

Capstone Project- The Battle of the Neighborhoods (Week 1)

Business Problem

This project defined to find a safe and secure location for opening of commercial establishments in Vancouver, BC, Canada. Specifically, this report will be targeted to stakeholders interested in opening any business place i.e. Liquor Store in the Vancouver City.

Following are the steps that will be taken:

- The first task would be to choose the safest borough by analysing crime data.
- Shortlisting a neighbourhood, where liquor stores are not amongst the most common venues
- As close to the city as possible.

Data

Based on the business problem, factors that will influence the analysis are:

- Reviewing the safest borough based on crime statistics
- Reviewing the most common venues
- Selecting the right neighbourhood within the borough

For the capstone project we will be using 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 the findings.

To extract/generate the required information, following are the steps and data sources that will be needed:

- Step 1: Using data set from Kaggle containing the Vancouver Crimes from 2003 to 2019: A dataset consisting of the crime statistics of each Neighbourhood in Vancouver along with type of crime, recorded year, month and hour.
- Step 2: Gathering additional data list to officially categorized boroughs in Vancouver from Wikipedia. This information will be used to map the existing data where each neighbourhood can be assigned with its respective borough.
- Step 3: Creating a new consolidated dataset for the neighborhoods, along with their boroughs, crime data and the respective neighbourhood's co-ordinates. The latitude and longitude data will be fetched using OpenCage Geocoder. To explore the neighbourhood by plotting it on maps using Folium and perform exploratory data analysis.
- Step 4: Creating a new consolidated dataset of the 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.