Coursera Capstone - REPORT

Contents

1. Introduction of usecase

1.1 Scenario and Background

1.2 Problem to be resolved

1.3 Interested Audience

2. Data used in usecase

2.1 Data of Current Location (current residence place)

2.2 Data required to resolve the problem

2.3 Data sources and data manipulation

2.4 How the data will be used to solve the problem

2.5 Mapping of Data

3. Methodology

3.1 The analysis and the strategy:

3.2 Data Science Methods, machine learing, mapping tools and exploratory data analysis.

3.3 Map of Mangalore with venues near residence place - for reference

3.4 Manhattan neighborhood

4. Results Section

4.1 Map of Manhattan neighborhoods with top 10 clustered venues.

5. Discussion Section

6. Conclusion Section

1. Introduction of usecase :

A description of the problem and a discussion of the background.

1.1 Scenario and Background

I currently live within walking distance to City Centre Mall Mangalore and I enjoy many places in the area, such as various famous cuisine restaurants, cafes, food shops and entertainment.

I have been offered a great opportunity to work for a leader firm in Manhattan, NY. I am very excited and I want to use this opportunity to practice my learnings in Coursera in order to answer relevant questions arisen. The key question is : How can I find a convenient and enjoyable place similar to mine now in Mangalore?

Certainly, I can use available real estate apps and Google but the idea is to use and apply myself the learned tools during the course. In order to make a comparison and evaluation of options in Manhattan NY, I must set some basis, therefore the apartment in Manhattan must meet the following demands:

desired location is near a mall in the Manhattan area and within 1.0 mile (1.6 km) radius top ammenities in the selected neighborhood shall be similar to current residence desirable to have venues such as hotel, clothing Store, theater, bookstore and bakery. As a reference, I have included a map of venues near current residence in Mangalore.

1.2 Problem to be resolved:

The challenge is to find a suitable apartment for rent in Manhattan NY that complies with the demands on location and venues. The data required to resolve this challenge is described in the following section 2, below.

1.3 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 used in usecase :

A description of the data and how it will be used to solve the problem

2.1 Data of Current Location

I Currently reside in the neighborhood of Ocean Pearl, Mangalore, India. I use Foursquare to identify the venues around the area of residence which are then shown in the Mangalore 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

2.2 Data Required to resolve the problem

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

List of neighborhoods of Manhattan with their geodata (latitud and longitud)
List of hotel, clothing Store, theater, bookstore and bakery in Manhattan with their address location
Venues for each Manhattan neighborhood (than can be clustered)

2.3 Data Source

Manhattan neighborhoods csv formed during the course.
A csv file was created which will be read in order to create a dataframe and its mapping.
The csv file 'manhattan_neghbour.csv' will be directly read to the Jupiter Notebook for convenience.
The clustering of neighborhoods and mapping will be 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 neighborhoods and clustered in groups (as per Course LAB)

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 Neighborhoods
Manhattan map of clustered venues and neighborhoods

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.

3. 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.

3.1 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 a candidate places for accomodation. The choice is made based on the demands imposed : similar venues to Mangalore, India. This visual approach and maps with popups labels allow quick identification of location, thus making the selection very easy.

The procesing of these DATA and its mapping will allow to answer the key questions to make a decision:

What are the venues of the best place to live? How venues distribute among Manhattan neighborhoods ?

3.2 Methodology followed - Data Science Methods, machine learing, mapping tools and exploratory data analysis

Mangalore Map - Current residence and venues in neighborhood for comparison to future Manhattan place

3.3 Map of Mangalore with venues near residence place - for reference

In [75]:
```python
map_nd = folium.Map(location=[latitude, longitude], zoom_start=14)

for lat, lng, label in zip(NDnearby_venues['lat'], NDnearby_venues['lng'], NDn
earby_venues['name']):
    label = folium.Popup(label, parse_html=True)
    folium.RegularPolygonMarker(
        [lat, lng],
        number_of_sides=4,
        radius=10,
        popup=label,
        color='blue',
        fill_color='#0f0f0f',
        fill_opacity=0.7,
    ).add_to(map_nd)

map_nd
```

