THE BATTLE OF EDMONTON NEIGHBOURHOODS - PRESENTATION

BY AARTHI VELUSAMY, NOVEMBER 2020

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.

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.

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

1. Edmonton Crime data (via Opendata),

2. list of neighbourhoods and boroughs in Edmonton (via Wikipedia),

3. Edmonton Postal Codes (via Wikipedia),

4. the Geographical location of the communities (via **Geocoder package**) and Edmonton Venue data (via **Foursquare**).

5. The Venue data will help find which neighbourhood is best suitable to open a Grocery Store.

Data Cleaning

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.
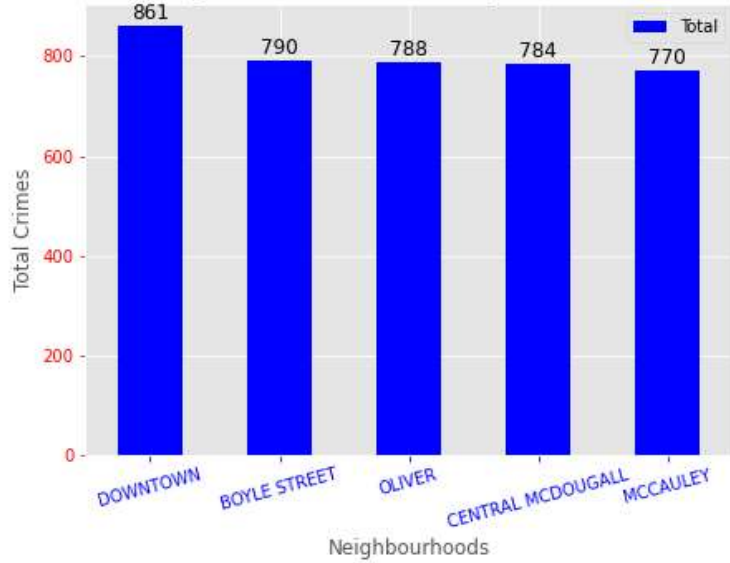4. We are pivoting the table to represent the data in a better understanding format.

