

Pandas Profiling Report

Pandas Profiling Report

- [Overview](#)
- [Variables](#)
- [Interactions](#)
- [Correlations](#)
- [Missing values](#)
- [Sample](#)
- [Duplicate rows](#)

Overview

Brought to you by [YData](#)

- Overview
- Alerts 7
- Reproduction

Dataset statistics

Number of variables	14
Number of observations	660679
Missing cells	19440
Missing cells (%)	0.2%
Duplicate rows	19
Duplicate rows (%)	< 0.1%
Total size in memory	70.6 MiB
Average record size in memory	112.0 B

Variable types

Text	2
Categorical	7
DateTime	1
Numeric	4

Alerts	Duplicates
Dataset has 19 (< 0.1%) duplicate rows	
Accident_Severity is highly imbalanced (58.0%)	Imbalance
Light_Conditions is highly imbalanced (51.8%)	Imbalance
Road_Surface_Conditions is highly imbalanced (52.0%)	Imbalance
Weather_Conditions is highly imbalanced (64.9%)	Imbalance
Vehicle_Type is highly imbalanced (60.2%)	Imbalance
Weather_Conditions has 14128 (2.1%) missing values	Missing

Reproduction

Analysis started	2025-10-05 13:58:55.694516
Analysis finished	2025-10-05 13:59:58.511240
Duration	1 minute and 2.82 seconds
Software version	ydata-profiling vv4.12.1
Download configuration	config.json

Variables

Select Columns Index Accident_Severity Accident Date Latitude Light_Conditions
District Area Longitude Number_of_Casualties Number_of_Vehicles
Road_Surface_Conditions Road_Type Urban_or_Rural_Area Weather_Conditions
Vehicle_Type

[Index](#)

Text

Distinct	421020
Distinct (%)	63.7%
Missing	0
Missing (%)	0.0%
Memory size	5.0 MiB

2025-10-05T19:30:00.020685image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
More details

- Overview
- Words
- Characters

Length

Max length	13
Median length	13
Mean length	11.183803
Min length	8

Characters and Unicode

Total characters	7388904
Distinct characters	38
Distinct categories	4 ?
Distinct scripts	2 ?
Distinct blocks	1 ?

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

Unique

Unique	420889 ?
Unique (%)	63.7%

Sample

1st row	200701BS64157
2nd row	200701BS65737
3rd row	200701BS66127
4th row	200701BS66128
5th row	200701BS66837

Value	Count	Frequency (%)
2.01e+12	239478	36.2%
2.01e+239	4	< 0.1%
2.01e+282	4	< 0.1%
2.01e+175	4	< 0.1%
2.01e+260	4	< 0.1%
2.01e+292	4	< 0.1%
2.01e+199	4	< 0.1%
2.01e+197	4	< 0.1%
2.01e+98	4	< 0.1%
2.01e+86	4	< 0.1%
Other values (421010)	421165	63.7%

2025-10-05T19:30:02.112222image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
0	1964672	26.6%
2	1222229	16.5%
1	1019189	13.8%
7	312183	4.2%
4	284181	3.8%
E	283347	

Value	Count	Frequency (%)
		3.8%
9	270527	3.7%
8	268977	3.6%
3	265171	3.6%
+	239903	3.2%
Other values (28)	1258525	17.0%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	6023433	81.5%
Uppercase Letter	885665	12.0%
Math Symbol	239903	3.2%
Other Punctuation	239903	3.2%

Most frequent character per category

Uppercase Letter

Value	Count	Frequency (%)
E	283347	32.0%
C	62372	7.0%
D	61464	6.9%
A	52750	6.0%
B	46644	5.3%
W	30465	3.4%
K	29051	3.3%
G	27237	3.1%
F	26931	3.0%
L	24665	2.8%
Other values (16)	240739	27.2%

Decimal Number

Value	Count	Frequency (%)
0	1964672	32.6%
2	1222229	20.3%
1	1019189	16.9%
7	312183	5.2%
4	284181	4.7%
9	270527	4.5%
8	268977	4.5%
3	265171	4.4%
5	222607	3.7%
6	193697	3.2%

Math Symbol

Value	Count	Frequency (%)
+	239903	100.0%

Other Punctuation

Value	Count	Frequency (%)
.	239903	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Common	6503239	88.0%
Latin	885665	12.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
E	283347	32.0%
C	62372	7.0%
D	61464	6.9%
A	52750	6.0%
B	46644	

Value	Count	Frequency (%)
		5.3%
W	30465	3.4%
K	29051	3.3%
G	27237	3.1%
F	26931	3.0%
L	24665	2.8%
Other values (16)	240739	27.2%

Common

Value	Count	Frequency (%)
0	1964672	30.2%
2	1222229	18.8%
1	1019189	15.7%
7	312183	4.8%
4	284181	4.4%
9	270527	4.2%
8	268977	4.1%
3	265171	4.1%
+	239903	3.7%
.	239903	3.7%
Other values (2)	416304	6.4%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	7388904	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
0	1964672	26.6%
2	1222229	16.5%

Value	Count	Frequency (%)
1	1019189	13.8%
7	312183	4.2%
4	284181	3.8%
E	283347	3.8%
9	270527	3.7%
8	268977	3.6%
3	265171	3.6%
+	239903	3.2%
Other values (28)	1258525	17.0%

[Accident_Severity](#)

Categorical

Imbalance

Distinct		3
Distinct (%)		< 0.1%
Missing		0
Missing (%)		0.0%
Memory size		5.0 MiB
Slight	563801	
Serious	88217	
Fatal	8661	

More details

- Overview
- Categories
- Words
- Characters

Length

Max length	7
Median length	6
Mean length	6.1204155
Min length	5

Characters and Unicode

Total characters	4043630
Distinct characters	13
Distinct categories	2 ?
Distinct scripts	1 ?
Distinct blocks	1 ?

