CALGARY COMMUNITIES

Where would you want to live?

DATA 604 – FALL 2025

L02 – Group 2

Thanusha Balasubramanian, Elise Beaupre

Table of Contents

Introduction & Background

Our project focused on real estate in Calgary communities. This topic is relevant as the residential real estate landscape is undergoing rapid change in the city. According to Statistics Canada's most recent population estimates, Calgary recorded the highest growth rate among Canadian census metropolitan areas at 5.97%, underscoring strong immigration to the city, both from within Canada and internationally (Statistics Canada, 2025). A key driver of this growth has been Calgary's relative affordability in both ownership and rental markets when compared to larger urban centers such as Toronto, Vancouver, and Montreal (Lundy, 2025).

This topic is meaningful for many individuals, as a home is the largest tangible asset they will acquire. Home ownership is often treated as a major life milestone and is closely linked to perceptions of financial stability and long-term wealth building. As Calgary grows, understanding which communities offer the best balance of value, safety, amenities, and future potential becomes increasingly important for prospective buyers. Accordingly, our guiding research question was:

Which community in Calgary is the most ideal to buy a house in?

To investigate which communities are "the most ideal", we defined the following sub areas of focus:

1. Property Value - (addressed by Elise and Thanusha individually)
    i. Which communities are most and least affordable in 2025?
    ii. What are the values of the quartiles for median assessed value by community?
    iii. What is the actual range of values in each Assessment Class?
    iv. Can we classify communities into 'Low', 'Medium', and 'High' categories based on median assessed value?
    v. Which communities have the highest and lowest median assessed values in each assessment class?
2. Amenities (Addressed by Thanusha)
    i. What types of community services/amenities are there?
    ii. Which communities have the most services?
    iii. How many communities have zero listed community services?
    iv. How many communities have at least 2 listed community services?
3. Crime Statistics (Addressed by Elise)
    i. What are the top 5 and bottom 5 communities in terms of the total number of crimes in 2024?
    ii. Which communities have the least stable crime patterns?
    iii. Which communities have shown the greatest increase or decrease in crime from 2018 to 2024?
    iv. Which communities have the highest rates of violent crimes in 2024?
4. Combined Analysis Results (Addressed by Elise & Thanusha together)
    i. Which communities have had the largest change in home value from 2024 – 2025
    ii. Are the communities with the most amenities the most expensive?
    iii. At each price level (low, medium, high) which communities have the most and least crime?
    iv. Are communities with the highest crime rates the most affordable?
    v. At each price level (low, medium, high) which communities have the most amenities?

Datasets

We have 5 datasets from the City of Calgary. All 5 datasets from the City of Calgary contain information licensed under the Open Government Licence – City of Calgary which allows public usage of the data (City of Calgary, 2025a). Appendices A to E provides details on each dataset, including which columns are going to be used.

See the table below for information for a summary of the tables (with their sizes) and descriptions.

DATASET	DESCRIPTION	ROWS	COLUMNS	SIZE of CSV	CITATION
Community Points	Breakdown of each community in Calgary, including a geographical point representing the centroid of the community.	312	10	50 KB	City of Calgary, 2025b
Community Services	Number of services and amenities in each community. Amenities include attractions, health clinics, libraries, and community centres.	211	5	22 KB	City of Calgary, 2025c
Community Crime Statistics	Crime counts for each community from 2018-2024	77.1K	5	4161 KB	City of Calgary, 2025d
Current Year Property Assessments (Parcel)	The assessed values of residential, non-residential and farm land properties in Calgary in 2025	589K	21	288,904 KB	City of Calgary, 2025e
Historical Property Assessments (Parcel)	The historical assessed values of residential, non-residential and farm land properties in Calgary (2005-2024)	9.62M	19	N/A – dataset too large to download as csv	City of Calgary, 2025f

Data Cleaning Methodology

API Technology & Data Import

Our project relied on datasets from the City of Calgary Open Data Portal. The portal allows data to be downloaded as CSV files, but it also exposes each dataset through the Socrata Open Data API (SODA), which was the approach used for this project. SODA provides an API endpoint for every dataset and supports query parameters (using Socrata's SoQL language) for filtering, selecting specific columns, aggregating, and limiting the number of records.

Because several of the datasets we need (especially the historical and current year property assessment tables) are very large, pulling the data directly via the API was more efficient than repeatedly downloading static CSVs.

We accessed the data in Python using pandas together with Socrata's official sodapy library. This approach let us do the following:

1. Authenticate using the API Token which was required for large datasets
2. Filter out requests to include only required data
3. Load the results straight into a DataFrame for cleaning and analysis

Python Data Cleaning & Load into Database

Once the data was in Python we performed the datacleaning steps of dropping any unneeded columns (such as the polygon which contained the geographic point information) and normalizing empty values (NA and N/A to be

pulled as null). Lastly, we converted all the data formats from python objects to number and string types so the data could be loaded into SQL. We did not need to rename and standardize columns as the City of Calgary data was homogeneously formatted overall (Refer to Appendix J for code used to clean data and load the data into SQL).

SQL Methodology for Data Filtering

The original list of communities included residual sub-areas, which are undeveloped areas that are part of the city's planning and development efforts, rather than established communities (City of Calgary, 2019). The communities also included parks, and other non-residential land uses such as the Airport. As we are focusing on established communities where someone could buy a home, these non-residential 'communities' should be removed.

We used the Community_Points table from the City of Calgary (City of Calgary, 2025c) to filter the Community Crime, Community Services, and Property Assessments tables by filtering for class_code = 1 in the community points dataset.

Additionally, we filtered the property assessments tables in the following ways in order to only get property assessments for the residential property within a community (some parcels are mixed use and include a portion of residential and commercial property):

- WHERE assessment_class = 'RE'
- AND re_assessed_value IS NOT NULL
- AND land_use_designation NOT LIKE 'DC%'
- AND land_use_designation NOT LIKE 'C%'
- AND land_use_designation NOT LIKE 'S%'
- AND land_use_designation NOT LIKE 'I%'

Data Analysis - Current Year Property Assessments

Query 1: Which communities are most and least affordable in 2025?

This query provides the top and bottom communities in terms of the median assessed value of residential properties. Median was chosen to reduce the skew that can result from outliers, compared to mean, which is less robust to outliers. Non-residential properties, such as industrial and commercial properties were filtered out as they are not the focus of this investigation (See Appendix F, Query 1).

comm_name	Median_assessed_value ($)
RICARDO RANCH	73000.0
LEWISBURG	146750.0
KEYSTONE HILLS	170000.0
HOTCHKISS	194750.0
LINCOLN PARK	222000.0
...	...
EAGLE RIDGE	1985000.0
BRITANNIA	2050000.0

BEL-AIRE	2480000.0
MEDICINE HILL	8020000.0
QUEENS PARK VILLAGE	42860000.0

The top 5 most affordable communities are mostly brand-new, master-planned suburban developments on Calgary's edges. The least affordable communities tend to be affluent, established communities. Queens Park Village appears to be an outlier, as there is only one parcel of residential property in the whole community in this dataset. The extremely large median property assessment value is due to that one property being a series of townhomes, but each townhome is not represented as its own parcel.

Query 2: What are the values of the quartiles for median assessed value by community?

This informs where the cutoffs will be for the "Low", "Medium", and "High" categories of assessment class. The "Low" class will consist of communities in the bottom 25% of median values, "Medium" will consist of the middle 50% of median values, and "High" will consist of the top 25% of median values (see Appendix F, Query 2).

q1	q2	q3
525000.0	640500.0	768500.0

The median value of all community median assessments is $640,500.00. The assessment classes will have the following boundaries:

- "Low" = Communities with median assessed value <= $525,000.00
- "Medium" = Communities with median assessed value > $525,000.00 and <= $768,500.00
- "High" = Communities with median assessed value > $768,500.00

Query 3: What is the actual range of values in each Assessment Class?

quartile	min_value	max_value	n_communities
1	73000.0	525000.0	51
2	525750.0	640500.0	51
3	641000.0	768500.0	51
4	769750.0	42860000.0	51

- The "Low" assessment class has median assessed values that range from $73,000.00 to $525,000.00, and includes 51 communities.
- The "Medium" assessment class has median assessed values that range from $525,750.00 to $768,500.00, and includes 102 communities.
- The "High" assessment class has median assessed values that range from $769,750.00 to $42,860,000.00, and includes 51 communities.

Query 4: Classify communities into ‘Low’, ‘Medium’, and ‘High’ categories based on median assessed value.

This query actually assigns “Low”, “Medium”, and “High” assessment classes to each community based on the previously mentioned parameters (See Appendix F, Query 4).

community	median_assessed_value	assessment_class
RICARDO RANCH	73000.0	Low
LEWISBURG	146750.0	Low
KEYSTONE HILLS	170000.0	Low
HOTCHKISS	194750.0	Low
LINCOLN PARK	222000.0	Low
...	...	...
EAGLE RIDGE	1985000.0	High
BRITANNIA	2050000.0	High
BEL-AIRE	2480000.0	High
MEDICINE HILL	8020000.0	High
QUEENS PARK VILLAGE	42860000.0	High

Query 5: Which communities have the highest and lowest median assessed values in each assessment class?

This query can help potential property buyers at each budget level know the most and least affordable communities in their budget level (refer to Appendix F, Query 5).

assessment_class	lowest_community	lowest_value	highest_community	highest_value
Low	RICARDO RANCH	73000.0	WHITEHORN	525000.0
Medium	BEDDINGTON HEIGHTS	525750.0	CAMBRIAN HEIGHTS	768500.0
High	CAPITOL HILL	769750.0	QUEENS PARK VILLAGE	42860000.0

Data Analysis - Community Services

Query 1: What types of Community Services Are There?

This first query helps identify which types of community services are available and how many there are of each type. The query also filters out community services which do not have a community code listed with them (Refer to Appendix G, Query 1)

type	service_count
Community Centre	107
Attraction	38
Library	21
Court	9
PHS Clinic	8
Social Dev Ctr	8
Visitor Info	5
Hospital	5

From the results, the most common type of community services are community centres, followed by attractions.

Query 2: Which Communities Have the Most Services?

This query incorporates a left join, where the table on the left is the community points table and the table on the right is the community services table and we filter for only communities in the community points table which have a class code of 1 (residential community). A left join was used, as we want to include communities that have no community services (ie. are not listed in the community services list) but are considered residential communities in the community points dataset.

This is performed to ensure we only look at amenities in residential communities. There were amenities listed that are in non-residential communities (Refer to Appendix G, Query 2). See below for the table with the results of the query.

comm_code	community_name	service_count	service_rank
DNC	DOWNTOWN COMMERCIAL CORE	24	1
BLN	BELTLINE	6	2
FLN	FOREST LAWN	5	3
UOC	UNIVERSITY OF CALGARY	4	4
HUN	HUNTINGTON HILLS	4	4

The downtown commercial core, and beltline are listed as having the most amenities. The surprising result is that the University of Calgary is listed despite having filtered Community Points for residential communities. This may be due to the limitations in classification in the community points table (1 = Residential, 2 = Industrial, 3 = Major Park, 4 = Residual Sub Area) where although the University of Calgary is not a residential neighborhood, it is the closest classification of the ones listed.

Query 3: Which communities have zero amenities listed?

This query finds all the communities which have zero amenities listed. It builds on the same logic as the previous query, but instead we wrap the query in a select all and filter for service_count of zero (Appendix G, Query 3). There are 100 communities listed as having zero amenities.

comm_code	community_name	service_count	service_rank
EVN	EVANSTON	0	122
BLM	BELMONT	0	122
ROY	ROYAL OAK	0	122
CUR	CURRIE BARRACKS	0	122
HKS	HOTCHKISS	0	122
...	...	...	...
COA	COACH HILL	0	122
SAD	SADDLE RIDGE	0	122
MAC	MACEWAN GLEN	0	122
LIV	LIVINGSTON	0	122
HUX	HUXLEY	0	122

100 rows x 4 columns

This result is not surprising given the nature of the data we are working with. As our dataset does not include schools, or private amenities the data is limited. Additionally, a community service may be located in one community but may be easily accessible to neighboring communities. Our analysis only matches on a community code basis (due to the limitations of time, and constraints of this class, we did not analyze the data on a geographic basis).

Query 4: How many communities have at least 2 amenities listed?

As we saw from the results of the previous query there are many communities which have zero community services. There are also a large number of communities with 1 service. We wanted to filter for communities with 2 or more amenities as that seems more rare in the city. In total there are 30 communities with 2 or more listed community services. This query also incorporates a left join from community points to the community services dataset, so we retain communities that have no services. We do however, apply a filtering WHERE condition to filter for service_count greater than one (Appendix G, Query 4).

comm_code	community_name	service_count	service_rank
DNC	DOWNTOWN COMMERCIAL CORE	24	1
BLN	BELTLINE	6	2
FLN	FOREST LAWN	5	3

UOC	UNIVERSITY OF CALGARY	4	4
HUN	HUNTINGTON HILLS	4	4
ING	INGLEWOOD	3	6
HOU	HOUNSFIELD HEIGHTS/BRIAR HILL	3	6
BRD	BRIDGELAND/RIVERSIDE	3	6
CAP	CAPITOL HILL	2	9
RAN	RANCHLANDS	2	9
KIN	KINGSLAND	2	9
SOC	SOUTH CALGARY	2	9
SHN	SHAWNESSY	2	9
BOW	BOWNESS	2	9
SNA	SUNALTA	2	9
SOW	SOUTHWOOD	2	9
ACA	ACADIA	2	9
STA	ST. ANDREWS HEIGHTS	2	9
MON	MONTGOMERY	2	9
CED	CEDARBRAE	2	9
REN	RENFREW	2	9
SPR	SPRUCE CLIFF	2	9
BRE	BRENTWOOD	2	9
ARB	ARBOUR LAKE	2	9
SET	SETON	2	9
DNW	DOWNTOWN WEST END	2	9
MOP	MOUNT PLEASANT	2	9
HIL	HILLHURST	2	9
RCK	ROSSCARROCK	2	9
PIN	PINERIDGE	2	9

Data Analysis - Crime Statistics

Query 1: What are the top 5 and bottom 5 communities in terms of the total number of crimes in 2024?

This query helps to identify which communities are the safest and least safe in the most recent data, which would be an important metric to prospective buyers (See Appendix H, Query 1)

community	total_crime_2024	crime_rank_desc	crime_rank_asc
BELTLINE	1147.0	1	214
DOWNTOWN COMMERCIAL CORE	760.0	2	213
FOREST LAWN	410.0	3	212
POINT MCKAY	4.0	212	3
DIAMOND COVE	2.0	213	2
TWINHILLS	1.0	214	1

The results of the query show that the Beltline, the Downtown Commercial Core, and Forest Lawn had the highest total number of crimes in 2024. Additionally, the communities with the fewest crimes in 2024 (Point McKay, Diamond Cove, and Twinhills) each had less than 5 total crimes.

Query 2: Which communities have the least stable crime patterns?

This query looks at the volatility of monthly crime patterns by community. Even if a community has relatively low numbers of crimes, spikes in monthly crime rates can indicate instability, which is something that potential homebuyers may want to avoid (refer to Appendix H, Query 2).

community	monthly_crime_avg	monthly_crime_stddev
BELTLINE	15.9075	17.9143
DOWNTOWN COMMERCIAL CORE	12.6377	13.4612
DOWNTOWN EAST VILLAGE	4.9111	5.8021
FOREST LAWN	5.3815	4.4082
HILLHURST	4.1573	4.0141

Not only do the Beltline, the Downtown Commercial Core, and Forest Lawn have the highest total crime rates in 2024, they also have some of the most volatility in terms of month-to-month changes in number of crimes. Downtown East Village and Hillhurst also rank in the top five communities in terms of month-to-month instability in crime rates, but they have smaller total number of crimes per month.

Query 3: Which communities have shown the greatest increase or decrease in crime from 2018 to 2024?

This query helps to identify which communities prospective home buyers may want to avoid because they have rising crime rates, or to keep an eye on because the crime rates are falling (see Appendix H, Query 3).

community	crime_2018	crime_2024	crime_change
BELTLINE	1729.0	1147.0	-582.0
DOWNTOWN COMMERCIAL CORE	1149.0	760.0	-389.0
BOWNESS	494.0	240.0	-254.0
BANKVIEW	348.0	108.0	-240.0
PINERIDGE	417.0	193.0	-224.0
...	...	...	...
CARRINGTON	8.0	48.0	40.0
LIVINGSTON	4.0	58.0	54.0
CORNERSTONE	17.0	133.0	116.0
SETON	48.0	179.0	131.0
DOWNTOWN EAST VILLAGE	133.0	385.0	252.0

Despite having the highest total number of crimes in 2024, the Beltline and the Downtown Commercial Core have actually shown the greatest decrease in total crime compared to 2018. This indicates that despite the high crime rates, these communities may be worth keeping an eye on as the crime rates are falling. Bowness, Bankview, and Pineridge also ranked in the top 5 communities in terms of the greatest decrease in total crime from 2018 to 2024. All 5 communities with the greatest decreases in crime rate from 2018 to 2024 are well-established communities (City of Calgary, 2025b), perhaps reflecting revitalization efforts.

The communities with the largest increase in crime from 2018 to 2024 tended to be developing neighborhoods, either established in the 2000s (Seton, Cornerstone) or currently in the building-out phase (Livingston, Carrigton) (City of Calgary, 2025b). The only truly established community in the top 5 greatest increases in crime was Downtown East Village.

Query 4: Which communities have the highest rates of *violent* crimes in 2024?

There are 9 categories of crime included in the table:

1. Assault (Non-domestic)
2. Break & Enter - Commercial
3. Break & Enter - Dwelling
4. Theft FROM Vehicle
5. Theft OF Vehicle
6. Violence 'Other' (Non-domestic)
7. Street Robbery
8. Commercial Robbery
9. Break & Enter - Other Premises

In this analysis, "Assault (Non-domestic)", "Violence 'Other' (Non-domestic)", "Street Robbery", and "Commercial Robbery" are considered violent crimes. It would be important for home buyers to not only know how much crime occurs overall, but specifically how many violent crimes. High violent crime rates would likely be a stronger deterrent for a home buyer than high property crime rates (See Appendix H, Query 4).

community	violent_crime_2024	total_crime_2024	pct_violent_crime_2024
BELTLINE	472.0	1147.0	41.15
DOWNTOWN COMMERCIAL CORE	389.0	760.0	51.18
DOWNTOWN EAST VILLAGE	255.0	385.0	66.23
FOREST LAWN	151.0	410.0	36.83
MARLBOROUGH	136.0	309.0	44.01

The 3 communities with the highest rates of total crime in 2024 (Beltline, Downtown Commercial Core, Forest Lawn) also have some of the highest numbers of violent crime. Downtown East Village, which had the greatest increase in crime from 2018 to 2024, also has a high number of violent crimes in 2024, suggesting it may be an area to avoid.

Reordering the results to be sorted by the *percentage* of crime in the community that is violent gives the following:

community	violent_crime_2024	total_crime_2024	pct_violent_crime_2024
DOWNTOWN EAST VILLAGE	255.0	385.0	66.23
ST. ANDREWS HEIGHTS	47.0	89.0	52.81
DOWNTOWN COMMERCIAL CORE	389.0	760.0	51.18
COUNTRY HILLS VILLAGE	30.0	60.0	50.00
EAGLE RIDGE	23.0	50.0	46.00

Downtown East Village and Downtown Commercial Core make an appearance again, confirming that these communities not only have high absolute numbers of violent crime, but also that a high *percentage* of the total crimes in these areas are violent. Other communities such as St. Andrew Heights, Country Hills Village, and Eagle Ridge show up, showing that despite these communities having relatively fewer crimes overall, a high proportion of those crimes are violent.

Data Analysis - Combined Data Queries

Query 1: Which Communities Had the Greatest Changes in Assessed Property Value from 2024-2025?

This query helps potential property values understand which communities are rapidly increasing in value, and which are decreasing (see Appendix I, Query 1).

First the (pre-calculated) historical median assessed value for each community was pulled from the Historical Property Assessments data. Then the current median assessed value for each community was calculated from the Current Property Assessments dataset using a windowing function. The historical (2024) medians were inner-joined with the current (2025) medians on the community code. An inner join was chosen to select for only the communities present in both tables, as it would not make sense to calculate a change when the community is missing from one of the datasets.

comm_code	comm_name	median_2024	median_2025	value_change	pct_change
HSN	HASKAYNE	405000.0	713500.0	308500.0	76.17
GAG	GARRISON GREEN	206500.0	325250.0	118750.0	57.51
HKS	HOTCHKISS	127000.0	195500.0	68500.0	53.94
COA	COACH HILL	372500.0	535500.0	163000.0	43.76
MIS	MISSION	193500.0	262500.0	69000.0	35.66
...	...	...	...	...	...
HIL	HILLHURST	775000.0	518000.0	-257000.0	-33.16
CUR	CURRIE BARRACKS	634000.0	341500.0	-292500.0	-46.14
DIS	DISCOVERY RIDGE	988000.0	532000.0	-456000.0	-46.15
LEB	LEWISBURG	1296376.0	147000.0	-1149376.0	-88.66
MDH	MEDICINE HILL	6100000.0	592000.0	-5508000.0	-90.30

Looking at year-over-year changes, the strongest growth was in newer suburban communities like Haskayne, Garrison Green and Hotchkiss, where values jumped more than 50%. Established areas like Coach Hill and inner-city neighbourhoods like Mission also saw appreciation. On the other end, several high-value communities, like Hillhurst, Currie Barracks and Discovery Ridge, saw decreases of over 30%. Lewisburg and Medicine Hill showed extreme declines, but these outliers are likely due to large parcels being divided into smaller parcels as the communities develop.

Query 2: Are Communities with the most services the most expensive?

The query to get the results first builds a subquery (q4) that counts the number of community services linked to each community point (Community_Points), using a LEFT JOIN to ensure communities are included even if they have no services. It then ranks communities by service count. A second subquery (q5) calculates the median re-assessed property value for each community from Current_Property_Assessments, restricted to residential assessments. Finally, the main query performs an INNER JOIN between q4 and q5 on comm_code, ensuring only communities that have both service information and property assessment data appear in the results. See Appendix I, Query 2.

comm_code	community_name	service_count	service_rank	median_re_assessed_value
FLN	FOREST LAWN	5	3	467000

HUN	HUNTINGTON HILLS	4	4	578000
BRD	BRIDGELAND/RIVERSIDE	3	6	499000
HOU	HOUNSFIELD HEIGHTS/BRIAR HILL	3	6	903000
ING	INGLEWOOD	3	6	607000
...	...	...	...	...
WAL	WALDEN	0	122	561750
WBN	WOODBINE	0	122	646500
WSP	WEST SPRINGS	0	122	893000
WWO	WOLF WILLOW	0	122	522500
YKV	YORKVILLE	0	122	787250

These results indicate that some communities which are relatively lowly priced (for example, Forest Lawn) have a high count of amenities. Additionally, of the 100 communities with no amenities which were analyzed in the individual analysis portion for Community Services, the pricing varies significantly.

Query 3: At Each Price Level (Low, Medium, High), which communities have the most amenities?

To answer this question we first created a table that had the service count, service, rank, re_median_assessed_value and property_value_class for all communities (refer to Appendix I, Query 3.1). The query uses a LEFT JOIN between Community_Points and Community_Service so that all communities of class_code = 1 are included, even if they have no associated services; this retains service counts per community. It then applies an INNER JOIN between that result and Current_Property_Assessments, ensuring only communities with valid residential property assessment data are kept. Together, the joins retain community identifiers, names, service counts and ranks, along with median property values for those communities.

The results of that query were loaded into a table called “Communities_service_and_value” and we queried that table to get the top 5 communities for each price level. The query used to get the top 5 communities in terms of amenities at each price level did not use any joins (refer to Appendix I, Query 3.2)

comm_code	community_name	service_count	service_rank	median_re_assessed_value	property_value_class
HOU	HOUNSFIELD HEIGHTS/BRIAR HILL	3	6	903000	High
HIL	HILLHURST	2	9	880250	High
MOP	MOUNT PLEASANT	2	9	781750	High
STA	ST. ANDREWS HEIGHTS	2	9	1000000	High
ALT	ALTADORE	1	31	1000000	High

FLN	FOREST LAWN	5	3	467000	Low
BRD	BRIDGELAND/RIVERSIDE	3	6	499000	Low
PIN	PINERIDGE	2	9	481000	Low
SET	SETON	2	9	461250	Low
SNA	SUNALTA	2	9	253000	Low
HUN	HUNTINGTON HILLS	4	4	578000	Medium
ING	INGLEWOOD	3	6	607000	Medium
ACA	ACADIA	2	9	595000	Medium
ARB	ARBOUR LAKE	2	9	683000	Medium
BOW	BOWNESS	2	9	541500	Medium

The listing above indicates that it is possible to find communities with similar levels of amenities at all three price points (especially, when looking for communities with 2 or 3 amenities). However, the residential only community with the highest amount of amenities is in the “low” price category. This indicates that community services do not necessarily have a linear relationship with the assessed property values.

Query 4: Do the most affordable communities have the highest crime rates?

This query allows potential property buyers to see both the median property value and the crime burden of each community, ranked side by side (See Appendix I, Query 4). This would allow them to identify communities that are affordable and safe, expensive but safe, or affordable but high-crime, by comparing the various ranks.

Aggregated crime counts per community were computed from the Community Crime data, then the median assessed values per community were computed from the Current Property Assessment data. Then, the community medians were LEFT-JOINed to the crime totals on community name. The LEFT JOIN ensures that communities without crimes in 2024 still show up. Ranking was done to create an affordability rank (1=lowest median property value assessment) and a crime_rank (1=most crimes).

community	median_assessed_value	total_crime_2024	affordability_rank	crime_rank
DOWNTOWN WEST END	37000.0	84.0	1	81
EAU CLAIRE	45500.0	45.0	2	131
RICARDO RANCH	66000.0	NaN	3	210
DOWNTOWN COMMERCIAL CORE	146500.0	760.0	4	2
LEWISBURG	147000.0	NaN	5	210
...	...	...	...	...

BEL-AIRE	2480000.0	5.0	212	205
HUXLEY	11497750.0	NaN	213	210
OSPREY HILL	12960200.0	NaN	214	210
UNIVERSITY OF CALGARY	24770000.0	21.0	215	170
QUEENS PARK VILLAGE	42860000.0	7.0	216	202

The most affordable communities did not have the highest crime rates. For example, Downtown West End and Eau Claire, the most affordable communities, had the 81st and 131st most crimes. The Downtown Commercial Core, which has high numbers of crimes (ranked #2 in total crimes), did not have the absolute cheapest assessments, although it was indeed high on the affordability ranking (ranked #4 in affordability). Wealthy communities such as Bel-Aire and Huxley showed low crime counts. These results show that the pattern between crime rates and assessed property value is not a direct, straightforward relationship.

Query 5: At each price level (low, medium, high) which communities have the most and least crime?

This query would allow potential property buyers to know what the best and worst options are at their budget level (See Appendix I, Query 5).

community	price_band	total_crime_2024	crime_extreme
MAYFAIR	High	NaN	Least Crime
HILLHURST	High	131.0	Most Crime
RICARDO RANCH	Low	NaN	Least Crime
FOREST LAWN	Low	410.0	Most Crime
MORAINE	Medium	NaN	Least Crime
SADDLE RIDGE	Medium	309.0	Most Crime

Crime totals were aggregated per community from the Community Crime data. Community medians were calculated from the Property Assessment data. A ‘banded’ data table was created, with each community assigned an assessment class (Low, Medium, High) based on median assessed value. The banded table was LEFT JOINED to the total crime data on community name. This ensured that every community’s assessment class data was matched with its crime totals, but any communities that lacked crime data were still included.

Query 6: What are the top 5 communities that are the most affordable, have the fewest crimes, and have the most amenities?

This query provides a composite ranking method to rank the communities by the best overall balance of affordability (low median property assessed value), safety (low crime rates), and access to amenities.
The Community Crime dataset was INNER-JOINed to the Community Points dataset to match each crime count to its community code. The community code was used to LEFT JOIN the calculated crime totals to each community. The amenity counts were also LEFT-JOINed to the data using community code to attach the number

of amenities/services to each community. The LEFT JOINS ensure communities without amenities or without crimes still appear (with 0 via COALESCE).

comm_name	median_assessed_value	assessment_class	service_count	service_rank	total_crime_2024	safety_rank	affordability_rank	overall_score	overall_rank
SPRUCE CLIFF	362000.0	Low	2	6	40.0	70	15	91	**1**
RUTLAND PARK	398500.0	Low	1	27	25.0	47	18	92	**2**
RICARDO RANCH	73000.0	Low	0	113	0.0	1	1	115	**3**
LEWISBURG	146750.0	Low	0	113	0.0	1	2	116	**4**
KEYSTONE HILLS	170000.0	Low	0	113	0.0	1	3	117	**5**
...	...	...	...	...	...	...	...	...	**...**
AUBURN BAY	687500.0	Medium	0	113	113.0	158	123	394	**199**
SPRINGBANK HILL	819250.0	High	0	113	72.0	119	167	399	**201**
ASPEN WOODS	1250000.0	High	0	113	57.0	99	194	406	**202**
SADDLE RIDGE	616500.0	Medium	0	113	309.0	202	93	408	**203**
MAHOGANY	763500.0	Medium	0	113	123.0	166	150	429	**204**

The top ranked communities were Spruce Cliff and Rutland Park, as they both were relatively affordable (in the “Low” assessment class), had modest crime counts, and at least some services. Ricardo Ranch, Lewisburg, and Keystone Hills also rank highly overall, as they were very affordable, and had nearly zero crimes, likely due to their developing nature. However, as they are developing they also lack amenities, which may limit their appeal despite their affordability/safety.

In addition to the analysis provided in the combined queries section we exported the table resulting from Query 6 (Appendix K) and attempted to create a linear model of median_assessed_value as explained by service count and total crime rate. The resulting model was not statistically significant overall, and neither of the predictors were statistically significant indicating that the relationship between crime, services and property value is non-linear when aggregating on a community basis.

Limitations & Future Directions

The Community Services dataset includes attractions, libraries, community centres, etc., but it does not include information on other amenities such as schools, retail, and transit. We considered including transit information, but no transit dataset could be found that had the information by community. A future direction could be to map the addresses of each transit station to the community they belong to, and use the resulting community transit data to investigate the role of transportation in choosing a community to live in. Additionally, amenities were only counted for the actual community they reside in. This does not take into account situations where an amenity may be in close proximity to multiple communities. A future direction could be to use the centroid points of the communities and calculate the distance to each amenity.

The community crime statistics are aggregated to communities, which may hide hotspots within neighbourhoods. Additionally, there are no population counts for the communities, so crime rates per capita can not be directly calculated. Future analyses could join the community crime dataset with a community population dataset.

Another limitation of our analysis is that the Current Year and Historical Property Assessments datasets rely on the assessments made by the government for taxation purposes, which may differ from the actual market values of properties. Access to real estate data in Calgary (and Canada more broadly) is difficult to gain as a lot of data is proprietary to the Canadian Real Estate Association (CREA). CREA has a database of active listings in Canada with granular details about homes (such as number of bedrooms, number of bathrooms etc..) however it is restricted to licensed realtors and the technologists working on behalf of licensed realtors.

Personal Reflection

Thanusha's Reflection

I have previous familiarity with using SQL both in school during undergrad, and in the workplace. The previous experience allowed me to attempt more challenging queries on this project than I typically use on a day to day basis. I had queries with multiple common table expressions (CTE), and joins. I was less familiar with the window functions used like Rank(), Median(), and Percentile(). They were a little tricky to use as we are not able to filter using them.

Another thing I found challenging was learning how to use the Socrata Open Data API (SODA). I initially pulled tables directly from SODA without pre-querying. For the historical property assessments dataset I could not do that as it was too large, so I had to run a query off of the City of Calgary Open Data Site using SODA and connect to the end point for my query. In terms of data sets, with more time and/or resources it would be engaging to learn how to work with the geographic coordinate data, so rather than matching communities to amenities by community code alone, we would search for communities within a geographic radius and have an amenity be tied to multiple communities.

Elise's Reflection

I had never used SQL previously, so there was a steep learning curve on this project. One of the challenges I kept running up against was finding out certain functions do not run in MariaDB and require workarounds. For example, I found that MariaDB does not allow the MEDIAN() function to act as an aggregate function, it only works as a windowing function. Additionally, PERCENTILE_CONT() would not work in the version of MariaDB we have. I learned to be creative and create workarounds that accomplished the same intended results. Reflecting on our choice of datasets, I would have definitely included a dataset with community population counts, as I found there was no way to account for the potential that the high crime rates are just due to high population counts. I learned about common table expressions (CTEs), which I used to clean up my queries. I also learned how to use CASE logic to classify observations.

Something I found difficult was considering how to create a composite score that could balance the dimensions of affordability, safety, and services. Next time, I would normalize the ranks before combining them, so that each factor contributes equally. I would also pay closer attention to the join conditions (community name vs code) to avoid missing data. Additionally, I would use visualisation earlier in the process to spot patterns rather than trying to see relationships through tables alone. It would be interesting to incorporate historical trends of crime and property values over time together to see if patterns persist. Mapping communities and amenities to their actual spatial data would allow for greater insights into the effect of amenities, as we could then take into account amenities close to multiple communities.

Conclusion

This project examined the question of which Calgary communities are the most ideal for homeownership by integrating residential property assessments, crime statistics, and community amenities using City of Calgary open data. Through structured SQL analysis and cross-dataset joins, we demonstrated that property value, safety, and access to amenities interact in complex and non-linear ways.

While lower-priced communities were not necessarily associated with higher crime rates, and highly serviced communities were not always the most expensive, certain neighbourhoods (such as Spruce Cliff and Rutland Park) emerged as offering a strong balance of affordability, relatively low crime, and access to amenities.

Conversely, some high-value and high-amenity areas experienced greater crime volatility, underscoring the importance of considering multiple dimensions rather than relying on a single indicator when evaluating neighbourhood desirability. Overall, this analysis highlights that “ideal” communities depend on household priorities, and that data-driven, multi-factor evaluation provides a more nuanced and realistic framework for prospective buyers navigating Calgary’s rapidly evolving housing market.

References

Lundy, M. (2025, January 18). Calgary's population growth is in a league of its own. The Globe and Mail. https://www.theglobeandmail.com/business/article-calgarys-population-growth-is-in-a-league-of-its-own/

Statistics Canada. (2025). Table 17-10-0148-01 Population estimates, July 1, by census metropolitan area and census agglomeration, 2021 boundaries [Data table]. https://doi.org/10.25318/1710014801-eng

City of Calgary. (2025a). Open Government License - City of Calgary. data.calgary.ca. https://data.calgary.ca/stories/s/Open-Calgary-Terms-of-Use/u45n-7awa

City of Calgary. (2025b). Community Points [Data set]. Open Calgary. https://data.calgary.ca/Base-Maps/Community-Points/j9ps-fyst/about_data

City of Calgary. (2025c). Community Services [Data set]. Open Calgary. https://data.calgary.ca/Services-and-Amenities/Community-Services/x34e-bcjz/about_data

City of Calgary. (2025d). Community Crime Statistics [Data set]. Open Calgary. https://data.calgary.ca/Health-and-Safety/Community-Crime-Statistics/78gh-n26t/about_data

City of Calgary. (2025e). Current Year Property Assessments (Parcel) [Data set]. Open Calgary. https://data.calgary.ca/Government/Current-Year-Property-Assessments-Parcel-/4bsw-nn7w/about_data

City of Calgary. (2025f). Historical Property Assessments (Parcel) [Data set]. Open Calgary. https://data.calgary.ca/Government/Historical-Property-Assessments-Parcel-/4ur7-wsgc/about_data

Appendix A: Community Points Dataset

All information from the City of Calgary (2025b).

Column	Notes/Description	Usage in Project
CLASS	Community class	N/A - not used
CLASS_CODE	Residential, Industrial, Major Park, or Residual Sub Area	Was used to filter for class_code of "1" indicating residential areas
COMM_CODE	A three-character code associated with the community	Was used as a legend to ensure each table has both the community name and the community code for clarity.
NAME	Name of community	
SECTOR	Section of the city (N, S, NW, etc.)	N/A - not used
SRG	Suburban Residential Growth	N/A - not used
COMM_STRUCTURE	Used to identify life-cycle patterns	N/A - not used
longitude	Longitude	N/A - not used
latitude	Latitude	N/A - not used
POINT	Geographical centroid	N/A - not used

Appendix B: Community Services Dataset

All information from the City of Calgary (2025c).

Column Name	Notes/Description	Usage in Project
TYPE	Type of community service/amenity	Used in Group By to analyze how many amenities of each type there were
NAME	Name of particular amenity	N/A - not used
ADDRESS	Address of amenity	N/A - not used
COMM_CODE	A three character code associated with the community	Was used as a key to join to other tables, and as a code to summarize by community.
POINT	Geographical point of amenity	N/A - not used

Appendix C: Community Crime Statistics

All information from the City of Calgary (2025d).

Column Name	Notes/Description	Usage in Project
Community	Name of community	Used to group data by community, then used as a key to join with other tables.
Category	Category of crime	Used to group by crime type, not just overall crime rates
Crime Count	Number of times the crime occurred	Used to find crime counts in each community
Year	The year the crime occurred	Used to filter for the year of interest
Month	The month the crime occurred	N/A - not used

Appendix D: Current Year Property Assessments (Parcel) Dataset

All information from the City of Calgary (2025e).

Column Name	Notes/Description	Usage in Project
ROLL_YEAR	Assessment roll year	Used when joining with the Historical Property Assessment table
ROLL_NUMBER	Assessment account roll number	N/A - not used
ADDRESS	Civic address	N/A - not used
ASSESSED_VALUE	Total assessed value of the property	N/A - not used
ASSESSMENT_CLASS	Predominant assessment class code	Used to filter for assessment class of “RE” (residential)
ASSESSMENT_CLASS_DESCRIPTION	Description of predominant assessment class	N/A - not used
RE_ASSESSED_VALUE	Residential property assessed value	Used to get the values of only the residential portions of property that was mixed use
NR_ASSESSED_VALUE	Non-residential property assessed value	N/A - not used
FL_ASSESSED_VALUE	Farmland property assessed value	N/A - not used
COMM_CODE	Community code	Used to group by community, and can be used to join with the Community Services table.
COMM_NAME	Community name	Used to group by community, and can be used to join with the Community Crime Statistics table
YEAR_OF_CONSTRUCTION	Actual year of construction	N/A - not used
LAND_USE_DESIGNATION	Land use designation	Used - used to filter for residential property values only.
PROPERTY_TYPE	Property type; land-only or land-and-improvement (parcel with building)	N/A - not used
LAND_SIZE_SM	Parcel/land size in square meters	N/A - not used
LAND_SIZE_SF	Parcel/land size in square feet	N/A - not used
LAND_SIZE_AC	Parcel/land size in square acres	N/A - not used
MOD_DATE	Last modified date	N/A - not used
SUB_PROPERTY_USE	Sub-property type	N/A - not used
MULTIPOLYGON	Property boundaries	N/A - not used
UNIQUE_KEY	Unique key	N/A - not used

Appendix E: Historical Property Assessments (Parcel) Dataset

All information from the City of Calgary (2025f).

Column Name	Notes/Description	Usage in Project
ROLL_YEAR	Assessment roll year	Used to group data by Year
ROLL_NUMBER	Assessment account roll number	N/A - not used
ADDRESS	Civic address	N/A - not used
ASSESSED_VALUE	Total assessed value of the property	N/A - not used
ASSESSMENT_CLASS	Predominant assessment class code	Used to filter for only residential properties
ASSESSMENT_CLASS_DESCRIPTION	Description of predominant assessment class	N/A - not used
RE_ASSESSED_VALUE	Residential property assessed value	Used to get the values of only the residential portions of property that was mixed use
NR_ASSESSED_VALUE	Non-residential property assessed value	N/A - not used
FL_ASSESSED_VALUE	Farmland property assessed value	N/A - not used
COMM_CODE	Community code	Used to group by community, and used to join with other tables..
COMM_NAME	Community name	Used to group by community, and can be used to join with other tables.
YEAR_OF_CONSTRUCTION	Actual year of construction	N/A - not used
LAND_USE_DESIGNATION	Land use designation	Used to filter directly via SODA api on City of Calgary server for residential property values only.
PROPERTY_TYPE	Property type; land-only or land-and-improvement (parcel with building)	N/A - not used
LAND_SIZE_SM	Parcel/land size in square meters	N/A - not used
LAND_SIZE_SF	Parcel/land size in square feet	N/A - not used
LAND_SIZE_AC	Parcel/land size in square acres	N/A - not used
SUB_PROPERTY_USE	Sub-property type	N/A - not used
MULTIPOLYGON	Boundaries of property	N/A - not used

Appendix F - Property Assessment Only Queries

Creation of a view that filters out unwanted retail types:

```
CREATE
	OR replace VIEW filtered_current_property_assessments AS

SELECT cpa.*
	,(cpa.re_assessed_value / cpa.land_size_sf) AS value_per_sqft
FROM current_property_assessments cpa
LEFT JOIN community_points cp ON cpa.comm_name = cp.name
WHERE cp.class = "residential"
	AND cpa.assessment_class_description = 'Residential'
	AND cpa.assessment_class = 'RE'
	AND cpa.re_assessed_value IS NOT NULL
	AND cpa.land_use_designation NOT LIKE 'DC%'
	AND cpa.land_use_designation NOT LIKE 'C%'
	AND cpa.land_use_designation NOT LIKE 'S%'
	AND cpa.land_use_designation NOT LIKE 'I%';
```

Query 1: Which communities are most and least affordable in 2025?

```
WITH community_medians
AS (
	SELECT DISTINCT comm_name
		,MEDIAN(re_assessed_value) OVER (PARTITION BY comm_name) AS
median_assessed_value
	FROM filtered_Current_Property_Assessments
	WHERE roll_year = 2025
		AND assessment_class = 'RE'
		AND re_assessed_value IS NOT NULL
		AND land_use_designation NOT LIKE 'DC%'
		AND land_use_designation NOT LIKE 'C%'
		AND land_use_designation NOT LIKE 'S%'
		AND land_use_designation NOT LIKE 'I%'
	)
SELECT comm_name
	,median_assessed_value
FROM community_medians
ORDER BY median_assessed_value;
```

Query 2: What are the values of the quartiles for median assessed value by community?

```
WITH community_medians AS (
  SELECT
    DISTINCT comm_name,
    MEDIAN(re_assessed_value) OVER (PARTITION BY comm_name) AS median_assessed_value
  FROM filtered_Current_Property_Assessments
  WHERE roll_year = 2025
    AND assessment_class = 'RE'
```

```
    AND re_assessed_value IS NOT NULL
    AND land_use_designation NOT LIKE 'DC%'
    AND land_use_designation NOT LIKE 'C%'
    AND land_use_designation NOT LIKE 'S%'
    AND land_use_designation NOT LIKE 'I%'
),
ranked AS (
  SELECT
    comm_name,
    median_assessed_value,
    ROW_NUMBER() OVER (ORDER BY median_assessed_value) AS pos
  FROM community_medians
),
totals AS (
  SELECT COUNT(*) AS m FROM community_medians
)
SELECT
  MAX(CASE WHEN r.pos = FLOOR(0.25 * (t.m + 1)) THEN r.median_assessed_value END) AS q1,
  MAX(CASE WHEN r.pos = FLOOR(0.50 * (t.m + 1)) THEN r.median_assessed_value END) AS q2,
  MAX(CASE WHEN r.pos = FLOOR(0.75 * (t.m + 1)) THEN r.median_assessed_value END) AS q3
FROM ranked r
CROSS JOIN totals t;
```

Query 3: What is the actual range of values in each Assessment Class?

```
WITH community_medians AS (
  SELECT
    DISTINCT comm_name,
    MEDIAN(re_assessed_value) OVER (PARTITION BY comm_name) AS median_assessed_value
  FROM filtered_Current_Property_Assessments
  WHERE roll_year = 2025
    AND assessment_class = 'RE'
    AND re_assessed_value IS NOT NULL
    AND land_use_designation NOT LIKE 'DC%'
    AND land_use_designation NOT LIKE 'C%'
    AND land_use_designation NOT LIKE 'S%'
    AND land_use_designation NOT LIKE 'I%'
),
assessment_class AS (
  SELECT
    comm_name,
    median_assessed_value,
    NTILE(4) OVER (ORDER BY median_assessed_value) AS quartile
  FROM community_medians
)
SELECT
  quartile,
  MIN(median_assessed_value) AS min_value,
  MAX(median_assessed_value) AS max_value,
  COUNT(*) AS n_communities
FROM assessment_class
```

```
GROUP BY quartile
ORDER BY quartile;
```

Query 4: Can we classify communities into ‘Low’, ‘Medium’, and ‘High’ categories based on median assessed value?

```
WITH community_medians AS (
  SELECT DISTINCT
    comm_name,
    MEDIAN(re_assessed_value) OVER (PARTITION BY comm_name) AS median_assessed_value
  FROM filtered_Current_Property_Assessments
  WHERE roll_year = 2025
    AND assessment_class = 'RE'
    AND re_assessed_value IS NOT NULL
    AND land_use_designation NOT LIKE 'DC%'
    AND land_use_designation NOT LIKE 'C%'
    AND land_use_designation NOT LIKE 'S%'
    AND land_use_designation NOT LIKE 'I%'
),
banded AS (
  SELECT
    comm_name,
    median_assessed_value,
    NTILE(4) OVER (ORDER BY median_assessed_value) AS quartile
  FROM community_medians
)
SELECT
  b.comm_name AS community,
  b.median_assessed_value,
  CASE
    WHEN quartile=1 THEN 'Low'
    WHEN quartile IN (2, 3) THEN 'Medium'
    ELSE 'High'
  END AS assessment_class
FROM banded b
```

Query 5: Which communities have the highest and lowest median assessed values in each assessment class?

```
WITH community_medians AS (
  SELECT
    DISTINCT comm_name,
    MEDIAN(re_assessed_value) OVER (PARTITION BY comm_name) AS median_assessed_value
  FROM filtered_Current_Property_Assessments
  WHERE roll_year = 2025
   AND assessment_class = 'RE'
   AND re_assessed_value IS NOT NULL
   AND land_use_designation NOT LIKE 'DC%'
   AND land_use_designation NOT LIKE 'C%'
   AND land_use_designation NOT LIKE 'S%'
```

```
    AND land_use_designation NOT LIKE 'I%'
),
banded AS (
  SELECT
    comm_name,
    median_assessed_value,
    NTILE(4) OVER (ORDER BY median_assessed_value) AS quartile
  FROM community_medians
),
classified AS (
  SELECT
    comm_name,
    median_assessed_value,
    CASE
      WHEN quartile = 1 THEN 'Low'
      WHEN quartile IN (2,3) THEN 'Medium'
      ELSE 'High'
    END AS assessment_class
  FROM banded
),
extremes AS (
  SELECT
    assessment_class,
    comm_name,
    median_assessed_value,
    ROW_NUMBER() OVER (PARTITION BY assessment_class ORDER BY median_assessed_value ASC)
AS rn_low,
    ROW_NUMBER() OVER (PARTITION BY assessment_class ORDER BY median_assessed_value DESC)
AS rn_high
  FROM classified
)
SELECT assessment_class,
  MAX(CASE WHEN rn_low = 1 THEN comm_name END) AS lowest_community,
  MAX(CASE WHEN rn_low = 1 THEN median_assessed_value END) AS lowest_value,
  MAX(CASE WHEN rn_high = 1 THEN comm_name END) AS highest_community,
  MAX(CASE WHEN rn_high = 1 THEN median_assessed_value END) AS highest_value
FROM extremes
GROUP BY assessment_class
ORDER BY lowest_value;
```

Appendix G - Amenity Queries

Query 1: What types of Community Services Are There?

```
SELECT type, COUNT(name) AS service_count
FROM Community_Service
WHERE comm_code IS NOT NULL
GROUP BY type
ORDER BY service_count DESC;
```

Query 2: Which Communities Have the Most Services?

```
SELECT q3.comm_code
	,q3.community_name
	,COUNT(q3.service_name) AS service_count
	,RANK() OVER (
		ORDER BY COUNT(q3.service_name) DESC
		) AS service_rank
FROM (
	SELECT cp.class
		,cp.class_code
		,cp.comm_code
		,cp.name AS community_name
		,cp.sector
		,cp.srg
		,cs.type AS service_type
		,cs.name AS service_name
		,cs.address
	FROM Community_Points AS cp
	LEFT JOIN Community_Service AS cs ON cp.comm_code = cs.comm_code
	WHERE cp.class_code = 1
	) AS q3
GROUP BY q3.comm_code
	,q3.community_name
ORDER BY service_count DESC LIMIT 5;
```

Query 3: Which Communities Have Zero Amenities Listed?

```
SELECT *
FROM (
	SELECT q3.comm_code
		,q3.community_name
		,COUNT(q3.service_name) AS service_count
		,RANK() OVER (
			ORDER BY COUNT(q3.service_name) DESC
			) AS service_rank
	FROM (
		SELECT cp.class
			,cp.class_code
			,cp.comm_code
			,cp.name AS community_name
```

```
                ,cp.sector
                ,cp.srg
                ,cs.type AS service_type
                ,cs.name AS service_name
                ,cs.address
            FROM Community_Points AS cp
            LEFT JOIN Community_Service AS cs ON cp.comm_code = cs.comm_code
            WHERE cp.class_code = 1
            ) AS q3
        GROUP BY q3.comm_code
            ,q3.community_name
        ORDER BY service_count DESC
        ) AS ranked
WHERE ranked.service_count = 0;
```

Query 4: How many communities have at least 2 amenities listed?

```
SELECT *
FROM (
        SELECT q3.comm_code
            ,q3.community_name
            ,COUNT(q3.service_name) AS service_count
            ,RANK() OVER (
                ORDER BY COUNT(q3.service_name) DESC
                ) AS service_rank
        FROM (
            SELECT cp.class
                ,cp.class_code
                ,cp.comm_code
                ,cp.name AS community_name
                ,cp.sector
                ,cp.srg
                ,cs.type AS service_type
                ,cs.name AS service_name
                ,cs.address
            FROM Community_Points AS cp
            LEFT JOIN Community_Service AS cs ON cp.comm_code = cs.comm_code
            WHERE cp.class_code = 1
            ) AS q3
        GROUP BY q3.comm_code
            ,q3.community_name
        ORDER BY service_count DESC
        ) AS ranked
WHERE ranked.service_count > 1;
```

Appendix H - Crime Statistics Queries

Query 1: What are the top 5 and bottom 5 communities in terms of the total number of crimes in 2024?

```
WITH crime_totals AS (
    SELECT community, SUM(crime_count) AS total_crime_2024
    FROM filtered_Community_Crime
    WHERE year=2024
    GROUP BY community
),
ranked_crime AS (
    SELECT *,
        RANK() OVER (ORDER BY total_crime_2024 DESC) AS crime_rank_desc,
        RANK() OVER (ORDER BY total_crime_2024 ASC) AS crime_rank_asc
        FROM crime_totals
)
SELECT community, total_crime_2024, crime_rank_desc, crime_rank_asc
FROM ranked_crime
WHERE crime_rank_desc <= 3 OR crime_rank_asc <= 3
ORDER BY total_crime_2024 DESC;
```

Query 2: Which communities have the least stable crime patterns?

```
SELECT
    community,
    AVG(crime_count) AS monthly_crime_avg,
    STDDEV(crime_count) AS monthly_crime_stddev
FROM filtered_Community_Crime
GROUP BY community
ORDER BY monthly_crime_stddev DESC
LIMIT 5;
```

Query 3: Which communities have shown the greatest increase or decrease in crime from 2018 to 2024?

```
SELECT
    community,
    SUM(CASE WHEN year=2018 THEN crime_count ELSE 0 END) AS crime_2018,
    SUM(CASE WHEN year=2024 THEN crime_count ELSE 0 END) AS crime_2024,
    SUM(CASE WHEN year=2024 THEN crime_count ELSE 0 END) - SUM(CASE WHEN year=2018 THEN
crime_count ELSE 0 END) AS crime_change
FROM filtered_Community_Crime
GROUP BY community
ORDER BY crime_change ASC;
```

Query 4: Which communities have the highest rates of violent crimes in 2024?

```
SELECT
   community,
   SUM(CASE WHEN category IN (
      'Assault (Non-domestic)',
      'Violence  \\'Other\\' (Non-domestic)',
      'Street Robbery',
      'Commercial Robbery'
      ) THEN crime_count ELSE 0 END) AS violent_crime_2024,
   SUM(crime_count) as total_crime_2024,
   ROUND(100 * SUM(CASE WHEN category IN (
      'Assault (Non-domestic)',
      'Violence  \\'Other\\' (Non-domestic)',
      'Street Robbery',
      'Commercial Robbery'
      ) THEN crime_count ELSE 0 END) / SUM(crime_count), 2) AS pct_violent_crime_2024
FROM filtered_Community_Crime
WHERE year = 2024
GROUP BY community
ORDER BY violent_crime_2024 DESC;
```

Appendix I - COMBINED Analysis

Query 1: Which Communities Had the Greatest Changes in Assessed Property Value from 2024-2025?

```
WITH hist_2024 AS (
    SELECT
        comm_code,
        comm_name,
        median_re_assessed_value AS median_2024
    FROM filtered_Historical_Property_Assessments
    WHERE roll_year = 2024
),
curr_2025_raw AS (
    SELECT
        comm_code,
        comm_name,
        MEDIAN(re_assessed_value) OVER (PARTITION BY comm_code) AS median_2025
    FROM filtered_Current_Property_Assessments
    WHERE roll_year = 2025
        AND re_assessed_value IS NOT NULL
),
curr_2025 AS (
    SELECT DISTINCT
        comm_code,
        comm_name,
        median_2025
    FROM curr_2025_raw
)
SELECT
    h.comm_code,
    h.comm_name,
    h.median_2024,
    c.median_2025,
    (c.median_2025 - h.median_2024) AS value_change,
    ROUND(((c.median_2025 - h.median_2024) / h.median_2024) * 100, 2) AS pct_change
FROM hist_2024 h
JOIN curr_2025 c
    ON h.comm_code = c.comm_code
ORDER BY pct_change DESC;
```

Query 2: Are communities with the most services the most expensive?

```
SELECT q4.comm_code
        ,q4.community_name
        ,q4.service_count
        ,q4.service_rank
        ,q5.median_re_assessed_value
FROM (
        SELECT cp.comm_code
                ,cp.name AS community_name
```

```
		,COUNT(cs.name) AS service_count
		,RANK() OVER (
			ORDER BY COUNT(cs.name) DESC
			) AS service_rank
	FROM Community_Points AS cp
	LEFT JOIN Community_Service AS cs ON cp.comm_code = cs.comm_code
	WHERE cp.class_code = 1
	GROUP BY cp.comm_code
		,cp.name
	) AS q4
INNER JOIN (
	SELECT DISTINCT comm_code
		,MEDIAN(re_assessed_value) OVER (PARTITION BY comm_code) AS
median_re_assessed_value
	FROM Current_Property_Assessments
	WHERE assessment_class = 'RE'
		AND re_assessed_value IS NOT NULL
	) AS q5 ON q4.comm_code = q5.comm_code
ORDER BY q4.service_rank;
```

Query 3: At Each Price Level (Low, Medium, High), which communities have the most amenities?

Query 3.1 - This first query is used to get all the service counts, ranks and price bands for all communities.

```
WITH q6
AS (
	SELECT q4.comm_code
		,q4.community_name
		,q4.service_count
		,q4.service_rank
		,q5.median_re_assessed_value
	FROM (
		SELECT cp.comm_code
			,cp.name AS community_name
			,COUNT(cs.name) AS service_count
			,RANK() OVER (
				ORDER BY COUNT(cs.name) DESC
				) AS service_rank
		FROM Community_Points AS cp
		LEFT JOIN Community_Service AS cs ON cp.comm_code = cs.comm_code
		WHERE cp.class_code = 1
		GROUP BY cp.comm_code
			,cp.name
		) AS q4
	INNER JOIN (
		SELECT DISTINCT comm_code
			,MEDIAN(re_assessed_value) OVER (PARTITION BY comm_code) AS
median_re_assessed_value
```

```
		FROM Current_Property_Assessments
		WHERE assessment_class = 'RE'
			AND re_assessed_value IS NOT NULL
		) AS q5 ON q4.comm_code = q5.comm_code
	)
	,thresholds
AS (
	SELECT PERCENTILE_CONT(0.25) WITHIN
	GROUP (
			ORDER BY median_re_assessed_value
			) OVER () AS q1
		,PERCENTILE_CONT(0.75) WITHIN
	GROUP (
			ORDER BY median_re_assessed_value
			) OVER () AS q3
	FROM q6 LIMIT 1
	)
SELECT q6.comm_code
	,q6.community_name
	,q6.service_count
	,q6.service_rank
	,q6.median_re_assessed_value
	,CASE
		WHEN q6.median_re_assessed_value < thresholds.q1
			THEN 'Low'
		WHEN q6.median_re_assessed_value <= thresholds.q3
			THEN 'Medium'
		ELSE 'High'
		END AS property_value_class
FROM q6
	,thresholds
ORDER BY q6.service_rank;
```

Query 3.2 - After the results from the prior query (3.1) are loaded into a table called "Communities_service_and_value" the following query is run to get the summary for each price level:

```
WITH ranked
AS (
	SELECT comm_code
		,community_name
		,service_count
		,service_rank
		,median_re_assessed_value
		,property_value_class
		,ROW_NUMBER() OVER (
			PARTITION BY property_value_class ORDER BY service_count DESC
			) AS rank_within_class
	FROM Communities_service_and_value
	)
SELECT comm_code
	,community_name
```

```
        ,service_count
        ,service_rank
        ,median_re_assessed_value
        ,property_value_class
FROM ranked
WHERE rank_within_class <= 5
ORDER BY property_value_class
        ,service_count DESC;
```

Query 4: Are communities with the lowest crime rates the most expensive?

```
WITH crime_totals AS (
   SELECT
      community,
      SUM(crime_count) AS total_crime_2024
   FROM filtered_Community_Crime
   WHERE year = 2024
   GROUP BY community
),
community_medians AS (
   SELECT DISTINCT
      comm_name,
      MEDIAN(re_assessed_value) OVER (PARTITION BY comm_name) AS median_assessed_value
   FROM filtered_Current_Property_Assessments
   WHERE roll_year = 2025
     AND assessment_class = 'RE'
     AND re_assessed_value IS NOT NULL
),
joined AS (
   SELECT
      cm.comm_name AS community,
      cm.median_assessed_value,
      ct.total_crime_2024,
      RANK() OVER (ORDER BY cm.median_assessed_value ASC) AS affordability_rank,
      RANK() OVER (ORDER BY ct.total_crime_2024 DESC) AS crime_rank
   FROM community_medians cm
   LEFT JOIN crime_totals ct ON ct.community = cm.comm_name
)
SELECT
   community,
   median_assessed_value,
   total_crime_2024,
   affordability_rank,
   crime_rank
FROM joined
ORDER BY crime_rank;
```

Query 5: At each price level (low, medium, high) which communities have the most and least crime?

```
WITH crime_totals AS (
```

```
    SELECT
        community,
        SUM(crime_count) AS total_crime_2024
    FROM filtered_Community_Crime
    WHERE year = 2024
    GROUP BY community
),
community_medians AS (
    SELECT DISTINCT
        comm_name,
        MEDIAN(re_assessed_value) OVER (PARTITION BY comm_name) AS median_assessed_value
    FROM filtered_Current_Property_Assessments
    WHERE roll_year = 2025
),
banded AS (
    SELECT
        comm_name,
        median_assessed_value,
        NTILE(4) OVER (ORDER BY median_assessed_value) AS quartile
    FROM community_medians
),
joined AS (
    SELECT
        b.comm_name AS community,
        b.median_assessed_value,
        CASE
            WHEN quartile=1 THEN 'Low'
            WHEN quartile IN (2, 3) THEN 'Medium'
            ELSE 'High'
        END AS price_band,
        ct.total_crime_2024
    FROM banded b
    LEFT JOIN crime_totals ct
      ON ct.community = b.comm_name
),
ranked AS (
    SELECT
        community,
        price_band,
        total_crime_2024,
        ROW_NUMBER() OVER (PARTITION BY price_band ORDER BY total_crime_2024 DESC) AS rank_desc,
        ROW_NUMBER() OVER (PARTITION BY price_band ORDER BY total_crime_2024 ASC)  AS rank_asc
    FROM joined
)
SELECT community, price_band, total_crime_2024,
      CASE WHEN rank_desc = 1 THEN 'Most Crime'
           WHEN rank_asc = 1 THEN 'Least Crime'
      END AS crime_extreme
FROM ranked
WHERE rank_desc = 1 OR rank_asc = 1
ORDER BY price_band, crime_extreme;
```

Query 6: What are the top 5 communities that are the most affordable, have the fewest crimes, and have the most amenities?

```
WITH community_medians AS (
    SELECT DISTINCT
        comm_code,
        comm_name,
        MEDIAN(re_assessed_value) OVER (PARTITION BY comm_code) AS median_assessed_value
    FROM filtered_Current_Property_Assessments
    WHERE roll_year = 2025
        AND assessment_class = 'RE'
        AND re_assessed_value IS NOT NULL
        AND land_use_designation NOT LIKE 'DC%'
        AND land_use_designation NOT LIKE 'C%'
        AND land_use_designation NOT LIKE 'S%'
        AND land_use_designation NOT LIKE 'I%'
),
assessment_bands AS (
    SELECT
        comm_code,
        comm_name,
        median_assessed_value,
        NTILE(4) OVER (ORDER BY median_assessed_value) AS quartile
    FROM community_medians
),
crime_totals AS (
    SELECT cp.comm_code, SUM(fc.crime_count) AS total_crime_2024
    FROM filtered_Community_Crime fc
    INNER JOIN Community_Points cp
        ON fc.community = cp.name
    WHERE fc.year = 2024
    GROUP BY cp.comm_code
),
service_counts AS (
    SELECT
        s.comm_code,
        COUNT(*) AS service_count
    FROM Community_Service s
    GROUP BY s.comm_code
),
combined AS (
    SELECT
        ab.comm_code,
        ab.comm_name,
        ab.median_assessed_value,
        CASE
            WHEN ab.quartile = 1 THEN 'Low'
            WHEN ab.quartile IN (2,3) THEN 'Medium'
            ELSE 'High'
        END AS assessment_class,
```

```
        COALESCE(sc.service_count, 0) AS service_count,
        RANK() OVER (ORDER BY COALESCE(sc.service_count, 0) DESC) AS service_rank,
        COALESCE(ct.total_crime_2024, 0) AS total_crime_2024,
        RANK() OVER (ORDER BY COALESCE(ct.total_crime_2024, 0) ASC) AS safety_rank,
        RANK() OVER (ORDER BY ab.median_assessed_value ASC) AS affordability_rank
    FROM assessment_bands ab
    LEFT JOIN service_counts sc
        ON ab.comm_code = sc.comm_code
    LEFT JOIN crime_totals ct
        ON ab.comm_code = ct.comm_code
)
SELECT
    comm_code,
    comm_name,
    median_assessed_value,
    assessment_class,
    service_count,
    service_rank,
    total_crime_2024,
    safety_rank,
    affordability_rank,
    (service_rank + safety_rank + affordability_rank) AS overall_score,
    RANK() OVER (ORDER BY (service_rank + safety_rank + affordability_rank)) AS overall_rank
FROM combined
ORDER BY overall_rank;
```

Appendix J - Python Notebook of Data Cleaning & App Token

Please refer to Data-604-AppendixJ-DataCleaningImport.zip which contains Appendix J which is the jupyter notebook used to pull data from the City of Calgary server, clean the data in pandas and load the data into SQL environment. The zip file also contains an Calgary_App_Token which is required to connect to the SODA api.

Appendix K – Overall Comparison Table CSV Extract

Please refer to Appendix_K_Overall_Comparison.csv. The report contained an abbreviated version of the table of final rankings of Calgary communities (Best, and Worst). The overall ranking of every community is listed in the csv file.

Appendix L - R-Markdown File of Linear Regression Model

Please refer to Data-604-Appendix-L-Regression_Results.pdf for the r-markdown file of the linear regression which was performed. The regression was performed using the dataset from Appendix K.