Coursera Capstone: The Battle of the Neighborhoods

Recommendation on the Toronto City Neighborhoods for the next Expansion

1. Introduction

The Outdoor Recreation and Professional Services Inc is located in Rosedale, Downtown Toronto. The client had a successful run in the present neighborhood of the Toronto City. They offer outdoor activity and recreation products and services to their customer
.
Now with its existing business maturing in Rosendale, they would like to expand their operations in another neighborhood.

The bulk of the clients existing business comes from the venues in the existing neighborhood. The client's business operating model is therefore proven and as such the client wants to reduce the risk and accentuate the possibility of success by investing in neighborhoods with similar venue profile.

The objective therefore is to facilitate the client's selection of the neighborhoods by proposing neighborhoods that are similar to the present one.

2. Data Description

2.1: Requirements and Sources for the Project

Considering the objective of the project, to find neighborhoods similar to existing one, there is a need to conduct location analytics. All the neighborhoods within the Tornonto city are in the scope of the project.

The following data set was ascertained to be used for the initial iteration for the project:
A. List of all the boroughs associated with Toronto

The following link was found to contain all the relevant borough's along with the postal codes https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'

Columns used: PostalCode Borough Neighborhood

	PostalCode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

B. List of all the boroughs associated with Toronto

In order to explore the venues for the boroughs, the geo coordinates for the boroughs were also required. The same was pre-processed and made available postal code wise.

The csv file is to be downloaded from http://cocl.us/Geospatial_data'

Columns used: PostalCode Latitude Longitude

	PostalCode	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

C. Data from Foursquare

FOURSQUARE api with the Explore endpoint is planned to be used to fetch the details for the venues near the latitude and longitude coordinates for each borough.

In order to call the api, the developer access to Foursquare Developers was registered and the client_id and client secret keys were generated as a pre-requisite. The api call returns json with a multitude of keys and value pairs. The relevant data sets are to be extract by writing suitable functions. The json data can then be transformed into data frame using some of the existing python packages.

All the above data sets were progressively used, combined and transformed for the effective analysis, exploration, clustering and final recommendations to meet the project objective.

Image shows the sample data:

	PostalCode	Borough	Neighborhood	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
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
7	M3B	North York	Don Mills	43.745906	-79.352188
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
10	M6B	North York	Glencairn	43.709577	-79.445073
11	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...	43.650943	-79.554724

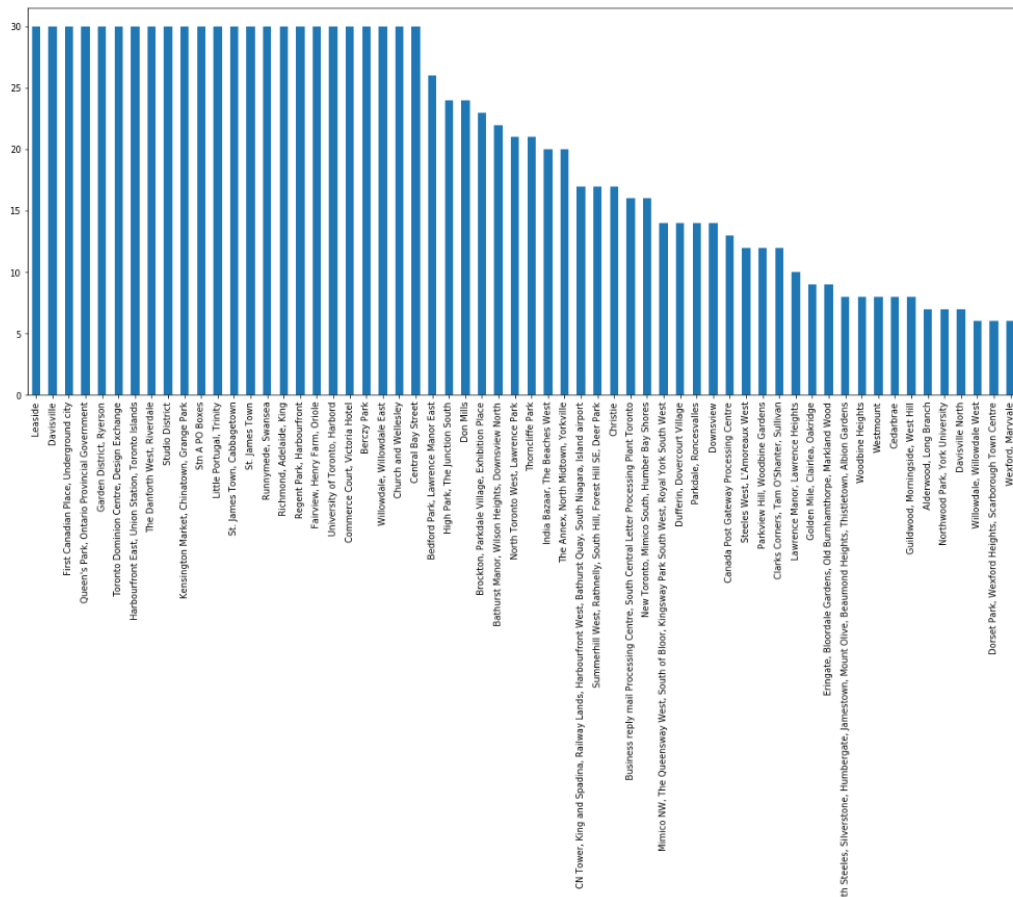## 2.2 Features
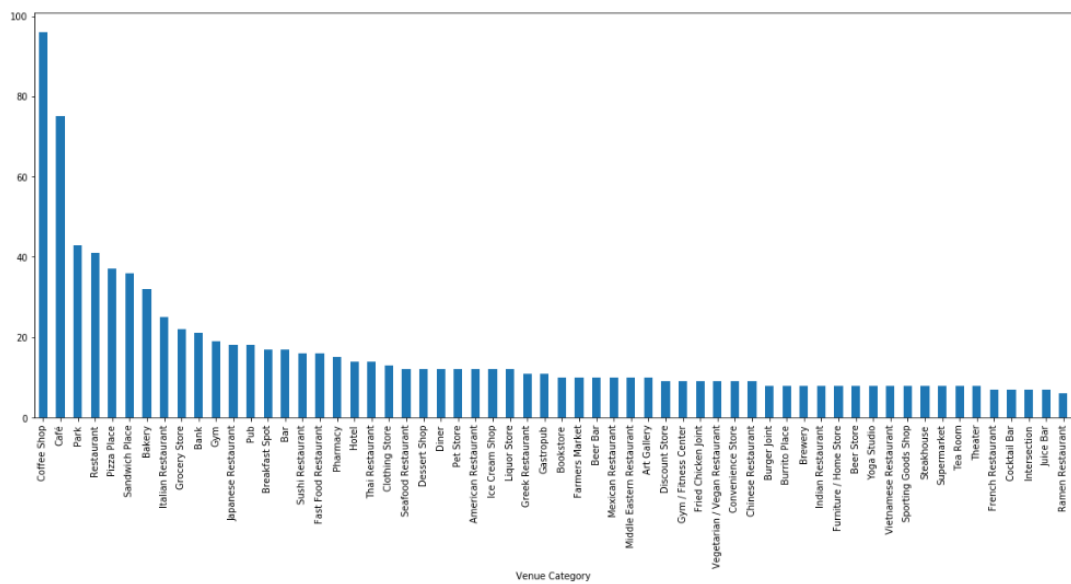
For the machine learning model, the following features are required for each of the neighborhood.

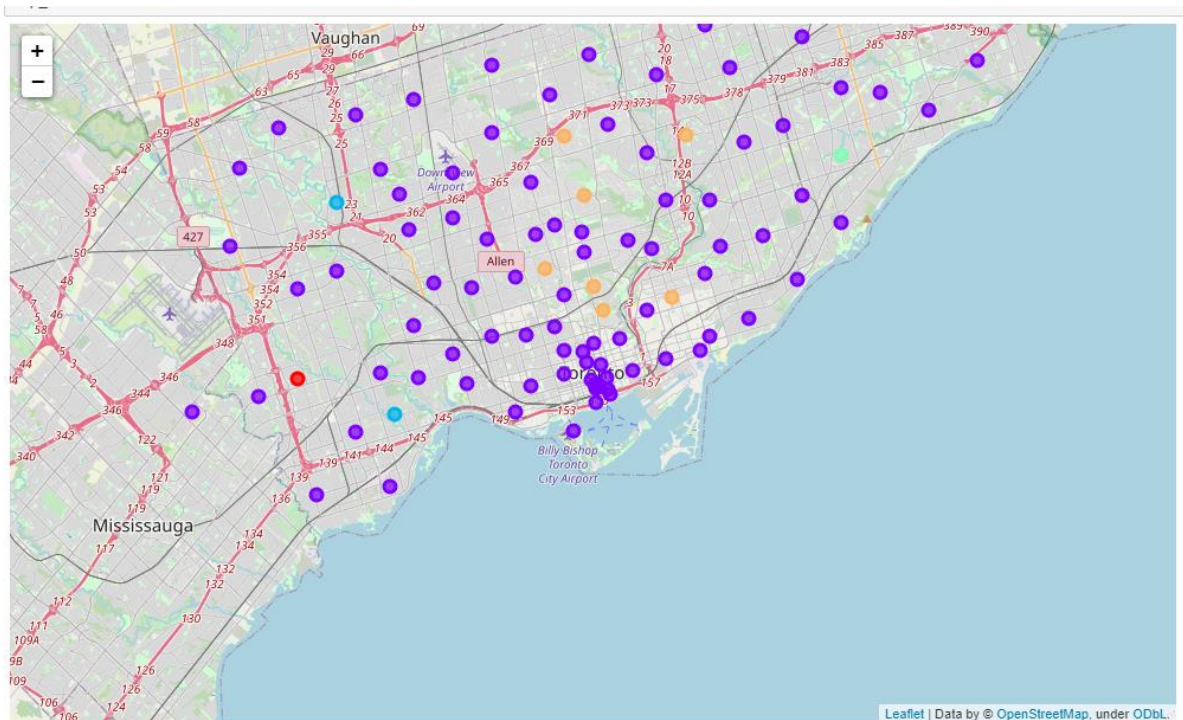| Sr. No.	Feature	Source
1	Neighborhood	Wikipedia page
2	Latitude	Enriched using the Postal Code Geo Coding file
3	Longitude	Enriched using the Postal Code Geo Coding file
4	Venue	Derived using the FourSquare API with Explore endpoint
5	Venue Latitude	Derived using the FourSquare API with Explore endpoint
6	Venue Longitude	Derived using the FourSquare API with Explore endpoint
7	Venue Category	Derived using the FourSquare API with Explore endpoint

Sample Records

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
1	Parkwoods	43.753259	-79.329656	Brookbanks Pool	43.751389	-79.332184	Pool
2	Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
3	Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
4	Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant

3. Methodology

Data was scraped from the wikipage that contained the updated information on the Postal Codes for each of the Borough's Neighborhood of the Toronto City.
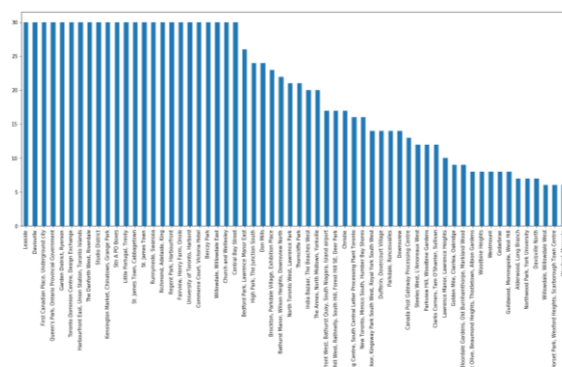
To facilitate the usage of the foursquare api, the latitude as well as longitude codes for each of the neighborhood is required. This data shall be made available by joining the postal code

data set to the mapping file available in the pre-processed csv file. This will create an intermediate table with the following sample data set:

	PostalCode	Borough	Neighborhood	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
5	M9A	Etobicoke	Islington Avenue, Humber Valley Village	43.667856	-79.532242
6	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
7	M3B	North York	Don Mills	43.745906	-79.352188
8	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
10	M6B	North York	Glencairn	43.709577	-79.445073
11	M9B	Etobicoke	West Deane Park, Princess Gardens, Martin Grov...	43.650943	-79.554724

Thereafter using the Foursquare api with explore endpoint, the details of the venues located within 500 meters of each of the neighborhood shall be obtained. As the data from the api is returned as json, further munging is required to extract relevant data set and transform the same into a data frame.

Sample output expected:

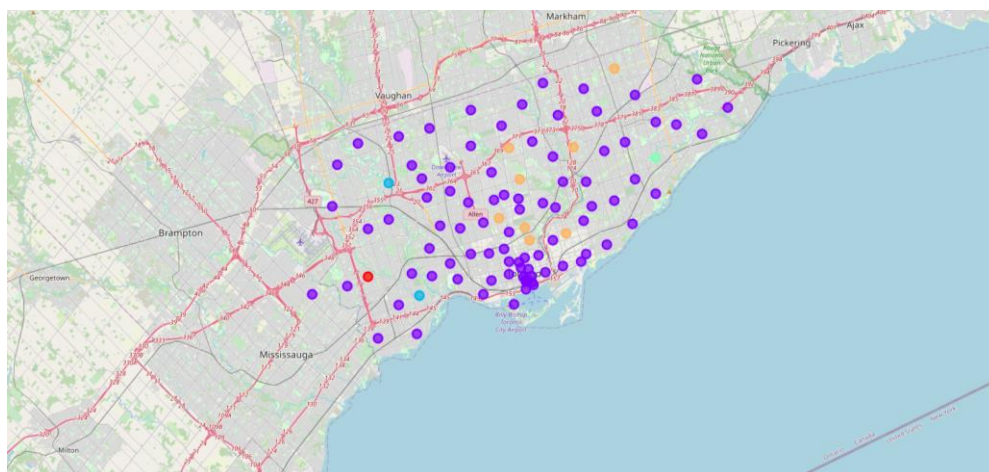| Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
Parkwoods	43.753259	-79.329656	Brookbanks Pool	43.751389	-79.332184	Pool
Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant

Exploratory Data analysis is carried out on the above data set leading to the following details:

Count of the venues per neighborhood:

Top Venue Categories

The clustering algorithm used in K-Means. Based on the venue profile for each of the neighborhood, K-means shall help ascertain the similarity amongst the neighborhoods by grouping them into the clusters.

Once the clusters are assigned to the neighborhoods, it will be easy to locate other neighborhoods that are similar to the client's present business location, Rosendale. Thereby the other neighborhoods belonging to the same cluster as Rosendale shall be the recommended locations for the next expansion location.
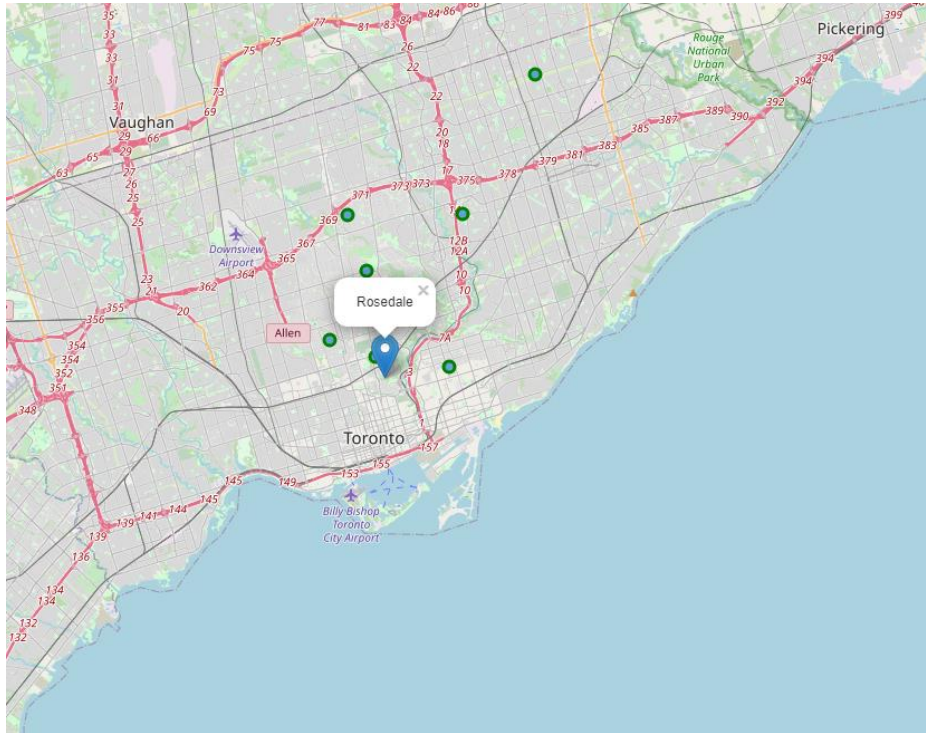
4. Result

