

# Identify the best location to open a Grocery Store in Calgary, AB Canada

Data Science IBM Capstone Project

## 1. Description of the Problem

The energy capital, nature, mountains in addition to the fourth highest population density makes the city of Calgary an attractive place for a business. However, for newcomers, immigrants and even people from the Eastern provinces of Canada could be a big challenge to select the best neighborhood to set up the business to have the maximum rate of return specifically in this uncertain and tough economic times. This study aims to answer most of the questions to select the best neighborhood in Calgary to set up specifically a grocery store. The study would use the neighborhood coordinates, population density in each neighborhood and four-square API to get the existing grocery store in each neighborhood. The study is designed in particular for an investor who is interested to start up an Asian grocery store in Calgary and have not been to Calgary before. It is believed that this study along with the economic analysis would help to optimize the location for the investor.

## 2. Data utilized for Calgary City

The Calgary neighborhood data, latitude and longitude and postal code will be scrapped from Wikipedia using beautiful soup library. The population density of each neighborhood will also be extracted from Wikipedia and the two tables to be merged.

Now that we have built a data frame of the postal code of each neighborhood along with the neighborhood name, population density, in order to utilize the Foursquare location data, we need to get the latitude and the longitude coordinates of each neighborhood.

For latitude and longitude of metro stations, venues, Geo coder Python package will be utilized to get the geographical coordinates of a given postal code. In order to extract the latitude and longitude for all the neighborhoods, a while loop will be employed in order to make sure that I get the coordinates for all of Calgary neighborhoods. Machine learning algorithms will be applied including K-Means to cluster the neighborhood. Then simple 2-D plots are generated to see the population density within various neighborhood and ranking is done based on the author knowledge to rank the neighborhood first within each cluster and then on global level to start the business.

Following information to be used in order to make the business decision for the selection of neighborhood:

- Neighborhood locations and population density (Population/area) for each neighborhood
- Top 100 venues of each neighborhood and filter the venue data specifically for Grocery store
- Data wrangling to arrange the data and drop with a minimum threshold of population density to be uneconomic to start a business in any specific neighborhood
- Machine learning algorithm to cluster the neighborhood
- Generate plots for each neighborhood to understand the best neighborhood with each cluster and then rank them globally over the entire City of Calgary

	Name	Quadrant	Sector	Population (2012)	Area (km2)	Population density
0	Abbeydale	NE/SE	Northeast	5,917	1.7	3,480.6
1	Acadia	SE	South	10,705	3.9	2,744.9
2	Albert Park/Radisson Heights	SE	East	6,234	2.5	2,493.6
3	Altadore	SW	Centre	9,116	2.9	3,143.4
4	Applewood Park	SE/NE	East	6,498	1.6	4,061.3

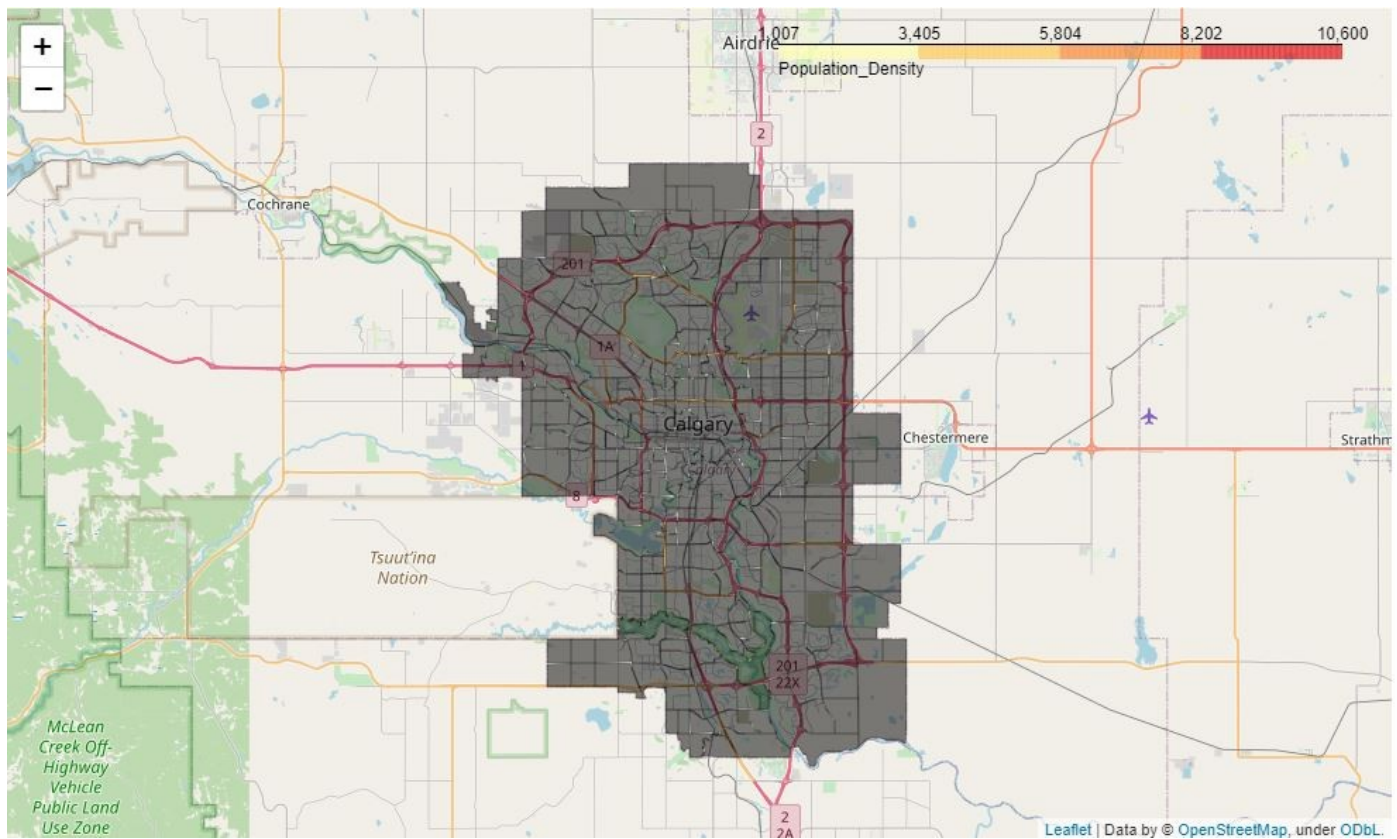
**Table 1** City of Calgary data set for neighborhood and population density

