Coursera Capstone project

Coursera IBM Data Science Certification

Abhijit R. Parab

Mar. 25th, 2020

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

1.1 Scenario and Background

I am currently living in India, Goregaon suburb in Mumbai city. I also enjoy great venues and attractions, such as international cuisine, entertainment and shopping. I have an offer to move to work to Manhattan NY
and I would like to move if I can find a place to live similar with similar venues.

1.2 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 (<=1.0 mile, 1.6 km) from a subway metro station in Manhattan
- Venues and amenities as in my current residence.

1.3 Interested Audience

I believe the methodology, tools and strategy used in this project is relevant for a person or entity considering moving to a major city in US, Europe or Asia. Europe, US or Asia, Likewise, it can be helpful approach to explore the opening of a new business. 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 for a person developing Data Science skills.

2.0 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 (as in Course Lab). 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://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-
- 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 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

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.

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

# 4.0 Execution and Results

My place in India

Venues around Neighborhood in

	name	categories	lat	lng
0	Satkar Rice Plate	Seafood Restaurant	19.165101	72.850714
1	Sai Veg World	Indian Restaurant	19.164901	72.850406
2	Sapre	Fast Food Restaurant	19.166621	72.847547
3	Anupam Stationery	Bookstore	19.165411	72.850501
4	Jay Prakash Restaurant & Bar	Bar	19.164831	72.850855
5	Shri Datta Snacks	Snack Place	19.163932	72.852673
6	Sudama Restaurant	Indian Restaurant	19.162545	72.846586
7	Jumbo Xerox	Design Studio	19.164172	72.851540
8	Udipi Vihar	Vegetarian / Vegan Restaurant	19.166498	72.854199
9	McDonald's	Fast Food Restaurant	19.165117	72.851015

