

Understanding Austin Crash Report Data - Crash Level Records

Y-Data Profiling

No. of Columns: 54

No. of Observations: 147750

Column Name: crash_id

Column Description: TxDOT C.R.I.S. system-generated unique identifying number for a crash

Type: Integer

Distinct values: 147750 (100%)

Missing: None

Natural Key

Column Name: crash_fatal_fl

Column Description: Fatal Crash Identifier - Indicates that the crash involved one or more fatalities

Type: Boolean

Missing: None

"Y" values: 866 (0.6%)

"N" values: 146884 (99.4%)

Column Name: crash_date

Column Description: Date of crash

Type: Date

Missing: None

Distinct values: 144667 (97.9%)

Minimum date of crash: 2014-03-26 06:41:00

Maximum date of crash: 2024-03-11 22:05:00

Column Name: crash_time

Column Description: Crash Time - Time crash occurred

Type: Date

Missing: None

Column Name: case_id

Column Description: Case ID

Type: String (It has an alpha numeric value)

Distinct values: 145678 (99.9%)

Missing: 1858 (1.3%)

Additional comments: This field can be ignored for analysis purpose as we

Column Name: rpt_latitude

Column Description: Reported Latitude

Type: Double

Distinct values: 7976 (77.5%)

Missing: 137456 (93%)

Median: 30.29528 (At this latitude a maximum number of crashes are reported)

Additional comments: There are outliers present in this field

Column Name: rpt_longitude

Column Description: Reported Longitude

Type: Double

Distinct values: 7264 (70.6%)

Missing: 137456 (93%)

Median: -97.73109 (At this longitude a maximum number of crashes are reported)

Additional comments: There are outliers present in this field

Column Name: rpt_block_num
Column Description: Reported Block Number (road on which crash occurred)
Type: String (It has an alpha numeric value)
Missing: 19611 (13.3%)
Additional comments: This field can be ignored for analysis purpose as the data is not consistent

Column Name: rpt_street_pfx
Column Description: Reported Street Prefix (road on which crash occurred)
Type: String
Distinct Values: 8
Values: N (North), E (East), S (South), W (West), SW (South-West), NE (North-East), SE (South-East), NW (North-West), NaN
Missing: 67805 (45.9%)
Additional comments: Maximum number of crashes occurred at N (North) side with count of 26382 (33%)

Column Name: rpt_street_name
Column Description: Reported Street Name (road on which crash occurred)
Type: String
Distinct Values: 9794 (6.6%)
Missing: 3 (<0.1%)

Column Name: rpt_street_sfx
Column Description: Reported Street Suffix (road on which crash occurred)
Type: String
Distinct Values: 18
Missing: 50340 (34.1%)

Column Name: crash_speed_limit
Column Description: Speed Limit
Speed Limit range: 0 to 85
Additional comments: There are outliers present in this field

Column Name: road_constr_zone_fl
Column Description: Construction Zone - Indicates whether the crash occurred in or was related to a construction, maintenance, or utility work zone, regardless of whether or not workers were actually present at the time of the crash
Type: Boolean
"Y" values: 7847 (5.3%)
"N" values: 139901 (94.7%)

Column Name: latitude
Column Description: Derived Latitude map coordinate of the crash
Type: Double
Distinct values: 96355 (66.2%)
Missing: 2243 (1.5%)
Median: 30.284661 (Maximum number of crashes are occurred at this latitude)

Column Name: longitude
Column Description: Derived Longitude map coordinate of the crash
Type: Double
Distinct values: 96230 (66.1%)
Missing: 2243 (1.5%)
Median: -97.736145 (Maximum number of crashes are occurred at this longitude)

Column Name: street_name

Column Description: Derived Street Name - Name of the road crash occurred on, as determined by the Locator application.

Type: String

Distinct Values: 4630 (3.1%)

Missing: 2

Column Name: street_nbr

Column Description: Derived Street Number - Block number of primary street where crash occurred as determined by the Locator application

Type: Integer

Missing: 19611 (13.3%)

Column Name: street_name_2

Column Description: Derived Street Name 2 - The road name for the secondary road related to the crash location (If applicable)

Type: String

Distinct Values: 3396 (5.1%)

Missing: 81474 (55.1%)

Column Name: street_nbr_2

Column Description: Derived Street Number 2 - Block number of secondary street related to the crash location as determined by the Locator application (If applicable)

Type: Integer

Missing: 147750 (100%)

Additional comments: This field doesn't contain any values, so it can be ignored.

Column Name: crash_sev_id

Column Description: Crash Severity - Most severe injury suffered by any one person involved in the crash

Type: Integer

Distinct values: 8

0=UNKNOWN, 1=INCAPACITATING INJURY, 2=NON-INCAPACITATING INJURY, 3=POSSIBLE INJURY, 4=KILLED, 5=NOT INJURED

Median: 5=NOT INJURED

Additional comments: There are outliers present in this field

Column Name: sus_serious_injry_cnt

Column Description: Total Suspected Serious Injury Count

Type: Integer

Minimum: 0

Maximum: 10

Missing: None

Column Name: nonincap_injry_cnt

Column Description: Total Non-incapacitating Injury Count

Type: Integer

Minimum: 0

Maximum: 14

Missing: 1

Column Name: poss_injry_cnt

Column Description: Total Possible Injury Count

Type: Integer

Minimum: 0

Maximum: 20

Missing: 1

Column Name: non_injry_cnt
Column Description: Total Not Injured Count
Type: Integer
Minimum: 0
Maximum: 56
Missing: 1

Column Name: unkn_injry_cnt
Column Description: Total Unknown Injury Count
Type: Integer
Minimum: 0
Maximum: 41
Missing: 2

Column Name: tot_injry_cnt
Column Description: Total Injury Count
Type: Integer
Minimum: 0
Maximum: 21
Missing: 2

Column Name: death_cnt
Column Description: Total Death Count
Type: String
Distinct values: 5
Missing: None
Additional comments: Maximum number of deaths count is 4. Data type should be changed.

Column Name: contrib_factr_p1_id
Column Description: The first factor for a given vehicle which the officer felt possibly contributed to the crash
Type: Integer
Distinct values: 70
Missing: 119143 (80.6%)
Median: 20

Column Name: contrib_factr_p2_id
Column Description: The second factor for a given vehicle which the officer felt possibly contributed to the crash
Type: Integer
Distinct values: 65
Missing: 143235 (96.9%)
Median: 22

Column Name: units_involved
Column Description: Mode of units involved in crash
Type: String
Missing: 7

Column Name: atd_mode_category_metadata
Column Description: Description of units involved in crash
Type: String
Missing: 7
Sample value: [{"mode_id": 1, "mode_desc": "Passenger car", "unit_id": 2259431, "death_cnt": 0, "sus_serious_injry_cnt": 0, "nonincap_injry_cnt": 0, "poss_injry_cnt": 0, "non_injry_cnt": 4, "unkn_injry_cnt": 0, "tot_injry_cnt": 0}]
Additional comments: It is a multivalued attribute.

Column Name: pedestrian_fl
Column Description: Pedestrian involved crash flag
Type: Boolean
Missing: 144245 (97.6%)
"Y" values: 3505 (2.4%)
Additional comments: There are no "N" values in this field, instead there are null values.

Column Name: motor_vehicle_fl
Column Description: Motor vehicle involved crash flag
Type: Boolean
Missing: 1116 (0.8%)
"Y" values: 146634 (99.2%)
Additional comments: There are no "N" values in this field, instead there are null values.

Column Name: motorcycle_fl
Column Description: Motorcycle involved crash flag
Type: Boolean
Missing: 144148 (97.6%)
"Y" values: 3602 (2.4%)
Additional comments: There are no "N" values in this field, instead there are null values.

Column Name: bicycle_fl
Column Description: Bicyclist involved crash flag
Type: Boolean
Missing: 145306 (98.3%)
"Y" values: 2444 (1.7%)
Additional comments: There are no "N" values in this field, instead there are null values.

Column Name: other_fl
Column Description: Other involved crash flag
Type: Boolean
Missing: 142905 (96.7%)
"Y" values: 4845 (3.3%)
Additional comments: There are no "N" values in this field, instead there are null values.

Column Name: point
Column Description: Point datatype created with crash latitude and longitude to enable request of GeoJSON
Type: String
Distinct values: 97739 (67.2%)
Missing: 2243 (1.5%)
Additional comments: The type of this field is text which needs to be changed to numeric to capture the location.

Column Name: apd_confirmed_fatality
Column Description: APD Fatality flag
Type: Boolean
Missing: None
"Y" values: 842 (0.6%)
"N" values: 146908 (99.4%)

Column Name: apd_confirmed_death_count
Column Description: APD Fatality flag
Type: String
Missing: None
Distinct values: 5
Minimum: 0
Maximum: 4
Additional comments: This field contains values in text type that should be changed to numeric type for analysis.

Column Name: motor_vehicle_serious_injury_count
Column Description: N/A
Type: Integer
Distinct values: 6
Minimum: 0
Maximum: 5

Column Name: bicycle_death_count
Column Description: N/A
Type: String
"0" values: 147721 (>99.9%)
"1" values: 29 (<0.1%)

Column Name: pedestrian_death_count
Column Description: N/A
Type: String
Distinct values: 3
Additional comments: This field contains values in text type that should be changed to numeric type for analysis.

Column Name: pedestrian_serious_injury_count
Column Description: N/A
Type: String
Distinct values: 5
Additional comments: This field contains values in text type that should be changed to numeric type for analysis.

Column Name: motorcycle_death_count
Column Description: N/A
Type: String
Distinct values: 3
Additional comments: This field contains values in text type that should be changed to numeric type for analysis.

Column Name: motorcycle_serious_injury_count
Column Description: N/A
Type: String
Distinct values: 3
Additional comments: This field contains values in text type that should be changed to numeric type for analysis.

Column Name: motorcycle_serious_injury_count
Column Description: N/A
Type: String
Distinct values: 3
Additional comments: This field contains values in text type that should be changed to numeric type for analysis.

Column Name: other_death_count
Column Description: N/A
Type: String
Additional comments: This field contains values in text type that should be changed to numeric type for analysis.

Column Name: other_serious_injury_count
Column Description: N/A
Type: String
Distinct: 3
Additional comments: This field contains values in text type that should be changed to numeric type for analysis.

Column Name: onsys_fl

Column Description: Flag indicates whether primary road of crash was on the TxDOT highway system.

Type: Boolean

Missing: 0

"Y" values: 73041 (49.4%)

"N" values: 74709 (50.6%)

Column Name: private_dr_fl

Column Description: Flag indicating whether crash occurred on a private drive or road/private property/parking lot.

Type: Boolean

Missing: 0

"N" values: 147750 (100%)

"Y" values: 0

Column Name: micromobility_serious_injury_count

Column Description: N/A

Type: String

Distinct: 3

Missing: 0

Additional comments: This field contains values in text type that should be changed to numeric type for analysis.

Column Name: micromobility_death_count

Column Description: N/A

Type: String

Missing: 0

"1" values: 6

"0" values: 147744

Additional comments: This field contains values in text type that should be changed to numeric type for analysis.

Column Name: micromobility_fl

Column Description: N/A

Type: Boolean

Missing: 147439 (99.8%)

"Y" values: 311 (0.2%)

Additional comments: There are no "N" values in this field, instead there are null values.