Overview

Dataset statistics

Number of variables	18
Number of observations	73954
Missing cells	984628
Missing cells (%)	74.0%
Duplicate rows	6352
Duplicate rows (%)	8.6%
Total size in memory	43.3 MiB
Average record size in memory	614.1 B

Variable types

Numeric	6
Text	5
Categorical	6
DateTime	1

Alerts

Dataset has 6352 (8.6%) duplicate rows	**Duplicates**
`Sentiment` is highly overall correlated with `Sentiment_Polarity` and 1 other fields (Sentiment_Polarity, Sentiment_Subjectivity)	**High correlation**
`Sentiment_Polarity` is highly overall correlated with `Sentiment`	**High correlation**
`Sentiment_Subjectivity` is highly overall correlated with `Sentiment`	**High correlation**
`Type` is highly imbalanced (60.4%)	**Imbalance**
`Content Rating` is highly imbalanced (63.6%)	**Imbalance**
`Unnamed: 0` has 64295 (86.9%) missing values	**Missing**
`Category` has 64295 (86.9%) missing values	**Missing**
`Rating` has 65758 (88.9%) missing values	**Missing**
`Reviews` has 64295 (86.9%) missing values	**Missing**
`Size` has 65522 (88.6%) missing values	**Missing**
`Installs` has 64295 (86.9%) missing values	**Missing**
`Type` has 64295 (86.9%) missing values	**Missing**
`Price` has 64295 (86.9%) missing values	**Missing**
`Content Rating` has 64295 (86.9%) missing values	**Missing**
`Genres` has 64295 (86.9%) missing values	**Missing**
`Last Updated` has 64295 (86.9%) missing values	**Missing**
`Current Ver` has 64303 (86.9%) missing values	**Missing**
`Android Ver` has 64297 (86.9%) missing values	**Missing**
`Translated_Review` has 36527 (49.4%) missing values	**Missing**
`Sentiment` has 36522 (49.4%) missing values	**Missing**
`Sentiment_Polarity` has 36522 (49.4%) missing values	**Missing**
`Sentiment_Subjectivity` has 36522 (49.4%) missing values	**Missing**
`Reviews` is highly skewed (γ1 = 26.57345876)	**Skewed**
`Sentiment_Polarity` has 5163 (7.0%) zeros	**Zeros**
`Sentiment_Subjectivity` has 4416 (6.0%) zeros	**Zeros**

Reproduction

Analysis started	2024-09-23 10:32:35.952652
Analysis finished	2024-09-23 10:32:43.614408
Duration	7.66 seconds
Software version	ydata-profiling vv4.10.0 (https://github.com/ydataai/ydata-profiling)
Download configuration	config.json (data:text/plain;charset=utf-8,%7B%22title%22%3A%20%22p4%20Profiling%20Report%22%2C%20%22dataset%22%3A%20%7B%22description%22%3A%20%22%22%2C%20%22creator%22%3A%20%22%22%2C%20%22author%22%3A%20%22%22%2C

Variables

Select Columns

Unnamed: 0
Real number (ℝ)
MISSING

Distinct	9659
Distinct (%)	100.0%
Missing	64295
Missing (%)	86.9%
Infinite	0
Infinite (%)	0.0%
Mean	5666.1729
Minimum	0
Maximum	10840
Zeros	1
Zeros (%)	< 0.1%
Negative	0
Negative (%)	0.0%
Memory size	577.9 KiB
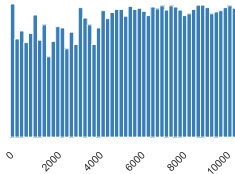
Quantile statistics

Minimum	0
5-th percentile	577.9
Q1	3111.5
median	5814
Q3	8327.5
95-th percentile	10347.1
Maximum	10840
Range	10840
Interquartile range (IQR)	5216

Descriptive statistics

Standard deviation	3102.3629
Coefficient of variation (CV)	0.54752351
Kurtosis	-1.1422172
Mean	5666.1729
Median Absolute Deviation (MAD)	2596
Skewness	-0.11950732
Sum	54729564
Variance	9624655.3
Monotonicity	Strictly increasing

Histogram with fixed size bins (bins=50)
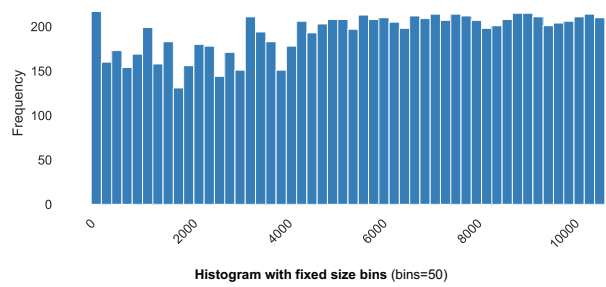
Value	Count	Frequency (%)
7483	1	< 0.1%
7485	1	< 0.1%
7486	1	< 0.1%
7487	1	< 0.1%
7488	1	< 0.1%
7489	1	< 0.1%
7490	1	< 0.1%
7491	1	< 0.1%
7492	1	< 0.1%
7493	1	< 0.1%
Other values (9649)	9649	13.0%
(Missing)	64295	86.9%

Value	Count	Frequency (%)
0	1	< 0.1%
1	1	< 0.1%
2	1	< 0.1%
3	1	< 0.1%
4	1	< 0.1%
5	1	< 0.1%
6	1	< 0.1%
7	1	< 0.1%
8	1	< 0.1%
9	1	< 0.1%

Value	Count	Frequency (%)
10840	1	< 0.1%
10839	1	< 0.1%
10838	1	< 0.1%
10837	1	< 0.1%
10836	1	< 0.1%
10835	1	< 0.1%
10834	1	< 0.1%
10833	1	< 0.1%
10832	1	< 0.1%
10831	1	< 0.1%

App
Text

Distinct	9713
Distinct (%)	13.1%
Missing	0
Missing (%)	0.0%
Memory size	6.2 MiB

Length

Max length	194
Median length	81
Mean length	23.8528
Min length	1

Characters and Unicode

Total characters	1764010	
Distinct characters	556	
Distinct categories	18 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	16 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	27 (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	8639	?
Unique (%)	11.7%	

Sample

1st row	Photo Editor & Candy Camera & Grid & ScrapBook
2nd row	Coloring book moana
3rd row	U Launcher Lite – FREE Live Cool Themes, Hide Apps
4th row	Sketch - Draw & Paint
5th row	Pixel Draw - Number Art Coloring Book

Value	Count	Frequency (%)
	25154	8.4%
free	5611	1.9%
for	3197	1.1%
app	3013	1.0%
news	2662	0.9%
google	2622	0.9%
dating	2412	0.8%
live	2233	0.7%
a	2192	0.7%
camera	2033	0.7%
Other values (9624)	247325	82.9%

Most occurring characters

Value	Count	Frequency (%)
	224499	12.7%
e	147089	8.3%
a	114385	6.5%
o	106849	6.1%
r	98108	5.6%
i	85594	4.9%
t	77110	4.4%
l	72856	4.1%
s	68609	3.9%
n	68535	3.9%
Other values (546)	700376	39.7%

Most occurring characters

Value	Count	Frequency (%)
	224499	12.7%
e	147089	8.3%
a	114385	6.5%
o	106849	6.1%
r	98108	5.6%
i	85594	4.9%
t	77110	4.4%
l	72856	4.1%
s	68609	3.9%
n	68535	3.9%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	1157782	65.6%
Uppercase Letter	303400	17.2%
Space Separator	224500	12.7%
Other Punctuation	34058	1.9%
Dash Punctuation	20583	1.2%
Decimal Number	13252	0.8%
Other Letter	4649	0.3%
Other Symbol	1851	0.1%
Math Symbol	1596	0.1%
Close Punctuation	1062	0.1%
Other values (8)	1277	0.1%

Most frequent character per category

Other Letter

Value	Count	Frequency (%)
房	278	6.0%
找	160	3.4%
屋	159	3.4%
料	121	2.6%
工	120	2.6%
無	81	1.7%
履	80	1.7%
新	80	1.7%
歷	80	1.7%
作	80	1.7%
Other values (278)	3410	73.3%

Lowercase Letter

Value	Count	Frequency (%)
e	147089	12.7%
a	114385	9.9%
o	106849	9.2%
r	98108	8.5%
i	85594	7.4%
t	77110	6.7%
l	72856	6.3%
s	68609	5.9%
n	68535	5.9%
c	34299	3.0%
Other values (80)	284348	24.6%

Uppercase Letter

Value	Count	Frequency (%)
C	33732	11.1%
S	26194	8.6%
F	25298	8.3%
A	23051	7.6%
B	22843	7.5%
D	20976	6.9%
P	15516	5.1%
G	15226	5.0%
E	14360	4.7%
T	13819	4.6%
Other values (42)	92385	30.4%

Other Symbol

Value	Count	Frequency (%)
™	786	42.5%
®	472	25.5%

♥		162	8.8%
		86	4.6%
✔		82	4.4%
°		44	2.4%
		41	2.2%
		41	2.2%
		41	2.2%
😊		41	2.2%
Other values (34)		55	3.0%

Other Punctuation

Value	Count	Frequency (%)
,	9979	29.3%
&	9676	28.4%
:	9159	26.9%
.	2411	7.1%
'	1013	3.0%
!	869	2.6%
`	323	0.9%
/	285	0.8%
#	91	0.3%
?	79	0.2%
Other values (11)	173	0.5%

Decimal Number

Value	Count	Frequency (%)
2	2865	21.6%
0	2488	18.8%
1	2279	17.2%
8	1487	11.2%
3	1386	10.5%
4	848	6.4%
5	798	6.0%
9	441	3.3%
6	392	3.0%
7	268	2.0%

Math Symbol

Value	Count	Frequency (%)
+	1534	96.1%
		50	3.1%
~	5	0.3%
>	2	0.1%
≡	1	0.1%
↔	1	0.1%
∞	1	0.1%
		1	0.1%
×	1	0.1%

Nonspacing Mark

Value	Count	Frequency (%)
	85	91.4%
◌	2	2.2%
◌	1	1.1%
◌	1	1.1%
◌	1	1.1%
◌	1	1.1%
◌	1	1.1%
◌	1	1.1%

Spacing Mark

Value	Count	Frequency (%)
◌	4	26.7%

	2	13.3%
ஂ	2	13.3%
௮	2	13.3%
௦	2	13.3%
௠	2	13.3%
ஂா	1	6.7%
௴	1	6.7%
௣	1	6.7%

Dash Punctuation

Value	Count	Frequency (%)
-	18182	88.3%
–	1979	9.6%
—	376	1.8%
‐	42	0.2%
-	2	< 0.1%
—	1	< 0.1%
～	1	< 0.1%

Close Punctuation

Value	Count	Frequency (%)
)	1033	97.3%
]	22	2.1%
）	3	0.3%
】	3	0.3%
」	1	0.1%

Open Punctuation

Value	Count	Frequency (%)
(	1029	97.3%
[	22	2.1%
（	3	0.3%
【	3	0.3%
「	1	0.1%

Final Punctuation

Value	Count	Frequency (%)
'	97	98.0%
»	1	1.0%
"	1	1.0%

Space Separator

Value	Count	Frequency (%)
	224499	> 99.9%
	1	< 0.1%

Modifier Letter

Value	Count	Frequency (%)
ー	8	100.0%

Connector Punctuation

Value	Count	Frequency (%)
_	2	100.0%

Initial Punctuation

Value	Count	Frequency (%)
"	1	100.0%

Modifier Symbol

Value	Count	Frequency (%)
`	1	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	1458668	82.7%
Common	298071	16.9%
Han	3578	0.2%
Cyrillic	2507	0.1%
Katakana	379	< 0.1%
Hiragana	328	< 0.1%
Hangul	236	< 0.1%
Arabic	86	< 0.1%
Inherited	85	< 0.1%
Bengali	18	< 0.1%
Other values (6)	54	< 0.1%

Most frequent character per script

Han

Value	Count	Frequency (%)
房	278	7.8%
找	160	4.5%
屋	159	4.4%
料	121	3.4%
工	120	3.4%
無	81	2.3%
履	80	2.2%
新	80	2.2%
歷	80	2.2%
作	80	2.2%
Other values (107)	2339	65.4%

Common

Value	Count	Frequency (%)
	224499	75.3%
-	18182	6.1%
,	9979	3.3%
&	9676	3.2%
:	9159	3.1%
2	2865	1.0%
0	2488	0.8%
.	2411	0.8%
1	2279	0.8%
–	1979	0.7%
Other values (100)	14554	4.9%

Latin

Value	Count	Frequency (%)
e	147089	10.1%
a	114385	7.8%
o	106849	7.3%
r	98108	6.7%
i	85594	5.9%
t	77110	5.3%
l	72856	5.0%
s	68609	4.7%
n	68535	4.7%
c	34299	2.4%
Other values (91)	585234	40.1%

Hangul

Value	Count	Frequency (%)
일	41	17.4%
음	41	17.4%

	40	16.9%
메	40	16.9%
다	40	16.9%
인	6	2.5%
카	4	1.7%
방	3	1.3%
송	3	1.3%
포	2	0.8%
용	2	0.8%
Other values (49)	54	22.9%

Katakana

Value	Count	Frequency (%)
レ	43	11.3%
ル	43	11.3%
リ	43	11.3%
ジ	42	11.1%
ナ	41	10.8%
シ	41	10.8%
オ	40	10.6%
ピ	40	10.6%
イ	6	1.6%
ゲ	5	1.3%
Other values (26)	35	9.2%

Cyrillic

Value	Count	Frequency (%)
и	242	9.7%
а	241	9.6%
т	239	9.5%
с	201	8.0%
р	159	6.3%
у	158	6.3%
е	124	4.9%
в	121	4.8%
ь	117	4.7%
о	83	3.3%
Other values (25)	822	32.8%

Arabic

Value	Count	Frequency (%)
ا	13	15.1%
ر	9	10.5%
ل	7	8.1%
ب	5	5.8%
ق	5	5.8%
ح	5	5.8%
ف	4	4.7%
ع	4	4.7%
ن	4	4.7%
ي	4	4.7%
Other values (16)	26	30.2%

Ethiopic

Value	Count	Frequency (%)
ከ	1	7.7%
መ	1	7.7%
ኖ	1	7.7%
ና	1	7.7%
ች	1	7.7%
ሎ	1	7.7%
ስ	1	7.7%
ዋ	1	7.7%

		1	7.7%
౪		1	7.7%
౪		1	7.7%
Other values (3)		3	23.1%

Hiragana

Value		Count	Frequency (%)
で		80	24.4%
く		41	12.5%
を		41	12.5%
が		40	12.2%
み		40	12.2%
に		40	12.2%
し		40	12.2%
け		2	0.6%
な		1	0.3%
む		1	0.3%
Other values (2)		2	0.6%

Bengali

Value		Count	Frequency (%)
া		4	22.2%
falseি		2	11.1%
ব		2	11.1%
়়		2	11.1%
ল		2	11.1%
ট		1	5.6%
ড		1	5.6%
প		1	5.6%
়়		1	5.6%
র		1	5.6%

