Exploring Neighbourhoods in Scarborough, Toronto

Immigration to Canada

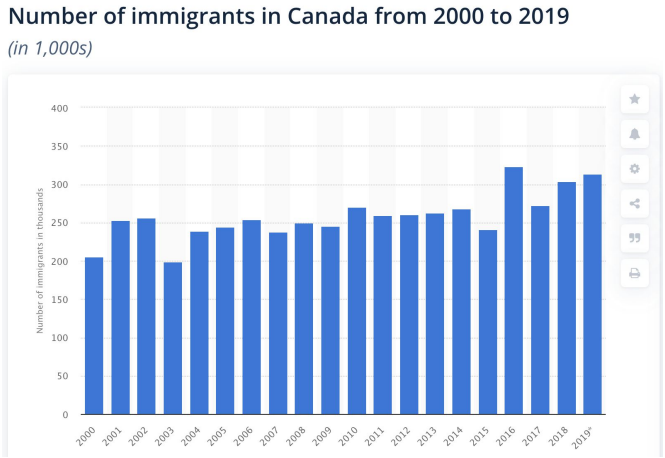


Fig 1. Nubmer of immigrants in Canada from 2000-2019

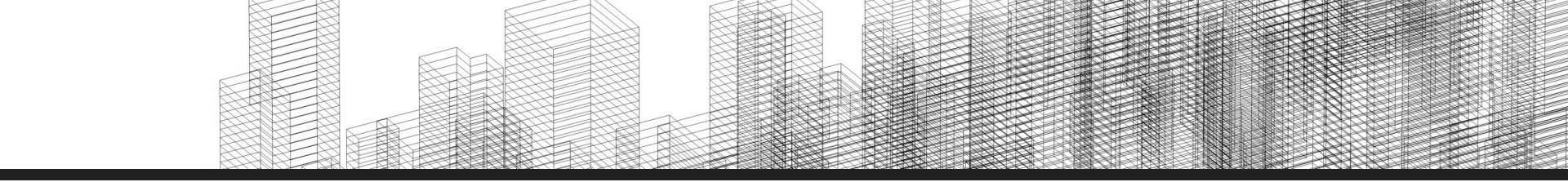
Canada's appeal as a destination to immigrants has been increasing over the past few years (see fig 1 for more). In 2019, a total of 313,580 have immigrated to the country (Erin Duffin, statista). The aim of this capstone project is to analyze and compare different neighborhoods in Toronto, so that immigrants are able to better understand the what facilities are accessible in what areas, so they can choose neighbourhoods that best fit their needs. This project will only focus on Toronto, because it is a relatively popular destination for immigrants in Canada.

Data Description

The data used in this project will be the Scarborough dataset scrapped from Wikipedia in Week 3 of this Capstone Project (it can be found in the same github repository that this document is placed in). The link to the raw data is as follows: "https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M".

This dataset consists of postal codes, latitude data and longitude data. Apart from the geographical data (postal codes, latitude, longitude), this project will also require data on the different facilities available in the area. I will collect the following from Foursquare:

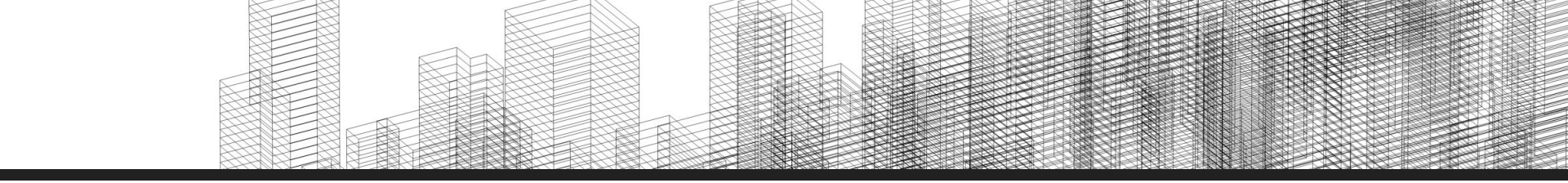
- Neighborhood
- Neighborhood Latitude
- Neighborhood Longitude
- Venues (eg. restaurants, cafes, etc.)
- Name of the venue
- Venue Latitude
- Venue Longitude
- Venue Category



