Capstone Project - The Battle of Neighbourhoods (Final Report)

6th March 2020

What to be accomplished in the final part of project

1) Introduction where you discuss the business problem and who would be interested in this project.

2) Data where you describe the data that will be used to solve the problem and the source of the data.

3) Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.

4) Results section where you discuss the results.

5) Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

6) Conclusion section where you conclude the report.

---

1. Introduction:

My Use Case: I am working @ XYZ Company and residing in Schaumburg, Chicago area. My location is walking distance from Woodfield mall and very close to all basic required amenities. I enjoy many amenities and venues in the area, such as various international and Indian restaurants, cafes, food shops and entertainment. I have been offered a great opportunity to work for a same company XYZ in New York City.

I am excited and I want to use this opportunity to trail run what I have had learned so far in Coursera in order to get answer to relevant potential questions. Question could be how can I find a convenient and enjoyable place like I have in Schaumburg, Chicago? Idea is to use and apply my learning during the course. In order to make a comparison and evaluation of the rental options in NY, I must set some basis, therefore the apartment in NY must meet the following demands:

- Apartment must be 1 bedrooms
- Location is near a metro station in the NY suburb and within 1.0 mile
- Price of rent not exceed $XXXXX per month
- Amenities in the neighborhood should be similar to current residence
- Venues such as coffee shops, Indian restaurants, gym and Asian food shops
- I have included a map of venues near current residence in Schaumburg, Chicago.

Problem Statement:

The challenge is to find a suitable apartment/studio for rent in or around NY city that complies with the demands on location, price and venues. The data required to resolve this challenge is described in the following section

2. Data Section: Description of Data

The following data is required to answer the issues of the problem: • List of Boroughs and neighborhoods of NY with their GeoData (latitude and longitude)

• List of Metro stations/Bus Station in and around NY with their address location
• List of apartments for rent in surrounding area with their addresses and price
• List of apartment for rent with additional information, such as price, address, area
• Venues for each neighbourhood

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 neighborhoods and clustered in groups
• 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
• Create a map that depicts the average rental price around a radius of 1.0 mile around each subway station
• Addresses from rental locations will be converted to GeoData (latitude and longitude) 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 apartment rent around a mile radius from each subway metro station?
• What is the area of suburb with best rental pricing that meets criteria established?
• What is the distance from work place and the tentative future home?
• What are the venues of the two best places to live?
• How venues distribute among neighborhoods and around metro stations?
• Any other interesting statistical data findings of the real estate and overall data.

Few data reference links which are leverage for demonstration.
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

A list of places for rent was collected by web-browsing real estate companies in NYC
http://www.rentmanhattan.com/index.cfm?page=search&state=results https://www.nestpick.com/search?city=new-york&page=1&order=relevance&district=manhattan&gclid=CjwKCAiAjNjgBRAgEiwAGLlf2hkP3A-cPxjZYkURqQEswQK2jKQEpv_MvKcrIhRWRzNkc_r-fGi0lxoCA7cQAvD_BwE&type=apartment&display=listhttps://www.realtor.com/apartments/Manhattan_NY

- Latitude and Longitude data
- Venues
- Flat/Condo Prices
- Rental Prices
- Subway/Metro Info

3. Methodology section:

This section will represents the main component of the report how the data is gathered, prepared for analysis. .

The strategy is based on mapping described data description section in order to facilitate the choice. The choice is made based on the location near a subway, rental price and similar venues to Schaumburg chicago. This visual approach and maps with popups labels allow quick identification of location, price and feature, thus making the selection very easy.
Processing of these DATA and its mapping will allow to answer the key questions to make a decision:

• Cost of available rental places that meet the demands?
• Cost of rent around a mile radius from each subway metro station?
• Area of NYC with best rental pricing that meets criteria established?
• What are the venues of the two best places to live?
• How venues distribute among NYC neighborhoods and around metro stations?
• Any other interesting statistical data findings of the real estate and overall data.

