

Examining median house prices in Scarborough, Canada

Battle of Neighborhoods - Applied Data Science Capstone

October 14, 2019

1) Introduction

1.1) The location: Being a very popular destination, Scarborough is place of choice among immigrants. It is home to many religious affinities and many places of worship. It also includes a number of natural landmarks.

1.2) Problem description: Each year many people migrate various states of Canada, and usually a new immigrant is faced with the problem of searching for a new dwelling and with reasonable prices. If the immigrant is accompanied by the family, naturally they would want to look for the best possible schools in the neighborhood in addition to searching for recreational and trending areas in the neighborhood. The project aims to develop easiness for the new mover by creating an analysis of features for a neighborhood as a comparative analysis between neighborhoods. It would help people make an informed decision before moving to a new country, state, city.

The project aims to facilitate the choice of neighborhood in Scarborough out of many neighborhoods based on the examination of various facilities available in and around the area.

2) Data:

2.1) Data source: The project will use location data provided by Foursquare API, more specifically the places API which grants the user the ability to perform location search, location sharing and provides details about businesses. To facilitate comparison between schools in each neighborhood in Scarborough, the data was obtained from www.greatschools.org.

I will subset the data from previous week (Week 2) to make it Scarborough specific, and all the neighborhoods within Scarborough with all the location (Latitude and Longitude)

coordinates. A snippet of the data is provided below.

	Postalcode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.811525	-79.195517
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.785665	-79.158725
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.765815	-79.175193
3	M1G	Scarborough	Woburn	43.768369	-79.217590
4	M1H	Scarborough	Cedarbrae	43.769688	-79.239440
5	M1J	Scarborough	Scarborough Village	43.743125	-79.231750
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.726276	-79.263625
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.713054	-79.285055
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.724235	-79.227925
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.696770	-79.259967
10	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...	43.759975	-79.268974
11	M1R	Scarborough	Maryvale, Wexford	43.750710	-79.300560
12	M1S	Scarborough	Agincourt	43.793940	-79.267976
13	M1T	Scarborough	Clarks Corners, Sullivan, Tam O'Shanter	43.784725	-79.299066
14	M1V	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...	43.817685	-79.280187
15	M1W	Scarborough	L'Amoreaux West	43.800883	-79.320740
16	M1X	Scarborough	Upper Rouge	43.834215	-79.216701

2.2) Data Cleaning: Data wrangling has been conducted on the dataset already because of Peer-graded assignment in Week 3. Data scraped from the Wikipedia Page was scraped using Python Library BeautifulSoup4. From the raw data all the relevant data was stored in a Pandas data frame and using another Python library called geocoders the data frame appended with location coordinates for each Borough and their respective neighbourhoods.

3) Methodology

To compare the similarities of two cities, I decided explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods. To be able to do that, I used unsupervised machine learning technique called K-means clustering.

But before I could apply the technique the data was transformed so as to facilitate the process of algorithm application and to achieve the required objective. For each neighborhood within Scarborough the nearby Venues were searched and through the analysis there came to be 151 unique categories of venues e.g. Korean Restaurants, Coffee shops, Gyms, and Supermarkets.

One-hot encoding was applied for all venues in each neighborhood as shown below to facilitate the application of machine learning algorithm

	Zoo Exhibit	American Restaurant	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Automotive Shop	BBQ Joint	Badminton Court	Bakery	Bank	Bar	Baseball Field	Beach	Beer Store	Big Box Store	Bistro	Bookstore	Bowling Alley
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Further exploratory analysis was conducted in the form of 5 most common venues based on each neighborhood as shown in the snippet below

----Agincourt----

```

      venue  freq
0  Chinese Restaurant  0.12
1  Indian Restaurant  0.06
2  Caribbean Restaurant  0.04
3  Gym / Fitness Center  0.04
4           Bakery  0.04

```

----Agincourt North, L'Amoreaux East, Milliken, Steeles East----

```

      venue  freq
0  Chinese Restaurant  0.17
1    Coffee Shop  0.06
2  Korean Restaurant  0.05
3    Dessert Shop  0.04
4    Pizza Place  0.04

