



# Capstone Project

---

# Introduction

**Toronto** is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 in 2016. Current to 2016, the Toronto census metropolitan area (CMA), of which the majority is within the Greater Toronto Area (GTA), held a population of 5,928,040, making it Canada's most populous CMA. The city is the anchor of the Golden Horseshoe, an urban agglomeration of 9,245,438 people (as of 2016) surrounding the western end of Lake Ontario.

Toronto is an international center of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.

# Problem Description

Now let me explain the context of this Capstone project through a scenario. Say you live on the west side of the city of Toronto in Canada. You love your neighbourhood, mainly because of all the great amenities and other types of venues that exist in the neighbourhood, such as gourmet fast food joints, pharmacies, parks, graduate schools and so on. Now say you receive a job offer from a great company on the other side of the city with great career prospects. However, given the far distance from your current place you unfortunately must move if you decide to accept the offer.

Wouldn't it be great if you are able to determine neighbourhoods on the other side of the city that are the same as your current neighbourhood, and if not perhaps similar neighbourhoods that are at least closer to your new job?

# Target Audience

This information provided by this report would be useful for people who are interested in relocating to a different part of the city and are interested in finding new neighbourhoods that are highly similar to their existing neighbourhood.

# Data Description

1. Neighborhoods in Toronto — Wikipedia.
2. Using Geopy to get geological location by address name.
3. Using Foursquare API to get the most common venues of given Borough of Toronto.
4. Using Foursquare API to get the venues' record of given venues of Toronto.

# Methodology

---

## 1. Scrape the Wikipedia page and gathering data into a Pandas data frame

	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

# Methodology

---

## 2. Generating a map of Toronto and plotting the Neighbourhood data on it

We first filter the data to find boroughs containing the word “Toronto”

	PostalCode	Borough	Neighborhood	Latitude	Longitude
37	M4E	East Toronto	The Beaches	43.676357	-79.293031
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
42	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572
43	M4M	East Toronto	Studio District	43.659526	-79.340923
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

# Methodology

---

## 3. Utilizing Foursquare API to explore the neighbourhoods

	name	categories	lat	lng
0	Glen Manor Ravine	Trail	43.676821	-79.293942
1	The Big Carrot Natural Food Market	Health Food Store	43.678879	-79.297734
2	Grover Pub and Grub	Pub	43.679181	-79.297215
3	Upper Beaches	Neighborhood	43.680563	-79.292869



# Results

---

We use the matplotlib and folium packages to visualize the clusters on a map of Toronto.

