# **Crime Analysis for Neighborhoods in Vancouver**

## **Introduction:**

As everyone knows, Vancouver is one of the best place to live in. But which Neighbourhood should you choose and which characters should you consider? Well, most people may say price and safety! Well, there are a lot of websites that can help you to find houses under the budget very easily. So, this project will help you to find how safety that house is.



In details, it will give you Neighbourhood guides in Vancouver based on the crimes reported in each Neighbourhood and clustering those Neighbourhoods into different groups.

## **Data:**

In this report, two datasets will be used. The first one is the crime.csv which includes all crimes reported in Vancouver. It also shows the detail of each crime such as: crime type, date, and Neighborhood. This data is directly download from Kaggle, <https://www.kaggle.com/agilesifaka/vancouver-crime-report>

The Second dataset is the location data, which combined with a list of Neighborhoods and their GPS location. When I’m doing the research, there’ no direct files that contain this information, so I manually create a Location.xlsx file getting the location data from Wikipedia. The purpose of getting this location data is helping to combine the crime data with the Foursquare location data.

**Data CleanUp:**

Based on the Vancouver Neighborhoods that identified in Location.xlsx dataset, some rows in crime.csv will be deleted (ie. rows that have Neighborhood values: ‘Musqueam’, ‘Stanley Park’, or ‘NaN’).

Also, in Neighborhood column of crime.csv table, ‘Central Business District’ will be changed to ‘Downtown’ just for convenience.

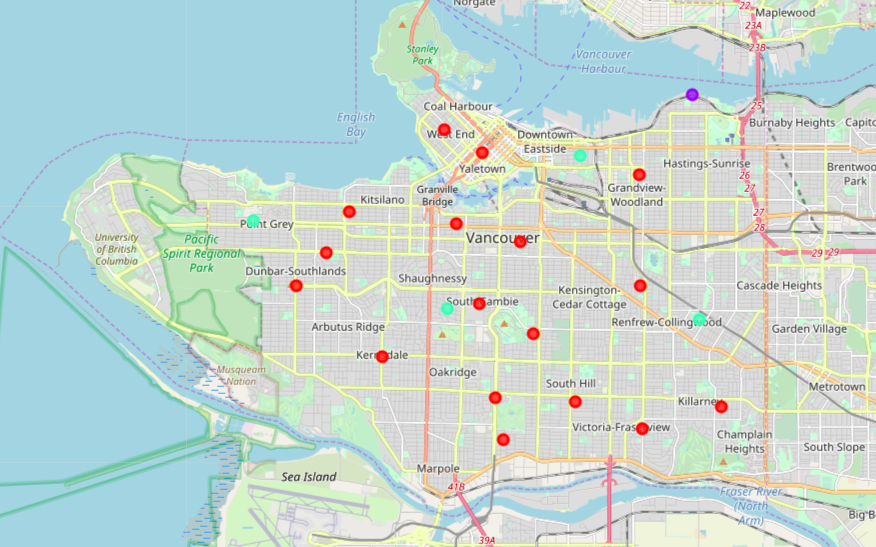
After that, pandas’ merge function is used to join crime table to Location table based on their Neighborhoods. Once two tables are merged, some columns with useless information are dropped ( ie. 'MONTH', 'DAY', 'HOUR', 'MINUTE', 'HUNDRED\_BLOCK', 'X', 'Y' ).

## **Methodology:**

Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why,

Based on the purpose of this report, the first things to do is to cluster all Vancouver Neighborhoods into different groups. I’m using ‘Foursquare’ to request information about each Neighborhood in Vancouver. More specifically, getting all the venues in the Neighborhood with details of each venue like their name, category and location. The data gathered from ‘Foursquare’ includes 84 unique venues categories in total.