Out[75]:
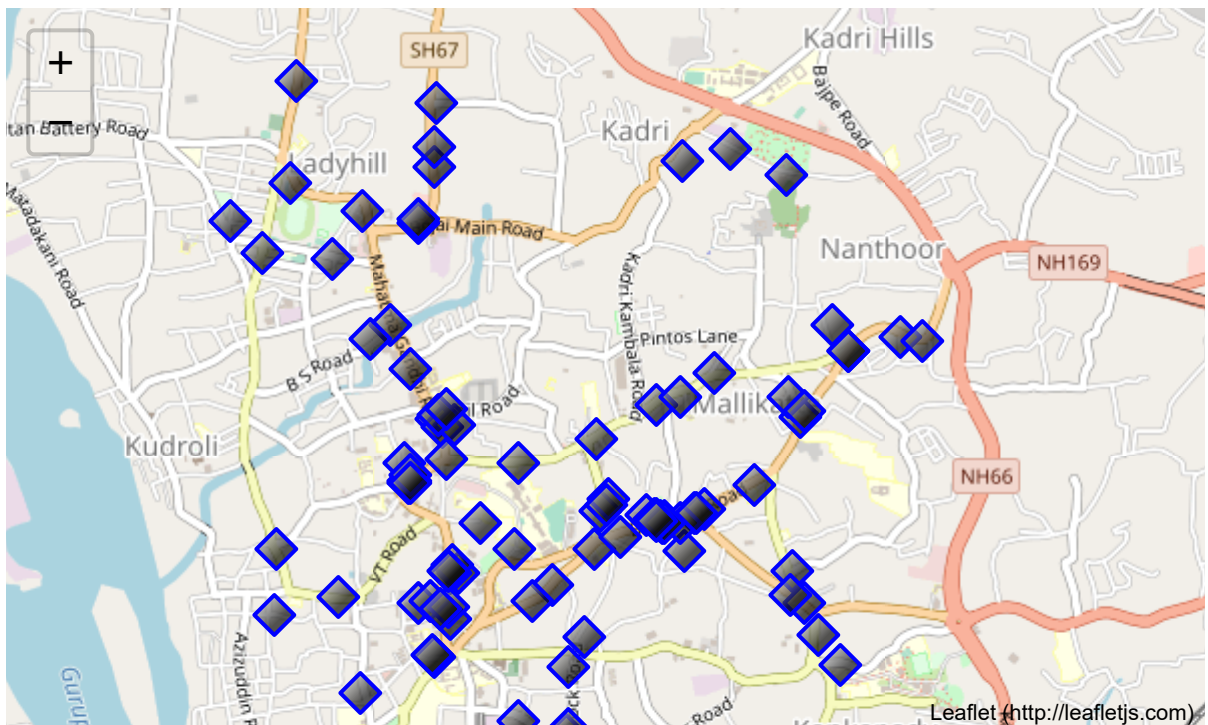
This was all about the data and how it will be used to solve the problem.

3.4 Manhattan neighborhood

Cluster neighborhood data was produced with Foursquare during course lab work.
A csv file was produced containing the neighborhoods around the 40 Boroughs.
Now, the csv file is just read for convenience and consolidation of report.

```
In [77]:  final_df=pd.read_csv('manhattan_neghbour.csv')
          final_df.head()
```

Out[77]:

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688

Grouped rows by neighborhood and by taking the mean of the frequency of occurrence of each category

Group all the rows based on the neighborhood for better understanding of the actual values

