Introduction

For my capstone project I will be looking at different neighbourhoods in the city of London, Ontario and deciding the best location to open a new bank. I will be looking at the geographical layout of the city, splitting the neighbourhood up into clusters and comparing each cluster's current need for a new bank. Identifying the best place in a community to open a new bank can be mutually beneficial for both a bank scouting out potentially successful locations and a community lacking access to more banking options.

It's easy to understand the importance of a bank's location to the bank itself. Clearly, the customers and thus a bank's profits can be greatly affected by the location of a bank. What can be a bit less obvious is how important banks can be to a community. A bank and all of its services can be incredibly important to the structure of a healthy community economy. Even during a time where day to day banking needs are being taken care of digitally, a brick and mortar location can still serve its own purposes for customers, and give options to those who may not get them through other mediums. Increasing community access to brick and mortar banking locations increases the access households have to credit products that can be used to build wealth or savings accounts to invest in the future. Alternatively, without access to basic financial services, communities can become a target for more predatory, high-cost credit options.

Interest: Obviously, any banks looking to open new locations, which could include both new banks looking to open its first brick and mortar locations in communities, as well as existing banks that are looking to add locations in a city or town in which they already reside. Along with banks, governments-local and beyond-may be interested in studying neighbourhoods that are under-serviced in terms of banking.

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

Data:

To solve the problem we will use the following data:

- list of neighbourhoods in London, Ontario. We chose to look specifically at the city of London, Ontario to ask the question of where would be the best location for a new bank. London is a medium sized city located in southwestern Ontario, surrounded by rural communities, both of which can be really affected by the (lack of) presence of banking locations.
- latitude and longitude coordinates of those neighbourhoods which will allow us to plot the map and obtain venue data, such as number of banks in each neighbourhood
- venue data, such as the number of banks in each neighbourhood, which we will use for clustering the neighbourhoods and ultimately deciding on the neighbourhoods in most need of another bank-and thus the best location for a bank

This project will make use of several data science skills, including web scraping, data cleaning, data wrangling, machine learning techniques such as K-means clustering, Foursquare

API, and map visualization with Folium. We will take a look at the neighbourhoods in London, Ontario and group them together using k-means clustering. After they are grouped together into k groups, we will analyze these clusters and find which neighbourhood clusters have the least amount of banks, signally an opportunity for a new location.

The Wikipedia page: https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_London,_Ontario contains a list of neighbourhoods in London Ontario, with a total of 9 neighbourhoods. I will use web scraping techniques from python and beautiful soup packages to extract the data from the wikipedia page containing all of the London neighbourhoods. I will then get the latitude and longitude coordinates of the neighbourhoods with the Python Geocoder package.

Once I've got the geographical coordinates of the neighbourhoods I will use the Foursquare API to get the venue data for the London neighbourhoods, specifically the number of banks in each neighbourhood, which will allow me to identify which area has the least amount of banks, and thus the greatest need for another brick-and-mortar banking location.