COUSERA CAPSTONE PROJECT

IBM DATA SCIENCE CERTIFICATION

K PAVAN YASWANTH

FINAL REPORT

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.

INTRODUCTION

1.1 SCENARIO AND BACKGROUND

I AM CURRENTLY LIVING IN SINGAPORE, WITHIN WALKING DISTANCE TO DOWNTOWN "TELOK AYER MRT METRO STATION" . 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.

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

METHODOLOGY

THE STRATEGY TO FIND THE ANSWER:

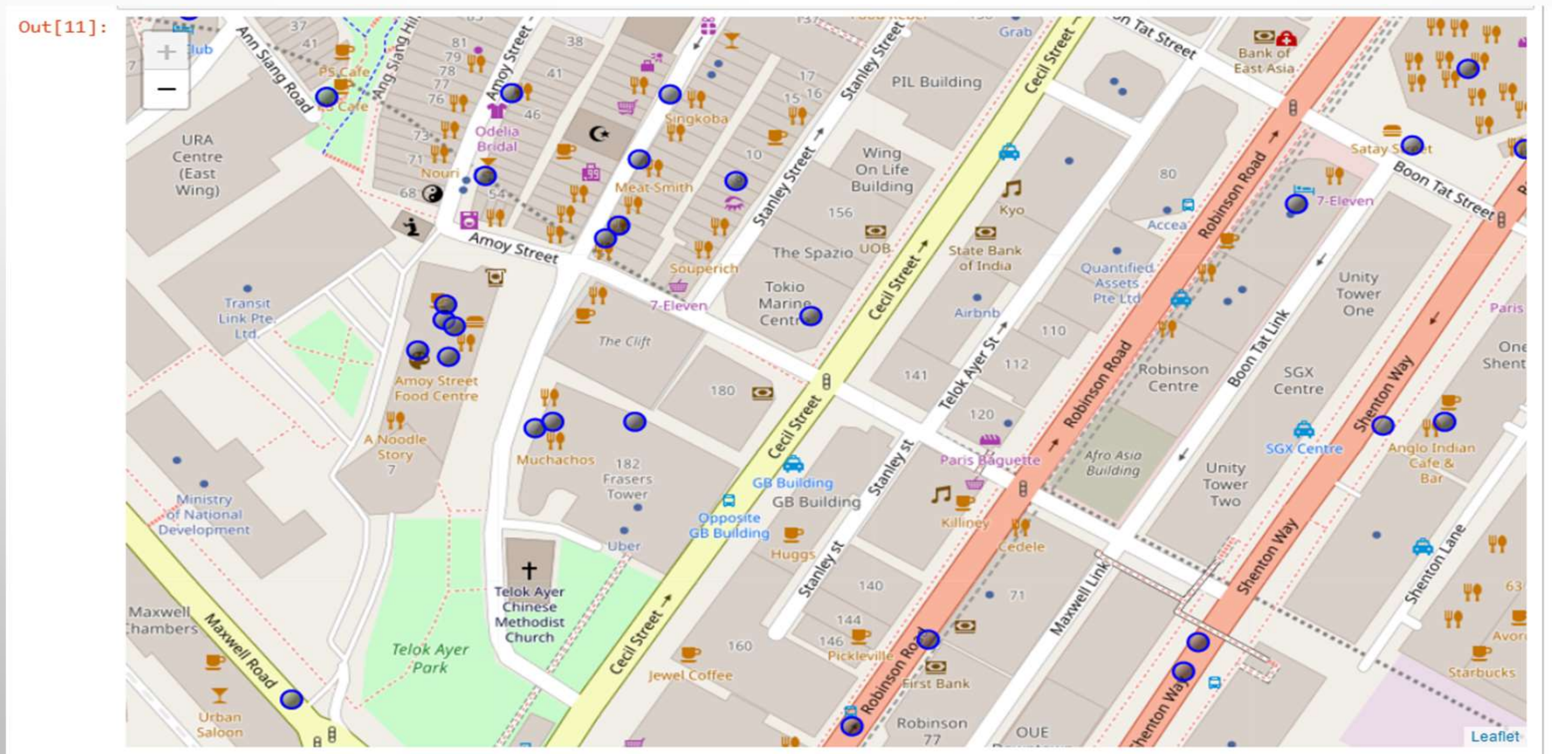
THE STRATEGY IS BASED ON MAPPING THE DESCRIBED DATA IN SECTION 2.0, IN ORDER TO FACILITATE THE CHOICE OF AT LEAST TWO CANDIDATE PLACES FOR RENT. THE INFORMATION WILL BE CONSOLIDATED IN ONE MAP WHERE ONE CAN SEE THE DETAILS OF THE APARTMENT, THE CLUSTER OF VENUES IN THE NEIGHBORHOOD AND THE RELATIVE LOCATION 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

