Applied Data Science Capstone Project

The Battle of Edmonton Neighborhoods

Aarthi Velusamy

11/13/2020

Table of Contents

Solution- GitHub Link: https://github.com/Aarthiv18/Capstone/tree/main/Capstone%20Project%20-%20The%20Battle%20of%20Neighborhoods

1. Introduction

1.1 Background

The City of Edmonton, which focuses on this study, is a young, vibrant city with an exceptional quality of life that welcomes people of all abilities and backgrounds. It is North America's northernmost metropolitan area, with a population of over one million. The Greater Edmonton area had a population of 1,321,426 in 2016, making its sixth-largest metro area in Canada home to approximately one-third of Alberta population. Known as the 'Gateway to the North,' the City's robust growth is associated with large-scale oil sands projects and large-scale diamond mining operations. The City has several commercial establishments, including one of the largest shopping malls globally, 'West Edmonton Mall.' Edmonton is the largest metropolitan hub, brings many opportunities for entrepreneurs to start their business. As no exception, at the risk of population rate and people from different backgrounds, several crime hotspots are seen on the Metropolitan area's periphery. These hotspots are mostly associated with commercial properties. No wonder entrepreneurs looking to establish business venues are also concerned about neighbourhood crime rates before finalizing the location. We look to address this issue by examining Edmonton neighbourhoods' crime rate model and find them the safest, colossal demand and less competitive neighbourhood in the City that best suits their business needs.

1.2 Business Problem

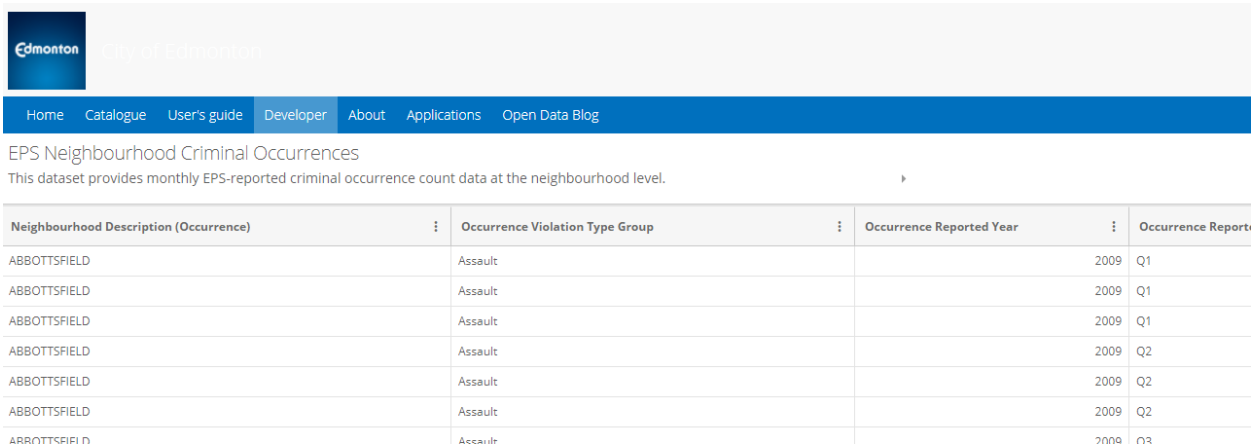
This project aims to determine the optimal solution for establishing a commercial spot in Edmonton, Canada. This project's target audiences are the stakeholders/entrepreneurs interested in opening commercial sites like Grocery Store in Edmonton. The initial step would be to choose the safest borough (geographical region) by analyzing EPS (Edmonton Police Service) crime datasets and shortlisting the neighbourhoods where grocery stores are not amongst the most common venues. We will use Foursquare location data and regional clustering of venue information to determine what might be the 'best' neighbourhood in Edmonton to open a grocery store.

2. Data Overview

This project works with five sets of data. The data web scraped is the fusion of multisource which will provide the real-world Edmonton Crime data (via Opendata), list of neighbourhoods and boroughs in Edmonton (via Wikipedia), Edmonton Postal Codes (via Wikipedia), the Geographical location of the communities (via **Geocoder package**) and Edmonton Venue data (via **Foursquare**). The Venue data will help find which neighbourhood is best suitable to open a Grocery Store.

2.1 Data Acquisition

Source 1: EPS Neighbourhood Criminal Occurrences dataset via Opendata.
https://dashboard.edmonton.ca/dataset/EPS-Neighbourhood-Criminal-Occurrences/xthe-mnvi/data



Neighbourhood Description (Occurrence)	Occurrence Violation Type Group	Occurrence Reported Year	Occurrence Report
ABBOTTSFIELD	Assault	2009	Q1
ABBOTTSFIELD	Assault	2009	Q1
ABBOTTSFIELD	Assault	2009	Q1
ABBOTTSFIELD	Assault	2009	Q2
ABBOTTSFIELD	Assault	2009	Q2
ABBOTTSFIELD	Assault	2009	Q2
ABBOTTSFIELD	Assault	2009	Q3

Following are the properties of the dataset:

```
[6]: edm_crime_df.dtypes

[6]: Neighbourhood Description (Occurrence)    object
     Occurrence Violation Type Group           object
     Occurrence Reported Year                   int64
     Occurrence Reported Quarter               object
     Occurrence Reported Month                  int64
     # Occurrences                              int64
     dtype: object
```

Source 2: List of neighbourhoods and boroughs in Edmonton
https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Edmonton

West	Riverview	Grandisle · River's Edge · Stillwater · The Uplands · White Birch	
	West Jasper Place (area)	Aldergrove · Belmead · Callingwood North · Callingwood South · Dechene · Donsdale · Gariep	
	Cameron Heights · Edgemont · Place LaRue		
Southwest	Heritage Valley	Allard · Blackburne · Blackmud Creek · Callaghan · Cashman · Cavanagh · Chappelle · Desrochers ·	
	Kaskitayo	Blue Quill	Blue Quill (neighbourhood) · Blue Quill Estates · Skyrattler · Sweet Grass
		Yellowbird	Bearspaw · Keheewin
		Ermineskin · Steinhauer · Twin Brooks	
	Riverbend	Brander Gardens · Brookside · Bulyea Heights · Carter Crest · Falconer Heights · Henderson Estates	
	Terwillegar Heights	Haddow · Hodgson · Leger · Mactaggart · Magrath Heights · South Terwillegar · Terwillegar Towne	
	Windermere (area)	Ambleside · Glenridding Heights · Glenridding Ravine · Keswick · Windermere	
	Ellerslie (area)	Ellerslie · The Orchards at Ellerslie · Summerside	
	The Meadows	Aster · Larkspur · Laurel · Maple · Silver Berry · Tamarack · Wild Rose	

Dataset properties:

```
edm_neigh_bor.dtypes
```

```
Neighbourhood Description (Occurrence)    object
Borough                                   object
Area                                      object
dtype: object
```

Source 3: Edmonton Postal Codes
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_T

The third data source is a Wikipedia page that contains the Postcode of the City of Edmonton in a wiki table. To scrape the data from the URL, pandas/BeautifulSoup has been used.