4. Result: Selection of Apartment

Using the map view it was easy to explore all possible option and make a good decision.

First option is very close all required amenities and transport facilities as well. There is not much difference between cluster 2 and 3 outcomes. However based on Schaumburg Chicago venues, Cluster 2 type of venues is a closer resemblance to my current place and desire place.

5. Discussion

Though the learning was in very fast forward mode and I can understand the usage of lab work and doing actual implementation will differ specifically in my current DWH and Data management project but I am pretty much satisfied the the was course is structured.

I am planning to go thru few lab exercises again that will help me to have more grip popular libraries.

Capstone project is a very good opportunity to practice and apply the Data Science tools and methodologies at high level. All these module boost the confidence to grab exciting opportunity in the area of data science.

6. Conclusion

I got what I wanted to learn and familiar myself with all function and libraries. Spending time and money is worth and I can and will recommend mu colleague to go thru this course and enhance the confidence.

This project has shown a practical application to resolve a real most encountered situation using Data Science tools.¶

All function and libraries is very great and return info with very small code.

*** End of Project and last module Happy Learning 06 March 2020 ***

***** *Few data pulled for decision making depicted in report below* *****

Map of Schaumburg Chicago and NYC with venues near residence place - for reference

	name	categories	lat	lng
0	Maxfield's Pancake House	Breakfast Spot	42.029146	-88.061800
1	Schaumburg Prairie Center for the Arts	Performing Arts Venue	42.025993	-88.067020
2	Bawarchi Biryani Point	Indian Restaurant	42.027498	-88.059736
3	Volkening Heritage Farm	Farm	42.023838	-88.061172
4	Patel Brothers	Grocery Store	42.027654	-88.059611
5	Baskin-Robbins	Ice Cream Shop	42.027741	-88.059312
6	The UPS Store	Shipping Store	42.027485	-88.059264
7	Redbox	Video Store	42.028480	-88.061477
8	Dunkin'	Donut Shop	42.027776	-88.059315
9	Walgreens	Pharmacy	42.028743	-88.061441
10	Clay Oven	Indian Restaurant	42.027409	-88.059532
11	Schaumburg Liquors	Convenience Store	42.027790	-88.060036
12	Schaumburg and Plum Grove Intersection	Intersection	42.028094	-88.060465
13	Septemberfest Craft Fair	Art Gallery	42.025576	-88.068764

```
manhattan_data.head()
```

Out[49]:

	Sr No	Borough	Neighborhood	Latitude	Longitude	Cluster Labels
0	1	Manhattan	Marble Hill	40.876551	-73.910660	2
1	2	Manhattan	Chinatown	40.715618	-73.994279	2
2	3	Manhattan	Washington Heights	40.851903	-73.936900	4
3	4	Manhattan	Inwood	40.867684	-73.921210	3
4	5	Manhattan	Hamilton Heights	40.823604	-73.949688	0

