# COURSERA CAPSTONE PROJECT

## Coursera IBM Data Science Certification

Bh.Hrushikesh Reddy

Apr – 17th,2020

# REPORT CONTENT

→ Introduction Section

→ Data Section

→ Methodology section

→ Results section

→ Discussion section

→ Conclusion section

# INTRODUCTION

## 1.1 Scenario and Background

I'm a Data Employee residing in Richmond Town in Bangalore. I currently live within walking distance to Richmond Town Bus-station and in this place, I enjoy many pleasurableness and places in that area, such as various international cousin restaurants, cafes, malls, Theatres, food shops and entertainment. I have been offered a great opportunity to work in the branch in Manhattan, New York. I am very interested and I want to use this opportunity.

## 1.2 Problem to be resolved:

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

- Apartment must be 2(or)3 bedrooms.
- The location required is near a metro station in the Manhattan area and within 1.0-mile radius.
- Price of rent not exceed $7,000 per month.
- Top Facilities in the selected neighbourhood shall be similar to current residence.
- Required 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 Singapore. (Which is similar to my location in Bangalore didn't get the exact location of my residence in Bangalore so I used Singapore Location).

## 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.

# DATA SECTION

## 2.1 Data of Current Situation

I Currently reside in the neighbourhood of 'MacCallum Street' in Downtown Singapore.(As I'm comparing this with my current residence in Richmond town in Bangalore) 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 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 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.
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.
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 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 becasues it takes time to run. With the use of geolocator = Nominatim() , it was possible to determine the latitude and longiude 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 neighborhoods in general and a cluster is created to later be able to search for the venues depending of the location shown.

## 2.4 Data used to solve 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 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 radious 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.

# METHODOLOGY

## 3.1 THE STRATEGY TO FIND THE ANSWER:

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.

## 3.2 THE TOOLS:

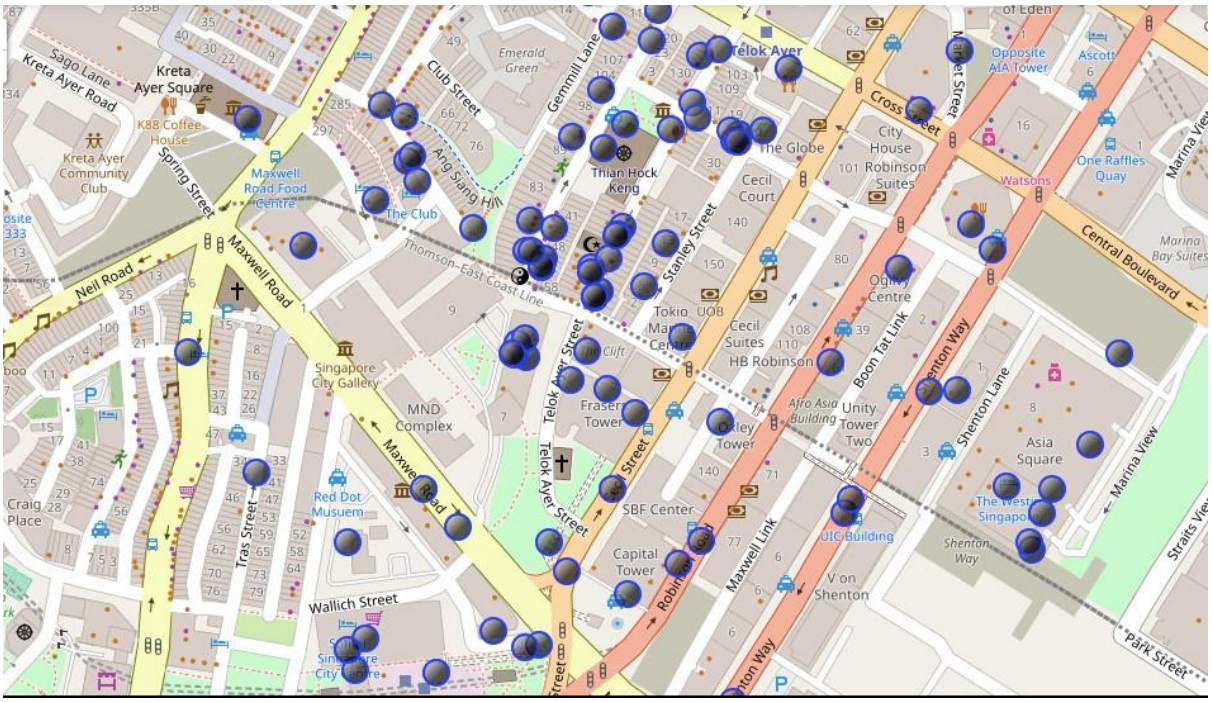
→Web-scraping of sites is used to consolidate data-frame information which was saved as csv files for convenience and to simply the report. Geodata was obtained by coding a program to use Nominatim to get latitude and longitude of subway stations and also for each of (144 units) the apartments for rent 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.