Postal Code	Borough	Neighborhood
T1A	Medicine Hat	Central Medicine Hat
T2A	Calgary	Penbrooke Meadows, Marlborough
T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley
T4A	Airdrie	East Airdrie
T5A	Edmonton	West Clareview, East Londonderry
T6A	Edmonton	North Capilano
T7A	Drayton Valley	Not assigned
T8A	Sherwood Park	West Sherwood Park
T9A	Wetaskiwin	Not assigned

Source 4: The Geographical location of the neighbourhoods (via **Geocoder package**)

The retrieval of Geographical coordinates of the Edmonton neighbourhoods requires Geocoder package registration and credentials. A glimpse of the dataset after retrieving the latitude and longitude coordinates from OpenCage API.

	Neighbourhood	Borough	Latitude	Longitude
0	AMBLESIDE	Southwest	53.429904	-113.599337
1	BEARSPAW	Southwest	53.443342	-113.501130
2	BLACKBURNE	Southwest	53.431102	-113.501211
3	BLACKMUD CREEK	Southwest	53.424864	-113.498876
4	BLUE QUILL	Southwest	53.459598	-113.526675
5	BRANDER GARDENS	Southwest	53.494351	-113.577737
6	BROOKSIDE	Southwest	53.491199	-113.566713

Dataset properties:

```
sw_neig_geo.dtypes
```

```
Neighbourhood     object
Borough           object
Latitude         float64
Longitude        float64
dtype: object
```

Source 5: Venue Data using Foursquare

The fourth set of the data source is generated from the FourSquare API. The generated data has a list of columns listed below:

Neighbourhood
Neighbourhood Latitude
Neighborhood Longitude
Venue
Venue Latitude
Venue Longitude
Venue Category

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
AMBLESIDE	53.429904	-113.599337	Rental Gladiators	53.433188	-113.601723	Construction & Landscaping
AMBLESIDE	53.429904	-113.599337	Ambleside Park	53.432911	-113.604124	Park
BEARSPAW	53.443342	-113.501130	Bears Paw Dog Park	53.440195	-113.504602	Outdoors & Recreation
BEARSPAW	53.443342	-113.501130	Bearspaw Lake	53.440056	-113.504465	Lake

To obtain this data, it was required to make an account where it would provide a 'Secret Key' and 'Client ID' to develop the app.

2.2 Data Cleaning

As the following data wrangling step, cleansing and merging, the data were required to perform the analysis. When we pulled data from Wikipedia, some neighbourhoods were not assigned any Boroughs; therefore, the following assumptions were made:

1. Only the rows that have an assigned borough will be processed.
2. More than one neighbourhood can exist in an area. These rows were processed to structure the data.

3. Using the Neighborhood boroughs collected from Wikipedia, we merged the Edmonton crime dataset and borough dataset. After joining the tables, the data frame is further cleaned by dropping records with inconsistent or invalid data like 'Not Entered' 'Nan' values.

	Neighbourhood	Crime Type	Year	Quarter	Month	Number Of Occurences	Borough	Area
0	ABBOTTSFIELD	Assault	2009	Q1	1	3	East	Beverly
1	ABBOTTSFIELD	Assault	2009	Q1	2	5	East	Beverly
2	ABBOTTSFIELD	Assault	2009	Q1	3	2	East	Beverly
3	ABBOTTSFIELD	Assault	2009	Q2	4	2	East	Beverly
4	ABBOTTSFIELD	Assault	2009	Q2	5	6	East	Beverly

4. We are pivoting the table to represent the data in a better understanding format.
To determine the total number of neighbourhoods under each borough and area:

		Neighbourhood
Borough	Area	
Central	Central core north	8
	Central core south	4
East	Beverly	5
	Greater Highlands	6
	The Avenue	7
North	Castle Downs	10
	Dickinsfield	2
	Goodridge Corners	1
	Lake District	8
	Londonderry	2
	The Palisades	6
	Transportation and utility corridor	2
Northeast	Casselman-Steele Heights	4
	Clareview	8
	Hermitage	3
	Horse Hill	1
	Northeast Industrial	4

Edmonton region(boroughs) and their crime type count determined by pivoting the table:

	Region	Assault	Break and Enter	Homicide	Robbery	Sexual Assaults	Theft from Vehicle	Theft Of Vehicle	Theft Over $5000
1	Southeast	3227	3856	22	794	712	4617	3278	644
2	West	2670	2917	21	1080	757	3475	2437	303
3	South	1865	2507	10	696	503	2726	1891	243
4	North	1741	2026	19	593	467	2656	1893	256
5	East	1920	1853	40	1029	658	2006	1756	196
6	Northeast	1731	1548	8	481	413	2167	1641	212
7	Central	1277	1335	53	967	772	1372	1175	330
8	Southwest	1170	1839	11	335	246	2223	1143	286
9	Northwest	742	1876	5	122	112	2124	1519	497

Similarly, neighbourhoods and their crime type count representation by pivoting the table:

	Neighbourhood	YearAssault	YearBreak and Enter	YearHomicide	YearRobbery	YearSexual Assaults	YearTheft From Vehicle	YearTheft Of Vehicle	YearTheft Over $5000	Total
0	All	16343	19757	189	6097	4640	23366	16733	2967	90092
1	DOWNTOWN	131	131	12	131	117	131	130	78	861
2	BOYLE STREET	131	126	10	117	113	131	125	37	790
3	OLIVER	131	130	7	119	97	130	129	45	788
4	CENTRAL MCDOUGALL	131	125	7	119	102	131	126	43	784
5	MCCAULEY	131	127	10	125	100	130	126	21	770
6	ALBERTA AVENUE	131	131	12	116	81	131	130	18	750
7	STRATHCONA	130	130	2	109	75	131	126	28	731
8	EASTWOOD	131	130	3	121	65	126	128	18	722
9	QUEEN MARY PARK	131	128	4	94	69	131	130	26	713
10	SUMMERLEA	130	104	0	90	91	130	122	44	711
11	SPRUCE AVENUE	130	126	1	90	58	129	119	37	690
12	EMPIRE PARK	129	109	0	87	58	130	117	33	663
13	PARKDALE	130	123	3	84	47	129	124	17	657

5. We also have to fetch the latitude and longitude data to analyze the crime data; to achieve this, a dictionary was initialized and converted to a data frame.

	Neighbourhood	Borough	Latitude	Longitude
0	AMBLESIDE	Southwest	53.429904	-113.599337
1	BEARSPAW	Southwest	53.443342	-113.501130
2	BLACKBURNE	Southwest	53.431102	-113.501211
3	BLACKMUD CREEK	Southwest	53.424864	-113.498876
4	BLUE QUILL	Southwest	53.459598	-113.526675
5	BRANDER GARDENS	Southwest	53.494351	-113.577737

2.3 Feature Selection

Now that we have obtained the different neighbourhoods and their respective geometric coordinates for Edmonton city, it is time to come up with various venues that the communities have to offer.

Foursquare API provides an enormous database consisting of venues from all around the world. Having signed in for a Foursquare developer, using the Client ID and Client Secret, it is possible to make API requests to retrieve venue details.

By feeding the function with Neighborhood name and its geometric coordinates, using Foursquare API, different venues (restaurants, coffee shops, gym, Yoga, etc.) were extracted.

