

# Overview

## Dataset statistics

<b>Number of variables</b>	30
<b>Number of observations</b>	2083363
<b>Missing cells</b>	3373
<b>Missing cells (%)</b>	< 0.1%
<b>Duplicate rows</b>	174167
<b>Duplicate rows (%)</b>	8.4%
<b>Total size in memory</b>	476.8 MiB
<b>Average record size in memory</b>	240.0 B

## Variable types

<b>Categorical</b>	24
<b>Numeric</b>	6

## Warnings

modalidade_de_apoio has constant value "REEMBOLSÁVEL"	Constant
forma_de_apoio has constant value "INDIRETA"	Constant
Dataset has 174167 (8.4%) duplicate rows	Duplicates
cliente has a high cardinality: 397313 distinct values	High cardinality
cpf_cnpj has a high cardinality: 98152 distinct values	High cardinality
municipio has a high cardinality: 5140 distinct values	High cardinality
data_da_contratacao has a high cardinality: 5528 distinct values	High cardinality
instrumento_financeiro has a high cardinality: 80 distinct values	High cardinality
subsetor_cnae_codigo has a high cardinality: 1640 distinct values	High cardinality
subsetor_cnae_nome has a high cardinality: 1604 distinct values	High cardinality
instituicao_financeira_credenciada has a high cardinality: 127 distinct values	High cardinality
cnpj_do_agente_financeiro has a high cardinality: 127 distinct values	High cardinality
valor_da_operacao_em_reais is highly correlated with valor_desembolsado_reais	High correlation
valor_desembolsado_reais is highly correlated with valor_da_operacao_em_reais	High correlation
custo_financeiro is highly correlated with modalidade_de_apoio and 1 other fields (modalidade_de_apoio, forma_de_apoio)	High correlation
subsetor_cnae_agrupado is highly correlated with setor_cnae and 4 other fields (setor_cnae, subsetor_bndes, modalidade_de_apoio, forma_de_apoio, setor_bndes)	High correlation
uf is highly correlated with modalidade_de_apoio and 1 other fields (modalidade_de_apoio, forma_de_apoio)	High correlation
area_operacional is highly correlated with modalidade_de_apoio and 1 other fields (modalidade_de_apoio, forma_de_apoio)	High correlation
setor_cnae is highly correlated with subsetor_cnae_agrupado and 3 other fields (subsetor_cnae_agrupado, subsetor_bndes, modalidade_de_apoio, forma_de_apoio)	High correlation
subsetor_bndes is highly correlated with subsetor_cnae_agrupado and 4 other fields (subsetor_cnae_agrupado, setor_cnae, modalidade_de_apoio, forma_de_apoio, setor_bndes)	High correlation
porte_do_cliente is highly correlated with modalidade_de_apoio and 1 other fields (modalidade_de_apoio, forma_de_apoio)	High correlation
inovacao is highly correlated with modalidade_de_apoio and 1 other fields (modalidade_de_apoio, forma_de_apoio)	High correlation

modalidade\_de\_apoio is highly correlated with custo\_financeiro and 14 other fields  
(custo\_financeiro, subsetor\_cnae\_agrupado, uf, area\_operacional, setor\_cnae,  
subsetor\_bndes, porte\_do\_cliente, inovacao, instrumento\_financeiro, forma\_de\_apoio,  
setor\_bndes, fonte\_de\_recurso\_desembolsos, produto, situacao\_da\_operacao,  
natureza\_do\_cliente)

High correlation

instrumento\_financeiro is highly correlated with modalidade\_de\_apoio and 1 other  
fields (modalidade\_de\_apoio, forma\_de\_apoio)

High correlation

forma\_de\_apoio is highly correlated with custo\_financeiro and 14 other fields  
(custo\_financeiro, subsetor\_cnae\_agrupado, uf, area\_operacional, setor\_cnae,  
subsetor\_bndes, porte\_do\_cliente, inovacao, modalidade\_de\_apoio, instrumento\_financeiro,  
setor\_bndes, fonte\_de\_recurso\_desembolsos, produto, situacao\_da\_operacao,  
natureza\_do\_cliente)

High correlation

setor\_bndes is highly correlated with subsetor\_cnae\_agrupado and 3 other fields  
(subsetor\_cnae\_agrupado, subsetor\_bndes, modalidade\_de\_apoio, forma\_de\_apoio)

High correlation

fonte\_de\_recurso\_desembolsos is highly correlated with modalidade\_de\_apoio and 1  
other fields (modalidade\_de\_apoio, forma\_de\_apoio)

High correlation

produto is highly correlated with modalidade\_de\_apoio and 1 other fields  
(modalidade\_de\_apoio, forma\_de\_apoio)

High correlation

situacao\_da\_operacao is highly correlated with modalidade\_de\_apoio and 1 other fields  
(modalidade\_de\_apoio, forma\_de\_apoio)

High correlation

natureza\_do\_cliente is highly correlated with modalidade\_de\_apoio and 1 other fields  
(modalidade\_de\_apoio, forma\_de\_apoio)

High correlation

valor\_da\_operacao\_em\_reais is highly skewed ( $y_1 = 253.0599823$ )

Skewed

valor\_desembolsado\_reais is highly skewed ( $y_1 = 256.9119925$ )

Skewed

prazo\_carencia\_meses has 309940 (14.9%) zeros

Zeros

## Reproduction

# Variables

## cliente

Categorical

HIGH

CARDINALITY

Distinct	397313
Distinct (%)	19.1%
Missing	0
Missing (%)	0.0%
Memory size	15.9 MiB

## Length

<b>Max length</b>	55
<b>Median length</b>	30
<b>Mean length</b>	31.99611158
<b>Min length</b>	3

## Characters and Unicode

<b>Total characters</b>	66659515	
<b>Distinct characters</b>	80	
<b>Distinct categories</b>	15 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

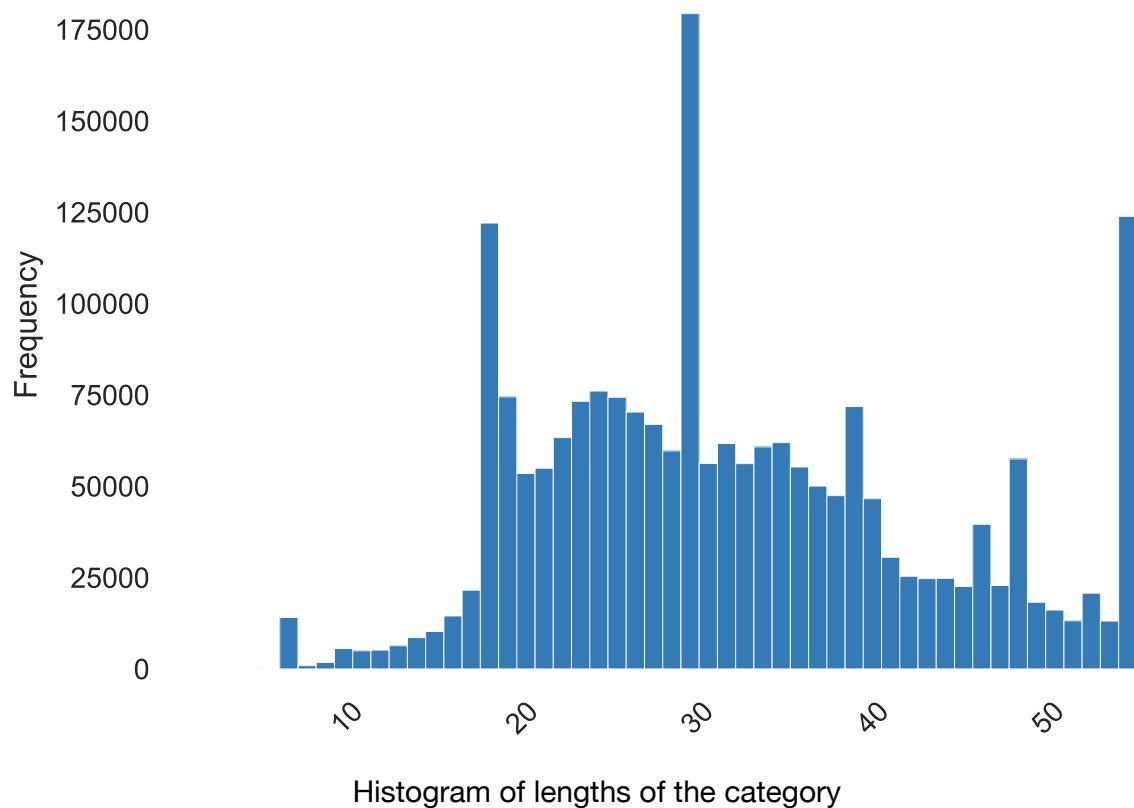
## Unique

<b>Unique</b>	193235	
<b>Unique (%)</b>	9.3%	

## Sample

<b>1st row</b>	BANCO COOPERATIVO SICOOB S.A.
<b>2nd row</b>	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL
<b>3rd row</b>	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL
<b>4th row</b>	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL
<b>5th row</b>	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL

Value	Count	Frequency (%)
BANCO DO BRASIL SA	87942	4.2%
BANCO COOPERATIVO SICREDI S.A.	42745	2.1%
BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL	39286	1.9%
COOPERATIVA CENTRAL DE CREDITO RURAL COM INT (01401771)	32178	1.5%
BANCO DO ESTADO DO RIO GRANDE DO SUL SA	24866	1.2%
BANCO BRADESCO S.A.	18265	0.9%
BANCO DE DESENVOLVIMENTO DO ESPIRITO SANTO S/A	16849	0.8%
INDUSTRIAS ROMI S A	14754	0.7%
COOPERATIVA CENTRAL DE CREDITO RURAL COM INTERACAO SOLI	14035	0.7%
BANCO SANTANDER (BRASIL) S.A.	12442	0.6%
Other values (397303)	1780001	85.4%



Value	Count	Frequency (%)
Itda	1239113	11.8%
de	559896	5.3%
e	454301	4.3%
me	274516	2.6%
transportes	271098	2.6%
	264949	2.5%
do	258251	2.5%
banco	256112	2.4%
comercio	176286	1.7%
s.a	143335	1.4%
Other values (161478)	6586224	62.8%

## Most occurring characters

Value	Count	Frequency (%)
	8534230	12.8%
A	7651860	11.5%
E	5549057	8.3%
O	5281545	7.9%
R	4727129	7.1%
T	4328709	6.5%
S	4145484	6.2%
I	4128240	6.2%
D	3454519	5.2%
L	3186999	4.8%
Other values (70)	15671743	23.5%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	56650708	85.0%
Space Separator	8534230	12.8%
Other Punctuation	711841	1.1%
Decimal Number	426992	0.6%
Dash Punctuation	230696	0.3%
Open Punctuation	52187	0.1%
Close Punctuation	52186	0.1%
Math Symbol	242	< 0.1%
Modifier Symbol	151	< 0.1%
Lowercase Letter	130	< 0.1%
Other values (5)	152	< 0.1%

## Most frequent character per category

Value	Count	Frequency (%)
A	7651860	13.5%
E	5549057	9.8%
O	5281545	9.3%
R	4727129	8.3%
T	4328709	7.6%
S	4145484	7.3%
I	4128240	7.3%
D	3454519	6.1%
L	3186999	5.6%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
C	2897057	5.1%
Other values (18)	11300109	19.9%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
.	489366	68.7%
/	112972	15.9%
&	79893	11.2%
,	25700	3.6%
'	2596	0.4%
?	539	0.1%
%	515	0.1%
:	100	< 0.1%
"	60	< 0.1%
!	35	< 0.1%
Other values (4)	65	< 0.1%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
á	32	24.6%
ç	28	21.5%
a	25	19.2%
é	24	18.5%
o	6	4.6%
í	5	3.8%
ô	4	3.1%
e	2	1.5%
û	2	1.5%

Value	Count	Frequency (%)
s	1	0.8%

Value	Count	Frequency (%)
1	113794	26.7%
0	93592	21.9%
7	80649	18.9%
4	46230	10.8%
2	19147	4.5%
9	16079	3.8%
3	15656	3.7%
8	14408	3.4%
6	13791	3.2%
5	13646	3.2%

Value	Count	Frequency (%)
+	200	82.6%
=	32	13.2%
~	7	2.9%
¬	3	1.2%

Value	Count	Frequency (%)
(	52186	> 99.9%
[	1	< 0.1%

Value	Count	Frequency (%)
)	52185	> 99.9%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
]	1	< 0.1%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
	16	72.7%
	6	27.3%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
	8534230	100.0%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
-	230696	100.0%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
o	50	100.0%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
:	57	100.0%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
-	18	100.0%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
'	151	100.0%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
\$	5	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	56650888	85.0%
Common	10008627	15.0%

## Most frequent character per script

Value	Count	Frequency (%)
A	7651860	13.5%
E	5549057	9.8%
O	5281545	9.3%
R	4727129	8.3%
T	4328709	7.6%
S	4145484	7.3%
I	4128240	7.3%
D	3454519	6.1%
L	3186999	5.6%
C	2897057	5.1%
Other values (30)	11300289	19.9%

Value	Count	Frequency (%)
.	8534230	85.3%
-	489366	4.9%
1	230696	2.3%
/	113794	1.1%
0	112972	1.1%
	93592	0.9%

Value	Count	Frequency (%)
7	80649	0.8%
&	79893	0.8%
(	52186	0.5%
)	52185	0.5%
Other values (30)	169064	1.7%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	66658874	> 99.9%
None	641	< 0.1%

## Most frequent character per block

Value	Count	Frequency (%)
A	8534230	12.8%
E	7651860	11.5%
O	5549057	8.3%
R	5281545	7.9%
T	4727129	7.1%
S	4328709	6.5%
I	4145484	6.2%
D	4128240	6.2%
L	3454519	5.2%
Other values (57)	3186999	4.8%
	15671102	23.5%

Value	Count	Frequency (%)
Ç	403	62.9%
:	57	8.9%
º	50	7.8%
á	32	5.0%
ç	28	4.4%
é	24	3.7%

Value	Count	Frequency (%)
i	22	3.4%
É	10	1.6%
í	5	0.8%
ô	4	0.6%
Other values (3)	6	0.9%





**cpf\_cnpj**

Categorical

HIGH

CARDINALITY

<b>Distinct</b>	98152
<b>Distinct (%)</b>	4.7%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	15.9 MiB

## Length

<b>Max length</b>	18
<b>Median length</b>	18
<b>Mean length</b>	18
<b>Min length</b>	18

## Characters and Unicode

<b>Total characters</b>	37500534	
<b>Distinct characters</b>	14	
<b>Distinct categories</b>	3 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	1 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

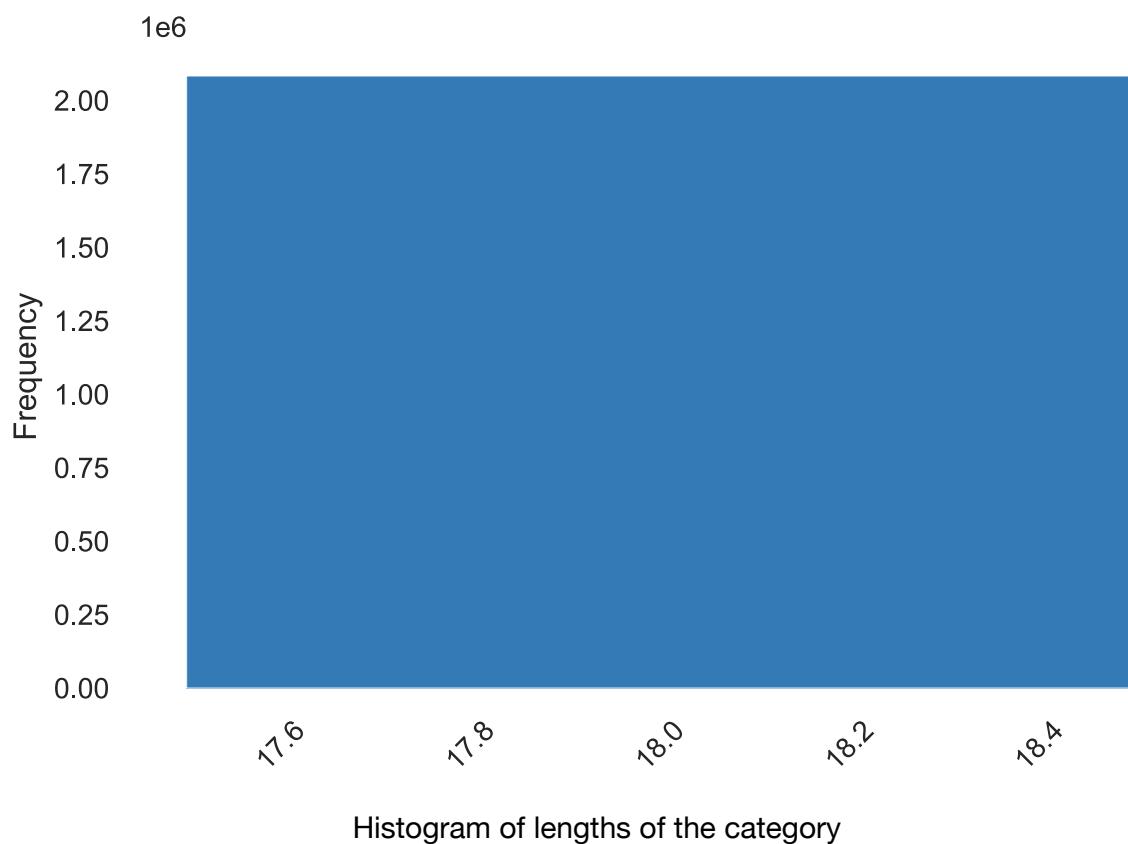
## Unique

<b>Unique</b>	3573	?
<b>Unique (%)</b>	0.2%	

## Sample

<b>1st row</b>	**.*38.232/0001-**
<b>2nd row</b>	**.*16.560/0001-**
<b>3rd row</b>	**.*16.560/0001-**
<b>4th row</b>	**.*16.560/0001-**
<b>5th row</b>	**.*16.560/0001-**

Value	Count	Frequency (%)
**.*00.000/0001-**	87954	4.2%
**.*81.521/0001-**	42746	2.1%
**.*16.560/0001-**	39319	1.9%
**.*01.771/0001-**	32188	1.5%
**.*02.067/0001-**	24879	1.2%
**.*46.948/0001-**	18279	0.9%
**.*45.829/0001-**	16858	0.8%
**.*20.428/0001-**	14755	0.7%
**.*02.627/0001-**	14060	0.7%
**.*00.888/0001-**	12450	0.6%
Other values (98142)	1779875	85.4%



<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
00.000/0001	87954	4.2%
81.521/0001	42746	2.1%
16.560/0001	39319	1.9%
01.771/0001	32188	1.5%
02.067/0001	24879	1.2%
46.948/0001	18279	0.9%
45.829/0001	16858	0.8%
20.428/0001	14755	0.7%
02.627/0001	14060	0.7%
00.888/0001	12450	0.6%
Other values (98142)	1779875	85.4%

## Most occurring characters

Value	Count	Frequency (%)
*	10416815	27.8%
0	7758109	20.7%
.	4166726	11.1%
1	3154875	8.4%
/	2083363	5.6%
-	2083363	5.6%
2	1043055	2.8%
8	1033216	2.8%
6	1000455	2.7%
5	992650	2.6%
Other values (4)	3767907	10.0%

## Most occurring categories

Value	Count	Frequency (%)
Decimal Number	18750267	50.0%
Other Punctuation	16666904	44.4%
Dash Punctuation	2083363	5.6%

## Most frequent character per category

Value	Count	Frequency (%)
0	7758109	41.4%
1	3154875	16.8%
2	1043055	5.6%
8	1033216	5.5%
6	1000455	5.3%
5	992650	5.3%
7	979515	5.2%
4	959005	5.1%
3	918769	4.9%
9	910618	4.9%

Value	Count	Frequency (%)
*	10416815	62.5%
.	4166726	25.0%
/	2083363	12.5%

Value	Count	Frequency (%)
-	2083363	100.0%



## Most occurring scripts

Value	Count	Frequency (%)
Common	37500534	100.0%

## Most frequent character per script

Value	Count	Frequency (%)
*	10416815	27.8%
0	7758109	20.7%
.	4166726	11.1%
1	3154875	8.4%
/	2083363	5.6%
-	2083363	5.6%
2	1043055	2.8%
8	1033216	2.8%
6	1000455	2.7%
5	992650	2.6%
Other values (4)	3767907	10.0%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	37500534	100.0%

## Most frequent character per block

Value	Count	Frequency (%)
*	10416815	27.8%
0	7758109	20.7%
.	4166726	11.1%
1	3154875	8.4%
/	2083363	5.6%
-	2083363	5.6%
2	1043055	2.8%
8	1033216	2.8%
6	1000455	2.7%
5	992650	2.6%
Other values (4)	3767907	10.0%





**uf**

Categorical

HIGH

CORRELATION (This variable has a high correlation with 2 fields: modalidade\_de\_apoio, forma\_de\_apoio).

Distinct	27
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	15.9 MiB

## Length

<b>Max length</b>	3
<b>Median length</b>	3
<b>Mean length</b>	3
<b>Min length</b>	3

## Characters and Unicode

<b>Total characters</b>	6250089	
<b>Distinct characters</b>	18	
<b>Distinct categories</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

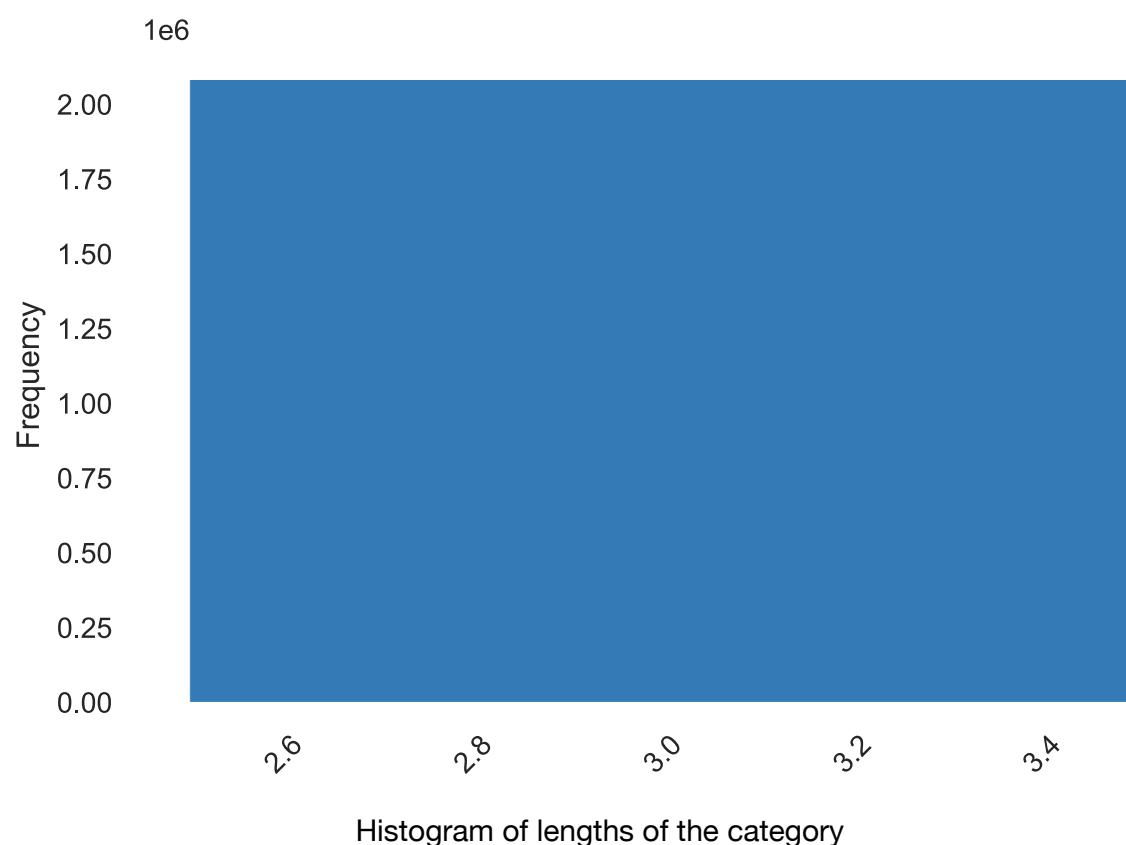
## Unique

<b>Unique</b>	0	?
<b>Unique (%)</b>	0.0%	

## Sample

<b>1st row</b>	SP
<b>2nd row</b>	RS
<b>3rd row</b>	RS
<b>4th row</b>	RS
<b>5th row</b>	RS

Value	Count	Frequency (%)
SP	434734	20.9%
RS	288488	13.8%
PR	275219	13.2%
MG	232436	11.2%
SC	216139	10.4%
RJ	79365	3.8%
GO	73032	3.5%
BA	72898	3.5%
ES	62616	3.0%
MT	53894	2.6%
Other values (17)	294542	14.1%



Value	Count	Frequency (%)
sp	434734	20.9%
rs	288488	13.8%
pr	275219	13.2%
mg	232436	11.2%
sc	216139	10.4%
rj	79365	3.8%
go	73032	3.5%
ba	72898	3.5%
es	62616	3.0%
mt	53894	2.6%
Other values (17)	294542	14.1%

## Most occurring characters

Value	Count	Frequency (%)
	2083363	33.3%
S	1042852	16.7%
P	814644	13.0%
R	679619	10.9%
M	354473	5.7%
G	305468	4.9%
C	253913	4.1%
A	155054	2.5%
E	154206	2.5%
O	109065	1.7%
Other values (8)	297432	4.8%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	4166726	66.7%
Space Separator	2083363	33.3%

## Most frequent character per category

Value	Count	Frequency (%)
S	1042852	25.0%
P	814644	19.6%
R	679619	16.3%
M	354473	8.5%
G	305468	7.3%
C	253913	6.1%
A	155054	3.7%
E	154206	3.7%
O	109065	2.6%
B	86104	2.1%
Other values (7)	211328	5.1%

Value	Count	Frequency (%)
	2083363	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	4166726	66.7%
Common	2083363	33.3%

## Most frequent character per script

Value	Count	Frequency (%)
S	1042852	25.0%
P	814644	19.6%
R	679619	16.3%
M	354473	8.5%
G	305468	7.3%
C	253913	6.1%
A	155054	3.7%
E	154206	3.7%
O	109065	2.6%
B	86104	2.1%
Other values (7)	211328	5.1%

Value	Count	Frequency (%)
	2083363	100.0%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	6250089	100.0%

## Most frequent character per block

Value	Count	Frequency (%)
	2083363	33.3%
S	1042852	16.7%
P	814644	13.0%
R	679619	10.9%
M	354473	5.7%
G	305468	4.9%
C	253913	4.1%
A	155054	2.5%
E	154206	2.5%
O	109065	1.7%
Other values (8)	297432	4.8%





## municipio

Categorical

HIGH

CARDINALITY

Distinct	5140
Distinct (%)	0.2%
Missing	0
Missing (%)	0.0%
Memory size	15.9 MiB

## Length

<b>Max length</b>	32
<b>Median length</b>	9
<b>Mean length</b>	10.50826764
<b>Min length</b>	3

## Characters and Unicode

<b>Total characters</b>	21892536	
<b>Distinct characters</b>	29	
<b>Distinct categories</b>	4 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

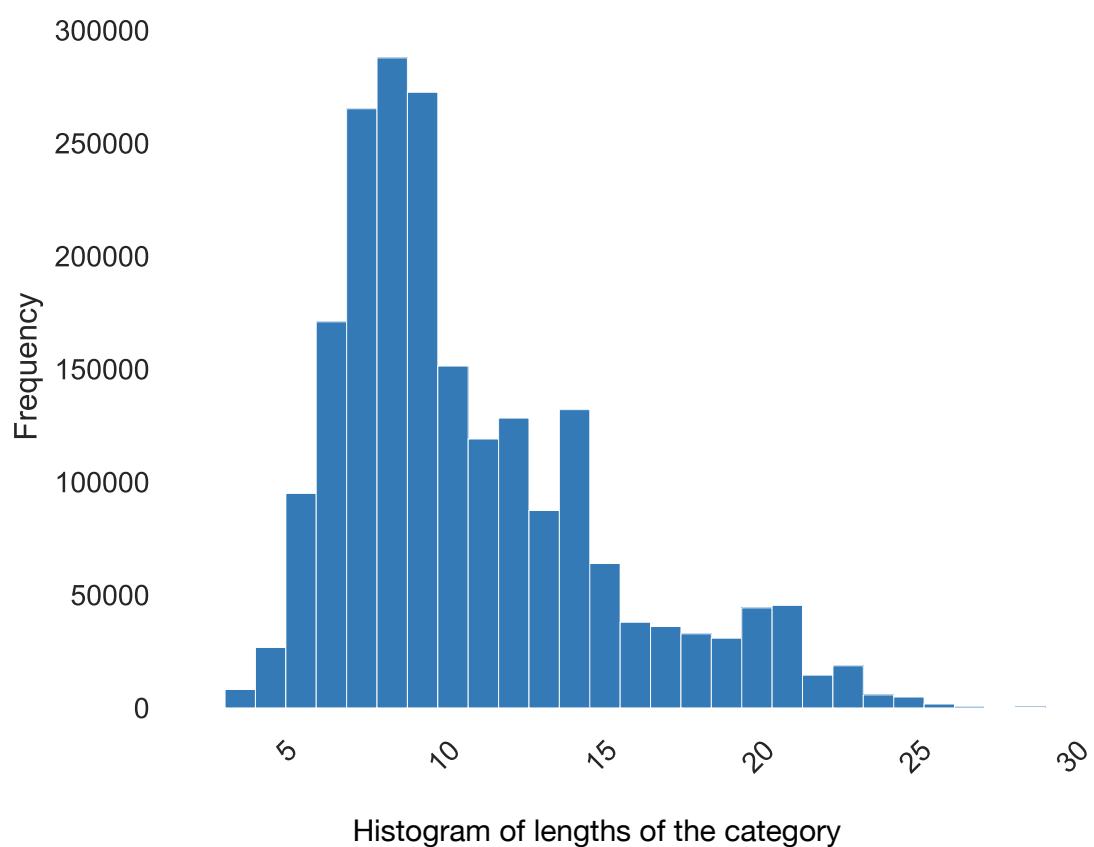
## Unique

<b>Unique</b>	101	?
<b>Unique (%)</b>	< 0.1%	

## Sample

<b>1st row</b>	PEDREGULHO
<b>2nd row</b>	TRES DE MAIO
<b>3rd row</b>	ERECHIM
<b>4th row</b>	ERECHIM
<b>5th row</b>	HUMAITA

Value	Count	Frequency (%)
SAO PAULO	72463	3.5%
BELO HORIZONTE	32333	1.6%
CURITIBA	31242	1.5%
RIO DE JANEIRO	27641	1.3%
GOIANIA	16161	0.8%
MARINGA	15910	0.8%
FORTALEZA	15455	0.7%
RECIFE	15342	0.7%
PORTO ALEGRE	14995	0.7%
SALVADOR	14552	0.7%
Other values (5130)	1827269	87.7%



Value	Count	Frequency (%)
sao	192539	5.5%
do	167345	4.8%
de	98125	2.8%
rio	74271	2.1%
paulo	74169	2.1%
sul	63163	1.8%
santa	37591	1.1%
nova	35451	1.0%
campo	35153	1.0%
belo	34540	1.0%
Other values (3865)	2661033	76.6%

## Most occurring characters

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
A	3635252	16.6%
O	2280591	10.4%
I	1717815	7.8%
R	1682212	7.7%
E	1448435	6.6%
	1390017	6.3%
S	1268368	5.8%
N	1117875	5.1%
U	881837	4.0%
T	839433	3.8%
Other values (19)	5630701	25.7%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	20480314	93.5%
Space Separator	1390017	6.3%
Other Punctuation	11469	0.1%
Dash Punctuation	10736	< 0.1%

## Most frequent character per category

Value	Count	Frequency (%)
A	3635252	17.7%
O	2280591	11.1%
I	1717815	8.4%
R	1682212	8.2%
E	1448435	7.1%
S	1268368	6.2%
N	1117875	5.5%
U	881837	4.3%
T	839433	4.1%
L	836799	4.1%
Other values (16)	4771697	23.3%

Value	Count	Frequency (%)
-	1390017	100.0%

Value	Count	Frequency (%)
-	10736	100.0%

**Value****Count Frequency (%)**

11469 100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	20480314	93.5%
Common	1412222	6.5%

## Most frequent character per script

Value	Count	Frequency (%)
A	3635252	17.7%
O	2280591	11.1%
I	1717815	8.4%
R	1682212	8.2%
E	1448435	7.1%
S	1268368	6.2%
N	1117875	5.5%
U	881837	4.3%
T	839433	4.1%
L	836799	4.1%
Other values (16)	4771697	23.3%

Value	Count	Frequency (%)
'	1390017	98.4%
-	11469	0.8%
-	10736	0.8%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	21892536	100.0%

## Most frequent character per block

Value	Count	Frequency (%)
A	3635252	16.6%
O	2280591	10.4%
I	1717815	7.8%
R	1682212	7.7%
E	1448435	6.6%
	1390017	6.3%
S	1268368	5.8%
N	1117875	5.1%
U	881837	4.0%
T	839433	3.8%
Other values (19)	5630701	25.7%

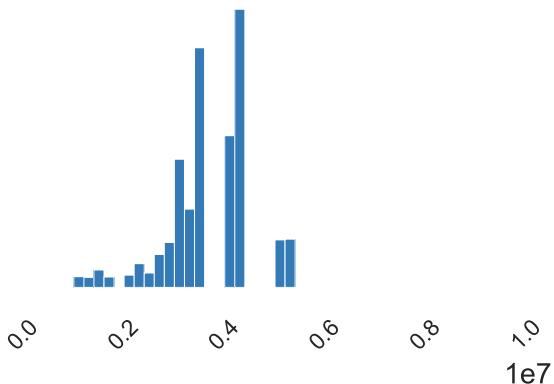




## municipio\_codigo

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

<b>Distinct</b>	5418
<b>Distinct (%)</b>	0.3%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	3664379.874
<b>Minimum</b>	0
<b>Maximum</b>	9999999
<b>Zeros</b>	26
<b>Zeros (%)</b>	< 0.1%
<b>Memory size</b>	15.9 MiB



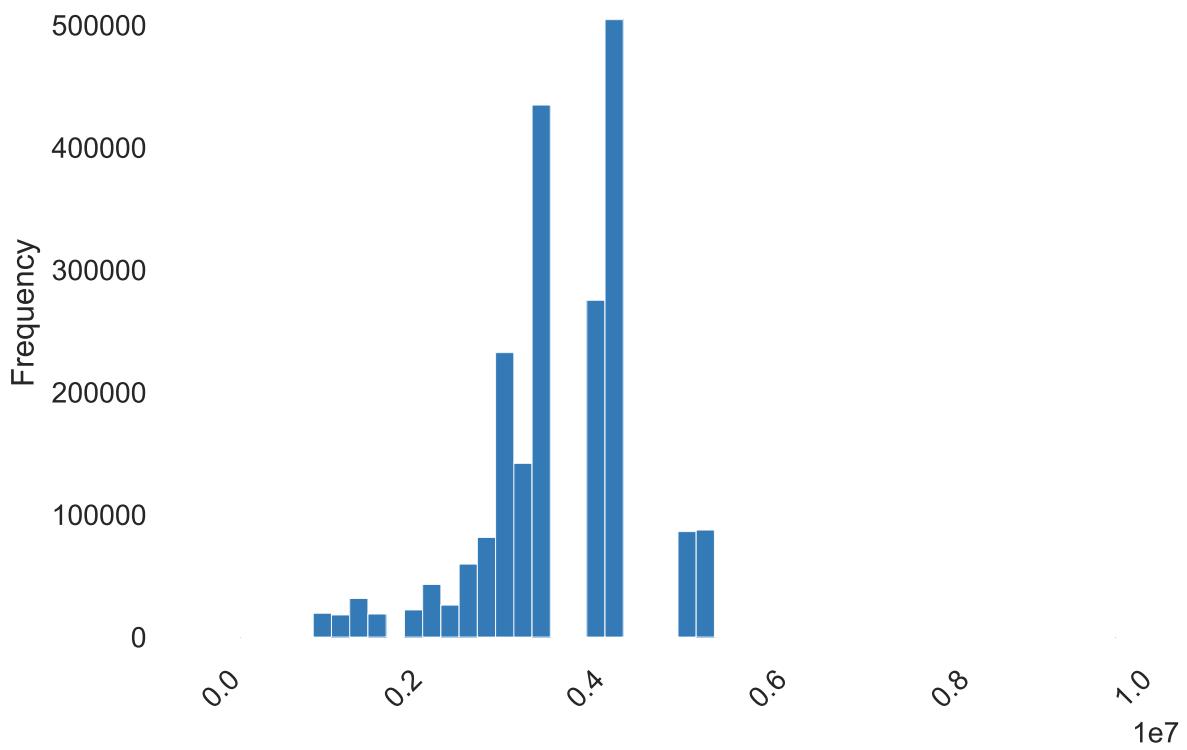
## Quantile statistics

<b>Minimum</b>	0
<b>5-th percentile</b>	2111300
<b>Q1</b>	3167004
<b>median</b>	3550308
<b>Q3</b>	4213609

95-th percentile	5107602
Maximum	9999999
Range	9999999
Interquartile range (IQR)	1046605

## Descriptive statistics

Standard deviation	838415.3134
Coefficient of variation (CV)	0.2288014186
Kurtosis	0.9026096602
Mean	3664379.874
Median Absolute Deviation (MAD)	573500
Skewness	-0.5651104449
Sum	$7.634233447 \times 10^{12}$
Variance	$7.029402377 \times 10^{11}$
Monotocity	Not monotonic



**Histogram with fixed size bins (bins=50)**

Value	Count	Frequency (%)
3550308	72463	3.5%
3106200	32333	1.6%
4106902	31242	1.5%
3304557	27641	1.3%
5208707	16161	0.8%
4115200	15910	0.8%
2304400	15455	0.7%
2611606	15342	0.7%
4314902	14995	0.7%
2927408	14552	0.7%
Other values (5408)	1827269	87.7%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
0	26	< 0.1%
1100015	228	< 0.1%
1100023	1570	0.1%
1100031	17	< 0.1%
1100049	956	< 0.1%

Value	Count	Frequency (%)
9999999	122	< 0.1%
5300108	14454	0.7%
5222302	34	< 0.1%
5222203	30	< 0.1%
5222054	401	< 0.1%





## **data\_da\_contratacao**

Categorical

HIGH

CARDINALITY

---

<b>Distinct</b>	5528
<b>Distinct (%)</b>	0.3%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	15.9 MiB

## Length

<b>Max length</b>	10
<b>Median length</b>	10
<b>Mean length</b>	10
<b>Min length</b>	10

## Characters and Unicode

<b>Total characters</b>	20833630	
<b>Distinct characters</b>	11	
<b>Distinct categories</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	1 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

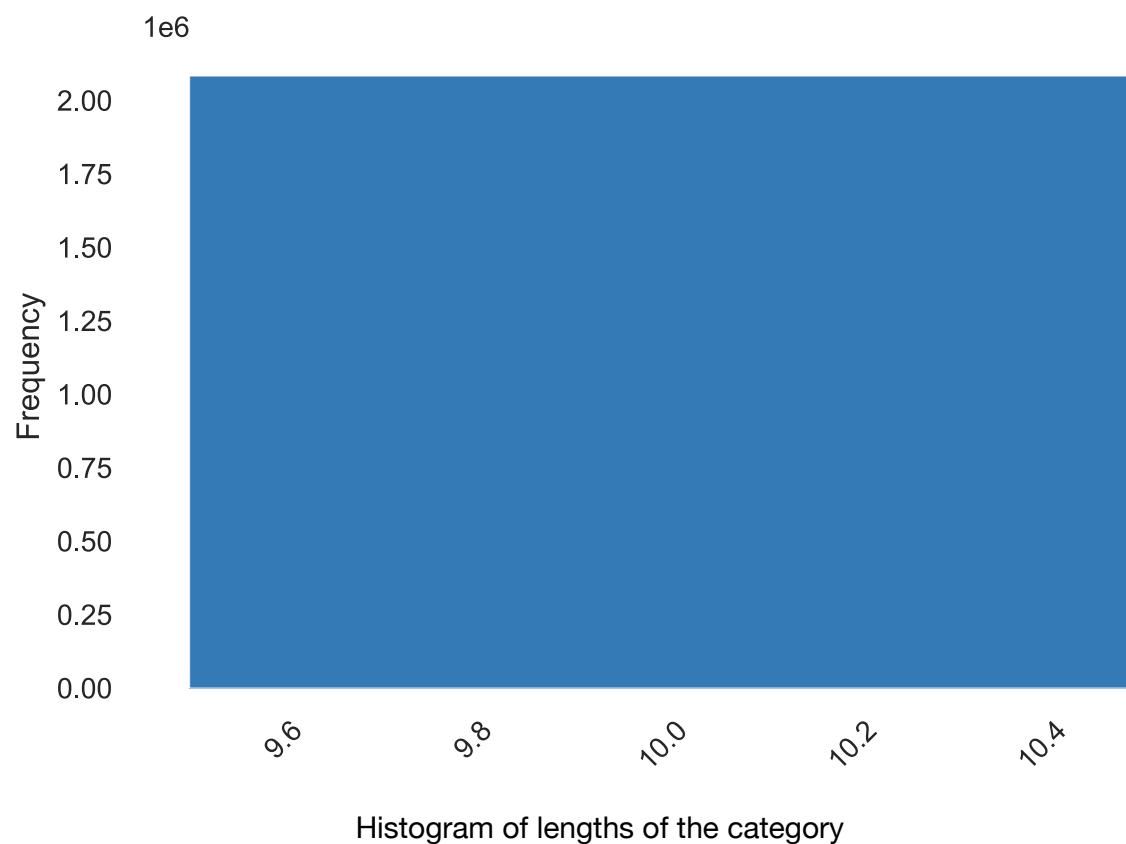
## Unique

<b>Unique</b>	293	?
<b>Unique (%)</b>	< 0.1%	

## Sample

<b>1st row</b>	2002-01-02
<b>2nd row</b>	2002-01-02
<b>3rd row</b>	2002-01-02
<b>4th row</b>	2002-01-02
<b>5th row</b>	2002-01-02

Value	Count	Frequency (%)
2004-06-30	8281	0.4%
2013-12-27	5280	0.3%
2013-12-30	5051	0.2%
2012-12-17	4478	0.2%
2005-06-20	4341	0.2%
2010-03-12	4063	0.2%
2010-06-02	3848	0.2%
2005-06-30	3778	0.2%
2011-10-19	3730	0.2%
2013-11-06	3633	0.2%
Other values (5518)	2036880	97.8%



<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
2004-06-30	8281	0.4%
2013-12-27	5280	0.3%
2013-12-30	5051	0.2%
2012-12-17	4478	0.2%
2005-06-20	4341	0.2%
2010-03-12	4063	0.2%
2010-06-02	3848	0.2%
2005-06-30	3778	0.2%
2011-10-19	3730	0.2%
2013-11-06	3633	0.2%
Other values (5518)	2036880	97.8%

## Most occurring characters

Value	Count	Frequency (%)
0	5529049	26.5%
-	4166726	20.0%
2	3643291	17.5%
1	3412098	16.4%
3	723404	3.5%
4	627673	3.0%
5	569495	2.7%
6	560446	2.7%
9	550369	2.6%
8	527330	2.5%

## Most occurring categories

Value	Count	Frequency (%)
Decimal Number	16666904	80.0%
Dash Punctuation	4166726	20.0%

## Most frequent character per category

Value	Count	Frequency (%)
0	5529049	33.2%
2	3643291	21.9%
1	3412098	20.5%
3	723404	4.3%
4	627673	3.8%
5	569495	3.4%
6	560446	3.4%
9	550369	3.3%
8	527330	3.2%
7	523749	3.1%

Value	Count	Frequency (%)
-	4166726	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Common	20833630	100.0%

## Most frequent character per script

Value	Count	Frequency (%)
0	5529049	26.5%
-	4166726	20.0%
2	3643291	17.5%
1	3412098	16.4%
3	723404	3.5%
4	627673	3.0%
5	569495	2.7%
6	560446	2.7%
9	550369	2.6%
8	527330	2.5%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	20833630	100.0%

## Most frequent character per block

Value	Count	Frequency (%)
0	5529049	26.5%
-	4166726	20.0%
2	3643291	17.5%
1	3412098	16.4%
3	723404	3.5%
4	627673	3.0%
5	569495	2.7%
6	560446	2.7%
9	550369	2.6%
8	527330	2.5%





## valor\_da\_operacao\_em\_reais

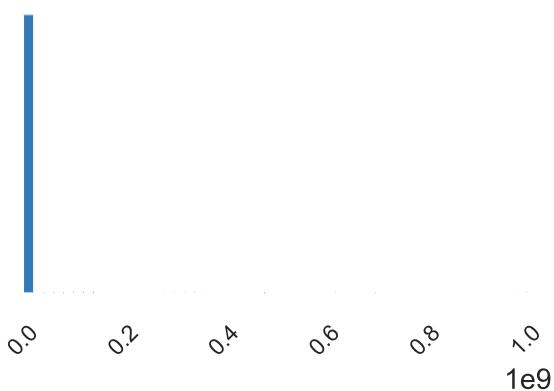
Real number ( $\mathbb{R}_{\geq 0}$ )

HIGH

CORRELATION (This variable has a high correlation with 1 fields: valor\_desembolsado\_reais).

SKEWED

Distinct	226554
Distinct (%)	10.9%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	273316.6764
Minimum	8
Maximum	1000000000
Zeros	0
Zeros (%)	0.0%
Memory size	15.9 MiB



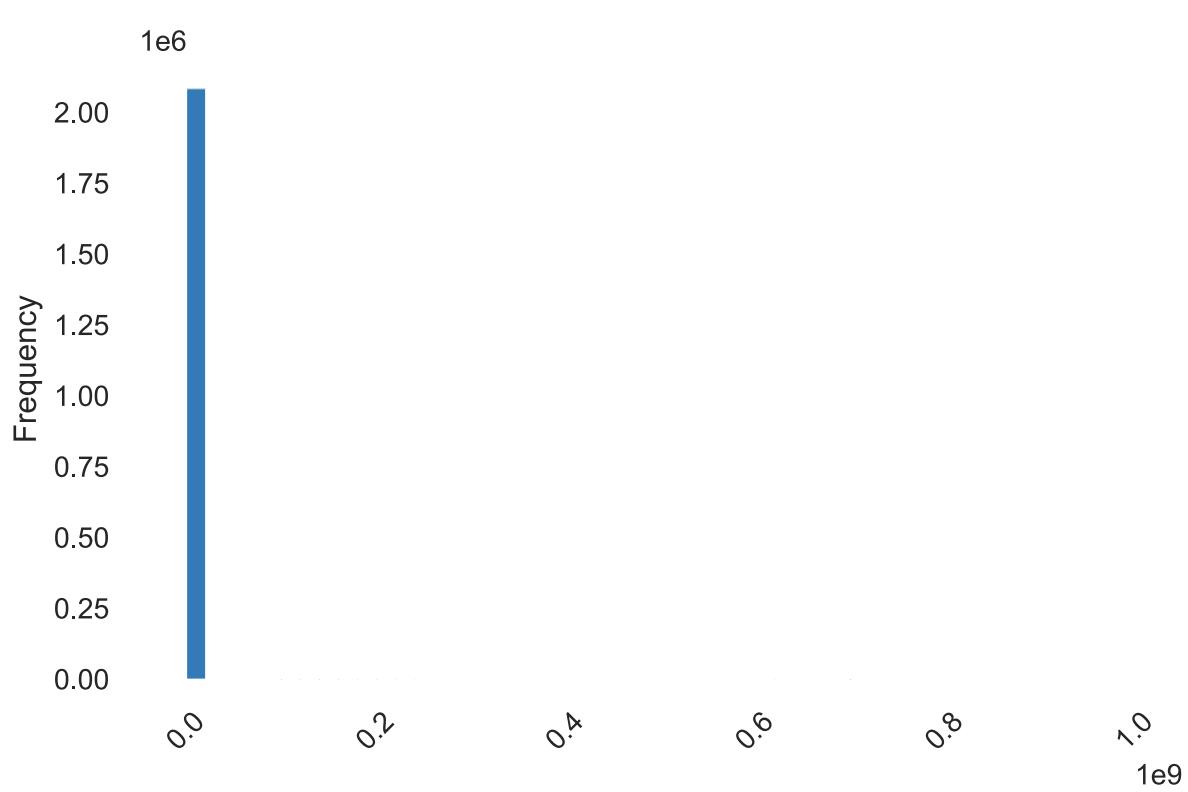
## Quantile statistics

Minimum	8
5-th percentile	5160
Q1	41785

<b>median</b>	112500
<b>Q3</b>	233000
<b>95-th percentile</b>	857500
<b>Maximum</b>	1000000000
<b>Range</b>	999999992
<b>Interquartile range (IQR)</b>	191215

## Descriptive statistics

<b>Standard deviation</b>	1671683.305
<b>Coefficient of variation (CV)</b>	6.116287255
<b>Kurtosis</b>	108744.3808
<b>Mean</b>	273316.6764
<b>Median Absolute Deviation (MAD)</b>	84830
<b>Skewness</b>	253.0599823
<b>Sum</b>	$5.69417851 \times 10^{11}$
<b>Variance</b>	$2.794525072 \times 10^{12}$
<b>Monotocity</b>	Not monotonic



**Histogram with fixed size bins** (bins=50)

Value	Count	Frequency (%)
100000	23593	1.1%
10000	22497	1.1%
200000	21403	1.0%
5000	14706	0.7%
50000	14229	0.7%
150000	12592	0.6%
18000	10700	0.5%
30000	9972	0.5%
4000	9939	0.5%
6000	9812	0.5%
Other values (226544)	1933920	92.8%

Value	Count	Frequency (%)
8	1	< 0.1%
10	2	< 0.1%
18	1	< 0.1%
31	1	< 0.1%
40	1	< 0.1%

Value	Count	Frequency (%)
10000000000	1	< 0.1%
714938313	1	< 0.1%
601544284	1	< 0.1%
600000000	1	< 0.1%
500000000	2	< 0.1%





## valor\_desembolsado\_reais

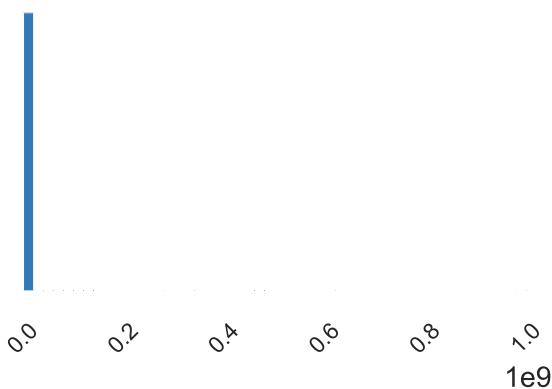
Real number ( $\mathbb{R}_{\geq 0}$ )

HIGH

CORRELATION (This variable has a high correlation with 1 fields: valor\_da\_operacao\_em\_reais)

SKEWED

Distinct	229424
Distinct (%)	11.0%
Missing	2937
Missing (%)	0.1%
Infinite	0
Infinite (%)	0.0%
Mean	268115.0582
Minimum	8
Maximum	999999999
Zeros	0
Zeros (%)	0.0%
Memory size	15.9 MiB



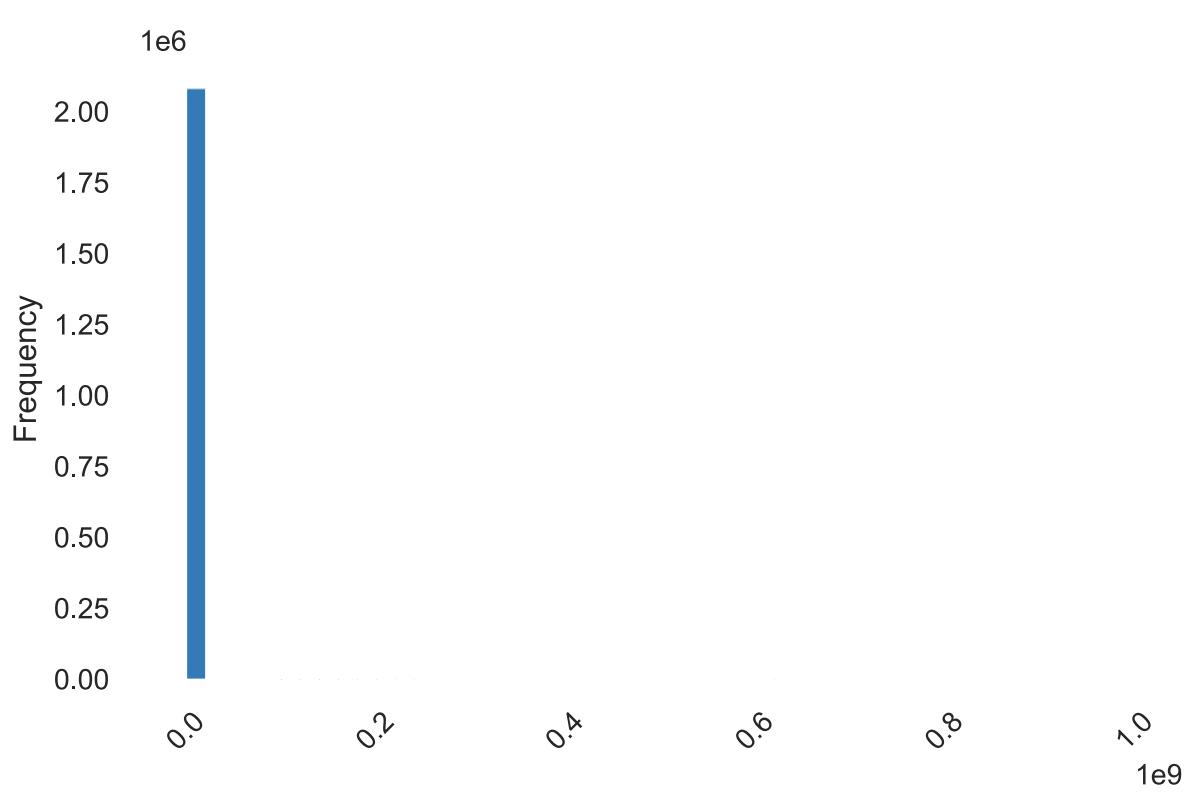
## Quantile statistics

Minimum	8
5-th percentile	5130
Q1	41590

<b>median</b>	112400
<b>Q3</b>	232000
<b>95-th percentile</b>	843734.5
<b>Maximum</b>	999999999
<b>Range</b>	999999991
<b>Interquartile range (IQR)</b>	190410

## Descriptive statistics

<b>Standard deviation</b>	1594244.404
<b>Coefficient of variation (CV)</b>	5.946120353
<b>Kurtosis</b>	115126.2986
<b>Mean</b>	268115.0582
<b>Median Absolute Deviation (MAD)</b>	84600
<b>Skewness</b>	256.9119925
<b>Sum</b>	$5.577935381 \times 10^{11}$
<b>Variance</b>	$2.541615221 \times 10^{12}$
<b>Monotocity</b>	Not monotonic



**Histogram with fixed size bins** (bins=50)

Value	Count	Frequency (%)
100000	23570	1.1%
10000	22479	1.1%
200000	21342	1.0%
5000	14703	0.7%
50000	14188	0.7%
150000	12532	0.6%
18000	10696	0.5%
30000	9945	0.5%
4000	9937	0.5%
6000	9811	0.5%
Other values (229414)	1931223	92.7%

Value	Count	Frequency (%)
8	1	< 0.1%
10	2	< 0.1%
18	1	< 0.1%
31	1	< 0.1%
40	1	< 0.1%

Value	Count	Frequency (%)
9999999999	1	< 0.1%
601544284	1	< 0.1%
6000000000	1	< 0.1%
5000000000	2	< 0.1%
481202202	1	< 0.1%





## **fonte\_de\_recurso\_desembolsos**

Categorical

HIGH

CORRELATION (This variable has a high correlation with 2 fields: modalidade\_de\_apoio, forma\_de\_apoio)

<b>Distinct</b>	9
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	15.9 MiB

## Length

<b>Max length</b>	45
<b>Median length</b>	25
<b>Mean length</b>	26.48261825
<b>Min length</b>	1

## Characters and Unicode

<b>Total characters</b>	55172907	
<b>Distinct characters</b>	21	
<b>Distinct categories</b>	4 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

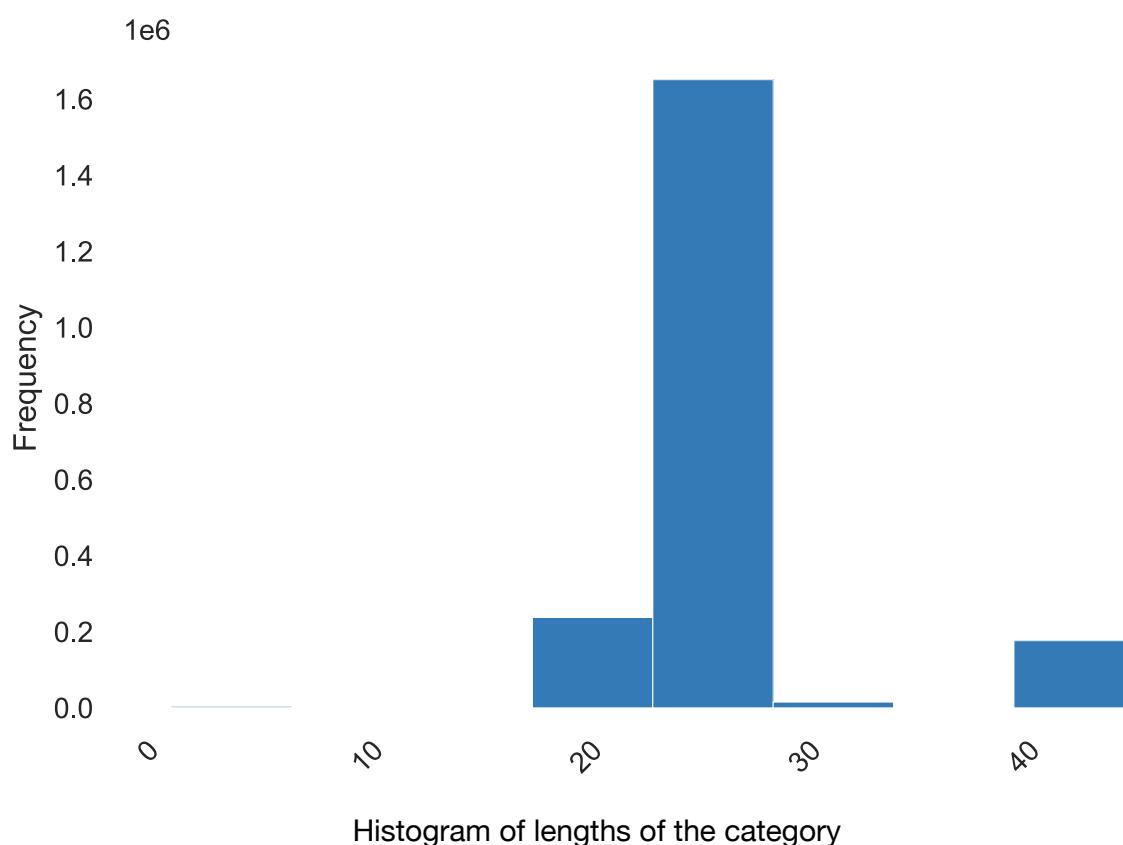
## Unique

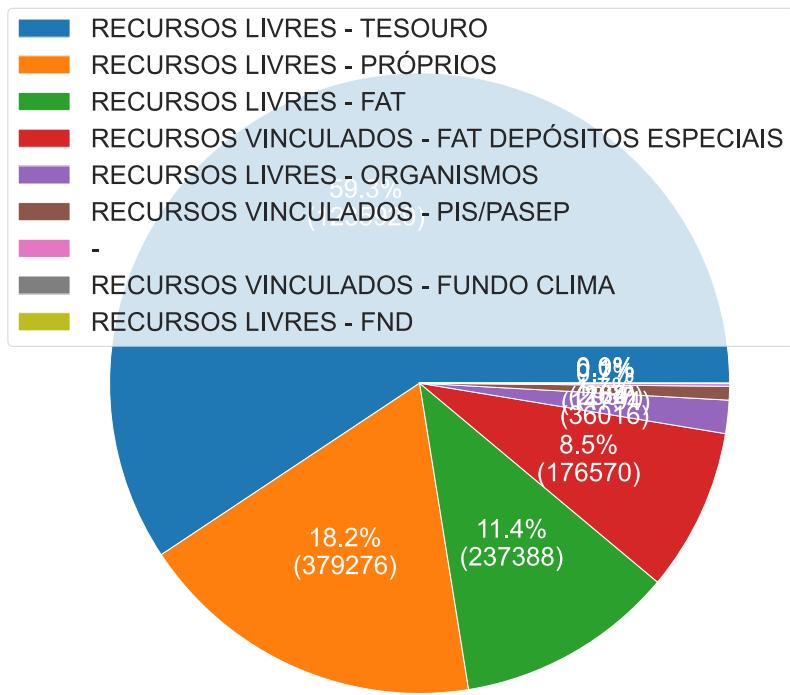
<b>Unique</b>	0	?
<b>Unique (%)</b>	0.0%	

## Sample

<b>1st row</b>	RECURSOS LIVRES - TESOURO
<b>2nd row</b>	RECURSOS VINCULADOS - FAT DEPÓSITOS ESPECIAIS
<b>3rd row</b>	RECURSOS VINCULADOS - FAT DEPÓSITOS ESPECIAIS
<b>4th row</b>	RECURSOS VINCULADOS - FAT DEPÓSITOS ESPECIAIS
<b>5th row</b>	RECURSOS VINCULADOS - FAT DEPÓSITOS ESPECIAIS

Value	Count	Frequency (%)
RECURSOS LIVRES - TESOURO	1235930	59.3%
RECURSOS LIVRES - PRÓPRIOS	379277	18.2%
RECURSOS LIVRES - FAT	237389	11.4%
RECURSOS VINCULADOS - FAT DEPÓSITOS ESPECIAIS	176571	8.5%
RECURSOS LIVRES - ORGANISMOS	36017	1.7%
RECURSOS VINCULADOS - PIS/PASEP	14794	0.7%
-	2937	0.1%
RECURSOS VINCULADOS - FUNDO CLIMA	263	< 0.1%
RECURSOS LIVRES - FND	185	< 0.1%





Value	Count	Frequency (%)
	2083363	24.0%
recursos	2080426	24.0%
livres	1888798	21.8%
tesouro	1235930	14.2%
fat	413960	4.8%
próprios	379277	4.4%
vinculados	191628	2.2%
especiais	176571	2.0%
depósitos	176571	2.0%
organismos	36017	0.4%
Other values (4)	15505	0.2%

## Most occurring characters

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
S	8664391	15.7%
R	8080151	14.6%
	6594683	12.0%
E	5749661	10.4%
O	5372059	9.7%
U	3508247	6.4%
I	3040490	5.5%
C	2448888	4.4%
-	2083363	3.8%
L	2080689	3.8%
Other values (11)	7550285	13.7%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	46480067	84.2%
Space Separator	6594683	12.0%
Dash Punctuation	2083363	3.8%
Other Punctuation	14794	< 0.1%

## Most frequent character per category

Value	Count	Frequency (%)
S	8664391	18.6%
R	8080151	17.4%
E	5749661	12.4%
O	5372059	11.6%
U	3508247	7.5%
I	3040490	6.5%
C	2448888	5.3%
L	2080689	4.5%
V	2080426	4.5%
T	1826461	3.9%
Other values (8)	3628604	7.8%

Value	Count	Frequency (%)
-	6594683	100.0%

Value	Count	Frequency (%)
-	2083363	100.0%

Value	Count	Frequency (%)
/	14794	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	46480067	84.2%
Common	8692840	15.8%

## Most frequent character per script

Value	Count	Frequency (%)
S	8664391	18.6%
R	8080151	17.4%
E	5749661	12.4%
O	5372059	11.6%
U	3508247	7.5%
I	3040490	6.5%
C	2448888	5.3%
L	2080689	4.5%
V	2080426	4.5%
T	1826461	3.9%
Other values (8)	3628604	7.8%

Value	Count	Frequency (%)
	6594683	75.9%
-	2083363	24.0%
/	14794	0.2%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	54617059	99.0%
None	555848	1.0%

## Most frequent character per block

Value	Count	Frequency (%)
S	8664391	15.9%
R	8080151	14.8%
	6594683	12.1%
E	5749661	10.5%
O	5372059	9.8%
U	3508247	6.4%
I	3040490	5.6%
C	2448888	4.5%
-	2083363	3.8%
L	2080689	3.8%
Other values (10)	6994437	12.8%

Value	Count	Frequency (%)
Ó	555848	100.0%





**custo\_financeiro**

Categorical

HIGH

CORRELATION (This variable has a high correlation with 2 fields: modalidade\_de\_apoio, forma\_de\_apoio)

Distinct	11
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	15.9 MiB

## Length

<b>Max length</b>	12
<b>Median length</b>	9
<b>Mean length</b>	7.003982023
<b>Min length</b>	3

## Characters and Unicode

<b>Total characters</b>	14591837	
<b>Distinct characters</b>	29	
<b>Distinct categories</b>	6 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

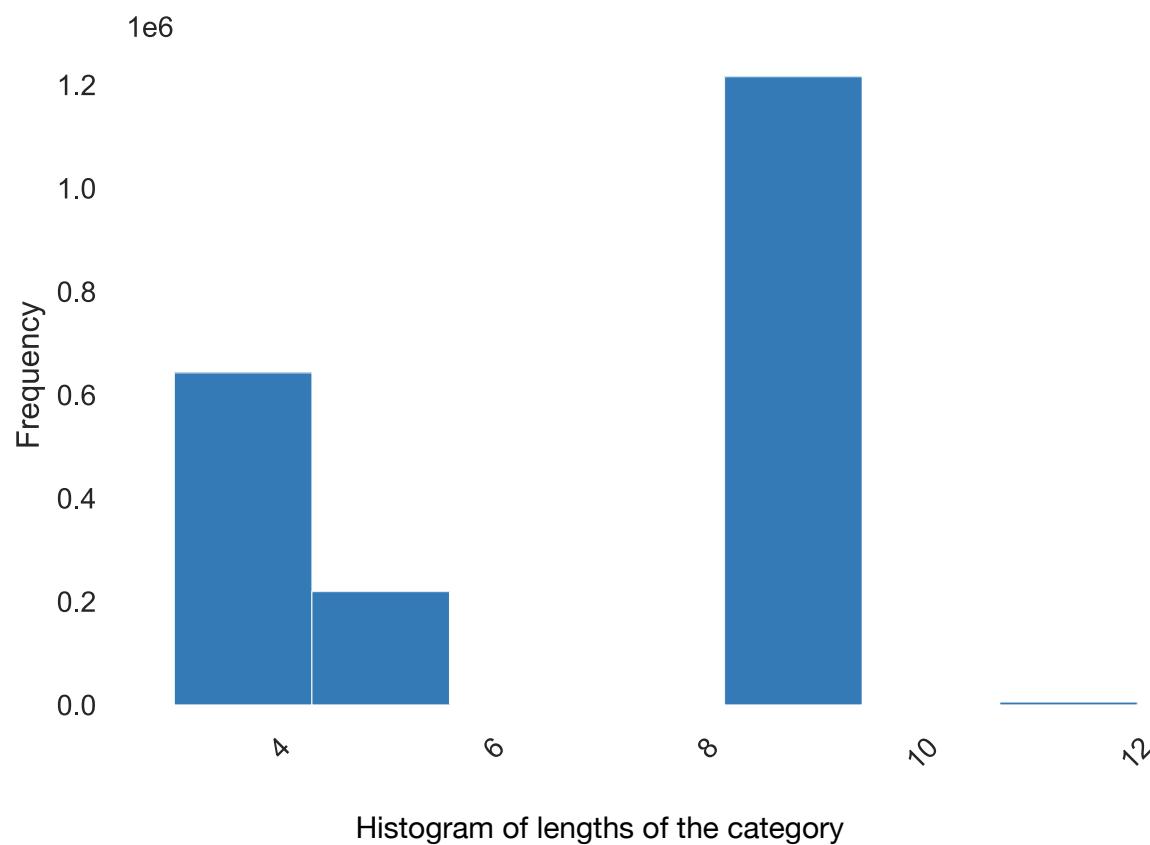
## Unique

<b>Unique</b>	0	?
<b>Unique (%)</b>	0.0%	

## Sample

<b>1st row</b>	TAXA FIXA
<b>2nd row</b>	TAXA FIXA
<b>3rd row</b>	TAXA FIXA
<b>4th row</b>	TAXA FIXA
<b>5th row</b>	TAXA FIXA

Value	Count	Frequency (%)
TAXA FIXA	1215892	58.4%
TJLP	563382	27.0%
TJ462	159600	7.7%
TLP	79331	3.8%
SELIC	56629	2.7%
US\$ / CESTA	5280	0.3%
TJ453	2973	0.1%
75% da SELIC	230	< 0.1%
90% da SELIC	28	< 0.1%
OUTROS	16	< 0.1%



Value	Count	Frequency (%)
fixa	1215892	36.7%
taxa	1215892	36.7%
tjlp	563382	17.0%
tj462	159600	4.8%
tlp	79331	2.4%
selic	56887	1.7%
cesta	5280	0.2%
us	5280	0.2%
	5280	0.2%
tj453	2973	0.1%
Other values (5)	534	< 0.1%

## Most occurring characters

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
A	3652958	25.0%
X	2431784	16.7%
T	2026474	13.9%
I	1272779	8.7%
	1226968	8.4%
F	1215894	8.3%
J	725955	5.0%
L	699600	4.8%
P	642713	4.4%
4	162573	1.1%
Other values (19)	534139	3.7%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	12865300	88.2%
Space Separator	1226968	8.4%
Decimal Number	488235	3.3%
Other Punctuation	5538	< 0.1%
Currency Symbol	5280	< 0.1%
Lowercase Letter	516	< 0.1%

## Most frequent character per category

Value	Count	Frequency (%)
A	3652958	28.4%
X	2431784	18.9%
T	2026474	15.8%
I	1272779	9.9%
F	1215894	9.5%
J	725955	5.6%
L	699600	5.4%
P	642713	5.0%
S	67463	0.5%
C	62167	0.5%
Other values (5)	67513	0.5%

Value	Count	Frequency (%)
4	162573	33.3%
6	159600	32.7%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
2	159600	32.7%
5	3203	0.7%
3	2973	0.6%
7	230	< 0.1%
9	28	< 0.1%
0	28	< 0.1%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
/	5280	95.3%
%	258	4.7%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
d	258	50.0%
a	258	50.0%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
	1226968	100.0%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
\$	5280	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	12865816	88.2%
Common	1726021	11.8%

## Most frequent character per script

Value	Count	Frequency (%)
A	3652958	28.4%
X	2431784	18.9%
T	2026474	15.8%
I	1272779	9.9%
F	1215894	9.5%
J	725955	5.6%
L	699600	5.4%
P	642713	5.0%
S	67463	0.5%
C	62167	0.5%
Other values (7)	68029	0.5%

Value	Count	Frequency (%)
	1226968	71.1%
4	162573	9.4%
6	159600	9.2%
2	159600	9.2%
\$	5280	0.3%
/	5280	0.3%

Value	Count	Frequency (%)
5	3203	0.2%
3	2973	0.2%
%	258	< 0.1%
7	230	< 0.1%
Other values (2)	56	< 0.1%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	14591837	100.0%

## Most frequent character per block

Value	Count	Frequency (%)
A	3652958	25.0%
X	2431784	16.7%
T	2026474	13.9%
I	1272779	8.7%
	1226968	8.4%
F	1215894	8.3%
J	725955	5.0%
L	699600	4.8%
P	642713	4.4%
4	162573	1.1%
Other values (19)	534139	3.7%

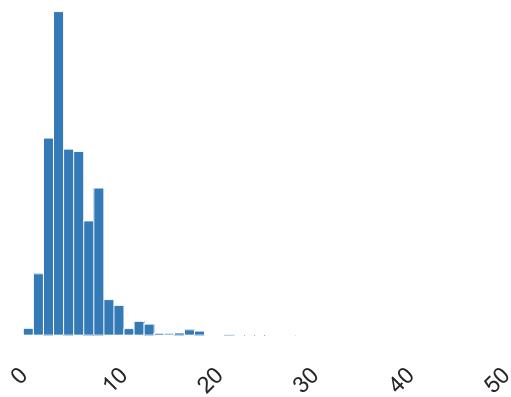




# juros

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

<b>Distinct</b>	1973
<b>Distinct (%)</b>	0.1%
<b>Missing</b>	436
<b>Missing (%)</b>	< 0.1%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	5.628104086
<b>Minimum</b>	0.33
<b>Maximum</b>	53
<b>Zeros</b>	0
<b>Zeros (%)</b>	0.0%
<b>Memory size</b>	15.9 MiB



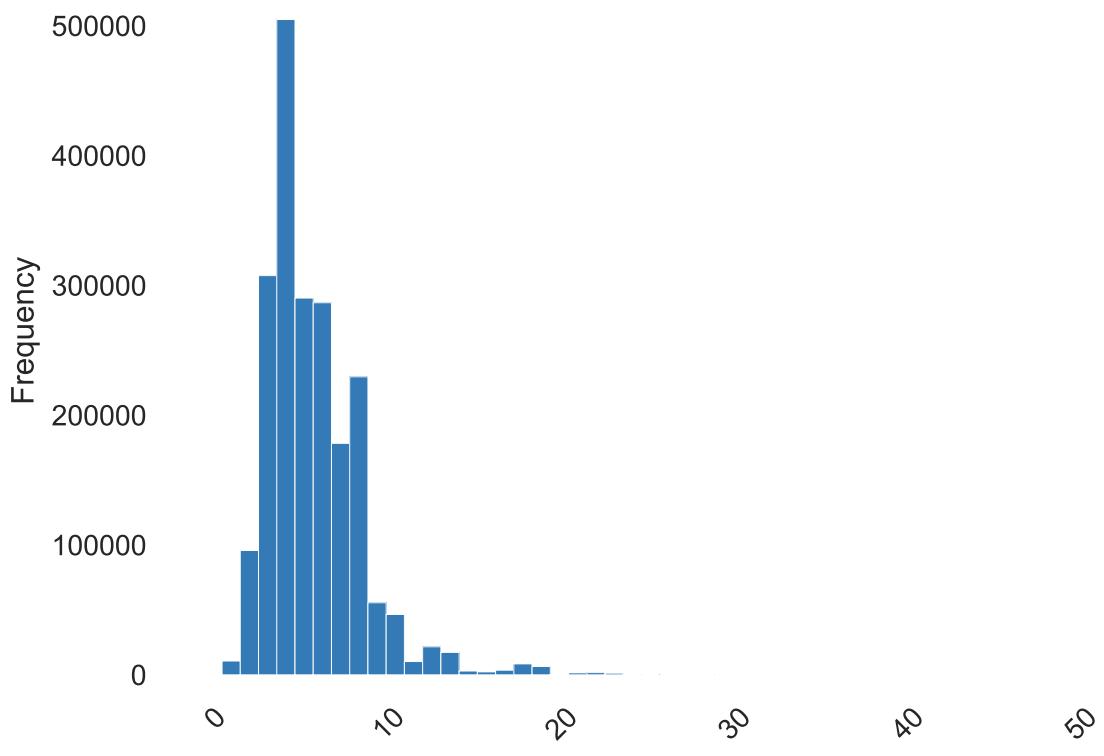
## Quantile statistics

<b>Minimum</b>	0.33
<b>5-th percentile</b>	2.4
<b>Q1</b>	3.9
<b>median</b>	5.1
<b>Q3</b>	7

<b>95-th percentile</b>	10
<b>Maximum</b>	53
<b>Range</b>	52.67
<b>Interquartile range (IQR)</b>	3.1

## Descriptive statistics

<b>Standard deviation</b>	2.781144032
<b>Coefficient of variation (CV)</b>	0.4941529135
<b>Kurtosis</b>	5.16200662
<b>Mean</b>	5.628104086
<b>Median Absolute Deviation (MAD)</b>	1.6
<b>Skewness</b>	1.674332276
<b>Sum</b>	11722929.96
<b>Variance</b>	7.734762125
<b>Monotocity</b>	Not monotonic



**Histogram with fixed size bins (bins=50)**

Value	Count	Frequency (%)
4	174756	8.4%
6	153060	7.3%
3	134376	6.4%
8.75	123765	5.9%
5.5	115703	5.6%
4.5	115418	5.5%
7	110181	5.3%
2.5	93453	4.5%
8	70755	3.4%
3.9	60636	2.9%
Other values (1963)	930824	44.7%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
0.33	1	< 0.1%
0.59	6	< 0.1%
0.73	3	< 0.1%
0.75	2	< 0.1%
0.8	114	< 0.1%

Value	Count	Frequency (%)
53	1	< 0.1%
28.21	1	< 0.1%
27.51	1	< 0.1%
27.1	1	< 0.1%
27.05	1	< 0.1%



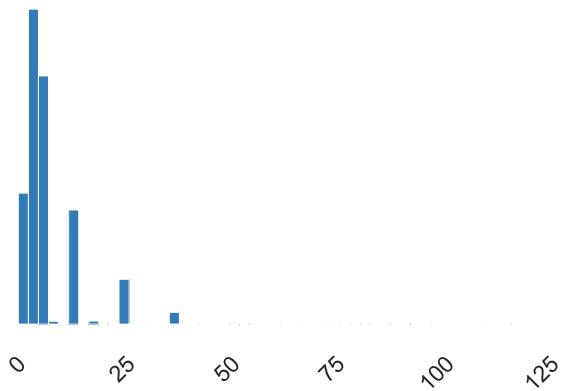


## prazo\_carencia\_meses

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

ZEROS

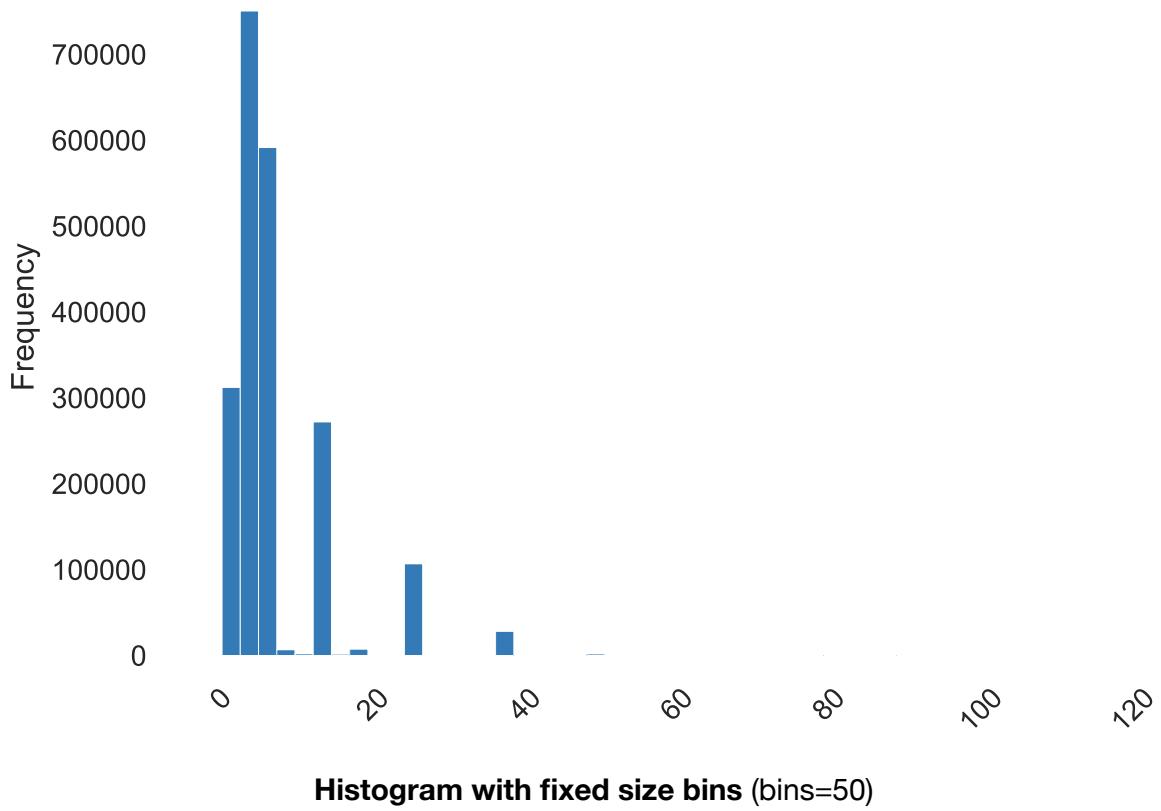
<b>Distinct</b>	57
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	6.296293541
<b>Minimum</b>	0
<b>Maximum</b>	120
<b>Zeros</b>	309940
<b>Zeros (%)</b>	14.9%
<b>Memory size</b>	15.9 MiB



<b>95-th percentile</b>	24
<b>Maximum</b>	120
<b>Range</b>	120
<b>Interquartile range (IQR)</b>	3

## Descriptive statistics

<b>Standard deviation</b>	6.92397936
<b>Coefficient of variation (CV)</b>	1.099691321
<b>Kurtosis</b>	13.82370159
<b>Mean</b>	6.296293541
<b>Median Absolute Deviation (MAD)</b>	3
<b>Skewness</b>	2.849542919
<b>Sum</b>	13117465
<b>Variance</b>	47.94149017
<b>Monotocity</b>	Not monotonic



Value	Count	Frequency (%)
3	749616	36.0%
6	589245	28.3%
0	309940	14.9%
12	271805	13.0%
24	106714	5.1%
36	28094	1.3%
18	7303	0.4%
9	6590	0.3%
5	1905	0.1%
48	1809	0.1%
Other values (47)	10342	0.5%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
0	309940	14.9%
1	1135	0.1%
2	1049	0.1%
3	749616	36.0%
4	826	< 0.1%

Value	Count	Frequency (%)
120	1	< 0.1%
108	1	< 0.1%
96	230	< 0.1%
95	2	< 0.1%
90	23	< 0.1%

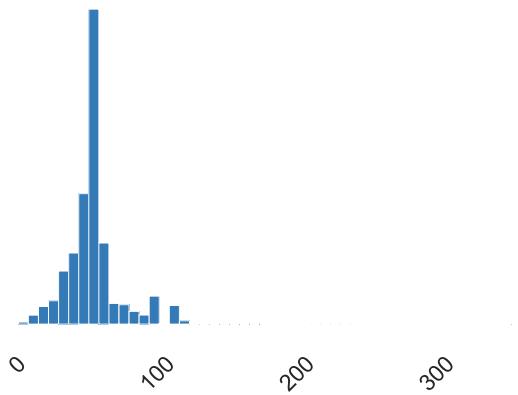




## prazo\_amortizacao\_meses

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

<b>Distinct</b>	162
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Infinite</b>	0
<b>Infinite (%)</b>	0.0%
<b>Mean</b>	53.87461091
<b>Minimum</b>	0
<b>Maximum</b>	360
<b>Zeros</b>	764
<b>Zeros (%)</b>	< 0.1%
<b>Memory size</b>	15.9 MiB



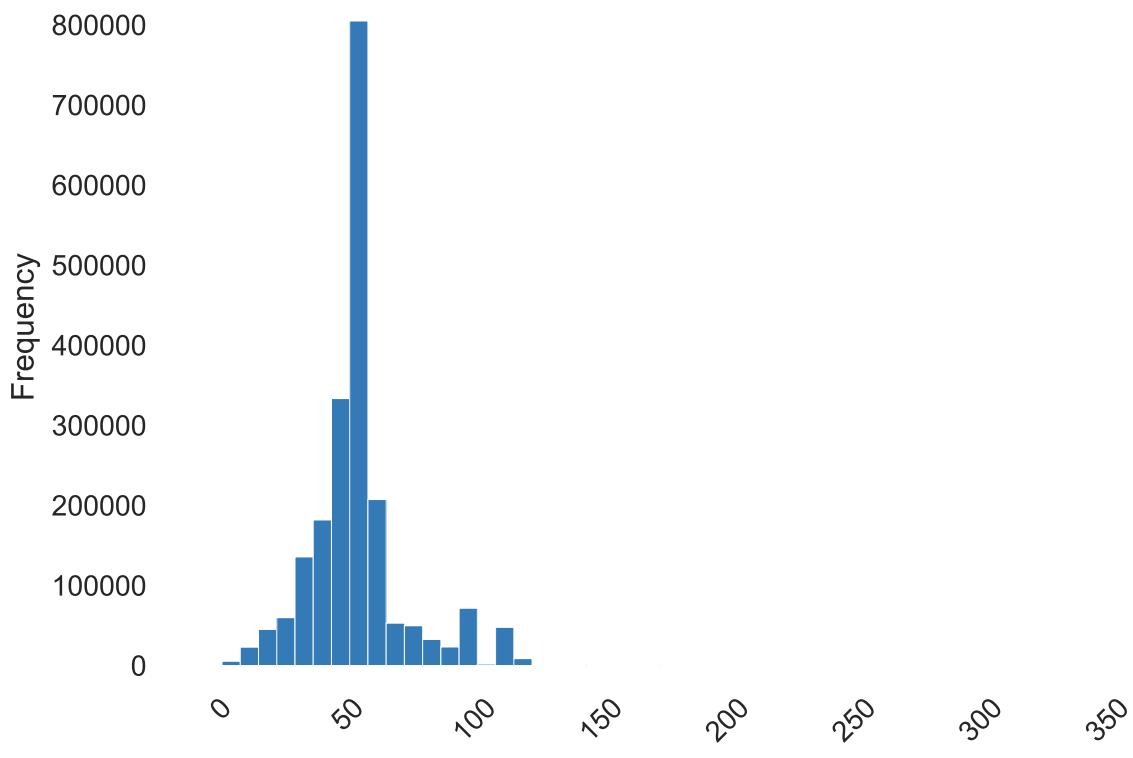
## Quantile statistics

<b>Minimum</b>	0
<b>5-th percentile</b>	24
<b>Q1</b>	45
<b>median</b>	54
<b>Q3</b>	57

<b>95-th percentile</b>	96
<b>Maximum</b>	360
<b>Range</b>	360
<b>Interquartile range (IQR)</b>	12

## Descriptive statistics

<b>Standard deviation</b>	18.66029
<b>Coefficient of variation (CV)</b>	0.3463651929
<b>Kurtosis</b>	2.607520269
<b>Mean</b>	53.87461091
<b>Median Absolute Deviation (MAD)</b>	6
<b>Skewness</b>	0.9448430367
<b>Sum</b>	112240371
<b>Variance</b>	348.2064228
<b>Monotocity</b>	Not monotonic



**Histogram with fixed size bins (bins=50)**

Value	Count	Frequency (%)
57	468410	22.5%
54	327863	15.7%
48	213696	10.3%
60	204193	9.8%
45	116621	5.6%
36	113295	5.4%
33	82099	3.9%
96	69933	3.4%
42	62683	3.0%
24	57053	2.7%
Other values (152)	367517	17.6%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
0	764	< 0.1%
1	560	< 0.1%
2	113	< 0.1%
3	831	< 0.1%
4	379	< 0.1%

Value	Count	Frequency (%)
360	1	< 0.1%
234	3	< 0.1%
231	3	< 0.1%
230	2	< 0.1%
228	8	< 0.1%





## ~~modalidade\_de\_apoio~~

Categorical

CONSTANT

HIGH CORRELATION (This variable has a high correlation with 15 fields: *custo financeiro, subsetor\_cnae agrupado, uf, area\_operacional, setor\_cnae, subsetor\_bndes, porte do cliente, inovacao, instrumento financeiro, forma de apoio, setor\_bndes, fonte de recurso desembolsos, produto, situacao da operacao, natureza do cliente*).

REJECTED

Distinct	1
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	15.9 MiB

## Length

Max length	12
Median length	12
Mean length	12
Min length	12

## Characters and Unicode

Total characters	25000356	
Distinct characters	9	
Distinct categories	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
Distinct scripts	1 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
Distinct blocks	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

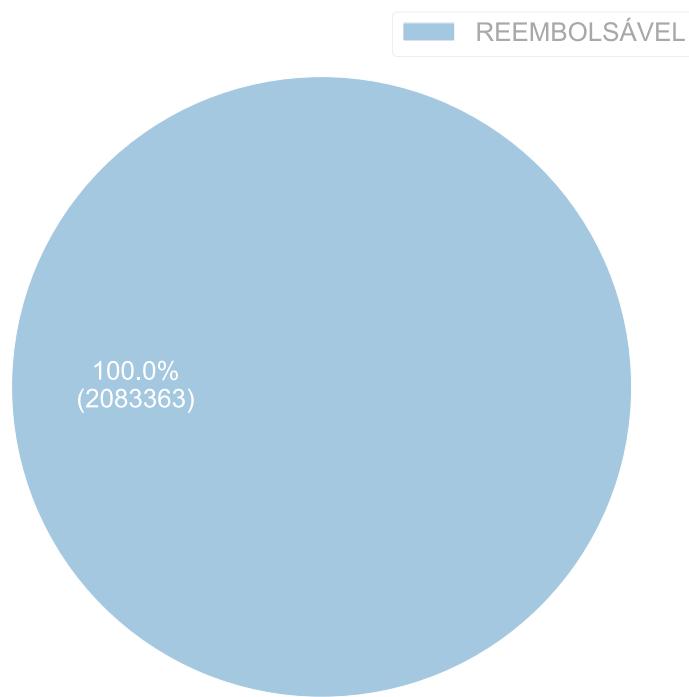
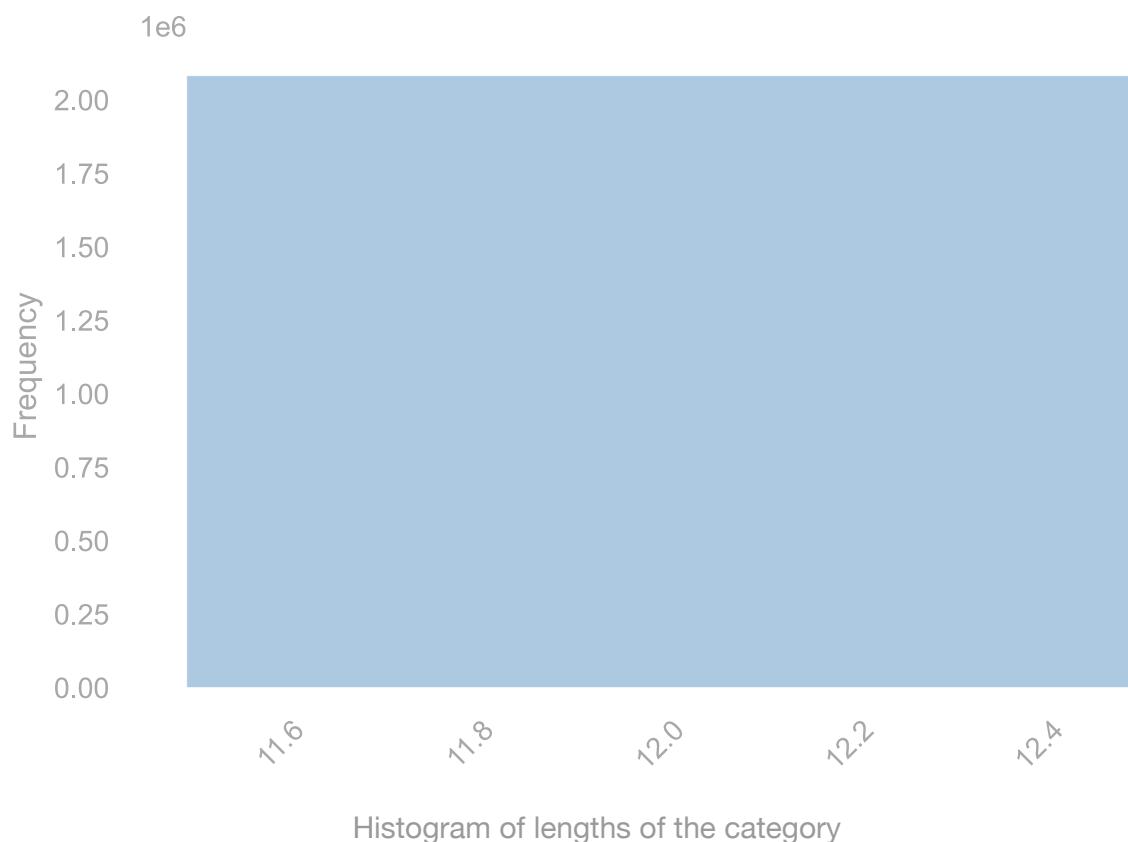
## Unique

Unique	0	?
Unique (%)	0.0%	

## Sample

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

Value	Count	Frequency (%)
REEMBOLSÁVEL	2083363	100.0%



Value	Count	Frequency (%)
-------	-------	---------------

reembolsável	2083363	100.0%
--------------	---------	--------

## Most occurring characters

Value	Count	Frequency (%)
E	6250089	25.0%
L	4166726	16.7%
R	2083363	8.3%
M	2083363	8.3%
B	2083363	8.3%
O	2083363	8.3%
S	2083363	8.3%
Á	2083363	8.3%
V	2083363	8.3%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	25000356	100.0%

## Most frequent character per category

Value	Count	Frequency (%)
E	6250089	25.0%
L	4166726	16.7%
R	2083363	8.3%
M	2083363	8.3%
B	2083363	8.3%
O	2083363	8.3%
S	2083363	8.3%
Á	2083363	8.3%
V	2083363	8.3%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	25000356	100.0%

## Most frequent character per script

Value	Count	Frequency (%)
E	6250089	25.0%
L	4166726	16.7%
R	2083363	8.3%
M	2083363	8.3%
B	2083363	8.3%
O	2083363	8.3%
S	2083363	8.3%
Á	2083363	8.3%
V	2083363	8.3%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	22916993	91.7%
None	2083363	8.3%

## Most frequent character per block

Value	Count	Frequency (%)
E	6250089	27.3%
L	4166726	18.2%
R	2083363	9.1%
M	2083363	9.1%
B	2083363	9.1%
O	2083363	9.1%
S	2083363	9.1%
V	2083363	9.1%

Value	Count	Frequency (%)
Á	2083363	100.0%





## forma\_de\_apoio

Categorical

CONSTANT

HIGH CORRELATION (This variable has a high correlation with 15 fields: custo\_financeiro, subsetor\_cnae\_agrupado, uf, area\_operacional, setor\_cnae, subsetor\_bndes, porte\_do\_cliente, inovacao, modalidade\_de\_apoio, instrumento\_financeiro, setor\_bndes, fonte\_de\_recurso\_desembolsos, produto, situacao\_da\_operacao, natureza\_do\_cliente)

REJECTED

Distinct	1
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	15.9 MiB

## Length

Max length	8
Median length	8
Mean length	8
Min length	8

## Characters and Unicode

Total characters	16666904	
Distinct characters	7	
Distinct categories	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
Distinct scripts	1 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
Distinct blocks	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

## Unique

Unique	0	?
Unique (%)	0.0%	

## Sample

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

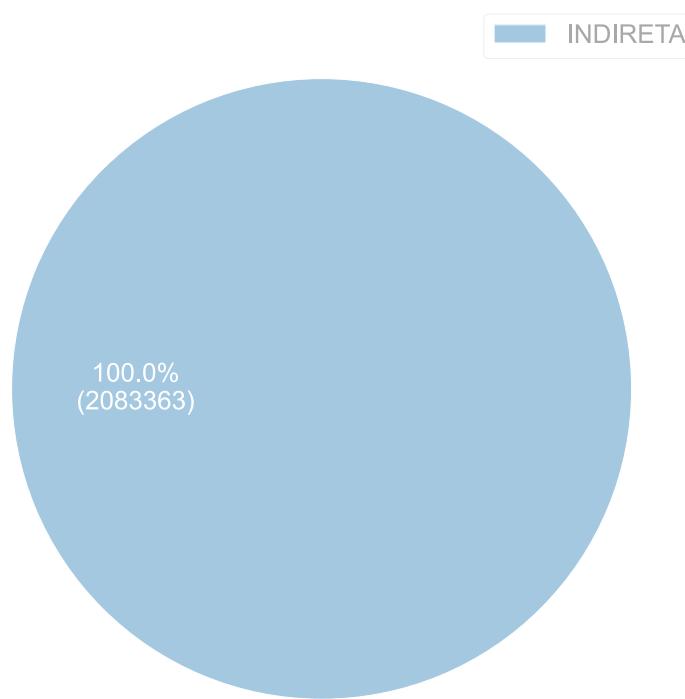
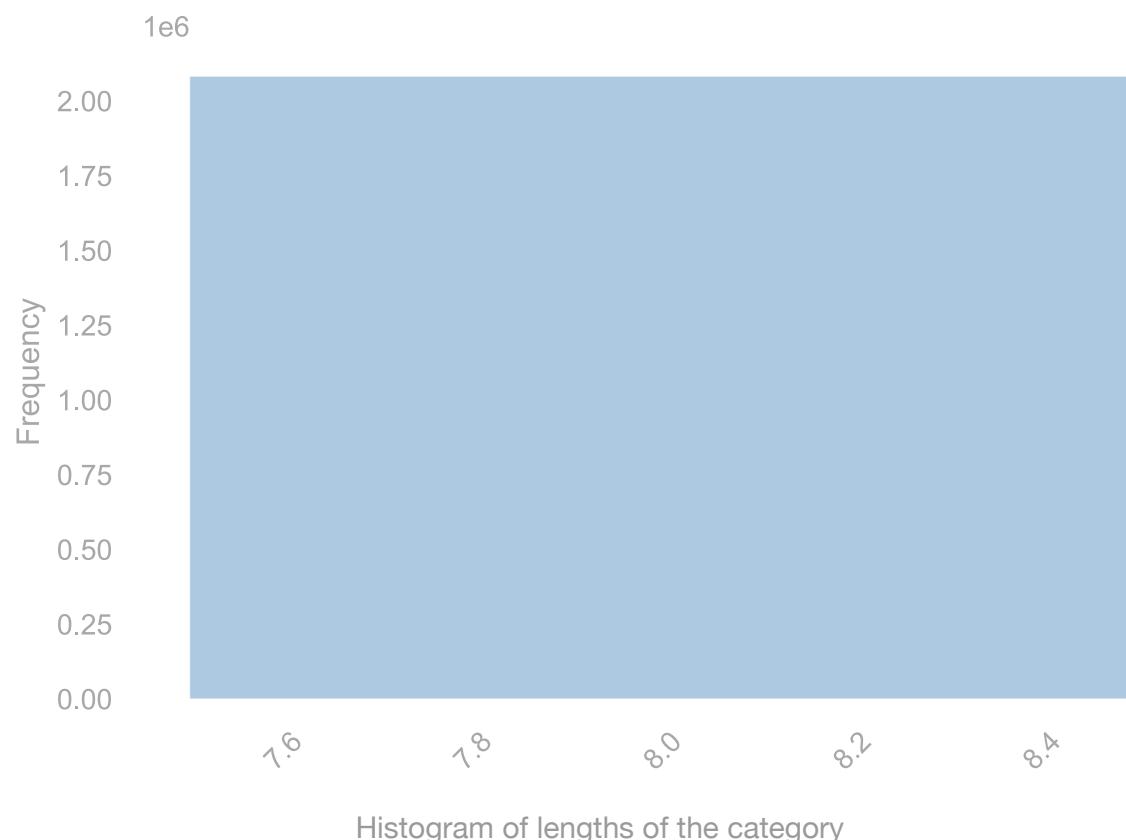
Value

Count Frequency (%)

INDIRETA

2083363

100.0%



Value	Count	Frequency (%)
indireta	2083363	100.0%

## Most occurring characters

Value	Count	Frequency (%)
I	4166726	25.0%
N	2083363	12.5%
D	2083363	12.5%
R	2083363	12.5%
E	2083363	12.5%
T	2083363	12.5%
A	2083363	12.5%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	16666904	100.0%

## Most frequent character per category

Value	Count	Frequency (%)
I	4166726	25.0%
N	2083363	12.5%
D	2083363	12.5%
R	2083363	12.5%
E	2083363	12.5%
T	2083363	12.5%
A	2083363	12.5%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	16666904	100.0%

## Most frequent character per script

Value	Count	Frequency (%)
I	4166726	25.0%
N	2083363	12.5%
D	2083363	12.5%
R	2083363	12.5%
E	2083363	12.5%
T	2083363	12.5%
A	2083363	12.5%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	16666904	100.0%

## Most frequent character per block

Value	Count	Frequency (%)
I	4166726	25.0%
N	2083363	12.5%
D	2083363	12.5%
R	2083363	12.5%
E	2083363	12.5%
T	2083363	12.5%
A	2083363	12.5%





## produto

Categorical

HIGH

CORRELATION (This variable has a high correlation with 2 fields: modalidade\_de\_apoio, forma\_de\_apoio)

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	15.9 MiB

## Length

<b>Max length</b>	21
<b>Median length</b>	12
<b>Mean length</b>	13.14974971
<b>Min length</b>	12

## Characters and Unicode

<b>Total characters</b>	27395702	
<b>Distinct characters</b>	19	
<b>Distinct categories</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

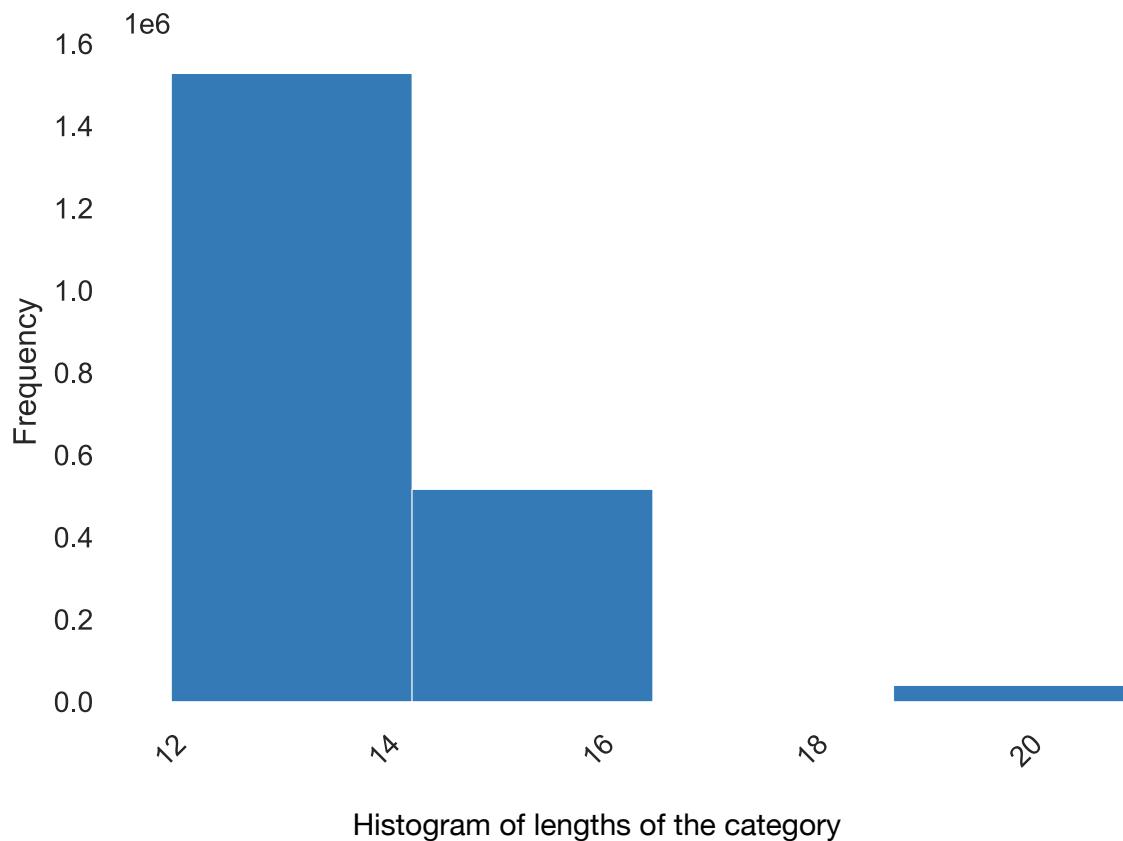
## Unique

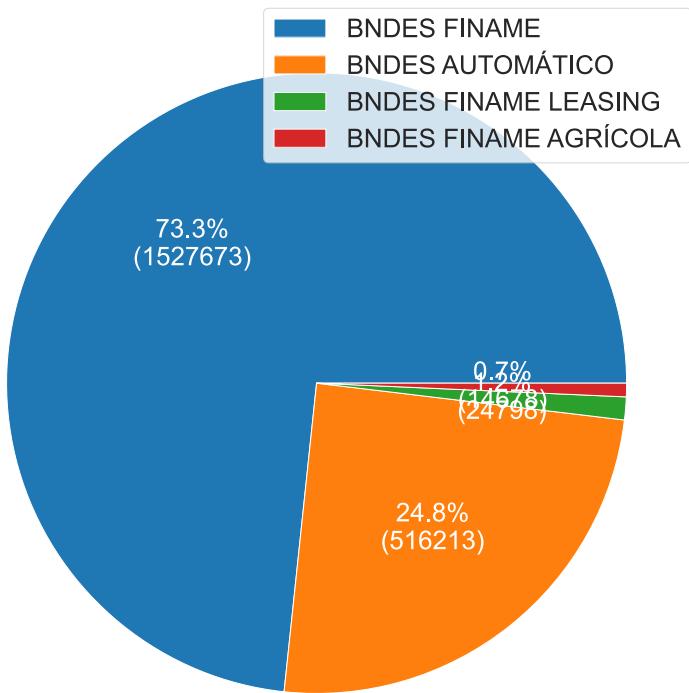
<b>Unique</b>	0	?
<b>Unique (%)</b>	0.0%	

## Sample

<b>1st row</b>	BNDES AUTOMÁTICO
<b>2nd row</b>	BNDES AUTOMÁTICO
<b>3rd row</b>	BNDES AUTOMÁTICO
<b>4th row</b>	BNDES AUTOMÁTICO
<b>5th row</b>	BNDES AUTOMÁTICO

Value	Count	Frequency (%)
BNDES FINAME	1527673	73.3%
BNDES AUTOMÁTICO	516213	24.8%
BNDES FINAME LEASING	24799	1.2%
BNDES FINAME AGRÍCOLA	14678	0.7%





Value	Count	Frequency (%)
bnedes	2083363	49.5%
finame	1567150	37.3%
automático	516213	12.3%
leasing	24799	0.6%
agrícola	14678	0.3%

## Most occurring characters

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
N	3675312	13.4%
E	3675312	13.4%
A	2137518	7.8%
	2122840	7.7%
S	2108162	7.7%
I	2108162	7.7%
B	2083363	7.6%
D	2083363	7.6%
M	2083363	7.6%
F	1567150	5.7%
Other values (9)	3751157	13.7%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	25272862	92.3%
Space Separator	2122840	7.7%

## Most frequent character per category

Value	Count	Frequency (%)
N	3675312	14.5%
E	3675312	14.5%
A	2137518	8.5%
S	2108162	8.3%
I	2108162	8.3%
B	2083363	8.2%
D	2083363	8.2%
M	2083363	8.2%
F	1567150	6.2%
O	1047104	4.1%
Other values (8)	2704053	10.7%

Value	Count	Frequency (%)
	2122840	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	25272862	92.3%
Common	2122840	7.7%

## Most frequent character per script

Value	Count	Frequency (%)
N	3675312	14.5%
E	3675312	14.5%
A	2137518	8.5%
S	2108162	8.3%
I	2108162	8.3%
B	2083363	8.2%
D	2083363	8.2%
M	2083363	8.2%
F	1567150	6.2%
O	1047104	4.1%
Other values (8)	2704053	10.7%

Value	Count	Frequency (%)
	2122840	100.0%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	26864811	98.1%
None	530891	1.9%

## Most frequent character per block

Value	Count	Frequency (%)
N	3675312	13.7%
E	3675312	13.7%
A	2137518	8.0%
	2122840	7.9%
S	2108162	7.8%
I	2108162	7.8%
B	2083363	7.8%
D	2083363	7.8%
M	2083363	7.8%
F	1567150	5.8%
Other values (7)	3220266	12.0%

Value	Count	Frequency (%)
Á	516213	97.2%
í	14678	2.8%





## instrumento\_financeiro

Categorical

HIGH

CARDINALITY

HIGH\_CORRELATION (This variable has a high correlation with 2 fields: modalidade\_de\_apoio, forma\_de\_apoio)

Distinct	80
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	15.9 MiB

## Length

<b>Max length</b>	65
<b>Median length</b>	28
<b>Mean length</b>	23.48336896
<b>Min length</b>	5

## Characters and Unicode

<b>Total characters</b>	48924382	
<b>Distinct characters</b>	60	
<b>Distinct categories</b>	5 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	3 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

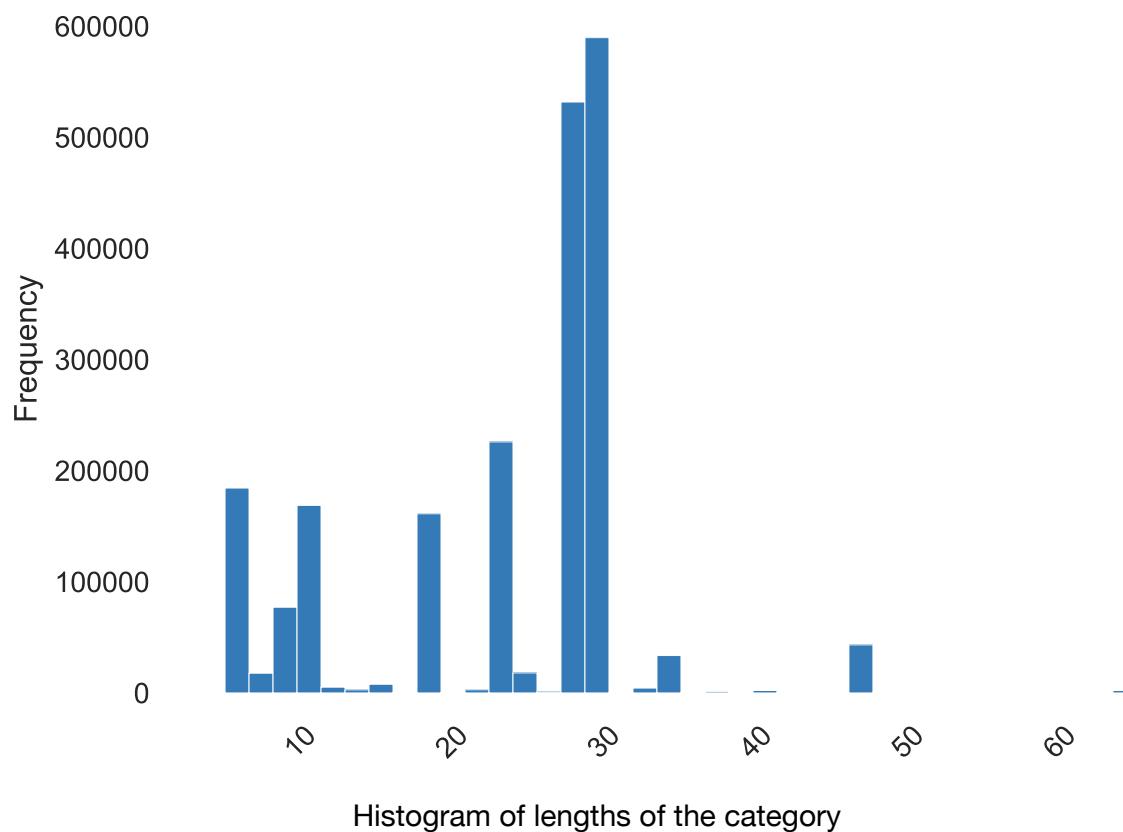
## Unique

<b>Unique</b>	5	?
<b>Unique (%)</b>	< 0.1%	

## Sample

<b>1st row</b>	OUTROS
<b>2nd row</b>	PRONAF INVESTIMENTO
<b>3rd row</b>	PRONAF INVESTIMENTO
<b>4th row</b>	PRONAF INVESTIMENTO
<b>5th row</b>	PRONAF INVESTIMENTO

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
BK AQUISIÇÃO E COMERCIALIZAÇÃO	589349	28.3%
PSI - BK - Ônibus e Caminhão	530980	25.5%
PSI - BK - Demais Itens	225933	10.8%
OUTROS	184381	8.9%
PRONAF INVESTIMENTO	160735	7.7%
BNDES GIRO	80381	3.9%
MODERAGRO	68409	3.3%
MODERINFRA	50139	2.4%
LINHA EMPRÉSTIMO PARA MICRO E PEQUENAS EMPRESAS	43278	2.1%
BNDES PROCAMINHONEIRO - Equalizado	33661	1.6%
Other values (70)	116117	5.6%



Value	Count	Frequency (%)
	1578436	17.6%
bk	1352201	15.1%
e	1174671	13.1%
psi	762909	8.5%
aquisição	590070	6.6%
comercialização	589349	6.6%
caminhão	530980	5.9%
ônibus	530980	5.9%
demais	226028	2.5%
itens	225933	2.5%
Other values (107)	1384039	15.5%

## Most occurring characters

Value	Count	Frequency (%)
	6862233	14.0%
I	4055456	8.3%
O	3036849	6.2%
A	2427318	5.0%
E	2249286	4.6%
S	2070122	4.2%
C	1881624	3.8%
R	1698579	3.5%
-	1581694	3.2%
B	1525013	3.1%
Other values (50)	21536208	44.0%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	31033279	63.4%
Lowercase Letter	9433760	19.3%
Space Separator	6862233	14.0%
Dash Punctuation	1581846	3.2%
Other Punctuation	13264	< 0.1%

## Most frequent character per category

Value	Count	Frequency (%)
I	4055456	13.1%
O	3036849	9.8%
A	2427318	7.8%
E	2249286	7.2%
S	2070122	6.7%
C	1881624	6.1%
R	1698579	5.5%
B	1525013	4.9%
K	1352201	4.4%
P	1255301	4.0%
Other values (20)	9481530	30.6%

Value	Count	Frequency (%)
i	1340650	14.2%
n	1288556	13.7%
e	983221	10.4%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
s	982949	10.4%
a	867773	9.2%
m	756913	8.0%
u	588790	6.2%
o	584111	6.2%
ã	531144	5.6%
b	530980	5.6%
Other values (14)	978673	10.4%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
,	12532	94.5%
:	721	5.4%
&	11	0.1%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
-	1581694	> 99.9%
-	152	< 0.1%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
	6862233	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	40467039	82.7%
Common	8457343	17.3%

## Most frequent character per script

Value	Count	Frequency (%)
I	4055456	10.0%
O	3036849	7.5%
A	2427318	6.0%
E	2249286	5.6%
S	2070122	5.1%
C	1881624	4.6%
R	1698579	4.2%
B	1525013	3.8%
K	1352201	3.3%
i	1340650	3.3%
Other values (44)	18829941	46.5%

Value	Count	Frequency (%)
,	6862233	81.1%
-	1581694	18.7%
,	12532	0.1%
:	721	< 0.1%
-	152	< 0.1%
&	11	< 0.1%



## Most occurring blocks

Value	Count	Frequency (%)
ASCII	45436377	92.9%
None	3487853	7.1%
Punctuation	152	< 0.1%

## Most frequent character per block

Value	Count	Frequency (%)
I	6862233	15.1%
O	4055456	8.9%
A	3036849	6.7%
E	2427318	5.3%
S	2249286	5.0%
C	2070122	4.6%
R	1881624	4.1%
-	1698579	3.7%
B	1581694	3.5%
Other values (38)	1525013	3.4%
	18048203	39.7%

Value	Count	Frequency (%)
Ç	1184500	34.0%
Ã	1180751	33.9%
ã	531144	15.2%
Ô	530980	15.2%
É	52567	1.5%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
Ú	5224	0.1%
Á	2401	0.1%
ç	164	< 0.1%
Ó	49	< 0.1%
Ê	48	< 0.1%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
-	152	100.0%





## inovacao

Categorical

HIGH

CORRELATION (This variable has a high correlation with 2 fields: modalidade\_de\_apoio, forma\_de\_apoio)

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	15.9 MiB

## Length

<b>Max length</b>	3
<b>Median length</b>	3
<b>Mean length</b>	3
<b>Min length</b>	3

## Characters and Unicode

<b>Total characters</b>	6250089	
<b>Distinct characters</b>	6	
<b>Distinct categories</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	1 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

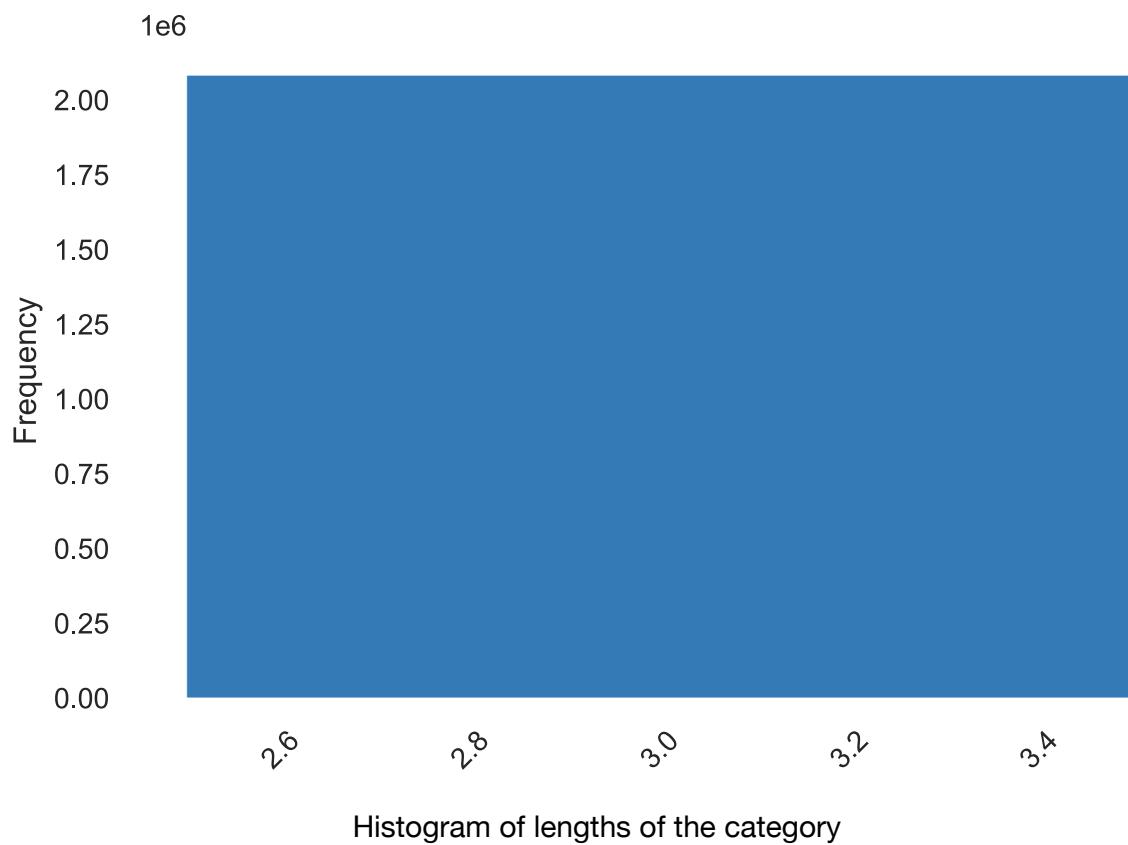
## Unique

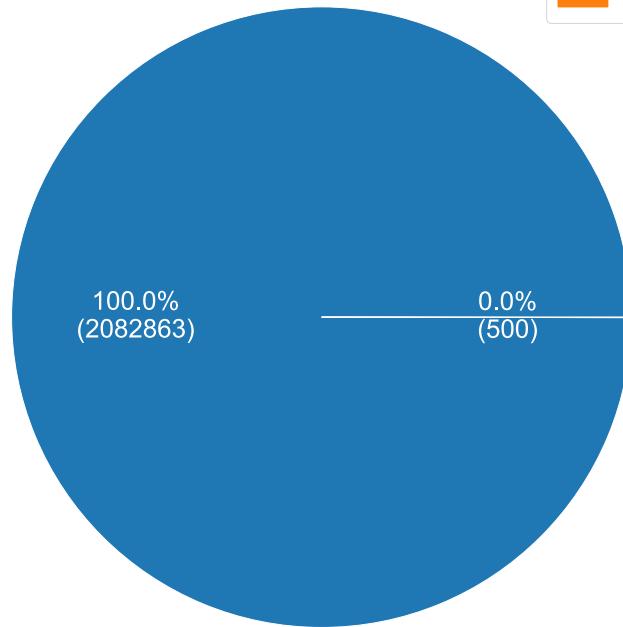
<b>Unique</b>	0	?
<b>Unique (%)</b>	0.0%	

## Sample

<b>1st row</b>	NÃO
<b>2nd row</b>	NÃO
<b>3rd row</b>	NÃO
<b>4th row</b>	NÃO
<b>5th row</b>	NÃO

Value	Count	Frequency (%)
NÃO	2082863	> 99.9%
SIM	500	< 0.1%





Value	Count	Frequency (%)
não	2082863	> 99.9%
sim	500	< 0.1%

## Most occurring characters

Value	Count	Frequency (%)
N	2082863	33.3%
Ã	2082863	33.3%
O	2082863	33.3%
S	500	< 0.1%
I	500	< 0.1%
M	500	< 0.1%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	6250089	100.0%

## Most frequent character per category

Value	Count	Frequency (%)
N	2082863	33.3%
Ã	2082863	33.3%
O	2082863	33.3%
S	500	< 0.1%
I	500	< 0.1%
M	500	< 0.1%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	6250089	100.0%

## Most frequent character per script

Value	Count	Frequency (%)
N	2082863	33.3%
Ã	2082863	33.3%
O	2082863	33.3%
S	500	< 0.1%
I	500	< 0.1%
M	500	< 0.1%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	4167226	66.7%
None	2082863	33.3%

## Most frequent character per block

Value	Count	Frequency (%)
N	2082863	50.0%
O	2082863	50.0%
S	500	< 0.1%
I	500	< 0.1%
M	500	< 0.1%

Value	Count	Frequency (%)
Ã	2082863	100.0%





## area\_operacional

Categorical

HIGH

CORRELATION (This variable has a high correlation with 2 fields: modalidade\_de\_apoio, forma\_de\_apoio)

Distinct	3
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	15.9 MiB

## Length

<b>Max length</b>	44
<b>Median length</b>	35
<b>Mean length</b>	35.00184989
<b>Min length</b>	28

## Characters and Unicode

<b>Total characters</b>	72921559	
<b>Distinct characters</b>	17	
<b>Distinct categories</b>	3 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

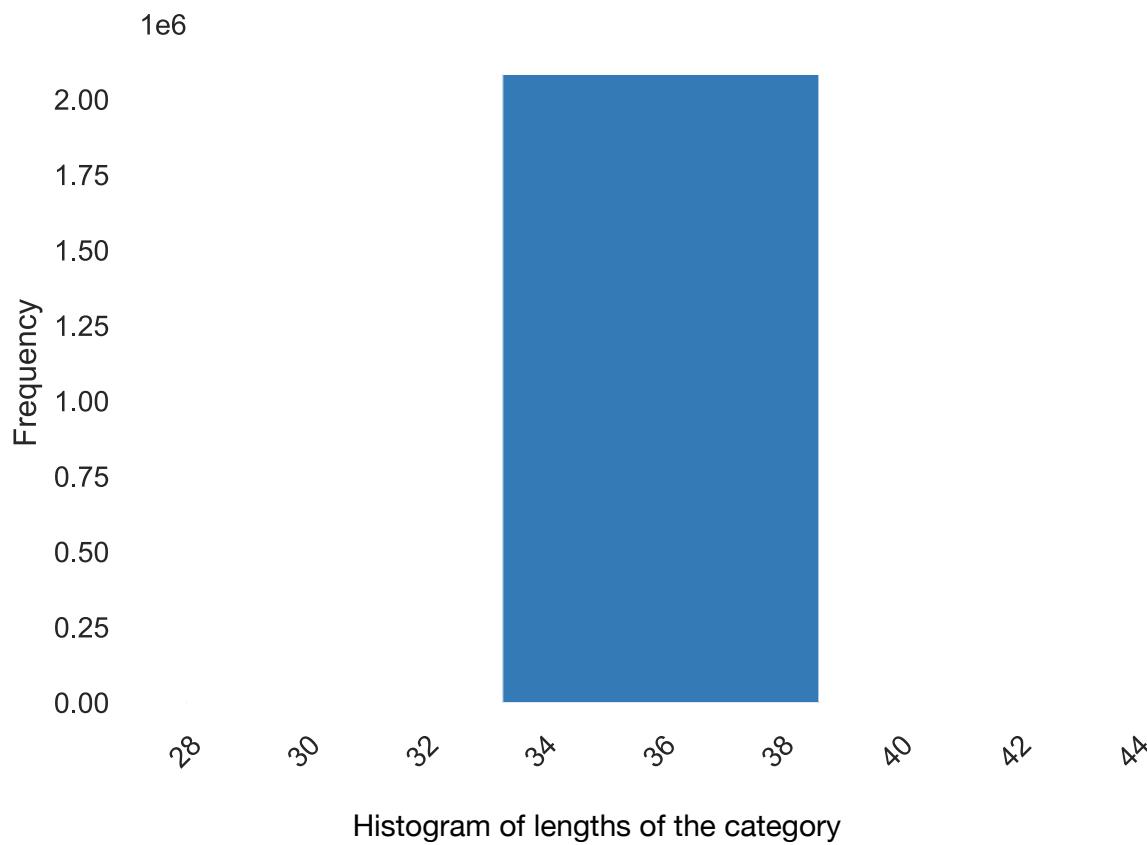
## Unique

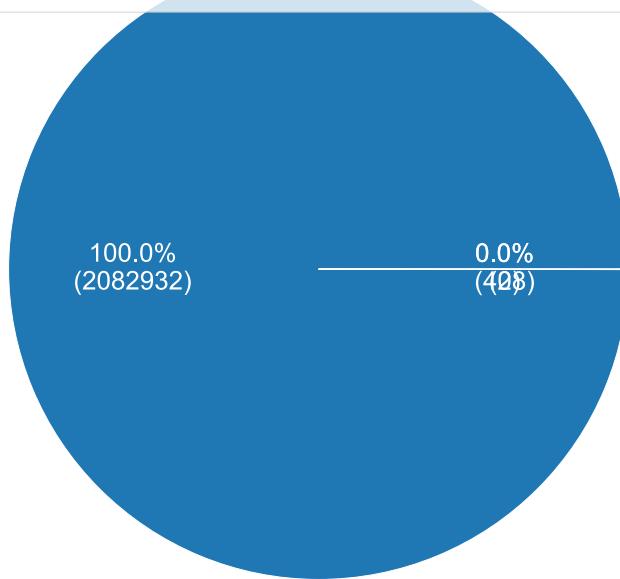
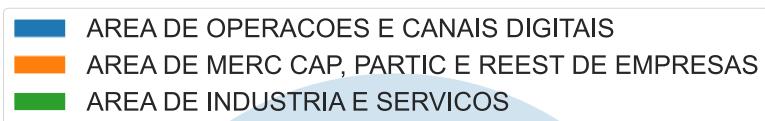
<b>Unique</b>	1	?
<b>Unique (%)</b>	< 0.1%	

## Sample

<b>1st row</b>	AREA DE OPERACOES E CANAIS DIGITAIS
<b>2nd row</b>	AREA DE OPERACOES E CANAIS DIGITAIS
<b>3rd row</b>	AREA DE OPERACOES E CANAIS DIGITAIS
<b>4th row</b>	AREA DE OPERACOES E CANAIS DIGITAIS
<b>5th row</b>	AREA DE OPERACOES E CANAIS DIGITAIS

Value	Count	Frequency (%)
AREA DE OPERACOES E CANAIS DIGITAIS	2082933	> 99.9%
AREA DE MERC CAP, PARTIC E REEST DE EMPRESAS	429	< 0.1%
AREA DE INDUSTRIA E SERVICOS	1	< 0.1%





Value	Count	Frequency (%)
de	2083792	16.7%
area	2083363	16.7%
e	2083363	16.7%
digitais	2082933	16.7%
operacoes	2082933	16.7%
canais	2082933	16.7%
reest	429	< 0.1%
cap	429	< 0.1%
empresas	429	< 0.1%
partic	429	< 0.1%
Other values (3)	431	< 0.1%

## Most occurring characters

Value	Count	Frequency (%)
A	12499746	17.1%
E	10418530	14.3%
	10418101	14.3%
I	8332164	11.4%
S	6250089	8.6%
R	4168014	5.7%
C	4167154	5.7%
D	4166726	5.7%
O	4165867	5.7%
P	2084220	2.9%
Other values (7)	6250948	8.6%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	62503029	85.7%
Space Separator	10418101	14.3%
Other Punctuation	429	< 0.1%

## Most frequent character per category

Value	Count	Frequency (%)
A	12499746	20.0%
E	10418530	16.7%
I	8332164	13.3%
S	6250089	10.0%
R	4168014	6.7%
C	4167154	6.7%
D	4166726	6.7%
O	4165867	6.7%
P	2084220	3.3%
T	2083792	3.3%
Other values (5)	4166727	6.7%

Value	Count	Frequency (%)
,	10418101	100.0%

Value	Count	Frequency (%)
,	429	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	62503029	85.7%
Common	10418530	14.3%

## Most frequent character per script

Value	Count	Frequency (%)
A	12499746	20.0%
E	10418530	16.7%
I	8332164	13.3%
S	6250089	10.0%
R	4168014	6.7%
C	4167154	6.7%
D	4166726	6.7%
O	4165867	6.7%
P	2084220	3.3%
T	2083792	3.3%
Other values (5)	4166727	6.7%

Value	Count	Frequency (%)
	10418101	> 99.9%
,	429	< 0.1%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	72921559	100.0%

## Most frequent character per block

Value	Count	Frequency (%)
A	12499746	17.1%
E	10418530	14.3%
	10418101	14.3%
I	8332164	11.4%
S	6250089	8.6%
R	4168014	5.7%
C	4167154	5.7%
D	4166726	5.7%
O	4165867	5.7%
P	2084220	2.9%
Other values (7)	6250948	8.6%





**setor\_cnae**

Categorical

HIGH

CORRELATION (This variable has a high correlation with 4 fields: subsetor\_cnae\_agrupado, subsetor\_bndes, modalidade\_de\_apoio, forma\_de\_apoio)

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	15.9 MiB

## Length

<b>Max length</b>	26
<b>Median length</b>	19
<b>Mean length</b>	20.19330285
<b>Min length</b>	19

## Characters and Unicode

<b>Total characters</b>	42069980	
<b>Distinct characters</b>	21	
<b>Distinct categories</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

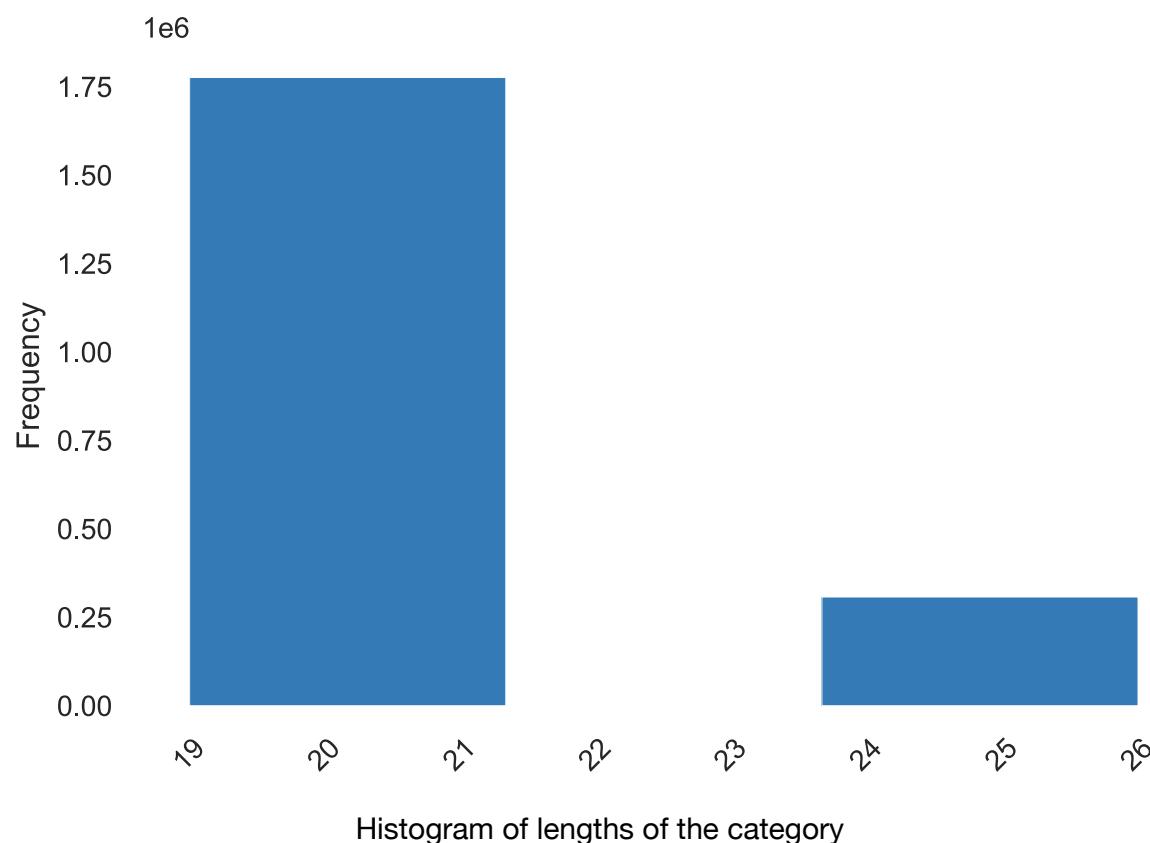
## Unique

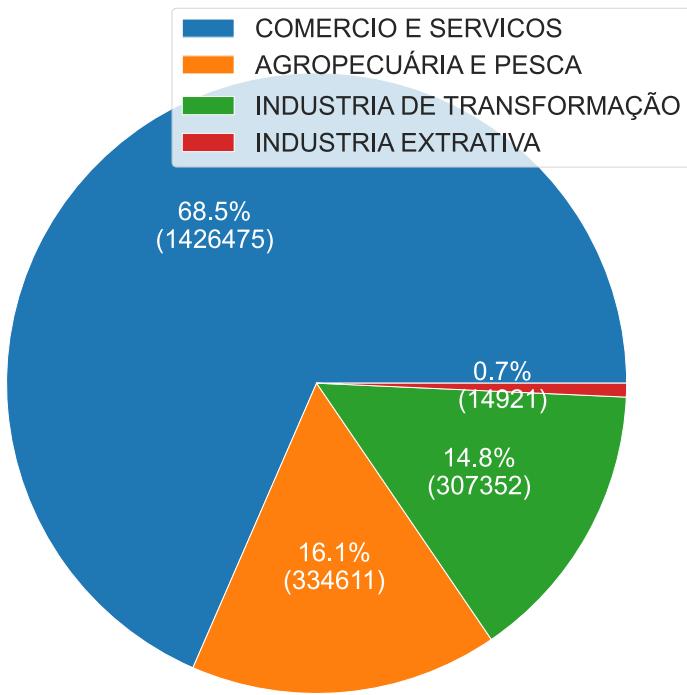
<b>Unique</b>	0	?
<b>Unique (%)</b>	0.0%	

## Sample

<b>1st row</b>	AGROPECUÁRIA E PESCA
<b>2nd row</b>	AGROPECUÁRIA E PESCA
<b>3rd row</b>	AGROPECUÁRIA E PESCA
<b>4th row</b>	AGROPECUÁRIA E PESCA
<b>5th row</b>	AGROPECUÁRIA E PESCA

Value	Count	Frequency (%)
COMERCIO E SERVICOS	1426476	68.5%
AGROPECUÁRIA E PESCA	334612	16.1%
INDUSTRIA DE TRANSFORMAÇÃO	307353	14.8%
INDUSTRIA EXTRATIVA	14922	0.7%





Value	Count	Frequency (%)
e	1761088	28.2%
comercio	1426476	22.9%
servicos	1426476	22.9%
pesca	334612	5.4%
agropecuária	334612	5.4%
industria	322275	5.2%
transformação	307353	4.9%
de	307353	4.9%
extrativa	14922	0.2%

## Most occurring characters

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
E	5605539	13.3%
O	5228746	12.4%
C	4948652	11.8%
R	4474079	10.6%
	4151804	9.9%
I	3847036	9.1%
S	3817192	9.1%
A	1970661	4.7%
M	1733829	4.1%
V	1441398	3.4%
Other values (11)	4851044	11.5%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	37918176	90.1%
Space Separator	4151804	9.9%

## Most frequent character per category

Value	Count	Frequency (%)
E	5605539	14.8%
O	5228746	13.8%
C	4948652	13.1%
R	4474079	11.8%
I	3847036	10.1%
S	3817192	10.1%
A	1970661	5.2%
M	1733829	4.6%
V	1441398	3.8%
P	669224	1.8%
Other values (10)	4181820	11.0%

Value	Count	Frequency (%)
	4151804	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	37918176	90.1%
Common	4151804	9.9%

## Most frequent character per script

Value	Count	Frequency (%)
E	5605539	14.8%
O	5228746	13.8%
C	4948652	13.1%
R	4474079	11.8%
I	3847036	10.1%
S	3817192	10.1%
A	1970661	5.2%
M	1733829	4.6%
V	1441398	3.8%
P	669224	1.8%
Other values (10)	4181820	11.0%

Value	Count	Frequency (%)
	4151804	100.0%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	41120662	97.7%
None	949318	2.3%

## Most frequent character per block

Value	Count	Frequency (%)
E	5605539	13.6%
O	5228746	12.7%
C	4948652	12.0%
R	4474079	10.9%
	4151804	10.1%
I	3847036	9.4%
S	3817192	9.3%
A	1970661	4.8%
M	1733829	4.2%
V	1441398	3.5%
Other values (8)	3901726	9.5%

Value	Count	Frequency (%)
Á	334612	35.2%
Ç	307353	32.4%
Ã	307353	32.4%





## **subsetor\_cnae\_agrupado**

Categorical

HIGH

CORRELATION (This variable has a high correlation with 5 fields: setor\_cnae, subsetor\_bndes, modalidade\_de\_apoio, forma\_de\_apoio, setor\_bndes)

<b>Distinct</b>	44
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	15.9 MiB

## Length

<b>Max length</b>	31
<b>Median length</b>	20
<b>Mean length</b>	16.66774585
<b>Min length</b>	4

## Characters and Unicode

<b>Total characters</b>	34724965	
<b>Distinct characters</b>	49	
<b>Distinct categories</b>	4 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

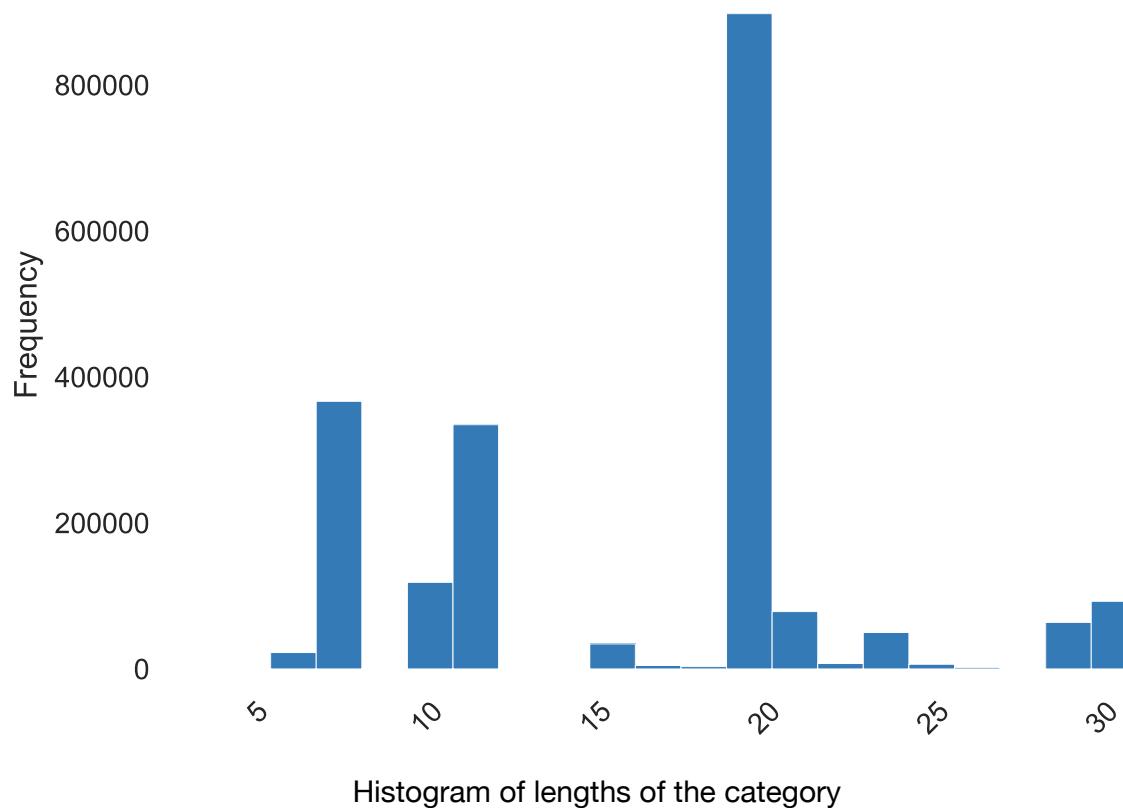
## Unique

<b>Unique</b>	0	?
<b>Unique (%)</b>	0.0%	

## Sample

<b>1st row</b>	Agropecuária
<b>2nd row</b>	Agropecuária
<b>3rd row</b>	Agropecuária
<b>4th row</b>	Agropecuária
<b>5th row</b>	Agropecuária

Value	Count	Frequency (%)
Transporte terrestre	805177	38.6%
Agropecuária	334612	16.1%
Comércio	322068	15.5%
Construção	109010	5.2%
Produtos Alimentícios	77527	3.7%
Ativ imobil, profissional e adm	77399	3.7%
Ativ aux transporte e entrega	43273	2.1%
Mineral não metálico	26936	1.3%
Borracha e plástico	24464	1.2%
Produto de metal	23699	1.1%
Other values (34)	239198	11.5%



Value	Count	Frequency (%)
transporte	850234	21.4%
terrestre	805177	20.3%
agropecuária	334612	8.4%
comércio	322068	8.1%
e	263162	6.6%
ativ	135278	3.4%
construção	109010	2.7%
produtos	82292	2.1%
alimentícios	77527	2.0%
imobil	77399	2.0%
Other values (77)	911009	23.0%

## Most occurring characters

Value	Count	Frequency (%)
r	5742453	16.5%
e	4492028	12.9%
t	3250914	9.4%
o	2989078	8.6%
s	2322891	6.7%
a	1899639	5.5%
	1884405	5.4%
i	1612225	4.6%
n	1349938	3.9%
p	1346217	3.9%
Other values (39)	7835177	22.6%

## Most occurring categories

Value	Count	Frequency (%)
Lowercase Letter	30509716	87.9%
Uppercase Letter	2176356	6.3%
Space Separator	1884405	5.4%
Other Punctuation	154488	0.4%

## Most frequent character per category

Value	Count	Frequency (%)
r	5742453	18.8%
e	4492028	14.7%
t	3250914	10.7%
o	2989078	9.8%
s	2322891	7.6%
a	1899639	6.2%
i	1612225	5.3%
n	1349938	4.4%
p	1346217	4.4%
c	956990	3.1%
Other values (21)	4547343	14.9%

Value	Count	Frequency (%)
T	816965	37.5%
A	558303	25.7%
C	465742	21.4%
P	106535	4.9%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
M	94180	4.3%
B	33400	1.5%
E	26320	1.2%
I	19446	0.9%
Á	14100	0.6%
Q	13820	0.6%
Other values (5)	27545	1.3%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
,	150960	97.7%
.	3528	2.3%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
	1884405	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	32686072	94.1%
Common	2038893	5.9%

## Most frequent character per script

Value	Count	Frequency (%)
r	5742453	17.6%
e	4492028	13.7%
t	3250914	9.9%
o	2989078	9.1%
s	2322891	7.1%
a	1899639	5.8%
i	1612225	4.9%
n	1349938	4.1%
p	1346217	4.1%
c	956990	2.9%
Other values (36)	6723699	20.6%

Value	Count	Frequency (%)
,	1884405	92.4%
.	150960	7.4%
	3528	0.2%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	33441982	96.3%
None	1282983	3.7%

## Most frequent character per block

Value	Count	Frequency (%)
r	5742453	17.2%
e	4492028	13.4%
t	3250914	9.7%
o	2989078	8.9%
s	2322891	6.9%
a	1899639	5.7%
	1884405	5.6%
i	1612225	4.8%
n	1349938	4.0%
p	1346217	4.0%
Other values (29)	6552194	19.6%

Value	Count	Frequency (%)
á	431668	33.6%
é	322390	25.1%
ã	170946	13.3%
ç	157632	12.3%
í	111881	8.7%
ó	37862	3.0%

Value	Count	Frequency (%)
ú	23463	1.8%
Á	14100	1.1%
ê	9933	0.8%
õ	3108	0.2%





**subsetor\_cnae\_codigo**

Categorical

HIGH

CARDINALITY

<b>Distinct</b>	1640
<b>Distinct (%)</b>	0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	15.9 MiB

## Length

<b>Max length</b>	8
<b>Median length</b>	8
<b>Mean length</b>	8
<b>Min length</b>	8

## Characters and Unicode

<b>Total characters</b>	16666904	
<b>Distinct characters</b>	29	
<b>Distinct categories</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

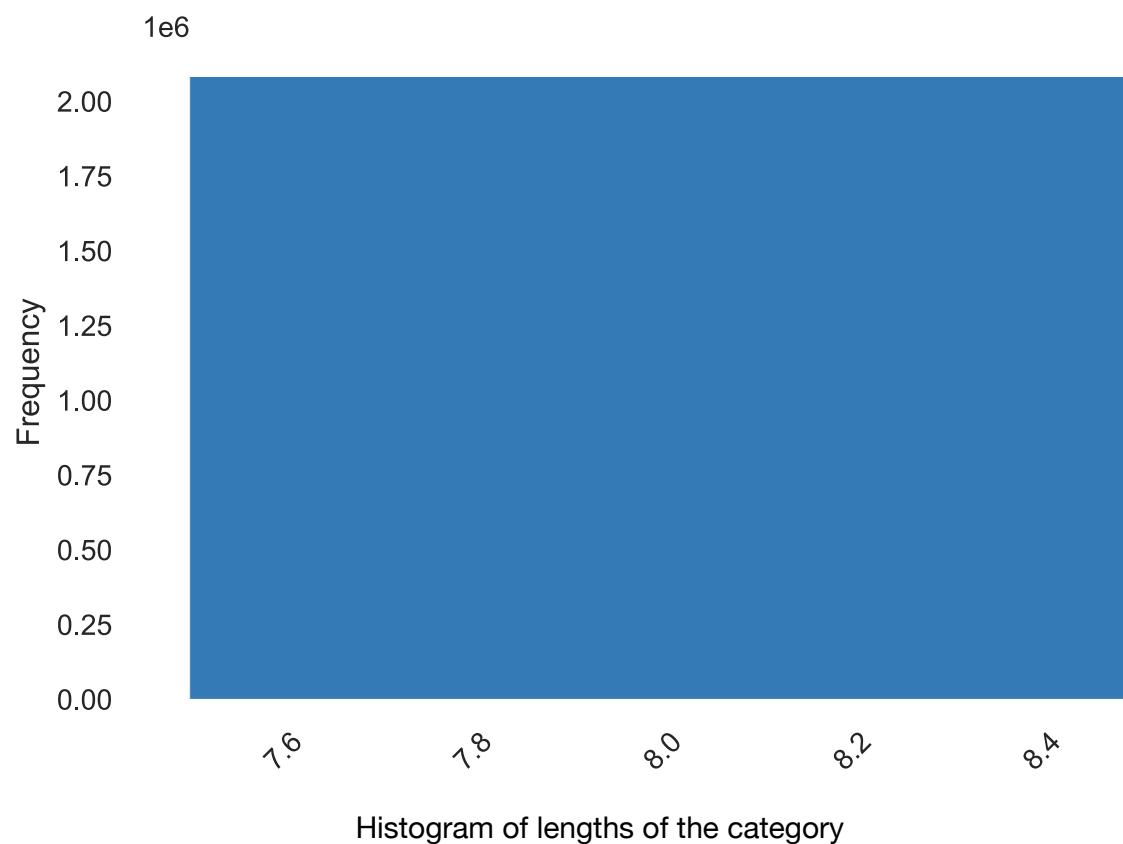
## Unique

<b>Unique</b>	54	?
<b>Unique (%)</b>	< 0.1%	

## Sample

<b>1st row</b>	A0119900
<b>2nd row</b>	A0100000
<b>3rd row</b>	A0100000
<b>4th row</b>	A0100000
<b>5th row</b>	A0100000

Value	Count	Frequency (%)
H4930202	614726	29.5%
A0100000	142064	6.8%
H4930201	35378	1.7%
H4930203	32979	1.6%
F4313400	32540	1.6%
A0115600	27244	1.3%
H5211799	26447	1.3%
G4744099	24336	1.2%
H4930200	24213	1.2%
H4921301	23875	1.1%
Other values (1630)	1099561	52.8%



Value	Count	Frequency (%)
h4930202	614726	29.5%
a0100000	142064	6.8%
h4930201	35378	1.7%
h4930203	32979	1.6%
f4313400	32540	1.6%
a0115600	27244	1.3%
h5211799	26447	1.3%
g4744099	24336	1.2%
h4930200	24213	1.2%
h4921301	23875	1.1%
Other values (1630)	1099561	52.8%

## Most occurring characters

Value	Count	Frequency (%)
0	4609952	27.7%
2	2311829	13.9%
1	1693045	10.2%
4	1646562	9.9%
9	1504208	9.0%
3	1377520	8.3%
H	849303	5.1%
7	496894	3.0%
5	399784	2.4%
A	334612	2.0%
Other values (19)	1443195	8.7%

## Most occurring categories

Value	Count	Frequency (%)
Decimal Number	14583541	87.5%
Uppercase Letter	2083363	12.5%

## Most frequent character per category

Value	Count	Frequency (%)
H	849303	40.8%
A	334612	16.1%
G	322068	15.5%
C	307353	14.8%
F	109010	5.2%
N	66673	3.2%
B	14922	0.7%
E	14100	0.7%
I	12268	0.6%
K	10452	0.5%
Other values (9)	42602	2.0%

Value	Count	Frequency (%)
0	4609952	31.6%
2	2311829	15.9%
1	1693045	11.6%
4	1646562	11.3%
9	1504208	10.3%
3	1377520	9.4%

Value	Count	Frequency (%)
7	496894	3.4%
5	399784	2.7%
6	333380	2.3%
8	210367	1.4%

## Most occurring scripts

Value	Count	Frequency (%)
Common	14583541	87.5%
Latin	2083363	12.5%

## Most frequent character per script

Value	Count	Frequency (%)
H	849303	40.8%
A	334612	16.1%
G	322068	15.5%
C	307353	14.8%
F	109010	5.2%
N	66673	3.2%
B	14922	0.7%
E	14100	0.7%
I	12268	0.6%
K	10452	0.5%
Other values (9)	42602	2.0%

Value	Count	Frequency (%)
0	4609952	31.6%
2	2311829	15.9%
1	1693045	11.6%
4	1646562	11.3%
9	1504208	10.3%
3	1377520	9.4%

Value	Count	Frequency (%)
7	496894	3.4%
5	399784	2.7%
6	333380	2.3%
8	210367	1.4%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	16666904	100.0%

## Most frequent character per block

Value	Count	Frequency (%)
0	4609952	27.7%
2	2311829	13.9%
1	1693045	10.2%
4	1646562	9.9%
9	1504208	9.0%
3	1377520	8.3%
H	849303	5.1%
7	496894	3.0%
5	399784	2.4%
A	334612	2.0%
Other values (19)	1443195	8.7%





**subsetor\_cnae\_nome**

Categorical

HIGH

CARDINALITY

<b>Distinct</b>	1604
<b>Distinct (%)</b>	0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	15.9 MiB

## Length

<b>Max length</b>	65
<b>Median length</b>	60
<b>Mean length</b>	50.5719157
<b>Min length</b>	5

## Characters and Unicode

<b>Total characters</b>	105359658	
<b>Distinct characters</b>	34	
<b>Distinct categories</b>	7 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

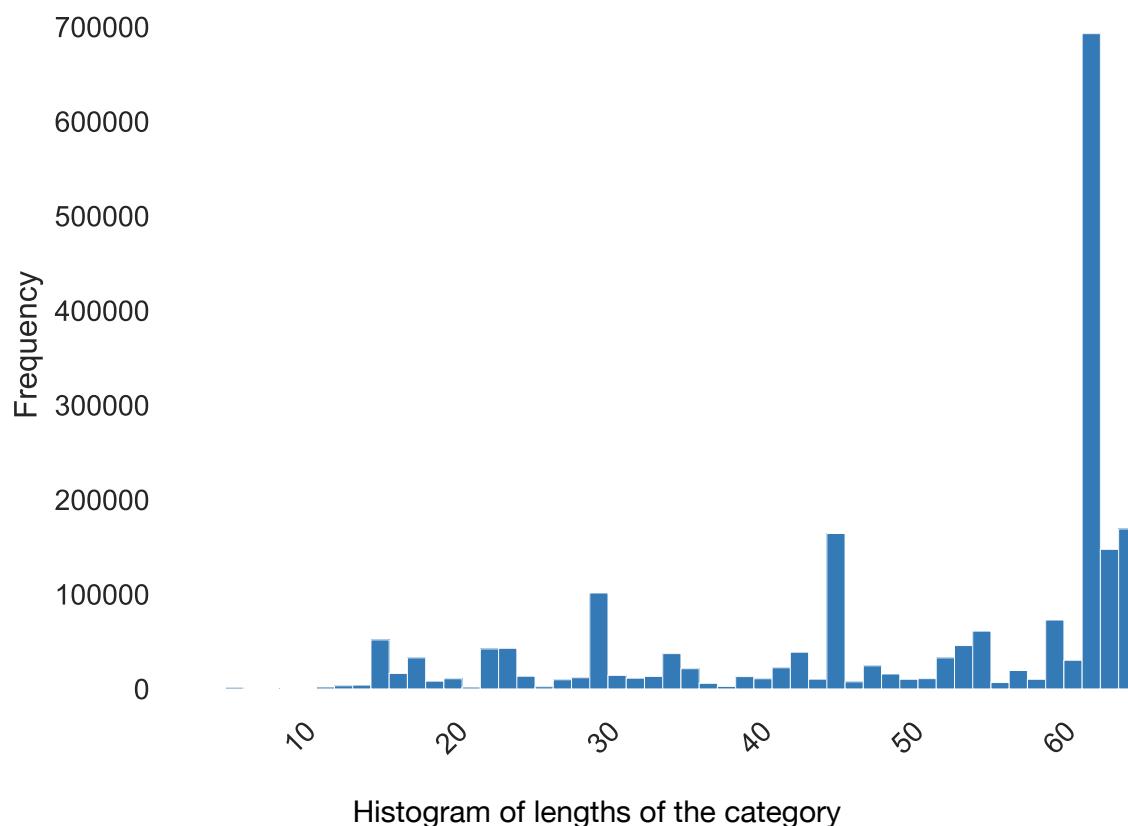
## Unique

<b>Unique</b>	51	?
<b>Unique (%)</b>	< 0.1%	

## Sample

<b>1st row</b>	CULT PLANTAS LAVOURA TEMPORARIA NAO ESPECIFICADAS ANTERIORMENTE
<b>2nd row</b>	AGRICULTURA, PECUARIA E SERVICOS RELACIONADOS
<b>3rd row</b>	AGRICULTURA, PECUARIA E SERVICOS RELACIONADOS
<b>4th row</b>	AGRICULTURA, PECUARIA E SERVICOS RELACIONADOS
<b>5th row</b>	AGRICULTURA, PECUARIA E SERVICOS RELACIONADOS

Value	Count	Frequency (%)
TRANSP ROD CARGA EXC PROD PERIG MUDANCA INTER-MUN-EST-NA...	614726	29.5%
AGRICULTURA, PECUARIA E SERVICOS RELACIONADOS	142064	6.8%
TRANSP ROD CARGA EXCETO PRODUTOS PERIGOSOS E MUDANCAS, ...	35378	1.7%
TRANSPORTE RODOVIARIO DE PRODUTOS PERIGOSOS	32979	1.6%
OBRAS DE TERRAPLENAGEM	32540	1.6%
CULTIVO DE SOJA	27244	1.3%
DEPOSITO MERCADORIA P/TERCEIRO EXC ARMAZEM GERAL GUARDA-...	26447	1.3%
COMERCIO VAREJISTA DE MATERIAIS DE CONSTRUCAO EM GERAL	24336	1.2%
TRANSPORTE RODOVIARIO DE CARGA	24213	1.2%
TRANSP RODOV COL PASSAG C/ITINERARIO FIXO, MUNICIPAL	23875	1.1%
Other values (1594)	1099561	52.8%



Value	Count	Frequency (%)
de	1113989	7.8%
transp	738944	5.2%
carga	681943	4.8%
rod	673595	4.7%
exc	671809	4.7%
prod	656788	4.6%
inter-mun-est-nacional	639738	4.5%
perig	614726	4.3%
mudanca	614726	4.3%
e	512521	3.6%
Other values (2054)	7419286	51.7%

## Most occurring characters

Value	Count	Frequency (%)
	12256299	11.6%
A	11937607	11.3%
E	9201338	8.7%
R	9146272	8.7%
O	7856703	7.5%
C	6731271	6.4%
I	6426683	6.1%
N	5893313	5.6%
S	5287903	5.0%
T	4980106	4.7%
Other values (24)	25642163	24.3%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	90471114	85.9%
Space Separator	12256299	11.6%
Dash Punctuation	2066921	2.0%
Other Punctuation	542969	0.5%
Open Punctuation	11125	< 0.1%
Close Punctuation	11125	< 0.1%
Lowercase Letter	105	< 0.1%

## Most frequent character per category

Value	Count	Frequency (%)
A	11937607	13.2%
E	9201338	10.2%
R	9146272	10.1%
O	7856703	8.7%
C	6731271	7.4%
I	6426683	7.1%
N	5893313	6.5%
S	5287903	5.8%
T	4980106	5.5%
D	4495877	5.0%
Other values (16)	18514041	20.5%

Value	Count	Frequency (%)
,	338119	62.3%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
/	204503	37.7%
.	347	0.1%
<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
	12256299	100.0%
<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
-	2066921	100.0%
<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
(	11125	100.0%
<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
)	11125	100.0%
<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
á	105	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	90471219	85.9%
Common	14888439	14.1%

## Most frequent character per script

Value	Count	Frequency (%)
A	11937607	13.2%
E	9201338	10.2%
R	9146272	10.1%
O	7856703	8.7%
C	6731271	7.4%
I	6426683	7.1%
N	5893313	6.5%
S	5287903	5.8%
T	4980106	5.5%
D	4495877	5.0%
Other values (17)	18514146	20.5%

Value	Count	Frequency (%)
	12256299	82.3%
-	2066921	13.9%
,	338119	2.3%
/	204503	1.4%
(	11125	0.1%
)	11125	0.1%

**Value****Count Frequency (%)**

347

&lt; 0.1%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	105358693	> 99.9%
None	965	< 0.1%

## Most frequent character per block

Value	Count	Frequency (%)
	12256299	11.6%
A	11937607	11.3%
E	9201338	8.7%
R	9146272	8.7%
O	7856703	7.5%
C	6731271	6.4%
I	6426683	6.1%
N	5893313	5.6%
S	5287903	5.0%
T	4980106	4.7%
Other values (22)	25641198	24.3%

Value	Count	Frequency (%)
Ç	860	89.1%
á	105	10.9%





## **setor\_bndes**

Categorical

HIGH

CORRELATION (This variable has a high correlation with 4 fields: subsetor\_cnae\_agrupado, subsetor\_bndes, modalidade\_de\_apoio, forma\_de\_apoio)

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	15.9 MiB

## Length

<b>Max length</b>	17
<b>Median length</b>	15
<b>Mean length</b>	14.12354304
<b>Min length</b>	9

## Characters and Unicode

<b>Total characters</b>	29424467	
<b>Distinct characters</b>	19	
<b>Distinct categories</b>	3 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

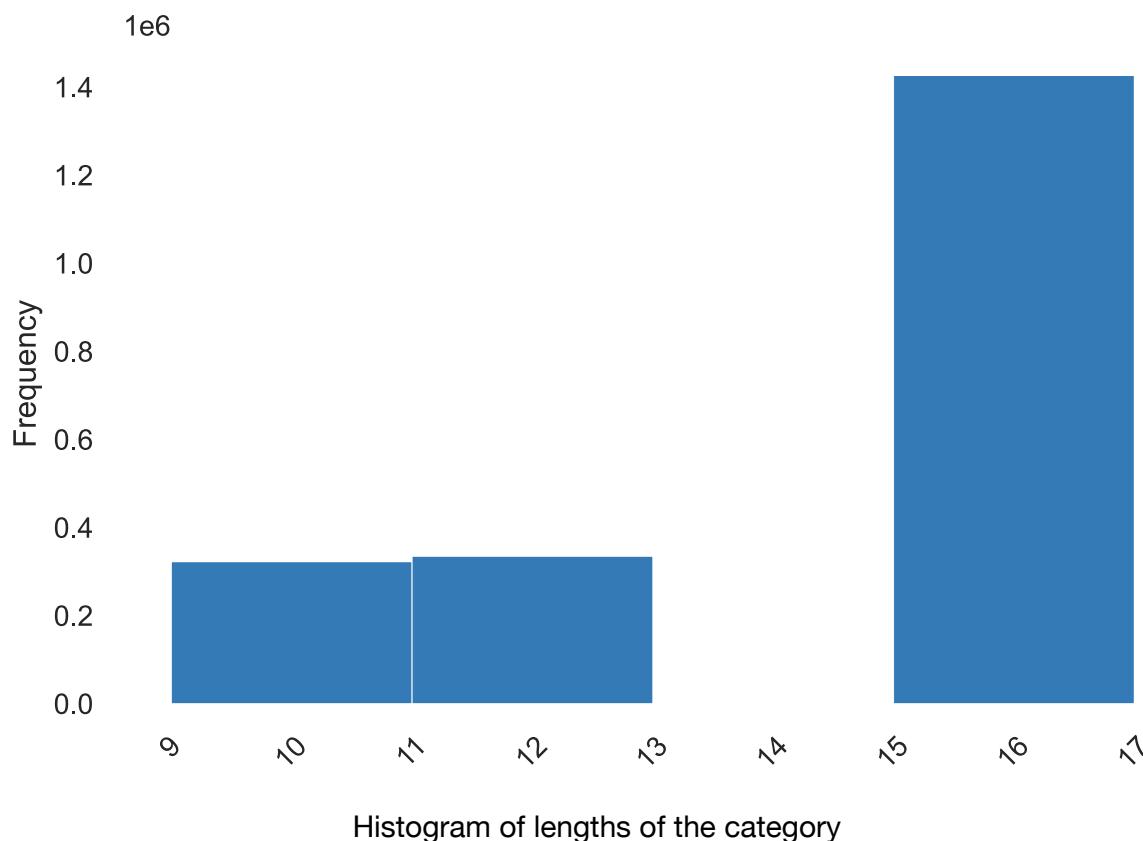
## Unique

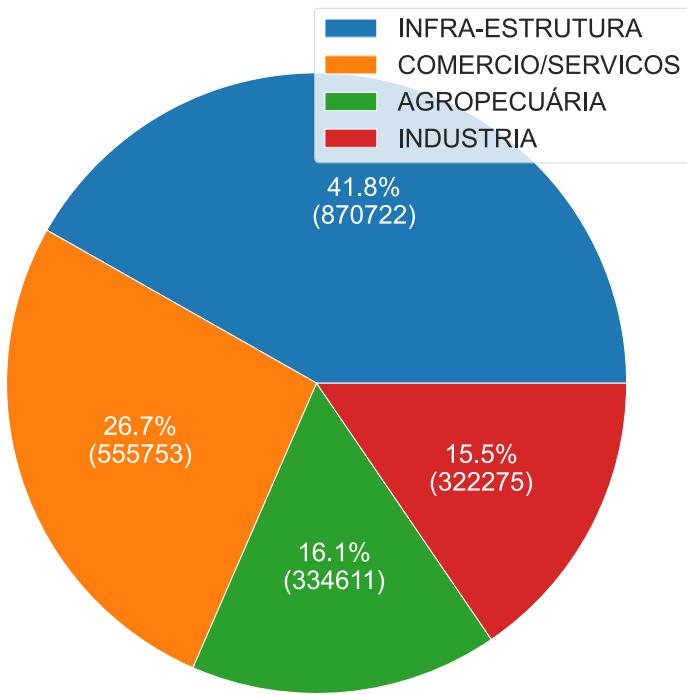
<b>Unique</b>	0	?
<b>Unique (%)</b>	0.0%	

## Sample

<b>1st row</b>	AGROPECUÁRIA
<b>2nd row</b>	AGROPECUÁRIA
<b>3rd row</b>	AGROPECUÁRIA
<b>4th row</b>	AGROPECUÁRIA
<b>5th row</b>	AGROPECUÁRIA

Value	Count	Frequency (%)
INFRA-ESTRUTURA	870722	41.8%
COMERCIO/SERVICOS	555754	26.7%
AGROPECUÁRIA	334612	16.1%
INDUSTRIA	322275	15.5%





Value	Count	Frequency (%)
infra-estrutura	870722	41.8%
comercio/servicos	555754	26.7%
agropecuária	334612	16.1%
industria	322275	15.5%

## Most occurring characters

Value	Count	Frequency (%)
R	4715173	16.0%
I	2961392	10.1%
A	2732943	9.3%
U	2398331	8.2%
E	2316842	7.9%
S	2304505	7.8%
T	2063719	7.0%
O	2001874	6.8%
C	2001874	6.8%
N	1192997	4.1%
Other values (9)	4734817	16.1%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	27997991	95.2%
Dash Punctuation	870722	3.0%
Other Punctuation	555754	1.9%

## Most frequent character per category

Value	Count	Frequency (%)
R	4715173	16.8%
I	2961392	10.6%
A	2732943	9.8%
U	2398331	8.6%
E	2316842	8.3%
S	2304505	8.2%
T	2063719	7.4%
O	2001874	7.2%
C	2001874	7.2%
N	1192997	4.3%
Other values (7)	3308341	11.8%

Value	Count	Frequency (%)
-	870722	100.0%

Value	Count	Frequency (%)
/	555754	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	27997991	95.2%
Common	1426476	4.8%

## Most frequent character per script

Value	Count	Frequency (%)
R	4715173	16.8%
I	2961392	10.6%
A	2732943	9.8%
U	2398331	8.6%
E	2316842	8.3%
S	2304505	8.2%
T	2063719	7.4%
O	2001874	7.2%
C	2001874	7.2%
N	1192997	4.3%
Other values (7)	3308341	11.8%

Value	Count	Frequency (%)
-	870722	61.0%
/	555754	39.0%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	29089855	98.9%
None	334612	1.1%

## Most frequent character per block

Value	Count	Frequency (%)
R	4715173	16.2%
I	2961392	10.2%
A	2732943	9.4%
U	2398331	8.2%
E	2316842	8.0%
S	2304505	7.9%
T	2063719	7.1%
O	2001874	6.9%
C	2001874	6.9%
N	1192997	4.1%
Other values (8)	4400205	15.1%

Value	Count	Frequency (%)
Á	334612	100.0%





## **subsetor\_bndes**

Categorical

HIGH

CORRELATION (This variable has a high correlation with 5 fields: subsetor\_cnae\_agrupado, setor\_cnae, modalidade\_de\_apoio, forma\_de\_apoio, setor\_bndes)

<b>Distinct</b>	20
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	15.9 MiB

## Length

<b>Max length</b>	23
<b>Median length</b>	19
<b>Mean length</b>	17.85249618
<b>Min length</b>	6

## Characters and Unicode

<b>Total characters</b>	37193230	
<b>Distinct characters</b>	31	
<b>Distinct categories</b>	3 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

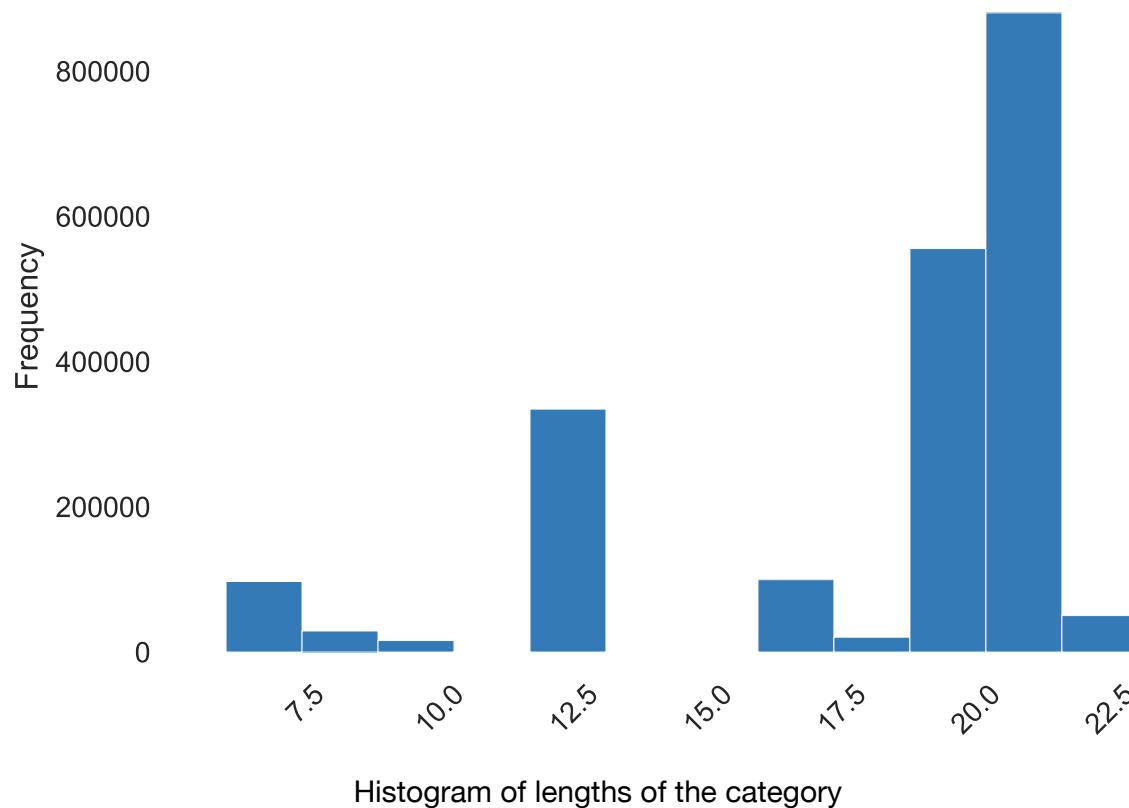
## Unique

<b>Unique</b>	0	?
<b>Unique (%)</b>	0.0%	

## Sample

<b>1st row</b>	AGROPECUÁRIA
<b>2nd row</b>	AGROPECUÁRIA
<b>3rd row</b>	AGROPECUÁRIA
<b>4th row</b>	AGROPECUÁRIA
<b>5th row</b>	AGROPECUÁRIA

Value	Count	Frequency (%)
TRANSPORTE RODOVIÁRIO	804461	38.6%
COMÉRCIO E SERVIÇOS	555754	26.7%
AGROPECUÁRIA	334612	16.1%
OUTRAS	96566	4.6%
ALIMENTO E BEBIDA	86463	4.2%
ATV. AUX. TRANSPORTES	42643	2.0%
METALURGIA E PRODUTOS	33281	1.6%
MECÂNICA	29057	1.4%
QUÍMICA E PETROQUÍMICA	25356	1.2%
TÊXTIL E VESTUÁRIO	19534	0.9%
Other values (10)	55636	2.7%



Value	Count	Frequency (%)
transporte	815024	18.2%
rodoviário	804461	18.0%
e	727555	16.2%
serviços	555754	12.4%
comércio	555754	12.4%
agropecuária	334612	7.5%
outras	96566	2.2%
bebida	86463	1.9%
alimento	86463	1.9%
transportes	43578	1.0%
Other values (23)	375109	8.4%

## Most occurring characters

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
O	5584963	15.0%
R	5364485	14.4%
I	3458341	9.3%
E	2860568	7.7%
	2397976	6.4%
S	2190240	5.9%
T	2156153	5.8%
A	2123584	5.7%
C	1586639	4.3%
V	1452304	3.9%
Other values (21)	8017977	21.6%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	34695612	93.3%
Space Separator	2397976	6.4%
Other Punctuation	99642	0.3%

## Most frequent character per category

Value	Count	Frequency (%)
O	5584963	16.1%
R	5364485	15.5%
I	3458341	10.0%
E	2860568	8.2%
S	2190240	6.3%
T	2156153	6.2%
A	2123584	6.1%
C	1586639	4.6%
V	1452304	4.2%
P	1280541	3.7%
Other values (19)	6637794	19.1%

Value	Count	Frequency (%)
.	2397976	100.0%

Value	Count	Frequency (%)
.	99642	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	34695612	93.3%
Common	2497618	6.7%

## Most frequent character per script

Value	Count	Frequency (%)
O	5584963	16.1%
R	5364485	15.5%
I	3458341	10.0%
E	2860568	8.2%
S	2190240	6.3%
T	2156153	6.2%
A	2123584	6.1%
C	1586639	4.6%
V	1452304	4.2%
P	1280541	3.7%
Other values (19)	6637794	19.1%

Value	Count	Frequency (%)
	2397976	96.0%
.	99642	4.0%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	34797678	93.6%
None	2395552	6.4%

## Most frequent character per block

Value	Count	Frequency (%)
O	5584963	16.0%
R	5364485	15.4%
I	3458341	9.9%
E	2860568	8.2%
	2397976	6.9%
S	2190240	6.3%
T	2156153	6.2%
A	2123584	6.1%
C	1586639	4.6%
V	1452304	4.2%
Other values (12)	5622425	16.2%

Value	Count	Frequency (%)
Á	1159241	48.4%
Ç	559836	23.4%
É	558734	23.3%
Í	50712	2.1%
Â	29057	1.2%
Ê	19534	0.8%

Value	Count	Frequency (%)
Ú	14356	0.6%
Õ	3108	0.1%
Ã	974	< 0.1%





## porte\_do\_cliente

Categorical

HIGH

CORRELATION (This variable has a high correlation with 2 fields: modalidade\_de\_apoio, forma\_de\_apoio)

Distinct	4
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	15.9 MiB

## Length

<b>Max length</b>	7
<b>Median length</b>	5
<b>Mean length</b>	5.586141733
<b>Min length</b>	5

## Characters and Unicode

<b>Total characters</b>	11637961	
<b>Distinct characters</b>	14	
<b>Distinct categories</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	1 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

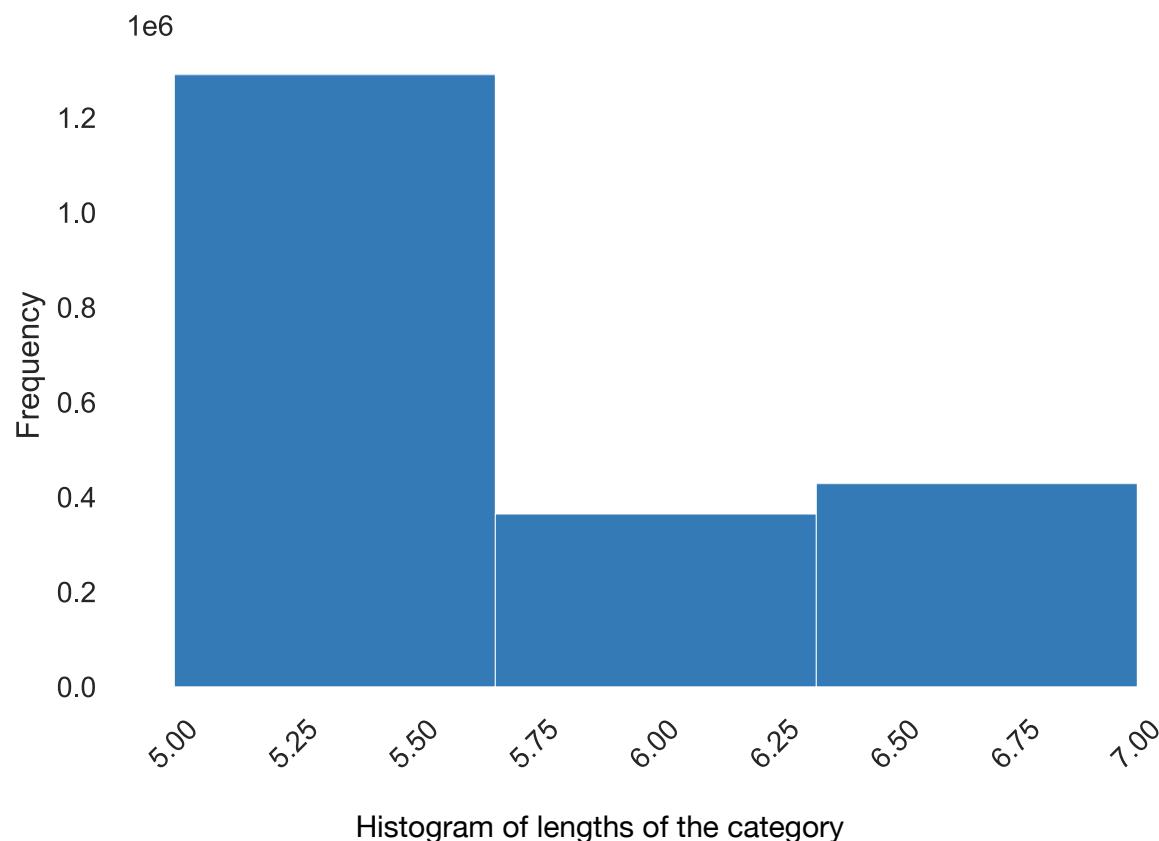
## Unique

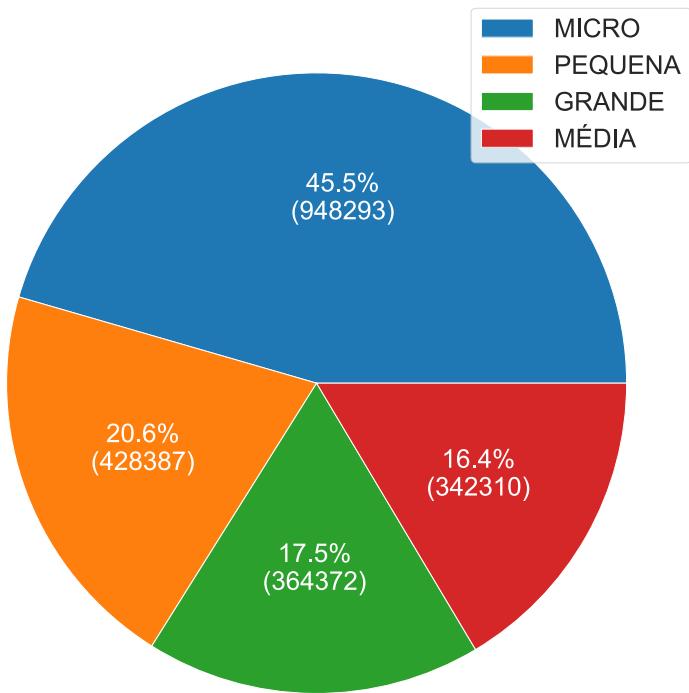
<b>Unique</b>	0	?
<b>Unique (%)</b>	0.0%	

## Sample

<b>1st row</b>	MICRO
<b>2nd row</b>	MICRO
<b>3rd row</b>	MICRO
<b>4th row</b>	MICRO
<b>5th row</b>	MICRO

Value	Count	Frequency (%)
MICRO	948294	45.5%
PEQUENA	428387	20.6%
GRANDE	364372	17.5%
MÉDIA	342310	16.4%





Value	Count	Frequency (%)
micro	948294	45.5%
pequena	428387	20.6%
grande	364372	17.5%
média	342310	16.4%

## Most occurring characters

Value	Count	Frequency (%)
R	1312666	11.3%
M	1290604	11.1%
I	1290604	11.1%
E	1221146	10.5%
A	1135069	9.8%
C	948294	8.1%
O	948294	8.1%
N	792759	6.8%
D	706682	6.1%
P	428387	3.7%
Other values (4)	1563456	13.4%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	11637961	100.0%

## Most frequent character per category

Value	Count	Frequency (%)
R	1312666	11.3%
M	1290604	11.1%
I	1290604	11.1%
E	1221146	10.5%
A	1135069	9.8%
C	948294	8.1%
O	948294	8.1%
N	792759	6.8%
D	706682	6.1%
P	428387	3.7%
Other values (4)	1563456	13.4%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	11637961	100.0%

## Most frequent character per script

Value	Count	Frequency (%)
R	1312666	11.3%
M	1290604	11.1%
I	1290604	11.1%
E	1221146	10.5%
A	1135069	9.8%
C	948294	8.1%
O	948294	8.1%
N	792759	6.8%
D	706682	6.1%
P	428387	3.7%
Other values (4)	1563456	13.4%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	11295651	97.1%
None	342310	2.9%

## Most frequent character per block

Value	Count	Frequency (%)
R	1312666	11.6%
M	1290604	11.4%
I	1290604	11.4%
E	1221146	10.8%
A	1135069	10.0%
C	948294	8.4%
O	948294	8.4%
N	792759	7.0%
D	706682	6.3%
P	428387	3.8%
Other values (3)	1221146	10.8%

Value	Count	Frequency (%)
É	342310	100.0%





## **natureza\_do\_cliente**

Categorical

HIGH

CORRELATION (This variable has a high correlation with 2 fields: modalidade\_de\_apoio, forma\_de\_apoio)

<b>Distinct</b>	5
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	15.9 MiB

## Length

<b>Max length</b>	48
<b>Median length</b>	7
<b>Mean length</b>	7.065633305
<b>Min length</b>	7

## Characters and Unicode

<b>Total characters</b>	14720279	
<b>Distinct characters</b>	23	
<b>Distinct categories</b>	3 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	2 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

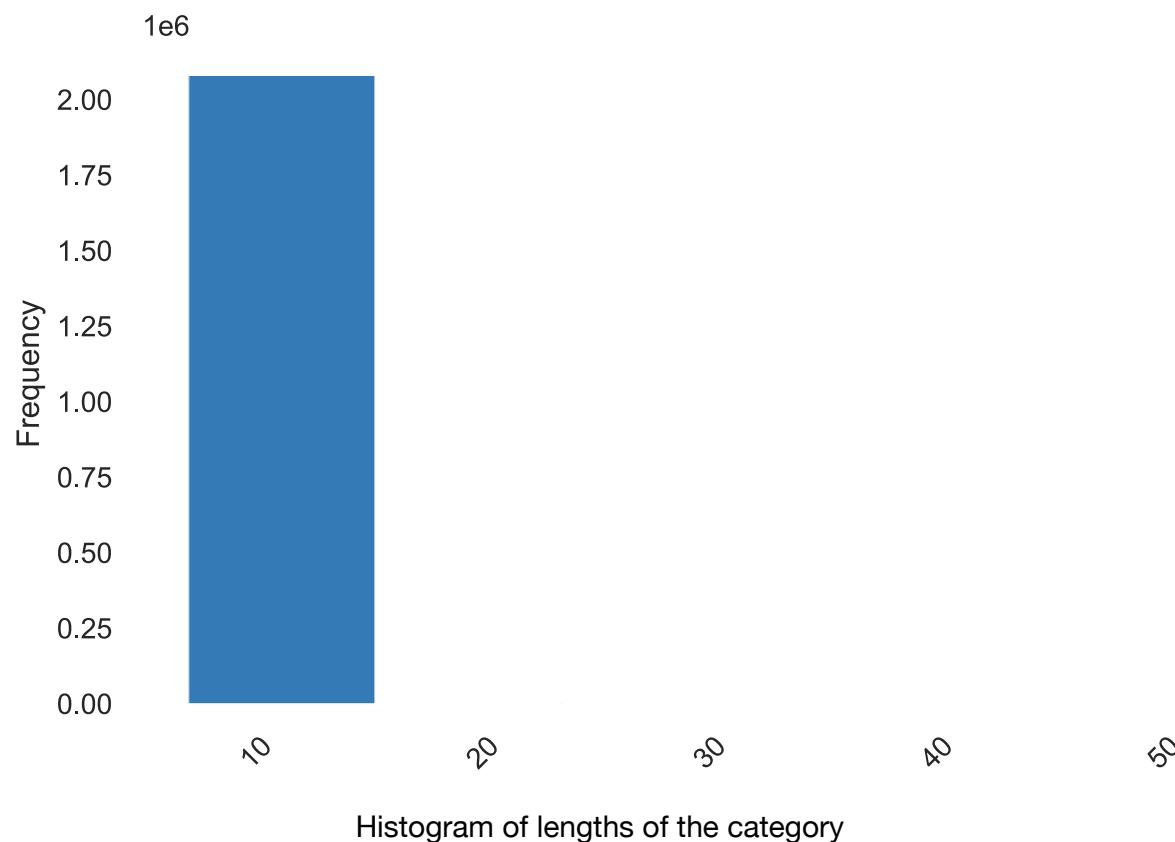
## Unique

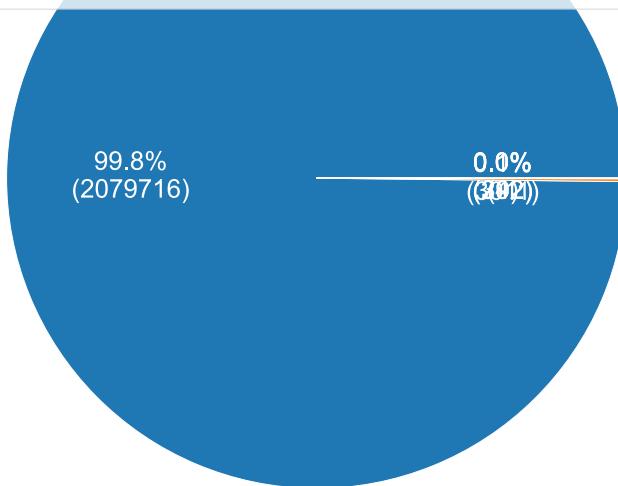
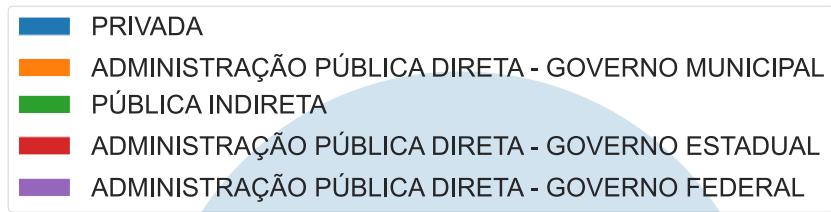
<b>Unique</b>	1	?
<b>Unique (%)</b>	< 0.1%	

## Sample

<b>1st row</b>	PRIVADA
<b>2nd row</b>	PRIVADA
<b>3rd row</b>	PRIVADA
<b>4th row</b>	PRIVADA
<b>5th row</b>	PRIVADA

Value	Count	Frequency (%)
PRIVADA	2079716	99.8%
ADMINISTRAÇÃO PÚBLICA DIRETA - GOVERNO MUNICIPAL	3011	0.1%
PÚBLICA INDIRETA	392	< 0.1%
ADMINISTRAÇÃO PÚBLICA DIRETA - GOVERNO ESTADUAL	243	< 0.1%
ADMINISTRAÇÃO PÚBLICA DIRETA - GOVERNO FEDERAL	1	< 0.1%





Value	Count	Frequency (%)
privada	2079716	99.0%
pública	3647	0.2%
direta	3255	0.2%
administração	3255	0.2%
governo	3255	0.2%
	3255	0.2%
municipal	3011	0.1%
indireta	392	< 0.1%
estadual	243	< 0.1%
federal	1	< 0.1%

## Most occurring characters

Value	Count	Frequency (%)
A	4176734	28.4%
I	2099934	14.3%
R	2089874	14.2%
D	2086862	14.2%
P	2086374	14.2%
V	2082971	14.2%
	16667	0.1%
N	9913	0.1%
O	9765	0.1%
E	7147	< 0.1%
Other values (13)	54038	0.4%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	14700357	99.9%
Space Separator	16667	0.1%
Dash Punctuation	3255	< 0.1%

## Most frequent character per category

Value	Count	Frequency (%)
A	4176734	28.4%
I	2099934	14.3%
R	2089874	14.2%
D	2086862	14.2%
P	2086374	14.2%
V	2082971	14.2%
N	9913	0.1%
O	9765	0.1%
E	7147	< 0.1%
T	7145	< 0.1%
Other values (11)	43638	0.3%

Value	Count	Frequency (%)
-	16667	100.0%

Value	Count	Frequency (%)
-	3255	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	14700357	99.9%
Common	19922	0.1%

## Most frequent character per script

Value	Count	Frequency (%)
A	4176734	28.4%
I	2099934	14.3%
R	2089874	14.2%
D	2086862	14.2%
P	2086374	14.2%
V	2082971	14.2%
N	9913	0.1%
O	9765	0.1%
E	7147	< 0.1%
T	7145	< 0.1%
Other values (11)	43638	0.3%

Value	Count	Frequency (%)
-	16667	83.7%
-	3255	16.3%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	14710122	99.9%
None	10157	0.1%

## Most frequent character per block

Value	Count	Frequency (%)
A	4176734	28.4%
I	2099934	14.3%
R	2089874	14.2%
D	2086862	14.2%
P	2086374	14.2%
V	2082971	14.2%
	16667	0.1%
N	9913	0.1%
O	9765	0.1%
E	7147	< 0.1%
Other values (10)	43881	0.3%

Value	Count	Frequency (%)
Ú	3647	35.9%
Ç	3255	32.0%
Ã	3255	32.0%





## **instituicao\_financeira\_credenciada**

Categorical

HIGH

CARDINALITY

<b>Distinct</b>	127
<b>Distinct (%)</b>	< 0.1%
<b>Missing</b>	0
<b>Missing (%)</b>	0.0%
<b>Memory size</b>	15.9 MiB

## Length

<b>Max length</b>	55
<b>Median length</b>	19
<b>Mean length</b>	24.19055057
<b>Min length</b>	8

## Characters and Unicode

<b>Total characters</b>	50397698	
<b>Distinct characters</b>	40	
<b>Distinct categories</b>	7 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	2 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

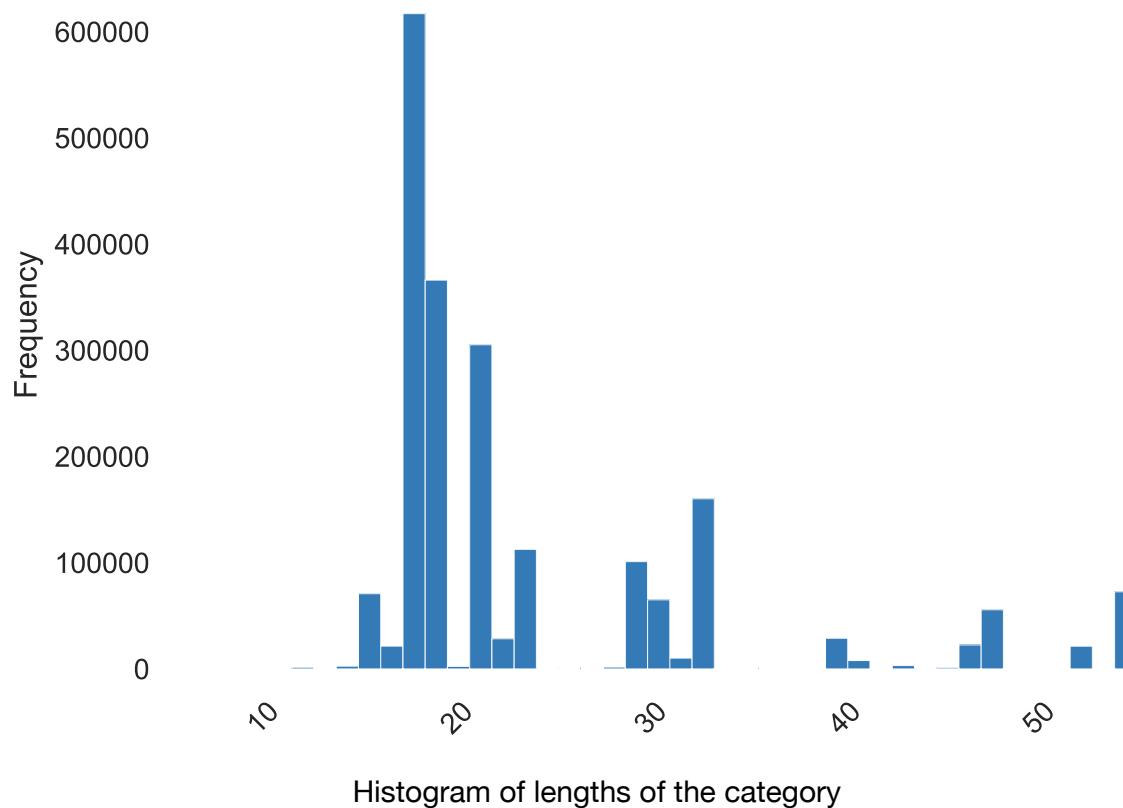
## Unique

<b>Unique</b>	2	?
<b>Unique (%)</b>	< 0.1%	

## Sample

<b>1st row</b>	BANCO COOPERATIVO SICOOB S.A.
<b>2nd row</b>	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL
<b>3rd row</b>	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL
<b>4th row</b>	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL
<b>5th row</b>	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL

Value	Count	Frequency (%)
BANCO BRADESCO S.A.	344854	16.6%
BANCO DO BRASIL SA	306240	14.7%
ITAU UNIBANCO S.A.	283149	13.6%
BANCO VOLKSWAGEN S.A.	275842	13.2%
BANCO MERCEDES-BENZ DO BRASIL S/A	125310	6.0%
BANCO SANTANDER (BRASIL) S.A.	78572	3.8%
BANCO COOPERATIVO SICREDI S.A.	62097	3.0%
BANCO VOLVO (BRASIL) S.A	60580	2.9%
BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL	55588	2.7%
CAIXA ECONOMICA FEDERAL	51357	2.5%
Other values (117)	439774	21.1%



Value	Count	Frequency (%)
banco	1642994	20.0%
s.a	1342663	16.3%
do	607943	7.4%
brasil	591896	7.2%
bradesco	350846	4.3%
sa	344800	4.2%
unibanco	283149	3.4%
itau	283149	3.4%
volkswagen	275842	3.4%
de	190450	2.3%
Other values (206)	2310752	28.1%

## Most occurring characters

Value	Count	Frequency (%)
A	6811382	13.5%
	6141121	12.2%
O	4567912	9.1%
S	3921239	7.8%
N	3575376	7.1%
C	3151933	6.3%
B	3098082	6.1%
E	2710947	5.4%
.	2623713	5.2%
I	2231513	4.4%
Other values (30)	11564480	22.9%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	40521599	80.4%
Space Separator	6141121	12.2%
Other Punctuation	2815011	5.6%
Decimal Number	372723	0.7%
Open Punctuation	186091	0.4%
Close Punctuation	186091	0.4%
Dash Punctuation	175062	0.3%

## Most frequent character per category

Value	Count	Frequency (%)
A	6811382	16.8%
O	4567912	11.3%
S	3921239	9.7%
N	3575376	8.8%
C	3151933	7.8%
B	3098082	7.6%
E	2710947	6.7%
I	2231513	5.5%
R	2166365	5.3%
D	1818664	4.5%
Other values (16)	6468186	16.0%

Value	Count	Frequency (%)
1	137982	37.0%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
0	91420	24.5%
7	91420	24.5%
4	44858	12.0%
8	3408	0.9%
2	1931	0.5%
9	1704	0.5%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
.	2623713	93.2%
/	189187	6.7%
,	2111	0.1%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
)	6141121	100.0%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
(	186091	100.0%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
)	186091	100.0%

<b>Value</b>	<b>Count</b>	<b>Frequency (%)</b>
-	175062	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	40521599	80.4%
Common	9876099	19.6%

## Most frequent character per script

Value	Count	Frequency (%)
A	6811382	16.8%
O	4567912	11.3%
S	3921239	9.7%
N	3575376	8.8%
C	3151933	7.8%
B	3098082	7.6%
E	2710947	6.7%
I	2231513	5.5%
R	2166365	5.3%
D	1818664	4.5%
Other values (16)	6468186	16.0%

Value	Count	Frequency (%)
.	6141121	62.2%
-	2623713	26.6%
/	189187	1.9%
(	186091	1.9%
)	186091	1.9%
_	175062	1.8%

Value	Count	Frequency (%)
1	137982	1.4%
0	91420	0.9%
7	91420	0.9%
4	44858	0.5%
Other values (4)	9154	0.1%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	50397698	100.0%

## Most frequent character per block

Value	Count	Frequency (%)
A	6811382	13.5%
	6141121	12.2%
O	4567912	9.1%
S	3921239	7.8%
N	3575376	7.1%
C	3151933	6.3%
B	3098082	6.1%
E	2710947	5.4%
.	2623713	5.2%
I	2231513	4.4%
Other values (30)	11564480	22.9%





## cnpj\_do\_agente\_financeiro

Categorical

HIGH

CARDINALITY

Distinct	127
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	15.9 MiB

## Length

<b>Max length</b>	18
<b>Median length</b>	18
<b>Mean length</b>	18
<b>Min length</b>	18

## Characters and Unicode

<b>Total characters</b>	37500534	
<b>Distinct characters</b>	12	
<b>Distinct categories</b>	3 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	1 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

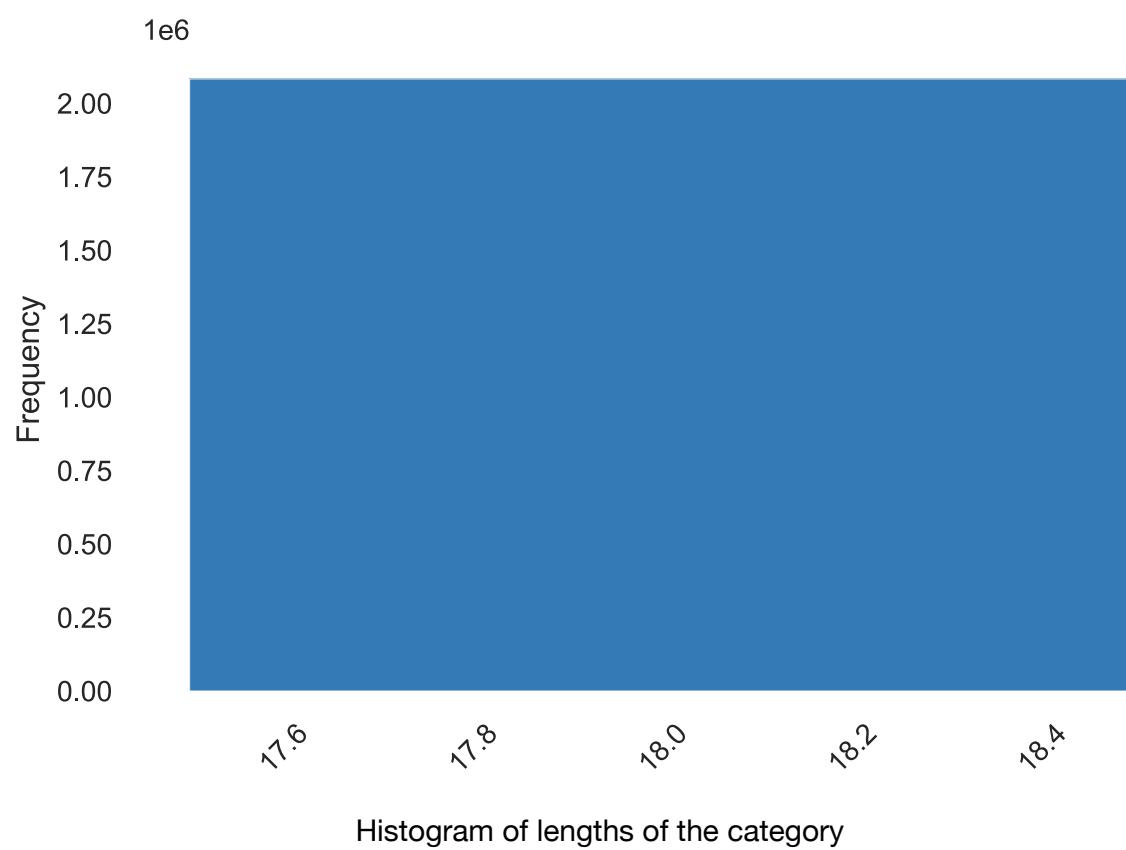
## Unique

<b>Unique</b>	2	?
<b>Unique (%)</b>	< 0.1%	

## Sample

<b>1st row</b>	02.038.232.0001-64
<b>2nd row</b>	92.816.560.0001-37
<b>3rd row</b>	92.816.560.0001-37
<b>4th row</b>	92.816.560.0001-37
<b>5th row</b>	92.816.560.0001-37

Value	Count	Frequency (%)
60.746.948.0001-12	344854	16.6%
00.000.000.0001-91	306240	14.7%
60.701.190.0001-04	283149	13.6%
59.109.165.0001-49	275842	13.2%
60.814.191.0001-57	125310	6.0%
90.400.888.0001-42	78572	3.8%
01.181.521.0001-55	62097	3.0%
58.017.179.0001-70	60580	2.9%
92.816.560.0001-37	55588	2.7%
00.360.305.0001-04	51357	2.5%
Other values (117)	439774	21.1%



Value	Count	Frequency (%)
60.746.948.0001-12	344854	16.6%
00.000.000.0001-91	306240	14.7%
60.701.190.0001-04	283149	13.6%
59.109.165.0001-49	275842	13.2%
60.814.191.0001-57	125310	6.0%
90.400.888.0001-42	78572	3.8%
01.181.521.0001-55	62097	3.0%
58.017.179.0001-70	60580	2.9%
92.816.560.0001-37	55588	2.7%
00.360.305.0001-04	51357	2.5%
Other values (117)	439774	21.1%

## Most occurring characters

Value	Count	Frequency (%)
0	12018596	32.0%
.	6250089	16.7%
1	5169293	13.8%
9	2359146	6.3%
-	2083363	5.6%
6	1943908	5.2%
4	1927806	5.1%
7	1521249	4.1%
5	1383951	3.7%
8	1269832	3.4%
Other values (2)	1573301	4.2%

## Most occurring categories

Value	Count	Frequency (%)
Decimal Number	29167082	77.8%
Other Punctuation	6250089	16.7%
Dash Punctuation	2083363	5.6%

## Most frequent character per category

Value	Count	Frequency (%)
0	12018596	41.2%
1	5169293	17.7%
9	2359146	8.1%
6	1943908	6.7%
4	1927806	6.6%
7	1521249	5.2%
5	1383951	4.7%
8	1269832	4.4%
2	1058976	3.6%
3	514325	1.8%

Value	Count	Frequency (%)
.	6250089	100.0%

Value	Count	Frequency (%)
-	2083363	100.0%

## Most occurring scripts

Value	Count	Frequency (%)
Common	37500534	100.0%

## Most frequent character per script

Value	Count	Frequency (%)
0	12018596	32.0%
.	6250089	16.7%
1	5169293	13.8%
9	2359146	6.3%
-	2083363	5.6%
6	1943908	5.2%
4	1927806	5.1%
7	1521249	4.1%
5	1383951	3.7%
8	1269832	3.4%
Other values (2)	1573301	4.2%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	37500534	100.0%

## Most frequent character per block

Value	Count	Frequency (%)
0	12018596	32.0%
.	6250089	16.7%
1	5169293	13.8%
9	2359146	6.3%
-	2083363	5.6%
6	1943908	5.2%
4	1927806	5.1%
7	1521249	4.1%
5	1383951	3.7%
8	1269832	3.4%
Other values (2)	1573301	4.2%





## situacao\_da\_operacao

Categorical

HIGH

CORRELATION (This variable has a high correlation with 2 fields: modalidade\_de\_apoio, forma\_de\_apoio)

Distinct	2
Distinct (%)	< 0.1%
Missing	0
Missing (%)	0.0%
Memory size	15.9 MiB

## Length

<b>Max length</b>	9
<b>Median length</b>	9
<b>Mean length</b>	8.513459728
<b>Min length</b>	5

## Characters and Unicode

<b>Total characters</b>	17736627	
<b>Distinct characters</b>	8	
<b>Distinct categories</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_character_property#General_Category">https://en.wikipedia.org/wiki/Unicode_character_property#General_Category</a> )	?
<b>Distinct scripts</b>	1 ( <a href="https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode">https://en.wikipedia.org/wiki/Script_(Unicode)#List_of_scripts_in_Unicode</a> )	?
<b>Distinct blocks</b>	1 ( <a href="https://en.wikipedia.org/wiki/Unicode_block">https://en.wikipedia.org/wiki/Unicode_block</a> )	?

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

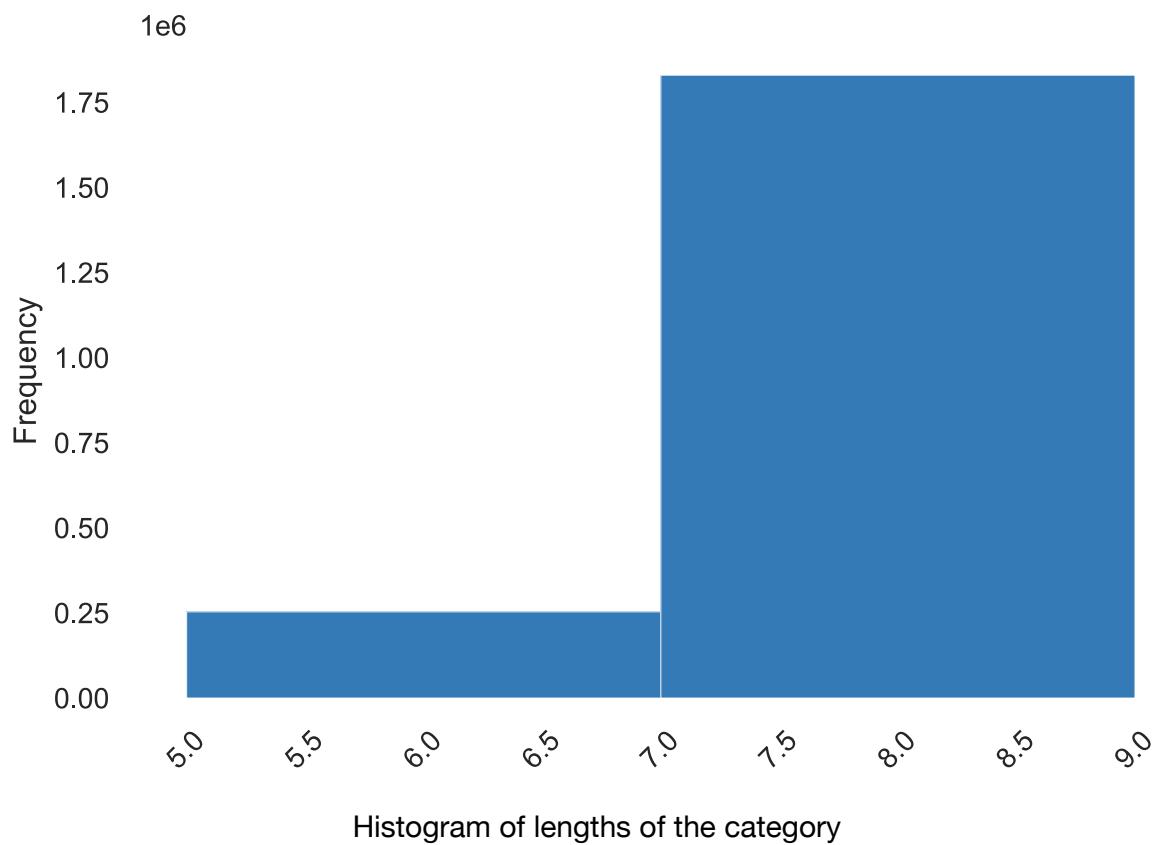
## Unique

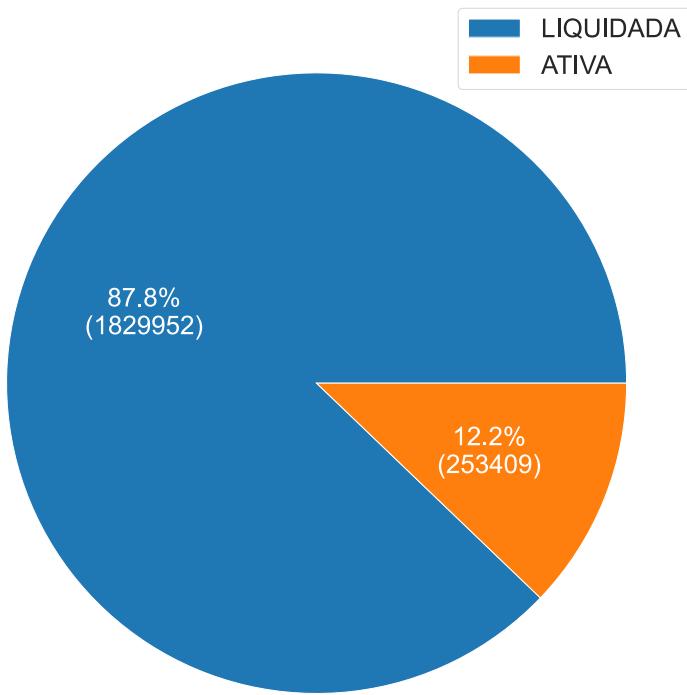
<b>Unique</b>	0	?
<b>Unique (%)</b>	0.0%	

## Sample

<b>1st row</b>	LIQUIDADA
<b>2nd row</b>	LIQUIDADA
<b>3rd row</b>	LIQUIDADA
<b>4th row</b>	LIQUIDADA
<b>5th row</b>	LIQUIDADA

Value	Count	Frequency (%)
LIQUIDADA	1829953	87.8%
ATIVA	253410	12.2%





Value	Count	Frequency (%)
liquidada	1829953	87.8%
ativa	253410	12.2%

## Most occurring characters

Value	Count	Frequency (%)
A	4166726	23.5%
I	3913316	22.1%
D	3659906	20.6%
L	1829953	10.3%
Q	1829953	10.3%
U	1829953	10.3%
T	253410	1.4%
V	253410	1.4%

## Most occurring categories

Value	Count	Frequency (%)
Uppercase Letter	17736627	100.0%

## Most frequent character per category

Value	Count	Frequency (%)
A	4166726	23.5%
I	3913316	22.1%
D	3659906	20.6%
L	1829953	10.3%
Q	1829953	10.3%
U	1829953	10.3%
T	253410	1.4%
V	253410	1.4%

## Most occurring scripts

Value	Count	Frequency (%)
Latin	17736627	100.0%

## Most frequent character per script

Value	Count	Frequency (%)
A	4166726	23.5%
I	3913316	22.1%
D	3659906	20.6%
L	1829953	10.3%
Q	1829953	10.3%
U	1829953	10.3%
T	253410	1.4%
V	253410	1.4%

## Most occurring blocks

Value	Count	Frequency (%)
ASCII	17736627	100.0%

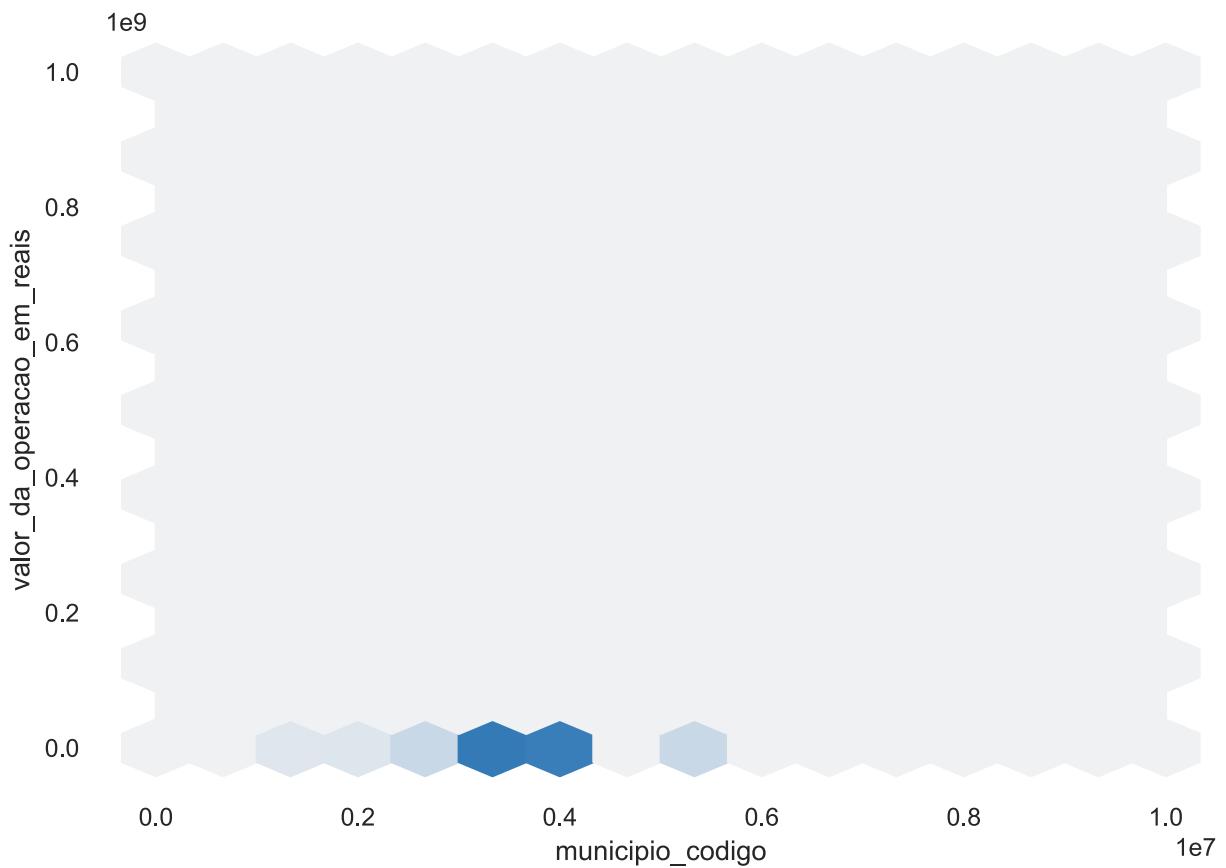
## Most frequent character per block

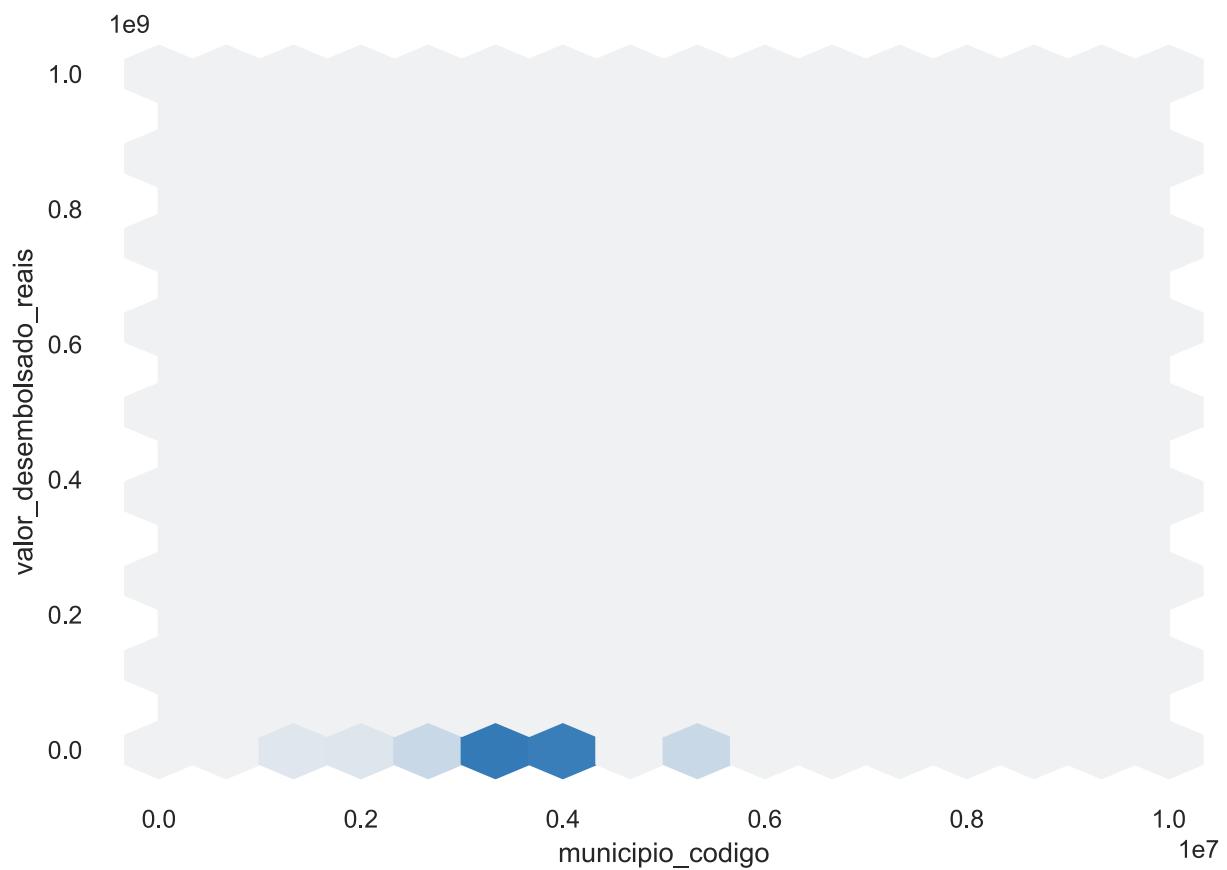
Value	Count	Frequency (%)
A	4166726	23.5%
I	3913316	22.1%
D	3659906	20.6%
L	1829953	10.3%
Q	1829953	10.3%
U	1829953	10.3%
T	253410	1.4%
V	253410	1.4%

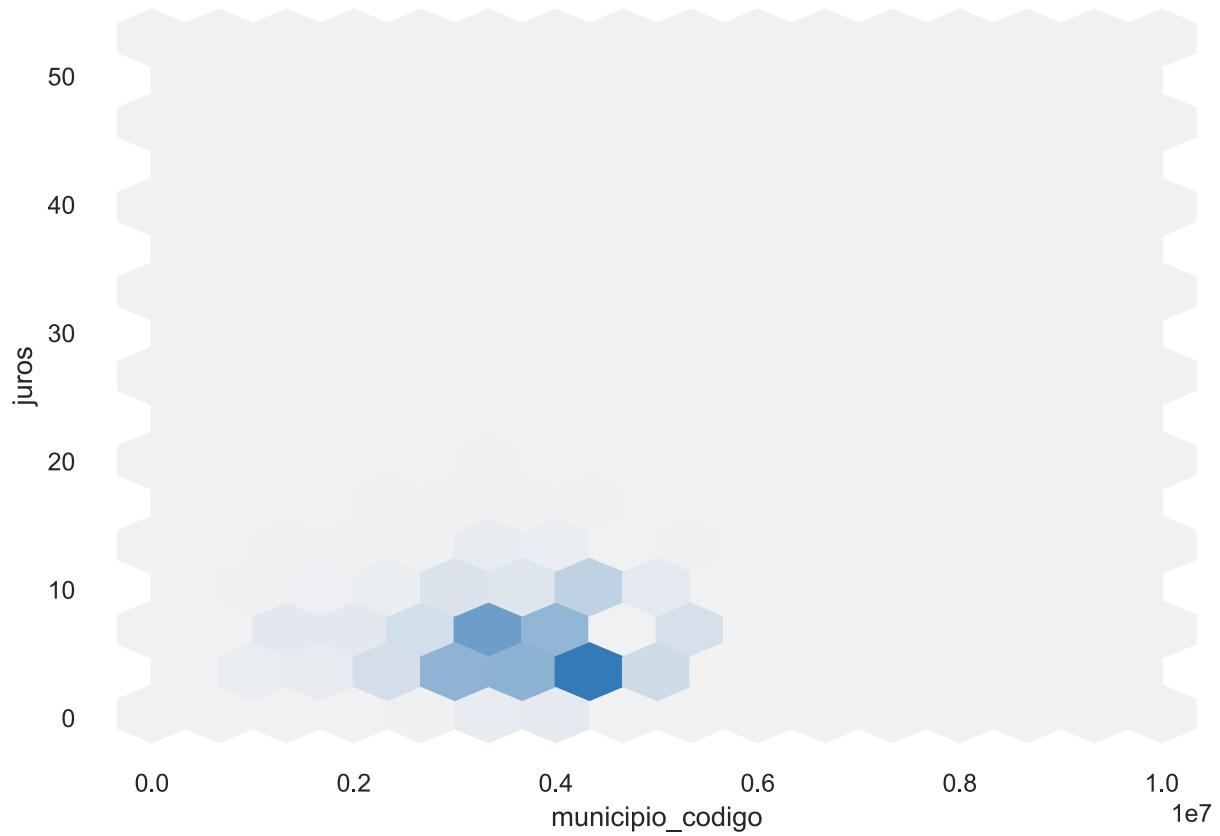


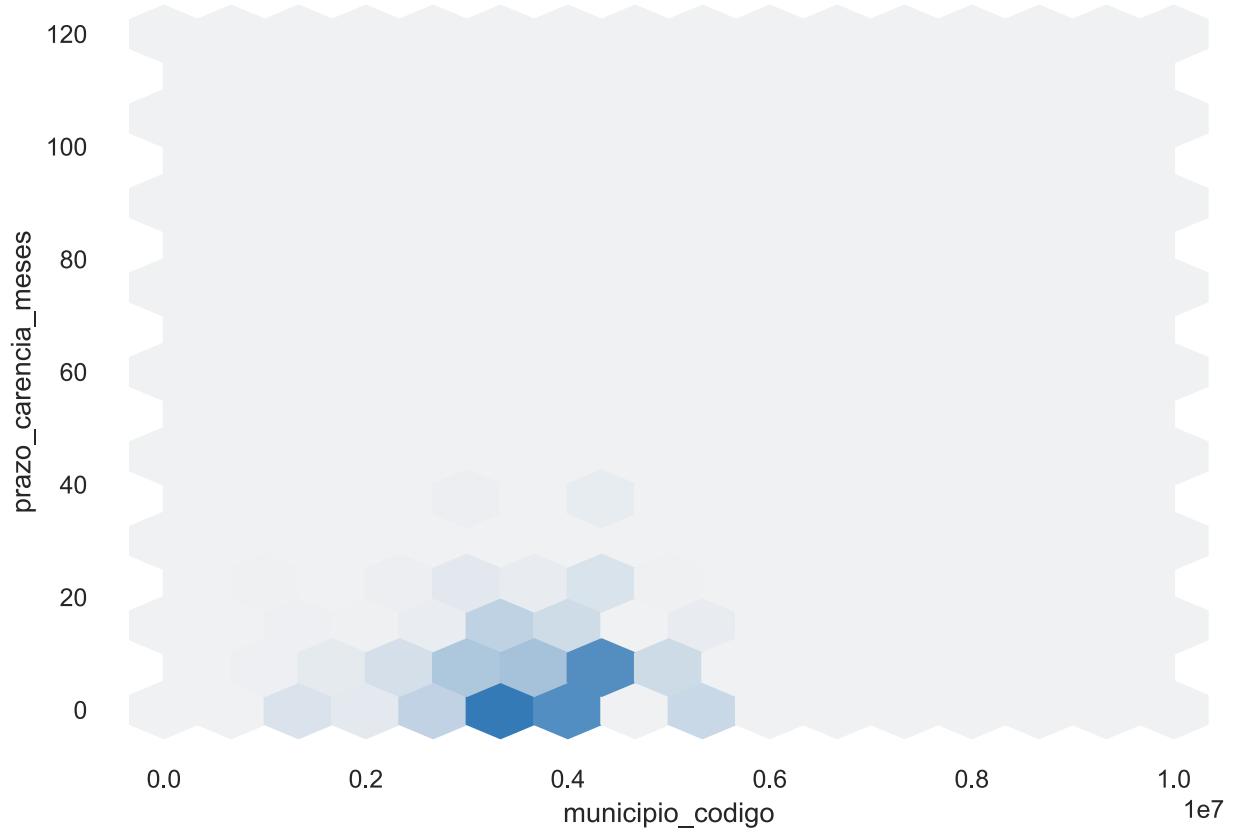


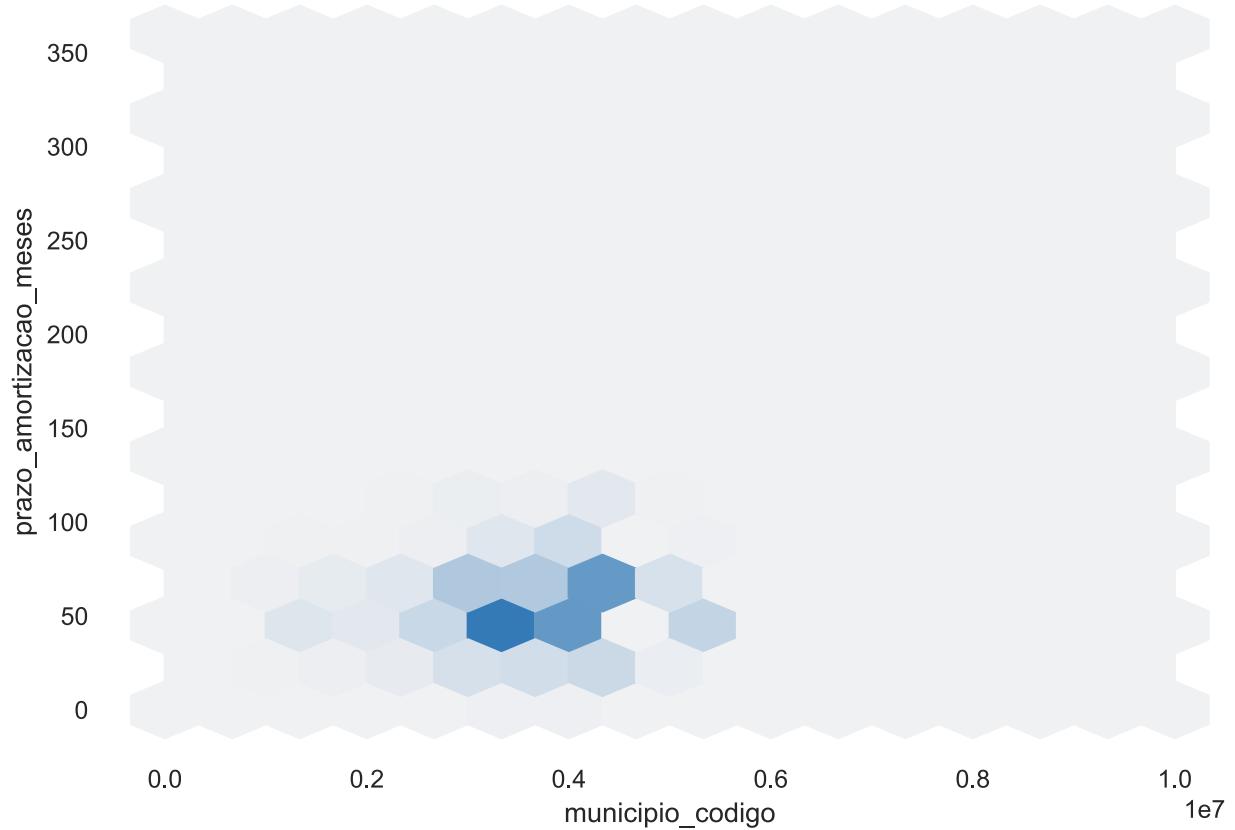
# Interactions



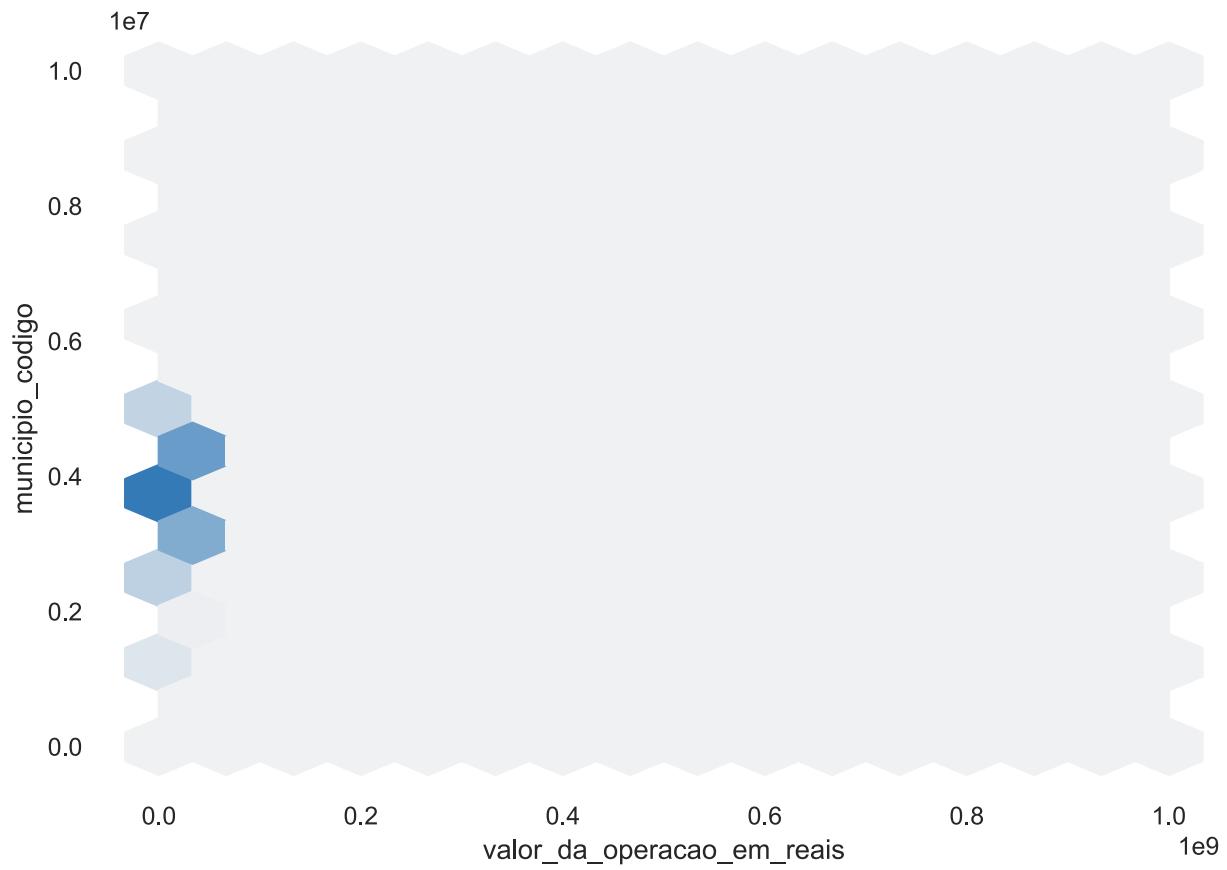




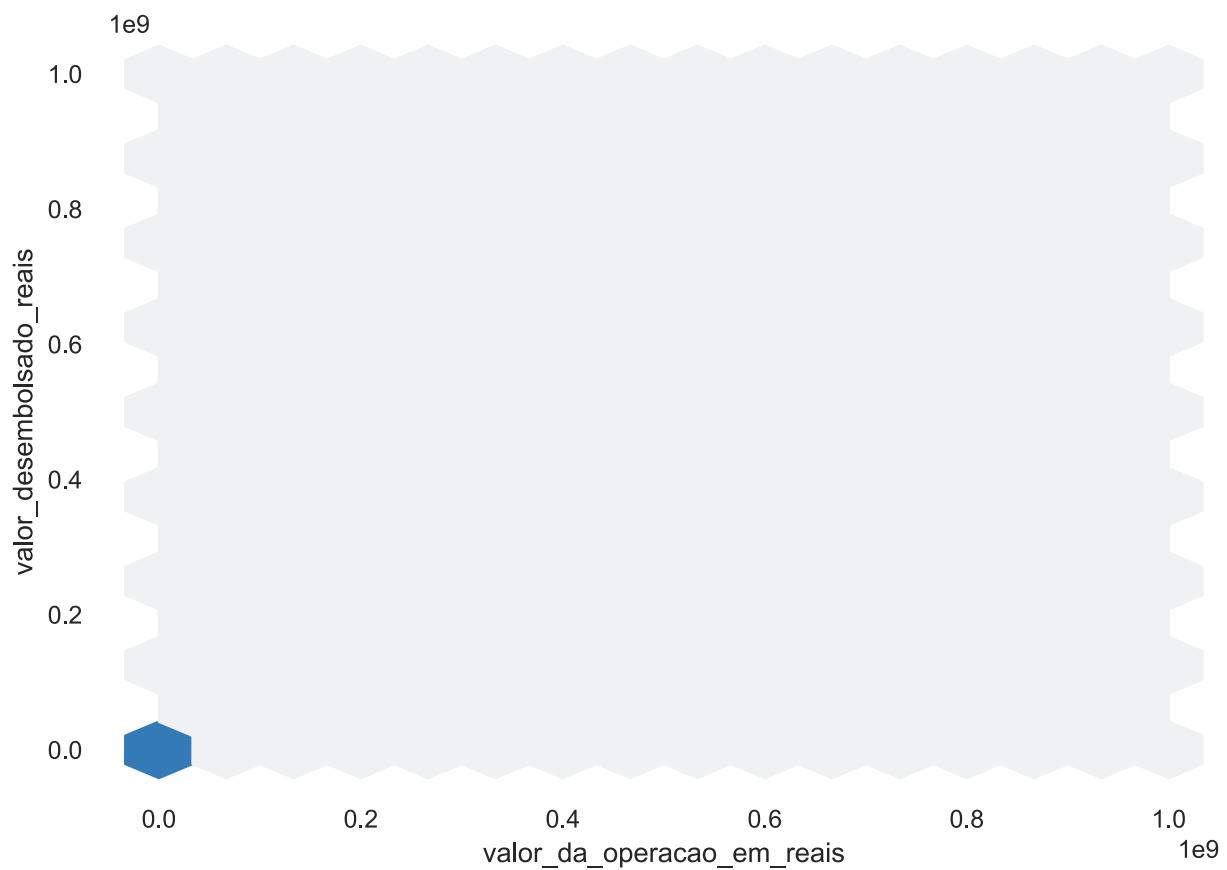


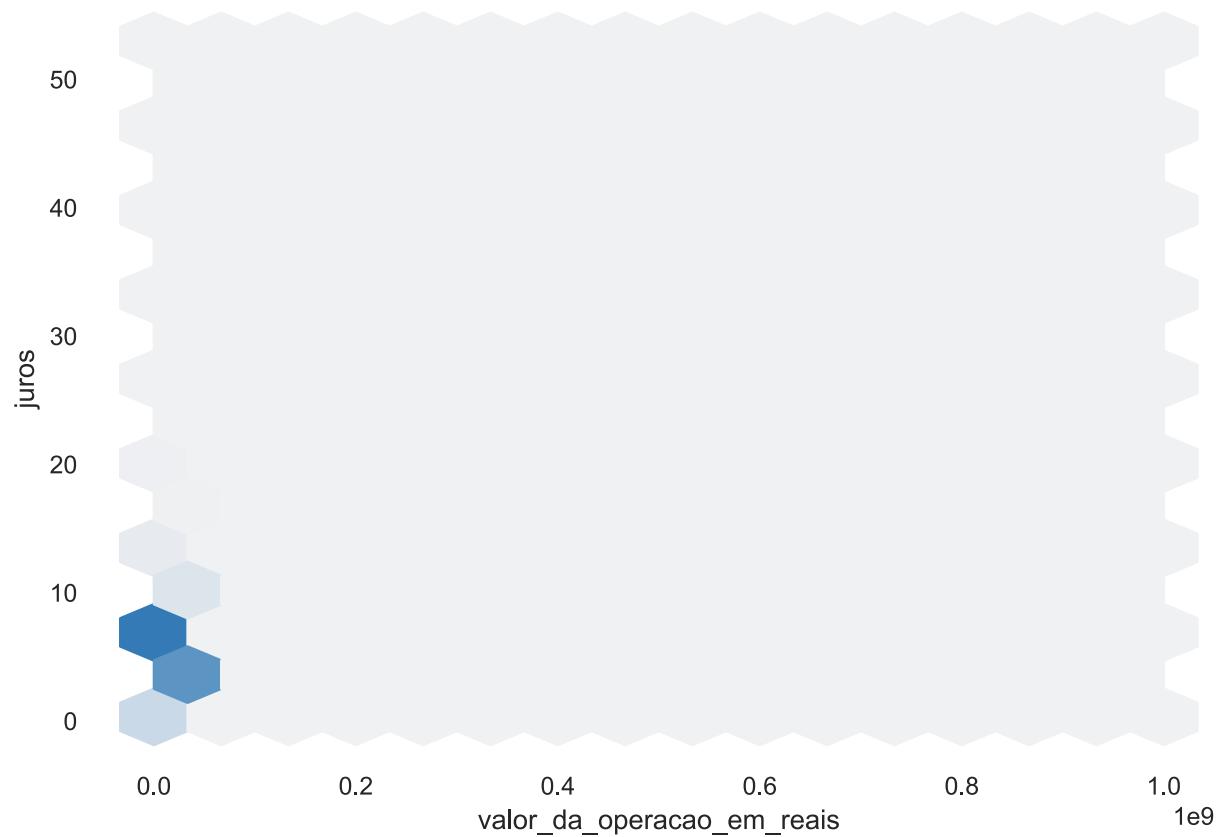


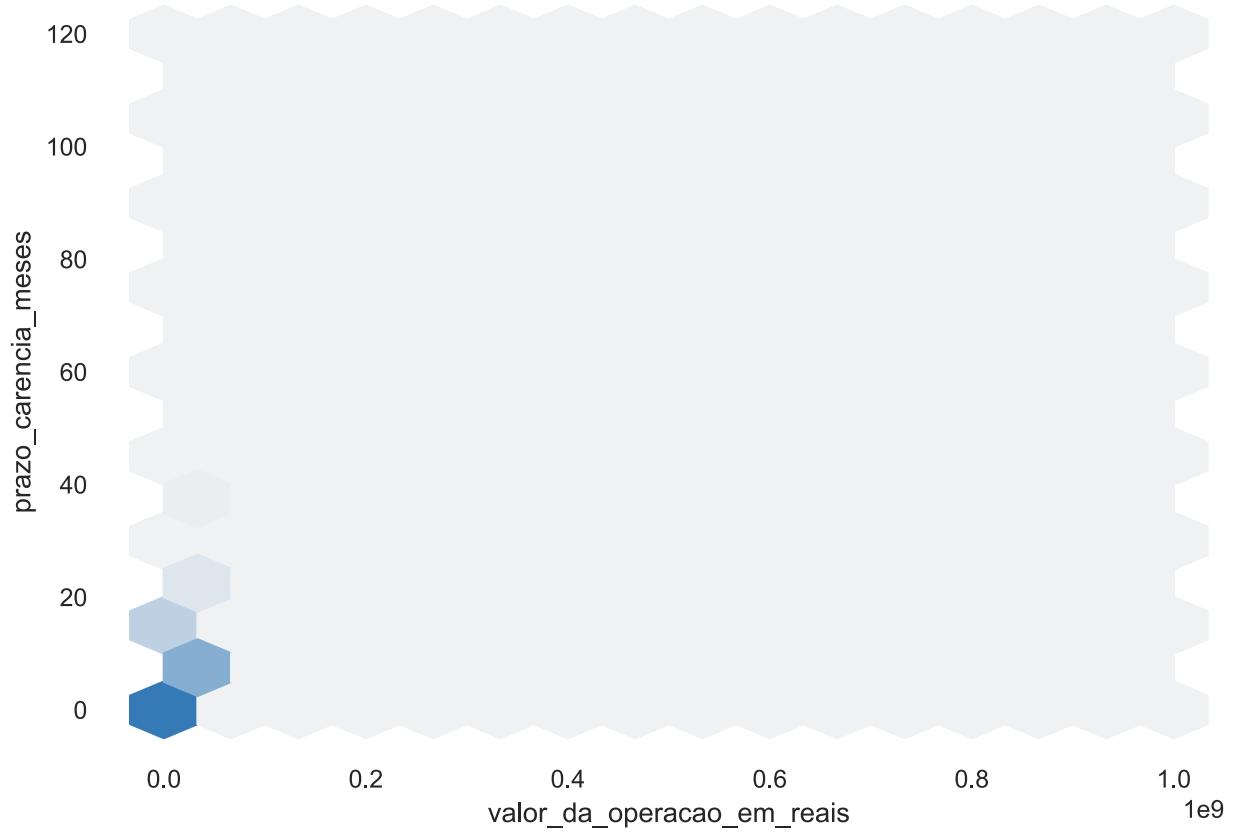


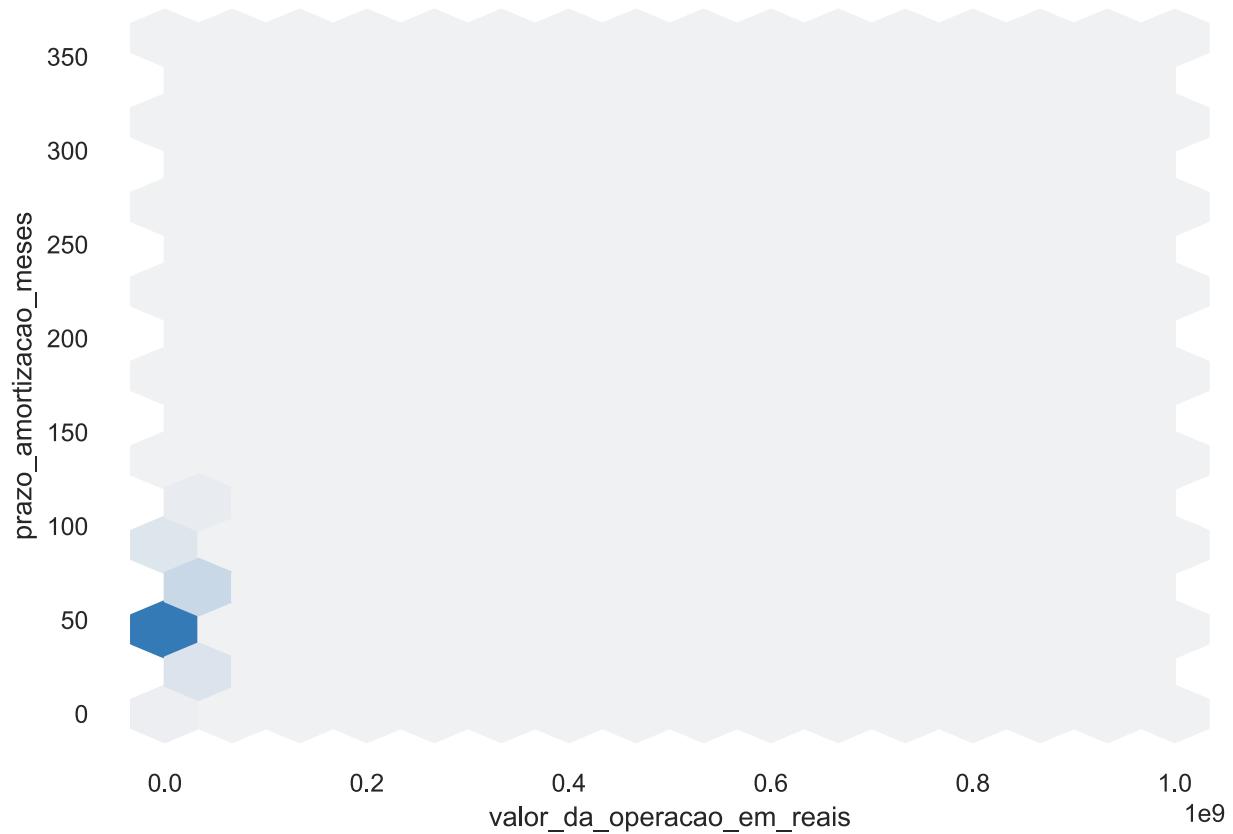




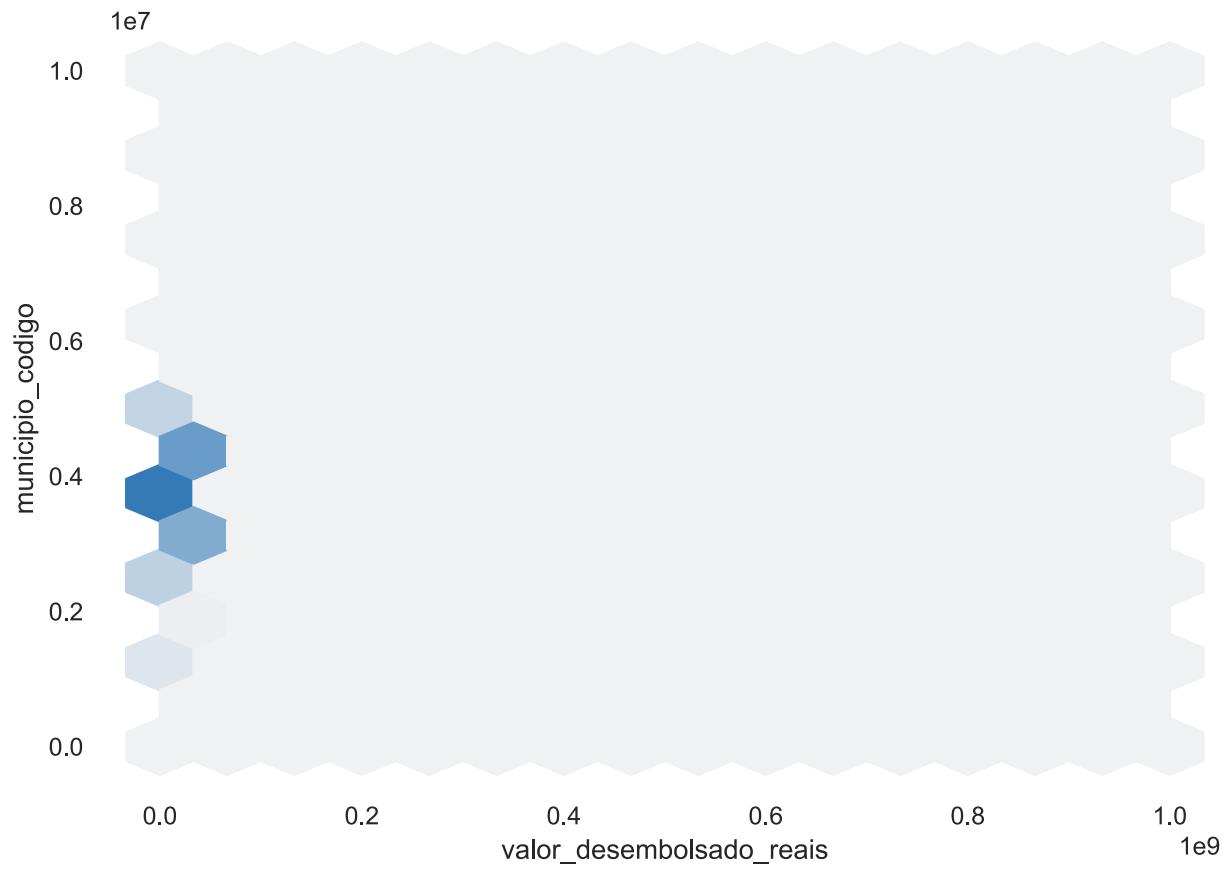


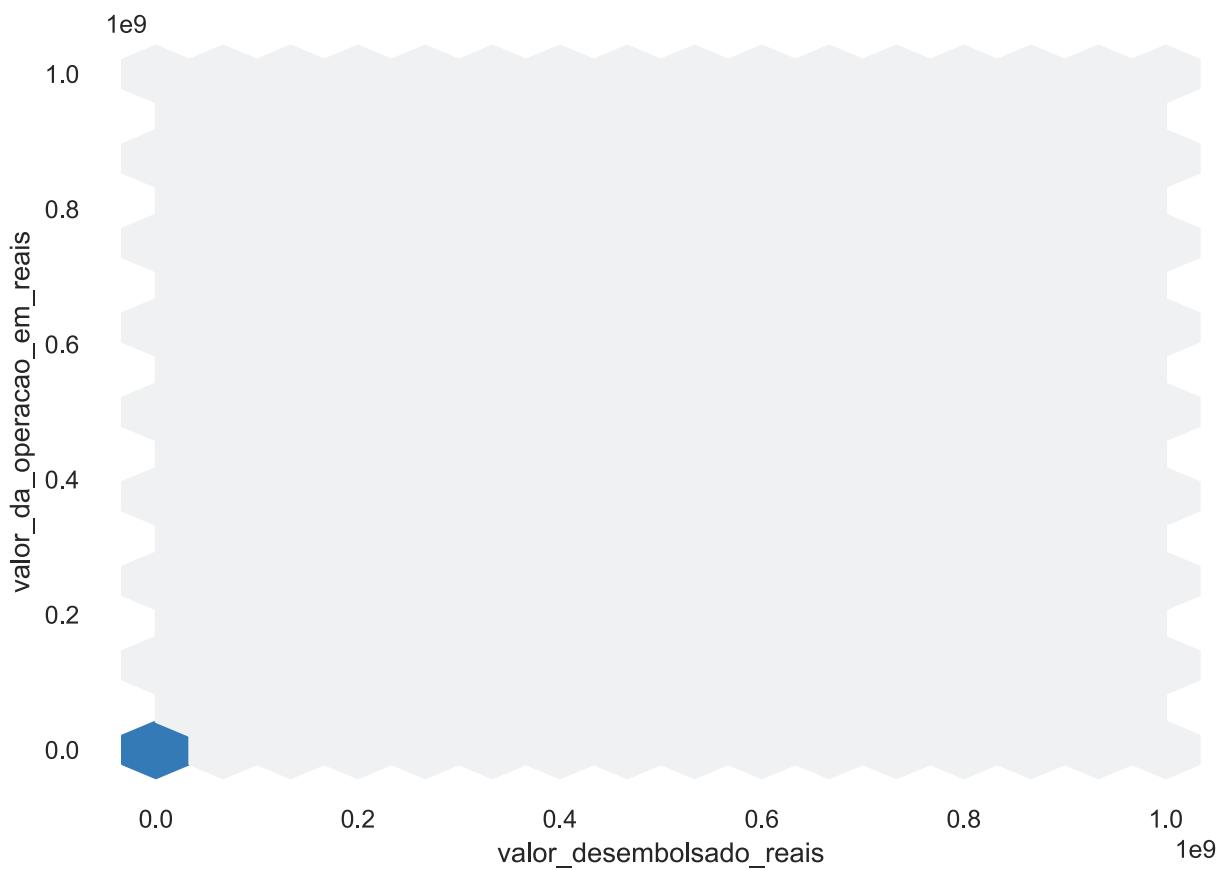




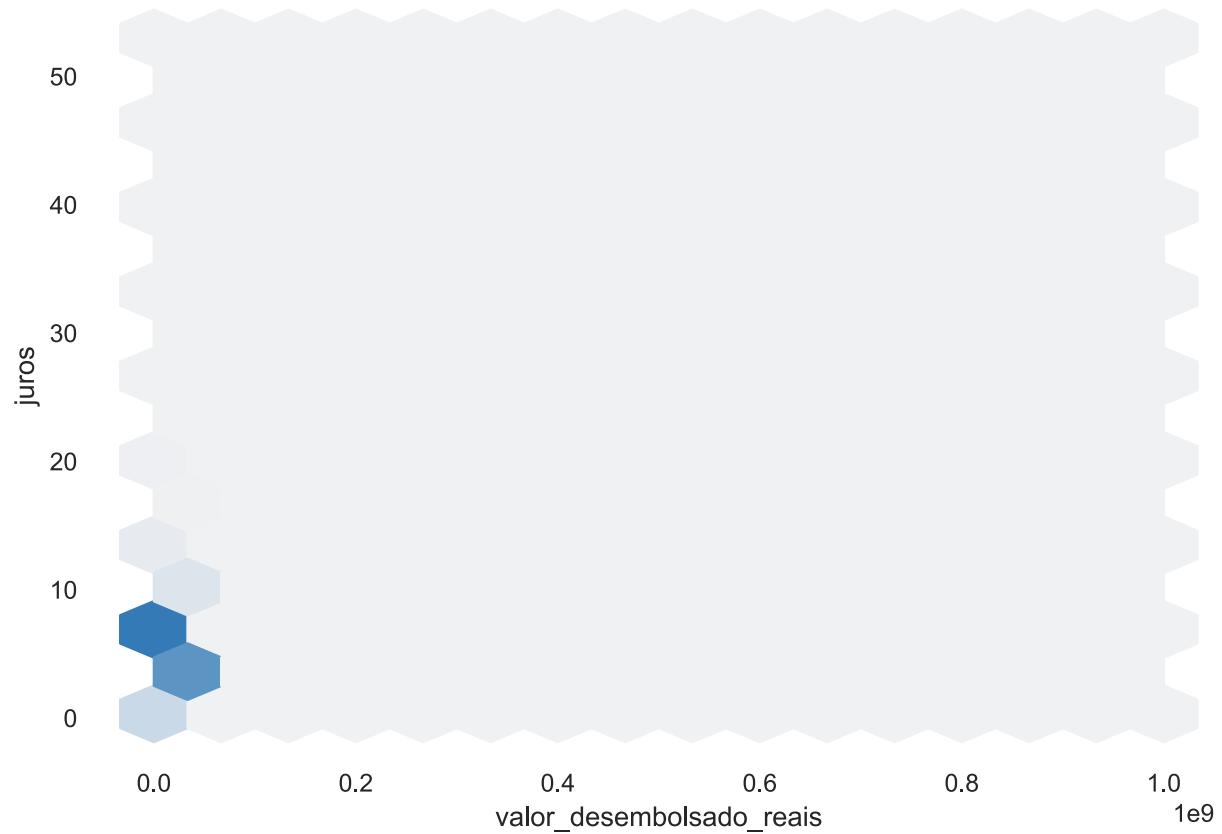


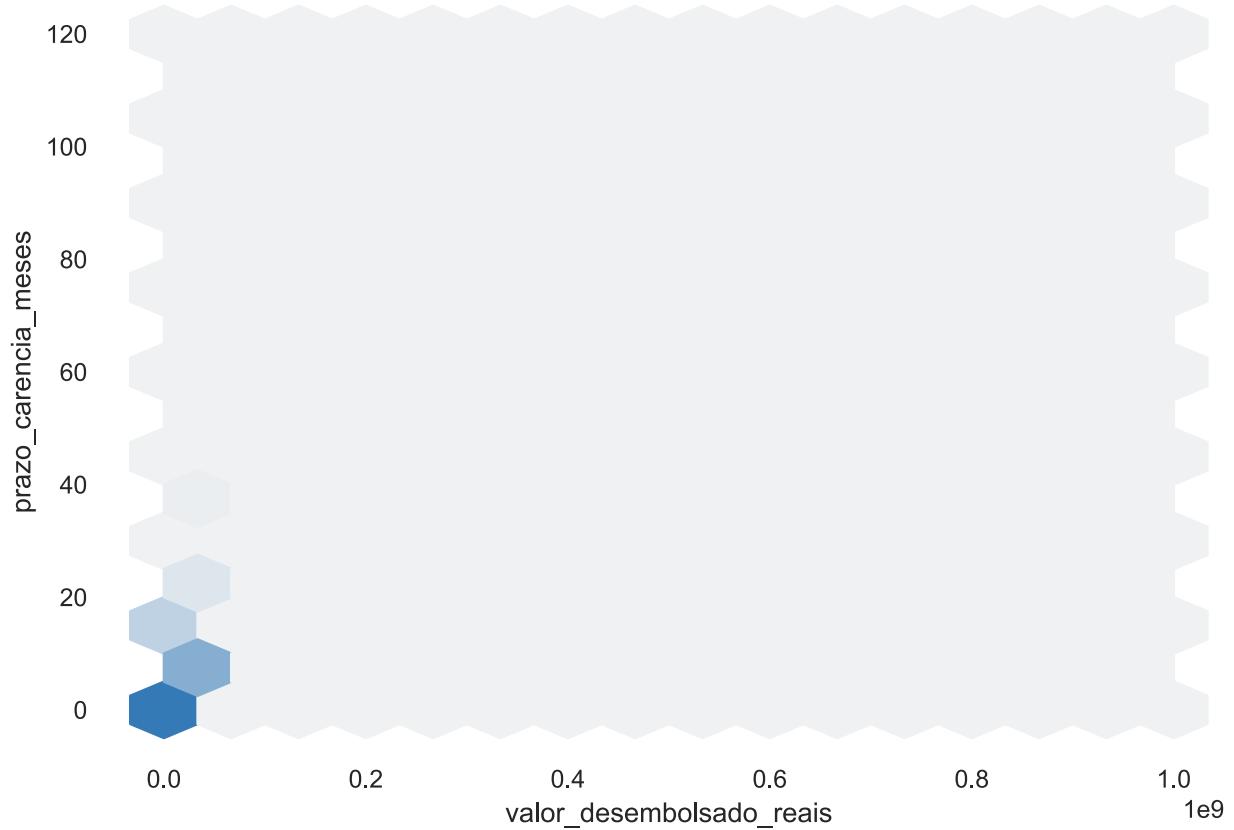


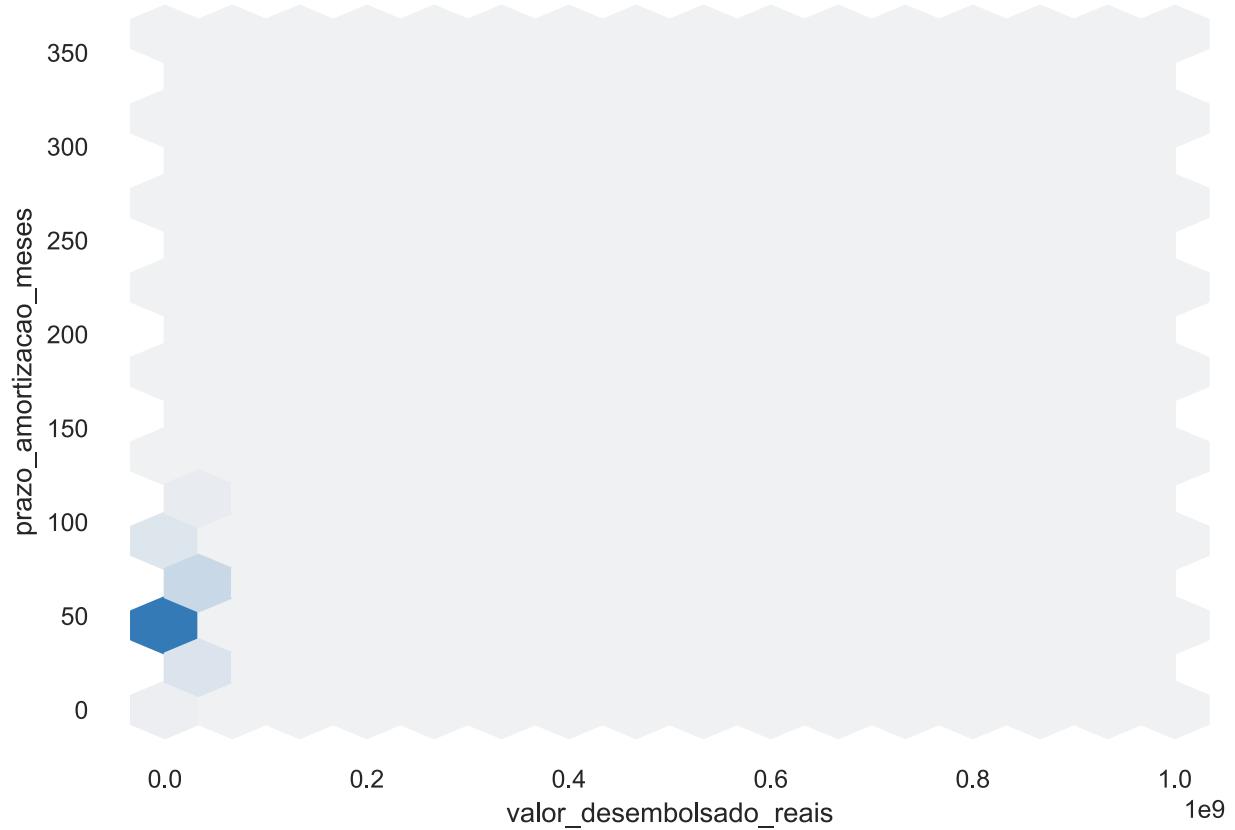




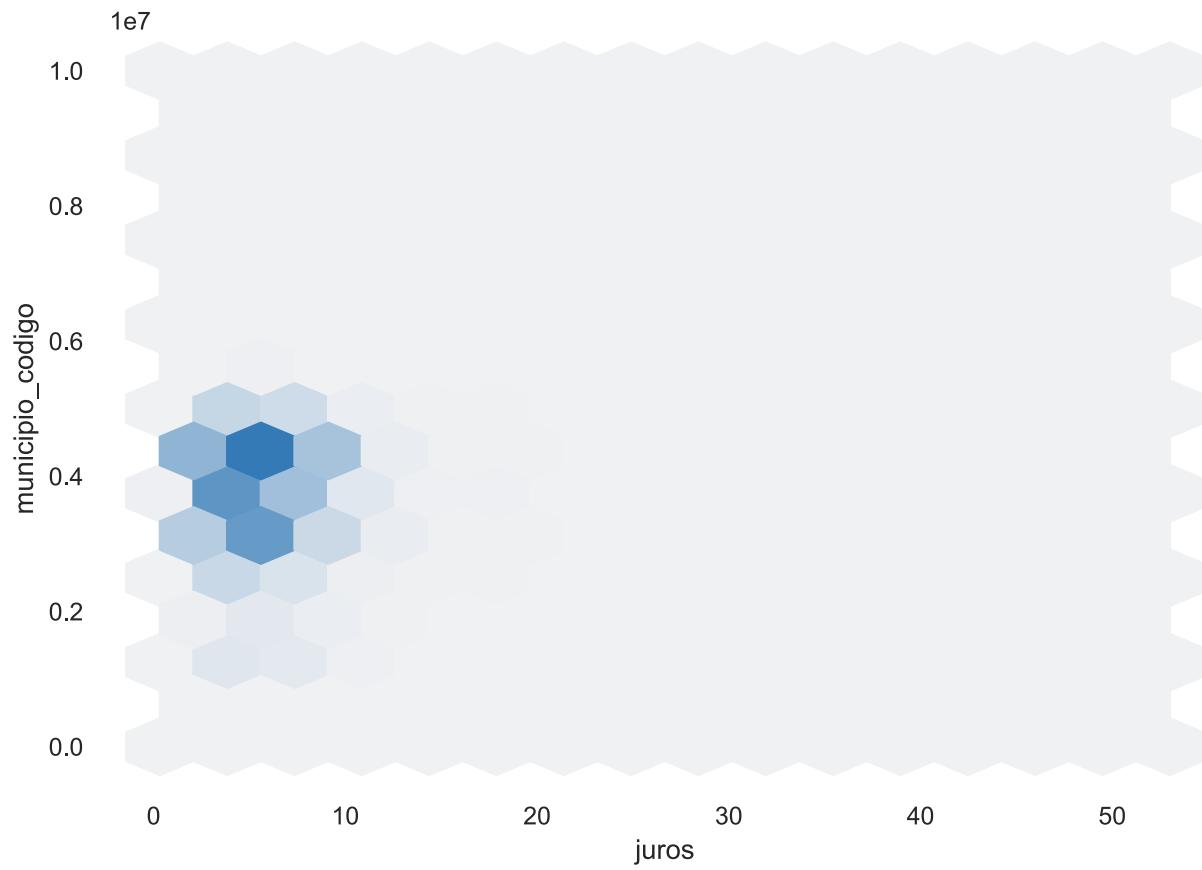


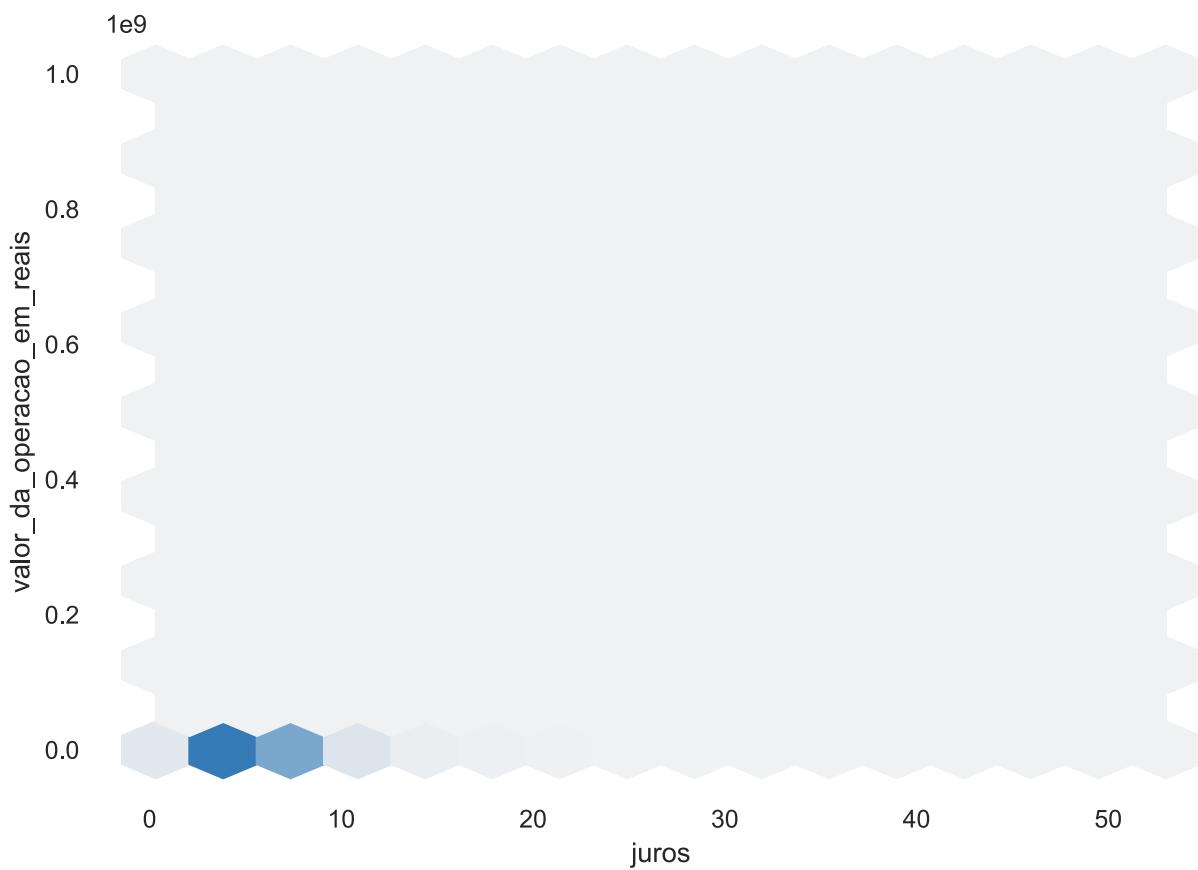


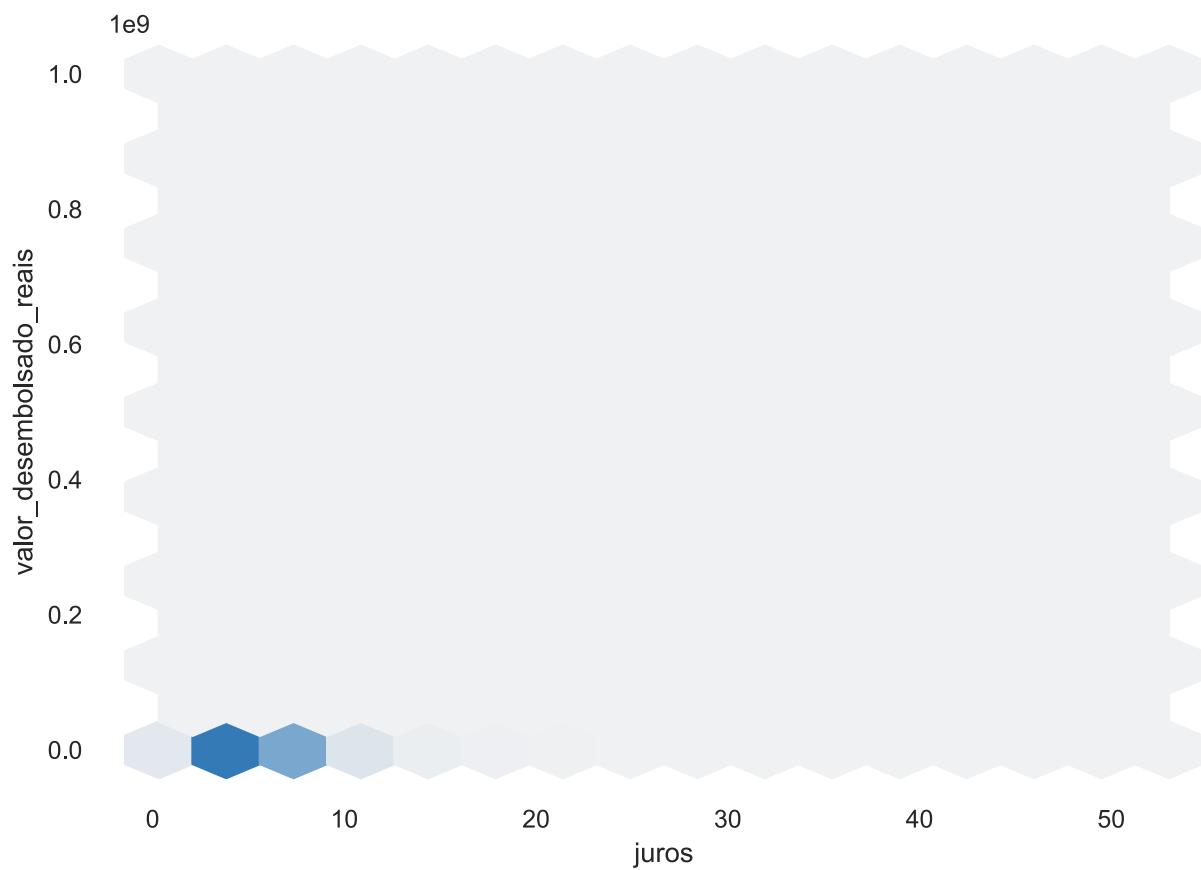




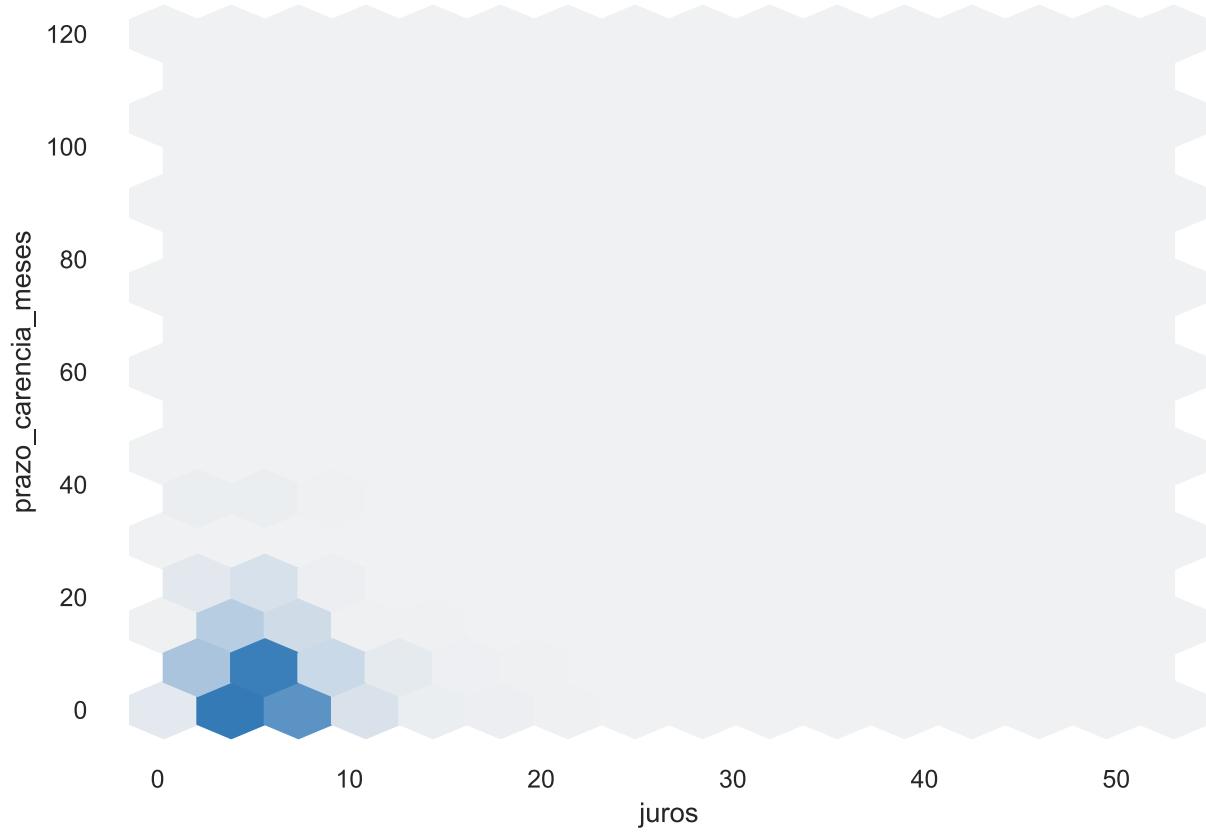


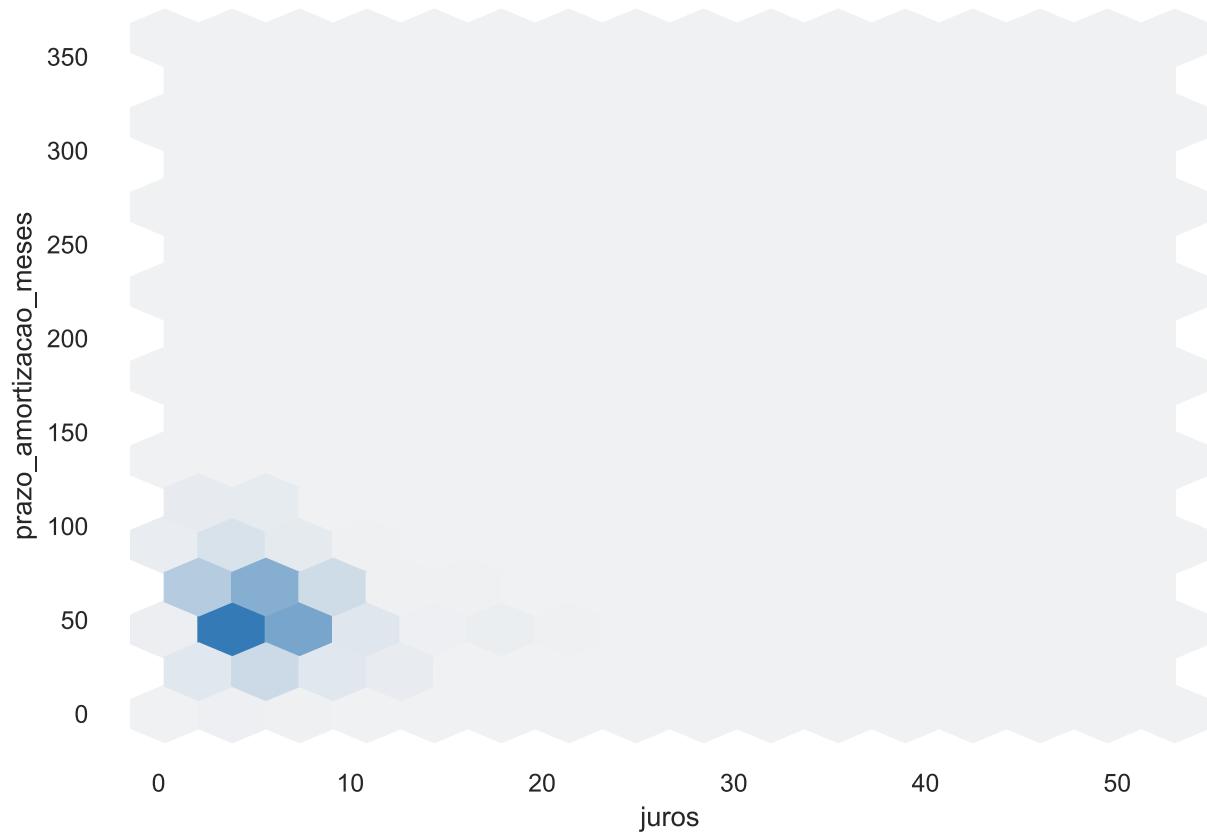




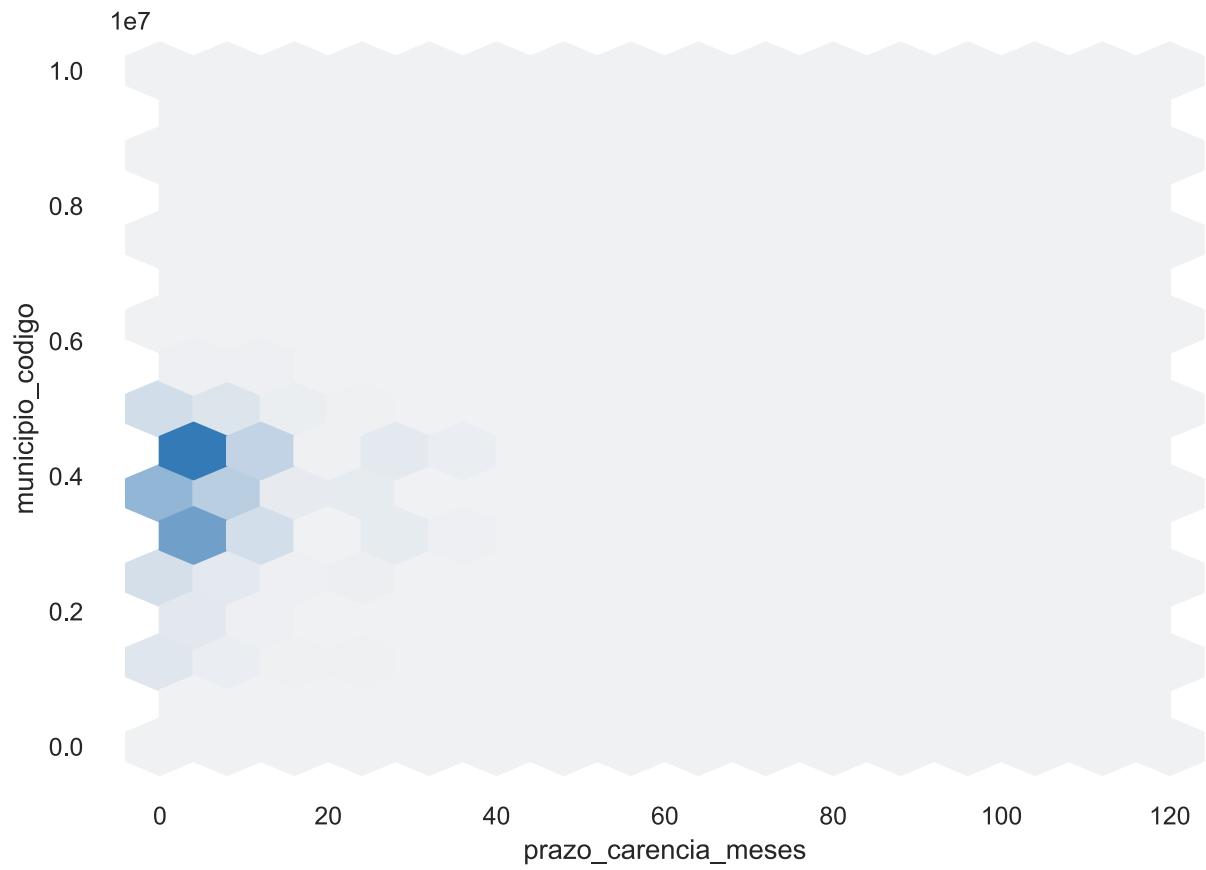


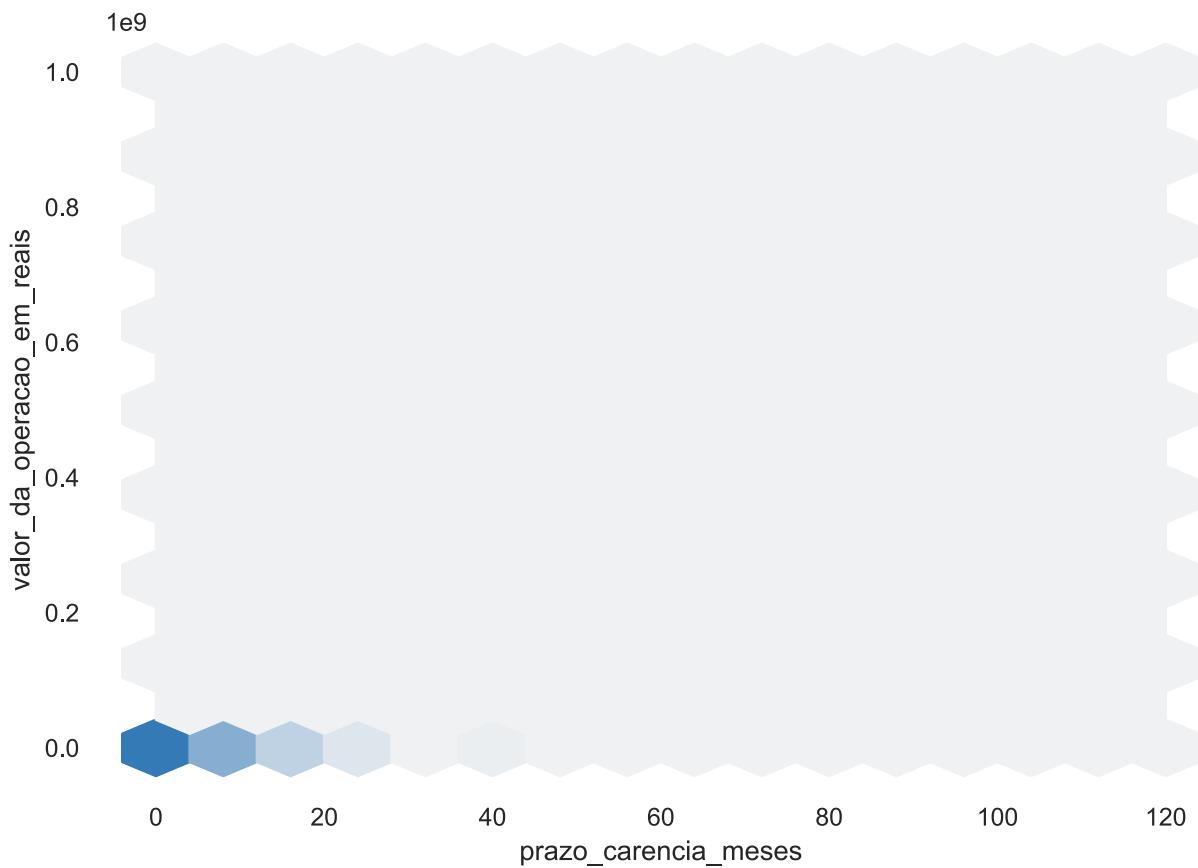


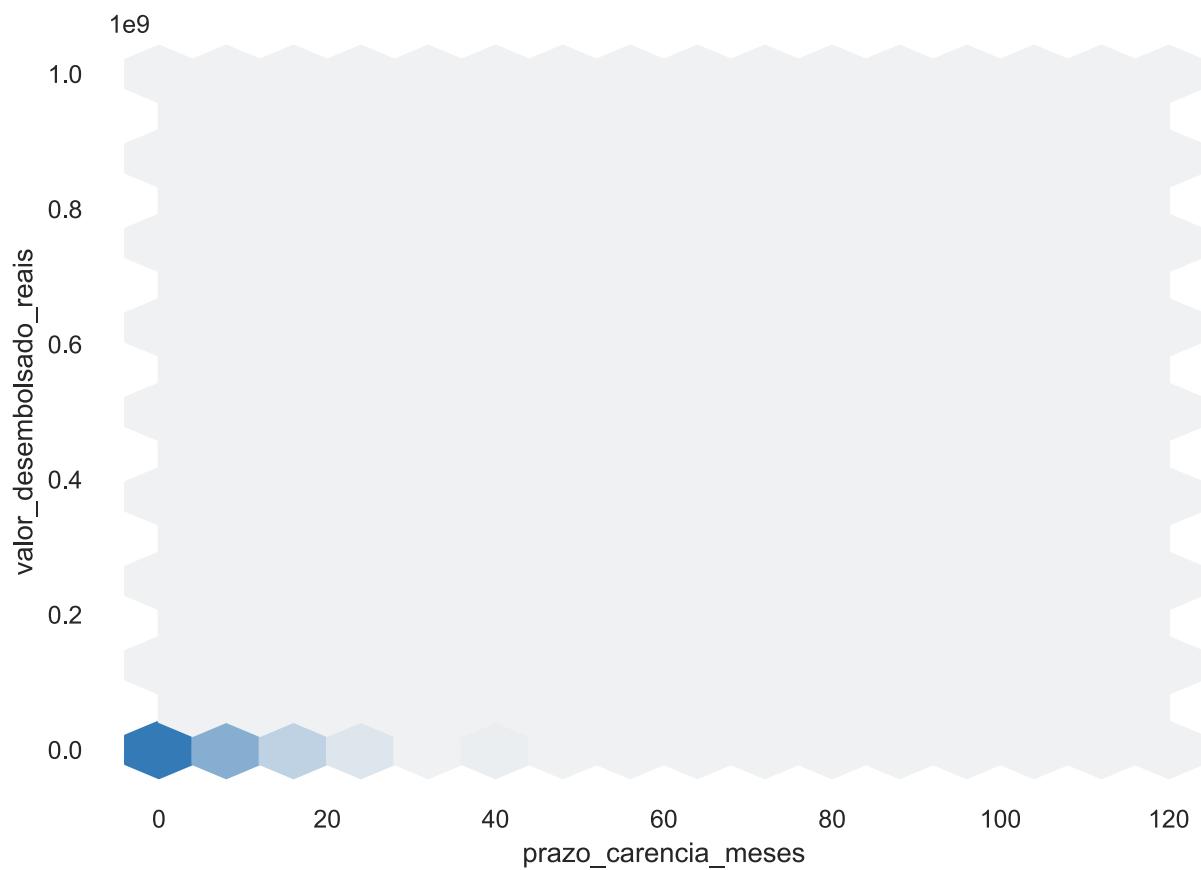


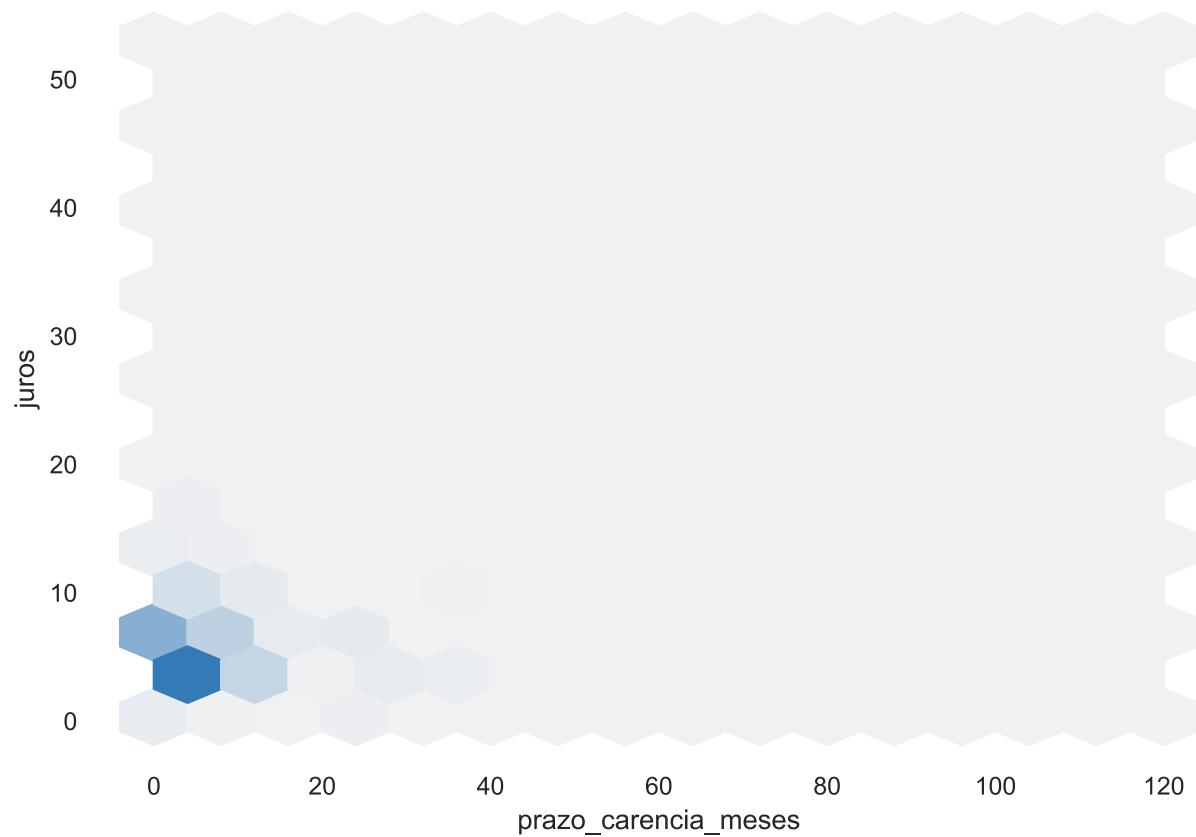




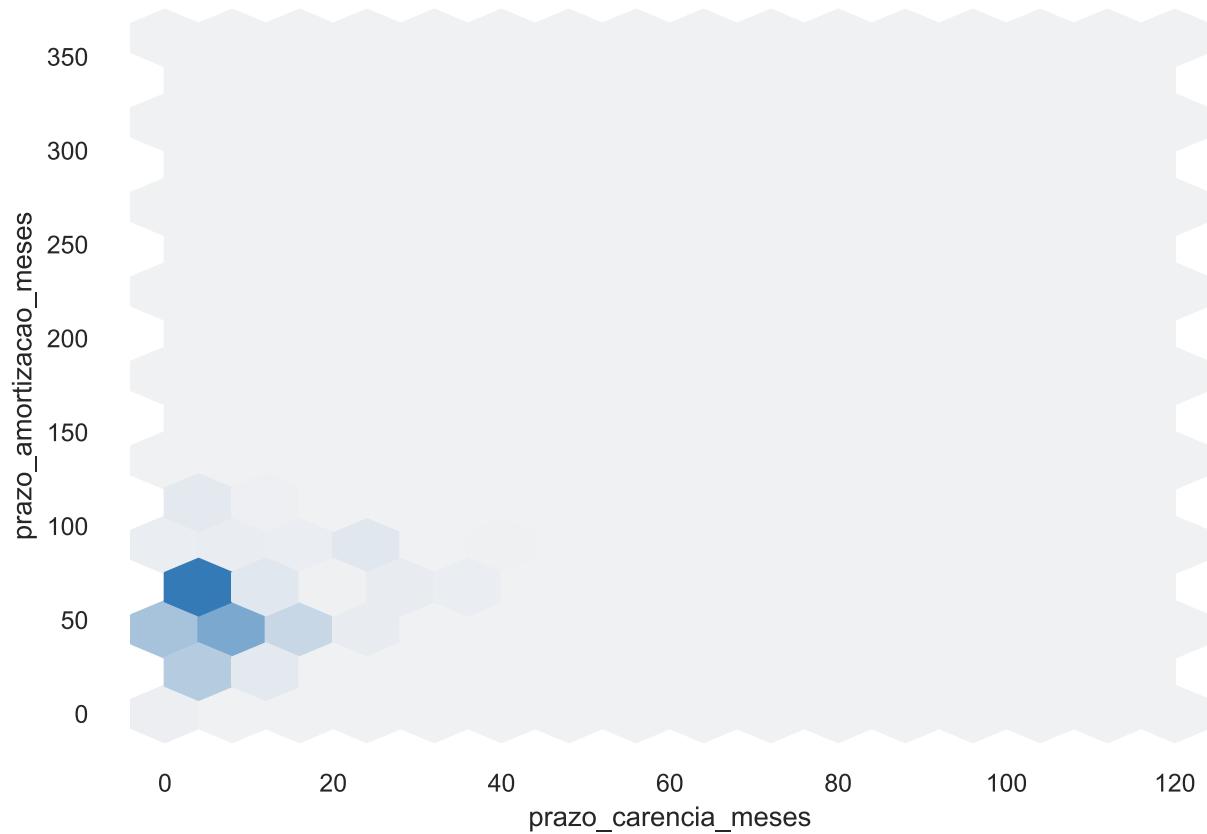




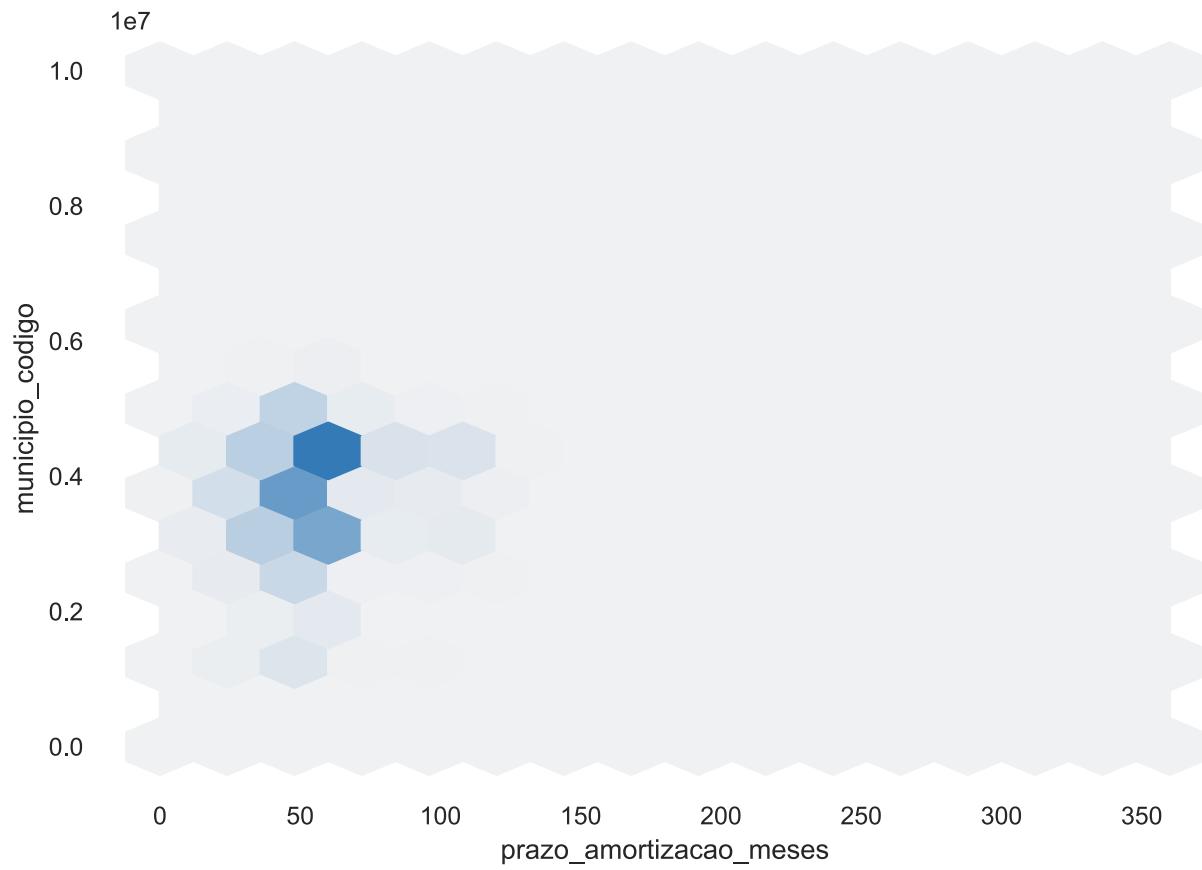


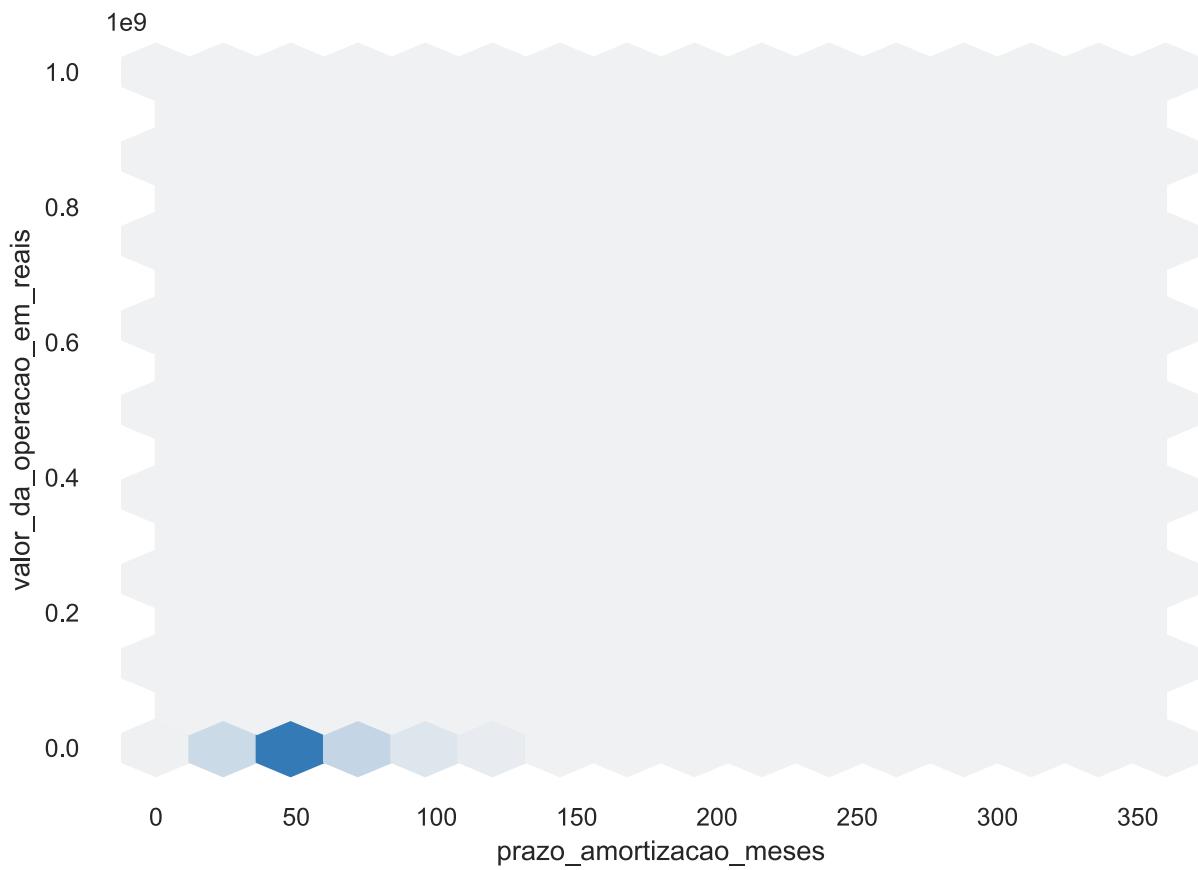


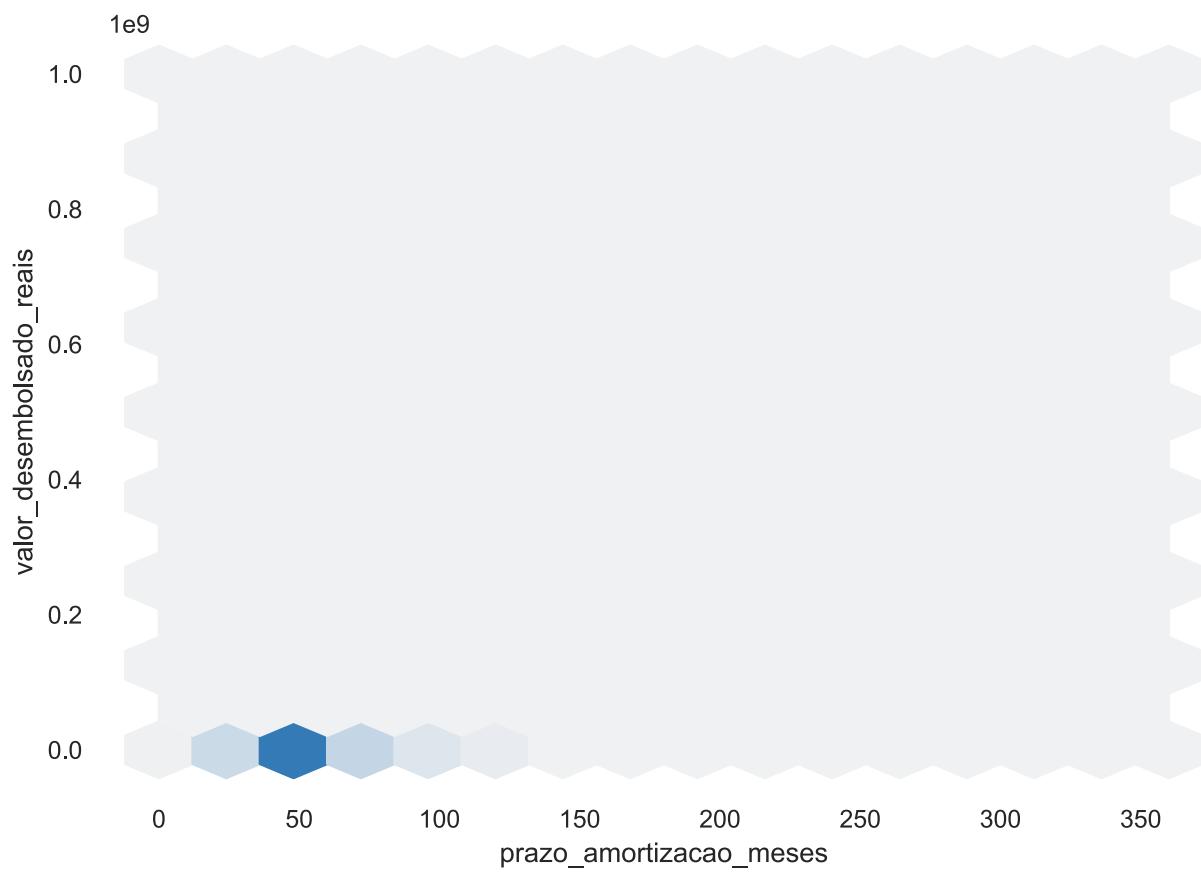


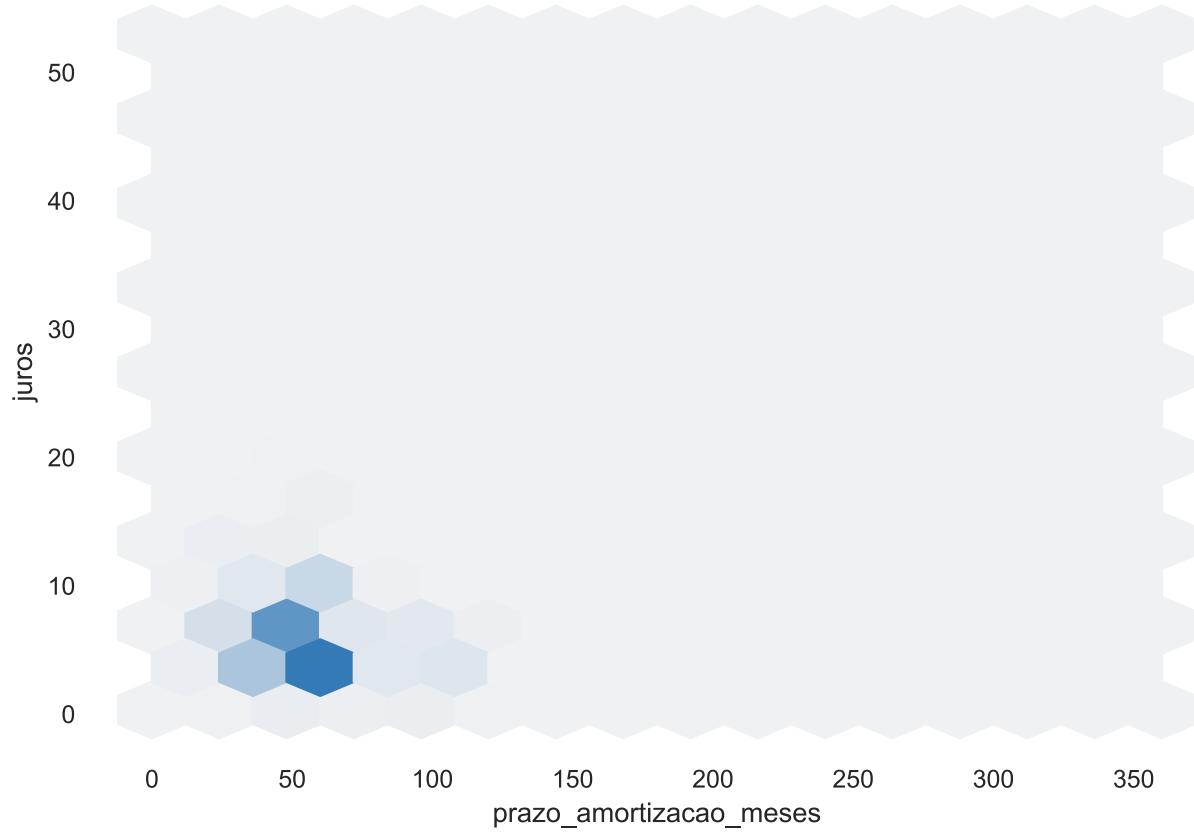


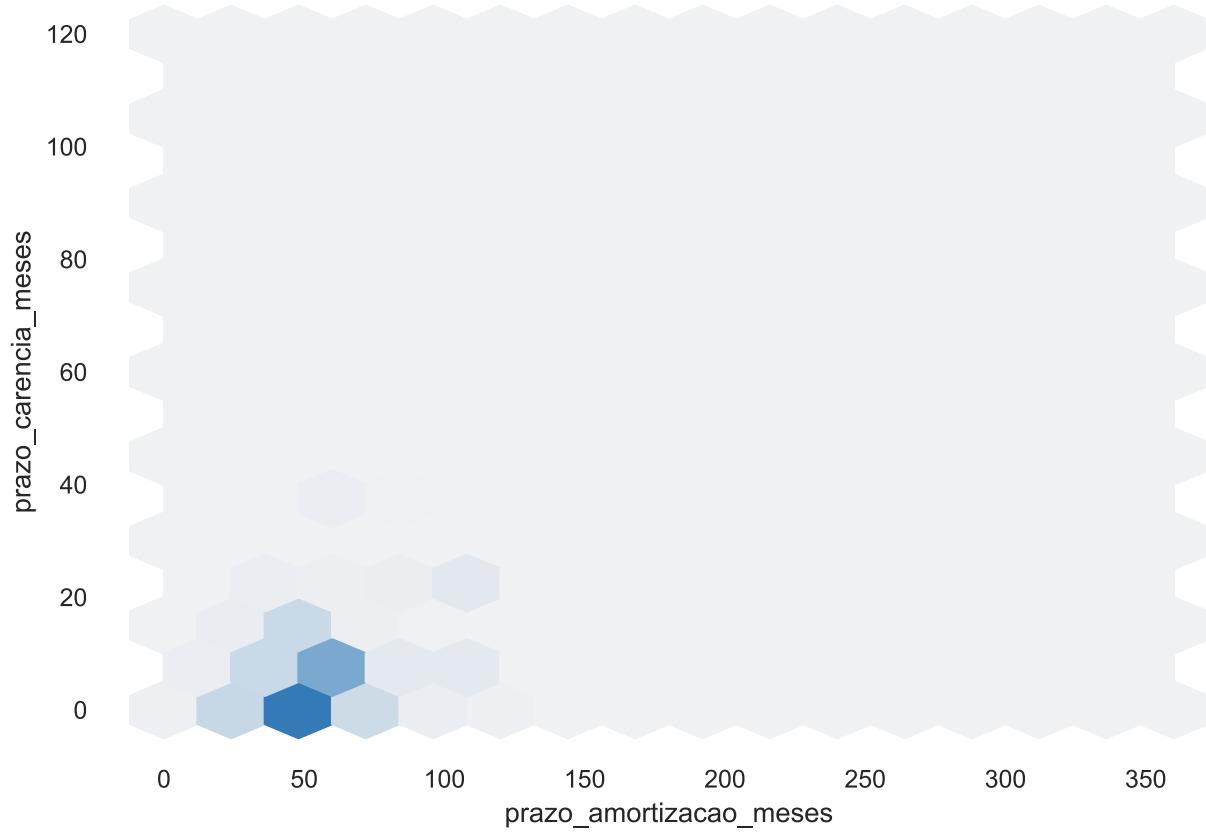








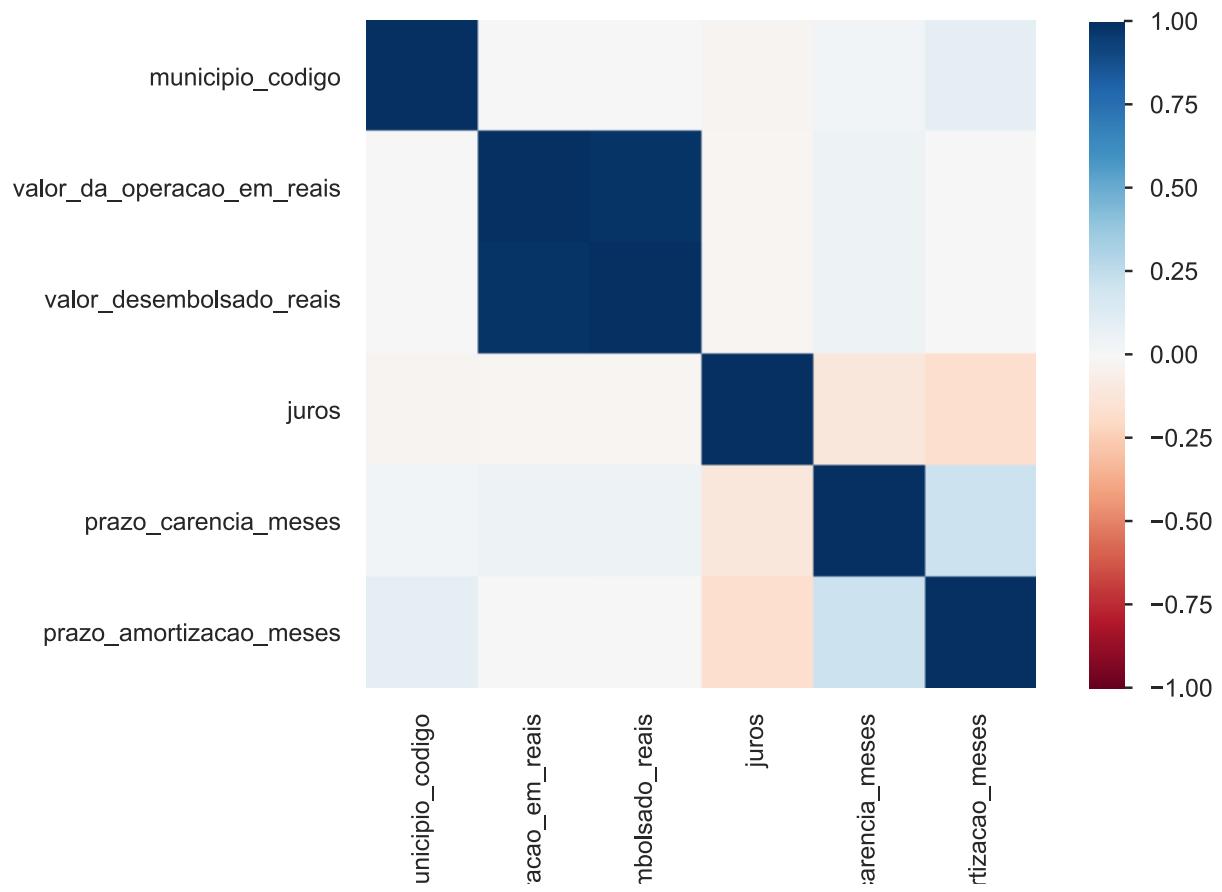


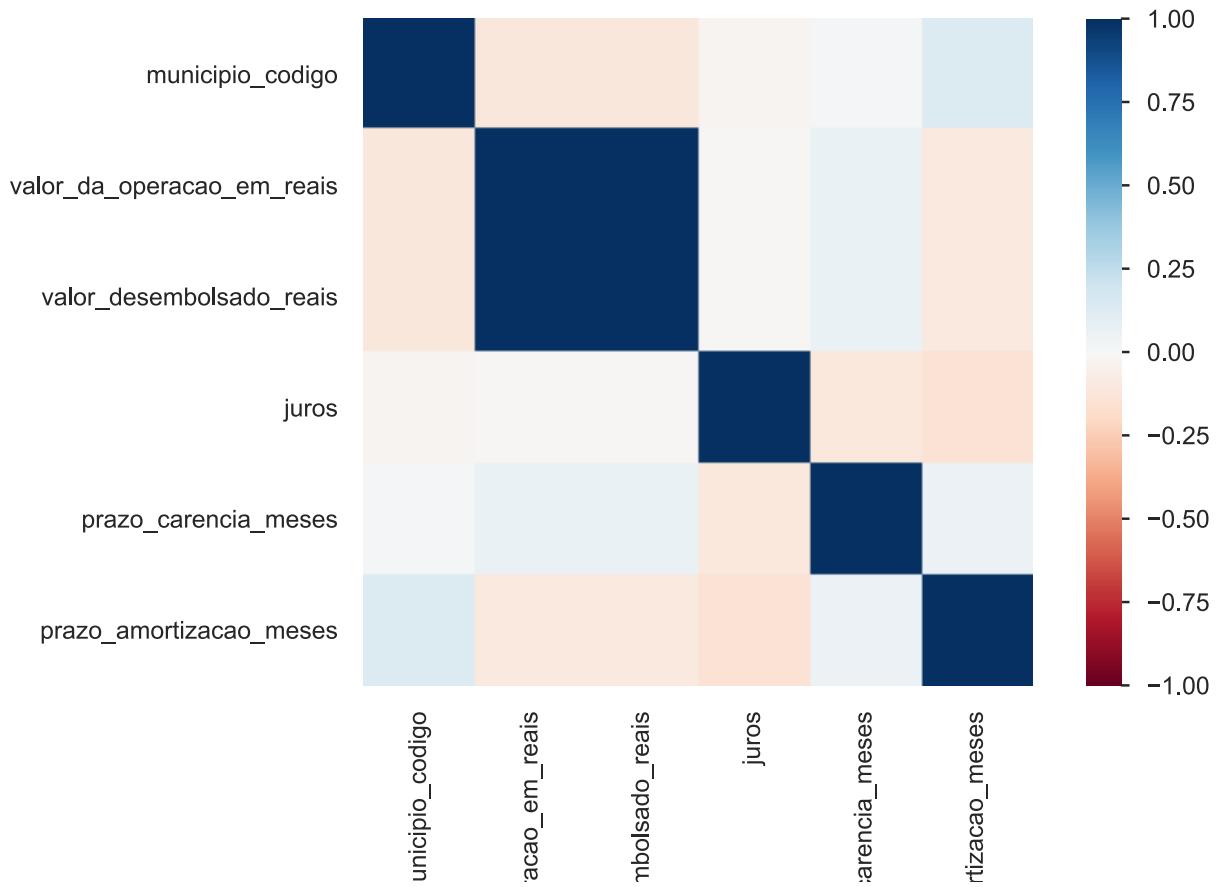


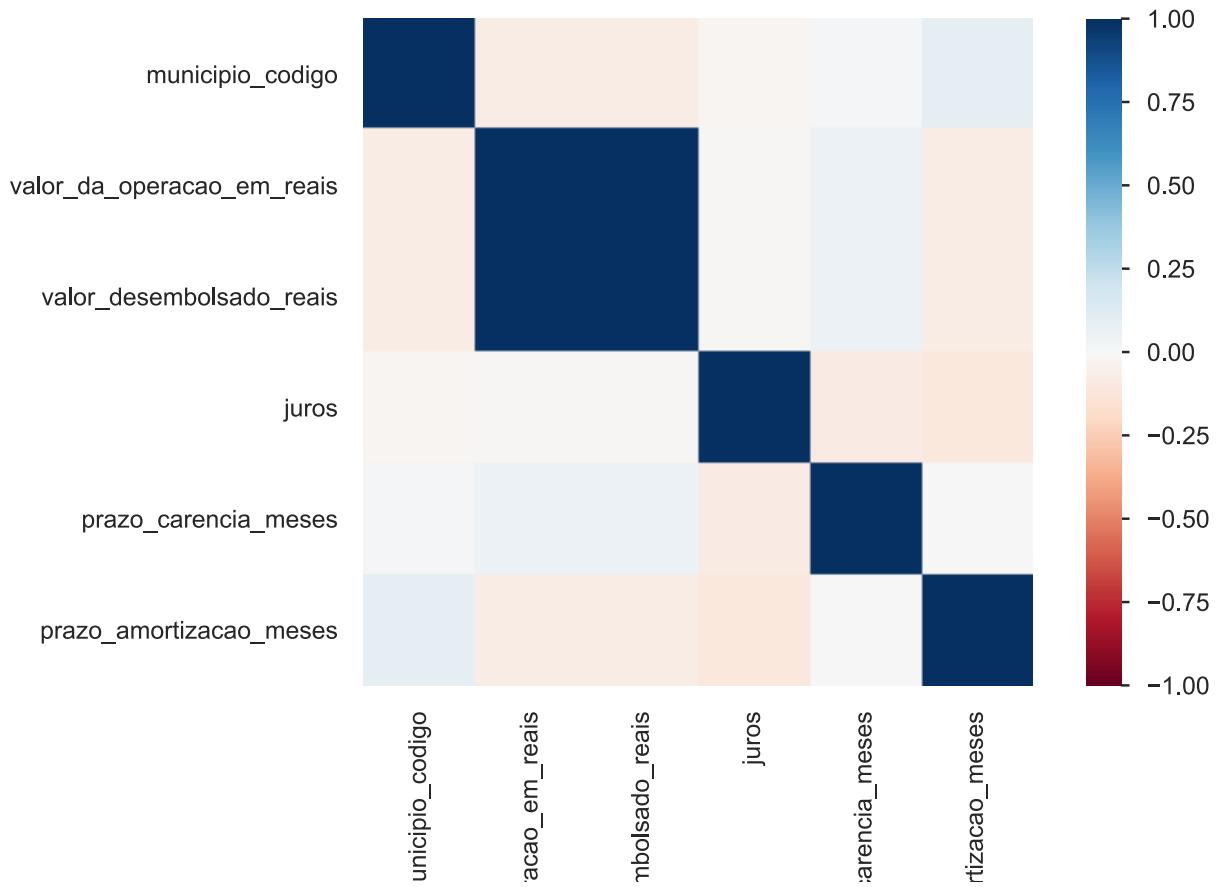


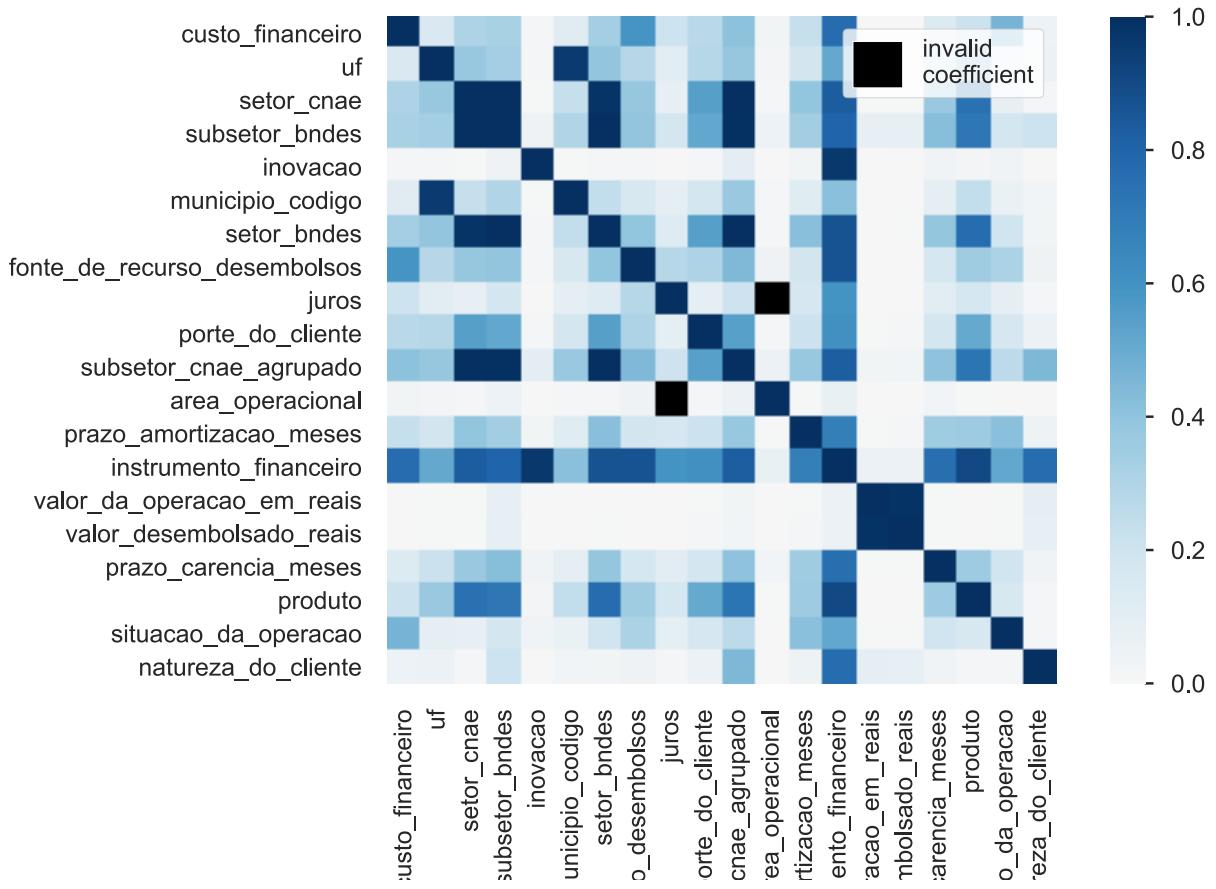


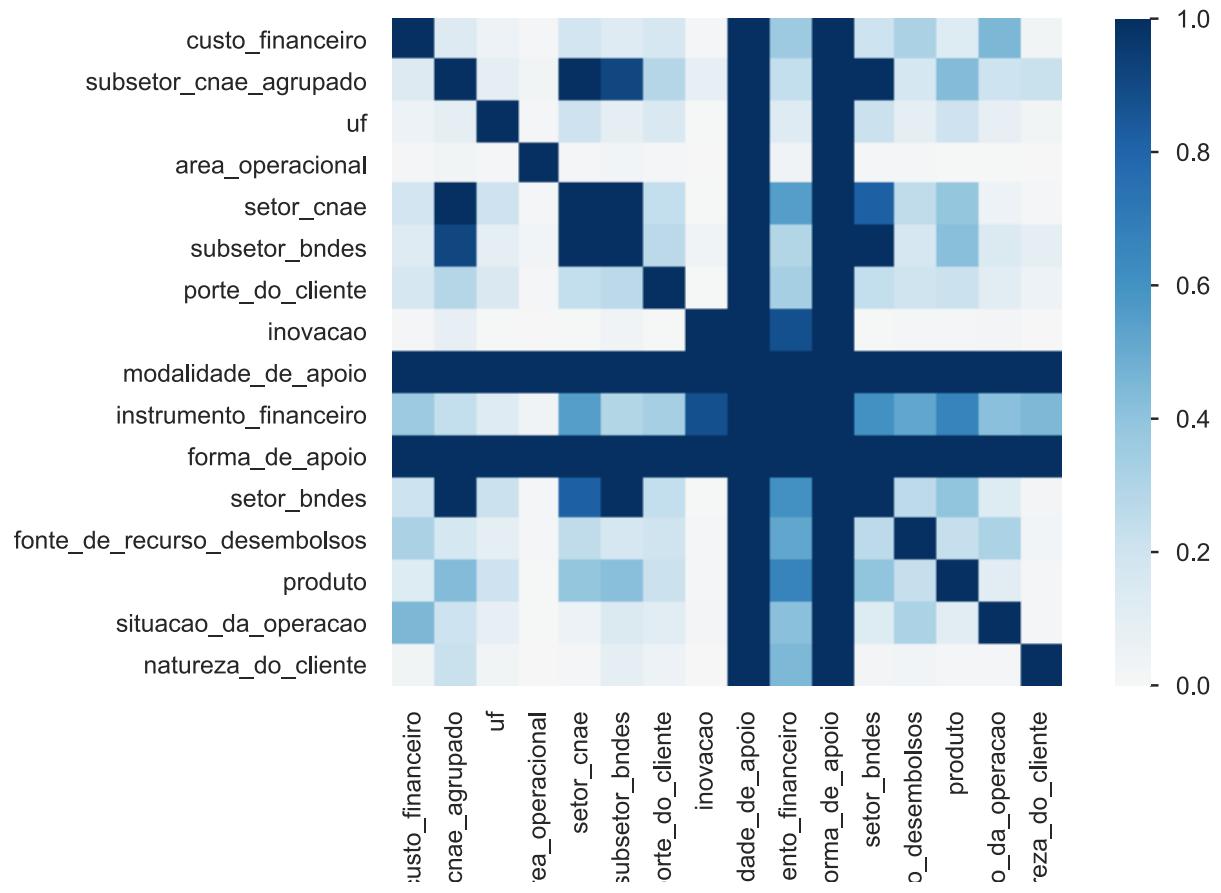
# Correlations



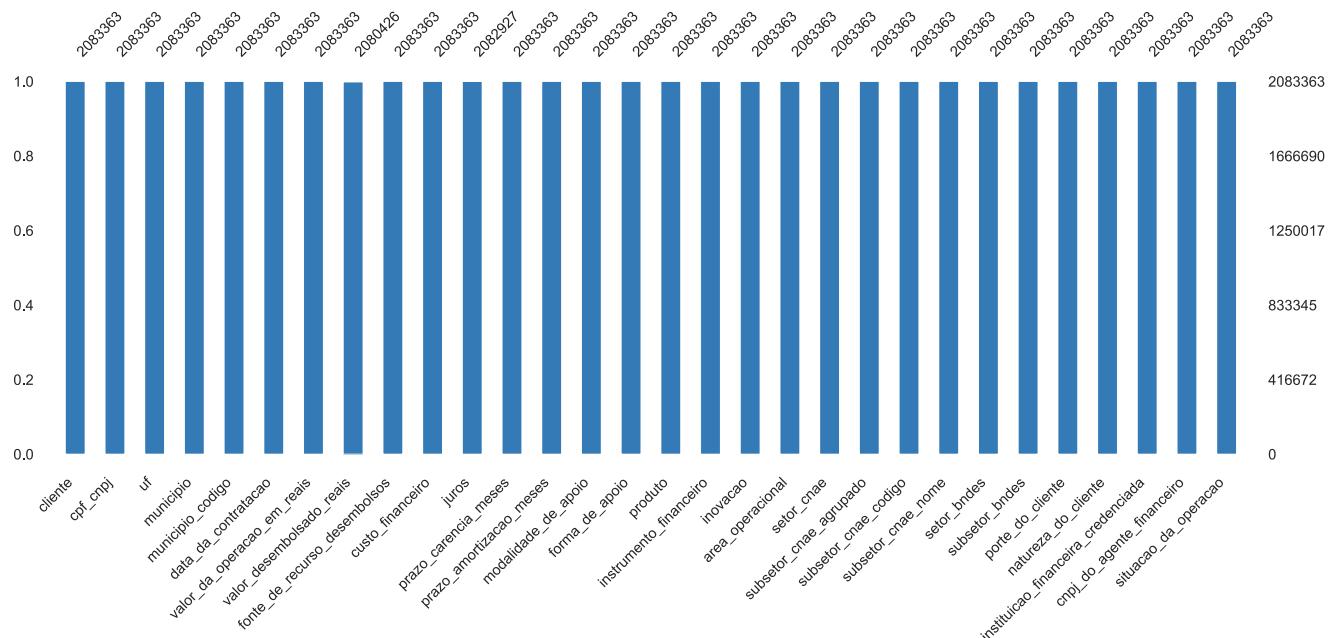








# Missing values



A simple visualization of nullity by column.

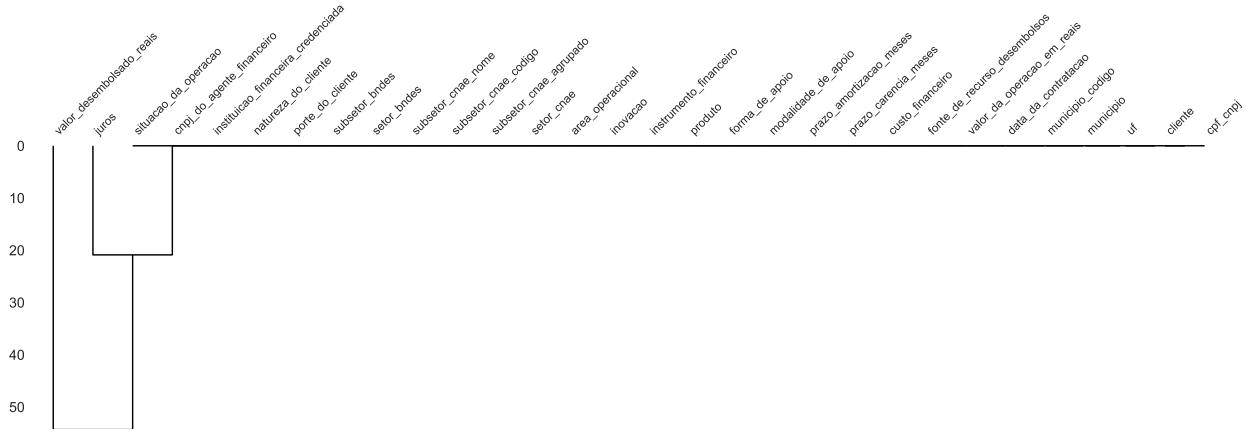
cliente  
cpf\_cnpj  
uf  
municipio  
municipio\_codigo  
data\_da\_contratacao  
valor\_da\_operacao\_em\_reais  
fonte\_de\_recurso\_desembolsos  
custo\_financiero  
juros  
prazo\_carencia\_meses  
prazo\_amortizacao\_meses  
modalidade\_de\_apoio  
forma\_de\_apoio  
produto  
instrumento\_financiero  
inovacao  
area\_operacional  
setor\_cnae  
subsetor\_cnae\_agrupado  
subsetor\_cnae\_codigo  
subsetor\_cnae\_nome  
selor\_bndes  
porte\_do\_cliente  
natureza\_do\_cliente  
instituicao\_financa\_credenciada  
cnpj\_do\_agente\_financiero  
situação\_da\_operação

2083363

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



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



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

# Sample

## First rows

	cliente	cpf_cnpj	uf	municipio
0	BANCO COOPERATIVO SICOOB S.A.	**.*38.232/0001-**	SP	PEDREGULH
1	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL	**.*16.560/0001-**	RS	TRES DE MAI
2	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL	**.*16.560/0001-**	RS	ERECHIM
3	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL	**.*16.560/0001-**	RS	ERECHIM
4	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL	**.*16.560/0001-**	RS	HUMAITA
5	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL	**.*16.560/0001-**	RS	VERANOPOL
6	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL	**.*16.560/0001-**	RS	VERANOPOL
7	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL	**.*16.560/0001-**	RS	VERANOPOL
8	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL	**.*16.560/0001-**	RS	VISTA ALEGRI
9	BANCO REGIONAL DE DESENVOLVIMENTO DO EXTREMO SUL	**.*16.560/0001-**	RS	VISTA ALEGRI

## Last rows

	cliente	cpf_cnpj	uf
2083353	INDUSTRIAS ROMI S A	**.*20.428/0001-**	SP
2083354	J A MORAES DE LARA - LAJES	**.*97.609/0001-**	PR
2083355	JULIANE POMAGERSKI 07689672957	**.*68.334/0001-**	PR
2083356	JUSSARA FATIMA PACHECO 65098404991	**.*83.609/0001-**	PR

	cliente	cpf_cnpj	uf
2083357	L DELFINO TOMAZELLI - TRANSPORTES	**.*12.975/0001-**	PR
2083358	LUIZ HENRIQUE PEREIRA 09189608941	**.*40.920/0001-**	SC
2083359	LUMA - COMERCIO E PRODUCAO DE ARTEFATOS PLASTICOS EIREL	**.*94.946/0001-**	PR
2083360	LUPATINI & KRAEMER LTDA	**.*74.998/0001-**	RS
2083361	INDUSTRIAS ROMI S A	**.*20.428/0001-**	SP
2083362	VLP TRANSPORTES LTDA	**.*44.788/0001-**	PR

## Duplicate rows

## Most frequent

	cliente	cpf_cnpj	uf	municipio
12515	BANCO DO BRASIL SA	**.*00.000/0001-**	RS	SANTA CRUZ DO SUL
16410	BANCO DO ESTADO DO RIO GRANDE DO SUL SA	**.*02.067/0001-**	RS	HULHA NEGRA
12463	BANCO DO BRASIL SA	**.*00.000/0001-**	RS	SANTA CRUZ DO SUL
16398	BANCO DO ESTADO DO RIO GRANDE DO SUL SA	**.*02.067/0001-**	RS	HERVAL
16525	BANCO DO ESTADO DO RIO GRANDE DO SUL SA	**.*02.067/0001-**	RS	JOIA
42037	HORIZONTE EXPRESS TRANSPORTES LTDA	**.*65.584/0001-**	PE	OLINDA
42040	HORIZONTE EXPRESS TRANSPORTES LTDA	**.*65.584/0001-**	PE	OLINDA
16385	BANCO DO ESTADO DO RIO GRANDE DO SUL SA	**.*02.067/0001-**	RS	HERVAL
17073	BANCO DO ESTADO DO RIO GRANDE DO SUL SA	**.*02.067/0001-**	RS	SANTANA DO LIVRAM
12566	BANCO DO BRASIL SA	**.*00.000/0001-**	RS	SANTA CRUZ DO SUL

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