Khmer

Value		Count	Frequency (%)
៣		2	15.4%
ន		2	15.4%
ឹ		1	7.7%
ៀ		1	7.7%
ង		1	7.7%
ំ		1	7.7%
ម		1	7.7%
ក		1	7.7%
គ		1	7.7%
ល		1	7.7%

Devanagari

Value		Count	Frequency (%)
़्		2	18.2%
र		2	18.2%
ा		2	18.2%
म		1	9.1%
ग		1	9.1%
ो		1	9.1%
त		1	9.1%
प		1	9.1%

Greek

Value		Count	Frequency (%)
σ		2	28.6%
Σ		1	14.3%
β		1	14.3%
A		1	14.3%
E		1	14.3%
K		1	14.3%

Myanmar

Value	Count	Frequency (%)
ဗ	1	20.0%
သ	1	20.0%
ပ	1	20.0%
င	1	20.0%
၃	1	20.0%

Hebrew

Value	Count	Frequency (%)
ו	1	20.0%
ק	1	20.0%
פ	1	20.0%
ה	1	20.0%
ת	1	20.0%

Inherited

Value	Count	Frequency (%)
	85	100.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	1751188	99.3%
CJK	3578	0.2%
Cyrillic	2507	0.1%
Punctuation	2463	0.1%
None	1822	0.1%
Letterlike Symbols	787	< 0.1%
Katakana	427	< 0.1%
Hiragana	328	< 0.1%
Hangul	236	< 0.1%
Misc Symbols	169	< 0.1%
Other values (17)	505	< 0.1%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
	224499	12.8%
e	147089	8.4%
a	114385	6.5%
o	106849	6.1%
r	98108	5.6%
i	85594	4.9%
t	77110	4.4%
l	72856	4.2%
s	68609	3.9%
n	68535	3.9%
Other values (77)	687554	39.3%

Punctuation

Value	Count	Frequency (%)
–	1979	80.3%
—	376	15.3%
'	97	3.9%
•	5	0.2%
-	2	0.1%
—	1	< 0.1%
"	1	< 0.1%
"	1	< 0.1%
‼	1	< 0.1%

Letterlike Symbols

Value	Count	Frequency (%)
™	786	99.9%
℠	1	0.1%

None

Value	Count	Frequency (%)
®	472	25.9%
`	323	17.7%
é	232	12.7%
	86	4.7%
á	59	3.2%
ú	46	2.5%
ó	44	2.4%
°	44	2.4%
–	42	2.3%
ã	42	2.3%
Other values (82)	432	23.7%

CJK

Value	Count	Frequency (%)

房	278	7.8%
找	160	4.5%
屋	159	4.4%
料	121	3.4%
工	120	3.4%
無	81	2.3%
履	80	2.2%
新	80	2.2%
歴	80	2.2%
作	80	2.2%
Other values (107)	2339	65.4%

Cyrillic

Value	Count	Frequency (%)
и	242	9.7%
а	241	9.6%
т	239	9.5%
с	201	8.0%
р	159	6.3%
у	158	6.3%
е	124	4.9%
в	121	4.8%
ь	117	4.7%
о	83	3.3%
Other values (25)	822	32.8%

Misc Symbols

Value	Count	Frequency (%)
♥	162	95.9%
★	5	3.0%
♪	2	1.2%

VS

Value	Count	Frequency (%)
	85	100.0%

Dingbats

Value	Count	Frequency (%)
✔	82	91.1%
❤	5	5.6%
❧	3	3.3%

Hiragana

Value	Count	Frequency (%)
で	80	24.4%
く	41	12.5%
を	41	12.5%
が	40	12.2%
み	40	12.2%
に	40	12.2%
し	40	12.2%
け	2	0.6%
な	1	0.3%
む	1	0.3%
Other values (2)	2	0.6%

Katakana

Value	Count	Frequency (%)
レ	43	10.1%
ル	43	10.1%
リ	43	10.1%
ジ	42	9.8%
ナ	41	9.6%

	41	9.6%
シ	41	9.6%
オ	40	9.4%
・	40	9.4%
ピ	40	9.4%
ー	8	1.9%
Other values (28)	46	10.8%

Hangul

Value	Count	Frequency (%)
일	41	17.4%
음	41	17.4%
메	40	16.9%
다	40	16.9%
인	6	2.5%
카	4	1.7%
방	3	1.3%
송	3	1.3%
포	2	0.8%
용	2	0.8%
Other values (49)	54	22.9%

Emoticons

Value	Count	Frequency (%)
😄	41	91.1%
😃	1	2.2%
😆	1	2.2%
😊	1	2.2%
😉	1	2.2%

Latin Ext Additional

Value	Count	Frequency (%)
ứ	40	32.5%
ọ	40	32.5%
Ớ	40	32.5%
ệ	2	1.6%
ậ	1	0.8%

Arabic

Value	Count	Frequency (%)
ا	13	15.1%
ر	9	10.5%
ل	7	8.1%
ب	5	5.8%
ق	5	5.8%
ح	5	5.8%
ف	4	4.7%
ع	4	4.7%
ن	4	4.7%
ي	4	4.7%
Other values (16)	26	30.2%

Bengali

Value	Count	Frequency (%)
ো	4	22.2%
কি	2	11.1%
ব	2	11.1%
োং	2	11.1%
ল	2	11.1%
ট	1	5.6%
ও	1	5.6%
প	1	5.6%
ং	1	5.6%
র	1	5.6%

IPA Ext

Value	Count	Frequency (%)
ə	2	100.0%

Khmer

Value	Count	Frequency (%)
ក	2	15.4%
ន	2	15.4%
ា	1	7.7%
េ្	1	7.7%
៘	1	7.7%
ៀ	1	7.7%
ធ	1	7.7%
ត	1	7.7%
ស	1	7.7%
ឈ	1	7.7%

Devanagari

Value	Count	Frequency (%)
्	2	18.2%
र	2	18.2%
ा	2	18.2%
म	1	9.1%
ग	1	9.1%
ो	1	9.1%
त	1	9.1%
प	1	9.1%

Geometric Shapes

Value	Count	Frequency (%)
◆	2	66.7%
►	1	33.3%

Myanmar

Value	Count	Frequency (%)
ိ	1	20.0%
သ	1	20.0%
ိ	1	20.0%
င	1	20.0%
ု	1	20.0%

Hebrew

Value	Count	Frequency (%)
ו	1	20.0%
ק	1	20.0%
פ	1	20.0%
ה	1	20.0%
ת	1	20.0%

Misc Technical

Value	Count	Frequency (%)
	1	100.0%

Math Operators

Value	Count	Frequency (%)
≡	1	50.0%
∞	1	50.0%

Arrows

Value	Count	Frequency (%)
↔	1	100.0%

Ethiopic

Value	Count	Frequency (%)

		1	7.7%
ሐ		1	7.7%
መ		1	7.7%
ዮ		1	7.7%
ፍ		1	7.7%
ቅ		1	7.7%
ዱ		1	7.7%
ሰ		1	7.7%
ዌ		1	7.7%
ፐ		1	7.7%
ያ		1	7.7%
Other values (3)		3	23.1%

Enclosed Alphanum Sup

Value	Count	Frequency (%)
⬚	1	50.0%
⬚	1	50.0%

Category

Categorical

`MISSING`

Distinct	33
Distinct (%)	0.3%
Missing	64295
Missing (%)	86.9%
Memory size	4.5 MiB

Length

Max length	19
Median length	16
Mean length	9.0314732
Min length	4

Characters and Unicode

Total characters	87235	
Distinct characters	24	
Distinct categories	2 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	2 (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	0	?
Unique (%)	0.0%	

Sample

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

Common Values

Value	Count	Frequency (%)
FAMILY	1832	2.5%
GAME	959	1.3%
TOOLS	827	1.1%
BUSINESS	420	0.6%
MEDICAL	395	0.5%
PERSONALIZATION	376	0.5%
PRODUCTIVITY	374	0.5%
LIFESTYLE	369	0.5%
FINANCE	345	0.5%
SPORTS	325	0.4%
Other values (23)	3437	4.6%
(Missing)	64295	86.9%

Length

Histogram of lengths of the category
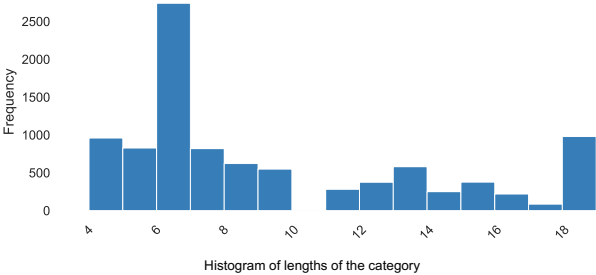
Value	Count	Frequency (%)
family	1832	19.0%
game	959	9.9%
tools	827	8.6%
business	420	4.3%
medical	395	4.1%
personalization	376	3.9%
productivity	374	3.9%
lifestyle	369	3.8%
finance	345	3.6%
sports	325	3.4%
Other values (23)	3437	35.6%

Most occurring characters

Value	Count	Frequency (%)
A	9258	10.6%
I	7836	9.0%
E	7124	8.2%
N	6463	7.4%
O	6411	7.3%
S	5950	6.8%
L	5684	6.5%
T	5168	5.9%
M	4517	5.2%
_	3229	3.7%
Other values (14)	25595	29.3%

Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	84006	96.3%
Connector Punctuation	3229	3.7%

Most frequent character per category

Uppercase Letter

Value	Count	Frequency (%)
A	9258	11.0%
I	7836	9.3%
E	7124	8.5%
N	6463	7.7%
O	6411	7.6%
S	5950	7.1%
L	5684	6.8%
T	5168	6.2%
M	4517	5.4%
F	3168	3.8%
Other values (13)	22427	26.7%

Connector Punctuation

Value	Count	Frequency (%)
_	3229	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	84006	96.3%
Common	3229	3.7%

Most frequent character per script

Latin

Value	Count	Frequency (%)
A	9258	11.0%
I	7836	9.3%
E	7124	8.5%
N	6463	7.7%
O	6411	7.6%
S	5950	7.1%
L	5684	6.8%
T	5168	6.2%
M	4517	5.4%
F	3168	3.8%
Other values (13)	22427	26.7%

Common

Value	Count	Frequency (%)
_	3229	100.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	87235	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
A	9258	10.6%
I	7836	9.0%
E	7124	8.2%
N	6463	7.4%
O	6411	7.3%
S	5950	6.8%
L	5684	6.5%
T	5168	5.9%
M	4517	5.2%
_	3229	3.7%
Other values (14)	25595	29.3%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	87235	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
A	9258	10.6%
I	7836	9.0%
E	7124	8.2%

Rating
Real number (ℝ)

MISSING

Distinct	39
Distinct (%)	0.5%
Missing	65758
Missing (%)	88.9%
Infinite	0
Infinite (%)	0.0%
Mean	4.173243
Minimum	1
Maximum	5
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	577.9 KiB
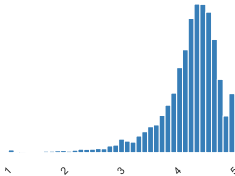
Quantile statistics

Minimum	1
5-th percentile	3.1
Q1	4
median	4.3
Q3	4.5
95-th percentile	4.8
Maximum	5
Range	4
Interquartile range (IQR)	0.5

Descriptive statistics

Standard deviation	0.53662473
Coefficient of variation (CV)	0.12858698
Kurtosis	5.1208215
Mean	4.173243
Median Absolute Deviation (MAD)	0.3
Skewness	-1.7439651
Sum	34203.9
Variance	0.2879661
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=39)
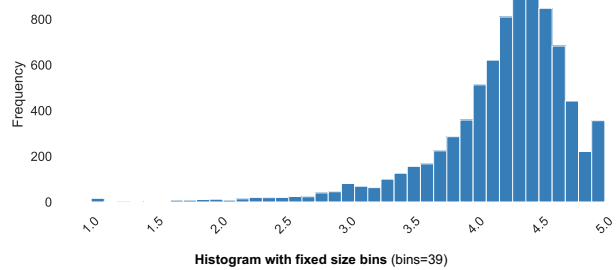
Value	Count	Frequency (%)
4.3	897	1.2%
4.4	895	1.2%
4.5	848	1.1%
4.2	810	1.1%
4.6	683	0.9%
4.1	621	0.8%
4	513	0.7%
4.7	442	0.6%
3.9	359	0.5%
3.8	286	0.4%
Other values (29)	1842	2.5%
(Missing)	65758	88.9%

Value	Count	Frequency (%)
1	16	< 0.1%
1.2	1	< 0.1%
1.4	3	< 0.1%
1.5	3	< 0.1%
1.6	4	< 0.1%
1.7	8	< 0.1%
1.8	8	< 0.1%
1.9	11	< 0.1%
2	12	< 0.1%
2.1	8	< 0.1%

Value	Count	Frequency (%)
5	271	0.4%
4.9	85	0.1%
4.8	221	0.3%
4.7	442	0.6%
4.6	683	0.9%
4.5	848	1.1%
4.4	895	1.2%
4.3	897	1.2%
4.2	810	1.1%
4.1	621	0.8%

Reviews
Real number (ℝ)

`MISSING  SKEWED`

Distinct	5330
Distinct (%)	55.2%
Missing	64295
Missing (%)	86.9%
Infinite	0
Infinite (%)	0.0%
Mean	216592.56
Minimum	0
Maximum	78158306
Zeros	593
Zeros (%)	0.8%
Negative	0
Negative (%)	0.0%
Memory size	577.9 KiB

Quantile statistics

Minimum	0
5-th percentile	0
Q1	25
median	967
Q3	29401
95-th percentile	641881.5
Maximum	78158306
Range	78158306
Interquartile range (IQR)	29376

Descriptive statistics

Standard deviation	1831319.6
Coefficient of variation (CV)	8.4551359
Kurtosis	907.62244
Mean	216592.56
Median Absolute Deviation (MAD)	967
Skewness	26.573459
Sum	2.0920676×10^9
Variance	3.3537314×10^{12}
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=50)
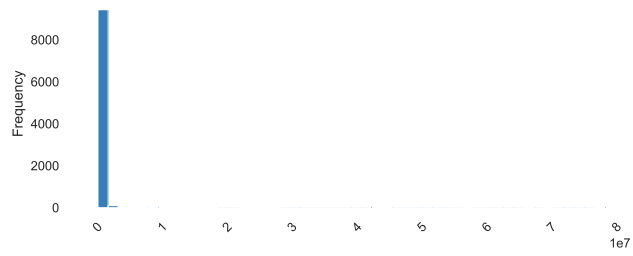
Value	Count	Frequency (%)
0	593	0.8%
1	272	0.4%
2	213	0.3%
3	170	0.2%
4	137	0.2%
5	108	0.1%
6	94	0.1%
7	88	0.1%
8	72	0.1%
9	64	0.1%
Other values (5320)	7848	10.6%
(Missing)	64295	86.9%

Value	Count	Frequency (%)
0	593	0.8%
1	272	0.4%
2	213	0.3%
3	170	0.2%
4	137	0.2%
5	108	0.1%
6	94	0.1%
7	88	0.1%
8	72	0.1%
9	64	0.1%

