

Battle of Neighborhoods – Report

Topic: Toronto Health Seminar - Searching Venues of interest for Delegates and Families

Introduction

Toronto has been hosting its famous health congregation Seminar since 2010. Its the time of the year when health professionals converge in hordes to participate, learn and share the latest technologies and breakthroughs in their field.

It's a Five-day event and the professionals are also accompanied by their families.

This year is no exception. However, with the COVID-19 threat the seminar has been postponed to August, considering that the threat is reduced by that time.

Lights and Action the event management company entrusted with the logistics of this event has employed me to provide them with data points and map that would assist them in arriving at probable logistics with regards to:

1. Hotels as Accommodation
2. Shopping malls and Event places
3. Parks, Restaurant, Cafeteria for family recreation

This Project Therefore would have the objective of looking at such data points, clean them, merge them into probable sets and visualize the same with maps.

Data Description

Foursquare API call would be used to complete the project in the following manner:

- First the parameters of the Foursquare account would be defined ex: Client ID etc.

```
CLIENT_ID = 'XXXXXXXX' # your Foursquare ID
CLIENT_SECRET = 'XXXXXXXX' # your Foursquare Secret code
VERSION = '20180604'
LIMIT = 30
print('Your credentials:')
print('Foursquare_ID: ' + CLIENT_ID)
print('Foursquare_Secret: ' + CLIENT_SECRET)
```

- Next step is to define the geographical area, get the coordinates of the same. In our case the same is Toronto and we would be using Nominatim function from Geopy Library to achieve the same

```
# define the city and get its latitude & longitude
city = 'Toronto'
geolocator = Nominatim(user_agent="My-application")
location = geolocator.geocode(city)
latitude = location.latitude
longitude = location.longitude
print("Geographical Coordinates of the City of {} is {},{}".format(city, latitude, longitude))
```

- Individual Search URL would have to be created specifying the search criterion as 'Hotel','Restaurant','Park' etc.. and queried to the Foursquare API to get the data.

For example below is for Hotels Search:

```
# search for hotels
search_query = 'Hotel'
radius = 500

# Define the corresponding URL
url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&v={}&query={}&radius={}&limit={}'\
.format(CLIENT_ID, CLIENT_SECRET, latitude, longitude, VERSION, search_query, radius, LIMIT)
url
```

Methodology

On query of the Foursquare API using each individual URL the get response received provides us with a JSON data file as outlined above.

Each of the individual JSON file is then converted into a dataframe using the Pandas library.

```
# assign relevant part of JSON to venues
venues = results['response']['venues']

# transform venues into a dataframe
dataframe = json_normalize(venues)
dataframe.head()
```

Once the dataframe is obtained the tasks involve cleaning and more cleaning such that

1. All null data points are ignored
2. Un-necessary columns dropped
3. Relevant results are bundled together.

Cleaning Sample 1: Deleting un-necessary Columns:

```
# delete unnecessary columns
clean_dataframe2= clean_dataframe.drop(['cc', 'city', 'country', 'crossStreet', 'distance', 'formattedAddress',\
                                         'labeledLatLngs','neighborhood', 'id'], axis=1)
clean_dataframe2
```

Cleaning Sample 2: Null Data points ignored

```
# delete rows with none values
clean_dataframe3 = clean_dataframe2.dropna(axis=0, how='any', thresh=None, subset=None, inplace=False)
clean_dataframe3
```

Cleaning Sample 3: Bundling similar category value together

```
# delete rows which its category is not Hotel or Event Space
array= ['Hotel', 'Event Space']
hotel_dataframe= clean_dataframe3.loc[clean_dataframe3['categories'].isin(array)]
hotel_dataframe
```

Next Step was to create two merged data frame with

1. Point of interest of Delegates – Merged Data frame consisting of Hotel, Shopping centre and Cafeteria clusters
2. Point of interest of Families of Delegates – Merged Data frame consisting of Park, Restaurant and Cafeteria cluster

```
# create dataframe of Park, Restaurant and Cafeteria
park_neighbourhood_df = pd.concat([df_park, df_Restaurant, df_Cafeteria,], ignore_index=True)
park_neighbourhood_df
```

