## Introduction

Toronto is the largest cities in Canada. Downtown Toronto is urban and very densely populated with a variety of residential and commercial buildings. An estimated 2.83 million people live in Toronto as of 2020 and this number is rapidly increasing. As such there are plenty of people who eat, sleep, and navigate the city daily. One of the key necessities for more people is shelter. In this scenario we are approached by a development company that wants to build an apartment complex to maximize the desirability to live there. This in turn, will help to increase the value of the building Therefore, our problem of choice will be trying to decide where we should build a new apartment complex in Toronto.

## Data

The main dataset we will be focused on is the foursquare dataset. From the foursquare dataset we will obtain location data and points of interests that are related to our gyms. Some of the most important factors will be derived from the following data in foursquare are as follows:

1. Neighbourhoods

Neighborhood data is key as this is how we will divide the candidate areas for the developments. As from the lab in week 3 the data of the neighbourhoods in Toronto was scraped from <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M> The is read into a pandas data frame using the read\_html () method. This gives us the data in format that can be easily processed. Using the head method, we can view the data as follows.

Graphical user interface, text

Description automatically generated

1. Location Data from GeoPy

From the geocoder library and GeoPy we obtain data on the geographical position of the neighbourhoods. This allows us to make visualization with Folium and the locations are also important in determining their relative distance to other points of interest. The first few rows are shown below. Graphical user interface, text, application, email

Description automatically generated

1. Foursquare Venue Data

Venue data from Foursquare will help to determine the types of amenities that will be available to individuals who purchase a unit. This data can help to predict the expected value of our condo location depending on which neighbourhood we decide to build in. The venue data has been extracted using the Foursquare API. This data contains venue recommendations for all neighborhoods in Toronto and is used to study the popular venues of different neighborhoods as well as build the unsupervised learning model to cluster neighborhoods. Sample data is below. From the venue we can see that there are distinct categories that can be either beneficial or a detriment to the value of our apartment development.

Table

Description automatically generated

## Methodology

## Results

## Discussion

## Conclusion