Value	Count	Frequency (%)
78158306	1	< 0.1%
69119316	1	< 0.1%
66577313	1	< 0.1%
56642847	1	< 0.1%
44891723	1	< 0.1%
42916526	1	< 0.1%
27722264	1	< 0.1%
25655305	1	< 0.1%
24900999	1	< 0.1%
23133508	1	< 0.1%

Size
Real number (ℝ)

MISSING

Distinct	191
Distinct (%)	2.3%
Missing	65522
Missing (%)	88.6%
Infinite	0
Infinite (%)	0.0%
Mean	20.395327
Minimum	0
Maximum	100
Zeros	25
Zeros (%)	< 0.1%
Negative	0
Negative (%)	0.0%
Memory size	577.9 KiB
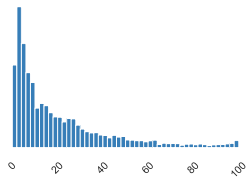
Quantile statistics

Minimum	0
5-th percentile	1.3
Q1	4.6
median	12
Q3	28
95-th percentile	70
Maximum	100
Range	100
Interquartile range (IQR)	23.4

Descriptive statistics

Standard deviation	21.827509
Coefficient of variation (CV)	1.0702211
Kurtosis	2.3378992
Mean	20.395327
Median Absolute Deviation (MAD)	9
Skewness	1.6458211
Sum	171973.4
Variance	476.44017
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=50)
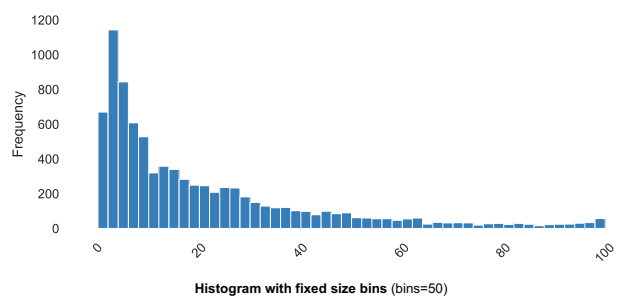
Value	Count	Frequency (%)
11	182	0.2%
12	181	0.2%
13	177	0.2%
14	177	0.2%
15	163	0.2%
17	146	0.2%
26	142	0.2%
10	138	0.2%
16	137	0.2%
19	130	0.2%
Other values (181)	6859	9.3%
(Missing)	65522	88.6%

Value	Count	Frequency (%)
0	25	< 0.1%
0.1	36	< 0.1%
0.2	48	0.1%
0.3	31	< 0.1%
0.4	28	< 0.1%
0.5	27	< 0.1%
0.6	25	< 0.1%
0.7	24	< 0.1%
0.8	22	< 0.1%
0.9	30	< 0.1%

Value	Count	Frequency (%)
100	14	< 0.1%
99	30	< 0.1%
98	13	< 0.1%
97	11	< 0.1%
96	23	< 0.1%
95	15	< 0.1%
94	15	< 0.1%
93	13	< 0.1%
92	12	< 0.1%
91	18	< 0.1%

Installs
Categorical

MISSING

Distinct	21
Distinct (%)	0.2%
Missing	64295
Missing (%)	86.9%
Memory size	4.5 MiB

Length

Max length	14
Median length	11
Mean length	7.6002692
Min length	1

Characters and Unicode

Total characters	73411	
Distinct characters	5	
Distinct categories	3 (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	1	?
Unique (%)	< 0.1%	

Sample

1st row	10,000+
2nd row	500,000+
3rd row	5,000,000+
4th row	50,000,000+
5th row	100,000+

Common Values

Value	Count	Frequency (%)
1,000,000+	1417	1.9%
100,000+	1112	1.5%
10,000+	1031	1.4%
10,000,000+	937	1.3%
1,000+	888	1.2%
100+	710	1.0%
5,000,000+	607	0.8%
500,000+	505	0.7%
50,000+	469	0.6%
5,000+	468	0.6%
Other values (11)	1515	2.0%
(Missing)	64295	86.9%

Length

Histogram of lengths of the category
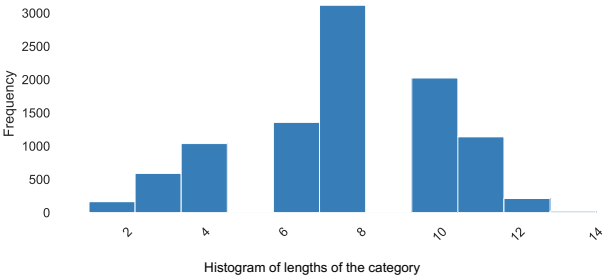
Value	Count	Frequency (%)
1,000,000	1417	14.7%
100,000	1112	11.5%
10,000	1031	10.7%
10,000,000	937	9.7%
1,000	888	9.2%
100	710	7.4%
5,000,000	607	6.3%
500,000	505	5.2%
50,000	469	4.9%
5,000	468	4.8%
Other values (10)	1515	15.7%

Most occurring characters

Value	Count	Frequency (%)
0	42826	58.3%
,	11283	15.4%
+	9658	13.2%
1	6755	9.2%
5	2889	3.9%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	52470	71.5%
Other Punctuation	11283	15.4%
Math Symbol	9658	13.2%

Most frequent character per category

Decimal Number

Value	Count	Frequency (%)
0	42826	81.6%
1	6755	12.9%
5	2889	5.5%

Other Punctuation

Value	Count	Frequency (%)
,	11283	100.0%

Math Symbol

Value	Count	Frequency (%)
+	9658	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Common	73411	100.0%

Most frequent character per script

Common

Value	Count	Frequency (%)
0	42826	58.3%
,	11283	15.4%
+	9658	13.2%
1	6755	9.2%
5	2889	3.9%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	73411	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
0	42826	58.3%
,	11283	15.4%
+	9658	13.2%
1	6755	9.2%
5	2889	3.9%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	73411	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)

Type
Categorical

IMBALANCE MISSING

Distinct	2
Distinct (%)	< 0.1%
Missing	64295
Missing (%)	86.9%
Memory size	4.5 MiB

Length

Max length	4
Median length	4
Mean length	4
Min length	4

Characters and Unicode

Total characters	38636	
Distinct characters	7	
Distinct categories	2 (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	0	?
Unique (%)	0.0%	

Sample

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

Common Values

Value	Count	Frequency (%)
Free	8903	12.0%
Paid	756	1.0%
(Missing)	64295	86.9%

Length

Histogram of lengths of the category

Common Values (Plot)

92.2%
(8903)
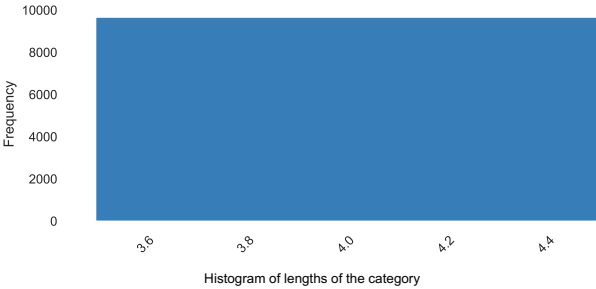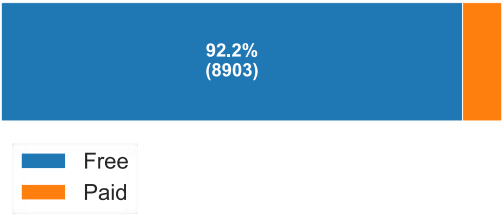
Free
Paid

Value	Count	Frequency (%)
free	8903	92.2%
paid	756	7.8%

Most occurring characters

Value	Count	Frequency (%)
e	17806	46.1%
F	8903	23.0%
r	8903	23.0%
P	756	2.0%
a	756	2.0%
i	756	2.0%
d	756	2.0%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	28977	75.0%
Uppercase Letter	9659	25.0%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
e	17806	61.4%
r	8903	30.7%
a	756	2.6%
i	756	2.6%
d	756	2.6%

Uppercase Letter

Value	Count	Frequency (%)
F	8903	92.2%
P	756	7.8%

Most occurring categories

Most frequent character per category

Most occurring scripts

Value	Count	Frequency (%)
Latin	38636	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
e	17806	46.1%
F	8903	23.0%
r	8903	23.0%
P	756	2.0%
a	756	2.0%
i	756	2.0%
d	756	2.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	38636	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
e	17806	46.1%
F	8903	23.0%
r	8903	23.0%
P	756	2.0%
a	756	2.0%
i	756	2.0%
d	756	2.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	38636	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
e	17806	46.1%
F	8903	23.0%
r	8903	23.0%

Price
Text

`MISSING`

Distinct	92
Distinct (%)	1.0%
Missing	64295
Missing (%)	86.9%
Memory size	2.5 MiB

Length

Max length	7
Median length	1
Mean length	1.3230148
Min length	1

Characters and Unicode

Total characters	12779	
Distinct characters	12	
Distinct categories	3 (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.
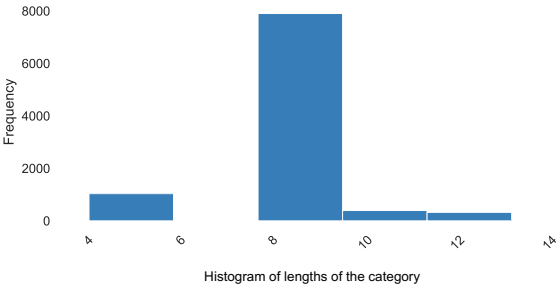
Unique

Unique	57	?
Unique (%)	0.6%	

Sample

1st row	0
2nd row	0
3rd row	0
4th row	0
5th row	0

Value	Count	Frequency (%)
0	8903	92.2%
0.99	145	1.5%
2.99	124	1.3%
1.99	73	0.8%
4.99	70	0.7%
3.99	57	0.6%
1.49	46	0.5%
5.99	26	0.3%
2.49	25	0.3%
9.99	19	0.2%
Other values (82)	171	1.8%

Most occurring characters

Value	Count	Frequency (%)
0	9100	71.2%
9	1389	10.9%
$	756	5.9%
.	756	5.9%
4	210	1.6%
2	180	1.4%
1	179	1.4%
3	95	0.7%
5	46	0.4%
6	28	0.2%
Other values (2)	40	0.3%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	11267	88.2%
Currency Symbol	756	5.9%
Other Punctuation	756	5.9%

Most frequent character per category

Decimal Number

Value	Count	Frequency (%)
0	9100	80.8%
9	1389	12.3%
4	210	1.9%
2	180	1.6%
1	179	1.6%
3	95	0.8%
5	46	0.4%
6	28	0.2%
7	22	0.2%
8	18	0.2%

Currency Symbol

Value	Count	Frequency (%)
$	756	100.0%

Other Punctuation

Value	Count	Frequency (%)
.	756	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Common	12779	100.0%

Most frequent character per script

Common

Value	Count	Frequency (%)
0	9100	71.2%
9	1389	10.9%
$	756	5.9%
.	756	5.9%
4	210	1.6%
2	180	1.4%
1	179	1.4%
3	95	0.7%
5	46	0.4%
6	28	0.2%
Other values (2)	40	0.3%

Most occurring scripts

Value	Count	Frequency (%)
Common	12779	100.0%

Most frequent character per script

Common

Value	Count	Frequency (%)
	9100	71.2%
	1389	10.9%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	12779	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
0	9100	71.2%
9	1389	10.9%
$	756	5.9%
.	756	5.9%
4	210	1.6%
2	180	1.4%
1	179	1.4%
3	95	0.7%
5	46	0.4%
6	28	0.2%
Other values (2)	40	0.3%

Content Rating
Categorical

Distinct	6
Distinct (%)	0.1%
Missing	64295
Missing (%)	86.9%
Memory size	4.5 MiB

Length

Max length	15
Median length	8
Mean length	7.7876592
Min length	4

Characters and Unicode

Total characters	75221	
Distinct characters	23	
Distinct categories	5 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	2 (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	0	?
Unique (%)	0.0%	

Sample

1st row	Everyone
2nd row	Everyone
3rd row	Everyone
4th row	Teen
5th row	Everyone

Common Values

Value	Count	Frequency (%)
Everyone	7903	10.7%
Teen	1036	1.4%
Mature 17+	393	0.5%
Everyone 10+	322	0.4%
Adults only 18+	3	< 0.1%
Unrated	2	< 0.1%
(Missing)	64295	86.9%

Length

Histogram of lengths of the category

Common Values (Plot)

- Everyone
- Teen
- Mature 17+
- Everyone 10+
- Adults only 18+
- Unrated
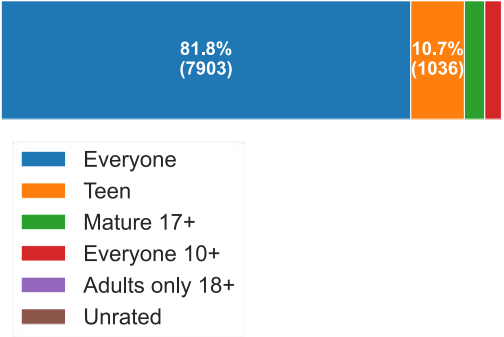
Value	Count	Frequency (%)
everyone	8225	79.2%
teen	1036	10.0%
mature	393	3.8%
17	393	3.8%
10	322	3.1%
adults	3	< 0.1%
only	3	< 0.1%
18	3	< 0.1%
unrated	2	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
e	18917	25.1%
n	9266	12.3%
r	8620	11.5%
y	8228	10.9%
o	8228	10.9%
E	8225	10.9%
v	8225	10.9%
T	1036	1.4%
	721	1.0%
+	718	1.0%
Other values (13)	3037	4.0%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	62687	83.3%
Uppercase Letter	9659	12.8%
Decimal Number	1436	1.9%
Space Separator	721	1.0%
Math Symbol	718	1.0%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
e	18917	30.2%
n	9266	14.8%
r	8620	13.8%
y	8228	13.1%
o	8228	13.1%
v	8225	13.1%
t	398	0.6%
u	396	0.6%
a	395	0.6%
l	6	< 0.1%
Other values (2)	8	< 0.1%

Uppercase Letter

Value	Count	Frequency (%)
E	8225	85.2%
T	1036	10.7%
M	393	4.1%
A	3	< 0.1%
U	2	< 0.1%

Decimal Number

Value	Count	Frequency (%)
1	718	50.0%
7	393	27.4%
0	322	22.4%
8	3	0.2%

Space Separator

Value	Count	Frequency (%)
	721	100.0%

Math Symbol

Value	Count	Frequency (%)
+	718	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	72346	96.2%
Common	2875	3.8%

Most frequent character per script

Latin

Value	Count	Frequency (%)
e	18917	26.1%
n	9266	12.8%
r	8620	11.9%
y	8228	11.4%
o	8228	11.4%
E	8225	11.4%
v	8225	11.4%
T	1036	1.4%
t	398	0.6%
u	396	0.5%
Other values (7)	807	1.1%

