




FEBRUARY 4, 2022

MILESTONE 1

FINDING DATA AND WRITING DATA PLAN

KAJAL RAJAN
MAHESH KUMAR AMDA
CHAHAT KAUR CHHABRA
AMANDEEPSINGH SALUGA



Question 1. What dataset will you use for your course project? (Describe your dataset, and either includes a link to where I can find it online or submit it as a spreadsheet along with your report).

Answer:

We will be working with 'Major Crime Indicators' open-source dataset volunteered by The Toronto Police Service on their website portal-

[Major Crime Indicators | Toronto Police Service Public Safety Data Portal.](#)

The dataset gives detailed information about all types of major crimes committed across Toronto. The offense includes Assault, Break and Enter, Auto Theft, Robbery, and Theft Over. The data provided can be at both offense and victim level, thus one occurrence number can have several records associated with the various Major Crime Indicators (MCIs) used to categorize the occurrence. This dataset doesn't contain incidences that have been considered *unfounded*. According to Statistics Canada, the term *unfounded* is used for incidences that have been reported but did not occur nor were attempted. Moreover, this initiative has been undertaken by The Toronto Police Department Service to help in promoting public safety and awareness among citizens.

Question 2. Describe the dataset. What kind of data does it contain?

Answer:

The dataset 'Major Crime Indicators' is an excel file that gives us details about the major crimes committed (MCI). The crimes that MCI comprises include Assault, Break, auto theft, robbery, and theft-over. The purpose of this dataset is to give people knowledge about public safety and awareness. Because this information is provided at the offender and/or victim level, a single occurrence number may contain multiple rows of information related to the numerous MCIs used to categorize the incident.

The dataset has a different column to give details about the crimes that are committed. Columns include parameters like Occurrence Date, Reported Date, Premise type, offense, division, and many more. The dataset contains both numerical and categorical information. The MCI indicated in the dataset is between the periods of 2014.01.01 to 2021.06.30.

Question 3: is there anything about your data that you don't understand? (i.e., what a column heading means)? How will you find this out?

Answer:

a) The following columns are unclear in terms of the information they contain

–

- Hood_ID
- ucr_ext
- ucr_code
- Division

b) What is the primary key of the dataset – which field uniquely identifies a record?

c) Exclusions in the data – what data is excluded?

Upon further search on the website, we came across the documentation on the dataset. It has information on the fields, what data is included (for e.g., it doesn't include offences that are unfounded), which field is the unique identifier, etc.

The documentation will be used during further analysis to understand the data better.

Question 4: What are some questions you hope to answer with your data? List at least three.

Answer: Questions that we hope to answer with the data are as following:

- a) What are some Major Crimes committed across Toronto that contributes most to the total crime rate? Based on the offence type mentioned in the data.
- b) Will this initiative help in promoting safety and awareness among citizens?
- c) Will this Initiative and analysis help the Toronto Police in reducing the crime rate and ease the crime investigation process? If it helps, then How? Explain.