Methodology

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.
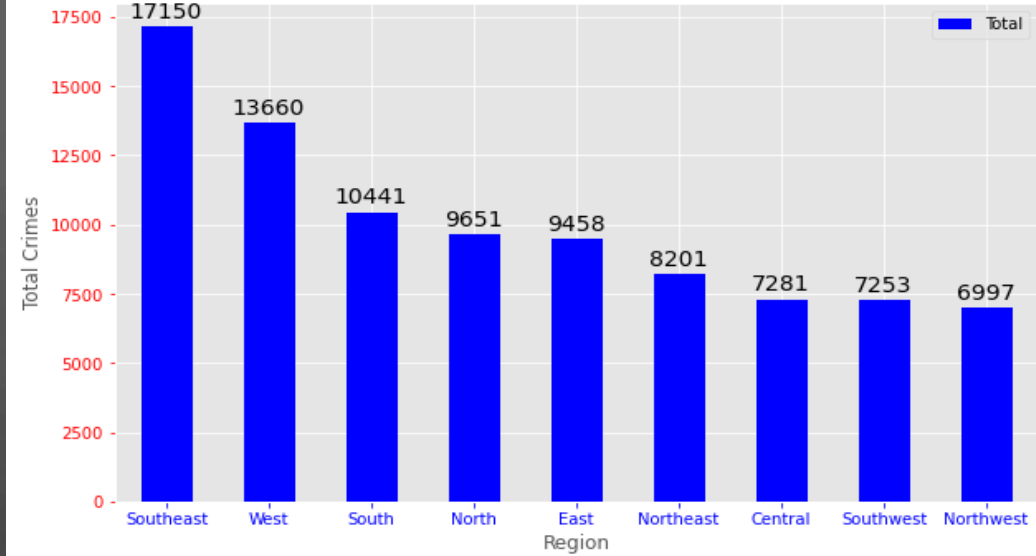
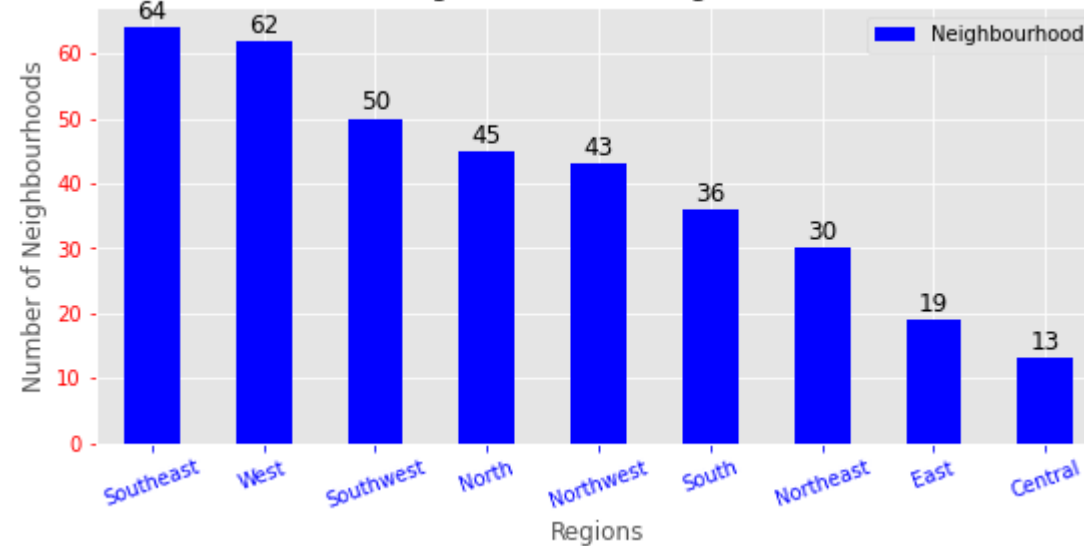|	Assault	Break and Enter	Homicide	Robbery	Sexual Assaults	Theft from Vehicle	Theft Of Vehicle	Theft Over $5000	Total
count	10.000000	10.000000	10.00000	10.000000	10.000000	10.00000	10.000000	10.000000	10.000000
mean	3268.600000	3951.400000	37.80000	1219.400000	928.000000	4673.20000	3346.600000	593.400000	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

