**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 Killed or Seriously Injured (KSI) dataset will be used in this project. This dataset includes all traffic collisions events where a person was either Killed or Seriously Injured (KSI) from 2006 – 2021. This dataset is a subset from all traffic collision events. The source of the data comes from police reports where an officer attended an event related to a traffic collision. The KSI data only includes events where a person sustained a major or fatal injury in a traffic collision event.

The following link is the raw data found in Toronto Police- Public Safety Portal; the CSV data file will be attached with this assignment.

<https://data.torontopolice.on.ca/datasets/TorontoPS::ksi/explore?location=43.722421%2C-79.380682%2C11.97>

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

The dataset contains all date, time, location of KSI collision events happened from 2006 to 2021 around GTA, which has 17488 rows in total. The other data such as road Classification, light condition, environment condition, type of vehicle , driver and passengers are included in dataset. Following list is columns name and its correlate data type.

|  |  |  |
| --- | --- | --- |
| **Field** | **Field Name** | **Description** |
| 1 | INDEX\_ | Unique Identifier |
| 2 | ACCNUM | Accident Number |
| 3 | YEAR | Year Collision Occurred |
| 4 | DATE | Date Collision Occurred |
| 5 | TIME | Time Collision Occurred |
| 7 | STREET1 | Street Collision Occurred |
| 8 | STREET2 | Street Collision Occurred |
| 9 | OFFSET | Distance and direction of the Collision |
| 10 | ROAD\_CLASS | Road Classification |
| 11 | DISTRICT | City District |
| 12 | WARDNUM | City of Toronto Ward collision occurred |
| 13 | LATITUDE | Latitude |
| 14 | LONGITUDE | Longitude |
| 15 | LOCCOORD | Location Coordinate |
| 16 | ACCLOC | Collision Location |
| 17 | TRAFFCTL | Traffic Control Type |
| 18 | VISIBILITY | Environment Condition |
| 19 | LIGHT | Light Condition |
| 20 | RDSFCOND | Road Surface Condition |
| 21 | ACCLASS | Classification of Accident |
| 22 | IMPACTYPE | Initial Impact Type |
| 23 | INVTYPE | Involvement Type |
| 24 | INVAGE | Age of Involved Party |
| 25 | INJURY | Severity of Injury |
| 26 | FATAL\_NO | Sequential Number |
| 27 | INITDIR | Initial Direction of Travel |
| 28 | VEHTYPE | Type of Vehicle |
| 29 | MANOEUVER | Vehicle Manouever |
| 30 | DRIVACT | Apparent Driver Action |
| 31 | DRIVCOND | Driver Condition |
| 32 | PEDTYPE | Pedestrian Crash Type - detail |
| 33 | PEDACT | Pedestrian Action |
| 34 | PEDCOND | Condition of Pedestrian |
| 35 | CYCLISTYPE | Cyclist Crash Type - detail |
| 36 | CYCACT | Cyclist Action |
| 37 | CYCCOND | Cyclist Condition |
| 38 | PEDESTRIAN | Pedestrian Involved In Collision |
| 39 | CYCLIST | Cyclists Involved in Collision |
| 40 | AUTOMOBILE | Driver Involved in Collision |
| 41 | MOTORCYCLE | Motorcyclist Involved in Collision |
| 42 | TRUCK | Truck Driver Involved in Collision |
| 43 | TRSN\_CITY\_VEH | Transit or City Vehicle Involved in Collision |
| 44 | EMERG\_VEH | Emergency Vehicle Involved in Collision |
| 45 | PASSENGER | Passenger Involved in Collision |
| 46 | SPEEDING | Speeding Related Collision |
| 47 | AG\_DRIV | Aggressive and Distracted Driving Collision |
| 48 | REDLIGHT | Red Light Related Collision |
| 49 | ALCOHOL | Alcohol Related Collision |
| 50 | DISABILITY | Medical or Physical Disability Related Collision |
| 51 | HOOD\_158 | Unique ID for City of Toronto Neighbourhood (new) |
| 52 | NEIGHBOURHOOD\_158 | City of Toronto Neighbourhood name (new) |
| 53 | HOOD\_140 | Unique ID for City of Toronto Neighbourhood (old) |
| 54 | NEIGHBOURHOOD\_140 | City of Toronto Neighbourhood name (old) |
| 55 | DIVISION | Toronto Police Service Division |
| 56 | ObjectID | Unique Identifier (auto generated) |

**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?**

The most of attributes are clear, however more consideration to be made in use of few variables. For example: The “ACCNUM” is a unique identification for each traffic collision event. Since the data includes every person involved in a collision event, this identification is duplicated.

**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 evaluate how different environmental variables (like road surface condition or light condition) caused KSI collisions. Second, to identify the locations with highest KSI collision frequency. Third, to find which type of collision (Red light related or speed related and etc.) is the most common reason of KSI collision. Forth, evaluate the trend of KSI collisions during this period and fitting any correlation which predicts KSI collisions in future.