The Battle of the Neighbourhoods

Applied Data Science Capstone Project Report

Introduction

A novel system helps new immigrants and visitors to find the best neighbourhoods for them to stay in Toronto, based on their preferences and requirements.

Example

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.

The system will recommend 5 neighbourhoods to John

Data

- 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

- Integrate data from the three data sources and produce a data frame
- Calculate the distance between each neighbourhood and the University of Toronto
- Sort the data frame using the distances
- Filter out neighbourhoods that do not have coffee shops
- Recommend top 5 neighbourhoods to John

Results
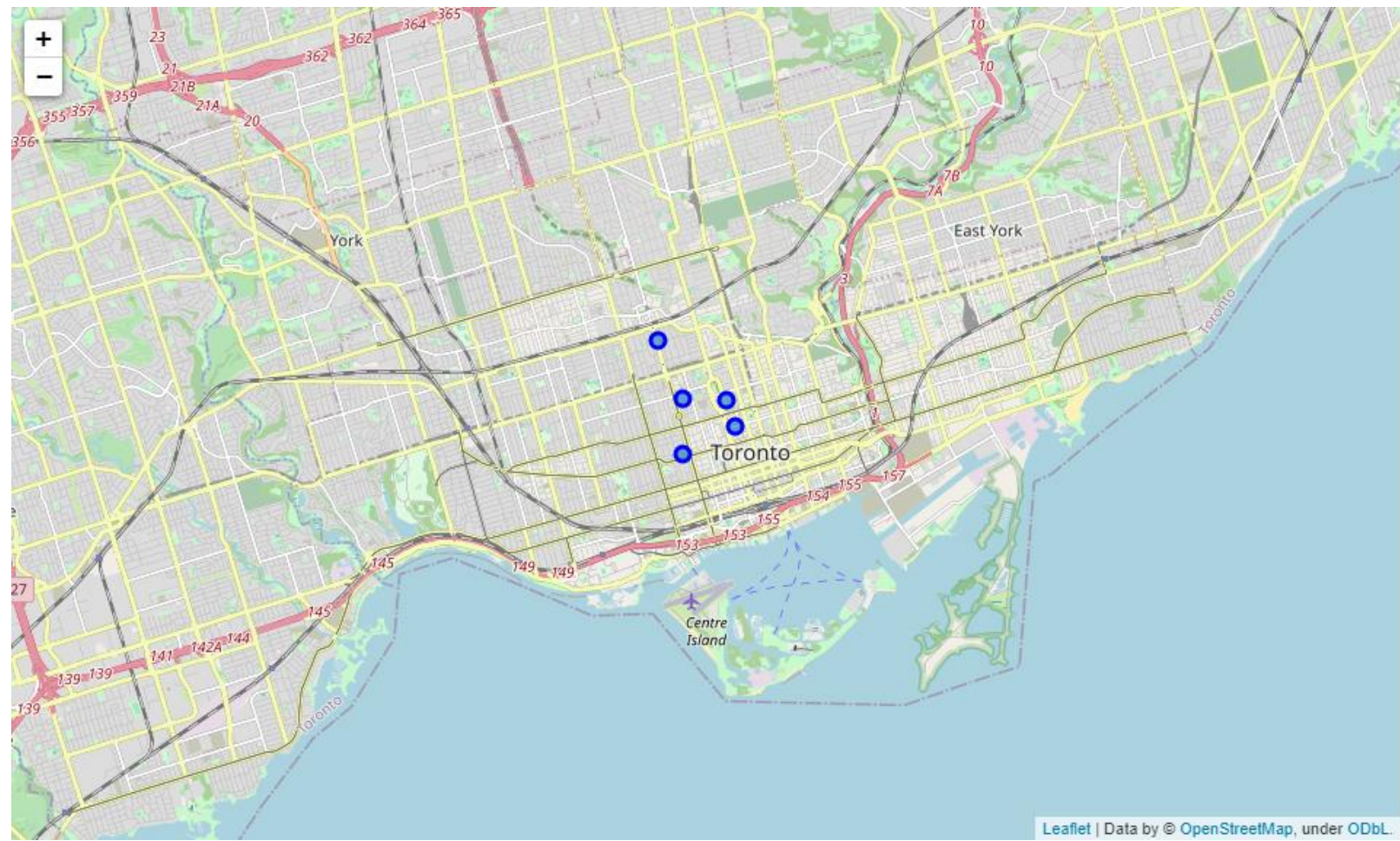
	Postal Code	Borough	Neighbourhood	Latitude	Longitude	Distance	Coffee Shop
0	M5S	Downtown Toronto	University of Toronto, Harbord	43.662696	-79.400049	0.191289	1
1	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662301	-79.389494	0.766481	9
2	M5R	Central Toronto	The Annex, North Midtown, 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

Disucssion

- John's case is relatively simple since he has only two items in his criteria list. Problem can be much 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.