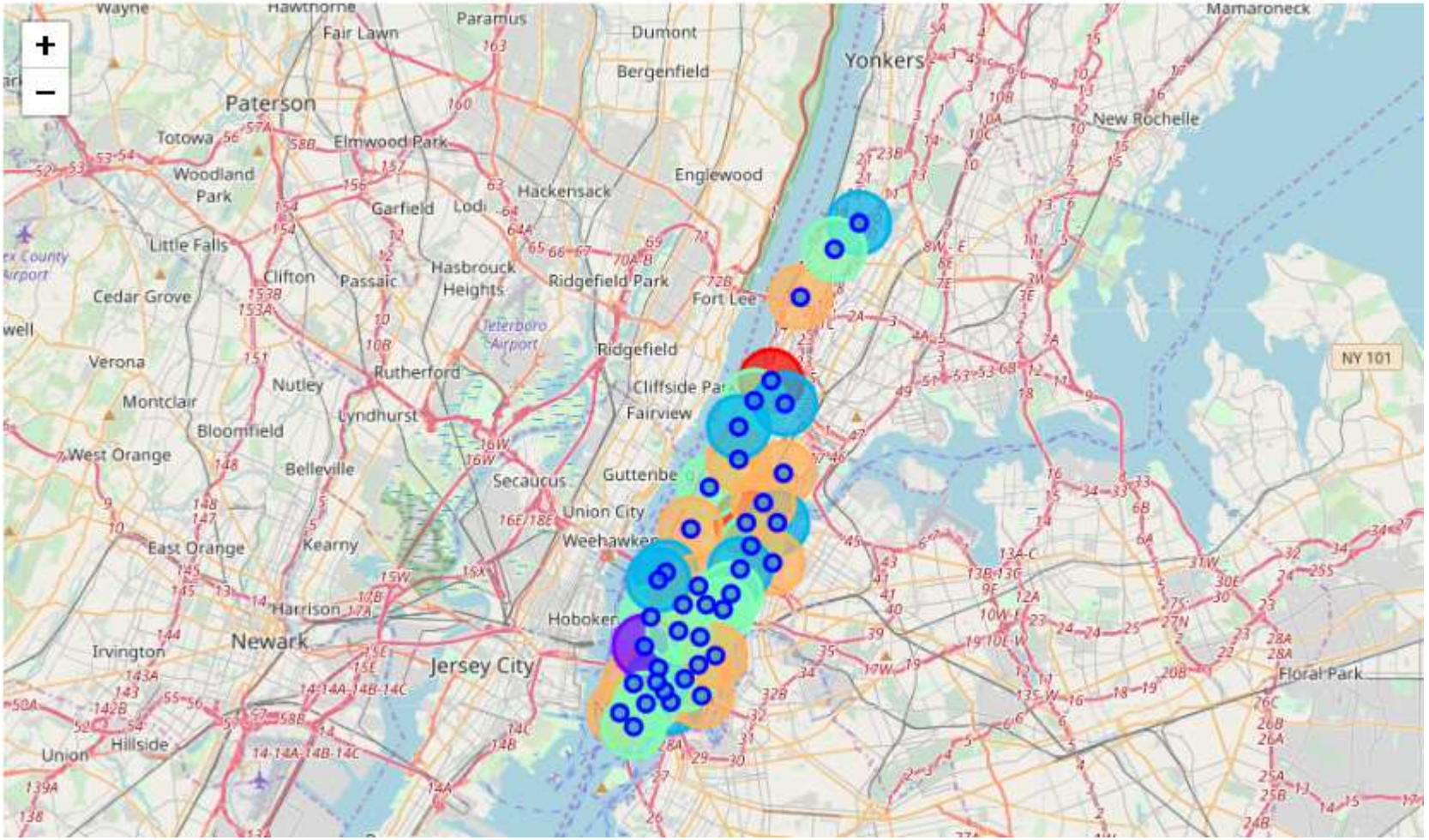
In [50]:
```
manhattan_data.tail()
```

Out[50]:

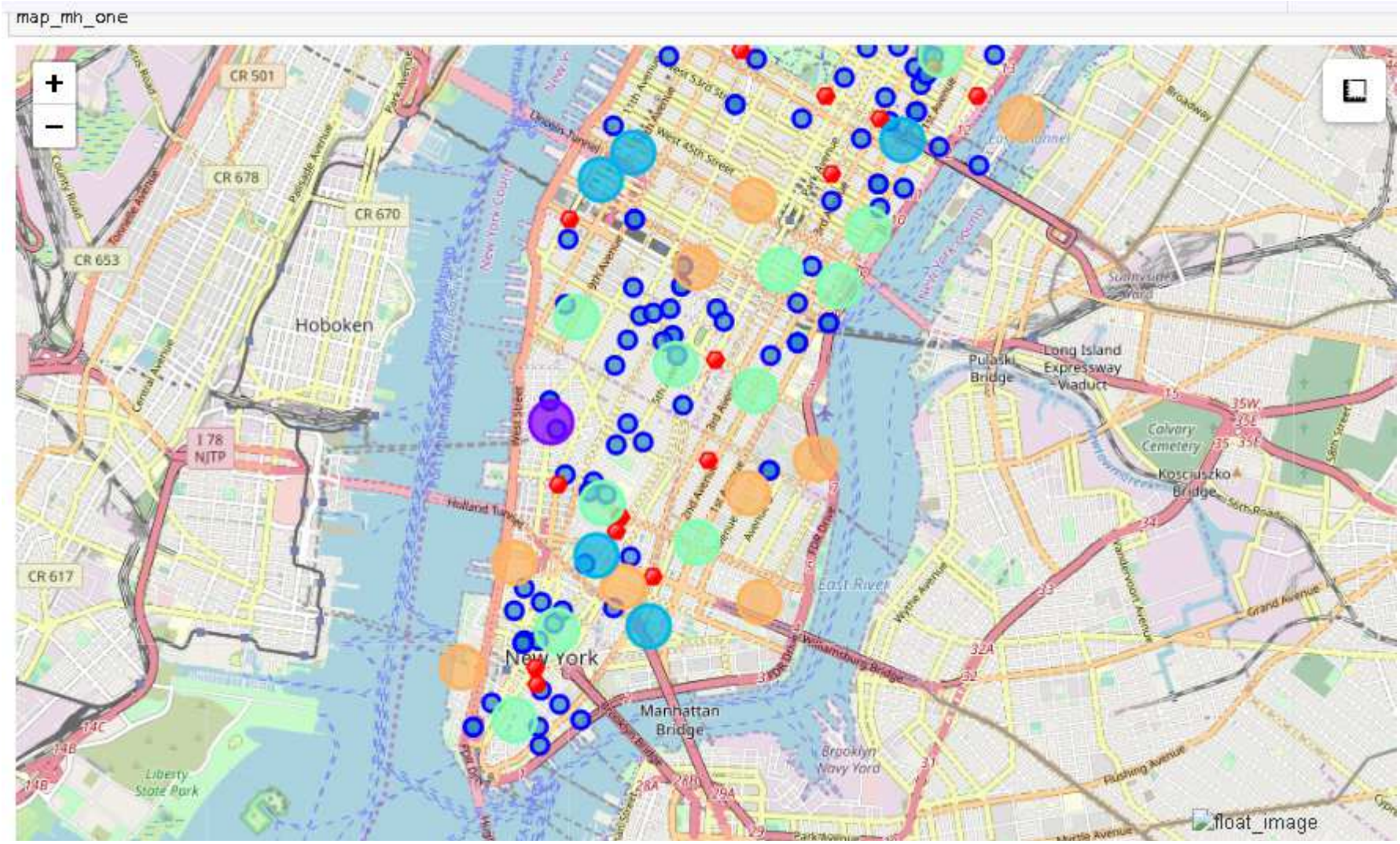
	Sr No	Borough	Neighborhood	Latitude	Longitude	Cluster Labels
35	36	Manhattan	Turtle Bay	40.752042	-73.967708	3
36	37	Manhattan	Tudor City	40.746917	-73.971219	3
37	38	Manhattan	Stuyvesant Town	40.731000	-73.974052	4
38	39	Manhattan	Flatiron	40.739673	-73.990947	3
39	40	Manhattan	Hudson Yards	40.756658	-74.000111	2

```
map_clusters
```

:



Conventions: Red dots are Subway stations, Blue dots are apartments available for rent, Bubbles are the clusters of venues



**** Rent pattern at destination location. ****

Out[73]: <matplotlib.axes._subplots.AxesSubplot at 0x7f3fee6eef28>