Presentation for IBM Capstone

Battle of the Neighbourhoods in Scarborough, Toronto

By Alex Turner

Introduction

The purpose of this project is to use the skills we have learnt on this course to look at the various neighbourhoods in Scarborough, Toronto to see if there is a way we can look at the data available to make more informed choices about where to move to. Hopefully using these skills in this way will mean that I can use these skills to make more informed choices about where to move to in my own Country.

I chose to use Scarborough in Canada due to the high amount of immigration that takes place in this area. Not only does this mean that a lot of the data collection for what we want to look at has already been done in this area, but it also means that more people could be influenced by the work I will cover in this final project. This project will focus primarily on housing prices and school ratings as influencing factors, however we will also look at local amenities and venues as a secondary influencing factor.

Data Selection

For this project I have gathered Postal Code, Neighbourhood and Borough information from this link:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

I will also be using School Rating Data provided by https://www.greatschools.org/

Data Selection - Foursquare API Data

We will be using data provided by the application/website Foursquare. Foursquare is a location data provider that helps to power popular tools such as Google Maps. The location data provided by this will include lots of useful information for us such as venue names, locations, category and if we wanted - menus and photos. As this provides such as wealth of information we will be using the Foursquare API data as our sole source of data for the location based data outside that already mentioned above. For each neighbourhood we have chosen a radius of 100 metres to look around in the Foursquare API to provide venue information so as to not have repeated information too much or an overwhelming amount of information. We want to get a feel for what it's like when stepping out of the door if you were to live here, which I feel a 100 metre radius will provide. The information we will be getting for each venue will be Neighbourhood, Neighbourhood Latitude and Longitude, Venue, Venue Name, Venue Latitude and Longitude as well as Category.

Clustering Approach
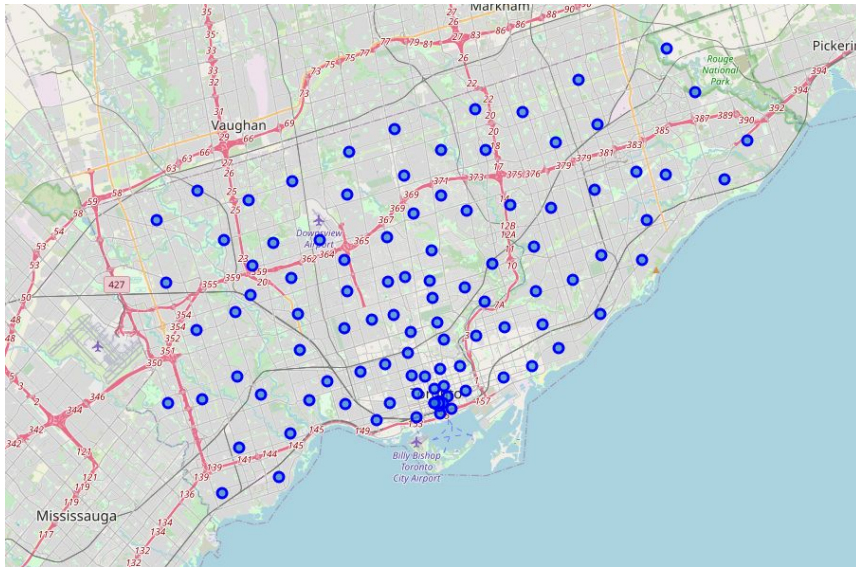
In the figure on the left we can see the 103 neighbourhoods plotted onto Scarborough, if we were to look at venue information for each of these neighbourhoods - it would take a long time to interpret the results!

Clustering Approach
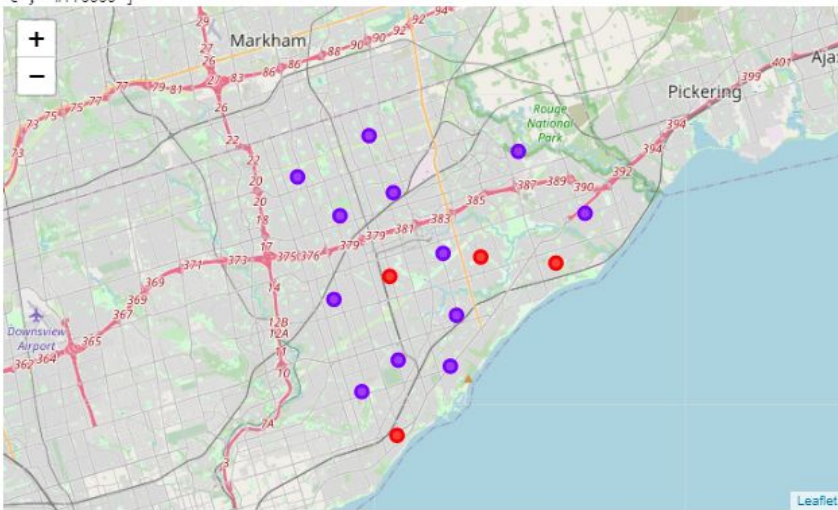
So instead, in order to effectively compare cities or areas of cities we broke our dataset up into neighbourhoods, segmented them and grouped them into clusters so that we can find similar/desirable neighbourhoods. Clustering data in this manner is a form of unsupervised machine learning known as k-means clustering algorithm.
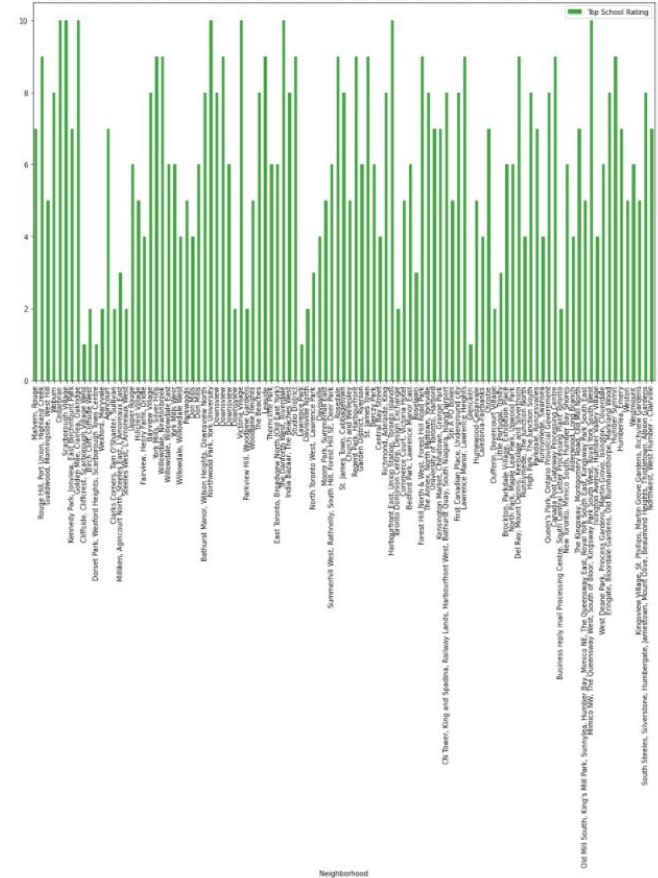
Clustering Approach - Nearby Venues

	Postalcode	Borough	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	Scarborough	Malvern, Rouge	43.81139	-79.19662	1	Zoo Exhibit	Paintball Field	Fast Food Restaurant	Event Space	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Escape Room	Ethiopian Restaurant
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.78574	-79.15875	1	Bar	Construction & Landscaping	Fish & Chips Shop	Falafel Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Escape Room	Ethiopian Restaurant	Event Space
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.76575	-79.17470	0	Park	Gym / Fitness Center	Athletics & Sports	Gymnastics Gym	Yoga Studio	Doner Restaurant	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School
3	M1G	Scarborough	Woburn	43.76812	-79.21761	0	Coffee Shop	Chinese Restaurant	Park	Fast Food Restaurant	Event Space	Dumpling Restaurant	Eastern European Restaurant	Electronics Store	Elementary School	Escape Room
4	M1H	Scarborough	Cedarbrae	43.76944	-79.23892	1	Thai Restaurant	Bakery	Caribbean Restaurant	Gas Station	Athletics & Sports	Hakka Restaurant	Bank	Playground	Fish & Chips Shop	Dumpling Restaurant

The table above shows the clustered neighbourhoods and their most common venues within each cluster. We can use this to give an idea of what each neighbourhood "feels like" when making our own personal comparisons to where we would want to live.

Housing Prices for each Neighbourhood

We were able to make a graph out of the average housing prices for each neighbourhood so that we can consider this when looking at where we want to move. While many would assume the cheapest areas would be the most desirable, many might consider more affluent areas if they had lower crime rates/better amenities as they may be more of an investment when they go to sell their property in the future. It is difficult to read the neighbourhoods on this slide, so if you want a closer look please look at my final report where you can see the graph easier:

https://github.com/Shwoople/Coursera_Capstone/blob/main/Week%205/Final%20Report%20for%20IBM%20Capstone%20Project%20-%20Battle%20of%20the%20Neighbourhoods%20in%20Scarborough%2C%20Toronto.pdf

School Rating for each Neighbourhood

On the right we also plotted a graph of the average school rating for the nearest school in each neighbourhood. While this may not be a big factor for those without children in the age range to attend school, this would be a huge factor to consider for families with children still in education. Similar to the previous slide, a better view of the graph can be found on my final report:

https://github.com/Shwoople/Coursera_Capstone/blob/main/Week%205/Final%20Report%20for%20IBM%20Capstone%20Project%20-%20Battle%20of%20the%20Neighbourhoods%20in%20Scarborough%2C%20Toronto.pdf

Discussion

The purpose for this project was to gather information in a digestible manner so that people can make more informed decisions around what neighbourhood may be most suitable for them. The combination of the figures and table shown in the previous slides can help people who are unfamiliar with the Scarborough area get an idea of what each neighbourhood feels like and their own suitability to each neighbourhood. The idea of desirability for each neighbourhood would be largely based on our own personal circumstances, and so interpretation of these graphs could be taken in a variety of ways.

Discussion - Personal Choice

For my own personal circumstances, if I were moving to Scarborough I would want to consider school ratings, with a low housing price as I am still young. As I do not currently have children I would also strongly consider local amenities. One such neighbourhood that jumps out to me would be Cedarbrae as it has an average housing price of just over $200,000, a school rating of 10/10 and most importantly has its most common venue as "Thai Restaurant" and second most common as "Bakery". As a lover of food I would strongly consider Cederbrae given it these factors, as well as having a nearby athletics and sports venue to work off the food I would have eaten at the common bakeries and Thai restaurants.

Conclusion

For this project we have used k-means clustering to group together neighbourhoods to more easily look at local amenities when comparing neighbourhoods. For each neighbourhood we have considered the average housing price as well as school rating. Using the results provided above we have created tools that would allow anyone to more easily compare these neighbourhoods to make more informed decisions about suitable areas to move to within the Scarborough region. I feel that the tools that this practical has had me use will also be applicable in my own personal circumstance, allowing me to use these skills to make more informed choices about any future moves I have.

Many thanks for listening to my Capstone Presentation