With K-Means clustering technique, the neighborhoods have been grouped as shown below:

Client's current location, that is Rosendale has been grouped into Cluster 5. Therefore keeping in line with the projects objective of reduce the risk and maximizing return, all the locations grouped in cluster 5 are the recommended locations for the client, Outdoor Recreation and Professional Services Inc.

Neighborhoods recommended for the next business Location those are similar to Rosendale:

- Parkwoods
- East Toronto, Broadview North (Old East York)
- Lawrence Park
- York Mills West
- Forest Hill North & West, Forest Hill Road Park
- Moore Park, Summerhill East
- Milliken, Agincourt North, Steeles East, L'Amoreaux East

PostalCode	Borough	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
M3A	North York	Parkwoods	43.753259	-79.329656	4.0	Park	Pool	Food & Drink Shop	Yoga Studio	Curling Ice	Drugstore	Donut Shop	Dog Run	Distribution Center	Discount Store
M4J	East York	East Toronto, Broadview North (Old East York)	43.685347	-79.338106	4.0	Park	Convenience Store	Rental Car Location	Yoga Studio	Curling Ice	Drugstore	Donut Shop	Dog Run	Distribution Center	Discount Store
M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	4.0	Park	Swim School	Bus Line	Yoga Studio	Deli / Bodega	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center
M2P	North York	York Mills West	43.752758	-79.400049	4.0	Park	Convenience Store	Yoga Studio	Dance Studio	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center	Discount Store
M5P	Central Toronto	Forest Hill North & West, Forest Hill Road Park	43.696948	-79.411307	4.0	Park	Jewelry Store	Trail	Sushi Restaurant	Yoga Studio	Dance Studio	Drugstore	Donut Shop	Dog Run	Distribution Center
M4T	Central Toronto	Moore Park, Summerhill East	43.689574	-79.383160	4.0	Park	Trail	Yoga Studio	Curling Ice	Eastern European Restaurant	Drugstore	Donut Shop	Dog Run	Distribution Center	Discount Store
M1V	Scarborough	Milliken, Agincourt North, Steeles East, L'Amo...	43.815252	-79.284577	4.0	Playground	Park	Yoga Studio	Curling Ice	Drugstore	Donut Shop	Dog Run	Distribution Center	Discount Store	Diner
M4W	Downtown Toronto	Rosedale	43.679563	-79.377529	4.0	Park	Playground	Trail	Yoga Studio	Curling Ice	Drugstore	Donut Shop	Dog Run	Distribution Center	Discount Store

5. Discussion

Based on the result above, looks like that are seven neighborhoods that have similarity with the client's existing business location. These locations can there after be further subjected to diligence on factors such as proximity, resource availability etc to further shortlist on one that leverages the synergy and reduces cost to get the maximum possible returns on investment.