Methodology

Libraries needed:

- Pandas, to create data frames;
- JSON, to handle JSON files;
- Scikit Learn, for k-means clustering;
- Geocoder, to retrieve location data;
- Beautiful Soup and Requests, to handle http requests;
- Matplotlib, to plot visuals;
- Folium, to create maps



Methodology pt.2

Segment the neighbourhoods, and group them into clusters. The algorithm I will use for this is k-Means clustering. Workflow is as follows:

- Mine nearby venues with Foursquare API
- Organize each neighbhourhood’s most common venues and put them into a dataframe
- Cluster them accordingly using the k-Means clustering algorithm Show results on a map

Results: Clustering

A sorted table of each neighbourhood, its most common venues, latitude, longitude, and assigned cluster.

Out[35]:

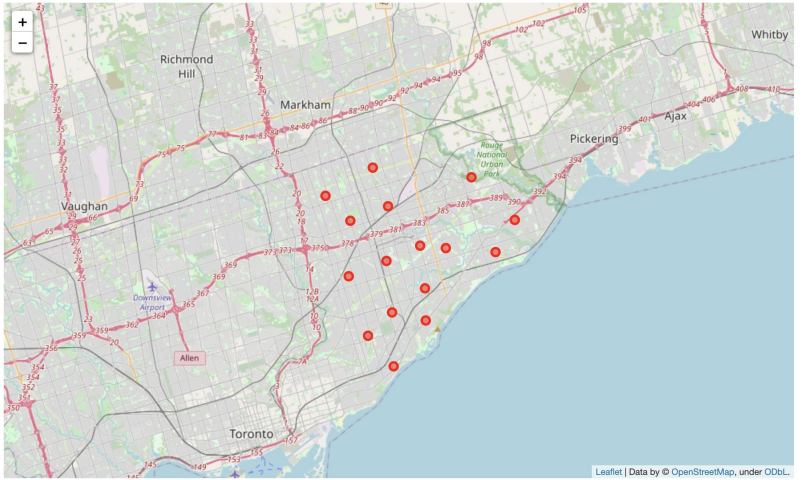
	Postcode	Borough	Neighbourhood	Latitude	Longitude	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Mo Comm Venue
0	M1B	Scarborough	Malvern, Rouge	43.81153	-79.19552	0	Eco Painting	Grizzly Bear Exhibit	Wendy's	AMJ Painting	한바탕 (Han Ba Tang) (Han Ba Tang)	GoodLife Fitness Toronto 137 Yonge Street	GoodLife Fitness Toronto Richmond and Bathurst	GoodLi Fitness Toronto Mount Pleasar and Da
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.78564	-79.15871	0	Organized By Catherine	Malt & Salt Fish & Chips	Royal Canadian Legion	Chris Effects Painting	GoodLife Fitness North York York Mills Centre	GoodLife Fitness Toronto Mount Pleasant and Da...	GoodLife Fitness Toronto King Liberty	GoodLi Fitness Toronto Bloor a Park
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.76575	-79.17520	0	Buckler Aquatics Ltd	Heron Park Community Centre	Heron Park	Peter Secor Park	GoodLife Fitness North York Madison Centre	GoodLife Fitness Toronto Mount Pleasant and Da...	GoodLife Fitness Toronto King Liberty	GoodLi Fitness Toronto Bloor a Park
3	M1G	Scarborough	Woburn	43.76820	-79.21761	0	Densgrove Park	Starbucks	Lucky Hakka	Aunty Mary's	한바탕 (Han Ba Tang) (Han Ba Tang)	GoodLife Fitness Toronto Mount Pleasant and Da...	GoodLife Fitness Toronto King Liberty	GoodLi Fitness Toronto Bloor a Park
4	M1H	Scarborough	Cedarbrae	43.76969	-79.23944	0	Centennial Recreation Centre	Drupati's Roti & Doubles	B&A Bakery	Sheridan Nurseries	TD Canada Trust	Federick Restaurant	Thai One On	CANBE Foods Inc

