# Coursera Capstone Project

**Coursera IBM Data Science Certification**

**Geesala Vijay Kumar**

7$^{th}$ July,2020

# Report Content

**1 Introduction Section:**

- Discussion of the "backgrounds situation" leading to the problem at hand:
- Problem to be resolved
- Audience for this project.

**2 Data Section:**

- Data of Current Situation (current residence place)
- Data required to resolve the problem
- Data sources and data manipulation

**3 Methodology section:**

- Process steps and strategy to resolve the problem
- Data Science Methods, machine learing, mapping tools and exploratory data analysis.

**4 Results section:**

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

**5 Discussion section:**

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

**6 Conclusion section:**

- Decisions 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 Downtown Singapore. I currently live within walking distance to Downtown "Telok Ayer MRT metro station" therefore I have access to good public transportation to work. Likewise, I enjoy many ammenities in the neighborhood , such as international cousine 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 renta 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 ammenities 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 'Mccallum Street' in Downtonw Singapore. 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 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 ( latitud and longitud. 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 ammenities in the Manhattan neighborhoods (e.g. top 10) 2.3 sources and manipulation The list of Manhattan neigbhorhood 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 algorythm was used to determine the geodata from Nominatim . The actual algorythm coding may be shown in 'markdown' mode becasues it takes time to run.

**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 Singaapproach and maps with popups labels allow quick identification of location, price and feature, thus making the selection very easy.pore. This visual

