

Coursera Capstone Project

Presentation

Coursera IBM Data Science Certification

Xavier Pamphile

01/05/2020

Report Content

1 Introduction Section :

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

2 Data Section:

- 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 learning, 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:

- Decision taken and Report Conclusion.

1 Introduction

- **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 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

- **2.1 Data Requirements**

- Geodata for current residence in Singapore with venues established using Foursquare.
- List of Manhattan (MH) neighborhoods with clustered venues established via Foursquare :
https://en.wikipedia.org/wiki/List_of_Manhattan_neighborhoods
- A list of Manhattan subway metro stations with addresses and geodata (latitude, longitude):
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.

- **2.2 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 website and Google map.
 - List of apartments for rent was consolidated from web-scraping real estate sites for MH.
- The geolocation (latitude, longitude) 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 apartments.

3 Methodology

The strategy to find the answer:

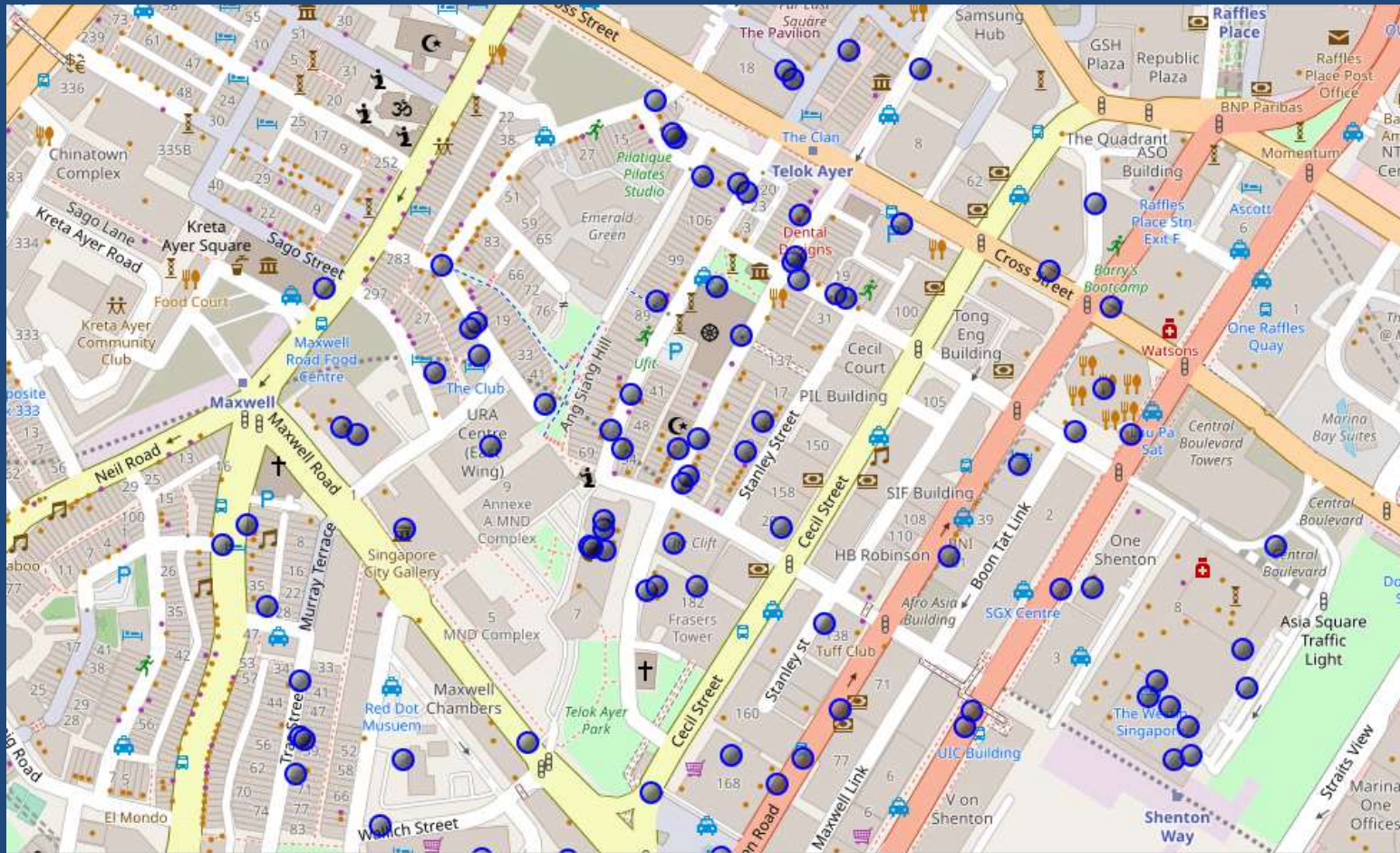
The strategy is based on mapping the described data in section 2.0 in order to facilitate the choice of at least two candidate places for rent. The information will be consolidated in one map where one can see the details of the apartment, the cluster of venues in the neighborhood and the relative location from a subway station and from a work place. A measurement tool icon will also be provided. The pop-ups 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 csv files for convenience and to simplify the report. Geodata was obtained by coding a program to use Nominatim to get the 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 pop-ups labels allow quick identification of location, price and feature, thus making the selection very easy.

