

# Lista de bibliotecas importadas para realização do projeto

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.pipeline import Pipeline
from sklearn.impute import SimpleImputer
from sklearn.compose import ColumnTransformer
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.preprocessing import StandardScaler, OneHotEncoder, OrdinalEncoder
from sklearn.model_selection import GridSearchCV, StratifiedKFold, train_test_sp
from sklearn.metrics import classification_report, ConfusionMatrixDisplay, f1_sc
from imblearn.over_sampling import SMOTE
from sklearn.decomposition import PCA
from sklearn.feature_selection import SelectKBest, f_classif
from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import LogisticRegression
from xgboost import XGBClassifier
```

C:\Users\wiltd\AppData\Local\Temp\ipykernel\_14384\1476442007.py:1: DeprecationWarning:  
Pyarrow will become a required dependency of pandas in the next major release of pandas (pandas 3.0),  
(to allow more performant data types, such as the Arrow string type, and better interoperability with other libraries)  
but was not found to be installed on your system.  
If this would cause problems for you,  
please provide us feedback at <https://github.com/pandas-dev/pandas/issues/54466>

```
import pandas as pd
```

## Importação dos dados para análise

```
In [2]: # abrindo o arquivo contendo do DataSet e transformando-os em DataFrame para mel
df = pd.read_csv('../credito-imoveis/application_train.csv')

print('planilha de treino: ', df.shape)
```

planilha de treino: (246008, 122)

```
In [3]: pd.options.display.max_rows = 160
```

```
In [4]: pd.options.display.max_columns = 160
```

## Realização da Exploração e Análise dos Dados - EDA (Exploration Data Analysis)

## Passo 1 - Entendendo os dados e preparando os dados:

### A. Conhecendo o df:

```
In [5]: df.head()
```

	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLA
0	456162	0	Cash loans	F	N	
1	134978	0	Cash loans	F	N	
2	318952	0	Cash loans	M	Y	
3	361264	0	Cash loans	F	N	
4	260639	0	Cash loans	F	N	

Constata-se que o dataset é grande com muitos dados e colunas de difícil entendimento.

### B. Verificando os tipos de dados contidos no df por coluna:

```
In [6]: df.info(verbose=True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 246008 entries, 0 to 246007
Data columns (total 122 columns):
 #   Column           Dtype  
 --- 
 0   SK_ID_CURR        int64  
 1   TARGET            int64  
 2   NAME_CONTRACT_TYPE object 
 3   CODE_GENDER       object 
 4   FLAG_OWN_CAR      object 
 5   FLAG_OWN_REALTY   object 
 6   CNT_CHILDREN      int64  
 7   AMT_INCOME_TOTAL  float64 
 8   AMT_CREDIT         float64 
 9   AMT_ANNUITY        float64 
 10  AMT_GOODS_PRICE   float64 
 11  NAME_TYPE_SUITE   object 
 12  NAME_INCOME_TYPE  object 
 13  NAME_EDUCATION_TYPE object 
 14  NAME_FAMILY_STATUS object 
 15  NAME_HOUSING_TYPE object 
 16  REGION_POPULATION_RELATIVE float64 
 17  DAYS_BIRTH         int64  
 18  DAYS_EMPLOYED      int64  
 19  DAYS_REGISTRATION  float64 
 20  DAYS_ID_PUBLISH   int64  
 21  OWN_CAR_AGE        float64 
 22  FLAG_MOBIL         int64  
 23  FLAG_EMP_PHONE     int64  
 24  FLAG_WORK_PHONE    int64  
 25  FLAG_CONT_MOBILE   int64  
 26  FLAG_PHONE          int64  
 27  FLAG_EMAIL          int64  
 28  OCCUPATION_TYPE    object 
 29  CNT_FAM_MEMBERS    float64 
 30  REGION_RATING_CLIENT int64  
 31  REGION_RATING_CLIENT_W_CITY int64  
 32  WEEKDAY_APPR_PROCESS_START object 
 33  HOUR_APPR_PROCESS_START int64  
 34  REG_REGION_NOT_LIVE_REGION int64  
 35  REG_REGION_NOT_WORK_REGION int64  
 36  LIVE_REGION_NOT_WORK_REGION int64  
 37  REG_CITY_NOT_LIVE_CITY   int64  
 38  REG_CITY_NOT_WORK_CITY  int64  
 39  LIVE_CITY_NOT_WORK_CITY int64  
 40  ORGANIZATION_TYPE   object 
 41  EXT_SOURCE_1        float64 
 42  EXT_SOURCE_2        float64 
 43  EXT_SOURCE_3        float64 
 44  APARTMENTS_AVG      float64 
 45  BASEMENTAREA_AVG    float64 
 46  YEARS_BEGINEXPLUATATION_AVG float64 
 47  YEARS_BUILD_AVG     float64 
 48  COMMONAREA_AVG      float64 
 49  ELEVATORS_AVG       float64 
 50  ENTRANCES_AVG       float64 
 51  FLOORSMAX_AVG       float64 
 52  FLOORSMIN_AVG       float64 
 53  LANDAREA_AVG        float64 
 54  LIVINGAPARTMENTS_AVG float64
```

55	LIVINGAREA_AVG	float64
56	NONLIVINGAPARTMENTS_AVG	float64
57	NONLIVINGAREA_AVG	float64
58	APARTMENTS_MODE	float64
59	BASEMENTAREA_MODE	float64
60	YEARS_BEGINEXPLUATATION_MODE	float64
61	YEARS_BUILD_MODE	float64
62	COMMONAREA_MODE	float64
63	ELEVATORS_MODE	float64
64	ENTRANCES_MODE	float64
65	FLOORSMAX_MODE	float64
66	FLOORSMIN_MODE	float64
67	LANDAREA_MODE	float64
68	LIVINGAPARTMENTS_MODE	float64
69	LIVINGAREA_MODE	float64
70	NONLIVINGAPARTMENTS_MODE	float64
71	NONLIVINGAREA_MODE	float64
72	APARTMENTS_MEDI	float64
73	BASEMENTAREA_MEDI	float64
74	YEARS_BEGINEXPLUATATION_MEDI	float64
75	YEARS_BUILD_MEDI	float64
76	COMMONAREA_MEDI	float64
77	ELEVATORS_MEDI	float64
78	ENTRANCES_MEDI	float64
79	FLOORSMAX_MEDI	float64
80	FLOORSMIN_MEDI	float64
81	LANDAREA_MEDI	float64
82	LIVINGAPARTMENTS_MEDI	float64
83	LIVINGAREA_MEDI	float64
84	NONLIVINGAPARTMENTS_MEDI	float64
85	NONLIVINGAREA_MEDI	float64
86	FONDKAPREMONT_MODE	object
87	HOUSETYPE_MODE	object
88	TOTALAREA_MODE	float64
89	WALLSMATERIAL_MODE	object
90	EMERGENCYSTATE_MODE	object
91	OBS_30_CNT_SOCIAL_CIRCLE	float64
92	DEF_30_CNT_SOCIAL_CIRCLE	float64
93	OBS_60_CNT_SOCIAL_CIRCLE	float64
94	DEF_60_CNT_SOCIAL_CIRCLE	float64
95	DAYS_LAST_PHONE_CHANGE	float64
96	FLAG_DOCUMENT_2	int64
97	FLAG_DOCUMENT_3	int64
98	FLAG_DOCUMENT_4	int64
99	FLAG_DOCUMENT_5	int64
100	FLAG_DOCUMENT_6	int64
101	FLAG_DOCUMENT_7	int64
102	FLAG_DOCUMENT_8	int64
103	FLAG_DOCUMENT_9	int64
104	FLAG_DOCUMENT_10	int64
105	FLAG_DOCUMENT_11	int64
106	FLAG_DOCUMENT_12	int64
107	FLAG_DOCUMENT_13	int64
108	FLAG_DOCUMENT_14	int64
109	FLAG_DOCUMENT_15	int64
110	FLAG_DOCUMENT_16	int64
111	FLAG_DOCUMENT_17	int64
112	FLAG_DOCUMENT_18	int64
113	FLAG_DOCUMENT_19	int64
114	FLAG_DOCUMENT_20	int64

```
115 FLAG_DOCUMENT_21           int64
116 AMT_REQ_CREDIT_BUREAU_HOUR float64
117 AMT_REQ_CREDIT_BUREAU_DAY  float64
118 AMT_REQ_CREDIT_BUREAU_WEEK float64
119 AMT_REQ_CREDIT_BUREAU_MON   float64
120 AMT_REQ_CREDIT_BUREAU_QRT  float64
121 AMT_REQ_CREDIT_BUREAU_YEAR float64
dtypes: float64(65), int64(41), object(16)
memory usage: 229.0+ MB
```

Conclui-se que a maior parte do dataset é constituído de dados numéricos, são poucas colunas com dados classificatórios.

### C. Verificando se há linhas duplicadas:

```
In [7]: df.duplicated().value_counts()
```

```
Out[7]: False    246008
Name: count, dtype: int64
```

Constata-se, portanto, que não há linhas duplicadas.

### D. Verificando se há ou não dados nulos:

```
In [8]: df.isnull().sum().head(60)
```

```
Out[8]: SK_ID_CURR          0
TARGET              0
NAME_CONTRACT_TYPE 0
CODE_GENDER         0
FLAG_OWN_CAR        0
FLAG_OWN_REALTY    0
CNT_CHILDREN        0
AMT_INCOME_TOTAL   0
AMT_CREDIT          0
AMT_ANNUITY         10
AMT_GOODS_PRICE     226
NAME_TYPE_SUITE     1048
NAME_INCOME_TYPE    0
NAME_EDUCATION_TYPE 0
NAME_FAMILY_STATUS   0
NAME_HOUSING_TYPE   0
REGION_POPULATION_RELATIVE 0
DAYS_BIRTH          0
DAYS_EMPLOYED       0
DAYS_REGISTRATION   0
DAYS_ID_PUBLISH    0
OWN_CAR_AGE         162359
FLAG_MOBIL          0
FLAG_EMP_PHONE      0
FLAG_WORK_PHONE     0
FLAG_CONT_MOBILE    0
FLAG_PHONE          0
FLAG_EMAIL          0
OCCUPATION_TYPE     77237
CNT_FAM_MEMBERS     1
REGION_RATING_CLIENT 0
REGION_RATING_CLIENT_W_CITY 0
WEEKDAY_APPR_PROCESS_START 0
HOUR_APPR_PROCESS_START 0
REG_REGION_NOT_LIVE_REGION 0
REG_REGION_NOT_WORK_REGION 0
LIVE_REGION_NOT_WORK_REGION 0
REG_CITY_NOT_LIVE_CITY 0
REG_CITY_NOT_WORK_CITY 0
LIVE_CITY_NOT_WORK_CITY 0
ORGANIZATION_TYPE    0
EXT_SOURCE_1         138803
EXT_SOURCE_2         544
EXT_SOURCE_3         48728
APARTMENTS_AVG      124955
BASEMENTAREA_AVG    144090
YEARS_BEGINEXPLUATATION_AVG 120096
YEARS_BUILD_AVG     163680
COMMONAREA_AVG      171978
ELEVATORS_AVG       131208
ENTRANCES_AVG       123944
FLOORSMAX_AVG       122483
FLOORSMIN_AVG       166999
LANDAREA_AVG        146087
LIVINGAPARTMENTS_AVG 168278
LIVINGAREA_AVG      123617
NONLIVINGAPARTMENTS_AVG 170914
NONLIVINGAREA_AVG   135860
APARTMENTS_MODE     124955
```

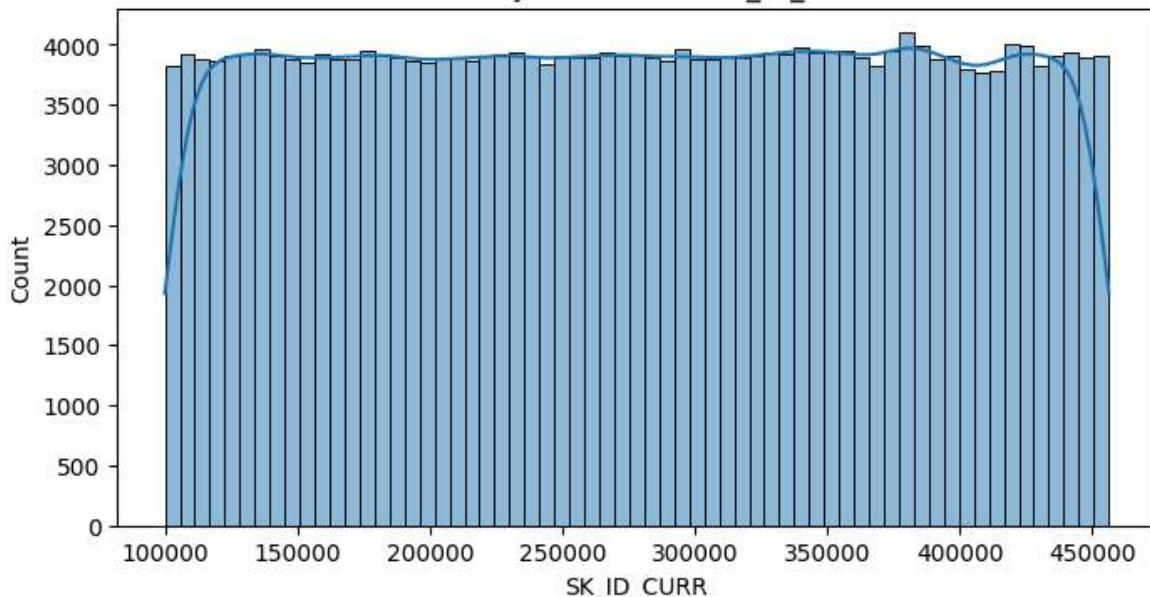
```
BASEMENTAREA_MODE           144090
dtype: int64
```

Foram encontrados dados nulos em algumas colunas que deverão ser tratados.

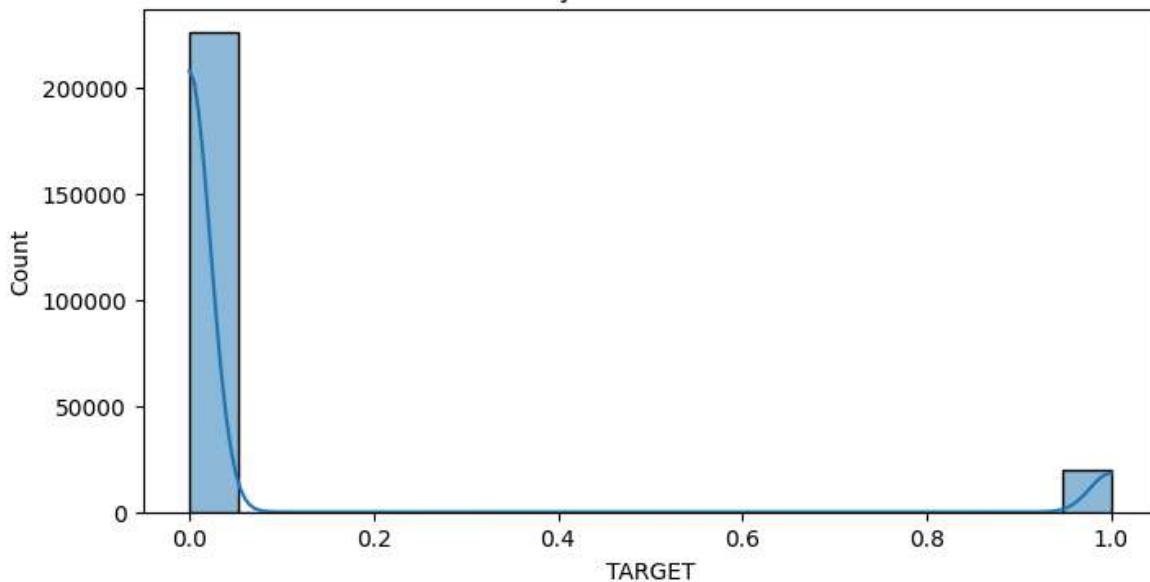
```
In [9]: # Visualização de distribuição para variáveis numéricas
num_cols = df.select_dtypes(include=['float64', 'int64']).columns
for col in num_cols:
    plt.figure(figsize=(8, 4))
    sns.histplot(df[col].dropna(), kde=True)
    plt.title(f"Distribuição da coluna {col}")
    plt.show()

# Visualização de outliers com boxplot
for col in num_cols:
    plt.figure(figsize=(8, 4))
    sns.boxplot(x=df[col])
    plt.title(f"Outliers na coluna {col}")
    plt.show()
```

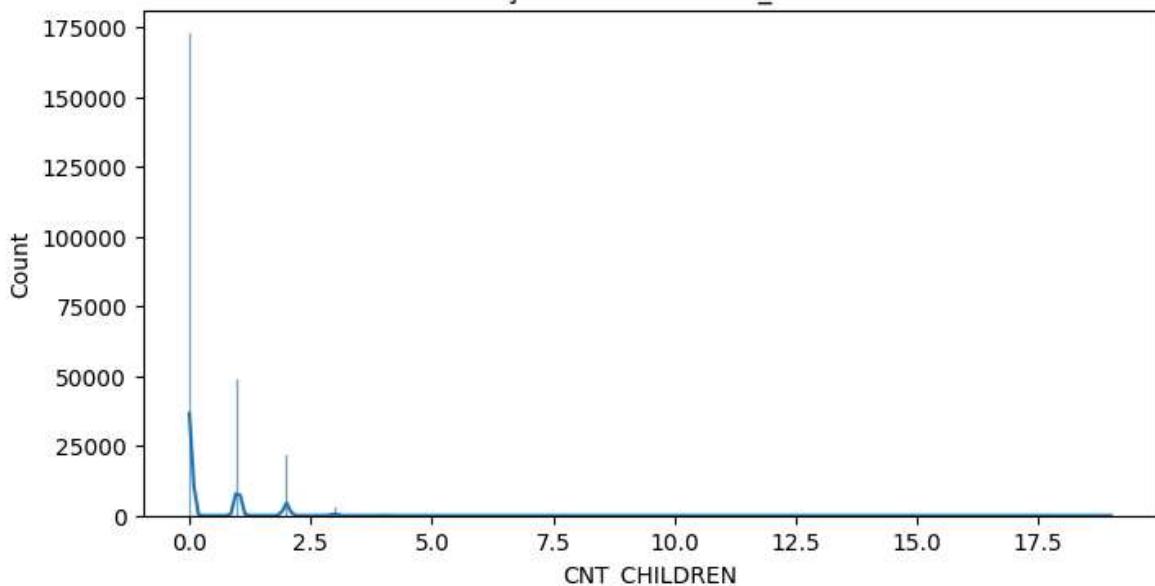
Distribuição da coluna SK\_ID\_CURR



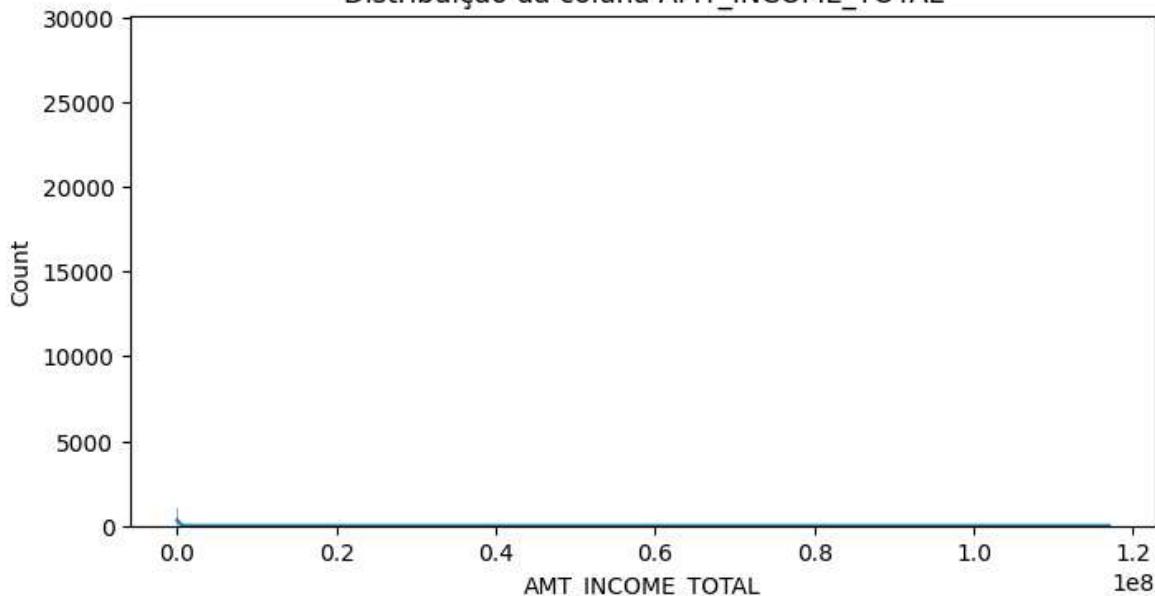
Distribuição da coluna TARGET



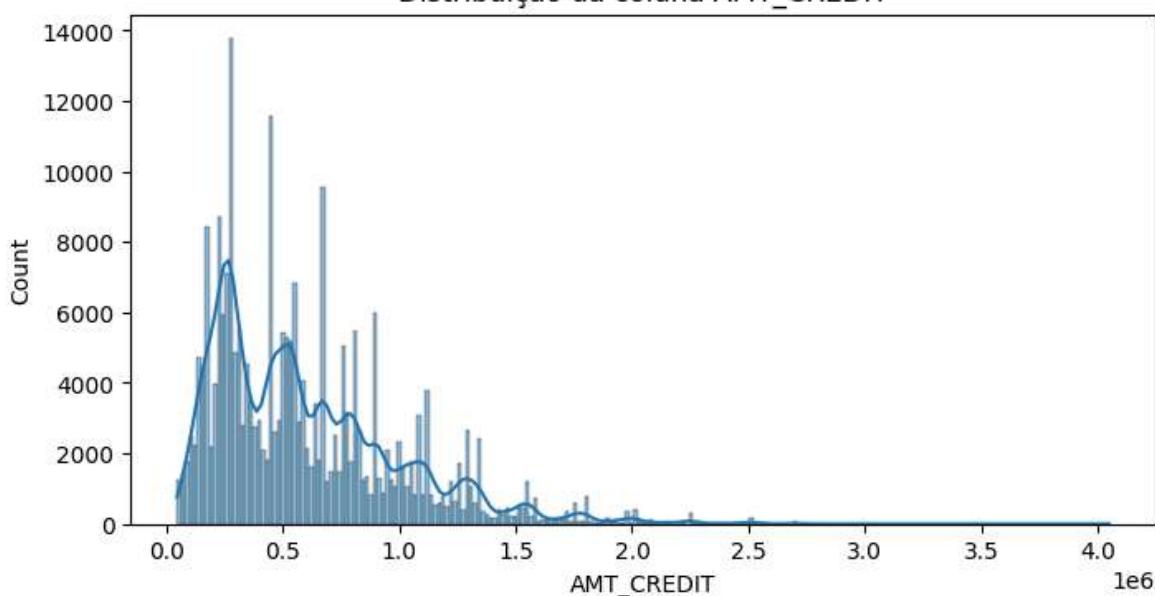
## Distribuição da coluna CNT\_CHILDREN



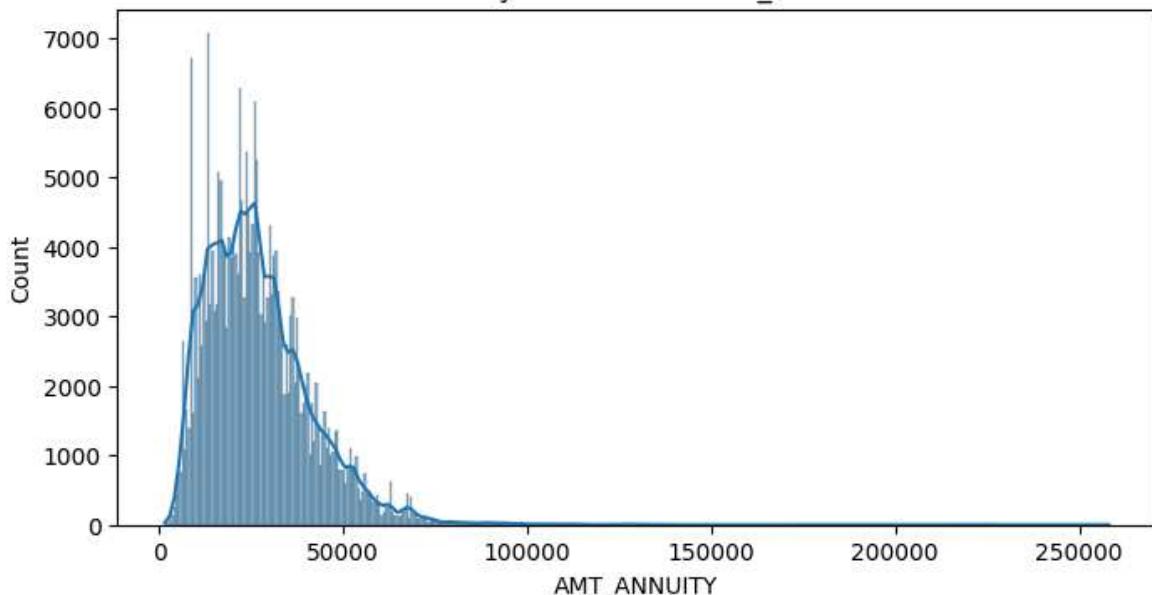
## Distribuição da coluna AMT\_INCOME\_TOTAL



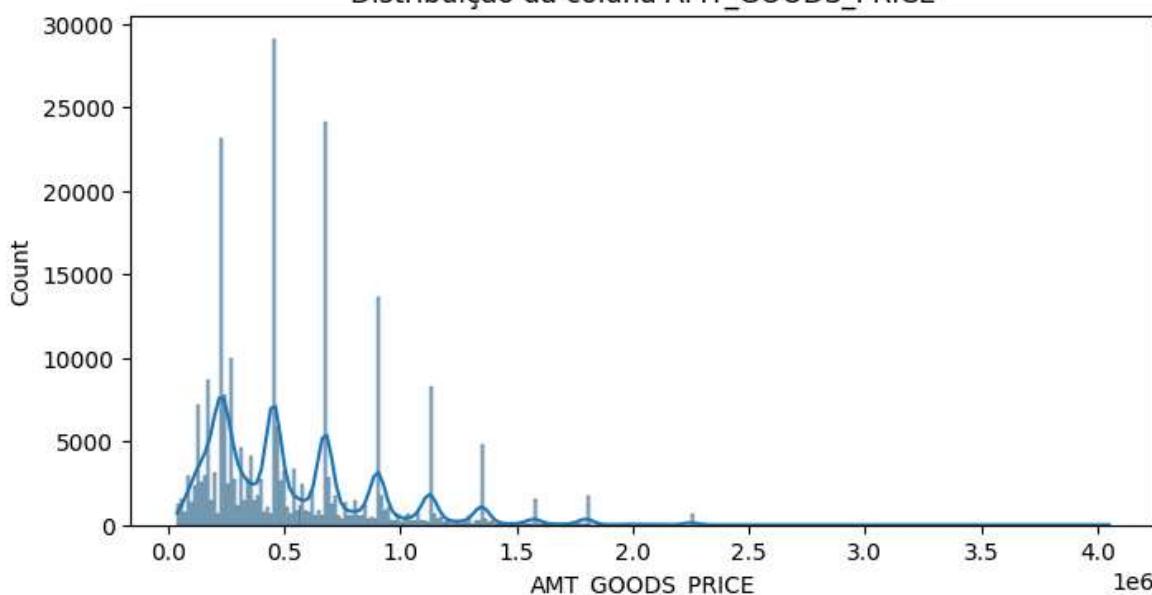
## Distribuição da coluna AMT\_CREDIT



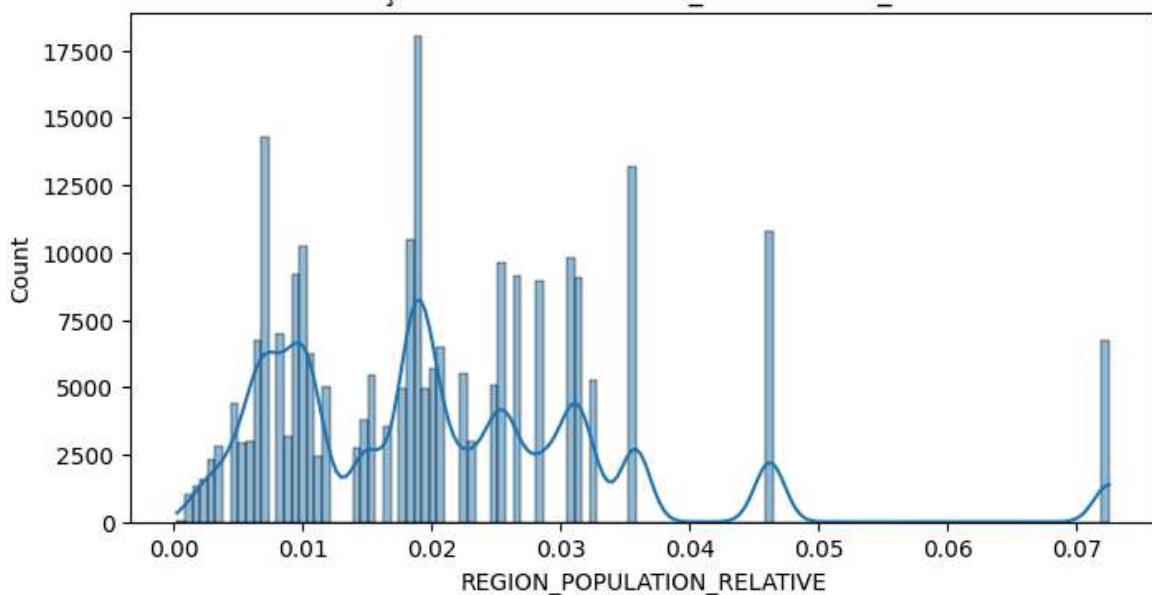
## Distribuição da coluna AMT\_ANNUITY



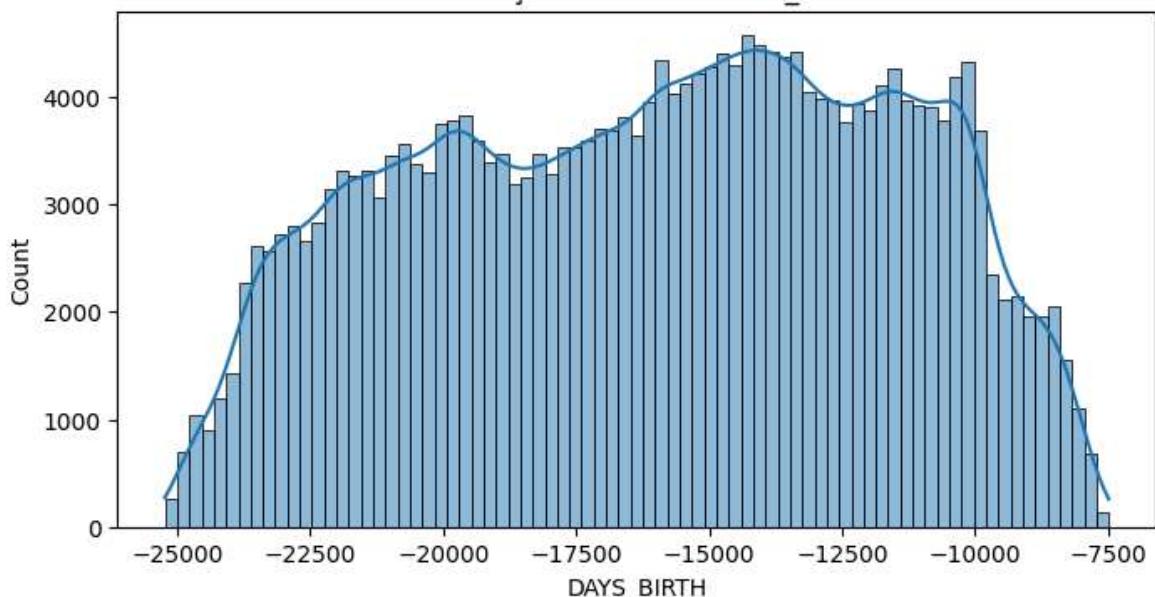
## Distribuição da coluna AMT\_GOODS\_PRICE



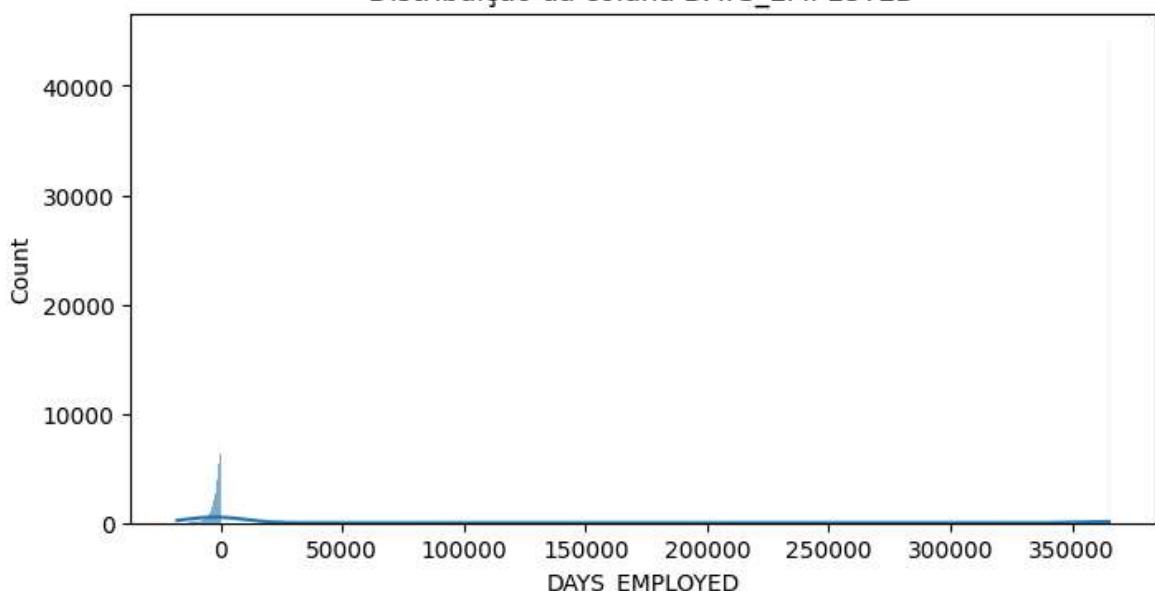
## Distribuição da coluna REGION\_POPULATION\_RELATIVE



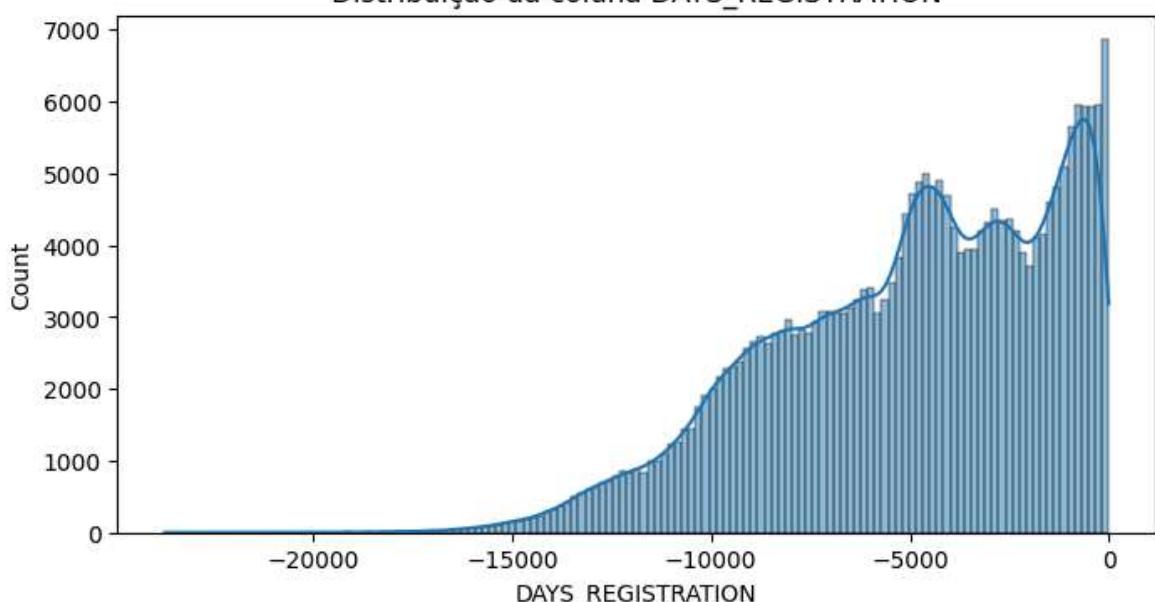
## Distribuição da coluna DAYS\_BIRTH



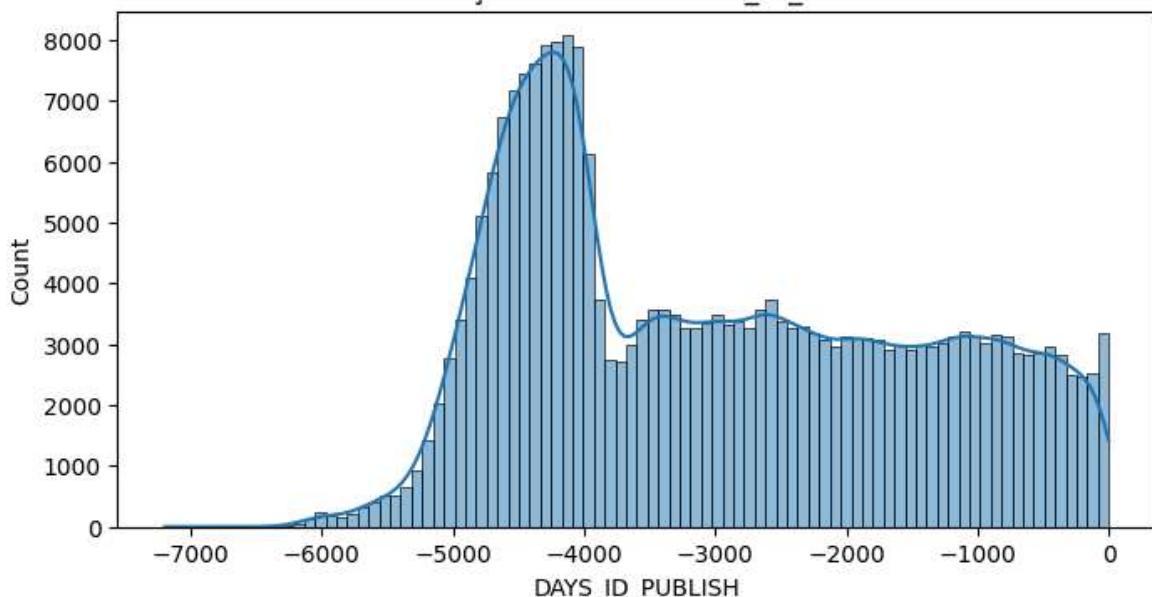
## Distribuição da coluna DAYS\_EMPLOYED



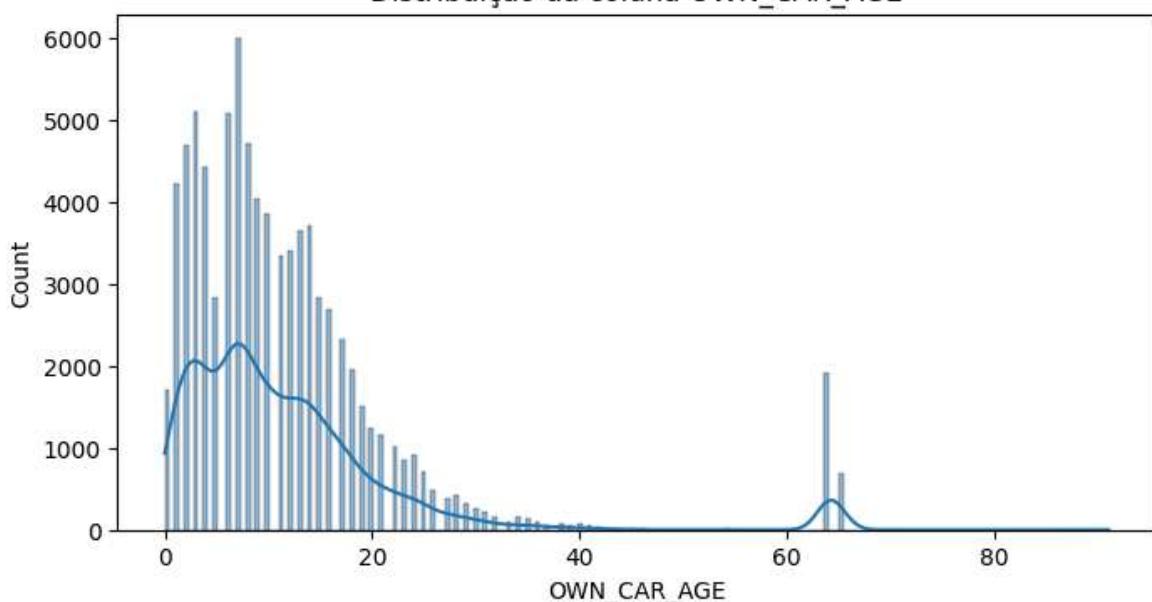
## Distribuição da coluna DAYS\_REGISTRATION



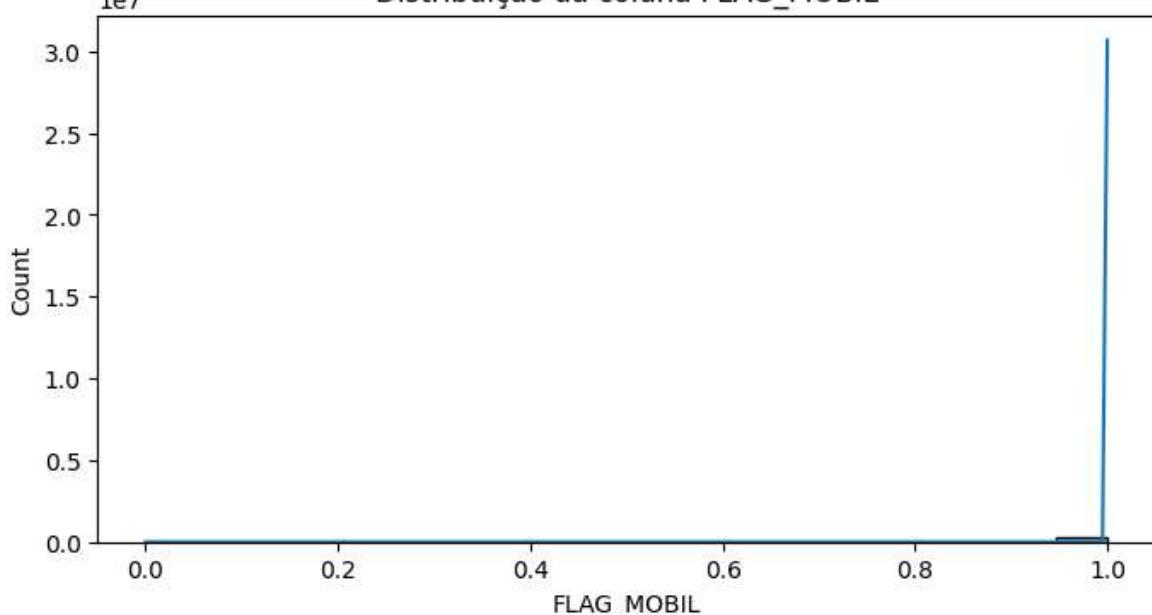
Distribuição da coluna DAYS\_ID\_PUBLISH



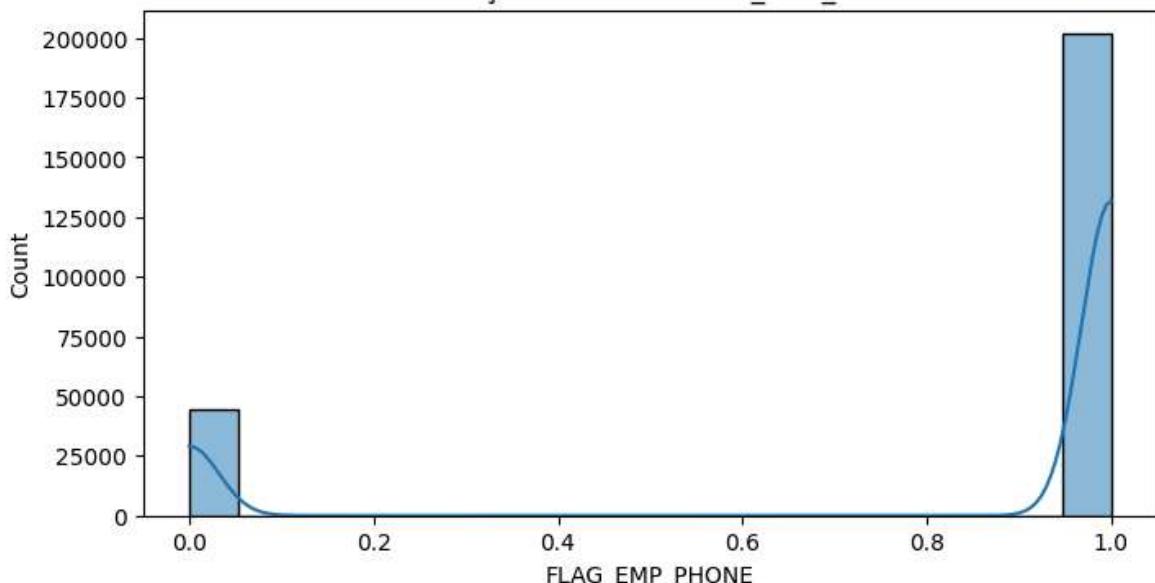
Distribuição da coluna OWN\_CAR\_AGE



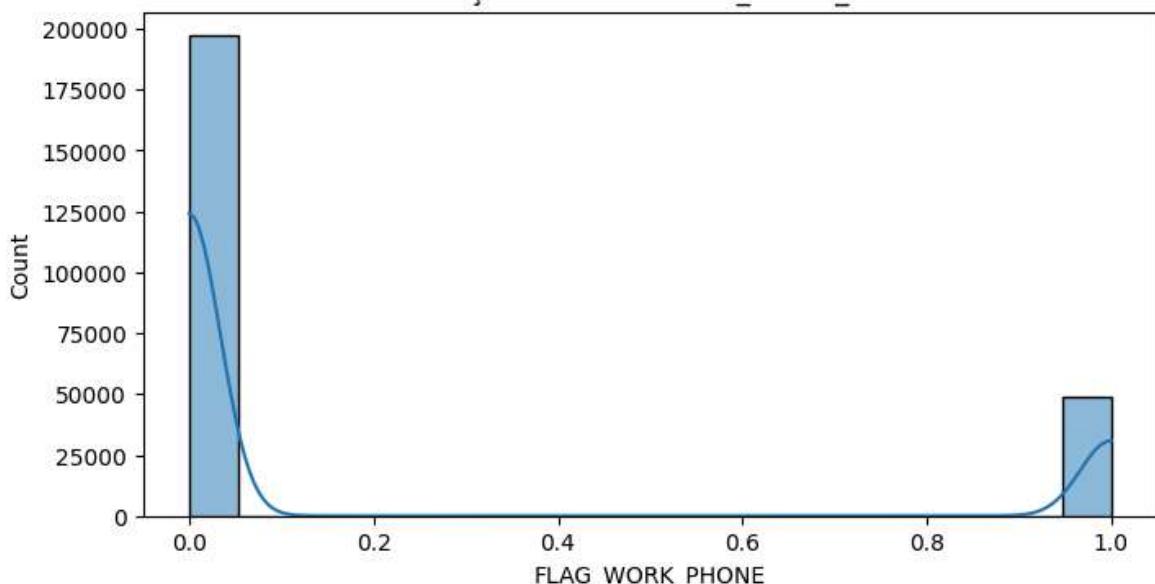
Distribuição da coluna FLAG\_MOBIL



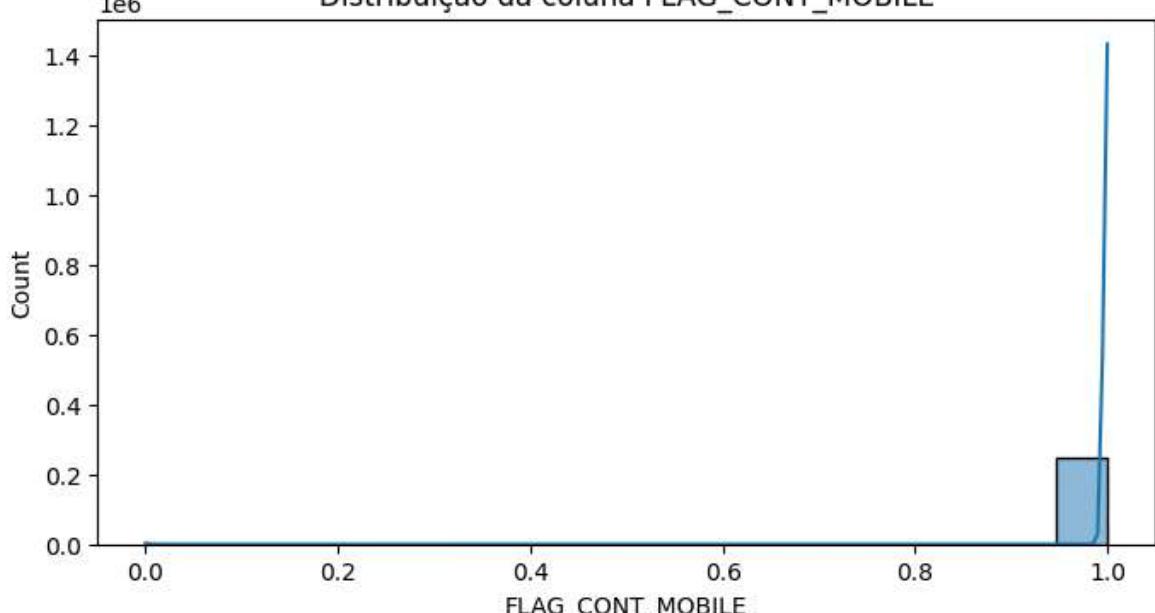
Distribuição da coluna FLAG\_EMP\_PHONE



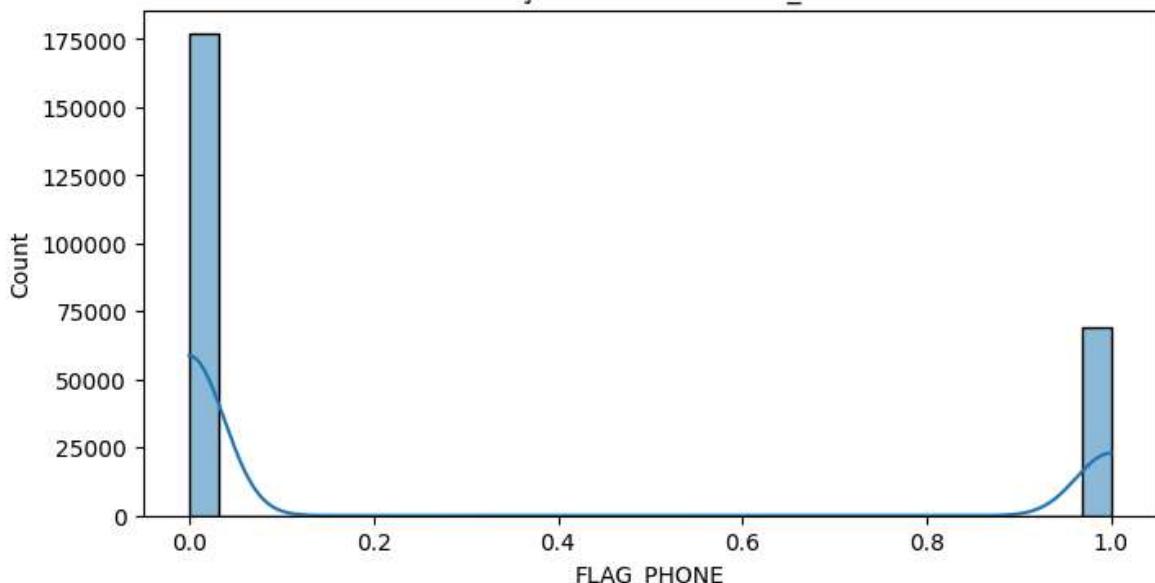
Distribuição da coluna FLAG\_WORK\_PHONE



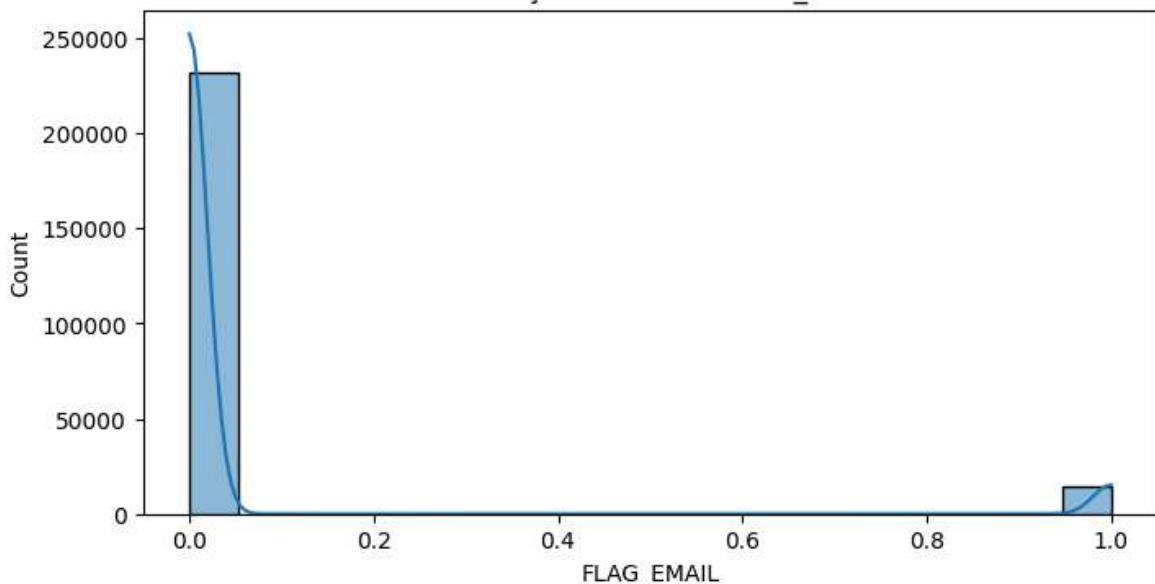
Distribuição da coluna FLAG\_CONT\_MOBILE



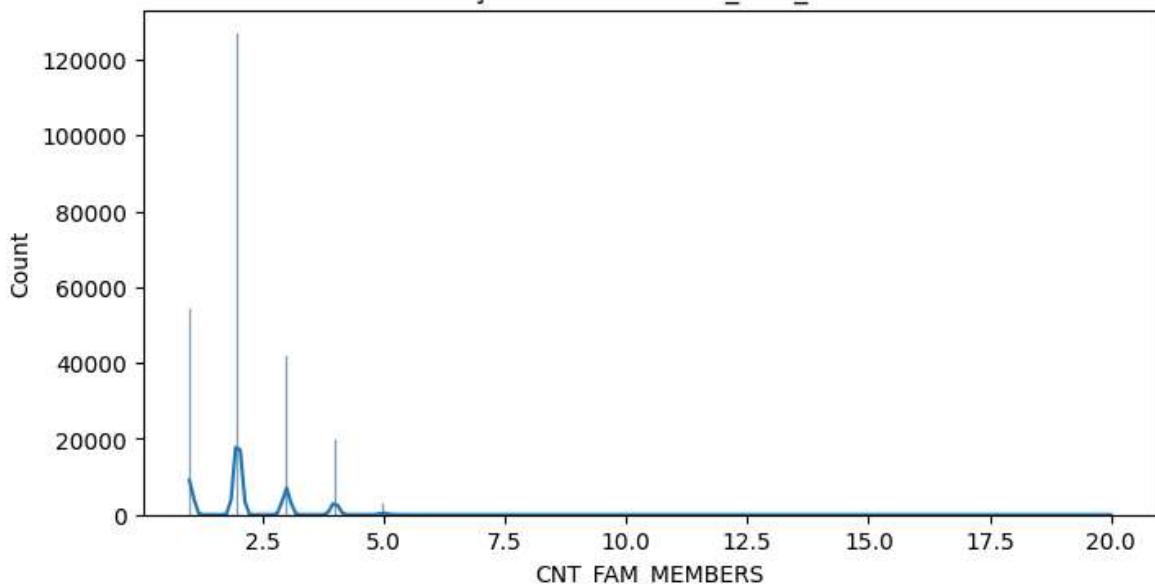
## Distribuição da coluna FLAG\_PHONE



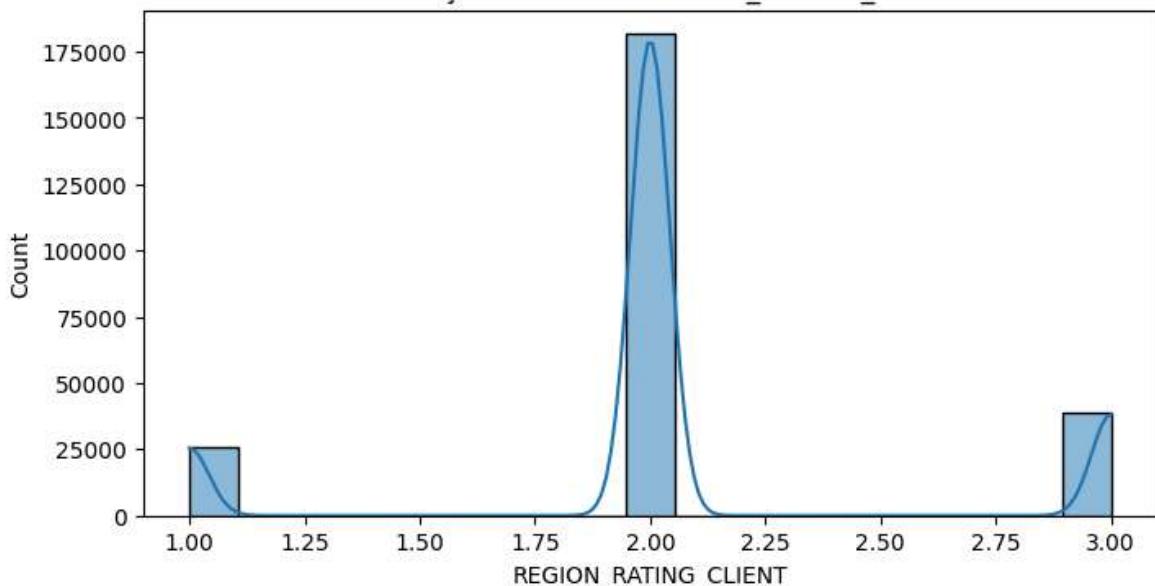
## Distribuição da coluna FLAG\_EMAIL



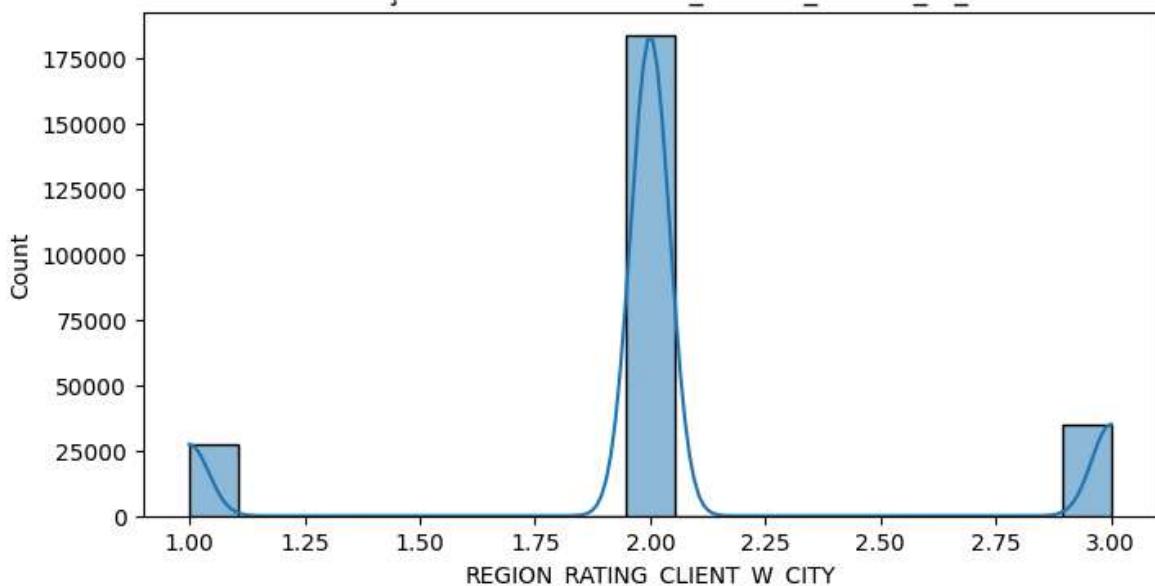
## Distribuição da coluna CNT\_FAM\_MEMBERS



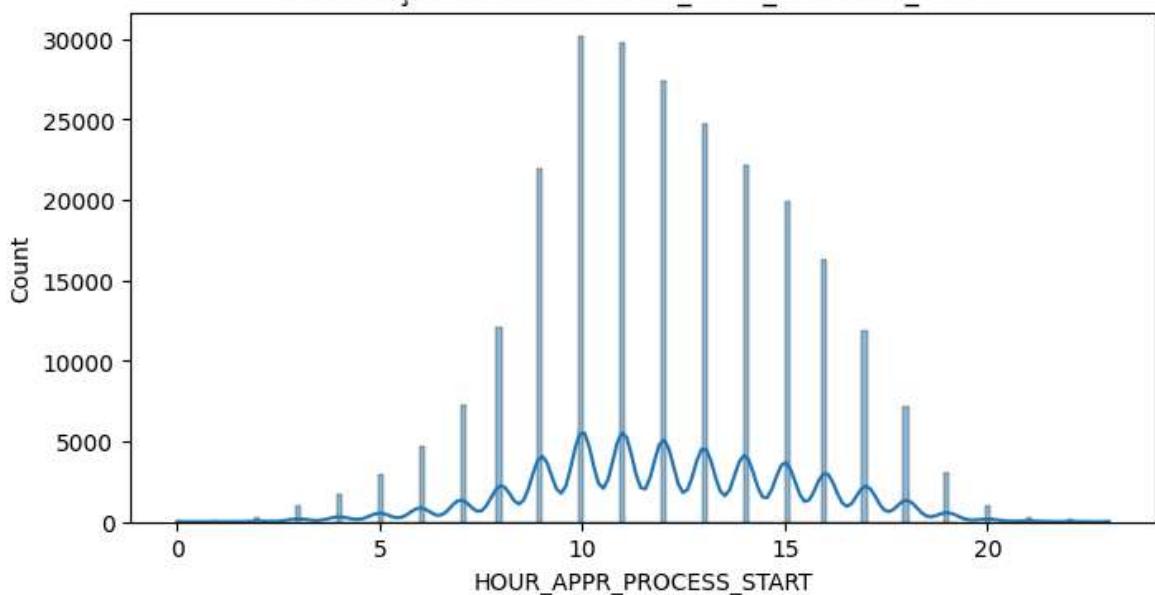
Distribuição da coluna REGION\_RATING\_CLIENT



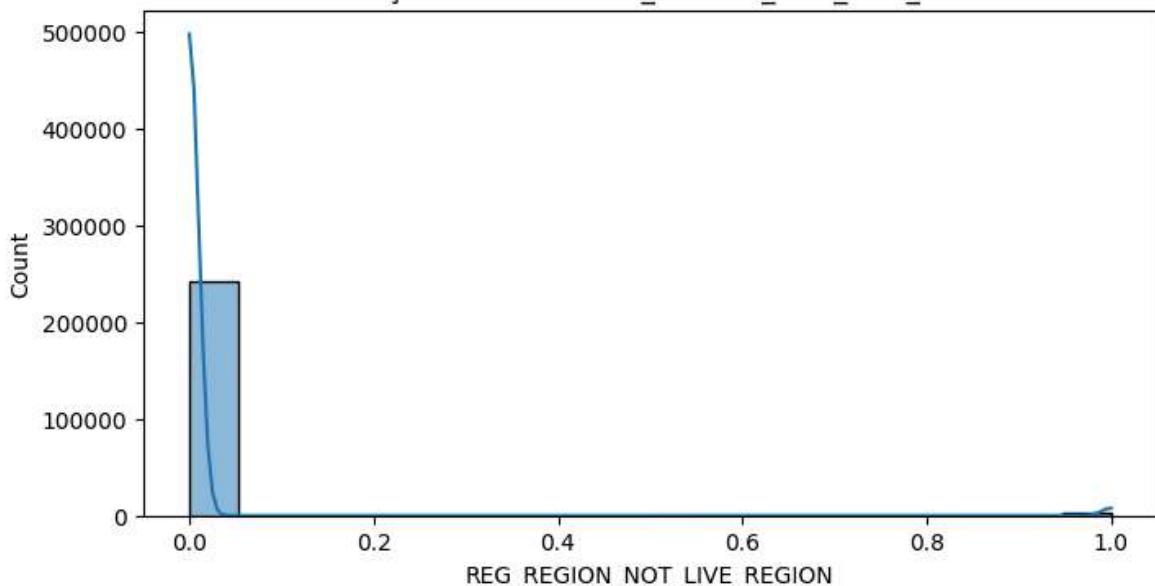
Distribuição da coluna REGION\_RATING\_CLIENT\_W\_CITY



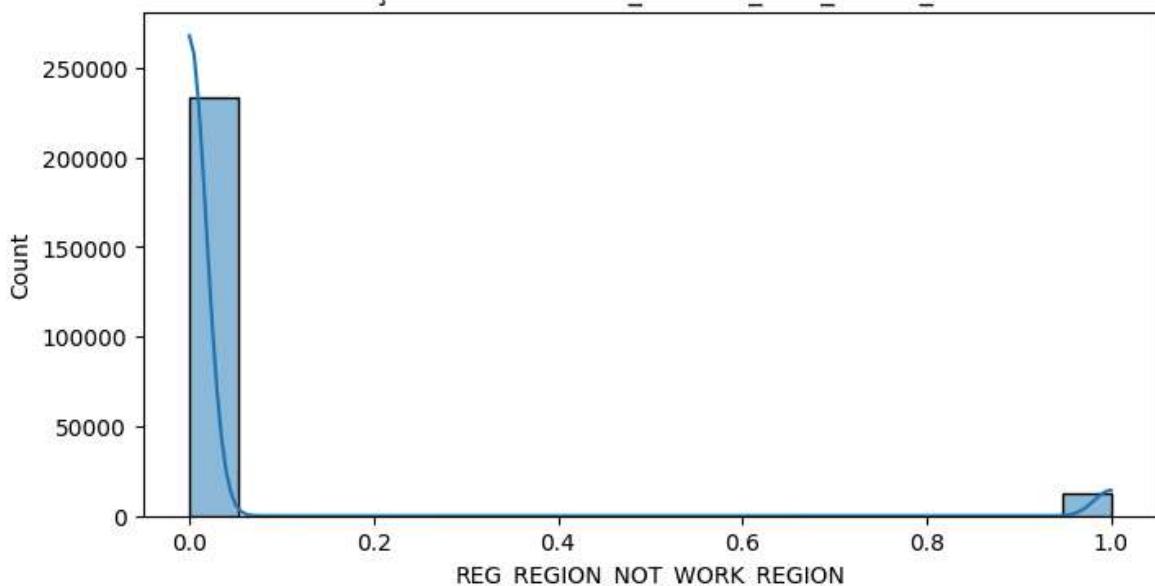
Distribuição da coluna HOUR\_APPR\_PROCESS\_START



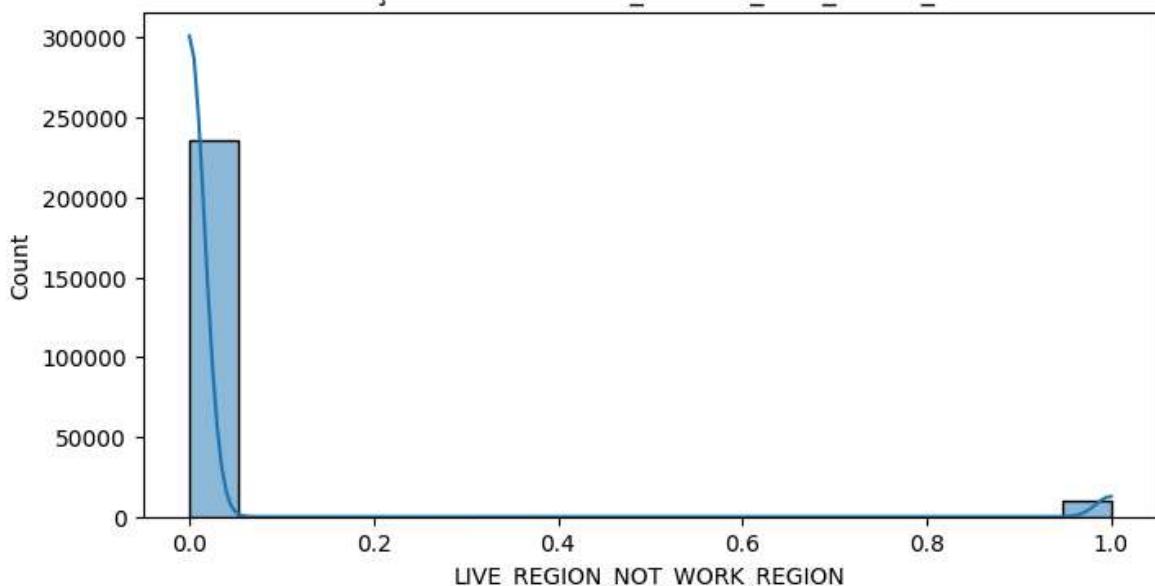
Distribuição da coluna REG\_REGION\_NOT\_LIVE\_REGION



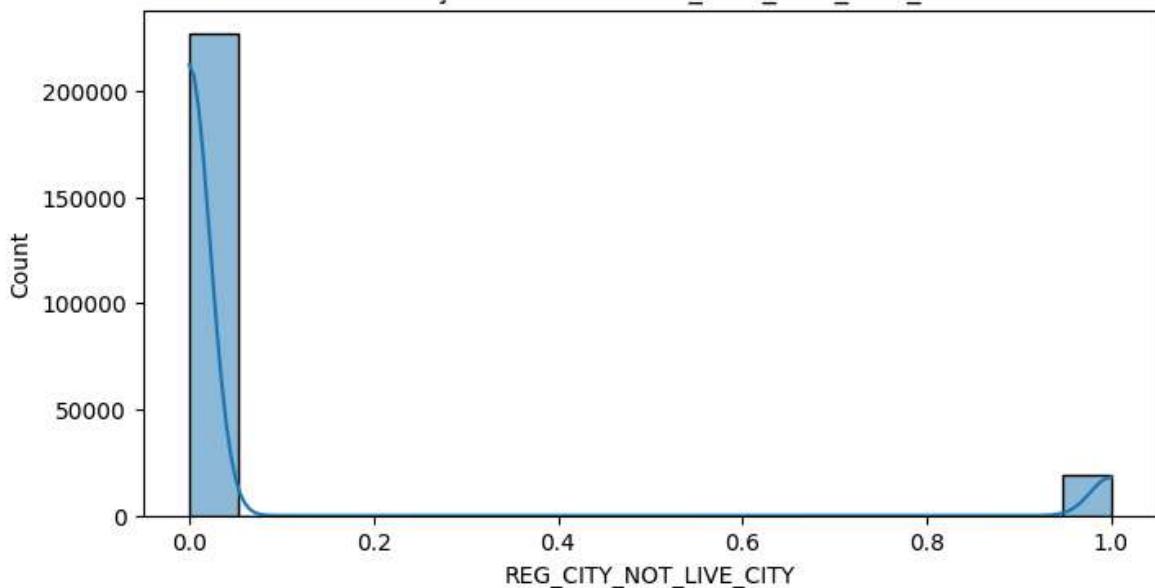
Distribuição da coluna REG\_REGION\_NOT\_WORK\_REGION



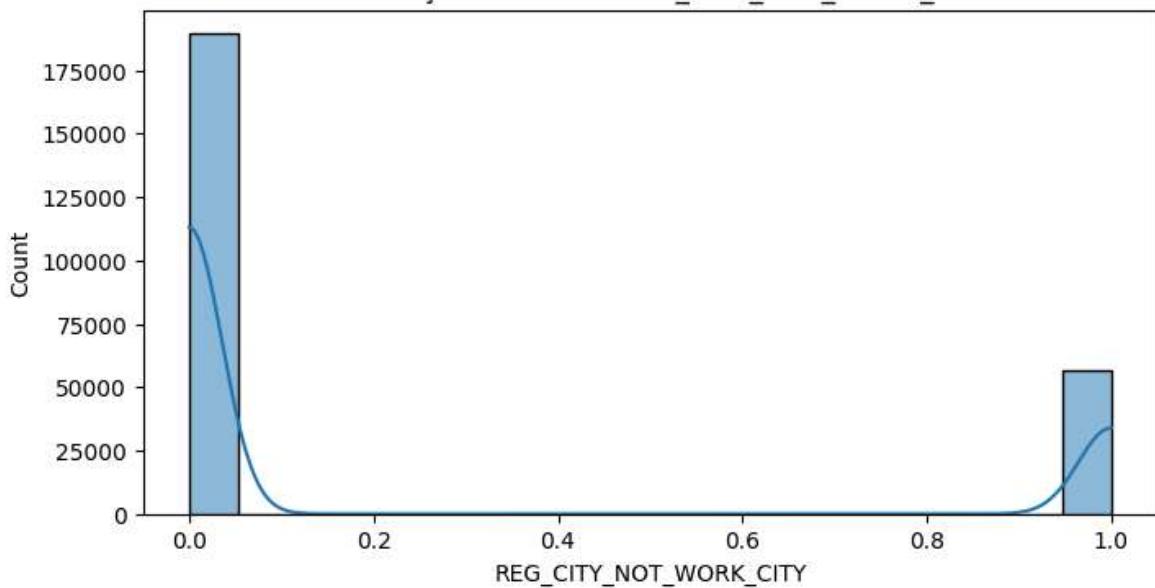
Distribuição da coluna LIVE\_REGION\_NOT\_WORK\_REGION



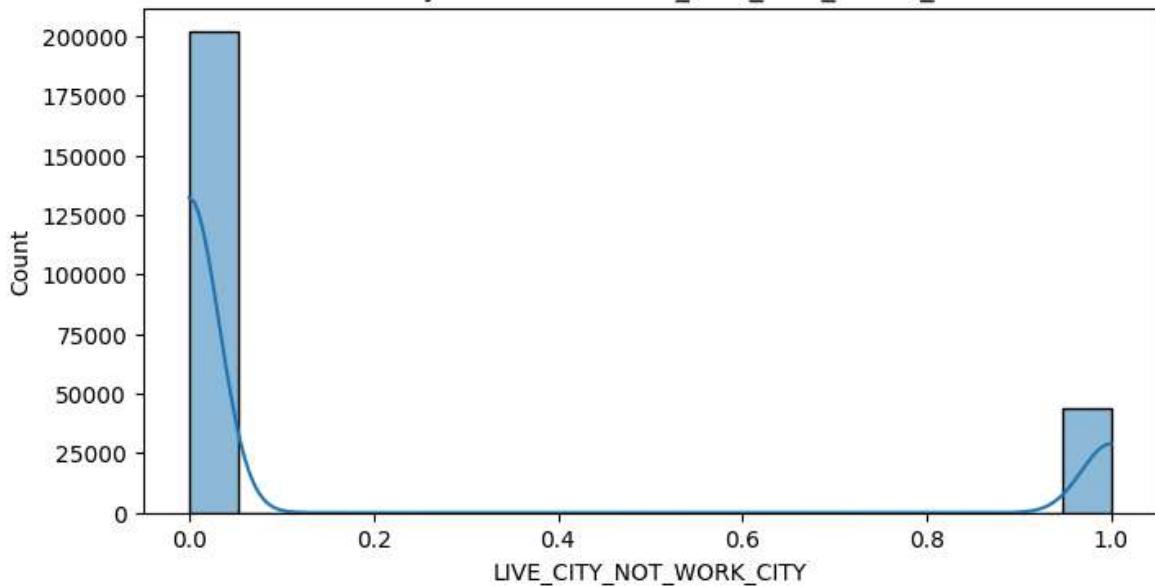
Distribuição da coluna REG\_CITY\_NOT\_LIVE\_CITY



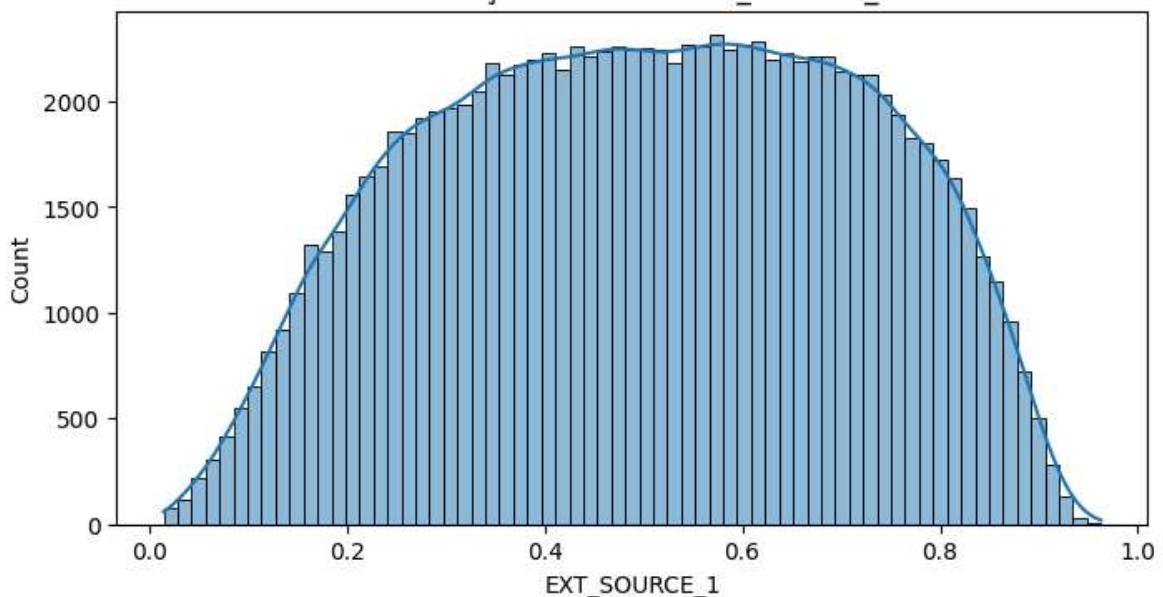
Distribuição da coluna REG\_CITY\_NOT\_WORK\_CITY



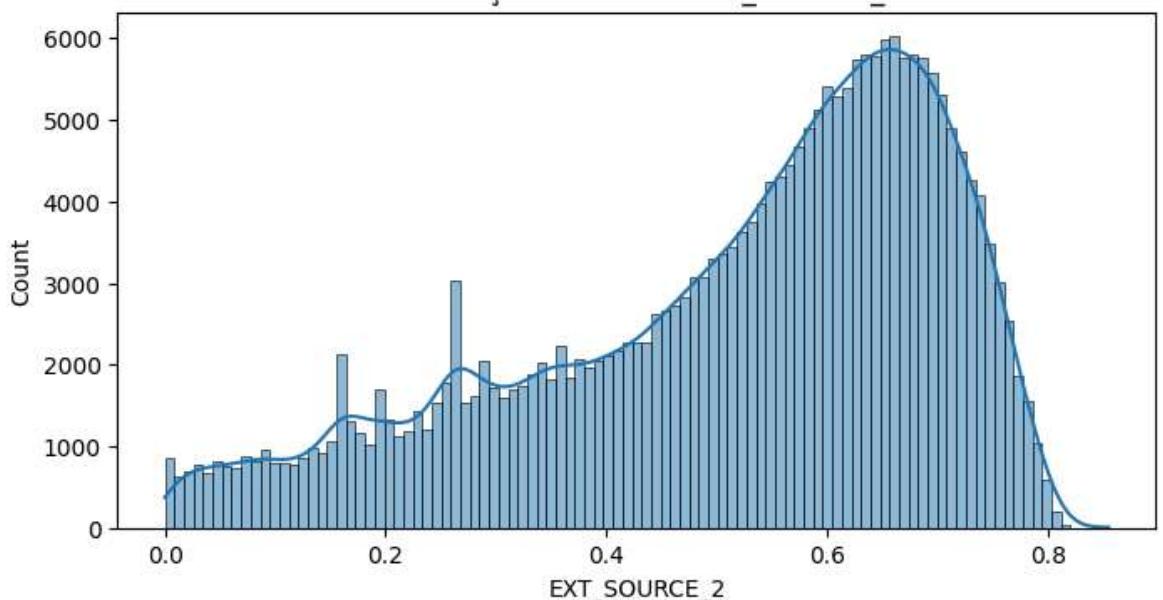
Distribuição da coluna LIVE\_CITY\_NOT\_WORK\_CITY



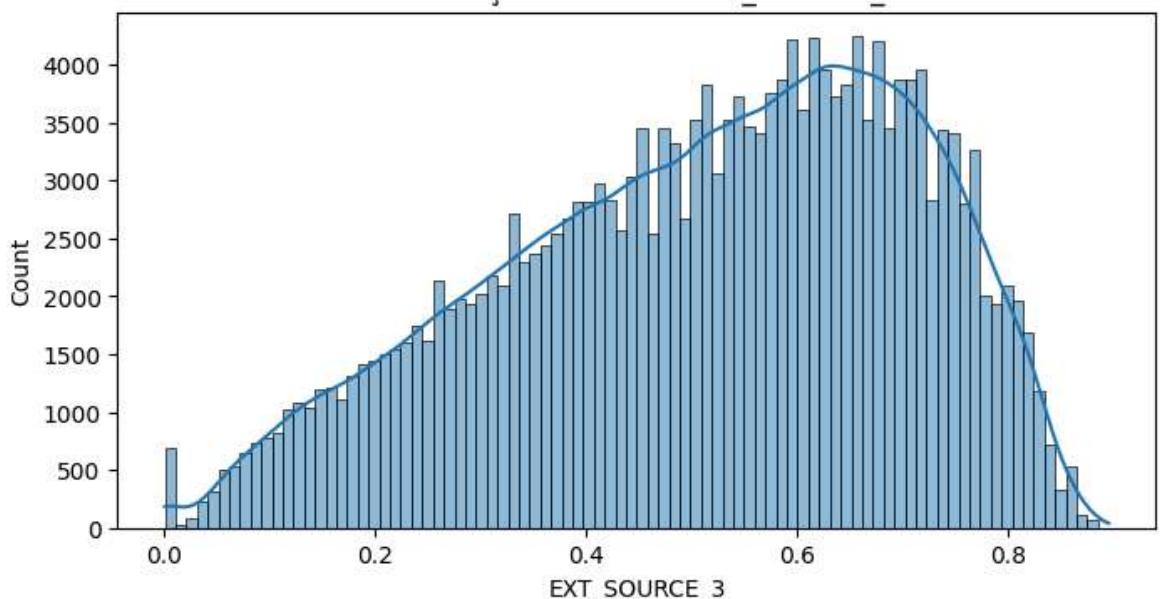
## Distribuição da coluna EXT\_SOURCE\_1



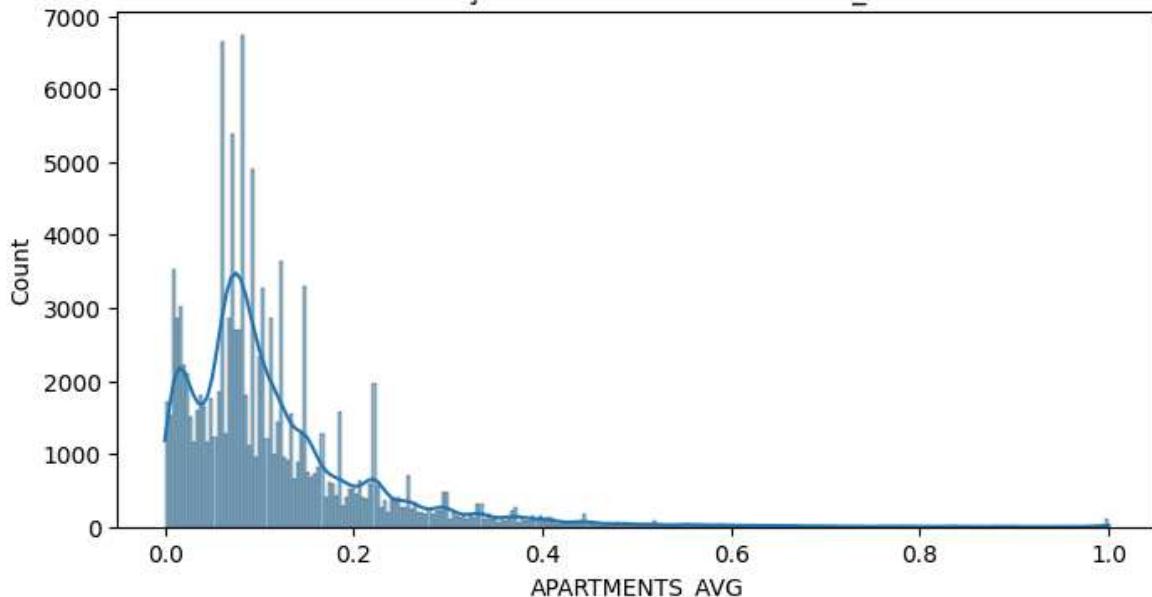
## Distribuição da coluna EXT\_SOURCE\_2



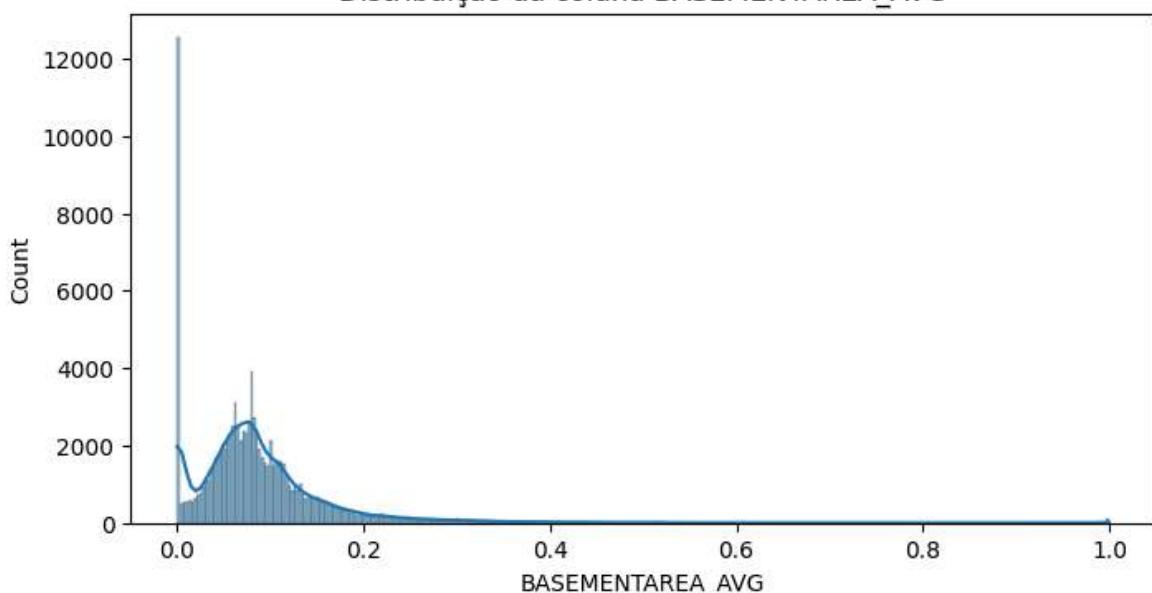
## Distribuição da coluna EXT\_SOURCE\_3



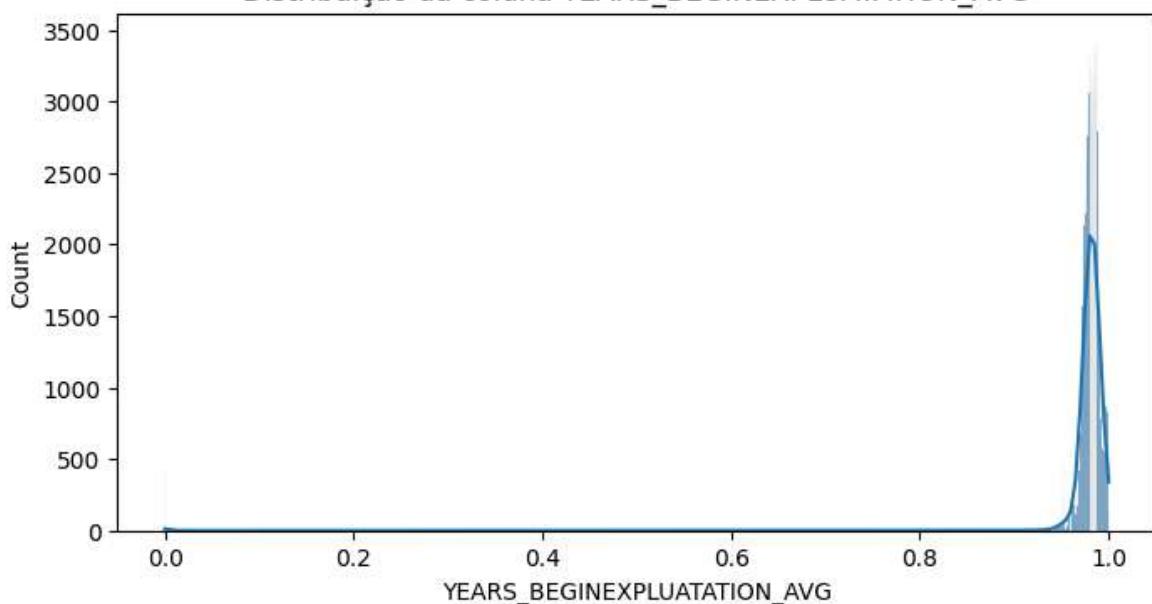
Distribuição da coluna APARTMENTS\_AVG



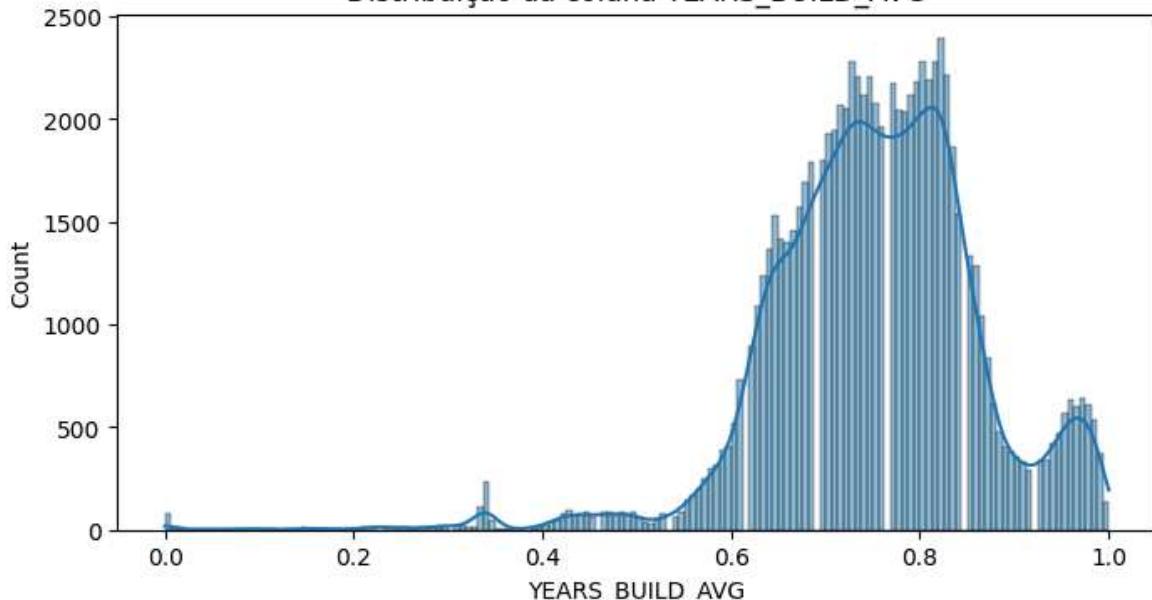
Distribuição da coluna BASEMENTAREA\_AVG



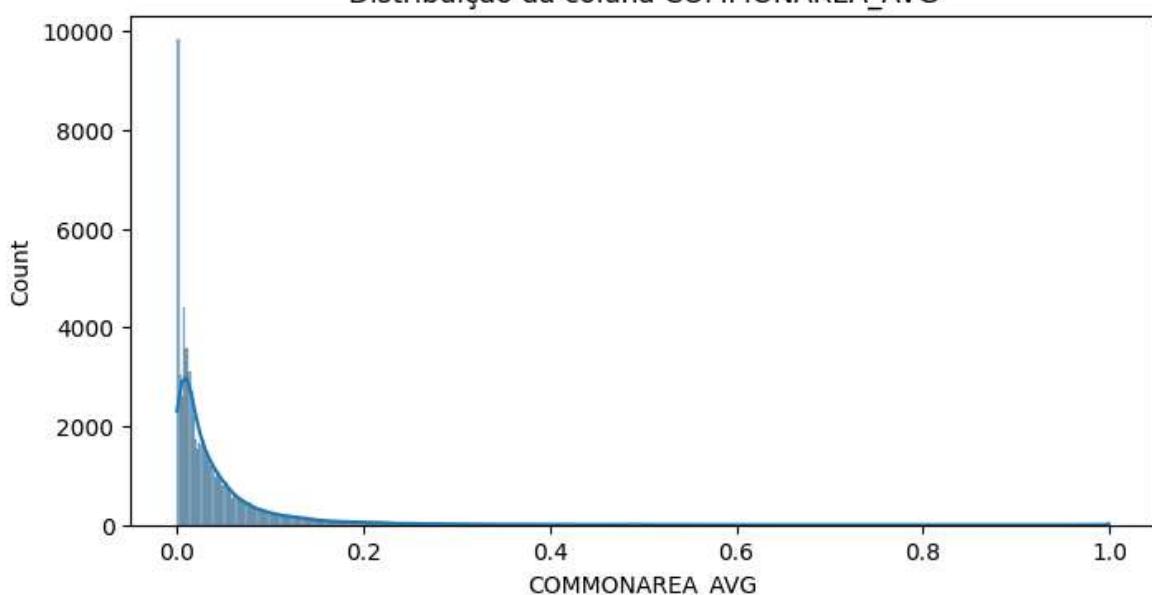
Distribuição da coluna YEARS\_BEGINEXPLUATATION\_AVG



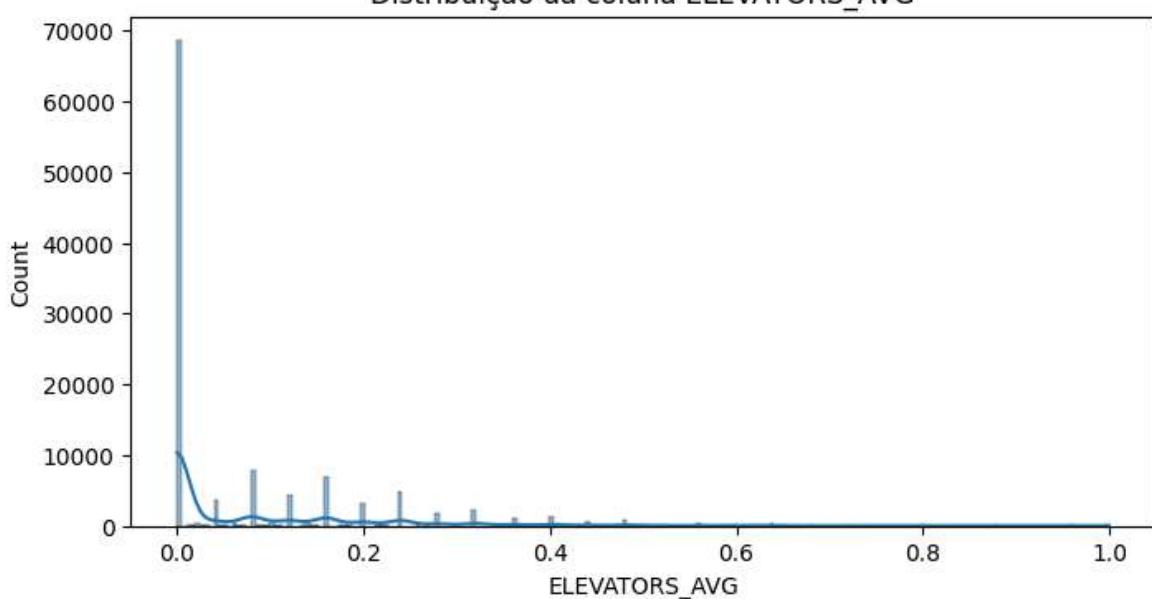
## Distribuição da coluna YEARS\_BUILD\_AVG



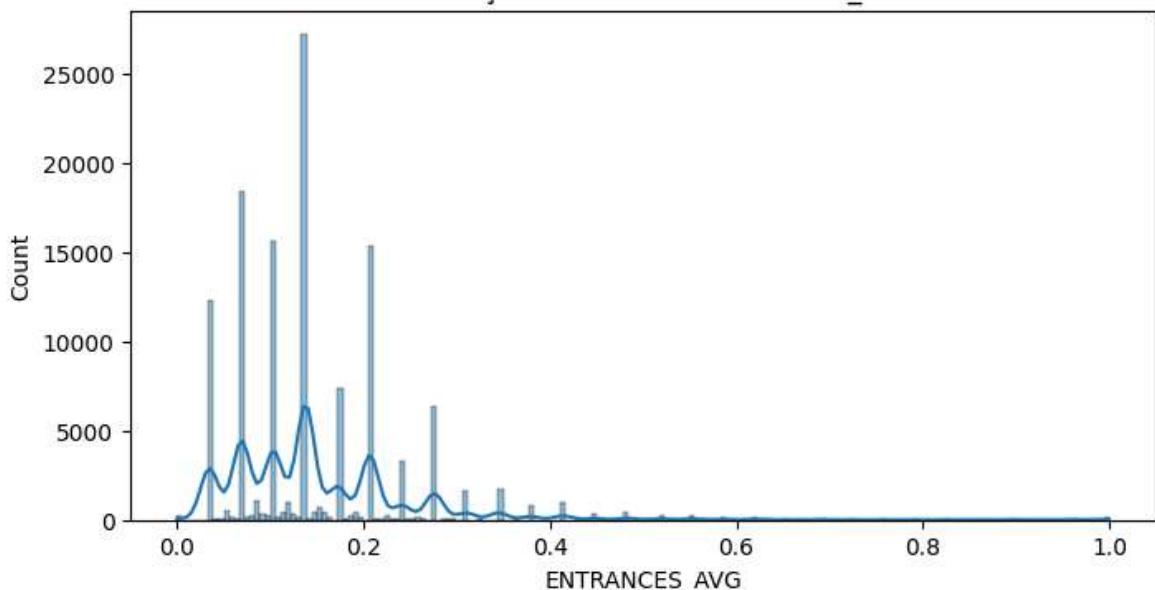
## Distribuição da coluna COMMONAREA\_AVG



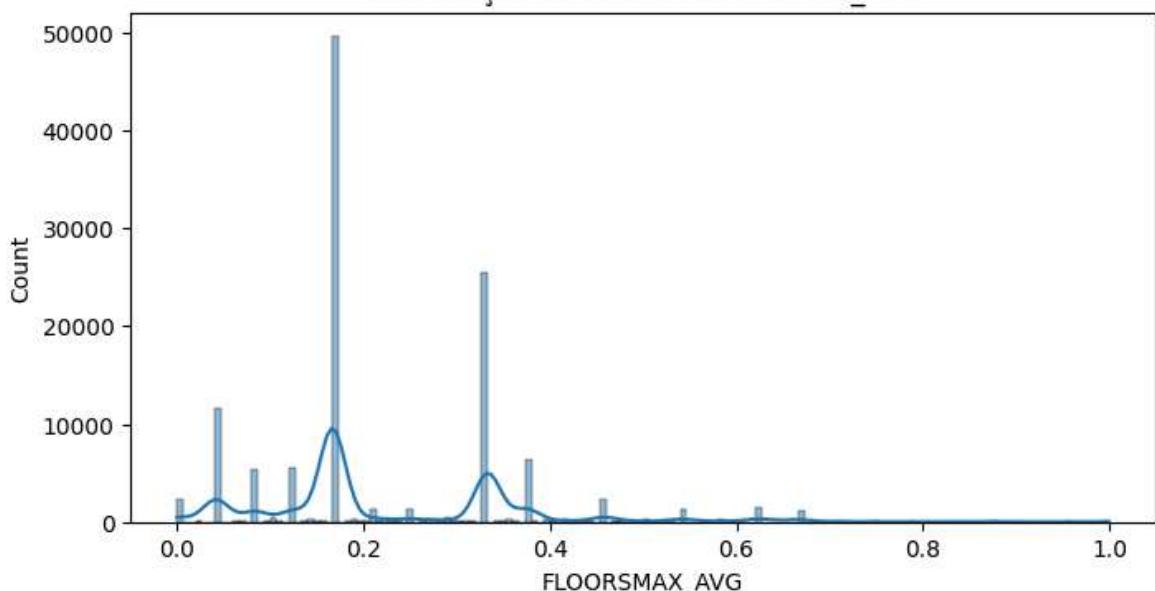
## Distribuição da coluna ELEVATORS\_AVG



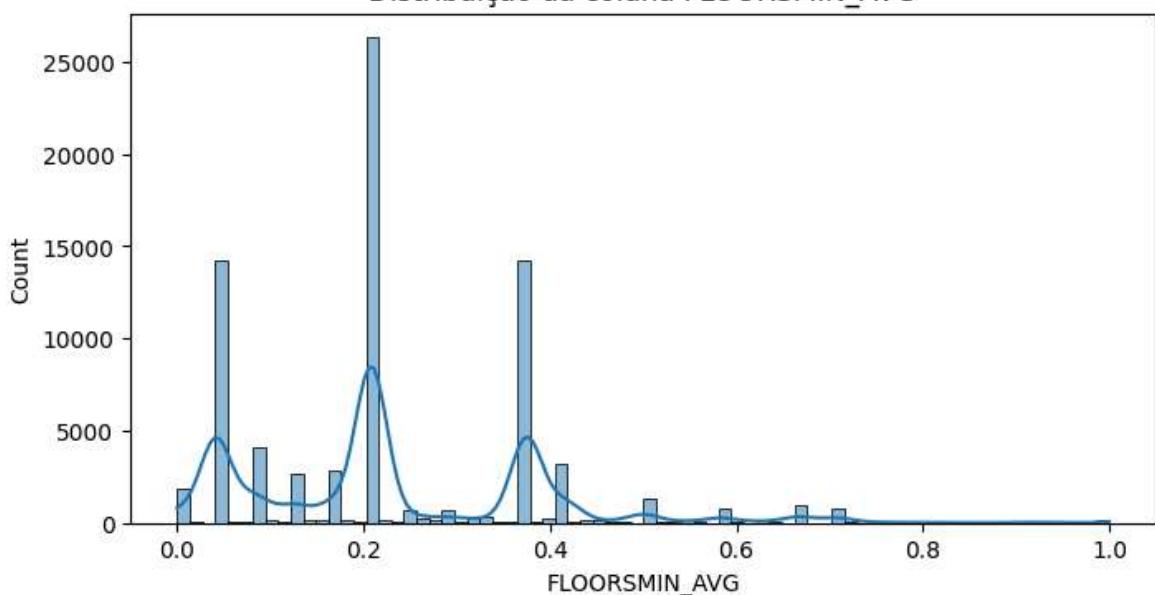
## Distribuição da coluna ENTRANCES\_AVG



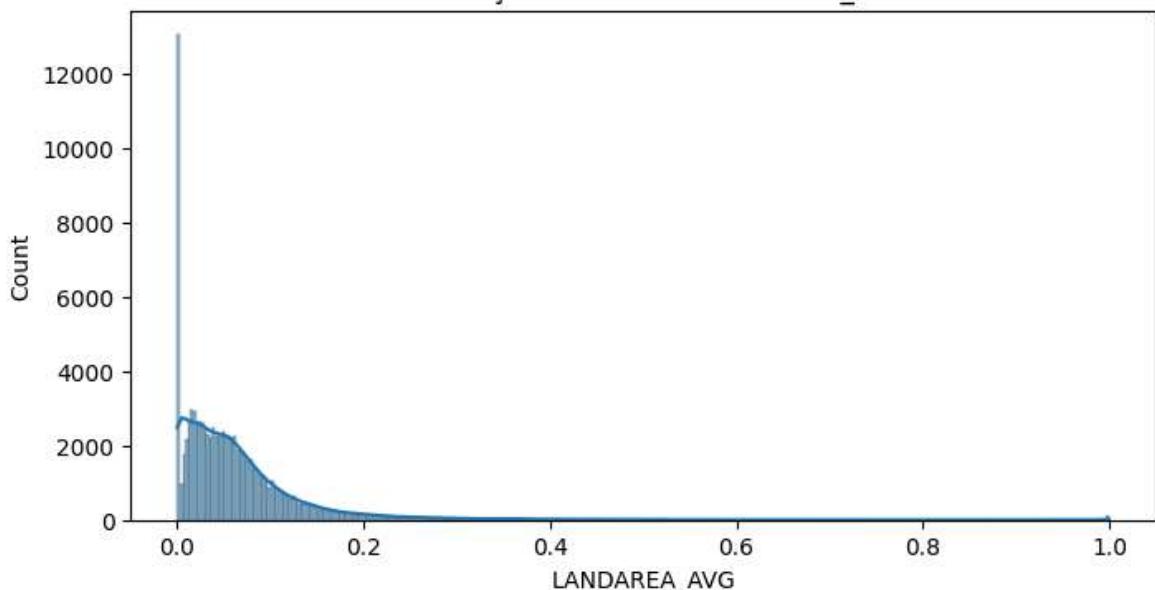
## Distribuição da coluna FLOORSMAX\_AVG



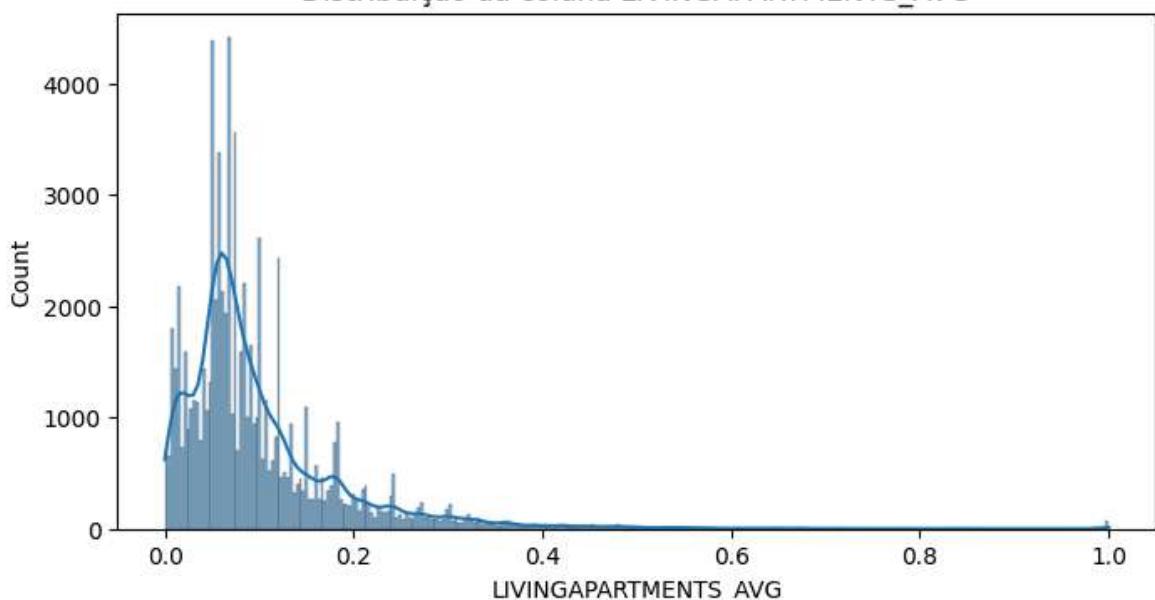
## Distribuição da coluna FLOORSMIN\_AVG



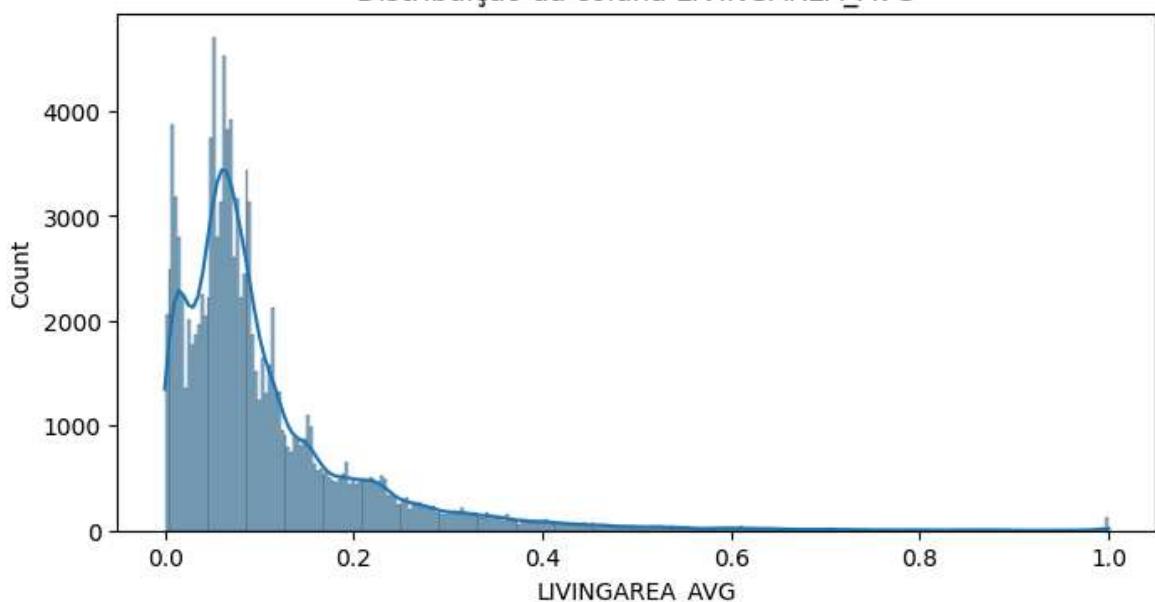
## Distribuição da coluna LANDAREA\_AVG



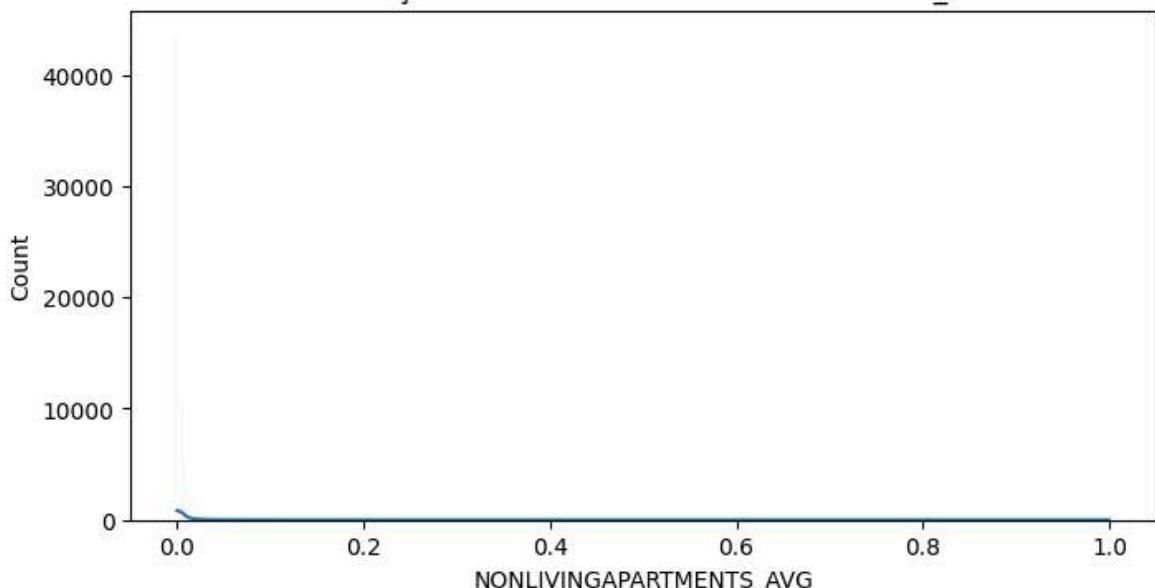
## Distribuição da coluna LIVINGAPARTMENTS\_AVG



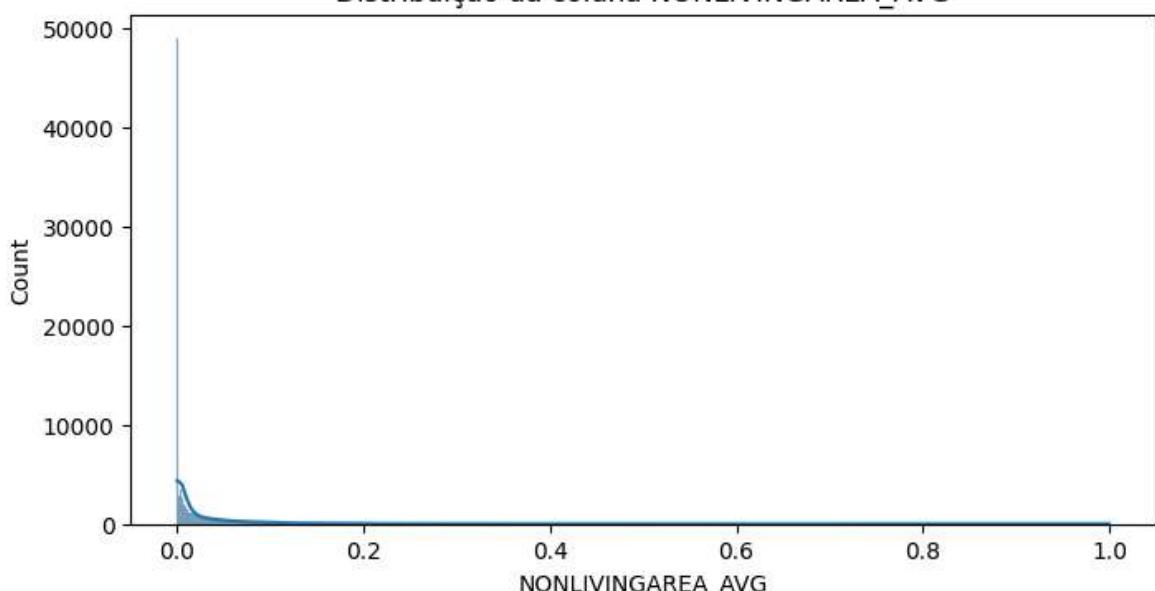
## Distribuição da coluna LIVINGAREA\_AVG



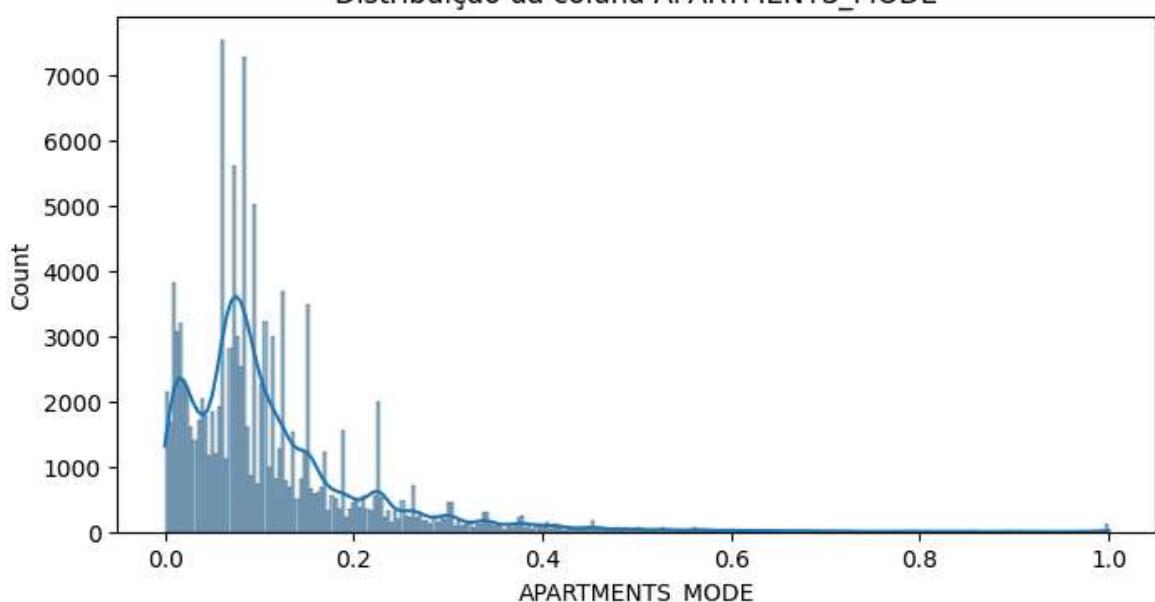
Distribuição da coluna NONLIVINGAPARTMENTS\_AVG



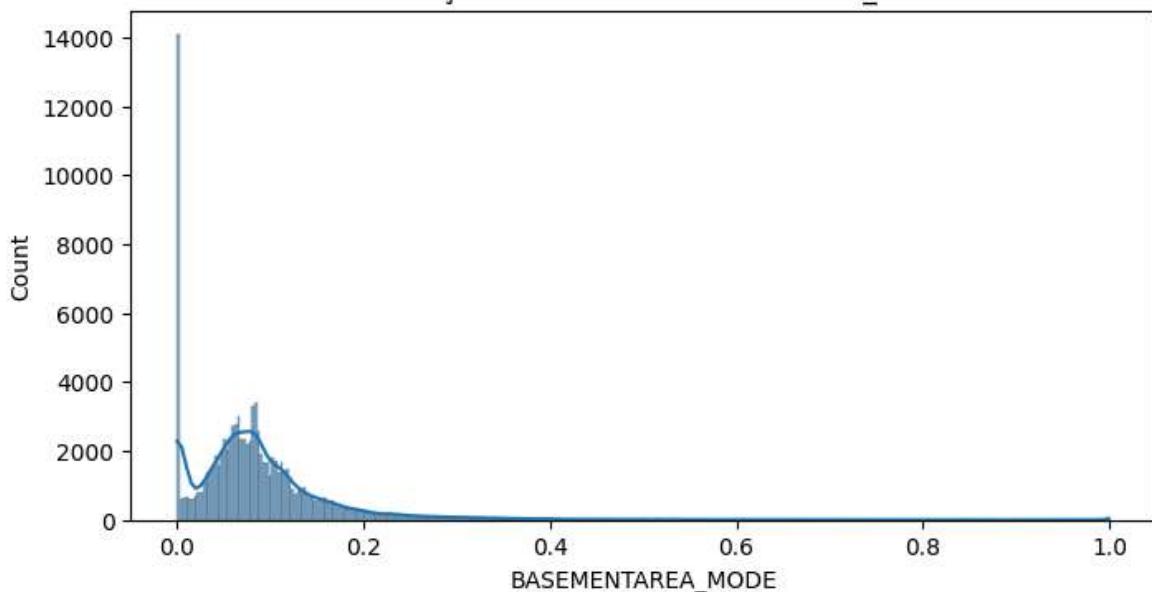
Distribuição da coluna NONLIVINGAREA\_AVG



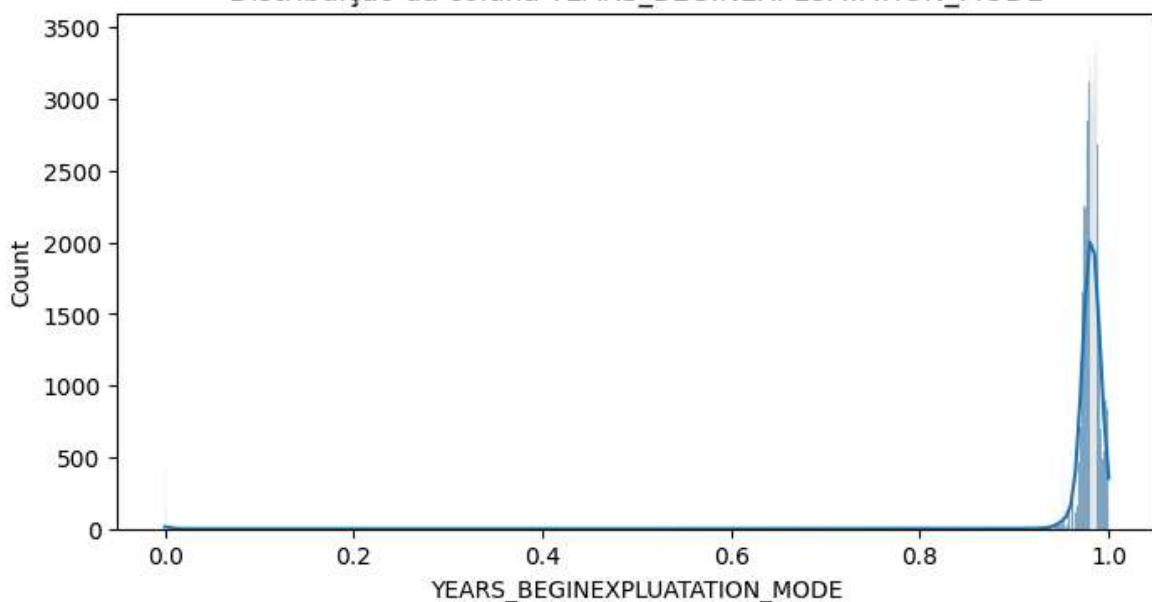
Distribuição da coluna APARTMENTS\_MODE



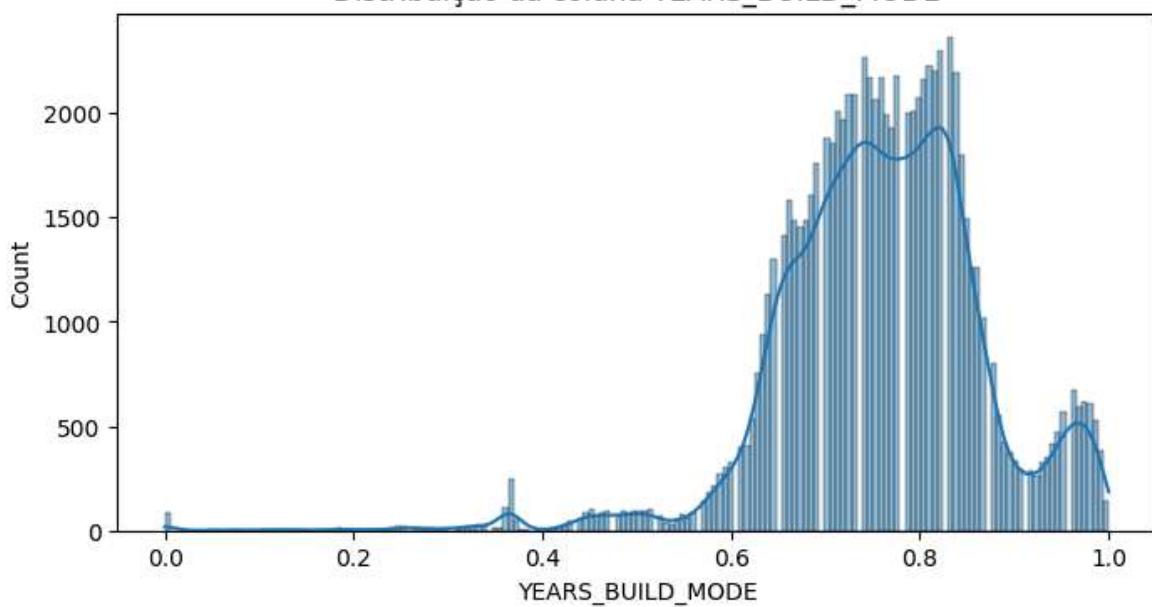
Distribuição da coluna BASEMENTAREA\_MODE



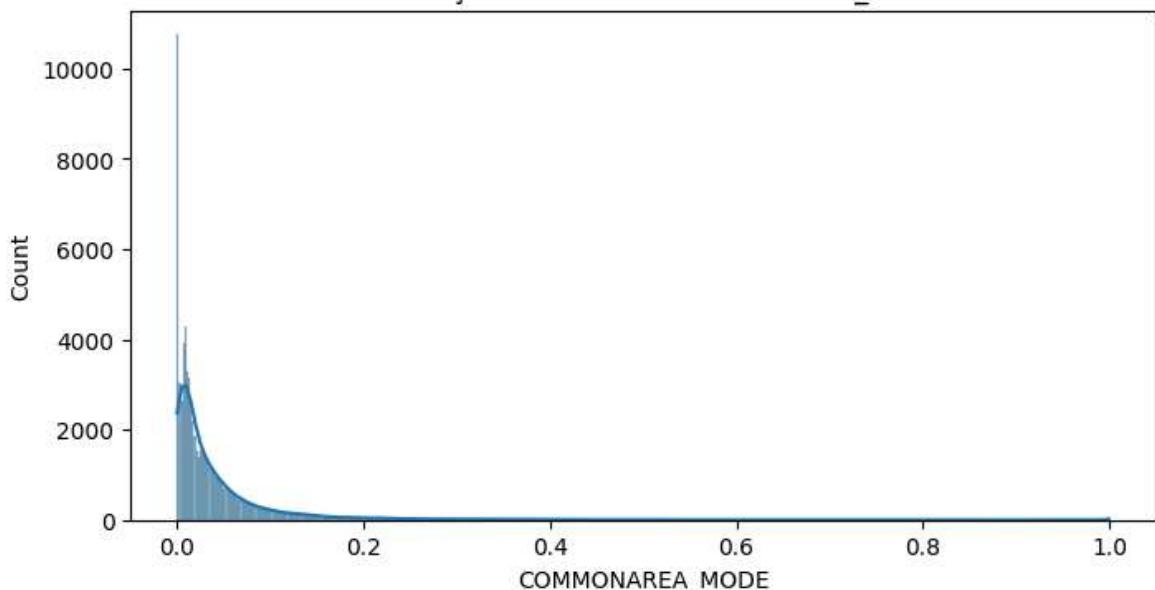
Distribuição da coluna YEARS\_BEGINEXPLUATATION\_MODE



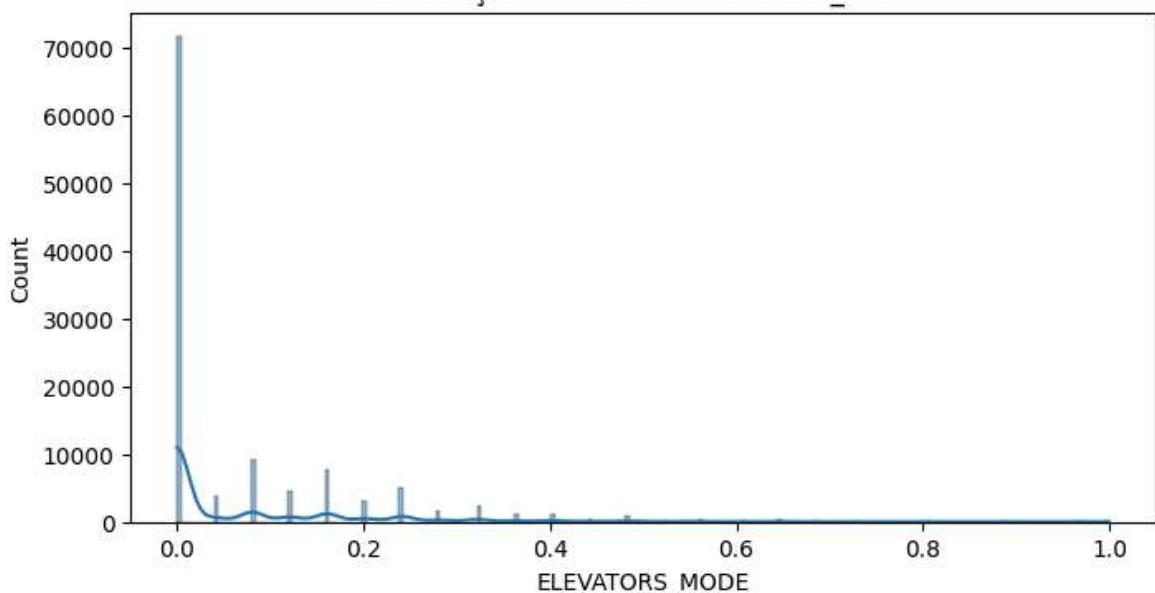
Distribuição da coluna YEARS\_BUILD\_MODE



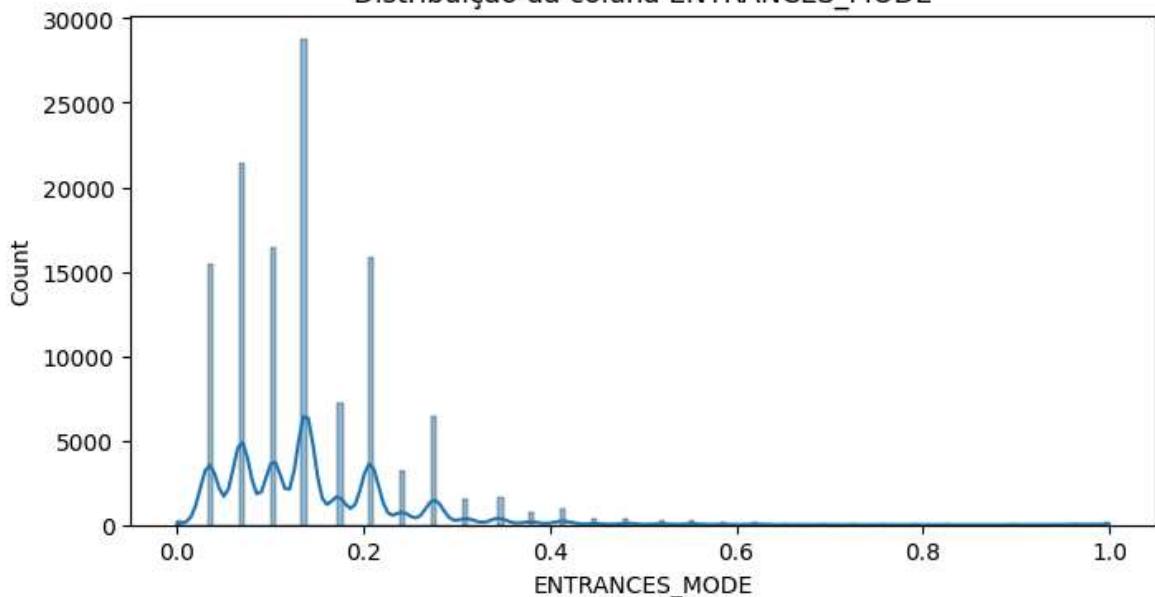
## Distribuição da coluna COMMONAREA\_MODE



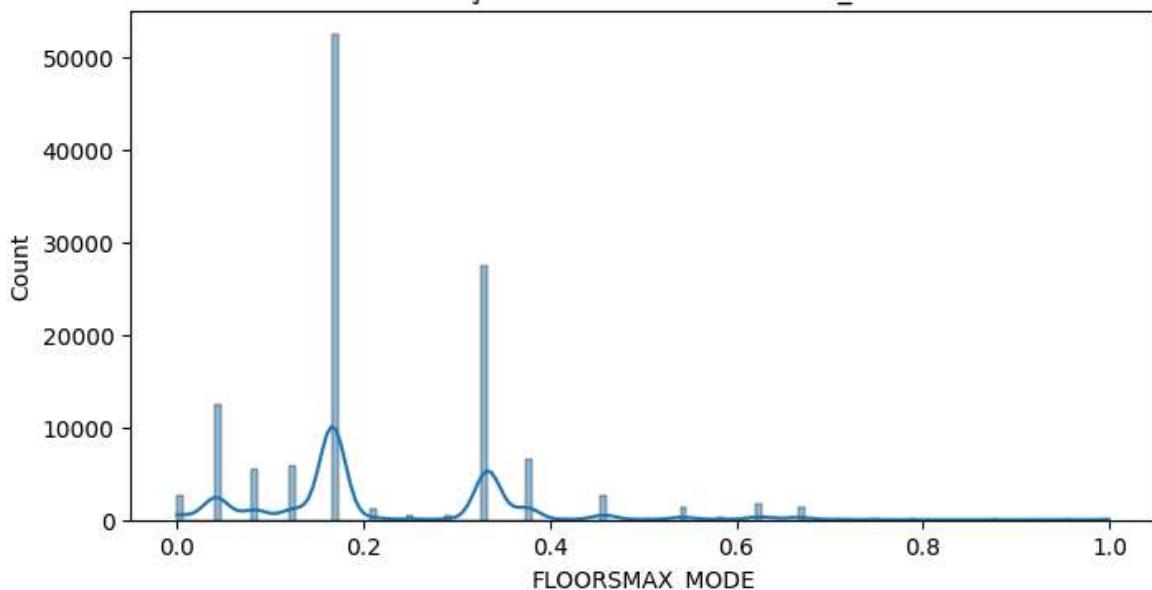
## Distribuição da coluna ELEVATORS\_MODE



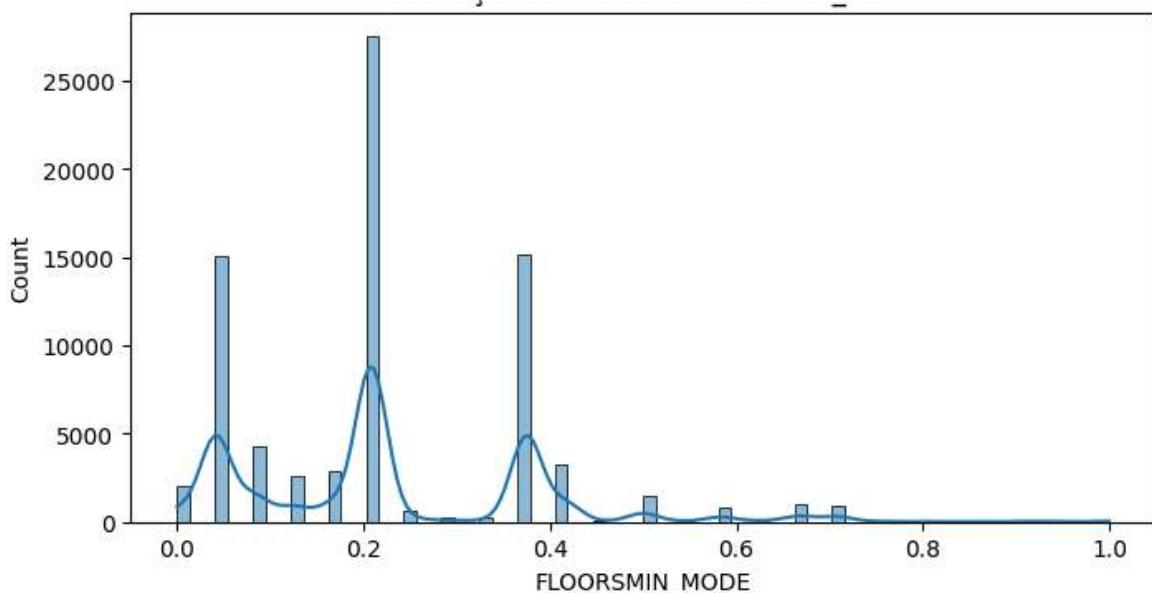
## Distribuição da coluna ENTRANCES\_MODE



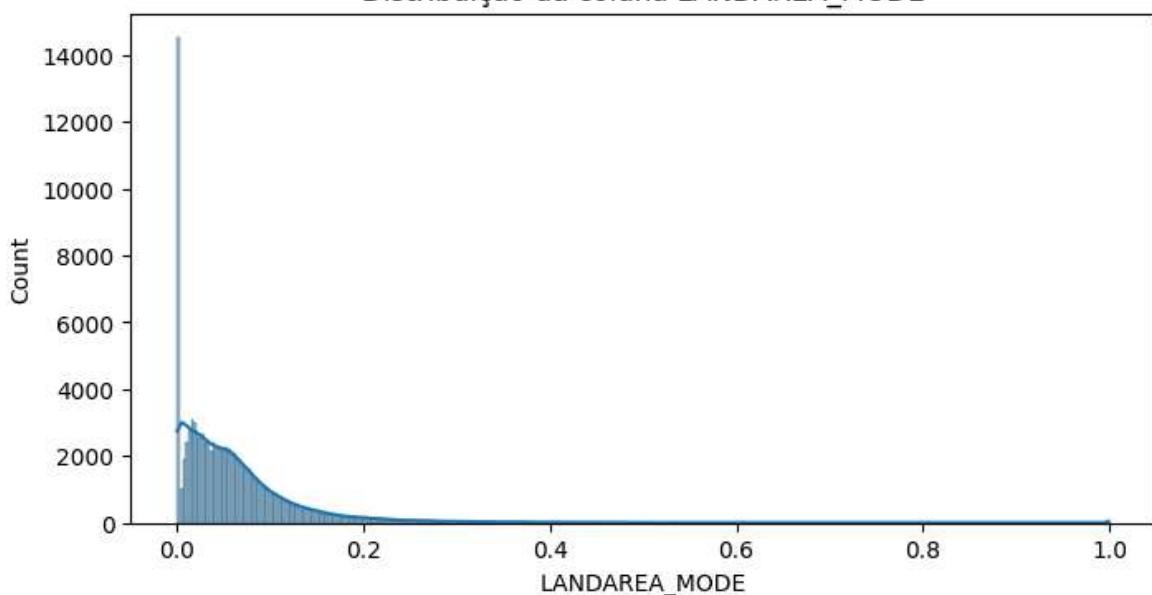
Distribuição da coluna FLOORSMAX\_MODE



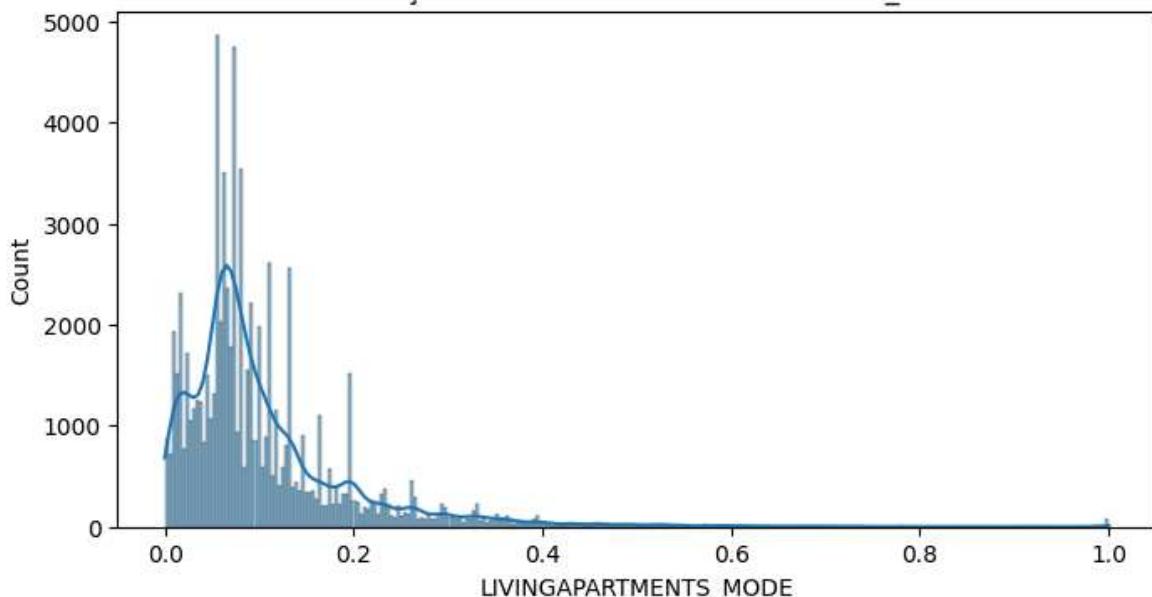
Distribuição da coluna FLOORSMIN\_MODE



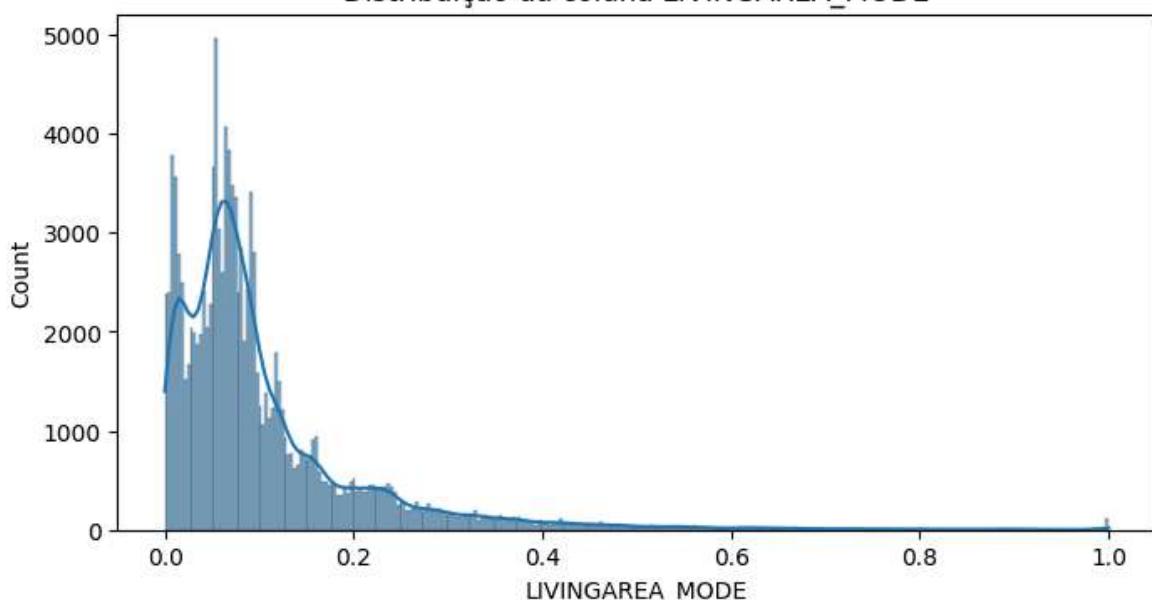
Distribuição da coluna LANDAREA\_MODE



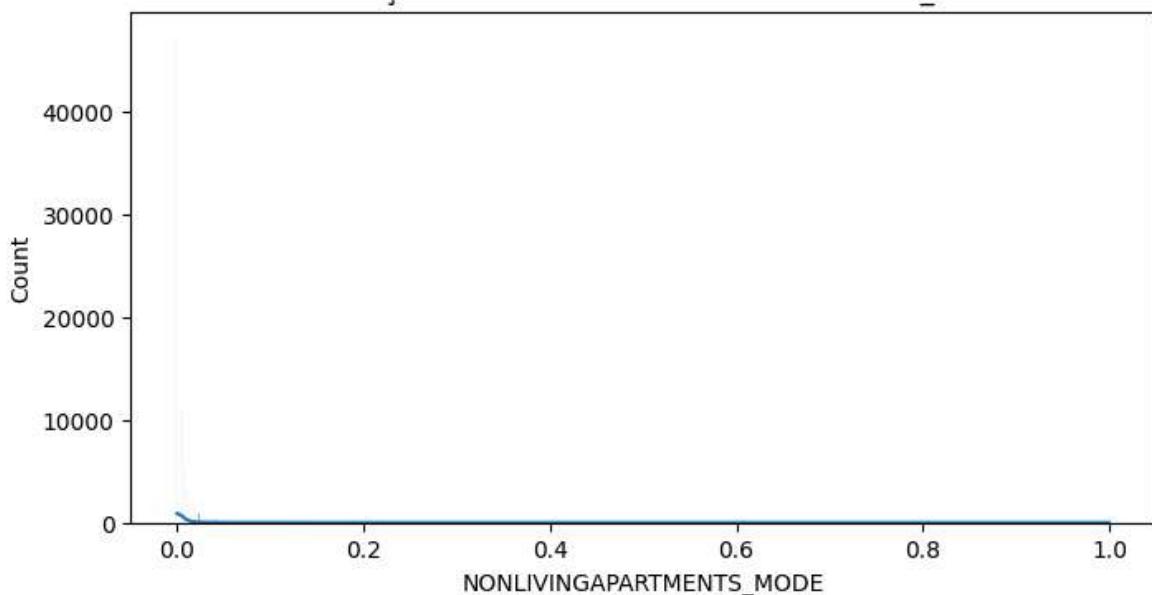
Distribuição da coluna LIVINGAPARTMENTS\_MODE



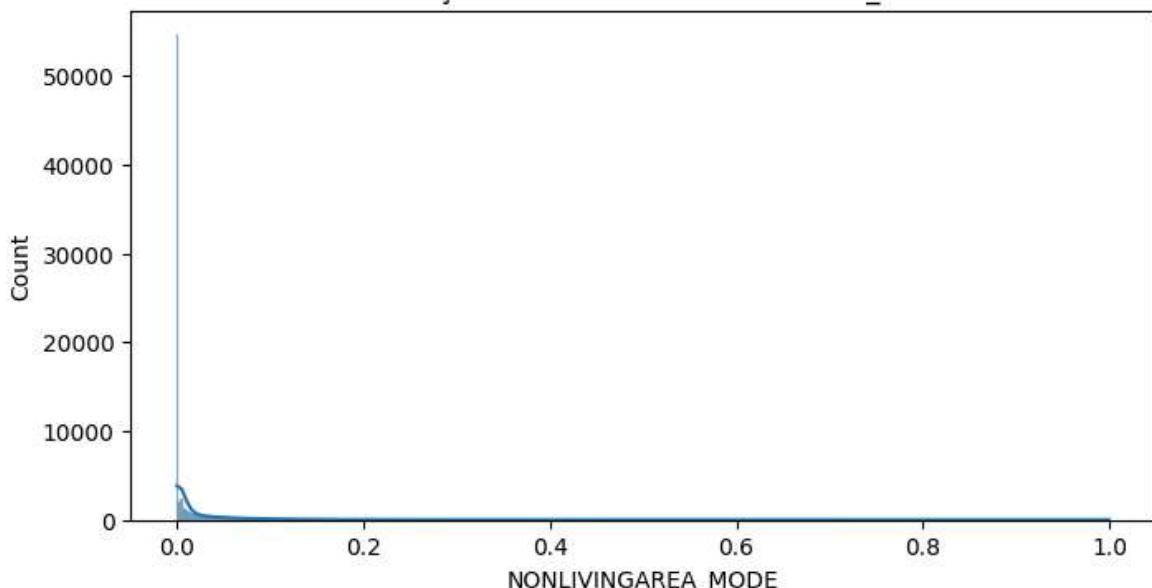
Distribuição da coluna LIVINGAREA\_MODE



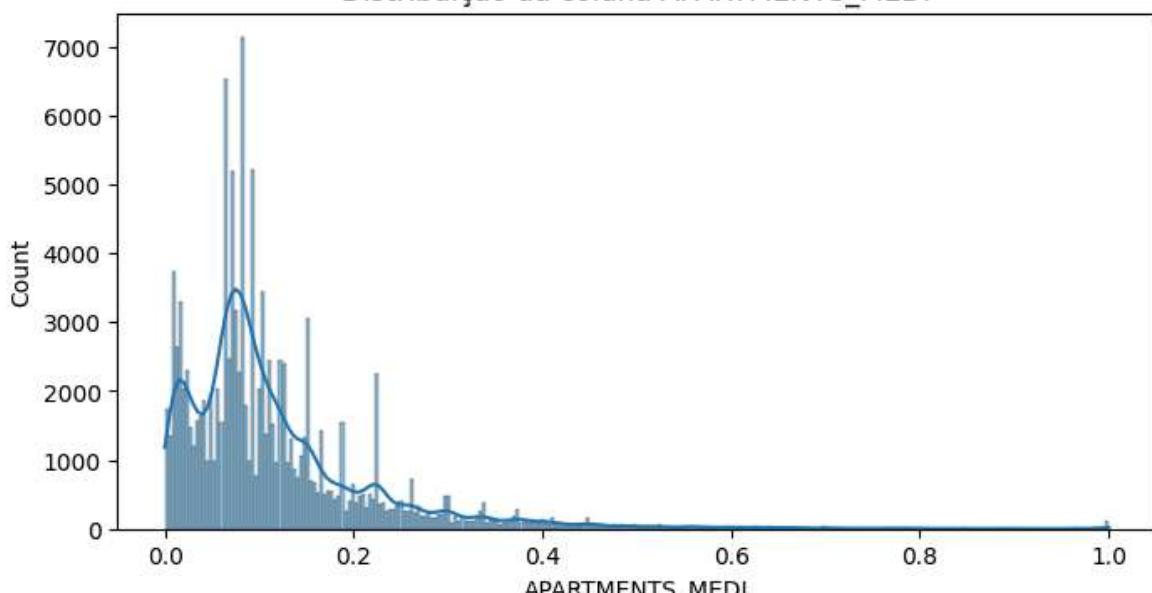
Distribuição da coluna NONLIVINGAPARTMENTS\_MODE



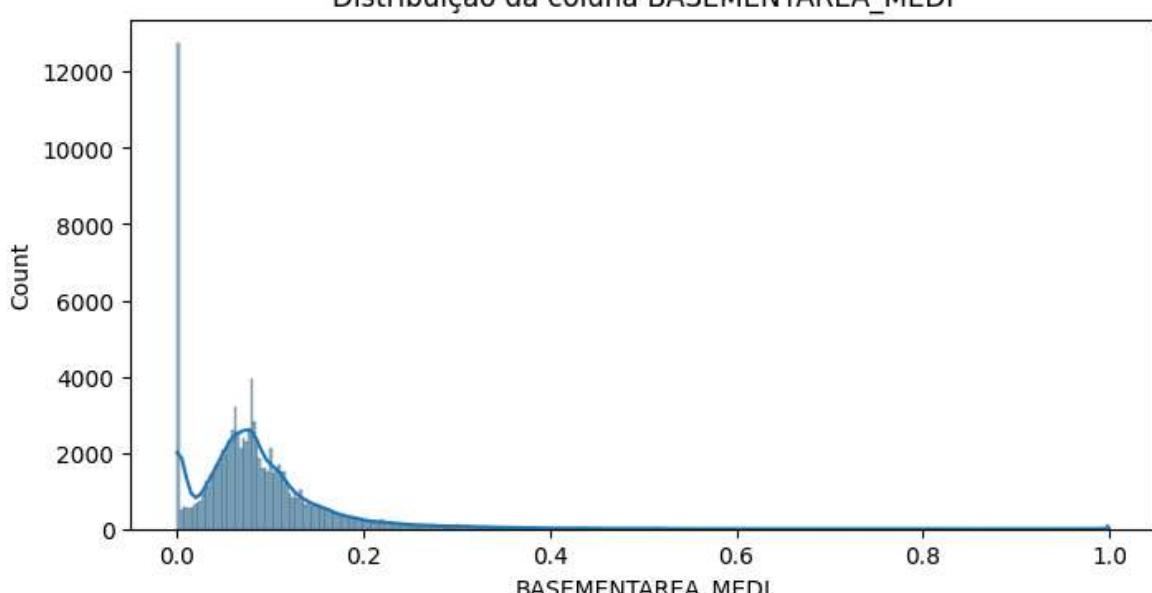
Distribuição da coluna NONLIVINGAREA\_MODE



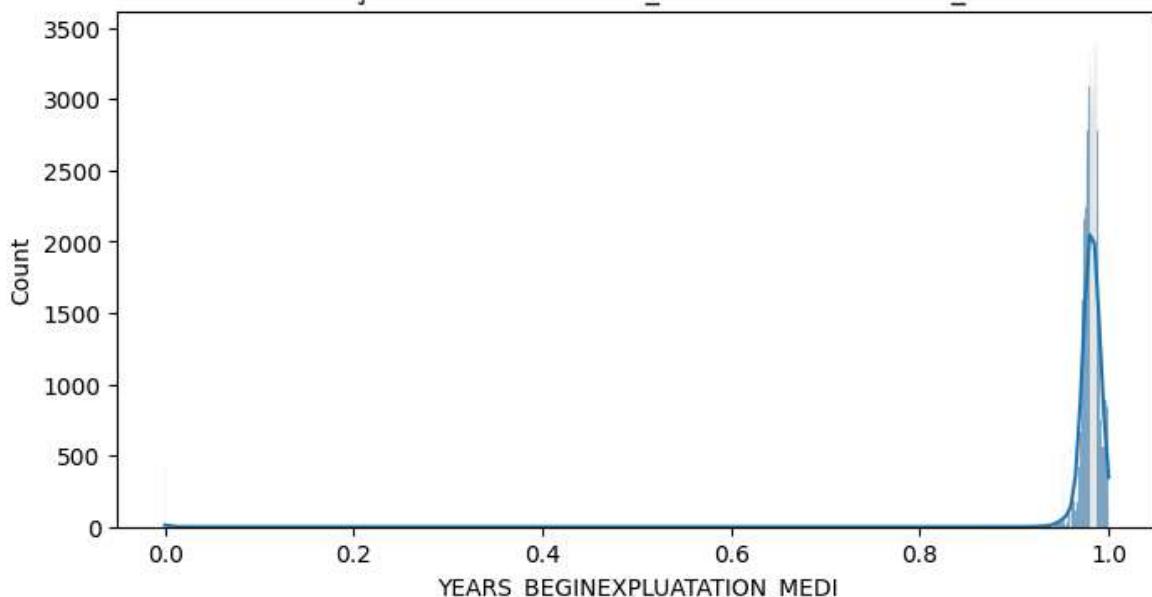
Distribuição da coluna APARTMENTS\_MEDI



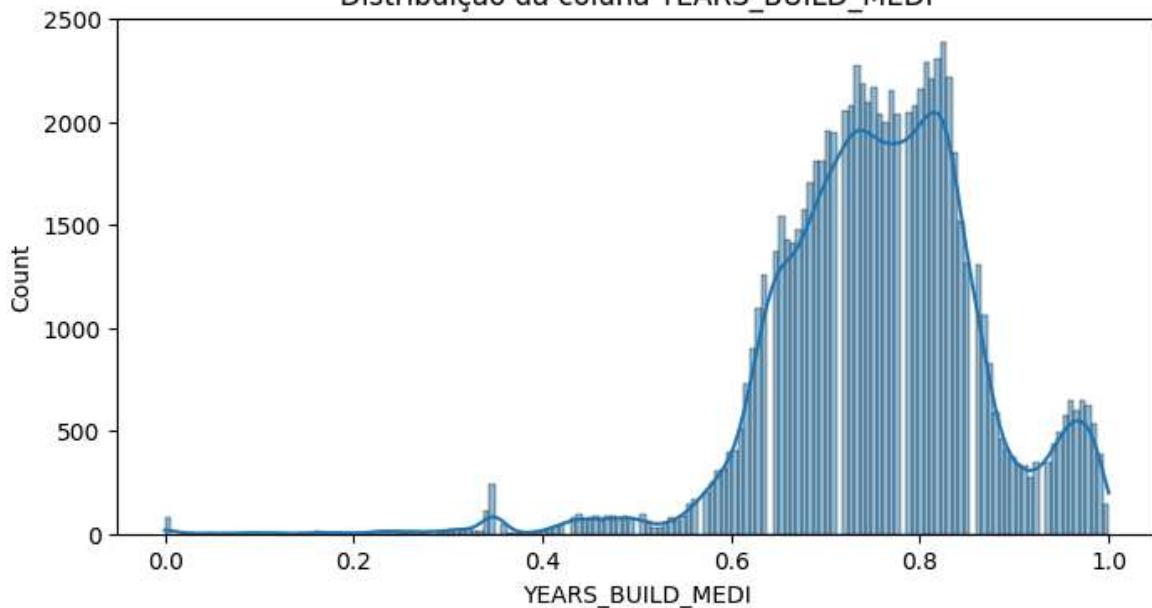
Distribuição da coluna BASEMENTAREA\_MEDI



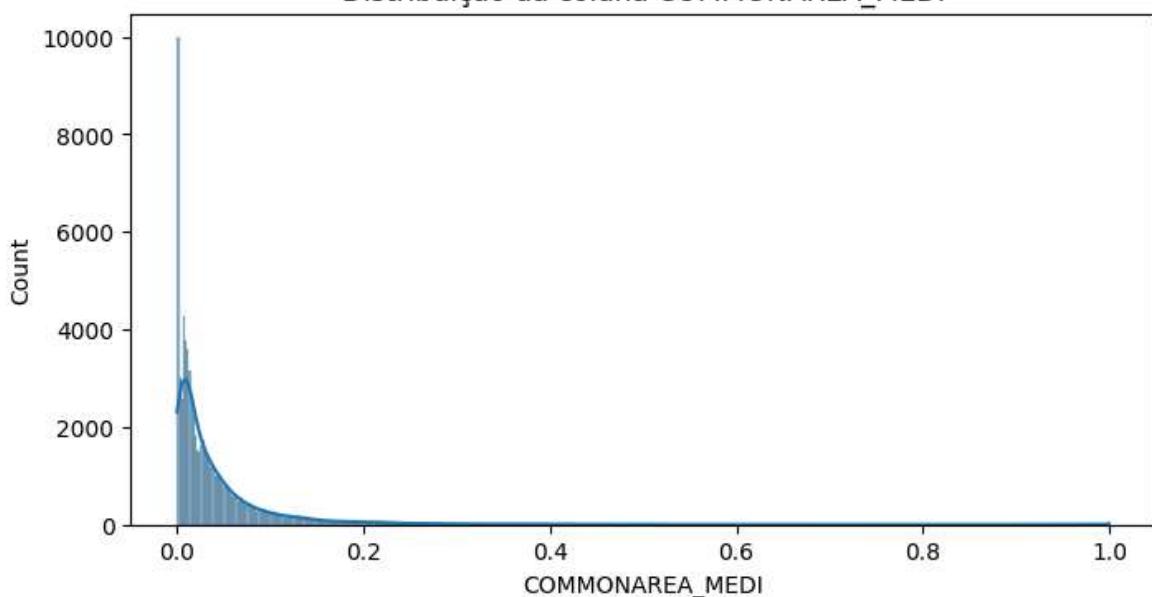
Distribuição da coluna YEARS\_BEGINEXPLUATATION\_MEDI



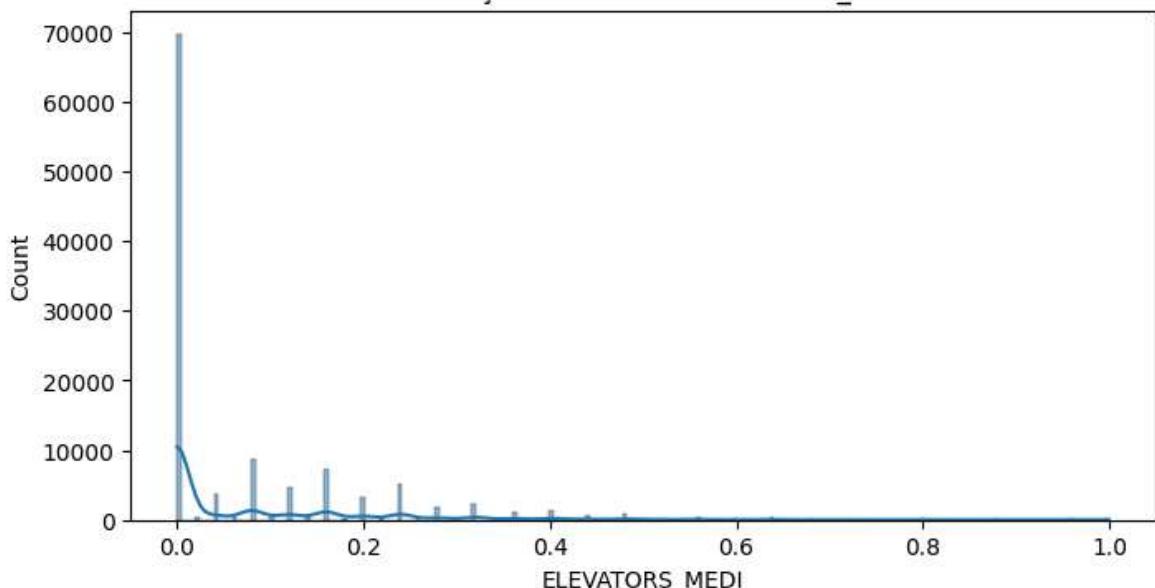
Distribuição da coluna YEARS\_BUILD\_MEDI



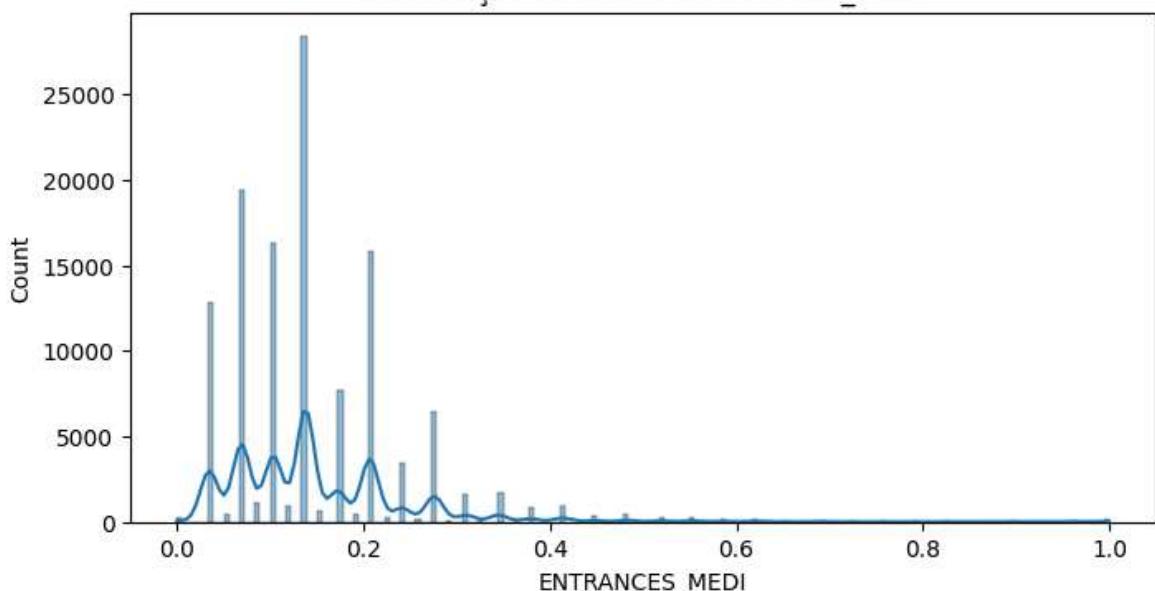
Distribuição da coluna COMMONAREA\_MEDI



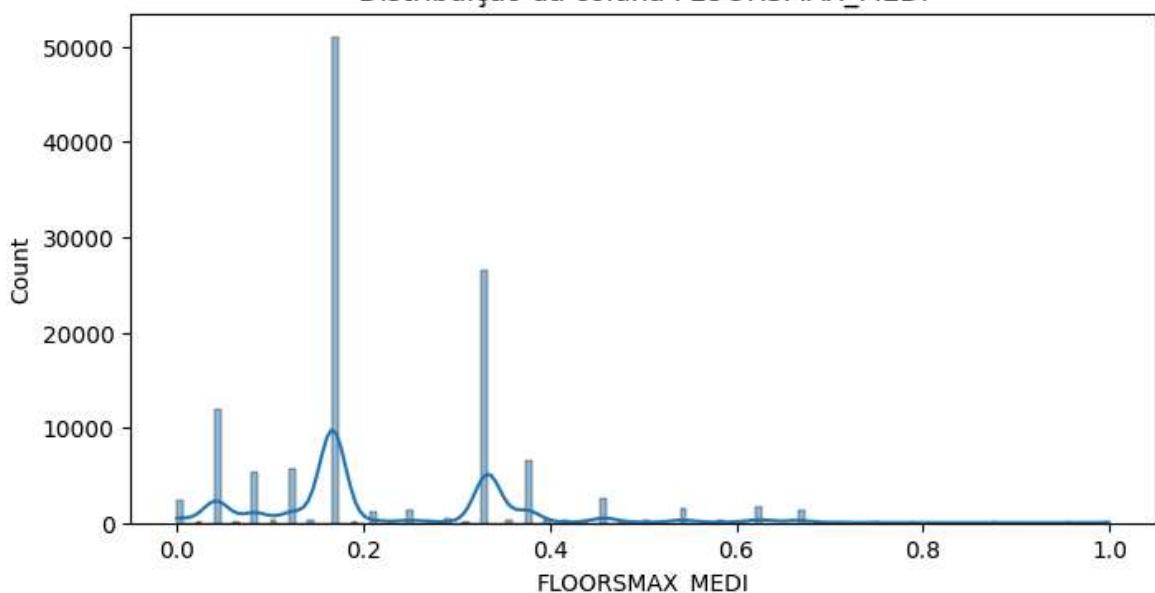
Distribuição da coluna ELEVATORS\_MEDI



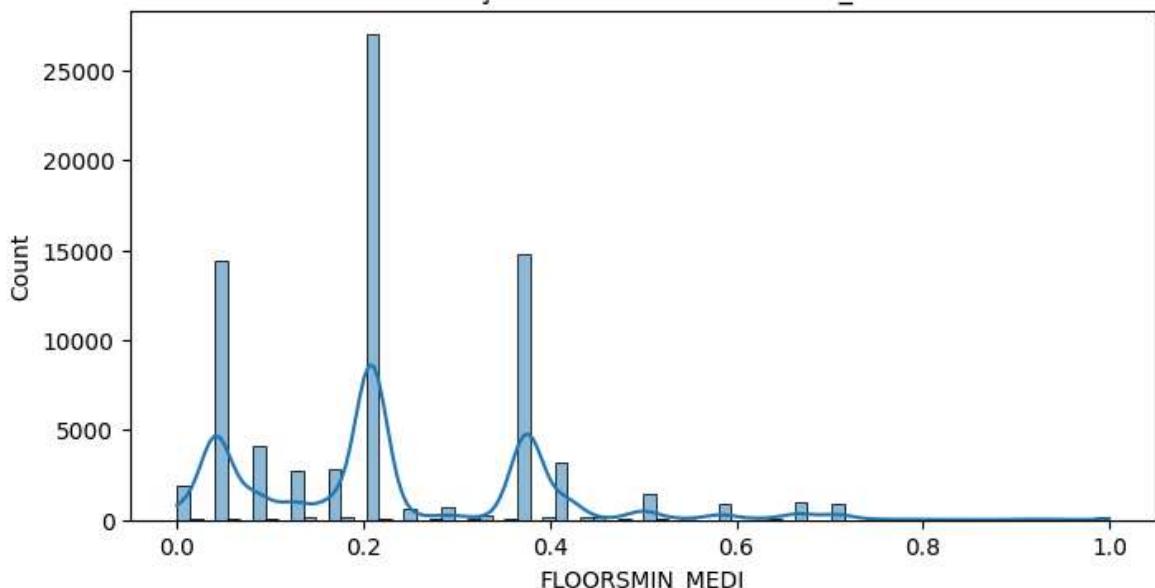
Distribuição da coluna ENTRANCES\_MEDI



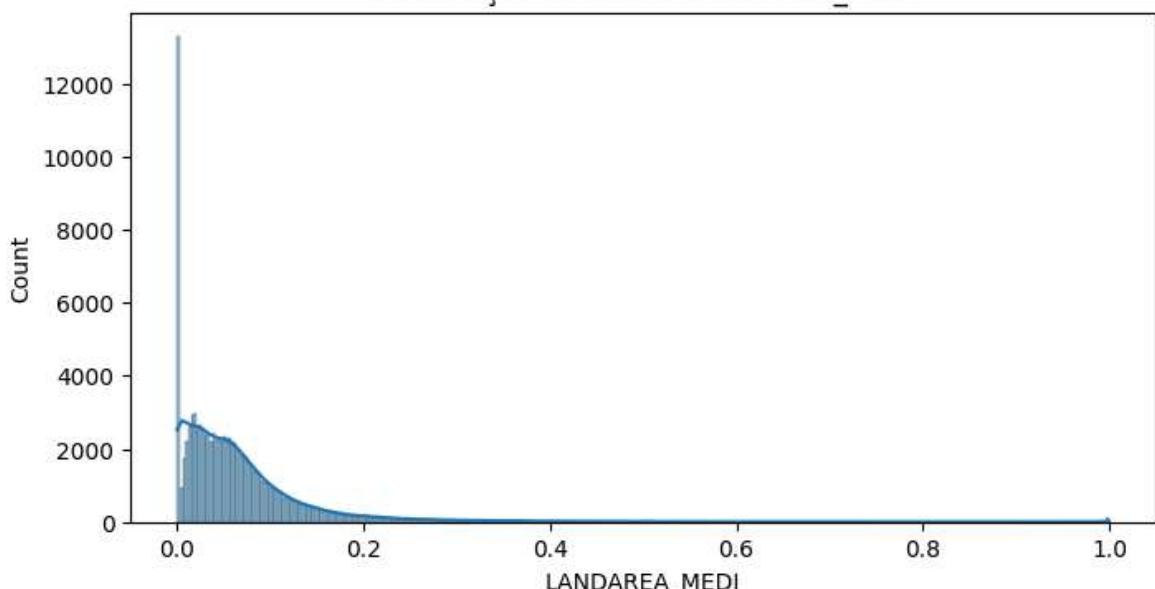
Distribuição da coluna FLOORSMAX\_MEDI



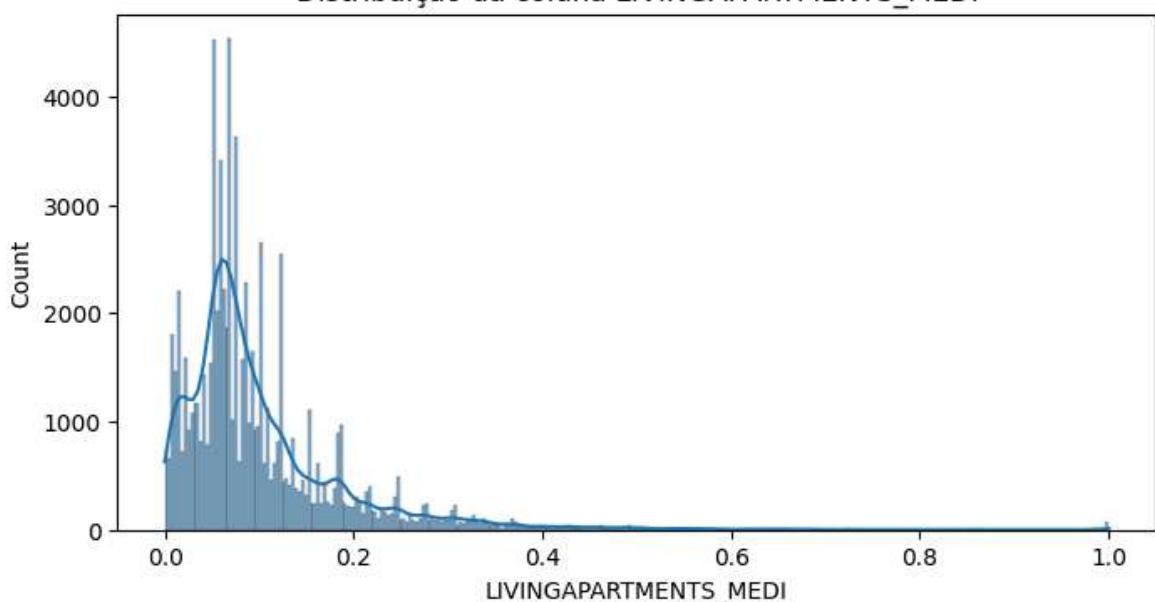
## Distribuição da coluna FLOORSMIN\_MEDI



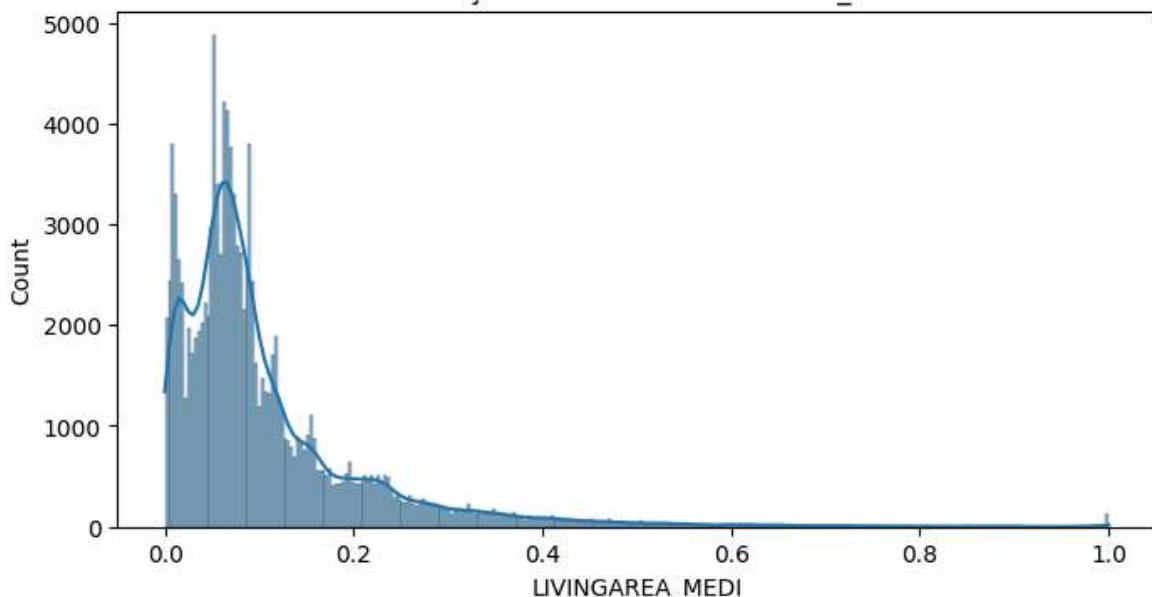
## Distribuição da coluna LANDAREA\_MEDI



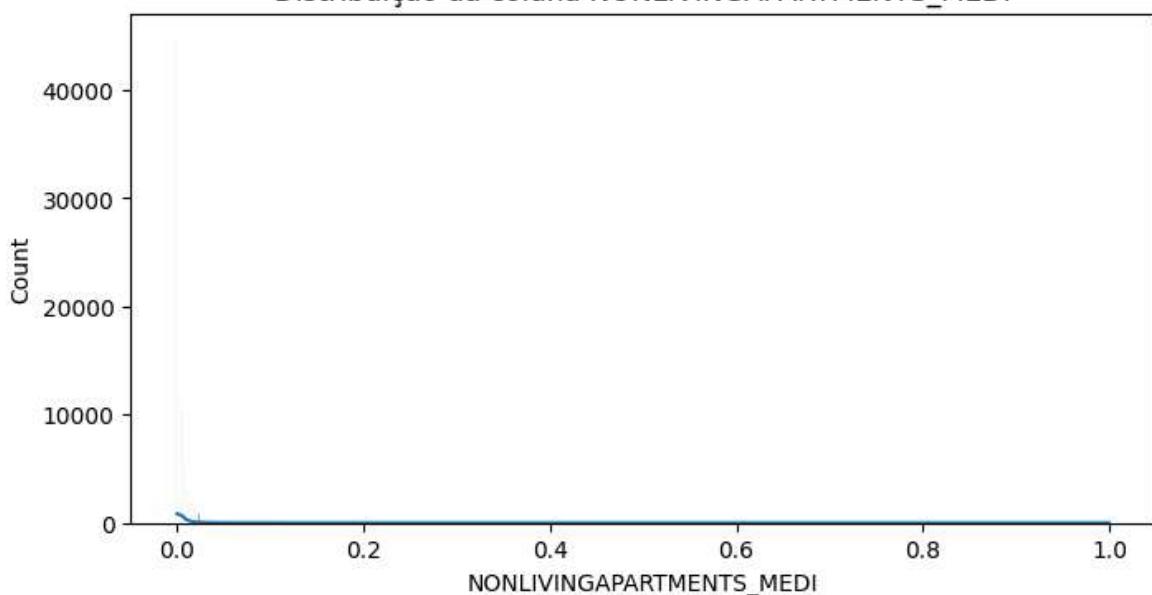
## Distribuição da coluna LIVINGAPARTMENTS\_MEDI



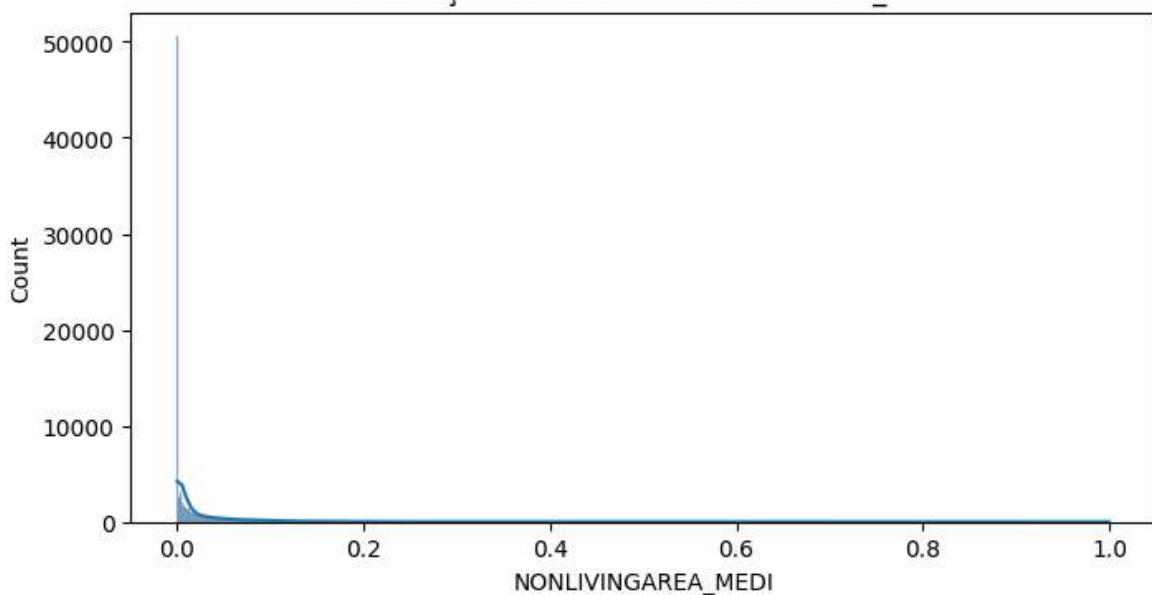
Distribuição da coluna LIVINGAREA\_MEDI



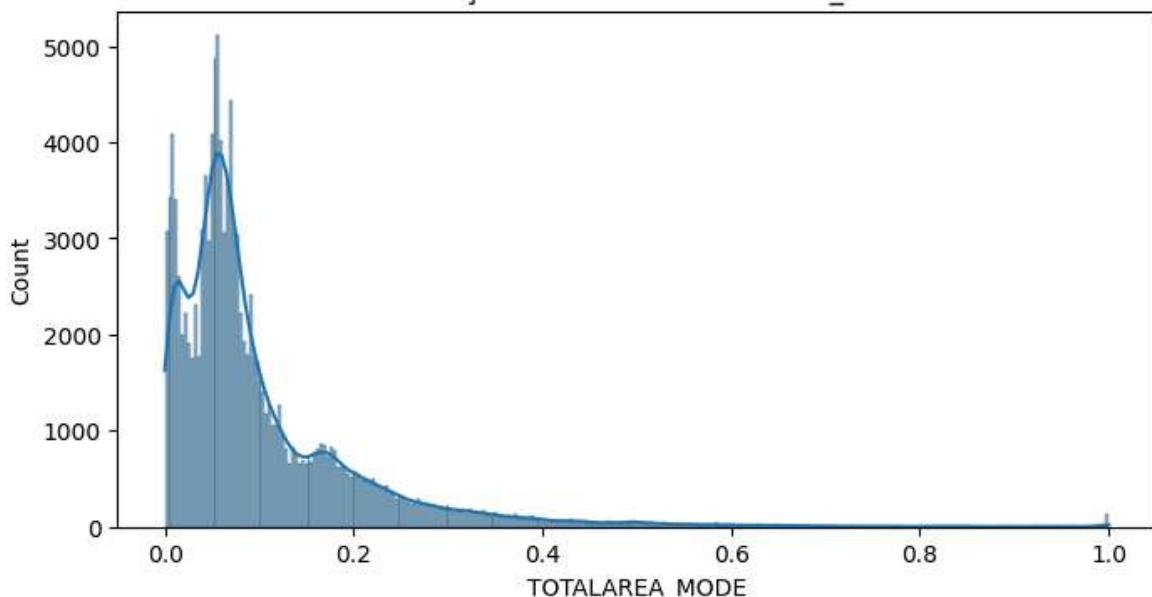
Distribuição da coluna NONLIVINGAPARTMENTS\_MEDI



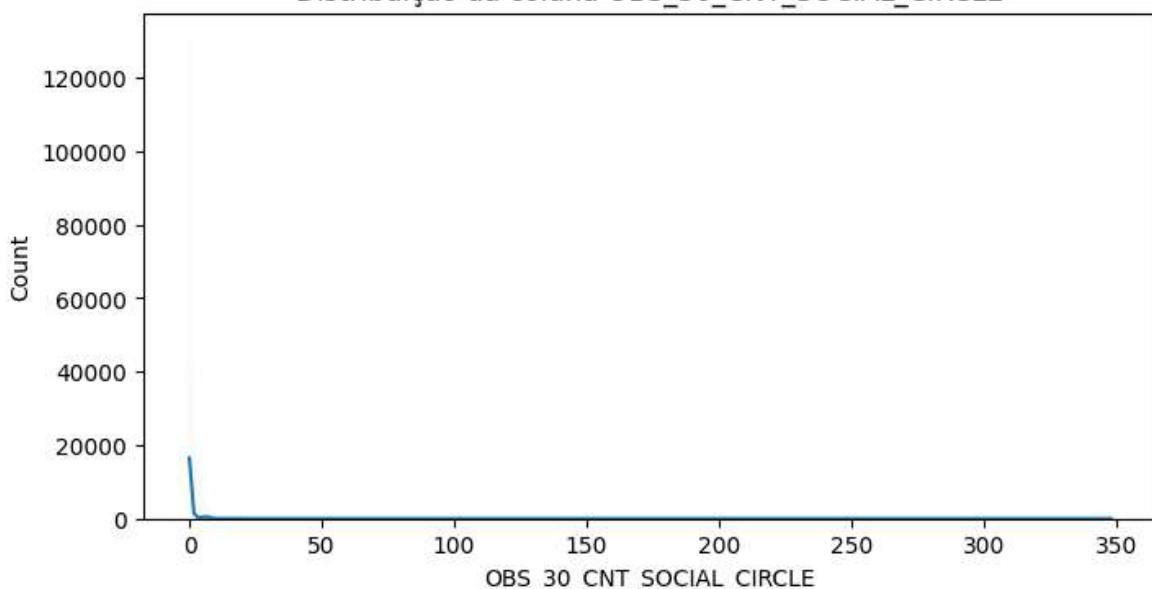
Distribuição da coluna NONLIVINGAREA\_MEDI



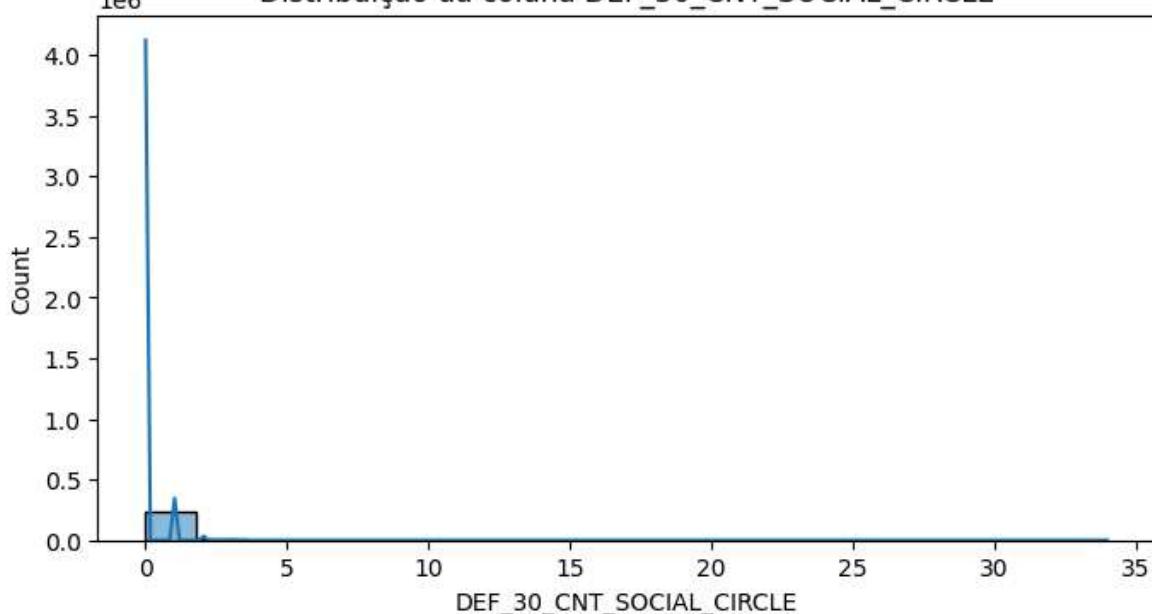
## Distribuição da coluna TOTALAREA\_MODE



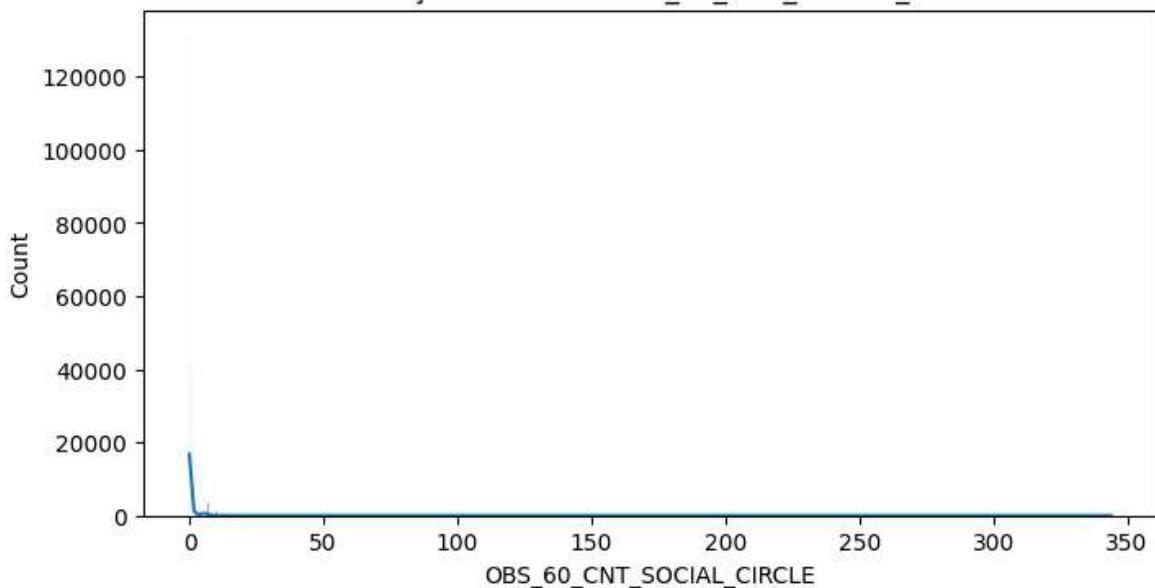
## Distribuição da coluna OBS\_30\_CNT\_SOCIAL\_CIRCLE



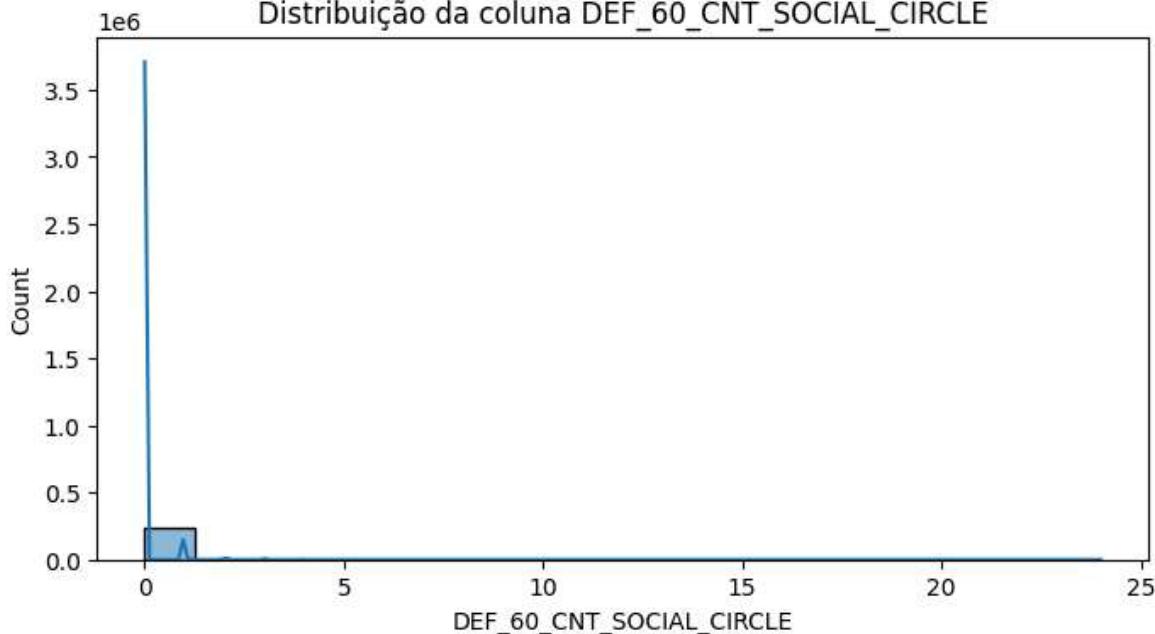
## Distribuição da coluna DEF\_30\_CNT\_SOCIAL\_CIRCLE



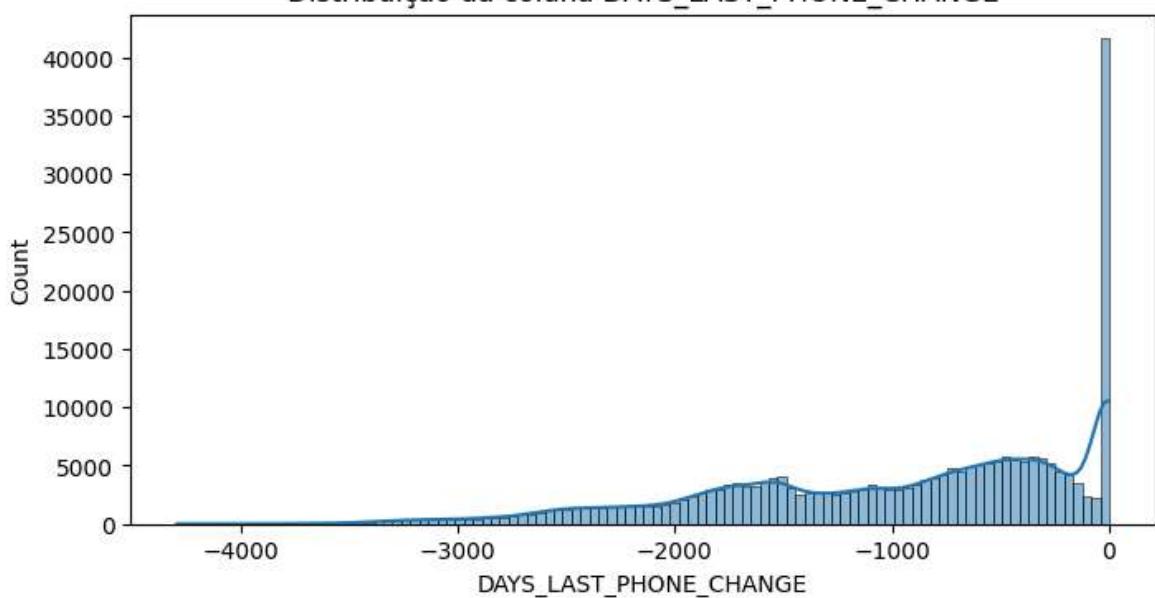
Distribuição da coluna OBS\_60\_CNT\_SOCIAL\_CIRCLE

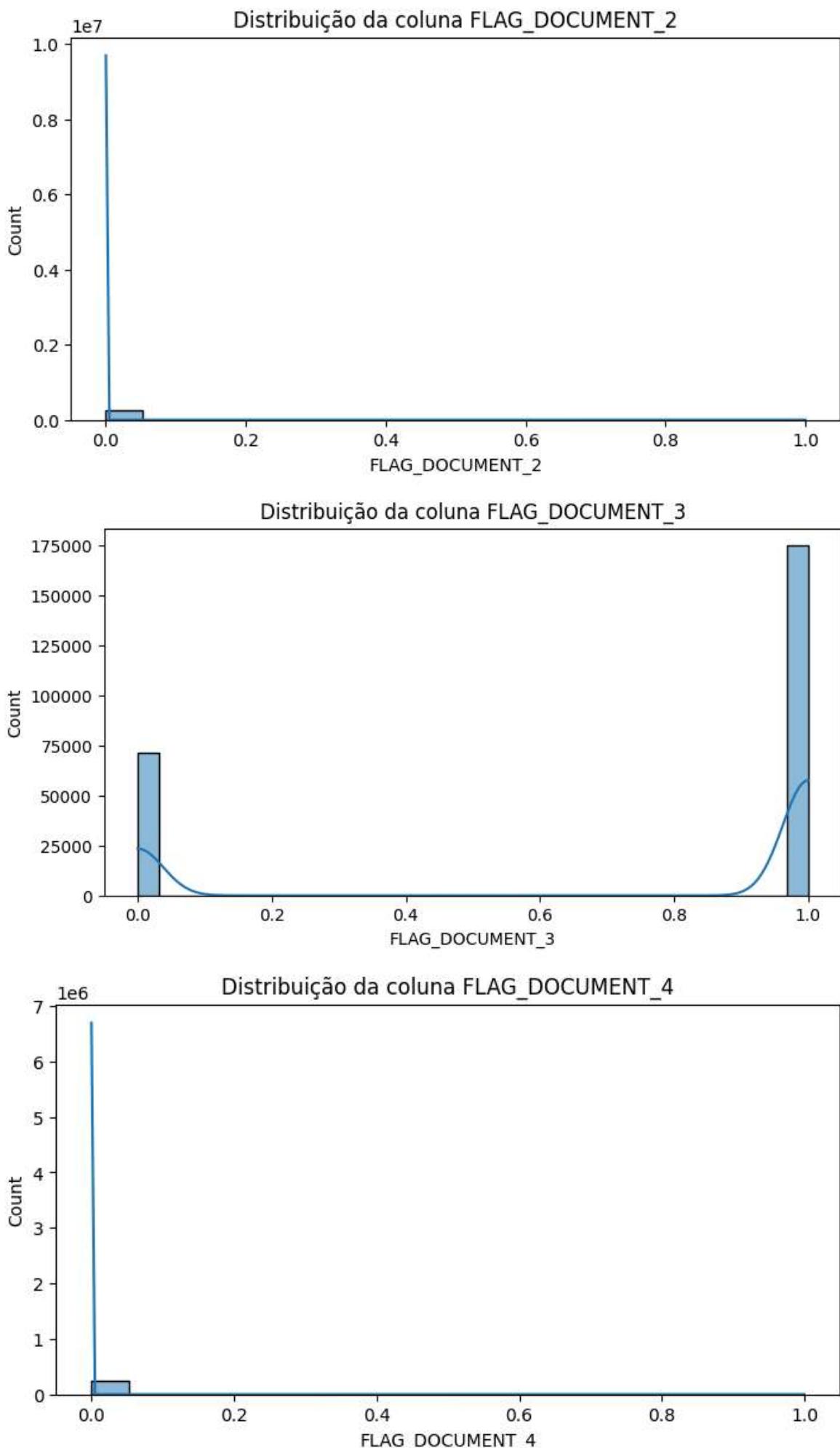


Distribuição da coluna DEF\_60\_CNT\_SOCIAL\_CIRCLE

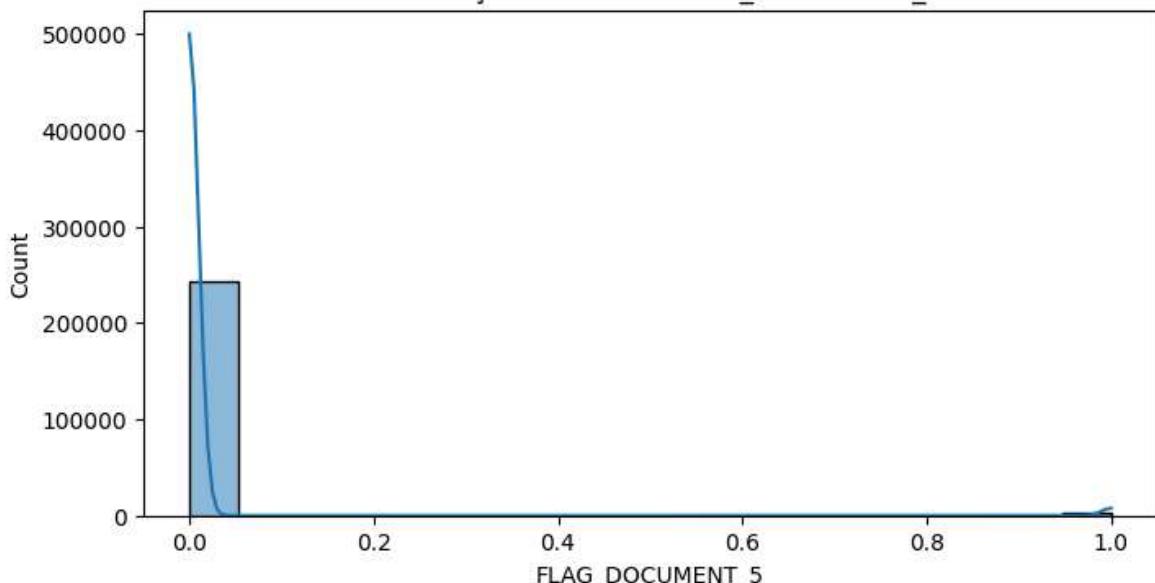


Distribuição da coluna DAYS\_LAST\_PHONE\_CHANGE

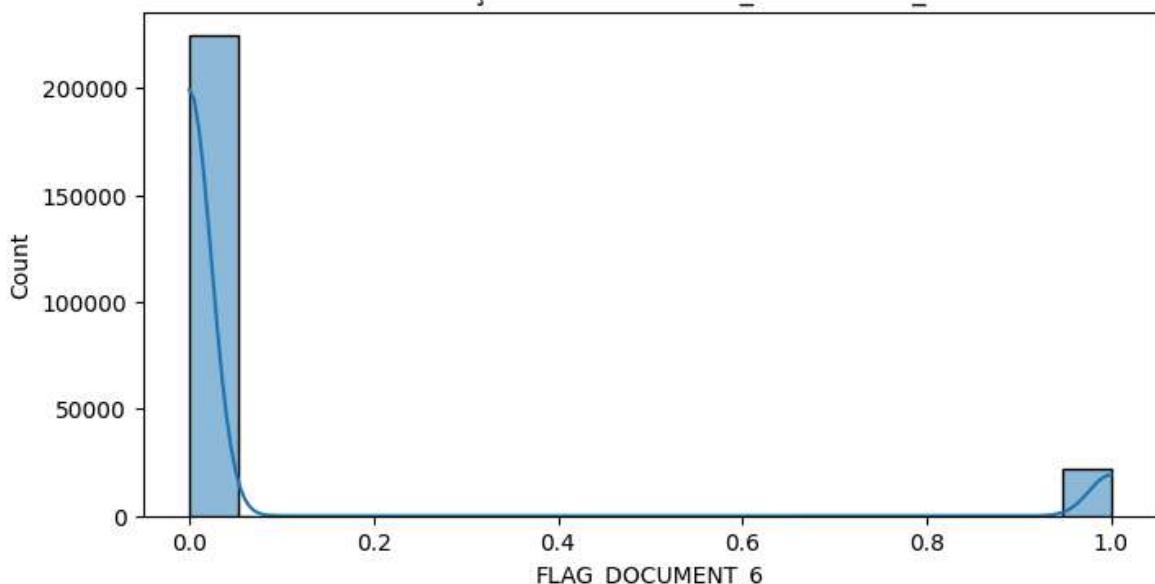




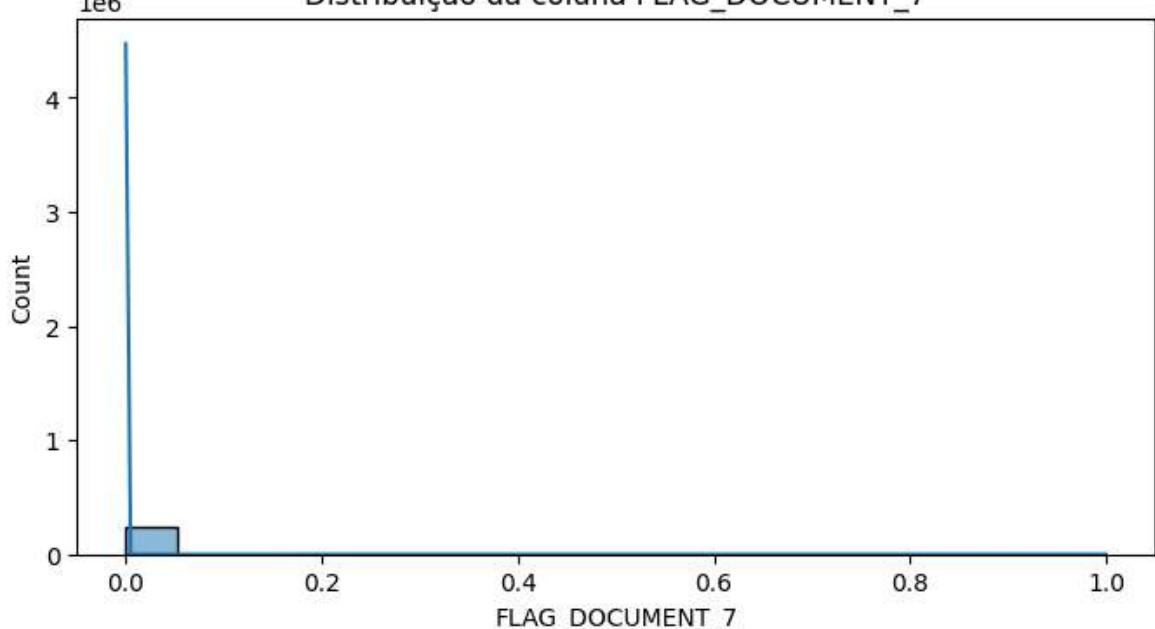
Distribuição da coluna FLAG\_DOCUMENT\_5



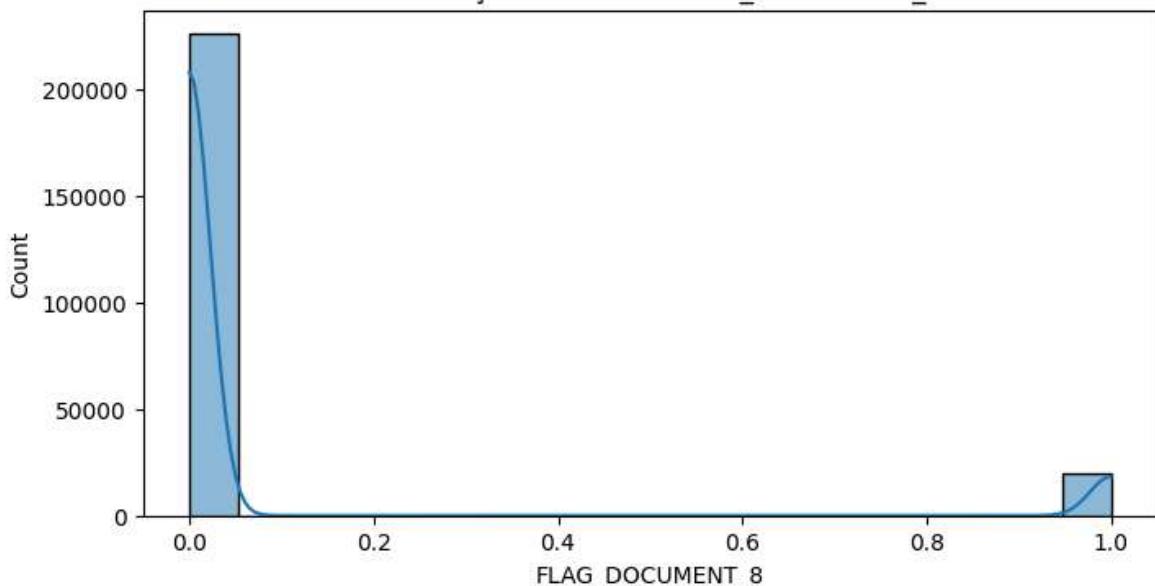
Distribuição da coluna FLAG\_DOCUMENT\_6



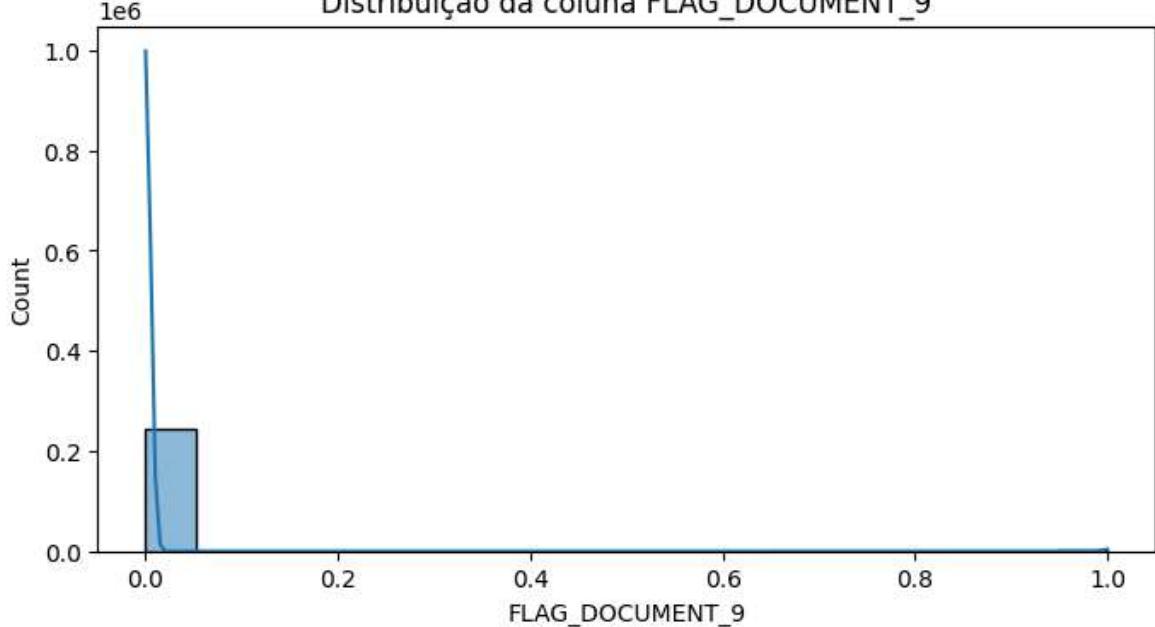
Distribuição da coluna FLAG\_DOCUMENT\_7



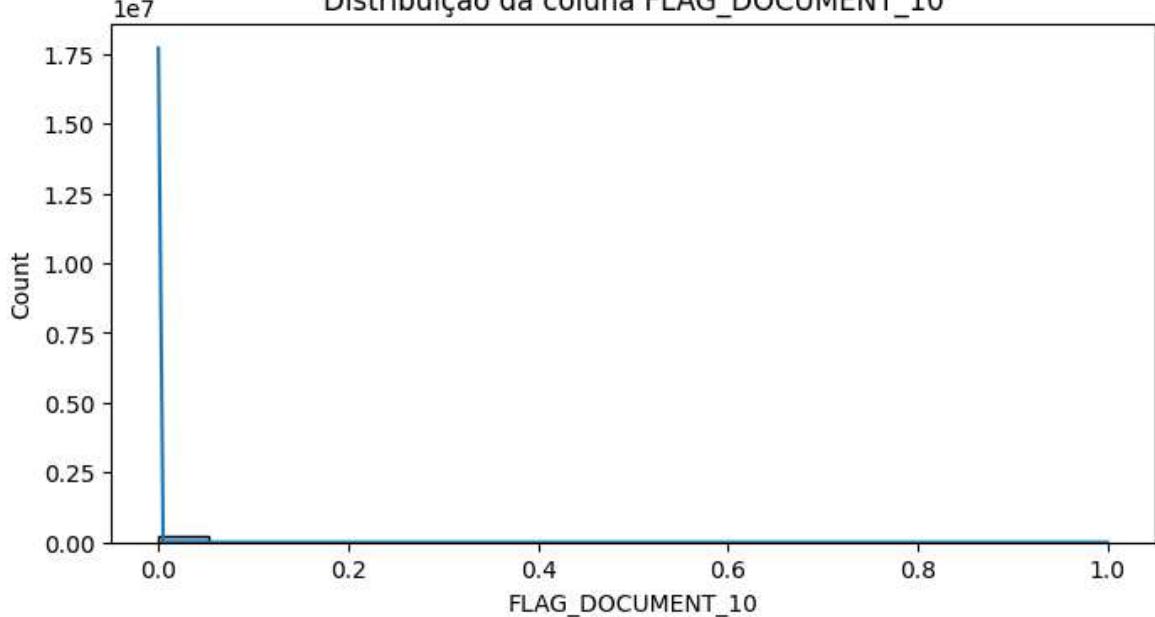
Distribuição da coluna FLAG\_DOCUMENT\_8

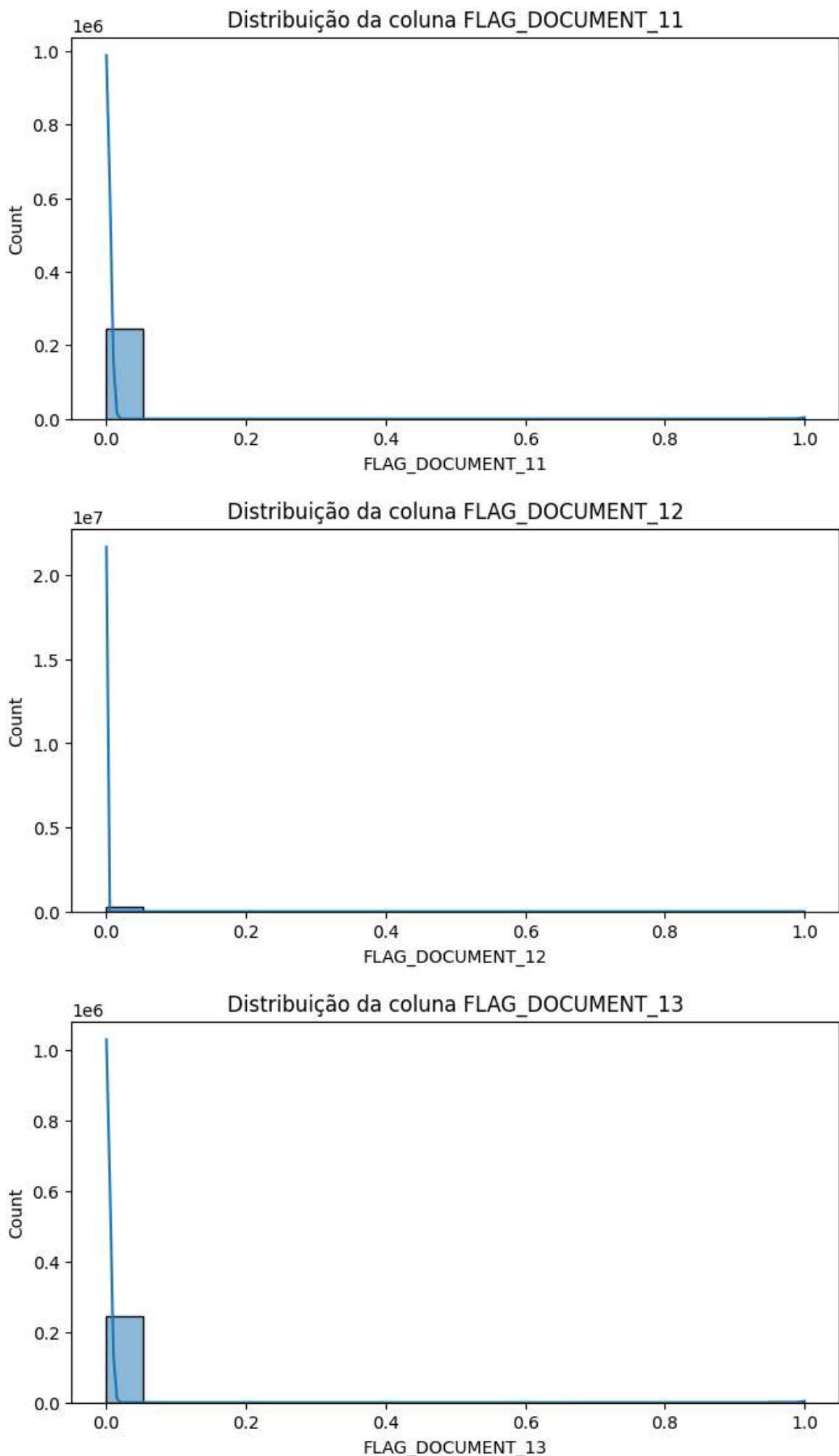


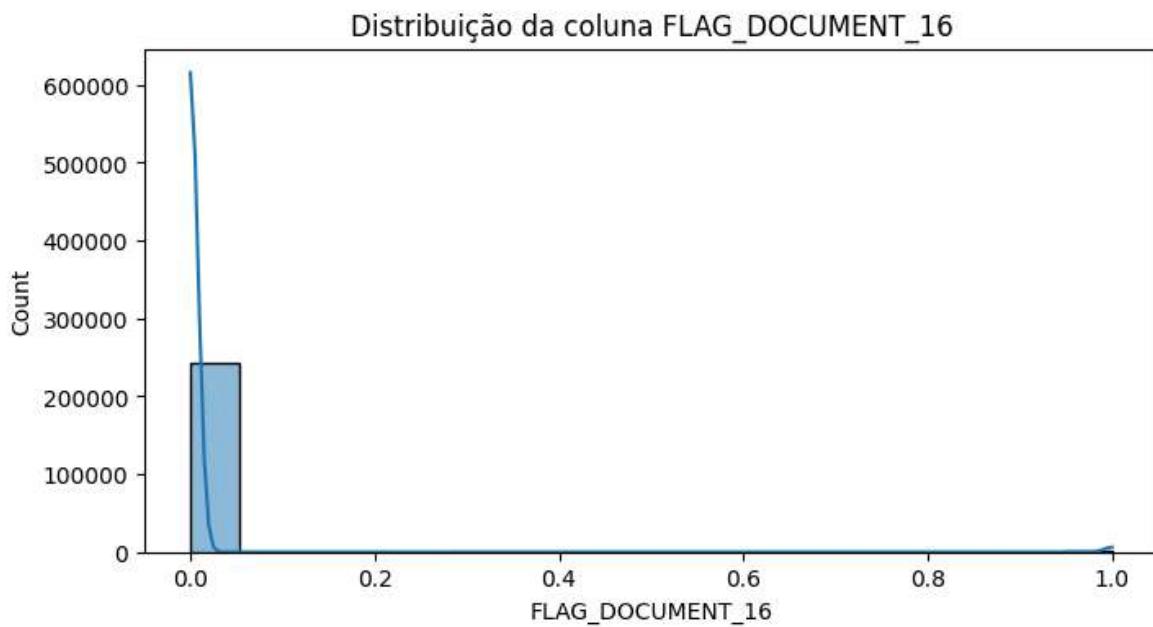
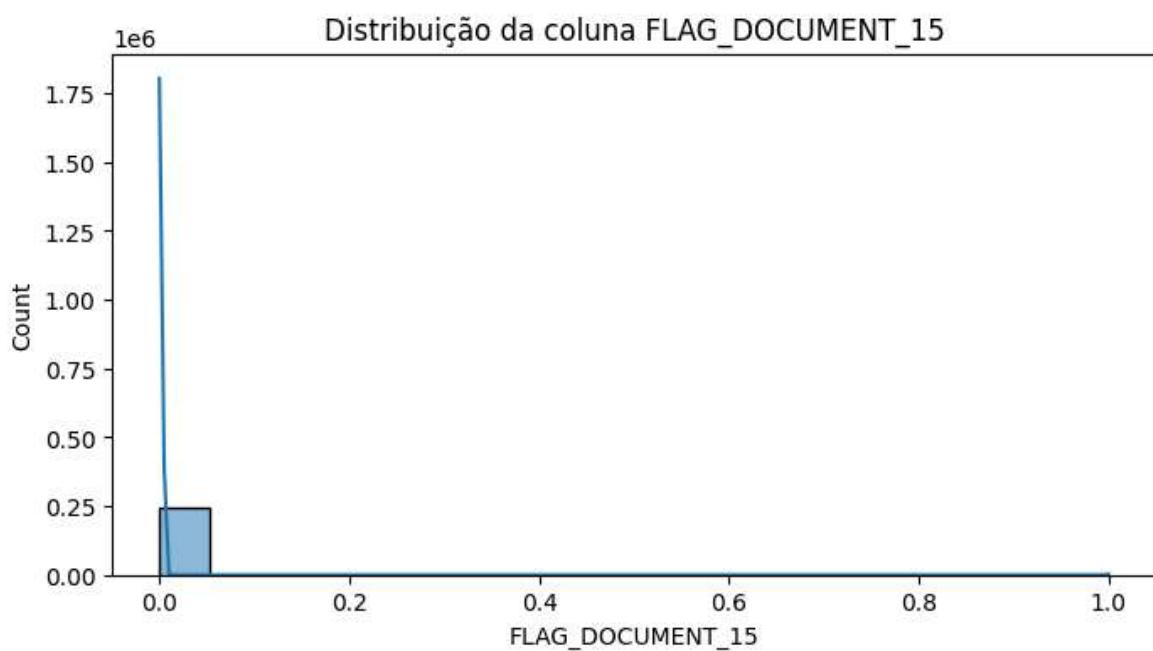
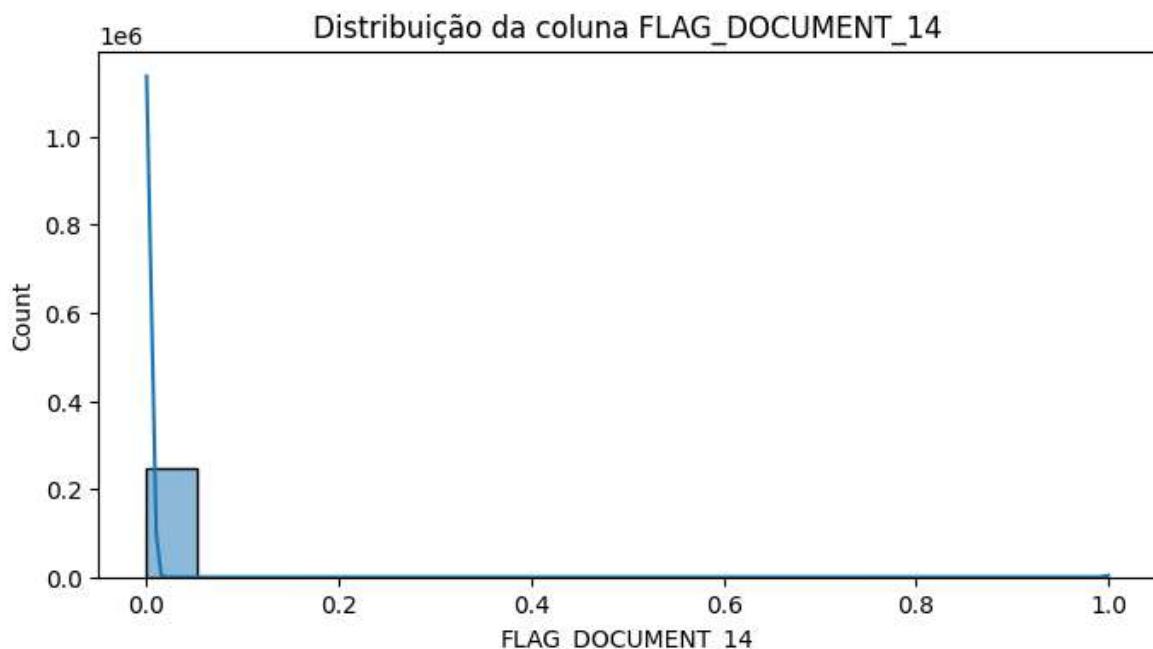
Distribuição da coluna FLAG\_DOCUMENT\_9

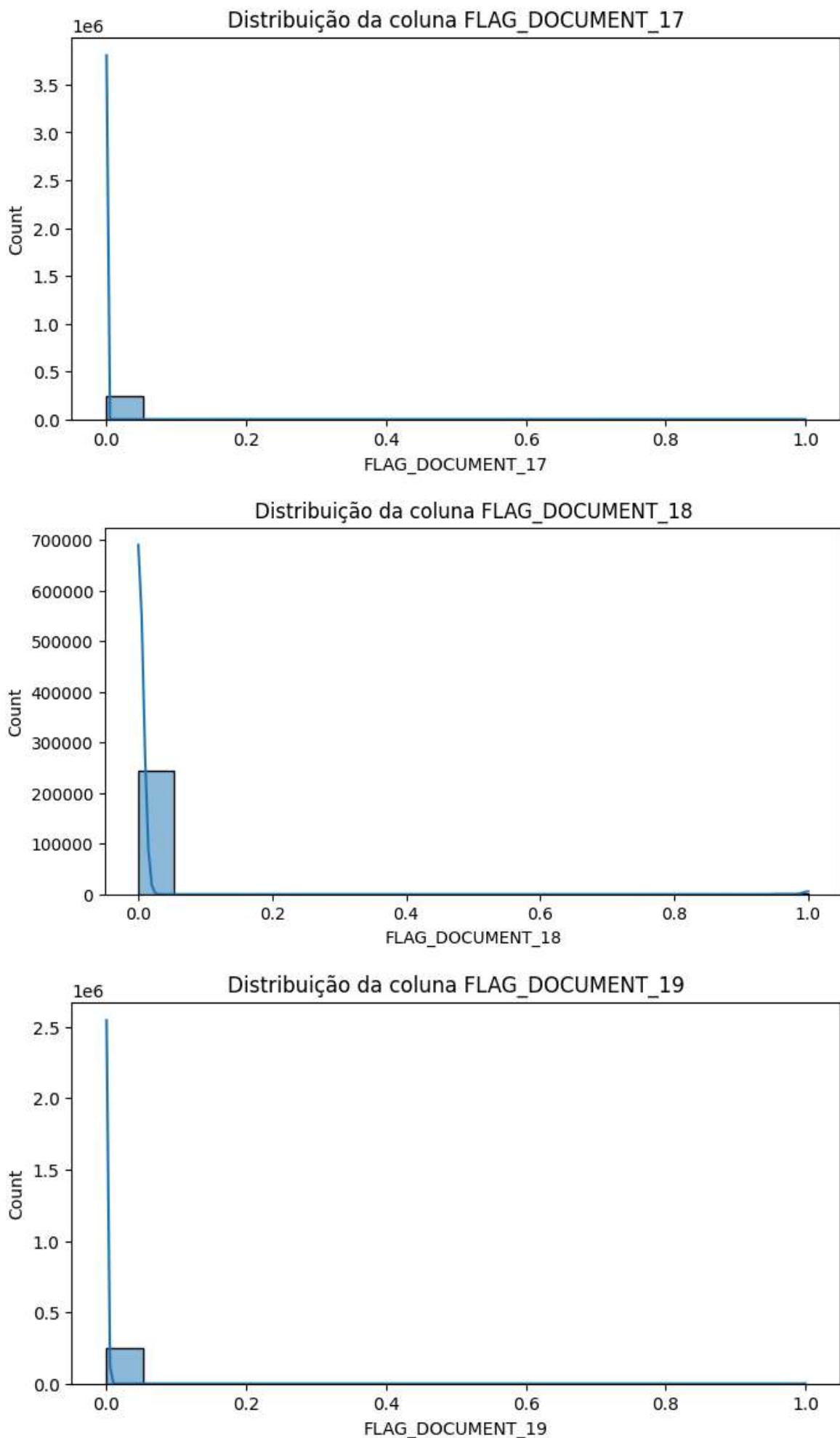


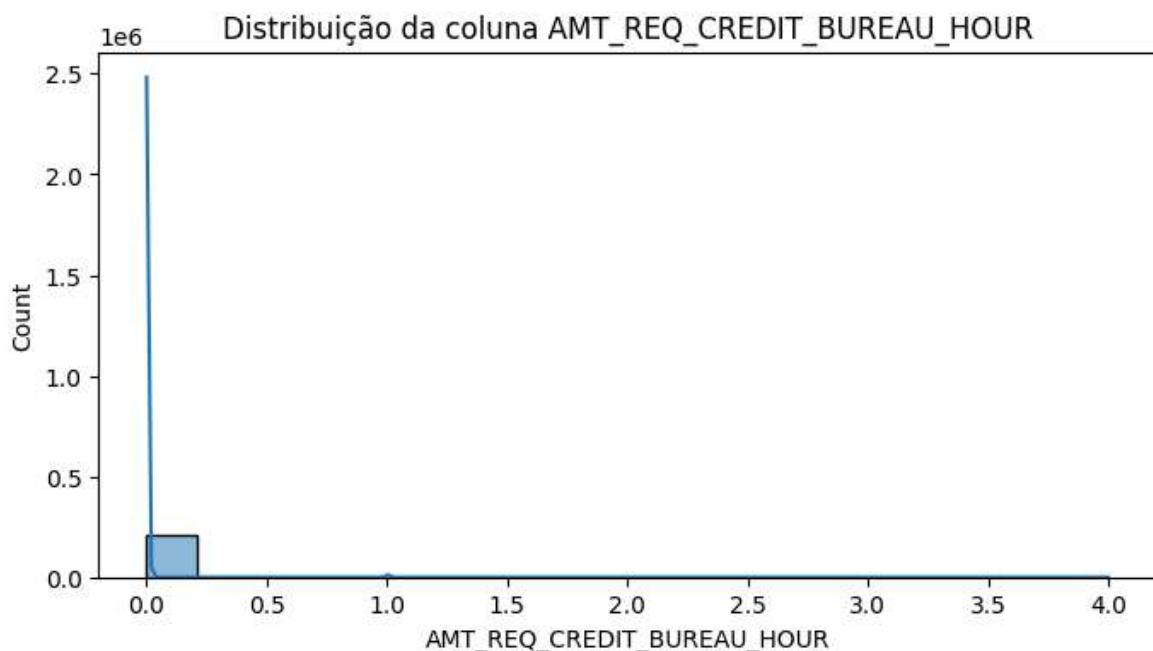
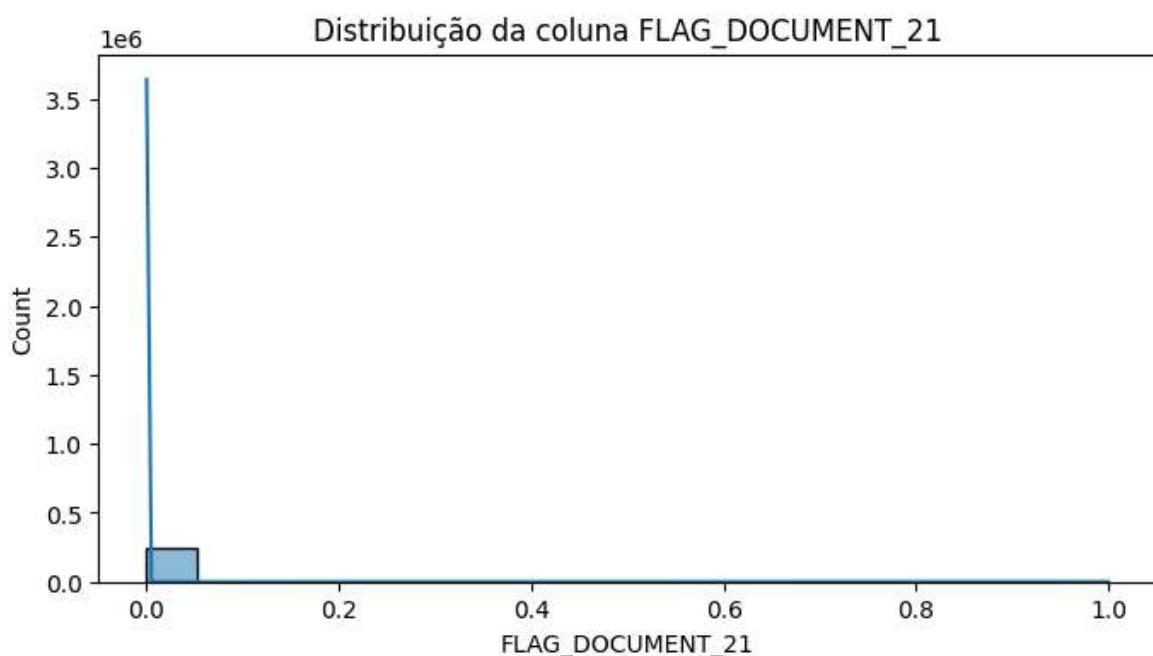
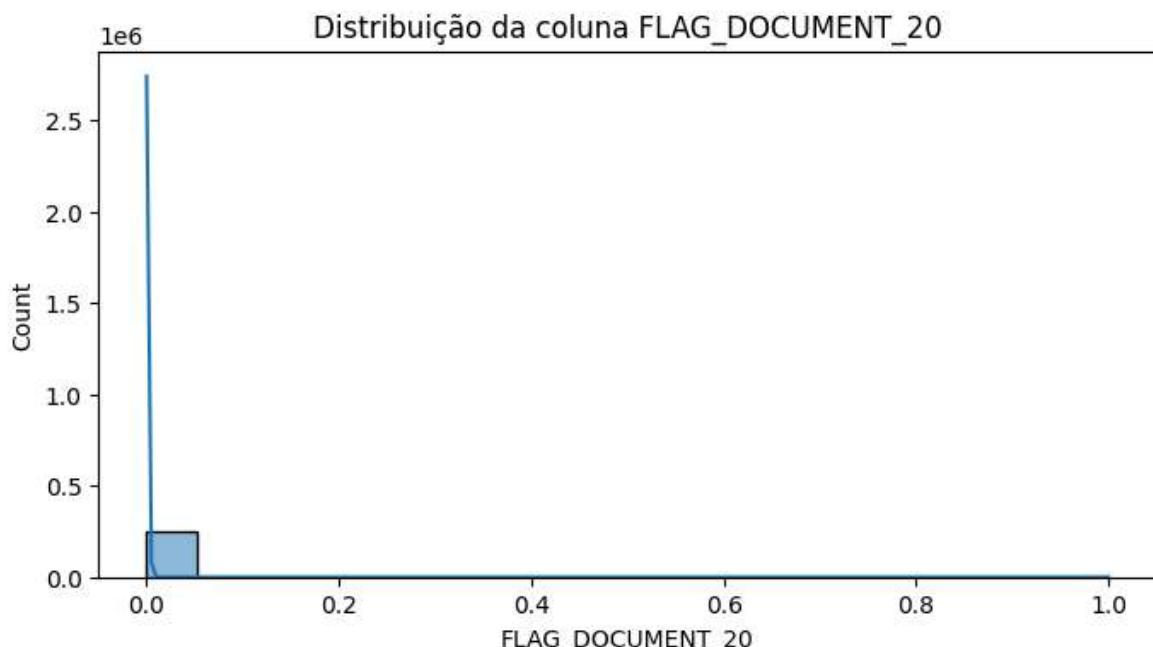
Distribuição da coluna FLAG\_DOCUMENT\_10

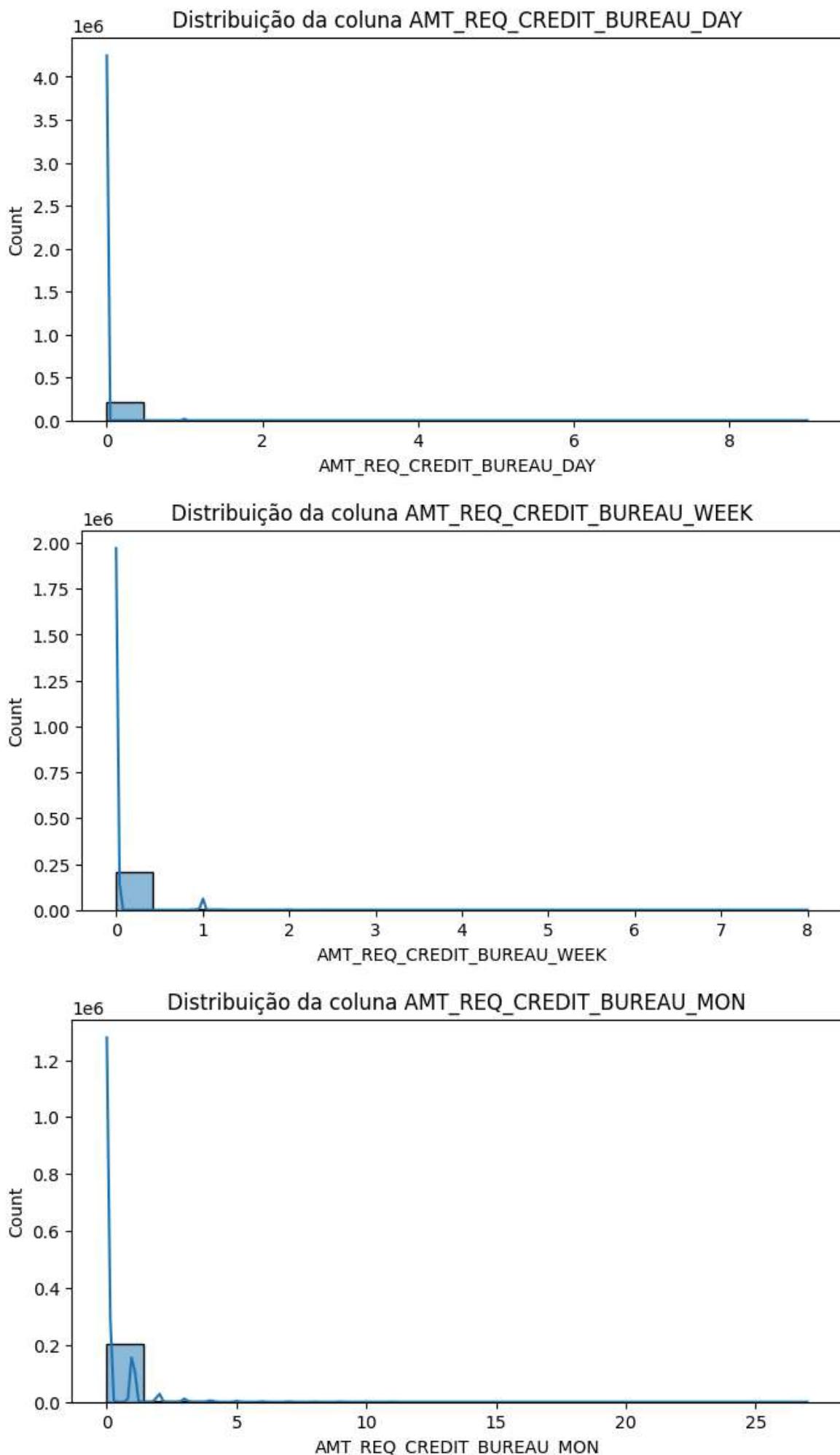


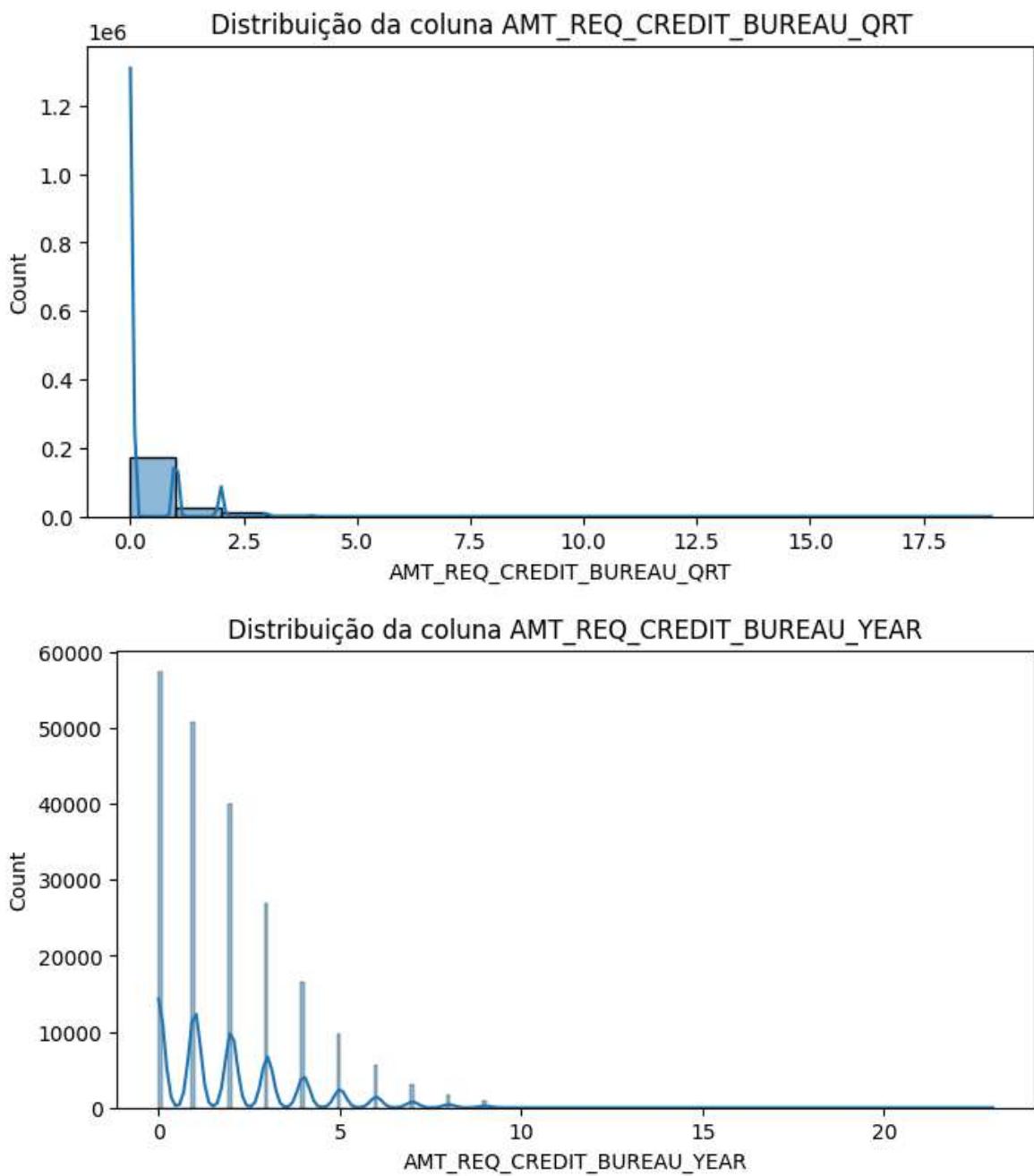








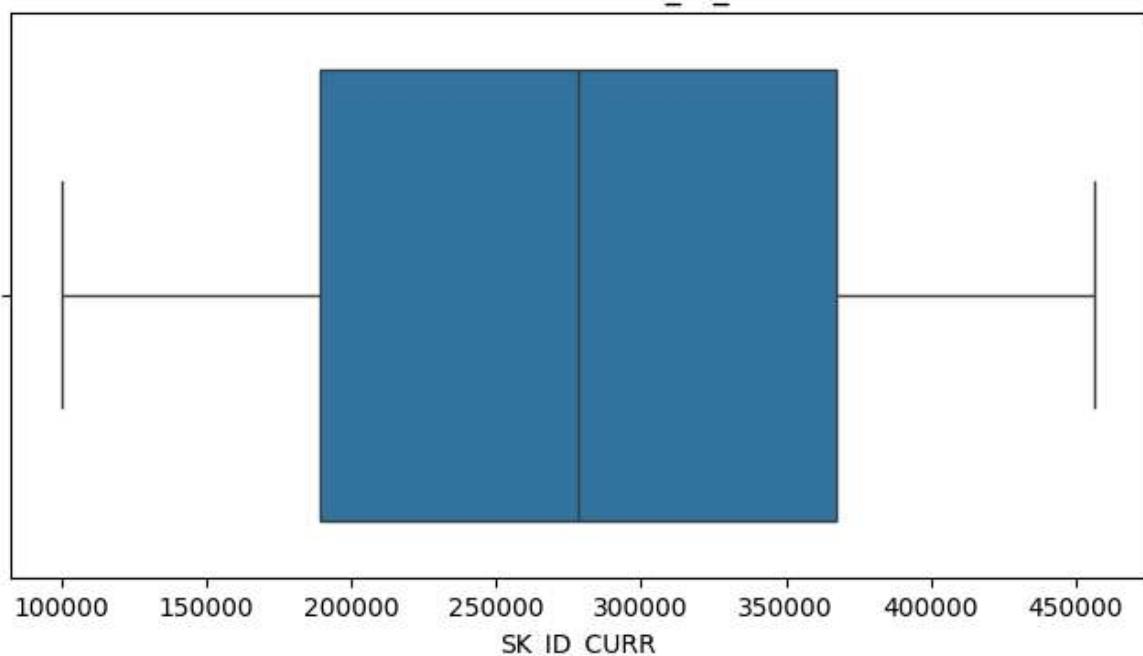




```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
```

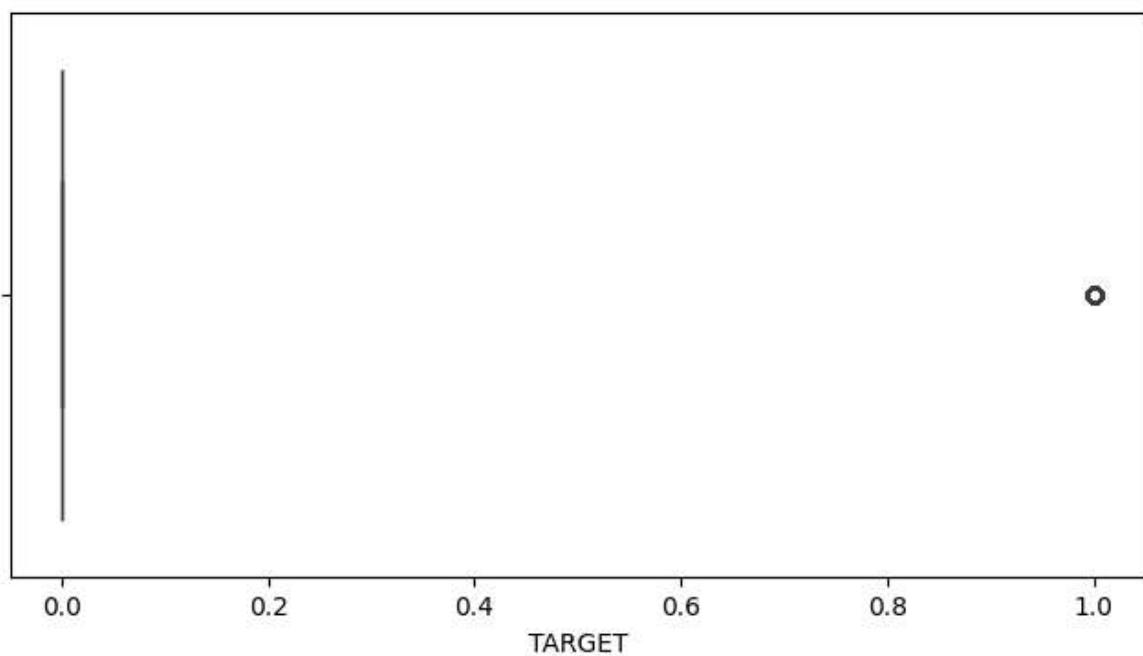
```
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna SK\_ID\_CURR



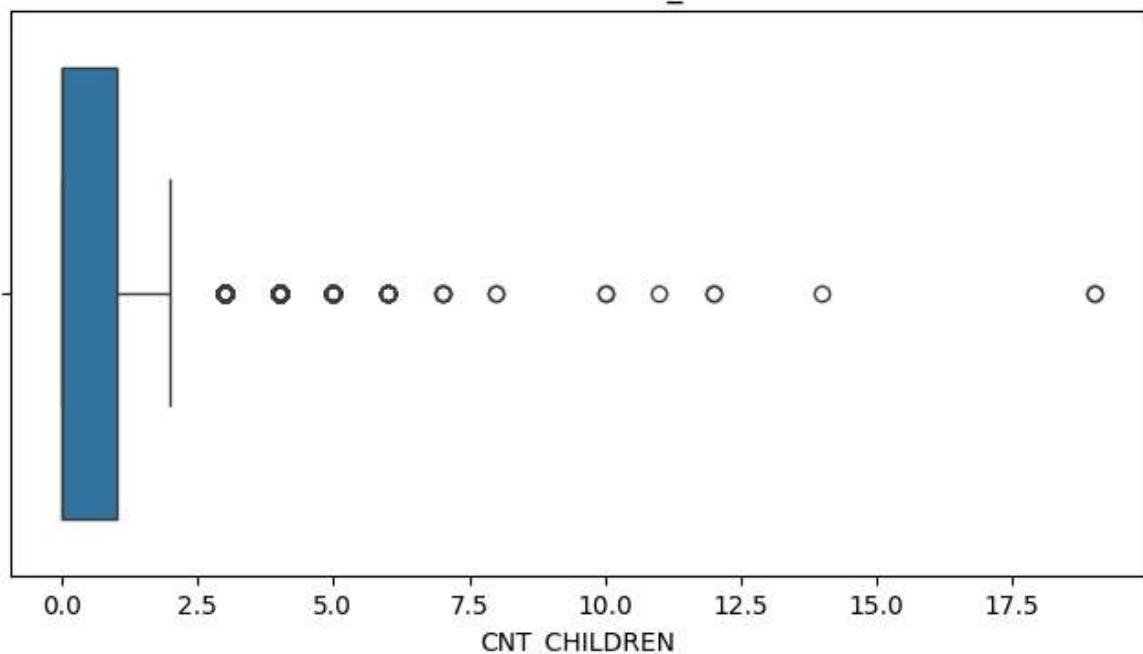
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna TARGET



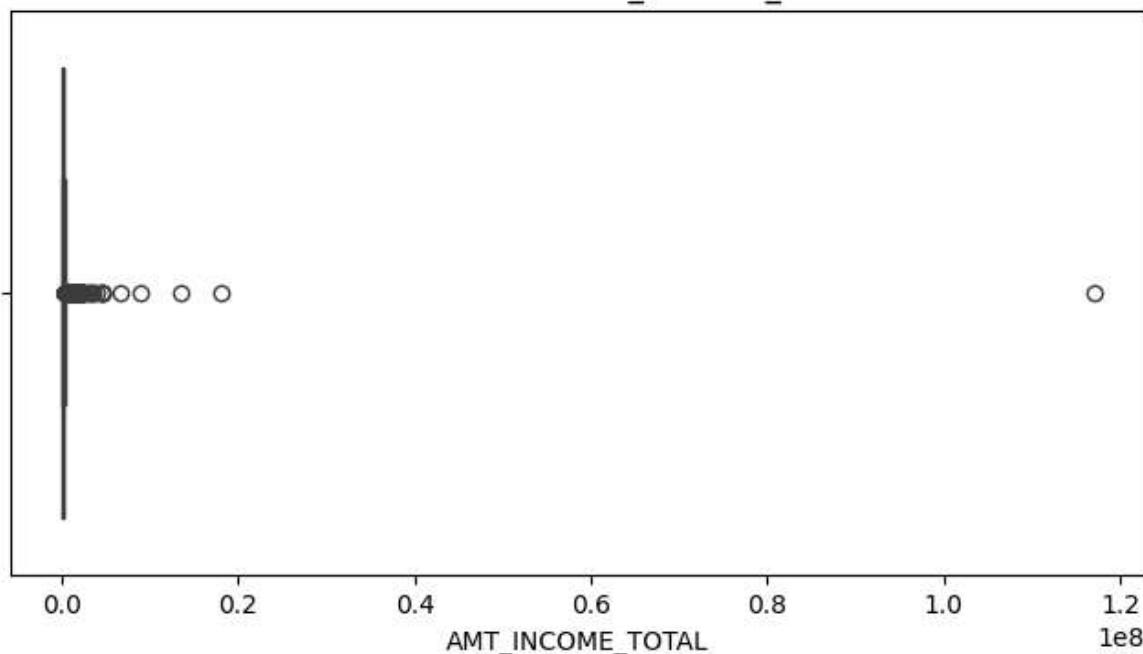
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna CNT\_CHILDREN



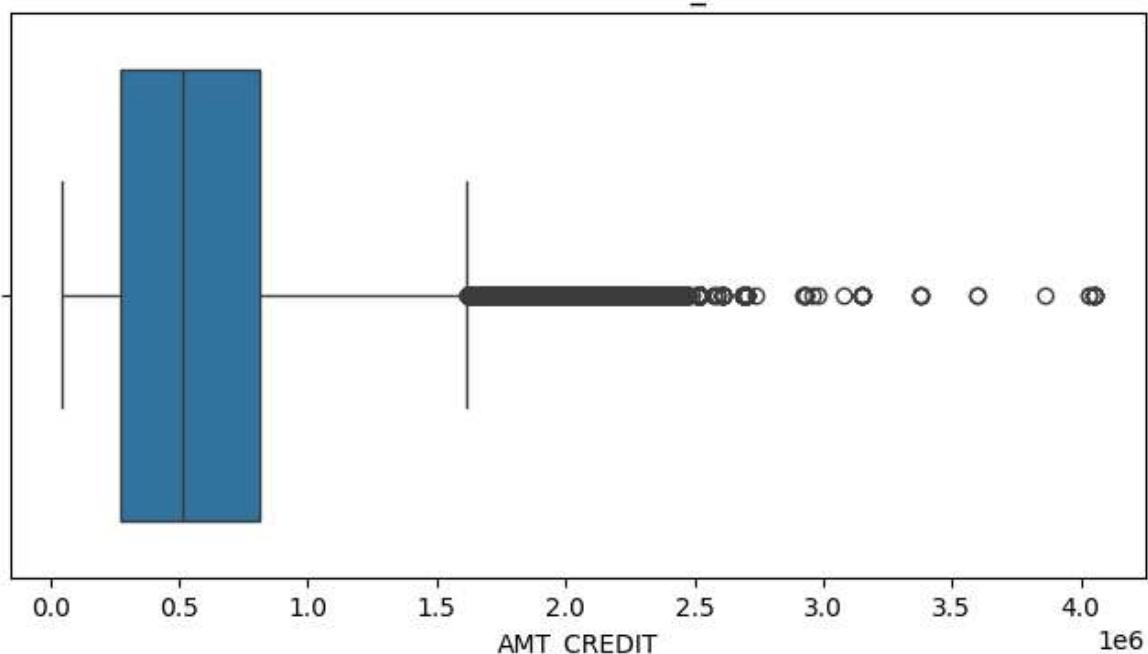
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna AMT\_INCOME\_TOTAL



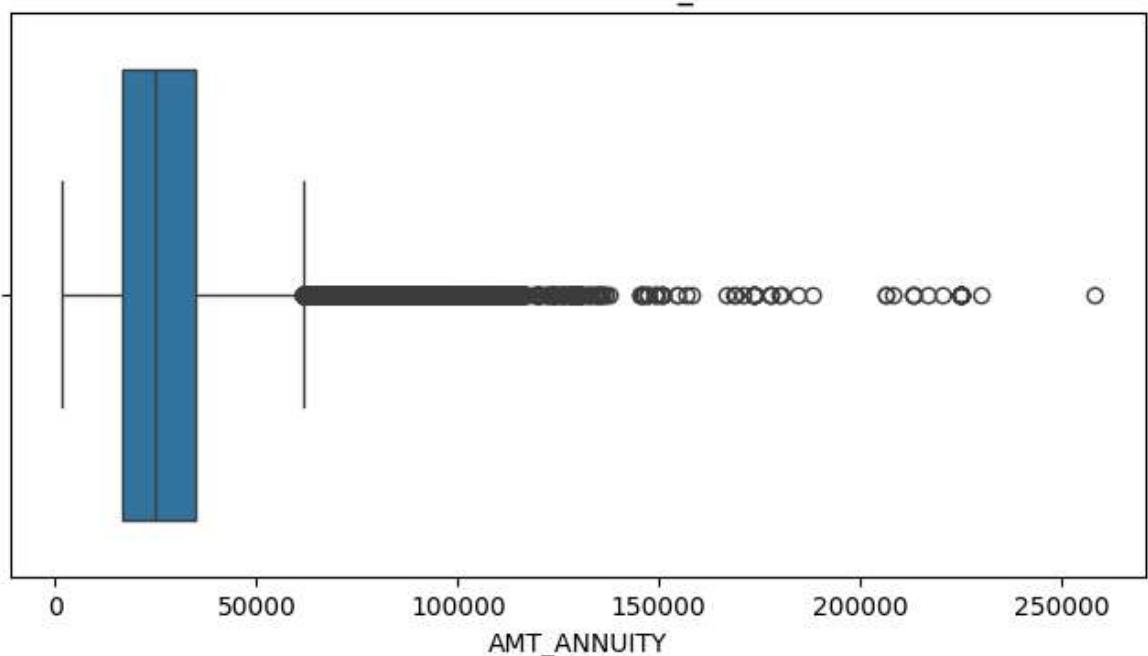
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna AMT\_CREDIT



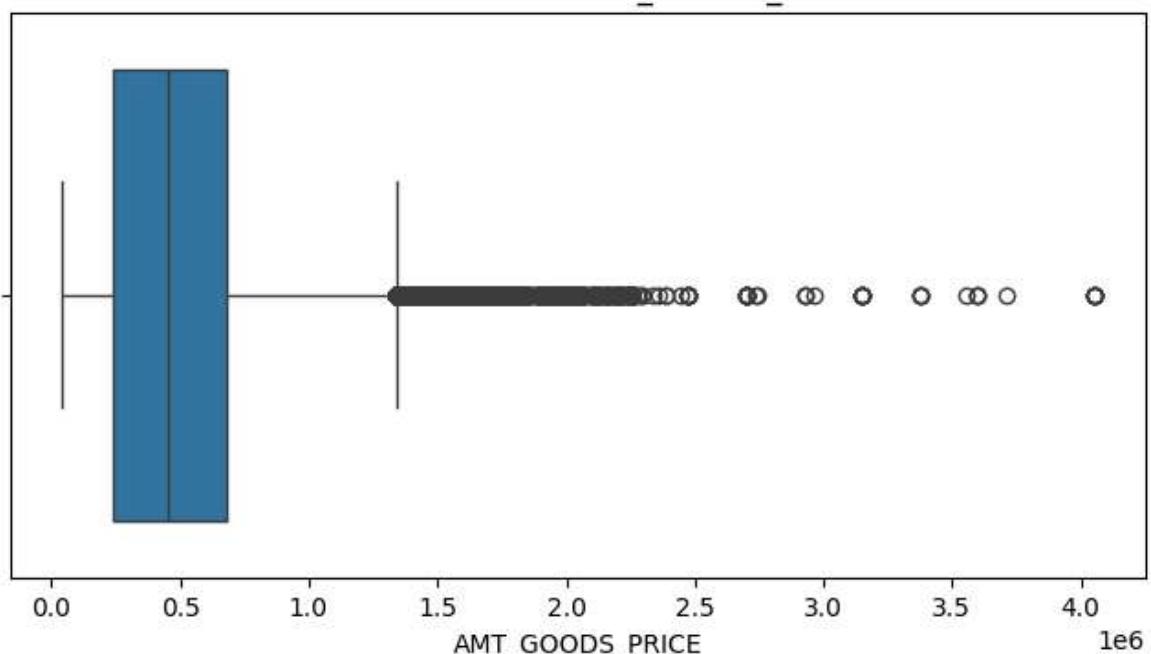
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna AMT\_ANNUITY



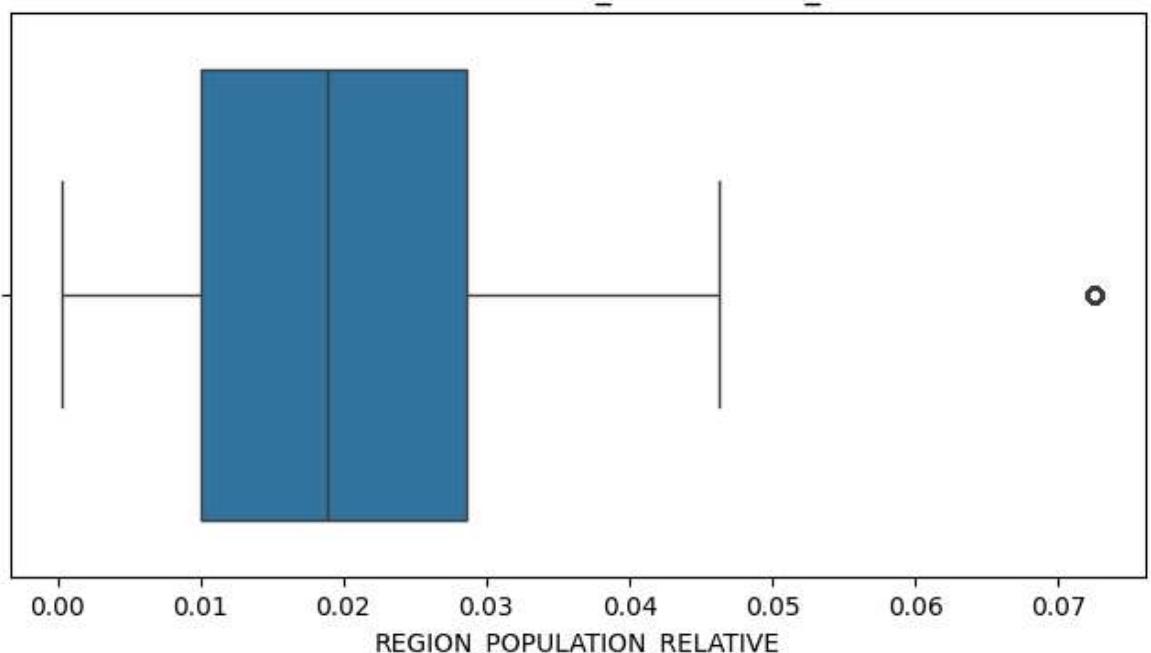
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna AMT\_GOODS\_PRICE



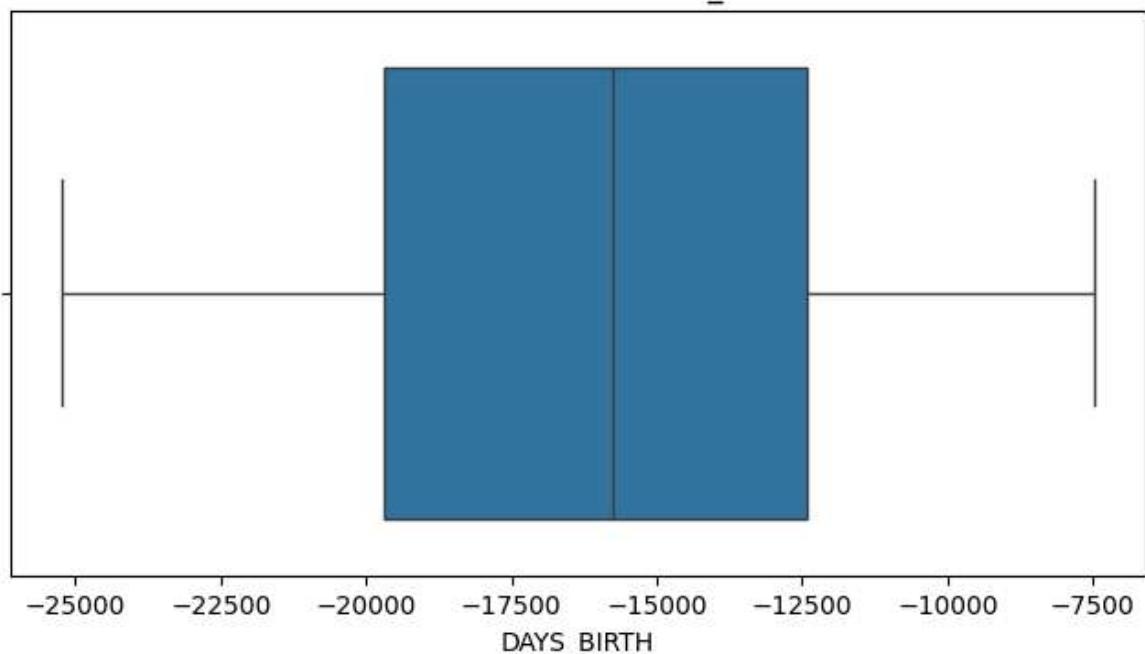
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna REGION\_POPULATION\_RELATIVE



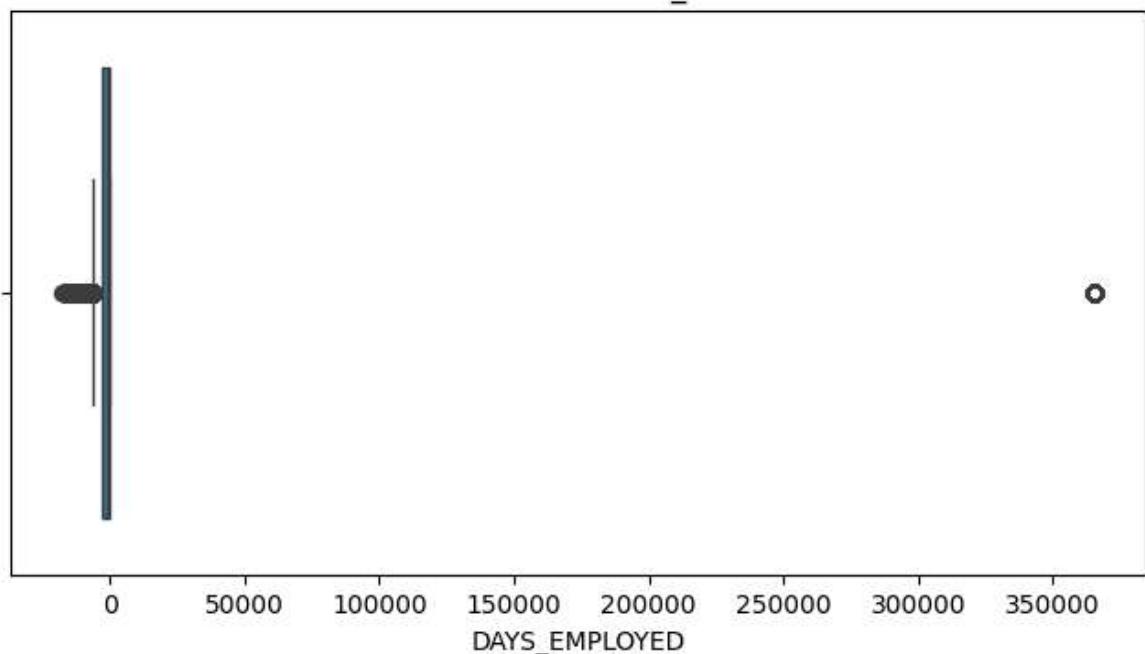
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna DAYS\_BIRTH



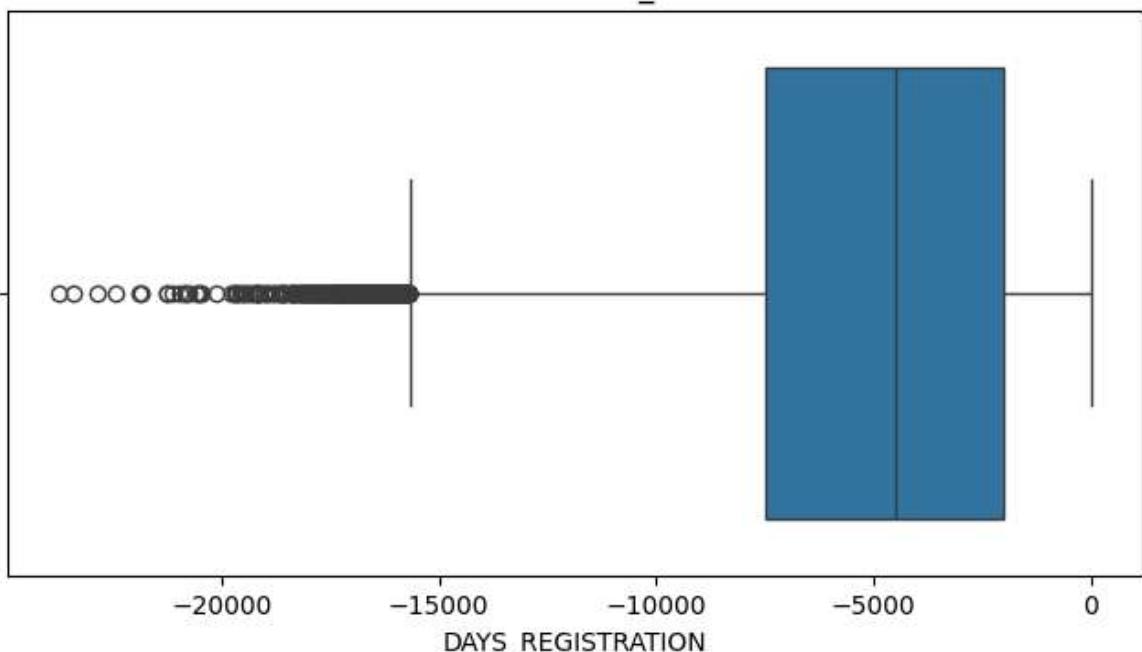
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna DAYS\_EMPLOYED



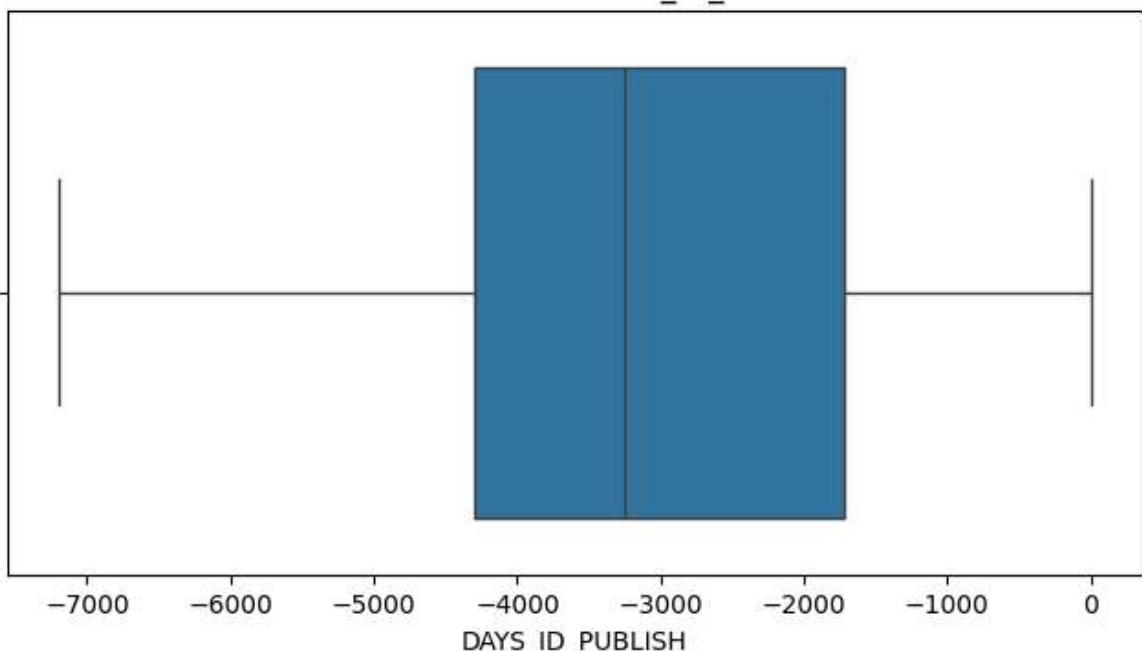
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna DAYS\_REGISTRATION



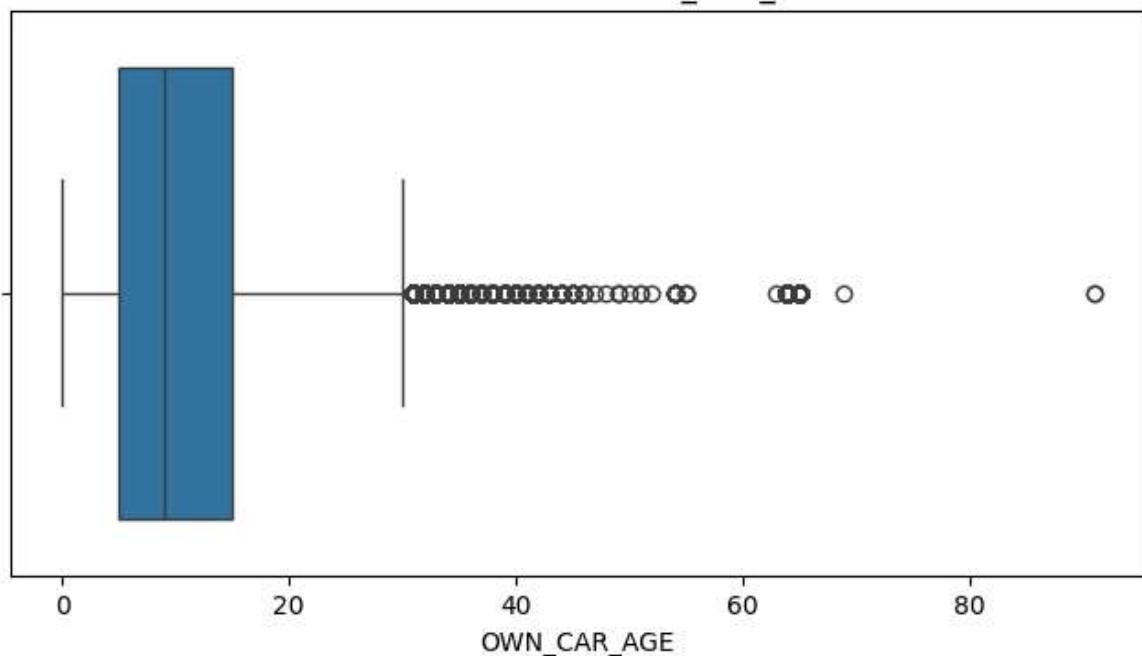
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna DAYS\_ID\_PUBLISH



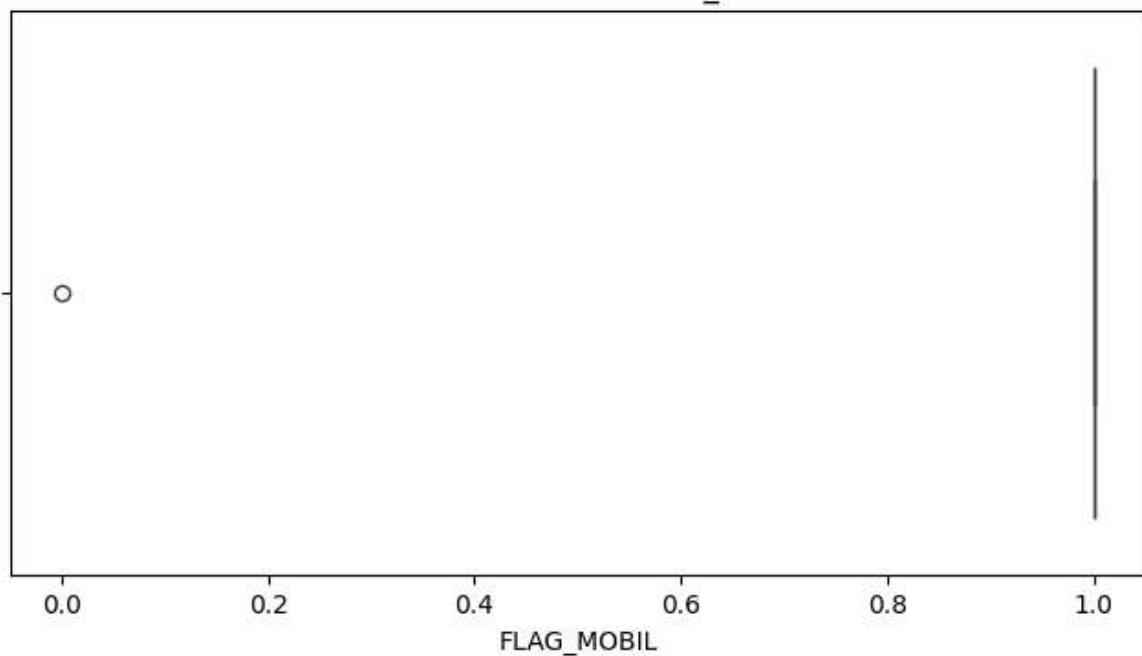
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna OWN\_CAR\_AGE



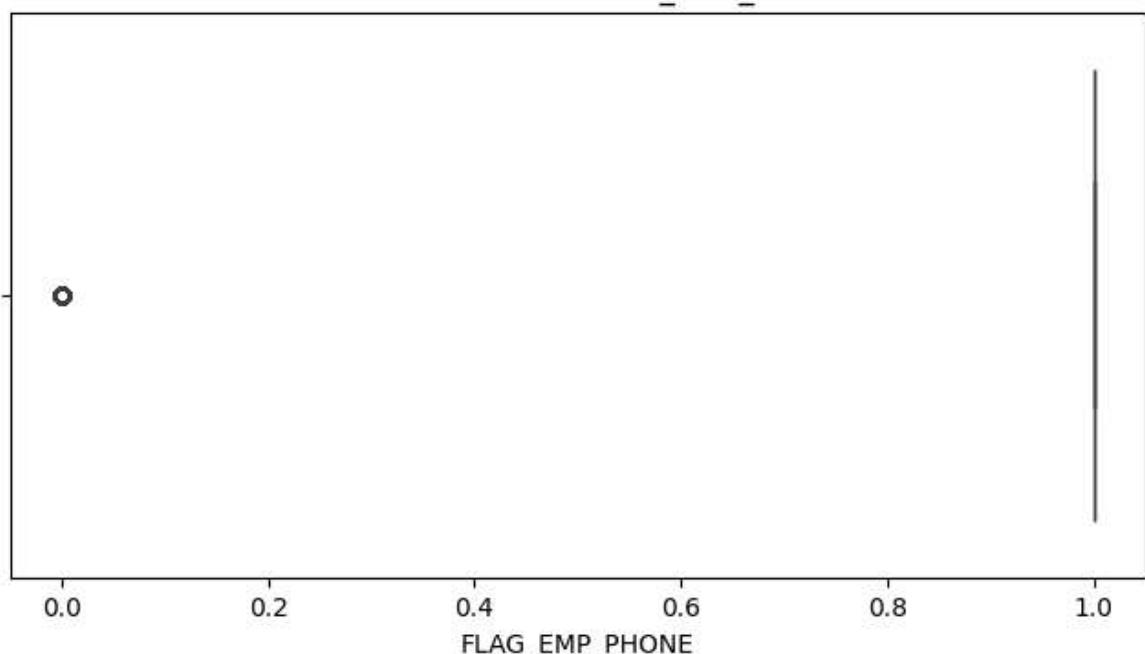
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna FLAG\_MOBIL



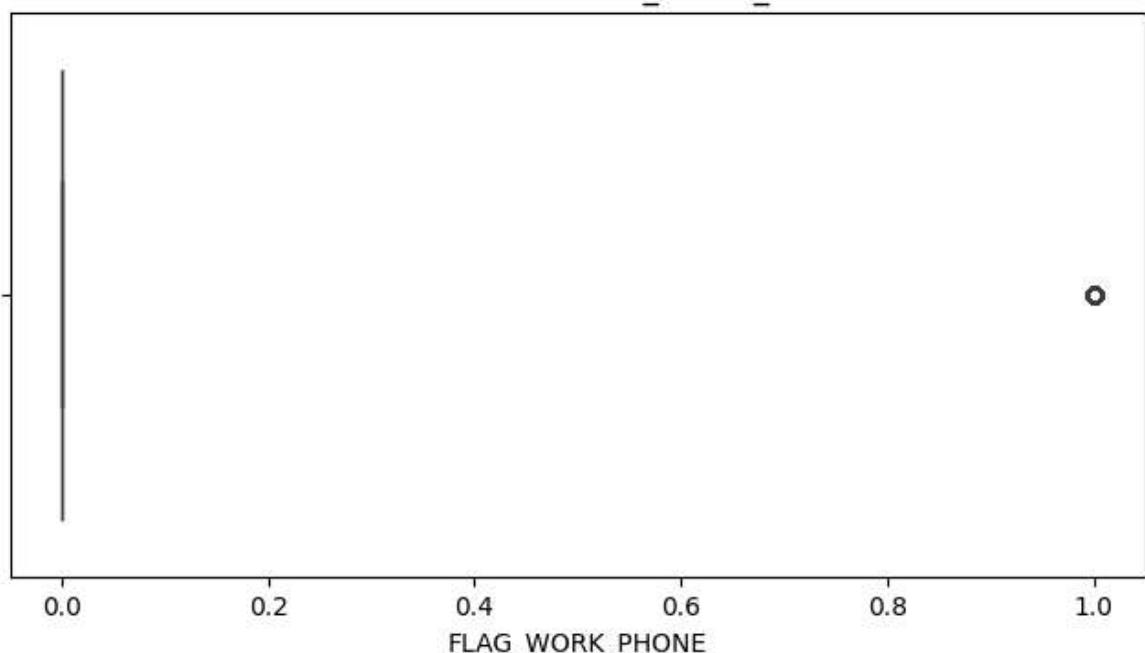
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna FLAG\_EMP\_PHONE



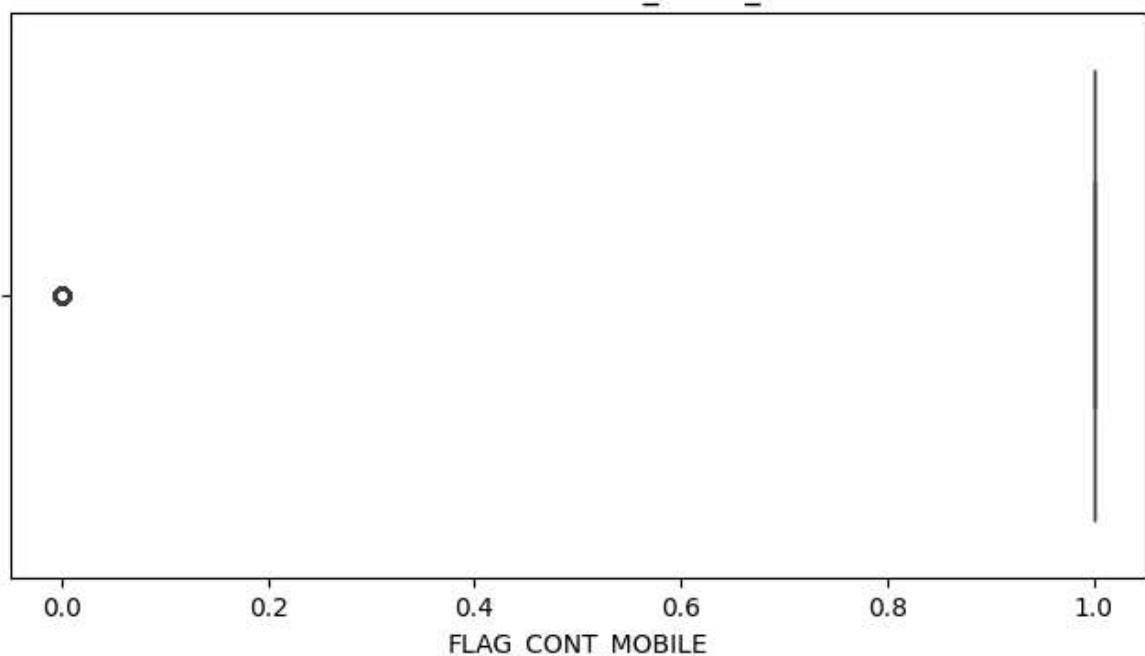
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna FLAG\_WORK\_PHONE



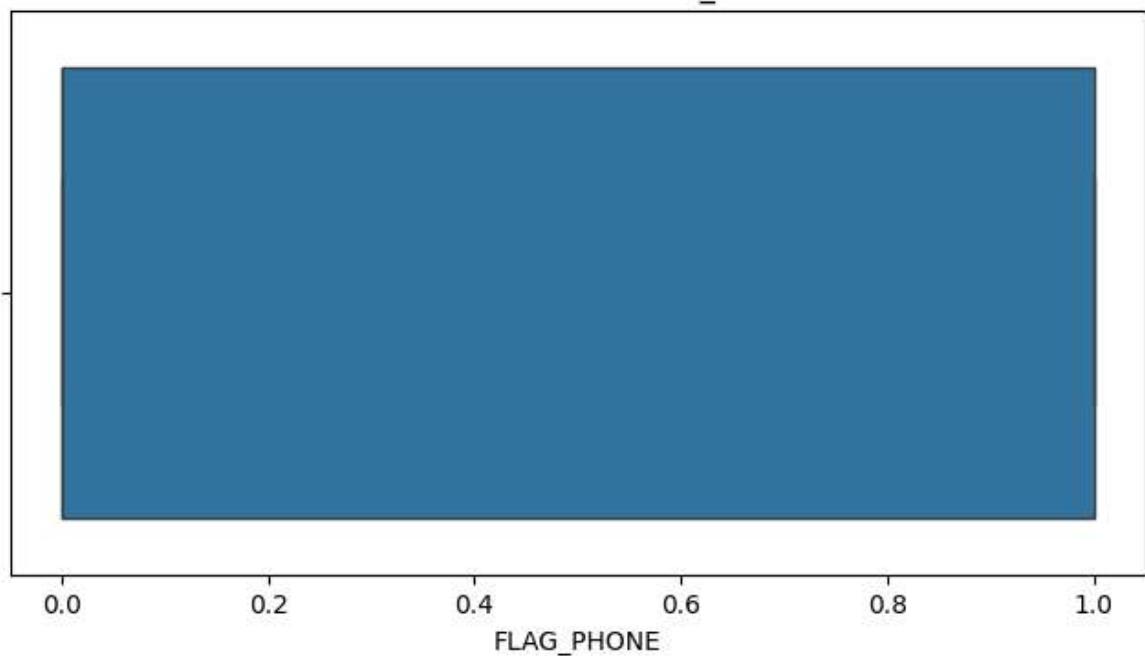
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_CONT\_MOBILE



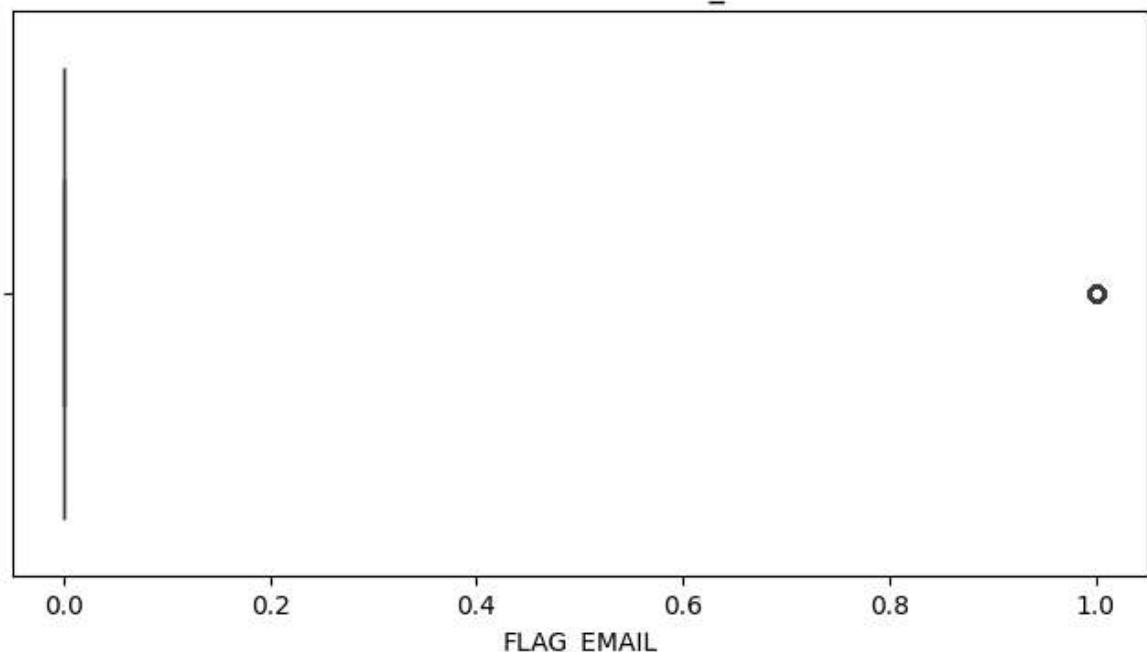
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_PHONE



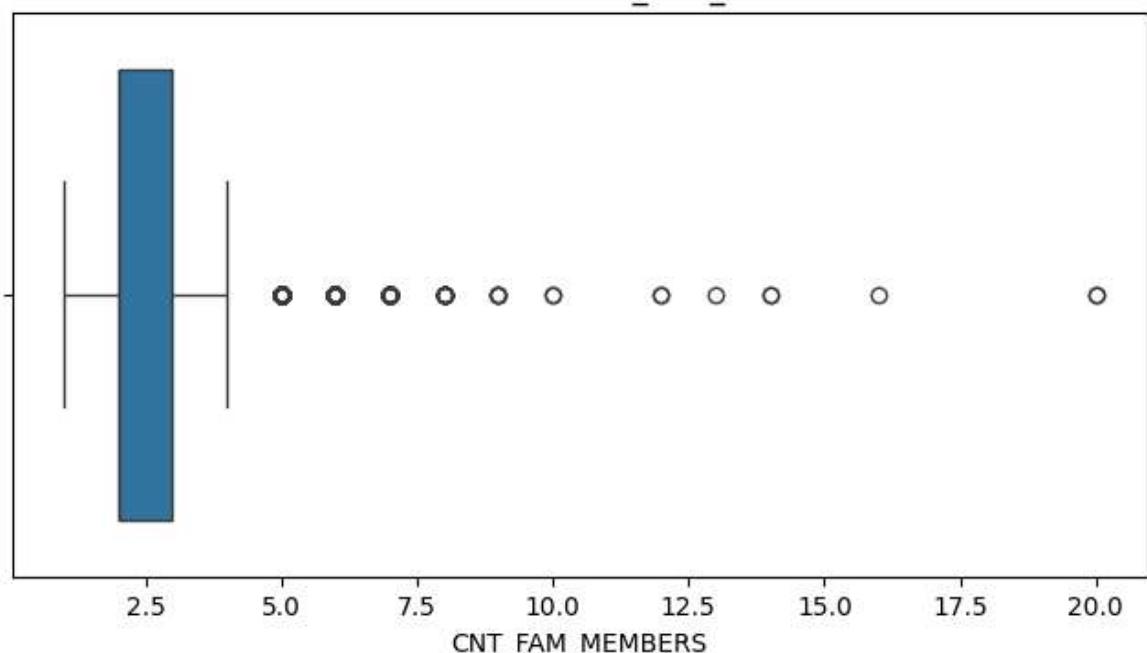
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_EMAIL



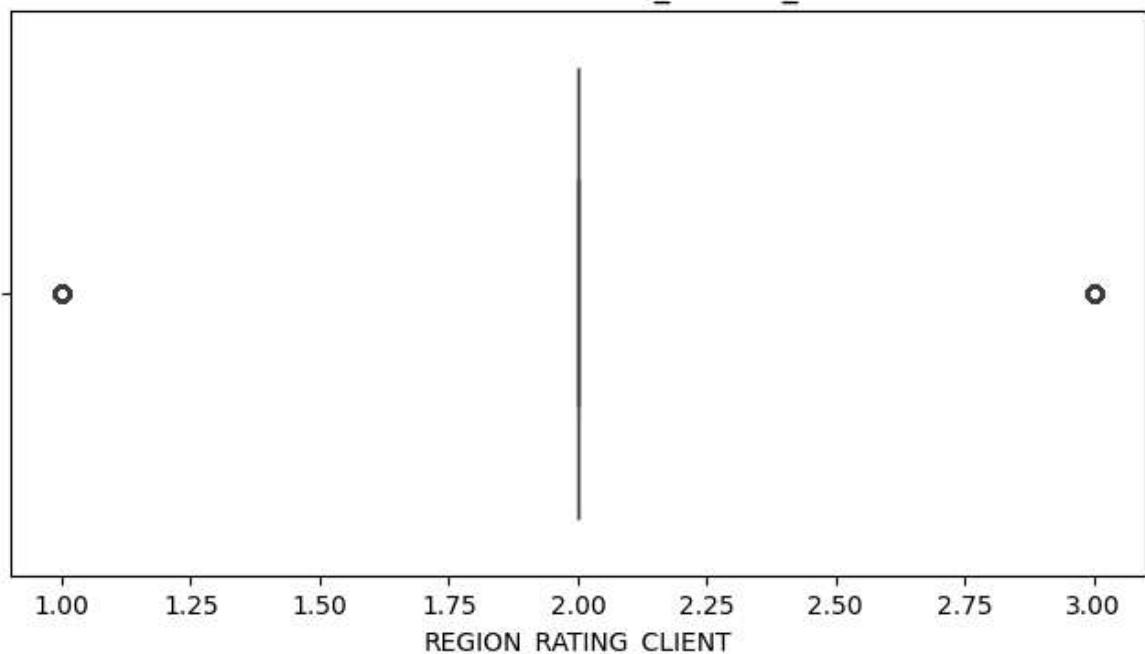
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna CNT\_FAM\_MEMBERS



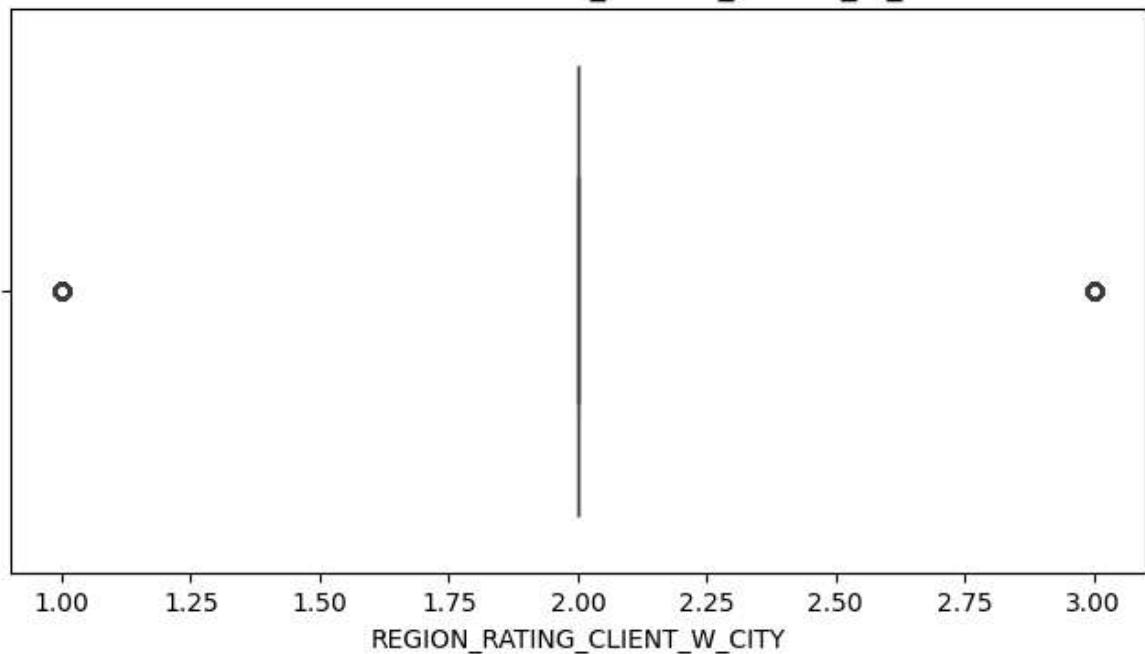
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna REGION\_RATING\_CLIENT



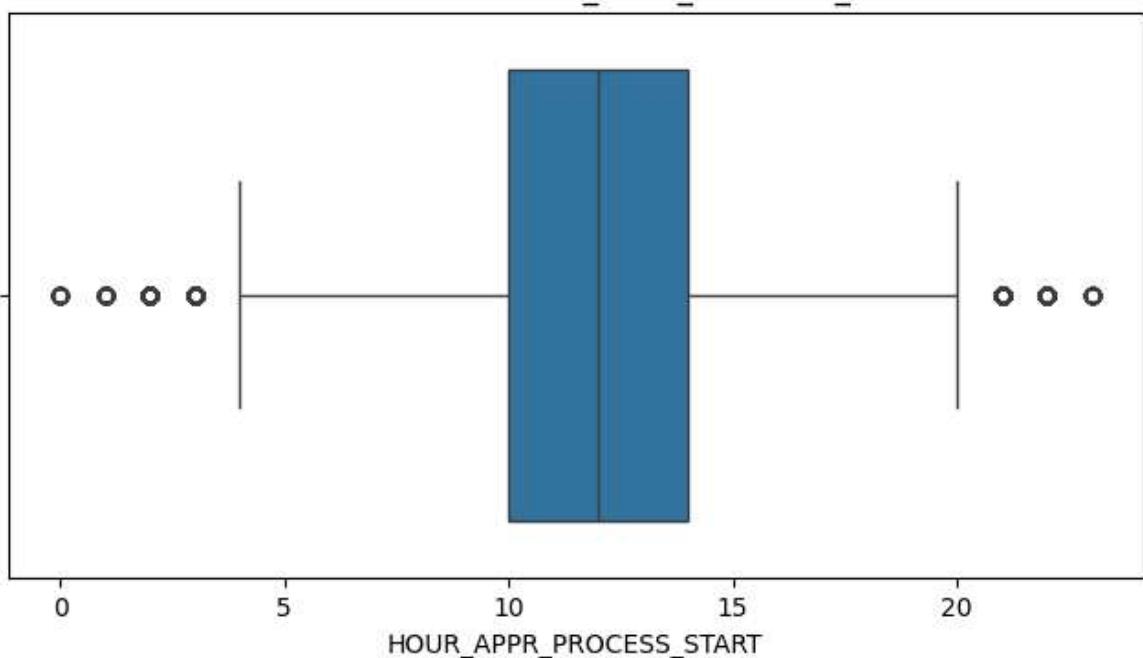
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna REGION\_RATING\_CLIENT\_W\_CITY



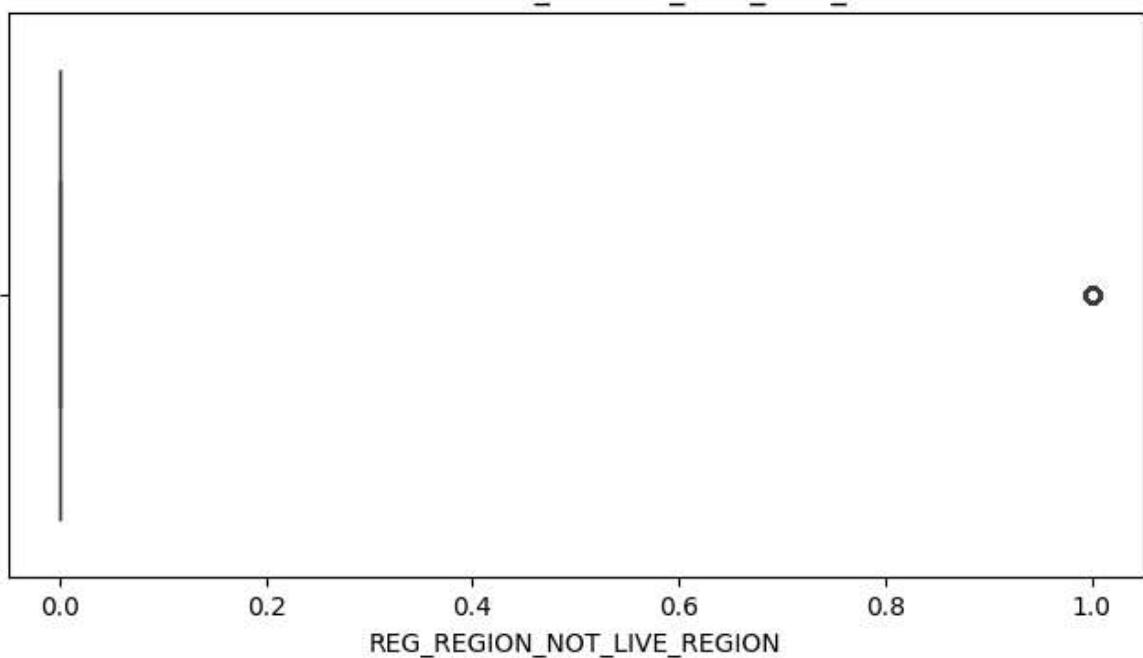
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna HOUR\_APPR\_PROCESS\_START



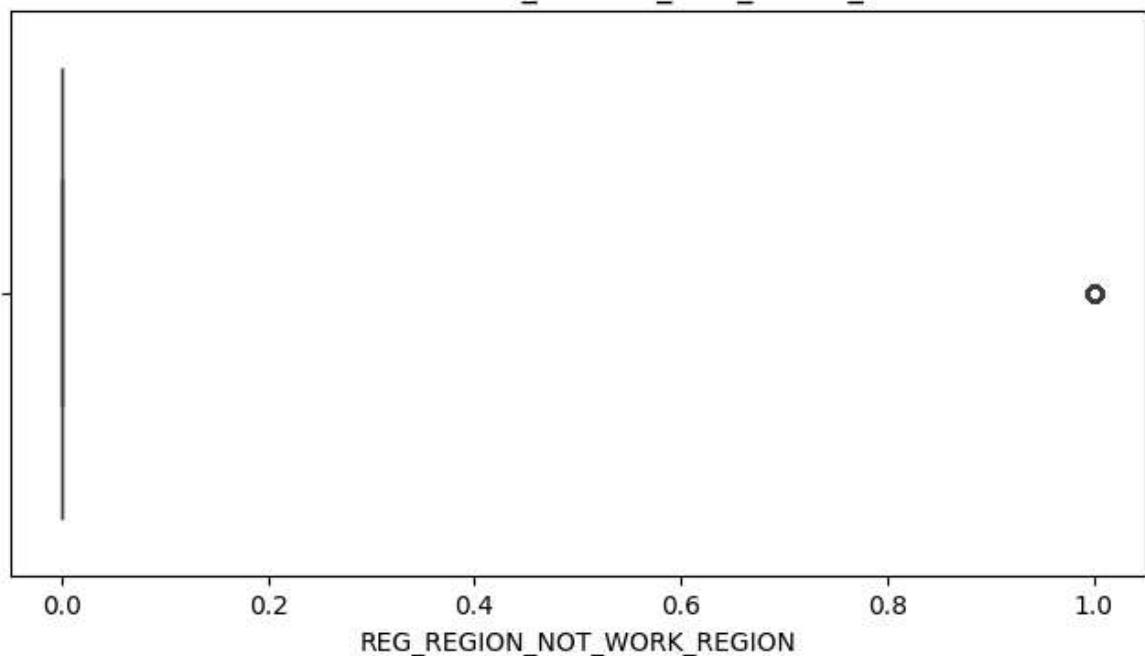
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna REG\_REGION\_NOT\_LIVE\_REGION



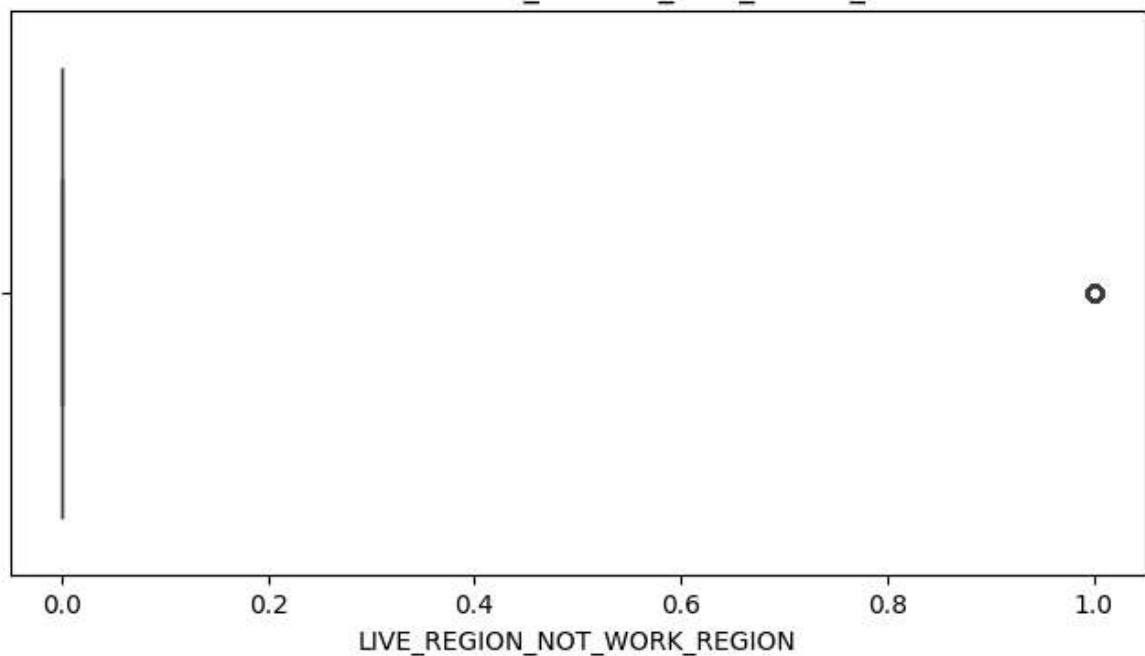
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna REG\_REGION\_NOT\_WORK\_REGION



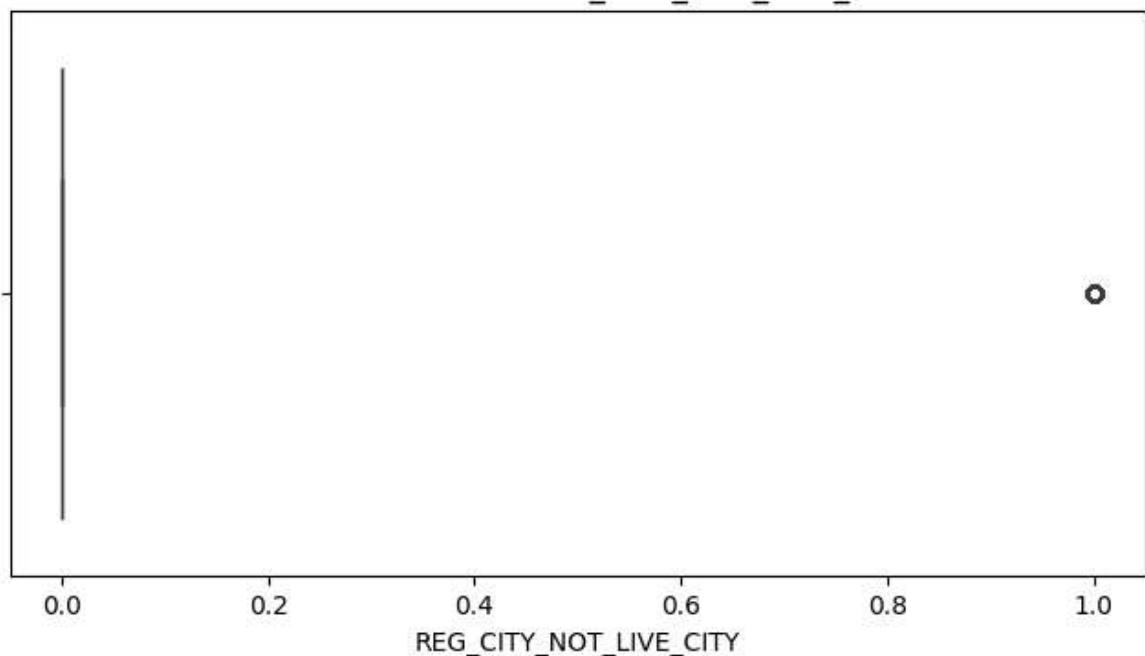
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna LIVE\_REGION\_NOT\_WORK\_REGION



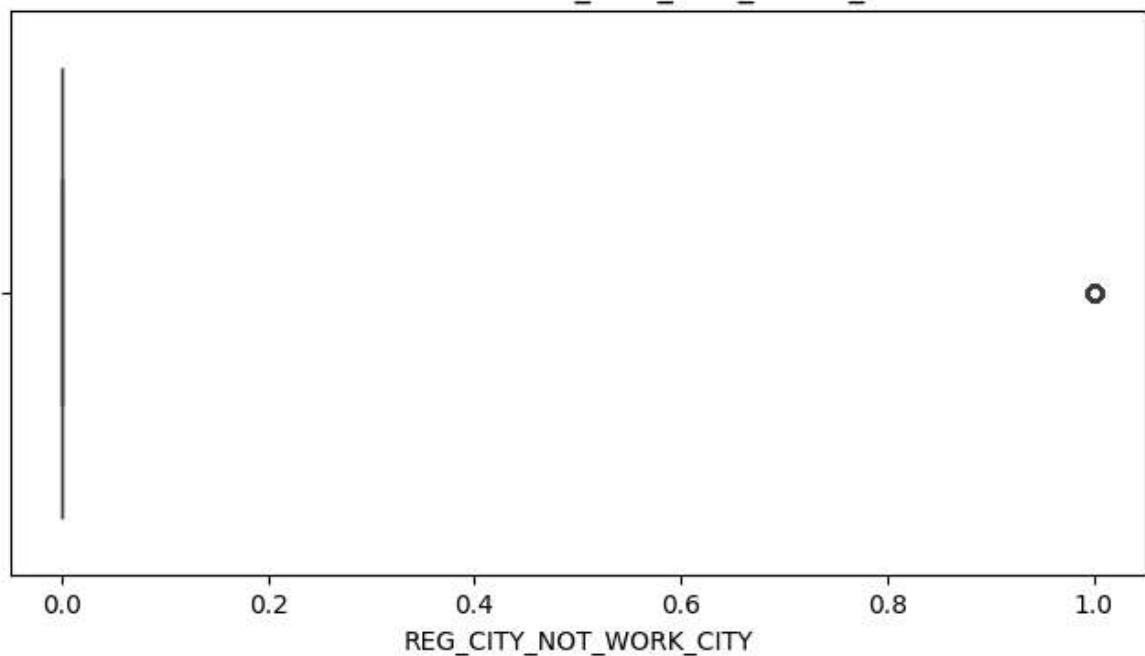
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna REG\_CITY\_NOT\_LIVE\_CITY



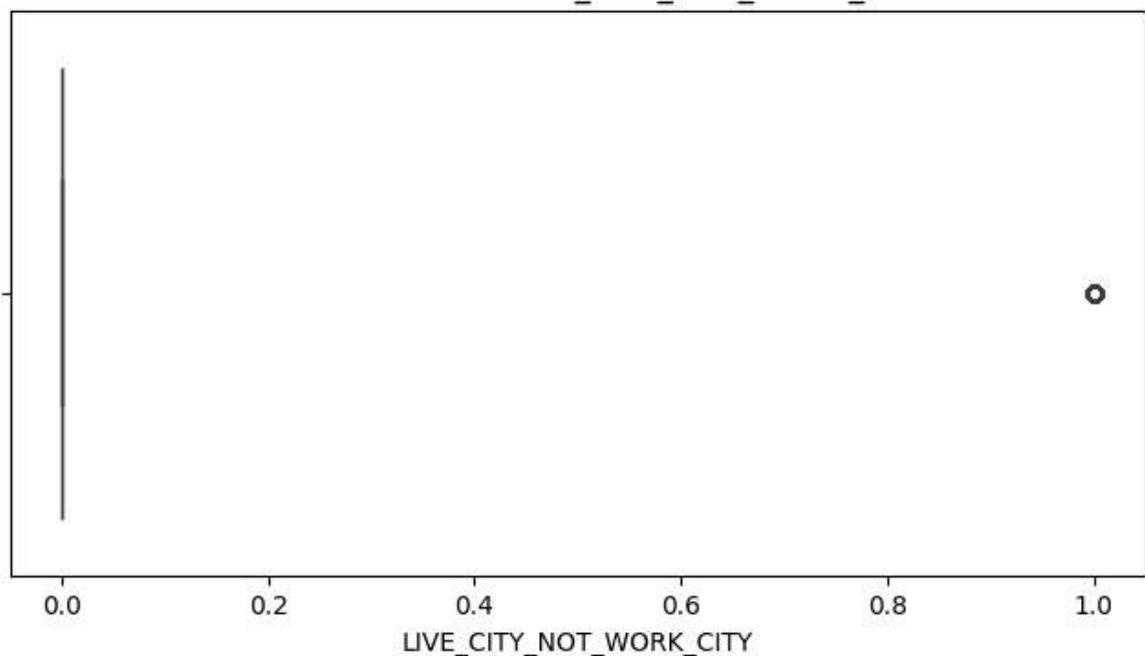
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna REG\_CITY\_NOT\_WORK\_CITY



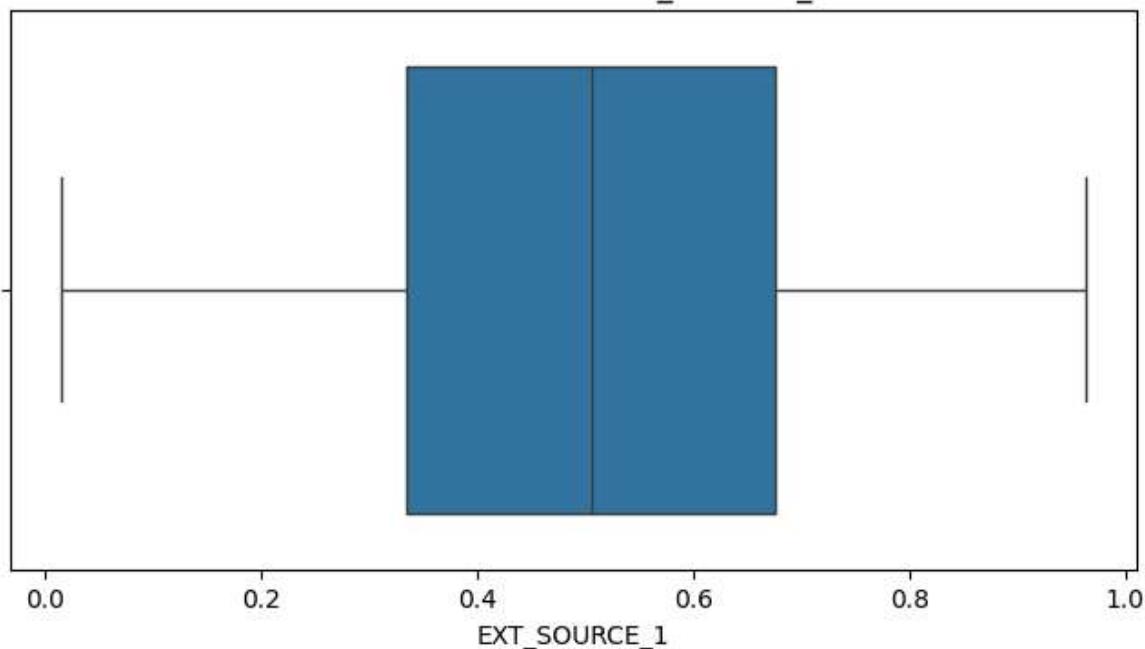
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna LIVE\_CITY\_NOT\_WORK\_CITY



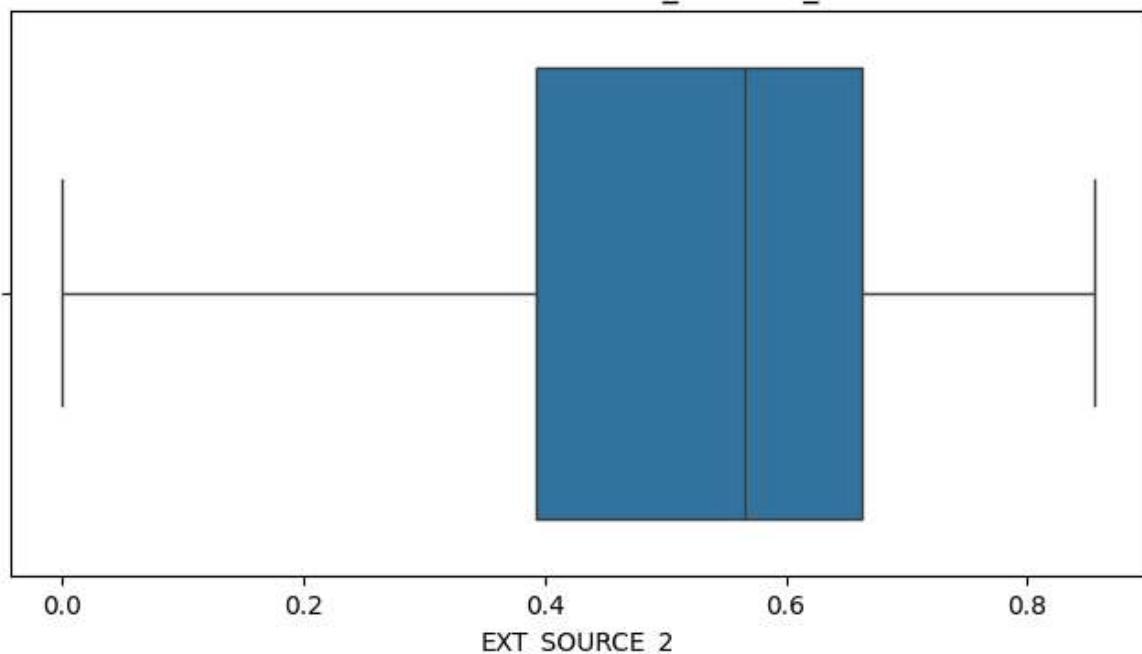
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna EXT\_SOURCE\_1



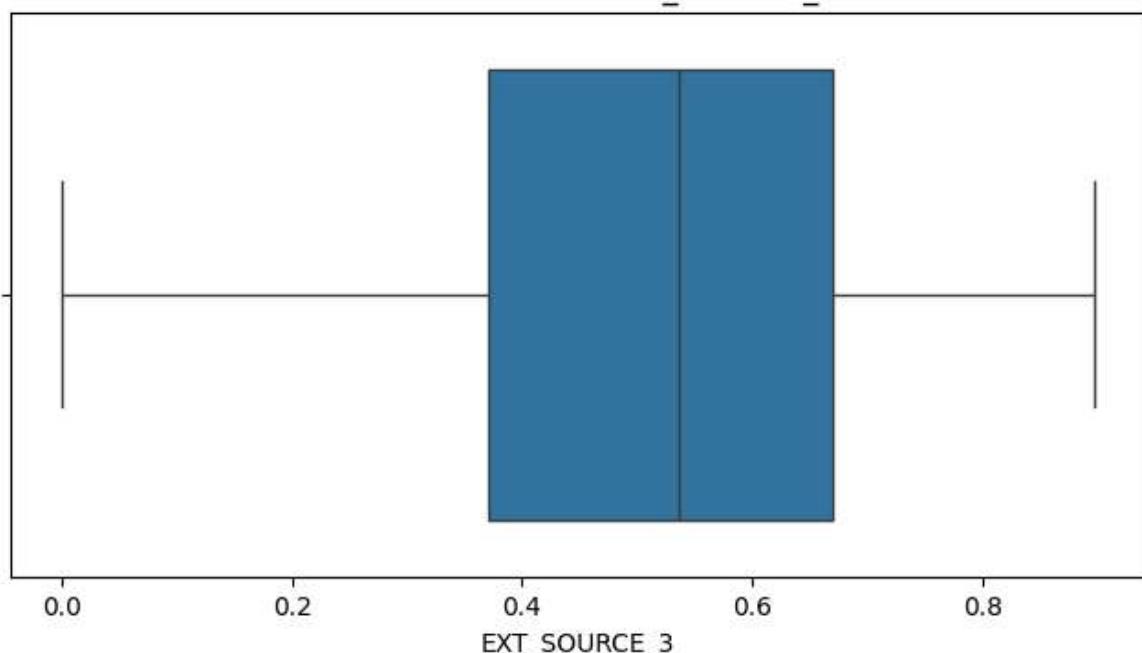
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna EXT\_SOURCE\_2



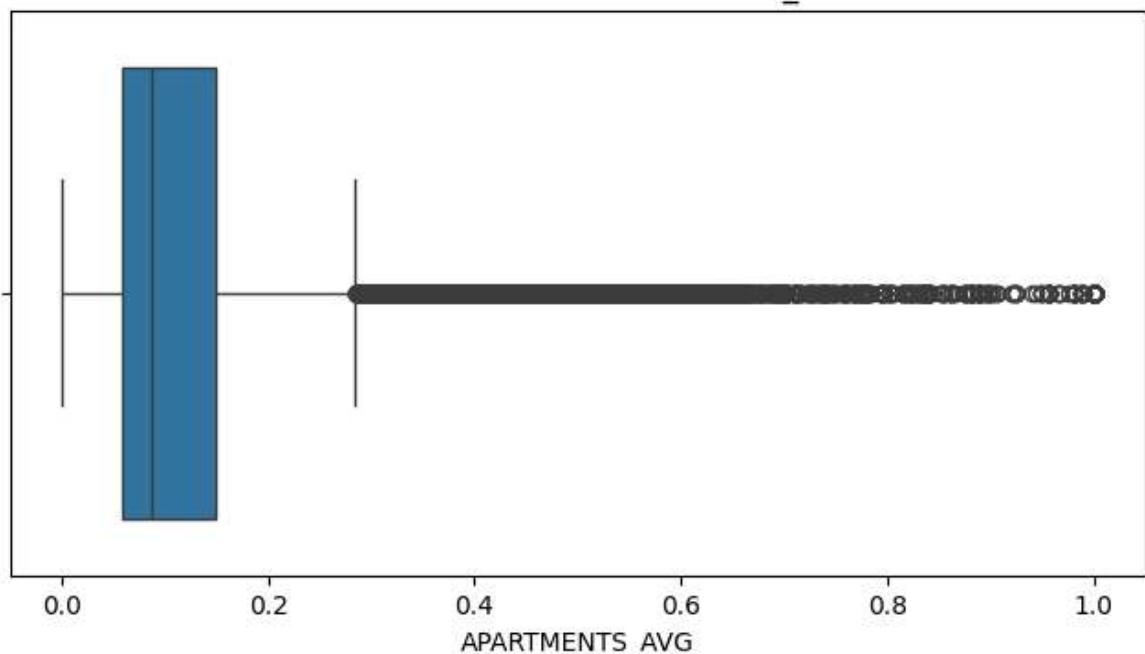
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna EXT\_SOURCE\_3



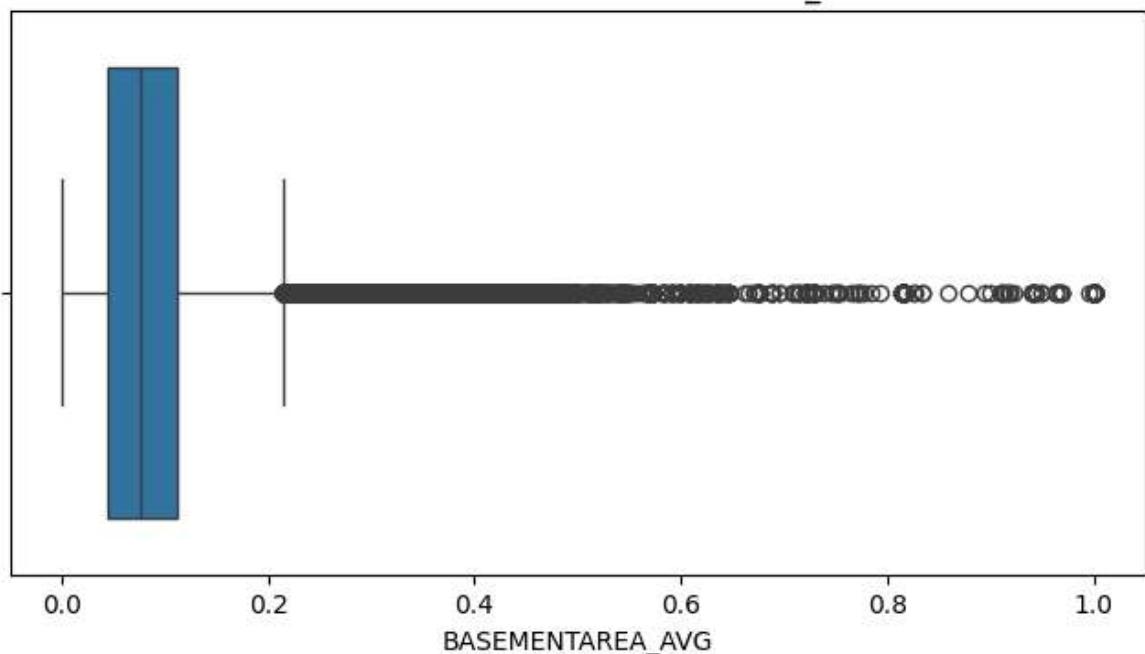
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna APARTMENTS\_AVG



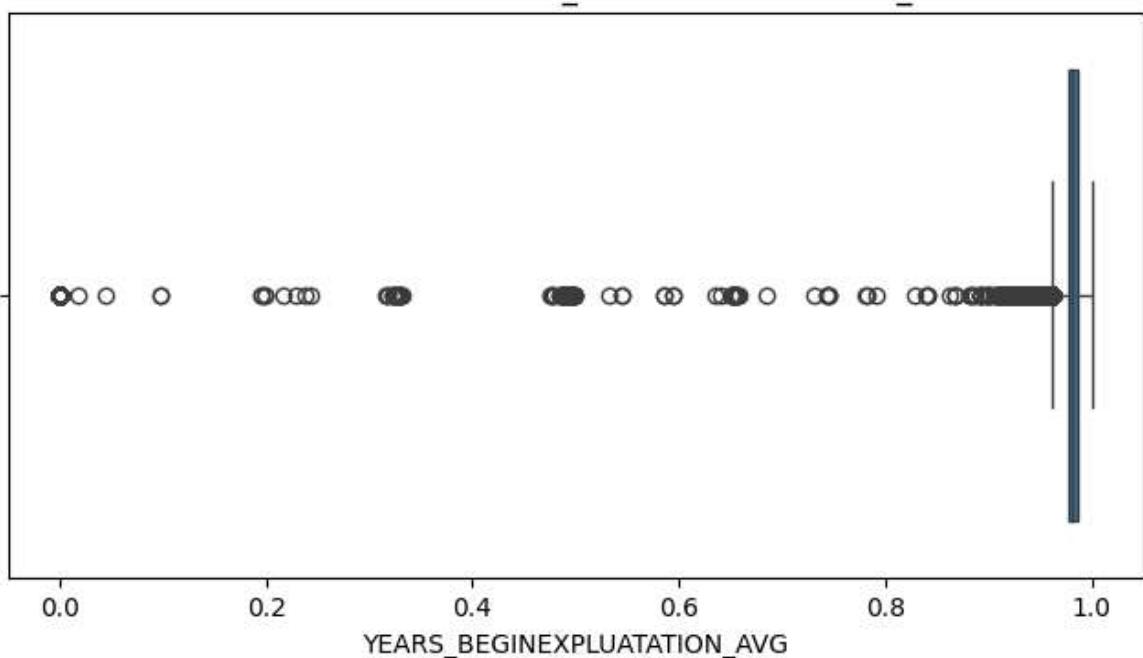
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna BASEMENTAREA\_AVG



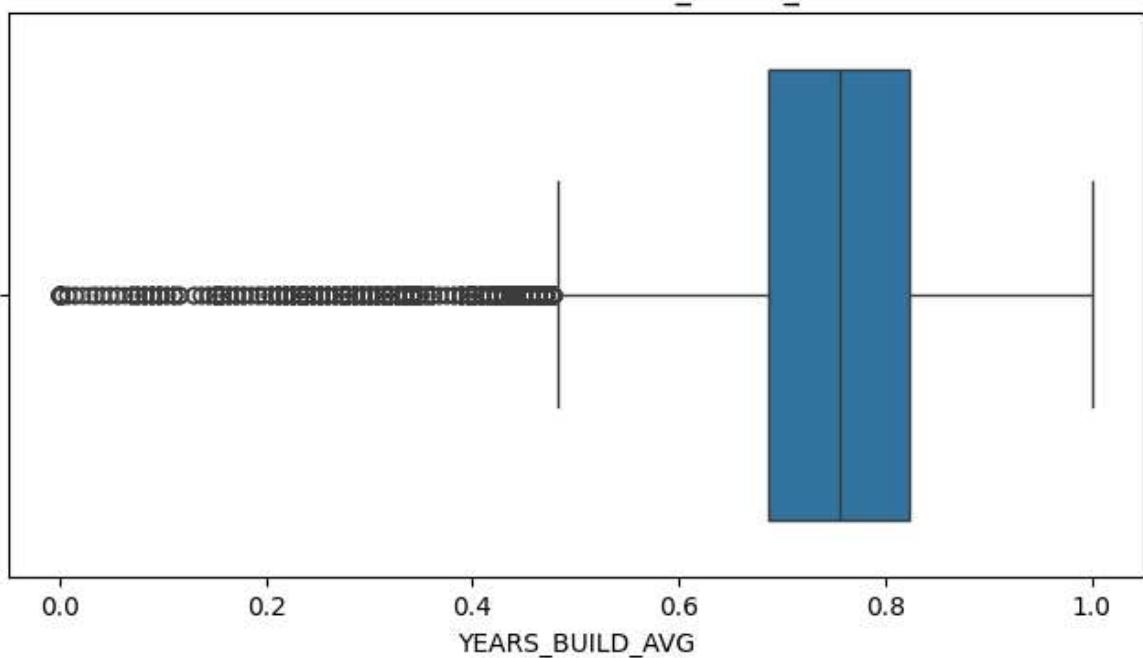
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna YEARS\_BEGINEXPLUATATION\_AVG



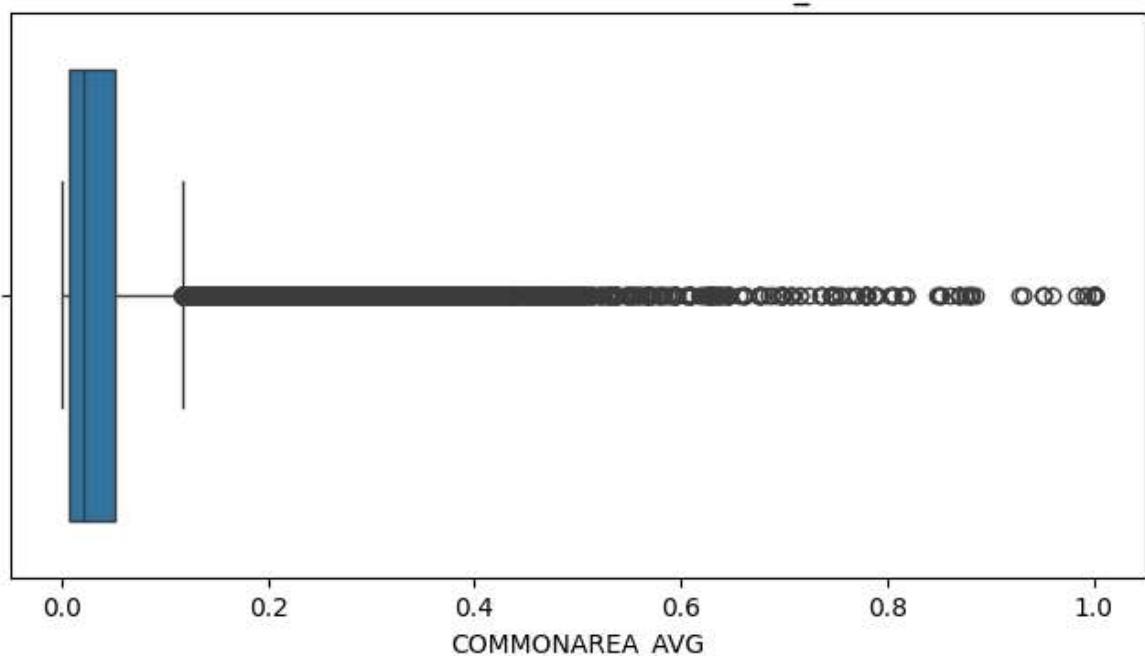
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna YEARS\_BUILD\_AVG



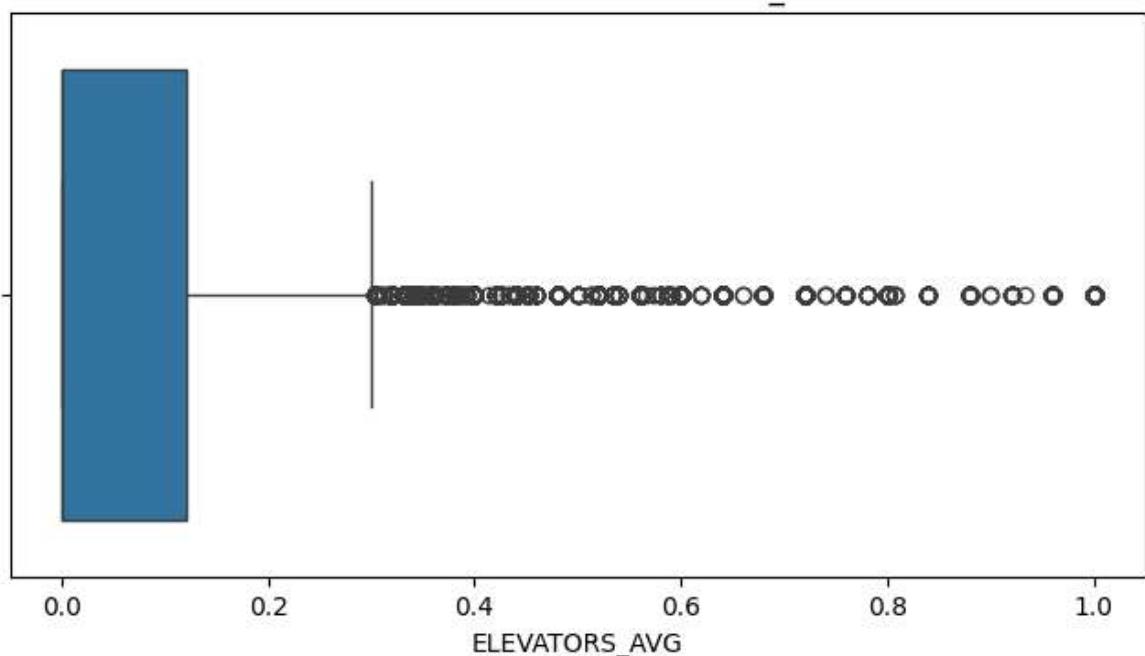
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna COMMONAREA\_AVG



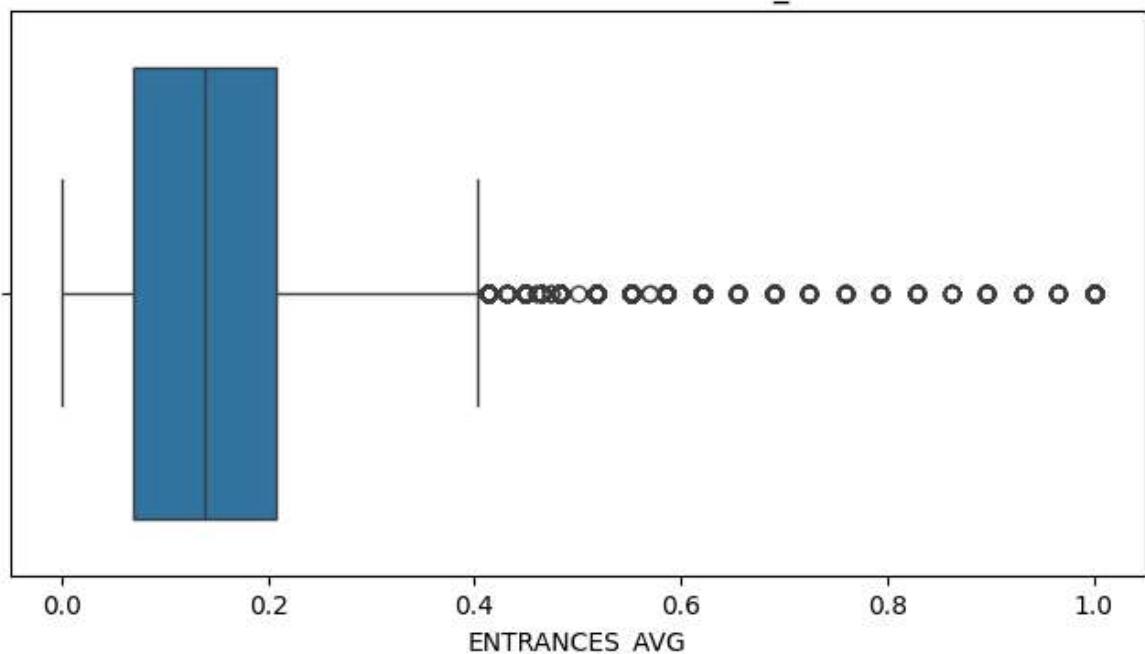
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\n\b\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna ELEVATORS\_AVG



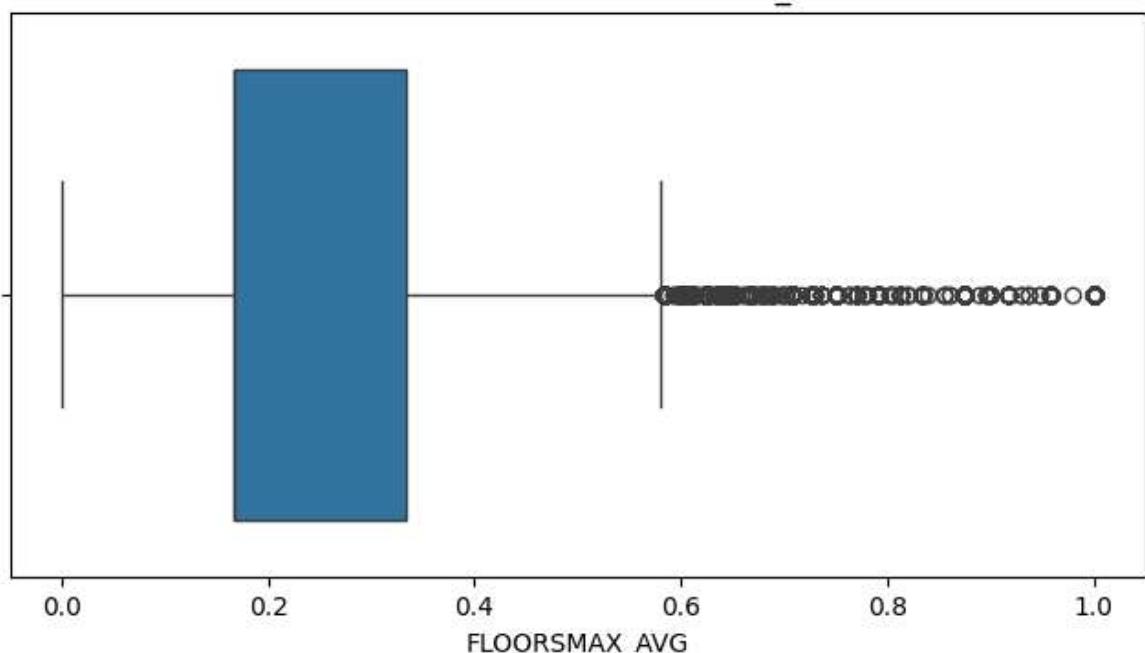
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\n\b\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna ENTRANCES\_AVG



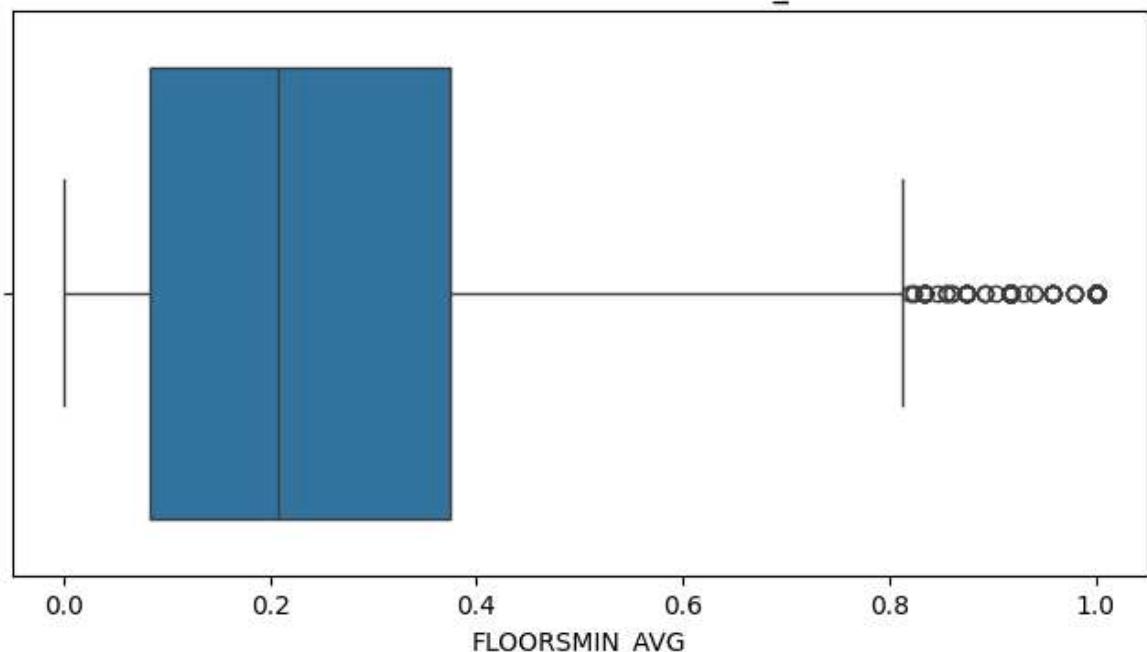
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna FLOORSMAX\_AVG



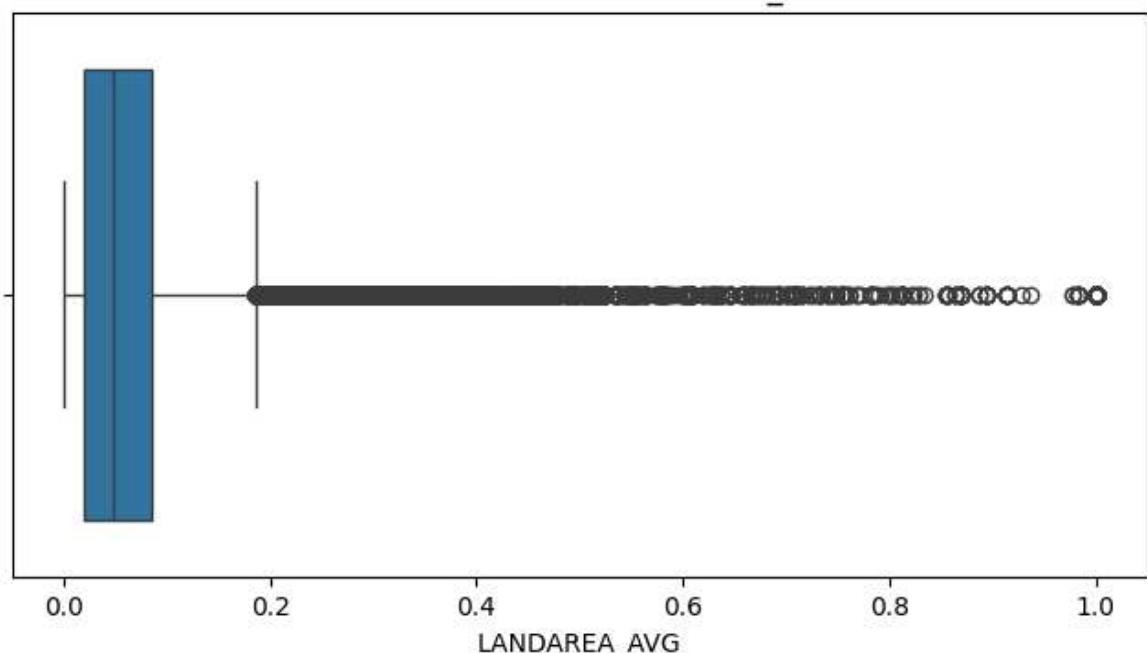
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna FLOORSMIN\_AVG



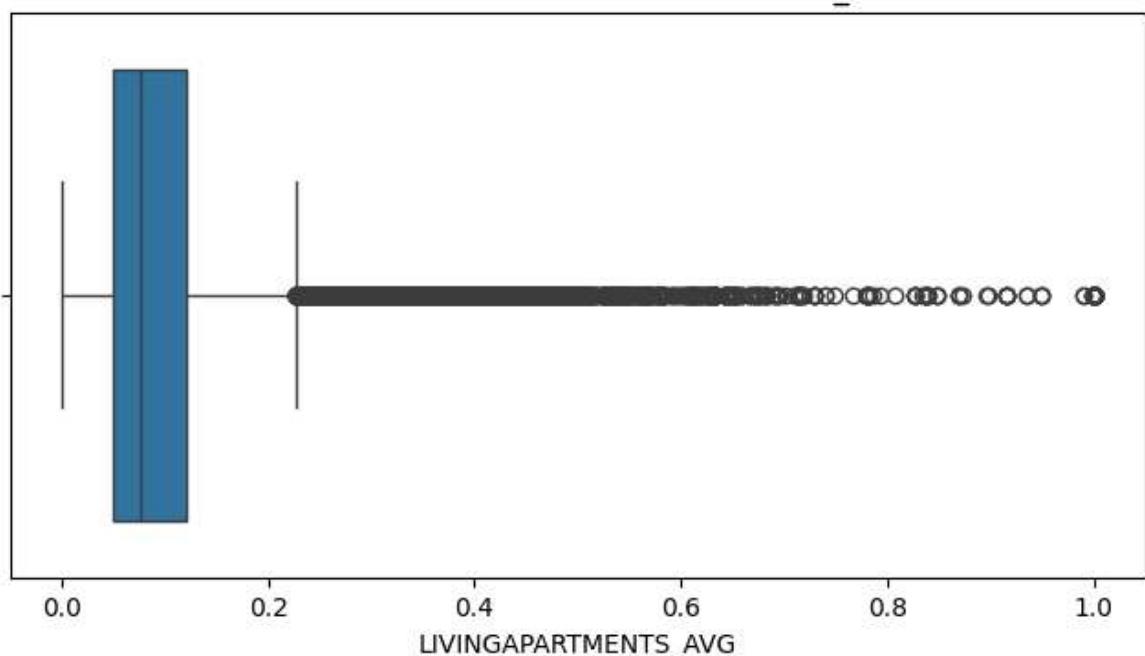
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna LANDAREA\_AVG



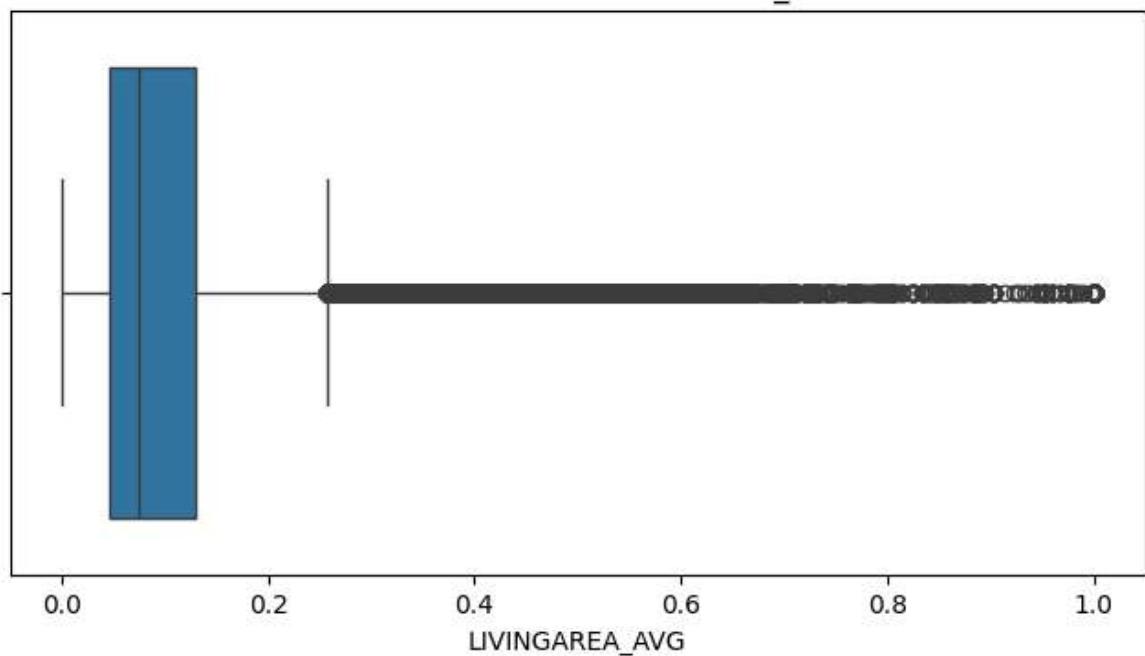
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna LIVINGAPARTMENTS\_AVG



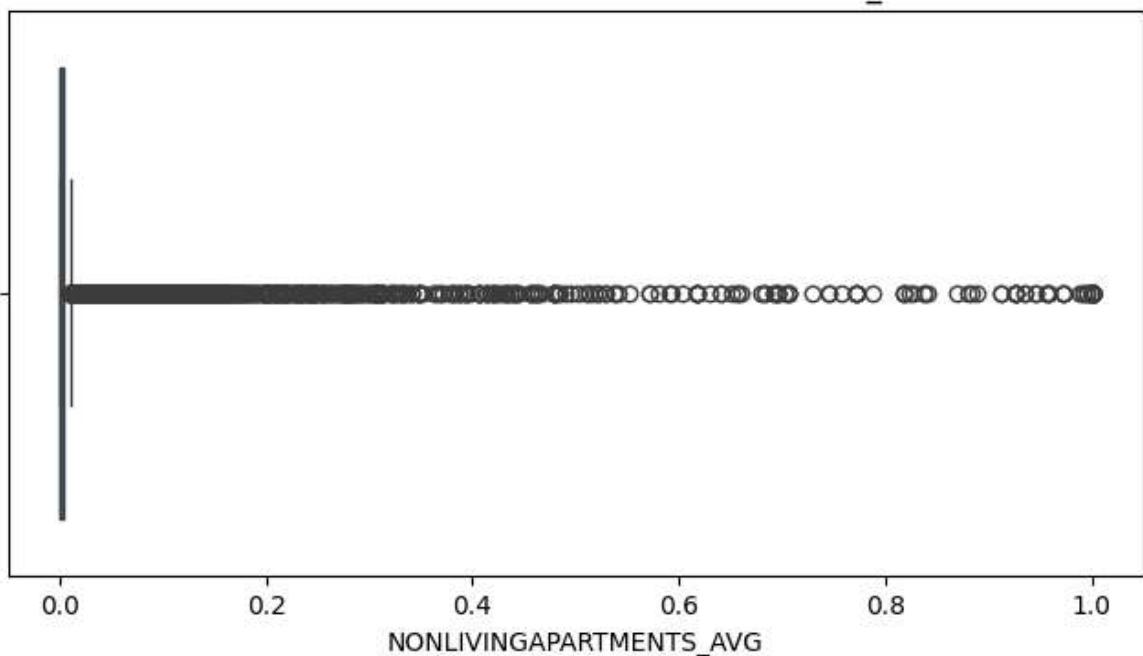
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna LIVINGAREA\_AVG



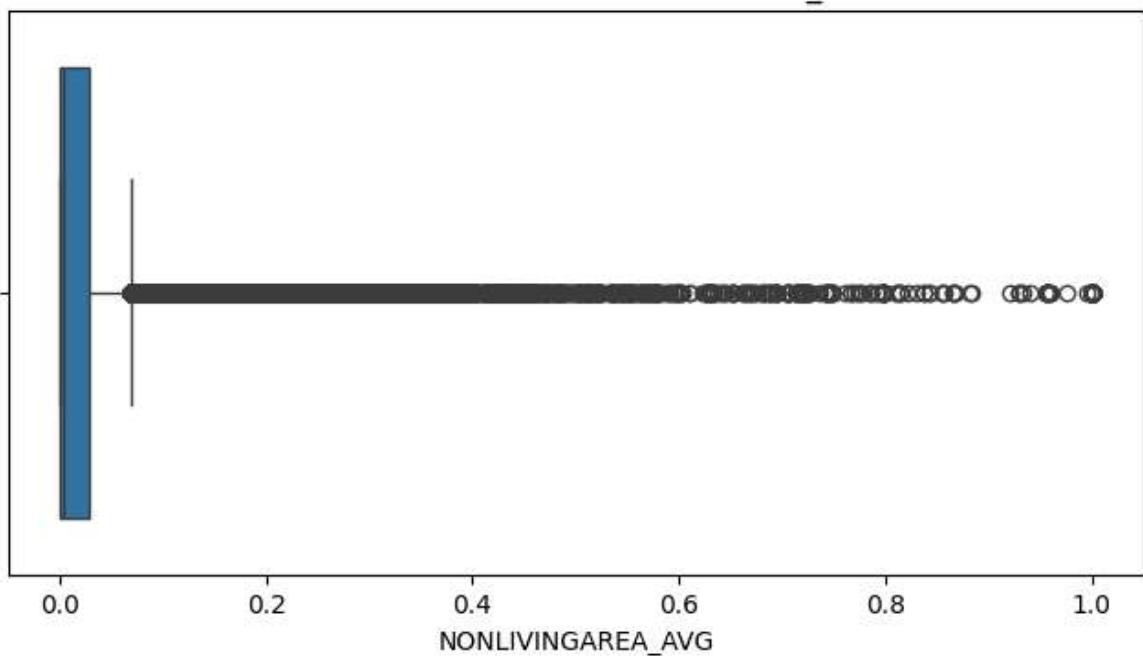
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna NONLIVINGAPARTMENTS\_AVG



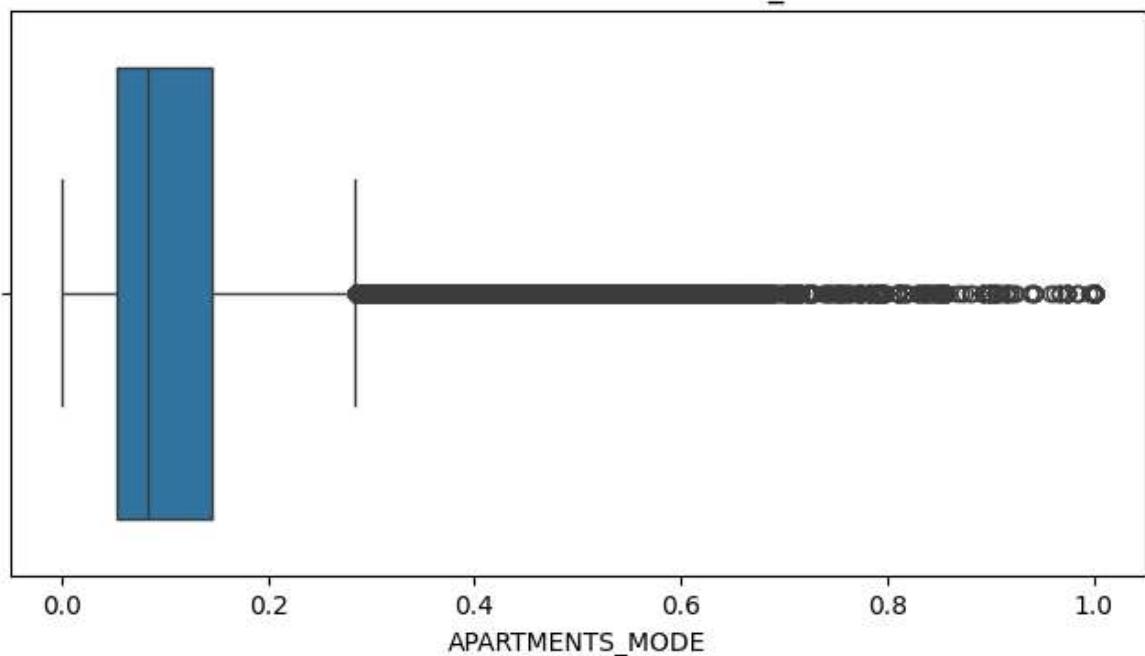
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna NONLIVINGAREA\_AVG



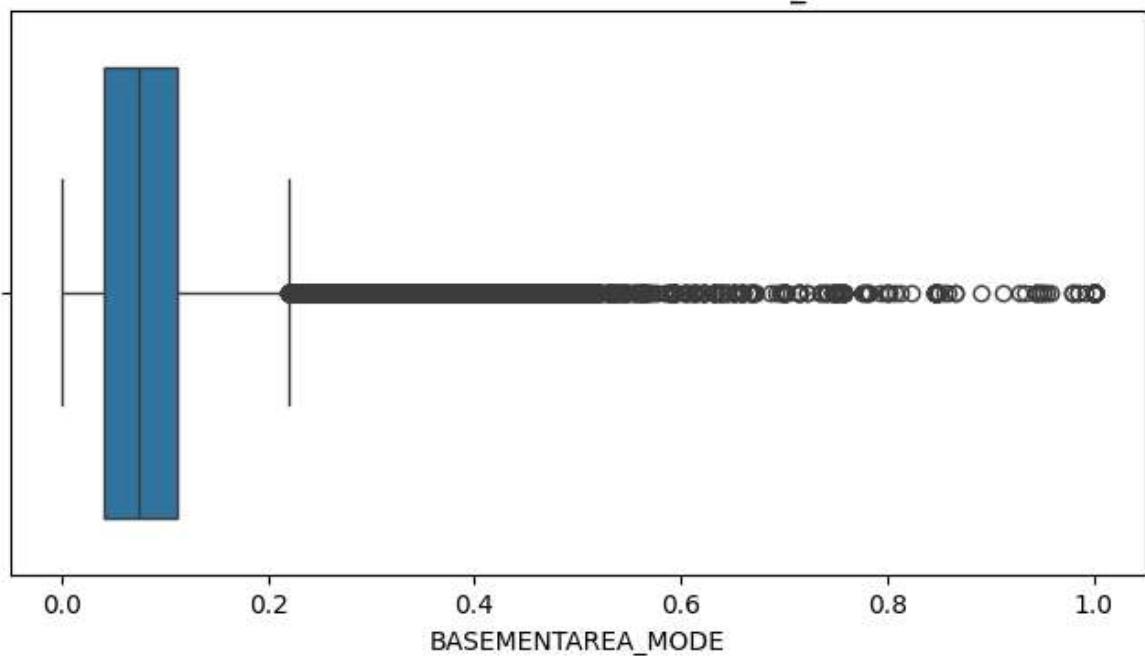
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna APARTMENTS\_MODE



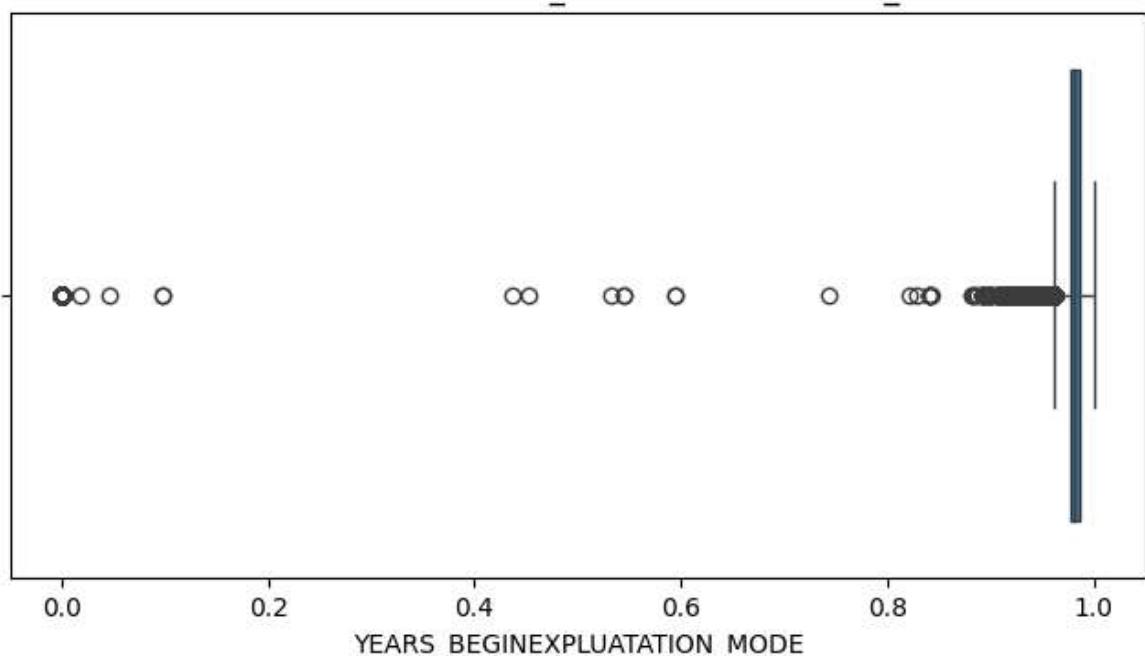
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\n\b\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna BASEMENTAREA\_MODE



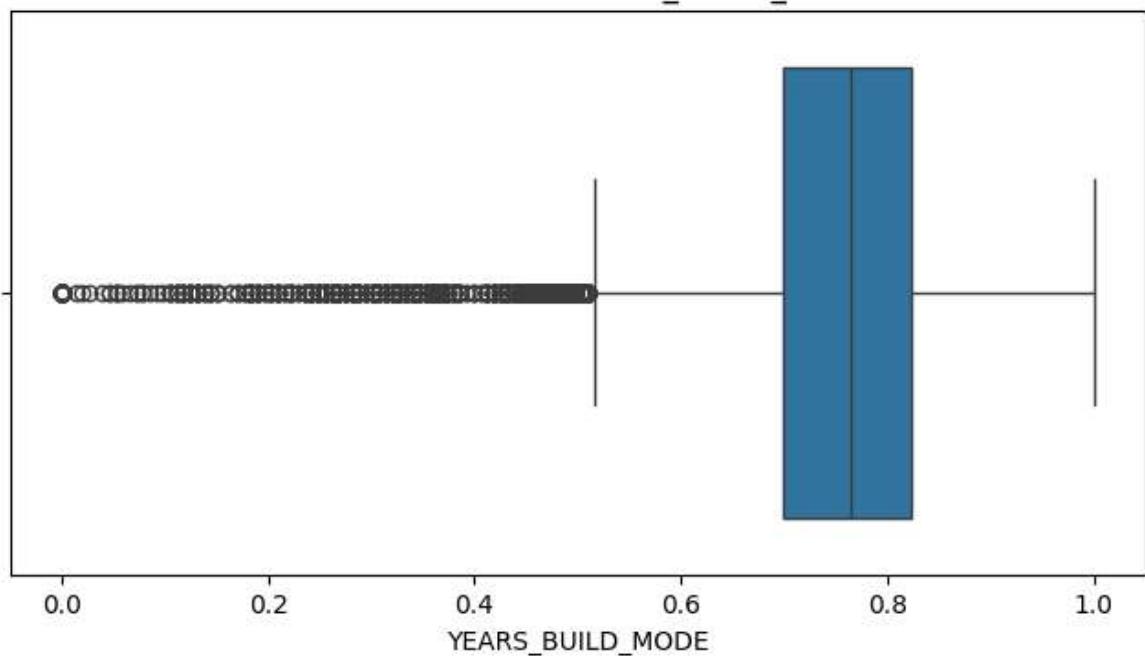
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\n\b\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna YEARS\_BEGINEXPLUATATION\_MODE



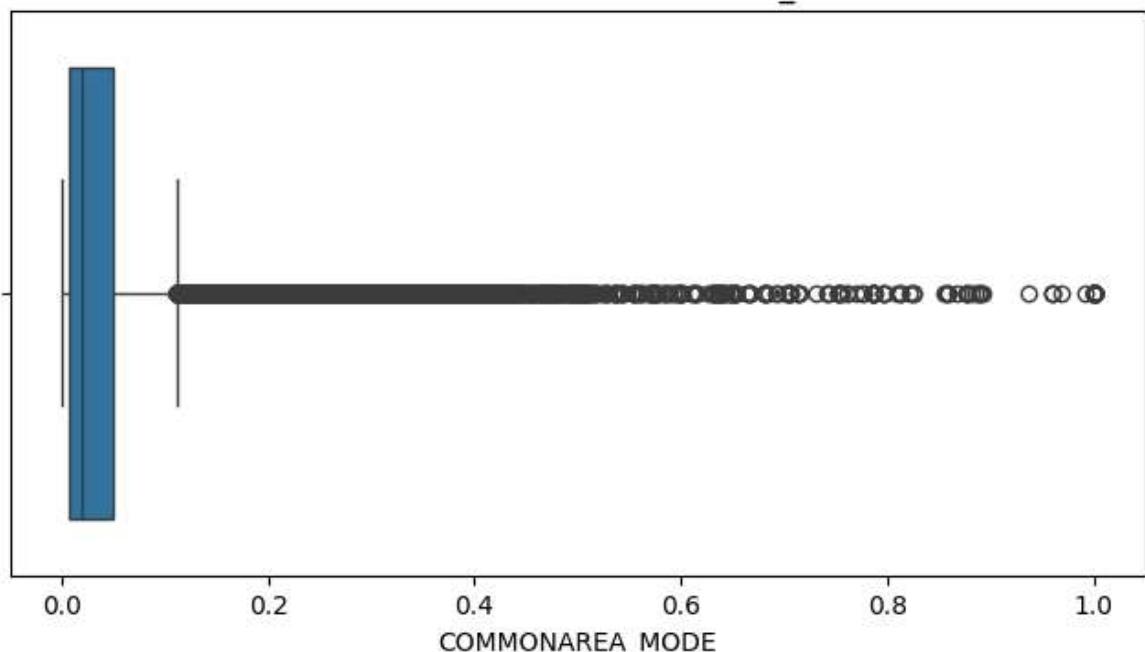
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna YEARS\_BUILD\_MODE



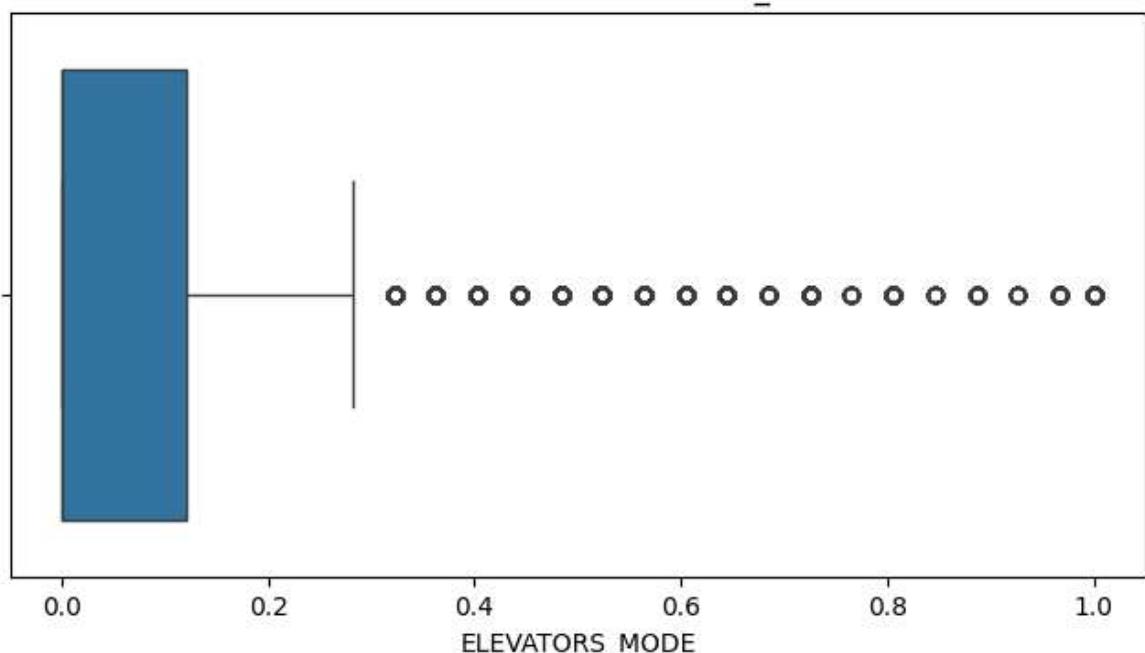
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna COMMONAREA\_MODE



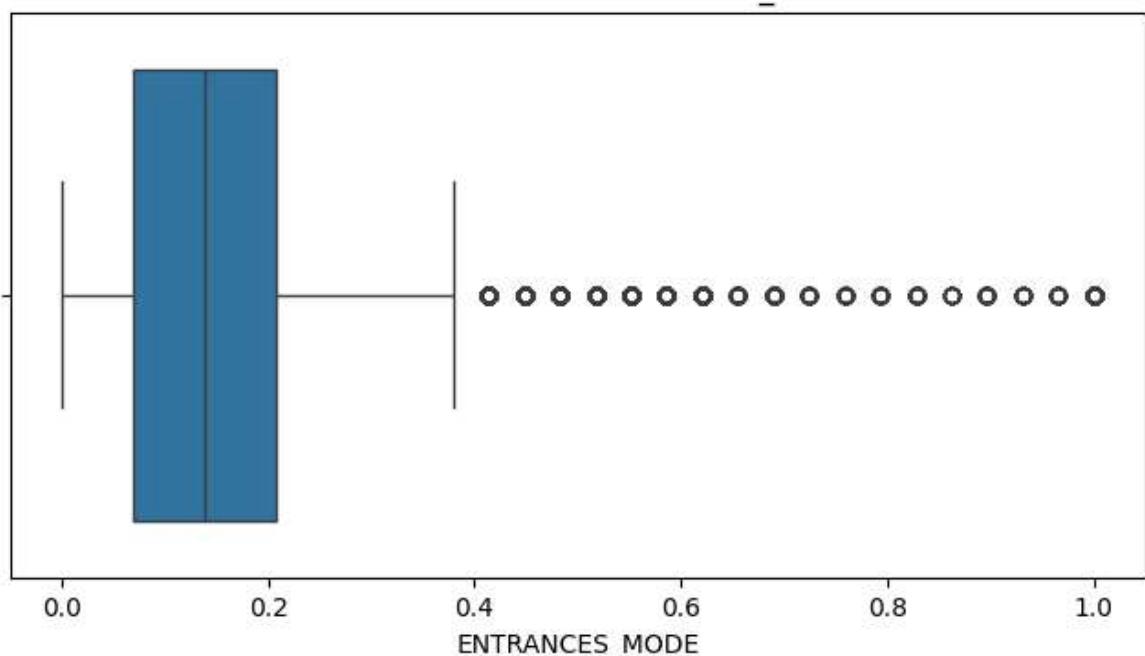
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna ELEVATORS\_MODE



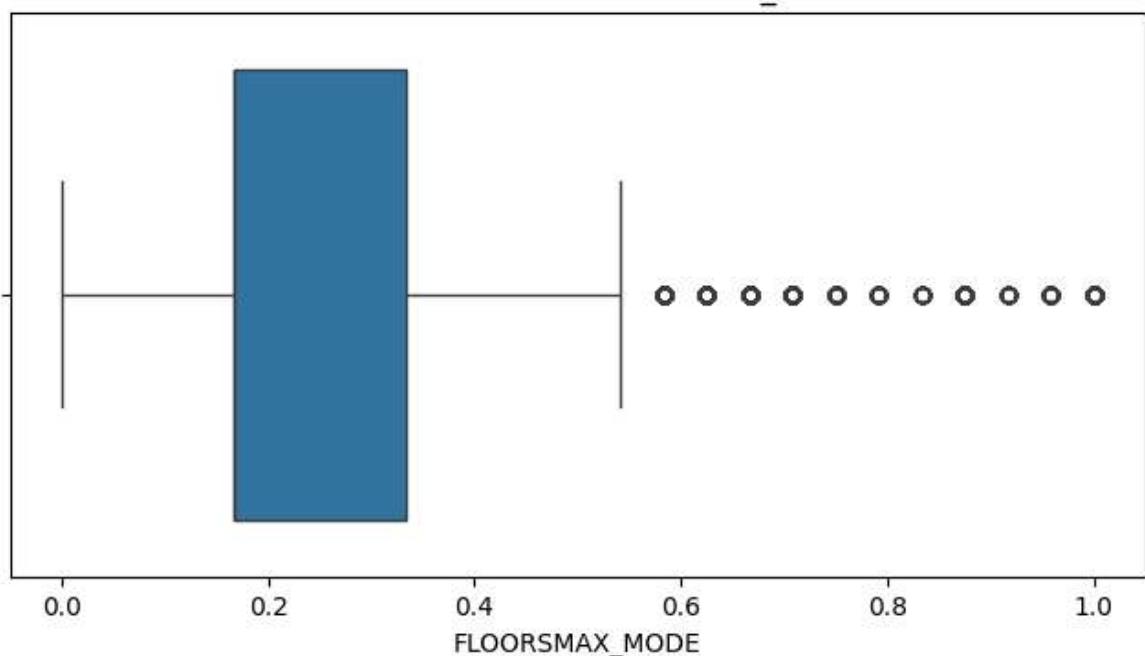
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna ENTRANCES\_MODE



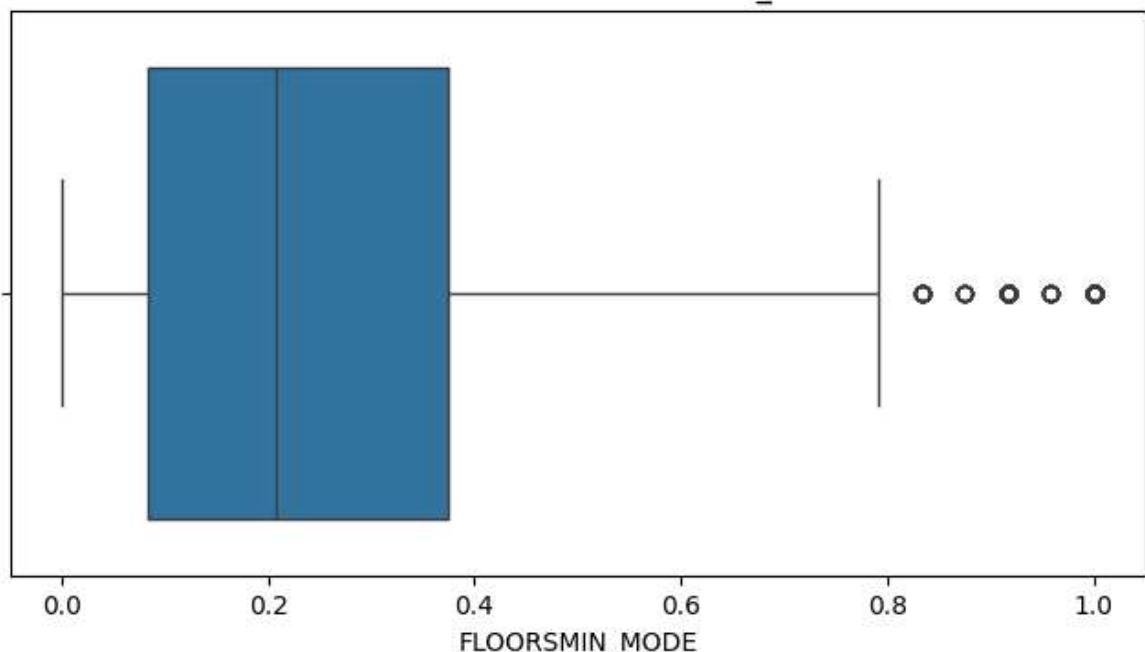
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna FLOORSMAX\_MODE



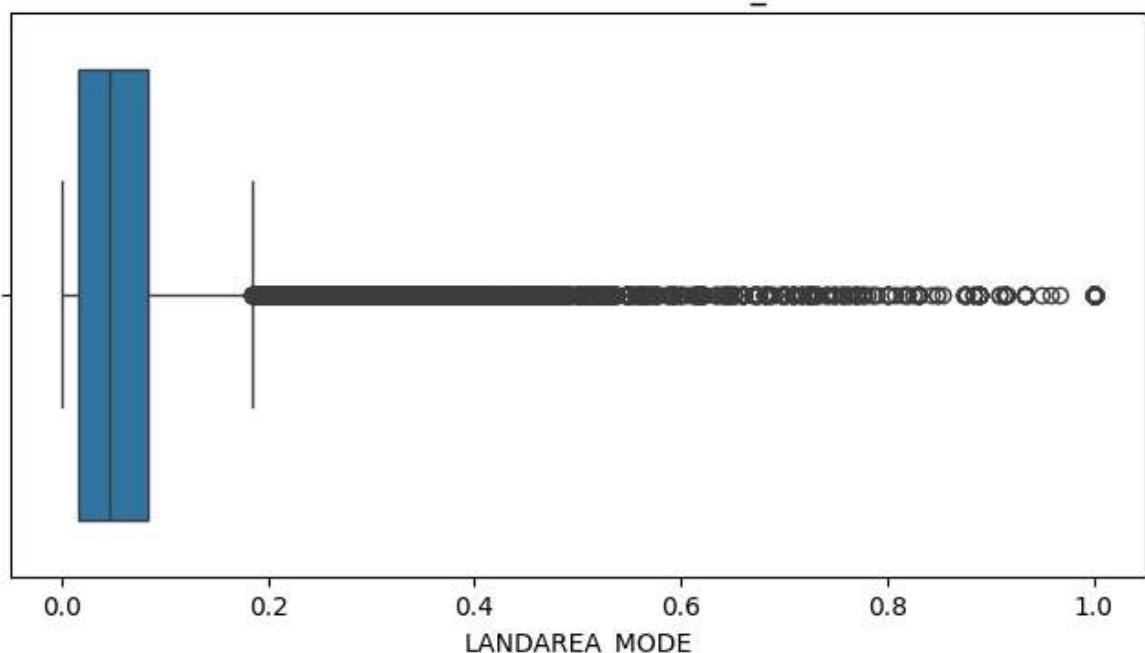
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna FLOORSMIN\_MODE



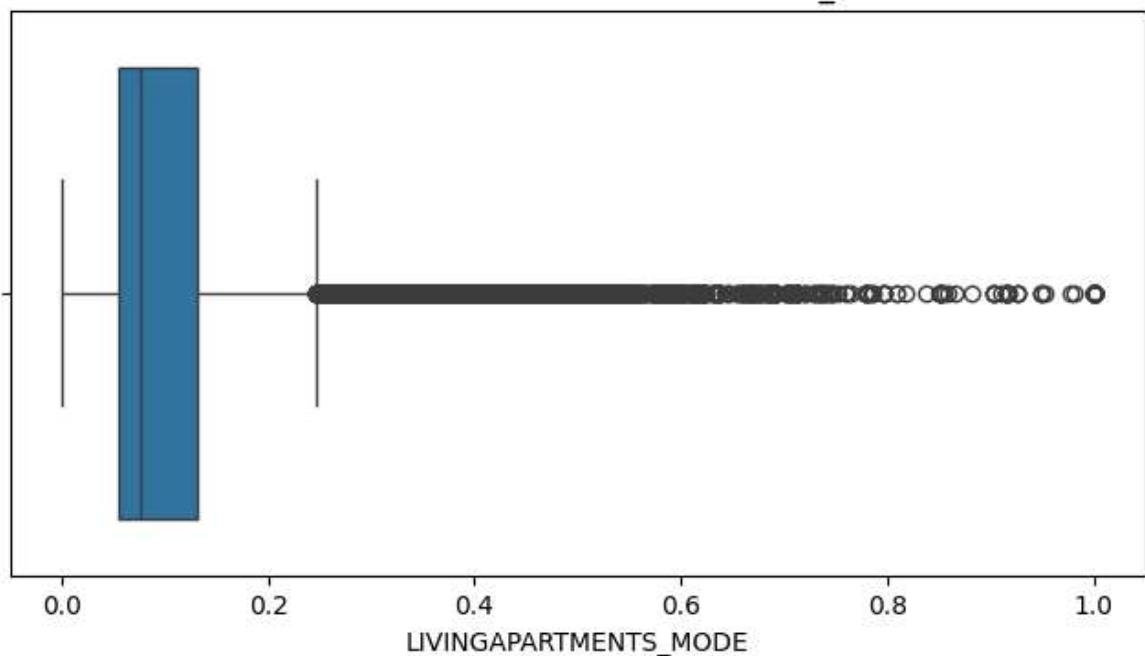
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna LANDAREA\_MODE



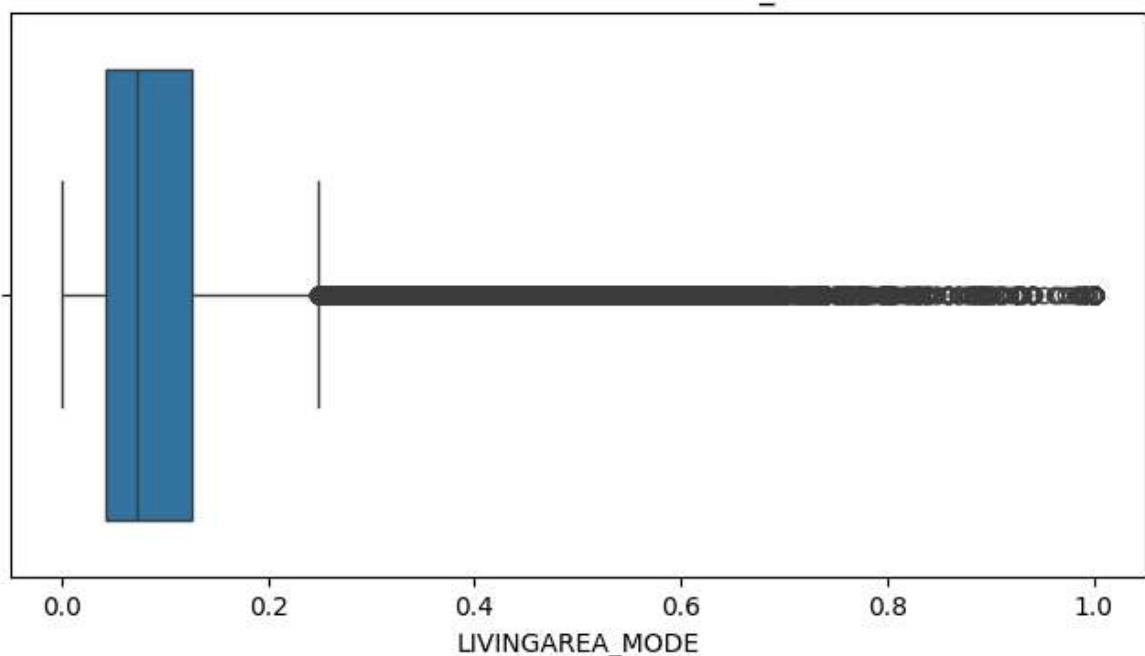
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna LIVINGAPARTMENTS\_MODE



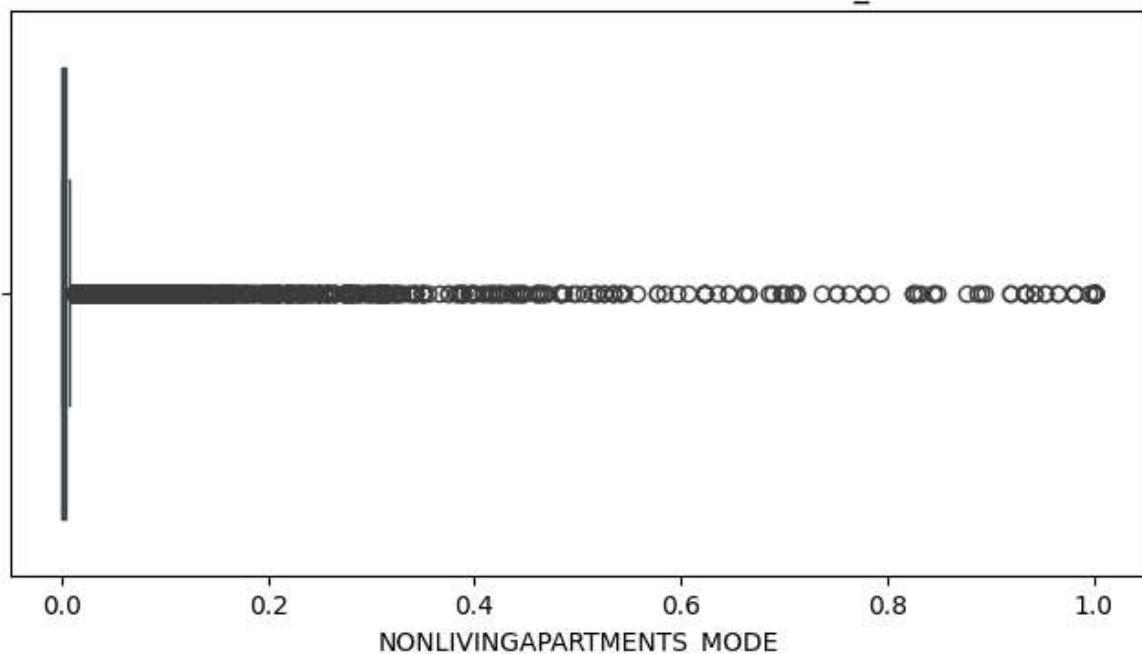
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna LIVINGAREA\_MODE



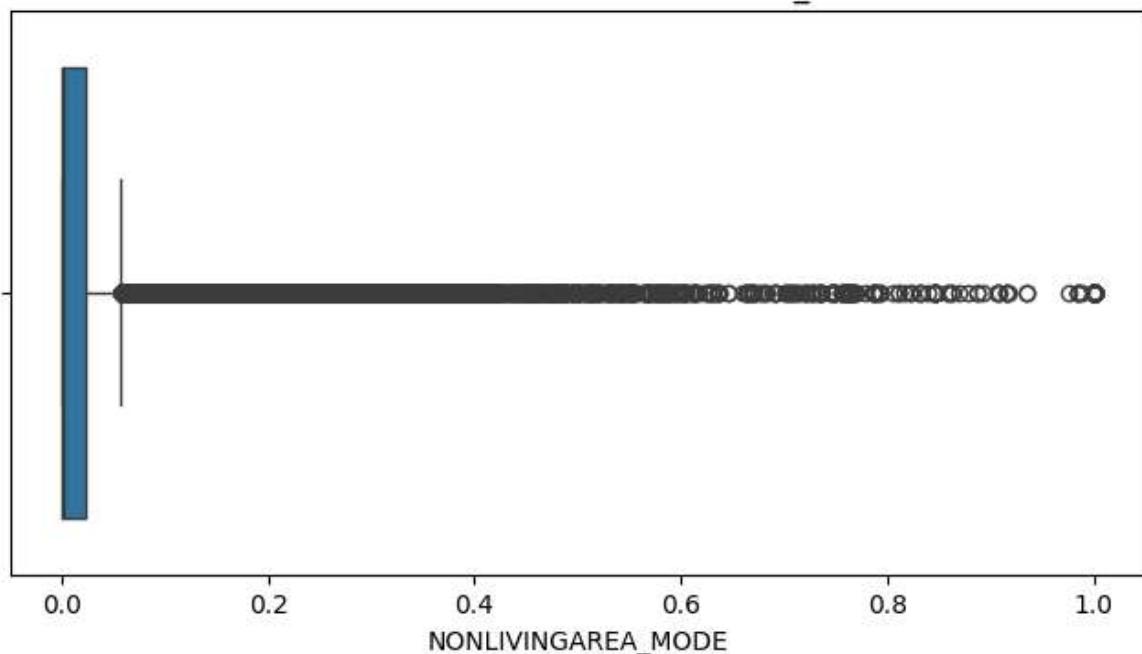
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna NONLIVINGAPARTMENTS\_MODE



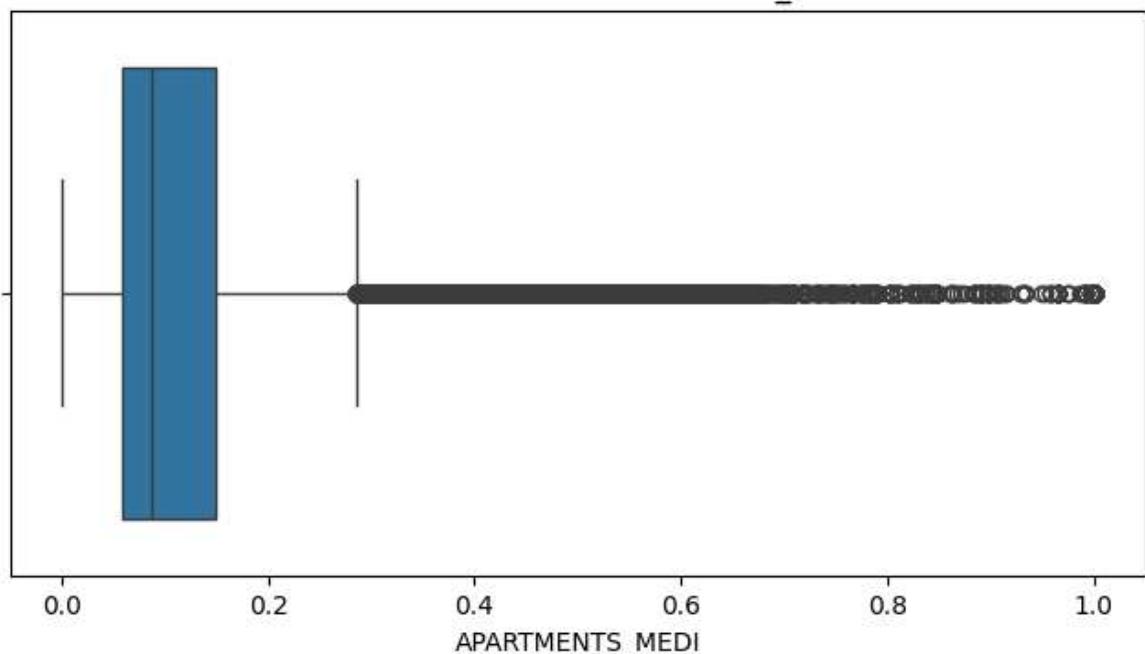
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna NONLIVINGAREA\_MODE



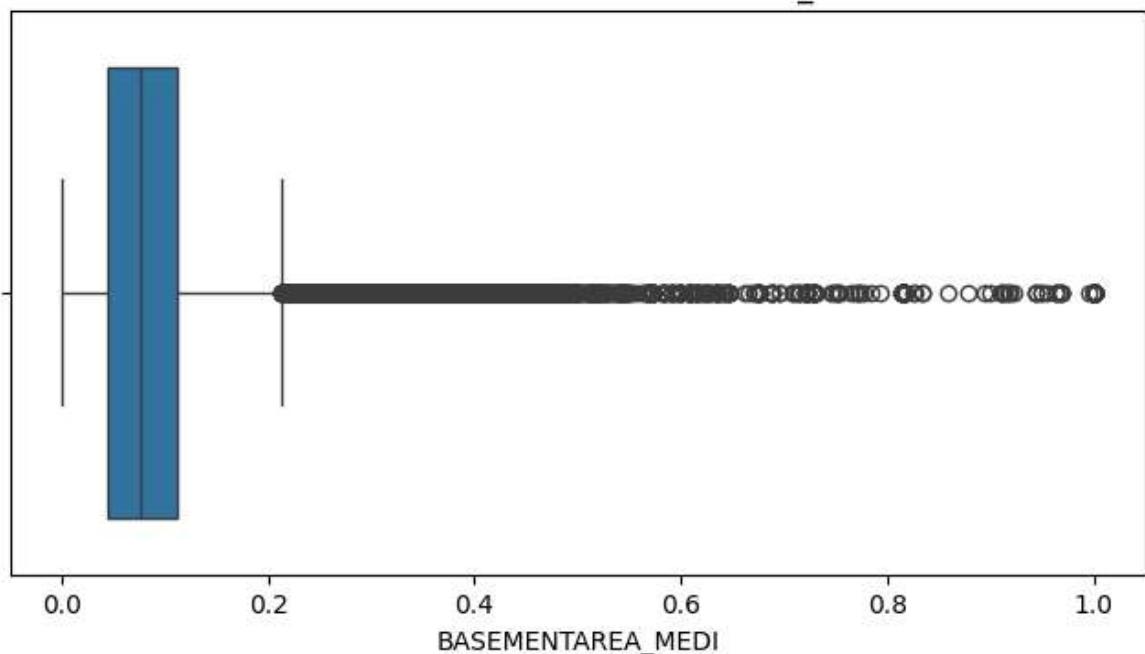
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna APARTMENTS\_MEDI



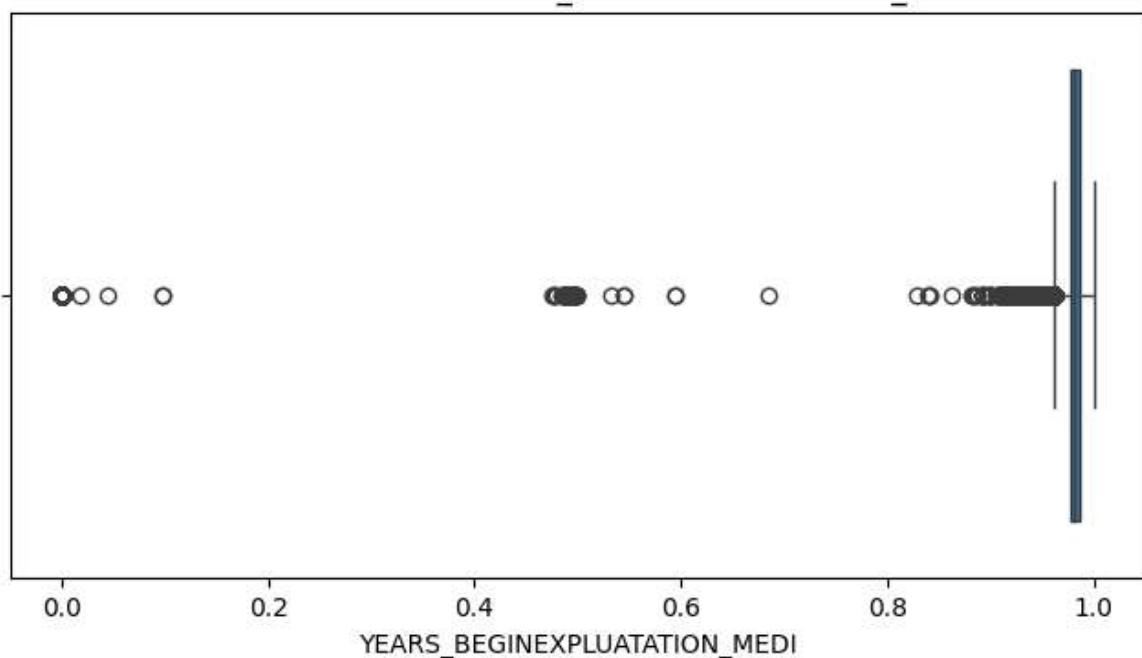
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna BASEMENTAREA\_MEDI



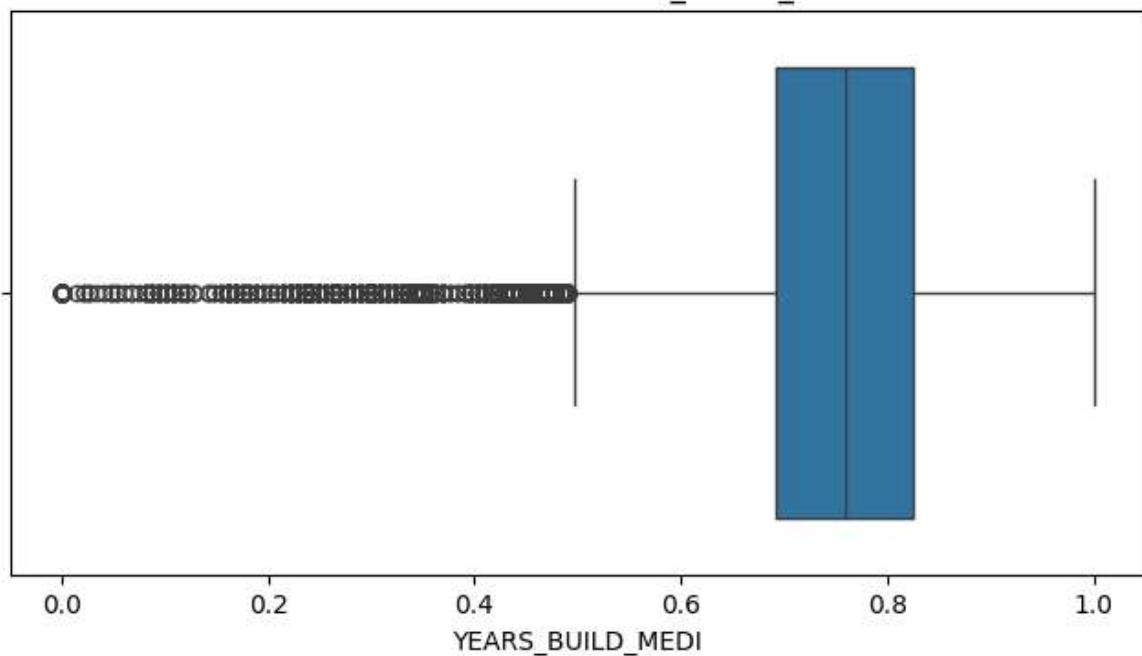
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna YEARS\_BEGINEXPLUATATION\_MEDI



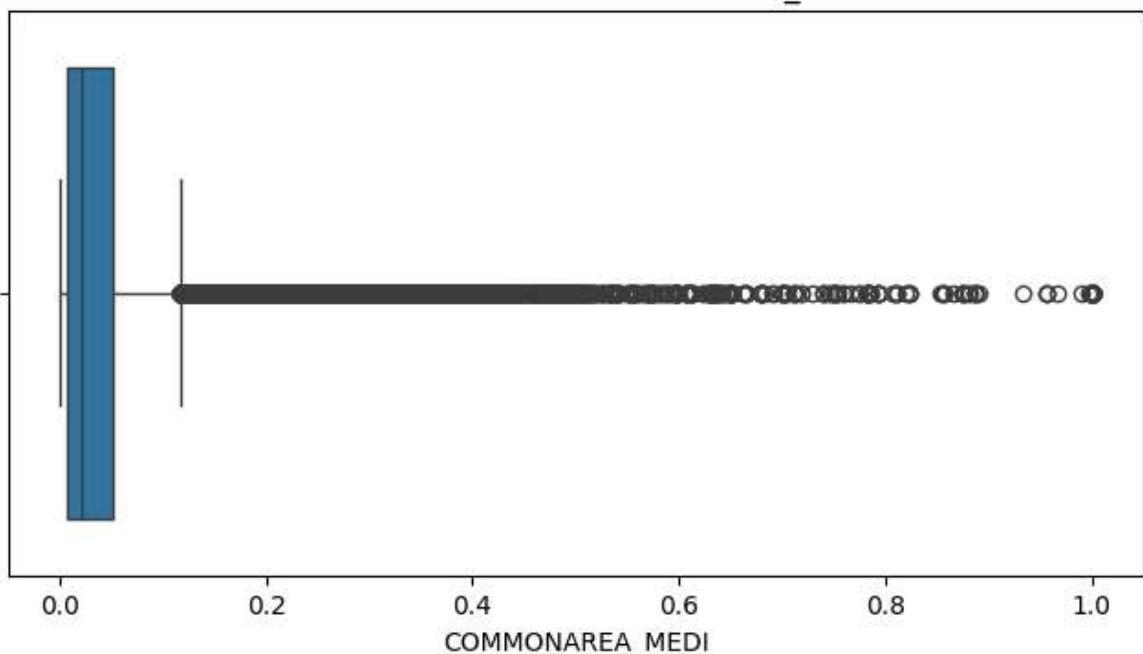
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna YEARS\_BUILD\_MEDI



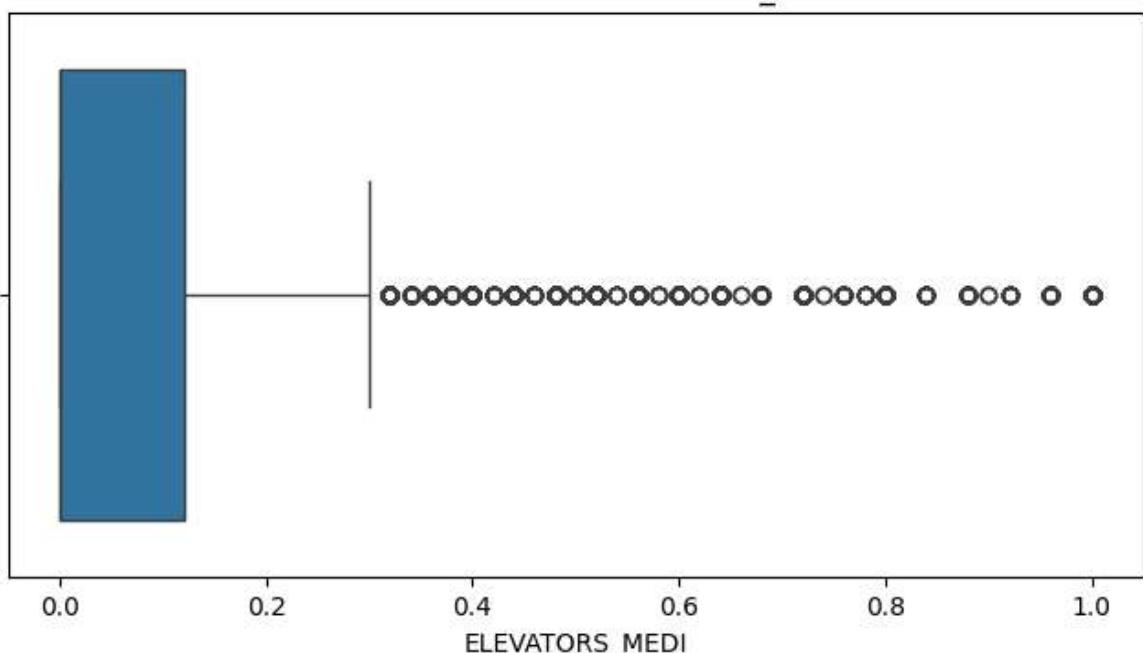
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna COMMONAREA\_MEDI



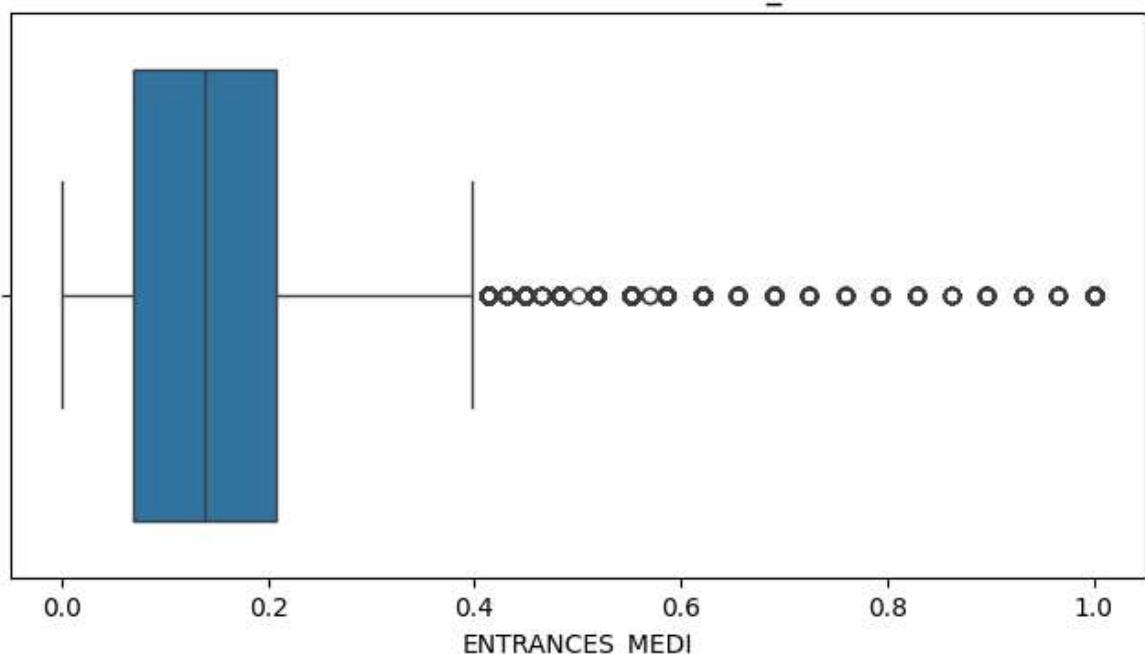
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna ELEVATORS\_MEDI



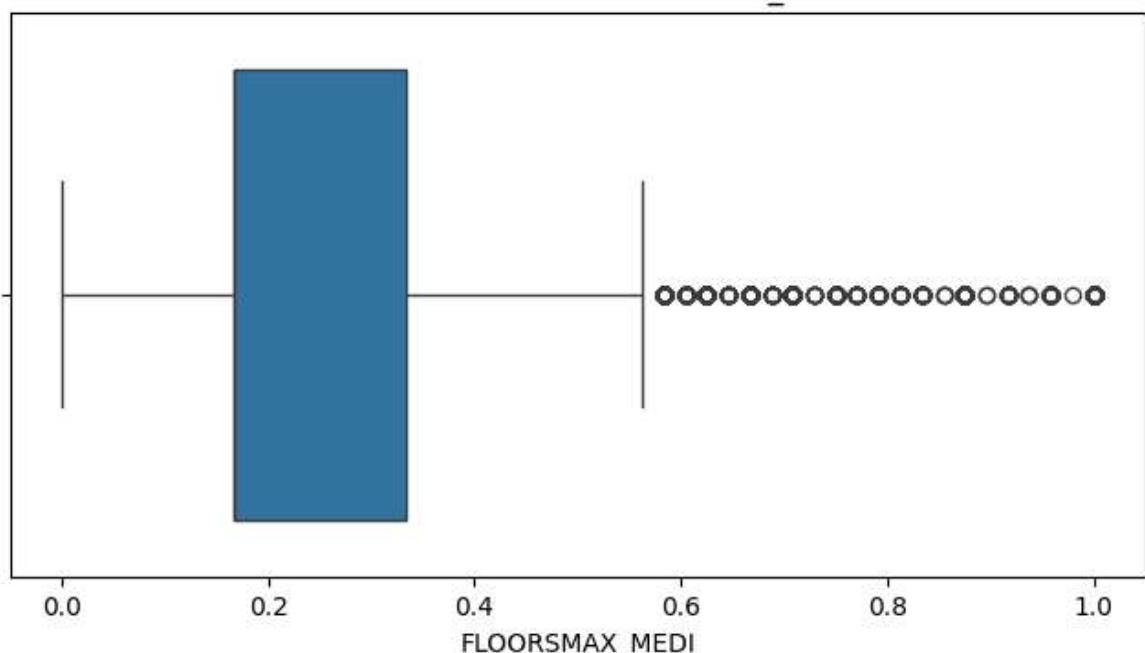
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna ENTRANCES\_MEDI



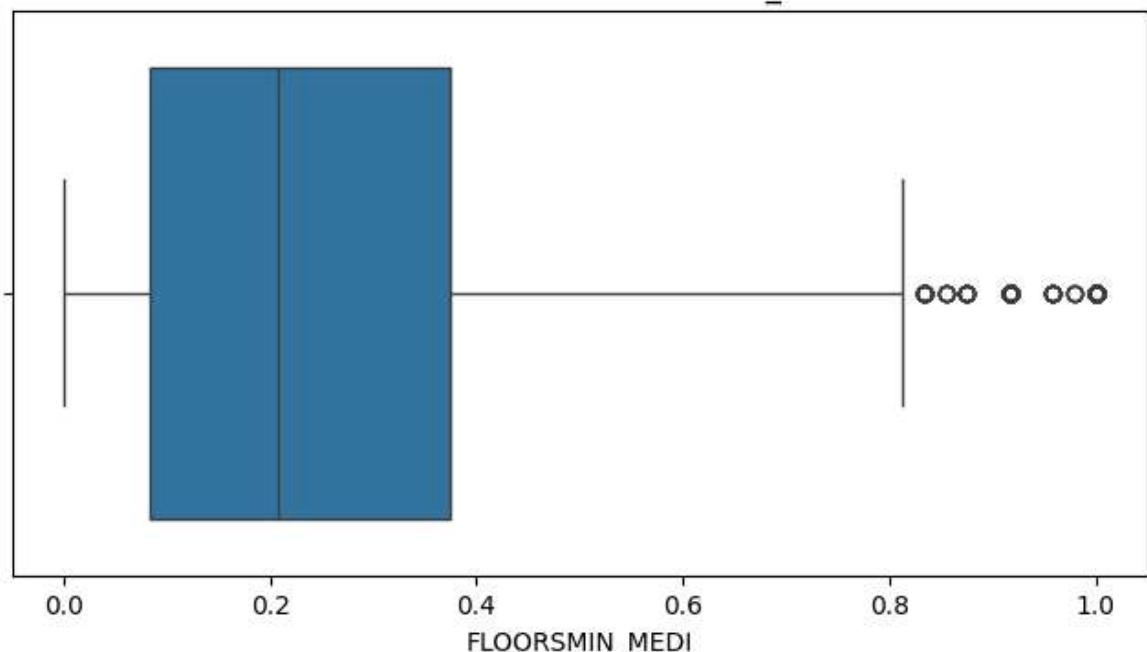
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna FLOORSMAX\_MEDI



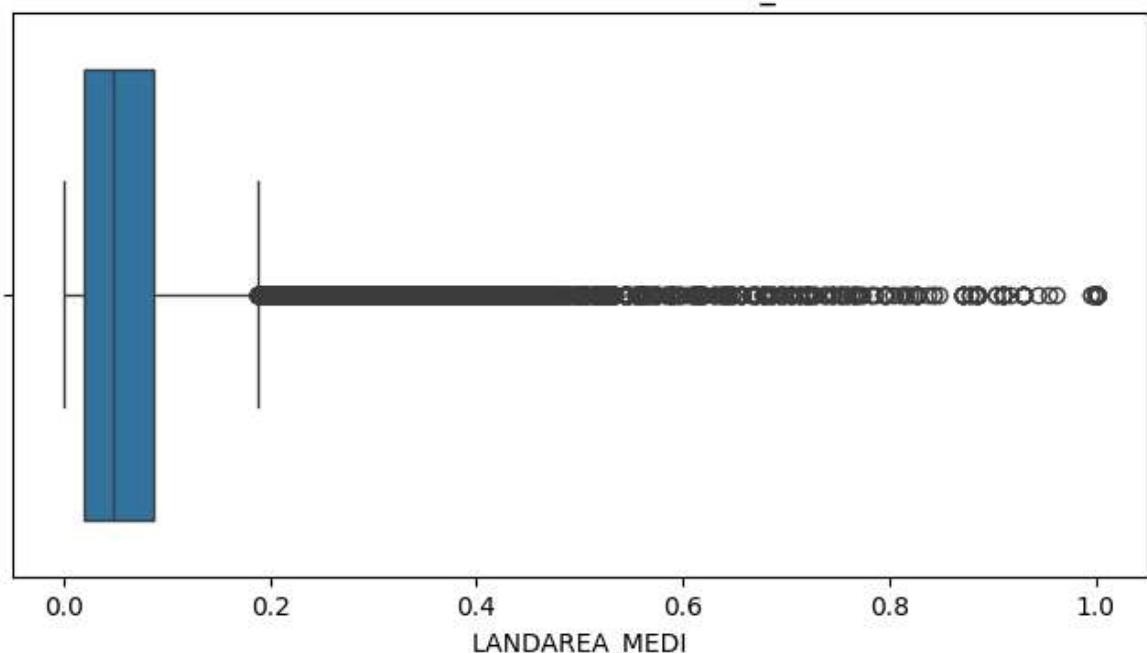
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna FLOORSMIN\_MEDI



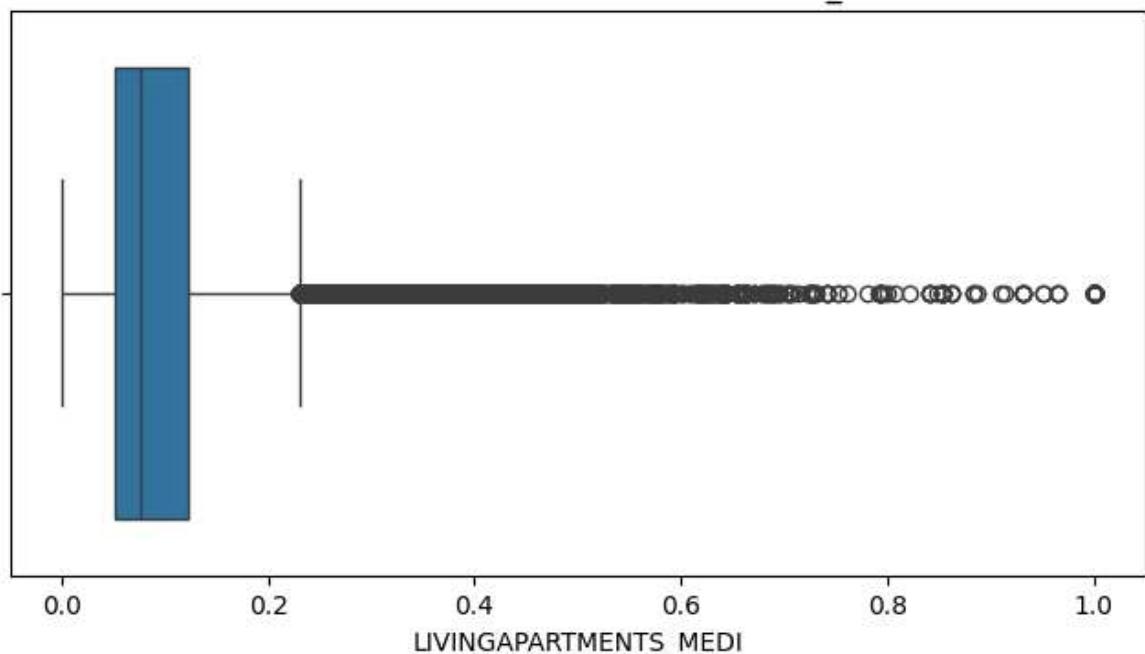
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna LANDAREA\_MEDI



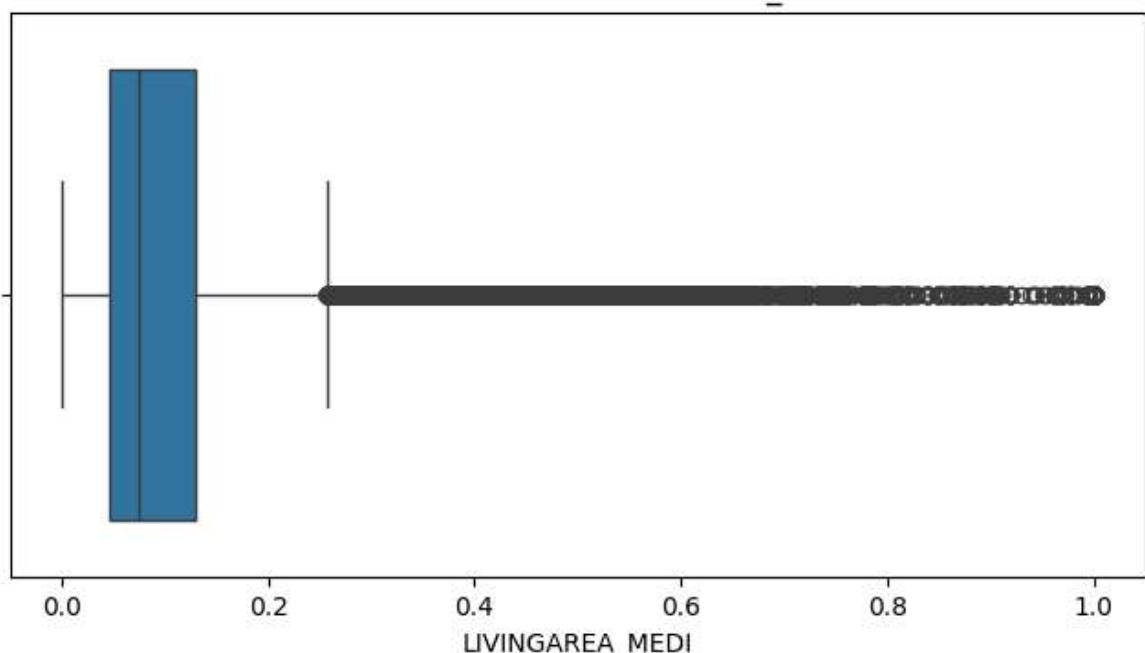
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna LIVINGAPARTMENTS\_MEDI



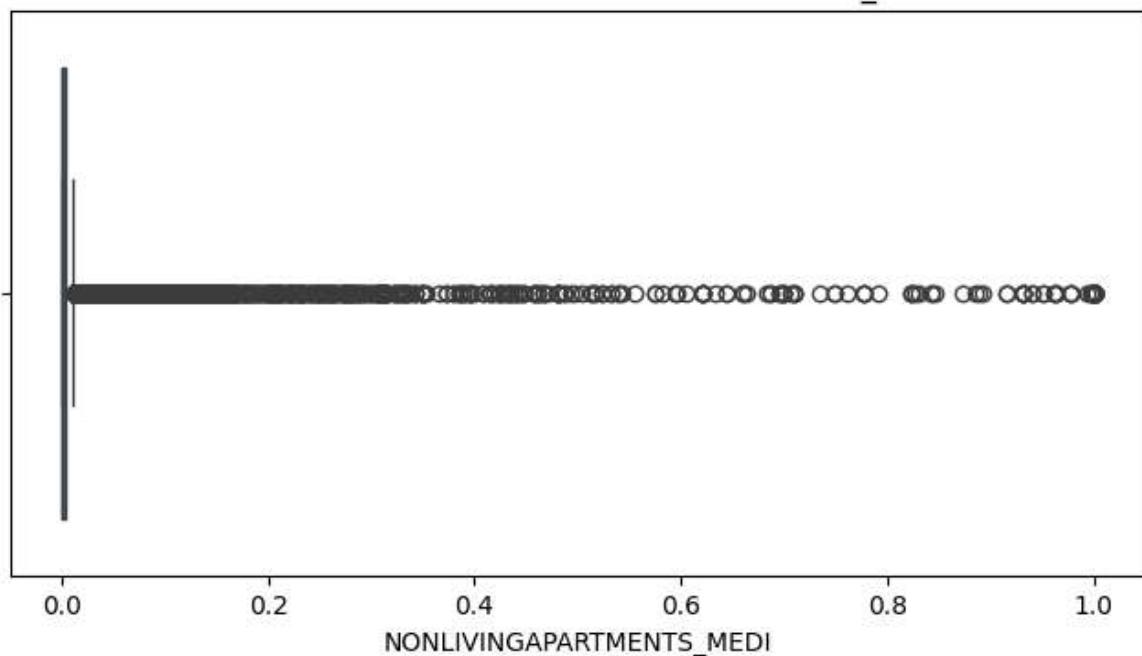
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna LIVINGAREA\_MEDI



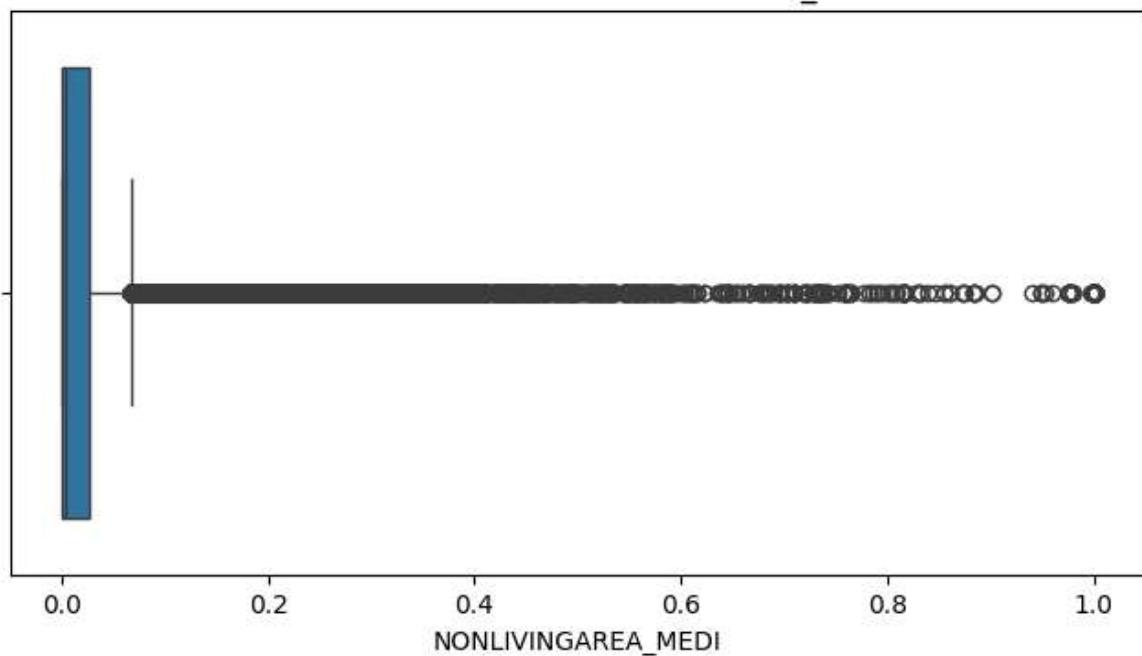
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna NONLIVINGAPARTMENTS\_MEDI



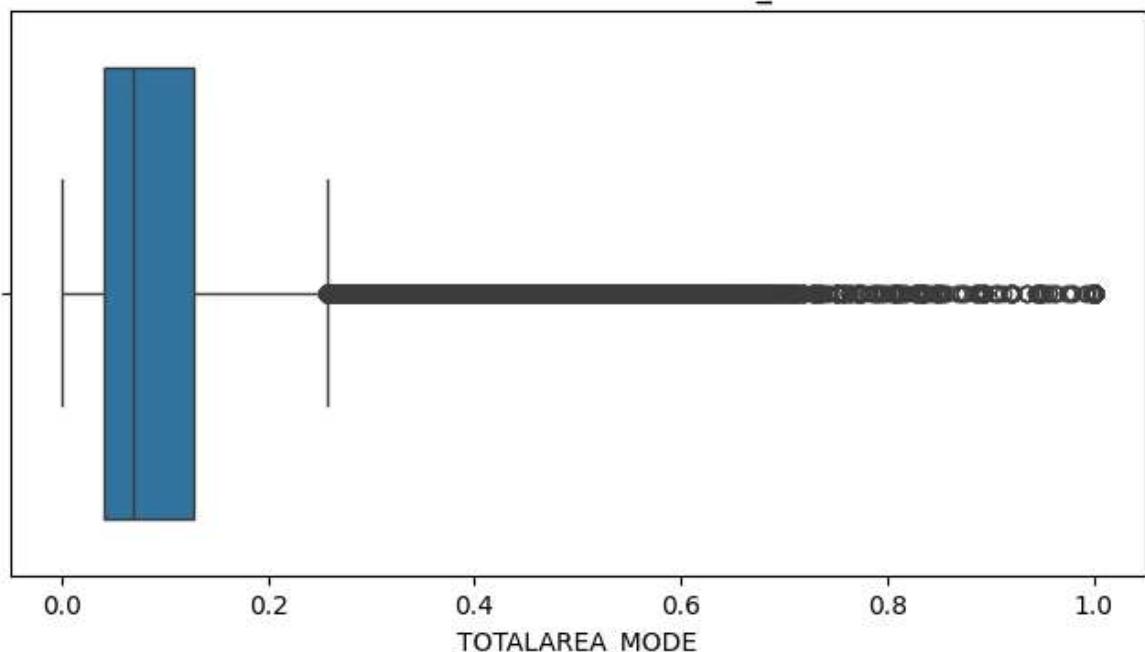
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna NONLIVINGAREA\_MEDI



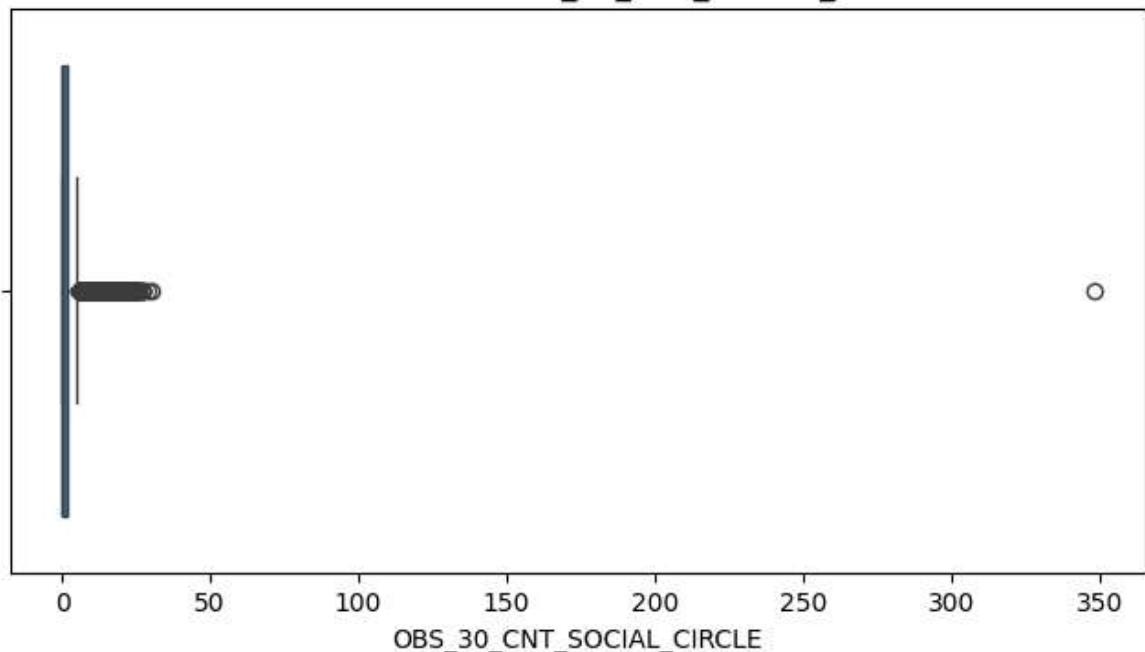
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna TOTALAREA\_MODE



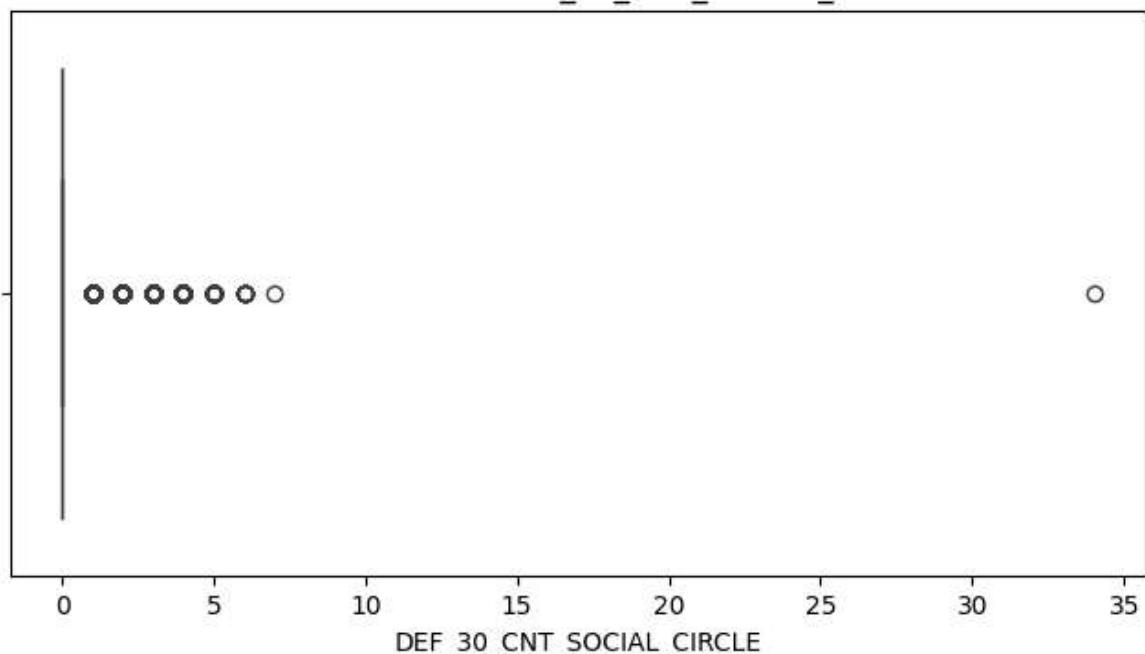
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna OBS\_30\_CNT\_SOCIAL\_CIRCLE



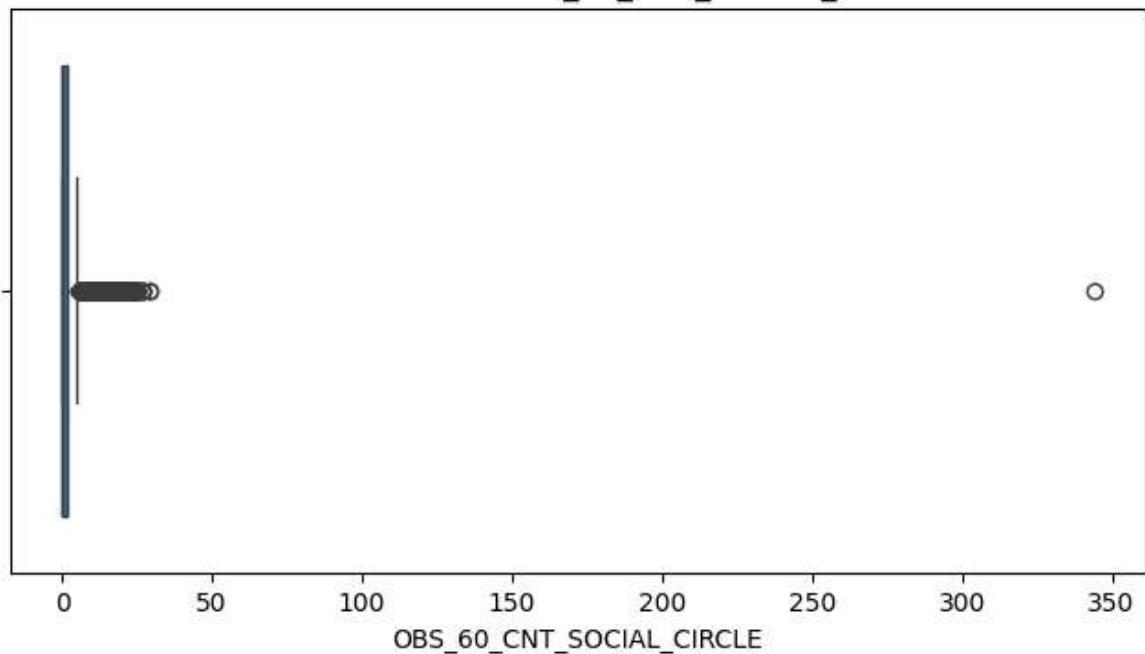
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna DEF\_30\_CNT\_SOCIAL\_CIRCLE



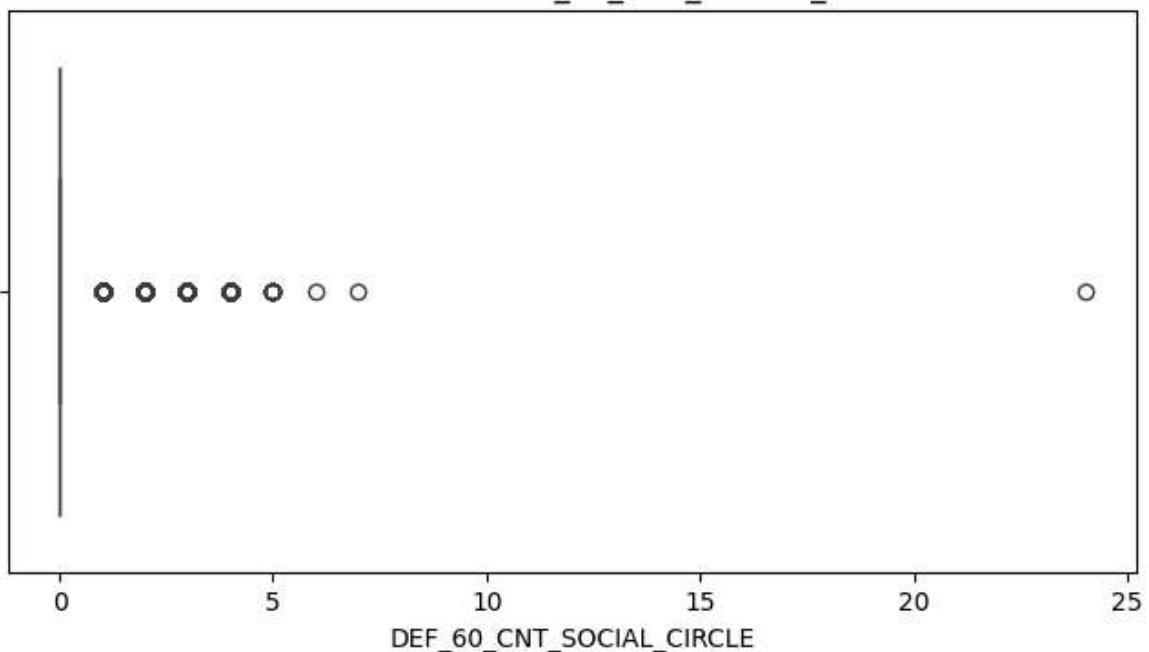
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna OBS\_60\_CNT\_SOCIAL\_CIRCLE



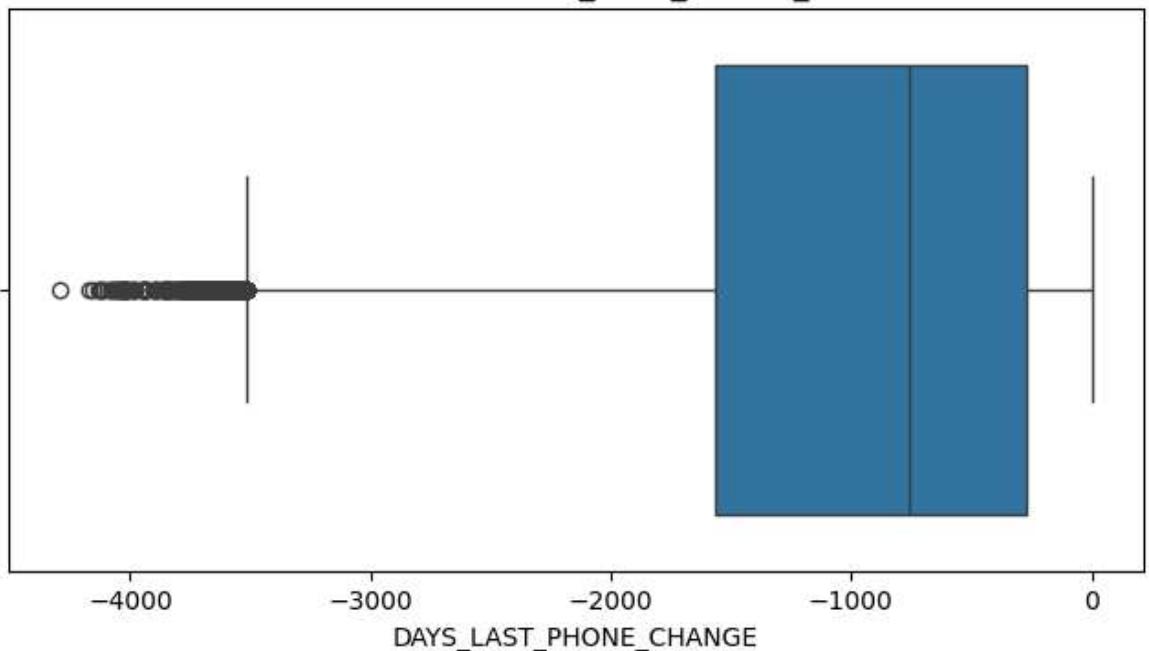
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna DEF\_60\_CNT\_SOCIAL\_CIRCLE



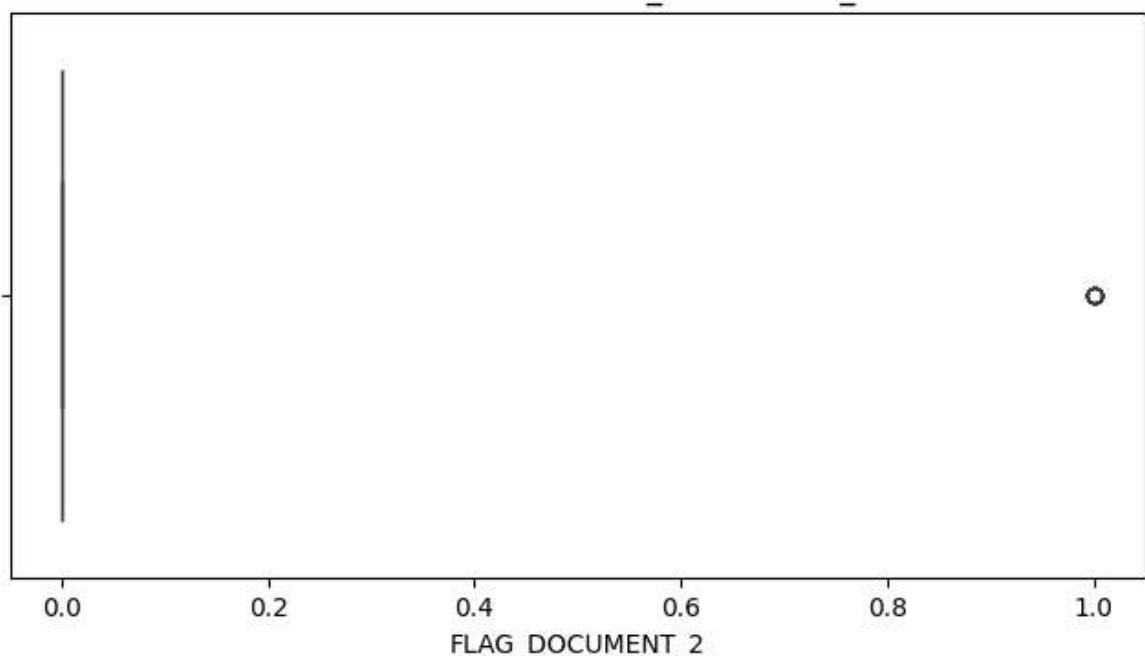
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna DAYS\_LAST\_PHONE\_CHANGE



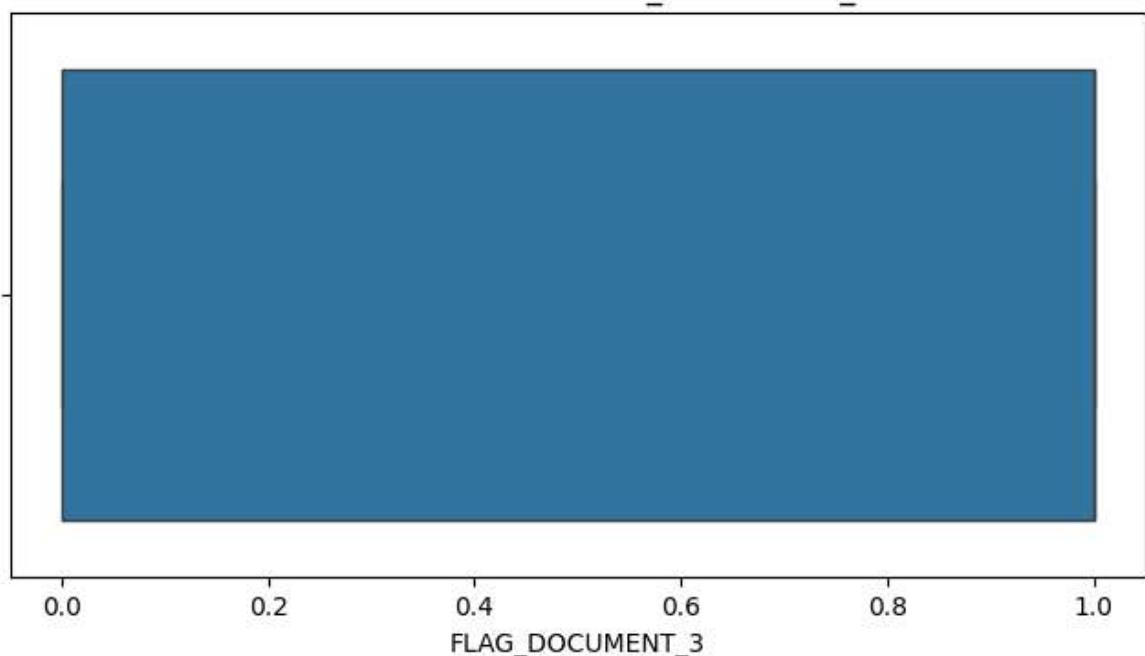
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_2



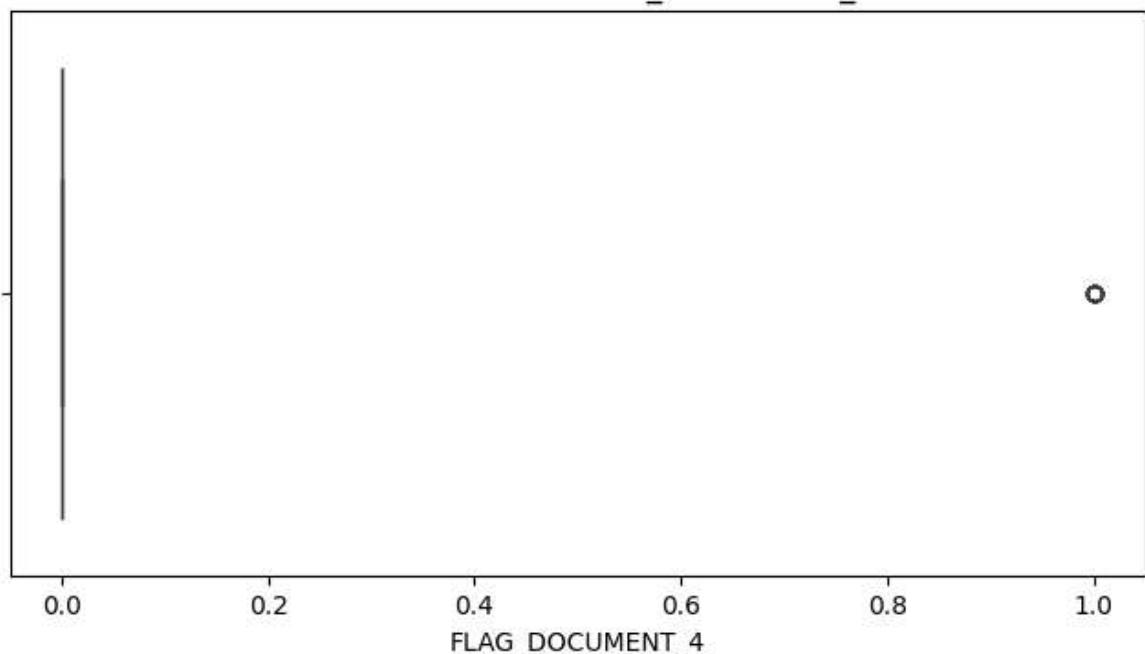
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\n\b\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_3



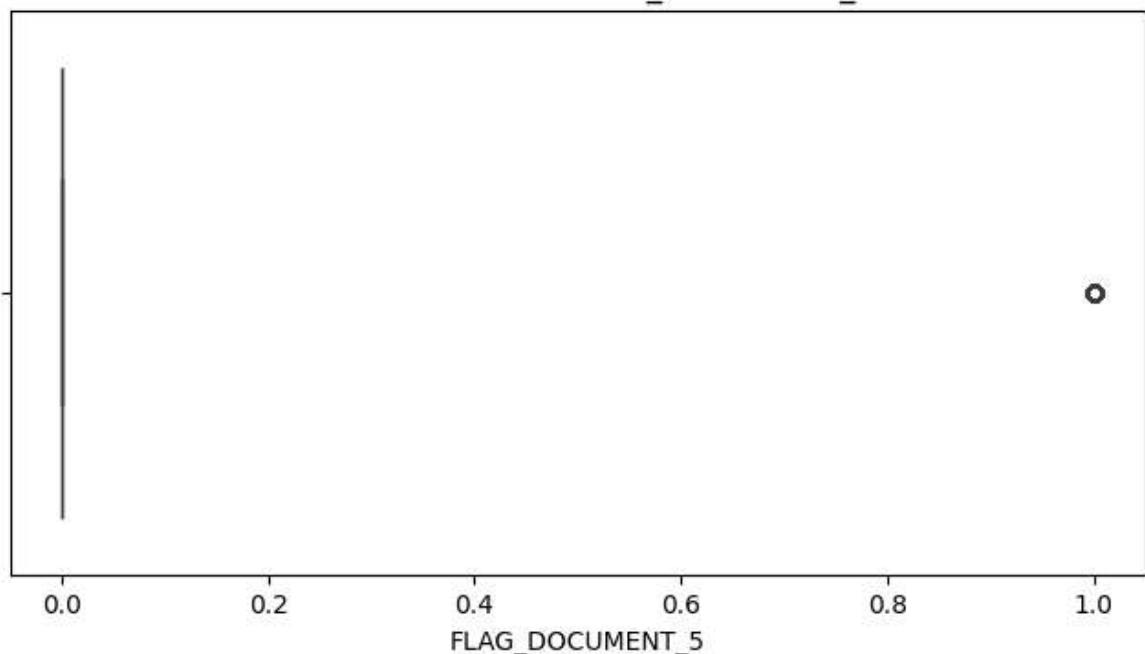
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\n\b\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_4



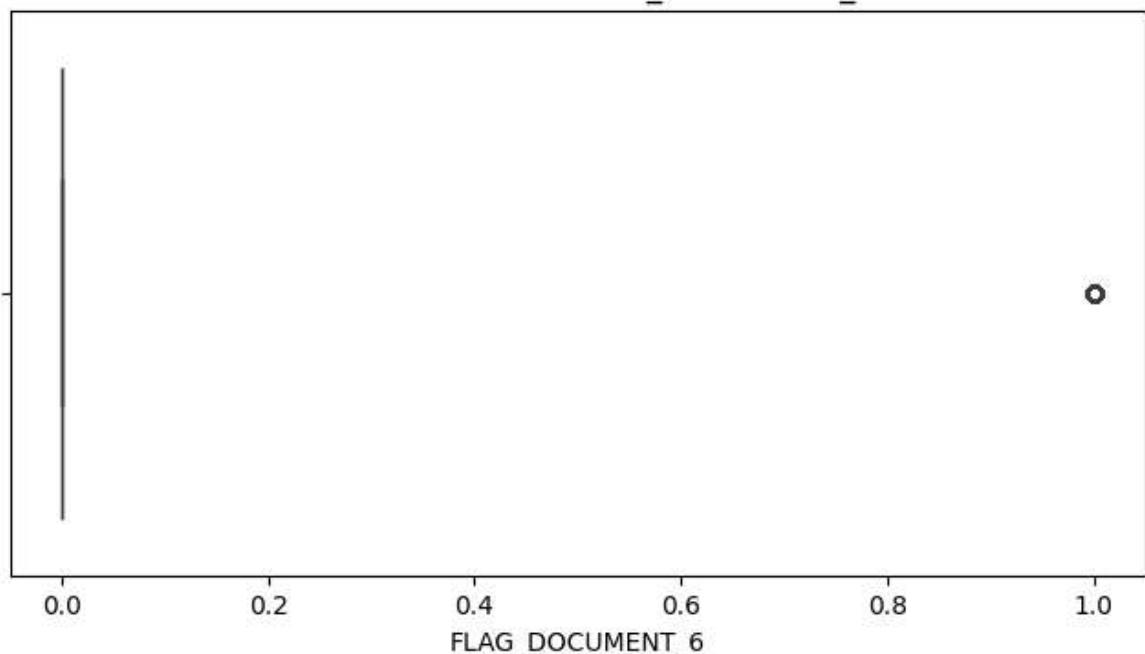
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_5



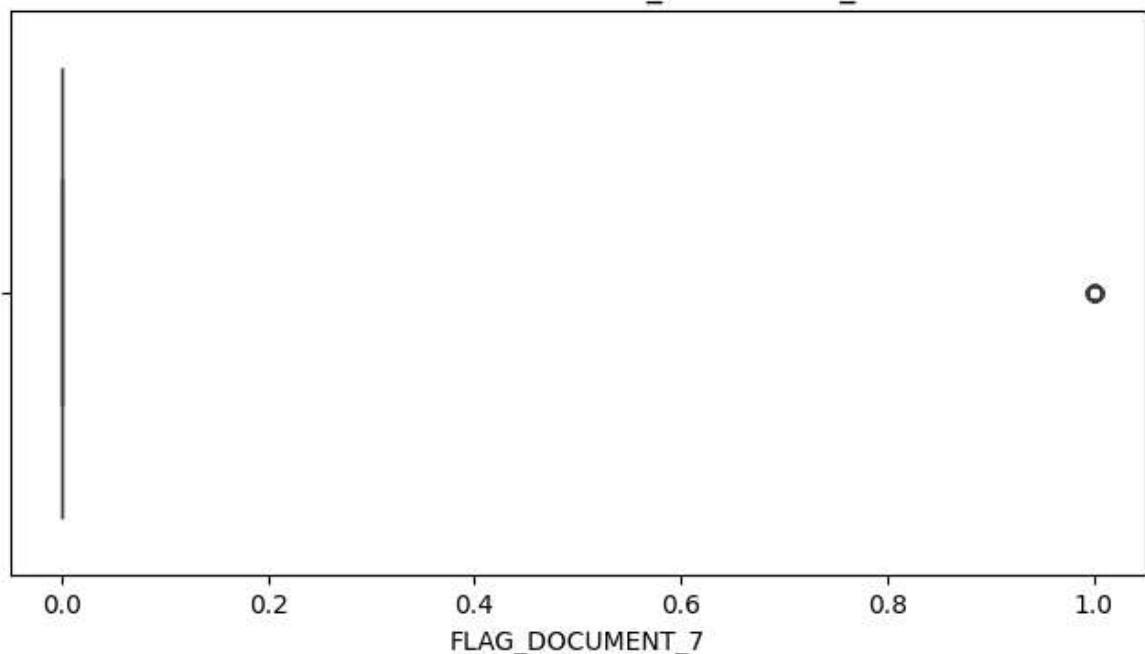
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_6



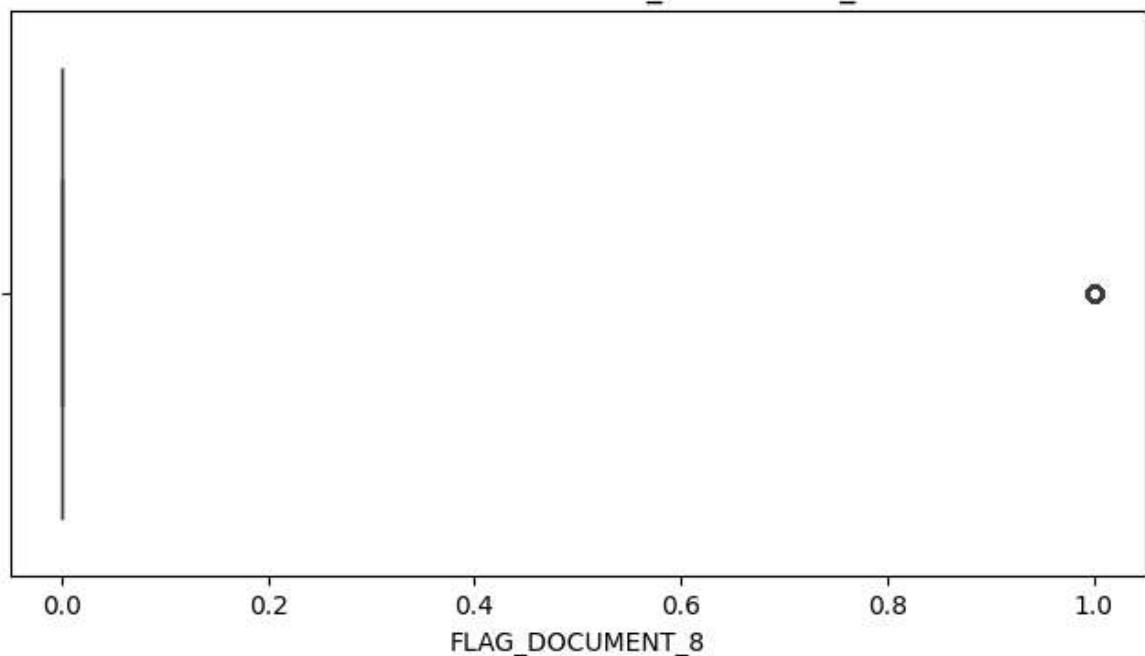
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_7



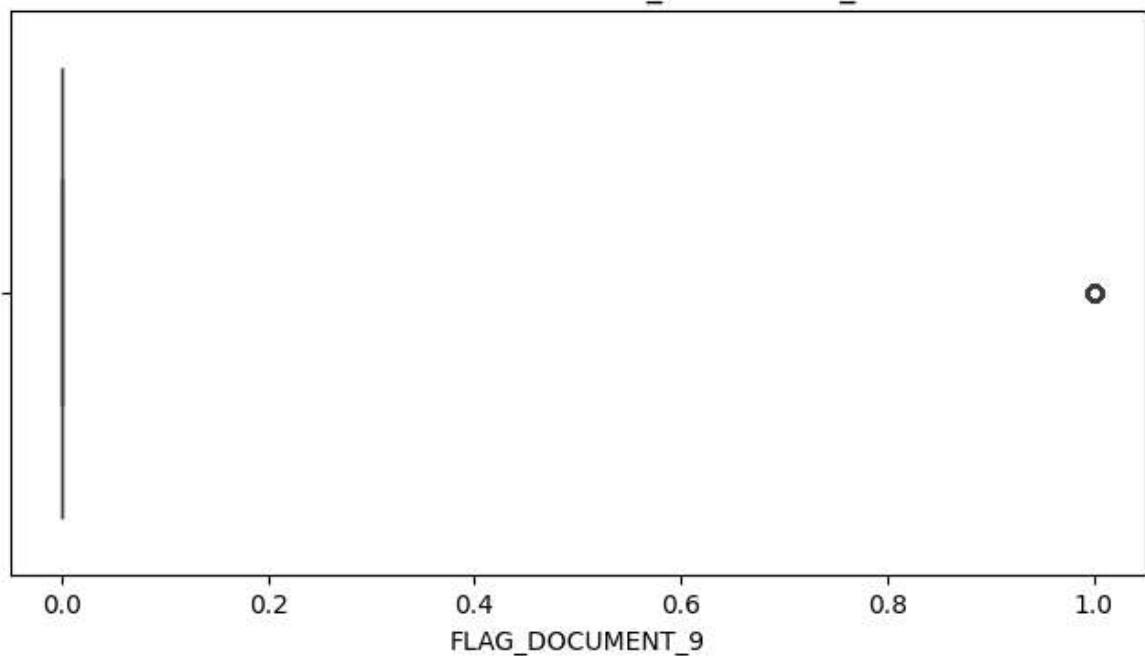
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_8



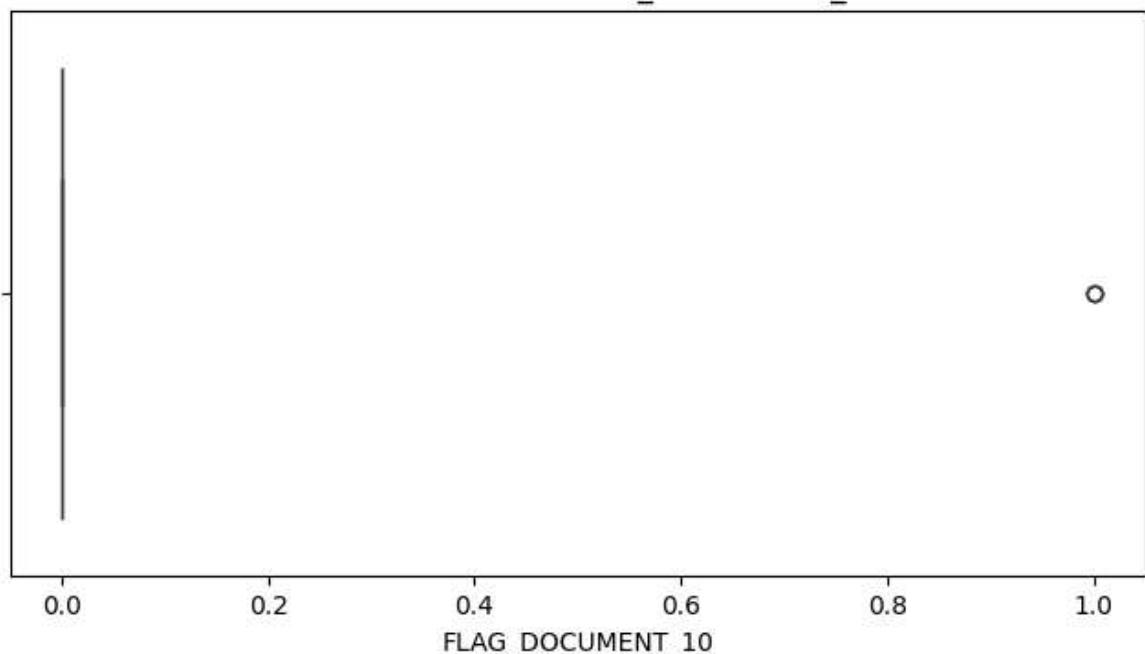
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_9



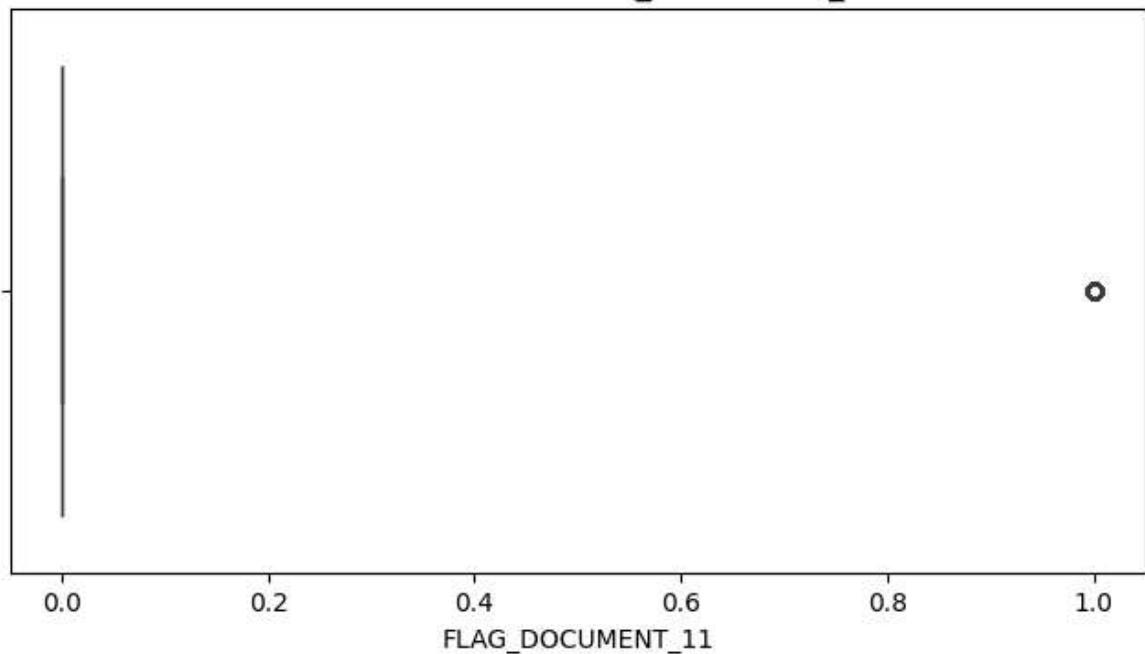
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_10



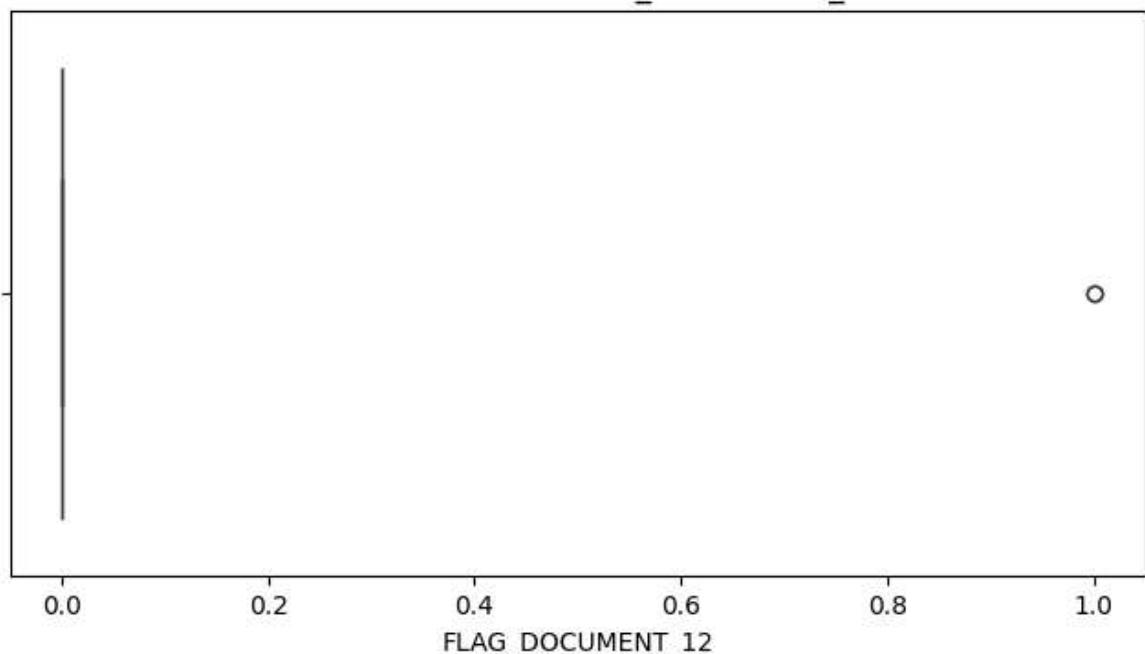
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_11



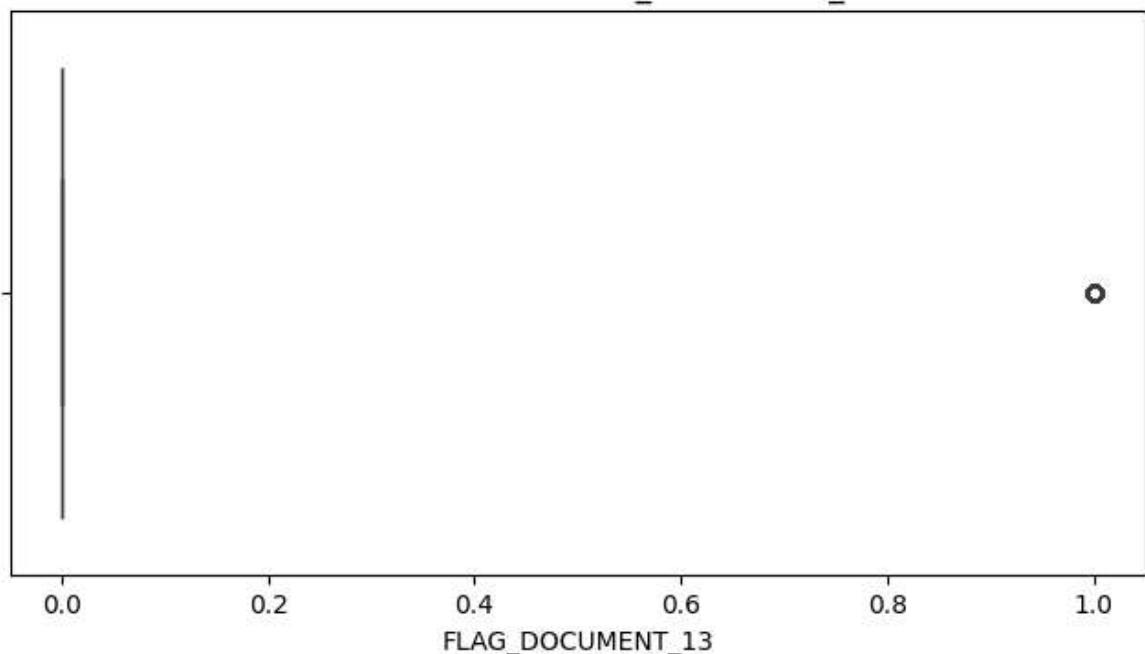
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_12



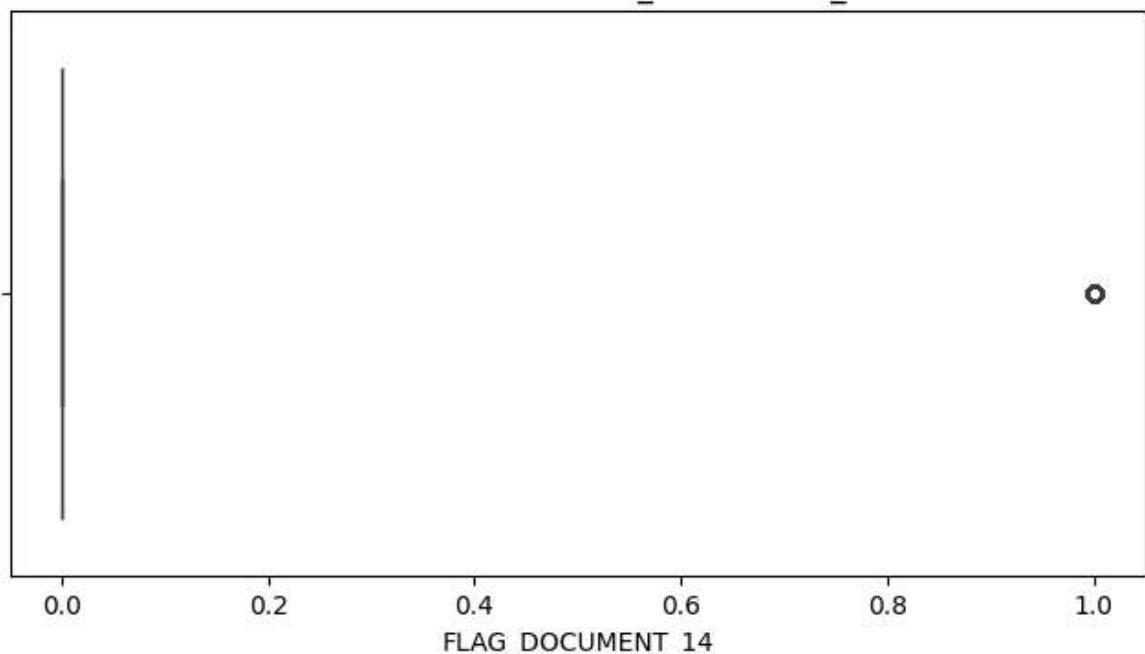
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_13



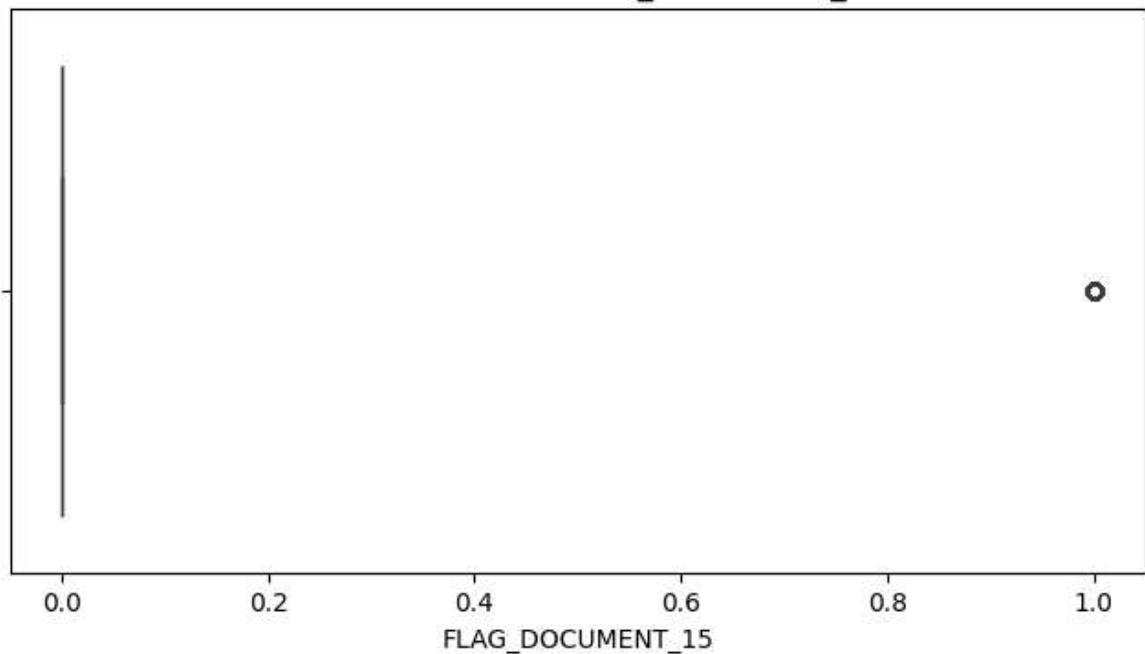
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_14



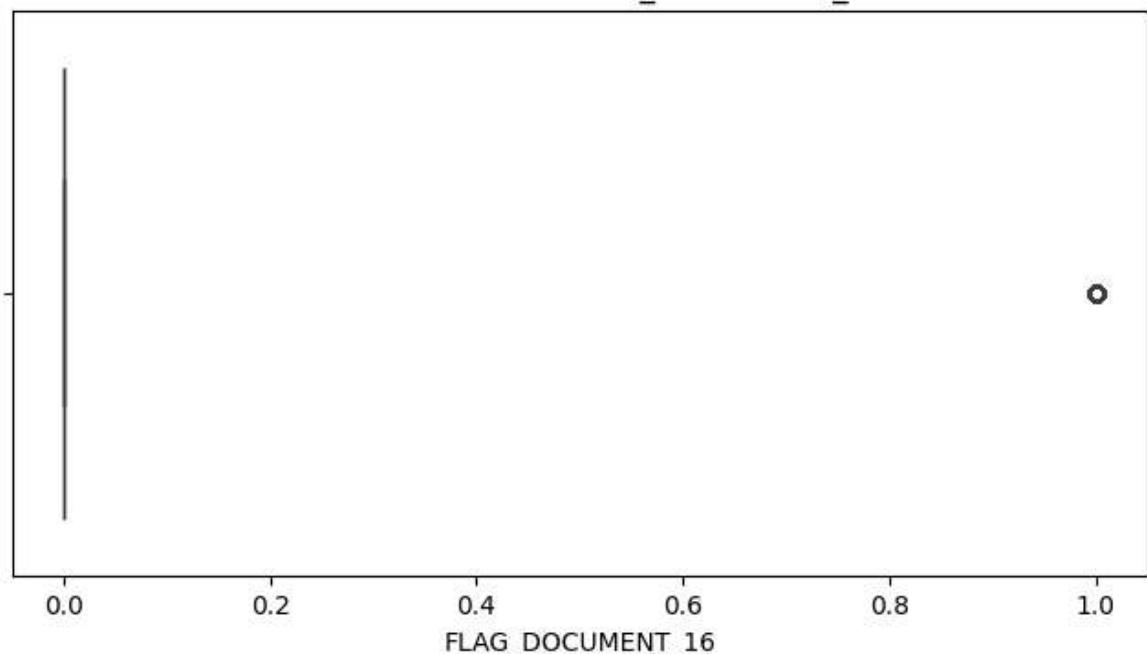
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_15



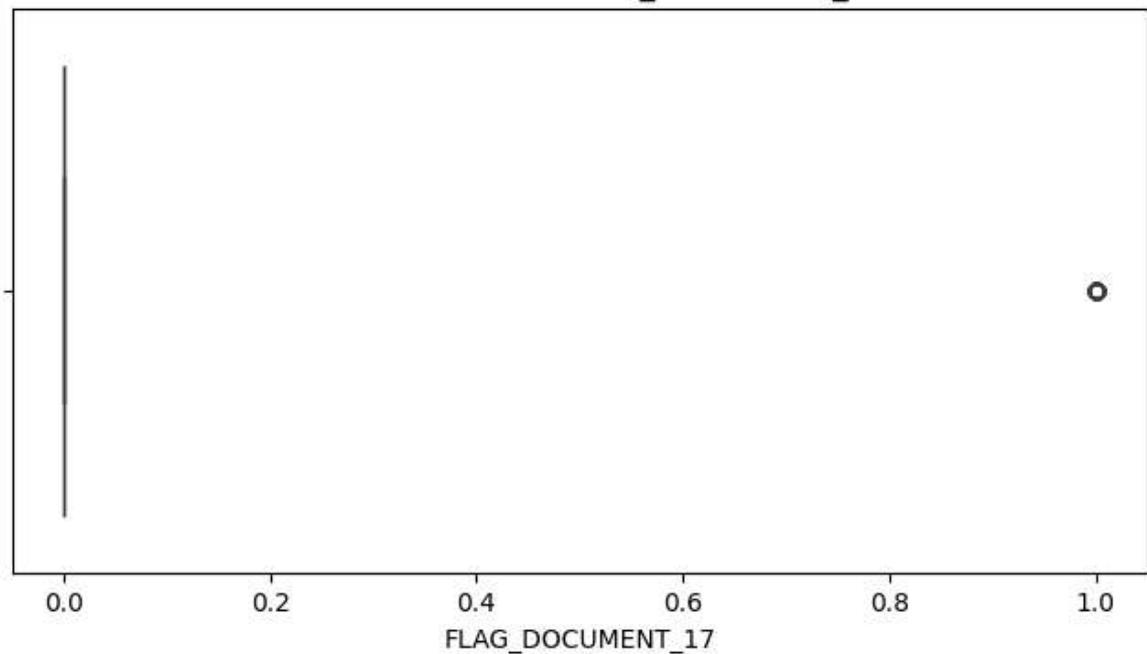
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\noncategorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_16



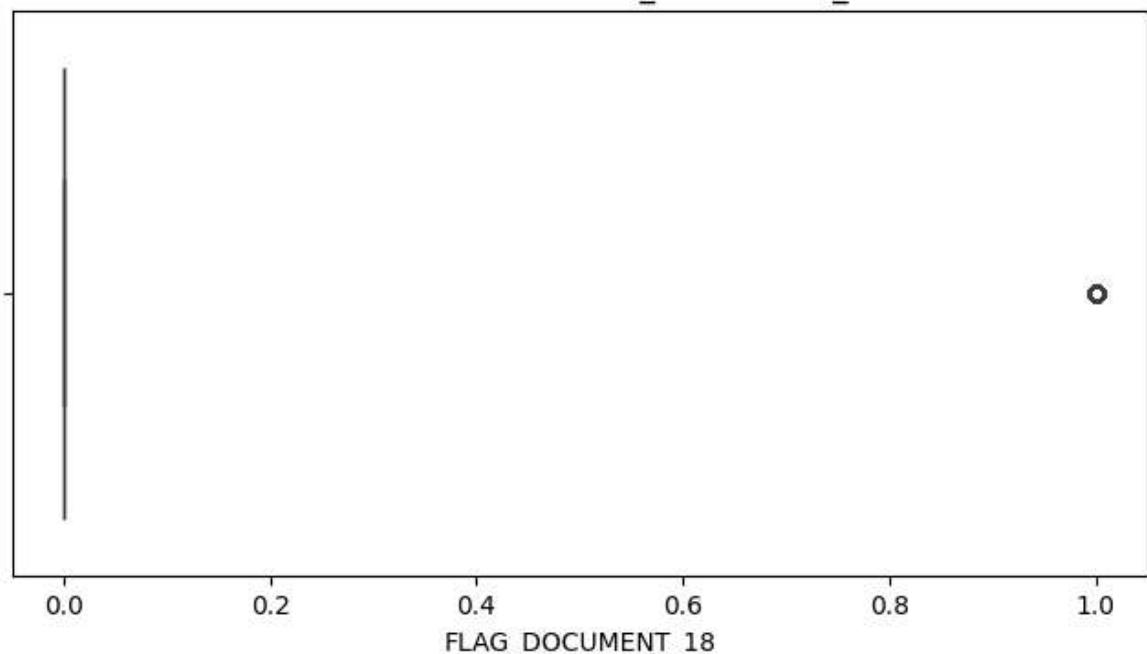
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\n\b\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_17



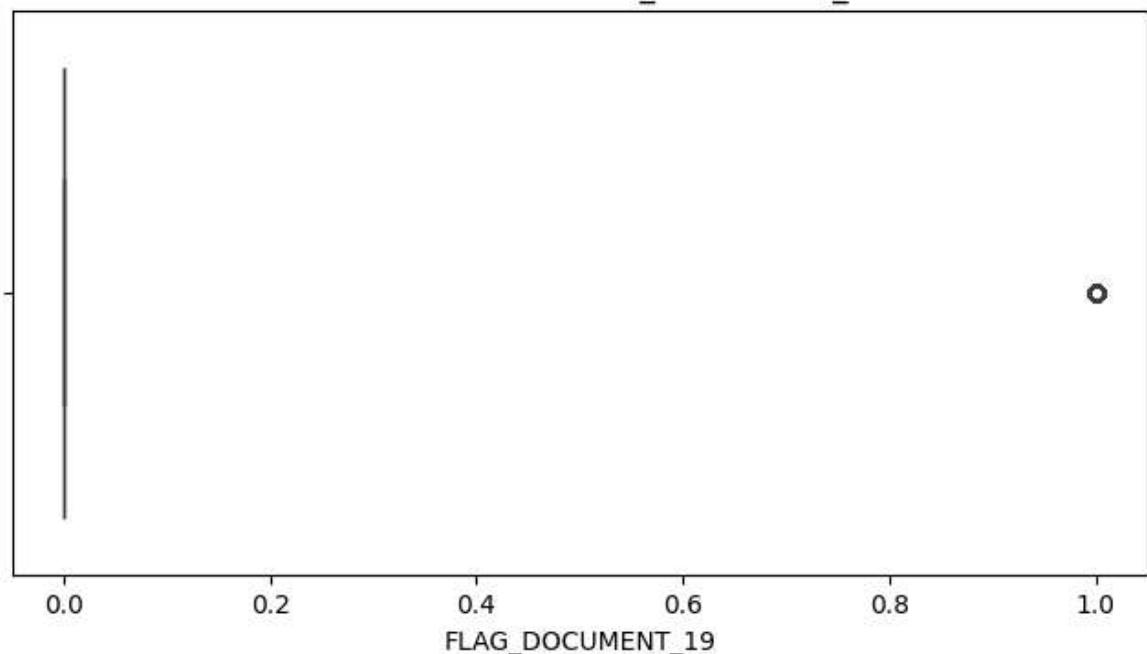
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\n\b\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_18



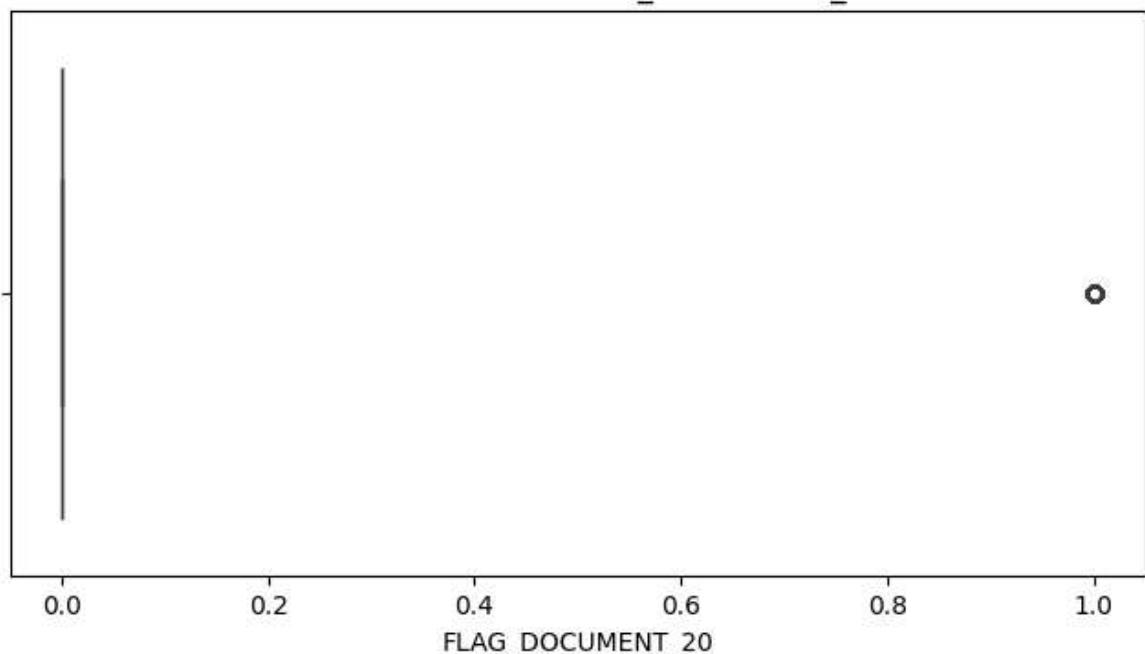
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_19



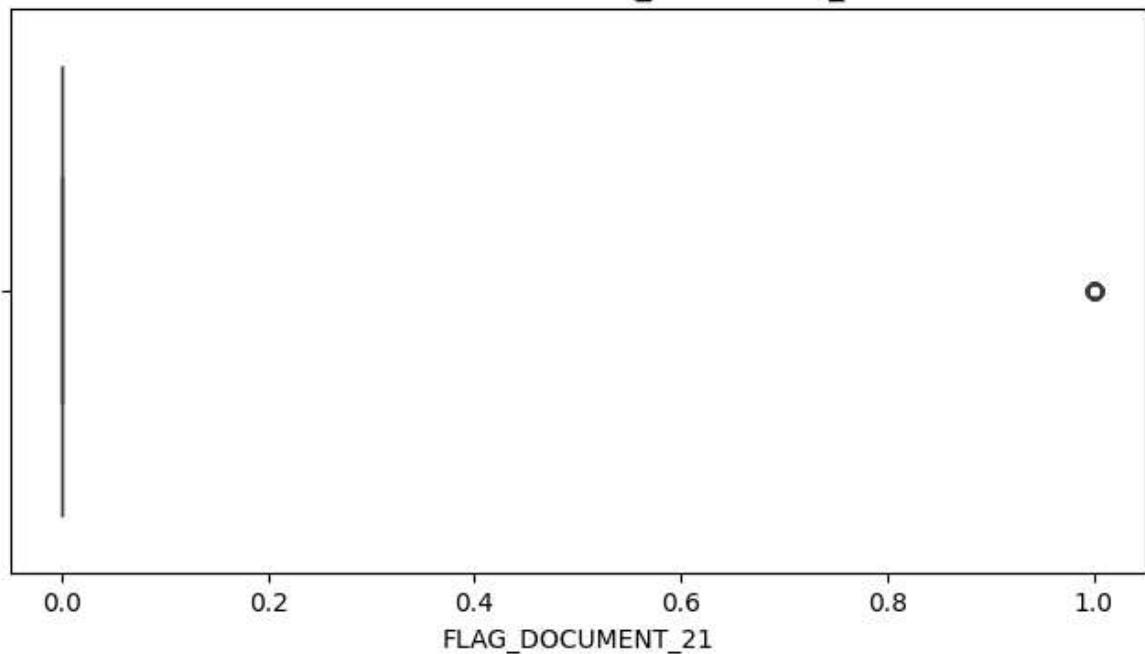
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_20



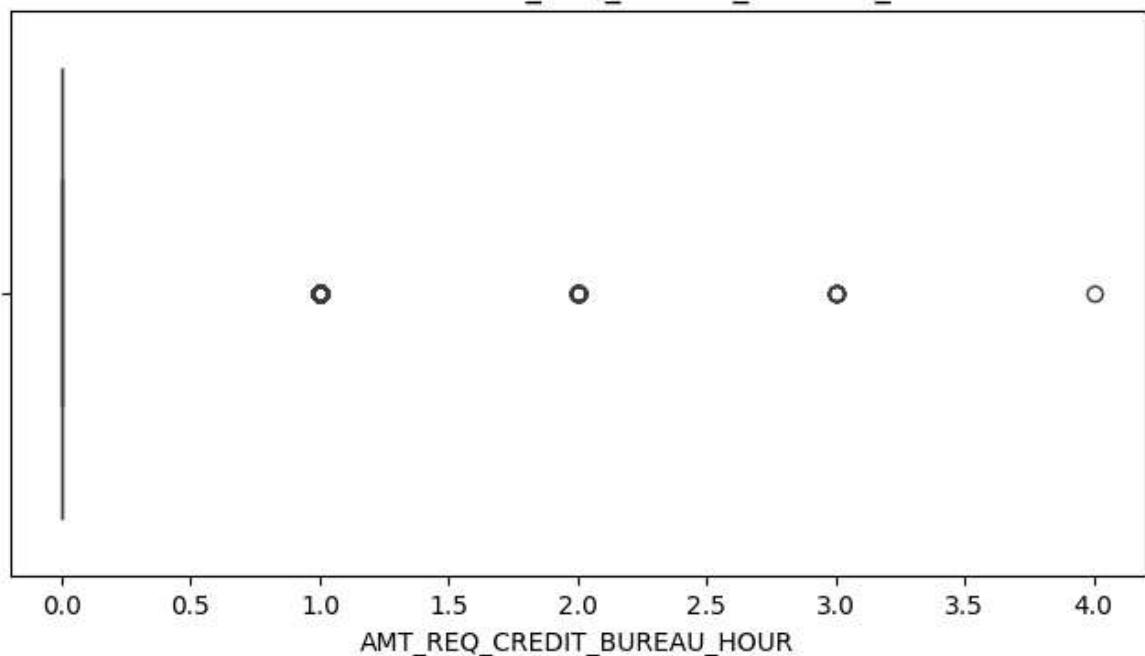
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna FLAG\_DOCUMENT\_21



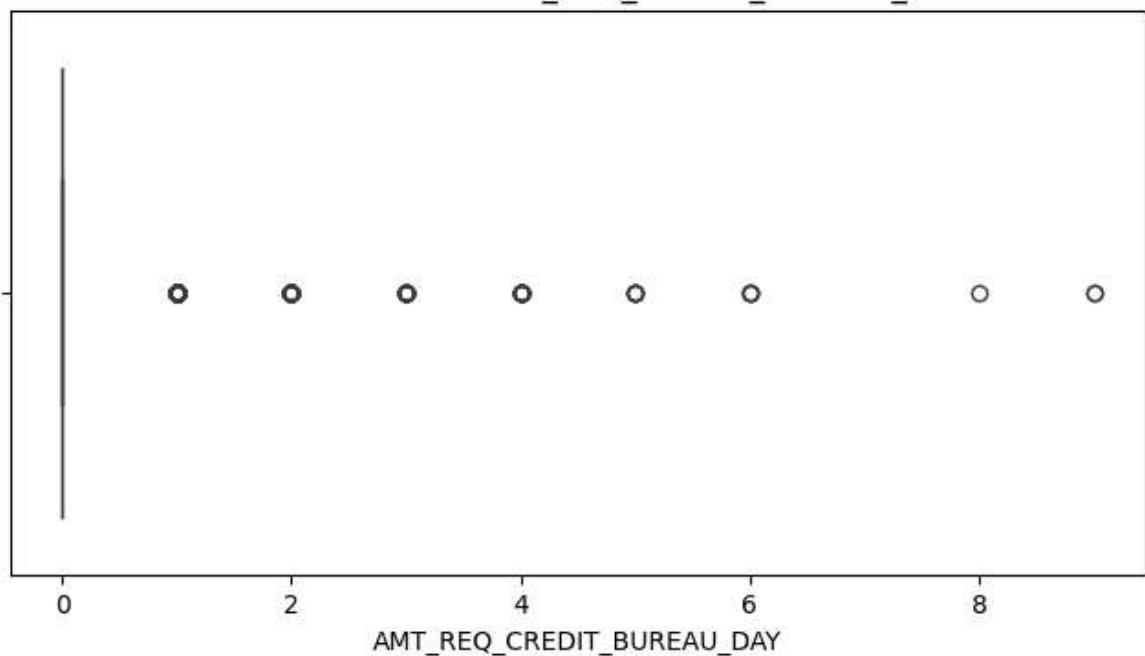
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna AMT\_REQ\_CREDIT\_BUREAU\_HOUR



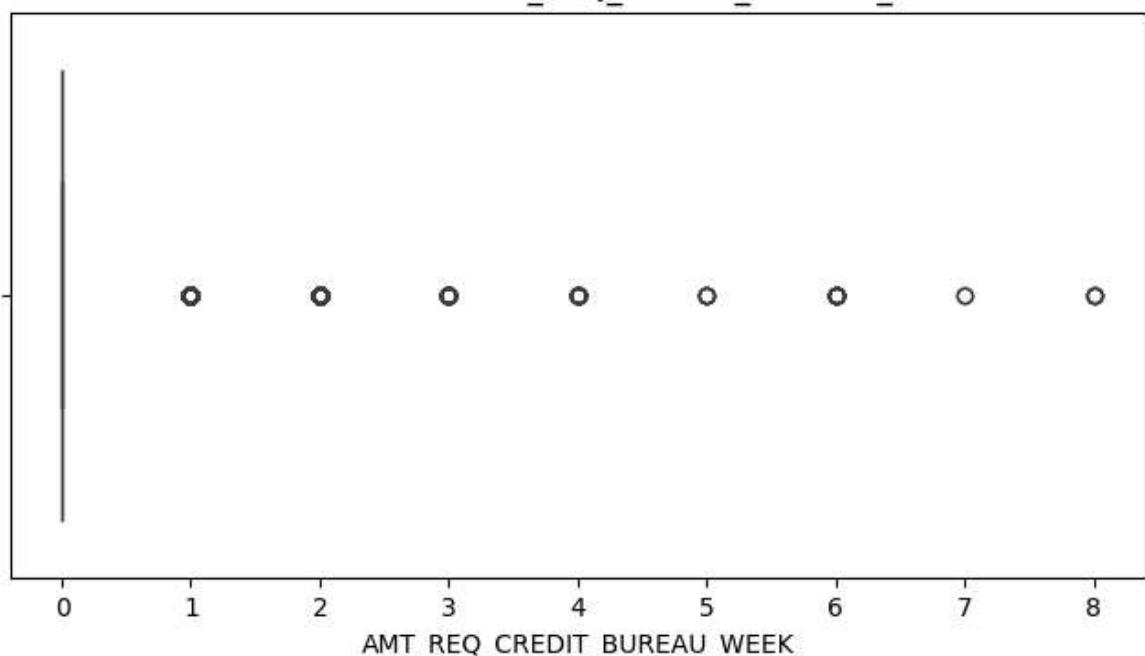
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

### Outliers na coluna AMT\_REQ\_CREDIT\_BUREAU\_DAY



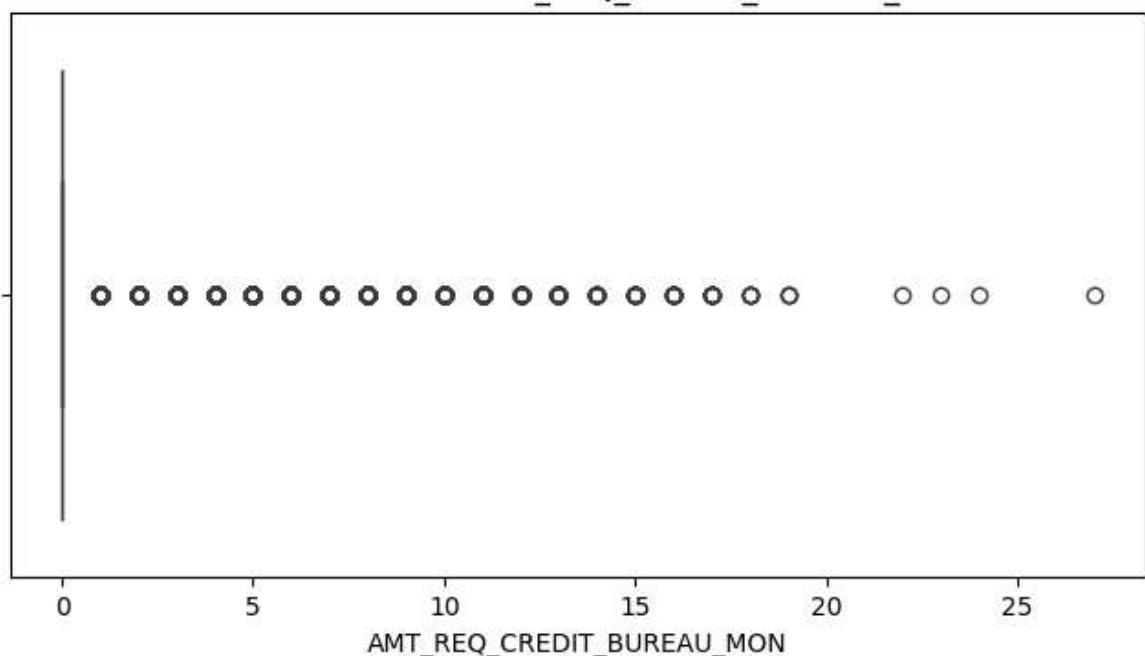
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna AMT\_REQ\_CREDIT\_BUREAU\_WEEK



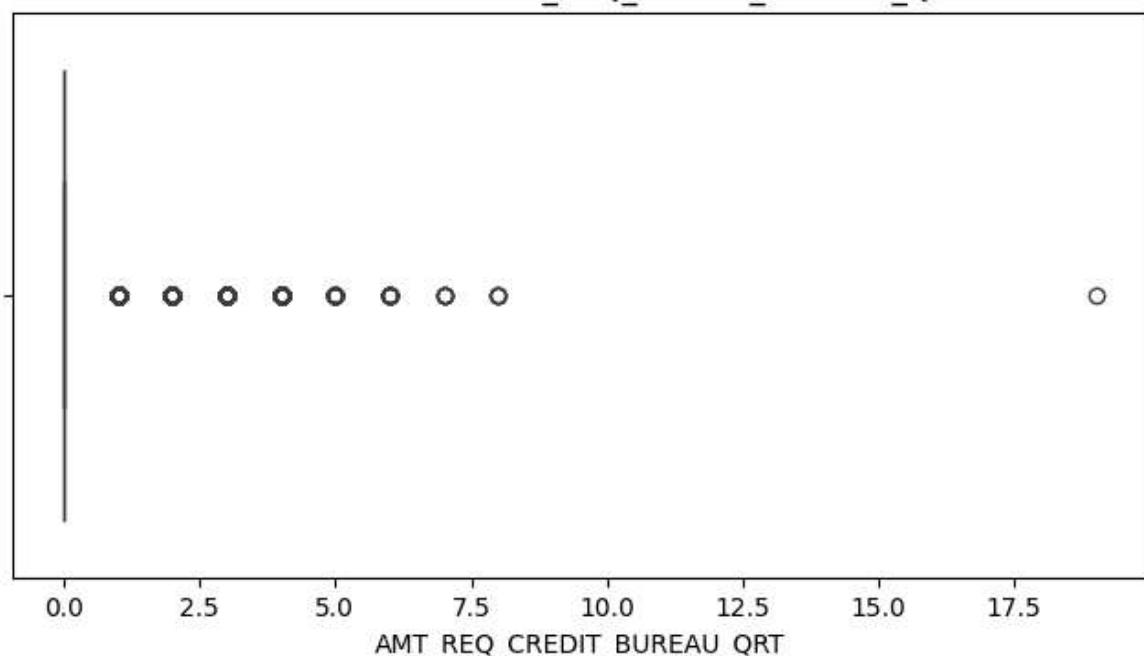
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

## Outliers na coluna AMT\_REQ\_CREDIT\_BUREAU\_MON



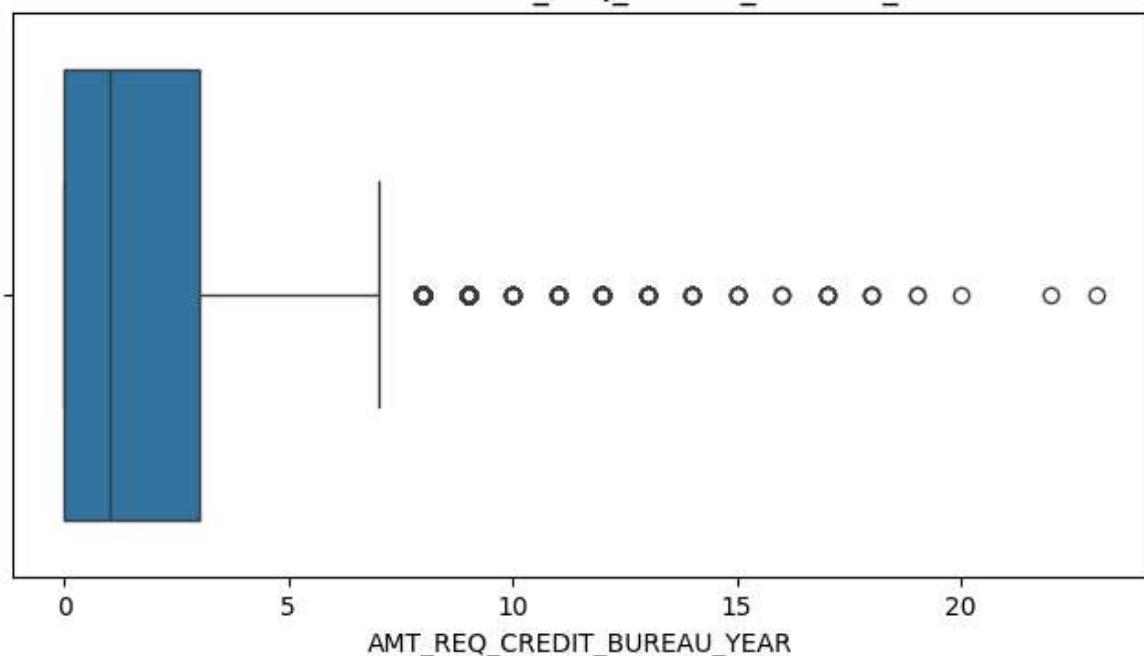
```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn
\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will
be removed in a future version of pandas.
    positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

Outliers na coluna AMT\_REQ\_CREDIT\_BUREAU\_QRT



```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\seaborn\categorical.py:640: FutureWarning: SeriesGroupBy.grouper is deprecated and will be removed in a future version of pandas.  
positions = grouped.grouper.result_index.to_numpy(dtype=float)
```

Outliers na coluna AMT\_REQ\_CREDIT\_BUREAU\_YEAR



## E. Tratando os dados nulos:

Quantificando os dados nulos:

```
In [10]: df.info(verbose=True, show_counts=True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 246008 entries, 0 to 246007
Data columns (total 122 columns):
 #   Column           Non-Null Count Dtype  
 --- 
 0   SK_ID_CURR       246008 non-null  int64  
 1   TARGET           246008 non-null  int64  
 2   NAME_CONTRACT_TYPE 246008 non-null  object  
 3   CODE_GENDER      246008 non-null  object  
 4   FLAG_OWN_CAR     246008 non-null  object  
 5   FLAG_OWN_REALTY  246008 non-null  object  
 6   CNT_CHILDREN     246008 non-null  int64  
 7   AMT_INCOME_TOTAL 246008 non-null  float64 
 8   AMT_CREDIT        246008 non-null  float64 
 9   AMT_ANNUITY       245998 non-null  float64 
 10  AMT_GOODS_PRICE   245782 non-null  float64 
 11  NAME_TYPE_SUITE   244960 non-null  object  
 12  NAME_INCOME_TYPE  246008 non-null  object  
 13  NAME_EDUCATION_TYPE 246008 non-null  object  
 14  NAME_FAMILY_STATUS 246008 non-null  object  
 15  NAME_HOUSING_TYPE 246008 non-null  object  
 16  REGION_POPULATION_RELATIVE 246008 non-null  float64 
 17  DAYS_BIRTH        246008 non-null  int64  
 18  DAYS_EMPLOYED     246008 non-null  int64  
 19  DAYS_REGISTRATION 246008 non-null  float64 
 20  DAYS_ID_PUBLISH   246008 non-null  int64  
 21  OWN_CAR_AGE       83649 non-null   float64 
 22  FLAG_MOBIL        246008 non-null  int64  
 23  FLAG_EMP_PHONE    246008 non-null  int64  
 24  FLAG_WORK_PHONE   246008 non-null  int64  
 25  FLAG_CONT_MOBILE  246008 non-null  int64  
 26  FLAG_PHONE         246008 non-null  int64  
 27  FLAG_EMAIL         246008 non-null  int64  
 28  OCCUPATION_TYPE   168771 non-null  object  
 29  CNT_FAM_MEMBERS   246007 non-null  float64 
 30  REGION_RATING_CLIENT 246008 non-null  int64  
 31  REGION_RATING_CLIENT_W_CITY 246008 non-null  int64  
 32  WEEKDAY_APPR_PROCESS_START 246008 non-null  object  
 33  HOUR_APPR_PROCESS_START 246008 non-null  int64  
 34  REG_REGION_NOT_LIVE_REGION 246008 non-null  int64  
 35  REG_REGION_NOT_WORK_REGION 246008 non-null  int64  
 36  LIVE_REGION_NOT_WORK_REGION 246008 non-null  int64  
 37  REG_CITY_NOT_LIVE_CITY 246008 non-null  int64  
 38  REG_CITY_NOT_WORK_CITY 246008 non-null  int64  
 39  LIVE_CITY_NOT_WORK_CITY 246008 non-null  int64  
 40  ORGANIZATION_TYPE   246008 non-null  object  
 41  EXT_SOURCE_1        107205 non-null  float64 
 42  EXT_SOURCE_2        245464 non-null  float64 
 43  EXT_SOURCE_3        197280 non-null  float64 
 44  APARTMENTS_AVG      121053 non-null  float64 
 45  BASEMENTAREA_AVG    101918 non-null  float64 
 46  YEARS_BEGINEXPLUATATION_AVG 125912 non-null  float64 
 47  YEARS_BUILD_AVG     82328 non-null   float64 
 48  COMMONAREA_AVG      74030 non-null   float64 
 49  ELEVATORS_AVG       114800 non-null  float64 
 50  ENTRANCES_AVG       122064 non-null  float64 
 51  FLOORSMAX_AVG       123525 non-null  float64 
 52  FLOORSMIN_AVG       79009 non-null   float64 
 53  LANDAREA_AVG        99921 non-null   float64 
 54  LIVINGAPARTMENTS_AVG 77730 non-null  float64
```

55	LIVINGAREA_AVG	122391	non-null	float64
56	NONLIVINGAPARTMENTS_AVG	75094	non-null	float64
57	NONLIVINGAREA_AVG	110148	non-null	float64
58	APARTMENTS_MODE	121053	non-null	float64
59	BASEMENTAREA_MODE	101918	non-null	float64
60	YEARS_BEGINEXPLUATATION_MODE	125912	non-null	float64
61	YEARS_BUILD_MODE	82328	non-null	float64
62	COMMONAREA_MODE	74030	non-null	float64
63	ELEVATORS_MODE	114800	non-null	float64
64	ENTRANCES_MODE	122064	non-null	float64
65	FLOORSMAX_MODE	123525	non-null	float64
66	FLOORSMIN_MODE	79009	non-null	float64
67	LANDAREA_MODE	99921	non-null	float64
68	LIVINGAPARTMENTS_MODE	77730	non-null	float64
69	LIVINGAREA_MODE	122391	non-null	float64
70	NONLIVINGAPARTMENTS_MODE	75094	non-null	float64
71	NONLIVINGAREA_MODE	110148	non-null	float64
72	APARTMENTS_MEDI	121053	non-null	float64
73	BASEMENTAREA_MEDI	101918	non-null	float64
74	YEARS_BEGINEXPLUATATION_MEDI	125912	non-null	float64
75	YEARS_BUILD_MEDI	82328	non-null	float64
76	COMMONAREA_MEDI	74030	non-null	float64
77	ELEVATORS_MEDI	114800	non-null	float64
78	ENTRANCES_MEDI	122064	non-null	float64
79	FLOORSMAX_MEDI	123525	non-null	float64
80	FLOORSMIN_MEDI	79009	non-null	float64
81	LANDAREA_MEDI	99921	non-null	float64
82	LIVINGAPARTMENTS_MEDI	77730	non-null	float64
83	LIVINGAREA_MEDI	122391	non-null	float64
84	NONLIVINGAPARTMENTS_MEDI	75094	non-null	float64
85	NONLIVINGAREA_MEDI	110148	non-null	float64
86	FONDKAPREMONT_MODE	77616	non-null	object
87	HOUSETYPE_MODE	122468	non-null	object
88	TOTALAREA_MODE	127171	non-null	float64
89	WALLSMATERIAL_MODE	120844	non-null	object
90	EMERGENCYSTATE_MODE	129247	non-null	object
91	OBS_30_CNT_SOCIAL_CIRCLE	245195	non-null	float64
92	DEF_30_CNT_SOCIAL_CIRCLE	245195	non-null	float64
93	OBS_60_CNT_SOCIAL_CIRCLE	245195	non-null	float64
94	DEF_60_CNT_SOCIAL_CIRCLE	245195	non-null	float64
95	DAYS_LAST_PHONE_CHANGE	246007	non-null	float64
96	FLAG_DOCUMENT_2	246008	non-null	int64
97	FLAG_DOCUMENT_3	246008	non-null	int64
98	FLAG_DOCUMENT_4	246008	non-null	int64
99	FLAG_DOCUMENT_5	246008	non-null	int64
100	FLAG_DOCUMENT_6	246008	non-null	int64
101	FLAG_DOCUMENT_7	246008	non-null	int64
102	FLAG_DOCUMENT_8	246008	non-null	int64
103	FLAG_DOCUMENT_9	246008	non-null	int64
104	FLAG_DOCUMENT_10	246008	non-null	int64
105	FLAG_DOCUMENT_11	246008	non-null	int64
106	FLAG_DOCUMENT_12	246008	non-null	int64
107	FLAG_DOCUMENT_13	246008	non-null	int64
108	FLAG_DOCUMENT_14	246008	non-null	int64
109	FLAG_DOCUMENT_15	246008	non-null	int64
110	FLAG_DOCUMENT_16	246008	non-null	int64
111	FLAG_DOCUMENT_17	246008	non-null	int64
112	FLAG_DOCUMENT_18	246008	non-null	int64
113	FLAG_DOCUMENT_19	246008	non-null	int64
114	FLAG_DOCUMENT_20	246008	non-null	int64

```
115 FLAG_DOCUMENT_21          246008 non-null  int64
116 AMT_REQ_CREDIT_BUREAU_HOUR 212836 non-null  float64
117 AMT_REQ_CREDIT_BUREAU_DAY   212836 non-null  float64
118 AMT_REQ_CREDIT_BUREAU_WEEK 212836 non-null  float64
119 AMT_REQ_CREDIT_BUREAU_MON   212836 non-null  float64
120 AMT_REQ_CREDIT_BUREAU_QRT   212836 non-null  float64
121 AMT_REQ_CREDIT_BUREAU_YEAR  212836 non-null  float64
dtypes: float64(65), int64(41), object(16)
memory usage: 229.0+ MB
```

In [11]: `df.isnull().sum().head(60)`

```
Out[11]: SK_ID_CURR          0
TARGET              0
NAME_CONTRACT_TYPE 0
CODE_GENDER         0
FLAG_OWN_CAR        0
FLAG_OWN_REALTY    0
CNT_CHILDREN        0
AMT_INCOME_TOTAL   0
AMT_CREDIT          0
AMT_ANNUITY         10
AMT_GOODS_PRICE     226
NAME_TYPE_SUITE     1048
NAME_INCOME_TYPE    0
NAME_EDUCATION_TYPE 0
NAME_FAMILY_STATUS   0
NAME_HOUSING_TYPE   0
REGION_POPULATION_RELATIVE 0
DAYS_BIRTH          0
DAYS_EMPLOYED       0
DAYS_REGISTRATION   0
DAYS_ID_PUBLISH    0
OWN_CAR_AGE         162359
FLAG_MOBIL          0
FLAG_EMP_PHONE      0
FLAG_WORK_PHONE     0
FLAG_CONT_MOBILE    0
FLAG_PHONE          0
FLAG_EMAIL          0
OCCUPATION_TYPE     77237
CNT_FAM_MEMBERS     1
REGION_RATING_CLIENT 0
REGION_RATING_CLIENT_W_CITY 0
WEEKDAY_APPR_PROCESS_START 0
HOUR_APPR_PROCESS_START 0
REG_REGION_NOT_LIVE_REGION 0
REG_REGION_NOT_WORK_REGION 0
LIVE_REGION_NOT_WORK_REGION 0
REG_CITY_NOT_LIVE_CITY 0
REG_CITY_NOT_WORK_CITY 0
LIVE_CITY_NOT_WORK_CITY 0
ORGANIZATION_TYPE    0
EXT_SOURCE_1         138803
EXT_SOURCE_2         544
EXT_SOURCE_3         48728
APARTMENTS_AVG      124955
BASEMENTAREA_AVG    144090
YEARS_BEGINEXPLUATATION_AVG 120096
YEARS_BUILD_AVG     163680
COMMONAREA_AVG      171978
ELEVATORS_AVG       131208
ENTRANCES_AVG       123944
FLOORSMAX_AVG       122483
FLOORSMIN_AVG       166999
LANDAREA_AVG        146087
LIVINGAPARTMENTS_AVG 168278
LIVINGAREA_AVG      123617
NONLIVINGAPARTMENTS_AVG 170914
NONLIVINGAREA_AVG   135860
APARTMENTS_MODE     124955
```

```
BASEMENTAREA_MODE           144090
dtype: int64
```

Visualizando a porcentagem de dados faltantes por coluna:

```
In [12]: nulos_por_coluna = df.isnull().mean()*100
nulos_por_coluna.head(60)
```

Out[12]:	SK_ID_CURR	0.000000
	TARGET	0.000000
	NAME_CONTRACT_TYPE	0.000000
	CODE_GENDER	0.000000
	FLAG_OWN_CAR	0.000000
	FLAG_OWN_REALTY	0.000000
	CNT_CHILDREN	0.000000
	AMT_INCOME_TOTAL	0.000000
	AMT_CREDIT	0.000000
	AMT_ANNUITY	0.004065
	AMT_GOODS_PRICE	0.091867
	NAME_TYPE_SUITE	0.426002
	NAME_INCOME_TYPE	0.000000
	NAME_EDUCATION_TYPE	0.000000
	NAME_FAMILY_STATUS	0.000000
	NAME_HOUSING_TYPE	0.000000
	REGION_POPULATION_RELATIVE	0.000000
	DAYS_BIRTH	0.000000
	DAYS_EMPLOYED	0.000000
	DAYS_REGISTRATION	0.000000
	DAYS_ID_PUBLISH	0.000000
	OWN_CAR_AGE	65.997447
	FLAG_MOBIL	0.000000
	FLAG_EMP_PHONE	0.000000
	FLAG_WORK_PHONE	0.000000
	FLAG_CONT_MOBILE	0.000000
	FLAG_PHONE	0.000000
	FLAG_EMAIL	0.000000
	OCCUPATION_TYPE	31.396133
	CNT_FAM_MEMBERS	0.000406
	REGION_RATING_CLIENT	0.000000
	REGION_RATING_CLIENT_W_CITY	0.000000
	WEEKDAY_APPR_PROCESS_START	0.000000
	HOUR_APPR_PROCESS_START	0.000000
	REG_REGION_NOT_LIVE_REGION	0.000000
	REG_REGION_NOT_WORK_REGION	0.000000
	LIVE_REGION_NOT_WORK_REGION	0.000000
	REG_CITY_NOT_LIVE_CITY	0.000000
	REG_CITY_NOT_WORK_CITY	0.000000
	LIVE_CITY_NOT_WORK_CITY	0.000000
	ORGANIZATION_TYPE	0.000000
	EXT_SOURCE_1	56.422149
	EXT_SOURCE_2	0.221131
	EXT_SOURCE_3	19.807486
	APARTMENTS_AVG	50.793064
	BASEMENTAREA_AVG	58.571266
	YEARS_BEGINEXPLUATATION_AVG	48.817925
	YEARS_BUILD_AVG	66.534422
	COMMONAREA_AVG	69.907483
	ELEVATORS_AVG	53.334851
	ENTRANCES_AVG	50.382101
	FLOORSMAX_AVG	49.788218
	FLOORSMIN_AVG	67.883565
	LANDAREA_AVG	59.383028
	LIVINGAPARTMENTS_AVG	68.403467
	LIVINGAREA_AVG	50.249179
	NONLIVINGAPARTMENTS_AVG	69.474976
	NONLIVINGAREA_AVG	55.225846
	APARTMENTS_MODE	50.793064

```
BASEMENTAREA_MODE           58.571266
dtype: float64
```

Constata-se do resultado acima que há colunas com uma porcentagem alta de valores nulos. Tais colunas não terão grande influencia na construção do aprendizado de máquina e por isso serão excluídas aquelas com o percentual acima de 20%.

Será utilizado o método dropna com os parâmetros thresh (parâmetro que apaga as colunas que não tenham o pelo menos a quantidade de valores não nulos passados, ou seja, se a coluna tiver valores não nulos menor daquele passado no parâmetro, será deletada) e o parâmetro axis (aponta qual serie será deletado linha (0) ou coluna (1)).

No presente caso, o thresh receberá o valor de 80% do tamanho do dataframe, ou seja, a coluna ou linha deverá conter pelo menos 80% das linhas preenchidas com valores não nulos, caso a porcentagem seja menor, a coluna ou linha será excluída. Já o axis receberá o valor 1, indicando que a coluna será descartada.

```
In [13]: df = df.dropna(thresh=len(df)*0.90, axis=1)
```

Após a exclusão das colunas com os parâmetros acima, restaram 72 colunas no dataframe, conforme verifica-se abaixo:

```
In [14]: df.shape[1]
```

```
Out[14]: 65
```

```
In [15]: df.info(verbose=True, show_counts=True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 246008 entries, 0 to 246007
Data columns (total 65 columns):
 #   Column           Non-Null Count Dtype  
 --- 
 0   SK_ID_CURR       246008 non-null int64  
 1   TARGET           246008 non-null int64  
 2   NAME_CONTRACT_TYPE 246008 non-null object  
 3   CODE_GENDER       246008 non-null object  
 4   FLAG_OWN_CAR      246008 non-null object  
 5   FLAG_OWN_REALTY    246008 non-null object  
 6   CNT_CHILDREN      246008 non-null int64  
 7   AMT_INCOME_TOTAL   246008 non-null float64 
 8   AMT_CREDIT         246008 non-null float64 
 9   AMT_ANNUITY        245998 non-null float64 
 10  AMT_GOODS_PRICE     245782 non-null float64 
 11  NAME_TYPE_SUITE     244960 non-null object  
 12  NAME_INCOME_TYPE    246008 non-null object  
 13  NAME_EDUCATION_TYPE 246008 non-null object  
 14  NAME_FAMILY_STATUS   246008 non-null object  
 15  NAME_HOUSING_TYPE    246008 non-null object  
 16  REGION_POPULATION_RELATIVE 246008 non-null float64 
 17  DAYS_BIRTH          246008 non-null int64  
 18  DAYS_EMPLOYED        246008 non-null int64  
 19  DAYS_REGISTRATION     246008 non-null float64 
 20  DAYS_ID_PUBLISH      246008 non-null int64  
 21  FLAG_MOBIL           246008 non-null int64  
 22  FLAG_EMP_PHONE        246008 non-null int64  
 23  FLAG_WORK_PHONE       246008 non-null int64  
 24  FLAG_CONT_MOBILE      246008 non-null int64  
 25  FLAG_PHONE            246008 non-null int64  
 26  FLAG_EMAIL             246008 non-null int64  
 27  CNT_FAM_MEMBERS       246007 non-null float64 
 28  REGION_RATING_CLIENT   246008 non-null int64  
 29  REGION_RATING_CLIENT_W_CITY 246008 non-null int64  
 30  WEEKDAY_APPR_PROCESS_START 246008 non-null object  
 31  HOUR_APPR_PROCESS_START 246008 non-null int64  
 32  REG_REGION_NOT_LIVE_REGION 246008 non-null int64  
 33  REG_REGION_NOT_WORK_REGION 246008 non-null int64  
 34  LIVE_REGION_NOT_WORK_REGION 246008 non-null int64  
 35  REG_CITY_NOT_LIVE_CITY   246008 non-null int64  
 36  REG_CITY_NOT_WORK_CITY    246008 non-null int64  
 37  LIVE_CITY_NOT_WORK_CITY   246008 non-null int64  
 38  ORGANIZATION_TYPE       246008 non-null object  
 39  EXT_SOURCE_2            245464 non-null float64 
 40  OBS_30_CNT_SOCIAL_CIRCLE 245195 non-null float64 
 41  DEF_30_CNT_SOCIAL_CIRCLE 245195 non-null float64 
 42  OBS_60_CNT_SOCIAL_CIRCLE 245195 non-null float64 
 43  DEF_60_CNT_SOCIAL_CIRCLE 245195 non-null float64 
 44  DAYS_LAST_PHONE_CHANGE   246007 non-null float64 
 45  FLAG_DOCUMENT_2          246008 non-null int64  
 46  FLAG_DOCUMENT_3          246008 non-null int64  
 47  FLAG_DOCUMENT_4          246008 non-null int64  
 48  FLAG_DOCUMENT_5          246008 non-null int64  
 49  FLAG_DOCUMENT_6          246008 non-null int64  
 50  FLAG_DOCUMENT_7          246008 non-null int64  
 51  FLAG_DOCUMENT_8          246008 non-null int64  
 52  FLAG_DOCUMENT_9          246008 non-null int64  
 53  FLAG_DOCUMENT_10         246008 non-null int64  
 54  FLAG_DOCUMENT_11         246008 non-null int64
```

```
55 FLAG_DOCUMENT_12           246008 non-null int64
56 FLAG_DOCUMENT_13           246008 non-null int64
57 FLAG_DOCUMENT_14           246008 non-null int64
58 FLAG_DOCUMENT_15           246008 non-null int64
59 FLAG_DOCUMENT_16           246008 non-null int64
60 FLAG_DOCUMENT_17           246008 non-null int64
61 FLAG_DOCUMENT_18           246008 non-null int64
62 FLAG_DOCUMENT_19           246008 non-null int64
63 FLAG_DOCUMENT_20           246008 non-null int64
64 FLAG_DOCUMENT_21           246008 non-null int64
dtypes: float64(13), int64(41), object(11)
memory usage: 122.0+ MB
```

Entretanto, ainda há dados nulos no dataframe a serem tratados, todavia, em quantitavo muito inferior a 7.326.502, facilitando o tratamento, conforme abaixo:

```
In [16]: df.isnull().sum().sum()
```

```
Out[16]: 5082
```

```
In [17]: nulos_por_coluna.sort_values(ascending=False).head(60)
```

Out[17]:	COMMONAREA_MEDI	69.907483
	COMMONAREA_AVG	69.907483
	COMMONAREA_MODE	69.907483
	NONLIVINGAPARTMENTS_MODE	69.474976
	NONLIVINGAPARTMENTS_AVG	69.474976
	NONLIVINGAPARTMENTS_MEDI	69.474976
	FONDKAPREMONT_MODE	68.449807
	LIVINGAPARTMENTS_MODE	68.403467
	LIVINGAPARTMENTS_AVG	68.403467
	LIVINGAPARTMENTS_MEDI	68.403467
	FLOORSMIN_AVG	67.883565
	FLOORSMIN_MODE	67.883565
	FLOORSMIN_MEDI	67.883565
	YEARS_BUILD_MEDI	66.534422
	YEARS_BUILD_MODE	66.534422
	YEARS_BUILD_AVG	66.534422
	OWN_CAR_AGE	65.997447
	LANDAREA_MEDI	59.383028
	LANDAREA_MODE	59.383028
	LANDAREA_AVG	59.383028
	BASEMENTAREA_MEDI	58.571266
	BASEMENTAREA_AVG	58.571266
	BASEMENTAREA_MODE	58.571266
	EXT_SOURCE_1	56.422149
	NONLIVINGAREA_MODE	55.225846
	NONLIVINGAREA_AVG	55.225846
	NONLIVINGAREA_MEDI	55.225846
	ELEVATORS_MEDI	53.334851
	ELEVATORS_AVG	53.334851
	ELEVATORS_MODE	53.334851
	WALLSMATERIAL_MODE	50.878020
	APARTMENTS_MEDI	50.793064
	APARTMENTS_AVG	50.793064
	APARTMENTS_MODE	50.793064
	ENTRANCES_MEDI	50.382101
	ENTRANCES_AVG	50.382101
	ENTRANCES_MODE	50.382101
	LIVINGAREA_AVG	50.249179
	LIVINGAREA_MODE	50.249179
	LIVINGAREA_MEDI	50.249179
	HOUSETYPE_MODE	50.217879
	FLOORSMAX_MODE	49.788218
	FLOORSMAX_MEDI	49.788218
	FLOORSMAX_AVG	49.788218
	YEARS_BEGINEXPLUATATION_MODE	48.817925
	YEARS_BEGINEXPLUATATION_MEDI	48.817925
	YEARS_BEGINEXPLUATATION_AVG	48.817925
	TOTALAREA_MODE	48.306153
	EMERGENCYSTATE_MODE	47.462278
	OCCUPATION_TYPE	31.396133
	EXT_SOURCE_3	19.807486
	AMT_REQ_CREDIT_BUREAU_HOUR	13.484114
	AMT_REQ_CREDIT_BUREAU_DAY	13.484114
	AMT_REQ_CREDIT_BUREAU_WEEK	13.484114
	AMT_REQ_CREDIT_BUREAU_MON	13.484114
	AMT_REQ_CREDIT_BUREAU_QRT	13.484114
	AMT_REQ_CREDIT_BUREAU_YEAR	13.484114
	NAME_TYPE_SUITE	0.426002
	OBS_30_CNT_SOCIAL_CIRCLE	0.330477

```
DEF_30_CNT_SOCIAL_CIRCLE      0.330477
dtype: float64
```

Continuando o tratamento dos dados nulos, podemos, ainda, excluir as linhas que contem 1 ou mais dados nulos, visto que a perda de dados em relação ao seu total (em torno de 20%) não será significativa para comprometer o aprendizado do modelo, posto que o volume de dados é alto.

```
In [18]: df = df.dropna(axis=0)
```

```
In [19]: df.shape[0]
```

```
Out[19]: 243597
```

Uma vez que, após a exclusão das linhas duplicadas, o dataframe permanecerá com o index das linhas originais, devemos reorganizá-lo, a fim, de facilitar a leitura das linhas, utilizando o método reset\_index com o parâmetro drop=True, a fim de não transformar o index em coluna, apenas para tê-lo como referência no momento de leitura e análise do dataframe.

```
In [20]: df = df.reset_index(drop=True)
df
```

Out[20]:

	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR
0	456162	0	Cash loans	F	N
1	134978	0	Cash loans	F	N
2	318952	0	Cash loans	M	Y
3	361264	0	Cash loans	F	N
4	260639	0	Cash loans	F	N
...	...	...	...	...	.
<b>243592</b>	242114	0	Cash loans	F	N
<b>243593</b>	452374	0	Cash loans	F	N
<b>243594</b>	276545	1	Revolving loans	M	N
<b>243595</b>	236776	1	Cash loans	M	Y
<b>243596</b>	454197	0	Cash loans	F	N

243597 rows × 65 columns

## F. Analisando as colunas e suas correlações:

A finalidade neste tópico é perceber quais colunas são as mais importantes para o dataset.

In [21]:

```
correlacao = df.select_dtypes(include=[float, int]).corr()
correlacao
```

Out[21]:

	SK_ID_CURR	TARGET	CNT_CHILDREN	AMT_INCOM
<b>SK_ID_CURR</b>	1.000000	-0.002705	-0.000443	-2.57
<b>TARGET</b>	-0.002705	1.000000	0.019270	-1.93
<b>CNT_CHILDREN</b>	-0.000443	0.019270	1.000000	1.28
<b>AMT_INCOME_TOTAL</b>	-0.002576	-0.001933	0.012842	1.000000
<b>AMT_CREDIT</b>	0.000511	-0.031044	0.002089	1.41
<b>AMT_ANNUITY</b>	0.000387	-0.013826	0.022162	1.73
<b>AMT_GOODS_PRICE</b>	0.000546	-0.039937	-0.002046	1.43
<b>REGION_POPULATION_RELATIVE</b>	0.000126	-0.036220	-0.022729	6.71
<b>DAYS_BIRTH</b>	-0.002065	0.079813	0.333085	2.54
<b>DAYS_EMPLOYED</b>	0.001837	-0.045052	-0.239913	-5.87
<b>DAYS_REGISTRATION</b>	-0.002743	0.041864	0.183963	2.49
<b>DAYS_ID_PUBLISH</b>	0.000063	0.051523	-0.026789	7.54
<b>FLAG_MOBIL</b>	NaN	NaN	NaN	
<b>FLAG_EMP_PHONE</b>	-0.001795	0.046090	0.240840	5.85
<b>FLAG_WORK_PHONE</b>	-0.000288	0.027286	0.055061	-1.51
<b>FLAG_CONT_MOBILE</b>	0.004535	-0.000444	-0.001790	-7.04
<b>FLAG_PHONE</b>	0.003330	-0.023746	-0.030352	4.60
<b>FLAG_EMAIL</b>	-0.000977	-0.002025	0.024006	3.48
<b>CNT_FAM_MEMBERS</b>	-0.001523	0.009221	0.878782	1.56
<b>REGION_RATING_CLIENT</b>	0.000832	0.059261	0.023579	-7.72
<b>REGION_RATING_CLIENT_W_CITY</b>	0.000484	0.061620	0.022976	-8.27
<b>HOUR_APPR_PROCESS_START</b>	-0.001190	-0.023045	-0.006879	3.30
<b>REG_REGION_NOT_LIVE_REGION</b>	0.000649	0.006833	-0.013293	2.86
<b>REG_REGION_NOT_WORK_REGION</b>	0.001446	0.006569	0.007732	5.63
<b>LIVE_REGION_NOT_WORK_REGION</b>	0.002685	0.001840	0.014188	5.24
<b>REG_CITY_NOT_LIVE_CITY</b>	-0.001328	0.045750	0.021526	3.29
<b>REG_CITY_NOT_WORK_CITY</b>	0.000321	0.050449	0.071205	5.66
<b>LIVE_CITY_NOT_WORK_CITY</b>	0.001339	0.031242	0.069489	7.32
<b>EXT_SOURCE_2</b>	0.001423	-0.161406	-0.017794	5.43
<b>OBS_30_CNT_SOCIAL_CIRCLE</b>	-0.002065	0.007796	0.013611	-1.21
<b>DEF_30_CNT_SOCIAL_CIRCLE</b>	-0.002073	0.029875	-0.001744	-1.22
<b>OBS_60_CNT_SOCIAL_CIRCLE</b>	-0.002114	0.007724	0.013172	-1.21
<b>DEF_60_CNT_SOCIAL_CIRCLE</b>	-0.000810	0.029540	-0.002711	-1.18

	SK_ID_CURR	TARGET	CNT_CHILDREN	AMT_INCOM
<b>DAY_S_LAST_PHONE_CHANGE</b>	0.001011	0.055199	-0.005703	-1.75
<b>FLAG_DOCUMENT_2</b>	0.001507	0.005120	0.002529	-1.22
<b>FLAG_DOCUMENT_3</b>	-0.003971	0.043626	0.056813	-1.41
<b>FLAG_DOCUMENT_4</b>	-0.004471	-0.002765	-0.004746	3.23
<b>FLAG_DOCUMENT_5</b>	-0.001306	-0.001400	-0.018136	1.49
<b>FLAG_DOCUMENT_6</b>	0.001566	-0.027862	-0.156985	-4.13
<b>FLAG_DOCUMENT_7</b>	-0.002311	-0.001065	-0.000262	5.89
<b>FLAG_DOCUMENT_8</b>	0.002357	-0.007932	0.052430	6.58
<b>FLAG_DOCUMENT_9</b>	0.001427	-0.003634	-0.002423	1.64
<b>FLAG_DOCUMENT_10</b>	0.001074	-0.000853	-0.001654	-1.77
<b>FLAG_DOCUMENT_11</b>	-0.001904	-0.004872	-0.004547	2.40
<b>FLAG_DOCUMENT_12</b>	-0.001177	-0.000853	0.000335	2.59
<b>FLAG_DOCUMENT_13</b>	-0.000894	-0.011751	0.004550	2.01
<b>FLAG_DOCUMENT_14</b>	-0.002353	-0.009688	-0.005643	1.91
<b>FLAG_DOCUMENT_15</b>	0.002542	-0.006688	0.005512	8.80
<b>FLAG_DOCUMENT_16</b>	-0.000064	-0.010632	0.009178	6.28
<b>FLAG_DOCUMENT_17</b>	0.001448	-0.002975	0.000490	1.66
<b>FLAG_DOCUMENT_18</b>	-0.000178	-0.008617	0.003098	2.72
<b>FLAG_DOCUMENT_19</b>	0.001631	-0.002715	0.001006	2.32
<b>FLAG_DOCUMENT_20</b>	0.000382	-0.000523	0.002597	4.41
<b>FLAG_DOCUMENT_21</b>	0.000427	0.003807	-0.001379	-8.22

```
In [22]: df.replace('Cash loans', np.nan, inplace=True)
```

```
In [ ]: df = df.apply(lambda col: pd.to_numeric(col, errors='coerce') if col.name != 'TA'
```

```
In [ ]: df.replace('Cash loans', np.nan, inplace=True)
```

```
In [ ]: df = df.drop(columns="FLAG_MOBIL")# Drop coluna não relaciona a nada
```

```
In [ ]: df.head()
```

	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLA
0	456162	0		NaN	NaN	NaN
1	134978	0		NaN	NaN	NaN
2	318952	0		NaN	NaN	NaN
3	361264	0		NaN	NaN	NaN
4	260639	0		NaN	NaN	NaN

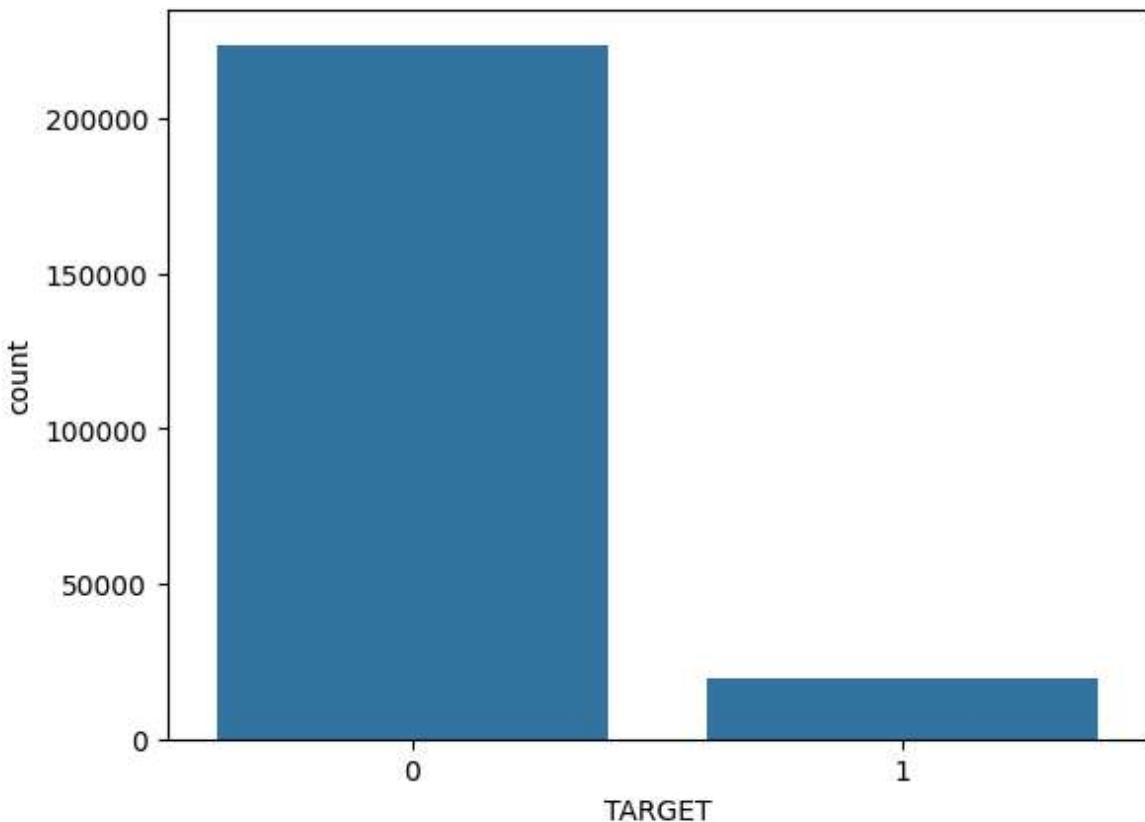
◀ ▶

```
In [ ]: df = pd.read_csv("Aplication_train.csv", index_col=0) # Transformando os dados em
```

```
In [ ]: df.TARGET.value_counts()
```

```
Out[ ]: TARGET
0    223759
1    19838
Name: count, dtype: int64
```

```
In [ ]: ax = sns.countplot(x="TARGET", data=df) #Target antes do balanceamento
```



```
In [ ]: # Redução de dimensionalidade com PCA
pca = PCA(n_components=10) # Ajuste o número de componentes conforme necessário
X_reduced = pca.fit_transform(df.drop(columns=["TARGET"]))
```

```
# Seleção de variáveis com SelectKBest
selector = SelectKBest(f_classif, k=20)
X_selected = selector.fit_transform(df.drop(columns=["TARGET"])), df["TARGET"])
```

```
In [ ]: smt = SMOTE()
```

```
In [ ]: def pipe_pre_process_german(drop_not_features=["Risk"], col_target="Risk",
                                inputer_num_strategy="mean",
                                encoding="ordinal"):
    ...
    - inputer_num_strategy (str): "mean", "median";
    - encoding (str): "onehot" para OneHotEncoder; "ordinal" OrdinalEncoder;
    ...

    df = pd.read_csv("Aplication_train.csv", index_col=0)

    X = df.drop(columns=drop_not_features)
    y = df[col_target]

    X, y = smt.fit_resample(X, y)

    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, ran
# =====

    pipe_features_num = Pipeline([("input_num", SimpleImputer(strategy=inputer_n
                                ("std", StandardScaler()))])

    features_num = X_train.select_dtypes(include=np.number).columns.tolist()
# =====

    if encoding == "onehot":

        pipe_features_cat = Pipeline([("input_cat", SimpleImputer(strategy="cons
                                ("onehot", OneHotEncoder(handle_unknown="i
        elif encoding == "ordinal":

            pipe_features_cat = Pipeline([("input_cat", SimpleImputer(strategy="cons
                                ("ordinal", OrdinalEncoder(handle_unknown=
                                    unknown_value=-
    else:

        raise ValueError("Únicos encodings disponíveis são 'ordinal' e 'onehot'")

    features_cat = X_train.select_dtypes(exclude=np.number).columns.tolist()
# =====

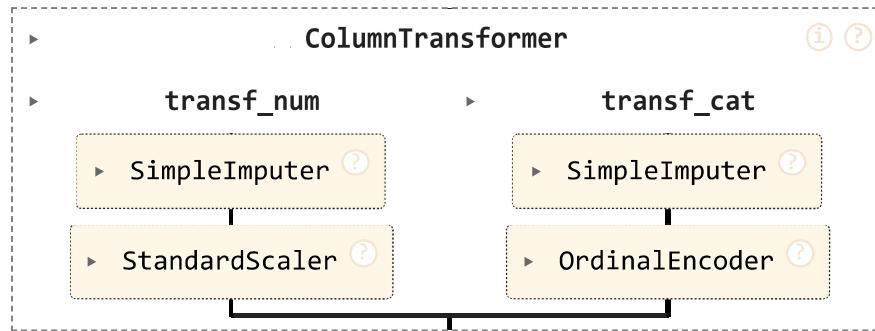
    pre_processador = ColumnTransformer([("transf_num", pipe_features_num, featu
                                ("transf_cat", pipe_features_cat, featu

    return X_train, X_test, y_train, y_test, pre_processador
```

```
In [ ]: X_train, X_test, y_train, y_test, pre_processador = pipe_pre_process_german(drop_
col_
inpu
enco
```

```
In [ ]: pre_processador
```

Out[ ]:

In [ ]: `def metricas_classificacao(estimator, X_train, X_test, y_train, y_test):`

```

# =====

print("\nMétricas de avaliação de treino:")

y_pred_train = estimator.predict(X_train)

ConfusionMatrixDisplay.from_predictions(y_train, y_pred_train)
plt.show()

print(classification_report(y_train, y_pred_train))

# =====

print("\nMétricas de avaliação de teste:")

y_pred_test = estimator.predict(X_test)

ConfusionMatrixDisplay.from_predictions(y_test, y_pred_test)
plt.show()

print(classification_report(y_test, y_pred_test))

# =====

f1_score_weighted = f1_score(y_test, y_pred_test, average='weighted')
print(f"F1-score weighted: {100*f1_score_weighted:.2f}")
return f1_score_weighted
  
```

In [ ]: `pipe = Pipeline([("pp", pre_processador), ("gb", GradientBoostingClassifier(random_state=42))])`

```

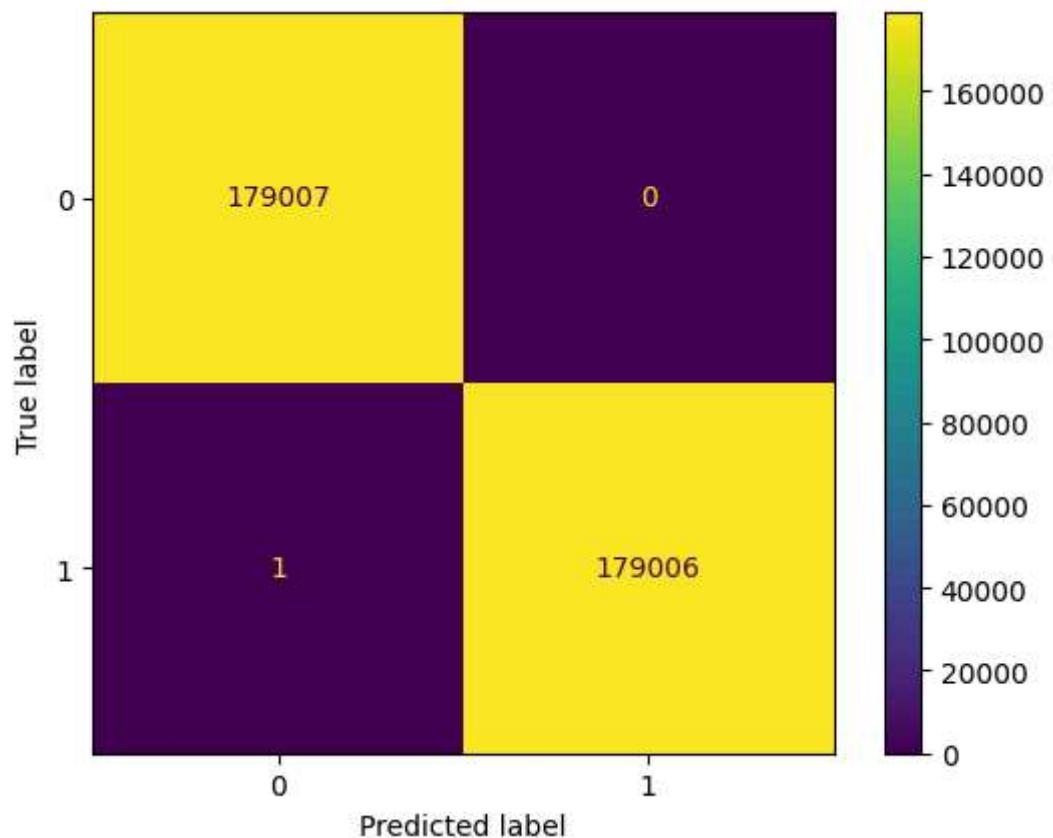
modelos = {
    "Random Forest": RandomForestClassifier(random_state=42),
    "Logistic Regression": LogisticRegression(max_iter=1000),
    "XGBoost": XGBClassifier(use_label_encoder=False, eval_metric='logloss')
}
# Testando diferentes modelos
for nome, modelo in modelos.items():
    print(f"\nModelo: {nome}")
    pipe = Pipeline([("pp", pre_processador), ("modelo", modelo)])
    pipe.fit(X_train, y_train)
    metricas_classificacao(pipe, X_train, X_test, y_train, y_test)

pipe.fit(X_train, y_train)

score = metricas_classificacao(pipe, X_train, X_test, y_train, y_test)
  
```

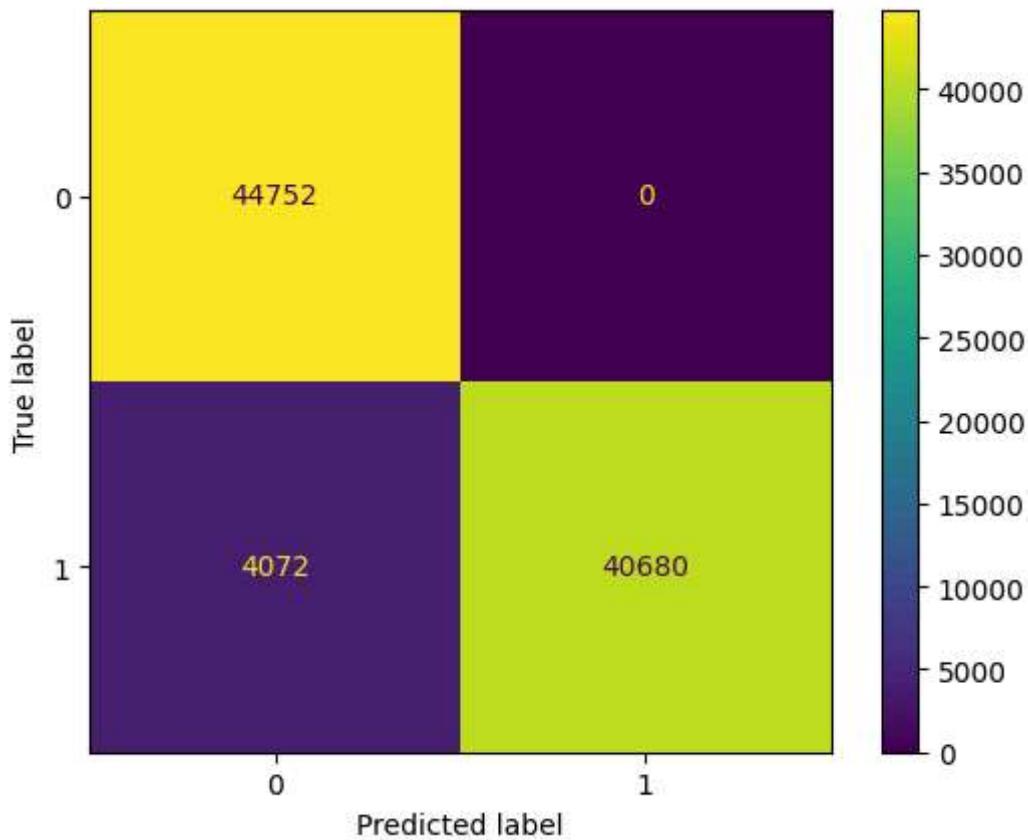
Modelo: Random Forest

Métricas de avaliação de treino:



	precision	recall	f1-score	support
0	1.00	1.00	1.00	179007
1	1.00	1.00	1.00	179007
accuracy			1.00	358014
macro avg	1.00	1.00	1.00	358014
weighted avg	1.00	1.00	1.00	358014

Métricas de avaliação de teste:

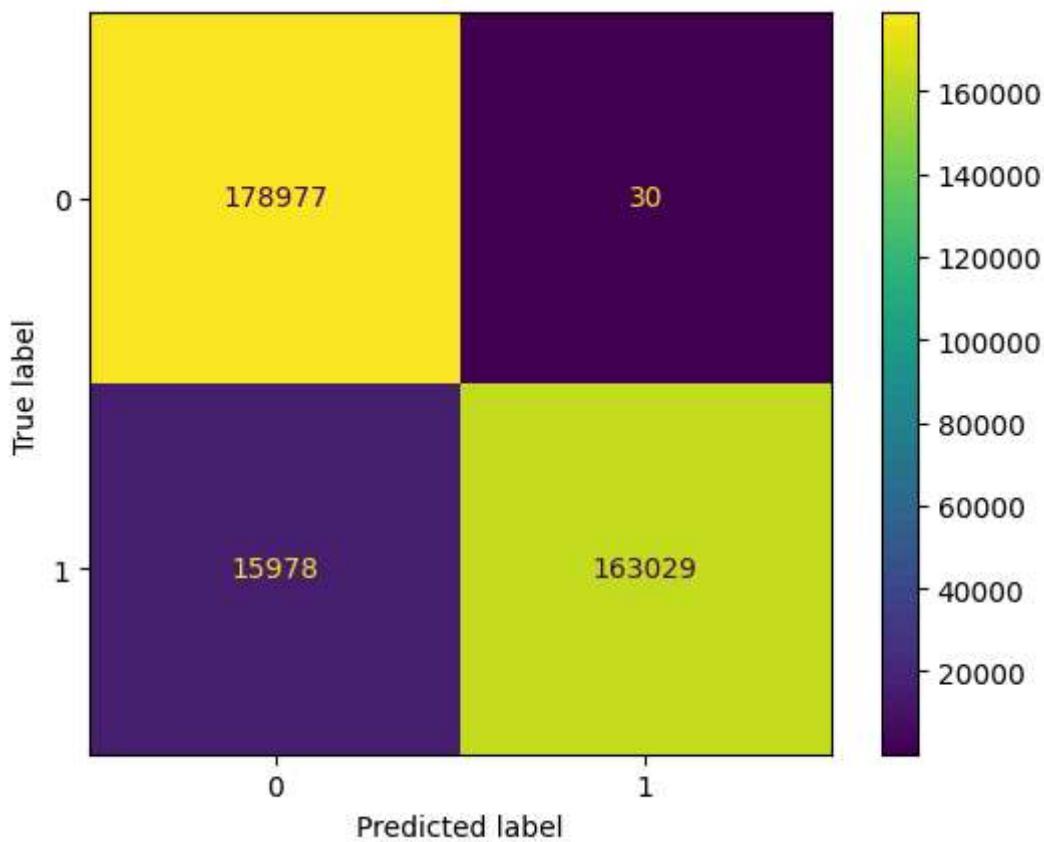


	precision	recall	f1-score	support
0	0.92	1.00	0.96	44752
1	1.00	0.91	0.95	44752
accuracy			0.95	89504
macro avg	0.96	0.95	0.95	89504
weighted avg	0.96	0.95	0.95	89504

F1-score weighted: 95.44

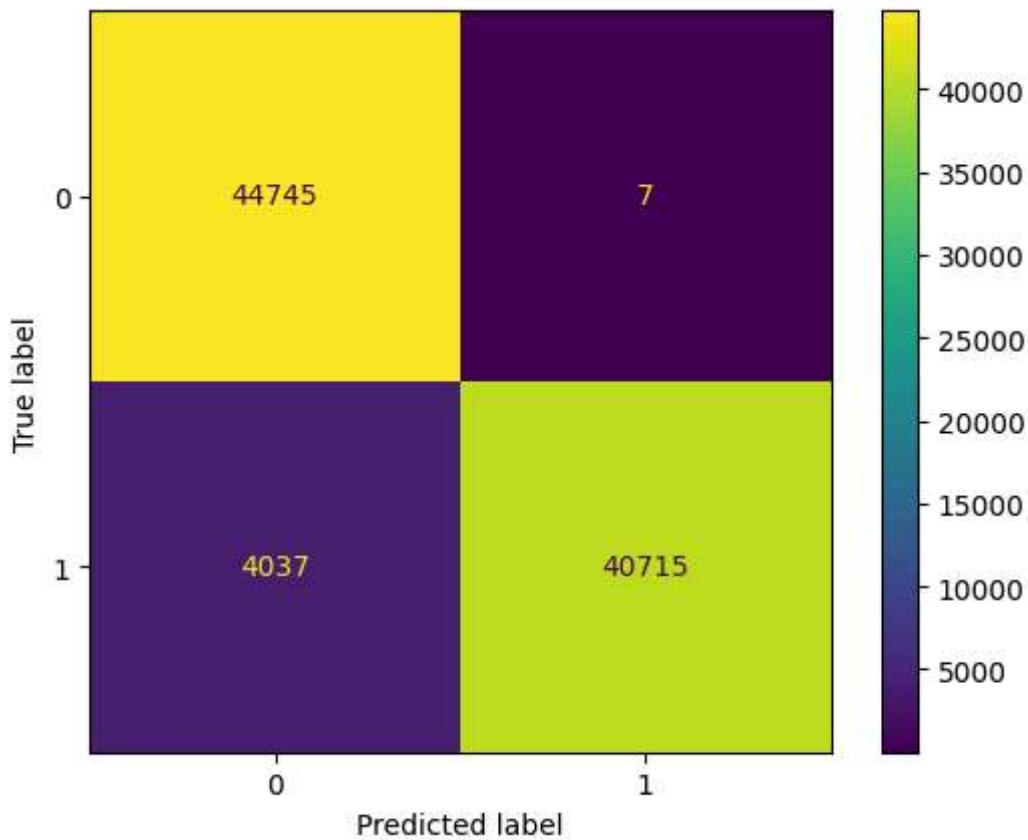
Modelo: Logistic Regression

Métricas de avaliação de treino:



	precision	recall	f1-score	support
0	0.92	1.00	0.96	179007
1	1.00	0.91	0.95	179007
accuracy			0.96	358014
macro avg	0.96	0.96	0.96	358014
weighted avg	0.96	0.96	0.96	358014

Métricas de avaliação de teste:



	precision	recall	f1-score	support
0	0.92	1.00	0.96	44752
1	1.00	0.91	0.95	44752
accuracy			0.95	89504
macro avg	0.96	0.95	0.95	89504
weighted avg	0.96	0.95	0.95	89504

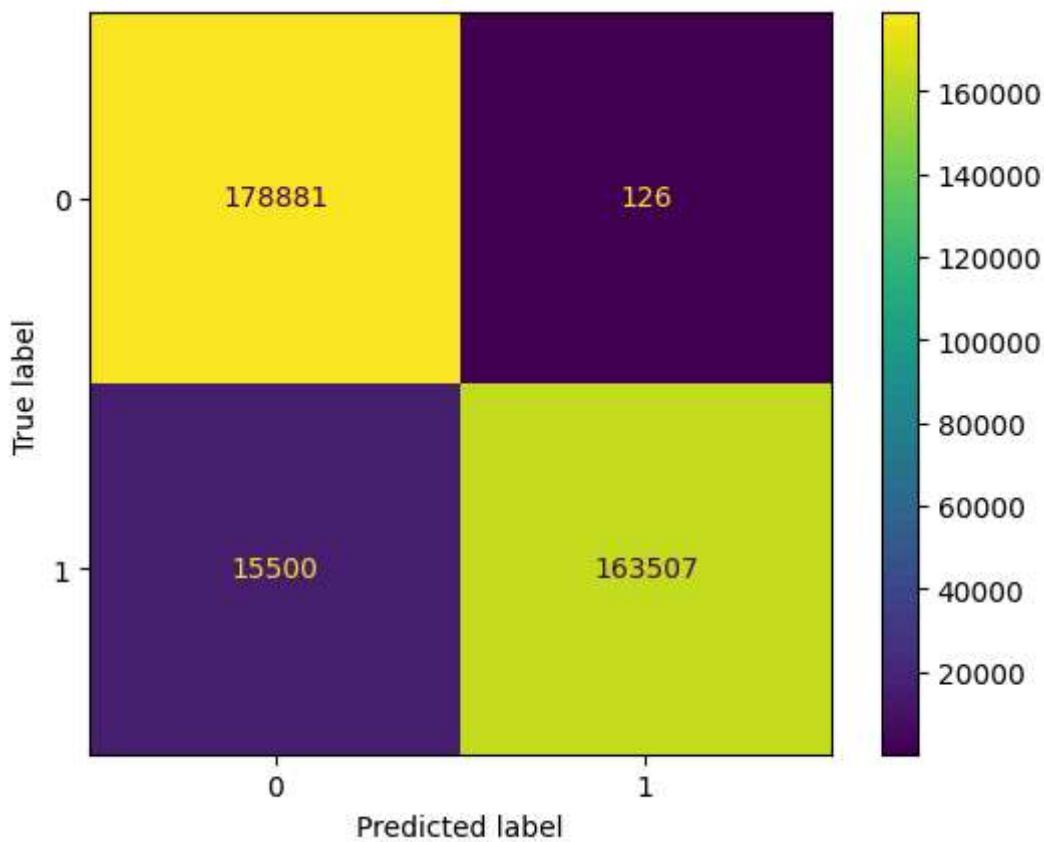
F1-score weighted: 95.47

Modelo: XGBoost

```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:39:25] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.
```

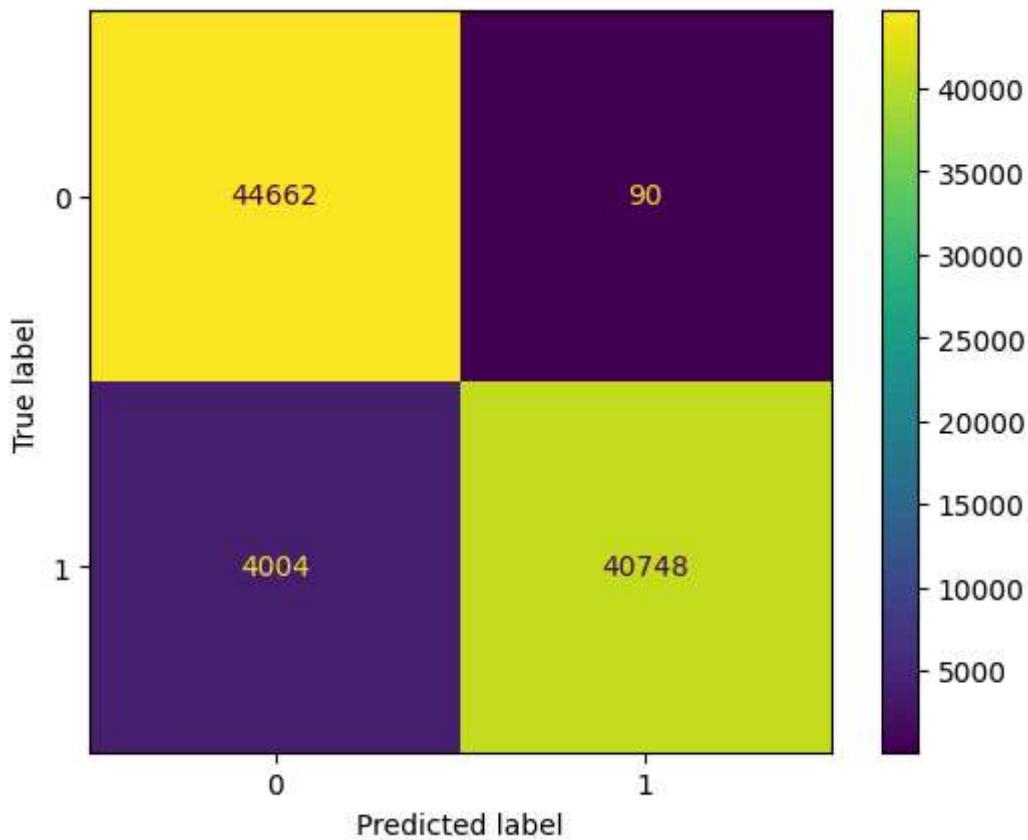
```
warnings.warn(msg, UserWarning)
```

Métricas de avaliação de treino:



	precision	recall	f1-score	support
0	0.92	1.00	0.96	179007
1	1.00	0.91	0.95	179007
accuracy			0.96	358014
macro avg	0.96	0.96	0.96	358014
weighted avg	0.96	0.96	0.96	358014

Métricas de avaliação de teste:



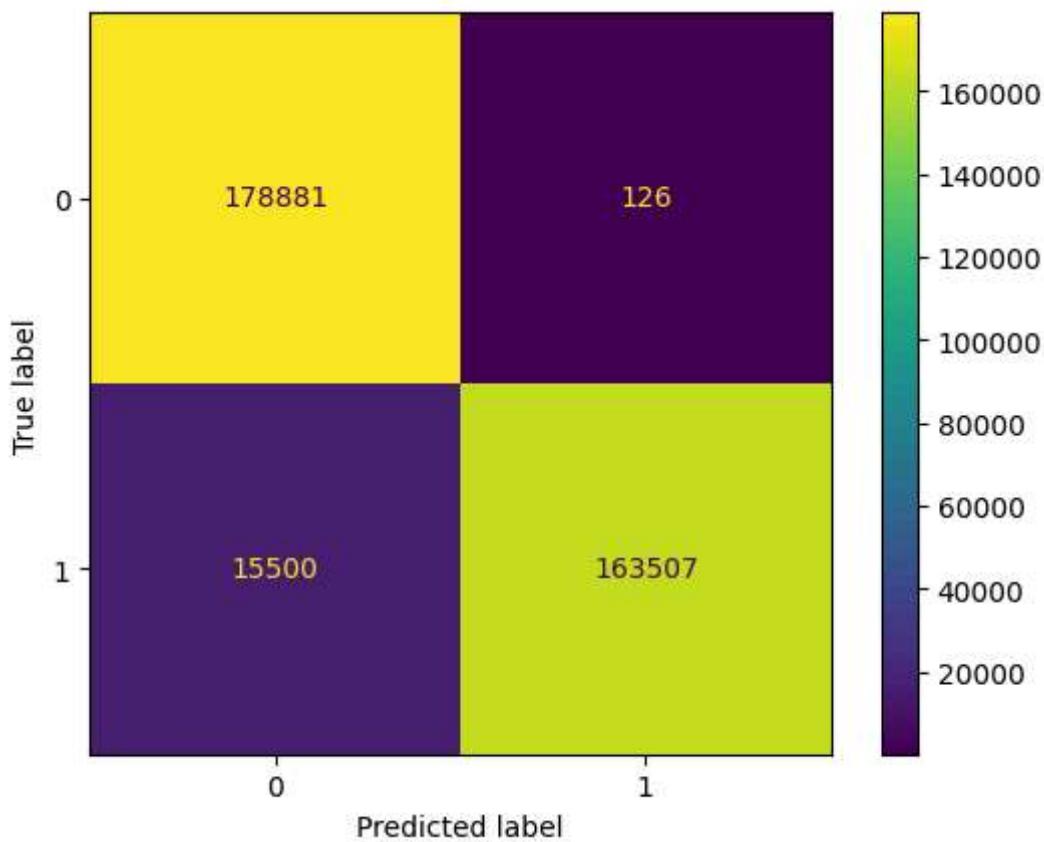
	precision	recall	f1-score	support
0	0.92	1.00	0.96	44752
1	1.00	0.91	0.95	44752
accuracy			0.95	89504
macro avg	0.96	0.95	0.95	89504
weighted avg	0.96	0.95	0.95	89504

F1-score weighted: 95.42

```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:39:35] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.
```

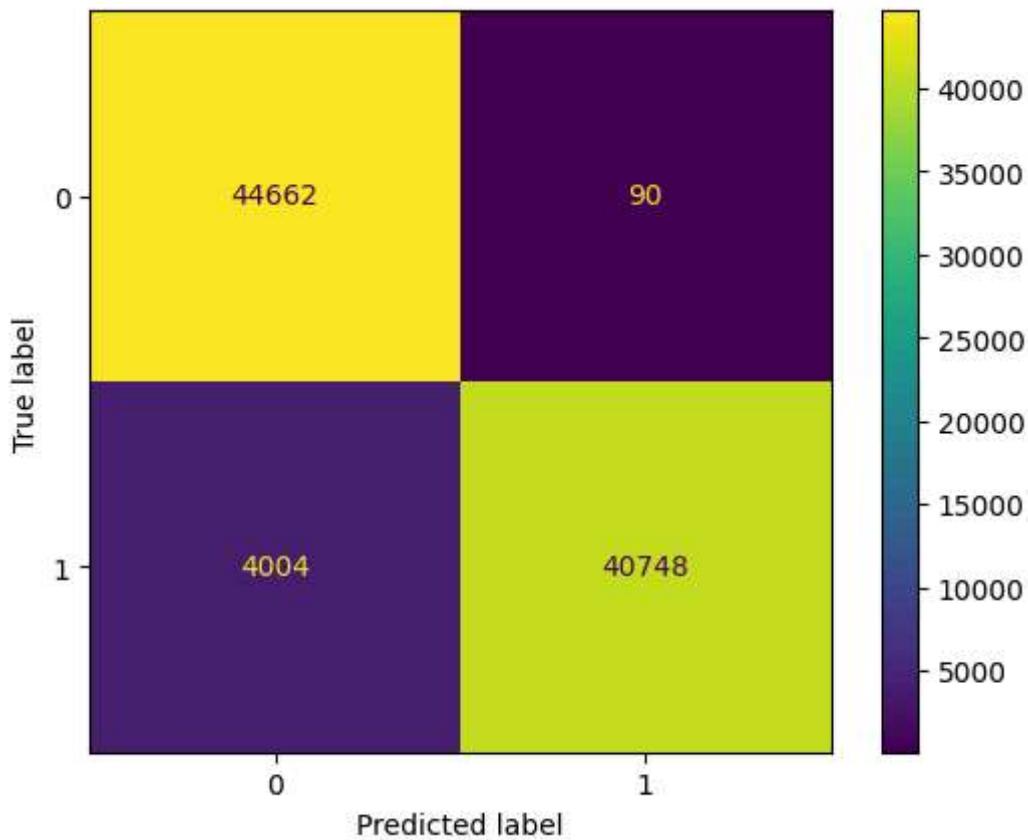
```
warnings.warn(smsg, UserWarning)
```

Métricas de avaliação de treino:



	precision	recall	f1-score	support
0	0.92	1.00	0.96	179007
1	1.00	0.91	0.95	179007
accuracy			0.96	358014
macro avg	0.96	0.96	0.96	358014
weighted avg	0.96	0.96	0.96	358014

Métricas de avaliação de teste:



	precision	recall	f1-score	support
0	0.92	1.00	0.96	44752
1	1.00	0.91	0.95	44752
accuracy			0.95	89504
macro avg	0.96	0.95	0.95	89504
weighted avg	0.96	0.95	0.95	89504

F1-score weighted: 95.42

```
In [ ]: from sklearn.model_selection import cross_val_score #Validação cruzada
scores = cross_val_score(pipe, X_train, y_train, cv=5, scoring='accuracy')
print("Acurácia média com validação cruzada:", scores.mean())
```

```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:40:16] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:40:21] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:40:27] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:40:33] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:40:39] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
Acurácia média com validação cruzada: 0.9549235511608455
```

```
In [ ]: from sklearn.model_selection import GridSearchCV
```

```
# Hiperparâmetros para busca
param_grid = {
    'modelo_n_estimators': [100, 200],
    'modelo_max_depth': [3, 5, 7],
    'modelo_learning_rate': [0.01, 0.1, 0.2]
}

grid_search = GridSearchCV(pipe, param_grid, cv=3, scoring='accuracy')
grid_search.fit(X_train, y_train)
print("Melhores parâmetros:", grid_search.best_params_)
print("Melhor score:", grid_search.best_score_)
```

```
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:56:24] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:56:29] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:56:34] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:56:39] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:56:45] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:56:51] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:56:57] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:57:03] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:57:08] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
```

```
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:57:14] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:57:21] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:57:28] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:57:35] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:57:42] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:57:49] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:57:55] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:58:04] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
```

```
\core.py:158: UserWarning: [21:58:13] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:58:23] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:58:27] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:58:32] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:58:36] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:58:42] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:58:48] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:58:54] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:58:59] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.
```

```
    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:59:04] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:59:09] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:59:16] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:59:22] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:59:29] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:59:35] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:59:40] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:59:46] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [21:59:53] WARNING: C:\buildkite-agent\builds\buildkit
```

```
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:00:01] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:00:08] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:00:12] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:00:17] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:00:22] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:00:28] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:00:34] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:00:40] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.
```

```
    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:00:45] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:00:50] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:00:54] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:01:01] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:01:07] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:01:14] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

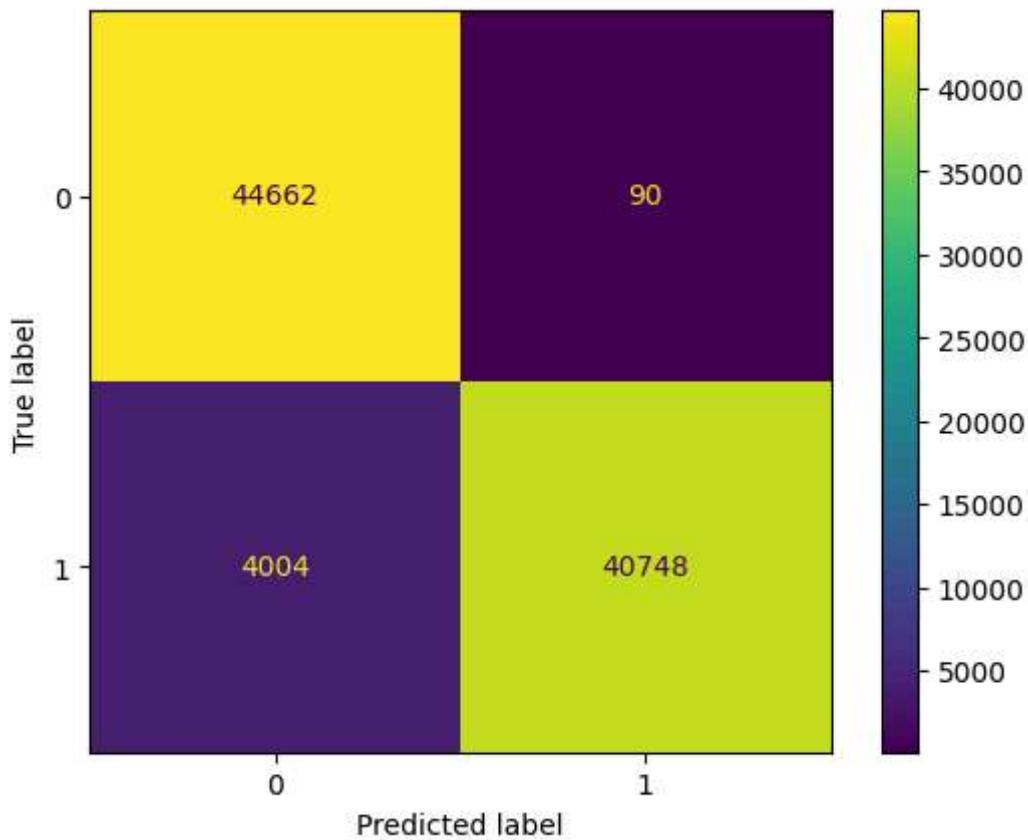
    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:01:19] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:01:25] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
\src\learner.cc:740:
Parameters: { "use_label_encoder" } are not used.

    warnings.warn(smsg, UserWarning)
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost
\core.py:158: UserWarning: [22:01:30] WARNING: C:\buildkite-agent\builds\buildkit
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows
```

```
\src\learner.cc:740:  
Parameters: { "use_label_encoder" } are not used.  
  
    warnings.warn(smsg, UserWarning)  
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost  
\core.py:158: UserWarning: [22:01:37] WARNING: C:\buildkite-agent\builds\buildkit  
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows  
\src\learner.cc:740:  
Parameters: { "use_label_encoder" } are not used.  
  
    warnings.warn(smsg, UserWarning)  
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost  
\core.py:158: UserWarning: [22:01:44] WARNING: C:\buildkite-agent\builds\buildkit  
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows  
\src\learner.cc:740:  
Parameters: { "use_label_encoder" } are not used.  
  
    warnings.warn(smsg, UserWarning)  
c:\Users\wiltd\AppData\Local\Programs\Python\Python312\Lib\site-packages\xgboost  
\core.py:158: UserWarning: [22:01:52] WARNING: C:\buildkite-agent\builds\buildkit  
e-windows-cpu-autoscaling-group-i-0ed59c031377d09b8-1\xgboost\xgboost-ci-windows  
\src\learner.cc:740:  
Parameters: { "use_label_encoder" } are not used.  
  
    warnings.warn(smsg, UserWarning)  
Melhores parâmetros: {'modelo_learning_rate': 0.1, 'modelo_max_depth': 7, 'mode  
lo_n_estimators': 200}  
Melhor score: 0.9550380711368828
```

```
In [ ]: from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay, classifica  
  
# Exibir a matriz de confusão  
y_pred = pipe.predict(X_test)  
cm = confusion_matrix(y_test, y_pred)  
ConfusionMatrixDisplay(cm).plot()  
plt.show()  
  
# Exibir o relatório de classificação  
print(classification_report(y_test, y_pred))
```



	precision	recall	f1-score	support
0	0.92	1.00	0.96	44752
1	1.00	0.91	0.95	44752
accuracy			0.95	89504
macro avg	0.96	0.95	0.95	89504
weighted avg	0.96	0.95	0.95	89504

```
In [ ]: ns_probs = [0 for _ in range(len(y_test))]

# predict probabilities

lr_probs = pipe.predict_proba(X_test)

# keep probabilities for the positive outcome only

lr_probs = lr_probs[:, 1]

# calculate scores

ns_auc = roc_auc_score(y_test, ns_probs)
lr_auc = roc_auc_score(y_test, lr_probs)

# summarize scores

print('padrão: ROC AUC=%3f' % (ns_auc))
print('gb: ROC AUC=%3f' % (lr_auc))
ns_fpr, ns_tpr, _ = roc_curve(y_test, ns_probs)
lr_fpr, lr_tpr, _ = roc_curve(y_test, lr_probs)

# plot the roc curve for the model
```

```
plt.plot(ns_fpr, ns_tpr, linestyle='--', label='padrão')
plt.plot(lr_fpr, lr_tpr, marker='.', label='gb')

# axis Labels

plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')

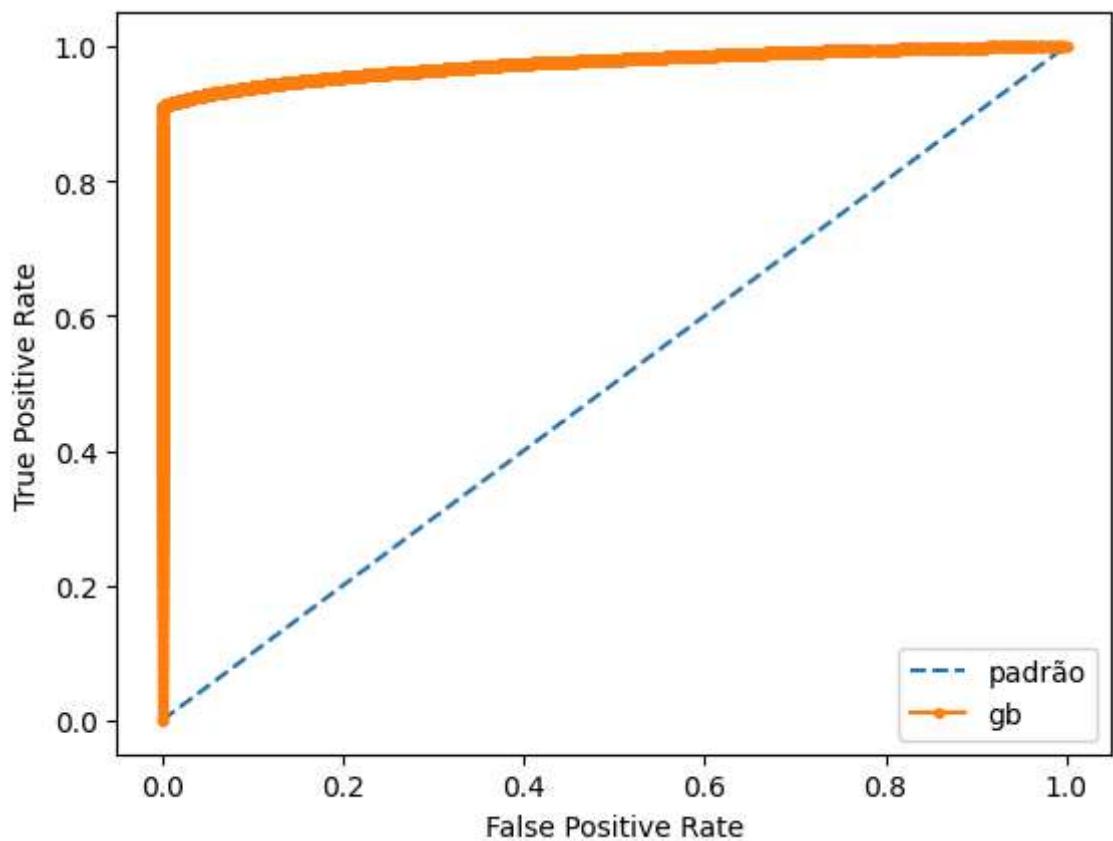
# show the legend

plt.legend()

# show the plot

plt.show()
```

padrão: ROC AUC=0.500  
gb: ROC AUC=0.974



In [ ]: