Report

Table of contents:

Introduction

The aim of this project is to find a safe and secure location for relocation of family of 4 in Edmonton, Canada. Specifically, this report will be focused on finding a safe and friendly neighborhood in Edmonton, Canada.

The task would be to choose the safest borough by analyzing crime data and short listing a neighborhood

We will make use of our data science tools to analyze data and focus on the safest borough and explore its neighborhoods and the most common venues in each neighborhood so that the best neighborhood where grocery store is not amongst the most common venue can be selected.

Data

Based on definition of our problem, factors that will influence our decision are:

- finding the safest borough based on crime statistics
- finding the most common venues
- choosing the right neighborhood within the borough

We will be using the geographical coordinates of Vancouver to plot neighborhoods in a borough that is safe and in the city's vicinity, and finally cluster our neighborhoods and present our findings.

Following data sources will be needed to extract/generate the required information:

Part 1: Using a real-world data set from Kaggle containing the Edmonton Crime Statistics from 2018: A dataset consisting of the crimes of each Neighborhood in Edmonton along with type of crime, recorded year, month, etc.

Part 2: Gathering additional information of the list of officially categorized boroughs in Edmonton from Wikipedia.: Borough information will be used to map the existing data where each neighborhood can be assigned with the right borough.

Part 3: Creating a new consolidated dataset of the Neighborhoods, along with their boroughs, crime data and the respective Neighborhood's co-ordinates.: This data will be fetched using OpenCageGeocoder to find the safest borough and explore the neighborhood by plotting it on maps using Folium and perform exploratory data analysis.

Part 4: Creating a new consolidated dataset of the Neighborhoods, boroughs, and the most common venues and the respective Neighborhood along with co-ordinates.: This data will be fetched using Four Square API to explore the neighborhood venues and to apply machine learning algorithm to cluster the neighborhoods and present the findings by plotting it on maps using Folium.

Methodology

Categorized the methodology section into two parts:

Exploratory Data Analysis: visualize the crime reports in different boroughs to identity the safest borough and normalize the neighborhoods of that borough. The data will be used to find most common venues in each neighborhood.

Modelling: To make the right choice for best neighborhood within a borough clustering similar neighborhoods using K - means clustering will be used which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size.

Results and Discussion

The objective of the business problem was to help a family of 4 to identify one of the safest borough in Edmonton. This has been achieved by first making use of Edmonton crime statistics to identify a safe borough. It was achieved by grouping the neighborhoods into clusters to provide relevant 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 Edmonton and later categorized them into different boroughs, this helped us group the neighborhoods into boroughs and choose the safest borough.