Final Step was to create two visualization of this data frame and project the venue points as being reflected in the data frame on the map of Toronto

```
# Generate map to visualize park + Restaurant + Cafeteria
park_map = folium.Map(location=[latitude, longitude], zoom_start=14)

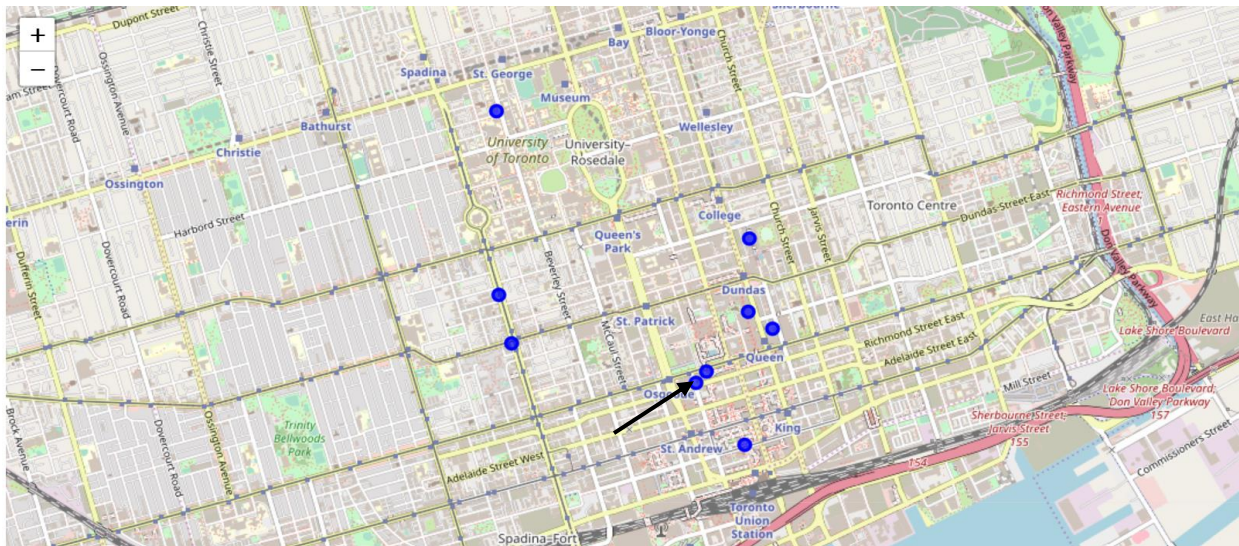
for lat, lng, name, categories, address in zip(park_neighbourhood_df['lat'], park_neighbourhood_df['lng'], park_neighbourhood_df['name'], park_neighbourhood_df['categories'],\park_neighbourhood_df['address']):
    label = '{} , {}'.format(name, address)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
```

```
fill=True,  
fill_color='blue',  
fill_opacity=0.7,  
parse_html=False).add_to(park_map)
```

