

Capstone Project - The Battle of Neighborhoods

1. Introduction

1.1 Background

As knowing a friend is going to move to live in Toronto, I want to do a explore and make a suggestion for where my friend should live in this city based on all the great amenities and other types of venues that exist in the neighborhood, such as gourmet fast food joints, pharmacies, parks, grad schools and so on.

1.2 Problem

The problem is how to identify the best place in Toronto to live.

1.3 Interest

Not only people who want to move the family to Toronto but also people who takes a short stay in Toronto for leisure or business would be interested in the result of this project to identify the best place in Toronto to book a hotel or buy/rent a house.

2. Data acquisition and cleaning

2.1 Data sources

Neighbourhood data are from Wikipedia for Postcode, Borough and Neighbourhood.

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

	Postcode	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned\n
1	M2A	Not assigned	Not assigned\n
2	M3A	North York	Parkwoods\n
3	M4A	North York	Victoria Village\n
4	M5A	Downtown Toronto	Harbourfront\n

Geospatial Coordinates data are from the CSV file.

https://cocl.us/Geospatial_data

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

Amenity data are from Foursquare by API request.

url =

```
'https://api.foursquare.com/v2/venues/explore?&client_id={} &client_secret={} &v={} &ll={} ,{} &radius={} &limit={} '.format(CLIENT_ID, _SECRET, VERSION, lat, lng, radius, LIMIT)
```

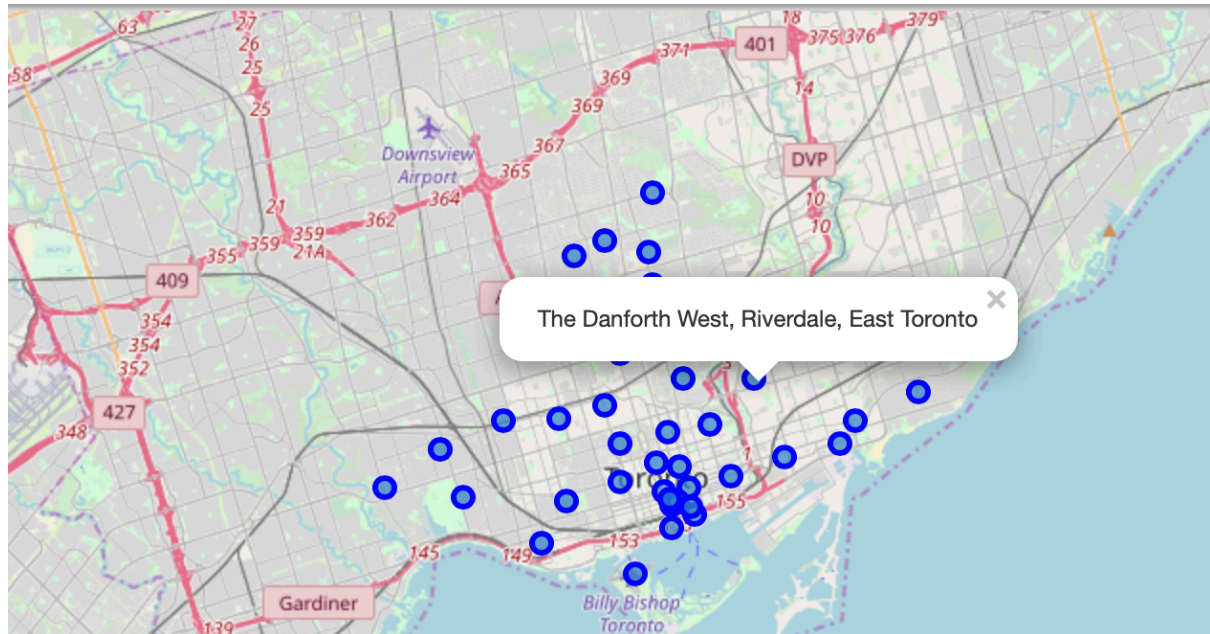
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Harbourfront, Regent Park	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Harbourfront, Regent Park	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Harbourfront, Regent Park	43.65426	-79.360636	Toronto Cooper Koo Family Cherry St YMCA Centre	43.653191	-79.357947	Gym / Fitness Center
3	Harbourfront, Regent Park	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Harbourfront, Regent Park	43.65426	-79.360636	Morning Glory Cafe	43.653947	-79.361149	Breakfast Spot

2.2 Data cleaning

Neighbourhood data are available on the web. I got them by using BeautifulSoup and put them into a dataframe. Then read the CSV file with Geospatial Coordinates into another dataframe. As both of the two dataframe have postal code, I can use the postal code as keys to consolidate two dataframes into one dataframe with Neighbourhood and Coordinates. Then I use the Coordinates to call Foursqare API to get Amenity data.

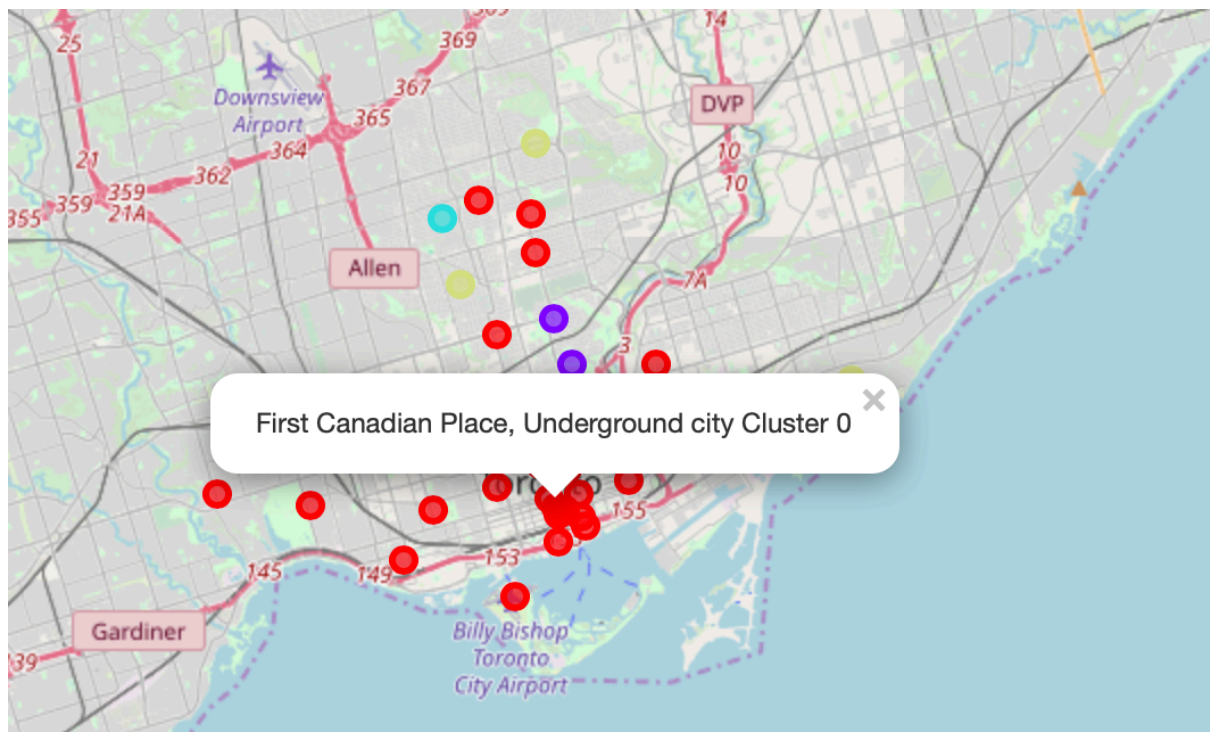
3. Exploratory Data Analysis

Once I got the Neighbourhood and Coordinates, I marked those Neighbourhood in the map with blue circles.



Then I used Foursquare API to get the nearby venues within 500 meter area. For this, I created an account on Foursquare API to get Client ID and Client Secret. I used this information to access locations on Foursquare API. I grouped them by neighborhoods and took the mean on the frequency occurrences of each venue category. This is the preprocessed data for Clustering.

I used K-Means Clustering Method to group different venues in group. I created 4 clusters to show the futures of different neighborhood Clusters.



I marked the 4 Clusters on the map with circles in different colors.

4. Result

The 4 clusters are :

Cluster 1 - Most Common Venues in this cluster are related to shop, park, Café, Store, etc.

Cluster 2 - Most Common Venues in this cluster are related to Playground, park, Trail, Building, etc.

Cluster 3 - Most Common Venues in this cluster are related to shop, Garden, Yoga Studio, Dive Bar, etc.

Cluster 4 - Most Common Venues in this cluster are related to park, Trail, Health Food Store, Lake, etc.

5. Conclusions

As clustered with different futures, the best place to live with family should be cluster 1 supported by the most important amenities for daily life. The other 3 clusters could be also taken into consideration because of diversity of people who prefers to those futures.