4 Execution and results

Current residence neighborhood in Singapore



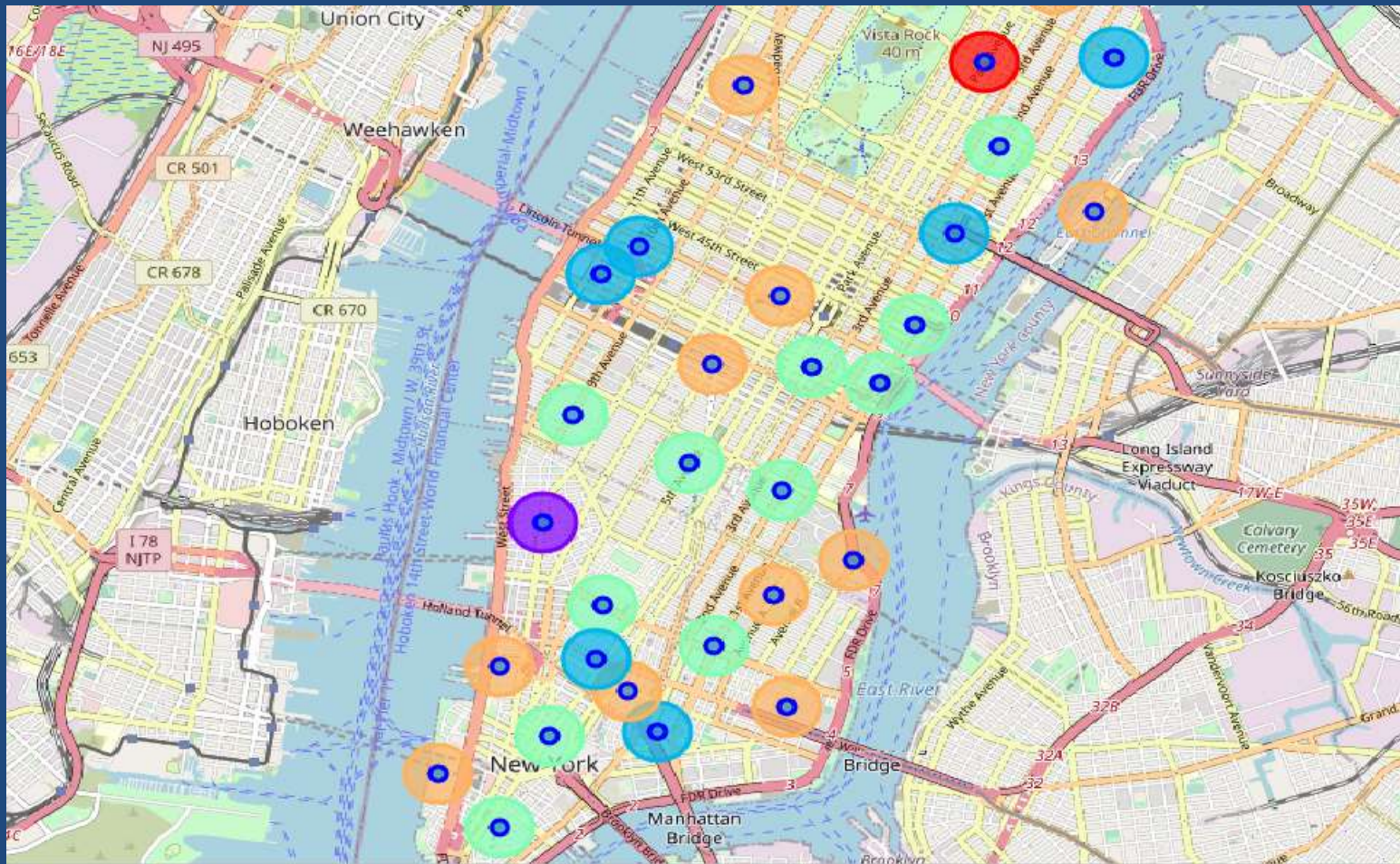
Venues around my neighborhood in Singapore

```
[11]: # Venues near current Singapore residence place  
SGnearby_venues.head(10)
```

```
[11]:
```

