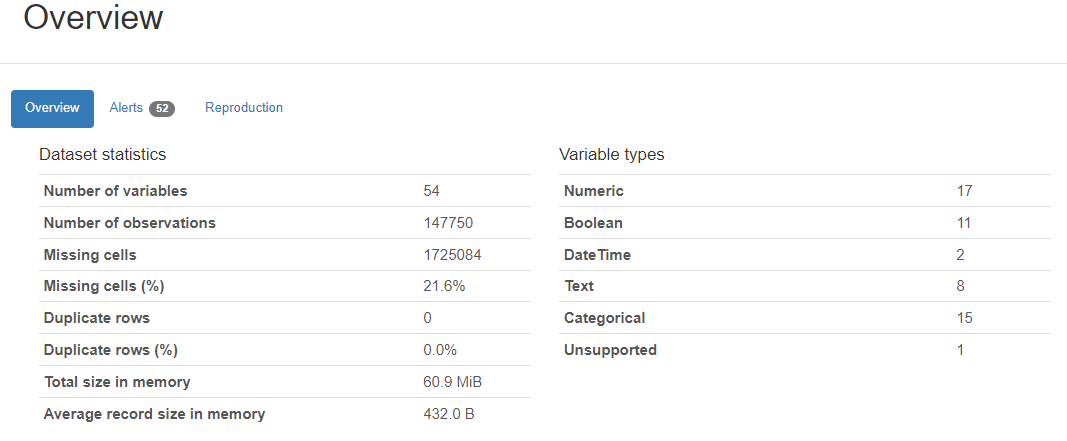
***Austin Dataset***

Crash data is obtained from the Texas Department of Transportation (TXDOT) Crash Record Information System (CRIS) database, which is populated by reports submitted by Texas Peace Officers throughout the state, including the Austin Police Department (APD), and maintained by TXDOT. This dataset contains crash-level records for crashes that have occurred in the last ten years. Crash data may take several days or weeks to be initially provided and finalized as it is furnished to the Austin Transportation & Public Works Department, therefore a two-week delay is implemented to help ensure more accurate and complete results.

**Dataset insights and information from Ydata Profiling**



**1. Statistics of the Austin Dataset**

* Number of variables/ columns = 54
* Duplicate Values = 0 (This means that the dataset has no duplicate values)
* Missing values/cells = 1725084(21.6%)

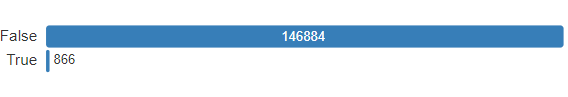
**2. Variables and their Types**

a. [Crash\_id](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_8062629517444432844) (Number)

* Unique values = 100%
* Null/ Missing Values = 0

b. [Crash\_fatal\_fl](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_511503428812552449) (Boolean)

* Distinct Values = 2
* Null Values = 0



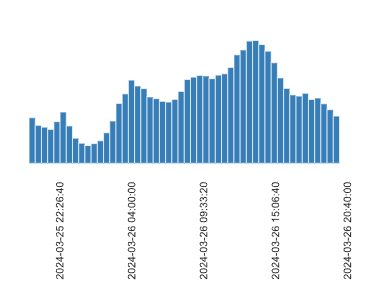
A large number of False values indicate that less proportion of the crash involved one or more fatalities

c.[crash\_date](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_4230749912199383634) (Date)

* Distinct Values = 144667 (97.9%)
* Null Values = 0

d. [Crash\_time](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-2724352977956582205) (Date)

* Distinct Values: 1440(1.0%)
* Missing/ Null Values: 0

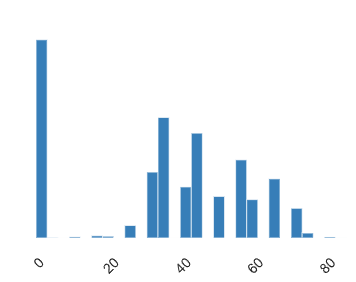


e. [case\_id](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_7544974539041288068) (Text)

* Distinct Values: 145678(99.9%)
* Missing/ Null Values: 1858(1.3%)

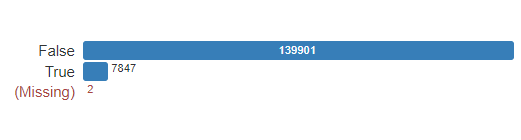
f. [Crash\_speed\_limit](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-4444400909102560329) (Number)

* Distinct Values: 28(< 0.1%)
* Null Values: 2 (<0.1%)
* Mean: 34.757526



g. [Road\_constr\_zone\_fl](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-286059902147586445) (Boolean)

* Distinct Values: 2(< 0.1%)
* Missing Values: 2(< 0.1%)



We can observe from the graph that majority of the crashes that occured, were not related to construction, maintenance, or utility work zone

h. [Latitude](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-7770175722083782698) (Number)

* Distinct Values: 96355(66.2%)
* Missing Values: 2243(1.5%)

i. [Longitude](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-5542726319050214150) (Number)

* Distinct Values: 96230(66.1%)
* Missing Values: 2243(1.5%)

j. [Street\_name](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-304667553749750569) (Text)

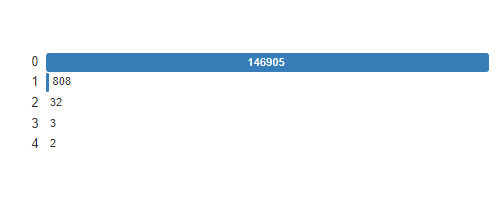
* Distinct: 4630(3.1%)
* Missing: 2(<0.1%)

k. [crash\_sev\_id](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_305674878568717201) (Number)

* Distinct Values: 8(< 0.1%)
* Missing Values: 0

l. [Death\_cnt](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-2120274068817753359) (Number)

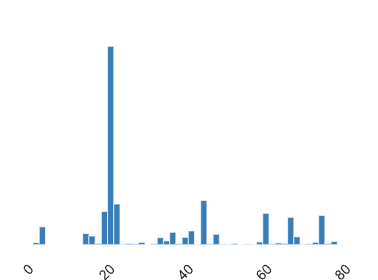
* Distinct Values: 5(<0.1%)
* Missing Values: 0



From the above graph, we can infer that in majority of the cases, the death count was 0, meaning most of the cases were not fatal

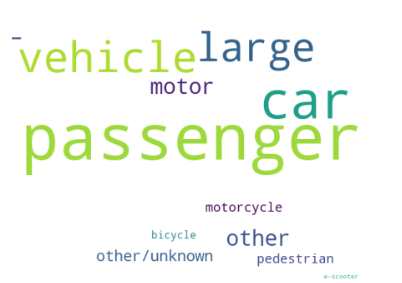
m. [contrib\_factr\_p1\_id](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-6250516818679765086) (Number)

* Distinct Values: 70(0.2%)
* Missing Values:119143(80.6%)



n. Units\_involved (Text)

* Distinct Values: 112(0.8%)
* Missing Values: 7(< 0.1%)

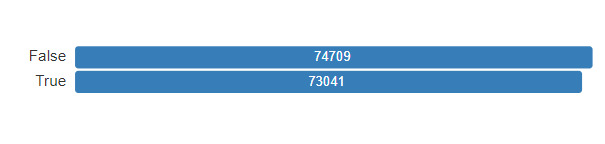


We get to know that the kind of units involved in crashes was Large passenger vehicle, Passenger car, Motor Vehicles, Motorcycle,pedestrian

|  |
| --- |

o. [Onsys\_fl](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_8065092790554518304) (Boolean)

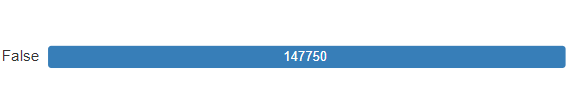
* Distinct Values: 2(< 0.1%)
* Missing Values: 0



We observe that half of the reported crashes occurred on the TxDOT highway system

p. [private\_dr\_fl](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-3532978264067053514) (Boolean)

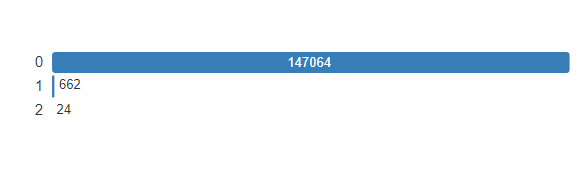
* Distinct Values: 1
* Missing Values: 0



It can be inferred that none of the reported crashes occurred on a private drive or road/private property/parking lot

q. [Motorcycle\_serious\_injury\_count](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_3255224391713412504) (Number)

* Distinct Values: 3(< 0.1%)
* Missing Values: 0



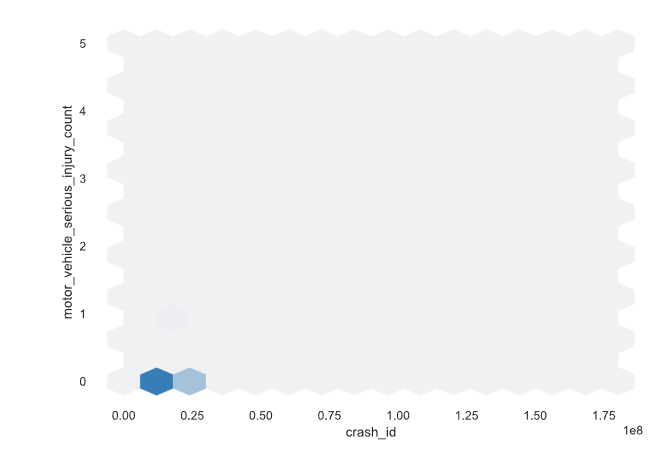
We can conclude that very few reported crashes had [motorcycle serious injury](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_3255224391713412504)

