

Finding a Comfortable Living Area in Manhattan, NY

Sandip Pattanayak

06th December, 2019

REPORT CONTENT

1. 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.

2. 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

3. 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.

4. Results section

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

5. Discussion section

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

6. Conclusion section

Decision taken and Report Conclusion

1. Introduction

Background:

I am a Learner data scientist residing in India. I currently live in Kolkata under the state West Bengal with my friend Sumit and we enjoy many amenities and venues in the area, such as various cuisine restaurants, cafes, food shops and entertainment. Sumit has been offered a great opportunity to work in Manhattan, NY. Although, He is a bit stressed about a comfortable place to live in Manhattan. Therefore, 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

Business Problem:

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 rental 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

As a reference, I have included a map of venues near current residence.

Interested Audience

I believe this is a relevant challenge with valid questions for anyone moving to other large city in US, EU or Asia. The same methodology can be applied in accordance to demands as applicable. This case is also applicable for anyone interested in exploring starting or locating a new business in any city. Lastly, it can also serve as a good practical exercise to develop Data Science skills.

2. Data Section:

Description of the data and its sources that will be used to solve the problem

2.1 Data of Current Situation

I currently reside in the Park Street area of Kolkata in India. I use Foursquare to identify the venues around the area of residence which are then shown in the India map shown in methodology and execution in section 3.0. It serves as a reference for comparison with the desired future location in Manhattan NY NEAR PARK AVENUE AND 53 STRT RD.

2.2 Description of the Data:

The following data is required to answer the issues of the problem:

1. List of Boroughs and neighbourhoods of Manhattan with their geodata (latitude and longitude)
2. List of Subway metro stations in Manhattan with their address location
3. List of apartments for rent in Manhattan area with their addresses and price
4. Preferably, a list of apartment for rent with additional information, such as price, address, area, # of beds, etc.
5. Venues for each Manhattan neighbourhood (than can be clustered)
6. Venues for subway metro stations, as needed

2.3 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 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 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 'mhtn_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 geodata from Nominatim .

A list of Manhattan subway metro stops was compiled in Numbers (Apple excel) and it was completed with Wikipedia data (https://en.wikipedia.org/wiki/List_of_New_York_City_Subway_stations_in_Manhattan) and information from NY Transit authority and Google maps (<https://www.google.com/maps/search/manhattan+subway+metro+stations/@40.7837297,-74.1033043,11z/data=!3m1!4b1>) for a final consolidated list of subway stops names and their address. The geo location was obtained via an algorithm using Nominatim. Details will be shown in the execution of methodology in section 3.0. The subway csv file is "mhtn_subway.csv".

A list of places for rent was collected by web-browsing real estate companies in Manhattan. A csv file was compiled with the rental place that indicated: areas of Manhattan, address, number of beds, area and monthly rental price. An algorithm was used to create all the geodata using Nominatim, as shown in section 3.0. With the use of geolocator = Nominatim() , it was possible to determine the latitude and longitude for the subway metro locations as well as for the geodata for each rental place listed. The loop algorithms used are shown in the execution of data in section 3.0 "Great circle" function from geolocator was used to calculate distances between two points, as in the case to

calculate average rent price for units around each subway station and at 1.6 km radius. Foursquare is used to find the avenues at Manhattan neighbourhoods in general and a cluster is created to later be able to search for the venues depending of the location shown.

2.4 How the data will be used to solve the problem

The data will be used as follows:

Use Foursquare and geopy data to map top 10 venues for all Manhattan neighbourhoods and clustered in groups (as per Course LAB) Use foursquare and geopy data to map the location of subway metro stations, separately and on top of the above clustered map in order to be able to identify the venues and amenities near each metro station, or explore each subway location separately Use Foursquare and geopy data to map the location of rental places, in some form, linked to the subway locations. Create a map that depicts, for instance, the average rental price per square ft., around a radius of 1.0 mile (1.6 km) around each subway station - or a similar metrics. I will be able to quickly point to the popups to know the relative price per subway area. Addresses from rental locations will be converted to geodata (lat, long) using Geopy-distance and Nominatim. Data will be searched in open data sources if available, from real estate sites if open to reading, libraries or other government agencies such as Metro New York MTA, etc. The processing of these DATA will allow to answer the key questions to make a decision:

2.5 Mapping of Data

The following maps were created to facilitate the analysis and the choice of the palace to live. Manhattan map of Neighbourhoods Manhattan subway metro locations Manhattan map of places for rent Manhattan map of clustered venues and neighbourhoods combined maps of Manhattan rent places with subway locations combined maps of Manhattan rent places with subway locations and venues clusters

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.

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: location near a subway, rental price and similar venues to Singapore. This visual approach and maps with popups labels allow quick identification of location, price and feature, thus making the selection very easy.

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 trade-offs between size and price and location?
8. Any other interesting statistical data findings of the real estate and overall data.

4. Execution and Results

4.1 Current Residence neighbourhood in Kolkata



4.2 Venues around Neighbourhood

```
SGnearby_venues.head(10)
```

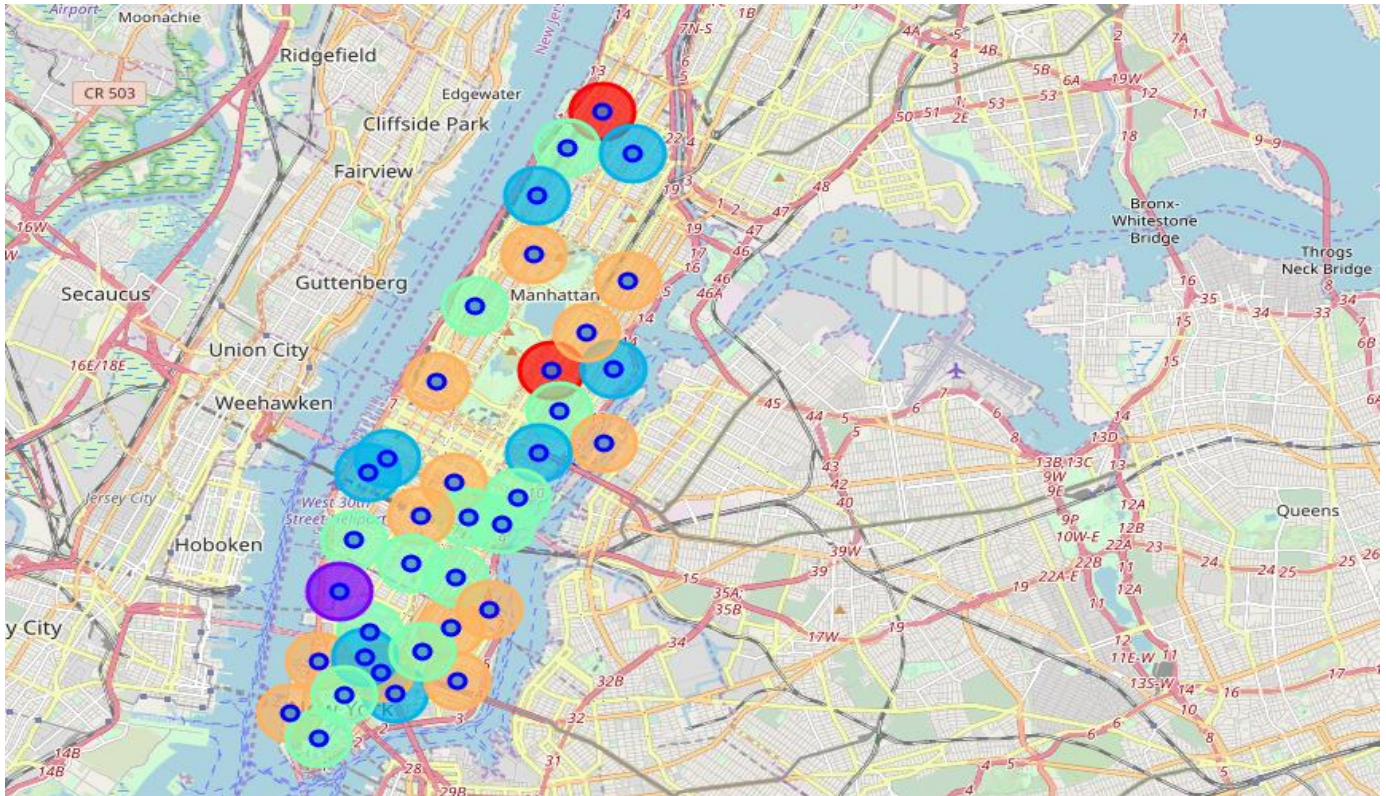
Out[65]:

click to scroll output; double click to hide

	name	categories	address	lat	lng
0	Aqua	Lounge	The Park	22.554734	88.352180
1	Oxford Bookstore	Bookstore	17, Park St	22.553652	88.351732
2	Tantra	Nightclub	The Park Hotel	22.553843	88.351459
3	Peter Cat	Indian Restaurant	18, Park St	22.552365	88.352544
4	Someplace Else	Pub	The Park	22.554250	88.351832
5	Trincas Tavern	Indian Restaurant	17B, Park St	22.553882	88.351506
6	Bar-B-Q	BBQ Joint	43, 47, 55, Park St	22.553125	88.352625
7	Flurys	Bakery	18, Park St	22.552786	88.352625
8	Mocambo	Restaurant	25B, Mirza Ghalib St	22.553206	88.353296
9	One Step Up	Restaurant	18A, Park St	22.553040	88.352372

4.3 Manhattan map- venue and cluster of Venues

popups allow to identify each neighbourhood and the cluster of venues around it in order to proceed to examine in more detail in the next cell



4.4 Geodata Manhattan Apartments for Rent

```
mh_rent.head()
```

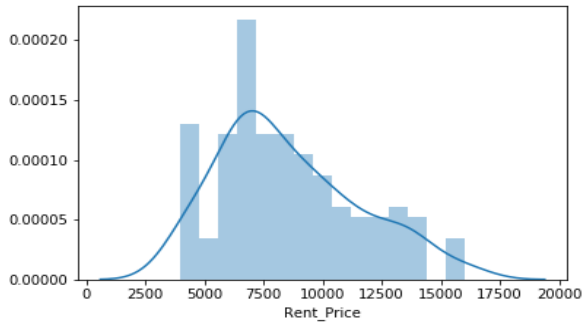
Out[18]:

	Address	Area	Price_per_ft2	Rooms	Area-ft2	Rent_Price	Lat	Long
0	West 105th Street	Upper West Side	2.94	5	3400	10000	40.799771	-73.966213
1	East 97th Street	Upper East Side	3.57	3	2100	7500	40.788585	-73.955277
2	West 105th Street	Upper West Side	1.89	4	2800	5300	40.799771	-73.966213
3	CARMINE ST.	West Village	3.03	2	1650	5000	40.730523	-74.001873
4	171 W 23RD ST.	Chelsea	3.45	2	1450	5000	40.744118	-73.995299

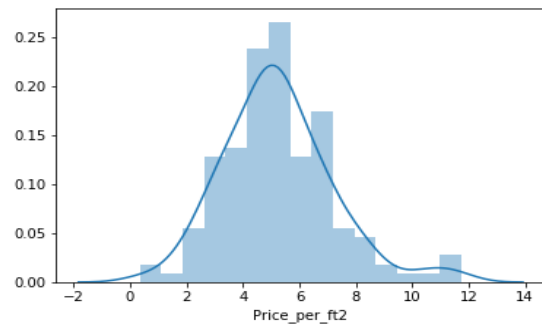
4.5 Rental Price statistics MH Apartments

Budget USD 7000 / Month is around the mean

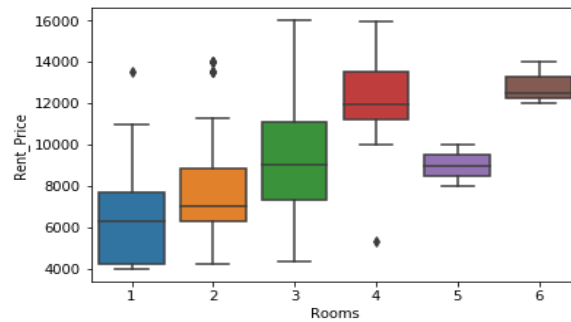
<matplotlib.axes._subplots.AxesSubplot at 0x7f947f874e80>



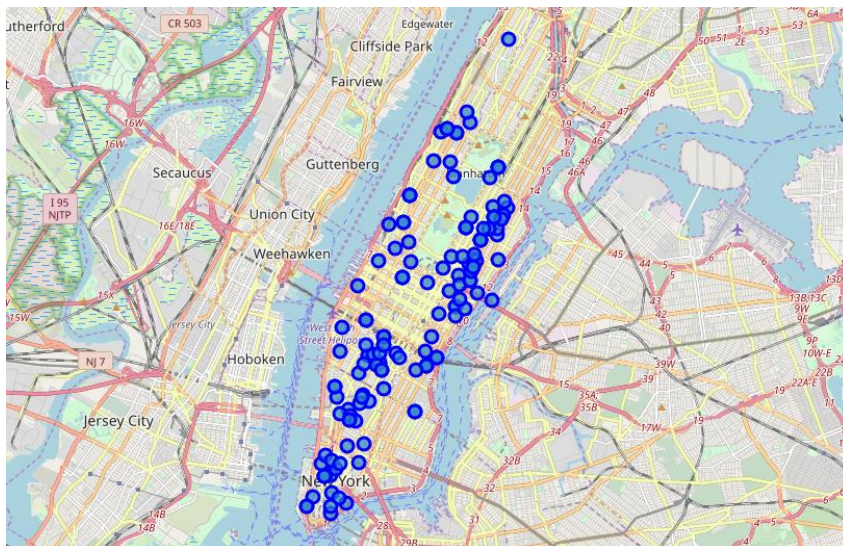
<matplotlib.axes._subplots.AxesSubplot at 0x7f1dbe9d3668>



