Final Project - GAGAN BHATIA

Coursera Capstone - REPORT CONTENT

- Introduction Section: Discussion of the business problem and the interested audience in this project.
- Data Section: Description of the data that will be used to solve the problem and the sources.
- Methodology section: Discussion and description of exploratory data analysis carried out, any inferential statistical testing performed, and if any machine learnings were used establishing the strategy and purposes.
- Results section: Discussion of the results.
- Discussion section: Elaboration and discussion on any observations noted and any recommendations suggested based on the results.
- Conclusion section: Report Conclusion.

Introduction

I am a data scientist currently residing in Downtown Vancouver. I currently live within walking distance to Downtown Skytrain station therefore I have access to good public transportation to work. Likewise, I enjoy many amenities in the neighborhood, such as international cuisine restaurants, cafes, food shops and entertainment. I have been offered a great opportunity to work in Manhattan, NY. Although, I am very excited about it, I am a bit stress toward the process to secure a comparable place to live in Manhattan. Of course, there are alternatives to achieve the answer using available Google and Social media tools, but it is rewarding doing it myself with learned tools.

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 rent a unit subject to the following conditions:

- Apartment with min 2 bedrooms with monthly rent not to exceed US$7000/month
- Unit located within walking distance (<=1.0-mile, 1.6 km) from a subway metro station in Manhattan
- Area with amenities and venues similar to the ones described for current location

I believe this is a relevant project for a person or entity considering moving to a major city in Europe, US or Asia, since the approach and methodologies used here are applicable in all cases. 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 toward the development of Data Science skills.

	name	categories	lat	lng
0	UBC Museum of Anthropology	History Museum	49.269246	-123.259563
1	Chan Centre for the Performing Arts	Concert Hall	49.269538	-123.255597
2	UBC Rose Garden	Garden	49.269165	-123.256381
3	Wreck Beach	Beach	49.261041	-123.261623
4	Mercante	Italian Restaurant	49.263765	-123.255148
5	Koerner's Pub	Pub	49.268790	-123.258201
6	Beaty Biodiversity Museum	Science Museum	49.263347	-123.251000
7	Loafe Café	Coffee Shop	49.265903	-123.250335
8	Great Dane Coffee	Café	49.270049	-123.250737
9	Cecil Green Park House	Scenic Lookout	49.271348	-123.257754
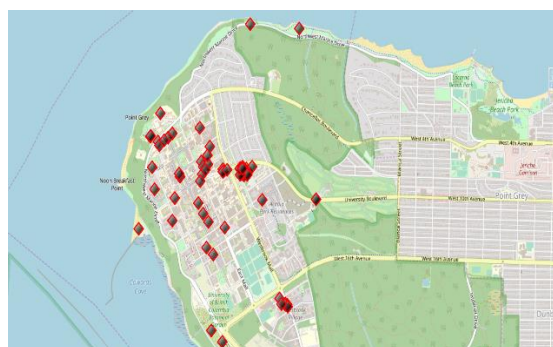
Current amenities available and its map

Data

Description of Data used

I Currently reside in the neighborhood of UBC in Vancouver. I use Foursquare to identify the venues around the area of residence which are then shown in the Vancouver 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

In order 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 amenities in the Manhattan neighborhoods (e.g. top 10) 2.3 sources and manipulation
- The list of Manhattan neighborhoods is worked out during Lab exercise during the course.

	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

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	Manhattan	Marble Hill	40.876551	-73.910660	2	Coffee Shop	Discount Store	Yoga Studio	Steakhouse	Supplement Shop	Tennis Stadium	Shoe Store	Gym	Bank	Seafood Restaurant
1	Manhattan	Chinatown	40.715618	-73.994279	2	Chinese Restaurant	Cocktail Bar	Dim Sum Restaurant	American Restaurant	Vietnamese Restaurant	Salon / Barbershop	Noodle House	Bakery	Bubble Tea Shop	Ice Cream Shop
2	Manhattan	Washington Heights	40.851903	-73.936900	4	Café	Bakery	Mobile Phone Shop	Pizza Place	Sandwich Place	Park	Gym	Latin American Restaurant	Tapas Restaurant	Mexican Restaurant
3	Manhattan	Inwood	40.867684	-73.921210	3	Mexican Restaurant	Lounge	Pizza Place	Café	Wine Bar	Bakery	American Restaurant	Park	Frozen Yogurt Shop	Spanish Restaurant
4	Manhattan	Hamilton Heights	40.823604	-73.949688	0	Mexican Restaurant	Coffee Shop	Café	Deli / Bodega	Pizza Place	Liquor Store	Indian Restaurant	Sushi Restaurant	Sandwich Place	Yoga Studio

