

Overview

Dataset statistics	
Number of variables	17
Number of observations	9301
Missing cells	18497
Missing cells (%)	11.7%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	1.1 MiB
Average record size in memory	129.0 B

Variable types	
Unsupported	5
Numeric	6
Categorical	3
Boolean	1
DateTime	2

Alerts

content has a high cardinality: 7541 distinct values	High cardinality
user_displayname has a high cardinality: 6461 distinct values	High cardinality
url has a high cardinality: 9301 distinct values	High cardinality
likeCount is highly correlated with retweetCount	High correlation
retweetCount is highly correlated with likeCount	High correlation
likeCount is highly correlated with retweetCount	High correlation
retweetCount is highly correlated with likeCount and 1 other fields (likeCount, replyCount)	High correlation
replyCount is highly correlated with retweetCount	High correlation
likeCount is highly correlated with retweetCount	High correlation
retweetCount is highly correlated with likeCount	High correlation
coordinates_longitude is highly correlated with user_verified	High correlation
user_verified is highly correlated with coordinates_longitude	High correlation
likeCount is highly correlated with retweetCount and 1 other fields (retweetCount, replyCount)	High correlation
retweetCount is highly correlated with likeCount and 1 other fields (likeCount, replyCount)	High correlation
replyCount is highly correlated with likeCount and 1 other fields (likeCount, retweetCount)	High correlation
coordinates_longitude has 9196 (98.9%) missing values	Missing
retweetedTweet has 9301 (100.0%) missing values	Missing
user_followersCount is highly skewed ($\gamma_1 = 35.38748941$)	Skewed
likeCount is highly skewed ($\gamma_1 = 33.35035369$)	Skewed
retweetCount is highly skewed ($\gamma_1 = 29.37350621$)	Skewed
replyCount is highly skewed ($\gamma_1 = 50.28653569$)	Skewed
url is uniformly distributed	Uniform
id has unique values	Unique
url has unique values	Unique
_id is an unsupported type, check if it needs cleaning or further analysis	Unsupported
retweetedTweet is an unsupported type, check if it needs cleaning or further analysis	Unsupported
sents is an unsupported type, check if it needs cleaning or further analysis	Unsupported
word_freq is an unsupported type, check if it needs cleaning or further analysis	Unsupported
common_words is an unsupported type, check if it needs cleaning or further analysis	Unsupported
likeCount has 4634 (49.8%) zeros	Zeros
retweetCount has 6102 (65.6%) zeros	Zeros
replyCount has 7610 (81.8%) zeros	Zeros

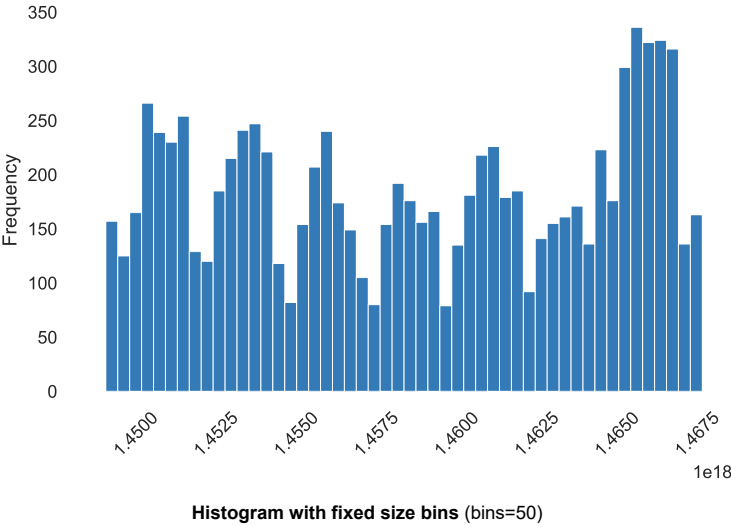
Reproduction

Analysis started	2021-12-07 04:06:15.619216
Analysis finished	2021-12-07 04:06:29.820077
Duration	14.2 seconds
Software version	pandas-profiling v3.1.0 (https://github.com/pandas-profiling/pandas-profiling)
Download configuration	config.json (data:text/plain;charset=utf-8,%7B%22title%22%3A%20%22Pandas%20Profiling%20Report%22%2C%20%22dataset%22%3A%20%7B%22description%22%3A%20%22%22%2C%22

Maximum	1.4676406×10^{18}
Range	$1.859249951 \times 10^{16}$
Interquartile range (IQR)	$1.074522648 \times 10^{16}$

Descriptive statistics

Standard deviation	$5.654354126 \times 10^{15}$
Coefficient of variation (CV)	0.003876301675
Kurtosis	-1.359956185
Mean	$1.458698161 \times 10^{18}$
Median Absolute Deviation (MAD)	$5.348558257 \times 10^{15}$
Skewness	-0.04535526896
Sum	$8.994704116 \times 10^{18}$
Variance	$3.197172058 \times 10^{31}$
Monotonicity	Not monotonic



Value	Count	Frequency (%)
1.466914816 × 10 ¹⁸	1	< 0.1%
1.454215064 × 10 ¹⁸	1	< 0.1%
1.45418927 × 10 ¹⁸	1	< 0.1%
1.454187821 × 10 ¹⁸	1	< 0.1%
1.454187499 × 10 ¹⁸	1	< 0.1%
1.454187467 × 10 ¹⁸	1	< 0.1%
1.454185677 × 10 ¹⁸	1	< 0.1%
1.454185547 × 10 ¹⁸	1	< 0.1%
1.454185332 × 10 ¹⁸	1	< 0.1%
1.454179664 × 10 ¹⁸	1	< 0.1%
Other values (9291)	9291	99.9%

Value	Count	Frequency (%)
$1.449048101 \times 10^{18}$	1	< 0.1%
$1.449048284 \times 10^{18}$	1	< 0.1%
$1.44905002 \times 10^{18}$	1	< 0.1%
$1.449050646 \times 10^{18}$	1	< 0.1%
$1.449058019 \times 10^{18}$	1	< 0.1%
$1.44906017 \times 10^{18}$	1	< 0.1%
$1.449060542 \times 10^{18}$	1	< 0.1%
$1.449061256 \times 10^{18}$	1	< 0.1%
$1.449061696 \times 10^{18}$	1	< 0.1%
$1.449062061 \times 10^{18}$	1	< 0.1%

Value	Count	Frequency (%)
1.4676406×10^{18}	1	< 0.1%
$1.467636237 \times 10^{18}$	1	< 0.1%
$1.467633654 \times 10^{18}$	1	< 0.1%
$1.467629891 \times 10^{18}$	1	< 0.1%
$1.467629428 \times 10^{18}$	1	< 0.1%
$1.467622343 \times 10^{18}$	1	< 0.1%
$1.467622325 \times 10^{18}$	1	< 0.1%
$1.467618572 \times 10^{18}$	1	< 0.1%
$1.467618376 \times 10^{18}$	1	< 0.1%
$1.467618082 \times 10^{18}$	1	< 0.1%

content

Categorical

HIGH

CARDINALITY

Distinct	7541
Distinct (%)	81.1%
Missing	0
Missing (%)	0.0%
Memory size	72.8 KiB

Length

Max length	299
Median length	209
Mean length	192.1533168
Min length	27

Characters and Unicode

Total characters	10436	
Distinct characters	1	
Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	1 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	1 (https://en.wikipedia.org/wiki/Unicode_block)	?

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

Unique

Unique	7009	?
Unique (%)	75.4%	

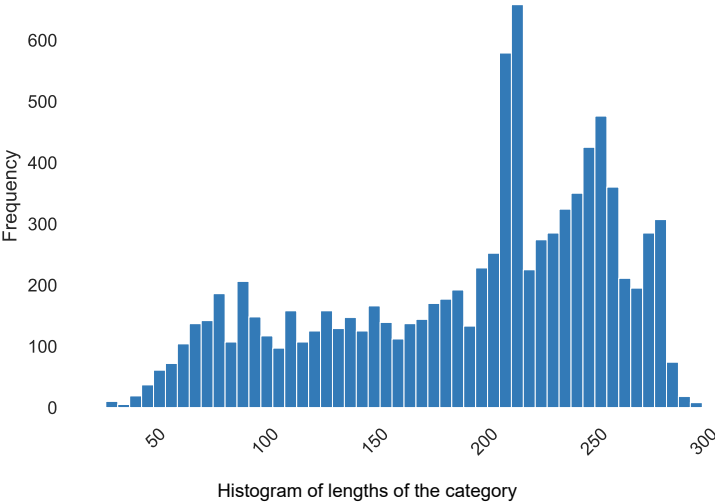
Sample

1st row	BC teachers and support staff have risen to the challenge of the COVID-19 pandemic, even as it has taken its toll. The sentiment that “this has been the hardest year of my teaching career,” has echoed across the province, with significant impacts on physical and mental health
2nd row	College students' mental health still struggling despite COVID-19 vaccines and no more lockdowns New study on college students shows almost 1/2 were at risk of clinical depression compared to a little over 1/3 prior to the pandemic, a 36 percent increase.
3rd row	Half (52%) of 126 surveyed U.S. mayors identified mental health challenges that are associated with trauma as the leading long-term consequence of the COVID-19 pandemic. (Boston University Initiative on Cities 11/22/2021)#DisabilityInfo #Zusia
4th row	NOVA Chemicals is proud to have donated to COVID-19 Recovery Fund. One example of our \$160,000 gift in action includes 3,408 people able to access subsidized counselling services to receive needed mental health support. #NOVAGivesBack
5th row	The physical effects of COVID-19 are well documented, but what about the effects on our mental health? From the psychology behind vaccine hesitancy to finding support in the middle of a pandemic, my guest Dr. Jeffrey Barkin & I talk mental health this Sunday at 11 AM

Common Values

Value	Count	Frequency (%)
The COVID-19 Pandemic has not been nice to many people. This is a suggestion to visit a mental health professional if Corona Virus...	54	0.6%
The COVID-19 Pandemic has not been enjoyable to many people. This is a reminder to visit a mental health professional if Corona V...	51	0.5%
The COVID-19 Pandemic has not been enjoyable to many people. This is a hint to visit a mental health professional if Corona Virus h...	51	0.5%
The COVID-19 Pandemic has not been pleasant to many people. This is a reminder to visit a mental health professional if Corona Vir...	51	0.5%
The COVID-19 Pandemic has not been welcoming to many people. This is a suggestion to visit a mental health professional if Coron...	50	0.5%
The COVID-19 Pandemic has not been delightful to many people. This is a suggestion to visit a mental health professional if Corona ...	50	0.5%
The COVID-19 Pandemic has not been pleasant to many people. This is a suggestion to visit a mental health professional if Corona ...	49	0.5%
The COVID-19 Pandemic has not been nice to many people. This is a reminder to visit a mental health professional if Corona Virus h...	49	0.5%
The COVID-19 Pandemic has not been delightful to many people. This is a reminder to visit a mental health professional if Corona Vi...	46	0.5%
The COVID-19 Pandemic has not been enjoyable to many people. This is a suggestion to visit a mental health professional if Corona...	46	0.5%
Other values (7531)	8804	94.7%

Length



Value	Count	Frequency (%)
the	11639	4.3%
health	8612	3.2%
covid-19	7961	2.9%
mental	7943	2.9%
to	7694	2.8%
of	7486	2.8%
and	6604	2.4%
a	5160	1.9%
in	3983	1.5%
pandemic	3395	1.2%
Other values (18234)	201501	74.1%

Most occurring characters

Value	Count	Frequency (%)
	10436	100.0%

Most occurring categories

Value	Count	Frequency (%)
Control	10436	100.0%

Most frequent character per category

Control

Value	Count	Frequency (%)
	10436	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Common	10436	100.0%

Most frequent character per script

Common

Value	Count	Frequency (%)
	10436	100.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	10436	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
	10436	100.0%

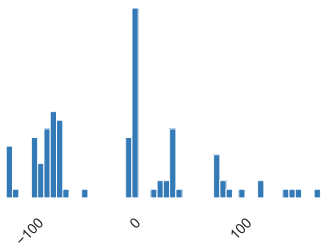
coordinates_longitude

Real number (\mathbb{R})

HIGH

CORRELATION (This variable has a high correlation with 1 fields: user_verified)
MISSING

Distinct	89
Distinct (%)	84.8%
Missing	9196
Missing (%)	98.9%
Infinite	0
Infinite (%)	0.0%
Mean	-20.11545905
Minimum	-122.7900653
Maximum	174.770789
Zeros	0
Zeros (%)	0.0%
Negative	74
Negative (%)	0.8%
Memory size	72.8 KiB



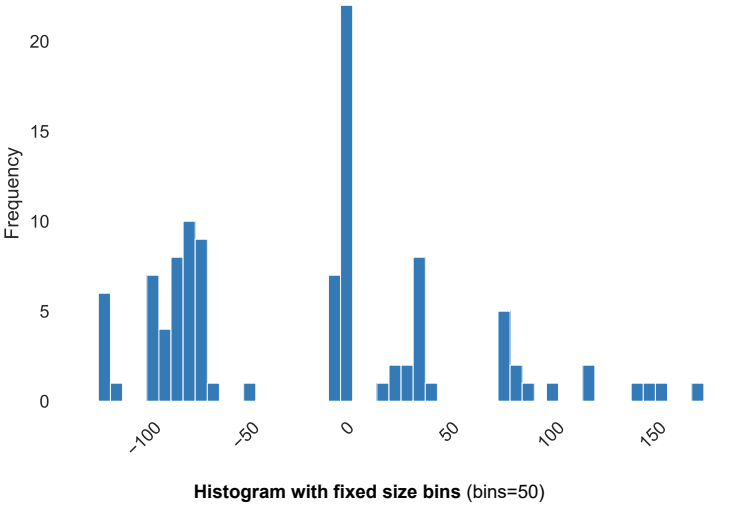
Quantile statistics

Minimum	-122.7900653
5-th percentile	-117.3014128
Q1	-79.810142
median	-4.2037009
Q3	24.78281
95-th percentile	116.8887239
Maximum	174.770789
Range	297.5608543
Interquartile range (IQR)	104.592952

Descriptive statistics

Standard deviation	71.36331265
Coefficient of variation (CV)	-3.547685015
Kurtosis	-0.2754998366
Mean	-20.11545905
Median Absolute Deviation (MAD)	69.8229741
Skewness	0.6133896277

Sum	-2112.1232
Variance	5092.722393
Monotonicity	Not monotonic



Value	Count	Frequency (%)
-76.353876	3	< 0.1%
-79.639319	3	< 0.1%
-2.9179613	2	< 0.1%
0.678274	2	< 0.1%
-73.728282	2	< 0.1%
78.426176	2	< 0.1%
36.8815837	2	< 0.1%
-95.823268	2	< 0.1%
120.9416422	2	< 0.1%
-1.4269717	2	< 0.1%
Other values (79)	83	0.9%
(Missing)	9196	98.9%

