

Overview

Dataset statistics		Variable types	
Number of variables	54	Numeric	17
Number of observations	147750	Boolean	11
Missing cells	1725084	DateTime	2
Missing cells (%)	21.6%	Text	8
Duplicate rows	0	Categorical	15
Duplicate rows (%)	0.0%	Unsupported	1
Total size in memory	60.9 MB		
Average record size in memory	432.0 B		

Alerts
pedestrian_fl has constant value ""
motor_vehicle_fl has constant value ""
motorcycle_fl has constant value ""
bicycle_fl has constant value ""
other_fl has constant value ""
other_death_count has constant value ""
private_dr_fl has constant value ""
micromobility_fl has constant value ""
crash_fatal_fl is highly imbalanced (94.8%)
road_constr_zone_fl is highly imbalanced (70.1%)
death_cnt is highly imbalanced (97.7%)
apd_confirmed_fatality is highly imbalanced (94.9%)
apd_confirmed_death_count is highly imbalanced (97.7%)
motor_vehicle_death_count is highly imbalanced (98.8%)
bicycle_death_count is highly imbalanced (99.7%)
bicycle_serious_injury_count is highly imbalanced (99.0%)
pedestrian_death_count is highly imbalanced (98.6%)
pedestrian_serious_injury_count is highly imbalanced (98.3%)
motorcycle_death_count is highly imbalanced (99.4%)
motorcycle_serious_injury_count is highly imbalanced (97.2%)
other_serious_injury_count is highly imbalanced (> 99.9%)
micromobility_serious_injury_count is highly imbalanced (99.8%)
micromobility_death_count is highly imbalanced (99.9%)
case_id has 1858 (1.3%) missing values
rpt_latitude has 137456 (93.0%) missing values
rpt_longitude has 137456 (93.0%) missing values
rpt_block_num has 19611 (13.3%) missing values
rpt_street_pfx has 67805 (45.9%) missing values

rpt_street_sfx has 50340 (34.1%) missing values	Missing
latitude has 2243 (1.5%) missing values	Missing
longitude has 2243 (1.5%) missing values	Missing
street_nbr has 87038 (68.9%) missing values	Missing
street_name_2 has 81474 (55.1%) missing values	Missing
street_nbr_2 has 147750 (100.0%) missing values	Missing
contrib_factr_p1_id has 119143 (80.6%) missing values	Missing
contrib_factr_p2_id has 143235 (96.9%) missing values	Missing
pedestrian_fl has 144245 (97.6%) missing values	Missing
motorcycle_fl has 144148 (97.6%) missing values	Missing
bicycle_fl has 145306 (98.3%) missing values	Missing
other_fl has 142905 (96.7%) missing values	Missing
point has 2243 (1.5%) missing values	Missing
micromobility_fl has 147439 (99.8%) missing values	Missing
crash_id has unique values	Unique
street_nbr_2 is an unsupported type, check if it needs cleaning or further analysis	Unsupported
crash_sev_id has 4851 (3.3%) zeros	Zeros
sus_serious_injry_cnt has 143326 (97.0%) zeros	Zeros
nonincap_injry_cnt has 117724 (79.7%) zeros	Zeros
poss_injry_cnt has 111987 (75.8%) zeros	Zeros
non_injry_cnt has 24343 (16.5%) zeros	Zeros
unkn_injry_cnt has 133316 (90.2%) zeros	Zeros
tot_injry_cnt has 82717 (56.0%) zeros	Zeros
motor_vehicle_serious_injury_count has 144928 (98.1%) zeros	Zeros

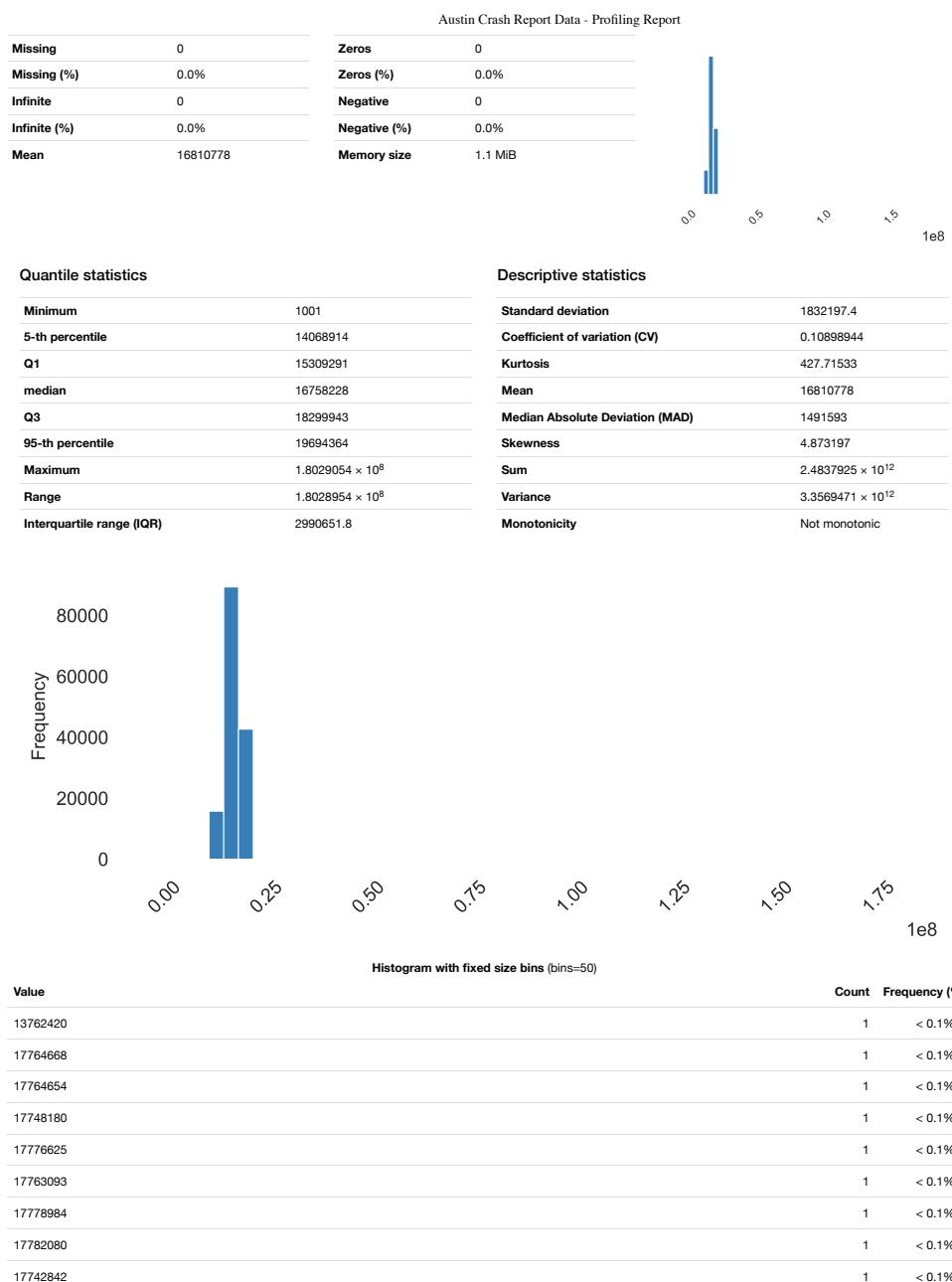
Reproduction

Analysis started	2024-04-08 03:20:27.310743
Analysis finished	2024-04-08 03:21:25.137161
Duration	57.83 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-configuration%20Profiling%20Report%22%2C%20%22dataset%22%3A%20%7B%22description%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_id	Real number (R)
UNIQUE	
Distinct	147750
Distinct (%)	100.0%
Minimum	1001
Maximum	1.8029054 × 10 ⁸



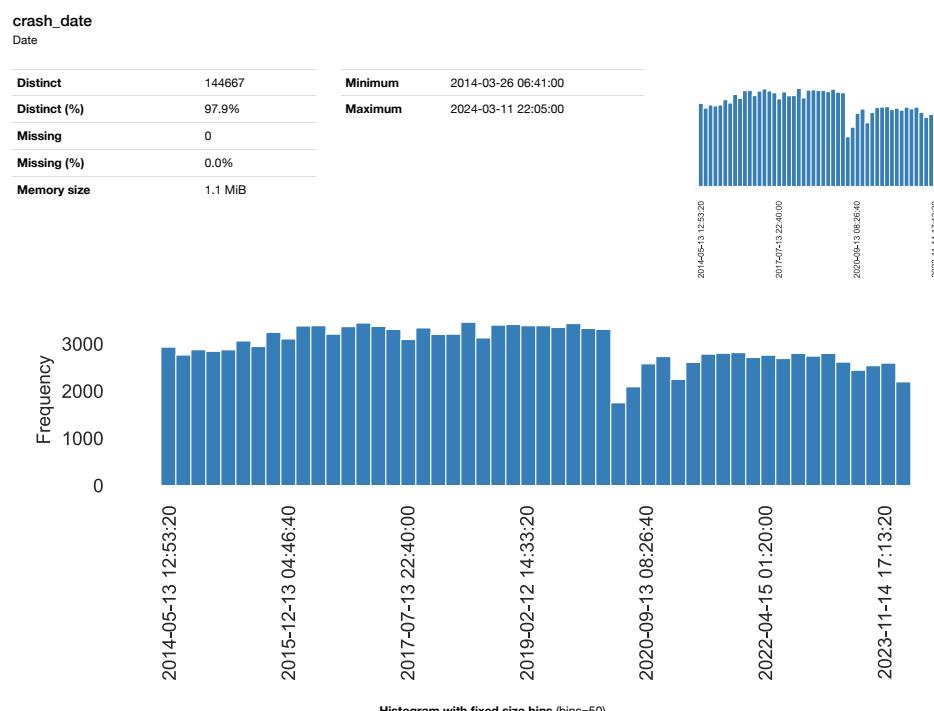
Austin Crash Report Data - Profiling Report

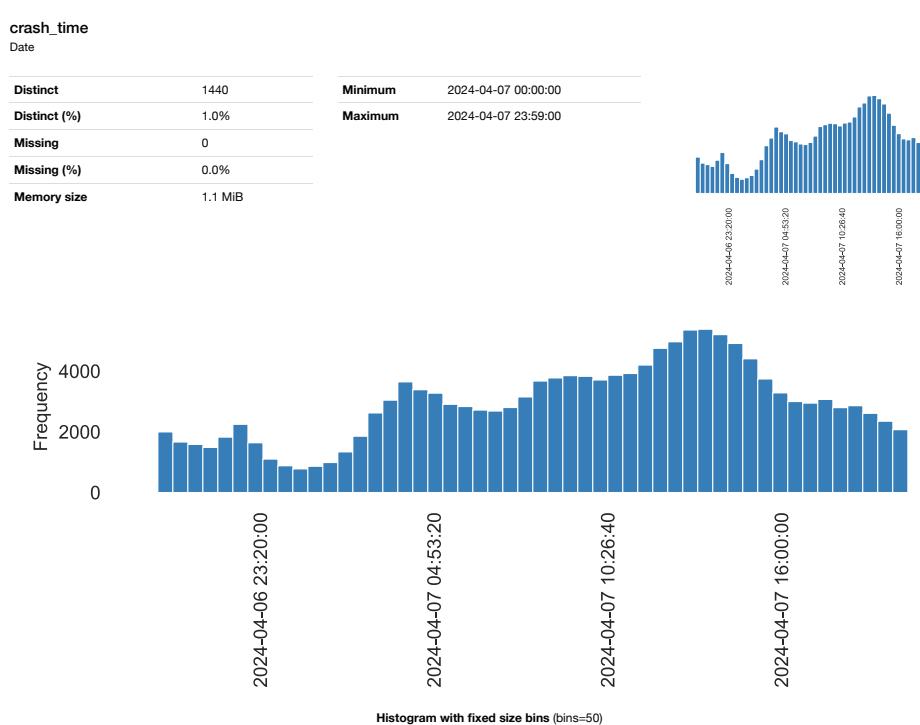
Value	Count	Frequency (%)
17757259	1	< 0.1%
Other values (147740)	147740	> 99.9%

Value	Count	Frequency (%)
1001	1	< 0.1%
13756880	1	< 0.1%
13756881	1	< 0.1%
13756928	1	< 0.1%
13756945	1	< 0.1%
13756946	1	< 0.1%
13756947	1	< 0.1%
13756956	1	< 0.1%
13756957	1	< 0.1%
13756958	1	< 0.1%

Value	Count	Frequency (%)
180290542	1	< 0.1%
20087189	1	< 0.1%
20085660	1	< 0.1%
20085537	1	< 0.1%
20085478	1	< 0.1%
20085477	1	< 0.1%
20085476	1	< 0.1%
20085475	1	< 0.1%
20085474	1	< 0.1%
20085473	1	< 0.1%

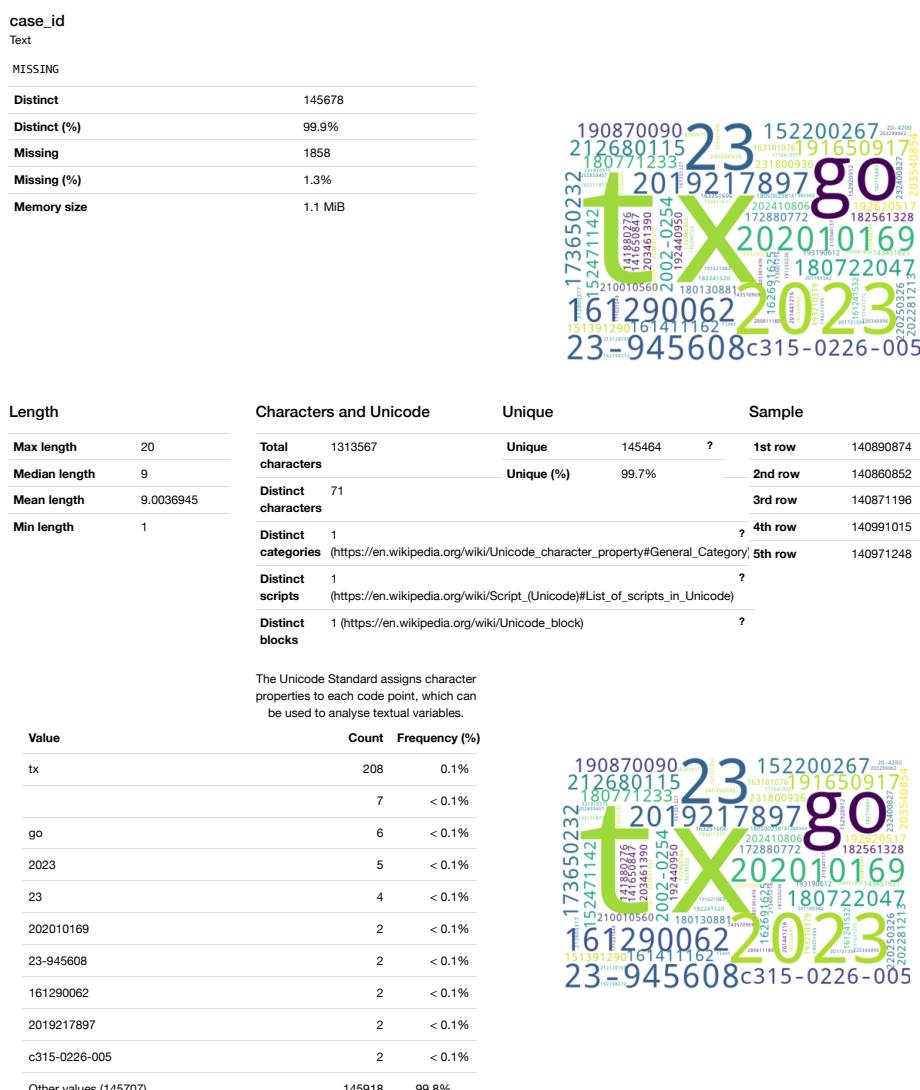
crash_fatal_flg	
Boolean	
IMBALANCE	
Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	144.4 KIB
Value	
False	146884 99.4%
True	866 0.6%





file:///Users/pratikkpoojari/Downloads/html/Austin_Profiling_Report.html

7/225



Montage

Value	Count	Frequency (%)
1	284315	21.6%
0	204676	15.6%
2	187630	14.3%
3	118323	9.0%
4	90263	6.9%
5	88855	6.8%
6	86587	6.6%
9	82968	6.3%
7	82780	6.3%
8	82526	6.3%
Other values (61)	4644	0.4%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	1313567	100.0%

Most frequent character per category*(unknown)*

Value	Count	Frequency (%)
1	284315	21.6%
0	204676	15.6%
2	187630	14.3%
3	118323	9.0%
4	90263	6.9%
5	88855	6.8%
6	86587	6.6%
9	82968	6.3%
7	82780	6.3%
8	82526	6.3%
Other values (61)	4644	0.4%

Most occurring scripts*(unknown)*

Value	Count	Frequency (%)
(unknown)	1313567	100.0%

Most frequent character per script*(unknown)*

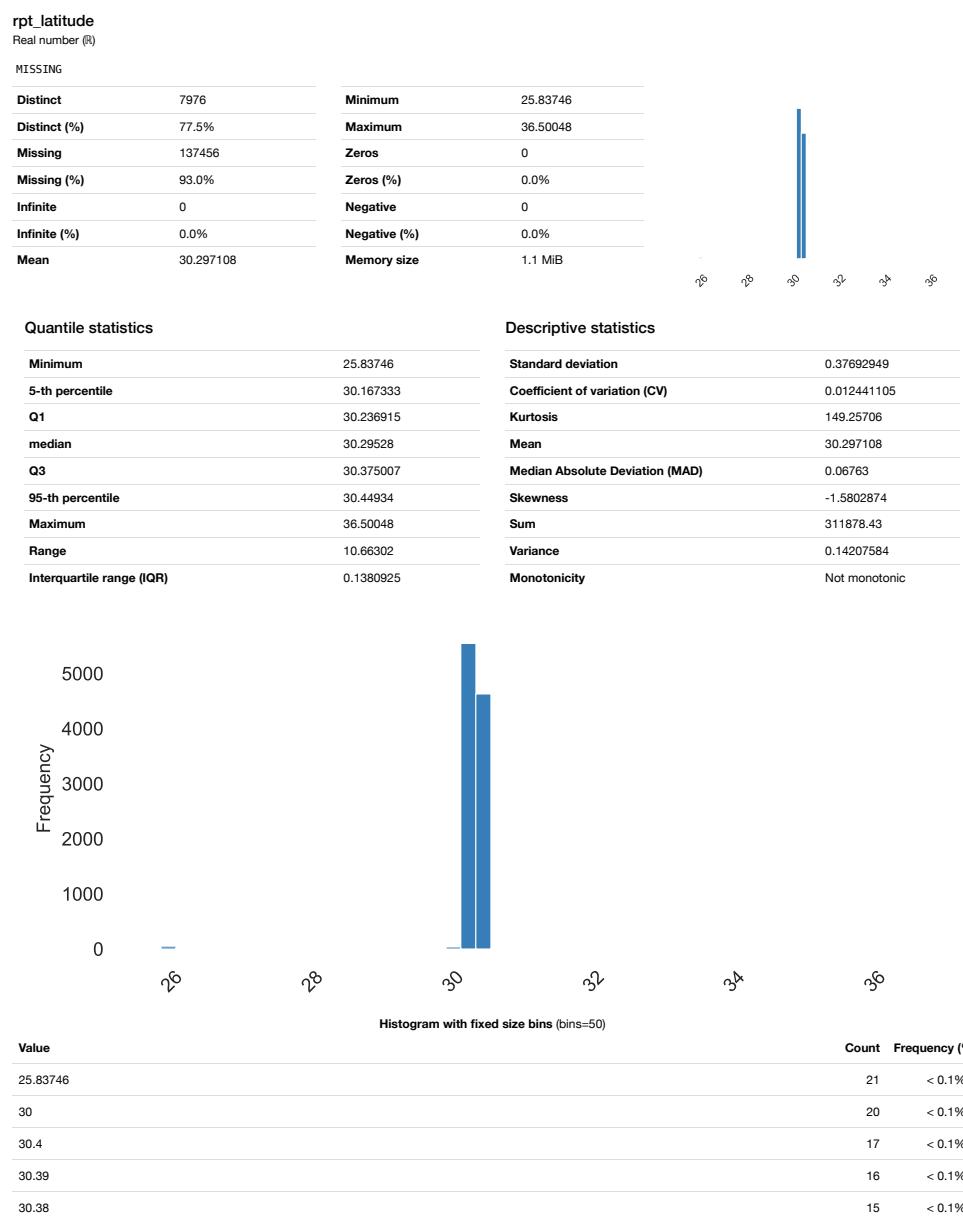
Value	Count	Frequency (%)
1	284315	21.6%
0	204676	15.6%
2	187630	14.3%
3	118323	9.0%
4	90263	6.9%
5	88855	6.8%
6	86587	6.6%
9	82968	6.3%
7	82780	6.3%
8	82526	6.3%
Other values (61)	4644	0.4%

Most occurring blocks*(unknown)*

Value	Count	Frequency (%)
(unknown)	1313567	100.0%

Most frequent character per block*(unknown)*

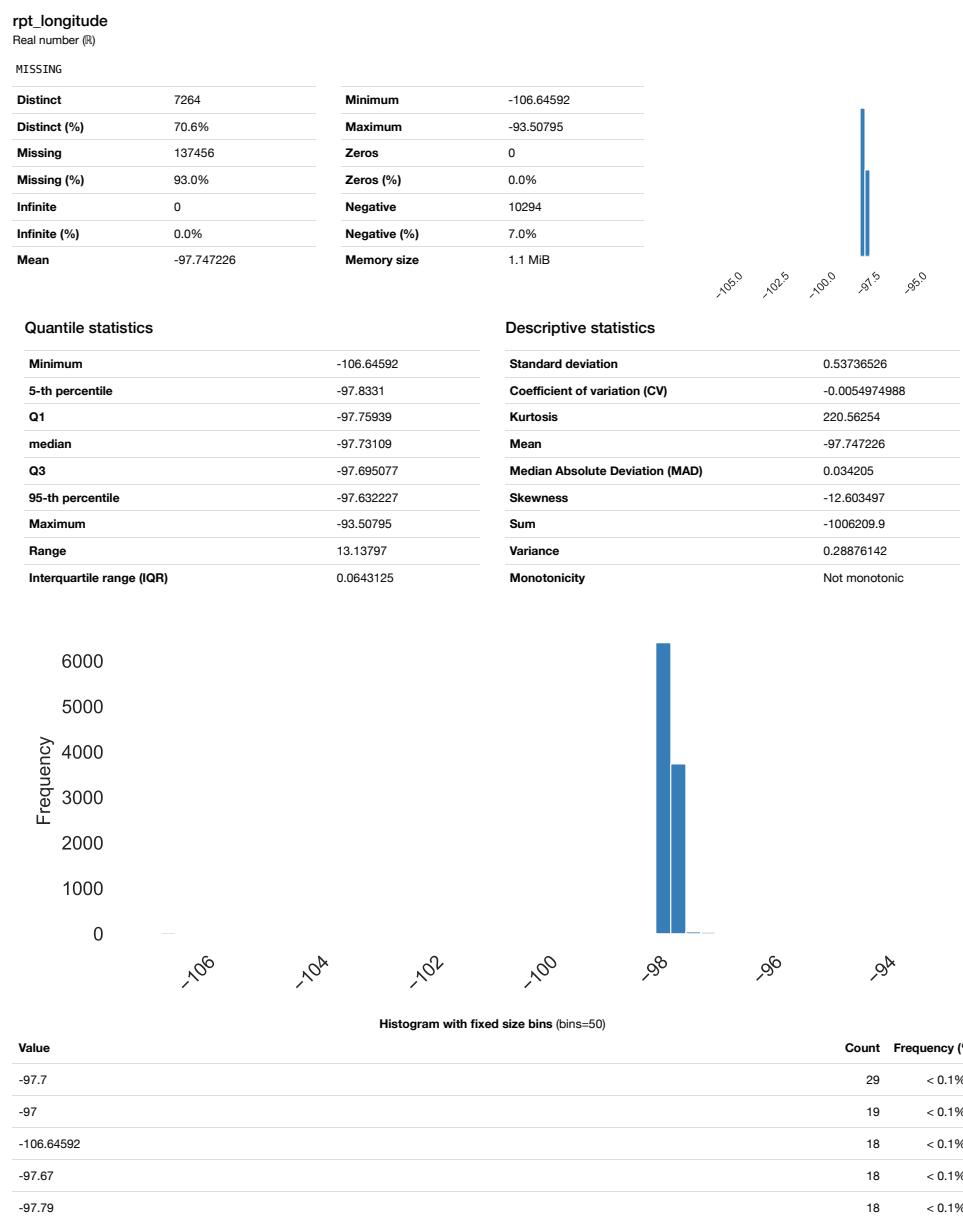
Value	Count	Frequency (%)
1	284315	21.6%
0	204676	15.6%
2	187630	14.3%
3	118323	9.0%
4	90263	6.9%
5	88855	6.8%
6	86587	6.6%
9	82968	6.3%
7	82780	6.3%
8	82526	6.3%
Other values (61)	4644	0.4%



Value	Count	Frequency (%)
30.41426	13	< 0.1%
30.36	13	< 0.1%
30.33426	12	< 0.1%
30.25943	12	< 0.1%
30.41428	12	< 0.1%
Other values (7966)	10143	6.9%
(Missing)	137456	93.0%

Value	Count	Frequency (%)
25.83746	21	< 0.1%
25.83747	1	< 0.1%
25.84	1	< 0.1%
25.84345	1	< 0.1%
25.84563	1	< 0.1%
25.85	1	< 0.1%
25.855	1	< 0.1%
25.9	6	< 0.1%
26	7	< 0.1%
26.1	1	< 0.1%

Value	Count	Frequency (%)
36.50048	3	< 0.1%
36.21798	1	< 0.1%
36.08	1	< 0.1%
36	5	< 0.1%
35.50078	1	< 0.1%
35.3508	1	< 0.1%
35.1	1	< 0.1%
35	1	< 0.1%
34	3	< 0.1%
33.505	1	< 0.1%

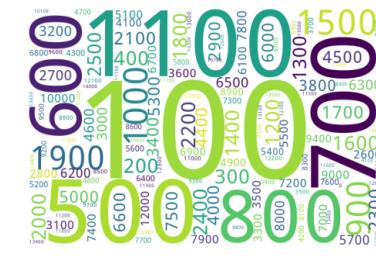


Value	Count	Frequency (%)
-97.69	15	< 0.1%
-97.71	14	< 0.1%
-97.79304	14	< 0.1%
-97.73	12	< 0.1%
-97.69621	12	< 0.1%
Other values (7254)	10125	6.9%
(Missing)	137456	93.0%

Value	Count	Frequency (%)
-106.64592	18	< 0.1%
-106.6459	3	< 0.1%
-106.64	1	< 0.1%
-106.5	3	< 0.1%
-106.45678	1	< 0.1%
-106	3	< 0.1%
-105.1	1	< 0.1%
-105	1	< 0.1%
-104.6	1	< 0.1%
-104	1	< 0.1%

Value	Count	Frequency (%)
-93.50795	6	< 0.1%
-93.6	2	< 0.1%
-93.7	2	< 0.1%
-93.7377	1	< 0.1%
-93.73884	1	< 0.1%
-93.775	1	< 0.1%
-93.786	1	< 0.1%
-94	6	< 0.1%
-94.72418	1	< 0.1%
-95	2	< 0.1%

rpt_block_num	
Text	
MISSING	
Distinct	4788
Distinct (%)	3.7%
Missing	19611
Missing (%)	13.3%
Memory size	1.1 MiB



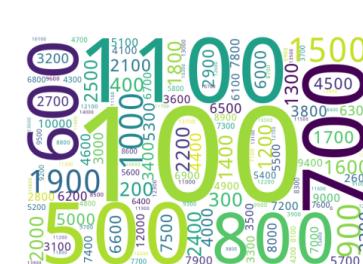
Length
Max length
Median length
Mean length
Min length

Characters and Unicode

Total characters	Unique
511734	2496
Distinct characters	1.9%
32	?
Distinct categories	1
(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	5th row
Distinct scripts	?
(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.

Value	Count	Frequency (%)
100	2467	1.9%
1100	2239	1.7%
700	2177	1.7%
500	2084	1.6%
800	2056	1.6%
600	2039	1.6%
1500	1914	1.5%
1000	1831	1.4%
1900	1727	1.3%
900	1704	1.3%
Other values (4767)	107923	84.2%



Most occurring characters

Value
0
1
2
3
5
4
6
7
8
9
Other values (22)

Count	Frequency (%)
258511	50.5%
62344	12.2%
33184	6.5%
26090	5.1%
25005	4.9%
23603	4.6%
22321	4.4%
21046	4.1%
19942	3.9%
19581	3.8%
107	< 0.1%

Most occurring categories

Value
(unknown)

Count	Frequency (%)
511734	100.0%

Most frequent character per category

Value
(unknown)

Count	Frequency (%)
258511	50.5%
62344	12.2%
33184	6.5%
26090	5.1%
25005	4.9%
23603	4.6%
22321	4.4%
21046	4.1%
19942	3.9%
19581	3.8%
107	< 0.1%

Most occurring scripts

Value
(unknown)

Count	Frequency (%)
511734	100.0%

Most frequent character per script

(unknown)		Count	Frequency (%)
Value			
0		258511	50.5%
1		62344	12.2%
2		33184	6.5%
3		26090	5.1%
5		25005	4.9%
4		23603	4.6%
6		22321	4.4%
7		21046	4.1%
8		19942	3.9%
9		19581	3.8%
Other values (22)		107	< 0.1%

Most occurring blocks

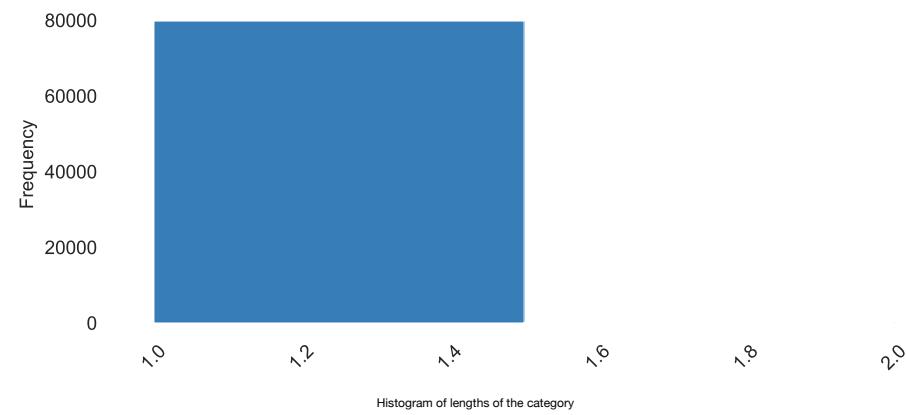
(unknown)		Count	Frequency (%)
		511734	100.0%

Most frequent character per block

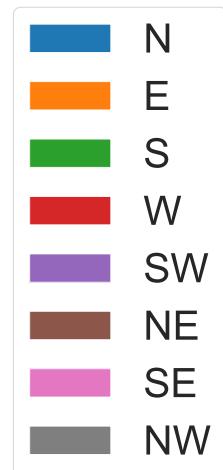
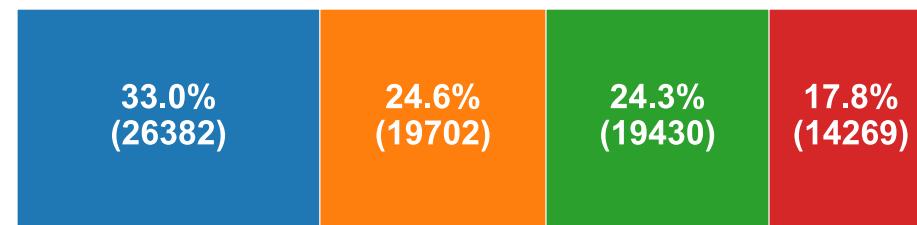
(unknown)		Count	Frequency (%)
Value			
0		258511	50.5%
1		62344	12.2%
2		33184	6.5%
3		26090	5.1%
5		25005	4.9%
4		23603	4.6%
6		22321	4.4%
7		21046	4.1%
8		19942	3.9%
9		19581	3.8%
Other values (22)		107	< 0.1%

rpt_street_pfx	Categorical					
MISSING						
Distinct						
	8					
Distinct (%)	< 0.1%					
Missing	67805					
Missing (%)	45.9%					
Memory size	1.1 MiB					
Length						
Max length	2					
Median length	1					
Mean length	1.0020264					
Min length	1					
Characters and Unicode						
Total characters	80107	Unique	0	?	1st row	W
		Unique (%)	0.0%		2nd row	S
Distinct characters	4				3rd row	E
					4th row	N
Distinct categories	1			?	5th row	S
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			?		
The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.						
Common Values						
Value						
		Count	Frequency (%)			
N		26382	17.9%			
E		19702	13.3%			
S		19430	13.2%			
W		14269	9.7%			
SW		49	< 0.1%			
NE		45	< 0.1%			
SE		38	< 0.1%			
NW		30	< 0.1%			
(Missing)		67805	45.9%			

Length



Common Values (Plot)



Value	Count	Frequency (%)
n	26382	33.0%
e	19702	24.6%
s	19430	24.3%
w	14269	17.8%
sw	49	0.1%
ne	45	0.1%
se	38	< 0.1%
nw	30	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
N	26457	33.0%
E	19785	24.7%
S	19517	24.4%
W	14348	17.9%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	80107	100.0%

Most frequent character per category

Value	Count	Frequency (%)
N	26457	33.0%
E	19785	24.7%
S	19517	24.4%
W	14348	17.9%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	80107	100.0%

Most frequent character per script

Value	Count	Frequency (%)
(unknown)	80107	100.0%

Value	Count	Frequency (%)
N	26457	33.0%
E	19785	24.7%
S	19517	24.4%
W	14348	17.9%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	80107	100.0%

Most frequent character per block

Value	Count	Frequency (%)
N	26457	33.0%
E	19785	24.7%
S	19517	24.4%
W	14348	17.9%

rpt_street_name

Text

Distinct	9794
Distinct (%)	6.6%
Missing	3
Missing (%)	< 0.1%
Memory size	1.1 MiB

**Length**

Max length	49
Median length	41
Mean length	10.456382
Min length	1

Characters and Unicode

Total characters	Unique	?	1st row	Sample
1544899	5596	?	3707 MANCHACA	
	Unique (%)	3.8%		
37			2nd row	PALM WAY TO MOPAC NB
Distinct characters			RAMP	
1			?	BALCONES CLUB DR
Distinct categories (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)		?	3rd row	
1		?	4th row	E US 290 HWY SVRD EB
Distinct scripts (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)		?	5th row	BEN WHITE
1 (https://en.wikipedia.org/wiki/Unicode_block_blocks)		?		

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

Value

Value	Count	Frequency (%)
35	18801	5.5%
ih	18084	5.3%
nb	12693	3.7%
sb	12675	3.7%
svrd	11976	3.5%
n	11393	3.3%
not	10178	3.0%
reported	10178	3.0%
e	8672	2.5%
mopac	8430	2.5%
Other values (4401)	218139	63.9%

**Most occurring characters**

Value	Count	Frequency (%)
E	193472	12.5%
R	127764	8.3%
A	125088	8.1%
N	98571	6.4%
S	95829	6.2%
T	95159	6.2%
O	76211	4.9%
D	74088	4.8%
L	64671	4.2%
Other values (27)	63313	4.1%
	530733	34.4%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	1544899	100.0%

Most frequent character per category*(unknown)*

Value	Count	Frequency (%)
E	193472	12.5%
R	127764	8.3%
A	125088	8.1%
N	98571	6.4%
S	95829	6.2%
T	95159	6.2%
O	76211	4.9%
D	74088	4.8%
L	64671	4.2%
Other values (27)	63313	4.1%
	530733	34.4%

Most occurring scripts*(unknown)*

Value	Count	Frequency (%)
(unknown)	1544899	100.0%

Most frequent character per script*(unknown)*

Value	Count	Frequency (%)
E	193472	12.5%
R	127764	8.3%
A	125088	8.1%
N	98571	6.4%
S	95829	6.2%
T	95159	6.2%
O	76211	4.9%
D	74088	4.8%
L	64671	4.2%
Other values (27)	63313	4.1%
	530733	34.4%

Most occurring blocks*(unknown)*

Value	Count	Frequency (%)
(unknown)	1544899	100.0%

Most frequent character per block*(unknown)*

Value	Count	Frequency (%)
E	193472	12.5%
R	127764	8.3%
A	125088	8.1%
N	98571	6.4%
S	95829	6.2%
T	95159	6.2%
O	76211	4.9%
D	74088	4.8%
L	64671	4.2%
Other values (27)	63313	4.1%
	530733	34.4%

rpt_street_sfx

Categorical

MISSING

Distinct	18
Distinct (%)	< 0.1%
Missing	50340
Missing (%)	34.1%

Memory size 1.1 MiB

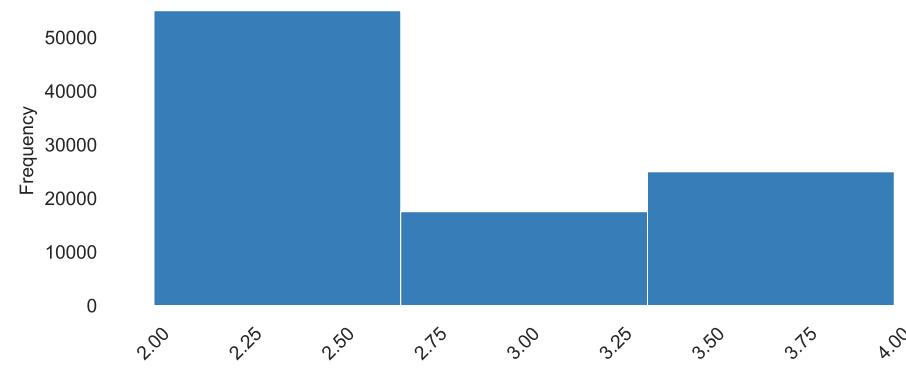
Length

	Characters and Unicode	Unique		Sample
Total characters	262194	Unique	0 ?	1st row RD
Distinct characters	20	Unique (%)	0.0%	2nd row DR
Distinct categories	1			3rd row BLVD
Distinct scripts	1		?	4th row RD
Distinct blocks	1	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?	5th row LN

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

Common Values

Value	Count	Frequency (%)
BLVD	17719	12.0%
RD	15192	10.3%
ST	14489	9.8%
LN	13962	9.4%
HWY	12308	8.3%
DR	11069	7.5%
EXPY	5193	3.5%
AVE	3825	2.6%
PKWY	1615	1.1%
TRL	511	0.3%
Other values (8)	1527	1.0%
(Missing)	50340	34.1%

Length**Value**

Value	Count	Frequency (%)
bld	17719	18.2%
rd	15192	15.6%
st	14489	14.9%
ln	13962	14.3%
hwy	12308	12.6%
dr	11069	11.4%
expy	5193	5.3%
ave	3825	3.9%
pkwy	1615	1.7%
trl	511	0.5%
Other values (8)	1527	1.6%

Most occurring characters

Value	Count	Frequency (%)
D	43980	16.8%
L	32690	12.5%
R	26994	10.3%
V	21614	8.2%
Y	19788	7.5%
B	17719	6.8%
T	15065	5.7%
W	14595	5.6%
S	14489	5.5%
N	13962	5.3%

Value	Count	Frequency (%)
Other values (10)	41298	15.8%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	262194	100.0%

Most frequent character per category

Value	Count	Frequency (%)
D	43980	16.8%
L	32690	12.5%
R	26994	10.3%
V	21614	8.2%
Y	19788	7.5%
B	17719	6.8%
T	15065	5.7%
W	14595	5.6%
S	14489	5.5%
N	13962	5.3%
Other values (10)	41298	15.8%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	262194	100.0%

Most frequent character per script

Value	Count	Frequency (%)
D	43980	16.8%
L	32690	12.5%
R	26994	10.3%
V	21614	8.2%
Y	19788	7.5%
B	17719	6.8%
T	15065	5.7%
W	14595	5.6%

Value	Count	Frequency (%)
S	14489	5.5%
N	13962	5.3%
Other values (10)	41298	15.8%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	262194	100.0%

Most frequent character per block

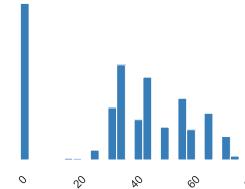
Value	Count	Frequency (%)
D	43980	16.8%
L	32690	12.5%
R	26994	10.3%
V	21614	8.2%
Y	19788	7.5%
B	17719	6.8%
T	15065	5.7%
W	14595	5.6%
S	14489	5.5%
N	13962	5.3%
Other values (10)	41298	15.8%

crash_speed_limit

Real number (R)

Distinct	28
Distinct (%)	< 0.1%
Missing	2
Missing (%)	< 0.1%
Infinite	0
Infinite (%)	0.0%
Mean	34.757526

Minimum	-1
Maximum	85
Zeros	8
Zeros (%)	< 0.1%
Negative	36104
Negative (%)	24.4%
Memory size	1.1 MiB

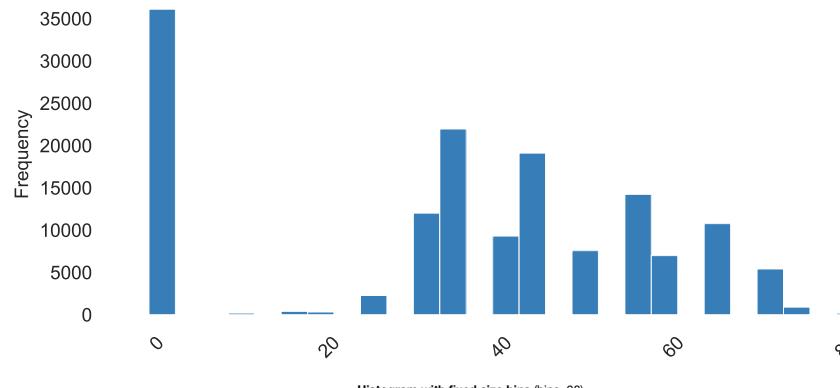


Quantile statistics

Minimum	-1
5-th percentile	-1
Q1	20
median	40
Q3	55
95-th percentile	65
Maximum	85
Range	86
Interquartile range (IQR)	35

Descriptive statistics

Standard deviation	23.232703
Coefficient of variation (CV)	0.66842222
Kurtosis	-1.0117395
Mean	34.757526
Median Absolute Deviation (MAD)	15
Skewness	-0.41792648
Sum	5135355
Variance	539.75848
Monotonicity	Not monotonic



Value	Count	Frequency (%)
-1	36104	24.4%
35	21947	14.9%
45	19101	12.9%
55	14227	9.6%
30	12004	8.1%

Value

Value	Count	Frequency (%)
65	10764	7.3%
40	9296	6.3%
50	7569	5.1%
60	6986	4.7%
70	5408	3.7%
Other values (18)	4342	2.9%

Value

Value	Count	Frequency (%)
-1	36104	24.4%
0	8	< 0.1%
5	67	< 0.1%
6	2	< 0.1%
7	1	< 0.1%
10	190	0.1%
15	392	0.3%
20	320	0.2%
24	2	< 0.1%
25	2270	1.5%

Value

Value	Count	Frequency (%)
85	6	< 0.1%
80	176	0.1%
79	1	< 0.1%
75	893	0.6%
70	5408	3.7%
66	1	< 0.1%
65	10764	7.3%
60	6986	4.7%
55	14227	9.6%
50	7569	5.1%

road_constr_zone_fl

Boolean

IMBALANCE

Distinct	2
Distinct (%)	< 0.1%
Missing	2
Missing (%)	< 0.1%
Memory size	288.7 KIB

Value

Value	Count	Frequency (%)
False	139901	94.7%
True	7847	5.3%
(Missing)	2	< 0.1%

94.7%
(139901)

	False
	True

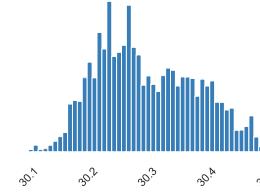
latitude

Real number (R)

MISSING

Distinct	96357
Distinct (%)	66.2%
Missing	2243
Missing (%)	1.5%
Infinite	0
Infinite (%)	0.0%
Mean	30.298798

Minimum	30.098737
Maximum	30.511625
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	1.1 MiB



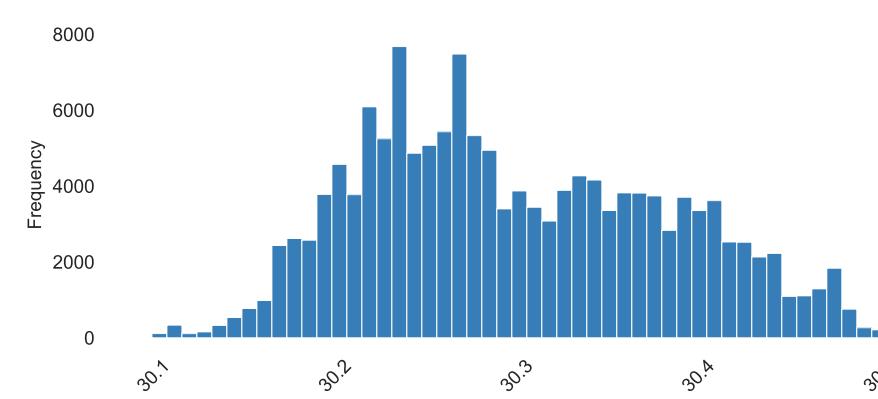
Quantile statistics

Minimum	30.098737
5-th percentile	30.177655
Q1	30.233272
median	30.284661
Q3	30.36443
95-th percentile	30.442425
Maximum	30.511625
Range	0.41288743
Interquartile range (IQR)	0.13115791

Descriptive statistics

Standard deviation	0.083391975
Coefficient of variation (CV)	0.0027523195
Kurtosis	-0.80132875
Mean	30.298798
Median Absolute Deviation (MAD)	0.06247375
Skewness	0.28570097
Sum	4408687.2
Variance	0.0069542214
Monotonicity	Not monotonic

Frequency



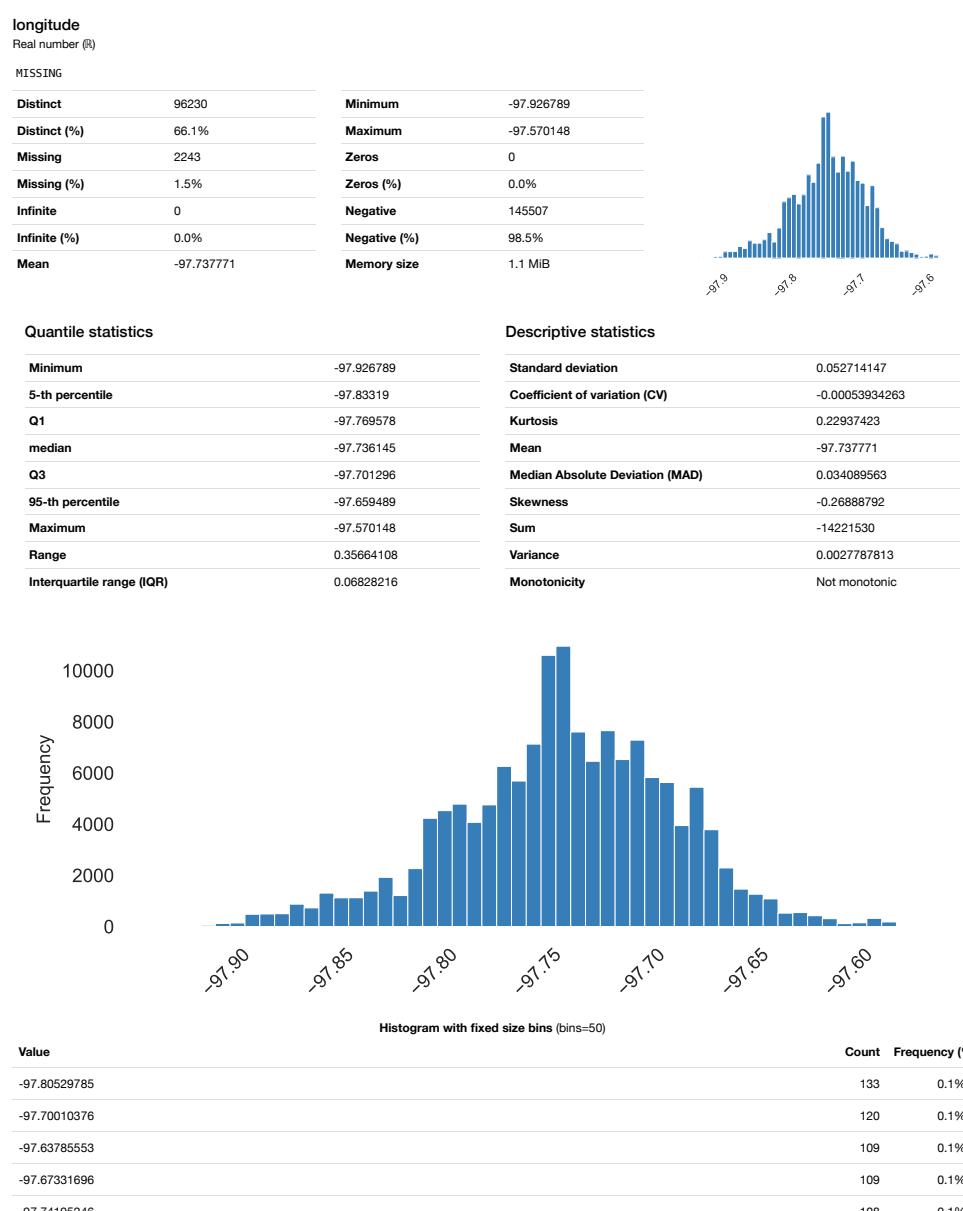
Value

Value	Count	Frequency (%)
30.24915123	133	0.1%
30.33917999	112	0.1%
30.20236015	109	0.1%
30.27753478	109	0.1%
30.32556534	108	0.1%

Value	Count	Frequency (%)
30.33882963	95	0.1%
30.34774163	91	0.1%
30.44802363	90	0.1%
30.23557091	88	0.1%
30.38954926	83	0.1%
Other values (96347)	144489	97.8%
(Missing)	2243	1.5%

Value	Count	Frequency (%)
30.0987373	1	< 0.1%
30.09889195	1	< 0.1%
30.0989809	1	< 0.1%
30.09899192	1	< 0.1%
30.09902058	1	< 0.1%
30.09906538	1	< 0.1%
30.09907731	1	< 0.1%
30.09908481	1	< 0.1%
30.09911847	1	< 0.1%
30.09919882	1	< 0.1%

Value	Count	Frequency (%)
30.51162473	1	< 0.1%
30.50985091	1	< 0.1%
30.50981473	1	< 0.1%
30.50938474	1	< 0.1%
30.50938473	1	< 0.1%
30.50775084	1	< 0.1%
30.50774627	1	< 0.1%
30.50707863	1	< 0.1%
30.50625252	1	< 0.1%
30.50616954	1	< 0.1%



Value	Count	Frequency (%)
-97.69958145	95	0.1%
-97.71247936	91	0.1%
-97.79304577	90	0.1%
-97.82463074	88	0.1%
-97.74504852	83	0.1%
Other values (96220)	144481	97.8%
(Missing)	2243	1.5%

Value	Count	Frequency (%)
-97.92678889	1	< 0.1%
-97.92658994	1	< 0.1%
-97.92493512	1	< 0.1%
-97.92453454	1	< 0.1%
-97.92448327	1	< 0.1%
-97.92442406	1	< 0.1%
-97.92431681	1	< 0.1%
-97.924013	1	< 0.1%
-97.92313723	1	< 0.1%
-97.92259045	1	< 0.1%

Value	Count	Frequency (%)
-97.57014781	1	< 0.1%
-97.57024774	1	< 0.1%
-97.5702853	1	< 0.1%
-97.57029177	1	< 0.1%
-97.57040391	1	< 0.1%
-97.57054233	1	< 0.1%
-97.57074071	1	< 0.1%
-97.57075081	1	< 0.1%
-97.57102709	1	< 0.1%
-97.57103645	1	< 0.1%

street_name

Text

Distinct	4630
Distinct (%)	3.1%
Missing	2
Missing (%)	< 0.1%
Memory size	1.1 MiB



Length	Characters and Unicode	Unique	Sample		
Max length	41				
Median length	40				
Mean length	9.2789682				
Min length	3				
	Total characters	1370949	Unique	2208	?
	Distinct characters		Unique (%)	1.5%	1st row
	Distinct categories	37			2nd row
	Distinct scripts	1	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?	3rd row
	Distinct blocks	1	(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?	4th row
	Distinct blocks	1	(https://en.wikipedia.org/wiki/Unicode_block)	?	5th row
					US0290
					US0290

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

Value	Count	Frequency (%)
ih0035	24841	8.9%
st	16185	5.8%
e	12954	4.6%
rd	12785	4.6%
dr	12668	4.5%
us0183	12105	4.3%
ln	11109	4.0%
w	10828	3.9%
blvd	10220	3.7%
sl0001	9023	3.2%
Other values (3316)	146218	52.4%



Austin Crash Report Data - Profiling Report

Value	Count	Frequency (%)
0	131188	9.6%
S	129302	9.4%
R	103032	7.5%
L	85099	6.2%
E	77592	5.7%
A	77389	5.6%
N	61412	4.5%
I	58169	4.2%
D	55131	4.0%
Other values (27)	534914	39.0%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	1370949	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
0	131188	9.6%
S	129302	9.4%
R	103032	7.5%
L	85099	6.2%
E	77592	5.7%
A	77389	5.6%
N	61412	4.5%
I	58169	4.2%
D	55131	4.0%
Other values (27)	534914	39.0%

Most occurring scripts

(unknown)

Value	Count	Frequency (%)
(unknown)	1370949	100.0%

Most frequent character per script

Austin Crash Report Data - Profiling Report

Value	Count	Frequency (%)
0	131188	9.6%
S	129302	9.4%
R	103032	7.5%
L	85099	6.2%
E	77592	5.7%
A	77389	5.6%
N	61412	4.5%
I	58169	4.2%
D	55131	4.0%
Other values (27)	534914	39.0%

Most occurring blocks

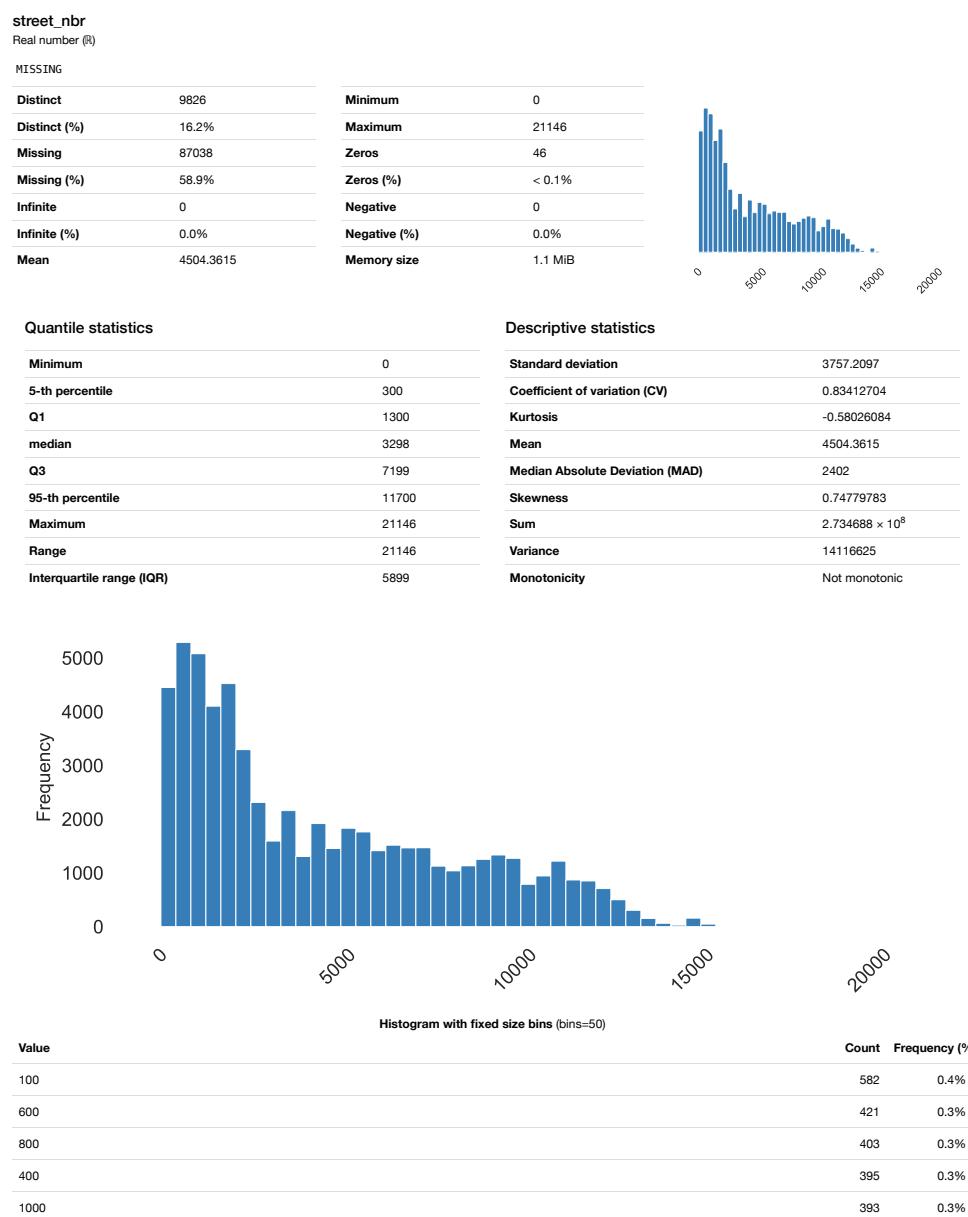
(unknown)

Value	Count	Frequency (%)
(unknown)	1370949	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
0	131188	9.6%
S	129302	9.4%
R	103032	7.5%
L	85099	6.2%
E	77592	5.7%
A	77389	5.6%
N	61412	4.5%
I	58169	4.2%
D	55131	4.0%
Other values (27)	534914	39.0%



Value	Count	Frequency (%)
900	392	0.3%
500	355	0.2%
1100	325	0.2%
1900	312	0.2%
300	299	0.2%
Other values (9816)	56835	38.5%
(Missing)	87038	58.9%

Value	Count	Frequency (%)
0	46	< 0.1%
1	5	< 0.1%
2	1	< 0.1%
4	20	< 0.1%
5	2	< 0.1%
7	1	< 0.1%
9	2	< 0.1%
10	2	< 0.1%
19	1	< 0.1%
21	1	< 0.1%

Value	Count	Frequency (%)
21146	1	< 0.1%
20032	1	< 0.1%
16799	1	< 0.1%
16600	1	< 0.1%
16501	1	< 0.1%
16400	1	< 0.1%
16300	2	< 0.1%
16200	1	< 0.1%
16157	3	< 0.1%
16126	1	< 0.1%

street_name_2	
Text	MISSING
<hr/>	
Distinct	3396
Distinct (%)	5.1%
Missing	81474
Missing (%)	55.1%
Memory size	1.1 MiB



Length	Characters and Unicode	Unique	Sample		
Max length	38	Total characters	Unique	1410	?
Median length	34		Unique (%)	2.1%	1st row
Mean length	10.161929	Distinct characters			VICTOR ST
Min length	3	Distinct categories (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)		?	3rd row SH0071
		Distinct scripts (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)		4th row	S 1ST ST
		Distinct blocks (https://en.wikipedia.org/wiki/Unicode_block)		5th row	COLLINFIELD DR

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

Value	Count	Frequency (%)
st	10325	7.7%
unknown	9867	7.3%
dr	8316	6.2%
ln	5983	4.4%
rd	5709	4.2%
e	5656	4.2%
w	4926	3.7%
blvd	4276	3.2%
not	2632	2.0%
reported	2632	2.0%
Other values (2635)	74436	55.2%



Most occurring characters

Value	Count	Frequency (%)
N	68482	10.2%
R	62626	9.3%
E	50876	7.6%
S	45809	6.8%
O	39042	5.8%
L	37239	5.5%
T	35991	5.3%
A	35302	5.2%
D	33381	5.0%
Other values (27)	33314	4.9%
	231430	34.4%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	673492	100.0%

Most frequent character per category

Most frequent character per category			
(unknown)	Value	Count	Frequency (%)
N		68482	10.2%
R		62626	9.3%
E		50876	7.6%
S		45809	6.8%
O		39042	5.8%
L		37239	5.5%
T		35991	5.3%
A		35302	5.2%
D		33381	5.0%
		33314	4.9%

Monte Carlo simulation

Value	Count	Frequency (%)
-------	-------	---------------

Most frequent character per script

(unknown)	Value	Count	Frequency (%)
N		68482	10.2%
R		62626	9.3%
E		50876	7.6%
S		45809	6.8%
O		39042	5.8%
L		37239	5.5%
T		35991	5.3%
A		35302	5.2%
D		33381	5.0%
Other values (27)		33314	4.9%
		231430	34.4%

Most occurring blocks

(unknown)	Value	Count	Frequency (%)
		673492	100.0%

Most frequent character per block

(unknown)	Value	Count	Frequency (%)
N		68482	10.2%
R		62626	9.3%
E		50876	7.6%
S		45809	6.8%
O		39042	5.8%
L		37239	5.5%
T		35991	5.3%
A		35302	5.2%
D		33381	5.0%
Other values (27)		33314	4.9%
		231430	34.4%

street_nbr_2	Unsupported
MISSING	REJECTED
Missing	147750
Missing (%)	100.0%
Memory size	1.1 MiB

crash_sev_idReal number (\mathbb{R})**ZEROS**

Distinct	8	Minimum	0
Distinct (%)	< 0.1%	Maximum	99
Missing	0	Zeros	4851
Missing (%)	0.0%	Zeros (%)	3.3%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	3.7065245	Memory size	1.1 MiB

**Quantile statistics**

Minimum	0
5-th percentile	1
Q1	2
median	5
Q3	5
95-th percentile	5
Maximum	99
Range	99
Interquartile range (IQR)	3

Descriptive statistics

Standard deviation	1.7536941
Coefficient of variation (CV)	0.47313705
Kurtosis	812.86388
Mean	3.7065245
Median Absolute Deviation (MAD)	0
Skewness	14.637743
Sum	547639
Variance	3.0754429
Monotonicity	Not monotonic

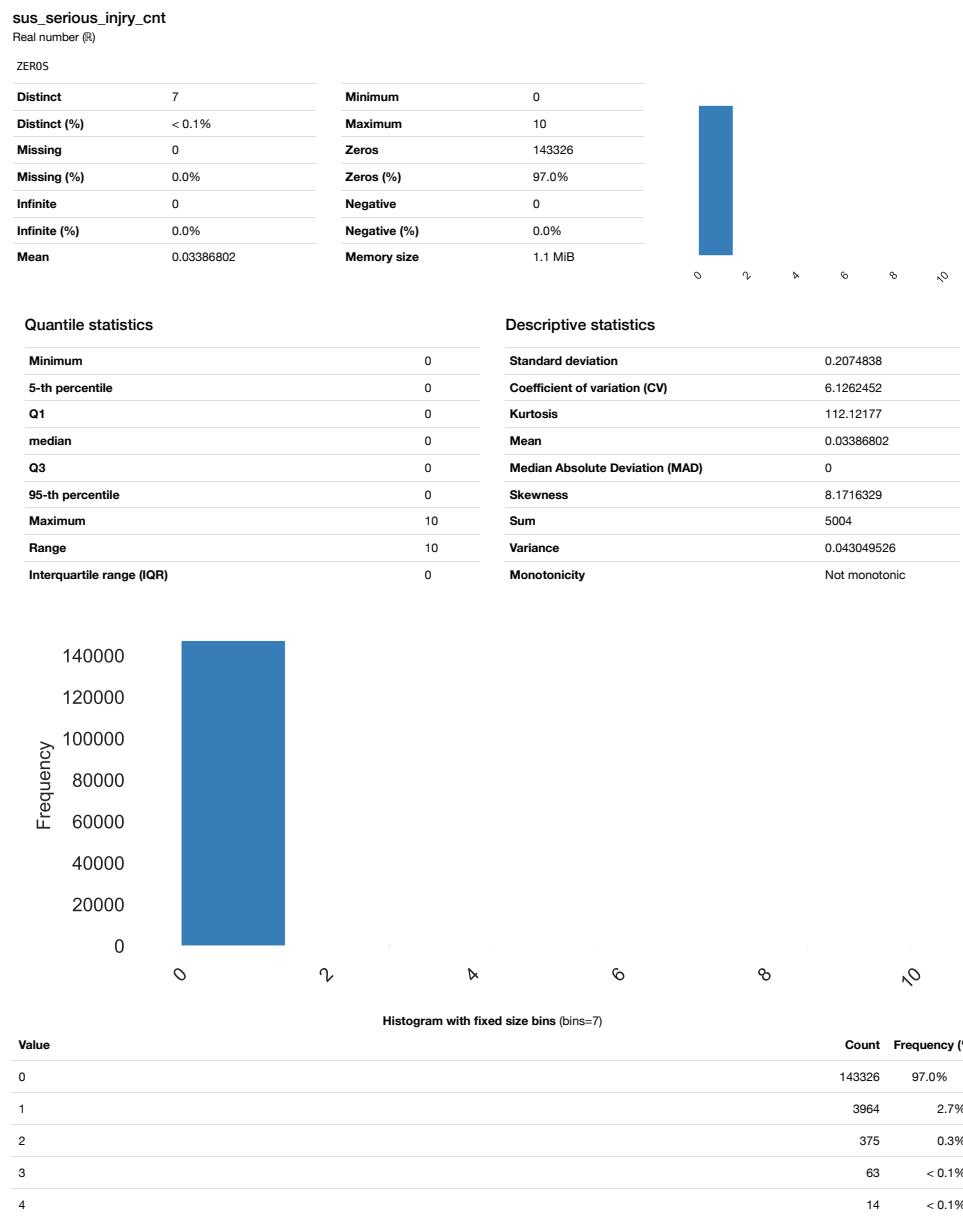


Histogram with fixed size bins (bins=8)		
Value	Count	Frequency (%)
5	77274	52.3%
3	31307	21.2%
2	29125	19.7%
0	4851	3.3%
1	4333	2.9%
4	846	0.6%
99	13	< 0.1%
94	1	< 0.1%
94	1	< 0.1%

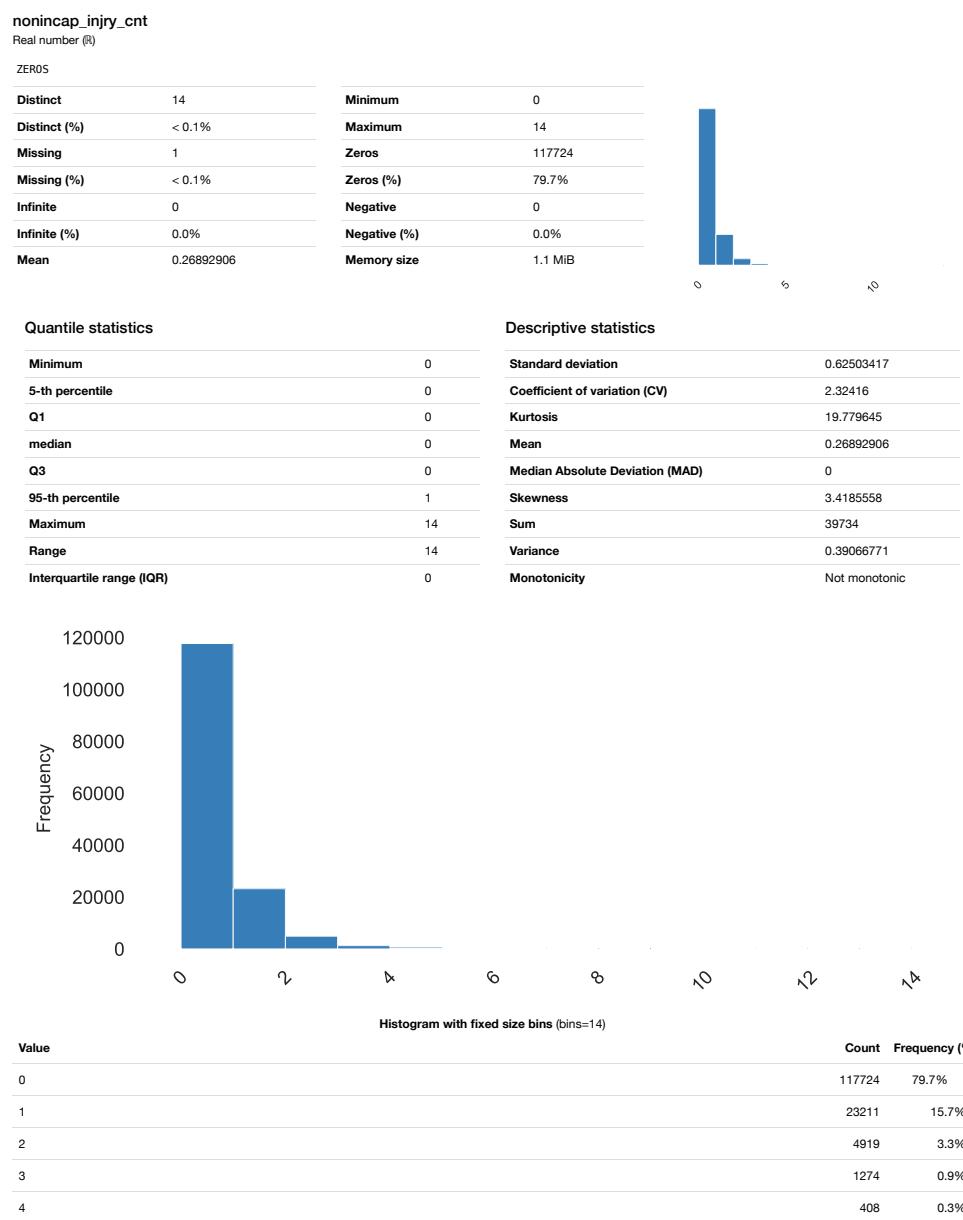
Value	Count	Frequency (%)
0	4851	3.3%
1	4333	2.9%
2	29125	19.7%
3	31307	21.2%
4	846	0.6%
5	77274	52.3%
94	1	< 0.1%
99	13	< 0.1%

Value	Count	Frequency (%)
99	13	< 0.1%
94	1	< 0.1%
5	77274	52.3%
4	846	0.6%
3	31307	21.2%

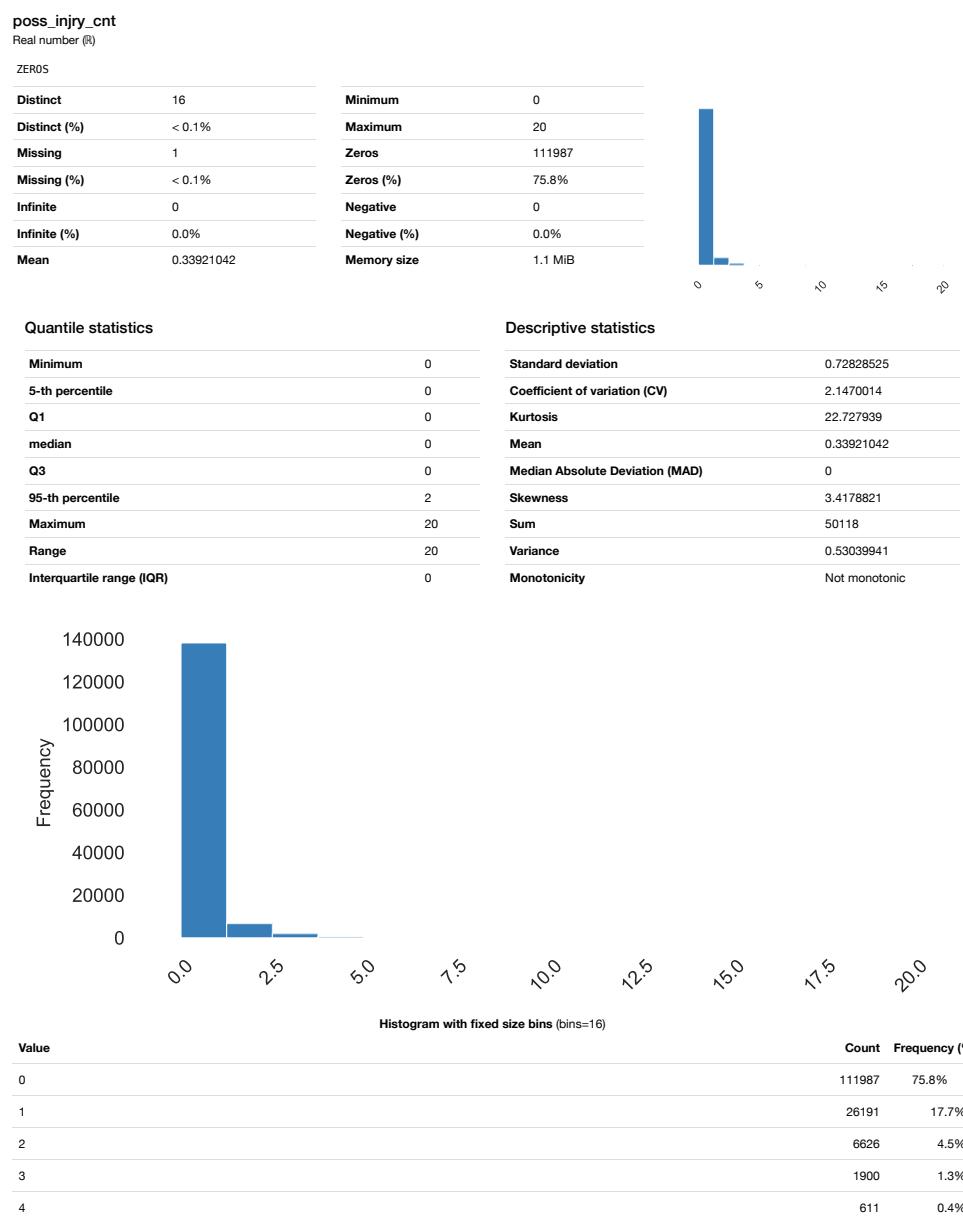
Value	Count	Frequency (%)
2	29125	19.7%
1	4333	2.9%
0	4851	3.3%



Value	Count	Frequency (%)
5	7	< 0.1%
10	1	< 0.1%
0	143326	97.0%
1	3964	2.7%
2	375	0.3%
3	63	< 0.1%
4	14	< 0.1%
5	7	< 0.1%
10	1	< 0.1%
0	143326	97.0%
1	3964	2.7%
2	375	0.3%
3	63	< 0.1%
4	14	< 0.1%
5	7	< 0.1%
10	1	< 0.1%
0	143326	97.0%



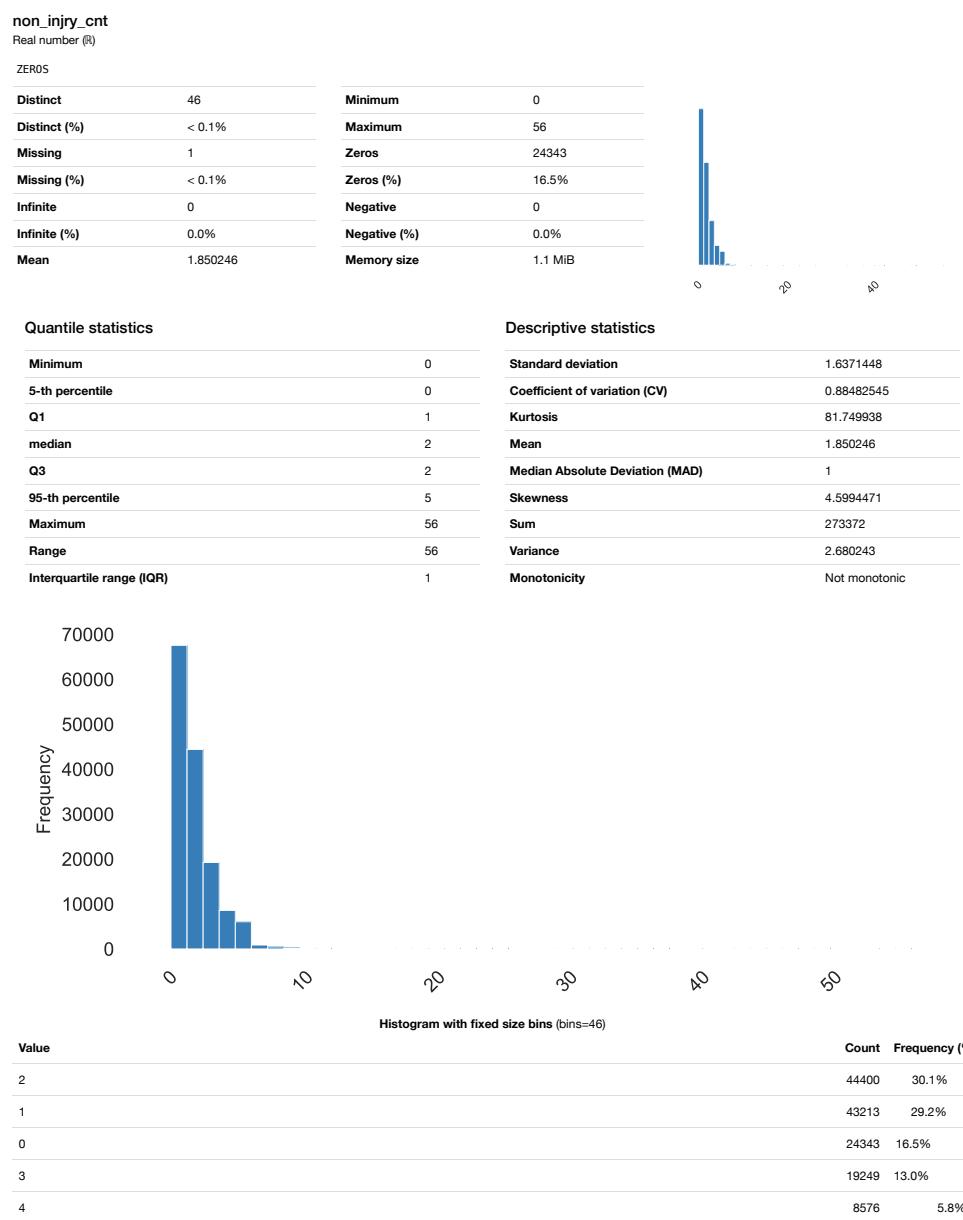
Value	Count	Frequency (%)
5	134	0.1%
6	38	< 0.1%
7	22	< 0.1%
9	8	< 0.1%
8	5	< 0.1%
Other values (4)	6	< 0.1%
0	117724	79.7%
1	23211	15.7%
2	4919	3.3%
3	1274	0.9%
4	408	0.3%
5	134	0.1%
6	38	< 0.1%
7	22	< 0.1%
8	5	< 0.1%
9	8	< 0.1%
14	1	< 0.1%
12	1	< 0.1%
11	1	< 0.1%
10	3	< 0.1%
9	8	< 0.1%
8	5	< 0.1%
7	22	< 0.1%
6	38	< 0.1%
5	134	0.1%
4	408	0.3%



Value	Count	Frequency (%)
5	250	0.2%
6	106	0.1%
7	38	< 0.1%
8	20	< 0.1%
9	9	< 0.1%
Other values (6)	11	< 0.1%

Value	Count	Frequency (%)
0	111987	75.8%
1	26191	17.7%
2	6626	4.5%
3	1900	1.3%
4	611	0.4%
5	250	0.2%
6	106	0.1%
7	38	< 0.1%
8	20	< 0.1%
9	9	< 0.1%

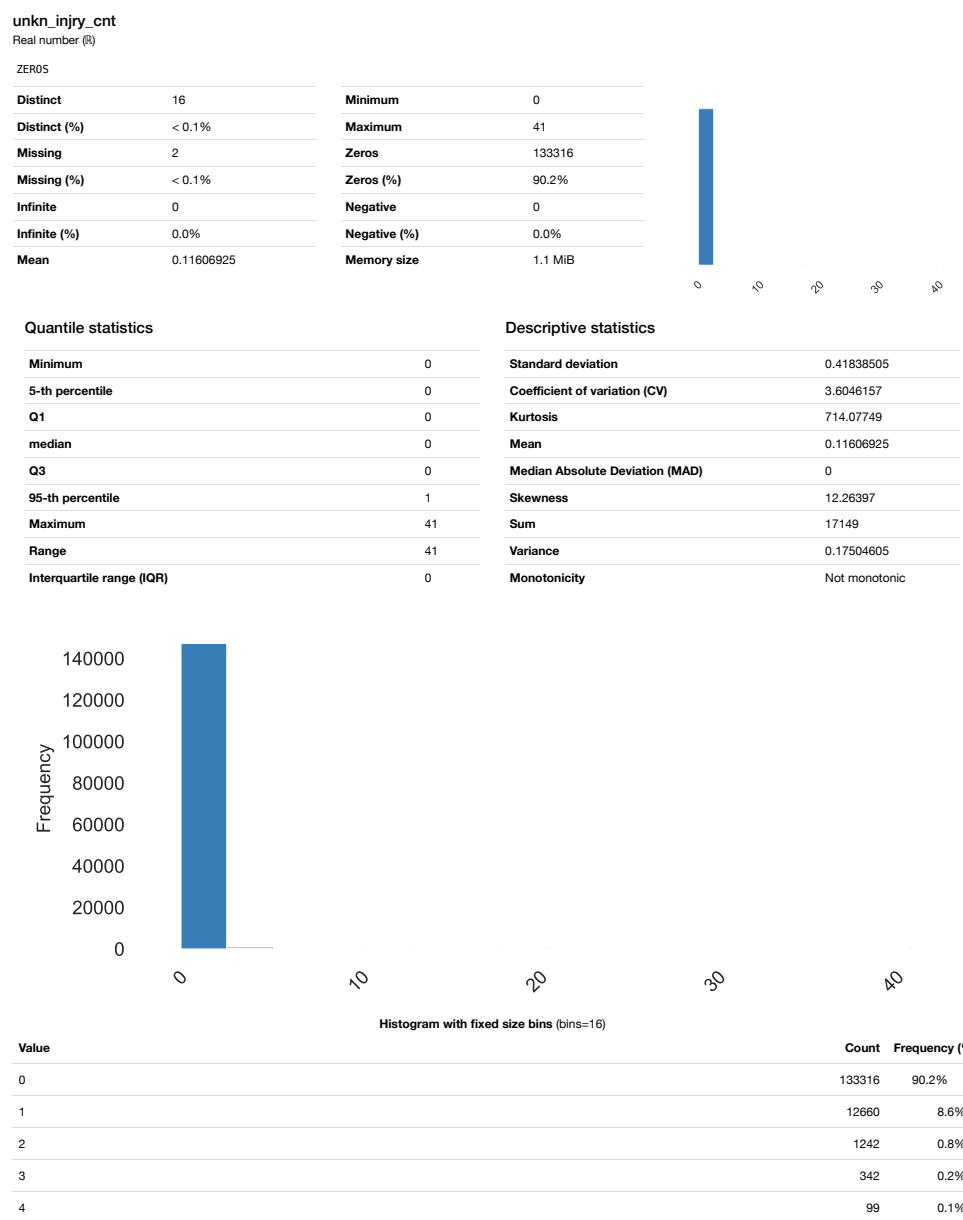
Value	Count	Frequency (%)
20	1	< 0.1%
18	1	< 0.1%
15	1	< 0.1%
12	2	< 0.1%
11	1	< 0.1%
10	5	< 0.1%
9	9	< 0.1%
8	20	< 0.1%
7	38	< 0.1%
6	106	0.1%



Value	Count	Frequency (%)
5	4199	2.8%
6	1853	1.3%
7	875	0.6%
8	467	0.3%
9	225	0.2%
Other values (36)	349	0.2%

Value	Count	Frequency (%)
0	24343	16.5%
1	43213	29.2%
2	44400	30.1%
3	19249	13.0%
4	8576	5.8%
5	4199	2.8%
6	1853	1.3%
7	875	0.6%
8	467	0.3%
9	225	0.2%

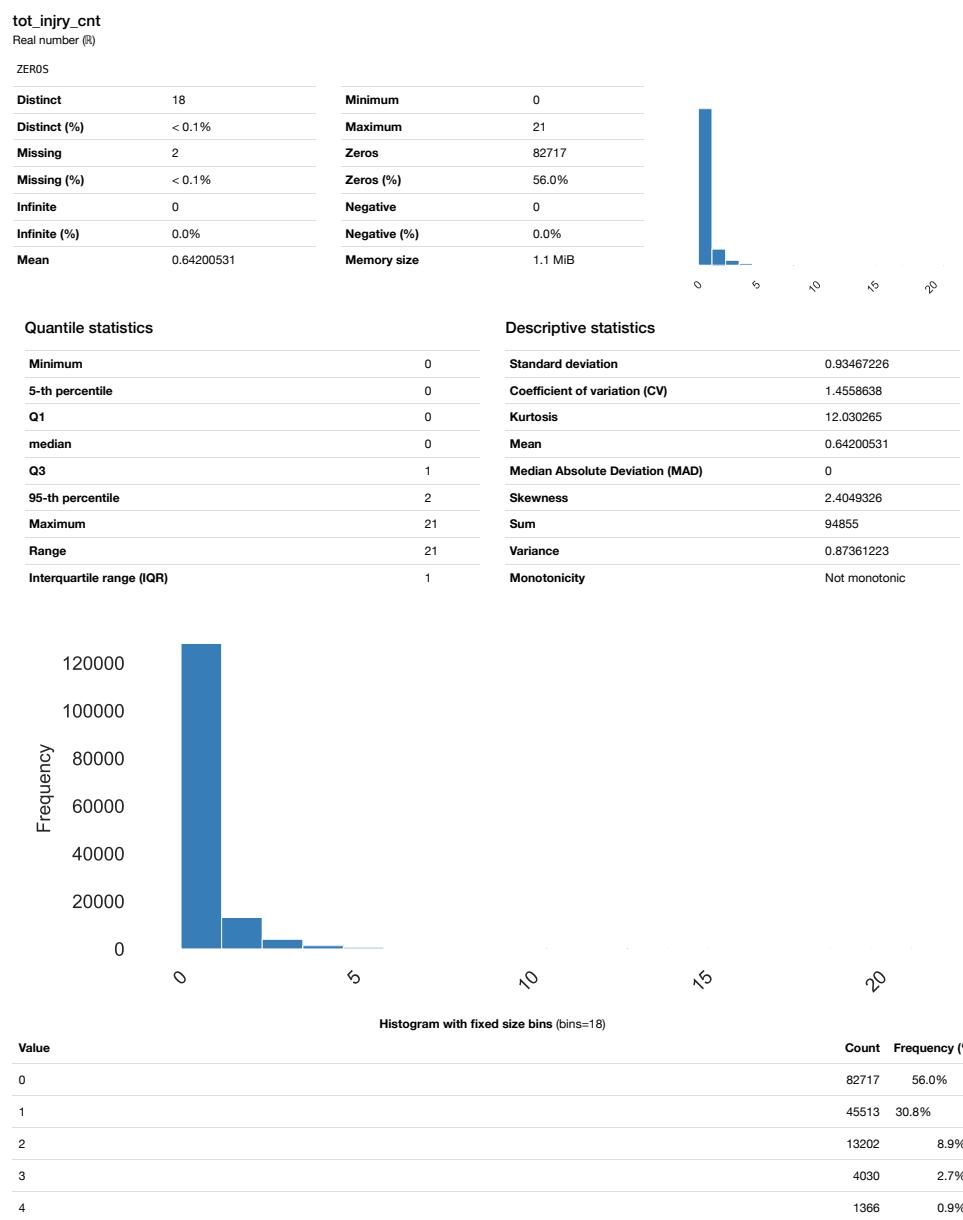
Value	Count	Frequency (%)
56	1	< 0.1%
53	1	< 0.1%
50	1	< 0.1%
48	2	< 0.1%
46	2	< 0.1%
44	1	< 0.1%
43	3	< 0.1%
42	1	< 0.1%
39	1	< 0.1%
38	1	< 0.1%



Value	Count	Frequency (%)
5	55	< 0.1%
6	14	< 0.1%
7	8	< 0.1%
8	4	< 0.1%
14	2	< 0.1%
Other values (6)	6	< 0.1%
(Missing)	2	< 0.1%

Value	Count	Frequency (%)
0	133316	90.2%
1	12660	8.6%
2	1242	0.8%
3	342	0.2%
4	99	0.1%
5	55	< 0.1%
6	14	< 0.1%
7	8	< 0.1%
8	4	< 0.1%
9	1	< 0.1%

Value	Count	Frequency (%)
41	1	< 0.1%
18	1	< 0.1%
17	1	< 0.1%
14	2	< 0.1%
13	1	< 0.1%
10	1	< 0.1%
9	1	< 0.1%
8	4	< 0.1%
7	8	< 0.1%
6	14	< 0.1%

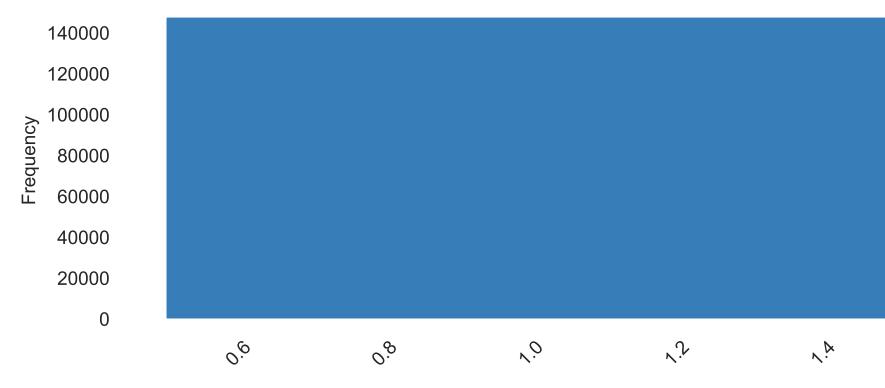


Value	Count	Frequency (%)
5	526	0.4%
6	213	0.1%
7	89	0.1%
8	49	< 0.1%
9	18	< 0.1%
Other values (8)	25	< 0.1%

Value	Count	Frequency (%)
0	82717	56.0%
1	45513	30.8%
2	13202	8.9%
3	4030	2.7%
4	1366	0.9%
5	526	0.4%
6	213	0.1%
7	89	0.1%
8	49	< 0.1%
9	18	< 0.1%

Value	Count	Frequency (%)
21	1	< 0.1%
18	1	< 0.1%
15	1	< 0.1%
14	1	< 0.1%
13	2	< 0.1%
12	6	< 0.1%
11	3	< 0.1%
10	10	< 0.1%
9	18	< 0.1%
8	49	< 0.1%

death_cnt						
Categorical						
IMBALANCE						
Distinct 5						
Distinct (%) < 0.1%						
Missing 0						
Missing (%) 0.0%						
Memory size 1.1 MiB						
Length						
Characters and Unicode						
Unique						
Sample						
Total characters	147750	Unique	0	?	1st row	0
Distinct characters	5	Unique (%)	0.0%		2nd row	0
Min length	1				3rd row	0
Distinct categories	1			?	4th row	0
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.						
Common Values						
Value		Count		Frequency (%)		
0		146905		99.4%		
1		808		0.5%		
2		32		< 0.1%		
3		3		< 0.1%		
4		2		< 0.1%		



Common Values (Plot)



Value	Count	Frequency (%)
0	146905	99.4%
1	808	0.5%
2	32	< 0.1%
3	3	< 0.1%
4	2	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
0	146905	99.4%
1	808	0.5%
2	32	< 0.1%
3	3	< 0.1%
4	2	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per category

Value	Count	Frequency (%)
0	146905	99.4%
1	808	0.5%
2	32	< 0.1%
3	3	< 0.1%
4	2	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per script

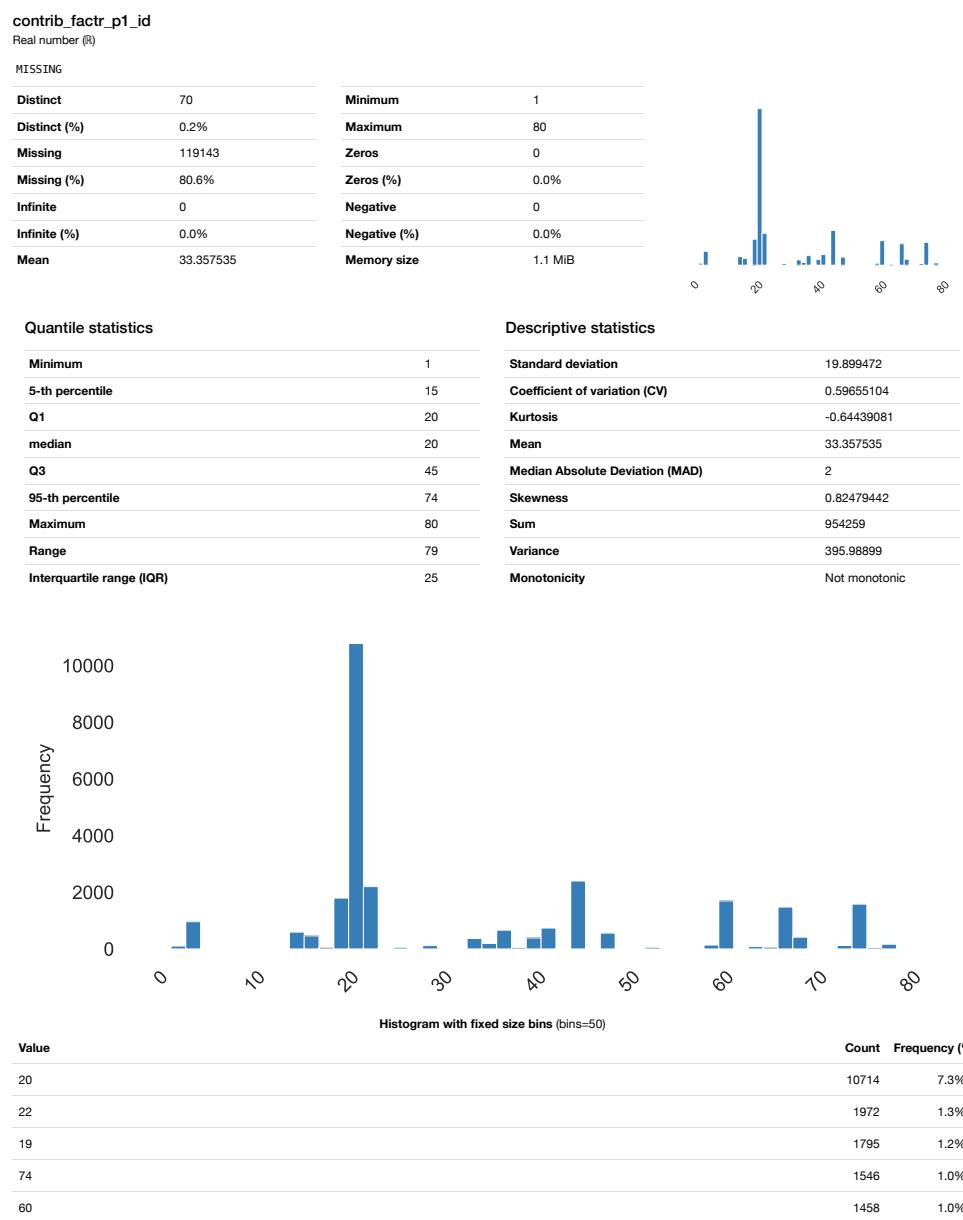
Value	Count	Frequency (%)
0	146905	99.4%
1	808	0.5%
2	32	< 0.1%
3	3	< 0.1%
4	2	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per block

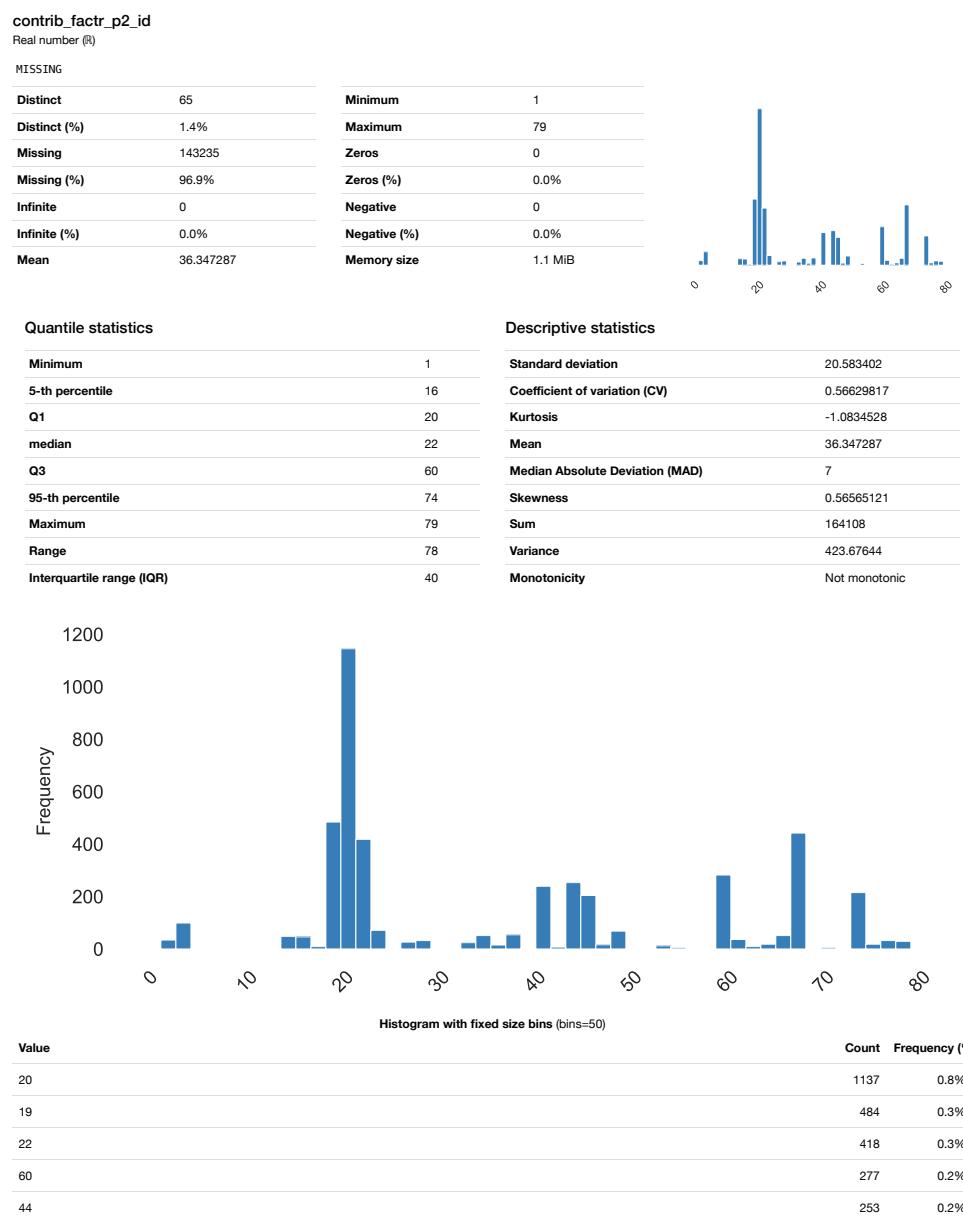
Value	Count	Frequency (%)
0	146905	99.4%
1	808	0.5%
2	32	< 0.1%
3	3	< 0.1%
4	2	< 0.1%



Value	Count	Frequency (%)
44	1424	1.0%
67	1282	0.9%
45	974	0.7%
4	907	0.6%
41	734	0.5%
Other values (60)	5801	3.9%
(Missing)	119143	80.6%

Value	Count	Frequency (%)
1	58	< 0.1%
2	44	< 0.1%
3	54	< 0.1%
4	907	0.6%
14	47	< 0.1%
15	547	0.4%
16	461	0.3%
17	54	< 0.1%
18	7	< 0.1%
19	1795	1.2%

Value	Count	Frequency (%)
80	1	< 0.1%
79	4	< 0.1%
78	89	0.1%
77	76	0.1%
76	52	< 0.1%
75	34	< 0.1%
74	1546	1.0%
73	122	0.1%
72	17	< 0.1%
71	14	< 0.1%



Value	Count	Frequency (%)
67	236	0.2%
68	206	0.1%
74	196	0.1%
45	194	0.1%
41	156	0.1%
Other values (55)	958	0.6%
(Missing)	143235	96.9%

Value	Count	Frequency (%)
1	17	< 0.1%
2	17	< 0.1%
3	5	< 0.1%
4	94	0.1%
14	2	< 0.1%
15	46	< 0.1%
16	46	< 0.1%
17	7	< 0.1%
18	2	< 0.1%
19	484	0.3%

Value	Count	Frequency (%)
79	1	< 0.1%
78	28	< 0.1%
77	19	< 0.1%
76	13	< 0.1%
75	18	< 0.1%
74	196	0.1%
73	19	< 0.1%
72	3	< 0.1%
71	5	< 0.1%
69	1	< 0.1%

units_involved

Text

Distinct	1112
Distinct (%)	0.8%
Missing	7
Missing (%)	< 0.1%
Memory size	1.1 MiB

vehicle
large
motor
car
passenger

motorcycle
bicycle
other
other/unknown
pedestrian
e-scooter

Length

	Characters and Unicode	Unique		Sample
Total characters	5545335	Unique	593	?
Distinct characters	33	Unique (%)	0.4%	1st row
Distinct categories	1			Passenger car
Distinct scripts	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?		2nd row
Distinct blocks	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?		3rd row
				Large passenger vehicle
				4th row
				Motor vehicle – other
				5th row
				Passenger car

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

Value

Value	Count	Frequency (%)
passenger	278045	31.6%
car	172791	19.6%
	157282	17.9%
vehicle	116957	13.3%
large	105254	12.0%
motor	11703	1.3%
-	11703	1.3%
other	11703	1.3%
other/unknown	5061	0.6%
pedestrian	3705	0.4%
Other values (6)	6556	0.7%

vehicle
large
motor
car
passenger

motorcycle
bicycle
other
other/unknown
pedestrian
e-scooter

Most occurring characters**Value**

Value	Count	Frequency (%)
e	925973	16.7%
	733017	13.2%
r	592303	10.7%
s	560112	10.1%
a	559810	10.1%
g	383299	6.9%
c	302423	5.5%
n	296948	5.4%
P	176496	3.2%
&	157282	2.8%
Other values (23)	857672	15.5%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	5545335	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
e	925973	16.7%
	733017	13.2%
r	592303	10.7%
s	560112	10.1%
a	559810	10.1%
g	383299	6.9%
c	302423	5.5%
n	296948	5.4%
P	176496	3.2%
&	157282	2.8%
Other values (23)	857672	15.5%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	5545335	100.0%

Most frequent character per script

(unknown)	Value	Count	Frequency (%)
e		925973	16.7%
r		733017	13.2%
s		592303	10.7%
a		560112	10.1%
g		559810	10.1%
c		383299	6.9%
n		302423	5.5%
P		296948	5.4%
&		176496	3.2%
Other values (23)		157282	2.8%
		857672	15.5%

Most occurring blocks

(unknown)	Value	Count	Frequency (%)
		5545335	100.0%

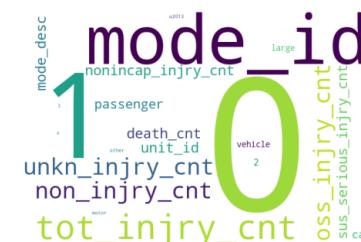
Most frequent character per block

(unknown)	Value	Count	Frequency (%)
e		925973	16.7%
r		733017	13.2%
s		592303	10.7%
a		560112	10.1%
g		559810	10.1%
c		383299	6.9%
n		302423	5.5%
P		296948	5.4%
&		176496	3.2%
Other values (23)		157282	2.8%
		857672	15.5%

atd_mode_category_metadata

Text

Distinct	147743
Distinct (%)	100.0%
Missing	7
Missing (%)	< 0.1%
Memory size	1.1 MiB

**Length**

Max length	3936
Median length	3325
Mean length	455.91029
Min length	214

Characters and Unicode

Total characters	67357554
Unique	147743
Unique (%)	100.0%
Distinct characters	52
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

147743

?

100.0%

?

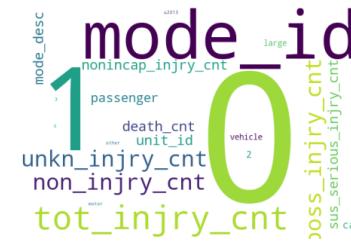
Sample

1st	[{"mode_id": 1, "mode_desc": "Passenger car", "unit_id": 2259431, "death_cnt": 0, "sus_serious_injry_cnt": 0, "nonincap_injry_cnt": 0, "poss_injry_cnt": 0, "non_injry_cnt": 4, "unkn_injry_cnt": 0, "tot_injry_cnt": 0}]
2nd	[{"mode_id": 2, "mode_desc": "Large passenger vehicle", "unit_id": 2260826, "death_cnt": 0, "sus_serious_injry_cnt": 0, "nonincap_injry_cnt": 0, "poss_injry_cnt": 0, "non_injry_cnt": 1, "unkn_injry_cnt": 0, "tot_injry_cnt": 0}]
3rd	[{"mode_id": 2, "mode_desc": "Large passenger vehicle", "unit_id": 2260877, "death_cnt": 0, "sus_serious_injry_cnt": 0, "nonincap_injry_cnt": 0, "poss_injry_cnt": 0, "non_injry_cnt": 1, "unkn_injry_cnt": 0, "tot_injry_cnt": 0}]
4th	[{"mode_id": 4, "mode_desc": "Motor vehicle u2013 other", "unit_id": 2262706, "death_cnt": 0, "sus_serious_injry_cnt": 0, "nonincap_injry_cnt": 0, "poss_injry_cnt": 0, "non_injry_cnt": 1, "unkn_injry_cnt": 0, "tot_injry_cnt": 0}]

```
5th [{"mode_id": 1, "mode_desc": "Passenger car", "unit_id": 2262540, "death_cnt": 0, "sus_serious_injy_cnt": 0, "nonincap_injy_cnt": 0, "poss_injy_cnt": 0, "non_injy_cnt": 1, "unkn_injy_cnt": 0, "tot_injy_cnt": 0}]

row
```

Value	Count	Frequency (%)
0	1750115	26.8%
1	493130	7.6%
mode_id	305025	4.7%
tot_injy_cnt	305025	4.7%
unkn_injy_cnt	305025	4.7%
non_injy_cnt	305025	4.7%
poss_injy_cnt	305025	4.7%
nonincap_injy_cnt	305025	4.7%
sus_serious_injy_cnt	305025	4.7%
death_cnt	305025	4.7%
Other values (305087)	1835508	28.2%

**Most occurring characters**

Value	Count	Frequency (%)
"	6710550	10.0%
n	6702473	10.0%
-	6371210	9.5%
i	5185425	7.7%
t	3478602	5.2%
:	3391473	5.0%
c	3050250	4.5%
,	3047648	4.5%
r	2902507	4.3%
Other values (42)	2727478	4.0%
	23789938	35.3%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	67357554	100.0%

Most frequent character per category

(*unknown*)

Value	Count	Frequency (%)
"	6710550	10.0%
n	6702473	10.0%
-	6371210	9.5%
i	5185425	7.7%
t	3478602	5.2%
:	3391473	5.0%
c	3050250	4.5%
,	3047648	4.5%
r	2902507	4.3%
Other values (42)	2727478	4.0%
	23789938	35.3%

Most occurring scripts

Value	Count	Frequency (%)
(<i>unknown</i>)	67357554	100.0%

Most frequent character per script

(*unknown*)

Value	Count	Frequency (%)
"	6710550	10.0%
n	6702473	10.0%
-	6371210	9.5%
i	5185425	7.7%
t	3478602	5.2%
:	3391473	5.0%
c	3050250	4.5%
,	3047648	4.5%
r	2902507	4.3%
Other values (42)	2727478	4.0%
	23789938	35.3%

Most occurring blocks

Value	Count	Frequency (%)
(<i>unknown</i>)	67357554	100.0%

Most frequent character per block*(unknown)*

Value	Count	Frequency (%)
*	6710550	10.0%
n	6702473	10.0%
-	6371210	9.5%
i	5185425	7.7%
t	3478602	5.2%
:	3391473	5.0%
c	3050250	4.5%
,	3047648	4.5%
r	2902507	4.3%
Other values (42)	2727478	4.0%
	23789938	35.3%

pedestrian_fl

Boolean

CONSTANT MISSING

Distinct	1
Distinct (%)	< 0.1%
Missing	144245
Missing (%)	97.6%
Memory size	288.7 KIB

Value

Value	Count	Frequency (%)
True	3505	2.4%
(Missing)	144245	97.6%



100.0%
(3505)



True

motor_vehicle_f1

Boolean

CONSTANT

Distinct	1
Distinct (%)	< 0.1%
Missing	1116
Missing (%)	0.8%

Memory size 288.7 KIB

Value

Value	Count	Frequency (%)
True	146634	99.2%
(Missing)	1116	0.8%

**100.0%
(146634)** True**motorcycle_f1**

Boolean

CONSTANT MISSING

Distinct	1
Distinct (%)	< 0.1%
Missing	144148
Missing (%)	97.6%

Memory size 288.7 KIB

Value

Value	Count	Frequency (%)
True	3602	2.4%
(Missing)	144148	97.6%

**100.0%
(3602)** True

bicycle_fl

Boolean

CONSTANT MISSING

Distinct	1
Distinct (%)	< 0.1%
Missing	145306
Missing (%)	98.3%

Memory size 288.7 KiB

Value

Value	Count	Frequency (%)
True	2444	1.7%
(Missing)	145306	98.3%

**100.0%
(2444)** True**other_fl**

Boolean

CONSTANT MISSING

Distinct	1
Distinct (%)	< 0.1%
Missing	142905
Missing (%)	96.7%

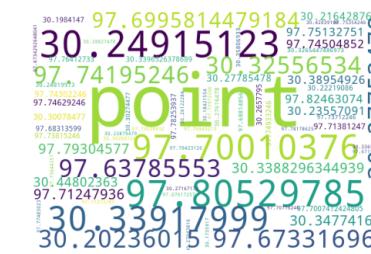
Memory size 288.7 KiB

Value

Value	Count	Frequency (%)
True	4845	3.3%
(Missing)	142905	96.7%

**100.0%
(4845)** True

point	
Text	
MISSING	
Distinct	97739
Distinct (%)	67.2%
Missing	2243
Missing (%)	1.5%
Memory size	1.1 MiB



Length	
Max length	45
Median length	32
Mean length	34.597174
Min length	22

Characters and Unicode		Unique			Sample	
Total characters	5034131	Unique	85923	?	1st row	POINT (-97.72445 30.404)
		Unique (%)	59.1%		2nd row	POINT (-97.78855804 30.43798421)
Distinct characters	20				3rd row	POINT (-97.662680859520 30.3276207970184)
					4th row	POINT (-97.76719944 30.22515795)
Distinct categories (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	1				5th row	POINT (-97.78425218 30.16984339)
					6th row	POINT (-97.72445 30.404)
Distinct scripts (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	1				7th row	POINT (-97.78425218 30.16984339)
					8th row	POINT (-97.72445 30.404)
Distinct blocks (https://en.wikipedia.org/wiki/Unicode_block)	1				9th row	POINT (-97.78425218 30.16984339)
					10th row	POINT (-97.72445 30.404)

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

Value	Count	Frequency (%)
point	145507	33.3%
97.80529785	133	< 0.1%
30.24915123	133	< 0.1%
97.70010376	120	< 0.1%
30.33917999	112	< 0.1%
97.63785553	109	< 0.1%
30.20236015	109	< 0.1%
97.67331696	109	< 0.1%
30.27753478	109	< 0.1%
30.32556534	108	< 0.1%
Other values (192616)	289972	66.4%



Most occurring characters

Value	Count	Frequency (%)
7	509037	10.1%
3	428304	8.5%
9	375986	7.5%
0	347207	6.9%
2	316857	6.3%
4	308324	6.1%
.	291014	5.8%
	291014	5.8%
6	277678	5.5%
8	253877	5.0%
Other values (10)	1634833	32.5%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	5034131	100.0%

Most frequent character per category

Most frequent character per category		
(unknown)	Count	Frequency (%)
7	509037	10.1%
3	428304	8.5%
9	375986	7.5%
0	347207	6.9%
2	316857	6.3%
4	308324	6.1%
.	291014	5.8%
	291014	5.8%
6	277678	5.5%
8	253877	5.0%

Montage

Value	Count	Frequency (%)
-------	-------	---------------

Most frequent character per script

(unknown)		Count	Frequency (%)
Value			
7		509037	10.1%
3		428304	8.5%
9		375986	7.5%
0		347207	6.9%
2		316857	6.3%
4		308324	6.1%
.		291014	5.8%
		291014	5.8%
6		277678	5.5%
8		253877	5.0%
Other values (10)		1634833	32.5%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	5034131	100.0%

Most frequent character per block

(unknown)	Count	Frequency (%)
Value		
7	509037	10.1%
3	428304	8.5%
9	375986	7.5%
0	347207	6.9%
2	316857	6.3%
4	308324	6.1%
.	291014	5.8%
	291014	5.8%
6	277678	5.5%
8	253877	5.0%
Other values (10)	1634833	32.5%

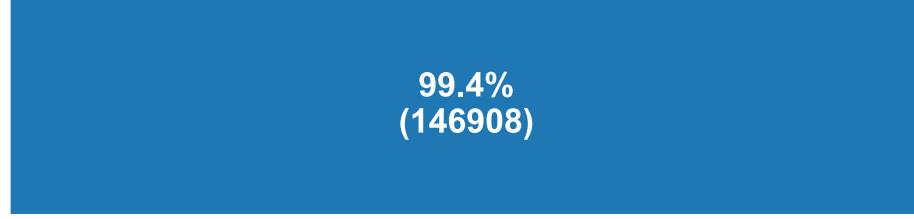
apd_confirmed_fatality
Boolean

IMBALANCE

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%

Memory size 144.4 KiB

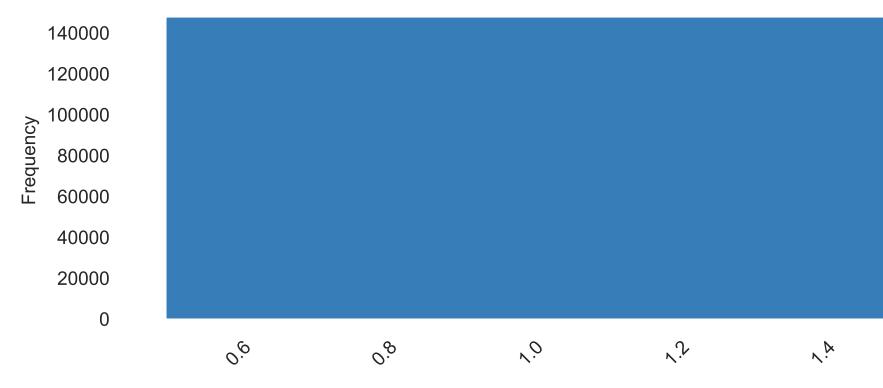
Value	Count	Frequency (%)
False	146908	99.4%
True	842	0.6%



99.4%
(146908)



apd_confirmed_death_count						
Categorical						
IMBALANCE						
Distinct 5						
Distinct (%) < 0.1%						
Missing 0						
Missing (%) 0.0%						
Memory size 1.1 MiB						
Length		Characters and Unicode	Unique	Sample		
Max length	1	Total characters 147750	Unique 0	?	1st row	0
Median length	1		Unique (%) 0.0%		2nd row	0
Mean length	1	Distinct characters 5			3rd row	0
Min length	1			?	4th row	0
		Distinct categories 1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)		?	5th row	0
		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.						
Common Values						
Value				Count	Frequency (%)	
0				146908	99.4%	
1				805	0.5%	
2				32	< 0.1%	
3				3	< 0.1%	
4				2	< 0.1%	



Common Values (Plot)



Value	Count	Frequency (%)
0	146908	99.4%
1	805	0.5%
2	32	< 0.1%
3	3	< 0.1%
4	2	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
0	146908	99.4%
1	805	0.5%
2	32	< 0.1%
3	3	< 0.1%
4	2	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per category

Value	Count	Frequency (%)
0	146908	99.4%
1	805	0.5%
2	32	< 0.1%
3	3	< 0.1%
4	2	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per script

Value	Count	Frequency (%)
0	146908	99.4%
1	805	0.5%
2	32	< 0.1%
3	3	< 0.1%
4	2	< 0.1%

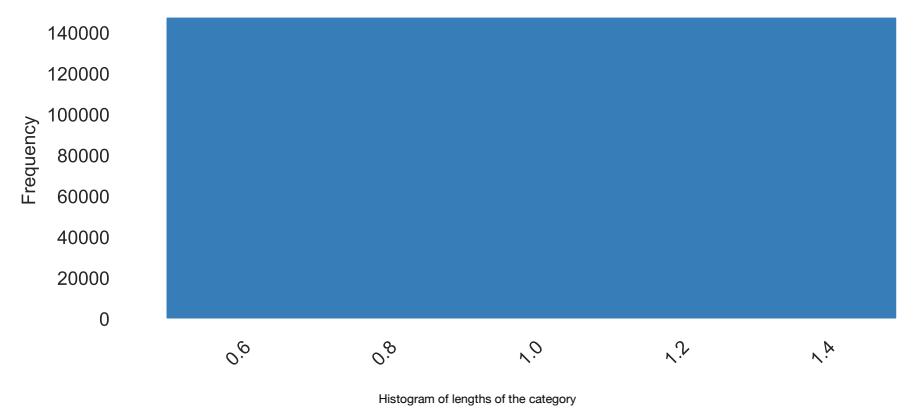
Most occurring blocks

Value	Count	Frequency (%)
(unknown)	147750	100.0%

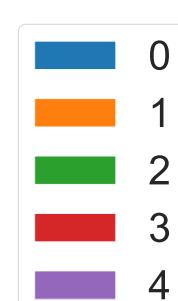
Most frequent character per block

Value	Count	Frequency (%)
0	146908	99.4%
1	805	0.5%
2	32	< 0.1%
3	3	< 0.1%
4	2	< 0.1%

motor_vehicle_death_count								
Categorical								
IMBALANCE								
Distinct 5								
Distinct (%) < 0.1%								
Missing 0								
Missing (%) 0.0%								
Memory size 1.1 MiB								
Length		Characters and Unicode		Unique		Sample		
Max length	1	Total characters	147750	Unique	0	?	1st row	0
Median length	1			Unique (%)	0.0%		2nd row	0
Mean length	1	Distinct characters	5				3rd row	0
Min length	1	Distinct categories	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?	1	4th row	0	
		Distinct scripts	(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?	1	5th row	0	
		Distinct blocks	(https://en.wikipedia.org/wiki/Unicode_block)	?	1			
The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.								
Common Values								
Value								
0								
Count 147372								
Frequency (%) 99.7%								
1								
Count 345								
Frequency (%) 0.2%								
2								
Count 28								
< 0.1%								
3								
Count 3								
< 0.1%								
4								
Count 2								
< 0.1%								
Length								



Common Values (Plot)



Value	Count	Frequency (%)
0	147372	99.7%
1	345	0.2%
2	28	< 0.1%
3	3	< 0.1%
4	2	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
0	147372	99.7%
1	345	0.2%
2	28	< 0.1%
3	3	< 0.1%
4	2	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per category

Value	Count	Frequency (%)
0	147372	99.7%
1	345	0.2%
2	28	< 0.1%
3	3	< 0.1%
4	2	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per script

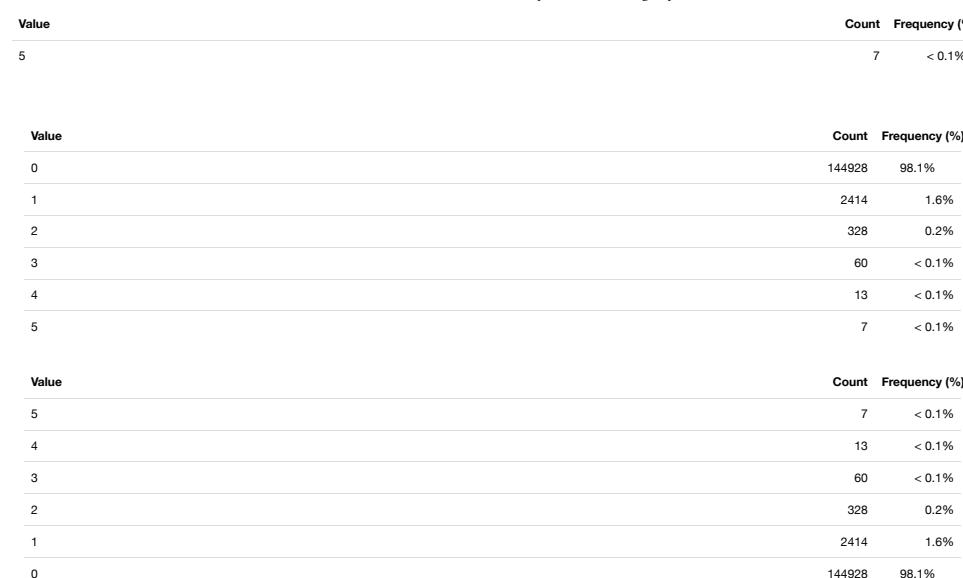
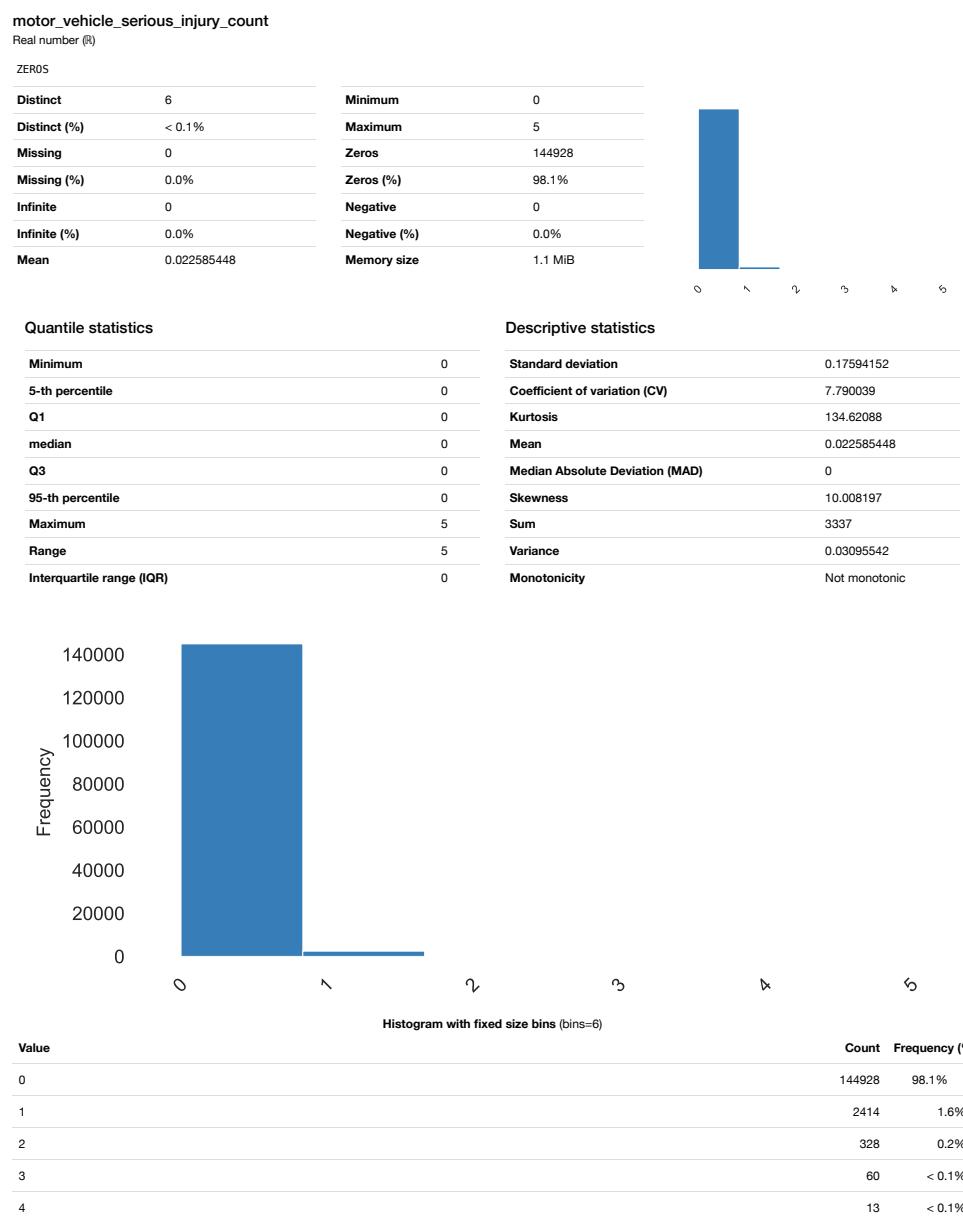
Value	Count	Frequency (%)
0	147372	99.7%
1	345	0.2%
2	28	< 0.1%
3	3	< 0.1%
4	2	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per block

Value	Count	Frequency (%)
0	147372	99.7%
1	345	0.2%
2	28	< 0.1%
3	3	< 0.1%
4	2	< 0.1%



bicycle_death_count
Categorical

IMBALANCE

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.1 MiB

Length

	Characters and Unicode	Unique	Sample
Total characters	147750	Unique 0 ? 1st row 0	
Distinct characters	2	Unique (%) 0.0% 2nd row 0	
Min length	1	3rd row 0	
Distinct categories	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	? 4th row 0	
Distinct scripts	(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	? 5th row 0	
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.

Common Values

Value	Count	Frequency (%)
0	147721	> 99.9%
1	29	< 0.1%

Length

Histogram of lengths of the category

Common Values (Plot)



Value	Count	Frequency (%)
0	147721	> 99.9%
1	29	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
0	147721	> 99.9%
1	29	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per category

Value	Count	Frequency (%)
0	147721	> 99.9%
1	29	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per script*(unknown)*

Value	Count	Frequency (%)
0	147721	> 99.9%
1	29	< 0.1%

Most occurring blocks*Value*

(unknown)	Count	Frequency (%)
	147750	100.0%

Most frequent character per block*(unknown)*

Value	Count	Frequency (%)
0	147721	> 99.9%
1	29	< 0.1%

bicycle_serious_injury_count*Categorical***IMBALANCE**

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.1 MiB

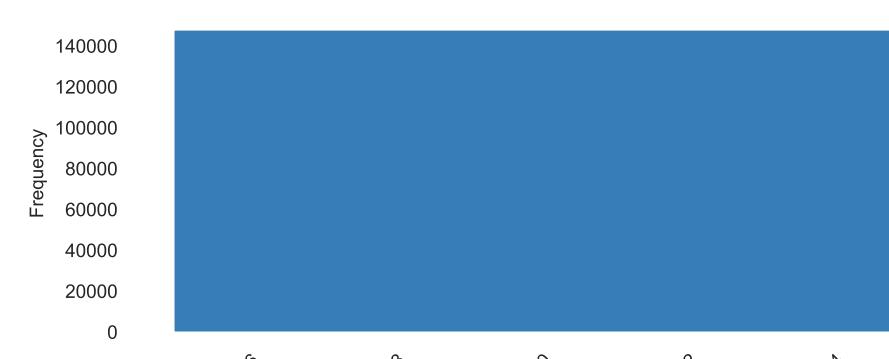
Length

	Characters and Unicode	Unique		Sample
Total characters	147750	Unique	1	? 1st row 0
Distinct characters	4	Unique (%)	< 0.1%	? 2nd row 0
Distinct scripts	1			? 3rd row 0
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)		? 4th row 0	? 5th row 0
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.

Common Values

Value	Count	Frequency (%)
0	147487	99.8%
1	260	0.2%
2		< 0.1%
3	1	< 0.1%

Length

Common Values (Plot)



Value	Count	Frequency (%)
0	147487	99.8%
1	260	0.2%
2	2	< 0.1%
3	1	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
0	147487	99.8%
1	260	0.2%
2	2	< 0.1%
3	1	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
0	147487	99.8%
1	260	0.2%
2	2	< 0.1%
3	1	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per script

Value	Count	Frequency (%)
0	147487	99.8%
1	260	0.2%
2	2	< 0.1%
3	1	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per block

Value	Count	Frequency (%)
0	147487	99.8%
1	260	0.2%
2	2	< 0.1%
3	1	< 0.1%

pedestrian_death_count
Categorical

IMBALANCE

Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.1 MiB

Length

	Characters and Unicode	Unique	Sample
Total characters	147750	Unique 0 ?	1st row 0
Distinct characters	3	Unique (%) 0.0%	2nd row 0
Min length	1		3rd row 0
Distinct categories	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	? 4th row	0
Distinct scripts	(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	? 5th row	0
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	? 6th row	0

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

Common Values

Value	Count	Frequency (%)
0	147431	99.8%
1	317	0.2%
2	2	< 0.1%

Length

Histogram of lengths of the category

Common Values (Plot)

Value

Value	Count	Frequency (%)
0	147431	99.8%
1	317	0.2%
2	2	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
0	147431	99.8%
1	317	0.2%
2	2	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per category

Value	Count	Frequency (%)
0	147431	99.8%
1	317	0.2%
2	2	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per script

(unknown)	Count	Frequency (%)
0	147431	99.8%
1	317	0.2%
2	2	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per block

(unknown)	Count	Frequency (%)
0	147431	99.8%
1	317	0.2%
2	2	< 0.1%

pedestrian_serious_injury_count

Categorical

IMBALANCE

Distinct	5
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.1 MiB

Length

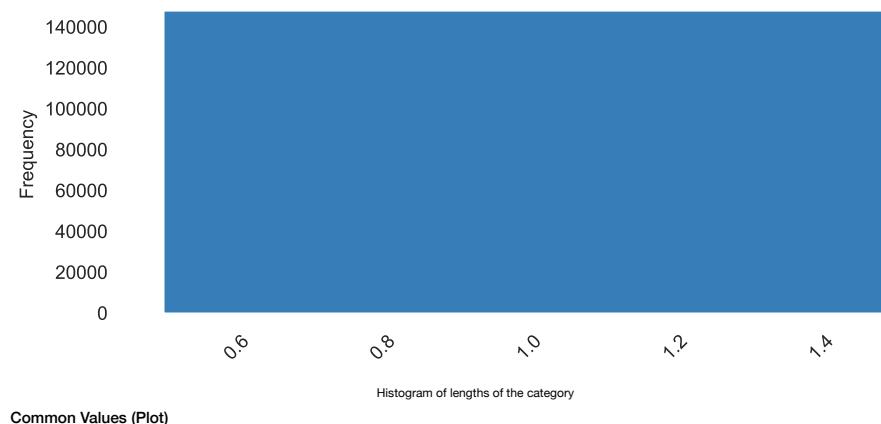
	Characters and Unicode	Unique		Sample
Total characters	147750	Unique	2	?
Distinct characters	5	Unique (%)	< 0.1%	1st row
Mean length	1			0
Min length	1	Distinct categories	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	5th row
Max length	1	1	(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Median length	1			2nd row
Distinct scripts	5			0
Distinct blocks	1			3rd row
				4th row
				0
				?
				0

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

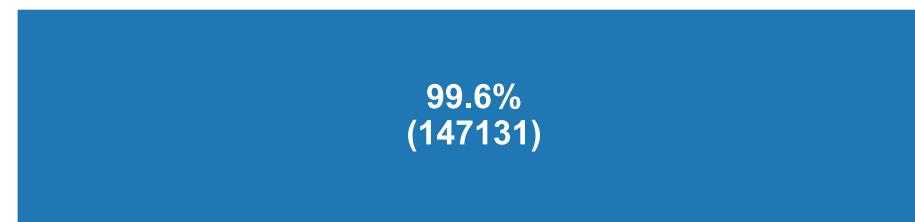
Common Values

Value	Count	Frequency (%)
0	147131	99.6%
1	606	0.4%
2	11	< 0.1%
3	1	< 0.1%
9	1	< 0.1%

Length



Common Values (Plot)



Value	Count	Frequency (%)
0	147131	99.6%
1	606	0.4%
2	11	< 0.1%
3	1	< 0.1%
9	1	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
0	147131	99.6%
1	606	0.4%
2	11	< 0.1%
3	1	< 0.1%
9	1	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per category

Value	Count	Frequency (%)
(unknown)	147131	99.6%
0	606	0.4%
2	11	< 0.1%
3	1	< 0.1%
9	1	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per script

Value	Count	Frequency (%)
(unknown)	147131	99.6%
0	606	0.4%
2	11	< 0.1%
3	1	< 0.1%
9	1	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	147750	100.0%
Most frequent character per block		
<i>(unknown)</i>		
Value	Count	Frequency (%)
0	147131	99.6%
1	606	0.4%
2	11	< 0.1%
3	1	< 0.1%
9	1	< 0.1%

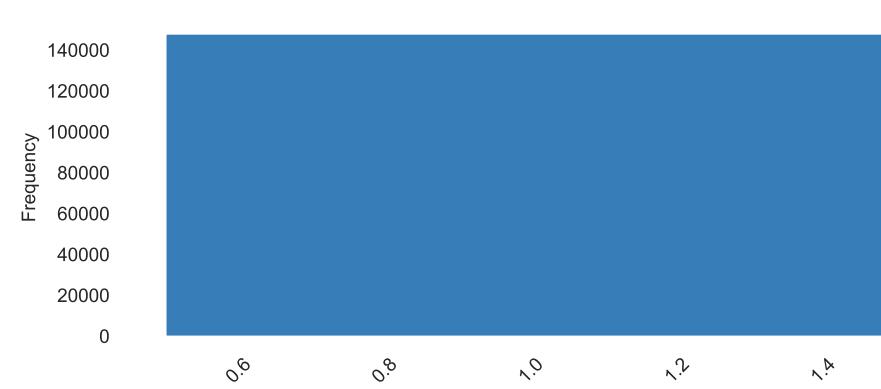
motorcycle_death_count	
Categorical	
IMBALANCE	
Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.1 MiB
Length	
Max length	1
Median length	1
Mean length	1
Min length	1
Characters and Unicode	
Total characters	147750
Distinct characters	3
Distinct categories	1
Distinct scripts	1
Distinct blocks	1
Unique	0
Unique (%)	0.0%
Sample	
1st row	0
2nd row	0
3rd row	0
4th row	0
5th row	0

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

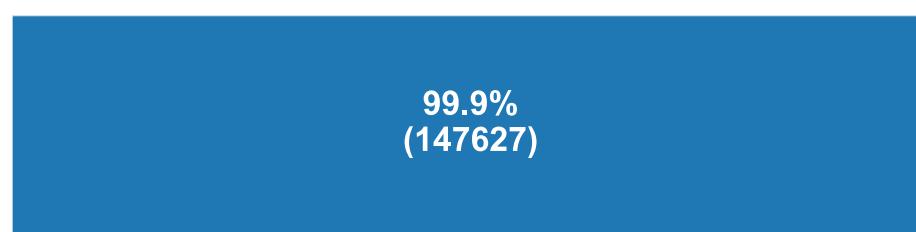
Common Values

Value	Count	Frequency (%)
0	147627	99.9%
1	121	0.1%
2	2	< 0.1%

Length



Common Values (Plot)



Value	Count	Frequency (%)
0	147627	99.9%
1	121	0.1%
2	2	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
0	147627	99.9%
1	121	0.1%
2	2	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per category

Value	Count	Frequency (%)
(unknown)	147627	99.9%
0	147627	99.9%
1	121	0.1%
2	2	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per script

Value	Count	Frequency (%)
(unknown)	147627	99.9%
0	147627	99.9%
1	121	0.1%
2	2	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per block

Value	Count	Frequency (%)
(unknown)	147627	99.9%
0	147627	99.9%
1	121	0.1%
2	2	< 0.1%

motorcycle_serious_injury_count

Categorical

IMBALANCE

Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.1 MiB

Length

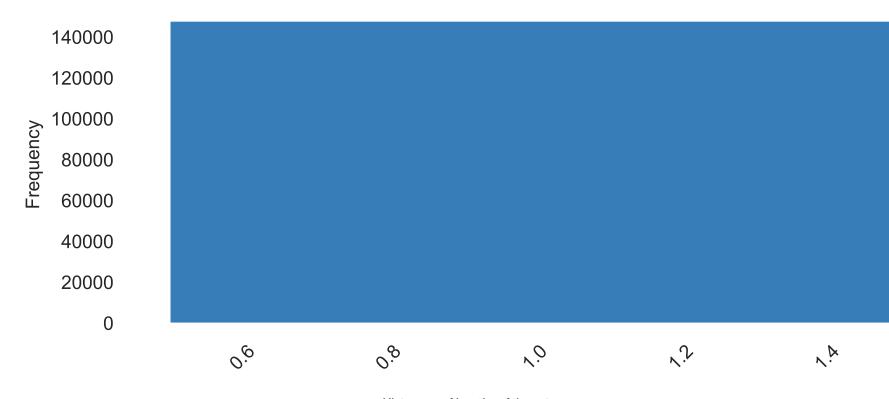
	Characters and Unicode	Unique	Sample
Max length	1	Total characters	147750
Median length	1	Unique	0
Mean length	1	Unique (%)	0.0%
Min length	1	Distinct characters	3
		?	1st row
		?	2nd row
		?	3rd row
		?	4th row
		?	5th row
		?	
		?	
		?	

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

Common Values

Value	Count	Frequency (%)
0	147064	99.5%
1	662	0.4%
2	24	< 0.1%

Length



Common Values (Plot)



Value

Value	Count	Frequency (%)
0	147064	99.5%
1	662	0.4%
2	24	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
0	147064	99.5%
1	662	0.4%
2	24	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per category

Value	Count	Frequency (%)
0	147064	99.5%
1	662	0.4%
2	24	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per script

(unknown)	Value	Count	Frequency (%)
0		147064	99.5%
1		662	0.4%
2		24	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per block

(unknown)	Value	Count	Frequency (%)
0		147064	99.5%
1		662	0.4%
2		24	< 0.1%

other_death_count

Categorical

CONSTANT

Distinct	1
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.1 MiB

Length

	Characters and Unicode	Unique		Sample
Total characters	147750	Unique	0	1st row
Distinct characters	1	Unique (%)	0.0%	2nd row
Distinct categories	1			3rd row
Distinct scripts	1			4th row
Distinct blocks	1			5th row

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

Common Values

Value	Count	Frequency (%)
0	147750	100.0%

Length**Common Values (Plot)**



Value	Count	Frequency (%)
0	147750	100.0%

Most occurring characters

Value	Count	Frequency (%)
0	147750	100.0%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per category

(unknown)	Value	Count	Frequency (%)
	0	147750	100.0%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per script

(unknown)	Value	Count	Frequency (%)
	0	147750	100.0%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per block

(unknown)	Value	Count	Frequency (%)
	0	147750	100.0%

other_serious_injury_count

Categorical

IMBALANCE

Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.1 MiB

Length

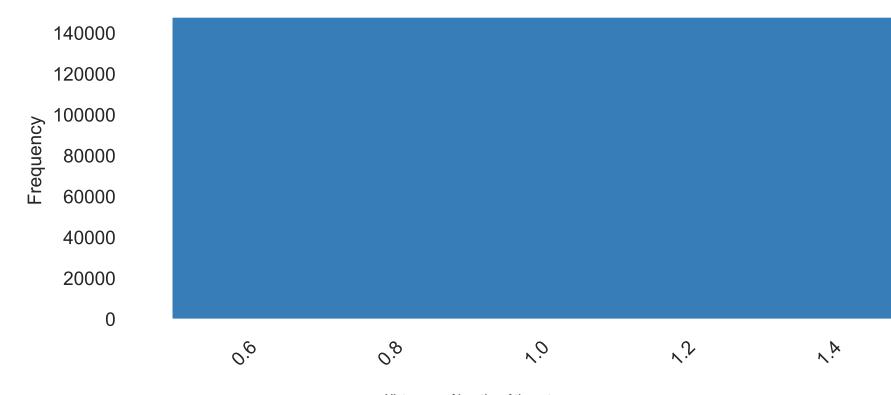
	Characters and Unicode	Unique		Sample
Max length	1	Total characters	147750	1st row
Median length	1	Unique (%)	< 0.1%	2nd row
Mean length	1	Distinct characters	3	3rd row
Min length	1	Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	4th row
		Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	5th row
		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.

Common Values

Value	Count	Frequency (%)
0	147746	> 99.9%
1	3	< 0.1%
3	1	< 0.1%

Length



Common Values (Plot)



Value

Value	Count	Frequency (%)
0	147746	> 99.9%
1	3	< 0.1%
3	1	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
0	147746	> 99.9%
1	3	< 0.1%
3	1	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per category

Value	Count	Frequency (%)
0	147746	> 99.9%
1	3	< 0.1%
3	1	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per script

Value	Count	Frequency (%)
(unknown)	147746	> 99.9%
0	3	< 0.1%
1	1	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per block

Value	Count	Frequency (%)
(unknown)	147746	> 99.9%
0	3	< 0.1%
1	1	< 0.1%

onsys_fl
Boolean

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	144.4 KIB

Value	Count	Frequency (%)
False	74709	50.6%
True	73041	49.4%

50.6%
(74709)

49.4%
(73041)



private_dr_fl

Boolean

CONSTANT

Distinct	1
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%

Memory size 144.4 KIB**Value**

	Count	Frequency (%)
False	147750	100.0%

100.0%
(147750)

■ False

micromobility_serious_injury_count

Categorical

IMBALANCE

Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%

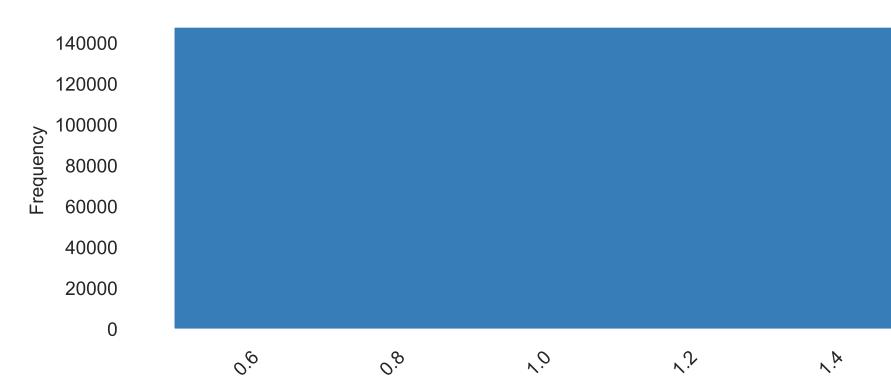
Memory size 1.1 MiB**Length**

	Characters and Unicode	Unique		Sample
Total characters	147750	Unique	0	?
		Unique (%)	0.0%	1st row
Distinct characters	3			0
				2nd row
Distinct categories	1			0
				3rd row
				4th row
				5th row
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.

Common Values

	Count	Frequency (%)
0	147708	> 99.9%
1	40	< 0.1%
2	2	< 0.1%

Length

Common Values (Plot)



Value	Count	Frequency (%)
0	147708	> 99.9%
1	40	< 0.1%
2	2	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
0	147708	> 99.9%
1	40	< 0.1%
2	2	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per category

Value	Count	Frequency (%)
(unknown)	147708	> 99.9%
0	147708	> 99.9%
1	40	< 0.1%
2	2	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per script

Value	Count	Frequency (%)
(unknown)	147708	> 99.9%
0	147708	> 99.9%
1	40	< 0.1%
2	2	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per block

Value	Count	Frequency (%)
(unknown)	147708	> 99.9%
0	147708	> 99.9%
1	40	< 0.1%
2	2	< 0.1%

micromobility_death_count
Categorical

IMBALANCE

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	1.1 MiB

Length

	Characters and Unicode	Unique	Sample
Total characters	147750	Unique 0 ? 1st row 0	
Distinct characters	2	Unique (%) 0.0% 2nd row 0	
Min length	1	3rd row 0	
Distinct categories (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	1	? 4th row 0	
Distinct scripts (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	1	? 5th row 0	
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	? 6th row 0	

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

Common Values

Value	Count	Frequency (%)
0	147744	> 99.9%
1	6	< 0.1%

Length

Histogram of lengths of the category

Common Values (Plot)



Value	Count	Frequency (%)
0	147744	> 99.9%
1	6	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
0	147744	> 99.9%
1	6	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per category

Value	Count	Frequency (%)
0	147744	> 99.9%
1	6	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per script*(unknown)*

Value	Count	Frequency (%)
0	147744	> 99.9%
1	6	< 0.1%

Most occurring blocks*(unknown)*

Value	Count	Frequency (%)
(unknown)	147750	100.0%

Most frequent character per block*(unknown)*

Value	Count	Frequency (%)
0	147744	> 99.9%
1	6	< 0.1%

micromobility_fl

Boolean

CONSTANT MISSING

Distinct	1
Distinct (%)	0.3%
Missing	147439
Missing (%)	99.8%
Memory size	288.7 KIB

Value	Count	Frequency (%)
True	311	0.2%
(Missing)	147439	99.8%

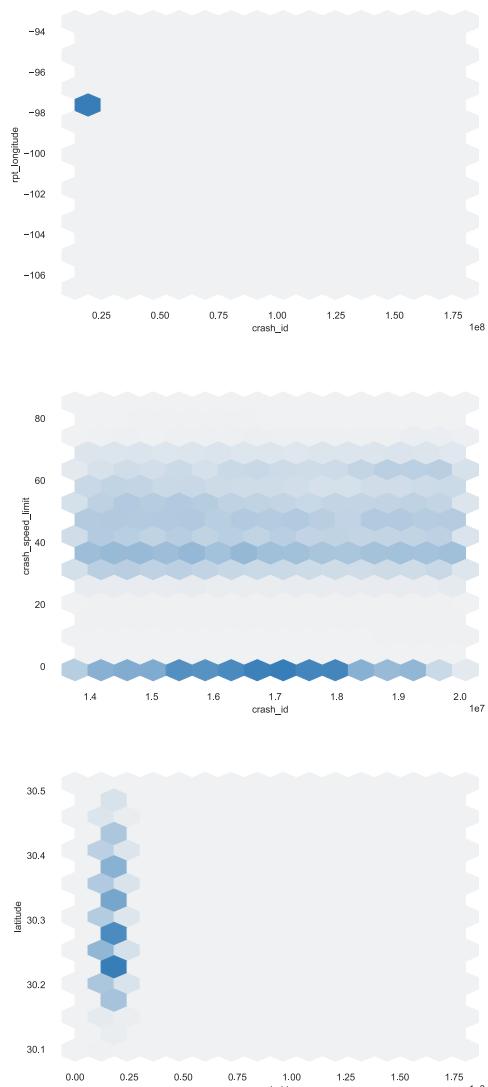
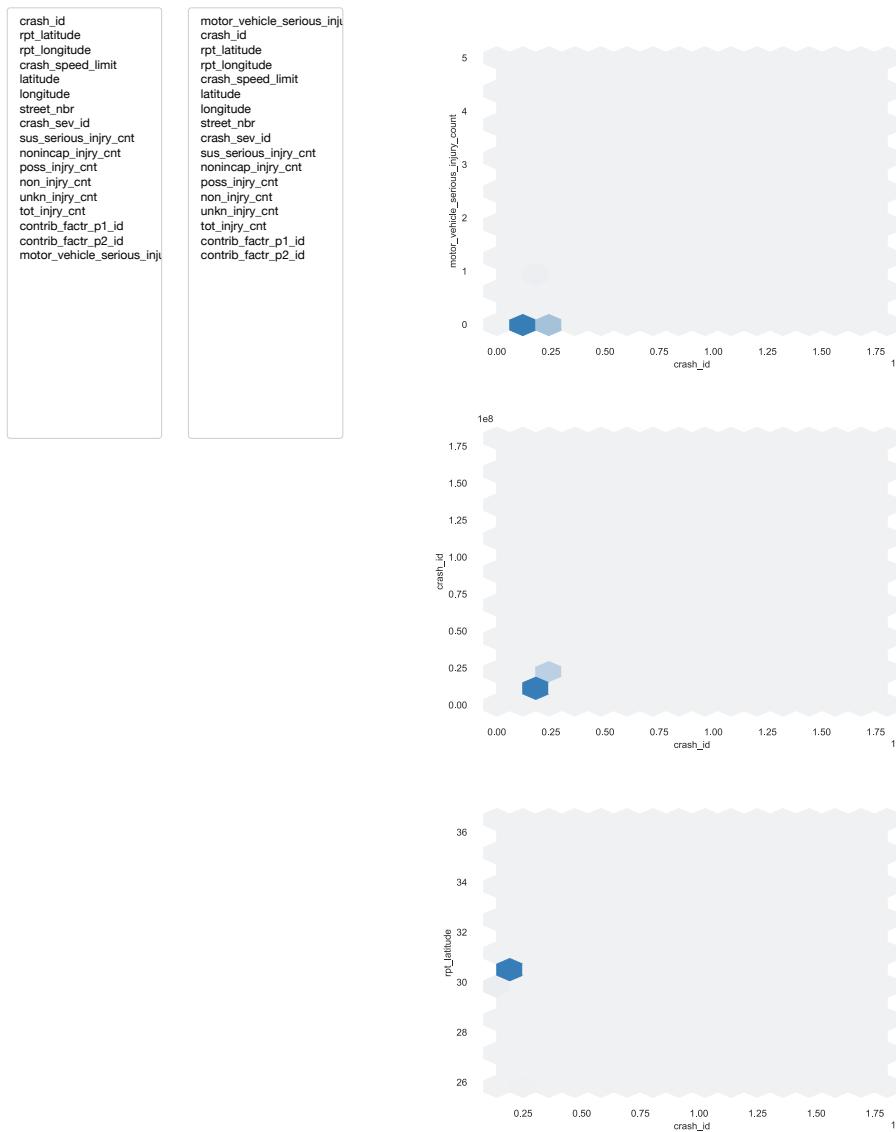


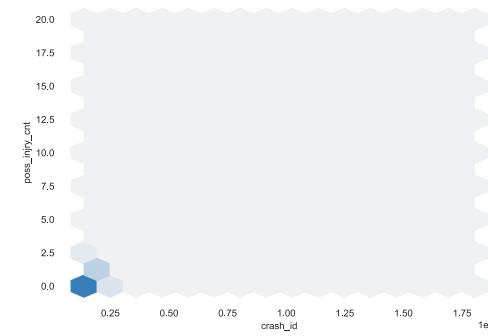
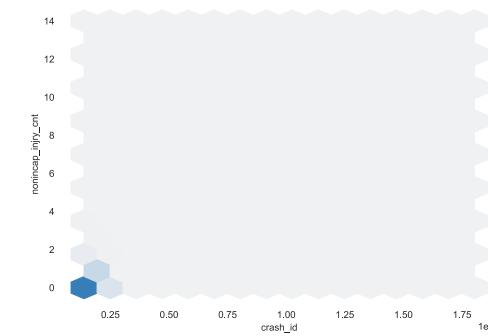
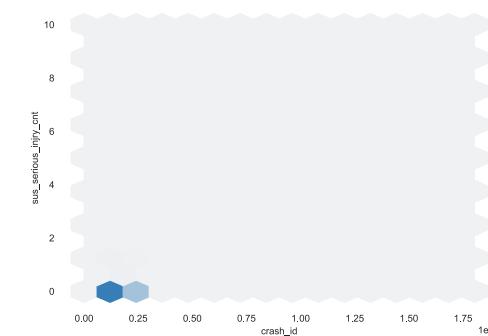
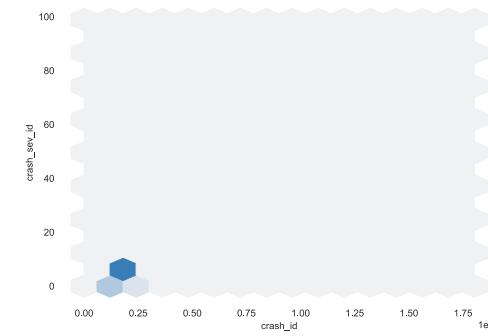
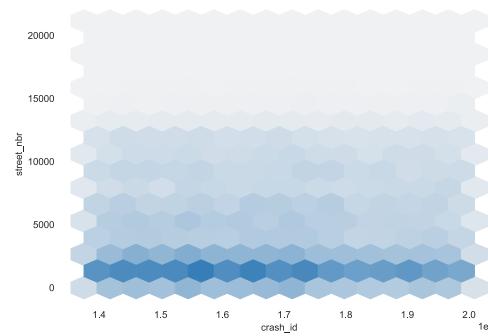
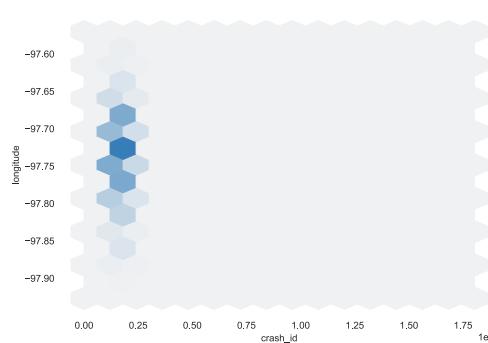
100.0%
(311)

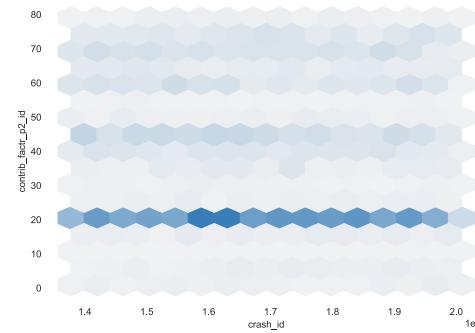
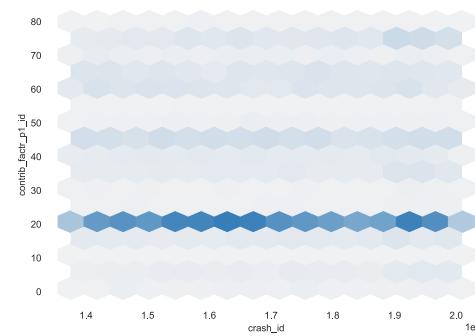
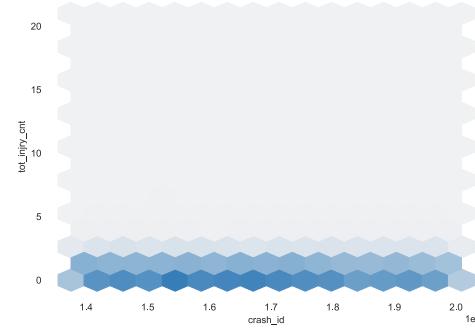
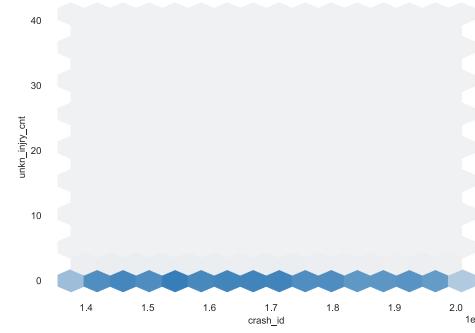
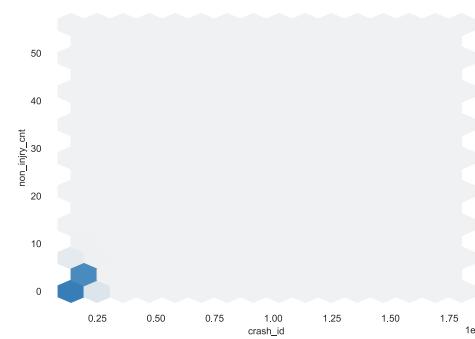


True

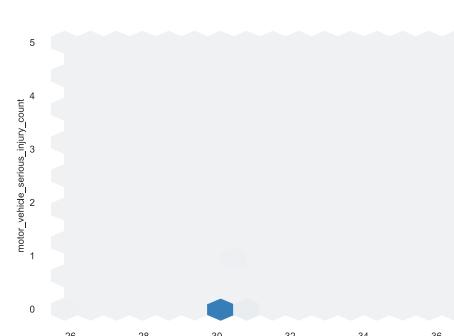
Interactions

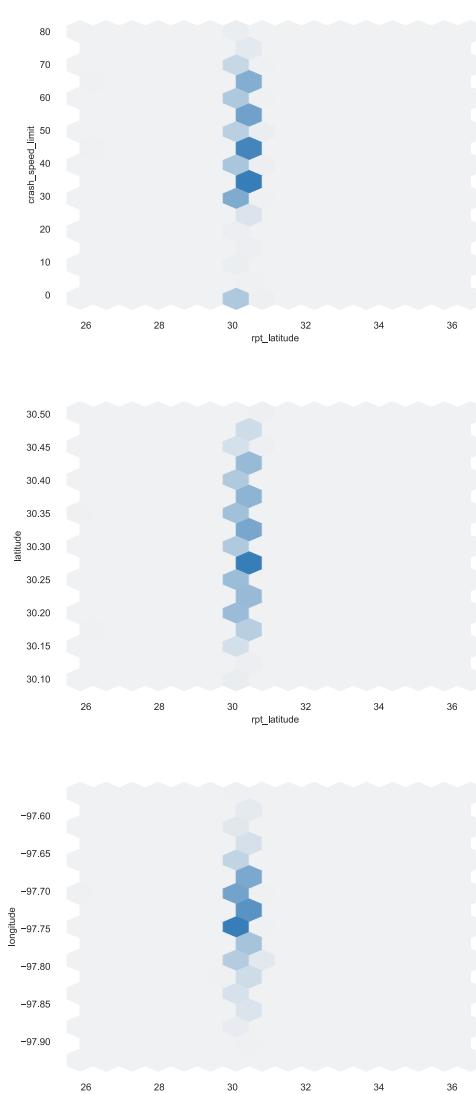
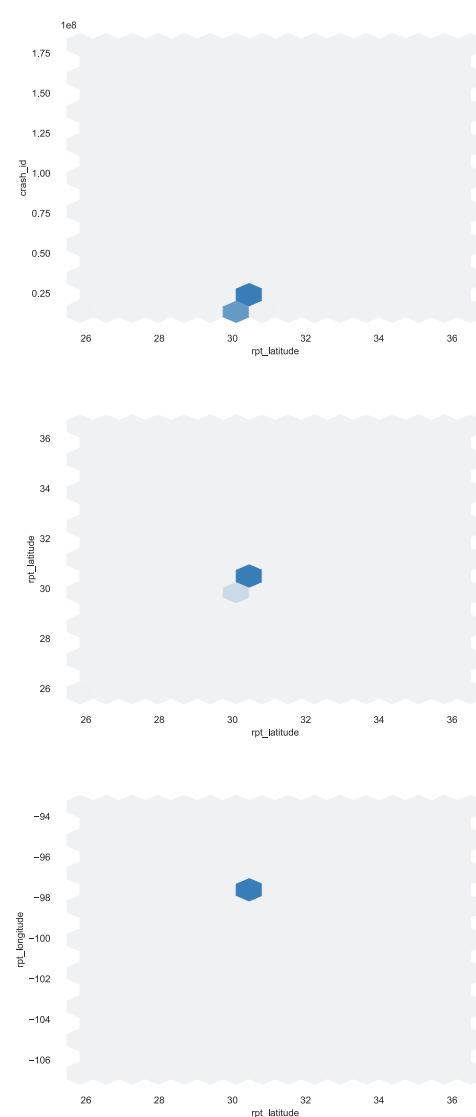


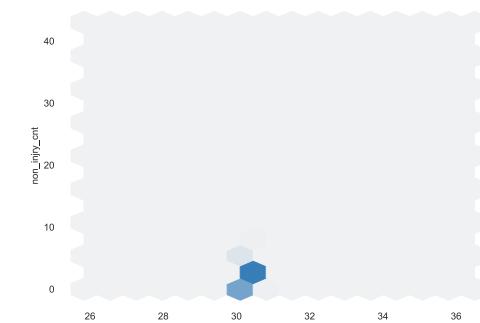
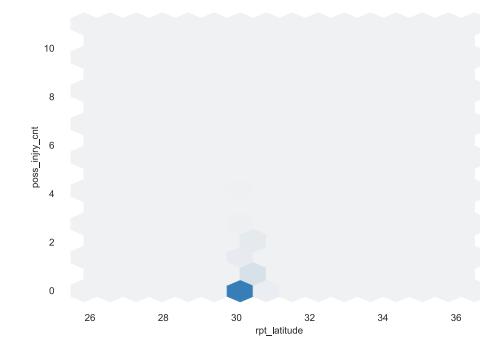
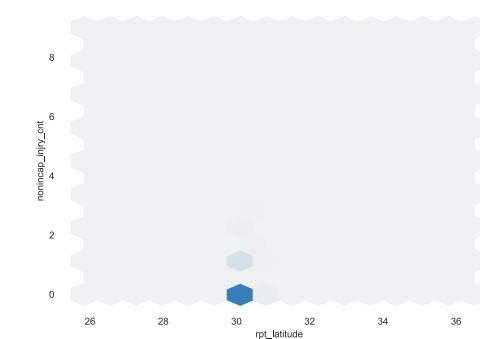
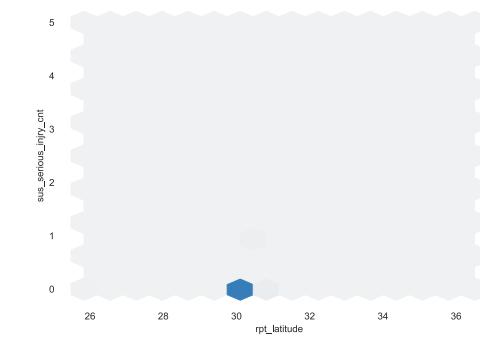
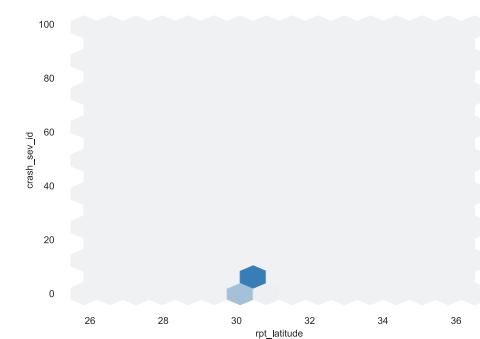
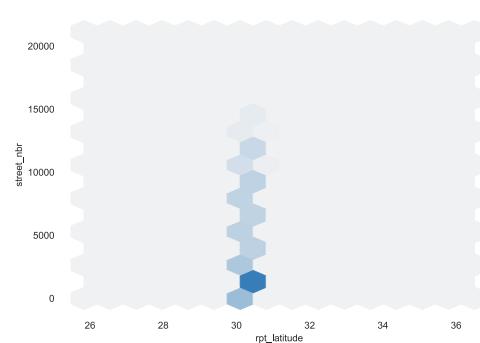


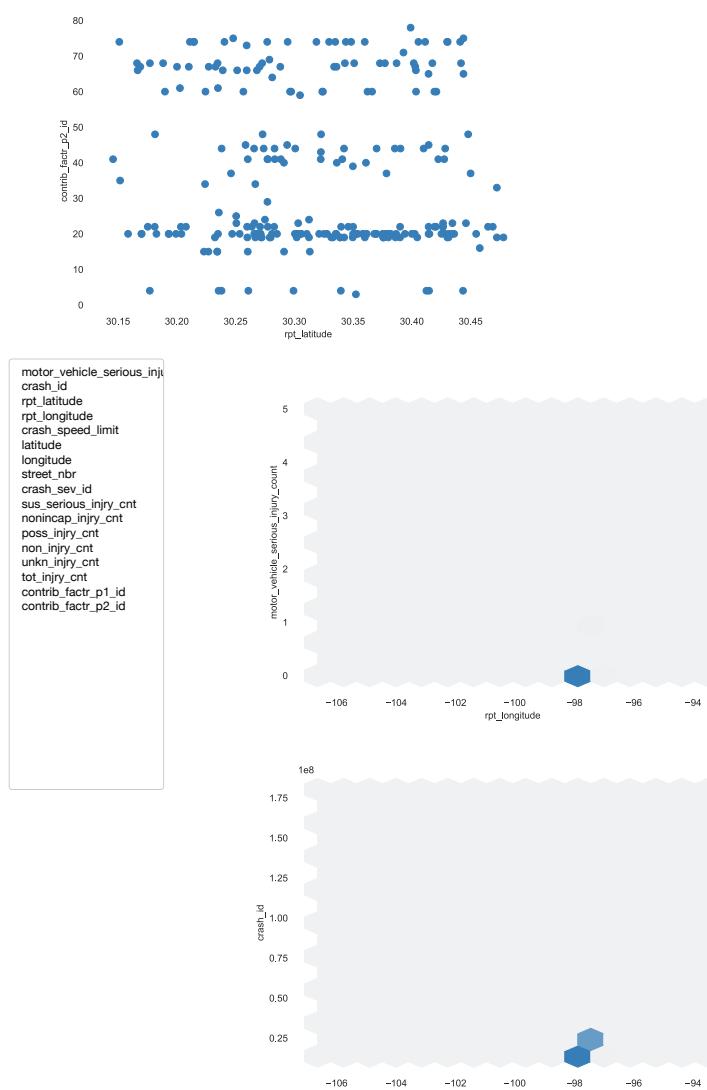
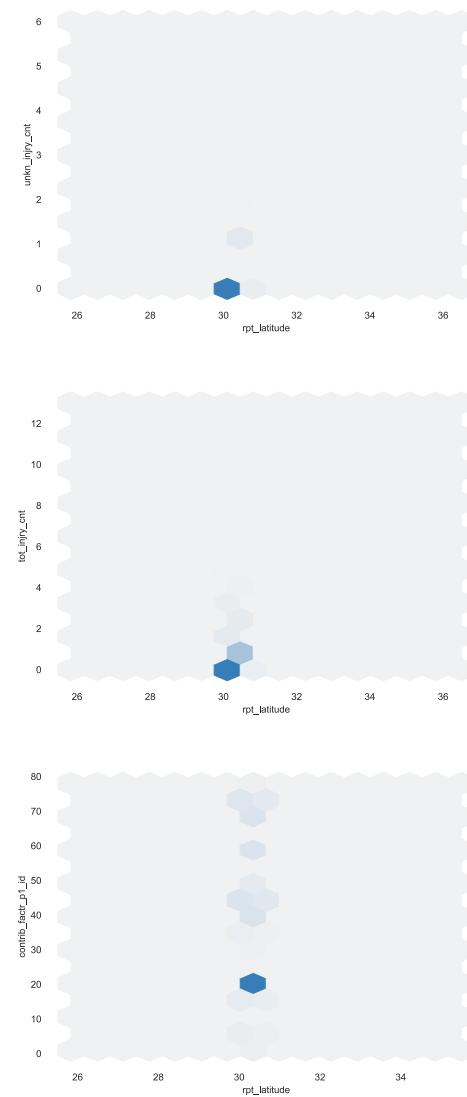


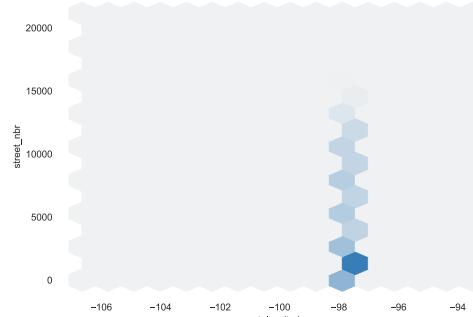
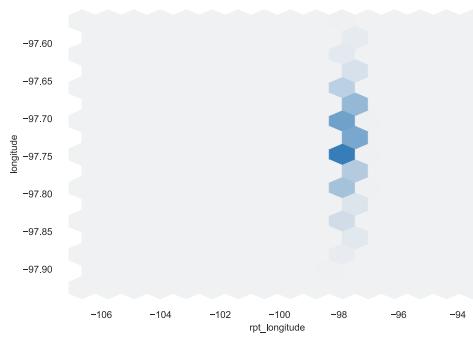
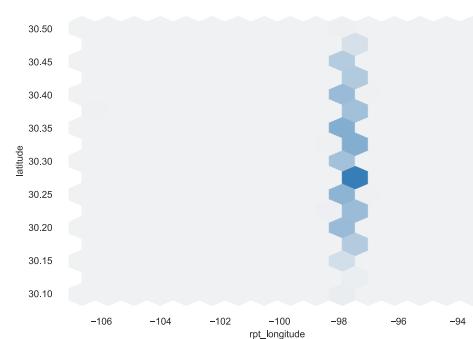
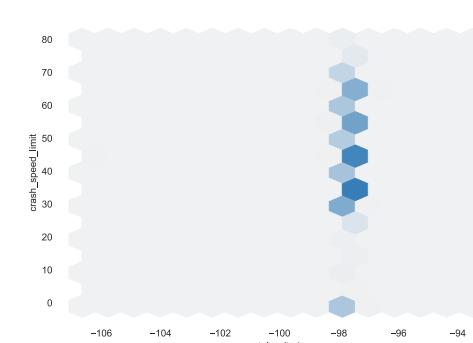
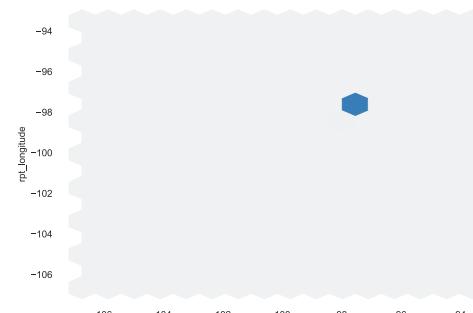
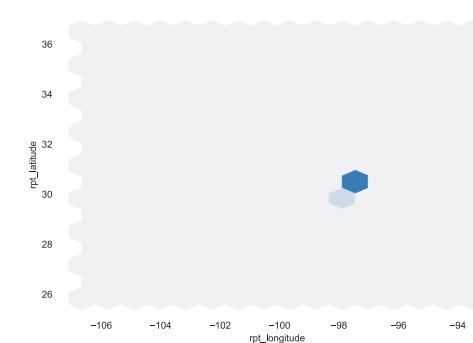
motor_vehicle_serious_inj
crash_id
rpt_latitude
rpt_longitude
crash_speed_limit
latitude
longitude
street_nbr
crash_sev_id
sus_serious_injry_cnt
noninjcap_injry_cnt
poss_injry_cnt
non_injry_cnt
unkn_injry_cnt
tot_injry_cnt
contrib_factr_p1_id
contrib_factr_p2_id

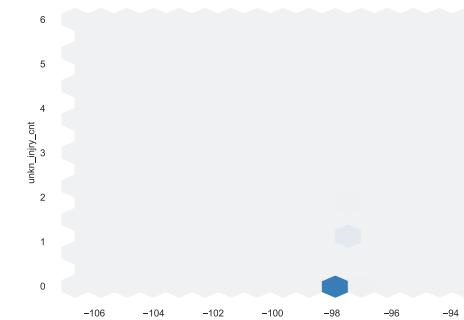
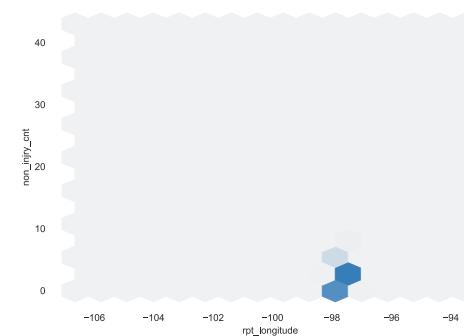
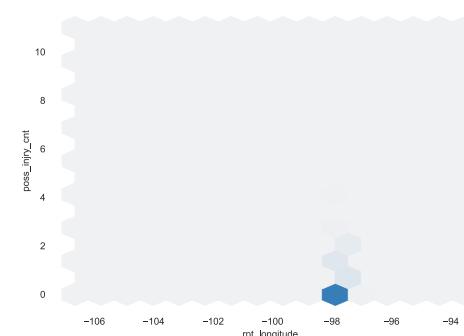
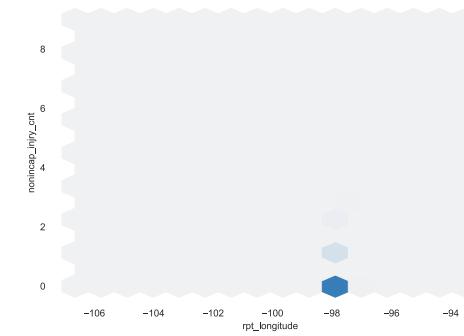
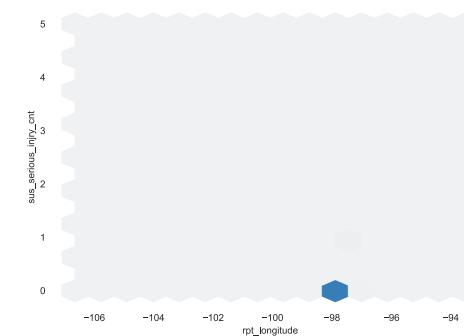
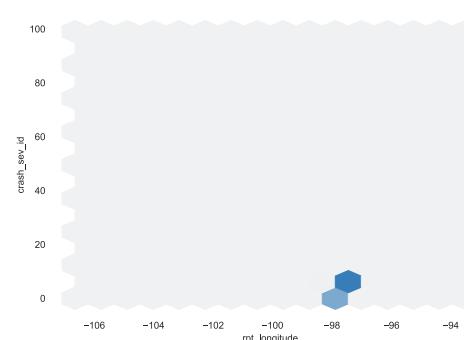


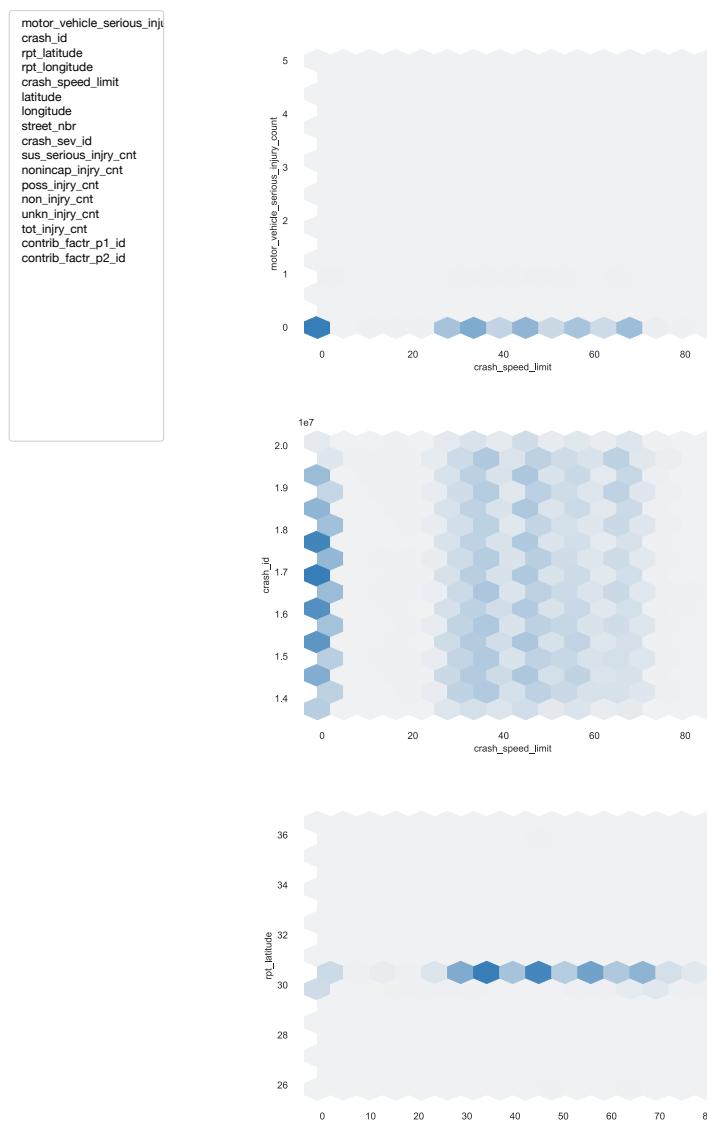
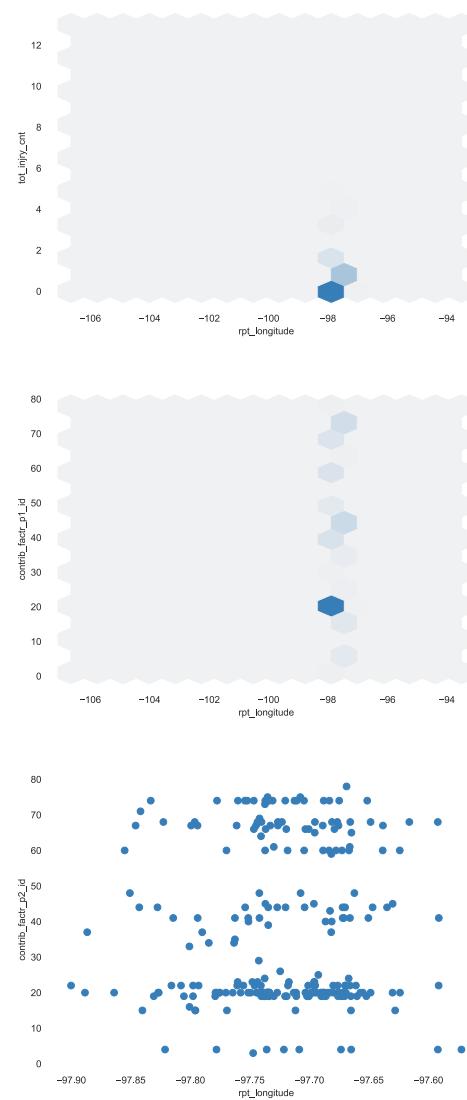


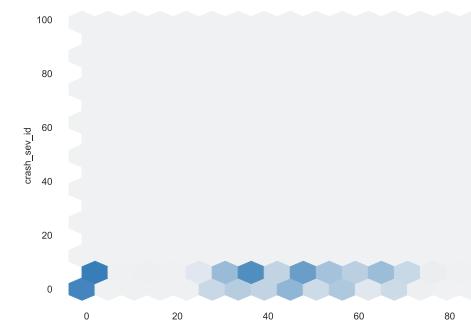
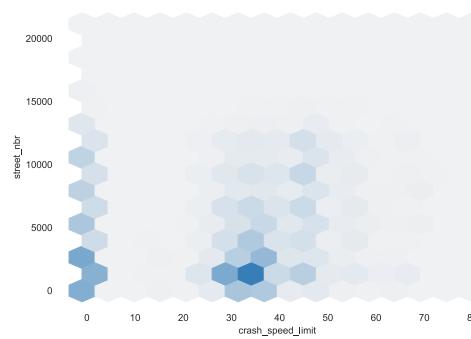
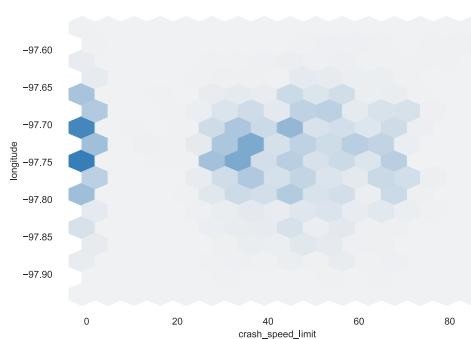
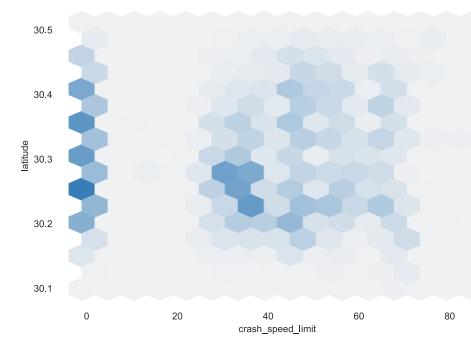
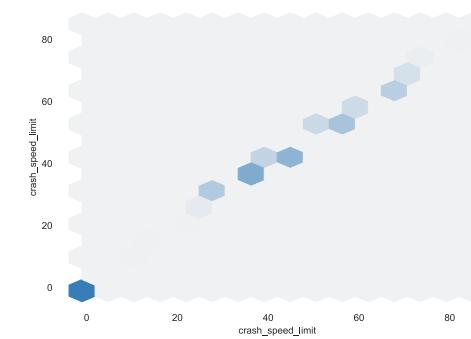
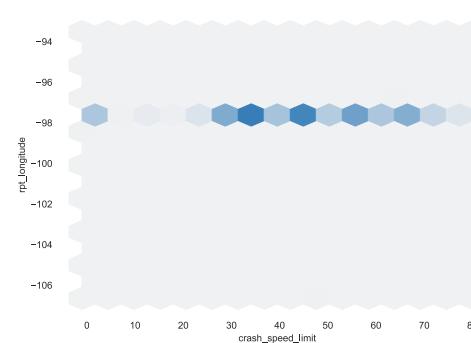


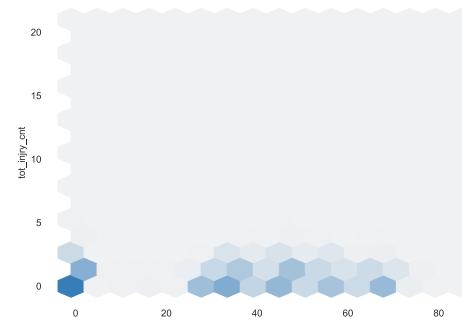
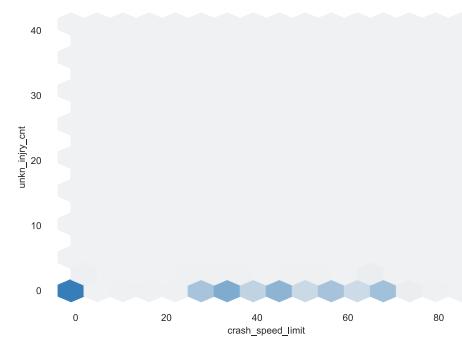
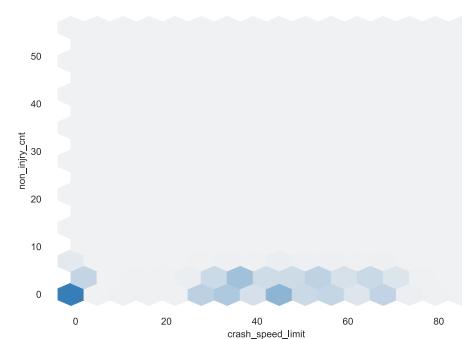
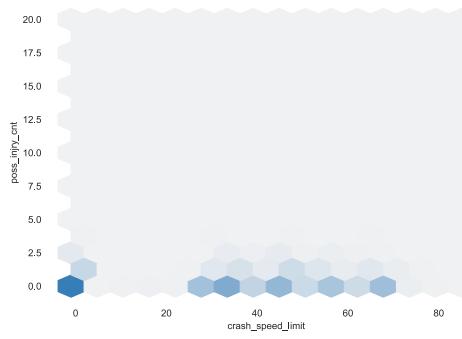
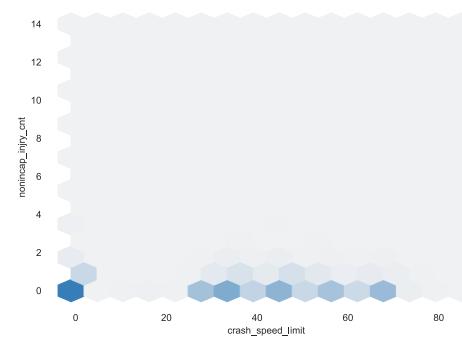
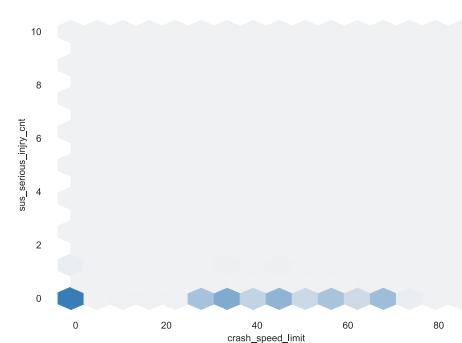


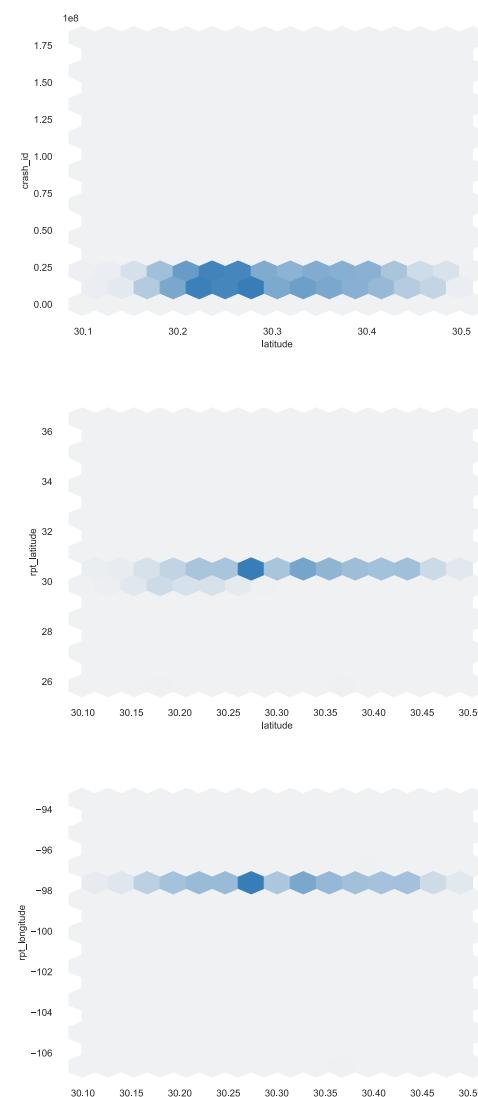
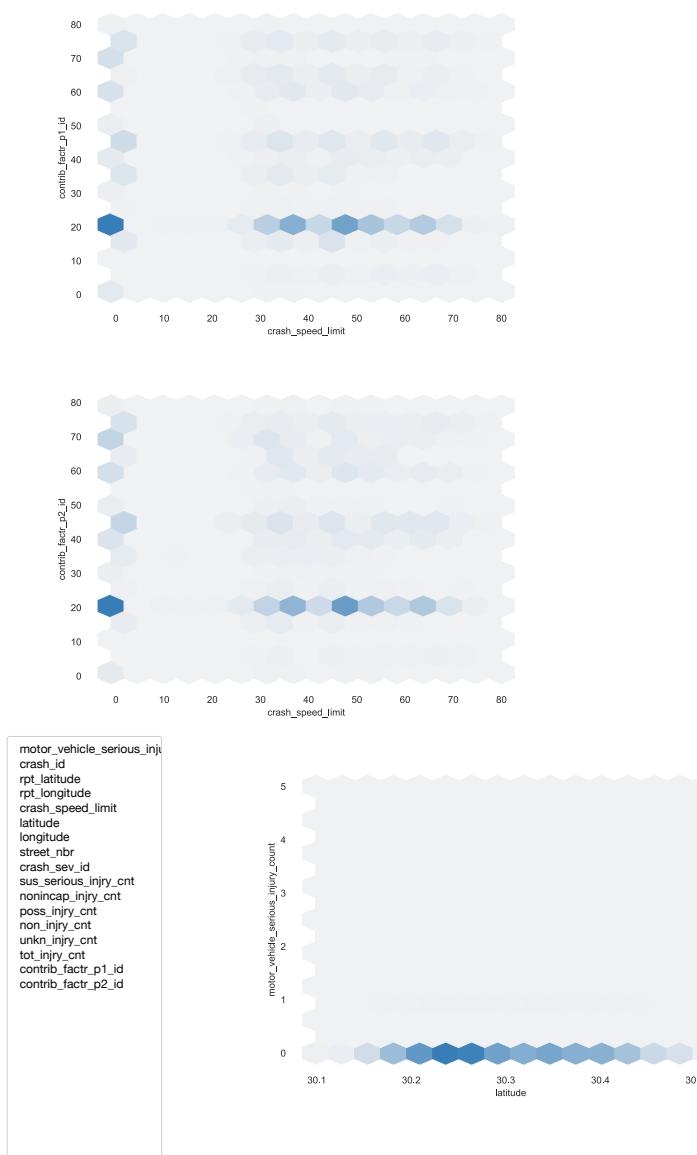


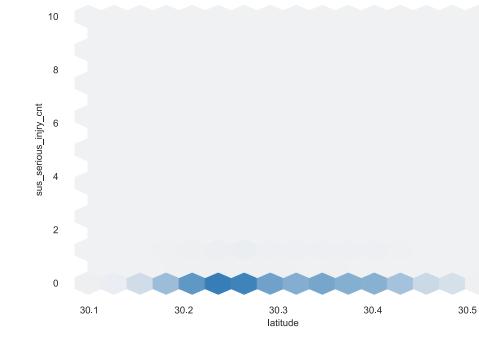
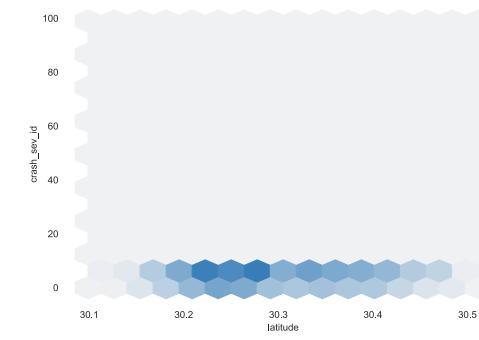
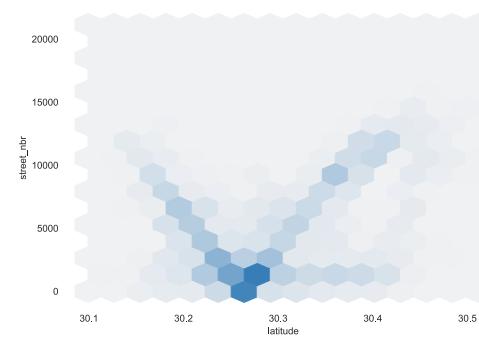
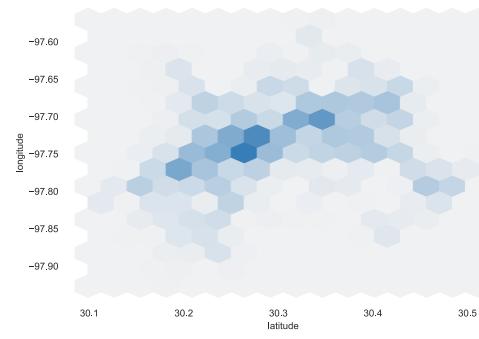
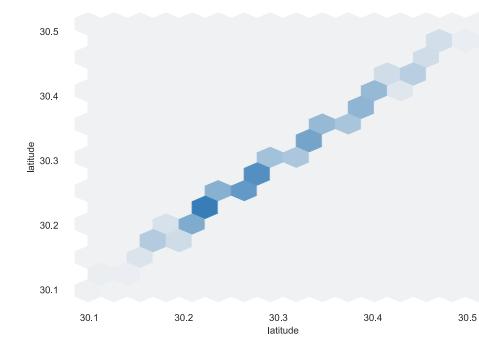
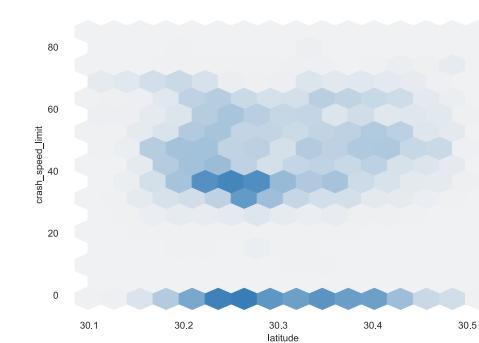


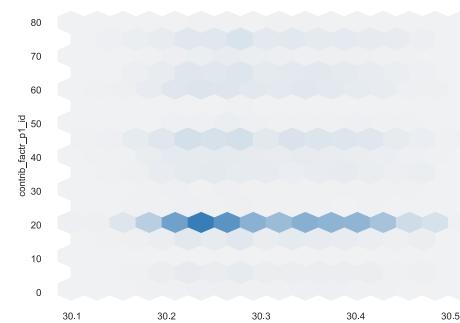
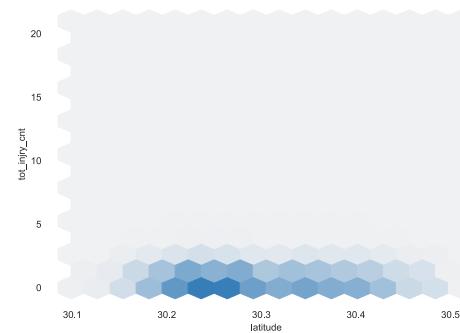
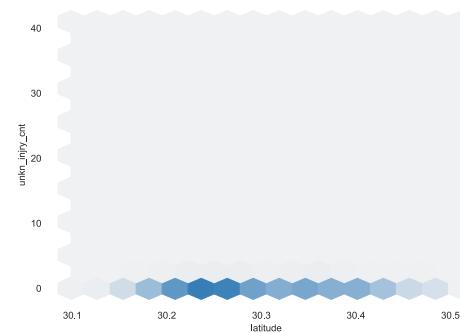
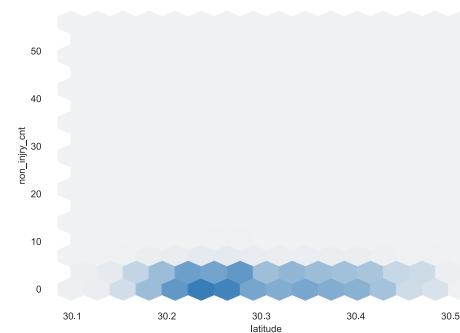
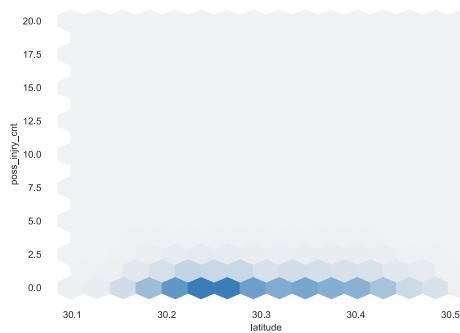
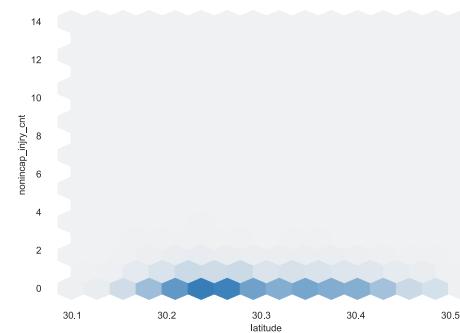


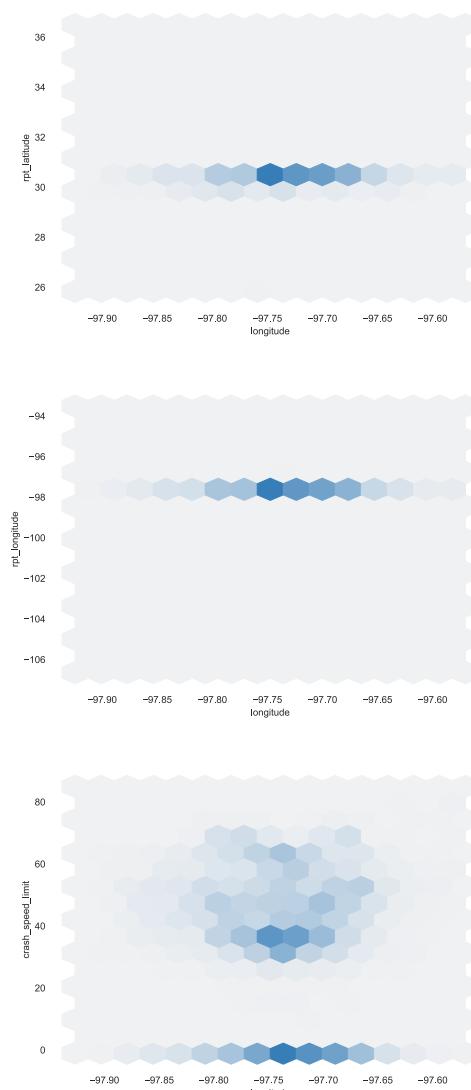
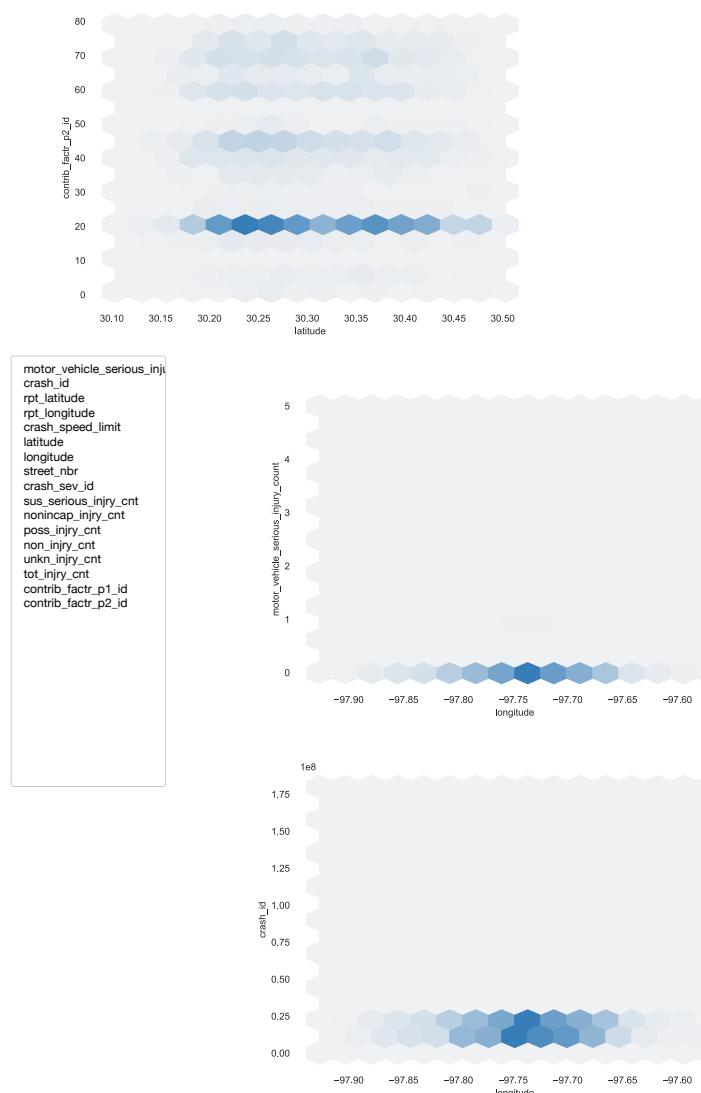


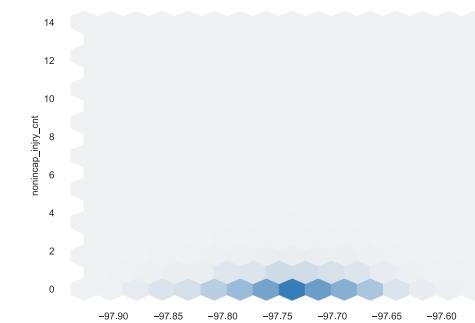
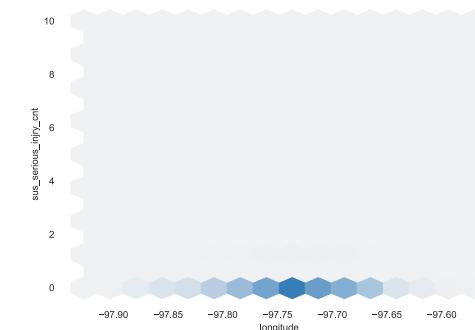
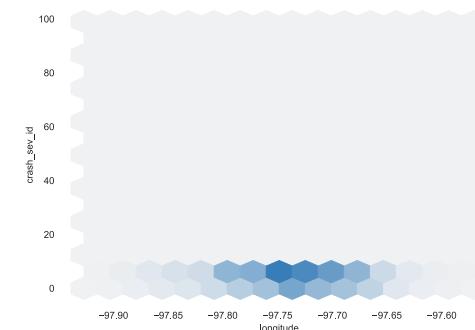
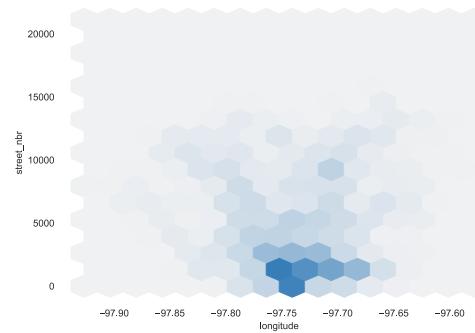
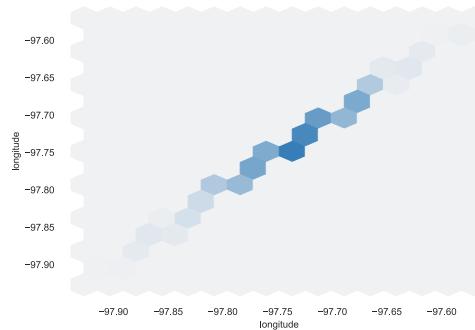
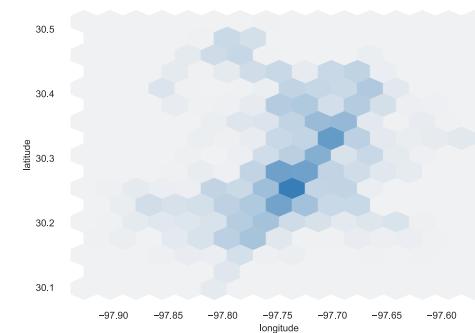


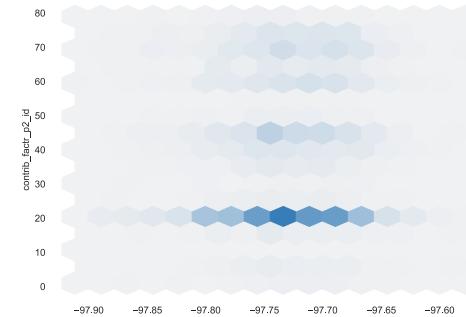
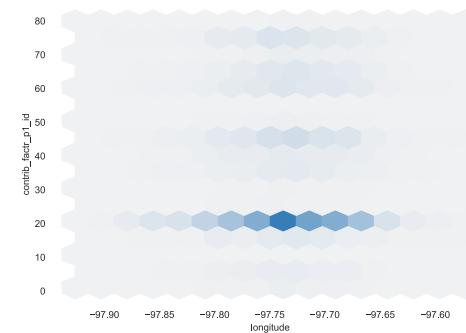
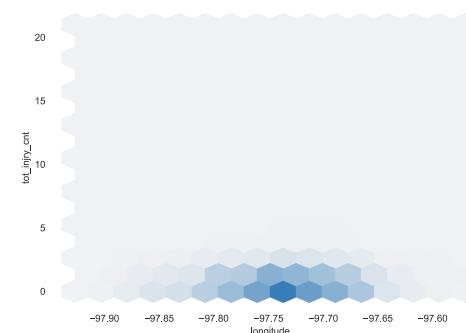
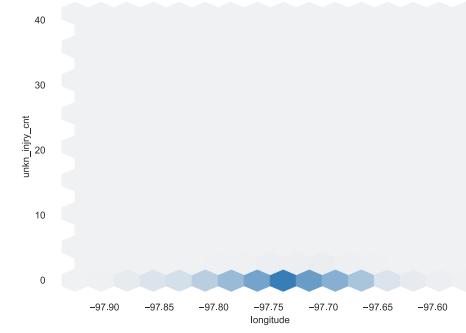
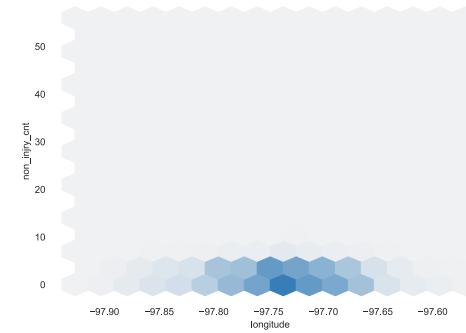
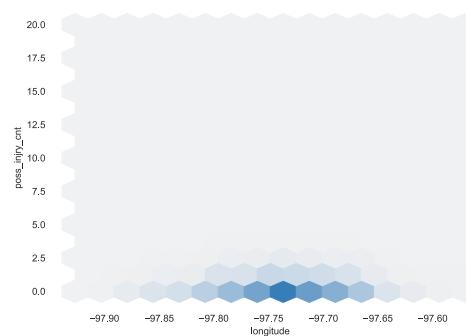


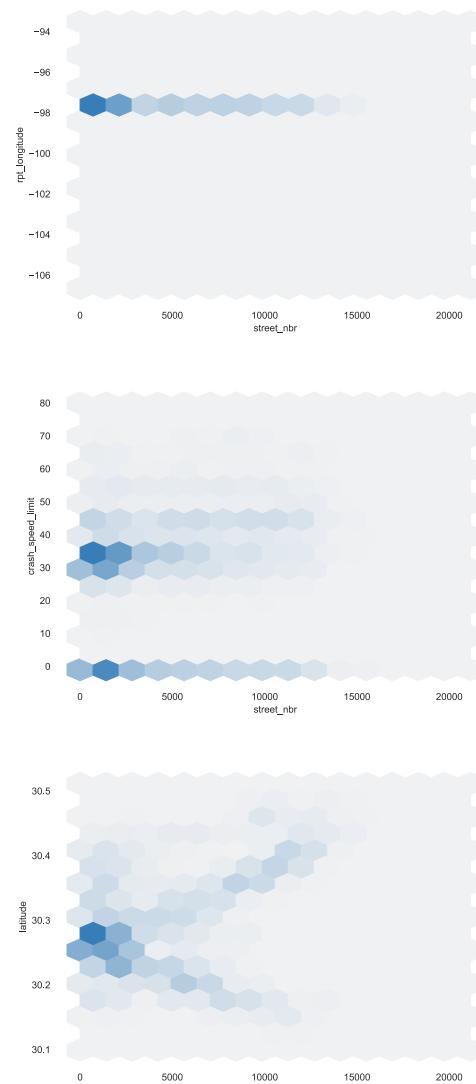
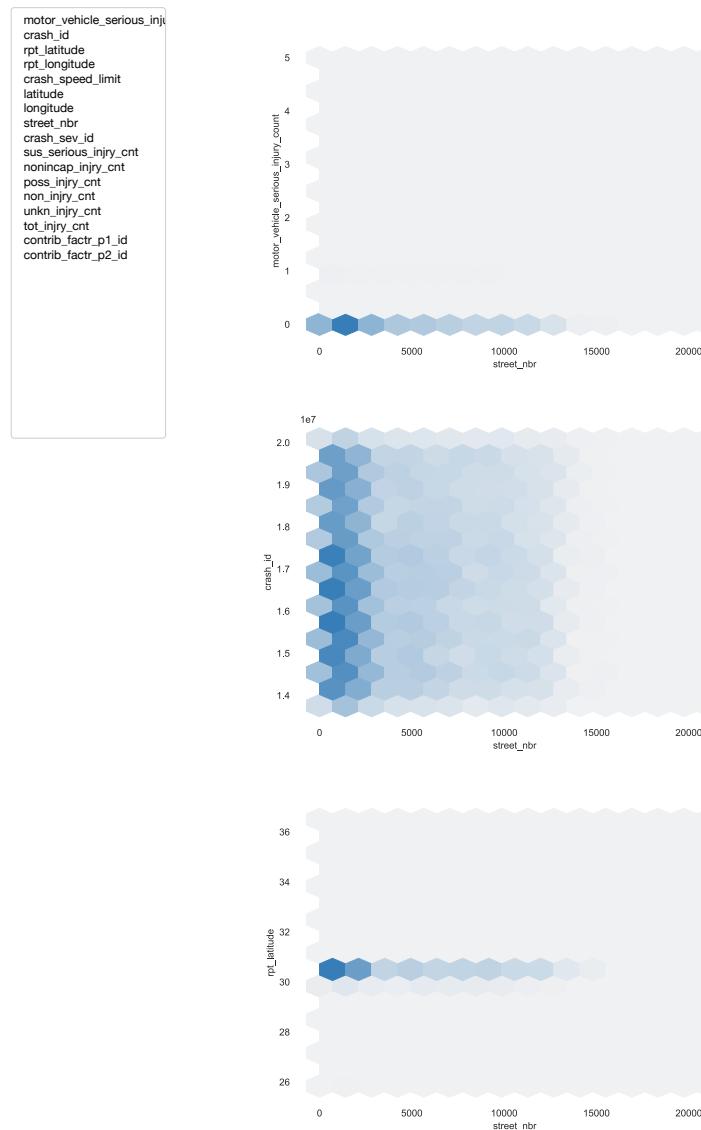


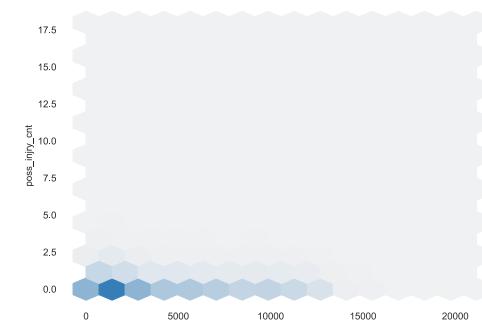
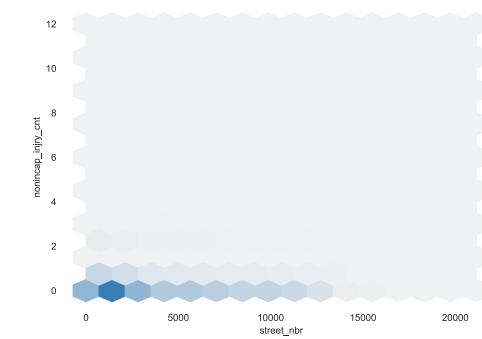
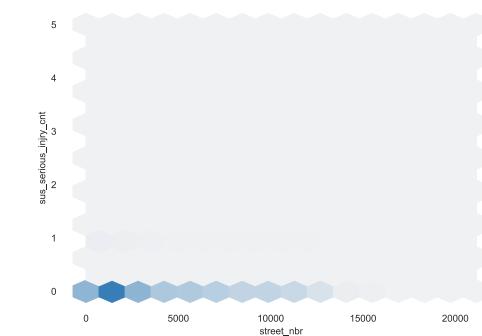
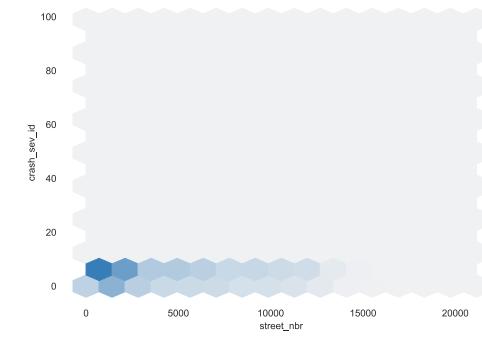
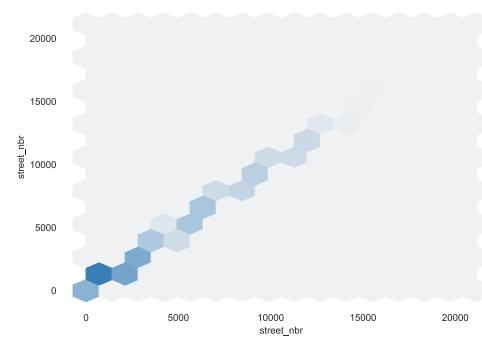
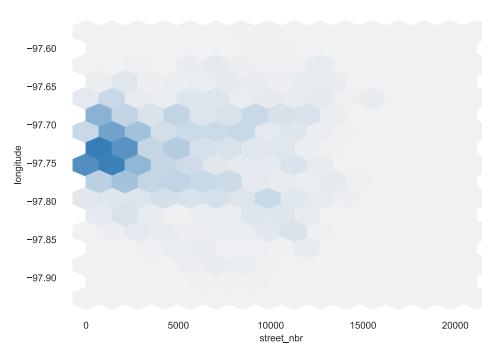


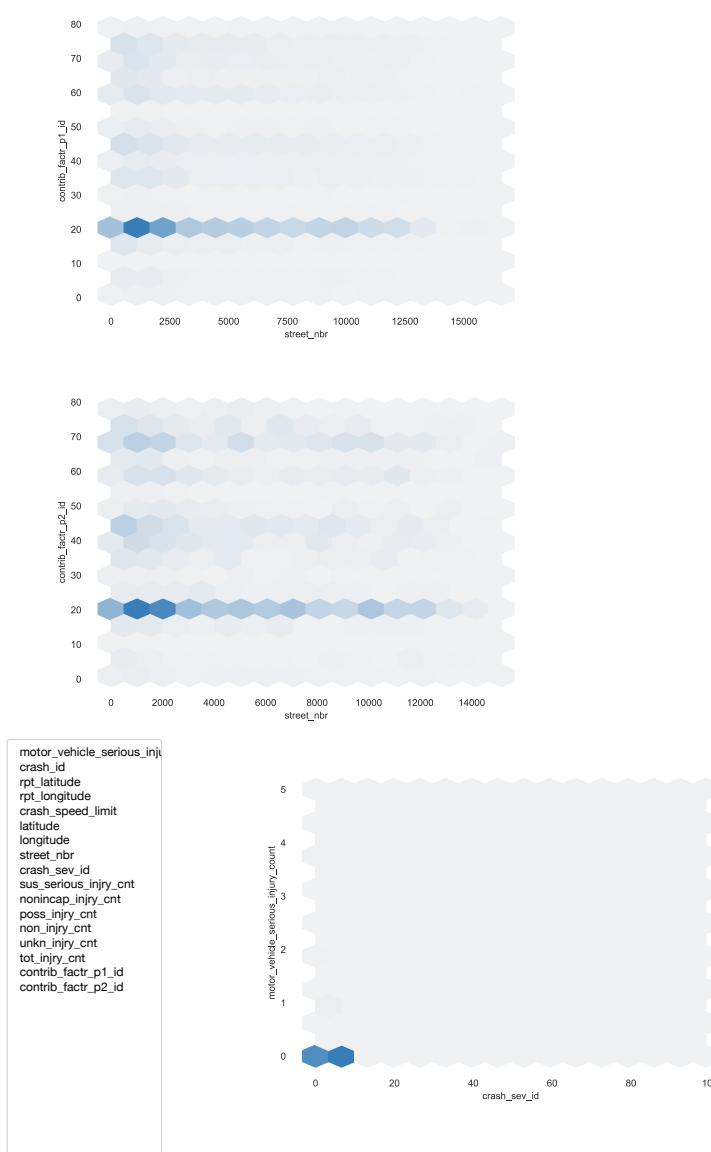
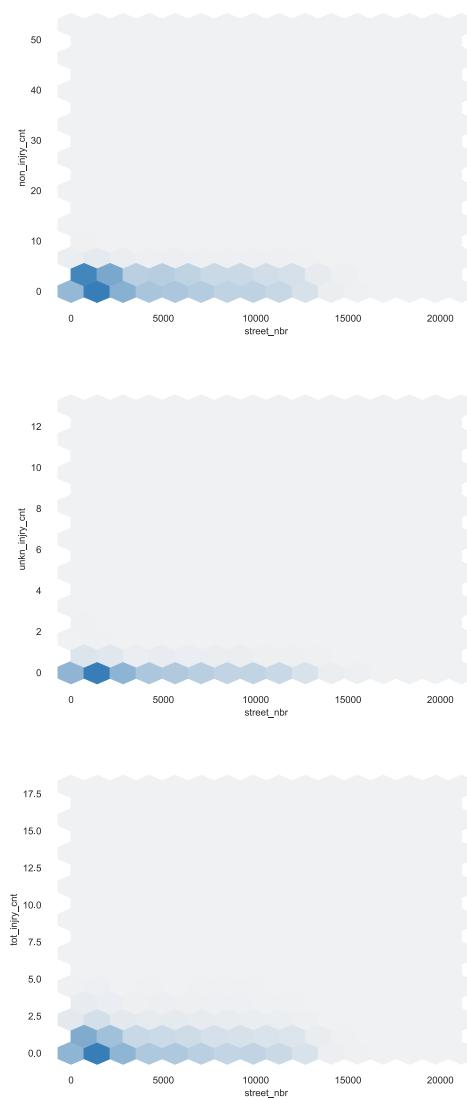


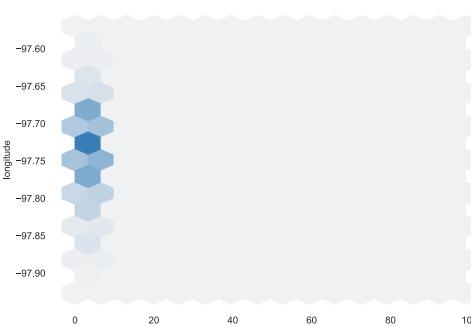
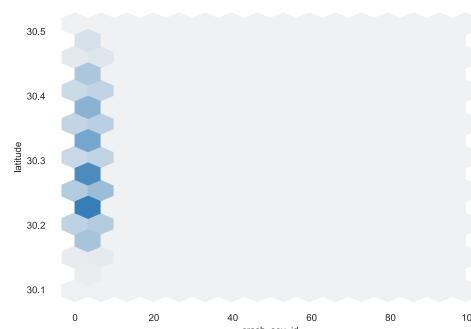
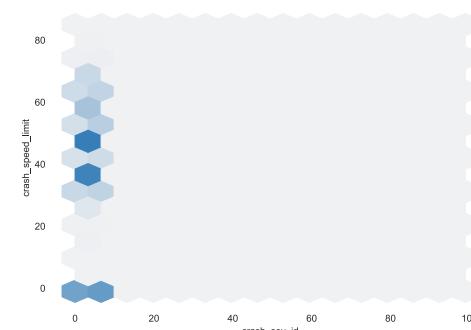
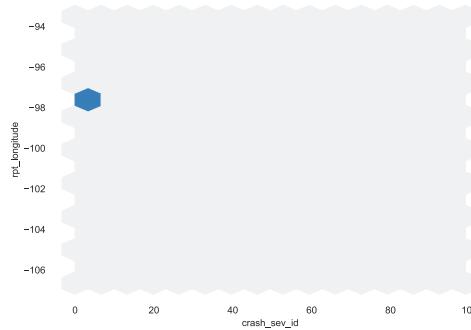
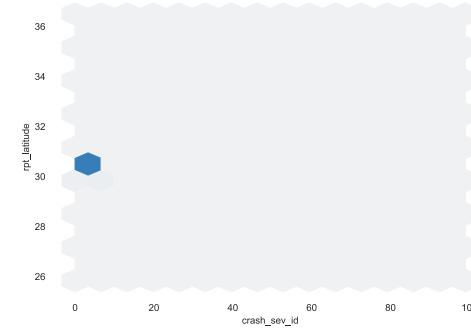
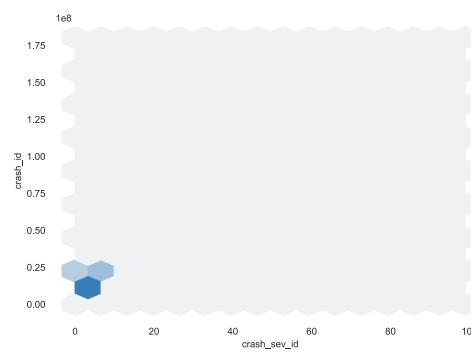


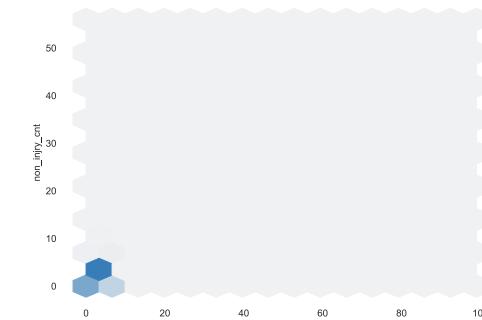
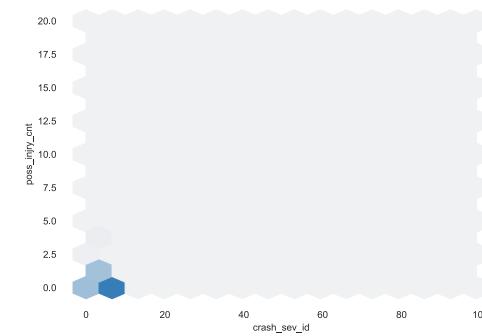
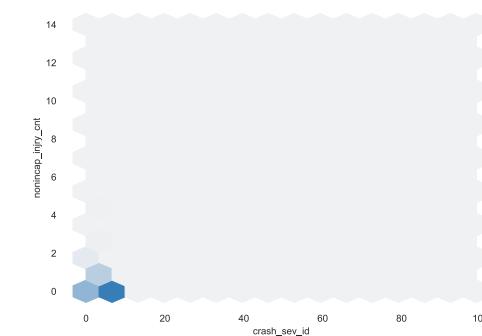
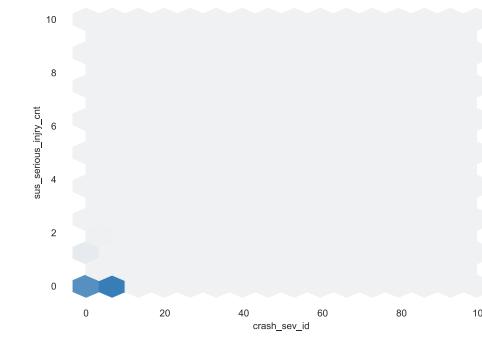
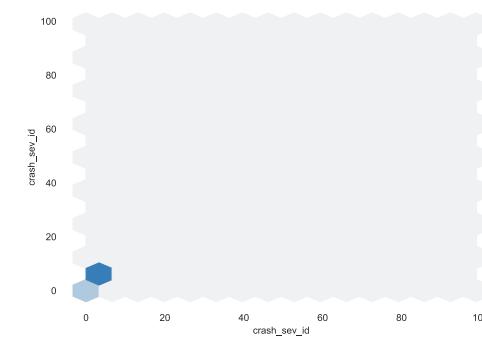
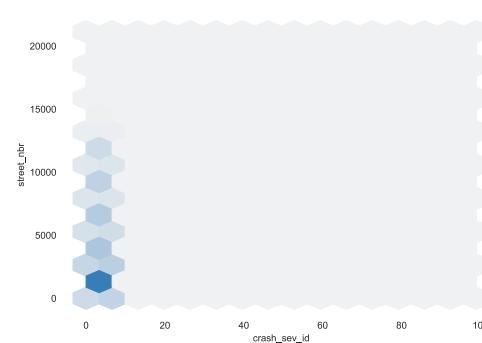


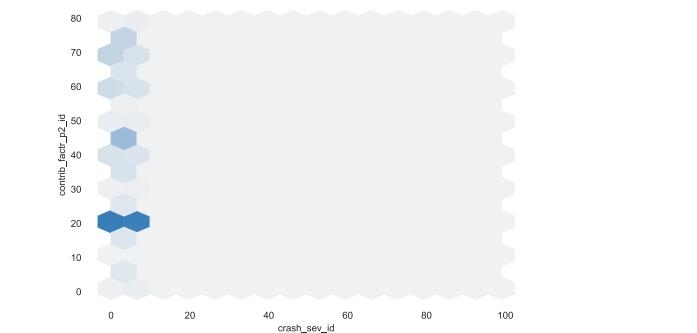
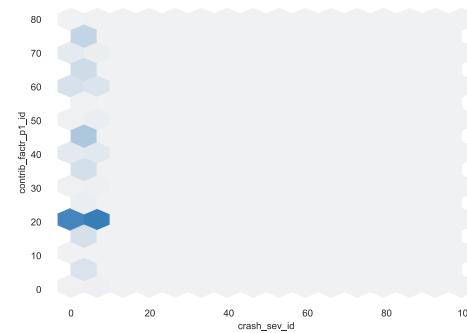
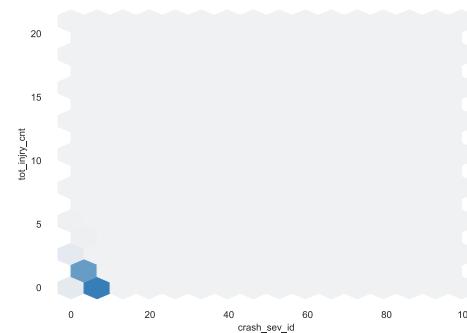
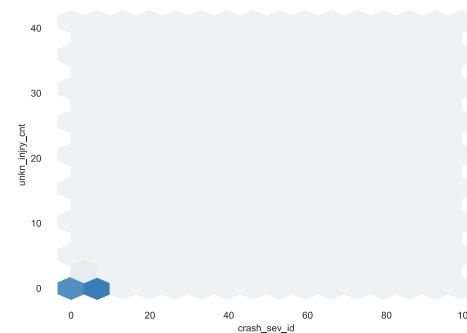




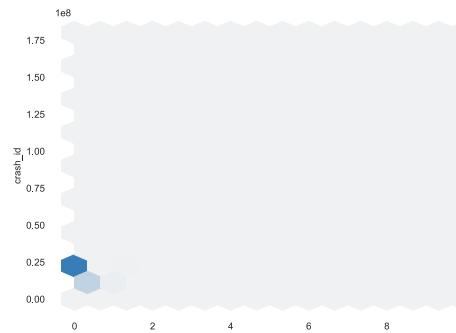
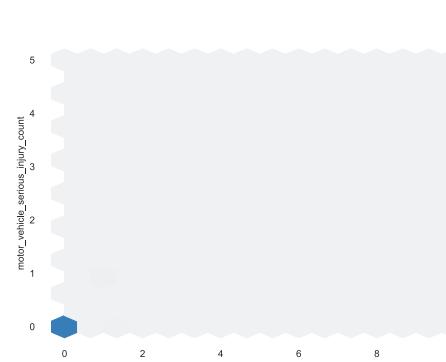


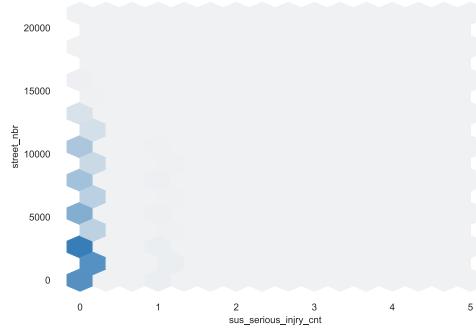
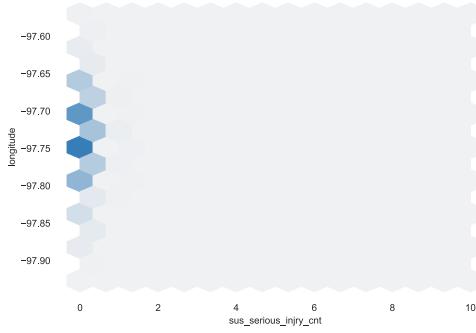
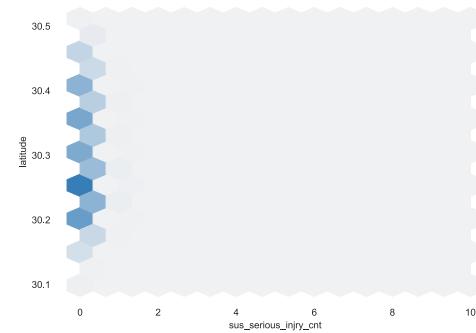
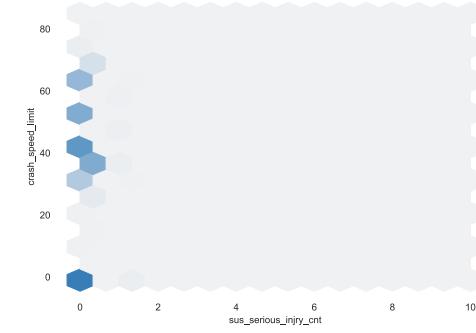
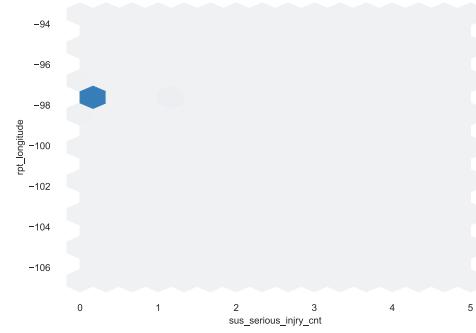
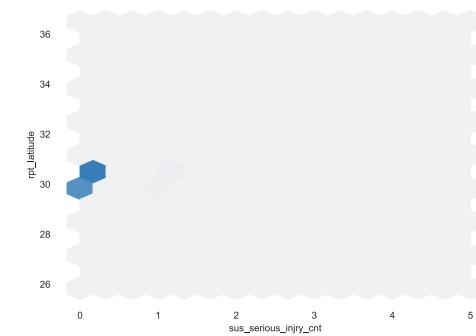


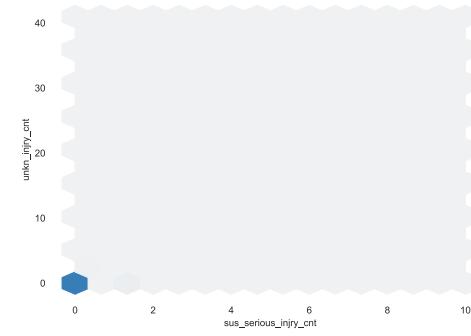
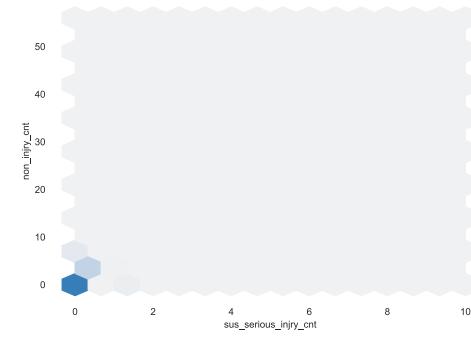
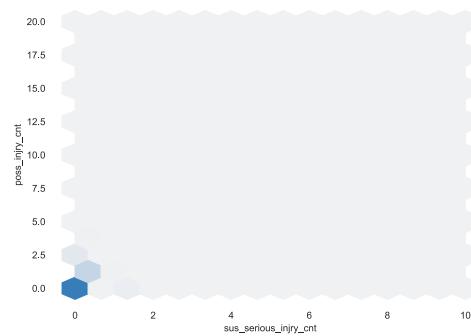
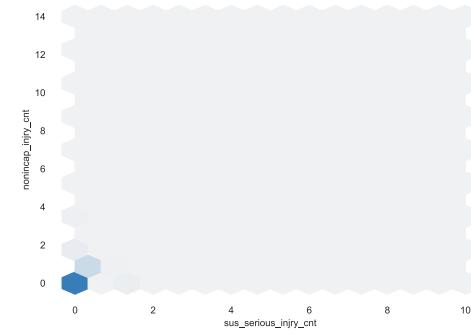
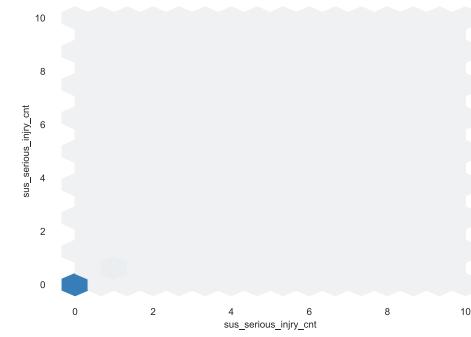
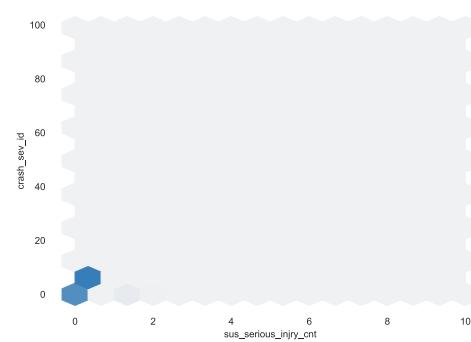


