The Battle of the Neighbourhoods

Applied Data Science Capstone Project Report

Introduction

Toronto is the capital city of Canadian province of Ontario. It is the most populous city in Canada and the fourth most populous city in North America. Toronto is one of the most multicultural urban areas in the world. Each year tens of thousands of newcomers from around the globe choose the city as their new home. Their diverse cultures and communities have helped create Toronto's identity as a vibrant global city.

By leveraging machine learning and big data, we will help new immigrants and visitors to find the best neighbourhoods for them to stay, based on their preferences and requirements.

An example of business case

John is moving to Toronto next year because he will be studying in the University of Toronto. Now he is looking for a place to rent. However, he does not know which neighbourhood suits him.

His search criteria are as below:

- Close to the campus.
- As a well-trained barista, he wants to find a part time job in a coffee shop while studying. If a neighbourhood has more coffee shops, he will have more chances to find a job.

Project Objective

The goal of the project is to identify neighbourhoods that meets John's criteria.

Data

following data will be used:

- Toronto neighbourhoods: The dataset has 10 boroughs and 103 neighbourhoods. (downloaded from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- Geographical coordinates: The dataset has coordinates of the 103 neighbourhoods. (downloaded from http://cocl.us/Geospatial_data)
- Venue data: (retrieved via Foursquare API by using following query https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={})

Methodology

I am using the basic methodology as taught in week 3 for this project.

1. Toronto neighbourhood dataset and geographical coordinates are merged to produce a data frame as below:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494

2. Find geographical coordinates of University of Toronto, Harbord:

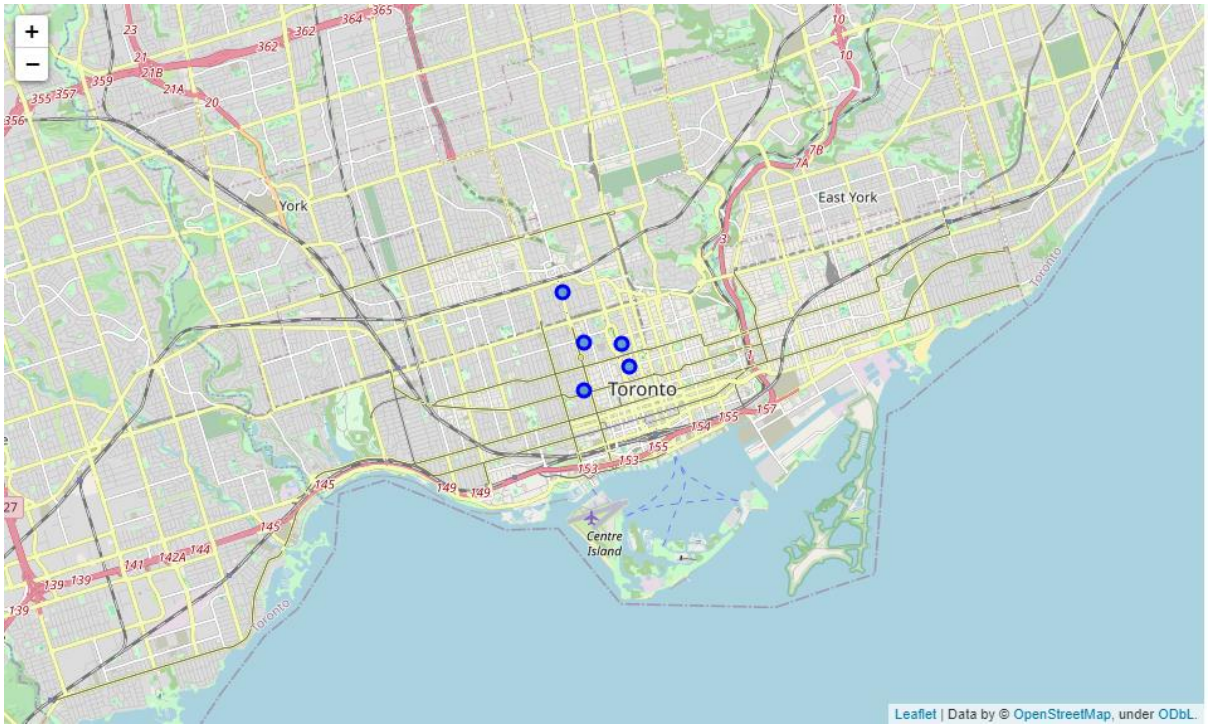
```
The geograpical coordinate of Toronto are 43.6534817, -79.3839347.
```

3. Add a new column to the data frame to store the distance between each neighbourhood to the University of Toronto and sort the data frame using this new column. Harbord. geopy.distance.geodesic() is used for distance calculation:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	Distance
0	M5S	Downtown Toronto	University of Toronto, Harbord	43.662696	-79.400049	0.191289
1	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	0.766481
2	M5R	Central Toronto	The Annex, North Midtown, Yorkville	43.672710	-79.405678	1.111517
3	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	1.137880
4	M5T	Downtown Toronto	Kensington Market, Chinatown, Grange Park	43.653206	-79.400049	1.215070

4. Retrieve venue data and add them to the data frame. Filter out neighbourhoods that do not have a coffee shop. Add the top 5 neighbourhoods to the recommendation list:

	Postal Code	Borough	Neighbourhood	Latitude	Longitude	Distance	Coffee Shop
0	M5S	Downtown Toronto	University of Toronto, Harbord	43.662696	-79.400049	0.191289	1
1	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	0.766481	9
2	M5R	Central Toronto	The Annex, North Midtown, Yorkville	43.672710	-79.405678	1.111517	2
3	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383	1.137880	11
4	M5T	Downtown Toronto	Kensington Market, Chinatown, Grange Park	43.653206	-79.400049	1.215070	4



Results

The outcome suggested 5 neighbourhoods to John. The neighbourhoods are:

- University of Toronto, Harbord
- Queen's Park, Ontario Provincial Government
- The Annex, North Midtown, Yorkville
- Central Bay Street
- Kensington Market, Chinatown, Grange Park

The neighbourhoods are in walking distance to the campus (less than 1.5 km) and have at least one coffee shop.

Discussion

John's case is relatively simple since he has only two items in his criteria list. Problem can be much more complicated if client has more requirements and preferences, for example, education resources, low crime rate, etc, then more data will be required.

Foursquare is a good source of venue data but we are not sure if they are up to date.

Conclusion

Since we find the top 5 neighbourhoods suite John, the goal of the project was met. We can extend the system to include more data such as education resource, public transports, park, crime rate to provide better and more accurate services to the users.