

# **Capstone Project - The Battle of Neighborhoods (Week 2)**

## **Yaw Anogyei Dapaa**

### **June 27, 2020**

## **1. Introduction**

### **1.1 Background & Problem**

There will always be the need for men and even women alike to have a haircut; this is because of the need to look good. Also, whilst some may think that not having but rather grooming their hair might be it when it comes to looking good, yet others cannot just do without having a good haircut. This might be weekly or fortnightly, or as the person who is involved wants it. It is because of this need that the barbershop business has continued to be on the rise.

Despite the essential service barbershops offer, they are susceptible to frequent burglaries and break-ins. Shop burglaries and break-ins are a problem for many communities. A survey found that stores lose as much to burglars and break-ins as they do to shoplifters. These losses impact the viability of businesses and, consequently, of the communities around them. The aim of this project is to find a safe and secure location for opening of commercial establishments in Vancouver, Canada

### **1.2 Interest**

Obviously, this report will be targeted to stakeholders interested in opening any business place like Barbershop in Vancouver City, Canada.

## **2. Data acquisition and cleaning**

### **2.1 Data sources**

Based on definition of our problem, factors that will influence our decision are:

- finding the safest borough based on crime statistics
- finding the most common venues
- choosing the right neighbourhood within the borough

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 our findings.

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

Part 1: Using a real world 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.

Part 2: Gathering additional information of the list of officially categorized boroughs in Vancouver from Wikipedia: Borough information will be used to map the existing data where each neighbourhood can be assigned with the right borough.

Part 3: Creating a new consolidated dataset of the Neighborhoods, along with their boroughs, crime data and the respective Neighbourhood's co-ordinates.: This data will be fetched using OpenCage Geocoder to find the safest borough and explore the neighbourhood by plotting it on maps using Folium and perform exploratory data analysis.

Part 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.

## 2.2 Data cleaning

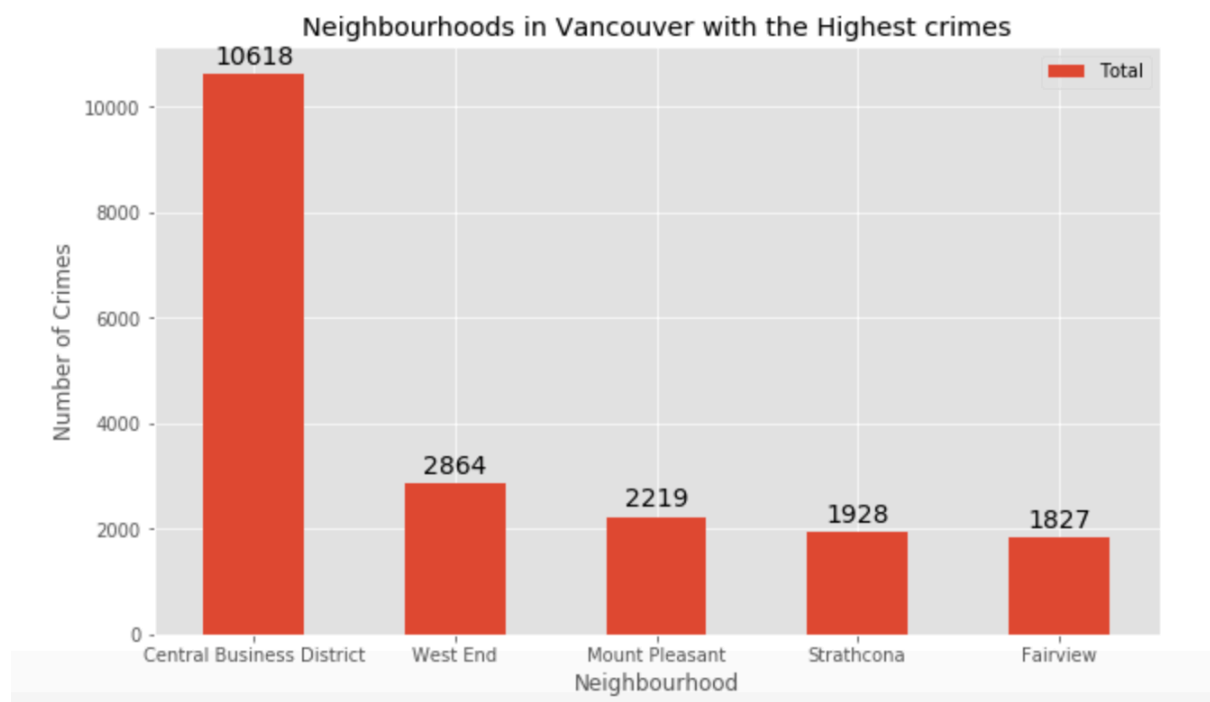
Due to sheer amount of data on crimes in Vancouver, it was not possible to process all of them. For this project we considered the recent crime report of the 2019 as the basis for this project.

