

## Overview

Dataset statistics		Variable types
Number of variables	48	Text
Number of observations	817723	Boolean
Missing cells	8268003	DateTime
Missing cells (%)	21.1%	Numeric
Duplicate rows	0	Categorical
Duplicate rows (%)	0.0%	
Total size in memory	299.5 MiB	
Average record size in memory	384.0 B	
<b>Alerts</b>		
INJURIES_UNKNOWN	has constant value ""	Constant
TRAFFIC_CONTROL_DEVICE	is highly imbalanced (61.7%)	Imbalance
DEVICE_CONDITION	is highly imbalanced (54.4%)	Imbalance
WEATHER_CONDITION	is highly imbalanced (66.2%)	Imbalance
ALIGNMENT	is highly imbalanced (92.0%)	Imbalance
ROADWAY_SURFACE_COND	is highly imbalanced (55.3%)	Imbalance
ROAD_DEFECT	is highly imbalanced (69.9%)	Imbalance
INTERSECTION RELATED_I	is highly imbalanced (72.4%)	Imbalance
NOT_RIGHT_OF WAY_I	is highly imbalanced (55.8%)	Imbalance
HIT_AND_RUN_I	is highly imbalanced (74.4%)	Imbalance
SEC_CONTRIBUTORY_CAUSE	is highly imbalanced (54.2%)	Imbalance
MOST_SEVERE_INJURY	is highly imbalanced (66.7%)	Imbalance
INJURIES_FATAL	is highly imbalanced (99.4%)	Imbalance
CRASH_DATE_EST_I	has 756594 (92.3%) missing values	Missing
LANE_CNT	has 618714 (75.7%) missing values	Missing
REPORT_TYPE	has 24314 (3.0%) missing values	Missing
INTERSECTION RELATED_I	has 630174 (77.1%) missing values	Missing
NOT_RIGHT_OF WAY_I	has 780015 (95.4%) missing values	Missing
HIT_AND_RUN_I	has 561774 (68.7%) missing values	Missing
PHOTOS_TAKEN_I	has 806948 (98.7%) missing values	Missing
STATEMENTS_TAKEN_I	has 799465 (97.8%) missing values	Missing
DOORING_I	has 815211 (99.7%) missing values	Missing
WORK_ZONE_I	has 813053 (99.4%) missing values	Missing
WORK_ZONE_TYPE	has 814105 (99.6%) missing values	Missing
WORKERS_PRESENT_I	has 816529 (99.9%) missing values	Missing
LANE_CNT	is highly skewed ( $\gamma_1 = 350.3035967$ )	Skewed
LATITUDE	is highly skewed ( $\gamma_1 = -116.2686453$ )	Skewed
LONGITUDE	is highly skewed ( $\gamma_1 = 127.289911$ )	Skewed

CRASH_RECORD_ID	has unique values	Unique
INJURIES_TOTAL	has 703431 (86.0%) zeros	Zeros
INJURIES_INCAPACITATING	has 801987 (98.1%) zeros	Zeros
INJURIES_NON_INCAPACITATING	has 749965 (91.7%) zeros	Zeros
INJURIES_REPORTED_NOT_EVIDENT	has 777695 (95.1%) zeros	Zeros
INJURIES_NO_INDICATION	has 17119 (2.1%) zeros	Zeros
CRASH_HOUR	has 17728 (2.2%) zeros	Zeros

## Variables

Select Columns 

CRASH_RECORD_ID	
Text	
UNIQUE	
<b>Distinct</b>	817723
<b>Distinct (%)</b>	100.0%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%

589#2d12eab3c91575a0f711d0f8078646138670580011798705765dc3c87f00b19afex3c187f205188401f7x2a93a3e3a2f11050  
004#d144030319113aa0d92cf7341962a857212b13781940fawee153d0d44804461f751239015812f97217ed20201c3c89412e2  
441f7724129be17a13727804607482212010b02f721774401a78175811647faff7a76153a94e7ed21c7ab309380707e0403913  
1ee2180ab01c82075165656275105bc3499300f7caaeewee5c95dfe15d011a2c204831ff19be0ab679f13100044151cf1a

Length	Characters and Unicode	Unique	Sample		
<b>Max length</b>	128	Total 104668544 characters	Unique	817723	?
<b>Median length</b>	128		Unique (%)	100.0%	
<b>Mean length</b>	128	Distinct 16 characters			1st row
<b>Min length</b>	128				5f54a59fc087b12ae5b1acff96a3caf4f42d37e79fbdb4106558b34b8a6d2b81af02cf1b576ecd7ced08ffd10fcfd940a847f613125b89d:
		Distinct 1 characters			2nd row
					?
					3rd row
					61fc8bc18eb522a6469b460e2134df3d15fb2e81fd93e9acf3dc7e631b9e1ba8h450a63af12bd90d1d2d9b127ea87f88d32e138a4eeb

Distinct scripts	1	4th row	004cd14d0303a9163aad69a2d7f341b7da2a8572b2ab3378594bfae8ac53dc604dd8d414f93c290b55862f9f2517ad32e6209cbc8034c2e26eb3c2bc9724390
Distinct blocks	1	5th row	(https://en.wikipedia.org/wiki/Scrow)

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

Value	Count	Frequency (%)
6c1659069e9c6285a650e70d6f9b574ed5f64c...	1	< 0.1%
569af2dc0ea9c5a55a01ff1d5fd678644c3dd76...	1	< 0.1%
4a1f7a24129e5e1d4a7a2fd44ab6f8822a20bcd...	1	< 0.1%
57cfcd70be67a9d66fd973426036aa20034e0f...	1	< 0.1%
1ee2180a89cc02c0b756f955b2755bb5cc9d9...	1	< 0.1%
61fc8c1eb522a6469b460e2134df3d15f82e81...	1	< 0.1%
004cd14d0303a9163aad69a2d7f341b7da2a85...	1	< 0.1%
a1d5f0ea90897745365a4ccb06cc60329a120d...	1	< 0.1%
b236c77d59e32b7b469a6e2f17f438b7457e1b...	1	< 0.1%
35156ce97cab22747495e92e8bbb16c57e0e60...	1	< 0.1%
Other values (817713)	817713	> 99.9%

#### Most occurring characters

Value	Count	Frequency (%)
2	6547788	6.3%
b	6545782	6.3%
d	6545680	6.3%
f	6545495	6.3%
c	6544094	6.3%
9	6542979	6.3%
0	6542034	6.3%
4	6541970	6.3%
8	6541267	6.2%
6	6540678	6.2%
Other values (6)	39230777	37.5%

#### Most occurring categories

Value	Count	Frequency (%)
(unknown)	104668544	100.0%

#### Most frequent character per category

(unknown)

Value	Count	Frequency (%)
2	6547788	6.3%
b	6545782	6.3%
d	6545680	6.3%
f	6545495	6.3%
c	6544094	6.3%
9	6542979	6.3%
0	6542034	6.3%
4	6541970	6.3%
8	6541267	6.2%
6	6540678	6.2%
Other values (6)	39230777	37.5%

#### Most occurring scripts

Value	Count	Frequency (%)
(unknown)	104668544	100.0%

#### Most frequent character per script

(unknown)

Value	Count	Frequency (%)
2	6547788	6.3%
b	6545782	6.3%
d	6545680	6.3%
f	6545495	6.3%
c	6544094	6.3%
9	6542979	6.3%
0	6542034	6.3%
4	6541970	6.3%
8	6541267	6.2%
6	6540678	6.2%
Other values (6)	39230777	37.5%

#### Most occurring blocks

Value	Count	Frequency (%)
(unknown)	104668544	100.0%

#### Most frequent character per block

(unknown)		Count	Frequency (%)
Value			
2		6547788	6.3%
b		6545782	6.3%
d		6545680	6.3%
f		6545495	6.3%
c		6544094	6.3%
9		6542979	6.3%
0		6542034	6.3%
4		6541970	6.3%
8		6541267	6.2%
6		6540678	6.2%
Other values (6)		39230777	37.5%

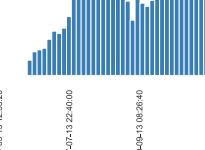
CRASH_DATE_EST_I		Count	Frequency (%)
<b>Boolean</b>			
<b>MISSING</b>			
<b>Distinct</b>	2		
<b>Distinct (%)</b>	< 0.1%		
<b>Missing</b>	756594		
<b>Missing (%)</b>	92.5%		
<b>Memory size</b>	1.6 MiB		
Value		Count	Frequency (%)
True		53267	6.5%
False		7862	1.0%
(Missing)		756594	92.5%



	True
	False

**CRASH\_DATE**  
 Date

Distinct	536888
Distinct (%)	65.7%
Missing	0
Missing (%)	0.0%
Memory size	6.2 MiB

 Minimum 2013-03-03 16:48:00  
 Maximum 2024-03-26 01:40:00


Frequency

0

2012-10-11 20:00:00

2014-05-13 12:53:20

2015-12-13 04:46:40

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

2023-11-14 17:13:20

2014-05-13 12:53:20

2017-07-13 22:40:00

2019-02-12 14:33:20

2020-09-13 08:26:40

2022-04-15 01:20:00

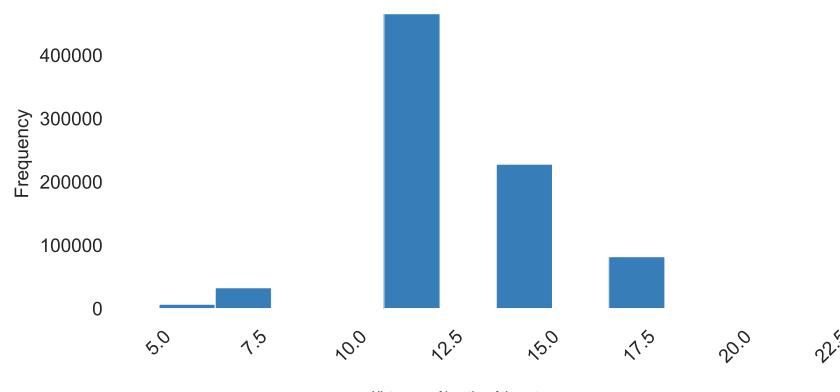
Value	Count	Frequency (%)
10	19118	2.3%
40	7811	1.0%
0	7437	0.9%
45	5446	0.7%
5	4711	0.6%
Other values (36)	1791	0.2%

Value	Count	Frequency (%)
0	7437	0.9%
1	41	< 0.1%
2	28	< 0.1%
3	199	< 0.1%
4	2	< 0.1%
5	4711	0.6%
6	7	< 0.1%
7	5	< 0.1%
8	2	< 0.1%
9	96	< 0.1%

Value	Count	Frequency (%)
99	66	< 0.1%
70	6	< 0.1%
65	18	< 0.1%
63	1	< 0.1%
62	1	< 0.1%
60	50	< 0.1%
55	808	0.1%
50	243	< 0.1%
49	1	< 0.1%
46	1	< 0.1%

TRAFFIC_CONTROL_DEVICE		Categorical			
<b>IMBALANCE</b>					
<b>Distinct</b>	19				
<b>Distinct (%)</b>	< 0.1%				
<b>Missing</b>	0				
<b>Missing (%)</b>	0.0%				
<b>Memory size</b>	6.2 MiB				
<b>Length</b>		<b>Characters and Unicode</b>		<b>Unique</b>	
<b>Total characters</b>	10025084	<b>Unique</b>	0	?	1st row OTHER
<b>Distinct characters</b>		<b>Unique (%)</b>	0.0%		2nd row TRAFFIC SIGNAL
<b>Distinct characters</b>	25				3rd row NO CONTROLS
<b>Distinct categories</b>	1			?	4th row NO CONTROLS
( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )					5th row NO CONTROLS
<b>Distinct scripts</b>	1			?	
( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )					
<b>Distinct blocks</b>	1	( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )		?	
The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.					
<b>Common Values</b>					
<b>Value</b>		<b>Count</b>		<b>Frequency (%)</b>	
NO CONTROLS		464816		56.8%	
TRAFFIC SIGNAL		226679		27.7%	
STOP SIGN/FLASHER		81158		9.9%	
UNKNOWN		32812		4.0%	
OTHER		5550		0.7%	
LANE USE MARKING		1226		0.1%	
YIELD		1207		0.1%	
OTHER REG. SIGN		908		0.1%	
OTHER WARNING SIGN		669		0.1%	
RAILROAD CROSSING GATE		532		0.1%	
Other values (9)		2166		0.3%	

**Length**



Value	Count	Frequency (%)
no	464867	29.1%
controls	464816	29.1%
signal	227007	14.2%
traffic	226679	14.2%
stop	81158	5.1%
sign/flasher	81158	5.1%
unknown	32812	2.1%
other	7308	0.5%
sign	2275	0.1%
crossing	1411	0.1%
Other values (18)	10328	0.6%

**Most occurring characters**

Value	Count	Frequency (%)
O	1520081	15.2%
N	1345892	13.4%
S	943198	9.4%
R	787081	7.9%
T	782096	7.8%
L	777993	7.8%
C	693869	6.9%
I	543866	5.4%
A	541696	5.4%

Value	Count	Frequency (%)
Other values (15)	1307651	13.0%

**Most occurring categories**

Value	Count	Frequency (%)
(unknown)	10025084	100.0%

**Most frequent character per category**

Value	Count	Frequency (%)
(unknown)	1520081	15.2%
O	1345892	13.4%
S	943198	9.4%
R	787081	7.9%
T	782096	7.8%
L	777993	7.8%
C	693869	6.9%
I	543866	5.4%
A	541696	5.4%
Other values (15)	1307651	13.0%

**Most occurring scripts**

Value	Count	Frequency (%)
(unknown)	10025084	100.0%

**Most frequent character per script**

Value	Count	Frequency (%)
(unknown)	1520081	15.2%
O	1345892	13.4%
S	943198	9.4%
R	787081	7.9%
T	782096	7.8%
L	777993	7.8%
C	693869	6.9%

Value	Count	Frequency (%)
I	543866	5.4%
A	541696	5.4%
Other values (15)	1307651	13.0%

**Most occurring blocks**

Value	Count	Frequency (%)
(unknown)	10025084	100.0%

**Most frequent character per block**

(unknown)	Count	Frequency (%)
O	1520081	15.2%
N	1345892	13.4%
S	943198	9.4%
R	787081	7.9%
T	782096	7.8%
L	781661	7.8%
C	777993	7.8%
I	693869	6.9%
A	543866	5.4%
Other values (15)	541696	5.4%
	1307651	13.0%

**DEVICE\_CONDITION**

Categorical

IMBALANCE

Distinct	8
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	6.2 MiB

**Length**

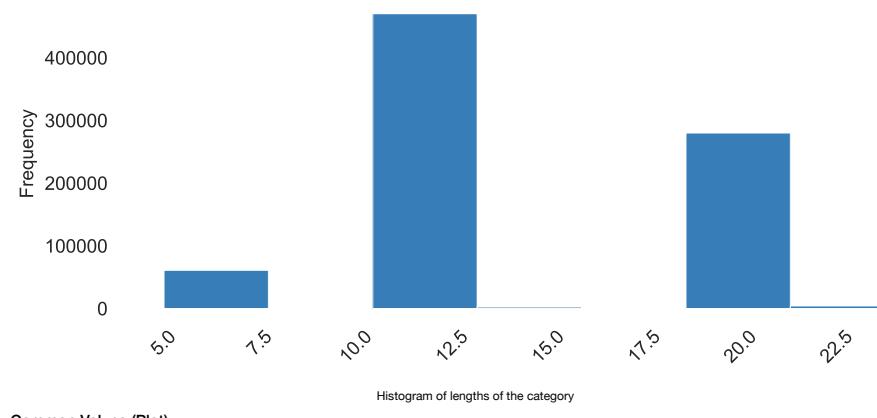
	Characters and Unicode	Unique		Sample
Total characters	11314190	Unique	0	? 1st row FUNCTIONING PROPERLY
Distinct characters	21	Unique (%)	0.0%	? 2nd row FUNCTIONING PROPERLY
Distinct categories	1			? 3rd row NO CONTROLS
Distinct scripts	1			? 4th row NO CONTROLS
Distinct blocks	1			? 5th row NO CONTROLS

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

**Common Values**

Value	Count	Frequency (%)
NO CONTROLS	470240	57.5%
FUNCTIONING PROPERLY	280006	34.2%
UNKNOWN	54603	6.7%
OTHER	6264	0.8%
FUNCTIONING IMPROPERLY	3849	0.5%
NOT FUNCTIONING	2382	0.3%
WORN REFLECTIVE MATERIAL	284	< 0.1%
MISSING	95	< 0.1%

**Length**



- NO CONTROLS
- FUNCTIONING PROPERLY
- UNKNOWN
- OTHER
- FUNCTIONING IMPROPERLY
- NOT FUNCTIONING
- WORN REFLECTIVE MATERIAL
- MISSING

Value	Count	Frequency (%)
no	470240	29.9%
controls	470240	29.9%
functioning	286237	18.2%
properly	280006	17.8%
unknown	54603	3.5%
other	6264	0.4%
improperly	3849	0.2%
not	2382	0.2%
worn	284	< 0.1%
reflective	284	< 0.1%
Other values (2)	379	< 0.1%

#### Most occurring characters

Value	Count	Frequency (%)
O	2044345	18.1%
N	1965761	17.4%
R	1045066	9.2%
T	765691	6.8%
C	756761	6.7%
L	754663	6.7%
I	577081	5.1%
P	567710	5.0%
S	470430	4.2%
Other values (11)	1609637	14.2%

#### Most occurring categories

Value	Count	Frequency (%)
(unknown)	11314190	100.0%

#### Most frequent character per category

Value	Count	Frequency (%)
(unknown)	2044345	18.1%
O	1965761	17.4%
R	1045066	9.2%

Value	Count	Frequency (%)
T	765691	6.8%
	757045	6.7%
C	756761	6.7%
L	754663	6.7%
I	577081	5.1%
P	567710	5.0%
S	470430	4.2%
Other values (11)	1609637	14.2%

**Most occurring scripts**

Value	Count	Frequency (%)
(unknown)	11314190	100.0%

**Most frequent character per script***(unknown)*

Value	Count	Frequency (%)
O	2044345	18.1%
N	1965761	17.4%
R	1045066	9.2%
T	765691	6.8%
	757045	6.7%
C	756761	6.7%
L	754663	6.7%
I	577081	5.1%
P	567710	5.0%
S	470430	4.2%
Other values (11)	1609637	14.2%

**Most occurring blocks**

Value	Count	Frequency (%)
(unknown)	11314190	100.0%

**Most frequent character per block***(unknown)*

Value	Count	Frequency (%)
O	2044345	18.1%

Value	Count	Frequency (%)
N	1965761	17.4%
R	1045066	9.2%
T	765691	6.8%
	757045	6.7%
C	756761	6.7%
L	754663	6.7%
I	577081	5.1%
P	567710	5.0%
S	470430	4.2%
Other values (11)	1609637	14.2%

**WEATHER\_CONDITION**

Categorical

IMBALANCE

Distinct	12
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%

Memory size

6.2 MiB

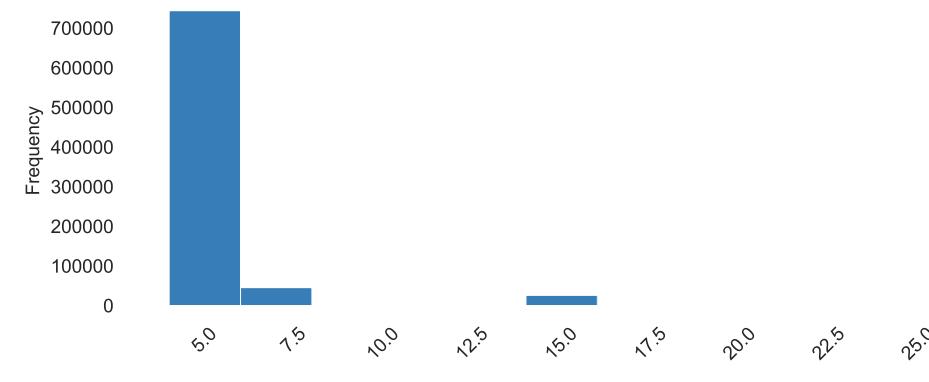
**Length**

	Characters and Unicode	Unique		Sample
Total characters	4371715	Unique	0	1st row
Distinct characters	25	Unique (%)	0.0%	2nd row
Distinct categories	1		?	3rd row
Distinct scripts	1		?	4th row
Distinct blocks	1	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?	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 (%)
CLEAR	641373	78.4%
RAIN	71391	8.7%
UNKNOWN	45108	5.5%
SNOW	28352	3.5%
CLOUDY/OVERCAST	24249	3.0%
OTHER	2601	0.3%
FREEZING RAIN/DRIZZLE	1702	0.2%
FOG/SMOKE/HAZE	1337	0.2%
SLEET/HAIL	1006	0.1%
BLOWING SNOW	444	0.1%
Other values (2)	160	< 0.1%

**Length**

Value	Count	Frequency (%)
clear	641373	78.2%
rain	71391	8.7%
unknown	45108	5.5%
snow	28796	3.5%
cloudy/overcast	24249	3.0%
other	2601	0.3%
freezing	1702	0.2%
rain/drizzle	1702	0.2%
fog/smoke/haze	1337	0.2%
sleet/hail	1006	0.1%
Other values (8)	1084	0.1%

**Most occurring characters**

Value	Count	Frequency (%)
R	745033	17.0%
A	741218	17.0%
C	690024	15.8%
E	678627	15.5%
L	669794	15.3%
N	239526	5.5%
O	128288	2.9%
I	78121	1.8%
W	74508	1.7%
U	69357	1.6%

Value	Count	Frequency (%)
Other values (15)	257219	5.9%

**Most occurring categories**

Value	Count	Frequency (%)
(unknown)	4371715	100.0%

**Most frequent character per category**

(unknown)	Value	Count	Frequency (%)
R		745033	17.0%
A		741218	17.0%
C		690024	15.8%
E		678627	15.5%
L		669794	15.3%
N		239526	5.5%
O		128288	2.9%
I		78121	1.8%
W		74508	1.7%
U		69357	1.6%
Other values (15)		257219	5.9%

**Most occurring scripts**

Value	Count	Frequency (%)
(unknown)	4371715	100.0%

**Most frequent character per script**

(unknown)	Value	Count	Frequency (%)
R		745033	17.0%
A		741218	17.0%
C		690024	15.8%
E		678627	15.5%
L		669794	15.3%
N		239526	5.5%
O		128288	2.9%
I		78121	1.8%

Value	Count	Frequency (%)
W	74508	1.7%
U	69357	1.6%
Other values (15)	257219	5.9%

**Most occurring blocks**

Value	Count	Frequency (%)
(unknown)	4371715	100.0%

**Most frequent character per block**

(unknown)	Value	Count	Frequency (%)
R		745033	17.0%
A		741218	17.0%
C		690024	15.8%
E		678627	15.5%
L		669794	15.3%
N		239526	5.5%
O		128288	2.9%
I		78121	1.8%
W		74508	1.7%
U		69357	1.6%
Other values (15)		257219	5.9%

**LIGHTING\_CONDITION**

Categorical

Distinct	6
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	6.2 MiB

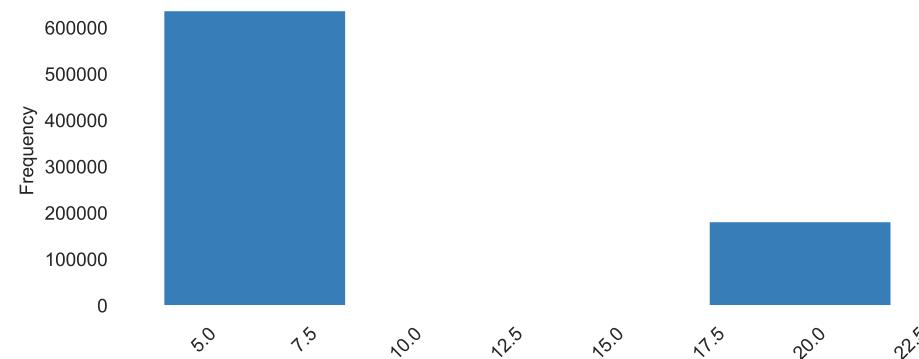
**Length****Characters and Unicode****Unique****Sample**

Total characters	8889772	Unique	0	?	1st row	DAYLIGHT
Distinct characters	18	Unique (%)	0.0%	?	2nd row	DAYLIGHT
Distinct categories	1	?	?	?	3rd row	DAYLIGHT
Distinct scripts	(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?	?	?	4th row	DAYLIGHT
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?	?	?	5th row	DAYLIGHT

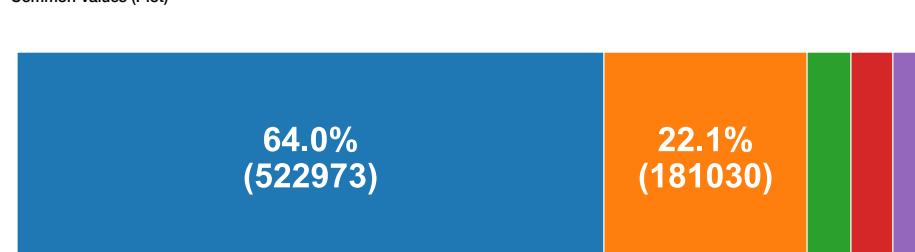
The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

**Common Values**

Value	Count	Frequency (%)
DAYLIGHT	522973	64.0%
DARKNESS, LIGHTED ROAD	181030	22.1%
DARKNESS	39092	4.8%
UNKNOWN	37360	4.6%
DUSK	23579	2.9%
DAWN	13689	1.7%

**Length**

Common Values (Plot)



Value	Count	Frequency (%)
road	181030	15.3%
unknown	37360	3.2%
dusk	23579	2.0%
dawn	13689	1.2%

**Most occurring characters**

Value	Count	Frequency (%)
D	1142423	12.9%
A	937814	10.5%
L	704003	7.9%
I	704003	7.9%
G	704003	7.9%
H	704003	7.9%
T	704003	7.9%
Y	522973	5.9%
S	463823	5.2%
R	401152	4.5%
Other values (8)	1901572	21.4%

**Most occurring categories**

Value	Count	Frequency (%)
(unknown)	8889772	100.0%

**Most frequent character per category**

Value	Count	Frequency (%)
(unknown)	1142423	12.9%
D	937814	10.5%
L	704003	7.9%
I	704003	7.9%
G	704003	7.9%
H	704003	7.9%
T	704003	7.9%
Y	522973	5.9%
S	463823	5.2%
R	401152	4.5%

Value	Count	Frequency (%)
Other values (8)	1901572	21.4%

**Most occurring scripts**

Value	Count	Frequency (%)
(unknown)	8889772	100.0%

**Most frequent character per script**

Value	Count	Frequency (%)
(unknown)	1142423	12.9%
D	937814	10.5%
L	704003	7.9%
I	704003	7.9%
G	704003	7.9%
H	704003	7.9%
T	704003	7.9%
Y	522973	5.9%
S	463823	5.2%
R	401152	4.5%
Other values (8)	1901572	21.4%

**Most occurring blocks**

Value	Count	Frequency (%)
(unknown)	8889772	100.0%

**Most frequent character per block**

Value	Count	Frequency (%)
(unknown)	1142423	12.9%
D	937814	10.5%
L	704003	7.9%
I	704003	7.9%
G	704003	7.9%
H	704003	7.9%
T	704003	7.9%
Y	522973	5.9%

Value	Count	Frequency (%)
S	463823	5.2%
R	401152	4.5%
Other values (8)	1901572	21.4%

### FIRST\_CRASH\_TYPE

Categorical

Distinct	18
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	6.2 MiB

#### Length

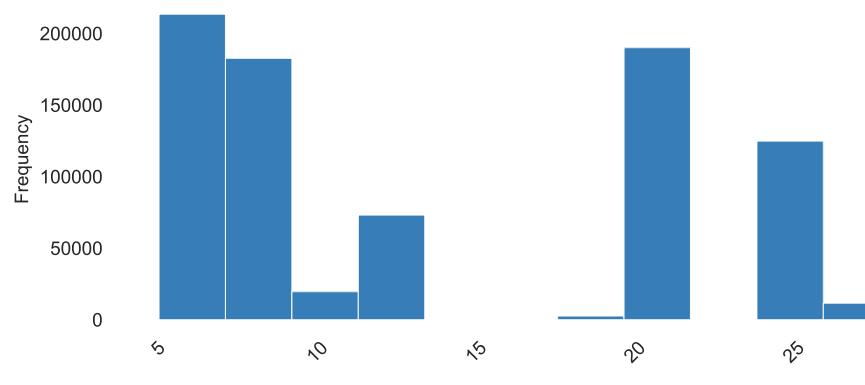
	Characters and Unicode	Unique	Sample
Total characters	11023148	Unique	0 ? 1st row REAR END
Distinct characters	25	Unique (%)	0.0% 2nd row PARKED MOTOR VEHICLE
Min length	5		3rd row PEDALCYCLIST
Max length	28		? 4th row PEDESTRIAN
Median length	20		(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category) 5th row FIXED OBJECT
Mean length	13.480296		?
Distinct categories	1		(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
Distinct scripts	1		?
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 (%)
PARKED MOTOR VEHICLE	190062	23.2%
REAR END	182653	22.3%
SIDESWIPE SAME DIRECTION	124798	15.3%
TURNING	116863	14.3%
ANGLE	88951	10.9%
FIXED OBJECT	38417	4.7%
PEDESTRIAN	19065	2.3%
PEDALCYCLIST	12307	1.5%
SIDESWIPE OPPOSITE DIRECTION	11529	1.4%
OTHER OBJECT	8102	1.0%
Other values (8)	24976	3.1%

#### Length



Value	Count	Frequency (%)
rear	198570	11.4%
parked	190062	10.9%
motor	190062	10.9%
vehicle	190062	10.9%
end	182653	10.5%
sideswipe	136327	7.8%
direction	136327	7.8%
same	124798	7.2%
turning	116863	6.7%
angle	88951	5.1%
Other values (15)	183121	10.5%

**Most occurring characters**

Value	Count	Frequency (%)
E	1744409	15.8%
R	1069188	9.7%
I	944043	8.6%
	920073	8.3%
D	727345	6.6%
N	684515	6.2%
A	641957	5.8%
O	634149	5.8%
T	566121	5.1%
S	447634	4.1%

Value	Count	Frequency (%)
Other values (15)	2643714	24.0%

**Most occurring categories**

Value	Count	Frequency (%)
(unknown)	11023148	100.0%

**Most frequent character per category**

(unknown)	Value	Count	Frequency (%)
E	E	1744409	15.8%
R	R	1069188	9.7%
I	I	944043	8.6%
	920073	8.3%	
D	D	727345	6.6%
N	N	684515	6.2%
A	A	641957	5.8%
O	O	634149	5.8%
T	T	566121	5.1%
S	S	447634	4.1%
Other values (15)	Other values (15)	2643714	24.0%

**Most occurring scripts**

(unknown)	Value	Count	Frequency (%)
(unknown)	(unknown)	11023148	100.0%

**Most frequent character per script**

(unknown)	Value	Count	Frequency (%)
E	E	1744409	15.8%
R	R	1069188	9.7%
I	I	944043	8.6%
	920073	8.3%	
D	D	727345	6.6%
N	N	684515	6.2%
A	A	641957	5.8%
O	O	634149	5.8%

Value	Count	Frequency (%)
T	566121	5.1%
S	447634	4.1%
Other values (15)	2643714	24.0%

**Most occurring blocks**

Value	Count	Frequency (%)
(unknown)	11023148	100.0%

**Most frequent character per block**

Value	Count	Frequency (%)
E	1744409	15.8%
R	1069188	9.7%
I	944043	8.6%
D	727345	6.6%
N	684515	6.2%
A	641957	5.8%
O	634149	5.8%
T	566121	5.1%
S	447634	4.1%
Other values (15)	2643714	24.0%

**TRAFFICWAY\_TYPE**

Categorical

Distinct	20
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	6.2 MiB

**Length**

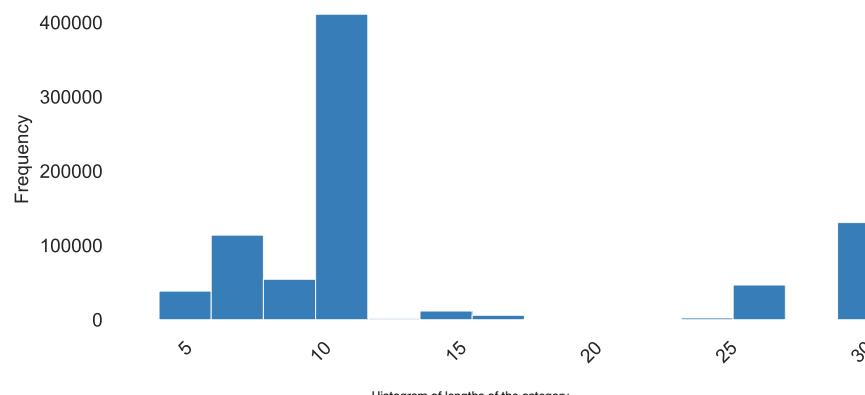
	Characters and Unicode	Unique		Sample
Total characters	11567277	Unique	0 ?	1st row OTHER
Distinct characters	28	Unique (%)	0.0%	2nd row DIVIDED - W/MEDIAN (NOT RAISED)
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)		? 3rd row	NOT DIVIDED
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)		? 4th row	ONE-WAY
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)		? 5th row	OTHER

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

**Common Values**

Value	Count	Frequency (%)
NOT DIVIDED	355079	43.4%
DIVIDED - W/MEDIAN (NOT RAISED)	130679	16.0%
ONE-WAY	104273	12.8%
PARKING LOT	55690	6.8%
FOUR WAY	51707	6.3%
DIVIDED - W/MEDIAN BARRIER	46724	5.7%
OTHER	22378	2.7%
ALLEY	13526	1.7%
T-INTERSECTION	10407	1.3%
UNKNOWN	9520	1.2%
Other values (10)	17740	2.2%

**Length**



Value	Count	Frequency (%)
divided	532482	27.1%
not	486338	24.8%
w/median	177403	9.0%
raised	130679	6.7%
one-way	104273	5.3%
parking	55690	2.8%
lot	55690	2.8%
four	51707	2.6%
way	51707	2.6%
Other values (22)	141247	7.2%

**Most occurring characters**

Value	Count	Frequency (%)
D	1909026	16.5%
I	1509514	13.0%
E	1082853	9.4%
N	906834	7.8%
O	751820	6.5%
T	621099	5.4%
A	592334	5.1%
V	536266	4.6%
R	437333	3.8%

Value	Count	Frequency (%)
Other values (18)	2073302	17.9%

**Most occurring categories**

Value	Count	Frequency (%)
(unknown)	11567277	100.0%

**Most frequent character per category**

(unknown)	Value	Count	Frequency (%)
D	D	1909026	16.5%
I	I	1509514	13.0%
E	E	1082853	9.4%
N	N	906834	7.8%
O	O	751820	6.5%
T	T	621099	5.4%
A	A	592334	5.1%
V	V	536266	4.6%
R	R	437333	3.8%
Other values (18)	Other values (18)	2073302	17.9%