*the screenshot was taken from github because I didn't want to run python all over again

A clustered map of the neighbourhoods

```
['#8000ff', '#4856fb', '#10a2f0', '#2adddd', '#62fbc4', '#9cfba4', '#d4dd80', '#ffa256', '#ff562c', '#ff0000']
```

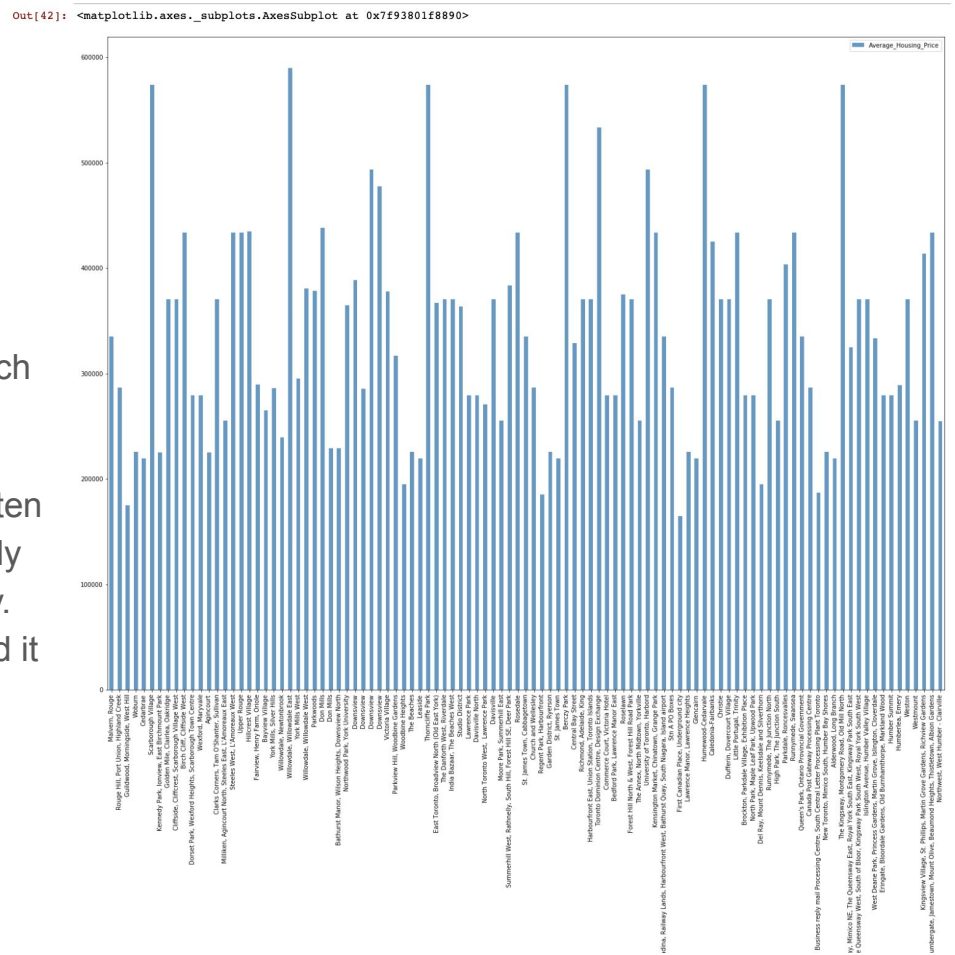
Out[38]:

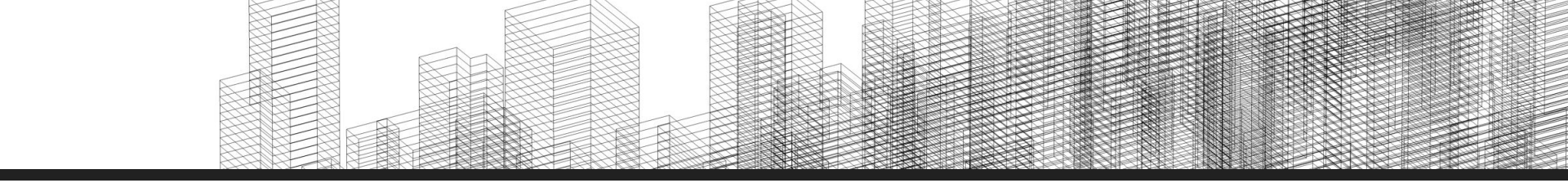


Results: House Prices

To the left is a bar graph of average house prices for each neighbourhood of Scarborough.

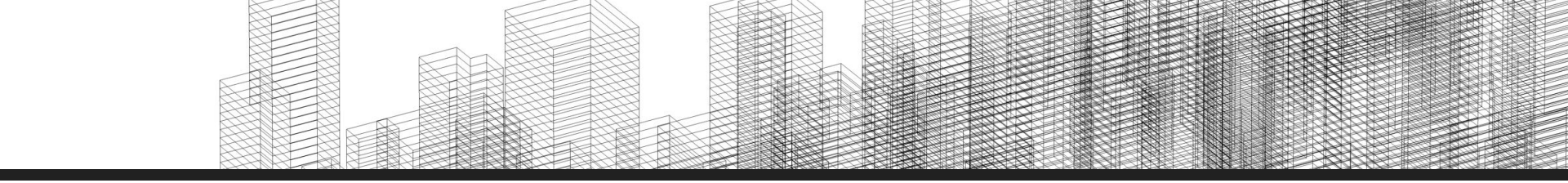
I chose to do this because I felt that house prices are often determining factors for immigrants who are not financially privileged, and it also (to an extent) reflects living quality. Therefore I think that it is an important topic to cover and it contributes to the aim of the project.





Discussion

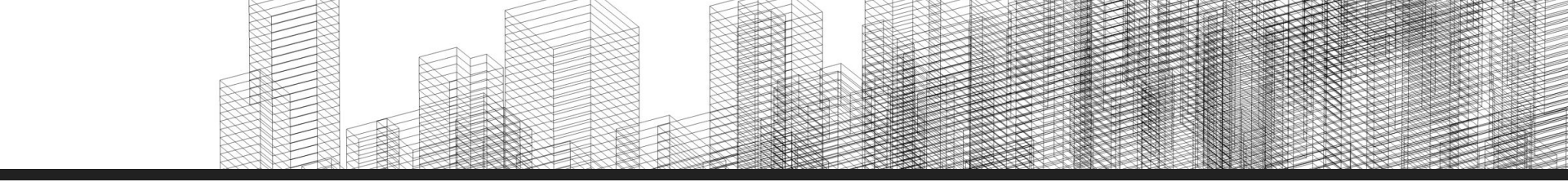
Toronto does have quite a lot of neighbourhoods to cover, and there are many other complex factors that can come into play when immigrants are deciding where to stay. This means that there are so many more methods to cluster and analyze neighbourhoods in Toronto. I set the kclusters to 10, because the area was quite diverse and 10 clusters was the amount that could represent the data most accurately.



Conclusion

In this project, I explored different neighbourhoods in Scarborough and clustered them into 10 different clusters using the k-Means clustering algorithm. I then used a bar graph to show the different house prices of each neighbourhood.

As the world become more globalized, more and more people will be moving around the world due to work, education, or simply just recreation. People can actually make sure that their stay is more comfortable by accessing such platforms where information about facilities and attractions are available.



References

Duffin, Erin. “Immigrants in Canada 2019.” Statista, 30 Oct. 2019, www.statista.com/statistics/443063/number-of-immigrants-in-canada/.