

Coursera Capstone project

Coursera IBM Data Science Certification

Final Report

Tejeshwar Parmar

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-york>
- 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:

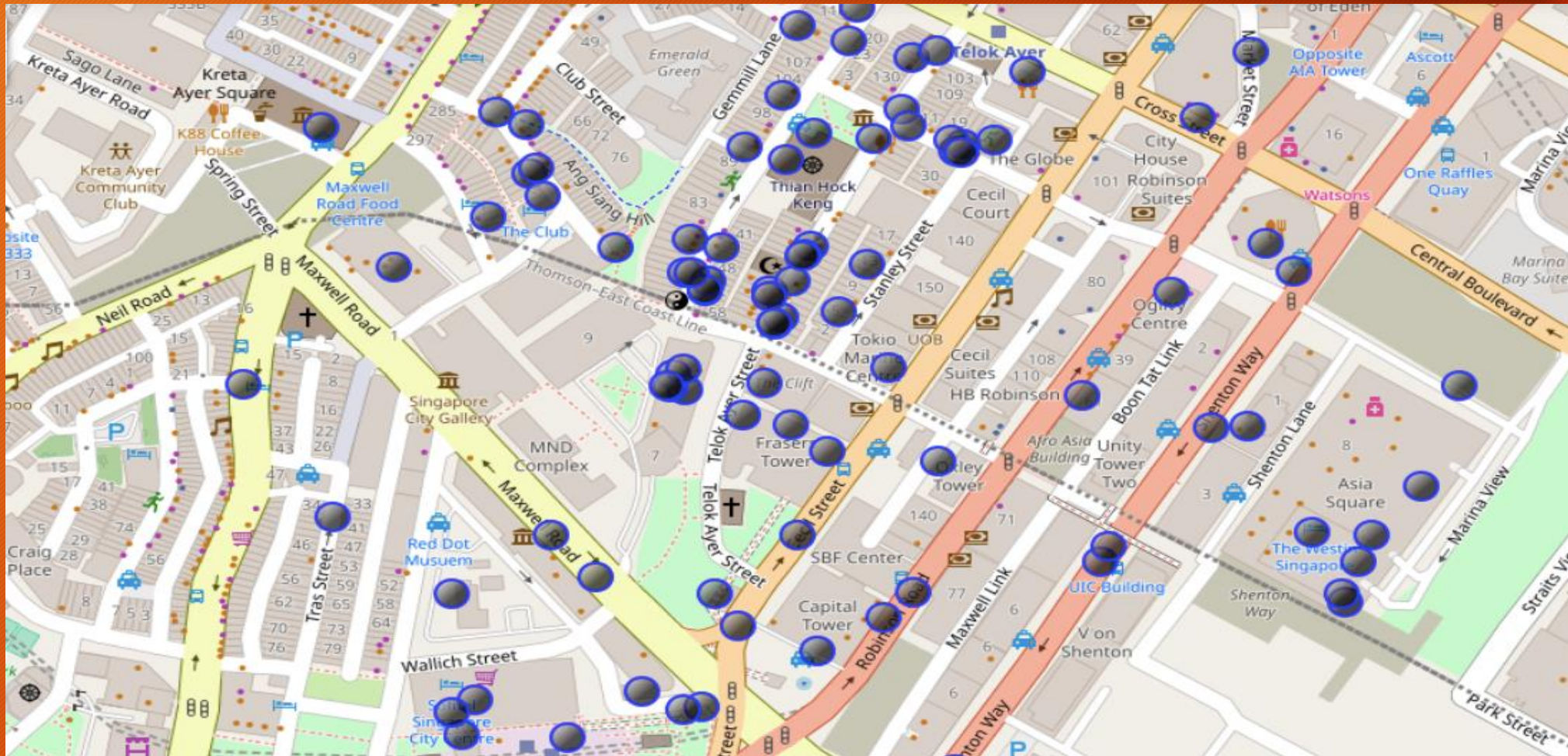
Web-scraping of sites is used to consolidate data-frame information which was saved as csv files for convenience and to simplify the report. Geodata was obtained by coding a program to use Nominatim to get latitude and longitude of subway stations and also for each of (144 units) the apartments for rent 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

