Importing Libraries and Dataset

```
# importing the required libraries/Modules

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

application_data = pd.read_csv("application_data.csv")
previous_application = pd.read_csv("previous_application.csv")
```

Dataset Explore

```
application_data.head()

   SK_ID_CURR  TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR  \
0      100002       1         Cash loans           M            N
1      100003       0         Cash loans           F            N
2      100004       0    Revolving loans           M            Y
3      100006       0         Cash loans           F            N
4      100007       0         Cash loans           M            N

  FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT
AMT_ANNUITY  \
0               Y             0          202500.0    406597.5
24700.5
1               N             0          270000.0   1293502.5
35698.5
2               Y             0           67500.0    135000.0
6750.0
3               Y             0          135000.0    312682.5
29686.5
4               Y             0          121500.0    513000.0
21865.5

   ...  FLAG_DOCUMENT_18 FLAG_DOCUMENT_19 FLAG_DOCUMENT_20
FLAG_DOCUMENT_21  \
0  ...                0                0                0
0
1  ...                0                0                0
0
2  ...                0                0                0
0
3  ...                0                0                0
0
```

```
4  ...                   0                 0                 0
0

  AMT_REQ_CREDIT_BUREAU_HOUR AMT_REQ_CREDIT_BUREAU_DAY  \
0                        0.0                       0.0
1                        0.0                       0.0
2                        0.0                       0.0
3                        NaN                       NaN
4                        0.0                       0.0

   AMT_REQ_CREDIT_BUREAU_WEEK  AMT_REQ_CREDIT_BUREAU_MON  \
0                         0.0                        0.0
1                         0.0                        0.0
2                         0.0                        0.0
3                         NaN                        NaN
4                         0.0                        0.0

   AMT_REQ_CREDIT_BUREAU_QRT  AMT_REQ_CREDIT_BUREAU_YEAR
0                        0.0                         1.0
1                        0.0                         0.0
2                        0.0                         0.0
3                        NaN                         NaN
4                        0.0                         0.0

[5 rows x 122 columns]
```

```
application_data.tail()
```

```
       SK_ID_CURR  TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR
\
49994      157871       0         Cash loans           F            N

49995      157872       0         Cash loans           M            N

49996      157873       0         Cash loans           M            N

49997      157874       0         Cash loans           F            N

49998      157875       0         Cash loans           F            N


      FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT  \
49994               N             0          180000.0   1206000.0
49995               N             0          126000.0   1125000.0
49996               N             1          112500.0    900000.0
49997               Y             0          270000.0    820638.0
49998               Y             0          117000.0    254700.0

       AMT_ANNUITY  ...  FLAG_DOCUMENT_18 FLAG_DOCUMENT_19
FLAG_DOCUMENT_20  \
49994      45936.0  ...                 0                 0
```

```
0
49995      47794.5  ...                 0                0
0
49996      26316.0  ...                 0                0
0
49997      34897.5  ...                 0                0
0
49998      14751.0  ...                 0                0
0

      FLAG_DOCUMENT_21 AMT_REQ_CREDIT_BUREAU_HOUR
AMT_REQ_CREDIT_BUREAU_DAY  \
49994                0                        0.0
0.0
49995                0                        0.0
0.0
49996                0                        0.0
0.0
49997                0                        0.0
0.0
49998                0                        0.0
0.0

       AMT_REQ_CREDIT_BUREAU_WEEK  AMT_REQ_CREDIT_BUREAU_MON  \
49994                         0.0                        0.0
49995                         0.0                        0.0
49996                         0.0                        0.0
49997                         0.0                        0.0
49998                         0.0                        0.0

       AMT_REQ_CREDIT_BUREAU_QRT  AMT_REQ_CREDIT_BUREAU_YEAR
49994                        0.0                         0.0
49995                        0.0                         0.0
49996                        0.0                         2.0
49997                        2.0                         4.0
49998                        0.0                         0.0

[5 rows x 122 columns]
```

```
application_data.columns
```

```
Index(['SK_ID_CURR', 'TARGET', 'NAME_CONTRACT_TYPE', 'CODE_GENDER',
       'FLAG_OWN_CAR', 'FLAG_OWN_REALTY', 'CNT_CHILDREN',
'AMT_INCOME_TOTAL',
       'AMT_CREDIT', 'AMT_ANNUITY',
       ...
       'FLAG_DOCUMENT_18', 'FLAG_DOCUMENT_19', 'FLAG_DOCUMENT_20',
       'FLAG_DOCUMENT_21', 'AMT_REQ_CREDIT_BUREAU_HOUR',
       'AMT_REQ_CREDIT_BUREAU_DAY', 'AMT_REQ_CREDIT_BUREAU_WEEK',
       'AMT_REQ_CREDIT_BUREAU_MON', 'AMT_REQ_CREDIT_BUREAU_QRT',
```

```
       'AMT_REQ_CREDIT_BUREAU_YEAR'],
      dtype='object', length=122)
```

```
application_data.shape
```

```
(49999, 122)
```

```
application_data.describe()
```

```
          SK_ID_CURR        TARGET  CNT_CHILDREN  AMT_INCOME_TOTAL  \
count   49999.000000  49999.000000  49999.000000      4.999900e+04   
mean   129013.210584      0.080522      0.419848      1.707676e+05   
std     16690.512048      0.272102      0.724039      5.318191e+05   
min    100002.000000      0.000000      0.000000      2.565000e+04   
25%    114570.500000      0.000000      0.000000      1.125000e+05   
50%    129076.000000      0.000000      0.000000      1.458000e+05   
75%    143438.500000      0.000000      1.000000      2.025000e+05   
max    157875.000000      1.000000     11.000000      1.170000e+08   

         AMT_CREDIT    AMT_ANNUITY  AMT_GOODS_PRICE  \
count  4.999900e+04   49998.000000     4.996100e+04   
mean   5.997006e+05   27107.377355     5.390600e+05   
std    4.024154e+05   14562.944435     3.698533e+05   
min    4.500000e+04    2052.000000     4.500000e+04   
25%    2.700000e+05   16456.500000     2.385000e+05   
50%    5.147775e+05   24939.000000     4.500000e+05   
75%    8.086500e+05   34596.000000     6.795000e+05   
max    4.050000e+06  258025.500000     4.050000e+06   

       REGION_POPULATION_RELATIVE    DAYS_BIRTH  DAYS_EMPLOYED  ...  \
count                49999.000000  49999.000000   49999.000000  ...   
mean                     0.020798 -16022.042081   63219.424488  ...   
std                      0.013761   4361.400270  140794.605668  ...   
min                      0.000533 -25184.000000  -17531.000000  ...   
25%                      0.010006 -19644.000000   -2786.000000  ...   
50%                      0.018850 -15731.000000   -1221.000000  ...   
75%                      0.028663 -12378.500000    -292.000000  ...   
max                      0.072508  -7680.000000  365243.000000  ...   

       FLAG_DOCUMENT_18  FLAG_DOCUMENT_19  FLAG_DOCUMENT_20
FLAG_DOCUMENT_21  \
count      49999.000000      49999.000000      49999.000000
49999.00000   
mean           0.008500          0.000700          0.000520
0.00038   
std            0.091805          0.026449          0.022798
0.01949   
min            0.000000          0.000000          0.000000
0.00000   
25%            0.000000          0.000000          0.000000
```

```
0.00000
50%            0.000000          0.000000          0.000000
0.00000
75%            0.000000          0.000000          0.000000
0.00000
max            1.000000          1.000000          1.000000
1.00000

       AMT_REQ_CREDIT_BUREAU_HOUR  AMT_REQ_CREDIT_BUREAU_DAY  \
count                43265.000000               43265.000000
mean                     0.007096                   0.007512
std                      0.087709                   0.107992
min                      0.000000                   0.000000
25%                      0.000000                   0.000000
50%                      0.000000                   0.000000
75%                      0.000000                   0.000000
max                      3.000000                   6.000000

       AMT_REQ_CREDIT_BUREAU_WEEK  AMT_REQ_CREDIT_BUREAU_MON  \
count                43265.000000               43265.000000
mean                     0.032382                   0.270288
std                      0.194080                   0.928560
min                      0.000000                   0.000000
25%                      0.000000                   0.000000
50%                      0.000000                   0.000000
75%                      0.000000                   0.000000
max                      6.000000                  24.000000

       AMT_REQ_CREDIT_BUREAU_QRT  AMT_REQ_CREDIT_BUREAU_YEAR
count               43265.000000                43265.000000
mean                    0.260973                    1.881035
std                     0.606996                    1.865054
min                     0.000000                    0.000000
25%                     0.000000                    0.000000
50%                     0.000000                    1.000000
75%                     0.000000                    3.000000
max                     8.000000                   25.000000

[8 rows x 106 columns]
```

```
application_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 49999 entries, 0 to 49998
Columns: 122 entries, SK_ID_CURR to AMT_REQ_CREDIT_BUREAU_YEAR
dtypes: float64(64), int64(42), object(16)
memory usage: 46.5+ MB
```

```
previous_application.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 49999 entries, 0 to 49998
Data columns (total 37 columns):
 #   Column                       Non-Null Count  Dtype  
---  ------                       --------------  -----  
 0   SK_ID_PREV                   49999 non-null  int64  
 1   SK_ID_CURR                   49999 non-null  int64  
 2   NAME_CONTRACT_TYPE           49999 non-null  object 
 3   AMT_ANNUITY                  39407 non-null  float64
 4   AMT_APPLICATION              49999 non-null  float64
 5   AMT_CREDIT                   49999 non-null  float64
 6   AMT_DOWN_PAYMENT             24801 non-null  float64
 7   AMT_GOODS_PRICE              39255 non-null  float64
 8   WEEKDAY_APPR_PROCESS_START   49999 non-null  object 
 9   HOUR_APPR_PROCESS_START      49999 non-null  int64  
 10  FLAG_LAST_APPL_PER_CONTRACT  49999 non-null  object 
 11  NFLAG_LAST_APPL_IN_DAY       49999 non-null  int64  
 12  RATE_DOWN_PAYMENT            24801 non-null  float64
 13  RATE_INTEREST_PRIMARY        165 non-null    float64
 14  RATE_INTEREST_PRIVILEGED     165 non-null    float64
 15  NAME_CASH_LOAN_PURPOSE       49999 non-null  object 
 16  NAME_CONTRACT_STATUS         49999 non-null  object 
 17  DAYS_DECISION                49999 non-null  int64  
 18  NAME_PAYMENT_TYPE            49999 non-null  object 
 19  CODE_REJECT_REASON           49999 non-null  object 
 20  NAME_TYPE_SUITE              25756 non-null  object 
 21  NAME_CLIENT_TYPE             49999 non-null  object 
 22  NAME_GOODS_CATEGORY          49999 non-null  object 
 23  NAME_PORTFOLIO               49999 non-null  object 
 24  NAME_PRODUCT_TYPE            49999 non-null  object 
 25  CHANNEL_TYPE                 49999 non-null  object 
 26  SELLERPLACE_AREA             49999 non-null  int64  
 27  NAME_SELLER_INDUSTRY         49999 non-null  object 
 28  CNT_PAYMENT                  39407 non-null  float64
 29  NAME_YIELD_GROUP             49999 non-null  object 
 30  PRODUCT_COMBINATION          49991 non-null  object 
 31  DAYS_FIRST_DRAWING           30839 non-null  float64
 32  DAYS_FIRST_DUE               30839 non-null  float64
 33  DAYS_LAST_DUE_1ST_VERSION    30839 non-null  float64
 34  DAYS_LAST_DUE                30839 non-null  float64
 35  DAYS_TERMINATION             30839 non-null  float64
 36  NFLAG_INSURED_ON_APPROVAL    30839 non-null  float64
dtypes: float64(15), int64(6), object(16)
memory usage: 14.1+ MB
```

```
previous_application.describe()
```

```
         SK_ID_PREV     SK_ID_CURR   AMT_ANNUITY  AMT_APPLICATION  \
count  4.999900e+04   49999.000000  39407.000000     4.999900e+04   
mean   1.922254e+06  278983.187604  15482.596847     1.688925e+05   
```

```
std    5.351980e+05  102780.124434   14530.971854     2.822035e+05
min    1.000001e+06  100007.000000       0.000000     0.000000e+00
25%    1.457920e+06  189919.500000    6122.835000     2.204550e+04
50%    1.920889e+06  279264.000000   10879.920000     7.155000e+04
75%    2.388632e+06  368527.500000   19669.140000     1.800000e+05
max    2.845367e+06  456254.000000  234478.395000     3.826372e+06

         AMT_CREDIT  AMT_DOWN_PAYMENT  AMT_GOODS_PRICE  \
count  4.999900e+04      2.480100e+04     3.925500e+04
mean   1.885429e+05      6.557571e+03     2.151414e+05
std    3.084736e+05      1.744458e+04     3.024993e+05
min    0.000000e+00      0.000000e+00     0.000000e+00
25%    2.605500e+04      0.000000e+00     4.941000e+04
50%    7.890750e+04      1.566000e+03     1.040175e+05
75%    1.981058e+05      7.875000e+03     2.250000e+05
max    4.104351e+06      1.035000e+06     3.826372e+06

       HOUR_APPR_PROCESS_START  NFLAG_LAST_APPL_IN_DAY
RATE_DOWN_PAYMENT  \
count             49999.000000            49999.000000
24801.000000
mean                 12.478330                0.996500
0.079083
std                   3.333012                0.059058
0.107658
min                   0.000000                0.000000
0.000000
25%                  10.000000                1.000000
0.000000
50%                  12.000000                1.000000
0.049732
75%                  15.000000                1.000000
0.108909
max                  23.000000                1.000000
0.944776

       ...  RATE_INTEREST_PRIVILEGED  DAYS_DECISION  SELLERPLACE_AREA
\
count  ...                165.000000   49999.000000      4.999900e+04

mean   ...                  0.787674    -900.112622      4.016558e+02

std    ...                  0.091985     786.531303      1.793772e+04

min    ...                  0.424419   -2922.000000     -1.000000e+00

25%    ...                  0.715645   -1335.000000     -1.000000e+00

50%    ...                  0.835095    -599.000000      1.000000e+01
```

```
75%    ...                   0.852537    -292.000000      1.000000e+02

max    ...                   0.867336      -2.000000      4.000000e+06


        CNT_PAYMENT  DAYS_FIRST_DRAWING  DAYS_FIRST_DUE  \
count  39407.000000        30839.000000    30839.000000
mean      15.555891       344485.142806    14217.240150
std       13.985174        84683.650627    73348.984383
min        0.000000        -2910.000000    -2891.000000
25%        6.000000       365243.000000    -1642.000000
50%       12.000000       365243.000000     -822.000000
75%       18.000000       365243.000000     -404.000000
max       60.000000       365243.000000   365243.000000

       DAYS_LAST_DUE_1ST_VERSION  DAYS_LAST_DUE  DAYS_TERMINATION  \
count               30839.000000   30839.000000      30839.000000
mean                31528.148611   76724.982101      81666.162586
std                103691.881189  149757.893751     153101.159809
min                 -2800.000000   -2850.000000      -2844.000000
25%                 -1270.000000   -1337.000000      -1293.000000
50%                  -366.000000    -536.000000       -500.000000
75%                   113.000000     -71.000000        -45.000000
max                365243.000000  365243.000000     365243.000000

       NFLAG_INSURED_ON_APPROVAL
count               30839.000000
mean                    0.322352
std                     0.467384
min                     0.000000
25%                     0.000000
50%                     0.000000
75%                     1.000000
max                     1.000000

[8 rows x 21 columns]
```

```
previous_application.head()
```

```
   SK_ID_PREV  SK_ID_CURR NAME_CONTRACT_TYPE  AMT_ANNUITY
AMT_APPLICATION  \
0     2030495      271877     Consumer loans     1730.430
17145.0
1     2802425      108129         Cash loans    25188.615
607500.0
2     2523466      122040         Cash loans    15060.735
112500.0
3     2819243      176158         Cash loans    47041.335
450000.0
4     1784265      202054         Cash loans    31924.395
```

```
337500.0

   AMT_CREDIT  AMT_DOWN_PAYMENT  AMT_GOODS_PRICE
WEEKDAY_APPR_PROCESS_START  \
0     17145.0               0.0          17145.0
SATURDAY
1    679671.0               NaN         607500.0
THURSDAY
2    136444.5               NaN         112500.0
TUESDAY
3    470790.0               NaN         450000.0
MONDAY
4    404055.0               NaN         337500.0
THURSDAY

   HOUR_APPR_PROCESS_START  ... NAME_SELLER_INDUSTRY  CNT_PAYMENT  \
0                       15  ...         Connectivity         12.0
1                       11  ...                  XNA         36.0
2                       11  ...                  XNA         12.0
3                        7  ...                  XNA         12.0
4                        9  ...                  XNA         24.0

   NAME_YIELD_GROUP       PRODUCT_COMBINATION  DAYS_FIRST_DRAWING  \
0            middle  POS mobile with interest            365243.0
1        low_action          Cash X-Sell: low            365243.0
2              high         Cash X-Sell: high            365243.0
3            middle       Cash X-Sell: middle            365243.0
4              high         Cash Street: high                 NaN

  DAYS_FIRST_DUE DAYS_LAST_DUE_1ST_VERSION  DAYS_LAST_DUE
DAYS_TERMINATION  \
0          -42.0                     300.0          -42.0             -
37.0
1         -134.0                     916.0       365243.0
365243.0
2         -271.0                      59.0       365243.0
365243.0
3         -482.0                    -152.0         -182.0            -
177.0
4            NaN                       NaN            NaN
NaN

  NFLAG_INSURED_ON_APPROVAL
0                        0.0
1                        1.0
2                        1.0
3                        1.0
4                        NaN

[5 rows x 37 columns]
```

```
previous_application.tail()
```

```
       SK_ID_PREV  SK_ID_CURR NAME_CONTRACT_TYPE  AMT_ANNUITY  \
49994     1171956      339569         Cash loans          NaN
49995     1904808      363980         Cash loans          NaN
49996     2331005      231295         Cash loans    22176.405
49997     1960897      346691         Cash loans          NaN
49998     1979352      363244         Cash loans    24909.390

       AMT_APPLICATION  AMT_CREDIT  AMT_DOWN_PAYMENT  AMT_GOODS_PRICE
\
49994              0.0         0.0               NaN              NaN

49995              0.0         0.0               NaN              NaN

49996         180000.0    216418.5               NaN         180000.0

49997              0.0         0.0               NaN              NaN

49998         360000.0    409896.0               NaN         360000.0


      WEEKDAY_APPR_PROCESS_START  HOUR_APPR_PROCESS_START  ...  \
49994                   THURSDAY                       11  ...
49995                     FRIDAY                       10  ...
49996                    TUESDAY                       12  ...
49997                  WEDNESDAY                       16  ...
49998                     FRIDAY                       18  ...

      NAME_SELLER_INDUSTRY  CNT_PAYMENT  NAME_YIELD_GROUP  \
49994                  XNA          NaN               XNA
49995                  XNA          NaN               XNA
49996                  XNA         12.0            middle
49997                  XNA          NaN               XNA
49998                  XNA         36.0              high

       PRODUCT_COMBINATION  DAYS_FIRST_DRAWING DAYS_FIRST_DUE  \
49994                 Cash                 NaN            NaN
49995                 Cash                 NaN            NaN
49996  Cash X-Sell: middle            365243.0         -670.0
49997                 Cash                 NaN            NaN
49998    Cash X-Sell: high                 NaN            NaN

      DAYS_LAST_DUE_1ST_VERSION  DAYS_LAST_DUE DAYS_TERMINATION  \
49994                       NaN            NaN              NaN
49995                       NaN            NaN              NaN
49996                    -340.0         -340.0           -338.0
49997                       NaN            NaN              NaN
49998                       NaN            NaN              NaN

      NFLAG_INSURED_ON_APPROVAL
```

```
49994                              NaN
49995                              NaN
49996                              1.0
49997                              NaN
49998                              NaN

[5 rows x 37 columns]
```

```
previous_application.shape
```

```
(49999, 37)
```

