Overview

Dataset statistics

Number of variables	29
Number of observations	2075427
Missing cells	17761579
Missing cells (%)	29.5%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	459.2 MiB
Average record size in memory	232.0 B

Variable types

DateTime	2
Categorical	6
Text	13
Numeric	8

Alerts

NUMBER OF PEDESTRIANS KILLED is highly imbalanced (99.6%)	Imbalance
NUMBER OF CYCLIST INJURED is highly imbalanced (92.3%)	Imbalance
NUMBER OF CYCLIST KILLED is highly imbalanced (99.9%)	Imbalance
CONTRIBUTING FACTOR VEHICLE 4 is highly imbalanced (90.8%)	Imbalance
CONTRIBUTING FACTOR VEHICLE 5 is highly imbalanced (89.9%)	Imbalance
BOROUGH has 645746 (31.1%) missing values	Missing
ZIP CODE has 645996 (31.1%) missing values	Missing
LATITUDE has 233626 (11.3%) missing values	Missing
LONGITUDE has 233626 (11.3%) missing values	Missing
LOCATION has 233626 (11.3%) missing values	Missing
ON STREET NAME has 440569 (21.2%) missing values	Missing
CROSS STREET NAME has 784436 (37.8%) missing values	Missing
OFF STREET NAME has 1727231 (83.2%) missing values	Missing
CONTRIBUTING FACTOR VEHICLE 2 has 321736 (15.5%) missing values	Missing
CONTRIBUTING FACTOR VEHICLE 3 has 1927163 (92.9%) missing values	Missing
CONTRIBUTING FACTOR VEHICLE 4 has 2041953 (98.4%) missing values	Missing
CONTRIBUTING FACTOR VEHICLE 5 has 2066358 (99.6%) missing values	Missing
VEHICLE TYPE CODE 2 has 396691 (19.1%) missing values	Missing
VEHICLE TYPE CODE 3 has 1932530 (93.1%) missing values	Missing
VEHICLE TYPE CODE 4 has 2043115 (98.4%) missing values	Missing
VEHICLE TYPE CODE 5 has 2066635 (99.6%) missing values	Missing
LATITUDE is highly skewed (γ1 = -20.43042564)	Skewed
NUMBER OF PERSONS KILLED is highly skewed (γ1 = 33.71743399)	Skewed
NUMBER OF MOTORIST KILLED is highly skewed (γ1 = 54.74414747)	Skewed
COLLISION_ID has unique values	Unique
NUMBER OF PERSONS INJURED has 1601221 (77.2%) zeros	Zeros
NUMBER OF PERSONS KILLED has 2072415 (99.9%) zeros	Zeros
NUMBER OF PEDESTRIANS INJURED has 1962919 (94.6%) zeros	Zeros

NUMBER OF MOTORIST INJURED has 1772939 (85.4%) zeros	Zeros
NUMBER OF MOTORIST KILLED has 2074246 (99.9%) zeros	Zeros

Reproduction

Analysis started	2024-04-08 03:25:54.954376
Analysis finished	2024-04-08 03:26:54.322263
Duration	59.37 seconds
Software version	ydata-profiling vv4.7.0 (https://github.com/ydataai/ydata-profiling)
Download configuration	config.json (data:text/plain;charset=utf-8,%7B%22title%22%3A%20%22Austin%20Crash%20Report%20Data%20-%20Profiling%20Report%22%2C%20%22dataset%22%3A%20%7B%22description%22%3A%20%22%2C%20%22creator%22%3A%20%22%22%2C%20%22author%22%3A%20%22%22%2C%20%22copyright_holder%22%3A%20%22%22

Variables
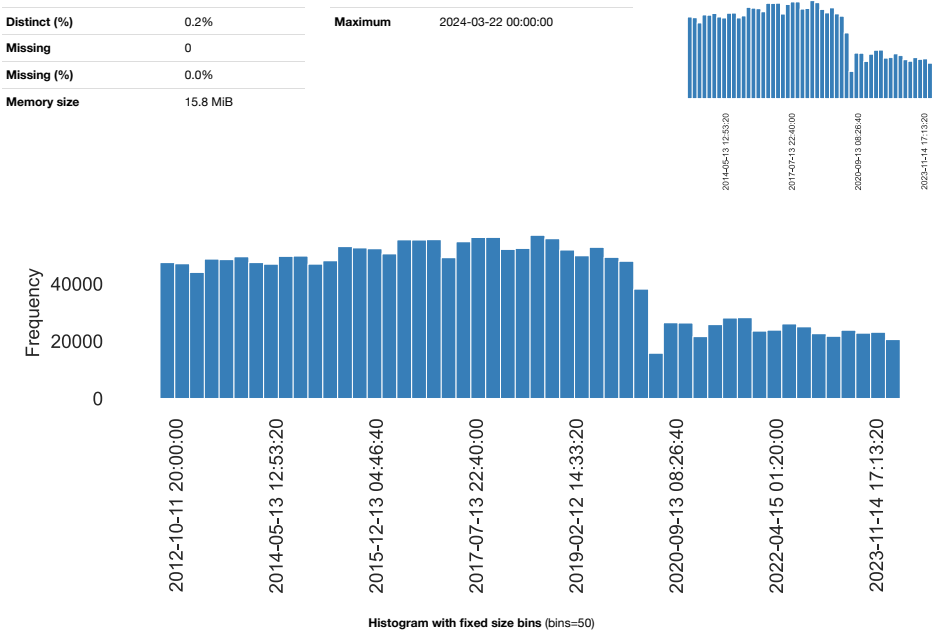
Select Columns ▼

CRASH DATE
Date

Distinct	4283	**Minimum**	2012-07-01 00:00:00	
Distinct (%)	0.2%	**Maximum**	2024-03-22 00:00:00	
Missing	0			
Missing (%)	0.0%			
Memory size	15.8 MiB			

Histogram with fixed size bins (bins=50)

CRASH TIME
Date

Distinct	1440	**Minimum**	2024-04-07 00:00:00	
Distinct (%)	0.1%	**Maximum**	2024-04-07 23:59:00	
Missing	0			
Missing (%)	0.0%			
Memory size	15.8 MiB			
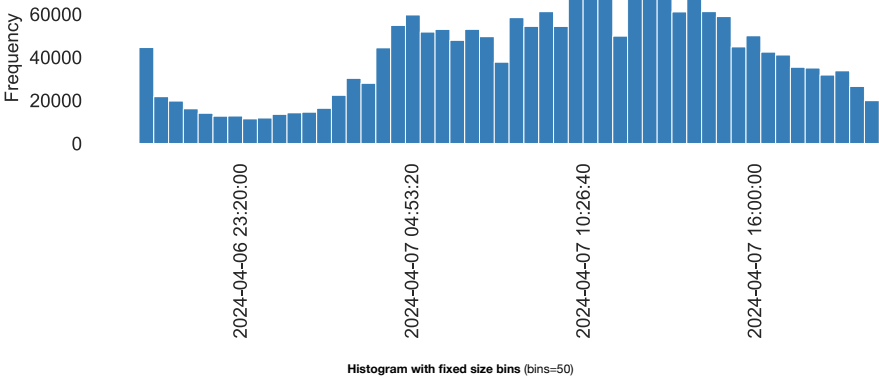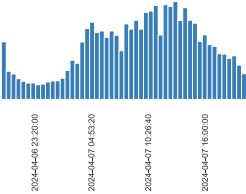
Histogram with fixed size bins (bins=50)

BOROUGH
Categorical

`MISSING`

Distinct	5
Distinct (%)	< 0.1%
Missing	645746
Missing (%)	31.1%
Memory size	15.8 MiB

Length

Max length	13
Median length	9
Mean length	7.4541209
Min length	5

Characters and Unicode

Total characters	10657015
Distinct characters	19
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category) ?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode) ?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block) ?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

Sample

1st row	BROOKLYN
2nd row	BROOKLYN
3rd row	BRONX
4th row	BROOKLYN
5th row	MANHATTAN

Common Values

Value	Count	Frequency (%)
BROOKLYN	454727	21.9%
QUEENS	383365	18.5%
MANHATTAN	320242	15.4%
BRONX	211335	10.2%
STATEN ISLAND	60012	2.9%
(Missing)	645746	31.1%

Length

Histogram of lengths of the category

Common Values (Plot)
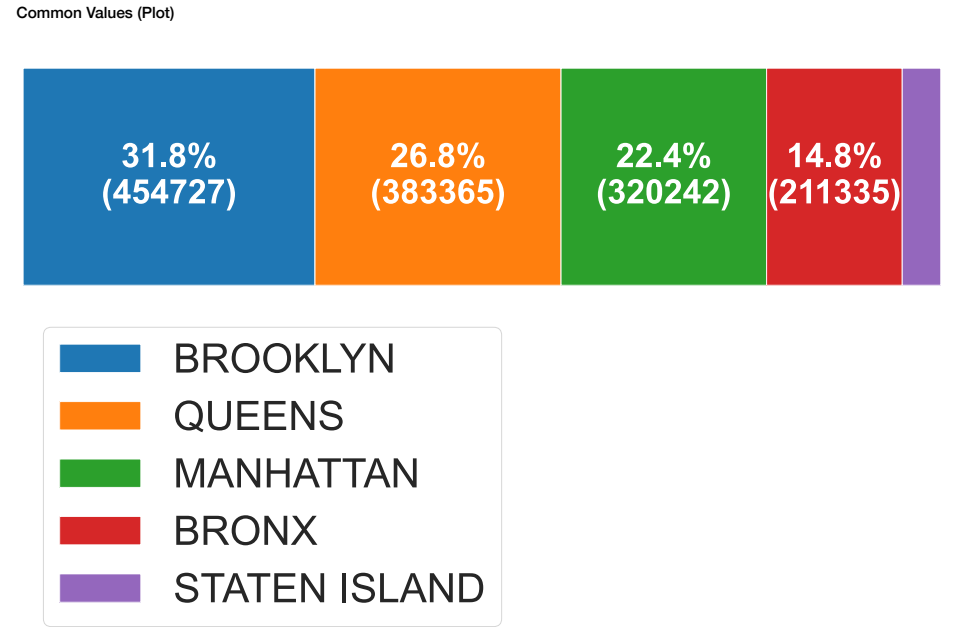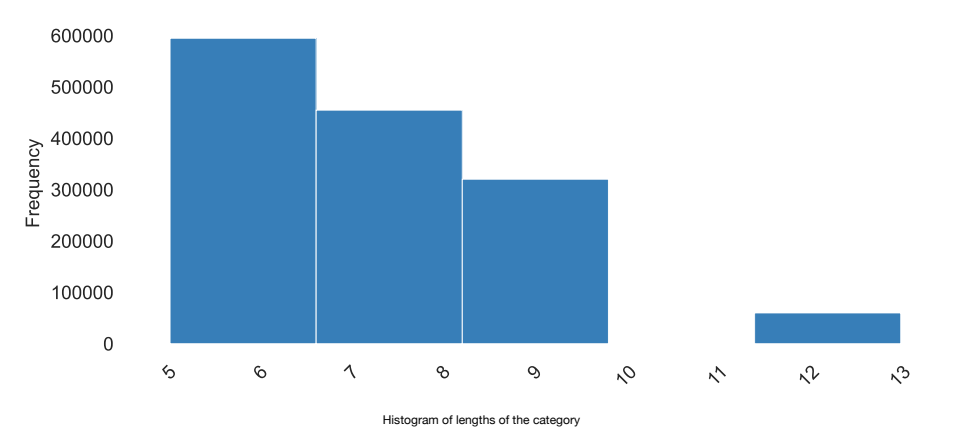
Value	Count	Frequency (%)
brooklyn	454727	30.5%
queens	383365	25.7%
manhattan	320242	21.5%
bronx	211335	14.2%
staten	60012	4.0%

Value	Count	Frequency (%)
island	60012	4.0%

Most occurring characters

Value	Count	Frequency (%)
N	1809935	17.0%
O	1120789	10.5%
A	1080750	10.1%
E	826742	7.8%
T	760508	7.1%
R	666062	6.2%
B	666062	6.2%
L	514739	4.8%
S	503389	4.7%
Y	454727	4.3%
Other values (9)	2253312	21.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	10657015	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
N	1809935	17.0%
O	1120789	10.5%
A	1080750	10.1%
E	826742	7.8%
T	760508	7.1%
R	666062	6.2%
B	666062	6.2%
L	514739	4.8%
S	503389	4.7%
Y	454727	4.3%
Other values (9)	2253312	21.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	10657015	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
N	1809935	17.0%
O	1120789	10.5%
A	1080750	10.1%
E	826742	7.8%
T	760508	7.1%
R	666062	6.2%
B	666062	6.2%
L	514739	4.8%
S	503389	4.7%
Y	454727	4.3%
Other values (9)	2253312	21.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	10657015	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
N	1809935	17.0%
O	1120789	10.5%
A	1080750	10.1%
E	826742	7.8%
T	760508	7.1%
R	666062	6.2%
B	666062	6.2%
L	514739	4.8%
S	503389	4.7%
Y	454727	4.3%
Other values (9)	2253312	21.1%

ZIP CODE
Text

MISSING

Distinct	235
Distinct (%)	< 0.1%
Missing	645996
Missing (%)	31.1%
Memory size	15.8 MiB

Length

Max length	5
Median length	5
Mean length	5
Min length	5

Characters and Unicode

Total characters	7147155	
Distinct characters	11	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	5	?
Unique (%)	< 0.1%	

Sample

1st row	11208
2nd row	11233
3rd row	10475
4th row	11207
5th row	10017

Value	Count	Frequency (%)
11207	27789	1.9%
11236	19259	1.3%
11101	19220	1.3%
11203	18372	1.3%
11234	18011	1.3%
11385	17924	1.3%
11208	17312	1.2%
10019	17258	1.2%
11212	17236	1.2%
11201	17146	1.2%
Other values (224)	1239862	86.7%

Most occurring characters

Value	Count	Frequency (%)
1	2772280	38.8%
0	1267960	17.7%
2	835731	11.7%
3	624685	8.7%
4	511341	7.2%
6	319786	4.5%
5	281189	3.9%
7	243220	3.4%
8	151585	2.1%
9	139168	1.9%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	7147155	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
1	2772280	38.8%
0	1267960	17.7%
2	835731	11.7%
3	624685	8.7%
4	511341	7.2%
6	319786	4.5%
5	281189	3.9%
7	243220	3.4%
8	151585	2.1%
9	139168	1.9%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	7147155	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
1	2772280	38.8%
0	1267960	17.7%
2	835731	11.7%
3	624685	8.7%
4	511341	7.2%
6	319786	4.5%
5	281189	3.9%
7	243220	3.4%
8	151585	2.1%
9	139168	1.9%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	7147155	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
1	2772280	38.8%
0	1267960	17.7%
2	835731	11.7%
3	624685	8.7%
4	511341	7.2%
6	319786	4.5%
5	281189	3.9%
7	243220	3.4%
8	151585	2.1%
9	139168	1.9%

LATITUDE
Real number (ℝ)

MISSING   SKEWED

Distinct	126594	**Minimum**	0
Distinct (%)	6.9%	**Maximum**	43.344444
Missing	233626	**Zeros**	4360
Missing (%)	11.3%	**Zeros (%)**	0.2%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	40.627693	**Memory size**	15.8 MiB

Quantile statistics

Minimum	0
5-th percentile	40.596622
Q1	40.6678
median	40.72083
Q3	40.769592
95-th percentile	40.86205
Maximum	43.344444
Range	43.344444
Interquartile range (IQR)	0.101792

Descriptive statistics

Standard deviation	1.9806568
Coefficient of variation (CV)	0.048751397
Kurtosis	416.08064
Mean	40.627693
Median Absolute Deviation (MAD)	0.051354
Skewness	-20.430426
Sum	74828126
Variance	3.9230014
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=50)
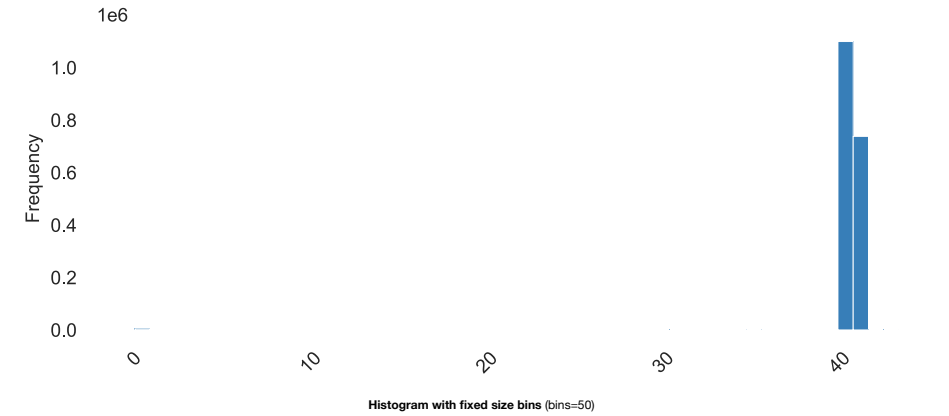
Value	Count	Frequency (%)
0	4360	0.2%
40.861862	883	< 0.1%
40.696033	762	< 0.1%
40.8047	692	< 0.1%
40.608757	671	< 0.1%

Value	Count	Frequency (%)
40.798256	627	< 0.1%
40.759308	622	< 0.1%
40.6960346	587	< 0.1%
40.675735	557	< 0.1%
40.658577	520	< 0.1%
Other values (126584)	1831520	88.2%
(Missing)	233626	11.3%

Value	Count	Frequency (%)
0	4360	0.2%
30.78418	1	< 0.1%
34.783634	1	< 0.1%
40.4989488	2	< 0.1%
40.4991346	1	< 0.1%
40.49931	1	< 0.1%
40.4994787	1	< 0.1%
40.499659	1	< 0.1%
40.49971	1	< 0.1%
40.49984	1	< 0.1%

Value	Count	Frequency (%)
43.344444	1	< 0.1%
42.64154	1	< 0.1%
42.318317	1	< 0.1%
42.107204	1	< 0.1%
41.91661	1	< 0.1%
41.34796	1	< 0.1%
41.258785	1	< 0.1%
41.12615	5	< 0.1%
41.12421	1	< 0.1%
41.061634	2	< 0.1%

LONGITUDE
Real number (ℝ)

MISSING

Distinct	98351	**Minimum**	-201.35999	
Distinct (%)	5.3%	**Maximum**	0	
Missing	233626	**Zeros**	4360	
Missing (%)	11.3%	**Zeros (%)**	0.2%	
Infinite	0	**Negative**	1837441	
Infinite (%)	0.0%	**Negative (%)**	88.5%	
Mean	-73.752129	**Memory size**	15.8 MiB	
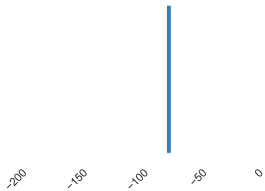
Quantile statistics

Minimum	-201.35999
5-th percentile	-74.03607
Q1	-73.97484
median	-73.92726
Q3	-73.866731
95-th percentile	-73.763239
Maximum	0
Range	201.35999
Interquartile range (IQR)	0.1081089

Descriptive statistics

Standard deviation	3.7233454
Coefficient of variation (CV)	-0.050484581
Kurtosis	440.66
Mean	-73.752129
Median Absolute Deviation (MAD)	0.0526217
Skewness	16.099628
Sum	-1.3583675×10^{8}
Variance	13.863301
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=50)
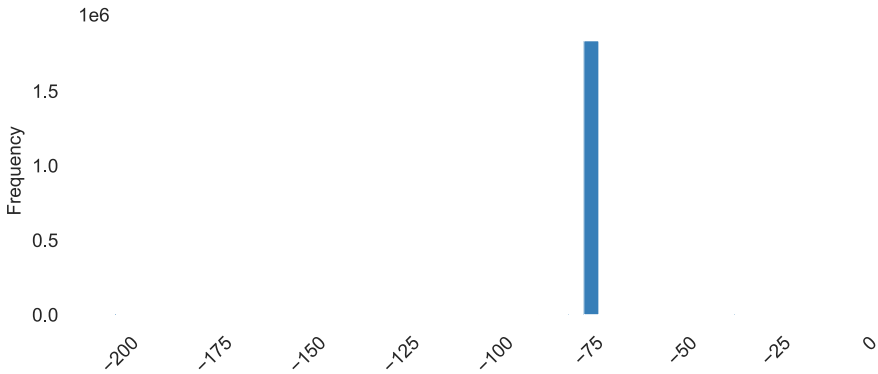
Value	Count	Frequency (%)
0	4360	0.2%
-73.89063	763	< 0.1%
-73.91282	719	< 0.1%
-73.98453	699	< 0.1%
-74.038086	672	< 0.1%

Value	Count	Frequency (%)
-73.89686	657	< 0.1%
-73.91243	654	< 0.1%
-73.9845292	587	< 0.1%
-73.94476	583	< 0.1%
-73.9112	576	< 0.1%
Other values (98341)	1831531	88.2%
(Missing)	233626	11.3%

Value	Count	Frequency (%)
-201.35999	1	< 0.1%
-201.23706	105	< 0.1%
-89.13527	1	< 0.1%
-86.76847	1	< 0.1%
-79.61955	1	< 0.1%
-79.00183	1	< 0.1%
-76.2634	1	< 0.1%
-76.02163	1	< 0.1%
-74.742	7	< 0.1%
-74.25496	1	< 0.1%

Value	Count	Frequency (%)
0	4360	0.2%
-32.768513	16	< 0.1%
-47.209625	3	< 0.1%
-73.66301	1	< 0.1%
-73.70055	2	< 0.1%
-73.700584	11	< 0.1%
-73.7005968	10	< 0.1%
-73.70061	4	< 0.1%
-73.70071	4	< 0.1%
-73.70073	1	< 0.1%

LOCATION
Text

MISSING

Distinct	283006
Distinct (%)	15.4%
Missing	233626
Missing (%)	11.3%
Memory size	15.8 MiB

Length

Max length	25
Median length	24
Mean length	22.779989
Min length	10

Characters and Unicode

Total characters	41956206
Distinct characters	16
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category) ?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode) ?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block) ?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	155498	?
Unique (%)	8.4%	

Sample

1st row	(40.667202, -73.8665)
2nd row	(40.683304, -73.917274)
3rd row	(40.709183, -73.956825)
4th row	(40.86816, -73.83148)
5th row	(40.67172, -73.8971)

Value	Count	Frequency (%)
0.0	8720	0.2%
40.861862	883	< 0.1%
73.89063	763	< 0.1%
40.696033	762	< 0.1%
73.91282	719	< 0.1%
73.98453	699	< 0.1%
40.8047	692	< 0.1%
74.038086	672	< 0.1%
40.608757	671	< 0.1%
73.89686	657	< 0.1%
Other values (224934)	3668364	99.6%

Most occurring characters

Value	Count	Frequency (%)
7	4595577	11.0%
4	3980471	9.5%
.	3683602	8.8%
3	3498540	8.3%
0	3400841	8.1%
9	2700094	6.4%
8	2648683	6.3%
6	2616640	6.2%
5	2094509	5.0%
(	1841801	4.4%
Other values (6)	10895448	26.0%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	41956206	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
7	4595577	11.0%
4	3980471	9.5%
.	3683602	8.8%
3	3498540	8.3%
0	3400841	8.1%
9	2700094	6.4%
8	2648683	6.3%
6	2616640	6.2%
5	2094509	5.0%
(	1841801	4.4%
Other values (6)	10895448	26.0%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	41956206	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
7	4595577	11.0%
4	3980471	9.5%
.	3683602	8.8%
3	3498540	8.3%
0	3400841	8.1%
9	2700094	6.4%
8	2648683	6.3%
6	2616640	6.2%
5	2094509	5.0%
(	1841801	4.4%
Other values (6)	10895448	26.0%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	41956206	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
7	4595577	11.0%
4	3980471	9.5%
.	3683602	8.8%
3	3498540	8.3%
0	3400841	8.1%
9	2700094	6.4%
8	2648683	6.3%
6	2616640	6.2%
5	2094509	5.0%
(	1841801	4.4%
Other values (6)	10895448	26.0%

ON STREET NAME
Text

`MISSING`

Distinct	18410
Distinct (%)	1.1%
Missing	440569
Missing (%)	21.2%
Memory size	15.8 MiB

Length

Max length	32
Median length	32
Mean length	29.630325
Min length	2

Characters and Unicode

Total characters	48441374	
Distinct characters	75	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	6537	?
Unique (%)	0.4%	

Sample

1st row	WHITESTONE EXPRESSWAY
2nd row	QUEENSBORO BRIDGE UPPER
3rd row	THROGS NECK BRIDGE
4th row	SARATOGA AVENUE
5th row	MAJOR DEEGAN EXPRESSWAY RAMP

Value	Count	Frequency (%)
avenue	608264	16.1%
street	520901	13.8%
east	153481	4.1%
boulevard	127014	3.4%
west	114792	3.0%
parkway	74643	2.0%
road	68123	1.8%
expressway	63293	1.7%
island	30410	0.8%
queens	27154	0.7%
Other values (5393)	1983965	52.6%

Most occurring characters

Value	Count	Frequency (%)
	27562630	56.9%
E	3672854	7.6%
A	1951050	4.0%
T	1831929	3.8%
R	1669600	3.4%
N	1427915	2.9%
S	1407885	2.9%
U	977757	2.0%
O	868930	1.8%
V	852133	1.8%
Other values (65)	6218691	12.8%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	48441374	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
	27562630	56.9%
E	3672854	7.6%
A	1951050	4.0%
T	1831929	3.8%
R	1669600	3.4%
N	1427915	2.9%
S	1407885	2.9%
U	977757	2.0%
O	868930	1.8%
V	852133	1.8%
Other values (65)	6218691	12.8%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	48441374	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
	27562630	56.9%
E	3672854	7.6%
A	1951050	4.0%
T	1831929	3.8%
R	1669600	3.4%
N	1427915	2.9%
S	1407885	2.9%
U	977757	2.0%
O	868930	1.8%
V	852133	1.8%
Other values (65)	6218691	12.8%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	48441374	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
	27562630	56.9%
E	3672854	7.6%
A	1951050	4.0%
T	1831929	3.8%
R	1669600	3.4%
N	1427915	2.9%
S	1407885	2.9%
U	977757	2.0%
O	868930	1.8%
V	852133	1.8%
Other values (65)	6218691	12.8%

CROSS STREET NAME
Text

MISSING

Distinct	20236
Distinct (%)	1.6%
Missing	784436
Missing (%)	37.8%
Memory size	15.8 MiB

Length

Max length	32
Median length	32
Mean length	22.706216
Min length	1

Characters and Unicode

Total characters	29313520
Distinct characters	76
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category) ?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode) ?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block) ?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	6201	?
Unique (%)	0.5%	

Sample

1st row	20 AVENUE
2nd row	DECATUR STREET
3rd row	EAST 43 STREET
4th row	EAST GATE PLAZA
5th row	west 80 street -west 81 street

Value	Count	Frequency (%)
avenue	565307	19.8%
street	459527	16.1%
east	112172	3.9%
west	71155	2.5%
boulevard	68647	2.4%
road	55544	1.9%
place	33946	1.2%
parkway	26605	0.9%
3	18757	0.7%
park	17426	0.6%
Other values (5483)	1426325	50.0%

Most occurring characters

Value	Count	Frequency (%)
	14115616	48.2%
E	2937153	10.0%
T	1453458	5.0%
A	1419427	4.8%
R	1147248	3.9%
N	1074756	3.7%
S	988831	3.4%
U	777244	2.7%
V	708819	2.4%
O	578382	2.0%
Other values (66)	4112586	14.0%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	29313520	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
	14115616	48.2%
E	2937153	10.0%
T	1453458	5.0%
A	1419427	4.8%
R	1147248	3.9%
N	1074756	3.7%
S	988831	3.4%
U	777244	2.7%
V	708819	2.4%
O	578382	2.0%
Other values (66)	4112586	14.0%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	29313520	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
	14115616	48.2%
E	2937153	10.0%
T	1453458	5.0%
A	1419427	4.8%
R	1147248	3.9%
N	1074756	3.7%
S	988831	3.4%
U	777244	2.7%
V	708819	2.4%
O	578382	2.0%
Other values (66)	4112586	14.0%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	29313520	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
	14115616	48.2%
E	2937153	10.0%
T	1453458	5.0%
A	1419427	4.8%
R	1147248	3.9%
N	1074756	3.7%
S	988831	3.4%
U	777244	2.7%
V	708819	2.4%
O	578382	2.0%
Other values (66)	4112586	14.0%

OFF STREET NAME
Text

`MISSING`

Distinct	225845
Distinct (%)	64.9%
Missing	1727231
Missing (%)	83.2%
Memory size	15.8 MiB

Length

Max length	40
Median length	40
Mean length	36.021158
Min length	8

Characters and Unicode

Total characters	12542423
Distinct characters	84
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category) ?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode) ?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block) ?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	176197	?
Unique (%)	50.6%	

Sample

1st row	1211 LORING AVENUE
2nd row	344 BAYCHESTER AVENUE
3rd row	2047 PITKIN AVENUE
4th row	480 DEAN STREET
5th row	878 FLATBUSH AVENUE

Value	Count	Frequency (%)
avenue	137975	11.9%
street	125856	10.9%
east	33204	2.9%
west	23966	2.1%
boulevard	22127	1.9%
road	16430	1.4%
lot	7881	0.7%
parking	7267	0.6%
of	6949	0.6%
parkway	6943	0.6%
Other values (27589)	769819	66.5%

Most occurring characters

Value	Count	Frequency (%)
	6866584	54.7%
E	796771	6.4%
T	436287	3.5%
A	408734	3.3%
R	339643	2.7%
N	298626	2.4%
S	285926	2.3%
1	276924	2.2%
U	203017	1.6%
V	189426	1.5%
Other values (74)	2440485	19.5%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	12542423	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
	6866584	54.7%
E	796771	6.4%
T	436287	3.5%
A	408734	3.3%
R	339643	2.7%
N	298626	2.4%
S	285926	2.3%
1	276924	2.2%
U	203017	1.6%
V	189426	1.5%
Other values (74)	2440485	19.5%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	12542423	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
	6866584	54.7%
E	796771	6.4%
T	436287	3.5%
A	408734	3.3%
R	339643	2.7%
N	298626	2.4%
S	285926	2.3%
1	276924	2.2%
U	203017	1.6%
V	189426	1.5%
Other values (74)	2440485	19.5%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	12542423	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
	6866584	54.7%
E	796771	6.4%
T	436287	3.5%
A	408734	3.3%
R	339643	2.7%
N	298626	2.4%
S	285926	2.3%
1	276924	2.2%
U	203017	1.6%
V	189426	1.5%
Other values (74)	2440485	19.5%

NUMBER OF PERSONS INJURED
Real number (ℝ)

ZEROS

Distinct	32	**Minimum**	0	
Distinct (%)	< 0.1%	**Maximum**	43	
Missing	18	**Zeros**	1601221	
Missing (%)	< 0.1%	**Zeros (%)**	77.2%	
Infinite	0	**Negative**	0	
Infinite (%)	0.0%	**Negative (%)**	0.0%	
Mean	0.30980159	**Memory size**	15.8 MiB	
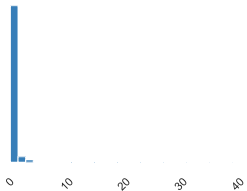
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	2
Maximum	43
Range	43
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	0.69996885
Coefficient of variation (CV)	2.2594102
Kurtosis	51.296075
Mean	0.30980159
Median Absolute Deviation (MAD)	0
Skewness	4.2602307
Sum	642965
Variance	0.4899564
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=32)
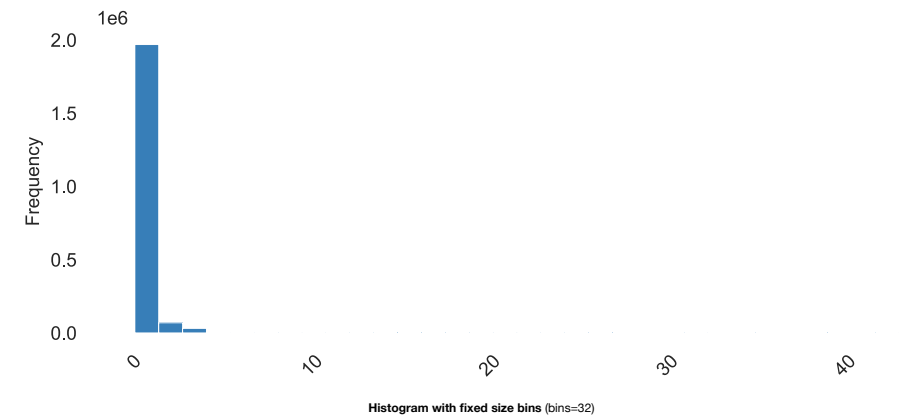
Value	Count	Frequency (%)
0	1601221	77.2%
1	368039	17.7%
2	69310	3.3%
3	22649	1.1%
4	8403	0.4%

Value	Count	Frequency (%)
5	3225	0.2%
6	1350	0.1%
7	574	< 0.1%
8	252	< 0.1%
9	129	< 0.1%
Other values (22)	257	< 0.1%

Value	Count	Frequency (%)
0	1601221	77.2%
1	368039	17.7%
2	69310	3.3%
3	22649	1.1%
4	8403	0.4%
5	3225	0.2%
6	1350	0.1%
7	574	< 0.1%
8	252	< 0.1%
9	129	< 0.1%

Value	Count	Frequency (%)
43	1	< 0.1%
40	1	< 0.1%
34	1	< 0.1%
32	1	< 0.1%
31	1	< 0.1%
27	1	< 0.1%
25	1	< 0.1%
24	3	< 0.1%
23	1	< 0.1%
22	3	< 0.1%

NUMBER OF PERSONS KILLED
Real number (ℝ)

SKEWED ZEROS

Distinct	7	**Minimum**	0
Distinct (%)	< 0.1%	**Maximum**	8
Missing	31	**Zeros**	2072415
Missing (%)	< 0.1%	**Zeros (%)**	99.9%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	0.0014951363	**Memory size**	15.8 MiB
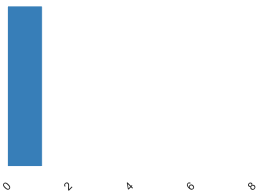
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	0
Maximum	8
Range	8
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	0.040773863
Coefficient of variation (CV)	27.271
Kurtosis	1937.399
Mean	0.0014951363
Median Absolute Deviation (MAD)	0
Skewness	33.717434
Sum	3103
Variance	0.0016625079
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=7)
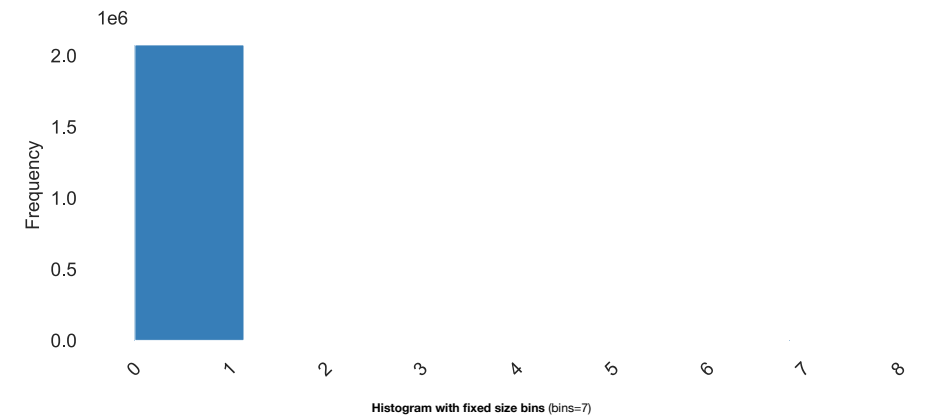
Value	Count	Frequency (%)
0	2072415	99.9%
1	2889	0.1%
2	74	< 0.1%
3	12	< 0.1%
4	3	< 0.1%

Value	Count	Frequency (%)
5	2	< 0.1%
8	1	< 0.1%
(Missing)	31	< 0.1%

Value	Count	Frequency (%)
0	2072415	99.9%
1	2889	0.1%
2	74	< 0.1%
3	12	< 0.1%
4	3	< 0.1%
5	2	< 0.1%
8	1	< 0.1%

Value	Count	Frequency (%)
8	1	< 0.1%
5	2	< 0.1%
4	3	< 0.1%
3	12	< 0.1%
2	74	< 0.1%
1	2889	0.1%
0	2072415	99.9%

NUMBER OF PEDESTRIANS INJURED
Real number (ℝ)

ZEROS

Distinct	14	**Minimum**	0	
Distinct (%)	< 0.1%	**Maximum**	27	
Missing	0	**Zeros**	1962919	
Missing (%)	0.0%	**Zeros (%)**	94.6%	
Infinite	0	**Negative**	0	
Infinite (%)	0.0%	**Negative (%)**	0.0%	
Mean	0.056549327	**Memory size**	15.8 MiB	
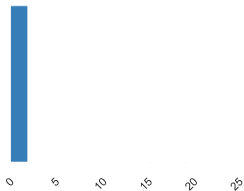
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	1
Maximum	27
Range	27
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	0.2440835
Coefficient of variation (CV)	4.3162936
Kurtosis	129.0936
Mean	0.056549327
Median Absolute Deviation (MAD)	0
Skewness	5.6862516
Sum	117364
Variance	0.059576754
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=14)
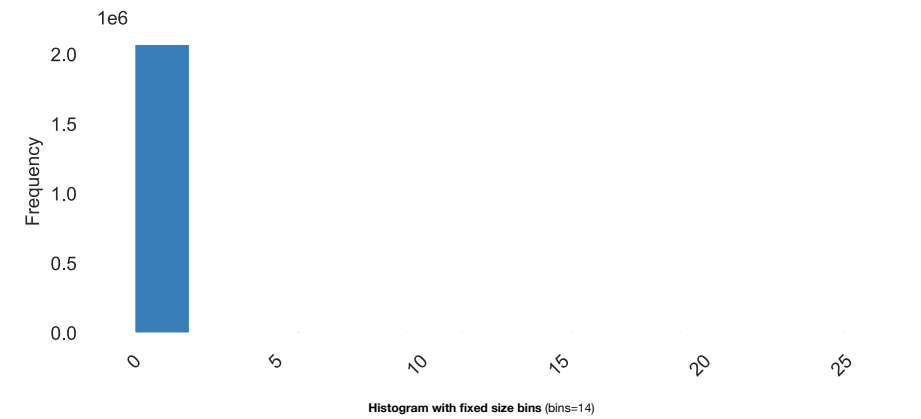
Value	Count	Frequency (%)
0	1962919	94.6%
1	108371	5.2%
2	3663	0.2%
3	365	< 0.1%
4	60	< 0.1%

Value	Count	Frequency (%)
5	26	< 0.1%
6	11	< 0.1%
7	4	< 0.1%
9	2	< 0.1%
8	2	< 0.1%
Other values (4)	4	< 0.1%

Value	Count	Frequency (%)
0	1962919	94.6%
1	108371	5.2%
2	3663	0.2%
3	365	< 0.1%
4	60	< 0.1%
5	26	< 0.1%
6	11	< 0.1%
7	4	< 0.1%
8	2	< 0.1%
9	2	< 0.1%

Value	Count	Frequency (%)
27	1	< 0.1%
19	1	< 0.1%
15	1	< 0.1%
13	1	< 0.1%
9	2	< 0.1%
8	2	< 0.1%
7	4	< 0.1%
6	11	< 0.1%
5	26	< 0.1%
4	60	< 0.1%

NUMBER OF PEDESTRIANS KILLED
Categorical

IMBALANCE

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	15.8 MiB

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	2075427		
Distinct characters	4		
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	1	?
Unique (%)	< 0.1%	?

Sample

1st row	0
2nd row	0
3rd row	0
4th row	0
5th row	0

Common Values

Value	Count	Frequency (%)
0	2073905	99.9%
1	1509	0.1%
2	12	< 0.1%
6	1	< 0.1%

Length

Histogram of lengths of the category

Common Values (Plot)

Value	Count	Frequency (%)
0	2073905	99.9%
1	1509	0.1%
2	12	< 0.1%
6	1	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
0	2073905	99.9%
1	1509	0.1%
2	12	< 0.1%
6	1	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	2075427	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
0	2073905	99.9%
1	1509	0.1%
2	12	< 0.1%
6	1	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	2075427	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
0	2073905	99.9%
1	1509	0.1%
2	12	< 0.1%
6	1	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	2075427	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
0	2073905	99.9%
1	1509	0.1%
2	12	< 0.1%
6	1	< 0.1%

NUMBER OF CYCLIST INJURED
Categorical

`IMBALANCE`

Distinct	5
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	15.8 MiB

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	2075427	
Distinct characters	5	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.
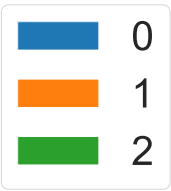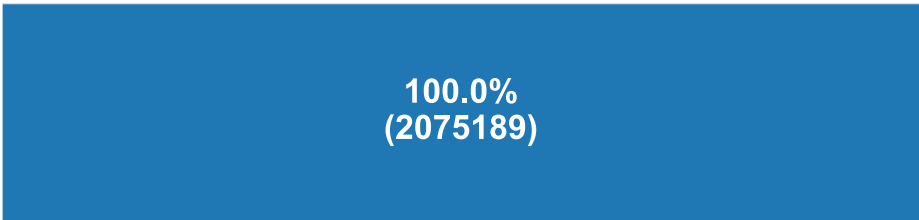
Unique

Unique	1	?
Unique (%)	< 0.1%	

Sample

1st row	0
2nd row	0
3rd row	0
4th row	0
5th row	0

Common Values

Value	Count	Frequency (%)
0	2020463	97.4%
1	54340	2.6%
2	600	< 0.1%
3	23	< 0.1%
4	1	< 0.1%

Length

Histogram of lengths of the category

Common Values (Plot)

97.4%
(2020463)

Value	Count	Frequency (%)
0	2020463	97.4%
1	54340	2.6%
2	600	< 0.1%
3	23	< 0.1%
4	1	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
0	2020463	97.4%
1	54340	2.6%
2	600	< 0.1%
3	23	< 0.1%
4	1	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	2075427	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
0	2020463	97.4%
1	54340	2.6%
2	600	< 0.1%
3	23	< 0.1%
4	1	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	2075427	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
0	2020463	97.4%
1	54340	2.6%
2	600	< 0.1%
3	23	< 0.1%
4	1	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	2075427	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
0	2020463	97.4%
1	54340	2.6%
2	600	< 0.1%
3	23	< 0.1%
4	1	< 0.1%

NUMBER OF CYCLIST KILLED
Categorical

IMBALANCE

Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	15.8 MiB

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	2075427	
Distinct characters	3	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	1	?
Unique (%)	< 0.1%	

Sample

1st row	0
2nd row	0
3rd row	0
4th row	0
5th row	0

Common Values

Value	Count	Frequency (%)
0	2075189	> 99.9%
1	237	< 0.1%
2	1	< 0.1%

Length

Histogram of lengths of the category

Common Values (Plot)

Value	Count	Frequency (%)
0	2075189	> 99.9%
1	237	< 0.1%
2	1	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
0	2075189	> 99.9%
1	237	< 0.1%
2	1	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	2075427	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
0	2075189	> 99.9%
1	237	< 0.1%
2	1	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	2075427	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
0	2075189	> 99.9%
1	237	< 0.1%
2	1	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	2075427	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
0	2075189	> 99.9%
1	237	< 0.1%
2	1	< 0.1%

NUMBER OF MOTORIST INJURED
Real number (ℝ)

ZEROS

Distinct	31	**Minimum**	0	
Distinct (%)	< 0.1%	**Maximum**	43	
Missing	0	**Zeros**	1772939	
Missing (%)	0.0%	**Zeros (%)**	85.4%	
Infinite	0	**Negative**	0	
Infinite (%)	0.0%	**Negative (%)**	0.0%	
Mean	0.22282162	**Memory size**	15.8 MiB	
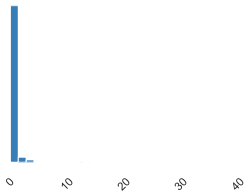
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	1
Maximum	43
Range	43
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	0.66109218
Coefficient of variation (CV)	2.9669122
Kurtosis	63.717057
Mean	0.22282162
Median Absolute Deviation (MAD)	0
Skewness	5.1266596
Sum	462450
Variance	0.43704287
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=31)
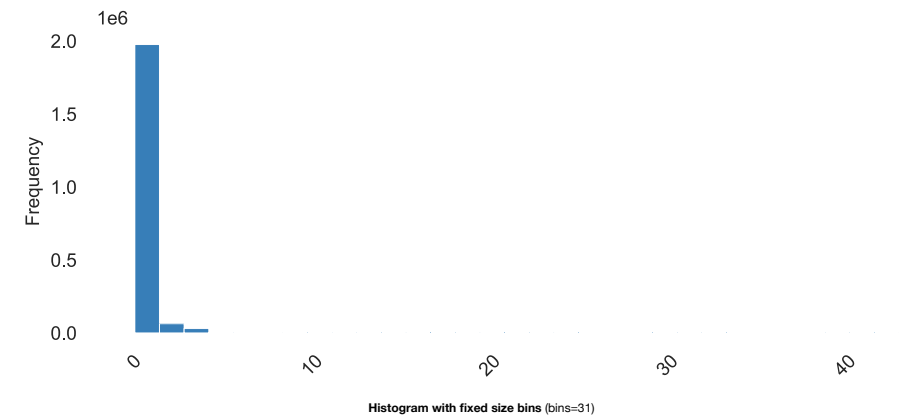
Value	Count	Frequency (%)
0	1772939	85.4%
1	203426	9.8%
2	63230	3.0%
3	21961	1.1%
4	8230	0.4%

Value	Count	Frequency (%)
5	3175	0.2%
6	1304	0.1%
7	548	< 0.1%
8	245	< 0.1%
9	123	< 0.1%
Other values (21)	246	< 0.1%

Value	Count	Frequency (%)
0	1772939	85.4%
1	203426	9.8%
2	63230	3.0%
3	21961	1.1%
4	8230	0.4%
5	3175	0.2%
6	1304	0.1%
7	548	< 0.1%
8	245	< 0.1%
9	123	< 0.1%

Value	Count	Frequency (%)
43	1	< 0.1%
40	1	< 0.1%
34	1	< 0.1%
31	1	< 0.1%
30	1	< 0.1%
25	1	< 0.1%
24	3	< 0.1%
23	1	< 0.1%
22	2	< 0.1%
21	1	< 0.1%

NUMBER OF MOTORIST KILLED
Real number (ℝ)

SKEWED   ZEROS

Distinct	6	**Minimum**	0
Distinct (%)	< 0.1%	**Maximum**	5
Missing	0	**Zeros**	2074246
Missing (%)	0.0%	**Zeros (%)**	99.9%
Infinite	0	**Negative**	0
Infinite (%)	0.0%	**Negative (%)**	0.0%
Mean	0.00061529507	**Memory size**	15.8 MiB
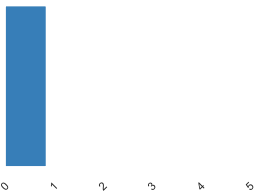
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	0
Maximum	5
Range	5
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	0.027135542
Coefficient of variation (CV)	44.101673
Kurtosis	4230.0939
Mean	0.00061529507
Median Absolute Deviation (MAD)	0
Skewness	54.744147
Sum	1277
Variance	0.00073633763
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=6)
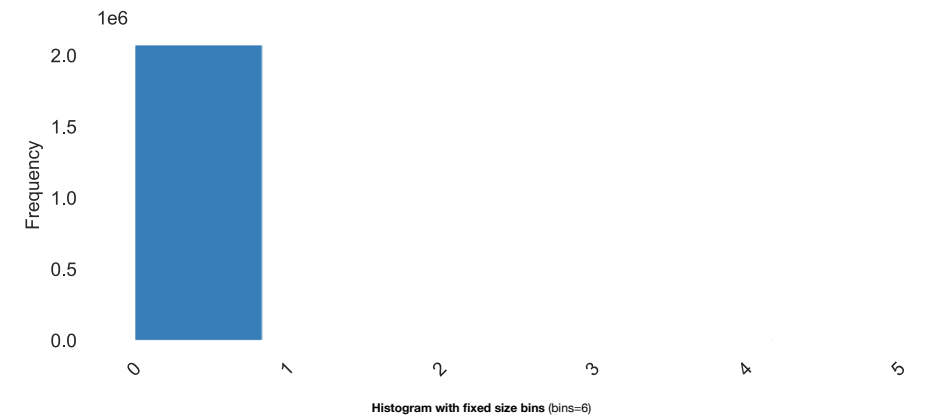
Value	Count	Frequency (%)
0	2074246	99.9%
1	1107	0.1%
2	58	< 0.1%
3	12	< 0.1%
4	2	< 0.1%

Value	Count	Frequency (%)
5	2	< 0.1%

Value	Count	Frequency (%)
0	2074246	99.9%
1	1107	0.1%
2	58	< 0.1%
3	12	< 0.1%
4	2	< 0.1%
5	2	< 0.1%

Value	Count	Frequency (%)
5	2	< 0.1%
4	2	< 0.1%
3	12	< 0.1%
2	58	< 0.1%
1	1107	0.1%
0	2074246	99.9%

CONTRIBUTING FACTOR VEHICLE 1
Text

Distinct	61
Distinct (%)	< 0.1%
Missing	6802
Missing (%)	0.3%
Memory size	15.8 MiB

Length

Max length	53
Median length	43
Mean length	19.504495
Min length	1

Characters and Unicode

Total characters	40347485
Distinct characters	55
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0 ?
Unique (%)	0.0%

Sample

1st row	Aggressive Driving/Road Rage
2nd row	Pavement Slippery
3rd row	Following Too Closely
4th row	Unspecified
5th row	Unspecified

Value	Count	Frequency (%)
unspecified	706732	17.1%
driver	447768	10.9%
inattention/distraction	415252	10.1%
too	162593	3.9%
closely	162593	3.9%
to	148089	3.6%
failure	129495	3.1%
yield	123304	3.0%
right-of-way	123304	3.0%
following	110930	2.7%
Other values (96)	1591210	38.6%

Most occurring characters

Value	Count	Frequency (%)
i	4541258	11.3%
e	4110099	10.2%
n	3507152	8.7%
t	2798284	6.9%
o	2379399	5.9%
r	2368411	5.9%
s	2097469	5.2%
	2052645	5.1%
a	1989702	4.9%
c	1555120	3.9%
Other values (45)	12947946	32.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	40347485	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
i	4541258	11.3%
e	4110099	10.2%
n	3507152	8.7%
t	2798284	6.9%
o	2379399	5.9%
r	2368411	5.9%
s	2097469	5.2%
	2052645	5.1%
a	1989702	4.9%
c	1555120	3.9%
Other values (45)	12947946	32.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	40347485	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
i	4541258	11.3%
e	4110099	10.2%
n	3507152	8.7%
t	2798284	6.9%
o	2379399	5.9%
r	2368411	5.9%
s	2097469	5.2%
	2052645	5.1%
a	1989702	4.9%
c	1555120	3.9%
Other values (45)	12947946	32.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	40347485	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
i	4541258	11.3%
e	4110099	10.2%
n	3507152	8.7%
t	2798284	6.9%
o	2379399	5.9%
r	2368411	5.9%
s	2097469	5.2%
	2052645	5.1%
a	1989702	4.9%
c	1555120	3.9%
Other values (45)	12947946	32.1%

CONTRIBUTING FACTOR VEHICLE 2
Text

`MISSING`

Distinct	61
Distinct (%)	< 0.1%
Missing	321736
Missing (%)	15.5%
Memory size	15.8 MiB

Length

Max length	53
Median length	11
Mean length	13.048611
Min length	1
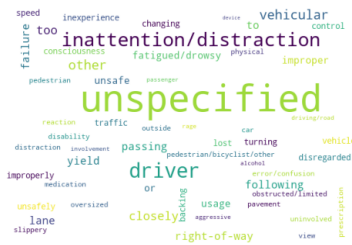
Characters and Unicode

Total characters	22883231	
Distinct characters	55	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	0	?
Unique (%)	0.0%	

Sample

1st row	Unspecified
2nd row	Unspecified
3rd row	Unspecified
4th row	Unspecified
5th row	Unspecified

Value	Count	Frequency (%)
unspecified	1476469	68.6%
driver	100961	4.7%
inattention/distraction	94252	4.4%
other	33129	1.5%
vehicular	32066	1.5%
too	27733	1.3%
closely	27733	1.3%
passing	21554	1.0%
to	21532	1.0%
lane	20107	0.9%
Other values (96)	295716	13.7%

Most occurring characters

Value	Count	Frequency (%)
i	3607436	15.8%
e	3511207	15.3%
n	2050908	9.0%
s	1757954	7.7%
c	1666318	7.3%
d	1549984	6.8%
p	1546225	6.8%
f	1532577	6.7%
U	1512982	6.6%
t	619191	2.7%
Other values (45)	3528449	15.4%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	22883231	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
i	3607436	15.8%
e	3511207	15.3%
n	2050908	9.0%
s	1757954	7.7%
c	1666318	7.3%
d	1549984	6.8%
p	1546225	6.8%
f	1532577	6.7%
U	1512982	6.6%
t	619191	2.7%
Other values (45)	3528449	15.4%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	22883231	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
i	3607436	15.8%
e	3511207	15.3%
n	2050908	9.0%
s	1757954	7.7%
c	1666318	7.3%
d	1549984	6.8%
p	1546225	6.8%
f	1532577	6.7%
U	1512982	6.6%
t	619191	2.7%
Other values (45)	3528449	15.4%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	22883231	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
i	3607436	15.8%
e	3511207	15.3%
n	2050908	9.0%
s	1757954	7.7%
c	1666318	7.3%
d	1549984	6.8%
p	1546225	6.8%
f	1532577	6.7%
U	1512982	6.6%
t	619191	2.7%
Other values (45)	3528449	15.4%

CONTRIBUTING FACTOR VEHICLE 3
Text

MISSING

Distinct	51
Distinct (%)	< 0.1%
Missing	1927163
Missing (%)	92.9%
Memory size	15.8 MiB

Length

Max length	53
Median length	11
Mean length	11.656053
Min length	1
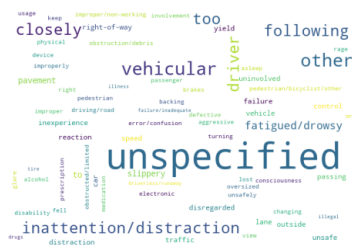
Characters and Unicode

Total characters	1728173
Distinct characters	55
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category) ?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode) ?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block) ?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

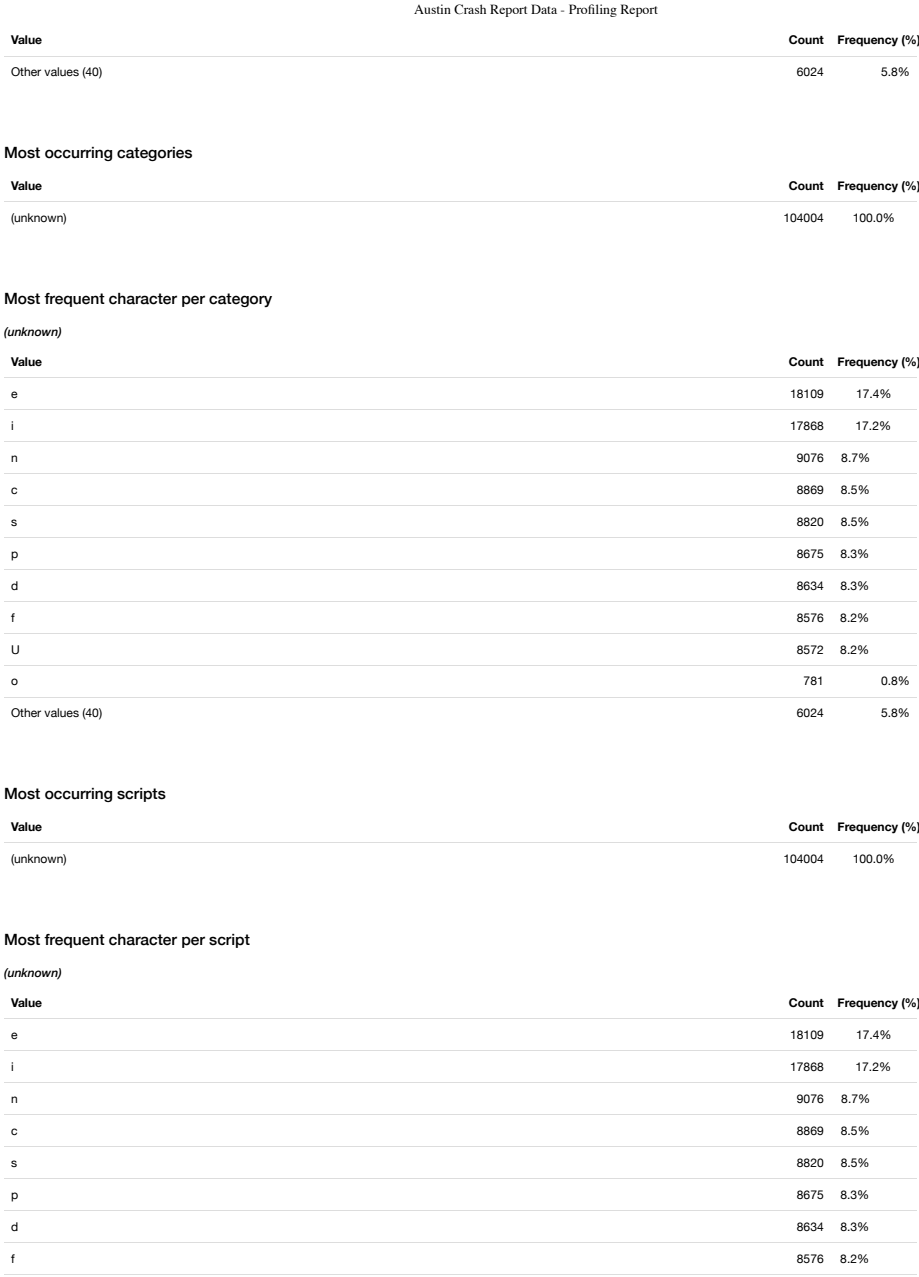|		
Unique	5	?
Unique (%)	< 0.1%	

Sample

1st row	Unspecified
2nd row	Unspecified
3rd row	Unspecified
4th row	Unspecified
5th row	Unspecified

Value	Count	Frequency (%)
unspecified	138219	85.8%
other	2813	1.7%
vehicular	2773	1.7%
driver	2131	1.3%
too	2011	1.2%
closely	2011	1.2%
following	1957	1.2%
inattention/distraction	1950	1.2%
fatigued/drowsy	853	0.5%
pavement	410	0.3%
Other values (79)	5908	3.7%

Most occurring characters

Value	Count	Frequency (%)
e	295337	17.1%
i	294017	17.0%
n	151554	8.8%
s	145163	8.4%
c	144599	8.4%
d	140321	8.1%
p	139879	8.1%
f	139124	8.1%
U	138882	8.0%
o	17264	1.0%
Other values (45)	122033	7.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	1728173	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
e	295337	17.1%
i	294017	17.0%
n	151554	8.8%
s	145163	8.4%
c	144599	8.4%
d	140321	8.1%
p	139879	8.1%
f	139124	8.1%
U	138882	8.0%
o	17264	1.0%
Other values (45)	122033	7.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	1728173	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
e	295337	17.1%
i	294017	17.0%
n	151554	8.8%
s	145163	8.4%
c	144599	8.4%
d	140321	8.1%
p	139879	8.1%
f	139124	8.1%
U	138882	8.0%
o	17264	1.0%
Other values (45)	122033	7.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	1728173	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
e	295337	17.1%
i	294017	17.0%
n	151554	8.8%
s	145163	8.4%
c	144599	8.4%
d	140321	8.1%
p	139879	8.1%
f	139124	8.1%
U	138882	8.0%
o	17264	1.0%
Other values (45)	122033	7.1%

CONTRIBUTING FACTOR VEHICLE 4
Categorical

`IMBALANCE   MISSING`

Distinct	41
Distinct (%)	0.1%
Missing	2041953
Missing (%)	98.4%
Memory size	15.8 MiB

Length

Max length	43
Median length	11
Mean length	11.489425
Min length	5

Characters and Unicode

Total characters	384597
Distinct characters	51
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	7	?
Unique (%)	< 0.1%	

Sample

1st row	Unspecified
2nd row	Unspecified
3rd row	Unspecified
4th row	Unspecified
5th row	Unspecified

Common Values

Value	Count	Frequency (%)
Unspecified	31577	1.5%
Other Vehicular	614	< 0.1%
Following Too Closely	390	< 0.1%
Driver Inattention/Distraction	275	< 0.1%
Fatigued/Drowsy	170	< 0.1%
Pavement Slippery	116	< 0.1%
Reaction to Uninvolved Vehicle	41	< 0.1%
Unsafe Speed	32	< 0.1%
Outside Car Distraction	28	< 0.1%
Driver Inexperience	27	< 0.1%
Other values (31)	204	< 0.1%
(Missing)	2041953	98.4%

Length

Histogram of lengths of the category

Value	Count	Frequency (%)
unspecified	31577	88.1%
other	623	1.7%
vehicular	614	1.7%
too	395	1.1%
closely	395	1.1%
following	390	1.1%
driver	302	0.8%
inattention/distraction	275	0.8%
fatigued/drowsy	170	0.5%
pavement	119	0.3%
Other values (64)	965	2.7%

Most occurring characters

Value	Count	Frequency (%)
e	66721	17.3%
i	66107	17.2%
n	33651	8.7%
c	32739	8.5%
s	32723	8.5%
p	31939	8.3%
d	31931	8.3%
f	31705	8.2%
U	31684	8.2%
o	3077	0.8%

Value	Count	Frequency (%)
Other values (41)	22320	5.8%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	384597	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
e	66721	17.3%
i	66107	17.2%
n	33651	8.7%
c	32739	8.5%
s	32723	8.5%
p	31939	8.3%
d	31931	8.3%
f	31705	8.2%
U	31684	8.2%
o	3077	0.8%
Other values (41)	22320	5.8%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	384597	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
e	66721	17.3%
i	66107	17.2%
n	33651	8.7%
c	32739	8.5%
s	32723	8.5%
p	31939	8.3%
d	31931	8.3%
f	31705	8.2%

Value	Count	Frequency (%)
U	31684	8.2%
o	3077	0.8%
Other values (41)	22320	5.8%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	384597	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
e	66721	17.3%
i	66107	17.2%
n	33651	8.7%
c	32739	8.5%
s	32723	8.5%
p	31939	8.3%
d	31931	8.3%
f	31705	8.2%
U	31684	8.2%
o	3077	0.8%
Other values (41)	22320	5.8%

CONTRIBUTING FACTOR VEHICLE 5
Categorical

`IMBALANCE  MISSING`

Distinct	30
Distinct (%)	0.3%
Missing	2066358
Missing (%)	99.6%
Memory size	15.8 MiB

Length

Max length	43
Median length	11
Mean length	11.468078
Min length	5

Characters and Unicode

Total characters	104004
Distinct characters	50
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category) ?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode) ?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block) ?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	11	?
Unique (%)	0.1%	

Sample

1st row	Unspecified
2nd row	Unspecified
3rd row	Unspecified
4th row	Unspecified
5th row	Unspecified

Common Values

Value	Count	Frequency (%)
Unspecified	8549	0.4%
Other Vehicular	178	< 0.1%
Following Too Closely	98	< 0.1%
Driver Inattention/Distraction	64	< 0.1%
Pavement Slippery	49	< 0.1%
Fatigued/Drowsy	41	< 0.1%
Reaction to Uninvolved Vehicle	12	< 0.1%
Alcohol Involvement	11	< 0.1%
Obstruction/Debris	10	< 0.1%
Driver Inexperience	10	< 0.1%
Other values (20)	47	< 0.1%
(Missing)	2066358	99.6%

Length

Histogram of lengths of the category

Value	Count	Frequency (%)
unspecified	8549	88.2%
other	180	1.9%
vehicular	178	1.8%
too	100	1.0%
closely	100	1.0%
following	98	1.0%
driver	74	0.8%
inattention/distraction	64	0.7%
pavement	50	0.5%
slippery	49	0.5%
Other values (47)	251	2.6%

Most occurring characters

Value	Count	Frequency (%)
e	18109	17.4%
i	17868	17.2%
n	9076	8.7%
c	8869	8.5%
s	8820	8.5%
p	8675	8.3%
d	8634	8.3%
f	8576	8.2%
U	8572	8.2%
o	781	0.8%

Value	Count	Frequency (%)
Other values (40)	6024	5.8%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	104004	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
e	18109	17.4%
i	17868	17.2%
n	9076	8.7%
c	8869	8.5%
s	8820	8.5%
p	8675	8.3%
d	8634	8.3%
f	8576	8.2%
U	8572	8.2%
o	781	0.8%
Other values (40)	6024	5.8%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	104004	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
e	18109	17.4%
i	17868	17.2%
n	9076	8.7%
c	8869	8.5%
s	8820	8.5%
p	8675	8.3%
d	8634	8.3%
f	8576	8.2%

Value	Count	Frequency (%)
U	8572	8.2%
o	781	0.8%
Other values (40)	6024	5.8%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	104004	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
e	18109	17.4%
i	17868	17.2%
n	9076	8.7%
c	8869	8.5%
s	8820	8.5%
p	8675	8.3%
d	8634	8.3%
f	8576	8.2%
U	8572	8.2%
o	781	0.8%
Other values (40)	6024	5.8%

COLLISION_ID
Real number (ℝ)

UNIQUE

Distinct	2075427	Minimum	22	
Distinct (%)	100.0%	Maximum	4712252	
Missing	0	Zeros	0	
Missing (%)	0.0%	Zeros (%)	0.0%	
Infinite	0	Negative	0	
Infinite (%)	0.0%	Negative (%)	0.0%	
Mean	3159627	Memory size	15.8 MiB	

Quantile statistics

Minimum	22
5-th percentile	104625.3
Q1	3154976.5
median	3673954
Q3	4193057.5
95-th percentile	4608219.7
Maximum	4712252
Range	4712230
Interquartile range (IQR)	1038081

Descriptive statistics

Standard deviation	1505149.9
Coefficient of variation (CV)	0.47636949
Kurtosis	-0.032800807
Mean	3159627
Median Absolute Deviation (MAD)	519041
Skewness	-1.2236319
Sum	6.5575751×10^{12}
Variance	2.2654762×10^{12}
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
4455765	1	< 0.1%
3176288	1	< 0.1%
3188747	1	< 0.1%
3176436	1	< 0.1%
3189909	1	< 0.1%

Value	Count	Frequency (%)
3187402	1	< 0.1%
3178392	1	< 0.1%
3183441	1	< 0.1%
3178566	1	< 0.1%
3185340	1	< 0.1%
Other values (2075417)	2075417	> 99.9%
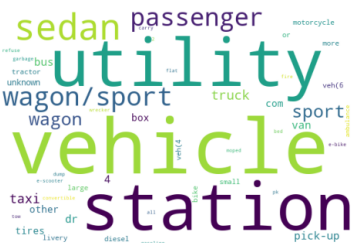
Value	Count	Frequency (%)
22	1	< 0.1%
23	1	< 0.1%
24	1	< 0.1%
25	1	< 0.1%
26	1	< 0.1%
27	1	< 0.1%
28	1	< 0.1%
29	1	< 0.1%
30	1	< 0.1%
31	1	< 0.1%

Value	Count	Frequency (%)
4712252	1	< 0.1%
4712247	1	< 0.1%
4712246	1	< 0.1%
4712245	1	< 0.1%
4712242	1	< 0.1%
4712241	1	< 0.1%
4712237	1	< 0.1%
4712235	1	< 0.1%
4712232	1	< 0.1%
4712231	1	< 0.1%
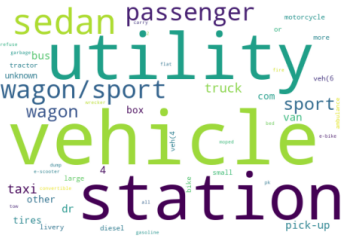
VEHICLE TYPE CODE 1
Text

Distinct	1631
Distinct (%)	0.1%
Missing	13691
Missing (%)	0.7%
Memory size	15.8 MiB

Length

Max length	38
Median length	35
Mean length	16.886453
Min length	1

Characters and Unicode

Total characters	34815408
Distinct characters	75
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category) ?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode) ?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block) ?

Unique

Unique	989	?
Unique (%)	< 0.1%	

Sample

1st row	Sedan
2nd row	Sedan
3rd row	Sedan
4th row	Sedan
5th row	Dump

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Value	Count	Frequency (%)
vehicle	880306	18.0%
utility	633851	13.0%
station	633808	13.0%
sedan	619493	12.7%
wagon/sport	453517	9.3%
passenger	416219	8.5%
	181665	3.7%
wagon	180354	3.7%
sport	180291	3.7%
truck	85920	1.8%
Other values (950)	616060	12.6%

Most occurring characters

Value	Count	Frequency (%)
	2832968	8.1%
S	2735641	7.9%
t	2300987	6.6%
i	1938153	5.6%
E	1818931	5.2%
a	1620452	4.7%
e	1611200	4.6%
n	1548461	4.4%
o	1436044	4.1%
T	1141718	3.3%
Other values (65)	15830853	45.5%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	34815408	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
	2832968	8.1%
S	2735641	7.9%
t	2300987	6.6%
i	1938153	5.6%
E	1818931	5.2%
a	1620452	4.7%
e	1611200	4.6%
n	1548461	4.4%
o	1436044	4.1%
T	1141718	3.3%
Other values (65)	15830853	45.5%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	34815408	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
	2832968	8.1%
S	2735641	7.9%
t	2300987	6.6%
i	1938153	5.6%
E	1818931	5.2%
a	1620452	4.7%
e	1611200	4.6%
n	1548461	4.4%
o	1436044	4.1%
T	1141718	3.3%
Other values (65)	15830853	45.5%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	34815408	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
	2832968	8.1%
S	2735641	7.9%
t	2300987	6.6%
i	1938153	5.6%
E	1818931	5.2%
a	1620452	4.7%
e	1611200	4.6%
n	1548461	4.4%
o	1436044	4.1%
T	1141718	3.3%
Other values (65)	15830853	45.5%

VEHICLE TYPE CODE 2
Text

`MISSING`

Distinct	1819
Distinct (%)	0.1%
Missing	396691
Missing (%)	19.1%
Memory size	15.8 MiB

Length

Max length	38
Median length	30
Mean length	16.08444
Min length	1
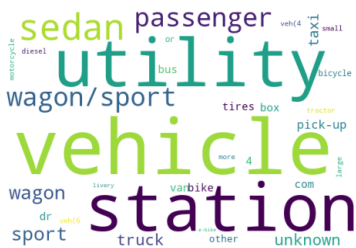
Characters and Unicode

|		
Total characters	27001529	
Distinct characters	73	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	1080	?
Unique (%)	0.1%	

Sample

1st row	Sedan
2nd row	Pick-up Truck
3rd row	Sedan
4th row	Tractor Truck Diesel
5th row	Sedan

Value	Count	Frequency (%)
vehicle	653746	17.1%
utility	466778	12.2%
station	466750	12.2%
sedan	435556	11.4%
wagon/sport	326546	8.5%
passenger	318612	8.3%
	141501	3.7%
wagon	140256	3.7%
sport	140204	3.7%
truck	85272	2.2%
Other values (1009)	655810	17.1%

Most occurring characters

Value	Count	Frequency (%)
	2165263	8.0%
S	2031182	7.5%
t	1665937	6.2%
E	1438671	5.3%
i	1431599	5.3%
e	1189958	4.4%
a	1165845	4.3%
n	1107454	4.1%
o	1060004	3.9%
T	919371	3.4%
Other values (63)	12826245	47.5%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	27001529	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
	2165263	8.0%
S	2031182	7.5%
t	1665937	6.2%
E	1438671	5.3%
i	1431599	5.3%
e	1189958	4.4%
a	1165845	4.3%
n	1107454	4.1%
o	1060004	3.9%
T	919371	3.4%
Other values (63)	12826245	47.5%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	27001529	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
	2165263	8.0%
S	2031182	7.5%
t	1665937	6.2%
E	1438671	5.3%
i	1431599	5.3%
e	1189958	4.4%
a	1165845	4.3%
n	1107454	4.1%
o	1060004	3.9%
T	919371	3.4%
Other values (63)	12826245	47.5%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	27001529	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
	2165263	8.0%
S	2031182	7.5%
t	1665937	6.2%
E	1438671	5.3%
i	1431599	5.3%
e	1189958	4.4%
a	1165845	4.3%
n	1107454	4.1%
o	1060004	3.9%
T	919371	3.4%
Other values (63)	12826245	47.5%

VEHICLE TYPE CODE 3
Text

`MISSING`

Distinct	260
Distinct (%)	0.2%
Missing	1932530
Missing (%)	93.1%
Memory size	15.8 MiB

Length

Max length	35
Median length	30
Mean length	17.679552
Min length	2

Characters and Unicode

Total characters	2526355
Distinct characters	62
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category) ?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode) ?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block) ?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	152 ?
Unique (%)	0.1%

Sample

1st row	Sedan
2nd row	Station Wagon/Sport Utility Vehicle
3rd row	Sedan
4th row	Sedan
5th row	Sedan

Value	Count	Frequency (%)
vehicle	64246	18.5%
utility	49457	14.2%
station	49455	14.2%
sedan	47158	13.6%
wagon/sport	36096	10.4%
passenger	27716	8.0%
	13439	3.9%
wagon	13359	3.8%
sport	13358	3.8%
truck	4339	1.3%
Other values (216)	28474	8.2%

Most occurring characters

Value	Count	Frequency (%)
	204635	8.1%
S	200575	7.9%
t	181889	7.2%
i	150270	5.9%
a	122930	4.9%
e	122469	4.8%
n	120231	4.8%
E	116403	4.6%
o	111274	4.4%
T	77028	3.0%
Other values (52)	1118651	44.3%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	2526355	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
	204635	8.1%
S	200575	7.9%
t	181889	7.2%
i	150270	5.9%
a	122930	4.9%
e	122469	4.8%
n	120231	4.8%
E	116403	4.6%
o	111274	4.4%
T	77028	3.0%
Other values (52)	1118651	44.3%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	2526355	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
	204635	8.1%
S	200575	7.9%
t	181889	7.2%
i	150270	5.9%
a	122930	4.9%
e	122469	4.8%
n	120231	4.8%
E	116403	4.6%
o	111274	4.4%
T	77028	3.0%
Other values (52)	1118651	44.3%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	2526355	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
	204635	8.1%
S	200575	7.9%
t	181889	7.2%
i	150270	5.9%
a	122930	4.9%
e	122469	4.8%
n	120231	4.8%
E	116403	4.6%
o	111274	4.4%
T	77028	3.0%
Other values (52)	1118651	44.3%

VEHICLE TYPE CODE 4
Text

MISSING

Distinct	101
Distinct (%)	0.3%
Missing	2043115
Missing (%)	98.4%
Memory size	15.8 MiB

Length

Max length	35
Median length	30
Mean length	17.97682
Min length	2
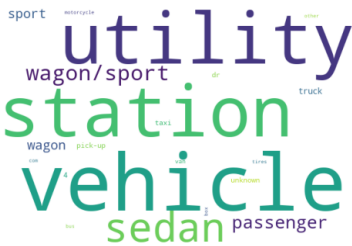
Characters and Unicode

Total characters	580867	
Distinct characters	57	
Distinct categories	1	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category) ?
Distinct scripts	1	(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode) ?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block) ?

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	45	?
Unique (%)	0.1%	

Sample

1st row	Station Wagon/Sport Utility Vehicle
2nd row	Sedan
3rd row	Station Wagon/Sport Utility Vehicle
4th row	Sedan
5th row	Sedan

Value	Count	Frequency (%)
vehicle	14893	18.9%
utility	11719	14.8%
station	11719	14.8%
sedan	11398	14.4%
wagon/sport	8867	11.2%
passenger	5970	7.6%
	2859	3.6%
sport	2852	3.6%
wagon	2852	3.6%
truck	798	1.0%
Other values (103)	5046	6.4%

Most occurring characters

Value	Count	Frequency (%)
	46717	8.0%
S	46409	8.0%
t	44549	7.7%
i	36568	6.3%
a	29793	5.1%
e	29584	5.1%
n	29274	5.0%
o	27071	4.7%
E	24669	4.2%
l	17966	3.1%
Other values (47)	248267	42.7%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	580867	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
	46717	8.0%
S	46409	8.0%
t	44549	7.7%
i	36568	6.3%
a	29793	5.1%
e	29584	5.1%
n	29274	5.0%
o	27071	4.7%
E	24669	4.2%
l	17966	3.1%
Other values (47)	248267	42.7%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	580867	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
	46717	8.0%
S	46409	8.0%
t	44549	7.7%
i	36568	6.3%
a	29793	5.1%
e	29584	5.1%
n	29274	5.0%
o	27071	4.7%
E	24669	4.2%
l	17966	3.1%
Other values (47)	248267	42.7%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	580867	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
	46717	8.0%
S	46409	8.0%
t	44549	7.7%
i	36568	6.3%
a	29793	5.1%
e	29584	5.1%
n	29274	5.0%
o	27071	4.7%
E	24669	4.2%
l	17966	3.1%
Other values (47)	248267	42.7%

VEHICLE TYPE CODE 5
Text

`MISSING`

Distinct	70
Distinct (%)	0.8%
Missing	2066635
Missing (%)	99.6%
Memory size	15.8 MiB

Length

Max length	35
Median length	30
Mean length	18.214058
Min length	2

Characters and Unicode

Total characters	160138
Distinct characters	54
Distinct categories	1 ? (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
Distinct scripts	1 ? (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
Distinct blocks	1 ? (https://en.wikipedia.org/wiki/Unicode_block)

The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Unique

Unique	31 ?
Unique (%)	0.4%

Sample

1st row	Station Wagon/Sport Utility Vehicle
2nd row	Station Wagon/Sport Utility Vehicle
3rd row	Sedan
4th row	Sedan
5th row	Station Wagon/Sport Utility Vehicle

Value	Count	Frequency (%)
vehicle	4020	18.5%
utility	3326	15.3%
station	3326	15.3%
sedan	3182	14.7%
wagon/sport	2524	11.6%
passenger	1487	6.8%
	804	3.7%
wagon	804	3.7%
sport	802	3.7%
truck	245	1.1%
Other values (68)	1196	5.5%

Most occurring characters

Value	Count	Frequency (%)
	12934	8.1%
S	12722	7.9%
t	12689	7.9%
i	10410	6.5%
a	8409	5.3%
e	8354	5.2%
n	8289	5.2%
o	7724	4.8%
E	6129	3.8%
l	5114	3.2%
Other values (44)	67364	42.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	160138	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
	12934	8.1%
S	12722	7.9%
t	12689	7.9%
i	10410	6.5%
a	8409	5.3%
e	8354	5.2%
n	8289	5.2%
o	7724	4.8%
E	6129	3.8%
l	5114	3.2%
Other values (44)	67364	42.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	160138	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
	12934	8.1%
S	12722	7.9%
t	12689	7.9%
i	10410	6.5%
a	8409	5.3%
e	8354	5.2%
n	8289	5.2%
o	7724	4.8%
E	6129	3.8%
l	5114	3.2%
Other values (44)	67364	42.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	160138	100.0%

Most frequent character per block

(unknown)

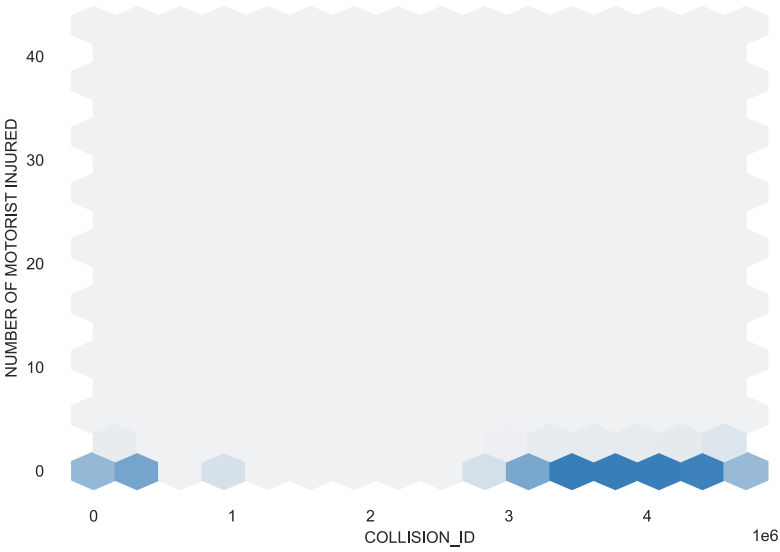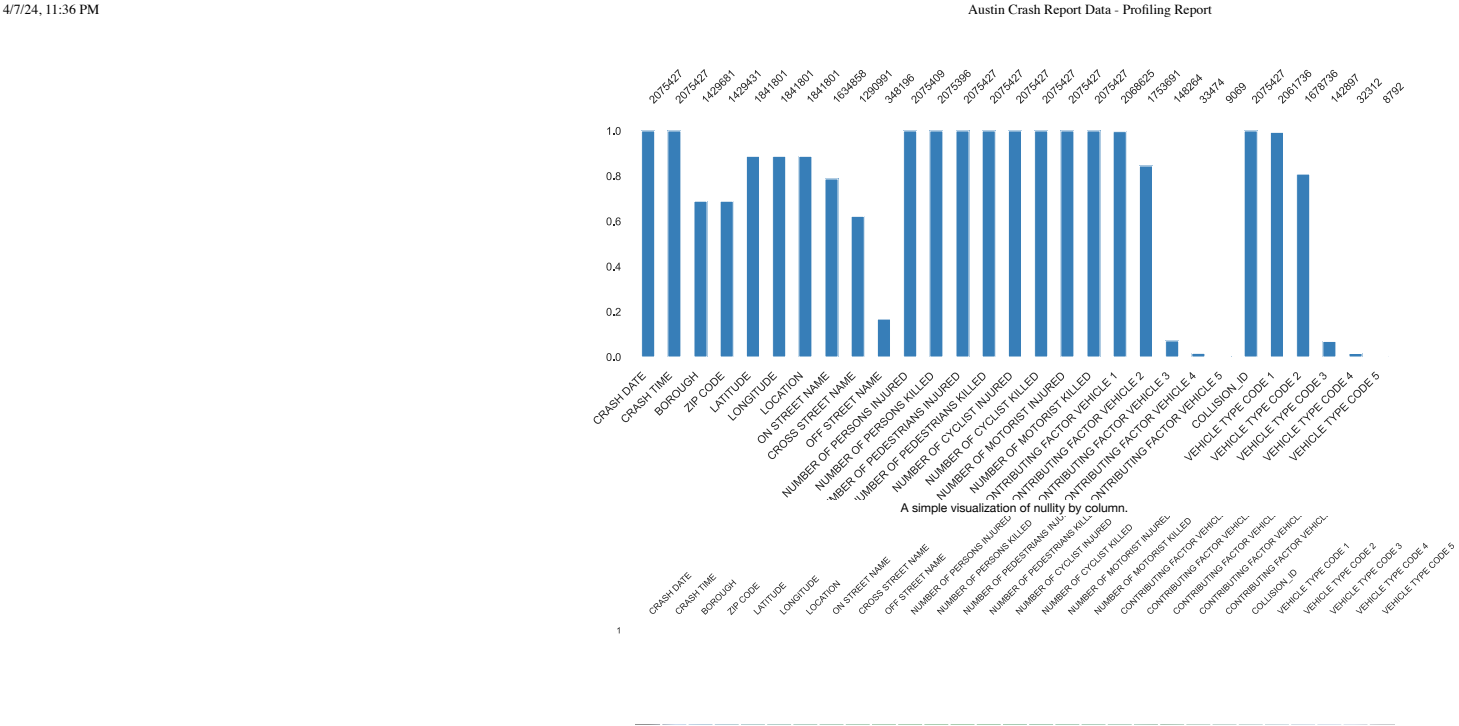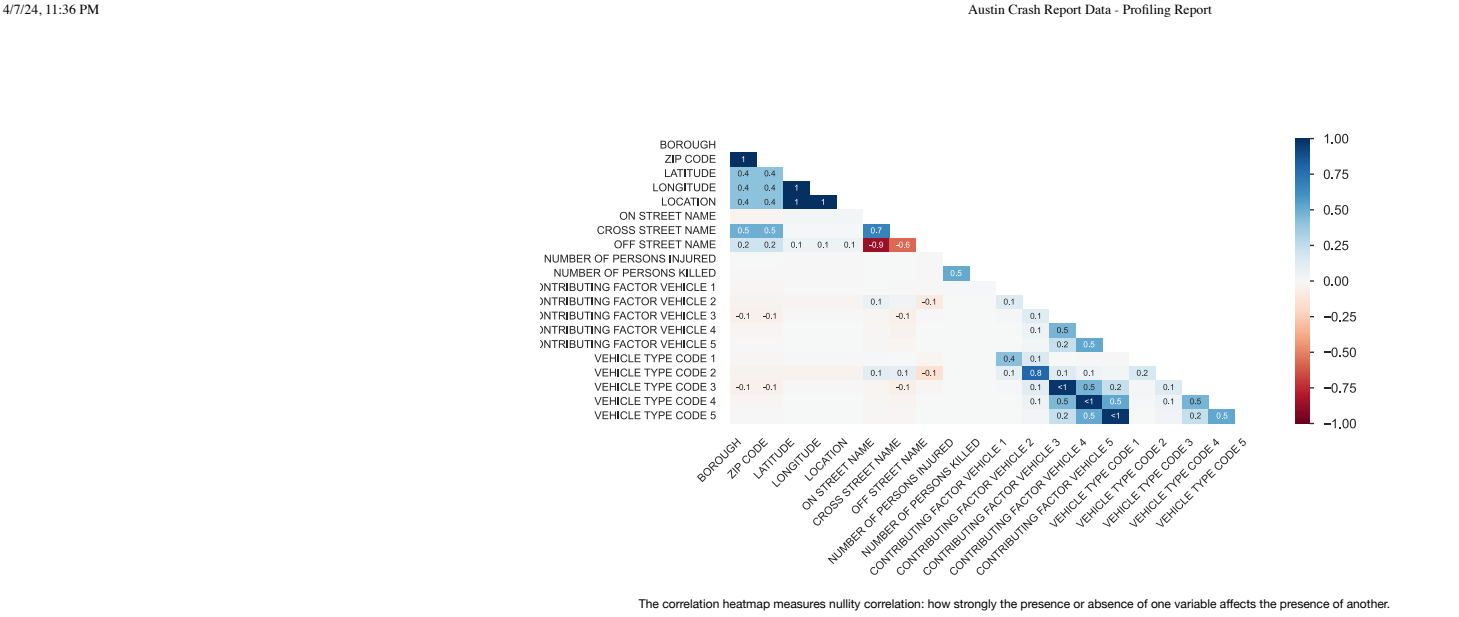| Value	Count	Frequency (%)
	12934	8.1%
S	12722	7.9%
t	12689	7.9%
i	10410	6.5%
a	8409	5.3%
e	8354	5.2%
n	8289	5.2%
o	7724	4.8%
E	6129	3.8%
l	5114	3.2%
Other values (44)	67364	42.1%

Interactions

Missing values

A simple visualization of nullity by column.
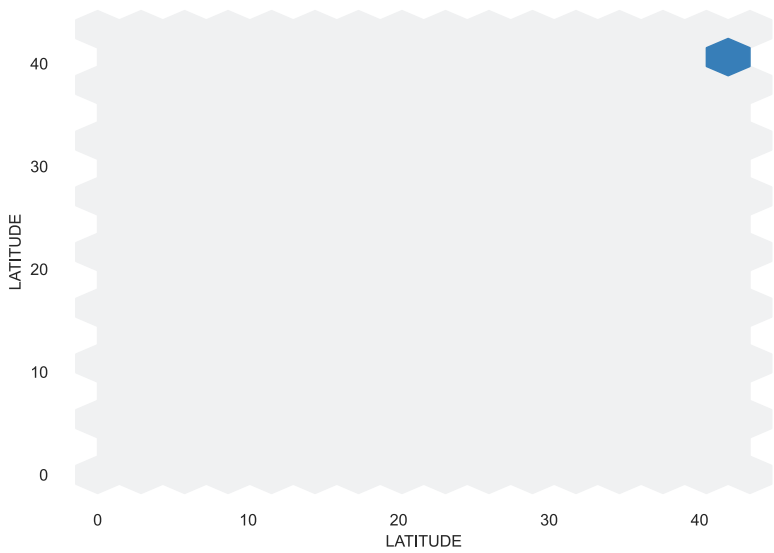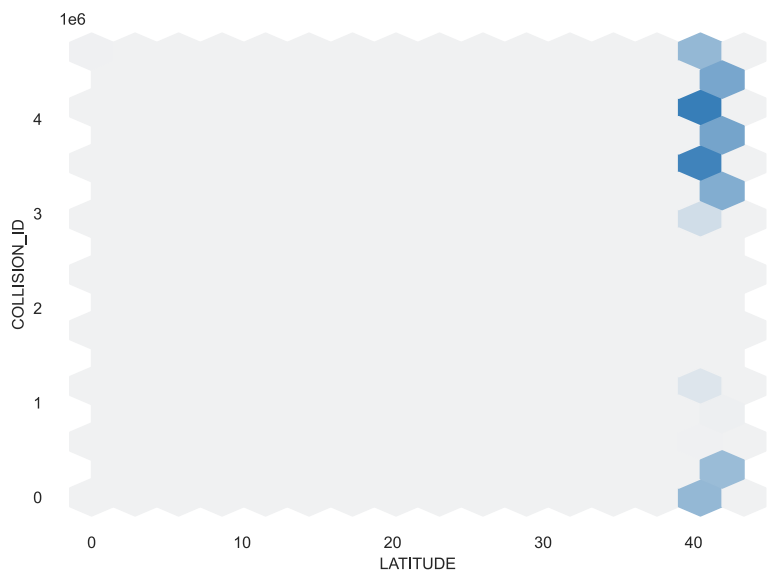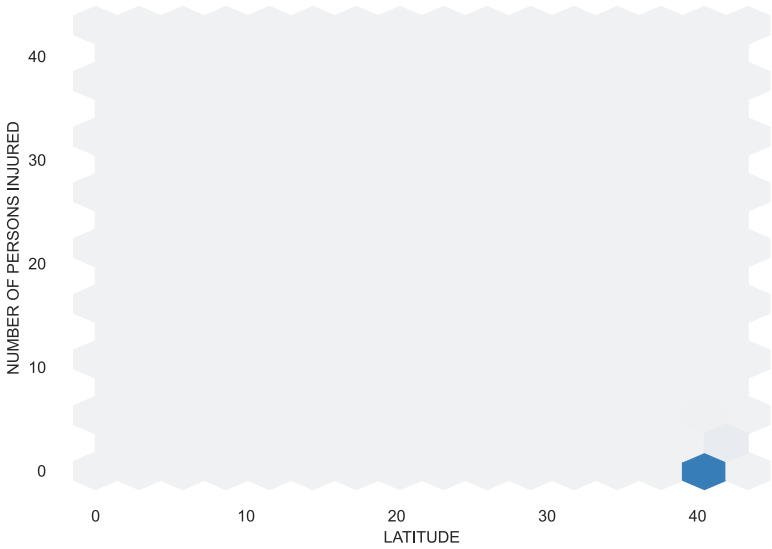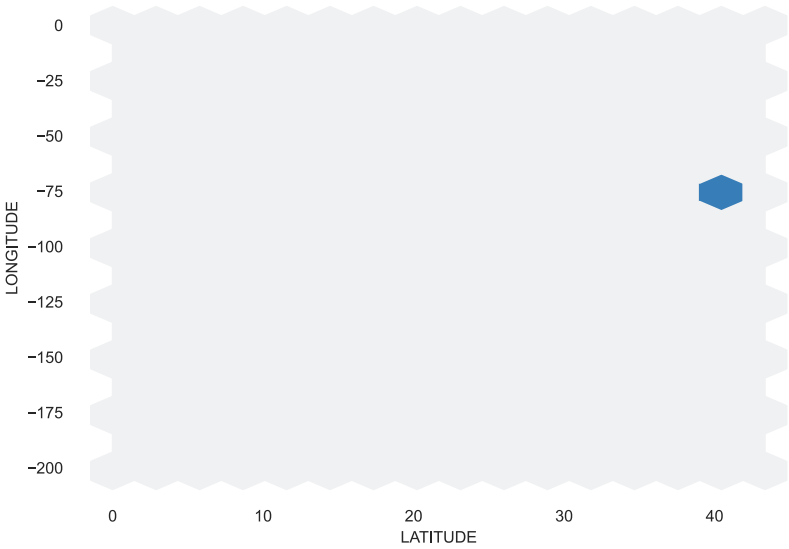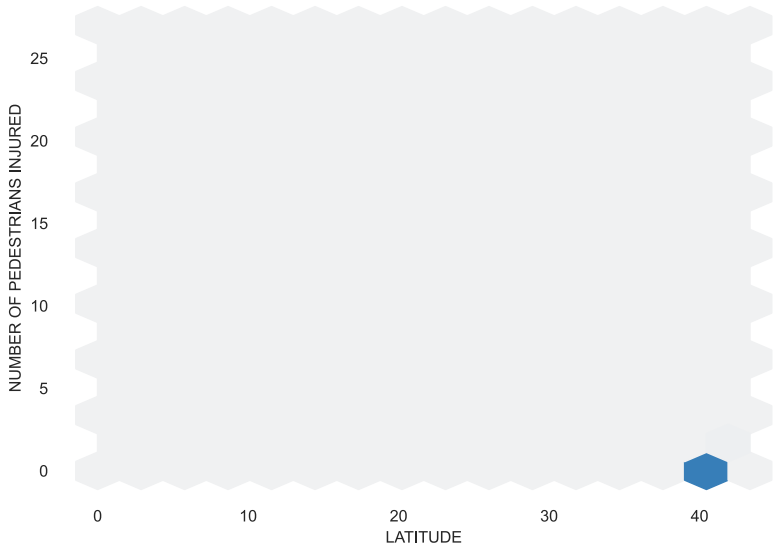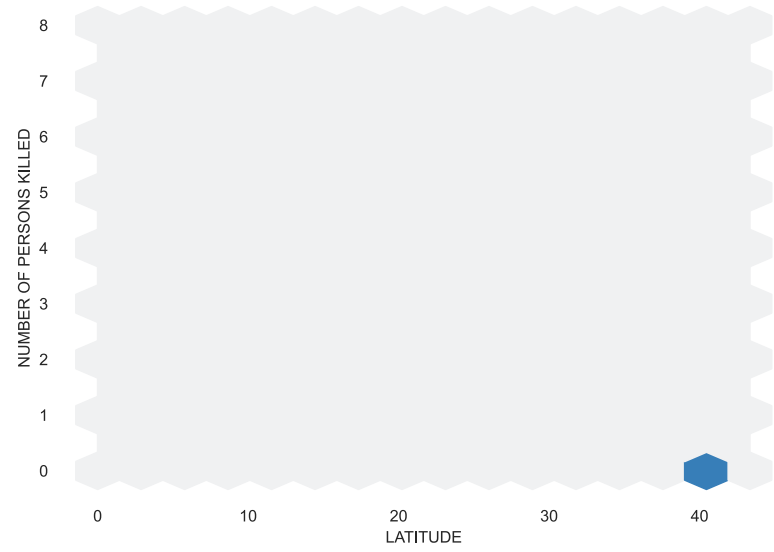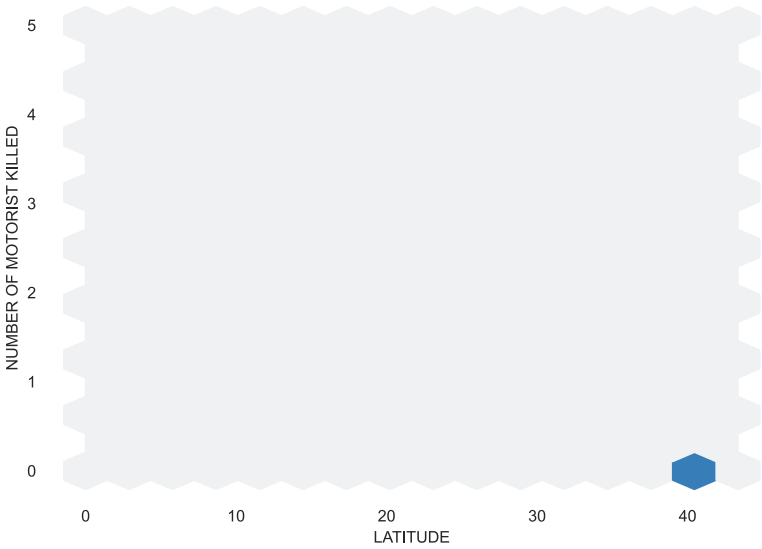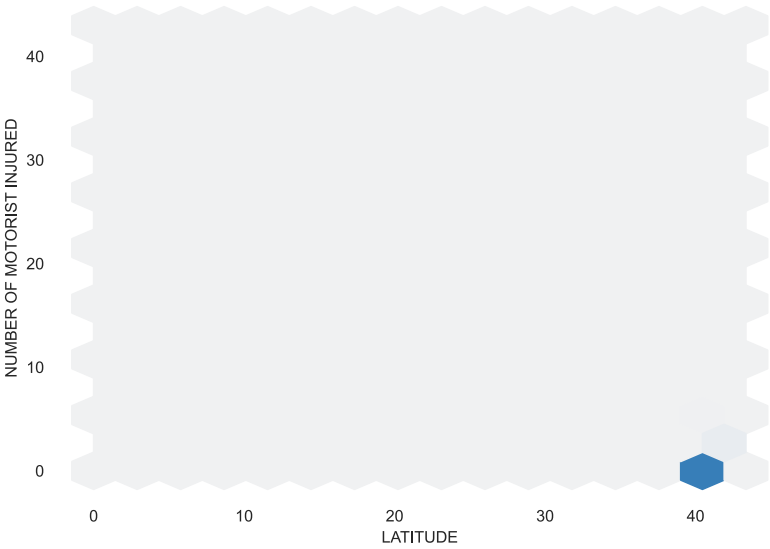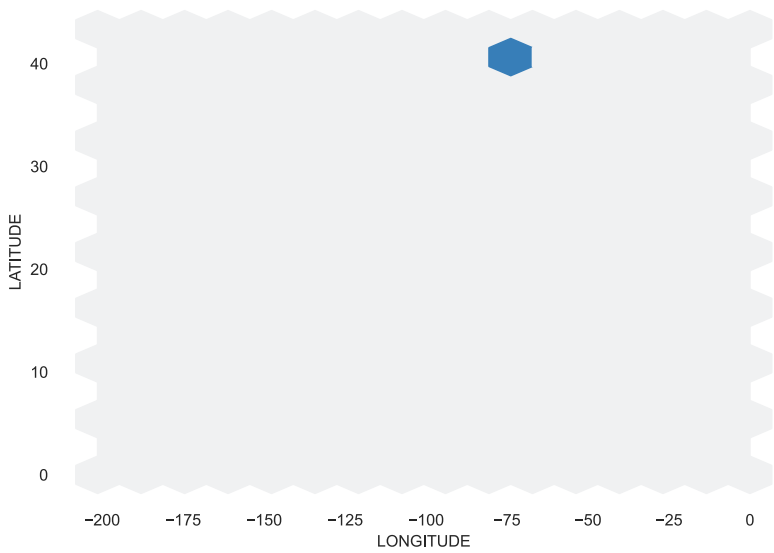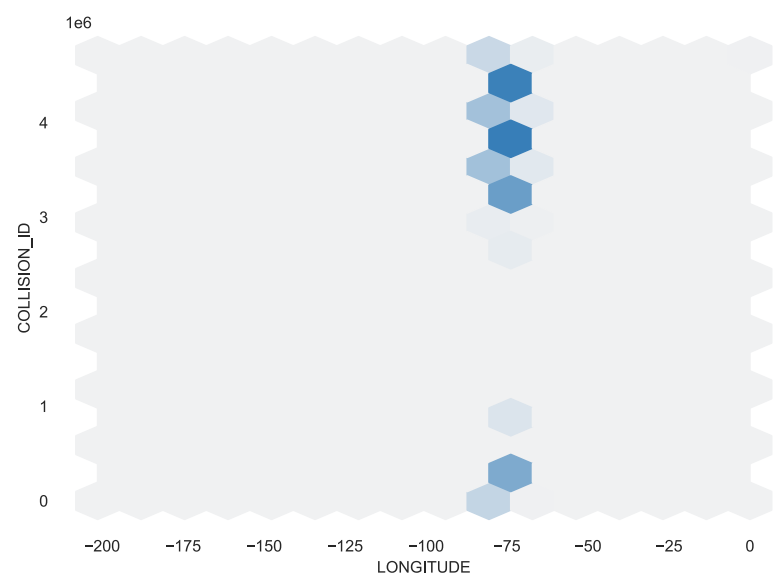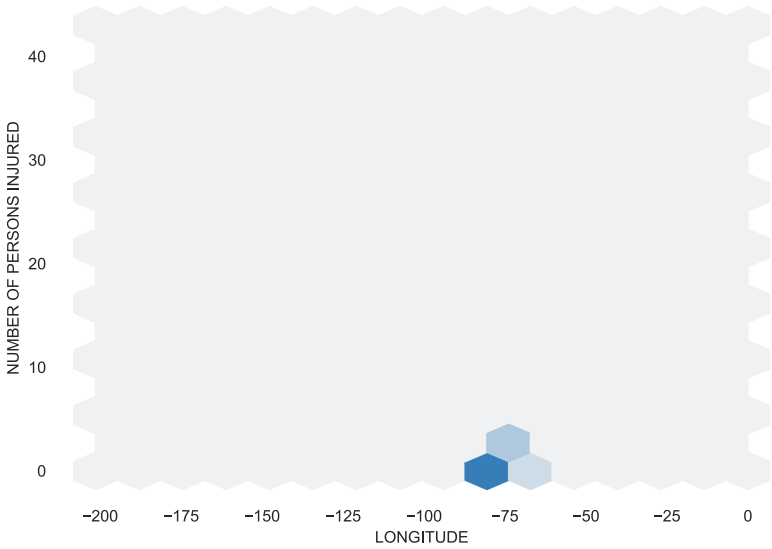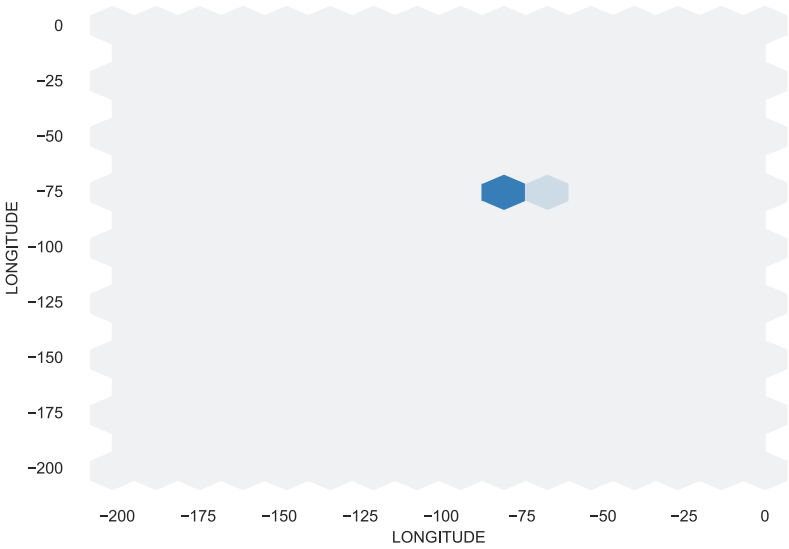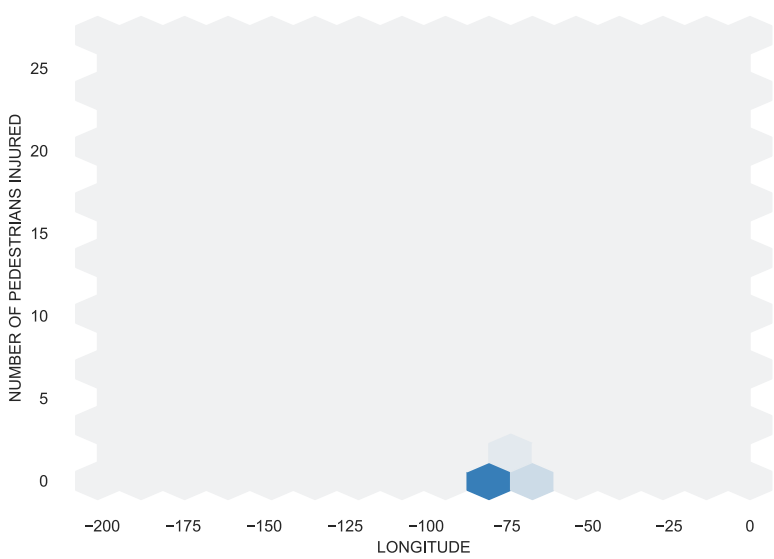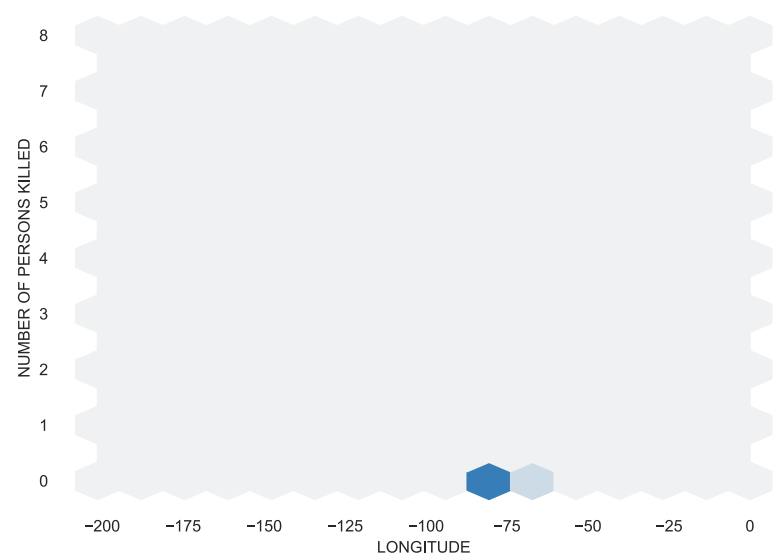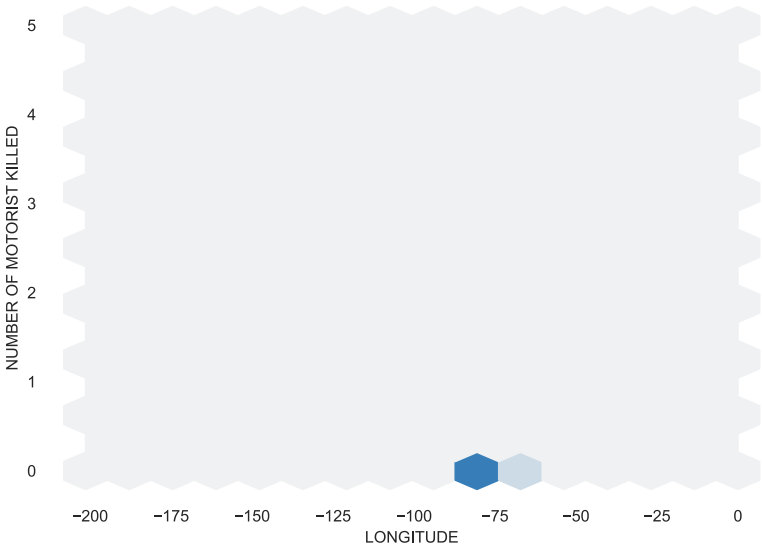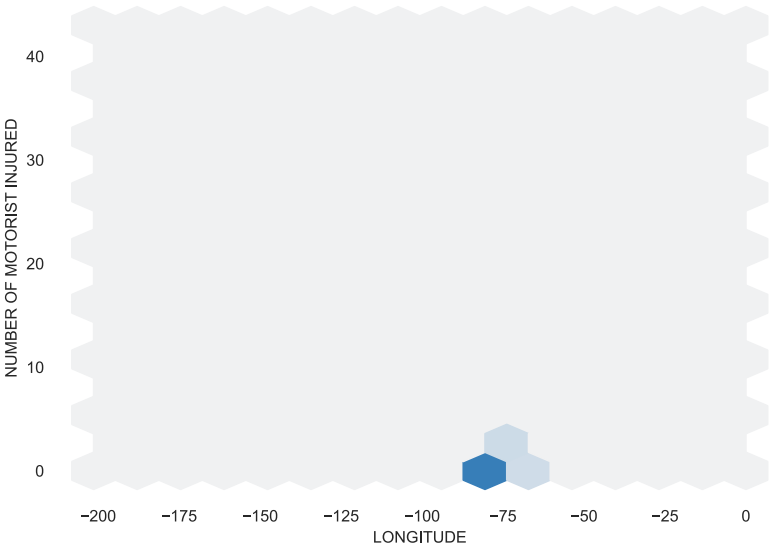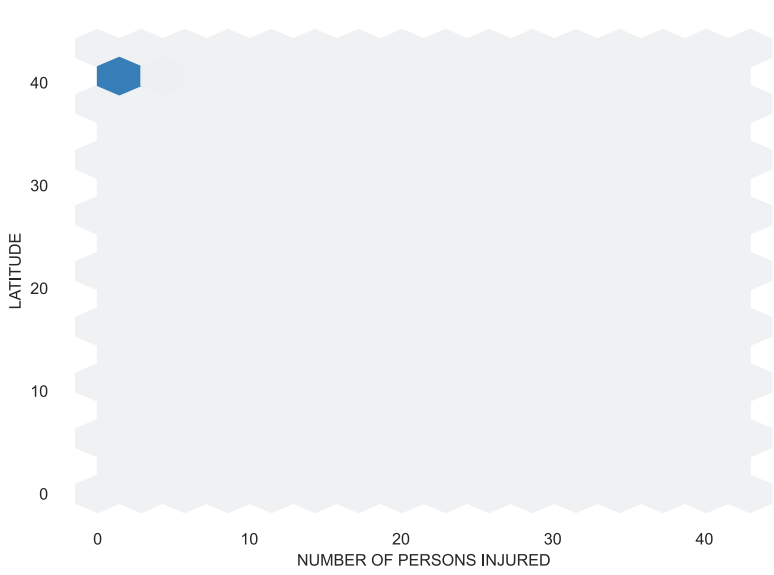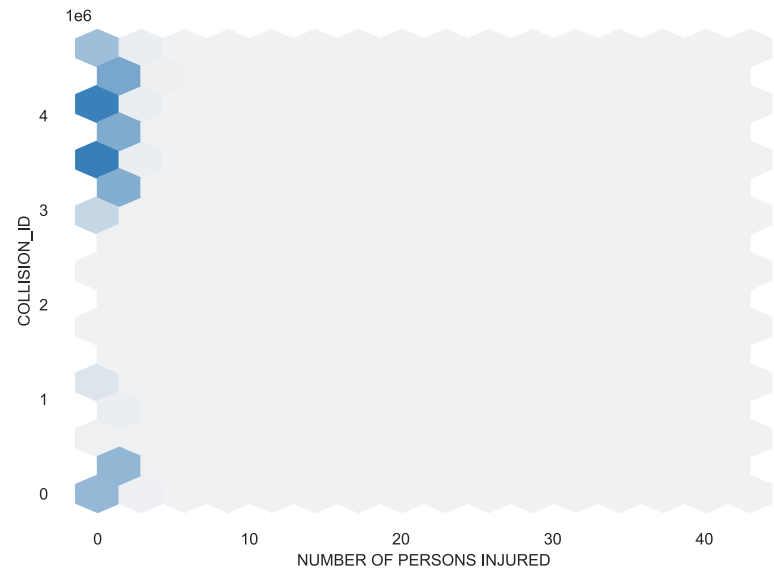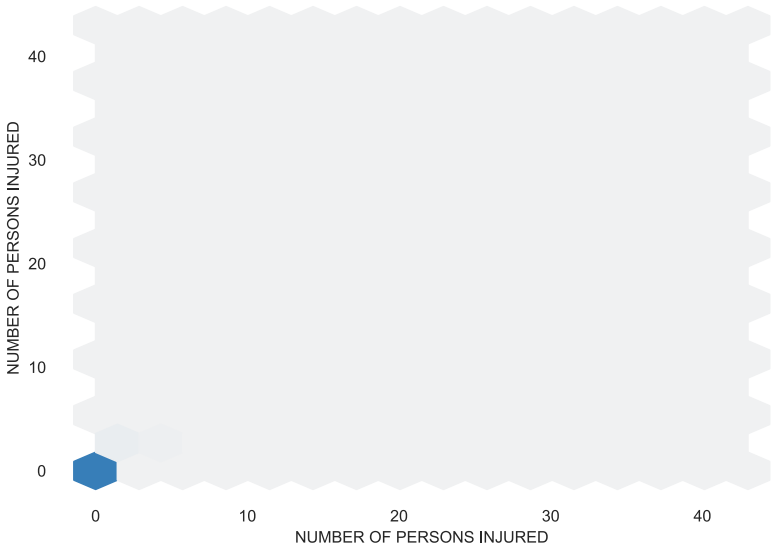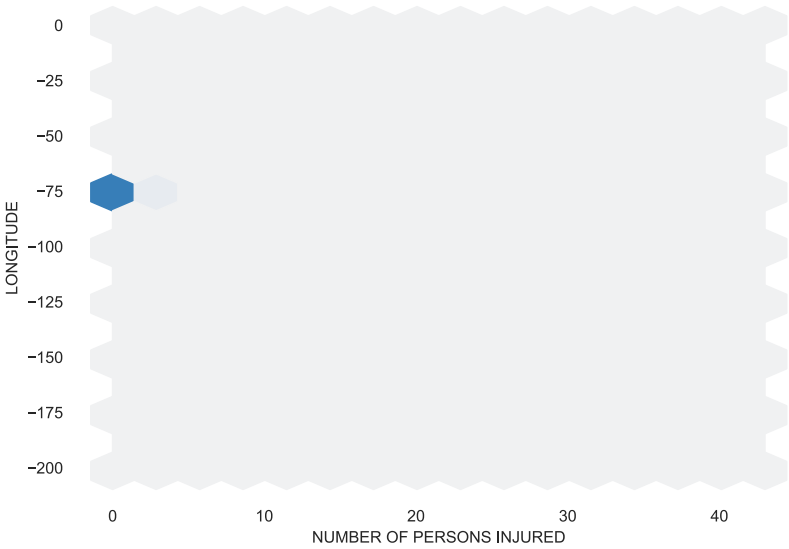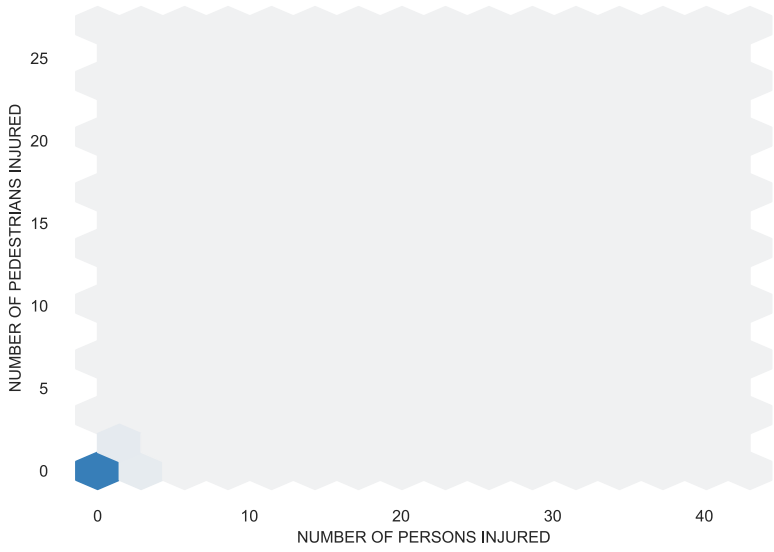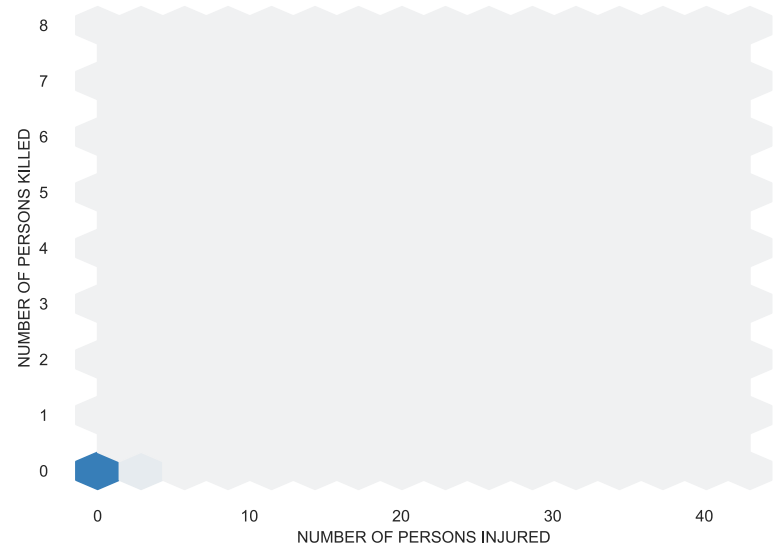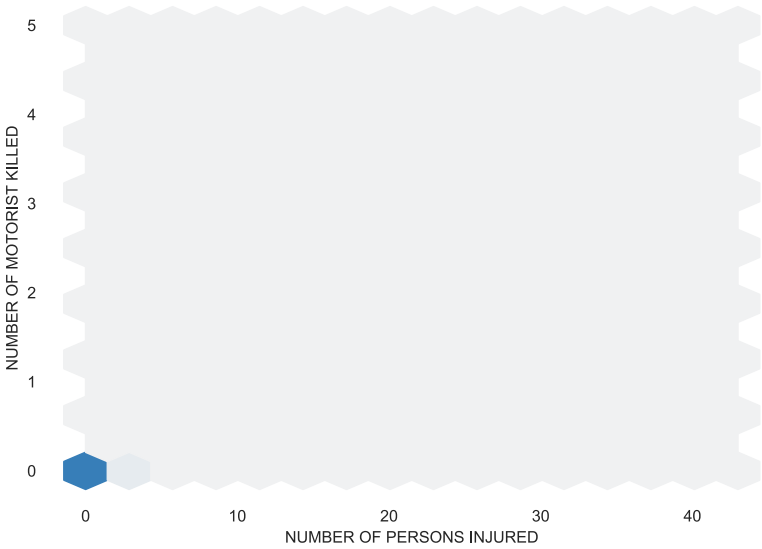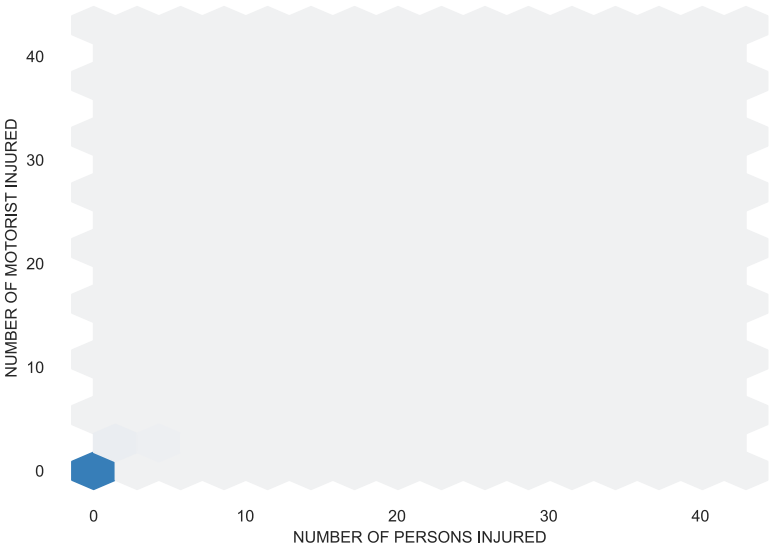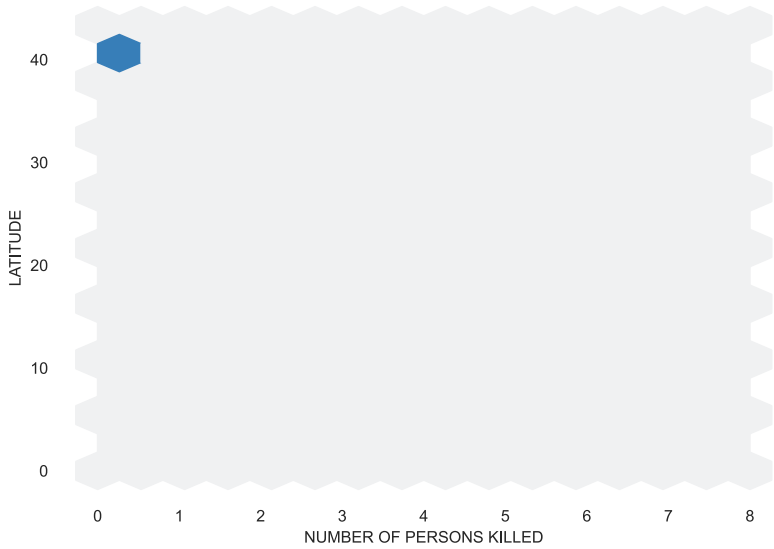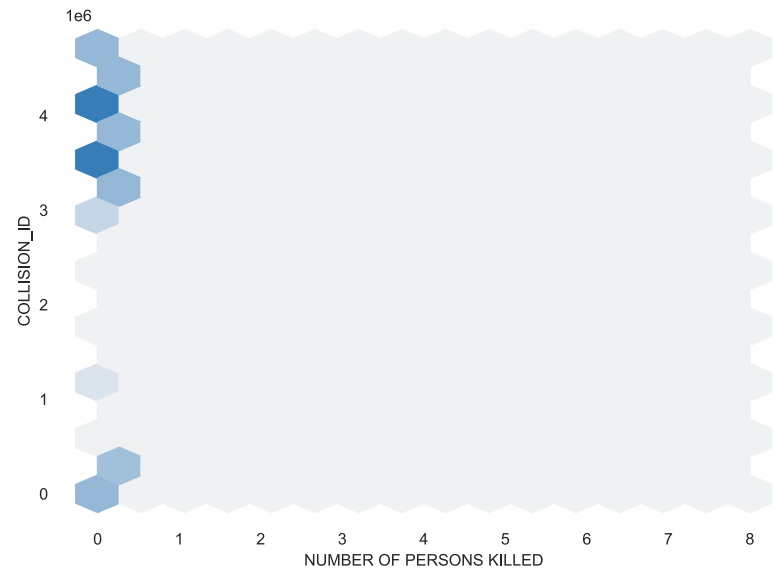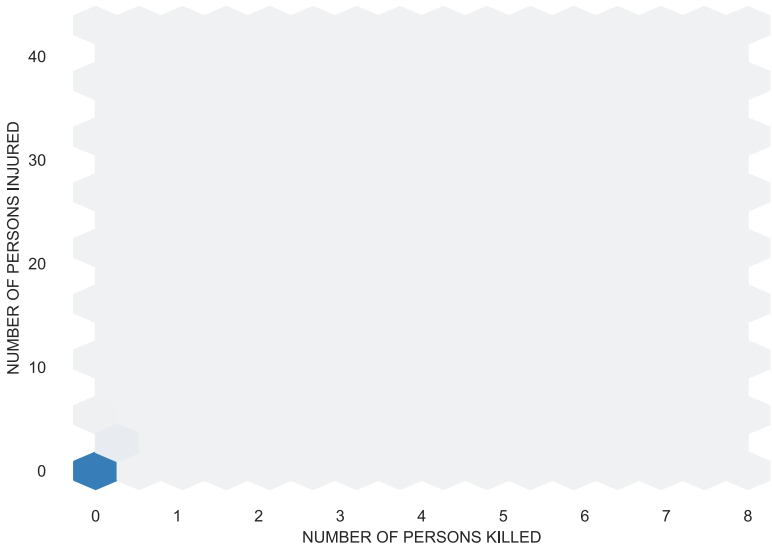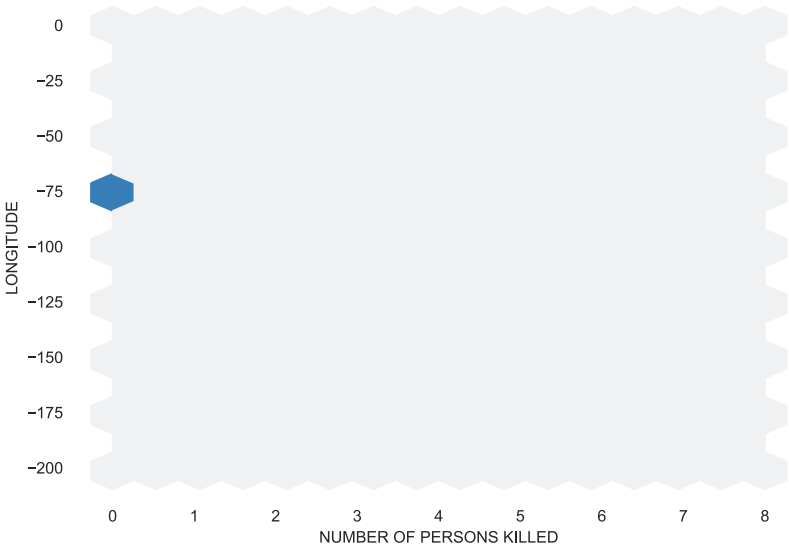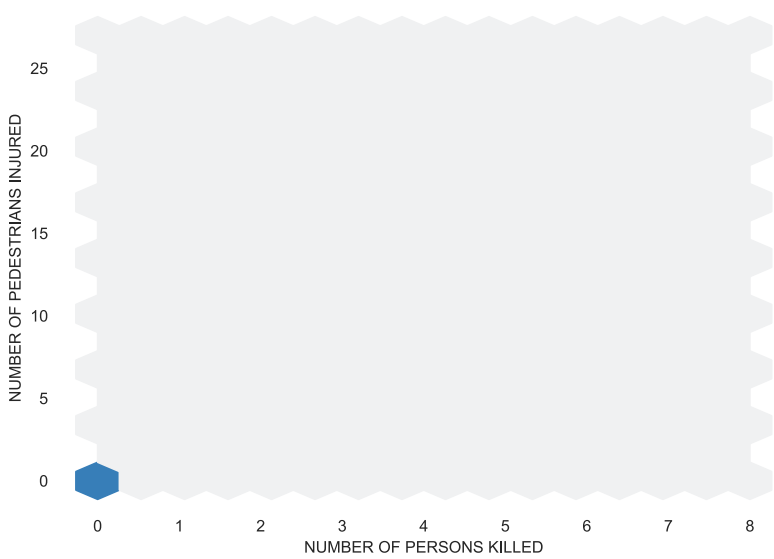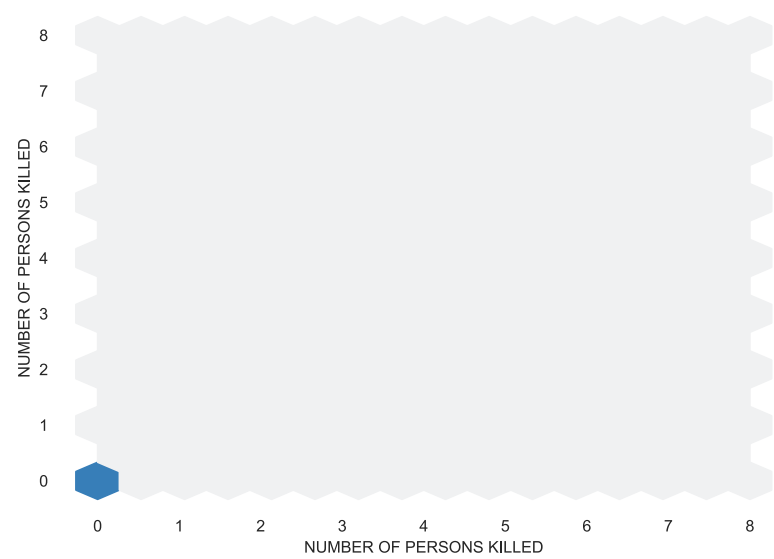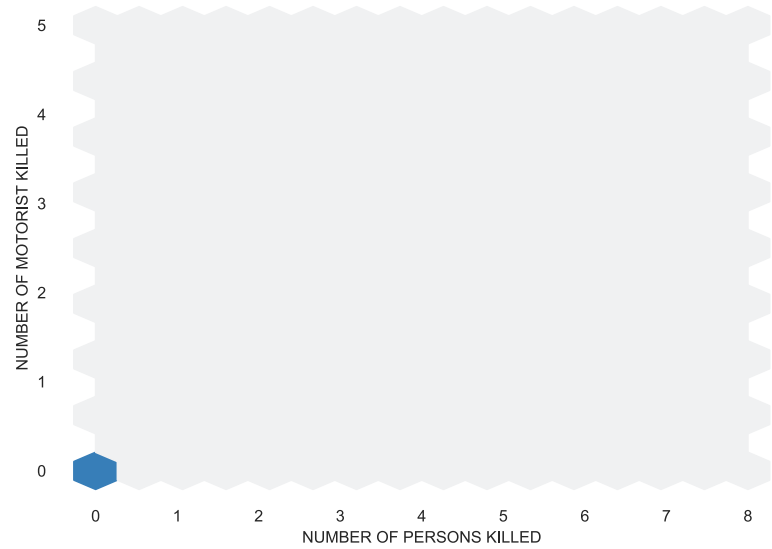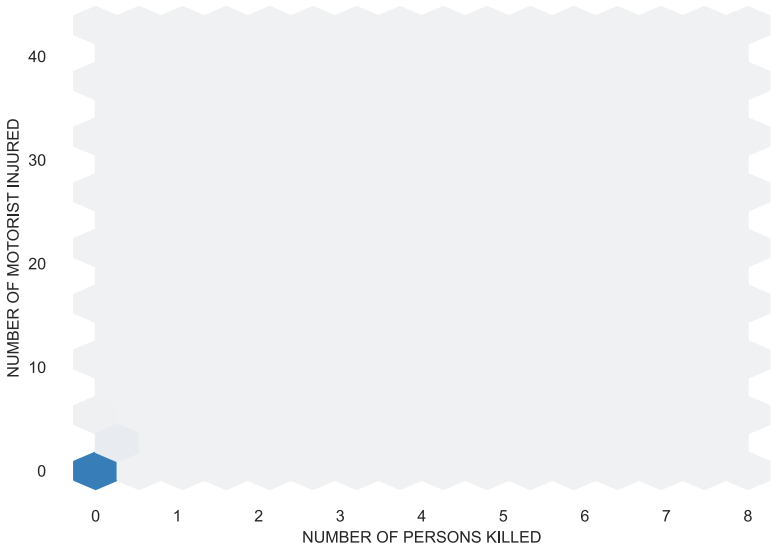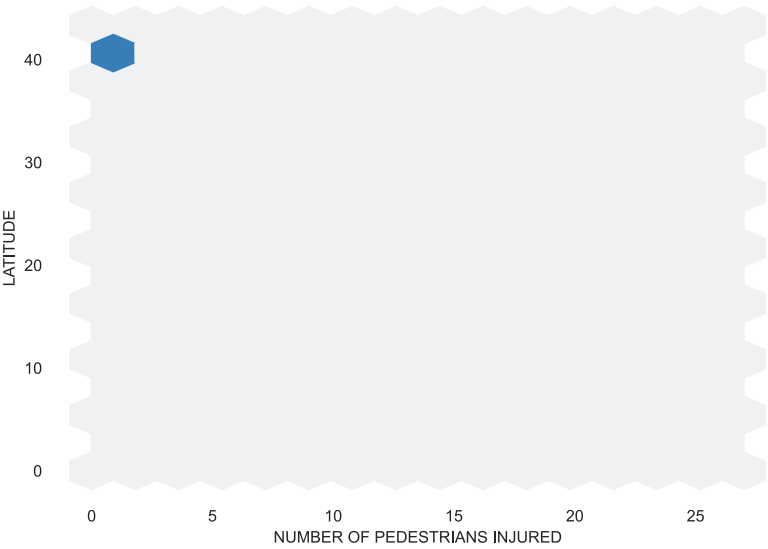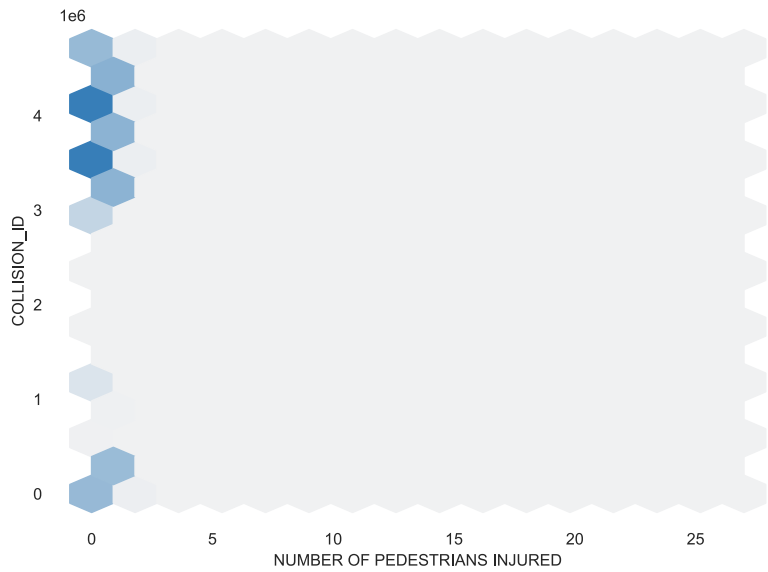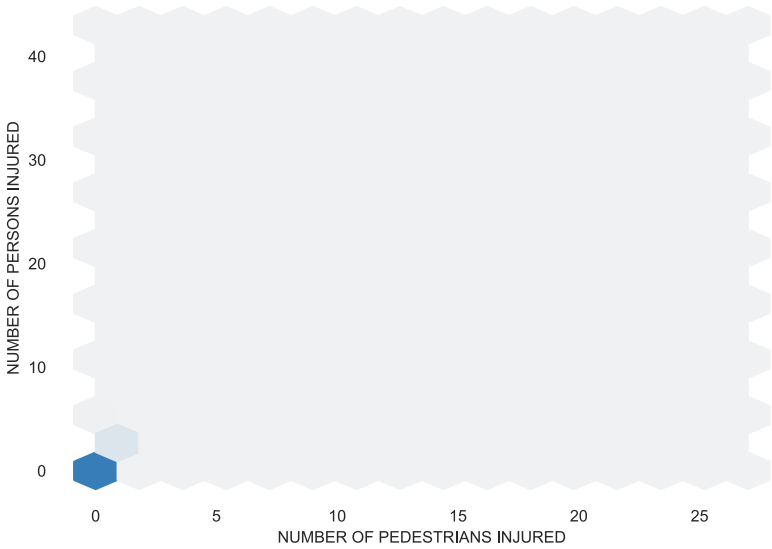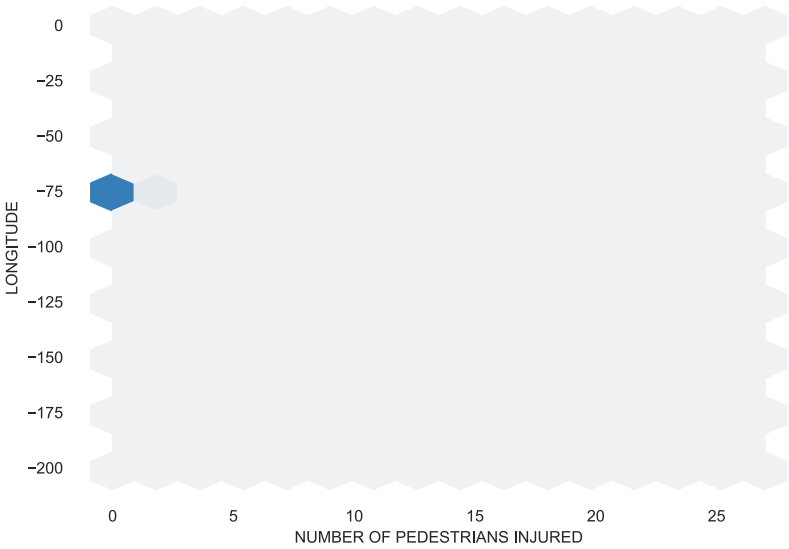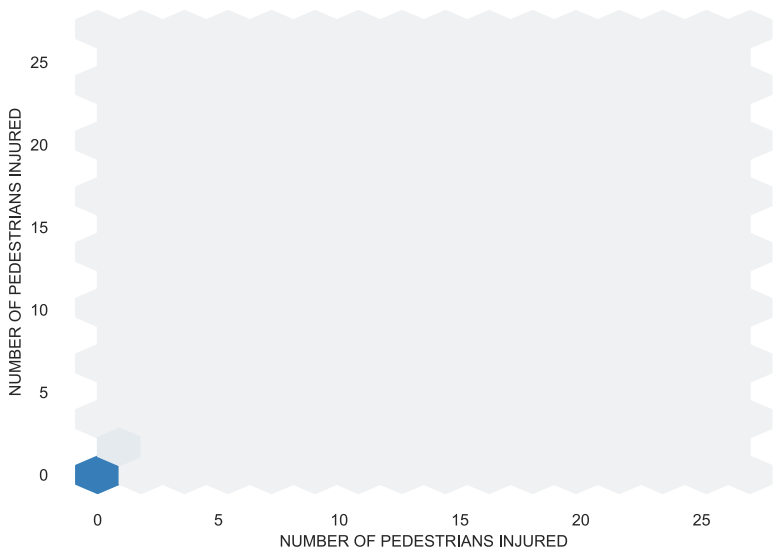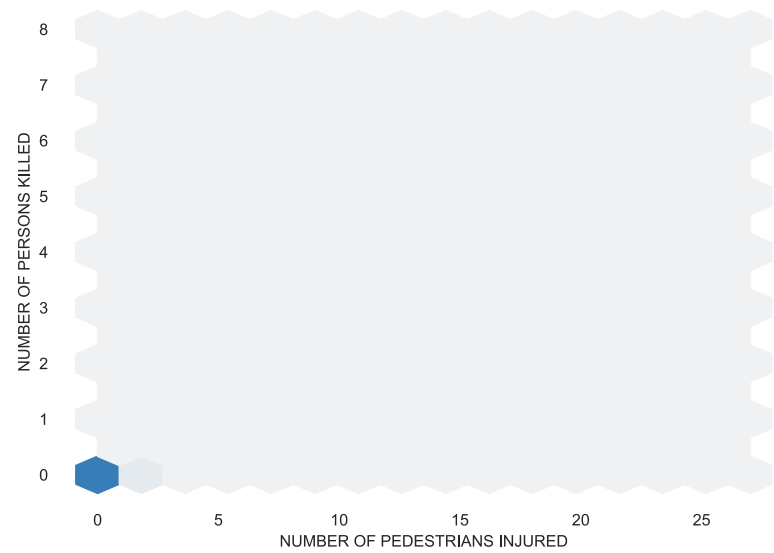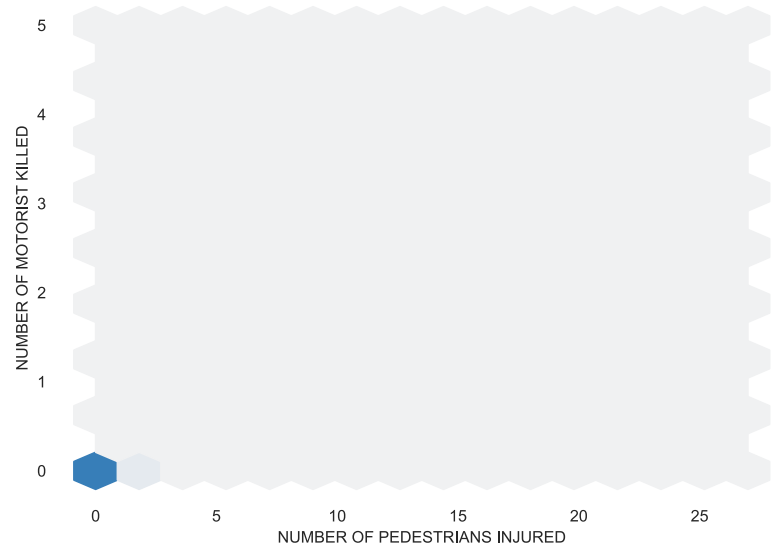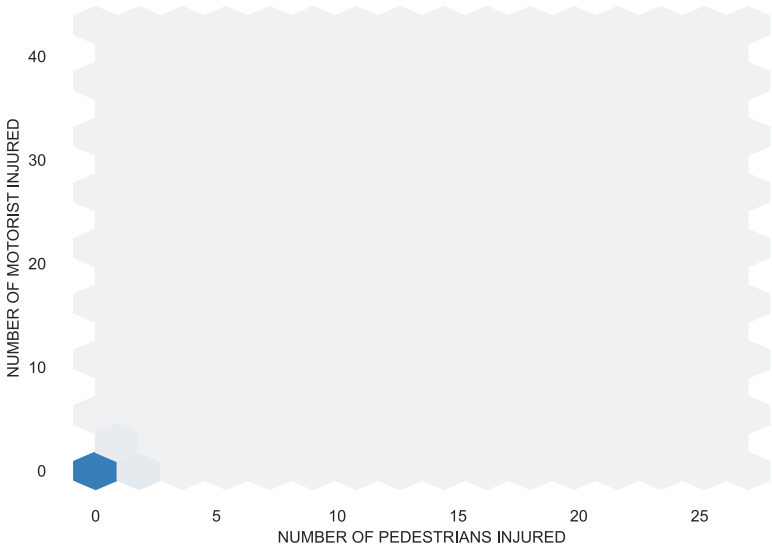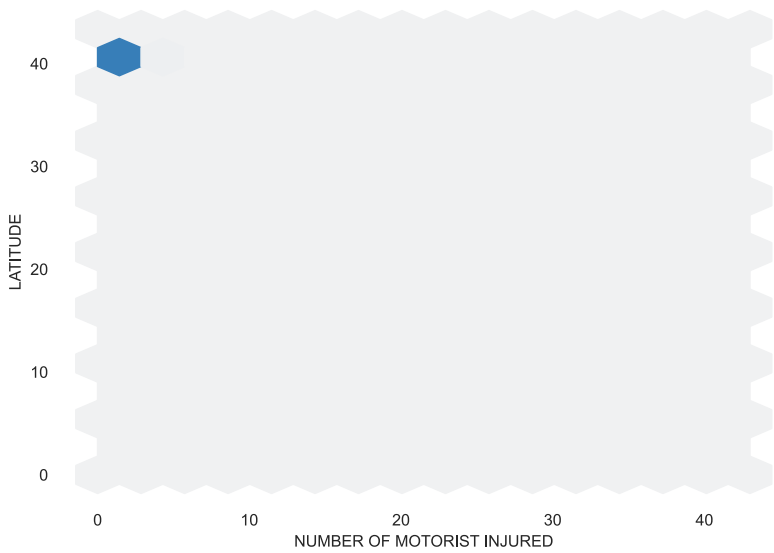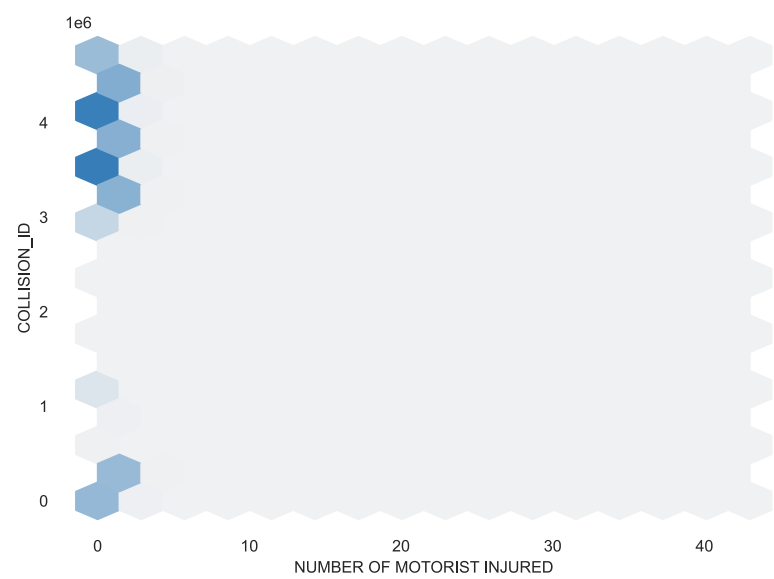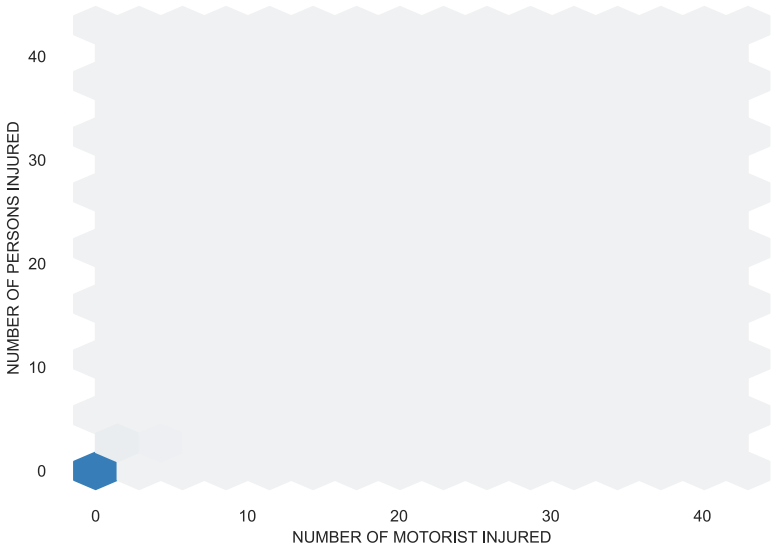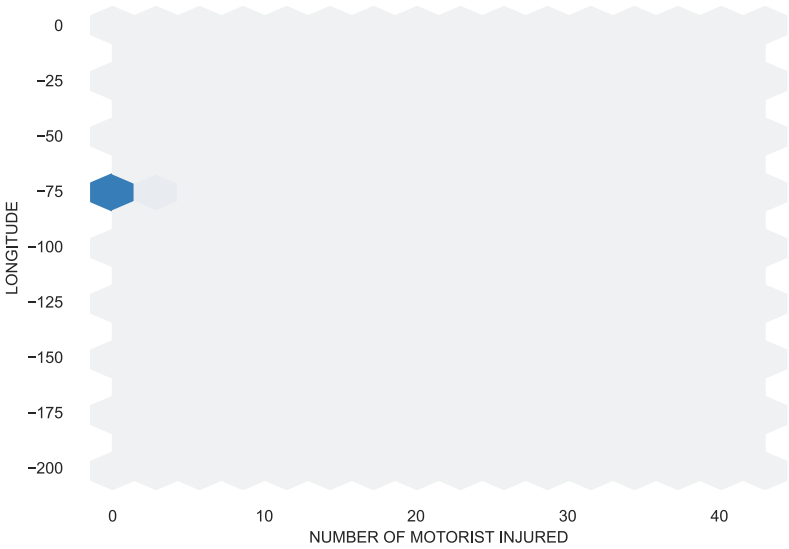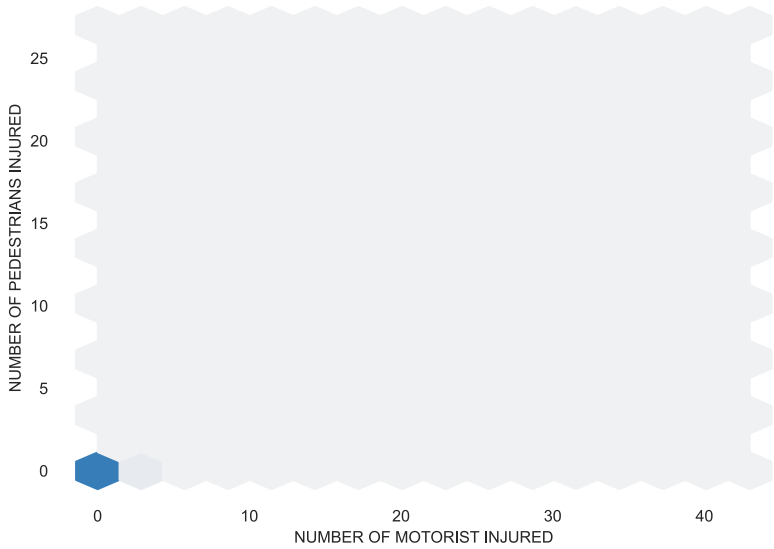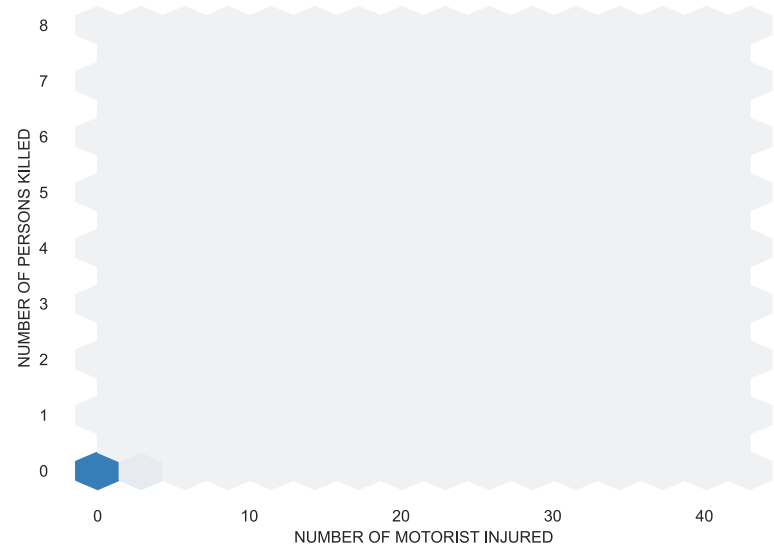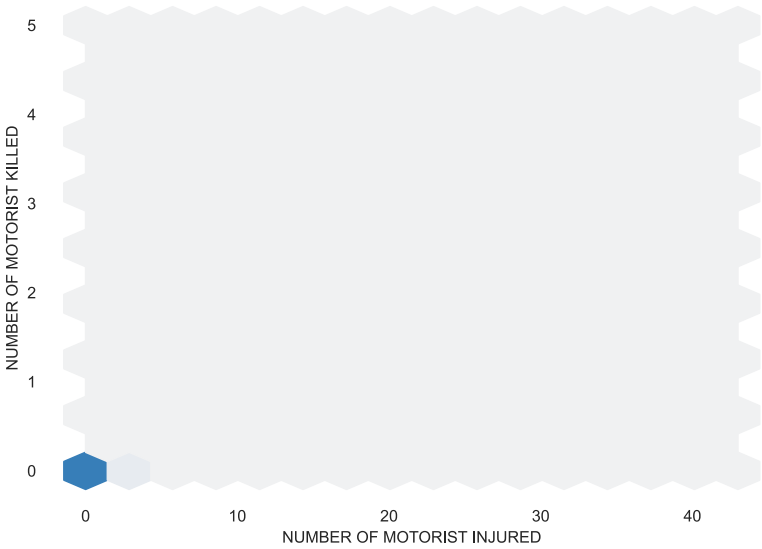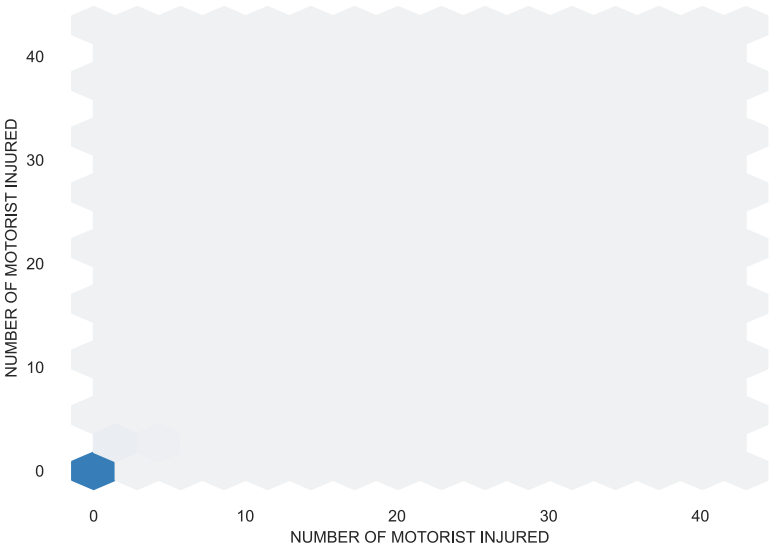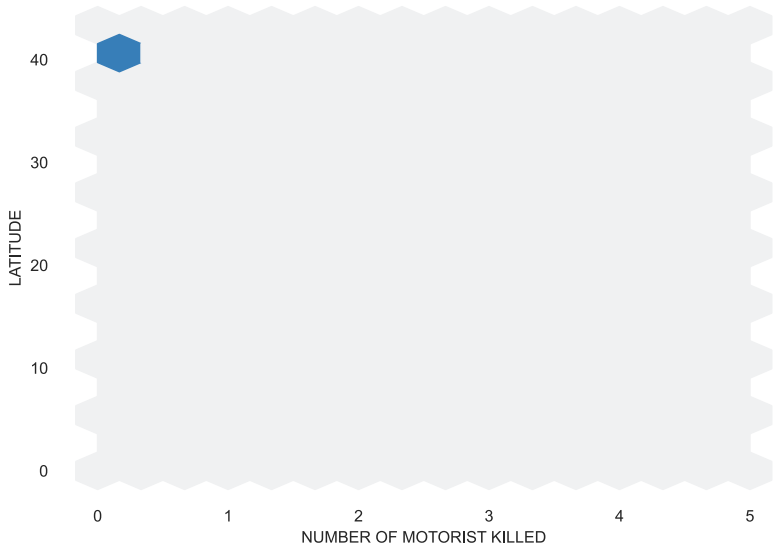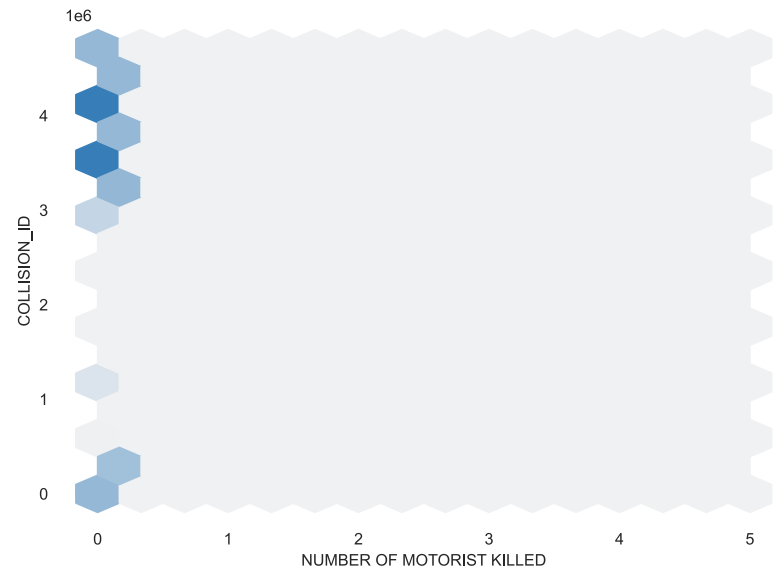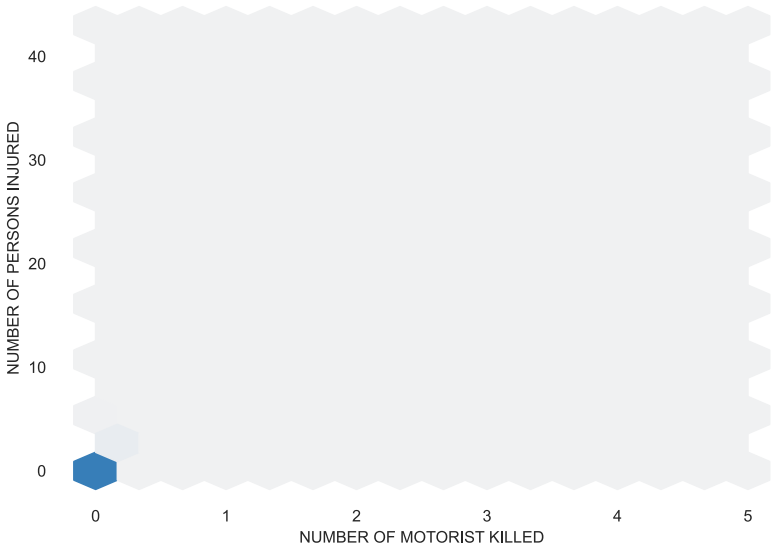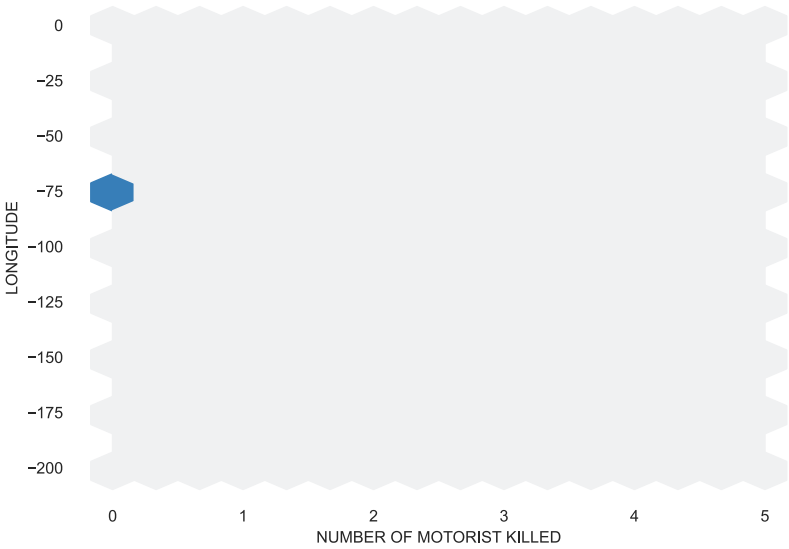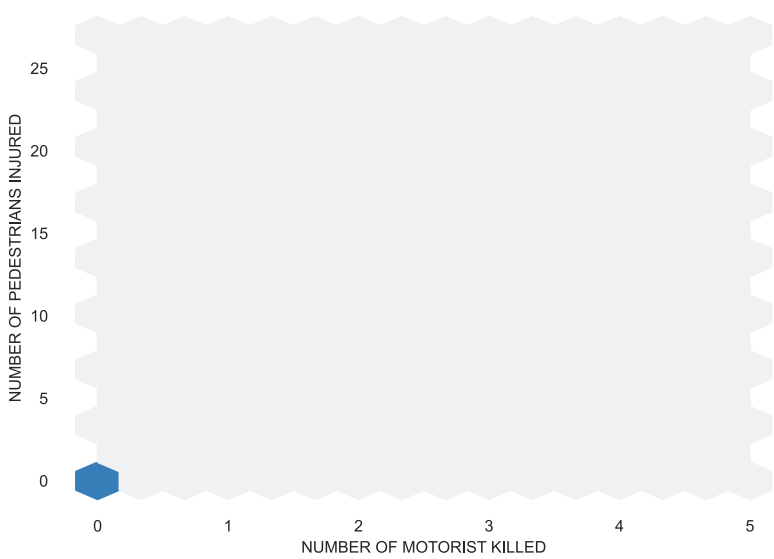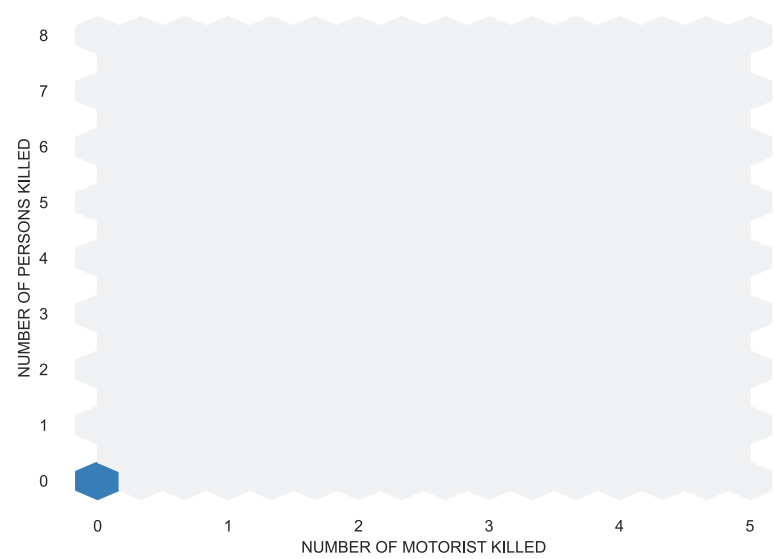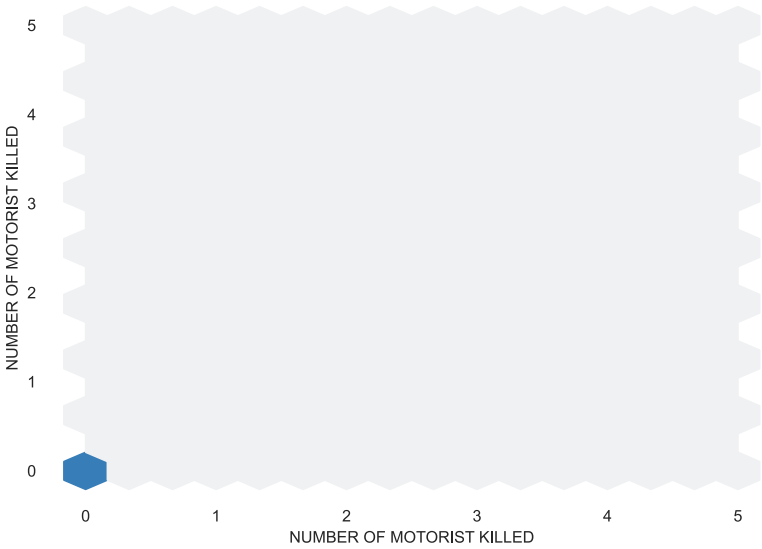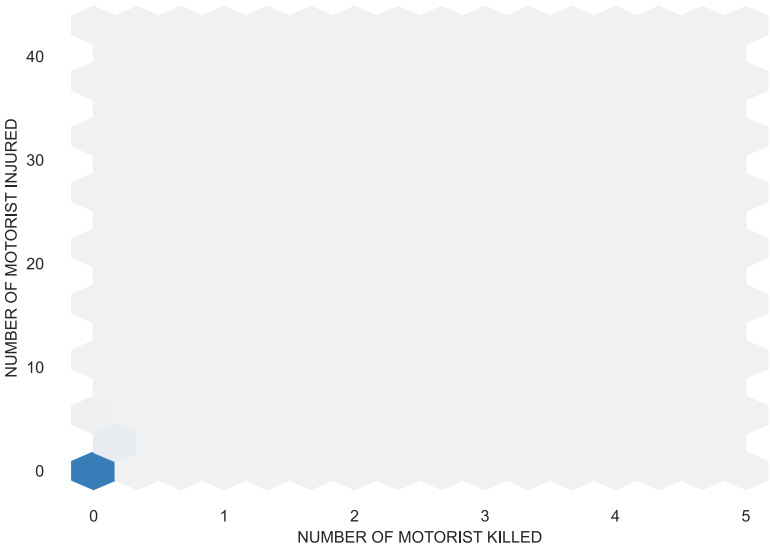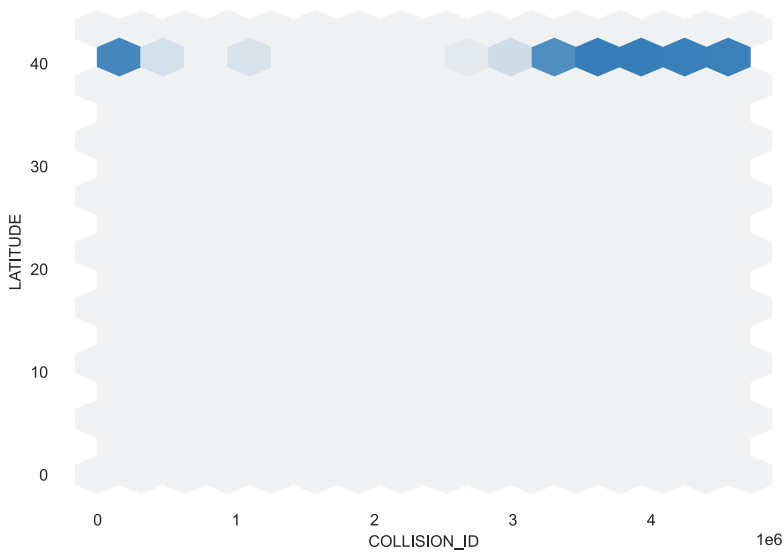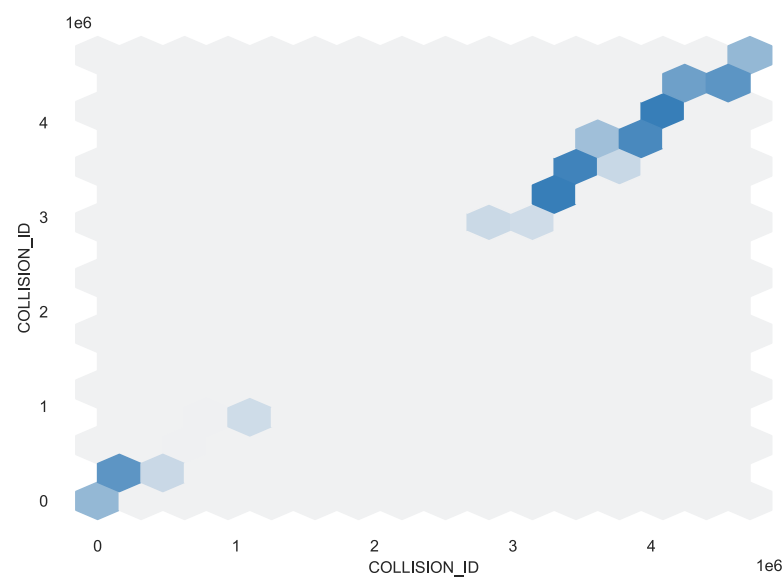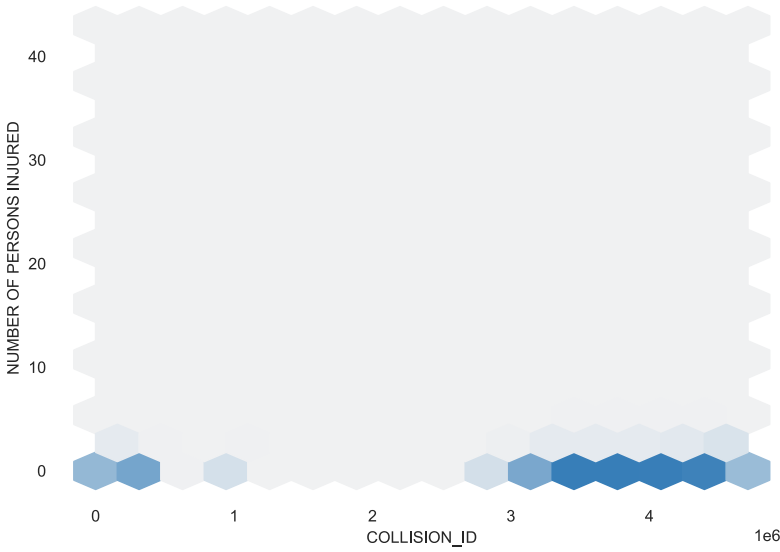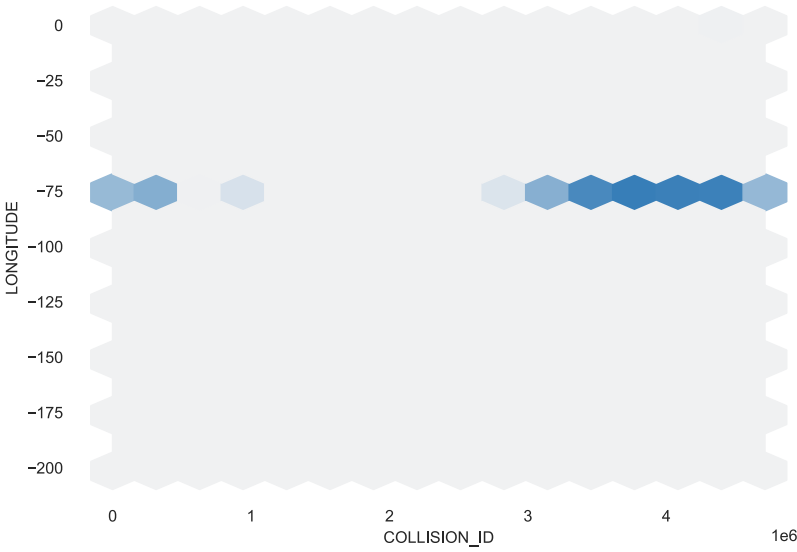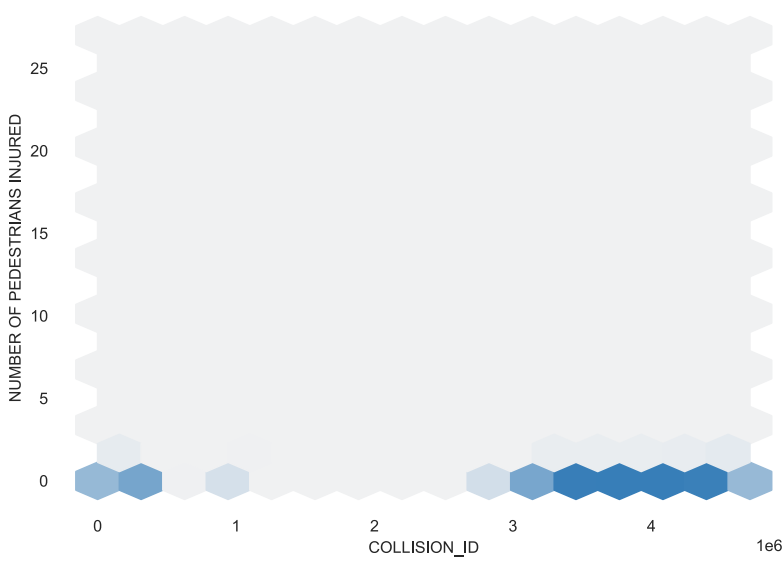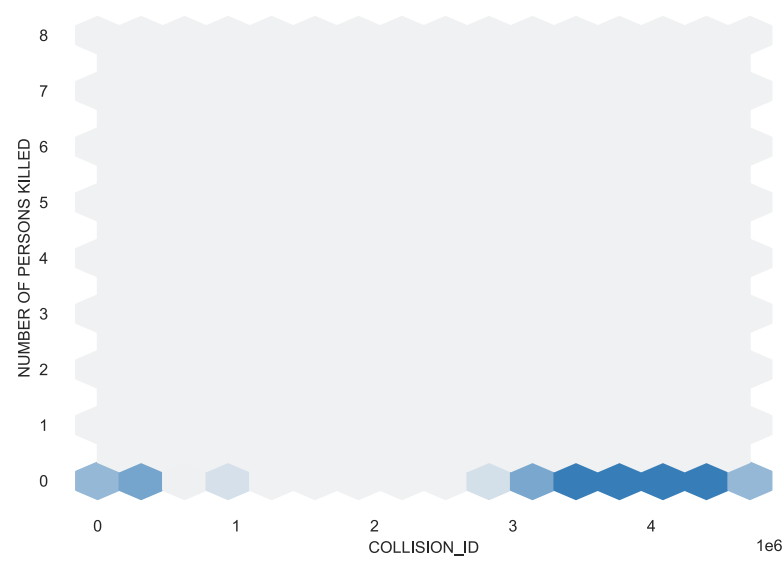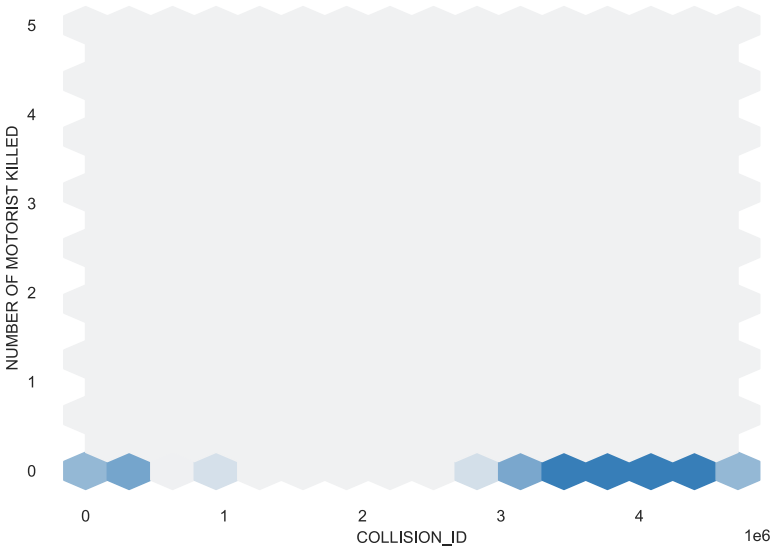
Nullity matrix is a data-dense display which lets you quickly visually pick out patterns in data completion.

The correlation heatmap measures nullity correlation: how strongly the presence or absence of one variable affects the presence of another.

Sample

	CRASH DATE	CRASH TIME	BOROUGH	ZIP CODE	LATITUDE	LONGITUDE	LOCATION	ON STREET NAME	CROS…
0	09/11/2021	2:39	NaN	NaN	NaN	NaN	NaN	WHITESTONE EXPRESSWAY	20 AVE…
1	03/26/2022	11:45	NaN	NaN	NaN	NaN	NaN	QUEENSBORO BRIDGE UPPER	NaN
2	06/29/2022	6:55	NaN	NaN	NaN	NaN	NaN	THROGS NECK BRIDGE	NaN
3	09/11/2021	9:35	BROOKLYN	11208	40.667202	-73.866500	(40.667202, -73.8665)	NaN	NaN
4	12/14/2021	8:13	BROOKLYN	11233	40.683304	-73.917274	(40.683304, -73.917274)	SARATOGA AVENUE	DECAT…
5	04/14/2021	12:47	NaN	NaN	NaN	NaN	NaN	MAJOR DEEGAN EXPRESSWAY RAMP	NaN
6	12/14/2021	17:05	NaN	NaN	40.709183	-73.956825	(40.709183, -73.956825)	BROOKLYN QUEENS EXPRESSWAY	NaN
7	12/14/2021	8:17	BRONX	10475	40.868160	-73.831480	(40.86816, -73.83148)	NaN	NaN
8	12/14/2021	21:10	BROOKLYN	11207	40.671720	-73.897100	(40.67172, -73.8971)	NaN	NaN
9	12/14/2021	14:58	MANHATTAN	10017	40.751440	-73.973970	(40.75144, -73.97397)	3 AVENUE	EAST 4…

	CRASH DATE	CRASH TIME	BOROUGH	ZIP CODE	LATITUDE	LONGITUDE	LOCATION	ON STREET NAME	CROSS S…
2075417	03/05/2024	20:40	QUEENS	11375	40.722622	-73.849144	(40.722622, -73.849144)	YELLOWSTONE BOULEVARD	GERARD …
2075418	03/05/2024	7:30	NaN	NaN	40.772953	-73.920280	(40.772953, -73.92028)	26 STREET	HOYT AVE…
2075419	03/05/2024	14:50	NaN	NaN	40.646000	-73.971750	(40.646, -73.97175)	CHURCH AVENUE	EAST 8 ST…
2075420	03/05/2024	14:00	NaN	NaN	40.722250	-74.005920	(40.72225, -74.00592)	CANAL STREET	AVENUE …

Austin Crash Report Data - Profiling Report

	CRASH DATE	CRASH TIME	BOROUGH	ZIP CODE	LATITUDE	LONGITUDE	LOCATION	ON STREET NAME	CROSS S
2075421	02/06/2024	12:37	BROOKLYN	11235	40.586670	-73.966156	(40.58667, -73.966156)	OCEAN PARKWAY	AVENUE 2
2075422	03/05/2024	17:22	QUEENS	11436	40.680477	-73.792100	(40.680477, -73.7921)	SUTPHIN BOULEVARD	FOCH BO
2075423	03/05/2024	17:00	BROOKLYN	11204	40.610786	-73.978820	(40.610786, -73.97882)	NaN	NaN
2075424	03/03/2024	17:50	NaN	NaN	40.675053	-73.947235	(40.675053, -73.947235)	SAINT MARKS AVENUE	NaN
2075425	03/05/2024	14:30	BROOKLYN	11207	40.677900	-73.892586	(40.6779, -73.892586)	MILLER AVENUE	FULTON S
2075426	03/05/2024	8:00	QUEENS	11385	40.706512	-73.878136	(40.706512, -73.878136)	EDSALL AVENUE	73 STREE