Coursera Capstone Project

Coursera IBM Data Science Certification

ARUNKUMAR P

Report Content

1. *Introduction Section :*

- The “business problem” to be solved by this project.

2. *Data Section:*

- Describe Data requirements and Sources needed to solve the problem

3. *Methodology section:*

- Main component of the report - data processing,discuss any data analysis and/or inferential statistical testing performed.

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.Introduction:

Scenario

I am a data scientist residing in Bangalore,India. I currently live within walking distance to Bangalore office and I enjoy many ammenities and venues in the area, such as various international cousine restaurants, cafes, food shops and entertainment. I have been offered a great opportunity to work for a company in Hyderabad also. I am very excited and I want to use this opportunity to use the data science skillset acquired here.The most difficult thing is to find a apartment in Bangalore.

The requirements are:

- :must be 2 bhk
- :should be near to metro station
- :rent should be minimum

Business Problem:

The challenge is to find a suitable apartment for rent in Bangalore that complies with the demands on location, price and venues.Interested Audience I believe this is a valid scenario with questions for anyone moving to city in India,Japan.

2. Data Section:

List of Boroughs and neighborhoods of Bangalore with their geodata. List of Subway metro stations with their address location. List of apartments for rent in area with their addresses and price.

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 whereone can visualize all details needed to make a selection of apartment.

2.1 Data of Current Situation

I Currently live in the neighborhood of 'Sai Street' in Bangalore. I used Foursquare to identify the venues around the area of residence which are shown in the Bangalore map shown in methodology and execution in section 3.0 . It can be used as a reference for comparison with the desired future location in Manhattan NY

2.2 *Data Required to resolve the problem*

For the purpose to make a good choice of a similar apartment in Manhattan NY, the following data is required: List/Information on neighborhoods 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 ammenities in the Manhattan neighborhoods.

3. Methodology section:

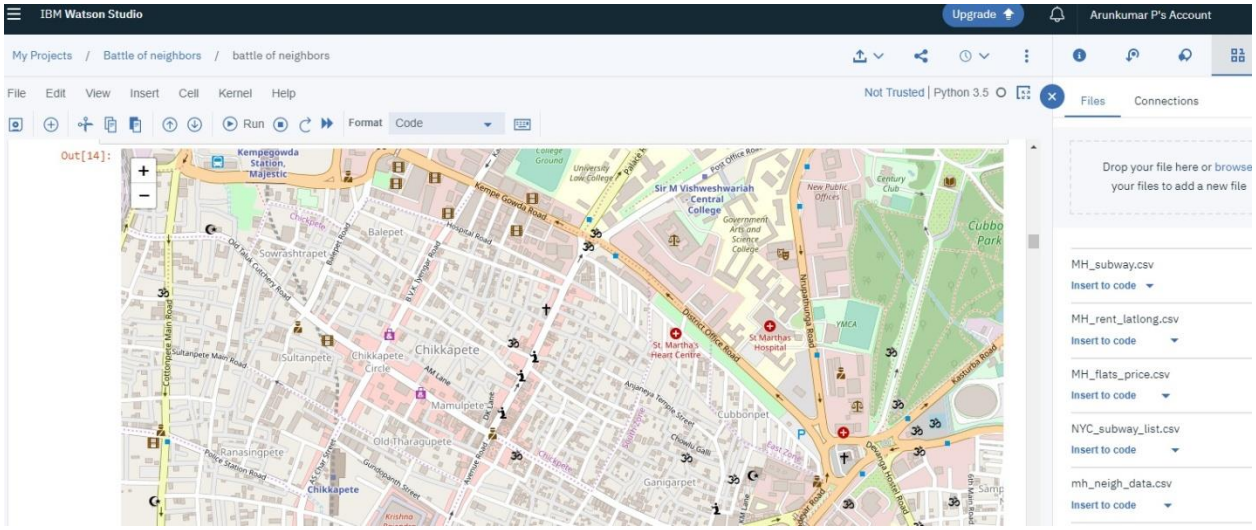
This section represents the main component of the report where the data is gathered, prepared for analysis. The tools described are used here and the Notebook cells indicates the execution of steps.

Analysis:-

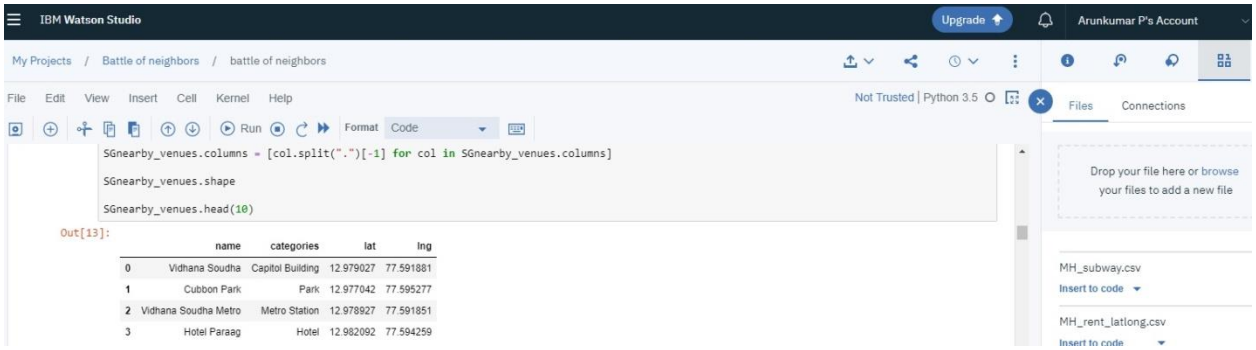
The analysis is done on mapping the above described data in section 2.0, in order to facilitate the choice of at least two chosen places for rent. The choice is made based on the demands imposed : location near a subway, rental price and similar venues to Bangalore.

4.0 Results

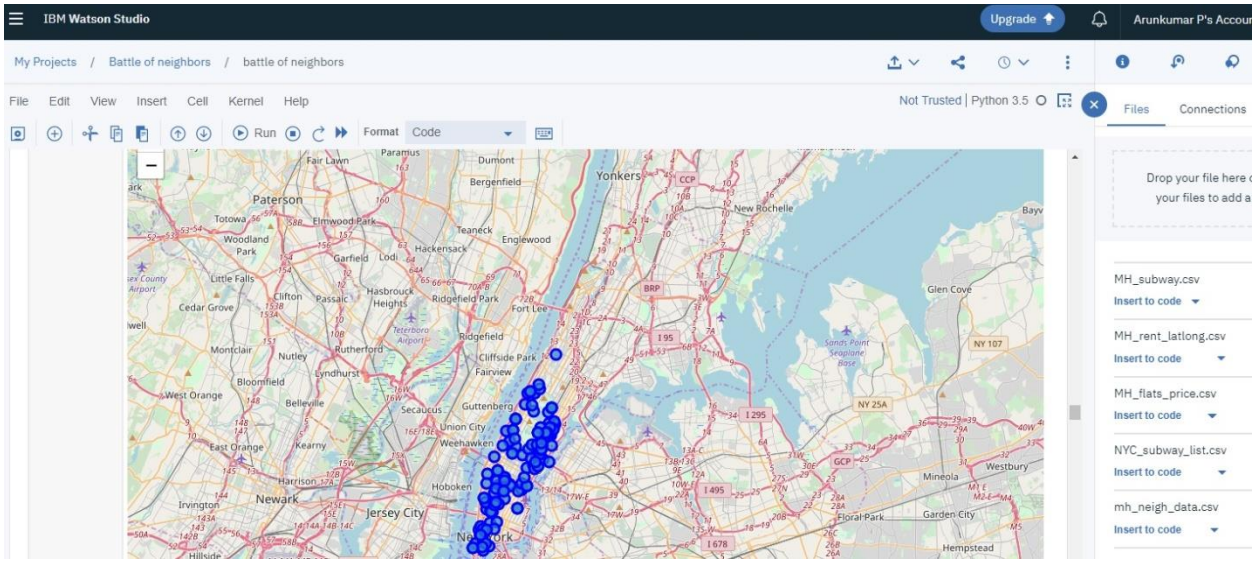
Present place



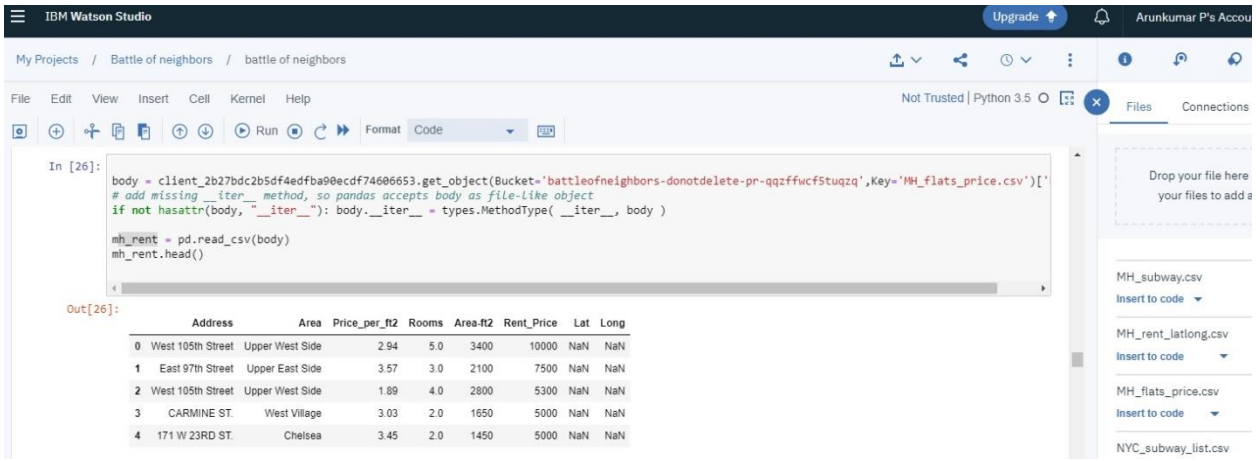
Neighborhood



Manhattan-Venues

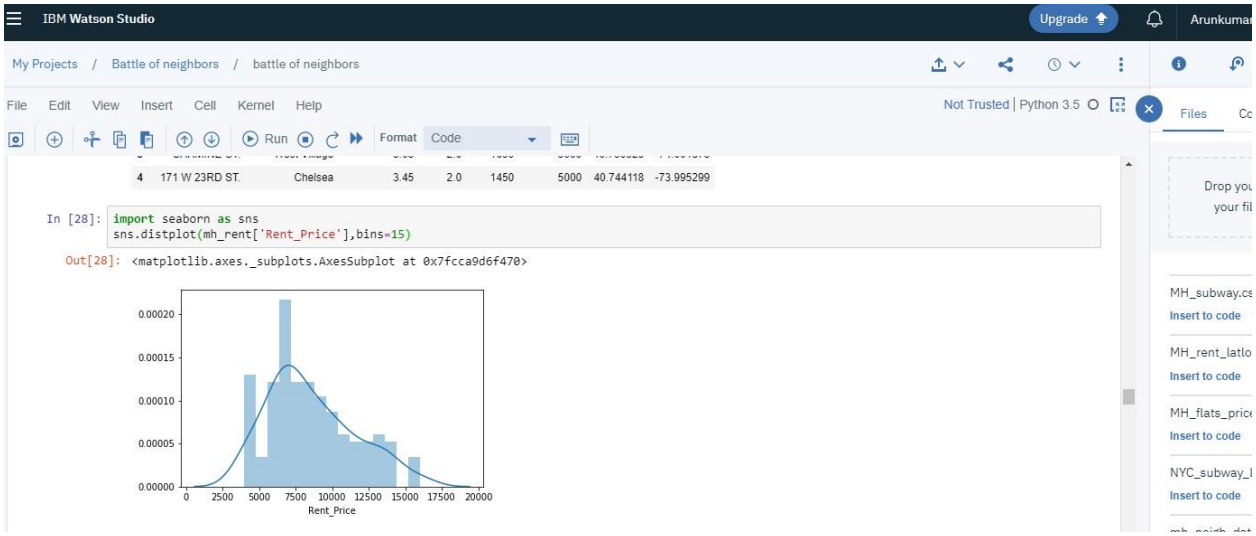


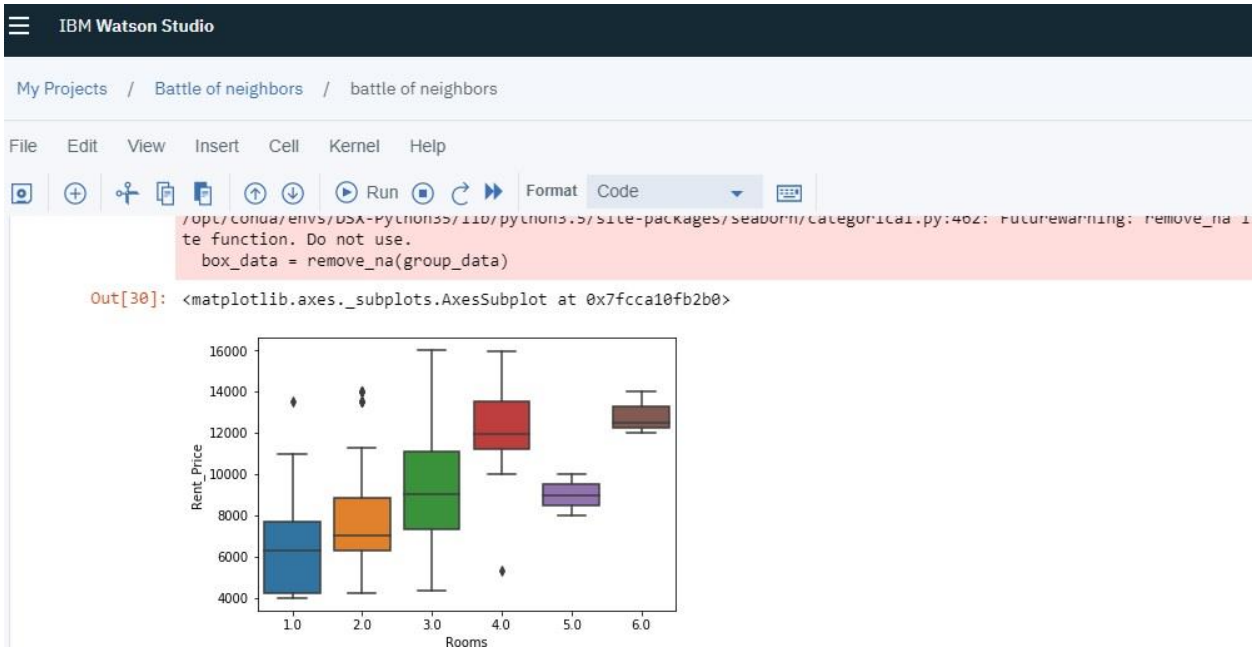
Space for rent



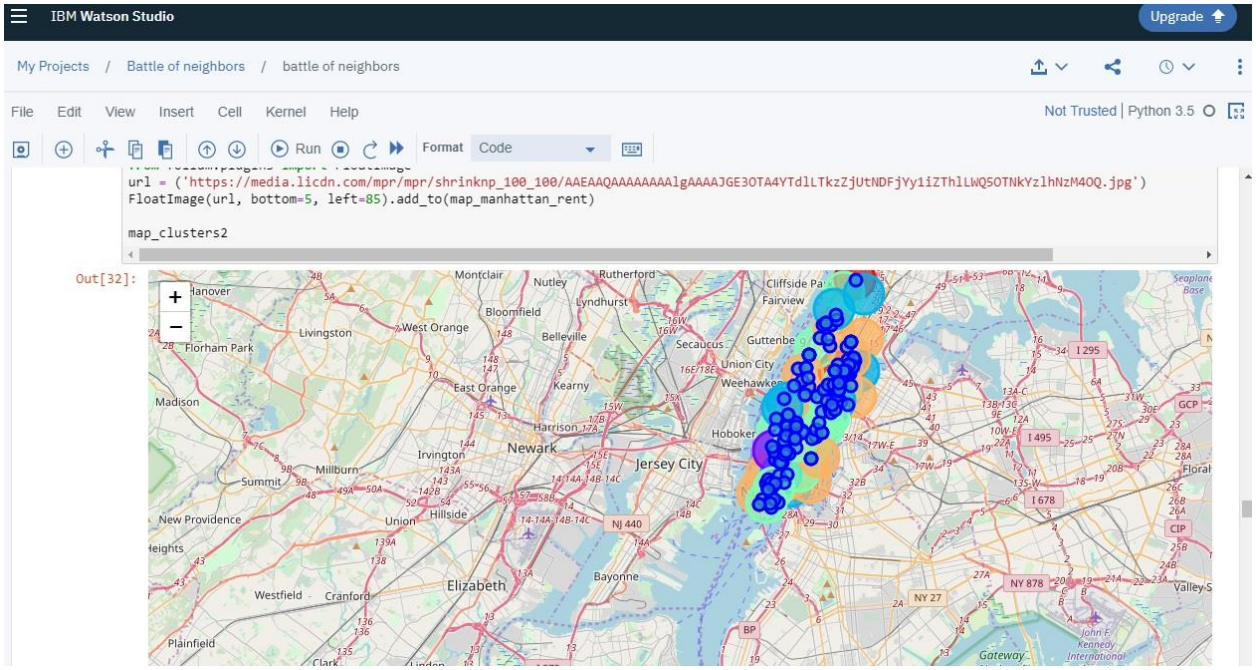
	Address	Area	Price_per_ft2	Rooms	Area-ft2	Rent_Price	Lat	Long
0	West 105th Street	Upper West Side	2.94	5.0	3400	10000	NaN	NaN
1	East 97th Street	Upper East Side	3.57	3.0	2100	7500	NaN	NaN
2	West 105th Street	Upper West Side	1.89	4.0	2800	5300	NaN	NaN
3	CARMINE ST.	West Village	3.03	2.0	1650	5000	NaN	NaN
4	171 W 23RD ST.	Chelsea	3.45	2.0	1450	5000	NaN	NaN

Statistics

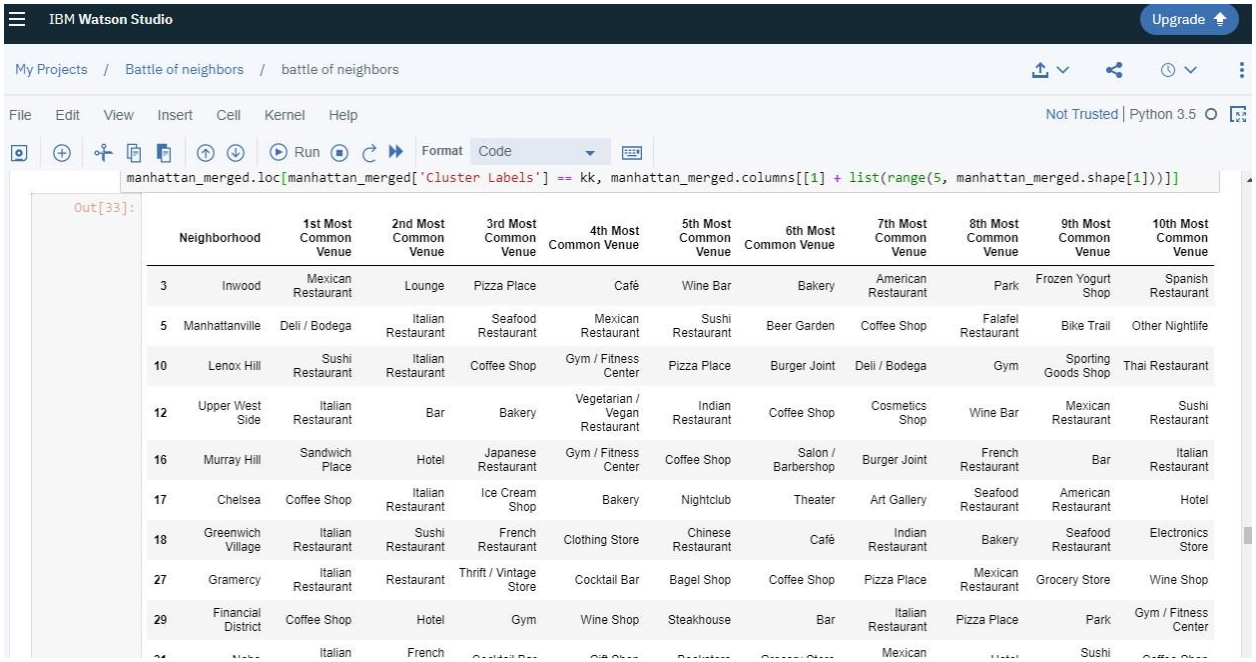

IBM Watson Studio
Upgrade
My Projects / Battle of neighbors / battle of neighbors
File Edit View Insert Cell Kernel Help
Not Trusted | Python 3.5
Run Format Code
4 171 W 23RD ST. Chelsea 3.45 2.0 1450 5000 40.744118 -73.995299
In [28]:
import seaborn as sns
sns.distplot(mh_rent['Rent_Price'],bins=15)
Out[28]: <matplotlib.axes._subplots.AxesSubplot at 0x7fcca9d6f470>
0.00020
0.00015
0.00010
0.00005
0.00000
0 2500 5000 7500 10000 12500 15000 17500 20000
Rent_Price
Files
MH_subway.cs
Insert to code
MH_rent_latlo
Insert to code
MH_flats_price
Insert to code
NYC_subway_l
Insert to code


IBM Watson Studio
My Projects / Battle of neighbors / battle of neighbors
File Edit View Insert Cell Kernel Help
Run Format Code
te function. Do not use.
box_data = remove_na(group_data)
Out[30]: <matplotlib.axes._subplots.AxesSubplot at 0x7fcca10fb2b0>
16000
14000
12000
10000
8000
6000
4000
Rent_Price
1.0 2.0 3.0 4.0 5.0 6.0
Rooms

Apartments for rent

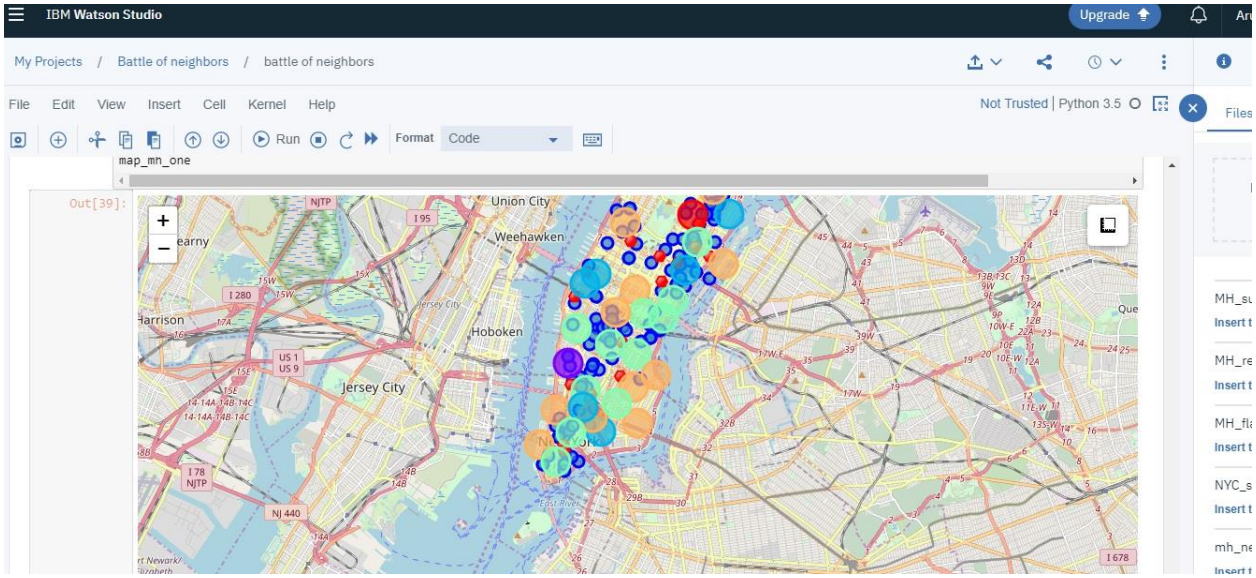


Venues



	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	French					Mexican			Sushi

Selection



The above dots are the available room.

5.0 DISCUSSION

Definitely this Capstone project has given me a golden opportunity to practice and apply the Data Science tools and methodologies learned.

6.0 CONCLUSIONS

I am greatly rewarded with these efforts, time and money spent. Surely this course is well worth 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.