

Introduction

The purpose of this Project is to help people in exploring better facilities around their neighborhood. It will help people make smart and efficient decisions on selecting a great neighborhood out of numbers of other neighborhoods in Scarborough, Toronto.

Lots of people are migrating to various states of Canada and need lots of research for good housing prices and reputed schools for their children. This project is for those people who are looking for better neighborhoods. For ease of access to Cafe, School, Supermarket, medical shops, grocery shops, mall, theatre, hospital, like minded people, etc.

This Project aims to create an analysis of features for people migrating to Scarborough to search for the best neighborhood as a comparative analysis between neighborhoods. The features include median housing price and better school according to ratings, crime rates of that particular area, road connectivity, weather conditions, good management for emergency, water resources both fresh and waste water and excrement conveyed in sewers and recreational facilities.

It will help people to get awareness of the area and neighborhood before moving to a new city, state, country or place for their work or to start a new fresh life.

Problem:

The major purpose of this project is to suggest a better neighborhood in a new city for the people who are shifting there. Social presence in society in terms of like minded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby .

1. Sorted list of house in terms of housing prices in ascending or descending order
2. Sorted list of schools in terms of location, fees, rating and reviews.

Data Description:

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

Will use the 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

Methodology

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.

Using credentials of Foursquare API, features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

