Capstone Project

Finding Nearest Neighborhood in Vancouver

- Anow Fanny Durairaj

Introduction - Business Problem

The business problem aims to target the stakeholders who wants to starts a business in Vancouver City in Canada. This problem will analysis the crime data for opening a store and shortlists the neighborhood where the kind of store is not as close to the city.

Task includes:

- **Analyze the crime data by choosing the safest borough**
- **Find the list of neighborhood**
- **Choose the close neighborhood to the city**

We have used the Data Science tools to analyze and focus on the data and explore the neighborhoods and the 10 most common venues in each neighborhood. This helps to find the best neighborhood that can be selected where the store is not the common venue.

Data

Based on definition of the problem, the following factors will influence our decision:

- **Find the safest borough based on crime statistics**
- **Find the most common venues**
- **Choose the right neighborhood 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.

The following data sources are extracted/generated:

1. A dataset consisting of the crime statistics of each Neighborhood in Vancouver along with type of crime, recorded year, month and hour.
2. Borough information is used to map the existing data where each neighborhood can be assigned with the right borough.
3. This data is fetched using OpenCage Geocoder to find the safest borough, explore the neighborhood by plotting it on maps using Folium, and perform exploratory data analysis.
4. This data is 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.

Vancouver Crime Report

The properties of the Crime report includes:

- TYPE - Crime type
- YEAR - Recorded year
- MONTH - Recorded month
- DAY - Recorded day
- HOUR - Recorded hour
- MINUTE - Recorded minute
- HUNDRED_BLOCK - Recorded block
- NEIGHBOURHOOD - Recorded neighborhood
- X - GPS longitude
- Y - GPS latitude

Data set URL: https://www.kaggle.com/agilesifaka/vancouver-crime-report/version/2

Reading from the Dataset

We have processed only possible data, as it was a huge dataset. We have restricted the data set to the recent crime report of the 2018.

Gathering additional information about the Neighborhood from Wikipedia.

As part of data set Borough that the neighborhood was part of was not categorized, so we will create a dictionary of Neighborhood and based on data in the following

[Wikipedia page](https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Vancouver).

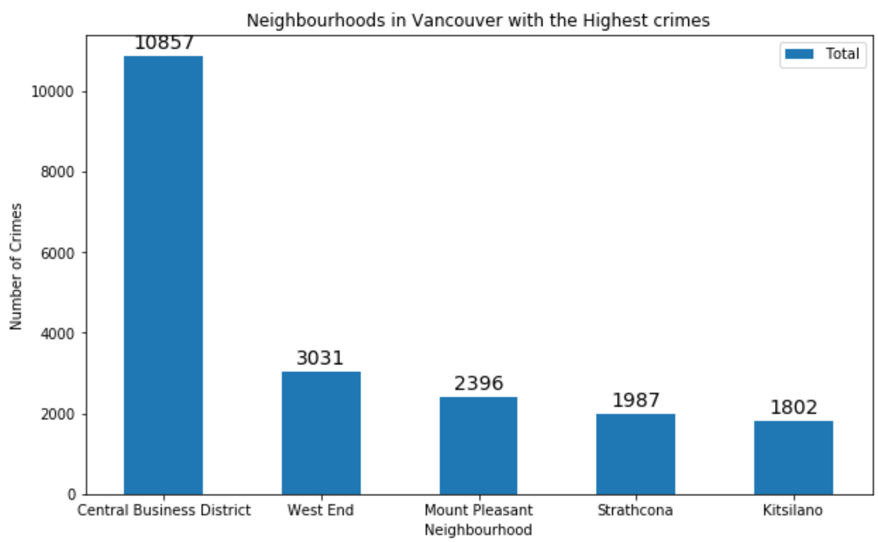
Type	Break and Enter Commercial	Break and Enter Residential/Other	Mischief	Other Theft	Theft from Vehicle	Theft of Bicycle	Theft of Vehicle	Vehicle Collision or Pedestrian Struck (with Fatality)	Vehicle Collision or Pedestrian Struck (with Injury)	All
Neighbourhood										
Arbutus Ridge	12	78	49	18	111	12	12	1	18	311
Central Business District	551	124	1812	2034	5301	640	165	0	230	10857
Dunbar-Southlands	8	106	81	31	199	16	9	1	23	474
Fairview	138	73	233	297	692	245	55	0	62	1795
Grandview-Woodland	148	162	304	215	634	110	123	0	65	1761
Hastings-Sunrise	48	117	195	107	607	52	74	0	70	1270
Kensington-Cedar Cottage	62	145	255	148	541	69	71	3	97	1391
Kerrisdale	24	97	49	9	172	13	11	0	42	417
Killarney	34	72	90	31	240	19	33	0	46	565
Kitsilano	106	165	320	154	755	189	51	1	61	1802
Marpole	44	125	134	75	290	34	39	0	87	828
Mount Pleasant	205	124	353	493	822	232	67	0	100	2396
Musqueam	0	4	3	0	4	2	2	0	2	17
Oakridge	19	123	64	63	164	18	18	0	30	499
Renfrew-Collingwood	91	156	243	472	569	37	92	0	102	1762
Riley Park	35	122	140	53	378	52	39	2	45	866
Shaughnessy	12	120	41	0	187	10	11	0	33	414
South Cambie	22	42	41	38	111	19	8	0	11	292
Stanley Park	6	2	8	0	109	14	3	0	12	154
Strathcona	160	124	527	81	821	108	76	2	88	1987
Sunset	37	93	175	105	382	18	63	1	93	967
Victoria-Fraserview	15	80	94	57	239	15	36	1	63	600
West End	230	72	460	455	1461	203	77	1	72	3031
West Point Grey	18	71	50	11	157	32	11	0	22	372
All	2025	2397	5721	4947	14946	2159	1146	13	1474	34828