Created a new dataframe and displaying the top 10 venues for each Neighborhood

Perform Clustering on the obtained dataframe

4. Results

4.1 Map of Manhattan neighborhoods with top 10 clustered venues.

Let's visualize the resulting clusters

In [87]:
```python
# create map of Manhattan using latitude and longitude values from Nominatim
latitude= 40.7308619
longitude= -73.9871558

map_mh_one = folium.Map(location=[latitude, longitude], zoom_start=13.3)


# set color scheme for the clusters
kclusters=5
x = np.arange(kclusters)
ys = [i+x+(i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(mh_merged['Latitude'], mh_merged['Longitude'], mh_merged['Neighborhood'], mh_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=15,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_mh_one)

    # Adds tool to the top right
from folium.plugins import MeasureControl
map_mh_one.add_child(MeasureControl())

# Measurement ruler icon tool to measure distances in map
from folium.plugins import FloatImage
url = ('https://media.licdn.com/mpr/mpr/shrinknp_100_100/AAEAAQAAAAAAAlgAAAAJ
GE3OTA4YTdlLTkzZjUtNDFjYy1iZThlLWQ5OTNkYzlhNzM4OQ.jpg')
FloatImage(url, bottom=5, left=85).add_to(map_mh_one)

map_mh_one
```
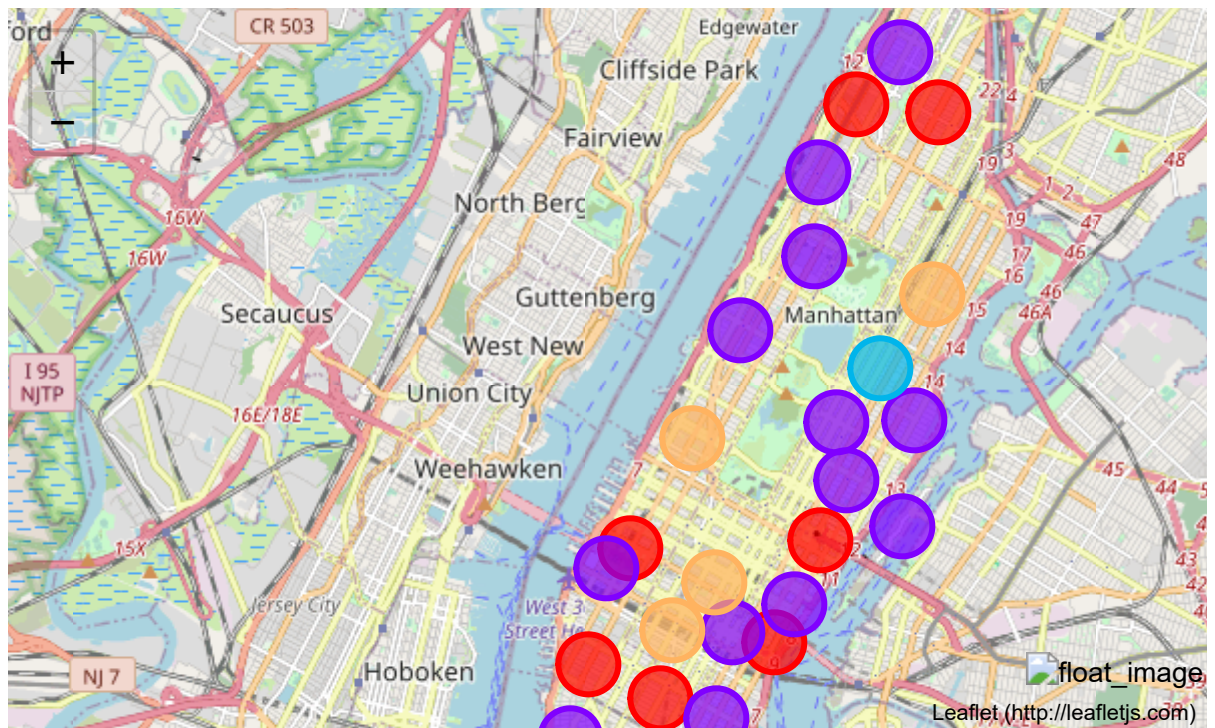
Out[87]:

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

Midtown Neighborhood, Cluster 4: Midtown District is having Hotel, Clothing Store, Theater, Bookstore, Bakery similar to Mangalore residence.

In [88]:
```python
## cluster_no is the cluster number to explore
cluster_no = 4
mh_merged.loc[mh_merged['Cluster Labels'] == cluster_no, mh_merged.columns[[1]
+ list(range(5, mh_merged.shape[1]))]]
```

Out[88]:

	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 Mo Commo Venu
37	Stuyvesant Town	Bar	Boat or Ferry	Park	Coffee Shop	Baseball Field	Gym / Fitness Center	Skatir Rir
21	Tribeca	Park	Italian Restaurant	American Restaurant	Spa	Greek Restaurant	Wine Bar	Ca
7	East Harlem	Mexican Restaurant	Bakery	Thai Restaurant	Spa	Sandwich Place	Deli / Bodega	Lat America Restaura
13	Lincoln Square	Café	Plaza	Gym / Fitness Center	Theater	Performing Arts Venue	Italian Restaurant	Conce H
15	Midtown	Hotel	Clothing Store	Sporting Goods Shop	Coffee Shop	Theater	Bookstore	Bake
33	Midtown South	Korean Restaurant	Hotel	Japanese Restaurant	American Restaurant	Cosmetics Shop	Hotel Bar	Gym Fitne Cent
19	East Village	Bar	Mexican Restaurant	Pizza Place	Ice Cream Shop	Cocktail Bar	Wine Bar	Vietnames Restaura

5. Discussion

Apartment Selection

The map above helped me to explore all possibilities since the popups provide the information needed for a good decision

Selecting the apartment in a location similar to the current one in Mangalore.

6. Conclusion

Midtown District is having Hotel, Clothing Store, Theater, Bookstore, Bakery is my preferabe choice for a future residence.

Based on current residence at Mangalore, I feel that Cluster 4 type of venues is a closer resemblance to my current place. Which means that Financial District is a better choice since it has Hotel, Clothing Store, Theater, Bookstore and is worth the convenience it provides.

Thank you!!!

In [ ]: