

Overview

Brought to you by YData (https://ydata.ai/?utm_source=opensource&utm_medium=ydataprofiling&utm_campaign=report)

Dataset statistics

Number of variables	13
Number of observations	6941
Missing cells	43971
Missing cells (%)	48.7%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	705.1 KiB
Average record size in memory	104.0 B

Variable types

Text	3
Numeric	4
Boolean	3
Categorical	3

Alerts

preventTargetGapPoints has constant value "True"	Constant
userFlaggedNewItem has constant value "True"	Constant
finalPrice is highly overall correlated with itemPrice and 2 other fields (itemPrice, userFlaggedBarcode, userFlaggedPrice)	High correlation
itemPrice is highly overall correlated with finalPrice and 2 other fields (finalPrice, userFlaggedBarcode, userFlaggedPrice)	High correlation
needsFetchReview is highly overall correlated with userFlaggedBarcode and 1 other fields (userFlaggedBarcode, userFlaggedPrice)	High correlation
partnerItemId is highly overall correlated with userFlaggedBarcode	High correlation
quantityPurchased is highly overall correlated with userFlaggedQuantity	High correlation
userFlaggedBarcode is highly overall correlated with finalPrice and 5 other fields (finalPrice, itemPrice, needsFetchReview, partnerItemId, userFlaggedPrice, userFlaggedQuantity)	High correlation
userFlaggedQuantity is highly overall correlated with quantityPurchased and 1 other fields (quantityPurchased, userFlaggedBarcode)	High correlation
barcode has 3851 (55.5%) missing values	Missing
description has 381 (5.5%) missing values	Missing
finalPrice has 174 (2.5%) missing values	Missing
itemPrice has 174 (2.5%) missing values	Missing
needsFetchReview has 6128 (88.3%) missing values	Missing
preventTargetGapPoints has 6583 (94.8%) missing values	Missing
quantityPurchased has 174 (2.5%) missing values	Missing
userFlaggedBarcode has 6604 (95.1%) missing values	Missing
userFlaggedNewItem has 6618 (95.3%) missing values	Missing
userFlaggedPrice has 6642 (95.7%) missing values	Missing
userFlaggedQuantity has 6642 (95.7%) missing values	Missing
partnerItemId has 145 (2.1%) zeros	Zeros

Reproduction

Analysis 2024-10-09 04:08:32.926241

Analysis finished 2024-10-09 04:10:20.818768

Duration 1 minute and 47.89 seconds

Software version ydata-profiling vv4.10.0 (<https://github.com/ydataai/ydata-profiling>)

Variables

Select Columns ▾

receiptId

Text

Distinct	679
Distinct (%)	9.8%
Missing	0
Missing (%)	0.0%
Memory size	54.4 KiB

600a1a8d0a7214ada2000008 600b2dd60a7214ada2000009
60099e3c0a7214aa89000135 600f0cc70a720f053500002c
60099e6ca720f05fa00013a 60023e8f0a720f05f300008b
600f24970a720f053500002f
600b42000a7214aa8900000d 600373700a720f05f3000091
600b720a7214aa8900000e 600b139e0a7214ada200000a
600f2fc80a720f0535000030
600ba8ae0a7214ada2000010 600b34ca0a7214ada200000c
600edb70a720f053500001d 60039f60a720f053500000e
600eda0c0a7214ada200001e 600260210a720f05f300008f
600370250a720f05f3000090 6002590c0a7214ada4c000082
600a1e270a720f0535000099
60049d9d0a720f05f3000094 6004a99e0a720f05f3000095
6004a5f20a7214ad4c000088

Length

Max length	24
Median length	24
Mean length	24
Min length	24

Characters and Unicode

Total characters	166584
Distinct characters	16
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	377
Unique (%)	5.4%

Sample

1st row	5ff1e1eb0a720f0523000575
2nd row	5ff1e1bb0a720f052300056b
3rd row	5ff1e1bb0a720f052300056b
4th row	5ff1e1f10a720f052300057a
5th row	5ff1e1ee0a7214ada100056f

Value	Count	Frequency (%)
600f2fc80a720f0535000030	459	6.6%
600f39c30a7214ada2000030	450	6.5%
600f24970a720f053500002f	381	5.5%
600f0cc70a720f053500002c	217	3.1%
600a1a8d0a7214ada2000008	203	2.9%
60049d9d0a720f05f3000094	194	2.8%
60025cb80a720f05f300008d	185	2.7%
600260210a720f05f300008f	183	2.6%
600a1e270a720f0535000009	176	2.5%
600edb570a720f053500001d	155	2.2%
Other values (669)	4338	62.5%

600a1a8d0a7214ada2000008
 600f0cc70a720f053500002c
 600f24970a720f053500002f
 600f0cc70a720f0535000030
 600f2fc80a720f0535000030
 600f39c30a7214ada2000030
 600edb570a720f053500004d
 600eda0c07214ada200004e
 600260210a720f05f300008f
 600370250a720f05f3000090
 600a1e270a720f0535000009
 60049d9d0a720f05f3000094
 6004a5f20a7214ad4c000088

Most occurring characters

Value	Count	Frequency (%)
0	58962	35.4%
a	14693	8.8%
2	13346	8.0%
f	11252	6.8%
7	9412	5.7%
5	8794	5.3%
3	7813	4.7%
6	7802	4.7%
1	6020	3.6%
4	5640	3.4%
Other values (6)	22850	13.7%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	166584	100.0%

Most frequent character per category

Value	Count	Frequency (%)
0	58962	35.4%
a	14693	8.8%
2	13346	8.0%
f	11252	6.8%
7	9412	5.7%
5	8794	5.3%
3	7813	4.7%
6	7802	4.7%
1	6020	3.6%
4	5640	3.4%
Other values (6)	22850	13.7%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	166584	100.0%

Most frequent character per script

Value	Count	Frequency (%)
0	58962	35.4%
a	14693	8.8%
2	13346	8.0%
f	11252	6.8%
7	9412	5.7%
5	8794	5.3%
3	7813	4.7%
6	7802	4.7%
1	6020	3.6%
4	5640	3.4%
Other values (6)	22850	13.7%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	166584	100.0%

Most frequent character per block

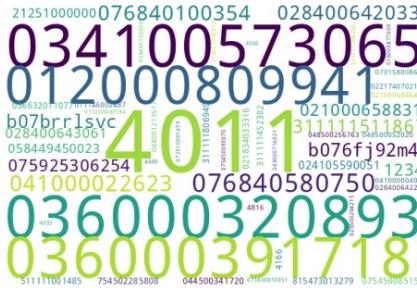
Value	Count	Frequency (%)
0	58962	35.4%
a	14693	8.8%
2	13346	8.0%
f	11252	6.8%
7	9412	5.7%
5	8794	5.3%
3	7813	4.7%
6	7802	4.7%
1	6020	3.6%
4	5640	3.4%
Other values (6)	22850	13.7%

barcode

Text

MISSING

Distinct	568
Distinct (%)	18.4%
Missing	3851
Missing (%)	55.5%
Memory size	54.4 KiB



Length

Max length	13
Median length	12
Mean length	11.067314
Min length	2

Characters and Unicode

Total characters	34198
Distinct characters	24
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.

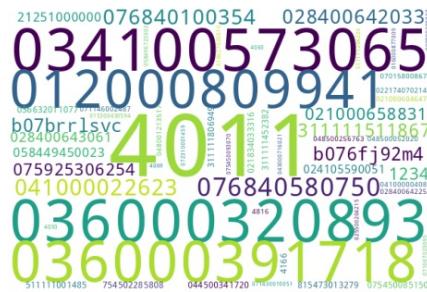
Unique

Unique	261
Unique (%)	8.4%

Sample

1st row	4011
2nd row	4011
3rd row	028400642255
4th row	4011
5th row	4011

Value	Count	Frequency (%)
4011	177	5.7%
036000320893	92	3.0%
034100573065	90	2.9%
036000391718	87	2.8%
012000809941	76	2.5%
076840580750	63	2.0%
041000022623	54	1.7%
076840100354	53	1.7%
028400642033	45	1.5%
31111511867	41	1.3%
Other values (558)	2312	74.8%



Most occurring characters

Value	Count	Frequency (%)
0	11212	32.8%
1	4353	12.7%
4	3987	9.0%
3	2690	7.9%
2	2582	7.6%
5	2310	6.8%
7	2083	6.1%
6	2026	5.9%
8	1890	5.5%
9	1471	4.3%
Other values (14)	494	1.4%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	34198	100.0%

Most frequent character per category

(*unknown*)

Value	Count	Frequency (%)
0	11212	32.8%
1	4353	12.7%
4	3087	9.0%
3	2690	7.9%
2	2582	7.6%
5	2310	6.8%
7	2083	6.1%
6	2026	5.9%
8	1890	5.5%
9	1471	4.3%
Other values (14)	494	1.4%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	34198	100.0%

Most frequent character per script

(*unknown*)

Value	Count	Frequency (%)
0	11212	32.8%
1	4353	12.7%
4	3087	9.0%
3	2690	7.9%
2	2582	7.6%
5	2310	6.8%
7	2083	6.1%
6	2026	5.9%
8	1890	5.5%
9	1471	4.3%
Other values (14)	494	1.4%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	34198	100.0%

Most frequent character per block

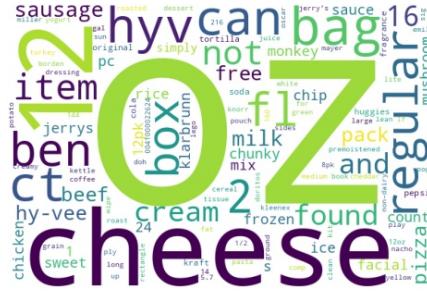
Value	Count	Frequency (%)
0	11212	32.8%
1	4353	12.7%
4	3087	9.0%
3	2690	7.9%
2	2582	7.6%
5	2310	6.8%
7	2083	6.1%
6	2026	5.9%
8	1890	5.5%
9	1471	4.3%
Other values (14)	494	1.4%

description

Text

MISSING

Distinct	1889
Distinct (%)	28.8%
Missing	381
Missing (%)	5.5%
Memory size	54.4 KiB



Length

Max length	155
Median length	92
Mean length	29.15122
Min length	2

Characters and Unicode

Total characters	191232
Distinct characters	83
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.

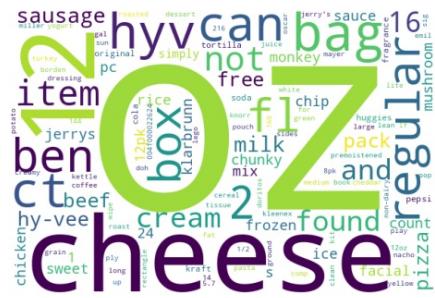
Unique

Unique	1136	?
Unique (%)	17.3%	

Sample

1st row	ITEM NOT FOUND
2nd row	ITEM NOT FOUND
3rd row	DORITOS TORTILLA CHIP SPICY SWEET CHILI REDUCED FAT BAG 1 OZ
4th row	ITEM NOT FOUND
5th row	ITEM NOT FOUND

Value	Count	Frequency (%)
oz	1209	3.5%
	931	2.7%
cheese	327	1.0%
12	321	0.9%
bag	276	0.8%
hyv	246	0.7%
can	241	0.7%
ct	237	0.7%
regular	215	0.6%
f1	211	0.6%
Other values (3183)	30149	87.7%



Most occurring characters

Value	Count	Frequency (%)
E	27810	14.5%
e	8222	4.3%
R	7865	4.1%
C	7158	3.7%
S	6724	3.5%
A	6612	3.5%
L	6575	3.4%
O	6297	3.3%
N	6266	3.3%
Other values (73)	5520	2.9%
	102183	53.4%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	191232	100.0%

Most frequent character per category

(*unknown*)

Value	Count	Frequency (%)
E	27810	14.5%
e	8222	4.3%
R	7865	4.1%
C	7158	3.7%
S	6724	3.5%
A	6612	3.5%
L	6575	3.4%
O	6297	3.3%
N	6266	3.3%
Other values (73)	5520	2.9%
	102183	53.4%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	191232	100.0%

Most frequent character per script

Value	Count	Frequency (%)
E	27810	14.5%
e	8222	4.3%
R	7865	4.1%
C	7158	3.7%
S	6724	3.5%
A	6612	3.5%
L	6575	3.4%
O	6297	3.3%
N	6266	3.3%
Other values (73)	5520	2.9%
	102183	53.4%

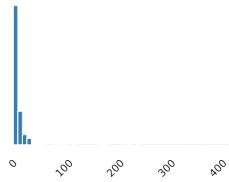
Most occurring blocks

Value	Count	Frequency (%)
(unknown)	191232	100.0%

Most frequent character per block

Value	Count	Frequency (%)
E	27810	14.5%
e	8222	4.3%
R	7865	4.1%
C	7158	3.7%
S	6724	3.5%
A	6612	3.5%
L	6575	3.4%
O	6297	3.3%
N	6266	3.3%
Other values (73)	5520	2.9%
	102183	53.4%

finalPrice	
Real number ()	
HIGH CORRELATION (This variable has a high overall correlation with 3 fields: itemPrice, userFlaggedBarcode, userFlaggedPrice) MISSING	
Distinct	823
Distinct (%)	12.2%
Missing	174
Missing (%)	2.5%
Infinite	0
Infinite (%)	0.0%
Mean	7.871661
Minimum	0
Maximum	441.58
Zeros	4
Zeros (%)	0.1%
Negative	0
Negative (%)	0.0%
Memory size	54.4 KiB

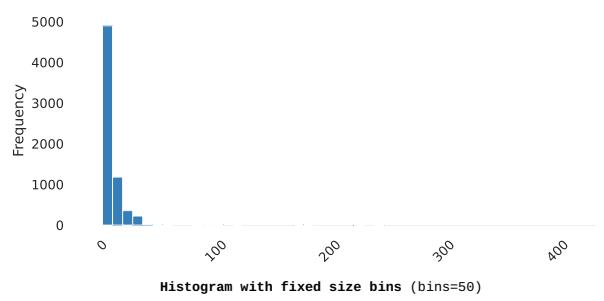


Quantile statistics

Minimum	0
5-th percentile	0.56
Q1	2.29
median	4.28
Q3	9.99
95-th percentile	26
Maximum	441.58
Range	441.58
Interquartile range (IQR)	7.7

Descriptive statistics

Standard deviation	14.656776
Coefficient of variation (CV)	1.8619674
Kurtosis	207.38946
Mean	7.871661
Median Absolute Deviation (MAD)	2.7
Skewness	11.383034
Sum	53267.53
Variance	214.82108
Monotonicity	Not monotonic

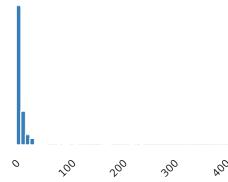


Value	Count	Frequency (%)
1	375	5.4%
9.99	355	5.1%
3.99	243	3.5%
4.99	195	2.8%
0.56	190	2.7%
2.99	179	2.6%
5.99	176	2.5%
3.49	139	2.0%
2.34	134	1.9%
5	124	1.8%
Other values (813)	4657	67.1%
(Missing)	174	2.5%

Value	Count	Frequency (%)
0	4	0.1%
0.16	1	< 0.1%
0.19	13	0.2%
0.25	2	< 0.1%
0.32	2	< 0.1%
0.48	3	< 0.1%
0.5	76	1.1%
0.54	66	1.0%
0.55	2	< 0.1%
0.56	190	2.7%

Value	Count	Frequency (%)
441.58	1	< 0.1%
245	3	< 0.1%
223.36	5	0.1%
180	6	0.1%
168.84	5	0.1%
115.96	1	< 0.1%
100.48	1	< 0.1%
100	6	0.1%
95.84	4	0.1%
82.34	1	< 0.1%

itemPrice	
Real number ()	
HIGH CORRELATION (This variable has a high overall correlation with 3 fields: finalPrice, userFlaggedBarcode, userFlaggedPrice)	MISSING
Distinct	823
Distinct (%)	12.2%
Missing	174
Missing (%)	2.5%
Infinite	0
Infinite (%)	0.0%
Mean	7.8721782
Minimum	0
Maximum	441.58
Zeros	4
Zeros (%)	0.1%
Negative	0
Negative (%)	0.0%
Memory size	54.4 KiB

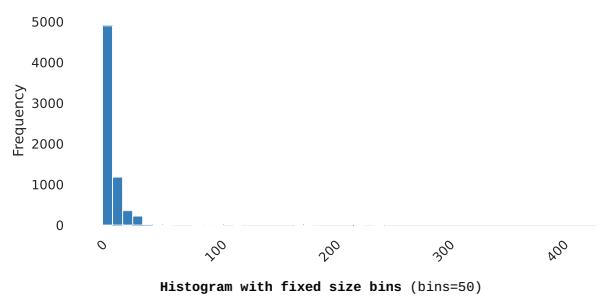


Quantile statistics

Minimum	0
5-th percentile	0.56
Q1	2.29
median	4.28
Q3	9.99
95-th percentile	26
Maximum	441.58
Range	441.58
Interquartile range (IQR)	7.7

Descriptive statistics

Standard deviation	14.656623
Coefficient of variation (CV)	1.8618256
Kurtosis	207.39662
Mean	7.8721782
Median Absolute Deviation (MAD)	2.7
Skewness	11.383294
Sum	53271.03
Variance	214.8166
Monotonicity	Not monotonic



Value	Count	Frequency (%)
1	375	5.4%
9.99	355	5.1%
3.99	243	3.5%
4.99	196	2.8%
0.56	190	2.7%
2.99	180	2.6%
5.99	176	2.5%
3.49	139	2.0%
2.34	134	1.9%
5	124	1.8%
Other values (813)	4655	67.1%
(Missing)	174	2.5%

Value	Count	Frequency (%)
0	4	0.1%
0.16	1	< 0.1%
0.19	13	0.2%
0.25	2	< 0.1%
0.32	2	< 0.1%
0.48	3	< 0.1%
0.5	76	1.1%
0.54	66	1.0%
0.55	2	< 0.1%
0.56	190	2.7%

Value	Count	Frequency (%)
441.58	1	< 0.1%
245	3	< 0.1%
223.36	5	0.1%
180	6	0.1%
168.84	5	0.1%
115.96	1	< 0.1%
100.48	1	< 0.1%
100	6	0.1%
95.84	4	0.1%
82.34	1	< 0.1%

needsFetchReview

Boolean

HIGH CORRELATION (This variable has a high overall correlation with 2 fields: userFlaggedBarcode, userFlaggedPrice) MISSING

Distinct

2

Distinct (%)

0.2%

Missing

6128

Missing (%)

88.3%

Memory size

54.4 KiB

Value**Count Frequency (%)**

False

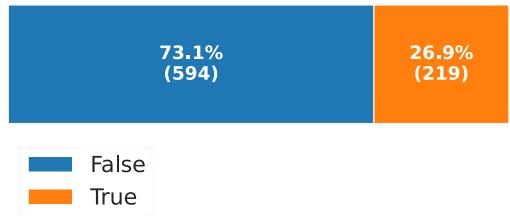
594 8.6%

True

219 3.2%

(Missing)

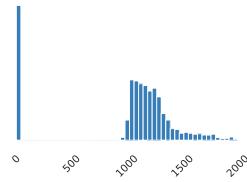
6128 88.3%



partnerItemId
Real number ()

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

Distinct	916
Distinct (%)	13.2%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	988.52428
Minimum	0
Maximum	2043
Zeros	145
Zeros (%)	2.1%
Negative	0
Negative (%)	0.0%
Memory size	54.4 KiB

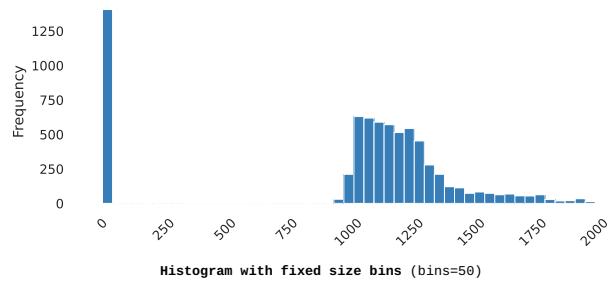


Quantile statistics

Minimum	0
5-th percentile	1
Q1	1027
median	1143
Q3	1274
95-th percentile	1644
Maximum	2043
Range	2043
Interquartile range (IQR)	247

Descriptive statistics

Standard deviation	527.38082
Coefficient of variation (CV)	0.53350316
Kurtosis	-0.13640789
Mean	988.52428
Median Absolute Deviation (MAD)	123
Skewness	-1.0174858
Sum	6861347
Variance	278130.53
Monotonicity	Not monotonic



Value	Count	Frequency (%)
1	531	7.7%
2	203	2.9%
0	145	2.1%
3	135	1.9%
4	113	1.6%
5	106	1.5%
6	34	0.5%
7	34	0.5%
8	34	0.5%
9	34	0.5%
Other values (906)	5572	80.3%

Value	Count	Frequency (%)
0	145	2.1%
1	531	7.7%
2	203	2.9%
3	135	1.9%
4	113	1.6%
5	106	1.5%
6	34	0.5%
7	34	0.5%
8	34	0.5%
9	34	0.5%

Value	Count	Frequency (%)
2043	1	< 0.1%
2040	1	< 0.1%
2036	1	< 0.1%
2033	1	< 0.1%
2029	1	< 0.1%
2026	1	< 0.1%
1986	1	< 0.1%
1983	1	< 0.1%
1980	1	< 0.1%
1976	1	< 0.1%

`preventTargetGapPoints`

Boolean

CONSTANT MISSING

Distinct

1

Distinct (%)

0.3%

Missing

6583

Missing (%)

94.8%

Memory size

54.4 KiB

Value

Count Frequency (%)

True

358

5.2%

(Missing)

6583

94.8%

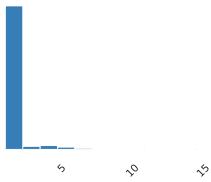


■ True

quantityPurchased
Real number ()

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

Distinct	13
Distinct (%)	0.2%
Missing	174
Missing (%)	2.5%
Infinite	0
Infinite (%)	0.0%
Mean	1.3861386
Minimum	1
Maximum	17
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	54.4 KiB

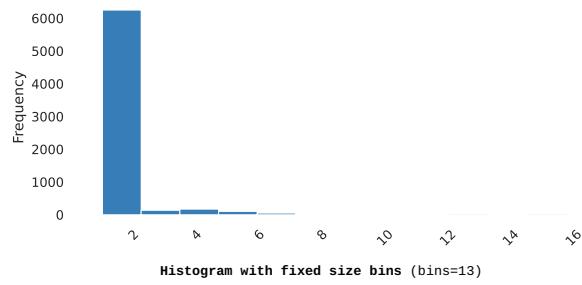


Quantile statistics

Minimum	1
5-th percentile	1
Q1	1
median	1
Q3	1
95-th percentile	4
Maximum	17
Range	16
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	1.2043632
Coefficient of variation (CV)	0.86886201
Kurtosis	36.002117
Mean	1.3861386
Median Absolute Deviation (MAD)	0
Skewness	5.1137362
Sum	9380
Variance	1.4504907
Monotonicity	Not monotonic



Value	Count	Frequency (%)
1	5628	81.1%
2	622	9.0%
4	170	2.4%
3	134	1.9%
5	101	1.5%
6	37	0.5%
8	22	0.3%
10	15	0.2%
7	13	0.2%
9	13	0.2%
Other values (3)	12	0.2%
(Missing)	174	2.5%

Value	Count	Frequency (%)
1	5628	81.1%
2	622	9.0%
3	134	1.9%
4	170	2.4%
5	101	1.5%
6	37	0.5%
7	13	0.2%
8	22	0.3%
9	13	0.2%
10	15	0.2%

Value	Count	Frequency (%)
17	3	< 0.1%
14	3	< 0.1%
12	6	0.1%
10	15	0.2%
9	13	0.2%
8	22	0.3%
7	13	0.2%
6	37	0.5%
5	101	1.5%
4	170	2.4%

userFlaggedBarcode

Categorical

HIGH CORRELATION (This variable has a high overall correlation with 6 fields: finalPrice, itemPrice, needsFetchReview, partnerItemId, userFlaggedPrice, userFlaggedQuantity) MISSING

Distinct	6
Distinct (%)	1.8%
Missing	6604
Missing (%)	95.1%
Memory size	54.4 KiB

Length

Max length	12
Median length	12
Mean length	8.7002967
Min length	4

Characters and Unicode

Total characters	2932
Distinct characters	10
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%	?

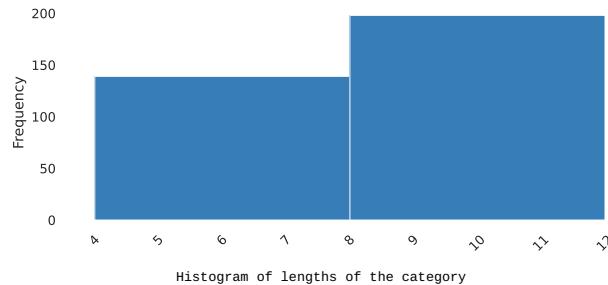
Sample

1st row	4011
2nd row	028400642255
3rd row	4011
4th row	4011
5th row	1234

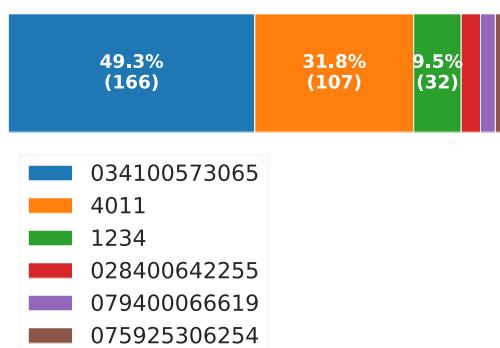
Common Values

Value	Count	Frequency (%)
034100573065	166	2.4%
4011	107	1.5%
1234	32	0.5%
028400642255	13	0.2%
079400066619	10	0.1%
075925306254	9	0.1%
(Missing)	6604	95.1%

Length



Common Values (Plot)



Value	Count	Frequency (%)
034100573065	166	49.3%
4011	107	31.8%
1234	32	9.5%
028400642255	13	3.9%
079400066619	10	3.0%
075925306254	9	2.7%

Most occurring characters

Value	Count	Frequency (%)
0	868	29.6%
1	422	14.4%
5	385	13.1%
3	373	12.7%
4	350	11.9%
6	218	7.4%
7	185	6.3%
2	89	3.0%
9	29	1.0%
8	13	0.4%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	2932	100.0%

Most frequent character per category

Value	Count	Frequency (%)
0	868	29.6%
1	422	14.4%
5	385	13.1%
3	373	12.7%
4	350	11.9%
6	218	7.4%
7	185	6.3%
2	89	3.0%
9	29	1.0%
8	13	0.4%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	2932	100.0%

Most frequent character per script

Value	Count	Frequency (%)
0	868	29.6%
1	422	14.4%
5	385	13.1%
3	373	12.7%
4	350	11.9%
6	218	7.4%
7	185	6.3%
2	89	3.0%
9	29	1.0%
8	13	0.4%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	2932	100.0%

Most frequent character per block

Value	Count	Frequency (%)
0	868	29.6%
1	422	14.4%
5	385	13.1%
3	373	12.7%
4	350	11.9%
6	218	7.4%
7	185	6.3%
2	89	3.0%
9	29	1.0%
8	13	0.4%

userFlaggedNewItem

Boolean

CONSTANT MISSING

Distinct

1

Distinct (%)

0.3%

Missing

6618

Missing (%)

95.3%

Memory size

54.4 KiB

Value**Count Frequency (%)**

True

323

4.7%

(Missing)

6618

95.3%



■ True

userFlaggedPrice

Categorical

HIGH CORRELATION (This variable has a high overall correlation with 4 fields: finalPrice, itemPrice, needsFetchReview, userFlaggedBarcode) MISSING

Distinct	13
Distinct (%)	4.3%
Missing	6642
Missing (%)	95.7%
Memory size	54.4 KiB

Length

Max length	5
Median length	5
Mean length	4.8862876
Min length	4

Characters and Unicode

Total characters	1461	
Distinct characters	11	
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%	

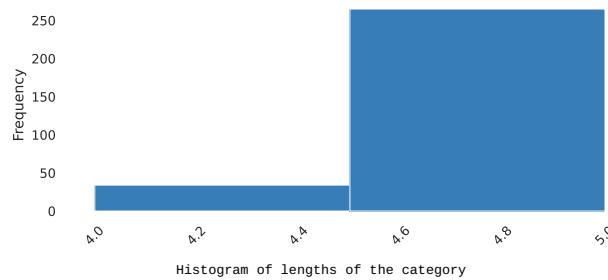
Sample

1st row	26.00
2nd row	10.00
3rd row	26.00
4th row	28.00
5th row	2.56

Common Values

Value	Count	Frequency (%)
29.00	142	2.0%
1.00	26	0.4%
10.00	22	0.3%
28.00	15	0.2%
20.00	14	0.2%
21.00	14	0.2%
27.00	13	0.2%
26.00	10	0.1%
25.00	10	0.1%
23.00	10	0.1%
Other values (3)	23	0.3%
(Missing)	6642	95.7%

Length



Value	Count	Frequency (%)
29.00	142	47.5%
1.00	26	8.7%
10.00	22	7.4%
28.00	15	5.0%
20.00	14	4.7%
21.00	14	4.7%
27.00	13	4.3%
26.00	10	3.3%
25.00	10	3.3%
23.00	10	3.3%
Other values (3)	23	7.7%

Most occurring characters

Value	Count	Frequency (%)
0	618	42.3%
.	299	20.5%
2	260	17.8%
9	142	9.7%
1	62	4.2%
6	18	1.2%
5	18	1.2%
8	15	1.0%
7	13	0.9%
3	10	0.7%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	1461	100.0%

Most frequent character per category

(*unknown*)

Value	Count	Frequency (%)
0	618	42.3%
.	299	20.5%
2	260	17.8%
9	142	9.7%
1	62	4.2%
6	18	1.2%
5	18	1.2%
8	15	1.0%
7	13	0.9%
3	10	0.7%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	1461	100.0%

Most frequent character per script

(*unknown*)

Value	Count	Frequency (%)
0	618	42.3%
.	299	20.5%
2	260	17.8%
9	142	9.7%
1	62	4.2%
6	18	1.2%
5	18	1.2%
8	15	1.0%
7	13	0.9%
3	10	0.7%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	1461	100.0%

Most frequent character per block

Value	Count	Frequency (%)
0	618	42.3%
.	299	20.5%
2	260	17.8%
9	142	9.7%
1	62	4.2%
6	18	1.2%
5	18	1.2%
8	15	1.0%
7	13	0.9%
3	10	0.7%

userFlaggedQuantity

Categorical

HIGH CORRELATION (This variable has a high overall correlation with 2 fields: quantityPurchased, userFlaggedBarcode) MISSING

Distinct	5
Distinct (%)	1.7%
Missing	6642
Missing (%)	95.7%
Memory size	54.4 KiB

Length

Max length	3
Median length	3
Mean length	3
Min length	3

Characters and Unicode

Total characters	897
Distinct characters	7
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%	

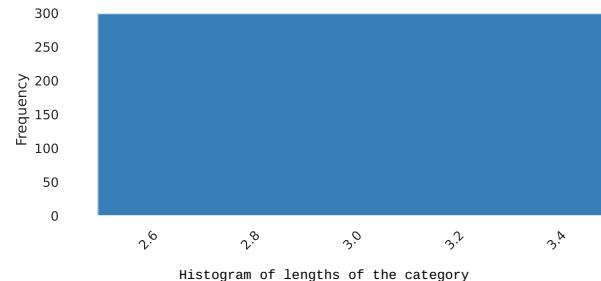
Sample

1st row	5.0
2nd row	1.0
3rd row	3.0
4th row	4.0
5th row	3.0

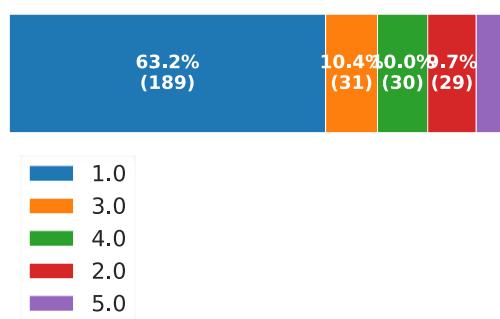
Common Values

Value	Count	Frequency (%)
1.0	189	2.7%
3.0	31	0.4%
4.0	30	0.4%
2.0	29	0.4%
5.0	20	0.3%
(Missing)	6642	95.7%

Length



Common Values (Plot)



Value	Count	Frequency (%)
1.0	189	63.2%
3.0	31	10.4%
4.0	30	10.0%
2.0	29	9.7%
5.0	20	6.7%

Most occurring characters

Value	Count	Frequency (%)
.	299	33.3%
0	299	33.3%
1	189	21.1%
3	31	3.5%
4	30	3.3%
2	29	3.2%
5	20	2.2%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	897	100.0%

Most frequent character per category

(*unknown*)

Value	Count	Frequency (%)
.	299	33.3%
0	299	33.3%
1	189	21.1%
3	31	3.5%
4	30	3.3%
2	29	3.2%
5	20	2.2%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	897	100.0%

Most frequent character per script

(*unknown*)

Value	Count	Frequency (%)
.	299	33.3%
0	299	33.3%
1	189	21.1%
3	31	3.5%
4	30	3.3%
2	29	3.2%
5	20	2.2%

Most occurring blocks

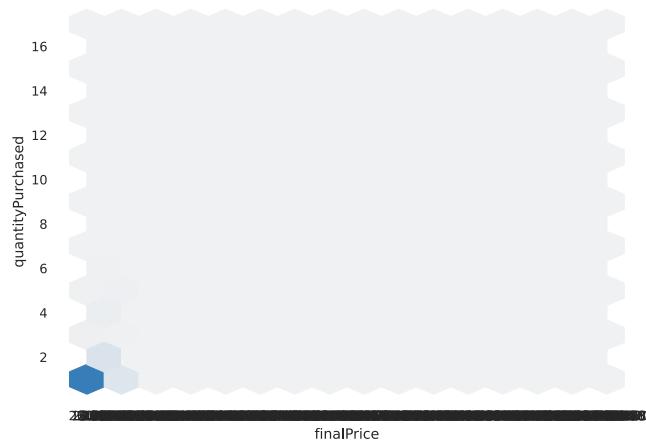
Value	Count	Frequency (%)
(unknown)	897	100.0%

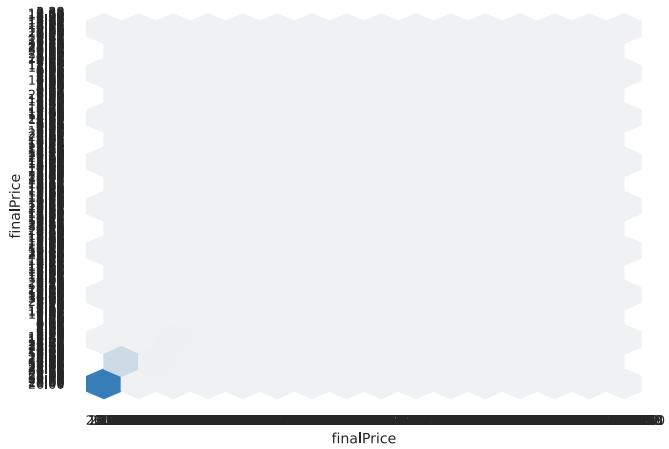
Most frequent character per block

(*unknown*)

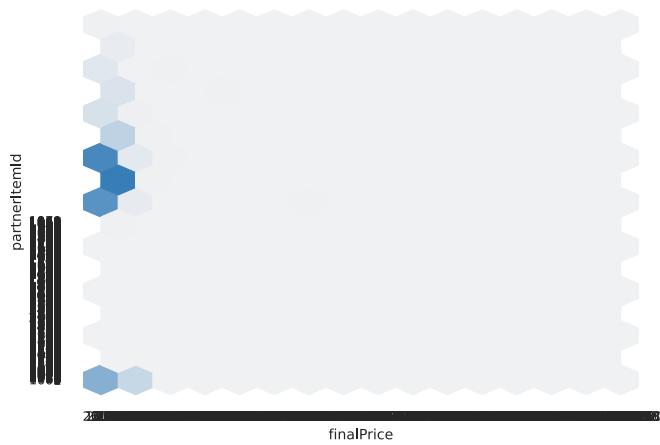
Value	Count	Frequency (%)
.	299	33.3%
0	299	33.3%
1	189	21.1%
3	31	3.5%
4	30	3.3%
2	29	3.2%
5	20	2.2%

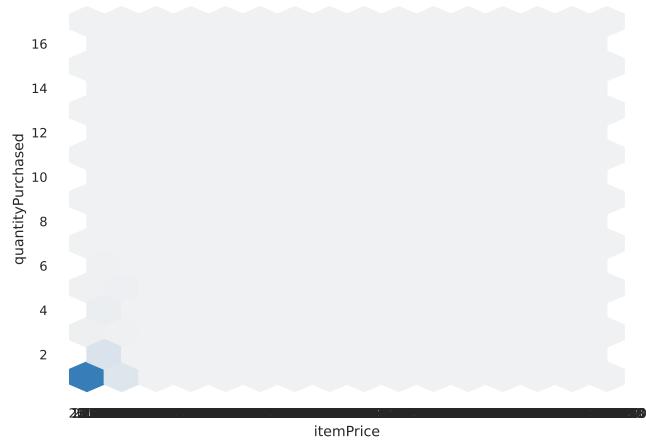
Interactions

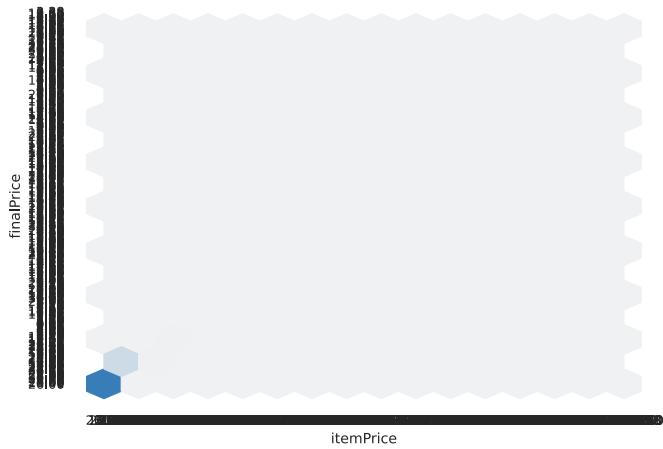


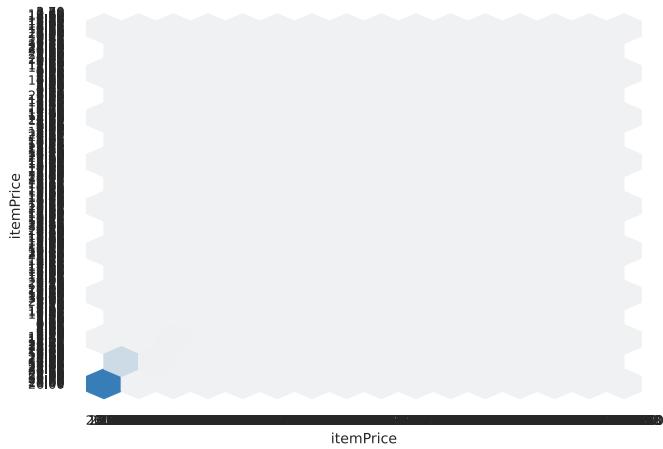


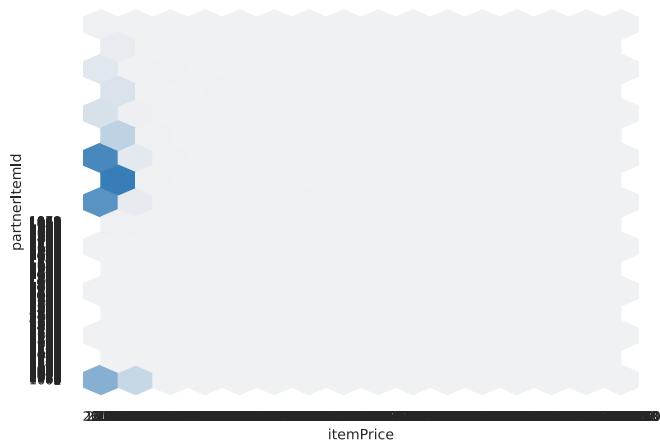


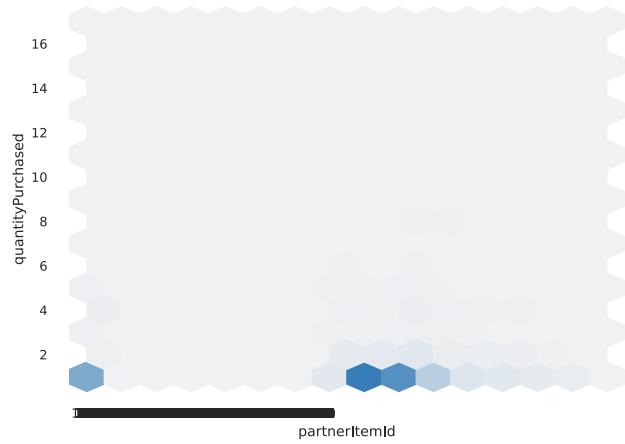


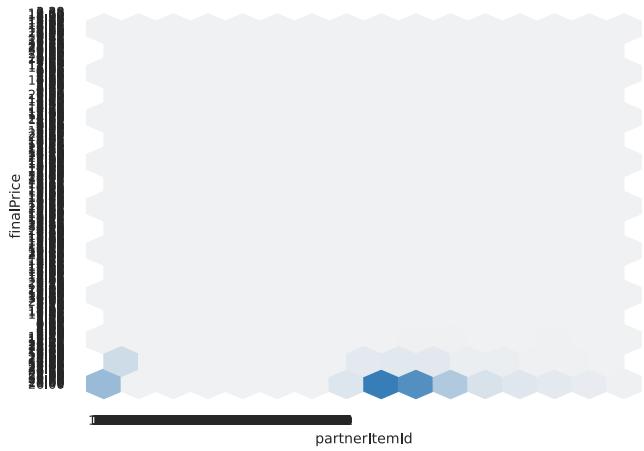


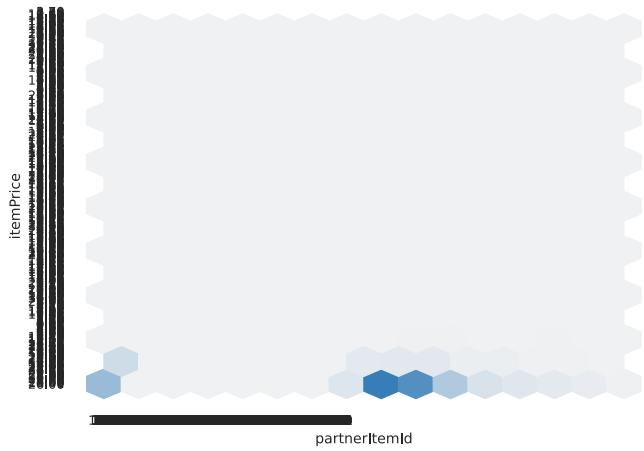


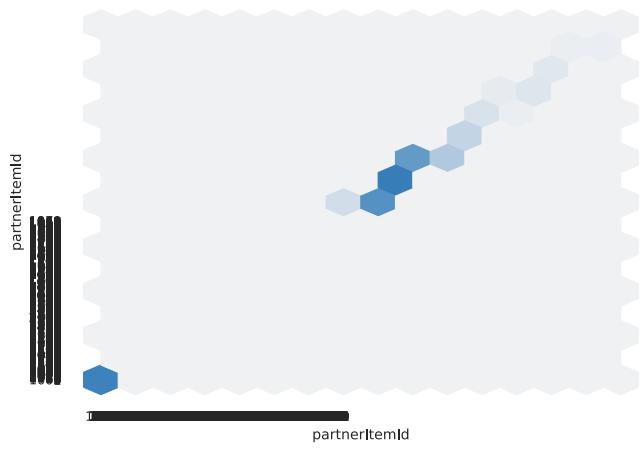


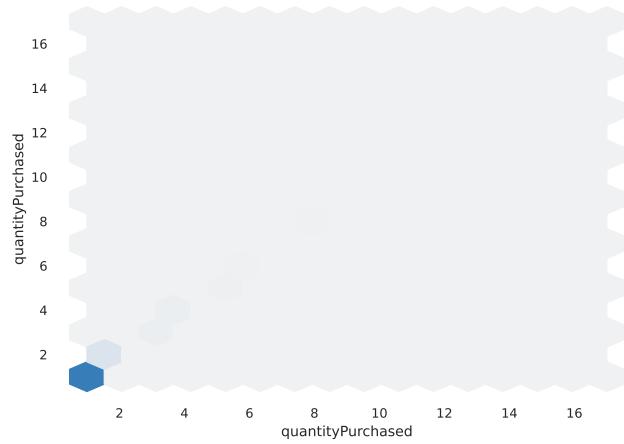


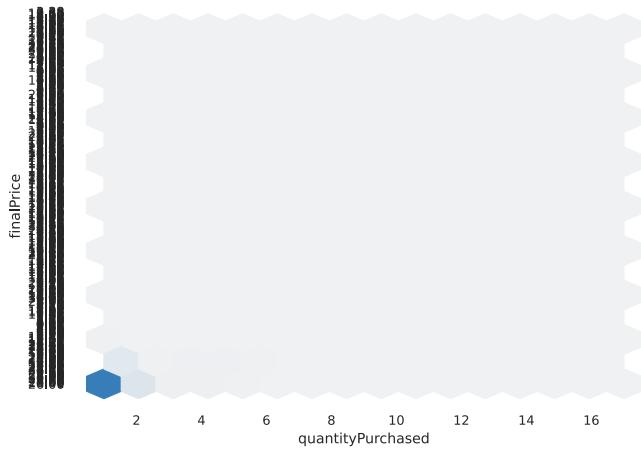


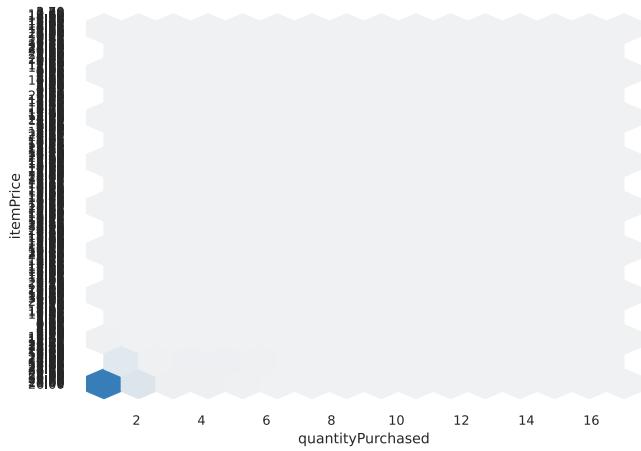


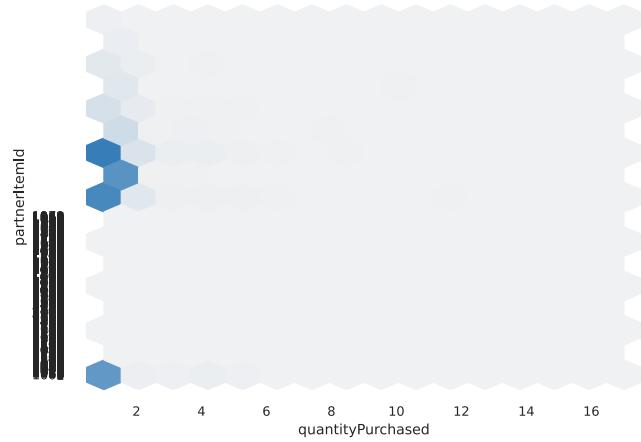




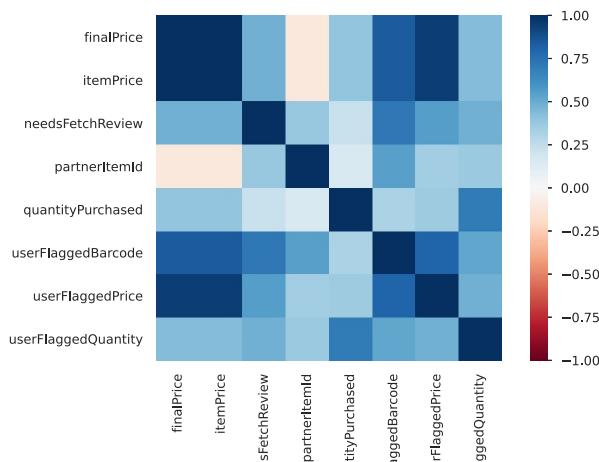






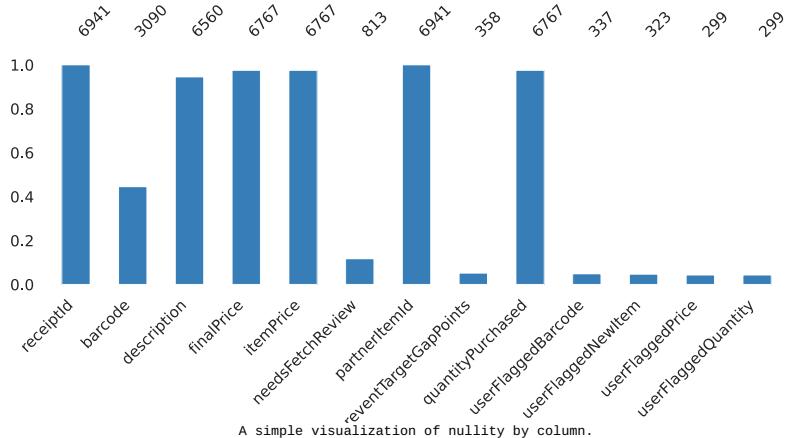


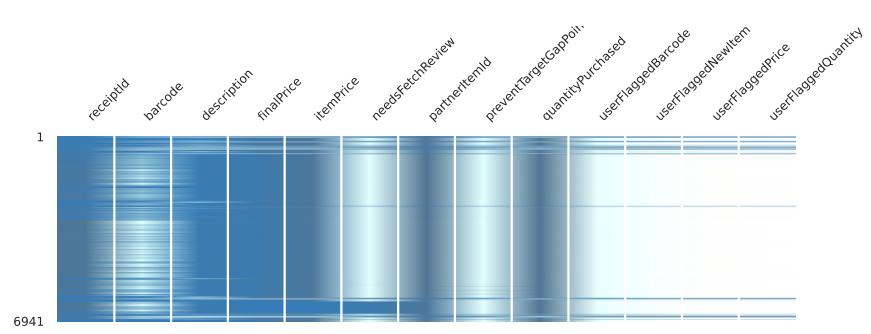
Correlations



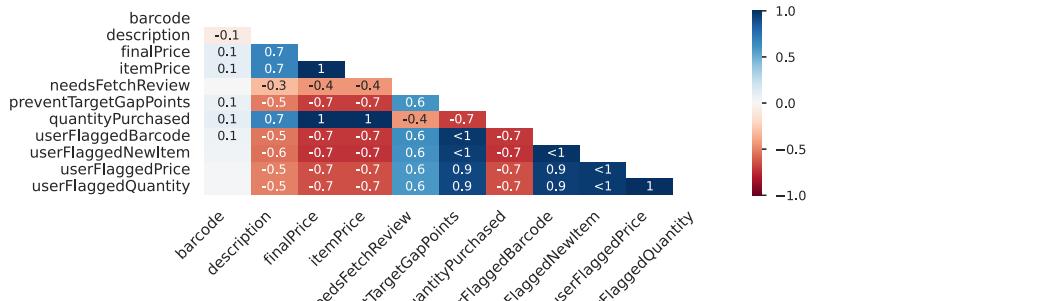
	finalPrice	itemPrice	needsFetchReview	partnerItemId	quantityPurchased	userFlaggedBarcode	userFlaggedPrice	userFlaggedQuantity
finalPrice	1.000	1.000	0.479	-0.103	0.398	0.830	0.943	0.432
itemPrice	1.000	1.000	0.479	-0.103	0.398	0.830	0.943	0.432
needsFetchReview	0.479	0.479	1.000	0.380	0.219	0.713	0.553	0.481
partnerItemId	-0.103	-0.103	0.380	1.000	0.150	0.544	0.345	0.371
quantityPurchased	0.398	0.398	0.219	0.150	1.000	0.318	0.357	0.701
userFlaggedBarcode	0.830	0.830	0.713	0.544	0.318	1.000	0.802	0.520
userFlaggedPrice	0.943	0.943	0.553	0.345	0.357	0.802	1.000	0.479
userFlaggedQuantity	0.432	0.432	0.481	0.371	0.701	0.520	0.479	1.000

Missing values





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



Sample

receiptId	barcode	description	finalPrice	itemPrice	needsFetchReview	partnerItemId	preventTargetGapPoi
0	5ff1e1eb0a720f0523000575	4011 ITEM NOT FOUND	26.00	26.00	False	1	True
1	5ff1e1bb0a720f052300056b	4011 ITEM NOT FOUND	1	1	NaN	1	NaN
2	5ff1e1bb0a720f052300056b	028400642255 DORITOS TORTILLA CHIP SPICY SWEET CHILI REDUCED FAT BAG 1 OZ	10.00	10.00	True	2	True
3	5ff1e1f10a720f052300057a	NaN NaN	NaN	NaN	False	1	True
4	5ff1e1ee0a7214ada100056f	4011 ITEM NOT FOUND	28.00	28.00	False	1	True
5	5ff1e1d20a7214ada1000561	4011 ITEM NOT FOUND	1	1	NaN	1	NaN
6	5ff1e1d20a7214ada1000561	1234 NaN	2.56	2.56	True	2	True
7	5ff1e1e40a7214ada1000566	4011 ITEM NOT FOUND	3.25	3.25	False	1	True
8	5ff1e1cd0a720f052300056f	NaN MSSN TORTLLA	2.23	2.23	NaN	1009	NaN
9	5ff1e1a40a720f0523000569	046000832517 Old El Paso Mild Chopped Green Chiles, 4.5 Oz	10.00	10.00	NaN	0	NaN

◀ ▶

receiptId	barcode	description
6931	B076FJ92M4	mueller austria hypergrind precision electric spice/coffee grinder millwith large grinding capacity and hd motor also for spices, herbs
6932	B07BRRLSVC	thindust summer face mask - sun protection neck gaiter for outdooractivities
6933	B076FJ92M4	mueller austria hypergrind precision electric spice/coffee grinder millwith large grinding capacity and hd motor also for spices, herbs
6934	B07BRRLSVC	thindust summer face mask - sun protection neck gaiter for outdooractivities
6935	B076FJ92M4	mueller austria hypergrind precision electric spice/coffee grinder millwith large grinding capacity and hd motor also for spices, herbs
6936	B07BRRLSVC	thindust summer face mask - sun protection neck gaiter for outdooractivities
6937	B076FJ92M4	mueller austria hypergrind precision electric spice/coffee grinder millwith large grinding capacity and hd motor also for spices, herbs
6938	B07BRRLSVC	thindust summer face mask - sun protection neck gaiter for outdooractivities
6939	B076FJ92M4	mueller austria hypergrind precision electric spice/coffee grinder millwith large grinding capacity and hd motor also for spices, herbs
6940	B07BRRLSVC	thindust summer face mask - sun protection neck gaiter for outdooractivities