One-Hot-Encoding to devise the feature set: After performing One-Hot-Encoding, the Edmonton dataset seemed to share 64 features. (One hot encoding is a process in which categorical values are transformed into Numerical values for Machine learning algorithms to make a better prediction.)

```
The Venue Categories are ['Construction & Landscaping' 'Park' 'Outdoors & Recreation' 'Lake'
 'Playground' 'Gift Shop' 'Coffee Shop' 'Chinese Restaurant' 'Hotel'
 'Convenience Store' 'Gym' 'Pizza Place' 'School' 'Café' 'Home Service'
 'Soccer Field' 'Pub' 'Baby Store' 'Insurance Office' 'Grocery Store'
 'Smoothie Shop' 'Asian Restaurant' 'Sandwich Place'
 'Fast Food Restaurant' 'Indian Restaurant' 'Yoga Studio' 'Restaurant'
 'Discount Store' 'Wine Shop' 'Japanese Restaurant' 'Ice Cream Shop'
 'Tea Room' 'Salon / Barbershop' 'Bus Station' 'Liquor Store' 'Spa'
 'Cheese Shop' 'Skate Park' 'Paper / Office Supplies Store'
 'Shopping Mall' 'Italian Restaurant' 'Shoe Store' 'Cosmetics Shop'
 'Speakeasy' 'Miscellaneous Shop' 'Roof Deck' 'Bridal Shop'
 'Business Service' 'Nature Preserve' 'Trail' 'Motorcycle Shop'
 'Rental Car Location' 'Performing Arts Venue' 'BBQ Joint' 'Gas Station'
 'Stadium' 'Massage Studio' 'Pharmacy' 'ATM' 'Dog Run' 'Dry Cleaner'
 'Garden Center' 'Moving Target' 'Electronics Store']
```

3. Methodology

3.1 Exploratory Data Analysis

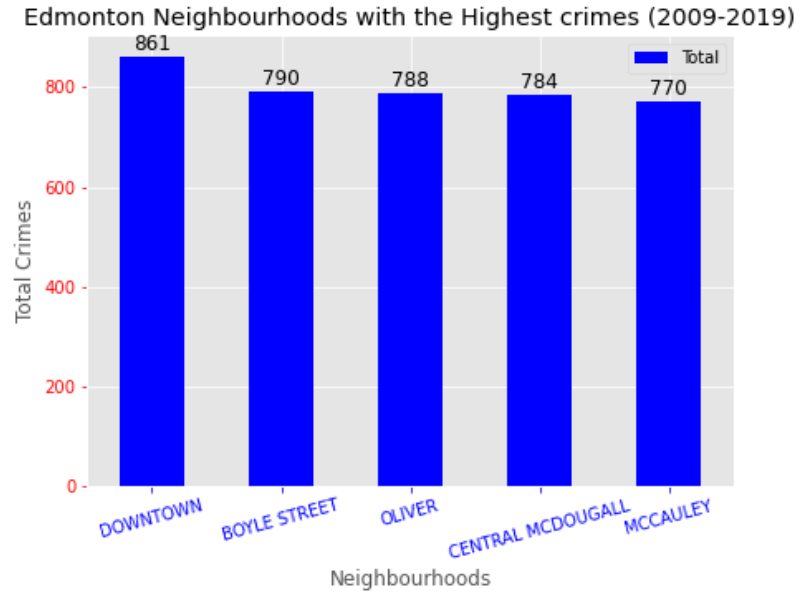
3.1.1 Statistical Summary of crimes

The describe function in python is used to get statistics of the crime data that returns the count, mean, standard deviation, minimum, maximum, 1st quartile (25%), 2nd quartile (50%), and the 3rd quartile (75%) for each of the major categories of crime.

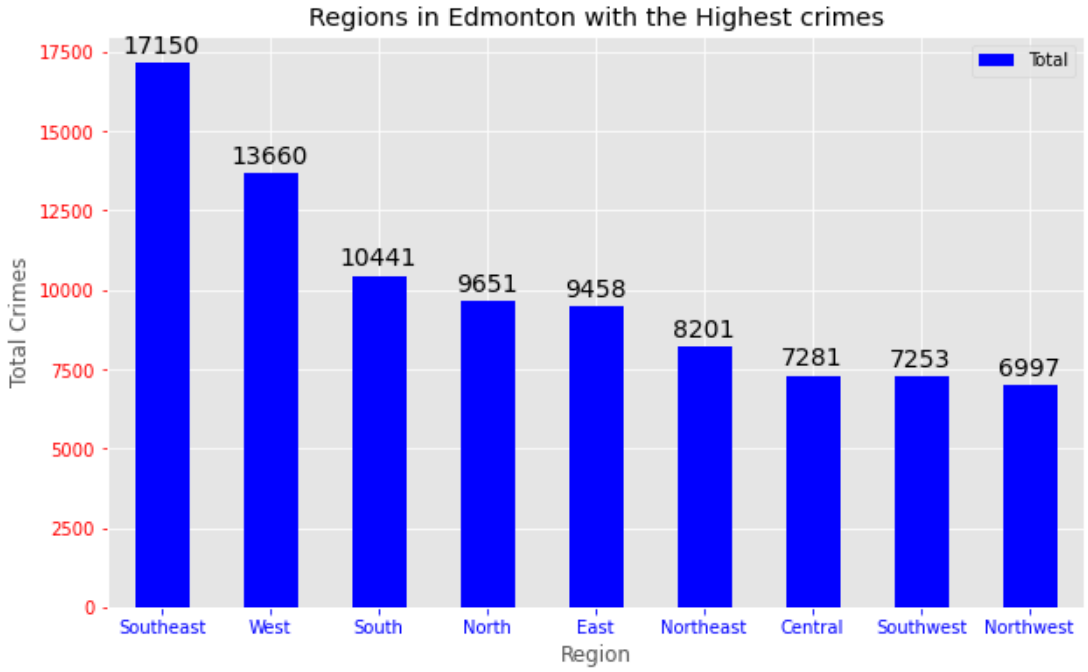
	Assault	Break and Enter	Homicide	Robbery	Sexual Assaults	Theft from Vehicle	Theft Of Vehicle	Theft Over $5000	Total
count	10.000000	10.000000	10.00000	10.000000	10.000000	10.00000	10.000000	10.000000	10.000000
mean	3268.600000	3951.400000	37.80000	1219.400000	928.000000	4673.20000	3346.600000	593.400000	18018.400000
std	4649.263722	5602.422238	55.21433	1741.306866	1322.515616	6629.18726	4744.709433	845.514203	25525.809606
min	742.000000	1335.000000	5.00000	122.000000	112.000000	1372.00000	1143.000000	196.000000	6997.000000
25%	1390.500000	1842.500000	10.25000	509.000000	426.500000	2134.75000	1549.500000	246.250000	7511.000000
50%	1803.000000	1951.000000	20.00000	745.000000	580.500000	2439.50000	1823.500000	294.500000	9554.500000
75%	2482.500000	2814.500000	35.50000	1013.500000	745.750000	3287.75000	2301.000000	455.250000	12855.250000
max	16343.000000	19757.000000	189.00000	6097.000000	4640.000000	23366.00000	16733.000000	2967.000000	90092.000000

3.1.2 Neighborhoods with the highest crime rates

Comparing crime rates among all the neighbourhoods, Downtown takes the significant chunk of the criminal records. But when it comes to the region, the crime rate is high in the Southeast region.