The information that was needed from the dataset did not contain any missing values or entries so much was not done on the cleaning of the data.

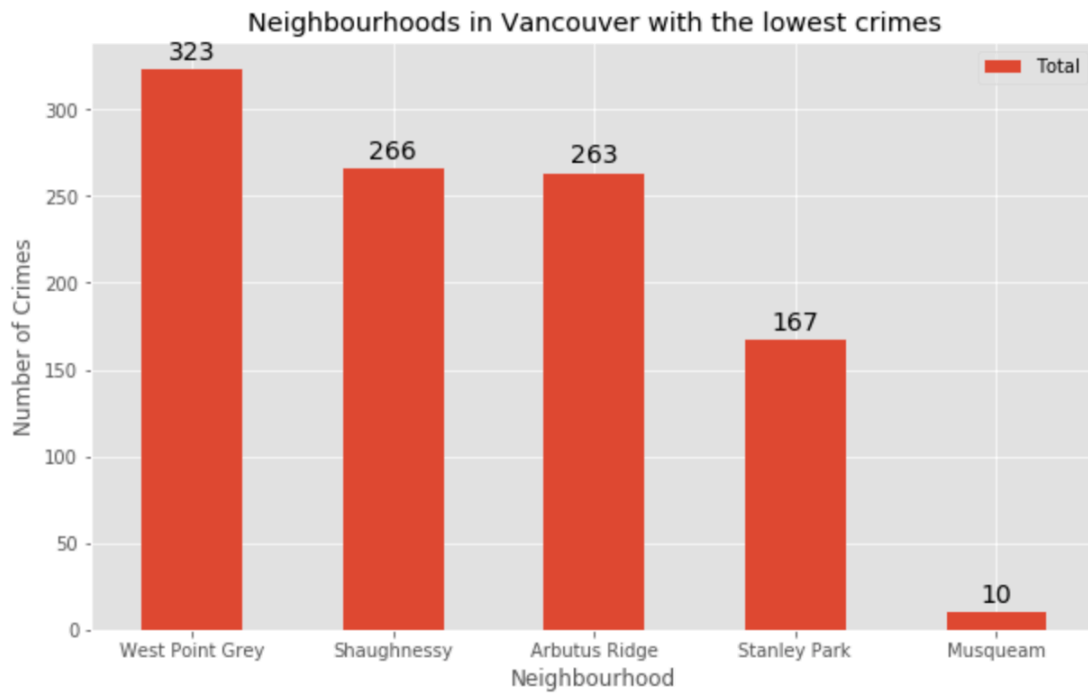
## 3. Methodology

### 3.1 Exploratory Data Analysis

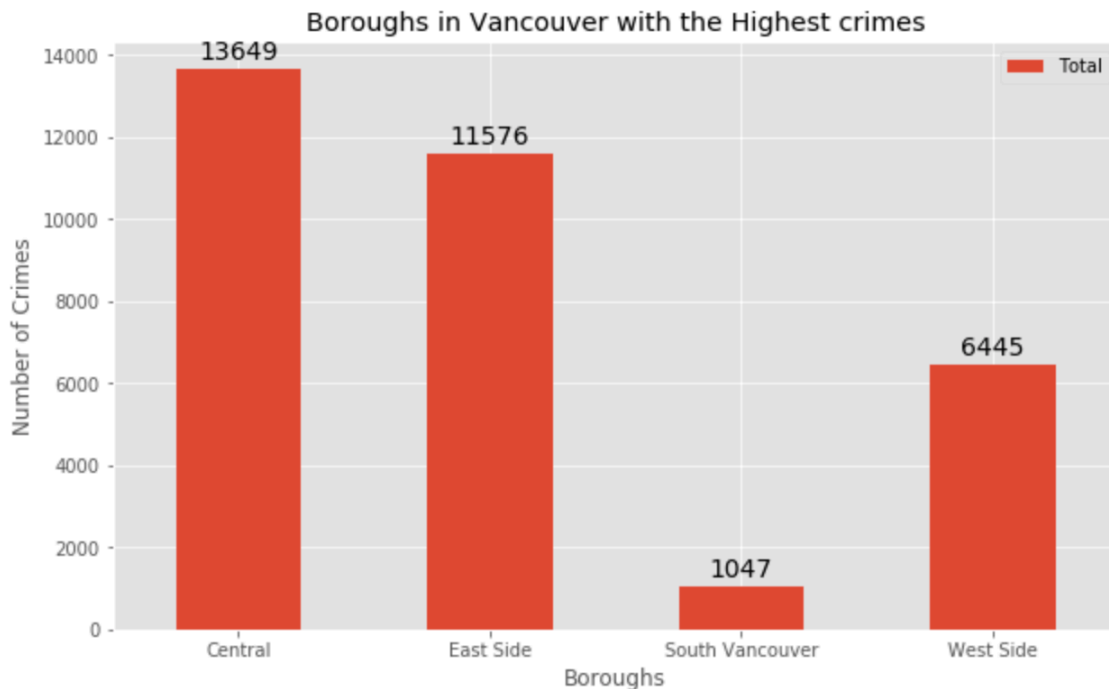
The first round of the exploratory analysis was to examine the top five(5) neighbourhood has the highest crime rates in Vancouver. From the analysis, we realize Central Business District has the highest crime rate with 10,618, followed by the West End Neighbourhood, Mount Pleasant, Strathcona and Fairview Neighbourhoods with crime rates of 2864, 2219, 1928, 1827 crimes respectively committed in the year 2019.



Examining the neighbourhoods with the lowest crime rates. With a crime rate of just 10, Musqueam neighbourhood has the lowest crime rate in Vancouver. 167, 263, 266 and 323 are the crime rates for Stanley Park, Arbutus Ridge, Shaughnessy and West Point Grey neighbourhoods respectively being the remaining four(4) neighbourhoods with the lowest crime rates in that order..



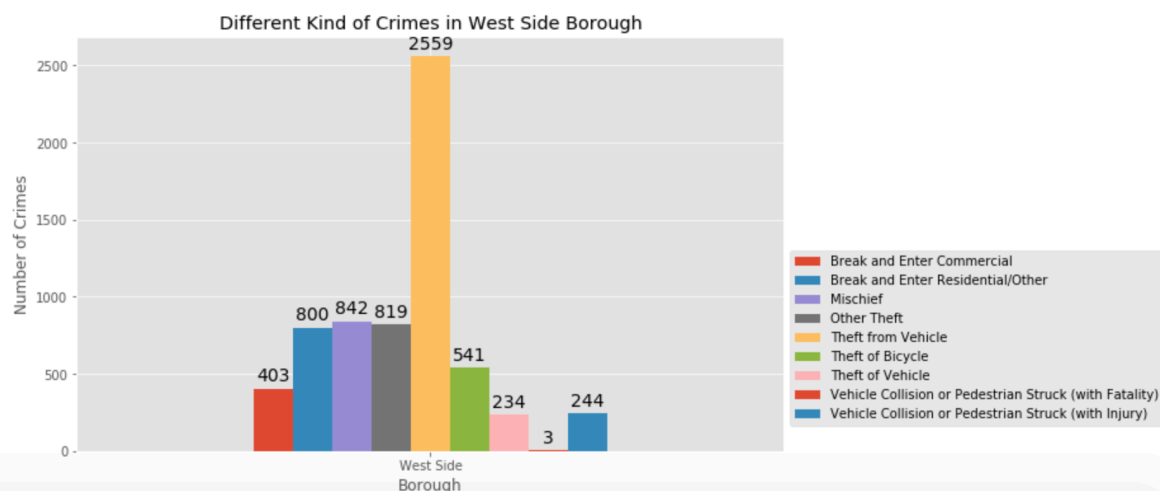
Next, I examine the crime rates of the four(4) boroughs in Vancouver. From the analysis, Central Borough has the highest crime with a rate of 13,649 and the lowest crime rate borough of 1,047 being that of South Vancouver.



Based on the exploratory data analysis, it is clear that the South Vancouver has the lowest crimes and the Neighbourhood with the lowest crime is Musqueam which, coincidentally is in South Vancouver.

But since South Vancouver has just three(3) neighborhoods and opening a commercial establishment like an Barbershop would not be viable, the next borough with lowest crime which is West Side is chosen because it has a high number of neighbourhoods as compared to the remaining neighbourhoods.

```
West Side      10
East Side      8
Central        3
South Vancouver 3
Name: Borough, dtype: int64
```



The different kind of crimes normally are committed in the West Side Borough of Vancouver is then analyzed to get a better understanding the type of crime this established will mostly be faced with.

It is realized the most committed crime in the West Side Borough is theft from vehicle. With the establishment as a Barbering Shop the mostly likely crime that will be faced will be break and enter commercial. From the analysis, it is realized that the West Side has the lowest crime rate of break and enter commercial as compared to the Central and East Side neighbourhood. This again helps justify why it is best the West Side is selected as the Borough for the commercial establishment.

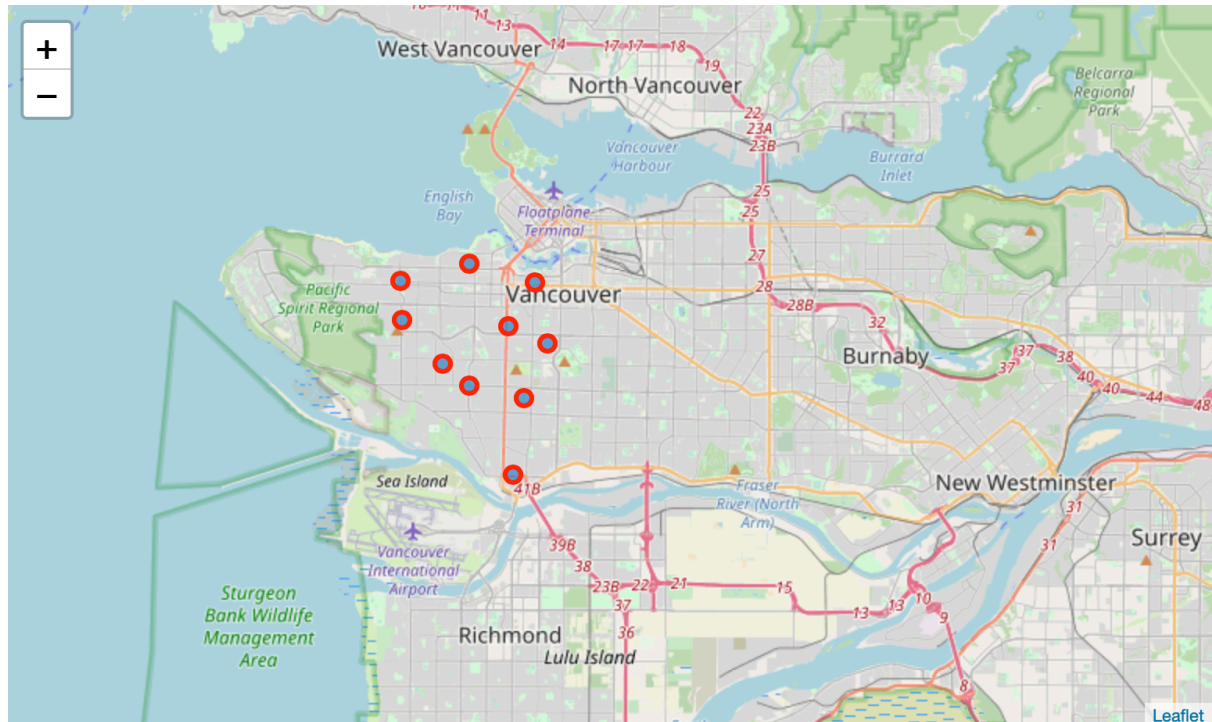
	Borough	Break and Enter Commercial	Break and Enter Residential/Other	Mischief	Other Theft	Theft from Vehicle	Theft of Bicycle	Theft of Vehicle	Vehicle Collision or Pedestrian Struck (with Fatality)	Vehicle Collision or Pedestrian Struck (with Injury)	Total
0	Central	774	192	2003	2473	7041	696	251	1	218	13649
1	East Side	749	881	1966	1492	4707	712	651	6	412	11576
2	South Vancouver	38	81	178	82	512	33	64	2	57	1047
3	West Side	403	800	842	819	2559	541	234	3	244	6445

