**Austin Crash Report Data**

***DAMG7370 Fall 2024 – Final Project – Motor Vehicle Collision Analysis***

**Datasets Overview:**

Austin Dataset:

This dataset compiles a decade of traffic crash records in Austin, TX, powering the Vision Zero Viewer dashboard. Data is sourced from the Texas Department of Transportation's Crash Record Information System, which includes reports from various law enforcement agencies, including the Austin Police Department. To ensure accuracy, the dataset is updated with a two-week delay, allowing time for submission, review, and finalization of crash records. However, some incidents may take longer to appear due to ongoing investigations.

Key attributes in the dataset (Summary) include:

* **Crash Date and Time:** When the incident occurred.
* **Crash Location:** Information on latitude, longitude, and address of the crash.
* **Severity Code:** The seriousness of the crash, ranging from minor property damage to fatal incidents.
* **Contributing Factors:** Possible causes, such as driver behavior, environmental conditions, or mechanical failures.
* **Environmental Conditions:** Weather, light conditions, and road surface status at the time of the crash.
* **Units and Persons Involved:** Number of vehicles, pedestrians, or other parties affected.
* **Injury and Fatality Data:** Details on the number and type of injuries or fatalities resulting from the crash.

The dataset is publicly accessible and serves multiple purposes, including traffic safety analysis, urban planning, and public awareness. It is part of Austin's open data initiative and adheres to transparency and data-driven decision-making principles.

Data Profiling in Alteryx

• Imported both the datasets separately into Alteryx.

• Utilized Alteryx tools for data profiling and to perform transformations.

• Studied the datatype of each column, their type (single valued or multi valued).

Data Profiling:

***Austin Dataset: 212,834 records***

***Columns: 43***

1.Numeric Field Records:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ColumnName** | **% Missi** | **Unique Values** | **Min** | **Mean** | **Median** | **Max** | **Std Dev** | **Description** |
| Micro\_Mobility\_By\_DeathCount | 0.0% | 2 | 0.000 | 0.000 | 0.000 | 1.000 | 0.005 | The number of victims in this crash who sustained fatal injuries and were occupants of a micromobility device such as an e-scooter |
| sus\_serious\_injry\_cnt | 0.0% | 9 | 0.000 | 0.035 | 0.000 | 14.000 | 0.214 | Total Suspected Serious Injury Count |
| motor\_vehicle\_serious\_injury\_count | 0.0% | 7 | 0.000 | 0.023 | 0.000 | 7.000 | 0.181 | The number of victims in this crash who sustained serious injuries and were occupants of a motor vehicle |
| motorcycle\_death\_count | 0.0% | 3 | 0.000 | 0.001 | 0.000 | 2.000 | 0.029 | The number of victims in this crash who sustained fatal injuries and were occupants of a motorcycle |
| motorcycle\_serious\_injury\_count | 0.0% | 3 | 0.000 | 0.005 | 0.000 | 2.000 | 0.073 | The number of victims in this crash who sustained serious injuries and were occupants of a motorcycle |
| other\_death\_count | 0.0% | 1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | The number of victims in this crash who sustained fatal injuries and were not categorized under any other mode of transportation |
| pedestrian\_death\_count | 0.0% | 3 | 0.000 | 0.002 | 0.000 | 2.000 | 0.046 | The number of victims in this crash who sustained fatal injuries and were pedestrians |
| Law enforcement fatality count | 0.0% | 5 | 0.000 | 0.003 | 0.000 | 4.000 | 0.058 | he the number of fatalities resulting from this crash, as reported by the investigating law enforcement agency. This may vary from the Vision Zero program's statistical reporting in certain circumstances. |
| pedestrian\_serious\_injury\_count | 0.0% | 5 | 0.000 | 0.004 | 0.000 | 9.000 | 0.068 | The number of victims in this crash who sustained serious injuries and were pedestrians |
| bicycle\_serious\_injury\_count | 0.0% | 5 | 0.000 | 0.002 | 0.000 | 14.000 | 0.054 | The number of victims in this crash who sustained serious injuries and were occupants of a bicycle |
| micromobilityserious\_injury\_count | 0.0% | 3 | 0.000 | 0.000 | 0.000 | 2.000 | 0.015 | The number of victims in this crash who sustained serious injuries and were occupants of a micromobility device such as an e-scooter |
| motor\_vehicle\_death\_count | 0.0% | 5 | 0.000 | 0.003 | 0.000 | 4.000 | 0.057 | The number of victims in this crash who sustained fatal injuries and were occupants of a motor vehicle |
| bicycle\_death\_count | 0.0% | 2 | 0.000 | 0.000 | 0.000 | 1.000 | 0.014 | The number of victims in this crash who sustained fatal injuries and were occupants of a bicycle |
| other\_serious\_injury\_count | 0.0% | 3 | 0.000 | 0.000 | 0.000 | 3.000 | 0.008 | The number of victims in this crash who sustained serious injuries and were not categorized under any other mode of transportation |
| death\_cnt | 0.0% | 5 | 0.000 | 0.006 | 0.000 | 4.000 | 0.080 | Total Death Count |
| crash\_sev\_id | 0.0% | 6 | 0.000 | 3.456 | 3.000 | 5.000 | 1.627 | Crash Severity - Most severe injury suffered by any one person involved in the crash ( 0=UNKNOWN, 1=INCAPACITATING INJURY, 2=NON-INCAPACITATING INJURY, 3=POSSIBLE INJURY, 4=KILLED, 5=NOT INJURED) |
| unkn\_injry\_cnt | 0.0% | 16 | 0.000 | 0.118 | 0.000 | 41.000 | 0.411 | Total Unknown Injury Count |
| poss\_injry\_cnt | 0.0% | 19 | 0.000 | 0.356 | 0.000 | 21.000 | 0.756 | Total Possible Injury Count |
| ID | 0.0% | 212,834 | 2.000 | 225,525.633 | 220,245.500 | 1,344,364.000 | 161,286.192 | The unique crash identifier within the Vision Zero crash database |
| latitude | 1.7% | 128,032 | 30.099 | 30.299 | 30.286 | 30.512 | 0.083 | Derived Latitude map coordinate of the crash |
| crash\_speed\_limit | 0.0% | 35 | -1.000 | 35.085 | 40.000 | 85.000 | 22.707 | Speed limit |
| Crash ID | 0.0% | 212,833 | 11,152,580.000 | 15,813,564.710 | 15,774,423.500 | 20,517,758.000 | 2,558,971.684 | TxDOT C.R.I.S. system-generated unique identifying number for a crash |
| longitude | 1.7% | 127,856 | -97.927 | -97.738 | -97.737 | -97.570 | 0.052 | Derived longitude map coordinate of the crash |
| tot\_injry\_cnt | 0.0% | 20 | 0.000 | 0.676 | 0.000 | 61.000 | 0.974 | Total Injury Count |
| non\_injry\_cnt | 0.0% | 51 | 0.000 | 1.849 | 2.000 | 64.000 | 1.690 | Total Not Injured Count |
| nonincap\_injry\_cnt | 0.0% | 16 | 0.000 | 0.285 | 0.000 | 57.000 | 0.653 | Total Non-incapacitating Injury Count |

