

Calgary Crime Analysis Report

Prepared by:
Yash Rajesh Patel

Southern Alberta Institute of Technology (SAIT), Calgary

Table of Contents

1. Introduction
2. Project Objectives
3. Dataset Source and Rationale
4. About the Dataset
5. Key Questions on the Data
6. Data Preparation and Cleaning
7. Data Modeling and Power BI Integration
8. Dashboard Design and Features
9. Key Insights and Findings
10. Challenges and Limitations
11. Future Work and Recommendations
12. Conclusion
13. References

➤ Introduction

The **Calgary Crime Analysis Project** was developed to explore, analyze, and visualize crime trends and patterns across various communities within the City of Calgary, Alberta. As one of Canada's largest and fastest-growing urban centers, Calgary faces challenges in maintaining community safety and equitable resource allocation. This project aims to address these challenges by applying **data analytics and visualization techniques** to identify significant insights that can aid **law enforcement, city planners, and policymakers** in making informed, data-driven decisions.

The analysis uses crime data obtained from the **City of Calgary's Open Data Portal**, covering the years **2018 to 2023**. This publicly available dataset provides comprehensive information about crime incidents, including their type, location, and time of occurrence. Such open data fosters transparency while allowing researchers and analysts to examine social and geographic patterns of criminal activity. By utilizing this dataset, the project contributes to the broader goal of improving public safety through evidence-based approaches.

To transform raw data into actionable insights, the project employs **Power BI**, a modern business intelligence platform that enables interactive and visual analysis. Power BI's dashboards and reports allow users to easily identify trends, compare communities, and observe how crime evolves over time. This interactivity ensures that even complex datasets can be interpreted intuitively, bridging the gap between technical analysis and real-world understanding.

➤ Project Objectives

The **Calgary Crime Analysis Project** was developed with the following key objectives:

- **Develop interactive dashboards:**

Create visually engaging and interactive Power BI dashboards to help users explore and understand Calgary's crime data more effectively.

- **Identify temporal and spatial trends:**

Analyze how crime patterns change over time and across different communities or sectors within Calgary.

- **Analyze crime rates relative to population:**

Compare crime incidents against community population sizes to identify areas with disproportionately high crime rates.

- **Support data-driven decision-making:**

Provide insights that assist law enforcement agencies and policymakers in allocating resources, improving safety strategies, and making informed decisions.

➤ Dataset Source and Rationale

The dataset was obtained from Kaggle, which hosts data provided by the City of Calgary Open Data Portal. It includes detailed records of crimes reported between 2018 and 2023. The dataset is publicly accessible, reliable, and ideal for community-level analysis.

- Source: Kaggle - Calgary Crime Statistics
- Provider: City of Calgary Open Data Portal
- Timeframe: 2018–2023
- URL: <https://www.kaggle.com/datasets/tcashion/calgary-crime-statistics>

➤ About the Dataset

The dataset used for this project contains **66,349 unique records** of reported crimes within the **City of Calgary**. It provides detailed information that supports both temporal (time-based) and spatial (location-based) analysis of crime patterns. The dataset's structure enables easy filtering, comparison, and visualization in Power BI.

- **Sector:** Represents the broader **geographical division** of Calgary, such as Northwest, Northeast, Southwest, and Southeast. It helps in identifying which parts of the city experience higher crime rates.
- **Community Name:** Refers to the **specific local community** where each crime occurred. This allows for a more detailed, community-level analysis of crime trends.
- **Category:** Indicates the **type of crime** reported, such as Assault, Theft, or Break and Enter. This helps classify incidents and identify the most common types of offenses.
- **Crime Count:** Shows the **number of incidents reported** within a given time frame for each community or category. It is a key measure for comparing crime intensity across areas.
- **Resident Count:** Provides the **population size** of each community, which is essential for calculating crime rates per 1,000 residents and ensuring fair comparisons.
- **Date, Year, Month:** Captures **when each crime occurred**, enabling time-based trend analysis and seasonal pattern identification.
- **Community Center Point:** Contains **geographic coordinates (latitude and longitude)** for mapping and spatial visualization in Power BI dashboards.