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

Unique

Unique	0 ?
Unique (%)	0.0%

Sample

1st row	Serious
2nd row	Serious
3rd row	Serious
4th row	Serious
5th row	Slight

Common Values

Value	Count	Frequency (%)
Slight	563801	85.3%
Serious	88217	13.4%
Fatal	8661	1.3%

Length

2025-10-05T19:30:02.590322image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
Histogram of lengths of the category
Common Values (Plot)

2025-10-05T19:30:02.948020image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

Value	Count	Frequency (%)
slight	563801	85.3%
serious	88217	13.4%
fatal	8661	1.3%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
S	652018	16.1%
i	652018	16.1%
l	572462	14.2%
t	572462	14.2%
g	563801	13.9%
h	563801	13.9%
e	88217	2.2%
r	88217	2.2%
o	88217	2.2%
u	88217	2.2%
Other values (3)	114200	2.8%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	3382951	83.7%
Uppercase Letter	660679	16.3%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
i	652018	19.3%
l	572462	16.9%
t	572462	16.9%
g	563801	16.7%
h	563801	16.7%
e	88217	2.6%
r	88217	2.6%
o	88217	2.6%
u	88217	2.6%
s	88217	2.6%

Uppercase Letter

Value	Count	Frequency (%)
S	652018	98.7%
F	8661	1.3%

Most occurring scripts

Value	Count	Frequency (%)
Latin	4043630	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
S	652018	16.1%
i	652018	16.1%
l	572462	14.2%
t	572462	14.2%
g	563801	13.9%
h	563801	13.9%
e	88217	2.2%
r	88217	2.2%
o	88217	2.2%
u	88217	2.2%
Other values (3)	114200	2.8%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	4043630	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
S	652018	16.1%
i	652018	16.1%
l	572462	14.2%
t	572462	14.2%
g	563801	13.9%
h	563801	13.9%
e	88217	2.2%

Value	Count	Frequency (%)
r	88217	2.2%
o	88217	2.2%
u	88217	2.2%
Other values (3)	114200	2.8%

[Accident Date](#)

Date

Distinct	1461
Distinct (%)	0.2%
Missing	0
Missing (%)	0.0%
Memory size	5.0 MiB
Minimum	2019-01-01 00:00:00
Maximum	2022-12-31 00:00:00
Invalid dates	0
Invalid dates (%)	0.0%

2025-10-05T19:30:03.332642image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
More details

- Histogram

2025-10-05T19:30:03.825685image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
Histogram with fixed size bins (bins=50)

[Latitude](#)

Real number (ℝ)

Distinct	511618
Distinct (%)	77.4%
Missing	25
Missing (%)	< 0.1%
Infinite	0
Infinite (%)	0.0%
Mean	52.553866
Minimum	49.91443
Maximum	60.757544
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	5.0 MiB

2025-10-05T19:30:04.308946image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

More details

- Statistics
- Histogram
- Common values
- Extreme values

Quantile statistics

Minimum	49.91443
5-th percentile	50.822129
Q1	51.49069
median	52.315641
Q3	53.453452
95-th percentile	55.615454
Maximum	60.757544
Range	10.843114
Interquartile range (IQR)	1.962762

Descriptive statistics

Standard deviation	1.4069222
Coefficient of variation (CV)	0.02677105
Kurtosis	0.93931855
Mean	52.553866
Median Absolute Deviation (MAD)	0.903863
Skewness	0.99848763
Sum	34719922
Variance	1.97943
Monotonicity	Not monotonic

2025-10-05T19:30:04.791568image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
52.458798	50	< 0.1%
52.949719	45	< 0.1%
53.083165	33	< 0.1%
52.967634	31	< 0.1%
52.93886	29	< 0.1%
52.944347	28	< 0.1%
52.989857	27	< 0.1%

Value	Count	Frequency (%)
51.496389	27	< 0.1%
52.940243	26	< 0.1%
54.96844	25	< 0.1%
Other values (511608)	660333	99.9%
(Missing)	25	< 0.1%

- Minimum 10 values
- Maximum 10 values

Value	Count	Frequency (%)
49.91443	1	< 0.1%
49.914488	1	< 0.1%
49.914949	1	< 0.1%
49.915159	1	< 0.1%
49.91573	1	< 0.1%
49.917703	1	< 0.1%
49.917913	1	< 0.1%
49.967938	1	< 0.1%
49.99465	1	< 0.1%
49.99533	1	< 0.1%
Value	Count	Frequency (%)
60.757544	1	< 0.1%
60.714772	1	< 0.1%
60.668921	1	< 0.1%
60.665748	1	< 0.1%
60.648412	1	< 0.1%
60.598055	1	< 0.1%
60.535623	1	< 0.1%
60.521829	1	< 0.1%
60.454772	1	< 0.1%
60.454326	1	< 0.1%

Light_Conditions

Categorical

Imbalance

Distinct	5
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%

Memory size	5.0 MiB
Daylight	484880
Darkness - lights lit	129335
Darkness - no lighting	37437
Darkness - lighting unknown	6484
Darkness - lights unlit	2543

More details

- Overview
- Categories
- Words
- Characters

Length

Max length	27
Median length	8
Mean length	11.582396
Min length	8

Characters and Unicode

Total characters	7652246
Distinct characters	18
Distinct categories	4 ?
Distinct scripts	2 ?
Distinct blocks	1 ?

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

Unique

Unique	0 ?
Unique (%)	0.0%

Sample

1st row	Darkness - lights lit
2nd row	Daylight
3rd row	Darkness - lighting unknown
4th row	Daylight
5th row	Darkness - lights lit

Common Values

Value	Count	Frequency (%)
Daylight	484880	73.4%

Value	Count	Frequency (%)
Darkness - lights lit	129335	19.6%
Darkness - no lighting	37437	5.7%
Darkness - lighting unknown	6484	1.0%
Darkness - lights unlit	2543	0.4%

Length

2025-10-05T19:30:05.195934image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
Histogram of lengths of the category
Common Values (Plot)

2025-10-05T19:30:05.581237image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

Value	Count	Frequency (%)
daylight	484880	40.8%
darkness	175799	14.8%
	175799	14.8%
lights	131878	11.1%
lit	129335	10.9%
lighting	43921	3.7%
no	37437	3.2%
unknown	6484	0.5%
unlit	2543	0.2%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
i	836478	10.9%
l	792557	10.4%
t	792557	10.4%
g	704600	9.2%

Value	Count	Frequency (%)
D	660679	8.6%
h	660679	8.6%
a	660679	8.6%
	527397	6.9%
y	484880	6.3%
s	483476	6.3%
Other values (8)	1048264	13.7%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	6288371	82.2%
Uppercase Letter	660679	8.6%
Space Separator	527397	6.9%
Dash Punctuation	175799	2.3%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
i	836478	13.3%
l	792557	12.6%
t	792557	12.6%
g	704600	11.2%
h	660679	10.5%
a	660679	10.5%
y	484880	7.7%
s	483476	7.7%
n	279152	4.4%
k	182283	2.9%
Other values (5)	411030	6.5%

Uppercase Letter

Value	Count	Frequency (%)
D	660679	100.0%

Space Separator

Value	Count	Frequency (%)
	527397	100.0%

Dash Punctuation

Value	Count	Frequency (%)
-	175799	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	6949050	90.8%
Common	703196	9.2%

Most frequent character per script

Latin

Value	Count	Frequency (%)
i	836478	12.0%
l	792557	11.4%
t	792557	11.4%
g	704600	10.1%
D	660679	9.5%
h	660679	9.5%
a	660679	9.5%
y	484880	7.0%
s	483476	7.0%
n	279152	4.0%
Other values (6)	593313	8.5%

Common

Value	Count	Frequency (%)
	527397	75.0%
-	175799	25.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	7652246	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
i	836478	10.9%
l	792557	10.4%
t	792557	10.4%
g	704600	9.2%
D	660679	8.6%
h	660679	8.6%

Value	Count	Frequency (%)
a	660679	8.6%
	527397	6.9%
y	484880	6.3%
s	483476	6.3%
Other values (8)	1048264	13.7%

District Area

Text

Distinct	422
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Memory size	5.0 MiB

2025-10-05T19:30:06.509884image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
More details

- Overview
- Words
- Characters

Length

Max length	28
Median length	23
Mean length	10.782792
Min length	4

Characters and Unicode

Total characters	7123964
Distinct characters	52
Distinct categories	7 ?
Distinct scripts	2 ?
Distinct blocks	1 ?

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

Unique

Unique	0 ?
Unique (%)	0.0%

Sample

1st row	Kensington and Chelsea
2nd row	Kensington and Chelsea
3rd row	Kensington and Chelsea
4th row	Kensington and Chelsea

5th row		Kensington and Chelsea
Value	Count	Frequency (%)
and	35481	3.7%
south	25795	2.7%
of	23415	2.4%
north	23392	2.4%
east	21943	2.3%
city	20023	2.1%
west	14387	1.5%
birmingham	13491	1.4%
forest	9966	1.0%
upon	9436	1.0%
Other values (442)	761229	79.4%

2025-10-05T19:30:07.888845image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
e	717756	10.1%
r	547856	7.7%
o	497620	7.0%
n	451974	6.3%
a	448157	6.3%
t	427258	6.0%
i	363374	

Value	Count	Frequency (%)
		5.1%
s	349676	4.9%
h	334713	4.7%
l	310391	4.4%
Other values (42)	2675189	37.6%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	5882886	82.6%
Uppercase Letter	898087	12.6%
Space Separator	297879	4.2%
Other Punctuation	24384	0.3%
Dash Punctuation	20370	0.3%
Open Punctuation	179	< 0.1%
Close Punctuation	179	< 0.1%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
e	717756	12.2%
r	547856	9.3%
o	497620	8.5%
n	451974	7.7%
a	448157	7.6%
t	427258	7.3%
i	363374	6.2%
s	349676	5.9%
h	334713	

Value	Count	Frequency (%)
		5.7%
l	310391	5.3%
Other values (13)	1434111	24.4%

Uppercase Letter

Value	Count	Frequency (%)
S	111924	12.5%
C	94198	10.5%
B	92463	10.3%
W	77540	8.6%
H	63582	7.1%
L	55205	6.1%
N	53498	6.0%
T	46830	5.2%
E	45212	5.0%
D	42155	4.7%
Other values (12)	215480	24.0%

Other Punctuation

Value	Count	Frequency (%)
,	18145	74.4%
.	4798	19.7%
'	1441	5.9%

Space Separator

Value	Count	Frequency (%)
	297879	100.0%

Dash Punctuation

Value	Count	Frequency (%)
-	20370	100.0%

Open Punctuation

Value	Count	Frequency (%)
(179	100.0%

Close Punctuation

Value	Count	Frequency (%)
)	179	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	6780973	95.2%
Common	342991	4.8%

Most frequent character per script

Latin

Value	Count	Frequency (%)
e	717756	10.6%
r	547856	8.1%
o	497620	7.3%
n	451974	6.7%
a	448157	6.6%
t	427258	6.3%
i	363374	5.4%
s	349676	5.2%
h	334713	4.9%
l	310391	4.6%
Other values (35)	2332198	34.4%

Common

Value	Count	Frequency (%)
	297879	86.8%
-	20370	5.9%
,	18145	5.3%
.	4798	1.4%
'	1441	0.4%

Value	Count	Frequency (%)
(179	0.1%
)	179	0.1%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	7123964	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
e	717756	10.1%
r	547856	7.7%
o	497620	7.0%
n	451974	6.3%
a	448157	6.3%
t	427258	6.0%
i	363374	5.1%
s	349676	4.9%
h	334713	4.7%
l	310391	4.4%
Other values (42)	2675189	37.6%

[Longitude](#)

Real number (ℝ)

Distinct	529766
Distinct (%)	80.2%
Missing	26
Missing (%)	< 0.1%
Infinite	0
Infinite (%)	0.0%
Mean	-1.4312104
Minimum	-7.516225