Common

Value	Count	Frequency (%)
	721	25.1%
+	718	25.0%
1	718	25.0%
7	393	13.7%
0	322	11.2%
8	3	0.1%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	75221	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
e	18917	25.1%
n	9266	12.3%
r	8620	11.5%
y	8228	10.9%
o	8228	10.9%
E	8225	10.9%
v	8225	10.9%
T	1036	1.4%
	721	1.0%
+	718	1.0%
Other values (13)	3037	4.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	75221	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)

Genres
Text

MISSING

Distinct	118
Distinct (%)	1.2%
Missing	64295
Missing (%)	86.9%
Memory size	2.6 MiB

Length

Max length	37
Median length	32
Mean length	10.371674
Min length	4

Characters and Unicode

Total characters	100180	
Distinct characters	43	
Distinct categories	4 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	2 (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	27	?
Unique (%)	0.3%	

Sample

1st row	Art & Design
2nd row	Art & Design;Pretend Play
3rd row	Art & Design
4th row	Art & Design
5th row	Art & Design;Creativity

Value	Count	Frequency (%)
	1834	13.3%
tools	826	6.0%
entertainment	561	4.1%
education	510	3.7%
business	420	3.1%
medical	395	2.9%
personalization	376	2.7%
productivity	374	2.7%
lifestyle	368	2.7%
finance	345	2.5%
Other values (124)	7731	56.3%

Most occurring characters

Value	Count	Frequency (%)
i	8705	8.7%
o	8141	8.1%
e	8060	8.0%
n	7872	7.9%
a	7637	7.6%
t	7561	7.5%
s	5752	5.7%
l	4159	4.2%
	4081	4.1%
r	4029	4.0%
Other values (33)	34183	34.1%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	81573	81.4%
Uppercase Letter	12299	12.3%
Space Separator	4081	4.1%
Other Punctuation	2227	2.2%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
i	8705	10.7%
o	8141	10.0%
e	8060	9.9%
n	7872	9.7%
a	7637	9.4%
t	7561	9.3%
s	5752	7.1%
l	4159	5.1%
r	4029	4.9%
c	3804	4.7%
Other values (13)	15853	19.4%

Uppercase Letter

Value	Count	Frequency (%)
P	1651	13.4%
E	1601	13.0%
T	1085	8.8%
S	1081	8.8%
A	937	7.6%
M	834	6.8%
B	822	6.7%
F	747	6.1%
C	718	5.8%
L	673	5.5%
Other values (7)	2150	17.5%

Other Punctuation

Value	Count	Frequency (%)
&	1834	82.4%
;	393	17.6%

Space Separator

Value	Count	Frequency (%)
	4081	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	93872	93.7%
Common	6308	6.3%

Most frequent character per script

Latin

Value	Count	Frequency (%)
i	8705	9.3%
o	8141	8.7%
e	8060	8.6%
n	7872	8.4%
a	7637	8.1%
t	7561	8.1%
s	5752	6.1%
l	4159	4.4%
r	4029	4.3%
c	3804	4.1%
Other values (30)	28152	30.0%

Common

Value	Count	Frequency (%)
	4081	64.7%
&	1834	29.1%
;	393	6.2%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	100180	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
i	8705	8.7%
o	8141	8.1%
e	8060	8.0%
n	7872	7.9%
a	7637	7.6%
t	7561	7.5%
s	5752	5.7%
l	4159	4.2%
	4081	4.1%
r	4029	4.0%
Other values (33)	34183	34.1%

Most occurring blocks

ASCII

Value	Count	Frequency (%)

Last Updated
Date

MISSING

Distinct	1377
Distinct (%)	14.3%
Missing	64295
Missing (%)	86.9%
Memory size	577.9 KiB

Minimum	2010-05-21 00:00:00
Maximum	2018-08-08 00:00:00

Histogram with fixed size bins (bins=50)

Current Ver
Text
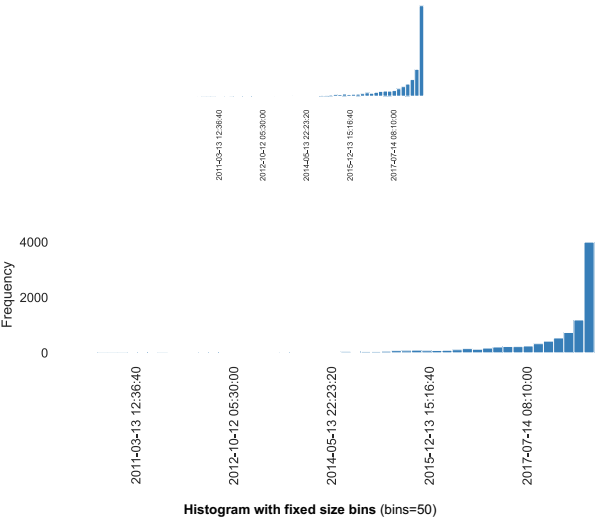
MISSING

Distinct	2769
Distinct (%)	28.7%
Missing	64303
Missing (%)	86.9%
Memory size	2.5 MiB

Length

Max length	50
Median length	39
Mean length	6.2222568
Min length	1

Characters and Unicode

Total characters	60051	
Distinct characters	79	
Distinct categories	12 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	2 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	3 (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	2011	?
Unique (%)	20.8%	

Sample

1st row	1.0.0
2nd row	2.0.0
3rd row	1.2.4
4th row	Varies with device
5th row	1.1

Value	Count	Frequency (%)
varies	1055	8.8%
device	1055	8.8%
with	1055	8.8%
1	833	7.0%
1.1	274	2.3%
1.2	183	1.5%
2	164	1.4%
1.3	142	1.2%
1.0.0	133	1.1%
1.0.1	116	1.0%
Other values (2810)	6926	58.0%

Most occurring characters

Value	Count	Frequency (%)
.	12831	21.4%
1	8228	13.7%
0	4048	6.7%
2	3842	6.4%
e	3295	5.5%
i	3251	5.4%
3	2467	4.1%
	2285	3.8%
4	1886	3.1%
5	1505	2.5%
Other values (69)	16413	27.3%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	26285	43.8%
Lowercase Letter	17016	28.3%
Other Punctuation	12844	21.4%
Space Separator	2285	3.8%
Uppercase Letter	1352	2.3%
Dash Punctuation	130	0.2%
Connector Punctuation	58	0.1%
Close Punctuation	35	0.1%
Open Punctuation	35	0.1%
Math Symbol	9	< 0.1%
Other values (2)	2	< 0.1%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
e	3295	19.4%
i	3251	19.1%
a	1170	6.9%
r	1161	6.8%
d	1147	6.7%
v	1112	6.5%
s	1102	6.5%
t	1100	6.5%
c	1096	6.4%
w	1071	6.3%
Other values (16)	1511	8.9%

Uppercase Letter

Value	Count	Frequency (%)
V	1075	79.5%
A	31	2.3%
C	23	1.7%
R	22	1.6%
B	21	1.6%
P	21	1.6%
E	18	1.3%
S	17	1.3%
T	16	1.2%
D	14	1.0%
Other values (15)	94	7.0%

Decimal Number

Value	Count	Frequency (%)
1	8228	31.3%
0	4048	15.4%
2	3842	14.6%
3	2467	9.4%
4	1886	7.2%
5	1505	5.7%
6	1259	4.8%
7	1142	4.3%
8	1025	3.9%
9	883	3.4%

Other Punctuation

Value	Count	Frequency (%)
.	12831	99.9%
,	4	< 0.1%
/	4	< 0.1%

	3	< 0.1%
:	3	< 0.1%
;	1	< 0.1%
'	1	< 0.1%

Dash Punctuation

Value	Count	Frequency (%)
-	129	99.2%
–	1	0.8%

Close Punctuation

Value	Count	Frequency (%)
)	34	97.1%
]	1	2.9%

Open Punctuation

Value	Count	Frequency (%)
(	34	97.1%
[	1	2.9%

Math Symbol

Value	Count	Frequency (%)
+	6	66.7%
|	3	33.3%

Space Separator

Value	Count	Frequency (%)
	2285	100.0%

Connector Punctuation

Value	Count	Frequency (%)
_	58	100.0%

Other Symbol

Value	Count	Frequency (%)
®	1	100.0%

Other Number

Value	Count	Frequency (%)
³	1	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Common	41683	69.4%
Latin	18368	30.6%

Most frequent character per script

Latin

Value	Count	Frequency (%)
e	3295	17.9%
i	3251	17.7%
a	1170	6.4%
r	1161	6.3%
d	1147	6.2%
v	1112	6.1%
s	1102	6.0%
t	1100	6.0%
c	1096	6.0%
V	1075	5.9%
Other values (41)	2859	15.6%

Common

Value	Count	Frequency (%)
.	12831	30.8%
1	8228	19.7%
0	4048	9.7%
2	3842	9.2%
3	2467	5.9%
	2285	5.5%
4	1886	4.5%
5	1505	3.6%
6	1259	3.0%
7	1142	2.7%
Other values (18)	2190	5.3%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	60047	> 99.9%
None	3	< 0.1%
Punctuation	1	< 0.1%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
.	12831	21.4%
1	8228	13.7%
0	4048	6.7%
2	3842	6.4%
e	3295	5.5%
i	3251	5.4%
3	2467	4.1%
	2285	3.8%
4	1886	3.1%
5	1505	2.5%
Other values (65)	16409	27.3%

Punctuation

Value	Count	Frequency (%)
–	1	100.0%

None

Value	Count	Frequency (%)
®	1	33.3%
³	1	33.3%
Ã	1	33.3%

Android Ver
Categorical

`MISSING`

Distinct	33
Distinct (%)	0.3%
Missing	64297
Missing (%)	86.9%
Memory size	4.6 MiB

Length

Max length	18
Median length	10
Mean length	11.168893
Min length	9

Characters and Unicode

Total characters	107858	
Distinct characters	28	
Distinct categories	6 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	2 (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	5	?
Unique (%)	0.1%	

Sample

1st row	4.0.3 and up
2nd row	4.0.3 and up
3rd row	4.0.3 and up
4th row	4.2 and up
5th row	4.4 and up

Common Values

Value	Count	Frequency (%)
4.1 and up	2202	3.0%
4.0.3 and up	1395	1.9%
4.0 and up	1285	1.7%
Varies with device	990	1.3%
4.4 and up	818	1.1%
2.3 and up	616	0.8%
5.0 and up	512	0.7%
4.2 and up	372	0.5%
2.3.3 and up	273	0.4%
2.2 and up	239	0.3%
Other values (23)	955	1.3%
(Missing)	64297	86.9%

Length

Histogram of lengths of the category
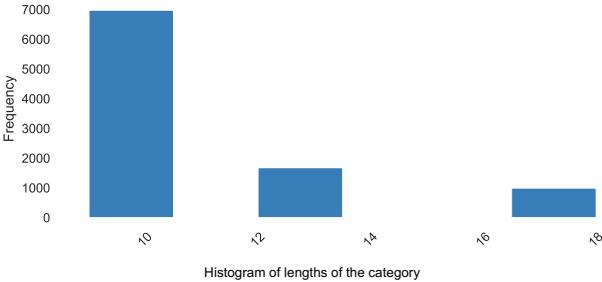
Value	Count	Frequency (%)
and	8658	29.9%
up	8658	29.9%
4.1	2203	7.6%
4.0.3	1397	4.8%
4.0	1285	4.4%
varies	990	3.4%
with	990	3.4%
device	990	3.4%
4.4	818	2.8%
2.3	616	2.1%
Other values (22)	2366	8.2%

Most occurring characters

Value	Count	Frequency (%)
	19314	17.9%
.	10359	9.6%
a	9648	8.9%
d	9648	8.9%
n	8658	8.0%
u	8658	8.0%
p	8658	8.0%
4	7136	6.6%
0	3576	3.3%
3	3057	2.8%
Other values (18)	19146	17.8%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	58140	53.9%
Space Separator	19314	17.9%
Decimal Number	19035	17.6%
Other Punctuation	10359	9.6%
Uppercase Letter	1001	0.9%
Dash Punctuation	9	< 0.1%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
a	9648	16.6%
d	9648	16.6%
n	8658	14.9%
u	8658	14.9%
p	8658	14.9%
i	2970	5.1%
e	2970	5.1%
r	990	1.7%
s	990	1.7%
w	990	1.7%
Other values (4)	3960	6.8%

Decimal Number

Value	Count	Frequency (%)
4	7136	37.5%
0	3576	18.8%
3	3057	16.1%
1	2528	13.3%
2	1949	10.2%
5	558	2.9%
6	171	0.9%
7	52	0.3%
8	8	< 0.1%

Uppercase Letter

Value	Count	Frequency (%)
V	990	98.9%
W	11	1.1%

Space Separator

Value	Count	Frequency (%)
	19314	100.0%

Other Punctuation

Value	Count	Frequency (%)
.	10359	100.0%

Dash Punctuation

Value	Count	Frequency (%)
-	9	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	59141	54.8%
Common	48717	45.2%

Most frequent character per script

Latin

Value	Count	Frequency (%)
a	9648	16.3%
d	9648	16.3%
n	8658	14.6%
u	8658	14.6%
p	8658	14.6%
i	2970	5.0%
e	2970	5.0%
r	990	1.7%
V	990	1.7%
s	990	1.7%
Other values (6)	4961	8.4%

Common

Value	Count	Frequency (%)
	19314	39.6%
.	10359	21.3%
4	7136	14.6%
0	3576	7.3%
3	3057	6.3%
1	2528	5.2%
2	1949	4.0%
5	558	1.1%
6	171	0.4%
7	52	0.1%
Other values (2)	17	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	107858	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
	19314	17.9%
.	10359	9.6%
a	9648	8.9%
d	9648	8.9%
n	8658	8.0%
u	8658	8.0%
p	8658	8.0%
4	7136	6.6%
0	3576	3.3%
3	3057	2.8%
Other values (18)	19146	17.8%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	107858	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
	19314	17.9%
.	10359	9.6%
a	9648	8.9%

Translated_Review
Text

Distinct	27994
Distinct (%)	74.8%
Missing	36527
Missing (%)	49.4%
Memory size	7.3 MiB

Length

Max length	2713
Median length	601
Mean length	113.21191
Min length	2

Characters and Unicode

Total characters	4237182	
Distinct characters	266	
Distinct categories	21 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	?
Distinct scripts	12 (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)	?
Distinct blocks	21 (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	22593	?
Unique (%)	60.4%	

Sample

1st row	I like eat delicious food. That's I'm cooking food myself, case "10 Best Foods" helps lot, also "Best Before (Shelf Life)"
2nd row	This help eating healthy exercise regular basis
3rd row	Works great especially going grocery store
4th row	Best idea us
5th row	Best way

Value	Count	Frequency (%)
i	33507	4.9%
game	9263	1.3%
it	8624	1.3%
the	6490	0.9%
like	5468	0.8%
good	5226	0.8%
this	5006	0.7%
great	4790	0.7%
app	4762	0.7%
get	4715	0.7%
Other values (25748)	598957	87.2%

Most occurring characters

Value	Count	Frequency (%)
	649381	15.3%
e	425348	10.0%
a	246382	5.8%
t	246062	5.8%
s	229870	5.4%
o	221200	5.2%
i	213536	5.0%
n	208620	4.9%
r	183720	4.3%
l	177263	4.2%
Other values (256)	1435800	33.9%