### 3.2 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.

Before we start modelling, a new consolidated dataset of the Neighborhoods, along with their boroughs, crime data and the respective co-ordinates of the neighbourhoods, latitudes and longitude is created.

	Neighbourhood	Borough	Latitude	Longitude
0	Fairview	West Side	49.264113	-123.126835
1	Kitsilano	West Side	49.269410	-123.155267
2	Marpole	West Side	49.209223	-123.136150
3	Oakridge	West Side	49.230829	-123.131134
4	Kerrisdale	West Side	49.234673	-123.155389
5	South Cambie	West Side	49.246685	-123.120915
6	Dunbar-Southlands	West Side	49.253460	-123.185044
7	West Point Grey	West Side	49.264484	-123.185433
8	Shaughnessy	West Side	49.251863	-123.138023
9	Arbutus Ridge	West Side	49.240968	-123.167001



Using a final data set containing the neighbourhoods, boroughs, and the most common venues and the respective Neighbourhood along with co-ordinates, all venues within the neighbourhoods by connecting to the Foursquare API.

Venue	
Neighbourhood	
Arbutus Ridge	5
Dunbar-Southlands	8
Fairview	26
Kerrisdale	36
Kitsilano	47
Marpole	35
Oakridge	7
Shaughnessy	3
South Cambie	19
West Point Grey	46

One Hot Encoding was used to analyze each venue and to help in finding similar neighborhoods in the safest 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.

	Neighbourhood	American Restaurant	Asian Restaurant	BBQ Joint	Bakery	Bank	Bar	Beach	Bookstore	Boutique	...
0	Arbutus Ridge	0.000000	0.000000	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.000000	...
1	Dunbar-Southlands	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...
2	Fairview	0.000000	0.076923	0.038462	0.000000	0.038462	0.000000	0.000000	0.000000	0.000000	...
3	Kerrisdale	0.000000	0.027778	0.000000	0.027778	0.027778	0.000000	0.000000	0.000000	0.027778	...
4	Kitsilano	0.042553	0.021277	0.000000	0.063830	0.000000	0.000000	0.021277	0.000000	0.000000	...
5	Marpole	0.000000	0.000000	0.000000	0.000000	0.028571	0.028571	0.000000	0.000000	0.000000	...
6	Oakridge	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...
7	Shaughnessy	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...
8	South Cambie	0.000000	0.000000	0.000000	0.000000	0.052632	0.000000	0.000000	0.000000	0.000000	...
9	West Point Grey	0.000000	0.021739	0.000000	0.043478	0.021739	0.021739	0.000000	0.043478	0.000000	...

10 rows × 93 columns

A cluster size of 5 for this project that will cluster the 10 neighborhoods into 5 clusters. The reason to conduct a K- means clustering is to cluster neighborhoods with similar venues together so that people can shortlist the area of their interests based on the venues/amenities around each neighborhood.

## 4. RESULTS

After running the K-means clustering we can access each cluster created to see which neighborhoods were assigned to each of the five clusters. Visualizing the clustered neighborhoods on a map using the folium library.