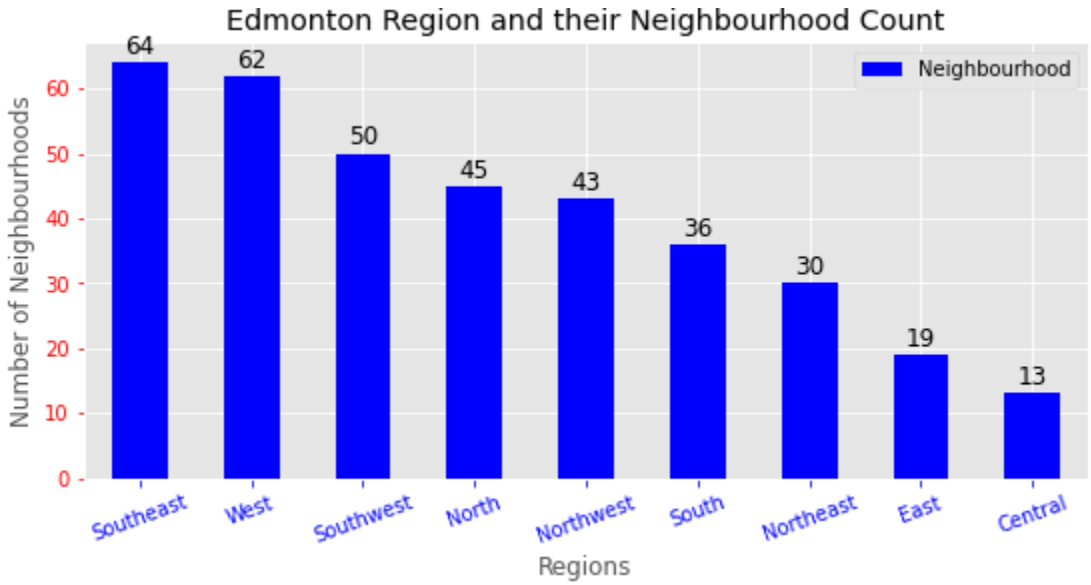


3.1.3 Boroughs (Regions) crime Analysis



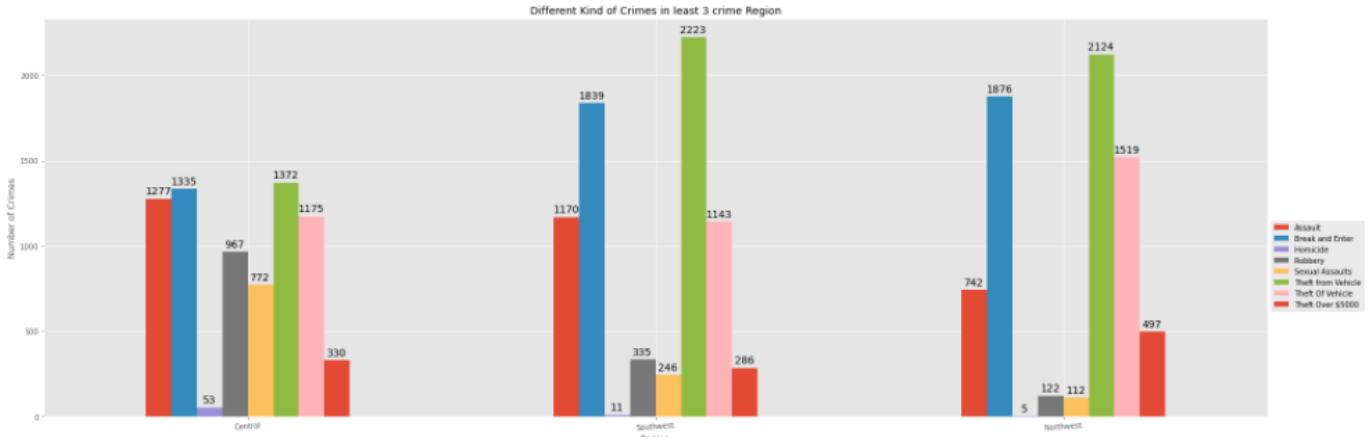
3.1.4 Neighborhoods in each region

Let's take a count of neighbourhoods in each area.



Based on exploratory data analysis, it is clear that Northwest has the lowest crimes. Next in place are Southwest and Central with minimal increase in crime rate compared to Northwest. Northwest has 43 Neighbourhoods, opening a commercial establishment in this region is an optimal idea. However, Southwest has 50 neighbourhoods with a bottom 2nd position in the crime rate, and Central has only 13 neighbourhoods with a bottom 3rd position in the crime rate. Let's compare different types of crime recorded in Northwest and Southwest to support our decision-making process further.

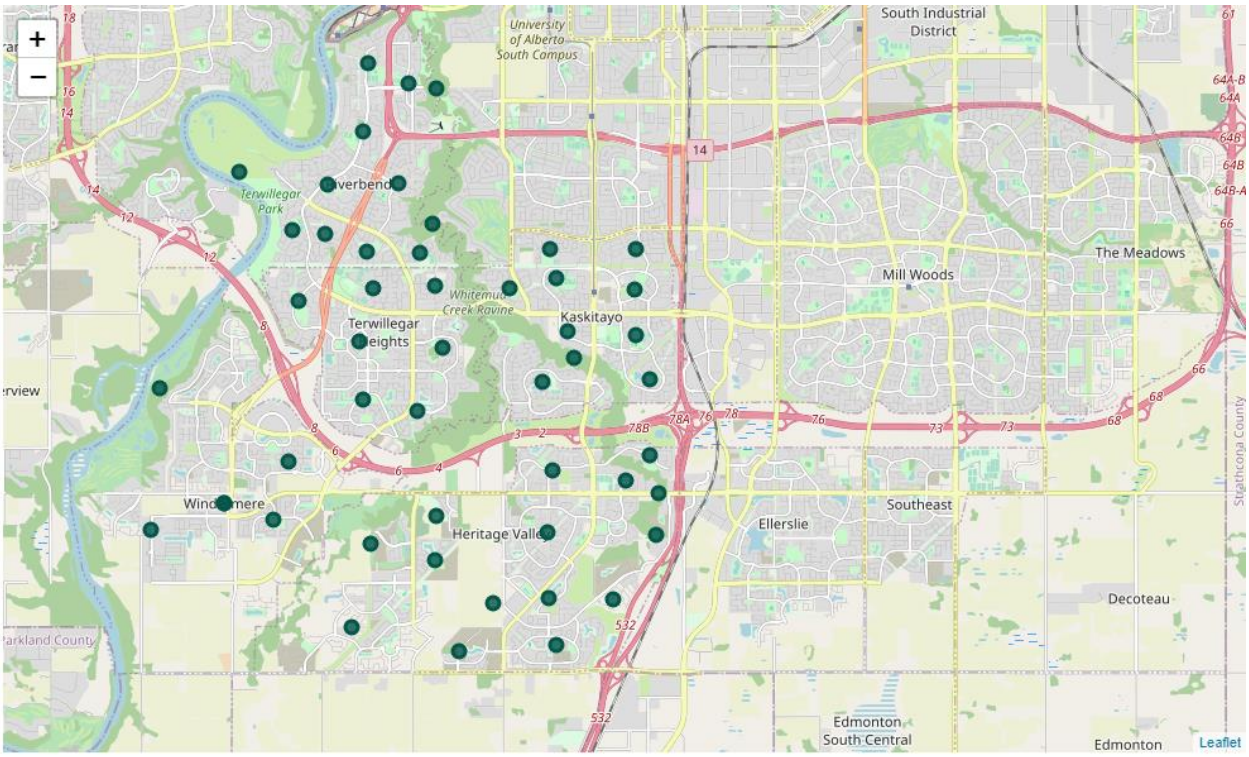
3.1.4 Crime Types for the least three regions



Southwest has more neighbourhoods than Northwest, and the crime type Break is also low compared to Northwest, which makes ***Southwest the ideal destination for opening commercial establishments like a Grocery store.***

3.1.5 Neighborhoods in Southwest, Edmonton, Canada

There are 50 neighbourhoods in the southside Edmonton borough, colour coded in the green circle, visualized on the map using Folium.



