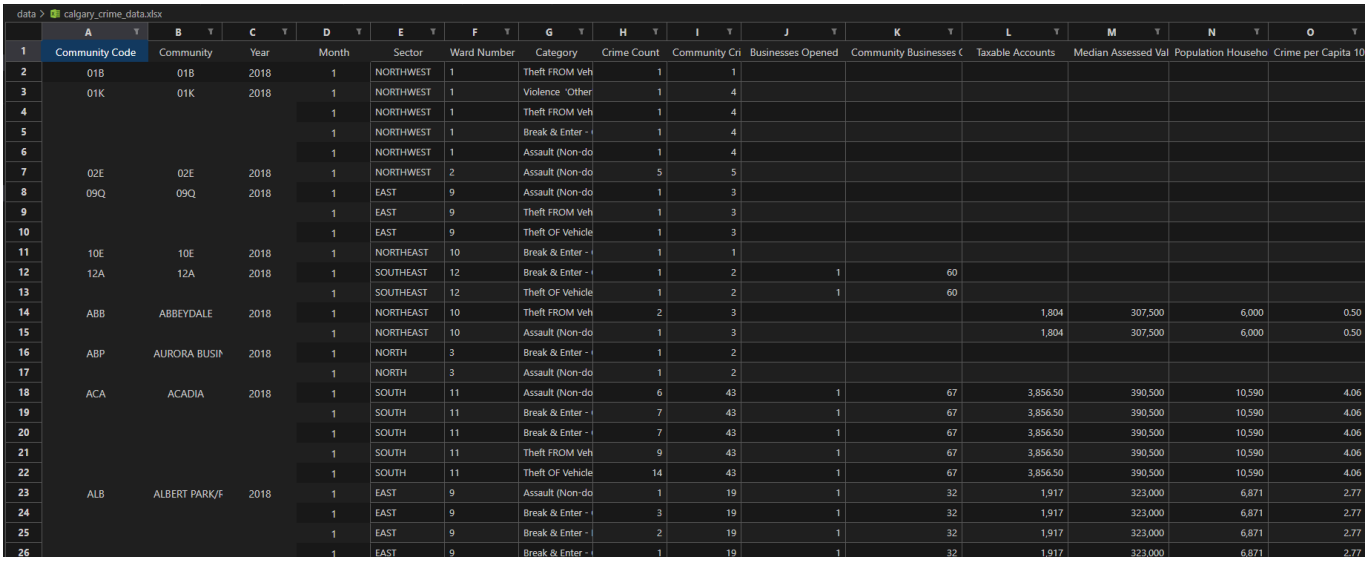
Report

ENSF 692 Spring 2025 - Final Project: Calgary Community Crime Statistics Dataset Analysis

Authors: Jason Chiu and Jason Tieh
Group: 1

Dataset Description

The program primarily uses the Calgary **Community Crime Statistics** Dataset [1], which provides detailed monthly crime statistics from 2018 to 2024. The dataset includes information such as crime count, crime category, year, month, and location by community. The final dataset when merged includes columns such as business count, median assessed value, sector, wards, population household; with additional columns created that include Crime per Capita 1000 or Community Crime MTD Total. The final multi-indexed dataframe is exported as an excel file and be be found under /data/calgary_crime_data.xlsx. A screenshot of the final dataframe is shown below:



data > calgary_crime_data.xlsx

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Community Code	Community	Year	Month	Sector	Ward Number	Category	Crime Count	Community Cri	Businesses Opened	Community Businesses (	Taxable Accounts	Median Assessed Val	Population Househo	Crime per Capita 10
2	01B	01B	2018	1	NORTHWEST	1	Theft FROM Veh	1	1						
3	01K	01K	2018	1	NORTHWEST	1	Violence 'Other	1	4						
4				1	NORTHWEST	1	Theft FROM Veh	1	4						
5				1	NORTHWEST	1	Break & Enter -	1	4						
6				1	NORTHWEST	1	Assault (Non-do	1	4						
7	02E	02E	2018	1	NORTHWEST	2	Assault (Non-do	5	5						
8	09Q	09Q	2018	1	EAST	9	Assault (Non-do	1	3						
9				1	EAST	9	Theft FROM Veh	1	3						
10				1	EAST	9	Theft OF Vehicle	1	3						
11	10E	10E	2018	1	NORTHEAST	10	Break & Enter -	1	1						
12	12A	12A	2018	1	SOUTHEAST	12	Break & Enter -	1	2	1	60				
13				1	SOUTHEAST	12	Theft OF Vehicle	1	2	1	60				
14	ABB	ABBEYDALE	2018	1	NORTHEAST	10	Theft FROM Veh	2	3			1,804	307,500	6,000	0.50
15				1	NORTHEAST	10	Assault (Non-do	1	3			1,804	307,500	6,000	0.50
16	ABP	AURORA BUSIN	2018	1	NORTH	3	Break & Enter -	1	2						
17				1	NORTH	3	Assault (Non-do	1	2						
18	ACA	ACADIA	2018	1	SOUTH	11	Assault (Non-do	6	43	1	67	3,856.50	390,500	10,590	4.06
19				1	SOUTH	11	Break & Enter -	7	43	1	67	3,856.50	390,500	10,590	4.06
20				1	SOUTH	11	Break & Enter -	7	43	1	67	3,856.50	390,500	10,590	4.06
21				1	SOUTH	11	Theft FROM Veh	9	43	1	67	3,856.50	390,500	10,590	4.06
22				1	SOUTH	11	Theft OF Vehicle	14	43	1	67	3,856.50	390,500	10,590	4.06
23	ALB	ALBERT PARK/F	2018	1	EAST	9	Assault (Non-do	1	19	1	32	1,917	323,000	6,871	2.77
24				1	EAST	9	Break & Enter -	3	19	1	32	1,917	323,000	6,871	2.77
25				1	EAST	9	Break & Enter -	2	19	1	32	1,917	323,000	6,871	2.77
26				1	EAST	9	Break & Enter -	1	19	1	32	1,917	323,000	6,871	2.77

User Interface Input and Output Summary

Input

Users interact with the program through a command-line interface (CLI) where they are prompted to:

1. Choose whether to view reference maps (sector, ward, community) [9][10].
2. Select the type of region: Sector, Ward, or Community.
3. Choose whether to view a reference table with available locations to choose from in the chosen region
4. Specify the exact name of the Sector, Ward, or Community.
5. Specify the year.

The user must enter a Sector, Ward, Community, and year that exists in the dataset. If there is an invalid input, the user is prompted for re-entry without terminating the program. The program is also designed to handle case-sensitive inputs or extra spaces.

After printing out a general stats table for the chosen region and year, plots are generated and further prompting appears:

6. Choose whether or not to save the generated plot as a png into the /images folder repeating for all 4 plots
7. Choose whether to re-run the visualizer with different chosen location and year repeating steps 1-6
8. Choose whether to export dataframe as an excel file for separate viewing thereafter terminating the program

Output

The program returns general statistics and data plots for the selected region and year, including:

- A text summary of basic descriptive statistics for the chosen region and year and how they compare to other locations of the same location type
- A bar chart of **crime count by category**. (A general line plot across all years in Calgary is shown for comparison) along with a corresponding pivot table of the same data but by month printed into the console
- A bar chart of **monthly crime counts**. (A general line plot across all years in Calgary is shown for comparison)
- A scatter plot comparing **crime per capita vs median assessed value** with the location highlighted for better viewing.
- A scatter plot comparing **total crime count vs number of businesses** with the location highlighted for better viewing.

The user is also prompted whether to analyze another region or quit. Upon quitting, the program thanks the user.

Specification Compliance

Data Handling

More than 3 seperate datasets are merged into a large dataset. All data is pulled directly from the .csv file without modification. The crime dataset is merged with:

- The *Communities by Wards* dataset [2] to retrieve the wards and sector column in the dataset
- The *Calgary Business Licences* [3] dataset to retrieve a count of the number of businesses in a community
- The *Assessments by Community* [4] dataset to retrieve the median assessed value of a community
- and The 2016, 2017, 2019, and 2021 *Census by Community* and *Federal Census Population by Community* datasets [5],[6],[7],[8] are merged and extrapolated to get the population of each community for a specified year.

Code Implementation

The program meets this requirement because:

1. A print_describe() function uses the describe() method to give statistics (count, mean, min, max, std. deviation, etc.) about the crime count, Crime per Capita 1000, Businesses Opened, and Median Assess Value

2. We added two or more columns to the main dataframe. We calculated the Crime per Capita (per 1000 residents) and summed the Business count for the community.
3. A masking operation, groupby operation, pivot table, and aggregation computation was used for a subset of the data when creating plots. For example: A masking operation is used based on the location and year, a group by operation and aggregation computation was used to get the total crime count for each month.

```
    subset = final_df[(final_df[location_type] == location) & (final_df['Year'] ==
year)]
    monthly_crime = subset.groupby('Month')['Crime Count'].sum().reset_index()
```

A Pivot table was also created to easily plot the total crime count per month for each year

```
    crime_month = final_df.groupby(['Year', 'Month'])['Crime
Count'].sum().reset_index()
    pivot_table = crime_month.pivot(index='Month', columns='Year', values='Crime
Count')
```

User Interface and Execution

The program meets this requirement because the user enters two required input values. The region and the year. Invalid inputs are handled and prompted for re-entry. Screenshots are provided for the expected execution, including handling of incorrect input:

```
Starting up Calgary Crime Statistics Visualizer...
Creating dataframe...

 --------- Start Calgary Crime Statistics Visualizer ---------

Welcome to Calgary Crime Statistic Visualizer!

In this program, you will be able to view Calgary's basic crime statistics from the years 2018 to 2024
Select a year and location/region within Calgary, and you will see data related to the input, including:
type and quantity of crime, crime per month, crime per capita 1000 vs median assessed value of the region,
and crime count vs the number of existing businesses of the region.

Based On current entire existing dataset, the following values have also been observed:

================ Overall Stats of the dataset oganized by Months ==================
================================================================================
       Crime Count  Crime per Capita 1000  Businesses Opened Median Assessed Value
count    84.000000              83.000000          84.000000                195.00
mean   2580.833333               2.465833         132.190476            496,711.97
std     576.279254               0.460747          62.728613            232,877.11
min       0.000000               1.711883          55.000000             42,000.00
25%    2228.750000               2.130566          87.000000            378,000.00
50%    2604.500000               2.444332         110.000000            451,000.00
75%    2950.000000               2.723601         154.250000            555,250.00
max    3899.000000               3.971015         313.000000          1,675,000.00
================================================================================


To begin the visualizer, first select the region type.

If you would like a map to see what sectors, wards, and communities you may see the map png files
found in the /data folder or if you wish to see the images from here, enter (Y/N) (Note: this process is slow): Y

Loading images. This may take a few seconds...

Data can be analyzed in either City Sectors, City Wards, or City Communities.
Type 'Sector' 'Ward' or 'Community' for the type of location you wish to analyze data on: region

region is not a valid location. Please try again.

Type 'Sector' 'Ward' or 'Community' for the type of location you wish to analyze data on: cOmmUniTy
Would you like to view a list of valid Communities? (Y/N): Y
Please enter a community by name or 3 character code to analyze data on: BELTLINE
Please enter the year of data to analyze (2018-2024): 2015

This year was not found in the data. Please try again.

Please enter the year of data to analyze (2018-2024): 2019

You have chosen the following region and year |  Community :  BELTLINE  for the year  2019

Based on these chosen fields, the following statistics can be seen:


Summary for Community: BELTLINE (2019)
------------------------------------------------------------
Total Population             : 25,129.0 (Avg: 6,336) | Rank: 2/202
Median Assessed Value        : 313,500.0 (Avg: 496,712) | Rank: 175/195
Number of Businesses         : 416.0 (Avg: 60) | Rank: 3/173
Total Crime Incidents        : 2,268.0 (Avg: 139) | Rank: 1/275
Crime per 1,000 residents    : 90.25 (Avg: inf) | Rank: 8/202
Business Density (/1000)     : 16.55 (Avg: 13) | Rank: 16/125
------------------------------------------------------------
*Note: The larger the value, the higher the rank
**Notes: If the location type is a sector or ward, the Median Assessed Value
         is an average of the communities within these regions
```

```
Summary for Community: BELTLINE (2019)
------------------------------------------------------------
Total Population              : 25,129.0 (Avg: 6,336) | Rank: 2/202
Median Assessed Value         : 313,500.0 (Avg: 496,712) | Rank: 175/195
Number of Businesses          : 416.0 (Avg: 60) | Rank: 3/173
Total Crime Incidents         : 2,268.0 (Avg: 139) | Rank: 1/275
Crime per 1,000 residents     : 90.25 (Avg: inf) | Rank: 8/202
Business Density (/1000)      : 16.55 (Avg: 13) | Rank: 16/125
------------------------------------------------------------
*Note: The larger the value, the higher the rank
**Notes: If the location type is a sector or ward, the Median Assessed Value
         is an average of the communities within these regions

Here is a plot showing the crime category and their total count for the chosen year and location:
Below is a pivot table for the same data but separated by month

Monthly Crime Count Table for Community BELTLINE (2019):
Category      1      2      3      4      5     6     7     8     9
1.0        39.0   58.0   24.0    6.0   16.0   6.0   1.0   0.0   2.0
2.0        38.0   69.0   20.0   14.0    6.0   0.0   2.0   1.0   1.0
3.0        76.0   46.0   29.0    6.0    8.0   4.0   3.0   0.0   0.0
4.0        74.0   40.0   31.0    8.0    4.0   3.0   4.0   1.0   1.0
5.0        69.0   57.0   27.0   17.0   15.0   4.0   2.0   2.0   3.0
6.0        70.0   32.0   27.0   12.0    8.0   1.0   0.0   1.0   2.0
7.0        90.0   44.0   45.0   21.0   11.0   8.0   7.0   5.0   1.0
8.0        70.0   48.0   30.0   16.0    9.0   1.0   1.0   3.0   1.0
9.0        98.0   61.0   37.0   17.0    3.0   4.0   6.0   3.0   0.0
10.0       96.0   79.0   46.0    8.0    8.0   3.0   2.0   1.0   2.0
11.0      111.0   42.0   39.0   16.0   11.0   4.0   5.0   4.0   3.0
12.0       75.0   33.0   22.0    8.0    8.0   7.0   3.0   1.0   1.0
Total     906.0  609.0  377.0  149.0  107.0  45.0  36.0  22.0  17.0

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

Would you like to save this plot as a png? (stored in /images) (Y/N): N

Here is a plot comparing the amount of crime per month for the chosen year and location:

Would you like to save this plot as a png? (stored in /images) (Y/N): N

Here is a plot comparing Crime per capita 1000 vs a locations communities, median assessed value:
Note: If the point is not highlighted, there is no information on median assessed value for this location

Would you like to save this plot as a png? (stored in /images) (Y/N): N

Here is a plot comparing Crime Count vs a locations communities business count to date total:
Note: If the point is not highlighted, there is no information on business count to date for this location

Would you like to save this plot as a png? (stored in /images) (Y/N): N

Would you like to visualize data for another location and/or time? Hit 'ENTER' to continue, otherwise enter 'Q' to quit: q

Would you like to export the indexed crime dataframe to an excel? (Y/N) (Note this may take over 10 minutes): N

Thank you for Using the Calgary Crime Statistics Visualizer.
------------------------------------------------------------------------------------
```

Plots are generated that depicts an aspect of the data. Screenshots of plots can be found in /screenshots. There are a total of four Plots, two examples are shown below:

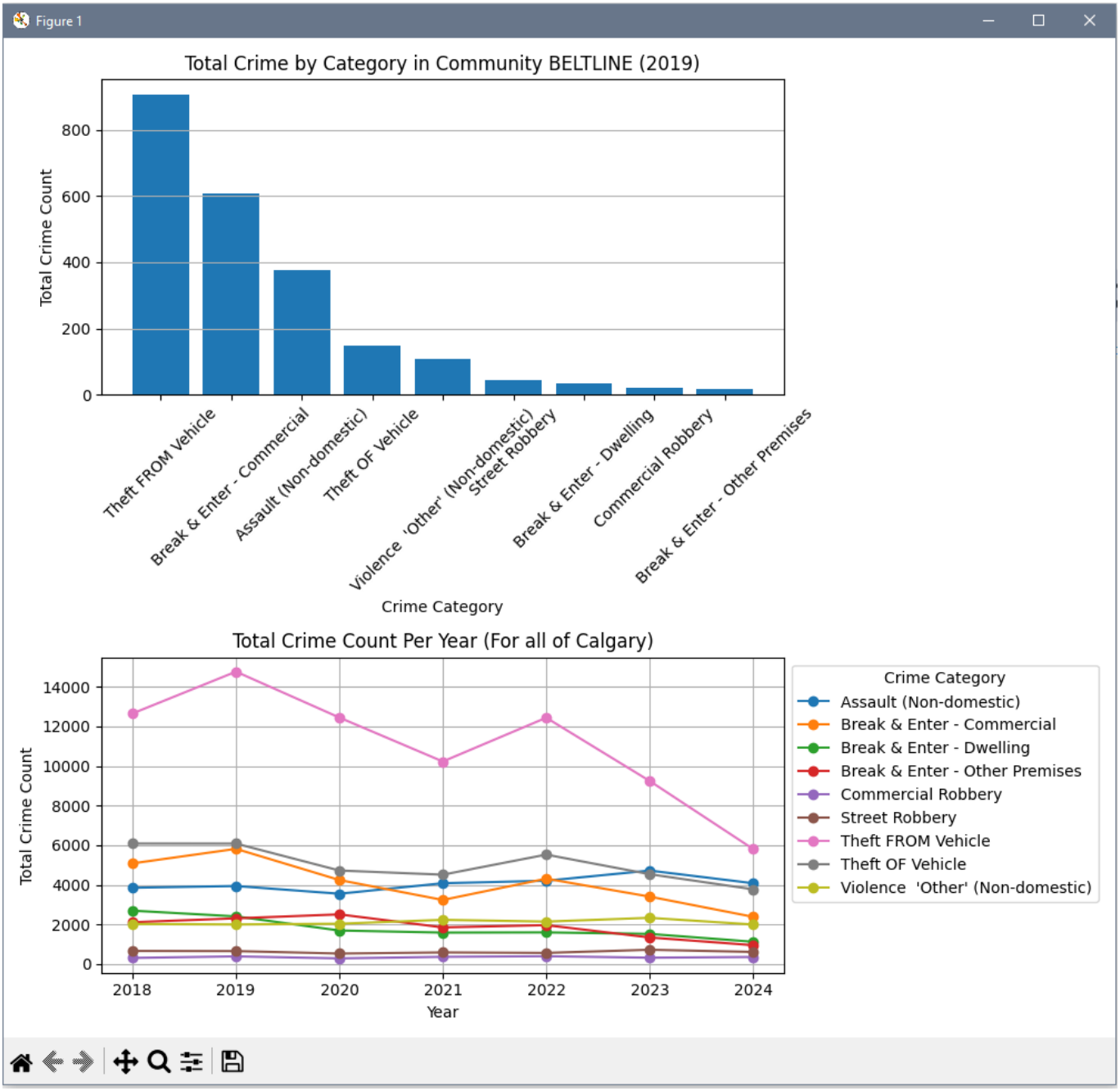
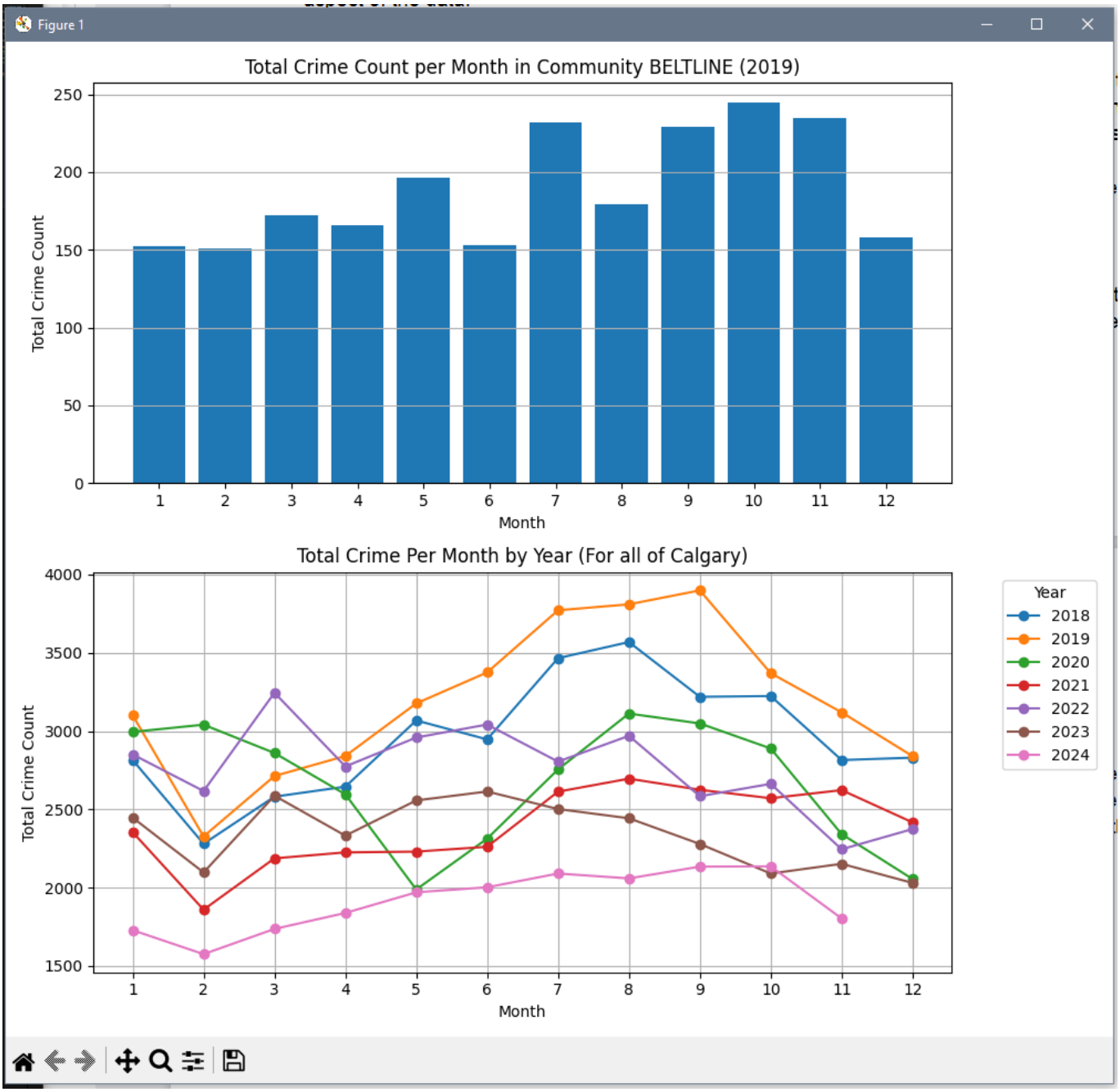


Figure 1
Total Crime by Category in Community BELTLINE (2019)
Total Crime Count
0
200
400
600
800
Theft FROM Vehicle
Break & Enter - Commercial
Assault (Non-domestic)
Theft OF Vehicle
Violence 'Other' (Non-domestic)
Street Robbery
Break & Enter - Dwelling
Commercial Robbery
Break & Enter - Other Premises
Crime Category
Total Crime Count Per Year (For all of Calgary)
Total Crime Count
0
2000
4000
6000
8000
10000
12000
14000
2018
2019
2020
2021
2022
2023
2024
Year
Crime Category
Assault (Non-domestic)
Break & Enter - Commercial
Break & Enter - Dwelling
Break & Enter - Other Premises
Commercial Robbery
Street Robbery
Theft FROM Vehicle
Theft OF Vehicle
Violence 'Other' (Non-domestic)



Commenting and Syntax

The program includes comments throughout the code for clarity. Classes, methods, and functions are documented using docstrings. variables and functions are lowercase and seperated by underscore (e.g. get_location(), get_year()).

Citation:

[1] The City of Calgary. Community Crime Statistics [Data set]. Calgary Open Data Portal. 2025. Available: https://data.calgary.ca/Health-and-Safety/Community-Crime-Statistics/78gh-n26t/about_data

[2] The City of Calgary. Communities by Ward [Data set]. Calgary Open Data Portal. 2025. Available: https://data.calgary.ca/Government/Communities-by-Ward/jd78-wxjp/about_data

[3] The City of Calgary. Calgary Business Licences [Data set]. Calgary Open Data Portal. 2025. Available: https://data.calgary.ca/Business-and-Economic-Activity/Calgary-Business-Licences/vdjc-pybd/about_data

[4] The City of Calgary. Assessments by Community [Data set]. Calgary Open Data Portal. 2025. Available: https://data.calgary.ca/Government/Assessments-by-Community/p84b-7zbi/about_data

[5] The City of Calgary. Census by Community 2016 [Data set]. Calgary Open Data Portal. 2025. Available: https://data.calgary.ca/Demographics/Census-by-Community-2016/hfwb-eab8/about_data

[6] The City of Calgary. Census by Community 2017 [Data set]. Calgary Open Data Portal. 2025. Available: https://data.calgary.ca/Demographics/Census-by-Community-2017/d8k6-ide9/about_data

[7] The City of Calgary. Census by Community 2019 [Data set]. Calgary Open Data Portal. 2025. Available: https://data.calgary.ca/browse?q=Census+by+Community+2019&sortBy=relevance&pageSize=20

[8] The City of Calgary. 2021 Federal Census Population and Dwellings by Community [Data set]. Calgary Open Data Portal. 2025. Available: https://data.calgary.ca/Demographics/2021-Federal-Census-Population-and-Dwellings-by-Co/f9wk-wej9/about_data

[9] Calgary Real Estate Board, City of Calgary Community and Sector Map 2025, CREB, Calgary, AB, Canada. [Online]. Available: https://www.creb.com/-/media/Public/CREBcom/Housing_Statistics/CITY_OF_CALGARY_MAP_2025.pdf. Accessed: Jun. 19, 2025.

[10] City of Calgary, City-wide Ward and Community Codes Map (42 × 56 inches), Election and Census Office, Calgary, AB, Canada. [Online]. Available: https://www.calgary.ca/content/dam/www/election/documents/ward-maps/city-wide-ward-map-poster-42x56-inches.pdf. Accessed: Jun. 19, 2025.