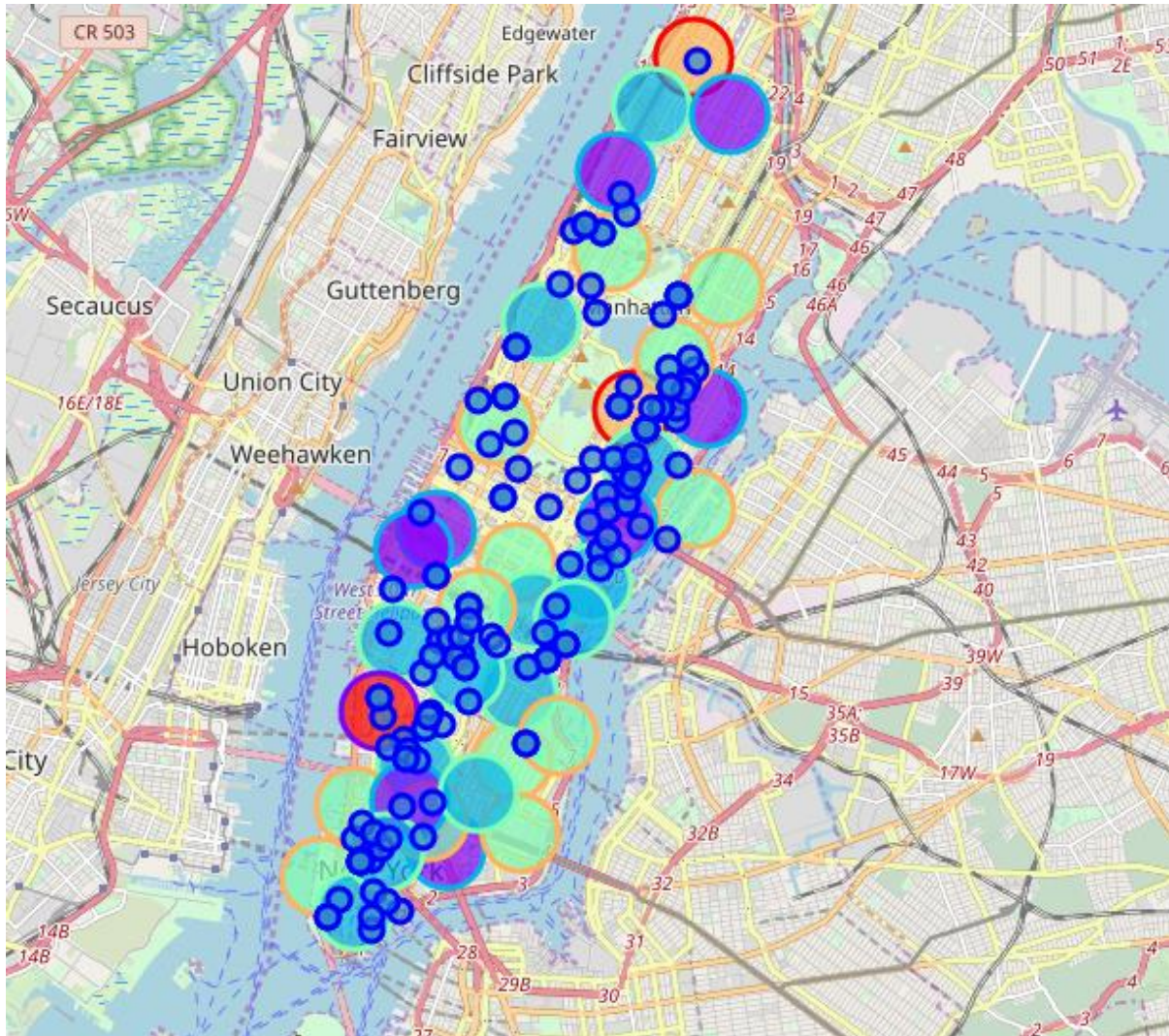
<matplotlib.axes._subplots.AxesSubplot at 0x7f947dd6a160>



4.6 Apartment for Rents In Manhattan



4.7 Manhattan Apts for rent with venue clusters



4.8 Venues of Cluster 3

	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
30	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
32	Civic Center	Gym / Fitness Center	Bakery	Italian Restaurant	Cocktail Bar	French Restaurant	Sandwich Place	Coffee Shop	Gym	Yoga Studio	Park
35	Turtle Bay	Italian Restaurant	Coffee Shop	Steakhouse	Wine Bar	Sushi Restaurant	Hotel	Noodle House	Indian Restaurant	Japanese Restaurant	French Restaurant
36	Tudor City	Café	Park	Pizza Place	Mexican Restaurant	Greek Restaurant	Sushi Restaurant	Hotel	Deli / Bodega	Diner	Dog Run
38	Flatiron	Italian Restaurant	American Restaurant	Gym	Gym / Fitness Center	Yoga Studio	Vegetarian / Vegan Restaurant	Bakery	Clothing Store	Cosmetics Shop	Cycle Studio

