The Battle of Neighborhoods: A Coursera Capstone Project

Where in Scarborough Should I Live? A Template Analysis of Median Housing Prices and School Ratings

Introduction

Local "neighborhood ranking" websites (ex: WalkScore, Neighborhood Scout, and AreaVibes) are very popular especially with people who are moving to an area they've never visited. To generate the neighborhood data for the Greater Toronto Area, many different parameters will need to be investigated for each city. In the project, I will analyze the city of Scarborough, Toronto, Canada as a template for future analysis of other cities.

This process can then be replicated, and the data presented and used by neighborhood ranking websites to give succinct information about a neighborhood of interest.

Neighborhood ranking websites often have premium membership services for clients interested in new neighborhoods. Creating a reproducible template for analyzing neighborhood data can eventually be used to expand the scope and clientele purchasing premium memberships.

Data

Data of interest to categorize each neighborhood in terms of "livability" include:

- Median housing price
- School rankings
- Crime rates
- Weather
- Recreational facilities

Scarborough's neighborhoods will be analyzed with these parameters to determine its livability score.

Neighborhood data is sourced from the Wikipedia page for Toronto, Canada. More information about data acquisition and cleaning can be found in the GitHub notebook entitled "Segmenting and Clustering Neighborhoods in Toronto," linked in the Jupyter notebook for this project.

Data for the Scarborough area is from the FourSquare API. The FourSquare API is a robust database used by many developers to scrape location data.

Methodology

FourSquare API credentials were used to mine information about places in the neighborhood of Scarborough. Due to free API request limitations, the number of places per neighborhood was set to 100 and the radius was set to 500.

The data analysis process consisted of five steps, outlined below:

1. Data acquisition and cleaning

2. Data preparation
3. Feature selection
4. Clustering

Data Acquisition and Cleaning

Data acquisition and cleaning consisted of two steps: obtaining postcodes for neighborhoods in Toronto and obtaining venues within those neighborhoods.

Information about data preparation and feature selection can be found in the attached raw Jupyter notebook where the data was analyzed. Features selected were the following:

1. Median housing price
2. School ratings

Clustering

The k-means clustering approach was used which is a form of unsupervised machine learning. This approach was chosen in order to group the neighborhoods of Scarborough into clusters. This approach is expandable and can be used with neighborhoods of very large cities such as Chicago and New York City. This makes it an ideal approach for a template analysis.

Results

Figure 1: Most Common Venues in Each Neighborhood of Scarborough (Note: full table can be found in report)

	Postcode	Neighborhood	Latitude	Longitude	Cluster Labels	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	M1B	Rouge, Malvern	43.811525	-79.195517	1	Zoo Exhibit	Construction & Landscaping	Fast Food Restaurant	Farmers Market	Department Store	Chinese Restaurant	Coffee Shop	College Stadium	Convenience Store	Deli / Bodega
1	M1C	Highland Creek, Rouge Hill, Port Union	43.785665	-79.158725	1	Moving Target	Bar	Zoo Exhibit	Department Store	Chinese Restaurant	Coffee Shop	College Stadium	Construction & Landscaping	Convenience Store	Deli / Bodega
2	M1E	Guildwood, Morningside, West Hill	43.765815	-79.175193	0	Park	Gym / Fitness Center	Athletics & Sports	Construction & Landscaping	Zoo Exhibit	Diner	Coffee Shop	College Stadium	Convenience Store	Deli / Bodega
3	M1G	Woburn	43.768369	-79.217590	1	Coffee Shop	Business Service	Park	Fast Food Restaurant	Construction & Landscaping	Zoo Exhibit	Deli / Bodega	Chinese Restaurant	College Stadium	Convenience Store
4	M1H	Cedarbrae	43.769688	-79.239440	1	Indian Restaurant	Gas Station	Athletics & Sports	Hakka Restaurant	Thai Restaurant	Bakery	Bank	Flower Shop	Caribbean Restaurant	Department Store
5	M1J	Scarborough Village	43.743125	-79.231750	1	Fast Food Restaurant	Sandwich Place	Big Box Store	Women's Store	Train Station	Indian Restaurant	Restaurant	Deli / Bodega	Chinese Restaurant	Coffee Shop
6	M1K	East Birchmount Park, Ionview, Kennedy Park	43.726276	-79.263625	1	Department Store	Coffee Shop	Bus Station	Convenience Store	Discount Store	Zoo Exhibit	Chinese Restaurant	College Stadium	Construction & Landscaping	Deli / Bodega
7	M1L	Clairlea, Golden Mile, Oakridge	43.713054	-79.285055	0	Intersection	Bus Line	Diner	Bakery	Gym	Fast Food Restaurant	Convenience Store	Metro Station	Coffee Shop	Park
8	M1M	Cliffcrest, Cliffside, Scarborough Village West	43.724235	-79.227925	1	Fast Food Restaurant	Pizza Place	Bank	Coffee Shop	Sandwich Place	Park	Discount Store	Bistro	Pharmacy	Farmers Market
9	M1N	Birch Cliff, Cliffside West	43.696770	-79.259967	0	Park	Gym	Baseball Field	Gym Pool	College Stadium	Skating Rink	Café	General Entertainment	Chinese Restaurant	Gas Station
10	M1P	Dorset Park, Scarborough Town Centre, Wexford ...	43.759975	-79.268974	2	Wine Shop	Park	Coffee Shop	Bakery	Zoo Exhibit	Department Store	Chinese Restaurant	College Stadium	Construction & Landscaping	Convenience Store
11	M1R	Maryvale, Wexford	43.750710	-79.300560	1	Hookah Bar	African Restaurant	Gas Station	Pizza Place	Athletics & Sports	Auto Garage	Coffee Shop	College Stadium	Construction & Landscaping	Convenience Store

Figure 2: Median Housing Prices in Each Neighborhood

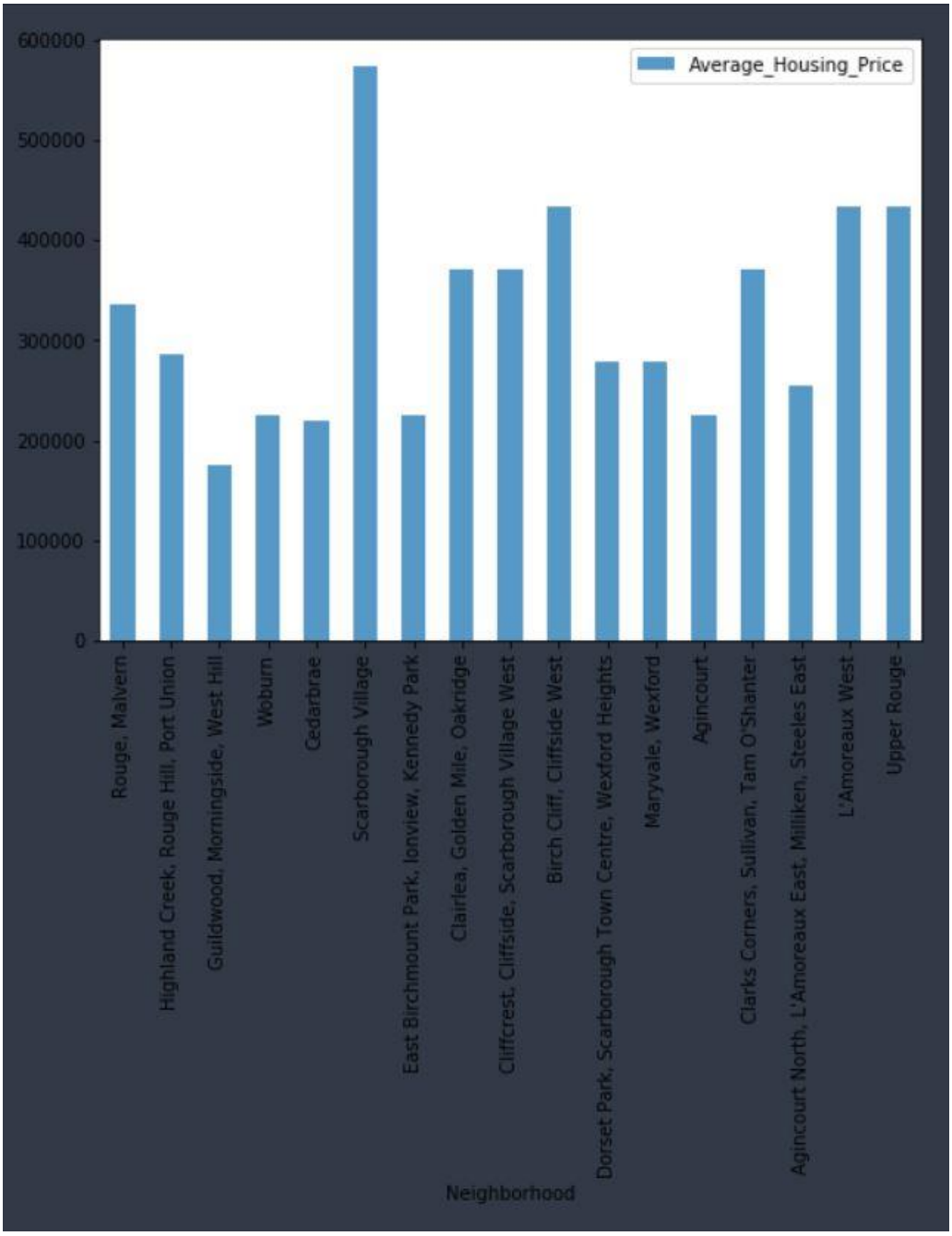


Figure 3: School Ratings of Each Neighborhood

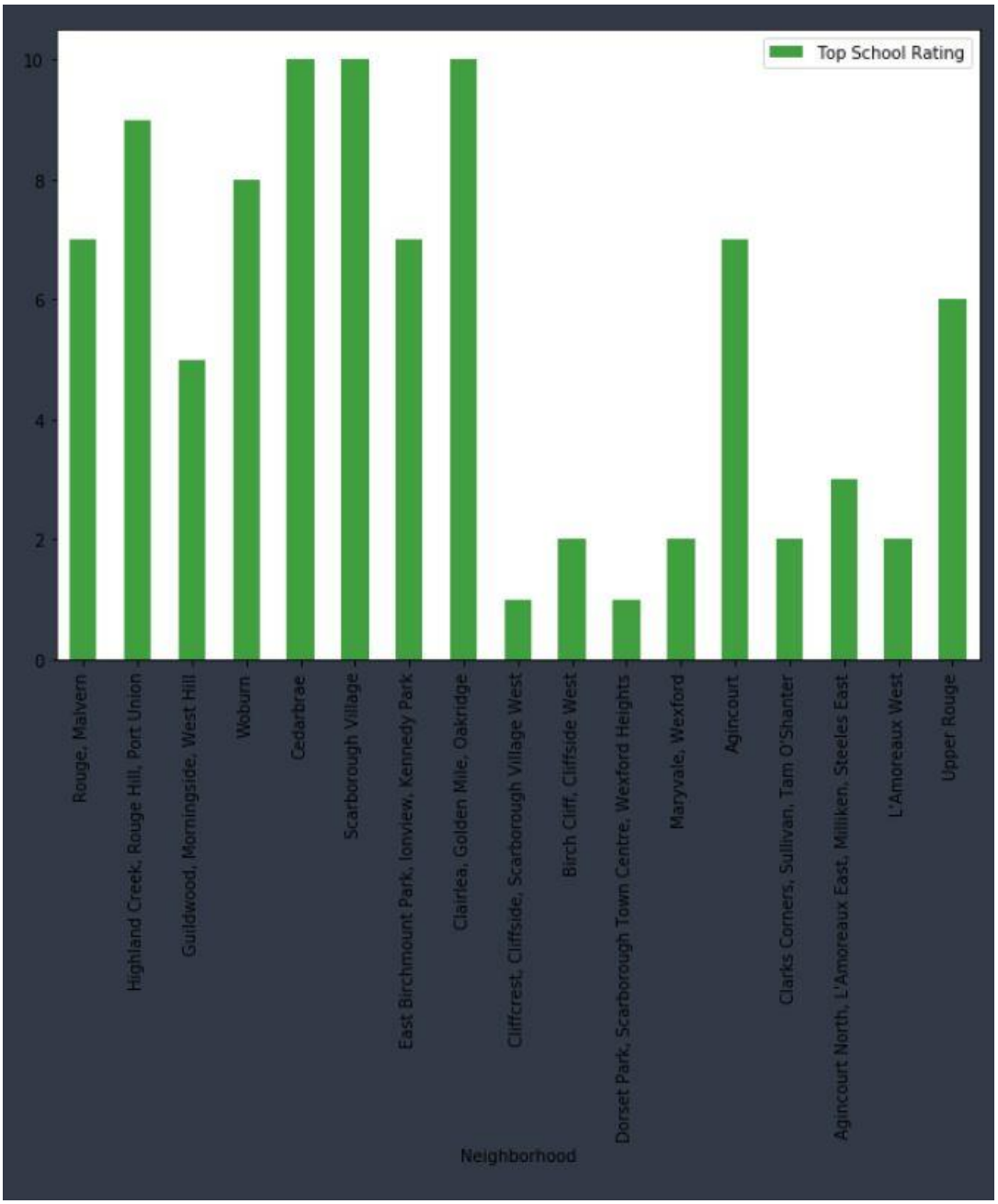
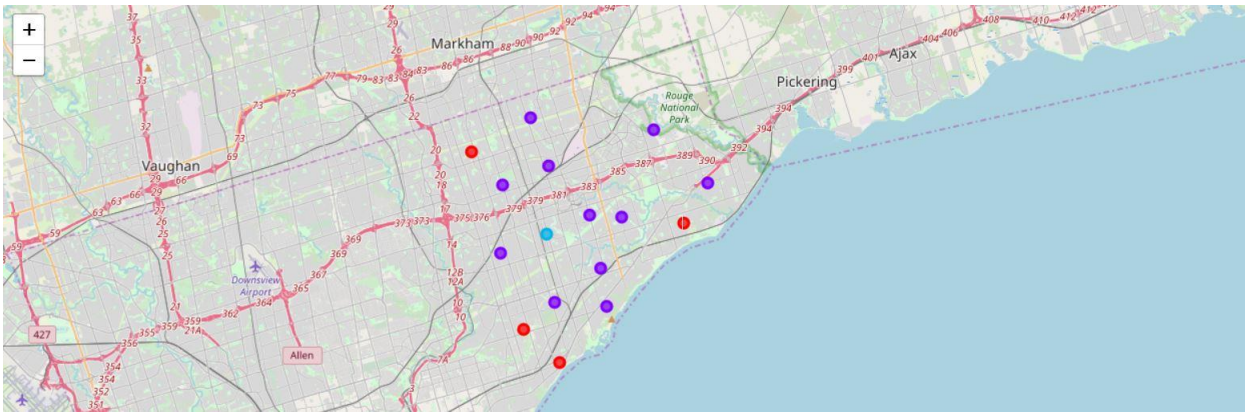


Figure 4: Cluster Map



Discussion

Depending on the client's interests, a lot of information can be used from this data analysis. For example, a client search engine could be created to tag interests such as children's activities, fine dining, bars, shopping, etc. including very niche interests such as hookah bars. A user interested in hookah bars, for instance, would be pointed to the neighborhood of Maryvale, Wexford, where hookah bars are the most common venue. A client with children may be interested in the neighborhood of Rouge, Malvern, where zoo exhibits are the most common venue.

The data can be merged together to create a comprehensive picture of each neighborhood. An area with low median housing prices and high school scores might be most interesting to a client with children. Childfree clients can easily filter out school ratings and have them not be considered in their results.

An interesting result is that high school ratings do not necessarily correlate with median housing prices, which is the case in many cities in the United States. This opens a lot of opportunities for parents who want to spend less on a house but live in an area with great schools.

I recommend continuing and expanding this analysis to include more "livability" factors stated in the introduction. Crime rates, for example, is probably a variable that most clients would be interested in.

Conclusion

Clustering neighborhoods based on their features is an extremely useful tool with strong business applications for real estate or neighborhood ranking websites. Expanding this analysis to include more features and cities could create a robust database that clients can use to decide where to move based on specific parameters.