

Focasting The Trend of Vehicle Related Crime In Calgary In 2025*

An Analysis of Calgary's Crime Count From 2018 to 2024

Haobo Ren

November 23, 2024

Vehicle-related crime is a long-standing challenge to urban security, so accurate prediction is necessary to effectively allocate police resources and develop security policies. This study examines trends in vehicle-related crime in Calgary, Canada, focusing on two types of crime: Theft of vehicle and Theft from vehicle. A Bayesian regression model is fit to historical crime data from January 2018 to October 2024, accounting for seasonal variations (months), long-term trends (years), and differences between crime categories. The usefulness of Bayesian modeling for crime trend forecasting is demonstrated by this analysis, providing law enforcement and policymakers practical insights to foresee and handle long-term and seasonal crime trends.

1 Introduction

Urban crime is an ongoing problem compromising public confidence, economic stability, and safety. Among the several types of crimes, vehicle-related offenses include theft of vehicles and theft from vehicles remain a major issue in urban areas across the country. These crimes disturb people's life but also burden local governments and law enforcement. Over the years, Calgary, Canada, has seen varying patterns in vehicle-related crimes, which raises issues regarding future developments of these trends. Designing effective preventive plans, maximizing resource allocation, and guaranteeing community safety all depend on an awareness of current trends and future crime count prediction. This study forecasts vehicle-related crime patterns in Calgary up until 2025, therefore addressing these issues.

Analyzing and projecting monthly crime counts for theft from vehicles and theft of vehicles in Calgary is the main aim of this study. I created a Bayesian regression model to assess the

*Code and data are available at: https://github.com/HaoboRrrr/calgary_vehicle_crime_forecast

effects of time (year and month) and crime category on reported crime counts using historical data spanning January 2018 through October 2024. Offering a comprehensive knowledge of how vehicle-related crimes have developed and may continue to alter in the near future, the model catches both long-term patterns and seasonal changes. The research also offers category-specific forecasts, therefore stressing variations in the seasonal and temporal trends of theft from and of vehicles.

The analysis reveals a continued decline in vehicle-related crimes in Calgary, with a mean change of -60.5 each year, and notable mean differences of -454.8 between the two categories. Theft from vehicles is anticipated to show cyclical surges in the summer months, indicating possible opportunistic habits. Conversely, vehicle theft exhibits a more stable decline throughout the course of the year. The data indicate that, although overall vehicle-related crimes are down, seasonal surges in certain categories persist as a problem for policymakers and law enforcement to tackle, for example, the crime count of Theft From Vehicle in August is expected larger than January by 159.4.

The study is significant for its role in proactive crime prevention and resource optimization. This study offers precise and comprehensible forecasts, providing policymakers and law enforcement with evidence-based insights to inform decision-making. This study addresses a significant research gap and enhances public safety in Calgary.

#Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2....

2 Data

2.1 Overview

The raw dataset obtained from Open Data Calgary(Calgary, n.d.) was recorded and updated monthly by the Calgary Police Service. The data is considered cumulative as late-reported incidents are often received well after an offence has occurred. An incident is either reported just after the crime happened, or reported on the Calgary Police Service(Service, n.d.).

The data analysis and visualization is done in R(R Core Team 2023) with the following Packages: tidyverse(Wickham et al. 2019), janitor(Firke 2023), arrow(Richardson et al. 2024), rstanarm(Goodrich et al. 2022), ggplot2(Wickham 2016), dplyr(Wickham et al. 2022), here, knitr.

2.2 Measurement

The data reflects reported crime incidents across Calgary, categorized by community, crime type, and temporal details (year and month). Each row in the dataset represents a summary of crime counts for a specific category and community within a given month and year. The

Calgary Police Service serves as the primary source, systematically recording incidents reported by the public. These reports may be filed in several ways, including: immediate Reporting and delayed Reporting.

Crime count quantifies the number of reported incidents of a specific crime and time period and is based on the most serious violation per incident. The count represents an aggregation of individual reports collected by the Calgary Police Services.

There are limitations in these measurements. Reporting Bias: Not all crimes are reported. Minor incidents or those involving uninsured vehicles may go unreported, leading to underestimation of true crime rates. Cumulative Nature: The inclusion of late-reported incidents makes it challenging to distinguish between crimes that occurred during the reported month versus those that occurred earlier.

By understanding how this dataset measures real-world crime phenomena, we can more confidently interpret the insights and predictions derived from it.

2.3 Data Examination

The raw data from Open Data has five columns and 75,595 rows, the column names are displayed below:

column_names
Community
Category
Crime Count
Year
Month

- Community: The dataset is spatially disaggregated into Calgary's various communities

Numeric Variable:

- Year: The year of each record indicates when the incidents were reported.

Categorical Variable:

- Category: Each entry is categorized into distinct crime types.
- month: The month of each incident recorded. Range from 1-12 repeatedly each year

Response Variable:

- Crime Count: This variable quantifies the number of reported incidents of a specific crime type (e.g., theft from or of vehicles) within a given community and time period.

2.4 Data Cleaning

Several steps were taken for better analysis of vehicle-related trend in Calgary. First, clean the names of all the columns. Second, the raw dataset contained granular records categorized by Community and specific details. These were aggregated to provide summarized counts of crimes (crime_count) by category, year, and month to align with analysis requirements. Third, a new time column was introduced in the cleaned data, combining year and month into a single date field (YYYY-MM-DD). This was essential for time-series analysis and visualization. Last, filter out the categories related to vehicle.

2.5 Cleaned Data

The first 6 rows of the cleaned dataset are shown in Table 2.

Table 2: Head of cleaned Calgary Crime data

category	crime_count	year	month	time
Theft FROM Vehicle	962	2018	1	2018-01-01
Theft FROM Vehicle	645	2018	2	2018-02-01
Theft FROM Vehicle	818	2018	3	2018-03-01
Theft FROM Vehicle	870	2018	4	2018-04-01
Theft FROM Vehicle	1063	2018	5	2018-05-01
Theft FROM Vehicle	1036	2018	6	2018-06-01

2.6 Response Variable

2.6.1 Crime count of Theft FROM Vehicle

Summary statistics of Crime count of Theft FROM Vehicle are shown in the Table 3

Table 3: Summary Statistic of crime_count(Theft FROM Vehicle)

crime_count
Min. : 120.0
1st Qu.: 788.2
Median : 916.0
Mean : 932.7
3rd Qu.: 1105.5
Max. : 1619.0

The response variable, `crime_count`, represents the number of reported incidents of theft from vehicles within Calgary. The summary statistics reveal a significant variability in the crime counts, ranging from a minimum of 120 to a maximum of 1,619 reported cases. The median value is 916, indicating that half of the recorded months experienced fewer than 916 thefts, while the other half exceeded this value. The mean crime count, at 932.7, is slightly higher than the median, suggesting the presence of months with particularly high crime counts that may be skewing the average. The interquartile range (IQR), spanning from 788.2 (1st quartile) to 1,105.5 (3rd quartile), highlights the central tendency of crime occurrences, with most months falling within this range. These statistics emphasize the fluctuating nature of vehicle thefts in Calgary and provide a foundational understanding for modeling and forecasting future trends.

2.6.2 Crime count of Theft OF Vehicle

Summary statistics of Crime count of Theft FROM Vehicle are shown in the Table 4

Table 4: Summary Statitic of `crime_count`(Theft OF Vehicle)

<code>crime_count</code>
Min. :146.0
1st Qu.:361.2
Median :424.5
Mean :422.5
3rd Qu.:469.0
Max. :651.0

The response variable, `crime_count`, for Theft OF Vehicle incidents reflects the monthly reported counts of vehicles being stolen in Calgary. The dataset's summary statistics indicate that the crime counts range from a minimum of 146 to a maximum of 651 incidents. The median count is 424.5, suggesting that half of the months have fewer than 425 thefts, while the other half exceed this value. Interestingly, the mean crime count is 422.5, closely aligned with the median, suggesting a relatively symmetric distribution of the data. The interquartile range (IQR) spans from 361.2 (1st quartile) to 469.0 (3rd quartile), indicating that most monthly counts are concentrated within this range. These figures highlight the scale and regularity of vehicle theft in Calgary and form an essential basis for analyzing and forecasting trends in such crimes.

2.7 Predictor Variable

2.7.1 year

The variable year represents the timeline over which the crime data is recorded, providing insight into how vehicle-related crime counts have changed annually. By analyzing the data across different years, we can observe any long-term trends, such as consistent increases, decreases, or fluctuations in vehicle thefts. This variable helps us understand whether certain years experienced higher crime rates due to factors like economic conditions, population changes, or other external influences.

2.7.2 month and category

The variables month and category provide a detailed view of how vehicle-related crimes are distributed across time and crime types. Month captures seasonal variations, showing how crime rates fluctuate throughout the year. For example, there might be a tendency for higher crime counts during certain seasons, such as summer or winter, influenced by weather conditions or holiday activities. Category distinguishes between “Theft FROM Vehicle” and “Theft OF Vehicle,” allowing us to compare the frequency and patterns of these two crime types. Together, these variables highlight differences in the nature and timing of vehicle-related crimes, offering valuable insights into their temporal and categorical distributions.

2.8 Trend of The Two Crime

The trend of Theft FROM Vehicle and Theft OF Vehicle are shown in [Figure 1](#)

The trends for “Theft FROM Vehicle” and “Theft OF Vehicle” reveal distinct patterns over time. “Theft FROM Vehicle” consistently shows higher crime counts compared to “Theft OF Vehicle,” indicating that it is a more prevalent issue. The trend for “Theft FROM Vehicle” demonstrates significant fluctuations, with peaks around 2019 and a general decline afterward, suggesting a reduction in these incidents in more recent years. In contrast, “Theft OF Vehicle” exhibits a relatively stable pattern with smaller variations over time. Although it remains less frequent than “Theft FROM Vehicle,” its counts appear to decrease gradually, particularly after 2022. Both categories show declining trends toward the end of the observed period, which could reflect successful crime prevention measures or other external influences.

3 Model

The goal of our modelling strategy is twofold. Firstly,...

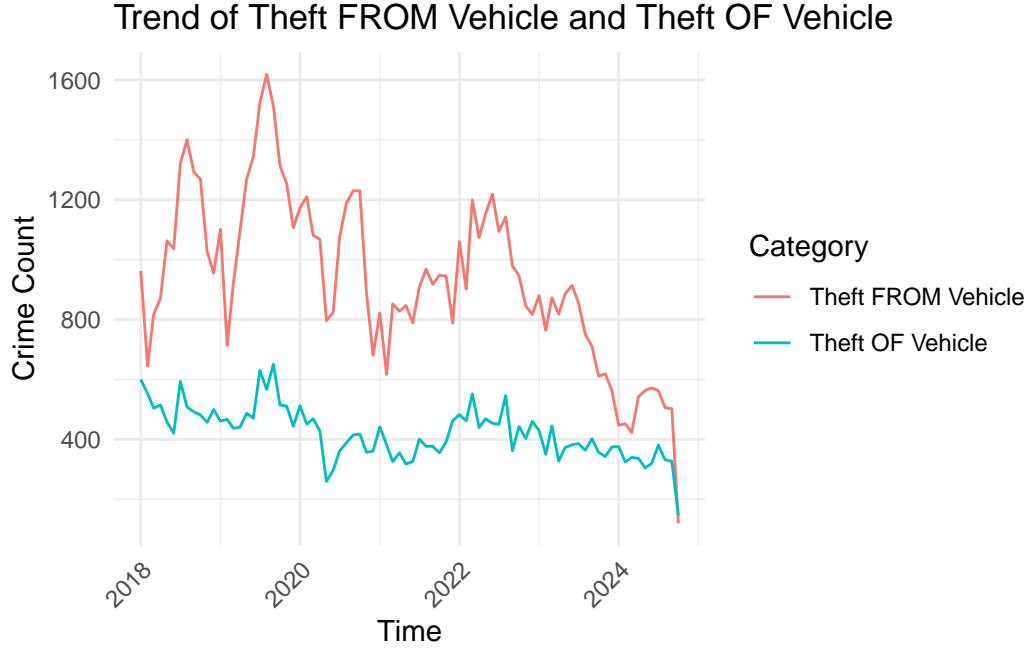


Figure 1

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in Appendix B.

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \quad (1)$$

$$\mu_i = \alpha + \beta_i + \gamma_i \quad (2)$$

$$\alpha \sim \text{Normal}(0, 2.5) \quad (3)$$

$$\beta \sim \text{Normal}(0, 2.5) \quad (4)$$

$$\gamma \sim \text{Normal}(0, 2.5) \quad (5)$$

$$\sigma \sim \text{Exponential}(1) \quad (6)$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ .

4 Results

Our results are summarized in

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In `?@fig-ppcheckandposteriorvsprior-1` we implement a posterior predictive check. This shows...

In `?@fig-ppcheckandposteriorvsprior-2` we compare the posterior with the prior. This shows...

Examining how the model fits, and is affected
by, the data

B.2 Diagnostics

`?@fig-stanareyouokay-1` is a trace plot. It shows... This suggests...

`?@fig-stanareyouokay-2` is a Rhat plot. It shows... This suggests...

Checking the convergence of the MCMC algo-
rithm

References

- Calgary, Open. n.d. *Open Calgary*. <https://data.calgary.ca/>.
- Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “rstanarm: Bayesian applied regression modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoş Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. <https://CRAN.R-project.org/package=arrow>.
- Service, Calgary Police. n.d. *The City of Calgary - Home Page*. <https://www.calgary.ca/cps.html>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*.