

Exploring Neighbourhoods in Scarborough, Toronto

Athena Wong

August 25, 2020

1 . Introduction

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). Immigrants often have to do a lot of research into house prices and other facilities in different areas before choosing where to stay in a country. Different people also come to Canada with different needs - some are looking for prestigious schools for their children for them to pursue higher education that may not be available in their home country, others come in search of job opportunities, and more. The aim of this capstone project is to analyze and compare different neighborhoods in Toronto, so that immigrants are able to better understand the different areas of Toronto. This will give immigrants a general understanding of 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.

Number of immigrants in Canada from 2000 to 2019

(in 1,000s)

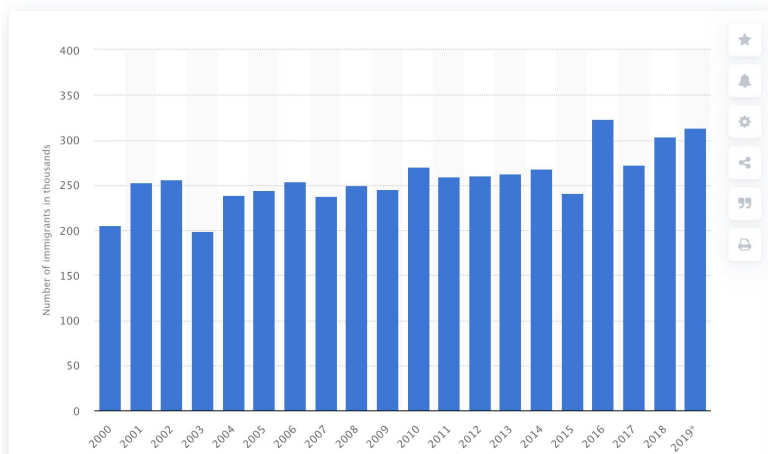


Fig 1. Number of immigrants in Canada from 2000 to 2019, shows a general upward trend.

Those who are considering immigration to Toronto would be interested in this project.

2. 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. This data will be sourced from Foursquare API. Foursquare is a crowdsourced location data provider.

I will use the following data from Foursquare API:

- Neighborhood
- Neighborhood Latitude
- Neighborhood Longitude
- Venues (eg. restaurants, cafes, etc.)
- Name of the venue (eg. name of restaurant, or name of a store, or school, etc.)
- Venue Latitude
- Venue Longitude
- Venue Category

All the data stated above are available on Foursquare, and it will therefore be the sole data source for this project.

3. Methodology

Since this project is about exploring the facilities in different neighbourhoods and assisting potential immigrants with their decision making on where to stay, the target audience will most likely compare the results for each neighbourhood covered in this project. Therefore, this project will go in the direction of a comparative analysis.

To compare the neighbourhoods, I chose to segment the neighbourhoods, and group them into clusters. The algorithm I will use for this is k-Means clustering.

Workflow:

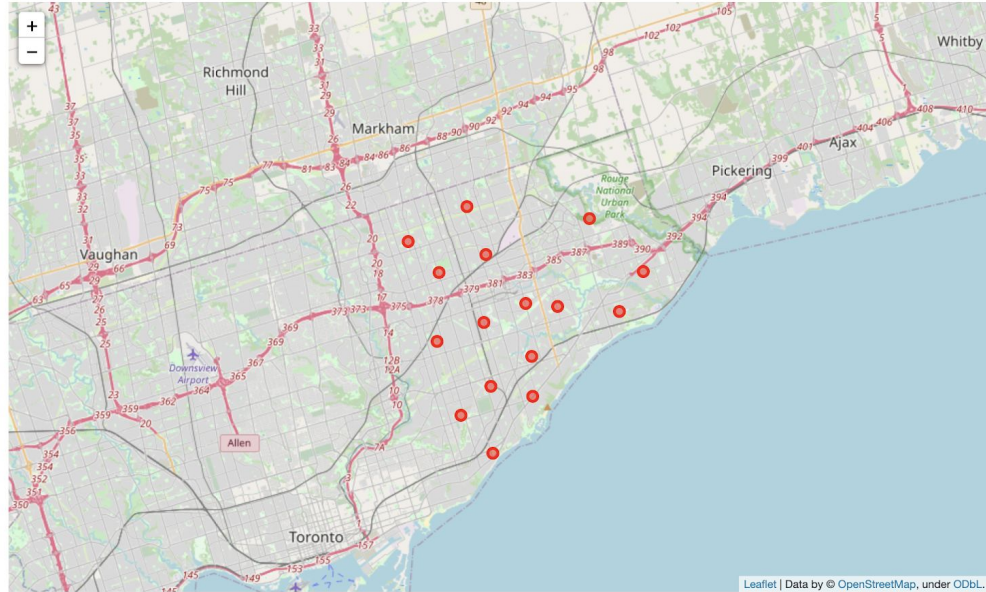
- Mine nearby venues with Foursquare API
 - Number of places per neighborhood will be set to 100
 - Radius set to 500
 - The above is due to request limitations.
- Organize each neighbourhood's most common venues and put them into a dataframe
- Cluster them accordingly using the k-Means clustering algorithm
- Show results on a map

4. Results

1. Map of clusters, as a result of k-Means clustering. See below.

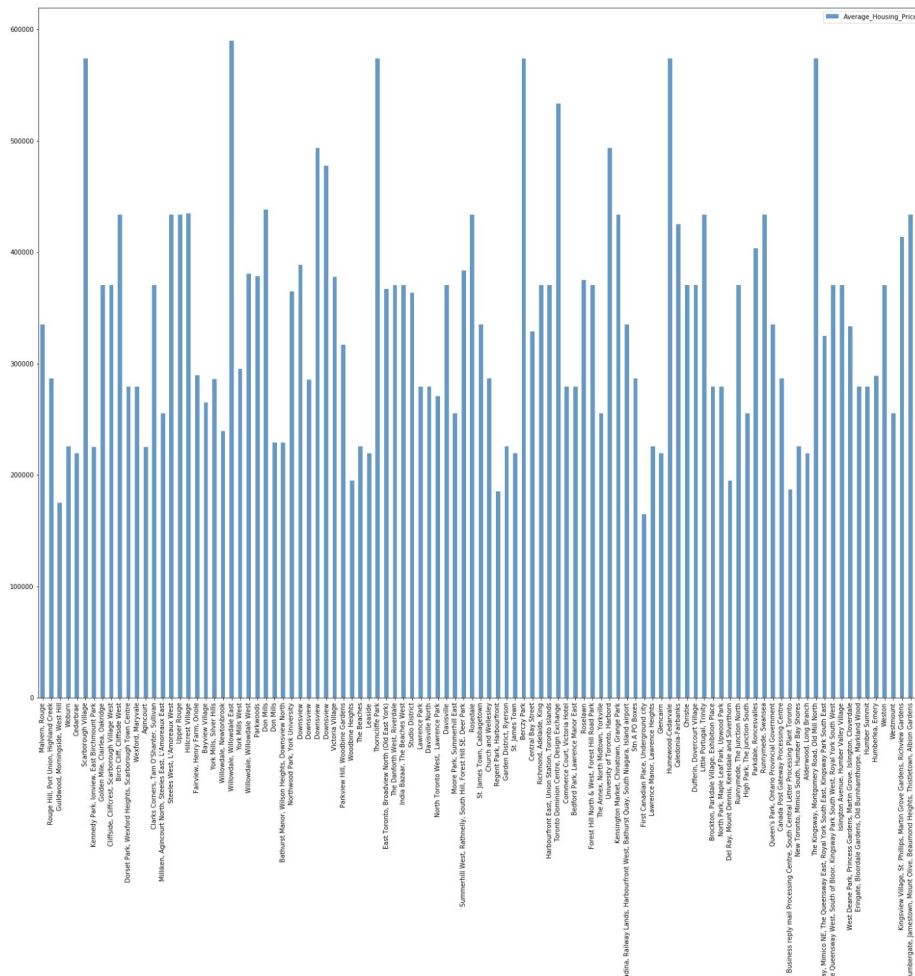
```
['#8000ff', '#4856fb', '#10a2f0', '#2adddd', '#62fbc4', '#9cfba4', '#d4dd80', '#ffa256', '#ff562c', '#ff0000']
```

Out[38]:



2. Averaging House Prices in Scarborough

Out[42]: <matplotlib.axes._subplots.AxesSubplot at 0x7f93801f8890>



The graph of average house prices is separate from the k-Means clustering. It is something that I did on top of the clustering, because I felt that only comparing the venues/facilities in the neighbourhood wasn't enough - for immigrants who are not financially privileged, house and living prices are often determining factors for where to stay.

5. Discussion Section

The neighbourhood with the highest house price is Willowdale, Willowdale East, and the lowest is First Canadian Place, underground city. The range is \$165000-58990. The gap between them is 424900. The dramatic gap definitely reflects that the house price may be a determining factor for those who aren't financially privileged.

A comparison of the top restaurants in these two places:

---- Willowdale, Willowdale East ----			---- First Canadian Place, Underground city ----		
	venue	freq		venue	freq
0	Subway	0.04	0	Starbucks	0.03
1	Tim Hortons	0.03	1	John & Sons Oyster House	0.02
2	Starbucks	0.02	2	Pilot Coffee Roasters	0.02
3	Shoppers Drug Mart	0.02	3	The National Club	0.01
4	McDonald's	0.02	4	Reds Wine Tavern	0.01

/

6. Conclusion Section

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.

```
Out[25]: Clothing Store      7
         Coffee Shop        5
         Restaurant         4
         Gas Station        2
         Tea Room           2
         Sandwich Place     2
         Pharmacy           2
         Electronics Store  2
         Bar                1
         Fish & Chips Shop   1
         Name: categories, dtype: int64
```

I also analyzed the different venues in the neighbourhoods, and did one hot encoding with the data.

7. Future Directions

I look forward to doing more projects, but maybe this time focusing on a different field, such as medicine and biology.

References

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