3.2 Modelling

We have defined the final dataset with neighbourhood and borough with their corresponding latitude and longitude; for Southwest Edmonton, we can find all the venues within a 500-meters radius of each community by connecting to the Foursquare API. The resulting response is in JSON format, which has to be converted to the Pandas data frame for further analysis.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	AMBLESIDE	53.429904	-113.599337	Rental Gladiators	53.433188	-113.601723	Construction & Landscaping
1	AMBLESIDE	53.429904	-113.599337	Ambleside Park	53.432911	-113.604124	Park
2	BEARSPAW	53.443342	-113.501130	Bears Paw Dog Park	53.440195	-113.504602	Outdoors & Recreation
3	BEARSPAW	53.443342	-113.501130	Bearspaw Lake	53.440056	-113.504465	Lake
4	BLACKBURNE	53.431102	-113.501211	Blackburne Park	53.429431	-113.497339	Playground
...	...	...	...	...	...	...	...
130	KESWICK AREA	53.419001	-113.636786	Rohit Communities | Keswick Single Family Home...	53.414668	-113.637663	Construction & Landscaping
131	HAYS RIDGE AREA	53.416785	-113.577041	Helping Hands Family Movers Inc	53.414494	-113.573356	Moving Target
132	BLACKMUD CREEK RAVINE	53.446738	-113.521741	7-Eleven	53.449085	-113.515486	Convenience Store
133	BLACKMUD CREEK RAVINE	53.446738	-113.521741	Bestbuy	53.446531	-113.517994	Electronics Store
134	BLACKMUD CREEK RAVINE	53.446738	-113.521741	Twin Brooks Park	53.443861	-113.526685	Playground

To analyze the data further, we need to convert categorical values to numerical data. One hot encoding is performed on the venue's data.

	Neighborhood	ATM	Asian Restaurant	BBQ Joint	Baby Store	Bridal Shop	Bus Station	Business Service	Café	Cheese Shop	...	Skate Park	Smoothie Shop	Soccer Field	Spa	Speakeasy	Stadium	Tea Room	Tr
0	AMBLESIDE	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	
1	AMBLESIDE	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	
2	BEARSPAW	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	
3	BEARSPAW	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	
4	BLACKBURNE	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	

5 rows × 65 columns

3.2.1 Group neighbourhoods by the average of the frequency of each venue

Then we grouped the venue's data by neighbourhood by taking the average of each venue category's frequency of occurrence.

	Neighborhood	ATM	Asian Restaurant	BBQ Joint	Baby Store	Bridal Shop	Bus Station	Business Service	Café	Cheese Shop	...	Skate Park	Smoothie Shop	Soccer Field	Spa	Speakeasy	Stadium	Tea Room	Tr
0	ALLARD	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	0.0	0.0	0.0	
1	AMBLESIDE	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	0.0	0.0	0.0	
2	BEARSPAW	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	0.0	0.0	0.0	
3	BLACKBURNE	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	0.0	0.0	0.0	
4	BLACKMUD CREEK	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	0.0	0.0	0.0	

3.2.2 Statistical Summary of crimes

	ATM	Asian Restaurant	BBQ Joint	Baby Store	Bridal Shop	Bus Station	Business Service	Café	Cheese Shop	Chinese Restaurant	...	Skate Park	Smoothie Shop	Soccer Field	
count	47.000000	47.000000	47.000000	47.000000	47.000000	47.000000	47.000000	47.000000	47.000000	47.000000	...	47.000000	47.000000	47.000000	4
mean	0.003546	0.002128	0.010638	0.005319	0.005319	0.004255	0.005319	0.005319	0.004255	0.009574	...	0.004255	0.002128	0.007092	
std	0.024311	0.014586	0.072932	0.036466	0.036466	0.029173	0.036466	0.036466	0.029173	0.046202	...	0.029173	0.014586	0.048622	
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	
75%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	
max	0.166667	0.100000	0.500000	0.250000	0.250000	0.200000	0.250000	0.250000	0.200000	0.250000	...	0.200000	0.100000	0.333333	

3.2.3 New Data frame storing neighbourhoods and the average grocery store in that neighbourhood

For later analysis in this project, we also created a new data frame that kept the neighbourhoods and their mean frequency of Grocery stores.

:

	Neighborhood	Grocery Store
0	ALLARD	0.0
1	AMBLESIDE	0.0
2	BEARSPAW	0.0
3	BLACKBURNE	0.0
4	BLACKMUD CREEK	0.0

3.2.4 Finding 15 common venues for each neighbourhood

After grouping the venue's data by neighbourhood and calculating their mean, we estimated 15 common venues for each neighbourhood.

	Neighborhood	1st Most Proximity Venue	2nd Most Proximity Venue	3rd Most Proximity Venue	4th Most Proximity Venue	5th Most Proximity Venue	6th Most Proximity Venue	7th Most Proximity Venue	8th Most Proximity Venue	9th Most Proximity Venue	10th Most Proximity Venue	11th Most Proximity Venue	12th Most Proximity Venue	13th Most Proximity Venue	14th Most Proximity Venue	15th Most Proximity Venue
0	ALLARD	Construction & Landscaping	Pizza Place	Massage Studio	Coffee Shop	Yoga Studio	Grocery Store	Garden Center	Gas Station	Gift Shop	Gym	Electronics Store	Home Service	Hotel	Ice Cream Shop	Indian Restaurant
1	AMBLESIDE	Park	Construction & Landscaping	Yoga Studio	Dry Cleaner	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant
2	BEARSPAW	Lake	Outdoors & Recreation	Japanese Restaurant	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	Electronics Store
3	BLACKBURNE	Construction & Landscaping	Playground	Yoga Studio	Dry Cleaner	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant
4	BLACKMUD CREEK	Hotel	Gift Shop	Coffee Shop	Chinese Restaurant	Yoga Studio	Grocery Store	Fast Food Restaurant	Garden Center	Gas Station	Gym	Dry Cleaner	Home Service	Ice Cream Shop	Indian Restaurant	Insurance Office
5	BLACKMUD CREEK RAVINE	Convenience Store	Playground	Electronics Store	Yoga Studio	Dry Cleaner	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center
6	BLUE QUILL	Convenience Store	Gym	Pizza Place	Yoga Studio	Dry Cleaner	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant
7	BLUE QUILL ESTATES	Trail	Bridal Shop	Business Service	Nature Preserve	Yoga Studio	Gym	Fast Food Restaurant	Garden Center	Gas Station	Gift Shop	Grocery Store	Home Service	Dry Cleaner	Hotel	Ice Cream Shop
8	BRANDER GARDENS	Café	School	Home Service	Chinese Restaurant	Grocery Store	Fast Food Restaurant	Garden Center	Gas Station	Gift Shop	Yoga Studio	Dry Cleaner	Gym	Hotel	Ice Cream Shop	Indian Restaurant
9	BROOKSIDE	Pub	Park	Soccer Field	Gift Shop	Dry Cleaner	Electronics Store	Fast Food Restaurant	Garden Center	Gas Station	Yoga Studio	Dog Run	Gym	Home Service	Hotel	Ice Cream Shop
10	BULYEA HEIGHTS	Pizza Place	Baby Store	Park	Yoga Studio	Grocery Store	Fast Food Restaurant	Garden Center	Gas Station	Gift Shop	Gym	Dry Cleaner	Home Service	Hotel	Ice Cream Shop	Indian Restaurant
11	CALLAGHAN	Construction & Landscaping	ATM	Pharmacy	Hotel	Coffee Shop	Gift Shop	Fast Food Restaurant	Garden Center	Gas Station	Grocery Store	Dry Cleaner	Gym	Home Service	Ice Cream Shop	Indian Restaurant