	name	categories	lat	lng
0	Napoleon Food & Wine Bar	Wine Bar	1.279925	103.847333
1	Native	Cocktail Bar	1.280135	103.846844
2	Freehouse	Beer Garden	1.281254	103.848513
3	Sofitel So Singapore	Hotel	1.280017	103.849813
4	Park Bench Deli	Deli / Bodega	1.279872	103.847287
5	Pepper Bowl	Asian Restaurant	1.279371	103.846710
6	Meat Smith	Southern / Soul Food Restaurant	1.280205	103.847410
7	Mellower Coffee	Café	1.277814	103.848188
8	PS.Cafe	Café	1.280468	103.846264
9	Anglo Indian Cafe & Bar	Indian Restaurant	1.279084	103.850127

Manhattan map – Neighborhoods and cluster of venues



Geodata

Manhattan apartments for rent

```
[20]: mh_rent=pd.read_csv('MH_rent_latlong.csv')
mh_rent.head()
```

```
[20]:
```

	Address	Area	Price_per_ft2	Rooms	Area-ft2	Rent_Price	Lat	Long
0	West 105th Street	Upper West Side	2.94	5.0	3400.0	10000.0	40.730862	-73.987156
1	East 97th Street	Upper East Side	3.57	3.0	2100.0	7500.0	40.730862	-73.987156
2	West 105th Street	Upper West Side	1.89	4.0	2800.0	5300.0	40.730862	-73.987156
3	CARMINE ST.	West Village	3.03	2.0	1650.0	5000.0	40.730862	-73.987156
4	171 W 23RD ST.	Chelsea	3.45	2.0	1450.0	5000.0	40.730862	-73.987156

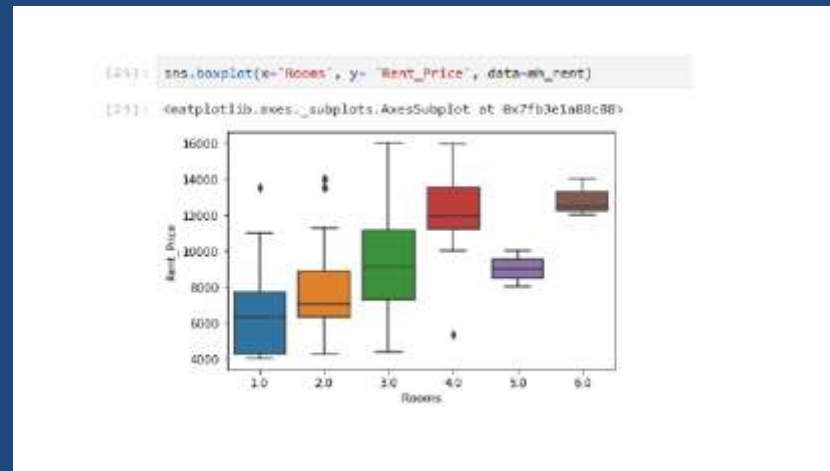
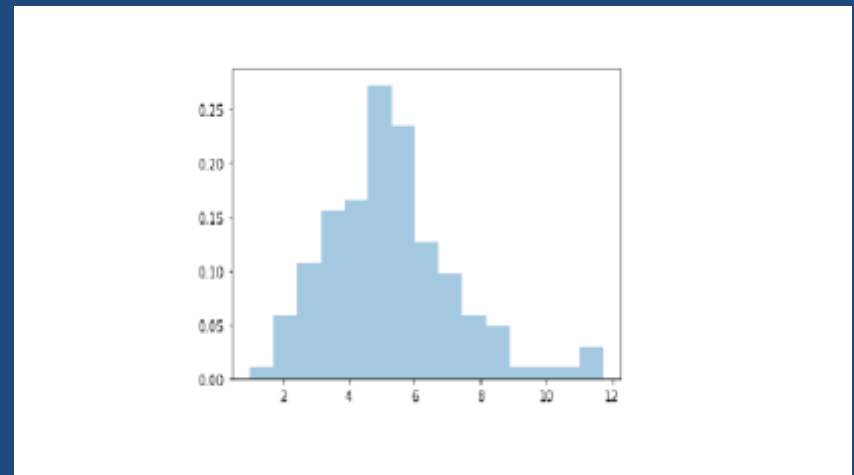
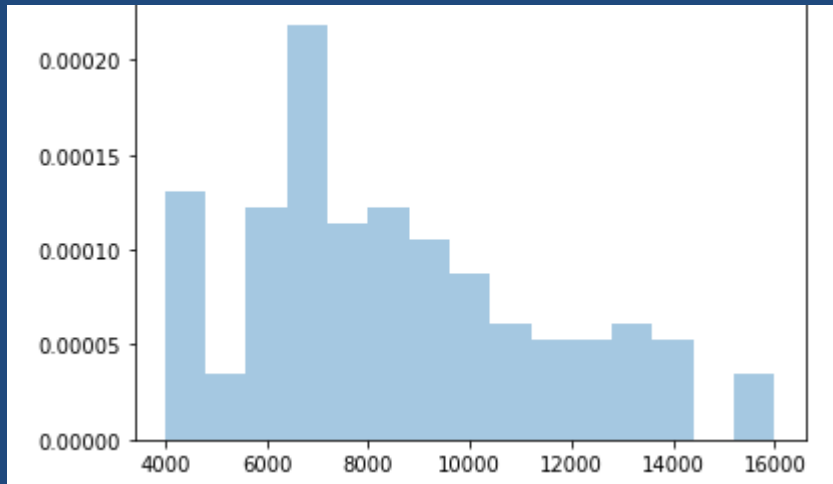
```
[21]: mh_rent.tail()
```

```
[21]:
```

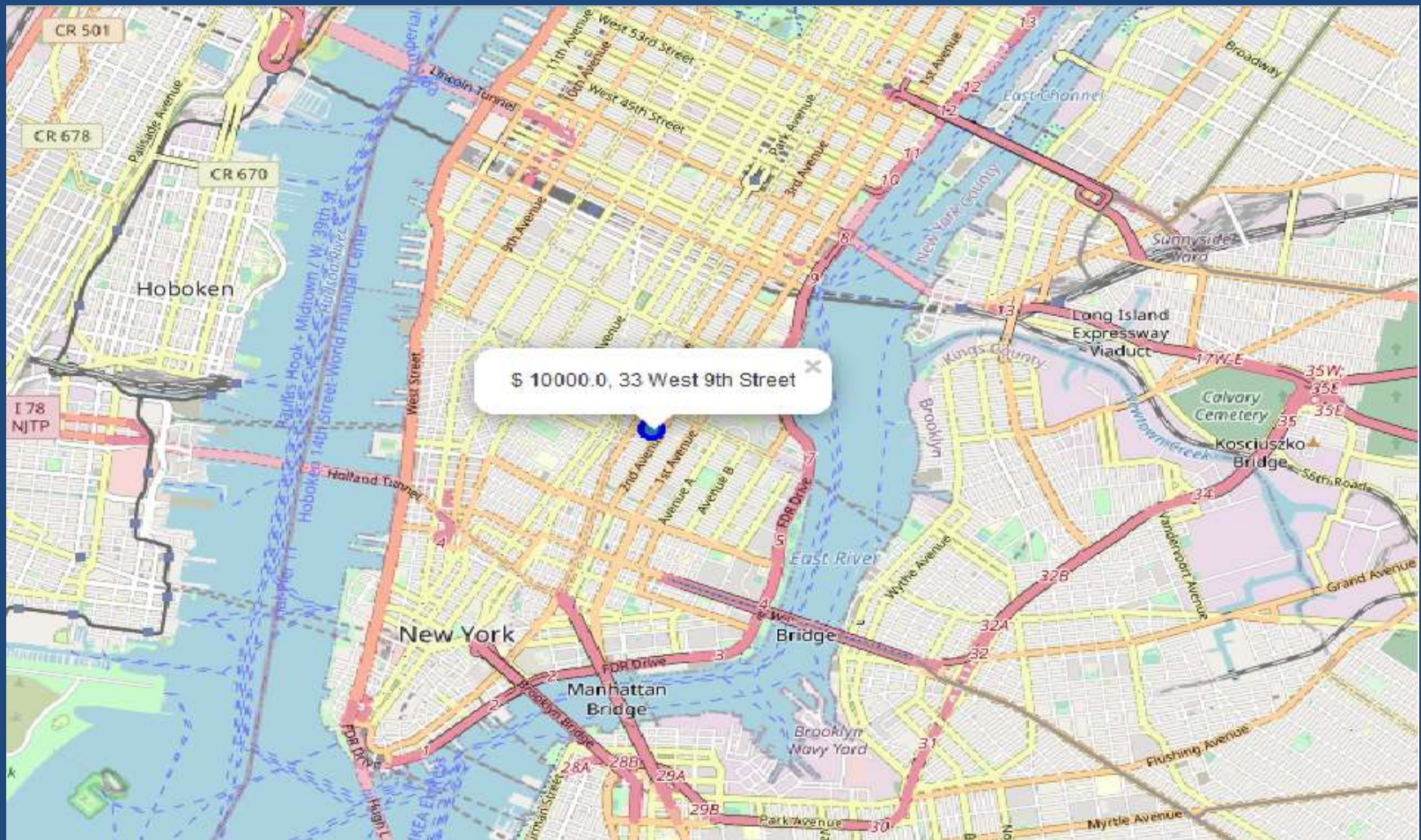
	Address	Area	Price_per_ft2	Rooms	Area-ft2	Rent_Price	Lat	Long
139	200 East 72nd Street	Rental in Lenox Hill	5.15	3.0	1700.0	8750.0	40.730862	-73.987156
140	50 Murray Street	No fee rental in Tribeca	7.11	2.0	1223.0	8700.0	40.730862	-73.987156
141	300 East 56th Street	No fee rental in Midtown East	3.87	3.0	2100.0	8118.0	40.730862	-73.987156
142	1930 Broadway	No fee rental in Central Park West	5.06	2.0	1600.0	8095.0	40.730862	-73.987156
143	33 West 9th Street	Rental in Greenwich Village	6.67	2.0	1500.0	10000.0	40.730862	-73.987156

