1. **A description of the problem and a discussion of the background.** (**15 marks**)

Calgary, Alberta, incorporated as a city in 1894, population 1,239,220 (2016 [census](https://www.thecanadianencyclopedia.ca/en/article/demographic-data-collection)) 1,096,833 (2011 census). The city of Calgary is situated on the Bow River in southern Alberta, about 220 km north of the American border at the meeting point of the Western prairies and mountain foothills. It is the financial centre of western Canada, based on its key role in the development of the region’s oil and gas industry.

In the recent past, Calgary has seen very polar economic conditions by virtue of Alberta’s dependence on Oil and Gas market. Like any other city, poor economic conditions lead to loss of jobs, which is often accompanied by an increase in crime rate. Consequently, many passive investors find it tough to justify investing in businesses in Calgary.

Therefore, I will be using data from City of Calgary’s website to determine clusters of types of businesses in communities most impacted by commercial crimes. While Calgary is an economic hub in the province of Alberta, it is still a growing city with a moderate number of businesses, however not as many as you would expect to find in metro cities like New York, Toronto, or Vancouver.

1. **A description of the data and how it will be used to solve the problem.** (**15 marks)**

We will be working with publicly available data repository managed by City of Calgary to find crime statistics in the various communities within Calgary.

<https://data.calgary.ca/Health-and-Safety/Community-Crime-Statistics/78gh-n26t>

I will scrape the data to only include types of crimes that have the potential to impact commercial businesses (e.g. commercial break-ins) and leave out data that has a low probability of affecting businesses (e.g. domestic violence).

Subsequently, using geospatial data of these communities, we will use FourSquare API to find the businesses within five kilometer radius of these communities. We will be using the category of businesses to determine which type of businesses are most impacted by commercial crimes in Calgary. Finally, K-means clustering will be used to compartmentalize the findings and we shall attempt to identify the theme of each cluster.