

Capstone Project - The Battle of Neighborhoods

1. Introduction

Introduction where you discuss the business problem and who would be interested in this project.

A New York based startup based on event management is planning its expansion in Toronto. However, unlike New York, the neighborhood data is not readily available on the internet.

The company is required to put together an application that offers good organized events and offers various services like a hotel of residence, meeting halls, places of landscapes to visit, stores for shopping, restaurants and cafes.

So, the company's purpose is to make a list of places of landscapes in Toronto, including the nearest restaurants, cafes and shopping stores for each place.

2. Data Description

For the Toronto neighborhood data, a Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto. We were required to scrape the Wikipedia page and wrangle the data, clean it, and then read it into a pandas dataframe so that it is in a structured format like the New York dataset.

Also, the data used in this project is provided by Foursquare location data. The data are grouped by landscape area, and each area included the information about this area and all information about restaurants, cafes, and stores which in this area.

2.1 Data sources

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

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.

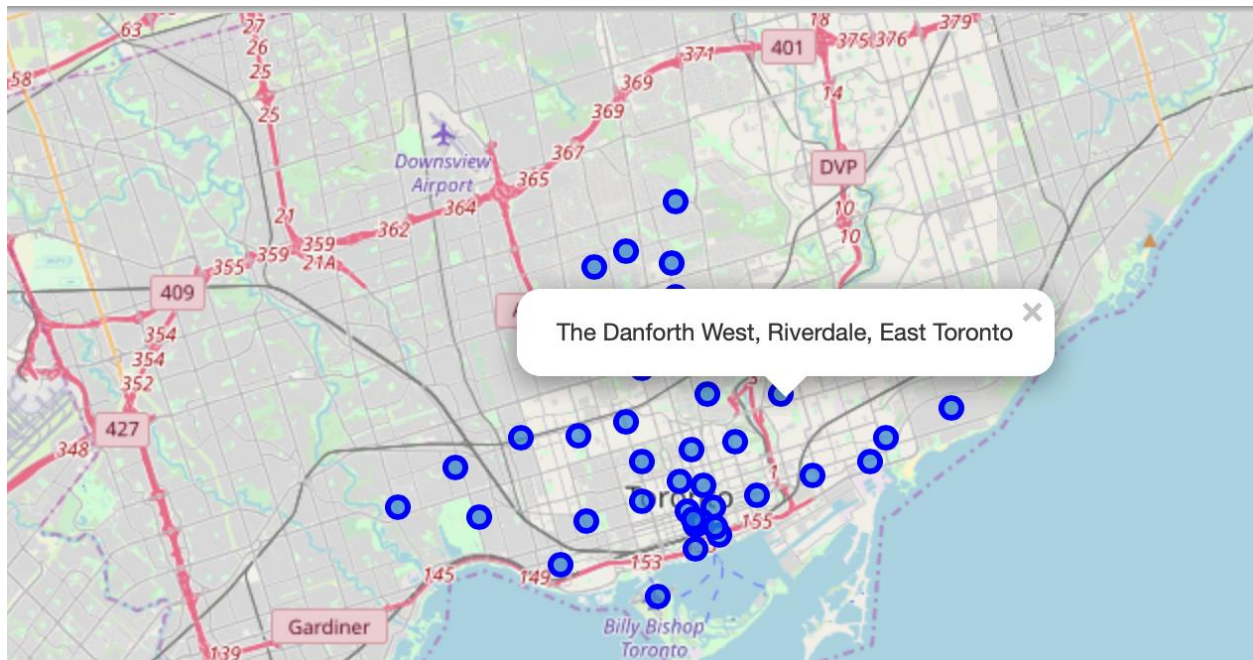
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={} &client_secret={} &v={} &ll={} ,{} &radius={} &limit={} '.format(CLIENT_ID, _SECRET, VERSION, lat, lng, radius, LIMIT)

	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 Foursquare API to get Amenity data.

3. Exploratory Data Analysis



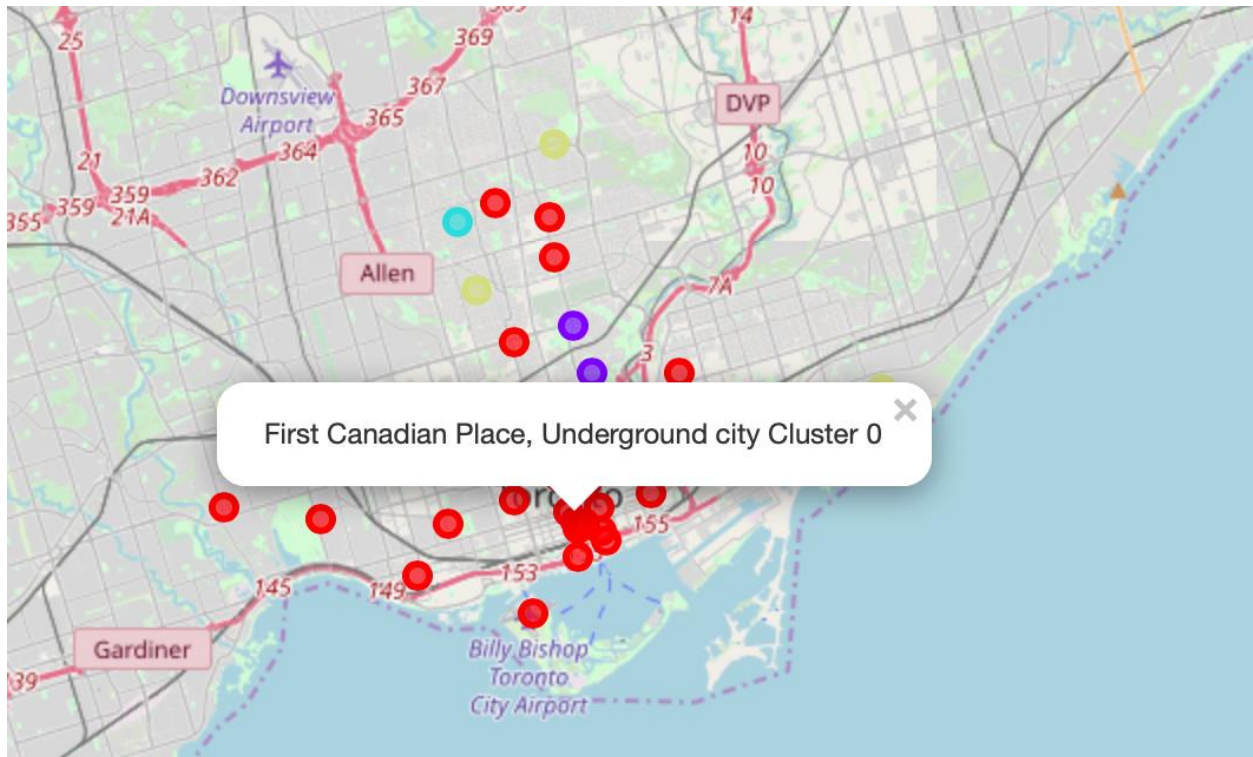
Once the Neighborhood and Coordinates were obtained, those Neighborhood in the map were marked

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.