Capstone Project - The Battle of Neighborhoods: Crime in Vanouver

Submitted by: Gowri Subramanian
Date: March 2020

Table of Contents

Introduction

Vancouver, a bustling west coast seaport in British Columbia, is among Canada's densest, most ethnically diverse cities. A popular filming location, it's surrounded by mountains, and also has thriving art, theatre and music scenes. Vancouver Art Gallery is known for its works by regional artists, while the Museum of Anthropology houses preeminent First Nations collections.

Vancouver is a coastal seaport city in western Canada, located in the Lower Mainland region of British Columbia. As the most populous city in the province, the 2016 census recorded 631,486 people in the city, up from 603,502 in 2011. The Greater Vancouver area had a population of 2,463,431 in 2016, making it the third-largest metropolitan area in Canada. Vancouver has the highest population density in Canada, with over 5,400 people per square kilometre which makes it the fifth-most densely populated city with over 250,000 residents in North America, behind New York City, Guadalajara, San Francisco, and Mexico City. Vancouver is one of the most ethnically and linguistically diverse cities in Canada: 52% of its residents are not native English speakers, 48.9% are native speakers of neither English nor French, and 50.6% of residents belong to visible minority groups.

According to bc.ctvnews.ca, Assaults and overall violent **crime:** There were 4,521 reported assaults in 2019,

compared to 4,064 in 2018. Due to that increase, **Vancouver's** overall **crime rate** was up 7.2 per cent. The number of "shots-fired incidents," as described by police, was up last year to 29 calls − 10 more than reported in 2018.

Problem Statement:

The idea of this study is to help people planning to settle either temporarily in Vancouver by renting a house or to settle permanently by buying a house so that they would be able to choose a safer location and stay and feel safe by providing data about the safe and unsafe areas, predominant type of crime as well as the crime trend over years in various regions.

Objective:

This project aims to select the safest and least safest areas in Vancouver based on the total crimes, explore the venues in these areas and finally using the decision tree classifier. This report will be targeted to the People who are looking to analyze crime incidents, as the city continues to be a leader in so many fields. The crime statistics will provide an insight for the people towards their approach in buying or renting houses in different areas of city. The most common venues in the safest areas will be explored.

Data Set URL and the Cognitive Class Lab:

To provide the stakeholders the necessary information I'll be downloading .csv file from https://www.kaggle.com/wosaku/crime-in-vancouver and used https://labs.cognitiveclass.ai/tools/jupyterlab/ to execute and show my work.

	TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED_BLOCK	NEIGHBOURHOOD	X	Y	Latitude	Longitude
0	Other Theft	2003	5	12	16.0	15.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763
1	Other Theft	2003	5	7	15.0	20.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763
2	Other Theft	2003	4	23	16.0	40.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763
3	Other Theft	2003	4	20	11.0	15.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763
4	Other Theft	2003	4	12	17.0	45.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763

3) Methodology

For this report I used a few bar graphs that could help a new investor to decide the best neighborhood to buy or rent a house in Vancouver based on its crime type, trend etc. In order to do that I've used the Kaggle platform https://www.kaggle.com/wosaku/crime-in-vancouver to derive data along with Decision Tree Classifier to predict the 2017 year along with its accuracy.

Firstly, necessary libraries were imported.

The Vancouver crime file.csv was loaded and it provides information about the following: TYPE, YEAR, MONTH,

DAY, HOUR, HUNDRED_BLOCK, NEIGHBOURHOOD, X, Y, Latitude, and Longitude.

	TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED_BLOCK	NEIGHBOURHOOD	X	Y	Latitude	Longitude	NeighbourhoodID	Incident
0	Other Theft	2003	5	12	16.0	15.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763	20	1
1	Other Theft	2003	5	7	15.0	20.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763	20	1
2	Other Theft	2003	4	23	16.0	40.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763	20	1
3	Other Theft	2003	4	20	11.0	15.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763	20	1
4	Other Theft	2003	4	12	17.0	45.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763	20	1

Then, cleaning of data was done by adding a new column NeighbourhoodId by grouping .groupby method to group the neighbourhood and a new column Incident = '1'. Then, dropped the minute column, since theft is not checked by every minute. (by using .drop method df.drop(['MINUTE'], axis=1).

Then the missing values of hour column was filled using 99- a default dummy and replaced missing values of NEIGHBOURHOOD and HUNDRED_BLOCK by N/A
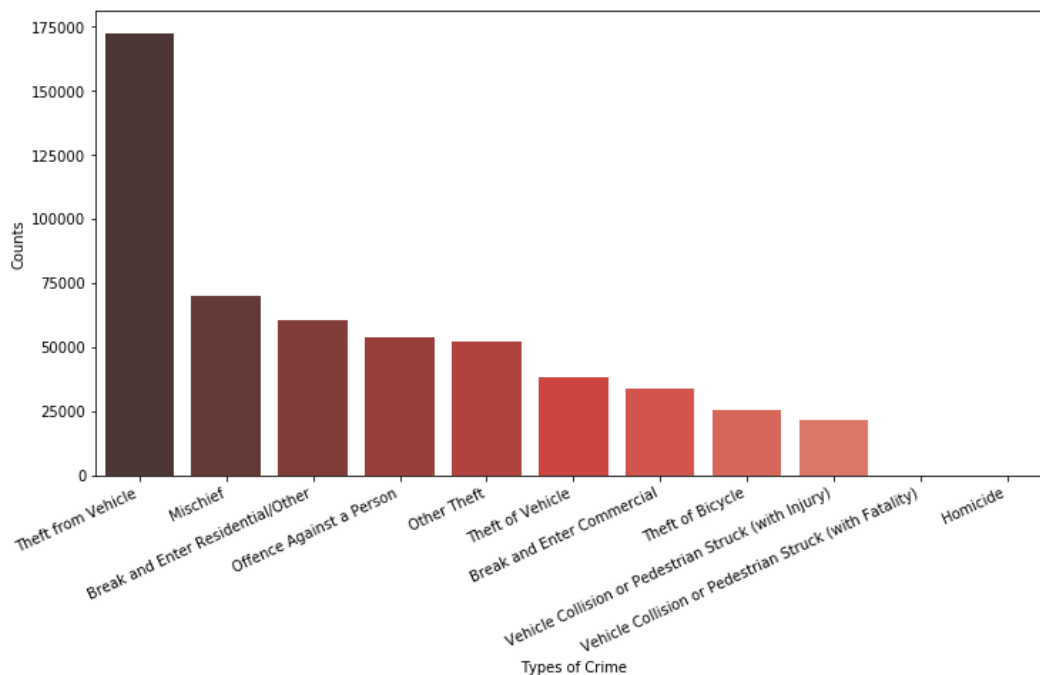
DATA EXPLORATORY ANALYSIS

Prevalence of crime types: Number of count of crimes over the period of years from 2003 to 2017 was graphed using matplotlib inline