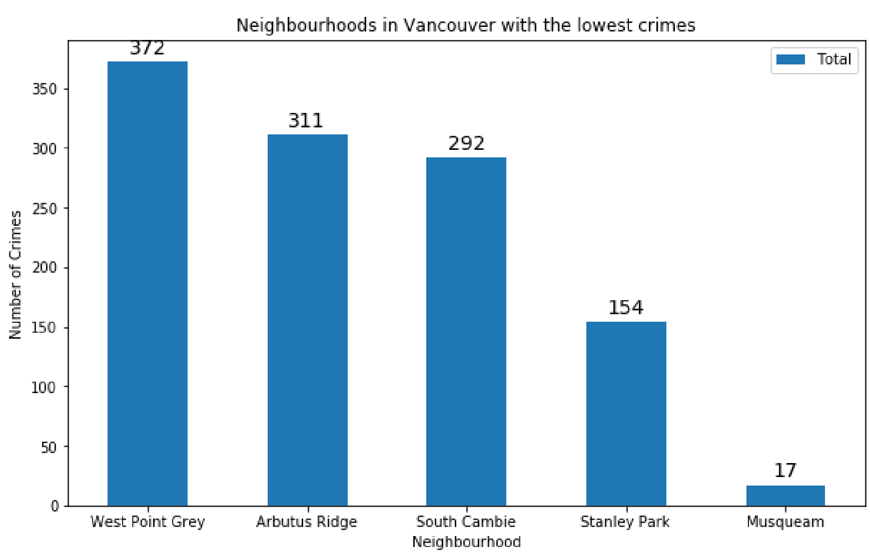
Methodology

Methodology section is categorized into two parts:

Exploratory Data Analysis

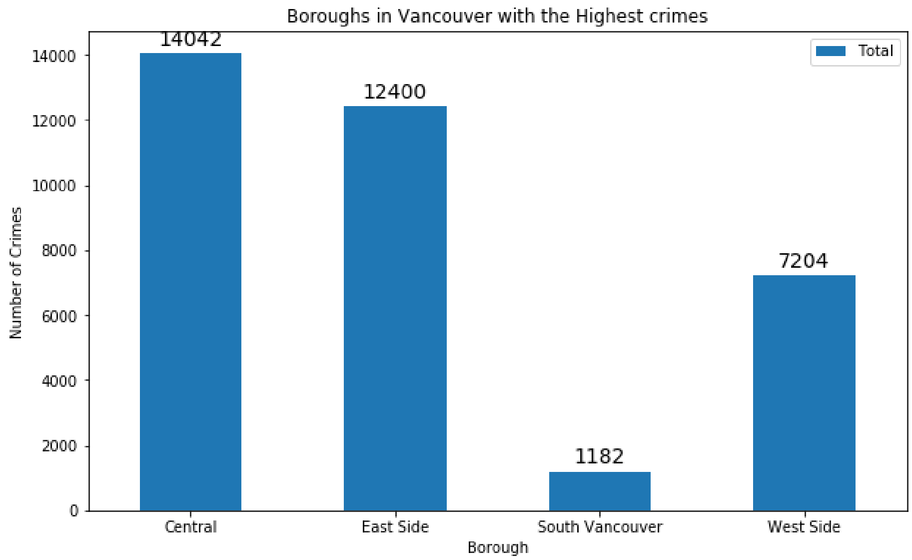
Visualize the crime reports in different Vancouver boroughs to identity the safest borough and normalize the neighborhoods of that borough. We will Use the resulting data and find 10 most common venues in each neighborhood.