We then took the dataset that provided the mean frequency of grocery store occurrences (3.2.3) and merged with the top 15 common venues dataset for later analysis, which would be the basis for analyzing new opportunities for opening a new grocery store in SW borough.

3.2.5 K-Means Clustering

The dataset is unlabeled, i.e. unsupervised, to help people find similar neighbourhoods in the safest borough, we will be clustering similar neighbourhoods using **K-Means Clustering,** a form of an unsupervised algorithm that clusters data based on predefined cluster size (kclusters=5). In this project, we used the cluster size of five to cluster 50 neighbourhoods into 5 clusters.

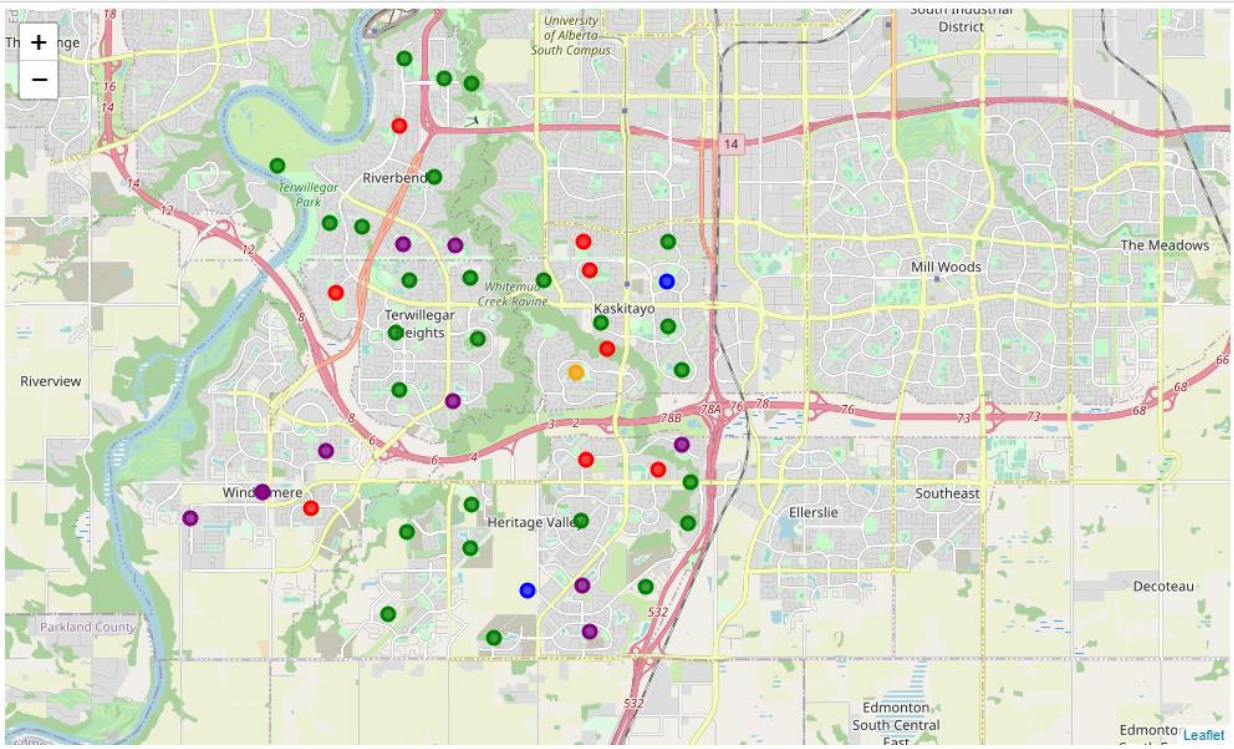
By clustering similar neighbourhoods, people can find the area of interest based on venues around the neighbourhood.

Mapping the neighbourhoods cluster using Folium

We mapped the clustered neighbourhood using Folium, and each neighbourhood was coloured based on their cluster.

- Cluster1: orange
- Cluster2: red
- Cluster3: purple
- Cluster4: blue
- Cluster5: green

The map below shows the clusters that had a similar mean frequency of grocery stores. We can see that the majority of the neighbourhood falls in cluster5. The remaining neighbourhoods are part of other clusters and have been represented in different clusters.

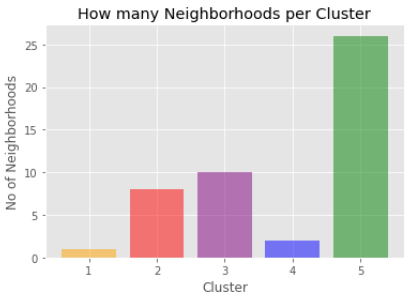


4. Results

4.1 Cluster Analysis

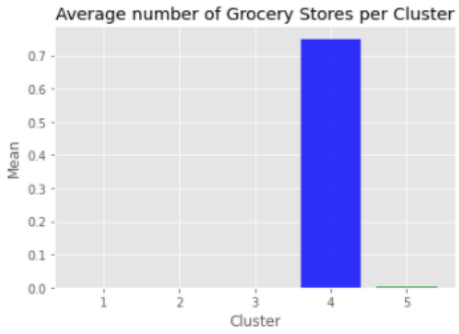
After running the K-means clustering, we can assess each cluster created.

There are 5 clusters defined. Here's the total number of neighbourhoods in each set. Cluster1 has the least neighbourhood (1), while Cluster5 has the most (26).



4	26
2	10
1	8
3	2
0	1

Let's compare the average grocery store in each cluster.



It is evident from the above two plots that there are only two neighbourhoods in cluster4 but has the highest number of grocery stores (0.75 average), whereas the cluster5 with 26 neighbourhoods has only (0.1) 2nd least average of grocery stores. In comparison, the rest of the clusters have 0 average grocery stores.

Let's analyze each cluster.

Cluster1 (Orange): Looking into the neighbourhoods in the first cluster, it is the smallest cluster with one neighbourhood. This is because of the unique venues it couldn't be clustered into a similar neighbourhood.

