Capstone Project - The Battle of the Neighborhoods (Week 2)

Applied Data Science Capstone by IBM/Coursera

Table of contents

Introduction: Business Problem

The aim of this project is to find a safe and secure location for opening of commercial establishments in Vancouver, Canada. Specifically, this report will be targeted to stakeholders interested in opening any business place like Barber Shop in Vancouver City, Canada.

The first task would be to choose the safest borough by analyzing crime data for opening a Barber Shop and short listing a neighborhoods, where Barber Shops are not amongst the most common venues, and yet as close to the city as possible.

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 Barber Shop is not amongst the most common venue can be selected.

Data

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 neighborhoods within the borough

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.

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 Neighborhood in Vancouver along with type of crime, recorded year, month and hour.

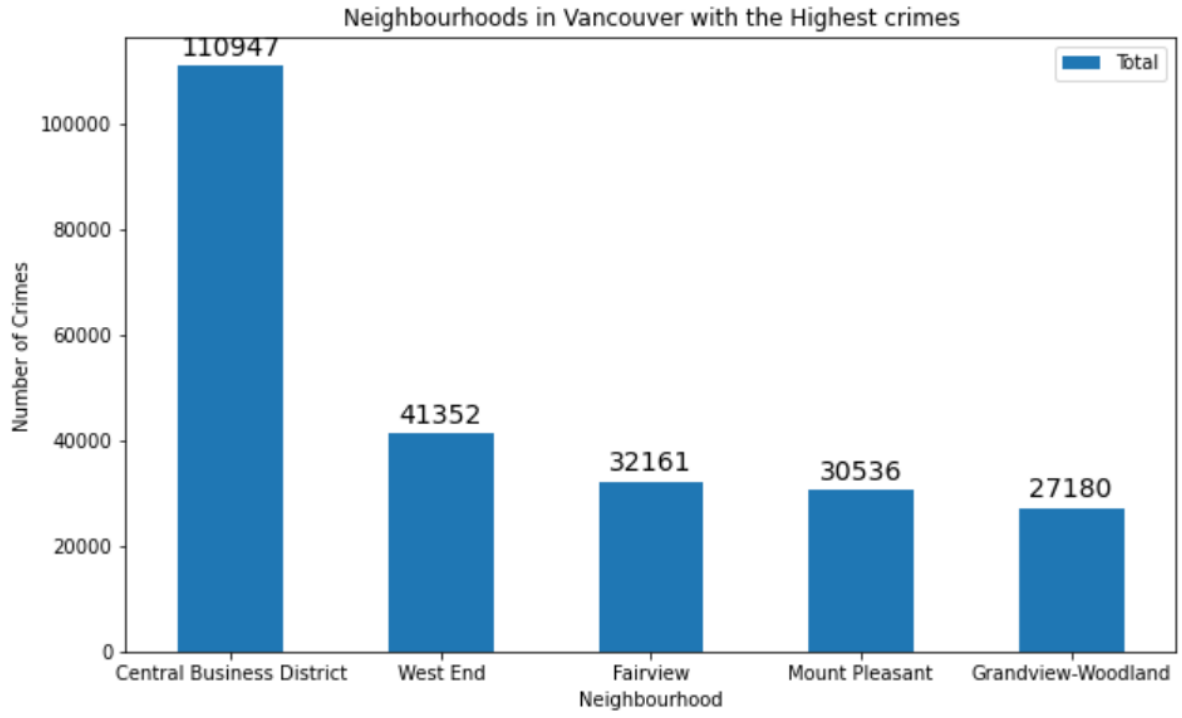
	TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED_BLOCK	NEIGHBOURHOOD	X	Y	Latitude	Longitude
0	Other Theft	2003	5	12	16.0	15.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763
1	Other Theft	2003	5	7	15.0	20.0	9XX TERMINAL AVE	Strathcona	493906.5	5457452.47	49.269802	-123.083763
2	Other Theft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

