Group 12 Project Proposal

Group Members

- Dhaval Potdar
- Divya Sharma
- Jiayi Zhou
- Jiechen Li

Dataset 3

Traffic Crashes Data

```
'data.frame':   763473 obs. of  49 variables:
 $ CRASH_RECORD_ID              : chr  "000ececd00a4474d1abebf13221e4a97efd79a0322f117291716
 $ RD_NO                        : chr  "JG360767" "JG413407" "JG360636" "JG335628" ...
 $ CRASH_DATE_EST_I             : chr  "" "" "" "" ...
 $ CRASH_DATE                   : chr  "07/29/2023 11:40:00 AM" "09/05/2023 02:23:00 PM" "07/
 $ POSTED_SPEED_LIMIT           : int  30 30 45 30 30 30 30 30 30 20 ...
 $ TRAFFIC_CONTROL_DEVICE       : chr  "NO CONTROLS" "TRAFFIC SIGNAL" "NO CONTROLS" "NO CONTI
 $ DEVICE_CONDITION             : chr  "NO CONTROLS" "FUNCTIONING PROPERLY" "NO CONTROLS" "NO
 $ WEATHER_CONDITION            : chr  "CLEAR" "CLEAR" "CLEAR" "CLEAR" ...
 $ LIGHTING_CONDITION           : chr  "DAYLIGHT" "DAYLIGHT" "DAYLIGHT" "DAYLIGHT" ...
 $ FIRST_CRASH_TYPE             : chr  "TURNING" "SIDESWIPE SAME DIRECTION" "SIDESWIPE SAME D
 $ TRAFFICWAY_TYPE              : chr  "NOT DIVIDED" "NOT DIVIDED" "NOT DIVIDED" "ONE-WAY" .
 $ LANE_CNT                     : int  NA NA NA NA NA NA NA NA NA NA ...
 $ ALIGNMENT                    : chr  "STRAIGHT AND LEVEL" "STRAIGHT AND LEVEL" "STRAIGHT AN
 $ ROADWAY_SURFACE_COND         : chr  "DRY" "DRY" "DRY" "DRY" ...
 $ ROAD_DEFECT                  : chr  "NO DEFECTS" "NO DEFECTS" "NO DEFECTS" "NO DEFECTS" .
 $ REPORT_TYPE                  : chr  "ON SCENE" "NOT ON SCENE (DESK REPORT)" "NOT ON SCENE
```

```
$ CRASH_TYPE                    : chr  "NO INJURY / DRIVE AWAY" "NO INJURY / DRIVE AWAY" "NO
$ INTERSECTION_RELATED_I        : chr  "" "" "" "" ...
$ NOT_RIGHT_OF_WAY_I            : chr  "" "" "" "" ...
$ HIT_AND_RUN_I                 : chr  "" "Y" "" "Y" ...
$ DAMAGE                        : chr  "OVER $1,500" "OVER $1,500" "$500 OR LESS" "$501 - $1
$ DATE_POLICE_NOTIFIED          : chr  "07/29/2023 11:43:00 AM" "09/06/2023 11:52:00 AM" "07/
$ PRIM_CONTRIBUTORY_CAUSE       : chr  "FAILING TO YIELD RIGHT-OF-WAY" "IMPROPER OVERTAKING/I
$ SEC_CONTRIBUTORY_CAUSE        : chr  "NOT APPLICABLE" "IMPROPER OVERTAKING/PASSING" "NOT AI
$ STREET_NO                     : int  804 3501 1465 5458 301 11100 1930 5242 16 1622 ...
$ STREET_DIRECTION              : chr  "N" "S" "E" "S" ...
$ STREET_NAME                   : chr  "KEDZIE AVE" "CALIFORNIA AVE" "130TH ST" "WELLS ST" .
$ BEAT_OF_OCCURRENCE            : int  1121 911 434 935 623 2234 2212 1622 1824 2533 ...
$ PHOTOS_TAKEN_I                : chr  "" "" "" "" ...
$ STATEMENTS_TAKEN_I            : chr  "" "" "" "" ...
$ DOORING_I                     : chr  "" "" "" "" ...
$ WORK_ZONE_I                   : chr  "" "" "" "" ...
$ WORK_ZONE_TYPE                : chr  "" "" "" "" ...
$ WORKERS_PRESENT_I             : chr  "" "" "" "" ...
$ NUM_UNITS                     : int  2 2 2 2 2 2 1 2 2 2 ...
$ MOST_SEVERE_INJURY            : chr  "NO INDICATION OF INJURY" "NO INDICATION OF INJURY" "I
$ INJURIES_TOTAL                : int  0 0 0 0 0 3 0 0 0 0 ...
$ INJURIES_FATAL                : int  0 0 0 0 0 0 0 0 0 0 ...
$ INJURIES_INCAPACITATING       : int  0 0 0 0 0 0 0 0 0 0 ...
$ INJURIES_NON_INCAPACITATING   : int  0 0 0 0 0 0 0 0 0 0 ...
$ INJURIES_REPORTED_NOT_EVIDENT: int  0 0 0 0 0 3 0 0 0 0 ...
$ INJURIES_NO_INDICATION        : int  2 2 2 3 2 2 1 3 2 1 ...
$ INJURIES_UNKNOWN              : int  0 0 0 0 0 0 0 0 0 0 ...
$ CRASH_HOUR                    : int  11 14 9 16 3 16 1 21 17 14 ...
$ CRASH_DAY_OF_WEEK             : int  7 3 7 2 1 3 7 3 7 7 ...
$ CRASH_MONTH                   : int  7 9 7 7 9 9 7 2 6 7 ...
$ LATITUDE                      : num  41.9 41.8 41.7 41.8 41.8 ...
$ LONGITUDE                     : num  -87.7 -87.7 -87.6 -87.6 -87.6 ...
$ LOCATION                      : chr  "POINT (-87.706709567706 41.895744504664)" "POINT (-87
```

Source -

Source of data - Link

Description

The data-set has data on Traffic Crashes in Chicago from 2015 onwards, and contains details of each crash, including:

- **Location:** The latitude and longitude of the crash
- **Date and time:** The date and time of the crash
- **Injuries:** The number and type of injuries that occurred
- **Damage:** The estimated damage cost of the crash
- **Crash Type:** The details about the type of crash - hit and run, no right of way, intersection related, causes etc
- **Conditions:** The details about the weather, lighting, traffic, traffic control devices, roads etc

This dataset can be used to identify the probabilities of crashes happening, and their eventual consequences, given the surrounding situations such as time of day, weather, road conditions, traffic conditions, lighting, roadway surface etc

Research Questions -

1. Given that there is a crash, based on the location, time, crash type, and surrounding road and traffic conditions, what is the possibility for an injury to occur?

- **Outcome Variable** - Injury Rate - The probability for an injury to occur, a continuous variable
- **Input/Explanatory Variables** - Location, Date, time of the day, crash type, and surrounding conditions

2. Given that there is a crash, based on the location, time, crash type, and surrounding road and traffic conditions, will there be damages to a vehicle?

- **Outcome Variable** - Damage or no damage (0 or 1) - categorical variable
- **Input/Explanatory Variables** - Location, Date, time of the day, crash type, and surrounding conditions