Years Trend
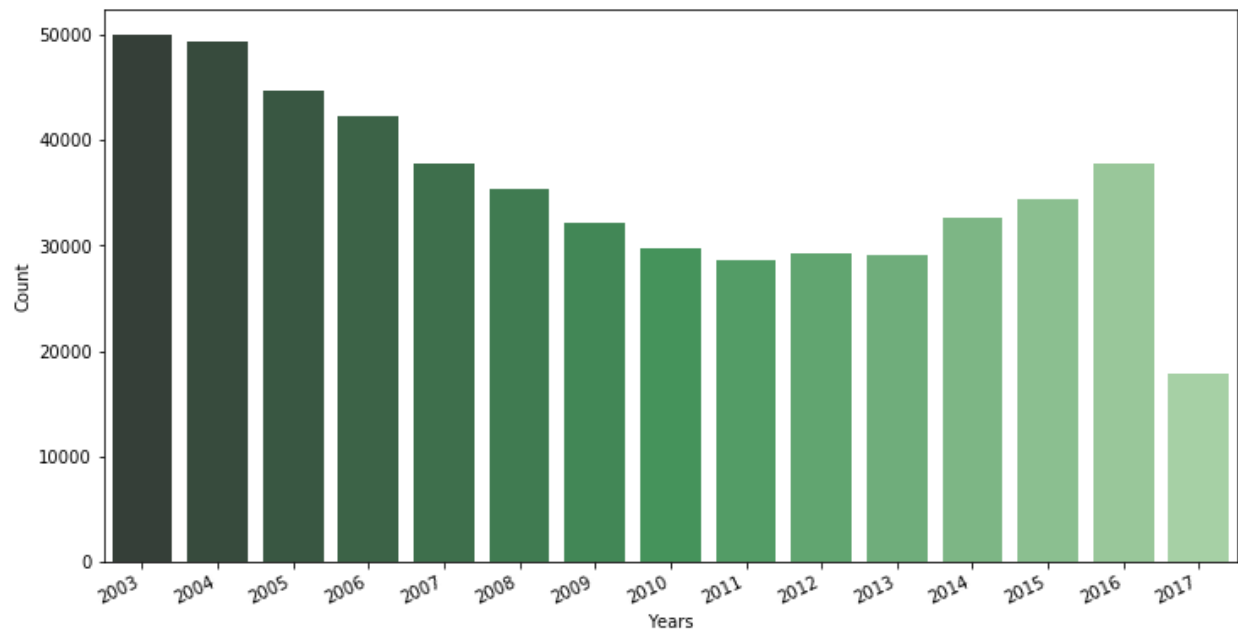
The counts of each type of crime was plotted and compared to see which crime is the most prevalent and which is least prevalent.

Types of Crime

The most safe and unsafe areas were evaluated. This is done by df.groupby(['NEIGHBOURHOOD','TYPE']).count()['NeighbourhoodID'].sort_values(ascending=True).head(5) and as well as = False.

```
NEIGHBOURHOOD   TYPE
Musqueam        Vehicle Collision or Pedestrian Struck (with Fatality)   1
                Other Theft                                              1
South Cambie    Vehicle Collision or Pedestrian Struck (with Fatality)   2
N/A             Break and Enter Commercial                               3
                Break and Enter Residential/Other                        3
Name: NeighbourhoodID, dtype: int64


NEIGHBOURHOOD             TYPE
N/A                       Offence Against a Person    54142
Central Business District Theft from Vehicle          48003
                          Other Theft                 19244
West End                  Theft from Vehicle          16904
Central Business District Mischief                    16672
Name: NeighbourhoodID, dtype: int64
```

Then, heat map was analysed.