A. Identify Missing Data and Deal with it Appropriately

Top Correlations
for Different Scenarios

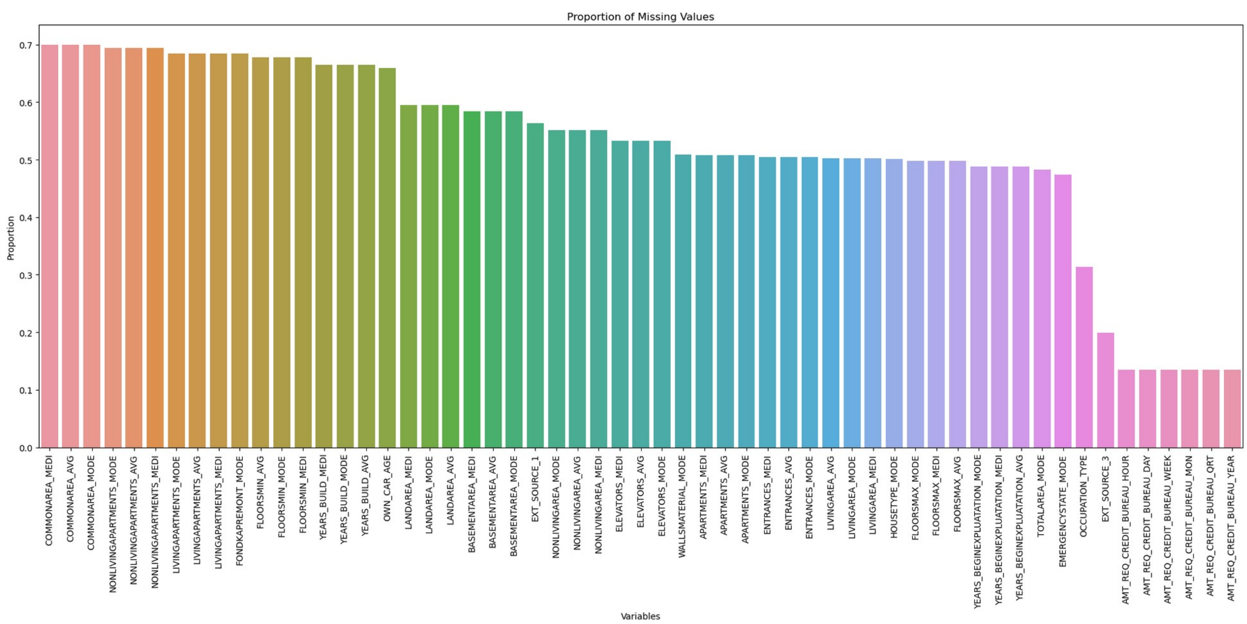
```
a = application_data.isnull().sum()
a
```

```
SK_ID_CURR                        0
TARGET                            0
NAME_CONTRACT_TYPE                0
CODE_GENDER                       0
FLAG_OWN_CAR                      0
                                ...
AMT_REQ_CREDIT_BUREAU_DAY      6734
AMT_REQ_CREDIT_BUREAU_WEEK     6734
AMT_REQ_CREDIT_BUREAU_MON      6734
AMT_REQ_CREDIT_BUREAU_QRT      6734
AMT_REQ_CREDIT_BUREAU_YEAR     6734
Length: 122, dtype: int64
```

Creating bar chart to visualize the proportion of missing values for each variable.

```
# Calculate the proportion of missing values for columns having
missing values more than 0
```

```python
mean_val = application_data.isnull().mean()

# Sorting the mean_val for better visualization
mean_val_sorted = mean_val.sort_values(ascending=False)

missing_var_mean = mean_val_sorted[mean_val_sorted>0.1]

# Create a bar chart using seaborn
plt.figure(figsize=(20, 10))
sns.barplot(x=missing_var_mean.index, y=missing_var_mean.values)
plt.title('Proportion of Missing Values')
plt.xlabel('Variables')
plt.ylabel('Proportion')
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```




```python
a = a[a>(len(application_data)*0.5)]
print(a)
len(a)
```

```
OWN_CAR_AGE                    32950
EXT_SOURCE_1                   28172
APARTMENTS_AVG                 25385
BASEMENTAREA_AVG               29199
YEARS_BUILD_AVG                33239
COMMONAREA_AVG                 34960
ELEVATORS_AVG                  26651
ENTRANCES_AVG                  25195
FLOORSMIN_AVG                  33894
```

