# CAPSTONE PRESENTATION

Analyzing Safest Neighbourhood in Vancouver

By: Ashish Rangari

Jan 30/2020

# Introduction

- Background:

Vancouver is a coastal seaport city in western Canada, located in the Lower Mainland region of British Columbia. The Greater Vancouver area had a population of 2,463,431 as in 2016, making it the third-largest metropolitan area in Canada [1]. Vancouver is no exception to crimes in different forms which are prevalent in other metropolitan cities. Criminal activity is an ongoing practice by offenders causing disruption of public peace, business owners. Therefore, it is important to consider the crime rate in the area before opening a business establishment. In this project, this very issue of finding a safe neighborhood is analyzed. For this purpose, the crime data of Vancouver City and finding the safest borough and a neighborhood within the borough is analyzed to resolve the business problem.

# Introduction

- 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 is catered towards individuals that are interested in opening any business place like liquor store in Vancouver City, Canada. The first step is to choose the safest borough by analyzing crime data for opening a liquor store and short listing a neighborhood, where liquor stores are not amongst the most common venues, and yet as close to the city as possible. Data science tools used to analyze data and focus on the safest borough and explore its neighborhoods and the 10 most common venues in each neighborhood. Then the best neighborhood where liquor stores are not amongst the most common venues can be selected.

# Data Acquisition and Cleaning

- Data Acquisition:
  - To make this project realistic and useful for the user, actual crime rate data set published on Kaggle datasets for Vancouver is used. This dataset included type of crime, recorded time and coordinates of the criminal activity along with neighborhoods. But the neighborhoods were not properly categorized into boroughs which were fetched from Wikipedia. Further the coordinates of the data were fetched using the OpenCage Geocoder API. Foursquare API is used to fetch venues for the listed neighborhoods.
  - The second source of data was extracted from a Wikipedia. This data did not require any scraping, as it was direct categorizations. The page contains additional information about the neighborhood and its boroughs.
  - The third data source was generated from OpenCage API. The data was generated as follows below are the list of columns:
    - Neighborhood: Name of the neighborhood in the Borough.
    - Borough: Name of the Borough.
    - Latitude: Latitude of the Borough.
    - Longitude: Longitude of the Borough.

# Data Acquisition and Cleaning

- Data Cleaning:
  - The data file extracted from Kaggle had close to 600,000 + data point. To simplify the project only 2018 crime data has been analyzed. The reference csv file is uploaded in the git repository.
  - It was observed that there was improper encoding of the co-ordinates of the crime record. Due to the erroneous nature of the information, these co-ordinates from the data couldn’t be used for plotting. Along with X,Y columns in the dataset which represented the GPS co-ordinates of the criminal activity, other fields such as month and hour in which the crime took place has been dropped because they were not in the scope of the problem.