Drop not user column minute, Hundred_Block, X, Y

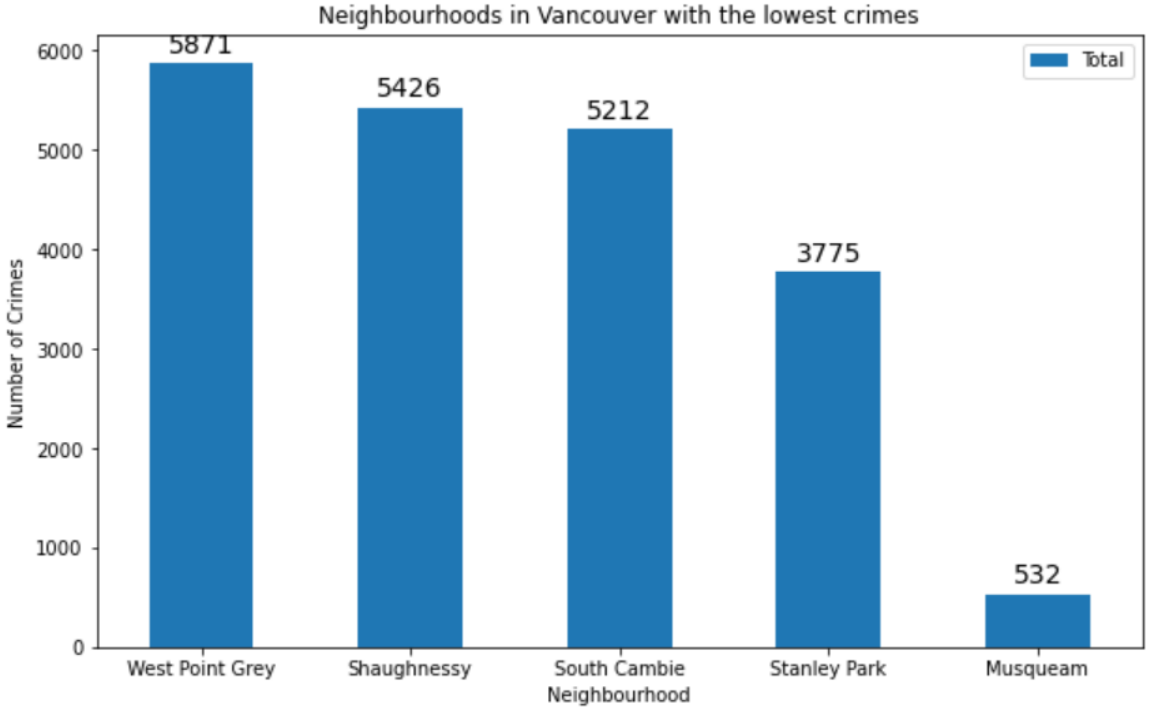
	TYPE	YEAR	MONTH	DAY	HOUR	NEIGHBOURHOOD	Latitude	Longitude
0	Other Theft	2003	5	12	16.0	Strathcona	49.269802	-123.083763
1	Other Theft	2003	5	7	15.0	Strathcona	49.269802	-123.083763
2	Other Theft	2003	4	23	16.0	Strathcona	49.269802	-123.083763
3	Other Theft	2003	4	20	11.0	Strathcona	49.269802	-123.083763
4	Other Theft	2003	4	12	17.0	Strathcona	49.269802	-123.083763

- 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 neighborhoods can be assigned with the right borough.

