

# Capstone Report

## Battle of Calgary Neighborhoods

### 1. Introduction: problem and background

Newcomers to a city often need some time and resources to get familiar with the neighborhoods and surroundings. For incoming residents of the city, their first thing might be to figure out a place to live, where it is safe and close to necessities. For newcomers who is looking to start their own business, they need to know their type of business, the regulations, and the best locations to attract the most customers.

Calgary is a western Canadian city in the province of Alberta. It is the third-largest municipality in Canada. This project aims to briefly explore the city, such as what the main neighborhoods are, what kind of venues it has, which area is safer to live in. In Section 1 of this project, I implemented machine learning to cluster the venues. In Section 2, I visualized crime data on top of the city map to identify the safer neighborhoods.

This project would benefit those who are interested in knowing some basic venue and crime information about the city of Calgary. The user of this report will be able to get certain level of idea of the neighborhoods in the city, the type of venues in each neighborhood, as well as the crime data in each sector. This should benefit the (potential) newcomers in terms of where to buy or live, or people who is curious about the city.

### 2. Data

In Section 1, I explored venue information. I found the provincial neighborhood data based on postal code on Wikipedia. It contains five columns: postal code, borough, neighborhood, latitude and longitude. I downloaded the table as a csv file and imported it into Python as a Panda dataframe (Table 1). I further restricted the dataframe to contain only the Calgary borough (Table 2), there are 35 neighborhoods in Calgary. I also changed the data type of latitude and longitude from object to float64.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	T1A	Medicine Hat	Central Medicine Hat	50.03646	-110.67925
1	T2A	Calgary	Penbrooke Meadows, Marlborough	51.04968	-113.96432
2	T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158
3	T4A	Airdrie	East Airdrie	51.27245	-113.98698
4	T5A	Edmonton	West Clareview, East Londonderry	53.5899	-113.4413

Table 1. Sample boroughs in Alberta, Canada.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
1	T2A	Calgary	Penbrooke Meadows, Marlborough	51.04968	-113.96432
2	T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158
10	T2B	Calgary	Forest Lawn, Dover, Erin Woods	51.0318	-113.9786
11	T3B	Calgary	Montgomery, Bowness, Silver Springs, Greenwood	51.0809	-114.1616
19	T2C	Calgary	Lynnwood Ridge, Ogden, Foothills Industrial, G...	50.9878	-114.0001

Table 2: Sample neighborhoods in Calgary, Alberta, Canada.

I used the geopy library to get the latitude and longitude values of Calgary.

I used Foursquare API to acquire nearby venue information in each neighborhood in a radius of 700 meters.

In [Section 2](#), I explore crime data. I obtained official crime map(geojson), crime statistics in 2020(csv) and city sector map (geojson) from <https://data.calgary.ca/>. The column descriptions of the crime statistics are shown below (Table 3).

Column Name	Description	Type
Sector	Calgary Planning Sector the community is located in	Plain Text
Community Name	The Calgary community name where the incident occurred	Plain Text
Category	The category of crime incident	Plain Text
Crime Count	The number of crimes	Number
Resident Count	The community resident count in the Civic Census of that year	Number
Date	The year and month the incidents occurred	Date & Time
Year	The year the incidents occurred	Plain Text
Month	The month the incidents occurred	Plain Text
ID	Unique id	Plain Text
Community Center Point	Center point of the community	Location

Table 3. The columns in the crime dataframe.

There are more than 40000 crimes recorded (Table 4). To reduce the processing time, I limited the data to contain the first 1000 rows in the dataframe. Note that one row can have more than one crime count.