Most occurring characters

Value	Count	Frequency (%)
	649381	15.3%
e	425348	10.0%
a	246382	5.8%
t	246062	5.8%
s	229870	5.4%
o	221200	5.2%
i	213536	5.0%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	3195554	75.4%
Space Separator	649381	15.3%
Uppercase Letter	199953	4.7%
Other Punctuation	159533	3.8%
Decimal Number	22979	0.5%
Dash Punctuation	3380	0.1%
Close Punctuation	2231	0.1%
Open Punctuation	1977	< 0.1%
Currency Symbol	660	< 0.1%
Math Symbol	600	< 0.1%
Other values (11)	934	< 0.1%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
e	425348	13.3%
a	246382	7.7%
t	246062	7.7%
s	229870	7.2%
o	221200	6.9%
i	213536	6.7%
n	208620	6.5%
r	183720	5.7%
l	177263	5.5%
d	124405	3.9%
Other values (48)	919148	28.8%

Other Letter

Value	Count	Frequency (%)
ي	7	7.1%
ا	6	6.1%
ఞ	5	5.1%
ن	5	5.1%
ں	5	5.1%
౦	4	4.0%
ب	3	3.0%
ح	3	3.0%
ۄ	3	3.0%
ৰ	3	3.0%
Other values (44)	55	55.6%

Uppercase Letter

Value	Count	Frequency (%)
I	53288	26.7%
T	19902	10.0%
A	13299	6.7%
S	10487	5.2%
E	9350	4.7%
N	8220	4.1%
O	7817	3.9%
G	7569	3.8%
W	6667	3.3%
P	6297	3.1%
Other values (22)	57057	28.5%

Other Punctuation

Value	Count	Frequency (%)
.	89228	55.9%
,	27929	17.5%

	Count	Frequency (%)
'	16074	10.1%
!	13994	8.8%
?	3561	2.2%
"	3288	2.1%
:	1617	1.0%
/	1537	1.0%
&	761	0.5%
*	619	0.4%
Other values (15)	925	0.6%

Other Symbol

Value	Count	Frequency (%)
❤	143	27.3%
☺	75	14.3%
♥	55	10.5%
♡	46	8.8%
☆	40	7.6%
★	33	6.3%
	24	4.6%
☹	22	4.2%
◇	17	3.2%
☆	15	2.9%
Other values (15)	54	10.3%

Decimal Number

Value	Count	Frequency (%)
0	4900	21.3%
1	4104	17.9%
2	3231	14.1%
5	2868	12.5%
3	2325	10.1%
4	1634	7.1%
8	1151	5.0%
9	960	4.2%
6	939	4.1%
7	866	3.8%

Nonspacing Mark

Value	Count	Frequency (%)
	45	65.2%
̊	6	8.7%
̊	4	5.8%
	3	4.3%
̦	2	2.9%
҇	2	2.9%
҉	2	2.9%
̥	2	2.9%
̦	2	2.9%
̦	1	1.4%

Math Symbol

Value	Count	Frequency (%)
+	404	67.3%
=	64	10.7%
~	60	10.0%
>	40	6.7%
<	15	2.5%
×	9	1.5%
|	4	0.7%
√	3	0.5%
÷	1	0.2%

Currency Symbol

Value	Count	Frequency (%)
$	617	93.5%
£	27	4.1%
¢	6	0.9%
€	5	0.8%
¥	2	0.3%
₩	1	0.2%
¤	1	0.2%
฿	1	0.2%

Spacing Mark

Value	Count	Frequency (%)
◌ం	9	50.0%
◌ः	3	16.7%
◌ৗ	2	11.1%
◌ୀ	1	5.6%
◌ం	1	5.6%
◌ುం	1	5.6%
◌ా ◌ౌ	1	5.6%

Dash Punctuation

Value	Count	Frequency (%)
-	3339	98.8%
—	34	1.0%
—	5	0.1%
–	2	0.1%

Open Punctuation

Value	Count	Frequency (%)
(	1943	98.3%
[	32	1.6%
{	1	0.1%
'	1	0.1%

Modifier Symbol

Value	Count	Frequency (%)
^	52	77.6%
`	12	17.9%
‾	2	3.0%
´	1	1.5%

Close Punctuation

Value	Count	Frequency (%)
)	2191	98.2%
]	35	1.6%
}	5	0.2%

Final Punctuation

Value	Count	Frequency (%)
'	20	90.9%
»	1	4.5%
"	1	4.5%

Initial Punctuation

Value	Count	Frequency (%)
"	3	60.0%
'	1	20.0%
«	1	20.0%

Other Number

Value	Count	Frequency (%)
¹	1	50.0%
²	1	50.0%

Space Separator

Value	Count	Frequency (%)
	649381	100.0%

Connector Punctuation

Value	Count	Frequency (%)
_	119	100.0%

Format

Value	Count	Frequency (%)
	8	100.0%

Enclosing Mark

Value	Count	Frequency (%)
	1	100.0%

Space Separator

Value	Count	Frequency (%)
	649381	100.0%

Connector Punctuation

Value	Count	Frequency (%)
_	119	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	3395490	80.1%
Common	841489	19.9%
Myanmar	61	< 0.1%
Inherited	49	< 0.1%
Arabic	46	< 0.1%
Kannada	16	< 0.1%
Cyrillic	8	< 0.1%
Greek	7	< 0.1%
Han	7	< 0.1%
Hangul	5	< 0.1%
Other values (2)	4	< 0.1%

Most frequent character per script

Common

Value	Count	Frequency (%)
	649381	77.2%
.	89228	10.6%
,	27929	3.3%
'	16074	1.9%
!	13994	1.7%
0	4900	0.6%
1	4104	0.5%
?	3561	0.4%
-	3339	0.4%
"	3288	0.4%
Other values (94)	25691	3.1%

Latin

Value	Count	Frequency (%)
e	425348	12.5%
a	246382	7.3%
t	246062	7.2%
s	229870	6.8%
o	221200	6.5%
i	213536	6.3%
n	208620	6.1%
r	183720	5.4%
l	177263	5.2%
d	124405	3.7%
Other values (69)	1119084	33.0%

Myanmar

Value	Count	Frequency (%)
ေ၁	9	14.8%
၆	6	9.8%
သ	5	8.2%
ပ	5	8.2%
�	4	6.6%
၆	4	6.6%
၁:	3	4.9%
ဝ	3	4.9%
၃	2	3.3%
၁	2	3.3%
Other values (13)	18	29.5%

Arabic

Value	Count	Frequency (%)
ي	7	15.2%
١	6	13.0%

ن	5	10.9%
ب	3	6.5%
ع	3	6.5%
ه	3	6.5%
ح	2	4.3%
د	2	4.3%
ت	2	4.3%
ف	2	4.3%
Other values (10)	11	23.9%

Kannada

Value	Count	Frequency (%)
ಂ	2	12.5%
ಕ	2	12.5%
ಾ	2	12.5%
ನ	2	12.5%
ಡ	1	6.2%
ದ	1	6.2%
ಂೂ	1	6.2%
ಟ	1	6.2%
ಯು	1	6.2%
ಥ	1	6.2%
Other values (2)	2	12.5%

Han

Value	Count	Frequency (%)
楼	2	28.6%
租	1	14.3%
搵	1	14.3%
喔	1	14.3%
言	1	14.3%
无	1	14.3%

Cyrillic

Value	Count	Frequency (%)
є	3	37.5%
я	2	25.0%
у	1	12.5%
т	1	12.5%
к	1	12.5%

Greek

Value	Count	Frequency (%)
ι	3	42.9%
ϊ	1	14.3%
χ	1	14.3%
α	1	14.3%
ν	1	14.3%

Hangul

Value	Count	Frequency (%)
픈	1	20.0%
슬	1	20.0%
게	1	20.0%
떨	1	20.0%
어	1	20.0%

Inherited

Value	Count	Frequency (%)
	45	91.8%
	3	6.1%
	1	2.0%

Malayalam

Value	Count	Frequency (%)
ട	1	33.3%
ைக	1	33.3%
ா	1	33.3%

Hiragana

Value	Count	Frequency (%)
し	1	100.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	4236223	> 99.9%
Misc Symbols	311	< 0.1%
Dingbats	174	< 0.1%
None	147	< 0.1%
Punctuation	97	< 0.1%
Myanmar	61	< 0.1%
Arabic	46	< 0.1%
VS	45	< 0.1%
Geometric Shapes	18	< 0.1%
Kannada	16	< 0.1%
Other values (11)	44	< 0.1%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
	649381	15.3%
e	425348	10.0%
a	246382	5.8%
t	246062	5.8%
s	229870	5.4%
o	221200	5.2%
i	213536	5.0%
n	208620	4.9%
r	183720	4.3%
l	177263	4.2%
Other values (85)	1434841	33.9%

Dingbats

Value	Count	Frequency (%)
❤	143	82.2%
✌	14	8.0%
✳	10	5.7%
❣	5	2.9%
✆	1	0.6%
✓	1	0.6%

Misc Symbols

Value	Count	Frequency (%)
☺	75	24.1%
♥	55	17.7%
♡	46	14.8%
☆	40	12.9%
★	33	10.6%
	24	7.7%
☹	22	7.1%
⚫	6	1.9%
♬	3	1.0%
♪	3	1.0%
Other values (3)	4	1.3%

VS

Value	Count	Frequency (%)
	45	100.0%

Punctuation

Value	Count	Frequency (%)
—	34	35.1%
'	20	20.6%
…	12	12.4%
	8	8.2%

	Count	Frequency (%)
—	5	5.2%
•	4	4.1%
‼	4	4.1%
"	3	3.1%
–	2	2.1%
‰	1	1.0%
Other values (4)	4	4.1%

None

Value	Count	Frequency (%)
£	27	18.4%
⁎	15	10.2%
﹡	13	8.8%
×	9	6.1%
¢	6	4.1%
ú	4	2.7%
ë	4	2.7%
à	4	2.7%
č	3	2.0%
í	3	2.0%
Other values (41)	59	40.1%

Geometric Shapes

Value	Count	Frequency (%)
◇	17	94.4%
◆	1	5.6%

Myanmar

Value	Count	Frequency (%)
၆၀	9	14.8%
၆	6	9.8%
တ	5	8.2%
ပ	5	8.2%
သ	4	6.6%
၉	4	6.6%
၀း	3	4.9%
ယ	3	4.9%
၀	2	3.3%
၈	2	3.3%
Other values (13)	18	29.5%

Arabic

Value	Count	Frequency (%)
ي	7	15.2%
ا	6	13.0%
ن	5	10.9%
ب	3	6.5%
ح	3	6.5%
ه	3	6.5%
ج	2	4.3%
د	2	4.3%
ت	2	4.3%
ق	2	4.3%
Other values (10)	11	23.9%

Currency Symbols

Value	Count	Frequency (%)
€	5	83.3%
₩	1	16.7%

Math Operators

Value	Count	Frequency (%)
√	3	100.0%

Diacriticals

Value	Count	Frequency (%)
	3	100.0%

Cyrillic

Value	Count	Frequency (%)
є	3	37.5%
я	2	25.0%
у	1	12.5%
т	1	12.5%
к	1	12.5%

Letterlike Symbols

Value	Count	Frequency (%)
ℓ	2	33.3%
™	2	33.3%
℅	2	33.3%

CJK

Value	Count	Frequency (%)
楼	2	28.6%
租	1	14.3%
搵	1	14.3%
喔	1	14.3%
言	1	14.3%
无	1	14.3%

Kannada

Value	Count	Frequency (%)
◌ಕ್	2	12.5%
ಠ	2	12.5%
◌ಾ	2	12.5%
ನ	2	12.5%
ಡ	1	6.2%
ದ	1	6.2%
◌ೌಲೀ	1	6.2%
ಟ	1	6.2%
ಯು	1	6.2%
ಧ	1	6.2%
Other values (2)	2	12.5%

Malayalam

Value	Count	Frequency (%)
ട	1	33.3%
ഌ	1	33.3%
ൻ	1	33.3%

Thai

Value	Count	Frequency (%)
฿	1	100.0%

Hiragana

Value	Count	Frequency (%)
し	1	100.0%

Hangul

Value	Count	Frequency (%)
픈	1	20.0%
숄	1	20.0%
게	1	20.0%
떨	1	20.0%
어	1	20.0%

Devanagari

Value	Count	Frequency (%)
I	1	100.0%

Sentiment
Categorical

Distinct	3
Distinct (%)	< 0.1%
Missing	36522
Missing (%)	49.4%
Memory size	4.5 MiB

Length

Max length	8
Median length	8
Mean length	7.8620699
Min length	7

Characters and Unicode

Total characters	294293	
Distinct characters	13	
Distinct categories	2 (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	0	?
Unique (%)	0.0%	

Sample

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

Common Values

Value	Count	Frequency (%)
Positive	23998	32.4%
Negative	8271	11.2%
Neutral	5163	7.0%
(Missing)	36522	49.4%

Length

Histogram of lengths of the category

Common Values (Plot)

64.1%
(23998)

22.1%
(8271)

13.8%
(5163)
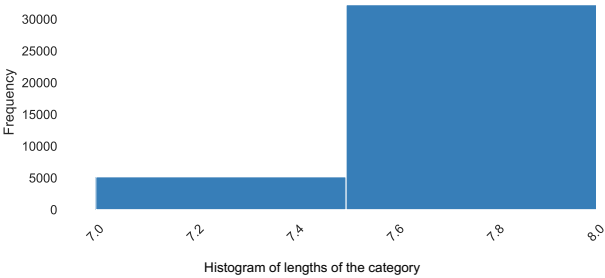
Positive
Negative
Neutral
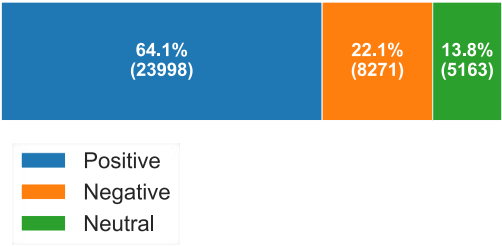
Value	Count	Frequency (%)
positive	23998	64.1%
negative	8271	22.1%
neutral	5163	13.8%

Most occurring characters

Value	Count	Frequency (%)
i	56267	19.1%
e	45703	15.5%
t	37432	12.7%
v	32269	11.0%
P	23998	8.2%
o	23998	8.2%
s	23998	8.2%
N	13434	4.6%
a	13434	4.6%
g	8271	2.8%
Other values (3)	15489	5.3%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	256861	87.3%
Uppercase Letter	37432	12.7%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
i	56267	21.9%
e	45703	17.8%
t	37432	14.6%
v	32269	12.6%
o	23998	9.3%
s	23998	9.3%
a	13434	5.2%
g	8271	3.2%
u	5163	2.0%
r	5163	2.0%

Uppercase Letter

Value	Count	Frequency (%)
P	23998	64.1%
N	13434	35.9%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	256861	87.3%
Uppercase Letter	37432	12.7%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)