EXECUTION AND RESULTS

CURRENT RESIDENCE NEIGHBORHOOD IN SINGAPORE



VENUES AROUND NEIGHBORHOOD

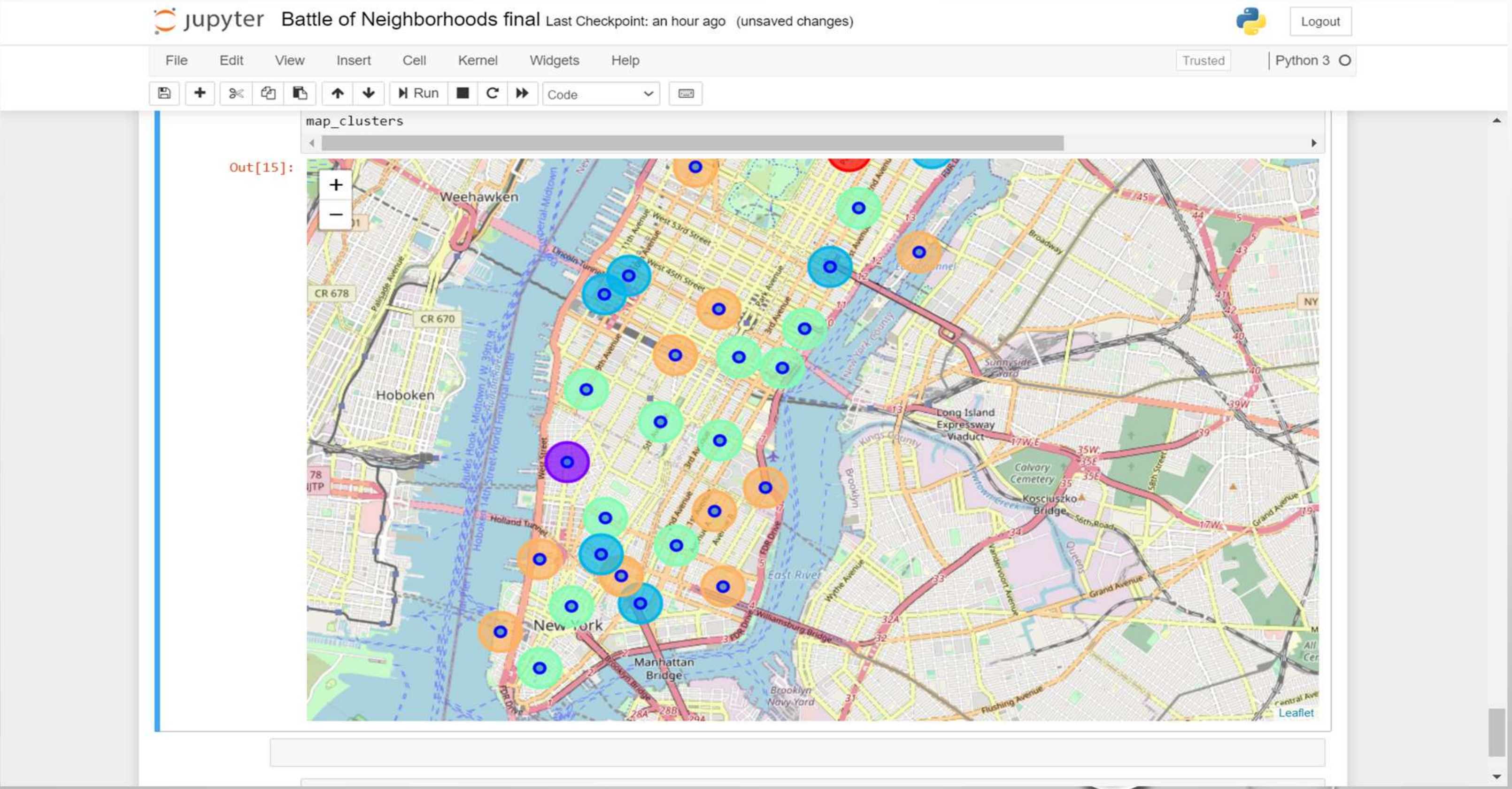
In [10]:

```
# Venues near current Singapore residence place
SGnearby_venues.head(10)
```