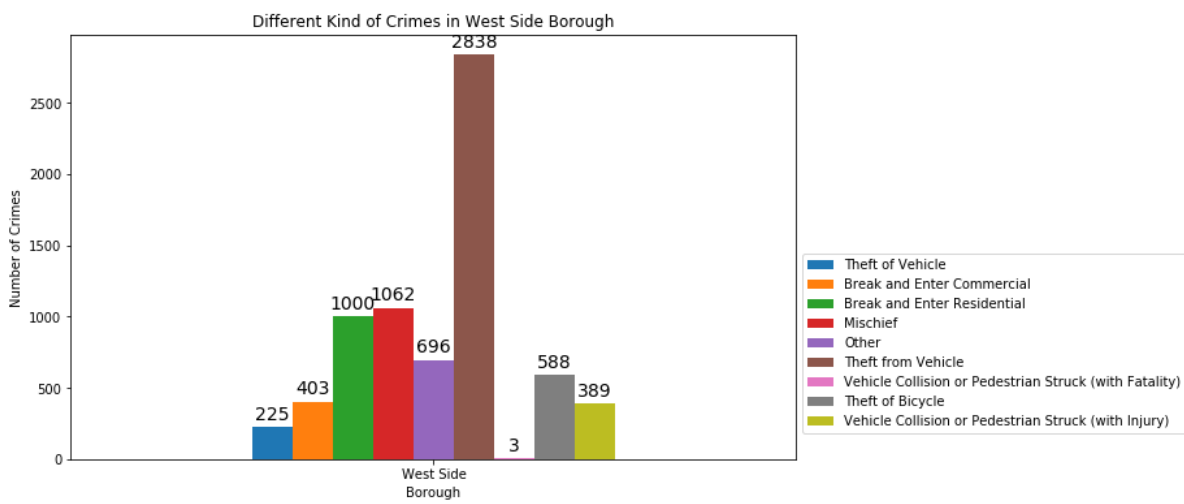


Based on exploratory data analysis it is clear that South Vancouver has the lowest crimes.

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



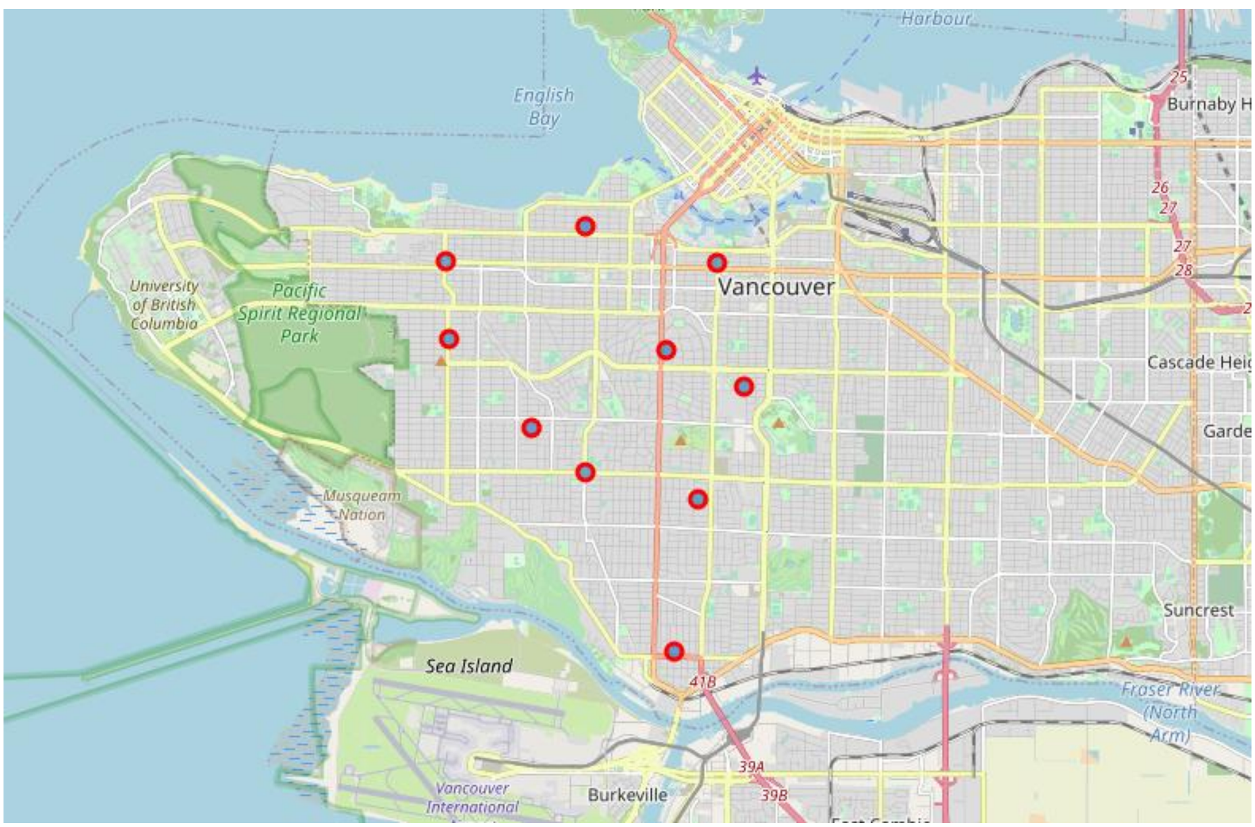
West side was chosen because crime type Break and enter Commercial is also low amongst other crimes types which makes West Side ideal destination for opening of commercial establishments.



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 OpenCage Geocoder to find the safest Borough, explore the neighborhood by plotting it on maps using Folium, and perform exploratory data analysis.

	Neighbourhood	Borough	Latitude	Longitude
0	Shaughnessy	West Side	49.251863	-123.138023
1	Fairview	West Side	49.264113	-123.126835
2	Oakridge	West Side	49.230829	-123.131134
3	Marpole	West Side	49.209223	-123.136150
4	Kitsilano	West Side	49.269410	-123.155267
5	Kerrisdale	West Side	49.234673	-123.155389
6	West Point Grey	West Side	49.264484	-123.185433
7	Arbutus Ridge	West Side	49.240968	-123.167001
8	South Cambie	West Side	49.246685	-123.120915
9	Dunbar-Southlands	West Side	49.253460	-123.185044

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 neighborhood venues and to apply machine-learning algorithm to cluster the neighbourhoods and present the findings by plotting it on maps using Folium.



- Setting Up Foursquare Credentials

```
In [63]: #Four Square Credentials

CLIENT_ID = 'XVY0YGK3DX5QGHMN2TGSK2EWA55P3JNPIVC5QVW5SGIGUI2L'
CLIENT_SECRET = 'T53Z3HT4W5DVALRIPBK2DPD4NFOCISMUTMNBLNW13KEJTAIJ'
VERSION = '20191101'
LIMIT = 100

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: XVY0YGK3DX5QGHMN2TGSK2EWA55P3JNPIVC5QVW5SGIGUI2L
CLIENT_SECRET:T53Z3HT4W5DVALRIPBK2DPD4NFOCISMUTMNBLNW13KEJTAIJ
```

- Defining a function to fetch top 10 venues around a given neighborhood and generating Venues

```
Shaughnessy
Fairview
Oakridge
Marpole
Kitsilano
Kerrisdale
West Point Grey
Arbutus Ridge
South Cambie
Dunbar-Southlands
```

- Data frame containing venues for each neighborhood in West Side

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

- Venue Count per neighborhood

	Venue
Neighbourhood	
Arbutus Ridge	5
Dunbar-Southlands	8
Fairview	27
Kerrisdale	42
Kitsilano	47
Marpole	31
Oakridge	10
Shaughnessy	2
South Cambie	17
West Point Grey	45

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. The process includes,

- One Hot Encoding to Analyze Each Neighborhood

	Neighbourhood	American Restaurant	Asian Restaurant	BBQ Joint	Bakery	Bank	Bar	Beach	Bistro	Bookstore	Boutique	Breakfast Spot	Bubble Tea Shop	Burger Joint	Bus Station	Bus Stop	Business Service	Café	Camera Store	Cheese Shop	Chinese Restaurant	Clothing Store	Coffee Shop	Convenience Store	Cosmetics Shop	Deli / Bodega	Dessert Shop	Dim Sum Restaurant	Dine
0	Shaughnessy	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Shaughnessy	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	Fairview	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	Fairview	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	Fairview	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

- Add neighborhood column back to data frame and move neighborhood column to the first column, and display the top 10 venues for each neighborhood.

	Neighbourhood	American Restaurant	Asian Restaurant	BBQ Joint	Bakery	Bank	Bar	Beach	Bistro	Bookstore	Boutique	Breakfast Spot	Bubble Tea Shop	Burger Joint	Bus Station	Bus Stop	Business Service	Café	Camera Store	Cheese Shop	Chinese Restaurant	Clothing Store	Coffee Shop	Convenience Store	Cosmetics Shop
0	Arbutus Ridge	0.000000	0.000000	0.000000	0.200000	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000
1	Dunbar-Southlands	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0.000000	0.125000	0.000000	0.000000
2	Fairview	0.000000	0.074074	0.037037	0.000000	0.037037	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0.037037	0.00000	0.037037	0.000000	0.148148	0.000000	0.000000
3	Kerrisdale	0.000000	0.023810	0.000000	0.023810	0.023810	0.000000	0.000000	0.02381	0.000000	0.02381	0.000000	0.023810	0.02381	0.000000	0.000000	0.02381	0.023810	0.000000	0.02381	0.071429	0.000000	0.095238	0.023810	0.000000
4	Kitsilano	0.042553	0.021277	0.000000	0.063830	0.000000	0.000000	0.021277	0.00000	0.000000	0.00000	0.021277	0.000000	0.00000	0.021277	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0.021277	0.042553	0.000000	0.000000
5	Marpole	0.000000	0.000000	0.000000	0.000000	0.032258	0.032258	0.000000	0.00000	0.000000	0.00000	0.000000	0.064516	0.00000	0.000000	0.032258	0.00000	0.032258	0.000000	0.00000	0.064516	0.000000	0.064516	0.000000	0.000000
6	Oakridge	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0.000000	0.00000	0.100000	0.000000	0.00000	0.000000	0.000000	0.000000	0.100000	0.000000
7	Shaughnessy	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.000000	0.000000
8	South Cambie	0.000000	0.000000	0.000000	0.000000	0.058824	0.000000	0.000000	0.00000	0.000000	0.00000	0.000000	0.000000	0.00000	0.000000	0.117647	0.00000	0.058824	0.000000	0.00000	0.000000	0.000000	0.294118	0.000000	0.000000
9	West Point Grey	0.000000	0.022222	0.000000	0.044444	0.022222	0.022222	0.000000	0.00000	0.044444	0.00000	0.000000	0.000000	0.00000	0.044444	0.000000	0.00000	0.066667	0.000000	0.00000	0.000000	0.000000	0.066667	0.022222	0.022222

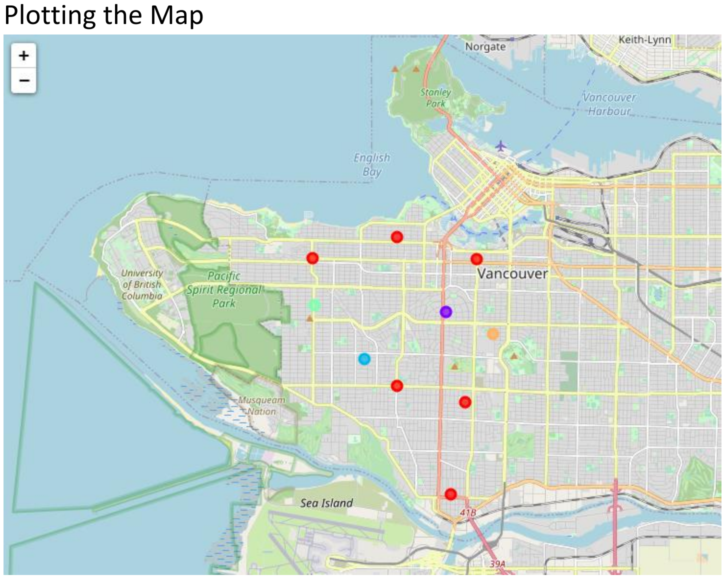
- Create columns according to number of top venues

	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	Nightlife Spot	Bakery	Pet Store	Grocery Store	Spa	Yoga Studio	Dessert Shop	Dim Sum Restaurant	Diner	Falafel Restaurant
1	Dunbar-Southlands	Sushi Restaurant	Indian Restaurant	Sporting Goods Shop	Ice Cream Shop	Italian Restaurant	Liquor Store	Coffee Shop	Gym Pool	Grocery Store	Cosmetics Shop
2	Fairview	Coffee Shop	Asian Restaurant	Park	Sushi Restaurant	Korean Restaurant	Restaurant	Falafel Restaurant	Salon / Barbershop	Sandwich Place	Nail Salon
3	Kerrisdale	Coffee Shop	Chinese Restaurant	Sandwich Place	Noodle House	Pharmacy	Tea Room	Sushi Restaurant	Cheese Shop	Gym Pool	Pizza Place
4	Kitsilano	Bakery	Food Truck	American Restaurant	Coffee Shop	Japanese Restaurant	Ice Cream Shop	French Restaurant	Sushi Restaurant	Thai Restaurant	Clothing Store

- Cluster Neighborhoods and merge toronto_grouped with Vancouver data to add latitude/longitude for each neighborhood

	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 Most Common Venue	10th Most Common Venue
0	Shaughnessy	West Side	49.251863	-123.138023	1	French Restaurant	Park	Yoga Studio	Gas Station	Deli / Bodega	Dessert Shop	Dim Sum Restaurant	Diner	Falafel Restaurant	Fast Food Restaurant
1	Fairview	West Side	49.264113	-123.126835	0	Coffee Shop	Asian Restaurant	Park	Sushi Restaurant	Korean Restaurant	Restaurant	Falafel Restaurant	Salon / Barbershop	Sandwich Place	Nail Salon
2	Oakridge	West Side	49.230829	-123.131134	0	Sporting Goods Shop	Fast Food Restaurant	Israeli Restaurant	Vietnamese Restaurant	Pharmacy	Café	Sandwich Place	Park	Sushi Restaurant	Convenience Store
3	Marpole	West Side	49.209223	-123.136150	0	Sushi Restaurant	Coffee Shop	Chinese Restaurant	Dessert Shop	Bubble Tea Shop	Pizza Place	Liquor Store	Falafel Restaurant	Bus Stop	Sandwich Place
4	Kitsilano	West Side	49.269410	-123.155267	0	Bakery	Food Truck	American Restaurant	Coffee Shop	Japanese Restaurant	Ice Cream Shop	French Restaurant	Sushi Restaurant	Thai Restaurant	Clothing Store

- Plotting the Map



Analysis

Examining the resulting Clusters

Cluster Analysis #1

vancouver_merged.loc[vancouver_merged['Cluster Labels'] == 0, vancouver_merged.columns[[1] + list(range(5, vancouver_merged.shape[1]))]]

Cluster Analysis #1

```
In [79]: vancouver_merged.loc[vancouver_merged['Cluster Labels'] == 0, vancouver_merged.columns[[1] + list(range(5, vancouver_merged.shape[1]))]]
Out[79]:
```