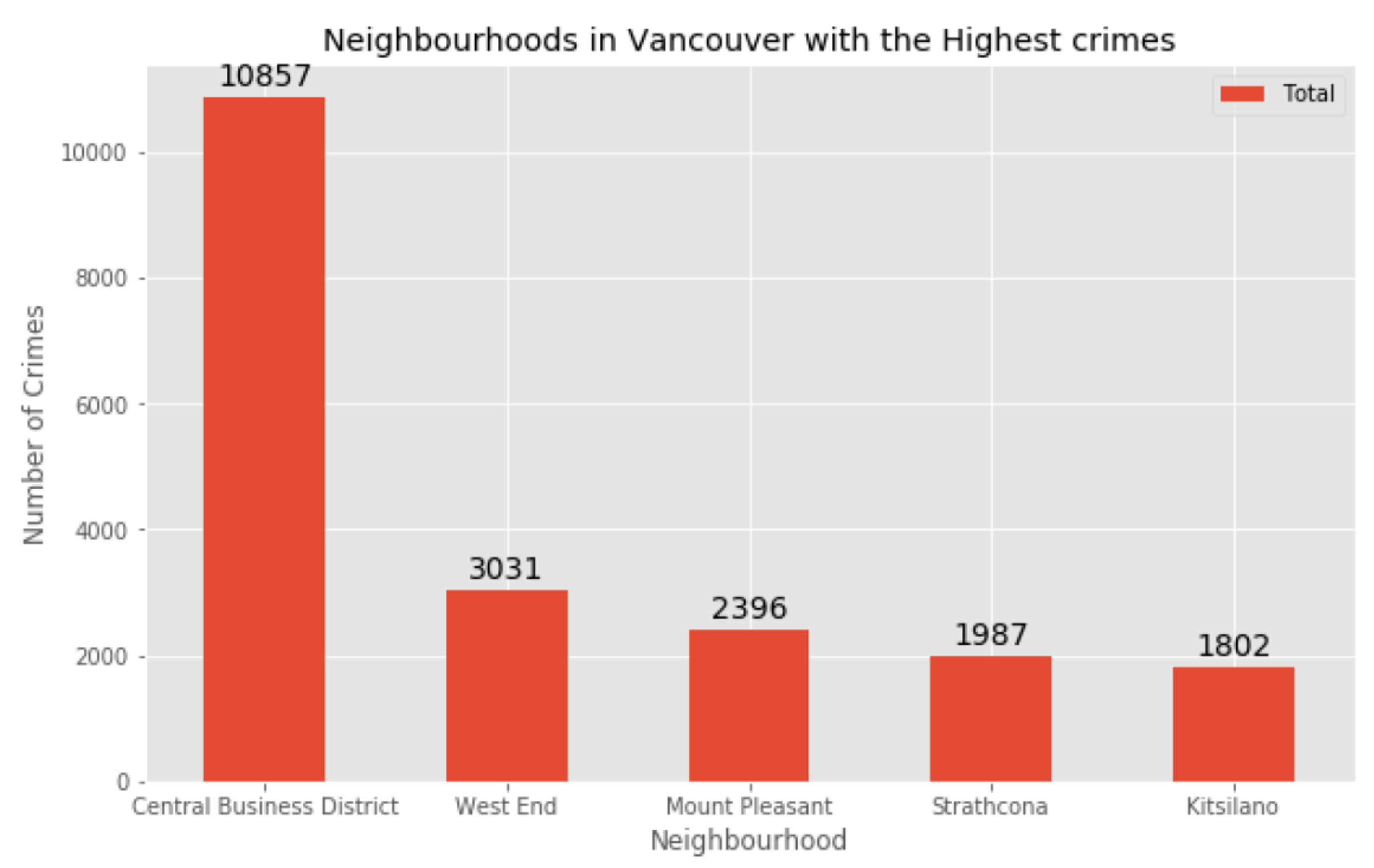
# Methodology

- Statistical crime rate summary:
  - The describe function in python is used extract statistics of the crime data. This function returns the mean, standard deviation, minimum, maximum, 1st quartile (25%), 2nd quartile (50%), and the 3rd quartile (75%) for each of the major categories of crimes.

| | 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) |
|---|---|---|---|---|---|---|---|---|---|
| **count** | 4.000000 | 4.000000 | 4.00000 | 4.000000 | 4.000000 | 4.000000 | 4.000000 | 4.000000 | 4.000000 |
| **mean** | 506.250000 | 599.250000 | 1430.25000 | 1236.750000 | 3736.500000 | 539.750000 | 286.500000 | 3.250000 | 368.500000 |
| **std** | 354.409721 | 488.189427 | 997.26572 | 1060.087221 | 2723.536977 | 353.955153 | 226.117226 | 3.304038 | 227.060198 |
| **min** | 49.000000 | 156.000000 | 187.00000 | 88.000000 | 483.000000 | 36.000000 | 71.000000 | 1.000000 | 111.000000 |
| **25%** | 314.500000 | 187.500000 | 843.25000 | 544.000000 | 2249.250000 | 450.000000 | 186.500000 | 1.000000 | 263.250000 |
| **50%** | 594.500000 | 599.000000 | 1627.00000 | 1185.000000 | 3796.000000 | 633.000000 | 235.000000 | 2.000000 | 351.500000 |
| **75%** | 786.250000 | 1010.750000 | 2214.00000 | 1877.750000 | 5283.250000 | 722.750000 | 335.000000 | 4.250000 | 456.750000 |
| **max** | 787.000000 | 1043.000000 | 2280.00000 | 2489.000000 | 6871.000000 | 857.000000 | 605.000000 | 8.000000 | 660.000000 |