Out[10]:

	name	categories	lat	lng
0	The Westin Singapore	Hotel	1.278275	103.850772
1	Pure Fitness	Gym	1.278631	103.851487
2	Lau Pa Sat Satay Street	Street Food Gathering	1.280261	103.850235
3	Anglo Indian Cafe & Bar	Indian Restaurant	1.279084	103.850127
4	Westin Infinity Pool	Pool	1.278057	103.851077
5	Napoleon Food & Wine Bar	Wine Bar	1.279925	103.847333
6	Sofitel So Singapore	Hotel	1.280017	103.849813
7	Lobby Lounge Westin	Bar	1.277811	103.850966
8	Mellower Coffee	Café	1.277814	103.848188
9	Cook & Brew	Gastropub	1.277842	103.851103

MANHATTAN MAP - NEIGHBORHOODS AND CLUSTER OF VENUES



GEODATA MANHATTAN APTS FOR RENT

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

Out[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	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

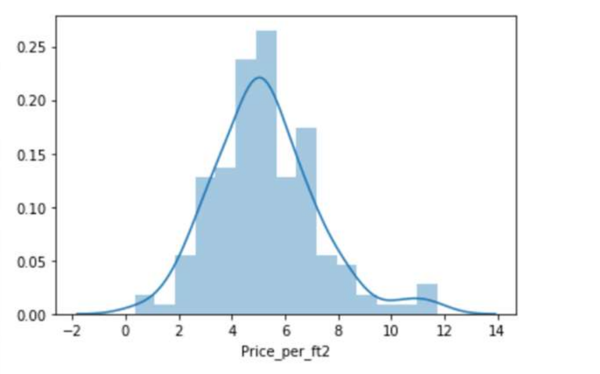
[21]:
```
mh_rent.tail()
```

Out[21]:

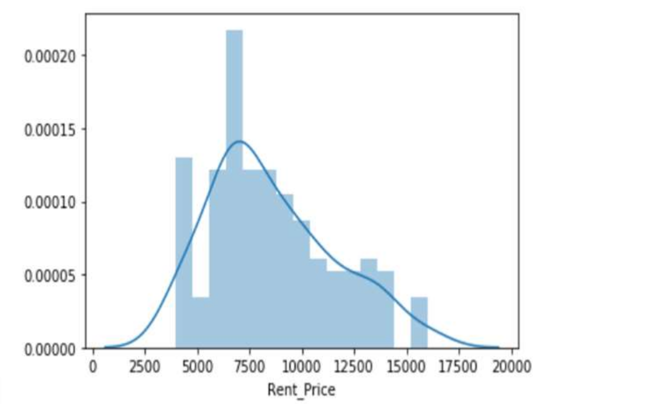
	Address	Area	Price_per_ft2	Rooms	Area-ft2	Rent_Price	Lat	Long
139	200 East 72nd Street	Rental in Lenox Hill	5.15	3.0	1700	8750	40.769465	-73.960339
140	50 Murray Street	No fee rental in Tribeca	7.11	2.0	1223	8700	40.714051	-74.009608
141	300 East 56th Street	No fee rental in Midtown East	3.87	3.0	2100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

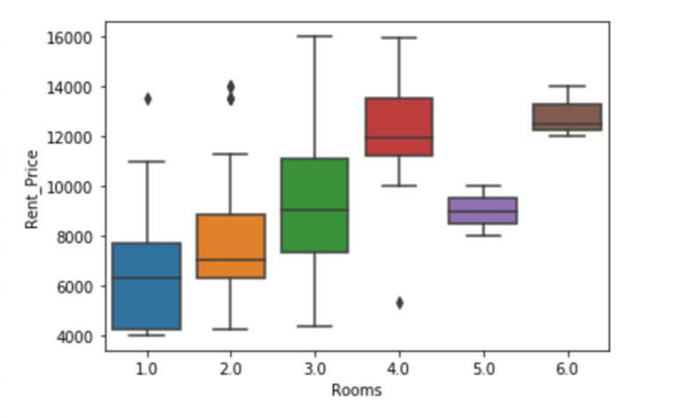
<matplotlib.axes._subplots.AxesSubplot at 0x140dfac6ac8>



<matplotlib.axes._subplots.AxesSubplot at 0x140dce90c48>

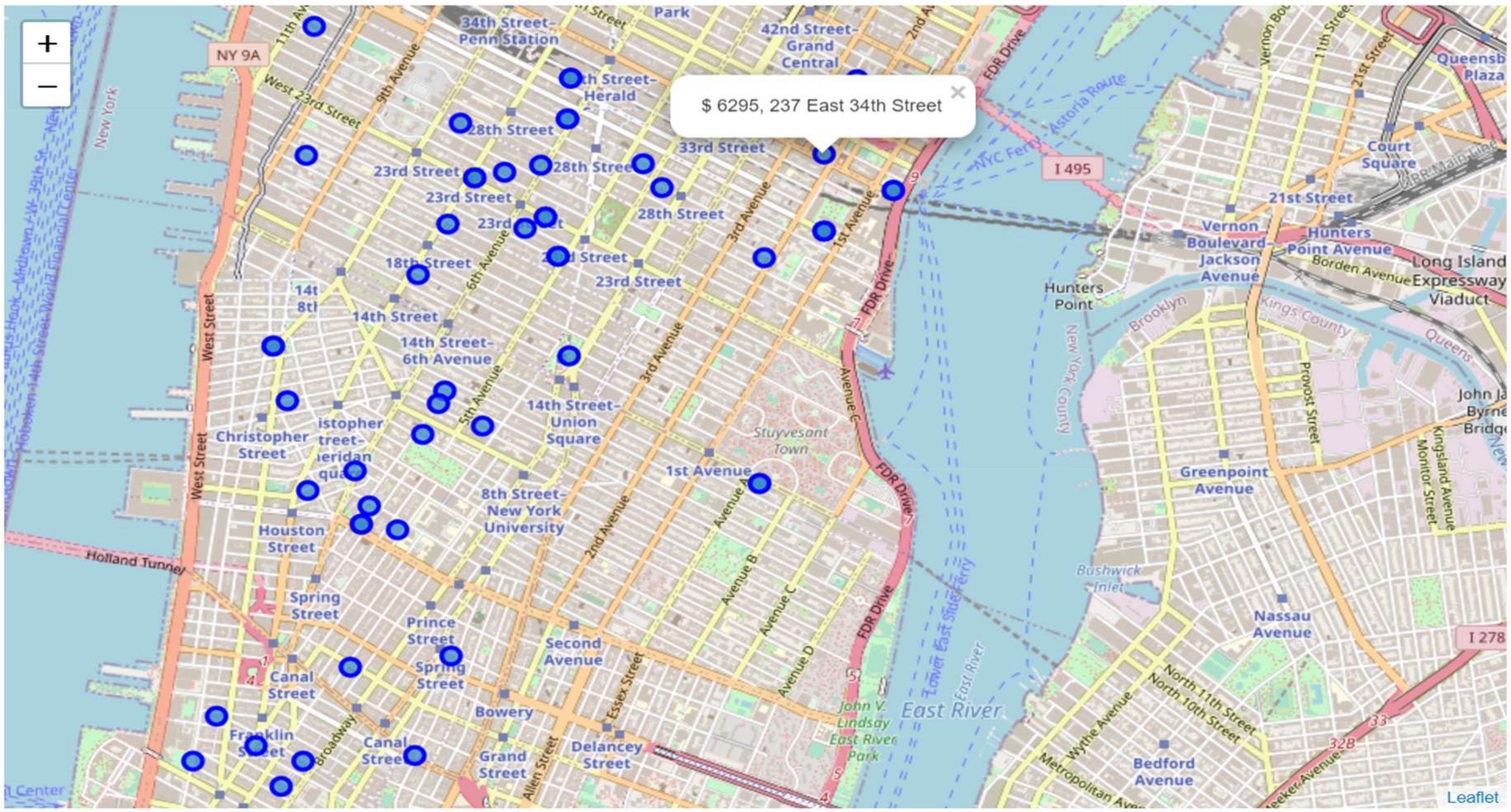


<matplotlib.axes._subplots.AxesSubplot at 0x140dfb63448>

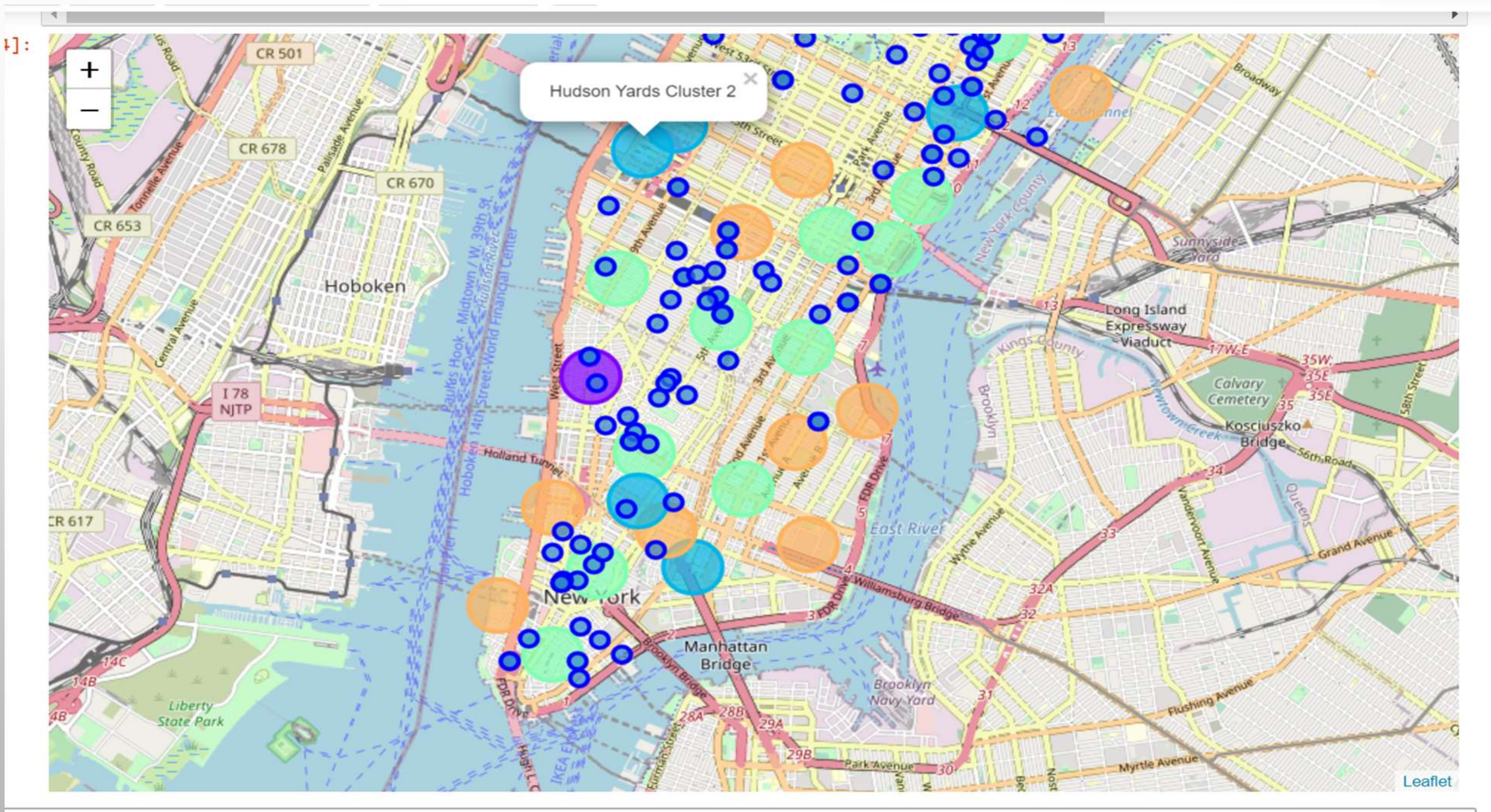


APARTMENTS FOR RENT IN MH

Out[23]:



MH APT FOR RENT WITH VENUE CLUSTERS



MH APT FOR RENT WITH VENUE CLUSTERS

In [25]:

```
## 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.
```

Out[25]:

	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 SUBWAY STATIONS GEODATA

In [28]:
```
mh=pd.read_csv('MH_subway.csv')
print(mh.shape)
mh.head()
```

```
(76, 4)
```

Out[28]:

	sub_station	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

In [29]:
```
# removing duplicate rows and creating new set mhsub1
mhsub1=mh.drop_duplicates(subset=['lat','long'], keep="last").reset_index(drop=True)
mhsub1.shape
```

Out[29]: (22, 4)

In [30]:
```
mhsub1.tail()
```

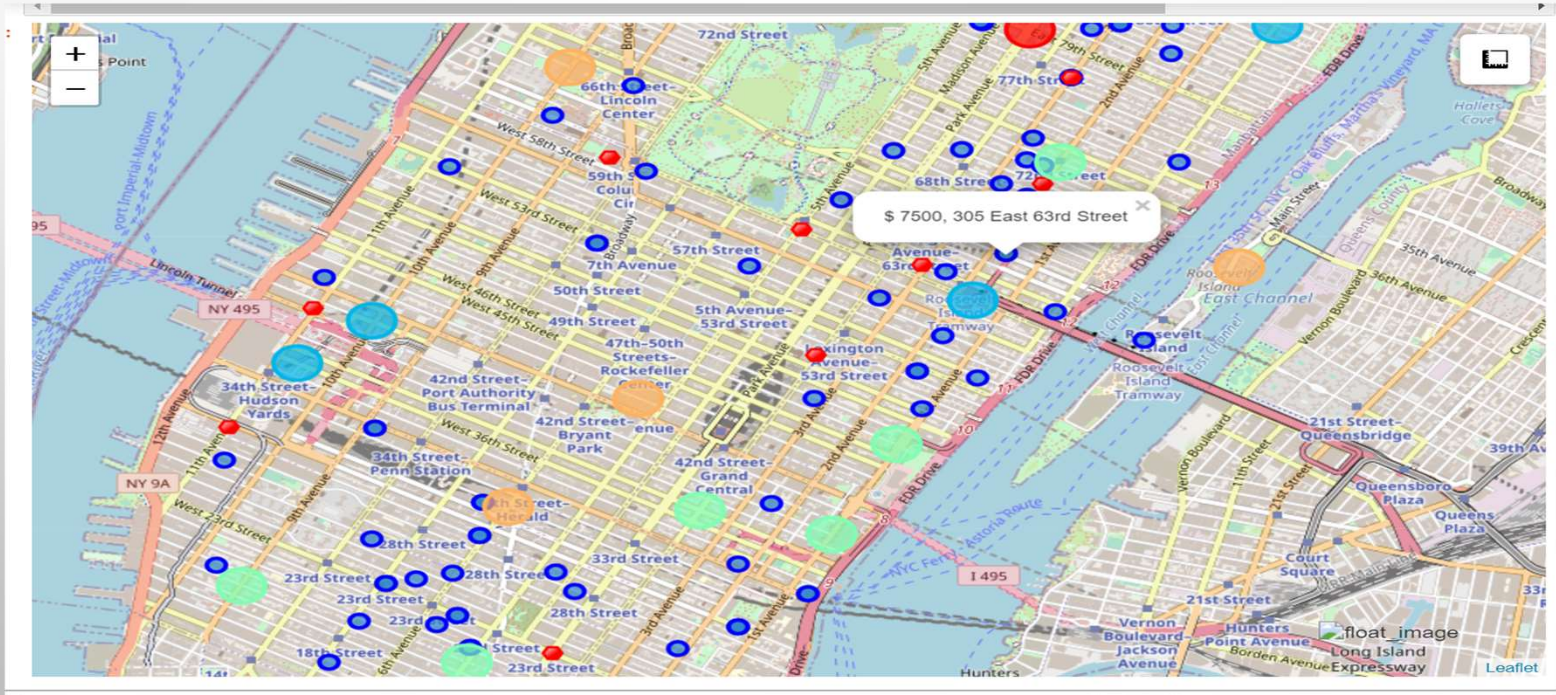
Out[30]:

	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

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



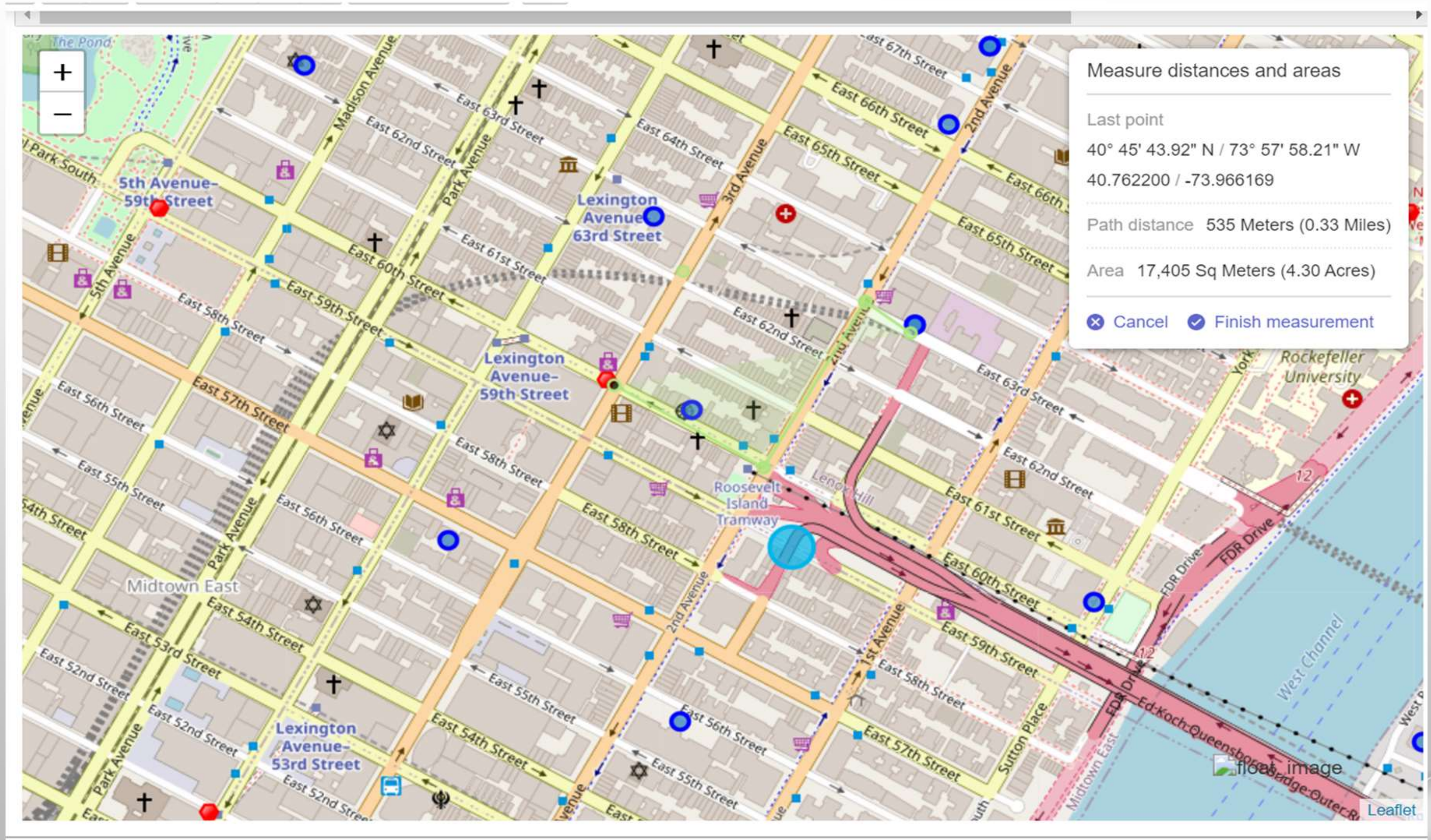
APARTMENT SELECTION

USING THE "ONE MAP" ON TOP OF, I USED TO BE ABLE TO EXPLORE ALL POTENTIALITIES SINCE THE POPUPS PROVIDE THE KNOWLEDGE REQUIRED FOR AN HONEST CALL.

