Coursera Applied Data Science Capstone Project

Exploring the opportunity for opening a new recreational center in Calgary, Alberta

By: Praise Okwa

May 2020

Introduction

For many cities like Calgary, recreational centers provide an opportunity for active living and recreation in a safe, inclusive environment. For cities they are incredibly important for a healthy, vibrant community. By creating a positive atmosphere, these facilities become essential to personal health and wellness, thereby reducing reliance on healthcare and other costly social services. This in turn boosts the local economy and can also help contribute to overall economic development.

For the city of Calgary, planning a development project like a recreational center, therefore requires details analysis in order to determine the success or failure of any proposed projects.

Objectives

My objective for this project is to analyse the city of Calgary using data science and propose the optimum location for a new recreational center. By using Segmentation and Clustering techniques, my aim is to provide insights to a city council/property developer considering the feasibility of such a development.

Data

In order to solve this problem, I will be making use of datasets available online. These datasets include:

- A list of all neighbourhoods within the City of Calgary including residential communities, industrial areas, major parks and residual areas by electoral ward.
- Datasets that show names and addresses for current recreation facilities, including amenities available at each location.
- Longitudinal and Latitudinal data for locations of neighbourhoods as well as current recreational facilities, as this will aid in collecting more data relating to those locations as well as allow for plotting a map.

Wikipedia provides a list of the neighbourhoods in Calgary along with information such as the type, population and dwellings for each neighbourhood. Web scraping will be used to collect this data and with LXML and beautifulsoup packages I will organize this data into a Pandas data frame. Python Geocoder will then be used to assign longitude and latitude coordinates to the neighbourhoods.

Once this process is completed the Foursquare API and datasets from the City of Calgary will then be used to explore the recreational venues in the city. Using machine learning; K-means clustering, and by visualizing the results on a map using Folium, I will be able to make a proposal to meet my objectives.