

Overview

Dataset statistics

Number of variables	50
Number of observations	101766
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	14.2 MiB
Average record size in memory	146.2 B

Variable types

Numeric	13
Categorical	37

Alerts

examide has constant value "No"	Constant
citoglipton has constant value "No"	Constant
medical_specialty has a high cardinality: 73 distinct values	High cardinality
diag_1 has a high cardinality: 717 distinct values	High cardinality
diag_2 has a high cardinality: 749 distinct values	High cardinality
diag_3 has a high cardinality: 790 distinct values	High cardinality
change is highly overall correlated with diabetesMed and 1 other fields (diabetesMed, insulin).	High correlation
diabetesMed is highly overall correlated with change and 1 other fields (change, insulin)	High correlation
encounter_id is highly overall correlated with patient_nbr	High correlation
insulin is highly overall correlated with change and 1 other fields (change, diabetesMed)	High correlation
patient_nbr is highly overall correlated with encounter_id	High correlation
race is highly imbalanced (55.9%)	Imbalance
weight is highly imbalanced (92.0%)	Imbalance
medical_specialty is highly imbalanced (54.2%)	Imbalance
max_glu_serum is highly imbalanced (81.2%)	Imbalance

A1Cresult is highly imbalanced (54.8%)	Imbalance
metformin is highly imbalanced (59.5%)	Imbalance
repaglinide is highly imbalanced (93.9%)	Imbalance
nateglinide is highly imbalanced (96.9%)	Imbalance
chlorpropamide is highly imbalanced (99.5%)	Imbalance
glimepiride is highly imbalanced (84.0%)	Imbalance
acetohexamide is highly imbalanced (> 99.9%)	Imbalance
glipizide is highly imbalanced (69.2%)	Imbalance
glyburide is highly imbalanced (72.3%)	Imbalance
tolbutamide is highly imbalanced (99.7%)	Imbalance
pioglitazone is highly imbalanced (80.2%)	Imbalance
rosiglitazone is highly imbalanced (82.2%)	Imbalance
acarbose is highly imbalanced (98.5%)	Imbalance
miglitol is highly imbalanced (99.7%)	Imbalance
troglitazone is highly imbalanced (> 99.9%)	Imbalance
tolazamide is highly imbalanced (99.7%)	Imbalance
glyburide-metformin is highly imbalanced (97.0%)	Imbalance
glipizide-metformin is highly imbalanced (99.8%)	Imbalance
glimepiride-pioglitazone is highly imbalanced (> 99.9%)	Imbalance
metformin-rosiglitazone is highly imbalanced (> 99.9%)	Imbalance
metformin-pioglitazone is highly imbalanced (> 99.9%)	Imbalance
number_emergency is highly skewed ($\gamma_1 = 22.85558215$)	Skewed
encounter_id has unique values	Unique
num_procedures has 46652 (45.8%) zeros	Zeros
number_outpatient has 85027 (83.6%) zeros	Zeros
number_emergency has 90383 (88.8%) zeros	Zeros
number_inpatient has 67630 (66.5%) zeros	Zeros

Reproduction

Analysis started	2024-07-01 03:15:25.345393
Analysis finished	2024-07-01 03:19:17.229436
Duration	3 minutes and 51.88 seconds
Software version	ydata-profiling v4.8.3 (https://github.com/ydataai/ydata-profiling)

Download config.json (data:text/plain;charset=utf-configuration 8,%7B%22title%22%3A%20%22YData%20Profiling%20Report%22%2C%20%22dataset%22%3A%20%7B%22description%22%3A%20%22%22%2C%20

Variables

Select Columns

encounter_id

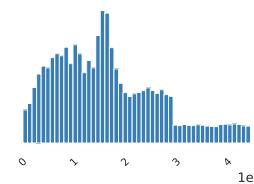
Real number (\mathbb{R})

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

UNIQUE

Distinct	101766
Distinct (%)	100.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1.6520165×10^8

Minimum	12522
Maximum	4.4386722×10^8
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	795.2 KiB

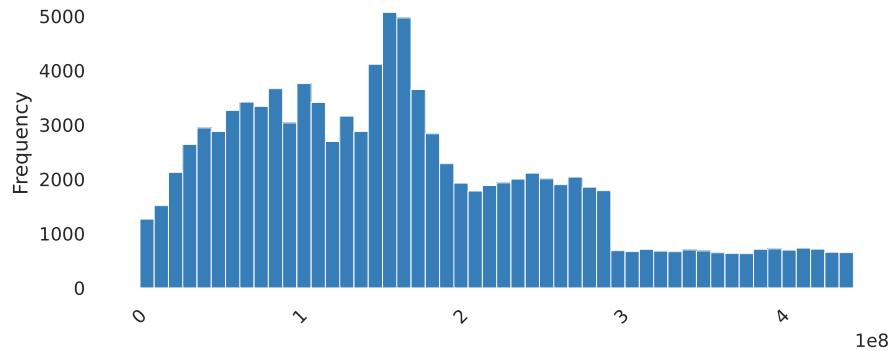


Quantile statistics

Minimum	12522
5-th percentile	27170784
Q1	84961194
median	1.5238899×10^8
Q3	2.3027089×10^8
95-th percentile	3.7896284×10^8
Maximum	4.4386722×10^8
Range	4.438547×10^8
Interquartile range (IQR)	1.4530969×10^8

Descriptive statistics

Standard deviation	1.026403×10^5
Coefficient of variation (CV)	0.62130311
Kurtosis	-0.10207139
Mean	1.6520165×10^8
Median Absolute Deviation (MAD)	70921143
Skewness	0.69914155
Sum	1.6811911×10^{13}
Variance	1.053503×10^{16}
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
2278392	1	< 0.1%
190792044	1	< 0.1%
190790070	1	< 0.1%
190789722	1	< 0.1%
190786806	1	< 0.1%
190785018	1	< 0.1%
190781412	1	< 0.1%
190775886	1	< 0.1%
190764504	1	< 0.1%
190760322	1	< 0.1%
Other values (101756)	101756	> 99.9%

Value	Count	Frequency (%)
12522	1	< 0.1%
15738	1	< 0.1%
16680	1	< 0.1%
28236	1	< 0.1%
35754	1	< 0.1%
36900	1	< 0.1%
40926	1	< 0.1%

Value	Count	Frequency (%)
42570	1	< 0.1%
55842	1	< 0.1%
62256	1	< 0.1%

Value	Count	Frequency (%)
443867222	1	< 0.1%
443857166	1	< 0.1%
443854148	1	< 0.1%
443847782	1	< 0.1%
443847548	1	< 0.1%
443847176	1	< 0.1%
443842778	1	< 0.1%
443842340	1	< 0.1%
443842136	1	< 0.1%
443842070	1	< 0.1%

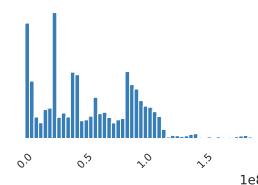
patient_nbr

Real number (\mathbb{R})

HIGH CORRELATION (This variable has a high overall correlation with 1 fields: encounter_id).

Distinct	71518
Distinct (%)	70.3%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	54330401

Minimum	135
Maximum	1.8950262×10^8
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	795.2 KiB



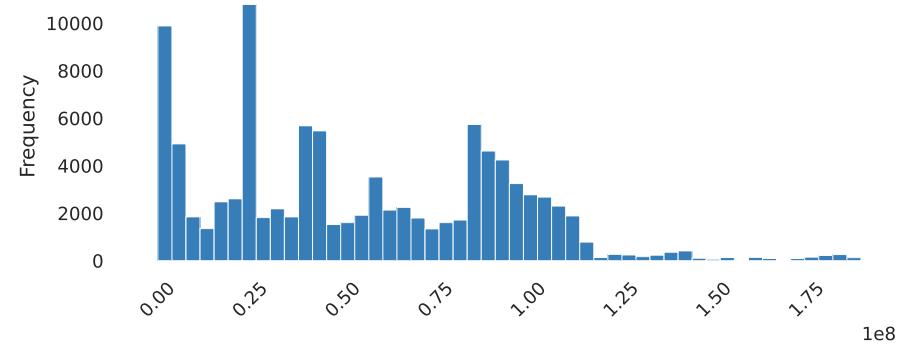
Quantile statistics

Minimum	135
5-th percentile	1456971.8
Q1	23413221
median	45505143
Q3	87545950
95-th percentile	1.1148027×10^8
Maximum	1.8950262×10^8
Range	1.8950248×10^8
Interquartile range (IQR)	64132729

Descriptive statistics

Standard deviation	38696359
Coefficient of variation (CV)	0.71224138
Kurtosis	-0.34737204
Mean	54330401
Median Absolute Deviation (MAD)	32950134
Skewness	0.47128072
Sum	5.5289876×10^{12}
Variance	1.4974082×10^{15}

Monotonicity Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
88785891	40	< 0.1%
43140906	28	< 0.1%
1660293	23	< 0.1%
88227540	23	< 0.1%
23199021	23	< 0.1%
23643405	22	< 0.1%
84428613	22	< 0.1%
92709351	21	< 0.1%
88789707	20	< 0.1%
29903877	20	< 0.1%
Other values (71508)	101524	99.8%

Value	Count	Frequency (%)
135	2	< 0.1%
378	1	< 0.1%
729	1	< 0.1%
774	1	< 0.1%
927	1	< 0.1%
1152	5	< 0.1%
1305	1	< 0.1%
1314	3	< 0.1%
1629	1	< 0.1%
2025	1	< 0.1%

Value	Count	Frequency (%)
189502619	1	< 0.1%
189481478	1	< 0.1%

Value	Count	Frequency (%)
189445127	1	< 0.1%
189365864	1	< 0.1%
189351095	1	< 0.1%
189349430	1	< 0.1%
189332087	1	< 0.1%
189298877	1	< 0.1%
189257846	2	< 0.1%
189215762	1	< 0.1%

race
Categorical

IMBALANCE

Distinct	6
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	100.1 KiB

Length	
Max length	15
Median length	9
Mean length	9.983521
Min length	5

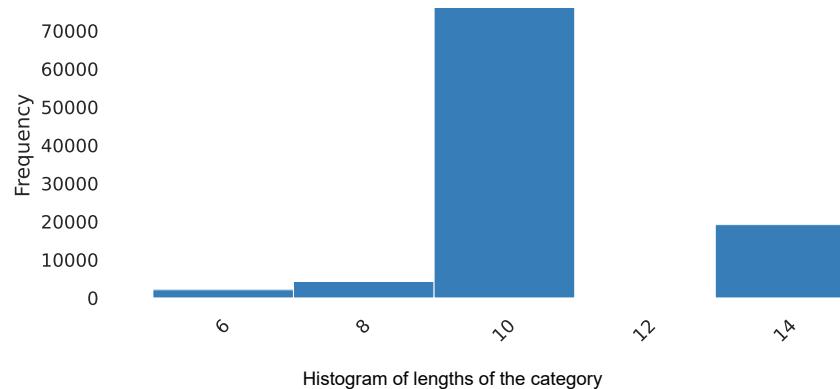
Characters and Unicode		Unique	Sample		
		Unique	0	?	
Total characters	1015983	Unique	0	?	1st row Caucasian
Distinct characters	19	Unique (%)	0.0%		2nd row Caucasian
Distinct categories	1				3rd row AfricanAmerican
https://en.wikipedia.org/wiki/Unicode_characters					?
Distinct scripts	1				4th row Caucasian
https://en.wikipedia.org/wiki/Script_(Unicode)					?
Distinct blocks	1	https://en.wikipedia.org/wiki/Unicode_blocks			5th row Caucasian
					?

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

Common Values

Value	Count	Frequency (%)
Caucasian	76099	74.8%
AfricanAmerican	19210	18.9%
Missing	2273	2.2%
Hispanic	2037	2.0%
Other	1506	1.5%
Asian	641	0.6%

Length



Common Values (Plot)



- Caucasian
- AfricanAmerican
- Missing
- Hispanic
- Other
- Asian

Value	Count	Frequency (%)
caucasian	76099	74.8%
africanamerican	19210	18.9%
missing	2273	2.2%
hispanic	2037	2.0%
other	1506	1.5%

Value	Count	Frequency (%)
asian	641	0.6%

Most occurring characters

Value	Count	Frequency (%)
a	269395	26.5%
i	123780	12.2%
n	119470	11.8%
c	116556	11.5%
s	83323	8.2%
C	76099	7.5%
u	76099	7.5%
r	39926	3.9%
A	39061	3.8%
e	20716	2.0%
Other values (9)	51558	5.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	1015983	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
a	269395	26.5%
i	123780	12.2%
n	119470	11.8%
c	116556	11.5%

Value	Count	Frequency (%)
s	83323	8.2%
C	76099	7.5%
u	76099	7.5%
r	39926	3.9%
A	39061	3.8%
e	20716	2.0%
Other values (9)	51558	5.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	1015983	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
a	269395	26.5%
i	123780	12.2%
n	119470	11.8%
c	116556	11.5%
s	83323	8.2%
C	76099	7.5%
u	76099	7.5%
r	39926	3.9%
A	39061	3.8%
e	20716	2.0%
Other values (9)	51558	5.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	1015983	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
a	269395	26.5%
i	123780	12.2%
n	119470	11.8%
c	116556	11.5%
s	83323	8.2%
C	76099	7.5%
u	76099	7.5%
r	39926	3.9%
A	39061	3.8%
e	20716	2.0%
Other values (9)	51558	5.1%

gender

Categorical

Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.8 KiB

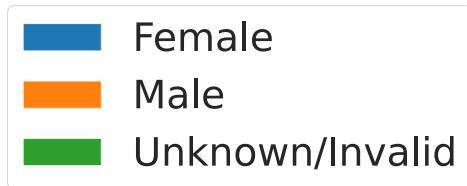
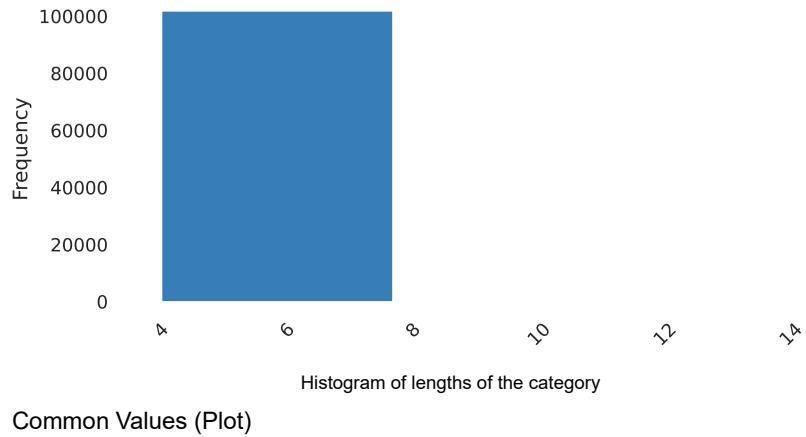
Length	Characters and Unicode	Unique			Sample		
		Unique	0	?	1st row	Female	
Max length	15	Total characters	516513	Unique (%)	0.0%	2nd row	Female
Median length	6	Distinct characters	16			3rd row	Female
Mean length	5.0754967	Distinct categories	1			4th row	Male
Min length	4	Distinct scripts	1	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)		5th row	Male
		Distinct blocks	1	(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)			?
							?

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

Common Values

Value	Count	Frequency (%)
Female	54708	53.8%
Male	47055	46.2%
Unknown/Invalid	3	< 0.1%

Length



Value	Count	Frequency (%)
female	54708	53.8%
male	47055	46.2%
unknown/invalid	3	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
e	156471	30.3%
a	101766	19.7%

Value	Count	Frequency (%)
I	101766	19.7%
F	54708	10.6%
m	54708	10.6%
M	47055	9.1%
n	12	< 0.1%
U	3	< 0.1%
k	3	< 0.1%
o	3	< 0.1%
Other values (6)	18	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	516513	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
e	156471	30.3%
a	101766	19.7%
I	101766	19.7%
F	54708	10.6%
m	54708	10.6%
M	47055	9.1%
n	12	< 0.1%
U	3	< 0.1%
k	3	< 0.1%
o	3	< 0.1%

Value	Count	Frequency (%)
Other values (6)	18	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	516513	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
e	156471	30.3%
a	101766	19.7%
l	101766	19.7%
F	54708	10.6%
m	54708	10.6%
M	47055	9.1%
n	12	< 0.1%
U	3	< 0.1%
k	3	< 0.1%
o	3	< 0.1%
Other values (6)	18	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	516513	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
e	156471	30.3%
a	101766	19.7%
l	101766	19.7%
F	54708	10.6%
m	54708	10.6%
M	47055	9.1%
n	12	< 0.1%
U	3	< 0.1%
k	3	< 0.1%
o	3	< 0.1%
Other values (6)	18	< 0.1%

age

Categorical

Distinct	10
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	100.4 KiB

Length	Characters and Unicode	Unique			Sample		
		Unique	0	?	1st row	[0-10)	
Max length	8	Total characters	714994	Unique (%)	0.0%	2nd row	[10-20)
Median length	7	Distinct characters	13			3rd row	[20-30)
Mean length	7.0258633	Distinct categories	1			4th row	[30-40)
Min length	6	Distinct scripts	1			5th row	[40-50)
		Distinct blocks	1				(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
							(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
							?
							?
							?

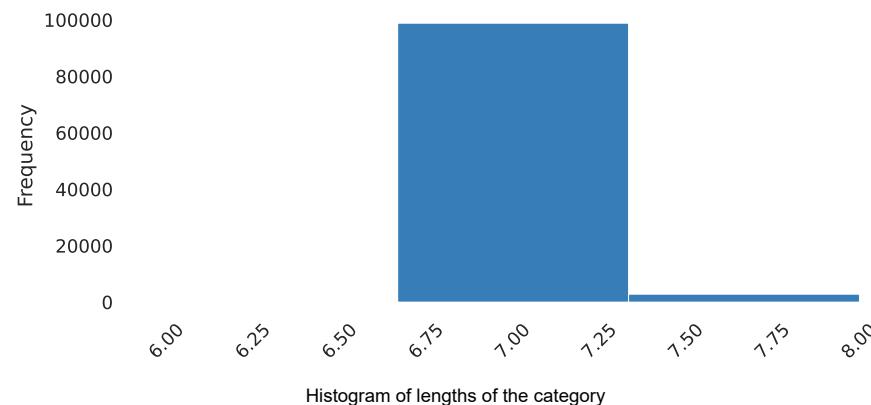
The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Common Values

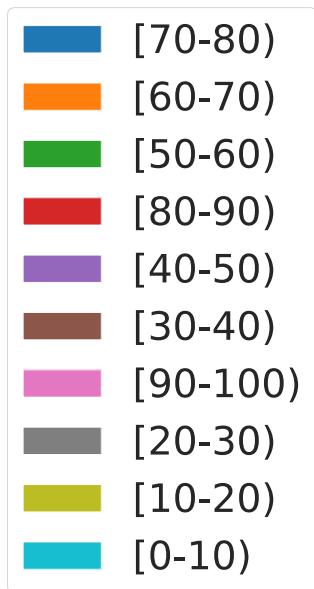
Value	Count	Frequency (%)
[70-80)	26068	25.6%
[60-70)	22483	22.1%
[50-60)	17256	17.0%
[80-90)	17197	16.9%
[40-50)	9685	9.5%
[30-40)	3775	3.7%
[90-100)	2793	2.7%

Value	Count	Frequency (%)
[20-30)	1657	1.6%
[10-20)	691	0.7%
[0-10)	161	0.2%

Length



Common Values (Plot)



Value	Count	Frequency (%)
70-80	26068	25.6%
60-70	22483	22.1%
50-60	17256	17.0%
80-90	17197	16.9%
40-50	9685	9.5%
30-40	3775	3.7%
90-100	2793	2.7%
20-30	1657	1.6%
10-20	691	0.7%
0-10	161	0.2%

Most occurring characters

Value	Count	Frequency (%)
0	206325	28.9%
[101766	14.2%
-	101766	14.2%
)	101766	14.2%
7	48551	6.8%
8	43265	6.1%
6	39739	5.6%
5	26941	3.8%
9	19990	2.8%
4	13460	1.9%
Other values (3)	11425	1.6%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	714994	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
0	206325	28.9%
[101766	14.2%
-	101766	14.2%
)	101766	14.2%
7	48551	6.8%
8	43265	6.1%

Value	Count	Frequency (%)
6	39739	5.6%
5	26941	3.8%
9	19990	2.8%
4	13460	1.9%
Other values (3)	11425	1.6%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	714994	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
0	206325	28.9%
[101766	14.2%
-	101766	14.2%
)	101766	14.2%
7	48551	6.8%
8	43265	6.1%
6	39739	5.6%
5	26941	3.8%
9	19990	2.8%
4	13460	1.9%
Other values (3)	11425	1.6%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	714994	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
0	206325	28.9%
[101766	14.2%
-	101766	14.2%
)	101766	14.2%
7	48551	6.8%
8	43265	6.1%
6	39739	5.6%
5	26941	3.8%
9	19990	2.8%
4	13460	1.9%
Other values (3)	11425	1.6%

weight

Categorical

IMBALANCE

Distinct	10
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	100.4 KiB

Length

		Characters and Unicode	Unique	?	Sample	
Max length	9		Unique	0	1st row	Missing
Median length	7	Total characters 715273	Unique (%)	0.0%	2nd row	Missing
Mean length	7.0286048	Distinct characters 14			3rd row	Missing
Min length	4				4th row	Missing
		Distinct categories 1			5th row	Missing
		(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)			?

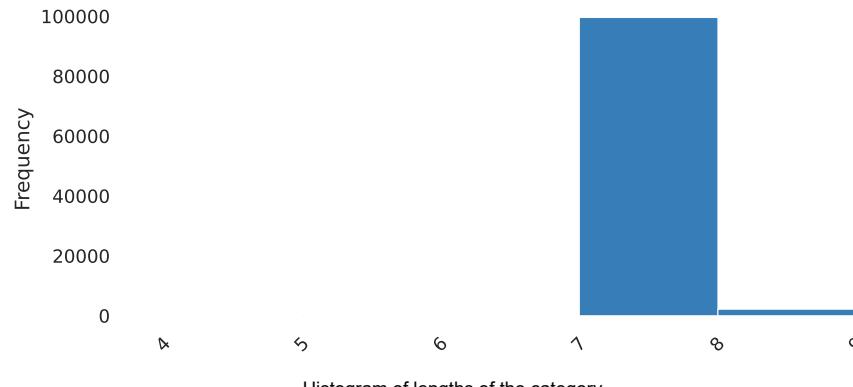
The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Common Values

Value	Count	Frequency (%)
Missing	98569	96.9%
[75-100)	1336	1.3%
[50-75)	897	0.9%
[100-125)	625	0.6%
[125-150)	145	0.1%
[25-50)	97	0.1%

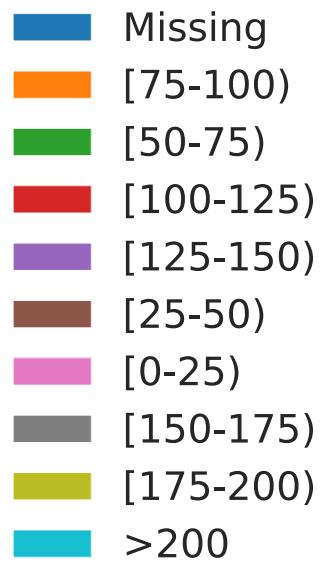
Value	Count	Frequency (%)
[0-25)	48	< 0.1%
[150-175)	35	< 0.1%
[175-200)	11	< 0.1%
>200	3	< 0.1%

Length



Histogram of lengths of the category

Common Values (Plot)



Value	Count	Frequency (%)
missing	98569	96.9%
75-100	1336	1.3%
50-75	897	0.9%
100-125	625	0.6%
125-150	145	0.1%
25-50	97	0.1%
0-25	48	< 0.1%
150-175	35	< 0.1%
175-200	11	< 0.1%
200	3	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
i	197138	27.6%
s	197138	27.6%
M	98569	13.8%
n	98569	13.8%
g	98569	13.8%
0	5172	0.7%
5	4368	0.6%
[3194	0.4%
-	3194	0.4%
)	3194	0.4%
Other values (4)	6168	0.9%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	715273	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
i	197138	27.6%
s	197138	27.6%
M	98569	13.8%
n	98569	13.8%
g	98569	13.8%
0	5172	0.7%

Value	Count	Frequency (%)
5	4368	0.6%
[3194	0.4%
-	3194	0.4%
)	3194	0.4%
Other values (4)	6168	0.9%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	715273	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
i	197138	27.6%
s	197138	27.6%
M	98569	13.8%
n	98569	13.8%
g	98569	13.8%
0	5172	0.7%
5	4368	0.6%
[3194	0.4%
-	3194	0.4%
)	3194	0.4%
Other values (4)	6168	0.9%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	715273	100.0%

Most frequent character per block

(unknown)

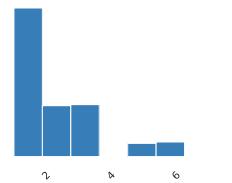
Value	Count	Frequency (%)
i	197138	27.6%
s	197138	27.6%
M	98569	13.8%
n	98569	13.8%
g	98569	13.8%
0	5172	0.7%
5	4368	0.6%
[3194	0.4%
-	3194	0.4%
)	3194	0.4%
Other values (4)	6168	0.9%

admission_type_id

Real number (\mathbb{R})

Distinct	8
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	2.0240061

Minimum	1
Maximum	8
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	795.2 KiB

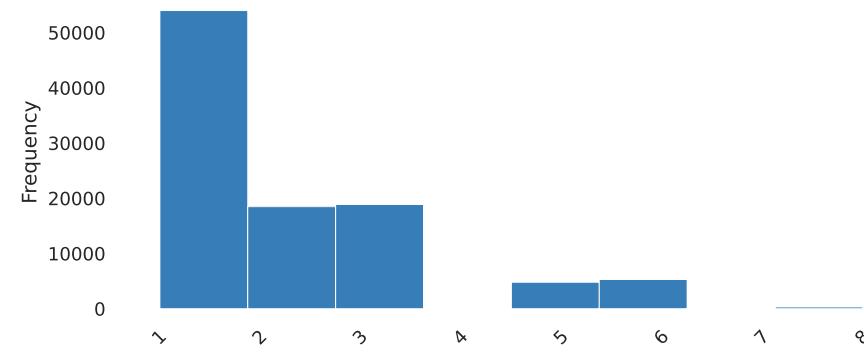


Quantile statistics

Minimum	1
5-th percentile	1
Q1	1
median	1
Q3	3
95-th percentile	6
Maximum	8
Range	7
Interquartile range (IQR)	2

Descriptive statistics

Standard deviation	1.4454028
Coefficient of variation (CV)	0.7141297
Kurtosis	1.9424761
Mean	2.0240061
Median Absolute Deviation (MAD)	0
Skewness	1.5919843
Sum	205975
Variance	2.0891893
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=8)

Value	Count	Frequency (%)
1	53990	53.1%
3	18869	18.5%
2	18480	18.2%
6	5291	5.2%
5	4785	4.7%
8	320	0.3%
7	21	< 0.1%
4	10	< 0.1%

Value	Count	Frequency (%)
1	53990	53.1%
2	18480	18.2%
3	18869	18.5%
4	10	< 0.1%
5	4785	4.7%
6	5291	5.2%
7	21	< 0.1%
8	320	0.3%

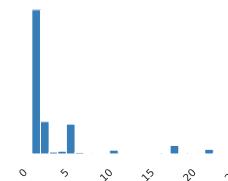
Value	Count	Frequency (%)
8	320	0.3%
7	21	< 0.1%
6	5291	5.2%
5	4785	4.7%
4	10	< 0.1%
3	18869	18.5%
2	18480	18.2%
1	53990	53.1%

discharge_disposition_id

Real number (\mathbb{R})

Distinct	26
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	3.7156418

Minimum	1
Maximum	28
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	795.2 KiB

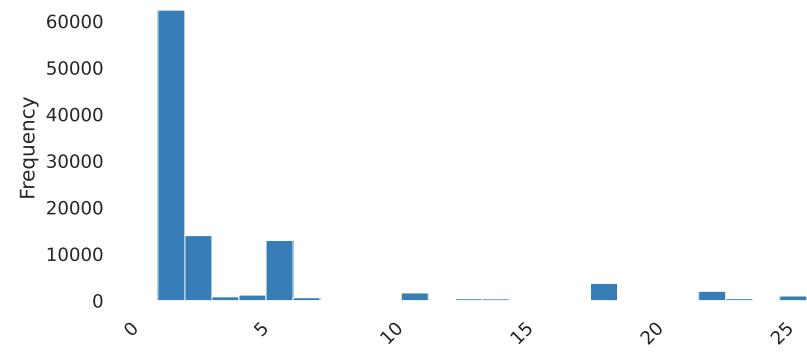


Quantile statistics

Minimum	1
5-th percentile	1
Q1	1
median	1
Q3	4
95-th percentile	18
Maximum	28
Range	27
Interquartile range (IQR)	3

Descriptive statistics

Standard deviation	5.2801655
Coefficient of variation (CV)	1.4210642
Kurtosis	6.0033468
Mean	3.7156418
Median Absolute Deviation (MAD)	0
Skewness	2.563067
Sum	378126
Variance	27.880148
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=26)

Value	Count	Frequency (%)
1	60234	59.2%
3	13954	13.7%
6	12902	12.7%
18	3691	3.6%
2	2128	2.1%
22	1993	2.0%
11	1642	1.6%
5	1184	1.2%
25	989	1.0%
4	815	0.8%
Other values (16)	2234	2.2%

Value	Count	Frequency (%)
1	60234	59.2%
2	2128	2.1%
3	13954	13.7%
4	815	0.8%
5	1184	1.2%
6	12902	12.7%
7	623	0.6%
8	108	0.1%
9	21	< 0.1%
10	6	< 0.1%

Value	Count	Frequency (%)
28	139	0.1%
27	5	< 0.1%
25	989	1.0%

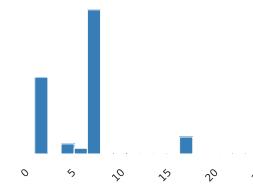
Value	Count	Frequency (%)
24	48	< 0.1%
23	412	0.4%
22	1993	2.0%
20	2	< 0.1%
19	8	< 0.1%
18	3691	3.6%
17	14	< 0.1%

admission_source_id

Real number (\mathbb{R})

Distinct	17
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	5.7544366

Minimum	1
Maximum	25
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	795.2 KiB

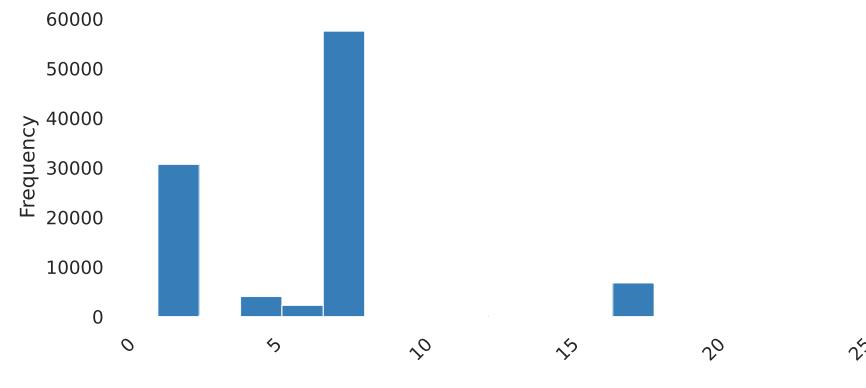


Quantile statistics

Minimum	1
5-th percentile	1
Q1	1
median	7
Q3	7
95-th percentile	17
Maximum	25
Range	24
Interquartile range (IQR)	6

Descriptive statistics

Standard deviation	4.0640808
Coefficient of variation (CV)	0.70625173
Kurtosis	1.7449894
Mean	5.7544366
Median Absolute Deviation (MAD)	0
Skewness	1.0299349
Sum	585606
Variance	16.516753
Monotonicity	Not monotonic



Value	Count	Frequency (%)
7	57494	56.5%
1	29565	29.1%
17	6781	6.7%
4	3187	3.1%
6	2264	2.2%
2	1104	1.1%
5	855	0.8%
3	187	0.2%
20	161	0.2%
9	125	0.1%
Other values (7)	43	< 0.1%

Value	Count	Frequency (%)
1	29565	29.1%
2	1104	1.1%
3	187	0.2%
4	3187	3.1%
5	855	0.8%
6	2264	2.2%
7	57494	56.5%
8	16	< 0.1%
9	125	0.1%
10	8	< 0.1%

Value	Count	Frequency (%)
25	2	< 0.1%
22	12	< 0.1%
20	161	0.2%

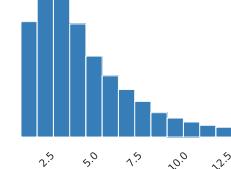
Value	Count	Frequency (%)
17	6781	6.7%
14	2	< 0.1%
13	1	< 0.1%
11	2	< 0.1%
10	8	< 0.1%
9	125	0.1%
8	16	< 0.1%

time_in_hospital

Real number (\mathbb{R})

Distinct	14
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	4.3959869

Minimum	1
Maximum	14
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	795.2 KiB

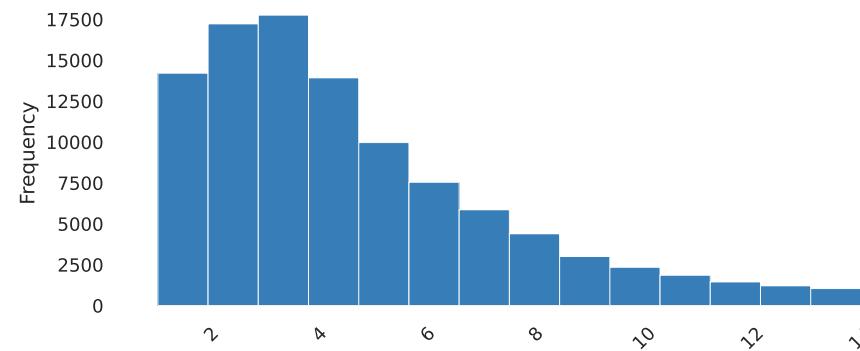


Quantile statistics

Minimum	1
5-th percentile	1
Q1	2
median	4
Q3	6
95-th percentile	11
Maximum	14
Range	13
Interquartile range (IQR)	4

Descriptive statistics

Standard deviation	2.9851078
Coefficient of variation (CV)	0.67905293
Kurtosis	0.85025084
Mean	4.3959869
Median Absolute Deviation (MAD)	2
Skewness	1.1339987
Sum	447362
Variance	8.9108684
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=14)

Value	Count	Frequency (%)
3	17756	17.4%
2	17224	16.9%
1	14208	14.0%
4	13924	13.7%
5	9966	9.8%
6	7539	7.4%
7	5859	5.8%
8	4391	4.3%
9	3002	2.9%
10	2342	2.3%
Other values (4)	5555	5.5%

Value	Count	Frequency (%)
1	14208	14.0%
2	17224	16.9%
3	17756	17.4%
4	13924	13.7%
5	9966	9.8%
6	7539	7.4%
7	5859	5.8%
8	4391	4.3%
9	3002	2.9%
10	2342	2.3%

Value	Count	Frequency (%)
14	1042	1.0%
13	1210	1.2%
12	1448	1.4%

Value	Count	Frequency (%)
11	1855	1.8%
10	2342	2.3%
9	3002	2.9%
8	4391	4.3%
7	5859	5.8%
6	7539	7.4%
5	9966	9.8%

payer_code

Categorical

Distinct	18
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	101.1 KiB

Length	Characters and Unicode	Unique			Sample	
		Unique	1	?	1st row	Missing
Max length	7	Total characters	404812	Unique (%) < 0.1%	2nd row	Missing
Median length	2	Distinct characters	20		3rd row	Missing
Mean length	3.9778708	Distinct categories	1		4th row	Missing
Min length	2	Distinct scripts	1	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	5th row	Missing ?
		Distinct blocks	1	(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)		?

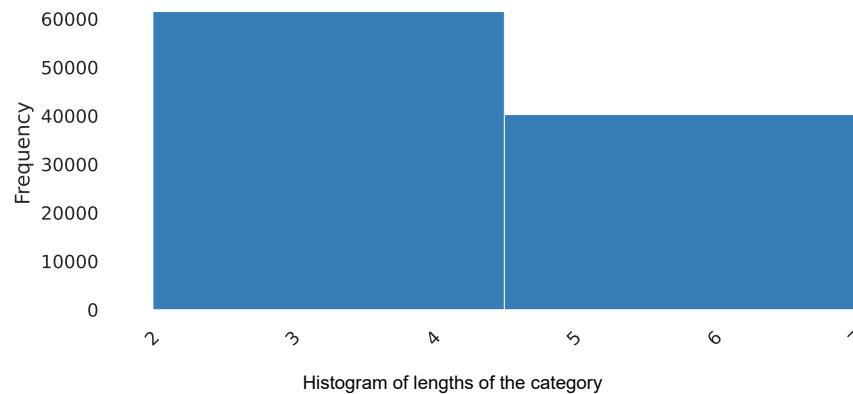
The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Common Values

Value	Count	Frequency (%)
Missing	40256	39.6%
MC	32439	31.9%
HM	6274	6.2%
SP	5007	4.9%
BC	4655	4.6%
MD	3532	3.5%
CP	2533	2.5%

Value	Count	Frequency (%)
UN	2448	2.4%
CM	1937	1.9%
OG	1033	1.0%
Other values (8)	1652	1.6%

Length



Value	Count	Frequency (%)
missing	40256	39.6%
mc	32439	31.9%
hm	6274	6.2%
sp	5007	4.9%
bc	4655	4.6%
md	3532	3.5%
cp	2533	2.5%
un	2448	2.4%
cm	1937	1.9%
og	1033	1.0%
Other values (8)	1652	1.6%

Most occurring characters

Value	Count	Frequency (%)
M	85066	21.0%
i	80512	19.9%
s	80512	19.9%
C	41845	10.3%
n	40256	9.9%
g	40256	9.9%
P	8211	2.0%
H	6420	1.6%
S	5062	1.3%
B	4655	1.1%
Other values (10)	12017	3.0%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	404812	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
M	85066	21.0%
i	80512	19.9%
s	80512	19.9%
C	41845	10.3%
n	40256	9.9%
g	40256	9.9%
P	8211	2.0%
H	6420	1.6%

Value	Count	Frequency (%)
S	5062	1.3%
B	4655	1.1%
Other values (10)	12017	3.0%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	404812	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
M	85066	21.0%
i	80512	19.9%
s	80512	19.9%
C	41845	10.3%
n	40256	9.9%
g	40256	9.9%
P	8211	2.0%
H	6420	1.6%
S	5062	1.3%
B	4655	1.1%
Other values (10)	12017	3.0%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	404812	100.0%

Most frequent character per block

(*unknown*)

Value	Count	Frequency (%)
M	85066	21.0%
i	80512	19.9%
s	80512	19.9%
C	41845	10.3%
n	40256	9.9%
g	40256	9.9%
P	8211	2.0%
H	6420	1.6%
S	5062	1.3%
B	4655	1.1%
Other values (10)	12017	3.0%

medical_specialty

Categorical

HIGH CARDINALITY IMBALANCE

Distinct	73
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Memory size	106.8 KiB

Length	Characters and Unicode	Unique			Sample	
		Unique	9	?	1st row	Pediatrics-Endocrinology
Max length	36	Total characters	1176171	Unique (%) < 0.1%	2nd row	Missing
Median length	33	Distinct characters	43		3rd row	Missing
Mean length	11.557603	Distinct categories	1	(https://en.wikipedia.org/wiki/Unicode_characters)	4th row	Missing category)
Min length	6	Distinct scripts	1	(https://en.wikipedia.org/wiki/Script_(Unicode_scripts))	5th row	Missing code)
	Distinct blocks	1	(https://en.wikipedia.org/wiki/Unicode_blocks)			?

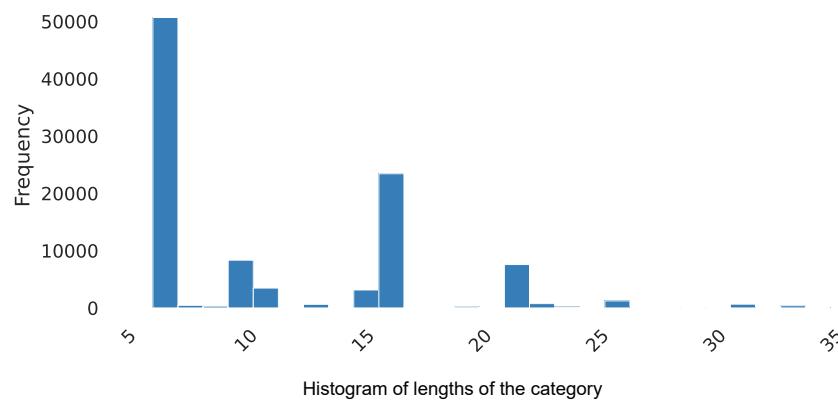
The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Common Values

Value	Count	Frequency (%)
Missing	49949	49.1%
InternalMedicine	14635	14.4%
Emergency/Trauma	7565	7.4%
Family/GeneralPractice	7440	7.3%
Cardiology	5352	5.3%
Surgery-General	3099	3.0%

Value	Count	Frequency (%)
Nephrology	1613	1.6%
Orthopedics	1400	1.4%
Orthopedics-Reconstructive	1233	1.2%
Radiologist	1140	1.1%
Other values (63)	8340	8.2%

Length



Value	Count	Frequency (%)
missing	49949	49.1%
internalmedicine	14635	14.4%
emergency/trauma	7565	7.4%
family/generalpractice	7440	7.3%
cardiology	5352	5.3%
surgery-general	3099	3.0%
nephrology	1613	1.6%
orthopedics	1400	1.4%
orthopedics-reconstructive	1233	1.2%
radiologist	1140	1.1%
Other values (63)	8340	8.2%

Most occurring characters

Value	Count	Frequency (%)
i	163206	13.9%
n	118747	10.1%
s	110536	9.4%
e	105151	8.9%
r	76899	6.5%
g	75545	6.4%
a	71149	6.0%
M	65004	5.5%
c	50007	4.3%
l	48871	4.2%
Other values (33)	291056	24.7%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	1176171	100.0%

Most frequent character per category

(*unknown*)

Value	Count	Frequency (%)
i	163206	13.9%
n	118747	10.1%
s	110536	9.4%
e	105151	8.9%
r	76899	6.5%
g	75545	6.4%
a	71149	6.0%

Value	Count	Frequency (%)
M	65004	5.5%
c	50007	4.3%
l	48871	4.2%
Other values (33)	291056	24.7%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	1176171	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
i	163206	13.9%
n	118747	10.1%
s	110536	9.4%
e	105151	8.9%
r	76899	6.5%
g	75545	6.4%
a	71149	6.0%
M	65004	5.5%
c	50007	4.3%
l	48871	4.2%
Other values (33)	291056	24.7%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	1176171	100.0%

Most frequent character per block

(unknown)

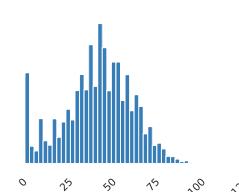
Value	Count	Frequency (%)
i	163206	13.9%
n	118747	10.1%
s	110536	9.4%
e	105151	8.9%
r	76899	6.5%
g	75545	6.4%
a	71149	6.0%
M	65004	5.5%
c	50007	4.3%
l	48871	4.2%
Other values (33)	291056	24.7%

num_lab_procedures

Real number (\mathbb{R})

Distinct	118
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	43.095641

Minimum	1
Maximum	132
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	795.2 KiB

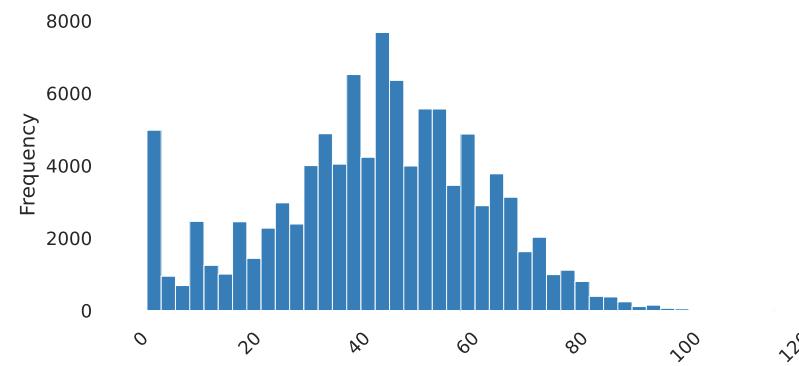


Quantile statistics

Minimum	1
5-th percentile	4
Q1	31
median	44
Q3	57
95-th percentile	73
Maximum	132
Range	131
Interquartile range (IQR)	26

Descriptive statistics

Standard deviation	19.674362
Coefficient of variation (CV)	0.45652789
Kurtosis	-0.24507352
Mean	43.095641
Median Absolute Deviation (MAD)	13
Skewness	-0.23654392
Sum	4385671
Variance	387.08053
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=50)

Value	Count	Frequency (%)
1	3208	3.2%
43	2804	2.8%
44	2496	2.5%
45	2376	2.3%
38	2213	2.2%
40	2201	2.2%
46	2189	2.2%
41	2117	2.1%
42	2113	2.1%
47	2106	2.1%
Other values (108)	77943	76.6%

Value	Count	Frequency (%)
1	3208	3.2%
2	1101	1.1%
3	668	0.7%
4	378	0.4%
5	286	0.3%
6	282	0.3%
7	323	0.3%
8	366	0.4%
9	933	0.9%
10	838	0.8%

Value	Count	Frequency (%)
132	1	< 0.1%
129	1	< 0.1%
126	1	< 0.1%

Value	Count	Frequency (%)
121	1	< 0.1%
120	1	< 0.1%
118	1	< 0.1%
114	2	< 0.1%
113	3	< 0.1%
111	3	< 0.1%
109	4	< 0.1%

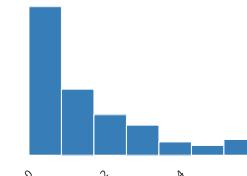
num_procedures

Real number (\mathbb{R})

ZEROS

Distinct	7
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	1.3397304

Minimum	0
Maximum	6
Zeros	46652
Zeros (%)	45.8%
Negative	0
Negative (%)	0.0%
Memory size	795.2 KiB

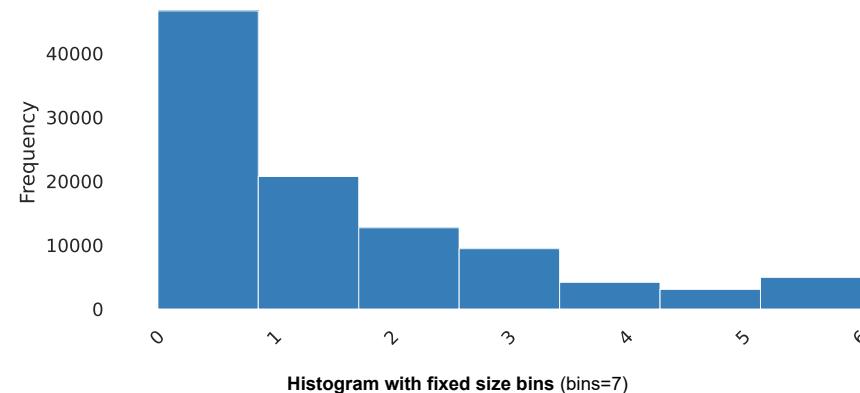


Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	1
Q3	2
95-th percentile	5
Maximum	6
Range	6
Interquartile range (IQR)	2

Descriptive statistics

Standard deviation	1.705807
Coefficient of variation (CV)	1.2732465
Kurtosis	0.8571103
Mean	1.3397304
Median Absolute Deviation (MAD)	1
Skewness	1.3164148
Sum	136339
Variance	2.9097775
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	46652	45.8%
1	20742	20.4%
2	12717	12.5%
3	9443	9.3%
6	4954	4.9%
4	4180	4.1%
5	3078	3.0%

Value	Count	Frequency (%)
0	46652	45.8%
1	20742	20.4%
2	12717	12.5%
3	9443	9.3%
4	4180	4.1%
5	3078	3.0%
6	4954	4.9%

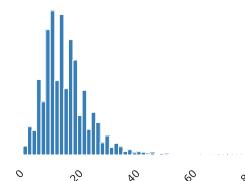
Value	Count	Frequency (%)
6	4954	4.9%
5	3078	3.0%
4	4180	4.1%
3	9443	9.3%
2	12717	12.5%
1	20742	20.4%
0	46652	45.8%

num_medications

Real number (\mathbb{R})

Distinct	75
Distinct (%)	0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	16.021844

Minimum	1
Maximum	81
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	795.2 KiB

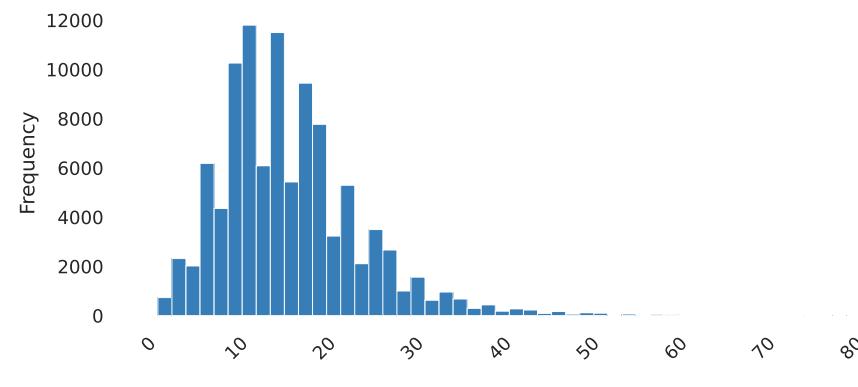


Quantile statistics

Minimum	1
5-th percentile	6
Q1	10
median	15
Q3	20
95-th percentile	31
Maximum	81
Range	80
Interquartile range (IQR)	10

Descriptive statistics

Standard deviation	8.1275662
Coefficient of variation (CV)	0.50728032
Kurtosis	3.4681549
Mean	16.021844
Median Absolute Deviation (MAD)	5
Skewness	1.3266721
Sum	1630479
Variance	66.057332
Monotonicity	Not monotonic



Value	Count	Frequency (%)
13	6086	6.0%
12	6004	5.9%
11	5795	5.7%
15	5792	5.7%
14	5707	5.6%
16	5430	5.3%
10	5346	5.3%
17	4919	4.8%
9	4913	4.8%
18	4523	4.4%
Other values (65)	47251	46.4%

Value	Count	Frequency (%)
1	262	0.3%
2	470	0.5%
3	900	0.9%
4	1417	1.4%
5	2017	2.0%
6	2699	2.7%
7	3484	3.4%
8	4353	4.3%
9	4913	4.8%
10	5346	5.3%

Value	Count	Frequency (%)
81	1	< 0.1%
79	1	< 0.1%
75	2	< 0.1%

Value	Count	Frequency (%)
74	1	< 0.1%
72	3	< 0.1%
70	2	< 0.1%
69	5	< 0.1%
68	7	< 0.1%
67	7	< 0.1%
66	5	< 0.1%

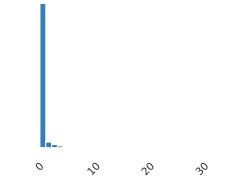
number_outpatient

Real number (\mathbb{R})

ZEROS

Distinct	39
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	0.36935715

Minimum	0
Maximum	42
Zeros	85027
Zeros (%)	83.6%
Negative	0
Negative (%)	0.0%
Memory size	795.2 KiB

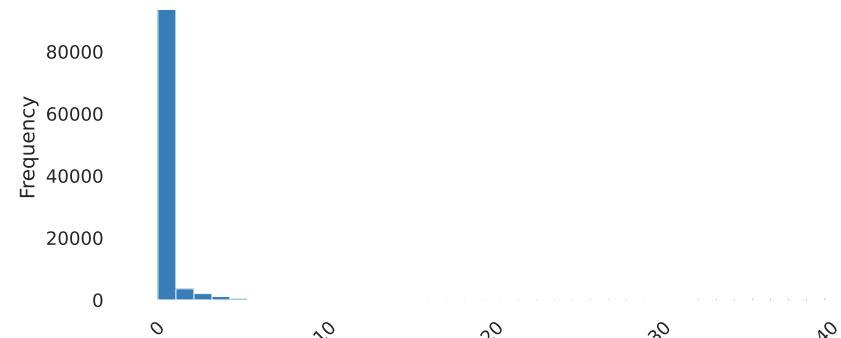


Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	2
Maximum	42
Range	42
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	1.2672651
Coefficient of variation (CV)	3.4310019
Kurtosis	147.90774
Mean	0.36935715
Median Absolute Deviation (MAD)	0
Skewness	8.8329589
Sum	37588
Variance	1.6059608
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	85027	83.6%
1	8547	8.4%
2	3594	3.5%
3	2042	2.0%
4	1099	1.1%
5	533	0.5%
6	303	0.3%
7	155	0.2%
8	98	0.1%
9	83	0.1%
Other values (29)	285	0.3%

Value	Count	Frequency (%)
0	85027	83.6%
1	8547	8.4%
2	3594	3.5%
3	2042	2.0%
4	1099	1.1%
5	533	0.5%
6	303	0.3%
7	155	0.2%
8	98	0.1%
9	83	0.1%

Value	Count	Frequency (%)
42	1	< 0.1%
40	1	< 0.1%

Value	Count	Frequency (%)
39	1	< 0.1%
38	1	< 0.1%
37	1	< 0.1%
36	2	< 0.1%
35	2	< 0.1%
34	1	< 0.1%
33	2	< 0.1%
29	2	< 0.1%

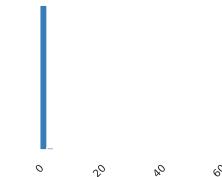
number_emergency

Real number (\mathbb{R})

SKEWED ZEROS

Distinct	33
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	0.19783621

Minimum	0
Maximum	76
Zeros	90383
Zeros (%)	88.8%
Negative	0
Negative (%)	0.0%
Memory size	795.2 KiB

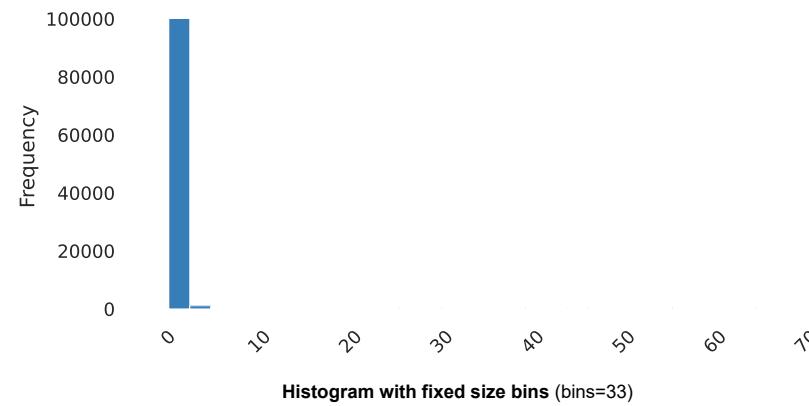


Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	0
95-th percentile	1
Maximum	76
Range	76
Interquartile range (IQR)	0

Descriptive statistics

Standard deviation	0.93047227
Coefficient of variation (CV)	4.7032455
Kurtosis	1191.6867
Mean	0.19783621
Median Absolute Deviation (MAD)	0
Skewness	22.855582
Sum	20133
Variance	0.86577864
Monotonicity	Not monotonic



Value	Count	Frequency (%)
0	90383	88.8%
1	7677	7.5%
2	2042	2.0%
3	725	0.7%
4	374	0.4%
5	192	0.2%
6	94	0.1%
7	73	0.1%
8	50	< 0.1%
10	34	< 0.1%
Other values (23)	122	0.1%

Value	Count	Frequency (%)
0	90383	88.8%
1	7677	7.5%
2	2042	2.0%
3	725	0.7%
4	374	0.4%
5	192	0.2%
6	94	0.1%
7	73	0.1%
8	50	< 0.1%
9	33	< 0.1%

Value	Count	Frequency (%)
76	1	< 0.1%
64	1	< 0.1%

Value	Count	Frequency (%)
63	1	< 0.1%
54	1	< 0.1%
46	1	< 0.1%
42	1	< 0.1%
37	1	< 0.1%
29	1	< 0.1%
28	1	< 0.1%
25	2	< 0.1%

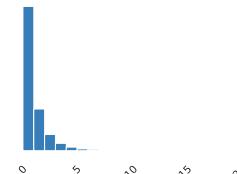
number_inpatient

Real number (\mathbb{R})

ZEROS

Distinct	21
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	0.63556591

Minimum	0
Maximum	21
Zeros	67630
Zeros (%)	66.5%
Negative	0
Negative (%)	0.0%
Memory size	795.2 KiB

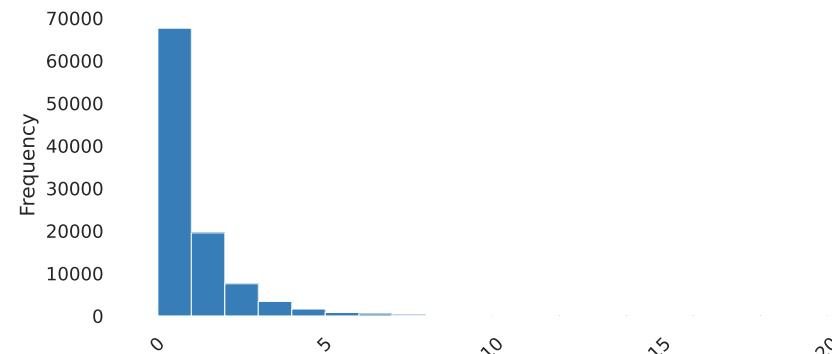


Quantile statistics

Minimum	0
5-th percentile	0
Q1	0
median	0
Q3	1
95-th percentile	3
Maximum	21
Range	21
Interquartile range (IQR)	1

Descriptive statistics

Standard deviation	1.2628633
Coefficient of variation (CV)	1.9869903
Kurtosis	20.719397
Mean	0.63556591
Median Absolute Deviation (MAD)	0
Skewness	3.614139
Sum	64679
Variance	1.5948237
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=21)

Value	Count	Frequency (%)
0	67630	66.5%
1	19521	19.2%
2	7566	7.4%
3	3411	3.4%
4	1622	1.6%
5	812	0.8%
6	480	0.5%
7	268	0.3%
8	151	0.1%
9	111	0.1%
Other values (11)	194	0.2%

Value	Count	Frequency (%)
0	67630	66.5%
1	19521	19.2%
2	7566	7.4%
3	3411	3.4%
4	1622	1.6%
5	812	0.8%
6	480	0.5%
7	268	0.3%
8	151	0.1%
9	111	0.1%

Value	Count	Frequency (%)
21	1	< 0.1%
19	2	< 0.1%
18	1	< 0.1%

Value	Count	Frequency (%)
17	1	< 0.1%
16	6	< 0.1%
15	9	< 0.1%
14	10	< 0.1%
13	20	< 0.1%
12	34	< 0.1%
11	49	< 0.1%

diag_1

Categorical

HIGH CARDINALITY

Distinct	717
Distinct (%)	0.7%
Missing	0
Missing (%)	0.0%
Memory size	257.1 KiB

Length		Characters and Unicode	Unique		Sample	
			Unique	82 ?	1st row	250.83
Max length	7	Total characters	323255	Unique (%) 0.1%	2nd row	276
Median length	3	Distinct characters	18		3rd row	648
Mean length	3.1764538	Distinct categories	1		4th row	8
Min length	1	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)			5th row	197 ?
		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)		?

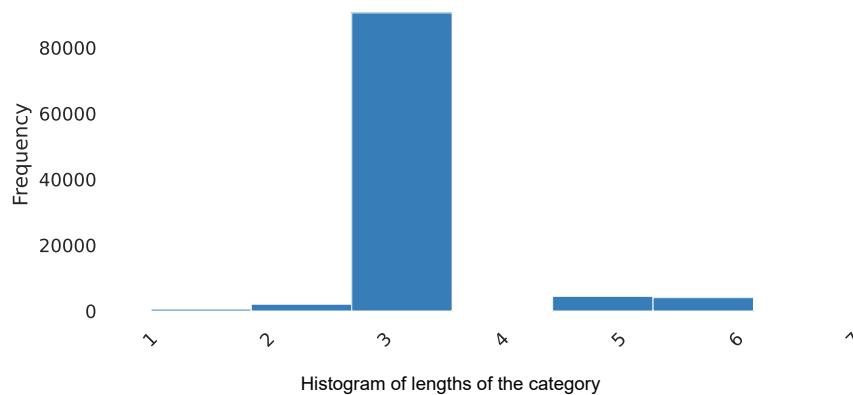
The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Common Values

Value	Count	Frequency (%)
428	6862	6.7%
414	6581	6.5%
786	4016	3.9%
410	3614	3.6%
486	3508	3.4%
427	2766	2.7%

Value	Count	Frequency (%)
491	2275	2.2%
715	2151	2.1%
682	2042	2.0%
434	2028	2.0%
Other values (707)	65923	64.8%

Length



Value	Count	Frequency (%)
428	6862	6.7%
414	6581	6.5%
786	4016	3.9%
410	3614	3.6%
486	3508	3.4%
427	2766	2.7%
491	2275	2.2%
715	2151	2.1%
682	2042	2.0%
434	2028	2.0%
Other values (707)	65923	64.8%

Most occurring characters

Value	Count	Frequency (%)
4	55457	17.2%
2	39876	12.3%
8	37949	11.7%
5	37131	11.5%
7	28668	8.9%
1	28106	8.7%
0	24960	7.7%
6	23198	7.2%
9	19978	6.2%
3	17618	5.5%
Other values (8)	10314	3.2%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	323255	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
4	55457	17.2%
2	39876	12.3%
8	37949	11.7%
5	37131	11.5%
7	28668	8.9%
1	28106	8.7%
0	24960	7.7%

Value	Count	Frequency (%)
6	23198	7.2%
9	19978	6.2%
3	17618	5.5%
Other values (8)	10314	3.2%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	323255	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
4	55457	17.2%
2	39876	12.3%
8	37949	11.7%
5	37131	11.5%
7	28668	8.9%
1	28106	8.7%
0	24960	7.7%
6	23198	7.2%
9	19978	6.2%
3	17618	5.5%
Other values (8)	10314	3.2%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	323255	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
4	55457	17.2%
2	39876	12.3%
8	37949	11.7%
5	37131	11.5%
7	28668	8.9%
1	28106	8.7%
0	24960	7.7%
6	23198	7.2%
9	19978	6.2%
3	17618	5.5%
Other values (8)	10314	3.2%

diag_2

Categorical

HIGH CARDINALITY

Distinct	749
Distinct (%)	0.7%
Missing	0
Missing (%)	0.0%
Memory size	259.1 KiB

Length		Characters and Unicode	Unique		Sample	
			Unique	124 ?	1st row	Missing
Max length	7	Total characters	324359	Unique (%) 0.1%	2nd row	250.01
Median length	3	Distinct characters	18		3rd row	250
Mean length	3.1873022	Distinct categories	1		4th row	250.43
Min length	1	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)			5th row	157 ?
		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)		?

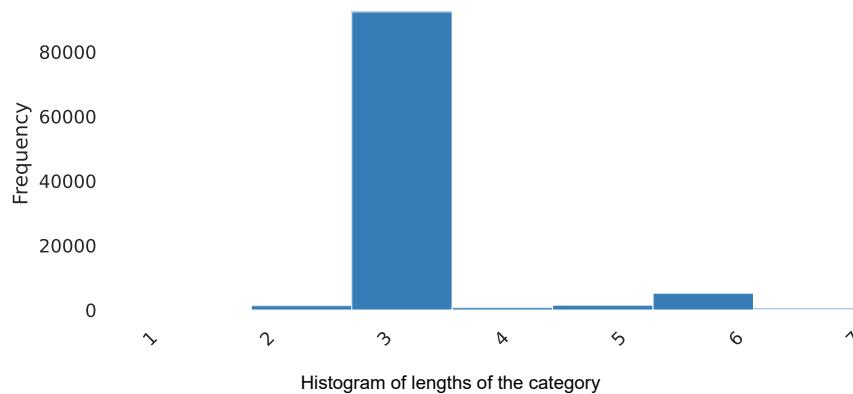
The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Common Values

Value	Count	Frequency (%)
276	6752	6.6%
428	6662	6.5%
250	6071	6.0%
427	5036	4.9%
401	3736	3.7%
496	3305	3.2%

Value	Count	Frequency (%)
599	3288	3.2%
403	2823	2.8%
414	2650	2.6%
411	2566	2.5%
Other values (739)	58877	57.9%

Length



Value	Count	Frequency (%)
276	6752	6.6%
428	6662	6.5%
250	6071	6.0%
427	5036	4.9%
401	3736	3.7%
496	3305	3.2%
599	3288	3.2%
403	2823	2.8%
414	2650	2.6%
411	2566	2.5%
Other values (739)	58877	57.9%

Most occurring characters

Value	Count	Frequency (%)
4	51155	15.8%
2	49765	15.3%
5	38176	11.8%
0	34046	10.5%
8	28711	8.9%
7	28654	8.8%
1	26158	8.1%
9	21842	6.7%
6	19990	6.2%
3	14097	4.3%
Other values (8)	11765	3.6%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	324359	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
4	51155	15.8%
2	49765	15.3%
5	38176	11.8%
0	34046	10.5%
8	28711	8.9%
7	28654	8.8%
1	26158	8.1%

Value	Count	Frequency (%)
9	21842	6.7%
6	19990	6.2%
3	14097	4.3%
Other values (8)	11765	3.6%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	324359	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
4	51155	15.8%
2	49765	15.3%
5	38176	11.8%
0	34046	10.5%
8	28711	8.9%
7	28654	8.8%
1	26158	8.1%
9	21842	6.7%
6	19990	6.2%
3	14097	4.3%
Other values (8)	11765	3.6%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	324359	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
4	51155	15.8%
2	49765	15.3%
5	38176	11.8%
0	34046	10.5%
8	28711	8.9%
7	28654	8.8%
1	26158	8.1%
9	21842	6.7%
6	19990	6.2%
3	14097	4.3%
Other values (8)	11765	3.6%

diag_3

Categorical

HIGH CARDINALITY

Distinct	790
Distinct (%)	0.8%
Missing	0
Missing (%)	0.0%
Memory size	277.6 KiB

Length		Characters and Unicode	Unique		Sample	
			Unique	122 ?	1st row	Missing
Max length	7	Total characters	325199	Unique (%) 0.1%	2nd row	255
Median length	3	Distinct characters	18		3rd row	V27
Mean length	3.1955565	Distinct categories	1		4th row	403
Min length	1	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)			5th row	250 ?
		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)		?

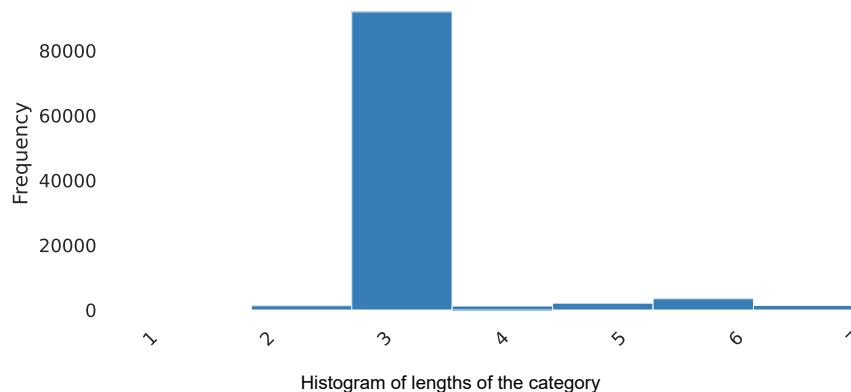
The Unicode Standard assigns character properties to each code point, which can be used to analyse textual variables.

Common Values

Value	Count	Frequency (%)
250	11555	11.4%
401	8289	8.1%
276	5175	5.1%
428	4577	4.5%
427	3955	3.9%
414	3664	3.6%

Value	Count	Frequency (%)
496	2605	2.6%
403	2357	2.3%
585	1992	2.0%
272	1969	1.9%
Other values (780)	55628	54.7%

Length



Value	Count	Frequency (%)
250	11555	11.4%
401	8289	8.1%
276	5175	5.1%
428	4577	4.5%
427	3955	3.9%
414	3664	3.6%
496	2605	2.6%
403	2357	2.3%
585	1992	2.0%
272	1969	1.9%
Other values (780)	55628	54.7%

Most occurring characters

Value	Count	Frequency (%)
2	51244	15.8%
4	49252	15.1%
5	41260	12.7%
0	39711	12.2%
7	26504	8.2%
1	24684	7.6%
8	23825	7.3%
9	17323	5.3%
6	16441	5.1%
3	14333	4.4%
Other values (8)	20622	6.3%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	325199	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
2	51244	15.8%
4	49252	15.1%
5	41260	12.7%
0	39711	12.2%
7	26504	8.2%
1	24684	7.6%
8	23825	7.3%

Value	Count	Frequency (%)
9	17323	5.3%
6	16441	5.1%
3	14333	4.4%
Other values (8)	20622	6.3%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	325199	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
2	51244	15.8%
4	49252	15.1%
5	41260	12.7%
0	39711	12.2%
7	26504	8.2%
1	24684	7.6%
8	23825	7.3%
9	17323	5.3%
6	16441	5.1%
3	14333	4.4%
Other values (8)	20622	6.3%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	325199	100.0%

Most frequent character per block

(unknown)

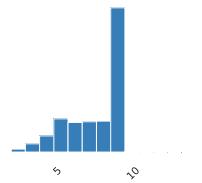
Value	Count	Frequency (%)
2	51244	15.8%
4	49252	15.1%
5	41260	12.7%
0	39711	12.2%
7	26504	8.2%
1	24684	7.6%
8	23825	7.3%
9	17323	5.3%
6	16441	5.1%
3	14333	4.4%
Other values (8)	20622	6.3%

number_diagnoses

Real number (\mathbb{R})

Distinct	16
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	7.4226068

Minimum	1
Maximum	16
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	795.2 KiB

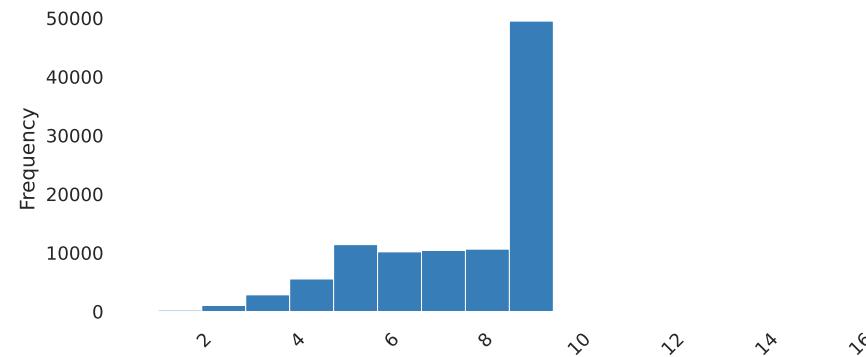


Quantile statistics

Minimum	1
5-th percentile	4
Q1	6
median	8
Q3	9
95-th percentile	9
Maximum	16
Range	15
Interquartile range (IQR)	3

Descriptive statistics

Standard deviation	1.9336001
Coefficient of variation (CV)	0.26050149
Kurtosis	-0.079056024
Mean	7.4226068
Median Absolute Deviation (MAD)	1
Skewness	-0.87674624
Sum	755369
Variance	3.7388095
Monotonicity	Not monotonic



Histogram with fixed size bins (bins=16)

Value	Count	Frequency (%)
9	49474	48.6%
5	11393	11.2%
8	10616	10.4%
7	10393	10.2%
6	10161	10.0%
4	5537	5.4%
3	2835	2.8%
2	1023	1.0%
1	219	0.2%
16	45	< 0.1%
Other values (6)	70	0.1%

Value	Count	Frequency (%)
1	219	0.2%
2	1023	1.0%
3	2835	2.8%
4	5537	5.4%
5	11393	11.2%
6	10161	10.0%
7	10393	10.2%
8	10616	10.4%
9	49474	48.6%
10	17	< 0.1%

Value	Count	Frequency (%)
16	45	< 0.1%
15	10	< 0.1%
14	7	< 0.1%

Value	Count	Frequency (%)
13	16	< 0.1%
12	9	< 0.1%
11	11	< 0.1%
10	17	< 0.1%
9	49474	48.6%
8	10616	10.4%
7	10393	10.2%

max_glu_serum

Categorical

IMBALANCE

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.9 KiB

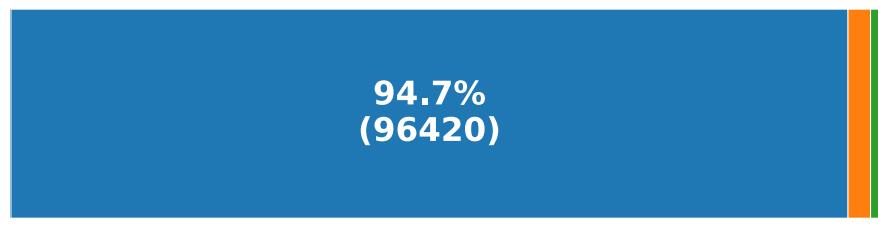
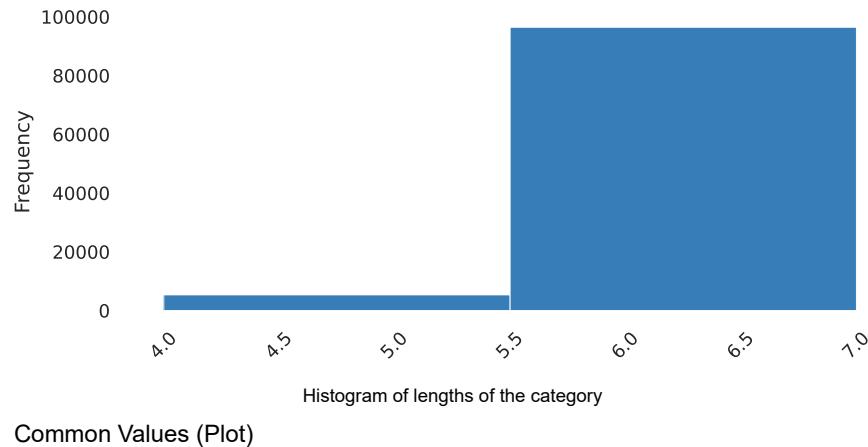
Length		Characters and Unicode	Unique		Sample	
			Unique	0 ?	1st row	No Test
Max length	7	Total characters	696324	Unique (%) 0.0%	2nd row	No Test
Median length	7	Distinct characters	13		3rd row	No Test
Mean length	6.8424032	Distinct categories	1		4th row	No Test
Min length	4	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)			5th row	No Test ?
		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.

Common Values

Value	Count	Frequency (%)
No Test	96420	94.7%
Norm	2597	2.6%
>200	1485	1.5%
>300	1264	1.2%

Length



Value	Count	Frequency (%)
no	96420	48.7%
test	96420	48.7%
norm	2597	1.3%
200	1485	0.7%
300	1264	0.6%

Most occurring characters

Value	Count	Frequency (%)
N	99017	14.2%
o	99017	14.2%
	96420	13.8%
T	96420	13.8%
e	96420	13.8%
s	96420	13.8%
t	96420	13.8%
0	5498	0.8%
>	2749	0.4%
r	2597	0.4%
Other values (3)	5346	0.8%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	696324	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
N	99017	14.2%
o	99017	14.2%
	96420	13.8%
T	96420	13.8%
e	96420	13.8%
s	96420	13.8%
t	96420	13.8%
0	5498	0.8%

Value	Count	Frequency (%)
>	2749	0.4%
r	2597	0.4%
Other values (3)	5346	0.8%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	696324	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
N	99017	14.2%
o	99017	14.2%
	96420	13.8%
T	96420	13.8%
e	96420	13.8%
s	96420	13.8%
t	96420	13.8%
0	5498	0.8%
>	2749	0.4%
r	2597	0.4%
Other values (3)	5346	0.8%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	696324	100.0%

Most frequent character per block

(*unknown*)

Value	Count	Frequency (%)
N	99017	14.2%
o	99017	14.2%
	96420	13.8%
T	96420	13.8%
e	96420	13.8%
s	96420	13.8%
t	96420	13.8%
0	5498	0.8%
>	2749	0.4%
r	2597	0.4%
Other values (3)	5346	0.8%

A1Cresult

Categorical

IMBALANCE

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.9 KiB

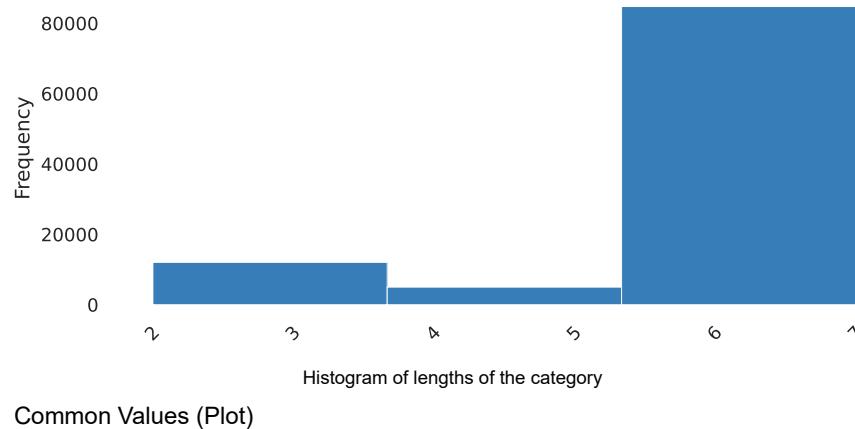
Length		Characters and Unicode	Unique		Sample	
Max length	7		Unique	0 ?	1st row	No Test
Median length	7	Total characters 637252	Unique (%)	0.0%	2nd row	No Test
Mean length	6.2619342	Distinct characters 12			3rd row	No Test
Min length	2	Distinct categories 1			4th row	No Test
		(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)			5th row	No Test ?
		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.

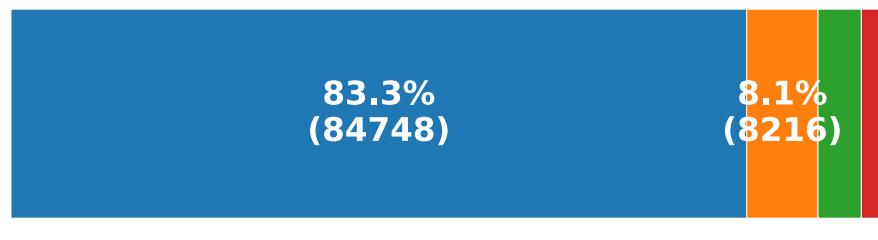
Common Values

Value	Count	Frequency (%)
No Test	84748	83.3%
>8	8216	8.1%
Norm	4990	4.9%
>7	3812	3.7%

Length



Common Values (Plot)



Value	Count	Frequency (%)
no	84748	45.4%
test	84748	45.4%
8	8216	4.4%
norm	4990	2.7%
7	3812	2.0%

Most occurring characters

Value	Count	Frequency (%)
N	89738	14.1%
o	89738	14.1%
	84748	13.3%
T	84748	13.3%
e	84748	13.3%
s	84748	13.3%
t	84748	13.3%
>	12028	1.9%
8	8216	1.3%
r	4990	0.8%
Other values (2)	8802	1.4%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	637252	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
N	89738	14.1%
o	89738	14.1%
	84748	13.3%
T	84748	13.3%
e	84748	13.3%
s	84748	13.3%
t	84748	13.3%
>	12028	1.9%

Value	Count	Frequency (%)
8	8216	1.3%
r	4990	0.8%
Other values (2)	8802	1.4%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	637252	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
N	89738	14.1%
o	89738	14.1%
	84748	13.3%
T	84748	13.3%
e	84748	13.3%
s	84748	13.3%
t	84748	13.3%
>	12028	1.9%
8	8216	1.3%
r	4990	0.8%
Other values (2)	8802	1.4%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	637252	100.0%

Most frequent character per block

(*unknown*)

Value	Count	Frequency (%)
N	89738	14.1%
o	89738	14.1%
	84748	13.3%
T	84748	13.3%
e	84748	13.3%
s	84748	13.3%
t	84748	13.3%
>	12028	1.9%
8	8216	1.3%
r	4990	0.8%
Other values (2)	8802	1.4%

metformin

Categorical

IMBALANCE

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.9 KiB

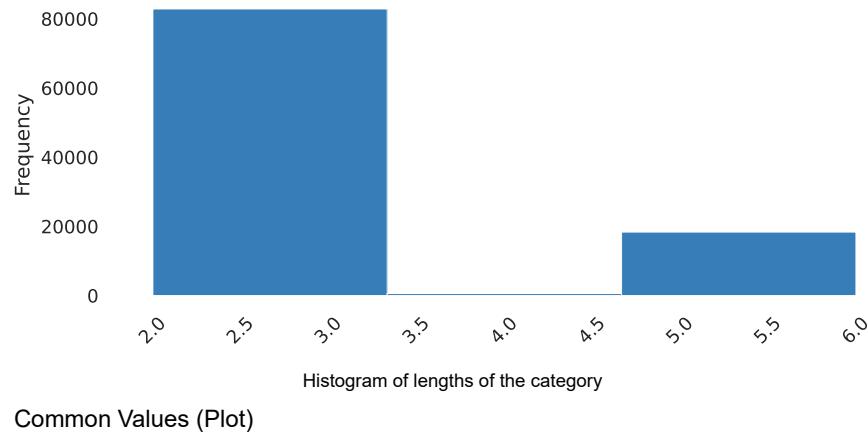
Length		Characters and Unicode	Unique	Sample		
Max length	6		Unique	0	?	1st row
Median length	2	Total characters	Unique (%)	278066	0.0%	2nd row
Mean length	2.7324057	Distinct characters		13		3rd row
Min length	2	Distinct categories		1		4th row
		(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				
		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.

Common Values

Value	Count	Frequency (%)
No	81778	80.4%
Steady	18346	18.0%
Up	1067	1.0%
Down	575	0.6%

Length



Value	Count	Frequency (%)
no	81778	80.4%
steady	18346	18.0%
up	1067	1.0%
down	575	0.6%

Most occurring characters

Value	Count	Frequency (%)
o	82353	29.6%
N	81778	29.4%
S	18346	6.6%
t	18346	6.6%
e	18346	6.6%
a	18346	6.6%
d	18346	6.6%
y	18346	6.6%
U	1067	0.4%
p	1067	0.4%
Other values (3)	1725	0.6%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	278066	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
o	82353	29.6%
N	81778	29.4%
S	18346	6.6%
t	18346	6.6%
e	18346	6.6%
a	18346	6.6%
d	18346	6.6%
y	18346	6.6%

Value	Count	Frequency (%)
U	1067	0.4%
p	1067	0.4%
Other values (3)	1725	0.6%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	278066	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
o	82353	29.6%
N	81778	29.4%
S	18346	6.6%
t	18346	6.6%
e	18346	6.6%
a	18346	6.6%
d	18346	6.6%
y	18346	6.6%
U	1067	0.4%
p	1067	0.4%
Other values (3)	1725	0.6%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	278066	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
o	82353	29.6%
N	81778	29.4%
S	18346	6.6%
t	18346	6.6%
e	18346	6.6%
a	18346	6.6%
d	18346	6.6%
y	18346	6.6%
U	1067	0.4%
p	1067	0.4%
Other values (3)	1725	0.6%

repaglinide

Categorical

IMBALANCE

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.9 KiB

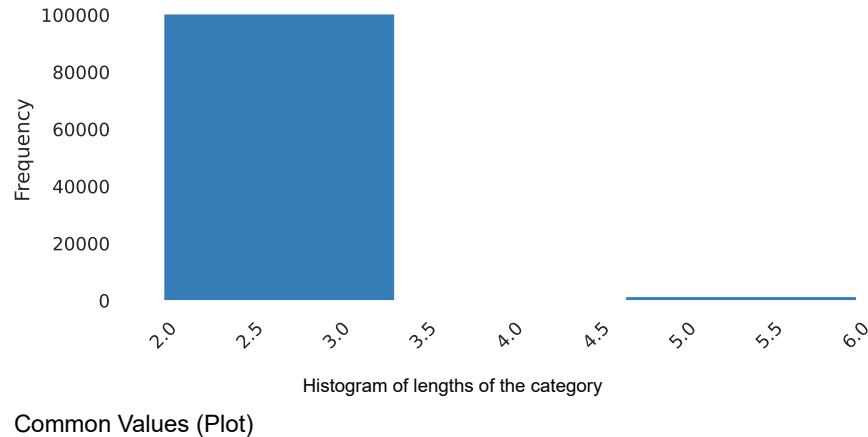
Length		Characters and Unicode	Unique	Sample		
Max length	6		Unique	0	?	1st row
Median length	2	Total characters	Unique (%)	209158	0.0%	2nd row
Mean length	2.0552837	Distinct characters		13		3rd row
Min length	2	Distinct categories		1		4th row
		(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				
		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.

Common Values

Value	Count	Frequency (%)
No	100227	98.5%
Steady	1384	1.4%
Up	110	0.1%
Down	45	< 0.1%

Length



Common Values (Plot)



Value	Count	Frequency (%)
no	100227	98.5%
steady	1384	1.4%
up	110	0.1%
down	45	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
o	100272	47.9%
N	100227	47.9%
S	1384	0.7%
t	1384	0.7%
e	1384	0.7%
a	1384	0.7%
d	1384	0.7%
y	1384	0.7%
U	110	0.1%
p	110	0.1%
Other values (3)	135	0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	209158	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
o	100272	47.9%
N	100227	47.9%
S	1384	0.7%
t	1384	0.7%
e	1384	0.7%
a	1384	0.7%
d	1384	0.7%
y	1384	0.7%

Value	Count	Frequency (%)
U	110	0.1%
p	110	0.1%
Other values (3)	135	0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	209158	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
o	100272	47.9%
N	100227	47.9%
S	1384	0.7%
t	1384	0.7%
e	1384	0.7%
a	1384	0.7%
d	1384	0.7%
y	1384	0.7%
U	110	0.1%
p	110	0.1%
Other values (3)	135	0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	209158	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
o	100272	47.9%
N	100227	47.9%
S	1384	0.7%
t	1384	0.7%
e	1384	0.7%
a	1384	0.7%
d	1384	0.7%
y	1384	0.7%
U	110	0.1%
p	110	0.1%
Other values (3)	135	0.1%

nateglinide

Categorical

IMBALANCE

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.9 KiB

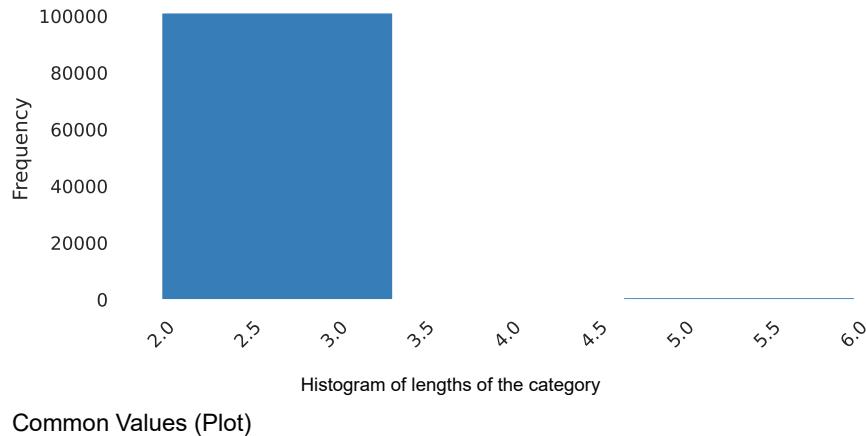
Length		Characters and Unicode	Unique	Sample		
Max length	6		Unique	0	?	1st row
Median length	2	Total characters	Unique (%)	206226	0.0%	2nd row
Mean length	2.0264725	Distinct characters		13		3rd row
Min length	2	Distinct categories		1		4th row
		(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				
		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.

Common Values

Value	Count	Frequency (%)
No	101063	99.3%
Steady	668	0.7%
Up	24	< 0.1%
Down	11	< 0.1%

Length



Common Values (Plot)

99.3%
(101063)



Value	Count	Frequency (%)
no	101063	99.3%
steady	668	0.7%
up	24	< 0.1%
down	11	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
o	101074	49.0%
N	101063	49.0%
S	668	0.3%
t	668	0.3%
e	668	0.3%
a	668	0.3%
d	668	0.3%
y	668	0.3%
U	24	< 0.1%
p	24	< 0.1%
Other values (3)	33	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	206226	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
o	101074	49.0%
N	101063	49.0%
S	668	0.3%
t	668	0.3%
e	668	0.3%
a	668	0.3%
d	668	0.3%
y	668	0.3%

Value	Count	Frequency (%)
U	24	< 0.1%
p	24	< 0.1%
Other values (3)	33	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	206226	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
o	101074	49.0%
N	101063	49.0%
S	668	0.3%
t	668	0.3%
e	668	0.3%
a	668	0.3%
d	668	0.3%
y	668	0.3%
U	24	< 0.1%
p	24	< 0.1%
Other values (3)	33	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	206226	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
o	101074	49.0%
N	101063	49.0%
S	668	0.3%
t	668	0.3%
e	668	0.3%
a	668	0.3%
d	668	0.3%
y	668	0.3%
U	24	< 0.1%
p	24	< 0.1%
Other values (3)	33	< 0.1%

chlorpropamide

Categorical

IMBALANCE

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.9 KiB

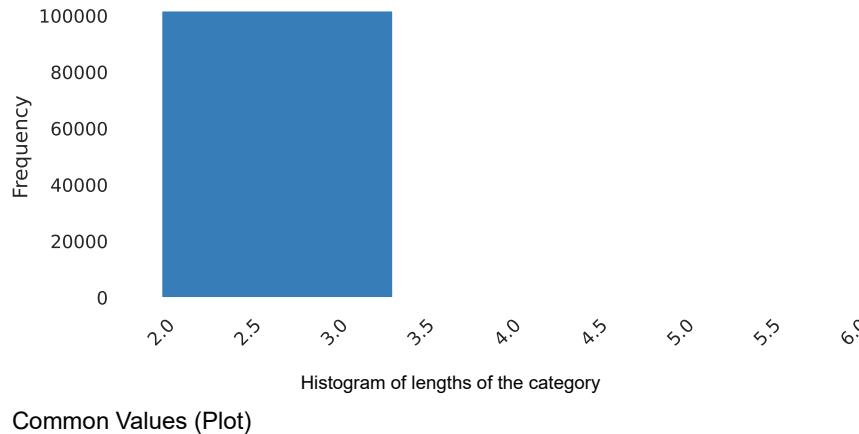
Length		Characters and Unicode	Unique			Sample	
			Unique	1	?	1st row	No
Max length	6	Total characters	203850	Unique (%)	< 0.1%	2nd row	No
Median length	2	Distinct characters	13			3rd row	No
Mean length	2.0031248	Distinct categories	1			4th row	No
Min length	2	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				5th row	?
		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.

Common Values

Value	Count	Frequency (%)
No	101680	99.9%
Steady	79	0.1%
Up	6	< 0.1%
Down	1	< 0.1%

Length



99.9%
(101680)



Value	Count	Frequency (%)
no	101680	99.9%
steady	79	0.1%
up	6	< 0.1%
down	1	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
o	101681	49.9%
N	101680	49.9%
S	79	< 0.1%
t	79	< 0.1%
e	79	< 0.1%
a	79	< 0.1%
d	79	< 0.1%
y	79	< 0.1%
U	6	< 0.1%
p	6	< 0.1%
Other values (3)	3	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	203850	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
o	101681	49.9%
N	101680	49.9%
S	79	< 0.1%
t	79	< 0.1%
e	79	< 0.1%
a	79	< 0.1%
d	79	< 0.1%
y	79	< 0.1%

Value	Count	Frequency (%)
U	6	< 0.1%
p	6	< 0.1%
Other values (3)	3	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	203850	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
o	101681	49.9%
N	101680	49.9%
S	79	< 0.1%
t	79	< 0.1%
e	79	< 0.1%
a	79	< 0.1%
d	79	< 0.1%
y	79	< 0.1%
U	6	< 0.1%
p	6	< 0.1%
Other values (3)	3	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	203850	100.0%

Most frequent character per block

(*unknown*)

Value	Count	Frequency (%)
o	101681	49.9%
N	101680	49.9%
S	79	< 0.1%
t	79	< 0.1%
e	79	< 0.1%
a	79	< 0.1%
d	79	< 0.1%
y	79	< 0.1%
U	6	< 0.1%
p	6	< 0.1%
Other values (3)	3	< 0.1%

glimepiride

Categorical

IMBALANCE

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.9 KiB

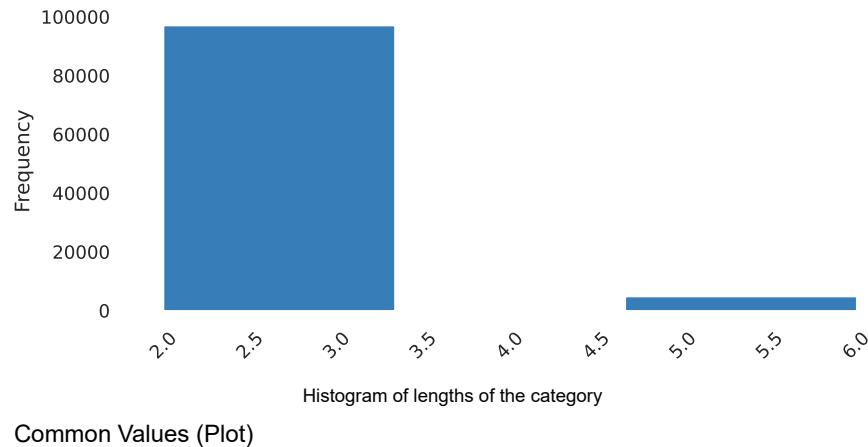
Length		Characters and Unicode	Unique	Sample		
Max length	6		Unique	0	?	1st row
Total characters	222600		Unique (%)	0.0%		2nd row
Median length	2					3rd row
Mean length	2.187371					4th row
Min length	2					5th row
		categories (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				?
		scripts (https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)				?
		blocks (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.

Common Values

Value	Count	Frequency (%)
No	96575	94.9%
Steady	4670	4.6%
Up	327	0.3%
Down	194	0.2%

Length



Value	Count	Frequency (%)
no	96575	94.9%
steady	4670	4.6%
up	327	0.3%
down	194	0.2%

Most occurring characters

Value	Count	Frequency (%)
o	96769	43.5%
N	96575	43.4%
S	4670	2.1%
t	4670	2.1%
e	4670	2.1%
a	4670	2.1%
d	4670	2.1%
y	4670	2.1%
U	327	0.1%
p	327	0.1%
Other values (3)	582	0.3%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	222600	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
o	96769	43.5%
N	96575	43.4%
S	4670	2.1%
t	4670	2.1%
e	4670	2.1%
a	4670	2.1%
d	4670	2.1%
y	4670	2.1%

Value	Count	Frequency (%)
U	327	0.1%
p	327	0.1%
Other values (3)	582	0.3%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	222600	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
o	96769	43.5%
N	96575	43.4%
S	4670	2.1%
t	4670	2.1%
e	4670	2.1%
a	4670	2.1%
d	4670	2.1%
y	4670	2.1%
U	327	0.1%
p	327	0.1%
Other values (3)	582	0.3%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	222600	100.0%

Most frequent character per block

(*unknown*)

Value	Count	Frequency (%)
o	96769	43.5%
N	96575	43.4%
S	4670	2.1%
t	4670	2.1%
e	4670	2.1%
a	4670	2.1%
d	4670	2.1%
y	4670	2.1%
U	327	0.1%
p	327	0.1%
Other values (3)	582	0.3%

acetohexamide

Categorical

IMBALANCE

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.7 KiB

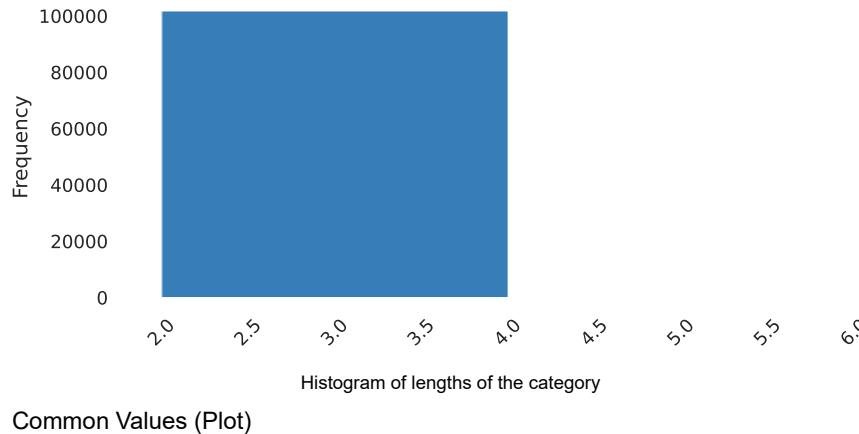
Length		Characters and Unicode	Unique	Sample		
Max length	6		Unique	1	?	1st row
Median length	2	Total characters	Unique (%)	< 0.1%		2nd row
Mean length	2.0000393	Distinct characters				3rd row
Min length	2	Distinct categories				4th row
		(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				5th row
		Distinct scripts				(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
		Distinct blocks				(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.

Common Values

Value	Count	Frequency (%)
No	101765	> 99.9%
Steady	1	< 0.1%

Length



Common Values (Plot)



Value	Count	Frequency (%)
no	101765	> 99.9%
steady	1	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
N	101765	50.0%
o	101765	50.0%
S	1	< 0.1%
t	1	< 0.1%

Value	Count	Frequency (%)
e	1	< 0.1%
a	1	< 0.1%
d	1	< 0.1%
y	1	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	203536	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
N	101765	50.0%
o	101765	50.0%
S	1	< 0.1%
t	1	< 0.1%
e	1	< 0.1%
a	1	< 0.1%
d	1	< 0.1%
y	1	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	203536	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
N	101765	50.0%
o	101765	50.0%
S	1	< 0.1%
t	1	< 0.1%
e	1	< 0.1%
a	1	< 0.1%
d	1	< 0.1%
y	1	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	203536	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
N	101765	50.0%
o	101765	50.0%
S	1	< 0.1%
t	1	< 0.1%
e	1	< 0.1%
a	1	< 0.1%
d	1	< 0.1%
y	1	< 0.1%

glipizide

Categorical

IMBALANCE

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.9 KiB

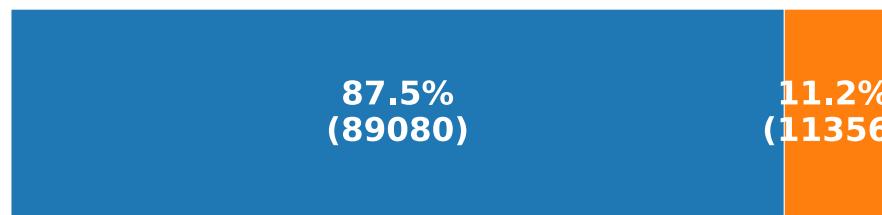
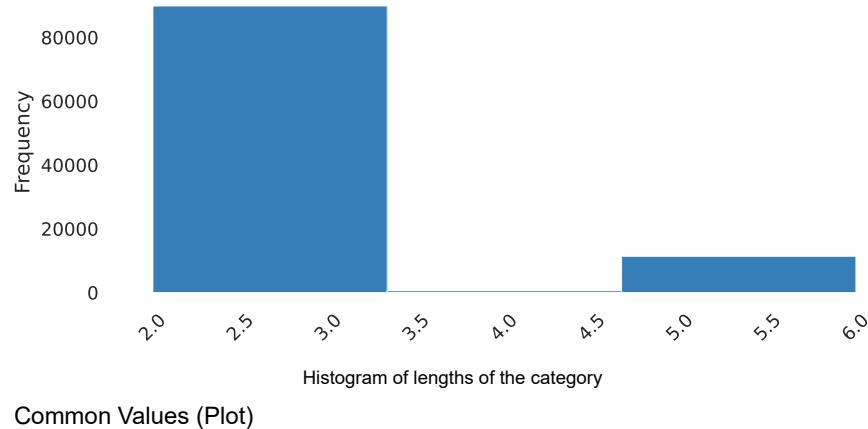
Length		Characters and Unicode	Unique		Sample	
Max length	6		Unique	0 ?	1st row	No
Median length	2	Total characters 250076	Unique (%) 0.0%		2nd row	No
Mean length	2.457363	Distinct characters 13			3rd row	Steady
Min length	2	Distinct categories 1			4th row	No
		(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)			5th row	Steady ?
		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.

Common Values

Value	Count	Frequency (%)
No	89080	87.5%
Steady	11356	11.2%
Up	770	0.8%
Down	560	0.6%

Length



Value	Count	Frequency (%)
no	89080	87.5%
steady	11356	11.2%
up	770	0.8%
down	560	0.6%

Most occurring characters

Value	Count	Frequency (%)
o	89640	35.8%
N	89080	35.6%
S	11356	4.5%
t	11356	4.5%
e	11356	4.5%
a	11356	4.5%
d	11356	4.5%
y	11356	4.5%
U	770	0.3%
p	770	0.3%
Other values (3)	1680	0.7%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	250076	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
o	89640	35.8%
N	89080	35.6%
S	11356	4.5%
t	11356	4.5%
e	11356	4.5%
a	11356	4.5%
d	11356	4.5%
y	11356	4.5%

Value	Count	Frequency (%)
U	770	0.3%
p	770	0.3%
Other values (3)	1680	0.7%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	250076	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
o	89640	35.8%
N	89080	35.6%
S	11356	4.5%
t	11356	4.5%
e	11356	4.5%
a	11356	4.5%
d	11356	4.5%
y	11356	4.5%
U	770	0.3%
p	770	0.3%
Other values (3)	1680	0.7%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	250076	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
o	89640	35.8%
N	89080	35.6%
S	11356	4.5%
t	11356	4.5%
e	11356	4.5%
a	11356	4.5%
d	11356	4.5%
y	11356	4.5%
U	770	0.3%
p	770	0.3%
Other values (3)	1680	0.7%

glyburide

Categorical

IMBALANCE

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.9 KiB

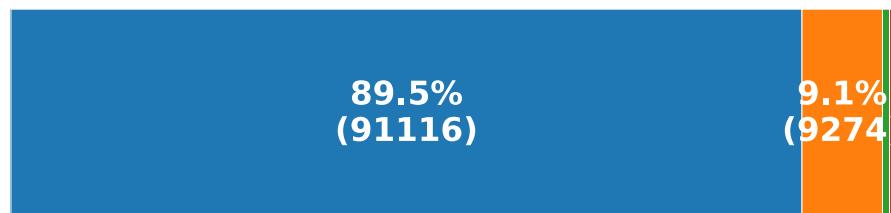
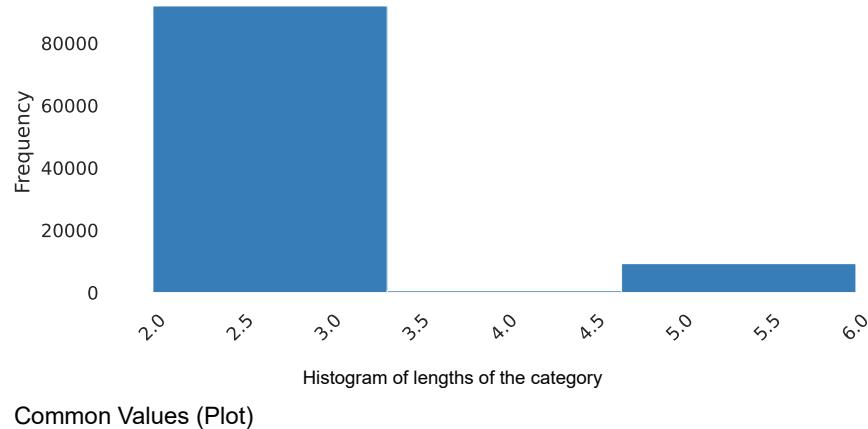
Length		Characters and Unicode	Unique	Sample		
Max length	6		Unique	0	?	1st row
Median length	2	Total characters	Unique (%)	241756	0.0%	2nd row
Mean length	2.3756068	Distinct characters		13		3rd row
Min length	2	Distinct categories		1		4th row
		(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				
		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.

Common Values

Value	Count	Frequency (%)
No	91116	89.5%
Steady	9274	9.1%
Up	812	0.8%
Down	564	0.6%

Length



Value	Count	Frequency (%)
no	91116	89.5%
steady	9274	9.1%
up	812	0.8%
down	564	0.6%

Most occurring characters

Value	Count	Frequency (%)
o	91680	37.9%
N	91116	37.7%
S	9274	3.8%
t	9274	3.8%
e	9274	3.8%
a	9274	3.8%
d	9274	3.8%
y	9274	3.8%
U	812	0.3%
p	812	0.3%
Other values (3)	1692	0.7%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	241756	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
o	91680	37.9%
N	91116	37.7%
S	9274	3.8%
t	9274	3.8%
e	9274	3.8%
a	9274	3.8%
d	9274	3.8%
y	9274	3.8%

Value	Count	Frequency (%)
U	812	0.3%
p	812	0.3%
Other values (3)	1692	0.7%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	241756	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
o	91680	37.9%
N	91116	37.7%
S	9274	3.8%
t	9274	3.8%
e	9274	3.8%
a	9274	3.8%
d	9274	3.8%
y	9274	3.8%
U	812	0.3%
p	812	0.3%
Other values (3)	1692	0.7%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	241756	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
o	91680	37.9%
N	91116	37.7%
S	9274	3.8%
t	9274	3.8%
e	9274	3.8%
a	9274	3.8%
d	9274	3.8%
y	9274	3.8%
U	812	0.3%
p	812	0.3%
Other values (3)	1692	0.7%

tolbutamide

Categorical

IMBALANCE

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.7 KiB

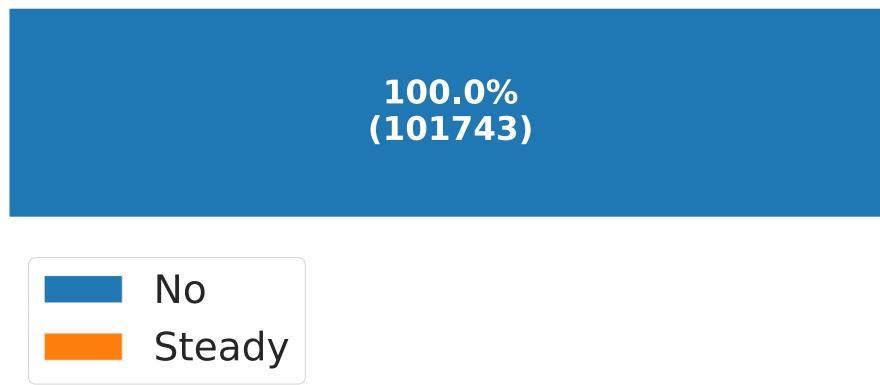
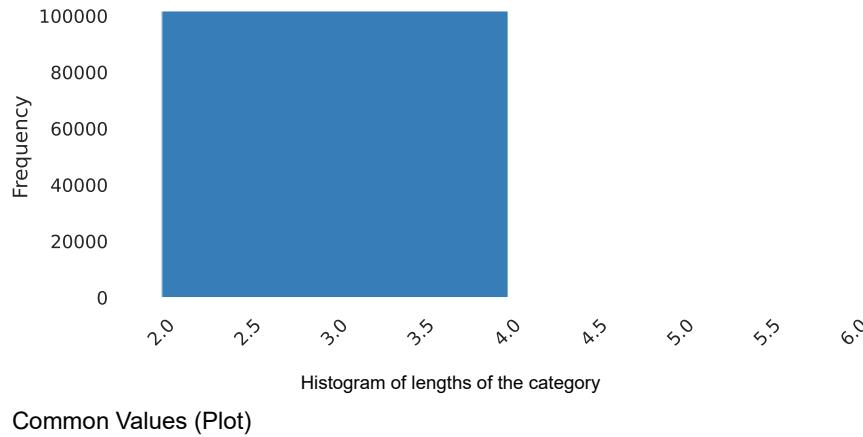
Length		Characters and Unicode	Unique	Sample		
Max length	6		Unique	0	?	1st row
			Total characters	203624	0.0%	2nd row
Median length	2					3rd row
Mean length	2.000904	Distinct characters	8			4th row
Min length	2	Distinct categories	1			5th row
			(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)			?
		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.

Common Values

Value	Count	Frequency (%)
No	101743	> 99.9%
Steady	23	< 0.1%

Length



Value	Count	Frequency (%)
no	101743	> 99.9%
steady	23	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
N	101743	50.0%
o	101743	50.0%
S	23	< 0.1%
t	23	< 0.1%

Value	Count	Frequency (%)
e	23	< 0.1%
a	23	< 0.1%
d	23	< 0.1%
y	23	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	203624	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
N	101743	50.0%
o	101743	50.0%
S	23	< 0.1%
t	23	< 0.1%
e	23	< 0.1%
a	23	< 0.1%
d	23	< 0.1%
y	23	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	203624	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
N	101743	50.0%
o	101743	50.0%
S	23	< 0.1%
t	23	< 0.1%
e	23	< 0.1%
a	23	< 0.1%
d	23	< 0.1%
y	23	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	203624	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
N	101743	50.0%
o	101743	50.0%
S	23	< 0.1%
t	23	< 0.1%
e	23	< 0.1%
a	23	< 0.1%
d	23	< 0.1%
y	23	< 0.1%

pioglitazone

Categorical

IMBALANCE

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.9 KiB

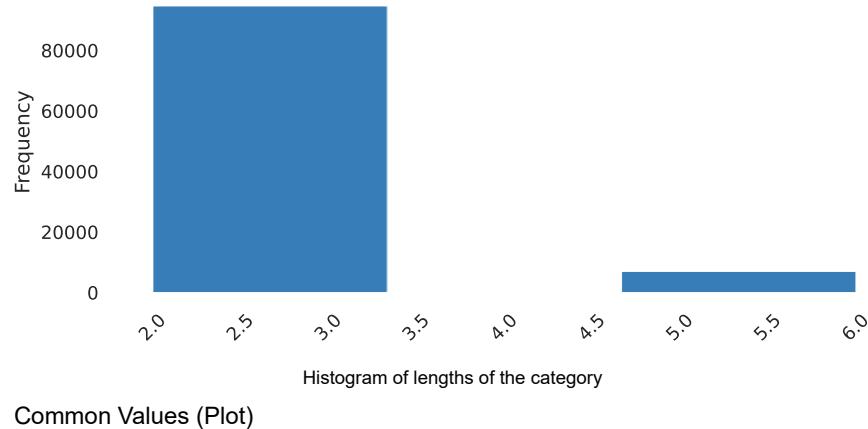
Length		Characters and Unicode	Unique	Sample		
Max length	6		Unique	0	?	1st row
Median length	2	Total characters	Unique (%)	231672	0.0%	2nd row
Mean length	2.2765167	Distinct characters		13		3rd row
Min length	2	Distinct categories		1		4th row
		(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				
		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.

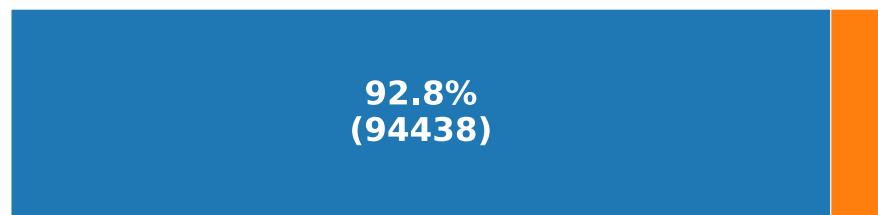
Common Values

Value	Count	Frequency (%)
No	94438	92.8%
Steady	6976	6.9%
Up	234	0.2%
Down	118	0.1%

Length



Common Values (Plot)



Value	Count	Frequency (%)
no	94438	92.8%
steady	6976	6.9%
up	234	0.2%
down	118	0.1%

Most occurring characters

Value	Count	Frequency (%)
o	94556	40.8%
N	94438	40.8%
S	6976	3.0%
t	6976	3.0%
e	6976	3.0%
a	6976	3.0%
d	6976	3.0%
y	6976	3.0%
U	234	0.1%
p	234	0.1%
Other values (3)	354	0.2%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	231672	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
o	94556	40.8%
N	94438	40.8%
S	6976	3.0%
t	6976	3.0%
e	6976	3.0%
a	6976	3.0%
d	6976	3.0%
y	6976	3.0%

Value	Count	Frequency (%)
U	234	0.1%
p	234	0.1%
Other values (3)	354	0.2%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	231672	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
o	94556	40.8%
N	94438	40.8%
S	6976	3.0%
t	6976	3.0%
e	6976	3.0%
a	6976	3.0%
d	6976	3.0%
y	6976	3.0%
U	234	0.1%
p	234	0.1%
Other values (3)	354	0.2%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	231672	100.0%

Most frequent character per block

(*unknown*)

Value	Count	Frequency (%)
o	94556	40.8%
N	94438	40.8%
S	6976	3.0%
t	6976	3.0%
e	6976	3.0%
a	6976	3.0%
d	6976	3.0%
y	6976	3.0%
U	234	0.1%
p	234	0.1%
Other values (3)	354	0.2%

rosiglitazone

Categorical

IMBALANCE

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.9 KiB

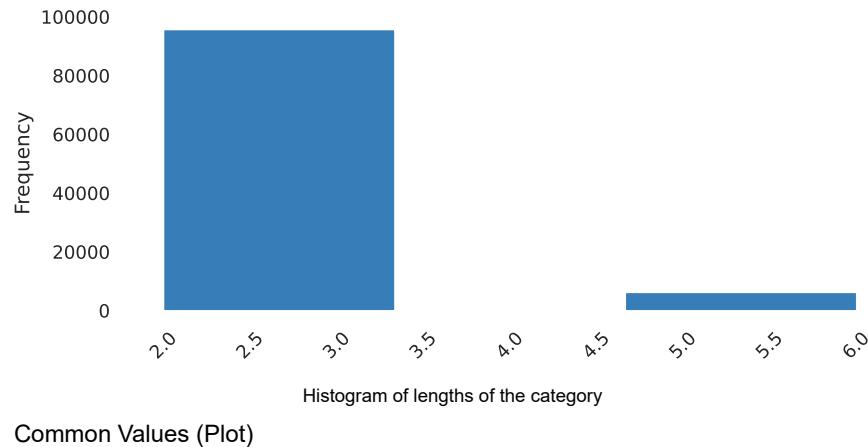
Length		Characters and Unicode	Unique		Sample		
			Unique	0 ?	1st row	No	
Max length	6	Total characters	228106	Unique (%)	0.0%	2nd row	No
Median length	2	Distinct characters	13			3rd row	No
Mean length	2.2414755	Distinct categories	1			4th row	No
Min length	2	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				5th row	?
		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.

Common Values

Value	Count	Frequency (%)
No	95401	93.7%
Steady	6100	6.0%
Up	178	0.2%
Down	87	0.1%

Length



Value	Count	Frequency (%)
no	95401	93.7%
steady	6100	6.0%
up	178	0.2%
down	87	0.1%

Most occurring characters

Value	Count	Frequency (%)
o	95488	41.9%
N	95401	41.8%
S	6100	2.7%
t	6100	2.7%
e	6100	2.7%
a	6100	2.7%
d	6100	2.7%
y	6100	2.7%
U	178	0.1%
p	178	0.1%
Other values (3)	261	0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	228106	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
o	95488	41.9%
N	95401	41.8%
S	6100	2.7%
t	6100	2.7%
e	6100	2.7%
a	6100	2.7%
d	6100	2.7%
y	6100	2.7%

Value	Count	Frequency (%)
U	178	0.1%
p	178	0.1%
Other values (3)	261	0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	228106	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
o	95488	41.9%
N	95401	41.8%
S	6100	2.7%
t	6100	2.7%
e	6100	2.7%
a	6100	2.7%
d	6100	2.7%
y	6100	2.7%
U	178	0.1%
p	178	0.1%
Other values (3)	261	0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	228106	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
o	95488	41.9%
N	95401	41.8%
S	6100	2.7%
t	6100	2.7%
e	6100	2.7%
a	6100	2.7%
d	6100	2.7%
y	6100	2.7%
U	178	0.1%
p	178	0.1%
Other values (3)	261	0.1%

acarbose

Categorical

IMBALANCE

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.9 KiB

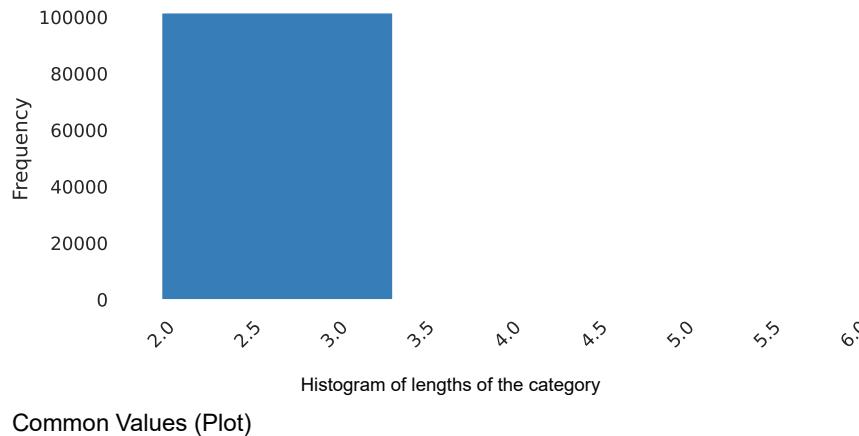
Length		Characters and Unicode	Unique	Sample		
Max length	6		Unique	0	?	1st row
Median length	2		Total characters	204718	0.0%	2nd row
Mean length	2.0116542		Distinct characters	13		3rd row
Min length	2		Distinct categories	1		4th row
			(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)			
			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.

Common Values

Value	Count	Frequency (%)
No	101458	99.7%
Steady	295	0.3%
Up	10	< 0.1%
Down	3	< 0.1%

Length



99.7%
(101458)



Value	Count	Frequency (%)
no	101458	99.7%
steady	295	0.3%
up	10	< 0.1%
down	3	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
o	101461	49.6%
N	101458	49.6%
S	295	0.1%
t	295	0.1%
e	295	0.1%
a	295	0.1%
d	295	0.1%
y	295	0.1%
U	10	< 0.1%
p	10	< 0.1%
Other values (3)	9	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	204718	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
o	101461	49.6%
N	101458	49.6%
S	295	0.1%
t	295	0.1%
e	295	0.1%
a	295	0.1%
d	295	0.1%
y	295	0.1%

Value	Count	Frequency (%)
U	10	< 0.1%
p	10	< 0.1%
Other values (3)	9	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	204718	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
o	101461	49.6%
N	101458	49.6%
S	295	0.1%
t	295	0.1%
e	295	0.1%
a	295	0.1%
d	295	0.1%
y	295	0.1%
U	10	< 0.1%
p	10	< 0.1%
Other values (3)	9	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	204718	100.0%

Most frequent character per block

(*unknown*)

Value	Count	Frequency (%)
o	101461	49.6%
N	101458	49.6%
S	295	0.1%
t	295	0.1%
e	295	0.1%
a	295	0.1%
d	295	0.1%
y	295	0.1%
U	10	< 0.1%
p	10	< 0.1%
Other values (3)	9	< 0.1%

miglitol

Categorical

IMBALANCE

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.9 KiB

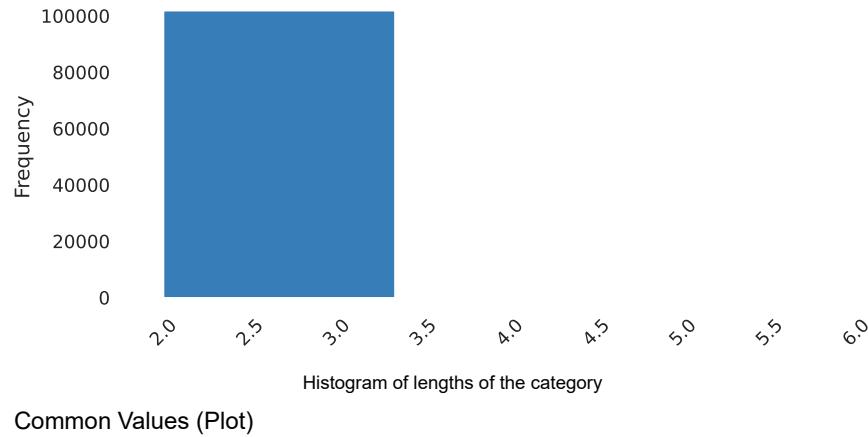
Length		Characters and Unicode	Unique		Sample	
Max length	6		Unique	0 ?	1st row	No
Total characters	203666		Unique (%)	0.0%	2nd row	No
Distinct characters	13				3rd row	No
Distinct categories	1				4th row	No
Distinct scripts	1	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				?
Distinct blocks	1	(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)				?

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

Common Values

Value	Count	Frequency (%)
No	101728	> 99.9%
Steady	31	< 0.1%
Down	5	< 0.1%
Up	2	< 0.1%

Length



100.0%
(101728)



Value	Count	Frequency (%)
no	101728	> 99.9%
steady	31	< 0.1%
down	5	< 0.1%
up	2	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
o	101733	50.0%
N	101728	49.9%
S	31	< 0.1%
t	31	< 0.1%
e	31	< 0.1%
a	31	< 0.1%
d	31	< 0.1%
y	31	< 0.1%
D	5	< 0.1%
w	5	< 0.1%
Other values (3)	9	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	203666	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
o	101733	50.0%
N	101728	49.9%
S	31	< 0.1%
t	31	< 0.1%
e	31	< 0.1%
a	31	< 0.1%
d	31	< 0.1%
y	31	< 0.1%

Value	Count	Frequency (%)
D	5	< 0.1%
w	5	< 0.1%
Other values (3)	9	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	203666	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
o	101733	50.0%
N	101728	49.9%
S	31	< 0.1%
t	31	< 0.1%
e	31	< 0.1%
a	31	< 0.1%
d	31	< 0.1%
y	31	< 0.1%
D	5	< 0.1%
w	5	< 0.1%
Other values (3)	9	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	203666	100.0%

Most frequent character per block

(*unknown*)

Value	Count	Frequency (%)
o	101733	50.0%
N	101728	49.9%
S	31	< 0.1%
t	31	< 0.1%
e	31	< 0.1%
a	31	< 0.1%
d	31	< 0.1%
y	31	< 0.1%
D	5	< 0.1%
w	5	< 0.1%
Other values (3)	9	< 0.1%

troglitazone

Categorical

IMBALANCE

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.7 KiB

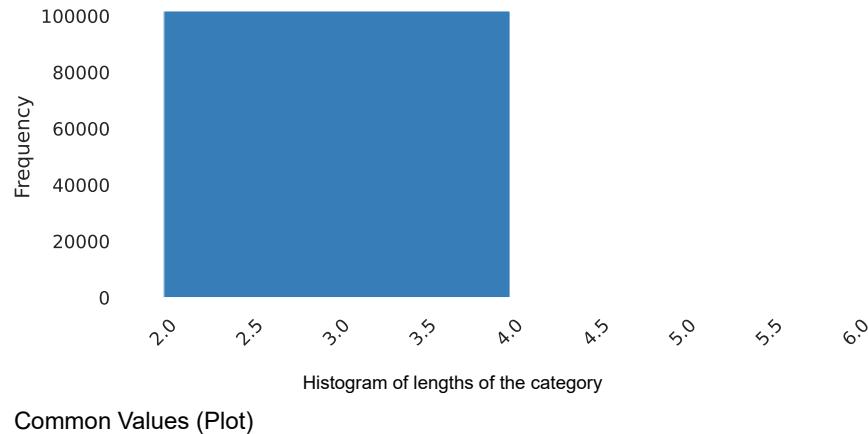
Length		Characters and Unicode	Unique	Sample		
Max length	6		Unique	0	?	1st row
Total characters	203544		Unique (%)	0.0%		2nd row
Median length	2					3rd row
Mean length	2.0001179	Distinct characters	8			4th row
Min length	2	Distinct categories	1			5th row
		(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				?
		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.

Common Values

Value	Count	Frequency (%)
No	101763	> 99.9%
Steady	3	< 0.1%

Length



Common Values (Plot)



Value	Count	Frequency (%)
no	101763	> 99.9%
steady	3	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
N	101763	50.0%
o	101763	50.0%
S	3	< 0.1%
t	3	< 0.1%

Value	Count	Frequency (%)
e	3	< 0.1%
a	3	< 0.1%
d	3	< 0.1%
y	3	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	203544	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
N	101763	50.0%
o	101763	50.0%
S	3	< 0.1%
t	3	< 0.1%
e	3	< 0.1%
a	3	< 0.1%
d	3	< 0.1%
y	3	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	203544	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
N	101763	50.0%
o	101763	50.0%
S	3	< 0.1%
t	3	< 0.1%
e	3	< 0.1%
a	3	< 0.1%
d	3	< 0.1%
y	3	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	203544	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
N	101763	50.0%
o	101763	50.0%
S	3	< 0.1%
t	3	< 0.1%
e	3	< 0.1%
a	3	< 0.1%
d	3	< 0.1%
y	3	< 0.1%

tolazamide

Categorical

IMBALANCE

Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.8 KiB

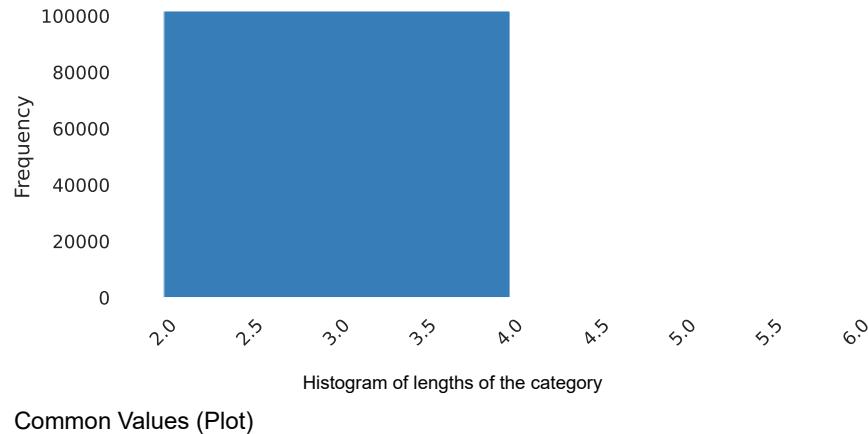
Length		Characters and Unicode	Unique			Sample	
Max length	6		Unique	1	?	1st row	No
Median length	2	Total characters 203684	Unique (%) < 0.1%			2nd row	No
Mean length	2.0014936	Distinct characters 10				3rd row	No
Min length	2	Distinct categories 1				4th row	No
		(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				5th row	?
		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.

Common Values

Value	Count	Frequency (%)
No	101727	> 99.9%
Steady	38	< 0.1%
Up	1	< 0.1%

Length



100.0%
(101727)



Value	Count	Frequency (%)
no	101727	> 99.9%
steady	38	< 0.1%
up	1	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
N	101727	49.9%
o	101727	49.9%

Value	Count	Frequency (%)
S	38	< 0.1%
t	38	< 0.1%
e	38	< 0.1%
a	38	< 0.1%
d	38	< 0.1%
y	38	< 0.1%
U	1	< 0.1%
p	1	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	203684	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
N	101727	49.9%
o	101727	49.9%
S	38	< 0.1%
t	38	< 0.1%
e	38	< 0.1%
a	38	< 0.1%
d	38	< 0.1%
y	38	< 0.1%
U	1	< 0.1%
p	1	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	203684	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
N	101727	49.9%
o	101727	49.9%
S	38	< 0.1%
t	38	< 0.1%
e	38	< 0.1%
a	38	< 0.1%
d	38	< 0.1%
y	38	< 0.1%
U	1	< 0.1%
p	1	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	203684	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
N	101727	49.9%
o	101727	49.9%
S	38	< 0.1%

Value	Count	Frequency (%)
t	38	< 0.1%
e	38	< 0.1%
a	38	< 0.1%
d	38	< 0.1%
y	38	< 0.1%
U	1	< 0.1%
p	1	< 0.1%

examide

Categorical

CONSTANT

Distinct	1
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.7 KiB

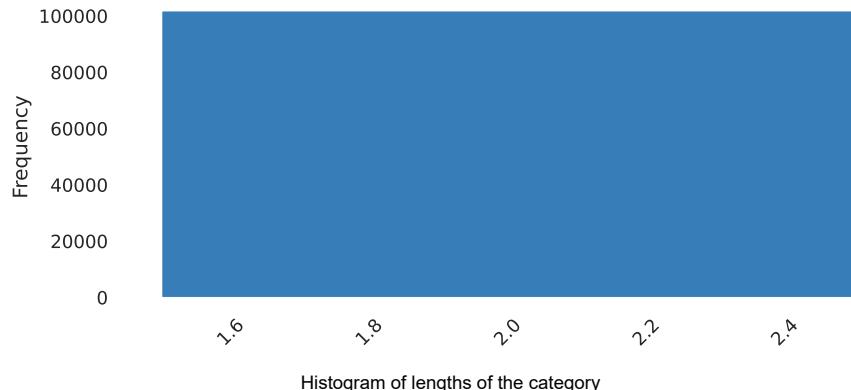
Length	Characters and Unicode	Unique		Sample	
		Unique	0 ?	1st row	No
Max length	2				
Median length	2	Total characters	203532 Unique (%) 0.0%	2nd row	No
Mean length	2			3rd row	No
Min length	2	Distinct characters	2	4th row	No
		Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	5th row	No ?
		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.

Common Values

Value	Count	Frequency (%)
No	101766	100.0%

Length



Common Values (Plot)

100.0%
(101766)

No

Value	Count	Frequency (%)
no	101766	100.0%

Most occurring characters

Value	Count	Frequency (%)
N	101766	50.0%
o	101766	50.0%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	203532	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
N	101766	50.0%
o	101766	50.0%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	203532	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
N	101766	50.0%
o	101766	50.0%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	203532	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
N	101766	50.0%
o	101766	50.0%

citoglipton

Categorical

CONSTANT

Distinct	1
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.7 KiB

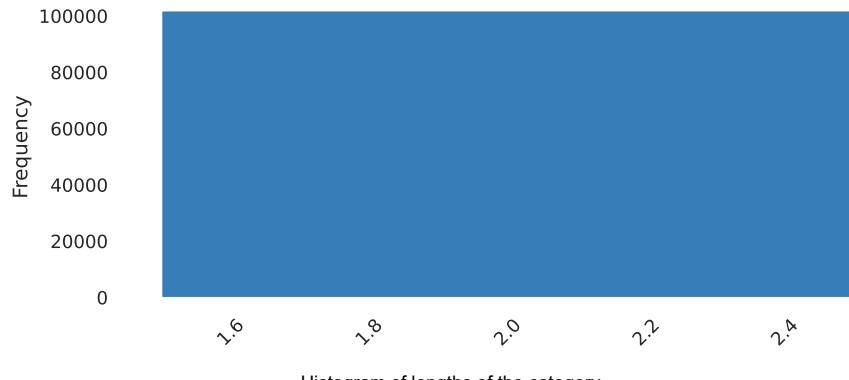
Length	Characters and Unicode	Unique		Sample	
		Unique	0 ?	1st row	No
Max length	2				
Median length	2	Total characters	203532 Unique (%) 0.0%	2nd row	No
Mean length	2			3rd row	No
Min length	2	Distinct characters	2	4th row	No
		Distinct categories	1 (https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	5th row	No ?
		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.

Common Values

Value	Count	Frequency (%)
No	101766	100.0%

Length



Common Values (Plot)

100.0%
(101766)

No

Value	Count	Frequency (%)
no	101766	100.0%

Most occurring characters

Value	Count	Frequency (%)
N	101766	50.0%
o	101766	50.0%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	203532	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
N	101766	50.0%
o	101766	50.0%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	203532	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
N	101766	50.0%
o	101766	50.0%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	203532	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
N	101766	50.0%
o	101766	50.0%

insulin

Categorical

HIGH CORRELATION (This variable has a high overall correlation with 2 fields: change, diabetesMed)

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.9 KiB

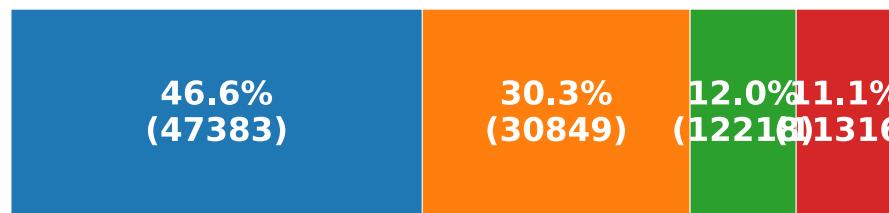
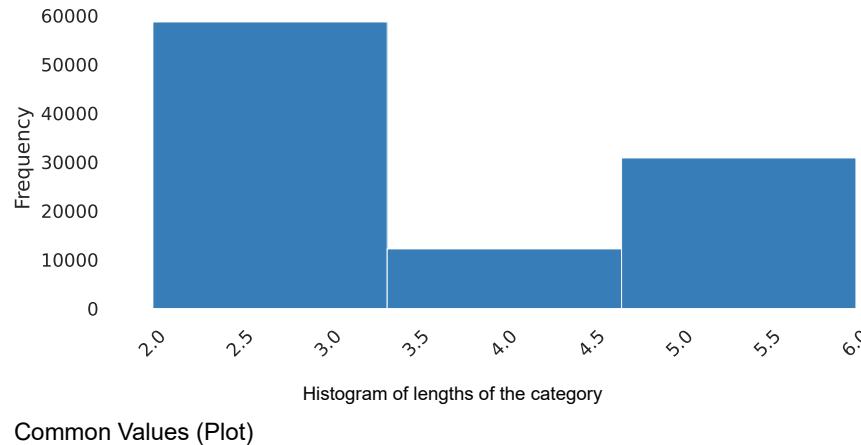
Length	Characters and Unicode	Unique		Sample	
		Unique	0 ?	1st row	No
Max length	6	Total characters	351364	Unique (%)	0.0%
Median length	2	Distinct characters	13	2nd row	Up
Mean length	3.4526659	Distinct categories	1	3rd row	No
Min length	2	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)		4th row	Up
		Distinct scripts	1	5th row	Steady ?
		(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.

Common Values

Value	Count	Frequency (%)
No	47383	46.6%
Steady	30849	30.3%
Down	12218	12.0%
Up	11316	11.1%

Length



Value	Count	Frequency (%)
no	47383	46.6%
steady	30849	30.3%
down	12218	12.0%
up	11316	11.1%

Most occurring characters

Value	Count	Frequency (%)
o	59601	17.0%
N	47383	13.5%
S	30849	8.8%
t	30849	8.8%
e	30849	8.8%
a	30849	8.8%
d	30849	8.8%
y	30849	8.8%
D	12218	3.5%
w	12218	3.5%
Other values (3)	34850	9.9%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	351364	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
o	59601	17.0%
N	47383	13.5%
S	30849	8.8%
t	30849	8.8%
e	30849	8.8%
a	30849	8.8%
d	30849	8.8%
y	30849	8.8%

Value	Count	Frequency (%)
D	12218	3.5%
w	12218	3.5%
Other values (3)	34850	9.9%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	351364	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
o	59601	17.0%
N	47383	13.5%
S	30849	8.8%
t	30849	8.8%
e	30849	8.8%
a	30849	8.8%
d	30849	8.8%
y	30849	8.8%
D	12218	3.5%
w	12218	3.5%
Other values (3)	34850	9.9%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	351364	100.0%

Most frequent character per block

(*unknown*)

Value	Count	Frequency (%)
o	59601	17.0%
N	47383	13.5%
S	30849	8.8%
t	30849	8.8%
e	30849	8.8%
a	30849	8.8%
d	30849	8.8%
y	30849	8.8%
D	12218	3.5%
w	12218	3.5%
Other values (3)	34850	9.9%

glyburide-metformin

Categorical

IMBALANCE

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.9 KiB

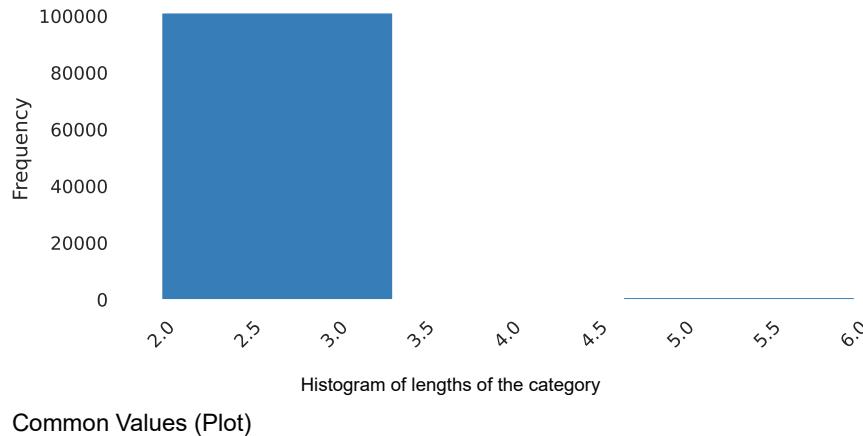
Length		Characters and Unicode	Unique		Sample		
			Unique	0 ?	1st row	No	
Max length	6	Total characters	206312	Unique (%)	0.0%	2nd row	No
Median length	2	Distinct characters	13			3rd row	No
Mean length	2.0273176	Distinct categories	1			4th row	No
Min length	2	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				5th row	?
		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.

Common Values

Value	Count	Frequency (%)
No	101060	99.3%
Steady	692	0.7%
Up	8	< 0.1%
Down	6	< 0.1%

Length



99.3%
(101060)

- █ No
- █ Steady
- █ Up
- █ Down

Value	Count	Frequency (%)
no	101060	99.3%
steady	692	0.7%
up	8	< 0.1%
down	6	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
o	101066	49.0%
N	101060	49.0%
S	692	0.3%
t	692	0.3%
e	692	0.3%
a	692	0.3%
d	692	0.3%
y	692	0.3%
U	8	< 0.1%
p	8	< 0.1%
Other values (3)	18	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	206312	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
o	101066	49.0%
N	101060	49.0%
S	692	0.3%
t	692	0.3%
e	692	0.3%
a	692	0.3%
d	692	0.3%
y	692	0.3%

Value	Count	Frequency (%)
U	8	< 0.1%
p	8	< 0.1%
Other values (3)	18	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	206312	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
o	101066	49.0%
N	101060	49.0%
S	692	0.3%
t	692	0.3%
e	692	0.3%
a	692	0.3%
d	692	0.3%
y	692	0.3%
U	8	< 0.1%
p	8	< 0.1%
Other values (3)	18	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	206312	100.0%

Most frequent character per block

(*unknown*)

Value	Count	Frequency (%)
o	101066	49.0%
N	101060	49.0%
S	692	0.3%
t	692	0.3%
e	692	0.3%
a	692	0.3%
d	692	0.3%
y	692	0.3%
U	8	< 0.1%
p	8	< 0.1%
Other values (3)	18	< 0.1%

glipizide-metformin

Categorical

IMBALANCE

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.7 KiB

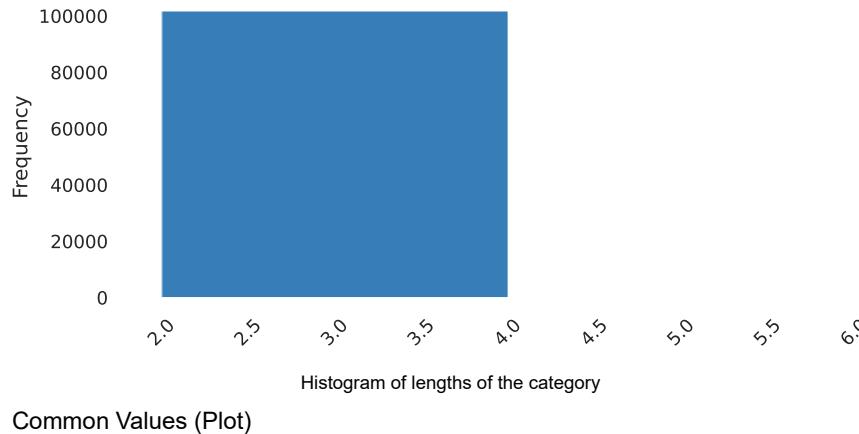
Length	Characters and Unicode	Unique		Sample	
		Unique	0 ?	1st row	No
Max length 6	Total 203584 characters	Unique (%)	0.0%	2nd row	No
Median length	8			3rd row	No
Mean length	2.000511			4th row	No
Min length 2	Distinct 1 categories	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)		5th row	No ?
	Distinct 1 scripts	(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)			?
	Distinct 1 blocks	(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.

Common Values

Value	Count	Frequency (%)
No	101753	> 99.9%
Steady	13	< 0.1%

Length



Common Values (Plot)



Value	Count	Frequency (%)
no	101753	> 99.9%
steady	13	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
N	101753	50.0%
o	101753	50.0%
S	13	< 0.1%
t	13	< 0.1%

Value	Count	Frequency (%)
e	13	< 0.1%
a	13	< 0.1%
d	13	< 0.1%
y	13	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	203584	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
N	101753	50.0%
o	101753	50.0%
S	13	< 0.1%
t	13	< 0.1%
e	13	< 0.1%
a	13	< 0.1%
d	13	< 0.1%
y	13	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	203584	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
N	101753	50.0%
o	101753	50.0%
S	13	< 0.1%
t	13	< 0.1%
e	13	< 0.1%
a	13	< 0.1%
d	13	< 0.1%
y	13	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	203584	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
N	101753	50.0%
o	101753	50.0%
S	13	< 0.1%
t	13	< 0.1%
e	13	< 0.1%
a	13	< 0.1%
d	13	< 0.1%
y	13	< 0.1%

glimepiride-pioglitazone

Categorical

IMBALANCE

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.7 KiB

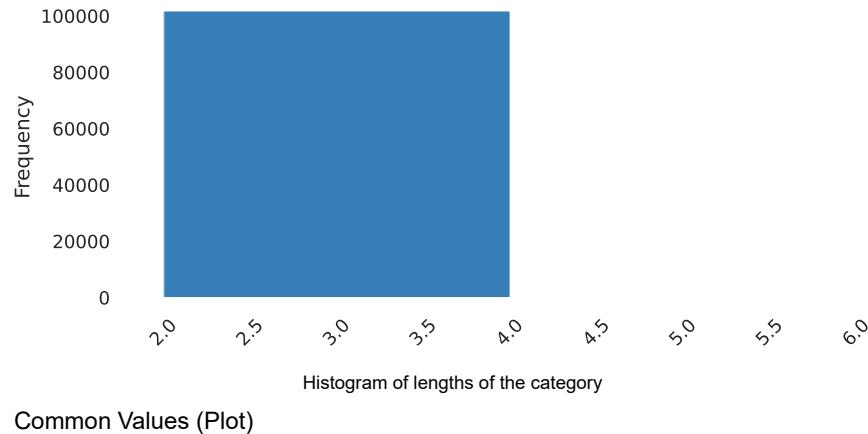
Length	Characters and Unicode	Unique			Sample	
		Unique	1	?	1st row	No
Max length	6	Total characters	203536	Unique (%)	< 0.1%	2nd row
Median length	2	Distinct characters	8			3rd row
Mean length	2.0000393	Distinct categories	1			4th row
Min length	2	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)				5th row
		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.

Common Values

Value	Count	Frequency (%)
No	101765	> 99.9%
Steady	1	< 0.1%

Length



Common Values (Plot)



Value	Count	Frequency (%)
no	101765	> 99.9%
steady	1	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
N	101765	50.0%
o	101765	50.0%
S	1	< 0.1%
t	1	< 0.1%

Value	Count	Frequency (%)
e	1	< 0.1%
a	1	< 0.1%
d	1	< 0.1%
y	1	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	203536	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
N	101765	50.0%
o	101765	50.0%
S	1	< 0.1%
t	1	< 0.1%
e	1	< 0.1%
a	1	< 0.1%
d	1	< 0.1%
y	1	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	203536	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
N	101765	50.0%
o	101765	50.0%
S	1	< 0.1%
t	1	< 0.1%
e	1	< 0.1%
a	1	< 0.1%
d	1	< 0.1%
y	1	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	203536	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
N	101765	50.0%
o	101765	50.0%
S	1	< 0.1%
t	1	< 0.1%
e	1	< 0.1%
a	1	< 0.1%
d	1	< 0.1%
y	1	< 0.1%

metformin-rosiglitazone

Categorical

IMBALANCE

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.7 KiB

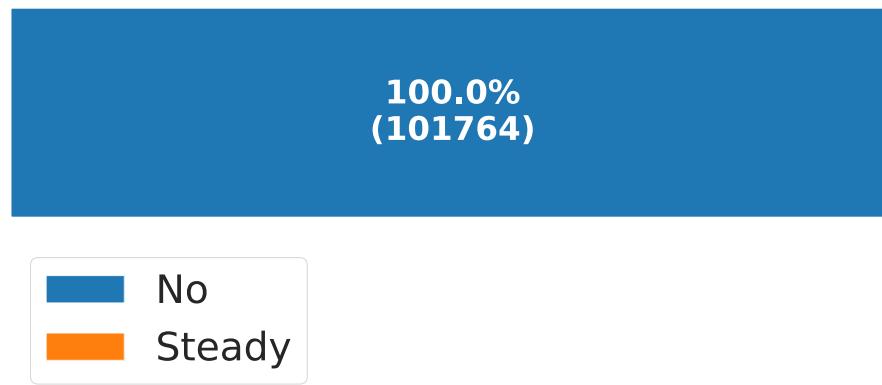
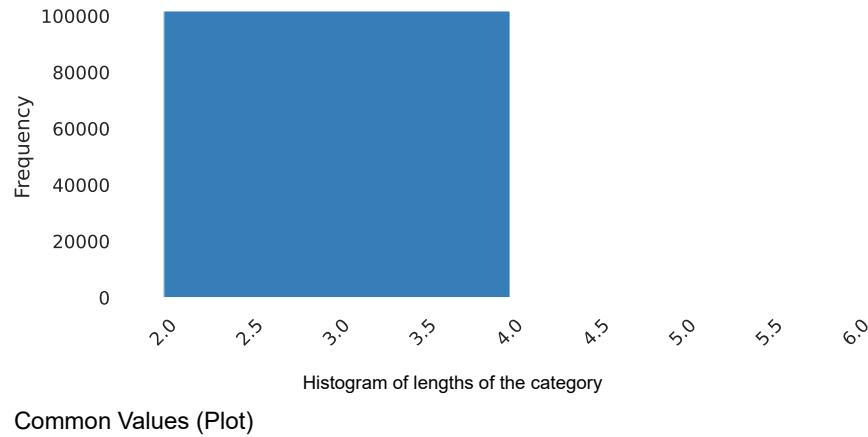
Length		Characters and Unicode	Unique		Sample		
			Unique	0 ?	1st row	No	
Max length	6	Total characters	203540	Unique (%)	0.0%	2nd row	No
Median length	2	Distinct characters	8			3rd row	No
Mean length	2.0000786	Distinct categories	1			4th row	No
Min length	2	Distinct scripts	1	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)			?
		Distinct blocks	1	(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)			?
				(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.

Common Values

Value	Count	Frequency (%)
No	101764	> 99.9%
Steady	2	< 0.1%

Length



Value	Count	Frequency (%)
no	101764	> 99.9%
steady	2	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
N	101764	50.0%
o	101764	50.0%
S	2	< 0.1%
t	2	< 0.1%

Value	Count	Frequency (%)
e	2	< 0.1%
a	2	< 0.1%
d	2	< 0.1%
y	2	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	203540	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
N	101764	50.0%
o	101764	50.0%
S	2	< 0.1%
t	2	< 0.1%
e	2	< 0.1%
a	2	< 0.1%
d	2	< 0.1%
y	2	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	203540	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
N	101764	50.0%
o	101764	50.0%
S	2	< 0.1%
t	2	< 0.1%
e	2	< 0.1%
a	2	< 0.1%
d	2	< 0.1%
y	2	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	203540	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
N	101764	50.0%
o	101764	50.0%
S	2	< 0.1%
t	2	< 0.1%
e	2	< 0.1%
a	2	< 0.1%
d	2	< 0.1%
y	2	< 0.1%

metformin-pioglitazone

Categorical

IMBALANCE

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.7 KiB

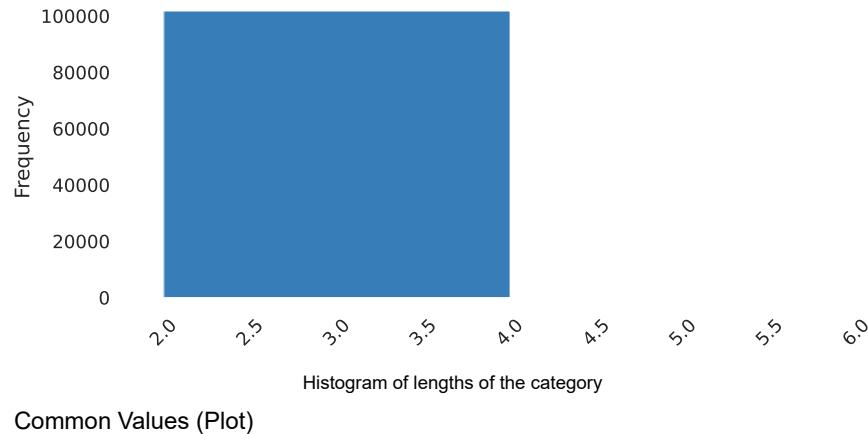
Length		Characters and Unicode	Unique			Sample	
Max length	6		Unique	1	?	1st row	No
Total characters	203536	Unique (%)	< 0.1%			2nd row	No
Distinct characters	8					3rd row	No
Distinct categories	1					4th row	No
Distinct scripts	1						?
Distinct blocks	1	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)					?

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

Common Values

Value	Count	Frequency (%)
No	101765	> 99.9%
Steady	1	< 0.1%

Length



Common Values (Plot)



Value	Count	Frequency (%)
no	101765	> 99.9%
steady	1	< 0.1%

Most occurring characters

Value	Count	Frequency (%)
N	101765	50.0%
o	101765	50.0%
S	1	< 0.1%
t	1	< 0.1%

Value	Count	Frequency (%)
e	1	< 0.1%
a	1	< 0.1%
d	1	< 0.1%
y	1	< 0.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	203536	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
N	101765	50.0%
o	101765	50.0%
S	1	< 0.1%
t	1	< 0.1%
e	1	< 0.1%
a	1	< 0.1%
d	1	< 0.1%
y	1	< 0.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	203536	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
N	101765	50.0%
o	101765	50.0%
S	1	< 0.1%
t	1	< 0.1%
e	1	< 0.1%
a	1	< 0.1%
d	1	< 0.1%
y	1	< 0.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	203536	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
N	101765	50.0%
o	101765	50.0%
S	1	< 0.1%
t	1	< 0.1%
e	1	< 0.1%
a	1	< 0.1%
d	1	< 0.1%
y	1	< 0.1%

change

Categorical

HIGH CORRELATION (This variable has a high overall correlation with 2 fields: diabetesMed, insulin)

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.7 KiB

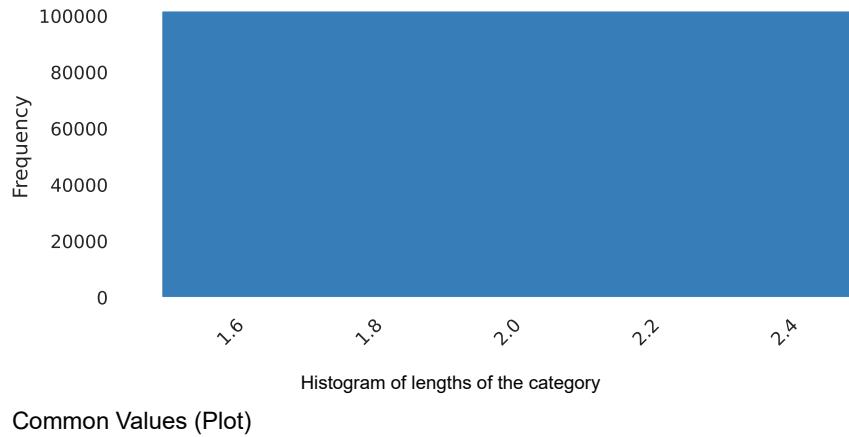
Length	Characters and Unicode	Unique		Sample	
		Unique	0 ?	1st row	No
Max length	2	Total characters	203532	Unique (%)	0.0%
Median length	2	Distinct characters	4	2nd row	Ch
Mean length	2	Distinct categories	1	3rd row	No
Min length	2	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)		4th row	Ch
	Distinct scripts	1		5th row	Ch ?
		(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.

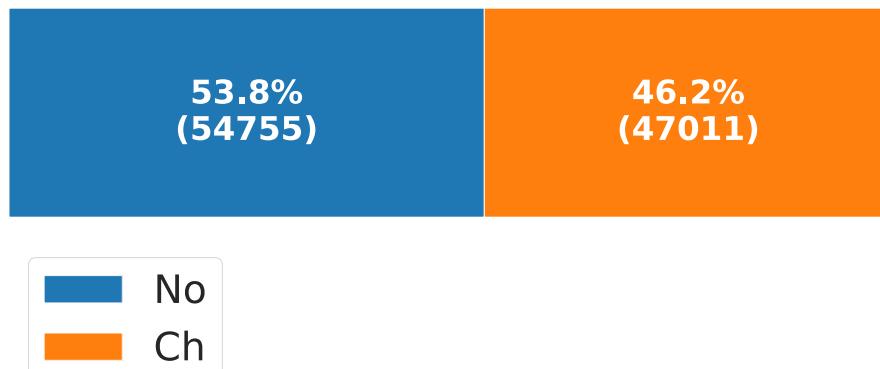
Common Values

Value	Count	Frequency (%)
No	54755	53.8%
Ch	47011	46.2%

Length



Common Values (Plot)



Value	Count	Frequency (%)
no	54755	53.8%
ch	47011	46.2%

Most occurring characters

Value	Count	Frequency (%)
N	54755	26.9%
o	54755	26.9%
C	47011	23.1%
h	47011	23.1%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	203532	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
N	54755	26.9%
o	54755	26.9%
C	47011	23.1%
h	47011	23.1%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	203532	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
N	54755	26.9%
o	54755	26.9%
C	47011	23.1%
h	47011	23.1%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	203532	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
N	54755	26.9%
o	54755	26.9%
C	47011	23.1%
h	47011	23.1%

diabetesMed

Categorical

HIGH CORRELATION (This variable has a high overall correlation with 2 fields: change, insulin)

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.7 KiB

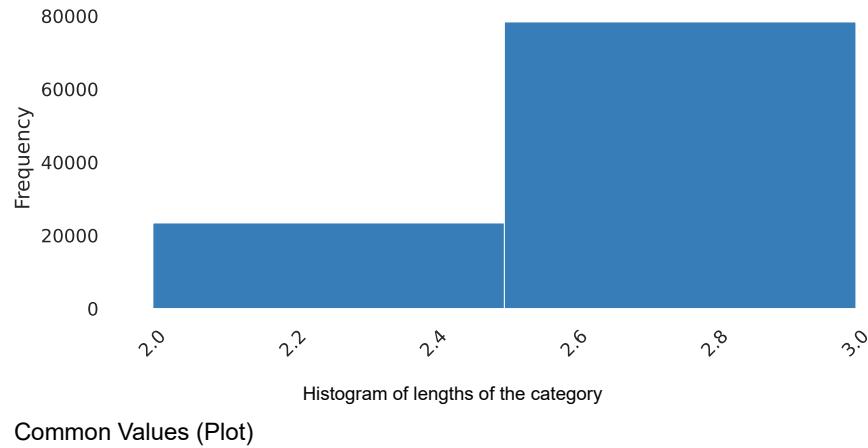
Length		Characters and Unicode	Unique		Sample	
			Unique	0 ?	1st row	No
Max length	3	Total characters	281895	Unique (%) 0.0%	2nd row	Yes
Median length	3	Distinct characters	5		3rd row	Yes
Mean length	2.7700312	Distinct categories	1		4th row	Yes
Min length	2	Distinct scripts	1	(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)	5th row	Yes ?
		Distinct blocks	1	(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)		?
				(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.