Rental price statistics MH apartments

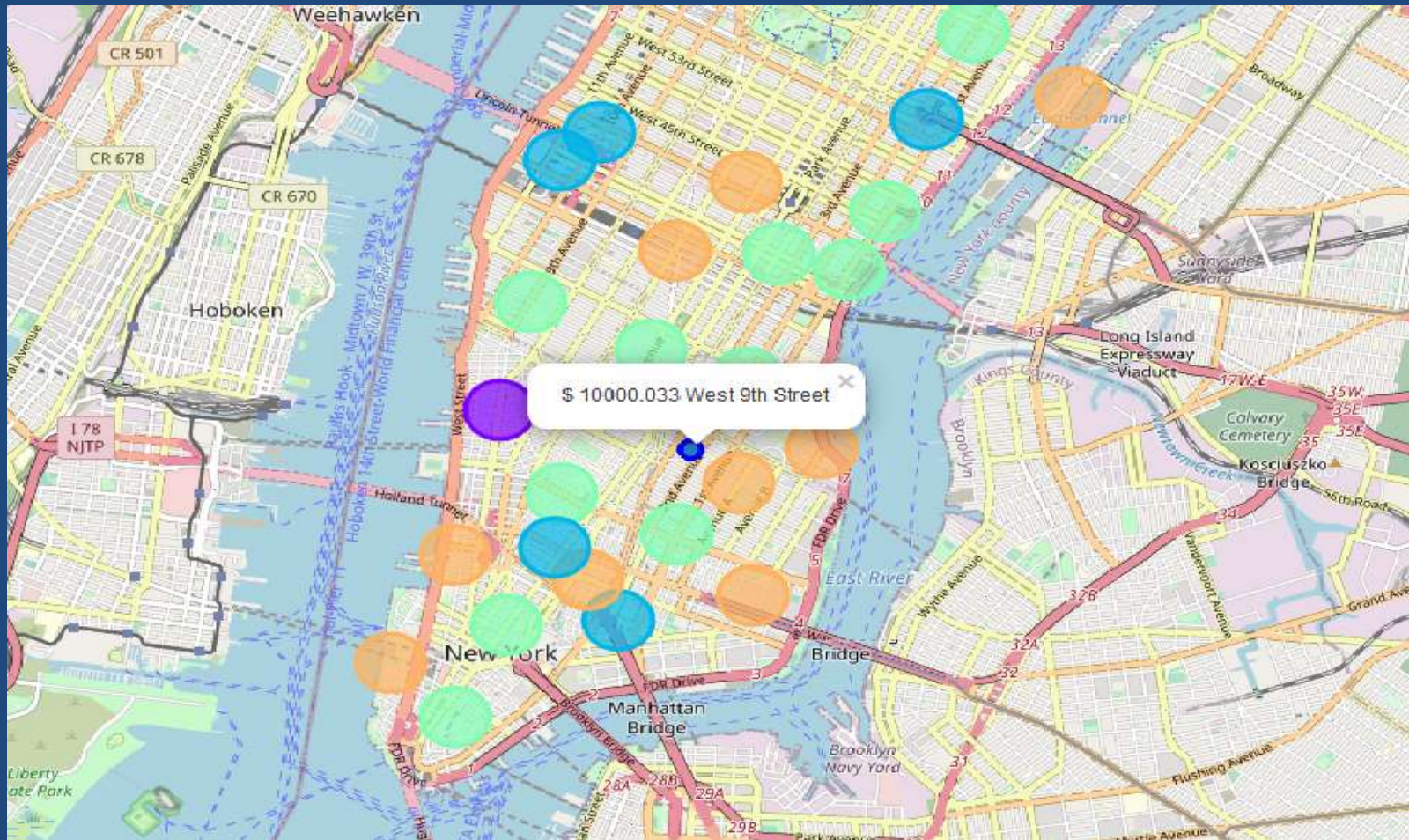
Budget US7000/month is around the mean



Apartments for rent in MH



Apartments for rent in MH with venues



Venues of cluster 3

```
[27]: ## 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_merged.shape[1]))]]
```

	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	Caf?	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	Caf?	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
32	Civic Center	Gym / Fitness Center	Bakery	Italian Restaurant	Cocktail Bar	French Restaurant	Sandwich Place	Coffee Shop	Gym	Yoga Studio	Park
35	Turtle Bay	Italian Restaurant	Coffee Shop	Steakhouse	Wine Bar	Sushi Restaurant	Hotel	Noodle House	Indian Restaurant	Japanese Restaurant	French Restaurant
36	Tudor City	Caf?	Park	Pizza Place	Mexican Restaurant	Greek Restaurant	Sushi Restaurant	Hotel	Deli / Bodega	Diner	Dog Run
38	Flatiron	Italian Restaurant	American Restaurant	Gym	Gym / Fitness Center	Yoga Studio	Vegetarian / Vegan Restaurant	Bakery	Clothing Store	Cosmetics Shop	Cycle Studio

Manhattan subway station Geodata

```
[30]:
```

	sub_station	sub_address	lat	long
0	Dyckman Street Subway Station	170 Nagle Ave New York NY 10034 USA	40.76429	-73.97282
1	57 Street Subway Station	New York NY 10106 USA	40.76429	-73.97282
2	Broad St	New York NY 10005 USA	40.76429	-73.97282
3	175 Street Station	807 W 177th St New York NY 10033 USA	40.76429	-73.97282
4	5 Av and 53 St	New York NY 10022 USA	40.76429	-73.97282

