# Introduction

## Background

Two of the most populated cities in Canada, apart from Toronto, are Montreal and Vancouver. Both cities are very multicultural, have hosted the Olympics, and are both homes to large universities, industry, and world-class attractions. However, located at opposite sides of this vast country they have drastically different cultural identities and local attractions. Montreal is Canada’s oldest established city and was originally founded in 1642 and boasts a large francophone community. Vancouver, in contrast, was founded much later in 1867 and hosts a large population of immigrants from eastern Asia. This difference in ethnographic make-up can greatly influence available cuisines and other amenities within a city. Locations of resources within a city can be a huge factor for visitors when selecting which city to visit, and also for locals choosing where in the city to live.

## Problem

Given how culturally and historically different Montreal and Vancouver are, it would be interesting to look at the similarities and differences between these two cities based on popular amenities and rental prices. Foursquare data can be helpful in this regard, to look at trending locations within these cities and compare neighbourhoods within each city as well as between Montreal and Vancouver.

## Audience

This study can potentially help people in determining which city to visit or live in based on their interests.

# Data acquisition and cleaning

## Data sources

Montreal is divided into 19 neighbourhoods and the dataset was found here: <https://en.wikipedia.org/wiki/Montreal>. This dataset included census data from 2016 which reported population, average monthly rent, and area of each neighbourhood.

Unfortunately, no equivalent table existed for Vancouver. The data was then sourced from a combination of sources. Neighbourhood designations for Vancouver were taken from: <https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Vancouver>. Other data including area, population and average monthly rent were taken from <https://vancouver.ca/news-calendar/areas-of-the-city.aspx>. Data was manually compiled from individual pdfs corresponding to 2016 census data in new table. The table was saved in .csv format and then converted into a pandas dataframe.

Datasets were chosen to represent only the metropolitan boundaries as opposed to the greater Montreal area or the greater Vancouver area which comprise the surrounding citys and towns. The 2016 census was the last available source of population and cost of living values as the long-form national census is only performed every 5 years.

## Data cleaning

The geographical locations of the neighbourhoods in both Vancouver and Montreal were sourced using geopy and the latitude and longitude positions were added to the datasets. The foursquare API was used to return the top 100 visited venues within each neighbourhood and then clustered according to the kmeans algorithm.