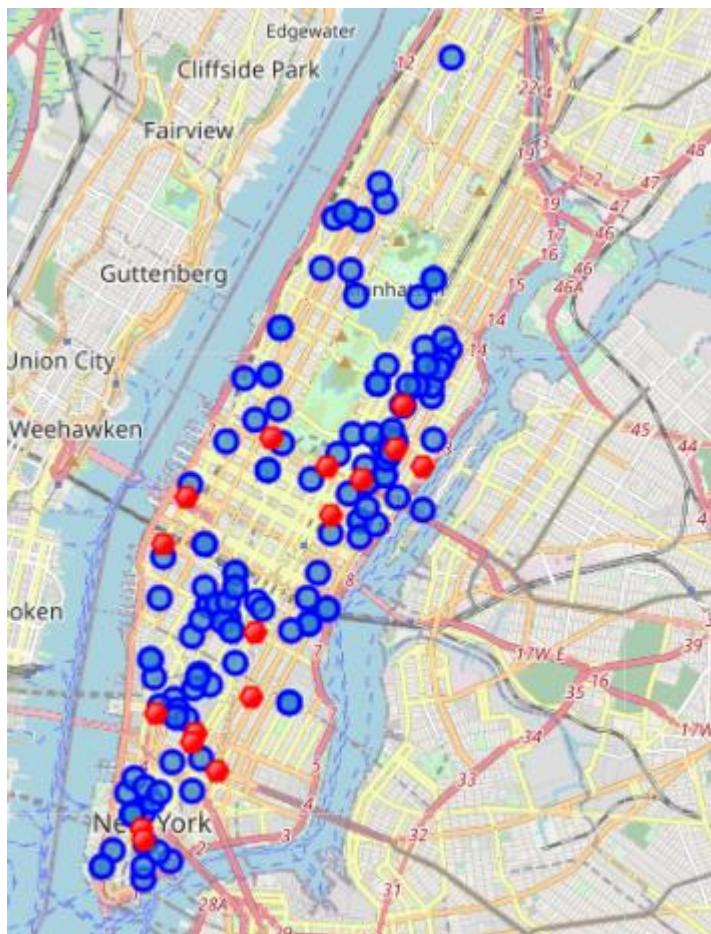


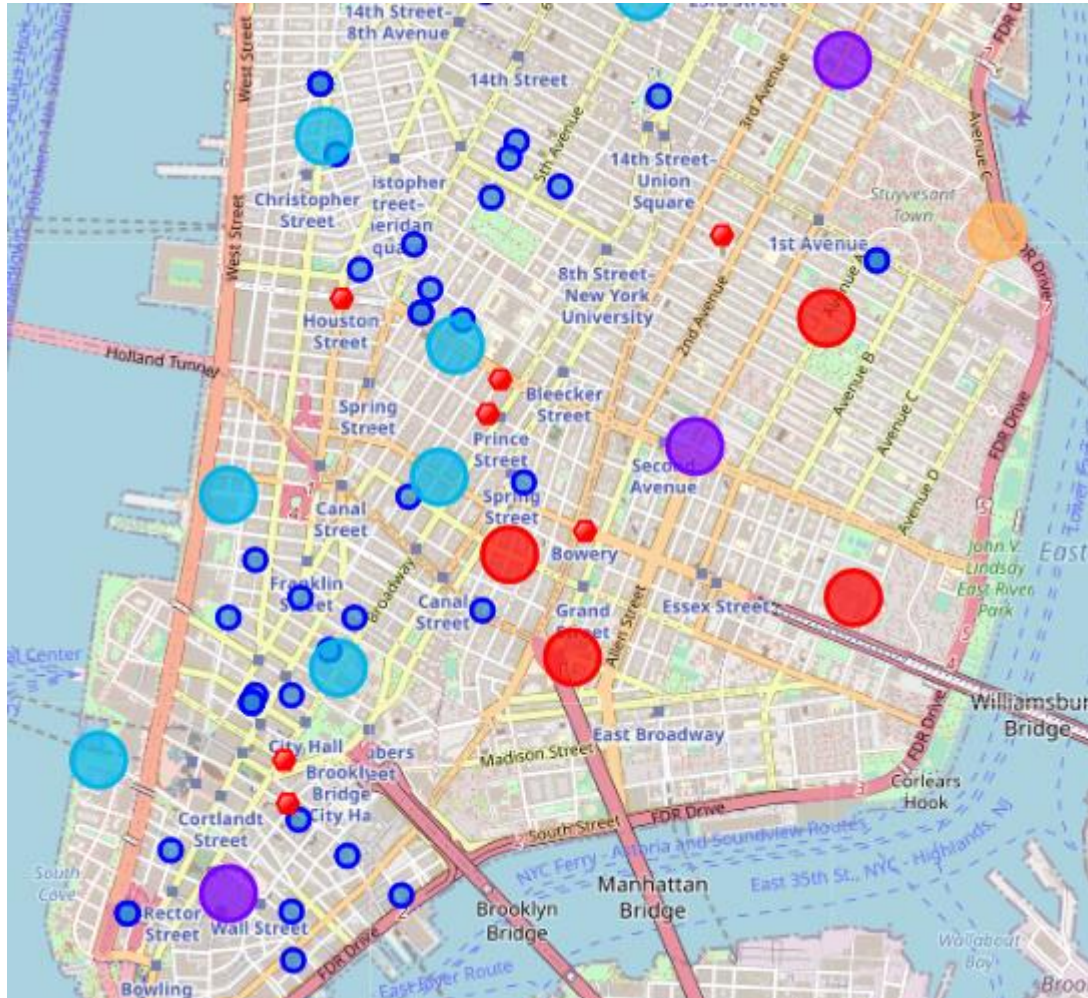
Figure 10: Map of rental apartments and subway stations

## Selected Apartment

The ONE consolidated map shows all information for decision:

Apartments address, price, neighborhood, cluster of venues and subway station nearby.

Blue dots=apartments , Red dots=Subway station, Bubbles=Cluster of Venues



## Final Result: 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 US\$7500 slightly above the US\$7000 budget. Apartment 1 is located 400 meters from subway station at 59th Street and work place (Park Ave and 53rd) is another 600 meters away. I can walk to work place and use subway for other places around. Venues for this apartment are as of Cluster 2 and it is located in a fine district in the East side of Manhattan.

Apartment 2 rent cost is US\$6935, just under the US\$7000 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 apartment 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.

- I will be able to walk to work
- Walk from home to work is less than 1 km
- Venues in Cluster 2 near future home