

Overview

Overview

Warnings14

Reproduction

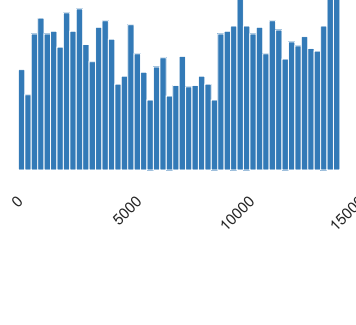
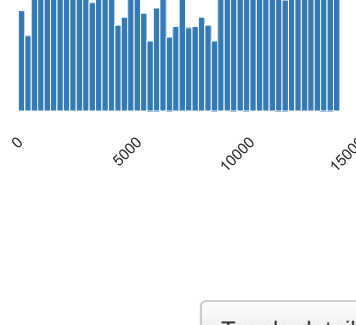


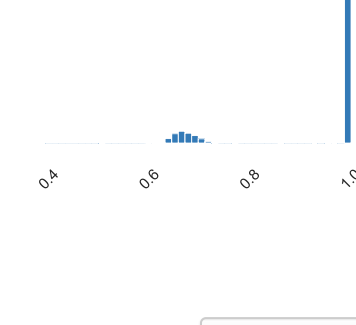
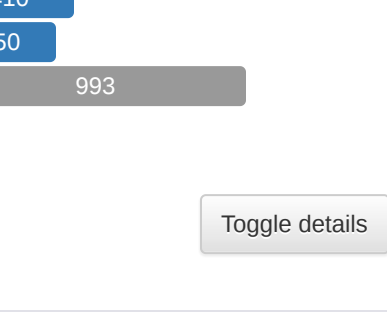
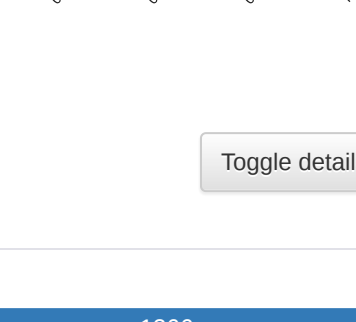
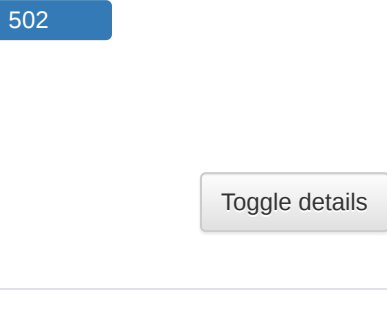
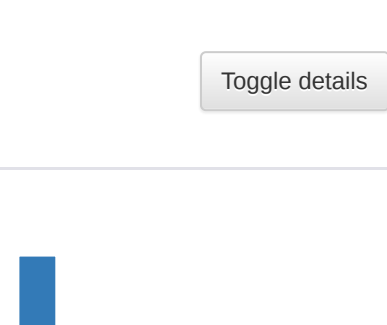
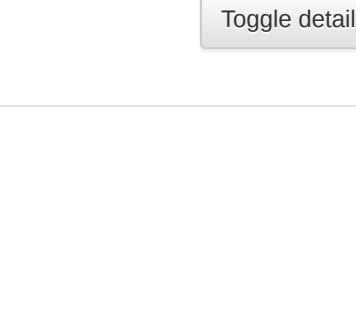
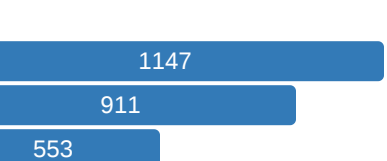
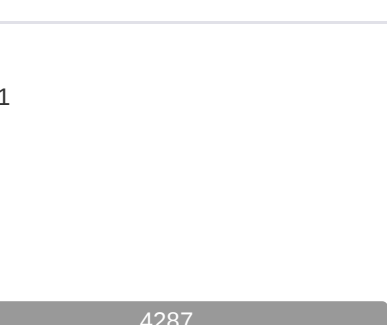
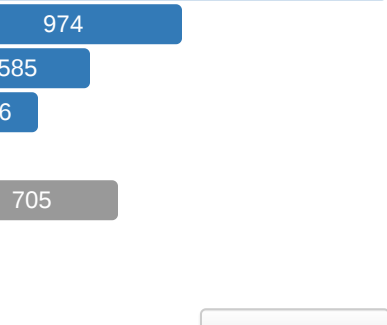
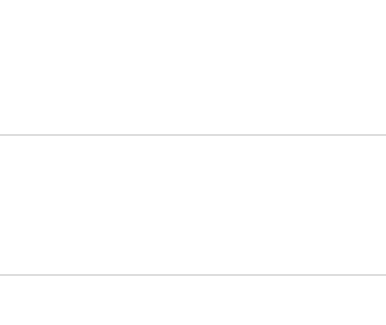
Dataset statistics

Number of variables	14
Number of observations	4674
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	3.2 MiB
Average record size in memory	713.3 B

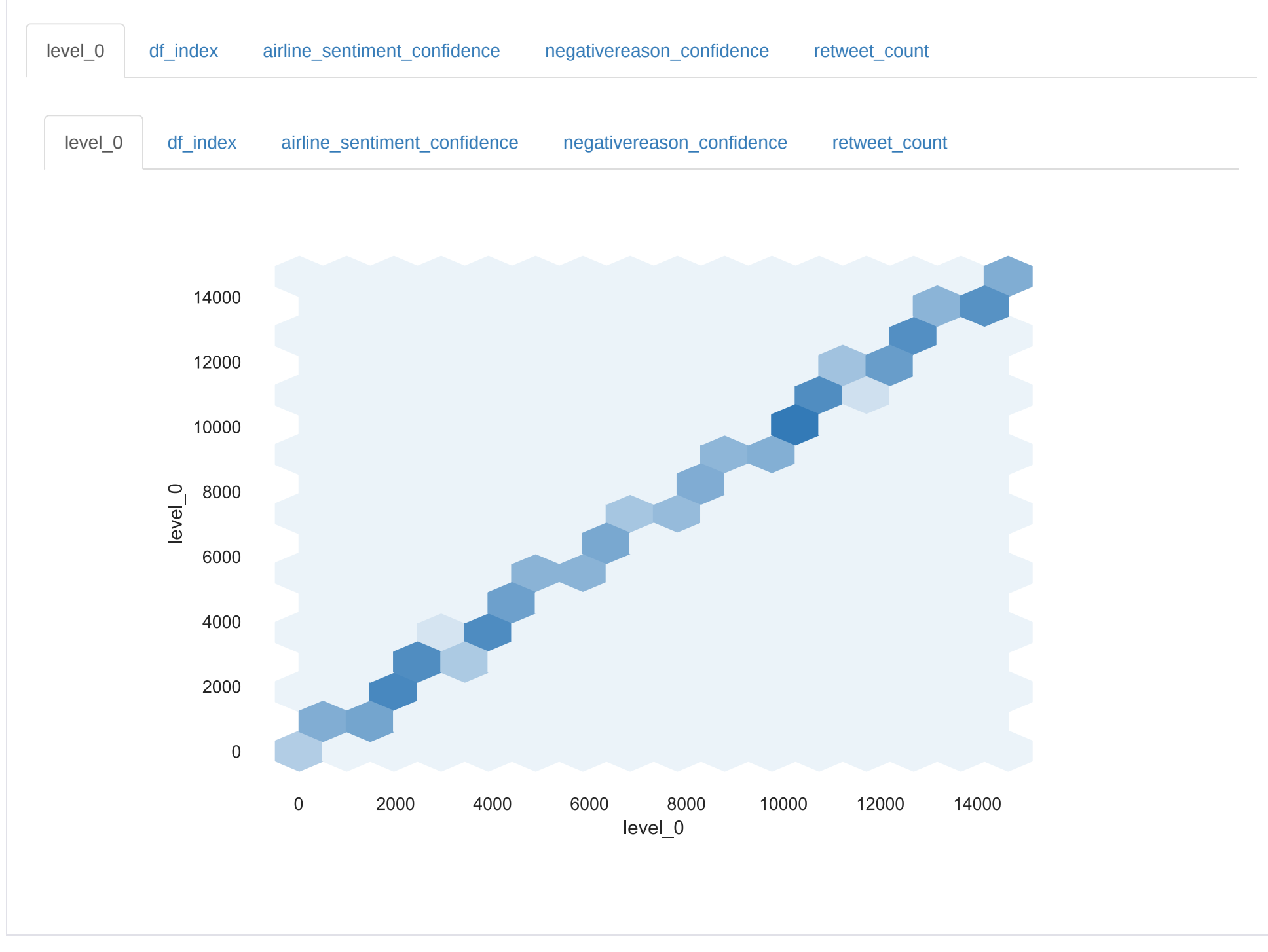
Variable types

CAT	9
NUM	5

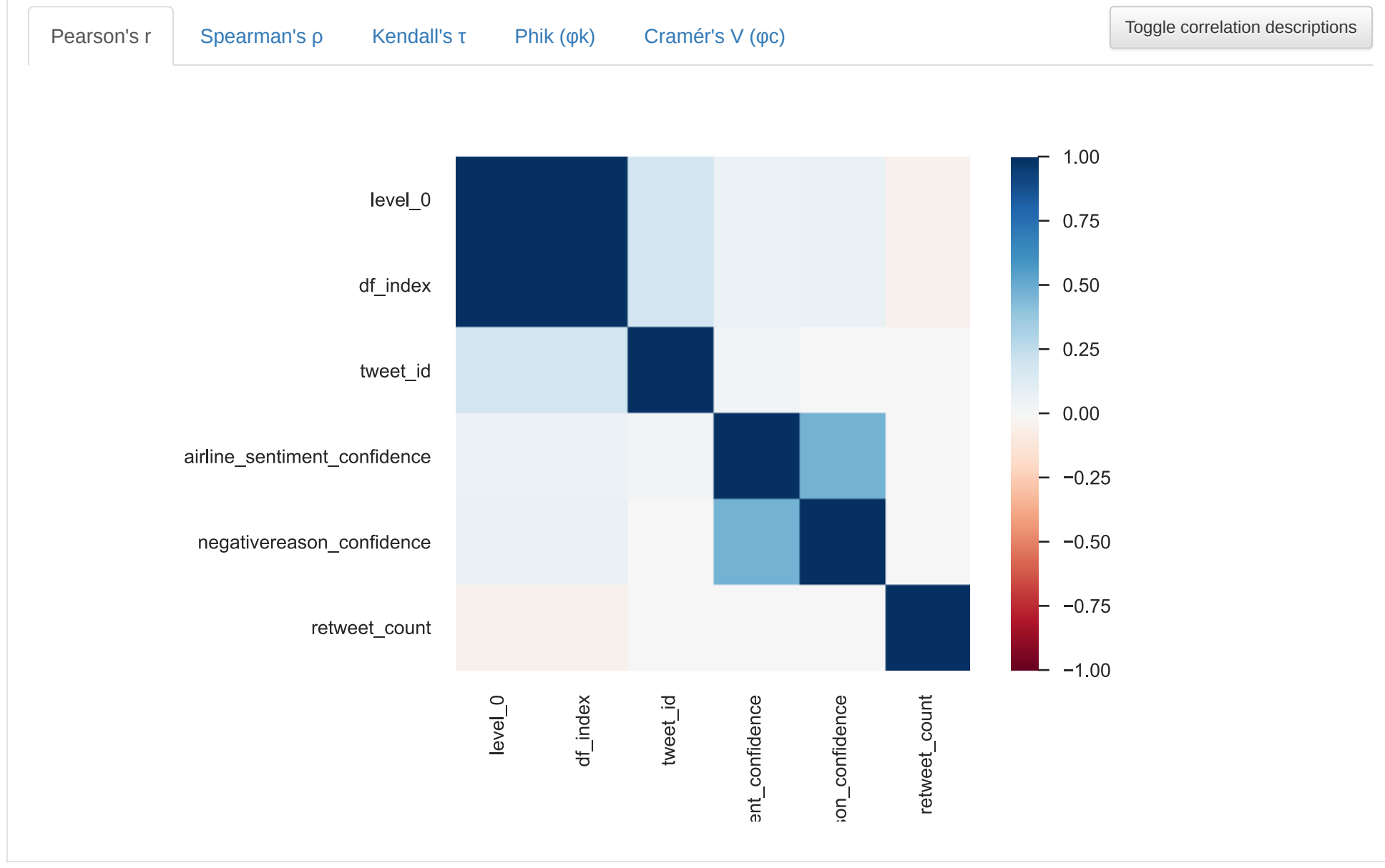
Variables

level_0 Real number (R ₃₂) HIGH CORRELATION UNIQUE	Distinct 4674 Distinct (%) 100.0% Missing 0 Missing (%) 0.0% Infinite 0 Infinite (%) 0.0%	Mean 7525.45721 Minimum 15 Maximum 14638 Zeros 0 Zeros (%) 0.0% Memory size 36.6 KiB		Toggle details
df_index Real number (R ₃₂) HIGH CORRELATION UNIQUE	Distinct 4674 Distinct (%) 100.0% Missing 0 Missing (%) 0.0% Infinite 0 Infinite (%) 0.0%	Mean 7525.45721 Minimum 15 Maximum 14638 Zeros 0 Zeros (%) 0.0% Memory size 36.6 KiB		Toggle details
tweet_id Categorical HIGH CORRELATION	Distinct 3 Distinct (%) 0.1% Missing 0 Missing (%) 0.0% Memory size 36.6 KiB	5.7e+17 2361 5.69e+17 1311 5.68e+17 1002		Toggle details
airline_sentiment Categorical CONSTANT REJECTED	Distinct 1 Distinct (%) < 0.1% Missing 0 Missing (%) 0.0% Memory size 36.6 KiB	negative 4674		Toggle details
airline_sentiment_confidence Real number (R ₃₂)	Distinct 524 Distinct (%) 11.2% Missing 0 Missing (%) 0.0% Infinite 0 Infinite (%) 0.0%	Mean 0.930673727 Minimum 0.3394 Maximum 1 Zeros 0 Zeros (%) 0.0% Memory size 36.6 KiB		Toggle details
negative_reason Categorical	Distinct 10 Distinct (%) 0.2% Missing 0 Missing (%) 0.0% Memory size 36.6 KiB	Customer Service Issue 1476 Late Flight 836 Can't Tell 609 Cancelled Flight 410 Lost Luggage 350 Other values (5) 993		Toggle details
negative_reason_confidence Real number (R ₃₂)	Distinct 1135 Distinct (%) 24.3% Missing 0 Missing (%) 0.0% Infinite 0 Infinite (%) 0.0%	Mean 0.7285897732 Minimum 0.3122 Maximum 1 Zeros 0 Zeros (%) 0.0% Memory size 36.6 KiB		Toggle details
airline Categorical	Distinct 6 Distinct (%) 0.1% Missing 0 Missing (%) 0.0% Memory size 36.6 KiB	United 1366 US Airways 1085 American 986 Southwest 647 Delta 502		Toggle details
name Categorical HIGH CARDINALITY	Distinct 2620 Distinct (%) 56.1% Missing 0 Missing (%) 0.0% Memory size 36.6 KiB	otsiday 28 rossj987 23 chagaga2013 18 scoobydoo9749 18 jasemccarty 15 Other values (2615) 4572		Toggle details
retweet_count Real number (R ₃₂) SKewed ZEROS	Distinct 9 Distinct (%) 0.2% Missing 0 Missing (%) 0.0% Infinite 0 Infinite (%) 0.0%	Mean 0.08857509628 Minimum 0 Maximum 22 Zeros 4366 Zeros (%) 93.4% Memory size 36.6 KiB		Toggle details
text Categorical HIGH CARDINALITY UNIFORM	Distinct 4624 Distinct (%) 98.9% Missing 0 Missing (%) 0.0% Memory size 36.6 KiB	@AmericanAir why am I contin... 2 @AmericanAir robocalls me wit... 2 @AmericanAir I'm frustrated by ... 2 @AmericanAir they will just say ... 2 @AmericanAir extremely upset ... 2 Other values (4619) 4664		Toggle details
tweet_created Categorical HIGH CORRELATION	Distinct 9 Distinct (%) 0.2% Missing 0 Missing (%) 0.0% Memory size 36.6 KiB	2015-02-22 1147 2015-02-23 911 2015-02-21 553 2015-02-20 470 2015-02-17 438 Other values (4) 1155		Toggle details
tweet_location Categorical HIGH CARDINALITY	Distinct 1588 Distinct (%) 34.0% Missing 0 Missing (%) 0.0% Memory size 36.6 KiB	Washington, DC 101 Boston, MA 85 New York, NY 80 NYC 63 New York City 58 Other values (1583) 4287		Toggle details
timezone Categorical HIGH CARDINALITY	Distinct 58 Distinct (%) 1.2% Missing 0 Missing (%) 0.0% Memory size 36.6 KiB	Eastern Time (US & Canada) 1852 Central Time (US & Canada) 974 Pacific Time (US & Canada) 595 Quito 366 Mountain Time (US & Canada) 152 Other values (53) 705		Toggle details

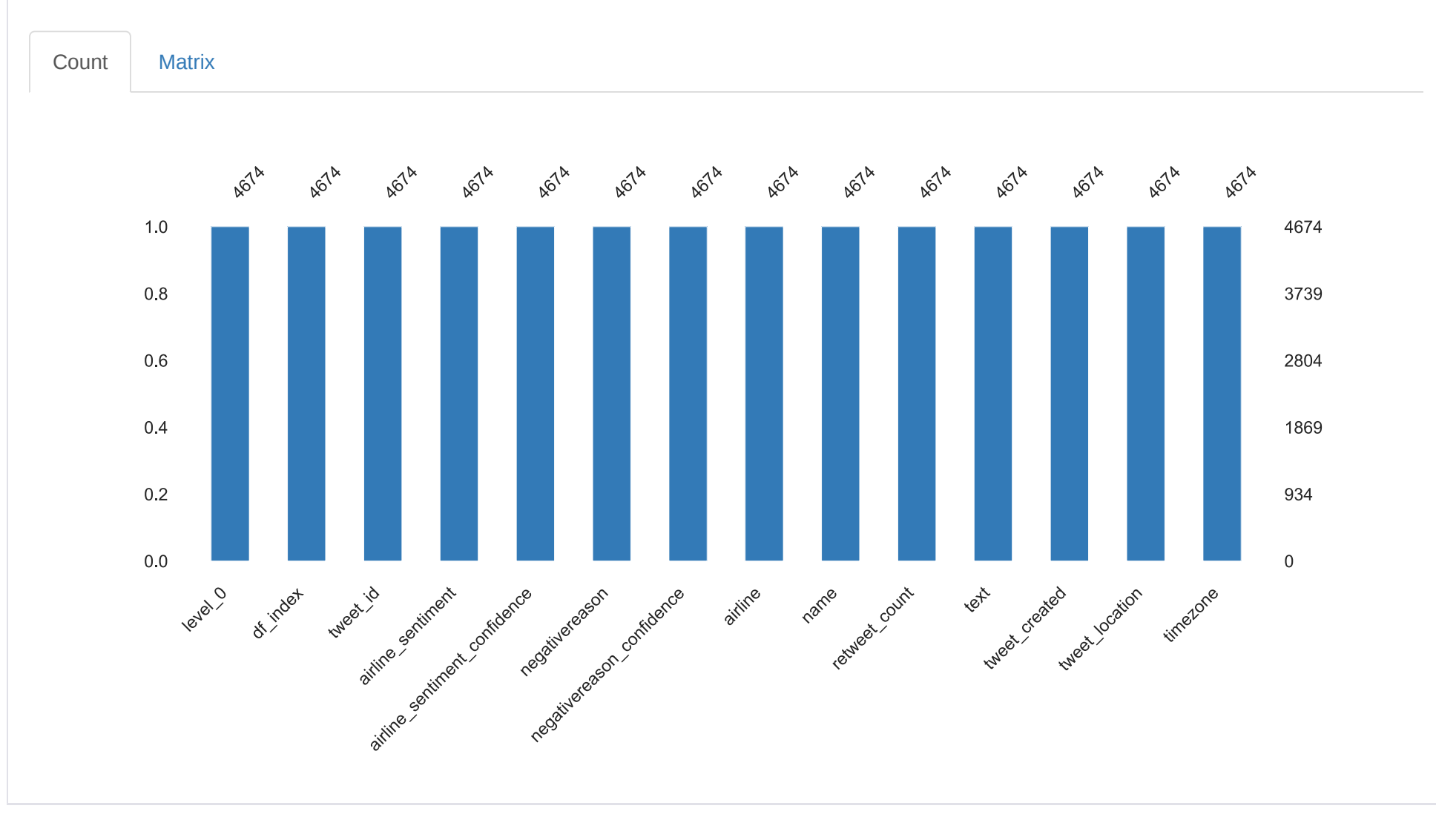
Interactions



Correlations



Missing values



Sample

First rows						
level_0	df_index	tweet_id	airline_sentiment	airline_sentiment_confidence	negativereason	negativereason_confidence
0	15	15	5.700000e+17	negative	0.6842	Late Flight 0.3684
1	17	17	5.700000e+17	negative	1.0000	Bad Flight 1.0000
2	24	24	5.700000e+17	negative	1.0000	Customer Service Issue 0.3557
3	32	32	5.700000e+17	negative	1.0000	Customer Service Issue 1.0000
4	39	39	5.700000e+17	negative	1.0000	Customer Service Issue 1.0000
5	41	41	5.700000e+17	negative	0.6688	Flight Booking Problems 0.6688
6	55	55	5.700000e+17	negative	0.6939	Flight Booking Problems 0.6939
7	61	61	5.700000e+17	negative	1.0000	Customer Service Issue 1.0000
8	66	66	5.700000e+17	negative	1.0000	Customer Service Issue 1.0000
9	67	67	5.700000e+17	negative	1.0000	Late Flight 0.6789

Last rows						
level_0	df_index	tweet_id	airline_sentiment	airline_sentiment_confidence	negativereason	negativereason_confidence
4664	14613	14613	5.700000e+17	negative	1.0	Customer Service Issue 0.7255
4665	14615	14615	5.700000e+17	negative	1.0	Cancelled Flight 1.0000
4666	14616	14616	5.700000e+17	negative	1.0	Cancelled Flight 1.0000
4667	14620	14620	5.700000e+17	negative	1.0	Customer Service Issue 1.0000
4668	14621	14621	5.700000e+17	negative	1.0	Customer Service Issue 1.0000
4669	14626	14626	5.700000e+17	negative	1.0	Late Flight 1.0000
4670	14629	14629	5.700000e+17	negative	1.0	Customer Service Issue 1.0000
4671	14631	14631	5.700000e+17	negative	1.0	Bad Flight 1.0000
4672	14633	14633	5.700000e+17	negative	1.0	Cancelled Flight 1.0000
4673	14638	14638	5.700000e+17	negative	1.0	Customer Service Issue 0.6659