Data Visualization

Different Kind of Crimes in least 3 crime Region
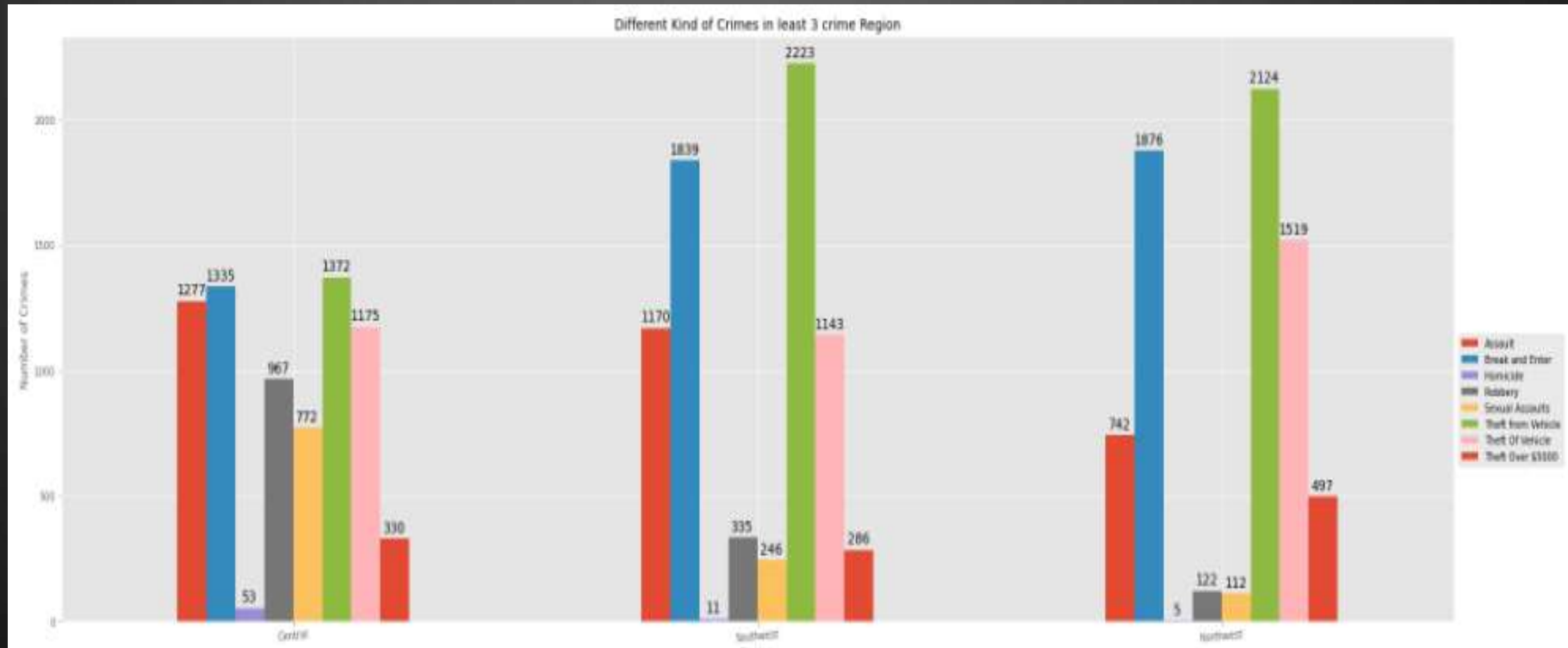
Northwest has the lowest crimes. Next in place are Southwest and Central with minimal increase in crime rate compared to Northwest. Southwest has 50 Neighbourhoods, opening a commercial establishment in this region is an optimal idea.. Southwest has more neighbourhoods than Northwest, and the crime type Break is also low compared to Northwest, which makes **Southwest the ideal destination for opening commercial establishments like a Grocery store.**
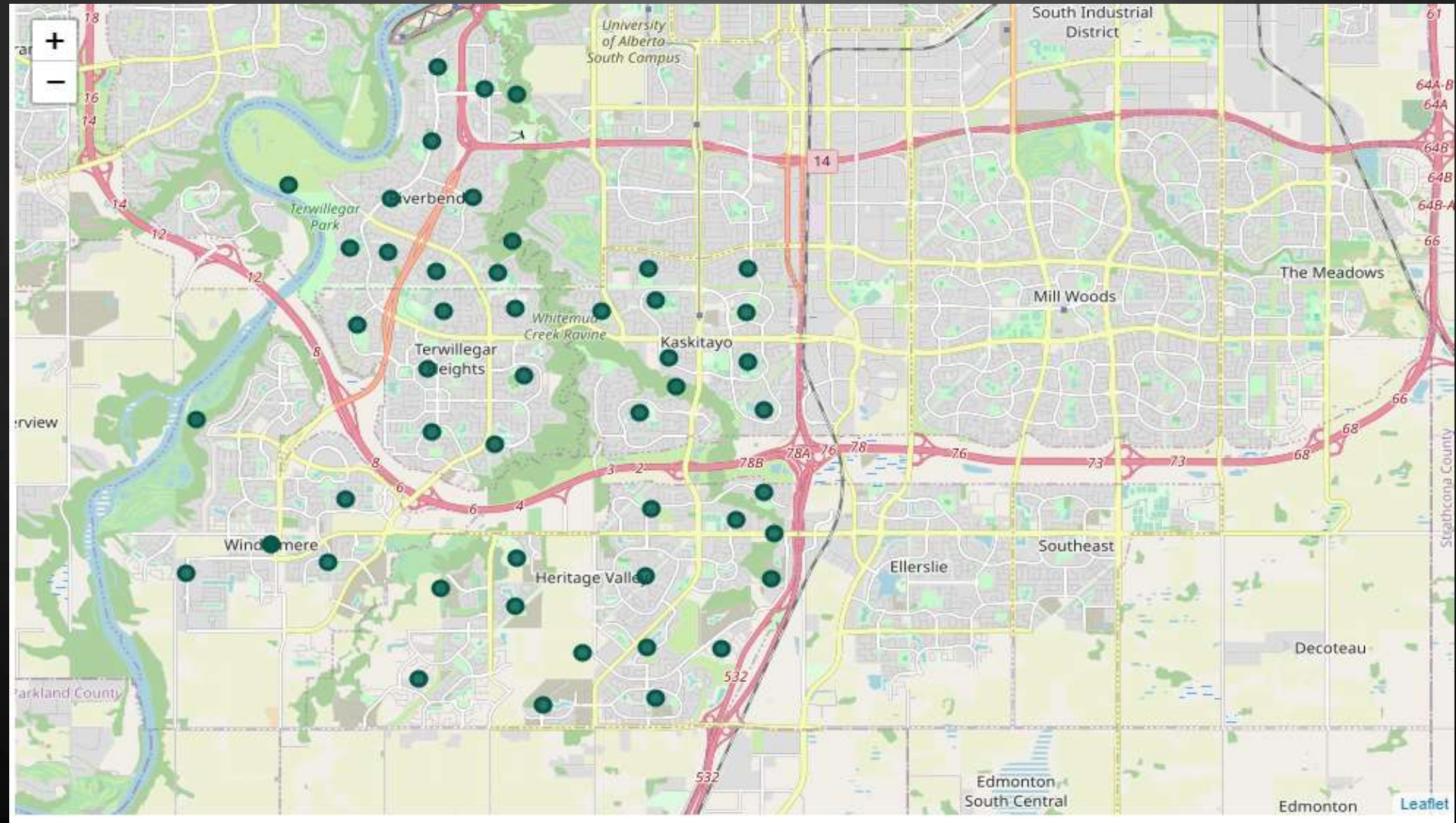
Modelling