Value	Count	Frequency (%)
-122.7900653	1	< 0.1%
-122.3385555	1	< 0.1%
-118.668404	1	< 0.1%
-118.2920577	1	< 0.1%
-118.201978	1	< 0.1%
-118.198346	1	< 0.1%
-113.71368	1	< 0.1%
-98.778559	1	< 0.1%
-97.239256	1	< 0.1%
-95.823268	2	< 0.1%

Value	Count	Frequency (%)
174.770789	1	< 0.1%
152.6685228	1	< 0.1%
150.5209286	1	< 0.1%
144.5937419	1	< 0.1%
120.9416422	2	< 0.1%
100.6770508	1	< 0.1%
88.094457	1	< 0.1%
80.1851996	1	< 0.1%
80.0633414	1	< 0.1%
78.426176	2	< 0.1%

user_displayname
Categorical

HIGH

CARDINALITY

Distinct	6461
Distinct (%)	69.5%
Missing	0
Missing (%)	0.0%
Memory size	72.8 KiB

Length

Max length	50
Median length	14
Mean length	16.0063434
Min length	1

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

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

Unique

Unique	5479	?
Unique (%)	58.9%	

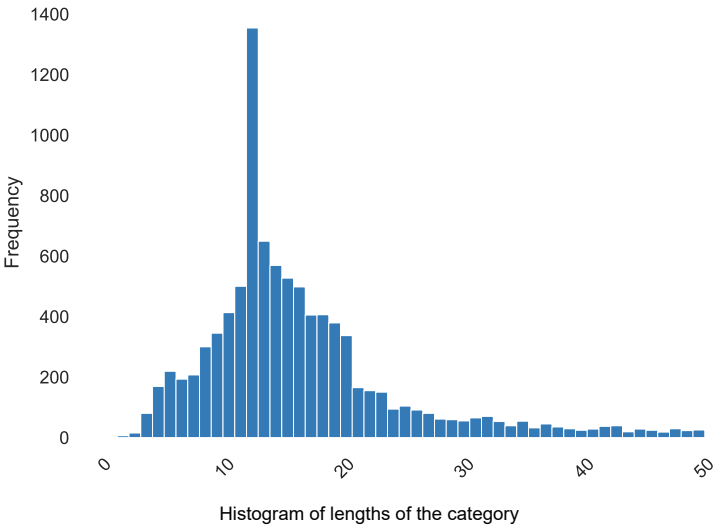
Sample

1st row	VESTA
2nd row	UCLASemelFriends
3rd row	Zusia:Dr Jackie Volk
4th row	NOVA Chemicals
5th row	Steveoe

Common Values

Value	Count	Frequency (%)
Victoria Dee	767	8.2%
CovidHelpBot	36	0.4%
UCL Open: Environment	30	0.3%
Incoming Book	29	0.3%
iWeller.com	23	0.2%
Mental Daily	23	0.2%
The European Sting	20	0.2%
Paperbirds_Coronavirus	19	0.2%
MyJournals	18	0.2%
Patricia Farrell, Ph.D.	17	0.2%
Other values (6451)	8319	89.4%

Length



Value	Count	Frequency (%)
victoria	772	3.4%
dee	768	3.4%
health	393	1.7%
the	347	1.5%
news	316	1.4%
	301	1.3%
of	208	0.9%
dr	158	0.7%
mental	156	0.7%
for	136	0.6%
Other values (8711)	18922	84.2%

Most occurring characters

Value	Count	Frequency (%)
No values found.		

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

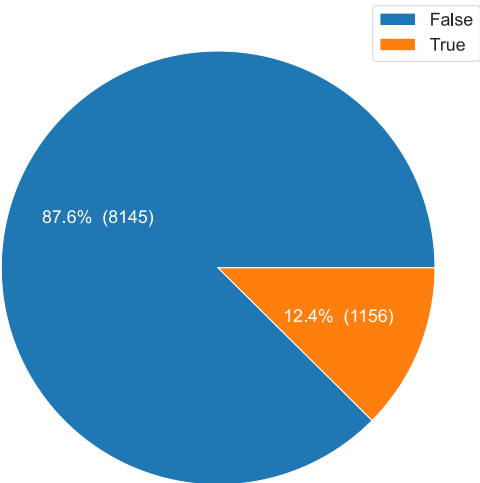
user_verified
Boolean

HIGH

CORRELATION (This variable has a high correlation with 1 fields: coordinates.longitude)

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	9.2 KiB

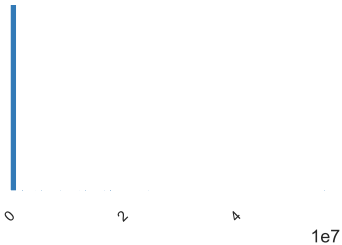
Value	Count	Frequency (%)
False	8145	87.6%
True	1156	12.4%



user_followersCount
Real number ($\mathbb{R}_{\geq 0}$)

SKEWED

Distinct	3906
Distinct (%)	42.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	69762.62036
Minimum	0
Maximum	55539298
Zeros	82
Zeros (%)	0.9%
Negative	0
Negative (%)	0.0%
Memory size	72.8 KiB



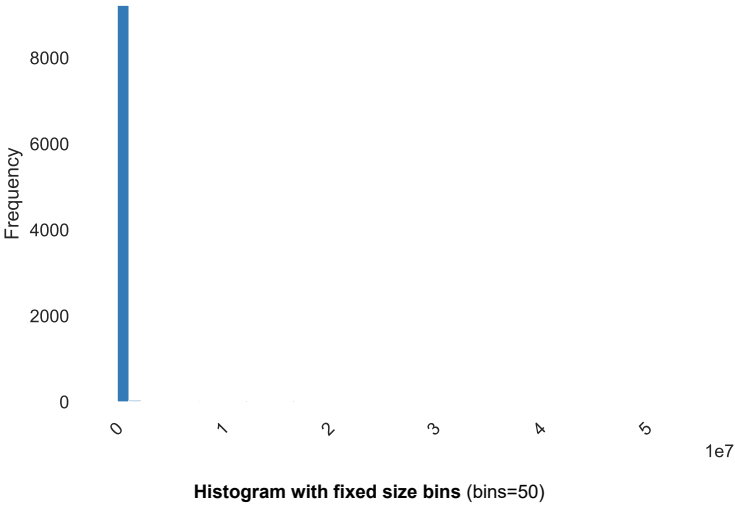
Quantile statistics

Minimum	0
5-th percentile	4
Q1	96
median	851
Q3	4928
95-th percentile	86625
Maximum	55539298
Range	55539298
Interquartile range (IQR)	4832

Descriptive statistics

Standard deviation	896423.8587
Coefficient of variation (CV)	12.84962999
Kurtosis	1779.075113
Mean	69762.62036
Median Absolute Deviation (MAD)	847
Skewness	35.38748941
Sum	648862132

Variance	8.035757345 × 10 ¹¹
Monotonicity	Not monotonic



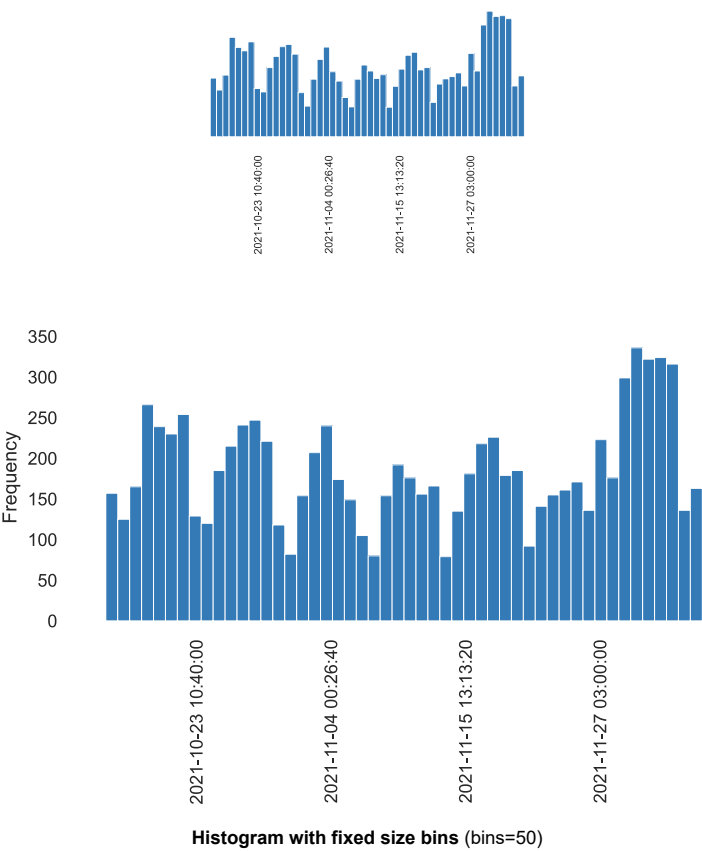
Value	Count	Frequency (%)
4	795	8.5%
0	82	0.9%
3	74	0.8%
1	66	0.7%
2	53	0.6%
6	48	0.5%
5	38	0.4%
10	37	0.4%
640	34	0.4%
7	32	0.3%
Other values (3896)	8042	86.5%

Value	Count	Frequency (%)
0	82	0.9%
1	66	0.7%
2	53	0.6%
3	74	0.8%
4	795	8.5%
5	38	0.4%
6	48	0.5%
7	32	0.3%
8	31	0.3%
9	26	0.3%

Value	Count	Frequency (%)
55539298	1	< 0.1%
24061090	1	< 0.1%
24054399	1	< 0.1%
18513404	1	< 0.1%
18508660	1	< 0.1%
16981223	2	< 0.1%
13430328	1	< 0.1%
13429425	1	< 0.1%
12950392	2	< 0.1%
8986646	1	< 0.1%

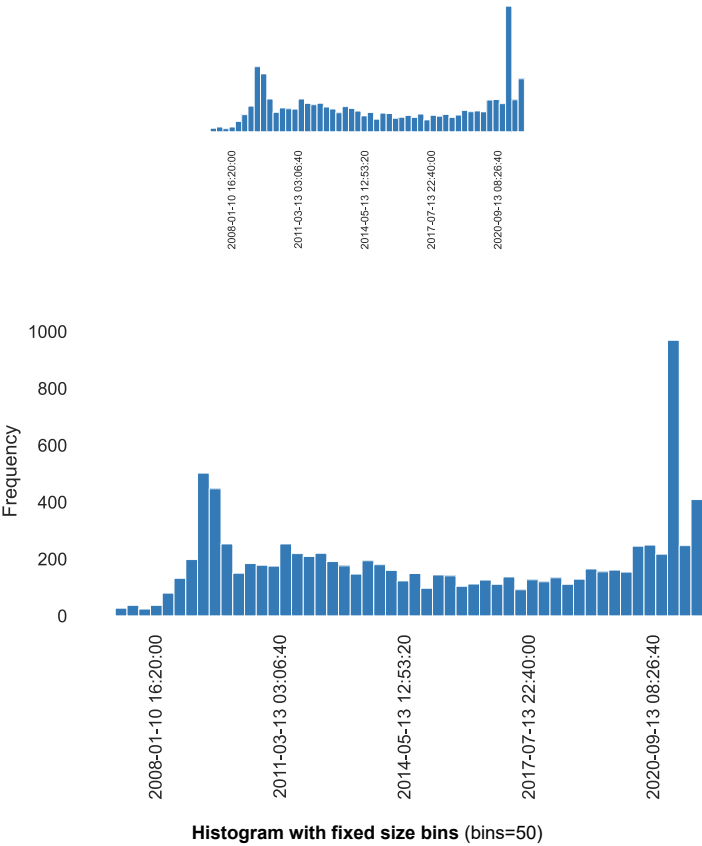
date
Date

Distinct	9203
Distinct (%)	98.9%
Missing	0
Missing (%)	0.0%
Memory size	72.8 KiB
Minimum	2021-10-15 16:22:43
Maximum	2021-12-05 23:42:41



user_created
Date

Distinct	6476
Distinct (%)	69.6%
Missing	0
Missing (%)	0.0%
Memory size	72.8 KiB
Minimum	2006-12-18 23:06:15
Maximum	2021-12-05 10:41:23



likeCount

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 1 fields: retweetCount)
HIGH CORRELATION (This variable has a high correlation with 1 fields: retweetCount)
HIGH CORRELATION (This variable has a high correlation with 1 fields: retweetCount)
HIGH CORRELATION (This variable has a high correlation with 2 fields: retweetCount, replyCount)
SKEWED
ZEROS

Distinct	135
Distinct (%)	1.5%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	4.974626384
Minimum	0
Maximum	2327
Zeros	4634
Zeros (%)	49.8%
Negative	0
Negative (%)	0.0%
Memory size	72.8 KiB



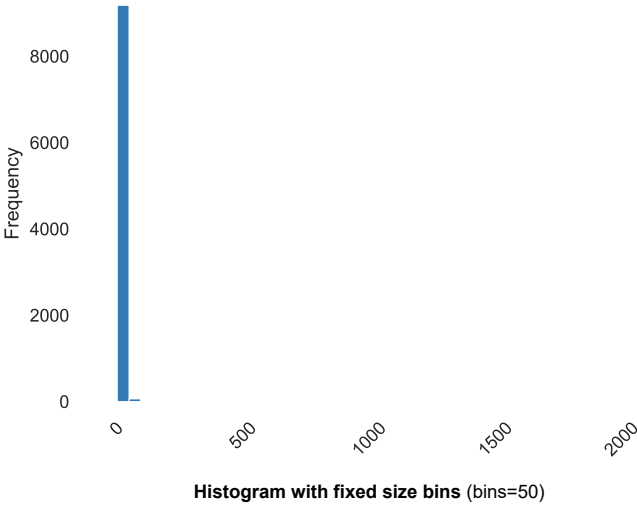
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	1
Q3	3
95-th percentile	13
Maximum	2327
Range	2327
Interquartile range (IQR)	3

Descriptive statistics

Standard deviation	41.27433008
Coefficient of variation (CV)	8.296970847
Kurtosis	1511.399914

Mean	4.974626384
Median Absolute Deviation (MAD)	1
Skewness	33.35035369
Sum	46269
Variance	1703.570324
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	4634	49.8%
1	1497	16.1%
2	812	8.7%
3	487	5.2%
4	341	3.7%
5	242	2.6%
6	197	2.1%
7	160	1.7%
8	117	1.3%
9	104	1.1%
Other values (125)	710	7.6%