# EXECUTION AND RESULTS

## Current Residence Neighborhood in Singapore

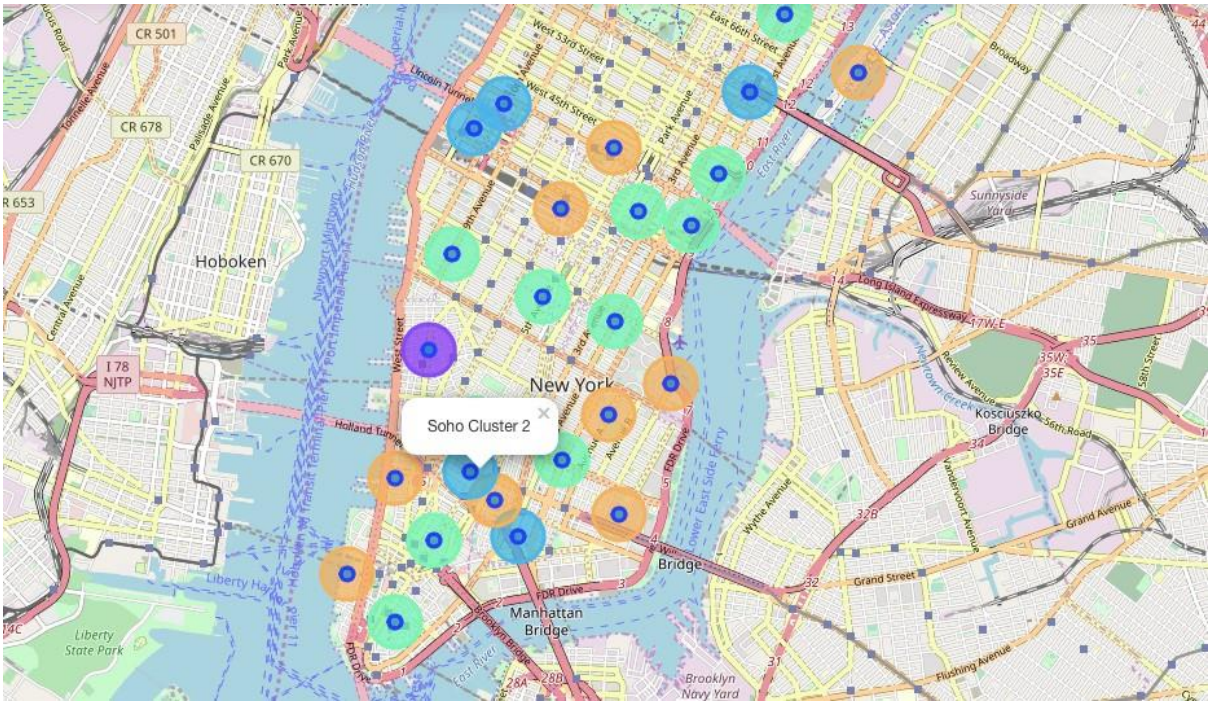



## Venues around in Neighborhood

```
# 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 – Clusters.




# Geodata Manhattan apt 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 |

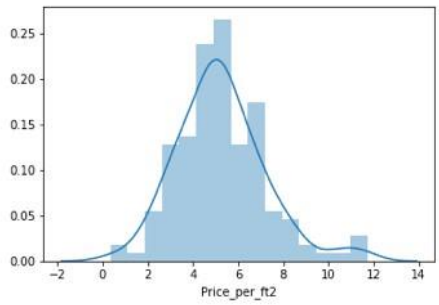
```
]: mh_rent.tail()
```

]:

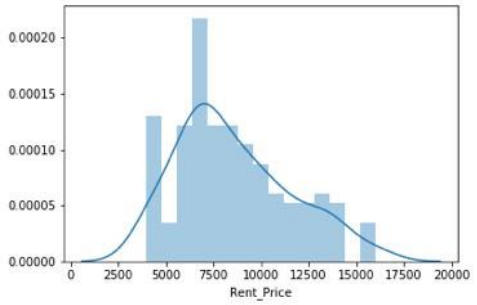
| | Address | Area | Price_per_ft2 | Rooms | Area-ft2 | Rent_Price | Lat | Long |
|---|---|---|---|---|---|---|---|---|
| 139 | 200 East 72nd Street | Rental in Lenox Hill | 5.15 | 3.0 | 1700 | 8750 | 40.769465 | -73.960339 |
| 140 | 50 Murray Street | No fee rental in Tribeca | 7.11 | 2.0 | 1223 | 8700 | 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 |

# Rents in Manhattan Apartments

```
<matplotlib.axes._subplots.AxesSubplot at 0x1a2415fc18>
```

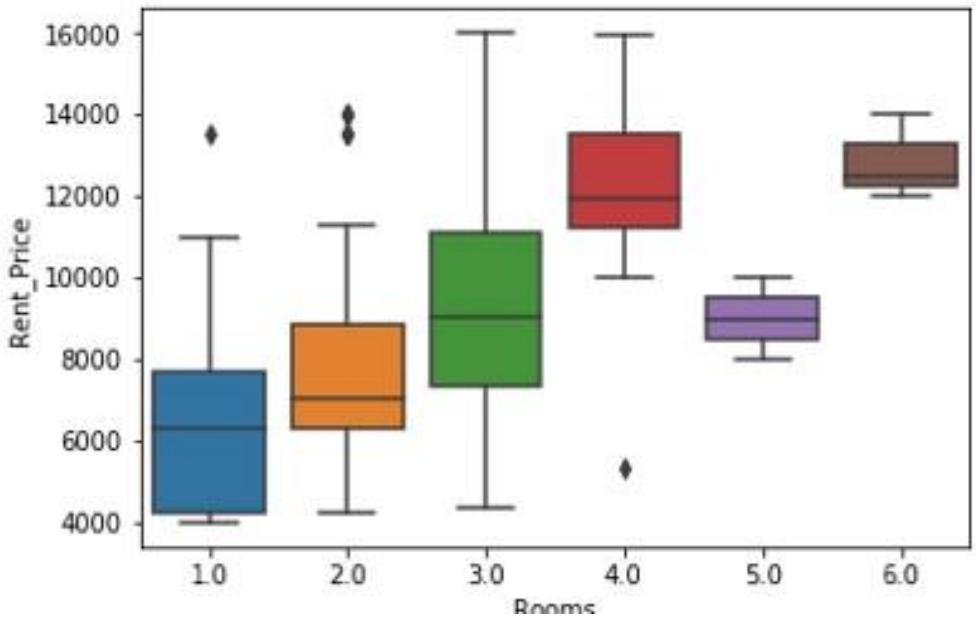



```
<matplotlib.axes._subplots.AxesSubplot at 0x1a25dd8400>
```

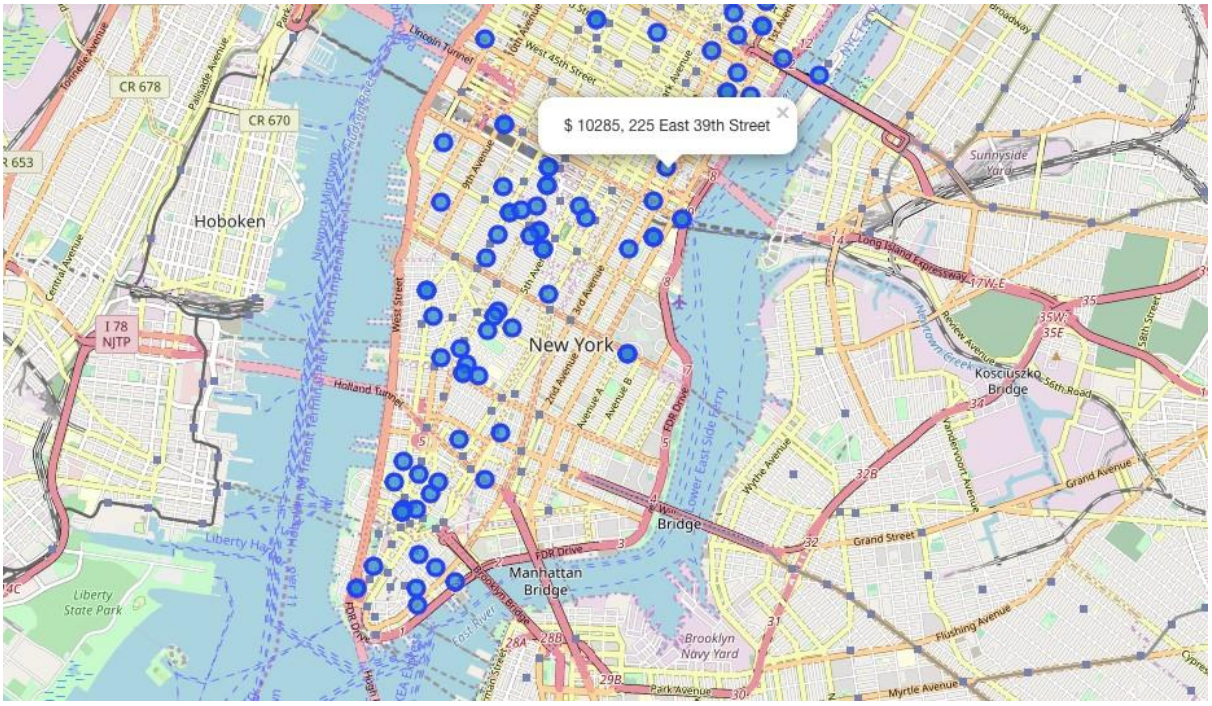



```
sns.boxplot(x='Rooms', y= 'Rent_Price', data=mh_rent)
```

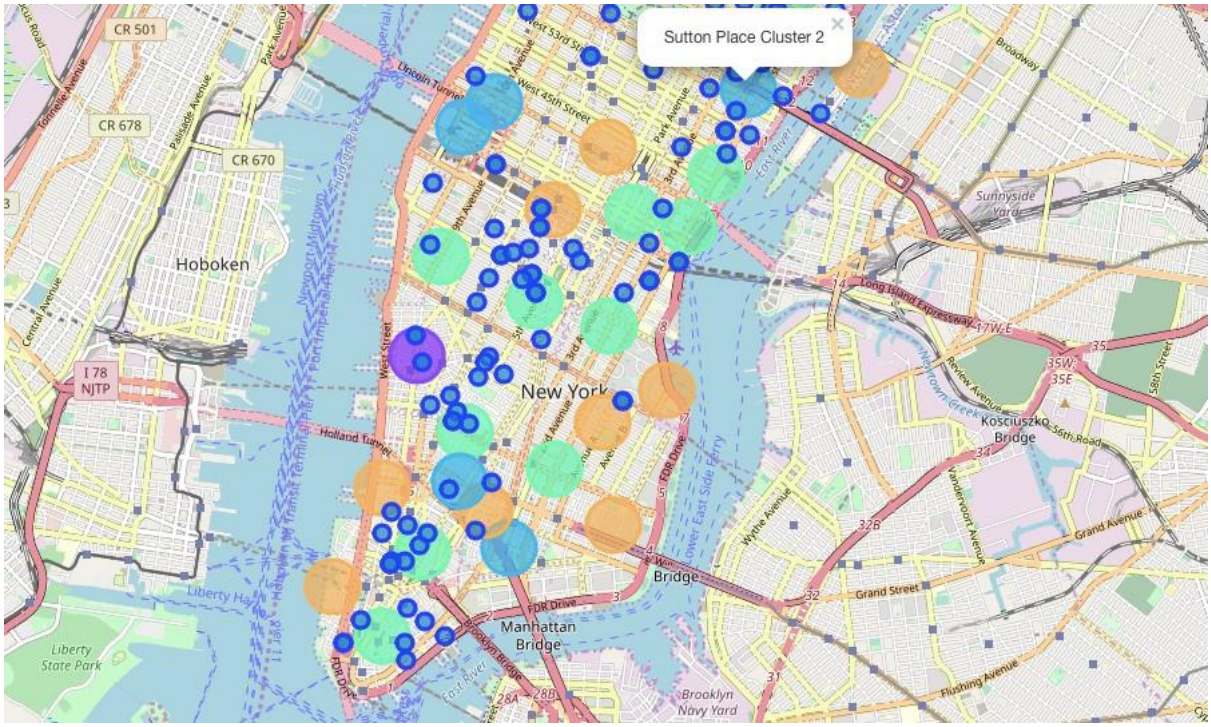
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a25f2a2b0>
```



# Apartments for Rent in Manhattan




## Manhattan apt for Rent-Clusters



# Manhattan Subway Stations GeoData

click to scroll output; double click to hide

| | | 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 |

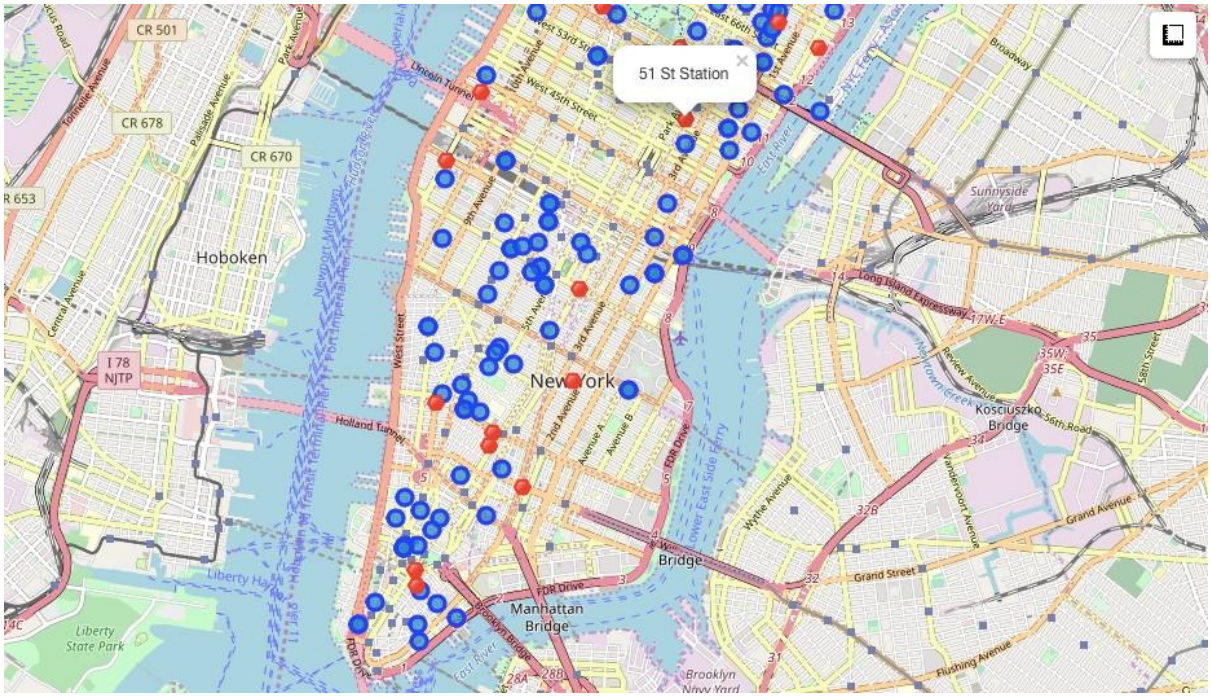
```
# removing duplicate rows and creating new set mhsub1
mhsub1=mh.drop_duplicates(subset=['lat','long'], keep="last").reset_index(drop=True)
mhsub1.shape
```

```
(22, 4)
```

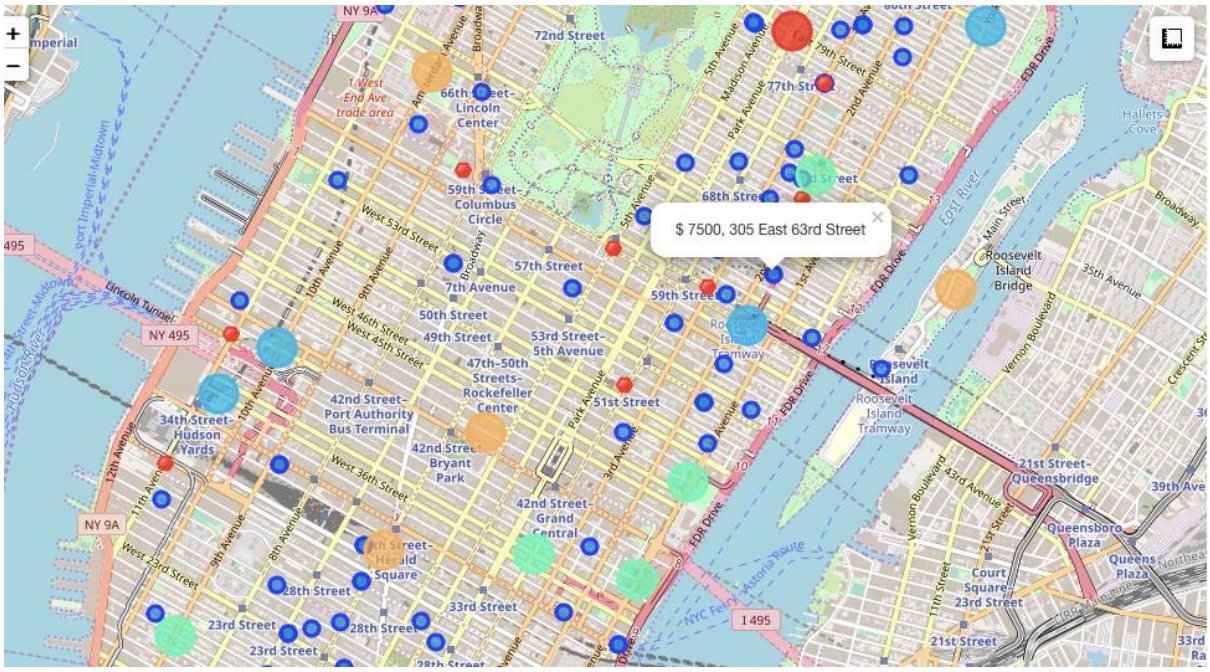
```
mhsub1.tail()
```

| | 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 |

# APT for Rent



## Selected Apartment




The one which is shown in the Graph is the selected Apartment.

# Apartment Selection

→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**.**

# DISCUSSION

**→I'm impressed with the overall 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 in this Course.**

**→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.**

# CONCLUSION

**→I feel rewarded with the efforts 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.**

---

Thanks COURSERA/IBM who helped me in this Course.

I also thank each and every friend who involved in my success by corrections.