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

Mapping the neighbourhoods cluster

After running the K-Means lustering, 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
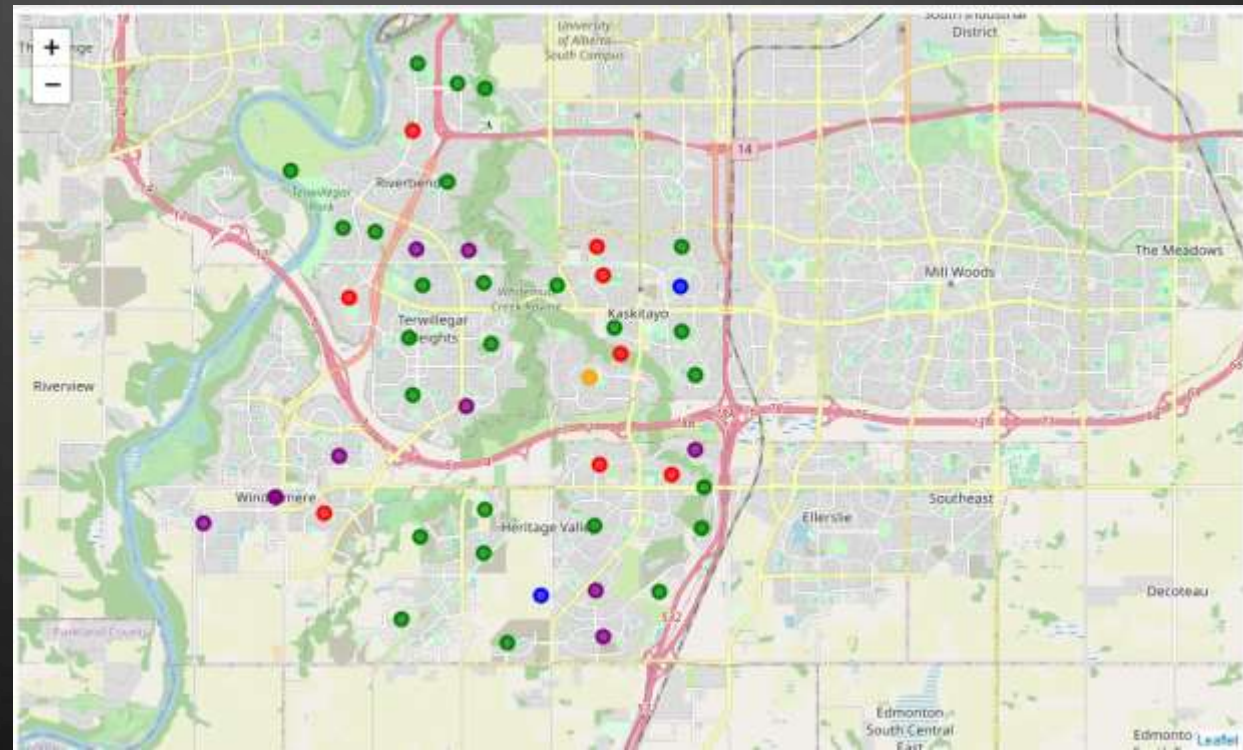
Results