The Quadrant longitude and latitude

	index	Population_2012	Area_km2	Population_density	latitude	longitude
Quadrant						
SW/SE	50.2	7198.3	2.0	4782.3	51.0	-114.1
SE/NE	22.0	6498.0	1.6	4061.3	51.0	-113.9
NE/NW	32.5	5421.0	1.5	3717.8	51.1	-114.1
NE	52.6	8757.6	2.5	3576.1	51.1	-114.0
NW/NE	73.0	9320.8	3.2	2929.0	51.1	-114.1
SW	98.5	4190.2	1.6	2725.2	51.0	-114.1
NE/SE	94.0	5755.5	2.4	2642.6	51.1	-114.0
SE	99.5	6876.2	2.9	2574.6	51.0	-114.0
NW	104.4	6451.9	2.7	2435.4	51.1	-114.1

**Table 2** City of Calgary data set based on Quadrant

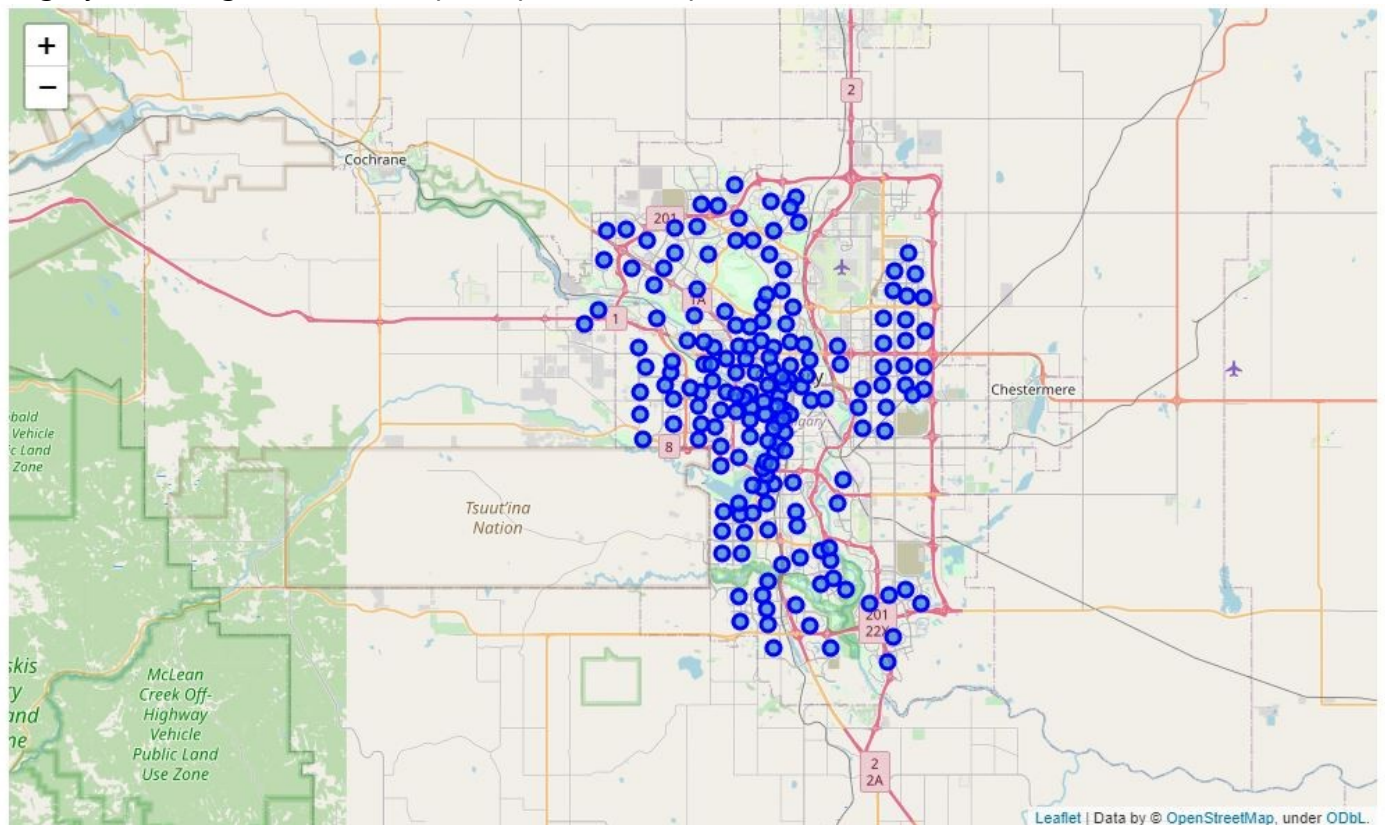
The neighborhood map is shown below



**Figure 1** Map of City of Calgary neighborhoods

Map of City of Calgary neighborhood boundaries

Calgary with neighborhoods superimposed on top



**Figure 2** Map of City of Calgary neighborhood population density



## Methodology

The Neighborhood data is web scrapped from Wikipedia using beautiful soup library. The names of neighborhood were used to extract the latitude and longitude using Geo coder library. The venues were extracted for each neighborhood using Four square API. The neighborhoods were then clustered using K-Mean Algorithm. Other plots are used to rank different neighborhood within each cluster and then compared globally for entire Calgary to determine the best cluster to start the business based on population density.

Then get the top 100 venues that are in Lower Mount Royal, Calgary within a radius of 1,000 meters.

For example, for the first neighborhood, Lower Mount Royal following grocery store data is extracted from Four square API.

	name	categories	address	cc	city	country	distance	formattedAddress	labeledLatLngs	lat	lng	postalCode
0	Kalamata Grocery	Grocery Store	1421 11th St. SW	CA	Calgary	Canada	337	[1421 11th St. SW, Calgary AB T2R 1G7, Canada]	[{"label": "display", "lat": 51.03940743942703...	51.039407	-114.089129	T2R 1G7
