Capstone Project

Coursera IBM Data Science Certification

Tarun Kumar

Content

1. Introduction
2. Data and Method section
3. Execution and Result/output section
4. Discussion & Conclusion

1.Introduction Section:

- Description of the Problem and Background

I am a IT Professional pursuing data science course from Coursera. I have been looking for an opportunity to work as Data scientist for analytic company based in New York.
Problem Statement : How can I find a convenient and less affordable place in New York with all amenities, I like to use and apply the tools and analytic technique which learned during the course for data science. In order to make a comparison and evaluation of the rental options in New York, I am considering city Manhattan in New York.

- Business Problem: The challenge is to find a suitable apartment for rent in NY that complies with the demands on location, price and venues.

  How to find an apartment/house for rent in Manhattan. Below are criteria.

  1.Studio or one-two bedroom apartment.

  2.Near to metro station(2 km) radius.

  3.Rent $2,500-4,000 per month.

  4.Food, coffee shops, restaurants Asian Thai, wine stores, gym, shopping and Health Clinic.

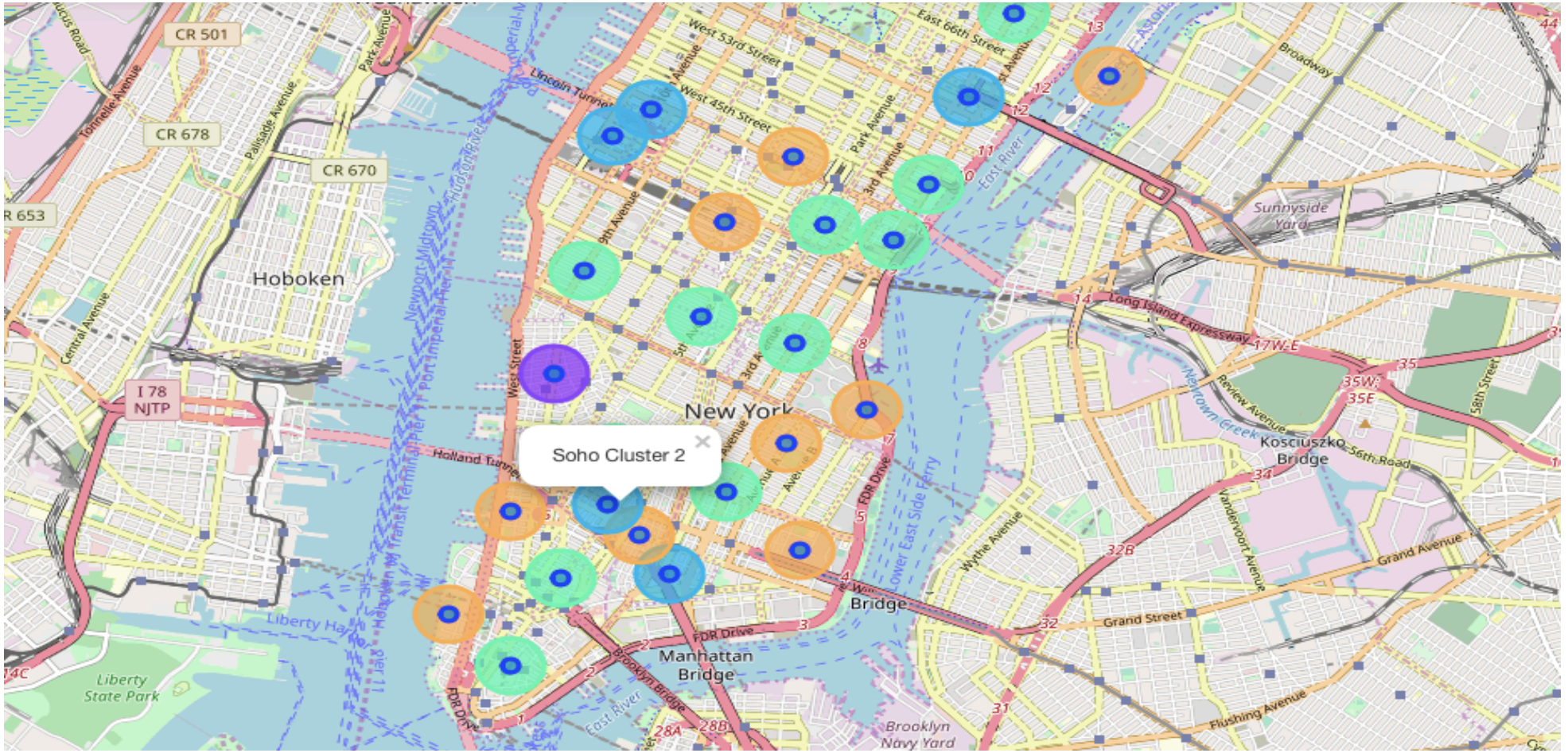
- Audience: This is a common challenge with valid questions for anyone moving to large city like New York in US. The same methodology can be applied in accordance to demands as applicable. This case is also applicable for anyone interested in exploring city in other country like Europe and Asia. This can be used as practice skills for data scientist.

