



# APPLIED DATA SCIENCE CAPSTONE

Madhumita Das

# BUSINESS PROBLEM

- The **objective** of this Capstone project is to propose to the stakeholders a safe place to start a mini-grocery store business venture in the city of Vancouver. Opening a mini-grocery store can be a lucrative business venture if it can be opened in a secure place with less crime and competition.
- The problem will be approached in two phases:
- First, the safe borough has been selected for opening the store by analyzing the crime data of Vancouver city neighborhoods.
- Second, the data science tools learned in the course, has been used to explore the neighborhood of the safest borough and the ten most common venues in each neighborhood. Based on the results of most common venues, proposal has been placed to the stakeholders as to in which neighborhood the store can be opened.

# TARGET AUDIENCE

- This project will be of interest to the stakeholders who would like to invest in a mini-grocery store business in the Vancouver City and would like to find out a safe and secure place with less competition to start with.

# DATA SOURCE

- The data sources that have been used are:
- I. Vancouver City Crime data from Kaggle: Since the data set is huge, for the purpose of this project, I have considered only the 2019 crime data from
- II. Further data has been scraped from Wikipedia to gather information on Boroughs in Vancouver.
- III. A consolidated data set will be created of neighborhoods, boroughs, coordinates and the crime data gathered before.
- IV. Foursquare API has been used to fetch that data and to find the most common venues in each neighborhood.
- V. Machine learning algorithm to be used to cluster the neighborhoods and finally select the best neighborhood to open the store.



# METHODOLOGY(Contd.)

---

- The Crime data table and the boroughs table have been merged and used to calculate the total number of crimes in each borough:

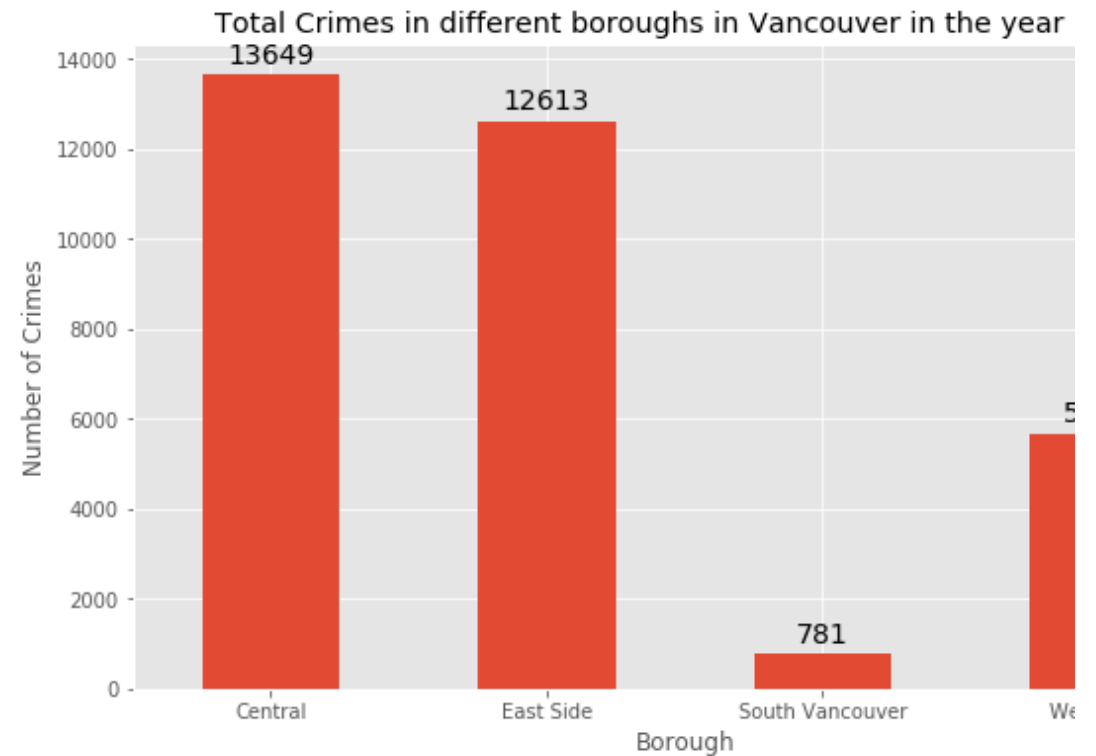
	Type	Year	Month	Day	Hour	Neighbourhood	Borough
0	Break and Enter Commercial	2019	3	7	2	Fairview	West Side
1	Break and Enter Commercial	2019	4	21	16	Fairview	West Side
2	Break and Enter Commercial	2019	10	26	0	Fairview	West Side
3	Break and Enter Commercial	2019	3	27	8	Fairview	West Side
4	Break and Enter Commercial	2019	7	13	1	Fairview	West Side

```
Central      13649
East Side    12613
West Side     5674
South Vancouver  781
Name: Borough, dtype: int64
```

# METHODOLOGY(Contd.)

The different crime types in each borough has been obtained as shown below:

A Bar graph has been created for the visual representation of the total crimes in each borough.