Value	Count	Frequency (%)
0	4634	49.8%
1	1497	16.1%
2	812	8.7%
3	487	5.2%
4	341	3.7%
5	242	2.6%
6	197	2.1%
7	160	1.7%
8	117	1.3%
9	104	1.1%

Value	Count	Frequency (%)
2327	1	< 0.1%
1730	1	< 0.1%
932	1	< 0.1%
887	1	< 0.1%
854	1	< 0.1%
717	1	< 0.1%
664	1	< 0.1%
607	1	< 0.1%
567	1	< 0.1%
523	1	< 0.1%

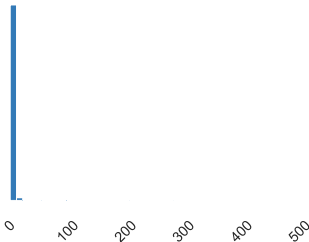
retweetCount

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 1 fields: likeCount)
HIGH CORRELATION (This variable has a high correlation with 2 fields: likeCount, replyCount)
HIGH CORRELATION (This variable has a high correlation with 1 fields: likeCount)
HIGH CORRELATION (This variable has a high correlation with 2 fields: likeCount, replyCount)
SKEWED
ZEROS

Distinct	73
Distinct (%)	0.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1.605956349
Minimum	0
Maximum	542
Zeros	6102
Zeros (%)	65.6%
Negative	0
Negative (%)	0.0%
Memory size	72.8 KiB



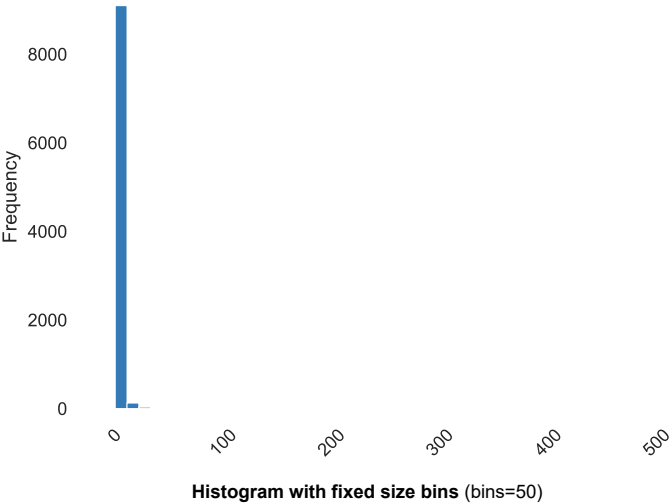
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	1
95-th percentile	6
Maximum	542
Range	542
Interquartile range (IQR)	1

Descriptive statistics

Standard deviation	9.813458117
Coefficient of variation (CV)	6.110663048
Kurtosis	1247.250964

Mean	1.605956349
Median Absolute Deviation (MAD)	0
Skewness	29.37350621
Sum	14937
Variance	96.30396022
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	6102	65.6%
1	1330	14.3%
2	648	7.0%
3	330	3.5%
4	210	2.3%
5	155	1.7%
6	108	1.2%
7	81	0.9%
8	61	0.7%
9	41	0.4%
Other values (63)	235	2.5%

Value	Count	Frequency (%)
0	6102	65.6%
1	1330	14.3%
2	648	7.0%
3	330	3.5%
4	210	2.3%
5	155	1.7%
6	108	1.2%
7	81	0.9%
8	61	0.7%
9	41	0.4%

Value	Count	Frequency (%)
542	1	< 0.1%
292	1	< 0.1%
275	1	< 0.1%
242	1	< 0.1%
207	1	< 0.1%
199	1	< 0.1%
196	1	< 0.1%
153	1	< 0.1%
124	1	< 0.1%
119	1	< 0.1%

url		
Categorical		
		HIGH
CARDINALITY		
UNIFORM		
UNIQUE		
<hr/>		
Distinct	9301	
<hr/>		
Distinct (%)	100.0%	
<hr/>		
Missing	0	
<hr/>		
Missing (%)	0.0%	
<hr/>		
Memory size	72.8 KiB	

Length

Max length	62
Median length	59
Mean length	58.61272981
Min length	49

Characters and Unicode

Total characters	0	
Distinct characters	0	
Distinct categories	0 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	0 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	0 (https://en.wikipedia.org/wiki/Unicode_block)	?

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

Unique

Unique	9301	?
Unique (%)	100.0%	

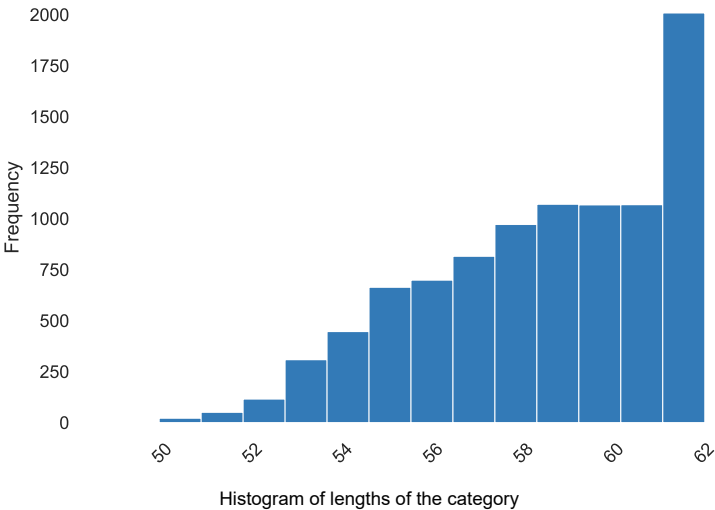
Sample

1st row	https://twitter.com/VESTA39/status/1466914816241455104
2nd row	https://twitter.com/UCLASemelFriend/status/1466908091300003840
3rd row	https://twitter.com/DrJackieVolk/status/1466905171452170246
4th row	https://twitter.com/NOVACchemicals/status/1466890614595801090
5th row	https://twitter.com/SteveWoods/status/1466883989600444429

Common Values

Value	Count	Frequency (%)
https://twitter.com/VESTA39/status/1466914816241455104	1	< 0.1%
https://twitter.com/SmithsonianMag/status/1454215063787741186	1	< 0.1%
https://twitter.com/TodayHeadline42/status/1454189269766004743	1	< 0.1%
https://twitter.com/Jake_1791/status/1454187821032304643	1	< 0.1%
https://twitter.com/LApodnetwork/status/1454187499132125188	1	< 0.1%
https://twitter.com/Victori44685362/status/1454187466626252805	1	< 0.1%
https://twitter.com/BehavioralHlth/status/1454185676526718979	1	< 0.1%
https://twitter.com/HarleyBendaWBOY/status/1454185546738241545	1	< 0.1%
https://twitter.com/Addiction_Pro/status/1454185332161843203	1	< 0.1%
https://twitter.com/WashTimes/status/1454179663547355138	1	< 0.1%
Other values (9291)	9291	99.9%

Length



Value	Count	Frequency (%)
https://twitter.com/vesta39/status/1466914816241455104	1	< 0.1%
https://twitter.com/deccanherald/status/1466004154367815681	1	< 0.1%
https://twitter.com/myjournals/status/1466874663317884929	1	< 0.1%
https://twitter.com/drjackievolk/status/1466905171452170246	1	< 0.1%
https://twitter.com/novachemicals/status/1466890614595801090	1	< 0.1%
https://twitter.com/stevoewoods/status/1466883989600444429	1	< 0.1%
https://twitter.com/andrewwalkom/status/1466879328247947265	1	< 0.1%
https://twitter.com/carolyn_bennett/status/1466878666642706440	1	< 0.1%
https://twitter.com/bibliotecasbuap/status/1466877152050176000	1	< 0.1%
https://twitter.com/snjreic/status/1466874394618257411	1	< 0.1%
Other values (9291)	9291	99.9%

Most occurring characters

Value	Count	Frequency (%)
No values found.		

Most occurring categories

Value	Count	Frequency (%)
No values found.		

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
No values found.		

Most frequent character per script

Most occurring blocks

Value	Count	Frequency (%)
No values found.		

Most frequent character per block

retweetedTweet
Unsupported

MISSING

REJECTED
UNSUPPORTED

Missing	9301
Missing (%)	100.0%
Memory size	72.8 KiB

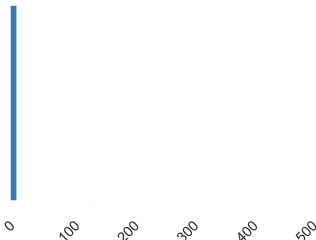
replyCount

Real number ($\mathbb{R}_{\geq 0}$)

HIGH

CORRELATION (This variable has a high correlation with 1 fields: retweetCount)
HIGH CORRELATION (This variable has a high correlation with 2 fields: likeCount, retweetCount)
SKEWED
ZEROS

Distinct	47
Distinct (%)	0.5%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	0.5921943877
Minimum	0
Maximum	529
Zeros	7610
Zeros (%)	81.8%
Negative	0
Negative (%)	0.0%
Memory size	72.8 KiB



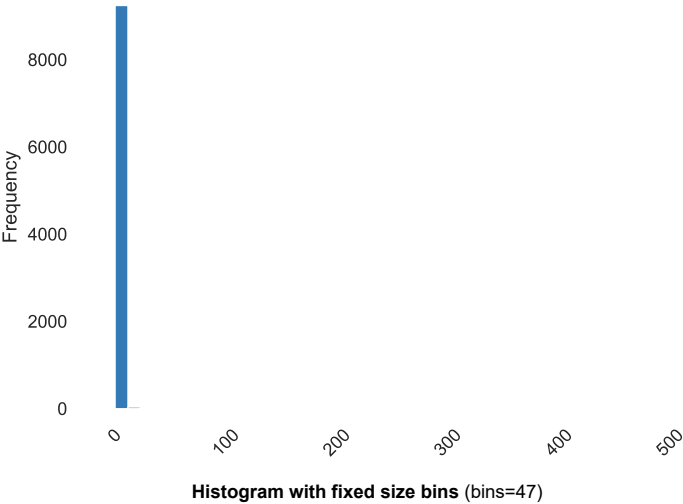
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	2

Maximum	529
Range	529
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	7.331653229
Coefficient of variation (CV)	12.38048415
Kurtosis	3214.693574
Mean	0.5921943877
Median Absolute Deviation (MAD)	0
Skewness	50.28653569
Sum	5508
Variance	53.75313907
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	7610	81.8%
1	1213	13.0%
2	223	2.4%
3	75	0.8%
4	36	0.4%
5	24	0.3%
6	17	0.2%
7	11	0.1%
10	9	0.1%
9	9	0.1%
Other values (37)	74	0.8%

Value	Count	Frequency (%)
0	7610	81.8%
1	1213	13.0%
2	223	2.4%
3	75	0.8%
4	36	0.4%
5	24	0.3%
6	17	0.2%
7	11	0.1%
8	6	0.1%
9	9	0.1%

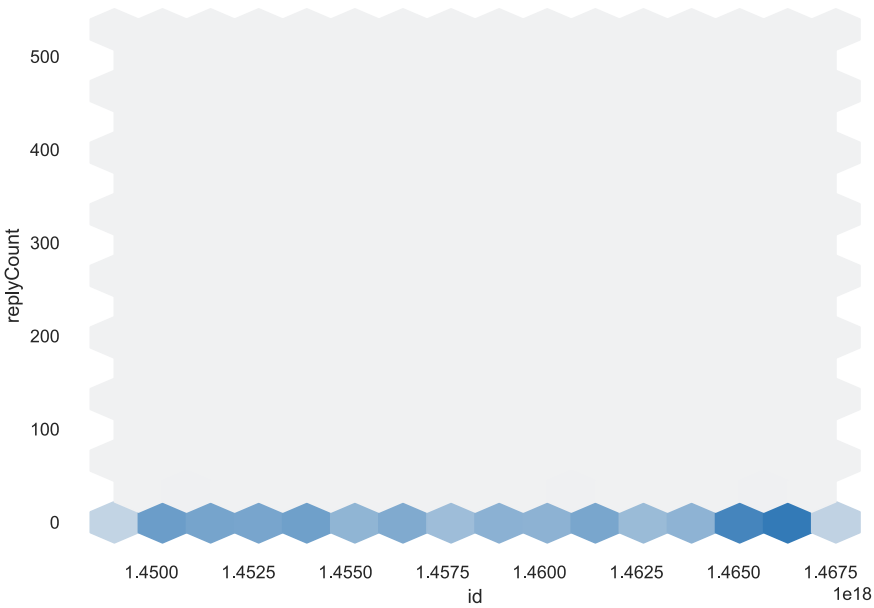
Value	Count	Frequency (%)
529	1	< 0.1%
289	1	< 0.1%
156	1	< 0.1%
135	1	< 0.1%
133	1	< 0.1%
108	1	< 0.1%
82	1	< 0.1%
73	1	< 0.1%
70	1	< 0.1%
68	1	< 0.1%

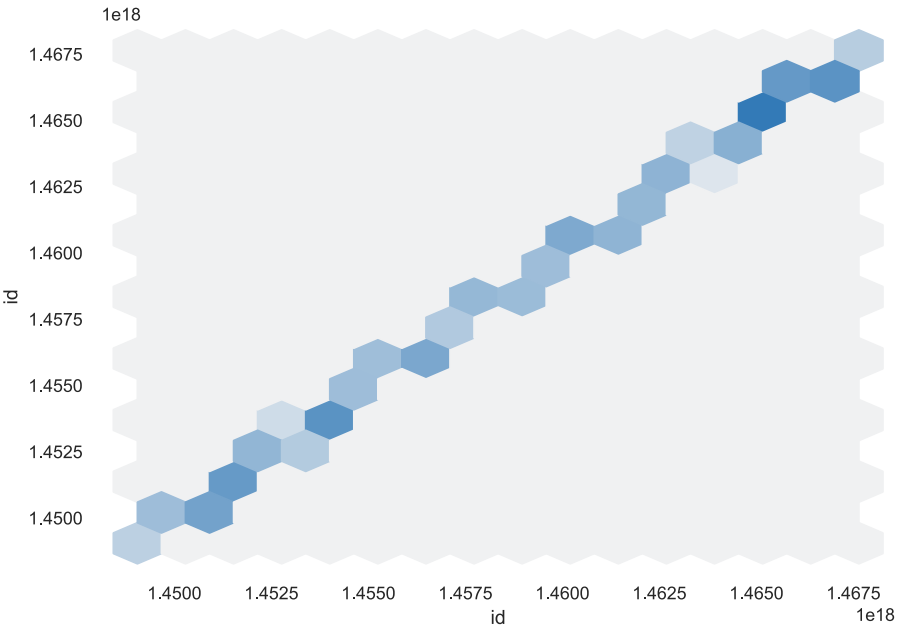
sents		
Unsupported		
UNSUPPORTED		REJECTED
Missing	0	
Missing (%)	0.0%	
Memory size	72.8 KiB	

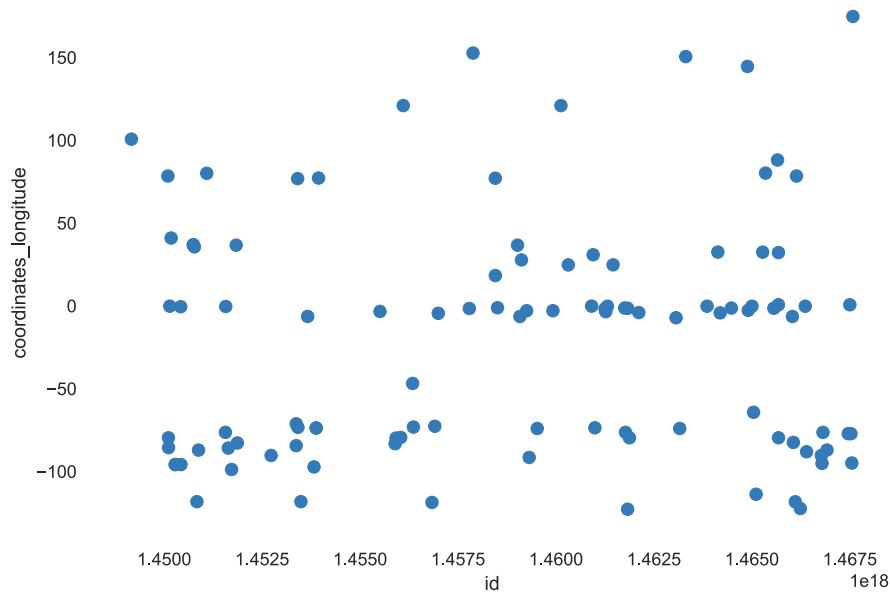
word_freq		
Unsupported		
UNSUPPORTED		REJECTED
Missing	0	
Missing (%)	0.0%	
Memory size	72.8 KiB	

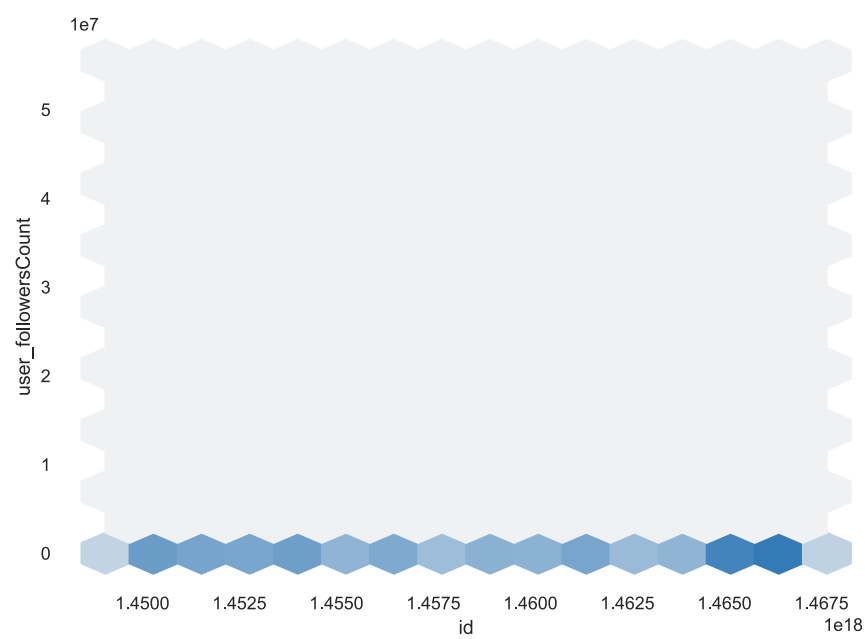
common_words		
Unsupported		
UNSUPPORTED		REJECTED
Missing	0	
Missing (%)	0.0%	
Memory size	72.8 KiB	

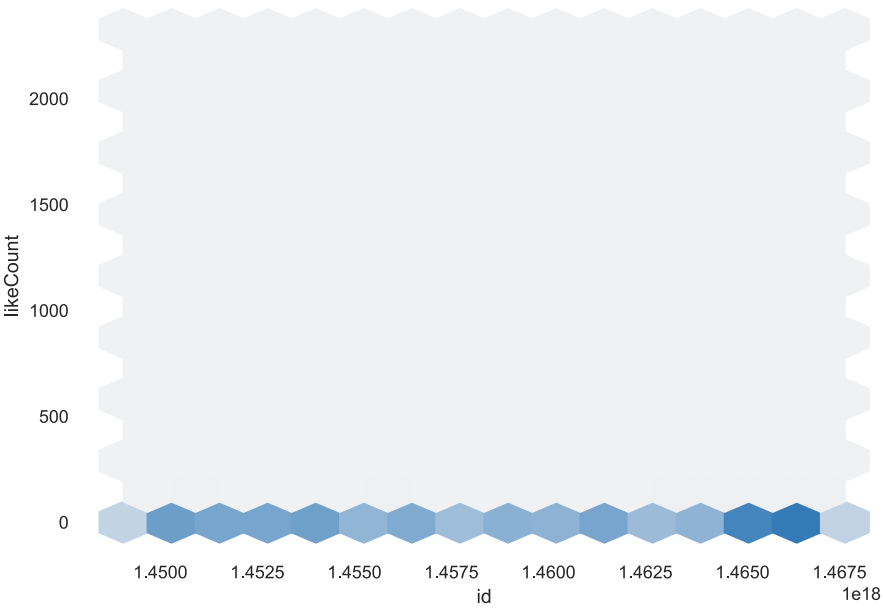
Interactions

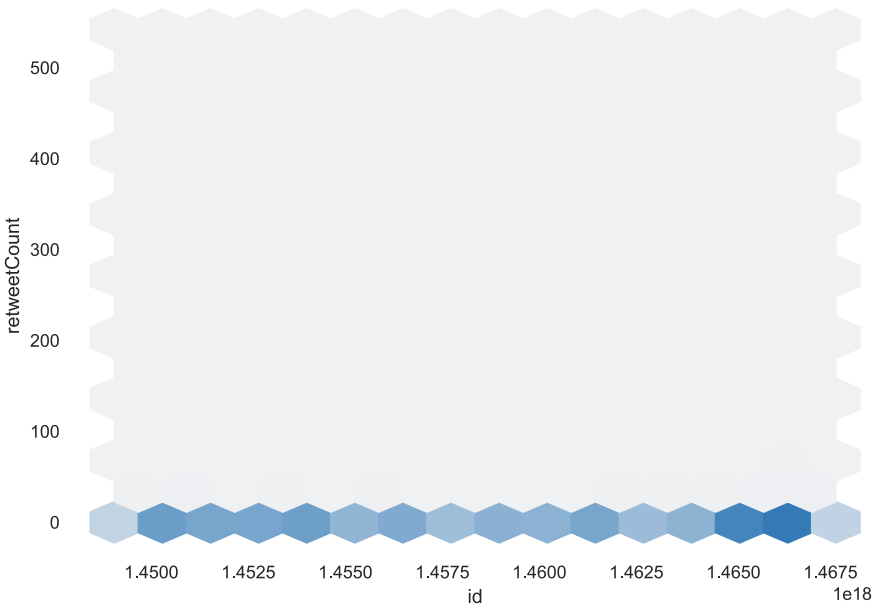


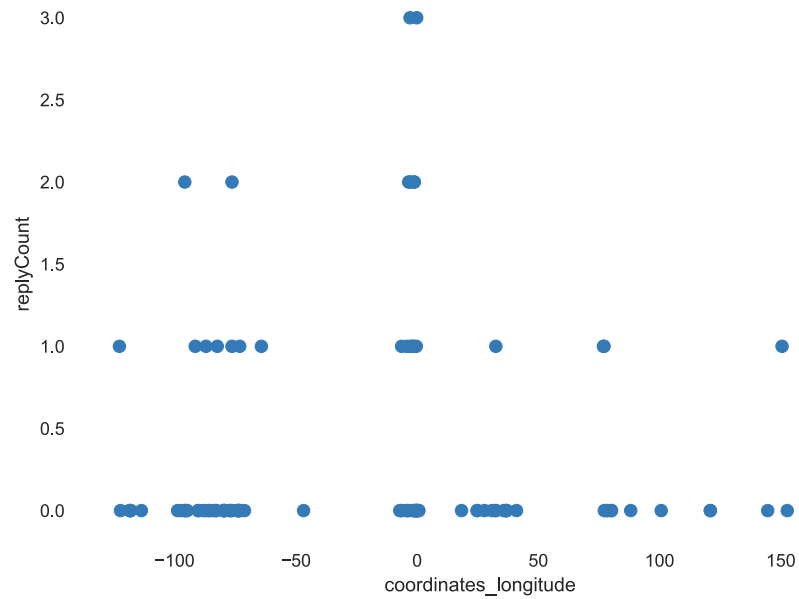


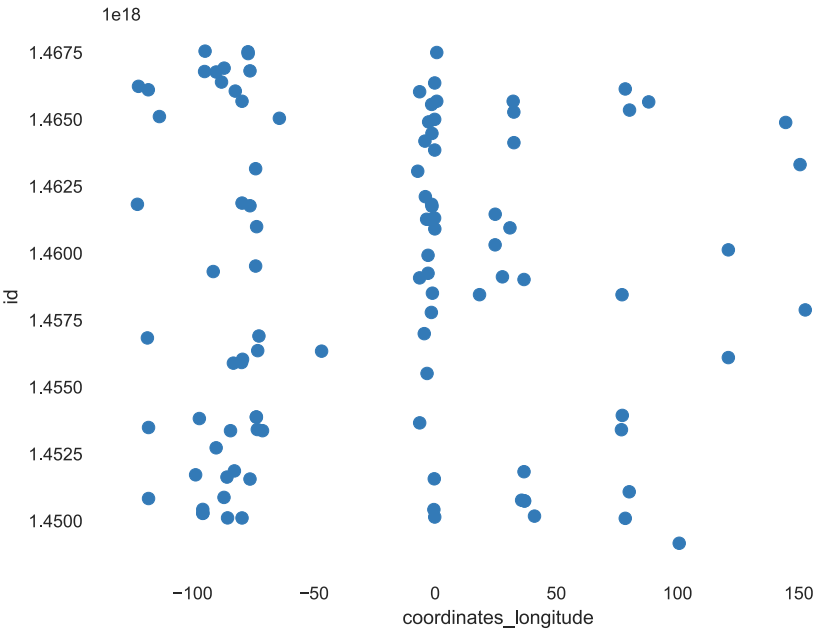


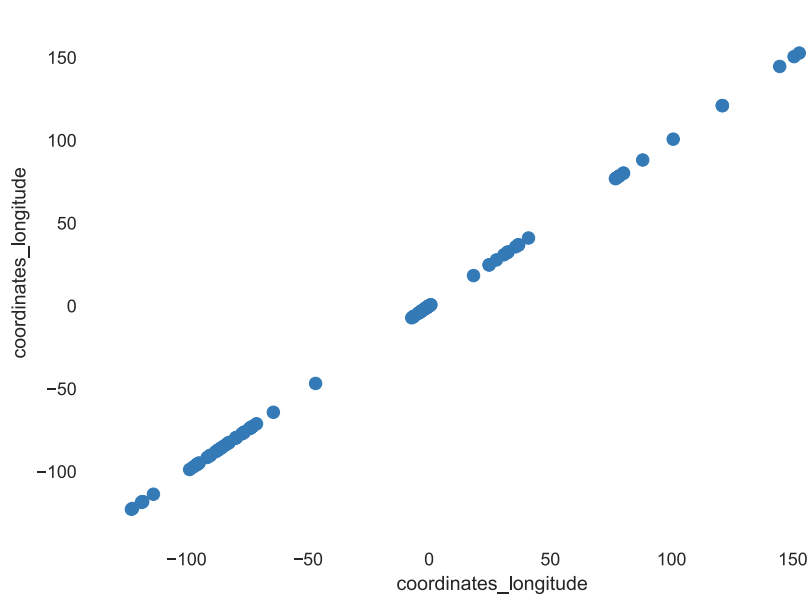


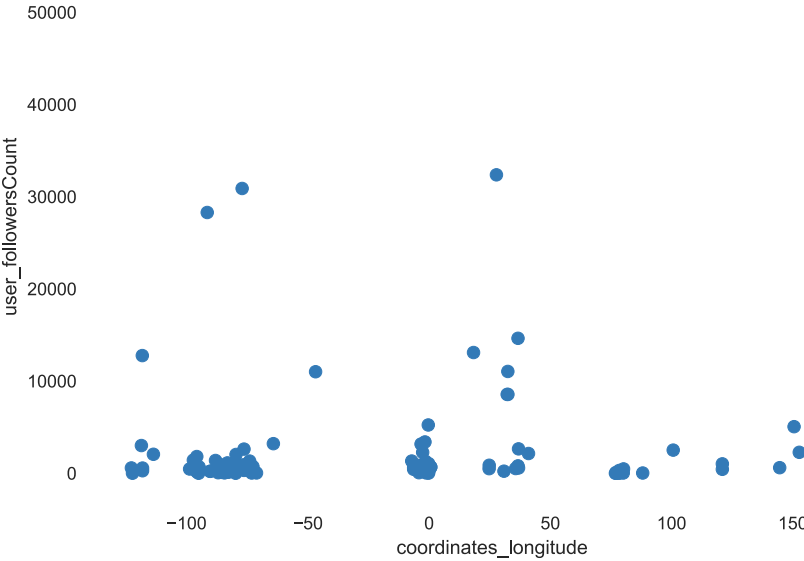


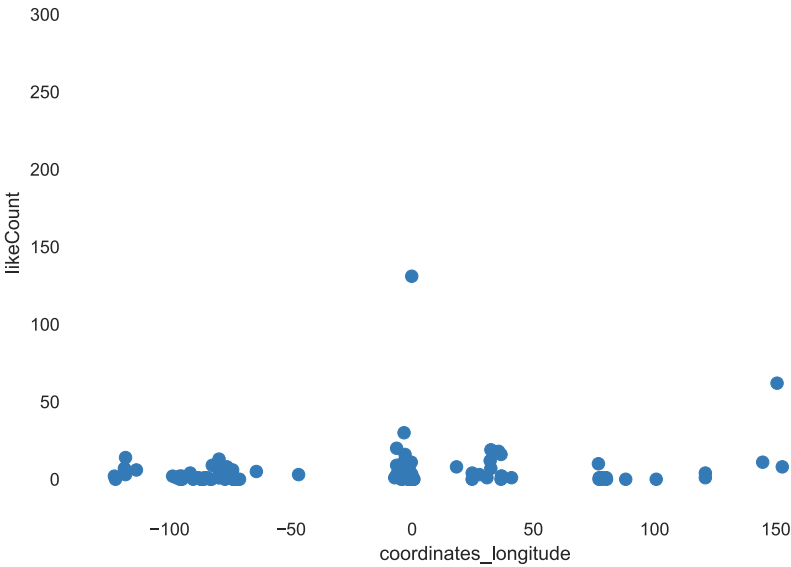


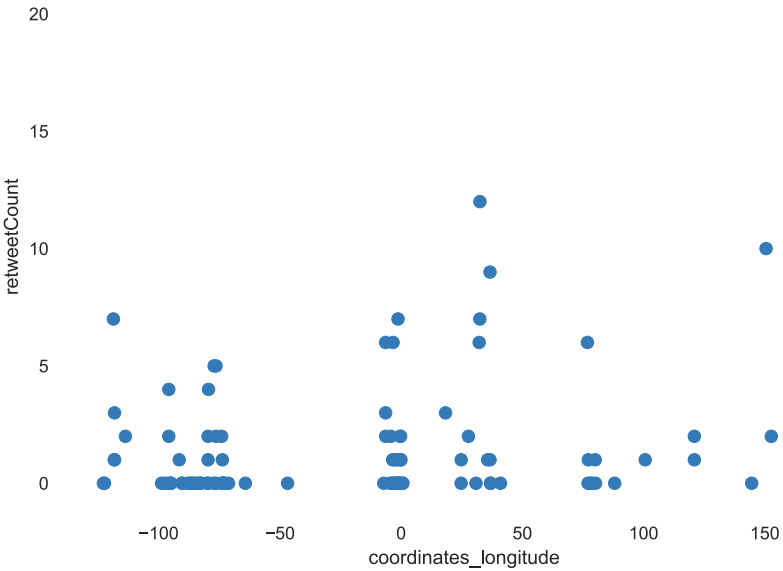


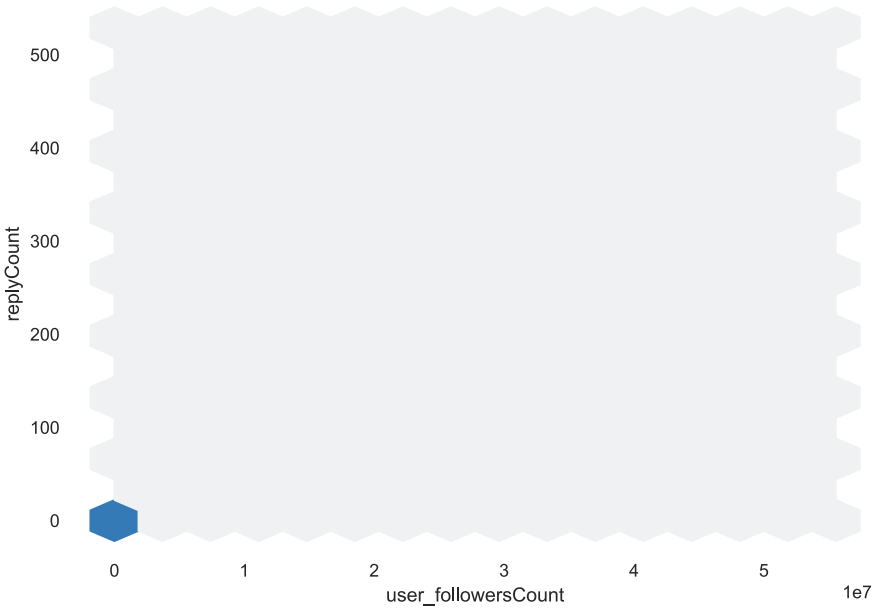


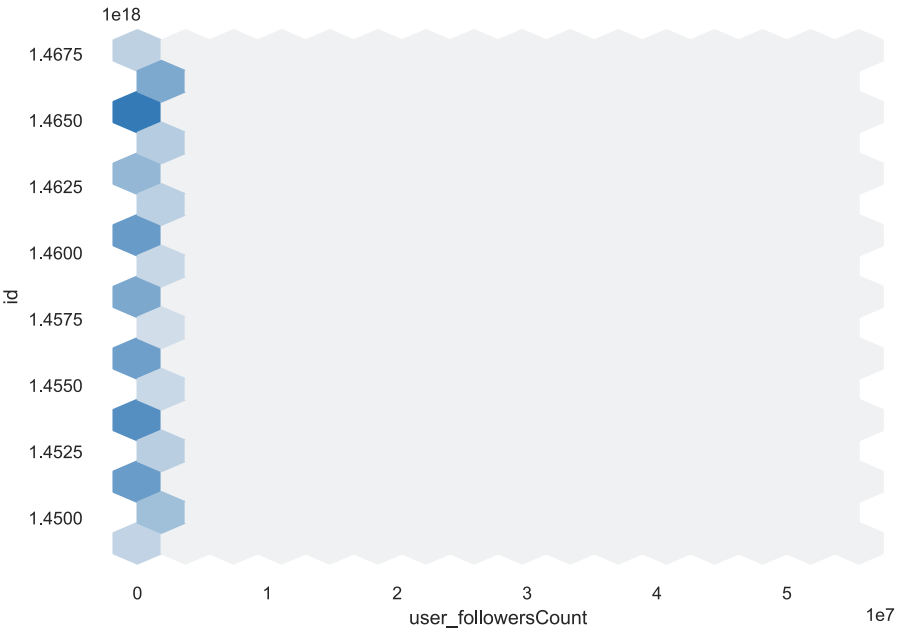


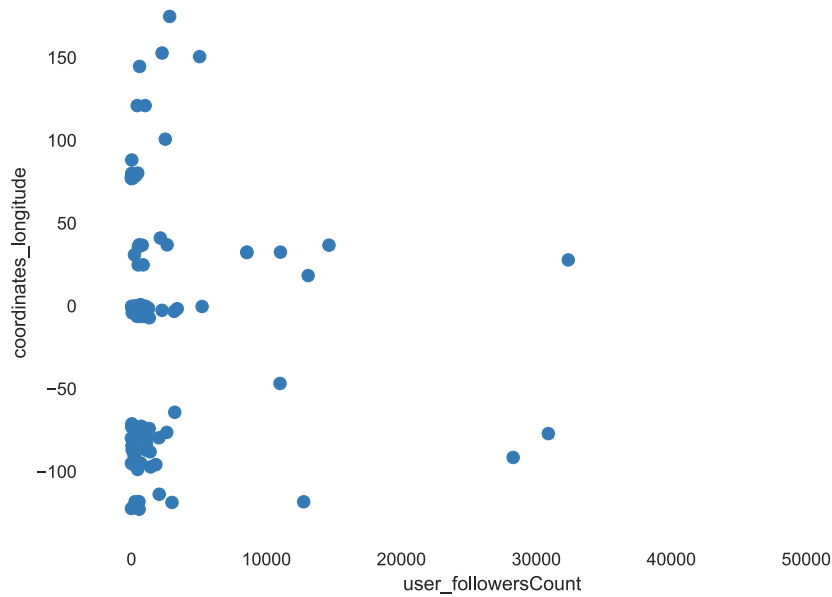


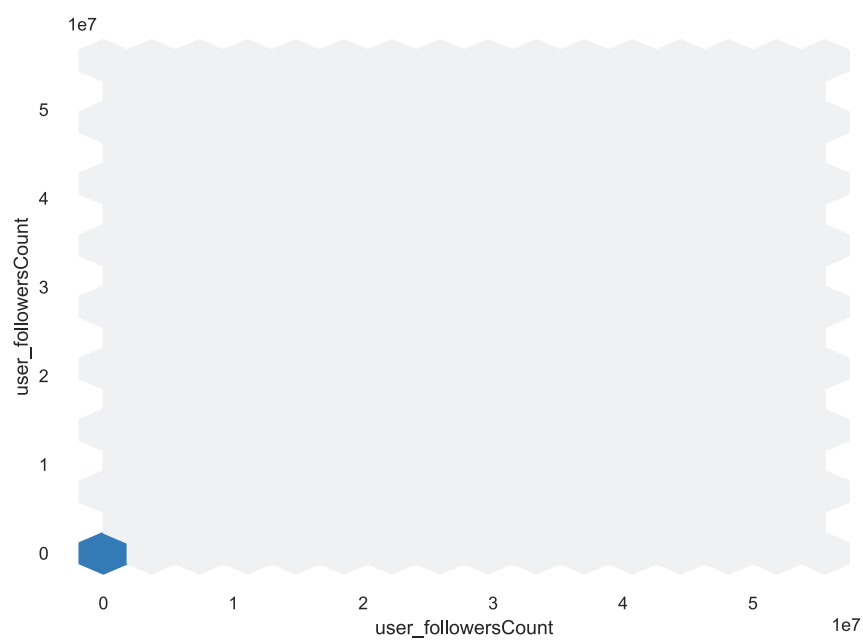


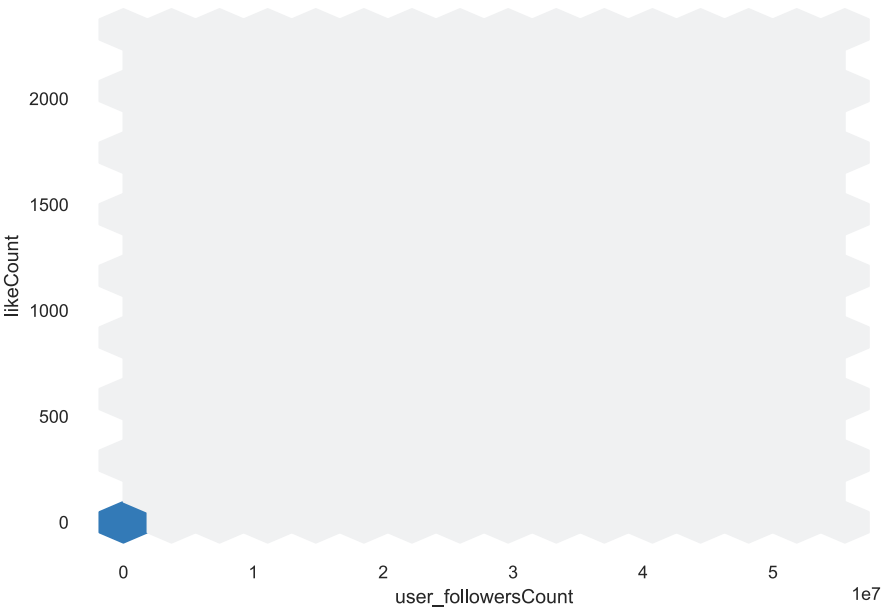


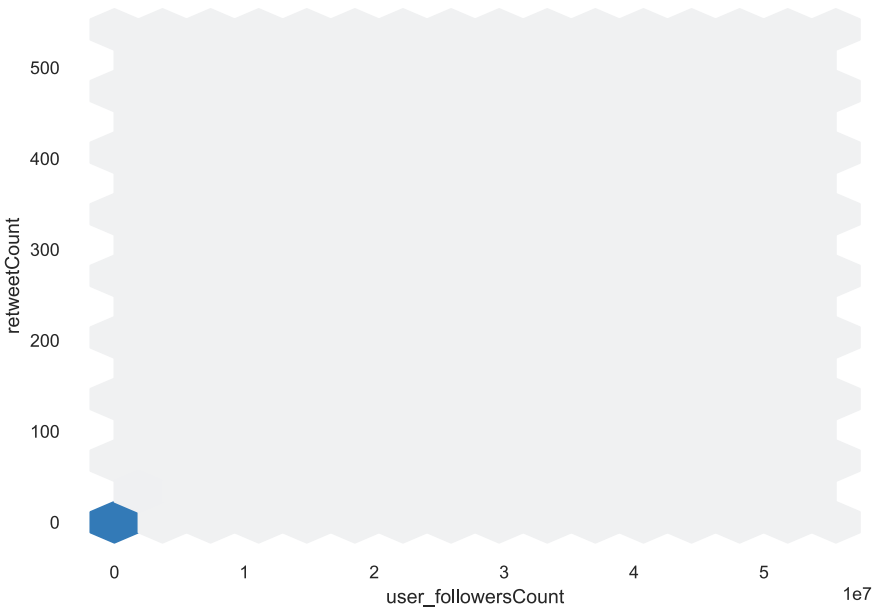


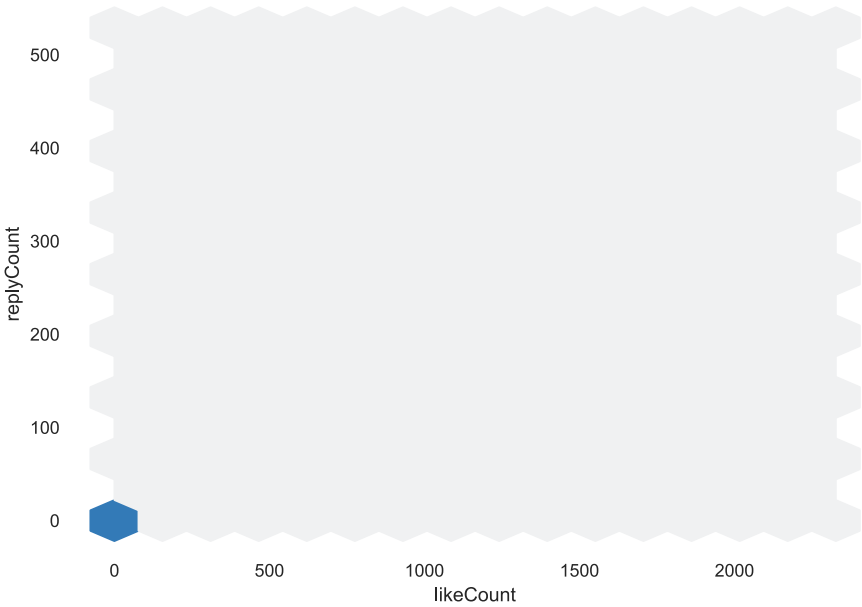


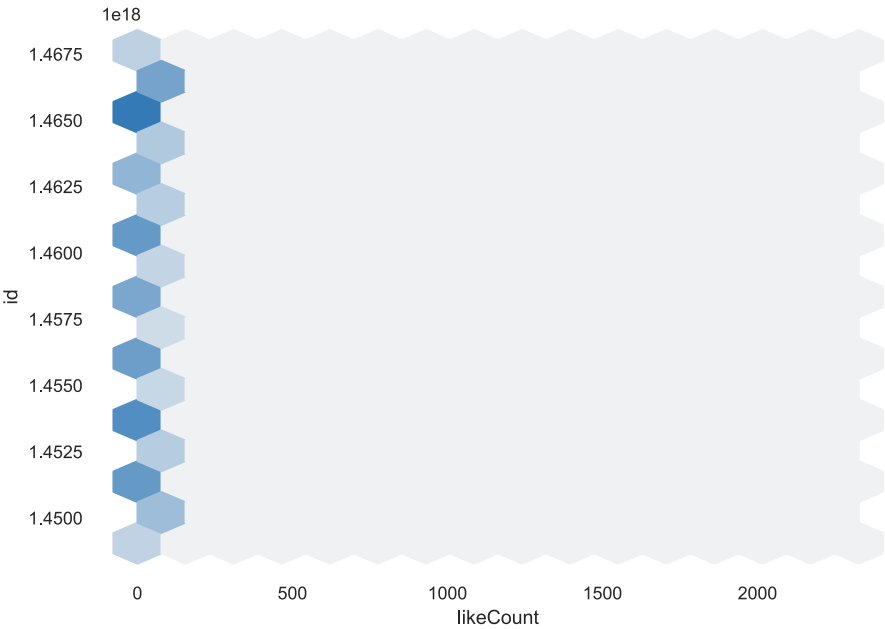


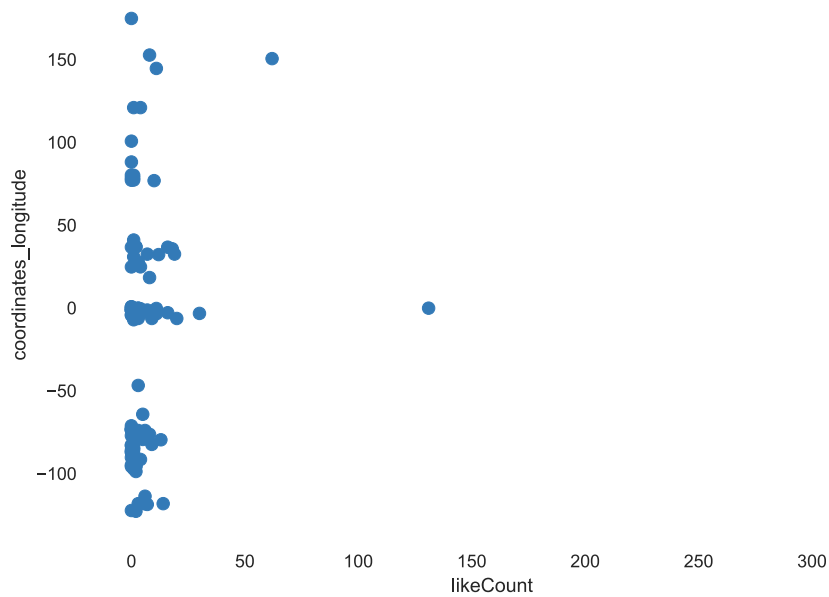


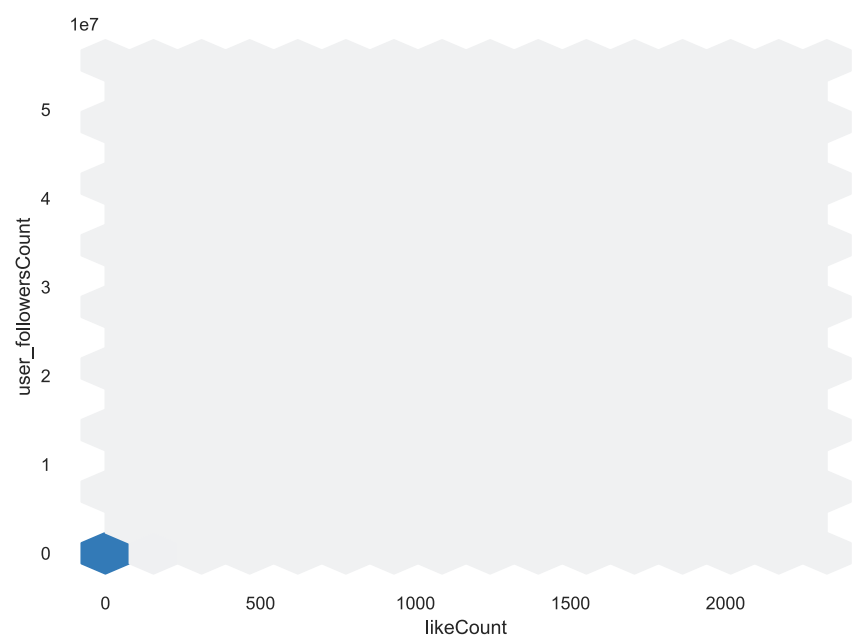


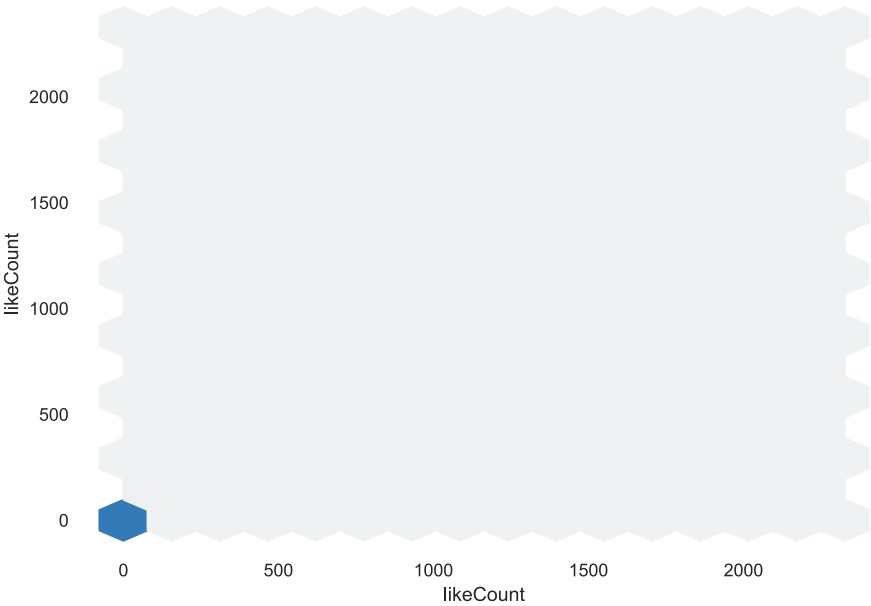


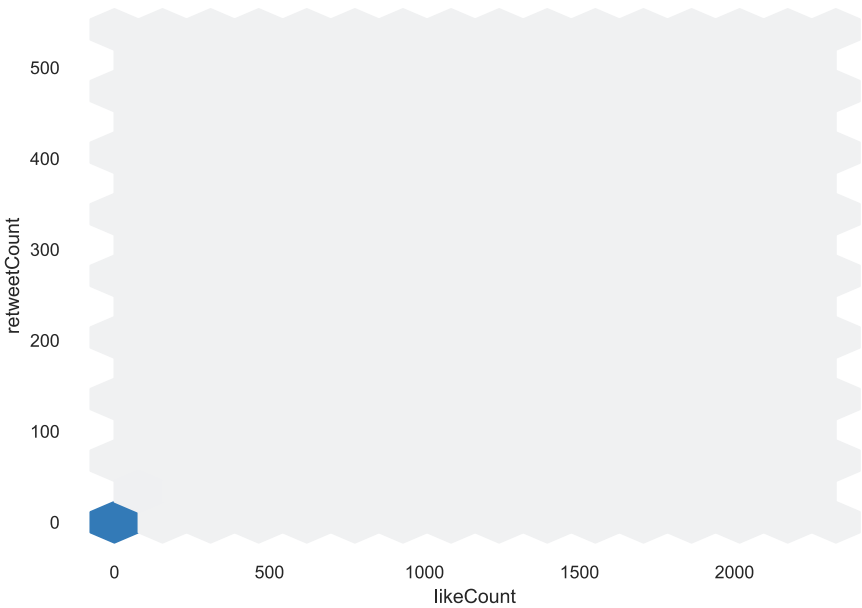


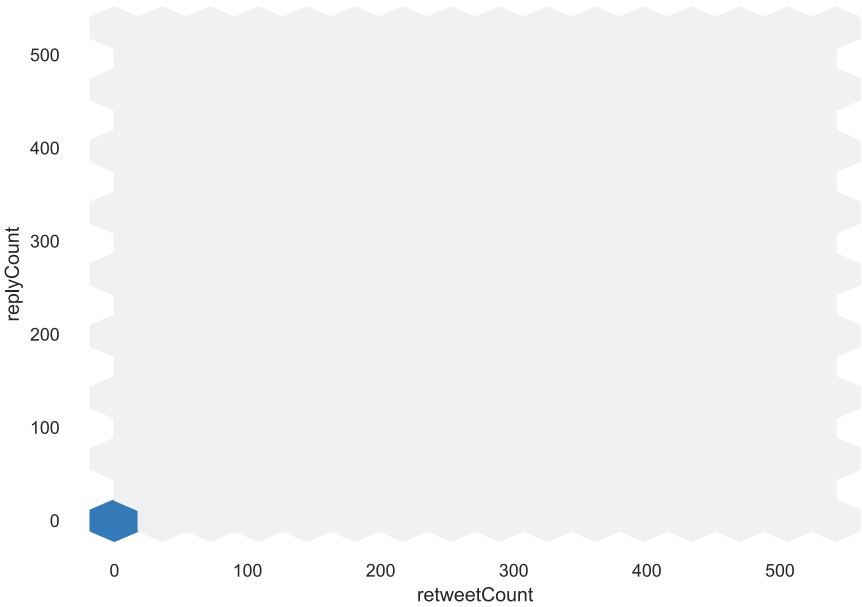


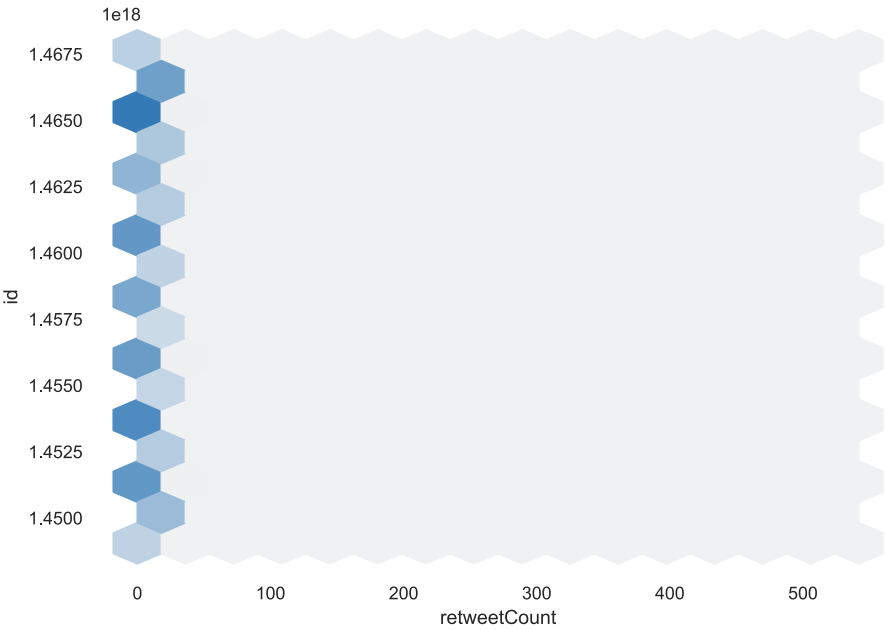


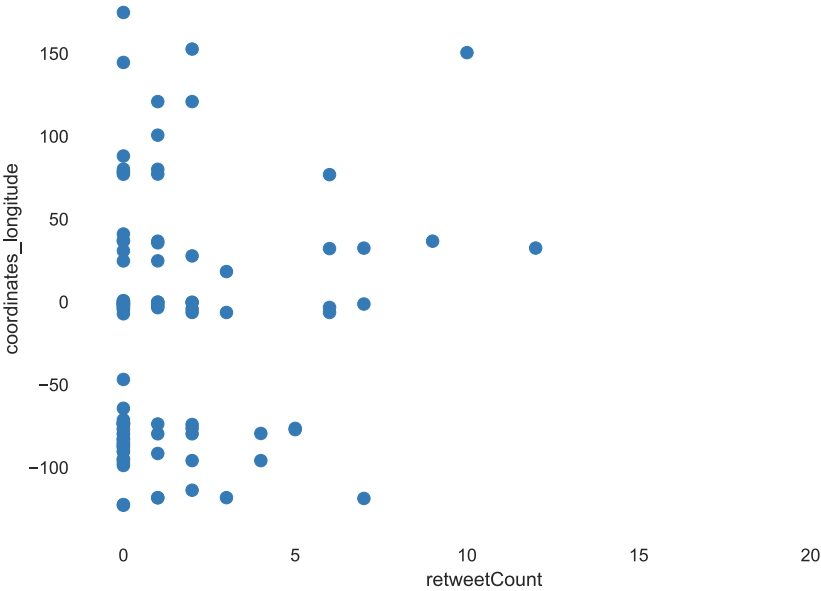


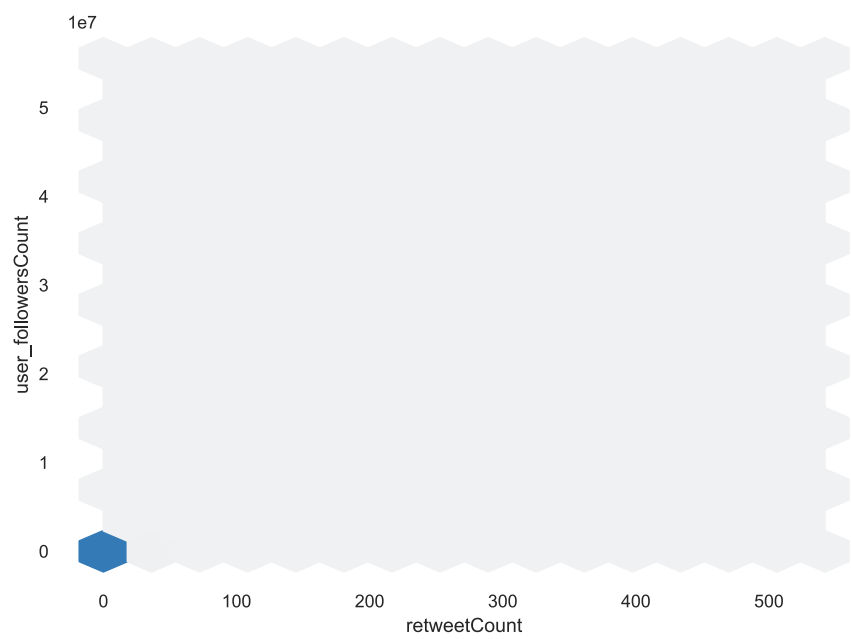


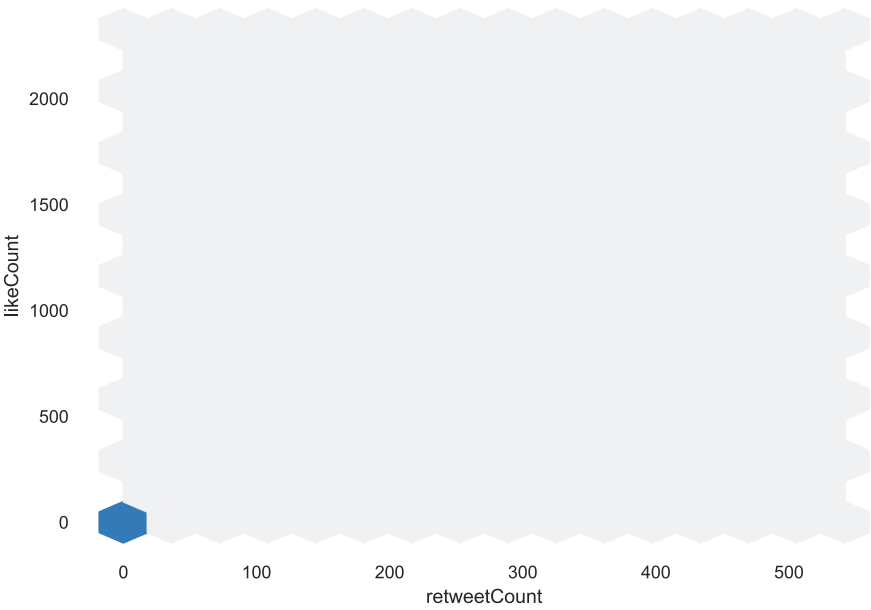


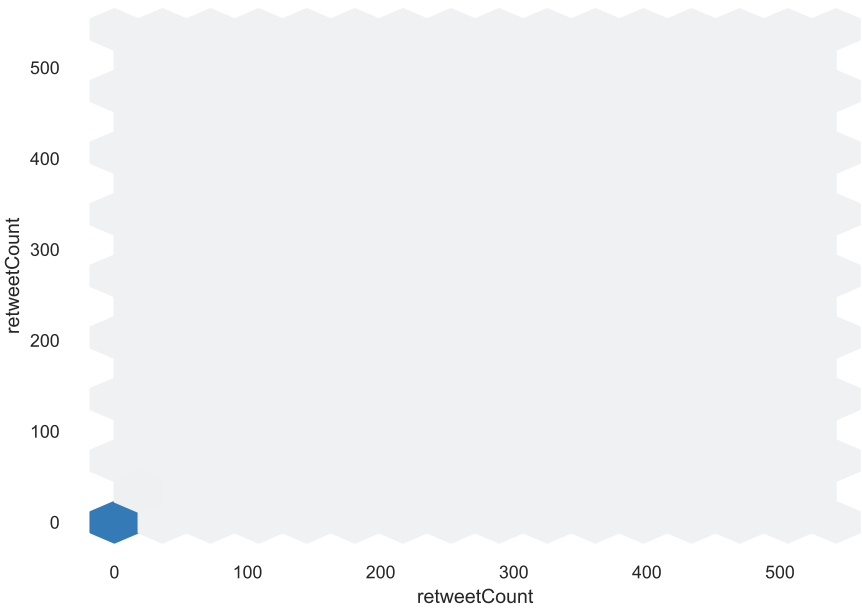


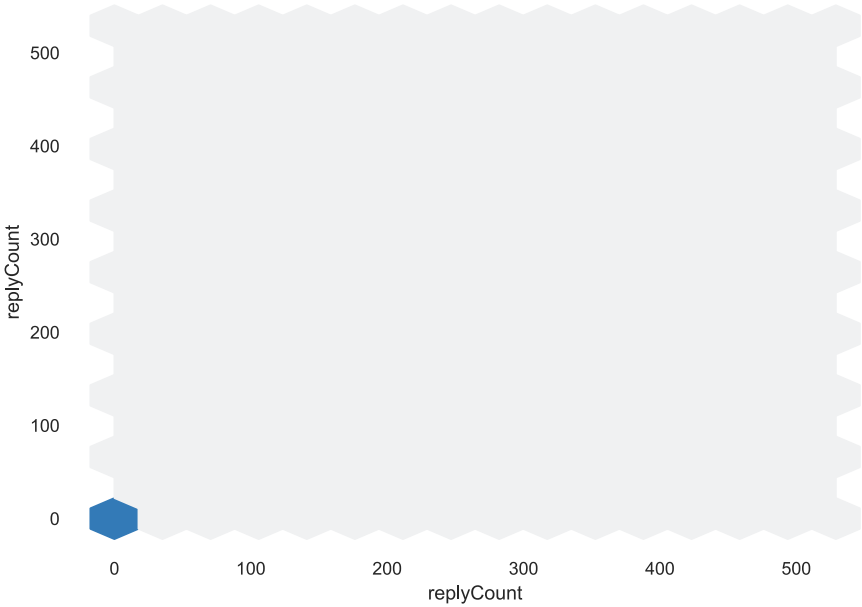


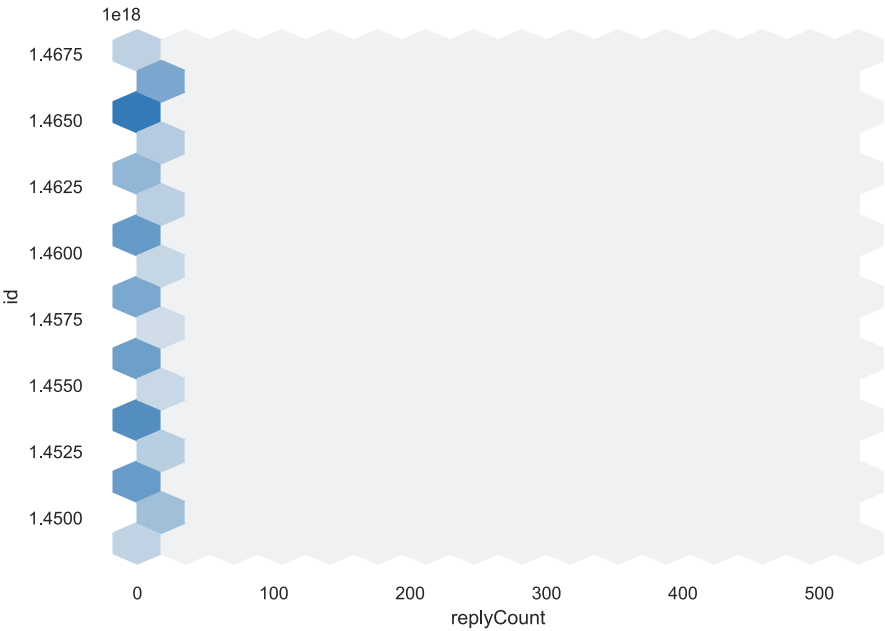


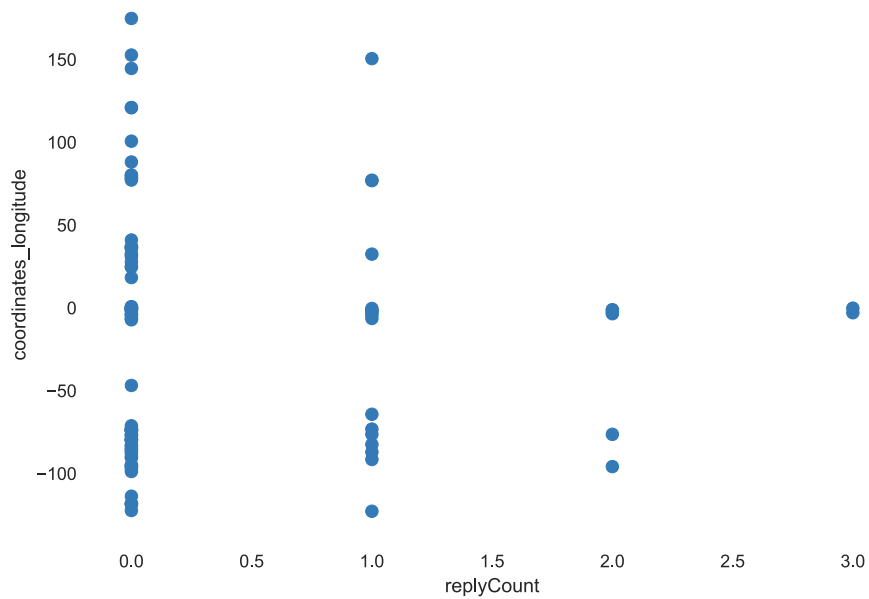


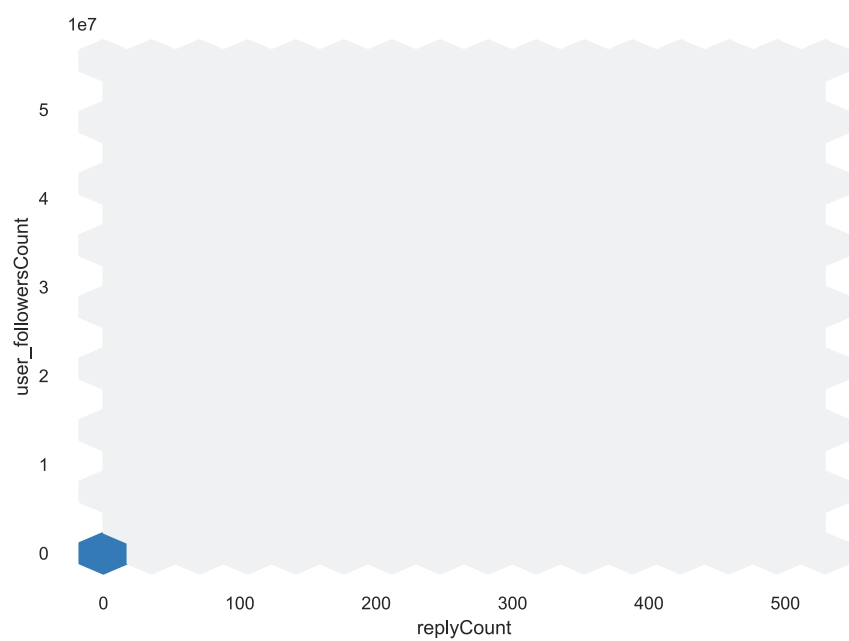


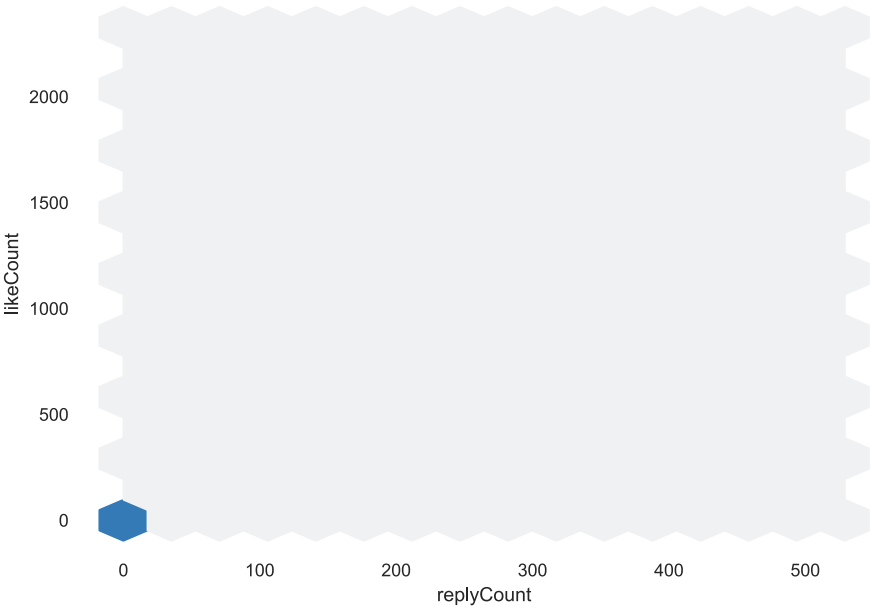


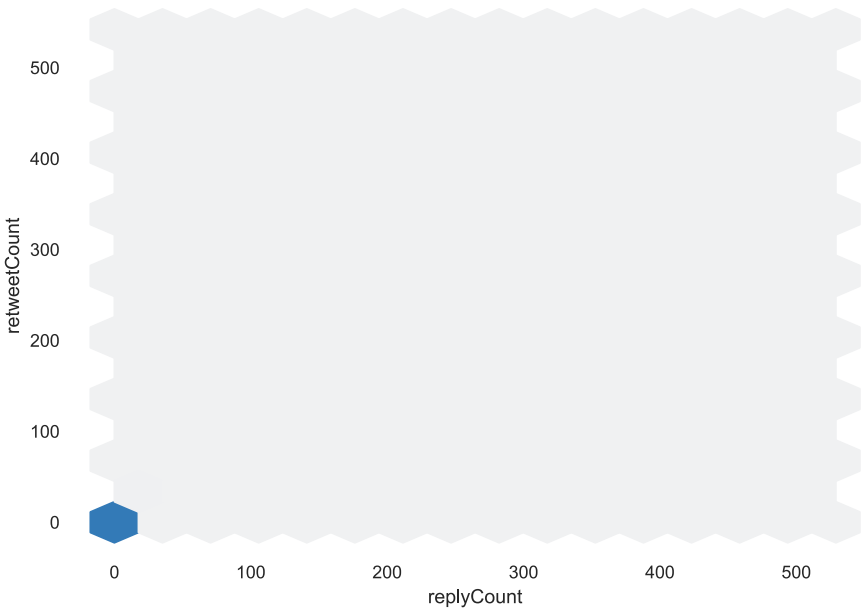




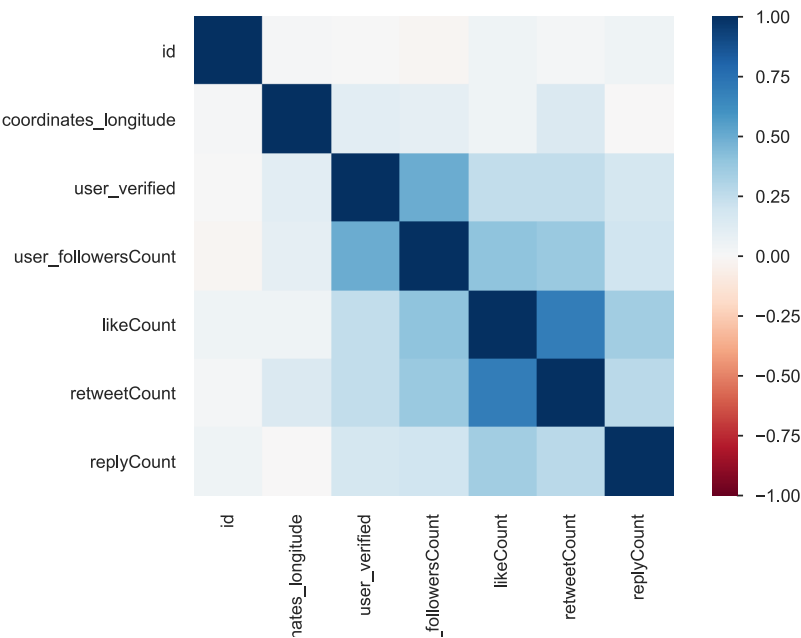


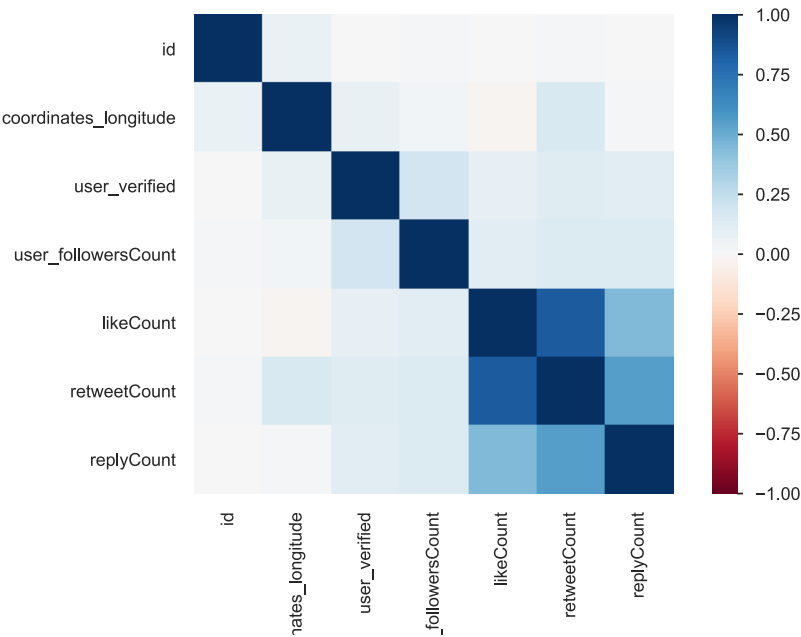


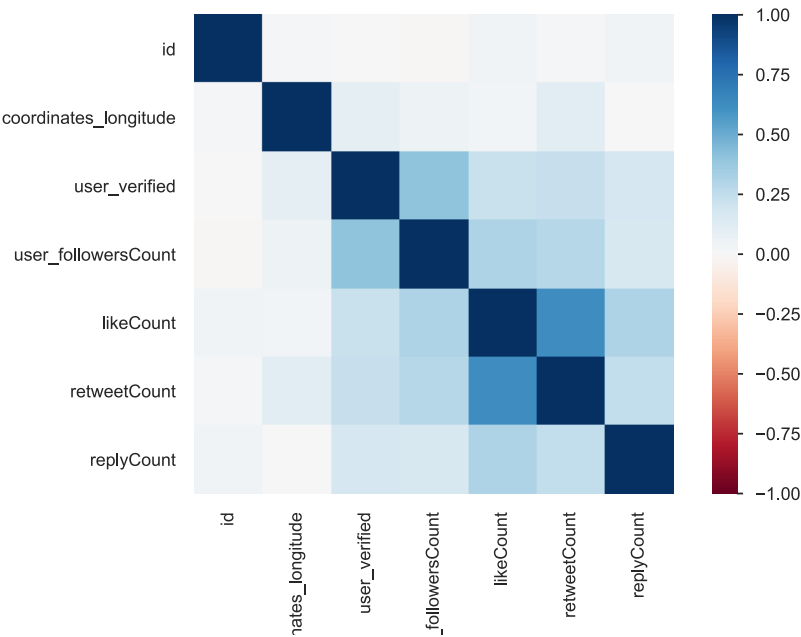


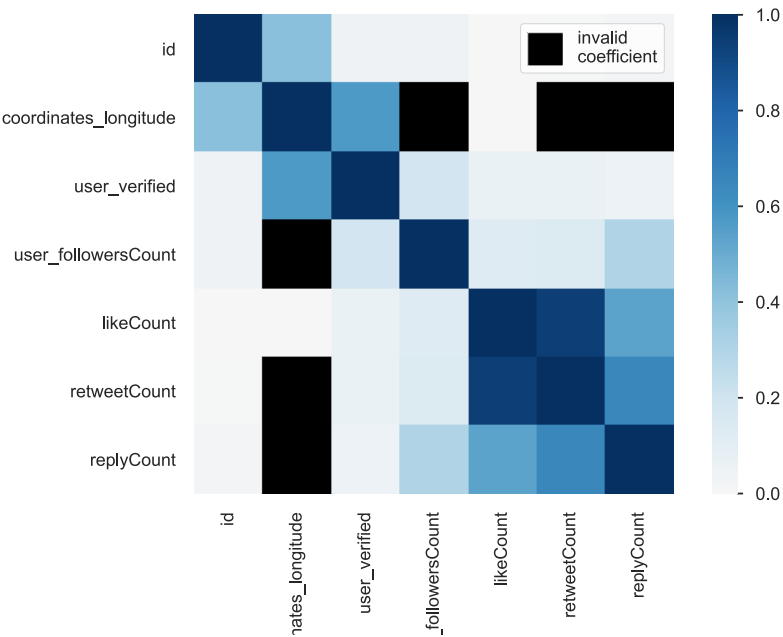


Correlations

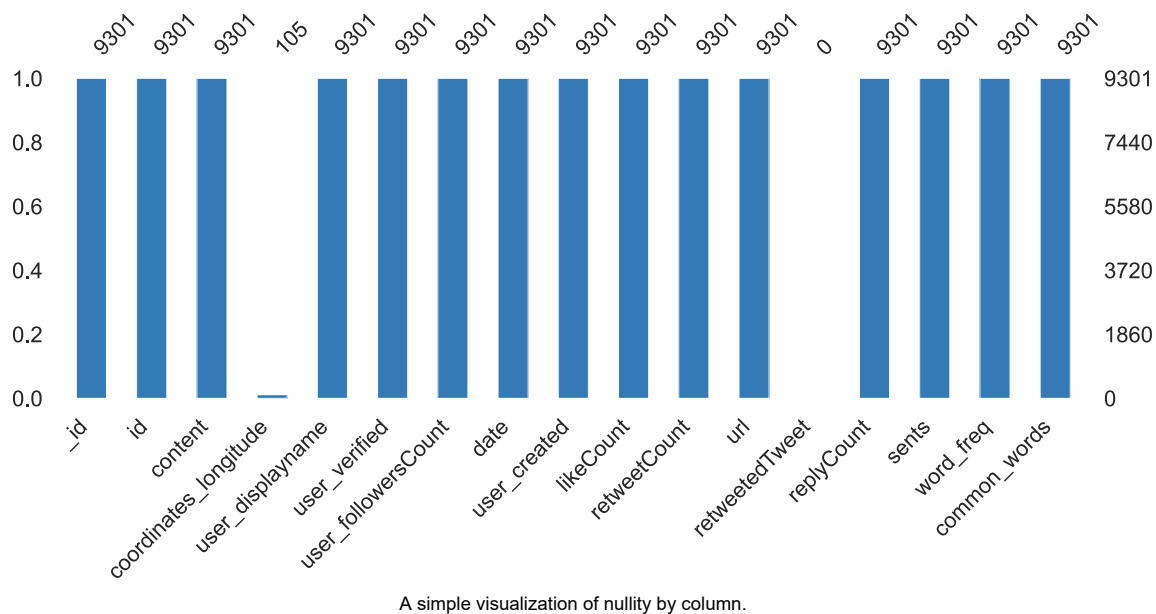


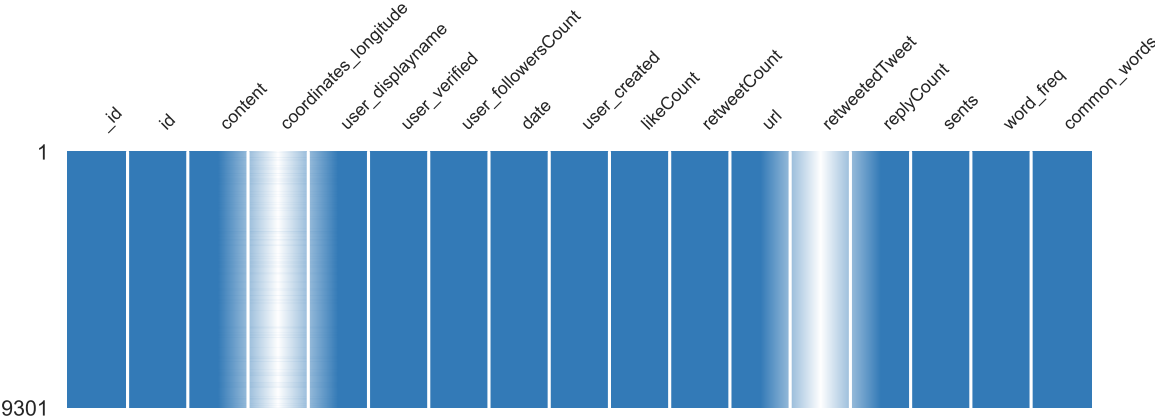




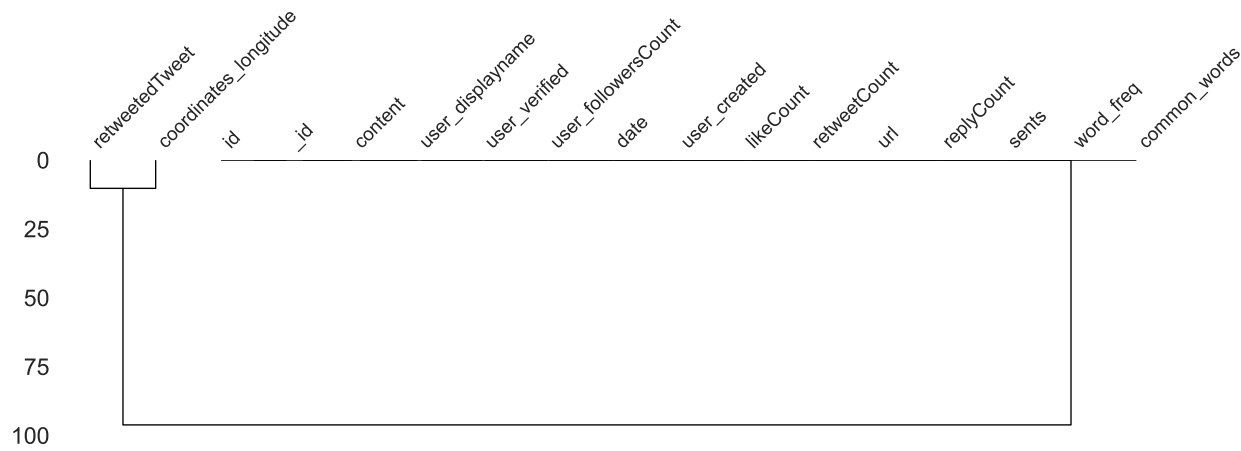


Missing values





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



The dendrogram allows you to more fully correlate variable completion, revealing trends deeper than the pairwise ones visible in the correlation heatmap.

Sample

First rows

	_id	id	content
0	61abc9ee61107103044678b6	1466914816241455104	BC teachers and support staff have risen to the challenge of the COVID-19 pandemic, even as it has taken
1	61abc9ee61107103044678eb	1466908091300003840	College students' mental health still struggling despite COVID-19 vaccines and no more lockdowns\n\nNew
2	61abc9ee6110710304467947	1466905171452170246	Half (52%) of 126 surveyed U.S. mayors identified mental health challenges that are associated with traum
3	61abcc1461107103044953ba	1466890614595801090	NOVA Chemicals is proud to have donated to COVID-19 Recovery Fund. One example of our \$160,000 gi
4	61abcc1461107103044953da	1466883989600444429	The physical effects of COVID-19 are well documented, but what about the effects on our mental health? f
5	61abcc1461107103044953fb	1466879328247947265	Is there any evidence to suggests those with significant mental health disorders (highest risk of suicide) wt
6	61abcc146110710304495422	1466878666642706440	Our Government remains committed to doing everything it can to support persons with disabilities. #Finish
7	61abcc14611071030449544b	1466877152050176000	#ArticuloDelDíaBibliotecasBUAP\n\nPsychiatric hospitalization in Korea, 2011-2020: the impact of the Mer
8	61abcc146110710304495477	1466874663317884929	IJERPH, Vol. 18, Pages 12748: Coping Strategies as a Mental Health Protection Factor of Spanish Nurses
9	61abcc1461107103044954ab	1466874394618257411	SAVE THE DATE! Thursday, December 16th, all are welcome to Healing In Conversations, a monthly foru

Last rows

	_id	id	content
9291	61aecb1611071030453565b	1463778594170277889	Covid: Cinema faces court for not checking Covid passes, and student depression increases
9292	61aecb161107103045356b7	1463778377945526275	Covid: Cinema faces court for not checking Covid passes, and student depression increases - BBC Ne
9293	61aecb161107103045356e6	1463777600724164609	Covid: Cinema faces court for not checking Covid passes, and student depression increases
9294	61aecb16110710304535725	1463776272924950535	Covid: Cinema faces court for not checking Covid passes, and student depression increases
9295	61aecb16110710304535788	1463776011883978752	BBC News - UK: Covid: Cinema faces court for not checking Covid passes, and student depression in
9296	61aecb261107103045358d5	1463775484819361797	Covid: Cinema faces court for not checking Covid passes, and student depression increases
9297	61aecb26110710304535918	1463775000243351552	Covid: Cinema faces court for not checking Covid passes, and student depression increases
9298	61aed121611071030459755b	1467175433678376969	#OmicronVarient and #coronavirus positive cases increases in India. Students don't take exam in cam
9299	61aed12161107103045975a4	1467174189513654277	#OmicronVarient and #coronavirus positive cases increases in India. Students don't take exam in cam
9300	61aed1216110710304597608	1467170911899897862	#OmicronVarient and #coronavirus positive cases increases in India. Students don't take exam in cam