Results

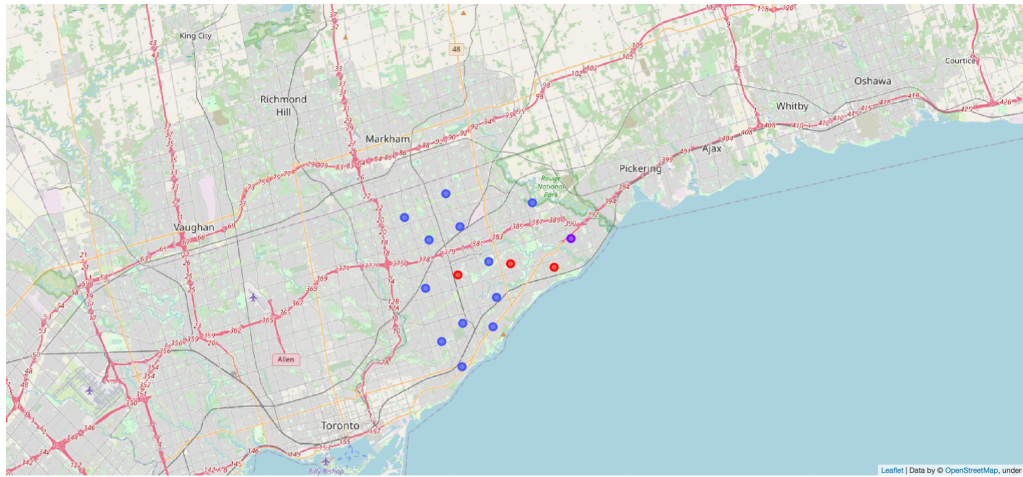


Fig. Cluster map in Scarborough

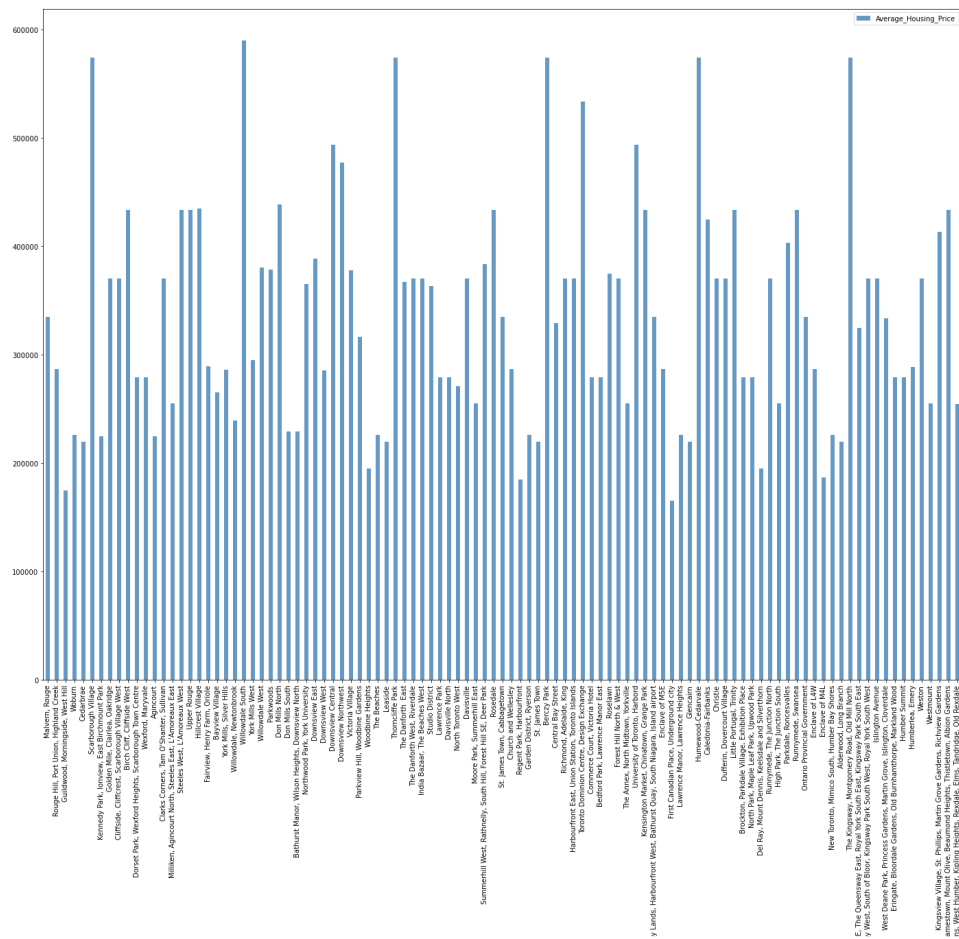


Fig. Average housing price

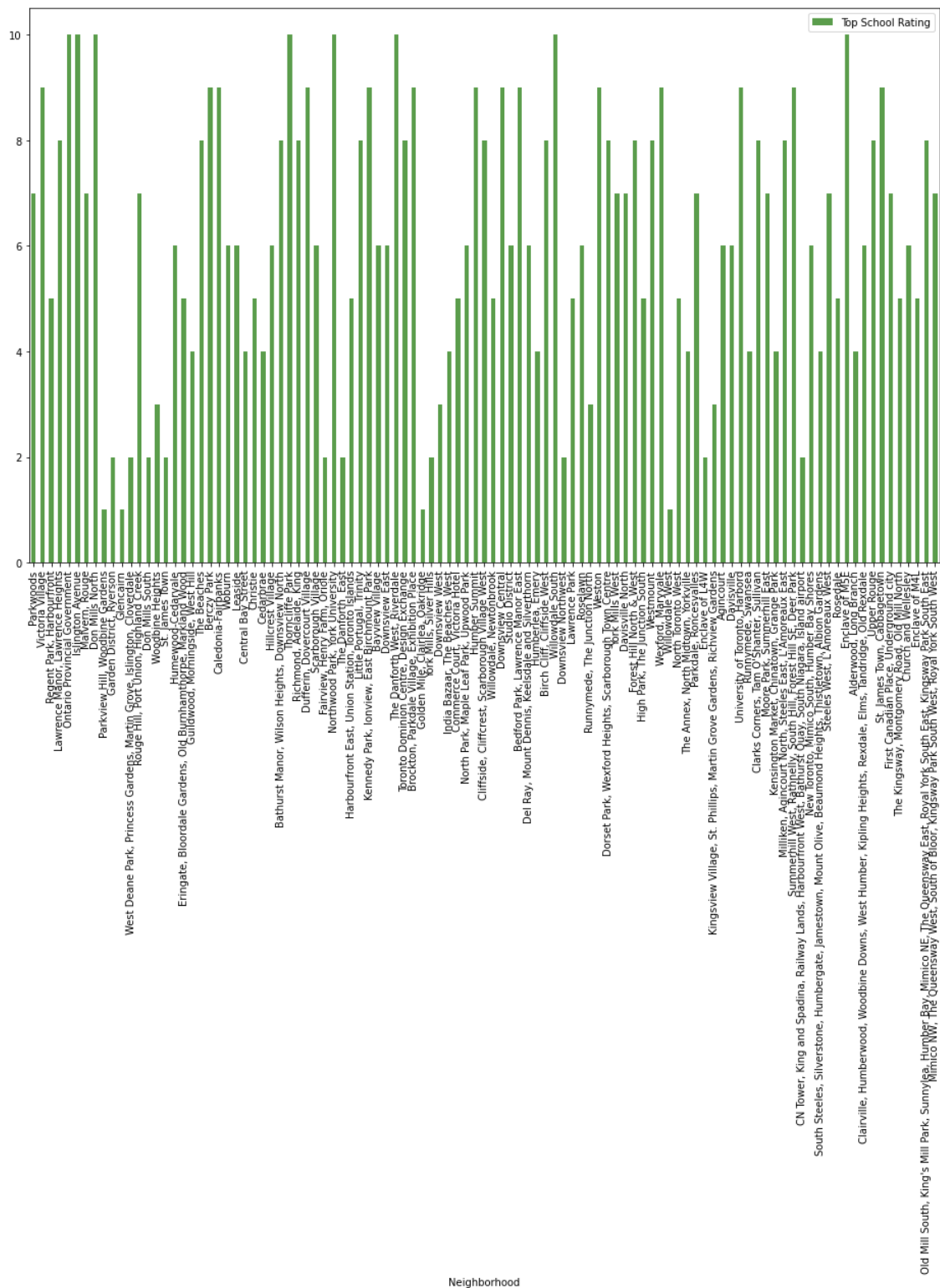


Fig. Top Schools rating

Discussion

The major purpose of this project is to suggest a better neighborhood in a new city for the people who are shifting there. Social presence in society in terms of like minded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby.

- Sorted list of house in terms of housing prices in a ascending or descending order
- Sorted list of schools in terms of location, fees, rating and reviews

Conclusion

Using the k-means cluster algorithm, the neighborhood was separated into 10 different clusters and for 103 different latitude and longitudes from the dataset, which have similar neighborhoods around them.

Using the charts above results presented to a particular neighborhood based on average house prices and school rating have been made.