Top Five Neighborhoods with highest crimes

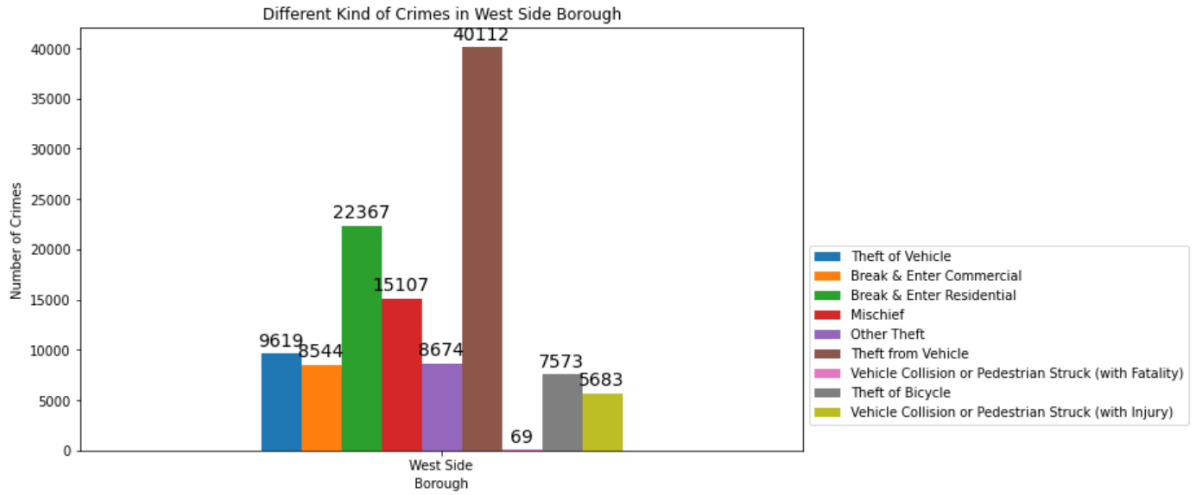


Top Five Neighborhoods with the lowest crimes

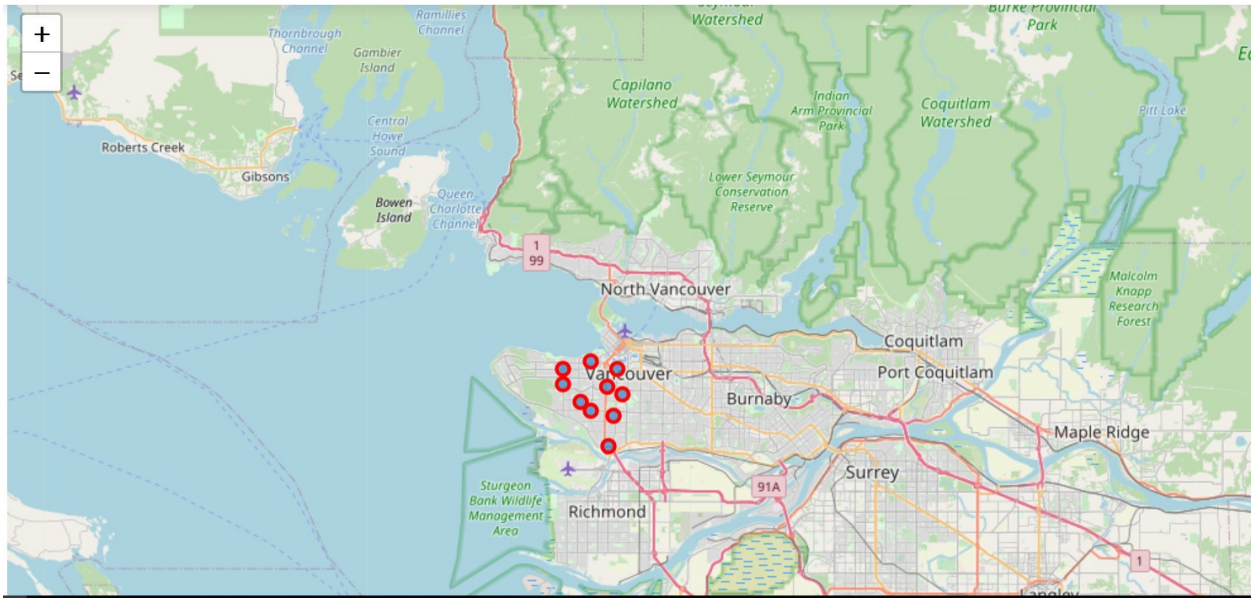


Based on visualized data analysis the South Vancouver has the lowest crimes and the East Side area has the highest crimes

Since South Vancouver has very little number of neighborhoods for opening the commercial area would not be viable, we can find the alternate with the next borough with second lowest crimes which is **West Side**.



- Part 3: Creating a 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 Open Cage Geocoder to find the safest borough and explore the neighborhoods 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 Neighborhood along with co-ordinates.: This data will be fetched using Four Square API to explore the neighborhoods venues and to apply machine learning algorithm to cluster the neighborhoods and present the findings by plotting it on maps using Folium.

