

# Toronto Major Crime Trends and Analysis

Alexander Guarasci

September 25, 2024

This paper investigates major crime trends in Toronto from 2014-2024, using data sourced from OpenDataToronto. Of particular interest, the data shows that assault, motor vehicle theft and the total number of major crimes have been steadily on the rise throughout the 10 years this data has been recorded. This research is significant as it provides insights into crime in Canada's largest city; it aims to help build an understanding of the nature of Toronto's crime scene.

## 1 Introduction

Crime data analysis plays a crucial role in understanding social behavior and informing public policy. The city of Toronto, like many major urban centers, faces significant challenges with crime, and accurate data collection is vital to monitor trends, allocate resources effectively, and implement preventative strategies. While many countries and cities publish crime data, not all provide detailed datasets like the "Major Crime Indicators" dataset made available by Toronto Police Services through OpenDataToronto. This paper seeks to analyze crime patterns in Toronto from 2014 to 2024, using this comprehensive dataset to uncover important trends and insights, ultimately showing that major crime is on the rise in Toronto.

Despite the availability of this dataset, there are limitations to its accuracy and coverage. Sexual assaults are notably absent, and certain crimes may be underreported or miscategorized. Nevertheless, the data provides a valuable glimpse into the most commonly reported crimes, allowing for the identification of trends and emerging issues. These patterns are crucial in both understanding the impact of socioeconomic factors on crime and providing recommendations for crime prevention strategies. A detailed analysis of how crime patterns evolved over time, especially through events like the COVID-19 pandemic, fills a gap in the current understanding of urban crime in Toronto.

This paper uses Python (Van Rossum and Drake 2009), a widely adopted tool for data analysis, visualization, and statistical investigation, alongside libraries like Matplotlib (Hunter 2007) and Seaborn (Waskom 2021), to visualize and analyze trends within the data. The paper provides insights into the types of crimes most frequently reported, changes in crime rates over time,

and how these trends differ based on location and time of year. In particular, we will focus on understanding the rise of assaults, motor vehicle thefts, and break-and-enters, and explore their geographical and temporal distribution.

The following sections will delve into the structure of the dataset, the methodology used for cleaning and analyzing the data, the resulting visualizations of major crime trends, and a discussion on the implications of these trends for public policy and law enforcement strategies. By the end of this paper, we aim to offer a clearer understanding of crime dynamics in Toronto.

## 2 Data

The data used in this paper was gathered from OpenDataToronto (Gelfand 2022). OpenDataToronto is a publicly funded data-gathering service that collects and publishes data about Toronto and the GTA. The specific dataset used in this paper is called “Major Crime Indicators”, and was published by Toronto Police Services. The data includes “all Major Crime Indicator offences reported by date” In order to perform the analysis, the programming language for data analysis, visualization and statistical investigation Python was used (Van Rossum and Drake 2009), along with the Python packages Matplotlib.pyplot (Hunter 2007), Seaborn (Waskom 2021), Pandas (team 2020), Requests(Reitz and Team 2024), Pytest(Krekel and team 2024), IO (Foundation 2024b), and Random (Foundation 2024a).

### 2.1 The Dataset

As mentioned, the dataset used in this paper is the “Major Crime Indicators” from OpenDataToronto. The dataset includes around 400,000 samples each with 26 observations, including the report date, the occurrence date, the neighbourhood where the crime took place, the crime location (whether it was in an apartment, a restaurant, outside, etc), and many more. No other datasets were considered because no other department has access to this information. There are a few interesting things to note with the dataset. The first is that sexual assaults were not included, it is not explained why this is done but it is an important factor to note when looking through the paper. Secondly, there are situations in which a crime has several instances of data. For example, if it was the case that someone committed breaking and entering (“B&E”) and then proceeded to assault the residents, this may be recorded as two separate instances (ie. one for B&E and one for assault). Lastly, the dataset does not include reported crimes that police have investigated and determined did not occur nor were attempted. These situations may result in inaccurate reporting, so it is highly likely that the data is not flawless and may leave some crimes out.

With respect to the cleaning of the dataset, the most important thing that was done was removing redundant columns (ie. there were 7 columns that had data on the occurrence date

which seemed unnecessary). Also, eliminating police-specific data points, like the ID of the crime which we have no use for in this analysis allowed for a more succinct analysis.

## 2.2 Variables of Interest

The cleaned data includes four variables. The date of the offence, the type of offence, the neighbourhood where the offence took place and the location type (ie. apartment, etc). The date of the offence needs no explanation, but it is important to note that although this data started being recorded in 2014, some people reported crimes back as far as 1966. Because this skews certain data points, all the analysis is done for crimes committed after 2014 (the older data is still in the cleaned dataset so people can examine it if they choose). The type of offence is also relatively straightforward, however, it is important to note that the criminal code distinguishes between crimes that might generally be lumped together and these specific offences are referenced in the dataset. For example, assault and assault with a weapon are two different offences, and so is B&E and B&E with intent. The neighbourhood unambiguous - these are the 158 social planning neighbourhoods within Toronto, and these are used for all relevant datasets. And the location type is also straightforward.

## 2.3 Measurement

The measurement of this data is not particularly complex and does not have too much room for error. The event takes place and then is reported to the police, the police categorize each instance, using the reported date, time, location, neighbourhood, and offence and this information is added to the dataset. The reporting method is somewhat flawed, it all depends on what the victim claims happened. Almost certainly some details (like the hour the incident took place) are not completely accurate, as a lot of these crimes were reported quite long after the fact. There is also the police categorization of these crimes, for example, what distinguishes assault from assault with a weapon, and what is considered a weapon? These can create uncertainty and measurement errors in the data, but ultimately, because of how broad these categorizations are and their legal definitions, it should not detract from the following analysis.

## 3 Data Visualization and Analysis

Our analysis explores the trends in major crime indicators in the GTA from 2014 to 2024.

The Figure 1 shows the overall breakdown of the top 20 types of major crimes committed over the study.

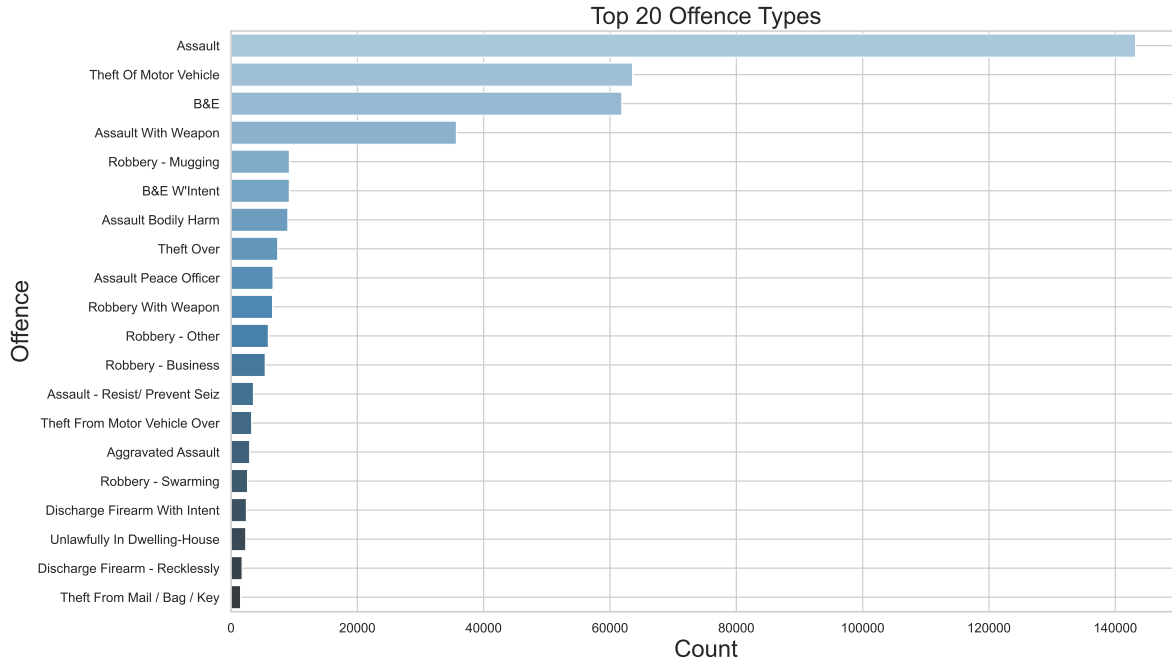


Figure 1

As you can see from the data, by far the most common type of major crime is assault with 144,101 occurrences from 2014-2024. There were 63,633 motor vehicle thefts and 61,954 reported B&E's. The graph only shows the top 20 most common major crimes committed over this period, but the data shows that there were reports of 657 occurrences of “Administering Noxious Thing” and 1 instance of “Hoax Terrorism Causing Bodily Harm”. The overwhelming amount of assaults relative to all other major crime indicators is particularly notable.

Figure 2 shows the trends in the top 5 major crimes from 2014 to 2024

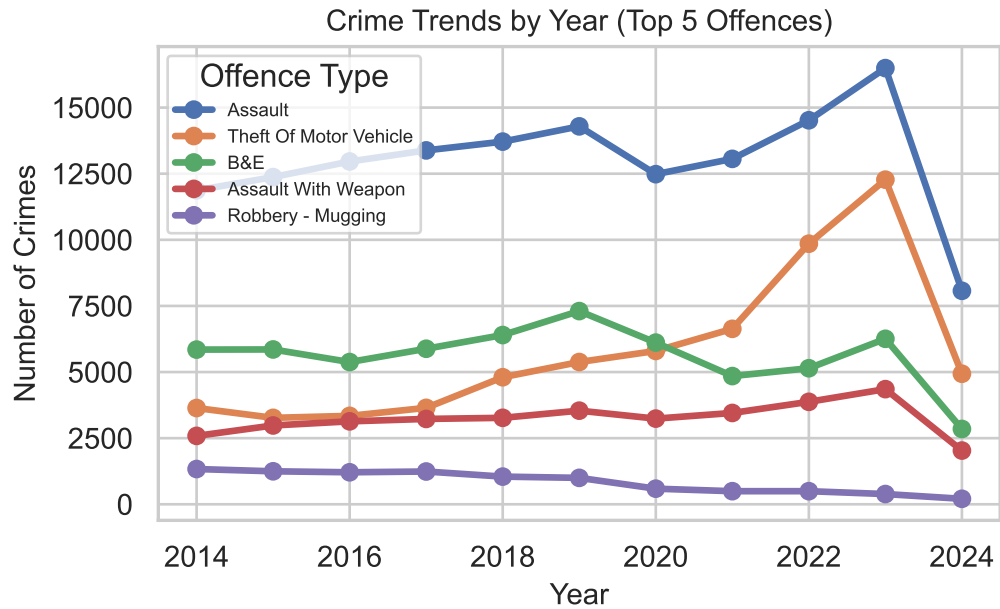


Figure 2

Figure 2 is somewhat misleading, but it is able to provide insights into the trends of these top 5 major crime indicators. The primary issue is that the 2024 data is not yet complete as the year is ongoing. The data is only up-to-date as of June 2024, therefore the decline in 2024 is not due to a decrease in crime but is instead due to a lack of data. With that explained, clearly the trend in assaults and theft of motor vehicles has been substantially increasing since the data started being collected. Whereas there has been some progress made with a slight decline in B&Es as well as robberies. This graph also shows how impactful COVID was in reducing violent crime, which is an interesting observation.

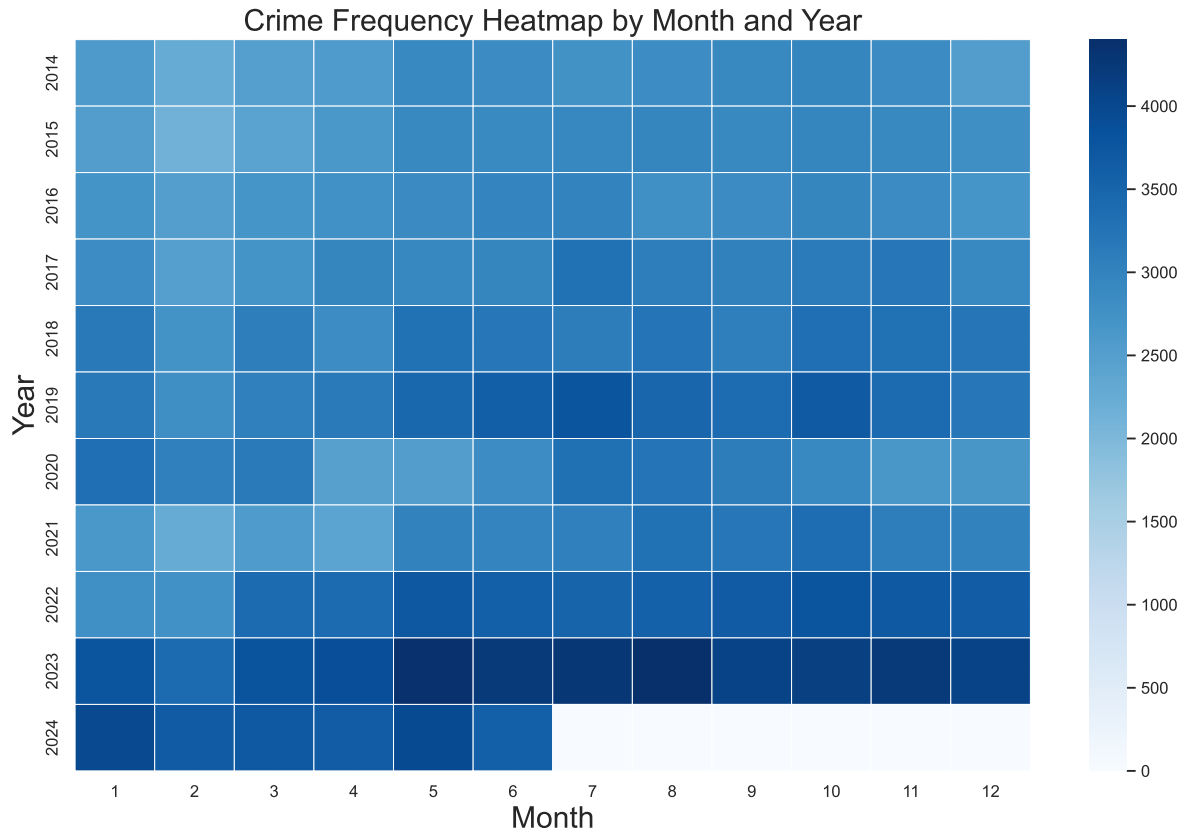


Figure 3

Figure 3 is more effective at showing the overall increasing trend in violence. The darker squares are months that have had more crimes. As you can see the last six months of 2024 have no data so there are zero reported crimes. This missing information skews the [chart] so it appears that there has been less crime in 2024, however if you examine the data on a month-by-month basis, it becomes clear that major crime is on the rise.

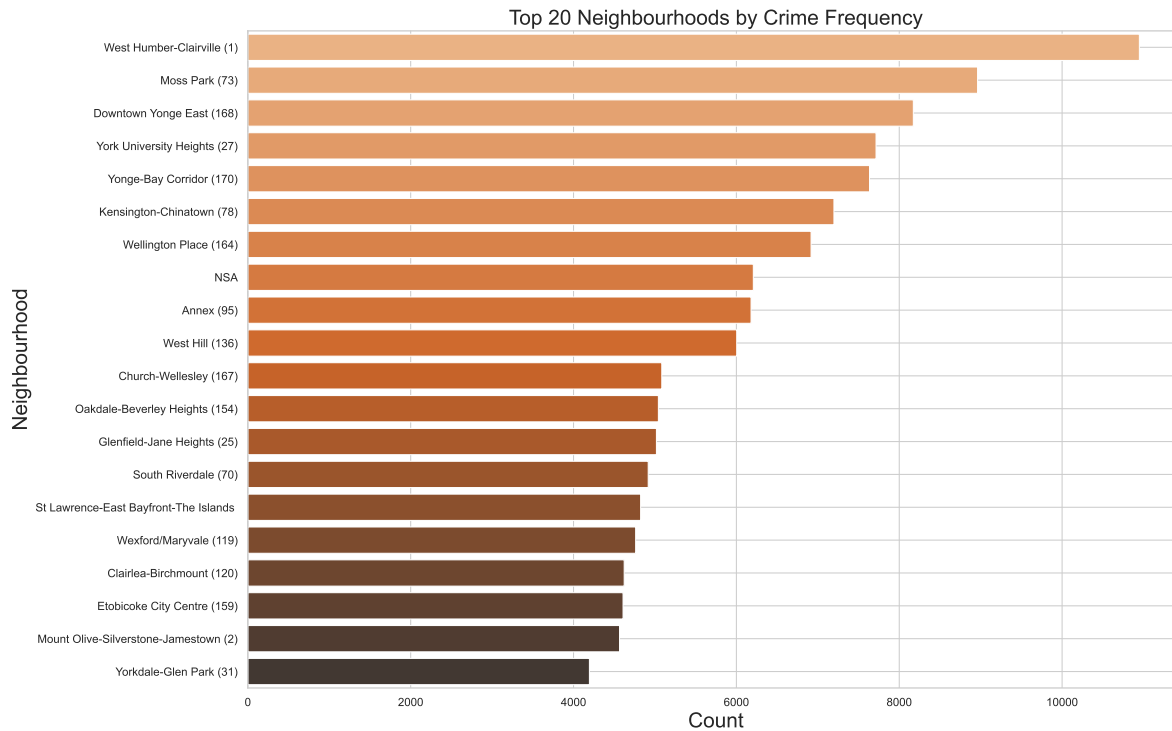


Figure 4

Figure 4 reveals the 20 neighbourhoods with the highest crime rates in the GTA. Understanding which neighbourhoods are most affected by crime is critical to enable local law enforcement and policymakers to develop targeted interventions. The frequency of major crimes is notably concentrated in certain areas, suggesting the need for enhanced security measures and community engagement in those neighbourhoods.

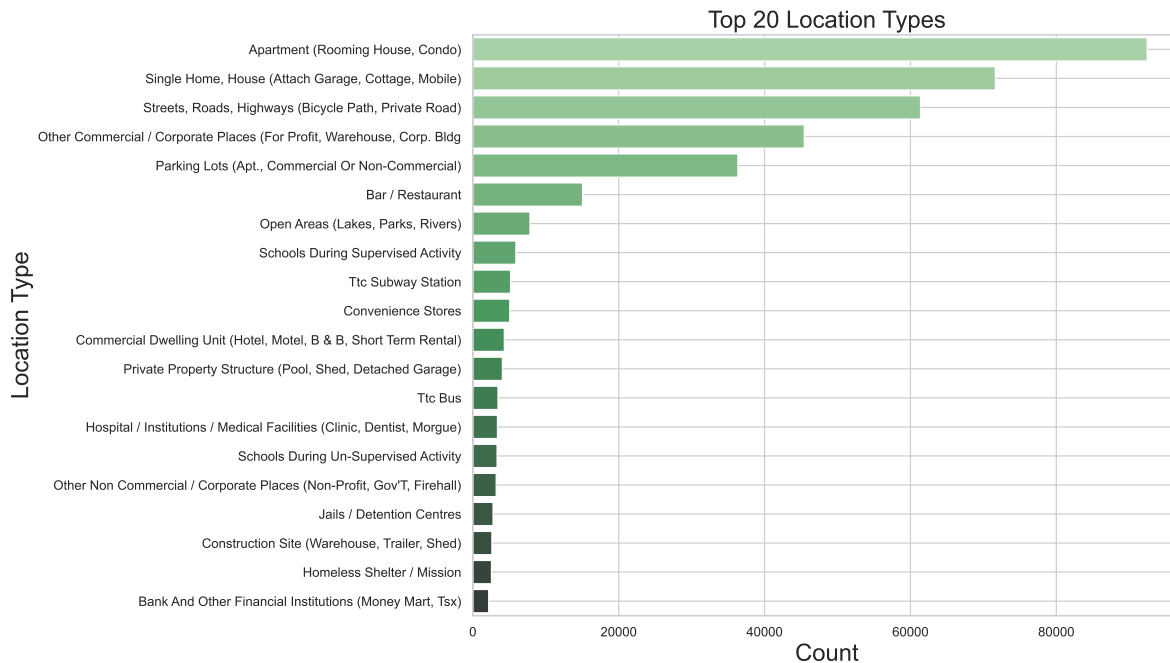


Figure 5

Figure 5 highlights the top 20 types of locations where major crimes were committed. Unsurprisingly, homes and apartments top the list. This raises an important question: Are these locations more prone to crime due to a higher likelihood of being victimized in your home, or do people spend more time at home and thus report crimes there more frequently? Further analysis would be needed to explore this in greater detail.

Each of the figures provided offers a distinct view of crime trends and patterns across Toronto, with a focus on different aspects such as offence types, neighbourhoods, and locations. These insights can be pivotal in addressing crime prevention and policy formulation moving forward.

## Conclusion

This study delved into the “Major Crime Indicators” dataset from OpenDataToronto, offering a comprehensive analysis of crime trends in Toronto between 2014 and 2024. The findings highlight persistent and growing challenges for public safety, with assaults being the most frequently reported major crime, followed by motor vehicle thefts and breaking and entering. The rise in violent crimes, particularly assaults, underscores an urgent need for targeted law enforcement and community-based interventions to address these issues.



The spatial analysis of crime further highlights certain Toronto neighborhoods as more vulnerable to specific types of crimes. This geographic concentration emphasizes the necessity of localized, context-specific crime prevention strategies. Insights from this analysis provide a foundation for policymakers to enhance resource allocation, law enforcement presence, and community engagement in high-crime areas.

Despite the limitations of the dataset—including the absence of sexual assault data and potential misclassification of incidents—this study offers valuable insights that can inform public policy, law enforcement strategies, and future research. The exclusion of sexual assaults, in particular, raises important questions about transparency in crime reporting and data completeness. Future work should push for more comprehensive data collection to include under-reported or excluded categories of crime.

Ultimately, the findings of this paper illustrate the importance of large-scale data analysis in understanding urban crime dynamics. By leveraging powerful tools like Python and its data analysis libraries (Pandas, Matplotlib, Seaborn, and more), we can unlock patterns that were previously inaccessible or unnoticed. Moving forward, it is crucial that law enforcement agencies and policymakers adopt data-driven approaches to address the complexities of crime in Toronto. Continuous monitoring, public transparency, and community collaboration are essential to reducing crime rates and enhancing safety across the city.

## References

- Foundation, Python Software. 2024a. “Random: A Module for Generating Pseudo-Random Numbers.” <https://docs.python.org/3/library/random.html>.
- . 2024b. “StringIO: A Module for Reading and Writing Strings as Files.” <https://docs.python.org/3/library/io.html>.
- Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- Hunter, J. D. 2007. “Matplotlib: A 2D Graphics Environment.” *Computing in Science & Engineering* 9 (3): 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
- Krekel, Holger, and pytest development team. 2024. “Pytest: Simple Powerful Testing with Python.” <https://pytest.org/>.
- Reitz, Kenneth, and The Requests Development Team. 2024. “Requests: HTTP for Humans.” <https://requests.readthedocs.io/>.
- team, The pandas development. 2020. “Pandas-Dev/Pandas: Pandas.” Zenodo. <https://doi.org/10.5281/zenodo.3509134>.
- Van Rossum, Guido, and Fred L. Drake. 2009. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.
- Waskom, Michael L. 2021. “Seaborn: Statistical Data Visualization.” *Journal of Open Source Software* 6 (60): 3021. <https://doi.org/10.21105/joss.03021>.