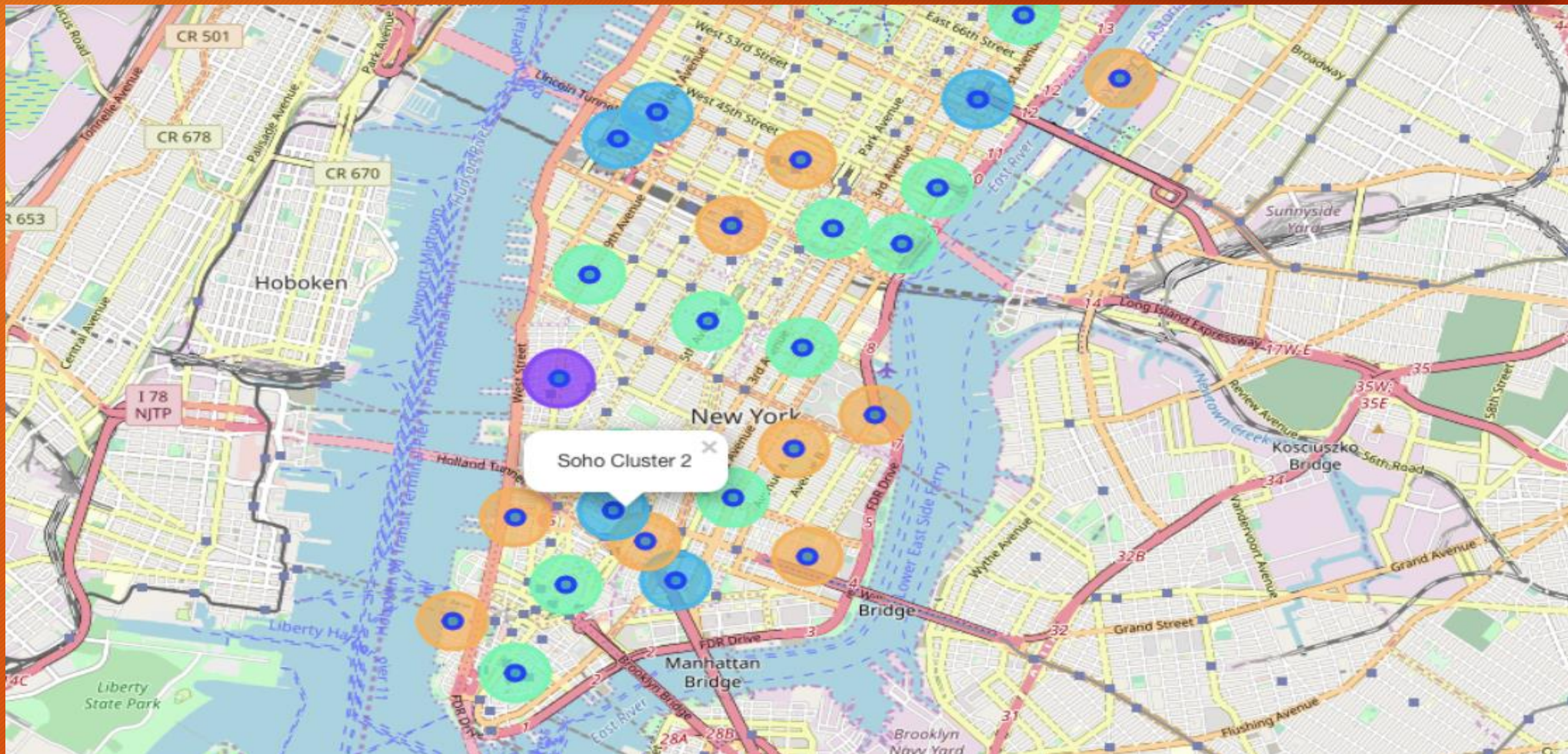


Venues around Neighborhood in

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

| | name | categories | lat | lng |
|---|--|-------------------|----------|------------|
| 0 | Napoleon Food & Wine Bar | Wine Bar | 1.279925 | 103.847333 |
| 1 | Park Bench Deli | Deli / Bodega | 1.279872 | 103.847287 |
| 2 | Native | Cocktail Bar | 1.280135 | 103.846844 |
| 3 | Muchachos | Burrito Place | 1.279175 | 103.847082 |
| 4 | Matt's The Chocolate Shop | Dessert Shop | 1.280462 | 103.846950 |
| 5 | Freehouse | Beer Garden | 1.281254 | 103.848513 |
