SELECTING THE SAFEST NEIGHBORHOOD IN VANCOUVER

Imam Maulana

Introduction

Background

Vancouver is a coastal seaport city in western Canada, located in the Lower Mainland region of British Columbia. The Greater Vancouver area had a population of 2,463,431 as in ,2106making it the third-largest metropolitan area in Canada. Crime in different forms is a prevalent distress to the people in Metropolitan cities and Vancouver is no exception. Crimes like break into commercial property to for theft are on rise and people thinking to enter into similar business should bear in mind criminal activity of the neighborhood before finalizing a location. We look to address this issue by analyzing the crime data of Vancouver City and finding the safest borough and a neighborhood with in the borough which best suits the requirements ofour business problem.

Introduction

Problem:

The aim of this project is to find a safe and secure location for opening of commercial establishments in Vancouver, Canada. Specifically, this report will be targeted to stakeholders interested in opening any business place like Grocery Store in Vancouver City, Canada. The first task would be to choose the safest borough by analyzing crime data for opening a grocery store and short listing a neighborhood, where grocery stores are not amongst the most common venues, and yet as close to the city as possible. We will make use of our data science tools to analyze data and focus on the safest borough and explore its neighborhoods and the 10most common venues in each neighborhood so that the best neighborhood where grocery store is not amongst the most common venue can be selected.

Data Acquisition

- To fetch the crime details of Vancouver I used real world data set published on Kaggle. Though this dataset included type of crime, recorded time and coordinates of the criminal activity along with neighborhood, the neighborhoods were not properly categorized into boroughs which I fetched from Wikipedia. Further the coordinates of the data has been fetched using the OpenCage Geocoder API. Foursquare API is used to fetch venues for the listed neighborhoods.
- The second source of data is based on data from a Wikipedia, which was not didn't require any scraping as it was direct categorizations. The page contains additional information about the neighborhood and its boroughs. The third data source is generated from OpenCage API
- The third data source is generated from OpenCage API

Data Cleaning

- Data from the kaggle data source was heavy file which Git could not accommodate. The Vancouver Crime report had close to ~600,000+ rows of information. Because of the sheer size of the dataset, we choose to take into consideration recent most crimes of the year 2018 which would greatly reduce the number of row in the dataset.
- Since the original data source couldn't be uploaded to git I processed the dataset in the runtime to filter the records of crimes that took place in the year 2018, created a new csv out of it using pandas and uploaded it to git hub repository.
- Due to improper encoding of the co-ordinates of the crime record, the exact same coordinates from the data couldn't be used for plotting because the co-ordinates seemed to be corrupted. Along with X,Y columns in the dataset which represented the GPS coordinates of the criminal activity, other fields such as month and hour in which the crime took place has been dropped because they were not in the scope of the problem.