

Providing Context to the Boroughs of Toronto

S Potter

February 2021

1 Business Problem

1.1 Introduction

The purpose of this project is to help people better understand their future neighbourhood when making a big life decision; relocating! The analysis produced in this project will help people making smart and efficient decisions when deciding on a neighbourhood to call home in Scarborough, Toronto.

Migration into Canada affects a very large pool of people who, prior to this analysis, need to undertake a tremendous amount of personal research to make a decision on where to locate their family. This project is for those people who are looking for an easy, one-stop shop to decide what suburb might appeal to them. This decision could be influenced by a range of available facilities in the area such as cafes, schools, supermarkets, malls, theatres or even hospitals.

This project aims to analyse a range of features across neighbourhoods in Scarborough for people migrating to the area. The features that will be comparatively analysed include median housing prices, school ratings, crime rates, road connectivity, weather conditions, emergency management, water utilities and recreational facilities.

It is hoped that this analysis will assist potential migrants better understand the neighbourhood before moving to Toronto.

1.2 The Problem Being Solved

The major purpose of this project is to create an aggregated source of information about the neighbourhoods in the Scarborough area. This data will not only enable potential migrants to better understand various neighbourhoods but will also display comparative analysis that will enable people considering relocating to the area to directly compare one suburb to another.

1.3 Background on the Location & Migration Trends

Scarborough is a overwhelmingly popular destination for new migrants in Canada. As a result, it is one of the most diverse and multicultural areas in the Greater Toronto Area. Although migration is a contentious political topic, the general trend of migration into Canada will continue into the foreseeable future.

2 Data Required

2.1 Geographic Data

The first set of data utilised in this project is the Scarborough region information available on Wikipedia. This data provides us with zip codes and longitude & latitude coordinates.

2.2. Facilities Data

To assess a neighbourhood for liveability, we first need to understand the facilities on offer in each suburb.

In order to gain this information, this project will utilise "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. This information includes venue names, locations, menus and even photos. Due to the extensive amounts of data available, the Foursquare platform will be used as to retrieve all data in relation to facilities within a given area.

After determining the list of neighbourhoods in the Scarborough area, our data analysis workbook will be connected to Foursquare via API to gather information about venues inside all relevant neighbourhoods. The information obtained from Foursquare, per venue, is as follows:

1. Neighbourhood
2. Neighbourhood Latitude
3. Neighbourhood Longitude
4. Venue
5. Name of the venue
6. Venue Latitude
7. Venue Longitude
8. Venue Category

3 Data Cleansing Activities

After using Beautiful Soup to scrape the Wikipedia data source, there were several data fields that required cleansing. Of the 180 rows, only 103 contained valid suburbs or neighbourhood combinations. This is because the web source lists all possible postcodes in sequential order, even if a particular postcode is not in use by the Toronto area.

It was helpful that any field that didn't have valid data in it had the words 'Not assigned'. This enabled me to easily drop any row where the words 'Not assigned' appeared in either the Borough or Neighbourhood fields.

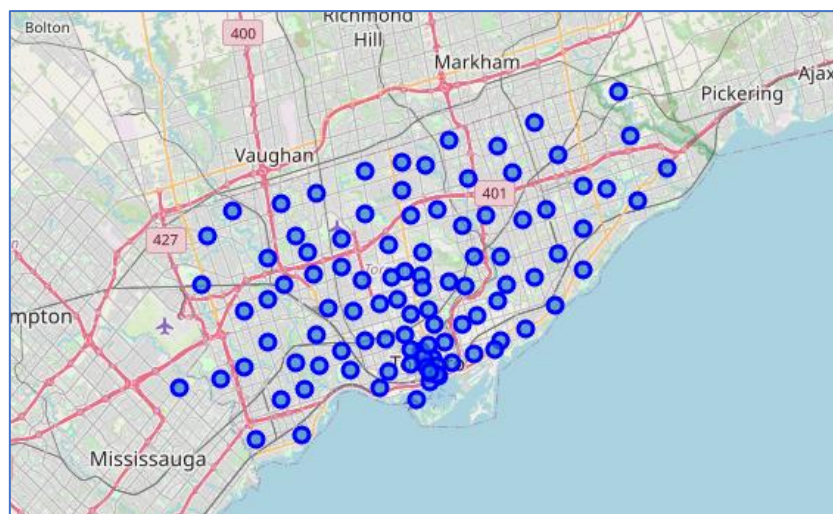
After doing this, I was left with a data frame of 103 rows in the format shown below:

