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

#### **Applied Data Science Capstone by IBM/Coursera**

#### **Table of contents**

* **Introduction: Business Problem**
* **Data**
* **Methodology**
* **Analysis**
* **Results and Discussion**
* **Conclusion**

#### **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**

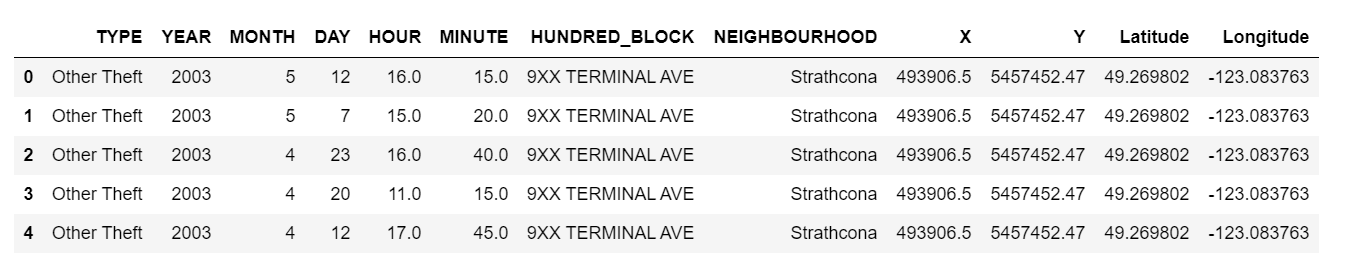
Based on definition of our problem, factors that will influence our decision are:

* finding the safest borough based on crime statistics
* finding the most common venues
* choosing the right neighborhoods within the borough

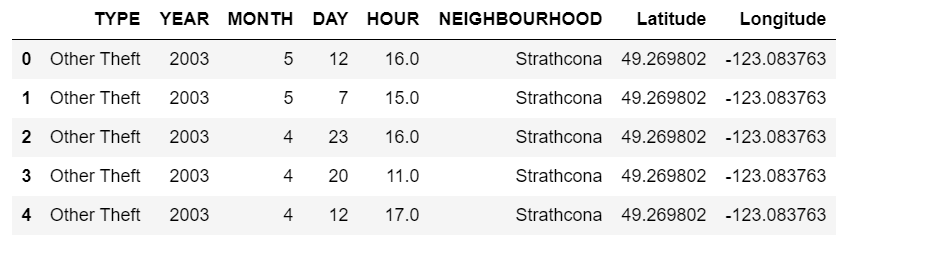
We will be using the geographical coordinates of Vancouver to plot neighborhoods in a borough that is safe and in the city's vicinity, and finally cluster our neighborhoods and present our findings.

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.