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

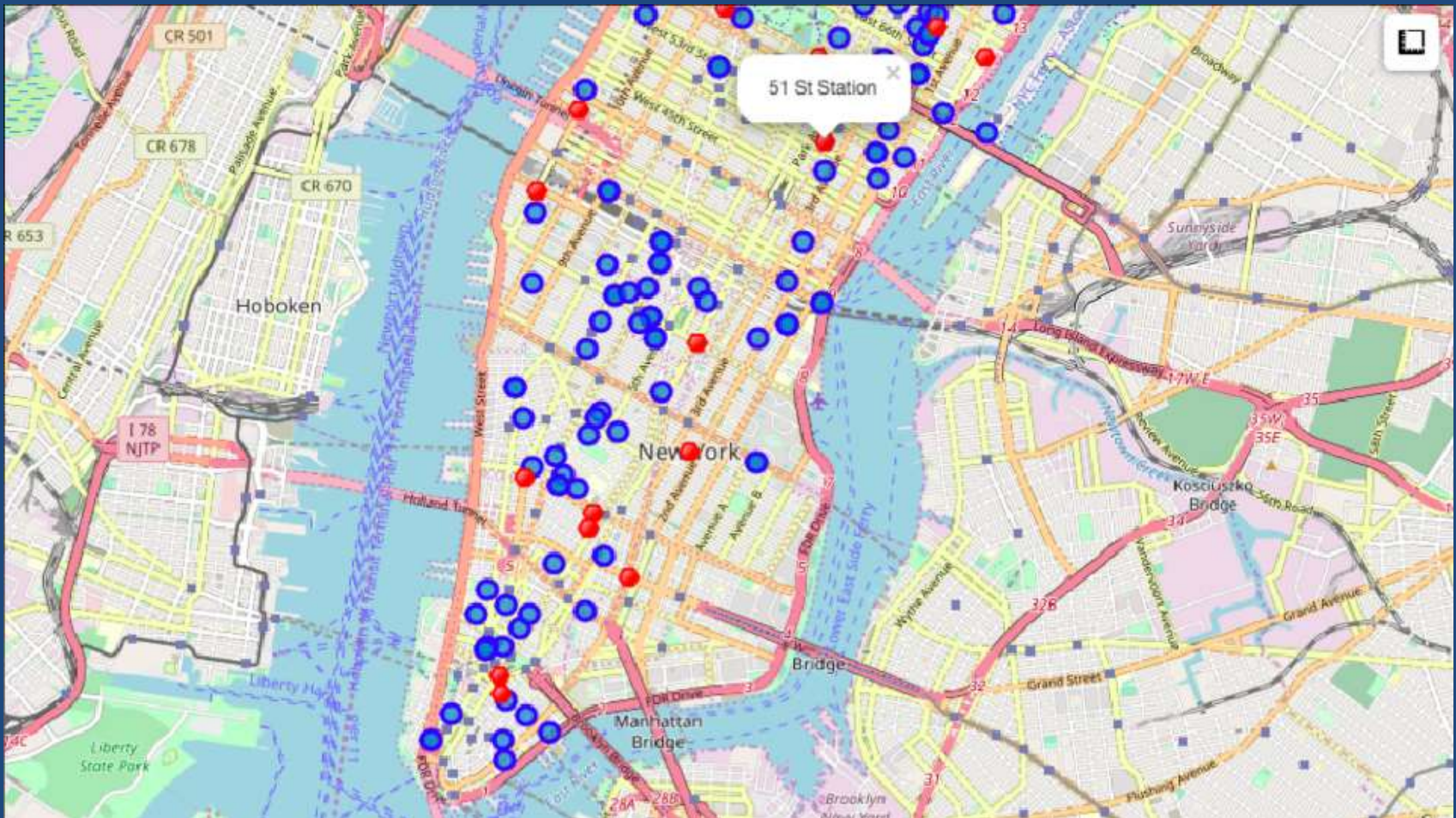
```
[32]: (76, 4)
```

```
[33]: mhsub1.tail()
```

```
[33]:
```

	sub_station	sub_address	lat	long
71	50 St Br	New York NY 10019 United States	40.76429	-73.97282
72	57 Street Station	New York NY 10019 United States	40.76429	-73.97282
73	23 Street Station	New York NY 10010 United States	40.76429	-73.97282
74	14 Street / 8 Av	New York NY 10014 United States	40.76429	-73.97282
75	MTA New York City	525 11th Ave New York NY 10018 USA	40.76429	-73.97282

Apartments for rent in blue and subway stations in red



Apartment Selection

- Using the "one map" above, I was able to explore all possibilities since the pop ups 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.

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.