Manhattan Map - Neighborhoods and Cluster of Venues
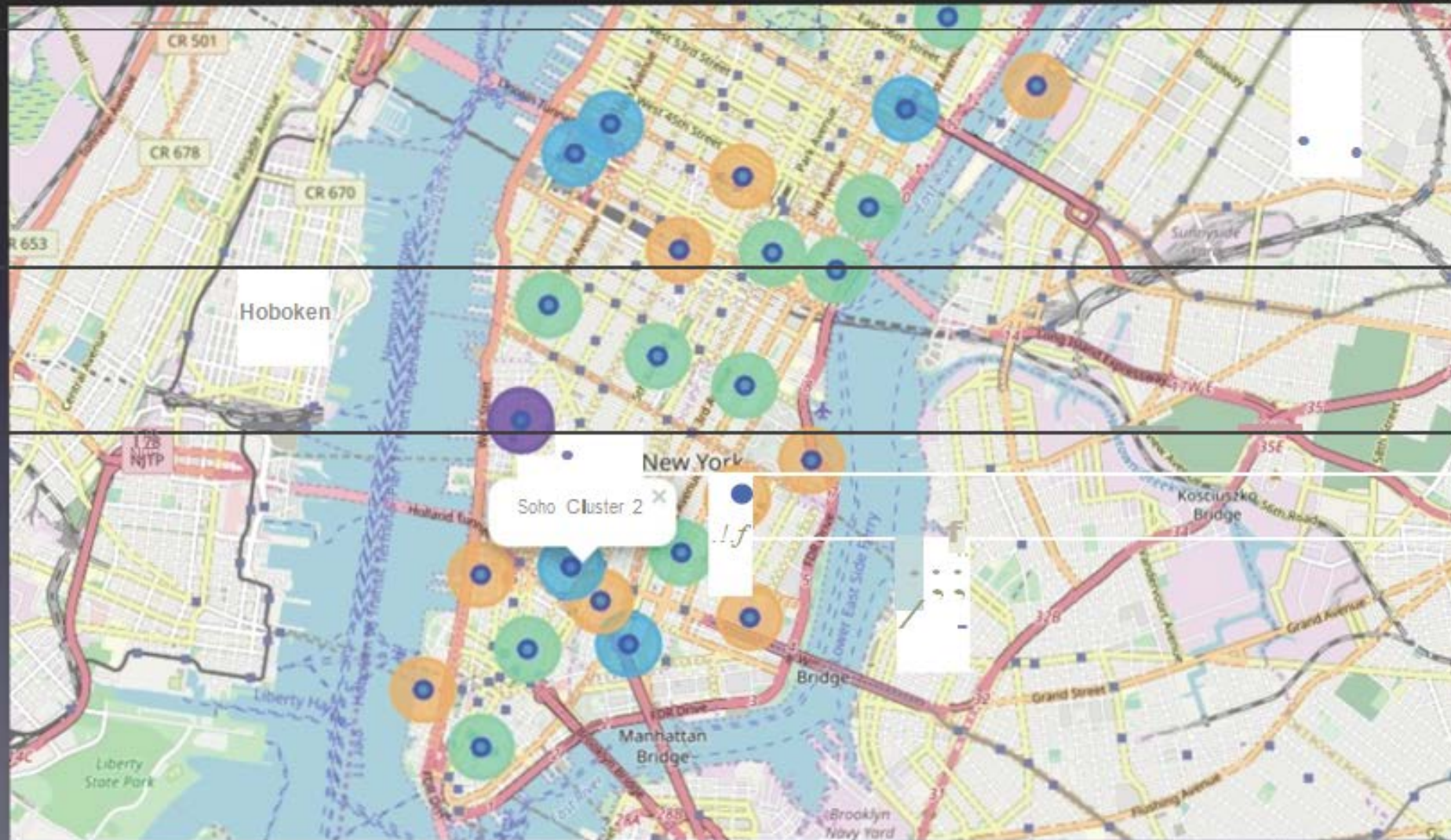
GeoData Manhattan apts for rent

```
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
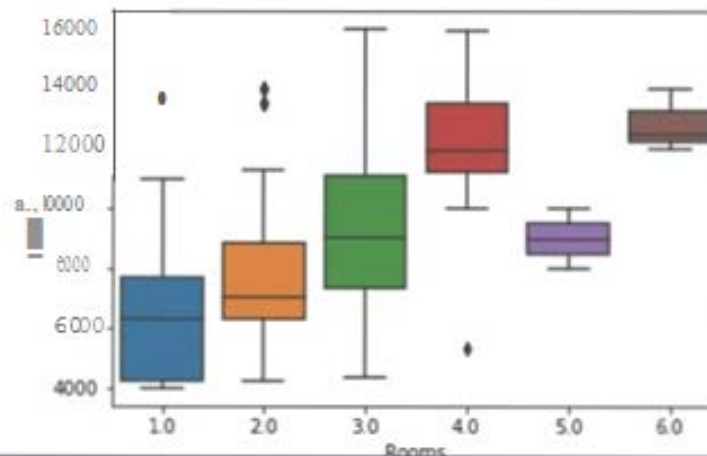## Budget US7000/month is around the mean
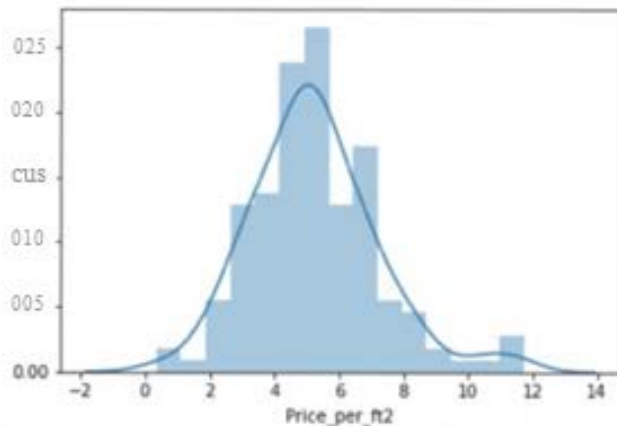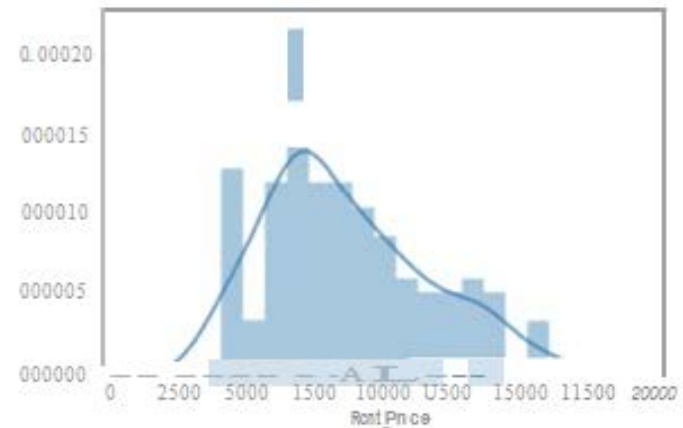
Apartments for Rent in MH

MH apts for rent with venue clusters
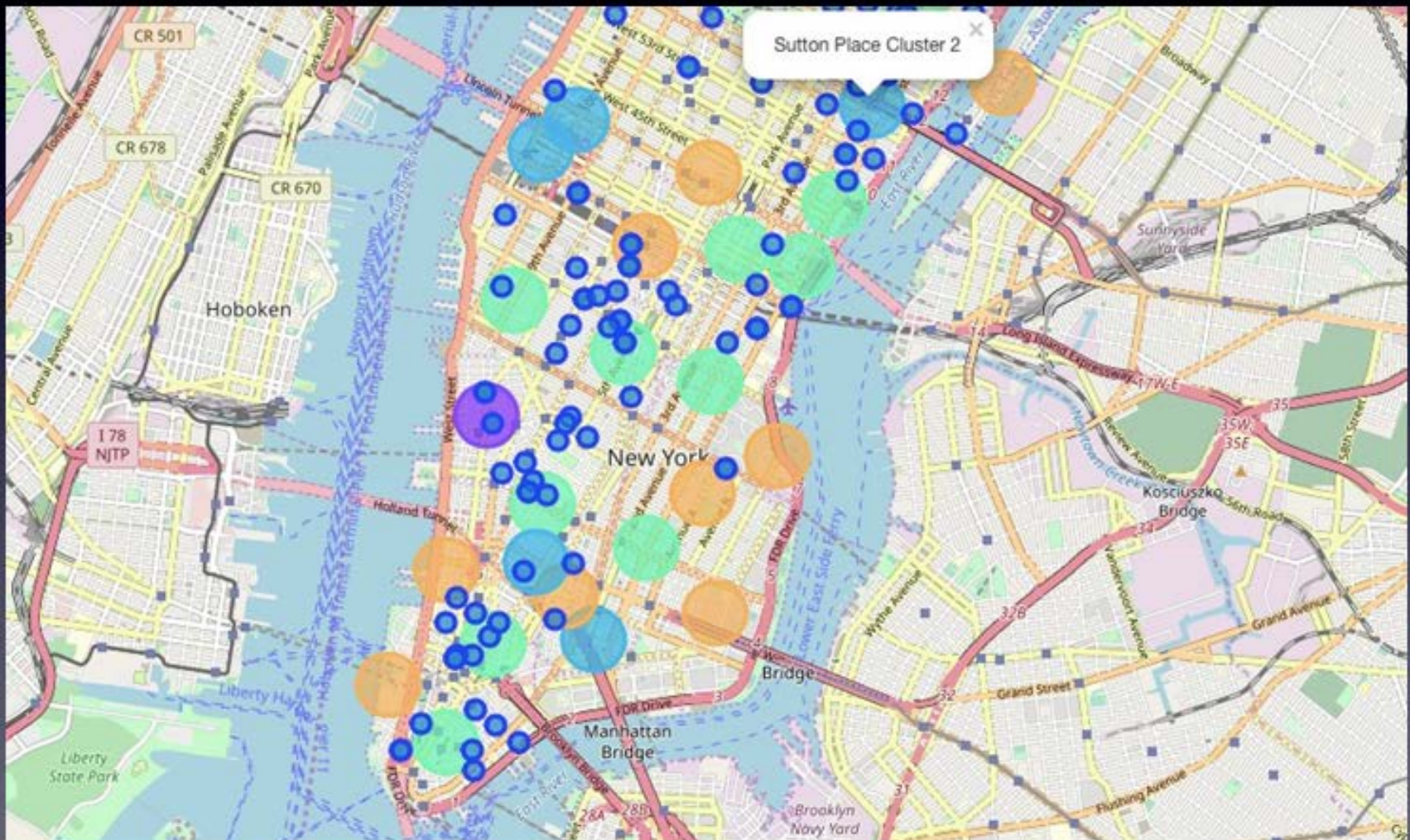
Manhattan subway stations geodata

		sub addntss	lat	long
0	Dyckman Stte« Subway S!lii!Oil	170 Nagle Ave.New York. NY 10034,USA	40.861857	-73.924509
1	57 Street Subway Slation	New Yofl(,NY 10106.USA	40.764250	-73.954525
2	Broad St	New York. NY 10005,USA	40.730862	-73.987156
3	175 Street Slation	807 W 1nth St, New YOI1<,NY 10033,USA	40.847991	-73.939785
4	5 Av IIId 53 St	New York. NY 10022.USA	40.784250	-73.954525

```
# removing duplicate rows and creating new set mbsubl
mhsubl-m.h.d.rop_duplicates(subset ('lat','long'], lteep-
"last") .reset_index(drop-True)
mhsubl.shape

(22, 4)
```

```
mhsubltail()
```

	sub station	aub_addrna	lat	long
17	190 Street Subway Station	Bennett Ave, New York,NY 10040,USA	40.858113	-73.932983
18	59 SH.exmgton Av Station	E 60th St, New Y011<,NY 10065, USA	40.762259	-73.966271
19	57 Street Station	New York,NY 100'9,United States	40.764250	-73.954525
20	14 Streel/8 Av	NewYOI1<,NY 10014,United States	40.730862	-73.987156
21	MTA New YOI1< C.ty	525 11th Ave,New York, NY 10018, USA	40.759809	-73.999282

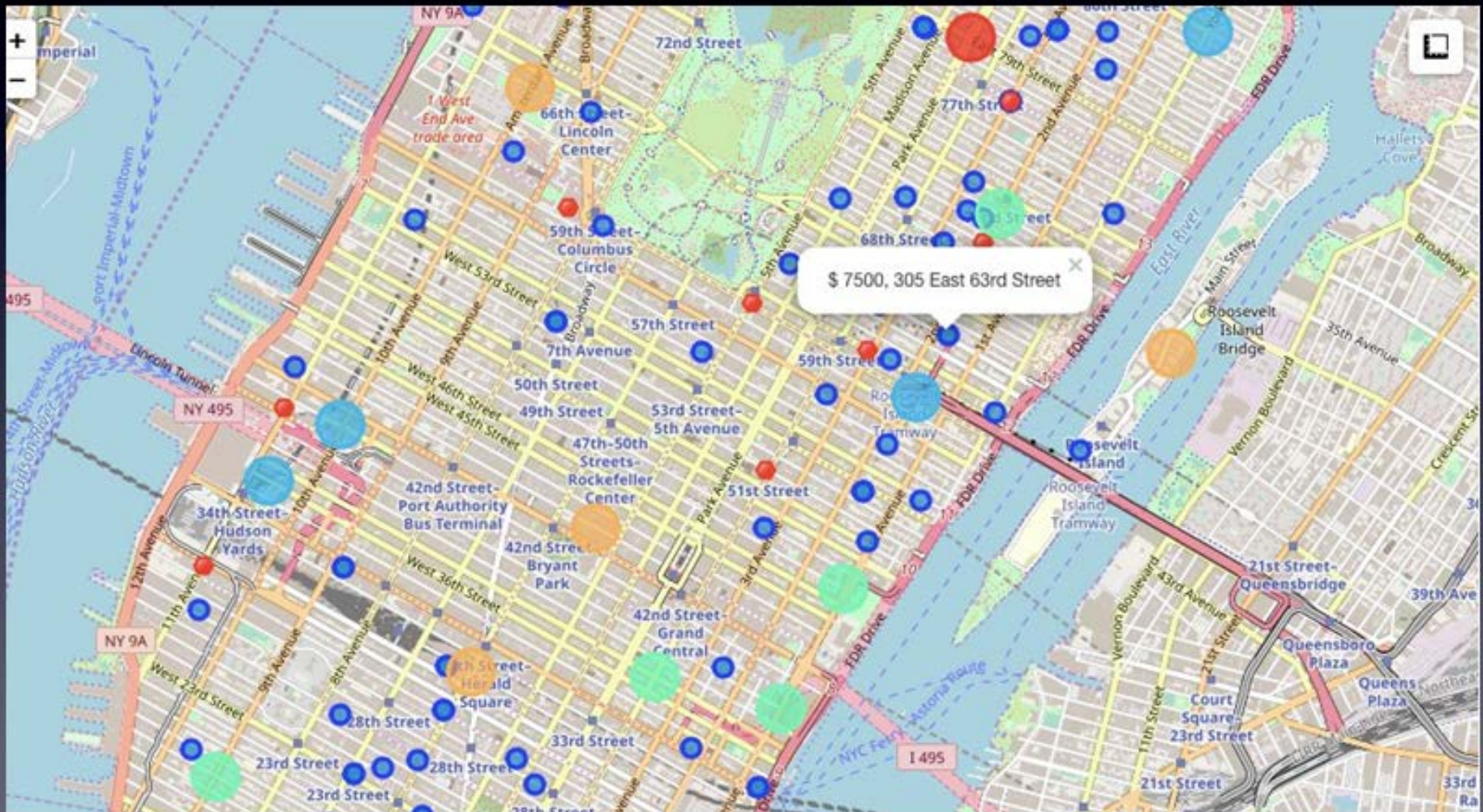# Apts for rent (blue) and subway stations (red)
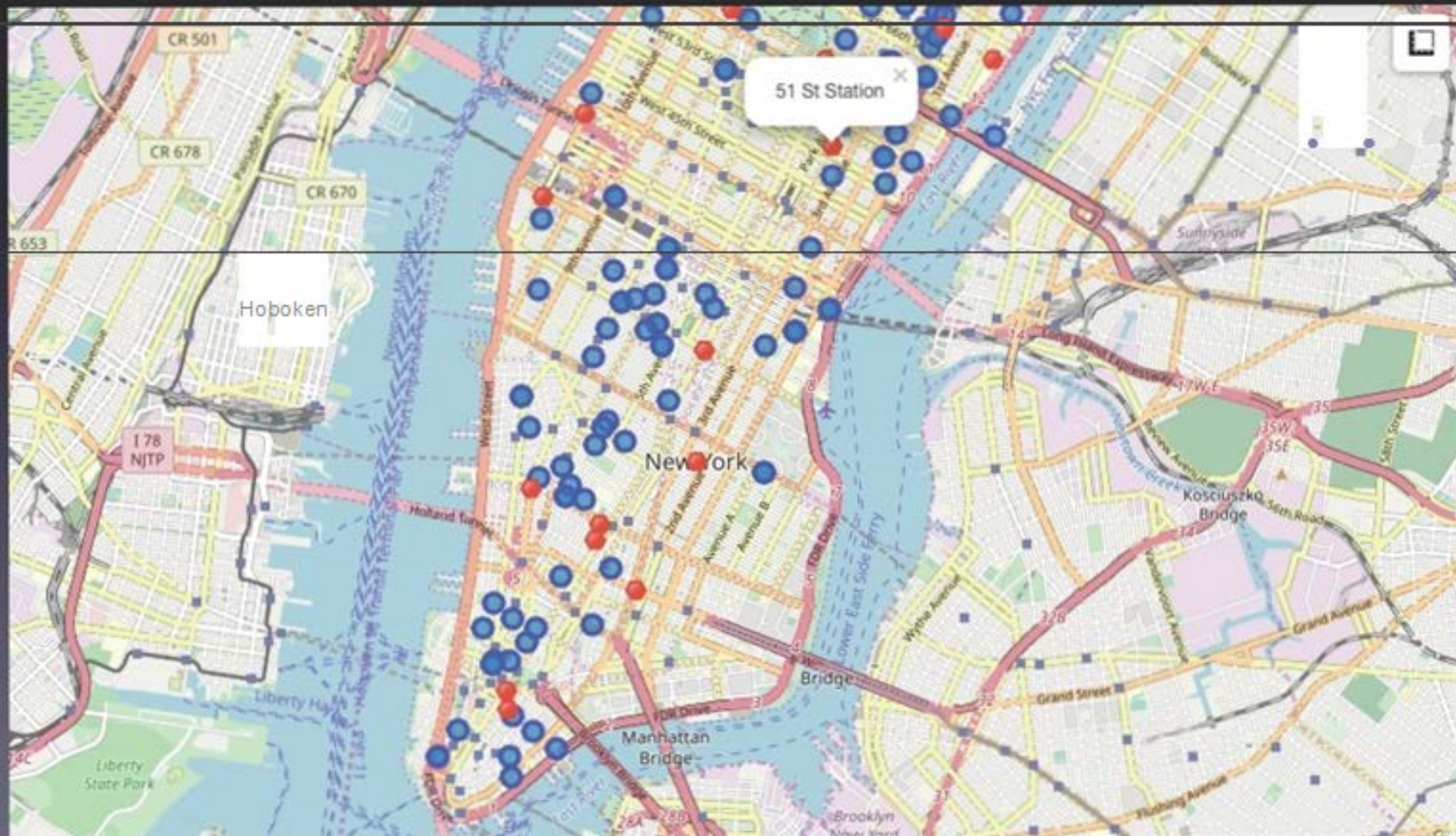
Selected Apartment!

The ONE consolidated map shows all information for decision:
Apartments address, price, neighborhood, 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 than 1 km!
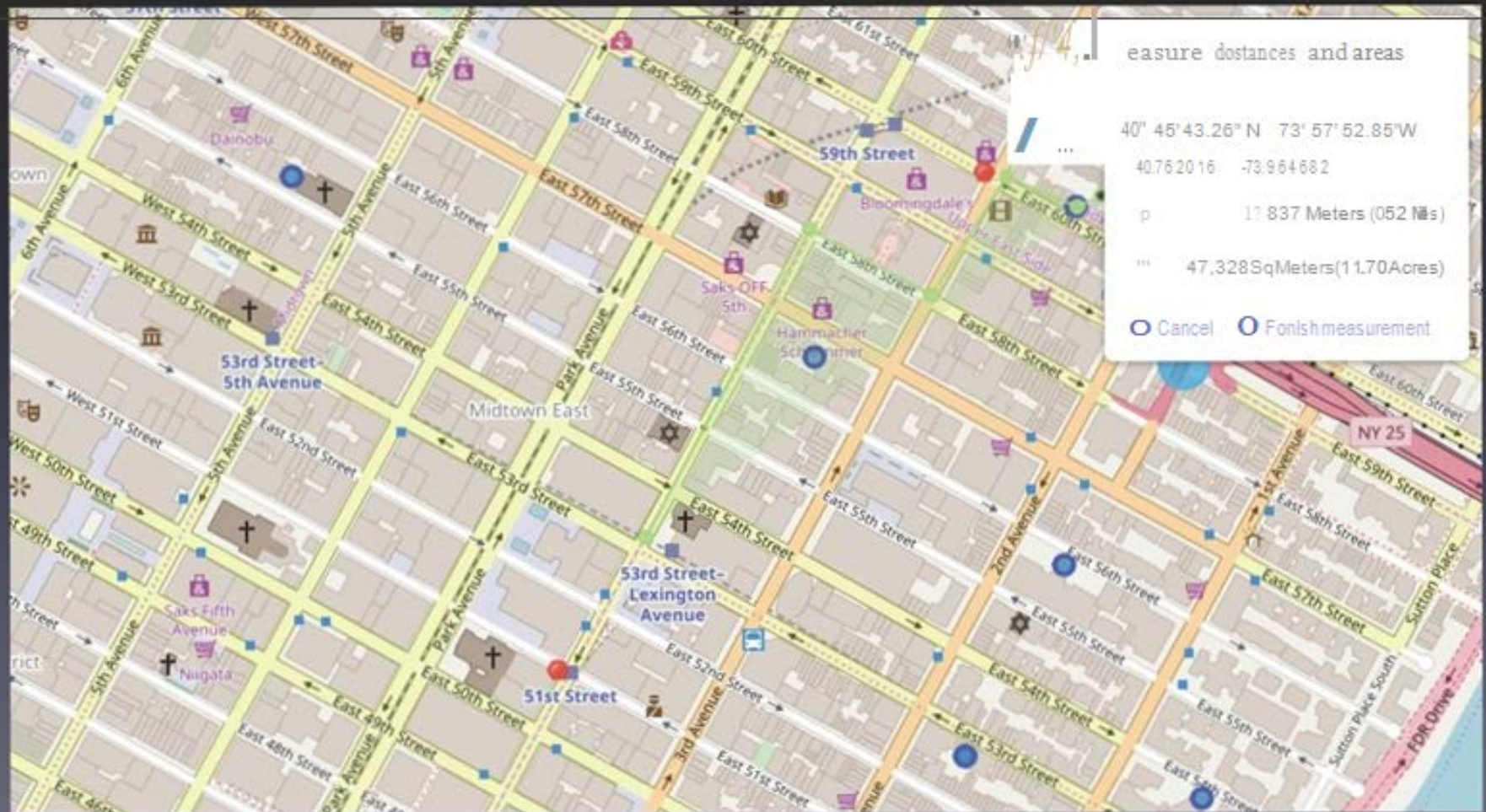
Venus in Cluster 2 near future home

```
## 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_n
```

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
MarbleHill	Coffee Shop	Discount Store	Yoga Studio	Steakhouse	Supplement Shop	Tennis Stadium	Shoe Store	Gym	Bank	Seafood Restaurant
Chinatown	Chinese Restaurant	Cocktail Bar	Dim Sum Restaurant	American Restaurant	Vietnamese Restaurant	Salon / Barbershop	Noodle House	Bakery	Bubble Tea Shop	Ice Cream Shop
Central Harlem	African Restaurant	Seafood Restaurant	French Restaurant	American Restaurant	Cosmetics Shop	Chinese Restaurant	Event Space	Liquor Store	Beer Bar	Gym/Fitness Center
Yorkville	Coffee Shop	Gym	Bar	Italian Restaurant	Sushi Restaurant	Pizza Place	Mexican Restaurant	Deli/Bodega	Japanese Restaurant	Pub
Clinton	Theater	Italian Restaurant	Coffee Shop	American Restaurant	Gym/Fitness Center	Hotel	Wine Shop	Spa	Gym	
Soho	Clothing Store	Boutique	Women's Store	Shoe Store	Men's Store	Furniture / Home Store	Italian Restaurant	Mediterranean Restaurant	Art Gallery	Design Studio
Morningside Heights		American Restaurant				Sandwich Place			Deli / Bodega	
Sutton Place	Gym / Fitness Center	Italian Restaurant	Furniture / Home Store	Indian Restaurant	Dessert Shop	American Restaurant	Bakery	Juice Bar	Boutique	Restaurant
Hudson Yards	Coffee Shop	Italian Restaurant	Hotel	Theater	American Restaurant	Cafe	Gym / Fitness Center	Thai Restaurant	Restaurant	Gym

5.0 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.

6.0 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.