	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	Sushi Restaurant	Korean Restaurant	Restaurant	Falafel Restaurant	Salon / Barbershop	Sandwich Place	Nail Salon
2	West Side	Sporting Goods Shop	Fast Food Restaurant	Israeli Restaurant	Vietnamese Restaurant	Pharmacy	Café	Sandwich Place	Park	Sushi Restaurant	Convenience Store
3	West Side	Sushi Restaurant	Coffee Shop	Chinese Restaurant	Dessert Shop	Bubble Tea Shop	Pizza Place	Liquor Store	Falafel Restaurant	Bus Stop	Sandwich Place
4	West Side	Bakery	Food Truck	American Restaurant	Coffee Shop	Japanese Restaurant	Ice Cream Shop	French Restaurant	Sushi Restaurant	Thai Restaurant	Clothing Store
5	West Side	Coffee Shop	Chinese Restaurant	Sandwich Place	Noodle House	Pharmacy	Tea Room	Sushi Restaurant	Cheese Shop	Gym Pool	Pizza Place
6	West Side	Sushi Restaurant	Japanese Restaurant	Coffee Shop	Café	Bookstore	Bakery	Bus Station	Pub	Sporting Goods Shop	Vegetarian / Vegan Restaurant

Cluster Analysis #2

vancouver_merged.loc[vancouver_merged['Cluster Labels'] == 1, vancouver_merged.columns[[1] + list(range(5, vancouver_merged.shape[1]))]]

Cluster Analysis #2

```
In [80]: vancouver_merged.loc[vancouver_merged['Cluster Labels'] == 1, vancouver_merged.columns[[1] + list(range(5, vancouver_merged.shape[1]))]]
Out[80]:
```

	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	French Restaurant	Park	Yoga Studio	Gas Station	Deli / Bodega	Dessert Shop	Dim Sum Restaurant	Diner	Falafel Restaurant	Fast Food Restaurant

Cluster Analysis #3

vancouver_merged.loc[vancouver_merged['Cluster Labels'] == 2, vancouver_merged.columns[[1] + list(range(5, vancouver_merged.shape[1]))]]

Cluster Analysis #3

```
In [81]: vancouver_merged.loc[vancouver_merged['Cluster Labels'] == 2, vancouver_merged.columns[[1] + list(range(5, vancouver_merged.shape[1]))]]
Out[81]:
```

	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
7	West Side	Nightlife Spot	Bakery	Pet Store	Grocery Store	Spa	Yoga Studio	Dessert Shop	Dim Sum Restaurant	Diner	Falafel Restaurant

Cluster Analysis #4

vancouver_merged.loc[vancouver_merged['Cluster Labels'] == 3, vancouver_merged.columns[[1] + list(range(5, vancouver_merged.shape[1]))]]

Cluster Analysis #4

```
In [82]: vancouver_merged.loc[vancouver_merged['Cluster Labels'] == 3, vancouver_merged.columns[[1] + list(range(5, vancouver_merged.shape[1]))]]
Out[82]:
```

	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	Sushi Restaurant	Indian Restaurant	Sporting Goods Shop	Ice Cream Shop	Italian Restaurant	Liquor Store	Coffee Shop	Gym Pool	Grocery Store	Cosmetics Shop

Cluster Analysis #5

vancouver_merged.loc[vancouver_merged['Cluster Labels'] == 4, vancouver_merged.columns[[1] + list(range(5, vancouver_merged.shape[1]))]]

Cluster Analysis #5

```
In [83]: vancouver_merged.loc[vancouver_merged['Cluster Labels'] == 4, vancouver_merged.columns[[1] + list(range(5, vancouver_merged.shape[1]))]]
Out[83]:
```

	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
8	West Side	Coffee Shop	Bus Stop	Light Rail Station	Shopping Mall	Vietnamese Restaurant	Bank	Grocery Store	Malay Restaurant	Park	Gift Shop

Results

The objective of the business problem helps the stakeholders identify one of the safest borough in Vancouver, an appropriate neighborhood within the borough to set up a commercial establishment a Grocery store.

Discussion

This was achieved by making use of Vancouver crime data to identify a safe Borough with considerable number of neighborhood. After selecting the Borough, the result chooses the right neighborhood where grocery shops were not among venues in a close proximity to each other. The result 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

Thus, we have explored the crime data to understand the different types of crimes in all neighborhoods of Vancouver. The data has categorized them into different Borough, which helps to group the neighborhoods into boroughs and choose the safest borough first. When the confirmed Borough number of neighborhoods is down, we can further shortlist the neighborhoods based on the common venues, to choose a neighborhood which best suits the business problem.