	Neighbourhood	Cluster_Label	1st Most Proximity Venue	2nd Most Proximity Venue	3rd Most Proximity Venue	4th Most Proximity Venue	5th Most Proximity Venue	6th Most Proximity Venue	7th Most Proximity Venue	8th Most Proximity Venue	9th Most Proximity Venue	10th Most Proximity Venue	11th Most Proximity Venue	12th Most Proximity Venue	13th Most Proximity Venue	14th Most Proximity Venue	15th Most Proximity Venue	Grocery Store
23	TWIN BROOKS	0	Playground	Yoga Studio	Japanese Restaurant	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	Electronics Store	0.0

Cluster2 (red): Upon closely examining these neighbourhoods, we can see that the most common venues are a Convenience store, Fast food, Gift shop, Pizza, Yoga Studio but has eight neighbourhoods least in number compared to cluster 3 and 5.

	Neighbourhood	Cluster_Label	1st Most Proximity Venue	2nd Most Proximity Venue	3rd Most Proximity Venue	4th Most Proximity Venue	5th Most Proximity Venue	6th Most Proximity Venue	7th Most Proximity Venue	8th Most Proximity Venue	9th Most Proximity Venue	10th Most Proximity Venue	11th Most Proximity Venue	12th Most Proximity Venue	13th Most Proximity Venue	14th Most Proximity Venue	15th Most Proximity Venue	Grocery Store
4	BLUE QUILL	1	Convenience Store	Gym	Pizza Place	Yoga Studio	Dry Cleaner	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	0.0
11	HADDOW	1	Yoga Studio	Pizza Place	Convenience Store	Park	Gift Shop	Fast Food Restaurant	Garden Center	Gas Station	Grocery Store	Dry Cleaner	Gym	Home Service	Hotel	Ice Cream Shop	Indian Restaurant	0.0
15	MACEWAN	1	Playground	Liquor Store	Convenience Store	Spa	Gift Shop	Electronics Store	Fast Food Restaurant	Garden Center	Gas Station	Yoga Studio	Grocery Store	Dry Cleaner	Home Service	Hotel	Ice Cream Shop	0.0
16	RAMSAY HEIGHTS	1	Park	Skate Park	Convenience Store	Cheese Shop	Dry Cleaner	Fast Food Restaurant	Garden Center	Gas Station	Gift Shop	Yoga Studio	Grocery Store	Gym	Home Service	Hotel	Ice Cream Shop	0.0
21	SWEET GRASS	1	Convenience Store	Gym	Pizza Place	Yoga Studio	Dry Cleaner	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	0.0
35	RICHFORD	1	Convenience Store	Stadium	Gas Station	Yoga Studio	Grocery Store	Electronics Store	Fast Food Restaurant	Garden Center	Gift Shop	Gym	Dog Run	Home Service	Hotel	Ice Cream Shop	Indian Restaurant	0.0
40	GLENRIDDING AREA	1	Convenience Store	Fast Food Restaurant	Yoga Studio	Dry Cleaner	Italian Restaurant	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	0.0
46	BLACKMUD CREEK RAVINE	1	Convenience Store	Playground	Electronics Store	Yoga Studio	Dry Cleaner	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	0.0

Cluster3 (purple): This cluster has mostly utility places Construction Landscaping, Dry Cleaner, Insurance office, Home service, some grocery stores in not proximity and some restaurants. This cluster can also be suggested for opening a grocery store.

	Neighbourhood	Cluster_Label	1st Most Proximity Venue	2nd Most Proximity Venue	3rd Most Proximity Venue	4th Most Proximity Venue	5th Most Proximity Venue	6th Most Proximity Venue	7th Most Proximity Venue	8th Most Proximity Venue	9th Most Proximity Venue	10th Most Proximity Venue	11th Most Proximity Venue	12th Most Proximity Venue	13th Most Proximity Venue	14th Most Proximity Venue	15th Most Proximity Venue	Grocery Store
0	AMBLESIDE	2	Park	Construction & Landscaping	Yoga Studio	Dry Cleaner	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	0.0
2	BLACKBURNE	2	Construction & Landscaping	Playground	Yoga Studio	Dry Cleaner	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	0.0
25	WINDERMERE	2	Construction & Landscaping	Pub	Yoga Studio	Dry Cleaner	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	0.0
27	CARTER CREST	2	Construction & Landscaping	Yoga Studio	Dry Cleaner	Italian Restaurant	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	0.0
32	MACTAGGART	2	BBQ Joint	Construction & Landscaping	Yoga Studio	Gym	Fast Food Restaurant	Garden Center	Gas Station	Gift Shop	Grocery Store	Home Service	Dry Cleaner	Hotel	Ice Cream Shop	Indian Restaurant	Insurance Office	0.0
34	OGILVIE RIDGE	2	Construction & Landscaping	Yoga Studio	Dry Cleaner	Italian Restaurant	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	0.0
36	ALLARD	2	Construction & Landscaping	Pizza Place	Massage Studio	Coffee Shop	Yoga Studio	Grocery Store	Garden Center	Gas Station	Gift Shop	Gym	Electronics Store	Home Service	Hotel	Ice Cream Shop	Indian Restaurant	0.0
37	CALLAGHAN	2	Construction & Landscaping	ATM	Pharmacy	Hotel	Coffee Shop	Gift Shop	Fast Food Restaurant	Garden Center	Gas Station	Grocery Store	Dry Cleaner	Gym	Home Service	Ice Cream Shop	Indian Restaurant	0.0
38	WINDERMERE AREA	2	Construction & Landscaping	Pub	Yoga Studio	Dry Cleaner	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	0.0
44	KESWICK AREA	2	Construction & Landscaping	Yoga Studio	Dry Cleaner	Italian Restaurant	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	0.0

Cluster4 (blue): Cluster4 has two neighbourhoods with proximity to grocery stores.

	Neighbourhood	Cluster_Label	1st Most Proximity Venue	2nd Most Proximity Venue	3rd Most Proximity Venue	4th Most Proximity Venue	5th Most Proximity Venue	6th Most Proximity Venue	7th Most Proximity Venue	8th Most Proximity Venue	9th Most Proximity Venue	10th Most Proximity Venue	11th Most Proximity Venue	12th Most Proximity Venue	13th Most Proximity Venue	14th Most Proximity Venue	15th Most Proximity Venue	Grocery Store
9	ERMINESKIN	3	Grocery Store	Coffee Shop	Yoga Studio	Dry Cleaner	Italian Restaurant	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	0.5
41	HERITAGE VALLEY TOWN CENTRE AREA	3	Grocery Store	Yoga Studio	Wine Shop	Italian Restaurant	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	Electronics Store	1.0

Cluster5 (green): The cluster5 is the most significant cluster with 26 neighbourhoods. The most commonly seen venues in this cluster are Gym, Gift Shop, Fast food, Restaurants. The grocery store is not among the most common venues, making this cluster of neighbourhoods an ideal destination to set up a grocery store.