**3. Interactions**

These are hexbin plots, which are a bivariate analog of histograms and are used for

visualizing the structure in datasets of two variables. These graphs are useful for large datasets. Each hexagon represents a bin with a count of observations falling within it. The coloring of the hexagons typically reflects the number of observations, with darker colors representing higher counts.

Graph: Crash\_id vs [motorcycle\_serious\_injury\_count](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_3255224391713412504)



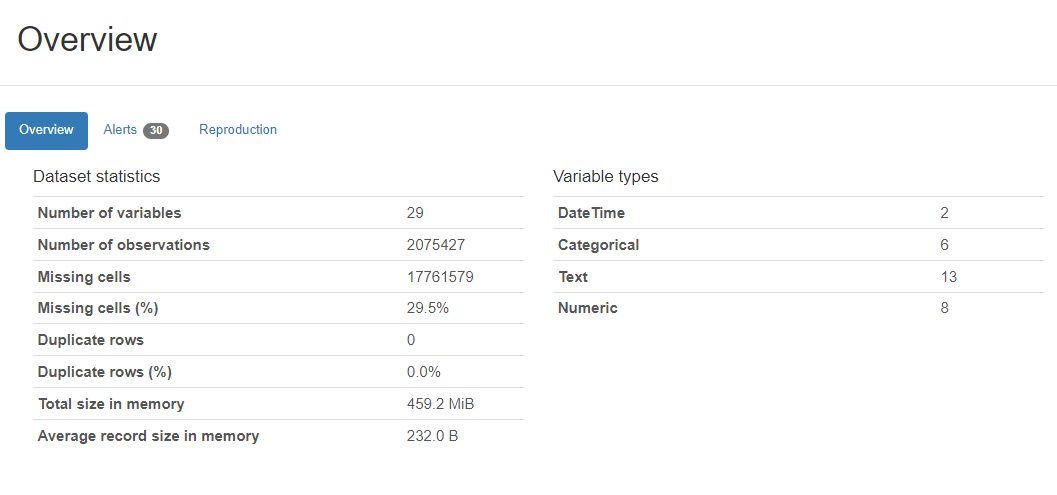
The concentration of darker hexagons along the lower range of 'Crash\_id'

the darkest bin is at a point where both Crash\_id and motorcycle\_serious\_injury\_count are relatively low, indicating a lower risk or severity associated with this type of crash

***New York Dataset***

The Motor Vehicle Collisions crash table contains details on the crash event. Each row represents a crash event. The Motor Vehicle Collisions data tables contain information from all police-reported motor vehicle collisions in NYC. The police report (MV104-AN) is required to be filled out for collisions where someone is injured or killed, or where there is at least $1000 worth of damage. ( It should be noted that the data is preliminary and subject to change when the MV-104AN forms are amended based on revised crash details

**Dataset insights and information from Ydata Profiling**



**1. Statistics of the New York Dataset -**

* Number of variables/ columns = 29
* Duplicate Values = 0 (This means that the dataset has no duplicate values)
* Missing values/cells = 17761579(29.5%)

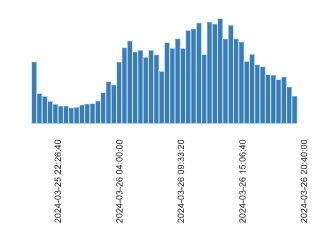
**2. Variables and their Types -**

a.[crash\_date](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_4230749912199383634) (Date)

* Distinct Values = 4283 (0.2%)
* Null Values = 0

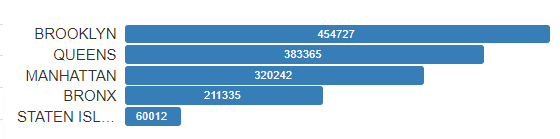
b. [Crash\_time](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-2724352977956582205) (Date)

* Distinct Values: 1440(1.0%)
* Missing/ Null Values: 0



c. [BOROUGH](http://localhost:8888/files/dadabi/NYC_Motor_Vehicle_Collision.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-3652720797137100151) (Text)

* Distinct Values: 5(< 0.1%)
* Missing/ Null Values: 645746(31.1%)



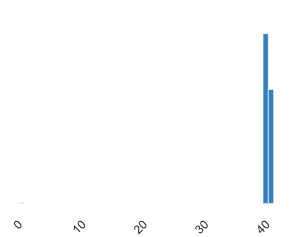
It can be inferred that the highest crashes occurred in Brooklyn and the lowest in Staten Island

d. [ZIP CODE](http://localhost:8888/files/dadabi/NYC_Motor_Vehicle_Collision.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-3134708091819391459) (Text)

* Distinct Values: 235(< 0.1%)
* Null Values: 645996 (31.1%)

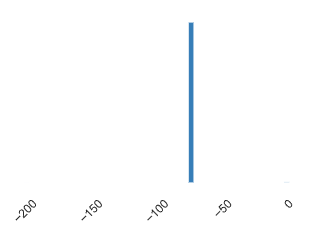
e. [Latitude](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-7770175722083782698) (Number)

* Distinct Values: 126594(66.2%)
* Missing Values: 2243(1.5%)



f. [Longitude](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-5542726319050214150) (Number)

* Distinct Values: 98351(5.3%)
* Missing Values:233626(11.3%)

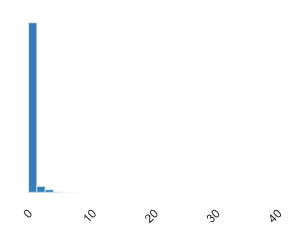


g. [ON STREET NAME](http://localhost:8888/files/dadabi/NYC_Motor_Vehicle_Collision.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-6366361020186796769) (Text)

* Distinct: 18410(1.1%)
* Missing: 440569(21.2%)

h. [NUMBER OF PERSONS INJURED](http://localhost:8888/files/dadabi/NYC_Motor_Vehicle_Collision.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-3675578601336335369) (Number)

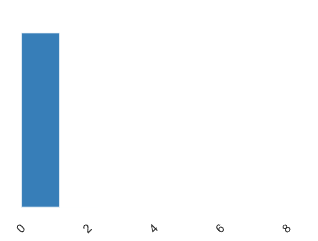
* Distinct Values: 32(< 0.1%)
* Missing Values: 18(<0.1%)



We can infer that less 5 people were injured for a reported incident

i. [NUMBER OF PERSONS KILLED](http://localhost:8888/files/dadabi/NYC_Motor_Vehicle_Collision.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-5849743554621528997) (Number)

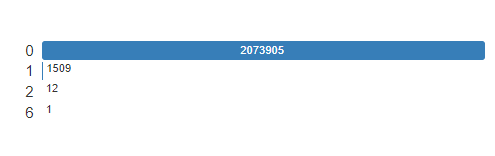
* Distinct Values: 7(<0.1%)
* Missing Values: 31(<0.1%)



We can see that overall less than 2 people were killed

j. [NUMBER OF PEDESTRIANS KILLED](http://localhost:8888/files/dadabi/NYC_Motor_Vehicle_Collision.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-2347588133968509666) (Number)

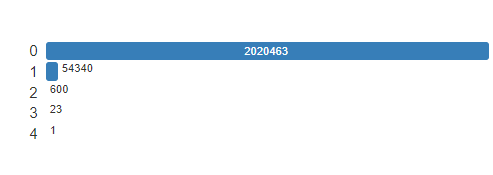
* Distinct Values: 4(<0.1%)
* Missing Values:0



We can observe that the portion of pedestrians killed was very less

k.[NUMBER OF CYCLIST INJURED](http://localhost:8888/files/dadabi/NYC_Motor_Vehicle_Collision.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-8953580185697914545) (Number)

* Distinct Values: 5(< 0.1%)
* Missing Values: 0

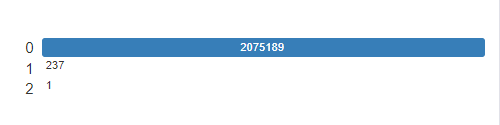


We can say that the majority of the cyclists were not injured, and very few were injured to varying degrees

|  |
| --- |

l. [NUMBER OF CYCLIST KILLED](http://localhost:8888/files/dadabi/NYC_Motor_Vehicle_Collision.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_7138066877296925588) (Number)

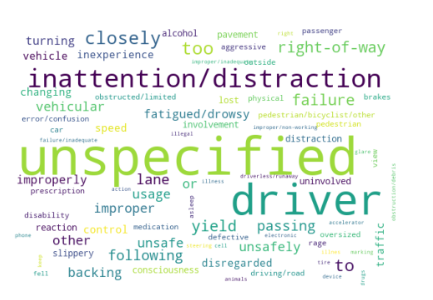
* Distinct Values: 3(< 0.1%)
* Missing Values: 0



Only a very few cyclists were killed

m. [CONTRIBUTING FACTOR VEHICLE 1](http://localhost:8888/files/dadabi/NYC_Motor_Vehicle_Collision.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_1631529107822336447) (Text)

* Distinct Values: 61(< 0.1%)
* Missing Values: 6802(0.3%)



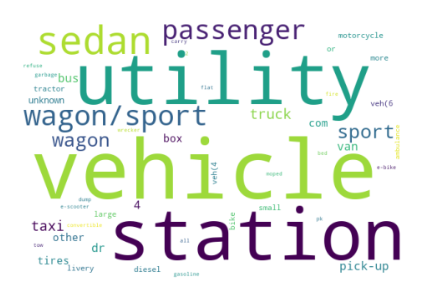
We can see the most common contributing factor 1 above

n. [COLLISION\_ID](http://localhost:8888/files/dadabi/NYC_Motor_Vehicle_Collision.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-1097198780355776366) (Number)

* Distinct Values: 2075427( 100%)
* Missing Values: 0

o. [VEHICLE TYPE CODE 1](http://localhost:8888/files/dadabi/NYC_Motor_Vehicle_Collision.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_4365054230512580298) (Number)

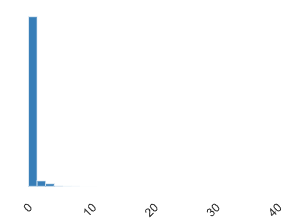
* Distinct Values: 1631(< 0.1%)
* Missing Values:13691(0.7%)



We can see some of the most commonly occurring [VEHICLE TYPE CODE 1](http://localhost:8888/files/dadabi/NYC_Motor_Vehicle_Collision.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_4365054230512580298) above

p. [NUMBER OF MOTORIST INJURED](http://localhost:8888/files/dadabi/NYC_Motor_Vehicle_Collision.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_3568190448261487108) (Number)

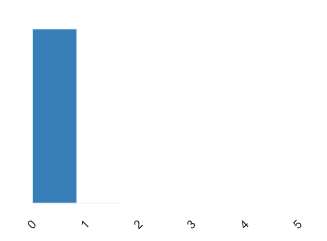
* Distinct Values: 31(< 0.1%)
* Missing Values: 0



It has been concluded that 10 motorists were injured

q. [NUMBER OF MOTORIST KILLED](http://localhost:8888/files/dadabi/NYC_Motor_Vehicle_Collision.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-7113287250596240054) (Number)

* Distinct Values: 6(< 0.1%)
* Missing Values: 0



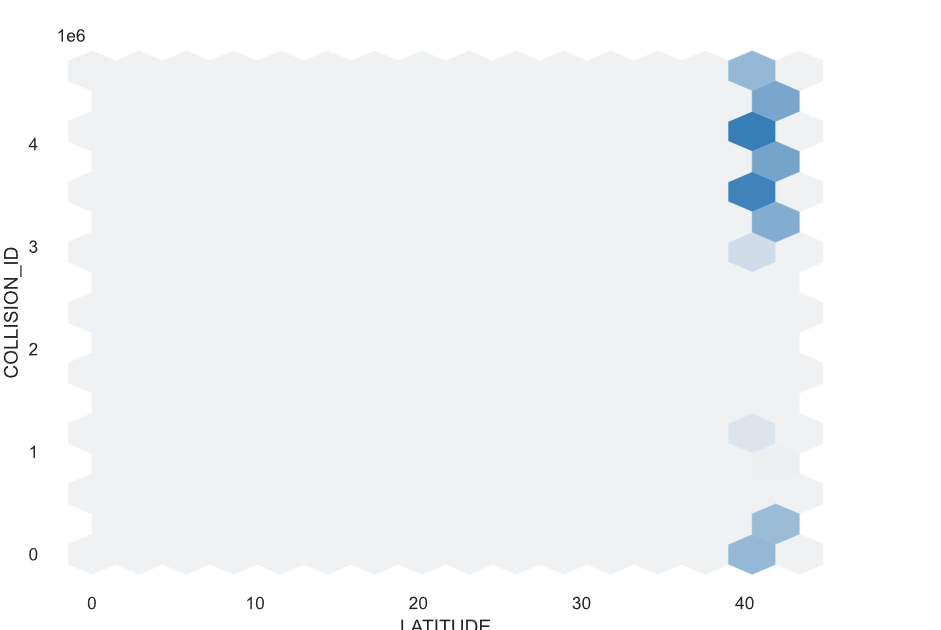
We can see that close 1 motorist was killed

**3. Interactions**

These are hexbin plots, which are a bivariate analog of histograms and are used for

visualizing the structure in datasets of two variables. These graphs are useful for large datasets. Each hexagon represents a bin with a count of observations falling within it. The coloring of the hexagons typically reflects the number of observations, with darker colors representing higher counts.

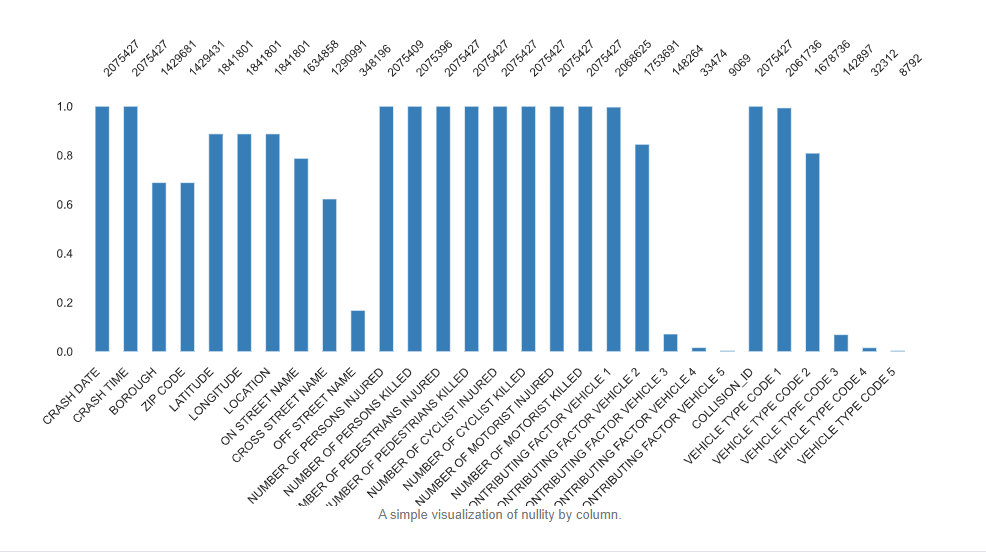
Graph: Latitude vs Collision\_id



The concentration of darker hexagons along the higher range of 'Collision\_id'

the darkest bin is at a point where both Collision\_id and Latitude are relatively high, indicating a higher risk or severity associated with this type of crash

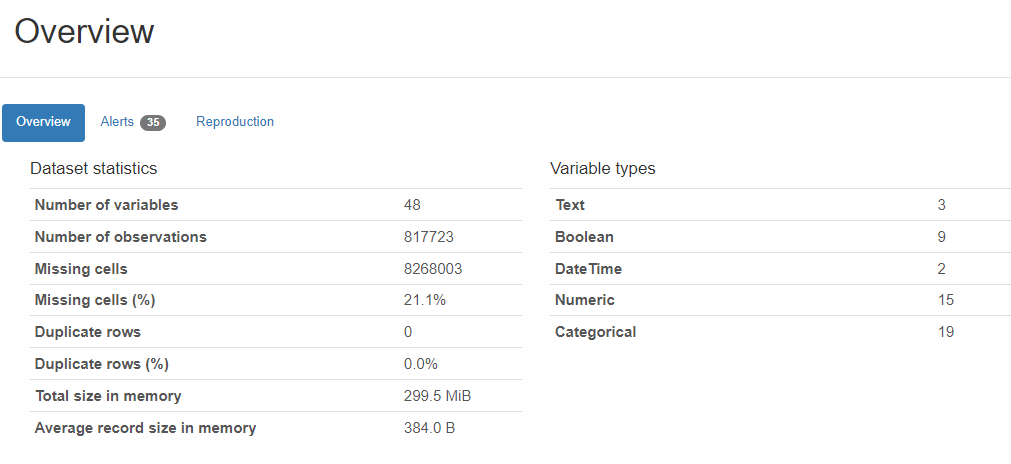
**4. Overall Null values**



***Chicago Dataset***

Chicago's crash data reflects incidents within city limits under CPD jurisdiction, sourced from E-Crash without personal identifiers. Self-reported and officer-recorded crashes, post-amendments, are included from 2015 onward. Parameters like street conditions and speed limits are officer-entered but may differ from actual conditions. Non-CPD-attended incidents and minor crashes are excluded. Data adhere to Illinois SR1050 format and statute for reportable crashes exceeding $1,500 damage or causing bodily harm on public roads involving moving vehicles (excluding bike dooring).