Cluster Analysis

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).
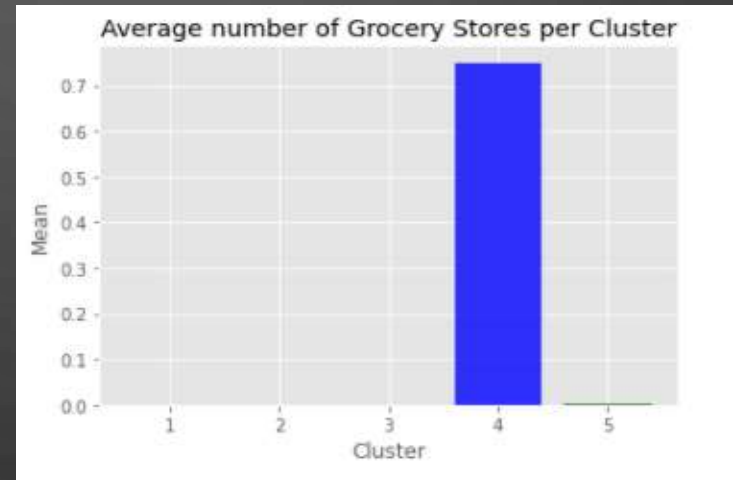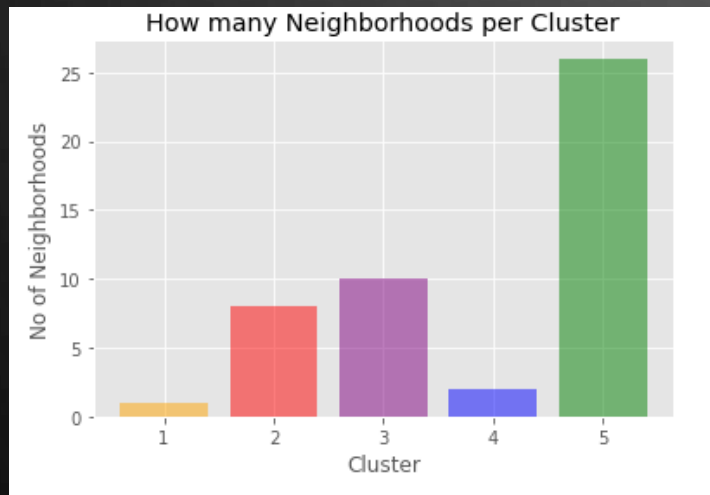
Discussion

The order of the average of Grocery stores in each cluster and entire southwest Edmonton neighbourhoods

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.

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.