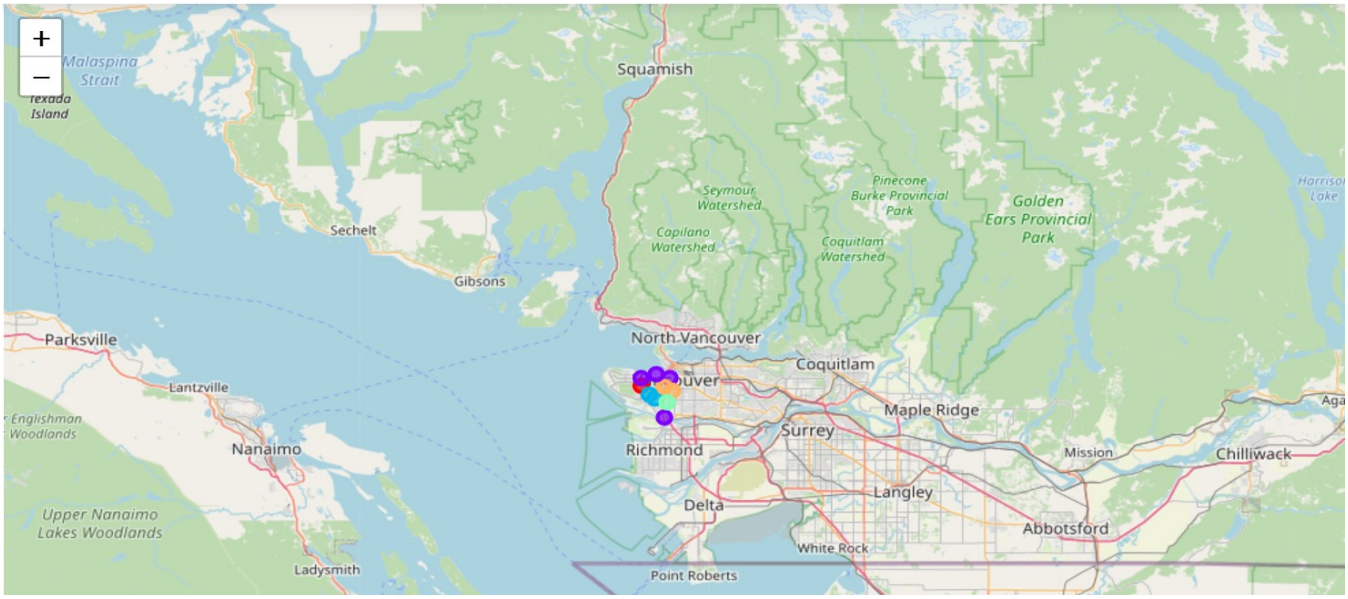
Data frame of venues for each neighborhood in West Side

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Category
0	Kerrisdale	49.234673	-123.155389	Nando's	Portuguese Restaurant
1	Kerrisdale	49.234673	-123.155389	Ajisai Sushi Bar	Japanese Restaurant
2	Kerrisdale	49.234673	-123.155389	Golden Ocean Seafood Restaurant	Chinese Restaurant
3	Kerrisdale	49.234673	-123.155389	Bufala	Pizza Place
4	Kerrisdale	49.234673	-123.155389	Minerva's Pizza & Steakhouse	Italian Restaurant

Now let's create the new data frame and display the top 10 venues for each neighborhood

	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	Sushi Restaurant	Bakery	Park	Coffee Shop	Pet Store	Burger Joint	Ice Cream Shop	Grocery Store	Italian Restaurant	Liquor Store
1	Dunbar-Southlands	Sushi Restaurant	Indian Restaurant	Coffee Shop	Bank	Cosmetics Shop	Pub	Sandwich Place	Café	Mexican Restaurant	Pizza Place
2	Fairview	Japanese Restaurant	Park	Coffee Shop	Furniture / Home Store	Café	Breakfast Spot	Chinese Restaurant	American Restaurant	Indian Restaurant	Restaurant
3	Kerrisdale	Coffee Shop	Chinese Restaurant	Sandwich Place	Pharmacy	Sushi Restaurant	Tea Room	Bus Stop	Japanese Restaurant	Liquor Store	Fast Food Restaurant
4	Kitsilano	Coffee Shop	Bakery	Yoga Studio	Japanese Restaurant	Vegetarian / Vegan Restaurant	Pizza Place	Board Shop	Café	Restaurant	Pharmacy

Finally, let's visualize the resulting clusters



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.

	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	Sushi Restaurant	Bakery	Park	Coffee Shop	Pet Store	Burger Joint	Ice Cream Shop	Grocery Store	Italian Restaurant	Liquor Store
1	Dunbar-Southlands	Sushi Restaurant	Indian Restaurant	Coffee Shop	Bank	Cosmetics Shop	Pub	Sandwich Place	Café	Mexican Restaurant	Pizza Place
2	Fairview	Japanese Restaurant	Park	Coffee Shop	Furniture / Home Store	Café	Breakfast Spot	Chinese Restaurant	American Restaurant	Indian Restaurant	Restaurant
3	Kerrisdale	Coffee Shop	Chinese Restaurant	Sandwich Place	Pharmacy	Sushi Restaurant	Tea Room	Bus Stop	Japanese Restaurant	Liquor Store	Fast Food Restaurant
4	Kitsilano	Coffee Shop	Bakery	Yoga Studio	Japanese Restaurant	Vegetarian / Vegan Restaurant	Pizza Place	Board Shop	Café	Restaurant	Pharmacy

- **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.