4.9 Manhattan Subway Station Geodata

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

Removing Duplicate Entries and Creating a New Dataframe

```
In [29]: # removing duplicate rows and creating new set mhsb1
mhsb1=mh.drop_duplicates(subset=['lat','long'], keep="last").reset_index(drop=True)
print(mhsb1.shape)
mhsb1.tail()
```

(22, 4)

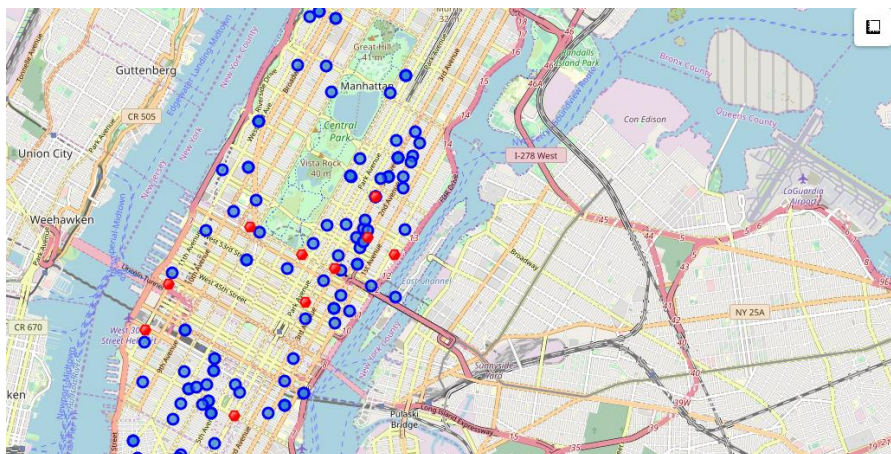
Out[29]:

	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

4.10 Apt for rent (blue) and subway station (red)

Now, we can visualize the desirable rental places and their nearest subway station. Popups display rental address and monthly rental price and the subway station name.

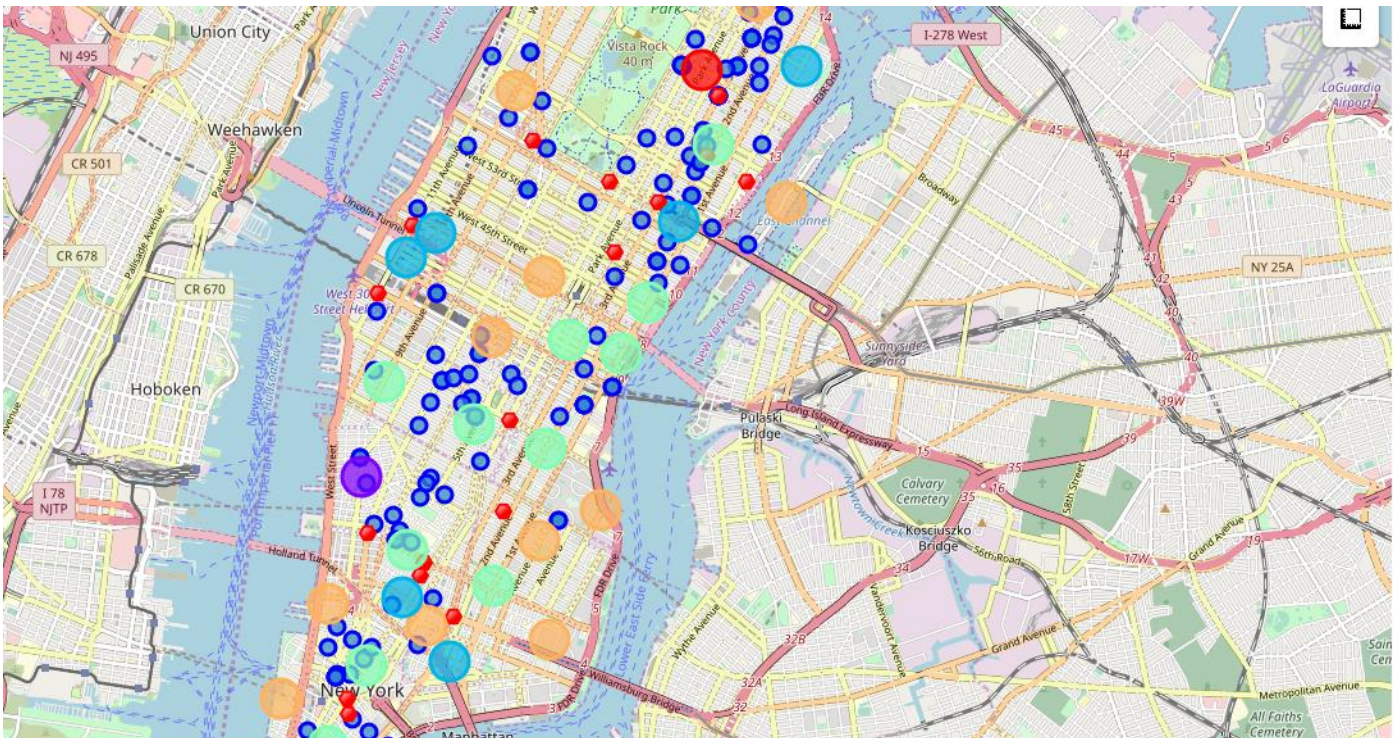
Notice that the icon in the top-right corner is a "ruler" that allows to measure the distance from a rental place to an specific subway station



4.11 Selected apartments

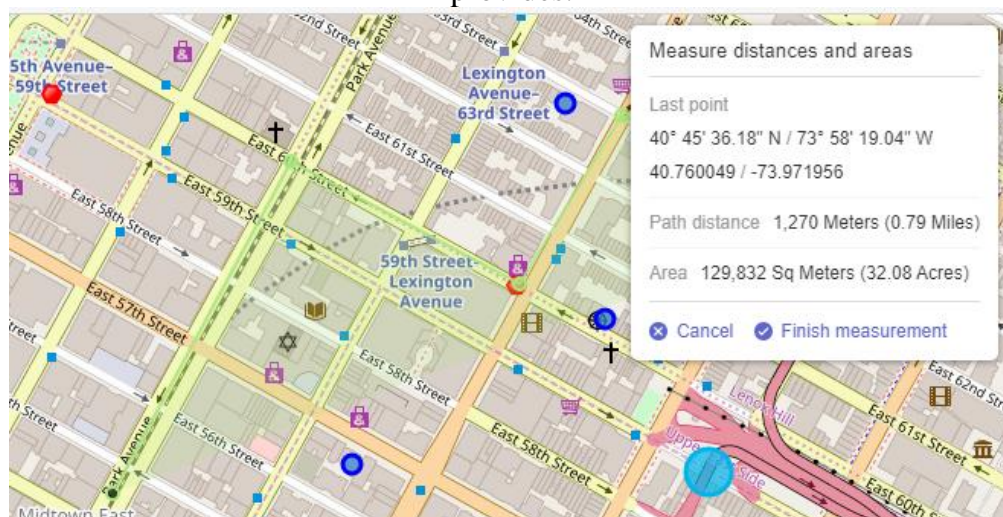
The one consolidated map that shows all information for decision: Apartment address, Price, Cluster of venues and subway station nearby.

Blue Dots = apts, Red dots = Subway stations, Bubbles = cluster of venues



4.12 Apartment Selection

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.



4.13 Venue is cluster 2 near Future living place

Venues for Apartment 1 - Cluster 2

```
## km is the cluster number to explore
```

```
km = 2
```

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

```
:
```

	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

5.0 DISCUSSION

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 US\$7500 slightly above the US\$7000 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 US\$6935, just under the US\$7000 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.

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.

6.0 CONCLUSIONS

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. It can be utilized effectively where people who are turning to new cities to start a business or work with their comfortable living options as per their desire.