**The Battle of Neighborhoods for Grocery Store in Vancouver, Canada**

**Sonal Tarve**

**Jan, 2021**

1. **Introduction /Business Problem:**

***Problem Background:***

Vancouver is highly populated city with high real estate prices. Opening any type of commercial establishment need to be analyzed in detail and carefully to get the maximum Return on Investment (ROI).

***Problem Description:***

The purpose of this project is to find a safe and secure location for stakeholders for opening of commercial establishments in **Grocery Store** in **Vancouver City**, Canada.

1. The first task would be to **choose the safest borough** by analyzing crime data for opening a grocery store.
2. The second task is **short listing a neighborhood**, where grocery stores are not amongst the most common venues, and yet **as close to many other types of venue** to attract shoppers.

I 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.

***Target Audience:***

Stakeholders and investors from a leading grocery chain store in Canada.

1. **Data**

Based on definition of the business problem defined above, the factors that will influence our decision are:

* Finding the safest borough based on crime statistics
* Choosing the right neighborhood within the safest borough
* Finding the most common venues where grocery stores are not existing or less.

I 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 the neighborhoods and present my findings.

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

**Section I - Vancouver Crimes of 2018 - Real world data set from Kaggle**

A dataset consisting of the crime statistics of each neighborhood in Vancouver along with type of crime, recorded year, month and hour. Vancouver Crime Report Data set URL: <https://www.kaggle.com/agilesifaka/vancouver-crime-report/version/2>. The data file is crime\_records.csv

**Section II - Using Wikipedia page for getting additional info about the neighborhood.**

Borough information will be used to map the existing data where each neighborhood can be assigned with the right borough. Link - <https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Vancouver>

**Section III - Using OpenCage Geocoder to find the safest borough and explore its neighborhood.**

This data will be fetched using OpenCage Geocoder to find the safest borough and explore the neighborhood along with their crime data and the respective neighborhood's co-ordinates by plotting it on maps using Folium and perform exploratory data analysis.

**Section IV - Using Four Square API to explore the neighborhood venues and to apply machine learning algorithm to cluster the neighborhoods**

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.

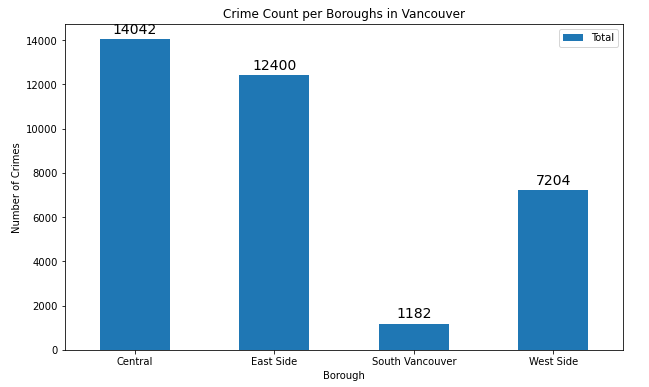
1. **Methodology**

***Analytic Approach:***

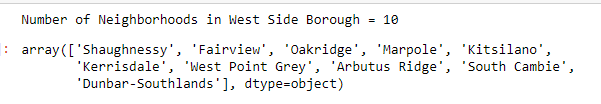
* Visualize the crime reports in different Vancouver boroughs to identity the safest borough and normalize the neighborhoods of that borough.
* First look for total crime count in various boroughs in Vancouver and from resulting data find the lower crimes borough and their number of neighborhoods.
* Then select the most suitable borough and look for 10 most common venues in each neighborhood.
* One hot Encoding to analyze each neighborhood
* Then clustering the neighborhoods to create map
* The last is to analyze the clusters to look for the suitable cluster neighborhood.

***Exploratory Data Analysis:***

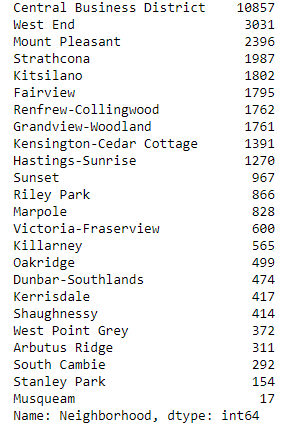
* Sorting the data by crime count per borough



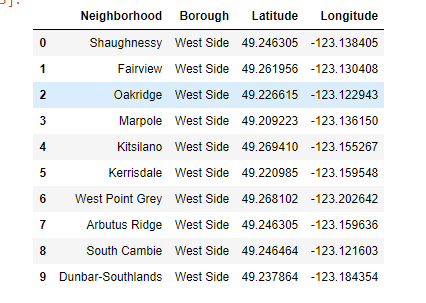
Since South Vancouver has very little number of neighborhoods and opening a commercial establishment would not be viable, I choose the next borough with lowest crime, which is **West Side**.