**Dataset insights and information from Ydata Profiling**



**1. Statistics of the Chicago Dataset**

* Number of variables/ columns = 54
* Duplicate Values = 0 (This means that the dataset has no duplicate values)
* Missing values/cells = 1725084(21.6%)

**2. Variables and their Types**

a. [CRASH\_RECORD\_ID](http://localhost:8888/files/Chicago_updated.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_5478876936627811781) (Number)

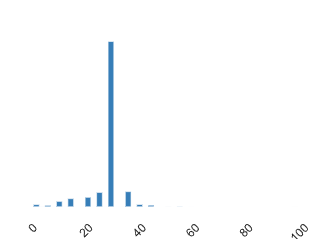
* Unique values = 817723(100%)
* Null/ Missing Values = 0

b. [CRASH\_DATE](http://localhost:8888/files/Chicago_updated.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_148211476398693882) (Date)

* Distinct Values = 536888(65.7%)
* Null Values = 0

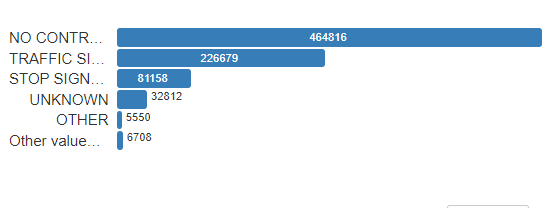
c.[POSTED\_SPEED\_LIMIT](http://localhost:8888/files/Chicago_updated.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_6959317117895227829) (Number)

* Distinct Values = 46(< 0.1%)
* Null Values = 0



It can be inferred that the posted speed limit was around 30 for most of the crashes

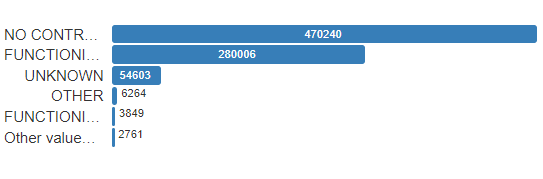
d. [TRAFFIC\_CONTROL\_DEVICE](http://localhost:8888/files/Chicago_updated.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_8392332745428515210) (Text)

* Distinct Values: 19(<0.1%)
* Missing/ Null Values: 0

It can be concluded that No controls and Traffic signal were the most used Traffic Control device

e. [DEVICE\_CONDITION](http://localhost:8888/files/Chicago_updated.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_3784990164001083652) (Text)

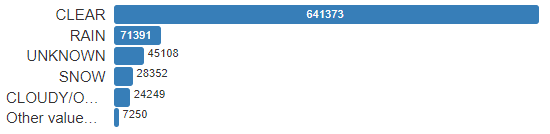
* Distinct Values: 8(< 0.1%)
* Missing/ Null Values: 0



We observe that No Control was the most prevalent device condition

f. [WEATHER\_CONDITION](http://localhost:8888/files/Chicago_updated.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-8819701442975962511) (Text)

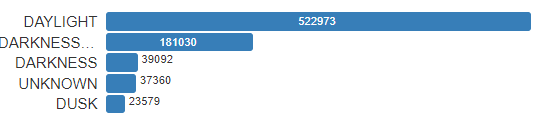
* Distinct Values: 12(< 0.1%)
* Null Values: 0



We can see that most crashes occur in clear weather

g. [LIGHTING\_CONDITION](http://localhost:8888/files/Chicago_updated.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-5297975436318682079) (Text)

* Distinct Values: 6(< 0.1%)
* Missing Values: 0



We can observe from the graph that the majority of the crashes occurred in daylight

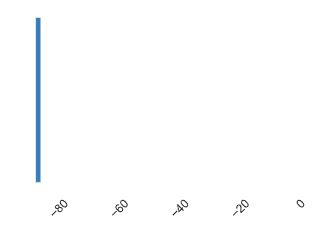
h. [Latitude](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-7770175722083782698) (Number)

* Distinct Values: 300091(37.0%)
* Missing Values: 5615(0.7%)



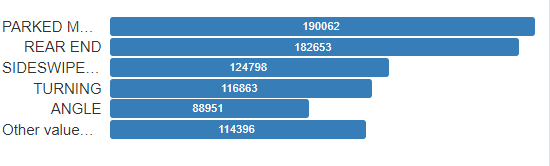
i. [Longitude](http://localhost:8888/files/dadabi/Austin%20Crash%20Data.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-5542726319050214150) (Number)

* Distinct Values: 300054(36.9%)
* Missing Values: 5615(0.7%)



j. [FIRST\_CRASH\_TYPE](http://localhost:8888/files/Chicago_updated.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-7734807368813384725) (Text)

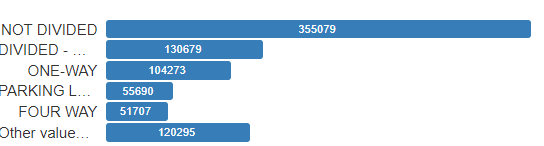
* Distinct: 18(< 0.1%)
* Missing: 0



We see that the first crash type is almost uniformly distributed for the reported crashes

k. [TRAFFICWAY\_TYPE](http://localhost:8888/files/Chicago_updated.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-6984484905137364408) (Text)

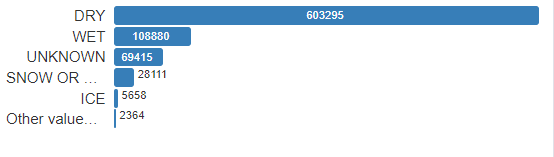
* Distinct Values: 20(< 0.1%)
* Missing Values: 0



We get to know that the majority of the crashes occurred in the not divided traffic way

l.[ROADWAY\_SURFACE\_COND](http://localhost:8888/files/Chicago_updated.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_8387671976944295281) (Text)

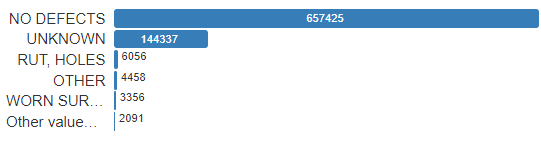
* Distinct Values: 7(<0.1%)
* Missing Values: 0



From the above graph, we can infer that in the majority of the cases, The road surface condition was dry

m. [ROAD\_DEFECT](http://localhost:8888/files/Chicago_updated.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_1228952997652438848) (Number)

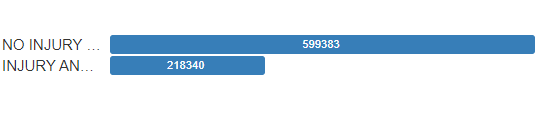
* Distinct Values: 7(<0.1%)
* Missing Values:0



Majority of the accidents occurred on the defect free roads

n.[CRASH\_TYPE](http://localhost:8888/files/Chicago_updated.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-5668625279846394632) (Text)

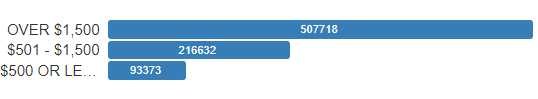
* Distinct Values: 2(0.8%)
* Missing Values: 7(< 0.1%)



We get to know that the type of severity of the collision in the majority of the crashes was no injury or drive-away

o.[DAMAGE](http://localhost:8888/files/Chicago_updated.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-7901118674464541333) (Text)

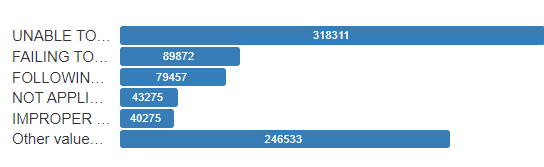
* Distinct Values: 3(< 0.1%)
* Missing Values: 0



We observe that in most of the crashes, the damaged cost was over $1500

p. [PRIM\_CONTRIBUTORY\_CAUSE](http://localhost:8888/files/Chicago_updated.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-6624414738326520447) (Text)

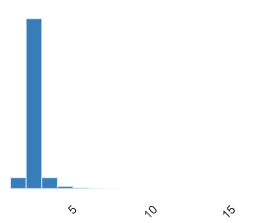
* Distinct Values: 40(< 0.1%)
* Missing Values: 0



It can be inferred that ‘Unable to determine’ was the most entered primary contributory cause

q.[NUM\_UNITS](http://localhost:8888/files/Chicago_updated.html?_xsrf=2%7Cab5aa700%7Cfb335e5cc8c7b5bf64f9de12d6933ee6%7C1709831693#pp_var_-6611276319895357180) (Number)

* Distinct Values: 17(< 0.1%)
* Missing Values: 0

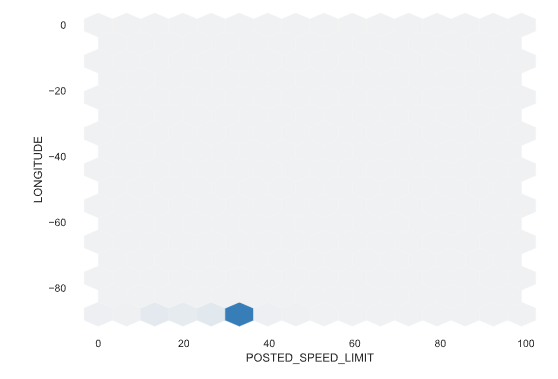


We can conclude that 5 or fewer units were involved in the crashes overall

**3. Interactions**

These are hexbin plots, which are a bivariate analog of histograms and are used for visualizing the structure in datasets of two variables. These graphs are useful for large datasets. Each hexagon represents a bin with a count of observations falling within it. The coloring of the hexagons typically reflects the number of observations, with darker colors representing higher counts.

Graph : posted\_speed\_limit vs Longitude

****

The concentration of darker hexagons along the lower range of 'posted\_speed\_limit'

in the darkest bin is at a point where both posted\_speed\_limit and Longitude are relatively low, indicating a lower risk or severity associated with this type of crash