**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.