**Project: which city is more suitable for living in Alberta?**

**Section 1**  
**Introduction:** Alberta is a province of Canada. With an estimated population of 4,067,175 people as of the 2016, it is Canada's fourth most populous province and the most populous of Canada's three prairie provinces. Alberta's capital, [Edmonton](https://en.wikipedia.org/wiki/Edmonton), is near the geographic center of the province and is the primary supply and service hub for Canada's [crude oil](https://en.wikipedia.org/wiki/Petroleum). About 290 km south of the capital is [Calgary](https://en.wikipedia.org/wiki/Calgary), the [largest city in Alberta](https://en.wikipedia.org/wiki/List_of_cities_in_Alberta). Calgary and Edmonton center Alberta's two [census metropolitan areas](https://en.wikipedia.org/wiki/Census_metropolitan_area), both of which have populations exceeding one million, while the province has [16 census agglomerations](https://en.wikipedia.org/wiki/List_of_census_agglomerations_in_Alberta).

**Problem Statement:** As in some people’s minds, Alberta may not as popular as other provinces like BC or Ontario in Canada. However, due to the development in recent years, it is also worth considering living. This project has been designed as someone moving to the Alberta, who is interested in deciding which city in Alberta is more suitable for living (having more convenience facilities like restaurants, shops etc.

**Libraries Employed:**

* Pandas: for manipulating data frames.
* Matplotlib: Python plotting module.
* BeautifulSoup with Requests: library handles online request, scraping http.
* Geocoder: retrieves locational data.
* FourSquareAPI: used to identify venues in the area.
* Folium: python data visualization library, used to visualize neighborhood clusters on an interactive map.
* JSON: library to handle JSON files.
* XML: separates data, allowing XML data to be displayed in plain text.
* Scikit Learn: mathematic library, employed for k-means clustering.

**Data**

* Using data from :

<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T> to get list of Borough, neighborhoods, coordinates in Alberta. For the missing data of coordinates, I get it from using pgecode. In this study, we will include every city in Alberta.

**Methodology:**

First, I scrape the data from website. For the missing values, I incorporated values from using pgecode. Then, I use Foursquare API to identify venues for multiple neighborhoods of Alberta. After that, I cluster multiple neighborhoods using K-means, finding the similarities of neighborhoods in each data set in order to decide which city has more neighborhoods (having more convenience facilities). Finally, I use latitude and longitude coordinates of those neighborhoods to plot the map, which could have a more visualized view of the result.