Coursera Capstone Project - The Battle of Neighborhoods

Introduction

Moving to a new place is a very exciting, time-consuming and stressing process. Among the reasons why people move to a new place are a new job, boring neighborhood, and etc. Finding an ideal place to live is very important for many people. When looking for a new place usually people pay attention the price and to the distance to workplace or to school. Since the searching process is very time-consuming, we rarely pay attention to amenities or places that are around the neighborhood we are planning to move. Time is a very expensive commodity, and unless one has a car then moving around a new city or neighborhood is hard. Hence, sometimes people end up feeling bored in a new neighborhood. An app or website showing which of the neighborhoods in a new city are ideal for us based on our preferences and hobbies and also are close to the workplace or school would be a good solution. The target audience is people who are moving to a new city or looking for a new place to live/rent.

Data

The main data source will be Foursquare API, a social location service that allows users to explore the world around them. Using it we will access the data about the places around the neighborhoods in a city. The data will include name, category, latitude, longitude, rating, and etc. To do this we will need some information about the neighborhoods in a particular city such as name, latitude, longitude, distance. This data can be scraped from Wikipedia.

First, if the person is moving from another city due to for example change of workplace and actually likes the current neighborhood, then he can enter the current neighborhood and we will find similar neighborhoods in a new city. Second, the person will be required to add information about their hobbies and the app will filter neighborhoods based on the requirements. Third, the address of the workplace or school can be entered, and the results will be sorted based on the distance.

The final result will be a map showing the filtered neighborhoods, as well as recommendations about the places the person should check out. These places will be sorted based on ratings.

Note: For capstone project we will assume that the person is moving to Toronto from New York and the workplace is IBM.

Example of User Data

```
current_borough_name = ['Manhattan']
current_neighborhood_name = ['Marble Hill']
current_neighborhood_latitude = [40.876551]
current_neighborhood_longitude = [-73.910660]
workplace = '185 Spadina Ave, Toronto, ON'
hobbies =set(['Yoga Studio', 'Sushi Restaurant', 'Coffee Shop'])
```

Note:

- In the app user will only be required to enter current neighborhood name, borough, workplace, and hobbies. Current neighborhood latitude and longitude will be extracted from the database.
- Below on the example of Toronto we will show how the data about cities will be stored in the database.

Example of Toronto Data

```
toronto_data.head(3)
```

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711

Methodology

1. The Beautiful Soup library was used to scrape the Toronto neighborhoods data from the Wikipedia. This dataset was merged with geocodes dataset, which contains latitudes and longitudes for each of the neighborhoods. The final result is shown above in the *Example of Toronto Data*.
2. To explore the neighborhoods: they were plotted on map along with the workplace (red circle) using visualization library Folium. For the simplicity of the use case only boroughs that contain the word Toronto were selected.

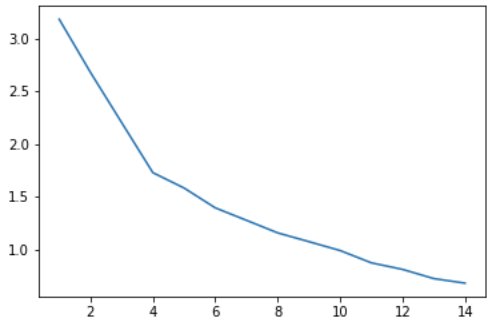
3. For each of the neighborhoods the information about the venues (ID, Name, Category, and etc.) was extracted from Foursquare API.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude
Marble Hill	40.876551	-73.91066	Arturo's	4b4429abf964a52037f225e3	40.874412	-73.910271
Marble Hill	40.876551	-73.91066	Bikram Yoga	4baf59e8f964a520a6f93be3	40.876844	-73.906204
Marble Hill	40.876551	-73.91066	Tibbett Diner	4b79cc46f964a520c5122fe3	40.880404	-73.908937

4. K-means clustering (from scikit- learn library) was used to identify neighborhoods in Toronto that are similar to Marble Hill in New York. The optimal number of clusters was 7 and it was selected by the Elbow Rule.

```
K = range(1,15)
all_grouped_clustering = all_grouped.drop('Neighborhood', 1)
wcss = [KMeans(n_clusters=i,random_state=0).fit(all_grouped_clustering ).inertia_ for i in K]
plt.plot(range(1,15),wcss)
```

```
[<matplotlib.lines.Line2D at 0x7ff2f91f7710>]
```



5. The neighborhoods that were similar to Marble Hill were filtered based on person's hobbies (in this case, they were Sushi, Yoga, and Coffee). Only neighborhoods that had venue category matching 2 or more of the person's hobbies were selected.
6. For each of the selected neighborhoods a list of popular and highly rated venues was generated based on the person's hobbies:

	Neighborhood	Venue and Rating
0	Adelaide, King, Richmond	[JaBistro-Sushi Restaurant : 8.8, Starbucks-Co...
1	Davisville	[Sakae Sushi-Sushi Restaurant : 7.6, Hokkaido ...
2	Deer Park, Forest Hill SE, Rathnelly, South Hi...	[Daeco Sushi-Sushi Restaurant : 7.4, Starbucks...
3	Harbourfront East, Toronto Islands, Union Station	[Pilot Coffee Roasters-Coffee Shop : 8.9, Balz...
4	Little Portugal, Trinity	[YogaSpace-Yoga Studio : -1.0, i deal coffee-C...
5	North Toronto West	[Barreworks-Yoga Studio : 7.7, Starbucks-Coffe...
6	St. James Town	[Fahrenheit Coffee-Coffee Shop : 9.2, Versus C...
7	Studio District	[Te Aro-Coffee Shop : 8.9, Mercury Espresso Ba...
8	The Danforth West, Riverdale	[Moksha Yoga Danforth-Yoga Studio : 8.0, Cafe ...

7. For each of the neighborhoods above the distance to workplace (IBM) was calculated using Geopy library.

```
import geopy.distance

coords_1 = (workplace_latitude, workplace_longitude)
coords_2 = (43.679557, -79.352188) # The Danforth West, Riverdale latitude and longitude

print("The distance from The Danforth West, Riverdale to IBM is {} km".format(np.round(geopy.dist
ance.vincenty(coords_1, coords_2).km,decimals=2)))
```

```
The distance from The Danforth West, Riverdale to IBM is 4.87 km
```

8. The final result was sorted based on the distance to workplace and the data frame that contains top 5 closest neighborhoods along with the venues and rating was generate

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Dist to Work	Venue and Rating
6	M5H	Downtown Toronto	Adelaide, King, Richmond	43.650571	-79.384568	0.96	[JaBistro-Sushi Restaurant : 8.8, Starbucks-Co...
7	M5J	Downtown Toronto	Harbourfront East, Toronto Islands, Union Station	43.640816	-79.381752	1.54	[Pilot Coffee Roasters-Coffee Shop : 8.9, Balz...
5	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418	1.71	[Fahrenheit Coffee-Coffee Shop : 9.2, Versus C...
8	M6J	West Toronto	Little Portugal, Trinity	43.647927	-79.419750	1.89	[YogaSpace-Yoga Studio : -1.0, i deal coffee-C...
4	M4V	Central Toronto	Deer Park, Forest Hill SE, Rathnelly, South Hi...	43.686412	-79.400049	4.09	[Daeco Sushi-Sushi Restaurant : 7.4, Starbucks...

Note:

Pandas and numpy were the main libraries used to manipulate the data.

Results

1. Marble Hill was assigned to cluster one. In cluster one there were 29 Toronto neighborhoods. A snapshot of the clusters is provided below:

In [41]:
```
similar_neighborhoods = all_merged.copy().loc[(all_merged['Cluster Labels'] == 1) & (all_merged['N
eighborhood'] != 'Marble Hill')]
print('There are {} other neighborhoods in that cluster:'.format(len(similar_neighborhoods)))
similar_neighborhoods
```

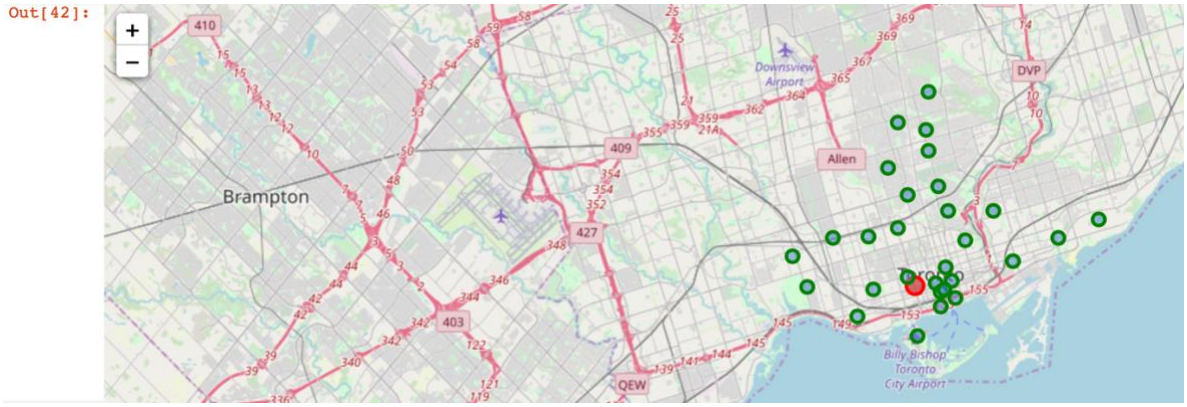
```
There are 29 other neighborhoods in that cluster:
```

