Capstone Project Report

Family Safe Neighborhoods in Canada

Introduction

I am a data scientist about to relocate to Canada with my family. I will relocating to Vancouver, more specifically, and would like to explore the safest neighborhoods there. I am considering starting a side business as well and would love to find the safest area to both raise my family and start my authentic middle eastern food shop business. My goal for this project is to find a safe neighborhood in Vancouver, Canada. To do that, I will start by choosing the safest borough by analysing crime data and short listing a neighbourhood. I will then analyse data and focus on the safest borough and explore its neighborhoods and the 10 most common venues in each neighborhood to explore the options for my potential business.

Data

Firstly, I need to find the aspects that will affect my choice. These are:

Finding the safest borough based on crime statistics

Choosing the right neighborhood within the borough

Finding the most common businesses in the area

To do that, I will utilize the geographical coordinates of Vancouver to plot neighbourhoods in a borough that is safe and in the city's vicinity, and finally cluster our neighborhoods and present our findings.

Data Sources:

Real world data set from Kaggle containing the Vancouver Crimes from 2003 to 2019: A dataset consisting of the crime statistics of each Neighbourhoof in Vancouver along with type of crime, recorded year, month and hour.

List of officially categorized boroughs in Vancouver from Wikipedia.: Borough information will be used to map the existing data where each neighbourhood can be assigned with the right borough.

Neighbourhood's co-ordinates.: This data will be fetched using OpenCage Geocoder to find the safest borough and explore the neighbourhood by plotting it on maps using Folium and perform exploratory data analysis.

Neighborhoods, boroughs, and the most common venues and the respective Neighbourhood along with co-ordinates.: This data will be fetched using Four Square API to explore the neighbourhood venues and to apply machine learning algorithm to cluster the neighbourhoods and present the findings by plotting it on maps using Folium.

Methodology

The methodology is divided into two parts:

1. Exploratory Data Analysis: Visualize the crime reports in different Vancouver boroughs to identify the safest borough and normalize the neighborhoods of that borough. We will Use the resulting data and find 10 most common venues in each neighborhood.
2. Modeling: To help me choose the right neighborhood within a borough we will be clustering similar neighborhoods using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. We will use K-Means clustering to address this problem so as to group data based on existing venues which will help in the decision making process.

Results and Discussion

The objective of the business problem was to help me identify one of the safest borough in Vancouver, and an appropriate neighborhood within the borough to bring up a family and set up a commercial establishment like my middle eastern shop. This has been achieved by first making use of Vancouver crime data to identify a safe borough with considerable number of neighborhood for any business to be viable. After selecting the borough it was imperative to choose the right neighborhood where shops were not among venues in a close proximity to each other. We achieved this by grouping the neighborhoods into clusters to assist the stakeholders by providing them with relavent data about venues and safety of a given neighborhood.

Conclusion

We have explored the crime data to understand different types of crimes in all neighborhoods of Vancouver and later categorized them into different boroughs, this helped us group the neighborhoods into boroughs and choose the safest borough first. Once we confirmed the borough the number of neighborhoods for consideration also comes down, we further shortlist the neighborhoods based on the common venues, to choose a neighborhood which best suits the business problem.