Datasheet for 'Major Crime Indicators'*

Kuiyao Qiao

April 22, 2024

The Toronto Major Crime Indicators dataset, created and maintained by Open Data Toronto, provides a comprehensive record of crime incidents reported to the Toronto Police Service from 2014 to 2021. The dataset contains 49,142 instances, each representing a unique reported crime with features such as offence type, location, date, and time. While the data has been preprocessed and aggregated to protect privacy, it offers valuable insights into crime patterns and trends in Toronto over time. The dataset is openly distributed under the Open Government License - Toronto. It is expected to be updated periodically, serving as a critical resource for researchers, policymakers, and the public to understand and address crime in the city.

Motivation

1. The dataset was likely created to provide public access to crucial crime statistics in Toronto from 2014 onwards. The specific task is to make this crime data readily available and transparent to the public. It fills a gap by providing detailed, up-to-date crime indicator statistics for the city of Toronto that may not have been easily accessible to the public otherwise.

2. Open Data Toronto created the dataset on behalf of the City of Toronto. Open Data Toronto is the City of Toronto's official open data portal and catalogue.

3. The City of Toronto likely funded the dataset's creation as part of its Open Data initiative to share data with the public freely. Information about an associated external grant needs to be provided.

4. The dataset is part of the City of Toronto's efforts to share municipal data openly through its Open Data Portal. Providing crime statistics helps inform the public and promotes transparency regarding public safety and crime in Toronto.

Composition

---

*Code and data are available at: https://github.com/QPP123/Toronto-Crime-Stats-Over-Years.

1. The instances represent reported crimes in the city of Toronto. There is only one type of instance in the dataset.

2. The dataset contains 49,142 instances (reported crimes).

3. The dataset is a sample covering reported crimes in Toronto from 2014-2021. It does not contain all possible crimes, only those reported to police. The sample is likely representative of the general distribution and trends in crime in Toronto over this period, but some crimes may be underreported. Representativeness could be validated by comparing it to other sources of crime statistics in Toronto.

4. Each instance consists of several features related to the reported crime, including the type of crime, police division where it occurred, premises type, neighbourhood, longitude, latitude, and date and time of occurrence. This is processed, structured data rather than raw, unprocessed data.

5. No explicit labels or targets are associated with each instance beyond the multiple categorical features describing the crime.

6. The dataset is complete, with no missing information. Any missing details about specific crimes were likely not included to protect privacy.

7. No relationships between individual crime instances are made explicit in this dataset. Each row represents a distinct reported crime.

8. There are no predefined recommended data splits, as this is an observational dataset, not one intended for machine learning purposes requiring train/validation/test splits. The data could be split by date ranges if desired.

9. There are no apparent signs of errors or redundancies, as this is a curated dataset from an official government source. Some noise and inconsistencies in how crimes are reported and categorized are possible.

10. The dataset is self-contained and does not rely on any external resources. All relevant information is included directly in the downloadable file.

11. The dataset does not contain any confidential information. Details identifying individuals, such as names and addresses, are not included.

12. The dataset is a matter-of-fact log of reported crimes and contains no directly offensive or threatening content, though some may find the information troubling.

13. The dataset does not directly identify any protected sub-populations. Demographic information about victims and perpetrators is not included.

14. It is not possible to directly identify individuals from the data provided. The closest would be the approximate crime location from latitude and longitude, but no directly identifying information is given.

15. The dataset does not appear to contain sensitive data about protected demographic classes. The only potentially sensitive information is the location of reported crimes.

16. This dataset provides critical insights into the level and nature of reported crime in Toronto over ten years but has limitations. It does not include unreported crimes or demographic details about those involved. Neighbourhood-level data may not reflect specific locations of crimes. Overall, it is a high-quality dataset that tracks crime reports in Toronto.

Collection process

1. Subjects likely reported the data, in this case, the Toronto Police Service (TPS). When a crime is reported to the police, officers collect details about the incident, which are logged into a records management system. Some of this data is then compiled into aggregate crime statistics. The data needs to be clarified whether or not it was validated.

2. The data was likely collected through manual entry by police officers and automated logging via records management software. Validation procedures are not specified.

3. The dataset samples all crimes reported to TPS from 2014-2024. The sampling strategy includes all crimes reported and logged in the system over this period, so it is more of a comprehensive sample than random.

4. The data collection was likely carried out by Toronto Police Service employees (officers and administrative staff) as part of their regular job duties, not crowdsourced or contracted out.

5. The data covers crimes reported from 2014-2021. This matches the timeframe when the crimes occurred; it's not historical data.

6. It's yet to be determined if any ethical review or institutional oversight was involved in collecting this government crime data. This type of official reporting may be exempt from academic-style review processes.

7. The data comes directly from the Toronto Police Service, the primary source that collects reports of crimes from individuals and generates these statistics for the city.

8. Since this data comes from crimes reported to law enforcement, individuals involved would have been notified that information about the incident was being collected as part of the regular police reporting process. The exact notification language needs to be provided.

9. Explicit consent from individuals to have data about the crimes they reported included in aggregate public datasets was likely not obtained, as reporting a crime to law enforcement implies consent for that information to be used for various law enforcement purposes. The specific language regarding consent and data use is not available.

10. There is no mechanism to revoke consent for this data, as that is generally not an option when data is reported to law enforcement.

11. It's yet to be determined if a formal data protection impact analysis was conducted. Still, the data has been compiled not to include any directly identifying personal information to protect the privacy of victims and others involved.

12. The collection of this data is a routine process conducted by law enforcement to track crime statistics. While individuals report specific incidents, the data is aggregated and anonymized before release.

Preprocessing/cleaning/labeling

1. The raw reported crime data was likely preprocessed to some degree to generate these standardized data fields and aggregate statistics suitable for public release. This may have involved data cleaning, standardizing offence categories, geocoding, and aggregating to geographic areas. However, the specific preprocessing steps could be more detailed.

2. The raw, unaggregated crime report data is not included, which will likely protect privacy. Only this summarized version is available.

3. The software used to preprocess the data internally needs to be provided, as this dataset comes directly from the police service.

4. This data has been collected through routine law enforcement processes and curated to provide public aggregate crime statistics. Still, many processing details are not readily available as it is an internal government process. The result is a clean and usable but somewhat opaque dataset.

Uses

1. This dataset is used by the City of Toronto, police, policymakers, researchers, and the general public to understand crime trends and patterns in Toronto over time. It could identify areas with higher crime rates, track changes in specific offences, and inform public safety initiatives and resource allocation.

2. There is no centralized repository linking to all uses of this dataset, but it is likely used in many internal city reports and analyses. As a prominent public dataset, it has likely been used in various research studies and media reports.

3. The dataset could be used for various other tasks related to understanding and responding to crime in Toronto, such as Identifying crime hot spots for targeted interventions, Examining relationships between crime and neighbourhood characteristics, Comparing crime rates across different premise types, Mapping and visualizing crime patterns, Developing predictive models of future crime trends, Evaluating the impact of policing strategies or policy changes.

4. The primary considerations around composition are that this data only includes crimes reported to police, not unreported crimes, so it only captures part of the extent of crime. Crime report data is also sensitive to changes in reporting behaviour, policing practices, and definitions of crimes over time. The lack of demographic data prevents crime analysis by protected classes and mitigates the risk of unfair characterizations of specific groups. Users should be cautious about interpreting raw counts, as they don't account for population differences across areas. Per-capita rates and statistical testing should be used for comparisons.

5. This dataset should not be used to make generalizations about overall crime prevalence or draw conclusions about specific individuals. It is an aggregate set of reported incidents, not a comprehensive view of all crime or a reflection on particular persons, and should be treated as such. With appropriate care and methods, it can provide valuable insights about crime patterns to inform public safety efforts but should be used for tasks other than individual risk assessment.

6. This dataset provides a valuable lens into crime patterns in Toronto but has some notable limitations in scope and granularity. Users should know the inherent constraints and potential biases in reported crime data. Pairing this with other datasets on neighbourhood characteristics, policing, and city policies could enable richer analyses to understand better the socioeconomic and enforcement context around these crime incidents.

Distribution

1. The dataset is distributed openly to the public by the City of Toronto via their Open Data Portal. It can be freely accessed and used by any third party.

2. The dataset is distributed as a downloadable CSV file from the Open Data Toronto website. It does not appear to have a digital object identifier (DOI).

3. The dataset is already available for public distribution as of the collection of this datasheet information (April 2023).

4. The dataset is distributed under the Open Government License - Toronto, allowing free use with minimal restrictions. The exact terms can be found here: https://open.toronto. ca/open-data-license/.[City of Toronto (2024)] There are no fees associated with this license.

5. There do not appear to be any additional IP-based restrictions imposed by third parties on this data. As an open government dataset, it is intended to be freely available.

6. no evident export controls or regulatory restrictions would apply to the use or distribution of this public crime statistics dataset.

7. This dataset exemplifies Toronto's commitment to open data and is a valuable resource for understanding crime patterns in the city. The open distribution enables broad public use and examination.

Maintenance

1. Open Data Toronto, which is part of the City of Toronto government, hosts and maintains the dataset.

2. Contact information for Open Data Toronto is provided on their website: https://open.toronto.ca/contact-us/. The current email is opendata@toronto.ca.

3. No public erratum for this dataset is apparent. Any data issues could be handled by issuing an updated dataset version.

4. This dataset is expected to be updated periodically with new data periods, but a specific update schedule needs to be provided. Updates would likely be communicated via the Open Data Toronto mailing list and social media channels.

5. This dataset aggregates crime incident reports and does not contain individual-level data. As such, there are no stated limits on data retention related to individuals. Retention policies for the disaggregated data are not specified.

6. Older dataset versions do not appear to be separately maintained. The downloadable file and API reflect the latest available data. The obsolescence of older data is implicit in providing an updated version, but there is no explicit communication process for this.

7. There is no evident mechanism for external contributions or extensions to this dataset. As an authoritative government data source, it is unlikely that the City of Toronto would accept or validate third-party contributions to be included in the official dataset. However, third parties could build on this data with their separate extensions.

8. As a regularly updated open dataset from the City of Toronto, this resource is likely to be maintained for the foreseeable future in a stable location for public access. However, specific long-term guarantees are not made explicit. Users should rely on Open Data Toronto channels for future dataset status updates.

References

City of Toronto. 2024. "Major Crime Indicators." City of Toronto Open Data Portal. 2024.
    https://open.toronto.ca/dataset/major-crime-indicators/.