	PostalCode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

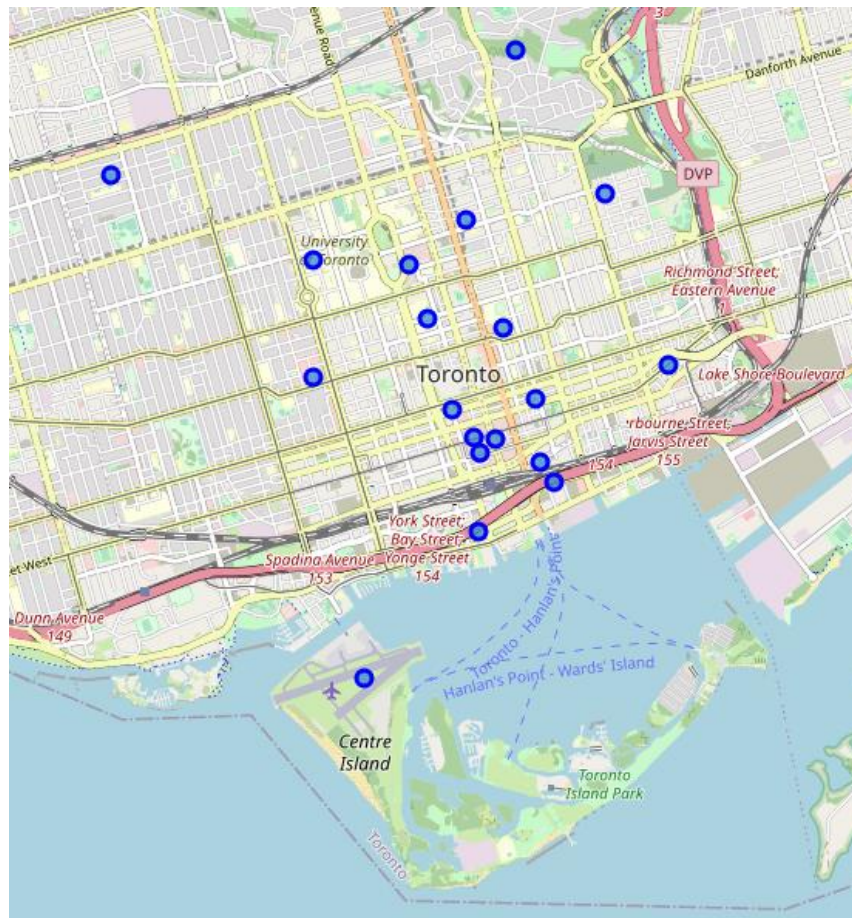
I was then able to add in two new columns to the data frame and use the geocoder function to determine input latitude and longitude into each row of data. The geocoder function does this by linking with Google Maps and using the postcode as an input for Google Maps to return the coordinates of any given suburb.

4 Visualisation of Suburbs

With this newfound data, I was able to generate a visualisation of all suburbs in the Toronto area. I achieved this by using the folium function, which utilised the latitude and longitude data obtained in the previous step.



We were also able to look more specifically at each Neighbourhood, to determine the suburbs in each. As an example, the below folium output shows the suburbs in the Downtown neighbourhood.



5 Facilities Data

The process to date has only allowed us to understand the proximity of suburbs to other suburbs, determine what suburbs fall into certain neighbourhoods and to determine general location of each suburb. This doesn't provide house buyers with enough information to make an informed decision as to what suburb they might like to live in. This decision mainly revolves around the facilities available in each suburb, which is where the following data really makes this project come to life.

After linking to the Foursquare database via API, we requested data of venues in close proximity to the latitudes and longitudes of each suburb. To appropriately manage Foursquare request limitations, we limited our search to only look for venues 100 metres from the coordinates provided. This is a key limitation that could be expanded in future projects.