```
LANDAREA_AVG                  29721
LIVINGAPARTMENTS_AVG          34226
LIVINGAREA_AVG                25137
NONLIVINGAPARTMENTS_AVG       34714
NONLIVINGAREA_AVG             27572
APARTMENTS_MODE               25385
BASEMENTAREA_MODE             29199
YEARS_BUILD_MODE              33239
COMMONAREA_MODE               34960
ELEVATORS_MODE                26651
ENTRANCES_MODE                25195
FLOORSMIN_MODE                33894
LANDAREA_MODE                 29721
LIVINGAPARTMENTS_MODE         34226
LIVINGAREA_MODE               25137
NONLIVINGAPARTMENTS_MODE      34714
NONLIVINGAREA_MODE            27572
APARTMENTS_MEDI               25385
BASEMENTAREA_MEDI             29199
YEARS_BUILD_MEDI              33239
COMMONAREA_MEDI               34960
ELEVATORS_MEDI                26651
ENTRANCES_MEDI                25195
FLOORSMIN_MEDI                33894
LANDAREA_MEDI                 29721
LIVINGAPARTMENTS_MEDI         34226
LIVINGAREA_MEDI               25137
NONLIVINGAPARTMENTS_MEDI      34714
NONLIVINGAREA_MEDI            27572
FONDKAPREMONT_MODE            34191
HOUSETYPE_MODE                25075
WALLSMATERIAL_MODE            25459
dtype: int64

41
```

```python
# Here we need OWN_CAR_AGE and EXT_SOURCE_1 columns thus removing that
entries from the series a

a.pop("OWN_CAR_AGE")
a.pop("EXT_SOURCE_1")
```

```
28172
```

This columns have more than 50% missing values

Discription of this columns :- Normalized information about building where the client lives, What is average (_AVG suffix), modus (_MODE suffix), median (_MEDI suffix) apartment size, common area, living area, age of building, number of elevators, number of entrances, state of the building, number of floor

We don't need this information for our analysis thus removing those columns

```
# drop those columns from the application_data

application_data = application_data.drop(columns=a.index)

# Shape of the dataset after dropping those columns

application_data.shape
```

```
(49999, 83)
```

```
application_data.head()
```

```
   SK_ID_CURR  TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR  \
0      100002       1         Cash loans           M            N
1      100003       0         Cash loans           F            N
2      100004       0    Revolving loans           M            Y
3      100006       0         Cash loans           F            N
4      100007       0         Cash loans           M            N

  FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT
AMT_ANNUITY  \
0               Y             0          202500.0    406597.5
24700.5
1               N             0          270000.0   1293502.5
35698.5
2               Y             0           67500.0    135000.0
6750.0
3               Y             0          135000.0    312682.5
29686.5
4               Y             0          121500.0    513000.0
21865.5

   ...  FLAG_DOCUMENT_18 FLAG_DOCUMENT_19 FLAG_DOCUMENT_20
FLAG_DOCUMENT_21  \
0  ...                 0                0                0
0
1  ...                 0                0                0
0
2  ...                 0                0                0
0
3  ...                 0                0                0
0
```

```
4  ...                   0                0                0
0

  AMT_REQ_CREDIT_BUREAU_HOUR AMT_REQ_CREDIT_BUREAU_DAY  \
0                        0.0                       0.0
1                        0.0                       0.0
2                        0.0                       0.0
3                        NaN                       NaN
4                        0.0                       0.0

   AMT_REQ_CREDIT_BUREAU_WEEK  AMT_REQ_CREDIT_BUREAU_MON  \
0                         0.0                        0.0
1                         0.0                        0.0
2                         0.0                        0.0
3                         NaN                        NaN
4                         0.0                        0.0

   AMT_REQ_CREDIT_BUREAU_QRT  AMT_REQ_CREDIT_BUREAU_YEAR
0                        0.0                         1.0
1                        0.0                         0.0
2                        0.0                         0.0
3                        NaN                         NaN
4                        0.0                         0.0

[5 rows x 83 columns]
```

```
# Let's see now which columns have null values so that we can work on
them

a = application_data.isnull().sum()
a = a[a>0]
print(a)
len(a)
```

```
AMT_ANNUITY                          1
AMT_GOODS_PRICE                     38
NAME_TYPE_SUITE                    192
OWN_CAR_AGE                      32950
OCCUPATION_TYPE                  15654
CNT_FAM_MEMBERS                      1
EXT_SOURCE_1                     28172
EXT_SOURCE_2                       126
EXT_SOURCE_3                      9944
YEARS_BEGINEXPLUATATION_AVG      24394
FLOORSMAX_AVG                    24875
YEARS_BEGINEXPLUATATION_MODE     24394
FLOORSMAX_MODE                   24875
YEARS_BEGINEXPLUATATION_MEDI     24394
FLOORSMAX_MEDI                   24875
TOTALAREA_MODE                   24148
```

```
EMERGENCYSTATE_MODE            23698
OBS_30_CNT_SOCIAL_CIRCLE         168
DEF_30_CNT_SOCIAL_CIRCLE         168
OBS_60_CNT_SOCIAL_CIRCLE         168
DEF_60_CNT_SOCIAL_CIRCLE         168
DAYS_LAST_PHONE_CHANGE             1
AMT_REQ_CREDIT_BUREAU_HOUR      6734
AMT_REQ_CREDIT_BUREAU_DAY       6734
AMT_REQ_CREDIT_BUREAU_WEEK      6734
AMT_REQ_CREDIT_BUREAU_MON       6734
AMT_REQ_CREDIT_BUREAU_QRT       6734
AMT_REQ_CREDIT_BUREAU_YEAR      6734
dtype: int64

28
```

Observation

1) AMT_REQ_CREDIT_BUREAU_HOUR/DAY/WEEK/MON/QRT/YEAR has the Number of enquiries to Credit Bureau about the client before application thus null values in this columns mean that there are 0 enquires thus filling null values in this columns with 0.

2) NAME_TYPE_SUITE has null values where it is not provided by applicant but if it is not porvided then it should be "Unaccompanied".

3) OCCUPATION_TYPE has null values where it is not provied by applicant let's replace it with "Unknown".

4) CNT_FAM_MEMBERS has 1 null value also its family_details are unknown so it is safe to replce its value to 0

```
# Filling those null values with the obersved value

keys = ['AMT_REQ_CREDIT_BUREAU_HOUR', "AMT_REQ_CREDIT_BUREAU_DAY",
"AMT_REQ_CREDIT_BUREAU_WEEK", "AMT_REQ_CREDIT_BUREAU_MON",
"AMT_REQ_CREDIT_BUREAU_QRT", "AMT_REQ_CREDIT_BUREAU_YEAR",
"NAME_TYPE_SUITE", "OCCUPATION_TYPE", "CNT_FAM_MEMBERS"]
values = [0,0,0,0,0,0,"Unaccompanied", "Unknown", 0]

filldict = dict(zip(keys, values))

application_data.fillna(value=filldict, inplace=True)

# Removing the Normalized information about building where the client
lives as there are lots of null values and we aren't gonna use them

application_data =
application_data.drop(columns=["YEARS_BEGINEXPLUATATION_AVG",
```

```
"FLOORSMAX_AVG", "YEARS_BEGINEXPLUATATION_MODE", "FLOORSMAX_MODE",
"YEARS_BEGINEXPLUATATION_MEDI","FLOORSMAX_MEDI","TOTALAREA_MODE","EMER
GENCYSTATE_MODE"])
```

Handling Duplicate values

```
# Finding duplicated values

application_data.duplicated().sum()

0
```

Changing the columns having neagtive values.

```
# This values might be entered negative accendentally .
# We can easily see that age can't have negative value.

application_data[['DAYS_BIRTH' ,'DAYS_EMPLOYED' ,'DAYS_REGISTRATION' ,
'DAYS_ID_PUBLISH']]

       DAYS_BIRTH  DAYS_EMPLOYED  DAYS_REGISTRATION  DAYS_ID_PUBLISH
0           -9461           -637              -3648            -2120
1          -16765          -1188              -1186             -291
2          -19046           -225              -4260            -2531
3          -19005          -3039              -9833            -2437
4          -19932          -3038              -4311            -3458
...           ...            ...                ...              ...
49994      -10667           -285              -2521            -3333
49995      -20211          -4651             -11281            -3722
49996      -10280          -1158              -8620            -2604
49997      -23485          -2181              -2662            -4200
49998      -19251         365243             -12934            -2783

[49999 rows x 4 columns]

# Using absolute fucntion to convert negative values to positive

application_data[['DAYS_BIRTH' ,'DAYS_EMPLOYED' ,'DAYS_REGISTRATION' ,
'DAYS_ID_PUBLISH']] =
application_data[['DAYS_BIRTH' ,'DAYS_EMPLOYED' ,'DAYS_REGISTRATION' ,
'DAYS_ID_PUBLISH']].abs()

application_data[['DAYS_BIRTH' ,'DAYS_EMPLOYED' ,'DAYS_REGISTRATION' ,
'DAYS_ID_PUBLISH']]

   DAYS_BIRTH  DAYS_EMPLOYED  DAYS_REGISTRATION  DAYS_ID_PUBLISH
0        9461            637               3648             2120
1       16765           1188               1186              291
2       19046            225               4260             2531
3       19005           3039               9833             2437
4       19932           3038               4311             3458
```

```
...          ...          ...          ...          ...
49994        10667        285          2521         3333
49995        20211        4651         11281        3722
49996        10280        1158         8620         2604
49997        23485        2181         2662         4200
49998        19251        365243       12934        2783

[49999 rows x 4 columns]
```

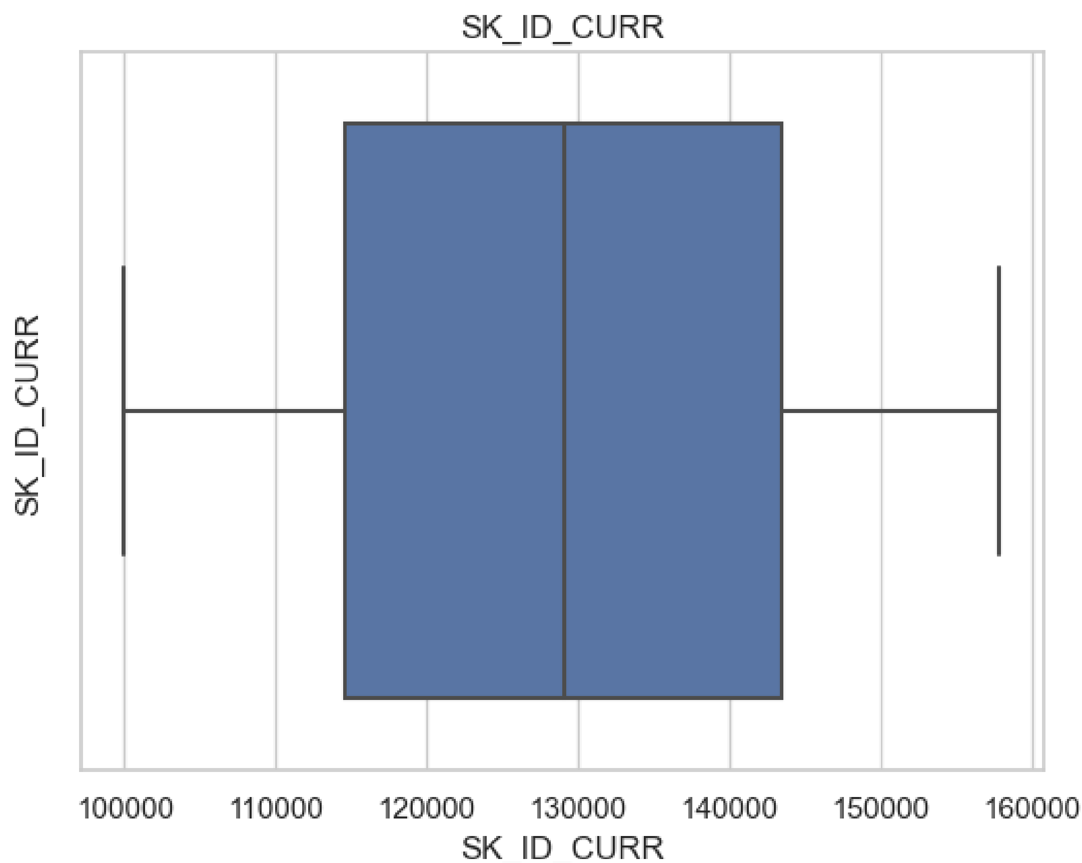
So we Identifed Missing values and Handled them.

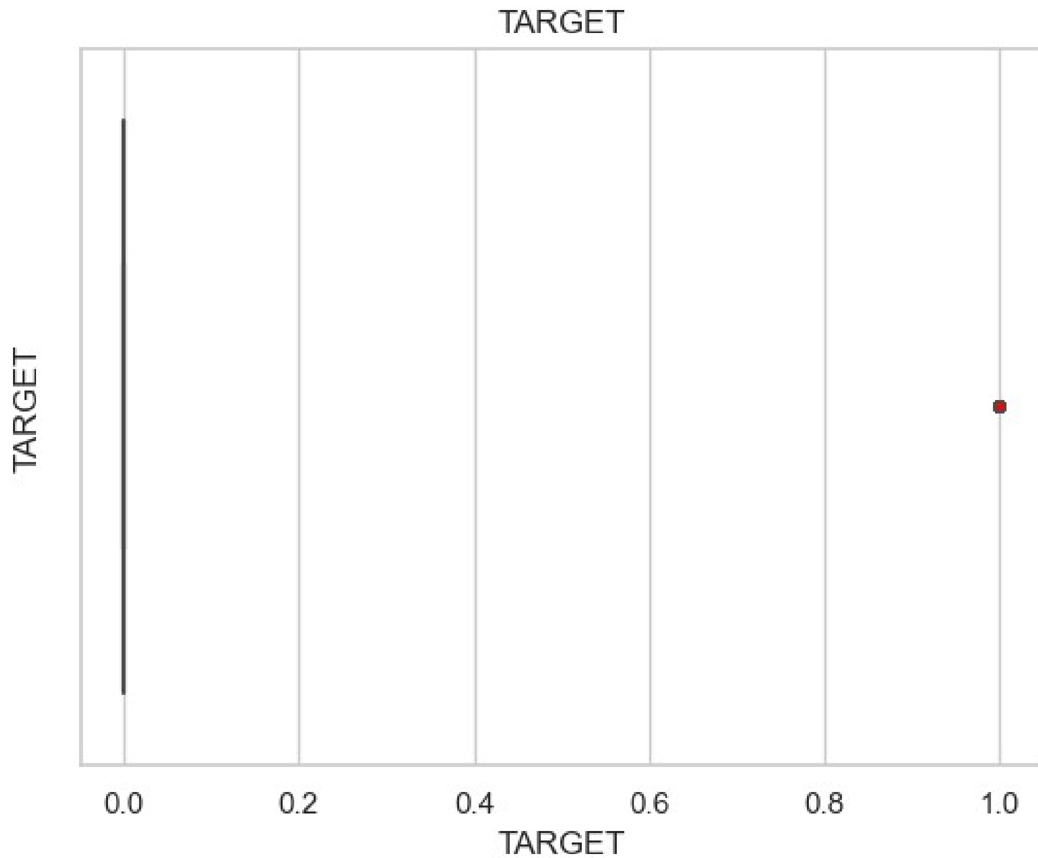
B. Identify Outliers in the Datase

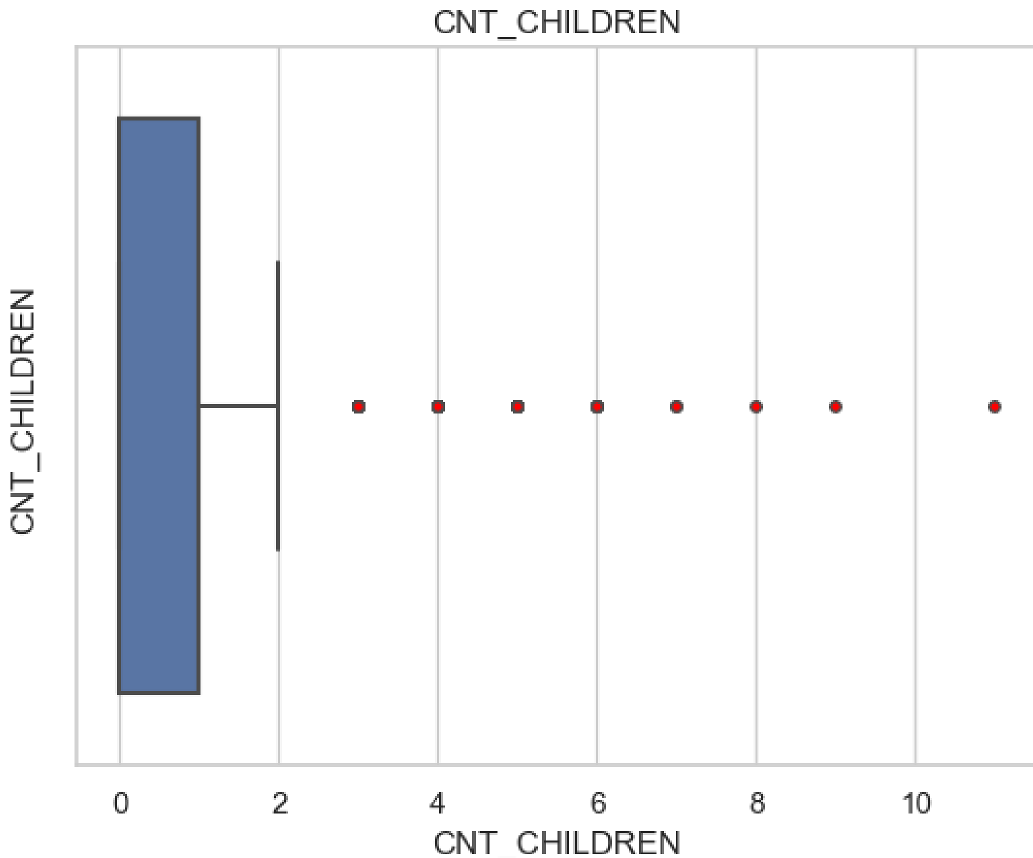
Top Correlations for Different Scenarios

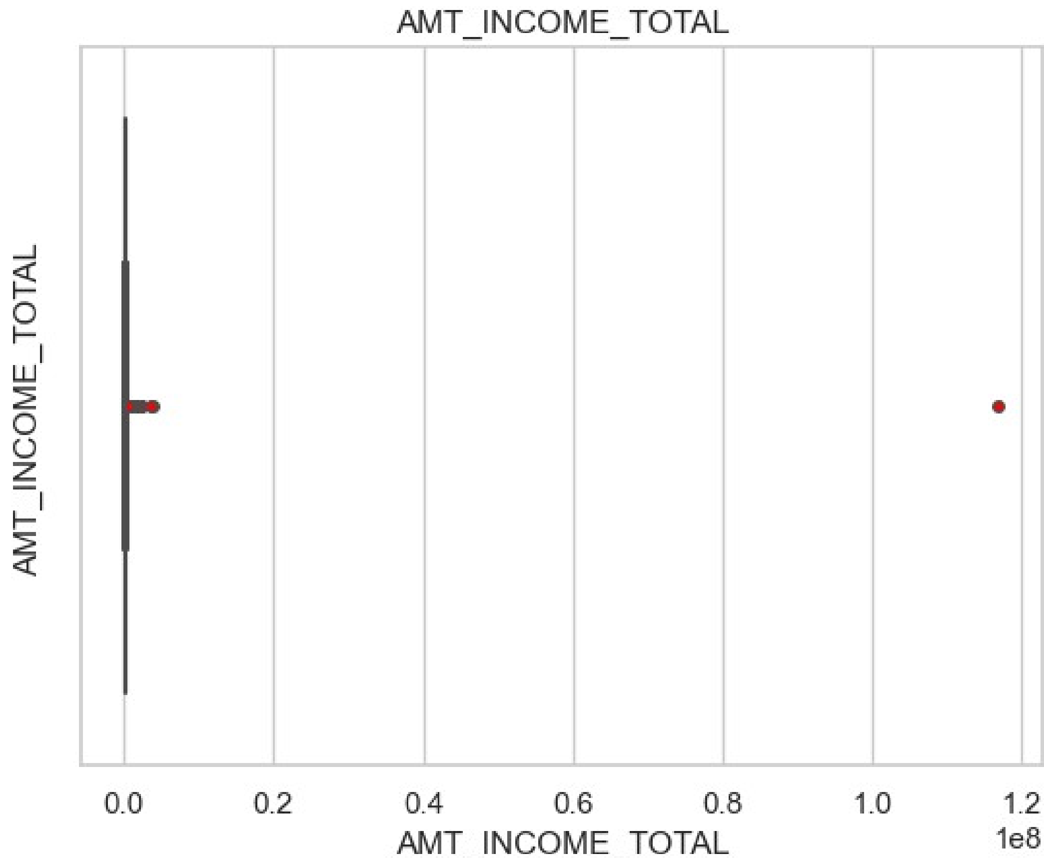
```
# Select numerical columns for analysis of Outliers
numerical_columns =
application_data.select_dtypes(include=[np.number]).columns

sns.set_theme(style="whitegrid")

# Created and empty dictuonary to store outliers as values and
variable name as keys
dictoutlier = {}


for i in numerical_columns:

    # Created an empty list to store outliers
    outliers = []

    # to Calculate IQR of data i
    q1 = np.percentile(application_data[i], 25)
    q3 = np.percentile(application_data[i], 75)
    iqr = q3 - q1

    # Define lower and upper bounds for outliers
    lower_bound = q1 - 1.5 * iqr
    upper_bound = q3 + 1.5 * iqr
```

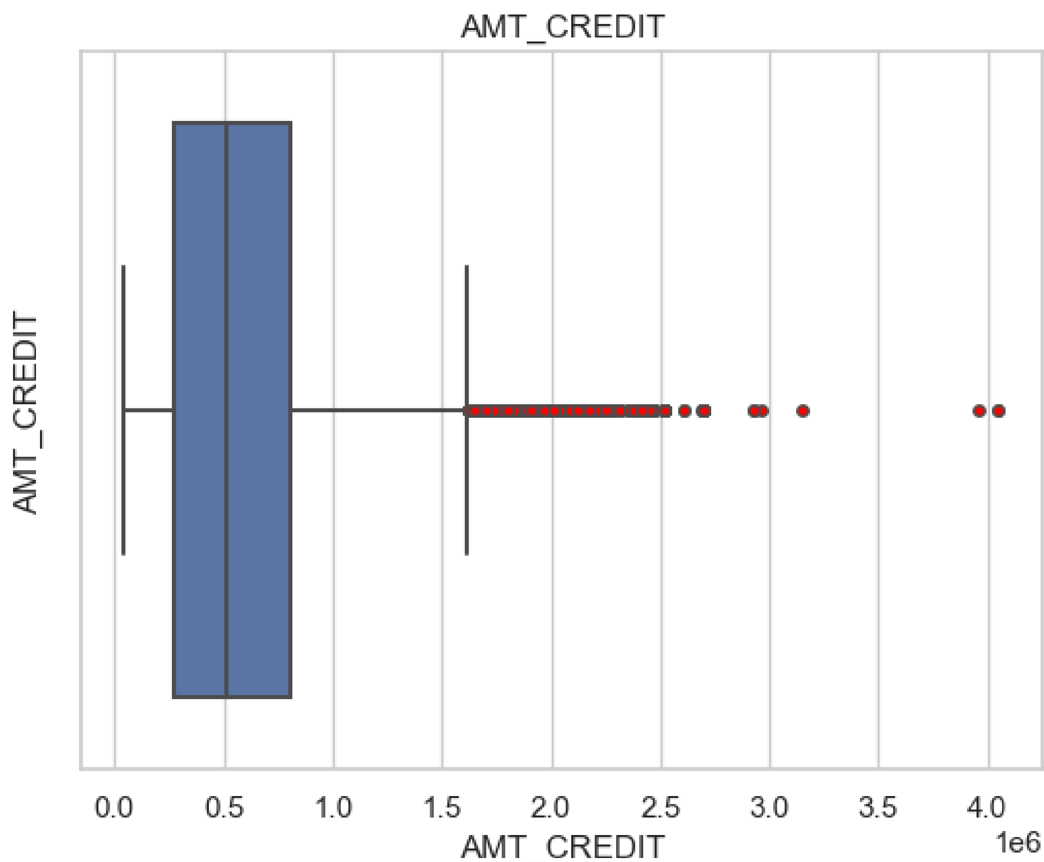
```
    # find and store outliers in the variable i
    outliers = [x for x in application_data[i] if x < lower_bound or x
> upper_bound]
    dictoutlier[i] = outliers

    # Plot the box plot with variable name as title and outliers
marked marked with red color dots
    sns.boxplot(x = application_data[i], flierprops=dict(marker='o',
markerfacecolor='red', markersize=4))
    plt.title(i)
    plt.ylabel(i)
    plt.show()
```

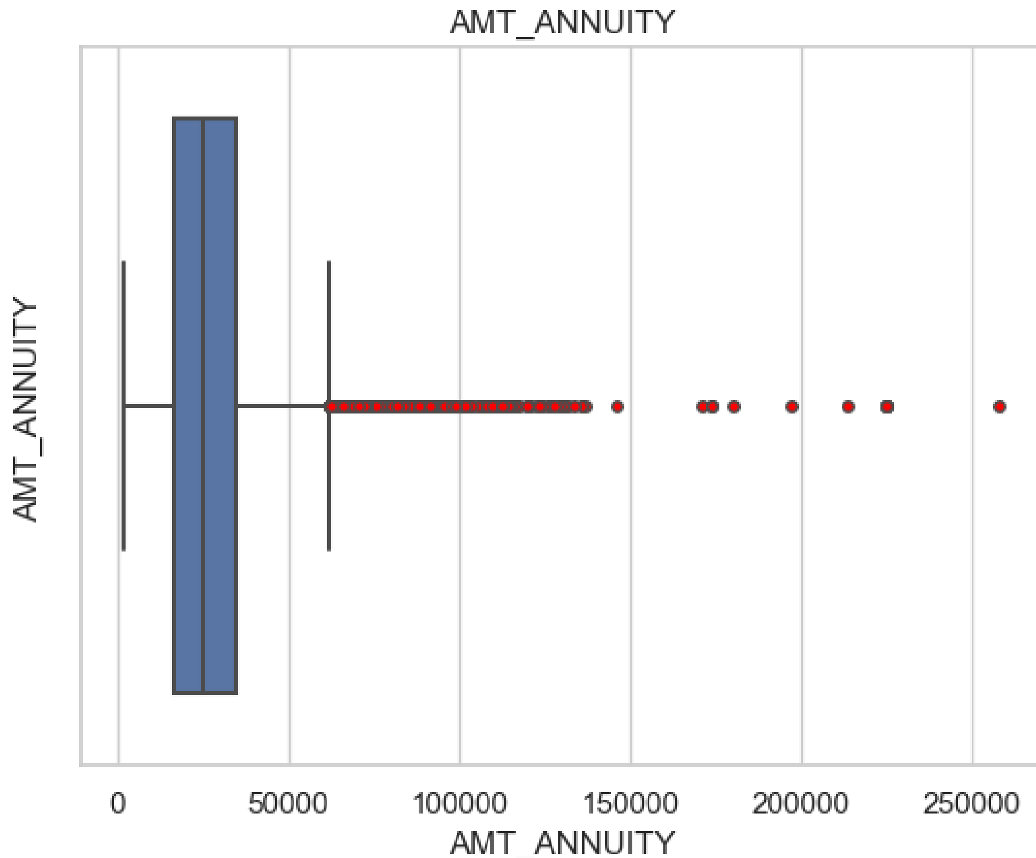


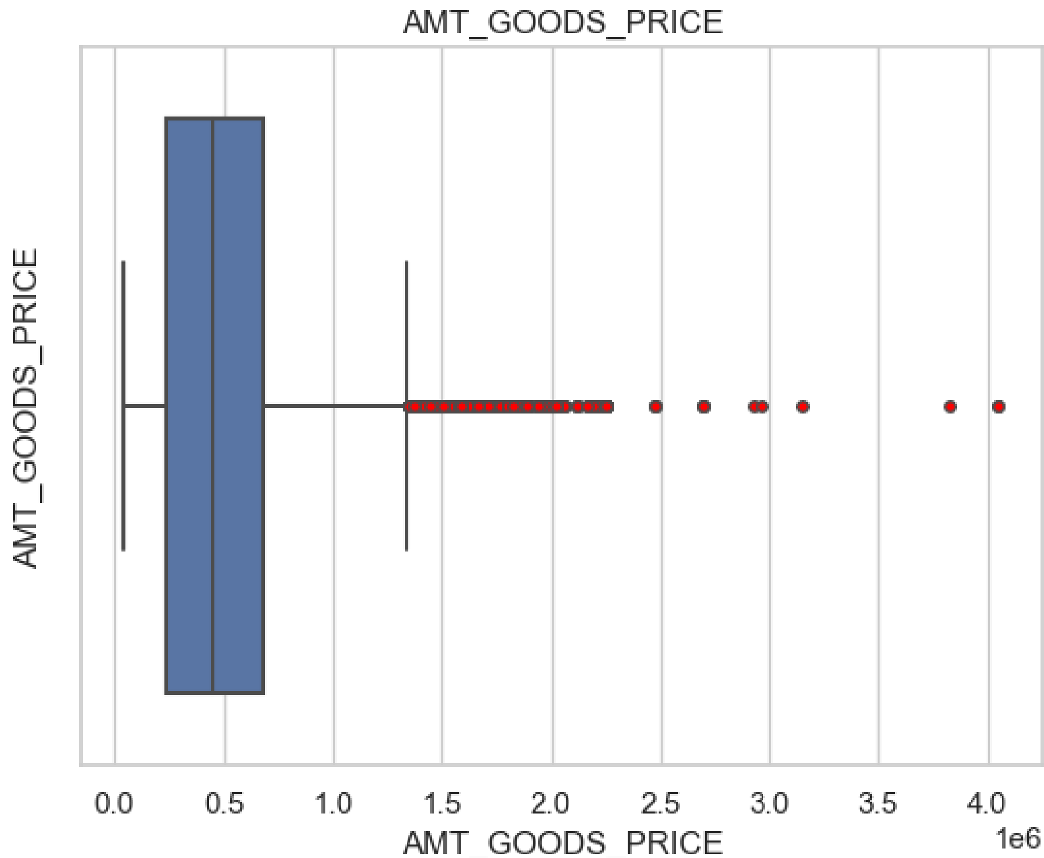

TARGET
TARGET
0.0
0.2
0.4
0.6
0.8
1.0
TARGET

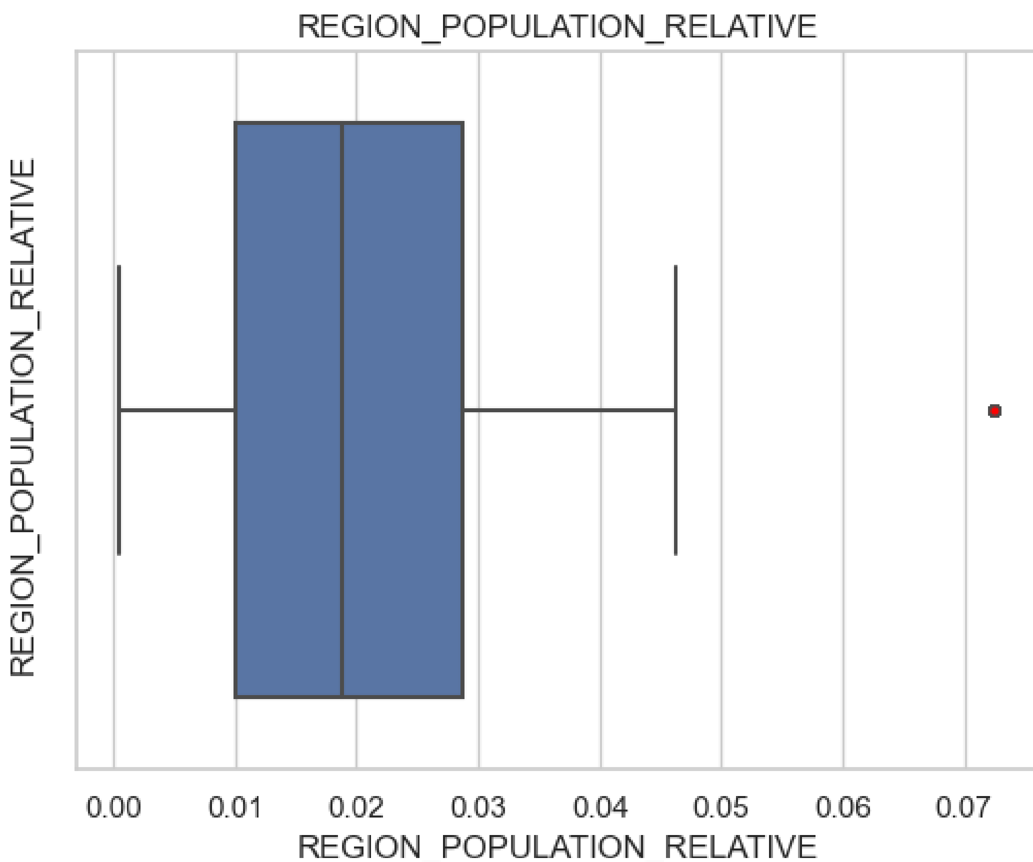

CNT_CHILDREN
CNT_CHILDREN
0
2
4
6
8
10
CNT_CHILDREN


AMT_INCOME_TOTAL
AMT_INCOME_TOTAL
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1e8
AMT_INCOME_TOTAL

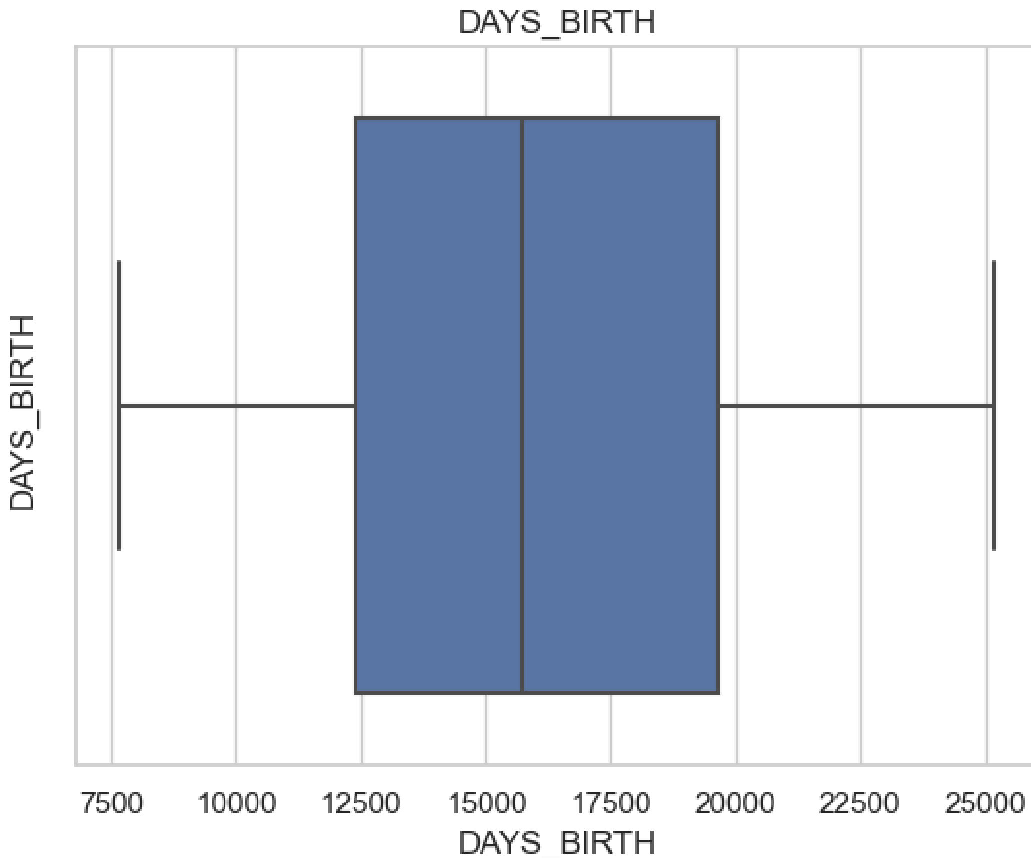

AMT_CREDIT
AMT_CREDIT
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
1e6
AMT_CREDIT

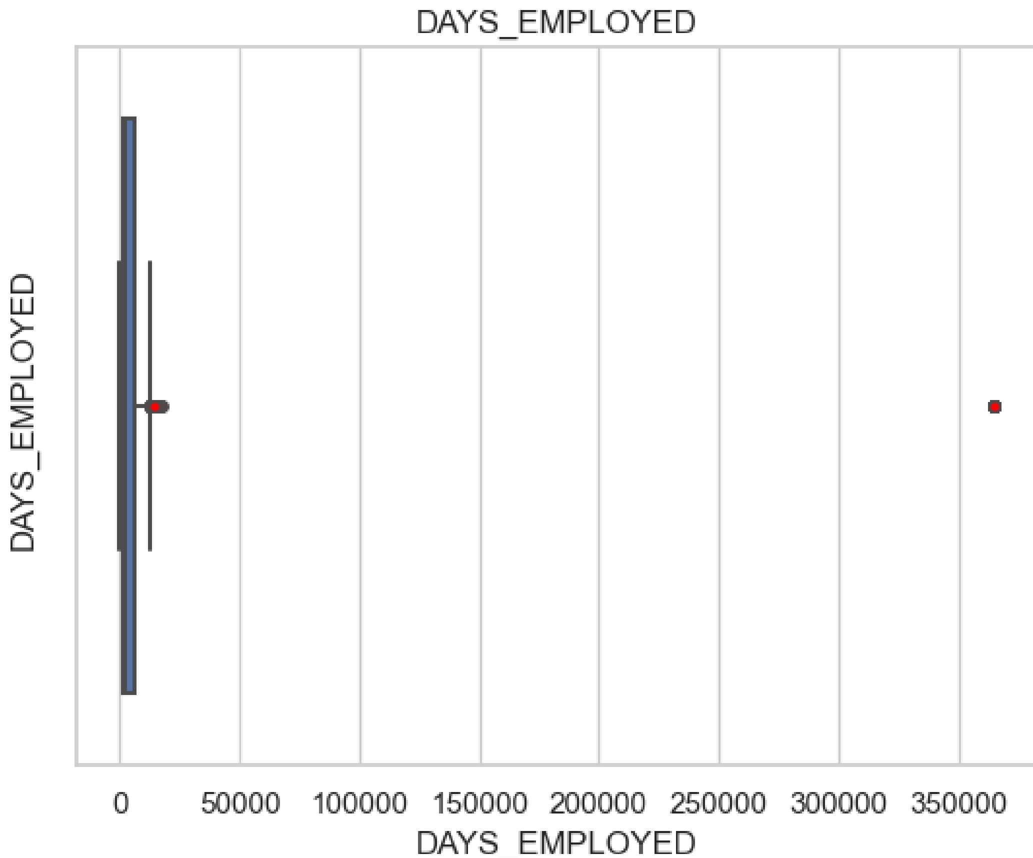

AMT_ANNUITY
AMT_ANNUITY
0
50000
100000
150000
200000
250000
AMT_ANNUITY

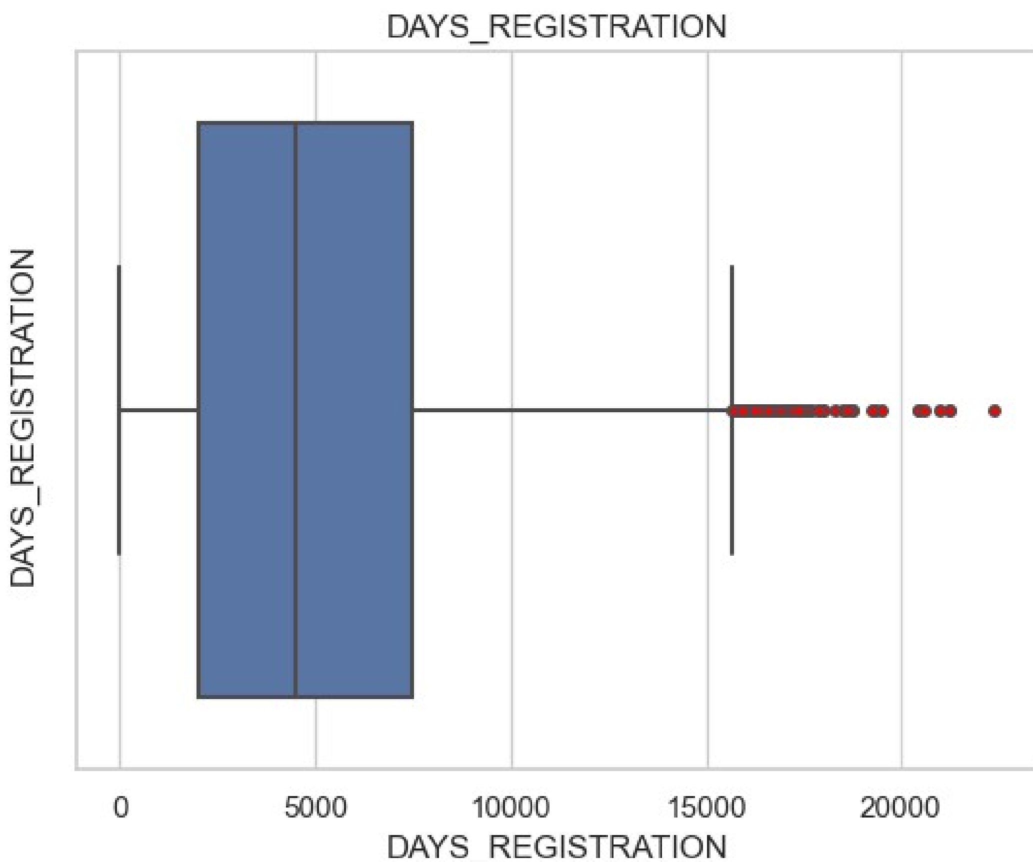

AMT_GOODS_PRICE
AMT_GOODS_PRICE
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
1e6
AMT_GOODS_PRICE



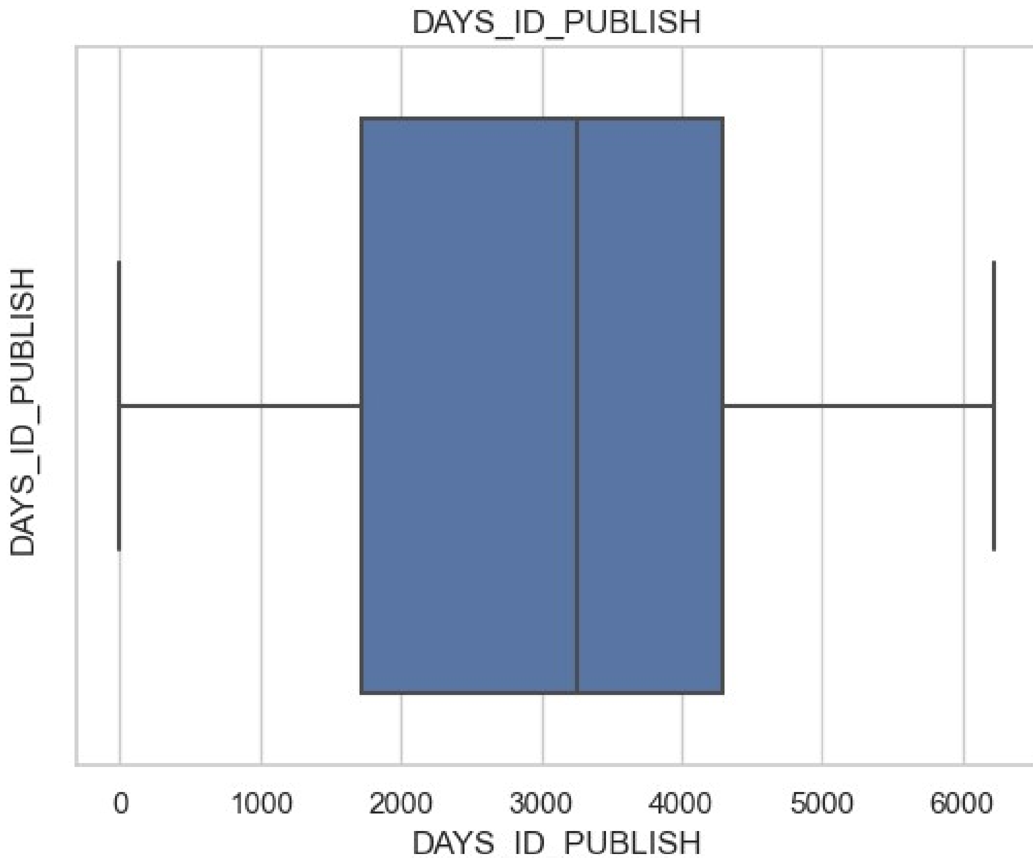
REGION_POPULATION_RELATIVE
REGION_POPULATION_RELATIVE
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
REGION_POPULATION_RELATIVE

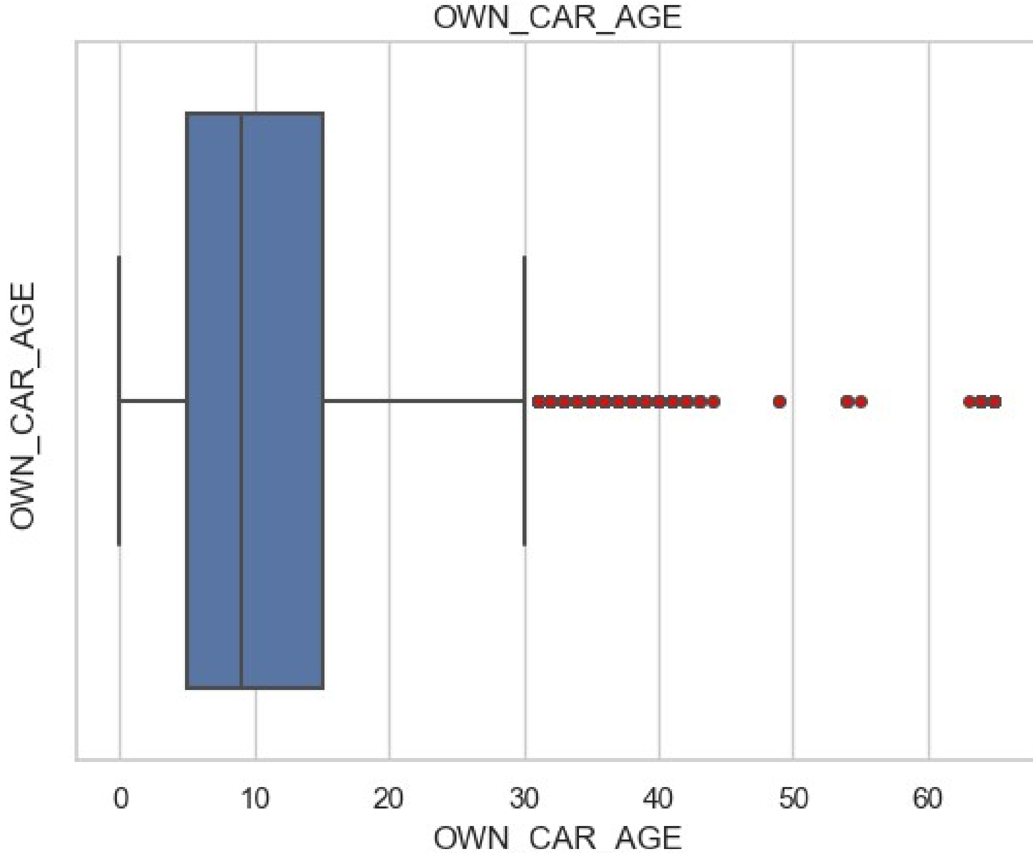

DAYS_BIRTH
DAYS_BIRTH
7500
10000
12500
15000
17500
20000
22500
25000
DAYS_BIRTH

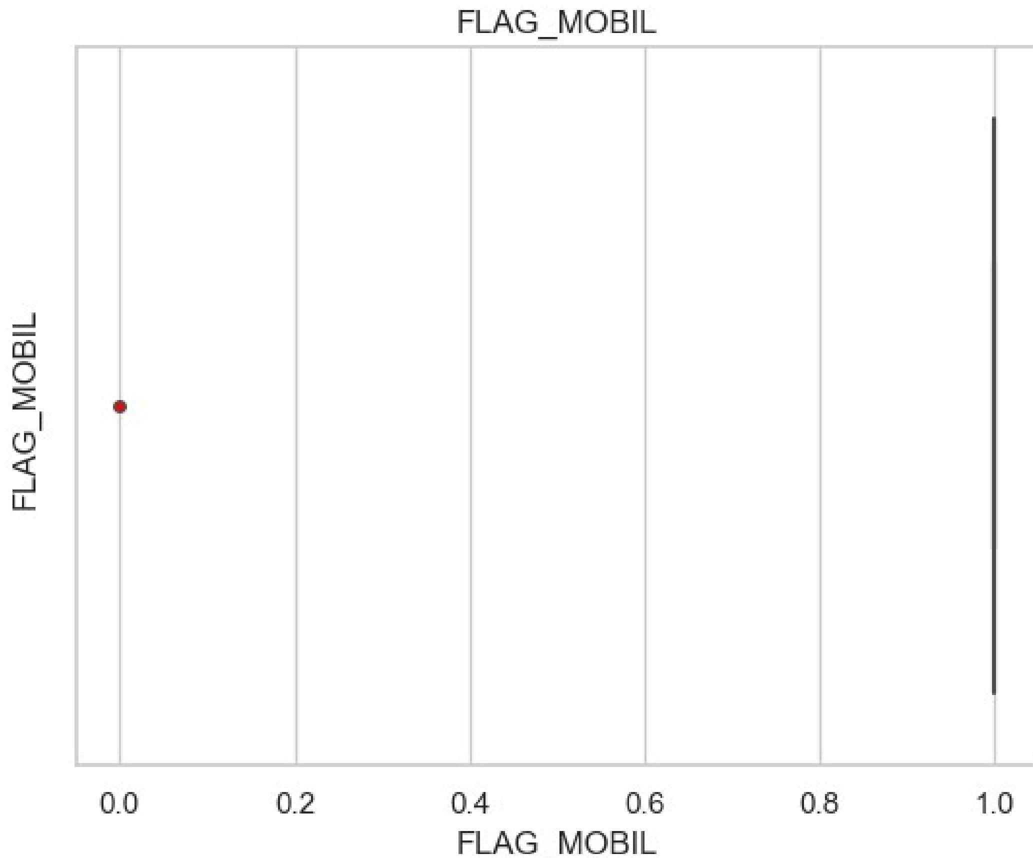

DAYS_EMPLOYED
DAYS_EMPLOYED
0
50000
100000
150000
200000
250000
300000
350000
DAYS_EMPLOYED

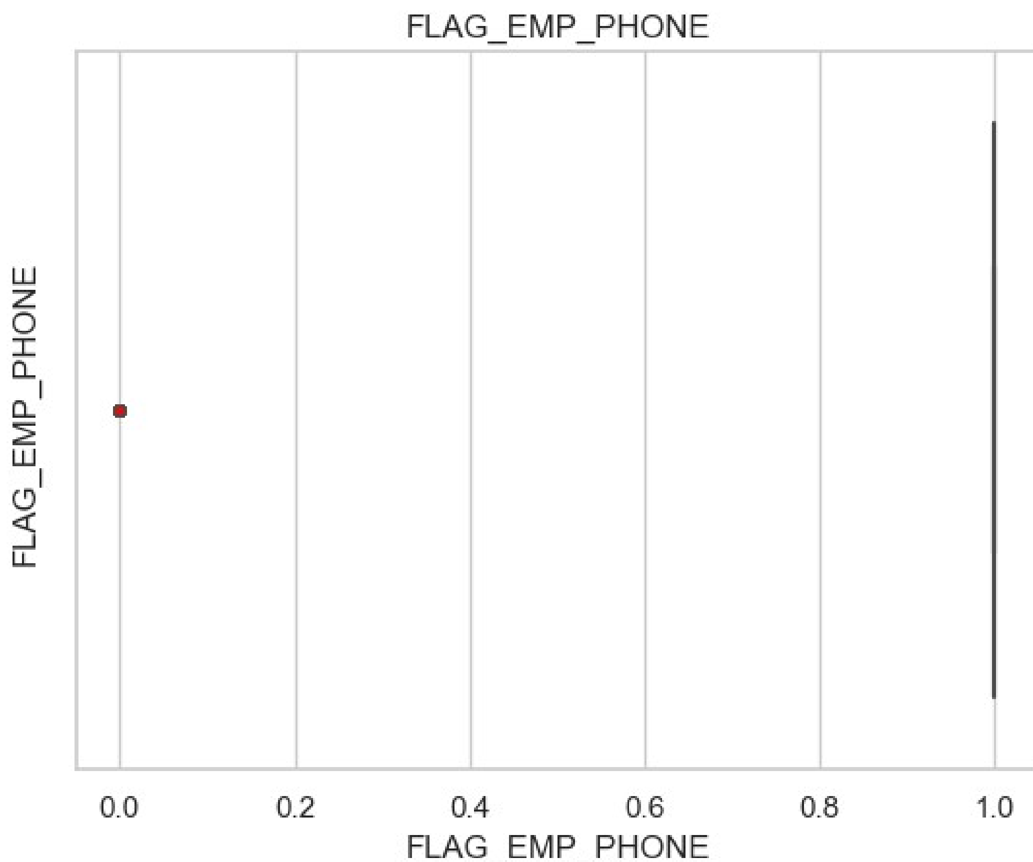


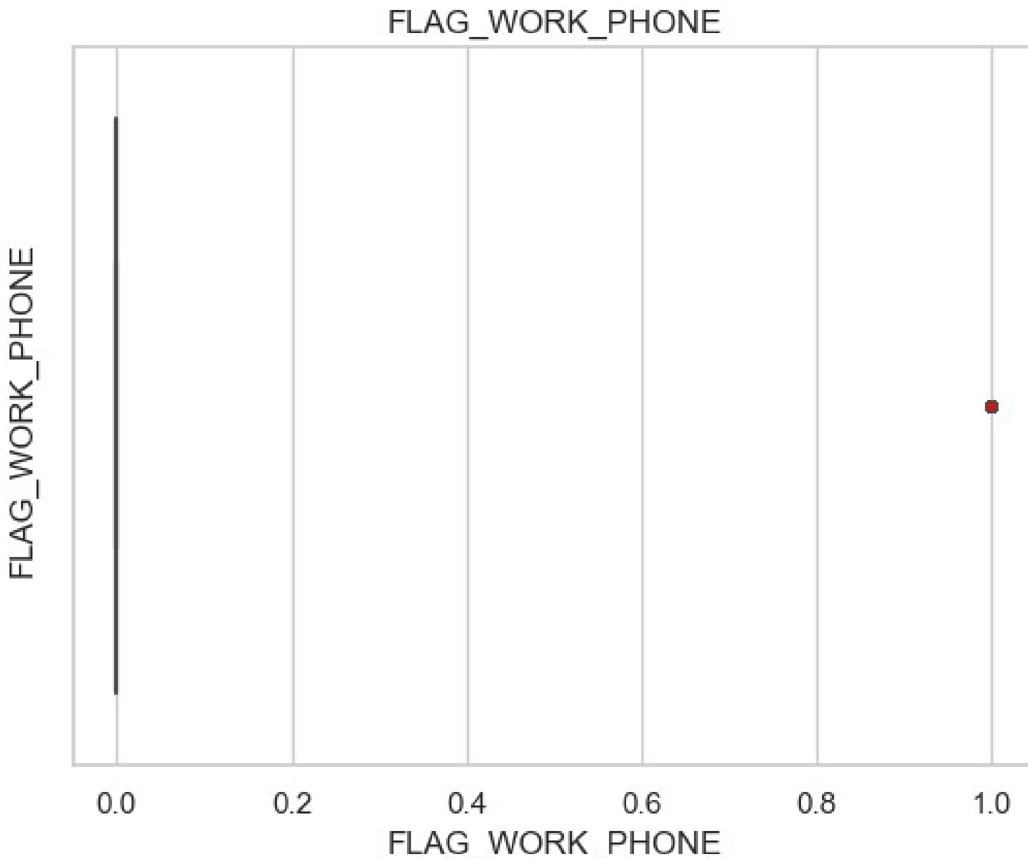
DAYS_REGISTRATION
DAYS_REGISTRATION
0
5000
10000
15000
20000
DAYS_REGISTRATION

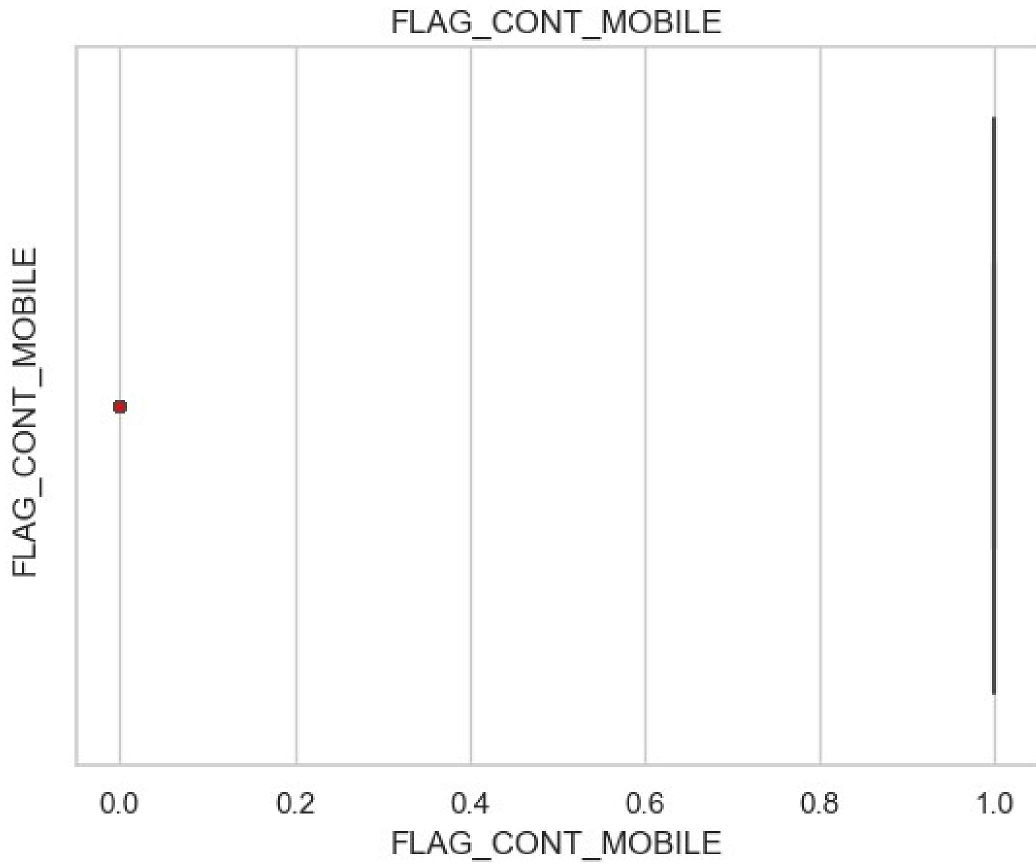

DAYS_ID_PUBLISH
DAYS_ID_PUBLISH
0
1000
2000
3000
4000
5000
6000
DAYS_ID_PUBLISH

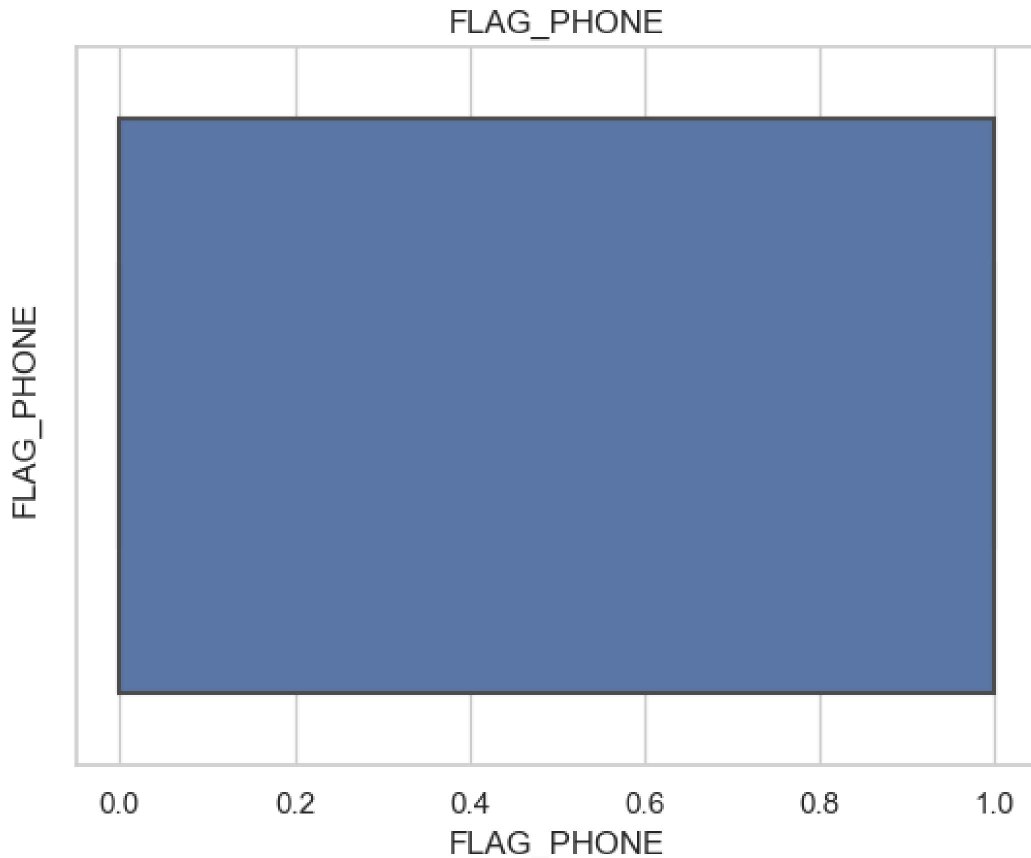

OWN_CAR_AGE
OWN_CAR_AGE
0
10
20
30
40
50
60
OWN_CAR_AGE