APARTMENT ONE RENT VALUE IS US7500 SLIGHTLY ON TOP OF THE US7000 BUDGET. APT ONE IS FOUND 400 METERS FROM THE DEPOT AT 59TH STREET AND GEOGRAPHIC POINT (PARK AVE AND 53RD) IS ANOTHER 600 METERS MEANS. I WILL WALK TO THE GEOGRAPHIC POINT AND USE SUBWAY FOR DIFFERENT PLACES AROUND. VENUES FOR THIS APT SPACE OF CLUSTER A PAIR OF AND IT'S SET DURING A FINE DISTRICT WITHIN THE EASTSIDE OF MANHATTAN.

APARTMENT A PAIR OF RENT VALUE IS US6935, JUST BELOW THE US7000 BUDGET. APT A PAIR OF IS FOUND SIXTY METERS FROM THE DEPOT AT DISCOVERER STREET, HOWEVER I WILL BE ABLE TO OUGHT TO RIDE THE SUBWAY DAILY TO WORK, PROBABLY 40-60 MIN RIDE. VENUES FOR THIS APT SPACE OF CLUSTER THREE.1 BASED ON CURRENT SINGAPORE VENUES, I FEEL THAT CLUSTER A PAIR OF FORM OF VENUES MAY BE A NEARER RESEMBLANCE TO MY CURRENT PLACE. MEANING THAT FLAT ONE MAY BE A BETTER OPTION SINCE THE ADDITIONAL MONTHLY RENT IS WELL WORTH THE CONVENIENCES IT PROVIDES.

WALK FROM HOME TO WORK IS LESS THAN 1 KM



VENUS IN CLUSTER 2 NEAR FUTURE HOME

In [35]:

```
## 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_merged.
```

Out[35]:

	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
0	Marble Hill	Coffee Shop	Discount Store	Yoga Studio	Steakhouse	Supplement Shop	Tennis Stadium	Shoe Store	Gym	Bank	Seafood Restaurant
1	Chinatown	Chinese Restaurant	Cocktail Bar	Dim Sum Restaurant	American Restaurant	Vietnamese Restaurant	Salon / Barbershop	Noodle House	Bakery	Bubble Tea Shop	Ice Cream Shop
6	Central Harlem	African Restaurant	Seafood Restaurant	French Restaurant	American Restaurant	Cosmetics Shop	Chinese Restaurant	Event Space	Liquor Store	Beer Bar	Gym / Fitness Center
9	Yorkville	Coffee Shop	Gym	Bar	Italian Restaurant	Sushi Restaurant	Pizza Place	Mexican Restaurant	Deli / Bodega	Japanese Restaurant	Pub
14	Clinton	Theater	Italian Restaurant	Coffee Shop	American Restaurant	Gym / Fitness Center	Hotel	Wine Shop	Spa	Gym	Indie Theater
23	Soho	Clothing Store	Boutique	Women's Store	Shoe Store	Men's Store	Furniture / Home Store	Italian Restaurant	Mediterranean Restaurant	Art Gallery	Design Studio
26	Morningside Heights	Coffee Shop	American Restaurant	Park	Bookstore	Pizza Place	Sandwich Place	Burger Joint	Café	Deli / Bodega	Tennis Court

DISCUSSION

IN GENERAL, I AM POSITIVELY IMPRESSED WITH THE OVERALL ORGANIZATION, CONTENT AND LAB WORKS PRESENTED DURING THE COURSERA IBM CERTIFICATION COURSE

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

- I FEEL REWARDED WITH THE EFFORTS, TIME AND CASH SPENT. I BELIEVE THIS COURSE WITH ALL THE TOPICS LINED IS WELL WORTHY OF APPRECIATION.
- THIS PROJECT HAS SHOWN PINE TREE STATE AN EMPLOYMENT TO RESOLVE A REAL SCENARIO THAT HAS IMPACTING PERSONAL AND MONEY IMPACT MISTREATMENT KNOWLEDGE SCIENCE TOOLS.
- THE MAPPING WITH GEOLOGICAL FORMATION COULD BE A TERRIBLY POWERFUL TECHNIQUE TO CONSOLIDATE DATA AND CREATE THE ANALYSIS AND CALL THOROUGHLY AND CONFIDENTLY. I'D SUGGEST FOR USE IN SIMILAR THINGS.
- ONE SHOULD KEEP UP WITH RECENT TOOLS FOR DS THAT CONTINUE TO APPEAR FOR APPLICATION IN MANY BUSINESS FIELDS.