Capstone Project - The Battle of Neighborhoods

By: Sharmili Tandulwadkar

August 15, 2020

Table of contents

Introduction: Business Problem

Discussion of the business problem and the audience who would be interested in this project

1. **Discussion of the business problem**

The Vancouver City is a major city in western Canada, located in the Lower Mainland region of British Columbia. The project aims to find a safe and secure location for opening of grocery store in Vancouver, 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.

2. **Interested Audience**

The project will be targeted to stakeholders interested in opening a Grocery Store in Vancouver City, Canada.

We will make use of our data science tools to analyze data and focus on the safest borough and explore its neighborhoods and the 10 most common venues in each neighborhood so that the best neighborhood where grocery store is not amongst the most common venue can be selected.

Data Description

Following data sources will be needed to extract/generate the required information:

1. Dataset from Kaggle containing the Vancouver Crimes from 2003 to 2019

https://www.kaggle.com/agilesifaka/vancouver-crime-report/version/2

2. The information about the Neighborhood from Wikipedia

https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Vancouver

Description of the Data:
- A dataset consisting of the crime statistics of each Neighborhood in Vancouver along with type of crime, recorded year, month and hour from Kaggle containing the Vancouver Crimes from 2003 to 2019
- Borough information will be used to map the existing data where each neighborhood can be assigned with the right borough from Wikipedia.
- The data will be fetched using OpenCage Geocoder to find the safest borough and explore the neighborhood by plotting it on maps using Folium and perform exploratory data analysis.
- The data will be fetched using Four Square API to explore the neighborhood venues and to apply machine learning algorithm to cluster the neighborhoods and present the findings by plotting it on maps using Folium.

How the data will be used to solve the problem:

We will be using the geographical coordinates of Vancouver to plot neighborhoods in a borough that is safe and in the city's vicinity, and finally cluster our neighborhoods and present our findings.

1. Using a real-world data set from Kaggle containing the Vancouver Crimes from 2003 to 2019

Properties of Vancouver Crime Report

- TYPE - Crime type
- YEAR - Recorded year
- MONTH - Recorded month
- DAY - Recorded day
- HOUR - Recorded hour
- MINUTE - Recorded minute
- HUNDRED_BLOCK - Recorded block
- NEIGHBOURHOOD - Recorded neighborhood
- X - GPS longitude
- Y - GPS latitude

2. Gathering additional information about the Neighborhood from Wiki

Gathered borough information from Wikipedia to map the existing data where each neighborhood can be assigned with the right borough

Data Cleaning:

Data downloaded or scraped from multiple sources were combined into one table. There were a lot of missing values, were removed for better analysis purpose.

The Vancouver Crimes from 2003 to 2019 was available. Only recent and complete data of 2018 was considered for the analysis purpose.

Methodology

Categorized the methodology section into two parts:

- **Exploratory Data Analysis**: Visualize the crime reports in different Vancouver boroughs to identity the safest borough and normalize the neighborhoods of that borough. We will Use the resulting data and find 10 most common venues in each neighborhood.
- **Modelling**: To help stakeholders choose the right 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.
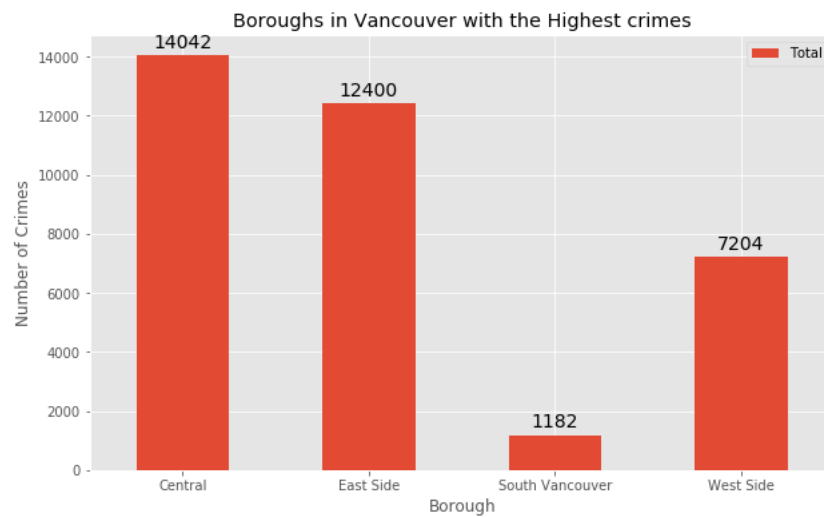
1. Exploratory Data Analysis:

- Five Neighborhoods with highest crime plot
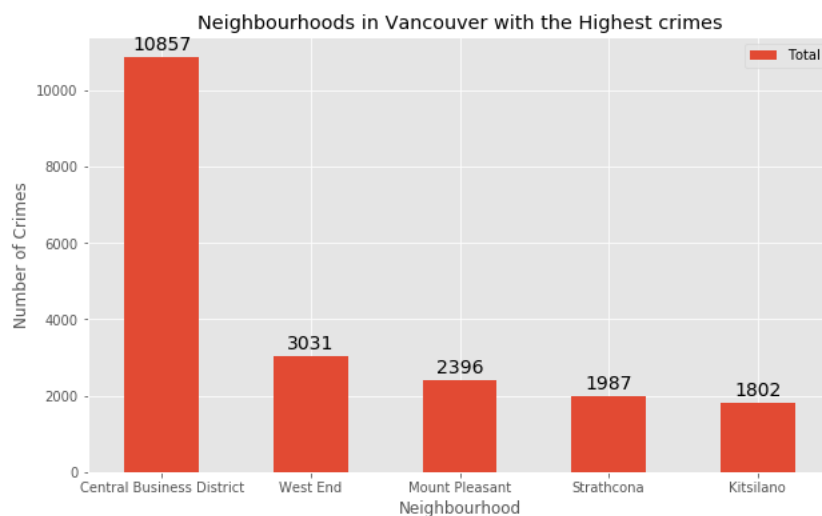
- Five Neighborhoods with lowest crime plot

Neighbourhoods in Vancouver with the lowest crimes

- Boroughs in Vancouver with the Highest Crimes
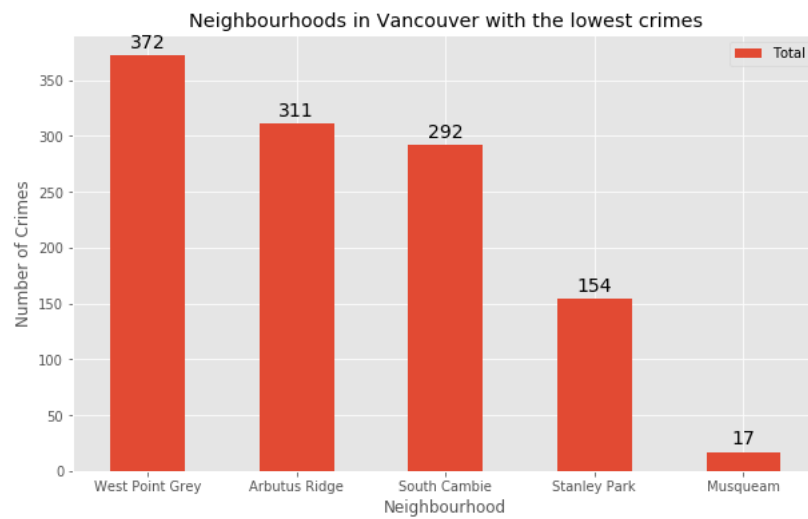
**Boroughs in Vancouver with the Highest crimes**

Based on exploratory data analysis it is clear that South Vancouver has the lowest crimes

Since South Vancouver has very little number of neighborhoods and opening a commercial establishment would not be viable, we can choose the next borough with lowest crime which is **West Side.**

