1. Introduction:

The main purpose of this project is to help people explore better facilities nearby. It will help them find wise and effective dicisions in Toranto. Many people are immigrating to all kinds of states in Canada, so they need to do a lot of researches to decide good housing price and choose the school for their children. This project is just for those who are looking for a better community.

1. Data Section:

Data Link: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

Will use Scarborough dataset which we scrapped from wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes.

Foursquare API Data: We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category
9. Methodology Section:

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

4. Results Section

**Map of Clusters in Scarborough**



**Average Housing Price by Clusters in Scarborough**



**School Ratings by Clusters in Scarborough**



5. Discussion Section:

Problems to be solved:

1. A list of schools considering the location, rating, fees and comments.
2. A list of house in terms of housing prices in a sorted order.

6. Conclusion Section:

With this project I learned a lot about cluster algorithm and its applications, with the methods I finally finish my project completing some critical issues. Also, this project helps me practice the ability to combine the knowledge with some real situation.