**Most occurring scripts**

(unknown)	Value	Count	Frequency (%)
(unknown)	(unknown)	11567277	100.0%

**Most frequent character per script**

(unknown)	Value	Count	Frequency (%)
D	D	1909026	16.5%
I	I	1509514	13.0%
E	E	1082853	9.4%
N	N	906834	7.8%
O	O	751820	6.5%
T	T	621099	5.4%
A	A	592334	5.1%

Value	Count	Frequency (%)
V	536266	4.6%
R	437333	3.8%
Other values (18)	2073302	17.9%

**Most occurring blocks**

Value	Count	Frequency (%)
(unknown)	11567277	100.0%

**Most frequent character per block**

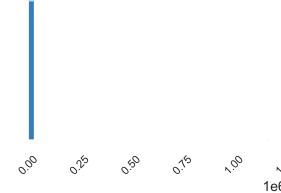
Value	Count	Frequency (%)
D	1909026	16.5%
I	1509514	13.0%
E	1146896	9.9%
N	1082853	9.4%
O	906834	7.8%
T	751820	6.5%
A	621099	5.4%
V	592334	5.1%
R	536266	4.6%
Other values (18)	437333	3.8%
	2073302	17.9%

**LANE\_CNT**  
Real number (R)

MISSING SKewed

Distinct	41
Distinct (%)	< 0.1%
Missing	618714
Missing (%)	75.7%
Infinite	0
Infinite (%)	0.0%
Mean	13.330161

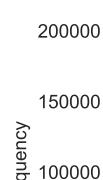
Minimum	0
Maximum	1191625
Zeros	8032
Zeros (%)	1.0%
Negative	0
Negative (%)	0.0%
Memory size	6.2 MiB

**Quantile statistics**

Minimum	0
5-th percentile	1
Q1	2
median	2
Q3	4
95-th percentile	4
Maximum	1191625
Range	1191625
Interquartile range (IQR)	2

**Descriptive statistics**

Standard deviation	2961.6012
Coefficient of variation (CV)	222.17295
Kurtosis	134678.59
Mean	13.330161
Median Absolute Deviation (MAD)	1
Skewness	350.3036
Sum	2652822
Variance	8771081.4
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=41)

Value	Count	Frequency (%)
2	91154	11.1%
4	49588	6.1%
1	32547	4.0%
3	8677	1.1%
0	8032	1.0%

Value	Count	Frequency (%)
6	4502	0.6%
5	1940	0.2%
8	1908	0.2%
7	184	< 0.1%
10	162	< 0.1%
Other values (31)	315	< 0.1%
(Missing)	618714	75.7%

Value	Count	Frequency (%)
0	8032	1.0%
1	32547	4.0%
2	91154	11.1%
3	8677	1.1%
4	49588	6.1%
5	1940	0.2%
6	4502	0.6%
7	184	< 0.1%
8	1908	0.2%
9	66	< 0.1%
Value	Count	Frequency (%)
1191625	1	< 0.1%
433634	1	< 0.1%
299679	1	< 0.1%
218474	1	< 0.1%
902	1	< 0.1%
400	1	< 0.1%
100	2	< 0.1%
99	108	< 0.1%
80	1	< 0.1%
60	3	< 0.1%

**ALIGNMENT**

Categorical

**IMBALANCE**

<b>Distinct</b>	6
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%

**Memory size** 6.2 MiB**Length**

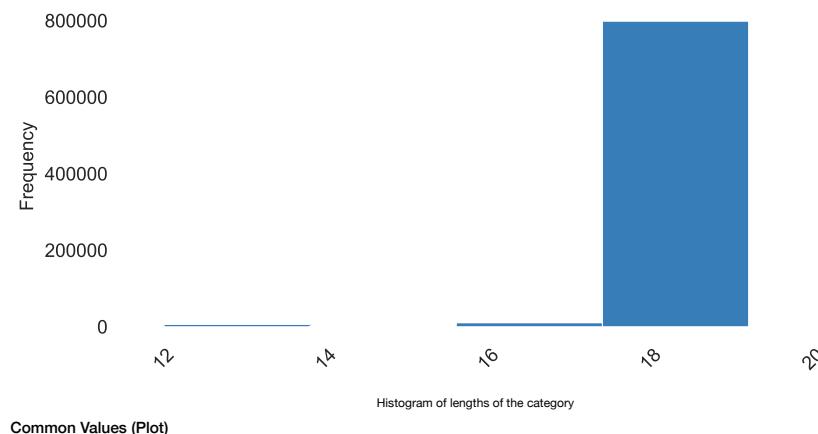
	Characters and Unicode	Unique		Sample
Total characters	14674910	Unique	0	? 1st row STRAIGHT AND LEVEL
Distinct characters	17	Unique (%)	0.0%	? 2nd row STRAIGHT AND LEVEL
Distinct categories	1			? 3rd row STRAIGHT AND LEVEL
Distinct scripts	( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )			? 4th row CURVE ON GRADE
Distinct blocks	1 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )			? 5th row STRAIGHT AND LEVEL

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

**Common Values**

Value	Count	Frequency (%)
STRAIGHT AND LEVEL	797888	97.6%
STRAIGHT ON GRADE	10191	1.2%
CURVE, LEVEL	5904	0.7%
STRAIGHT ON HILLCREST	2145	0.3%
CURVE ON GRADE	1231	0.2%
CURVE ON HILLCREST	364	< 0.1%

**Length**



Common Values (Plot)



- █ STRAIGHT AND LEVEL
- █ STRAIGHT ON GRADE
- █ CURVE, LEVEL
- █ STRAIGHT ON HILLCREST
- █ CURVE ON GRADE
- █ CURVE ON HILLCREST

Value	Count	Frequency (%)
straight	810224	33.1%
level	803792	32.8%
and	797888	32.6%

Value	Count	Frequency (%)
on	13931	0.6%
grade	11422	0.5%
curve	7499	0.3%
hillcrest	2509	0.1%

**Most occurring characters**

Value	Count	Frequency (%)
E	1629542	11.1%
T	1629014	11.1%
A	1622957	11.1%
L	1619534	11.0%
R	1612602	11.0%
G	831654	5.7%
S	821646	5.6%
I	812733	5.5%
H	812733	5.5%
Other values (7)	2469762	16.8%

**Most occurring categories**

Value	Count	Frequency (%)
(unknown)	14674910	100.0%

**Most frequent character per category**

Value	Count	Frequency (%)
E	1629542	11.1%
T	1629014	11.1%
A	1622957	11.1%
L	1619534	11.0%
R	1612602	11.0%
G	831654	5.7%
S	821646	5.6%
I	812733	5.5%
H	812733	5.5%

Value	Count	Frequency (%)
Other values (7)	2469762	16.8%

**Most occurring scripts**

Value	Count	Frequency (%)
(unknown)	14674910	100.0%

**Most frequent character per script**

Value	Count	Frequency (%)
E	1629542	11.1%
T	1629014	11.1%
A	1622957	11.1%
L	1619534	11.0%
R	1612602	11.0%
G	831654	5.7%
S	821646	5.6%
I	812733	5.5%
H	812733	5.5%
Other values (7)	2469762	16.8%

**Most occurring blocks**

Value	Count	Frequency (%)
(unknown)	14674910	100.0%

**Most frequent character per block**

Value	Count	Frequency (%)
E	1629542	11.1%
T	1629014	11.1%
A	1622957	11.1%
L	1619534	11.0%
R	1612602	11.0%
G	831654	5.7%
S	821646	5.6%
I	812733	5.5%

Value	Count	Frequency (%)
I	812733	5.5%
H	812733	5.5%
Other values (7)	2469762	16.8%

## ROADWAY\_SURFACE\_COND

Categorical

IMBALANCE

<b>Distinct</b>	7
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%

**Memory size** 6.2 MiB

## Length

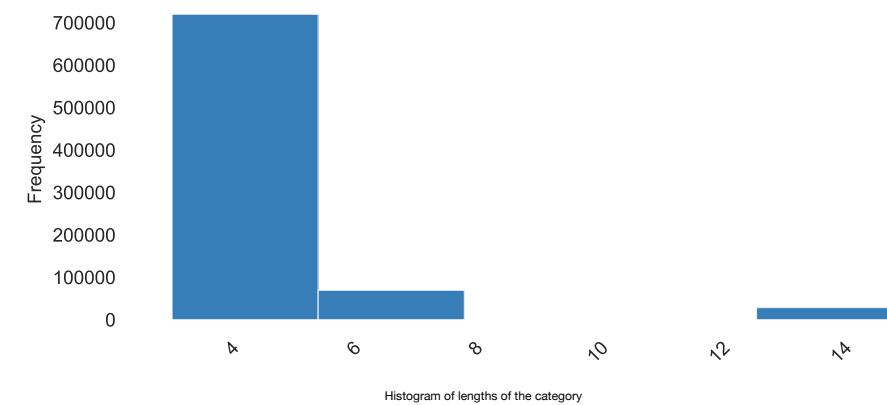
	Characters and Unicode	Unique		Sample
Total characters	3019697	Unique	0 ?	1st row DRY
Distinct characters	19	Unique (%)	0.0%	2nd row DRY
Distinct categories	1			3rd row DRY
Distinct scripts	1			? 4th row DRY
Distinct blocks	1	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)		? 5th row DRY

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

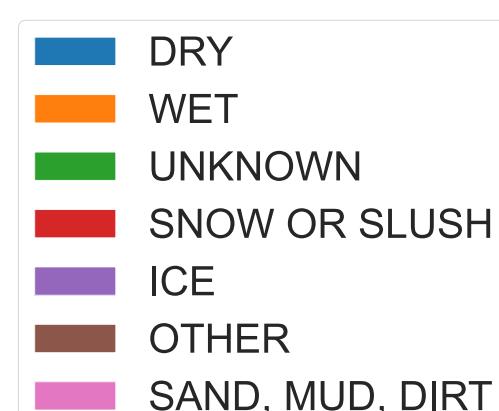
## Common Values

Value	Count	Frequency (%)
DRY	603295	73.8%
WET	108880	13.3%
UNKNOWN	69415	8.5%
SNOW OR SLUSH	28111	3.4%
ICE	5658	0.7%
OTHER	2061	0.3%
SAND, MUD, DIRT	303	< 0.1%

## Length



Common Values (Plot)



## Austin Crash Report Data - Profiling Report

Value	Count	Frequency (%)
wet	108880	12.4%
unknown	69415	7.9%
snow	28111	3.2%
or	28111	3.2%
slush	28111	3.2%
ice	5658	0.6%
other	2061	0.2%
sand	303	< 0.1%
mud	303	< 0.1%

## Most occurring characters

Value	Count	Frequency (%)
R	633770	21.0%
D	604204	20.0%
Y	603295	20.0%
N	236659	7.8%
W	206406	6.8%
O	127698	4.2%
E	116599	3.9%
T	111244	3.7%
U	97829	3.2%
S	84636	2.8%
Other values (9)	197357	6.5%

## Most occurring categories

Value	Count	Frequency (%)
(unknown)	3019697	100.0%

## Most frequent character per category

(unknown)	Value	Count	Frequency (%)
R		633770	21.0%
D		604204	20.0%
Y		603295	20.0%
N		236659	7.8%
W		206406	6.8%

## Austin Crash Report Data - Profiling Report

Value	Count	Frequency (%)
O	127698	4.2%
E	116599	3.9%
T	111244	3.7%
U	97829	3.2%
S	84636	2.8%
Other values (9)	197357	6.5%

## Most occurring scripts

Value	Count	Frequency (%)
(unknown)	3019697	100.0%

## Most frequent character per script

(unknown)	Value	Count	Frequency (%)
R		633770	21.0%
D		604204	20.0%
Y		603295	20.0%
N		236659	7.8%
W		206406	6.8%
O		127698	4.2%
E		116599	3.9%
T		111244	3.7%
U		97829	3.2%
S		84636	2.8%
Other values (9)		197357	6.5%

## Most occurring blocks

Value	Count	Frequency (%)
(unknown)	3019697	100.0%

## Most frequent character per block

(unknown)	Value	Count	Frequency (%)
R		633770	21.0%
D		604204	20.0%
Y		603295	20.0%

Value	Count	Frequency (%)
N	236659	7.8%
W	206406	6.8%
O	127698	4.2%
E	116599	3.9%
T	111244	3.7%
U	97829	3.2%
S	84636	2.8%
Other values (9)	197357	6.5%

**ROAD\_DEFECT**

Categorical

## IMBALANCE

Distinct	7
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	6.2 MiB

## Length

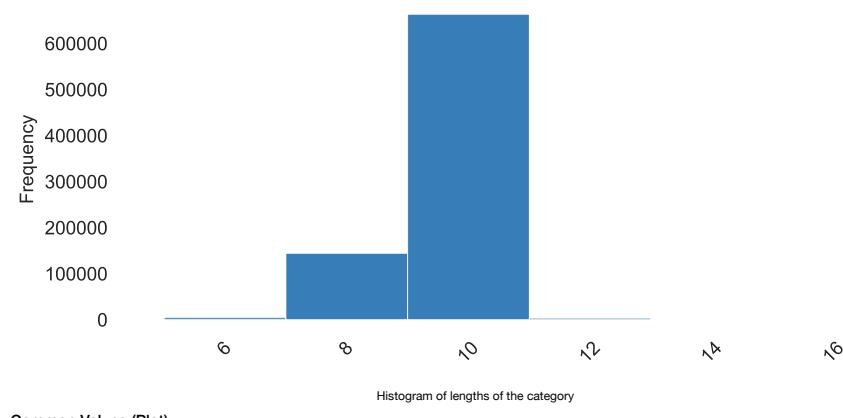
	Characters and Unicode	Unique		Sample
Total characters	7740328	Unique	0	? 1st row NO DEFECTS
Distinct characters	20	Unique (%)	0.0%	? 2nd row NO DEFECTS
Distinct categories	1			? 3rd row NO DEFECTS
Distinct scripts	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)			? 4th row NO DEFECTS
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)			? 5th row NO DEFECTS

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

## Common Values

Value	Count	Frequency (%)
NO DEFECTS	657425	80.4%
UNKNOWN	144337	17.7%
RUT, HOLES	6056	0.7%
OTHER	4458	0.5%
WORN SURFACE	3356	0.4%
SHOULDER DEFECT	1475	0.2%
DEBRIS ON ROADWAY	616	0.1%

## Length



Common Values (Plot)



Value	Count	Frequency (%)
no	657425	44.2%

Value	Count	Frequency (%)
defects	657425	44.2%
unknown	144337	9.7%
rut	6056	0.4%
holes	6056	0.4%
other	4458	0.3%
worn	3356	0.2%
surface	3356	0.2%
shoulder	1475	0.1%
defect	1475	0.1%
Other values (3)	1848	0.1%

## Most occurring characters

Value	Count	Frequency (%)
E	1333761	17.2%
N	1094408	14.1%
O	818339	10.6%
	669544	8.7%
T	669414	8.6%
S	668928	8.6%
F	662256	8.6%
C	662256	8.6%
D	661607	8.5%
U	155224	2.0%
Other values (10)	344591	4.5%

## Most occurring categories

Value	Count	Frequency (%)
(unknown)	7740328	100.0%

## Most frequent character per category

Value	Count	Frequency (%)
(unknown)	1333761	17.2%
E	1094408	14.1%
N	818339	10.6%
O	669544	8.7%

Value	Count	Frequency (%)
T	669414	8.6%
S	668928	8.6%
F	662256	8.6%
C	662256	8.6%
D	661607	8.5%
U	155224	2.0%
Other values (10)	344591	4.5%

**Most occurring scripts**

Value	Count	Frequency (%)
(unknown)	7740328	100.0%

**Most frequent character per script**

Value	Count	Frequency (%)
(unknown)		
E	1333761	17.2%
N	1094408	14.1%
O	818339	10.6%
	669544	8.7%
T	669414	8.6%
S	668928	8.6%
F	662256	8.6%
C	662256	8.6%
D	661607	8.5%
U	155224	2.0%
Other values (10)	344591	4.5%

**Most occurring blocks**

Value	Count	Frequency (%)
(unknown)	7740328	100.0%

**Most frequent character per block**

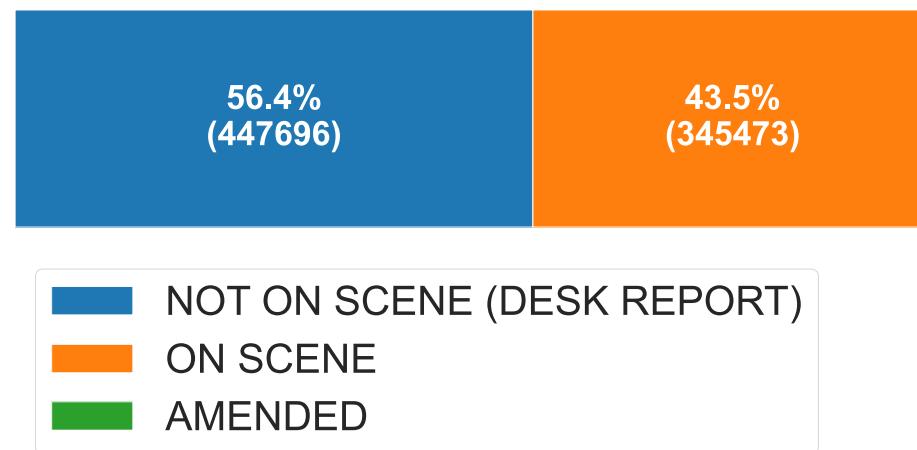
Value	Count	Frequency (%)
(unknown)		
E	1333761	17.2%
N	1094408	14.1%

Value	Count	Frequency (%)
O	818339	10.6%
	669544	8.7%
T	669414	8.6%
S	668928	8.6%
F	662256	8.6%
C	662256	8.6%
D	661607	8.5%
U	155224	2.0%
Other values (10)	344591	4.5%



Histogram of lengths of the category

Common Values (Plot)



Value	Count	Frequency (%)
on	793169	27.1%
scene	793169	27.1%
not	447696	15.3%
desk	447696	15.3%
report	447696	15.3%
amended	240	< 0.1%

**Most occurring characters**

Value	Count	Frequency (%)
E	2482210	17.2%
N	2034274	14.1%
O	1688561	11.7%
S	1240865	8.6%
T	895392	6.2%
R	895392	6.2%
C	793169	5.5%
D	448176	3.1%
(	447696	3.1%
Other values (5)	1343568	9.3%

**Most occurring categories**

Value	Count	Frequency (%)
(unknown)	14405560	100.0%

**Most frequent character per category**

(unknown)	Value	Count	Frequency (%)
E		2482210	17.2%
		2136257	14.8%
N		2034274	14.1%
O		1688561	11.7%
S		1240865	8.6%
T		895392	6.2%
R		895392	6.2%
C		793169	5.5%
D		448176	3.1%
(		447696	3.1%
Other values (5)		1343568	9.3%

**Most occurring scripts**

Value	Count	Frequency (%)
(unknown)	14405560	100.0%

**Most frequent character per script**

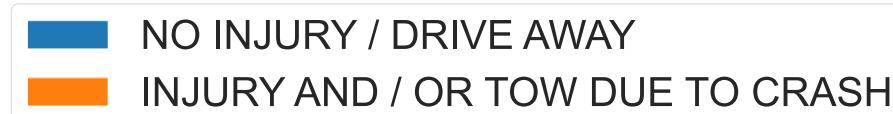
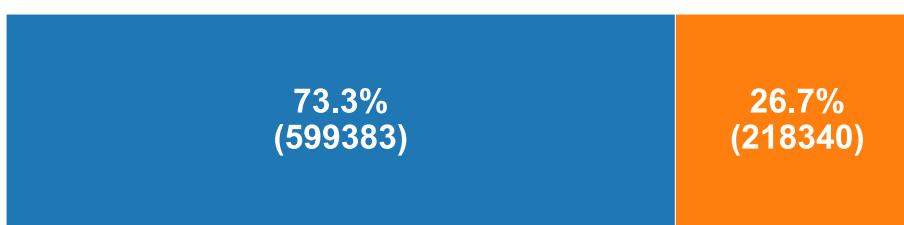
(unknown)	Value	Count	Frequency (%)
E		2482210	17.2%
		2136257	14.8%
N		2034274	14.1%
O		1688561	11.7%
S		1240865	8.6%
T		895392	6.2%
R		895392	6.2%
C		793169	5.5%
D		448176	3.1%
(		447696	3.1%
Other values (5)		1343568	9.3%

**Most occurring blocks**

Value	Count	Frequency (%)
(unknown)	14405560	100.0%

**Most frequent character per block**

(unknown)	Value	Count	Frequency (%)
E		2482210	17.2%
		2136257	14.8%
N		2034274	14.1%
O		1688561	11.7%
S		1240865	8.6%
T		895392	6.2%
R		895392	6.2%
C		793169	5.5%
D		448176	3.1%
(		447696	3.1%
Other values (5)		1343568	9.3%



Value	Count	Frequency (%)
injury	817723	17.2%
no	817723	17.2%
drive	599383	12.6%
away	599383	12.6%
and	218340	4.6%
or	218340	4.6%
tow	218340	4.6%
due	218340	4.6%
to	218340	4.6%

**Most occurring characters**

Value	Count	Frequency (%)
R	3925912	19.5%
N	1853786	9.2%
A	1635446	8.1%
Y	1635446	8.1%
I	1417106	7.0%
O	1417106	7.0%
D	1254403	6.2%
U	1036063	5.1%
/	1036063	5.1%
Other values (8)	817723	4.1%
	4144252	20.5%

**Most occurring categories**

Value	Count	Frequency (%)
(unknown)	20173306	100.0%

**Most frequent character per category**

Value	Count	Frequency (%)
R	3925912	19.5%
N	1853786	9.2%
A	1635446	8.1%
Y	1635446	8.1%
I	1417106	7.0%
O	1417106	7.0%
D	1254403	6.2%
U	1036063	5.1%
/	817723	4.1%
Other values (8)	4144252	20.5%

**Most occurring scripts**

Value	Count	Frequency (%)
(unknown)	20173306	100.0%

**Most frequent character per script**

Value	Count	Frequency (%)
R	3925912	19.5%
N	1853786	9.2%
A	1635446	8.1%
Y	1635446	8.1%
I	1417106	7.0%
O	1417106	7.0%
D	1254403	6.2%
U	1036063	5.1%
/	817723	4.1%
Other values (8)	4144252	20.5%

**Most occurring blocks**

Value	Count	Frequency (%)
(unknown)	20173306	100.0%

**Most frequent character per block**

Value	Count	Frequency (%)
R	3925912	19.5%
N	1853786	9.2%
A	1635446	8.1%
Y	1635446	8.1%
I	1417106	7.0%
O	1417106	7.0%
D	1254403	6.2%
U	1036063	5.1%
/	817723	4.1%
Other values (8)	4144252	20.5%

**INTERSECTION\_RELATED\_I**

Boolean

IMBALANCE MISSING

<b>Distinct</b>	2
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	630174
<b>Missing (%)</b>	77.1%

**Memory size** 1.6 MiB

Value	Count	Frequency (%)
True	178642	21.8%
False	8907	1.1%
(Missing)	630174	77.1%

**95.3%**  
**(178642)**

True
False

**NOT\_RIGHT\_OF WAY\_I**

Boolean

IMBALANCE MISSING

<b>Distinct</b>	2
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	780015
<b>Missing (%)</b>	95.4%

**Memory size** 1.6 MiB

Value	Count	Frequency (%)
True	34256	4.2%
False	3452	0.4%
(Missing)	780015	95.4%

**90.8%**  
**(34256)**

**9.2%**  
**(3452)**

True
False

**HIT\_AND\_RUN\_I**

Boolean

IMBALANCE MISSING

Distinct	2
Distinct (%)	< 0.1%
Missing	561774
Missing (%)	68.7%

Memory size 1.6 MiB

Value	Count	Frequency (%)
True	244955	30.0%
False	10994	1.3%
(Missing)	561774	68.7%

**95.7%**  
**(244955)**

True  
False

**DAMAGE**

Categorical

Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	6.2 MiB

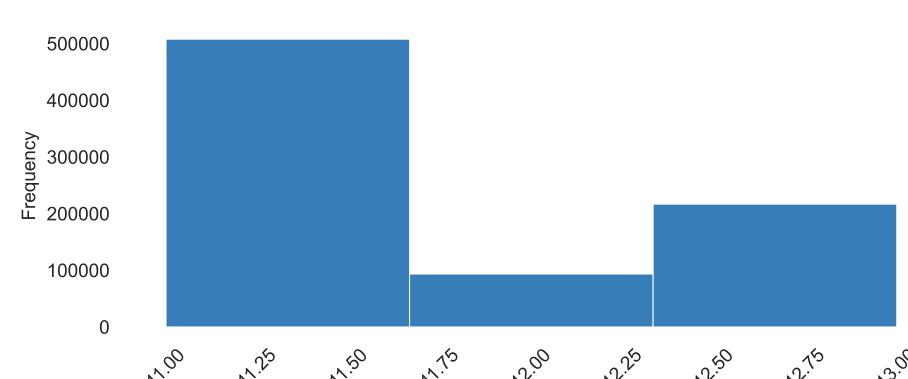
**Length****Characters and Unicode****Unique****Sample**

Total characters	9521590	Unique	0	?	1st row	OVER \$1,500
Distinct characters	13	Unique (%)	0.0%		2nd row	OVER \$1,500
Distinct categories	1				3rd row	\$501 - \$1,500
Distinct scripts	1			?	4th row	OVER \$1,500
Distinct blocks	1	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)		?	5th row	OVER \$1,500

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

**Common Values****Value****Count** **Frequency (%)**

OVER \$1,500	507718	62.1%
\$501 - \$1,500	216632	26.5%
\$500 OR LESS	93373	11.4%

**Length****Common Values (Plot)**



**OVER \$1,500**  
**501 – 1,500**  
**\$500 OR LESS**

Value	Count	Frequency (%)
1,500	724350	37.2%
over	507718	26.1%
501	216632	11.1%
	216632	11.1%
500	93373	4.8%
or	93373	4.8%
less	93373	4.8%

#### Most occurring characters

Value	Count	Frequency (%)
0	1852078	19.5%
	1127728	11.8%
\$	1034355	10.9%
5	1034355	10.9%
1	940982	9.9%
.	724350	7.6%
O	601091	6.3%
E	601091	6.3%
R	601091	6.3%
V	507718	5.3%
Other values (3)	496751	5.2%

#### Most occurring categories

Value	Count	Frequency (%)
(unknown)	9521590	100.0%

#### Most frequent character per category

Value	Count	Frequency (%)
0	1852078	19.5%
	1127728	11.8%
\$	1034355	10.9%
5	1034355	10.9%
1	940982	9.9%
.	724350	7.6%
O	601091	6.3%
E	601091	6.3%
R	601091	6.3%
V	507718	5.3%
Other values (3)	496751	5.2%

#### Most occurring scripts

Value	Count	Frequency (%)
(unknown)	9521590	100.0%

#### Most frequent character per script

Value	Count	Frequency (%)
0	1852078	19.5%
	1127728	11.8%
\$	1034355	10.9%
5	1034355	10.9%
1	940982	9.9%
.	724350	7.6%
O	601091	6.3%
E	601091	6.3%
R	601091	6.3%
V	507718	5.3%
Other values (3)	496751	5.2%

**Most occurring blocks**

Value	Count	Frequency (%)
(unknown)	9521590	100.0%

**Most frequent character per block**

(unknown)	Count	Frequency (%)
0	1852078	19.5%
\$	1127728	11.8%
5	1034355	10.9%
1	1034355	10.9%
,	940982	9.9%
O	724350	7.6%
E	601091	6.3%
R	601091	6.3%
V	507718	5.3%
Other values (3)	496751	5.2%

**DATE\_POLICE\_NOTIFIED**

Date

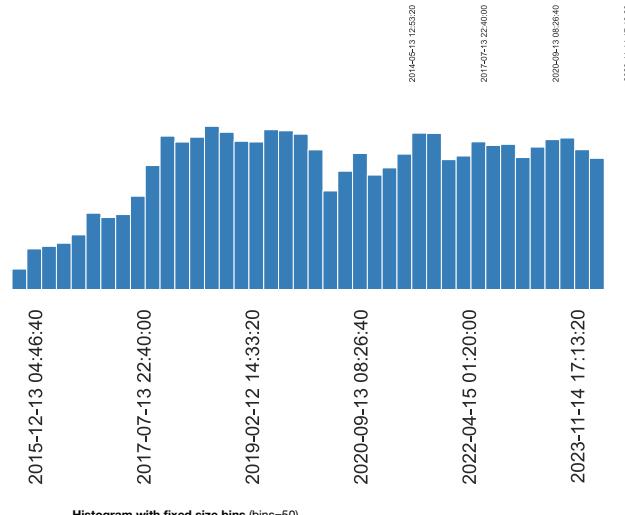
Distinct	620545
Distinct (%)	75.9%
Missing	0
Missing (%)	0.0%
Memory size	6.2 MiB

Minimum 2013-06-01 20:31:00

Maximum 2024-03-26 01:42:00

Frequency  
0  
10000  
20000

2014-05-13 12:53:20



**PRIM\_CONTRIBUTORY\_CAUSE**

Categorical

Distinct	40
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	6.2 MiB

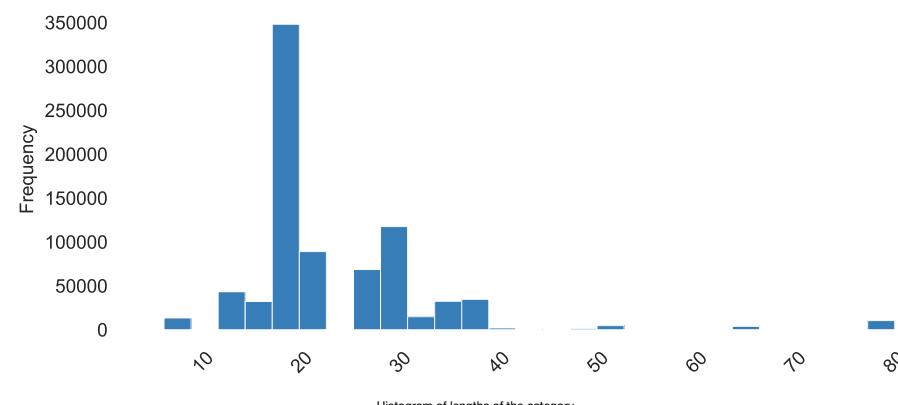
**Length****Characters and Unicode****Unique****Sample**

Total characters	19394375	Unique	0	?	1st row	FOLLOWING TOO CLOSELY
Distinct characters	33	Unique (%)	0.0%		2nd row	FAILING TO REDUCE SPEED
Distinct categories	1			?	3rd row	TO AVOID CRASH
Distinct scripts	1			?	4th row	FAILING TO REDUCE SPEED
Distinct blocks	1	( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?	5th row	UNABLE TO DETERMINE

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

**Common Values**

Value	Count	Frequency (%)
UNABLE TO DETERMINE	318311	38.9%
FAILING TO YIELD RIGHT-OF-WAY	89872	11.0%
FOLLOWING TOO CLOSELY	79457	9.7%
NOT APPLICABLE	43275	5.3%
IMPROPER OVERTAKING/PASSING	40275	4.9%
FAILING TO REDUCE SPEED TO AVOID CRASH	34546	4.2%
IMPROPER BACKING	32183	3.9%
IMPROPER LANE USAGE	29304	3.6%
DRIVING SKILLS/KNOWLEDGE/EXPERIENCE	27355	3.3%
IMPROPER TURNING/NO SIGNAL	27176	3.3%
Other values (30)	95969	11.7%

**Length****Value**

Value	Count	Frequency (%)
to	479178	17.8%
unable	318311	11.8%
determine	318311	11.8%
improper	128938	4.8%
failing	124418	4.6%
yield	90143	3.4%
right-of-way	89872	3.3%
closely	79457	3.0%
too	79457	3.0%
following	79457	3.0%
Other values (106)	902641	33.6%

**Most occurring characters****Value**

Value	Count	Frequency (%)
E	2379311	12.3%
I	1872460	9.7%
N	1640913	8.5%
O	1463185	7.5%
T	1408342	7.3%
L	1236970	6.4%
R	1181395	6.1%
A	1119969	5.8%
G	1068916	5.5%
	720628	3.7%

Value	Count	Frequency (%)
Other values (23)	5302286	27.3%

**Most occurring categories**

Value	Count	Frequency (%)
(unknown)	19394375	100.0%

**Most frequent character per category**

Value	Count	Frequency (%)
E	2379311	12.3%
I	1872460	9.7%
N	1640913	8.5%
O	1463185	7.5%
T	1408342	7.3%
L	1236970	6.4%
R	1181395	6.1%
A	1119969	5.8%
G	1068916	5.5%
Other values (23)	720628	3.7%
	5302286	27.3%

**Most occurring scripts**

Value	Count	Frequency (%)
(unknown)	19394375	100.0%

**Most frequent character per script**

Value	Count	Frequency (%)
E	2379311	12.3%
I	1872460	9.7%
N	1640913	8.5%
O	1463185	7.5%
T	1408342	7.3%
L	1236970	6.4%
R	1181395	6.1%
A	1119969	5.8%

Value	Count	Frequency (%)
A	1068916	5.5%
G	720628	3.7%
Other values (23)	5302286	27.3%

**Most occurring blocks**

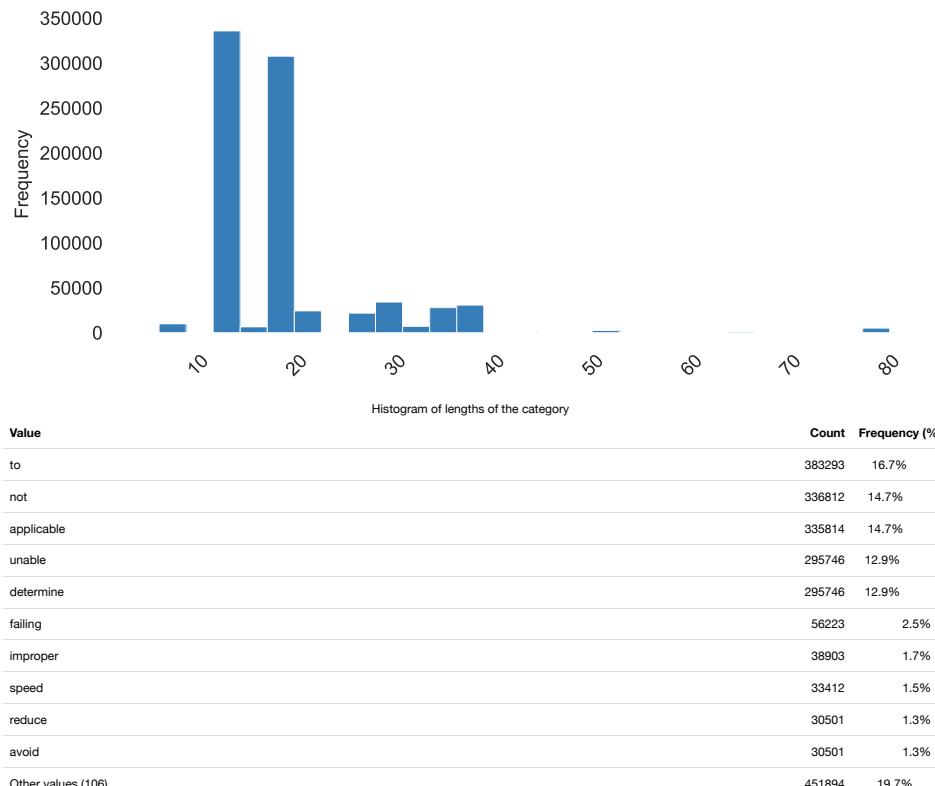
Value	Count	Frequency (%)
(unknown)	19394375	100.0%

