Capstone Project – The Battle of the Neighbourhoods

Introduction

- I am currently living in an apartment in a lovely suburb of Dublin, Ireland. During the Covid-19 lockdowns I have really come to value and appreciate the importance of having open spaces such as parks to spend time in. I love to spend my time going for runs, supporting local businesses by buying take out coffees and nice pastries, and sitting out reading in the park when the weather is good.
- Next year I would like to move to Ontario, Toronto to work for the summer. I would ideally be looking for bar work. I would like to be located fairly centrally in Toronto, but have the same kind of amenities available to me there as I do in Dublin (parks, coffee shops, bakeries).
- This project aims to analyse Toronto neighbourhoods and boroughs, with the aim of finding an area of Toronto that would be a good match for me to move to. This is quite a personal project, based on my own interests, but might also offer insights to anyone else looking to find out more about amenities in Toronto boroughs.

Data

The data that I have used for my project is:

- https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M – This dataset contains a list of Toronto postal codes in the province of Ontario, along with their corresponding neighbourhoods and boroughs. I used the beautiful soup function to extract this dataset from the Wikipedia page.
- https://cocl.us/Geospatial_data - This dataset contains the latitude and longitude of each postal code and was merged onto the above dataset.
- I also used the Foursquare API to access location data of places in Toronto by using the getNearbyVenues function.

Methodology

1. I extracted the Toronto dataset from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M using the BeautifulSoup function. I then formatted the dataset to drop anything where the borough was not assigned and to use the borough name for any missing neighbourhood.

```
source = requests.get('https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M').text
soup=BeautifulSoup(source, 'lxml')
table = str(soup.table)
display_html(table, raw=True)
```

```
df1 = df.replace('Not assigned', np.NaN)
df1 = df1.dropna(subset = ['Borough'], axis=0)
df1.head()
```

```
df1['Neighbourhood'] = np.where(df1['Neighbourhood'] == 'Not assigned',df1['Borough'], df1['Neighbourhood'])
df1.head()
```

Methodology continued

2. I then accessed the location data (latitudes and longitudes) at https://cocl.us/Geospatial_data and merged this dataset onto the dataset from step 1.

```
location_data = pd.read_csv('https://cocl.us/Geospatial_data')
location_data.head()
```

```
df2 = pd.merge(df1, location_data, on='Postal Code')
df2.head()
```

Methodology continued

3. I then used geolocator to extract the latitude and longitude of Toronto and used folium to plot the neighbourhoods of Toronto on the map.

```
address = 'Toronto'
geolocator = Nominatim(user_agent="toronto_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinates of Toronto are {}, {}.'.format(latitude, longitude))
```

```
map_toronto = folium.Map(location=[latitude, longitude], zoom_start=10)

# add markers to map
for lat, lng, borough, neighbourhood in zip(df2['Latitude'], df2['Longitude'], df2['Borough'], df2['Neighbourhood']):
    label = '{}, {}'.format(neighbourhood, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_toronto)

map_toronto
```

Methodology continued

4. From analysing the map produced in step 3, I decided that I was only interested in boroughs containing the word Toronto (as these were more central locations).

```
df3 = df2[df2['Borough'].str.contains('Toronto',regex=False)]
df3.shape
```

```
1]:  (40, 5)
```

```
map_toronto = folium.Map(location=[latitude, longitude], zoom_start=12)

# add markers to map
for lat, lng, borough, neighbourhood in zip(df3['Latitude'], df3['Longitude'], df3['Borough'], df3['Neighbourhood']):
    label = '{}, {}'.format(neighbourhood, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_toronto)

map_toronto
```

Methodology continued

5. I then accessed the Foursquare API and used one hot encoding to count the numbers of each venue type in each of the five areas containing Toronto in its name.

```
def getNearbyVenues(names, latitudes, longitudes, radius=500):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):
        print(name)

        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.format(
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

        nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['Neighborhood',
                  'Neighborhood Latitude',
                  'Neighborhood Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

Methodology continued

```
toronto_venues = getNearbyVenues(names=df3['Borough'],
                                 latitudes=df3['Latitude'],
                                 longitudes=df3['Longitude']
                                )
```

```
# one hot encoding
toronto_onehot = pd.get_dummies(toronto_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
toronto_onehot['Neighborhood'] = toronto_venues['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [toronto_onehot.columns[-1]] + list(toronto_onehot.columns[:-1])
toronto_onehot = toronto_onehot[fixed_columns]

toronto_onehot.head()
```

Methodology continued

6. I then grouped the areas and created a list of the top 10 venues in each of the areas.

```
toronto_grouped = toronto_onehot.groupby('Neighborhood').mean().reset_index()
toronto_grouped
```

```
def return_most_common_venues(row, num_top_venues):
    row_categories = row.iloc[1:]
    row_categories_sorted = row_categories.sort_values(ascending=False)

    return row_categories_sorted.index.values[0:num_top_venues]

num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = toronto_grouped['Neighborhood']

for ind in np.arange(toronto_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(toronto_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

Results

- The below table shows the results of the top 10 venues in each of the areas containing the word Toronto.
- In all but one area (Toronto/York), venues seem to be predominantly food based e.g. restaurants, cafes, bakeries.
- In terms of looking for a location that might have employment opportunities in a bar, East or West Toronto look to be the best options as they both have either ‘Bar’ or ‘Pub’ listed in their top 10 venues.
- Toronto/York seems to be more commercial than the other areas and so could probably be ruled out as a place to live.

	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 Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Central Toronto	Coffee Shop	Pizza Place	Sandwich Place	Park	Café	Restaurant	Gym	Sushi Restaurant	Dessert Shop	Clothing Store
1	Downtown Toronto	Coffee Shop	Café	Restaurant	Hotel	Japanese Restaurant	Italian Restaurant	Bakery	Park	Seafood Restaurant	Gym
2	East Toronto	Coffee Shop	Greek Restaurant	Brewery	Italian Restaurant	Park	Ice Cream Shop	Pub	American Restaurant	Bakery	Pizza Place
3	Toronto/York	Convenience Store	Breakfast Spot	Brewery	Bus Line	Women's Store	Distribution Center	Event Space	Ethiopian Restaurant	Escape Room	Electronics Store
4	West Toronto	Bar	Café	Coffee Shop	Bakery	Italian Restaurant	Restaurant	Breakfast Spot	Pizza Place	Park	Grocery Store

Discussion

- From these results I can tell that the area of West Toronto seems to be most ideally suited to my needs.
- The most common venue in the area are bars which would indicate that there should be plenty of bar tending work opportunities.
- The second and third most common venues are cafes/coffee shops, and the fourth is a bakery which means I would never be short of my morning coffee and pastry!
- I also mentioned that outdoor spaces such as parks are important to me – parks are the ninth most common venue in the area, indicating that there would be at least one park in my vicinity.

Conclusion

- In my discussion section I noted that the area of West Toronto seems to be most ideally suited to my needs.
- It is however worth noting a few limitations of this study, which would require further investigation before my move:
  - The study offers no indication of how safe the area would be.
  - It also doesn't tell me anything about the cost of living in West Toronto – maybe it would be too expensive for a bar tenders wage.
- Overall, the study offers a good high level overview of the amenities in the various areas of Toronto, however further investigation would be required before making the move.

Link to Github

- Final-Capstone-Project/Capstone Project - AC.ipynb at master · A-Corrigan/Final-Capstone-Project (github.com)