Each cluster is color coded for the ease of presentation, we can see that majority of the neighborhood falls in the red cluster which is the first cluster. Three neighborhoods have their own cluster (Purple, Blue and Orange), these are clusters two, three and five. The green cluster consists of two neighborhoods which is the 4th cluster.

Cluster 1: Looking into the neighborhoods in the first cluster

The cluster one is the biggest cluster with 6 of the 10 neighborhoods in the West Side borough. Upon closely examining these neighborhoods we can see that the most common venues in these neighborhoods are Coffee Shops, Bakery, Sushi Restaurant among others.

	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	Coffee Shop	Asian Restaurant	Park	Malay Restaurant	Sushi Restaurant	Salon / Barbershop	Sandwich Place	Falafel Restaurant	Chinese Restaurant	Nail Salon
1	West Side	Bakery	American Restaurant	Coffee Shop	Restaurant	Japanese Restaurant	Ice Cream Shop	French Restaurant	Sushi Restaurant	Food Truck	Thai Restaurant
2	West Side	Sushi Restaurant	Bus Stop	Dessert Shop	Vietnamese Restaurant	Pizza Place	Chinese Restaurant	Japanese Restaurant	Shanghai Restaurant	Bubble Tea Shop	Dim Sum Restaurant
4	West Side	Coffee Shop	Chinese Restaurant	Sandwich Place	Pharmacy	Sushi Restaurant	Tea Room	Liquor Store	Mediterranean Restaurant	Dessert Shop	Japanese Restaurant
5	West Side	Coffee Shop	Park	Sushi Restaurant	Malay Restaurant	Grocery Store	Gift Shop	Cantonese Restaurant	Café	Cafeteria	Bus Stop
7	West Side	Coffee Shop	Japanese Restaurant	Café	Sushi Restaurant	Vegetarian / Vegan Restaurant	Bakery	Bus Station	Bookstore	Pub	Sporting Goods Shop

Cluster 2: Looking into the neighborhood in the second cluster.

The second cluster has one neighborhood which consists of Venues such as French Restaurant, Bus Stop, Park, Yoga Studio, Gas Station.

	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
8	West Side	French Restaurant	Bus Stop	Park	Yoga Studio	Gas Station	Dessert Shop	Dim Sum Restaurant	Diner	Falafel Restaurant	Fast Food Restaurant

Cluster 3: Looking into the neighborhood in the third cluster.

The third cluster has one neighborhood which consists of venues such as Restaurants, Coffee shop and a Liquor Store.

	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
6	West Side	Sushi Restaurant	Coffee Shop	Liquor Store	Indian Restaurant	Ice Cream Shop	Sporting Goods Shop	Italian Restaurant	Beach	Bookstore	Dim Sum Restaurant

Cluster 4: Looking into the neighborhood in the fourth cluster.

The fourth cluster has also one neighborhood in it, this neighborhood has venues like Nightlife Spot, Bakery, Stores, Spa and others.

	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	Nightlife Spot	Bakery	Pet Store	Grocery Store	Spa	Yoga Studio	Dessert Shop	Dim Sum Restaurant	Diner	Falafel Restaurant

Cluster 5: Looking into the neighborhood in the fourth cluster.

The fifth cluster has one neighborhood which consists of venues such as Concert Hall, Restaurants, Pharmacy, Sandwich Place, Convenience Store and Food Truck.

	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
3	West Side	Concert Hall	Vietnamese Restaurant	Pharmacy	Sandwich Place	Fast Food Restaurant	Sushi Restaurant	Convenience Store	Food Truck	Cosmetics Shop	Deli / Bodega



## **5. Discussion**

The objective of the business problem was to help stakeholders identify one of the safest borough in Vancouver, and an appropriate neighborhood within the borough to set up viable a commercial establishment especially an Barbering Shop.

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 Barbering 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.

## **6. Conclusion**

I have explored the crime data to understand different types of crimes in all neighborhoods within Vancouver and categorizing them into the different boroughs. This helped us group the neighborhoods into boroughs and choose the safest borough first. Once I 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.