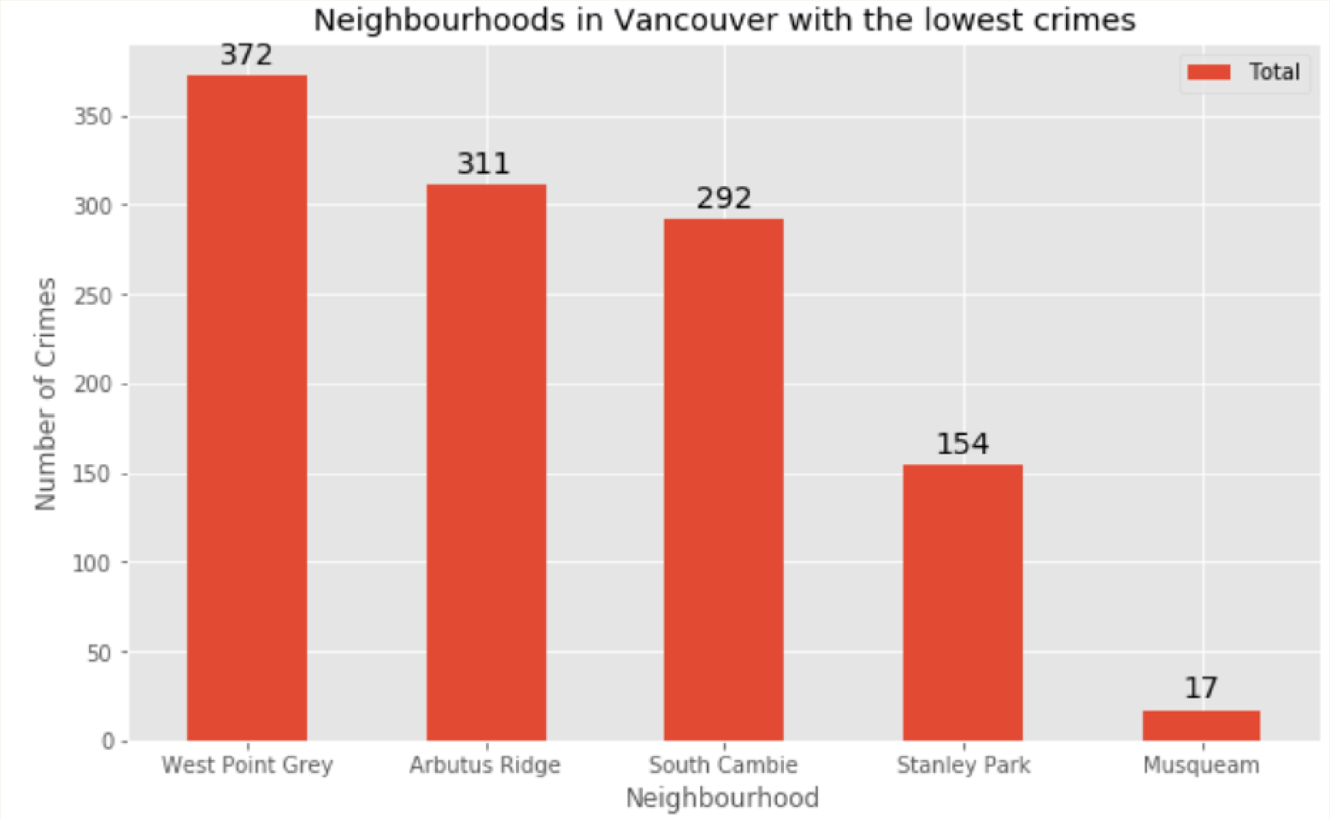
# Data Visualization

- Highest crime rate neighbourhood:



