

Capstone Project -- Battle of Neighbourhoods

1.0: Background

YourLovelyHome is a new start-up company providing agent services and housing consulting for customers who intent to rent or buy a property based on their specific demands. Different from other traditional consultation company, YourLovelyHome adopted the approach of machine learning technique combined the geographical information to satisfy the requirement from clients, making the recommendation more accurate and quantified.

1.1: Business Problem

The clients, Mr Smith and Mrs Smith, are a couple of retirees. They are living in the old Toronto area. After the retirement, they want to have a live in a less compacting and convenient place. They hope their life wouldn't deviate very much in term of the culture. As a result, they want to choose a place among Toronto or North York. Here are 3 main requirements purposed by them.

1. Health: there should be green environment nearby or places for exercise.
2. Convenience: The travel and should be convenient and some stores nearby.
3. Diversity: The type of venue should not be tilted to one type.

2.0: Data

Neighbourhood's information of Toronto and North York were extracted from the followed link:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

The location information including venue and venue categories within the neighbourhood were accessed through the Foursquare database.

3.0: Methodology

3.1: Summary

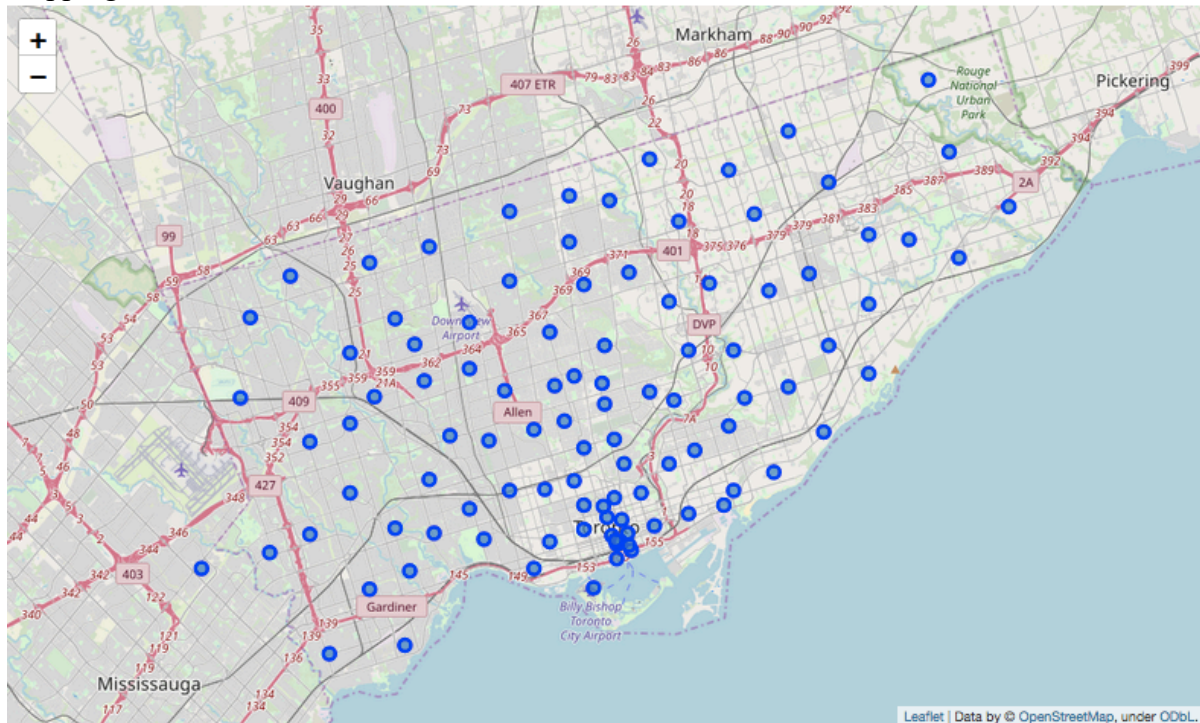
K-means clustering was applied to the data for segmentation. The neighbourhood clustered together could be treated as same type of categories. This is useful to identify the regional difference so that the places suitable for the clients were revealed.

Only the targeted cluster were needed and most of the location data were abandoned. Afterward, with a list of most common venues, final choices of neighbourhoods could be chosen and recommended to the clients based on their specific composition of venues.