Type of Crime By Year
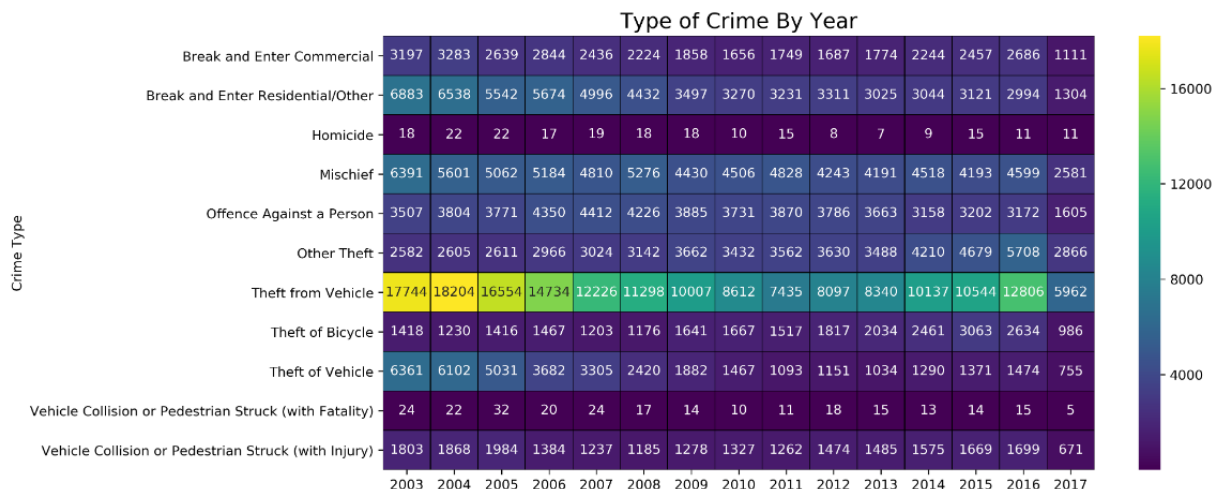
4) RESULTS

The study portrays majority of the crime type, and crime trend over years, also we could predict the 2017 crime using Decision Tree Classifier.

-By comparing the graphs, predominant type of crime is theft form vehicle, mischief, break and enter residential followed by others forms of crime.

 -During the years 2003 and 2004, highest crime rate was reported, and then gradually crime rate went down until 2011 then again it gradually went moderately up to 2016. the crime rate was lowest during 2017.

-Musqueum and South Cambie were the safest places to buy houses to settle in Vancouver.

-Central business district and Westend are the most unsafe places to buy houses to settle in Vancouver.

34]:		TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED_BLOCK	NEIGHBOURHOOD	X	Y	Latitude	Longitude	NeighbourhoodID	Incident
	514801	Break and Enter Residential/Other	2017	1	5	9.0	0.0	4XX E 6TH AVE	Mount Pleasant	493142.36	5456944.35	49.265223	-123.094259	12	1
	514947	Theft from Vehicle	2017	1	5	21.0	30.0	1XX W 1ST AVE	Mount Pleasant	492084.79	5457519.47	49.270384	-123.108807	12	1
	526353	Break and Enter Commercial	2017	1	5	5.0	45.0	2X W 7TH AVE	Mount Pleasant	492335.09	5456880.64	49.264641	-123.105353	12	1
	527070	Theft from Vehicle	2017	1	5	22.0	0.0	2XX E 6TH AVE	Mount Pleasant	492791.96	5456976.73	49.265511	-123.099076	12	1
	530182	Other Theft	2017	1	5	14.0	58.0	23XX CAMBIE ST	Mount Pleasant	491656.42	5456901.55	49.264820	-123.114682	12	1

```
35]: print ('Accuracy is', accuracy_score(y_test,y_pred_gn)*100, '%')
```
```
Accuracy is 42.867507455444006 %
```
```
36]: print ('Accuracy is', accuracy_score(y_test,y_pred_en)*100, '%')
```
```
Accuracy is 42.944299706493425 %
```

Analysing the heatmap,

-2010-2013 showed typically the lowest crime years.

-Crimes such as collisions-Vehicle collisions, mischief and homicide show constant over the years.

-Decision Tree Classifier gives a accuracy score of 42.86% and 42.94 % respectively.

5) DISCUSSION

When I started this study, I initially was planning on to use API and execute k- means clustering to find clusters of safer regions, so that it would give a clear idea of map showing a safer cluster neighbourhood. My future work would be to find the safe region clusters.

6) CONCLUSION

This report would be helpful for people planning to settle in Vancouver either temporarily for rent or for permanently settle in Vancouver by buying a house, by comparing the type of crimes in various locations, safe and unsafe areas, and crime trend. However, it would be necessary to analyze all variables such as cost of the house, size, rent amount, nearby grocery stores etc. Thus, further exploration is required.

Acknowledgements:

I immensely acknowledge

- Kaggle for letting me use the data fromhttps://www.kaggle.com/wosaku/crime-in-vancouver

- https://labs.cognitiveclass.ai/tools/jupyterlab/ for using the lab for writing the code as well as executing.

- IBM and its resources and coursera platform and the most valuable teachers without which I could not have attained this knowledge in Data Science.

- All other peers who gave valuable comments, clarifying the doubts as well as others who are part of this program.