|	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

A csv file was created which will be read in order to create a data frame and its mapping. The csv file 'mh_neigh_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 neighborhoods and mapping will be shown however.

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

- List of Boroughs and neighborhoods of Manhattan with their geodata (latitude and longitud)
- List of Subway metro stations in Manhattan with their address location
- List of apartments for rent in Manhattan area with their addresses and price
- Preferably, a list of apartment for rent with additional information, such as price, address, area, # of beds, etc.
- Venues for each Manhattan neighborhood (than can be clustered)
- Venues for subway metro stations, as needed

The data will be used as follows:

Use Foursquare and goopy data to map top 10 venues for all Manhattan neighborhoods 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.
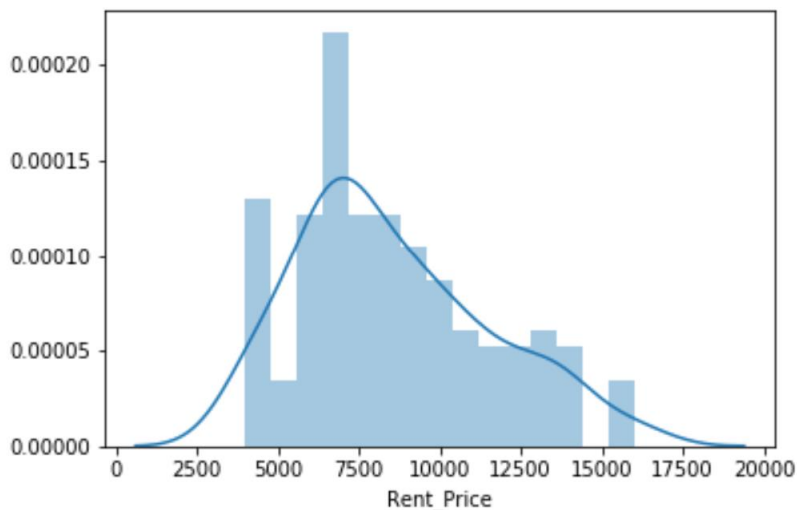
Addresses from rental locations will be converted to geodata(lat, long) using Geopy-distance and Nominatim. Data will be searched in open data sources if available, from real estate sites if open to reading, libraries or other government agencies such as Metro New York MTA, etc.
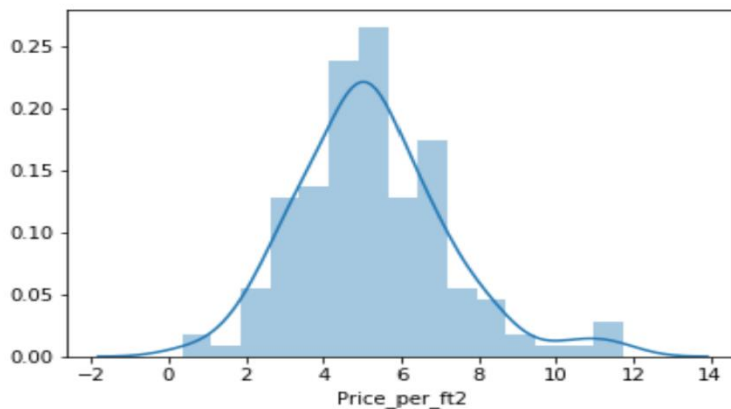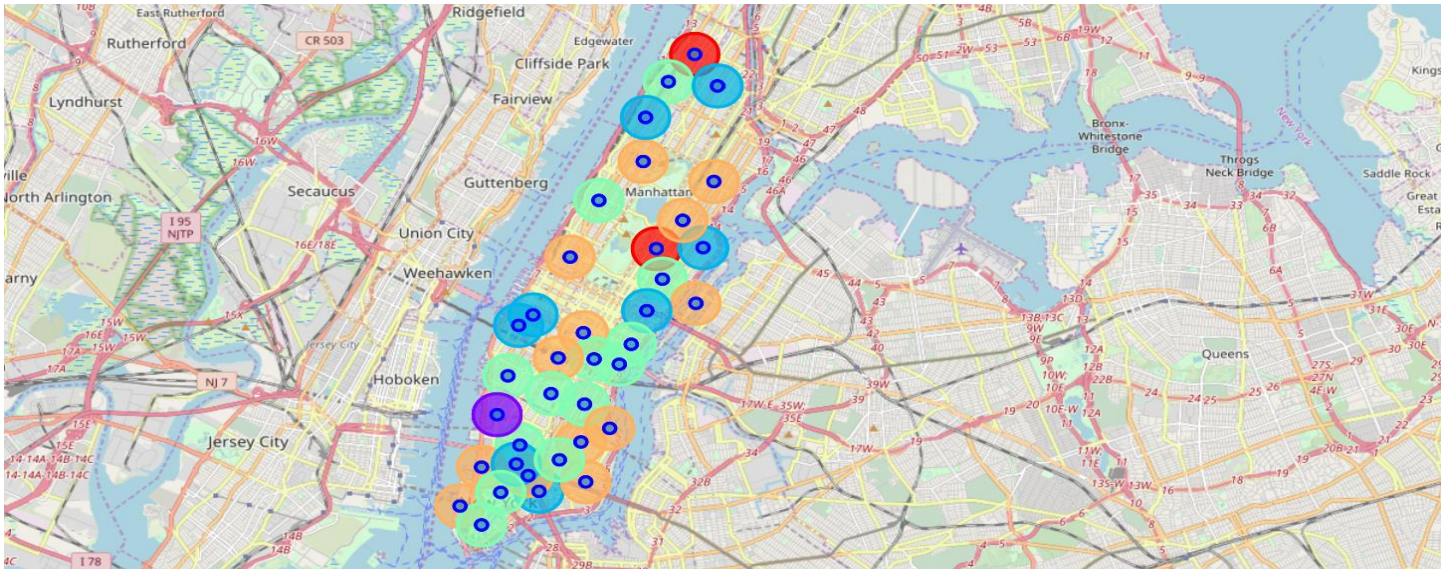

The processing of these DATA will allow to answer the key questions to make a decision

- what is the cost of rent (per square ft) around a mile radius from each subway metro station?
- what is the area of Manhattan with best rental pricing that meets criteria established?
- What is the distance from work place (Park Ave and 53 rd St) and the tentative future home?
- What are the venues of the two best places to live? How the prices compare?
- How venues distribute among Manhattan neighborhoods and around metro stations?
- Any other interesting statistical data findings of the real estate and overall data.
- Are there any Indian Restaurants Close by to the place

Methodology

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, eating options nearby and similar venues to Vancouver. This visual approach and maps with popups labels allow quick identification of location, price and feature, thus making the selection very easy


Clustering map

Map showing all the subways and rental spots
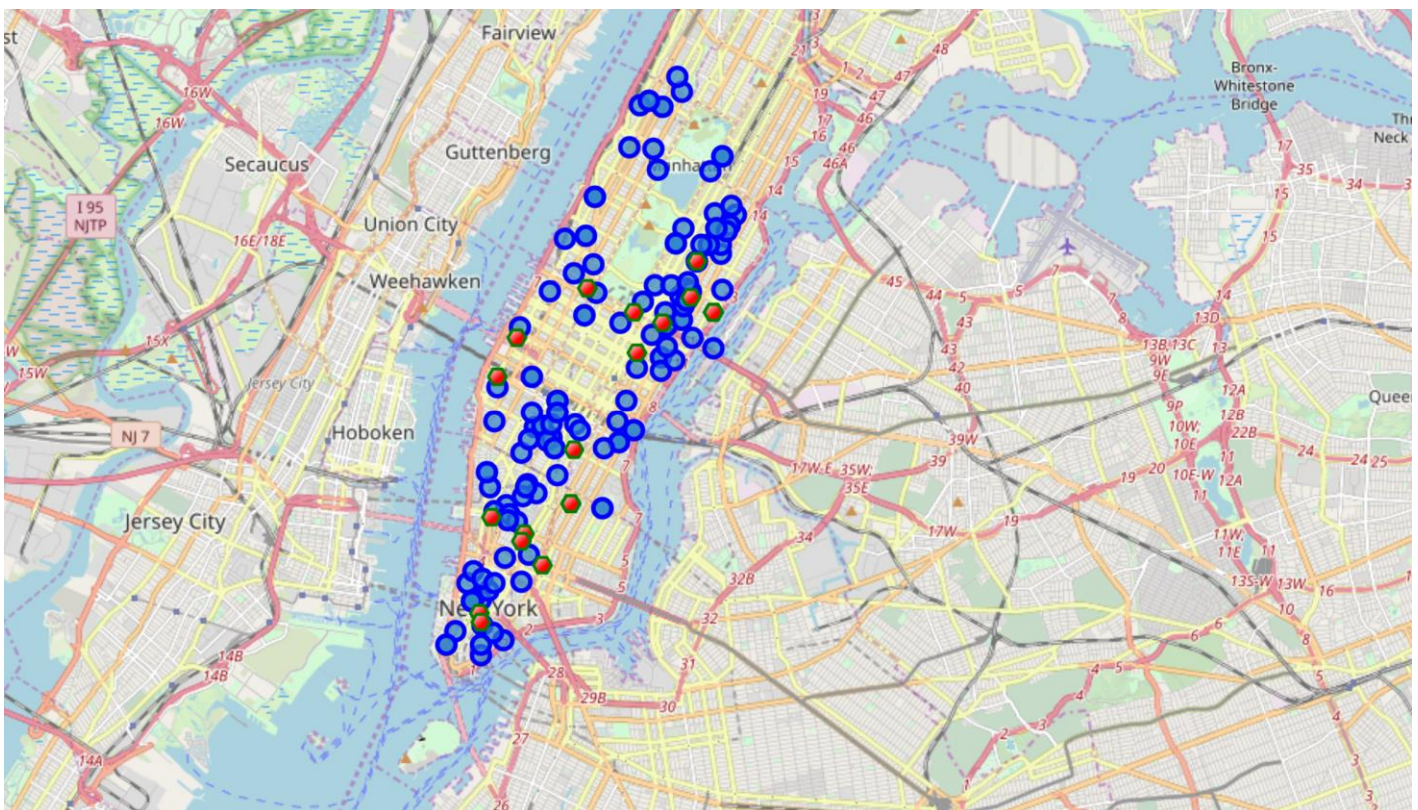
The processing of these DATA and its mapping will allow to answer the key questions to make a decision:

- what is the cost of available rental places that meet the demands?
- what is the cost of rent around a mile radius from each subway metro station?
- what is the area of Manhattan with best rental pricing that meets criteria established?
- What is the distance from work place (Park Ave and 53 rd St) and the tentative future rental home?
- What are the venues of the two best places to live? How the prices compare?
- How venues distribute among Manhattan neighborhoods and around metro stations?
- Are there tradeoffs between size and price and location?
- Any other interesting statistical data findings of the real estate and overall data.
- What are the eating options nearby
- how close is it form the place
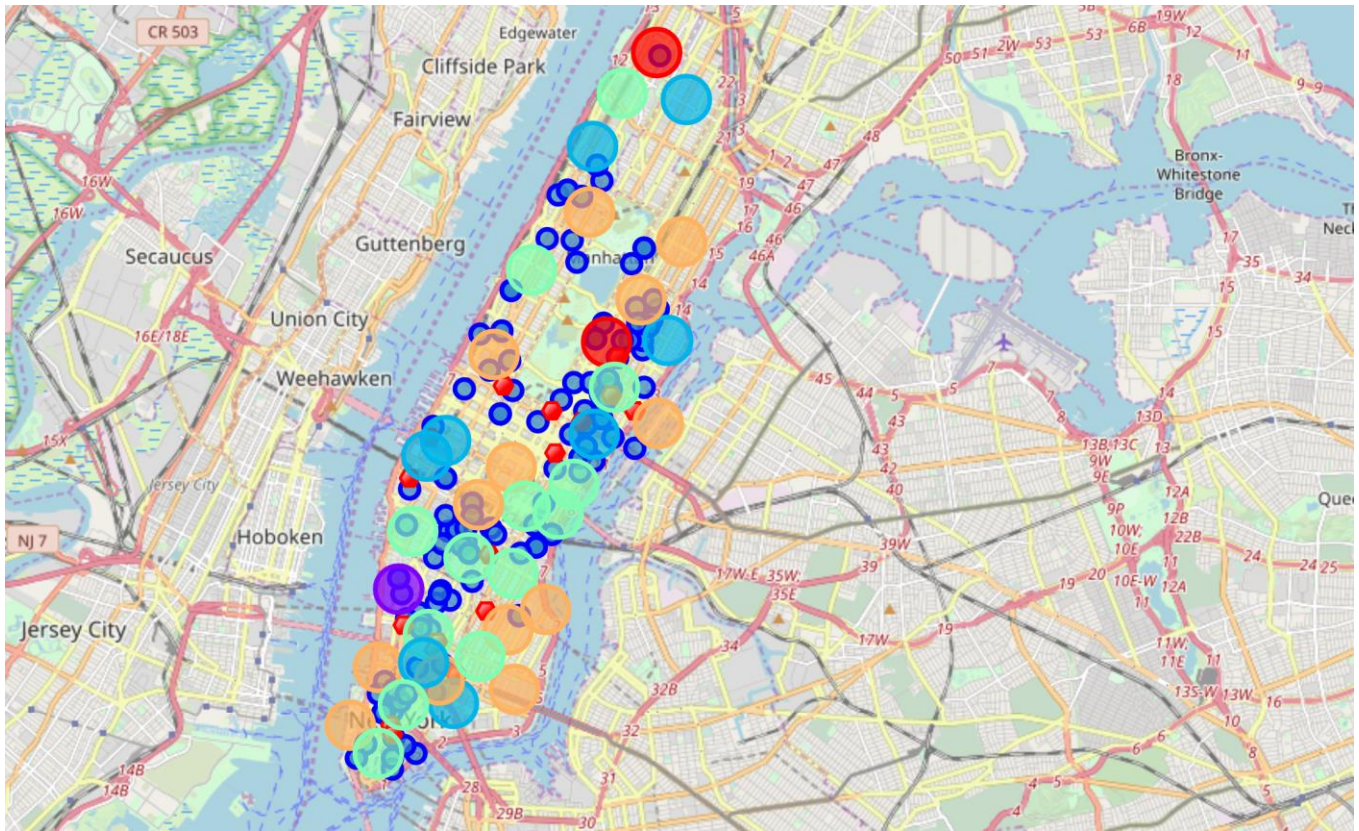
This Map includes Clustered data as well from clusters #2 and #3

Results

After examining, I have chosen two locations that meet the requirements which will assess to make a choice.

1. Apartment 1: 305 East 63rd Street in the Sutton Place Neighborhood and near 'subway 59th Street' station, Cluster # 2 Monthly rent: 7500 Dollars with a Indian restaurant nearby with good food options

2. Apartment 2: 19 Dutch Street in the Financial District Neighborhood and near 'Fulton Street Subway' station, Cluster # 3 Monthly rent: 6935 Dollars with a decent Italian restaurant.
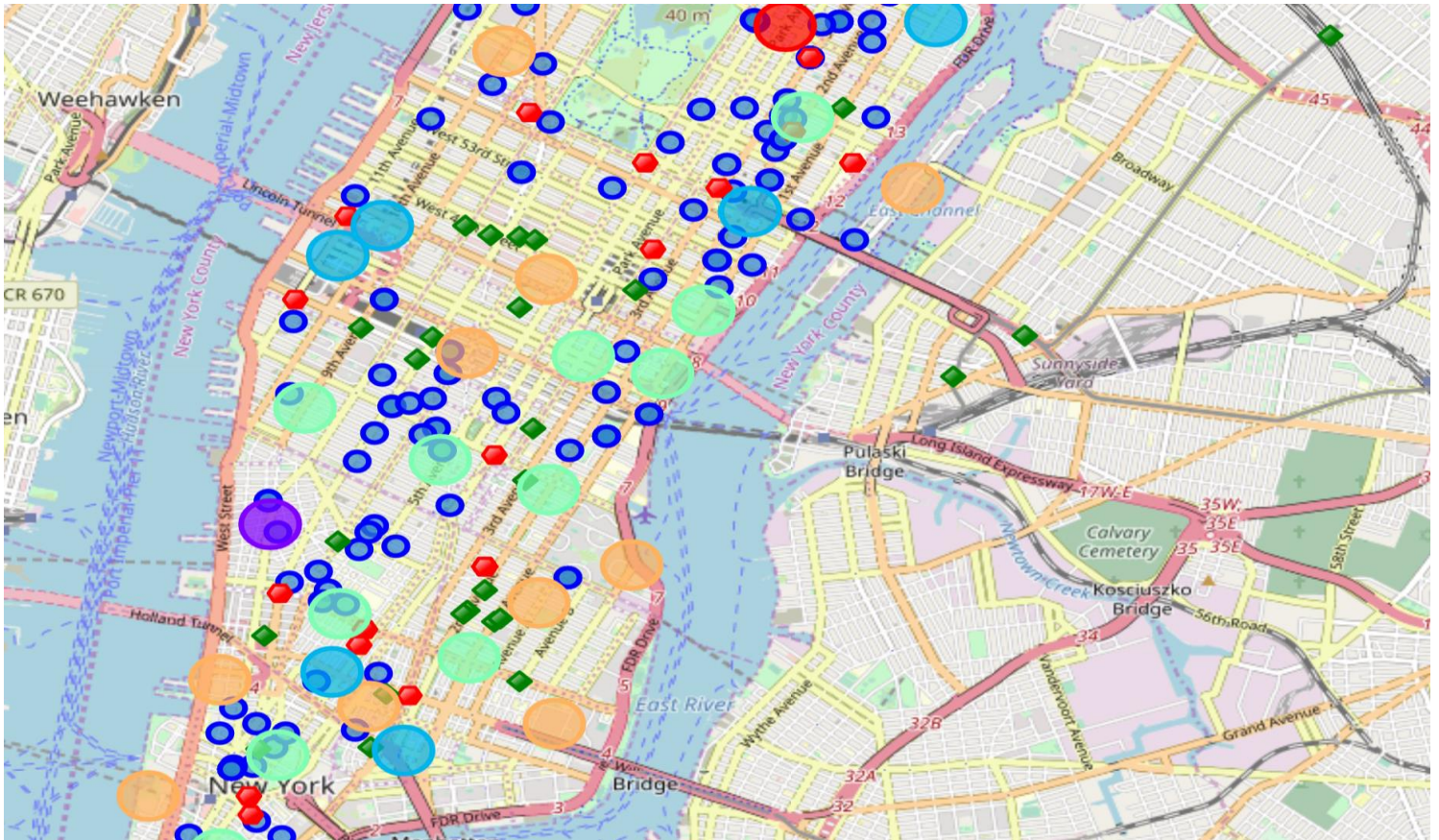
Final Map with all features analysed

Discussion

After Examining the Data I have two perfect options which are perfect for the person shifting. It maybe noted that a lot of external factors like weather were not considered during this report. We also did not consider traffic and other disabilities of a person.

Conclusion

For a person who is planning to shift from any place to Manhattan I believe that this report will be sufficient to understand the city. I have compared rent, food options, distance from subway using clustering and found that in cluster #3 we had the best place to stay found.

Thank you for taking the time to read, review and comment. I welcome all comments and suggestions. You can reach me at: email:gbhatia880@gmail.com¶