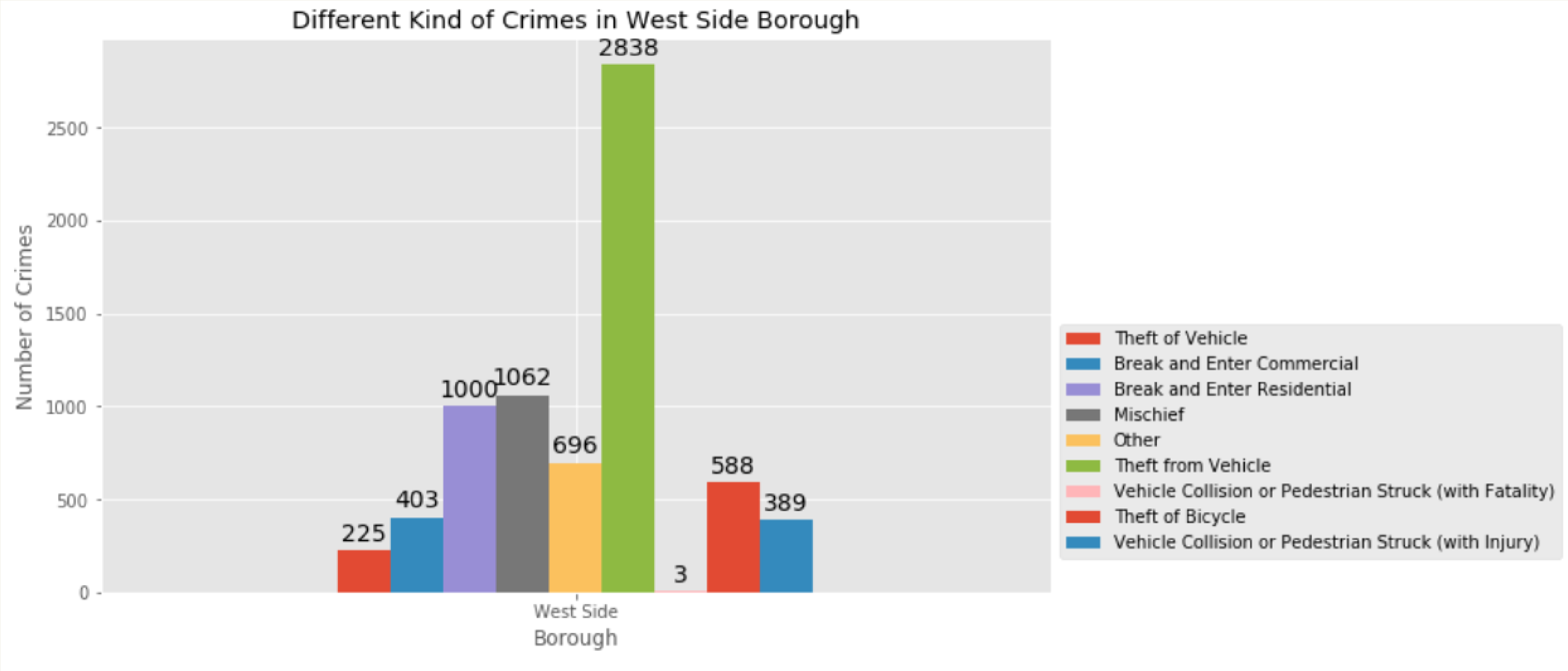
# Data Visualization

- Lowest crime rate neighbourhood:

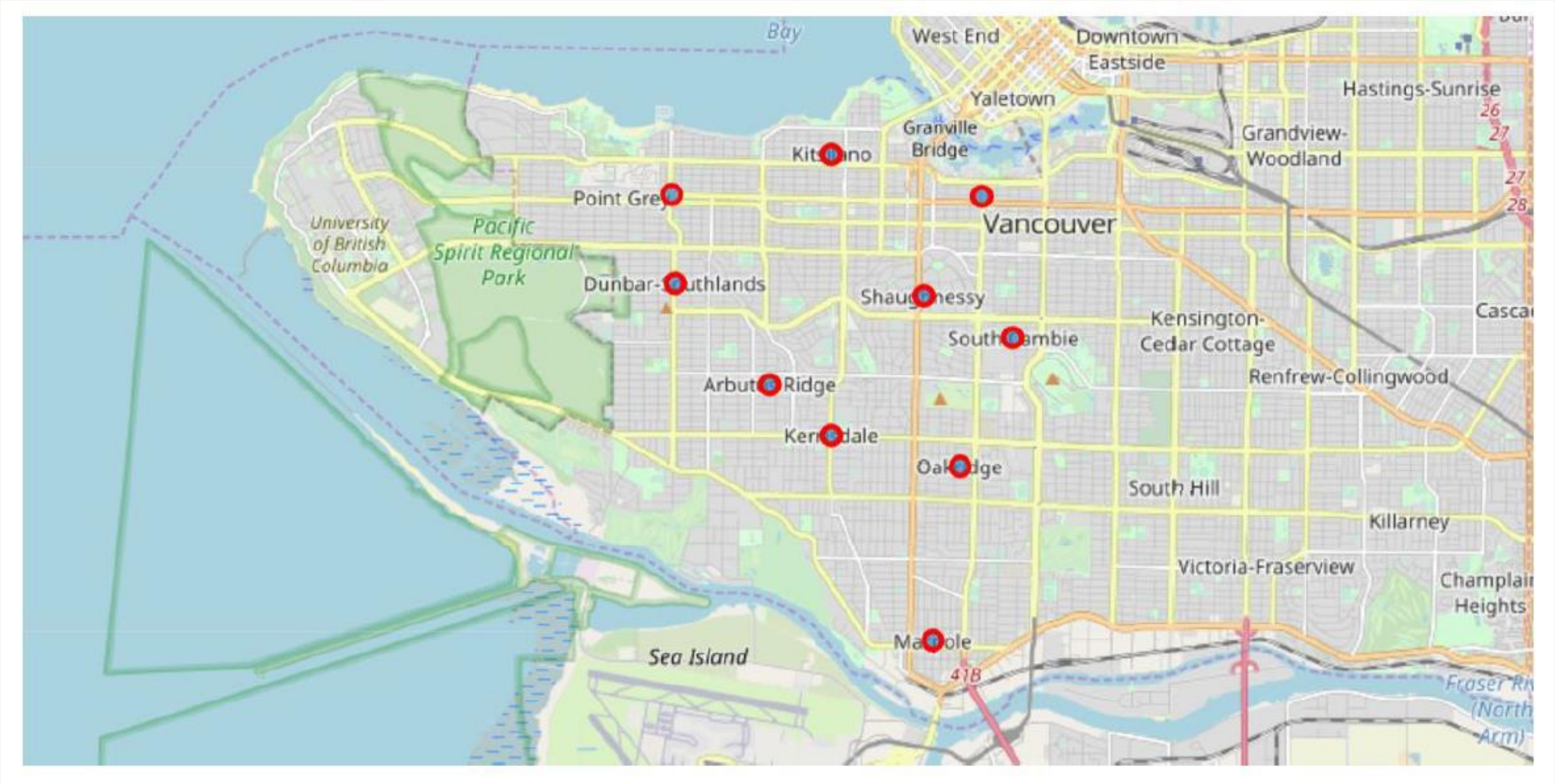


# Data Visualization

- Types of crime in West Side Borough:



# Neighborhoods in West Side Borough



# Modeling

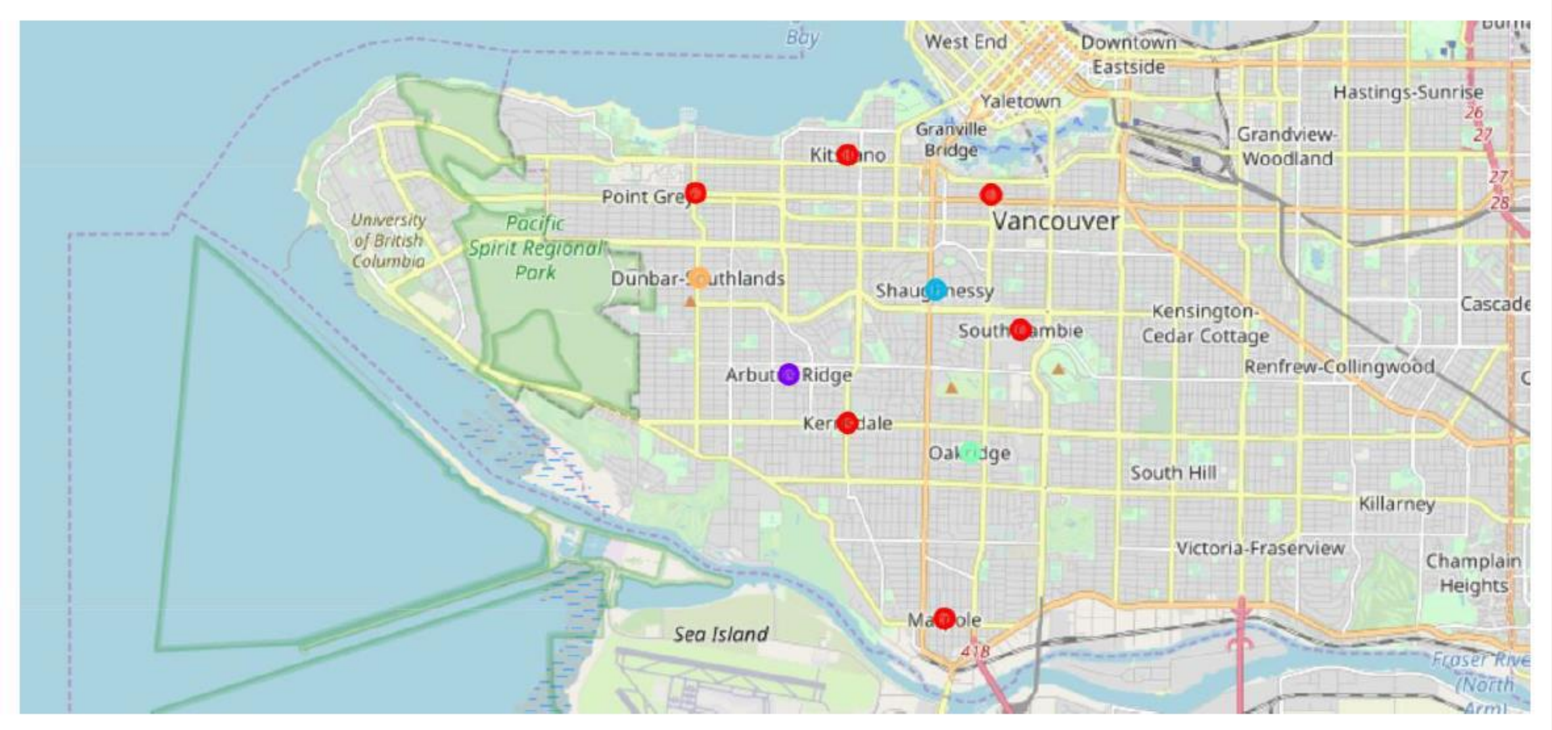
- By connecting to the FourSquare API and using the final dataset of neighborhood and borough along with latitude and longitude of neighborhoods in West Side Vancouver, all the venues within a 500-meter radius of each neighborhood can be found. This returns a response in json format containing all the venues in each neighborhood which was converted to a pandas data frame. This data frame contains all the venues along with their coordinates and category.

(229, 5)

| | Neighbourhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Category |
|---|---|---|---|---|---|
| 0 | Shaughnessy | 49.251863 | -123.138023 | Bus Stop 50209 (10) | Bus Stop |
| 1 | Shaughnessy | 49.251863 | -123.138023 | Angus Park | Park |
| 2 | Shaughnessy | 49.251863 | -123.138023 | Crepe & Cafe | French Restaurant |
| 3 | Fairview | 49.264113 | -123.126835 | Gyu-Kaku Japanese BBQ | BBQ Joint |
| 4 | Fairview | 49.264113 | -123.126835 | CRESCENT nail and spa | Nail Salon |

# Result

- Each cluster is color coded for the ease of presentation, it can be noted that majority of the neighborhood falls in the red cluster which belongs to the first cluster. Remaining neighborhood are part of remaining four clusters and has been represented with different colors.



# Result

- The data of Cluster contains the following Neighborhoods:

| | 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 | Asian Restaurant | Park | Chinese Restaurant | Sandwich Place | Indian Restaurant | Korean Restaurant | Malay Restaurant | Nail Salon | Fast Food Restaurant |
| 3 | West Side | Pizza Place | Chinese Restaurant | Sushi Restaurant | Japanese Restaurant | Lingerie Store | Noodle House | Dim Sum Restaurant | Falafel Restaurant | Plaza | Café |
| 4 | West Side | Bakery | Coffee Shop | Sushi Restaurant | American Restaurant | Thai Restaurant | Japanese Restaurant | Tea Room | Food Truck | French Restaurant | Ice Cream Shop |
| 5 | West Side | Coffee Shop | Chinese Restaurant | Pharmacy | Tea Room | Sushi Restaurant | Sandwich Place | Fast Food Restaurant | Noodle House | Dessert Shop | Pet Store |
| 6 | West Side | Japanese Restaurant | Coffee Shop | Café | Vegetarian / Vegan Restaurant | Bakery | Pub | Sushi Restaurant | Dessert Shop | Pizza Place | Pharmacy |
| 8 | West Side | Coffee Shop | Bus Stop | Malay Restaurant | Juice Bar | Cantonese Restaurant | Grocery Store | Sushi Restaurant | Park | Café | Bank |

# Discussion

- The objective of this project 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 liquor store. This has been achieved by first analyzing 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 liquor shops were not among venues in a close proximity to each other. This was achieved 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

- This project has explored the crime data to understand different types of crimes rate in all neighborhoods of Vancouver. Then categorized the data into different boroughs, this helped identify the safest borough. Once the borough was confirmed, the number of neighborhoods for consideration also were significantly reduced. Also, there was further shortlist the neighborhoods based on the common venues, and to choose a neighborhood which best resolves the business problem.