Yash Patel

Overall, this dataset offers a comprehensive foundation for analyzing Calgary's crime patterns by combining **location, population, and time-based factors**, making it ideal for interactive and data-driven insights.

➤ Key questions on the Data

Before beginning the analysis, several key questions were formulated to guide the exploration of Calgary's crime data. These questions help focus the analysis on understanding *when*, *where*, and *why* crime incidents occur more frequently, ensuring the findings are meaningful and actionable. The questions are divided into three main categories:

- **Time-Based Analysis:**

This area focuses on how crime patterns change over time. It explores whether there are noticeable increases or decreases in crime rates across the years and identifies specific **months or seasons** when crime incidents tend to spike. Understanding these temporal trends helps detect cyclical behavior, such as higher crime rates during summer months or holiday periods.

- **Location-Based Analysis:**

This aspect examines the **geographic distribution** of crime across Calgary's various sectors and communities. It aims to identify **which areas report the highest number of incidents** and **which types of crimes** (e.g., theft, assault, or break and enter) are most common in each region. This helps pinpoint high-risk neighborhoods and supports targeted interventions or resource allocation.

- **Comparative Analysis:**

This component compares **crime frequency relative to community population size**. It investigates whether larger populations naturally experience more crimes or if smaller communities sometimes have disproportionately high crime rates. Such comparisons provide deeper insight into whether crime is influenced by population density or other socioeconomic factors.

Together, these key questions establish the foundation for the entire analytical process, guiding data visualization, modeling, and the interpretation of insights in the Power BI dashboards.

➤ Data Preparation and Cleaning

Data preparation was carried out using **Power BI's Power Query Editor** to ensure the dataset was clean, consistent, and ready for analysis. The following steps were followed:

- **Imported the CSV dataset from Kaggle:**

The original Calgary Crime Statistics file was imported into Power BI to begin data transformation and visualization.

- **Removed duplicate records:**

Duplicate rows were identified and deleted to prevent inaccurate counts or misleading insights in the final analysis.

- **Formatted date fields and standardized community names:**

Date values were converted to a uniform format, and community names were standardized to maintain consistency across records.

- **Created calculated columns such as Crime Rate per 1,000 residents:**

New metrics were added to measure crime rates relative to population, allowing for more meaningful comparisons between communities.

- **Validated missing values and performed integrity checks:**

Data was reviewed for missing or incorrect entries to ensure the dataset's overall quality and reliability for analytical modeling.

➤ Data Modeling and Power BI Integration

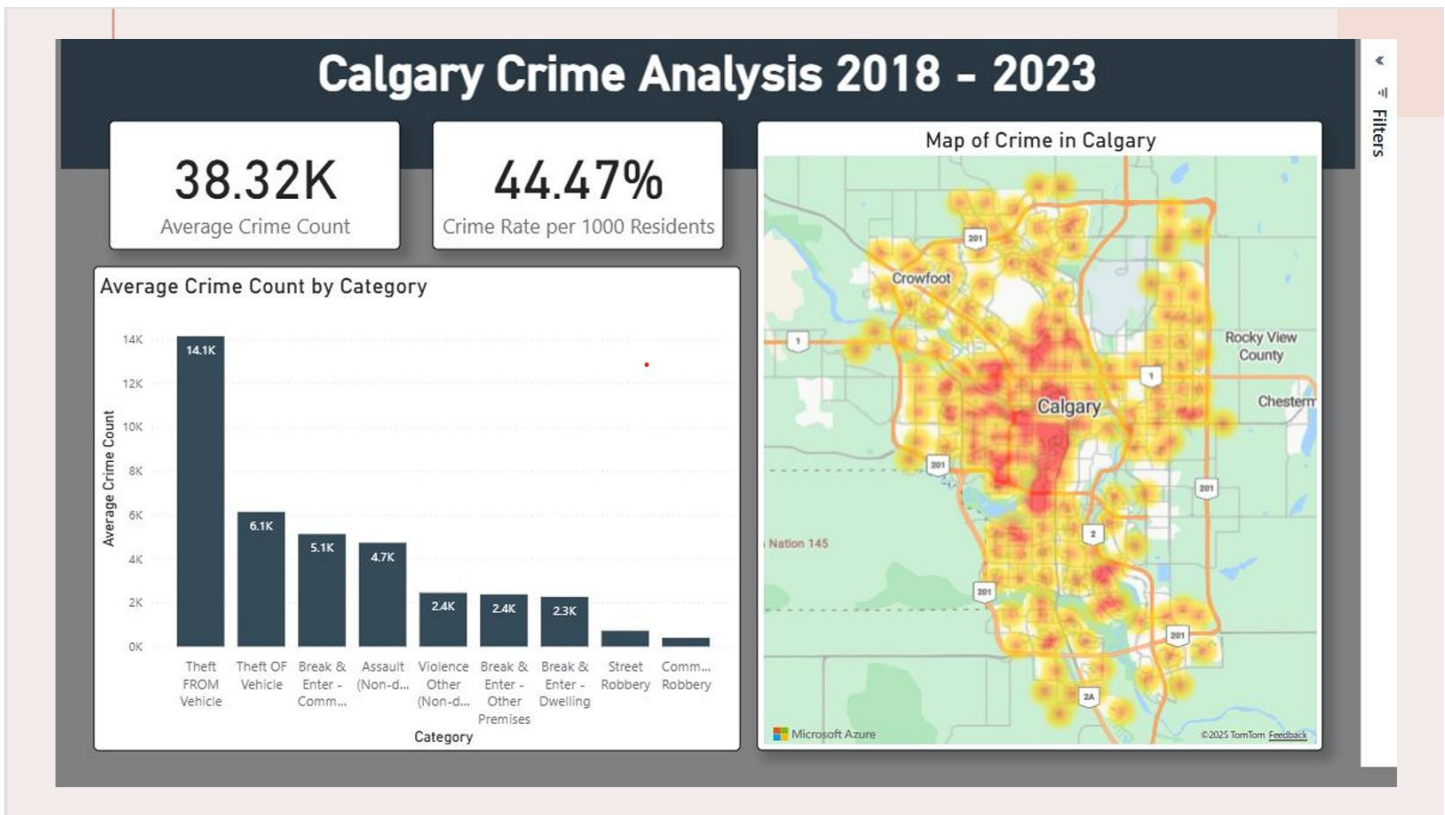
Data modeling in Power BI was central to enabling relational analysis. Relationships were established between various tables, such as Communities, Crime Categories, and Dates. This allowed users to dynamically filter and explore the dataset.

Key components of the data model included:

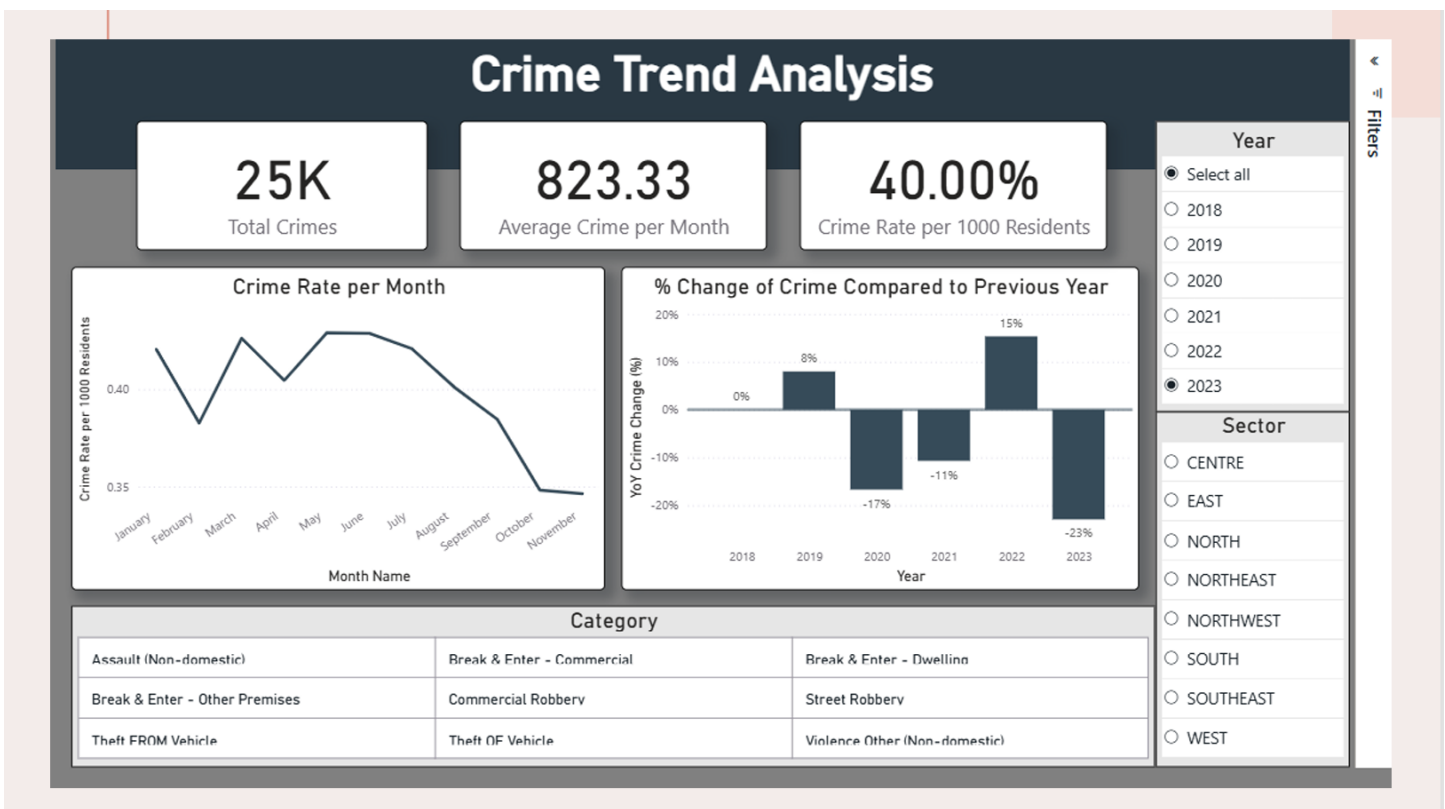
- Community table with population and location details.
- Crime data table with counts and categories.
- Date table supporting time-based trends and comparisons.

➤ Dashboard Design and Features

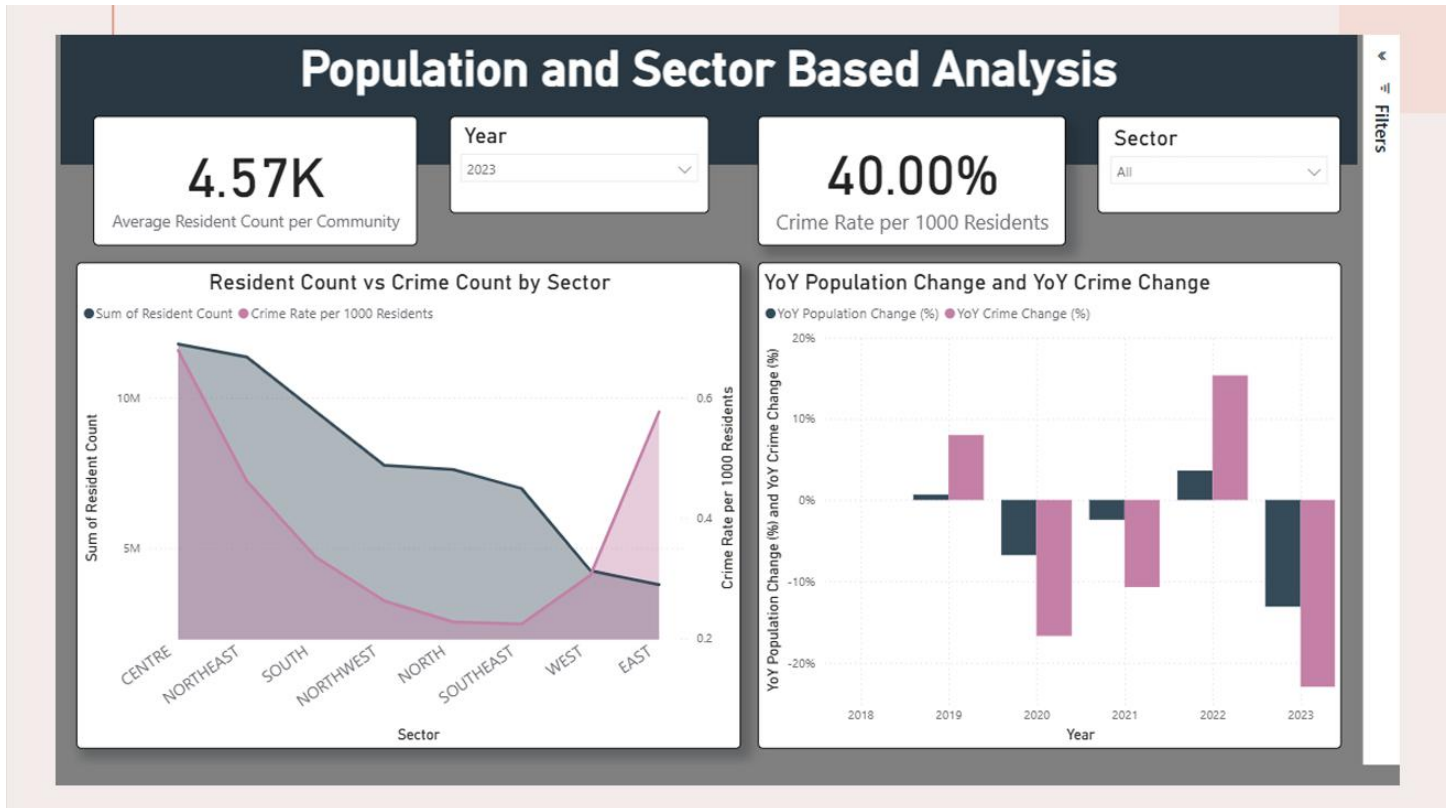
1. Calgary Crime Analysis 2018 - 2023



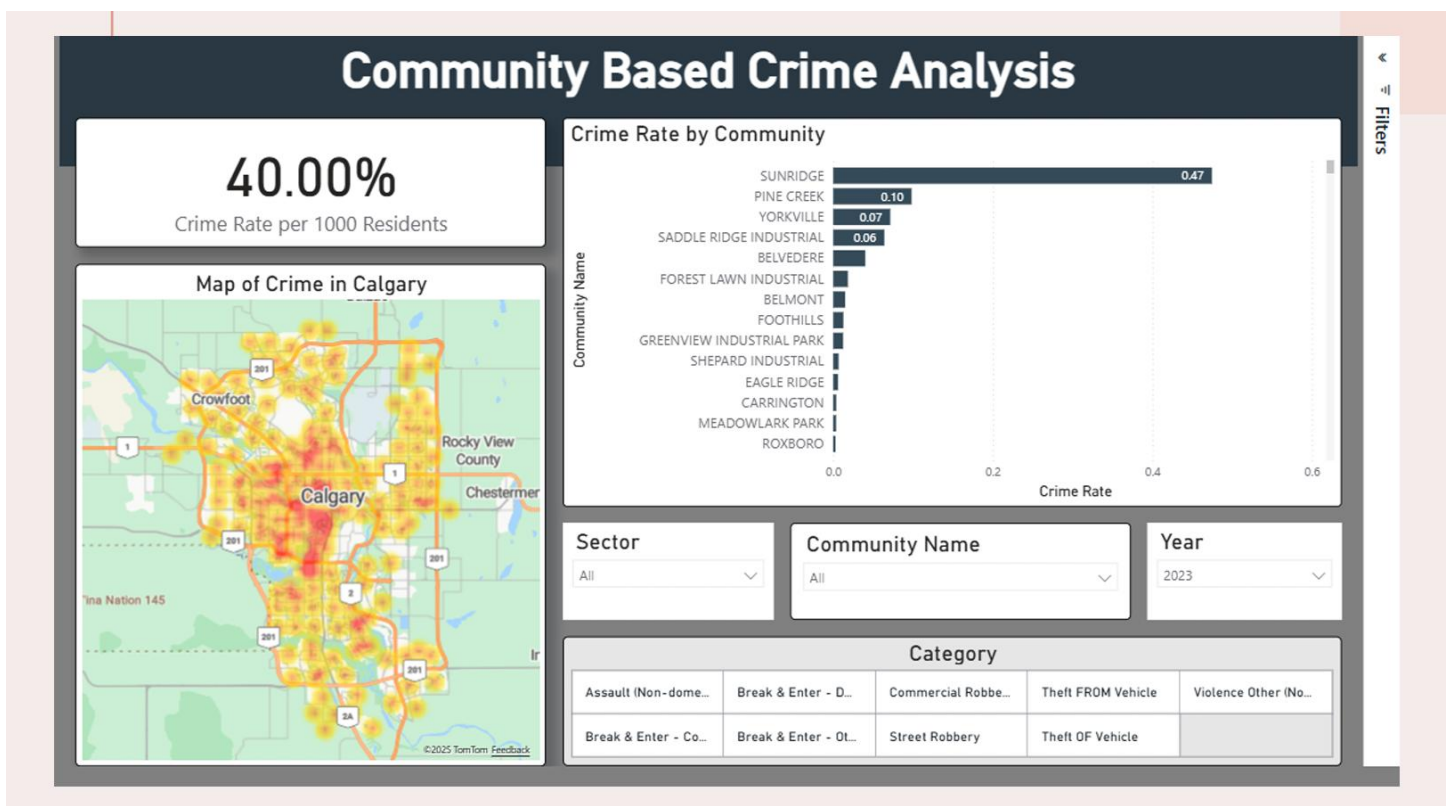
2. Crime Trend Analysis



3. Community Based Crime Analysis



4. Population and Sector Based Analysis



➤ Key Insights and Findings

The following insights were drawn from the analysis:

- Crime fluctuates throughout the year, peaking during the summer.
- Theft and assault are the most common crime categories.
- The Northeast sector is the most crime-prone area.
- Some smaller communities exhibit high crime density relative to population.
- Population size alone does not determine crime rate; other factors influence risk.

➤ Challenges and Limitations

While the project achieved its objectives, several limitations were identified:

- Limited data on demographic and socioeconomic variables restricted deeper correlation analysis.
- Incomplete or missing data for some communities.
- Power BI's geographic mapping precision was limited by coordinate granularity.
- Temporal resolution (monthly data) could be refined with daily incident data for better granularity.

➤ Future Work and Recommendations

Future iterations of this project could expand on the following areas:

- Integrate demographic data (income, age, employment) for deeper insights.
- Use machine learning models to predict future crime trends.
- Include real-time data updates for ongoing monitoring.
- Enhance dashboards with additional KPIs such as response time or case resolution rate.

➤ Conclusion

The **Calgary Crime Analysis Project** effectively utilized open data from the City of Calgary's Open Data Portal and the analytical capabilities of **Power BI** to uncover valuable insights about crime trends across the city. By transforming raw data into interactive dashboards, the project provided a clear and accessible way to explore how crime patterns vary over time and across different communities. This approach highlights the importance of using technology and open data to promote transparency, public awareness, and informed decision-making.

Through the analysis, the project successfully identified **high-risk sectors, seasonal crime trends, and population-based patterns**. The results revealed that certain sectors—particularly the Northeast—consistently experience higher crime rates, while incidents tend to increase during the summer months. Additionally, the comparison between population size and crime frequency showed that smaller communities can sometimes experience disproportionately higher crime rates. These findings emphasize that crime dynamics are influenced by multiple factors, not just population density.

Overall, the insights gained from this project can significantly aid **law enforcement agencies, city planners, and policymakers** in improving community safety strategies. By understanding where and when crimes are most likely to occur, authorities can allocate resources more effectively, design targeted interventions, and implement **data-informed public safety policies**. The Calgary Crime Analysis Project demonstrates how data analytics can turn information into action, ultimately supporting safer and more resilient communities.

➤ References

City of Calgary Open Data Portal. (2023). Calgary Crime Statistics Dataset. Retrieved from <https://www.kaggle.com/datasets/tcashion/calgary-crime-statistics>