Capstone Project - The Battle of Neighborhoods

Introduction

About the Project

I think people who are deciding to live in a specific area of Canada would like to have access to information like housing prices and school ratings. This projects aim to conduct a comparative analysis between neighborhoods. The features include house price and school ratings, weather conditions, crime rates and recreational facilities, which would help people to know more about the places before making the movement.

How to compare

In this project, the rank of neighborhoods is given by

1. Housing prices
2. School ratings

More specifically, the rank will be higher with higher housing prices and school ratings.

After we make the comparison, we can confidently recommend the best neighborhood for our clients or friends.

Data

Foursquare location data

We will need data about different venues in different neighborhoods. In order to get that information, we will use "Foursquare" for the locational information. Foursquare is a famous location data provider with various kinds of information of the venues and nearby. For example, it will provide venue names, locations, photos, and etc. As a result, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

In the first place, we will get the list of neighborhoods, we then use the Foursquare API to fetch information about venues in each neighborhood.

Specifically, the following information will be gathered.

- Neighborhood's name
- Neighborhood's Latitude and Longitude
- Venue's name
- Venue's Latitude and Longitude

Longitude and Latitude Data

We need geo-locational information about that specific borough and the neighborhoods in that borough. It is "Scarborough" in Toronto. The neighborhood data required will be:

1. Latitude and longitude for the location
2. School Ratings
3. House Prices (Median)

The latitude and longitude data can be gathered using the previous project. The location data of Scarborough would be like:

Postalcode	Borough	Neighborhood	Latitude	Longitude	
0	M1B	Scarborough	Rouge, Malvern	43.811650	-79.195561
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	43.785605	-79.158701
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.765690	-79.175299
3	M1G	Scarborough	Woburn	43.768216	-79.217610
4	M1H	Scarborough	Cedarbrae	43.769608	-79.239440
5	M1J	Scarborough	Scarborough Village	43.743085	-79.232172
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.726260	-79.263670
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.713213	-79.284910
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.723575	-79.234976
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.696690	-79.260069

Methodology

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
    1. Obtaining the post codes for neighborhoods
    2. Obtaining venues within the neighborhoods
2. Data cleaning
3. Feature selection
4. K-means clustering

To compare the similarities of two cities, we decided to explore neighborhoods, and group them into clusters, in which way we can find similarities between the two cities. Meanwhile, we use the unsupervised k-means clustering algorithm to make the clustering:

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	
											Bubble

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

Figure 1: Neighborhood Median Housing Prices

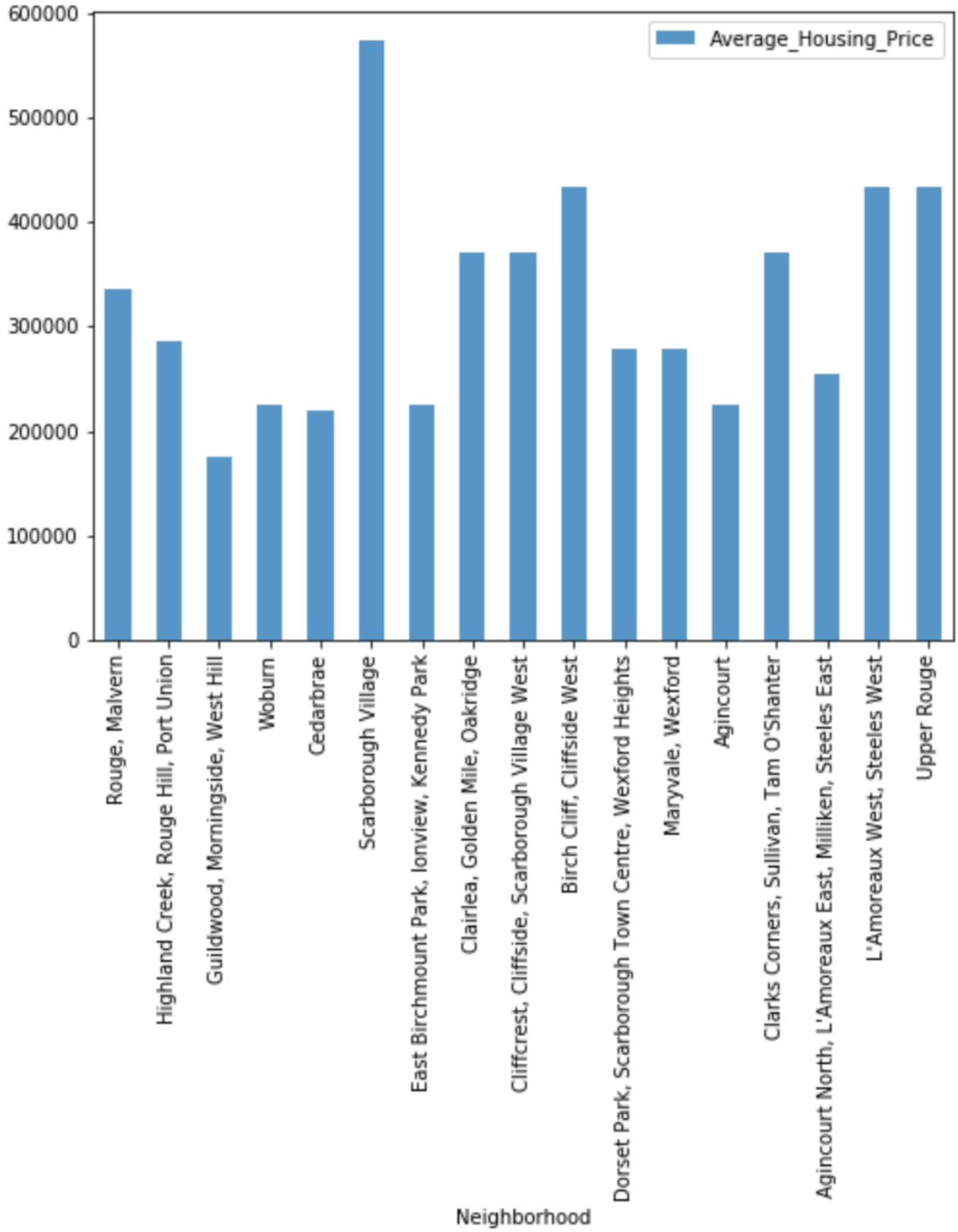
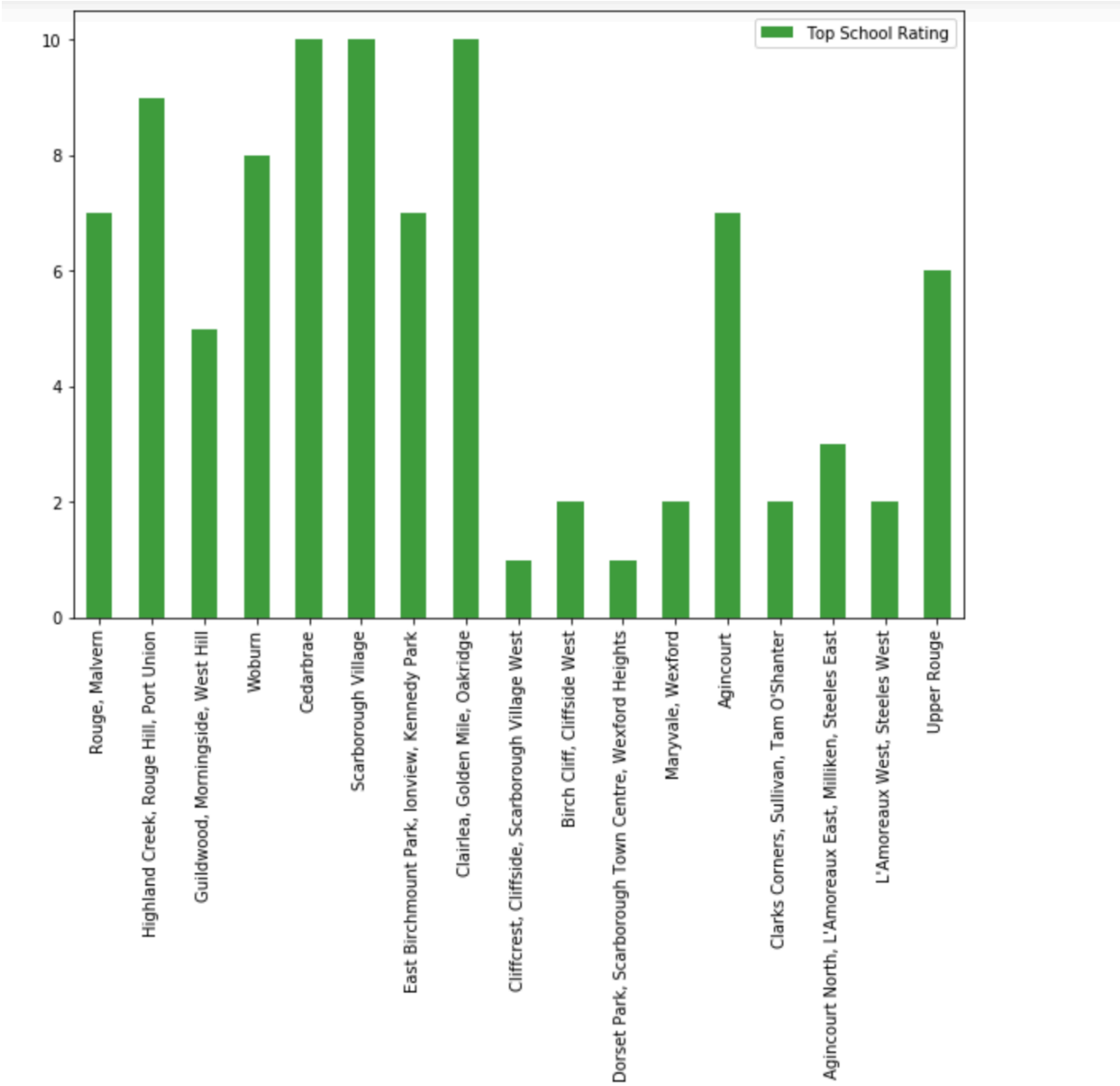
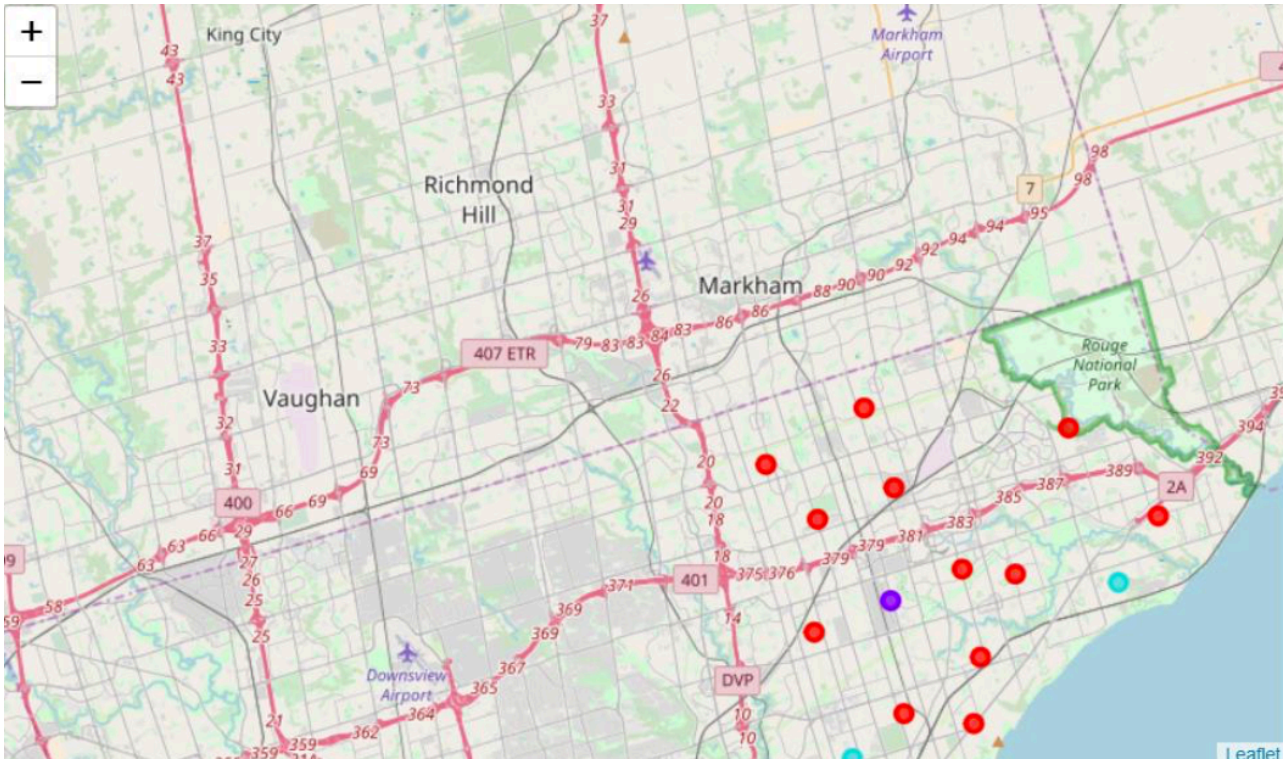


Figure 2: Neighborhood School Ratings



Conclusion

In this project, we use K-means cluster algorithm to separate the neighborhood into 3 clusters. Based on the clustering result, we compare them with the average house prices and school rating.


King City
Markham Airport
Richmond Hill
Markham
Vaughan
407 ETR
Rouge National Park
400
401
DVP
7
2A
Downsview Airport
Leaflet