After receiving the facilities data, we had to cleanse that data by drop many superfluous data fields. Once this was performed, we were left with a data frame of 1,222 facilities (rows) with four data points each (venue name, category, latitude, and longitude). An example of the data frame is shown below:

	name	categories	lat	lng
0	Roselle Desserts	Bakery	43.653447	-79.362017
1	Tandem Coffee	Coffee Shop	43.653559	-79.361809
2	Impact Kitchen	Restaurant	43.656369	-79.356980
3	The Distillery Historic District	Historic Site	43.650244	-79.359323
4	Cooper Koo Family YMCA	Distribution Center	43.653249	-79.358008

6 Neighbourhood Analysis

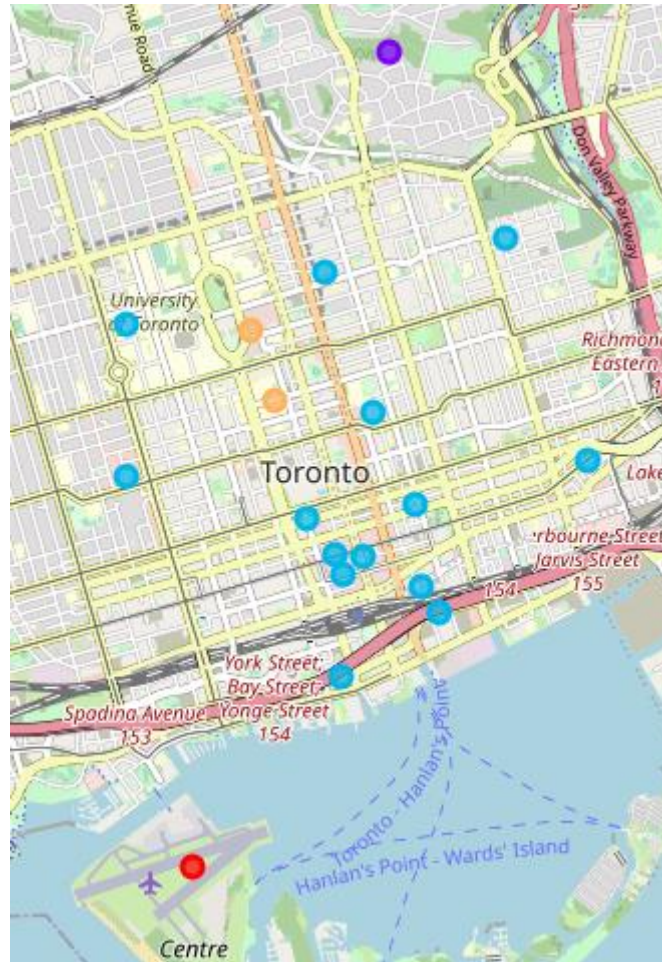
With all of the data required now obtained and cleansed, we were able to run some analysis on each Neighbourhood. This involved counting the number of facilities of every category in each Neighbourhood. To simplify the equation, we moved away from a pure count and instead took a percentage of venue categories in any one Neighbourhood. An example of this is below:

	Neighbourhood	Adult Boutique	Airport	Airport Food Court	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Antique Shop	Aquarium	...
0	Berczy Park	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.0	...
1	CN Tower, King and Spadina, Railway Lands, Har...	0.000000	0.066667	0.066667	0.133333	0.2	0.133333	0.000000	0.0	0.0	...
2	Central Bay Street	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.0	...
3	Christie	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.000000	0.0	0.0	...
4	Church and Wellesley	0.012658	0.000000	0.000000	0.000000	0.0	0.000000	0.012658	0.0	0.0	...

Using this data, we were then able to determine what the most common venue in any Neighbourhood was.

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Berczy Park	Coffee Shop	Cocktail Bar	Bakery	Cheese Shop
1	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Airport Lounge	Airport Terminal	Plane
2	Central Bay Street	Coffee Shop	Sandwich Place	Café	Italian Restaurant
3	Christie	Grocery Store	Café	Park	Nightclub
4	Church and Wellesley	Coffee Shop	Sushi Restaurant	Japanese Restaurant	Restaurant

This data can be visualised using the folium function to understand exact locations of various categories of venues. For example, the below shows different kinds of venues (represented by different coloured dots) in the Downtown Neighbourhood of Toronto.



7 Conclusion

The outcomes of this project are significant for migrants deciding to relocate to the Toronto area, particularly where those people are from a non-Canadian background. The output ensures that these people can find a Neighbourhood that best meets their needs. As an example, someone who plans to visit family overseas on a regular basis might prioritise the proximity to the Airport in the CN Tower Neighbourhood, whereas someone from Japan might value the significance of Japanese Restaurants in the Church and Wellesley Neighbourhood.