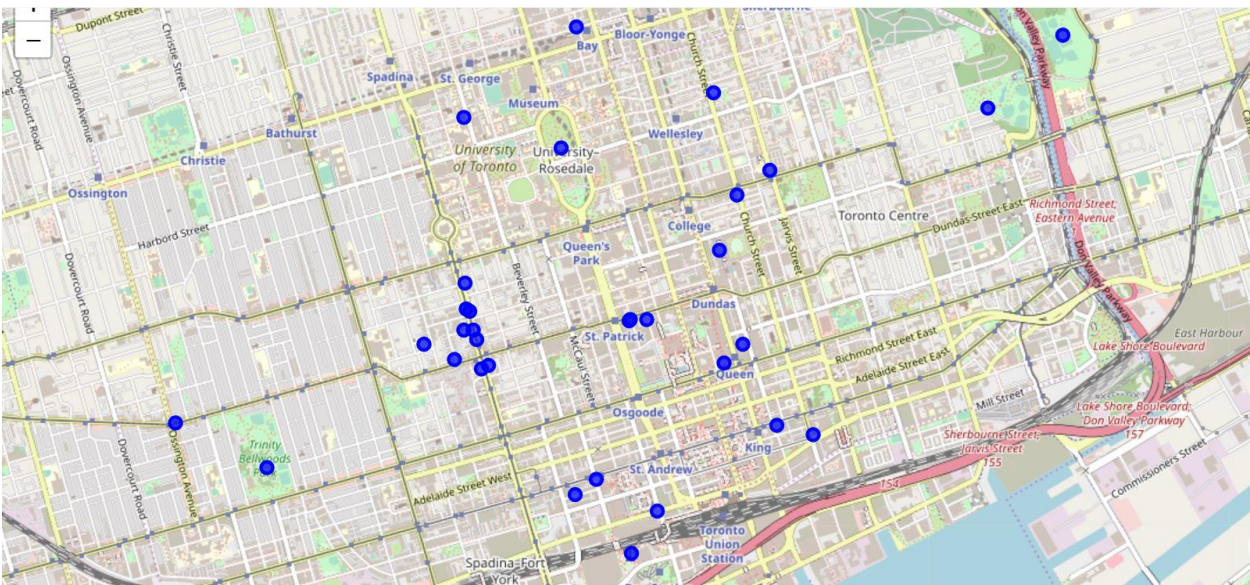
park_map

Results

Merged Dataframe 1 – Hotels, Shopping Store & Cafeteria



Merged Dataframe 2 – Parks, Restaurant and Cafeteria



The First Dataframe provides the ball point markers from the logistic point of view of the Delegates which includes Accommodation, place to hold the seminar, Point of shopping for daily essentials and souvenirs.

With regard to the place of accommodation, it would be desirable if the Seminar place is around the same location. Hence one of the prime locations if looked in conjunction with the analysis done at code stage is Hotel Sheraton which is in close proximity to the Grand Ballroom. This was arrived by dint of both having the same postal code.

```
In [13]: # choose the hotel which has the same postalCode with the event space
df_hotel = df_hotels[df_hotels.postalCode == 'M5H 2M9']
df_hotel
```

Out[13]:

	name	categories	address	lat	lng	postalCode	state
0	Sheraton Centre Toronto Hotel	Hotel	123 Queen Street West	43.650594	-79.384530	M5H 2M9	ON
19	Grand Ballroom	Event Space	123 Queen St. W	43.651217	-79.383771	M5H 2M9	ON

Considering the Delegates stay with their family in the aforementioned location(Marked with a black arrow in the first visualization) which is in Queens street, there are lots of recreational venues two or three streets away on both sides of the location as seen in the second visualization. There should therefore be ample choice for the families to select from to while away their time when the delegates are away on Seminar.

Discussion

The following points needs to be noted here regarding this project execution:

1. The Project Scope of a Toronto Health seminar is an Hypothetical problem created
2. This Project forms the final assignment (Capstone) for the IBM specialization in Data Science being done through Coursera
3. There is further scope of creating individual maps for each venue with labelling, However labelling was avoided to reduce the clutter and the intention was to provide for delegates and families and hence two visualization was thought.
4. Data limitation regarding the availability of the listings in the Foursquare database may result in many venues being seen as omitted. Results would have been more comprehensive had the dataset queried carried more inclusive results
5. Have felt the need of denoting by differing shapes the varied venues while imposing a merged data frame, something which I need to look up on.
6. Budget and Rating could have further streamlined the venues particularly with respect to Hotels and Restaurant. The assumption however was to produce an all inclusive set.
7. Full code is available at - https://github.com/MISHRA19/Coursera_Capstone-/blob/master/Final%20Week%20Assignment%20Notebook.ipynb

Conclusion

In this capstone project we have now defined a hypothetical problem statement about organization of a seminar and have queried data points using foursquare API to provide logistic choices for various venues of interest. We have also been able to suggest a probable location of choice for accommodation in the Results section of this report.