Overview

Dataset statistics

Number of variables	10
Number of observations	10605
Missing cells	1972
Missing cells (%)	1.9%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	869.9 KiB
Average record size in memory	84.0 B

Variable types

Numeric	7
Categorical	3

Alerts

`Gender` is highly correlated with `is_premium_member`	High correlation
`is_premium_member` is highly correlated with `Gender`	High correlation
`Gender` is highly correlated with `Men's_Clothing` and 3 other fields (Men's_Clothing, Kid's_Clothing, Home_&_Living, Beauty)	High correlation
`Men's_Clothing` is highly correlated with `Gender` and 1 other fields (Gender, Beauty)	High correlation
`Kid's_Clothing` is highly correlated with `Gender`	High correlation
`Home_&_Living` is highly correlated with `Gender` and 1 other fields (Gender, Beauty)	High correlation
`Beauty` is highly correlated with `Gender` and 2 other fields (Gender, Men's_Clothing, Home_&_Living)	High correlation
`Age` has 729 (6.9%) missing values	Missing
`Kid's_Clothing` has 648 (6.1%) missing values	Missing
`Home_&_Living` has 595 (5.6%) missing values	Missing

Reproduction

Analysis started	2022-06-14 21:45:44.885913
Analysis finished	2022-06-14 21:45:52.219085
Duration	7.33 seconds
Software version	pandas-profiling v3.2.0 (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%

Variables

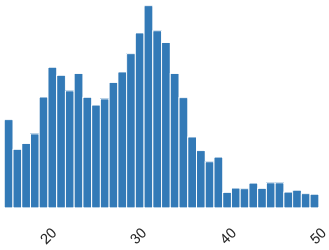
Age

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

MISSING

Distinct	36
Distinct (%)	0.4%
Missing	729
Missing (%)	6.9%
Infinite	0
Infinite (%)	0.0%
Mean	28.33880113

Minimum	15
Maximum	50
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	165.7 KiB



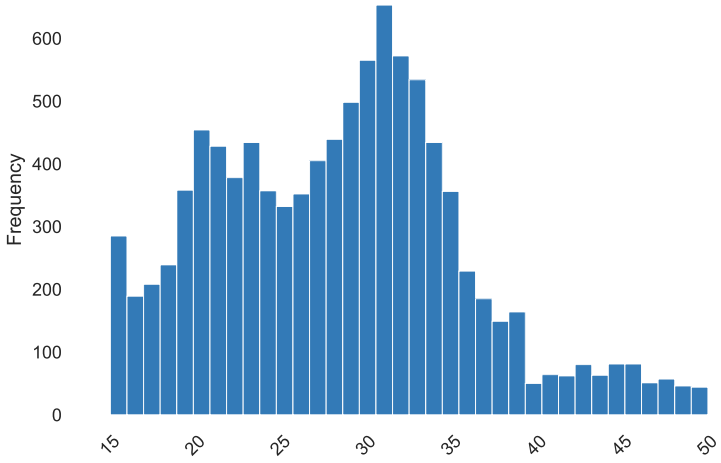
Quantile statistics

Minimum	15
5-th percentile	17
Q1	22
median	29
Q3	33
95-th percentile	43
Maximum	50
Range	35
Interquartile range (IQR)	11

Descriptive statistics

Standard deviation	7.548287568
Coefficient of variation (CV)	0.2663587472
Kurtosis	-0.1053307746
Mean	28.33880113
Median Absolute Deviation (MAD)	5
Skewness	0.3780147909
Sum	279874

Variance	56.97664521
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=36)

Value	Count	Frequency (%)
31	653	6.2%
32	572	5.4%
30	565	5.3%
33	534	5.0%
29	498	4.7%
20	454	4.3%
28	439	4.1%
34	434	4.1%
23	434	4.1%
21	428	4.0%
Other values (26)	4865	45.9%
(Missing)	729	6.9%

Value	Count	Frequency (%)
15	285	2.7%
16	189	1.8%
17	208	2.0%
18	239	2.3%
19	358	3.4%
20	454	4.3%
21	428	4.0%
22	378	3.6%
23	434	4.1%
24	357	3.4%

Value	Count	Frequency (%)
50	44	0.4%
49	46	0.4%
48	57	0.5%
47	51	0.5%
46	81	0.8%
45	81	0.8%
44	63	0.6%
43	80	0.8%
42	62	0.6%
41	64	0.6%

Gender

Categorical

HIGH CORRELATION (This variable has a high correlation with 1 fields: is premium member)
HIGH CORRELATION (This variable has a high correlation with 4 fields: Men’s Clothing, Kid’s Clothing, Home & Living, Beauty)

Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	165.7 KiB

Length

Max length	13
Median length	6
Mean length	6.145780292
Min length	4

Characters and Unicode

Total characters	65176	
Distinct characters	16	
Distinct categories	3 (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%	

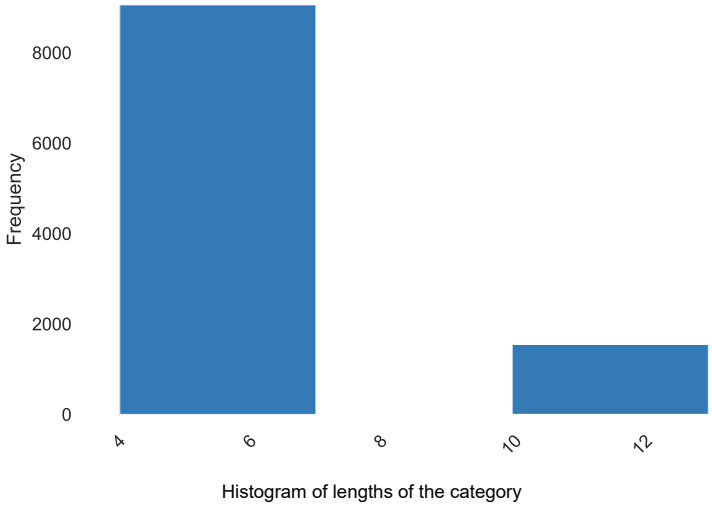
Sample

1st row	Not_Specified
2nd row	Male
3rd row	Not_Specified
4th row	Not_Specified
5th row	Female

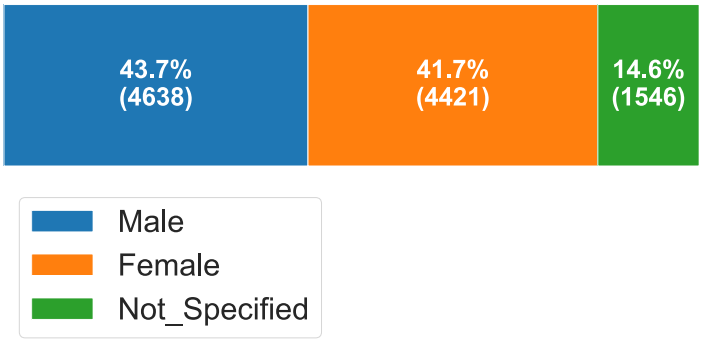
Common Values

Value	Count	Frequency (%)
Male	4638	43.7%
Female	4421	41.7%
Not_Specified	1546	14.6%

Length



Category Frequency Plot



Value	Count	Frequency (%)
male	4638	43.7%
female	4421	41.7%
not_specified	1546	14.6%

Most occurring characters

Value	Count	Frequency (%)
e	16572	25.4%
a	9059	13.9%
l	9059	13.9%
M	4638	7.1%
F	4421	6.8%
m	4421	6.8%
i	3092	4.7%
N	1546	2.4%
o	1546	2.4%
t	1546	2.4%
Other values (6)	9276	14.2%

Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	51479	79.0%
Uppercase Letter	12151	18.6%
Connector Punctuation	1546	2.4%

Most frequent character per category

Lowercase Letter

Value	Count	Frequency (%)
e	16572	32.2%
a	9059	17.6%
l	9059	17.6%
m	4421	8.6%
i	3092	6.0%
o	1546	3.0%
t	1546	3.0%
p	1546	3.0%
c	1546	3.0%
f	1546	3.0%

Uppercase Letter

Value	Count	Frequency (%)
M	4638	38.2%
F	4421	36.4%
N	1546	12.7%
S	1546	12.7%

Connector Punctuation

Value	Count	Frequency (%)
_	1546	100.0%

Most occurring scripts

Value	Count	Frequency (%)
Latin	63630	97.6%
Common	1546	2.4%

Most frequent character per script

Latin

Value	Count	Frequency (%)
e	16572	26.0%
a	9059	14.2%
l	9059	14.2%
M	4638	7.3%
F	4421	6.9%
m	4421	6.9%
i	3092	4.9%
N	1546	2.4%
o	1546	2.4%
t	1546	2.4%
Other values (5)	7730	12.1%

Common

Value	Count	Frequency (%)
_	1546	100.0%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	65176	100.0%

Most frequent character per block

ASCII

Value	Count	Frequency (%)
e	16572	25.4%
a	9059	13.9%
l	9059	13.9%
M	4638	7.1%
F	4421	6.8%
m	4421	6.8%
i	3092	4.7%
N	1546	2.4%
o	1546	2.4%
t	1546	2.4%
Other values (6)	9276	14.2%

is_premium_member

Categorical

HIGH

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

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	165.7 KiB

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	10605	
Distinct characters	2	
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	0	?
Unique (%)	0.0%	

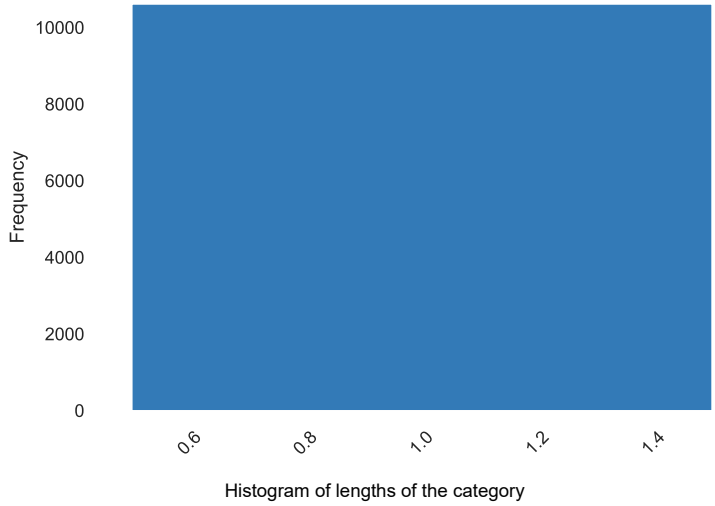
Sample

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

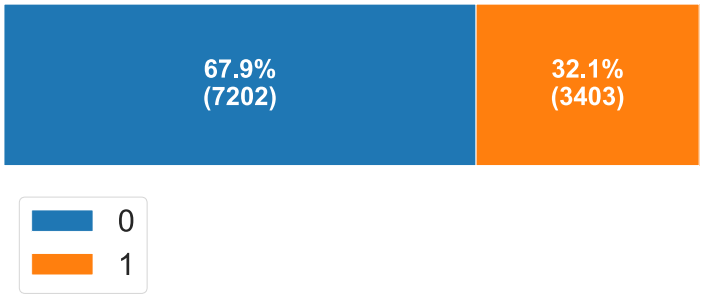
Common Values

Value	Count	Frequency (%)
0	7202	67.9%
1	3403	32.1%

Length



Category Frequency Plot



Value	Count	Frequency (%)
0	7202	67.9%
1	3403	32.1%

Most occurring characters

Value	Count	Frequency (%)
0	7202	67.9%
1	3403	32.1%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	10605	100.0%

Most frequent character per category

Decimal Number

Value	Count	Frequency (%)
0	7202	67.9%
1	3403	32.1%

Most occurring scripts

Value	Count	Frequency (%)
Common	10605	100.0%

Most frequent character per script

Common

Value	Count	Frequency (%)
0	7202	67.9%
1	3403	32.1%

Most occurring blocks

Value	Count	Frequency (%)
ASCII	10605	100.0%

Most frequent character per block

ASCII

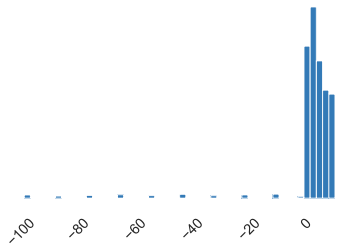
Value	Count	Frequency (%)
0	7202	67.9%
1	3403	32.1%

Women's_Clothing

Real number (ℝ)

Distinct	9556
Distinct (%)	90.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1.869138073

Minimum	-99
Maximum	9.993592355
Zeros	56
Zeros (%)	0.5%
Negative	458
Negative (%)	4.3%
Memory size	165.7 KiB

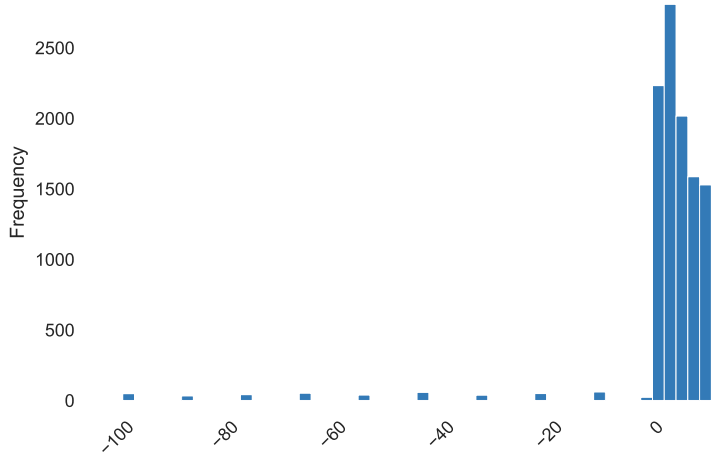


Quantile statistics

Minimum	-99
5-th percentile	0.0409648368
Q1	1.267084672
median	3.269669013
Q3	6.646784488
95-th percentile	8.997049264
Maximum	9.993592355
Range	108.9935924
Interquartile range (IQR)	5.379699816

Descriptive statistics

Standard deviation	12.64396768
Coefficient of variation (CV)	6.764598009
Kurtosis	35.58515565
Mean	1.869138073
Median Absolute Deviation (MAD)	2.269669013
Skewness	-5.722719083
Sum	19822.20926
Variance	159.8699188
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
3	169	1.6%
2	150	1.4%
1	129	1.2%
4	83	0.8%
-11	60	0.6%
-44	57	0.5%
0	56	0.5%
-66	51	0.5%
-22	49	0.5%
-99	48	0.5%
Other values (9546)	9753	92.0%

Value	Count	Frequency (%)
-99	48	0.5%
-88	32	0.3%
-77	41	0.4%
-66	51	0.5%
-55	38	0.4%
-44	57	0.5%
-33	37	0.3%
-22	49	0.5%
-11	60	0.6%
-1	22	0.2%

Value	Count	Frequency (%)
9.993592355	1	< 0.1%
9.991536724	1	< 0.1%
9.986885331	1	< 0.1%
9.985227427	1	< 0.1%
9.978396962	1	< 0.1%
9.97703636	1	< 0.1%
9.97451926	1	< 0.1%
9.971197654	1	< 0.1%
9.969167993	1	< 0.1%
9.966818042	1	< 0.1%

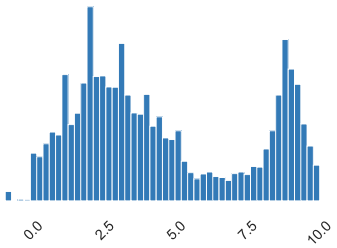
Men's_Clothing

Real number (ℝ)

HIGH

CORRELATION (This variable has a high correlation with 2 fields: Gender, Beauty)

Distinct	9788
Distinct (%)	92.3%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	4.470892531
Minimum	-1
Maximum	9.998657127
Zeros	94
Zeros (%)	0.9%
Negative	44
Negative (%)	0.4%
Memory size	165.7 KiB

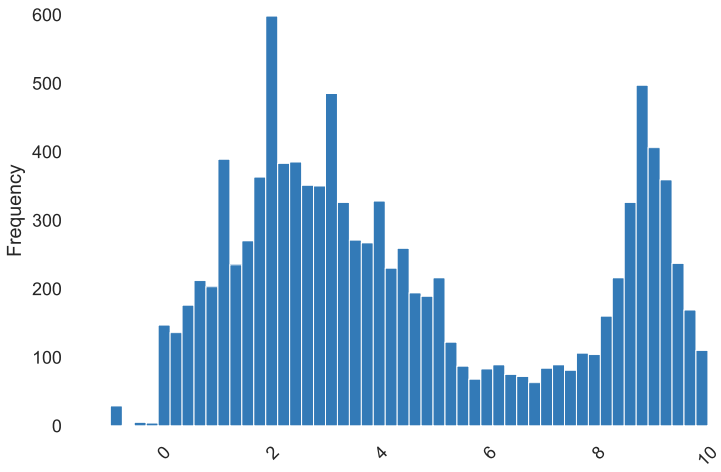


Quantile statistics

Minimum	-1
5-th percentile	0.5789903702
Q1	2.015025202
median	3.608578235
Q3	7.643128559
95-th percentile	9.328764133
Maximum	9.998657127
Range	10.99865713
Interquartile range (IQR)	5.628103357

Descriptive statistics

Standard deviation	2.988305066
Coefficient of variation (CV)	0.6683911646
Kurtosis	-1.184136791
Mean	4.470892531
Median Absolute Deviation (MAD)	1.922100489
Skewness	0.4362584887
Sum	47413.81529
Variance	8.929967166
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
2	189	1.8%
1	168	1.6%
3	160	1.5%
4	109	1.0%
0	94	0.9%
5	76	0.7%
-1	28	0.3%
3.00238395	1	< 0.1%
4.54988263	1	< 0.1%
8.59334966	1	< 0.1%
Other values (9778)	9778	92.2%

Value	Count	Frequency (%)
-1	28	0.3%
-0.995361655	1	< 0.1%
-0.57604872	1	< 0.1%
-0.553047749	1	< 0.1%
-0.431430727	1	< 0.1%
-0.423031187	1	< 0.1%
-0.39808373	1	< 0.1%
-0.343231053	1	< 0.1%
-0.210336054	1	< 0.1%
-0.179911573	1	< 0.1%

Value	Count	Frequency (%)
9.998657127	1	< 0.1%
9.993089419	1	< 0.1%
9.989066393	1	< 0.1%
9.988807781	1	< 0.1%
9.986359846	1	< 0.1%
9.984757189	1	< 0.1%
9.974687068	1	< 0.1%
9.972574219	1	< 0.1%
9.972248798	1	< 0.1%
9.967435796	1	< 0.1%

Kid’s_Clothing

Real number (ℝ)

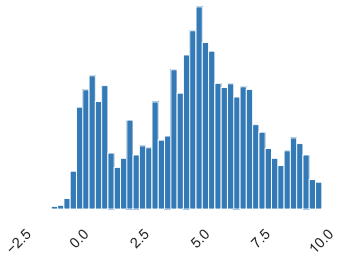
HIGH

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

MISSING

Distinct	9424
Distinct (%)	94.6%
Missing	648
Missing (%)	6.1%
Infinite	0
Infinite (%)	0.0%
Mean	4.525675141

Minimum	-3
Maximum	9.997131718
Zeros	45
Zeros (%)	0.4%
Negative	243
Negative (%)	2.3%
Memory size	165.7 KiB

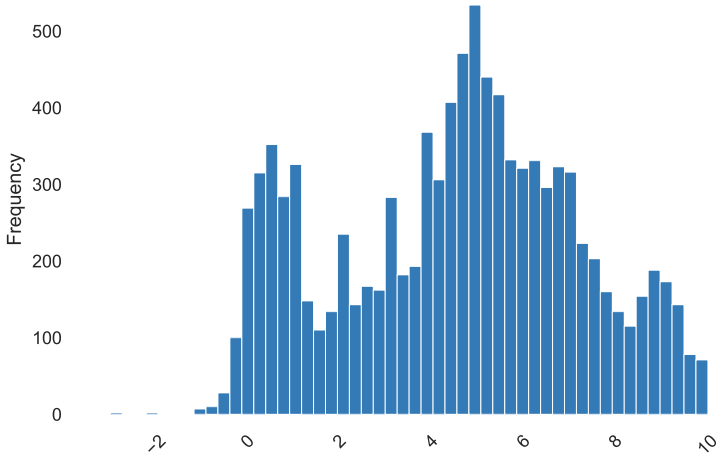


Quantile statistics

Minimum	-3
5-th percentile	0.1881122106
Q1	2.510596223
median	4.785558767
Q3	6.442305978
95-th percentile	8.909567221
Maximum	9.997131718
Range	12.99713172
Interquartile range (IQR)	3.931709755

Descriptive statistics

Standard deviation	2.642213867
Coefficient of variation (CV)	0.5838275582
Kurtosis	-0.8351713816
Mean	4.525675141
Median Absolute Deviation (MAD)	1.875363844
Skewness	-0.1039352094
Sum	45062.14738
Variance	6.981294118
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
3	138	1.3%
2	121	1.1%
1	111	1.0%
4	81	0.8%
0	45	0.4%
5	35	0.3%
-1	7	0.1%
-3	2	< 0.1%
-2	2	< 0.1%
9.52810022	1	< 0.1%
Other values (9414)	9414	88.8%
(Missing)	648	6.1%

Value	Count	Frequency (%)
-3	2	< 0.1%
-2	2	< 0.1%
-1.738291241	1	< 0.1%
-1	7	0.1%
-0.914823057	1	< 0.1%
-0.854238121	1	< 0.1%
-0.852646396	1	< 0.1%
-0.804528415	1	< 0.1%
-0.801865476	1	< 0.1%
-0.755548859	1	< 0.1%

Value	Count	Frequency (%)
9.997131718	1	< 0.1%
9.992701845	1	< 0.1%
9.9850969	1	< 0.1%
9.983727505	1	< 0.1%
9.963532959	1	< 0.1%
9.953272191	1	< 0.1%
9.952704853	1	< 0.1%
9.950593429	1	< 0.1%
9.949291858	1	< 0.1%
9.948478003	1	< 0.1%

Home_&_Living

Real number (ℝ)

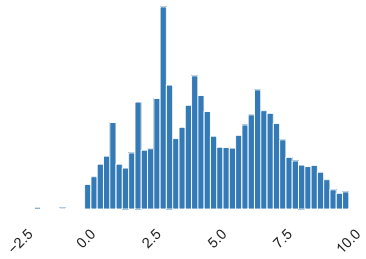
HIGH

CORRELATION (This variable has a high correlation with 2 fields: Gender, Beauty)

MISSING

Distinct	9555
Distinct (%)	95.5%
Missing	595
Missing (%)	5.6%
Infinite	0
Infinite (%)	0.0%
Mean	4.587931826

Minimum	-2
Maximum	9.993162903
Zeros	20
Zeros (%)	0.2%
Negative	10
Negative (%)	0.1%
Memory size	165.7 KiB

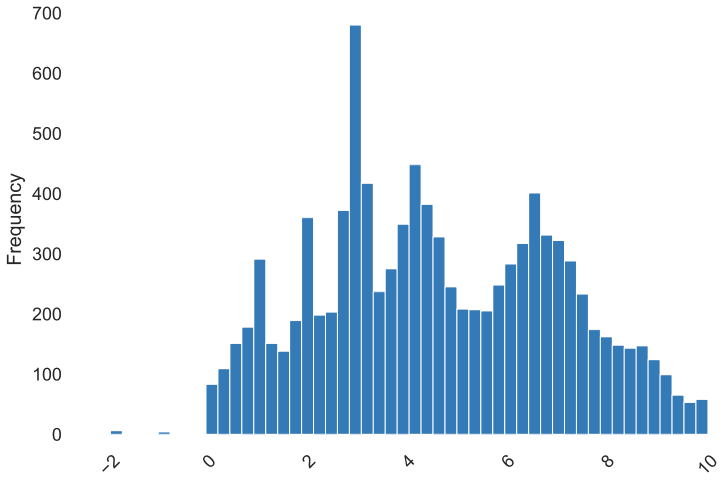


Quantile statistics

Minimum	-2
5-th percentile	0.8356122508
Q1	2.826419752
median	4.336520475
Q3	6.534797656
95-th percentile	8.619527757
Maximum	9.993162903
Range	11.9931629
Interquartile range (IQR)	3.708377904

Descriptive statistics

Standard deviation	2.386965527
Coefficient of variation (CV)	0.5202704874
Kurtosis	-0.8624109066
Mean	4.587931826
Median Absolute Deviation (MAD)	1.860994035
Skewness	0.1324565472
Sum	45925.19758
Variance	5.697604428
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
2	143	1.3%
3	128	1.2%
1	106	1.0%
4	44	0.4%
0	20	0.2%
5	12	0.1%
-2	6	0.1%
-1	4	< 0.1%
2.991463752	1	< 0.1%
5.158428449	1	< 0.1%
Other values (9545)	9545	90.0%
(Missing)	595	5.6%

Value	Count	Frequency (%)
-2	6	0.1%
-1	4	< 0.1%
0	20	0.2%
0.006363999	1	< 0.1%
0.007080552	1	< 0.1%
0.007560413	1	< 0.1%
0.008278316	1	< 0.1%
0.019453277	1	< 0.1%
0.021871199	1	< 0.1%
0.025831449	1	< 0.1%

Value	Count	Frequency (%)
9.993162903	1	< 0.1%
9.987811655	1	< 0.1%
9.978429283	1	< 0.1%
9.975933723	1	< 0.1%
9.975144983	1	< 0.1%
9.960149435	1	< 0.1%
9.950582866	1	< 0.1%
9.94626648	1	< 0.1%
9.945071339	1	< 0.1%
9.943942915	1	< 0.1%

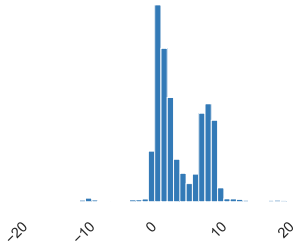
Beauty

Real number (ℝ)

HIGH

CORRELATION (This variable has a high correlation with 3 fields: Gender, Men's Clothing, Home & Living)

Distinct	9590
Distinct (%)	90.4%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	4.104913015
Minimum	-20.11429444
Maximum	26.56484548
Zeros	78
Zeros (%)	0.7%
Negative	222
Negative (%)	2.1%
Memory size	165.7 KiB

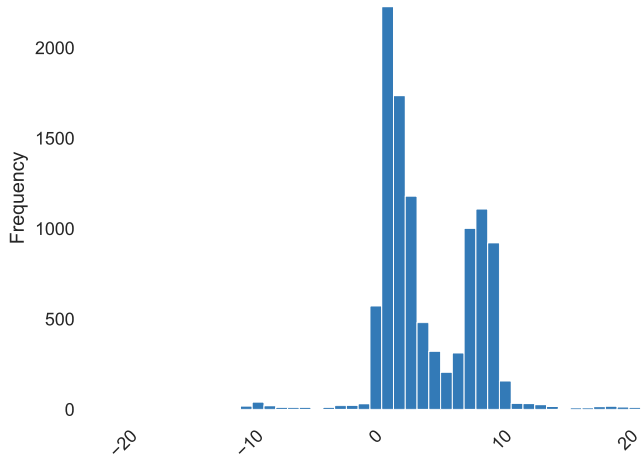


Quantile statistics

Minimum	-20.11429444
5-th percentile	0.2607677922
Q1	1.139450341
median	2.925936847
Q3	7.680576398
95-th percentile	9.510077377
Maximum	26.56484548
Range	46.67913992
Interquartile range (IQR)	6.541126057

Descriptive statistics

Standard deviation	3.783097012
Coefficient of variation (CV)	0.9216022357
Kurtosis	1.336495442
Mean	4.104913015
Median Absolute Deviation (MAD)	2.179986036
Skewness	0.3017209468
Sum	43532.60253
Variance	14.311823
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
3	263	2.5%
2	258	2.4%
1	224	2.1%
4	122	1.2%
0	78	0.7%
5	58	0.5%
-1	17	0.2%
-2	3	< 0.1%
8.272407324	1	< 0.1%
1.033421213	1	< 0.1%
Other values (9580)	9580	90.3%

Value	Count	Frequency (%)
-20.11429444	1	< 0.1%
-13.12891171	1	< 0.1%
-10.89758622	1	< 0.1%
-10.51398739	1	< 0.1%
-10.40693541	1	< 0.1%
-10.33904921	1	< 0.1%
-10.31875325	1	< 0.1%
-10.31526621	1	< 0.1%
-10.25881909	1	< 0.1%
-10.2555728	1	< 0.1%

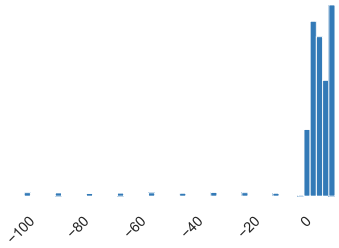
Value	Count	Frequency (%)
26.56484548	1	< 0.1%
21.02241148	1	< 0.1%
20.54926702	1	< 0.1%
20.49639069	1	< 0.1%
20.38993043	1	< 0.1%
20.27106068	1	< 0.1%
20.17834556	1	< 0.1%
20.15332571	1	< 0.1%
20.13967585	1	< 0.1%
20.08014588	1	< 0.1%

Electronics

Real number (ℝ)

Distinct	9486
Distinct (%)	89.4%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	2.78312252

Minimum	-99
Maximum	9.999038006
Zeros	58
Zeros (%)	0.5%
Negative	437
Negative (%)	4.1%
Memory size	165.7 KiB

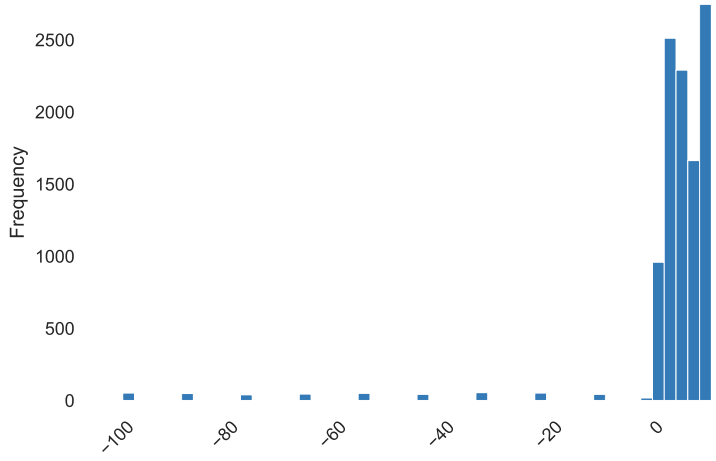


Quantile statistics

Minimum	-99
5-th percentile	0.1236793284
Q1	2.483113648
median	4.624030093
Q3	7.885223725
95-th percentile	9.37254741
Maximum	9.999038006
Range	108.999038
Interquartile range (IQR)	5.402110077

Descriptive statistics

Standard deviation	13.33407875
Coefficient of variation (CV)	4.791049857
Kurtosis	33.29531546
Mean	2.78312252
Median Absolute Deviation (MAD)	2.650374132
Skewness	-5.577660124
Sum	29515.01433
Variance	177.7976562
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
2	176	1.7%
1	153	1.4%
3	143	1.3%
4	101	1.0%
0	58	0.5%
-33	54	0.5%
-99	51	0.5%
-22	51	0.5%
-55	49	0.5%
-88	48	0.5%
Other values (9476)	9721	91.7%

Value	Count	Frequency (%)
-99	51	0.5%
-88	48	0.5%
-77	39	0.4%
-66	44	0.4%
-55	49	0.5%
-44	42	0.4%
-33	54	0.5%
-22	51	0.5%
-11	42	0.4%
-2	1	< 0.1%

Value	Count	Frequency (%)
9.999038006	29	0.3%
9.998377717	1	< 0.1%
9.990763628	1	< 0.1%
9.990607046	1	< 0.1%
9.990334567	1	< 0.1%
9.989220271	1	< 0.1%
9.988176922	1	< 0.1%
9.986931834	1	< 0.1%
9.977993062	1	< 0.1%
9.975847452	1	< 0.1%

Old_UI

Categorical

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	165.7 KiB

Length

Max length	1
Median length	1
Mean length	1
Min length	1

Characters and Unicode

Total characters	10605	
Distinct characters	2	
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	0	?
Unique (%)	0.0%	

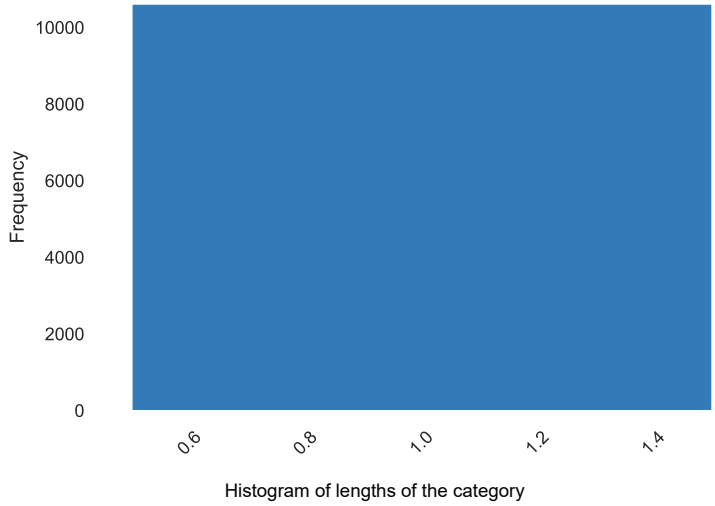
Sample

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

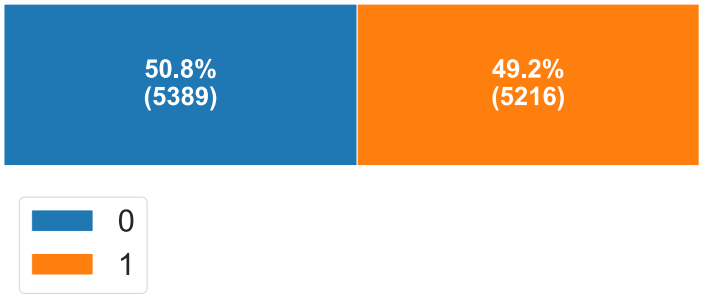
Common Values

Value	Count	Frequency (%)
0	5389	50.8%
1	5216	49.2%

Length



Category Frequency Plot



Value	Count	Frequency (%)
0	5389	50.8%
1	5216	49.2%

Most occurring characters

Value	Count	Frequency (%)
0	5389	50.8%
1	5216	49.2%

Most occurring categories

Value	Count	Frequency (%)
Decimal Number	10605	100.0%

Most frequent character per category

Decimal Number

Value	Count	Frequency (%)
0	5389	50.8%
1	5216	49.2%

Most occurring scripts

Value	Count	Frequency (%)
Common	10605	100.0%

Most frequent character per script

Common

Value	Count	Frequency (%)
0	5389	50.8%
1	5216	49.2%

Most occurring blocks

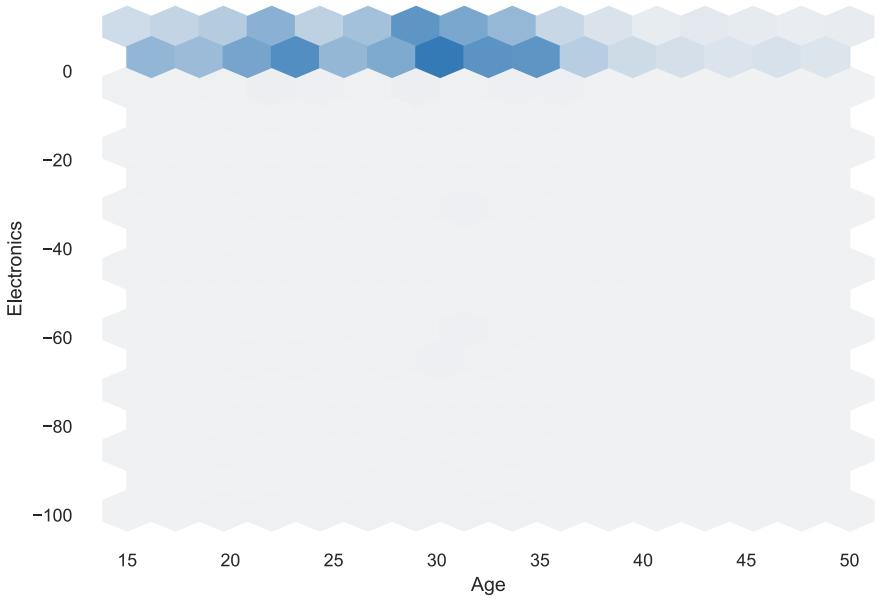
Value	Count	Frequency (%)
ASCII	10605	100.0%

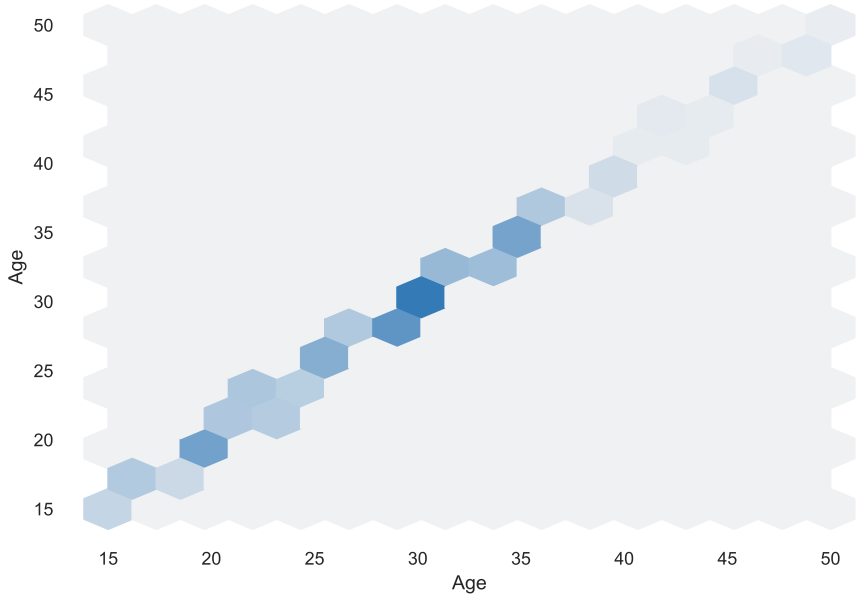
Most frequent character per block

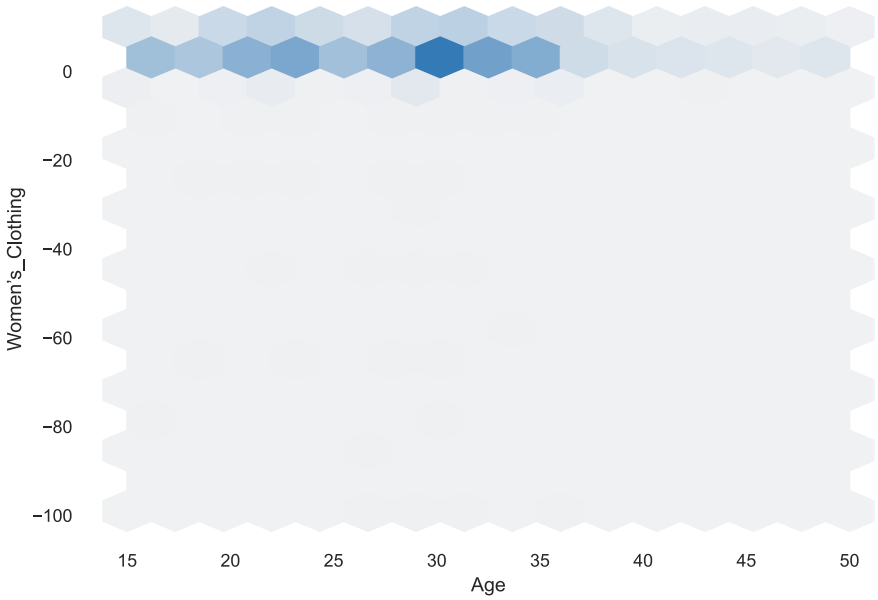
ASCII

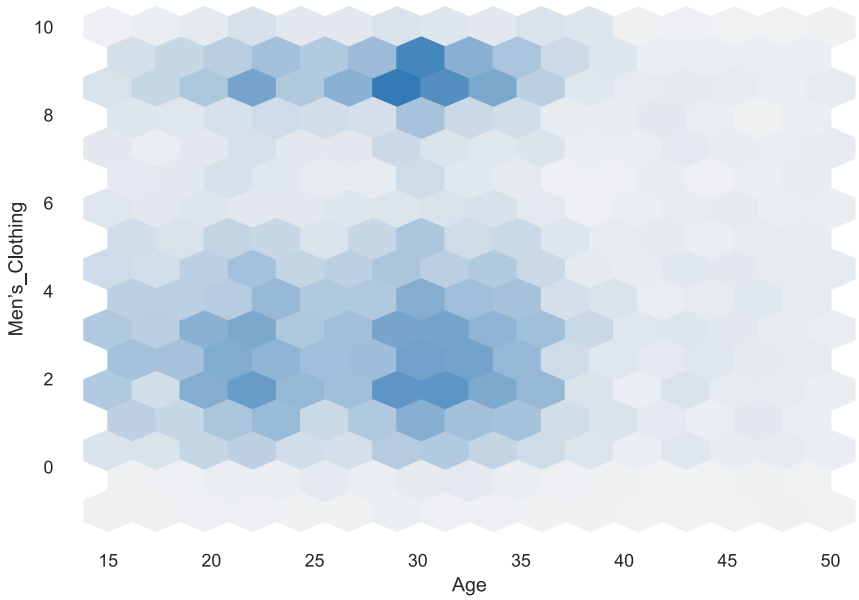
Value	Count	Frequency (%)
0	5389	50.8%
1	5216	49.2%

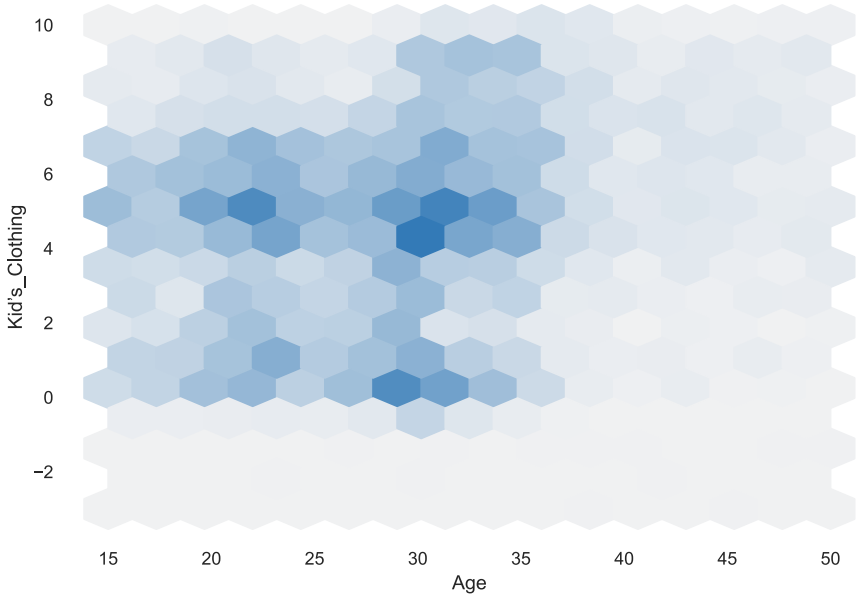
Interactions

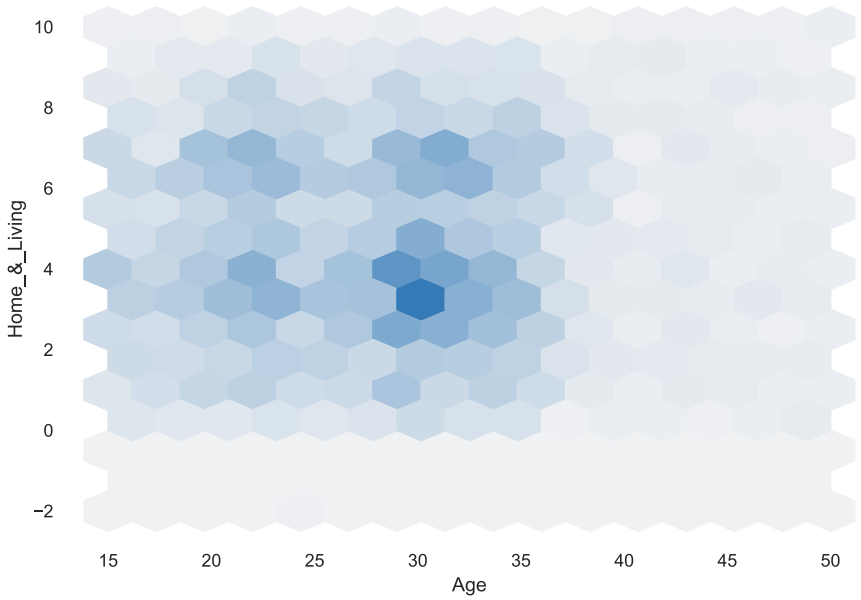


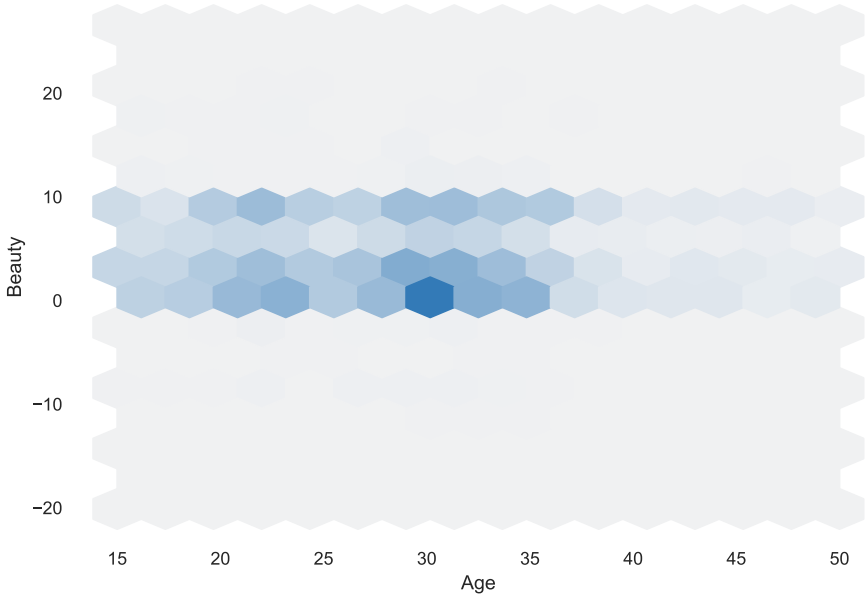

50
45
40
35
30
25
20
15
Age
15
20
25
30
35
40
45
50
Age

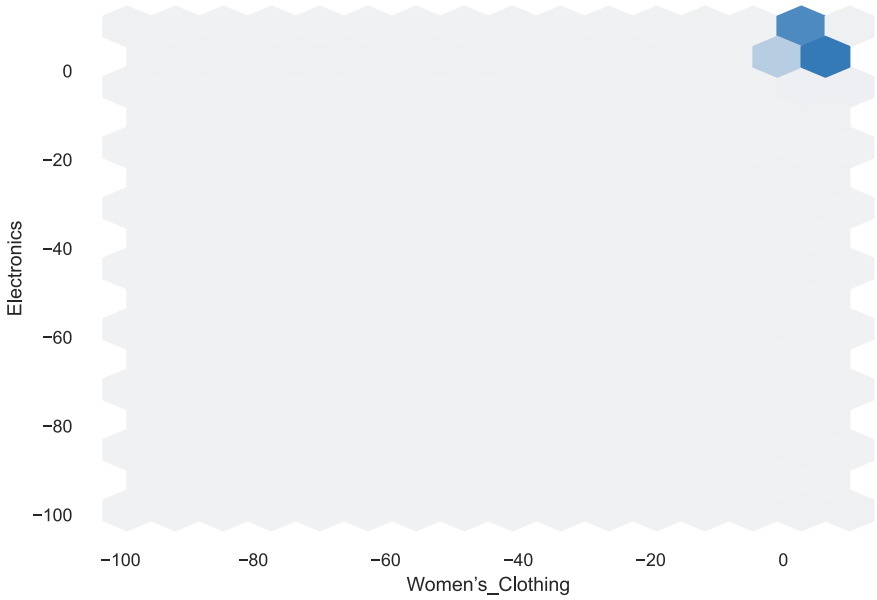

0
−20
−40
−60
−80
−100
Women's_Clothing
15
20
25
30
35
40
45
50
Age

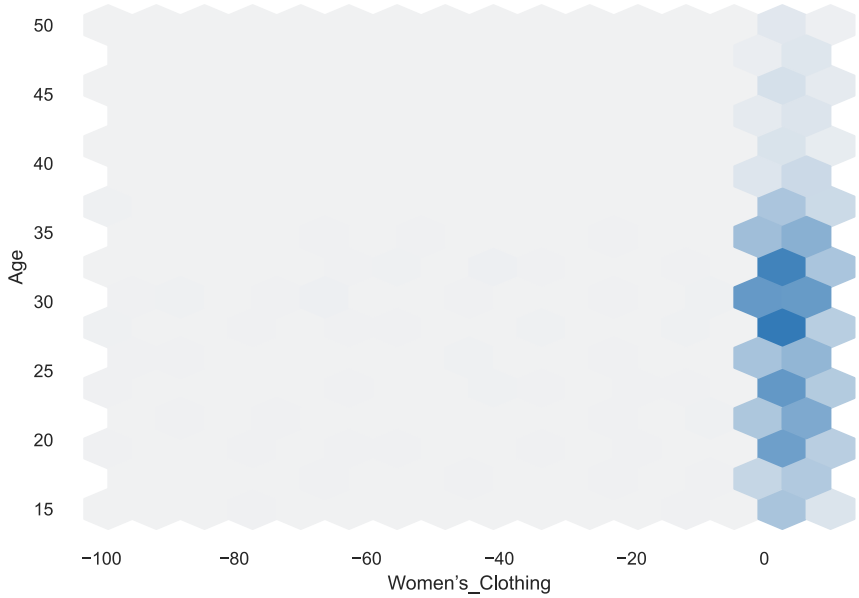

10
8
6
4
2
0
Men's_Clothing
15
20
25
30
35
40
45
50
Age

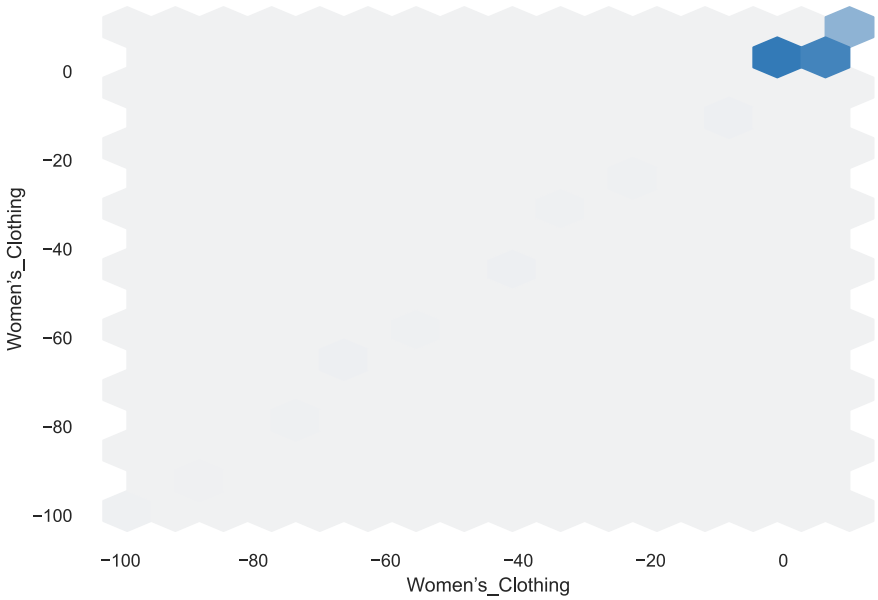

10
8
6
4
2
0
−2
Kid's_Clothing
15
20
25
30
35
40
45
50
Age

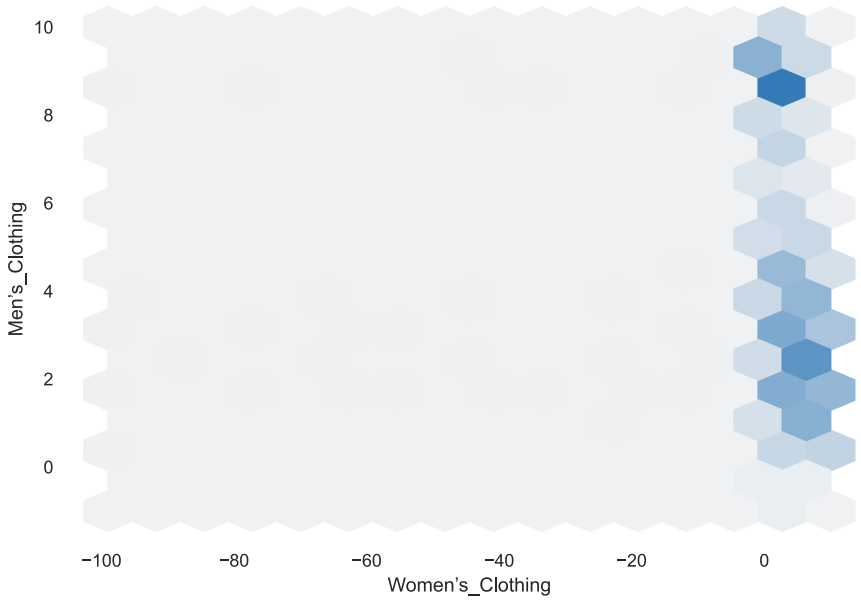

10
8
6
4
2
0
−2
Home_&_Living
15
20
25
30
35
40
45
50
Age

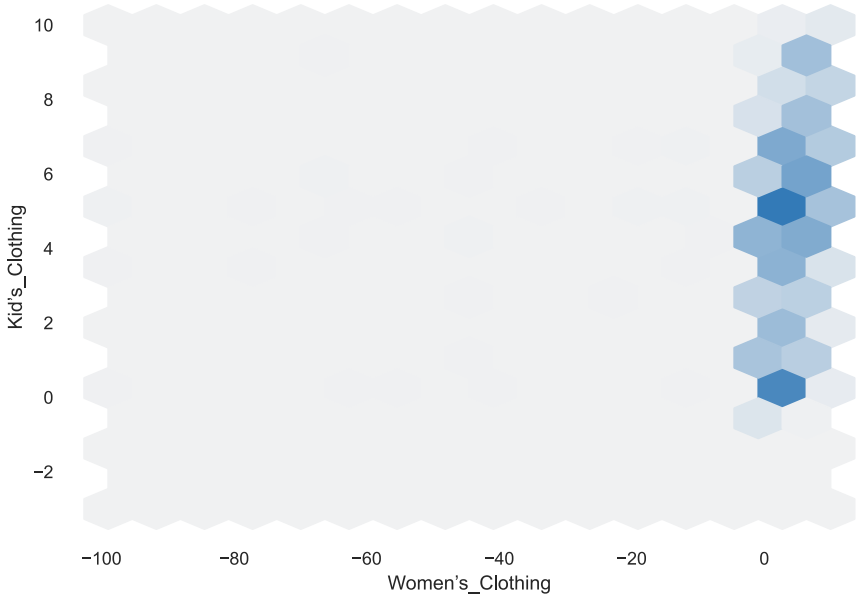

20
10
0
−10
−20
Beauty
15
20
25
30
35
40
45
50
Age

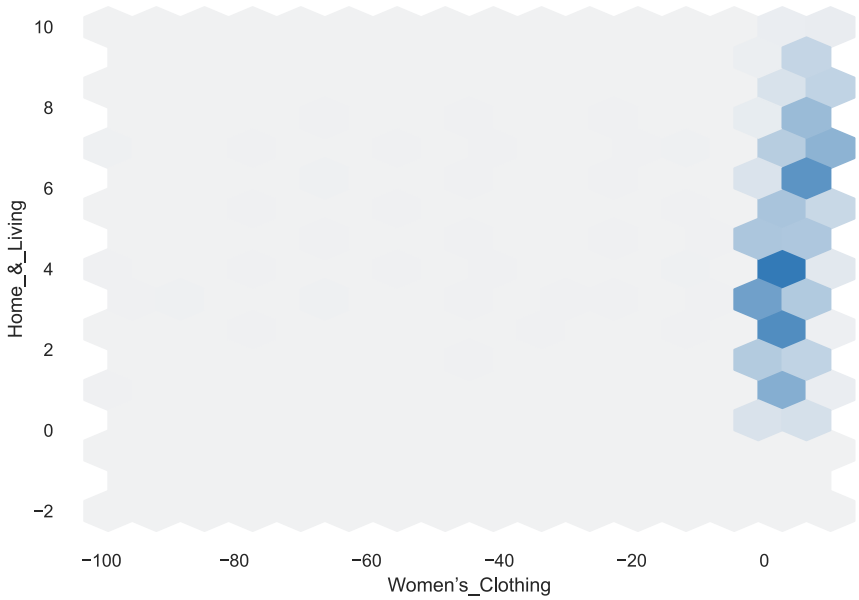

0
−20
−40
−60
−80
−100
Electronics
−100
−80
−60
−40
−20
0
Women’s_Clothing

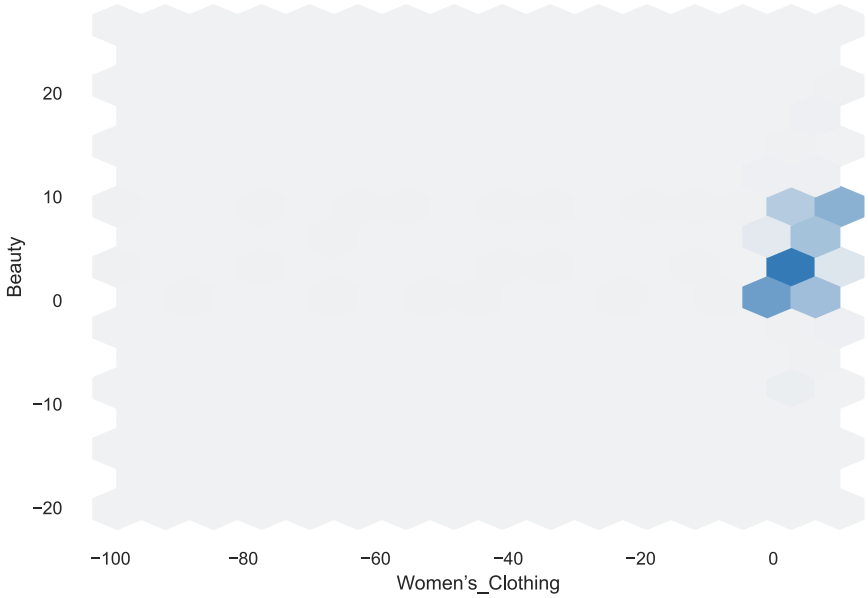

50
45
40
35
30
25
20
15
Age
−100
−80
−60
−40
−20
0
Women's_Clothing

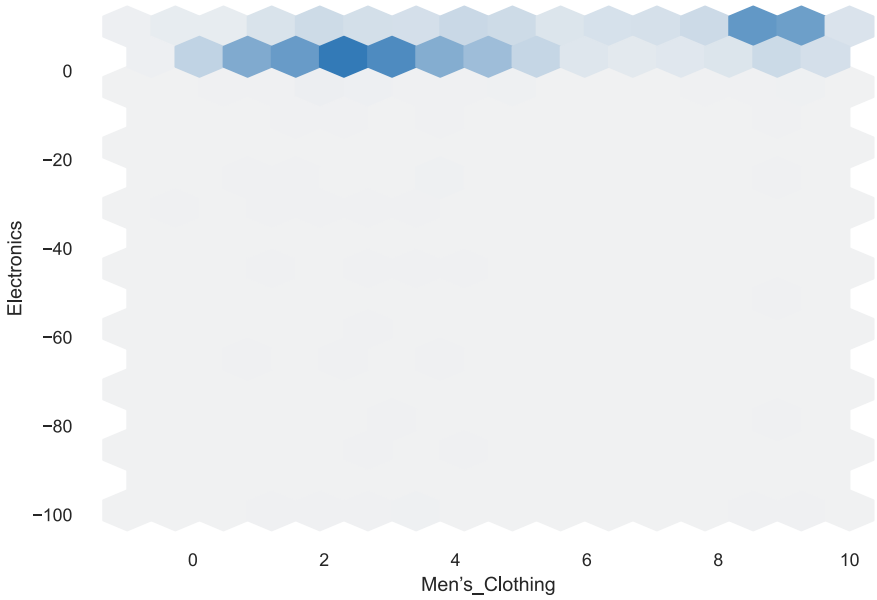

0
−20
−40
−60
−80
−100
Women's_Clothing
−100
−80
−60
−40
−20
0
Women's_Clothing

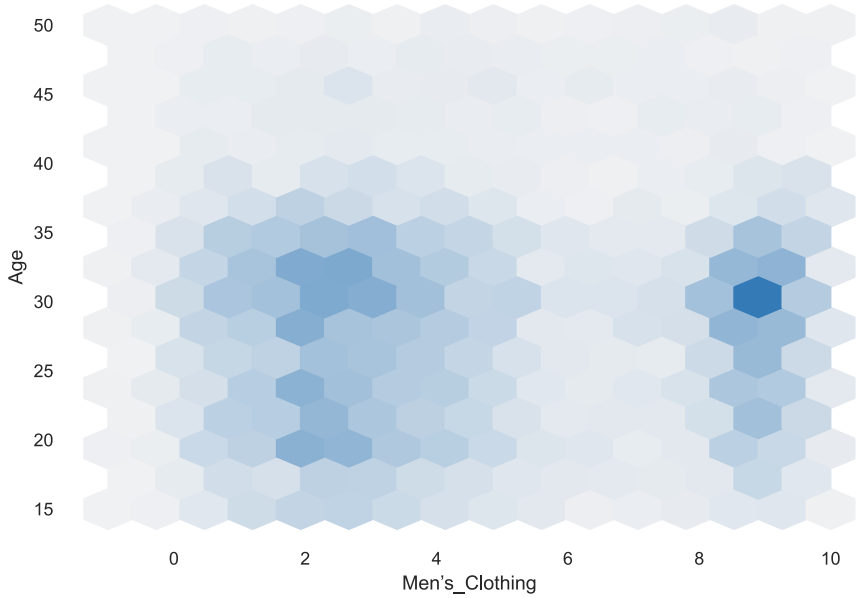

10
8
6
4
2
0
Men's_Clothing
−100
−80
−60
−40
−20
0
Women's_Clothing

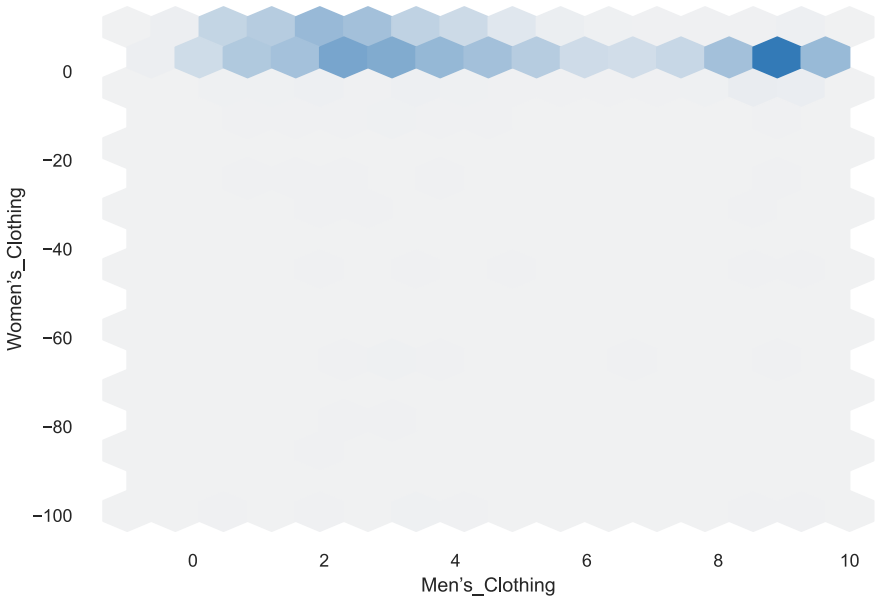

10
8
6
4
2
0
−2
Kid's_Clothing
−100
−80
−60
−40
−20
0
Women's_Clothing

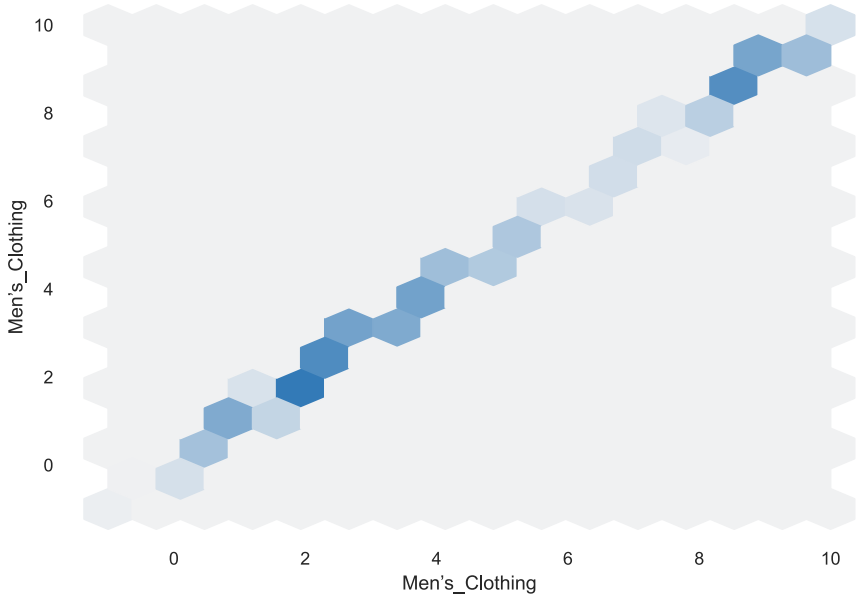

10
8
6
4
2
0
−2
Home_&_Living
−100
−80
−60
−40
−20
0
Women's_Clothing

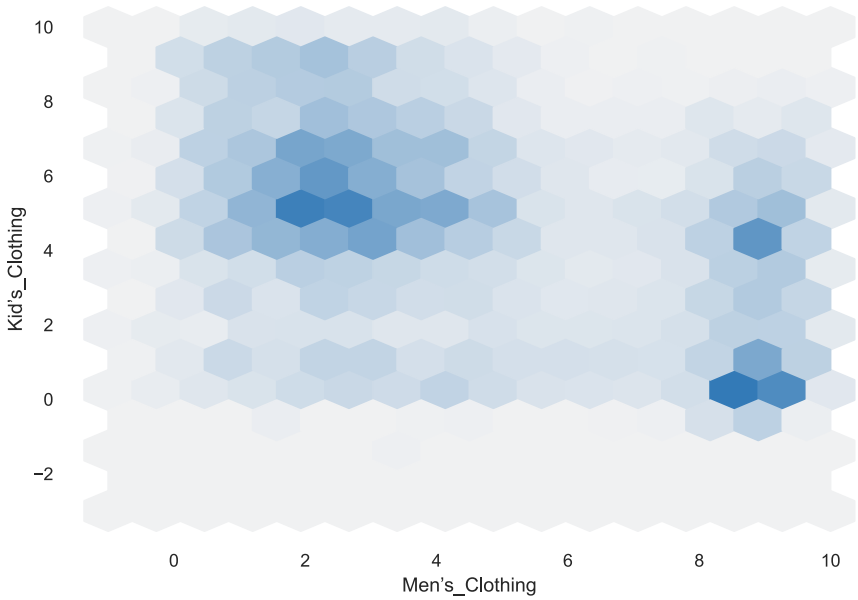

20
10
0
−10
−20
Beauty
−100
−80
−60
−40
−20
0
Women's_Clothing

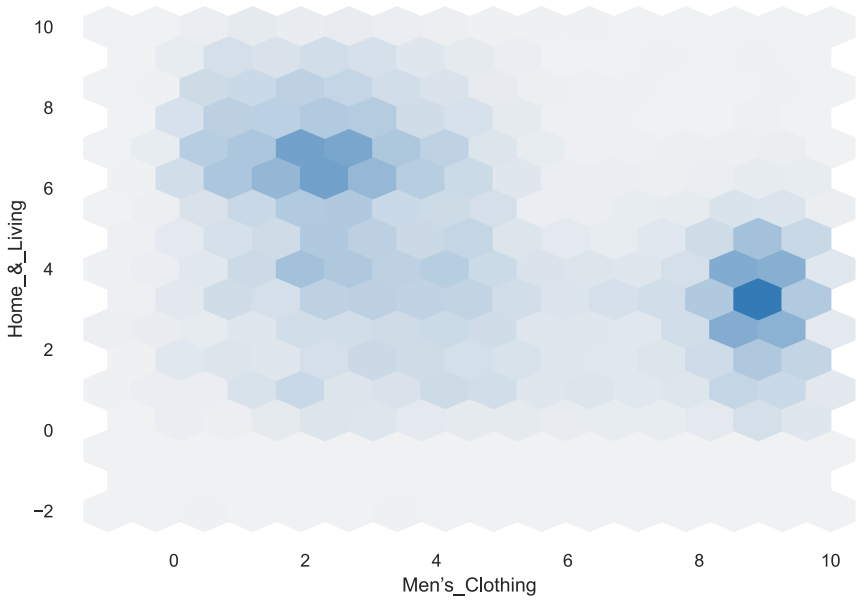

0
−20
−40
−60
−80
−100
Electronics
0
2
4
6
8
10
Men’s_Clothing

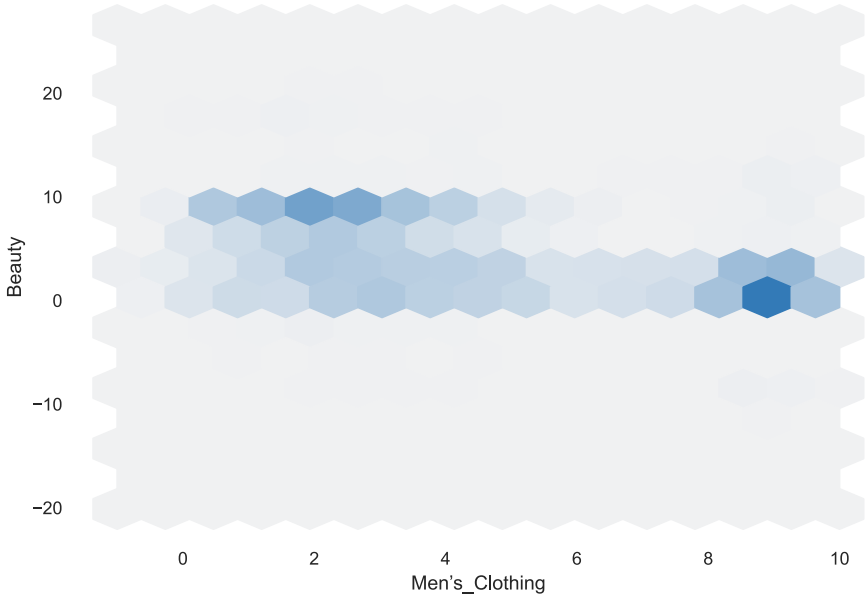

50
45
40
35
30
25
20
15
Age
0
2
4
6
8
10
Men’s_Clothing

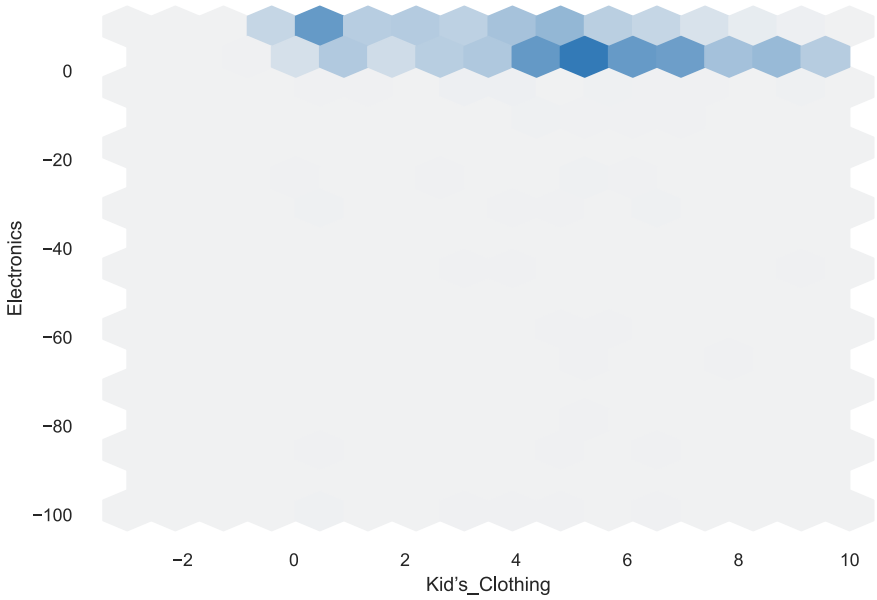

0
−20
−40
−60
−80
−100
Women's_Clothing
0
2
4
6
8
10
Men's_Clothing

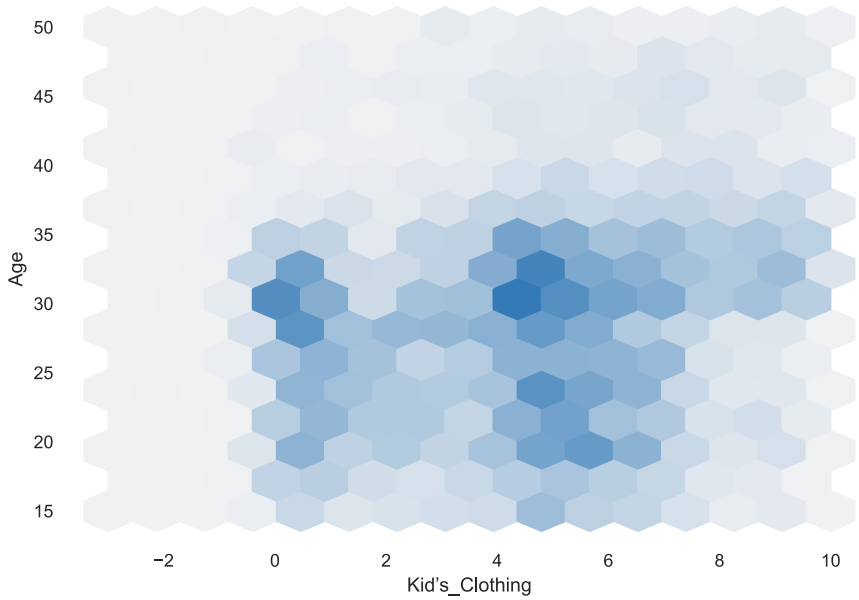

10
8
6
4
2
0
Men's_Clothing
0
2
4
6
8
10
Men's_Clothing

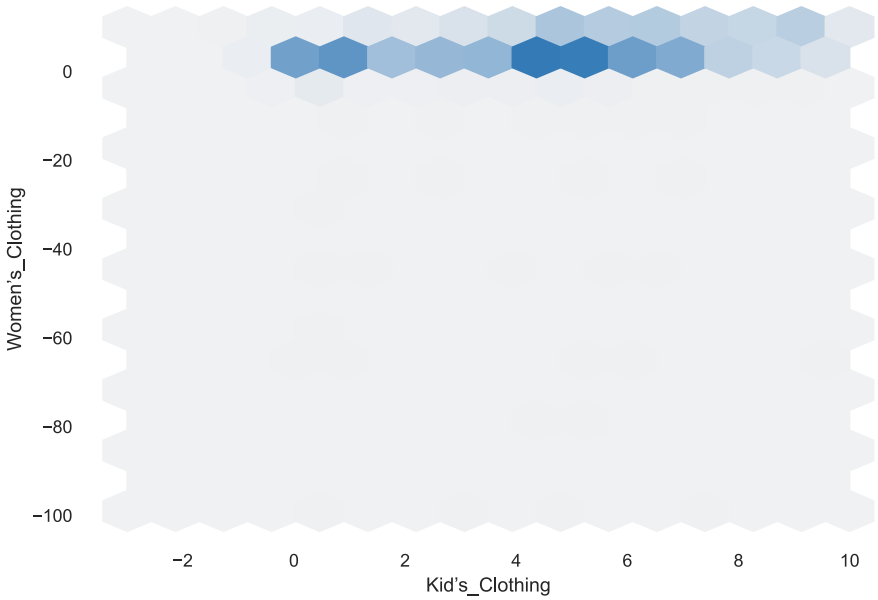

10
8
6
4
2
0
−2
Kid's_Clothing
0
2
4
6
8
10
Men's_Clothing

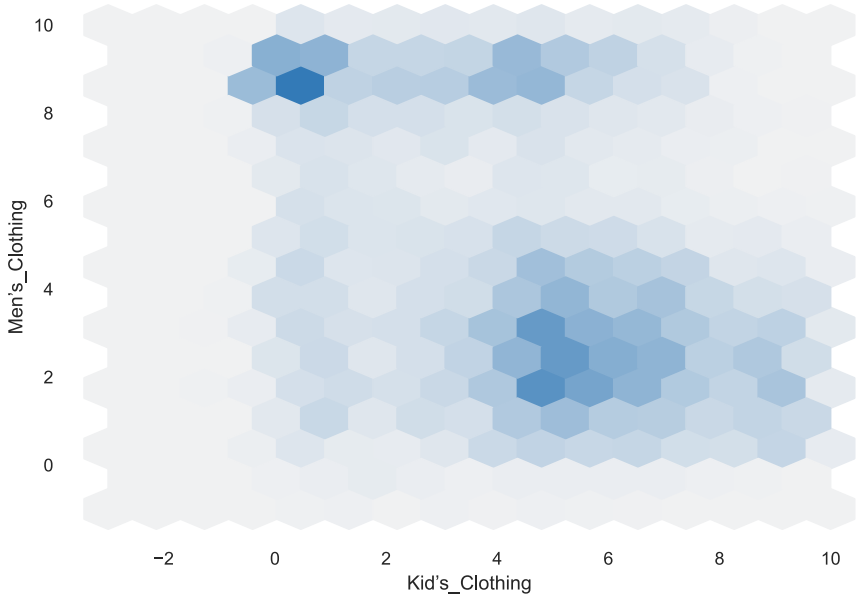

10
8
6
4
2
0
−2
Home_&_Living
0
2
4
6
8
10
Men's_Clothing

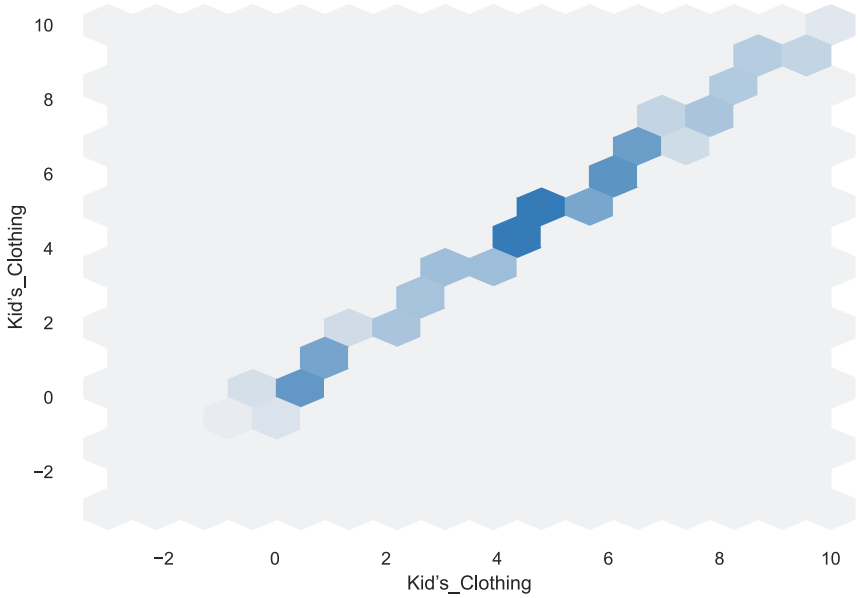

20
10
0
−10
−20
Beauty
0
2
4
6
8
10
Men’s_Clothing

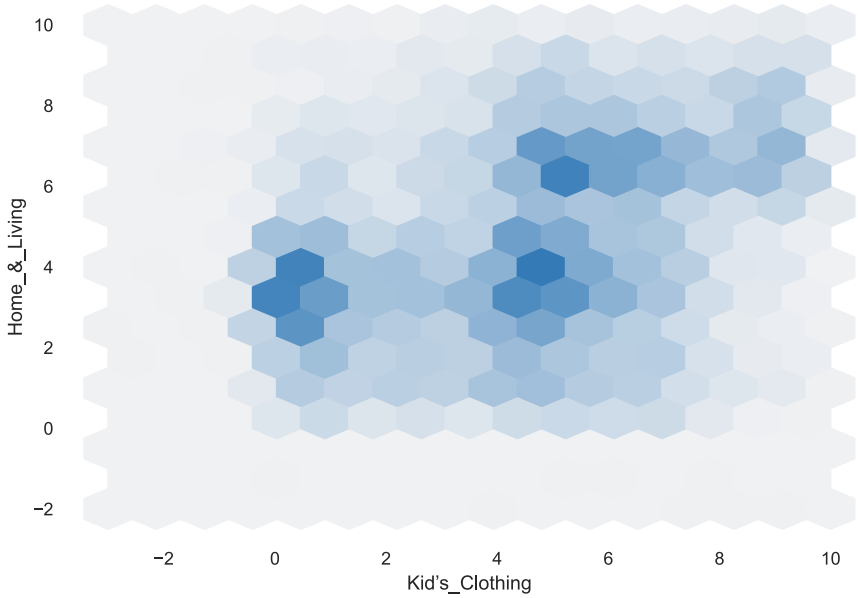

0
−20
−40
−60
−80
−100
Electronics
−2
0
2
4
6
8
10
Kid's_Clothing

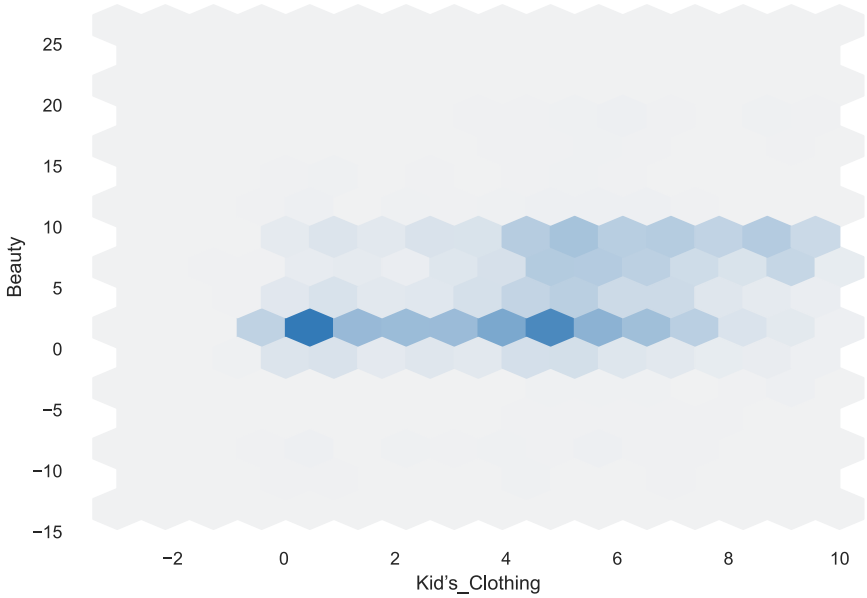

50
45
40
35
30
25
20
15
Age
−2
0
2
4
6
8
10
Kid's_Clothing

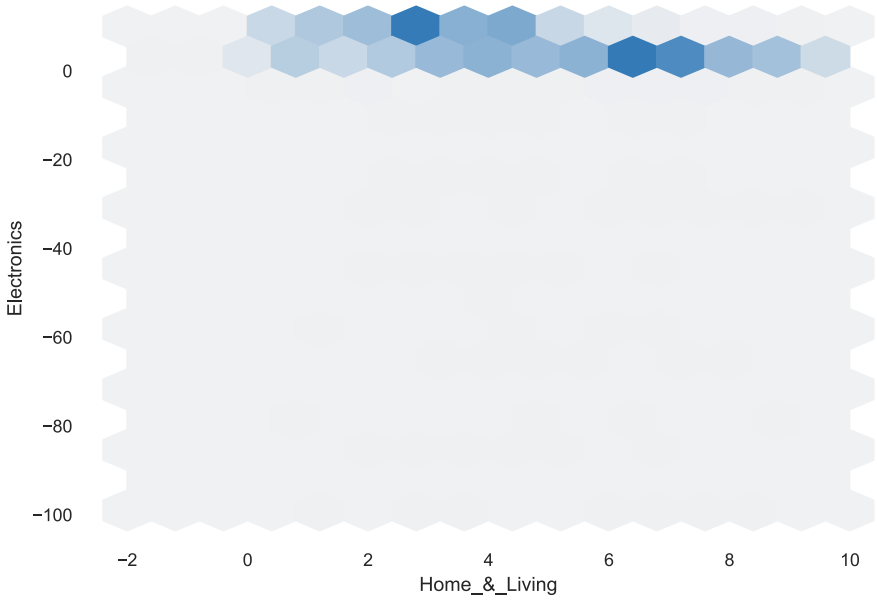

0
−20
−40
−60
−80
−100
Women's_Clothing
−2
0
2
4
6
8
10
Kid's_Clothing

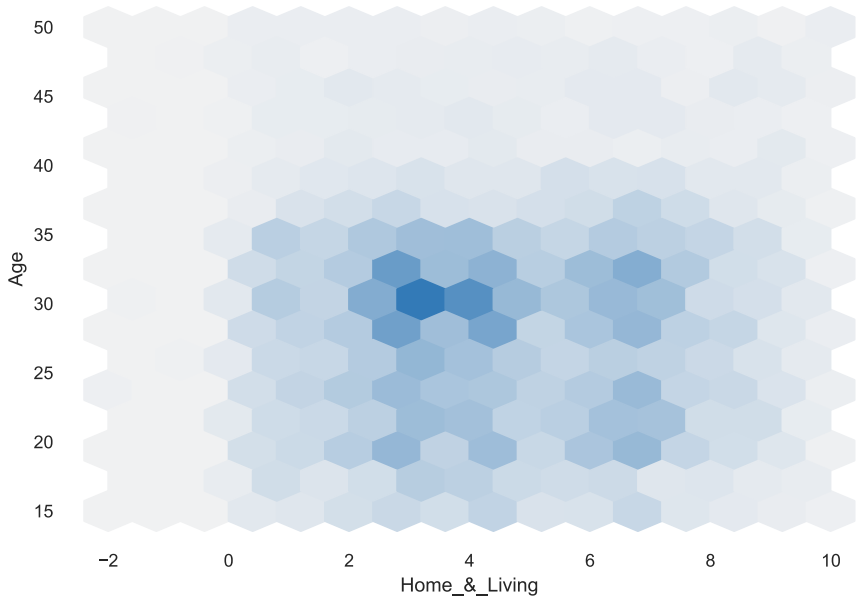

10
8
6
4
2
0
Men's_Clothing
−2
0
2
4
6
8
10
Kid's_Clothing

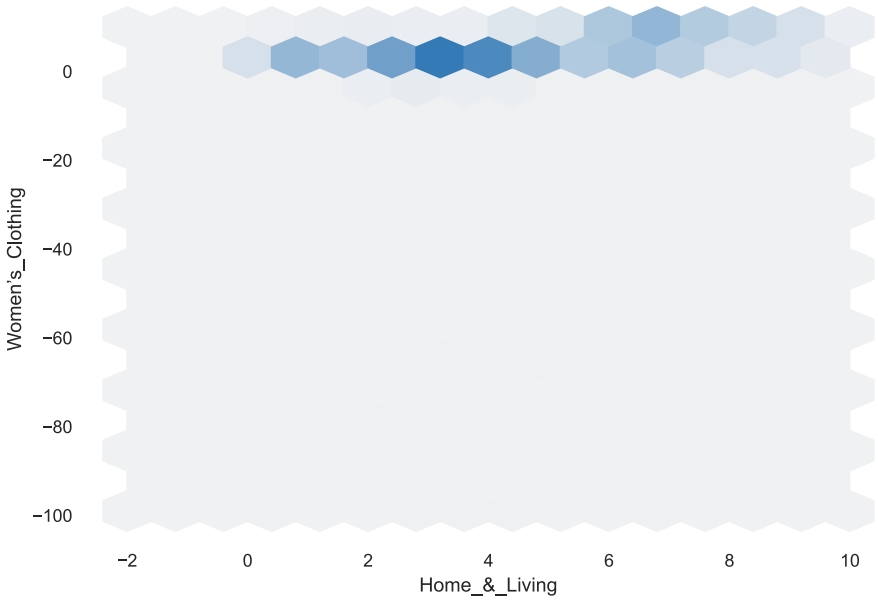

10
8
6
4
2
0
−2
Kid's_Clothing
−2
0
2
4
6
8
10
Kid's_Clothing

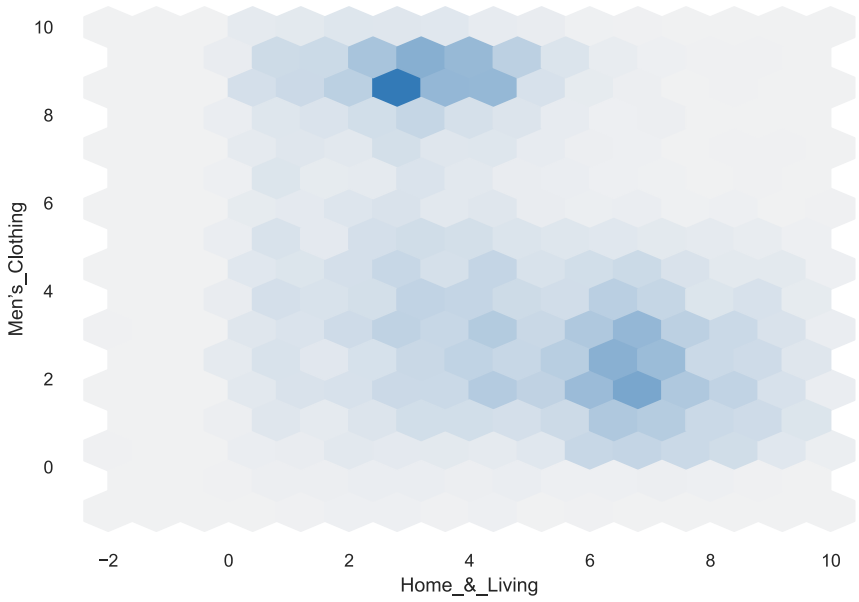

10
8
6
4
2
0
−2
Home_&_Living
−2
0
2
4
6
8
10
Kid's_Clothing

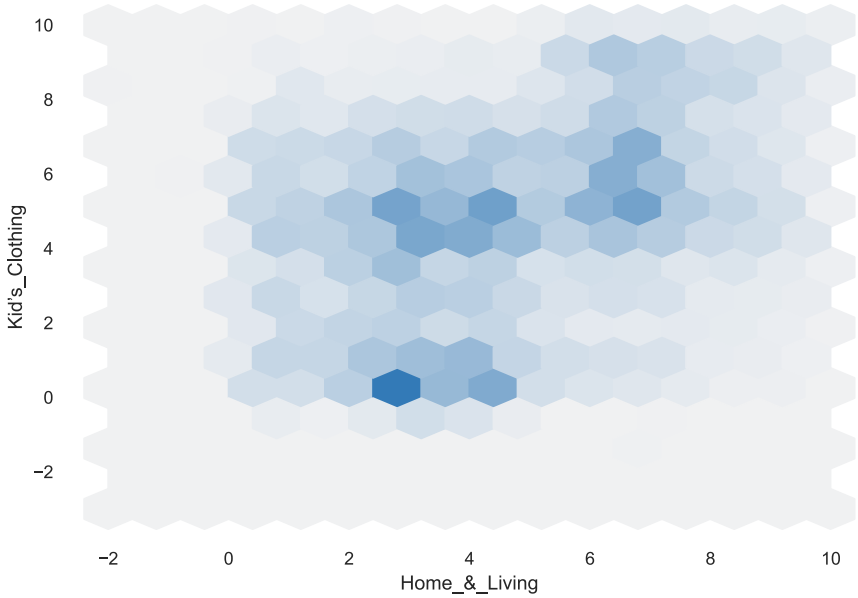

25
20
15
10
5
0
−5
−10
−15
Beauty
−2
0
2
4
6
8
10
Kid's_Clothing

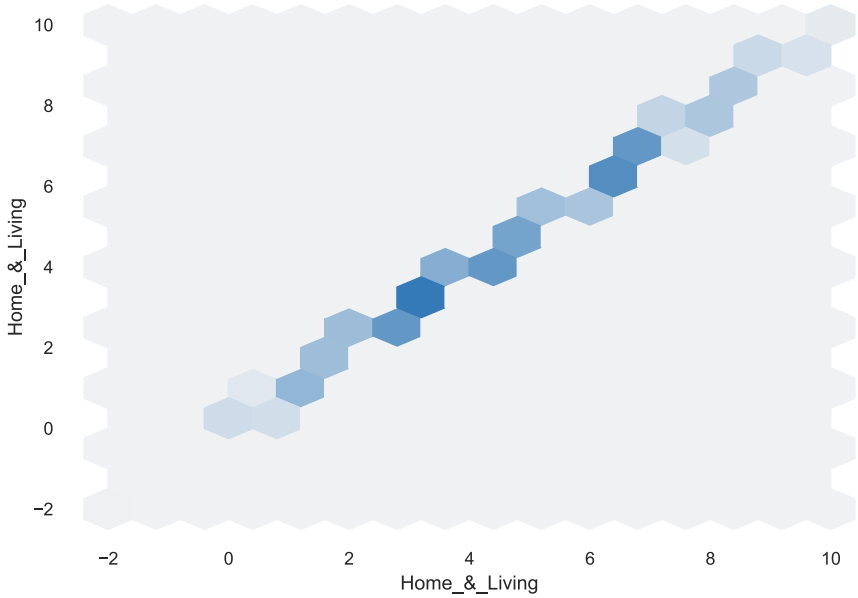

0
−20
−40
−60
−80
−100
Electronics
−2
0
2
4
6
8
10
Home_&_Living

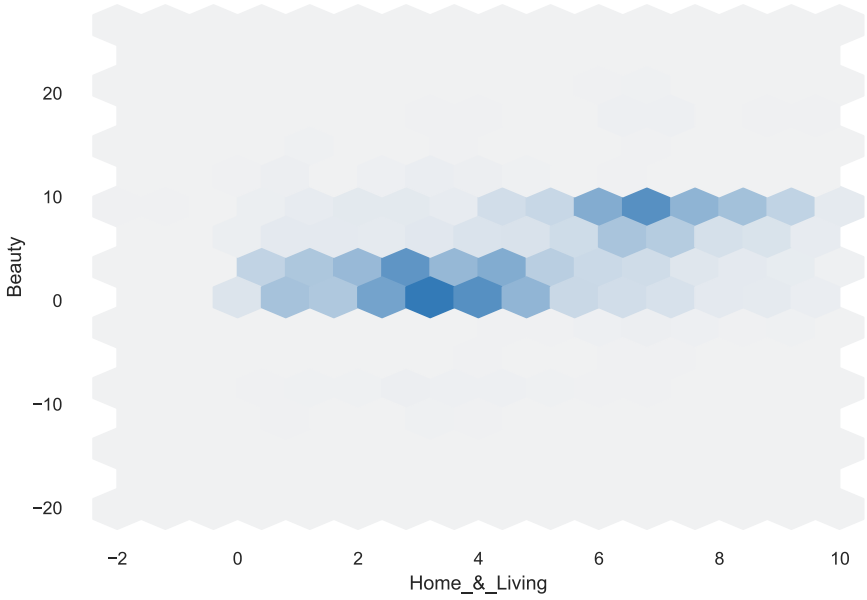

50
45
40
35
30
25
20
15
Age
−2
0
2
4
6
8
10
Home_&_Living

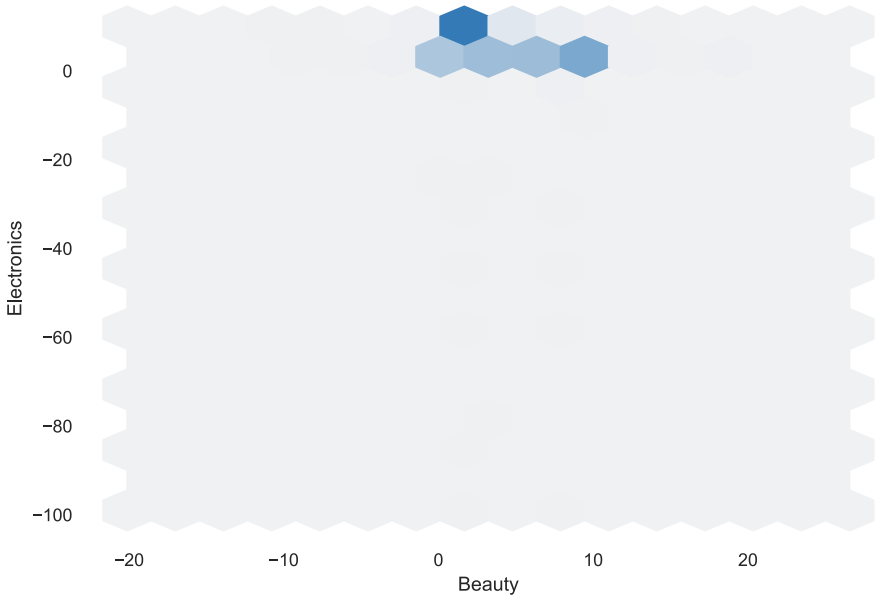

0
−20
−40
−60
−80
−100
Women's_Clothing
−2
0
2
4
6
8
10
Home_&_Living

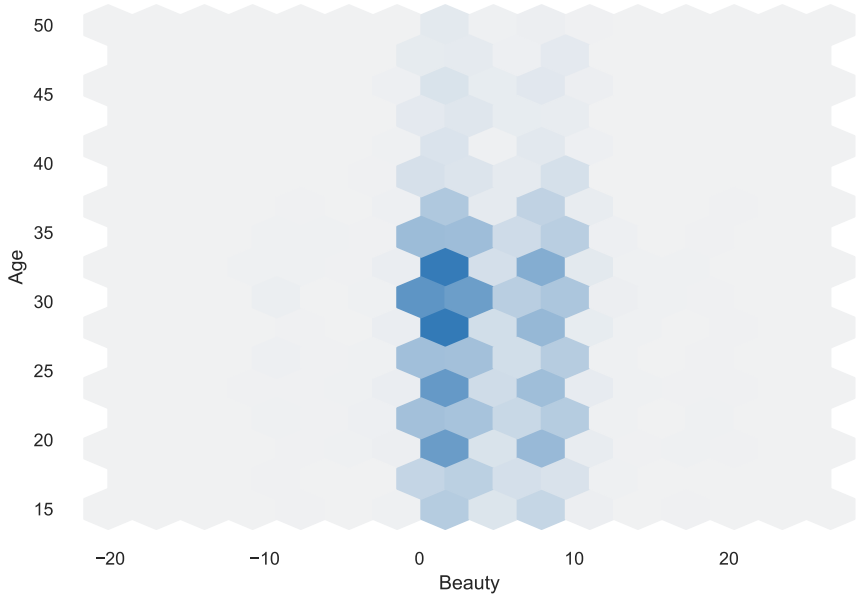

10
8
6
4
2
0
Men's_Clothing
−2
0
2
4
6
8
10
Home_&_Living

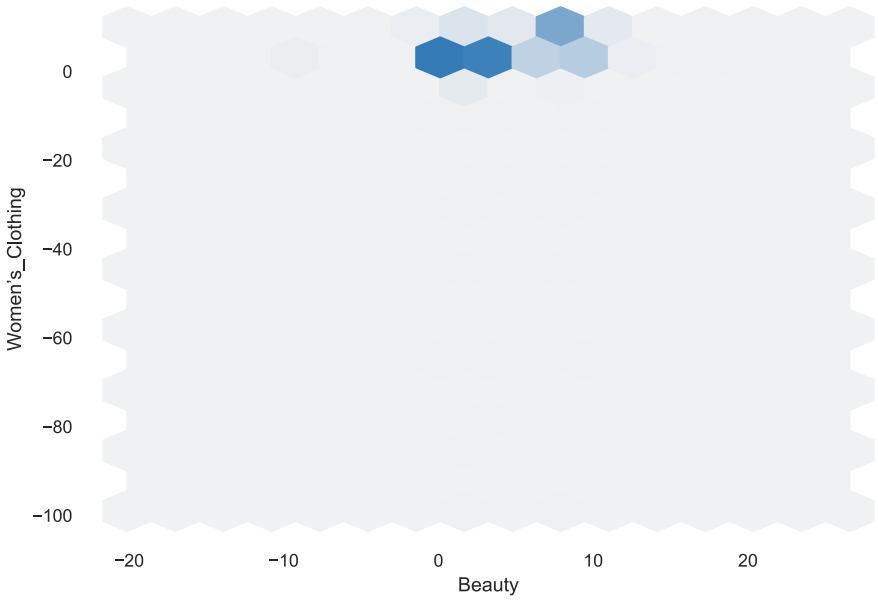

10
8
6
4
2
0
−2
Kid's_Clothing
−2
0
2
4
6
8
10
Home_&_Living

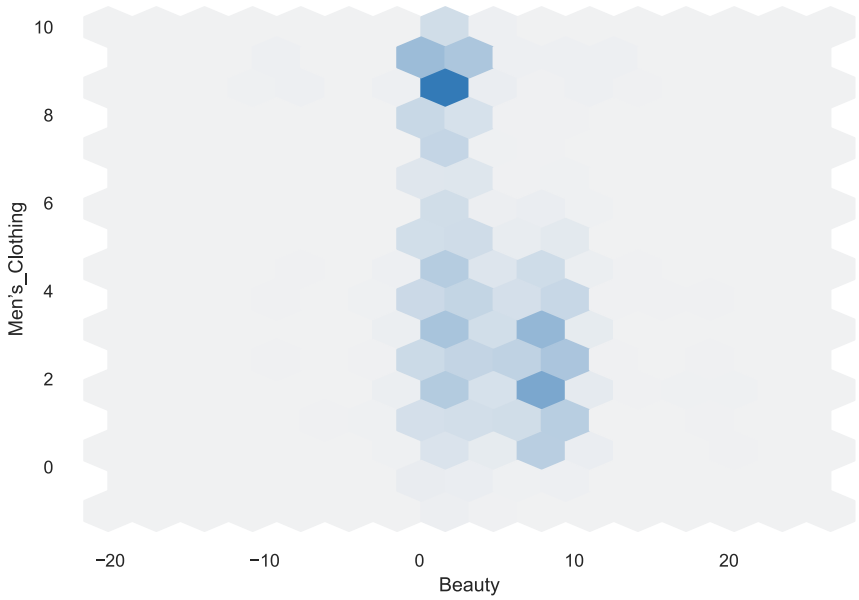

10
8
6
4
2
0
−2
Home_&_Living
−2
0
2
4
6
8
10
Home_&_Living

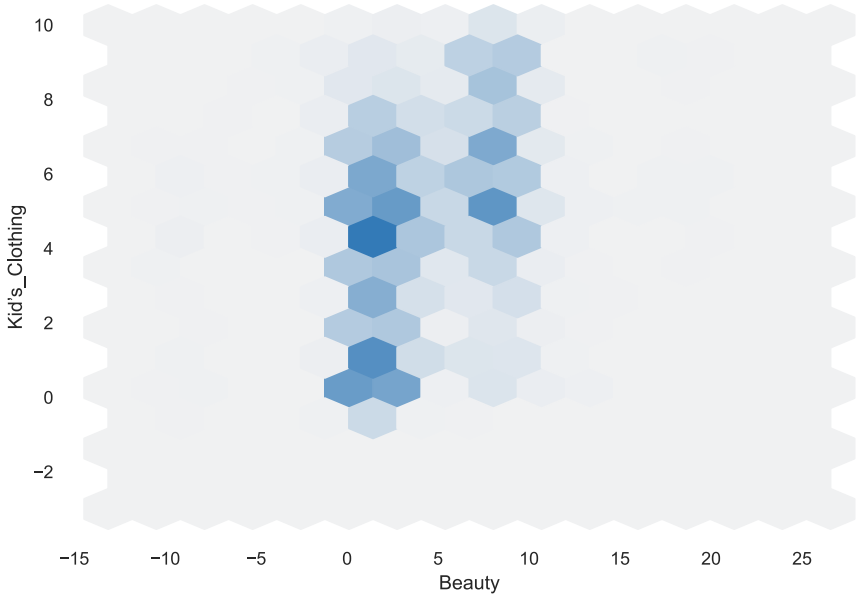

20
10
0
−10
−20
Beauty
−2
0
2
4
6
8
10
Home_&_Living

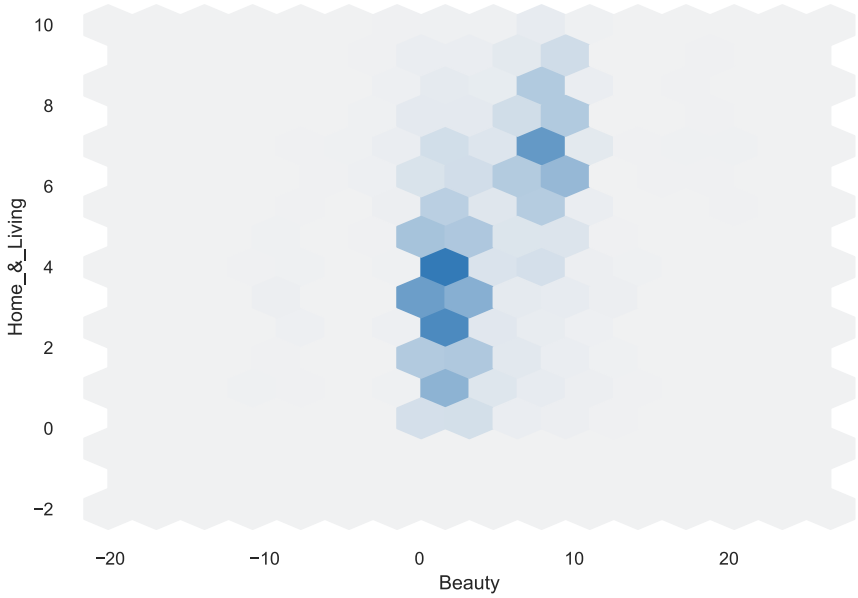

0
−20
−40
−60
−80
−100
Electronics
−20
−10
0
10
20
Beauty

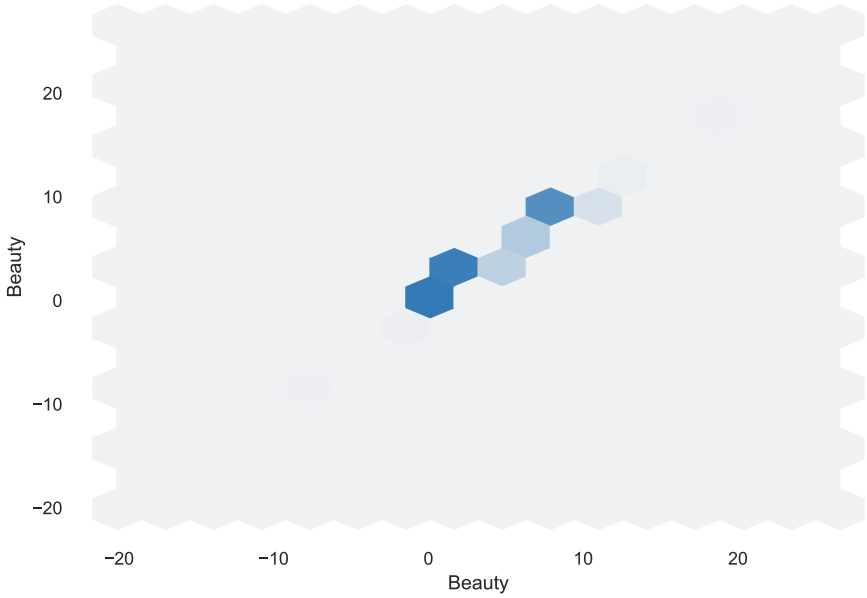

50
45
40
35
30
25
20
15
Age
−20
−10
0
10
20
Beauty

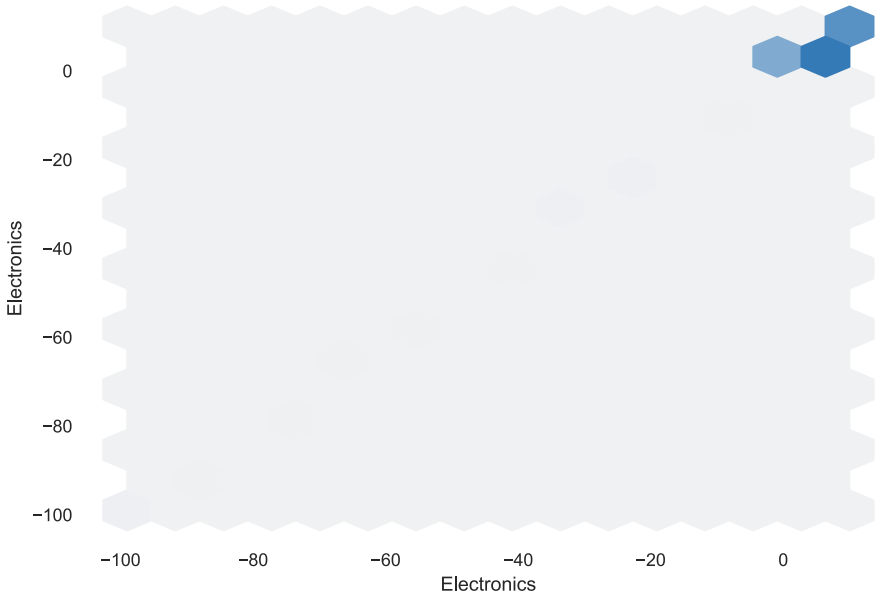

0
−20
−40
−60
−80
−100
Women's_Clothing
−20
−10
0
10
20
Beauty

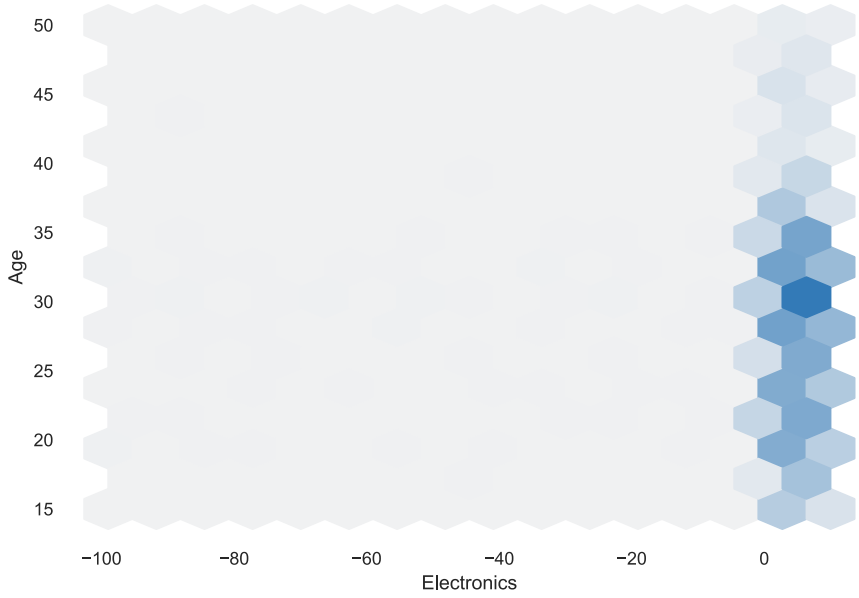

10
8
6
4
2
0
Men's_Clothing
−20
−10
0
10
20
Beauty

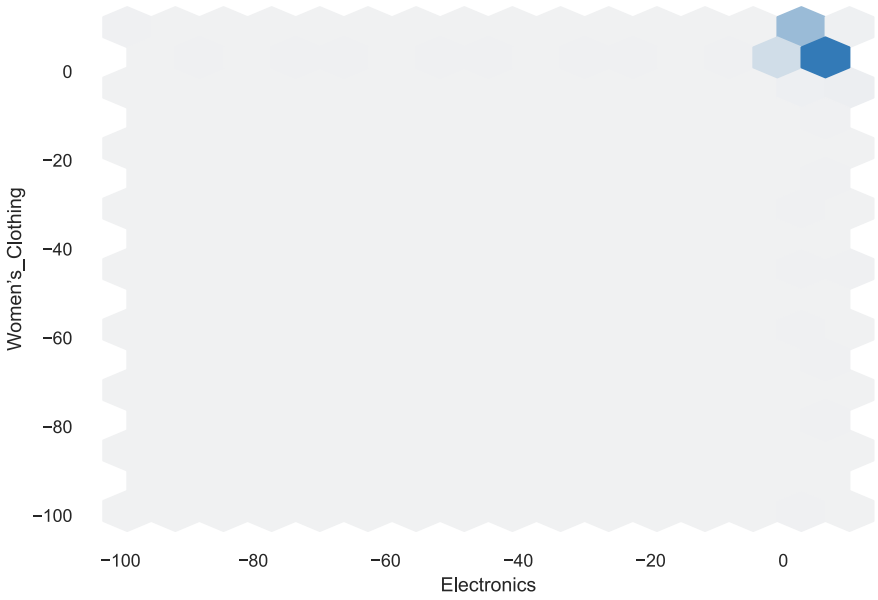

10
8
6
4
2
0
−2
Kid's_Clothing
−15
−10
−5
0
5
10
15
20
25
Beauty

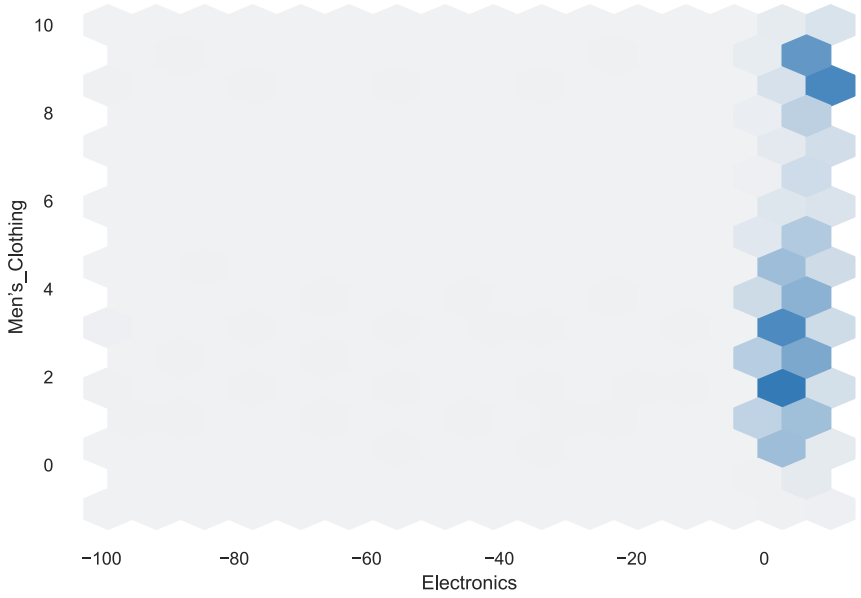

10
8
6
4
2
0
−2
Home_&_Living
−20
−10
0
10
20
Beauty

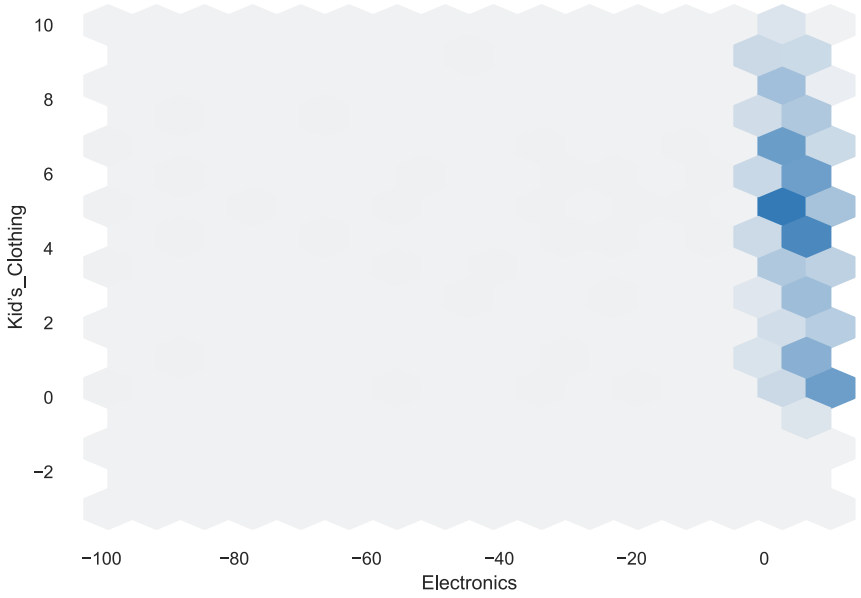

20
10
0
−10
−20
Beauty
−20
−10
0
10
20
Beauty

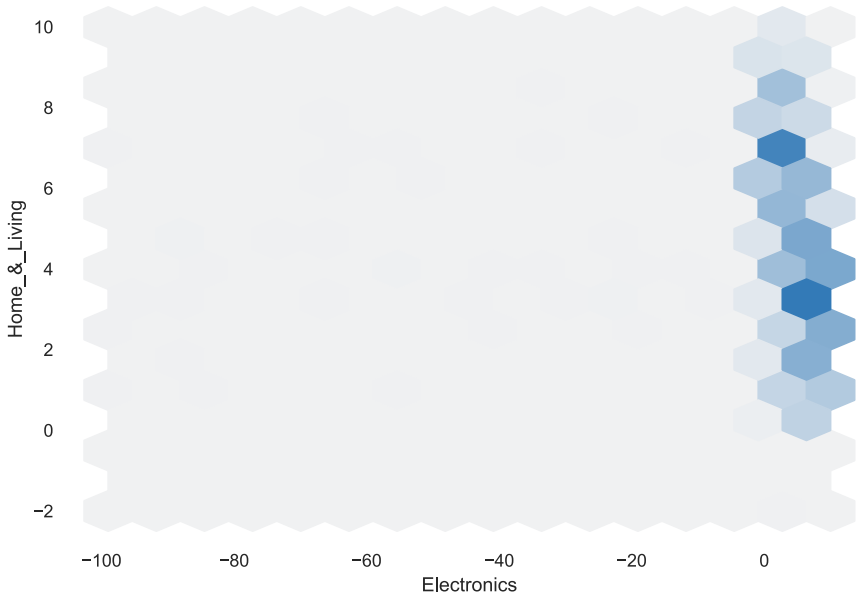

0
−20
−40
−60
−80
−100
Electronics
−100
−80
−60
−40
−20
0
Electronics

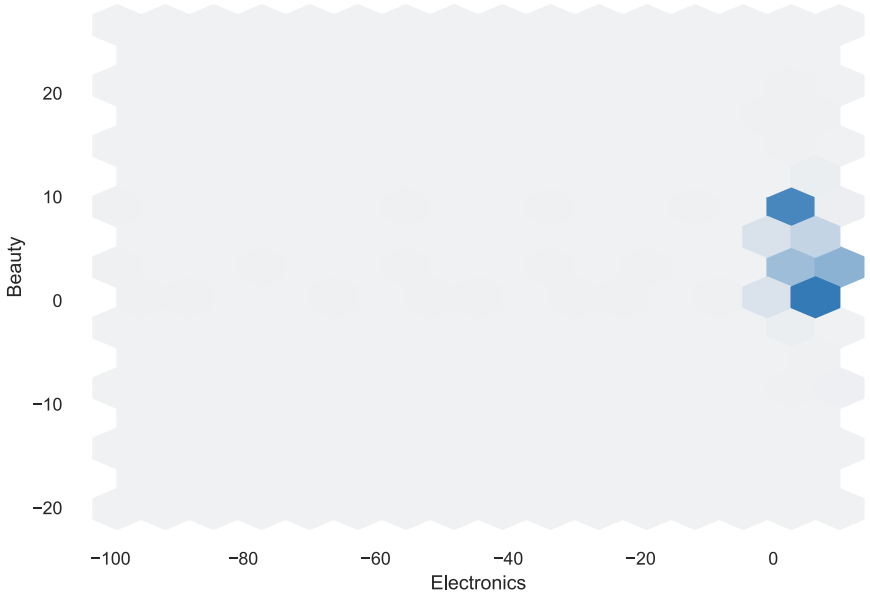

50
45
40
35
30
25
20
15
Age
−100
−80
−60
−40
−20
0
Electronics


0
−20
−40
−60
−80
−100
Women's_Clothing
−100
−80
−60
−40
−20
0
Electronics

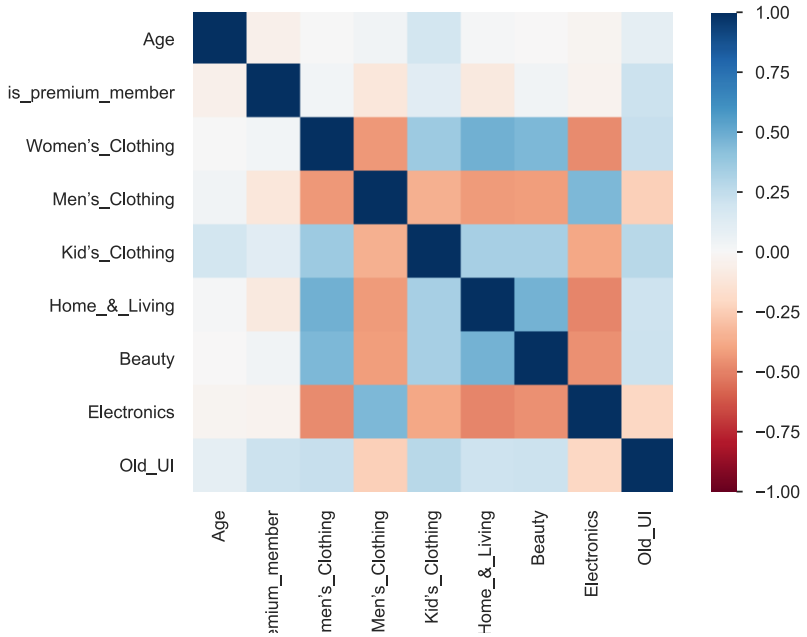

10
8
6
4
2
0
Men's_Clothing
−100
−80
−60
−40
−20
0
Electronics

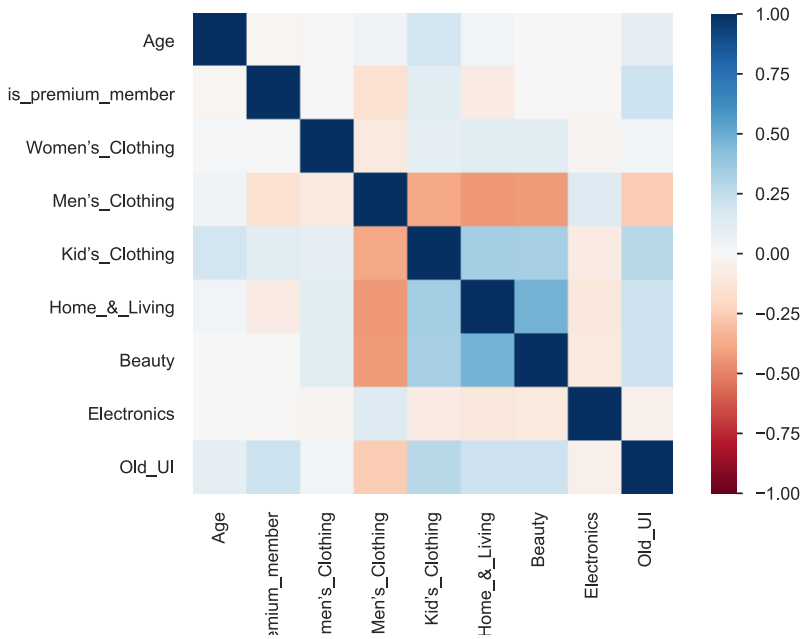

10
8
6
4
2
0
−2
Kid's_Clothing
−100
−80
−60
−40
−20
0
Electronics

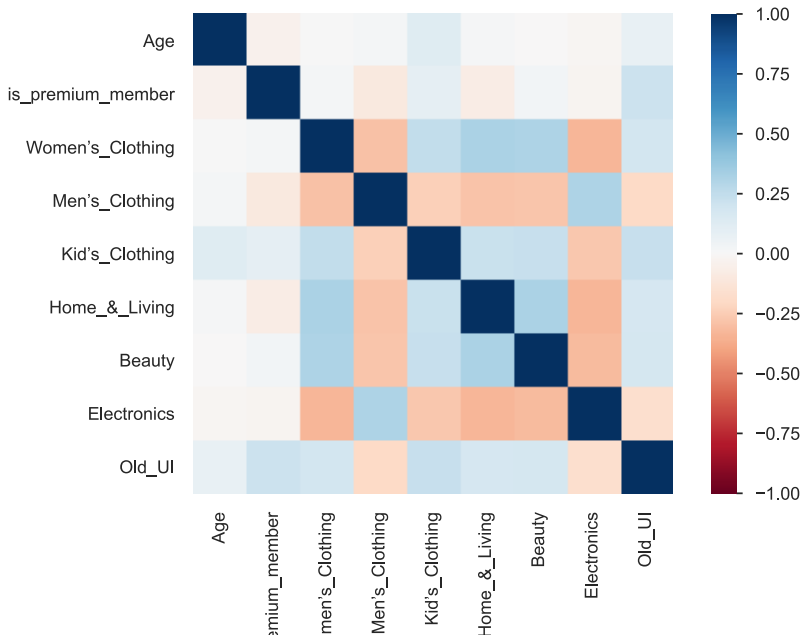

10
8
6
4
2
0
−2
Home_&_Living
−100
−80
−60
−40
−20
0
Electronics

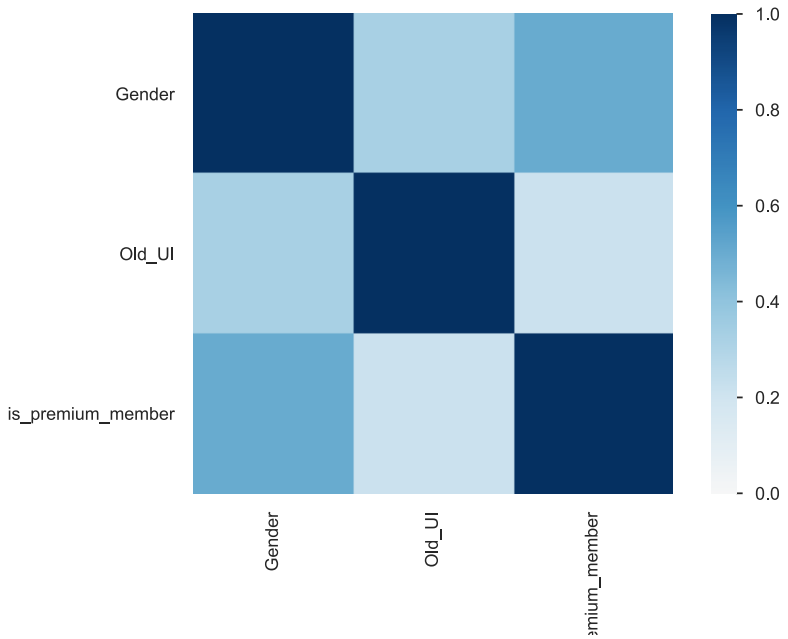

20
10
0
−10
−20
Beauty
−100
−80
−60
−40
−20
0
Electronics

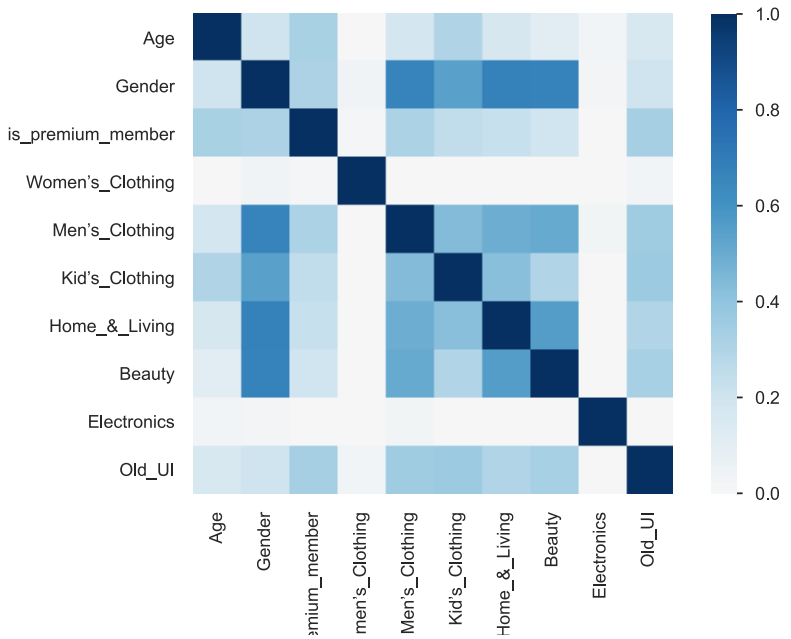
Correlations



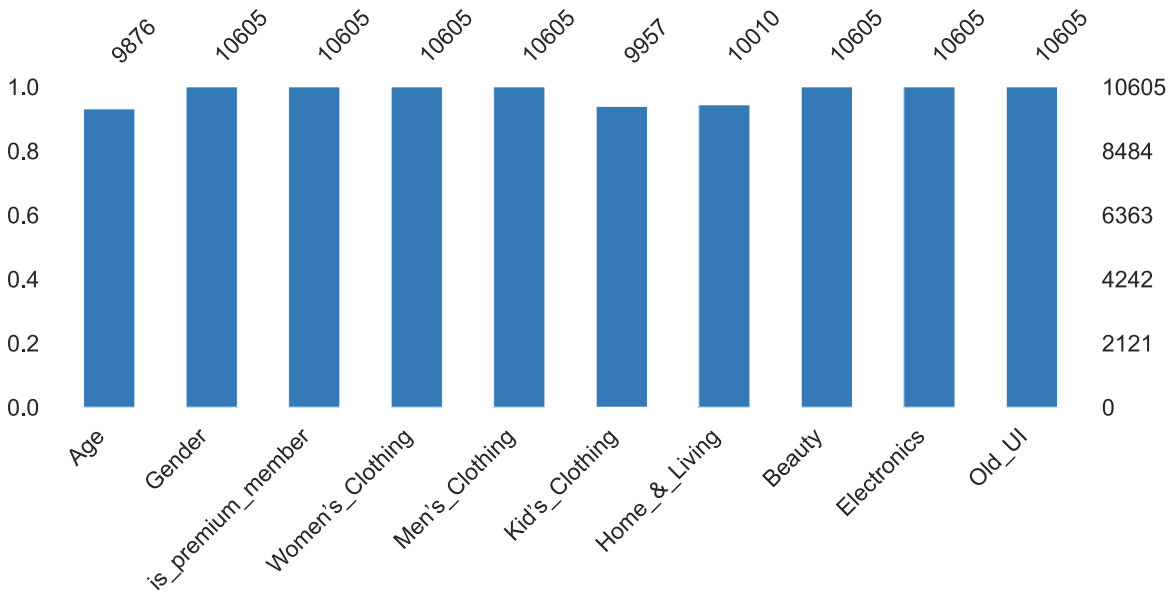

Age
is_premium_member
Women's_Clothing
Men's_Clothing
Kid's_Clothing
Home_&_Living
Beauty
Electronics
Old_UI
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
Age
mium_member
men's_Clothing
Men's_Clothing
Kid's_Clothing
Home_&_Living
Beauty
Electronics
Old_UI


Age
is_premium_member
Women's_Clothing
Men's_Clothing
Kid's_Clothing
Home_&_Living
Beauty
Electronics
Old_UI
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

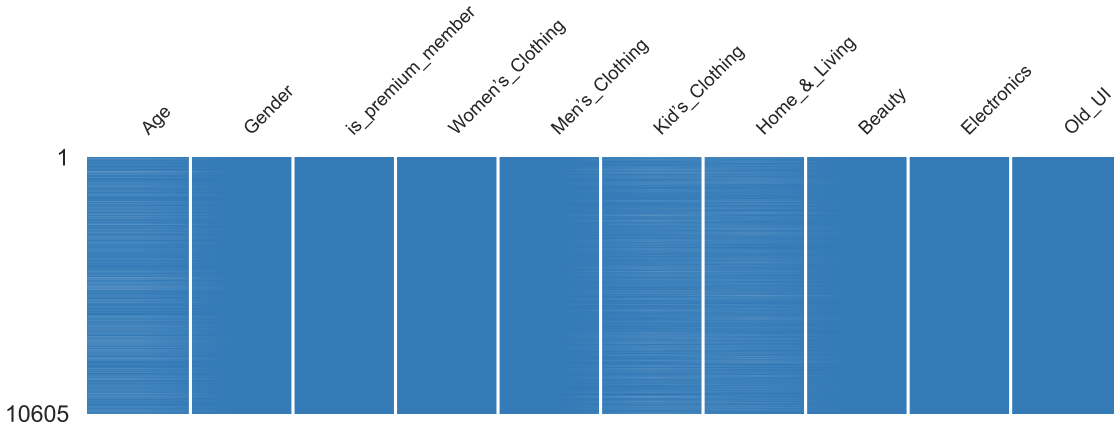

Gender
Old_UI
is_premium_member
1.0
0.8
0.6
0.4
0.2
0.0
Gender
Old_UI
mium_member


Age
Gender
is_premium_member
Women's_Clothing
Men's_Clothing
Kid's_Clothing
Home_&_Living
Beauty
Electronics
Old_UI
1.0
0.8
0.6
0.4
0.2
0.0

Missing values

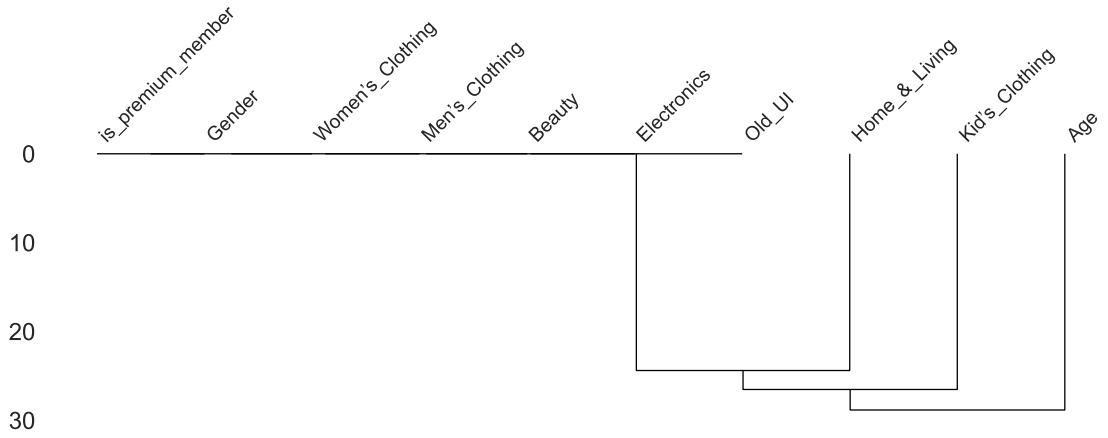


A simple visualization of nullity by column.



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.



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

	Age	Gender	is_premium_member	Women's_Clothing	Men's_Clothing	Kid's_Clothing	Home_&_Living	Beauty	Electronics	Ol
0	19.0	Not_Specified	1	3.445937	2.620136	5.457500	6.141905	8.106490	4.389933	1
1	31.0	Male	0	1.320304	9.025863	6.378695	2.825636	0.977430	7.789076	0
2	NaN	Not_Specified	1	3.793410	0.726490	3.957772	3.000000	2.589856	7.426173	1
3	29.0	Not_Specified	1	2.362948	2.701855	3.522246	3.121818	2.259882	5.970419	1
4	38.0	Female	0	7.568971	2.161103	7.535511	0.865851	4.638520	9.061972	1
5	30.0	Female	0	8.790859	4.219936	3.000000	6.395232	9.539830	1.066810	1
6	27.0	Female	0	7.496507	6.065087	6.757594	8.390672	2.000000	3.440012	0
7	27.0	Female	0	7.447626	0.724823	8.645873	7.822508	7.069303	2.882355	0
8	18.0	Female	0	7.573536	4.267514	4.359906	NaN	6.521518	0.482545	0
9	19.0	Female	1	5.061261	3.822283	NaN	7.377240	9.669255	2.410884	1

Last rows

	Age	Gender	is_premium_member	Women's_Clothing	Men's_Clothing	Kid's_Clothing	Home_&_Living	Beauty	Electronic
10595	27.0	Male	0	-77.000000	8.244964	3.000000	0.071109	3.000000	8.363963
10596	20.0	Male	0	0.029417	3.000000	NaN	3.000000	0.922374	9.905588
10597	50.0	Male	0	1.310550	4.426669	3.481056	1.265560	4.000000	7.563363
10598	35.0	Female	1	6.846009	1.083609	1.000000	9.038491	6.432147	1.657029
10599	30.0	Male	0	1.862447	8.968277	0.376080	3.417256	1.058207	4.948415
10600	41.0	Male	1	3.633796	4.537401	6.742788	1.186837	10.838269	5.551093
10601	23.0	Male	0	5.568309	4.000000	0.940724	6.450330	3.390740	8.884210
10602	NaN	Male	0	3.276963	9.081538	NaN	4.866557	1.597704	8.661640
10603	23.0	Not_Specified	1	2.665036	5.189818	5.301214	2.239331	1.000000	5.601689
10604	31.0	Female	0	5.048187	0.733494	5.893756	4.123734	8.734276	5.291852

Report generated with pandas-profiling (https://github.com/pandas-profiling/pandas-profiling).