Most occurring scripts

Value	Count	Frequency (%)
Latin	294293	100.0%

Most frequent character per script

Latin

Value	Count	Frequency (%)
i	56267	19.1%
e	45703	15.5%
t	37432	12.7%
v	32269	11.0%
P	23998	8.2%
o	23998	8.2%
s	23998	8.2%
N	13434	4.6%
a	13434	4.6%
g	8271	2.8%
Other values (3)	15489	5.3%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	294293	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
i	56267	19.1%
e	45703	15.5%
t	37432	12.7%
v	32269	11.0%
P	23998	8.2%
o	23998	8.2%
s	23998	8.2%
N	13434	4.6%
a	13434	4.6%
g	8271	2.8%
Other values (3)	15489	5.3%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	294293	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
i	56267	19.1%
e	45703	15.5%

Sentiment_Polarity
Real number (ℝ)

HIGH CORRELATION (This variable has a high overall correlation with 1 fields: Sentiment)   MISSING   ZEROS

Distinct	5410
Distinct (%)	14.5%
Missing	36522
Missing (%)	49.4%
Infinite	0
Infinite (%)	0.0%
Mean	0.18214631

Minimum	-1
Maximum	1
Zeros	5163
Zeros (%)	7.0%
Negative	8271
Negative (%)	11.2%
Memory size	577.9 KiB
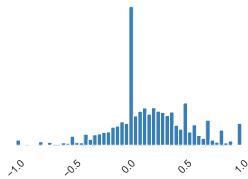
Quantile statistics

Minimum	-1
5-th percentile	-0.4
Q1	0
median	0.15
Q3	0.4
95-th percentile	0.8
Maximum	1
Range	2
Interquartile range (IQR)	0.4

Descriptive statistics

Standard deviation	0.35130098
Coefficient of variation (CV)	1.9286747
Kurtosis	0.64700243
Mean	0.18214631
Median Absolute Deviation (MAD)	0.195
Skewness	-0.10458074
Sum	6818.1008
Variance	0.12341238
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=50)
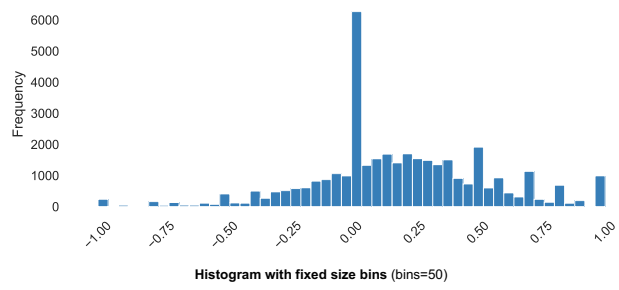
Value	Count	Frequency (%)
0	5163	7.0%
0.5	1524	2.1%
0.7	991	1.3%
1	959	1.3%
0.8	639	0.9%
0.6	623	0.8%
0.2	607	0.8%
0.3	568	0.8%
0.25	477	0.6%
0.4	445	0.6%
Other values (5400)	25436	34.4%
(Missing)	36522	49.4%

Value	Count	Frequency (%)
-1	234	0.3%
-0.98828125	2	< 0.1%
-0.9765625	1	< 0.1%
-0.955	1	< 0.1%
-0.9375	5	< 0.1%
-0.9333333333	1	< 0.1%
-0.91	28	< 0.1%
-0.9	15	< 0.1%
-0.9	1	< 0.1%
-0.8928571429	2	< 0.1%

Value	Count	Frequency (%)
1	959	1.3%
0.9921875	2	< 0.1%
0.98828125	2	< 0.1%
0.9765625	24	< 0.1%
0.975	1	< 0.1%
0.96875	3	< 0.1%
0.955	5	< 0.1%
0.9536743164	1	< 0.1%
0.95	3	< 0.1%
0.9375	6	< 0.1%

Sentiment_Subjectivity
Real number (ℝ)

HIGH CORRELATION (This variable has a high overall correlation with 1 fields: Sentiment) MISSING ZEROS

Distinct	4474
Distinct (%)	12.0%
Missing	36522
Missing (%)	49.4%
Infinite	0
Infinite (%)	0.0%
Mean	0.49270393

Minimum	0
Maximum	1
Zeros	4416
Zeros (%)	6.0%
Negative	0
Negative (%)	0.0%
Memory size	577.9 KiB
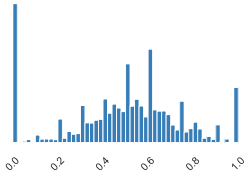
Quantile statistics

Minimum	0
5-th percentile	0
Q1	0.35714286
median	0.51428571
Q3	0.65
95-th percentile	0.94444444
Maximum	1
Range	1
Interquartile range (IQR)	0.29285714

Descriptive statistics

Standard deviation	0.25994901
Coefficient of variation (CV)	0.5275968
Kurtosis	-0.28273095
Mean	0.49270393
Median Absolute Deviation (MAD)	0.14458874
Skewness	-0.30634588
Sum	18442.893
Variance	0.06757349
Monotonicity	Not monotonic

Histogram with fixed size bins (bins=50)
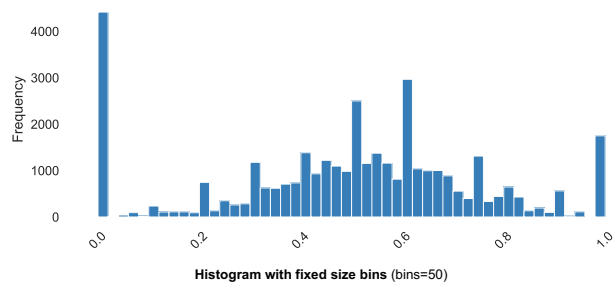
Value	Count	Frequency (%)
0	4416	6.0%
1	1749	2.4%
0.5	1647	2.2%
0.6	1166	1.6%
0.75	1140	1.5%
0.6	1060	1.4%
0.3	886	1.2%
0.4	667	0.9%
0.2	647	0.9%
0.8	514	0.7%
Other values (4464)	23540	31.8%
(Missing)	36522	49.4%

Value	Count	Frequency (%)
0	4416	6.0%
0.02222222222	1	< 0.1%
0.025	5	< 0.1%
0.03333333333	6	< 0.1%
0.04166666667	2	< 0.1%
0.05	35	< 0.1%
0.05555555556	3	< 0.1%
0.05833333333	3	< 0.1%
0.0625	9	< 0.1%
0.06666666667	67	0.1%

Value	Count	Frequency (%)
1	1749	2.4%
0.9875	1	< 0.1%
0.9759259259	1	< 0.1%
0.975	9	< 0.1%
0.9666666667	9	< 0.1%
0.962962963	1	< 0.1%
0.9583333333	2	< 0.1%
0.9555555556	1	< 0.1%
0.955	2	< 0.1%
0.9523809524	1	< 0.1%