2.DATA AND METHOD SECTION:

- Description of the data and its sources that will be used to solve the problem
- Description of the Data: The following data is required to answer the issues of the problem:
- List of Boroughs and neighborhoods of Manhattan with their geodata including latitude and longitude.
- List of metro stations in Manhattan with their address.
- List of apartments for rent in Manhattan area with their addresses and price range with additional information, such number of beds, etc.
- List of Venues for each Manhattan neighborhood
- List of Venues for metro stations.
- **Method to solve:**
- 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 20 venues for all Manhattan neighborhoods and cluster in groups.
- Use foursquare and geopy data to map the location of metro stations, with clustered map which will be able to identify the venues and amenities near each metro station.
- Use Foursquare and geopy data to map the location of rental places, in some form, linked to the near metro stations.
- create a map that depicts, for instance, the average rental price, around a radius of 2 km near each metro stations.
- I will be able to quickly point out the popups to know the relative price per metro station area. Addresses from rental locations will be converted to geodata(latitude , longitude) using Geopy-distance.
- Data will be searched in open data sources if available, from real estate sites if open to reading, libraries or other government agencies, etc.

3. Execution and Result/output

Manhattan Map - Neighborhoods and Cluster of Venues



Manhattan Geodata latitude longitude

```
]: 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

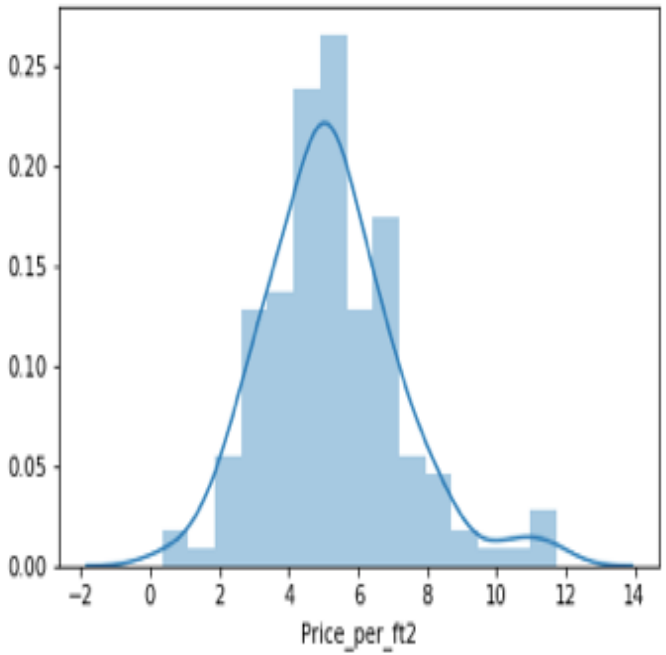
```
]: mh_rent.tail()
```

]:

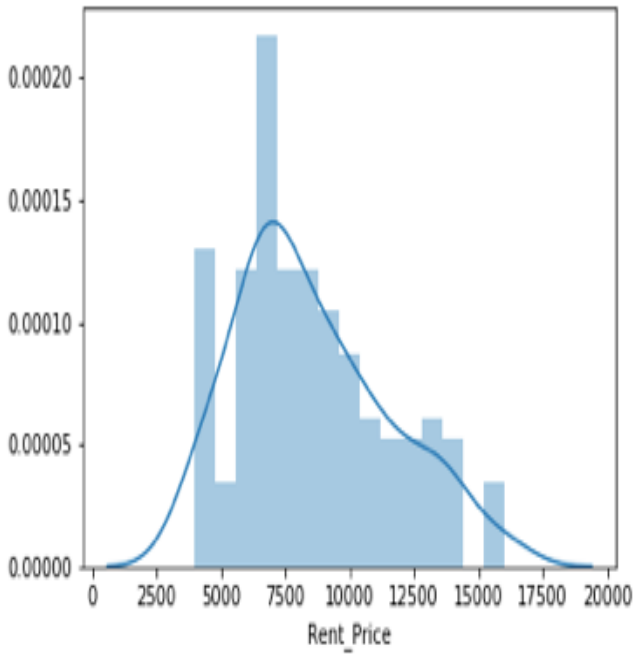
	Address	Area	Price_per_ft2	Rooms	Area-ft2	Rent_Price	Lat	Long
139	200 East 72nd Street	Rental in Lenox Hill	5.15	3.0	1700	8750	40.769465	-73.960339
140	50 Murray Street	No fee rental in Tribeca	7.11	2.0	1223	8700	40.714051	-74.009608
141	300 East 56th Street	No fee rental in Midtown East	3.87	3.0	2100	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

Manhattan rental statistics

```
<matplotlib.axes._subplots.AxesSubplot at 0x1a241
```

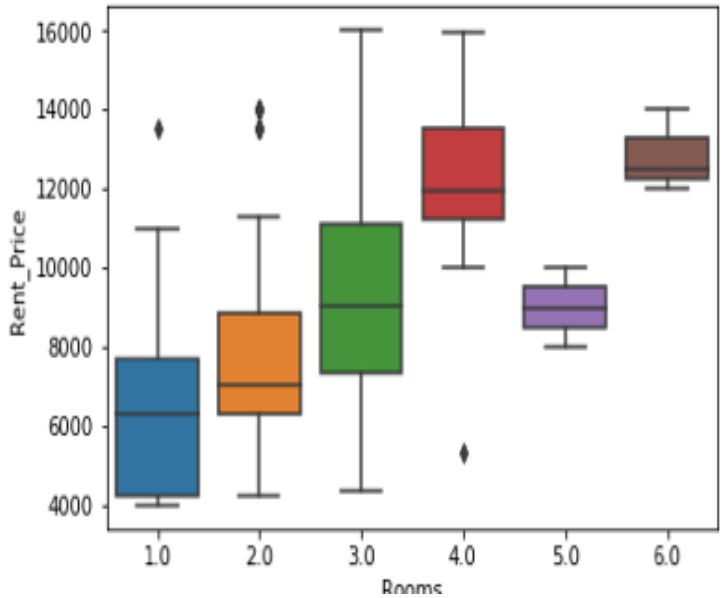



```
<matplotlib.axes._subplots.AxesSubplot at 0x1a25dd8400>
```




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

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

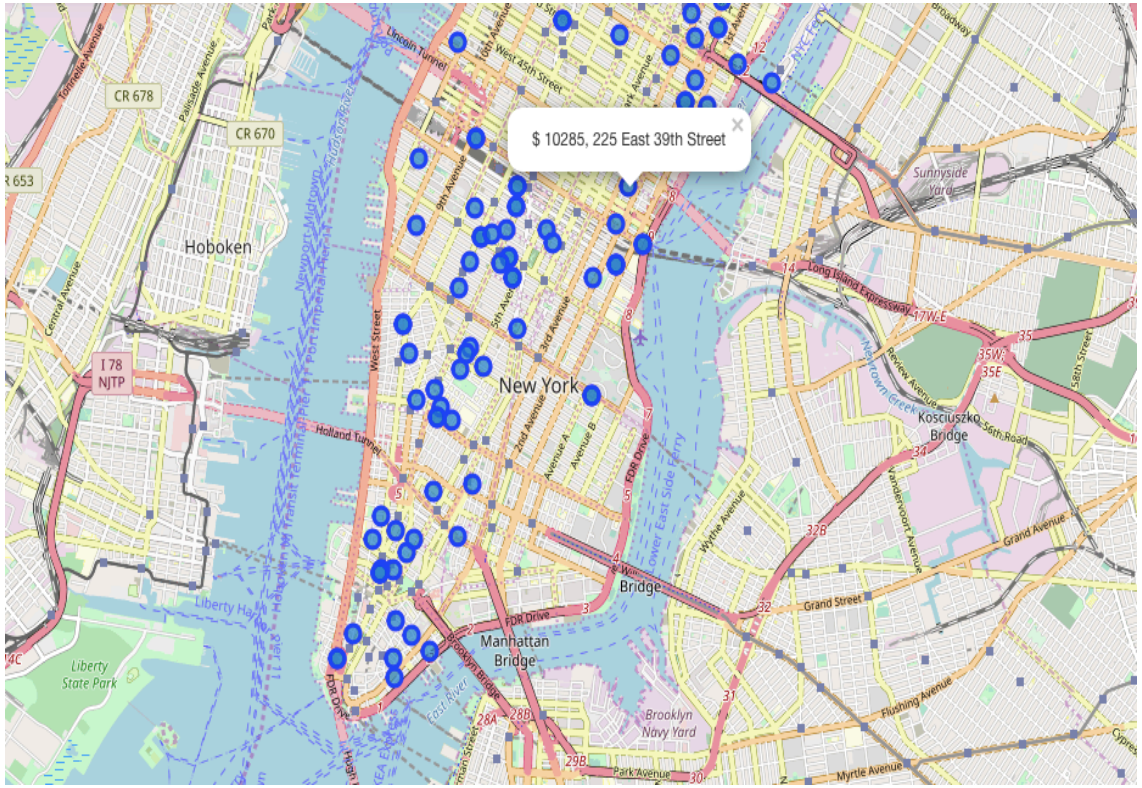


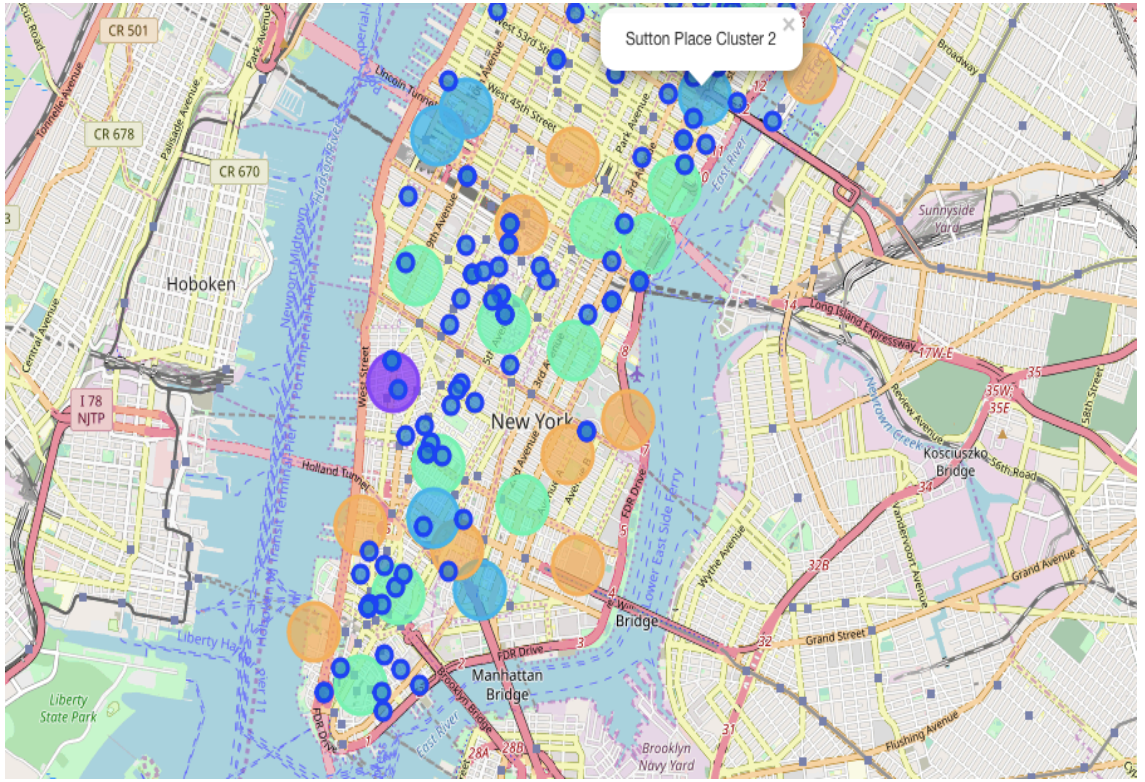
Manhattan venues of cluster (3)

```
## 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 Apartment for rent with venues cluster





Manhattan metro station & rental apartments

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

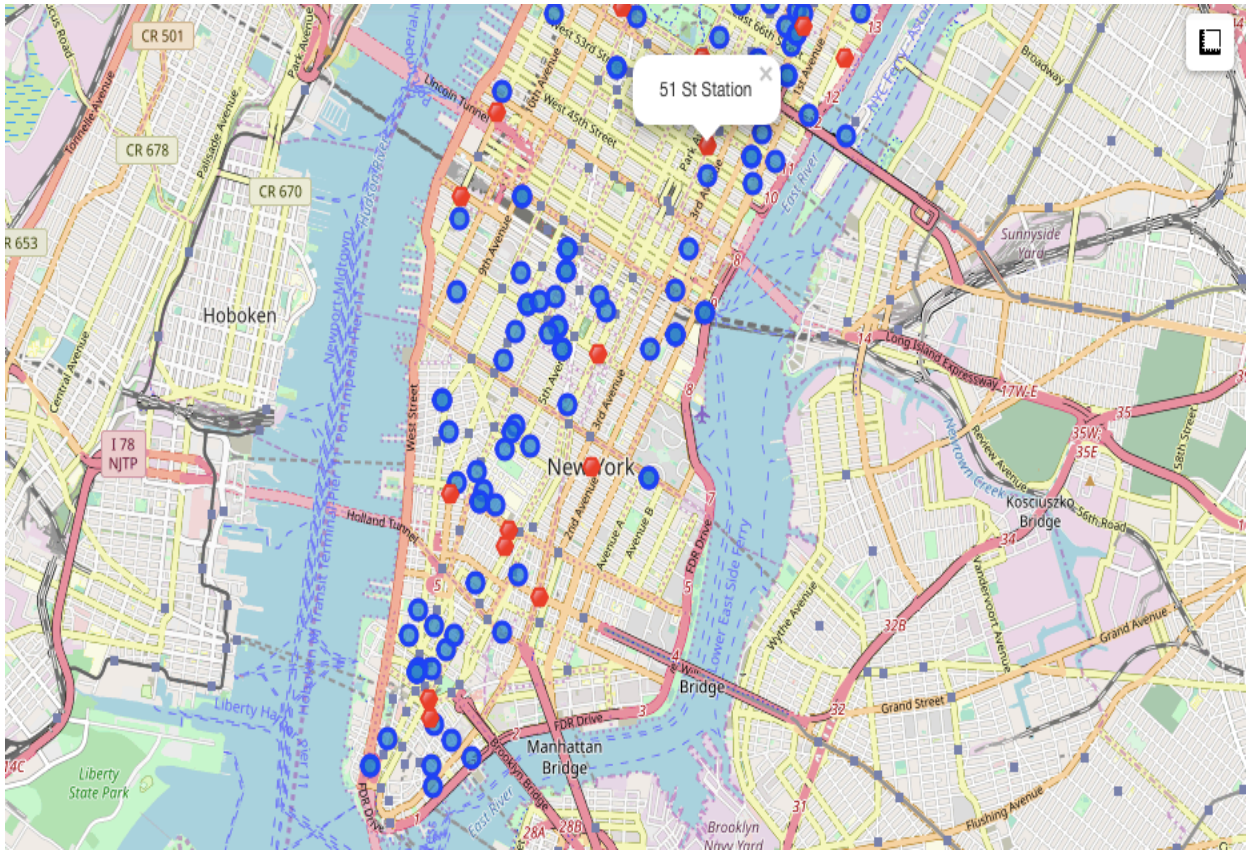
```
# removing duplicate rows and creating new set mhsub1
mhsub1=mh.drop_duplicates(subset=['lat','long'], keep="last").reset_index(drop=True)
mhsub1.shape
```

```
(22, 4)
```

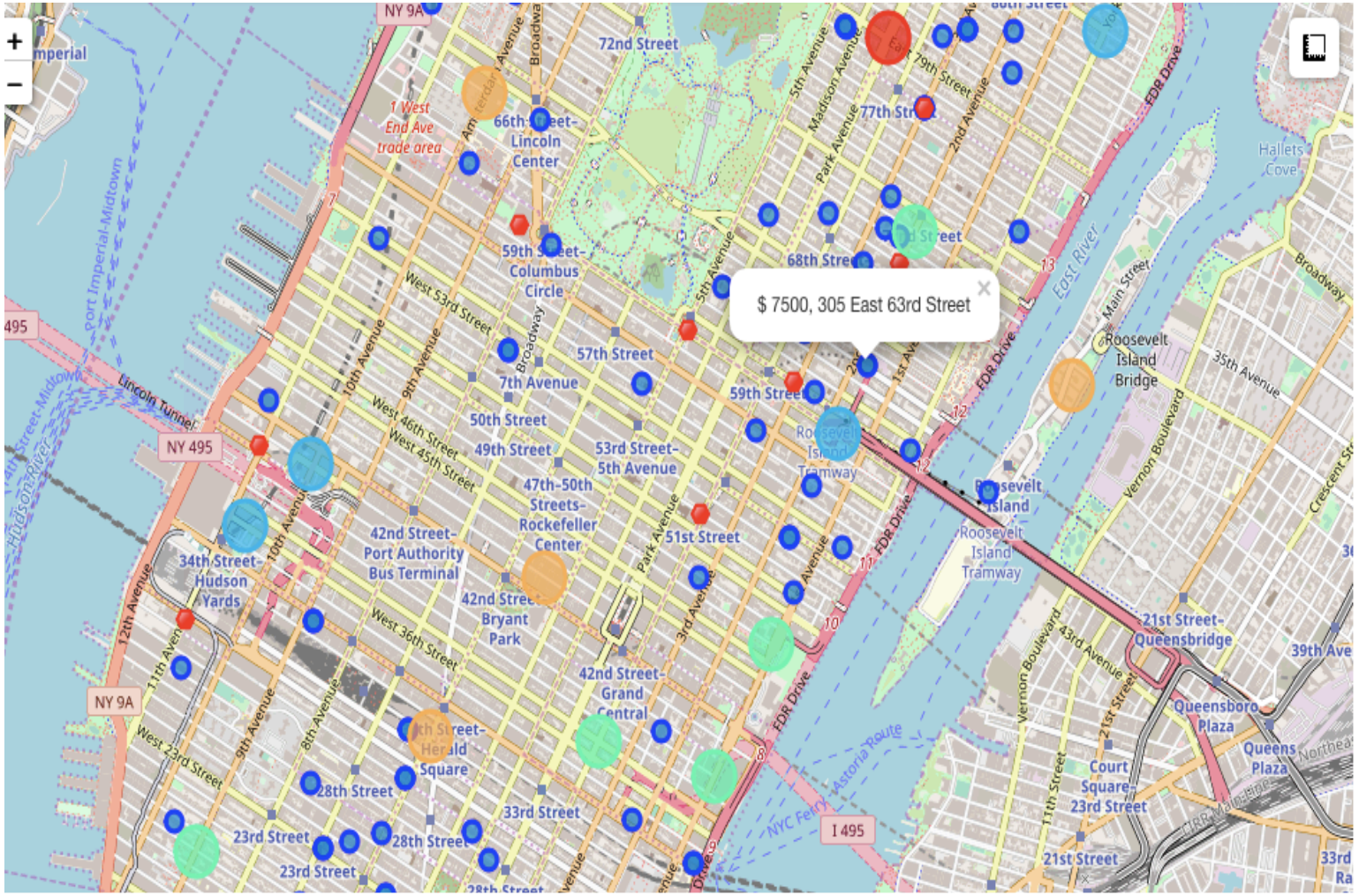
```
mhsub1.tail()
```

	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

- Red dot –Metro station
- Blue dot –Apartment for rent



Selected Apartment on one map shows
all information Apartments address, price, neighborhood, cluster of venues and metro station



- Red dot –Metro station
- Blue dot –Apartment for rent
- Bubbles – Cluster of Venues

- Apartment selection can done by Selecting the blue dot over all Area of Manhattan to find suitable Rental, selected apartment show Higher rental which is above the Budget.

4.Discussion & Conclusion

- DISCUSSION: 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.
- CONCLUSIONS: 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.

Thank You