Capstone Project - The Battle of Neighborhoods

Median House Prices and School Ratings analysis for Scarborough Canada

The Location

Scarborough is a popular destination for new immigrants in Canada to reside.

Project Description

Most people migrating to various states of Canada require search of a good housing prices as well as good rating schools for their children. The projects aim to create an analysis of features for a neighborhood as a comparative analysis between neighborhoods. The features include median house price and school ratings. This would help people to get awareness of the places before moving to new state, city or place for their work or to start a new life.
The aim of this Project is to help people explore different possibilities and take a better decision on choosing the best neighborhood out of many neighborhoods in Scarborough city based on the distribution of various facilities in and around that neighborhood.

Foursquare API

This project would use Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

Work Flow

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. Steps taken were:

1. Data acquisition and cleansing
2. Data preparation
3. Feature selection
4. Clustering

Data acquisition and cleansing

Data acquisition was a 2-step process:

1. Obtaining the postcodes for neighborhoods in Toronto
2. Obtaining venues within these neighborhoods

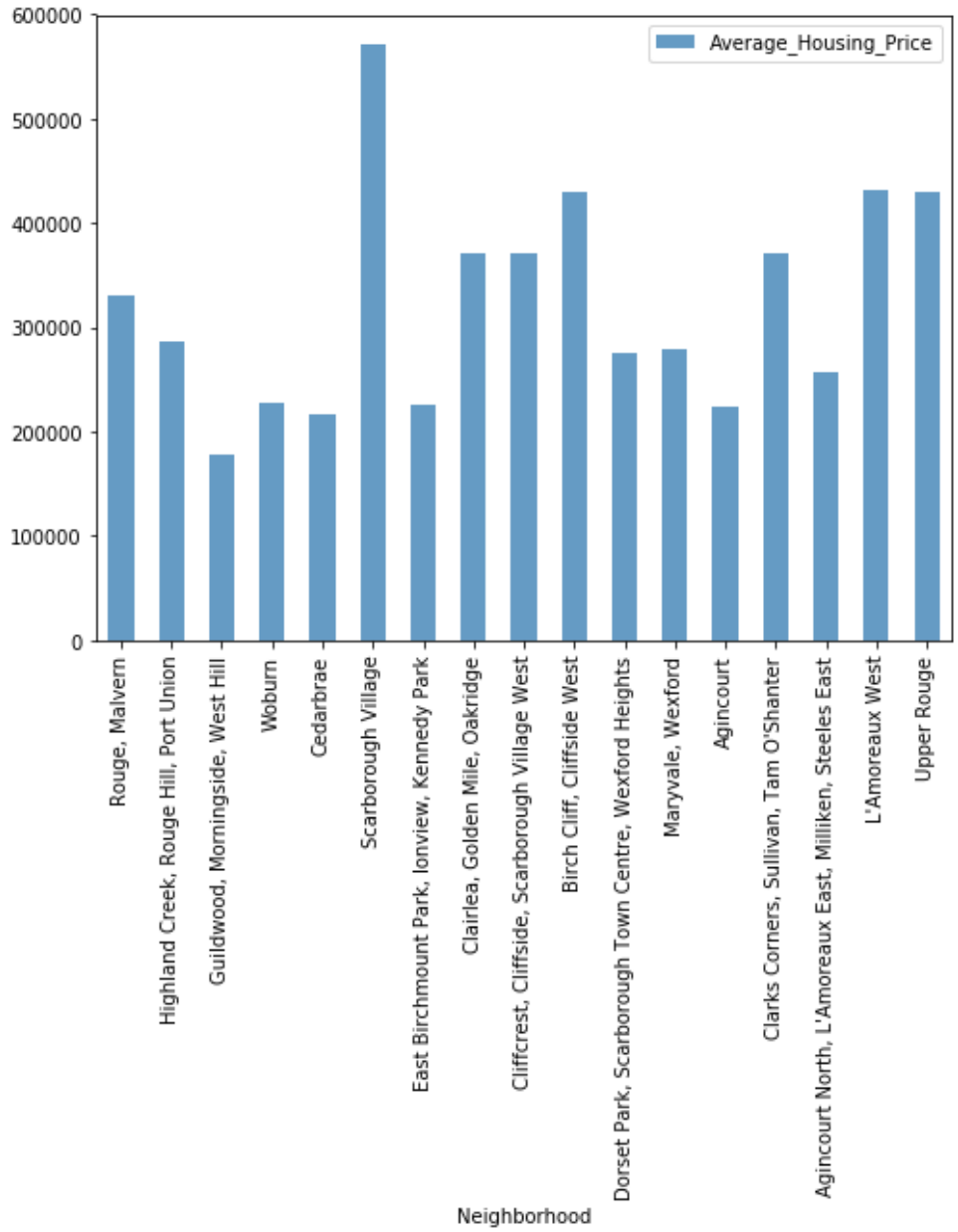
Clustering Approach