Maximum	1.76201
Zeros	0
Zeros (%)	0.0%
Negative	575313
Negative (%)	87.1%
Memory size	5.0 MiB

2025-10-05T19:30:08.298146image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
More details

- Statistics
- Histogram
- Common values
- Extreme values

Quantile statistics

Minimum	-7.516225
5-th percentile	-3.998785
Q1	-2.332291
median	-1.411667
Q3	-0.232869
95-th percentile	0.558791
Maximum	1.76201
Range	9.278235
Interquartile range (IQR)	2.099422

Descriptive statistics

Standard deviation	1.3833295
Coefficient of variation (CV)	-0.96654522
Kurtosis	-0.31056404
Mean	-1.4312104
Median Absolute Deviation (MAD)	1.08697
Skewness	-0.34731393
Sum	-945533.42
Variance	1.9136006
Monotonicity	Not monotonic

2025-10-05T19:30:08.739270image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
-0.977611	45	< 0.1%
-1.871043	36	< 0.1%
-0.816789	33	< 0.1%

Value	Count	Frequency (%)
-1.190861	31	< 0.1%
-1.216694	29	< 0.1%
-1.190402	28	< 0.1%
-1.999967	27	< 0.1%
-1.234393	27	< 0.1%
-3.143767	27	< 0.1%
-1.181848	26	< 0.1%
Other values (529756)	660344	99.9%

- Minimum 10 values
- Maximum 10 values

Value	Count	Frequency (%)
-7.516225	1	< 0.1%
-7.515933	1	< 0.1%
-7.504121	1	< 0.1%
-7.498973	1	< 0.1%
-7.497461	1	< 0.1%
-7.485126	1	< 0.1%
-7.465609	1	< 0.1%
-7.465103	1	< 0.1%
-7.460259	1	< 0.1%
-7.450015	1	< 0.1%

Value	Count	Frequency (%)
1.76201	1	< 0.1%
1.759398	1	< 0.1%
1.758345	1	< 0.1%
1.75819	1	< 0.1%
1.758106	1	< 0.1%
1.757915	1	< 0.1%
1.757907	1	< 0.1%
1.756307	1	< 0.1%
1.756285	1	< 0.1%
1.756049	1	< 0.1%

[Number_of_Casualties](#)

Real number (ℝ)

Distinct	36
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1.3570403
Minimum	1
Maximum	68
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	5.0 MiB

2025-10-05T19:30:09.148170image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
More details

- Statistics
- Histogram
- Common values
- Extreme values

Quantile statistics

Minimum	1
5-th percentile	1
Q1	1
median	1
Q3	1
95-th percentile	3
Maximum	68
Range	67
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	0.82484694
Coefficient of variation (CV)	0.60782791
Kurtosis	237.8686
Mean	1.3570403
Median Absolute Deviation (MAD)	0
Skewness	7.0813158
Sum	896568
Variance	0.68037248
Monotonicity	Not monotonic

2025-10-05T19:30:09.571678image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

Histogram with fixed size bins (bins=36)

Value	Count	Frequency (%)
1	504041	76.3%
2	107971	16.3%
3	30776	4.7%
4	11156	1.7%
5	4123	0.6%
6	1525	0.2%
7	539	0.1%
8	214	< 0.1%
9	112	< 0.1%
10	58	< 0.1%
Other values (26)	164	< 0.1%

- Minimum 10 values
- Maximum 10 values

Value	Count	Frequency (%)
1	504041	76.3%
2	107971	16.3%
3	30776	4.7%
4	11156	1.7%
5	4123	0.6%
6	1525	0.2%
7	539	0.1%
8	214	< 0.1%
9	112	< 0.1%
10	58	

Value	Count	Frequency (%)
		< 0.1%
Value	Count	Frequency (%)
68	1	< 0.1%
62	1	< 0.1%
48	1	< 0.1%
47	1	< 0.1%
45	1	< 0.1%
43	1	< 0.1%
42	2	< 0.1%
40	2	< 0.1%
32	1	< 0.1%
29	3	< 0.1%

Number_of_Vehicles

Real number (ℝ)

Distinct	19
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1.8312554
Minimum	1
Maximum	32
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	5.0 MiB

2025-10-05T19:30:09.965764image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
More details

- Statistics
- Histogram
- Common values
- Extreme values

Quantile statistics

Minimum	1
5-th percentile	1
Q1	1
median	2
Q3	2
95-th percentile	3
Maximum	32
Range	31
Interquartile range (IQR)	1

Descriptive statistics

Standard deviation	0.71526904
Coefficient of variation (CV)	0.39058945
Kurtosis	18.065456
Mean	1.8312554
Median Absolute Deviation (MAD)	0
Skewness	1.8026466
Sum	1209872
Variance	0.5116098
Monotonicity	Not monotonic

2025-10-05T19:30:10.356462image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

Histogram with fixed size bins (bins=19)

Value	Count	Frequency (%)
2	391995	59.3%
1	200787	30.4%
3	52806	8.0%
4	11300	1.7%
5	2464	0.4%
6	758	0.1%
7	296	< 0.1%
8	147	< 0.1%
9	57	< 0.1%
10	27	< 0.1%
Other values (9)	42	< 0.1%

- Minimum 10 values

- Maximum 10 values

Value	Count	Frequency (%)
1	200787	30.4%
2	391995	59.3%
3	52806	8.0%
4	11300	1.7%
5	2464	0.4%
6	758	0.1%
7	296	< 0.1%
8	147	< 0.1%
9	57	< 0.1%
10	27	< 0.1%

Value	Count	Frequency (%)
32	1	< 0.1%
28	1	< 0.1%
19	1	< 0.1%
16	4	< 0.1%
15	1	< 0.1%
14	9	< 0.1%
13	6	< 0.1%
12	7	< 0.1%
11	12	< 0.1%
10	27	< 0.1%

[Road_Surface_Conditions](#)

Categorical

Imbalance

Distinct	5
-----------------	---

Distinct (%)	< 0.1%
Missing	726
Missing (%)	0.1%
Memory size	5.0 MiB
Dry	447821
Wet or damp	186708
Frost or ice	18517
Snow	5890
Flood over 3cm. deep	1017

More details

- Overview
- Categories
- Words
- Characters

Length

Max length	20
Median length	3
Mean length	5.5509332
Min length	3

Characters and Unicode

Total characters	3663355
Distinct characters	23
Distinct categories	5 ?
Distinct scripts	2 ?
Distinct blocks	1 ?

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

Unique

Unique	0 ?
Unique (%)	0.0%

Sample

1st row	Dry
2nd row	Wet or damp
3rd row	Dry
4th row	Dry
5th row	Dry

Common Values

Value	Count	Frequency (%)
Dry	447821	67.8%
Wet or damp	186708	28.3%
Frost or ice	18517	2.8%
Snow	5890	0.9%
Flood over 3cm. deep	1017	0.2%
(Missing)	726	0.1%

Length

2025-10-05T19:30:10.793312image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
Histogram of lengths of the category
Common Values (Plot)

2025-10-05T19:30:11.183318image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

Value	Count	Frequency (%)
dry	447821	41.7%
or	205225	19.1%
wet	186708	17.4%
damp	186708	17.4%
frost	18517	1.7%
ice	18517	1.7%
snow	5890	0.5%
flood	1017	0.1%
over	1017	0.1%
3cm	1017	0.1%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
r	672580	18.4%
D	447821	12.2%

Value	Count	Frequency (%)
y	447821	12.2%
	413501	11.3%
o	232683	6.4%
e	208276	5.7%
t	205225	5.6%
d	188742	5.2%
m	187725	5.1%
p	187725	5.1%
Other values (13)	471256	12.9%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	2587867	70.6%
Uppercase Letter	659953	18.0%
Space Separator	413501	11.3%
Decimal Number	1017	< 0.1%
Other Punctuation	1017	< 0.1%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
r	672580	26.0%
y	447821	17.3%
o	232683	9.0%
e	208276	8.0%
t	205225	7.9%
d	188742	7.3%
m	187725	7.3%
p	187725	

Value	Count	Frequency (%)
		7.3%
a	186708	7.2%
c	19534	0.8%
Other values (6)	50848	2.0%

Uppercase Letter

Value	Count	Frequency (%)
D	447821	67.9%
W	186708	28.3%
F	19534	3.0%
S	5890	0.9%

Space Separator

Value	Count	Frequency (%)
	413501	100.0%

Decimal Number

Value	Count	Frequency (%)
3	1017	100.0%

Other Punctuation

Value	Count	Frequency (%)
.	1017	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	3247820	88.7%
Common	415535	11.3%

Most frequent character per script

Latin

Value	Count	Frequency (%)
r	672580	20.7%
D	447821	13.8%
y	447821	13.8%
o	232683	7.2%
e	208276	6.4%

Value	Count	Frequency (%)
t	205225	6.3%
d	188742	5.8%
m	187725	5.8%
p	187725	5.8%
a	186708	5.7%
Other values (10)	282514	8.7%

Common

Value	Count	Frequency (%)
	413501	99.5%
3	1017	0.2%
.	1017	0.2%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	3663355	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
r	672580	18.4%
D	447821	12.2%
y	447821	12.2%
	413501	11.3%
o	232683	6.4%
e	208276	5.7%
t	205225	5.6%
d	188742	5.2%
m	187725	5.1%
p	187725	5.1%
Other values (13)	471256	12.9%

Road_Type

Categorical

Distinct	5
Distinct (%)	< 0.1%
Missing	4520
Missing (%)	0.7%
Memory size	5.0 MiB
Single carriageway	492143
Dual carriageway	99424
Roundabout	43992
One way street	13559
Slip road	7041

More details

- Overview
- Categories
- Words
- Characters

Length

Max length	18
Median length	18
Mean length	16.981361
Min length	9

Characters and Unicode

Total characters	11142473
Distinct characters	22
Distinct categories	3 ?
Distinct scripts	2 ?
Distinct blocks	1 ?

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

Unique

Unique	0 ?
Unique (%)	0.0%

Sample

1st row	Single carriageway
2nd row	Single carriageway

3rd row	Single carriageway
4th row	Single carriageway
5th row	Dual carriageway

Common Values

Value	Count	Frequency (%)
Single carriageway	492143	74.5%
Dual carriageway	99424	15.0%
Roundabout	43992	6.7%
One way street	13559	2.1%
Slip road	7041	1.1%
(Missing)	4520	0.7%

Length

2025-10-05T19:30:11.606350image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
Histogram of lengths of the category

Common Values (Plot)

2025-10-05T19:30:11.995160image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

Value	Count	Frequency (%)
carriageway	591567	46.1%
single	492143	38.4%
dual	99424	7.8%
roundabout	43992	3.4%
one	13559	1.1%
way	13559	1.1%
street	13559	1.1%
slip	7041	0.5%
road	7041	0.5%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
a	1938717	17.4%
r	1203734	10.8%
e	1124387	10.1%
i	1090751	9.8%
g	1083710	9.7%
	625726	5.6%
y	605126	5.4%
w	605126	5.4%
l	598608	5.4%
c	591567	5.3%
Other values (12)	1675021	15.0%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	9860588	88.5%
Uppercase Letter	656159	5.9%
Space Separator	625726	5.6%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
a	1938717	19.7%
r	1203734	12.2%
e	1124387	11.4%
i	1090751	11.1%
g	1083710	11.0%
y	605126	6.1%
w	605126	6.1%
l	598608	6.1%
c	591567	6.0%
n	549694	

Value	Count	Frequency (%)
		5.6%
Other values (7)	469168	4.8%

Uppercase Letter

Value	Count	Frequency (%)
S	499184	76.1%
D	99424	15.2%
R	43992	6.7%
O	13559	2.1%

Space Separator

Value	Count	Frequency (%)
	625726	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	10516747	94.4%
Common	625726	5.6%

Most frequent character per script

Latin

Value	Count	Frequency (%)
a	1938717	18.4%
r	1203734	11.4%
e	1124387	10.7%
i	1090751	10.4%
g	1083710	10.3%
y	605126	5.8%
w	605126	5.8%
l	598608	5.7%
c	591567	5.6%
n	549694	5.2%
Other values (11)	1125327	10.7%

Common

Value	Count	Frequency (%)
	625726	100.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	11142473	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
a	1938717	17.4%
r	1203734	10.8%
e	1124387	10.1%
i	1090751	9.8%
g	1083710	9.7%
	625726	5.6%
y	605126	5.4%
w	605126	5.4%
l	598608	5.4%
c	591567	5.3%
Other values (12)	1675021	15.0%

[Urban_or_Rural_Area](#)

Categorical

Distinct		3
Distinct (%)		< 0.1%
Missing		15
Missing (%)		< 0.1%
Memory size		5.0 MiB
Urban	421663	
Rural	238990	
Unallocated	11	

More details

- Overview
- Categories
- Words
- Characters

Length

Max length	11
Median length	5
Mean length	5.0000999
Min length	5

Characters and Unicode

Total characters	3303386
Distinct characters	13
Distinct categories	2 ?
Distinct scripts	1 ?
Distinct blocks	1 ?

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

Unique

Unique	0 ?
Unique (%)	0.0%

Sample

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

Common Values

Value	Count	Frequency (%)
Urban	421663	63.8%
Rural	238990	36.2%
Unallocated	11	< 0.1%
(Missing)	15	< 0.1%

Length

2025-10-05T19:30:12.445764image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
Histogram of lengths of the category
Common Values (Plot)

2025-10-05T19:30:12.808359image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

Value	Count	Frequency (%)
urban	421663	63.8%
rural	238990	36.2%
unallocated	11	

Value	Count	Frequency (%)
		< 0.1%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
a	660675	20.0%
r	660653	20.0%
U	421674	12.8%
n	421674	12.8%
b	421663	12.8%
l	239012	7.2%
R	238990	7.2%
u	238990	7.2%
o	11	< 0.1%
c	11	< 0.1%
Other values (3)	33	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	2642722	80.0%
Uppercase Letter	660664	20.0%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
a	660675	25.0%
r	660653	25.0%
n	421674	16.0%
b	421663	16.0%
l	239012	9.0%
u	238990	

Value	Count	Frequency (%)
		9.0%
o	11	< 0.1%
c	11	< 0.1%
t	11	< 0.1%
e	11	< 0.1%

Uppercase Letter

Value	Count	Frequency (%)
U	421674	63.8%
R	238990	36.2%

Most occurring scripts

Value	Count	Frequency (%)
Latin	3303386	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
a	660675	20.0%
r	660653	20.0%
U	421674	12.8%
n	421674	12.8%
b	421663	12.8%
l	239012	7.2%
R	238990	7.2%
u	238990	7.2%
o	11	< 0.1%
c	11	< 0.1%
Other values (3)	33	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	3303386	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
a	660675	20.0%
r	660653	20.0%
U	421674	12.8%
n	421674	12.8%
b	421663	12.8%
l	239012	7.2%
R	238990	7.2%
u	238990	7.2%
o	11	< 0.1%
c	11	< 0.1%
Other values (3)	33	< 0.1%

[Weather_Conditions](#)

Categorical

Imbalance Missing

Distinct	8
Distinct (%)	< 0.1%
Missing	14128
Missing (%)	2.1%
Memory size	5.0 MiB
Fine no high winds	520885
Raining no high winds	79696
Other	17150
Raining + high winds	9615
Fine + high winds	8554
Other values (3)	10651

More details

- Overview
- Categories
- Words
- Characters

Length

Max length	21
Median length	18
Mean length	18.034958
Min length	5

Characters and Unicode

Total characters	11660520
Distinct characters	19
Distinct categories	4 ?
Distinct scripts	2 ?
Distinct blocks	1 ?

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

Unique

Unique	0 ?
Unique (%)	0.0%

Sample

1st row	Fine no high winds
2nd row	Raining no high winds
3rd row	Fine no high winds
4th row	Fine no high winds
5th row	Fine no high winds

Common Values

Value	Count	Frequency (%)
Fine no high winds	520885	78.8%
Raining no high winds	79696	12.1%
Other	17150	2.6%
Raining + high winds	9615	1.5%
Fine + high winds	8554	1.3%
Snowing no high winds	6238	0.9%
Fog or mist	3528	0.5%
Snowing + high winds	885	0.1%
(Missing)	14128	2.1%

Length

2025-10-05T19:30:13.232762image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
Histogram of lengths of the category
Common Values (Plot)

2025-10-05T19:30:13.680885image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

Value	Count	Frequency (%)
high	625873	24.7%
winds	625873	24.7%
no	606819	24.0%
fine	529439	20.9%
raining	89311	3.5%
	19054	0.8%
other	17150	0.7%
snowing	7123	0.3%
fog	3528	0.1%
or	3528	0.1%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
i	1970458	16.9%
n	1954999	16.8%
	1884675	16.2%
h	1268896	10.9%
g	725835	6.2%
w	632996	5.4%
s	629401	5.4%
d	625873	5.4%
o	620998	

Value	Count	Frequency (%)
		5.3%
e	546589	4.7%
Other values (9)	799800	6.9%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	9110240	78.1%
Space Separator	1884675	16.2%
Uppercase Letter	646551	5.5%
Math Symbol	19054	0.2%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
i	1970458	21.6%
n	1954999	21.5%
h	1268896	13.9%
g	725835	8.0%
w	632996	6.9%
s	629401	6.9%
d	625873	6.9%
o	620998	6.8%
e	546589	6.0%
a	89311	1.0%
Other values (3)	44884	0.5%

Uppercase Letter

Value	Count	Frequency (%)
F	532967	82.4%
R	89311	13.8%
O	17150	

Value	Count	Frequency (%)
		2.7%
S	7123	1.1%

Space Separator

Value	Count	Frequency (%)
	1884675	100.0%

Math Symbol

Value	Count	Frequency (%)
+	19054	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	9756791	83.7%
Common	1903729	16.3%

Most frequent character per script

Latin

Value	Count	Frequency (%)
i	1970458	20.2%
n	1954999	20.0%
h	1268896	13.0%
g	725835	7.4%
w	632996	6.5%
s	629401	6.5%
d	625873	6.4%
o	620998	6.4%
e	546589	5.6%
F	532967	5.5%
Other values (7)	247779	2.5%

Common

Value	Count	Frequency (%)
	1884675	99.0%
+	19054	

Value	Count	Frequency (%)
		1.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	11660520	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
i	1970458	16.9%
n	1954999	16.8%
	1884675	16.2%
h	1268896	10.9%
g	725835	6.2%
w	632996	5.4%
s	629401	5.4%
d	625873	5.4%
o	620998	5.3%
e	546589	4.7%
Other values (9)	799800	6.9%

[Vehicle_Type](#)

Categorical

Imbalance

Distinct	16
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	5.0 MiB
Car	497992
Van / Goods 3.5 tonnes mgw or under	34160
Bus or coach (17 or more pass seats)	25878
Motorcycle over 500cc	25657
Goods 7.5 tonnes mgw and over	

	17307
Other values (11)	59685

More details

- Overview
- Categories
- Words
- Characters

Length

Max length	37
Median length	3
Mean length	9.3513052
Min length	3

Characters and Unicode

Total characters	6178211
Distinct characters	47
Distinct categories	8 ?
Distinct scripts	2 ?
Distinct blocks	1 ?

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

Unique

Unique	0 ?
Unique (%)	0.0%

Sample

1st row	Car
2nd row	Car
3rd row	Taxi/Private hire car
4th row	Bus or coach (17 or more pass seats)
5th row	Other vehicle

Common Values

Value	Count	Frequency (%)
Car	497992	75.4%
Van / Goods 3.5 tonnes mgw or under	34160	5.2%
Bus or coach (17 or more pass seats)	25878	3.9%
Motorcycle over 500cc	25657	3.9%
Goods 7.5 tonnes mgw and over	17307	

Value	Count	Frequency (%)
		2.6%
Motorcycle 125cc and under	15269	2.3%
Taxi/Private hire car	13294	2.0%
Motorcycle over 125cc and up to 500cc	7656	1.2%
Motorcycle 50cc and under	7603	1.2%
Goods over 3.5t. and under 7.5t	6096	0.9%
Other values (6)	9767	1.5%

Length

2025-10-05T19:30:14.222253image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
Histogram of lengths of the category

Value	Count	Frequency (%)
car	511286	36.3%
or	85922	6.1%
under	63128	4.5%
goods	57563	4.1%
over	56716	4.0%
motorcycle	56185	4.0%
and	53931	3.8%
tonnes	51467	3.7%
mgw	51467	3.7%
	36136	2.6%
Other values (34)	384308	27.3%

- Characters
- Categories
- Scripts
- Blocks

Most occurring characters

Value	Count	Frequency (%)
r	837220	13.6%
	747430	12.1%
a	709713	11.5%
C	497992	8.1%
o	481029	7.8%
e	332981	5.4%
c	315027	5.1%
n	258121	4.2%
s	248316	4.0%
t	176244	2.9%
Other values (37)	1574138	25.5%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	4178833	67.6%
Space Separator	747430	12.1%
Uppercase Letter	708133	11.5%
Decimal Number	368922	6.0%
Other Punctuation	117209	1.9%
Open Punctuation	27854	0.5%
Close Punctuation	27854	0.5%
Dash Punctuation	1976	< 0.1%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
r	837220	20.0%
a	709713	17.0%
o	481029	11.5%
e	332981	

Value	Count	Frequency (%)
		8.0%
c	315027	7.5%
n	258121	6.2%
s	248316	5.9%
t	176244	4.2%
d	174827	4.2%
u	102538	2.5%
Other values (12)	542817	13.0%

Uppercase Letter

Value	Count	Frequency (%)
C	497992	70.3%
M	58161	8.2%
G	57563	8.1%
V	34160	4.8%
B	25878	3.7%
P	13491	1.9%
T	13294	1.9%
O	5637	0.8%
A	1947	0.3%
D	6	< 0.1%

Decimal Number

Value	Count	Frequency (%)
5	127500	34.6%
0	74229	20.1%
1	50779	13.8%
7	49281	13.4%

Value	Count	Frequency (%)
3	40256	10.9%
2	22925	6.2%
8	1976	0.5%
6	1976	0.5%

Other Punctuation

Value	Count	Frequency (%)
.	69755	59.5%
/	47454	40.5%

Space Separator

Value	Count	Frequency (%)
	747430	100.0%

Open Punctuation

Value	Count	Frequency (%)
(27854	100.0%

Close Punctuation

Value	Count	Frequency (%)
)	27854	100.0%

Dash Punctuation

Value	Count	Frequency (%)
-	1976	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	4886966	79.1%
Common	1291245	20.9%

Most frequent character per script

Latin

Value	Count	Frequency (%)
r	837220	17.1%
a	709713	14.5%
C	497992	10.2%
o	481029	9.8%
e	332981	6.8%

Value	Count	Frequency (%)
c	315027	6.4%
n	258121	5.3%
s	248316	5.1%
t	176244	3.6%
d	174827	3.6%
Other values (23)	855496	17.5%

Common

Value	Count	Frequency (%)
	747430	57.9%
5	127500	9.9%
0	74229	5.7%
.	69755	5.4%
1	50779	3.9%
7	49281	3.8%
/	47454	3.7%
3	40256	3.1%
(27854	2.2%
)	27854	2.2%
Other values (4)	28853	2.2%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	6178211	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
r	837220	13.6%

Value	Count	Frequency (%)
	747430	12.1%
a	709713	11.5%
C	497992	8.1%
o	481029	7.8%
e	332981	5.4%
c	315027	5.1%
n	258121	4.2%
s	248316	4.0%
t	176244	2.9%
Other values (37)	1574138	25.5%

Interactions

- Latitude
- Longitude
- Number_of_Casualties
- Number_of_Vehicles

- Number_of_Vehicles
- Latitude
- Longitude
- Number_of_Casualties

2025-10-05T19:29:48.409169image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
2025-10-05T19:29:42.756300image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
2025-10-05T19:29:44.592194image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
2025-10-05T19:29:46.538543image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

- Number_of_Vehicles
- Latitude
- Longitude
- Number_of_Casualties

2025-10-05T19:29:48.885915image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
2025-10-05T19:29:43.195988image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
2025-10-05T19:29:45.030280image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
2025-10-05T19:29:47.016640image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

- Number_of_Vehicles
- Latitude

- Longitude
- Number_of_Casualties

2025-10-05T19:29:49.326136image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
 2025-10-05T19:29:43.638254image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
 2025-10-05T19:29:45.469402image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
 2025-10-05T19:29:47.446465image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

- Number_of_Vehicles
- Latitude
- Longitude
- Number_of_Casualties

2025-10-05T19:29:49.788962image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
 2025-10-05T19:29:44.126083image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
 2025-10-05T19:29:46.059631image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
 2025-10-05T19:29:47.929863image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

Correlations

- Auto
- Heatmap
- Table

2025-10-05T19:30:14.545081image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

	Accident_Severity	Latitude	Light_Conditions	Longitude	Number_of_Casualties
Accident_Severity	1.000	0.028	0.060	0.010	0.025
Latitude	0.028	1.000	0.045	-0.388	0.047
Light_Conditions	0.060	0.045	1.000	0.037	0.002
Longitude	0.010	-0.388	0.037	1.000	-0.048
Number_of_Casualties	0.025	0.047	0.002	-0.048	1.000
Number_of_Vehicles	0.016	-0.031	0.007	0.014	0.226
Road_Surface_Conditions	0.016	0.048	0.111	0.040	0.003
Road_Type	0.042	0.026	0.040	0.029	0.008
Urban_or_Rural_Area	0.061	0.103	0.234	0.095	0.009
Vehicle_Type	0.001	0.016	0.001	0.012	0.000
Weather_Conditions	0.022	0.035	0.081	0.030	0.003

Missing values

- Count
- Matrix
- Heatmap

2025-10-05T19:29:50.852447image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>
 A simple visualization of nullity by column.
 2025-10-05T19:29:53.152779image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

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

2025-10-05T19:29:56.569526image/svg+xmlMatplotlib v3.8.4, <https://matplotlib.org/>

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

Sample

- First rows
- Last rows

	Index	Accident_Severity	Accident Date	Latitude	Light_Conditions	Dis
0	200701BS64157	Serious	05-06-2019	51.506187	Darkness - lights lit	Kensing
1	200701BS65737	Serious	02-07-2019	51.495029	Daylight	Kensing
2	200701BS66127	Serious	26-08-2019	51.517715	Darkness - lighting unknown	Kensing
3	200701BS66128	Serious	16-08-2019	51.495478	Daylight	Kensing
4	200701BS66837	Slight	03-09-2019	51.488576	Darkness - lights lit	Kensing
5	200701BS67159	Serious	18-09-2019	51.497750	Daylight	Kensing
6	200701BS67207	Serious	05-09-2019	51.501405	Daylight	Kensing
7	200701BS67370	Fatal	03-10-2019	51.482260	Darkness - lights lit	Kensing
8	200701BS67515	Slight	31-10-2019	51.493319	Darkness - lights lit	Kensing
9	200701BS67543	Slight	18-10-2019	51.484539	Daylight	Kensing
	Index	Accident_Severity	Accident Date	Latitude	Light_Conditions	District
660669	201091NM00903	Slight	24-01-2022	57.484296	Darkness - no lighting	Highlan
660670	201091NM00939	Slight	25-01-2022	57.574536	Darkness - lights lit	Highlan
660671	201091NM01053	Slight	29-01-2022	57.572162	Daylight	Highlan
660672	201091NM01085	Serious	29-01-2022	57.170725	Daylight	Highlan
660673	201091NM01190	Slight	02-02-2022	57.585152	Daylight	Highlan
660674	201091NM01760	Slight	18-02-2022	57.374005	Daylight	Highlan
660675	201091NM01881	Slight	21-02-2022	57.232273	Darkness - no lighting	Highlan
660676	201091NM01935	Slight	23-02-2022	57.585044	Daylight	Highlan
660677	201091NM01964	Serious	23-02-2022	57.214898	Darkness - no lighting	Highlan
660678	201091NM02142	Serious	28-02-2022	57.575210	Daylight	Highlan

Duplicate rows

Most frequently occurring

	Index	Accident_Severity	Accident Date	Latitude	Light_Conditions	District Area	Longitude	Number
0	2.01E+12	Slight	03-05-2020	51.015531	Daylight	Winchester	-1.341631	1
1	2.01E+12	Slight	04-03-2020	53.742266	Daylight	Kirklees	-1.675196	1
2	2.01E+12	Slight	05-04-2020	55.988302	Daylight	East Lothian	-2.527800	1

	Index	Accident_Severity	Accident Date	Latitude	Light_Conditions	District Area	Longitude	Number
3	2.01E+12	Slight	05-10-2019	51.360055	Daylight	Thanet	1.407403	1
4	2.01E+12	Slight	06-01-2021	53.648498	Daylight	Calderdale	-1.948984	1
5	2.01E+12	Slight	07-12-2020	50.870870	Darkness - no lighting	Brighton and Hove	-0.125952	1
6	2.01E+12	Slight	08-07-2022	53.009621	Daylight	East Staffordshire	-1.766123	1
7	2.01E+12	Slight	11-10-2020	50.868968	Daylight	Brighton and Hove	-0.147632	1
8	2.01E+12	Slight	12-02-2021	51.067126	Darkness - no lighting	Sedgemoor	-3.023829	1
9	2.01E+12	Slight	18-10-2019	53.970341	Daylight	Harrogate	-1.405573	1

Report generated by [YData](#).