| 6 | PS.Cafe | Café | 1.280468 | 103.846264 |
| 7 | 왕대박 Wang Dae Bak Korean BBQ Restaurant | Korean Restaurant | 1.281345 | 103.847551 |
| 8 | Ancient Therapy | Massage Studio | 1.280413 | 103.847481 |
| 9 | Oven & Fried Chicken | Korean Restaurant | 1.280479 | 103.847522 |

Manhattan Map - Neighborhoods and Cluster of Venues



GeoData Manhattan apts for rent

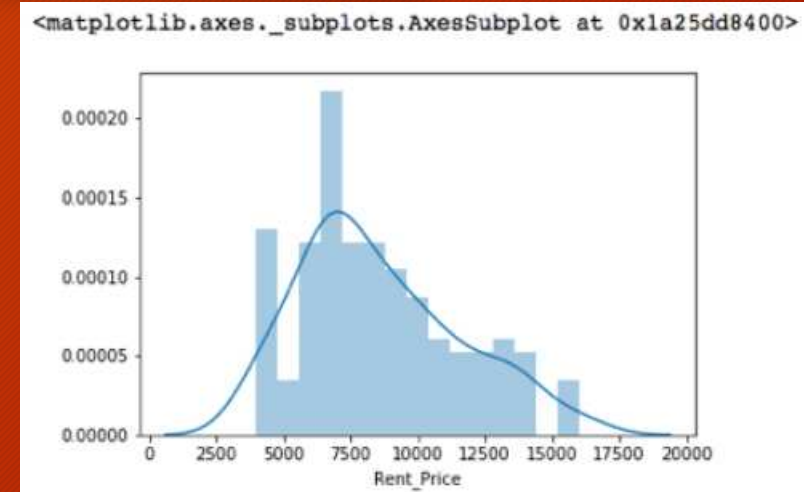
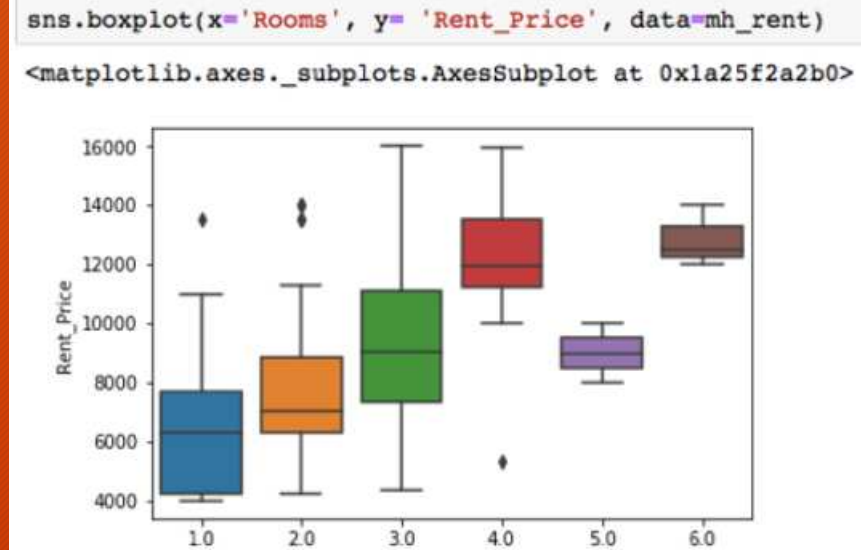
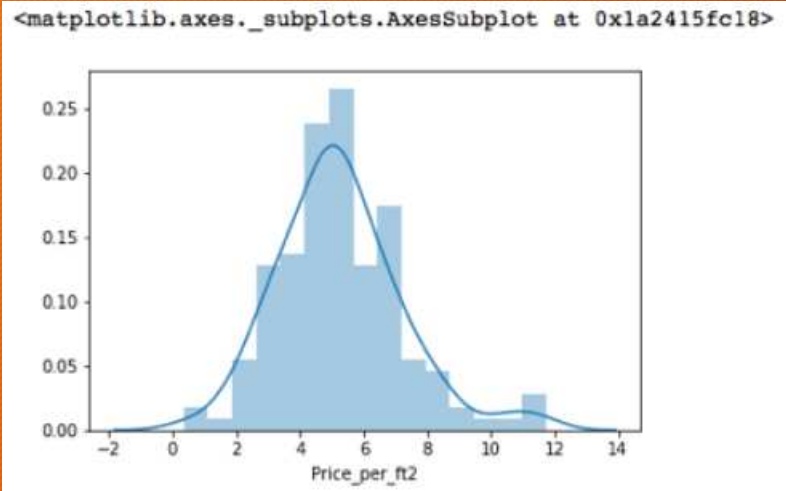
```
mh_rent=pd.read_csv('MH_rent_latlong.csv')
mh_rent.head()
```

| | 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 |

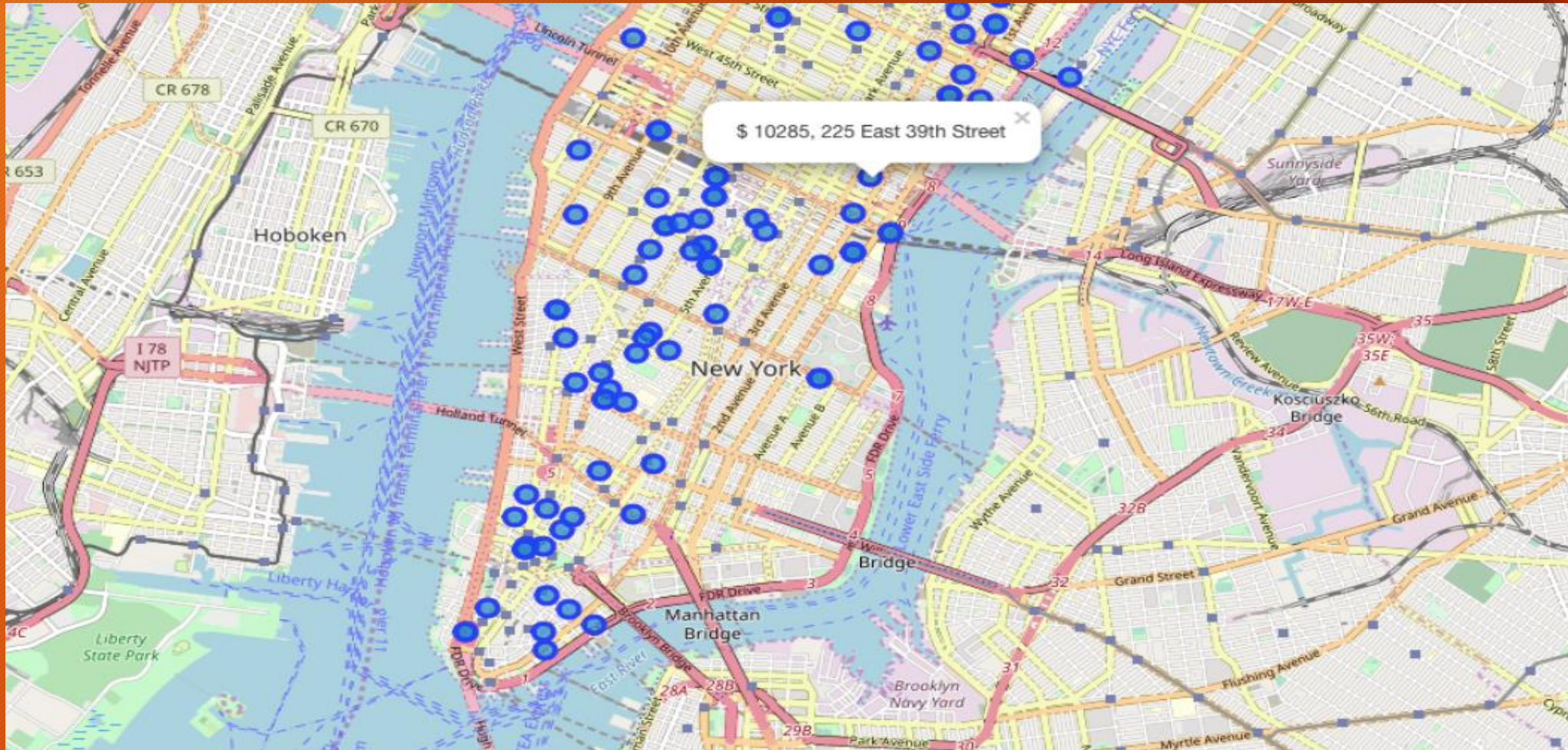
```
mh_rent.tail()
```

| | 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 |

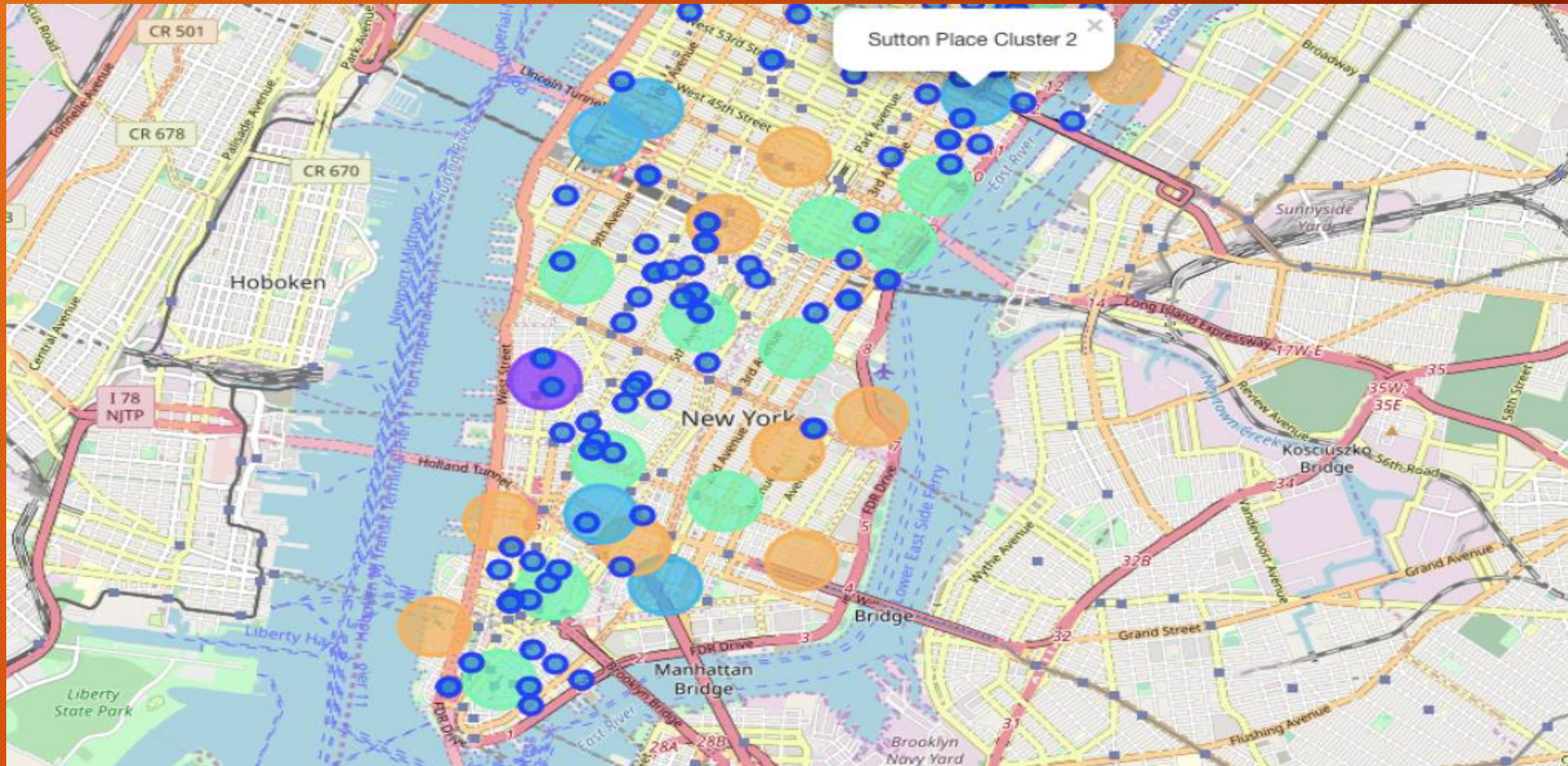
Rental Price Statistics MH Apartments



Apartments for Rent in MH



MH apt 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 |
| 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 |

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 | -73.924509 |
| 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 |