Now, let's explore West Side Borough

- Different types of crimes recorded in the West Side Borough

Different Kind of Crimes in West Side Borough
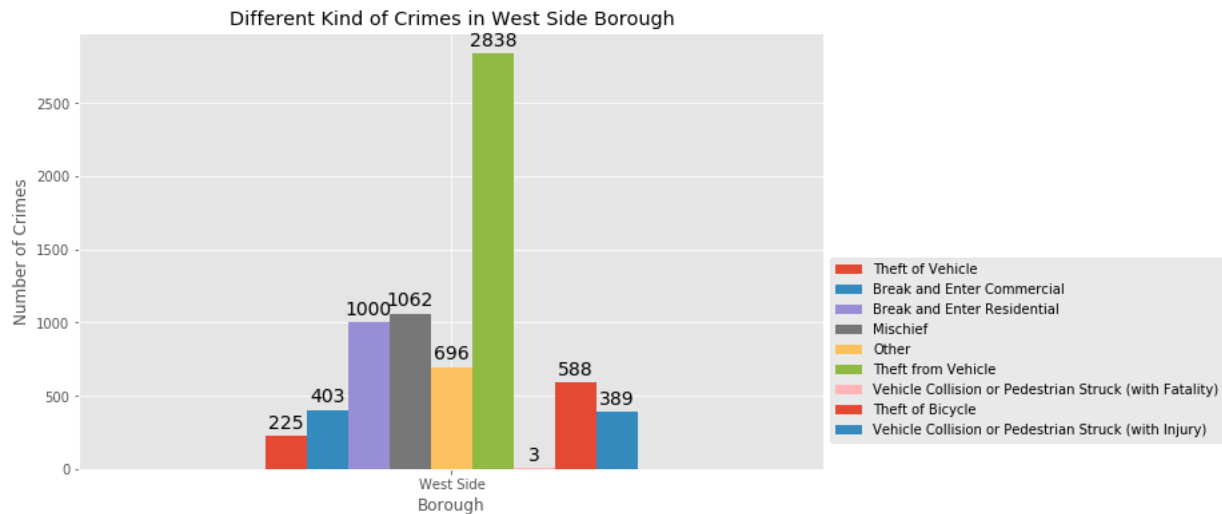
Now, created new consolidated dataset of the Neighborhoods, along with their boroughs, crime data and the respective Neighborhood's co-ordinates.

This data will be fetched using OpenCage Geocoder to find the safest borough and explore the neighborhood by plotting it on maps using Folium and perform exploratory data analysis. To plot this Vancouver City's West Side map used latitude and longitude to get visuals.

- Using Folium to plot Vancouver City's West Side Borough and its Neighborhoods
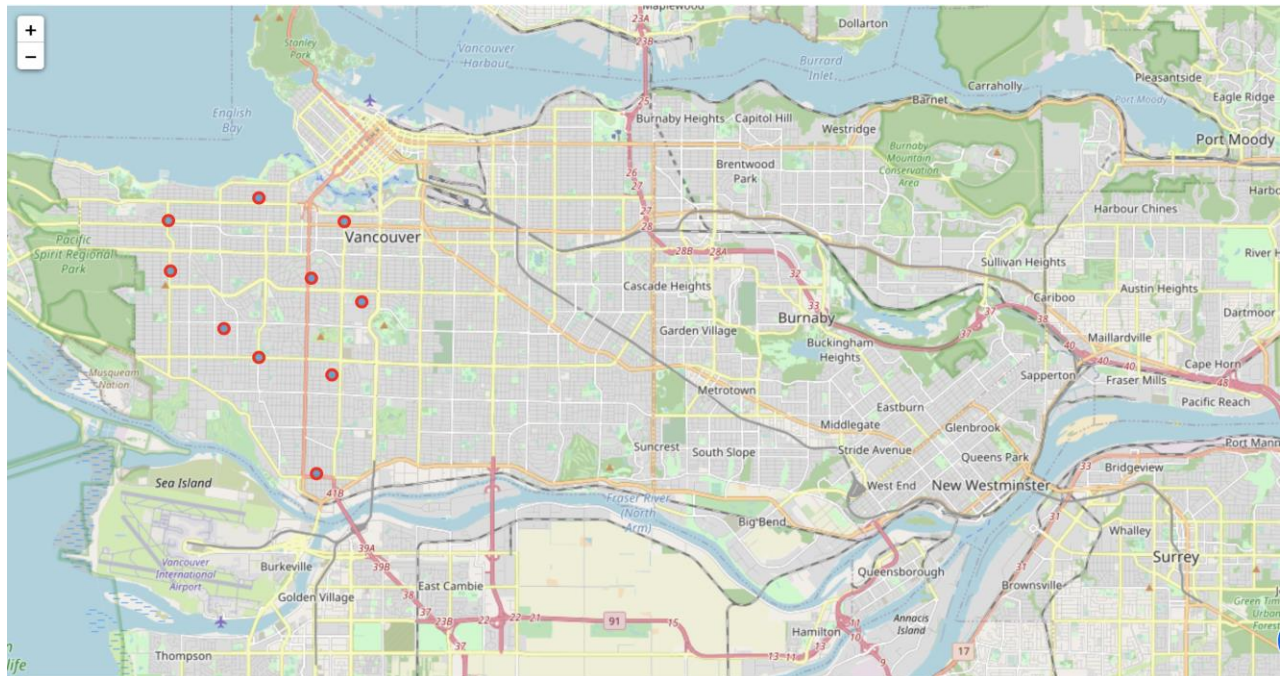
Then, creating a new consolidated dataset of the Neighborhoods, boroughs, and the most common venues and the respective Neighborhood along with co-ordinates:

This data will be fetched using Four Square API to explore the neighborhood venues and to apply machine learning algorithm to cluster the neighborhoods and present the findings by plotting it on maps using Folium.