1	Buy-Rite Grocery	Convenience Store	NaN	CA	NaN	Canada	562	[Canada]	[{"label": "display", "lat": 51.037731060453351...	51.037731	-114.079287	NaN
2	Lang Grocery	None	833 17ave SW	CA	Calgary	Canada	596	[833 17ave SW, Calgary AB T2T 0A1, Canada]	[{"label": "display", "lat": 51.041166, "lng":...	51.041166	-114.091712	T2T 0A1
3	Happy Together Convenience & Grocery Store	Convenience Store	348 14 Ave sw	CA	Calgary	Canada	1167	[348 14 Ave sw, Calgary AB T2R 0X8, Canada]	[{"label": "display", "lat": 51.03991057, "lng":...	51.039911	-114.071283	T2R 0X8

**Table 3** Grocery Stores for Lower Mount Royal neighborhood in Calgary

## Results

The function is created to extract the all the venues for all the neighborhood in Calgary. The venues were filtered to select the type of 'grocery' and 'super'. The resulting two arrays were then combined. This gives a total of 28 grocery stores in Calgary. However, the search limit of four square is reached. The neighborhoods were then divided based on the number of grocery store into 4 different clusters using K-Means algorithm. Further analysis was done using population density for each cluster to rank different neighborhood within each cluster and then compare the overall for the entire Calgary city.

There were total 1177 venues in city of Calgary. The results were filtered for grocery stores.

For example, for the Lower Mount Royal neighborhood, the following data was obtained which was then filtered by venue category type for grocery store only.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Lower Mount Royal ,Calgary	51.0	-114.1	Steeling Home	51.0	-114.1	Gift Shop
1	Lower Mount Royal ,Calgary	51.0	-114.1	Made By Marcus	51.0	-114.1	Ice Cream Shop
2	Lower Mount Royal ,Calgary	51.0	-114.1	Tubby Dog	51.0	-114.1	Hot Dog Joint
3	Lower Mount Royal ,Calgary	51.0	-114.1	Waves Coffee House	51.0	-114.1	Coffee Shop
4	Lower Mount Royal ,Calgary	51.0	-114.1	The Coup	51.0	-114.1	Vegetarian / Vegan Restaurant

**Table 4** All venue category for Lower Mount Royal neighborhood in Calgary

For Grocery store only, two type of keywords were used using "Grocery" and "Super" to select all the grocery store in Calgary. Following are the results:

index		Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue
0	15	Lower Mount Royal ,Calgary	51.036645	-114.087139	Kalamata Grocery
1	28	Lower Mount Royal ,Calgary	51.036645	-114.087139	Urban Fare
2	50	Lower Mount Royal ,Calgary	51.036645	-114.087139	lp foods
3	109	Bankview ,Calgary	51.033887	-114.099518	M&M Food Market
4	309	Forest Heights ,Calgary	51.047031	-113.970978	Lucky Supermarket
5	383	Sunnyside ,Calgary	51.057870	-114.078740	Safeway Kensington Calgary
6	456	Tuxedo Park ,Calgary	51.073788	-114.061348	Unimarket
7	468	Tuxedo Park ,Calgary	51.073788	-114.061348	Unimarket
8	485	Kingsland ,Calgary	50.986324	-114.077341	Yummy Russia
9	515	McKenzie Towne ,Calgary	50.916499	-113.964353	Sobeys - McKenzie Towne
10	541	Forest Lawn ,Calgary	51.033333	-113.966667	Merv's No Frills
11	597	Mayland Heights ,Calgary	51.060571	-114.011905	MRT Family Foods
12	618	Tuscany ,Calgary	51.125380	-114.244843	Sobeys - Tuscany
13	630	Huntington Hills ,Calgary	51.118724	-114.068288	Real Canadian Superstore
14	663	Country Hills Village ,Calgary	51.157500	-114.061525	Sobeys - Country Hills
15	690	Strathcona Park ,Calgary	51.047868	-114.184479	Sobeys - Strathcona Square
16	749	Arbour Lake ,Calgary	51.136786	-114.202355	Amaranth Whole Foods Market
17	804	Riverbend ,Calgary	50.974229	-114.014999	Sobeys - Riverbend
18	836	Southwood ,Calgary	50.957248	-114.083819	No Frills
19	880	Ramsay ,Calgary	51.037727	-114.041219	Alberta Corner Store
20	914	Lake Bonavista ,Calgary	50.939987	-114.051475	Safeway Bonavista Shopping Plaza
21	965	Bridgeland/Riverside ,Calgary	51.053801	-114.044310	Bridgeland Market
22	977	Varsity ,Calgary	51.090998	-114.155759	Planet Organic Market
23	1024	Willow Park ,Calgary	50.960293	-114.054645	Italian Centre Shop Ltd.
24	1072	Aspen Woods ,Calgary	51.043119	-114.210185	Blush Lane Organic Market
25	1079	Aspen Woods ,Calgary	51.043119	-114.210185	Safeway Aspen
26	1108	Montgomery ,Calgary	51.074802	-114.162474	Safeway Montgomery
27	1114	Manchester ,Calgary	51.006890	-114.068474	Unimarket Latin Grocery
28	1173	Scarboro/Sunalta West ,Calgary	51.041052	-114.114728	Casel Marche
29	88	Mission ,Calgary	51.031758	-114.066720	Safeway
30	290	Cliff Bungalow ,Calgary	51.034436	-114.073833	Safeway
31	548	Rutland Park ,Calgary	51.021847	-114.135943	Safeway Canada
32	1032	Cranston ,Calgary	50.875210	-113.965956	Real Canadian Superstore
33	1059	Roxboro ,Calgary	51.027124	-114.068708	Safeway
34	1080	Aspen Woods ,Calgary	51.043119	-114.210185	Safeway Canada

**Table 5** Filtered data set for the neighborhoods with grocery stores in Calgary

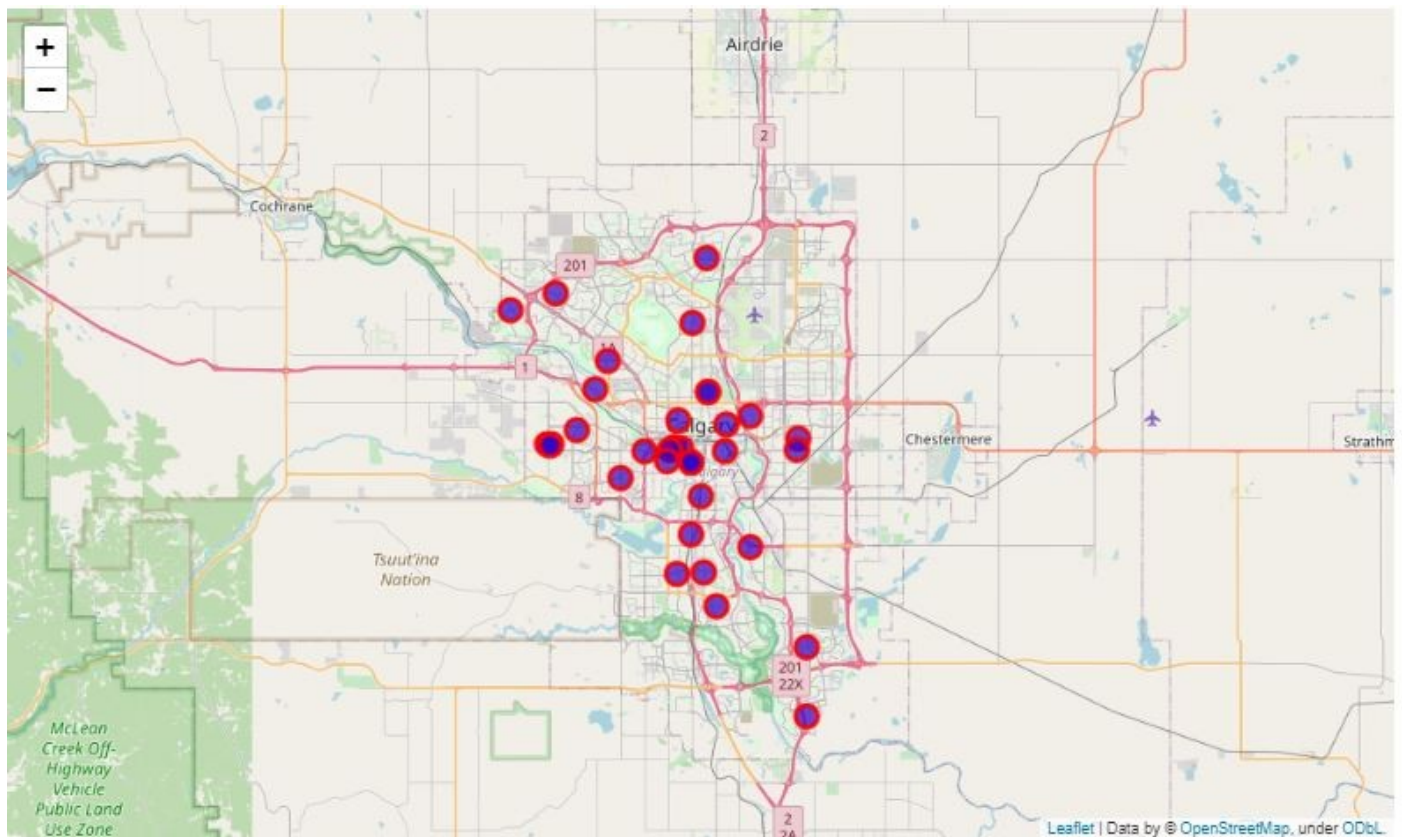
The following is the count of the stores by neighborhood.

	index	Neighborhood	Latitude	Neighborhood Longitude	Venue	
Neighborhood						
Arbour Lake ,Calgary	1		1		1	1
Aspen Woods ,Calgary	3		3		3	3
Bankview ,Calgary	1		1		1	1
Bridgeland/Riverside ,Calgary	1		1		1	1
Cliff Bungalow ,Calgary	1		1		1	1
Country Hills Village ,Calgary	1		1		1	1
Cranston ,Calgary	1		1		1	1
Forest Heights ,Calgary	1		1		1	1
Forest Lawn ,Calgary	1		1		1	1
Huntington Hills ,Calgary	1		1		1	1
Kingsland ,Calgary	1		1		1	1
Lake Bonavista ,Calgary	1		1		1	1
Lower Mount Royal ,Calgary	3		3		3	3
Manchester ,Calgary	1		1		1	1
Mayland Heights ,Calgary	1		1		1	1
McKenzie Towne ,Calgary	1		1		1	1
Mission ,Calgary	1		1		1	1
Montgomery ,Calgary	1		1		1	1
Ramsay ,Calgary	1		1		1	1
Riverbend ,Calgary	1		1		1	1
Roxboro ,Calgary	1		1		1	1
Rutland Park ,Calgary	1		1		1	1
Scarboro/Sunalta West ,Calgary	1		1		1	1
Southwood ,Calgary	1		1		1	1
Strathcona Park ,Calgary	1		1		1	1
Sunnyside ,Calgary	1		1		1	1
Tuscany ,Calgary	1		1		1	1
Tuxedo Park ,Calgary	2		2		2	2
Varsity ,Calgary	1		1		1	1
Willow Park ,Calgary	1		1		1	1

**Table 6** Neighborhood count of grocery stores in Calgary

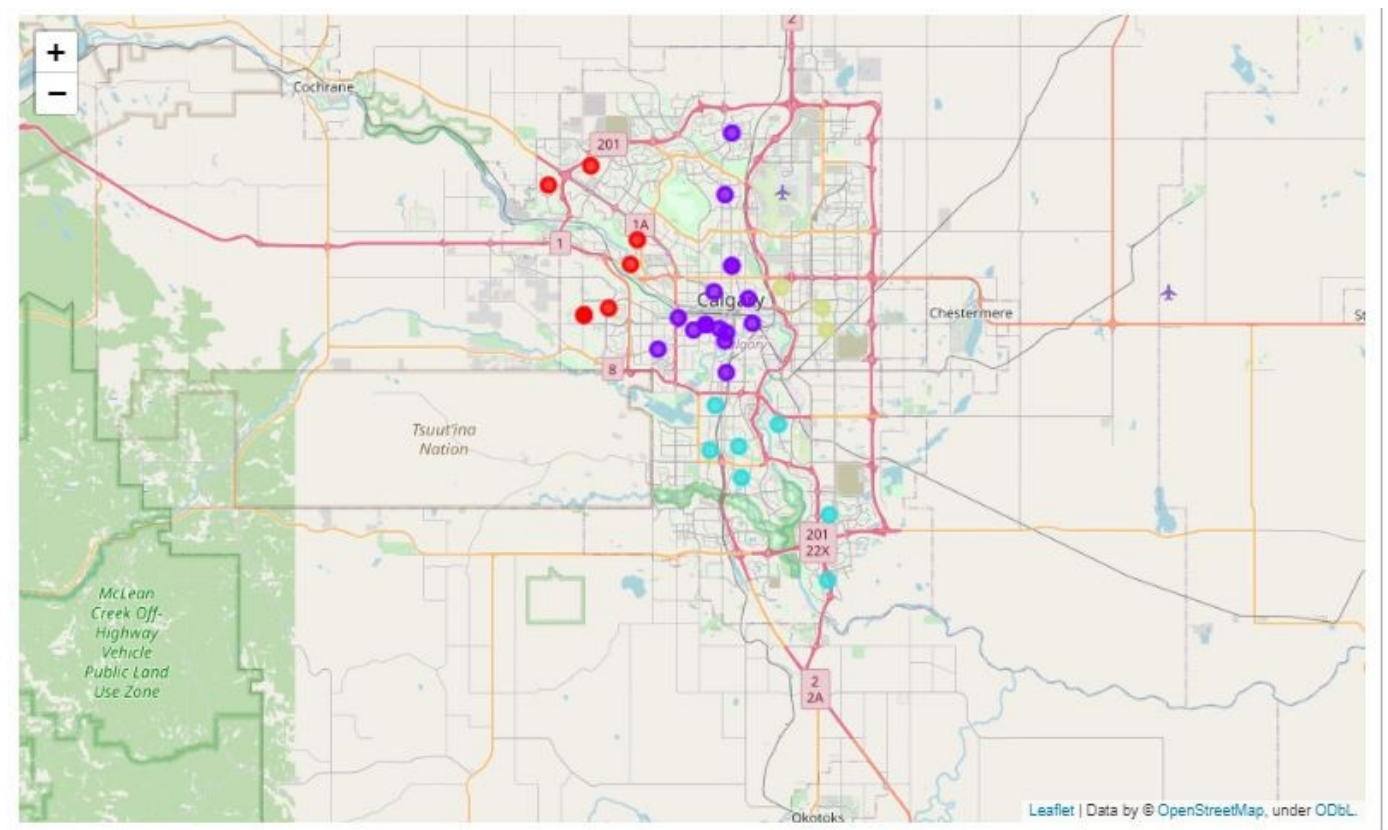


The map shows the neighborhood with grocery stores.



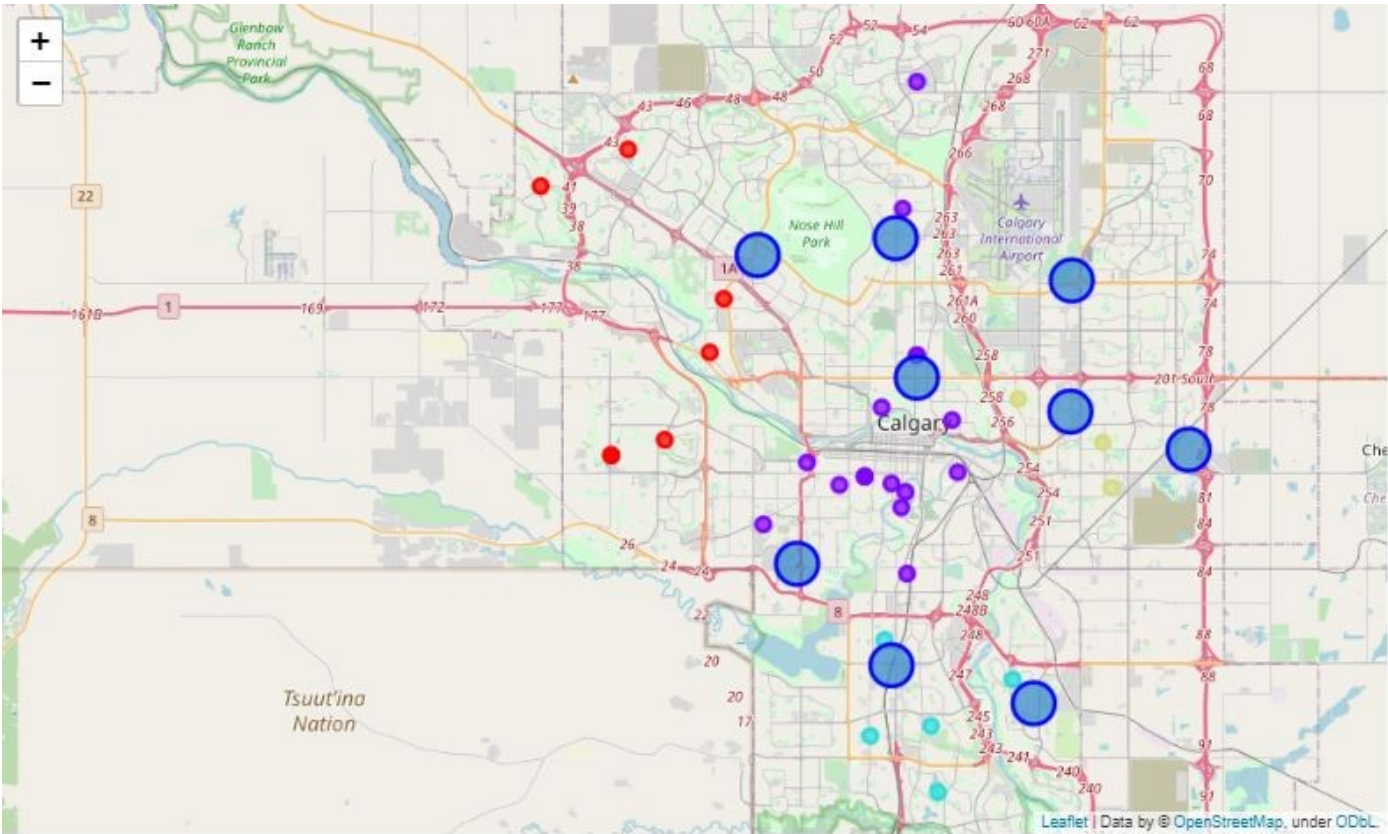
**Figure 3** Neighborhoods with Grocery stores in Calgary

Apply K-Means algorithm to cluster neighborhood. The maximum number of clusters selected is 4.



**Figure 4** Cluster of Neighborhoods in Calgary using K-Means Algorithm

The map with the small circles shows the K-Means cluster for grocery store in Calgary. The larger circle shows the population density of each quadrant. This is done to relate the population density and nearby existing grocery store.



**Figure 5** Cluster of Neighborhoods in Calgary using K-Means Algorithm and population density

The number of neighborhoods in each cluster are shown in the table below:

value_counts	
SW	64
NW	44
SE	32
NE	21
SW/SE	9
NW/NE	4
NE/SE	2
NE/NW	2
SE/NE	1

**Table 7** Count of neighborhoods in each quadrant

The best five locations for business in each quadrant are as follows:



	Neighborhood	Quadrant	Sector	Population_density	Postal_code
0	Lower Mount Royal ,Calgary	SW	Centre	10600.0	T2T 0K5
1	Bankview ,Calgary	SW	Centre	7458.6	T2T 4M8
2	Downtown West End ,Calgary	SW	Centre	5805.0	T2P 2B3
3	Cliff Bungalow ,Calgary	SW	Centre	4840.0	T2S 2B4
4	Evergreen ,Calgary	SW	South	4371.7	T2Y 5E2

**Table 8** Top five locations based on Population Density in SW neighborhood in Calgary

	Neighborhood	Quadrant	Sector	Population_density	Postal_code
0	Citadel ,Calgary	NW	Northwest	3776.7	T3G 4B8
1	Sunnyside ,Calgary	NW	Centre	3704.0	T2N 0E4
2	University Heights ,Calgary	NW	Northwest	3660.0	T2N 4B9
3	MacEwan Glen ,Calgary	NW	North	3642.9	T3K 3J9
4	Panorama Hills ,Calgary	NW	North	3531.3	T3K 5P5

**Table 9** Top five locations based on Population Density in NW neighborhood in Calgary

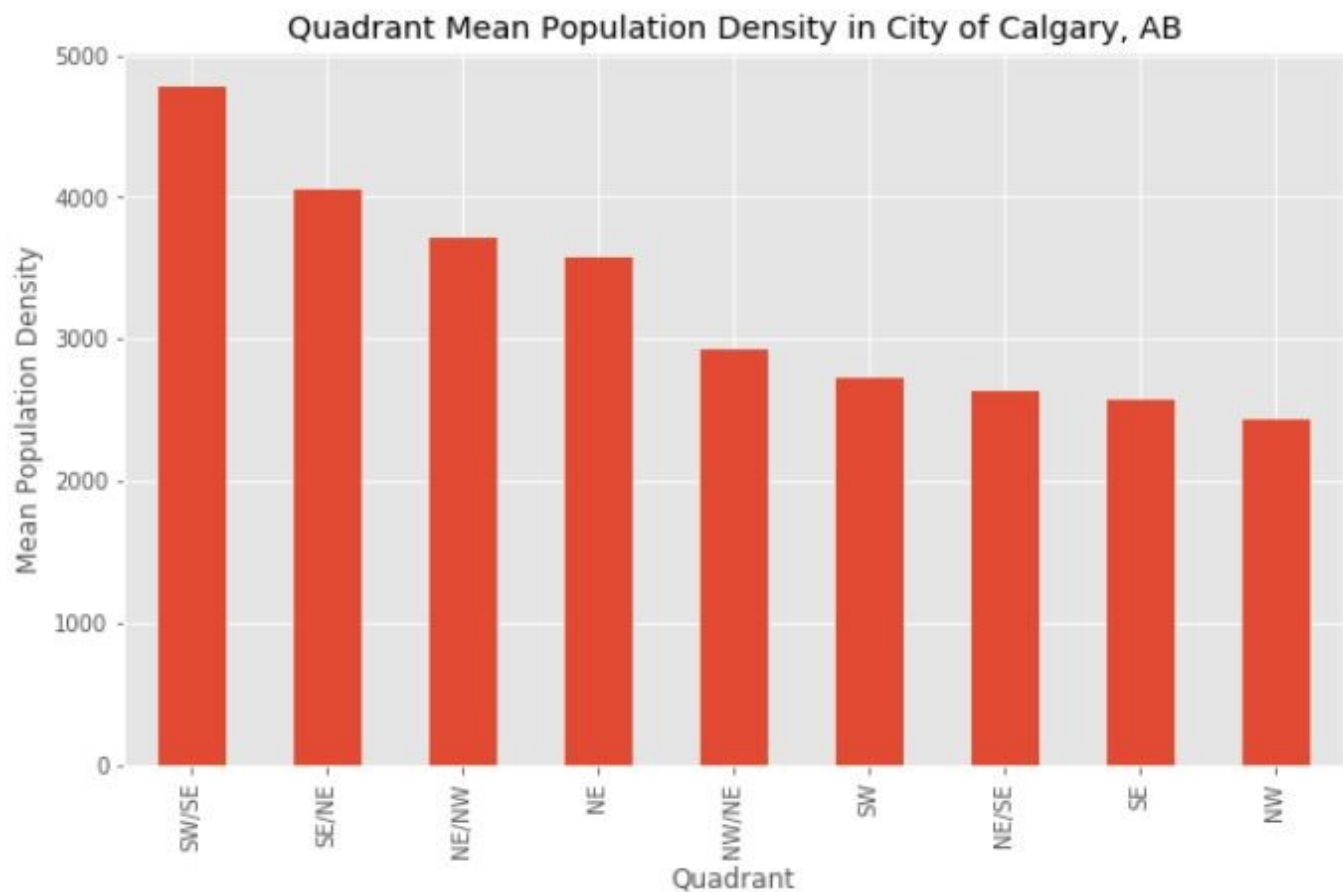
	Neighborhood	Quadrant	Sector	Population_density	Postal_code
0	Downtown East Village ,Calgary	SE	Centre	5564.0	T2G 1K7
1	Erin Woods ,Calgary	SE	East	4313.1	T2B 0L1
2	Penbrooke Meadows ,Calgary	SE	East	4273.5	T2A 6M7
3	Forest Heights ,Calgary	SE	East	4141.3	T2A 1N3
4	McKenzie Towne ,Calgary	SE	Southeast	3354.6	T2Z 3L4

**Table 10** Top five locations based on Population Density in SE neighborhood in Calgary

	Neighborhood	Quadrant	Sector	Population_density	Postal_code
0	Taradale ,Calgary	NE	Northeast	5807.2	T3J 3E6
1	Castleridge ,Calgary	NE	Northeast	5080.0	T3J 1P6
2	Martindale ,Calgary	NE	Northeast	5064.4	T3J 3G4
3	Falconridge ,Calgary	NE	Northeast	4718.6	T3J 1B1
4	Whitehorn ,Calgary	NE	Northeast	4558.5	T1Y 4S8

**Table 11** Top five locations based on Population Density in NE neighborhood in Calgary

The results for all the quadrants based on population density in Calgary are shown in the following figure:



**Figure 6** Quadrant Mean population density for City of Calgary

## Conclusion

Based on this analysis, the South West and South East has the highest potential to start up the grocery store with the highest population density and only three grocery stores. This study shows that best location to start up the grocery store for each quadrant in Calgary are as follows: **Lower Mount Royal** in the *South West*, **Citadel** and **Sunny side** in *North West*, **Downtown East Village** in *South East* and **Taradale** in *North East*.

Future studies to consider the economic analysis and evaluate the best location based on expected Internal rate of return. The availability of the data could be a challenge but based on household in the area, reasonable estimates can be made. The price of real estate, number of business opened in the past five years could be other attributes to select the best location of business.

In addition to the above, Machine learning algorithms to be applied and split the exiting business into training and testing data set and validate with on the testing data set in order to see the accuracy of the model. The model to be further tested with similar demographics globally to further improve the reliability of the model.

## References

- [https://en.wikipedia.org/wiki/List\\_of\\_neighbourhoods\\_in\\_Calgary](https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Calgary)
- <https://data.calgary.ca/Base-Maps/Community-Boundaries/ab7m-fwn6>
- IBM Course era- K-Means