Calgary Communities: Where would you want to live?

DATA 604 – Fall 2025: Thanusha Bala, Elise Beaupre

Introduction: Relevance & Background

Real Estate in Calgary

Calgary is one of the fastest growing cities in Canada. Based on Stats Canada's 2025 estimate of population growth Calgary is the metropolitan area with the largest growth rate (5.97%)

One of the reasons for the large influx is Calgary's relative affordability in the real estate and rental markets (in comparison to Toronto, Montreal and Vancouver)

Relevance

When you're choosing to buy a house in Calgary how do you decide which community to live in? What are your key concerns? Which communities are safe, have access to amenities and are affordable?

This topic is relevant because for most people their home is their largest tangible asset. Home ownership is considered a life "benchmark" and is associated with financial well-being.

Topics of Investigation

Where are the neighborhoods with the lowest crime?

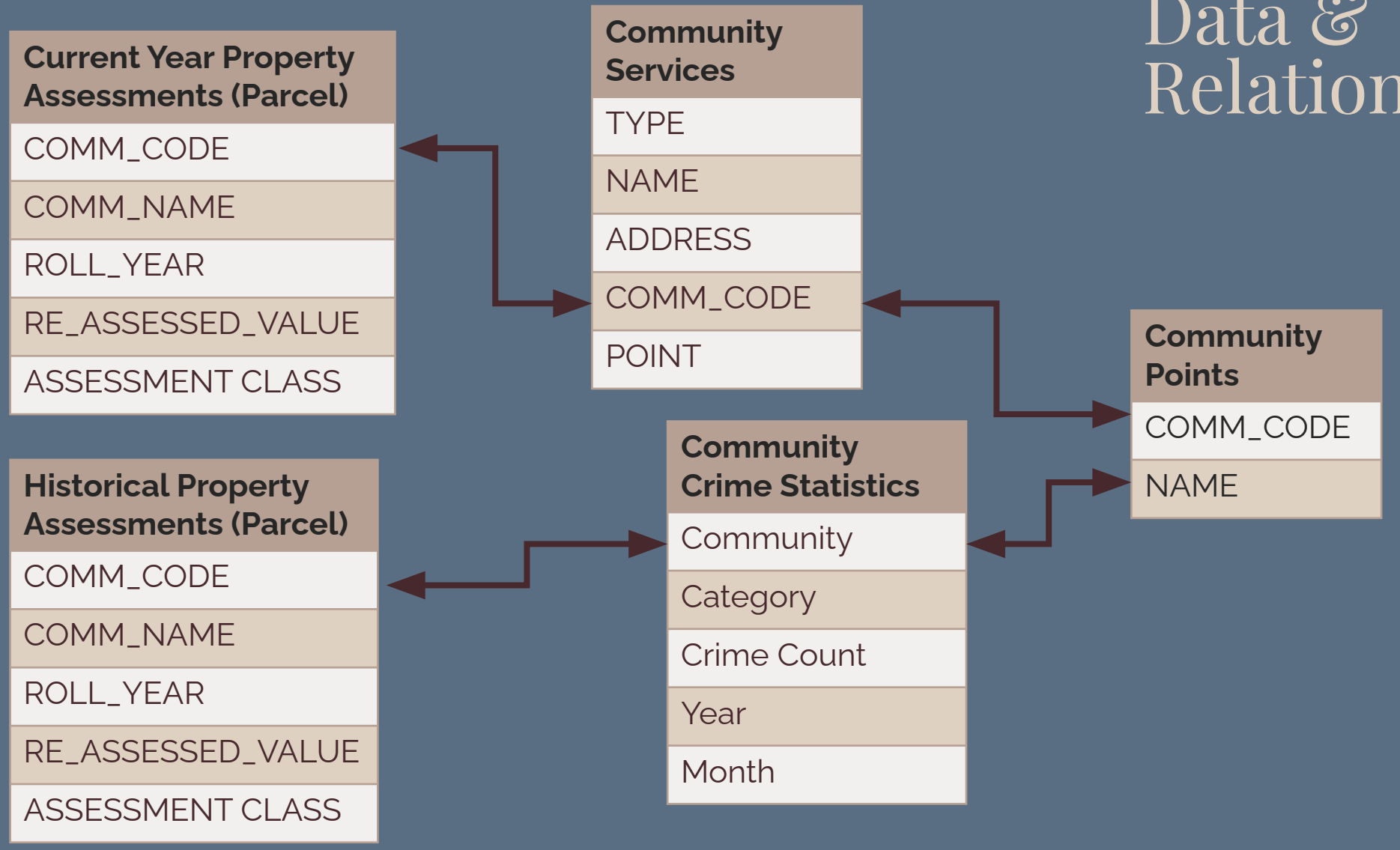
Where are the properties are the most affordable?

At each budget level, where is the best community?

Which communities have had the largest increase in property value since 2024?

Which communities have the most amenities?

Data & Relationships



Data Cleaning & Technology

Step	Details
Step 1: Load Data Using SODA	The Socrata Open Data API (SODA) is designed for accessing and interacting with datasets hosted on Socrata platforms, which are commonly used by government agencies and organizations for open data initiatives.
Step 2: Cleaning the Data for Import to SQL	When we connected to SODA using pandas, some of the data was not pulled in the correct form (ie. pulled as a Python object) so it had to be converted.
Step 3: Use SQLAlchemy to run queries for data analysis	We ran our queries via SQL Alchemy. This was challenging as some queries took longer to run than on SQL Server.

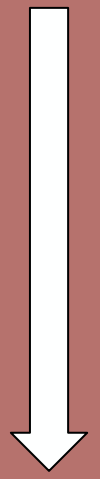
HOUSING AFFORDABILITY

Which Communities Are Most and Least Affordable in 2025?

```
pd.read_sql_query('''
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
)
SELECT
    comm_name,
    median_assessed_value
FROM community_medians
ORDER BY median_assessed_value
;''', engine)
```

	comm_name	median_assessed_value
0	RICARDO RANCH	73000.0
1	LEWISBURG	146750.0
2	KEYSTONE HILLS	170000.0
3	HOTCHKISS	194750.0
4	LINCOLN PARK	222000.0
...	...	...
199	EAGLE RIDGE	1985000.0
200	BRITANNIA	2050000.0
201	BEL-AIRE	2480000.0
202	MEDICINE HILL	8020000.0
203	QUEENS PARK VILLAGE	42860000.0

Least expensive



Most expensive

Which Communities Had the Greatest Changes in 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 AS (
    SELECT
        comm_code,
        comm_name,
                SUBSTRING_INDEX(
            SUBSTRING_INDEX(
                GROUP_CONCAT(re_assessed_value ORDER BY re_assessed_value),
                ',',
                FLOOR(COUNT(*)/2)
            ),
            ',',
            -1
        ) AS median_2025
    FROM filtered_Current_Property_Assessments
    WHERE roll_year = 2025
        AND assessment_class = 'RE'
        AND re_assessed_value IS NOT NULL
    GROUP BY comm_code, comm_name
)
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
WHERE c.median_2025 IS NOT NULL
ORDER BY pct_change DESC;
```

	comm_code	comm_name	median_2024	median_2025	value_change	pct_change
0	HSN	HASKAYNE	405000.0	710500	305500.0	75.43
1	HKS	HOTCHKISS	127000.0	194500	67500.0	53.15
2	COA	COACH HILL	372500.0	535500	163000.0	43.76
3	GRE	GREENWOOD/GREENBRIAR	319250.0	458500	139250.0	43.62
4	SKR	SKYVIEW RANCH	254500.0	347500	93000.0	36.54

	comm_code	comm_name	median_2024	median_2025	value_change	pct_change
194	LIV	LIVINGSTON	603500.0	642000	38500.0	6.38
195	WWO	WOLF WILLOW	522000.0	522500	500.0	0.10
196	CSC	CITYSCAPE	702500.0	681000	-21500.0	-3.06
197	HSD	HOMESTEAD	608500.0	576500	-32000.0	-5.26
198	LEB	LEWISBURG	1296376.0	146500	-1149876.0	-88.70

COMMUNITY SERVICES / AMENITIES

What types of Community Services Are There?

	type	service_count
0	Community Centre	107
1	Attraction	38
2	Library	21
3	Court	9
4	PHS Clinic	8
5	Social Dev Ctr	8
6	Visitor Info	5
7	Hospital	5

```
query1 = """
SELECT type, COUNT(name) AS service_count
FROM Community_Service
WHERE comm_code IS NOT NULL
GROUP BY type
ORDER BY service_count DESC;
"""

df1 = pd.read_sql_query(query1, engine)
df1
```

Which Communities Have the Most Services?

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

```
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;
```

Which Communities Have the Least Services

```
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;
```

	comm_code	community_name	service_count	service_rank
0	EVN	EVANSTON	0	122
1	BLM	BELMONT	0	122
2	ROY	ROYAL OAK	0	122
3	CUR	CURRIE BARRACKS	0	122
4	HKS	HOTCHKISS	0	122
...	...	...	...	...
95	COA	COACH HILL	0	122
96	SAD	SADDLE RIDGE	0	122
97	MAC	MACEWAN GLEN	0	122
98	LIV	LIVINGSTON	0	122
99	HUX	HUXLEY	0	122

Are the Communities with the Most Amenities the Most Expensive?

	comm_code	community_name	service_count	service_rank	median_re_assessed_value
0	FLN	FOREST LAWN	5	3	467000.0
1	HUN	HUNTINGTON HILLS	4	4	578000.0
2	BRD	BRIDGELAND/RIVERSIDE	3	6	499000.0
3	HOU	HOUNSFIELD HEIGHTS/BRIAR HILL	3	6	903000.0
4	ING	INGLEWOOD	3	6	607000.0

```
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
      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%'
) AS q5
ON q4.comm_code = q5.comm_code
ORDER BY q4.service_rank
LIMIT 5;
```

```
WITH ranked AS (
    SELECT
        comm_code,
        community_name,
        service_count,
        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,
    median_re_assessed_value,
    property_value_class
FROM ranked
WHERE rank_within_class <= 5 AND service_count >= 2
ORDER BY property_value_class, service_count DESC;
```

comm_code	community_name	service_count	median_re_assessed_value	property_value_class
HOU	HOUNSFIELD HEIGHTS/BRIAR HILL	3	903000	High
HIL	HILLHURST	2	880250	High
MOP	MOUNT PLEASANT	2	781750	High
STA	ST. ANDREWS HEIGHTS	2	1000000	High
FLN	FOREST LAWN	5	467000	Low
BRD	BRIDGELAND/RIVERSIDE	3	499000	Low
PIN	PINERIDGE	2	481000	Low
SET	SETON	2	461250	Low
SNA	SUNALTA	2	253000	Low
HUN	HUNTINGTON HILLS	4	578000	Medium
ING	INGLEWOOD	3	607000	Medium
ACA	ACADIA	2	595000	Medium
ARB	ARBOUR LAKE	2	683000	Medium
BOW	BOWNESS	2	541500	Medium

COMMUNITY CRIME STATISTICS

Which Communities Had the Highest Crime Rates in 2024?

```
pd.read_sql_query('''
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;
''', engine)
```

Communities with the MOST crimes in 2024

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

Communities with the FEWEST crimes in 2024

Which Communities Had the Highest Rates of Violent Crimes in 2024?

```
pd.read_sql_query('''
    SELECT
        community,
        SUM(crime_count) as violent_crime_2024
    FROM filtered_Community_Crime
    WHERE year = 2024 AND category IN (
        'Assault (Non-domestic)',
        'Violence  \\'Other\\' (Non-domestic)',
        'Street Robbery', 'Commercial Robbery'
        )
    GROUP BY community
    ORDER BY violent_crime_2024 DESC;
''', engine)
```

Communities with the highest NUMBER of VIOLENT crimes in 2024

	community	violent_crime_2024	total_crime_2024
0	BELTLINE	472.0	1147.0
1	DOWNTOWN COMMERCIAL CORE	389.0	760.0
2	DOWNTOWN EAST VILLAGE	255.0	385.0
3	FOREST LAWN	151.0	410.0
4	MARLBOROUGH	136.0	309.0

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

Communities with the highest PERCENTAGE of VIOLENT crimes in 2024

Which Communities' Crime Rates Changed the Most from 2018 to 2024?

```
pd.read_sql_query('''
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;
''', engine)
```

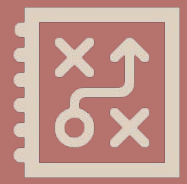
Communities with the greatest INCREASES in crime from 2018 to 2024

	community	crime_2018	crime_2024	crime_change
0	DOWNTOWN EAST VILLAGE	133.0	385.0	252.0
1	SETON	48.0	179.0	131.0
2	CORNERSTONE	17.0	133.0	116.0

Communities with the greatest DECREASES in crime from 2018 to 2024

213	BOWNESS	494.0	240.0	-254.0
214	DOWNTOWN COMMERCIAL CORE	1149.0	760.0	-389.0
215	BELTLINE	1729.0	1147.0	-582.0

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



```
df_crime_extremes = pd.read_sql_query("""
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;
```

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

Do the most expensive communities have the least crime?

```
pd.read_sql_query('''
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 DESC) AS value_rank,
        RANK() OVER (ORDER BY ct.total_crime_2024 ASC) 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,
    value_rank,
    crime_rank
FROM joined
WHERE value_rank <= 5
ORDER BY value_rank;
''', engine)
```

	community	median_assessed_value	total_crime_2024	value_rank	crime_rank
0	QUEENS PARK VILLAGE	42860000.0	7.0	1	10
1	MEDICINE HILL	8020000.0	10.0	2	15
2	BEL-AIRE	2480000.0	5.0	3	8
3	BRITANNIA	2050000.0	12.0	4	20
4	EAGLE RIDGE	1985000.0	50.0	5	87

The top 5 most expensive communities do not rank in the top 5 fewest crimes

CHALLENGES & CONCLUSIONS

CHALLENGES

- Change in scope regarding CMHC data & property taxes: we had originally planned on incorporating property taxes, but property taxes are a percentage based on assessed home value
- Working with the SODA API was initially difficult, especially as we couldn’t access it via the virtual machine
- It was hard to isolate for residential property only
- Due to the size of property assessments table the scripts took 15+ minute to run

CONCLUSIONS & FUTURE DIRECTION

- Property value does not have a linear straightforward relationship with amenities or crime
- Median home value in Calgary is 640,750
- There is an overlap in communities with the most crime and the most amenities, in the future we will be exploring the intersection of all 3 factors (property value, crime, and amenities)

References

Calgary's population growth is in a league of its own. (2025, October 12). *The Globe and Mail.* https://www.theglobeandmail.com/business/article-calgarys-population-growth-is-in-a-league-of-its-own/

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

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

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

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

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