Out[41]:

	PostalCode	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	M4E	East Toronto	The Beaches	43.676357	-79.293031	1	Astrologer	Coffee Shop	Pub	Yoga Studi
1	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188	1	Greek Restaurant	Coffee Shop	Ice Cream Shop	Italian Restaurant
2	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572	1	Park	Intersection	Pet Store	Pizza Place
3	M4M	East Toronto	Studio District	43.659526	-79.340923	1	Café	Coffee Shop	Yoga Studio	Bakery
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	1	Bus Line	Park	Swim School	Dim Sum Restaurant

Those neighborhoods were visualized on map as well:

Out[42]:



2. Those 29 neighborhoods were filtered, so that there are venues in each neighborhood that match at least 2 of the person's hobbies. Out of 29 only 9 neighborhoods contained venues that matched person's interests (Sushi, Yoga, Coffee). They were visualized on map:

Out[124]:

A snapshot of the 9 neighborhoods along with the venues sorted based on the rating:

```
Adelaide, King, Richmond
['JaBistro-Sushi Restaurant : 8.8', 'Starbucks-Coffee Shop : 8.4', 'M Square Coffee Co-Coffee Shop
: 8.3', 'Pilot Coffee Roasters-Coffee Shop : 8.2', 'Yuzu No Hana-Sushi Restaurant : 8.2', 'Dineen
@CommerceCourt-Coffee Shop : 7.9', 'Starbucks-Coffee Shop : 7.8', 'HotBlack Coffee-Coffee Shop : 7
.7']


Davisville
['Sakae Sushi-Sushi Restaurant : 7.6', 'Hokkaido Sushi-Sushi Restaurant : 7.5', 'Starbucks-Coffee
Shop : 7.4', 'Second Cup-Coffee Shop : 6.0']


Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West
['Daeco Sushi-Sushi Restaurant : 7.4', 'Starbucks-Coffee Shop : 6.4', 'Tim Hortons-Coffee Shop : 6
.3']
```

3. Finally, the top 5 closest to workplace neighborhoods were generated and visualized on map:

Out[189]:


```
Here are the 5 closest neighbourhoods that match your interests along with the top 5 places in each neighbourhood tha
t we hope you will enjoy:

Adelaide, King, Richmond is 0.96 km away from IBM and the places you should check out are :

JaBistro-Sushi Restaurant : 8.8
Starbucks-Coffee Shop : 8.4
M Square Coffee Co-Coffee Shop : 8.3
Pilot Coffee Roasters-Coffee Shop : 8.2
Yuzu No Hana-Sushi Restaurant : 8.2
--------------------------------------------------------------------------------

Harbourfront East, Toronto Islands, Union Station is 1.54 km away from IBM and the places you should check out are :

Pilot Coffee Roasters-Coffee Shop : 8.9
Balzac's Coffee-Coffee Shop : 8.1
Tim Hortons-Coffee Shop : 7.8
Starbucks-Coffee Shop : 7.7
Gonoe Sushi Japanese Restaurant-Sushi Restaurant : 7.6
--------------------------------------------------------------------------------

St. James Town is 1.71 km away from IBM and the places you should check out are :

Fahrenheit Coffee-Coffee Shop : 9.2
Versus Coffee-Coffee Shop : 8.1
Everyday Gourmet (Teas & Coffees)-Coffee Shop : 8.0
Starbucks-Coffee Shop : 7.8
Starbucks-Coffee Shop : 7.8
--------------------------------------------------------------------------------

Little Portugal, Trinity is 1.89 km away from IBM and the places you should check out are :

YogaSpace-Yoga Studio : -1.0
i deal coffee-Coffee Shop : -1.0
Jimmy's Coffee-Coffee Shop : -1.0
The Tampered Press-Coffee Shop : -1.0
--------------------------------------------------------------------------------

Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West is 4.09 km away from IBM and the places you should
check out are :

Daeco Sushi-Sushi Restaurant : 7.4
Starbucks-Coffee Shop : 6.4
Tim Hortons-Coffee Shop : 6.3
```

Discussion

During the analysis of the results it was identified that not always there will be neighborhoods that match all requirements (neighborhood similarity, hobbies, workplace distance). However, at least one of the criteria will always be satisfied. For example, in this case there was no neighborhood that satisfied all person's interests, however, there were that satisfied at least 2. The final choice should be made by the person.

Conclusion

This tool could be an add on to rental advertisement app or website (for example Kijiji), so that a person spends less time to search for a new place since he/she will already receive sorted results based on the indicated requirements.