```

----Birch Cliff, Cliffside West----

```

      venue  freq
0      Park  0.14
1  Coffee Shop  0.10
2  Sandwich Place  0.05
3  Grocery Store  0.05
4    Pizza Place  0.05

```

----Cedarbrae----

```

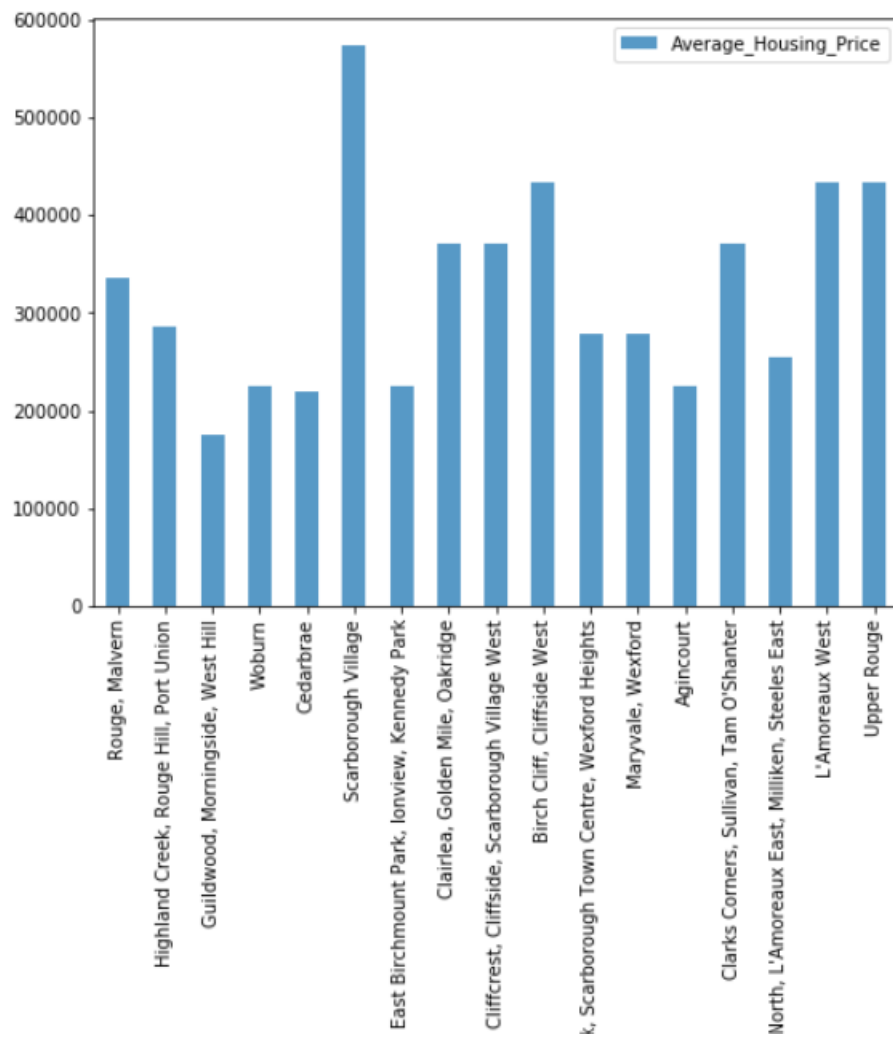
      venue  freq
0    Coffee Shop  0.11
1  Clothing Store  0.09
2  Fast Food Restaurant  0.05

```

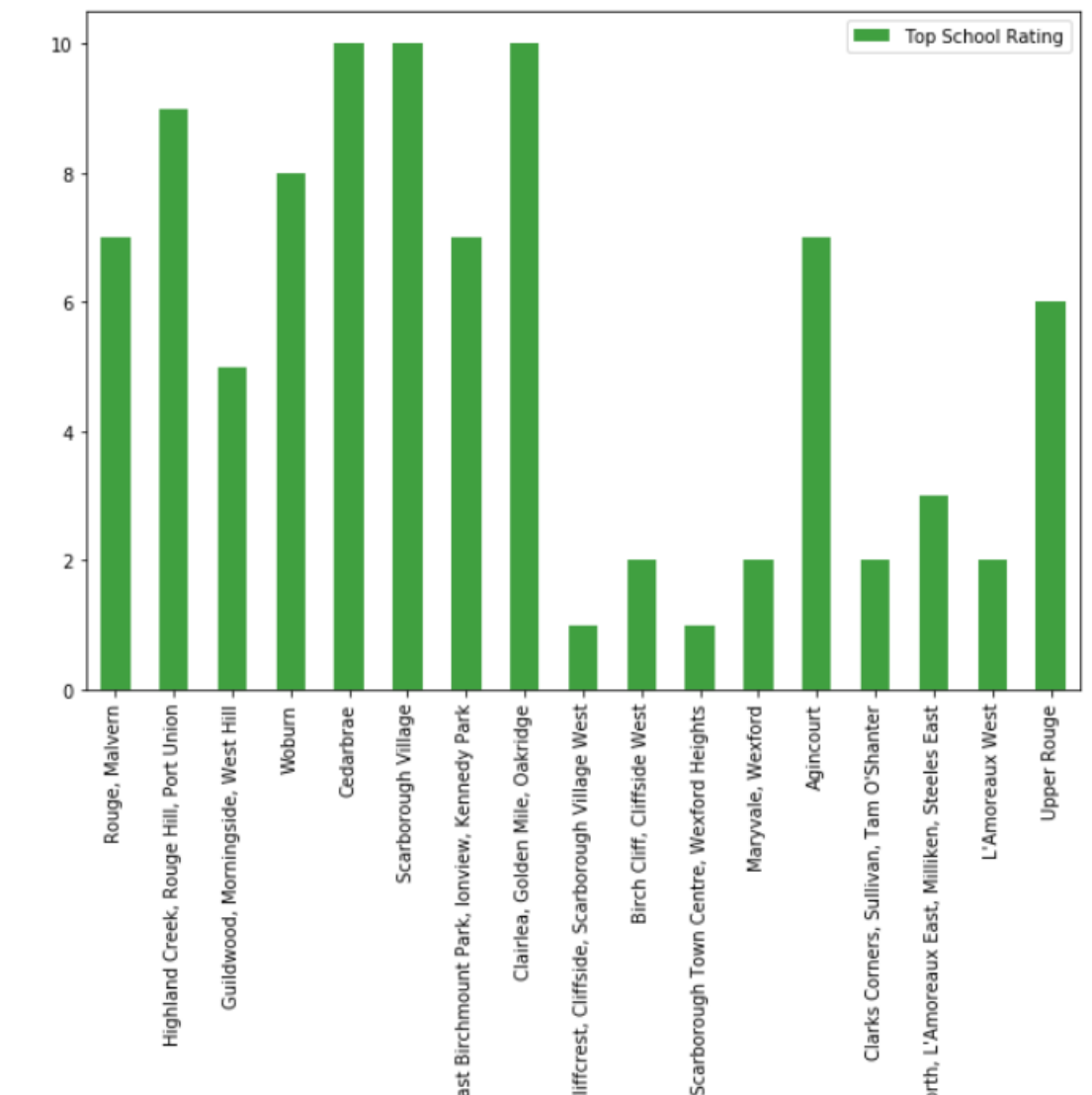
After that K-means algorithm was applied, the algorithm was initialized with 4 initial clusters, and the labels as the output of algorithm were merged with the original dataset which included data about the postal code, neighborhood, borough, location coordinates and the ten most common venues as shown below,

The entire motive of the project was to facilitate the moving decision of the user. For that I compared the average housing price in each neighborhood in Scarborough. The data for

the prices were searched online and then fit into a list. Attached is the screenshot



To further facilitate the moving decision, top school ratings were compared across each neighborhood in Scarborough. Attached is screenshot below,



Conclusion

In this project, through K-means clustering algorithm neighborhoods were separated into 4 clusters based on similar features. Using the graphs above a user can decide if they want to move into Scarborough. The graphs include average house prices, and average rating of schools based on each neighborhood and other data includes most common venues in the 2 km radius. The highest school ratings were found to be in three neighborhoods. And the housing prices were on average the lowest in Guildwood, Morningside, West Hill.