	Sector	Community Name	Category	Crime Count	Resident Count	Date	Year	Month	long	lat	ID	Community Center Point
0	NORTHWEST	02E	Assault (Non-domestic)	6	0.0	05/01/2020 12:00:00 AM	2020	MAY	-114.199362	51.161431	2020-MAY-02E-Assault (Non-domestic)	POINT (-114.199362020526 51.161430618799)
1	WEST	ASPEN WOODS	Street Robbery	1	9446.0	05/01/2020 12:00:00 AM	2020	MAY	-114.207876	51.045131	2020-MAY-ASPEN WOODS-Street Robbery	POINT (-114.207876270297 51.045130968851)
2	CENTRE	CAPITOL HILL	Street Robbery	1	4744.0	07/01/2020 12:00:00 AM	2020	JUL	-114.094652	51.072662	2020-JUL-CAPITOL HILL-Street Robbery	POINT (-114.094651621635 51.072661693566)
3	CENTRE	HIGHWOOD	Assault (Non-domestic)	1	2258.0	07/01/2020 12:00:00 AM	2020	JUL	-114.079901	51.092356	2020-JUL-HIGHWOOD-Assault (Non-domestic)	POINT (-114.07990104744 51.092355774905)
4	NORTH	02K	Theft OF Vehicle	1	0.0	04/01/2020 12:00:00 AM	2020	APR	-114.180588	51.190427	2020-APR-02K-Theft OF Vehicle	POINT (-114.180588398906 51.190427466876)

Table 4. Sample crime data for Calgary in 2020.

### 3. Methodology

In Section 1 where I explore venues in each neighborhood:

As described in the data section above, I created a dataframe which lists the neighborhoods and their location (Table 2). I used the Python folium library to visualize geographic details, as each neighborhood listed in the dataframe is represented as a blue circle on top of Calgary city map (Figure 1).



Figure 1. Representation of neighborhoods on Calgary map.

Next, I utilized the Foursquare API to explore the venues and cluster them. I created a function to list all the venues in each neighborhood in a radius of 700 meters. I then used the function and stored the result in a dataframe named `Calgary_venues` (Table 5). There are 679 venues

found in 165 unique venue categories. Table 6 shows the number of venues found in each neighborhood in a radius of 700 meters.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Penbrooke Meadows, Marlborough	51.04968	-113.96432	TD Canada Trust	51.052399	-113.970643	Bank
1	Penbrooke Meadows, Marlborough	51.04968	-113.96432	7-Eleven	51.045519	-113.957341	Convenience Store
2	Penbrooke Meadows, Marlborough	51.04968	-113.96432	7-Eleven	51.044357	-113.969612	Convenience Store
3	Penbrooke Meadows, Marlborough	51.04968	-113.96432	Gong Kee BBQ Noodle House	51.044477	-113.968857	Noodle House
4	Penbrooke Meadows, Marlborough	51.04968	-113.96432	Pho Rang Dong	51.044656	-113.969383	Vietnamese Restaurant

Table 5. Sample dataframe showing venues related information for all neighborhoods.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
	Neighborhood						
	Braeside, Cedarbrae, Woodbine	11	11	11	11	11	11
	Brentwood, Collingwood, Nose Hill	4	4	4	4	4	4
	Bridgeland, Greenview, Zoo, YYC	44	44	44	44	44	44
	City Centre, Calgary Tower	55	55	55	55	55	55
	Connaught, West Victoria Park	100	100	100	100	100	100
	Cranston, Auburn Bay, Mahogany	4	4	4	4	4	4

Table 6. Sample descriptive statistics on how many venues retrieved for each neighborhood.

I used one hot encoding to create a dataframe which shows list of top 10 venue category for each neighborhood (Table 7). There are 35 rows representing each neighborhood.

	Neighborhood	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	Braeside, Cedarbrae, Woodbine	Pharmacy	Pizza Place	Pub	Gas Station	Gym	Pool	Convenience Store	Coffee Shop	Hockey Arena	Hockey
1	Brentwood, Collingwood, Nose Hill	Building	Hockey Rink	Construction & Landscaping	Sporting Goods Shop	Fast Food Restaurant	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Food Truck	Food & Shop
2	Bridgeland, Greenview, Zoo, YYC	Chinese Restaurant	Coffee Shop	Scenic Lookout	Noodle House	Vietnamese Restaurant	Sandwich Place	Pub	Middle Eastern Restaurant	Bank	Burger
3	City Centre, Calgary Tower	Bar	Pub	Middle Eastern Restaurant	Sushi Restaurant	French Restaurant	Mediterranean Restaurant	Steakhouse	Coffee Shop	Italian Restaurant	Burger
4	Connaught, West Victoria Park	Pub	Restaurant	Italian Restaurant	Coffee Shop	Bar	Café	Vietnamese Restaurant	Sushi Restaurant	Brewery	Mediterranean Restaurant

Table 7. Sample dataframe showing the top 10 venue categories in all neighborhoods.

Next, I used the K-means algorithm (k=10) to cluster the neighborhoods into 10 clusters. I merged top 10 venue table with cluster labels, latitude, longitude for each neighborhood (Table 8).

	Postal Code	Borough	Neighborhood	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
1	T2A	Calgary	Penbrooke Meadows, Marlborough	51.04968	-113.964320	5	Convenience Store	Grocery Store	Vietnamese Restaurant	Noodle House	Bank	Yoga Studio
2	T3A	Calgary	Dalhousie, Edgemont, Hamptons, Hidden Valley	51.12606	-114.143158	5	Asian Restaurant	Café	Convenience Store	Gas Station	Yoga Studio	Philippine Restaurant
10	T2B	Calgary	Forest Lawn, Dover, Erin Woods	51.03180	-113.978600	0	Convenience Store	Sandwich Place	Falafel Restaurant	Smoke Shop	Fried Chicken Joint	Indonesian Restaurant
11	T3B	Calgary	Montgomery, Bowness, Silver Springs, Greenwood	51.08090	-114.161600	0	Clothing Store	Coffee Shop	Restaurant	Dessert Shop	Cosmetics Shop	Park

Table 8. Sample merged dataframe showing neighborhoods, location, cluster # and venues.

I visualized the cluster labels on the map using folium by having different colors denoting each cluster (Figure 2). Each colored circle represents each neighborhood with marker showing neighborhood name and cluster label.

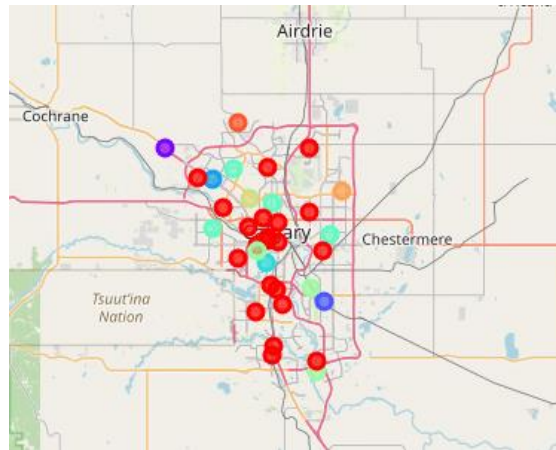


Figure 2. Visualize clustered neighborhoods on map.

Lastly, I viewed the detailed top 10 venues in each cluster (Table 9).

	Borough	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
11	Calgary	0	Restaurant	Clothing Store	Coffee Shop	Park	Accessories Store	Cosmetics Shop	Scenic Lookout	Shopping Mall	Snack Place	Korean Restaurant
19	Calgary	0	Diner	Theater	Convenience Store	Clothing Store	Pizza Place	Home Service	Filipino Restaurant	Fried Chicken Joint	French Restaurant	Forest
20	Calgary	0	Pub	Mexican Restaurant	Indian Restaurant	Vietnamese Restaurant	Pizza Place	Sandwich Place	Fast Food Restaurant	Electronics Store	Record Shop	Coffee Shop
28	Calgary	0	Coffee Shop	Noodle House	Japanese Restaurant	Chinese Restaurant	Sandwich Place	Vietnamese Restaurant	Scenic Lookout	Middle Eastern Restaurant	Pub	Pizza Place
29	Calgary	0	Bowling Alley	Sandwich Place	Coffee Shop	Japanese Restaurant	Sports Bar	Supermarket	Sushi Restaurant	Bakery	Chinese Restaurant	Hobby Shop

Table 9. Sample cluster 0 neighborhoods and venues.

In [Section 2](#) where I explore crime data in each sector:

First, to get an overall idea of the crime types and their occurrence, I calculated total number of crimes for the entire crime dataset on the Category column (Table 10).

	Category	Total
0	Assault (Non-domestic)	13064
1	Break & Enter - Commercial	17879
2	Break & Enter - Dwelling	8468
3	Break & Enter - Other Premises	7668
4	Commercial Robbery	1148
5	Street Robbery	2105
6	Theft FROM Vehicle	47213
7	Theft OF Vehicle	20905
8	Violence 'Other' (Non-domestic)	974
9	Violence Other (Non-domestic)	6467

Table 10. Crime count for each crime category.

Next, to create the choropleth map, I created another dataframe (Table 11) out of the crime data, which summarizes the crime by sector.

	Sector	Total
0	CENTRE	883
1	EAST	205
2	NORTH	199
3	NORTHEAST	295
4	NORTHWEST	192
5	SOUTH	325
6	SOUTHEAST	122
7	WEST	112

Table 11. Crime data by sector.

Next, I visualized the first 1000 rows of crimes on the Calgary map using folium (Figure 3). I also added pop-up text to each marker on the map.

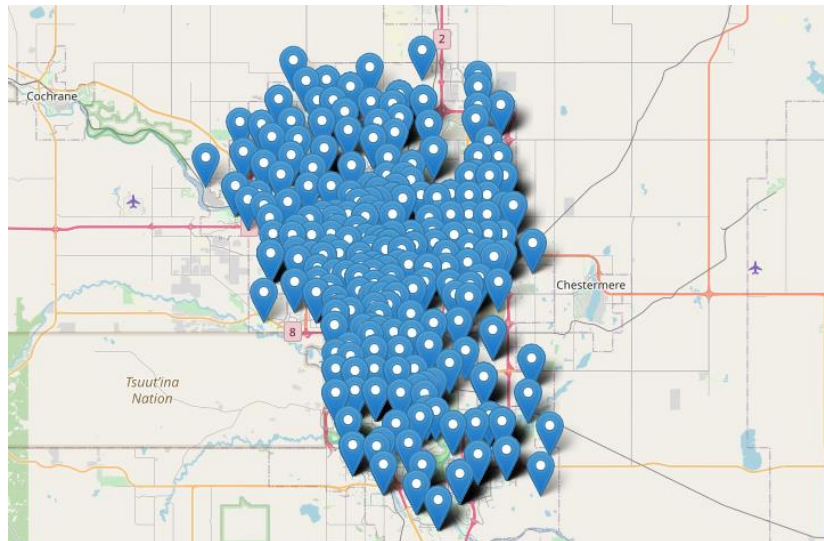


Figure 3. Visualize crimes on map.

As you can see, the map is pretty congested and hard to see clearly. Therefore, I clustered the crimes on the map, each colored circle shows the number of crimes in each nearby area (Figure 4). We can zoom in and out to view the crime count in a smaller area.

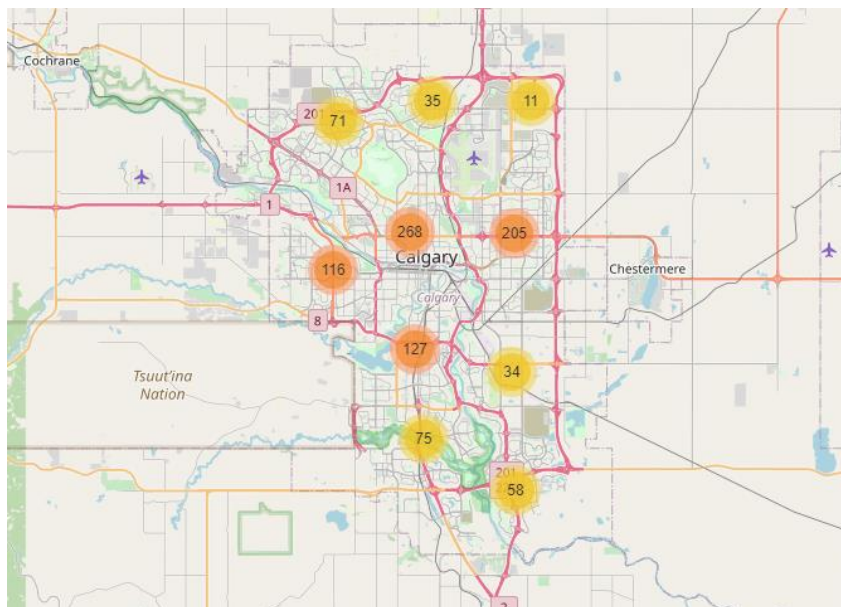


Figure 4. Clustered crime data with zoom in/out functionality.

Last but not least, I used city sector geojson map and the summary crime dataframe (Table 11) to create a choropleth map (Figure 5).



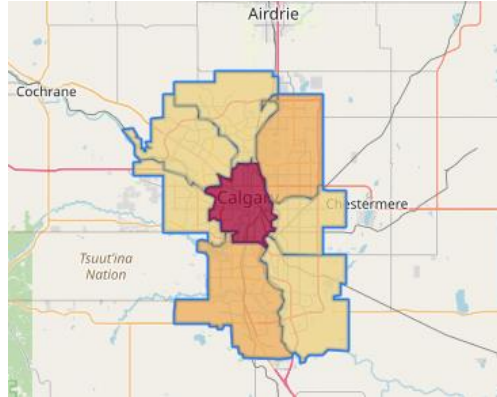


Figure 5. Choropleth map on crime data.

## 4. Results

In Section 1 where I explore venues in each neighborhood:

From Table 6 we can see that certain neighborhoods have more venues retrieved than other neighborhoods in the radius of 700 meters. This could indicate that some neighborhoods are more populated and business prosperous.

When examining Figure 2 and Table 9, it is obvious that certain clusters contains more neighborhoods, which means that some neighborhoods share the similar top 10 venues, while some other neighborhoods have unique venues. Cluster 0 contains the most neighborhoods, while Cluster 1, 3-9 only has one neighborhood. This could be due to city planning and development.

In Section 2 where I explore crime data in each sector:

Table 10 indicates that the most common crime categories are theft from/of vehicle, assault, and break & enter. Table 11 shows that CENTRE sector obviously has the most crime count, followed by SOUTH and NORTHEAST. This may indicate that you need to keep an eye on your vehicle, especially when you park at the CENTRE sector.

Figure 4 shows that most of the crimes in 2020 occurred near the city centre and surrounding areas. The number of crimes substantially decrease while we move to the north or south side of city.

The choropleth map anchors the result of Figure 4, when I map crimes by city sector.

## 5. Discussion

The aim of this project is to give general venue and crime introduction to the city of Calgary. This could be useful to those who are looking to relocate, start a business or choose a place to live. 22 out of 35 of the neighborhoods share the same top 10 venues, which are restaurant,



coffee shops, pubs, etc., while other neighborhoods might be more suitable for living or manufacturing. The new business starters can choose the neighborhood based on the type of business. The new comers who have special preference for a certain type of venue can also choose to locate near those venues.

As for crime information, most crimes occurred near the centre part of the city, which is the downtown area while being the older part of the city. When we move to newer/residential neighborhoods such as northern or southern part of the city, the crime count notably decrease. The safer neighborhoods might be more suitable to reside in.

This project only gives a glimpse of the city. When deciding where to start a business or where to live, it is important to consider multiple factors, and it is also based on personal preference of the surroundings.

## 6. Conclusion

This project gives users a brief geographical view of the city of Calgary. It shows where you can find the different types of venues, and what the safest neighborhood to live in based on crime data. The analysis of this city could be extended by including other relevant factors, such as cost of living, and nearby schools. The methodologies used in the report could be used to analyze any city, given the availability of data.