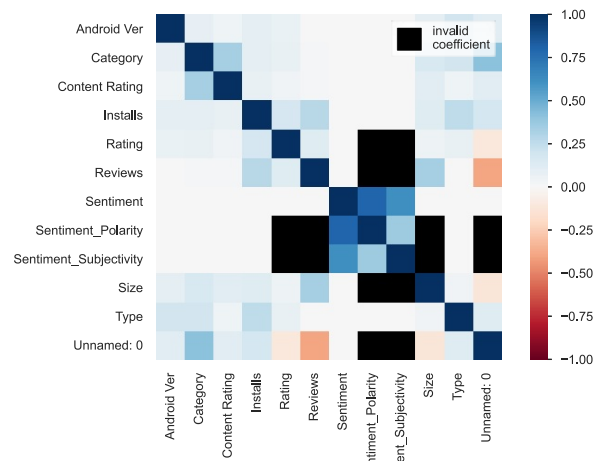# Interactions

Unnamed: 0

Sentiment_Polarity
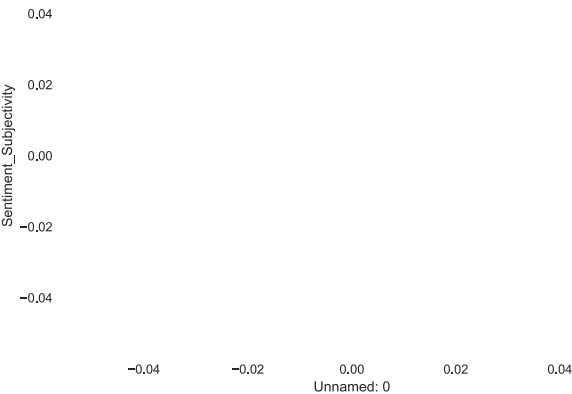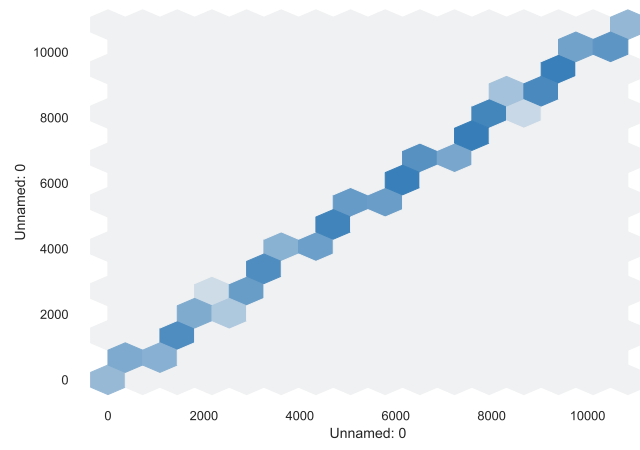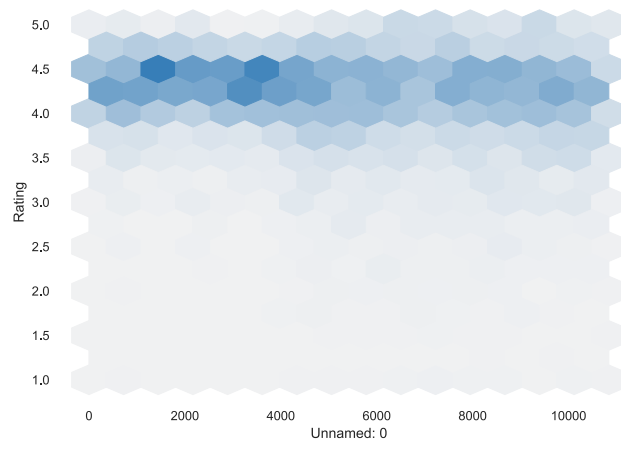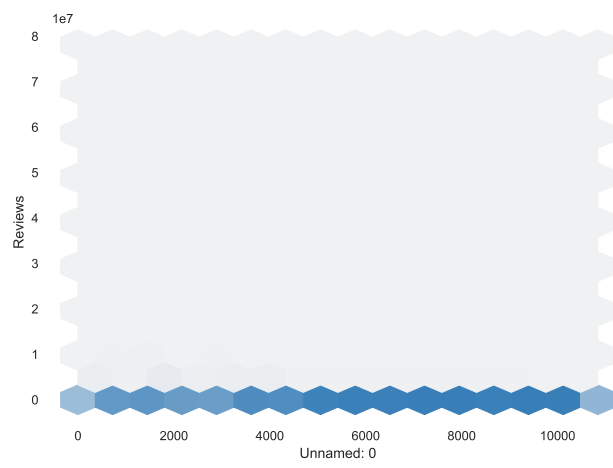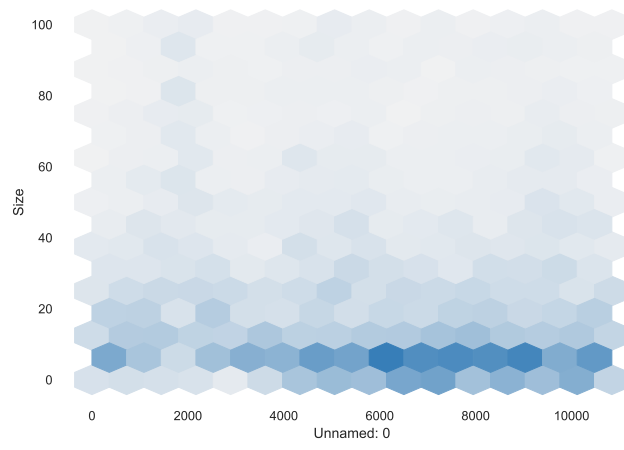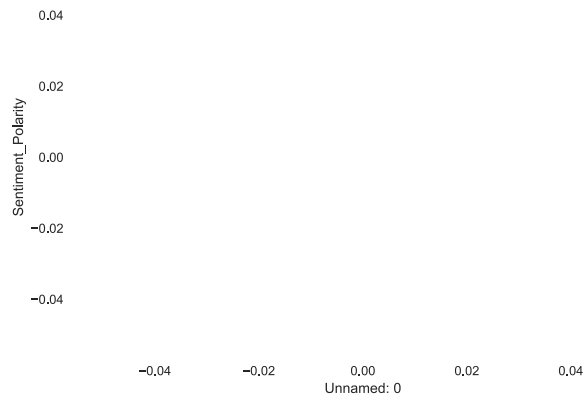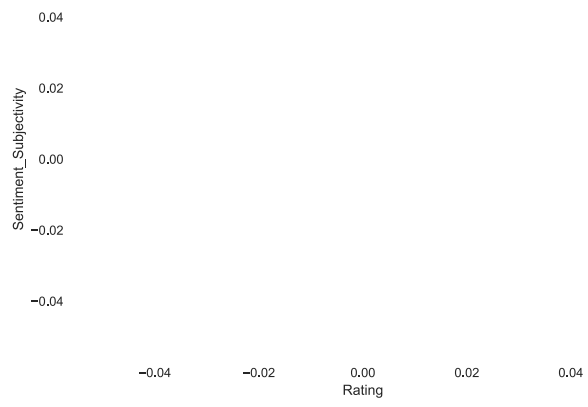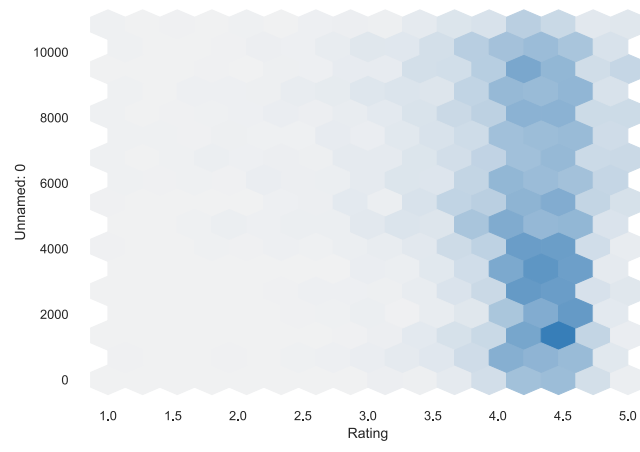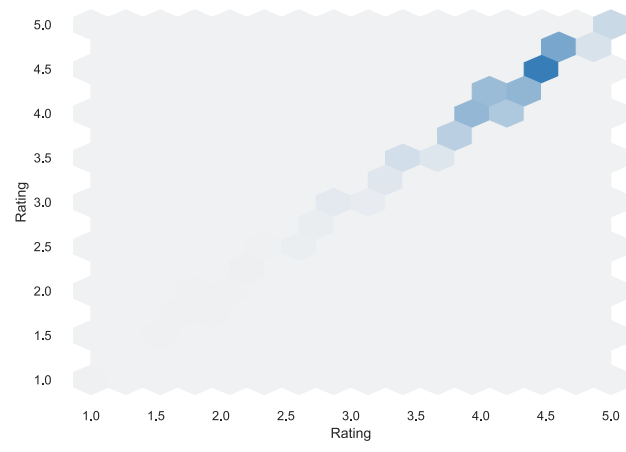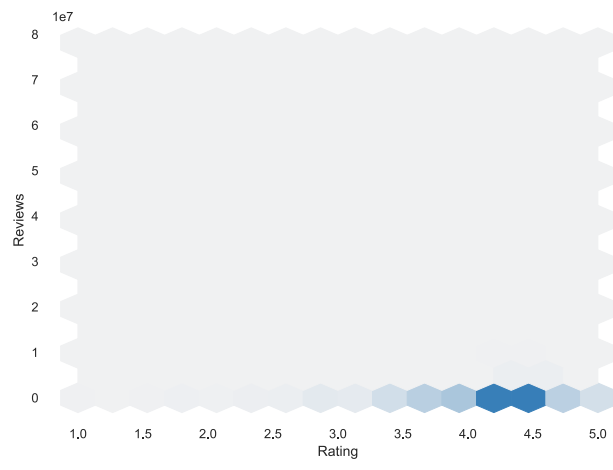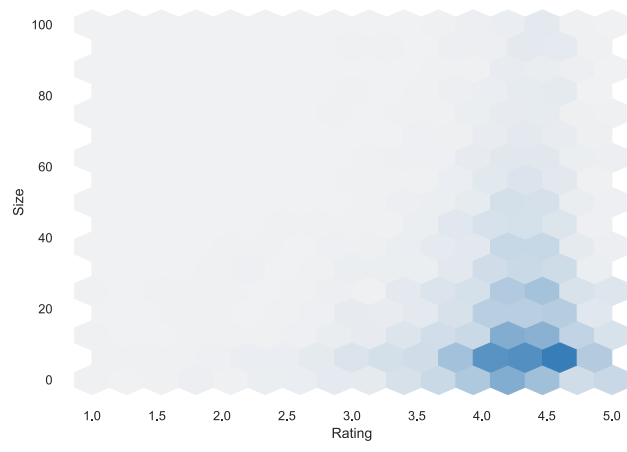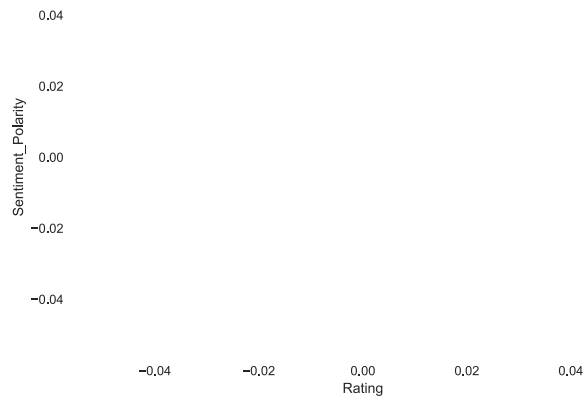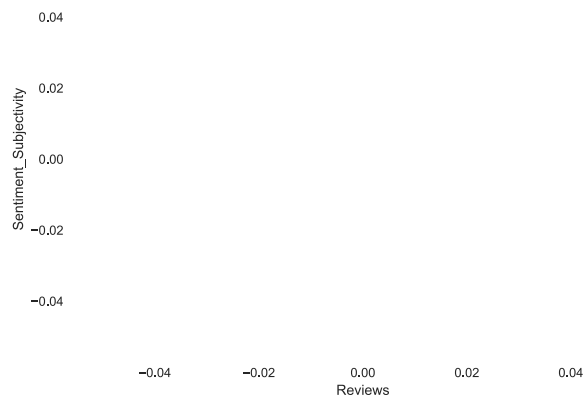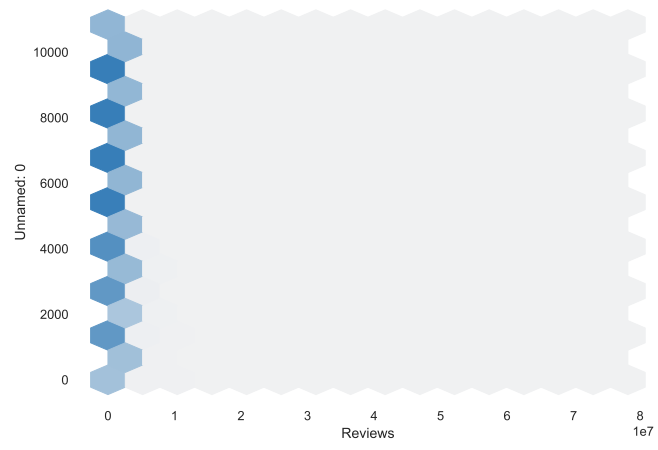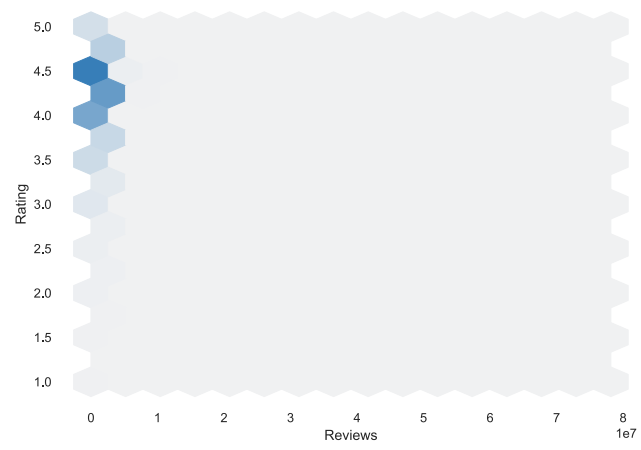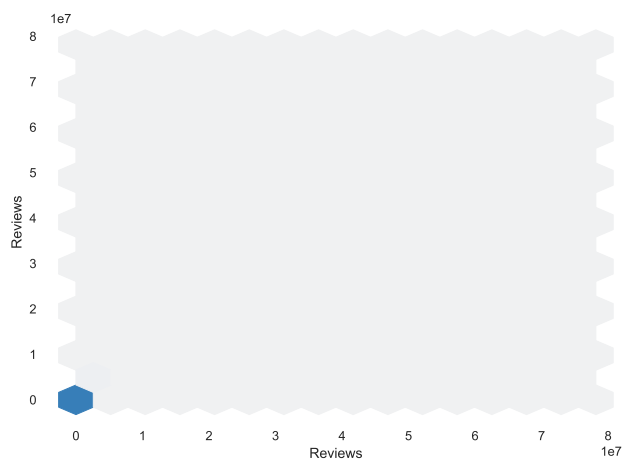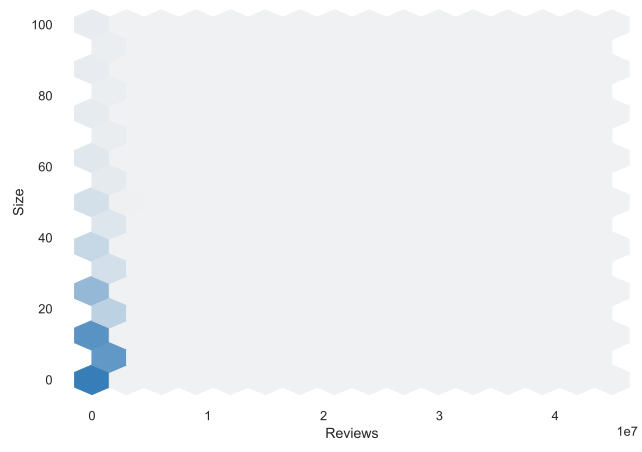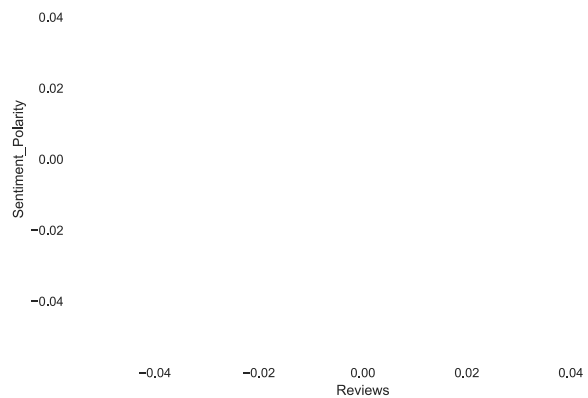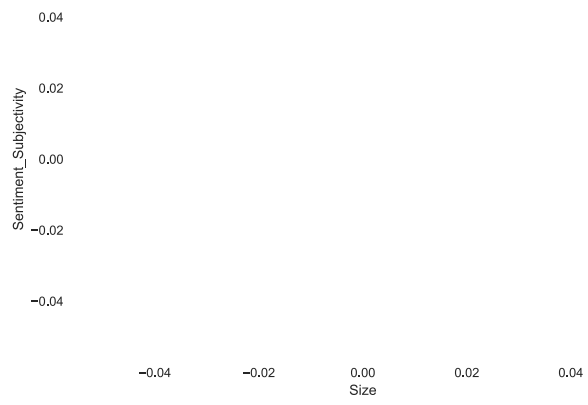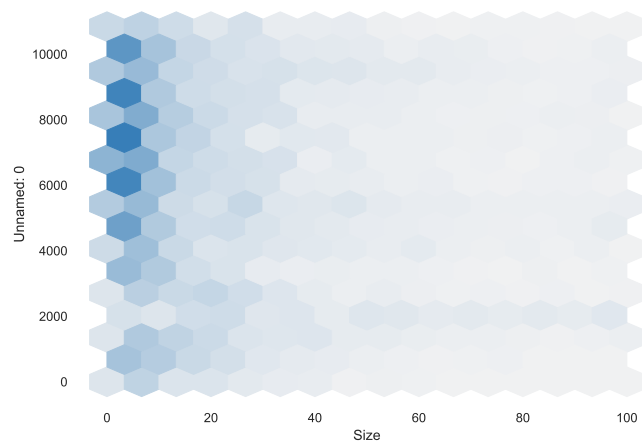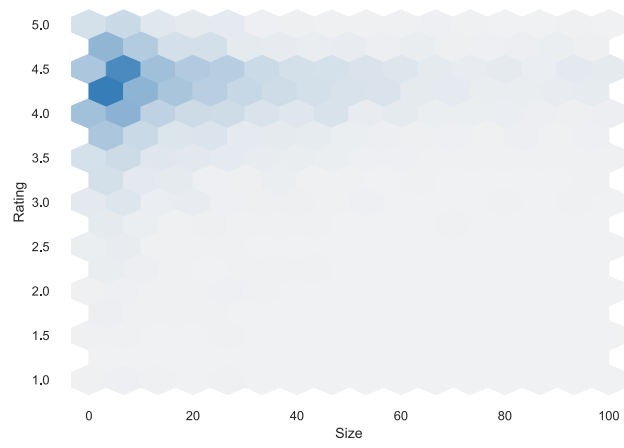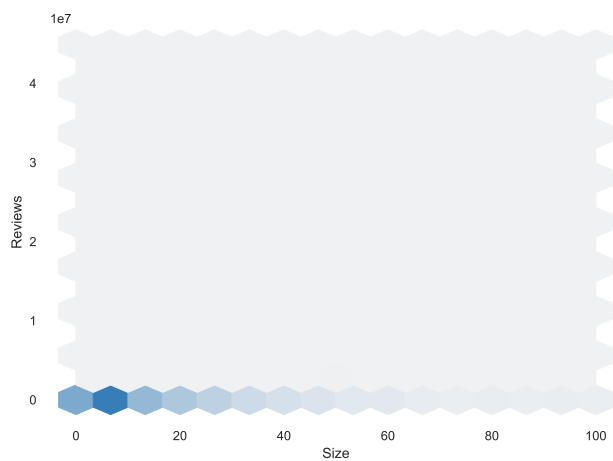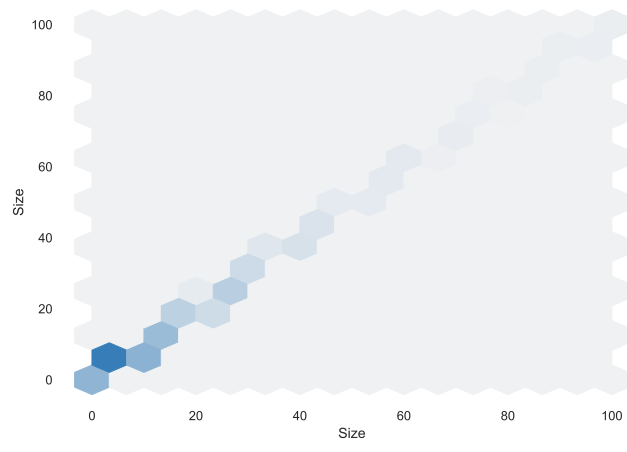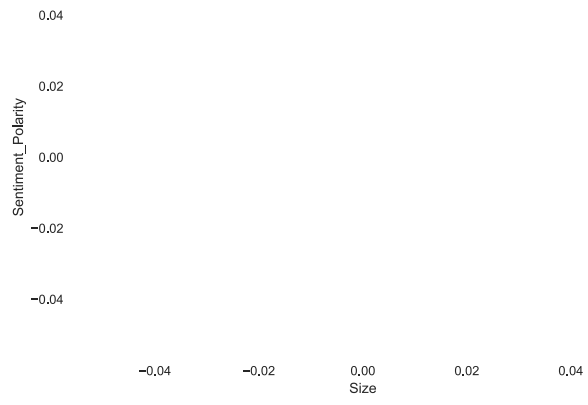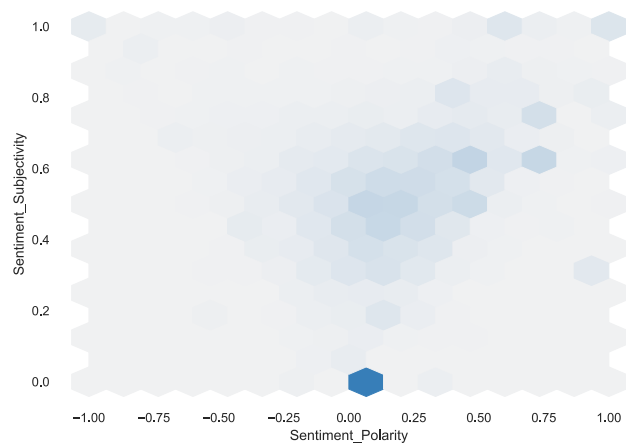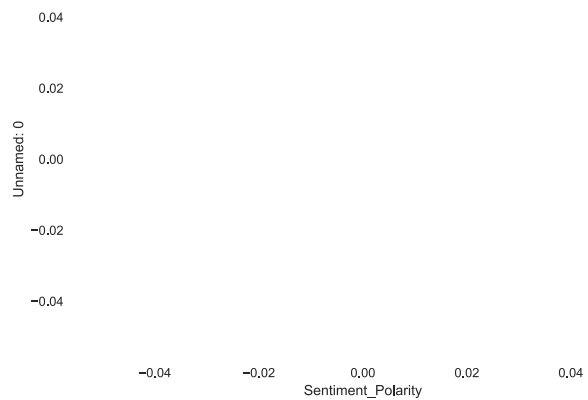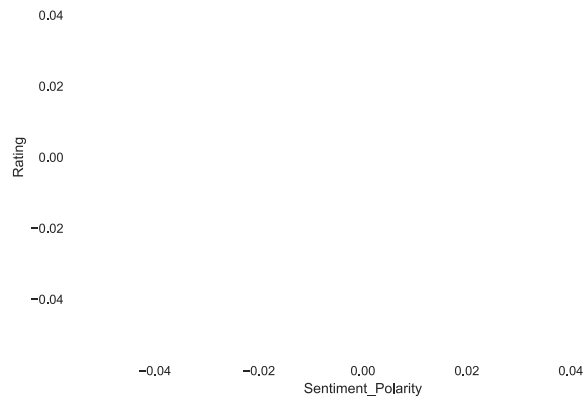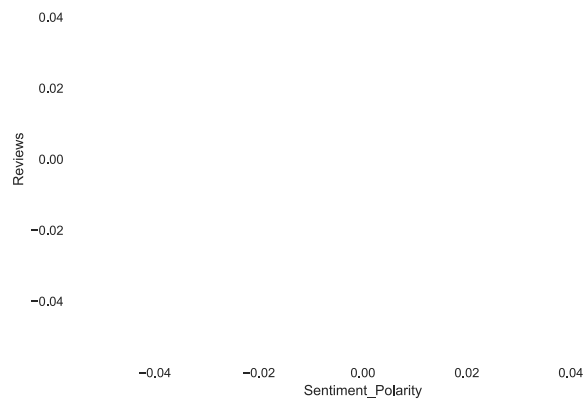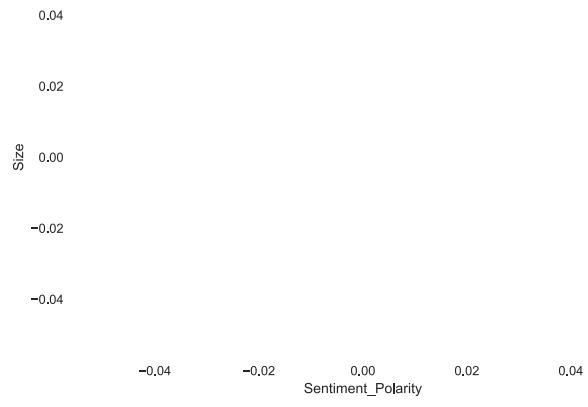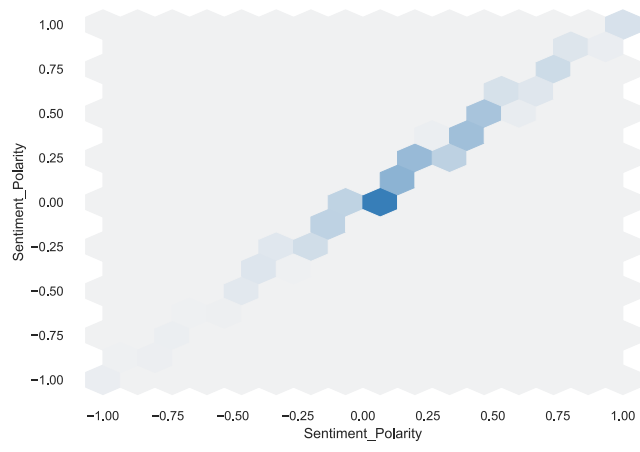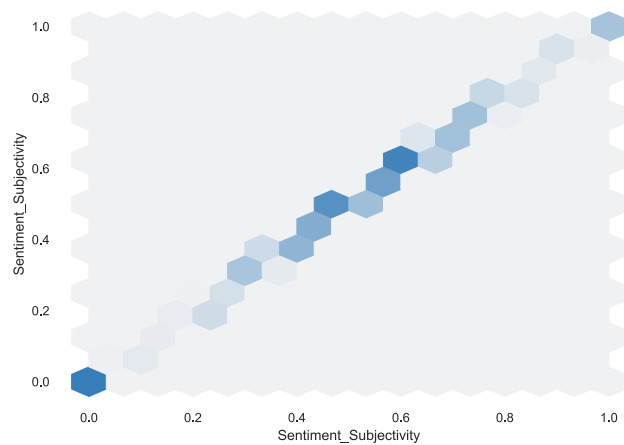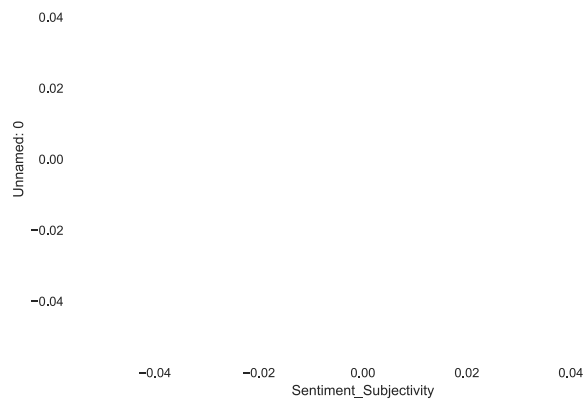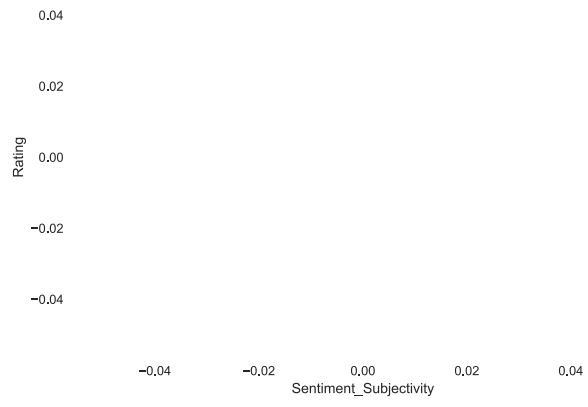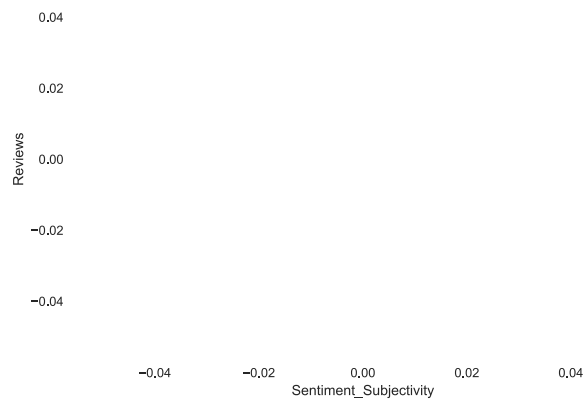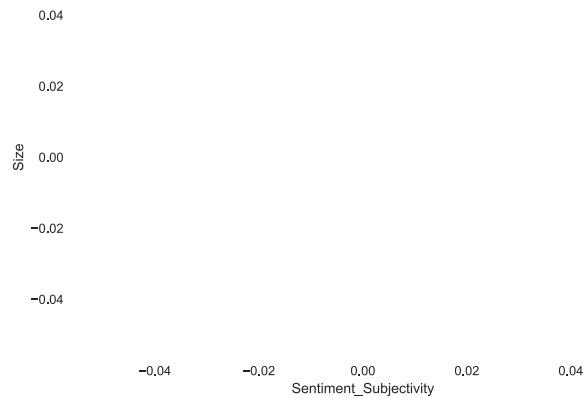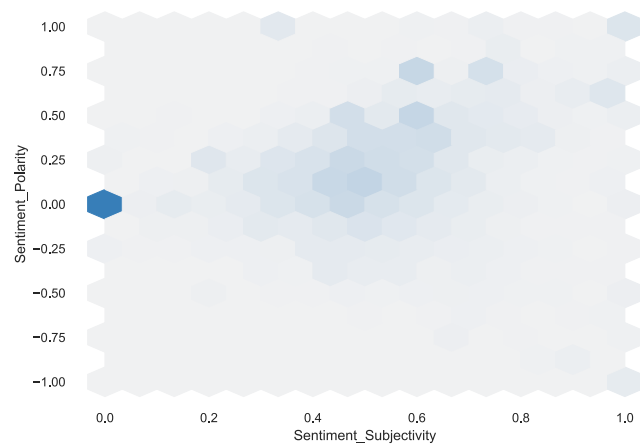
Correlations