2.Date Field:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Name** | **% Missing** | **Unique Values** | **Latest Date** | **Earliest Date** | **Interval** | **Description** |
| Crash\_TimeStamp | 0.0% | 205,819 | 11/13/2024 12:00 | 01/01/2010 06:00 | Unknown | The timestamp at which the crash occurred, in UTC time. |
| Crash\_TimeStamp\_(US/Central) | 0.0% | 205,760 | 11/13/2024 01:00 | 01/01/2010 01:00 | Unknown | The timestamp at which the crash occurred, in US/Central time. |

3.String Field:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Name** | **% Missin** | **Unique Values** | **Shortest Value** | **Longest Value** | **Min Value Count** | **Max Value Count** | **Description** |
| Primary address | 0.0% | 81,592 | LN | 6900 ED BLUESTEIN BLVD SB TO ED BLUESTEIN BLVD SVRD SB | 1 | 1,919 | The primary address where the crash is reported to have occurred |
| case\_id | 1.4% | 209,491 | 1 | 10000 BLK US HIGHWAY | 1 | 2,905 | Case ID |
| Secondary address | 0.0% | 65,404 | 1 | 7200 NE CURBLINE OF THE N IH 35 NB SVRD TURNAROUND TO N IH 35 SB | 1 | 7,748 | The secondary address where the crash is reported to have occurred. Typically the cross street or nearest cross street. |
| point | 1.7% | 129,955 | POINT (-97.65954 30.4) | POINT (-97.70786148611529 30.316989544922194) | 1 | 3,723 | Some values of this field have a small number of value counts. If A Point datatype created with crash latitude and longitude to enable request of GeoJSON. |
| units\_involved | 0.0% | 80 | Bicycle | Large passenger vehicle & Motor vehicle – other & Other/Unknown &Passenger car &Pedestrian | 1 | 75,183 | Mode of units involved in crash |
| rpt\_street\_name | 0.0% | 13,560 | 1 | MOPAC SOUTH BOUND SERVICE TO NORTH BOUND TURN AROUND | 1 | 14,247 | Reported Street Name (road on which crash occurred) |
| rpt\_block\_num | 13.4% | 5,820 | 9 | 9400-11300 | 1 | 28,626 | Reported Block Number (road on which crash occurred) |
| rpt\_street\_sfx | 30.6% | 21 |  | PKWY | 1 | 65,203 | Reported Street Suffix (road on which crash occurred) |

Inferences:

1. The Austin dataset has 3 unique columns namely the ID, Crash\_ID and Case\_ID. However, after the profiling is done it is found that Case\_ID column has 1.4% missing values which makes this column not suitable for assigning a PK to move forward.
2. Both the Crash Timestamp and Crash timestamp (US/Central) have multi-valued data separated by space. These columns can be either divided further for finer grain. Also, this must be cleaned further to make the values standardized into a proper format.
3. The crash speed limit column has the value -1 for accidents that they are not sure about or undergoing investigations. The count for these values are 38,244.
4. There are significant null values found in rpt\_street\_sfx and rpt\_block\_num which can be derived from the column “rpt\_street\_name”.
5. Latitude and Longitude has equal number of null values which needs to be addressed by giving a non-existent lat and long co-ordinates instead of populating them using “NA” or using 0.
6. There are no records where the lat and long column is empty, but the point column is not empty.
7. The column reported\_street\_prefix is entirely null, which can be removed down the line.
8. Fields like micromobility\_death\_count, bicycle\_death\_count, and motorcycle\_death\_count have limited unique values, indicating categorical data.