3.2: Procedures

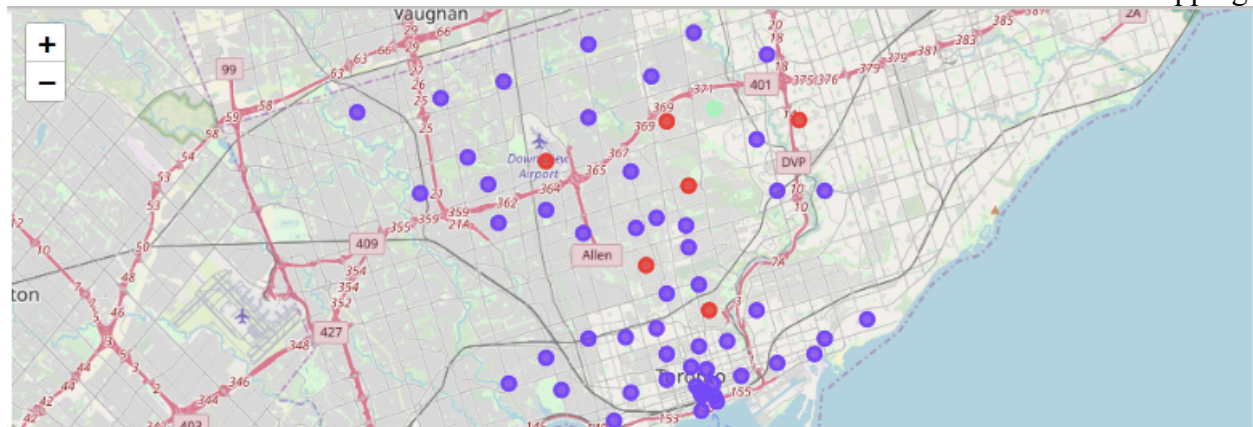
From the wiki data, there were 103 of neighbourhoods in Canada. By extracting the neighbourhood's data only for Tornado and North York, the neighbourhoods' number were reduced to 62. The latitude and longitude of the neighbourhoods were matched from

https://cocl.us/Geospatial_data. The distribution of the neighbourhoods were visualized by folium mapping.



The top 10 common venues with the 500 m radius from the centre of each neighbourhood were extracted from the Foursquare database. The radius was remained short to satisfy the requirement of being convenient. With onehot-encoding technique, 261 venues were listed correspondent to the neighbourhoods. The data were then transformed as a frequency of occurrence.

Based on the venue data, the neighbourhoods were clustered using K-means clustering with 3 clusters. The distribution of the neighbourhoods, with clusters this time, were visualized by folium mapping.



4.0: Results

There are 6 neighbourhoods belonged to “cluster 0”, 54 neighbourhoods belonged to “cluster 1” and only 1 neighbourhood belonged to “cluster 2”.

The pattern of common venues revealed that cluster 0 is the neighbourhoods’ group with park and unusual place like dog run nearby. Cluster 1 is the neighbourhoods’ group with plenty restaurants and café. Cluster 2 is the neighbourhoods’ group with highly uncommon restaurants like Filipino, Falafel, Ethiopian and Empanada restaurant. To fit the clients’ requirement, only neighbourhoods in cluster 0 were needed.

cluster 0:

| Neighborhood | Latitude | Longitude | Cluster Labels | 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 |
|-------------------------------------|-----------|------------|----------------|-----------------------|-----------------------|-----------------------|-----------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------------|
| York Mills West | 43.752758 | -79.400049 | 0 | Park | Convenience Store | Bank | Bar | Yoga Studio | Electronics Store | Doner Restaurant | Donut Shop | Dumpling Restaurant | Easterr European Restaurant |
| Parkwoods | 43.753259 | -79.329656 | 0 | Fast Food Restaurant | Park | Food & Drink Shop | Eastern European Restaurant | Discount Store | Dog Run | Doner Restaurant | Donut Shop | Dumpling Restaurant | Empanada Restaurant |
| CFB Toronto, Downsview East | 43.737473 | -79.464763 | 0 | Park | Airport | Other Repair Shop | Yoga Studio | Dumpling Restaurant | Diner | Discount Store | Dog Run | Doner Restaurant | Donut Shop |
| Lawrence Park | 43.728020 | -79.388790 | 0 | Park | Bus Line | Swim School | Yoga Studio | Dumpling Restaurant | Discount Store | Dog Run | Doner Restaurant | Donut Shop | Easterr European Restaurant |
| Rosedale | 43.679563 | -79.377529 | 0 | Park | Trail | Playground | Eastern European Restaurant | Diner | Discount Store | Dog Run | Doner Restaurant | Donut Shop | Dumpling Restaurant |
| Forest Hill North, Forest Hill West | 43.696948 | -79.411307 | 0 | Trail | Park | Sushi Restaurant | Jewelry Store | Dumpling Restaurant | Discount Store | Dog Run | Doner Restaurant | Donut Shop | Yoga Studio |

cluster 1 (partially):

| Cluster Labels | 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 |
|----------------|-----------------------|-----------------------|--------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1 | Pool | Dog Run | Mediterranean Restaurant | Golf Course | Dumpling Restaurant | Diner | Discount Store |
| 1 | Clothing Store | Fast Food Restaurant | Coffee Shop | Restaurant | Tea Room | Metro Station | Bakery |
| 1 | Café | Japanese Restaurant | Bank | Chinese Restaurant | Dumpling Restaurant | Dog Run | Doner Restaurant |
| 1 | Ramen Restaurant | Coffee Shop | Sushi Restaurant | Restaurant | Sandwich Place | Café | Japanese Restaurant |
| 1 | Pharmacy | Pizza Place | Grocery Store | Coffee Shop | Discount Store | Donut Shop | Dim Sum Restaurant |

cluster 2:

| Cluster Labels | 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 |
|----------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------------|------------------------|
| 2 | Cafeteria | Filipino Restaurant | Farmers Market | Falafel Restaurant | Event Space | Ethiopian Restaurant | Empanada Restaurant | Electronics Store | Eastern European Restaurant | Dumpling Restaurant |

There were 6 potential choices, 3 from North York and 3 from Tornado. The composition of venues was as follows:

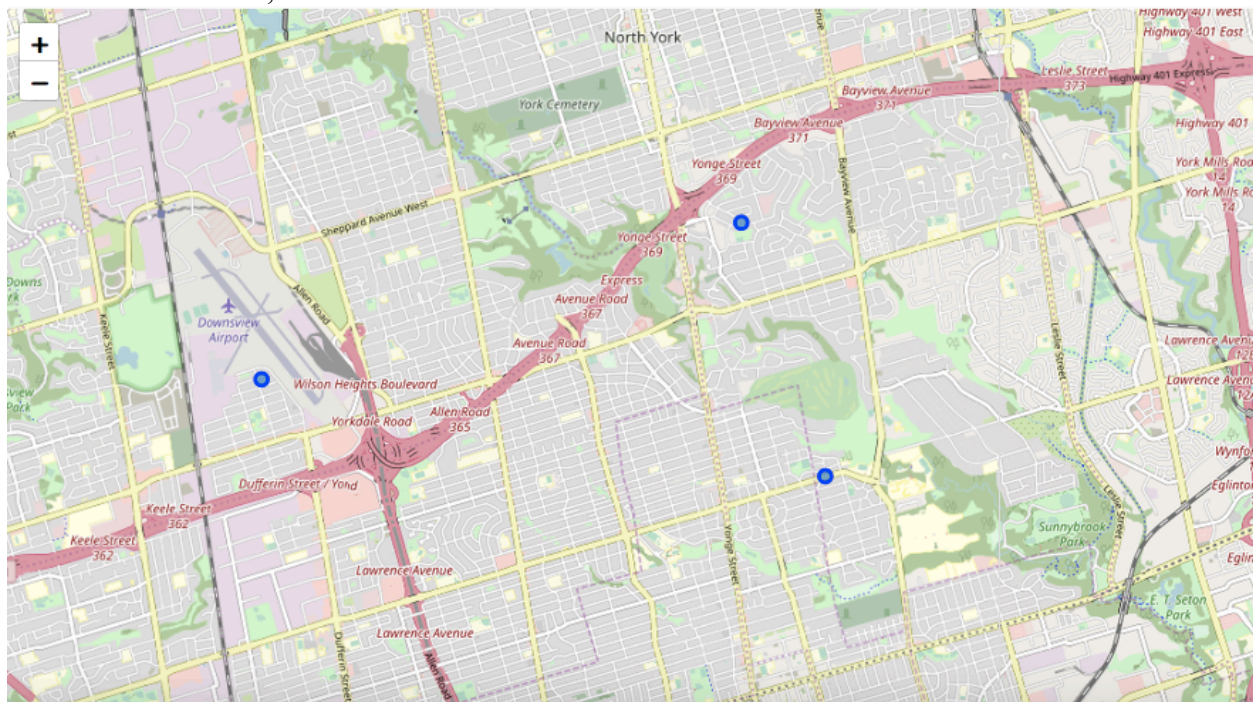
| | | Heath-related* | Store | Food & Drink | Service** | Seems not useful*** |
|-------------------|------------------------------|----------------|-------|--------------|-----------|---------------------|
| North York | York Mills West | 2 | 2 | 5 | 1 | 0 |
| | Parkwoods | 2 | 1 | 7 | 0 | 0 |
| | CFB Toronto & Downsview East | 3 | 2 | 4 | 1 | 0 |
| Toronto | Lawrence Park | 3 | 1 | 4 | 1 | 1 |
| | Rosedale | 3 | 1 | 5 | 0 | 1 |
| | Forest Hill | 4 | 1 | 4 | 0 | 1 |

*Heath-related: physically & mentally, such as dog run, park, Yoga studio

**Service: bank & travel

***Seems not useful: Jewelry Store, Playground, Swim School

The “cluster 0” included the neighbourhoods with health-related venues like dog run and park. Meanwhile, except high number of food & drink related venues in Parkwoods, the venues in rest neighbourhoods were diversified. The bank, airline and bus-line service contained a high add-on value to the place that give convenience to the clients. As a result, the final recommendation to the clients would be “York Mills West” and “CFB Toronto & Downsview East” in North York, as well as “Lawrence Park” in Toronto.



5.0: Discussion

It should be known that the radius of the covered area acted as a fundamental factor that would affect the recommendation result. As the clients are retirees and with old age, the radius of the neighbourhood are decided to be short. Moreover, to have a even more suitable recommendation, the client requirement should also be provided in details.

6.0 Conclusion

The final recommendation to the clients would be “York Mills West” and “CFB Toronto & Downsview East” in North York, as well as “Lawrence Park” in Toronto. All three of them fitted the requirement of the clients.