# 4.Execution and Results

# Current Residence Neighbourhood in Singapore
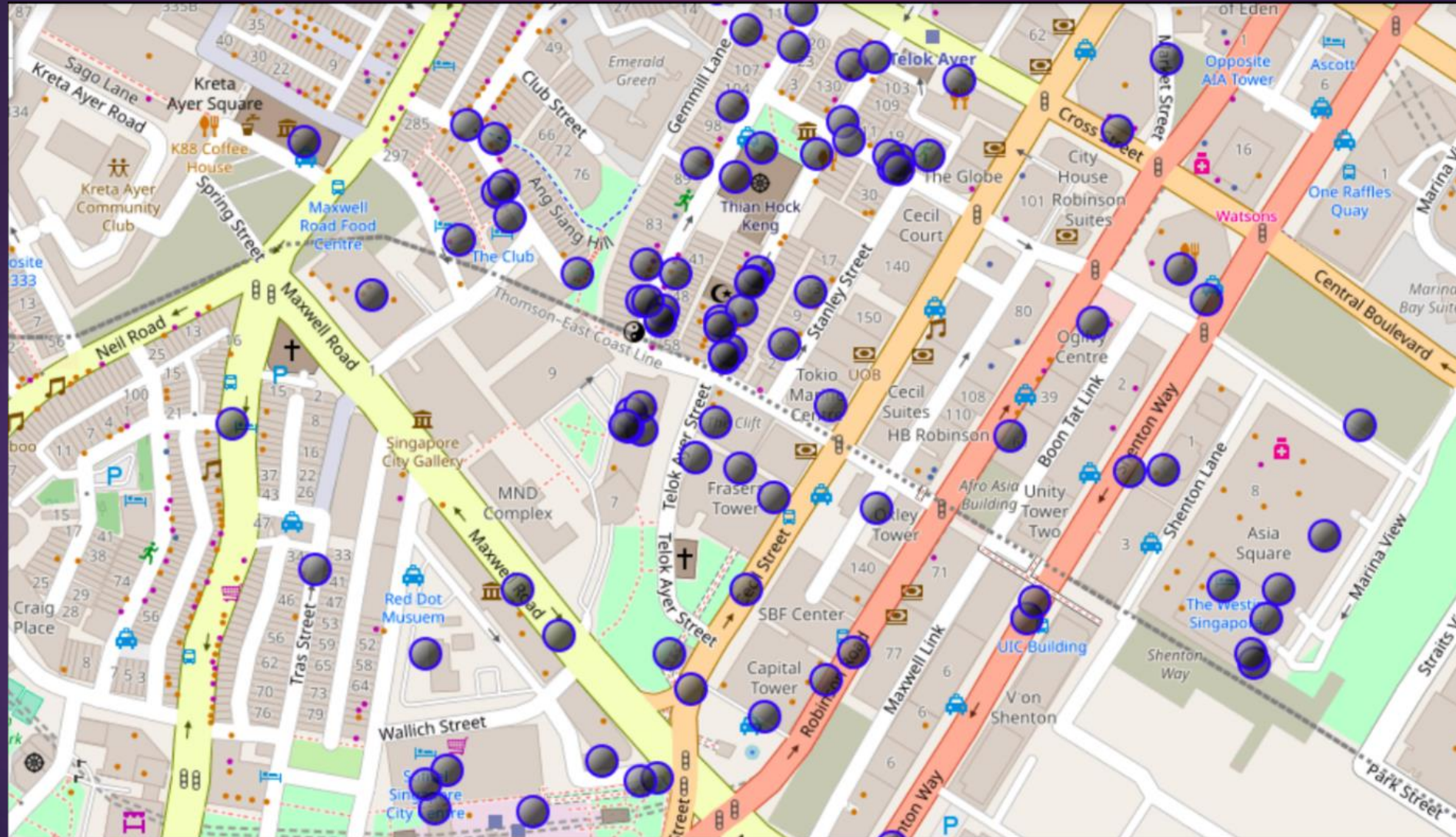
# Venues Around Neighbourhood in

```
# 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 – Neighbourhoods and clusters of venue
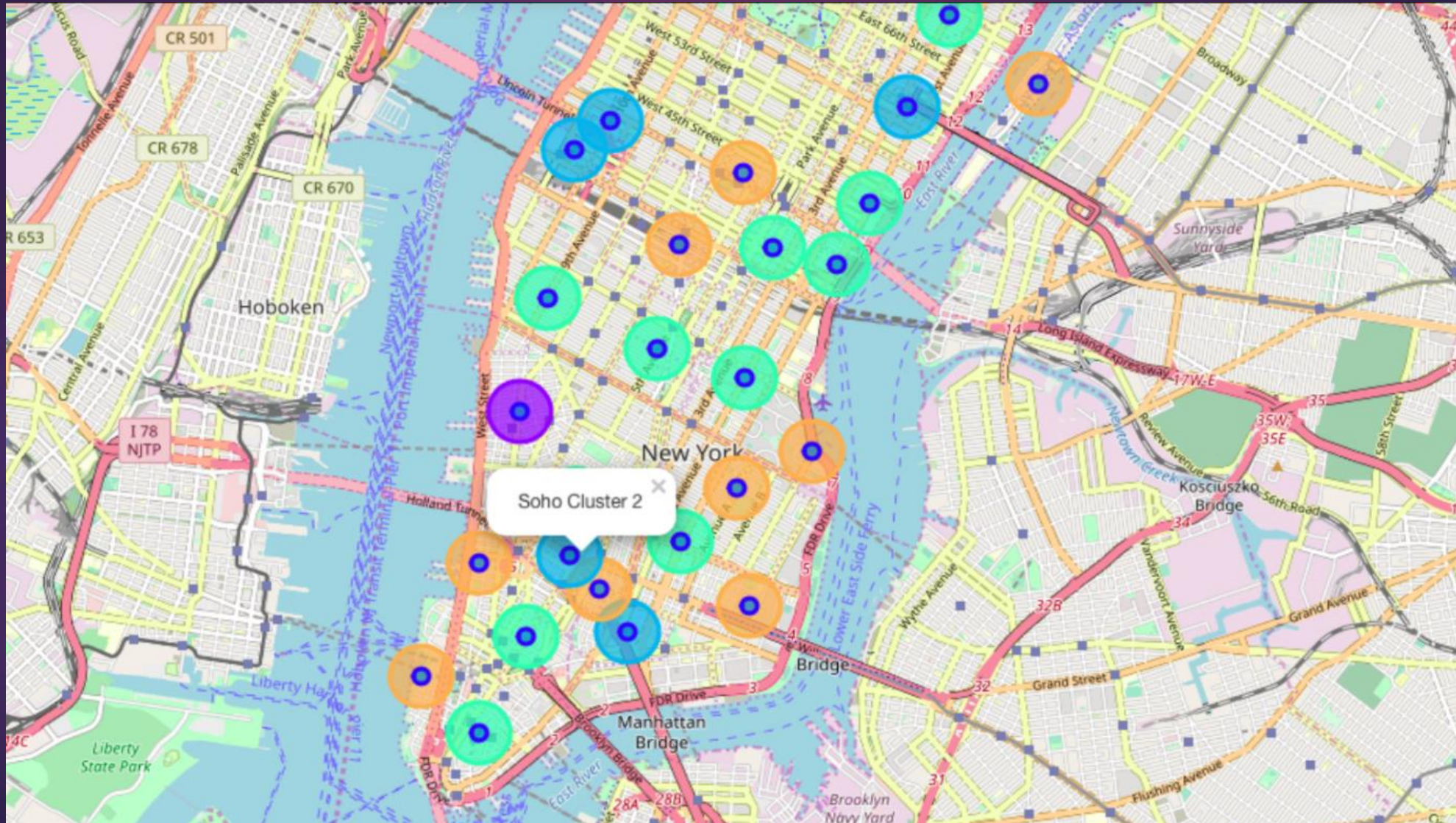
# GeoData Manhattan apts for Rent

```
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 MH Apartments

# Apartments For Rent in MH
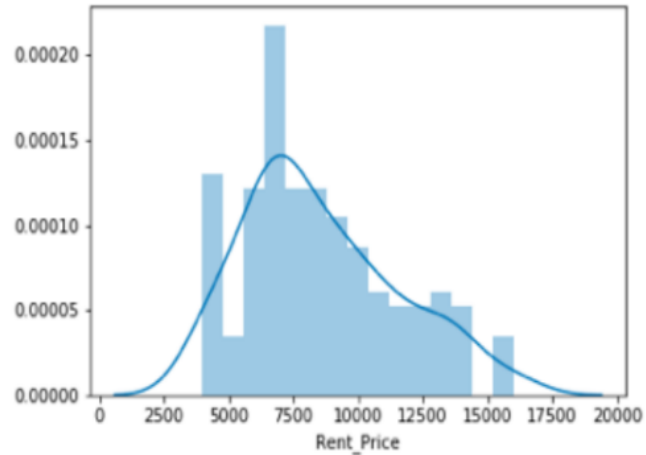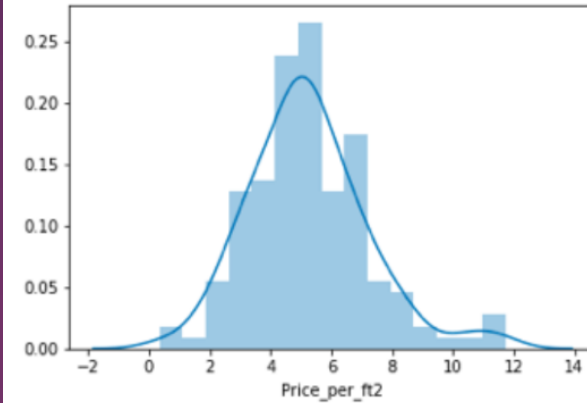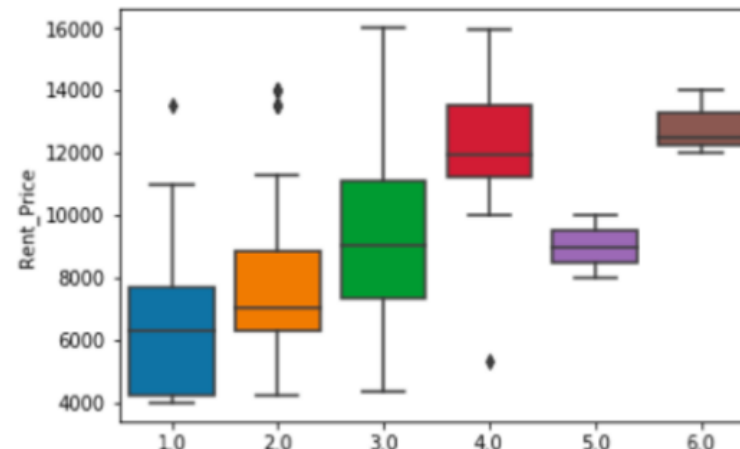
# MH apts for rent with venue clusters
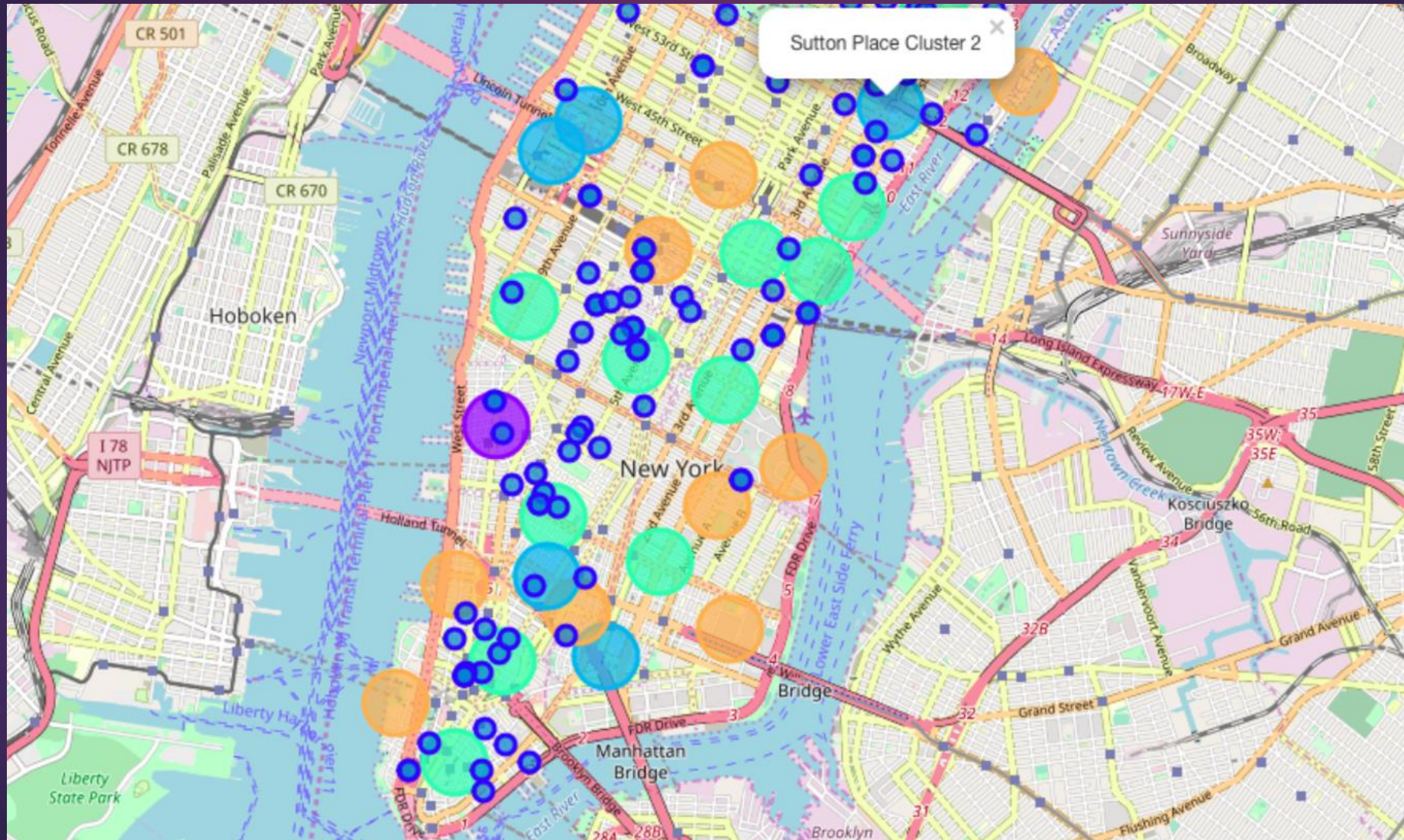
# Venues of Cluster 3

```
## 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_m
```

| | 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 |

# Manhattan subway stations geodata

```python
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 | 40.764250 | -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 |

```python
mhsub1=mh.drop_duplicates(subset=['lat','long'], keep="last").reset_index(drop=True)
mhsub1.shape
```

(22, 4)

```python
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 |

# Apts for rent (blue) and subway stations(red)
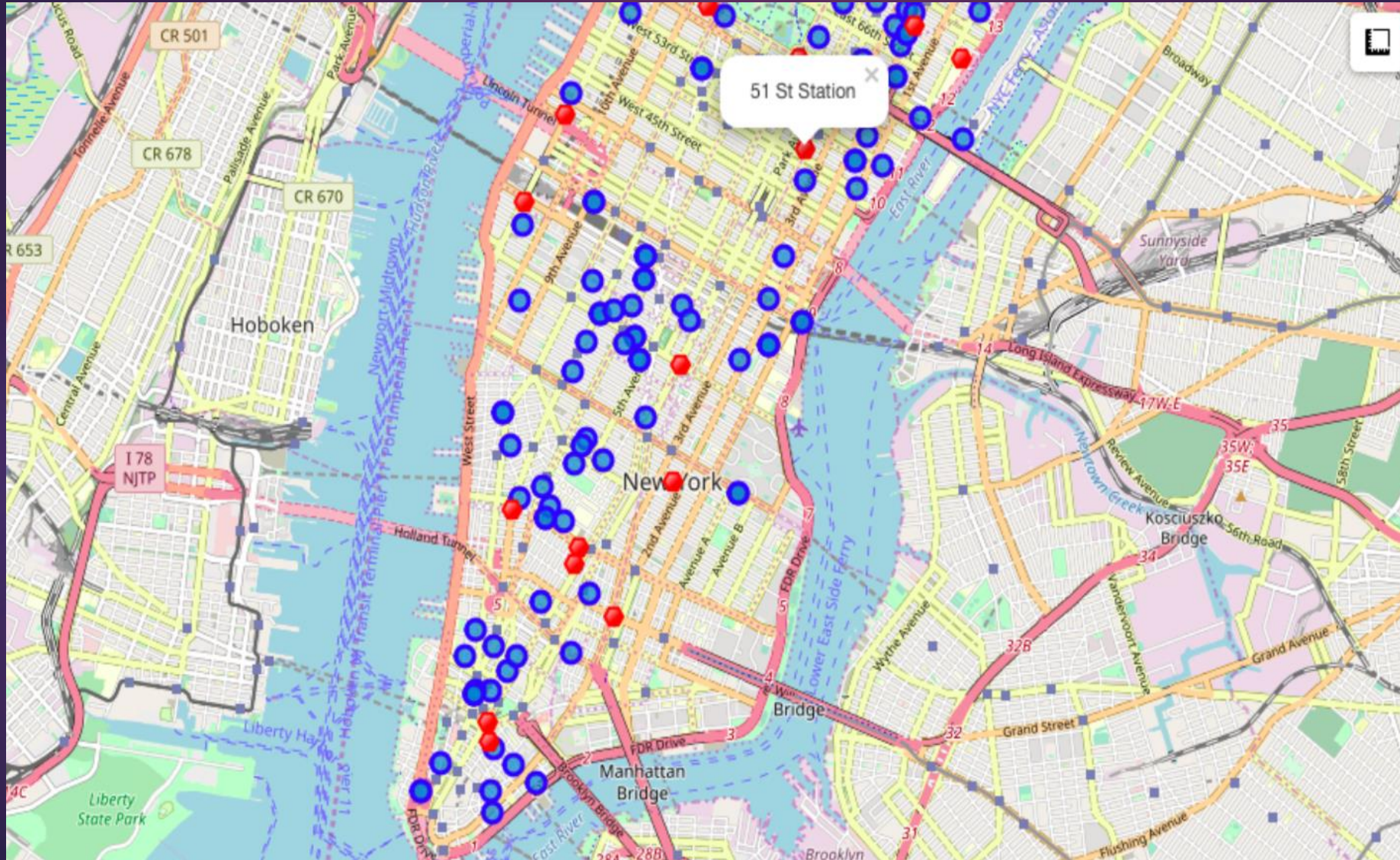
Selected Apartment!
The one consolidated map shows all information for decision.
