

Capstone Project – Best neighbourhood in Toronto for a migrant

Introduction

Last decade has seen an increase of migrants to Canada so becoming increasingly difficult the choice of buying a house in the country especially in big city as Toronto.

The migrants search both for a good housing prices and good rating schools for their children.

The aim of the project is to analyse several features for a neighbourhood such as school ratings, crime rates, median house prices, and compare them analysing with other neighbourhoods.

This study could support people to get knowledge about new places before moving to new country for working and life reasons.

The criteria selected to define a “good” neighbourhood, which is one with a tangible commercial presence inside a given community, are:

1. Compare median housing prices;
2. Compare school ratings.

The neighbourhood chose for the project is Scarborough in the Greater Toronto Area. Scarborough appears to be one of most popular destination to reside for new migrants in Canada, being a multicultural area and home of several religious groups.

The data of the project have been supported by **Foursquare** which is the most trusted, independent location data platform for understanding how people move through the real world. Especially their places API which provides the ability to perform location search, location sharing and details about a business.

Due to “*http*” request limitations the number of places per neighbourhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

As clustering approach, to compare the affinities of two cities, we choose to explore, segment and group neighbourhoods into clusters to find comparable neighbourhoods in important cities ad New York and Toronto. At this stage “*k-means clustering algorithm*” has been used.

Data

We need geo-locational information about that specific district and the neighbourhoods in that district. The borough analysed is "Scarborough" in Toronto.

This project require information of distinct neighbourhoods in Toronto, school ratings and median house prices. The data required are:

1. Neighbourhood 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. The location of Scarborough would be filtered using the same:

https://github.com/DDK79/Coursera_Capstone/blob/master/Segmenting%20and%20Clustering%20Neighborhoods%20in%20Toronto%20-%20Sections%201%2C2%2C3.ipynb

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

Furthermore, we need data about several venues in different neighbourhoods of that specific borough. In order to get that information we use “*Foursquare*” locational information. Foursquare is the most trusted, independent location data platform for understanding how people move through the real world. Especially their places API which provides the ability to perform location search, location sharing, venue names, details about a business, menus and photos.

Foursquare platform is the sole data source used because all information required can be acquired by its API. After finding the list of neighbourhoods, we then connect to the Foursquare API to gather information about venues within each neighbourhood. The radius of 100 metres has been chosen for each neighbourhood.

The data recovered from Foursquare has included information of venues within a specified distance of the longitude and latitude of the postcodes.

Methodology

Steps taken were:

1. Data acquisition and cleaning;
2. Data preparation;
3. Feature Selection;
4. Clustering.

Data acquisition and cleaning was a two-step process:

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

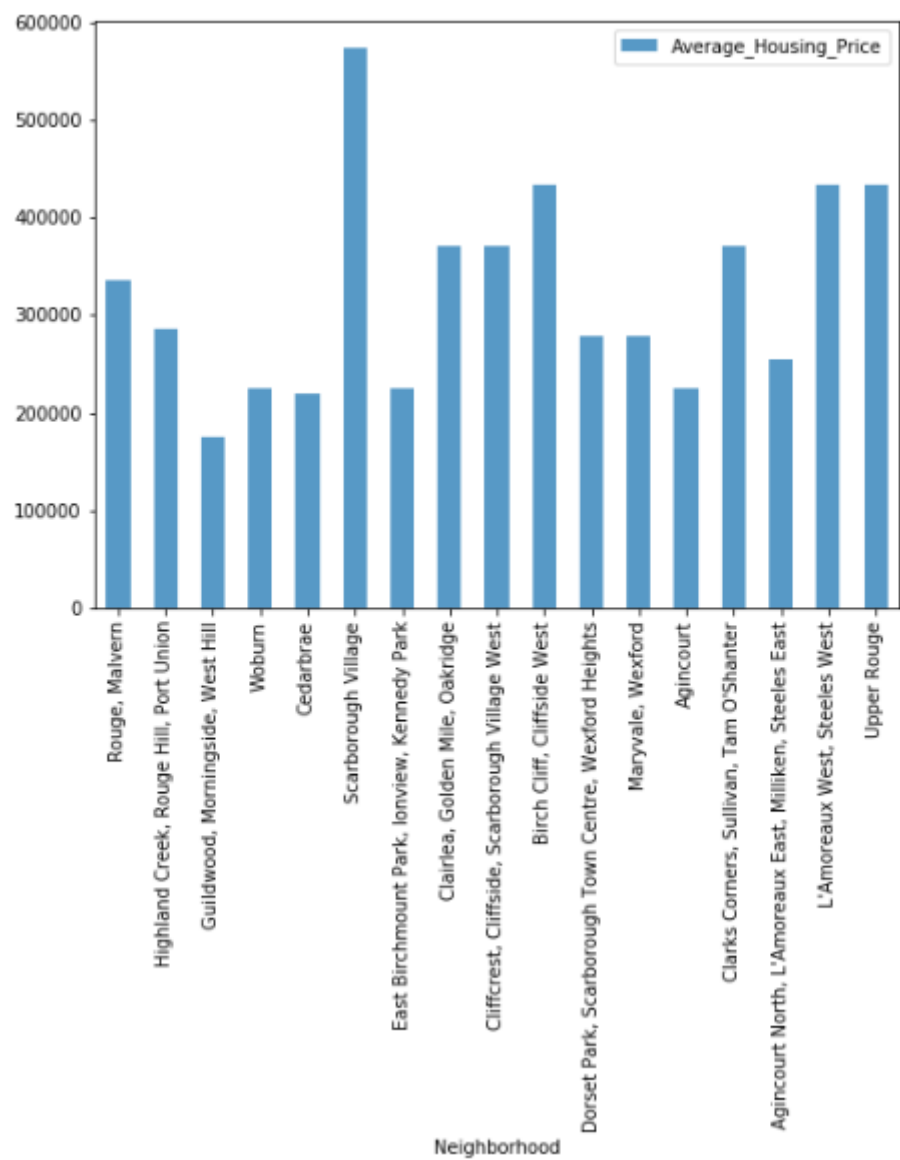
Clustering Approach:

As clustering approach, to compare the affinities of two cities, we choose to explore, segment and group neighbourhoods into clusters to find comparable neighbourhoods in important cities ad New York and Toronto. At this stage “*k-means clustering algorithm*” has been used.

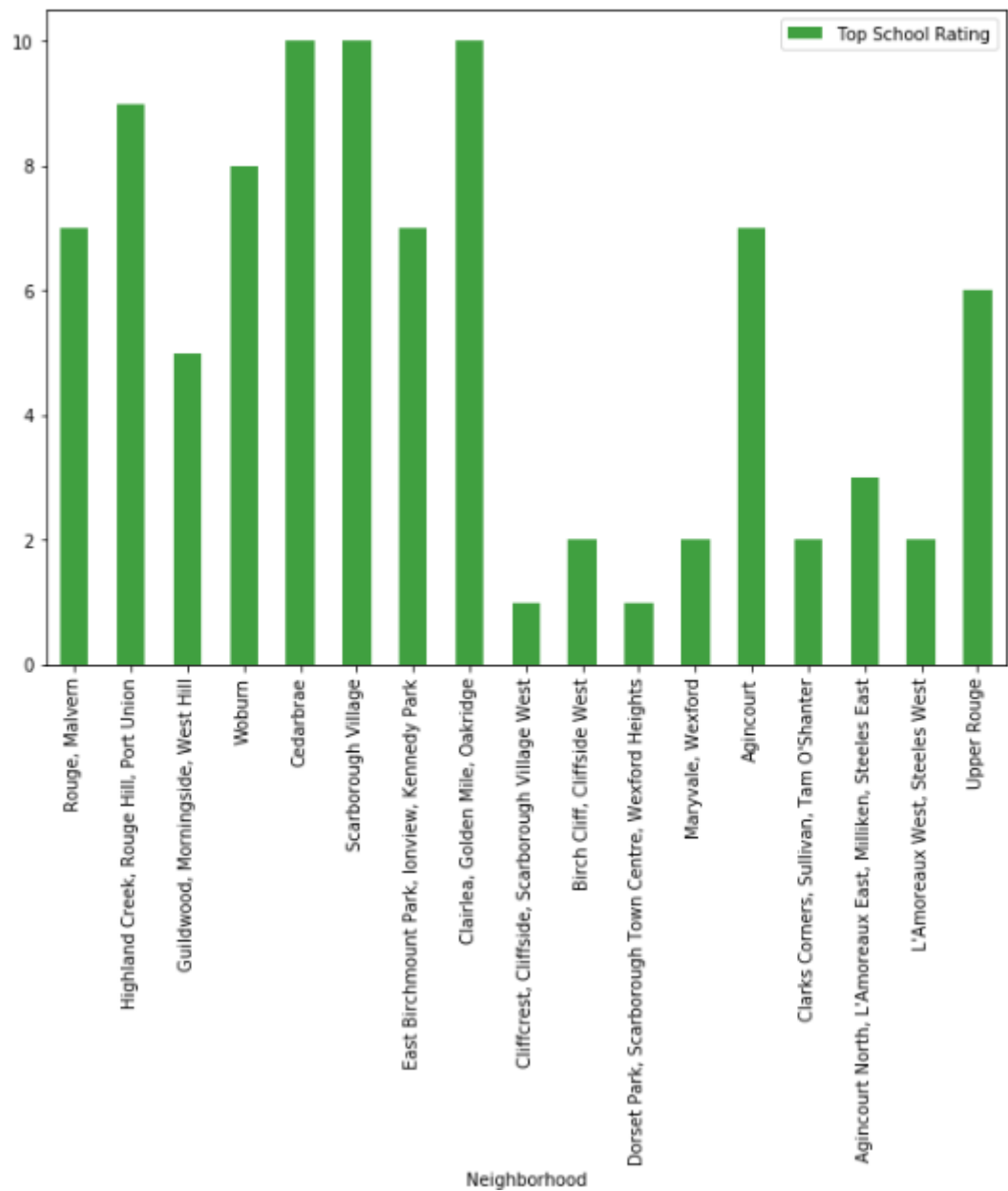
The figure below shows the “*Neighbourhood Most Common Venues*”:

	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	9th Most Common Venue	10th Most Common Venue
0	Agincourt	Badminton Court	Pool Hall	Breakfast Spot	Fabric Shop	Coffee Shop	Clothing Store	Latin American Restaurant	Lounge	Motorcycle Shop	Skating Rink
1	Alderwood, Long Branch	Pizza Place	Convenience Store	Pool	Pharmacy	Sandwich Place	Coffee Shop	Pub	Gym	Gas Station	Skating Rink
2	Bathurst Manor, Wilson Heights, Downsview North	Bank	Coffee Shop	Sandwich Place	Diner	Fried Chicken Joint	Frozen Yogurt Shop	Deli / Bodega	Sushi Restaurant	Gas Station	Restaurant
3	Bayview Village	Bank	Grocery Store	Skating Rink	Japanese Restaurant	Chinese Restaurant	Café	Event Space	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant
4	Bedford Park, Lawrence Manor East	Coffee Shop	Italian Restaurant	Sandwich Place	Restaurant	Bank	Bakery	Bagel Shop	Thai Restaurant	Café	Butcher

The figure below shows the “*Neighbourhood Median Housing Prices*”:



The figure below shows the “*Neighbourhood school ratings*”:



Conclusion

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