	Android Ver	Category	Content Rating	Installs	Rating	Reviews	Sentiment	Sentiment_Polarity	Sentiment_Subjectivity	Size	Type	Unnamed: 0
Android Ver	1.000	0.083	0.046	0.088	0.067	0.000	0.000	0.000	0.000	0.086	0.181	0.103
Category	0.083	1.000	0.329	0.098	0.077	0.016	0.000	0.000	0.000	0.151	0.180	0.409
Content Rating	0.046	0.329	1.000	0.075	0.032	0.016	0.000	0.000	0.000	0.102	0.041	0.113
Installs	0.088	0.098	0.075	1.000	0.168	0.276	0.000	0.000	0.000	0.118	0.252	0.172
Rating	0.067	0.077	0.032	0.168	1.000	0.120	0.000	NaN	NaN	0.050	0.071	-0.102
Reviews	0.000	0.016	0.016	0.276	0.120	1.000	0.000	NaN	NaN	0.330	0.000	-0.392
Sentiment	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.793	0.615	0.000	0.000	0.000
Sentiment_Polarity	0.000	0.000	0.000	0.000	NaN	NaN	0.793	1.000	0.352	NaN	0.000	NaN
Sentiment_Subjectivity	0.000	0.000	0.000	0.000	NaN	NaN	0.615	0.352	1.000	NaN	0.000	NaN
Size	0.086	0.151	0.102	0.118	0.050	0.330	0.000	NaN	NaN	1.000	0.039	-0.121

Missing values
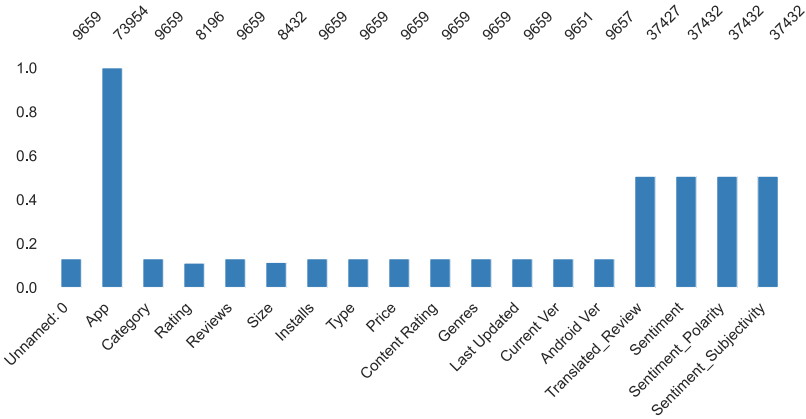
A simple visualization of nullity by column.

1

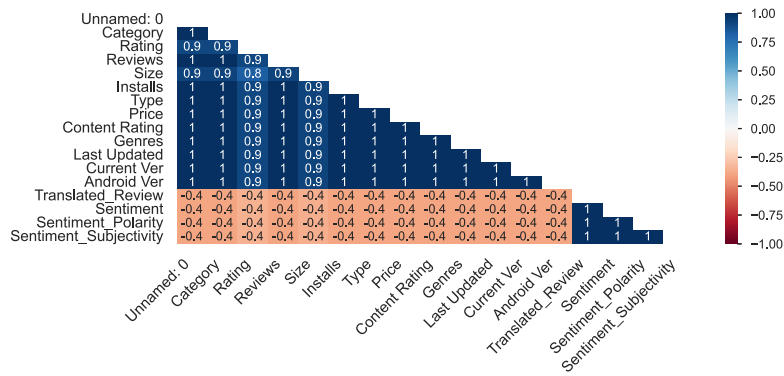Unnamed: 0 App Category Rating Reviews Size Installs Type Price Content Rating Genres Last Updated Current Ver Android Ver Translated_Review Sentiment Sentiment_Polarity Sentiment_Subjectivity

73954

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

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

Sample

	Unnamed: 0	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ve
0	0.0	Photo Editor & Candy Camera & Grid & ScrapBook	ART_AND_DESIGN	4.1	159.0	19.0	10,000+	Free	0	Everyone	Art & Design	January 7, 2018	1.0.0
1	1.0	Coloring book moana	ART_AND_DESIGN	3.9	967.0	14.0	500,000+	Free	0	Everyone	Art & Design;Pretend Play	January 15, 2018	2.0.0
2	2.0	U Launcher Lite – FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510.0	8.7	5,000,000+	Free	0	Everyone	Art & Design	August 1, 2018	1.2.4
3	3.0	Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644.0	25.0	50,000,000+	Free	0	Teen	Art & Design	June 8, 2018	Varies with
4	4.0	Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967.0	2.8	100,000+	Free	0	Everyone	Art & Design;Creativity	June 20, 2018	1.1
5	5.0	Paper flowers instructions	ART_AND_DESIGN	4.4	167.0	5.6	50,000+	Free	0	Everyone	Art & Design	March 26, 2017	1
6	6.0	Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178.0	19.0	50,000+	Free	0	Everyone	Art & Design	April 26, 2018	1.1
7	7.0	Infinite Painter	ART_AND_DESIGN	4.1	36815.0	29.0	1,000,000+	Free	0	Everyone	Art & Design	June 14, 2018	6.1.61.1
8	8.0	Garden Coloring Book	ART_AND_DESIGN	4.4	13791.0	33.0	1,000,000+	Free	0	Everyone	Art & Design	September 20, 2017	2.9.2
9	9.0	Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7	121.0	3.1	10,000+	Free	0	Everyone	Art & Design;Creativity	July 3, 2018	2.8

	Unnamed: 0	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver	Translated_Review	Sentiment	Sentim
73944	NaN	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
73945	NaN	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
73946	NaN	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
73947	NaN	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
73948	NaN	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
73949	NaN	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
73950	NaN	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
73951	NaN	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
73952	NaN	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
73953	NaN	Houzz Interior Design Ideas	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Duplicate rows

Most frequently occurring

	Unnamed: 0	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Rating	Genres	Last Updated	Current Ver	Android Ver	Translated_Review	Sen
3655	NaN	ESPN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1524	NaN	Bleacher Report: sports news, scores, & highlights	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2320	NaN	Calorie Counter by FatSecret	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5574	NaN	Granny	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2692	NaN	ClassDojo	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
6010	NaN	Hill Climb Racing	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1490	NaN	BeautyPlus - Easy Photo Editor & Selfie Camera	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2691	NaN	Clash of Clans	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3657	NaN	Earn to Die 2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3516	NaN	Dream League Soccer 2018	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN