**BAN140**

**Introduction to Data**  **Visualization**

**Course Project** – **Milestone No. 1**

**Find Your Data and Write Your Data Plan**

**Group ZAA.2231 10**

Sepehr Salehi

Roohalah Taraf

Rongzhao Yi

Conghan Zheng

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

The raw data includes all Major Crime Indicators (MCI) occurrences by reported date and related offences from 2014 to June 30, 2022 around GTA, which contains more than 30k rows. The Major Crime Indicators categories include Assault, Break and Enter, Auto Theft, Robbery and Theft Over. However, for this project, our group will only consider Robbery category crime data from 2021 to 2022 and eliminate data which coordinate has not been marked.

The following link is the raw data found in Public Safety Data Portal; the edited version will be attached with this assignment.

<https://data.torontopolice.on.ca/datasets/TorontoPS::major-crime-indicators-1/about>

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

The dataset contains all the reported and located robbery case occurred from Jan. 2021 to June. 2022 around GTA, which has 3636 rows in total. The data includes case location, reported time, occurrence time, premise type, neighbourhood and coordinates. Following list is columns name and its correlate data type.

|  |  |
| --- | --- |
| Name | Type |
| Index\_ | Number |
| event\_unique\_id | Text |
| Division | Text |
| occurrencedate | Date or Time |
| reporteddate | Date or Time |
| location\_type | Text |
| premises\_type | Text |
| ucr\_code | Number |
| ucr\_ext | Number |
| offence | Text |
| reportedyear | Number |
| reportedmonth | Text |
| reportedday | Number |
| reporteddayofyear | Number |
| reporteddayofweek | Text |
| reportedhour | Number |
| occurrenceyear | Number |
| occurrencemonth | Text |
| occurrenceday | Number |
| occurrencedayofyear | Number |
| occurrencedayofweek | Text |
| occurrencehour | Number |
| mci\_category | Text |
| Hood\_ID | Text |
| Neighbourhood | Text |
| Longitude | Number |
| Latitude | Number |

**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?**

Most of attributes are clear, however the meaning of coding index, event\_unique\_id , UCR\_code, UCR\_ext and object\_id are not defined. To find those out, we can take extra research about what these terms mean in police, or making assumptions by their correlations.

**Question 4. What are some questions you hope to answer with your data? List at least three (you don't need the answers to the questions at this milestone of the course project).**

First, we will try to understand when crime committed more (time, day of week, month) . Second, recognizing of the most dangerous neighbourhoods in GTA. Third, to find which location types have more rubbery occurrence. Forth, fitting any trend which predicts robbery occurrences in future.