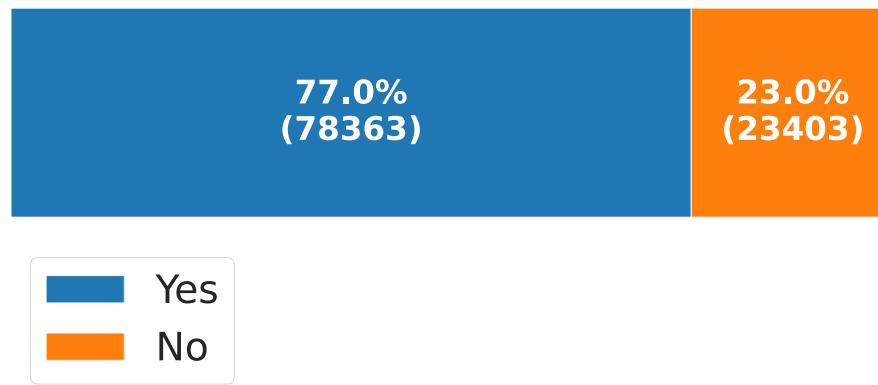
Common Values

Value	Count	Frequency (%)
Yes	78363	77.0%
No	23403	23.0%

Length



Common Values (Plot)



Value	Count	Frequency (%)
yes	78363	77.0%
no	23403	23.0%

Most occurring characters

Value	Count	Frequency (%)
Y	78363	27.8%
e	78363	27.8%
s	78363	27.8%
N	23403	8.3%

Value	Count	Frequency (%)
o	23403	8.3%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	281895	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
Y	78363	27.8%
e	78363	27.8%
s	78363	27.8%
N	23403	8.3%
o	23403	8.3%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	281895	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
Y	78363	27.8%
e	78363	27.8%
s	78363	27.8%
N	23403	8.3%

Value	Count	Frequency (%)
o	23403	8.3%

Most occurring blocks

Value	Count	Frequency (%)
(unknown)	281895	100.0%

Most frequent character per block

(unknown)

Value	Count	Frequency (%)
Y	78363	27.8%
e	78363	27.8%
s	78363	27.8%
N	23403	8.3%
o	23403	8.3%

readmitted

Categorical

Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	99.8 KiB

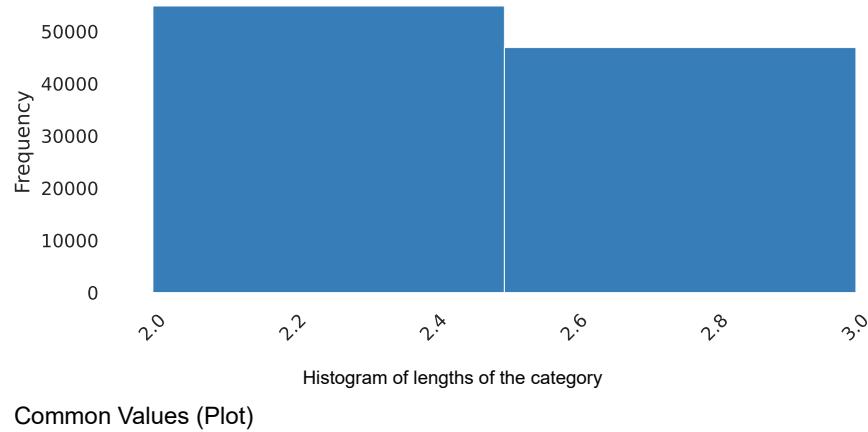
Length	Characters and Unicode	Unique			Sample		
		Unique	0	?	1st row	NO	
Max length	3	Total characters	250434	Unique (%)	0.0%	2nd row	>30
Median length	2	Distinct characters	6			3rd row	NO
Mean length	2.4608808	Distinct categories	1			4th row	NO
Min length	2	Distinct scripts	1			5th row	NO
		Distinct blocks	1				(https://en.wikipedia.org/wiki/Unicode_character_property#General_Category)
							(https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode)
							(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.

Common Values

Value	Count	Frequency (%)
NO	54864	53.9%
>30	35545	34.9%
<30	11357	11.2%

Length



Value	Count	Frequency (%)
no	54864	53.9%
30	46902	46.1%

Most occurring characters

Value	Count	Frequency (%)
N	54864	21.9%
O	54864	21.9%
3	46902	18.7%

Value	Count	Frequency (%)
0	46902	18.7%
>	35545	14.2%
<	11357	4.5%

Most occurring categories

Value	Count	Frequency (%)
(unknown)	250434	100.0%

Most frequent character per category

(unknown)

Value	Count	Frequency (%)
N	54864	21.9%
O	54864	21.9%
3	46902	18.7%
0	46902	18.7%
>	35545	14.2%
<	11357	4.5%

Most occurring scripts

Value	Count	Frequency (%)
(unknown)	250434	100.0%

Most frequent character per script

(unknown)

Value	Count	Frequency (%)
N	54864	21.9%

Value	Count	Frequency (%)
O	54864	21.9%
3	46902	18.7%
0	46902	18.7%
>	35545	14.2%
<	11357	4.5%

Most occurring blocks

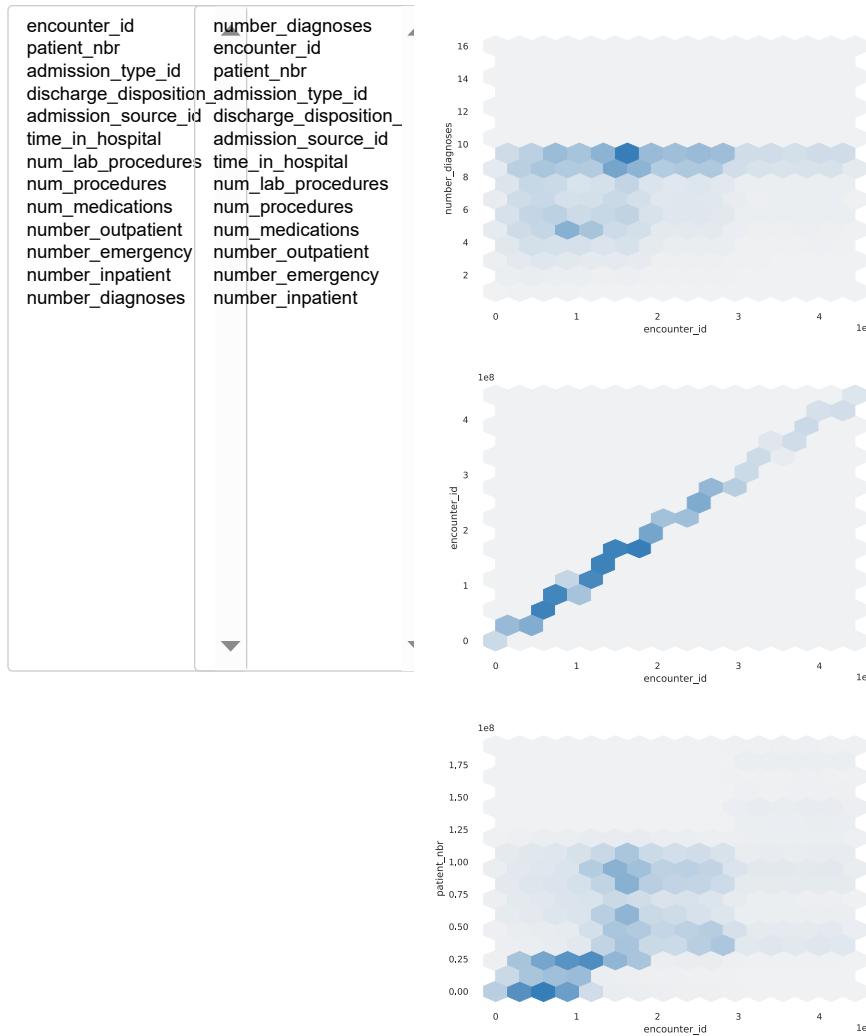
Value	Count	Frequency (%)
(unknown)	250434	100.0%

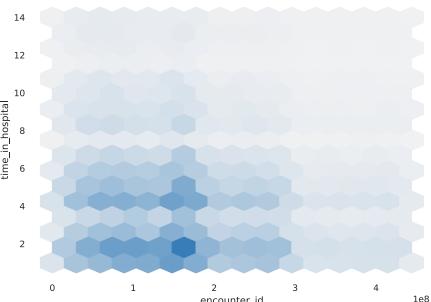
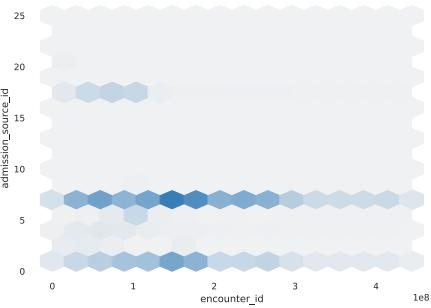
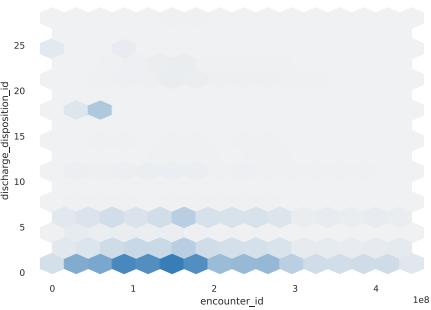
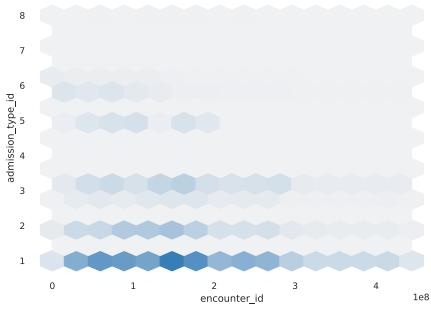
Most frequent character per block

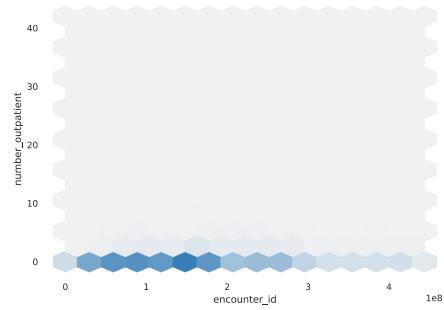
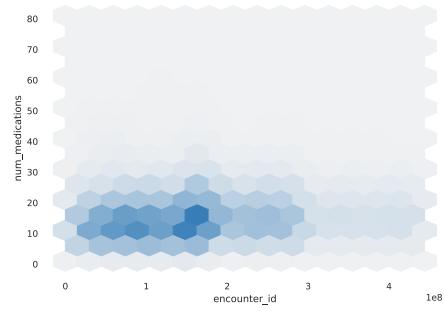
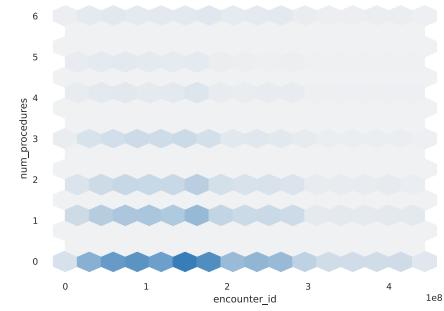
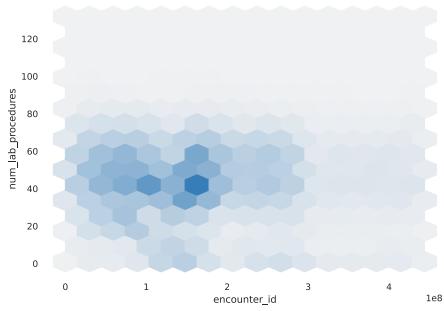
(unknown)

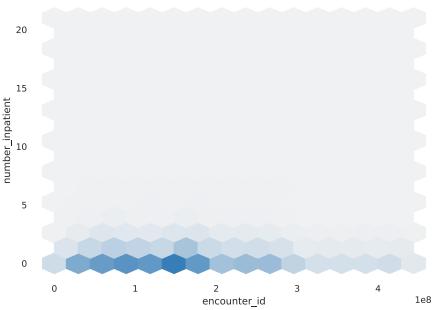
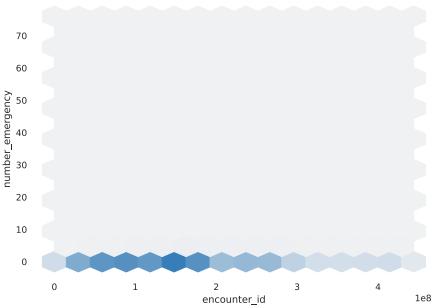
Value	Count	Frequency (%)
N	54864	21.9%
O	54864	21.9%
3	46902	18.7%
0	46902	18.7%
>	35545	14.2%
<	11357	4.5%

Interactions

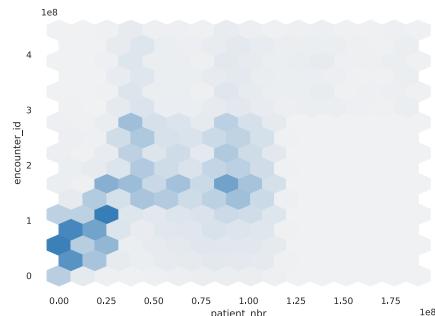
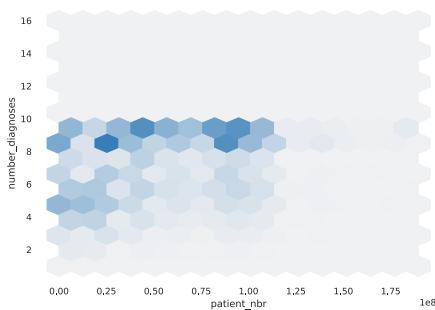


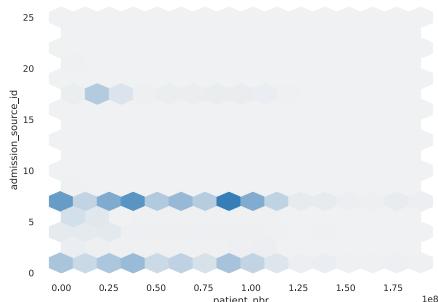
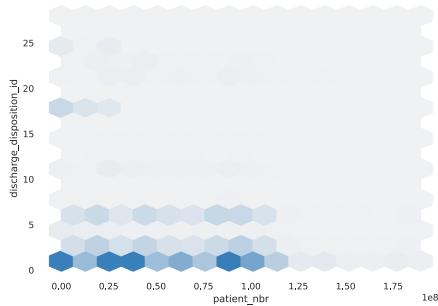
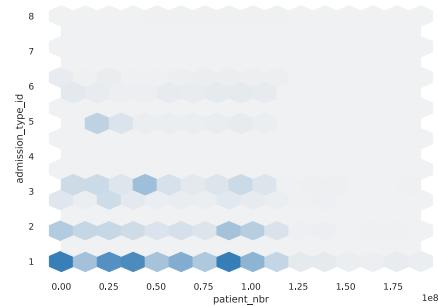
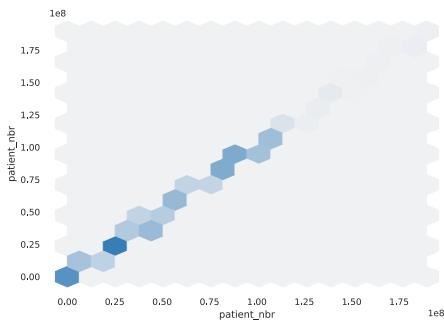


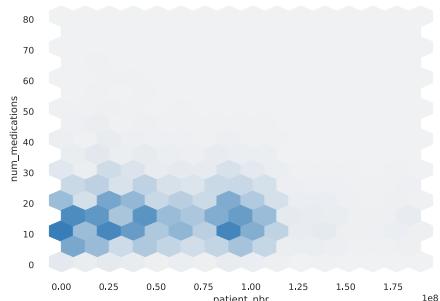
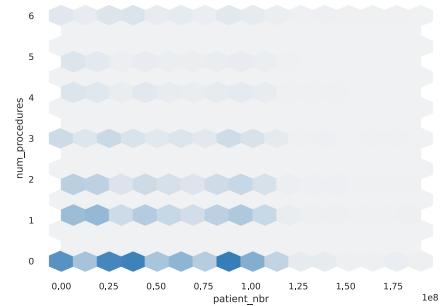
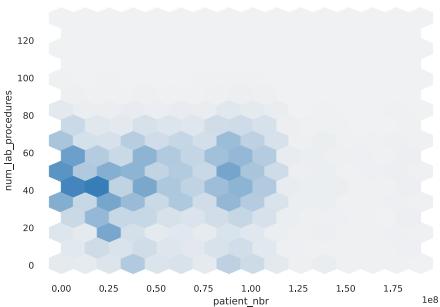
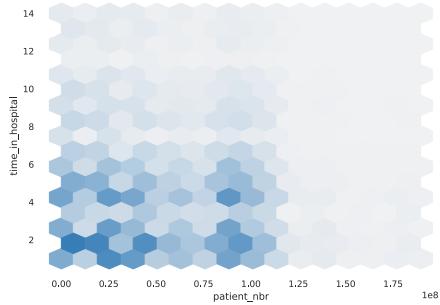


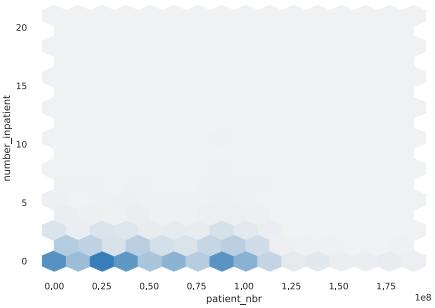
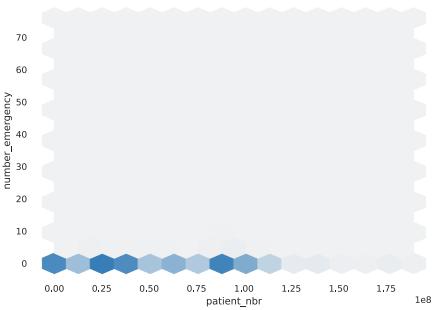
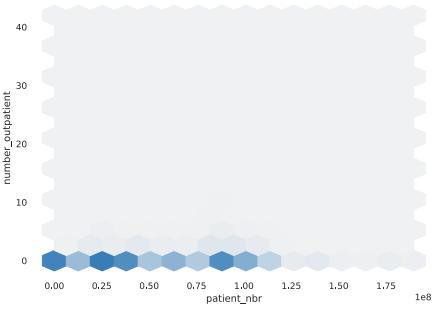


number_diagnoses
encounter_id
patient_nbr
admission_type_id
discharge_disposition_id
admission_source_id
time_in_hospital
num_lab_procedures
num_procedures
num_medications
number_outpatient
number_emergency
number_inpatient

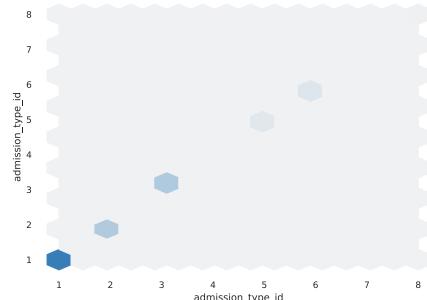
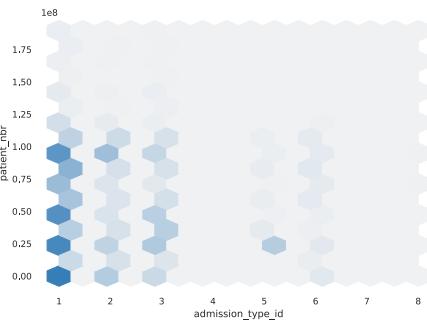
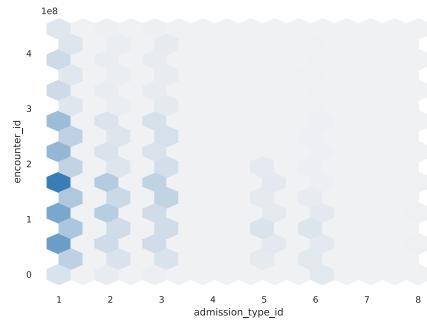
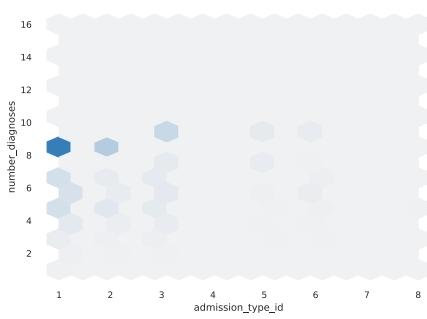


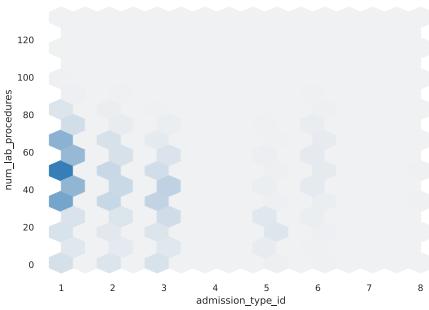
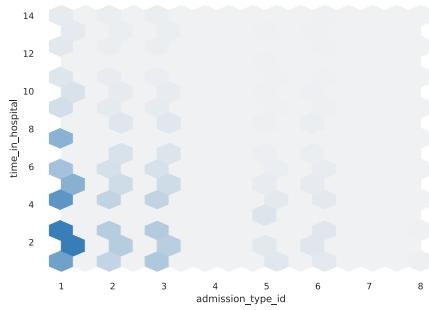
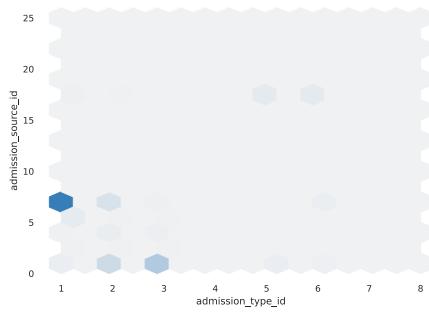
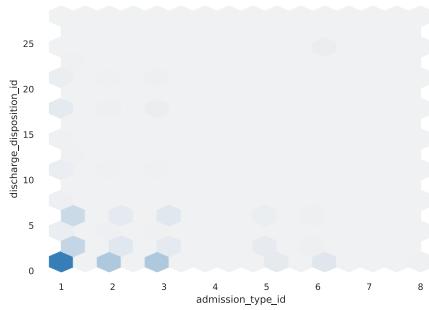


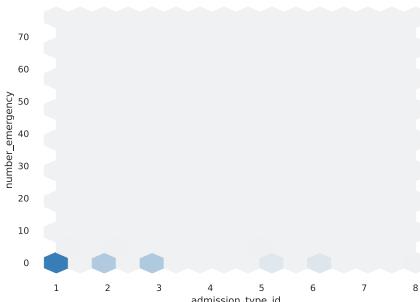
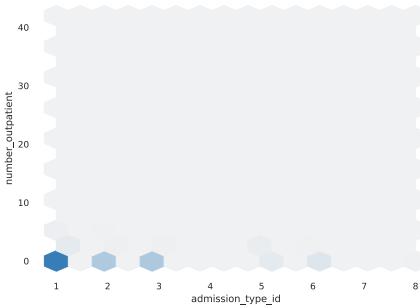
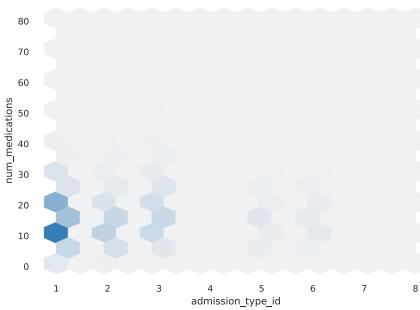
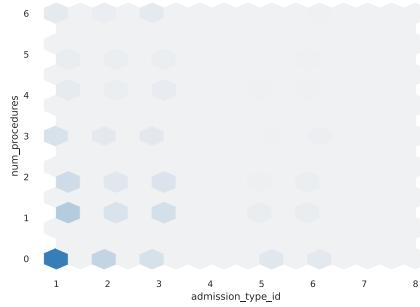


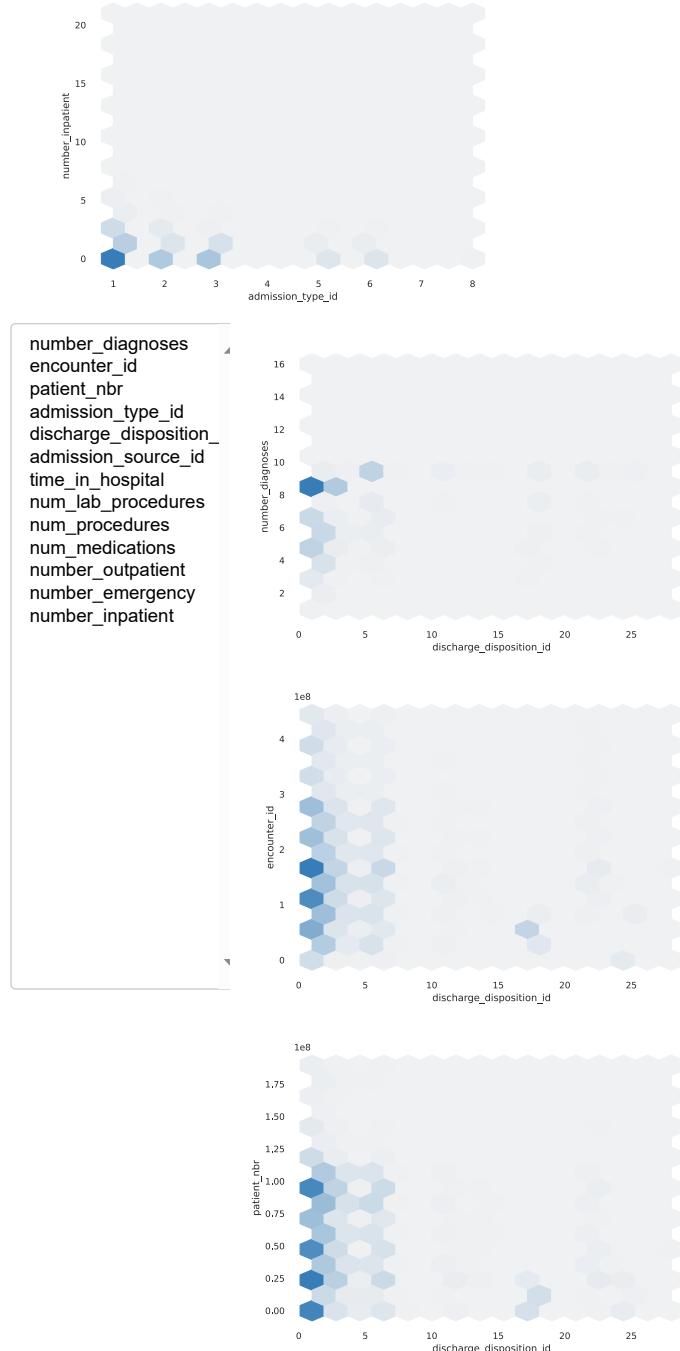


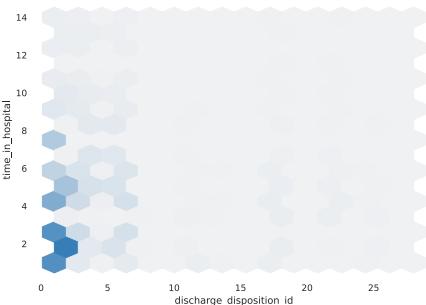
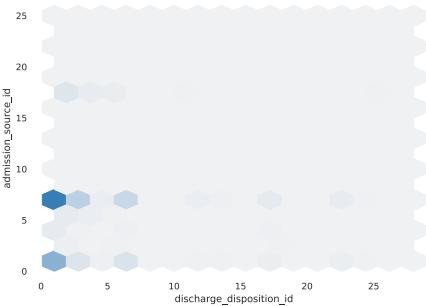
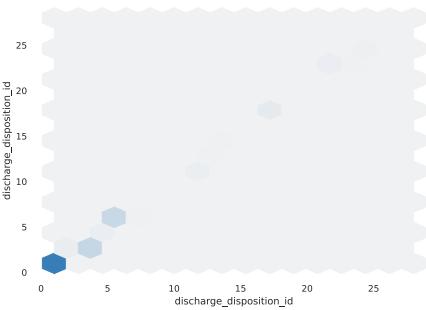
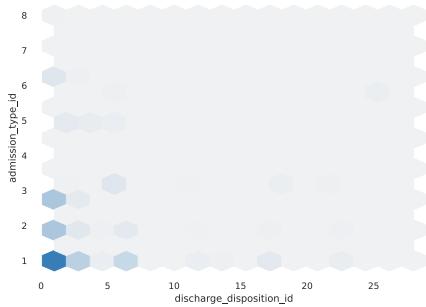
number_diagnoses
encounter_id
patient_nbr
admission_type_id
discharge_disposition_id
admission_source_id
time_in_hospital
num_lab_procedures
num_procedures
num_medications
number_outpatient
number_emergency
number_inpatient

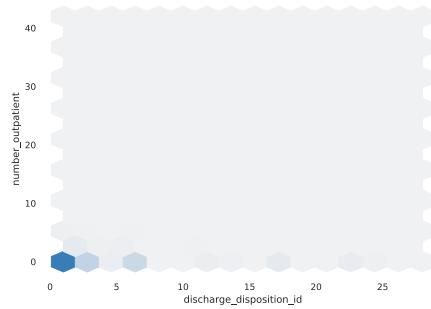
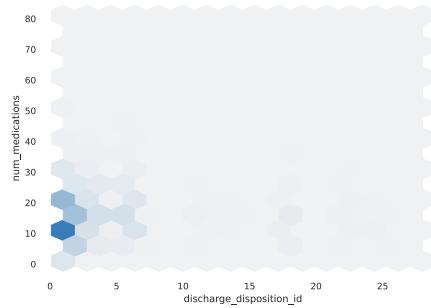
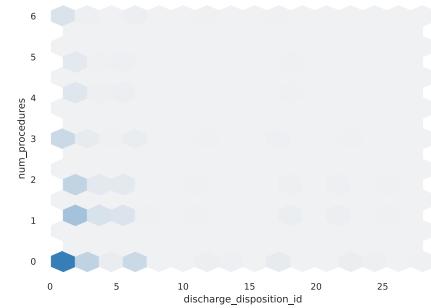
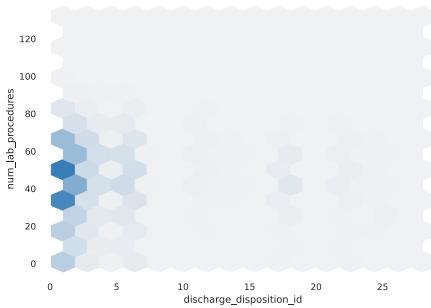


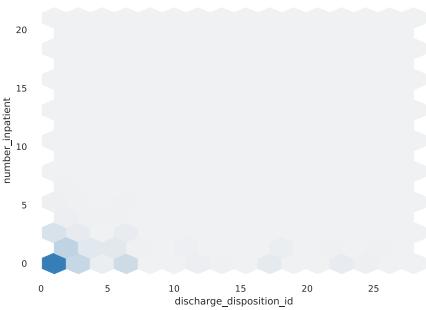
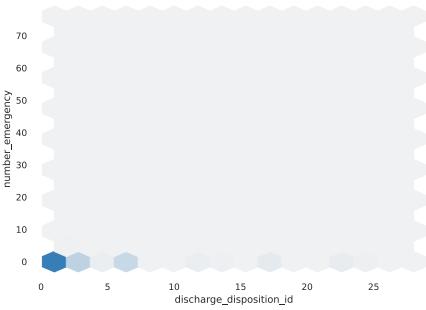




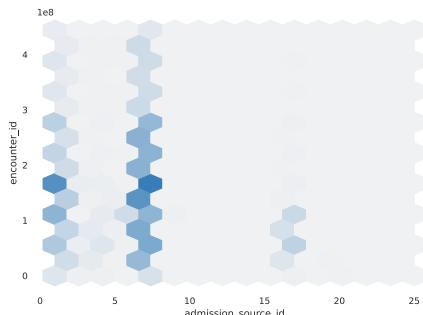
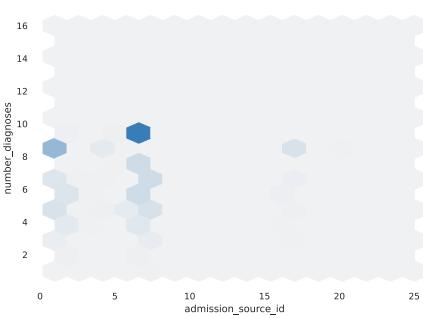


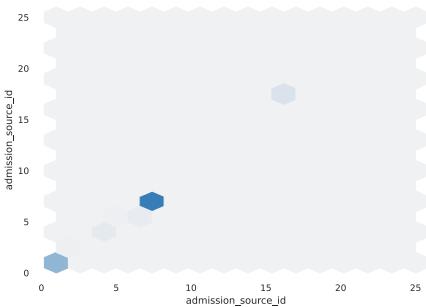
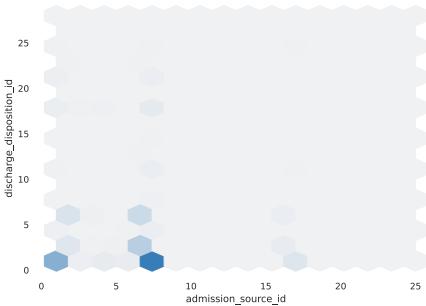
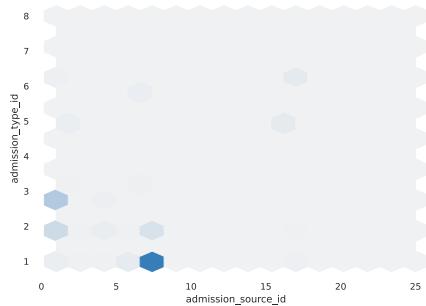
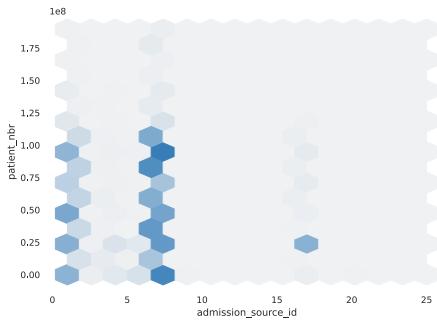


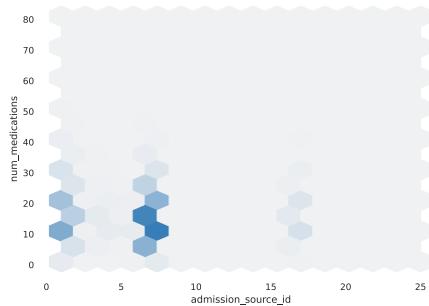
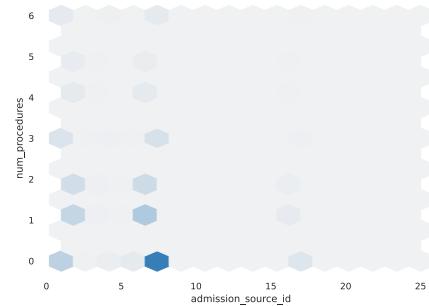
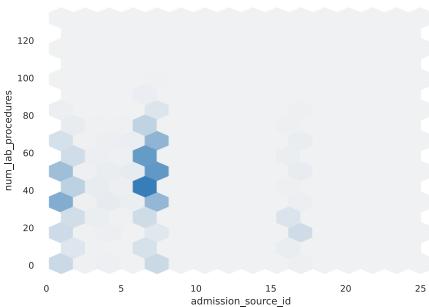
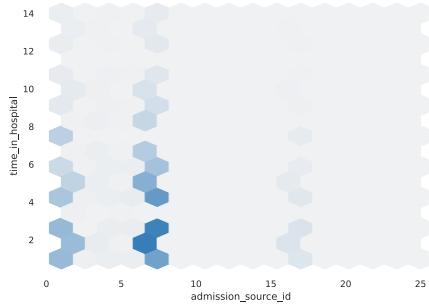


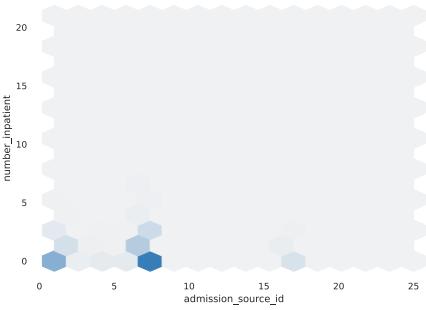
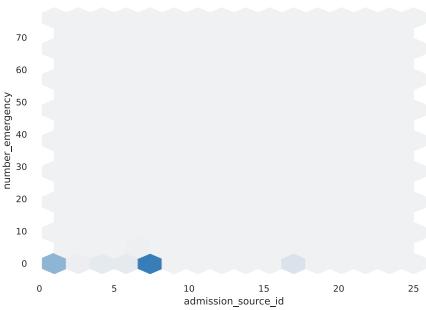
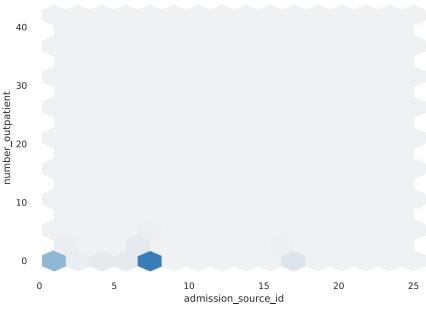


number_diagnoses
encounter_id
patient_nbr
admission_type_id
discharge_disposition_id
admission_source_id
time_in_hospital
num_lab_procedures
num_procedures
num_medications
number_outpatient
number_emergency
number_inpatient

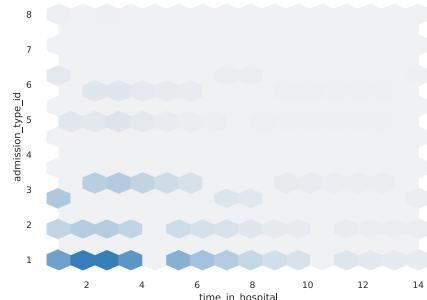
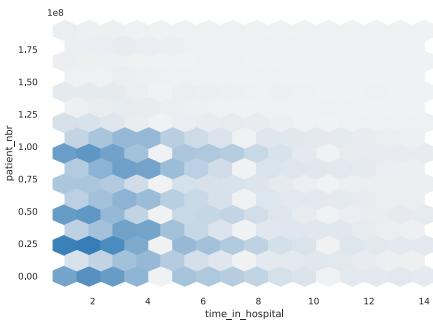
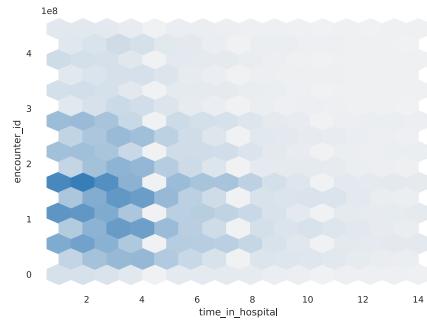
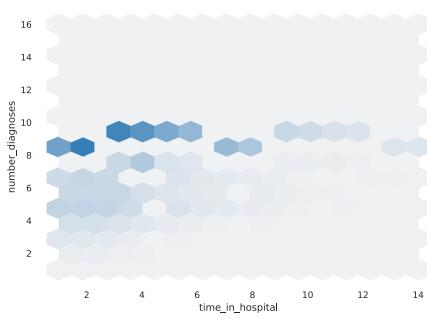


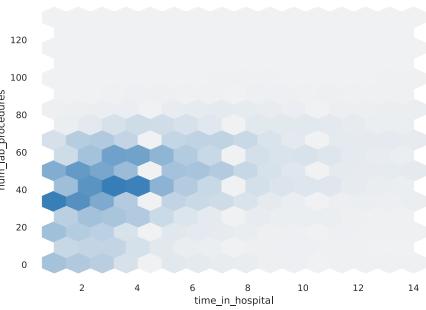
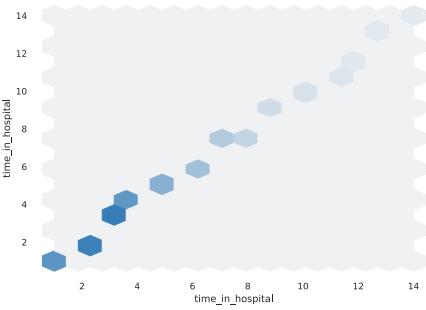
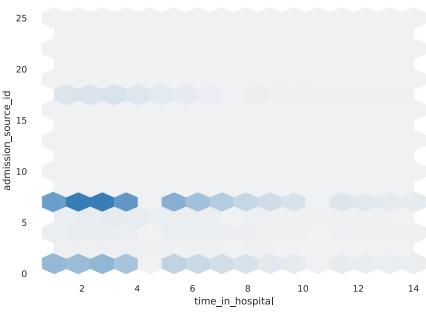
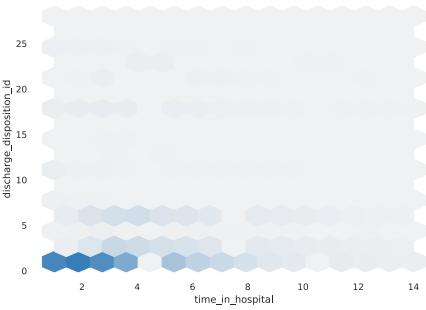


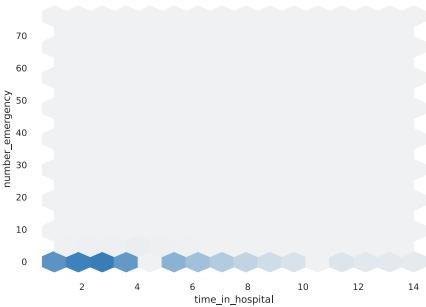
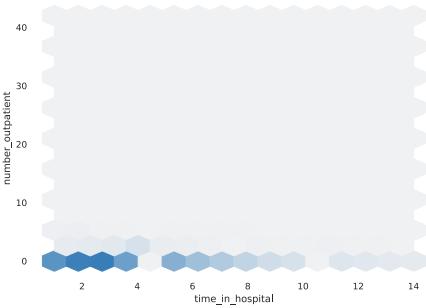
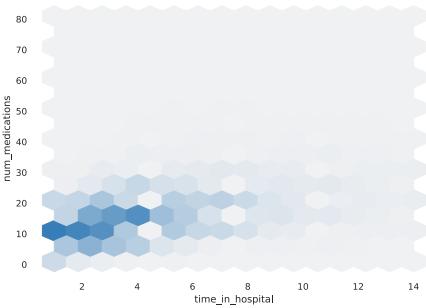
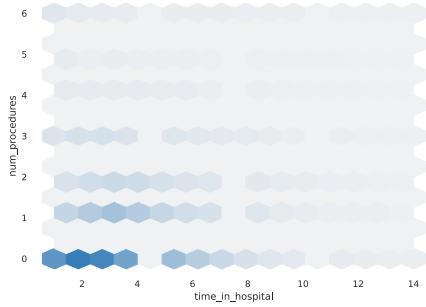


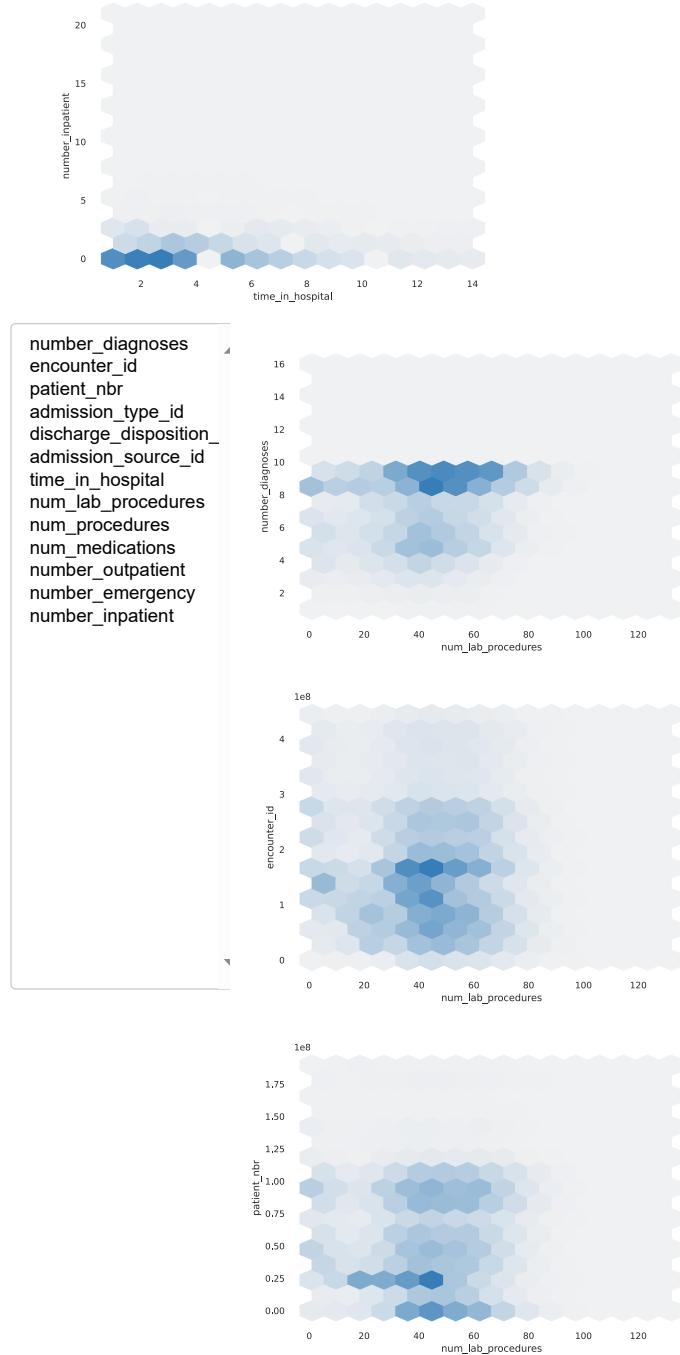


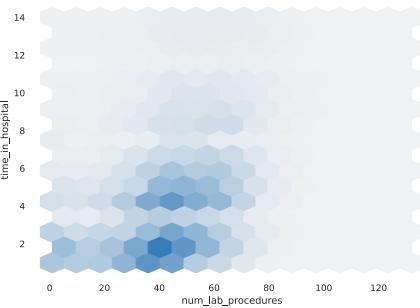
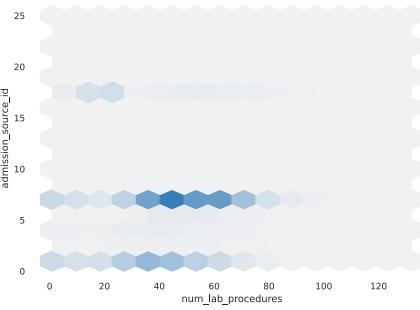
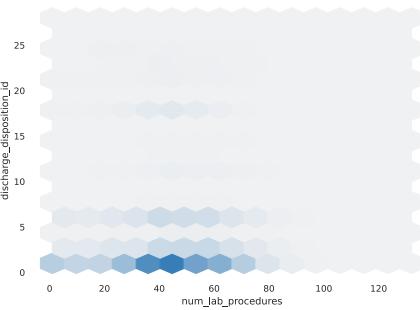
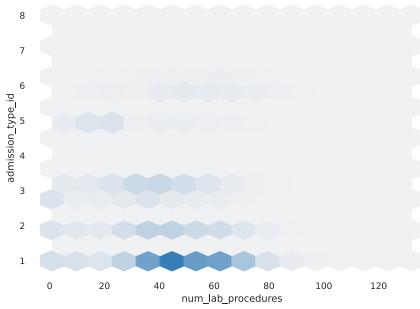
number_diagnoses
encounter_id
patient_nbr
admission_type_id
discharge_disposition_id
admission_source_id
time_in_hospital
num_lab_procedures
num_procedures
num_medications
number_outpatient
number_emergency
number_inpatient

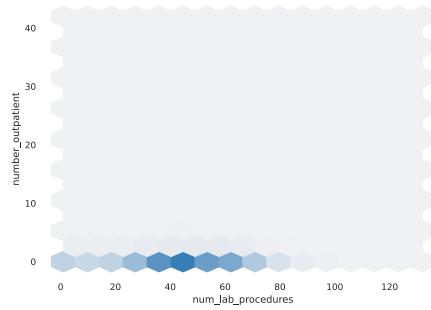
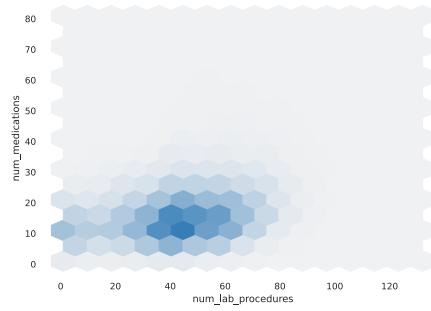
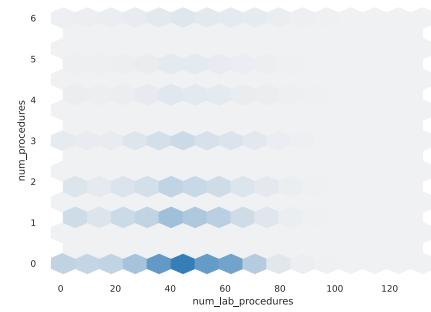
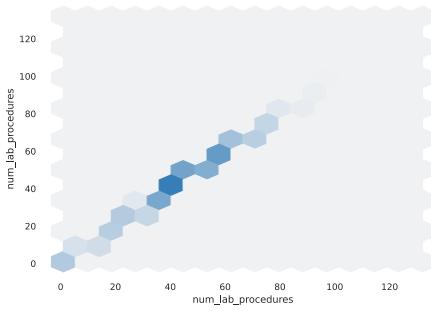


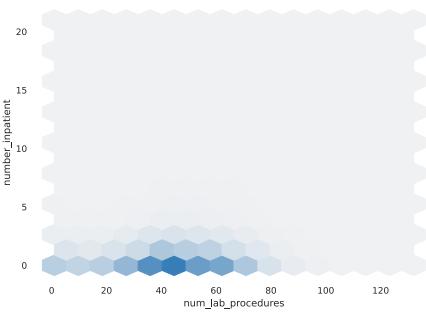
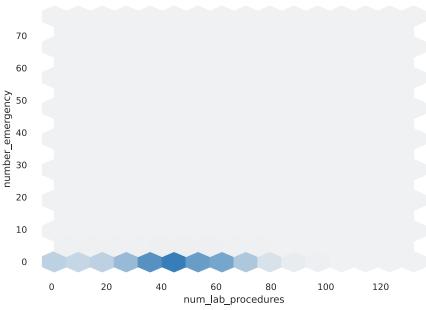




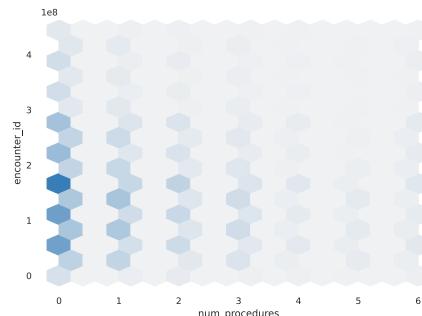
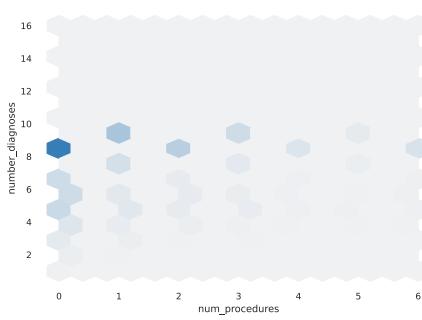


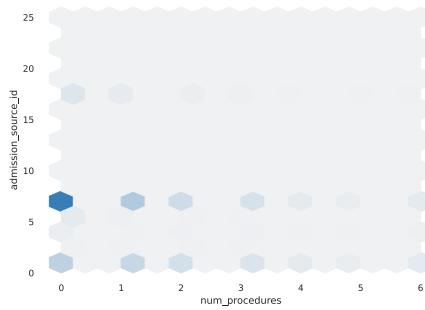
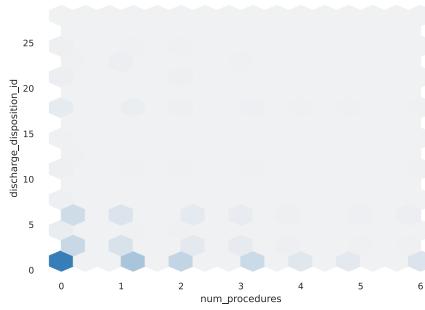
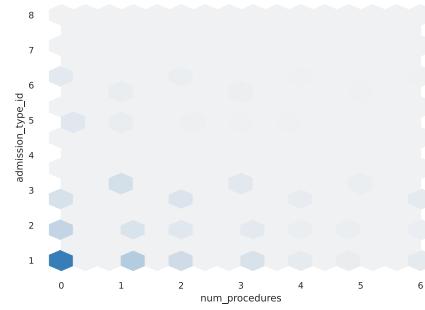
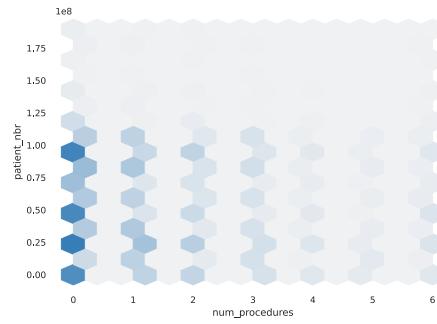


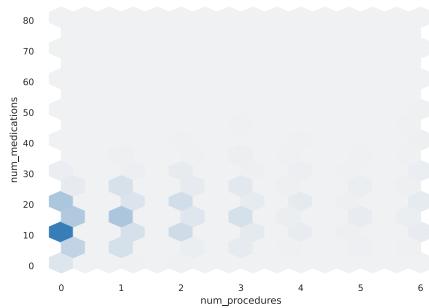
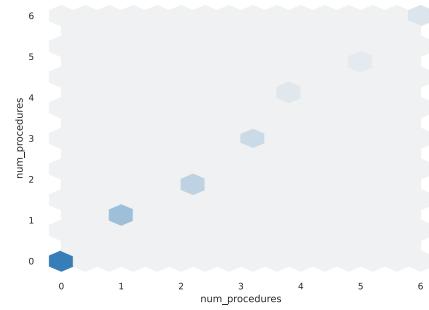
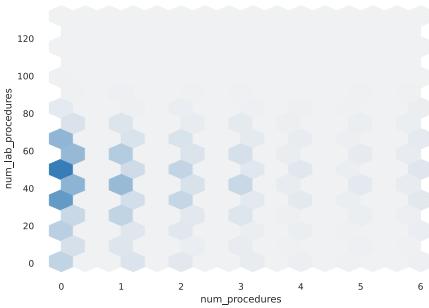
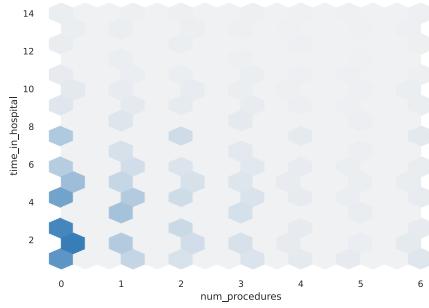


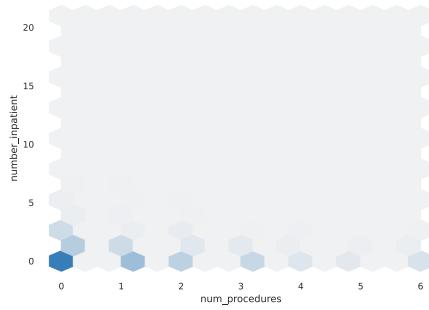
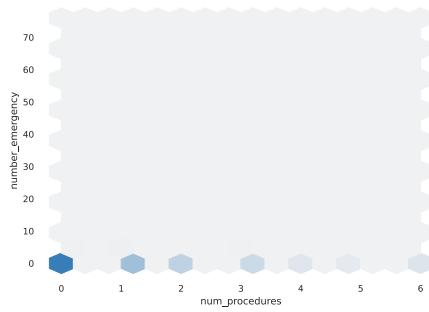
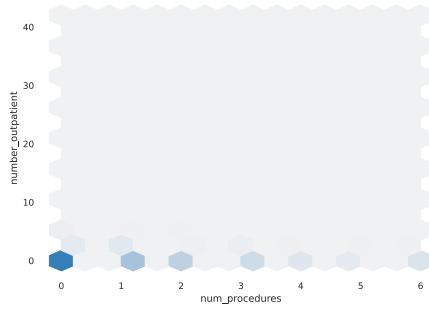


number_diagnoses
encounter_id
patient_nbr
admission_type_id
discharge_disposition_id
admission_source_id
time_in_hospital
num_lab_procedures
num_procedures
num_medications
number_outpatient
number_emergency
number_inpatient

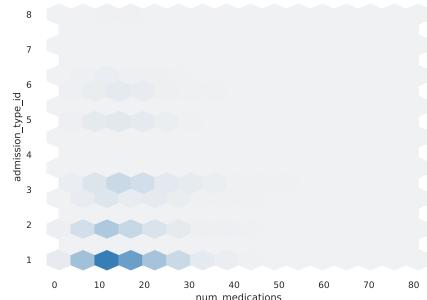
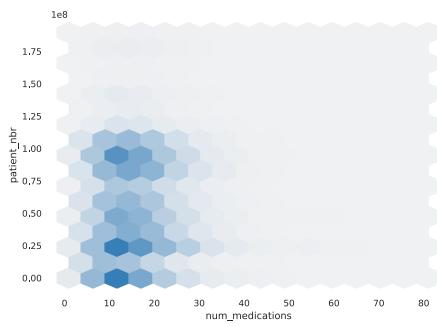
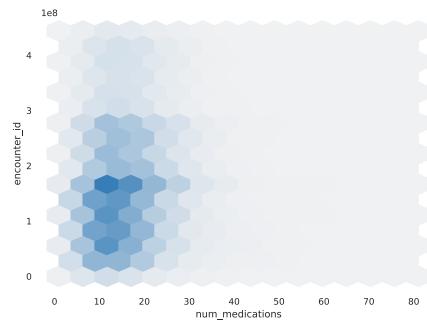
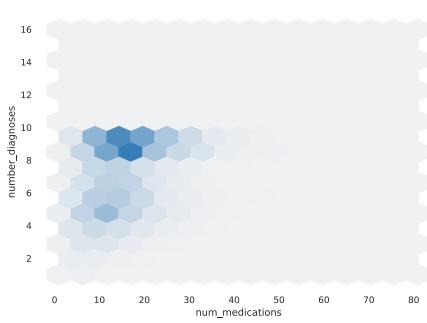


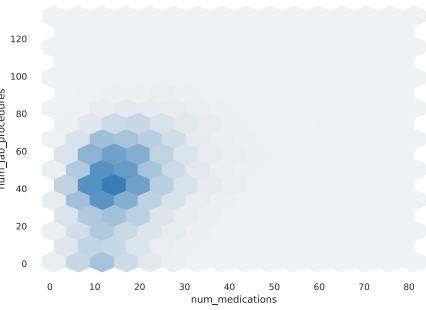
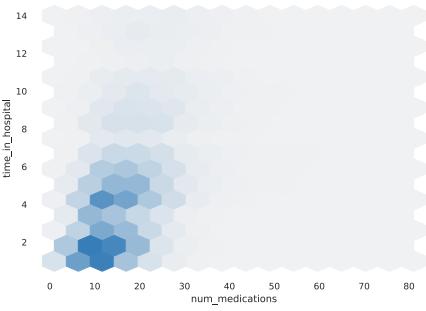
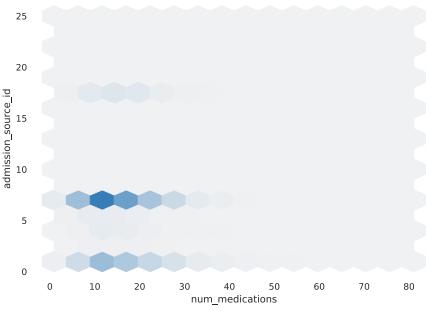
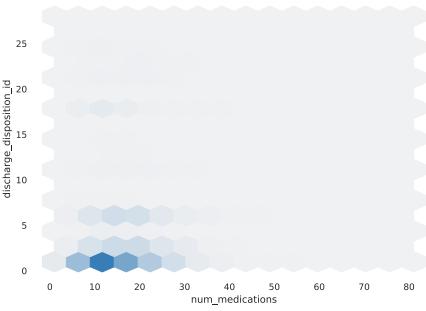


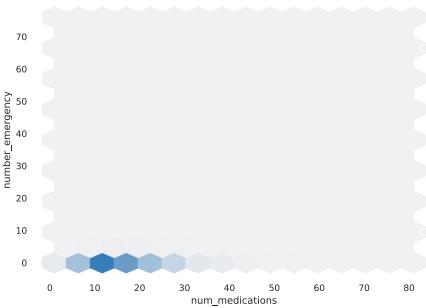
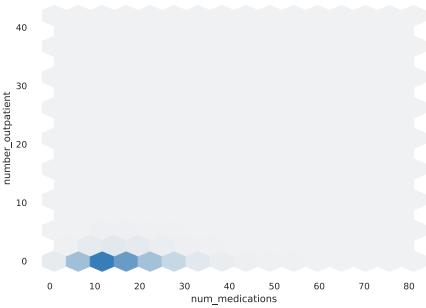
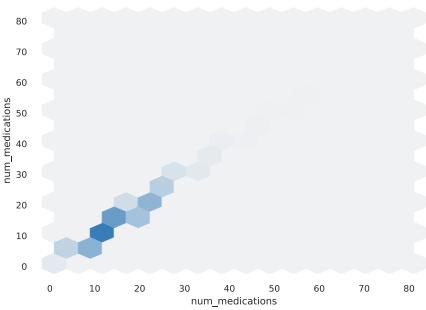
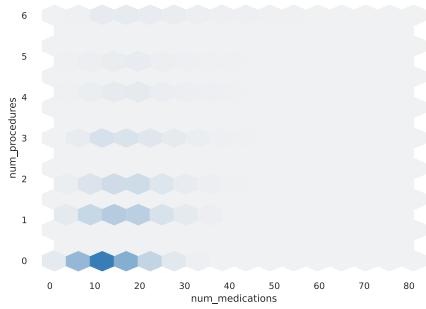


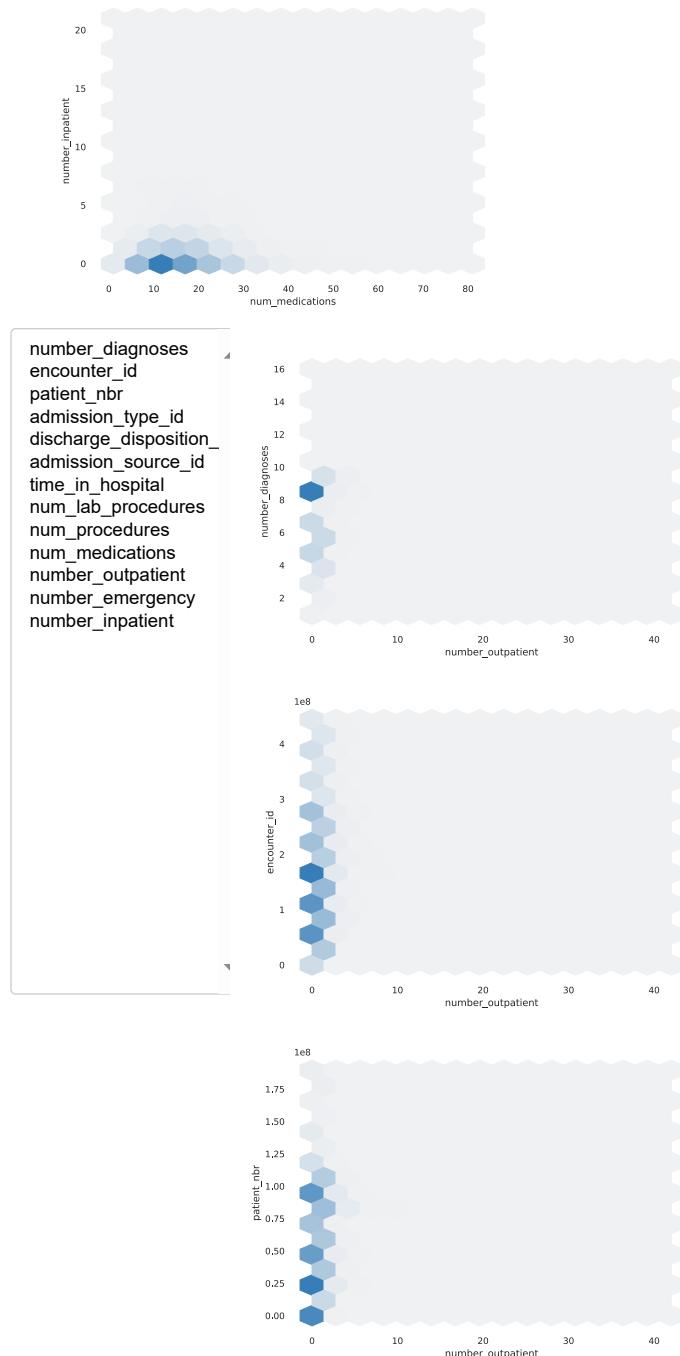


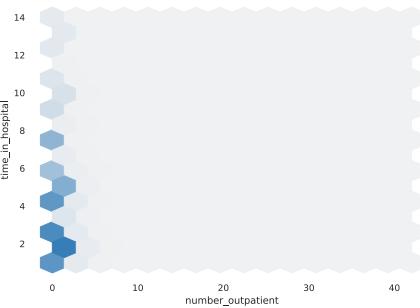
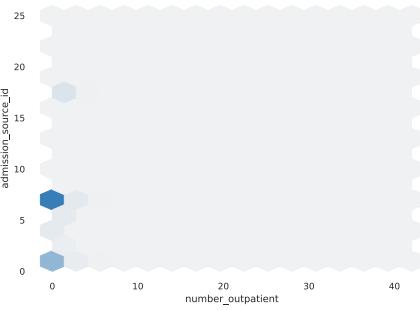
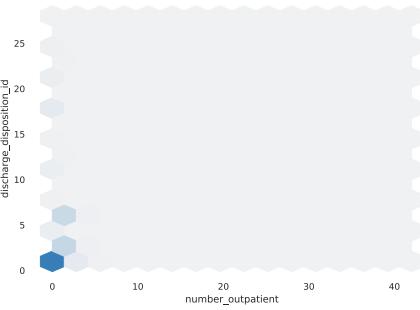
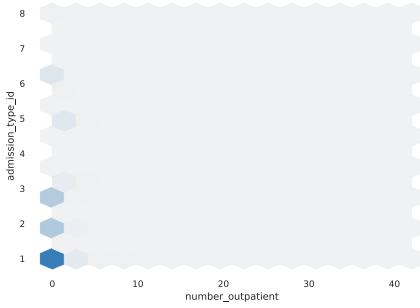
number_diagnoses
encounter_id
patient_nbr
admission_type_id
discharge_disposition_id
admission_source_id
time_in_hospital
num_lab_procedures
num_procedures
num_medications
number_outpatient
number_emergency
number_inpatient

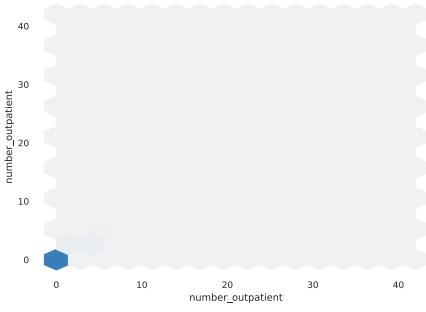
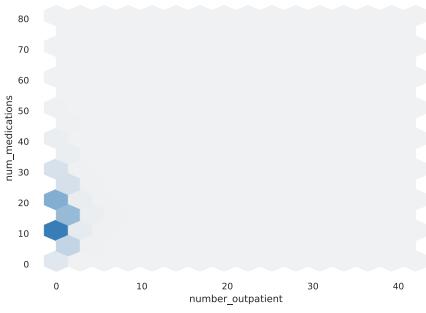
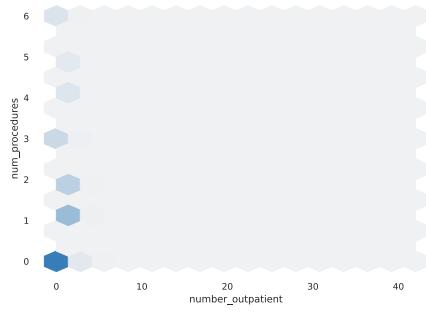
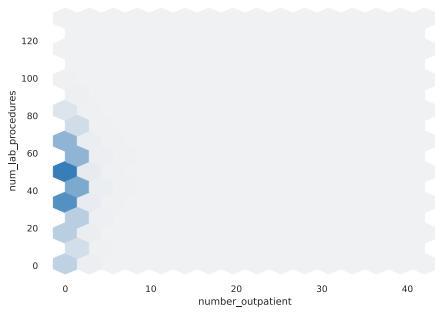


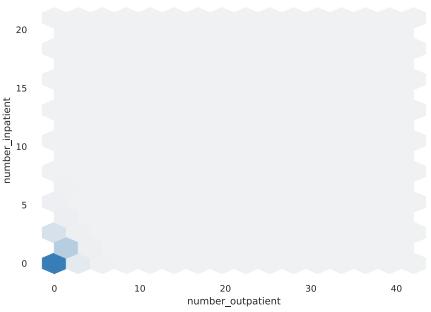
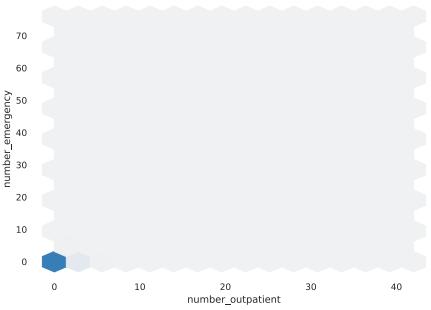




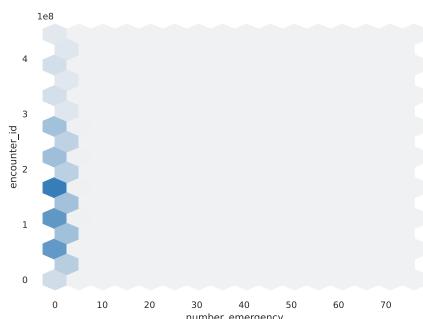
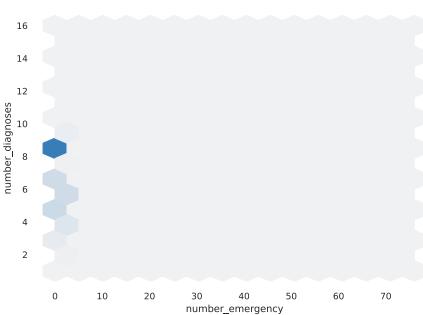


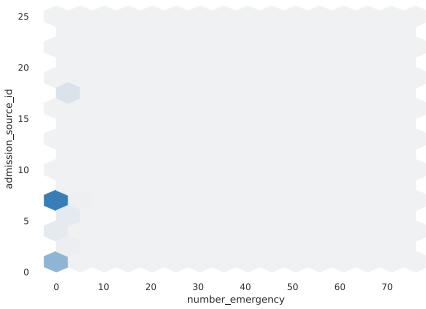
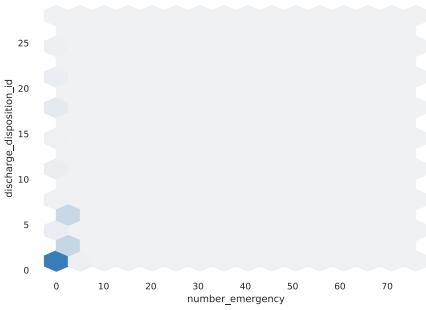
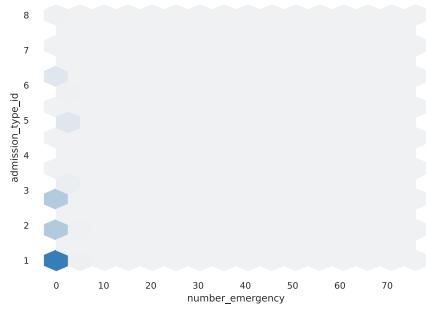
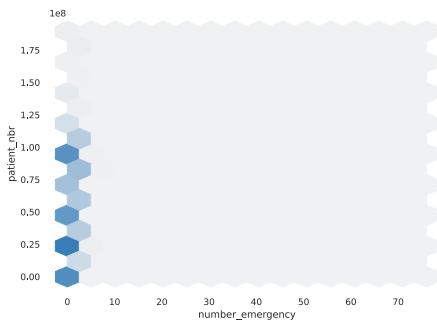


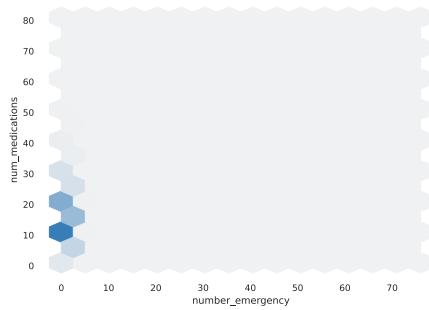
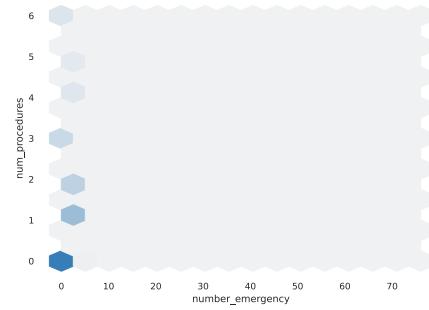
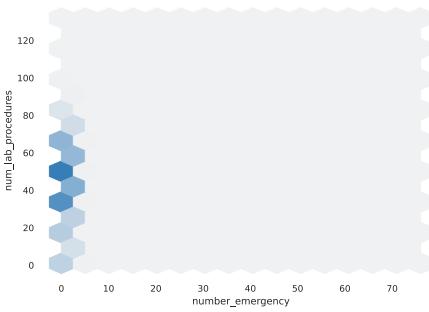
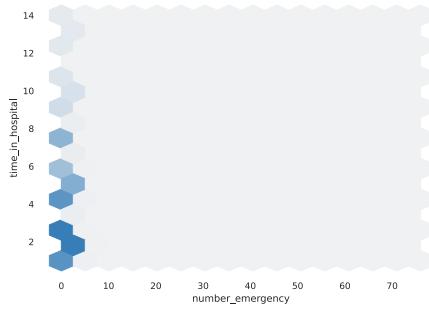


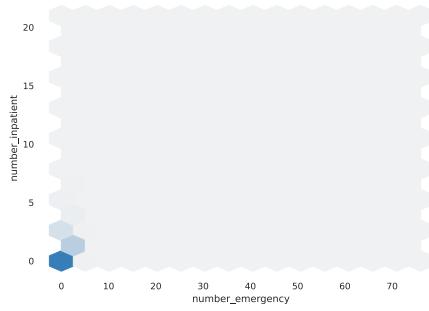
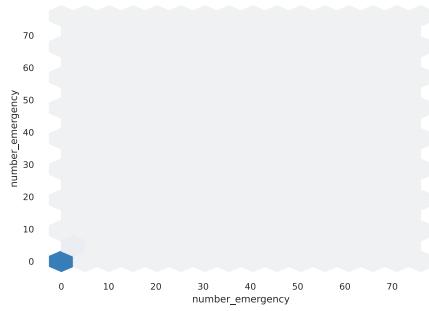
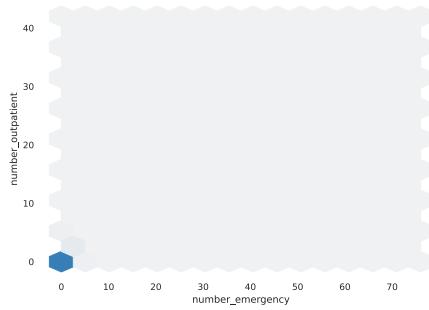


number_diagnoses
encounter_id
patient_nbr
admission_type_id
discharge_disposition_id
admission_source_id
time_in_hospital
num_lab_procedures
num_procedures
num_medications
number_outpatient
number_emergency
number_inpatient

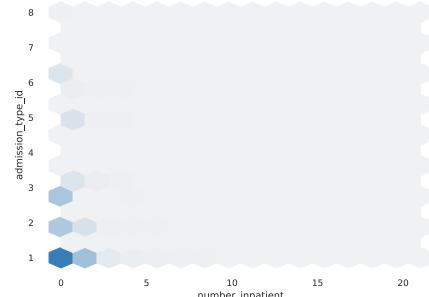
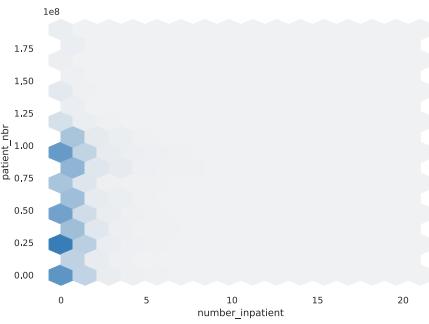
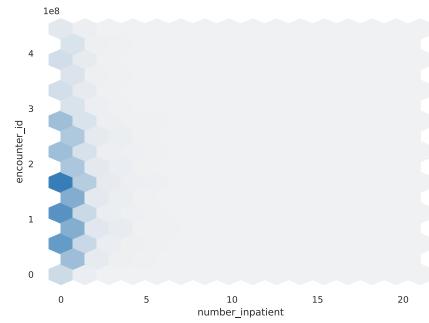
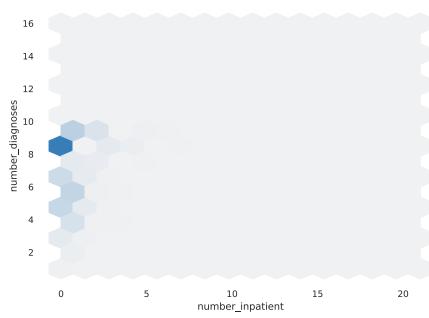


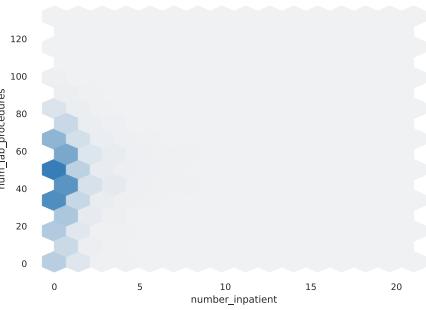
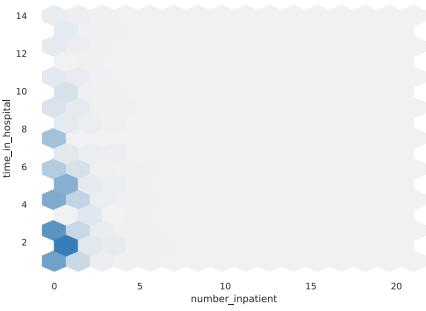
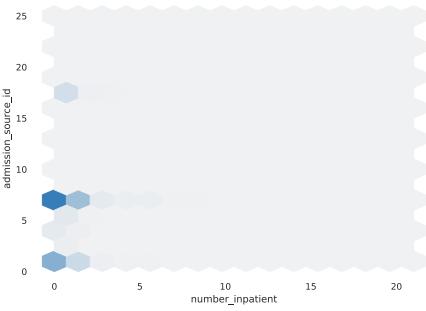
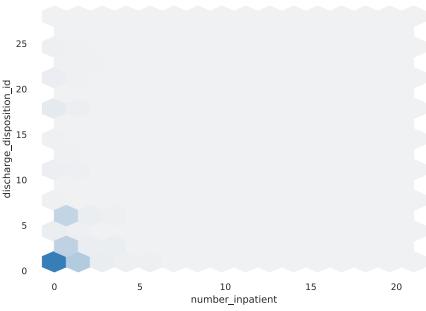


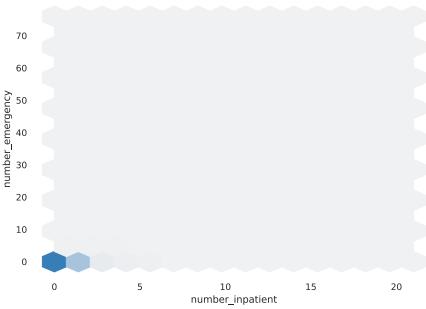
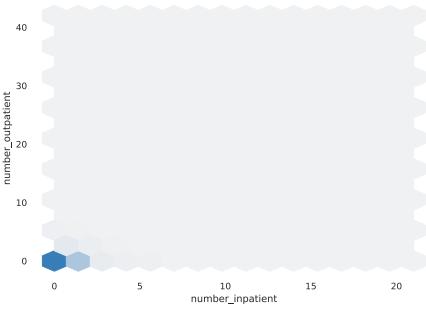
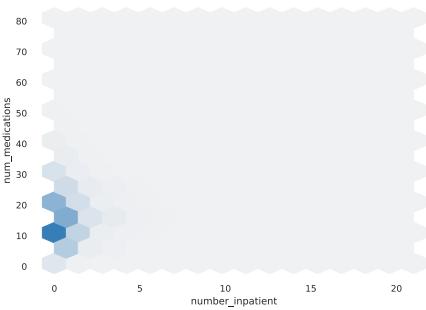
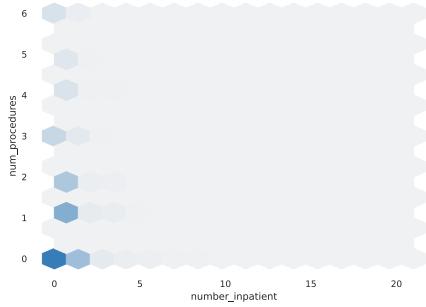


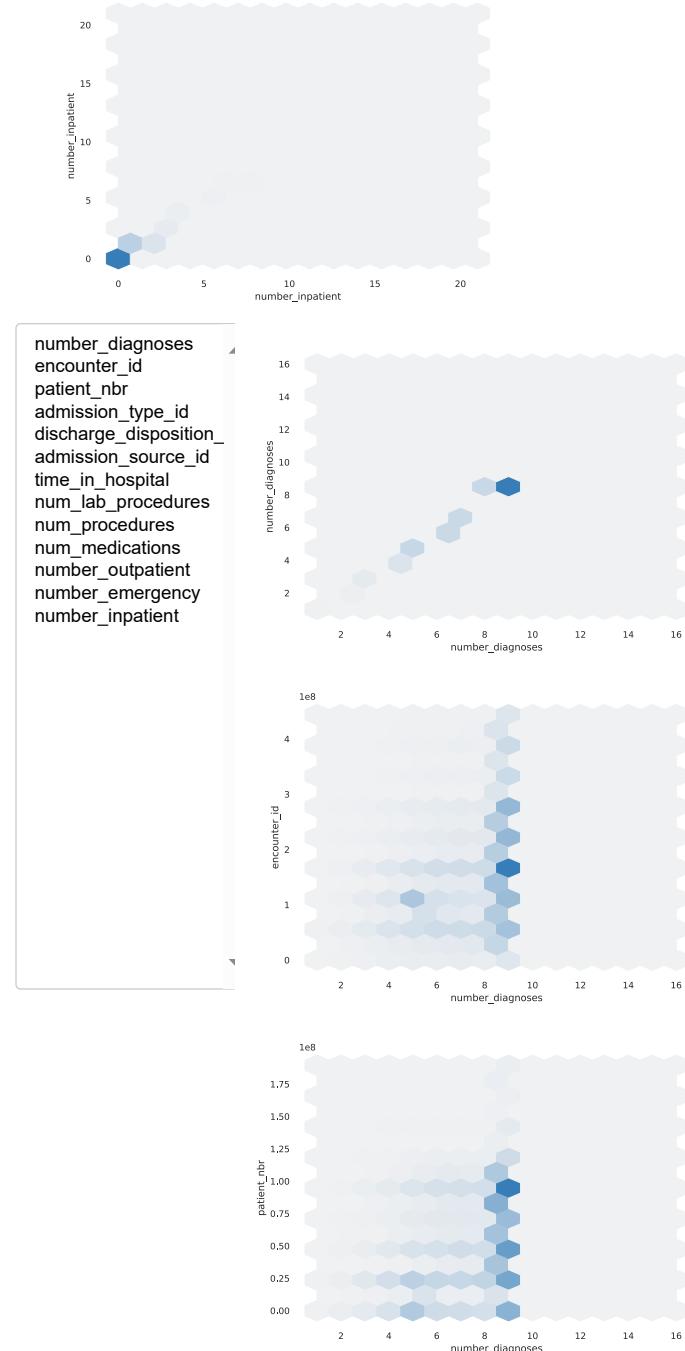


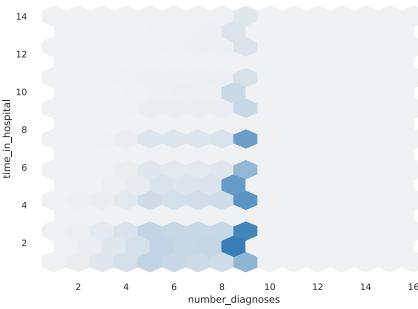
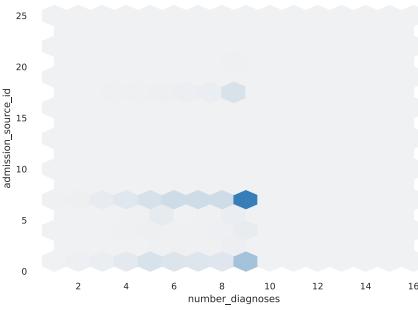
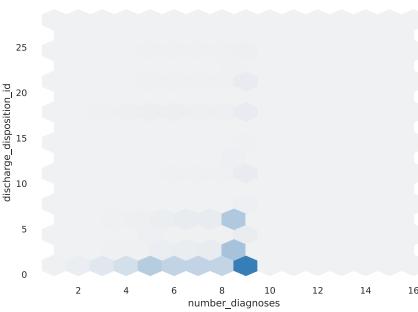
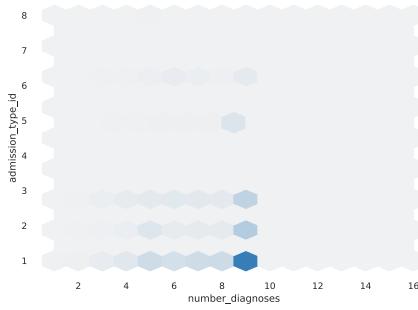
number_diagnoses
encounter_id
patient_nbr
admission_type_id
discharge_disposition_id
admission_source_id
time_in_hospital
num_lab_procedures
num_procedures
num_medications
number_outpatient
number_emergency
number_inpatient

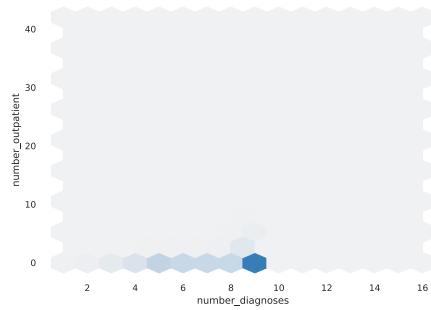
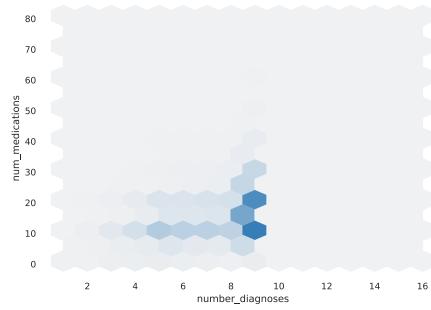
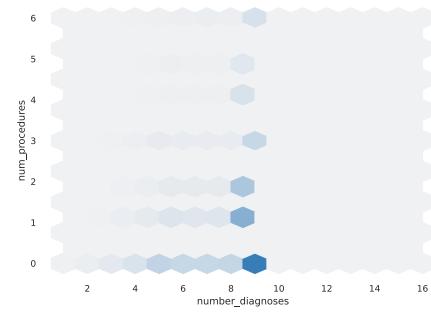
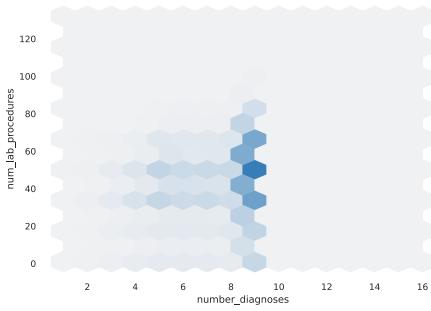


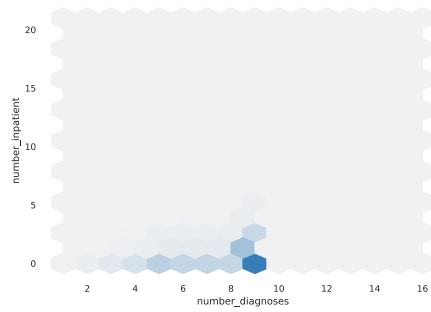
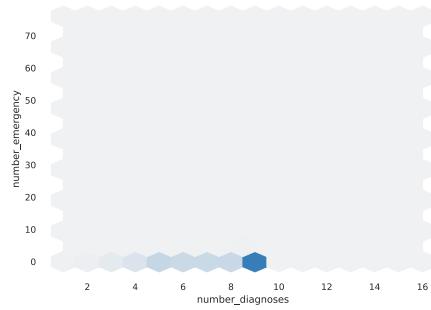




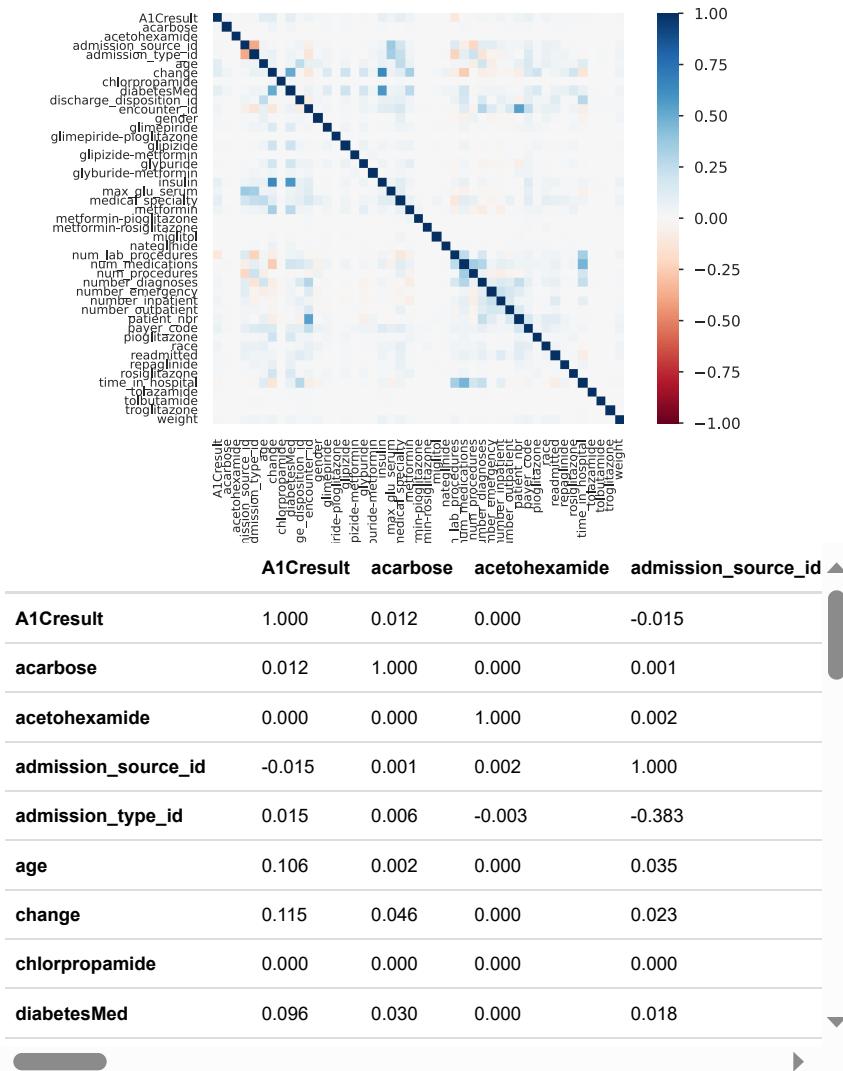




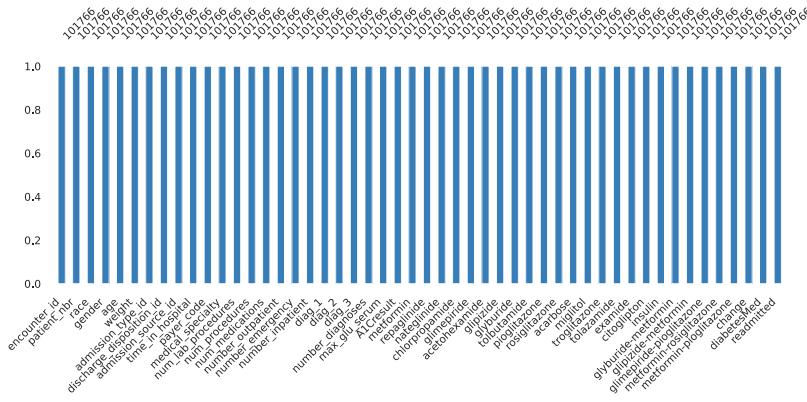




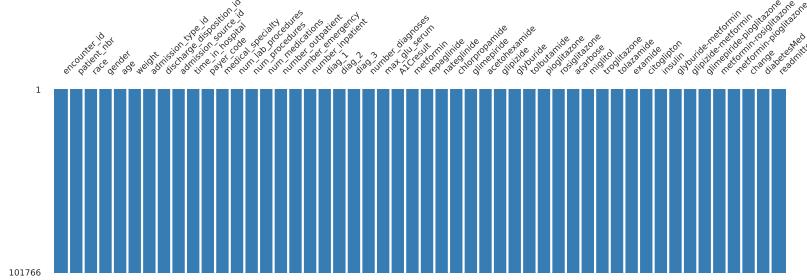
Correlations



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.

Sample

	encounter_id	patient_nbr	race	gender	age	weight	admission_type_id
0	2278392	8222157	Caucasian	Female	[0-10)	Missing	6
1	149190	55629189	Caucasian	Female	[10-20)	Missing	1
2	64410	86047875	AfricanAmerican	Female	[20-30)	Missing	1
3	500364	82442376	Caucasian	Male	[30-40)	Missing	1
4	16680	42519267	Caucasian	Male	[40-50)	Missing	1
5	35754	82637451	Caucasian	Male	[50-60)	Missing	2

	encounter_id	patient_nbr	race	gender	age	weight	admission_type_ic
6	55842	84259809	Caucasian	Male	[60-70)	Missing	3
7	63768	114882984	Caucasian	Male	[70-80)	Missing	1
8	12522	48330783	Caucasian	Female	[80-90)	Missing	2
9	15738	63555939	Caucasian	Female	[90-100)	Missing	3

	encounter_id	patient_nbr	race	gender	age	weight	admission_ty
101756	443842070	140199494	Other	Female	[60-70)	Missing	1
101757	443842136	181593374	Caucasian	Female	[70-80)	Missing	1
101758	443842340	120975314	Caucasian	Female	[80-90)	Missing	1
101759	443842778	86472243	Caucasian	Male	[80-90)	Missing	1
101760	443847176	50375628	AfricanAmerican	Female	[60-70)	Missing	1
101761	443847548	100162476	AfricanAmerican	Male	[70-80)	Missing	1
101762	443847782	74694222	AfricanAmerican	Female	[80-90)	Missing	1
101763	443854148	41088789	Caucasian	Male	[70-80)	Missing	1
101764	443857166	31693671	Caucasian	Female	[80-90)	Missing	2
101765	443867222	175429310	Caucasian	Male	[70-80)	Missing	1



Report generated by YData (https://ydata.ai/?utm_source=opensource&utm_medium=pandasprofiling&utm_campaign=report).