	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 Cor
0	Kerrisdale	West Side	49.234673	-123.155389	2	Coffee Shop	Chinese Restaurant	Sandwich Place	Pharmacy	Sushi Restaurant	Tea Room	Bus Stop	Japanese Restaurant	I
1	Dunbar-Southlands	West Side	49.253460	-123.185044	0	Sushi Restaurant	Indian Restaurant	Coffee Shop	Bank	Cosmetics Shop	Pub	Sandwich Place	Café	Me Resta
2	Fairview	West Side	49.264113	-123.126835	1	Japanese Restaurant	Park	Coffee Shop	Furniture / Home Store	Café	Breakfast Spot	Chinese Restaurant	American Restaurant	I Resta
3	West Point Grey	West Side	49.264484	-123.185433	1	Coffee Shop	Sushi Restaurant	Bakery	Greek Restaurant	Café	Vegetarian / Vegan Restaurant	Middle Eastern Restaurant	Bookstore	Jap Resta
4	Shaughnessy	West Side	49.251863	-123.138023	4	Coffee Shop	Breakfast Spot	Sushi Restaurant	Bus Stop	Furniture / Home Store	Bank	Park	Café	E

Analysis the resulting of:

Cluster 1

```
vcr_merged.loc[vcr_merged['Cluster Labels'] == 0, vcr_merged.columns[[1] + list(range(5, vcr_merged.shape[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	Sushi Restaurant	Indian Restaurant	Coffee Shop	Bank	Cosmetics Shop	Pub	Sandwich Place	Café	Mexican Restaurant	Pizza Place

Cluster 2

```
vcr_merged.loc[vcr_merged['Cluster Labels'] == 1, vcr_merged.columns[[1] + list(range(5, vcr_merged.shape[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
2	West Side	Japanese Restaurant	Park	Coffee Shop	Furniture / Home Store	Café	Breakfast Spot	Chinese Restaurant	American Restaurant	Indian Restaurant	Restaurant
3	West Side	Coffee Shop	Sushi Restaurant	Bakery	Greek Restaurant	Café	Vegetarian / Vegan Restaurant	Middle Eastern Restaurant	Bookstore	Japanese Restaurant	Bar
6	West Side	Coffee Shop	Bakery	Yoga Studio	Japanese Restaurant	Vegetarian / Vegan Restaurant	Pizza Place	Board Shop	Café	Restaurant	Pharmacy
8	West Side	Japanese Restaurant	Pizza Place	Chinese Restaurant	Sushi Restaurant	Café	Vietnamese Restaurant	Grocery Store	Bank	American Restaurant	Park

Cluster 3

```
vcr_merged.loc[vcr_merged['Cluster Labels'] == 2, vcr_merged.columns[[1] + list(range(5, vcr_merged.shape[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
0	West Side	Coffee Shop	Chinese Restaurant	Sandwich Place	Pharmacy	Sushi Restaurant	Tea Room	Bus Stop	Japanese Restaurant	Liquor Store	Fast Food Restaurant
5	West Side	Sushi Restaurant	Bakery	Park	Coffee Shop	Pet Store	Burger Joint	Ice Cream Shop	Grocery Store	Italian Restaurant	Liquor Store

Cluster 4

```
vcr_merged.loc[vcr_merged['Cluster Labels'] == 3, vcr_merged.columns[[1] + list(range(5, vcr_merged.shape[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
9	West Side	Bus Stop	Jewelry Store	Coffee Shop	Garden	Sporting Goods Shop	Sandwich Place	Optical Shop	Toy / Game Store	Bakery	Tea Room

Cluster 5

```
vcr_merged.loc[vcr_merged['Cluster Labels'] == 4, vcr_merged.columns[[1] + list(range(5, vcr_merged.shape[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
4	West Side	Coffee Shop	Breakfast Spot	Sushi Restaurant	Bus Stop	Furniture / Home Store	Bank	Park	Café	Burger Joint	Molecular Gastronomy Restaurant
7	West Side	Coffee Shop	Park	Bank	Garden	Chinese Restaurant	Bubble Tea Shop	Sushi Restaurant	Sandwich Place	Grocery Store	Greek Restaurant

Results and Discussion

The objective of the business problem was to help stakeholders identify one of the safest borough is the west side area in Vancouver and an appropriate neighborhood within the borough to set up a commercial area especially a Barber 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 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 categorized them into different boroughs this helped us to groups the neighborhoods into boroughs and choose the safest borough which is the best suits the business problem.