Safe and secure location in Vancouver

Efthymios Ioannis Kavour

November 2020

1 Introduction

1.1 Background

There are always going to be people with new ideas, appetite for work and the need for entrepreneurship is rising as long as the population does so. Apart from those, criminality in streets rises as well. This is something that makes certain places, streets or boroughs not the best choice for one to begin a business. Therefore it will be for the best if there we could predict (under the assumption that our data are correct and are up-to-date) the best location in terms of security. Vancouver is a city of $2,581,000$ estimated population which we will use as an example to find the solution for our problem.

1.2 Problem

Data that will contribute to find such a location in Vancouver as described in 1.1 are real data collected from $2003-2019$ about all Vancouver crimes (*from Kaggle*), additional information of the list of officially categorized boroughs in Vancouver (*from Wikipedia*), the creation of a new consolidated dataset of the Neighborhoods, along with their boroughs, crime data and the respective Neighbourhood's co-ordinates as well as a dataset of the Neighborhoods, boroughs, and the most common venues and the respective Neighbourhood along with co-ordinates. Remember this project primary objective is to indicate safe locations for one to open a business in Vancouver.

1.3 Interest

There is no need to say that everyone who is interested in Vancouver's business life is going to be interested in this project. Also, year by year, by making small configurations in python we will be able to predict safe locations as well. Hence this is not a project that will help us solve only the current problem but creates a solid ground for future technological developments.

2 Data acquisition and cleaning

2.1 Data sources

Initially, we will be using a real world data set (*from Kaggle*) containing the Vancouver Crimes from 2003 to 2019. The data contains crime type (*TYPE*), the recorded year

(*YEAR*), the recorded month (*MONTH*), the recorded day (*DAY*), the recorded hour (*HOUR*), the recorded minute (*MINUTE*), the recorded block (*HUNDRED BLOCK*), the recorded neighborhood (*NEIGHBOURHOOD*), the GPS longitude (X), and the GPS latitude (Y). Initially we changed the names of columns to lowercase and of course we searched to find the total number of crimes for each neighborhood separately.

2.2 Data merging and cleaning

At this point we collected more information concerning the neighborhood−boroughs from *Wikipedia* in order to categorize the data set constructed until now. We merged the crime data table with the boroughs. Now we are able to determine the number of crimes for each borough of Vancouver. At this point we decided to drop all the invalid data and we created a new data table containing the Boroughs and the number of a specific type of crime occurred. The types of crimes categorized as followed:

- Break and Enter Commercial
- Break and Enter Residential/Other
- Mischief
- Other Theft
- Theft from Vehicle
- Theft of Bicycle
- Theft of Vehicle
- Vehicle Collision or Pedestrian Struck (with Fatality)
- Vehicle Collision or Pedestrian Struck (with Injury)

3 Methodology

This section is divided into two parts:

- Exploratory Data Analysis; and
- Modelling

During *Exploratory Data Analysis* we will check and visualise the crime reports in different Vancouver boroughs to identify the safest one. Next, we will normalise the neighborhoods and fine the 10 most common venues in each of neighborhood.

In *Modelling* part, to help the stakeholders choose the right/safest neighborhood within a borough we will be clustering similar neighborhoods using K-means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. We will use K-means clustering to address this problem so as to group data based on existing venues which will help in the decision making process.

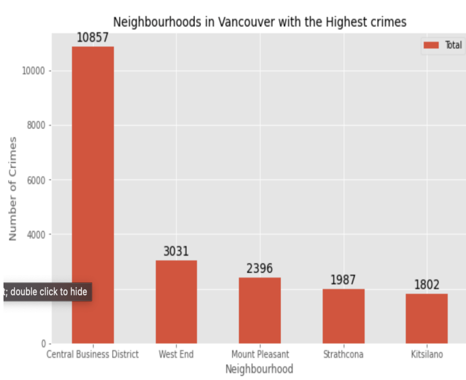
	Neighbourhood	YearBreak and Enter Commercial	YearBreak and Enter Residential/Other	YearMischief	YearOther Theft	YearTheft from Vehicle	YearTheft of Bicycle	YearTheft of Vehicle	YearVehicle Collision or Pedestrian Struck (with Fatality)	YearVehicle Collision or Pedestrian Struck (with Injury)	Total
0	Arbutus Ridge	12	78	49	18	111	12	12	1	18	311
1	Central Business District	551	124	1812	2034	5301	640	165	0	230	10857
2	Dunbar-Southlands	8	106	81	31	199	16	9	1	23	474
3	Fairview	138	73	233	297	692	245	55	0	62	1795
4	Grandview-Woodland	148	162	304	215	634	110	123	0	65	1761

Figure 1: Number of different type of crimes for each Neighbourhood

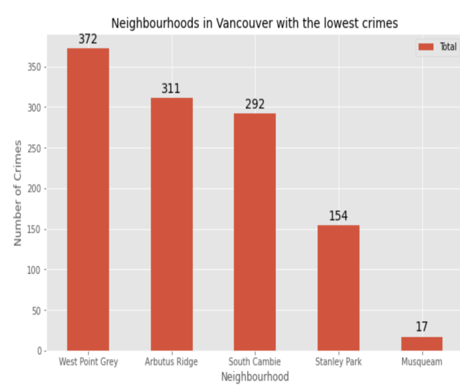
3.1 Exploratory Data Analysis

Initially we created a data-table as to see the relation of the Crime Types for each neighborhood separately. An example (extracted from the command .head()) is the following:

Of course the direct step is to use the ”*.describe* command as to check the basic statistical details. The first chart generated afterwards is one containing the neighborhoods in Vancouver that have the highest and a second chart with the lowest number of crimes:



(a) Highest number of crimes



(b) Lowest number of crimes

Figure 2: Neighbourhoods in Vancouver

Finally we will generate the chart with the Borough with the highest number of crimes. The results:

- ”Central” with *14042*
- ”East Side” with *12400*
- ”South Vancouver” with *1182*
- ’West Side” with *7204*

And as a result we see that South Vancouver has the lowest number of crimes. South Vancouver though has a small number of neighborhood and as a result this would harm our business (commercial establishment) as a result we should choose the next in line by taking the Boroughs in ascending order. Hence, West Side as we notice that the Break and enter Commercial crime type is low amongst other type of crimes. The next step from this point and on was to deepened our knowledge concerning the crime types at West Side of Vancouver.

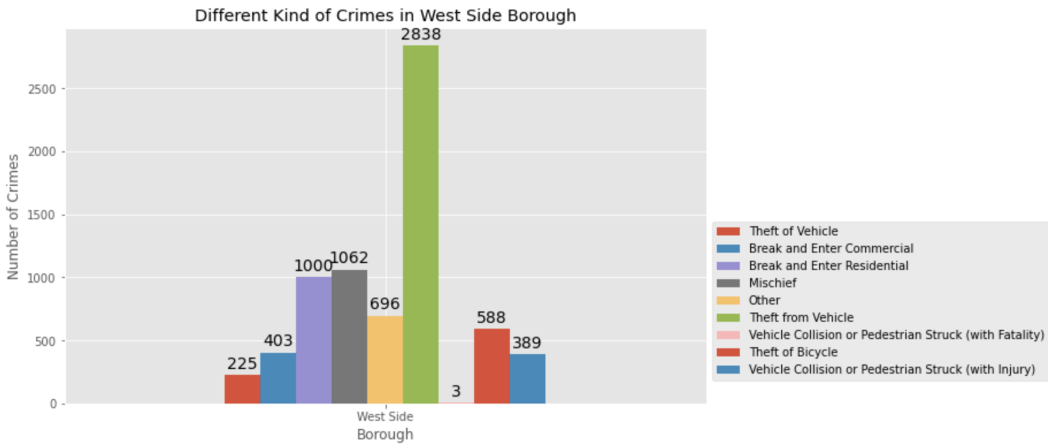


Figure 3: Type of crimes in West Side Borough

3.2 Modelling

We initially create a data set with the neighborhoods belonging to West Side. The new data set created containing the neighborhood, the borough, and the venues and this was done by the usage of Four Square API. The neighborhoods for which we generated venues are:

- Shaughnessy
- Fairview
- Oakridge
- Marpole
- Kitsilano
- Kerrisdale
- West Point Grey
- Arbutus Ridge
- South Cambie
- Dunbar-Southlands

We end up by finding the Top 5 most common venues across neighborhoods which are:

- Arbutus Ridge

1. Liquor Store 0.12
2. Seafood Restaurant 0.12
3. Fast Food Restaurant 0.12
4. Dance Studio 0.12
5. Coffee Shop 0.12

- Dunbar-Southlands
  1. Grocery Store 0.23
  2. Liquor Store 0.15
  3. Gym / Fitness Center 0.08
  4. Ski Area 0.08
  5. Japanese Restaurant 0.08
- Fairview
  1. Asian Restaurant 0.11
  2. Japanese Restaurant 0.11
  3. Coffee Shop 0.11
  4. Korean Restaurant 0.05
  5. Korean Restaurant 0.05
- Kerrisdale
  1. Park 0.25
  2. Golf Course 0.25
  3. Café 0.25
  4. Pool 0.25
  5. Massage Studio 0.00
- Kitsilano
  1. Bakery 0.07
  2. American Restaurant 0.04
  3. Coffee Shop 0.04
  4. Ice Cream Shop 0.04
  5. Thai Restaurant 0.04
- Marpole
  1. Sushi Restaurant 0.09
  2. Chinese Restaurant 0.06
  3. Vietnamese Restaurant 0.06
  4. Bubble Tea Shop 0.06

- Oakridge
  1. Bubble Tea Shop 0.2
  2. Sporting Goods Shop 0.1
  3. Vietnamese Restaurant 0.1
  4. Light Rail Station 0.1
  5. Bus Station 0.1
- Shaughnessy
  1. Bus Stop 0.4
  2. Video Game Store 0.2
  3. Chocolate Shop 0.2
  4. Park 0.2
  5. Noodle House 0.0
- South Cambie
  1. Coffee Shop 0.33
  2. Malay Restaurant 0.08
  3. Bus Stop 0.08
  4. Grocery Store 0.08
  5. Sushi Restaurant 0.08
- West Point Grey
  1. Disc Golf 0.11
  2. Harbor / Marina 0.11
  3. Sandwich Place 0.11
  4. Performing Arts Venue 0.11
  5. Park 0.11

We also choose to find the top 10 (most common) venues of each neighbohood.

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Arbutus Ridge	Coffee Shop	Grocery Store	Discount Store	Fast Food Restaurant	Sandwich Place	Seafood Restaurant	Dance Studio	Liquor Store	Gym	Greek Restaurant
1	Dunbar-Southlands	Grocery Store	Liquor Store	Japanese Restaurant	Pet Store	Coffee Shop	Café	Bus Stop	Ski Area	Gym	Gym / Fitness Center
2	Fairview	Asian Restaurant	Japanese Restaurant	Korean Restaurant	Salon / Barbershop	Pet Store	Pharmacy	Physical Therapist	Nail Salon	Coffee Shop	Restaurant
3	Kerrisdale	Pool	Park	Café	Golf Course	Food Truck	Disc Golf	Discount Store	Dog Run	Falafel Restaurant	Fast Food Restaurant
4	Kitsilano	Bakery	American Restaurant	Sushi Restaurant	Japanese Restaurant	Ice Cream Shop	French Restaurant	Coffee Shop	Food Truck	Thai Restaurant	Indian Restaurant

Figure 4: Top 10 venues per Neighbourhood.

Finally we begin the clustering in order to conclude to a borough and see if West Side is the outcome, which of course happens.

4 Results and Discussion

The objective of the business problem was to help stakeholders identify one of the safest (if not the safest, as we saw during this project) borough in Vancouver, and an appropriate neighborhood within the borough to set up a commercial establishment especially a Grocery store. This has been achieved by first making use of Vancouver crime data to identify a safe borough with considerable number of neighborhood for any business to be viable. After selecting the borough it was imperative to choose the right neighborhood where grocery shops were not among venues in a close proximity to each other. We achieved this by grouping the neighborhoods into clusters to assist the stakeholders by providing them with relevant data about venues and safety of a given neighborhood.

5 Conclusion

We have explored the crime data to understand different types of crimes in all neighborhoods of Vancouver and later categorized them into different boroughs, this helped us group the neighborhoods into boroughs and choose the safest borough first. Once we confirmed the borough the number of neighborhoods for consideration also comes down, we further shortlist the neighborhoods based on the common venues, to choose a neighborhood which best suits the business problem.