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

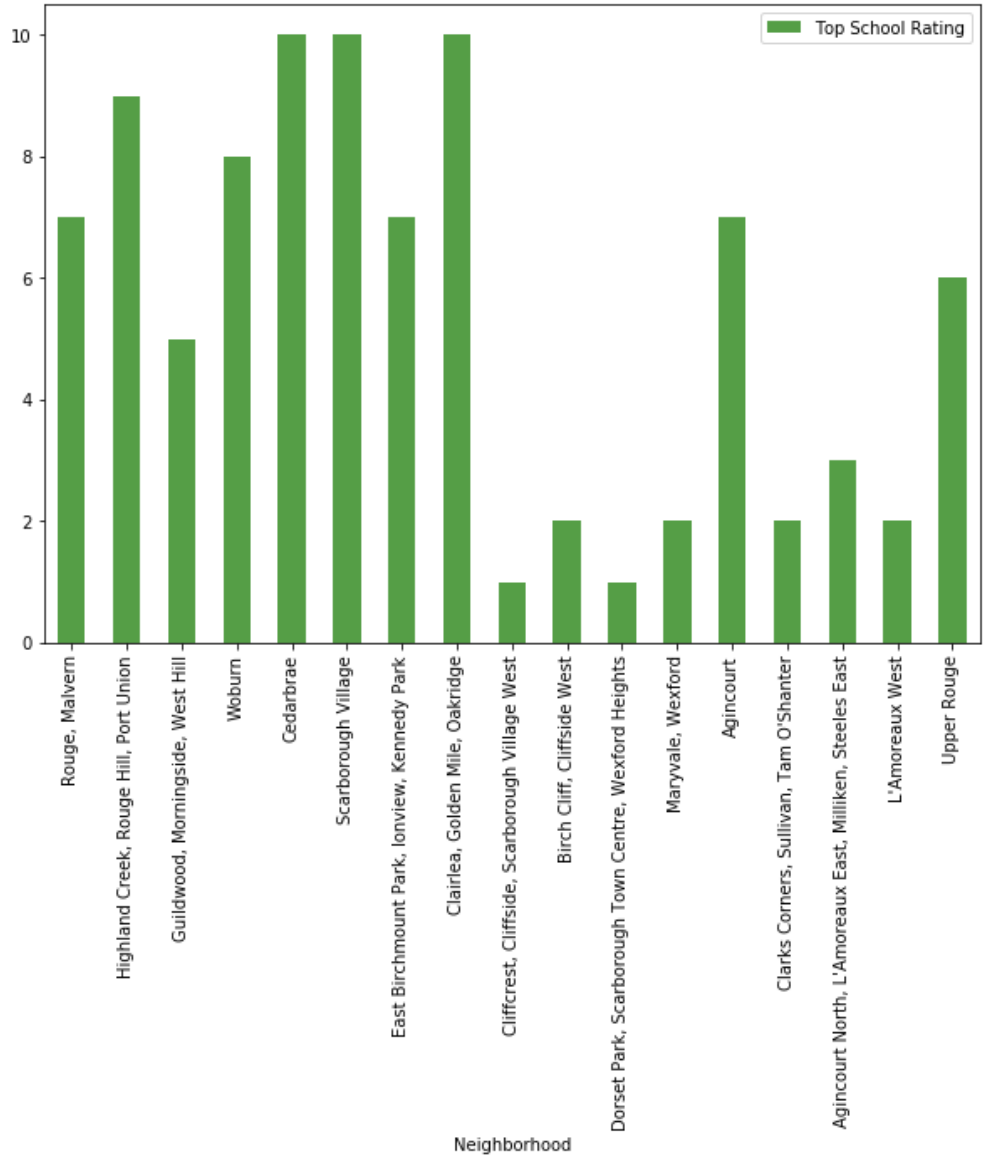
Neighborhood Common Venues?

	Neighborhood	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
0	Agincourt	Clothing Store	Breakfast Spot	Skating Rink	Lounge	Latin American Restaurant	College Stadium	Construction & Landscaping	Convenience Store	Department Store	Vietnamese Restaurant
1	Agincourt North, L'Amoreaux East, Milliken, St...	Arts & Crafts Store	Playground	Park	Vietnamese Restaurant	Clothing Store	Gas Station	Furniture / Home Store	Fried Chicken Joint	Fast Food Restaurant	Electronics Store
2	Birch Cliff, Cliffside West	General Entertainment	Skating Rink	Café	College Stadium	Vietnamese Restaurant	Clothing Store	Gas Station	Furniture / Home Store	Fried Chicken Joint	Fast Food Restaurant
3	Cedarbrae	Hakka Restaurant	Thai Restaurant	Athletics & Sports	Gas Station	Bakery	Bank	Caribbean Restaurant	Fried Chicken Joint	Discount Store	Convenience Store
4	Clairlea, Golden Mile, Oakridge	Bus Line	Bakery	Metro Station	Soccer Field	Fast Food Restaurant	Intersection	Park	Bus Station	Vietnamese Restaurant	Construction & Landscaping

Neighborhood Median house Prices



Neighborhood School Ratings



Conclusion

In this project, through a k-means cluster algorithm we separate the neighborhood into 03 clusters, which have similar neighborhoods around them. Using the charts above decision leading to a particular neighborhood based on average house prices and school rating can be made.