

Toronto Major Crime

Alexander Guarasci

Invalid Date

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

2 Data

The data used in this paper was gathered from OpenDataToronto (**opendatatoronto?**). 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 (**Python?**), along with the Python packages Matplotlib.pyplot (**Matplot?**) and Seaborn (**Seaborn?**).

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 that the crime took place in, where the crime took place (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

thing to note when looking through the paper. Secondly, there are instances in which a crime has several instances of data if it was the case that someone broke into a house (B&E) and proceeded to assault the residences, this may be recorded as two separate instances (ie. one for B&E and one for assault). Lastly, the dataset does not include occurrences that police have investigated and determined did not occur, nor was it attempted. This may result in inaccurate reporting, and police methods are often wrong 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, I removed quite a few redundant columns (ie. there were 7 columns that had data on the occurrence date which seemed unnecessary). Also, I eliminated police-specific data points, like the ID of the crime which we have no use for in this analysis.

2.2 Variables of Interest

The cleaned data includes four variables. The date of the offence, the type of offence, the neighbourhood the offence took place in 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 after 2014 (I left the older data in the cleaned dataset so people can examine it if they choose). The type of offence is also relatively straight forward, one aspect of it that is important is they do distinguish between different types of crimes. For example, assault and assault with a weapon are two different types of offences, and so is B&E and B&E with intent. The neighbourhood aspect is rather unambiguous, these are the 158 social planning neighbourhoods that Toronto is broken up into, and these are for almost 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, it is certain that 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 first graph shows the overall breakdown of the top 20 types of major crimes committed over the study

```
/var/folders/0d/ldk0yf0s73jfz51ll7sgchgr0000gn/T/ipykernel_12079/1327344260.py:27: FutureWarning
```

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign

```
sns.barplot(x=offence_counts.values, y=offence_counts.index, palette="Blues_d")
```