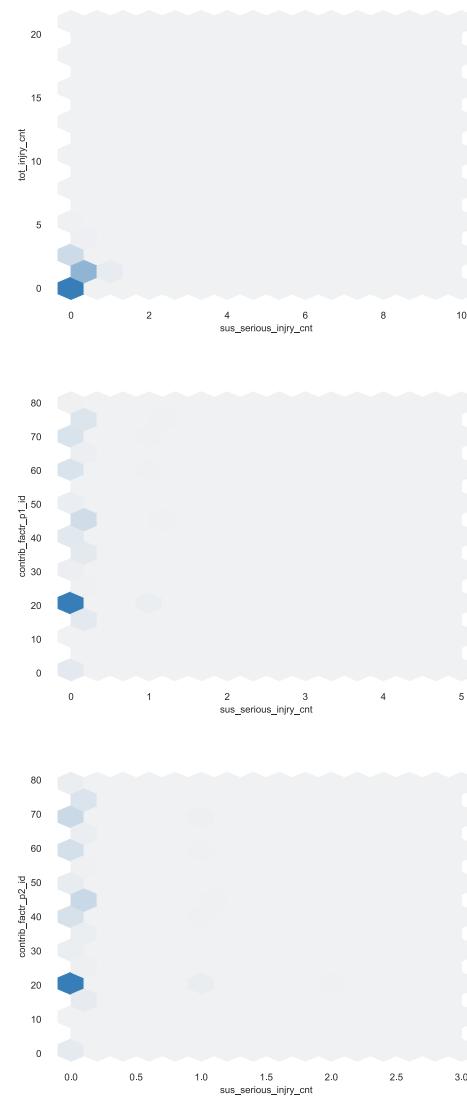


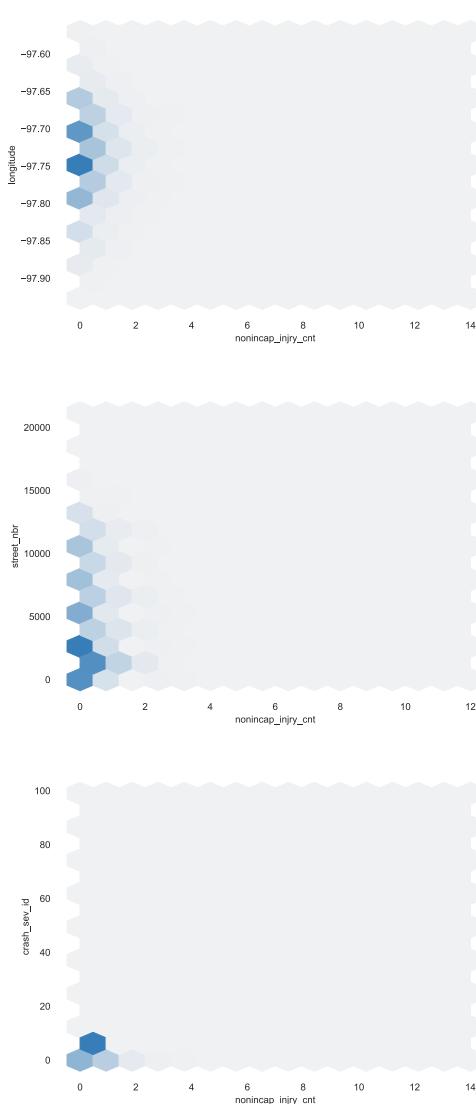
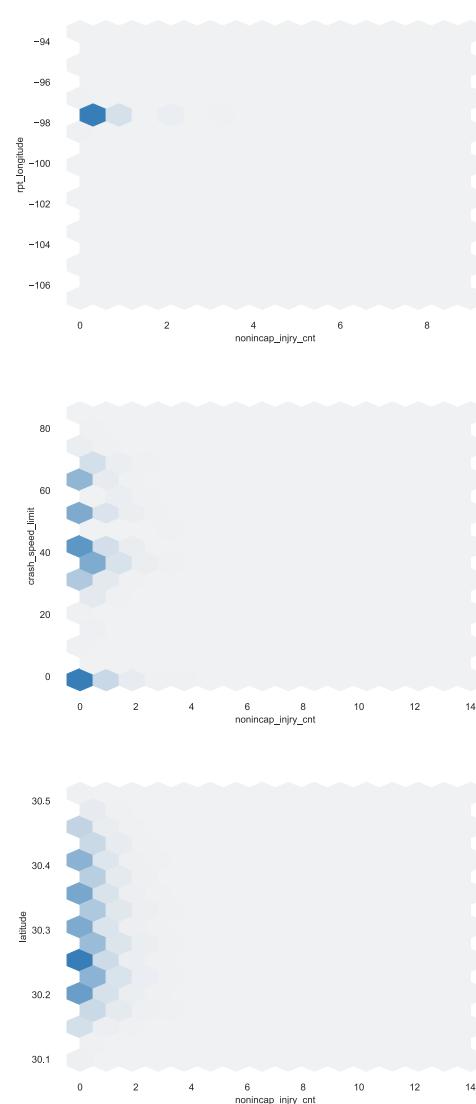
motor_vehicle_serious_inj
crash_id
rpt_latitude
rpt_longitude
crash_speed_limit
latitude
longitude
street_nbr
crash_sev_id
sus_serious_injury_cnt
noninjcap_injury_cnt
poss_injy_cnt
non_injy_cnt
unkn_injy_cnt
tot_injy_cnt
contrib_factr_p1_id
contrib_factr_p2_id

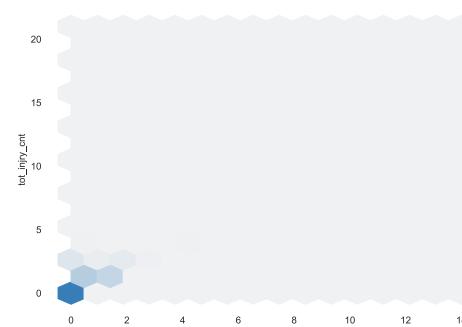
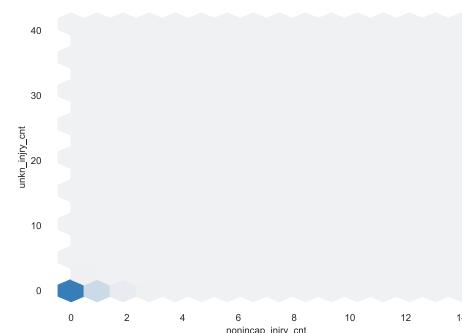
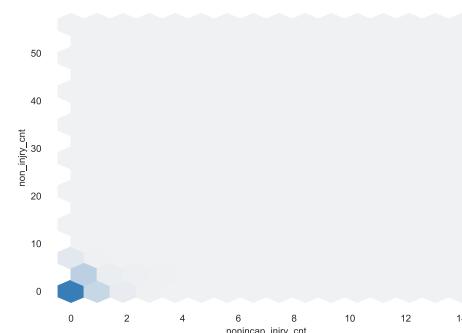
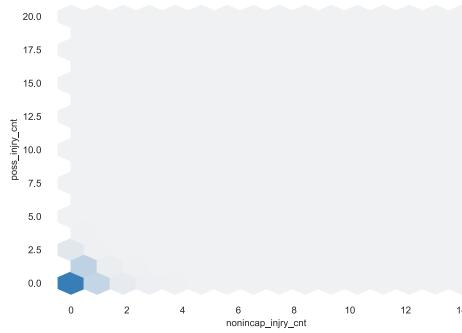
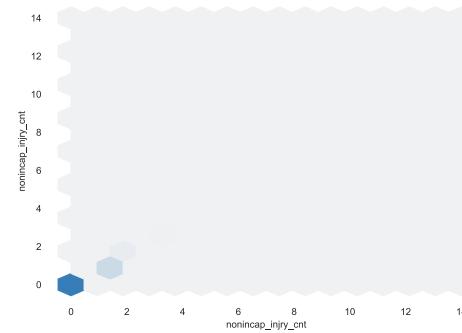
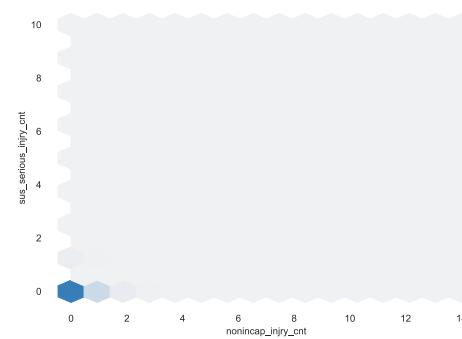


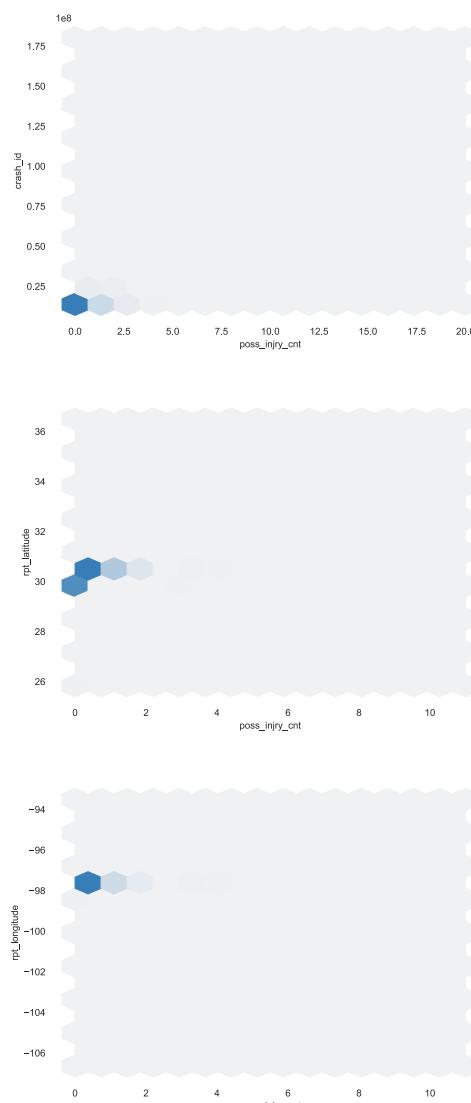
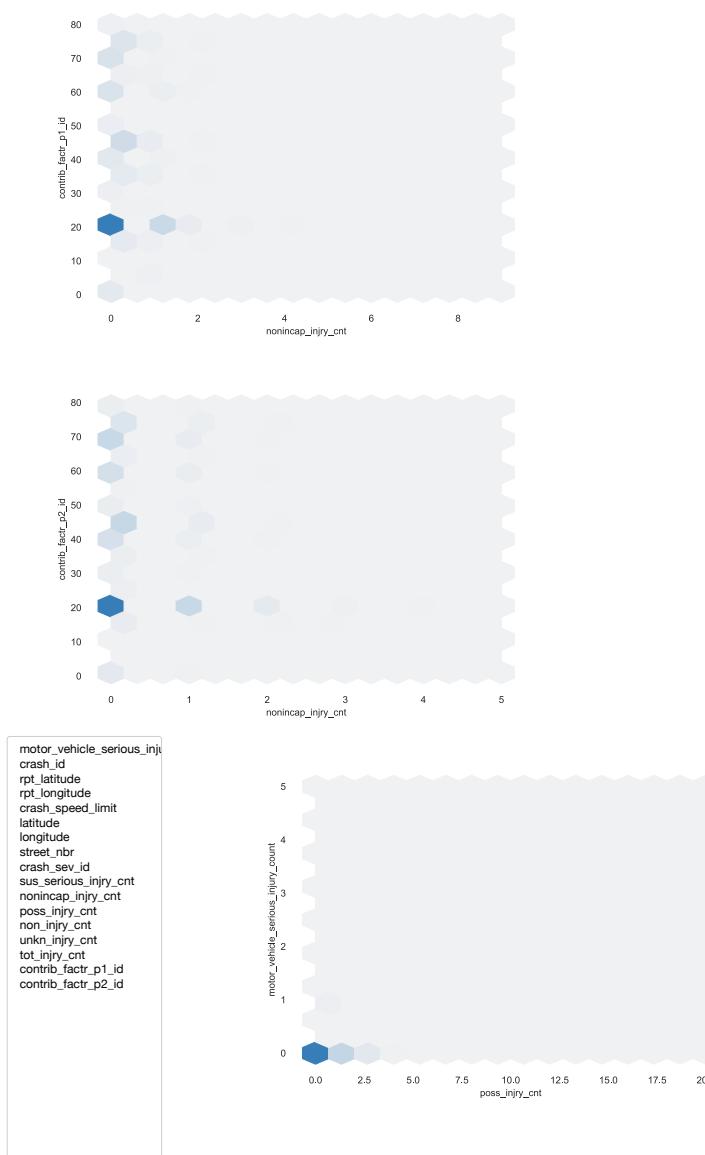


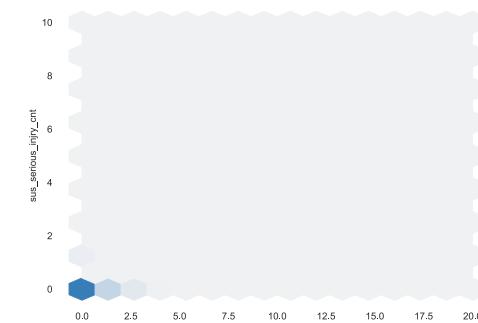
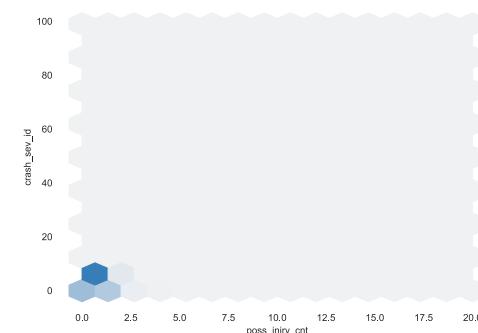
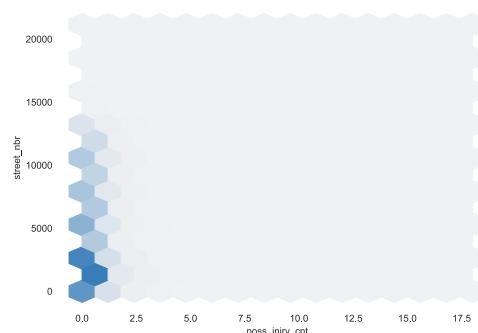
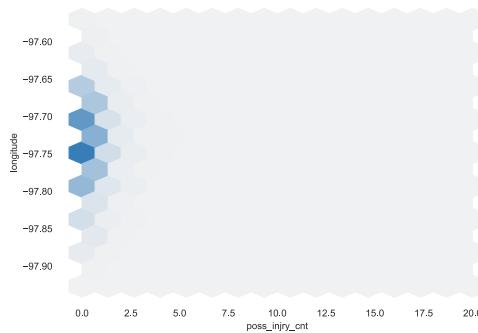
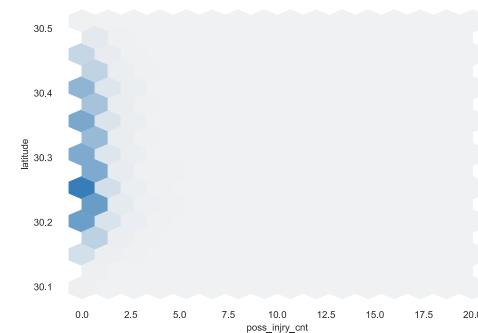
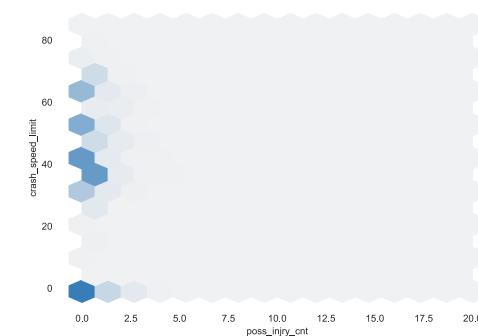


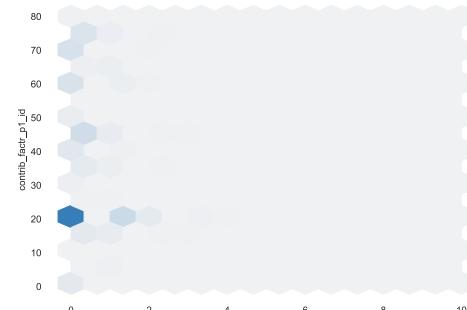
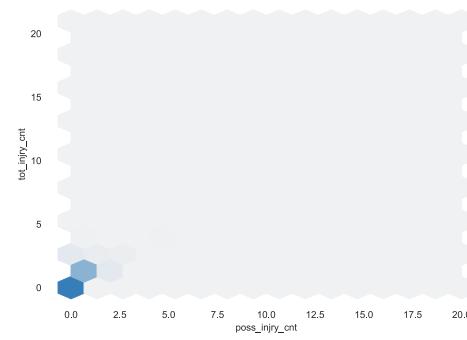
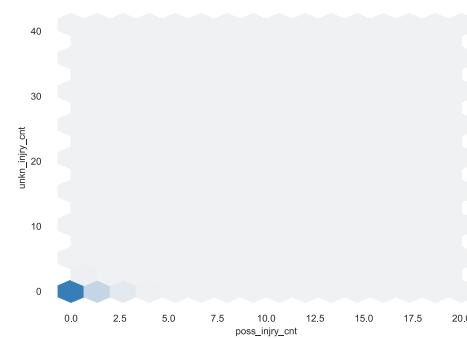
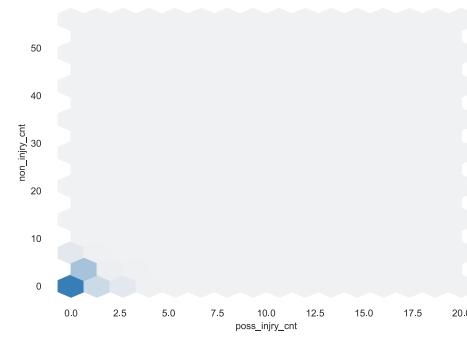
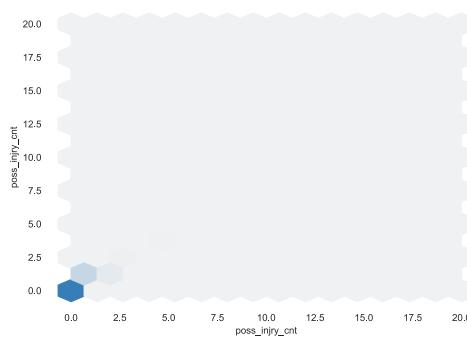
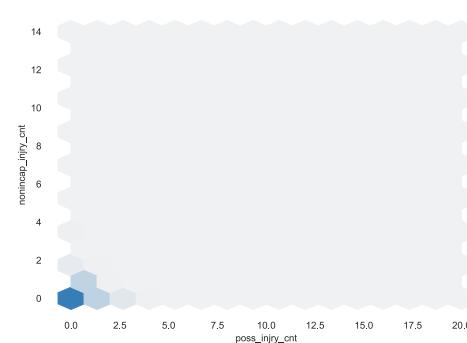


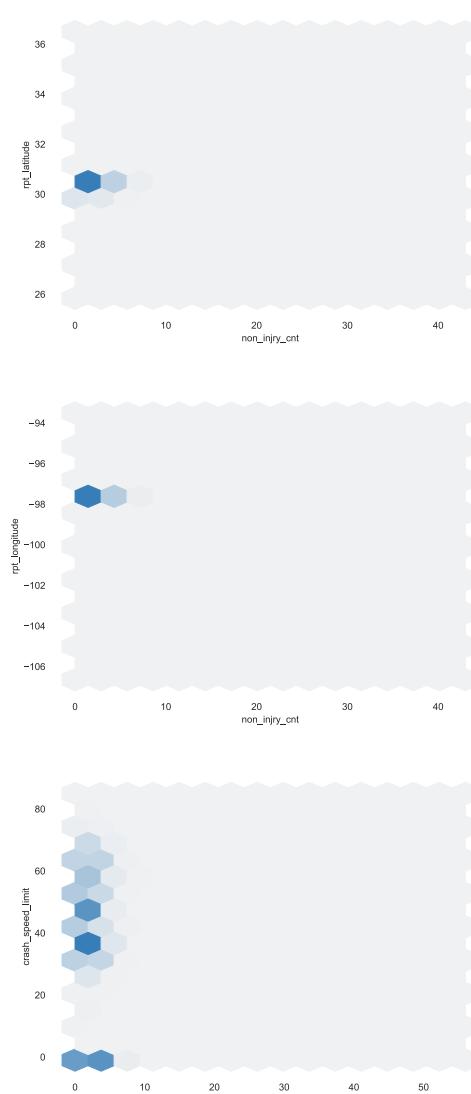


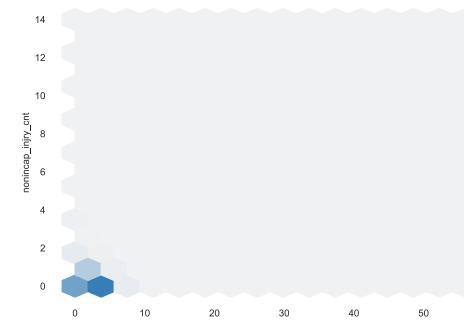
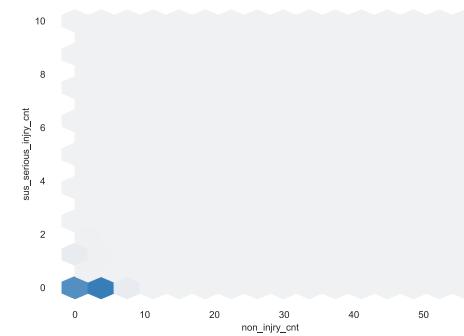
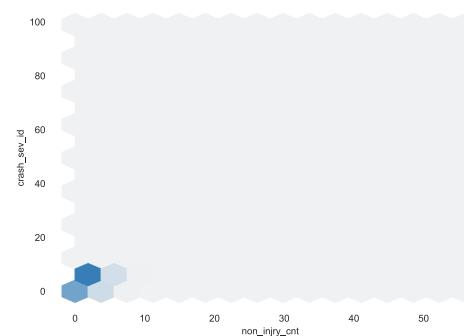
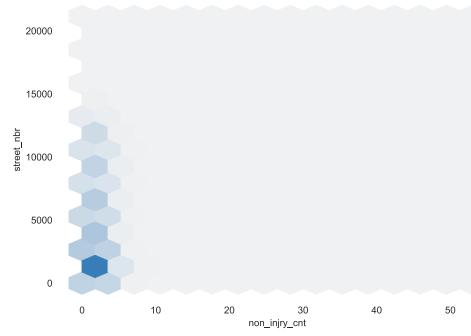
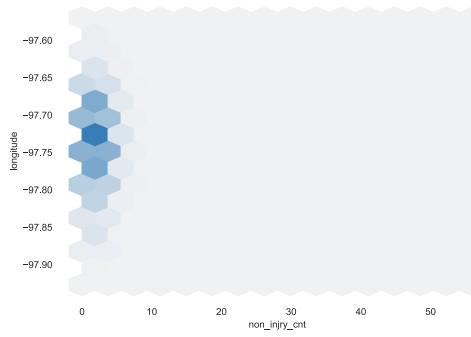
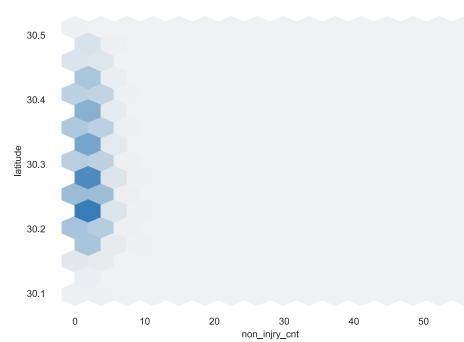


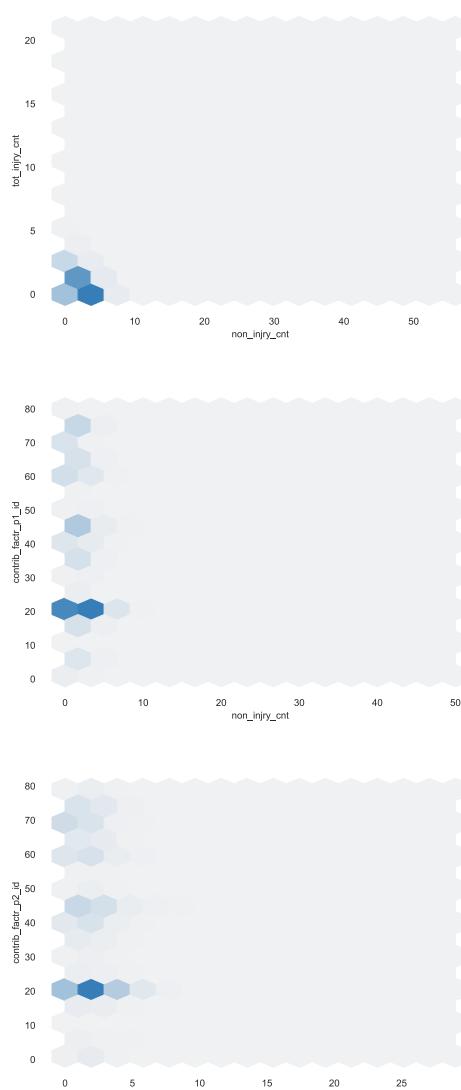
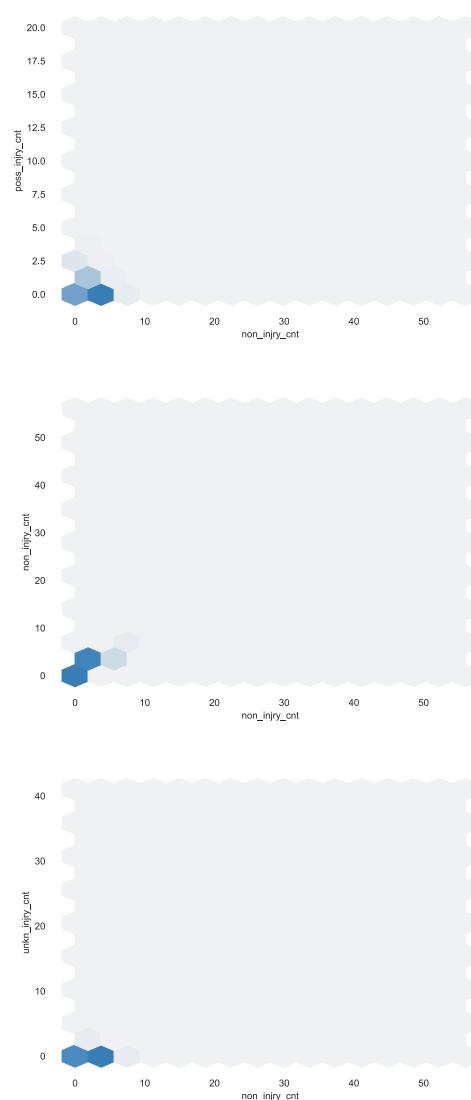


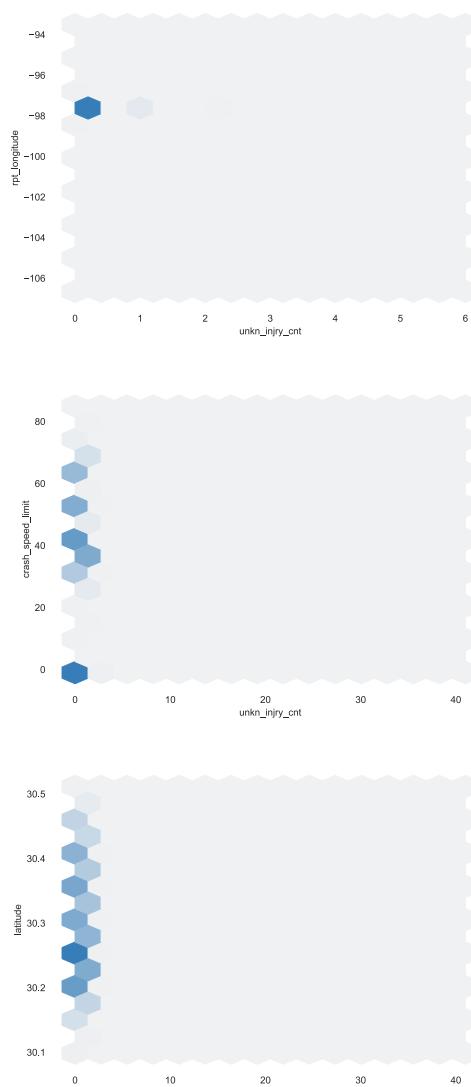
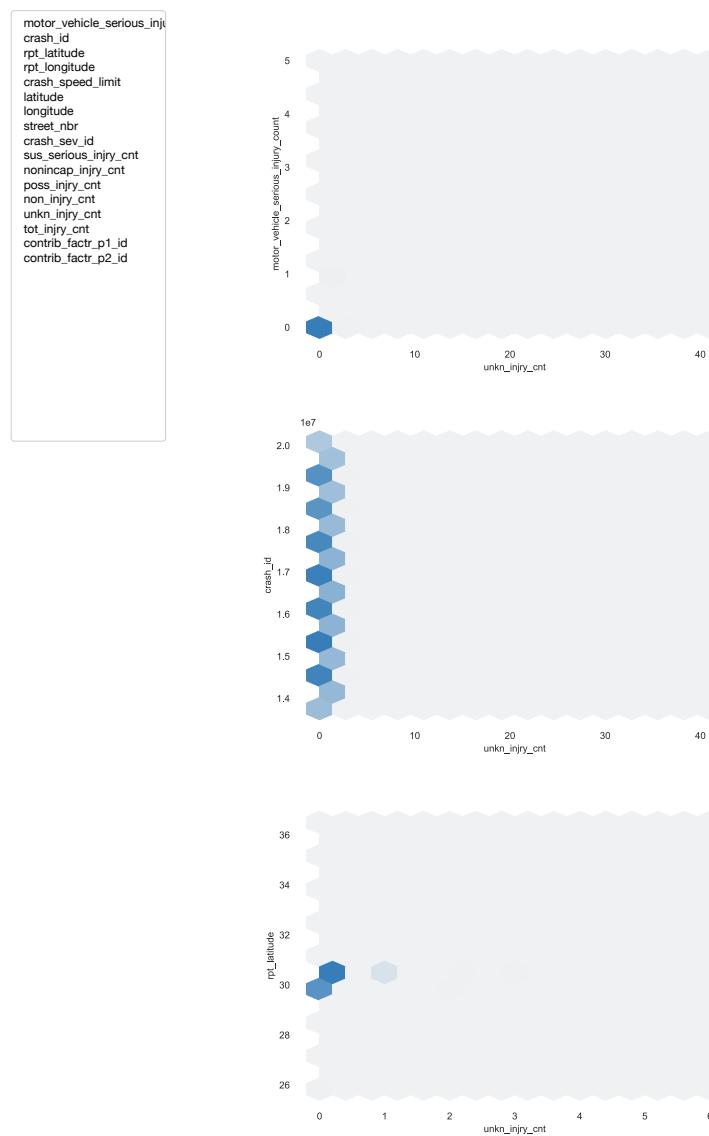


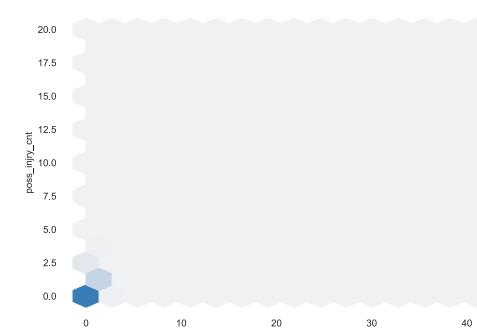
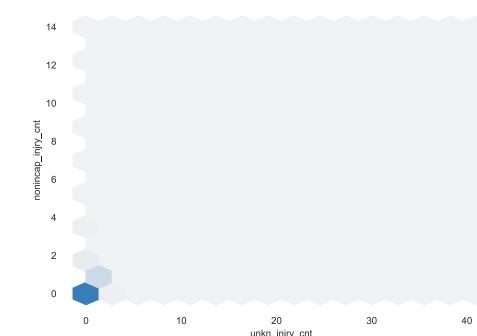
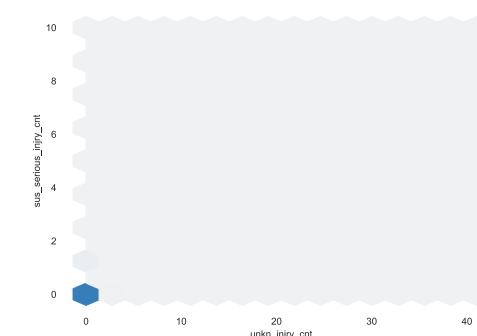
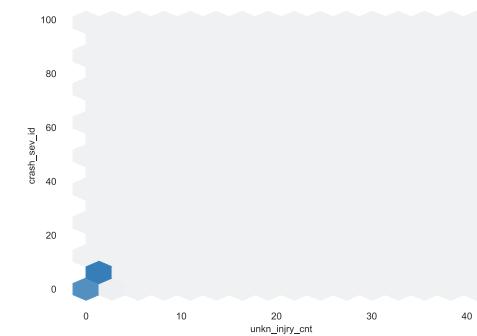
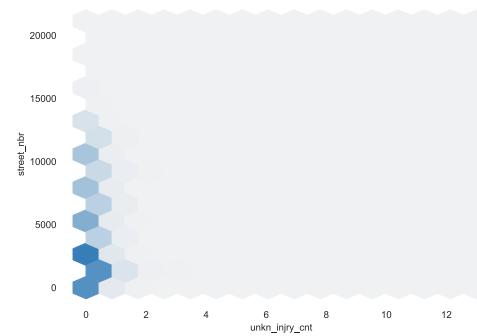
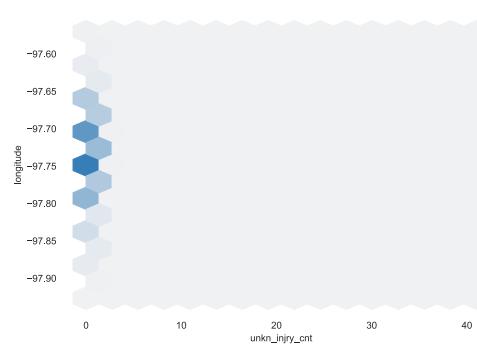


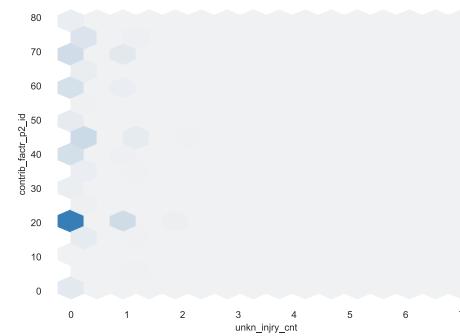
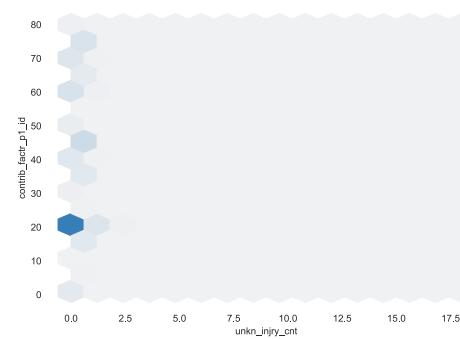
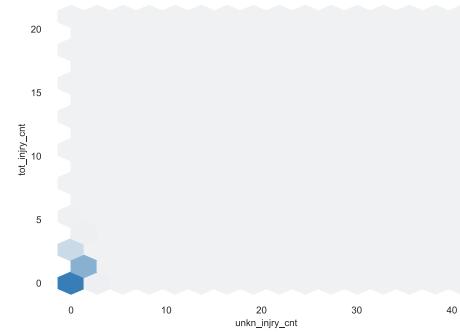
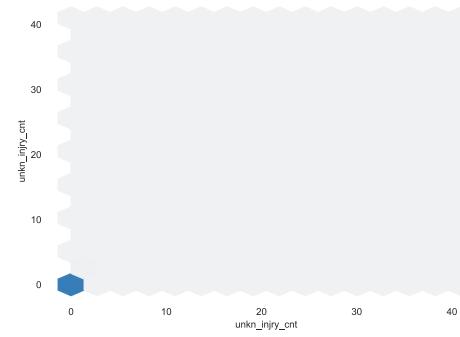
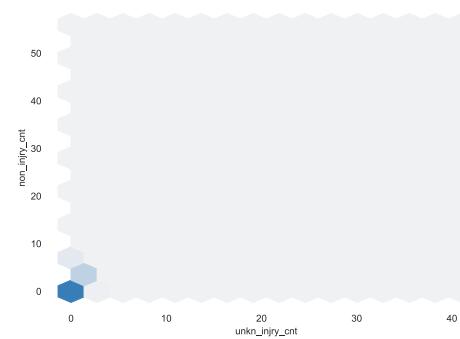




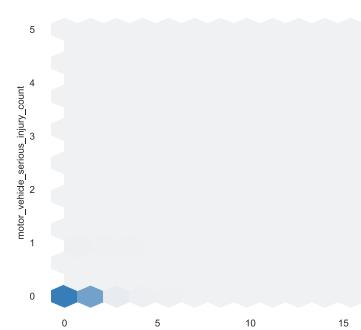


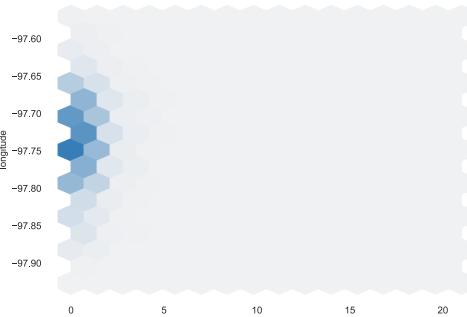
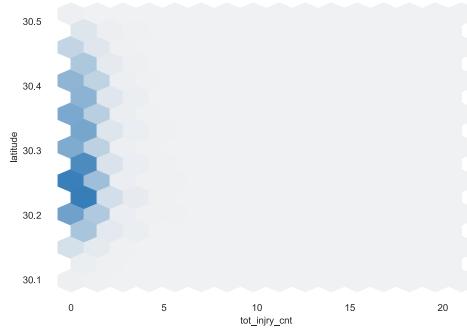
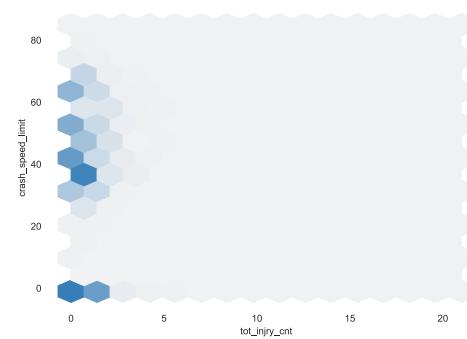
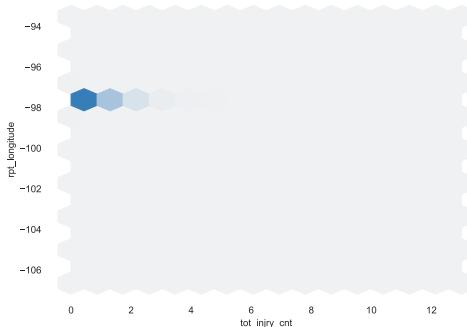
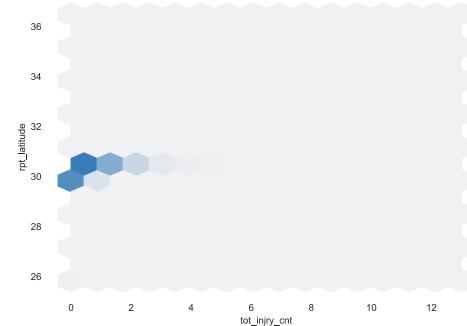
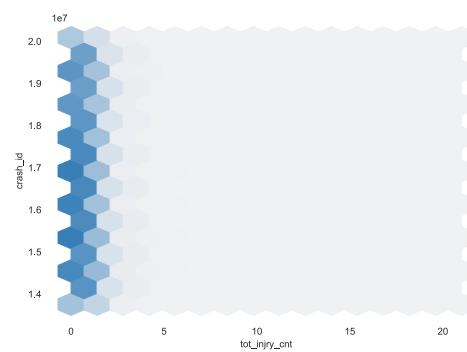


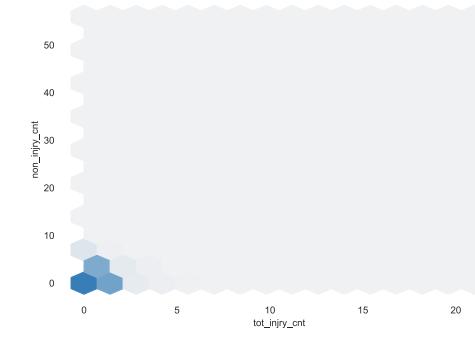
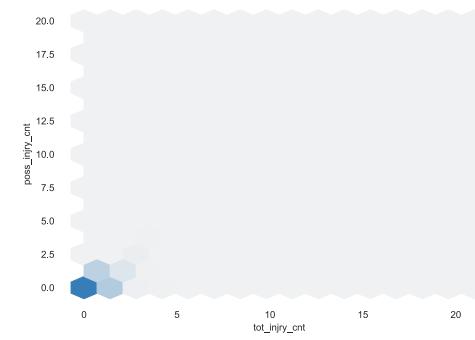
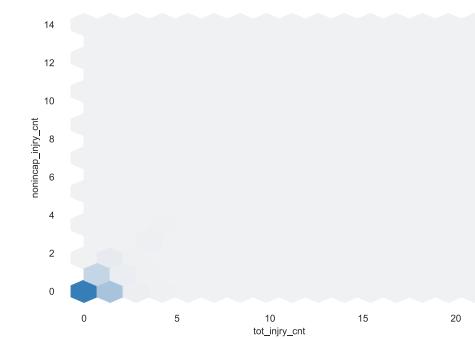
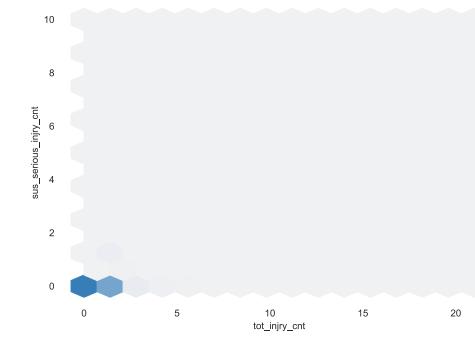
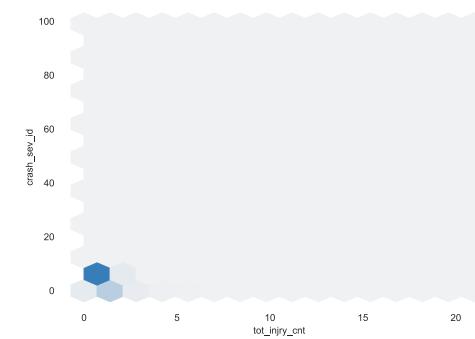
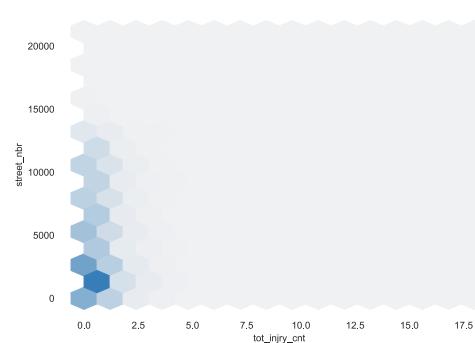


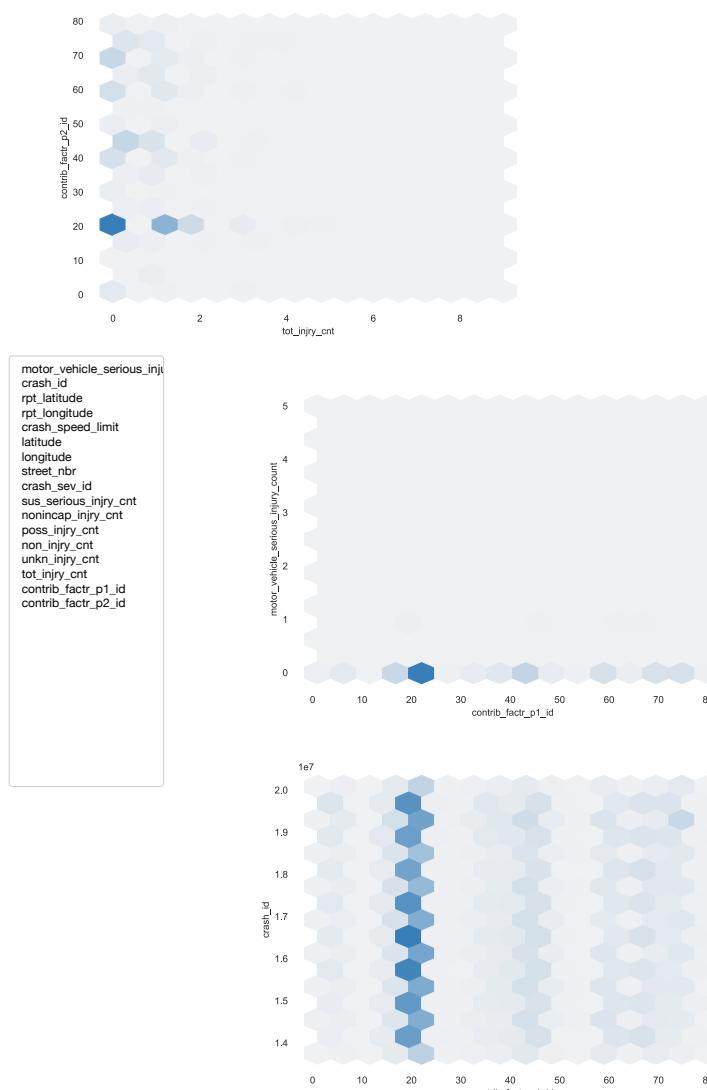
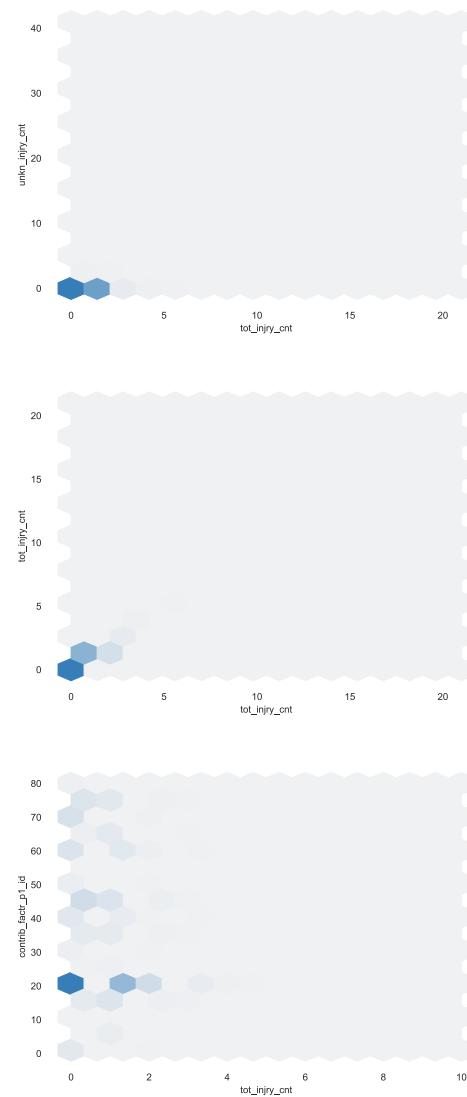


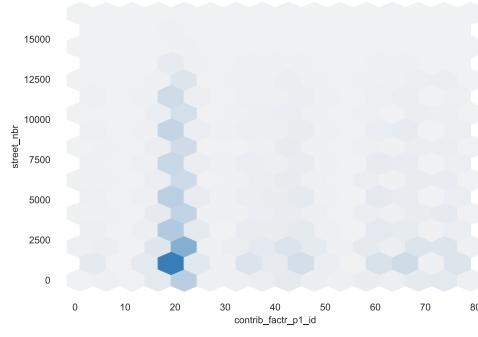
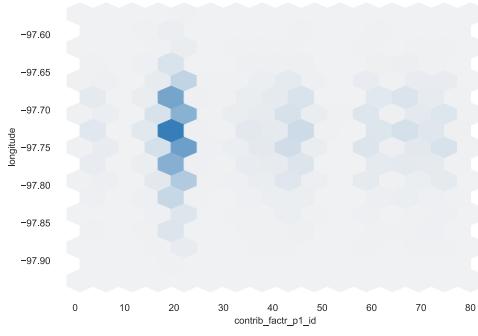
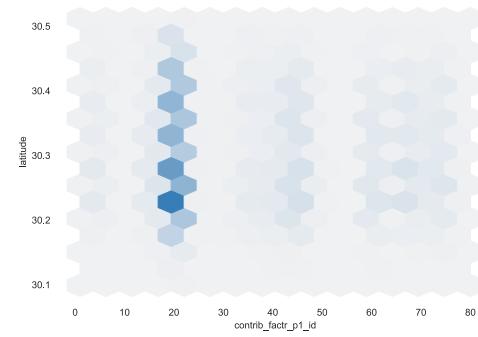
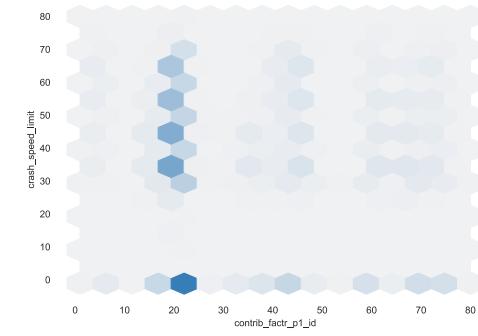
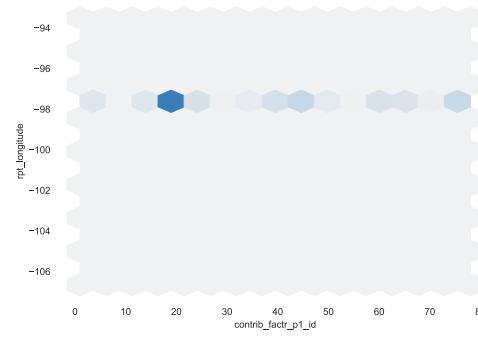
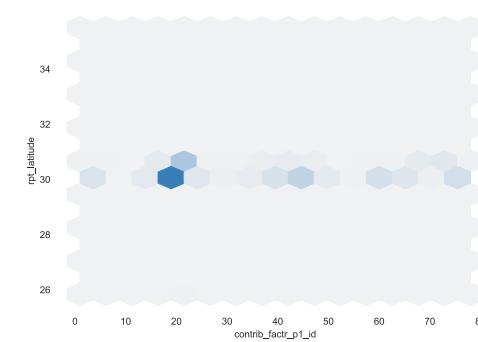
motor_vehicle_serious_inj
crash_id
rpt_latitude
rpt_longitude
crash_speed_limit
latitude
longitude
street_nbr
crash_sev_id
sus_serious_injry_cnt
noninjcap_injry_cnt
poss_injry_cnt
non_injry_cnt
unkn_injry_cnt
tot_injry_cnt
contrib_factr_p1_id
contrib_factr_p2_id

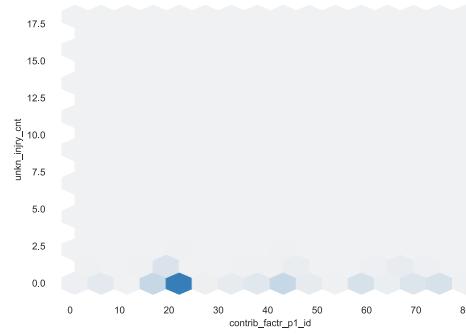
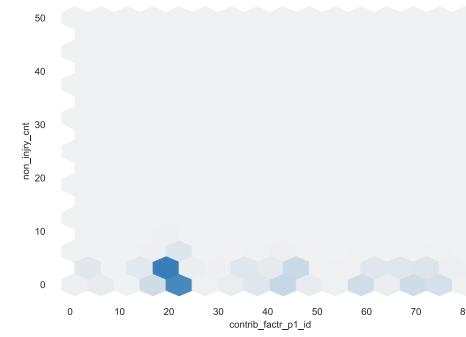
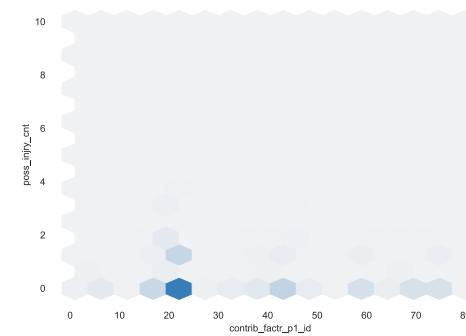
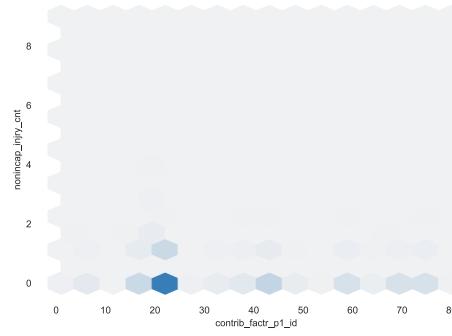
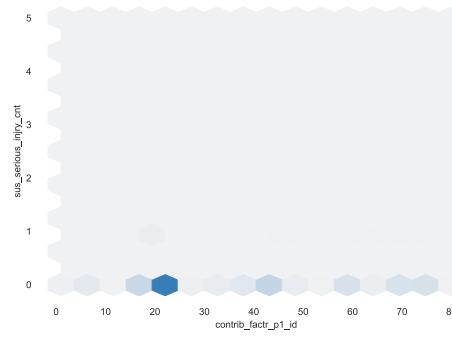
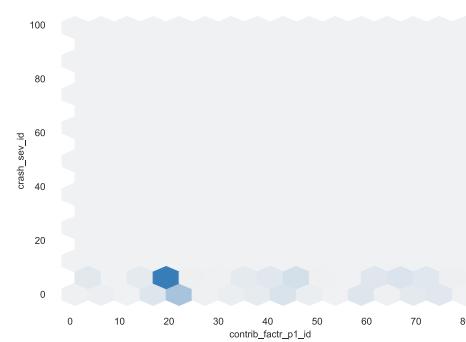


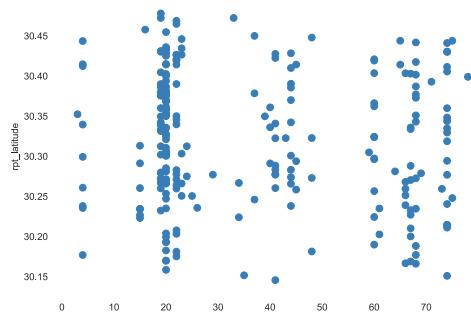
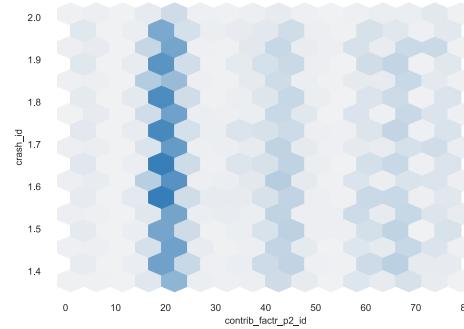
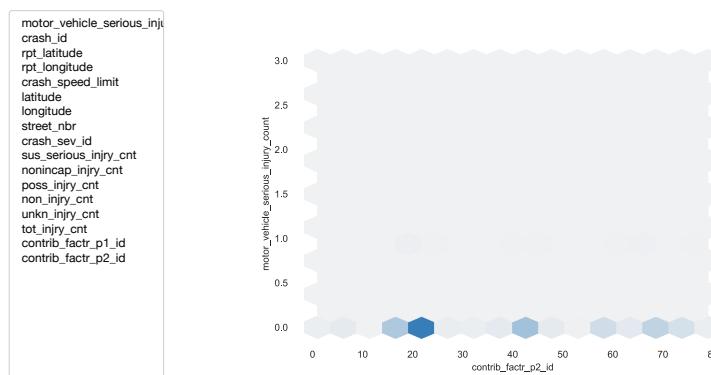
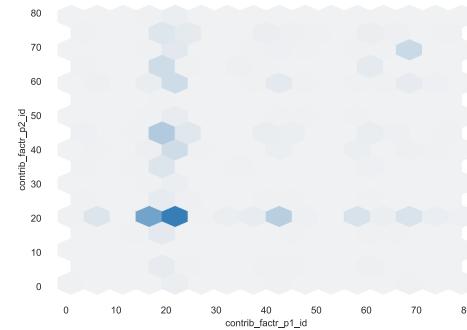
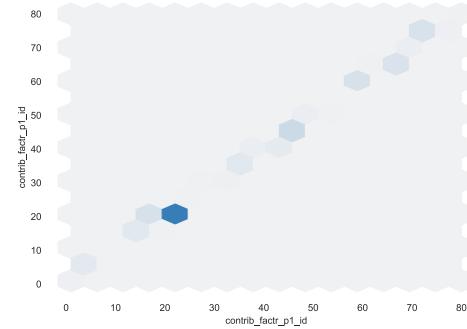
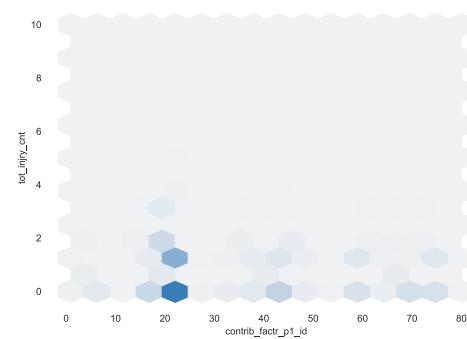


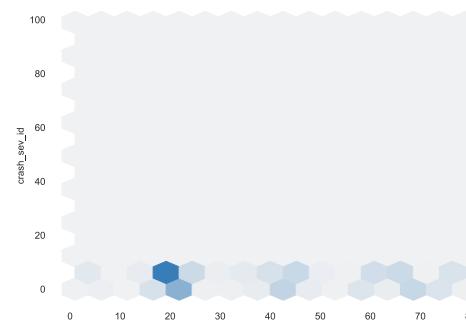
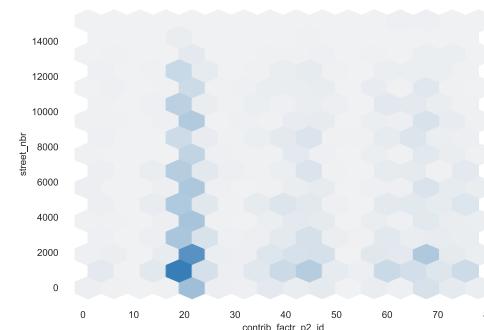
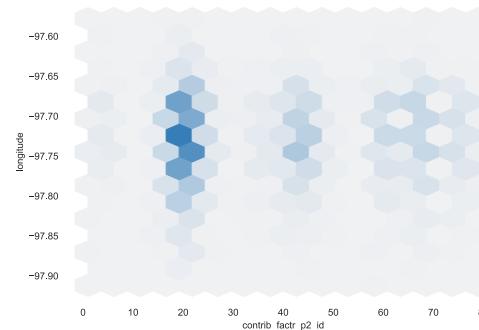
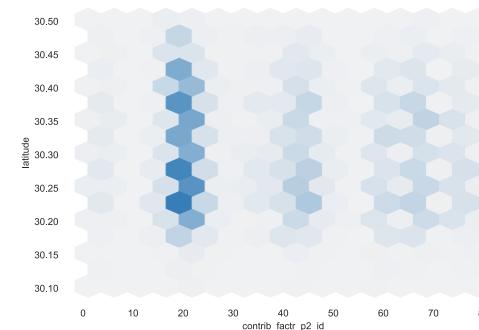
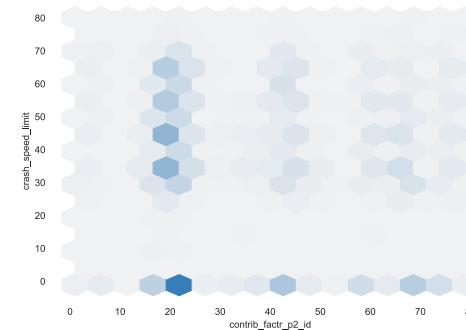
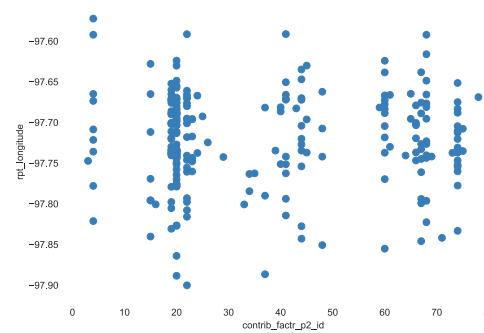


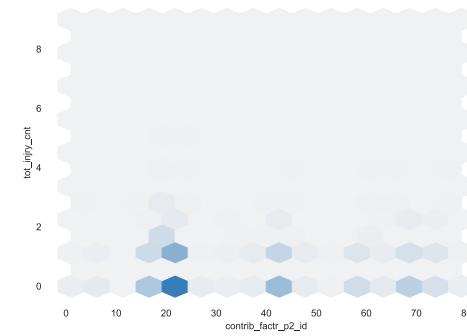
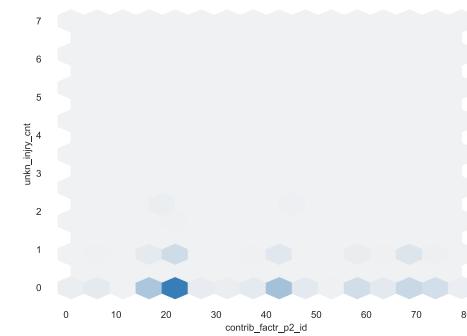
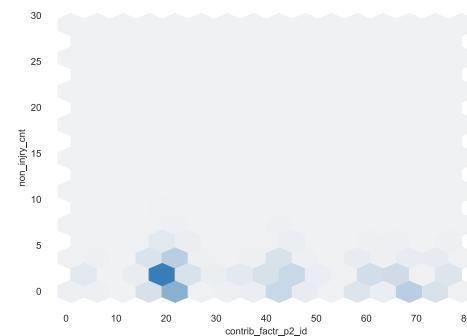
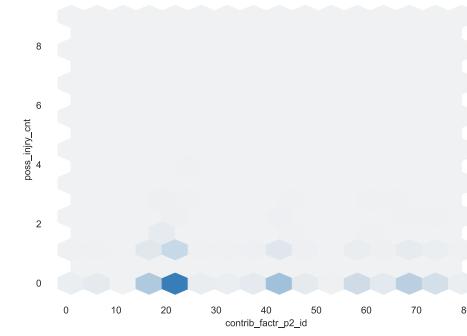
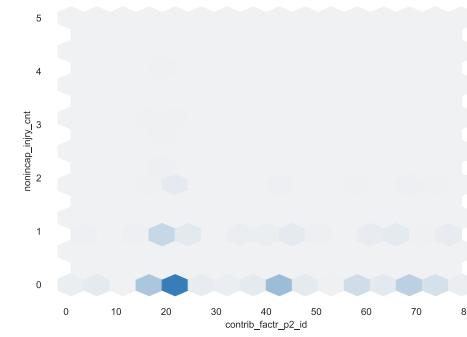
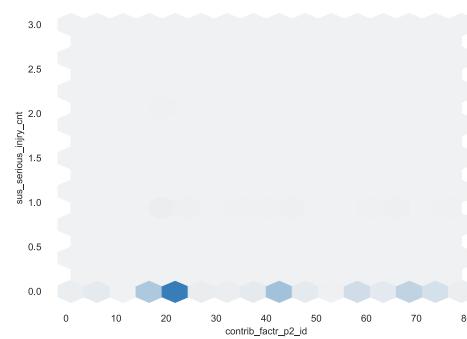


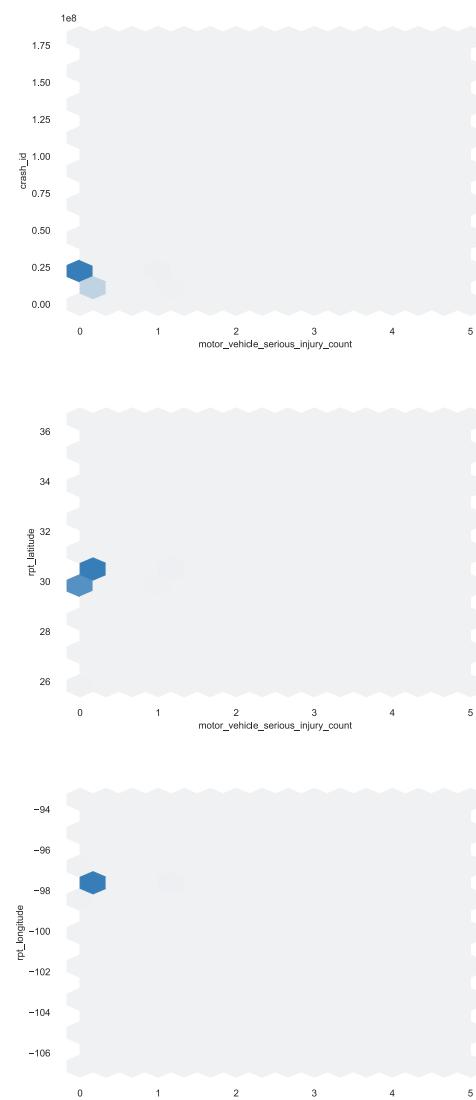
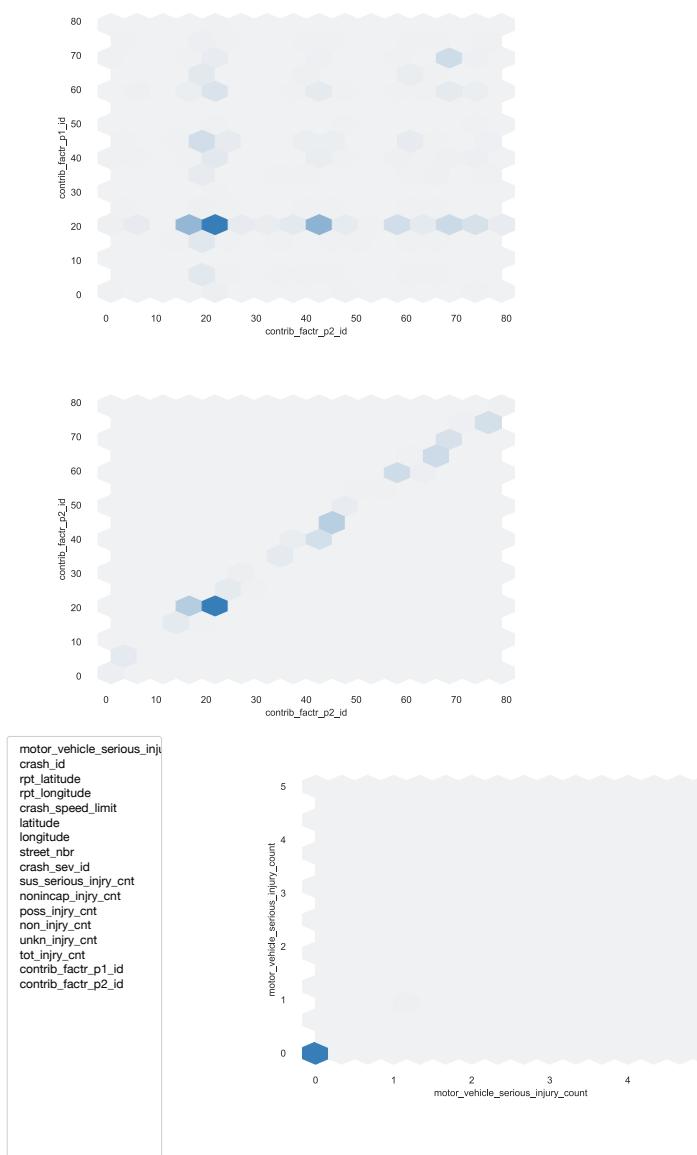


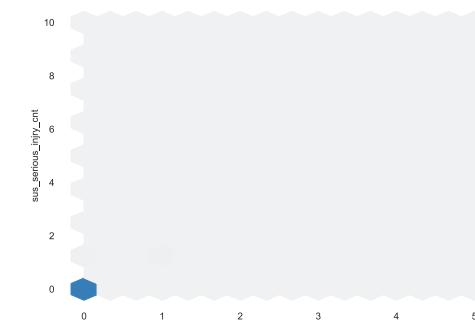
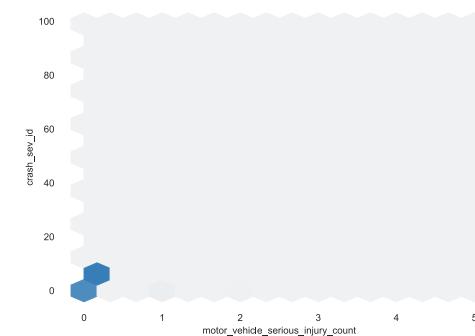
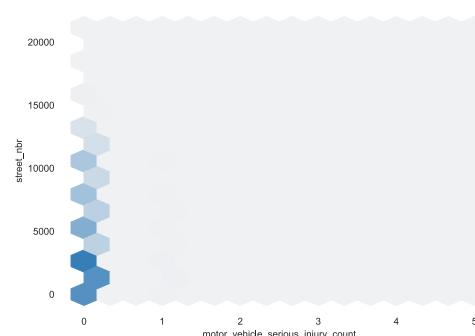
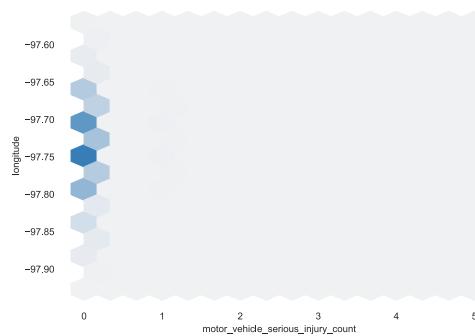
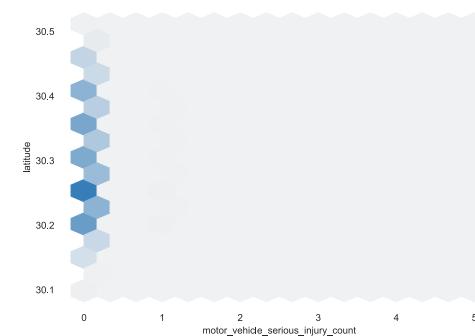
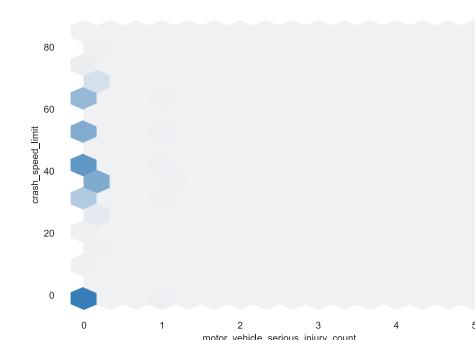


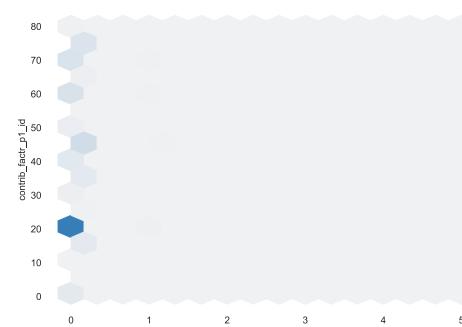
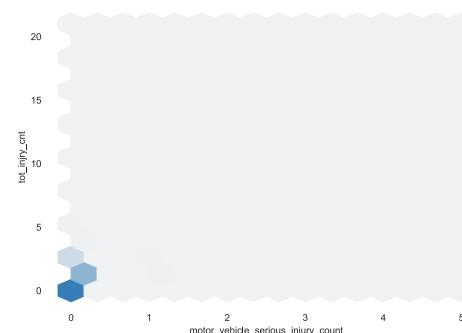
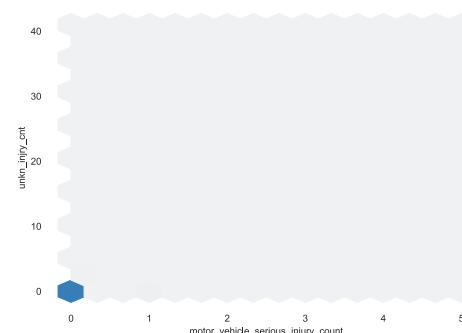
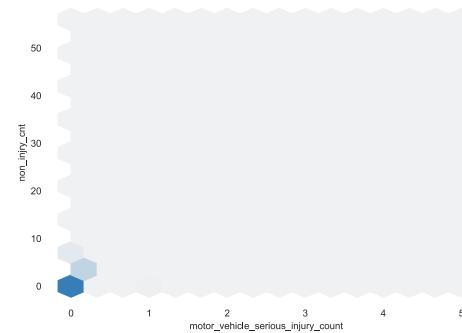
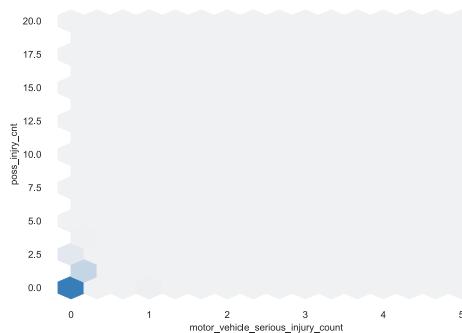
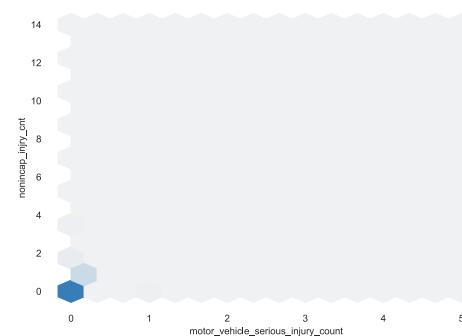


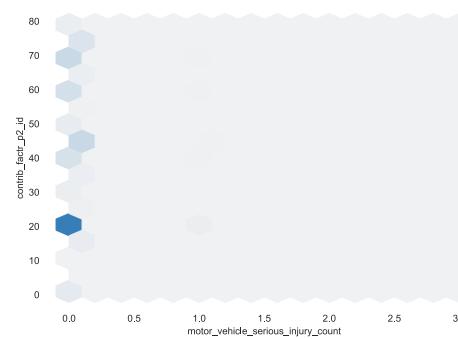




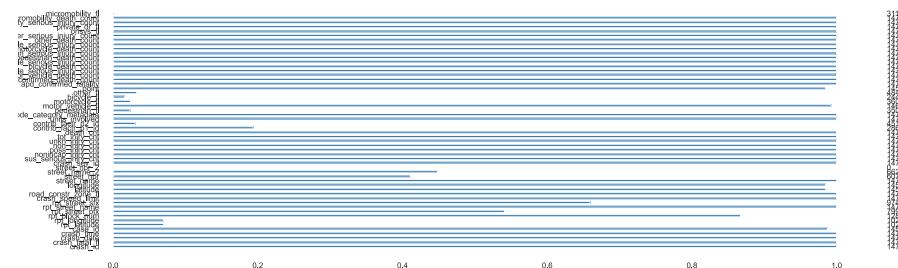




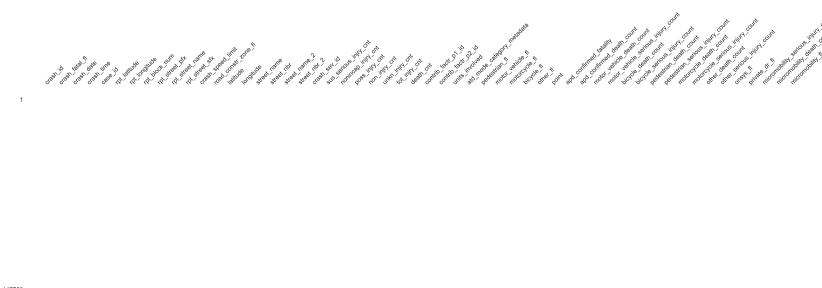




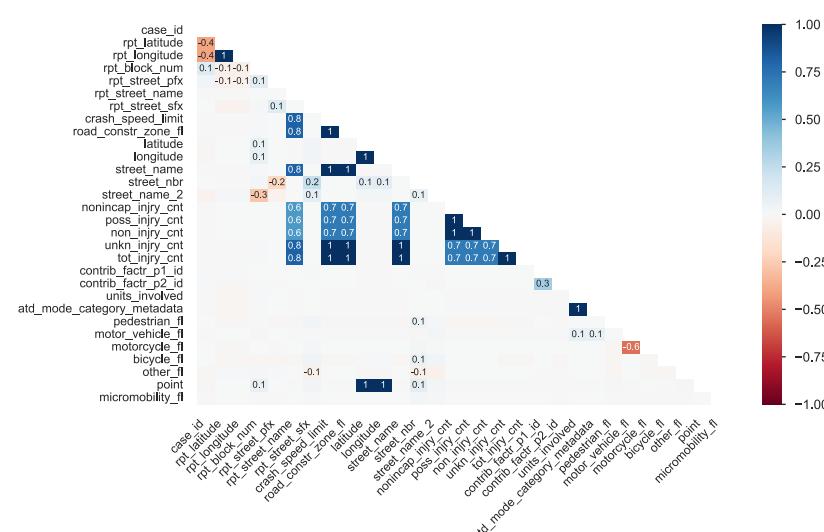
Missing values



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_id	crash_fatal_flg	crash_date	crash_time	case_id	rpt_latitude	rpt_longitude	rpt_block_num	rpt_street_pfx	rpt_street_name
0	13762420	N	03/30/2014	10:58:00 AM	10:58:00	140890874	NaN	NaN	NaN	3707 MANCHAC
1	13777334	N	03/27/2014	01:07:00 PM	13:07:00	140860852	NaN	NaN	3400	PALM WAY TO N
2	13777441	N	03/28/2014	03:42:00 PM	15:42:00	140871196	NaN	NaN	8704	BALCONES CLL
3	13797332	N	04/09/2014	02:09:00 PM	14:09:00	140991015	NaN	NaN	8000	E US 290 HWY 1
4	13795604	N	04/07/2014	06:00:00 PM	18:00:00	140971248	NaN	NaN	200	W BEN WHITE
5	13765070	N	03/31/2014	03:26:00 AM	03:26:00	140900191	NaN	NaN	8700	S IH 35 SVRD
6	13790426	N	04/04/2014	02:34:00 PM	14:34:00	140941160	NaN	NaN	4000	NaN S FM 973 RD
7	13795213	N	04/18/2014	02:06:00 AM	02:06:00	141080164	NaN	NaN	1500	E ANDERSON
8	13786430	N	04/04/2014	12:11:00 AM	00:11:00	140940020	NaN	NaN	3400	N MOPAC
9	13792606	N	04/18/2014	10:05:00 AM	10:05:00	141080445	NaN	NaN	700	S MOPAC

	crash_id	crash_fatal_flg	crash_date	crash_time	case_id	rpt_latitude	rpt_longitude	rpt_block_num	rpt_street_pfx	rpt_street_name
147740	20083275	N	03/07/2024	02:00:00 PM	14:00:00	240670719	NaN	NaN	4900	NaN LOCAL RD
147741	20061720	N	03/06/2024	07:35:00 AM	07:35:00	240660301	NaN	NaN	6300	MOPAC
147742	20048732	N	02/27/2024	05:12:00 PM	17:12:00	240580684	30.43730	-97.66906	NaN	NaN IH 35
147743	20082999	N	03/06/2024	03:32:00 PM	15:32:00	240660755	30.18018	-97.79516	NaN	NaN S 1ST ST
147744	20062080	N	03/06/2024	07:31:00 AM	07:31:00	240660295	30.34600	-97.71000	600	W E ANDERSON
147745	20060069	N	03/05/2024	03:23:00 AM	03:23:00	240650140	NaN	NaN	8800	NaN N IH 35 NI
147746	20056192	N	03/01/2024	09:28:00 PM	21:28:00	240611420	30.34404	-97.71144	7635	N GUADALLA
147747	20083436	N	03/10/2024	12:16:00 AM	00:16:00	240700030	NaN	NaN	3500	NaN MONTOPPI
147748	20049322	N	02/28/2024	01:27:00 AM	01:27:00	240590075	NaN	NaN	3800	NaN SPICEWOOD
147749	20047391	N	02/22/2024	07:57:00 PM	19:57:00	240531383	30.15390	-97.79219	10100	S SIH 35 SV

Report generated by YData (https://ydata.ai/?utm_source=opensource&utm_medium=pandasprofiling&utm_campaign=report).