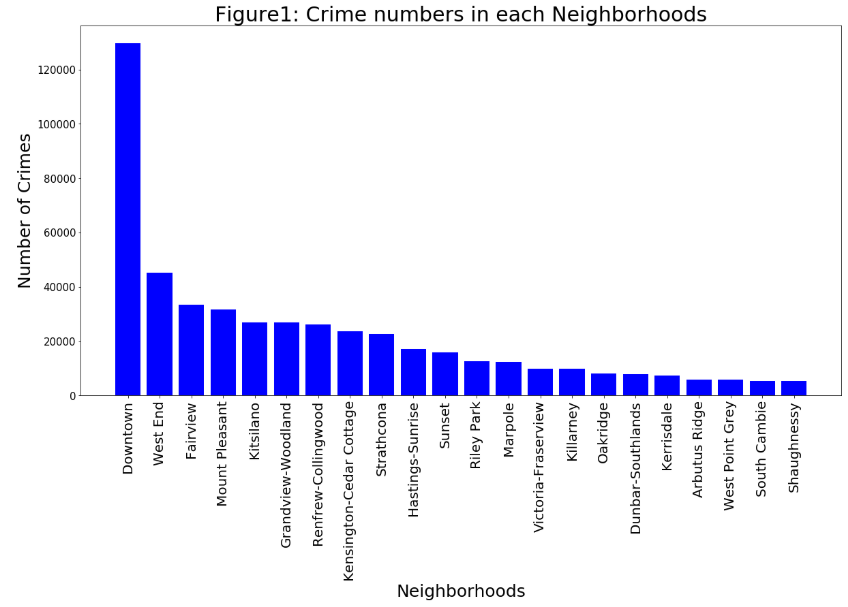
After that, I decided to cluster Neighborhoods into five groups based on venues around them. To do that, the python built-in function ‘KMeans’ is used to clusters Vancouver Neighborhoods into 5 groups based on the Venus categories in each Neighborhood.



It can be easily seen that Neighborhood ‘Shaughnessy’ is very different from all others as it’s top three venue categories are ‘Health & Beauty Service’, ‘Park’ and ‘Chocolate Shop’. ‘Hastings-Sunrise’ is another special Neighborhood that’s not similar to any other Neighborhoods. The top three venues categories in ‘Hastings-Sunrise’ are ‘Park’, ‘Chocolate Shop’ and ‘Concert hall’, which seems like a good place to live.

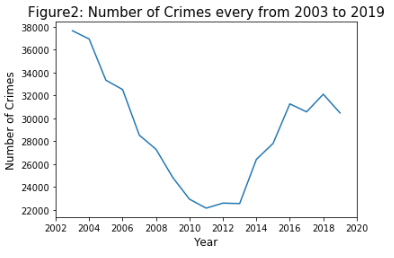


After cluster Neighborhoods into different groups, I want to do more calculation sfor Crime data. The first thing to do is find which Neighborhoods have the highest crimes in the past years. To do that, crimes in Vancouver are grouped by their Neighborhoods, and then summed up to get the total Crimes number in each Neighborhood. Here’s the figure to show the crimes data in Neighborhoods sorted from highest to lowest:

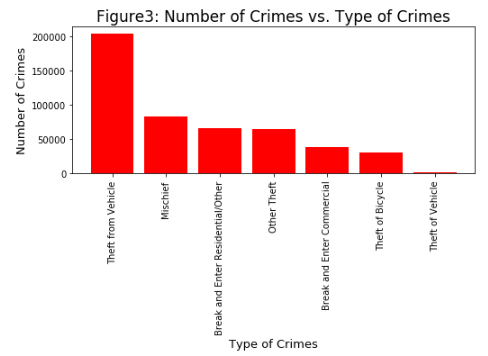


The Figure.1 shows that Downtown has much more crime rates that other regions. And the five most dangerous regions are ‘Downtown’, ‘West End’, ‘Fairview’, ‘Mount Pleasant’, ‘Kitsilano’. The next step is to find what type of crimes are happened in these places.

By looking at total crimes rates every year in Figure.2, we can find that the crime rate reach a local maximum in 2018, and local minimum from 2010 to 2013. But anyway, Vancouver becomes safer since 2003.



After grouping crime data by the crime types, Figure.3 shows that the top five crime types in Vancouver are: ‘Theft from Vehicle’, ‘Mischief’, ‘Break and Enter Residential/Other’, ‘Other Theft’ and ‘Break and Enter Commercial’. So, we can say that most crimes in Vancouver are financial related.



## **Results:**

Results section where you discuss the results,

## **Discussion:**

Discussion section where you discuss any observations you noted and any recommendations you can make based on the results,

## **Conclusion:**

Conclusion section where you conclude the report,