Capstone Project Report
(Toronto Neighborhood Clusters)

1. Introduction

When people would like to move to Toronto, or would like to move from one neighborhood to another, they are facing the questions such as: Which neighborhood could be the candidate? How the surroundings of this neighborhood look like? Is this neighborhood similar to where I lived before?

This project tries to cluster the neighborhoods in Toronto by venues, in order to answer these questions. By clustering the neighborhoods, similar neighborhoods will be grouped together. Also, checking the top 10 most common venues can help to provide a first picture of the surroundings and life in different neighborhood clusters. Therefore, those people who would like to relocate to one of the neighborhoods in Toronto could be interested in this project.

2. Data and Data Acquisition

We first need to obtain the list of neighborhoods in Toronto. This list can be obtained from Wikipedia page "List of postal codes of Canada: M". The coordinates of each neighborhood should be obtained, to prepare for the later exploration. The coordinates of each neighborhood can be obtained by using the "geocoder" in Python, i.e. searching the coordinates based on the name of neighborhood. Then, the venues of each neighborhood can be explored by calling Foursquare using the coordinates.

3. Methodology

The coordinates of three listed neighborhoods could not be found by "geocoder", e.g. wrong neighborhood name, no data found by "geocoder". Thus, these neighborhoods are dropped. In addition, different radius to explore neighborhood have been tried, and different number of clusters has been tried in the learning.

Clustering of the neighborhoods is an unsupervised learning process, since there are no existing "labels" for each neighborhood. Therefore, after obtaining the frequency of venues in each neighborhood, K means clustering method has been used as the learning method.

4. Results
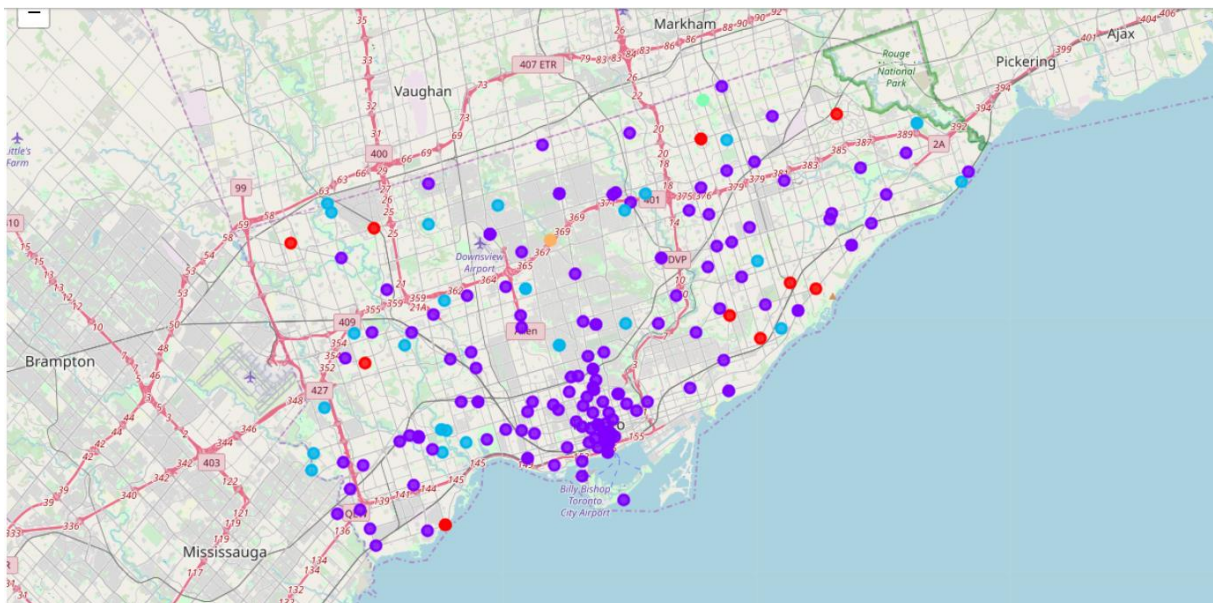
The neighborhoods in Toronto are clustered into 5 clusters, i.e.

- **Cluster 0**: Area with plenty and big variety of restaurants
- **Cluster 1**: Area with comprehensive living facilities, e.g. café, bar, hotel, Gym, etc.
- **Cluster 2**: Area for outdoor activities, e.g. park, playground, as well with restaurants
- **Cluster 3**: Area with playground and zoo
- **Cluster 4**: Area with plenty of wine shop

5. Discussion

Cluster 2

```
[40]: neighborhood_merged.loc[neighborhood_merged['Cluster Labels'] == 2, neighborhood_merged.columns[[0] + [11] + list(range(1, 10))]]
```

[40]:

	Neighborhood	Cluster Labels	1st Most Common Restaurant	2nd Most Common Restaurant	3rd Most Common Restaurant	4th Most Common Restaurant	5th Most Common Restaurant	6th Most Common Restaurant	7th Most Common Restaurant	8th Most Common Restaurant	9th Most Common Restaurant
5	Bathurst Manor	2	Playground	Convenience Store	Park	Baseball Field	Falafel Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	Ethiopian Restaurant
9	Birch Cliff	2	Park	Gym	Gym Pool	Business Service	Drugstore	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant
20	Clarks Corners	2	Park	Convenience Store	Caribbean Restaurant	Drugstore	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	Ethiopian Restaurant
43	Eringate	2	Park	Zoo	Falafel Restaurant	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	Ethiopian Restaurant	Event Space
48	Forest Hill North	2	Playground	Bank	Park	Drugstore	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	Ethiopian Restaurant
49	Forest Hill SE	2	Playground	Bank	Park	Drugstore	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	Ethiopian Restaurant
50	Forest Hill West	2	Playground	Bank	Park	Drugstore	Eastern European Restaurant	Egyptian Restaurant	Electronics Store	Empanada Restaurant	Ethiopian Restaurant
60	Henry Farm	2	Tennis Court	Lawyer	Restaurant	Park	Zoo	Event Space	Eastern European Restaurant	Egyptian Restaurant	Electronics Store
64	Humber Bay	2	Construction & Landscaping	Pool	Harbor / Marina	Park	Exhibit	Dumpling Restaurant	Eastern European Restaurant	Egyptian Restaurant	Electronics Store
	Humber		Construction &	Esennade			Dumpling	Egyptian		Ethiopian	

The venues of certain group of neighborhood are equivalent. This implies that these neighborhoods are very close to each other and the area of neighborhood is very small. This can also be confirmed by the names of these neighborhoods. Thus, the exploration results of them are the same.

There could be two ways to solve these problem: (1) Reduce the radius for exploration of these neighborhoods. (2) Group these neighborhoods together and then apply the exploration.

6. Conclusion

In this project, neighborhoods in Toronto have been clustered based on the frequency of the venues. K means clustering method has been applied for the clustering. The results show that there are 5 clusters of neighborhoods in Toronto, i.e. area with plenty and big variety of restaurants, area with comprehensive living facilities, e.g. café, bar, hotel, Gym, etc., area for outdoor activities, e.g. park, playground, as well with restaurants, area with playground and zoo and area with plenty of wine shop. The results can be further improved by adapting radius for exploration of each neighborhood or group close neighborhoods together before exploration.