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Category
0	Shaughnessy	49.251863	-123.138023	Angus Park	Park
1	Shaughnessy	49.251863	-123.138023	Crepe & Cafe	French Restaurant
2	Fairview	49.264113	-123.126835	Gyu-Kaku Japanese BBQ	BBQ Joint
3	Fairview	49.264113	-123.126835	CRESCENT nail and spa	Nail Salon
4	Fairview	49.264113	-123.126835	Charleson Park	Park

2. Modelling:

In summary of the exploratory analysis 94 unique categories were returned by Foursquare, then I created a table which shows list of top 10 venue category for each borough in below table.

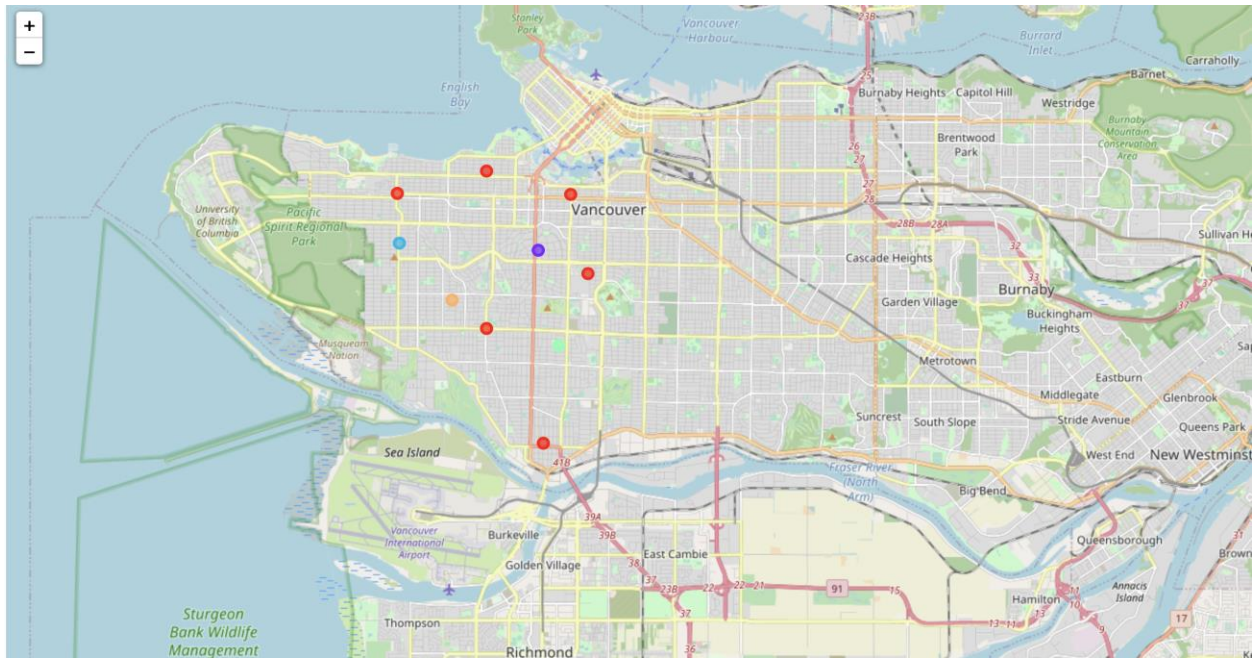|	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Arbutus Ridge	Nightlife Spot	Bakery	Pet Store	Grocery Store	Spa	Gas Station	Dessert Shop	Dim Sum Restaurant	Diner	Falafel Restaurant
1	Dunbar-Southlands	Sushi Restaurant	Indian Restaurant	Sporting Goods Shop	Ice Cream Shop	Coffee Shop	Italian Restaurant	Hobby Shop	Cosmetics Shop	Deli / Bodega	Dessert Shop
2	Fairview	Coffee Shop	Park	Asian Restaurant	Malay Restaurant	Pharmacy	Chinese Restaurant	Diner	Nail Salon	Falafel Restaurant	Restaurant
3	Kerrisdale	Coffee Shop	Chinese Restaurant	Sandwich Place	Pharmacy	Tea Room	Sushi Restaurant	Café	Hobby Shop	Portuguese Restaurant	Convenience Store
4	Kitsilano	Bakery	American Restaurant	Coffee Shop	Japanese Restaurant	Ice Cream Shop	Sushi Restaurant	Food Truck	French Restaurant	Thai Restaurant	Clothing Store

We have some common venue categories in boroughs. In this reason I used unsupervised learning K-means algorithm to cluster the boroughs. K-Means algorithm is one of the most common cluster methods of unsupervised learning.

First, I will run K-Means to cluster the boroughs into 5 clusters.

Here is the merged table with cluster labels for each borough.

	Neighbourhood	Borough	Latitude	Longitude	Cluster Labels	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	Shaughnessy	West Side	49.251863	-123.138023	1	Park	French Restaurant	Yoga Studio	Deli / Bodega	Dessert Shop	Dim Sum Restaurant	Diner	Falafel Restaurant	Fast Food Restaurant	Food Truck
1	Fairview	West Side	49.264113	-123.126835	0	Coffee Shop	Park	Asian Restaurant	Malay Restaurant	Pharmacy	Chinese Restaurant	Diner	Nail Salon	Falafel Restaurant	Restaurant
2	Oakridge	West Side	49.230829	-123.131134	4	Sushi Restaurant	Vietnamese Restaurant	Sandwich Place	Pharmacy	Park	Fast Food Restaurant	Convenience Store	Food Truck	Deli / Bodega	Dessert Shop
3	Marpole	West Side	49.209223	-123.136150	0	Sushi Restaurant	Dessert Shop	Japanese Restaurant	Vietnamese Restaurant	Pizza Place	Chinese Restaurant	Bus Stop	Liquor Store	Bubble Tea Shop	Plaza
4	Kitsilano	West Side	49.269410	-123.155267	0	Bakery	American Restaurant	Coffee Shop	Japanese Restaurant	Ice Cream Shop	Sushi Restaurant	Food Truck	French Restaurant	Thai Restaurant	Clothing Store

- Using Folium to plot clusters of Neighborhoods using latitude/longitude for each neighborhood

Analysis

Examining the resulting Clusters

Cluster 1:

	Borough	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
1	West Side	Coffee Shop	Park	Asian Restaurant	Malay Restaurant	Pharmacy	Chinese Restaurant	Diner	Nail Salon	Falafel Restaurant	Restaurant
3	West Side	Sushi Restaurant	Dessert Shop	Japanese Restaurant	Vietnamese Restaurant	Pizza Place	Chinese Restaurant	Bus Stop	Liquor Store	Bubble Tea Shop	Plaza
4	West Side	Bakery	American Restaurant	Coffee Shop	Japanese Restaurant	Ice Cream Shop	Sushi Restaurant	Food Truck	French Restaurant	Thai Restaurant	Clothing Store
5	West Side	Coffee Shop	Chinese Restaurant	Sandwich Place	Pharmacy	Tea Room	Sushi Restaurant	Café	Hobby Shop	Portuguese Restaurant	Convenience Store
6	West Side	Coffee Shop	Japanese Restaurant	Café	Bookstore	Pub	Sporting Goods Shop	Sushi Restaurant	Bus Station	Pizza Place	Vegetarian / Vegan Restaurant
8	West Side	Coffee Shop	Bus Stop	Shopping Mall	Cafeteria	Café	Sushi Restaurant	Grocery Store	Park	Liquor Store	Bank

Cluster 2:

	Borough	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	West Side	Park	French Restaurant	Yoga Studio	Deli / Bodega	Dessert Shop	Dim Sum Restaurant	Diner	Falafel Restaurant	Fast Food Restaurant	Food Truck

Cluster 3:

	Borough	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
9	West Side	Sushi Restaurant	Indian Restaurant	Sporting Goods Shop	Ice Cream Shop	Coffee Shop	Italian Restaurant	Hobby Shop	Cosmetics Shop	Deli / Bodega	Dessert Shop

Cluster 4:

	Borough	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
7	West Side	Nightlife Spot	Bakery	Pet Store	Grocery Store	Spa	Gas Station	Dessert Shop	Dim Sum Restaurant	Diner	Falafel Restaurant

Cluster 5:

	Borough	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
2	West Side	Sushi Restaurant	Vietnamese Restaurant	Sandwich Place	Pharmacy	Park	Fast Food Restaurant	Convenience Store	Food Truck	Deli / Bodega	Dessert Shop

Results and Discussion

The objective of the business problem was to help stakeholders identify one of the safest boroughs 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 neighborhoods 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.

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.