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

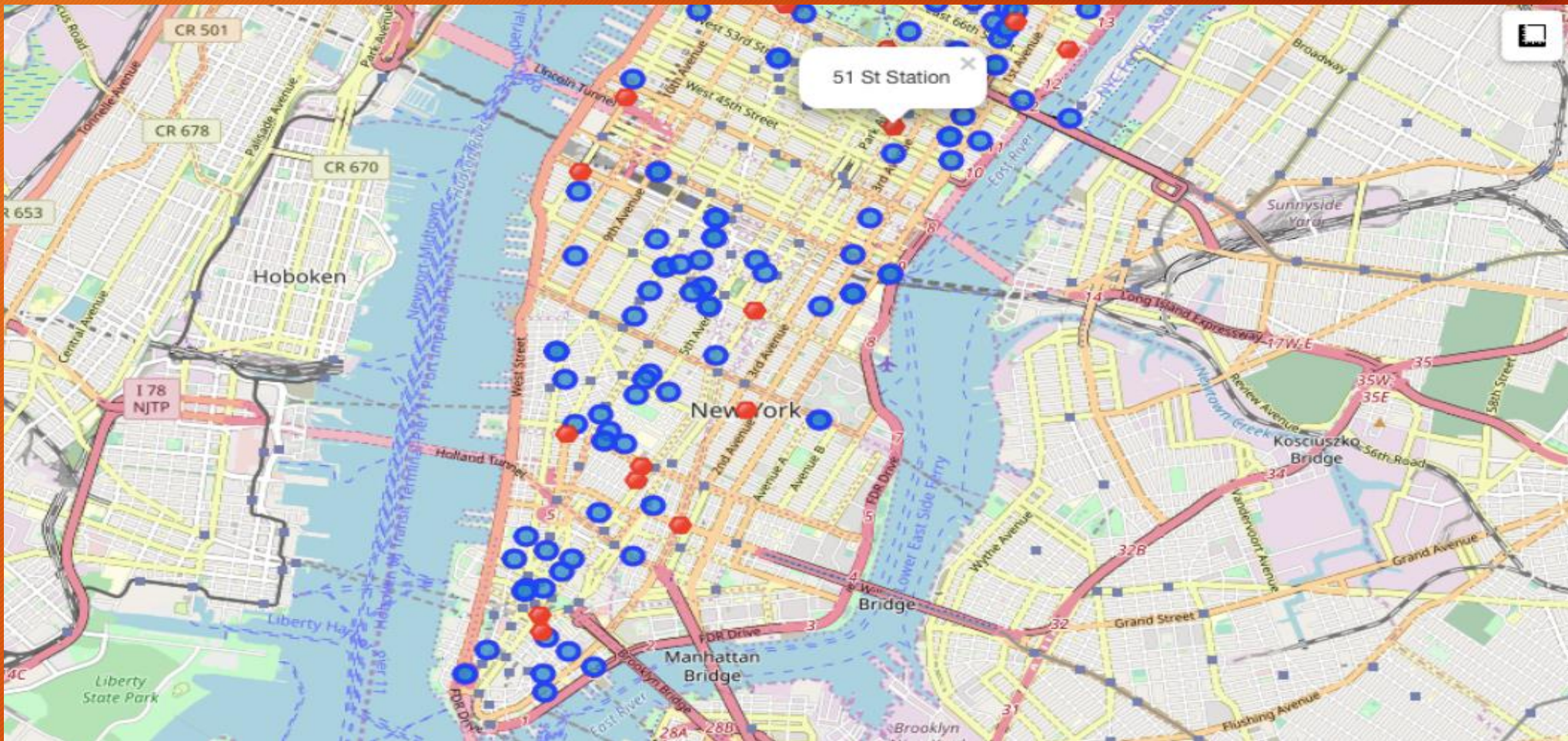
```
(22, 4)
```

```
: 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 | -73.999282 |

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

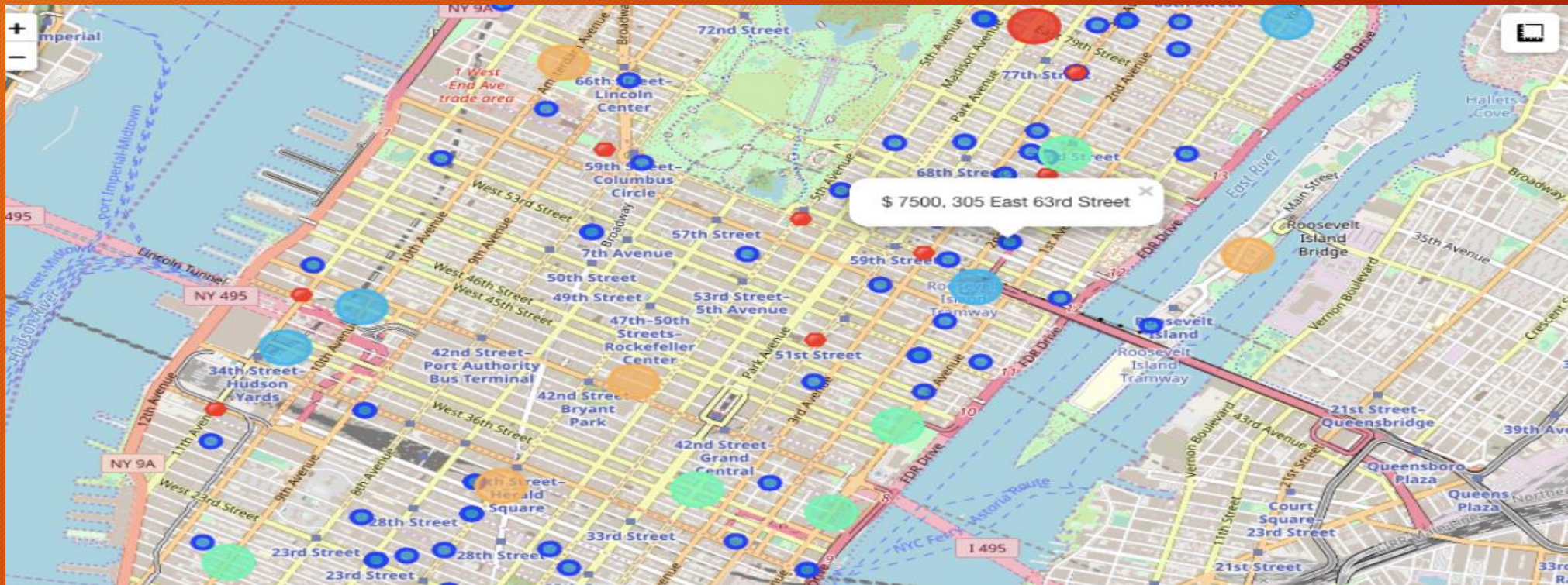


Selected Apartment!

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

```
## 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_
```

| | 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 |
| 34 | Sutton Place | Gym / Fitness Center | Italian Restaurant | Furniture / Home Store | Indian Restaurant | Dessert Shop | American Restaurant | Bakery | Juice Bar | Boutique | Sushi Restaurant |
| 39 | Hudson Yards | Coffee Shop | Italian Restaurant | Hotel | Theater | American Restaurant | Café | Gym / Fitness Center | Thai Restaurant | Restaurant | Gym |

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.

Conclusions

- 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.