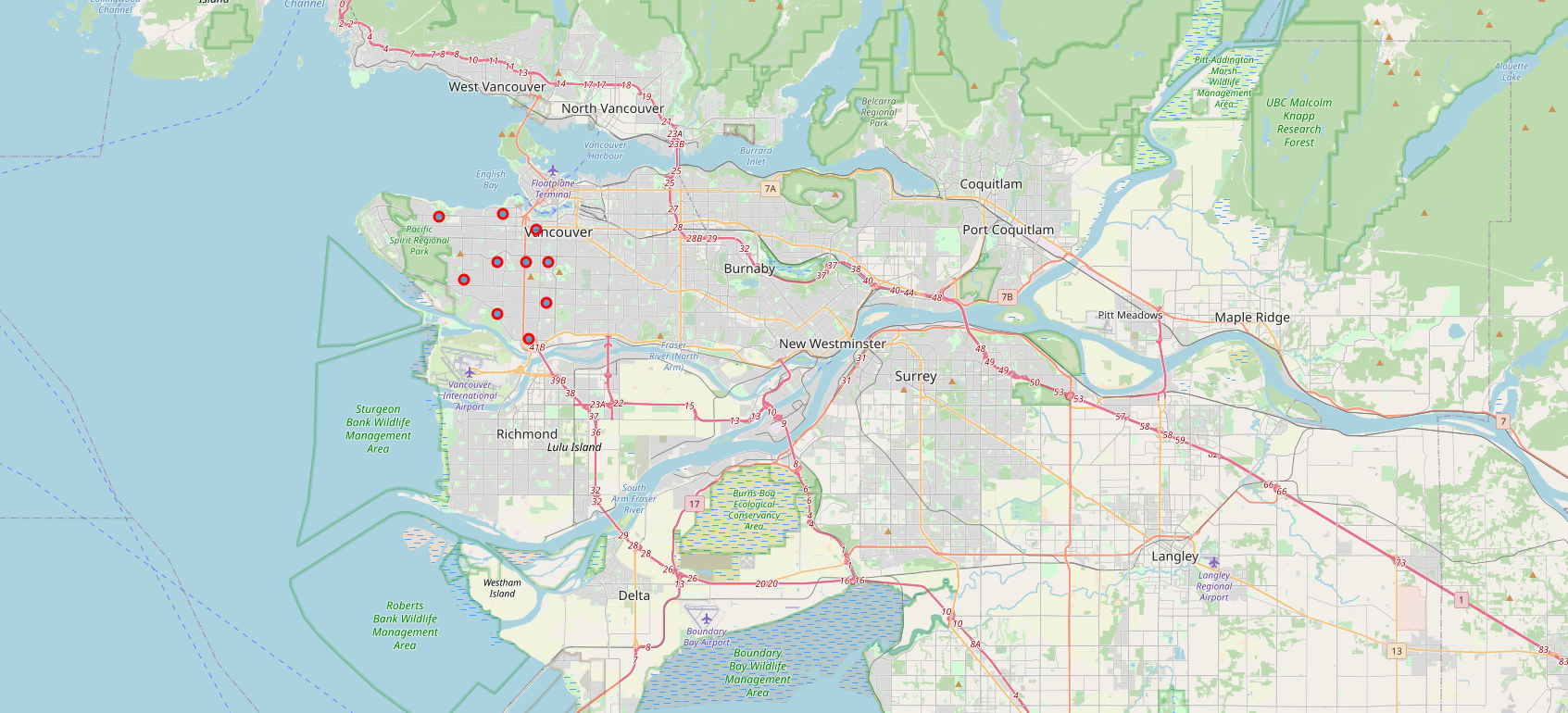
* Total crime count in various neighborhoods in Vancouver



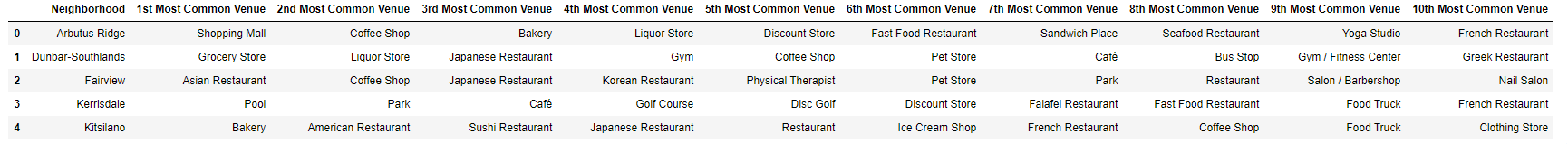
* Location of each neighborhood.



* Neighborhood Map



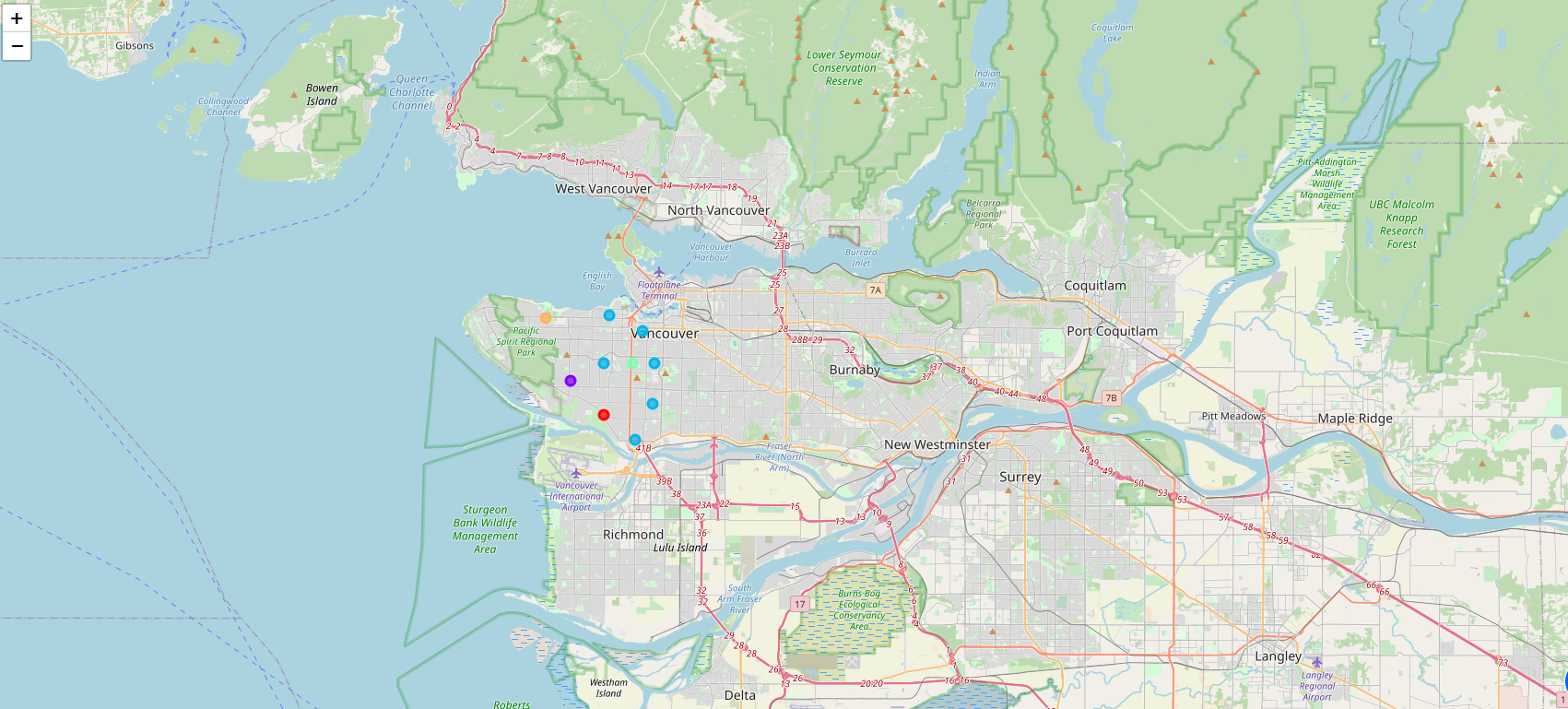
### Top 10 most common venues across neighborhoods



* Cluster Neighborhoods with label



* Cluster Map



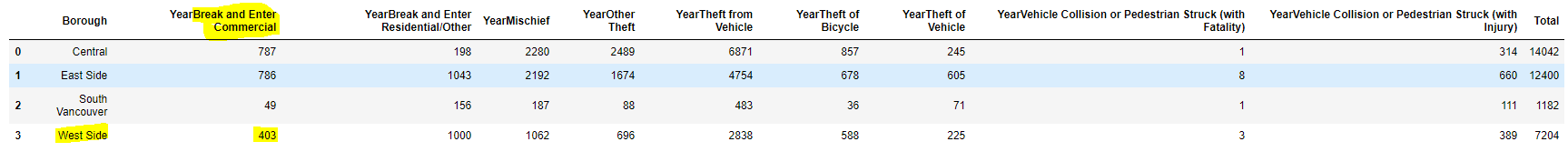
***Modelling/ inferential statistical testing/ machine learnings:***

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

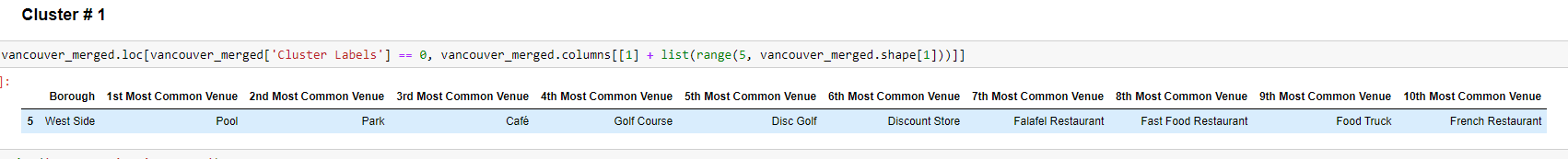
1. **Results**

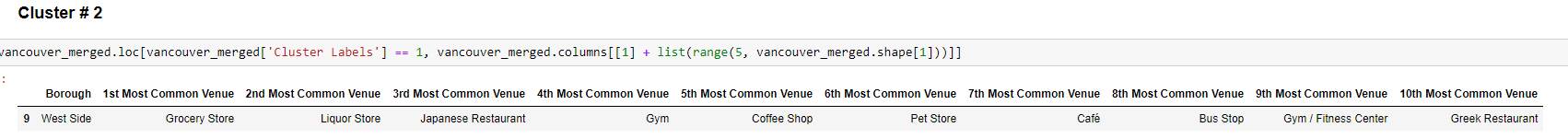
South Vancouver has very little number of neighborhoods and opening a commercial establishment would not be viable, I choose the next borough with lowest crime, which is **West Side**.

West side was also chosen because crime type Break and enter Commercial is also low amongst other crimes types which makes West Side ideal destination for opening of commercial establishments

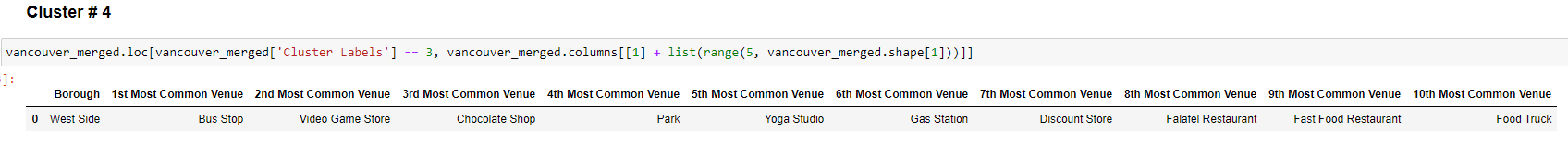


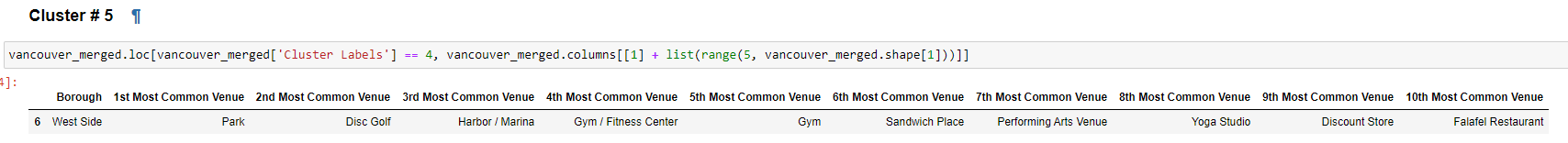
### **Analyzing the resulting Clusters**











1. **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 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 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.

1. **Conclusion**

There are only 2 clusters with Grocery Store, Cluster 2 & 3.

The first recommendation is Cluster 3, which has only one grocery store and it is situated in the 10th most common venue ranking. At the same time this cluster has lots of venues around to attract all types of shoppers.

The second recommendation is Cluster 1, 4 & 5. They do not have any grocery store and good other venues are around.