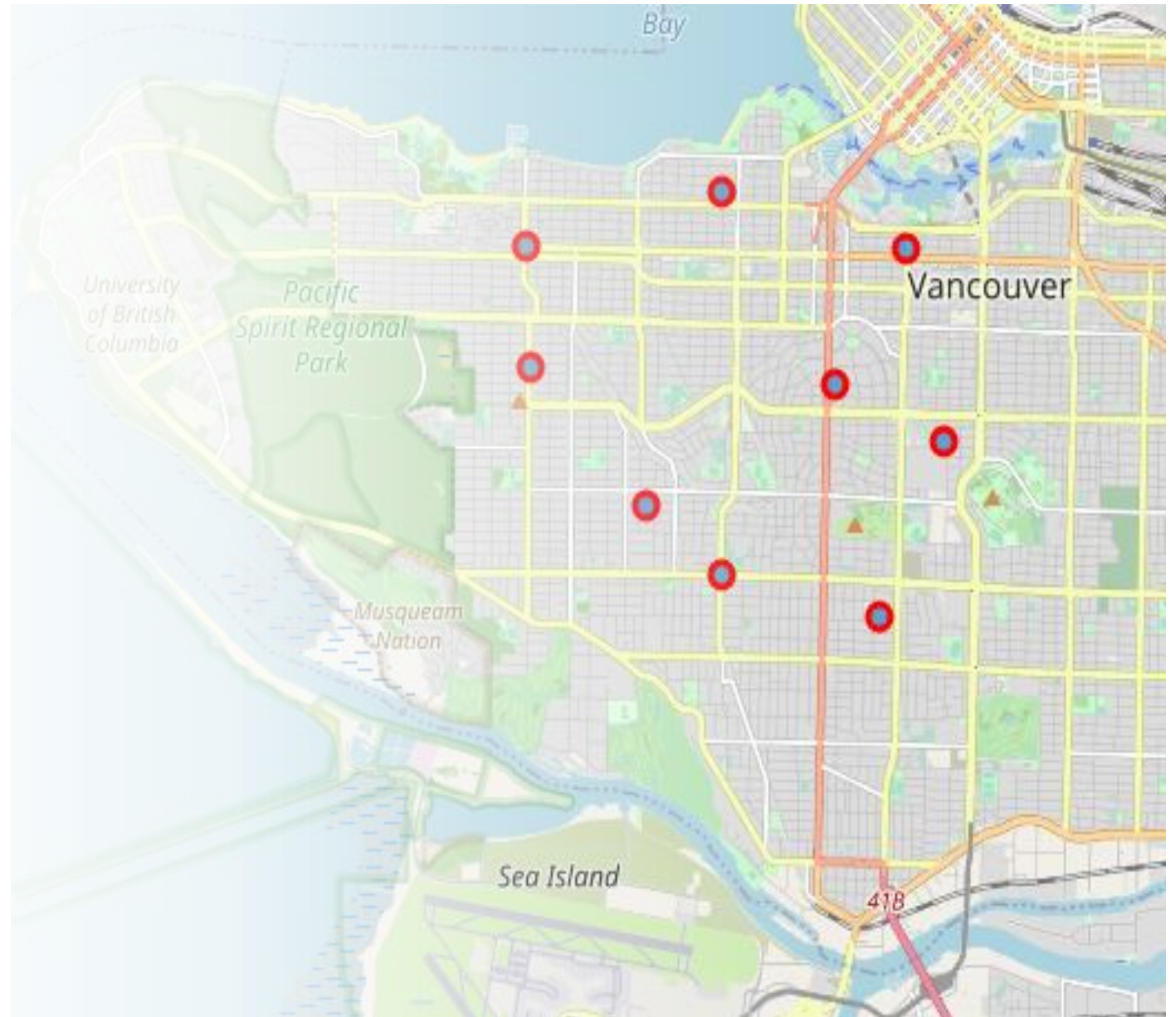


## FINDINGS

**As the results are analyzed, South Vancouver has the lowest crime rate out of the 4 boroughs in Vancouver. This may be because the number of neighborhoods is minimum there. Hence, we will consider the borough West Side as it has the next minimum crime rate after South Vancouver. Also, it has many neighborhoods and the 'Break and Enter Commercial' type crime is less comparatively.**



- 
- Folium library has been used to plot the neighborhoods of West Side in Vancouver.

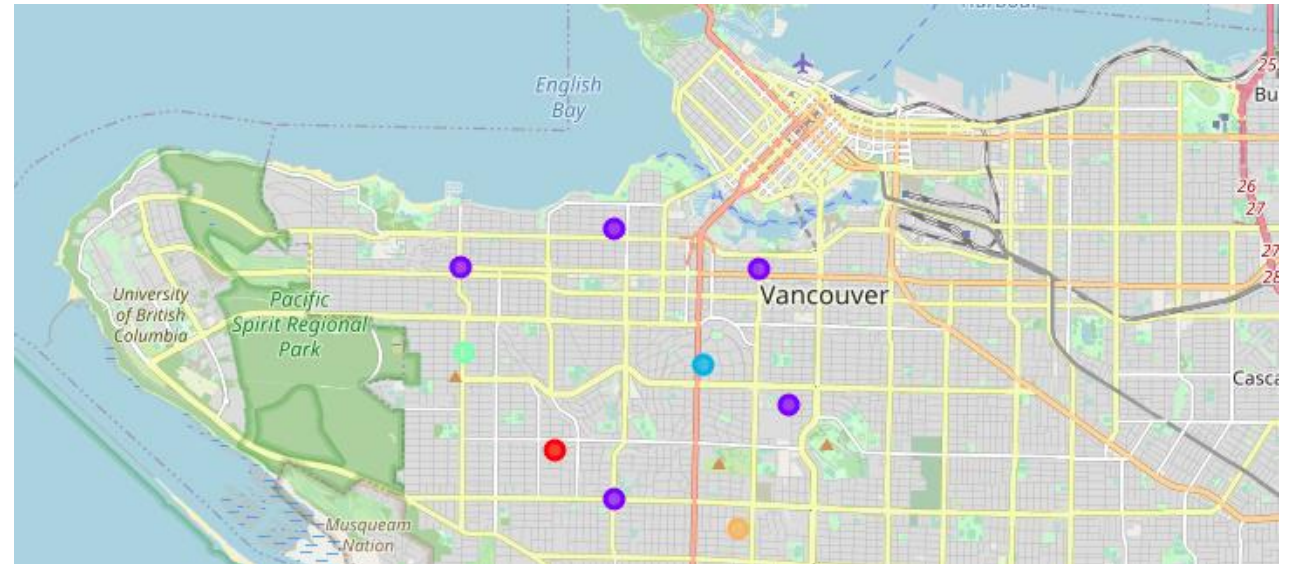


# ANALYSIS

- Foursquare API has been used to get the venues in each neighborhood in Vancouver and total number of venues in each neighborhood has been calculated.

Venue	
Neighbourhood	
Arbutus Ridge	5
Dunbar-Southlands	6
Fairview	26
Kerrisdale	40
Kitsilano	48
Oakridge	8
Shaughnessy	3
South Cambie	15
West Point Grey	40

- K- Means has been used to cluster the neighborhoods.



# RESULTS

```
vancouver_merged['Cluster Labels'] == 2, vancouver_merged.columns[[1] + list(range(5, vancouver_merge
```

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
Park	Bus Stop	French Restaurant	Yoga Studio	Dessert Shop	Diner	Falafel Restaurant	Fast Food Restaurant

```
vancouver_merged['Cluster Labels'] == 3, vancouver_merged.columns[[1] + list(range(5, vancouver_merge
```

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
Sushi Restaurant	Italian Restaurant	Indian Restaurant	Ice Cream Shop	Coffee Shop	Yoga Studio	French Restaurant	Deli / Bodega

```
vancouver_merged['Cluster Labels'] == 4, vancouver_merged.columns[[1] + list(range(5, vancouver_merge
```

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
Israeli Restaurant	Fast Food Restaurant	Café	Pharmacy	Sandwich Place	Sushi Restaurant	Convenience Store	Vietnamese Restaurant

- Each cluster has then been analyzed to find out the results:

# CONCLUSION

- **Based on the above analysis, the grocery stores are much less common in clusters(with labels 2, 3, 4) and therefore ideal for setting up one.**



THANK YOU