**Most frequent character per block**

Value	Count	Frequency (%)
E	2379311	12.3%
I	1872460	9.7%
N	1640913	8.5%
O	1463185	7.5%
T	1408342	7.3%
L	1236970	6.4%
R	1181395	6.1%
A	1119969	5.8%
G	1068916	5.5%
Other values (23)	720628	3.7%
	5302286	27.3%

SEC_CONTRIBUTORY_CAUSE			
Categorical			
IMBALANCE			
Distinct	40		
Distinct (%)	< 0.1%		
Missing	0		
Missing (%)	0.0%		
Memory size	6.2 MiB		
Length	Characters and Unicode	Unique	Sample
Max length	80	Total characters	15926593
Median length	75	Unique	0
Mean length	19.476758	Unique (%)	0.0%
Min length	6	Distinct characters	33
		Distinct categories	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )
		Distinct scripts	1 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )
		Distinct blocks	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )
The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.			
Common Values			
Value	Count	Frequency (%)	
NOT APPLICABLE	335814	41.1%	
UNABLE TO DETERMINE	295746	36.2%	
FAILING TO REDUCE SPEED TO AVOID CRASH	30501	3.7%	
FAILING TO YIELD RIGHT-OF-WAY	25722	3.1%	
DRIVING SKILLS/KNOWLEDGE/EXPERIENCE	25124	3.1%	
FOLLOWING TOO CLOSELY	21547	2.6%	
IMPROPER OVERTAKING/PASSING	12508	1.5%	
IMPROPER LANE USAGE	11477	1.4%	
WEATHER	9434	1.2%	
IMPROPER TURNING/NO SIGNAL	8350	1.0%	
Other values (30)	41500	5.1%	

Length



Most occurring characters

Value	Count	Frequency (%)
E	2092865	13.1%
	1471122	9.2%
L	1277925	8.0%
N	1264704	7.9%
A	1252400	7.9%
I	1174220	7.4%
T	1157151	7.3%
O	1032677	6.5%
P	833672	5.2%
B	648046	4.1%

Value	Count	Frequency (%)
Other values (23)	3721811	23.4%

**Most occurring categories**

Value	Count	Frequency (%)
(unknown)	15926593	100.0%

**Most frequent character per category**

Value	Count	Frequency (%)
E	2092865	13.1%
L	1471122	9.2%
N	1277925	8.0%
A	1264704	7.9%
I	1252400	7.9%
T	1174220	7.4%
O	1157151	7.3%
P	1032677	6.5%
B	833672	5.2%
Other values (23)	648046	4.1%
	3721811	23.4%

**Most occurring scripts**

Value	Count	Frequency (%)
(unknown)	15926593	100.0%

**Most frequent character per script**

Value	Count	Frequency (%)
E	2092865	13.1%
L	1471122	9.2%
N	1277925	8.0%
A	1264704	7.9%
I	1252400	7.9%
T	1174220	7.4%
O	1157151	7.3%
P	1032677	6.5%

Value	Count	Frequency (%)
P	833672	5.2%
B	648046	4.1%
Other values (23)	3721811	23.4%

**Most occurring blocks**

Value	Count	Frequency (%)
(unknown)	15926593	100.0%

**Most frequent character per block**

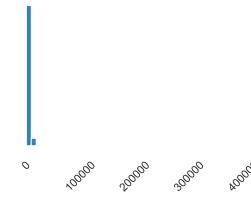
Value	Count	Frequency (%)
E	2092865	13.1%
L	1471122	9.2%
N	1277925	8.0%
A	1264704	7.9%
I	1252400	7.9%
T	1174220	7.4%
O	1157151	7.3%
P	1032677	6.5%
B	833672	5.2%
Other values (23)	648046	4.1%
	3721811	23.4%

**STREET\_NO**

Real number (R)

Distinct	11728
Distinct (%)	1.4%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	3689.6337

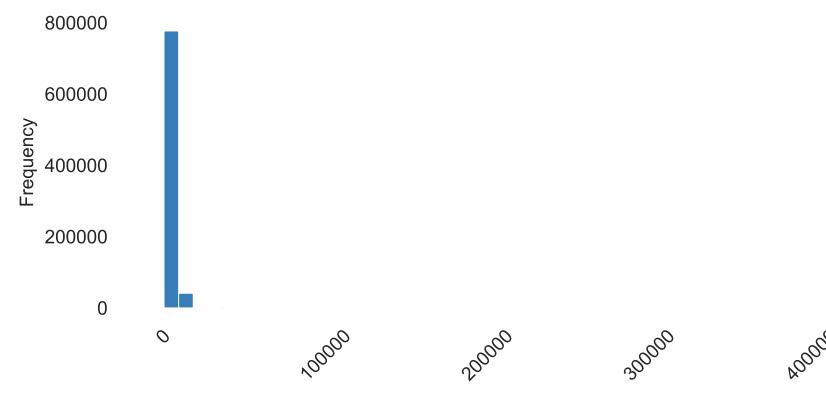
Minimum	0
Maximum	451100
Zeros	5
Zeros (%)	< 0.1%
Negative	0
Negative (%)	0.0%

**Quantile statistics**

Minimum	0
5-th percentile	150
Q1	1250
median	3201
Q3	5600
95-th percentile	9041
Maximum	451100
Range	451100
Interquartile range (IQR)	4350

**Descriptive statistics**

Standard deviation	2886.8844
Coefficient of variation (CV)	0.78243115
Kurtosis	705.36663
Mean	3689.6337
Median Absolute Deviation (MAD)	2101
Skewness	5.2745308
Sum	$3.0170984 \times 10^9$
Variance	8334101.4
Monotonicity	Not monotonic



Value	Count	Frequency (%)
1600	5262	0.6%
100	4886	0.6%
800	4779	0.6%
200	4525	0.6%
300	4114	0.5%

**Value**

Value	Count	Frequency (%)
2400	3998	0.5%
4700	3997	0.5%
1200	3927	0.5%
500	3924	0.5%
6300	3796	0.5%
Other values (11718)	774515	94.7%

**Value**

Value	Count	Frequency (%)
0	5	< 0.1%
1	3418	0.4%
2	1660	0.2%
3	667	0.1%
4	144	< 0.1%
5	599	0.1%
6	202	< 0.1%
7	166	< 0.1%
8	194	< 0.1%
9	175	< 0.1%

**Value**

Value	Count	Frequency (%)
451100	1	< 0.1%
34453	1	< 0.1%
13799	5	< 0.1%
13795	1	< 0.1%
13787	1	< 0.1%
13781	1	< 0.1%
13780	1	< 0.1%
13770	27	< 0.1%
13768	1	< 0.1%
13763	1	< 0.1%

### STREET\_DIRECTION

Categorical

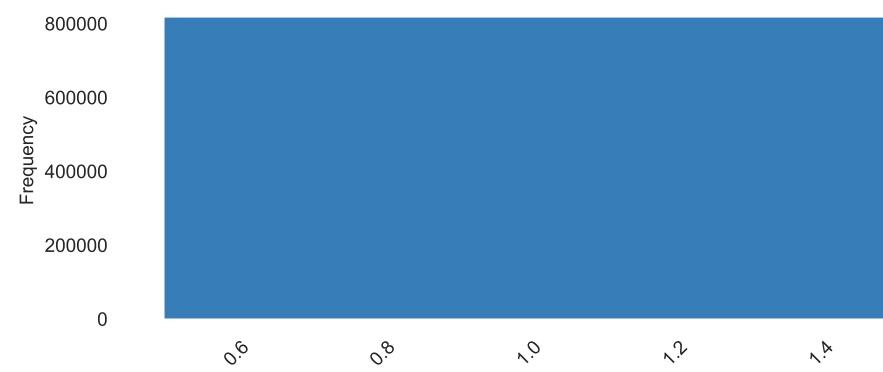
<b>Distinct</b>	4
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	4
<b>Missing (%)</b>	< 0.1%
<b>Memory size</b>	6.2 MiB
<b>Length</b>	
<b>Max length</b>	1
<b>Median length</b>	1
<b>Mean length</b>	1
<b>Min length</b>	1
<b>Characters and Unicode</b>	
Total characters	817719
Distinct characters	4
Distinct categories	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )
Distinct scripts	1 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )
Distinct blocks	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )
<b>Unique</b>	
Unique	0 ?
Unique (%)	0.0%
<b>Sample</b>	
1st row	W
2nd row	S
3rd row	N
4th row	W
5th row	W

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

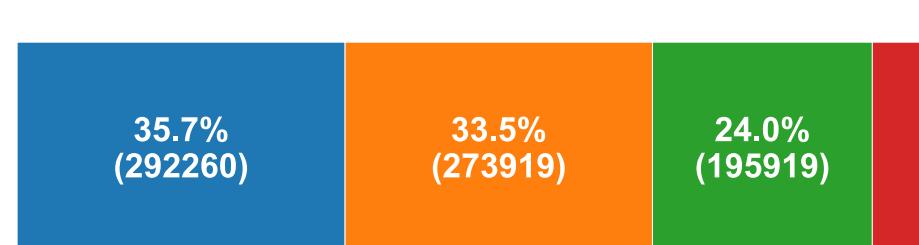
**Common Values**

Value	Count	Frequency (%)
W	292260	35.7%
S	273919	33.5%
N	195919	24.0%
E	55621	6.8%
(Missing)	4	< 0.1%

**Length**



Common Values (Plot)



Value	Count	Frequency (%)
w	292260	35.7%
s	273919	33.5%
n	195919	24.0%
e	55621	6.8%

Most occurring characters

## Austin Crash Report Data - Profiling Report

Value	Count	Frequency (%)
W	292260	35.7%
S	273919	33.5%
N	195919	24.0%
E	55621	6.8%

## Most occurring categories

Value	Count	Frequency (%)
(unknown)	817719	100.0%

## Most frequent character per category

Value	Count	Frequency (%)
W	292260	35.7%
S	273919	33.5%
N	195919	24.0%
E	55621	6.8%

## Most occurring scripts

Value	Count	Frequency (%)
(unknown)	817719	100.0%

## Most frequent character per script

Value	Count	Frequency (%)
W	292260	35.7%
S	273919	33.5%
N	195919	24.0%
E	55621	6.8%

## Most occurring blocks

Value	Count	Frequency (%)
(unknown)	817719	100.0%

## Most frequent character per block

Value	Count	Frequency (%)
(unknown)	817719	100.0%

## Austin Crash Report Data - Profiling Report

Value	Count	Frequency (%)
W	292260	35.7%
S	273919	33.5%
N	195919	24.0%
E	55621	6.8%



(unknown)		Count	Frequency (%)
E		1000603	11.5%
A		953307	10.9%
		930244	10.7%
T		620128	7.1%
S		576159	6.6%
R		559100	6.4%
V		506123	5.8%
N		438388	5.0%
L		407720	4.7%
O		365313	4.2%
Other values (30)		2375067	27.2%

**Most occurring blocks**

Value	Count	Frequency (%)
(unknown)	8732152	100.0%

**Most frequent character per block**

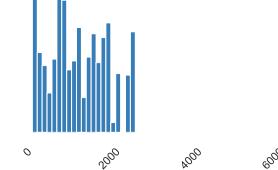
(unknown)	Count	Frequency (%)
E	1000603	11.5%
A	953307	10.9%
	930244	10.7%
T	620128	7.1%
S	576159	6.6%
R	559100	6.4%
V	506123	5.8%
N	438388	5.0%
L	407720	4.7%
O	365313	4.2%
Other values (30)	2375067	27.2%

**BEAT\_OF\_OCCURRENCE**

Real number (R)

Distinct	276
Distinct (%)	< 0.1%
Missing	5
Missing (%)	< 0.1%
Infinite	0
Infinite (%)	0.0%
Mean	1243.7345

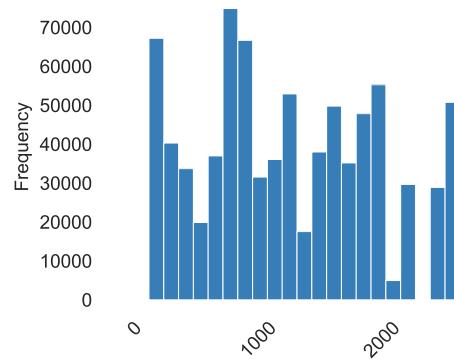
Minimum	111
Maximum	6100
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	6.2 MiB

**Quantile statistics**

Minimum	111
5-th percentile	131
Q1	714
median	1212
Q3	1822
95-th percentile	2513
Maximum	6100
Range	5989
Interquartile range (IQR)	1108

**Descriptive statistics**

Standard deviation	705.2713
Coefficient of variation (CV)	0.56705936
Kurtosis	-0.99710383
Mean	1243.7345
Median Absolute Deviation (MAD)	580
Skewness	0.17926577
Sum	$1.0170241 \times 10^9$
Variance	497407.61
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
1834	9942	1.2%
114	8369	1.0%
813	8217	1.0%
815	7798	1.0%
1831	7587	0.9%

Value	Count	Frequency (%)
122	7159	0.9%
833	6721	0.8%
834	6215	0.8%
2413	6142	0.8%
2512	6007	0.7%
Other values (266)	743561	90.9%

Value	Count	Frequency (%)
111	3557	0.4%
112	2497	0.3%
113	1908	0.2%
114	8369	1.0%
121	3831	0.5%
122	7159	0.9%
123	4979	0.6%
124	4277	0.5%
131	4862	0.6%
132	4970	0.6%

Value	Count	Frequency (%)
6100	4	< 0.1%
2535	2598	0.3%
2534	3472	0.4%
2533	5774	0.7%
2532	2293	0.3%
2531	2200	0.3%
2525	1791	0.2%
2524	2583	0.3%
2523	2860	0.3%
2522	3418	0.4%

**PHOTOS\_TAKEN\_I**

Boolean

MISSING

Distinct	2
Distinct (%)	< 0.1%
Missing	806948
Missing (%)	98.7%

Memory size 1.6 MiB

Value	Count	Frequency (%)
True	8110	1.0%
False	2665	0.3%
(Missing)	806948	98.7%

**75.3%**  
**(8110)**

**24.7%**  
**(2665)**



**STATEMENTS\_TAKEN\_I**

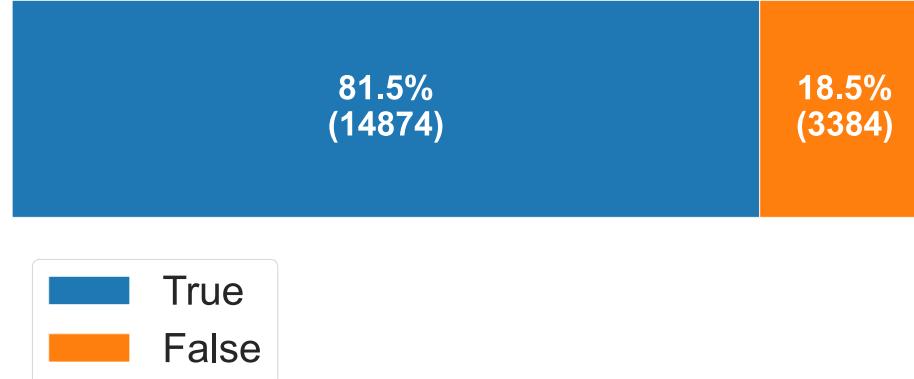
Boolean

MISSING

<b>Distinct</b>	2
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	799465
<b>Missing (%)</b>	97.8%

**Memory size** 1.6 MiB

Value	Count	Frequency (%)
True	14874	1.8%
False	3384	0.4%
(Missing)	799465	97.8%

**DOORING\_I**

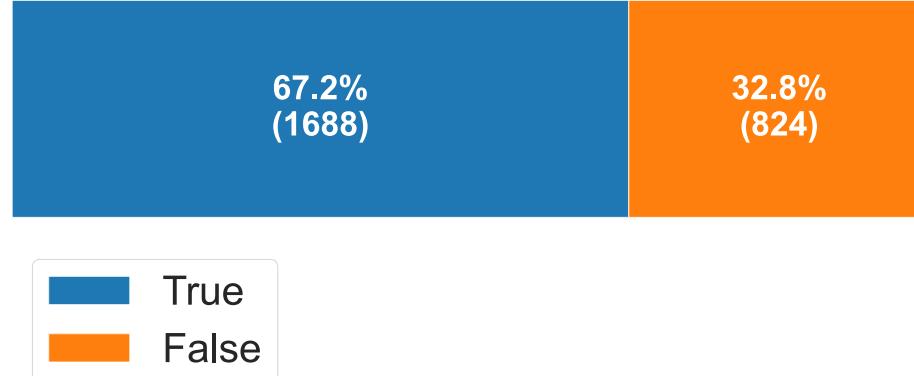
Boolean

MISSING

<b>Distinct</b>	2
<b>Distinct (%)</b>	0.1%
<b>Missing</b>	815211
<b>Missing (%)</b>	99.7%

**Memory size** 1.6 MiB

Value	Count	Frequency (%)
True	1688	0.2%
False	824	0.1%
(Missing)	815211	99.7%



**WORK\_ZONE\_I**

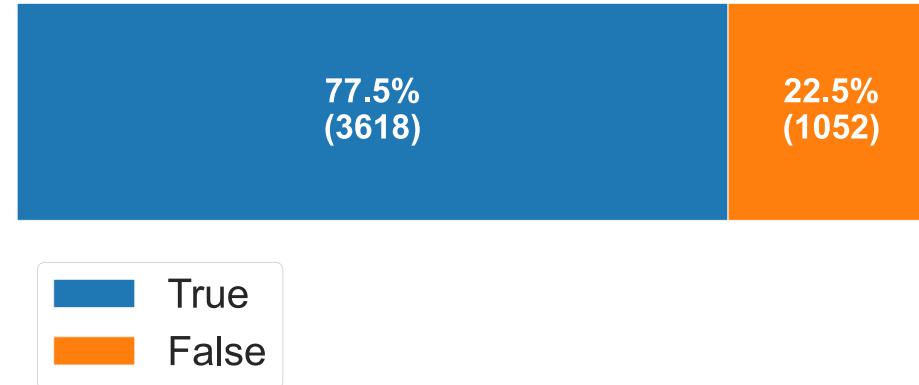
Boolean

MISSING

Distinct	2
Distinct (%)	< 0.1%
Missing	813053
Missing (%)	99.4%

Memory size 1.6 MiB

Value	Count	Frequency (%)
True	3618	0.4%
False	1052	0.1%
(Missing)	813053	99.4%

**WORK\_ZONE\_TYPE**

Categorical

MISSING

Distinct	4
Distinct (%)	0.1%
Missing	814105
Missing (%)	99.6%

Memory size 6.2 MiB

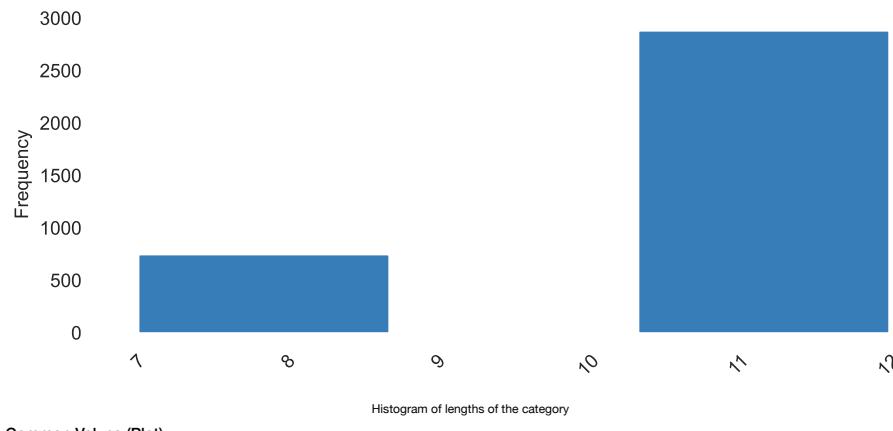
Length	Characters and Unicode	Unique	Sample
Max length	12	0	?
Median length	12	0.0%	1st row CONSTRUCTION
Mean length	10.872305	0.0%	2nd row CONSTRUCTION
Min length	7	0.0%	3rd row CONSTRUCTION
Distinct categories	15	0.0%	4th row CONSTRUCTION
Distinct scripts	1	0.0%	5th row CONSTRUCTION
Distinct blocks	1	0.0%	CONSTRUCTION

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

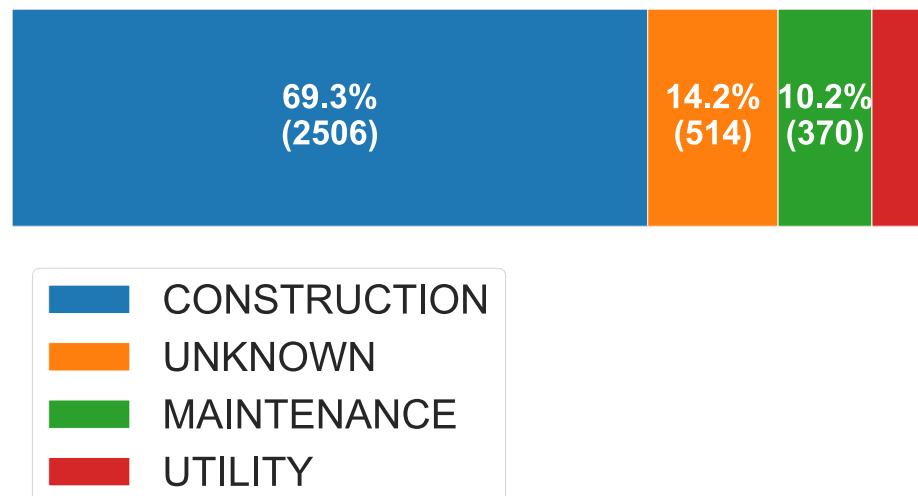
## Common Values

Value	Count	Frequency (%)
CONSTRUCTION	2506	0.3%
UNKNOWN	514	0.1%
MAINTENANCE	370	< 0.1%
UTILITY	228	< 0.1%
(Missing)	814105	99.6%

## Length



Common Values (Plot)



Value	Count	Frequency (%)
construction	2506	69.3%
unknown	514	14.2%
maintenance	370	10.2%
utility	228	6.3%

## Most occurring characters

Value	Count	Frequency (%)
N	7664	19.5%
T	5838	14.8%
O	5526	14.0%
C	5382	13.7%
I	3332	8.5%
U	3248	8.3%
S	2506	6.4%
R	2506	6.4%
A	740	1.9%
E	740	1.9%
Other values (5)	1854	4.7%

## Most occurring categories

Value	Count	Frequency (%)
(unknown)	39336	100.0%

## Most frequent character per category

(unknown)	Value	Count	Frequency (%)
N		7664	19.5%
T		5838	14.8%
O		5526	14.0%
C		5382	13.7%
I		3332	8.5%
U		3248	8.3%
S		2506	6.4%
R		2506	6.4%
A		740	1.9%
E		740	1.9%
Other values (5)		1854	4.7%

## Most occurring scripts

(unknown)	Value	Count	Frequency (%)
(unknown)		39336	100.0%

## Most frequent character per script

(unknown)	Value	Count	Frequency (%)
N		7664	19.5%
T		5838	14.8%
O		5526	14.0%
C		5382	13.7%
I		3332	8.5%
U		3248	8.3%
S		2506	6.4%
R		2506	6.4%
A		740	1.9%
E		740	1.9%
Other values (5)		1854	4.7%

**Most occurring blocks**

(unknown)	Value	Count	Frequency (%)
	(unknown)	39336	100.0%

**Most frequent character per block**

(unknown)	Value	Count	Frequency (%)
N		7664	19.5%
T		5838	14.8%
O		5526	14.0%
C		5382	13.7%
I		3332	8.5%
U		3248	8.3%
S		2506	6.4%
R		2506	6.4%
A		740	1.9%
E		740	1.9%
Other values (5)		1854	4.7%

**WORKERS\_PRESENT\_I**

Boolean

MISSING

Distinct	2
Distinct (%)	0.2%
Missing	816529
Missing (%)	99.9%

Memory size

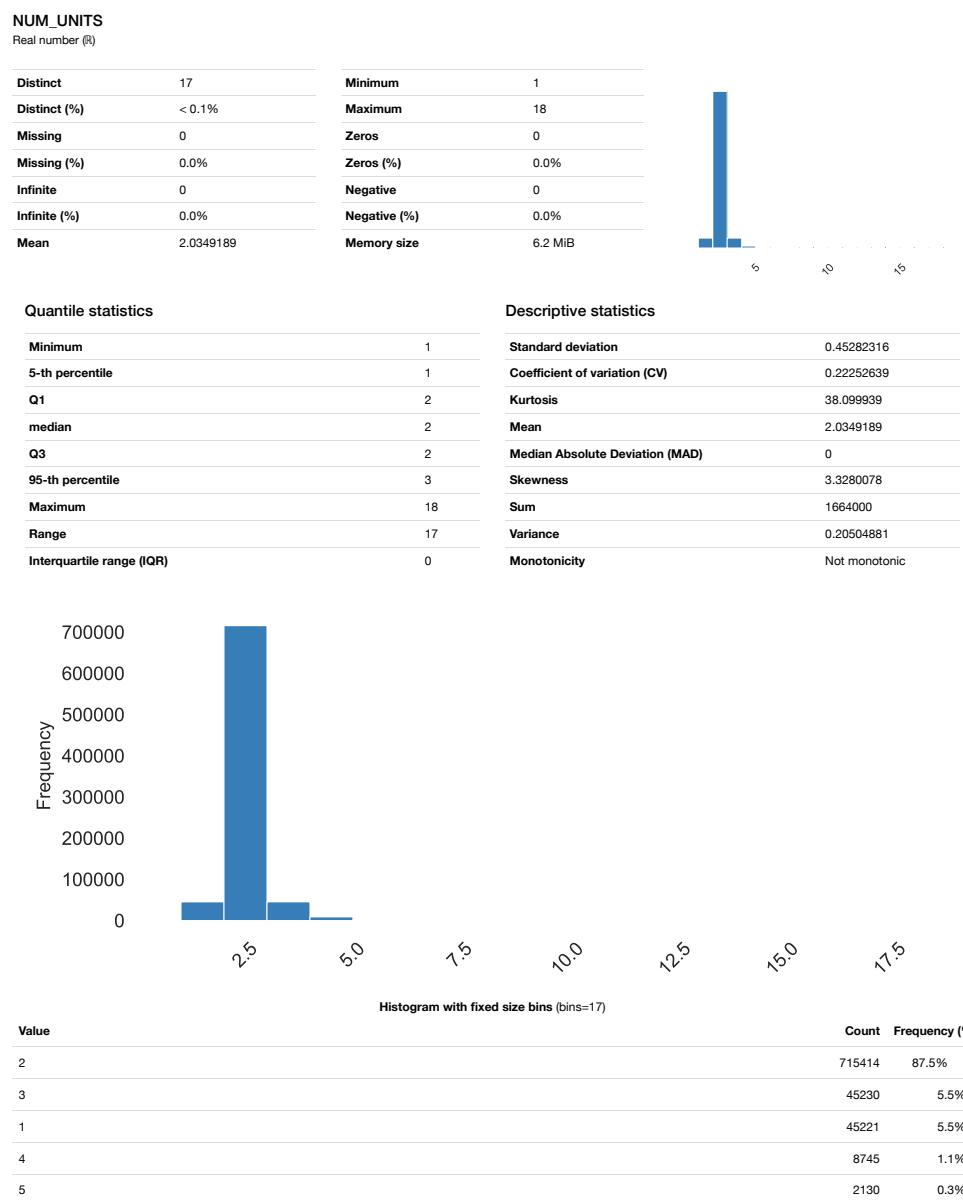
1.6 MiB

Value	Count	Frequency (%)
True	1056	0.1%
False	138	< 0.1%
(Missing)	816529	99.9%

**88.4%**  
**(1056)**

**11.6%**  
**(138)**





Value	Count	Frequency (%)
6	603	0.1%
7	210	< 0.1%
8	95	< 0.1%
9	38	< 0.1%
10	16	< 0.1%
Other values (7)	21	< 0.1%

Value	Count	Frequency (%)
1	45221	5.5%
2	715414	87.5%
3	45230	5.5%
4	8745	1.1%
5	2130	0.3%
6	603	0.1%
7	210	< 0.1%
8	95	< 0.1%
9	38	< 0.1%
10	16	< 0.1%

Value	Count	Frequency (%)
18	4	< 0.1%
16	1	< 0.1%
15	1	< 0.1%
14	2	< 0.1%
13	1	< 0.1%
12	5	< 0.1%
11	7	< 0.1%
10	16	< 0.1%
9	38	< 0.1%
8	95	< 0.1%

**MOST\_SEVERE\_INJURY**

Categorical

IMBALANCE

Distinct	5
Distinct (%)	< 0.1%
Missing	1792
Missing (%)	0.2%

Memory size

6.2 MiB

Length

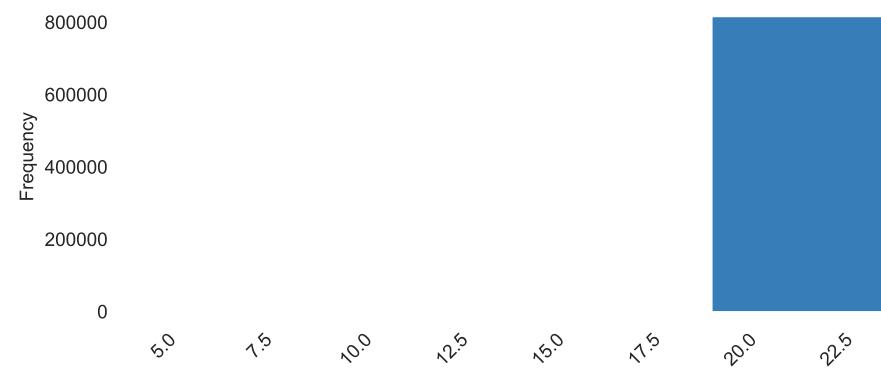
	Characters and Unicode	Unique	Sample
Total characters	18717388	Unique	0 ?
Distinct characters	19	Unique (%)	0.0%
Distinct categories	1		? 1st row NONINCAPACITATING INJURY
Distinct scripts	1		? 2nd row NO INDICATION OF INJURY
Distinct blocks	1		? 3rd row NONINCAPACITATING INJURY
(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)			? 4th row FATAL
(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)			? 5th row NO INDICATION OF INJURY

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

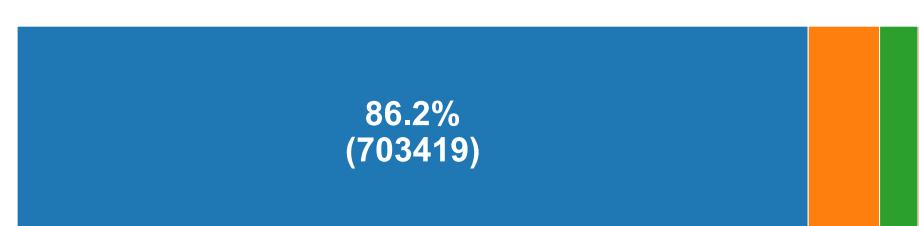
## Common Values

Value	Count	Frequency (%)
NO INDICATION OF INJURY	703419	86.0%
NONINCAPACITATING INJURY	63461	7.8%
REPORTED, NOT EVIDENT	34331	4.2%
INCAPACITATING INJURY	13821	1.7%
FATAL	899	0.1%
(Missing)	1792	0.2%

## Length



Common Values (Plot)



<span style="color: blue;">█</span>	NO INDICATION OF INJURY
<span style="color: orange;">█</span>	NONINCAPACITATING INJURY
<span style="color: green;">█</span>	REPORTED, NOT EVIDENT
<span style="color: red;">█</span>	INCAPACITATING INJURY
<span style="color: purple;">█</span>	FATAL

Value	Count	Frequency (%)
injury	780701	25.4%
no	703419	22.9%
indication	703419	22.9%
of	703419	22.9%
nonincapacitating	63461	2.1%

Value	Count	Frequency (%)
reported	34331	1.1%
not	34331	1.1%
evident	34331	1.1%
incapacitating	13821	0.4%
fatal	899	< 0.1%

**Most occurring characters**

Value	Count	Frequency (%)
N	3241106	17.3%
I	3157135	16.9%
	2256201	12.1%
O	2242380	12.0%
T	961875	5.1%
A	937063	5.0%
C	857983	4.6%
R	849363	4.5%
U	780701	4.2%
Y	780701	4.2%
Other values (9)	2652880	14.2%

**Most occurring categories**

Value	Count	Frequency (%)
(unknown)	18717388	100.0%

**Most frequent character per category**

(unknown)	Value	Count	Frequency (%)
N		3241106	17.3%
I		3157135	16.9%
	2256201	12.1%	
O		2242380	12.0%
T		961875	5.1%
A		937063	5.0%
C		857983	4.6%
R		849363	4.5%
U		780701	4.2%

Value	Count	Frequency (%)
Y	780701	4.2%
Other values (9)	2652880	14.2%

**Most occurring scripts**

Value	Count	Frequency (%)
(unknown)	18717388	100.0%

**Most frequent character per script**

(unknown)	Value	Count	Frequency (%)
N		3241106	17.3%
I		3157135	16.9%
	2256201	12.1%	
O		2242380	12.0%
T		961875	5.1%
A		937063	5.0%
C		857983	4.6%
R		849363	4.5%
U		780701	4.2%
Y		780701	4.2%
Other values (9)		2652880	14.2%

**Most occurring blocks**

Value	Count	Frequency (%)
(unknown)	18717388	100.0%

**Most frequent character per block**

(unknown)	Value	Count	Frequency (%)
N		3241106	17.3%
I		3157135	16.9%
	2256201	12.1%	
O		2242380	12.0%
T		961875	5.1%
A		937063	5.0%
C		857983	4.6%

Value	Count	Frequency (%)
R	849363	4.5%
U	780701	4.2%
Y	780701	4.2%
Other values (9)	2652880	14.2%

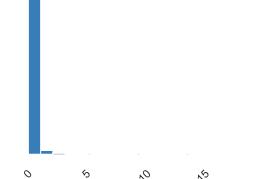
**INJURIES\_TOTAL**

Real number (R)

ZEROES

<b>Distinct</b>	20
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	1780
<b>Missing (%)</b>	0.2%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	0.18992381

<b>Minimum</b>	0
<b>Maximum</b>	21
<b>Zeros</b>	703431
<b>Zeros (%)</b>	86.0%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%
<b>Memory size</b>	6.2 MiB

**Quantile statistics**

<b>Minimum</b>	0
<b>5-th percentile</b>	0
<b>Q1</b>	0
<b>median</b>	0
<b>Q3</b>	0
<b>95-th percentile</b>	1
<b>Maximum</b>	21
<b>Range</b>	21
<b>Interquartile range (IQR)</b>	0

**Descriptive statistics**

<b>Standard deviation</b>	0.56603807
<b>Coefficient of variation (CV)</b>	2.9803429
<b>Kurtosis</b>	45.551048
<b>Mean</b>	0.18992381
<b>Median Absolute Deviation (MAD)</b>	0
<b>Skewness</b>	4.8263445
<b>Sum</b>	154967
<b>Variance</b>	0.32039909
<b>Monotonicity</b>	Not monotonic



Histogram with fixed size bins (bins=20)

Value	Count	Frequency (%)
0	703431	86.0%
1	84676	10.4%
2	18829	2.3%
3	5721	0.7%
4	2050	0.3%

Value	Count	Frequency (%)
5	723	0.1%
6	289	< 0.1%
7	110	< 0.1%
8	45	< 0.1%
9	26	< 0.1%
Other values (10)	43	< 0.1%
(Missing)	1780	0.2%

Value	Count	Frequency (%)
0	703431	86.0%
1	84676	10.4%
2	18829	2.3%
3	5721	0.7%
4	2050	0.3%
5	723	0.1%
6	289	< 0.1%
7	110	< 0.1%
8	45	< 0.1%
9	26	< 0.1%

Value	Count	Frequency (%)
21	4	< 0.1%
19	1	< 0.1%
17	1	< 0.1%
16	1	< 0.1%
15	7	< 0.1%
14	1	< 0.1%
13	2	< 0.1%
12	4	< 0.1%
11	8	< 0.1%
10	14	< 0.1%

**INJURIES\_FATAL**

Categorical

## IMBALANCE

Distinct	5
Distinct (%)	< 0.1%
Missing	1780
Missing (%)	0.2%

Memory size 6.2 MiB

## Length

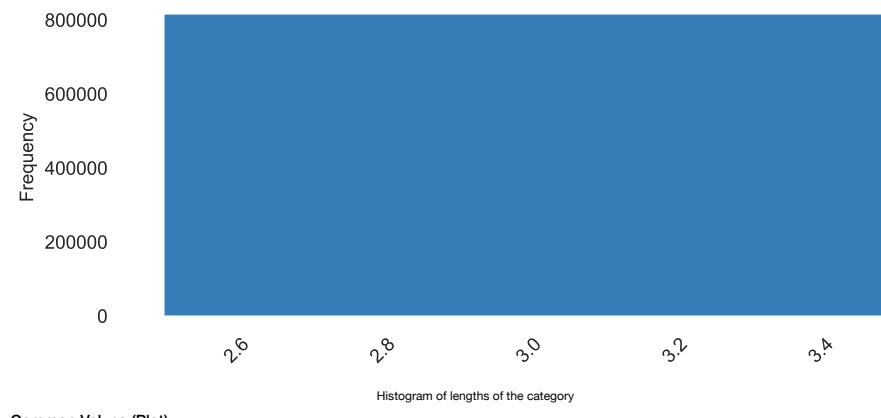
	Characters and Unicode	Unique		Sample
Total characters	2447829	Unique	1	1st row 0.0
Distinct characters	6	Unique (%)	< 0.1%	2nd row 0.0
Distinct categories	1			3rd row 0.0
Distinct scripts	1			4th row 1.0
Distinct blocks	1			5th row 0.0

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

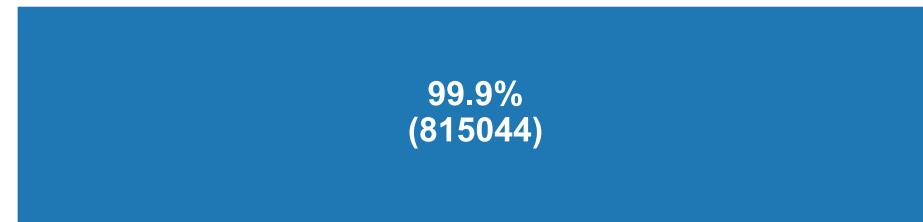
## Common Values

Value	Count	Frequency (%)
0.0	815044	99.7%
1.0	836	0.1%
2.0	54	< 0.1%
3.0	8	< 0.1%
4.0	1	< 0.1%
(Missing)	1780	0.2%

## Length



Common Values (Plot)



Value	Count	Frequency (%)
0.0	815044	99.9%
1.0	836	0.1%
2.0	54	< 0.1%
3.0	8	< 0.1%
4.0	1	< 0.1%

**Most occurring characters**

Value	Count	Frequency (%)
0	1630987	66.6%
.	815943	33.3%
1	836	< 0.1%
2	54	< 0.1%
3	8	< 0.1%
4	1	< 0.1%

**Most occurring categories**

Value	Count	Frequency (%)
(unknown)	2447829	100.0%

**Most frequent character per category**

Value	Count	Frequency (%)
0	1630987	66.6%
.	815943	33.3%
1	836	< 0.1%
2	54	< 0.1%
3	8	< 0.1%
4	1	< 0.1%

**Most occurring scripts**

Value	Count	Frequency (%)
(unknown)	2447829	100.0%

**Most frequent character per script**

Value	Count	Frequency (%)
0	1630987	66.6%
.	815943	33.3%
1	836	< 0.1%
2	54	< 0.1%
3	8	< 0.1%
4	1	< 0.1%

**Most occurring blocks**

Value	Count	Frequency (%)
(unknown)	244789	100.0%

**Most frequent character per block**

(unknown)	Value	Count	Frequency (%)
0		1630987	66.6%
.		815943	33.3%
1		836	< 0.1%
2		54	< 0.1%
3		8	< 0.1%
4		1	< 0.1%

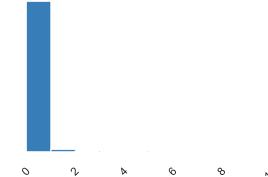
**INJURIES\_INCAPACITATING**

Real number (R)

ZEROES

<b>Distinct</b>	10
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	1780
<b>Missing (%)</b>	0.2%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	0.019908253

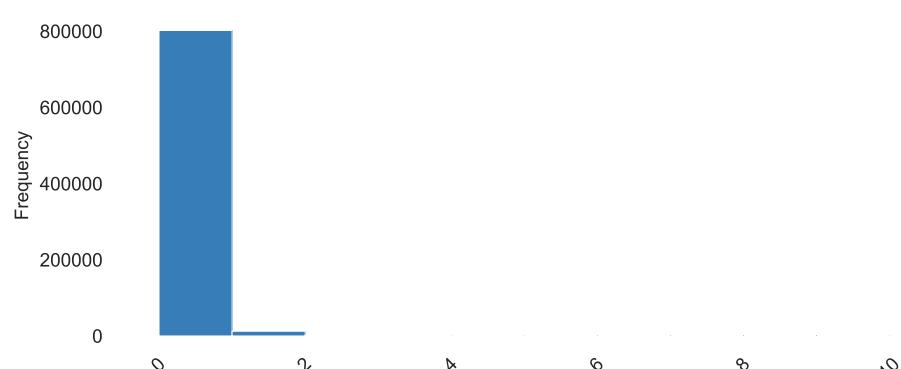
<b>Minimum</b>	0
<b>Maximum</b>	10
<b>Zeros</b>	801987
<b>Zeros (%)</b>	98.1%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%
<b>Memory size</b>	6.2 MiB

**Quantile statistics**

<b>Minimum</b>	0
<b>5-th percentile</b>	0
<b>Q1</b>	0
<b>median</b>	0
<b>Q3</b>	0
<b>95-th percentile</b>	0
<b>Maximum</b>	10
<b>Range</b>	10
<b>Interquartile range (IQR)</b>	0

**Descriptive statistics**

<b>Standard deviation</b>	0.16511357
<b>Coefficient of variation (CV)</b>	8.2937243
<b>Kurtosis</b>	193.15979
<b>Mean</b>	0.019908253
<b>Median Absolute Deviation (MAD)</b>	0
<b>Skewness</b>	11.32382
<b>Sum</b>	16244
<b>Variance</b>	0.02726249
<b>Monotonicity</b>	Not monotonic

**Frequency**

Value	Count	Frequency (%)
5	26	< 0.1%
6	7	< 0.1%
7	1	< 0.1%
10	1	< 0.1%
8	1	< 0.1%
(Missing)	1780	0.2%

Value	Count	Frequency (%)
0	801987	98.1%
1	12271	1.5%
2	1269	0.2%
3	282	< 0.1%
4	98	< 0.1%
5	26	< 0.1%
6	7	< 0.1%
7	1	< 0.1%
8	1	< 0.1%
10	1	< 0.1%

Value	Count	Frequency (%)
10	1	< 0.1%
8	1	< 0.1%
7	1	< 0.1%
6	7	< 0.1%
5	26	< 0.1%
4	98	< 0.1%
3	282	< 0.1%
2	1269	0.2%
1	12271	1.5%
0	801987	98.1%

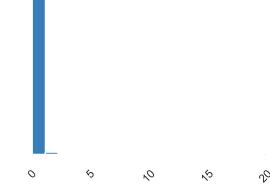
**INJURIES\_NON\_INCAPACITATING**

Real number (R)

ZEROES

<b>Distinct</b>	19
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	1780
<b>Missing (%)</b>	0.2%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	0.10711778

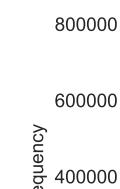
<b>Minimum</b>	0
<b>Maximum</b>	21
<b>Zeros</b>	749965
<b>Zeros (%)</b>	91.7%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%
<b>Memory size</b>	6.2 MiB

**Quantile statistics**

<b>Minimum</b>	0
<b>5-th percentile</b>	0
<b>Q1</b>	0
<b>median</b>	0
<b>Q3</b>	0
<b>95-th percentile</b>	1
<b>Maximum</b>	21
<b>Range</b>	21
<b>Interquartile range (IQR)</b>	0

**Descriptive statistics**

<b>Standard deviation</b>	0.42226344
<b>Coefficient of variation (CV)</b>	3.9420482
<b>Kurtosis</b>	79.874379
<b>Mean</b>	0.10711778
<b>Median Absolute Deviation (MAD)</b>	0
<b>Skewness</b>	6.3551825
<b>Sum</b>	87402
<b>Variance</b>	0.17830641
<b>Monotonicity</b>	Not monotonic



Histogram with fixed size bins (bins=19)

Value	Count	Frequency (%)
0	749965	91.7%
1	51685	6.3%
2	9835	1.2%
3	2873	0.4%
4	999	0.1%

Value	Count	Frequency (%)
5	348	< 0.1%
6	141	< 0.1%
7	46	< 0.1%
8	22	< 0.1%
10	9	< 0.1%
Other values (9)	20	< 0.1%
(Missing)	1780	0.2%

Value	Count	Frequency (%)
0	749965	91.7%
1	51685	6.3%
2	9835	1.2%
3	2873	0.4%
4	999	0.1%
5	348	< 0.1%
6	141	< 0.1%
7	46	< 0.1%
8	22	< 0.1%
9	6	< 0.1%

Value	Count	Frequency (%)
21	2	< 0.1%
19	1	< 0.1%
18	1	< 0.1%
16	1	< 0.1%
15	1	< 0.1%
14	1	< 0.1%
12	3	< 0.1%
11	4	< 0.1%
10	9	< 0.1%
9	6	< 0.1%

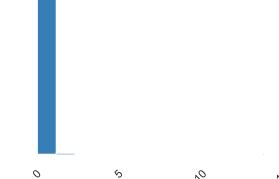
## INJURIES\_REPORTED\_NOT\_EVIDENT

Real number (R)

ZEROES

Distinct	13
Distinct (%)	< 0.1%
Missing	1780
Missing (%)	0.2%
Infinite	0
Infinite (%)	0.0%
Mean	0.061706516

Minimum	0
Maximum	15
Zeros	777695
Zeros (%)	95.1%
Negative	0
Negative (%)	0.0%
Memory size	6.2 MiB

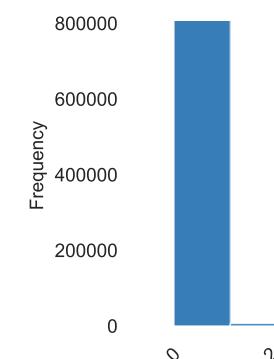


## Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	0
Maximum	15
Range	15
Interquartile range (IQR)	0

## Descriptive statistics

Standard deviation	0.31924399
Coefficient of variation (CV)	5.1735864
Kurtosis	96.640062
Mean	0.061706516
Median Absolute Deviation (MAD)	0
Skewness	7.6941999
Sum	50349
Variance	0.10191673
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=13)

Value	Count	Frequency (%)
0	777695	95.1%
1	29752	3.6%
2	6145	0.8%
3	1586	0.2%
4	496	0.1%

Value	Count	Frequency (%)
5	172	< 0.1%
6	45	< 0.1%
7	22	< 0.1%
8	11	< 0.1%
9	9	< 0.1%
Other values (3)	10	< 0.1%
(Missing)	1780	0.2%

Value	Count	Frequency (%)
0	777695	95.1%
1	29752	3.6%
2	6145	0.8%
3	1586	0.2%
4	496	0.1%
5	172	< 0.1%
6	45	< 0.1%
7	22	< 0.1%
8	11	< 0.1%
9	9	< 0.1%

Value	Count	Frequency (%)
15	2	< 0.1%
11	2	< 0.1%
10	6	< 0.1%
9	9	< 0.1%
8	11	< 0.1%
7	22	< 0.1%
6	45	< 0.1%
5	172	< 0.1%
4	496	0.1%
3	1586	0.2%

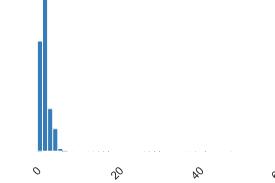
**INJURIES\_NO\_INDICATION**

Real number (R)

ZEROES

<b>Distinct</b>	48
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	1780
<b>Missing (%)</b>	0.2%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	2.0032404

<b>Minimum</b>	0
<b>Maximum</b>	61
<b>Zeros</b>	17119
<b>Zeros (%)</b>	2.1%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%
<b>Memory size</b>	6.2 MiB

**Quantile statistics**

<b>Minimum</b>	0
<b>5-th percentile</b>	1
<b>Q1</b>	1
<b>median</b>	2
<b>Q3</b>	2
<b>95-th percentile</b>	4
<b>Maximum</b>	61
<b>Range</b>	61
<b>Interquartile range (IQR)</b>	1

**Descriptive statistics**

<b>Standard deviation</b>	1.1571173
<b>Coefficient of variation (CV)</b>	0.57762278
<b>Kurtosis</b>	69.611204
<b>Mean</b>	2.0032404
<b>Median Absolute Deviation (MAD)</b>	1
<b>Skewness</b>	3.763925
<b>Sum</b>	1634530
<b>Variance</b>	1.3389205
<b>Monotonicity</b>	Not monotonic



Histogram with fixed size bins (bins=48)

Value	Count	Frequency (%)
2	379375	46.4%
1	250061	30.6%
3	103077	12.6%
4	38515	4.7%
0	17119	2.1%

Value	Count	Frequency (%)
5	16170	2.0%
6	6706	0.8%
7	2609	0.3%
8	1168	0.1%
9	484	0.1%
Other values (38)	659	0.1%
(Missing)	1780	0.2%

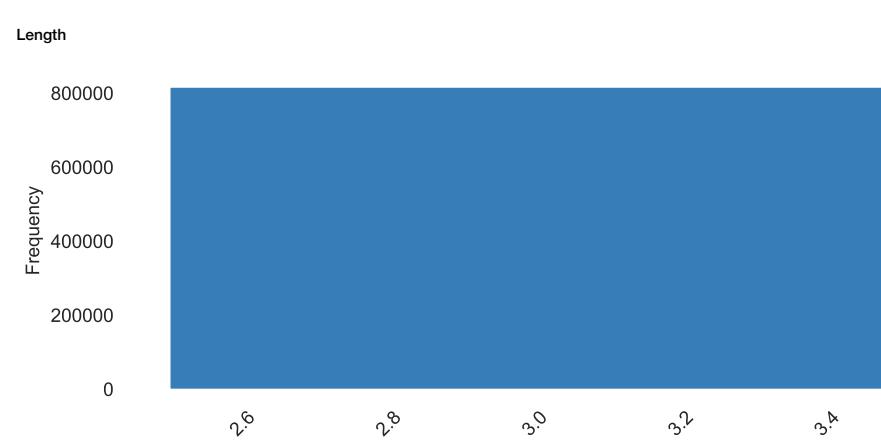
  

Value	Count	Frequency (%)
0	17119	2.1%
1	250061	30.6%
2	379375	46.4%
3	103077	12.6%
4	38515	4.7%
5	16170	2.0%
6	6706	0.8%
7	2609	0.3%
8	1168	0.1%
9	484	0.1%

Value	Count	Frequency (%)
61	1	< 0.1%
50	1	< 0.1%
48	1	< 0.1%
46	1	< 0.1%
45	2	< 0.1%
43	1	< 0.1%
42	3	< 0.1%
40	2	< 0.1%
39	1	< 0.1%
38	1	< 0.1%

INJURIES_UNKNOWN		
Categorical		
CONSTANT		
<b>Distinct</b>	1	
<b>Distinct (%)</b>	< 0.1%	
<b>Missing</b>	1780	
<b>Missing (%)</b>	0.2%	
<b>Memory size</b>	6.2 MiB	
<b>Length</b>		
<b>Max length</b>	3	
<b>Median length</b>	3	
<b>Mean length</b>	3	
<b>Min length</b>	3	
<b>Characters and Unicode</b>		
Total characters	2447829	
<b>Unique</b>	0	
<b>Unique (%)</b>	0.0%	
<b>Sample</b>		
1st row	0.0	
2nd row	0.0	
3rd row	0.0	
4th row	0.0	
5th row	0.0	
<b>Distinct scripts</b>	?	
( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?	
<b>Distinct blocks</b>	?	
The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.		
<b>Common Values</b>		
Value	Count	Frequency (%)
0.0	815943	99.8%
(Missing)	1780	0.2%



Common Values (Plot)



0.0

Value	Count	Frequency (%)
0.0	815943	100.0%

#### Most occurring characters

Value	Count	Frequency (%)
0	1631886	66.7%
.	815943	33.3%

#### Most occurring categories

Value	Count	Frequency (%)
(unknown)	2447829	100.0%

#### Most frequent character per category

Value	Count	Frequency (%)
0	1631886	66.7%
.	815943	33.3%

#### Most occurring scripts

Value	Count	Frequency (%)
(unknown)	2447829	100.0%

#### Most frequent character per script

Value	Count	Frequency (%)
(unknown)	2447829	100.0%

Value	Count	Frequency (%)
0	1631886	66.7%
.	815943	33.3%

#### Most occurring blocks

Value	Count	Frequency (%)
(unknown)	2447829	100.0%

#### Most frequent character per block

Value	Count	Frequency (%)
(unknown)	1631886	66.7%
.	815943	33.3%

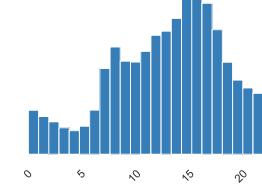
**CRASH\_HOUR**

Real number (R)

ZEROES

<b>Distinct</b>	24
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	13.198073

<b>Minimum</b>	0
<b>Maximum</b>	23
<b>Zeros</b>	17728
<b>Zeros (%)</b>	2.2%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%
<b>Memory size</b>	6.2 MiB

**Quantile statistics**

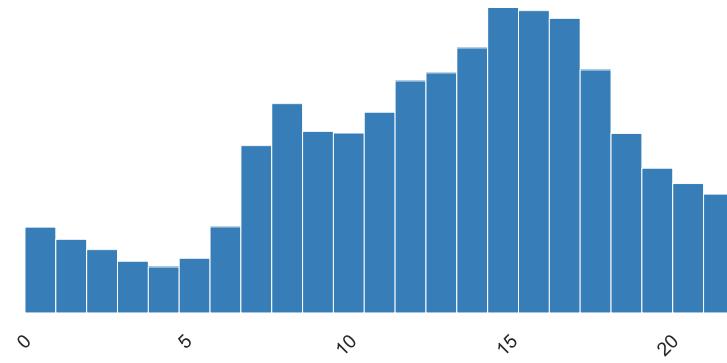
<b>Minimum</b>	0
<b>5-th percentile</b>	2
<b>Q1</b>	9
<b>median</b>	14
<b>Q3</b>	17
<b>95-th percentile</b>	22
<b>Maximum</b>	23
<b>Range</b>	23
<b>Interquartile range (IQR)</b>	8

**Descriptive statistics**

<b>Standard deviation</b>	5.5700507
<b>Coefficient of variation (CV)</b>	0.42203514
<b>Kurtosis</b>	-0.38952782
<b>Mean</b>	13.198073
<b>Median Absolute Deviation (MAD)</b>	4
<b>Skewness</b>	-0.42698356
<b>Sum</b>	10792368
<b>Variance</b>	31.025465
<b>Monotonicity</b>	Not monotonic

Frequency

60000  
50000  
40000  
30000  
20000  
10000  
0



Histogram with fixed size bins (bins=24)

Value

15  
16  
17  
14  
18

Count	Frequency (%)
63101	7.7%
62452	7.6%
60858	7.4%
54756	6.7%
50210	6.1%

Value

13  
12  
8  
11  
9  
Other values (14)

Count	Frequency (%)
49604	6.1%
47942	5.9%
43242	5.3%
41461	5.1%
37509	4.6%
306588	37.5%

Value

0  
1  
2  
3  
4  
5  
6  
7  
8  
9

Count	Frequency (%)
17728	2.2%
15210	1.9%
13085	1.6%
10681	1.3%
9509	1.2%
11295	1.4%
17792	2.2%
34569	4.2%
43242	5.3%
37509	4.6%

Value

23  
22  
21  
20  
19  
18  
17  
16  
15  
14

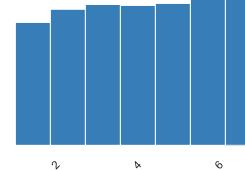
Count	Frequency (%)
21276	2.6%
24553	3.0%
26734	3.3%
29896	3.7%
37090	4.5%
50210	6.1%
60858	7.4%
62452	7.6%
63101	7.7%
54756	6.7%

## CRASH\_DAY\_OF\_WEEK

Real number (R)

Distinct	7
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	4.1228167

Minimum	1
Maximum	7
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	6.2 MiB



## Quantile statistics

Minimum	1
5-th percentile	1
Q1	2
median	4
Q3	6
95-th percentile	7
Maximum	7
Range	6
Interquartile range (IQR)	4

## Descriptive statistics

Standard deviation	1.9808553
Coefficient of variation (CV)	0.48046165
Kurtosis	-1.2397911
Mean	4.1228167
Median Absolute Deviation (MAD)	2
Skewness	-0.0772494
Sum	3371322
Variance	3.9237877
Monotonicity	Not monotonic

Frequency

120000  
100000  
80000  
60000  
40000  
20000  
0

1 2 3 4 5 6 7

Histogram with fixed size bins (bins=7)

Value	Count	Frequency (%)
6	132905	16.3%
7	121206	14.8%
5	117337	14.3%
3	116478	14.2%
4	115685	14.1%

## Value

Count Frequency (%)

2	112479	13.8%
1	101633	12.4%

## Value

Count Frequency (%)

1	101633	12.4%
2	112479	13.8%
3	116478	14.2%
4	115685	14.1%
5	117337	14.3%
6	132905	16.3%
7	121206	14.8%

## Value

Count Frequency (%)

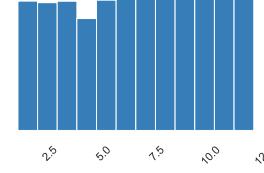
7	121206	14.8%
6	132905	16.3%
5	117337	14.3%
4	115685	14.1%
3	116478	14.2%
2	112479	13.8%
1	101633	12.4%

**CRASH\_MONTH**

Real number (R)

<b>Distinct</b>	12
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	6.6558235

<b>Minimum</b>	1
<b>Maximum</b>	12
<b>Zeros</b>	0
<b>Zeros (%)</b>	0.0%
<b>Negative</b>	0
<b>Negative (%)</b>	0.0%

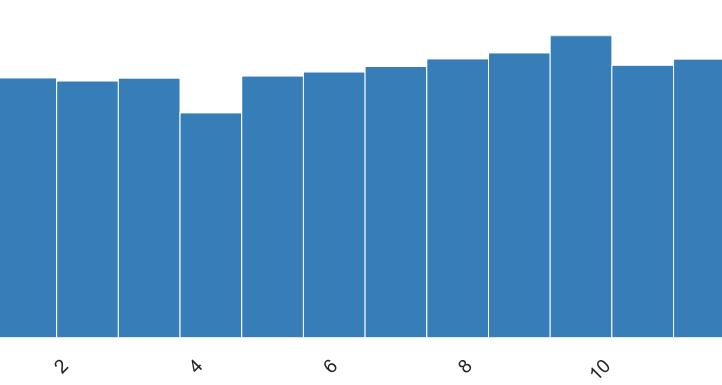
**Quantile statistics**

<b>Minimum</b>	1
<b>5-th percentile</b>	1
<b>Q1</b>	4
<b>median</b>	7
<b>Q3</b>	10
<b>95-th percentile</b>	12
<b>Maximum</b>	12
<b>Range</b>	11
<b>Interquartile range (IQR)</b>	6

**Descriptive statistics**

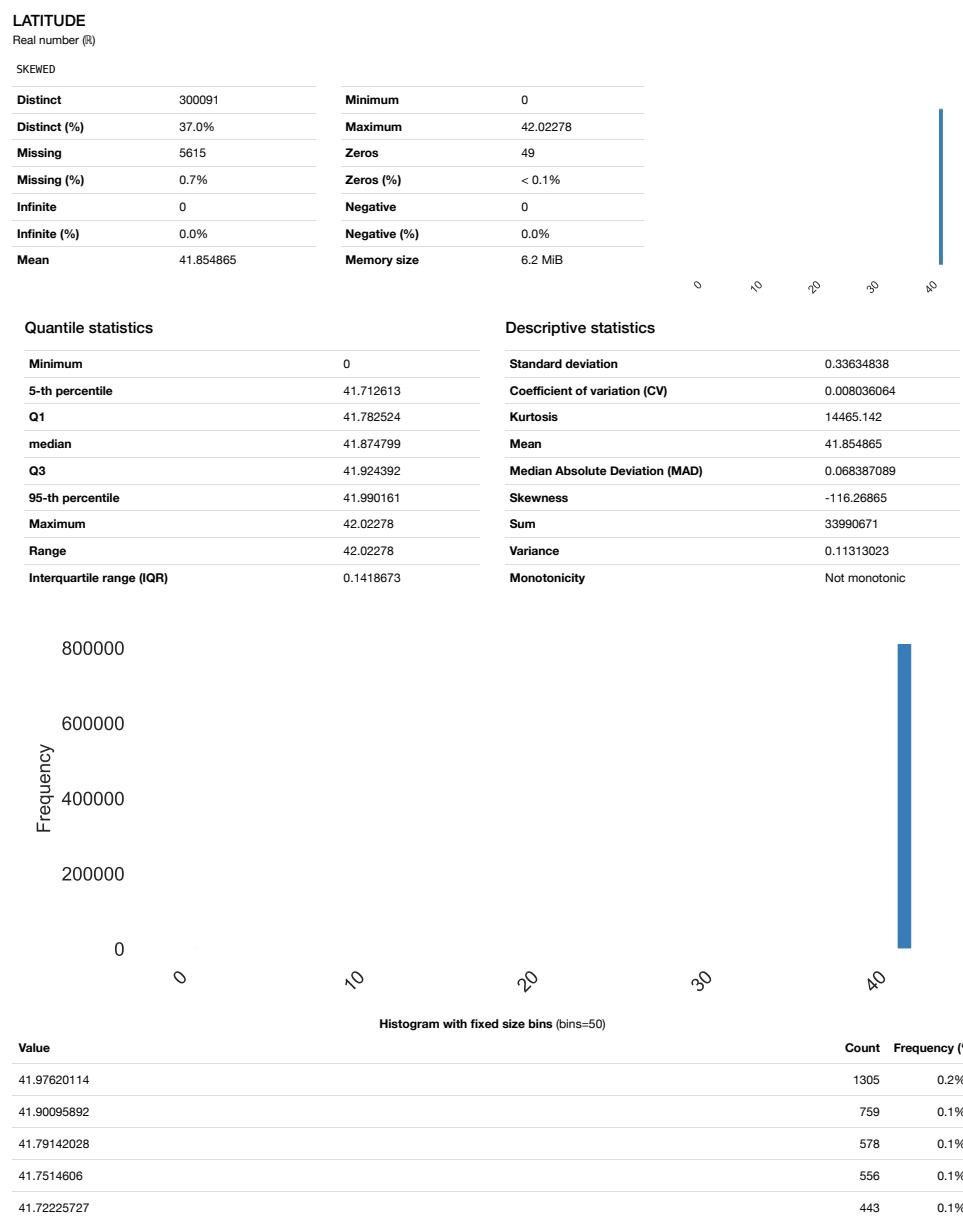
<b>Standard deviation</b>	3.4525324
<b>Coefficient of variation (CV)</b>	0.51872354
<b>Kurtosis</b>	-1.2109649
<b>Mean</b>	6.6558235
<b>Median Absolute Deviation (MAD)</b>	3
<b>Skewness</b>	-0.07976985
<b>Sum</b>	5442620
<b>Variance</b>	11.91998
<b>Monotonicity</b>	Not monotonic

Frequency

0  
20000  
40000  
60000  
80000

Histogram with fixed size bins (bins=12)

**Value**10  
9  
8  
12  
11**Count** **Frequency (%)**76799 9.4%  
72415 8.9%  
70898 8.7%  
70824 8.7%  
69260 8.5%**Value**7  
6  
5  
1  
3  
Other values (2)**Count** **Frequency (%)**68963 8.4%  
67560 8.3%  
66530 8.1%  
66052 8.1%  
65988 8.1%  
122434 15.0%**Value**1  
2  
3  
4  
5  
6  
7  
8  
9  
10**Count** **Frequency (%)**66052 8.1%  
65270 8.0%  
65988 8.1%  
57164 7.0%  
66530 8.1%  
67560 8.3%  
68963 8.4%  
70898 8.7%  
72415 8.9%  
76799 9.4%**Value**12  
11  
10  
9  
8  
7  
6  
5  
4  
3**Count** **Frequency (%)**70824 8.7%  
69260 8.5%  
76799 9.4%  
72415 8.9%  
70898 8.7%  
68963 8.4%  
67560 8.3%  
66530 8.1%  
57164 7.0%  
65988 8.1%



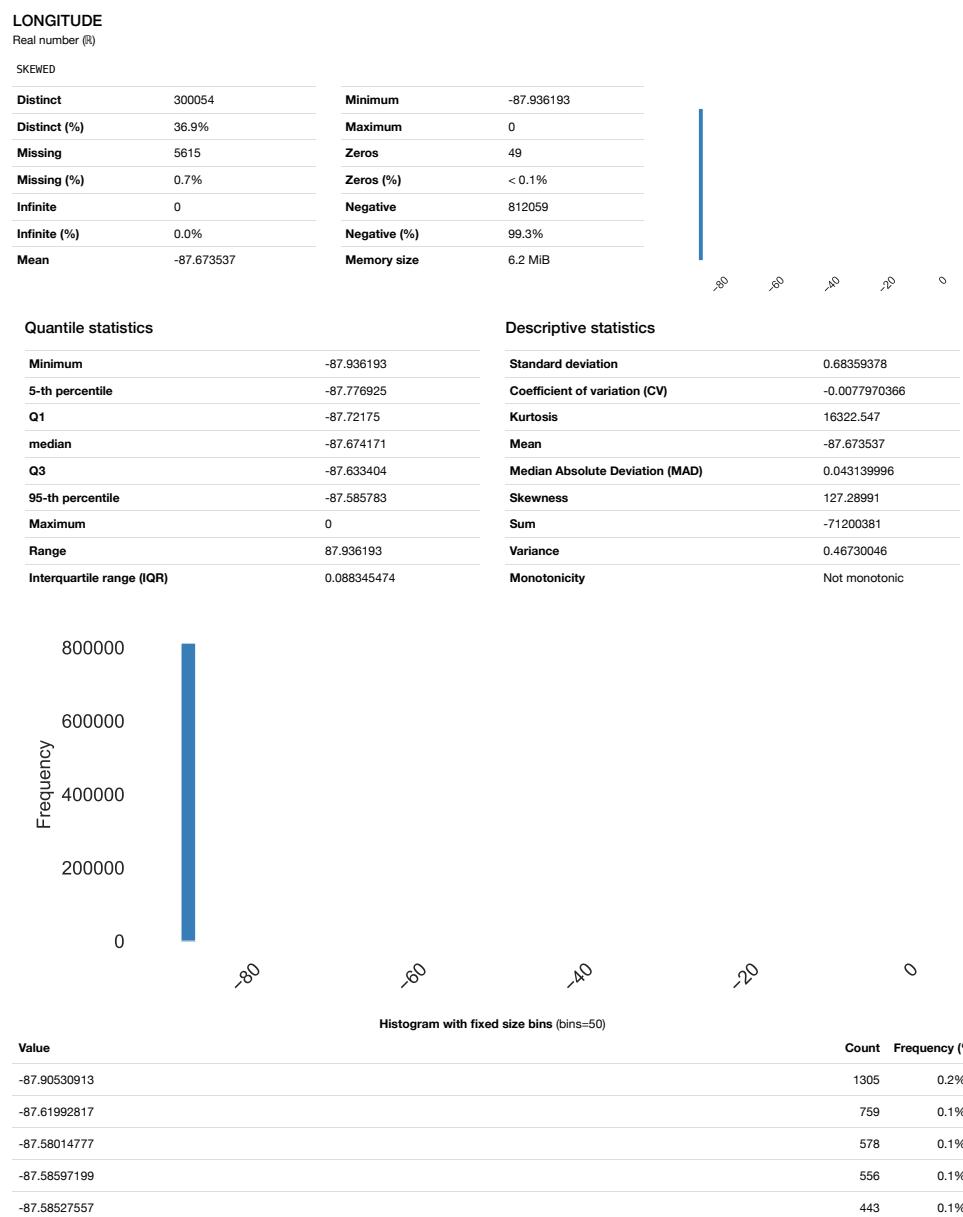
Value	Count	Frequency (%)
41.75466012	381	< 0.1%
41.88085605	329	< 0.1%
41.78932932	325	< 0.1%
41.90075297	306	< 0.1%
41.89680497	304	< 0.1%
Other values (300081)	806822	98.7%
(Missing)	5615	0.7%

Value	Count	Frequency (%)
0	49	< 0.1%
41.64467013	23	< 0.1%
41.64469152	4	< 0.1%
41.64469397	7	< 0.1%
41.64469408	1	< 0.1%
41.64470194	5	< 0.1%
41.64471103	1	< 0.1%
41.64471232	1	< 0.1%
41.64471457	1	< 0.1%
41.64471544	4	< 0.1%

Value	Count	Frequency (%)
42.02277986	9	< 0.1%
42.02275469	1	< 0.1%
42.02273632	1	< 0.1%
42.02272017	2	< 0.1%
42.02266893	1	< 0.1%
42.02266114	2	< 0.1%
42.02266027	7	< 0.1%
42.02265996	1	< 0.1%
42.02265785	1	< 0.1%
42.0226457	1	< 0.1%



Value	Count	Frequency (%)
-87.74138476	381	< 0.1%
-87.61763589	329	< 0.1%
-87.74164564	325	< 0.1%
-87.624235	306	< 0.1%
-87.61702742	304	< 0.1%
Other values (300044)	806822	98.7%
(Missing)	5615	0.7%

Value	Count	Frequency (%)
-87.93619295	1	< 0.1%
-87.93587692	1	< 0.1%
-87.93476313	3	< 0.1%
-87.93450972	1	< 0.1%
-87.93401422	1	< 0.1%
-87.93399393	54	< 0.1%
-87.9339765	3	< 0.1%
-87.93302828	8	< 0.1%
-87.92822117	3	< 0.1%
-87.92726168	13	< 0.1%

Value	Count	Frequency (%)
0	49	< 0.1%
-87.52458739	13	< 0.1%
-87.52458901	4	< 0.1%
-87.52464032	1	< 0.1%
-87.5246459	1	< 0.1%
-87.52467395	8	< 0.1%
-87.52467489	1	< 0.1%
-87.52467584	1	< 0.1%
-87.52467708	1	< 0.1%
-87.52468243	1	< 0.1%



(unknown)		Count	Frequency (%)
Value			
7		3089155	9.6%
8		2891746	9.0%
4		2599622	8.0%
1		2545102	7.9%
6		2307184	7.1%
9		2047841	6.3%
5		1864201	5.8%
2		1822984	5.6%
3		1785259	5.5%
Other values (10)		1624216	5.0%
		9727996	30.1%

**Most occurring blocks**

Value		Count	Frequency (%)
(unknown)		32305306	100.0%

**Most frequent character per block**

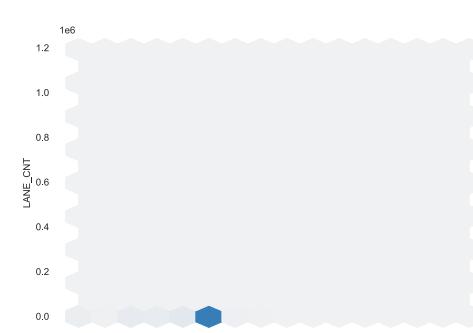
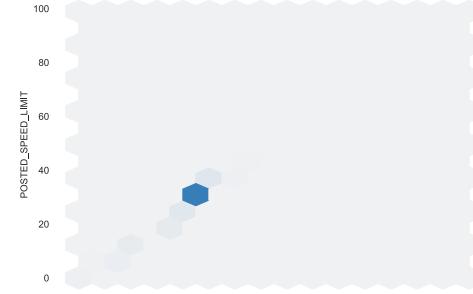
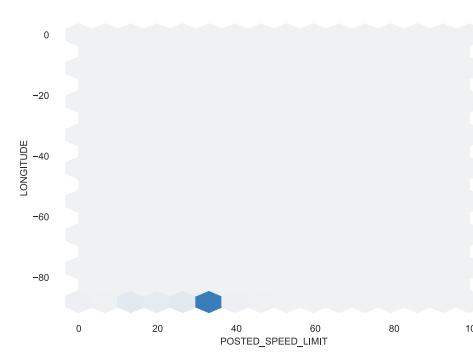
(unknown)		Count	Frequency (%)
Value			
7		3089155	9.6%
8		2891746	9.0%
4		2599622	8.0%
1		2545102	7.9%
6		2307184	7.1%
9		2047841	6.3%
5		1864201	5.8%
2		1822984	5.6%
3		1785259	5.5%
Other values (10)		1624216	5.0%
		9727996	30.1%

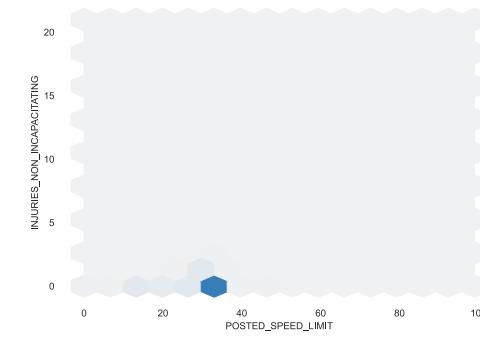
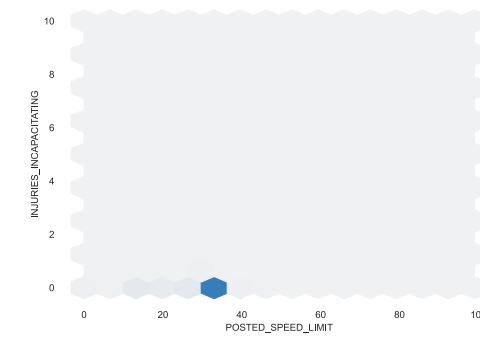
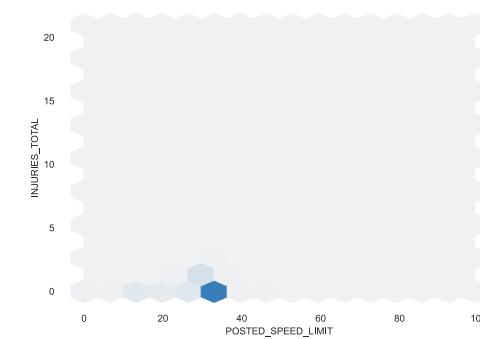
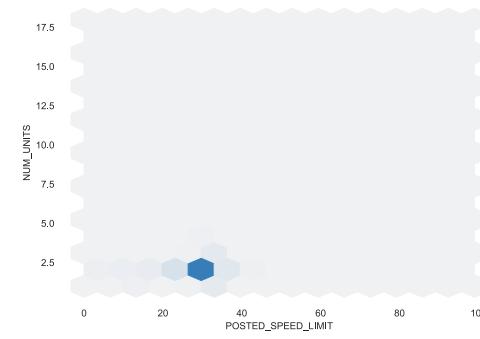
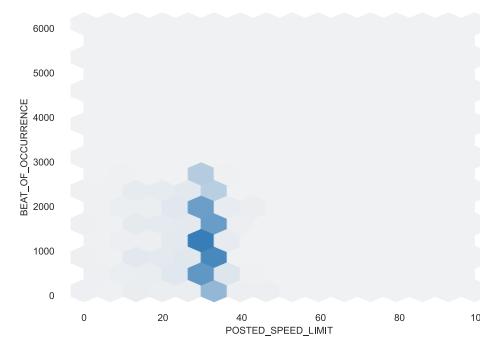
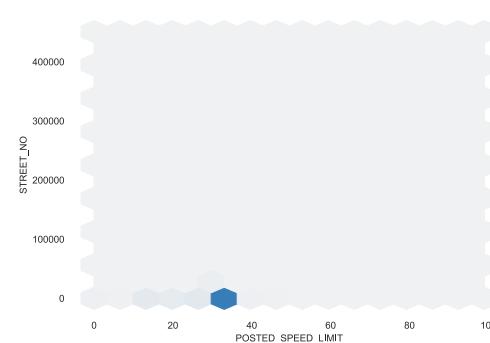
**Interactions**

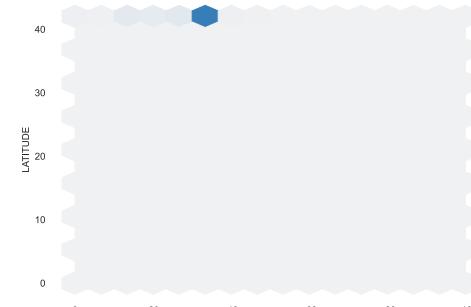
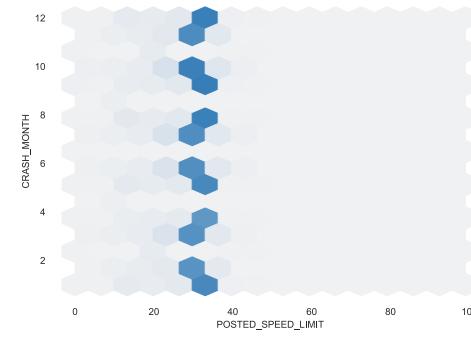
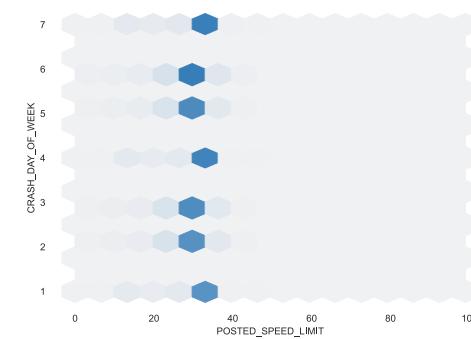
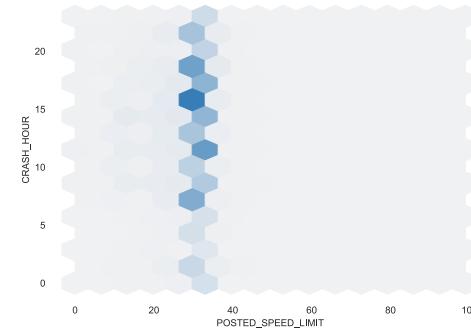
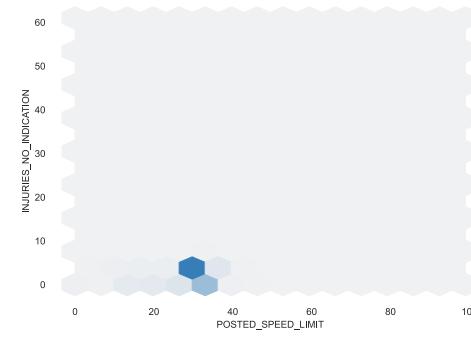
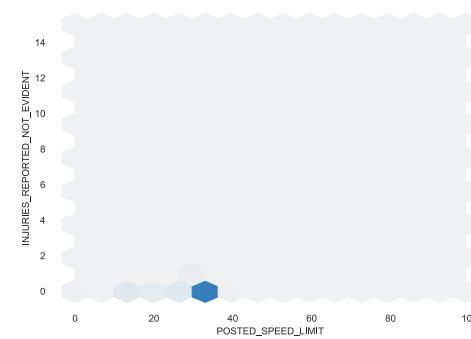
POSTED_SPEED_LIMIT
LANE_CNT
STREET_NO
BEAT_OF_OCCURRENCE
NUM_UNITS
INJURIES_TOTAL
INJURIES_INCAPACITATING
INJURIES_NON_INCAPACITATING
INJURIES_REPORTED_NC
INJURIES_NO_INDICATION
CRASH_HOUR
CRASH_DAY_OF_WEEK
CRASH_MONTH
LATITUDE
LONGITUDE

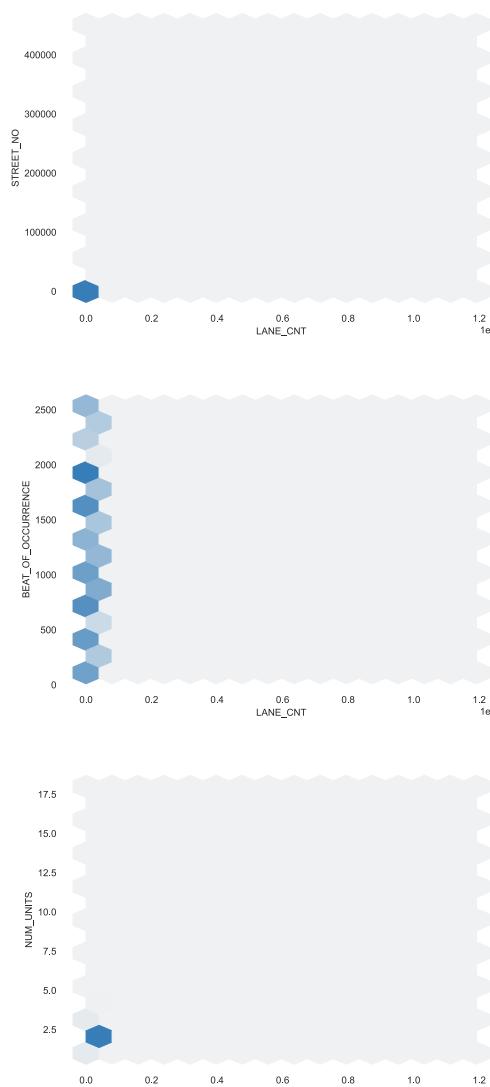
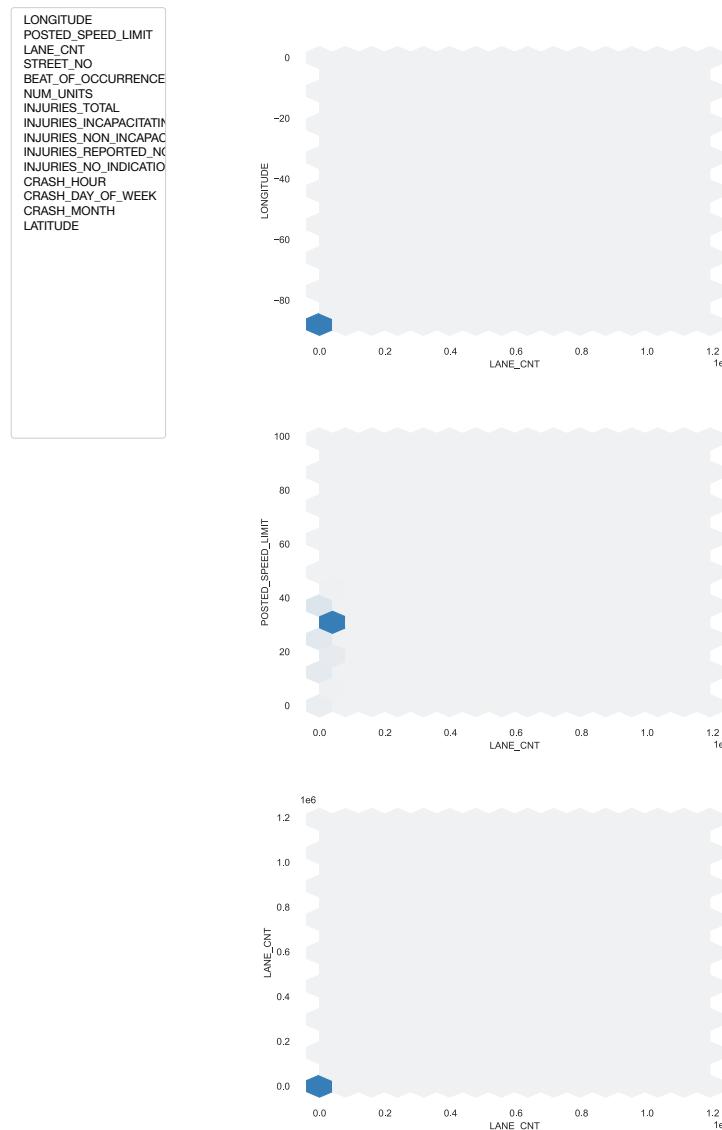
  

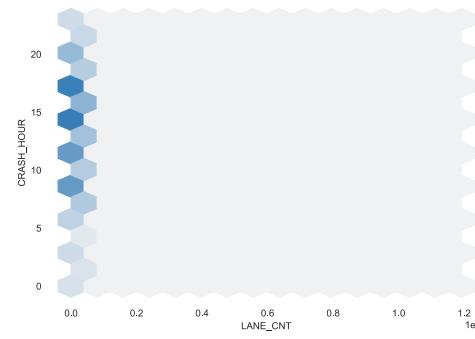
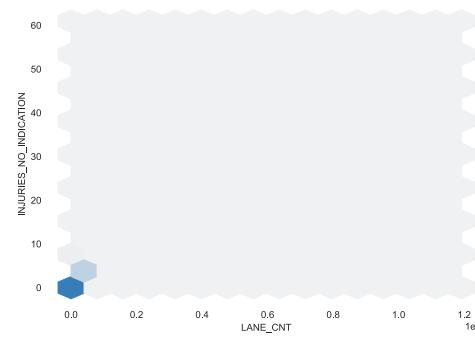
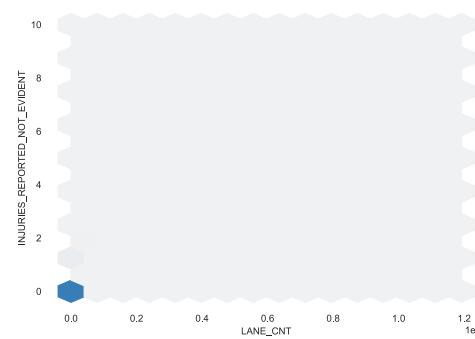
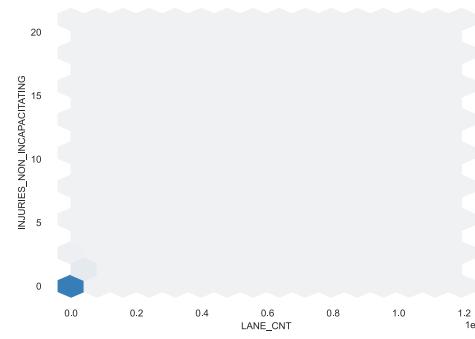
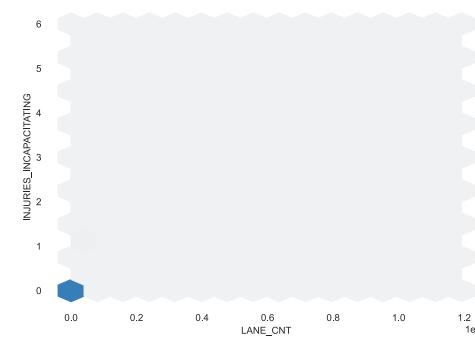
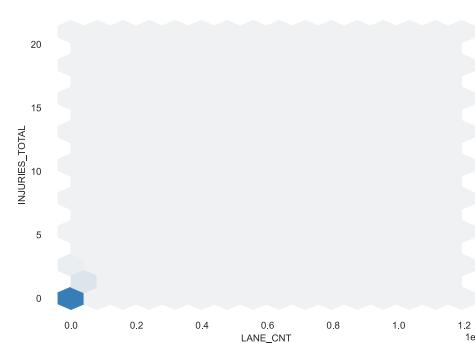
LONGITUDE
POSTED_SPEED_LIMIT
LANE_CNT
STREET_NO
BEAT_OF_OCCURRENCE
NUM_UNITS
INJURIES_TOTAL
INJURIES_INCAPACITATING
INJURIES_NON_INCAPACITATING
INJURIES_REPORTED_NC
INJURIES_NO_INDICATION
CRASH_HOUR
CRASH_DAY_OF_WEEK
CRASH_MONTH
LATITUDE

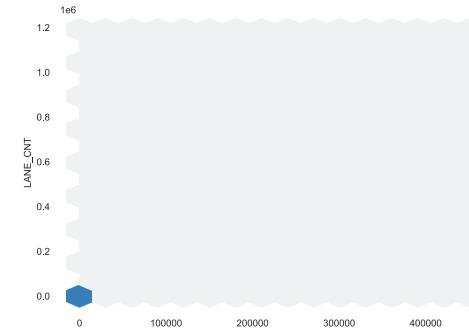
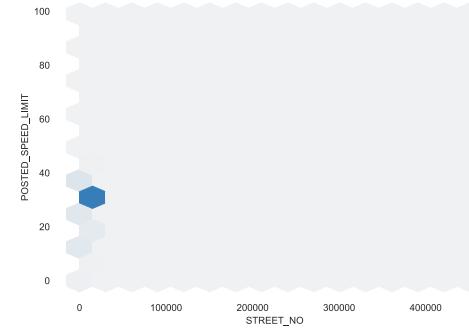
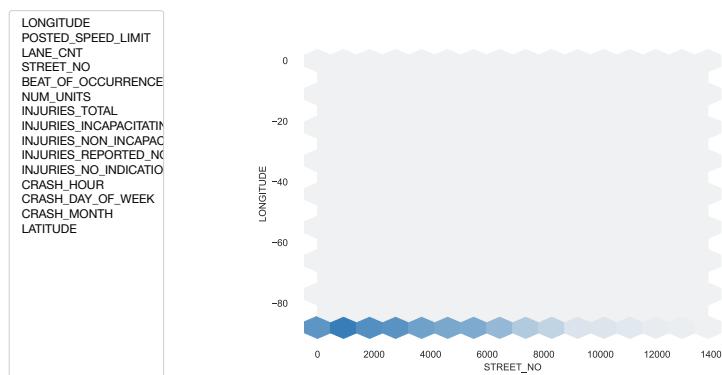
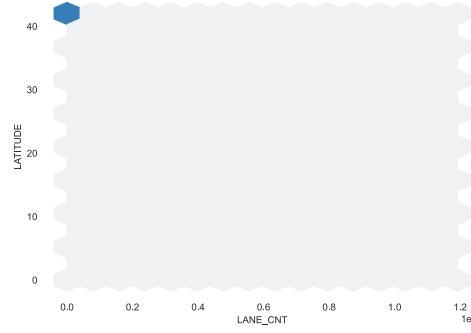
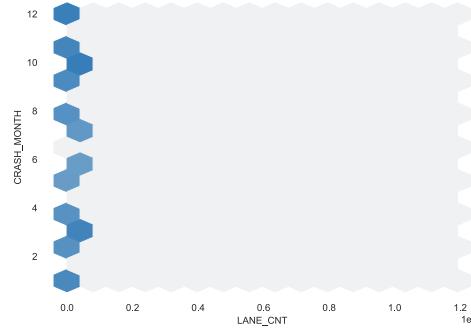
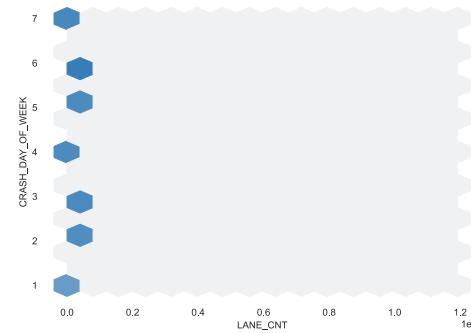


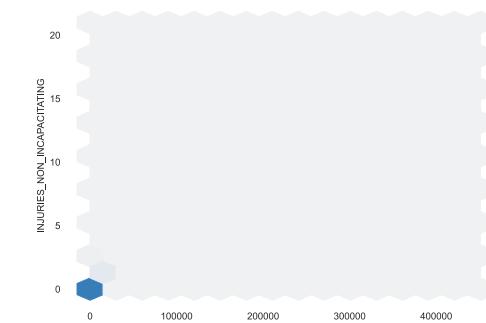
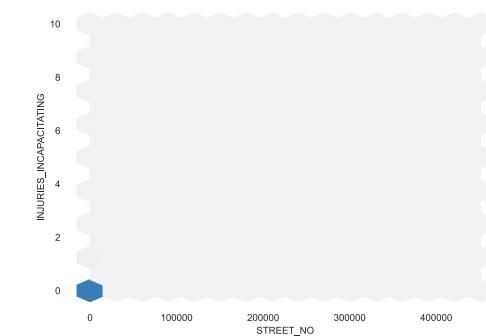
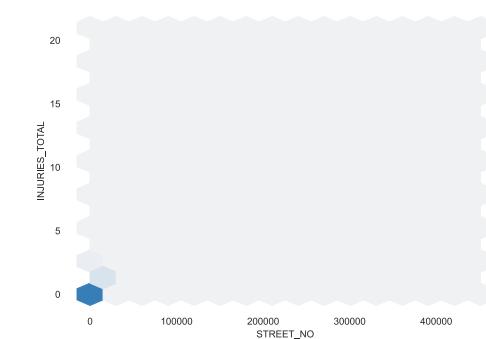
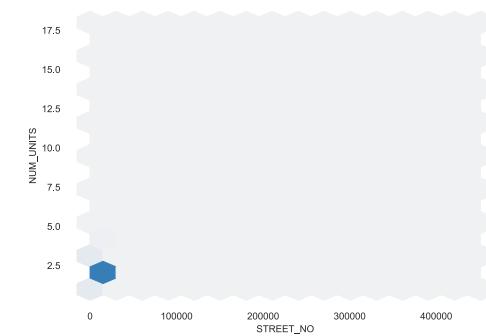
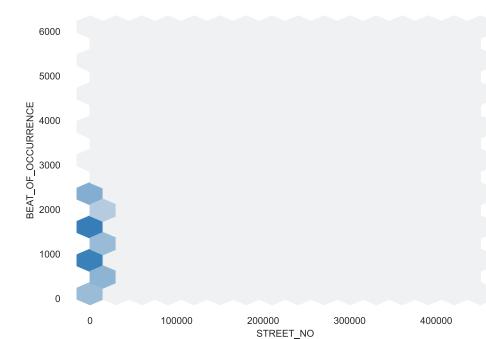
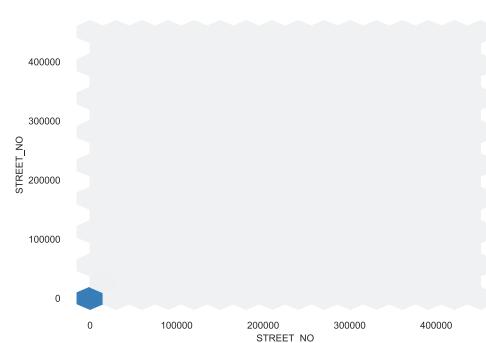


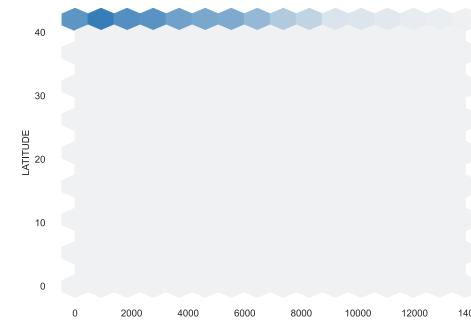
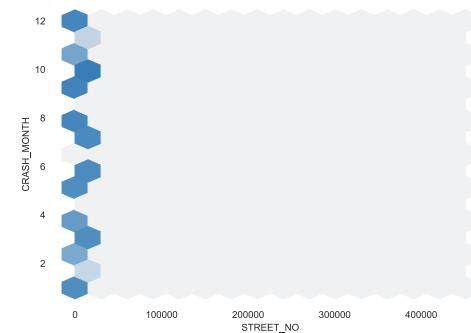
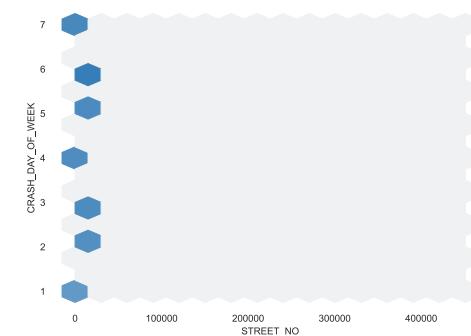
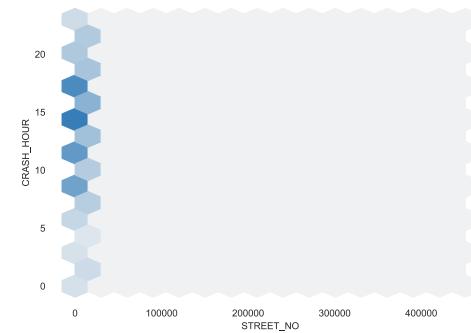
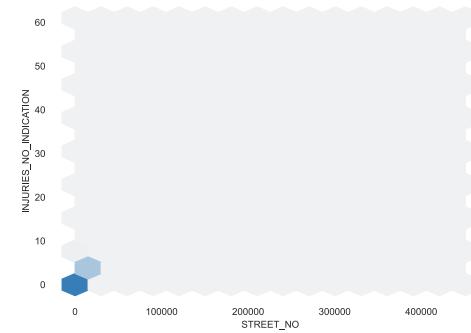
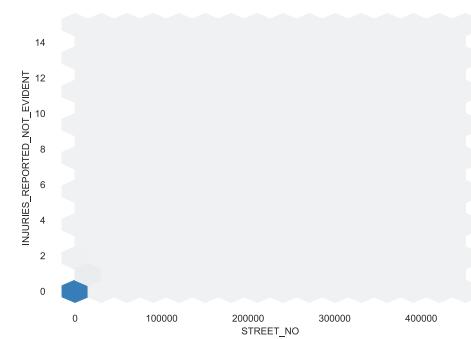


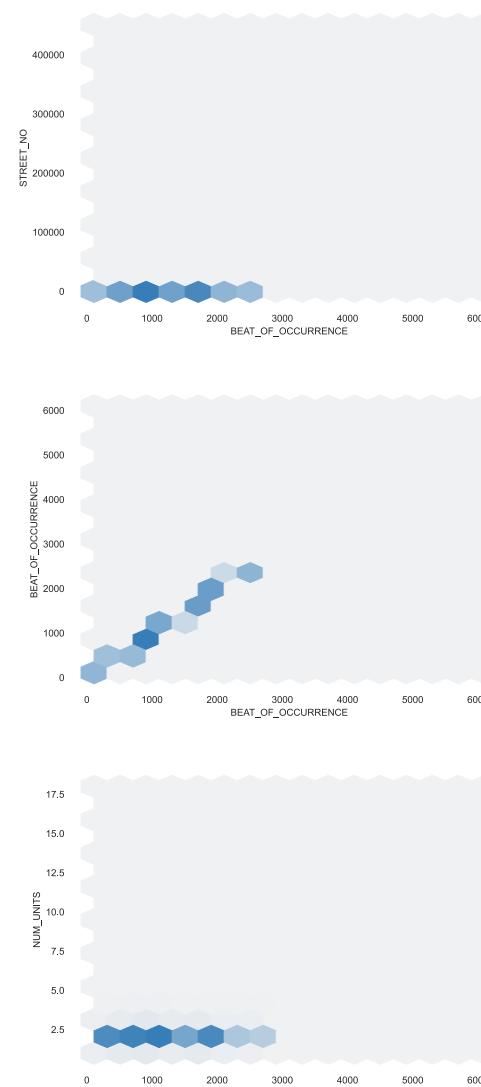
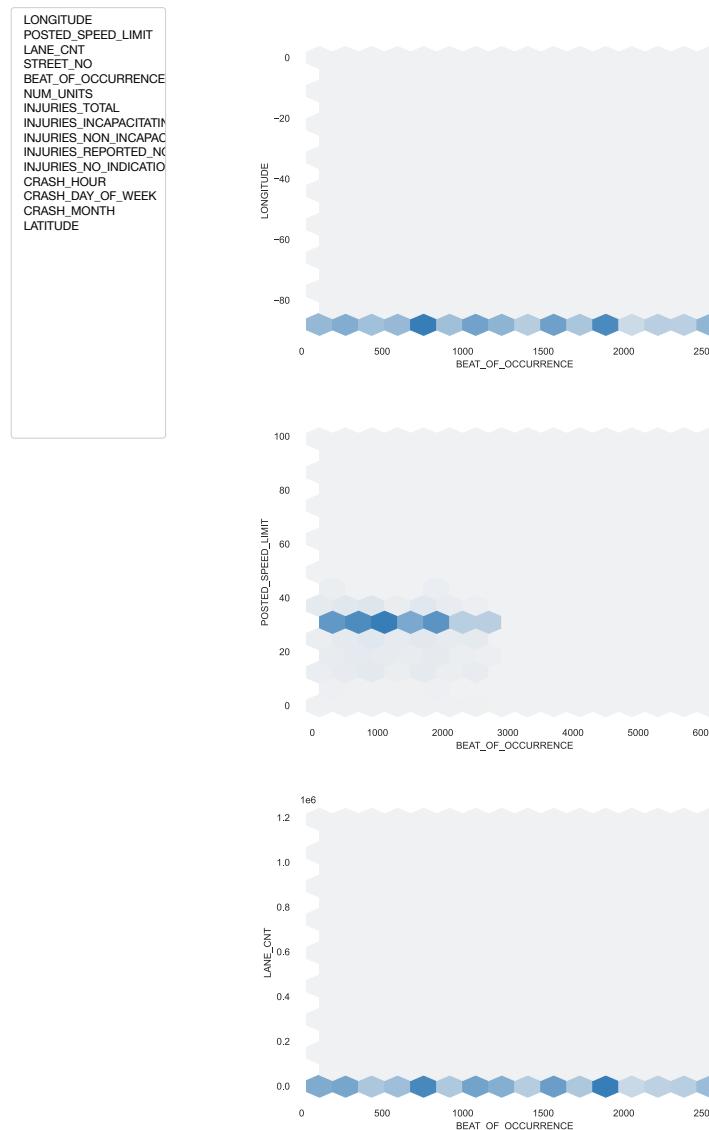


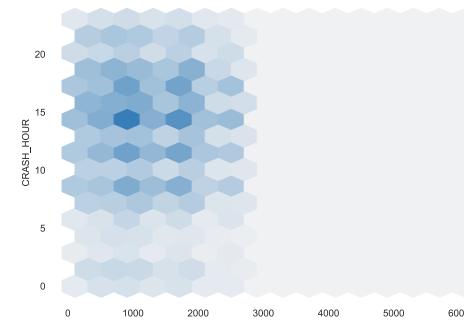
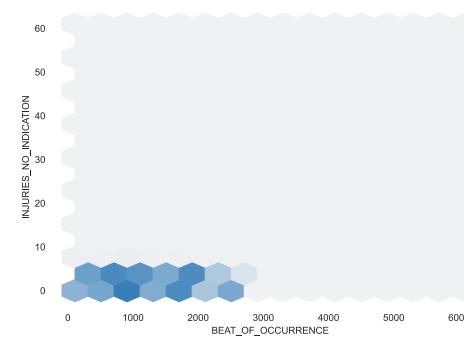
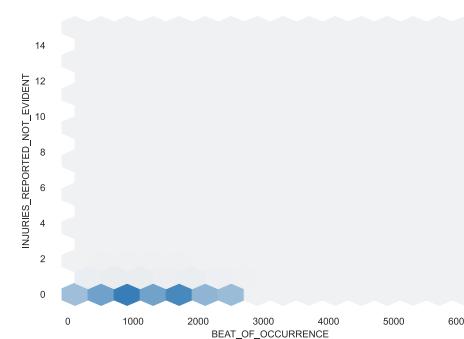
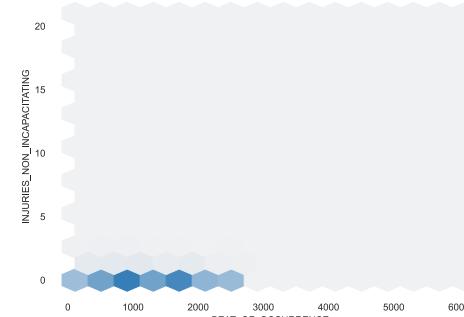
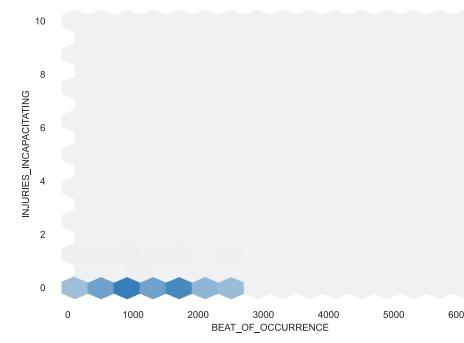
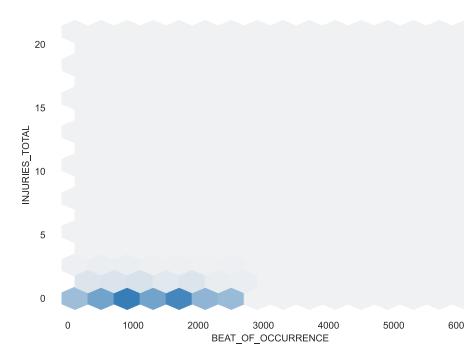


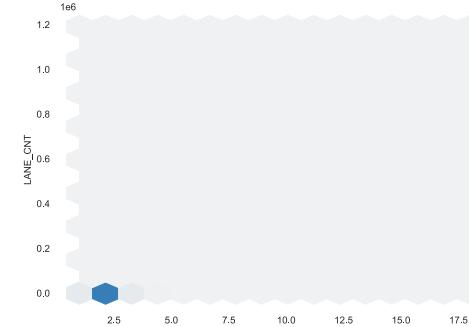
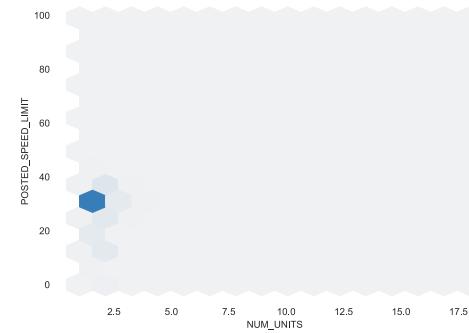
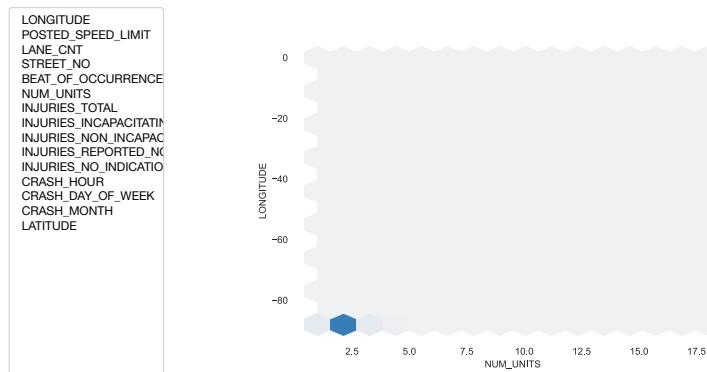
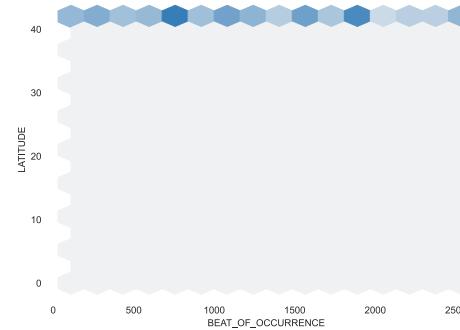
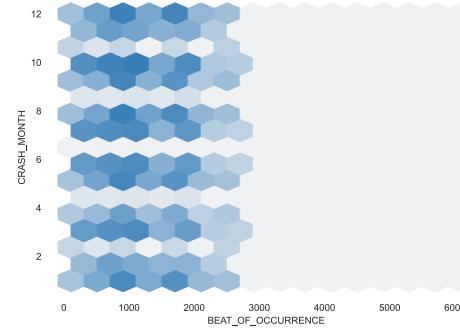
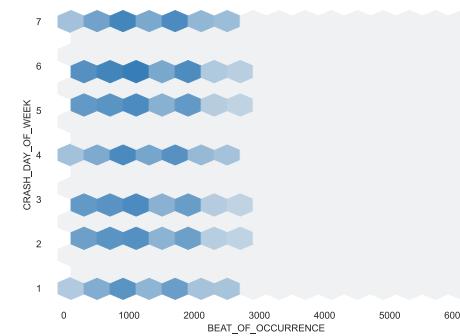


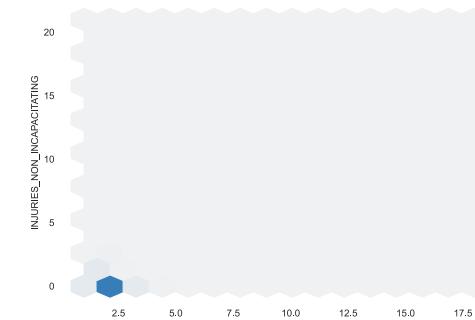
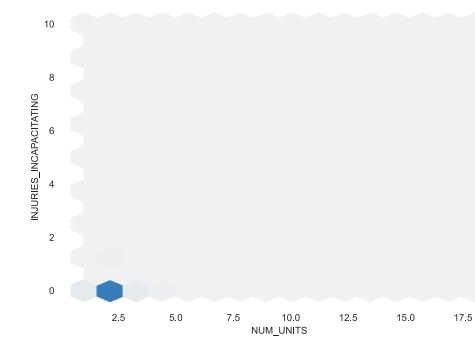
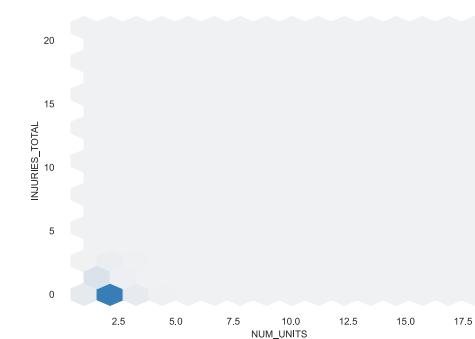
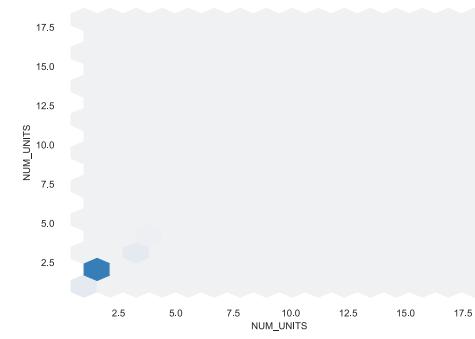
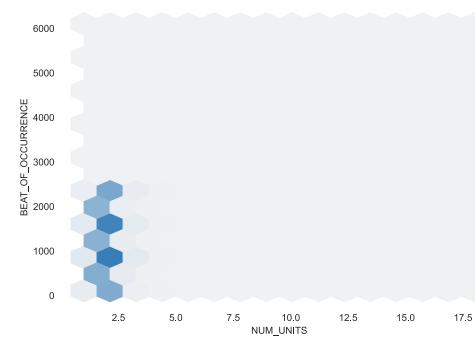
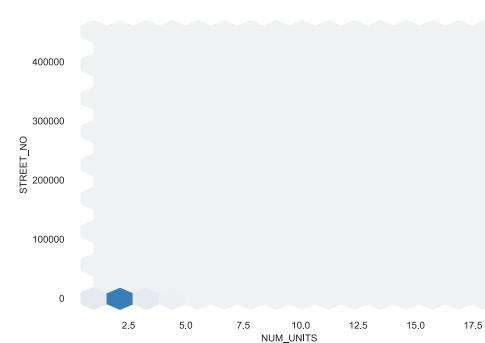


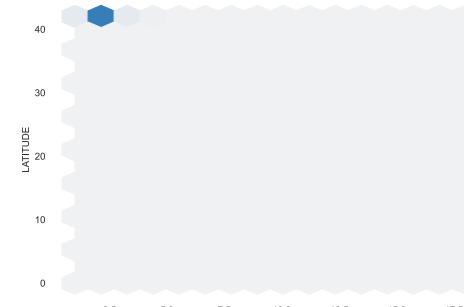
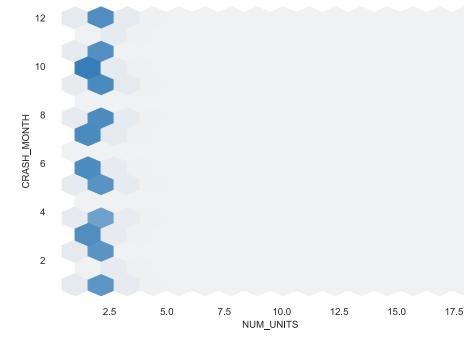
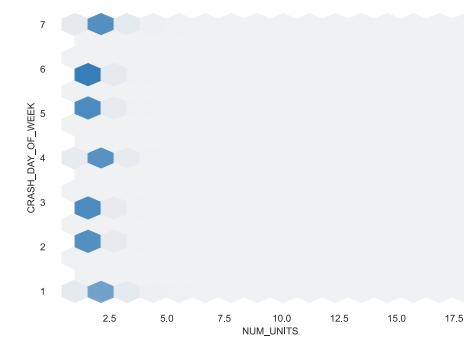
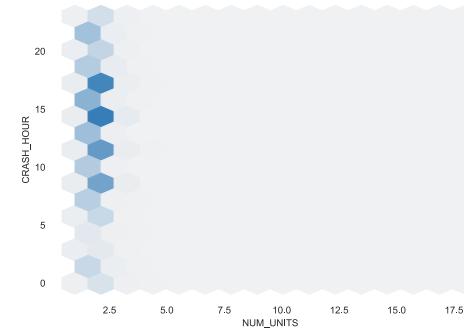
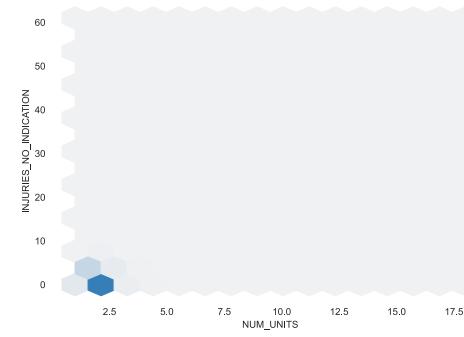
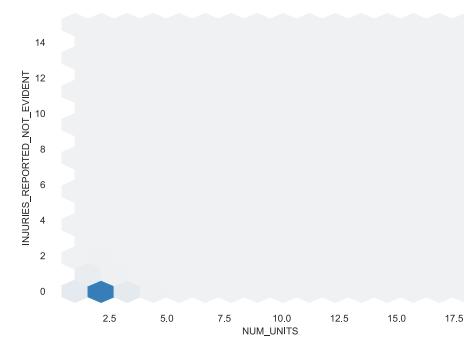


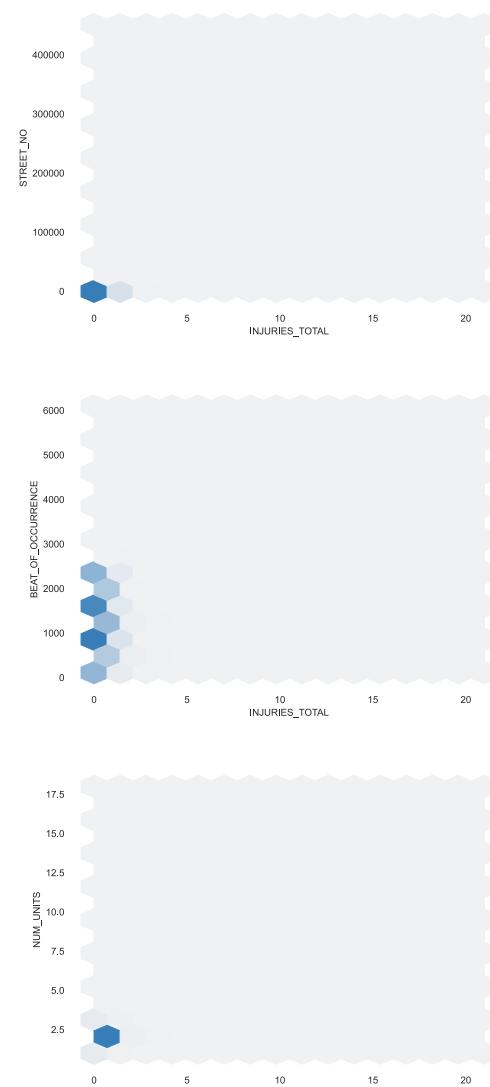
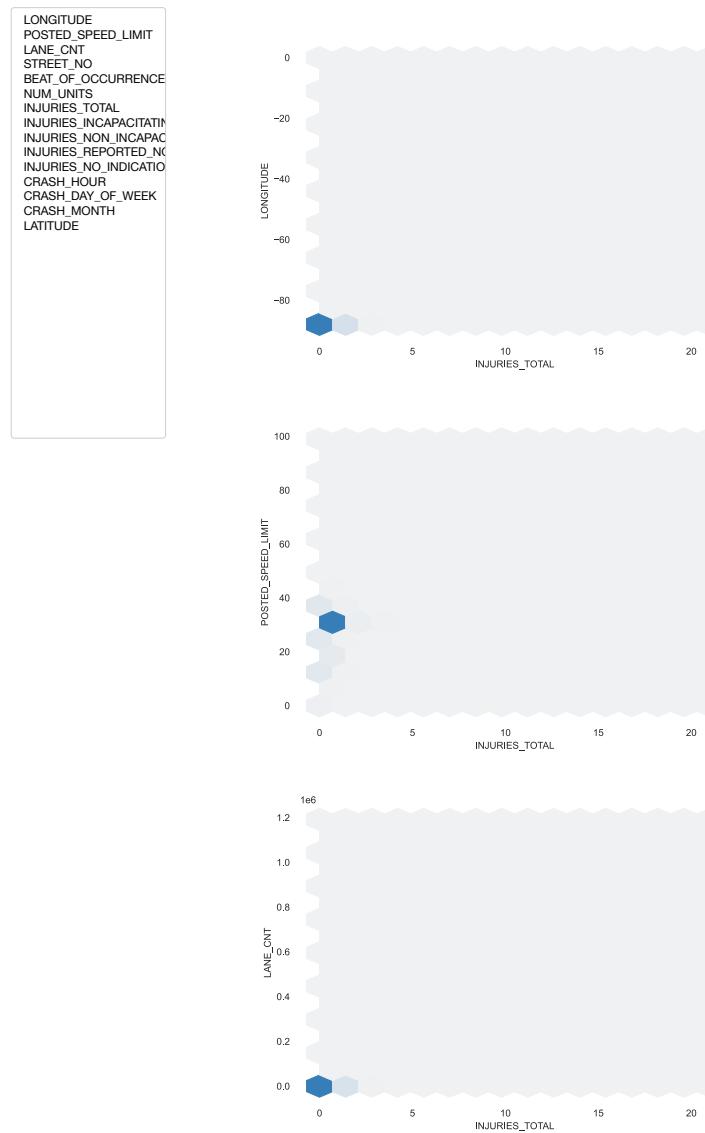


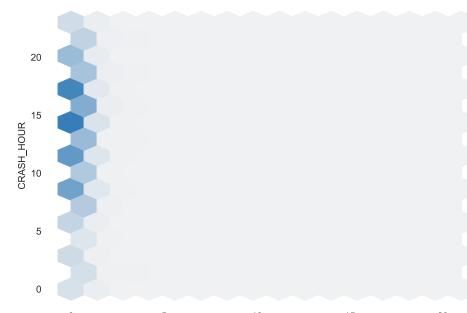
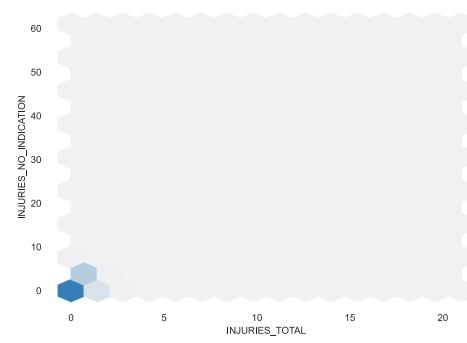
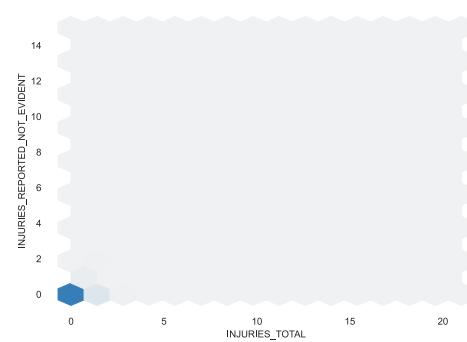
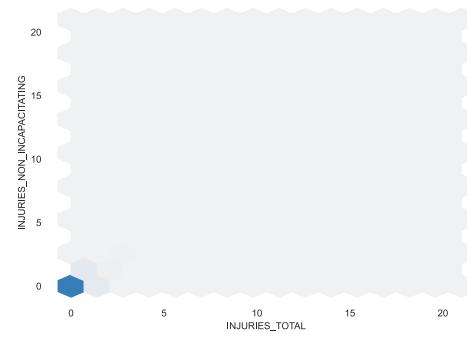
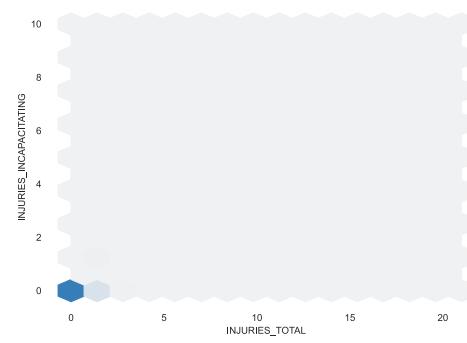
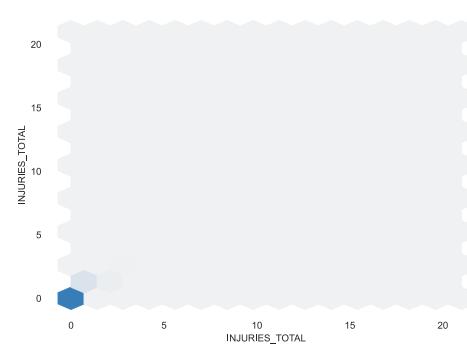


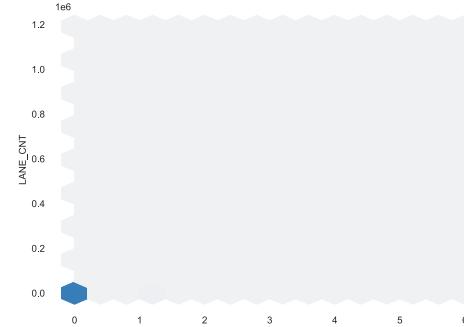
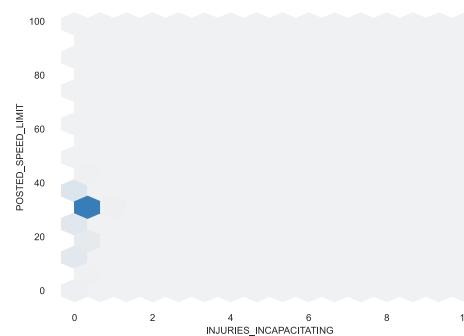
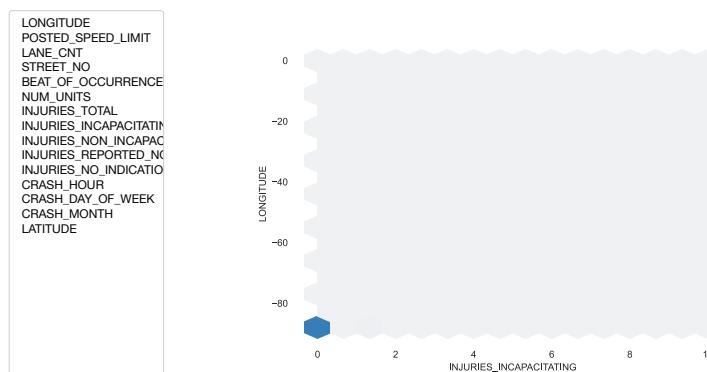
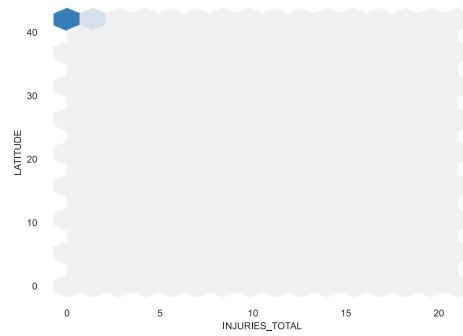
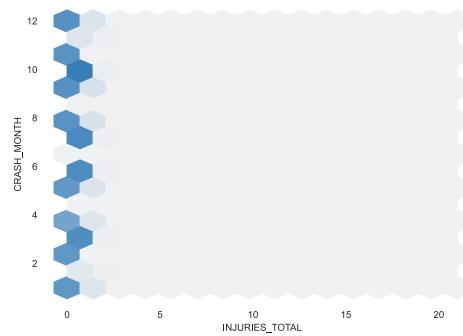
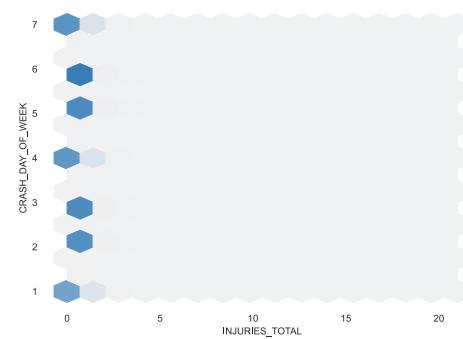


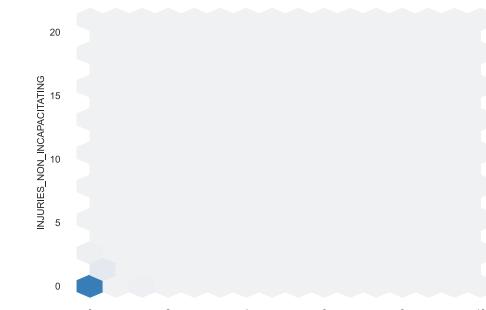
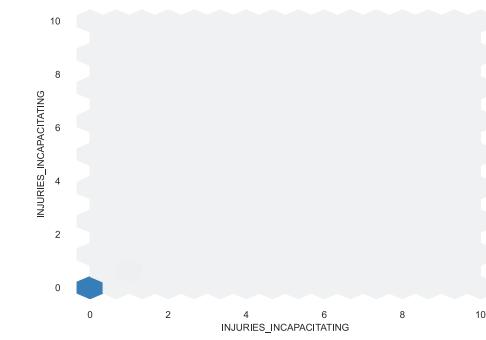
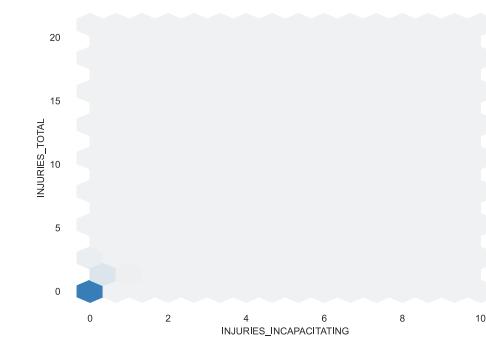
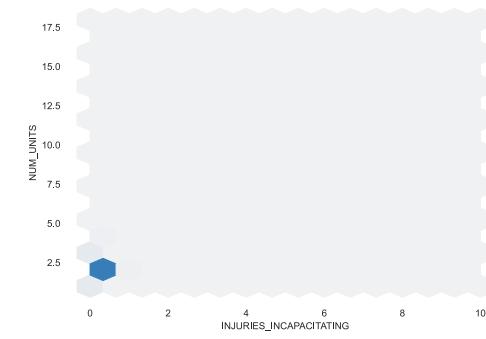
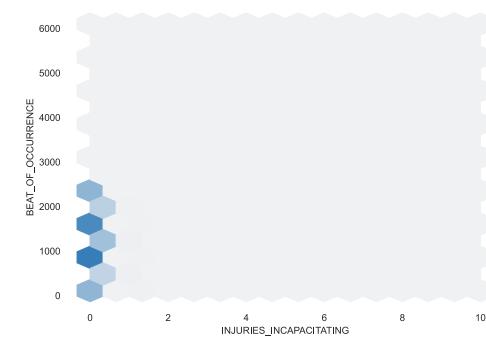
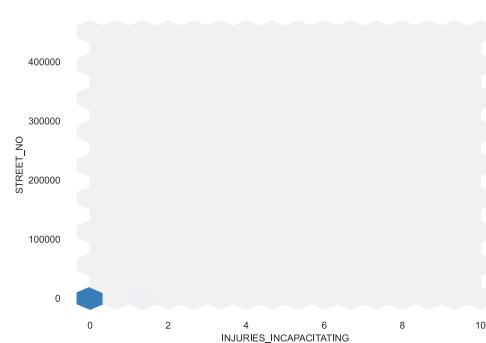


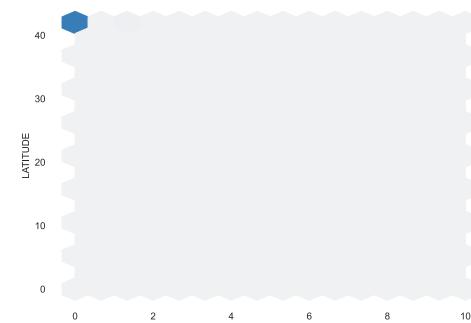
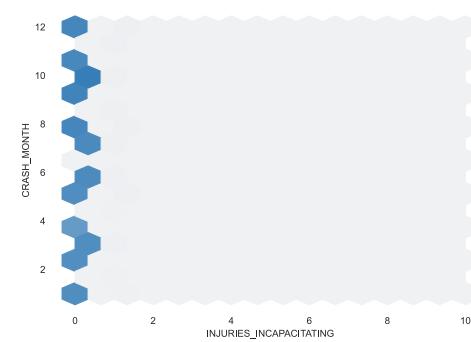
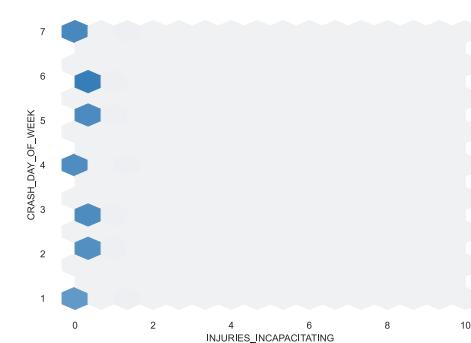
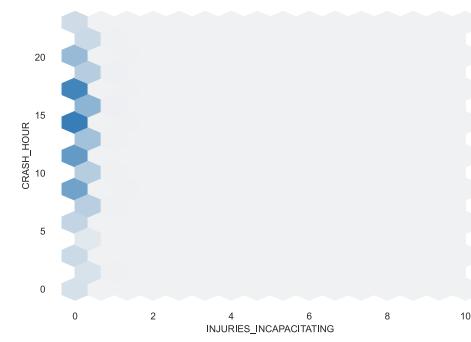
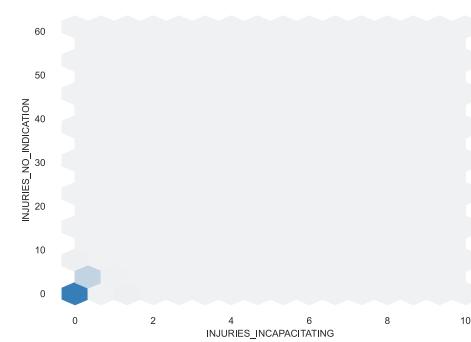
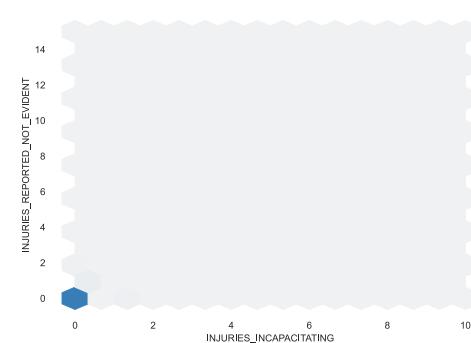


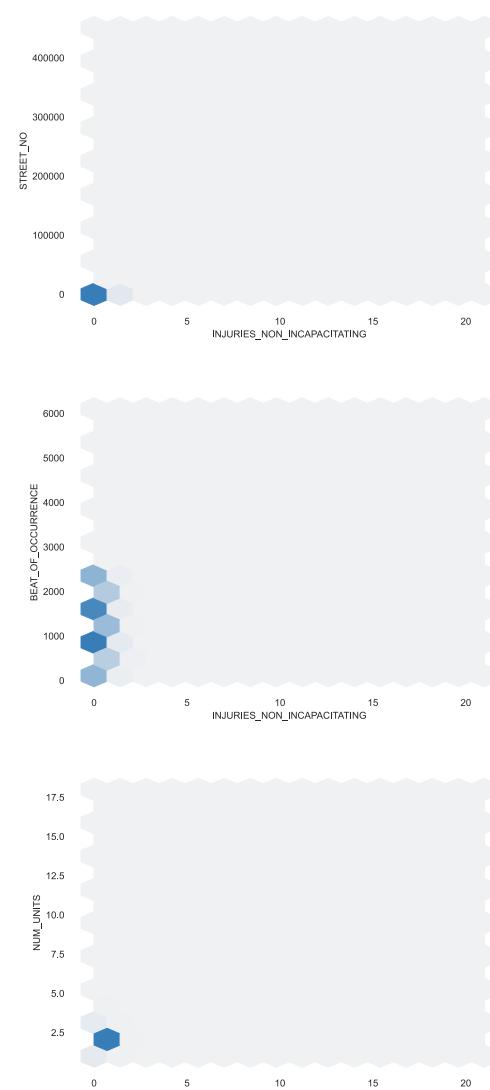
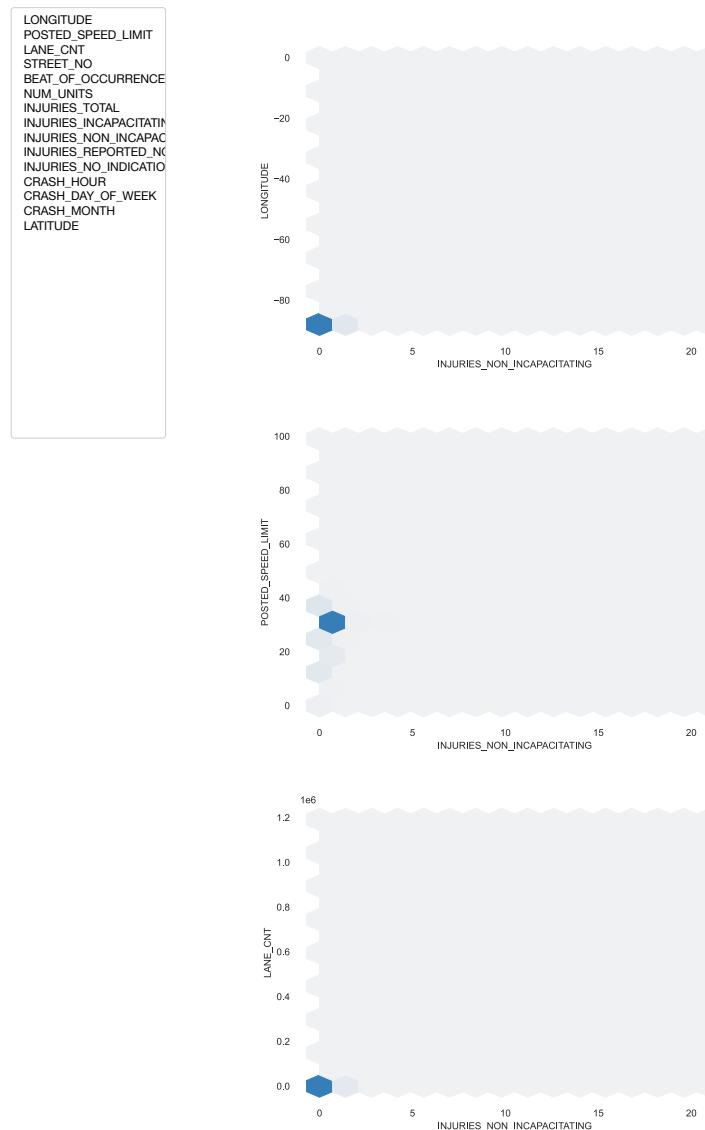


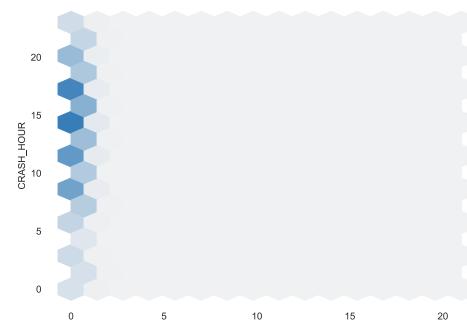
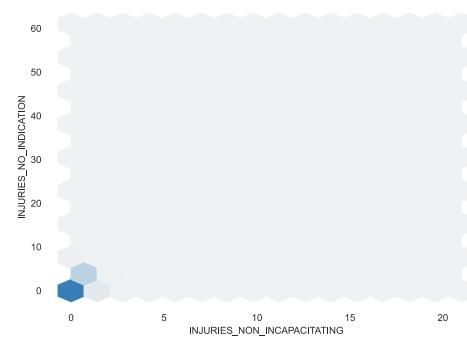


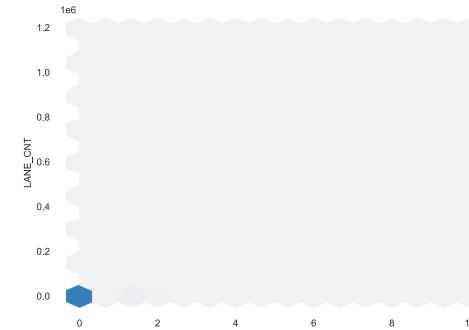
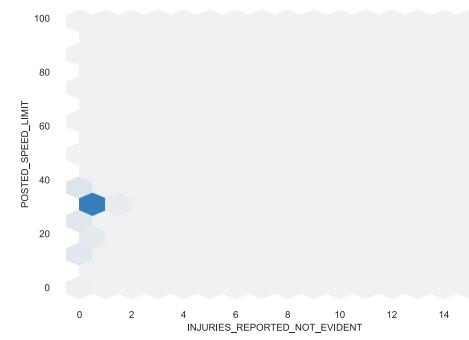
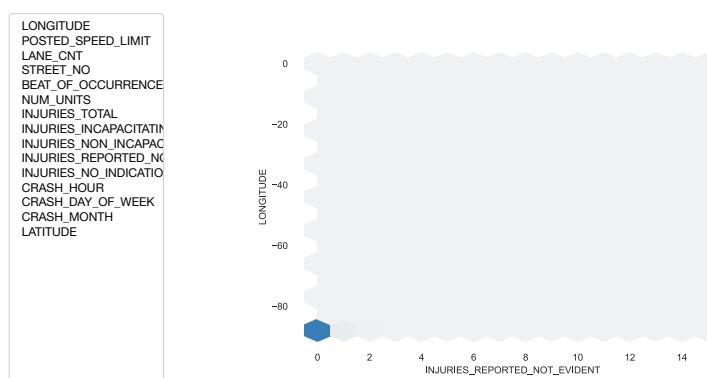
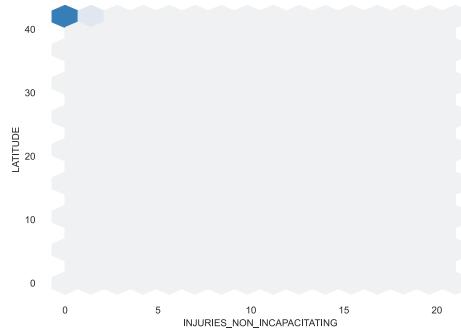
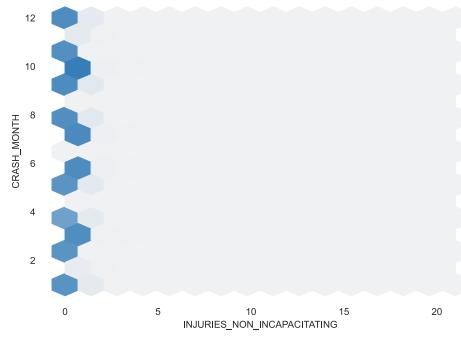
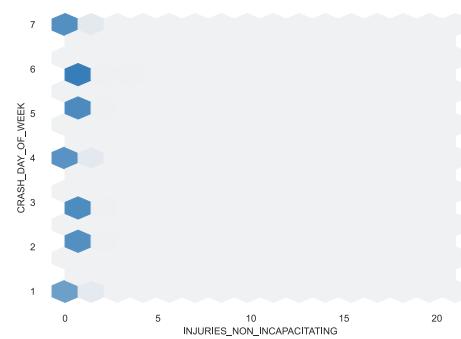


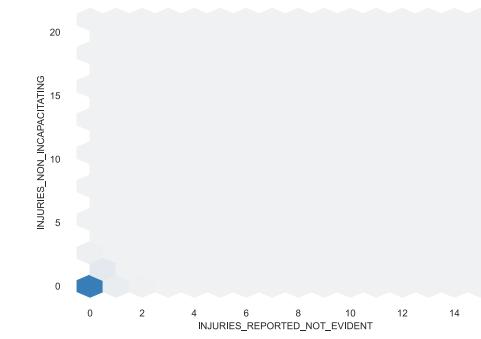
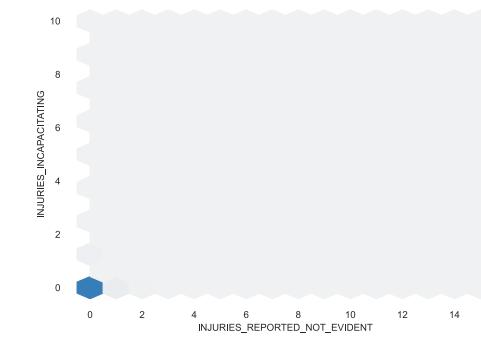
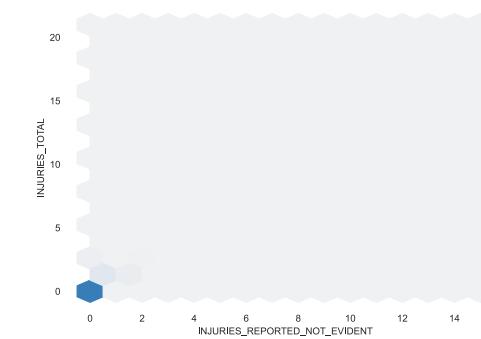
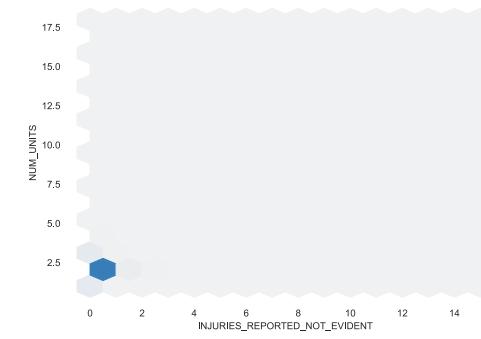
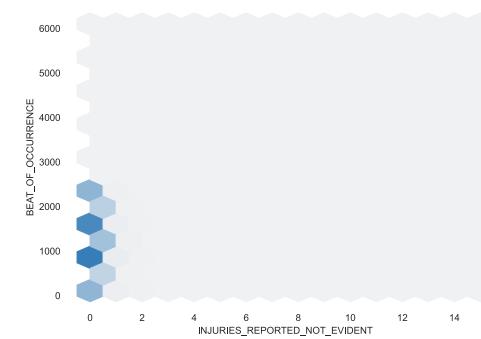
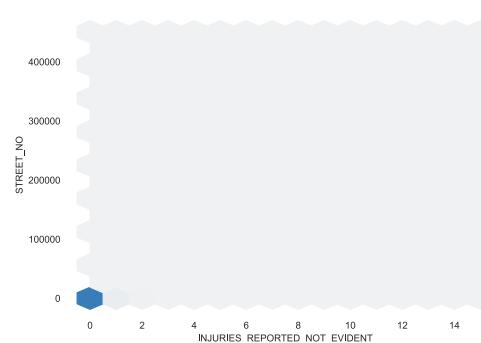


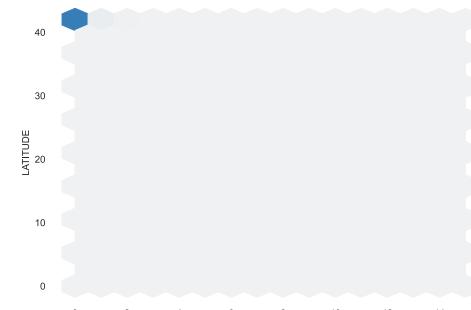
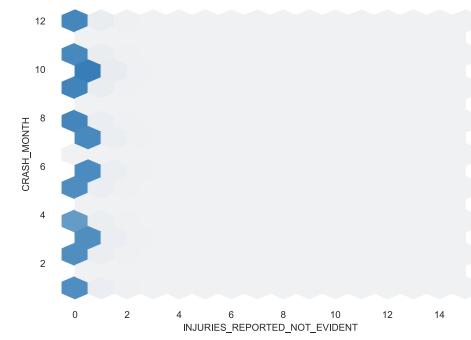
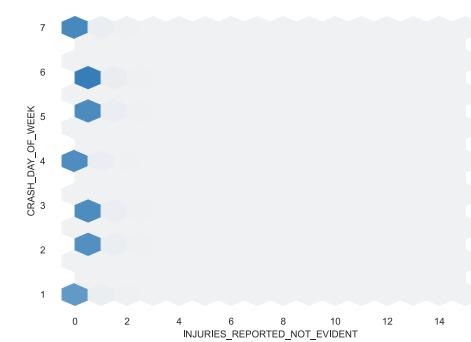
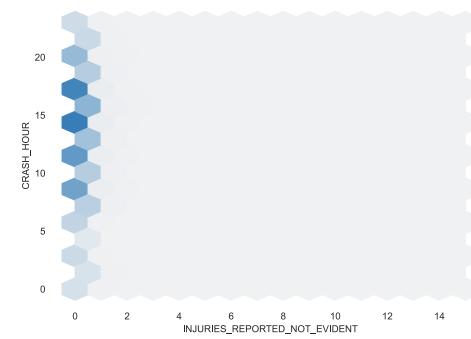
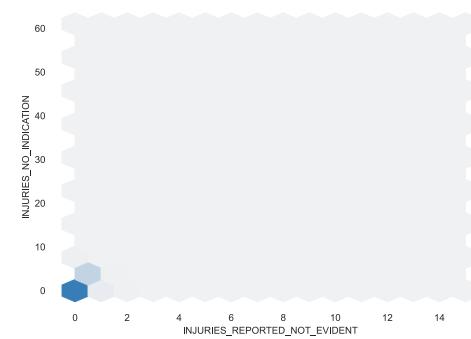
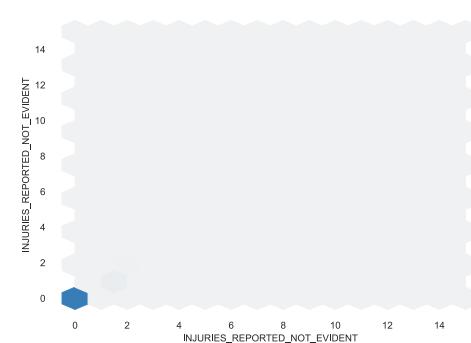


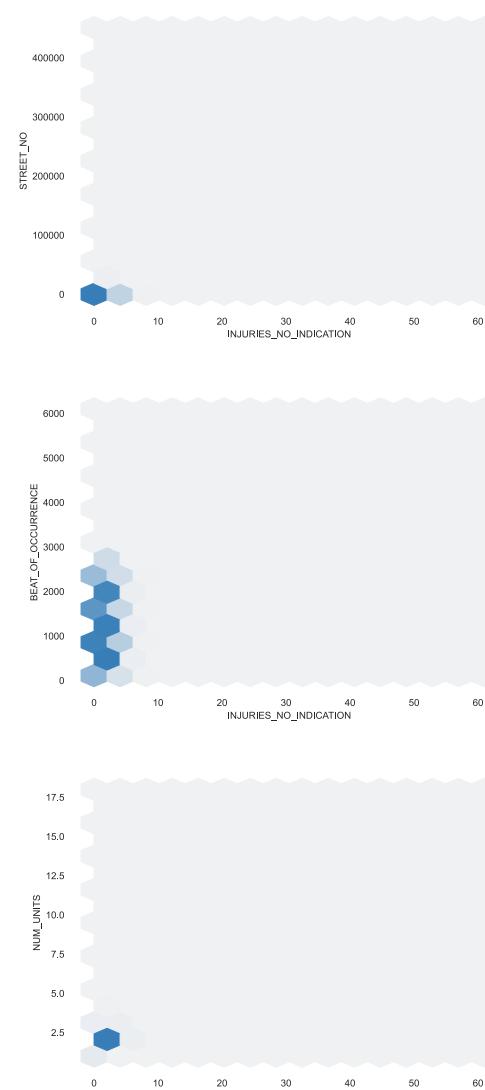
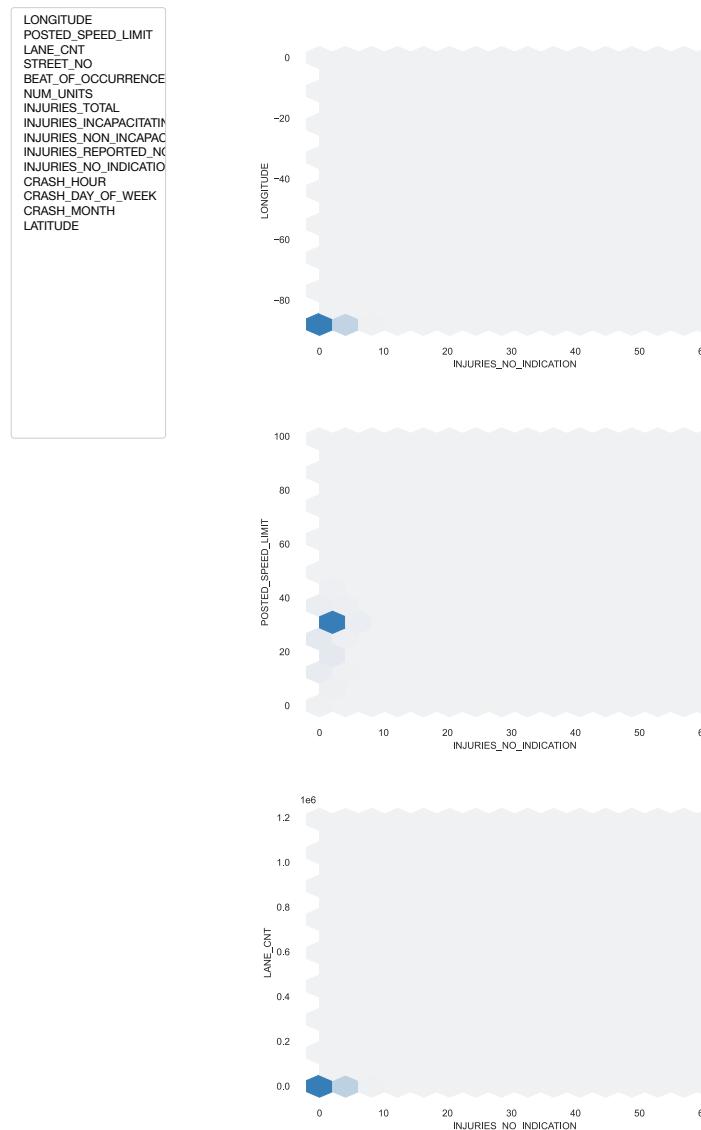


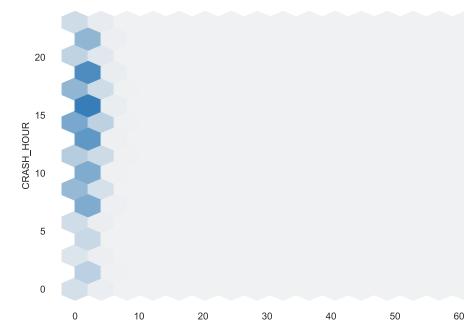
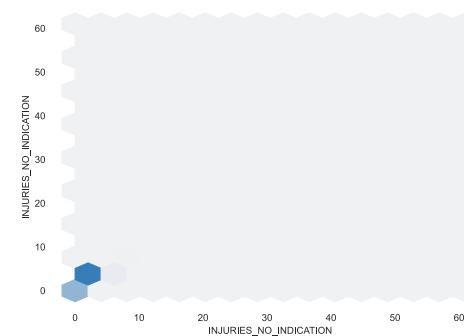
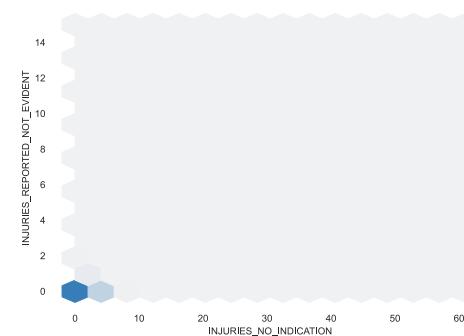
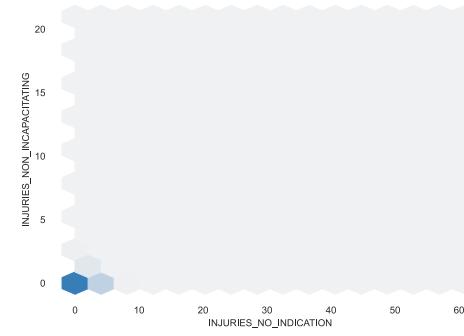
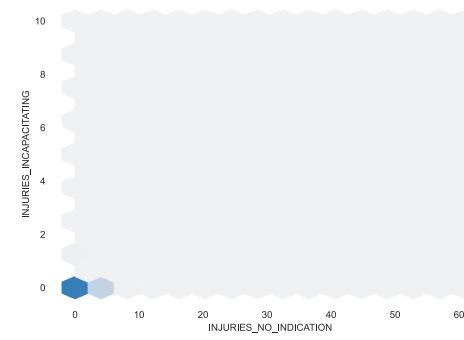
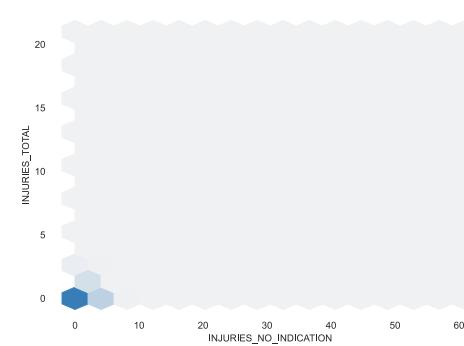


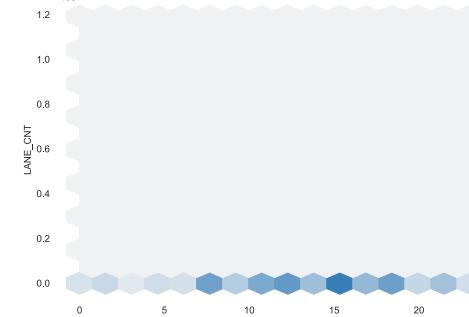
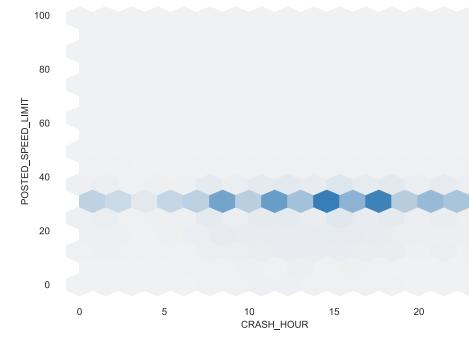
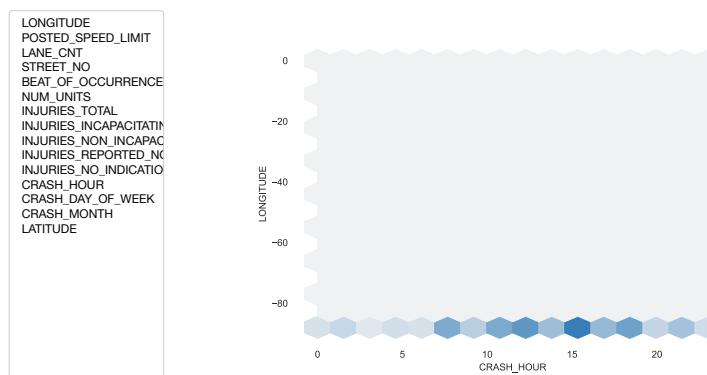
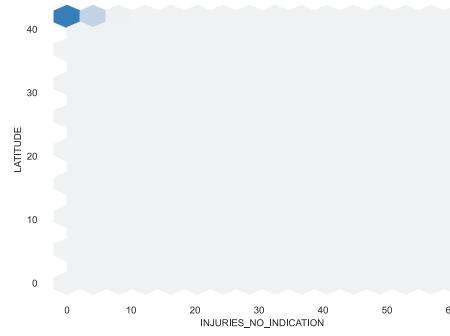
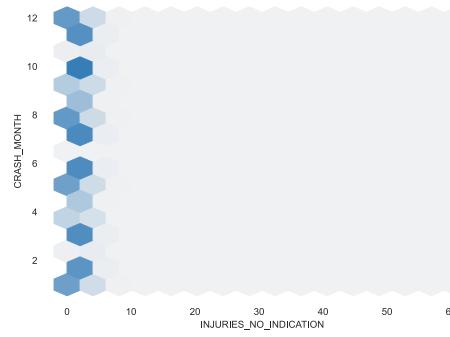
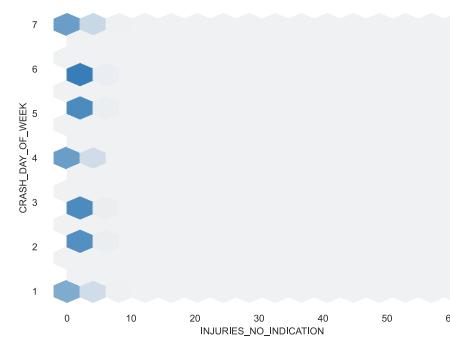


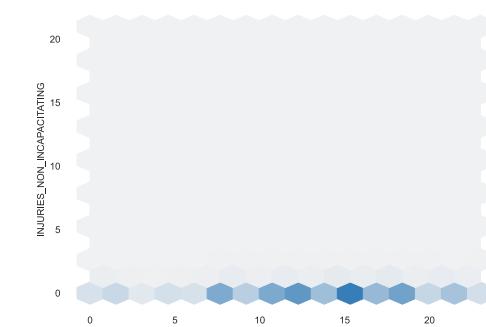
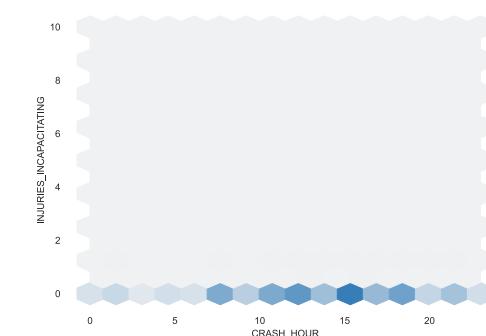
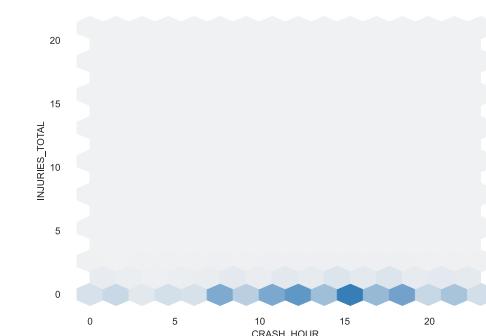
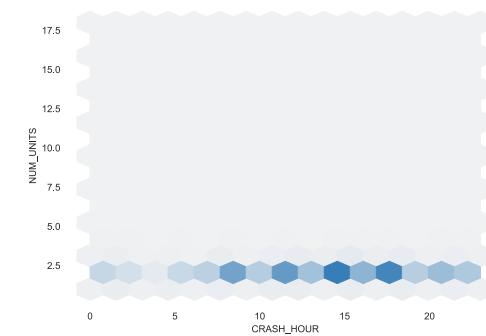
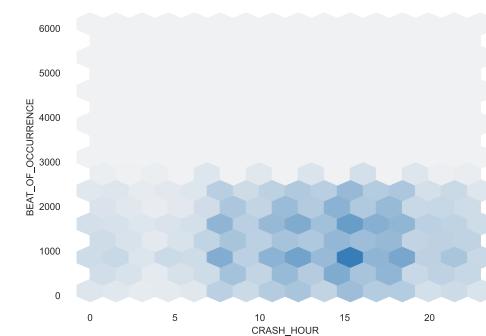
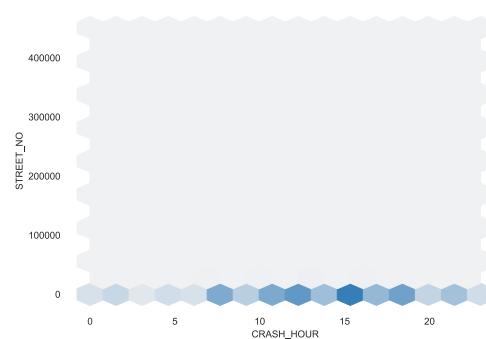


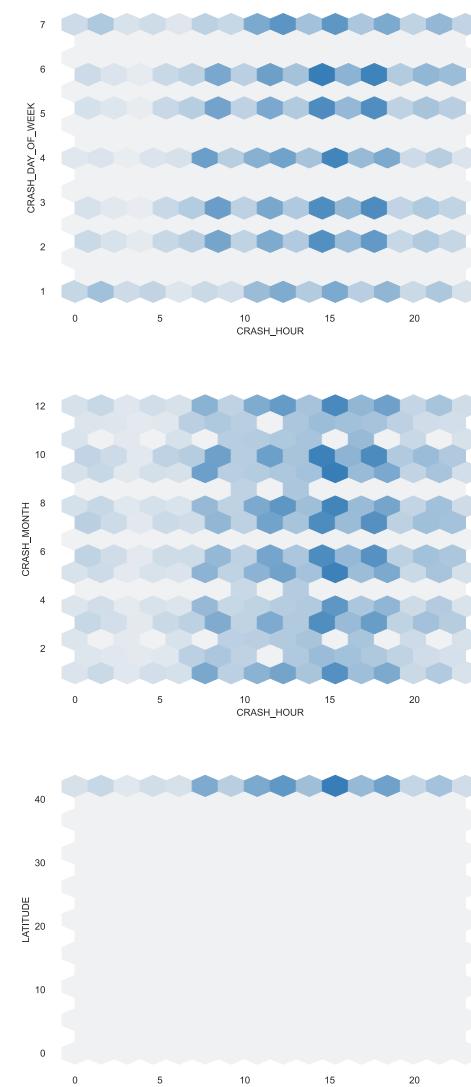
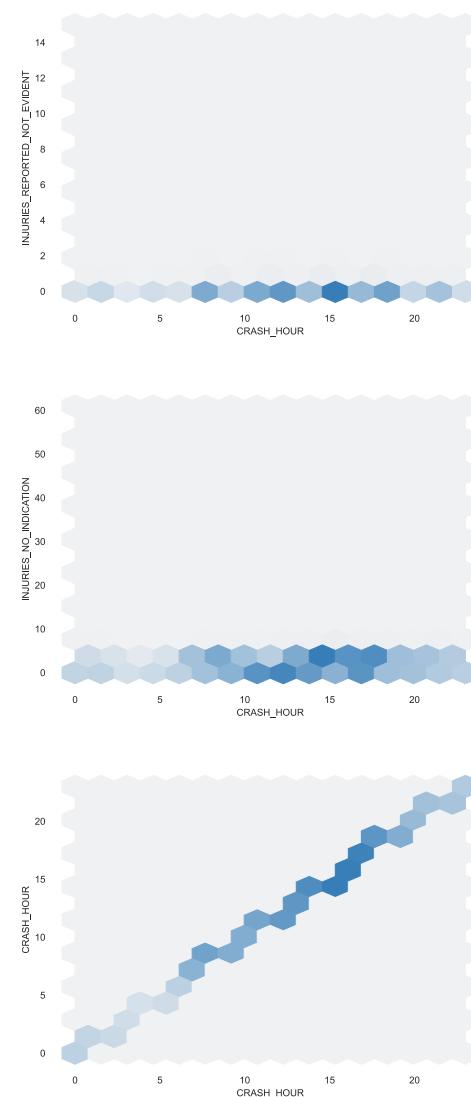


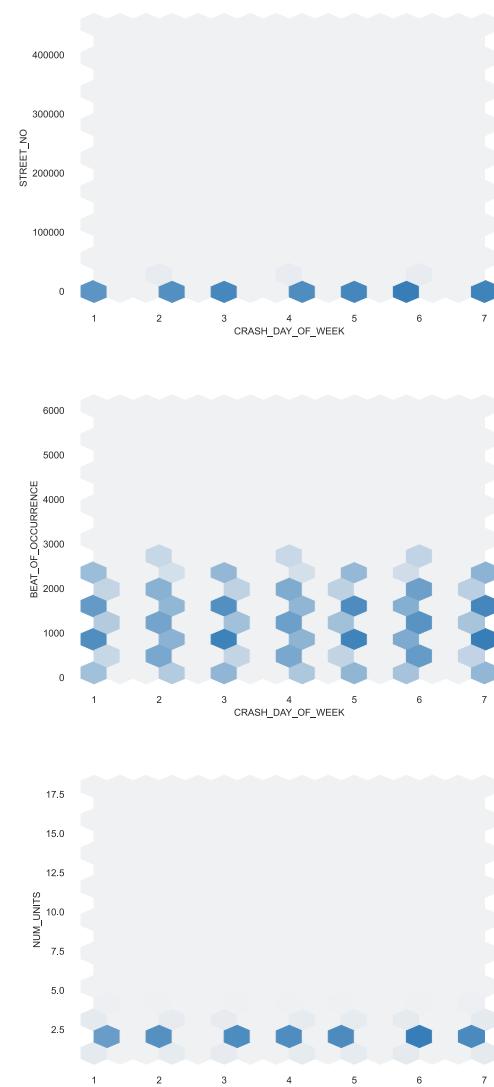
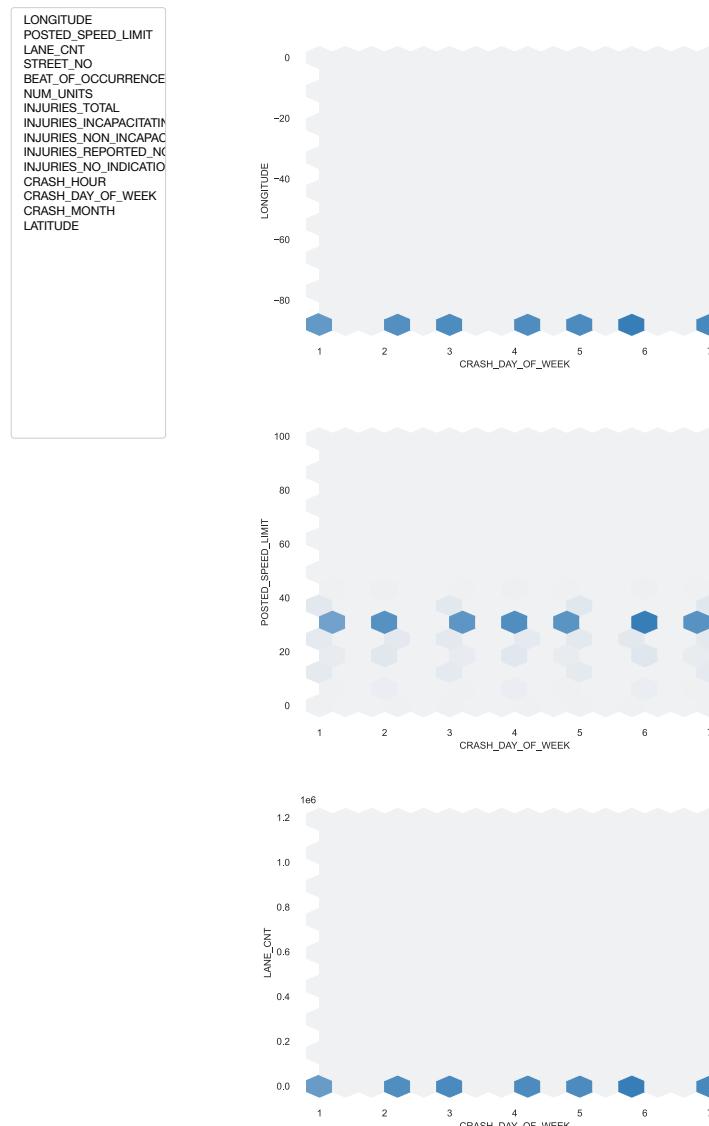


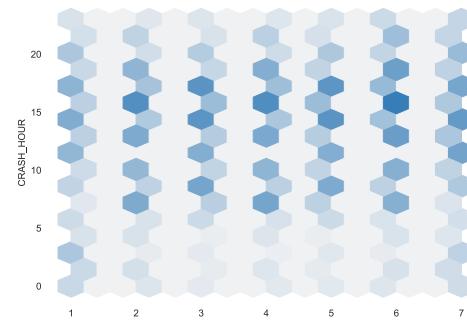
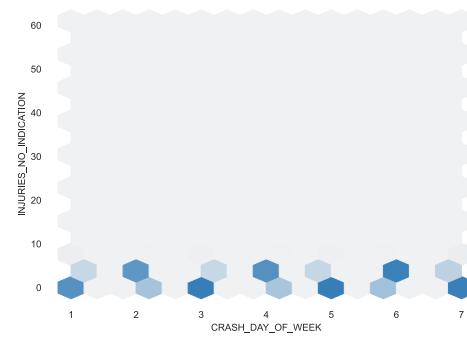
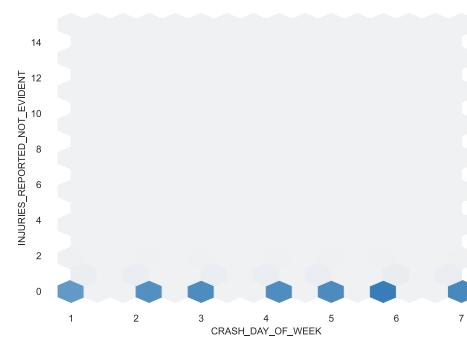
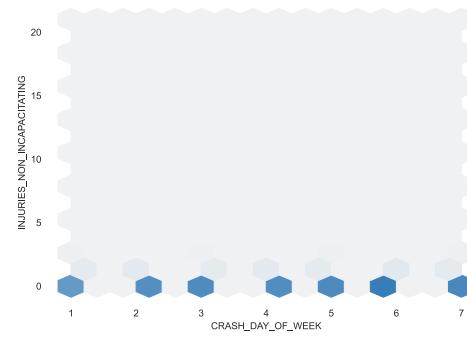
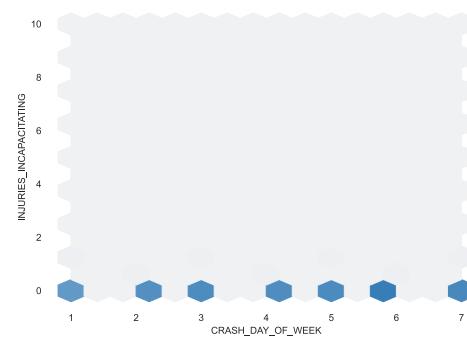
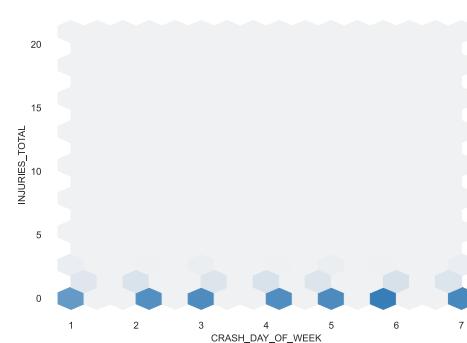


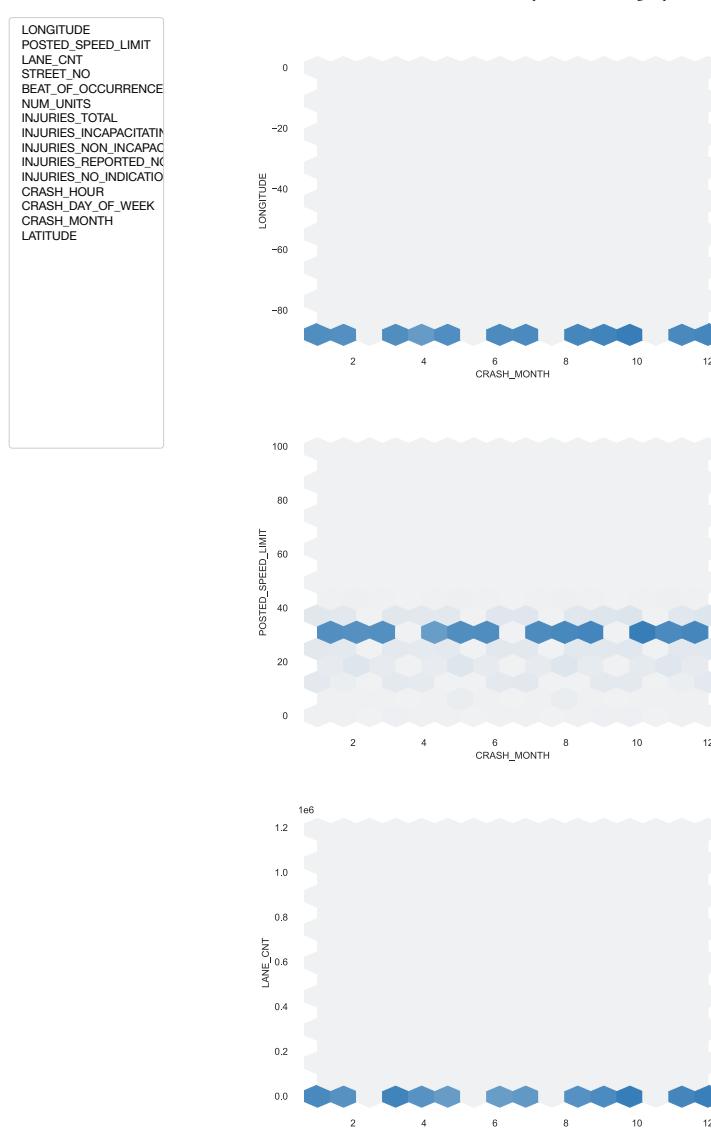
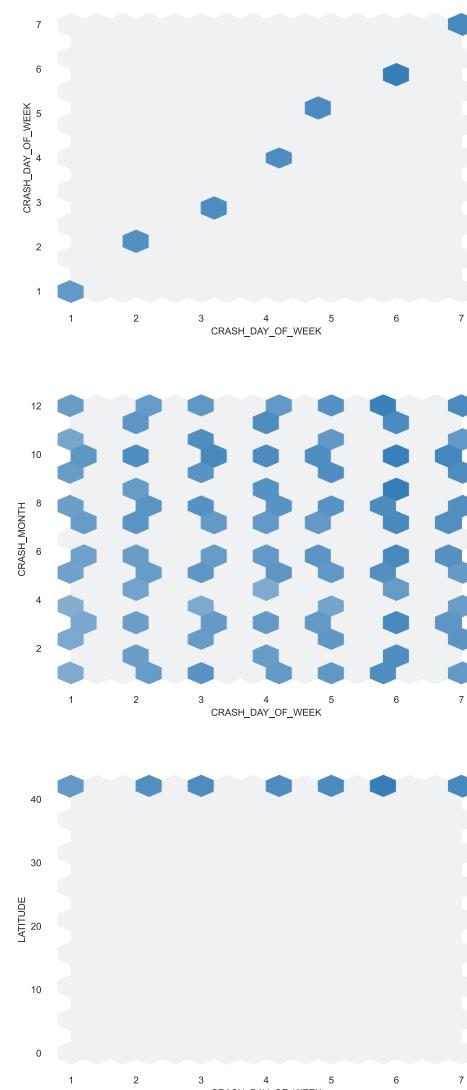


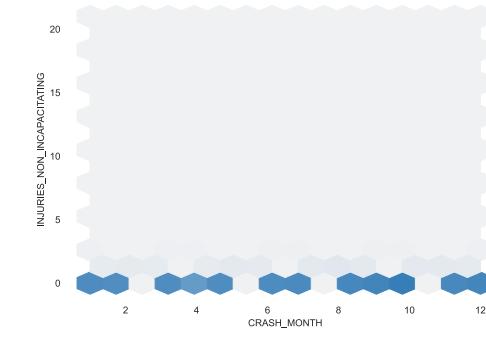
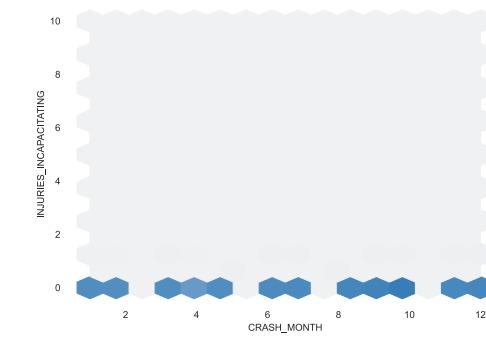
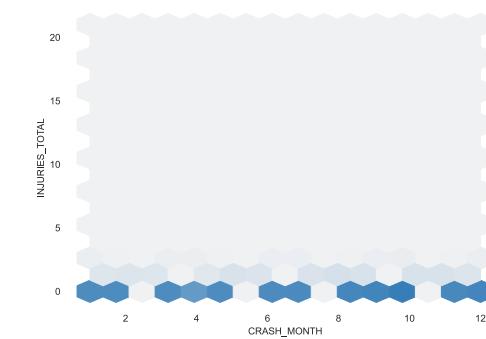
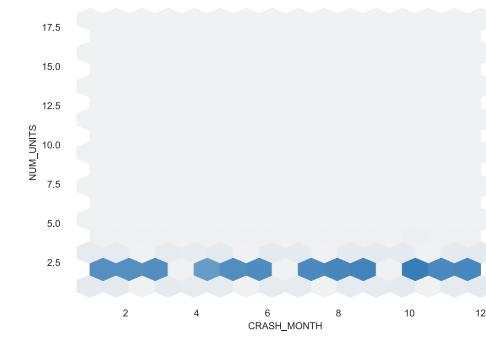
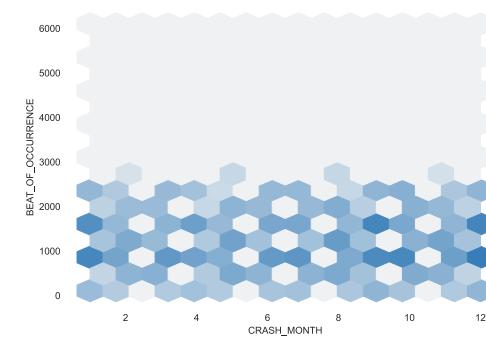
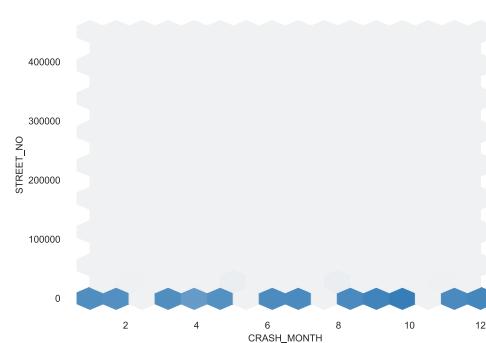


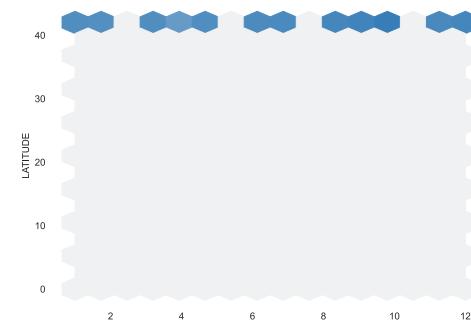
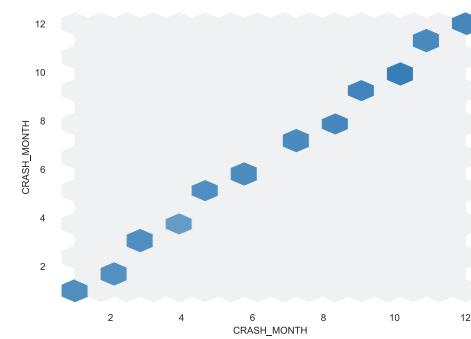
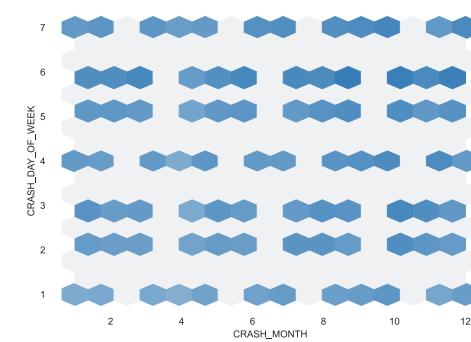
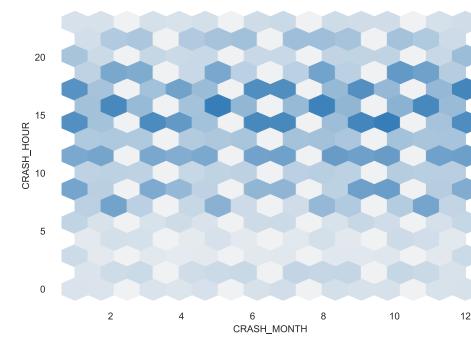
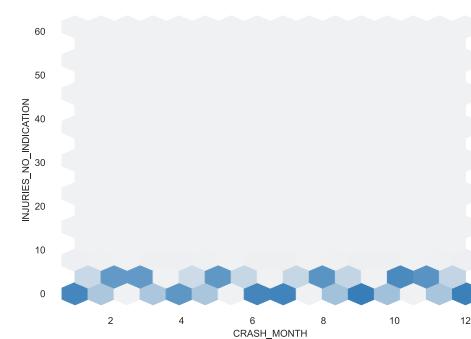
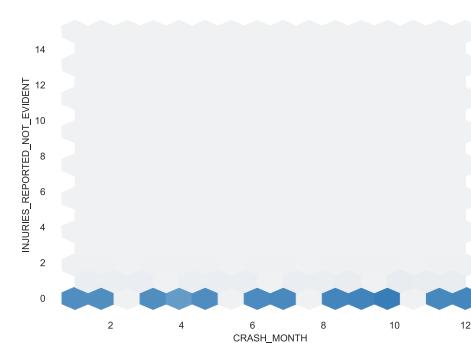


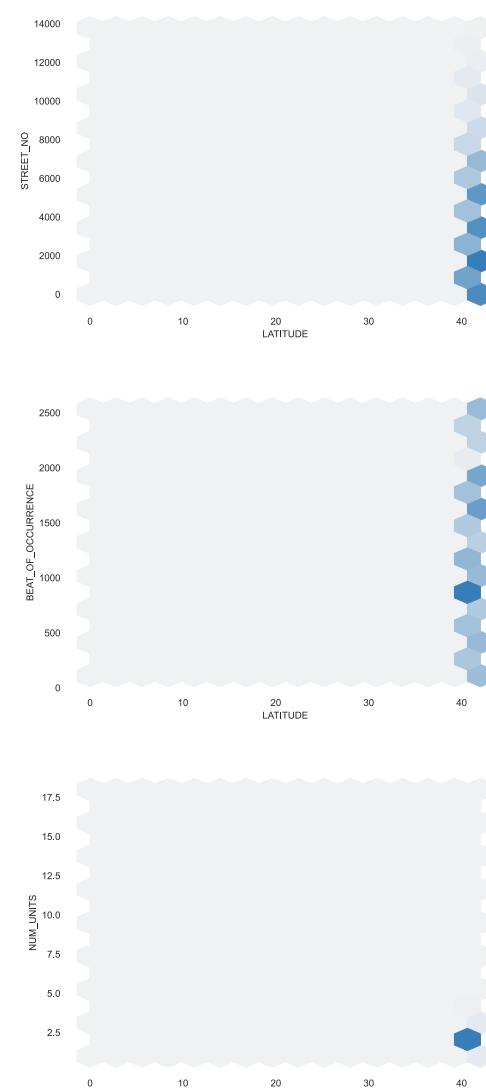
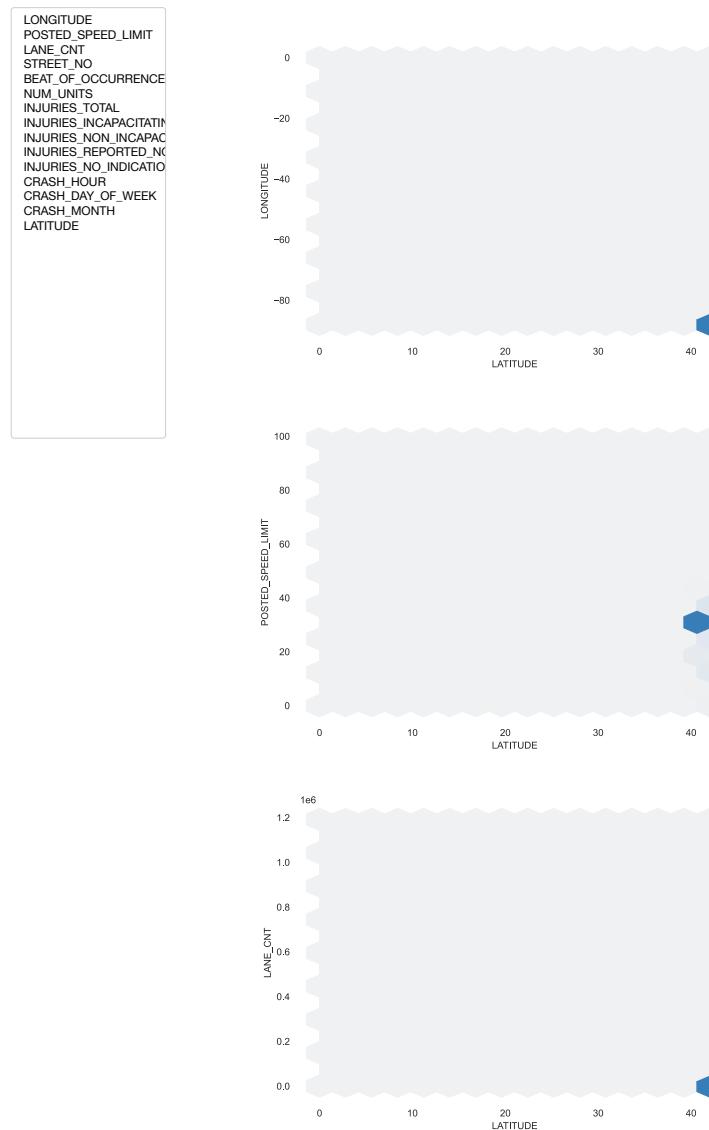


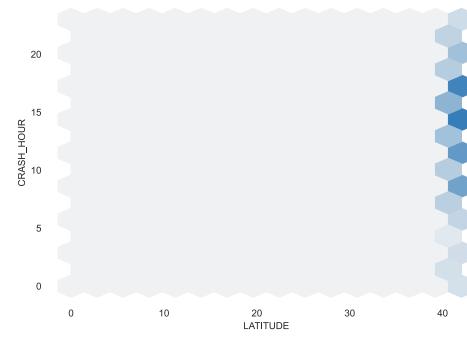
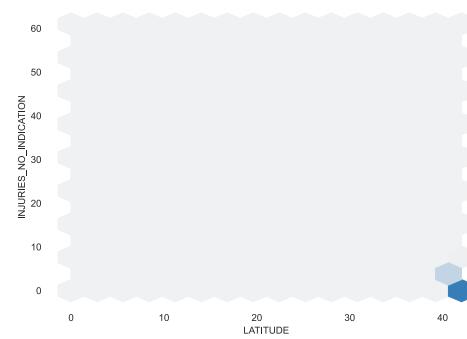
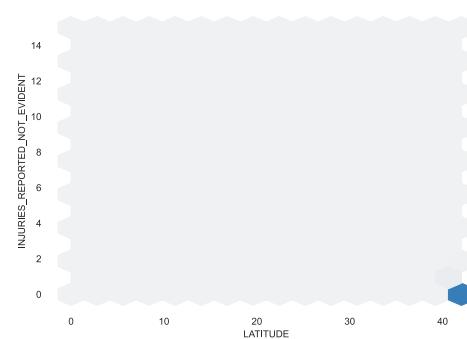
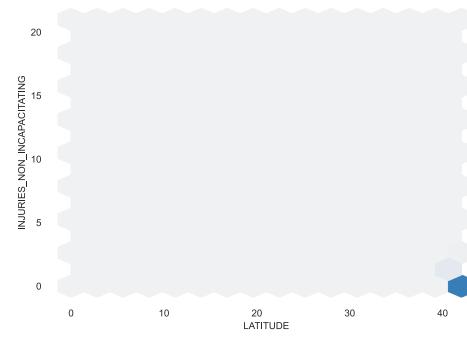
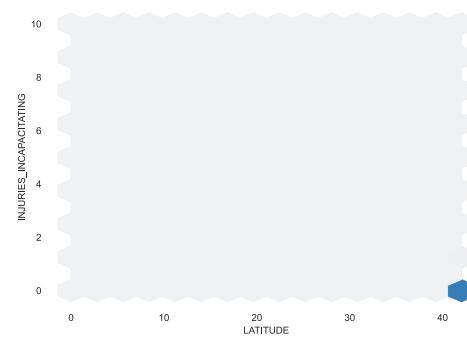
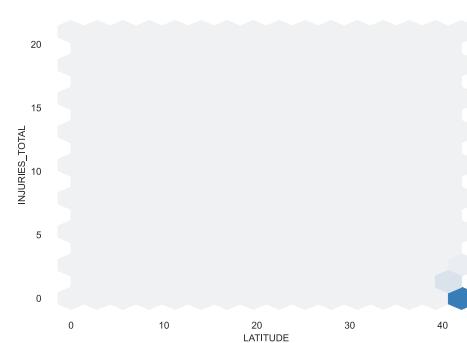




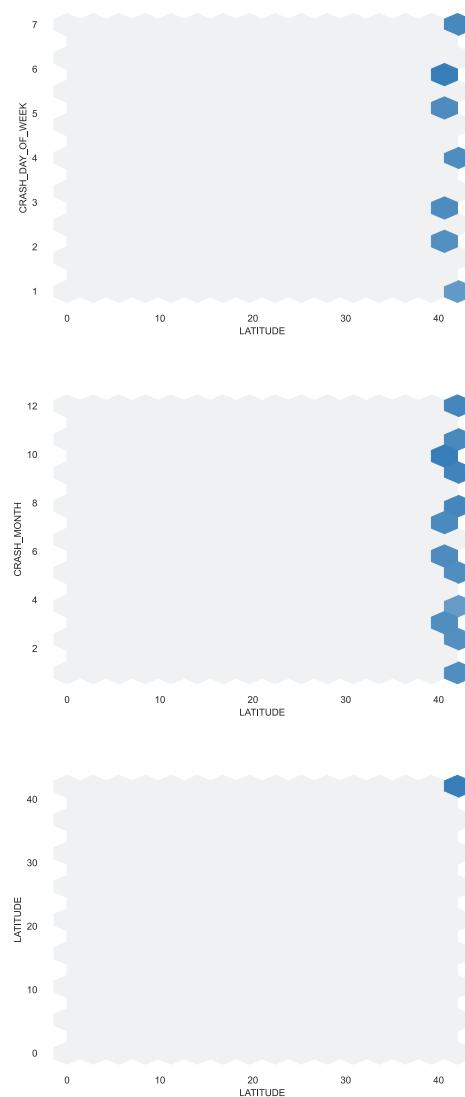




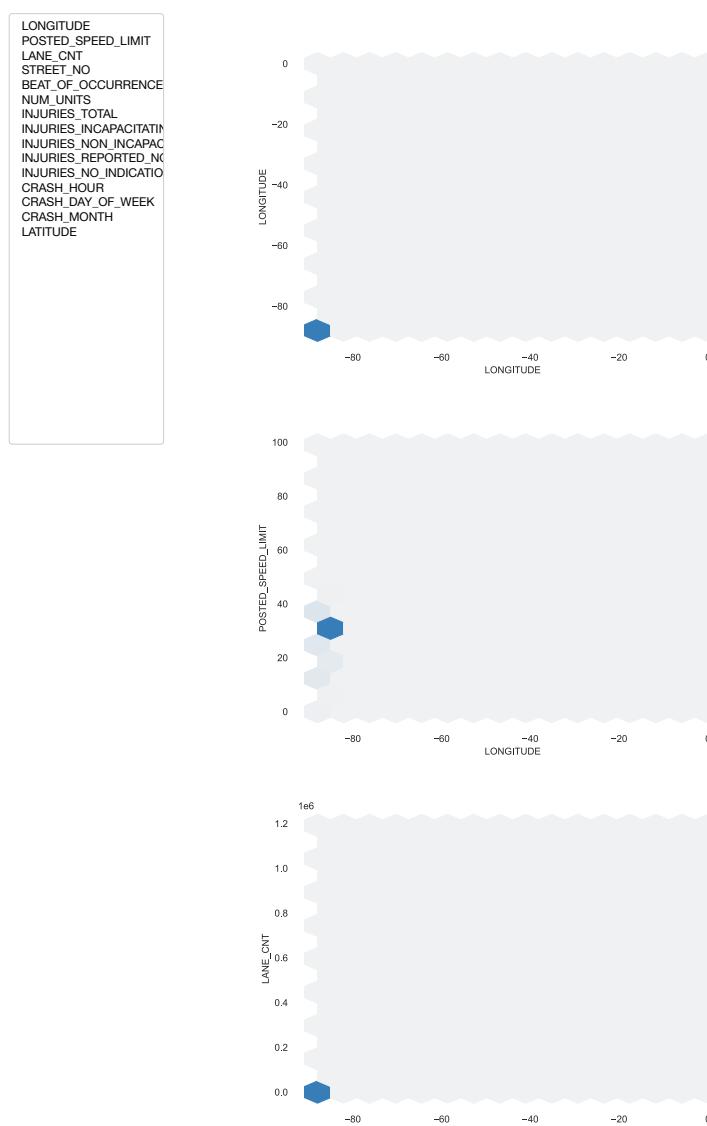


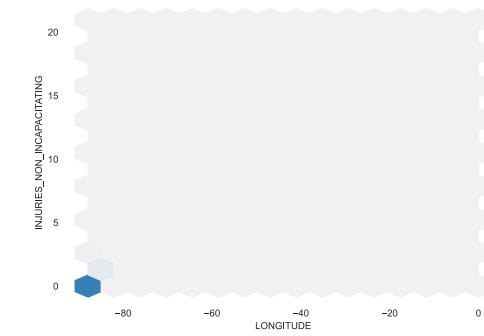
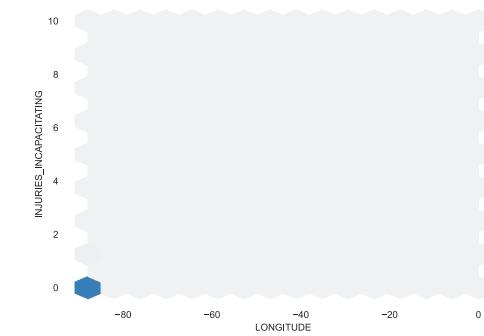
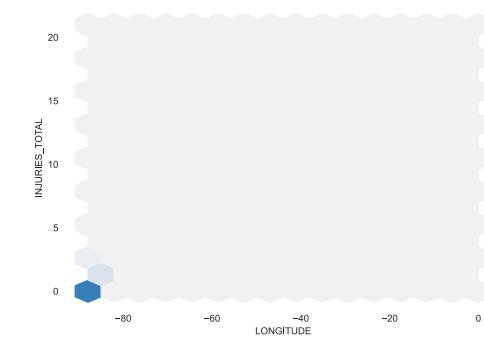
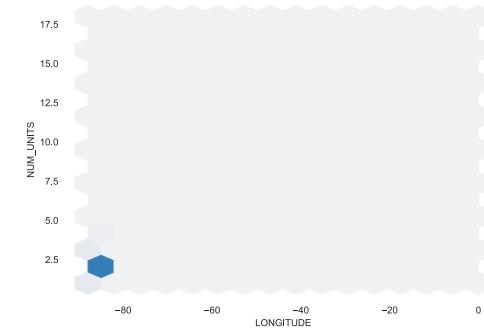
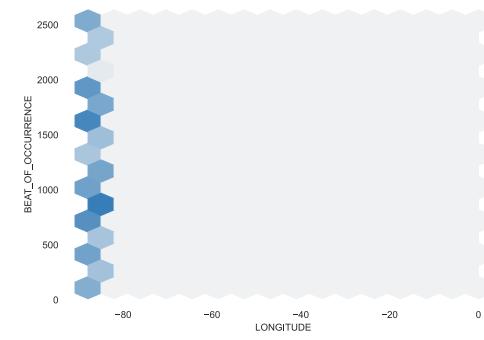
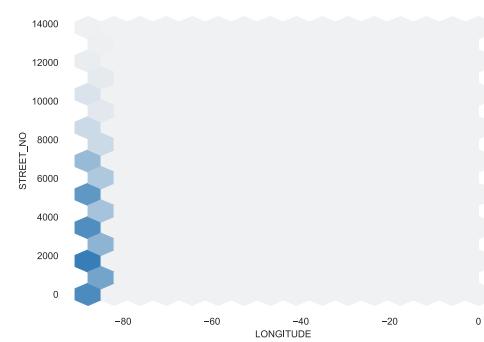


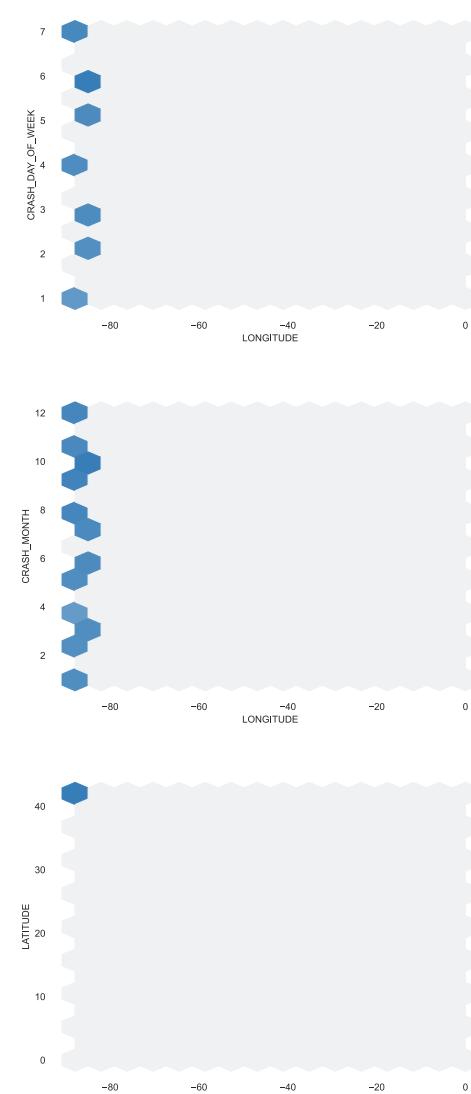
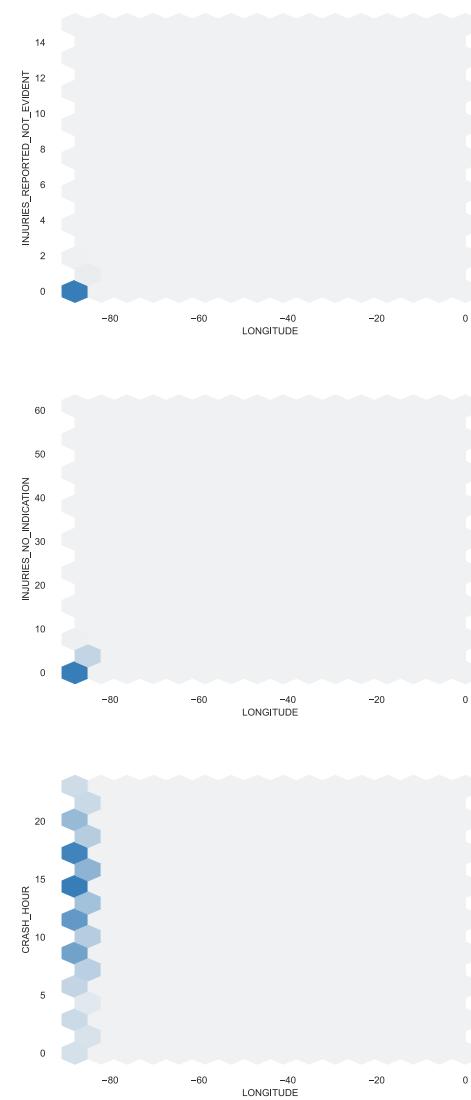
## Austin Crash Report Data - Profiling Report



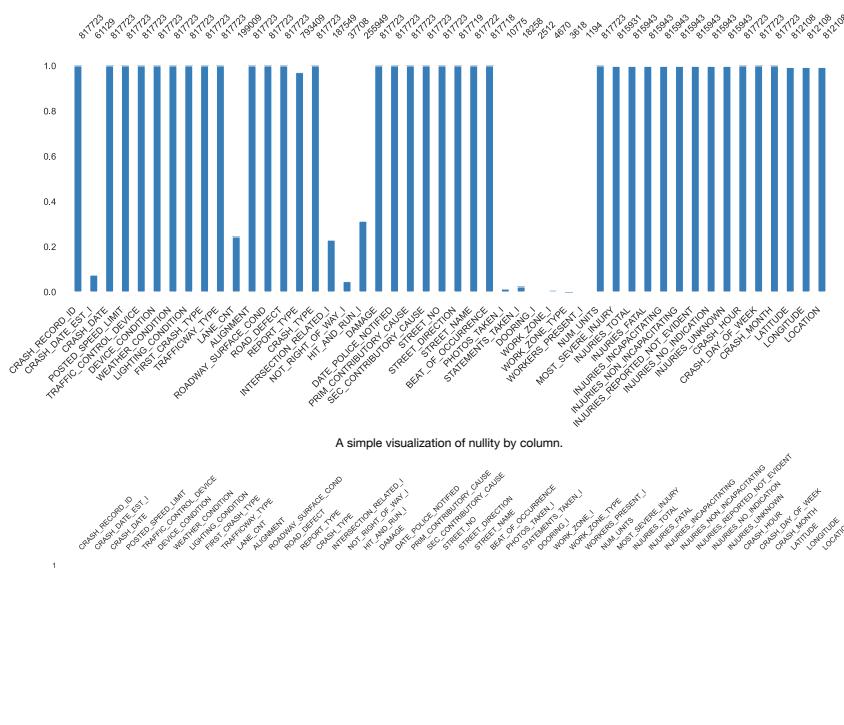
## Austin Crash Report Data - Profiling Report



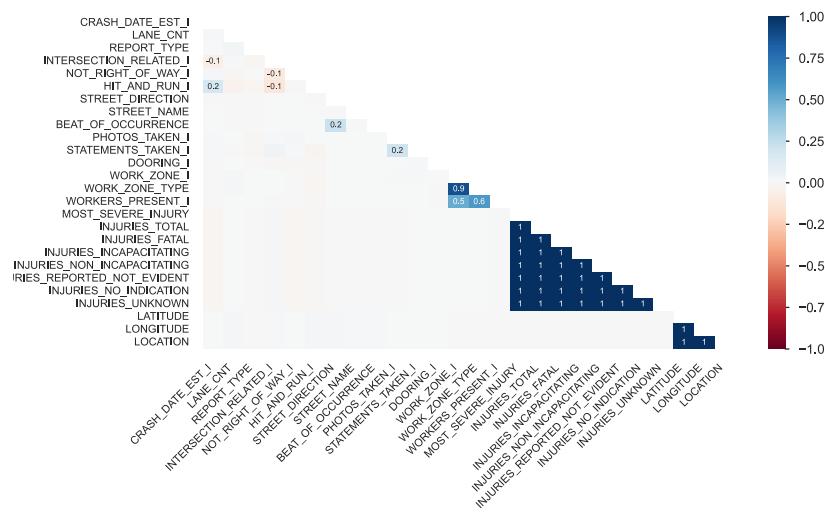




## Missing values



Nullity matrix is a data-dense display which lets you quickly visually pick out patterns in data completion.



## Sample

CRASH_RECORD_ID	CRASH
0 6c1659069e9c6285a650e70d6fb574ed5f64c12888479093dfeef179c0344ec6d2057eae224b5c05dfc278c0a237f8c22543f07def2e4a95a3849871c9345	NaN
1 5f54a59fcb087b12ae5b1acf9f96a3cafa4f2d37e79f8db4106558b34b8a6d2b1af2c91b576ecd7ced08ffd10cf940a84f7613125b89d33636e6075064e22	NaN
2 61fc8c1eb522a6469b460e2134df3d15f82e81fd93e9caf93dc7e631b9e1ba8b450a3fa12bd90d1d2d9b127ea287f88d32e138aeeba17799f536c08048934	NaN
3 004cd14d0303a9163aaad69a2d7f3417d7a2a8572b2ab3378594fbfae8ac53dc604dd8d414f93c290b55862f92517ad32e6209cbc8034c2e26eb3c2bc9724390	NaN
4 a1d5f0ea90897744536a4ccb6cc60329a120d89753fac2b02d69c9685d9cf7c763870a60abd01484a39ed1e6c09b1ba59f38214c03a83ccde1247f794e0287	NaN
5 b236c77d59e32b7b469a6e2f171438b7457e1bd8bc689b14cb4f5b1590cbe784f2b9e554b41925797251cbd3e93a3f4a131d1923b327673d441ae79c052f79c2	NaN
6 35156ce97cab22747495e92e8bb16c57e0e60dc3ce6d1f1852f2f7cece07cae825b073b286b1da52dfa58082ff6d763ecf11f3f06a223c7aed2b6c1e8c5972	NaN
7 0e208d23344f0d1b3a9fc4dbb07676a750dd7b3c397b5c398a33743fd0d49b8ce737c7740b1a77cdaebd61e8e79bdb284452d744b16668f9777258eecc28ff5	NaN
8 14386daec6219c6032b71612b28f0e4cd38e2898f39aae84deeb4efaf165253548cb27cb90088bd2d49d3e484a1167f5076969d4d3b8137a43a3dd94536ac19	NaN
9 359bf9f5872d646bb63576e55b1e0b480d93c2b935ab571dc26ddb48b7a328fbfe130ae70bbff9f03787041b6fb029ba02529da9a1f57494e385ec0e13ed834	NaN

CRASH_RECORD_ID	C
817713 5ee5d998db696abf1c85925f71d82015adb33f3b2cdad1942d7d5cf7f077d3ae22547160f024f0ebd1d4a6875fc58fd1e824aa2d1e0a9c53db3cb3d86efb3038	N

CRASH_RECORD_ID	C
817714	N
817715	Y
817716	N
817717	N
817718	N
817719	N
817720	N
817721	N
817722	N

Report generated by YData ([https://ydata.ai/?utm\\_source=opensource&utm\\_medium=pandasprofiling&utm\\_campaign=report](https://ydata.ai/?utm_source=opensource&utm_medium=pandasprofiling&utm_campaign=report)).