Coursera Capstone project
Coursera IBM Data Science Certification

NAME: Abdul Rehman

Report contents:

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 learning 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

I am currently living in Pakistan, within walking distance to a big market. I also enjoy great venues and attractions, such as international Cuisine, entertainment and shopping. I have an offer to move to work in Manhattan NY and I would like to move if I can find a place to live similar to 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 the US, Europe or Asia. Europe, US or Asia, Likewise, it can be a 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 that have 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 geodat via Nominatim for mapping with Folium.

- List of Subway stations was obtained via Wikipedia, NY Transit website 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:

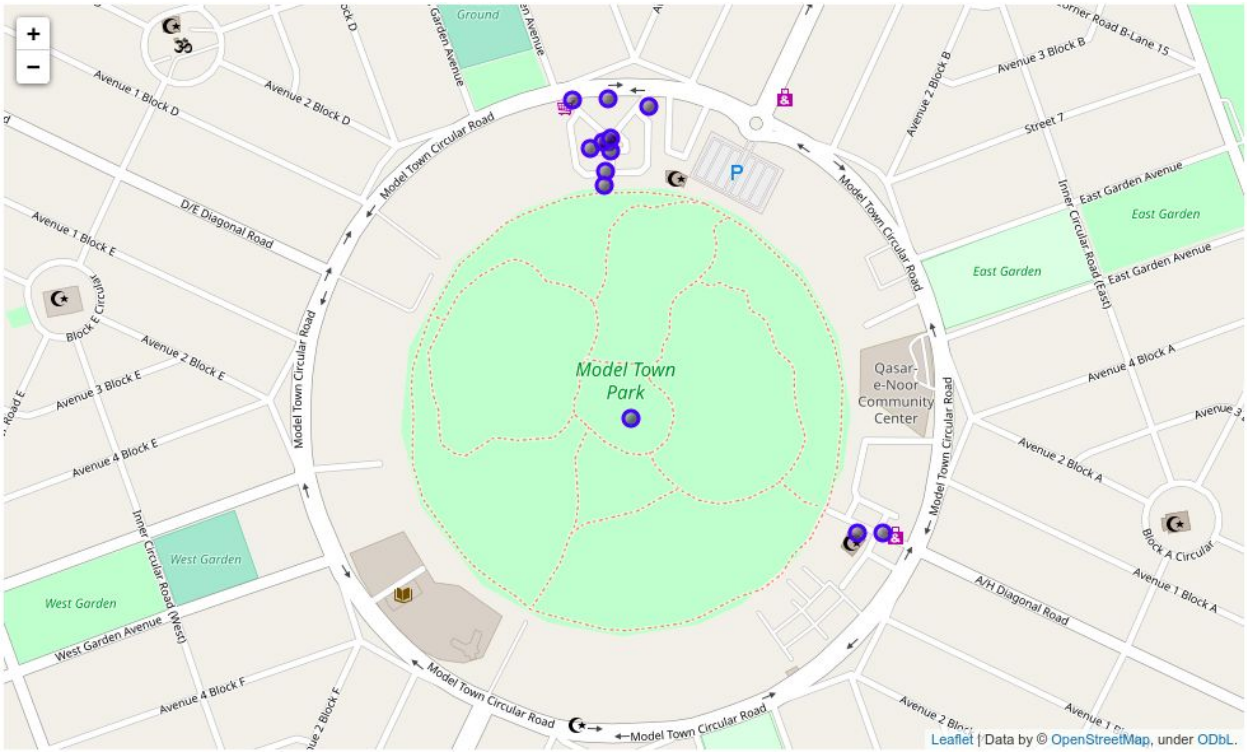
The strategy is based on mapping the described data in section 2.0, in order to facilitate the choice of at least two candidate places for rent. The information will be consolidated in ONE MAP where one can see the details of the apartment, the cluster of venues in the neighborhood and the relative location from a subway station and from my work place. A measurement tool icon will also be provided. The pop ups on the map items will display rent price, location and cluster of venues applicable.

The Tools:

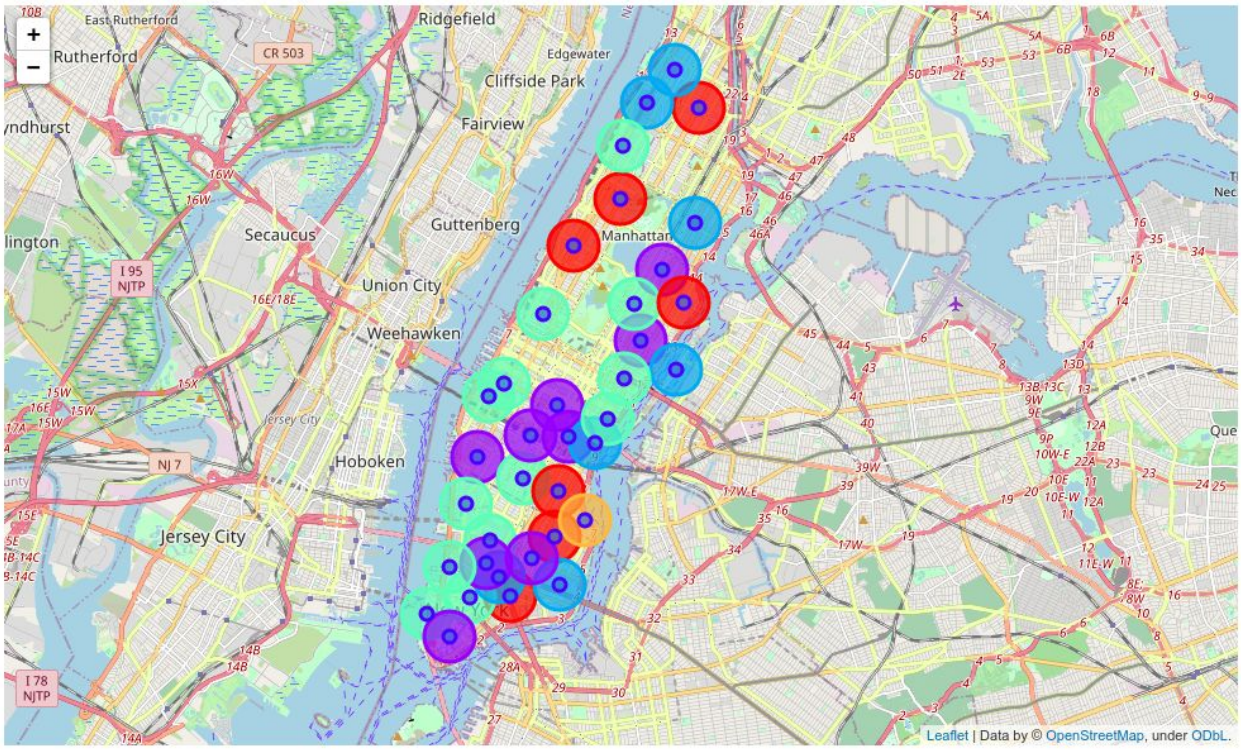
Web-scraping of sites is used to consolidate data-frame information which was saved as csv files for convenience and to simply the report. Geodata was obtained by coding a program to use Nominatim to get latitude and longitude of subway stations and also for each of (144 units) the apartments for rent listed.
Geopy_distance and Nominatim were used to establish relative distances. Seaborn graphic was used for general statistics on rental data.
Maps with pop ups labels allow quick identification of location, price and feature, thus making the selection very easy

4. Execution and Results

Venues around Neighborhood in Pakistan



Manhattan apartment for rent with venue clusters



After examining several cluster data , I concluded that cluster# 4 resembles closer Model Town, therefore providing guidance as to where to look for the future apartment

Venues of cluster 4

Cluster 4

```
In [86]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 3, manhattan_merged.columns[[1] + list(range(5, manhatta
         n_merged.shape[1]))]]
```

Out[86]:

	1st 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	Cluster Labels	Latitude	Longitude
0	Park	Boat or Ferry	Memorial Site	Food Court	Shopping Mall	Playground	BBQ Joint	3	40.711932	-74.016869
5	Coffee Shop	Spa	Italian Restaurant	Yoga Studio	Café	French Restaurant	Park	3	40.715229	-74.005415
6	Italian Restaurant	Wine Shop	American Restaurant	Hotel	Sandwich Place	Cocktail Bar	Gym	3	40.759101	-73.996119
10	Gym / Fitness Center	Furniture / Home Store	Vegetarian / Vegan Restaurant	Spa	Mediterranean Restaurant	Café	Salon / Barbershop	3	40.739673	-73.990947
12	Italian Restaurant	Ice Cream Shop	American Restaurant	Sandwich Place	Cosmetics Shop	Bubble Tea Shop	Dessert Shop	3	40.726933	-73.999914
14	American Restaurant	Café	Coffee Shop	Dog Run	Gym	Park	Burger Joint	3	40.756658	-74.000111
17	Plaza	Performing Arts Venue	Theater	Wine Shop	American Restaurant	Italian Restaurant	Coffee Shop	3	40.773529	-73.985338
25	Park	Deli / Bodega	Sandwich Place	Burger Joint	Donut Shop	Grocery Store	Greek Restaurant	3	40.808000	-73.963896
31	Italian Restaurant	Coffee Shop	Gym	Pizza Place	Spa	Beer Garden	Beer Bar	3	40.760280	-73.963556
32	American Restaurant	Wine Bar	Spa	Coffee Shop	Skate Park	Hotel	Men's Store	3	40.721522	-74.010683
34	Coffee Shop	Wine Bar	Seafood Restaurant	Japanese Restaurant	Café	Deli / Bodega	French Restaurant	3	40.752042	-73.967708
35	Italian Restaurant	Yoga Studio	Juice Bar	French Restaurant	Exhibit	Spa	Pizza Place	3	40.775639	-73.960508
38	Italian Restaurant	Cocktail Bar	Wine Bar	Jazz Club	Theater	French Restaurant	Coffee Shop	3	40.734434	-74.006180

After examining, I have chosen to look for an apartment in cluster 4 because it resembles very closely to my current location.

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 create 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 it for use in similar situations.

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