	Neighbourhood	Cluster_Label	1st Most Proximity Venue	2nd Most Proximity Venue	3rd Most Proximity Venue	4th Most Proximity Venue	5th Most Proximity Venue	6th Most Proximity Venue	7th Most Proximity Venue	8th Most Proximity Venue	9th Most Proximity Venue	10th Most Proximity Venue	11th Most Proximity Venue	12th Most Proximity Venue	13th Most Proximity Venue	14th Most Proximity Venue	15th Most Proximity Venue	Grocery Store
1	BEARSPAW	4	Lake	Outdoors & Recreation	Japanese Restaurant	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	Electronics Store	0.0
3	BLACKMUD CREEK	4	Hotel	Gift Shop	Coffee Shop	Chinese Restaurant	Yoga Studio	Grocery Store	Fast Food Restaurant	Garden Center	Gas Station	Gym	Dry Cleaner	Home Service	Ice Cream Shop	Indian Restaurant	Insurance Office	0.0
5	BRANDER GARDENS	4	Café	School	Home Service	Chinese Restaurant	Grocery Store	Fast Food Restaurant	Garden Center	Gas Station	Gift Shop	Yoga Studio	Dry Cleaner	Gym	Hotel	Ice Cream Shop	Indian Restaurant	0.0
6	BROOKSIDE	4	Pub	Park	Soccer Field	Gift Shop	Dry Cleaner	Electronics Store	Fast Food Restaurant	Garden Center	Gas Station	Yoga Studio	Dog Run	Gym	Home Service	Hotel	Ice Cream Shop	0.0
7	BULYEA HEIGHTS	4	Pizza Place	Baby Store	Park	Yoga Studio	Grocery Store	Fast Food Restaurant	Garden Center	Gas Station	Gift Shop	Gym	Dry Cleaner	Home Service	Hotel	Ice Cream Shop	Indian Restaurant	0.0
8	CHAPPELLE AREA	4	Insurance Office	Yoga Studio	Wine Shop	Italian Restaurant	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	Electronics Store	0.0
10	FALCONER HEIGHTS	4	Pizza Place	Coffee Shop	Smoothie Shop	Fast Food Restaurant	Sandwich Place	Grocery Store	Indian Restaurant	Asian Restaurant	Home Service	Gym	Hotel	Dry Cleaner	Gift Shop	Gas Station	Ice Cream Shop	0.1
12	HODGSON	4	Restaurant	Discount Store	Ice Cream Shop	Japanese Restaurant	Wine Shop	Bridal Shop	Bus Station	Italian Restaurant	Insurance Office	Indian Restaurant	Asian Restaurant	Hotel	Home Service	Gym	Grocery Store	0.0
13	KEHEEWIN	4	Pizza Place	Tea Room	Sandwich Place	Fast Food Restaurant	Yoga Studio	Gift Shop	Electronics Store	Garden Center	Gas Station	Grocery Store	Dog Run	Gym	Home Service	Hotel	Ice Cream Shop	0.0
14	LEGER	4	Gym	Salon / Barbershop	Ice Cream Shop	Bus Station	Coffee Shop	Gift Shop	Fast Food Restaurant	Garden Center	Gas Station	Yoga Studio	Grocery Store	Dry Cleaner	Home Service	Hotel	Indian Restaurant	0.0
17	RUTHERFORD	4	Paper / Office Supplies Store	Playground	Shopping Mall	Yoga Studio	Electronics Store	Fast Food Restaurant	Garden Center	Gas Station	Gift Shop	Dog Run	Grocery Store	Gym	Home Service	Hotel	Ice Cream Shop	0.0
18	SKYRATTLER	4	Italian Restaurant	Pizza Place	Shoe Store	Pub	Gas Station	Dry Cleaner	Electronics Store	Fast Food Restaurant	Garden Center	Yoga Studio	Grocery Store	Gym	Home Service	Hotel	Ice Cream Shop	0.0
19	SOUTH TERWILLEGAR	4	Liquor Store	Paper / Office Supplies Store	Yoga Studio	Dry Cleaner	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	0.0
20	STEINHAUER	4	Cosmetics Shop	Pizza Place	Park	Yoga Studio	Gift Shop	Fast Food Restaurant	Garden Center	Gas Station	Grocery Store	Dry Cleaner	Gym	Home Service	Hotel	Ice Cream Shop	Indian Restaurant	0.0
22	TERWILLEGAR TOWNE	4	Gym	Speakeasy	Park	Miscellaneous Shop	Playground	Hotel	Ice Cream Shop	Home Service	Dry Cleaner	Grocery Store	Gift Shop	Gas Station	Indian Restaurant	Garden Center	Fast Food Restaurant	0.0
24	WHITEMUD CREEK RAVINE NORTH	4	Roof Deck	Yoga Studio	Dog Run	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	Electronics Store	0.0
26	BLUE QUILL ESTATES	4	Trail	Bridal Shop	Business Service	Nature Preserve	Yoga Studio	Gym	Fast Food Restaurant	Garden Center	Gas Station	Gift Shop	Grocery Store	Home Service	Dry Cleaner	Hotel	Ice Cream Shop	0.0
28	CASHMAN	4	Rental Car Location	Motorcycle Shop	Hotel	Yoga Studio	Insurance Office	Indian Restaurant	Ice Cream Shop	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	Electronics Store	0.0
29	CAVANAGH	4	Restaurant	Japanese Restaurant	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	Electronics Store	Dry Cleaner	0.0
30	DESROCHERS AREA	4	Home Service	Yoga Studio	Wine Shop	Italian Restaurant	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	Electronics Store	0.0
31	HENDERSON ESTATES	4	Park	Performing Arts Venue	Yoga Studio	Dry Cleaner	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	0.0
33	MAGRATH HEIGHTS	4	Fast Food Restaurant	Yoga Studio	Wine Shop	Italian Restaurant	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Electronics Store	0.0
39	RIVER VALLEY TERWILLEGAR	4	Dog Run	Wine Shop	Italian Restaurant	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	Electronics Store	0.0
42	PAISLEY	4	Dry Cleaner	Yoga Studio	Wine Shop	Italian Restaurant	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	0.0
43	GRAYDON HILL	4	Garden Center	Yoga Studio	Wine Shop	Italian Restaurant	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Fast Food Restaurant	Electronics Store	0.0
45	HAYS RIDGE AREA	4	Moving Target	Yoga Studio	Japanese Restaurant	Insurance Office	Indian Restaurant	Ice Cream Shop	Hotel	Home Service	Gym	Grocery Store	Gift Shop	Gas Station	Garden Center	Fast Food Restaurant	Electronics Store	0.0

5. Discussion

The order of the average of Grocery stores in each cluster and entire southwest Edmonton neighbourhoods are as follows:

Clusters	# of neighbourhoods	Average grocery stores
1	1	0.0
2	8	0.0
3	10	0.0
4	2	0.75
5	26	0.1

The business problem's objective was to help stakeholders identify one of the safest boroughs in Edmonton and an appropriate neighborhood within the borough (region) to set up a commercial establishment, especially a grocery store. This has been achieved by first using Edmonton crime data to identify a safe borough with many neighbourhoods for any business to be viable. After selecting the borough, we chose the right neighbourhood where grocery shops were not among the venues close to each other. This is unsupervised data and the dataset having 65 features; many different approaches can be adopted, and we chose to go with K-means to achieve better results. Using K means we grouped the neighbourhood into clusters to assist the stakeholders by providing relevant data about a given neighbourhood's venues and safety.

6. Conclusion

In modern technology, with the increasing population in the Greater Edmonton Area and people looking to open new commercial establishments, having safe and less competitive neighbourhood recommendations based on location, venue combined with crime data is valuable nowadays. This project acts as an initial step to tackle complex real-life problems using data science and serve as an impressive tool for entrepreneurs and business people to provide detailed insight into the community. Furthermore, it can be used as a better tool if combined with neighbourhood population data.