Coursera Capstone Project

COURSERA IBM APPLIED DATA SCIENCE CAPSTONE FINAL PROJECT

VASANTHA GEETHIKA YALLAPRAGADA

REPORT CONTENT:-

Introduction Section :

1.1 Discussion of the "background situation" leading to the problem at hand:

1.2 Problem to be resolved

1.3 Audience for this project.

Data Section:

2.1 Data of Current Situation (current residence place)

2.2 Data required to resolve the problem

2.3 Data sources and data manipulation

Methodology section :

3.1 Process steps and strategy to resolve the problem

3.2 Data Science Methods, machine learning, mapping tools and exploratory data analysis.

Results section:

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

Discussion section :

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

Conclusion section :

Decision taken and Report Conclusion.

Introduction Section:

Discussion of the business problem and the audience who would be interested in this project.

1.1 Scenario and Background :-

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:

1.Apartment with min 2 bedrooms with monthly rent not to exceed US$7000/month

2.Unit located within walking distance (<=1.0 mile, 1.6 km) from a subway metro station in Manhattan

3.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.

Data Section :-

2.1 Data of Current Situation:-

Foursquare to identify the venues around the area of residence n week 2 applied capstone . 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 neighbourhoods form Manhattan with their Geo data (latitude and longitude. List/Information about the subway metro stations in Manhattan with Geo data. 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 neighbourhoods 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. 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 neighbourhoods and mapping will be shown however. An algorithm was used to determine the geo data .The actual algorithm coding may be shown in 'markdown' mode because it takes time to run.

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.

The processing of these DATA and its mapping will allow to answer the key questions to make a decision:

1.what is the cost of available rental places that meet the demands?

2.what is the cost of rent around a mile radius from each subway metro station?

3.what is the area of Manhattan with best rental pricing that meets criteria established?

4.What is the distance from work place (Park Ave and 53 rd St) and the tentative future rental home?

5.What are the venues of the two best places to live? How the prices compare?

6.How venues distribute among Manhattan neighbourhoods and around metro stations?

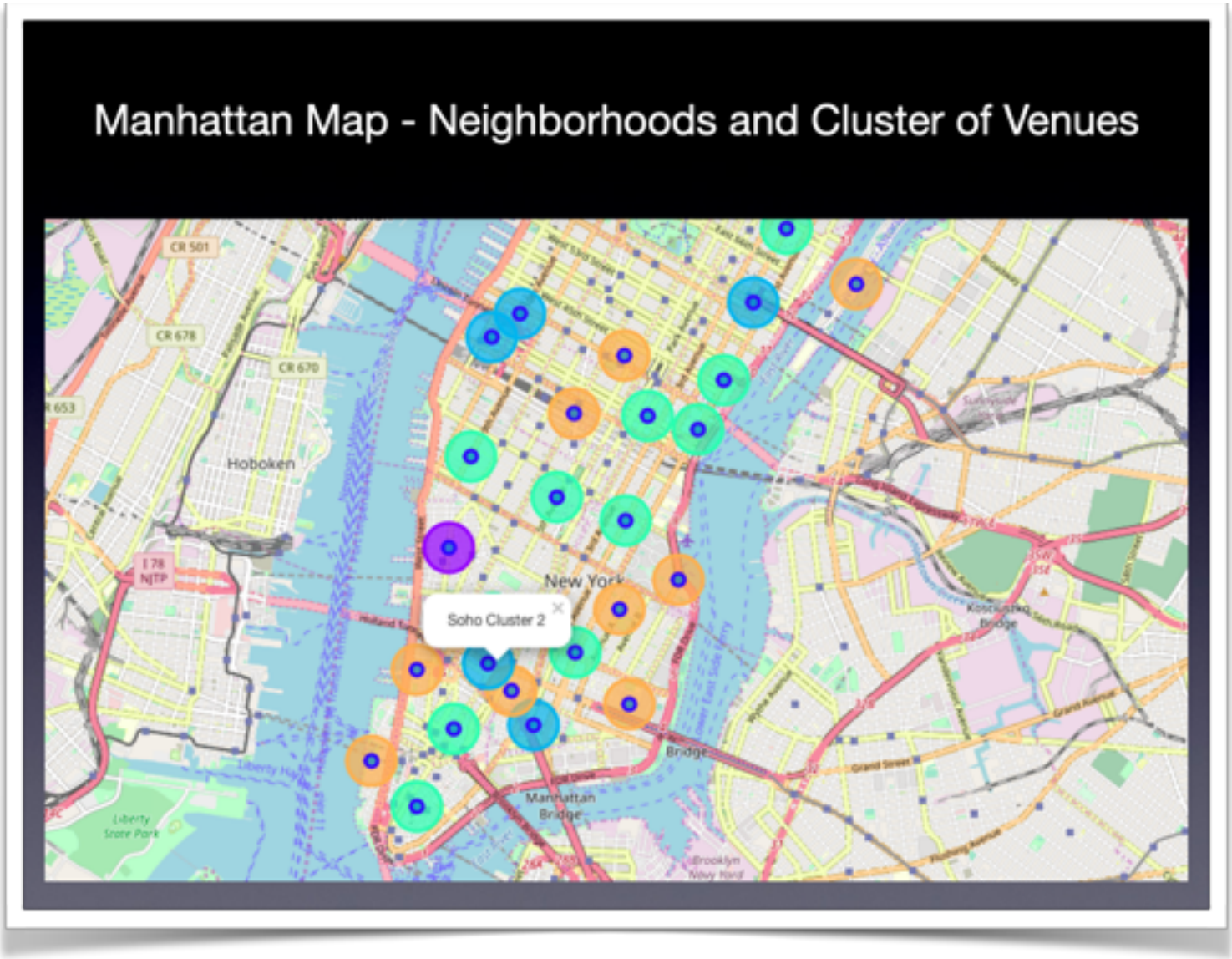
7.Are there tradeoffs between size and price and location?

8.Any other interesting statistical data findings of the real estate and overall data.

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 using Nominatim to get latitude and longitude . Geopy_distance and Nominatim were used to establish relative distances. Seaborn was used to visualise the data in the form of charts (bars,bubble,vertical) , folium used to display the maps

Execution and Results :-



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

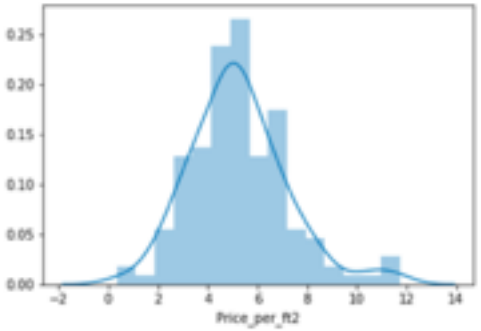
```
mh_rent.tail()
```

	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

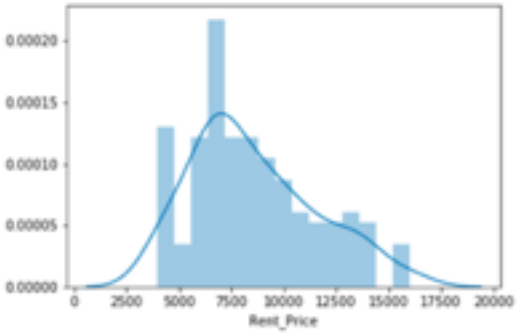
Rental Price Statistics MH Apartments

Budget US7000/month is around the mean

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

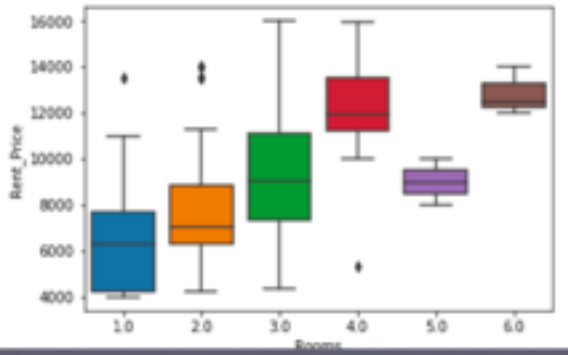



```
<matplotlib.axes._subplots.AxesSubplot at 0x1a25dd8400>
```

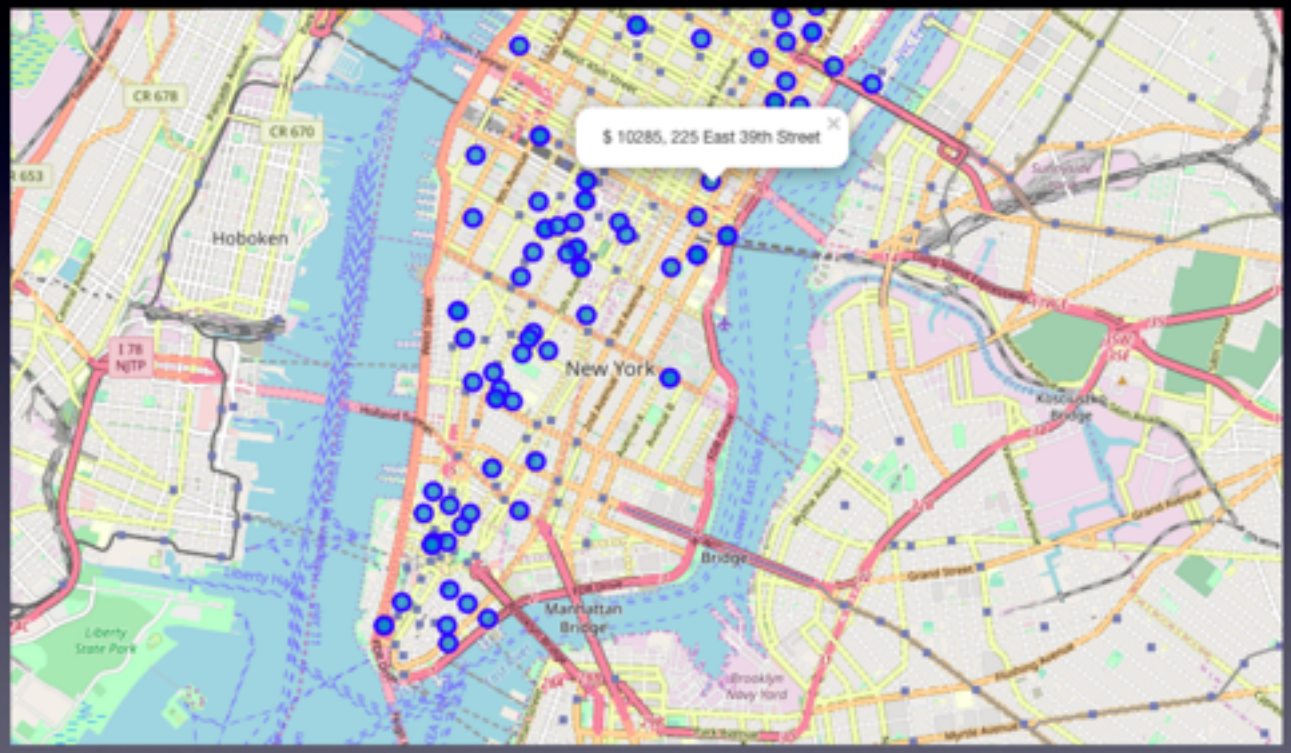



```
sns.boxplot(x='Rooms', y= 'Rent_Price', data=mh_rent)
```

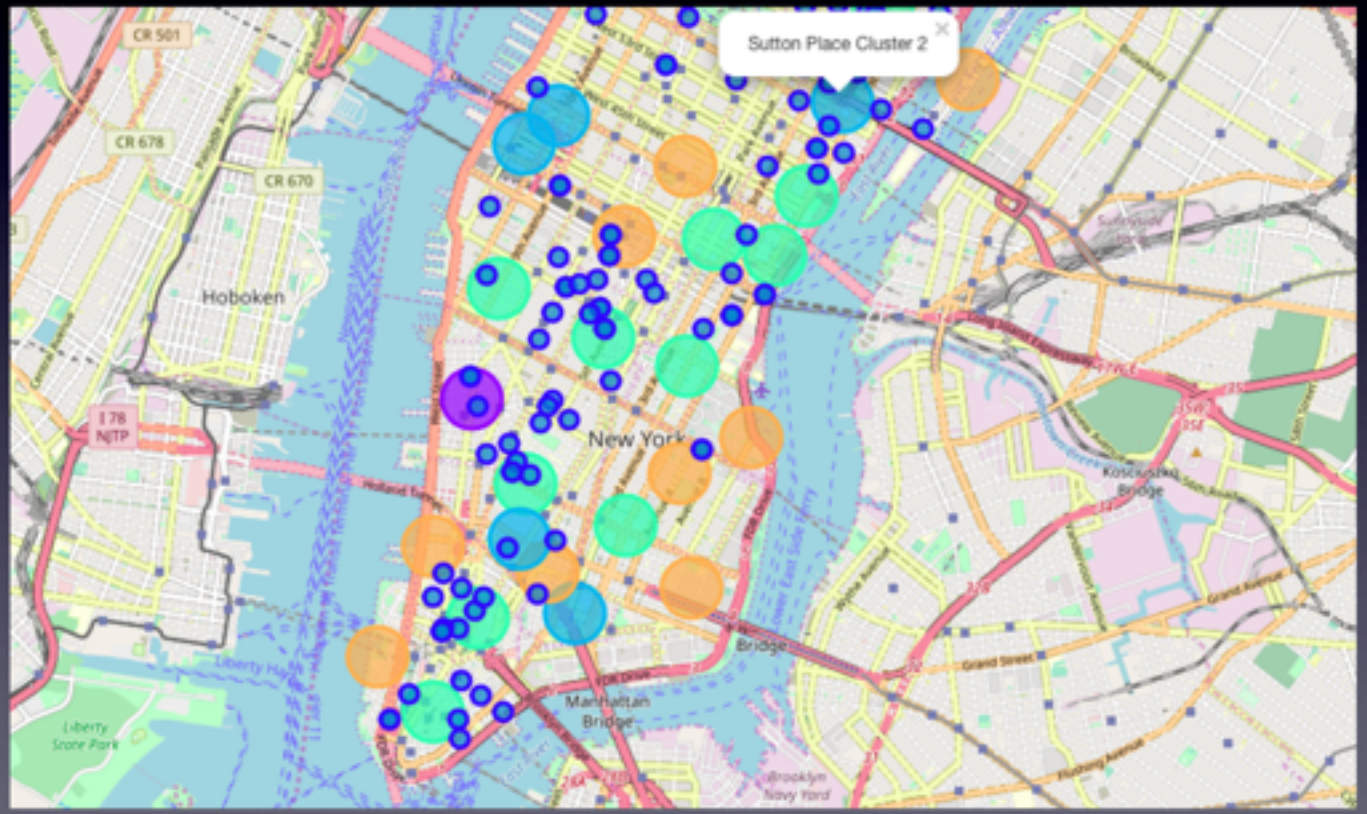
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a25f2a2b0>
```



Apartments for Rent in 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
6	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	Soho	Italian Restaurant	French Restaurant	Cocktail Bar	Gift Shop	Bookstore	Grocery Store	Mexican Restaurant	Hotel	Sushi Restaurant	Coffee Shop

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	-73.924609
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 mhsubl
mhsubl=mh.drop_duplicates(subset=['lat','long'], keep="last").reset_index(drop=True)
mhsubl.shape
```

```
(22, 4)
```

```
mhsubl.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	-73.996292

Apts for rent (blue) and subway stations (red)

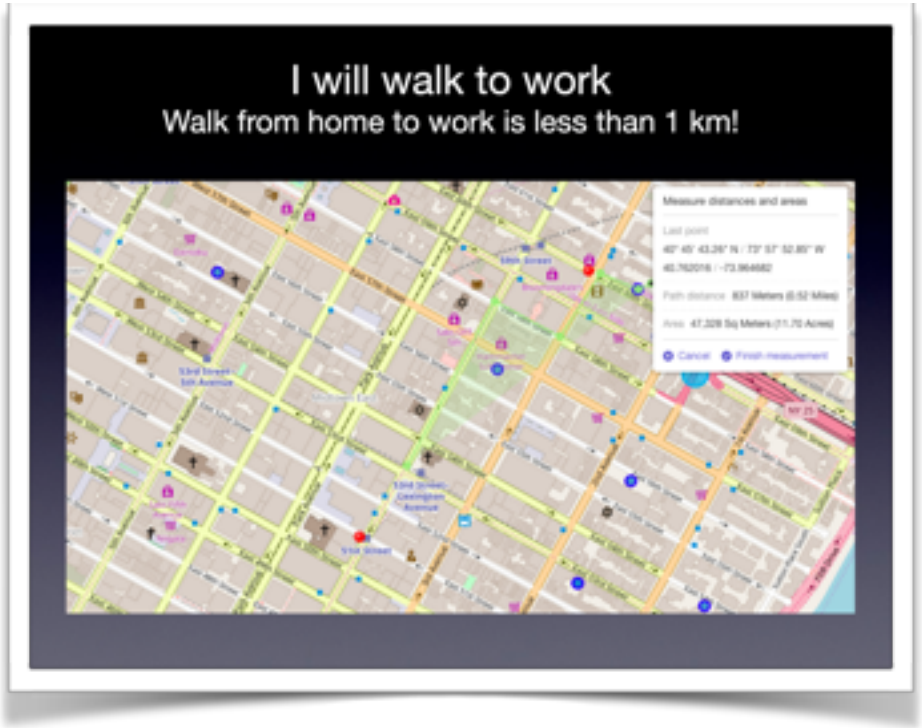
Selected Apartment!

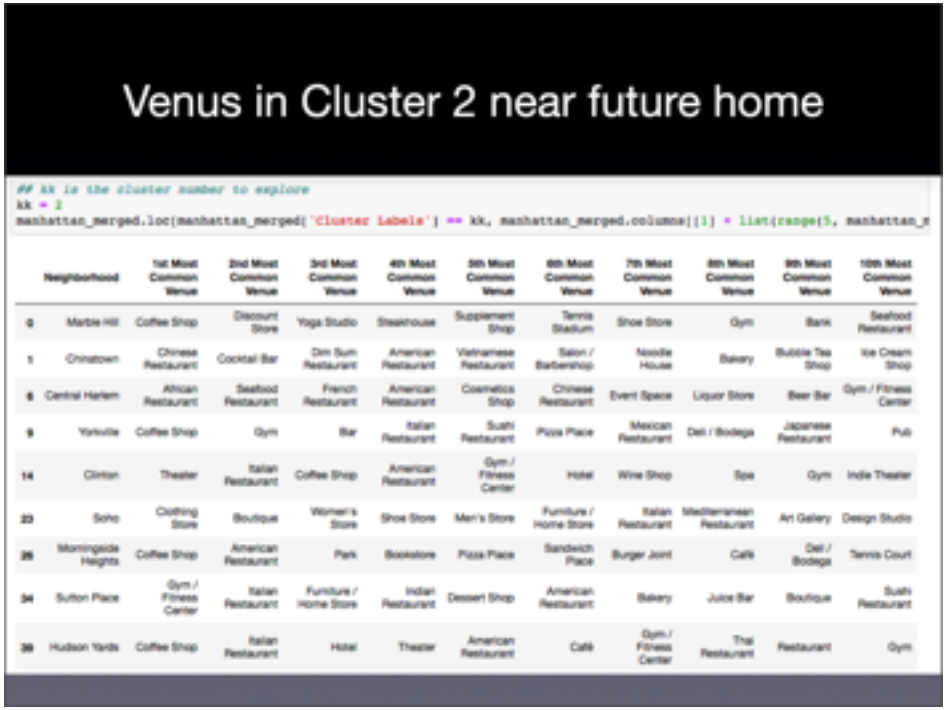
The ONE consolidated map shows all information for decision:
Apartments address, price, neighborhood, cluster of venues and subway station nearby.
Blue dots=apts , Red dots=Subway station, Bubbles=Cluster of Venues

Problem Resolution:-

After examining, I have chosen two locations that meet the requirements which will assess to make a choice.
Apartment 1: 305 East 63rd Street in the Sutton Place Neighbourhood and near 'subway 59th Street' station, Cluster # 2 Monthly rent : 7500 Dollars

Apartment 2: 19 Dutch Street in the Financial District Neighbourhood and near 'Fulton Street Subway' station, Cluster # 3 Monthly rent : 6935 Dollars





DISCUSSION

In general, I am positively impressed with the overall organisation, 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.

CONCLUSION

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.

With Foursquare, I was able to do a lot of analysis thanks to the available data. Thanks to the users of this service, it is really possible to generate interesting analyses and to expose them to offer other services.

Here is the end of data analysis .please do-not hesitate to give feedback