Capstone Project - The Battle of Neighborhoods

Analyzing Median House Prices and School Ratings for Scarborough Canada for Immigrants

The Location:

Scarborough is a popular destination for new immigrants in Canada to reside. As a result, it is one of the most diverse and multicultural areas in the Greater Toronto Area, being home to various religious groups and places of worship.

Project Description:

Many 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, crime rates, weather conditions, recreational facilities. This would help people to get awareness of the places before moving to a new country, 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. **Selection criteria** For the purposes of this project, the definition of a good neighborhood is one that has an appreciable commercial presence within a given community as well as:

1. Compare median housing prices

2. Compare school ratings

Longitude and Latitude Data:

We will need geo-locational information about that specific borough and the neighborhoods in that borough. It is "Scarborough" in Toronto. This project will require knowledge of the different neighborhoods in Toronto, school ratings and median house prices. As such the neighborhood data required will be:

1. Neighborhood location in terms of latitude and longitude

2. School Ratings

3. Median House Prices

Dataset comprising latitude and longitude, zip codes is already available through the previous notebook.

	Postalcode	Borough	Neighborhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

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

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 cleansing2. 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

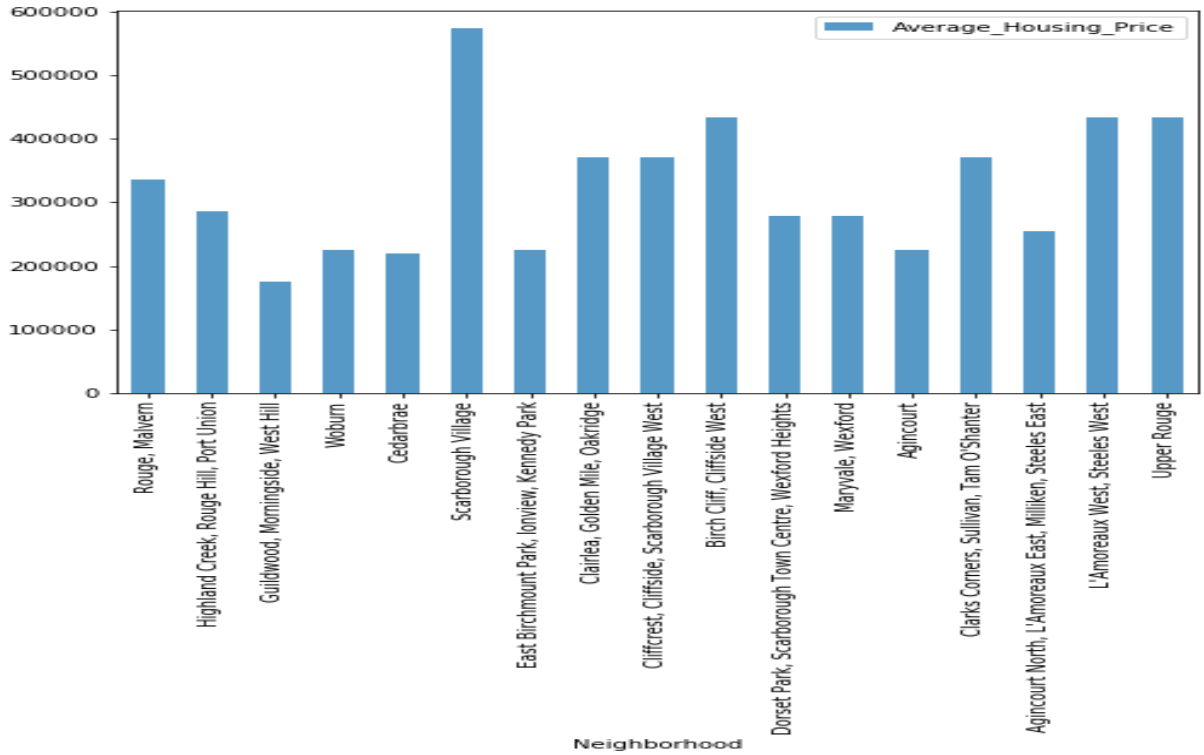
Clustering Approach

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods 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.

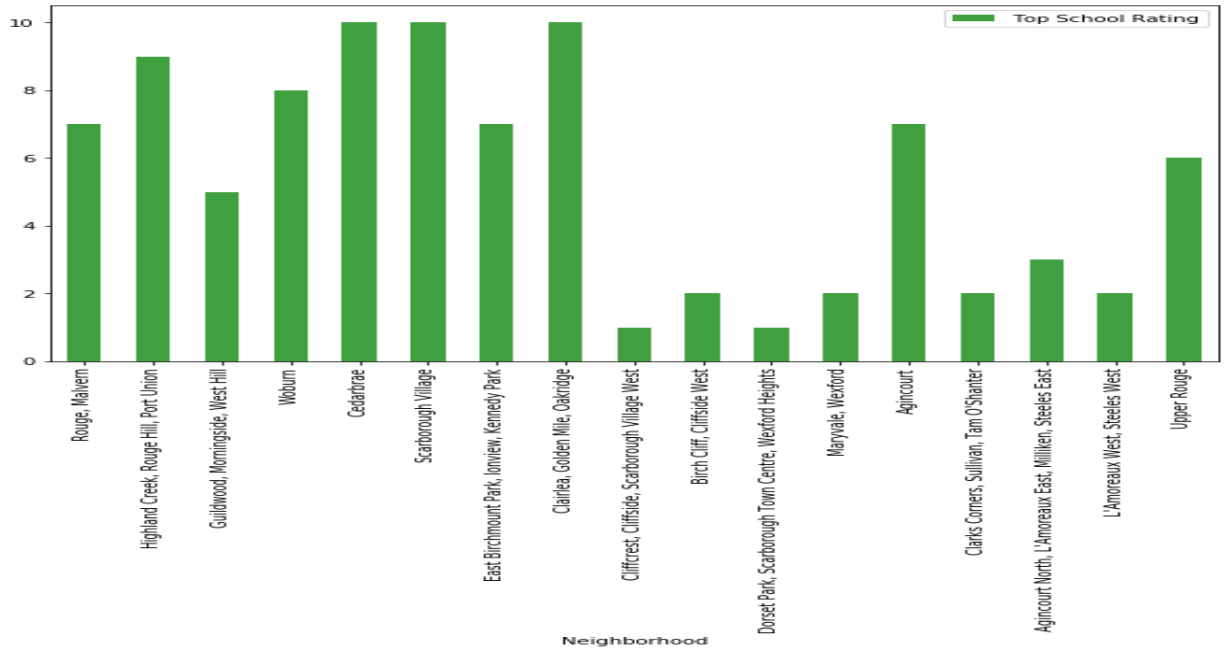
Neighborhood Most Common Venues

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
Agincourt	Shopping Mall	Chinese Restaurant	Supermarket	Pool	Breakfast Spot	Malay Restaurant	Mediterranean Restaurant	Hong Kong Restaurant
Agincourt North, L'Amoreaux East, Milliken, St...	Coffee Shop	Pharmacy	Sandwich Place	Zoo Exhibit	Construction & Landscaping	Convenience Store	Deli / Bodega	Department Store
Birch Cliff, Cliffside West	Park	Gym	Gym Pool	General Entertainment	Café	Skating Rink	College Stadium	Discount Store
Cedarbrae	Flower Shop	Athletics & Sports	Bakery	Thai Restaurant	Bank	Hakka Restaurant	Caribbean Restaurant	Indian Restaurant
Clairlea, Golden Mile, Oakridge	Coffee Shop	Bus Line	Diner	General Entertainment	Ice Cream Shop	Intersection	Metro Station	Convenience Store

Neighborhood Median Housing Prices



Neighborhood School Ratings



Libraries:

***Pandas**:* For creating and manipulating dataframes

***Folium**:* Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

***Scikit Learn**:* For importing k-means clustering

***JSON**:* Library to handle JSON files

***Geopy**:* To retrieve Location Data

***Requests**:* Library to handle http requests

***Matplotlib**:* Python Plotting Module

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