COURSERA CAPSTONE PROJECT

Coursera IBM Data Science Certification

ABQADER

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 - 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.

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 (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 METHODOLOGY

The Strategy to find the answer:

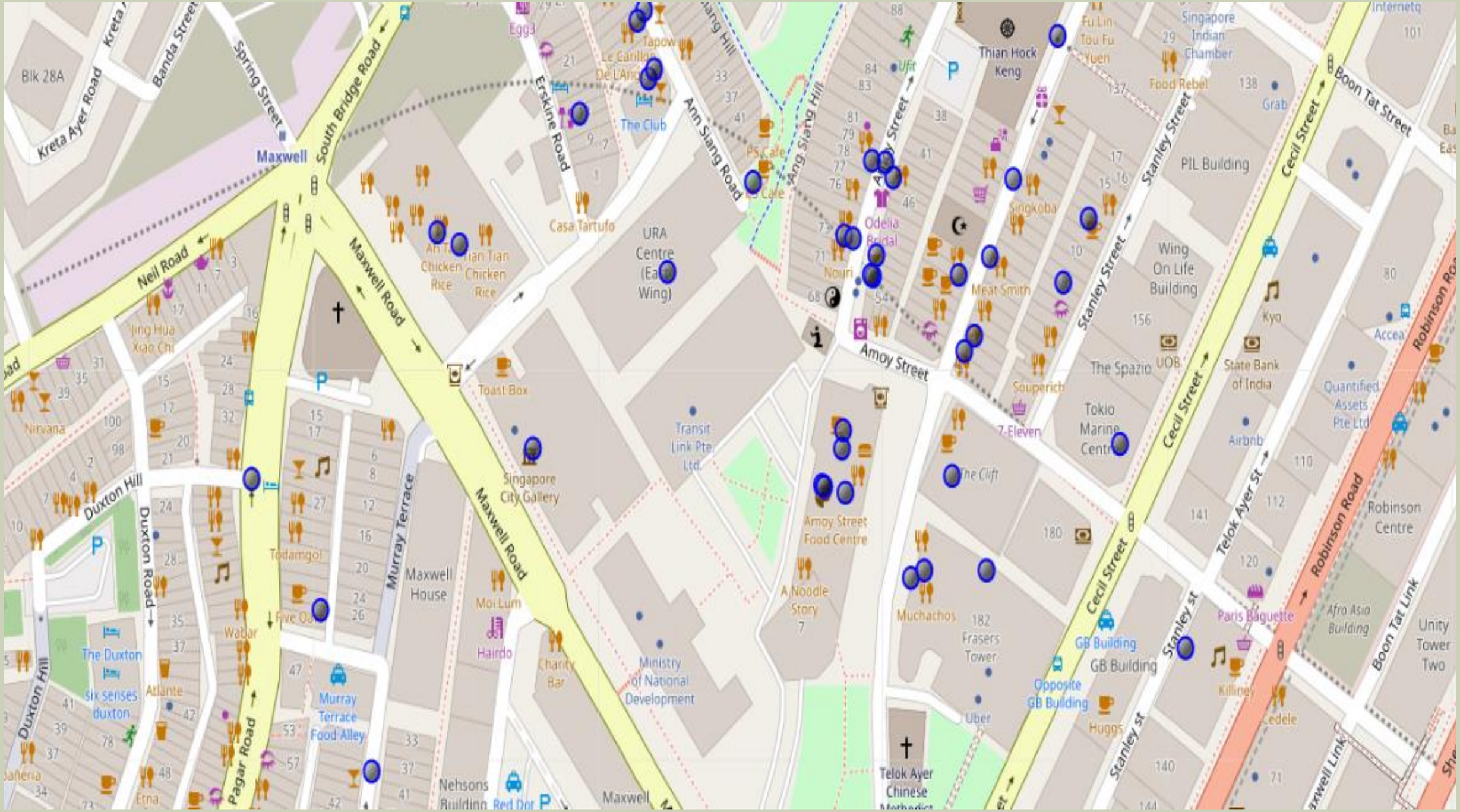
The strategy is based on mapping the described data in section 2.0, in order to facilitate the choice of at least two candidate places for rent. The information will be consolidated in ONE MAP where one can see the details of the apartment, the cluster of venues in the neighborhood and the relative location 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

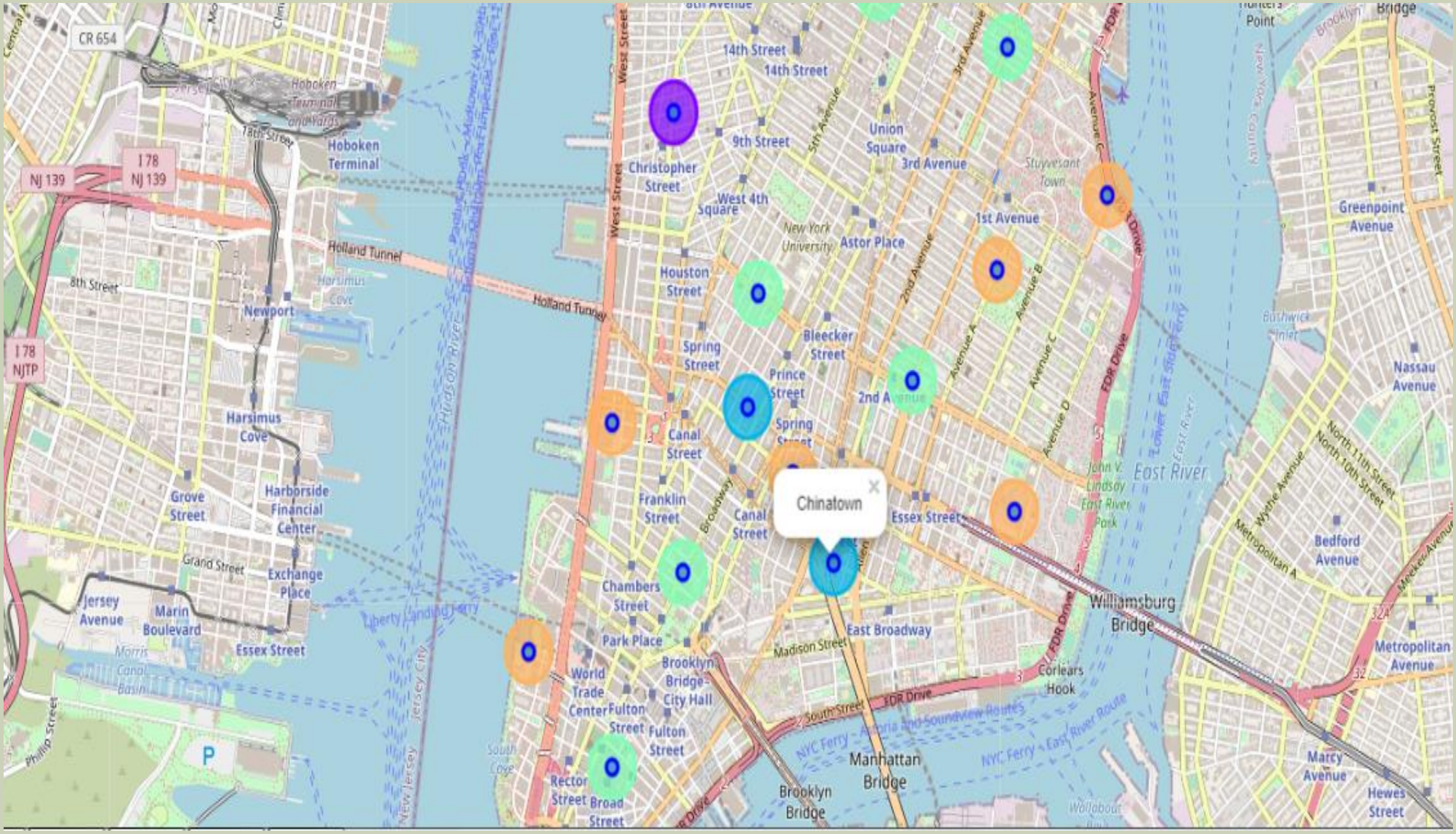
CURRENT RESIDENCE NEIGHBORHOOD IN SINGAPORE



VENUES AROUND NEIGHBORHOOD IN

lng	lat	categories	name	
103.847333	1.279925	Wine Bar	Napoleon Food & Wine Bar	0
103.847287	1.279872	Deli / Bodega	Park Bench Deli	1
103.846844	1.280135	Cocktail Bar	Native	2
103.849813	1.280017	Hotel	Sofitel So Singapore	3
103.846710	1.279371	Asian Restaurant	Pepper Bowl	4
103.848513	1.281254	Beer Garden	Freehouse	5
103.848188	1.277814	Café	Mellower Coffee	6
103.850127	1.279084	Indian Restaurant	Anglo Indian Cafe & Bar	7
103.847410	1.280205	Southern / Soul Food Restaurant	Meat Smith	8
103.850235	1.280261	Street Food Gathering	Lau Pa Sat Satay Street	9

MANHATTAN MAP - NEIGHBORHOODS AND CLUSTER OF VENUES



GEODATA MANHATTAN APTS FOR RENT

```
mh_rent=pd.read_excel('https://s3.eu-geo.objectstorage.softlayer.net/pythonbasicsfordatascienceproject-donotdelete-pr-zsoiumr6ebf53b/3333.xlsx?response-content-disposition=at
mh_rent.head()
```

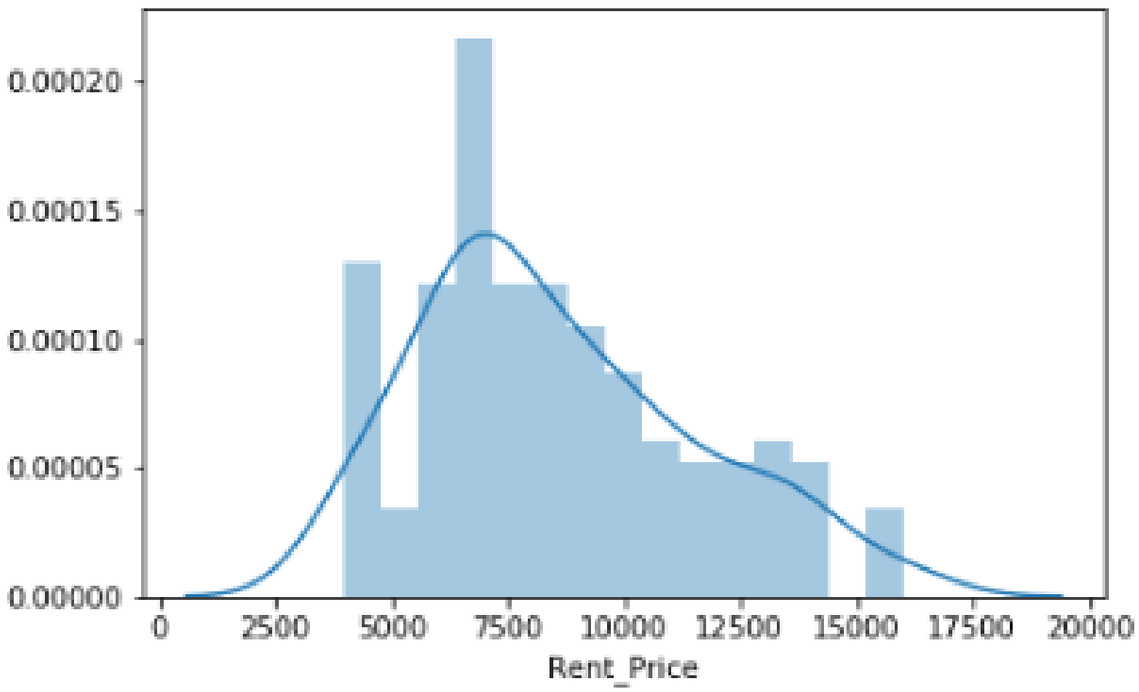
	Long	Lat	Rent_Price	Area-ft2	Rooms	Price_per_ft2	Area	Address
0	NaN	NaN	10000	3400	5	2.94	Upper West Side	West 105th Street
1	NaN	NaN	7500	2100	3	3.57	Upper East Side	East 97th Street
2	NaN	NaN	5300	2800	4	1.89	Upper West Side	West 105th Street
3	NaN	NaN	5000	1650	2	3.03	West Village	.CARMINE ST
4	NaN	NaN	5000	1450	2	3.45	Chelsea	.W 23RD ST 171

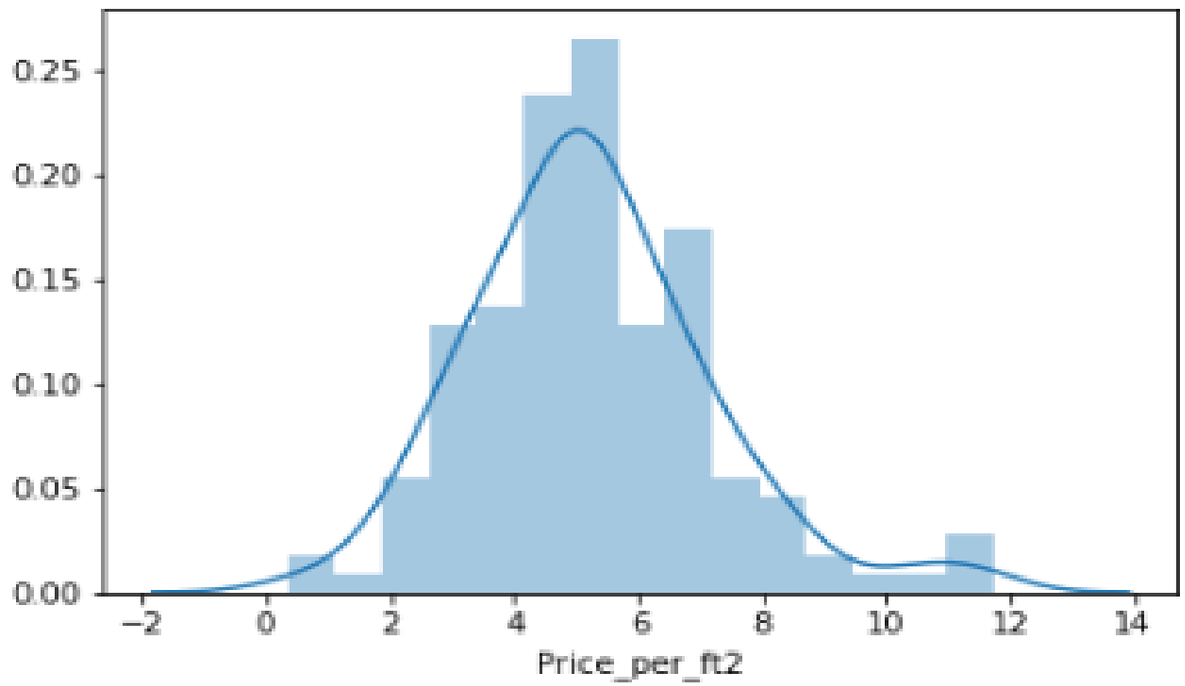
```
mh_rent.tail()
```

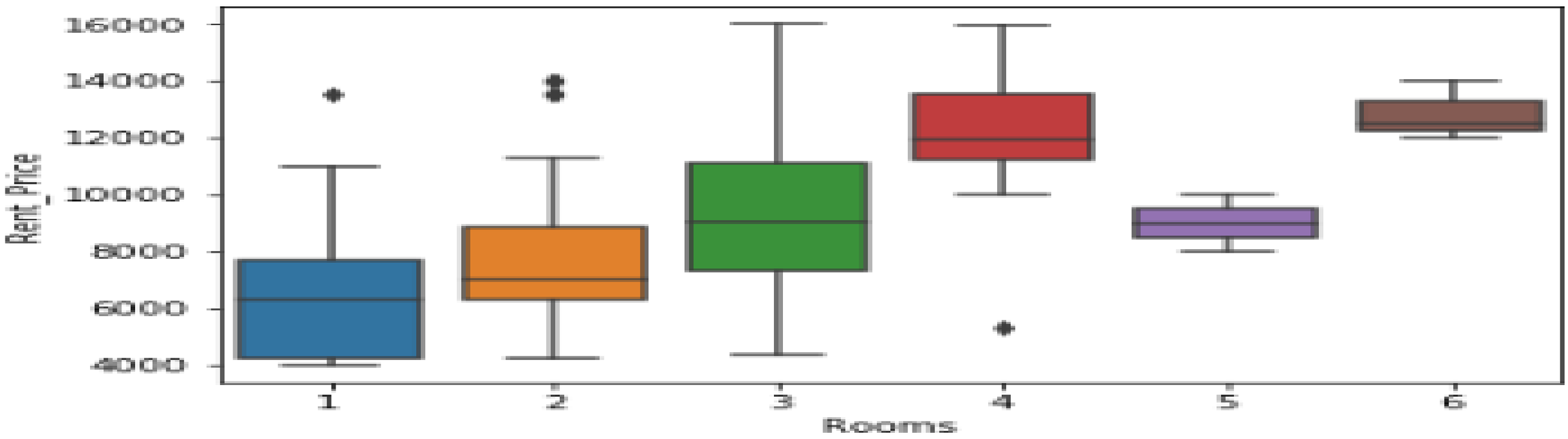
	Long	Lat	Rent_Price	Area-ft2	Rooms	Price_per_ft2	Area	Address
139	NaN	NaN	8750	1700	3	5.15	Rental in Lenox Hill	East 72nd Street 200
140	NaN	NaN	8700	1223	2	7.11	No fee rental in Tribeca	Murray Street 50
141	NaN	NaN	8118	2100	3	3.87	No fee rental in Midtown East	East 56th Street 300
142	NaN	NaN	8095	1600	2	5.06	No fee rental in Central Park West	Broadway 1930
143	NaN	NaN	10000	1500	2	6.67	Rental in Greenwich Village	West 9th Street 33

RENTAL PRICE STATISTICS MH APARTMENTS

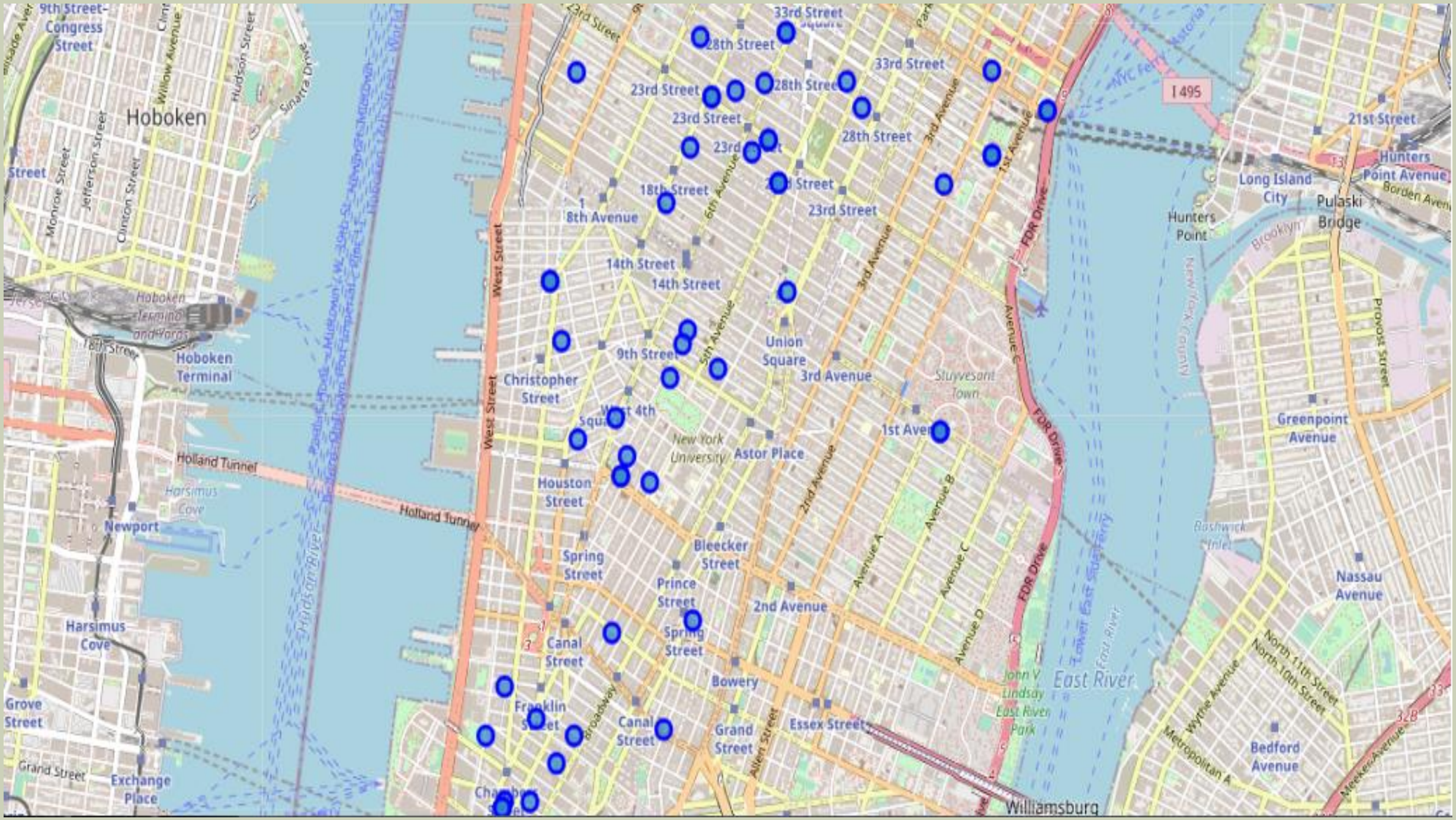
/MONTH IS AROUND THE MEAN7000BUDGET US



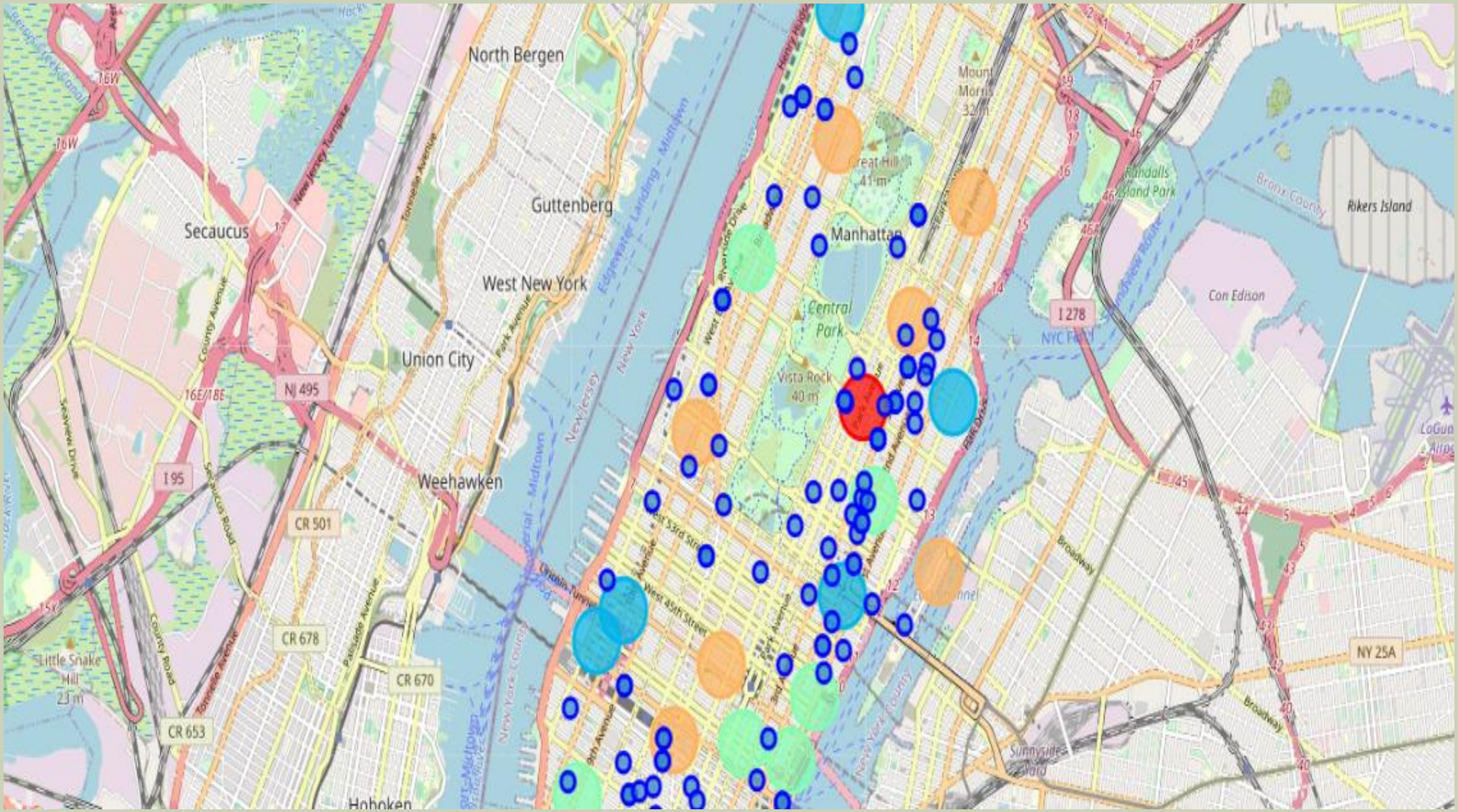




APARTMENTS FOR RENT IN MH



MH APTS FOR RENT WITH VENUE CLUSTERS



VENUES OF CLUSTER 3

10th Most Common Venue	9th Most Common Venue	8th Most Common Venue	7th Most Common Venue	6th Most Common Venue	5th Most Common Venue	4th Most Common Venue	3rd Most Common Venue	2nd Most Common Venue	1st Most Common Venue	Neighborhood	
Spanish Restaurant	Frozen Yogurt Shop	Park	American Restaurant	Bakery	Wine Bar	Café	Pizza Place	Lounge	Mexican Restaurant	Inwood	3
Other Nightlife	Bike Trail	Falafel Restaurant	Coffee Shop	Beer Garden	Sushi Restaurant	Mexican Restaurant	Seafood Restaurant	Italian Restaurant	Deli / Bodega	Manhattanville	5
Thai Restaurant	Sporting Goods Shop	Gym	Deli / Bodega	Burger Joint	Pizza Place	Gym / Fitness Center	Coffee Shop	Italian Restaurant	Sushi Restaurant	Lenox Hill	10
Sushi Restaurant	Mexican Restaurant	Wine Bar	Cosmetics Shop	Coffee Shop	Indian Restaurant	Vegetarian / Vegan Restaurant	Bakery	Bar	Italian Restaurant	Upper West Side	12
Italian Restaurant	Bar	French Restaurant	Burger Joint	Salon / Barbershop	Coffee Shop	Gym / Fitness Center	Japanese Restaurant	Hotel	Sandwich Place	Murray Hill	16
Hotel	American Restaurant	Seafood Restaurant	Art Gallery	Theater	Nightclub	Bakery	Ice Cream Shop	Italian Restaurant	Coffee Shop	Chelsea	17
Electronics Store	Seafood Restaurant	Bakery	Indian Restaurant	Café	Chinese Restaurant	Clothing Store	French Restaurant	Sushi Restaurant	Italian Restaurant	Greenwich Village	18
Wine Shop	Grocery Store	Mexican Restaurant	Pizza Place	Coffee Shop	Bagel Shop	Cocktail Bar	Thrift / Vintage Store	Restaurant	Italian Restaurant	Gramercy	27
Gym / Fitness Center	Park	Pizza Place	Italian Restaurant	Bar	Steakhouse	Wine Shop	Gym	Hotel	Coffee Shop	Financial District	29
Coffee Shop	Sushi Restaurant	Hotel	Mexican Restaurant	Grocery Store	Bookstore	Gift Shop	Cocktail Bar	French Restaurant	Italian Restaurant	Noho	31
Park	Yoga Studio	Gym	Coffee Shop	Sandwich Place	French Restaurant	Cocktail Bar	Italian Restaurant	Bakery	Gym / Fitness Center	Civic Center	32
French Restaurant	Japanese Restaurant	Indian Restaurant	Noodle House	Hotel	Sushi Restaurant	Wine Bar	Steakhouse	Coffee Shop	Italian Restaurant	Turtle Bay	35
Dog Run	Diner	Deli / Bodega	Hotel	Sushi Restaurant	Greek Restaurant	Mexican Restaurant	Pizza Place	Park	Café	Tudor City	36
Cycle Studio	Cosmetics Shop	Clothing Store	Bakery	Vegetarian / Vegan Restaurant	Yoga Studio	Gym / Fitness Center	Gym	American Restaurant	Italian Restaurant	Flatiron	38

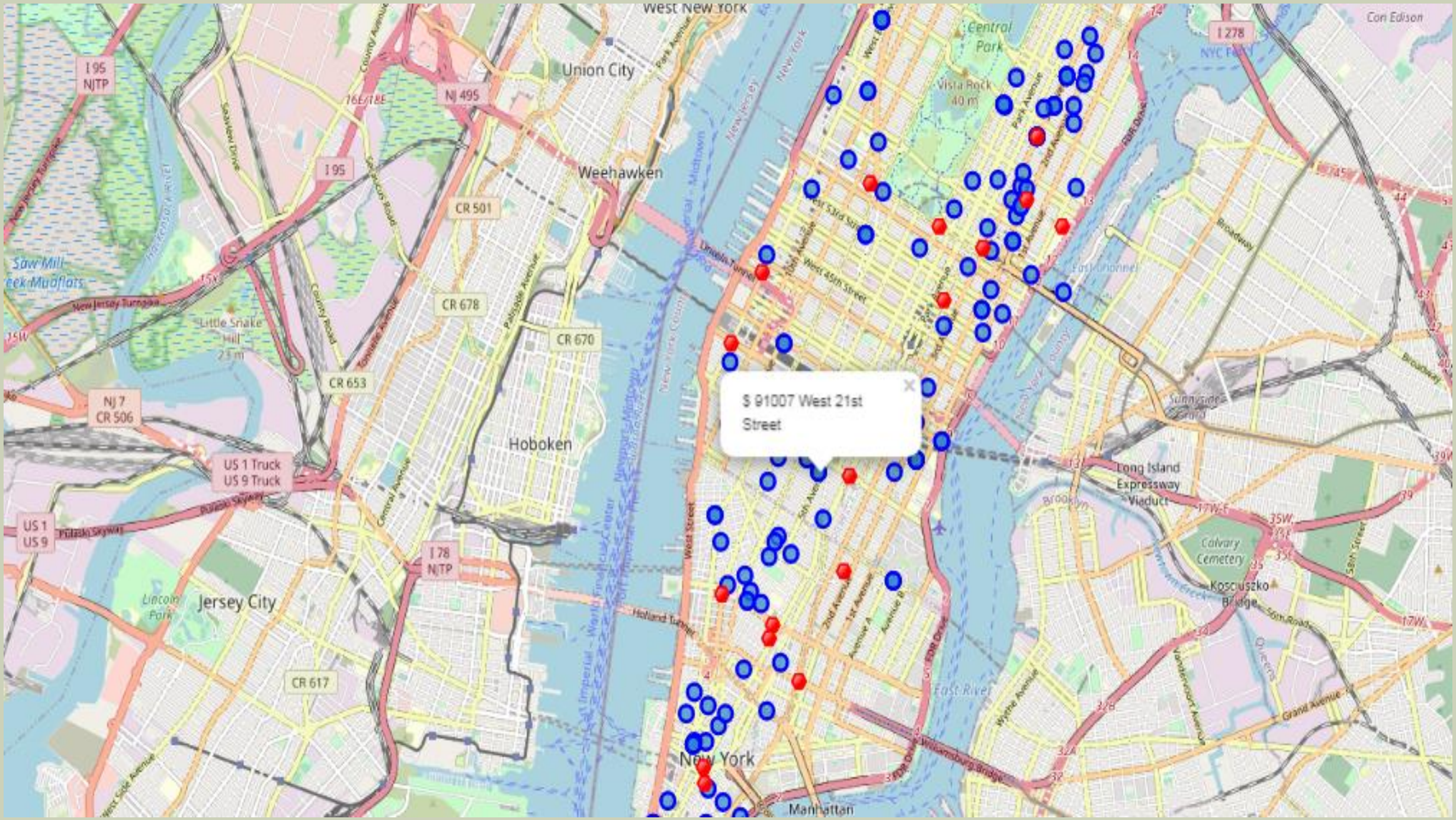
MANHATTAN SUBWAY STATIONS GEODATA

long	lat	sub_address	sub_station	
73.924509-	40.861857	Nagle Ave, New York, NY 10034, USA 170	Dyckman Street Subway Station	0
73.954525-	40.764250	New York, NY 10106, USA	Street Subway Station 57	1
73.987156-	40.730862	New York, NY 10005, USA	Broad St	2
73.939785-	40.847991	W 177th St, New York, NY 10033, USA 807	Street Station 175	3
73.954525-	40.764250	New York, NY 10022, USA	Av and 53 St 5	4

removing duplicate rows and creating new set mhsub1

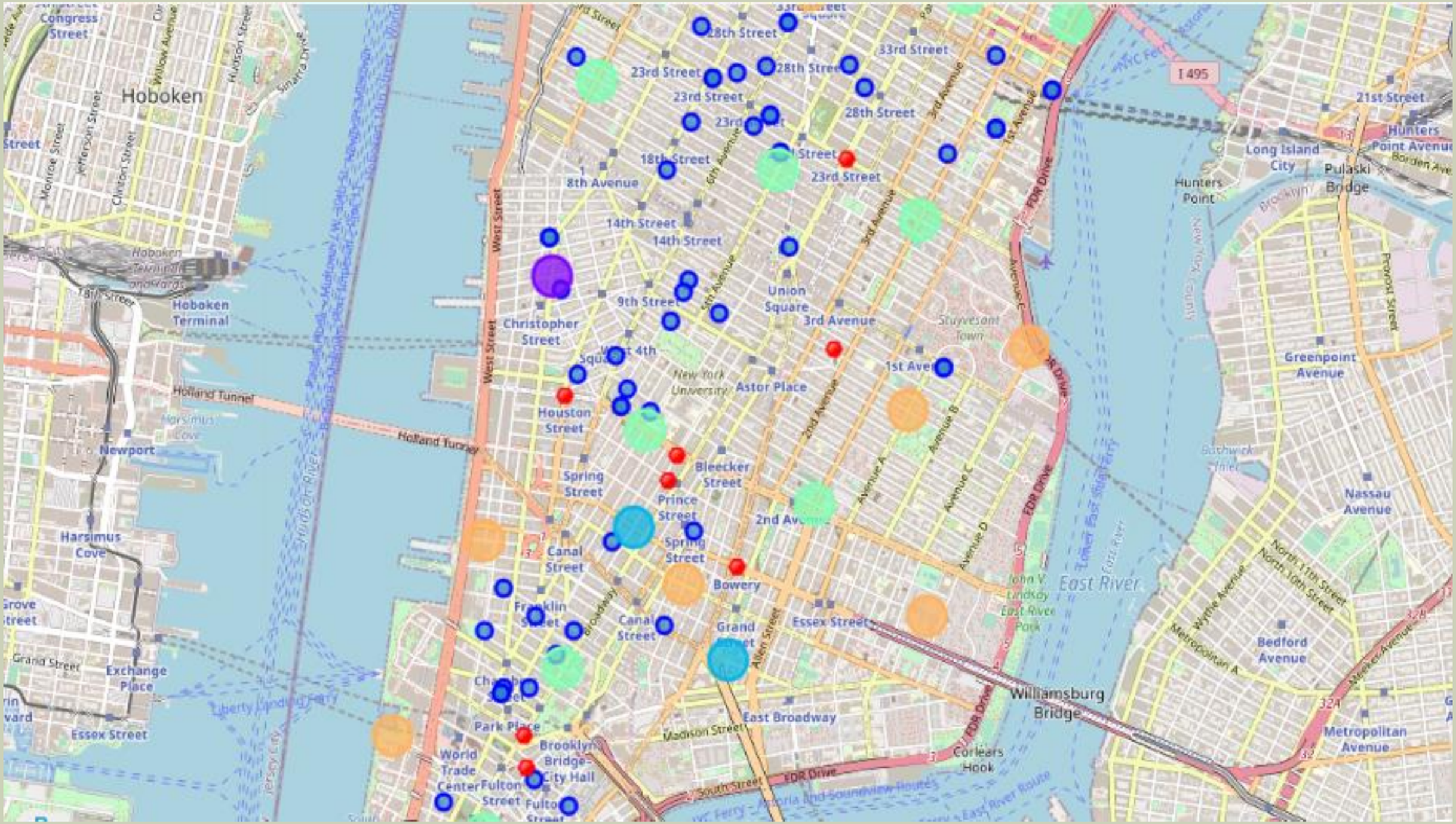
long	lat	sub_address	sub_station	
73.932983-	40.858113	Bennett Ave, New York, NY 10040, USA	Street Subway Station 190	17
73.966271-	40.762259	E 60th St, New York, NY 10065, USA	St-Lexington Av Station 59	18
73.954525-	40.764250	New York, NY 10019, United States	Street Station 57	19
73.987156-	40.730862	New York, NY 10014, United States	Street / 8 Av 14	20
73.999282-	40.759809	11th Ave, New York, NY 10018, USA 525	MTA New York City	21

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.

VENUS IN CLUSTER 2 NEAR FUTURE HOME

Venues for Apartment 1 - Cluster 2

```
kk = 2
manhattan_merged.loc[manhattan_merged['Cluster Labels'] == kk, manhattan_merged.columns[[1] + list(range(5, manhattan_merged.shape[1]))]]
```

10th Most Common Venue	9th Most Common Venue	8th Most Common Venue	7th Most Common Venue	6th Most Common Venue	5th Most Common Venue	4th Most Common Venue	3rd Most Common Venue	2nd Most Common Venue	1st Most Common Venue	Neighborhood	
Seafood Restaurant	Bank	Gym	Shoe Store	Tennis Stadium	Supplement Shop	Steakhouse	Yoga Studio	Discount Store	Coffee Shop	Marble Hill	0
Ice Cream Shop	Bubble Tea Shop	Bakery	Noodle House	Salon / Barbershop	Vietnamese Restaurant	American Restaurant	Dim Sum Restaurant	Cocktail Bar	Chinese Restaurant	Chinatown	1
Gym / Fitness Center	Beer Bar	Liquor Store	Event Space	Chinese Restaurant	Cosmetics Shop	American Restaurant	French Restaurant	Seafood Restaurant	African Restaurant	Central Harlem	6
Pub	Japanese Restaurant	Deli / Bodega	Mexican Restaurant	Pizza Place	Sushi Restaurant	Italian Restaurant	Bar	Gym	Coffee Shop	Yorkville	9
Indie Theater	Gym	Spa	Wine Shop	Hotel	Gym / Fitness Center	American Restaurant	Coffee Shop	Italian Restaurant	Theater	Clinton	14
Design Studio	Art Gallery	Mediterranean Restaurant	Italian Restaurant	Furniture / Home Store	Men's Store	Shoe Store	Women's Store	Boutique	Clothing Store	Soho	23
Tennis Court	Deli / Bodega	Café	Burger Joint	Sandwich Place	Pizza Place	Bookstore	Park	American Restaurant	Coffee Shop	Morningside Heights	26
Sushi Restaurant	Boutique	Juice Bar	Bakery	American Restaurant	Dessert Shop	Indian Restaurant	Furniture / Home Store	Italian Restaurant	Gym / Fitness Center	Sutton Place	34
Gym	Restaurant	Thai Restaurant	Gym / Fitness Center	Café	American Restaurant	Theater	Hotel	Italian Restaurant	Coffee Shop	Hudson Yards	39

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.

- One must keep abreast of new tools for DS that continue to appear for application in several business fields.