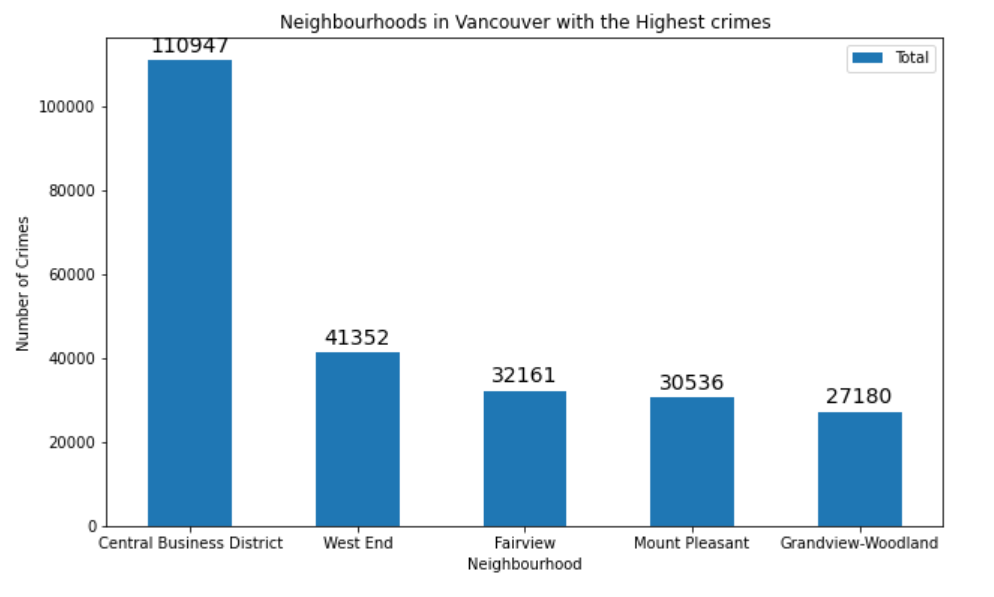


#### **Drop not user column minute, Hundred\_Block, X, Y**

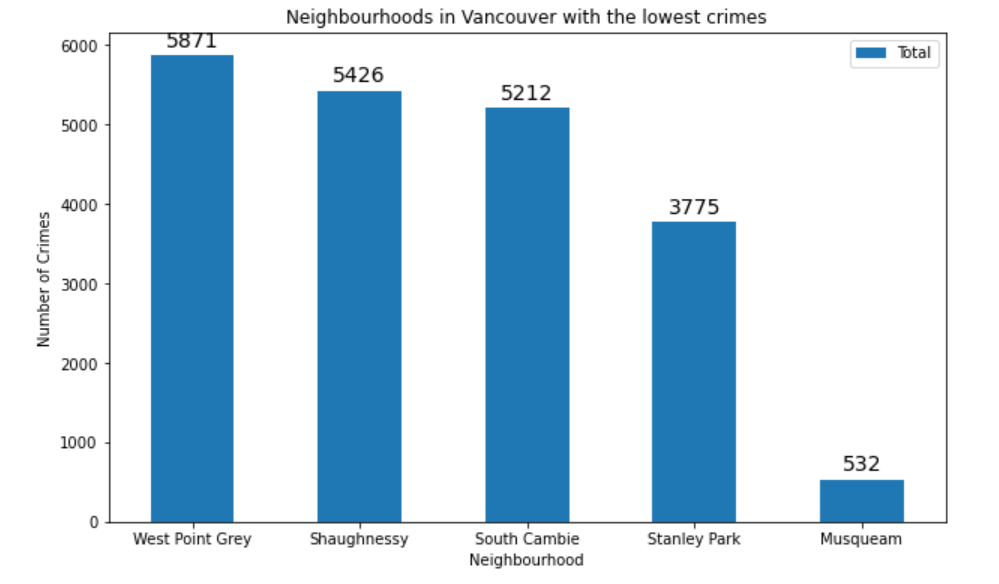


* Part 2: Gathering additional information of the list of officially categorized boroughs in Vancouver from Wikipedia: Borough information will be used to map the existing data where each neighborhoods can be assigned with the right borough.

#### **Top Five Neighborhoods with highest crimes**

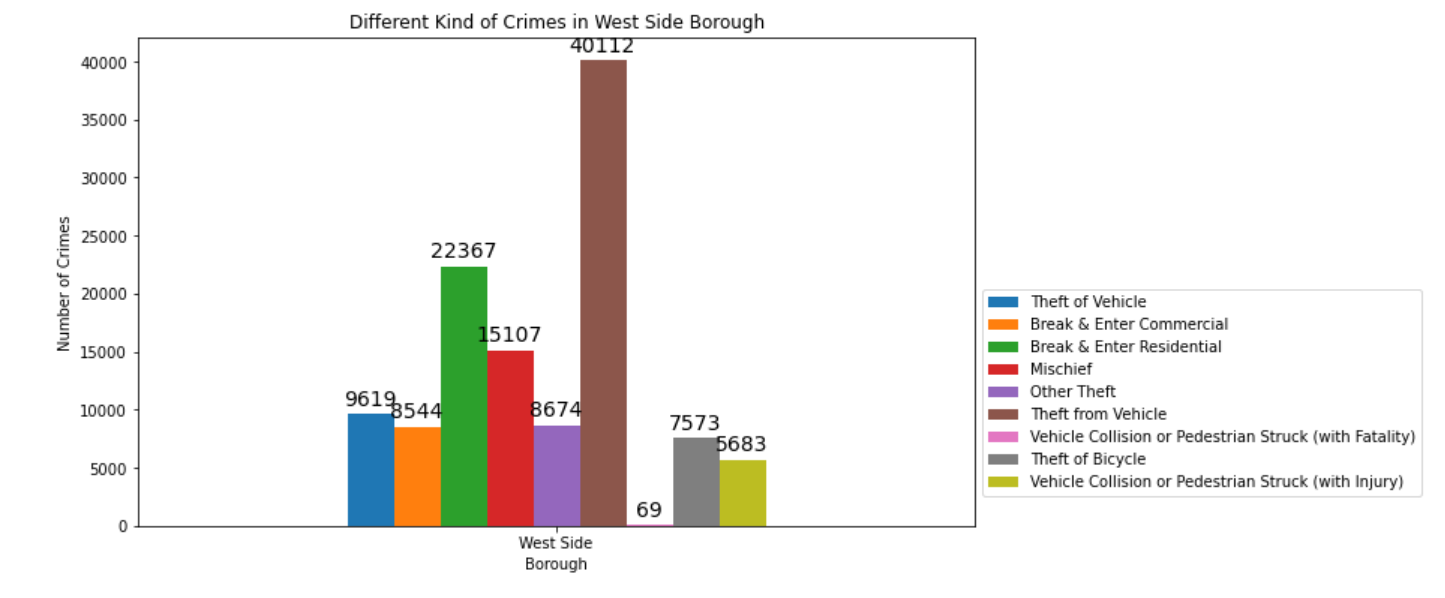


#### **Top Five Neighborhoods with the lowest crimes**

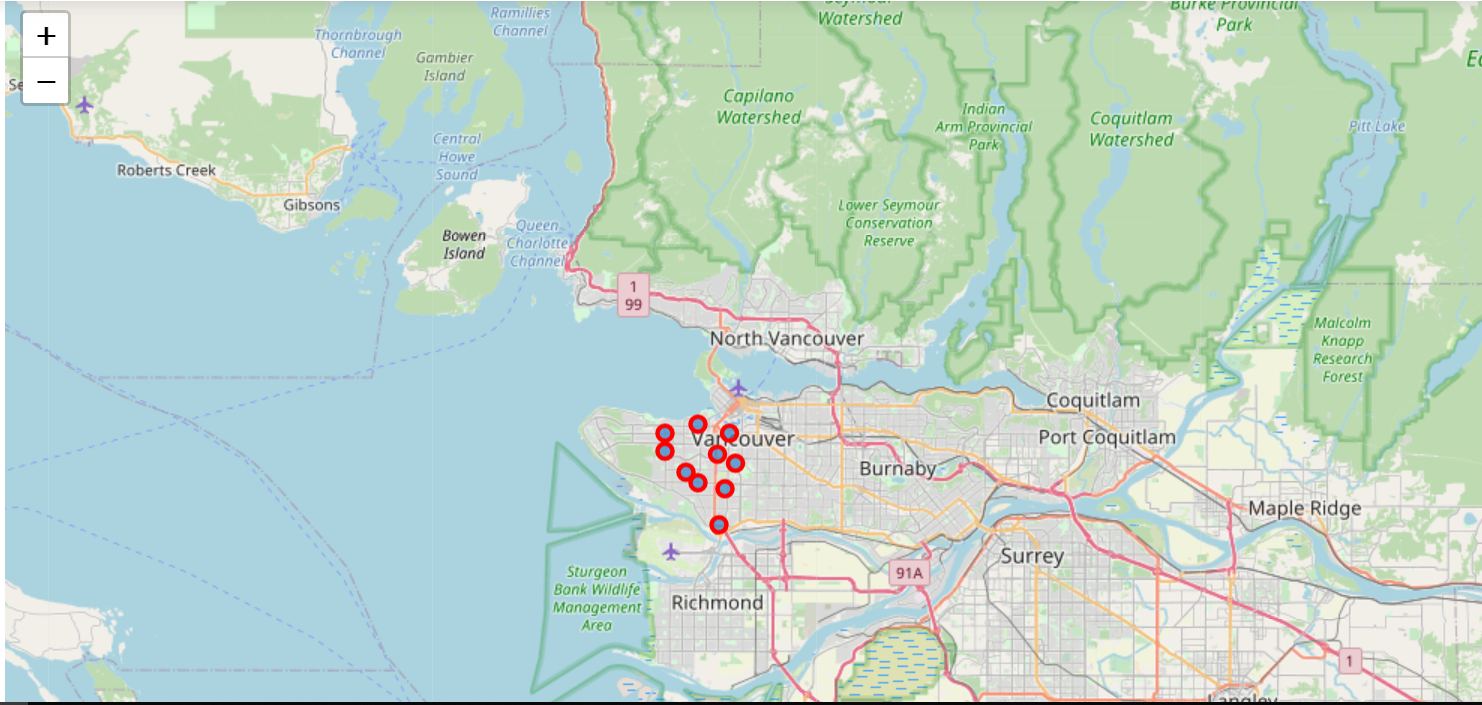


### Based on visualized data analysis the South Vancouver has the lowest crimes and the East Side area has the highest crimes

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**.

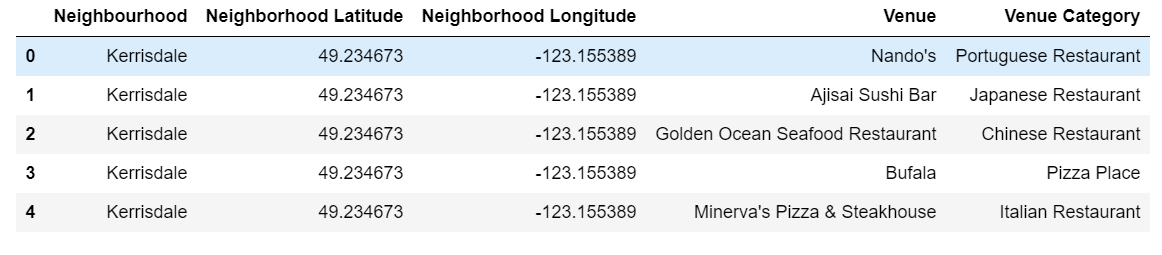


* Part 3: Creating a new consolidated dataset of the Neighborhoods, along with their boroughs, crime data and the respective Neighborhood’s co-ordinates.: This data will be fetched using Open Cage Geocoder to find the safest borough and explore the neighborhoods by plotting it on maps using Folium and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



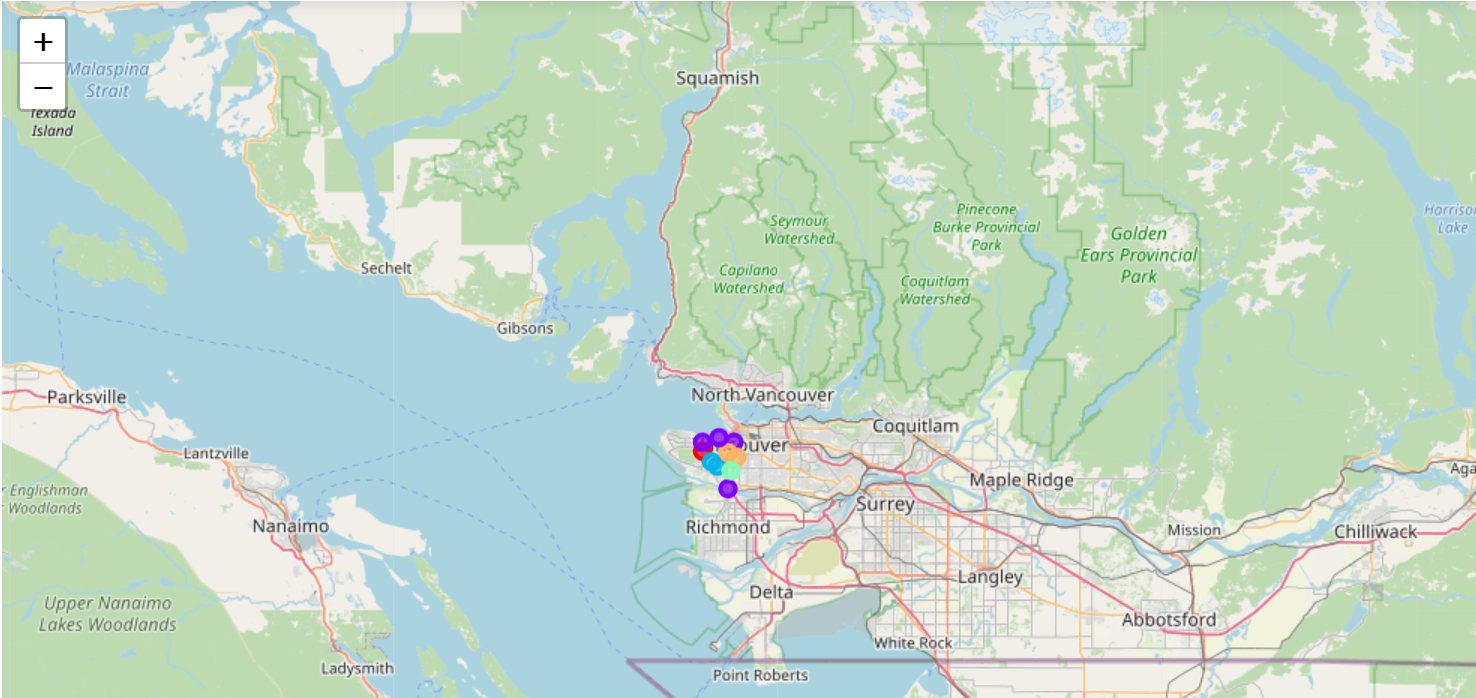
* Part 4: Creating a new consolidated dataset of the Neighborhoods, boroughs, and the most common venues and the respective Neighborhood along with co-ordinates.: This data will be fetched using Four Square API to explore the neighborhoods venues and to apply machine learning algorithm to cluster the neighborhoods and present the findings by plotting it on maps using Folium.

#### **Data frame of venues for each neighborhood in West Side**

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



#### **Finally, let's visualize the resulting clusters**

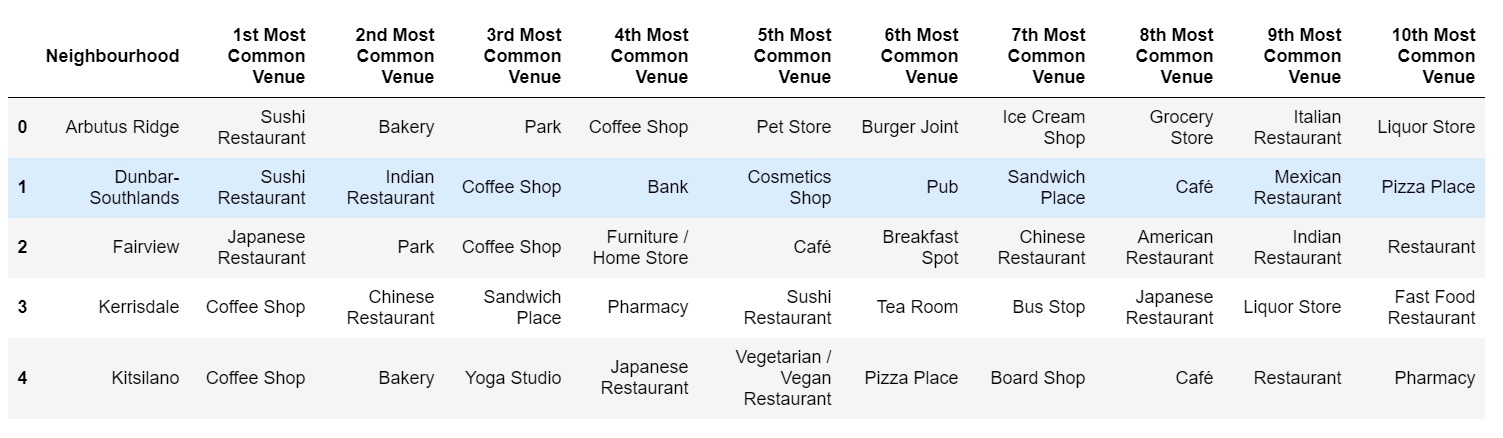


#### **Methodology**

Categorized the methodology section into two parts:

* **Exploratory Data Analysis:**

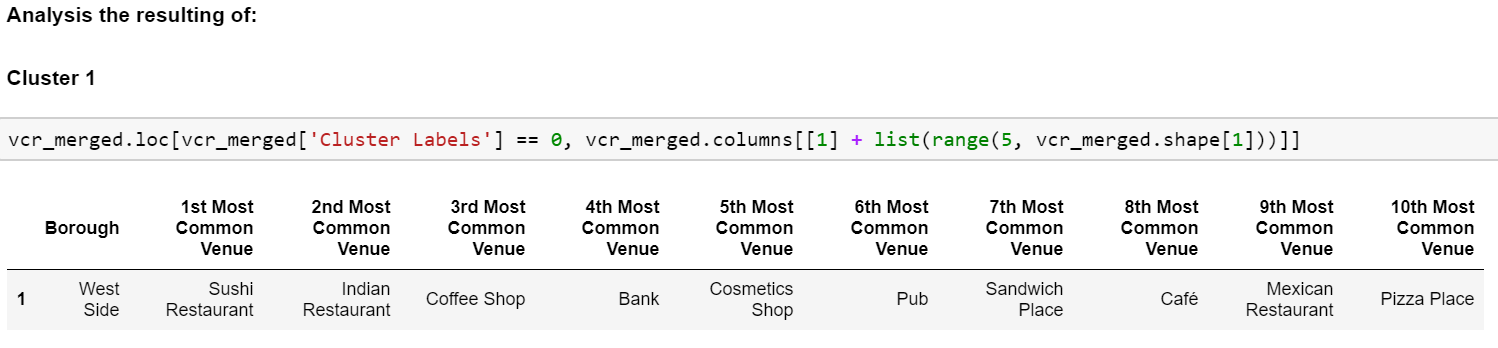
Visualize the crime reports in different Vancouver boroughs to identity the safest borough and normalize the neighborhoods of that borough. We will Use the resulting data and find 10 most common venues in each neighborhood.



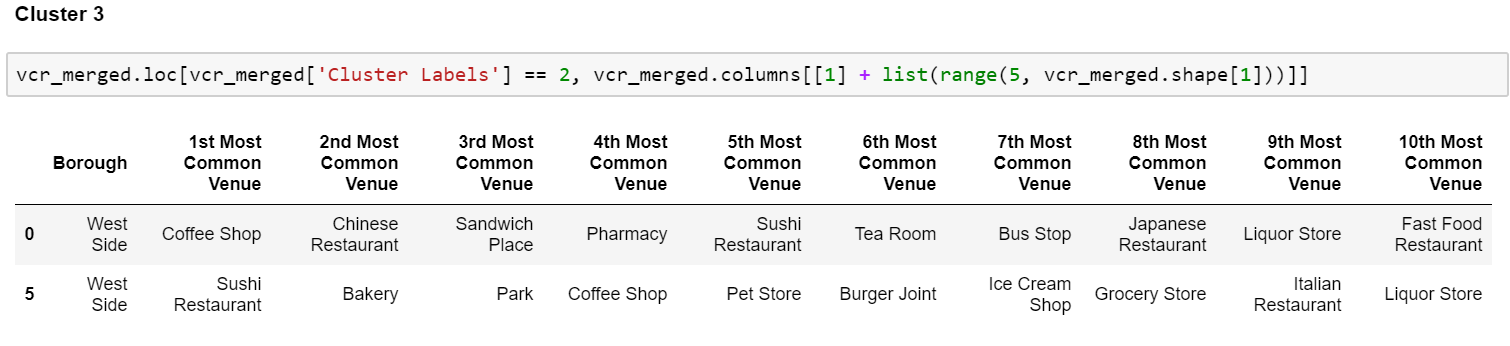
* **Modelling:**

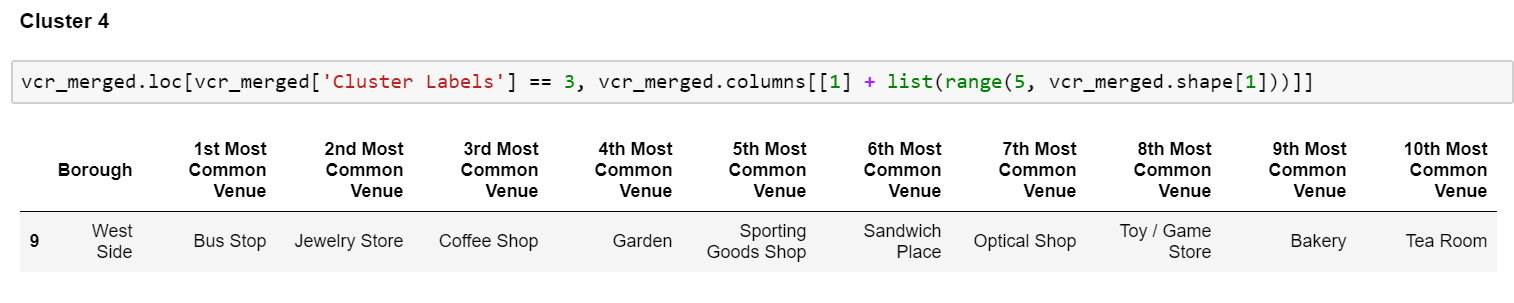
To help stakeholders choose the right neighborhood within a borough we will be clustering similar neighborhoods using K - means clustering which is a form of unsupervised machine learning algorithm that clusters data based on predefined cluster size. We will use K-Means clustering to address this problem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













#### **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.