Apartments address, price, neighourhood, cluster of venues and subway station nearby
Blue dots=apts , Red dots=Subway station, Bubbles=Cluster of Venues
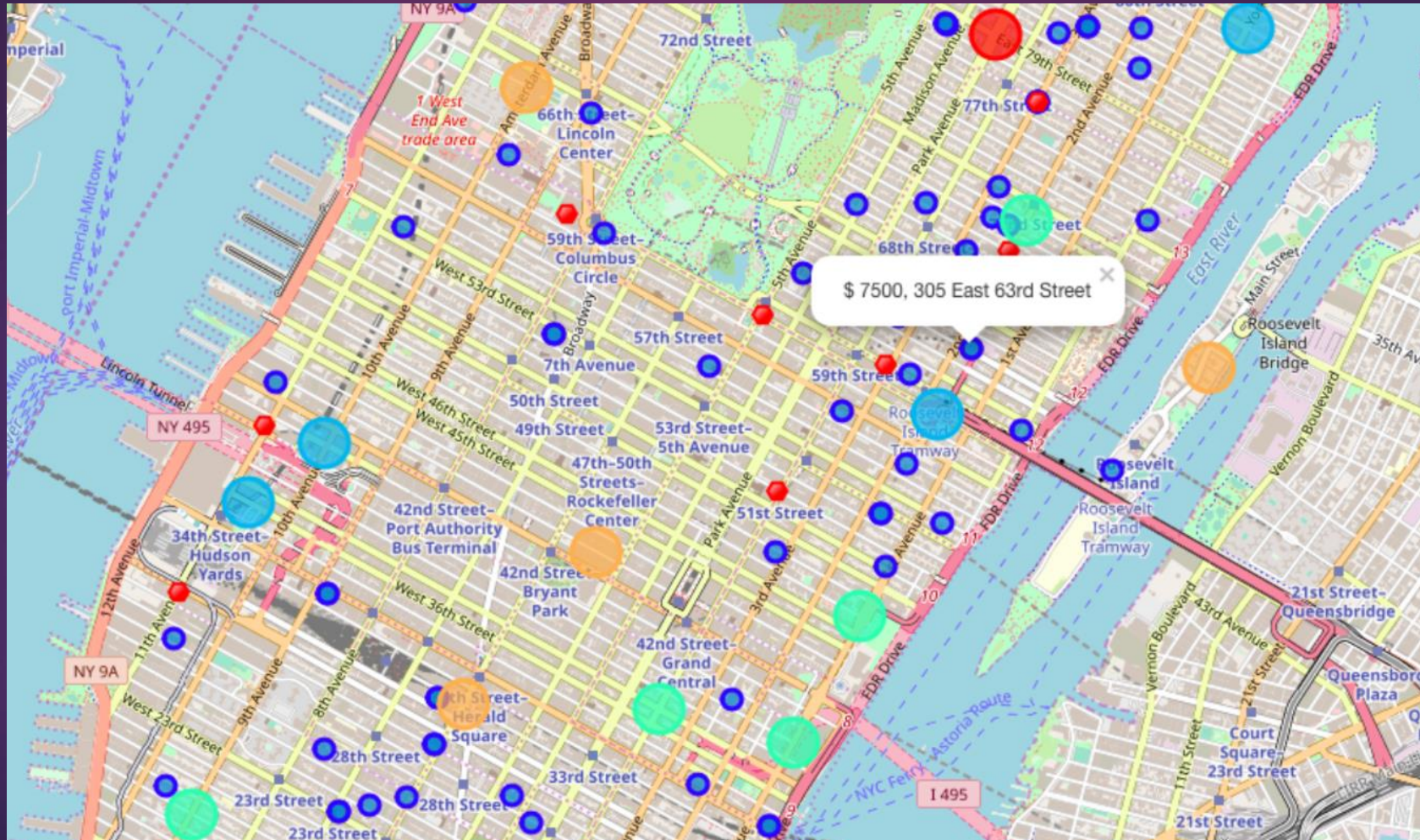
# Apartment Selection

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 around. 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.

# I will walk to work
## Walk from home to work is less then 1km!
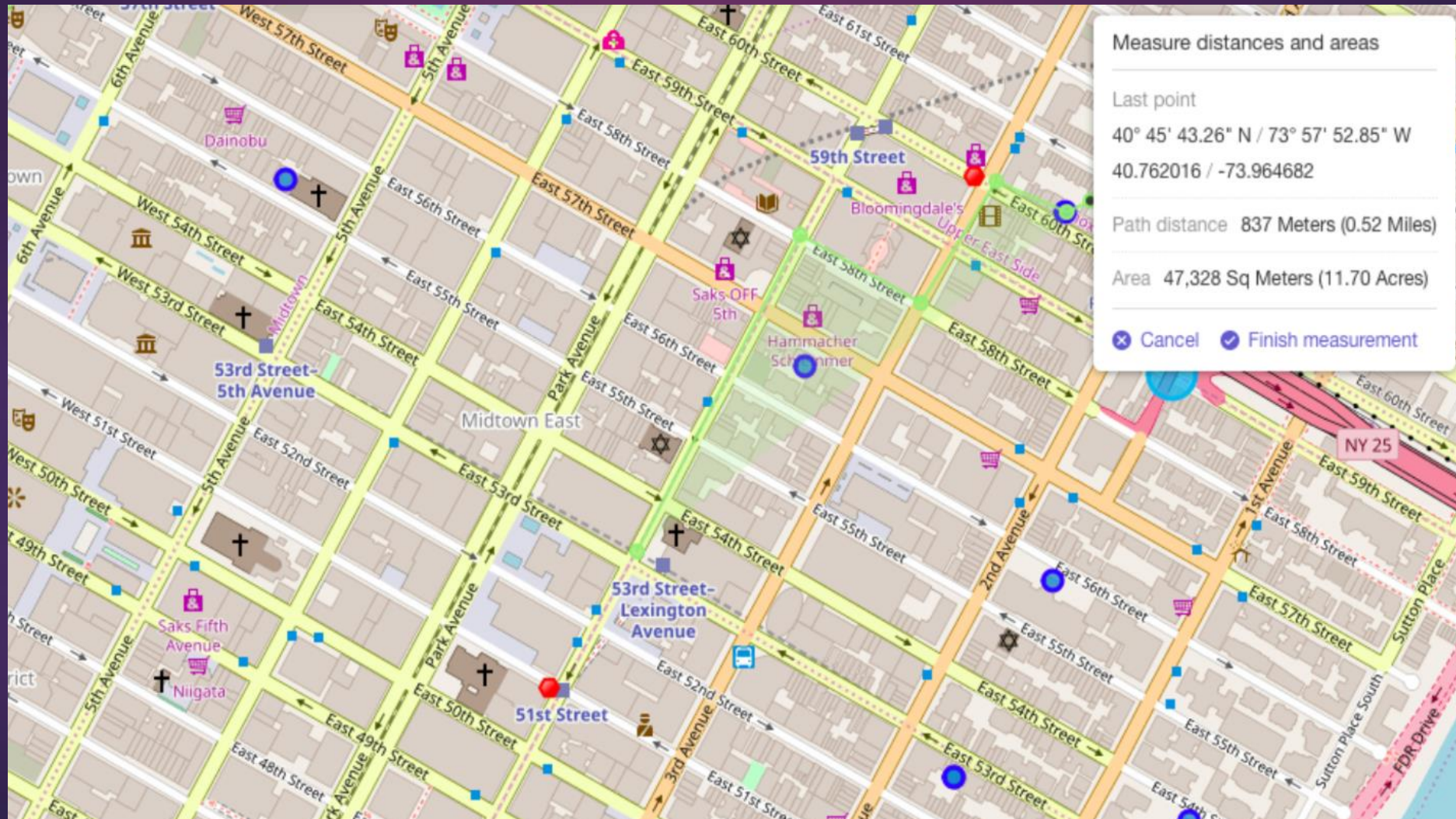
# Venus in Cluster 2 near future home

```
## 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_m
```

| | 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 |

## 5 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. 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. One must keep abreast of new tools for DS that continue to appear for application in several business fields.