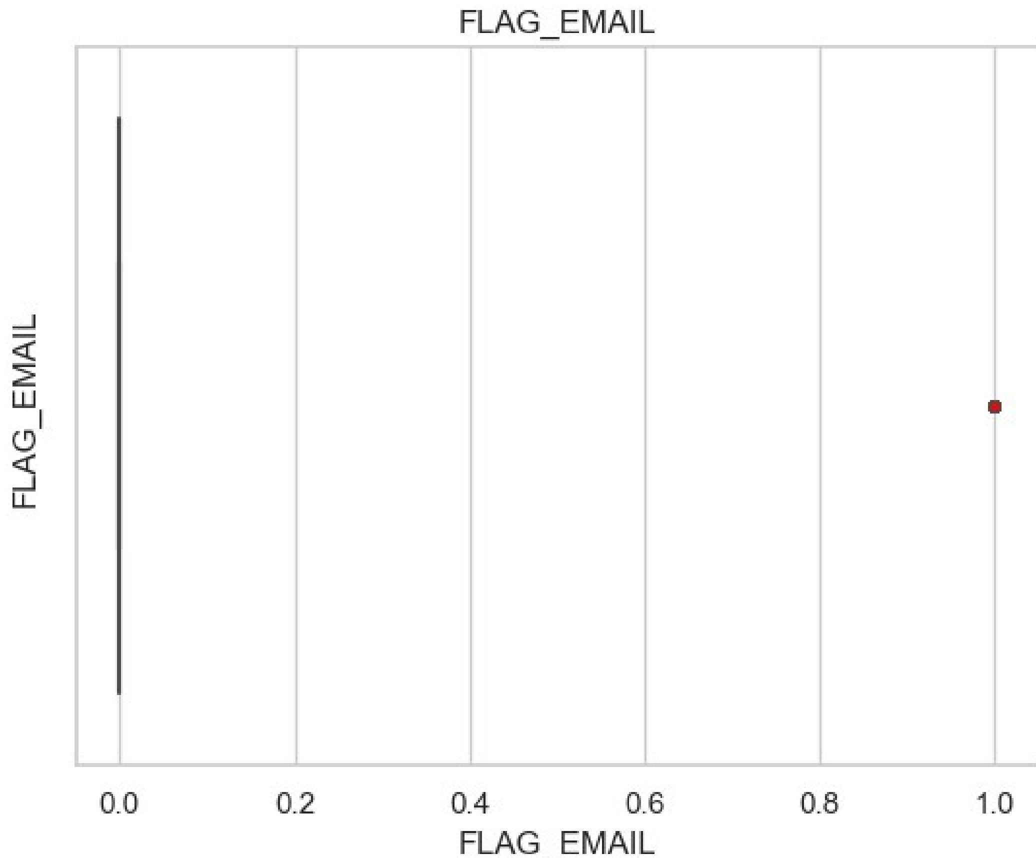

FLAG_MOBIL
FLAG_MOBIL
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_MOBIL

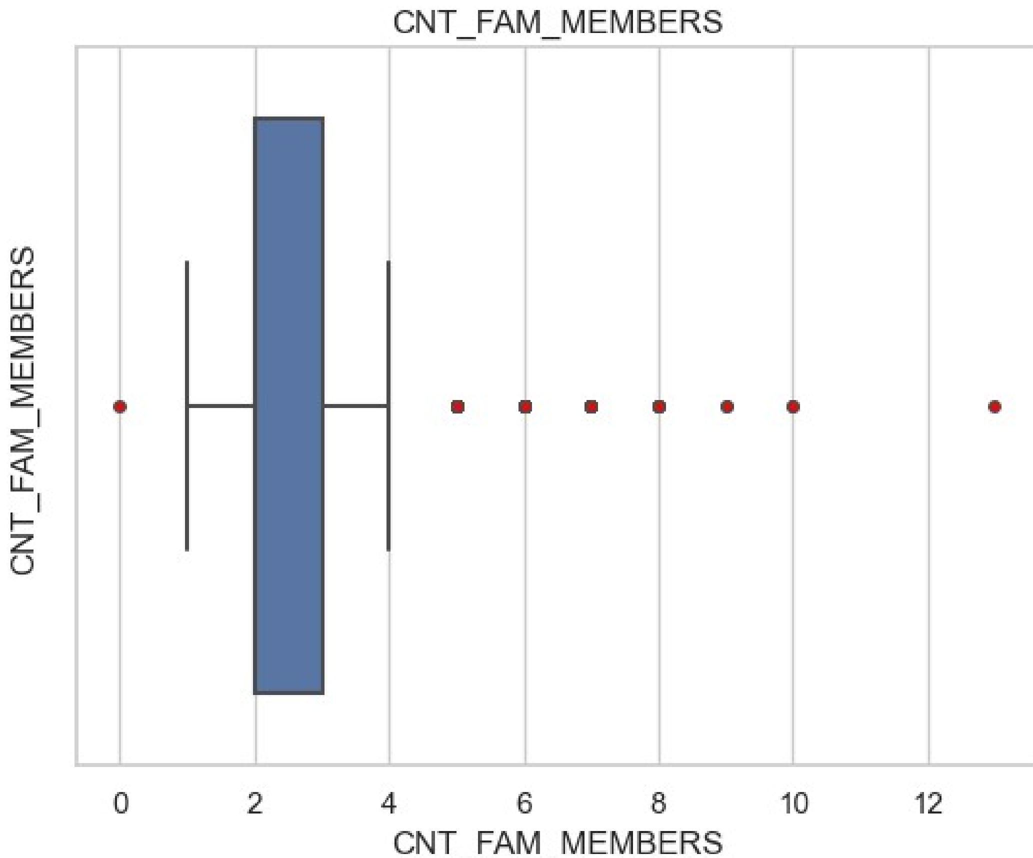

FLAG_EMP_PHONE
FLAG_EMP_PHONE
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_EMP_PHONE


FLAG_WORK_PHONE
FLAG_WORK_PHONE
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_WORK_PHONE

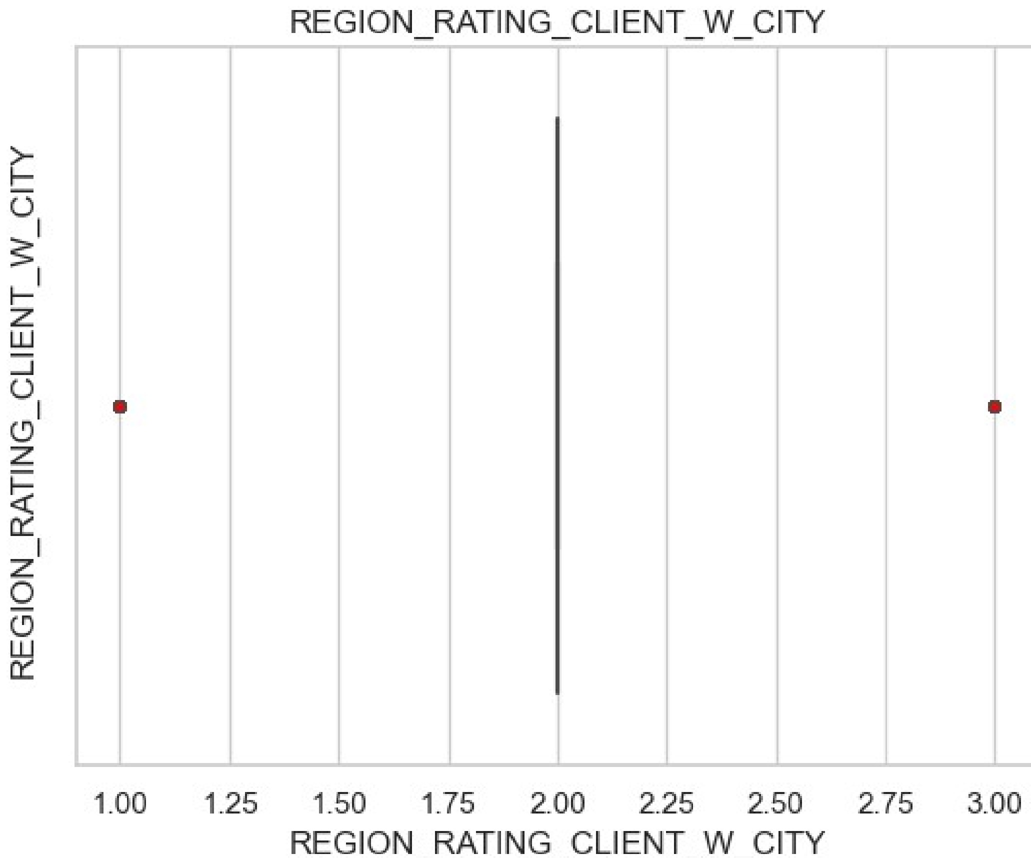

FLAG_CONT_MOBILE
FLAG_CONT_MOBILE
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_CONT_MOBILE

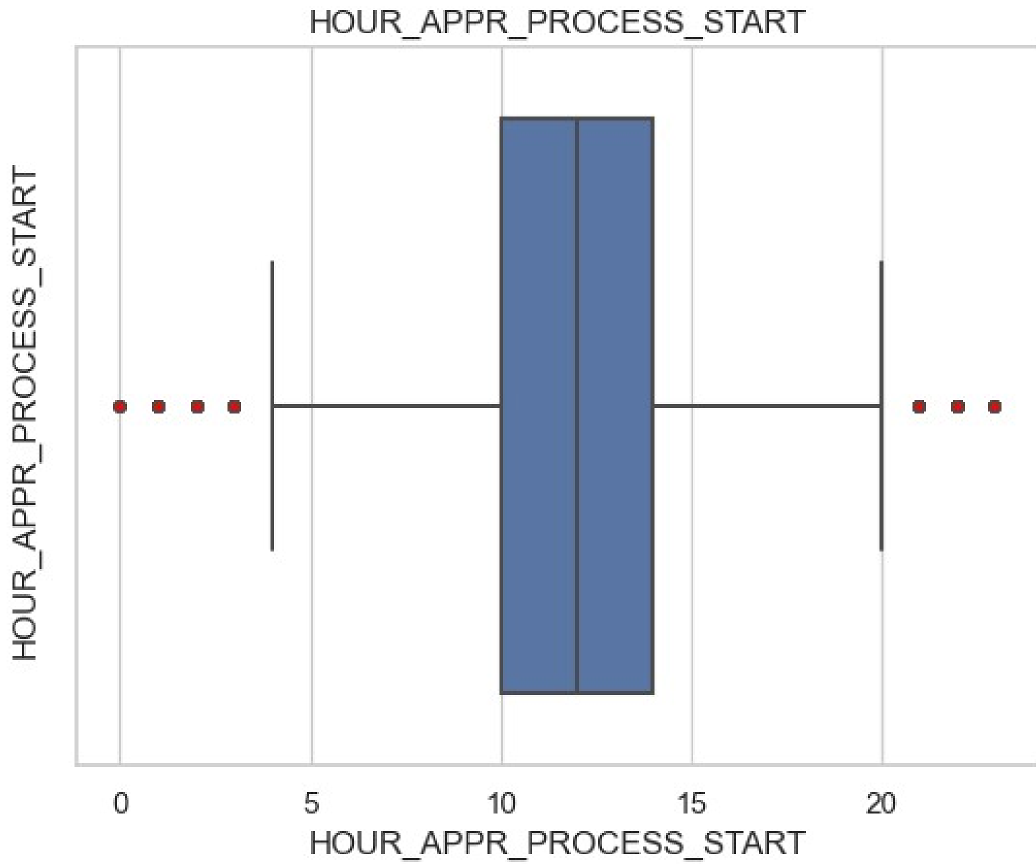

FLAG_PHONE
FLAG_PHONE
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_PHONE

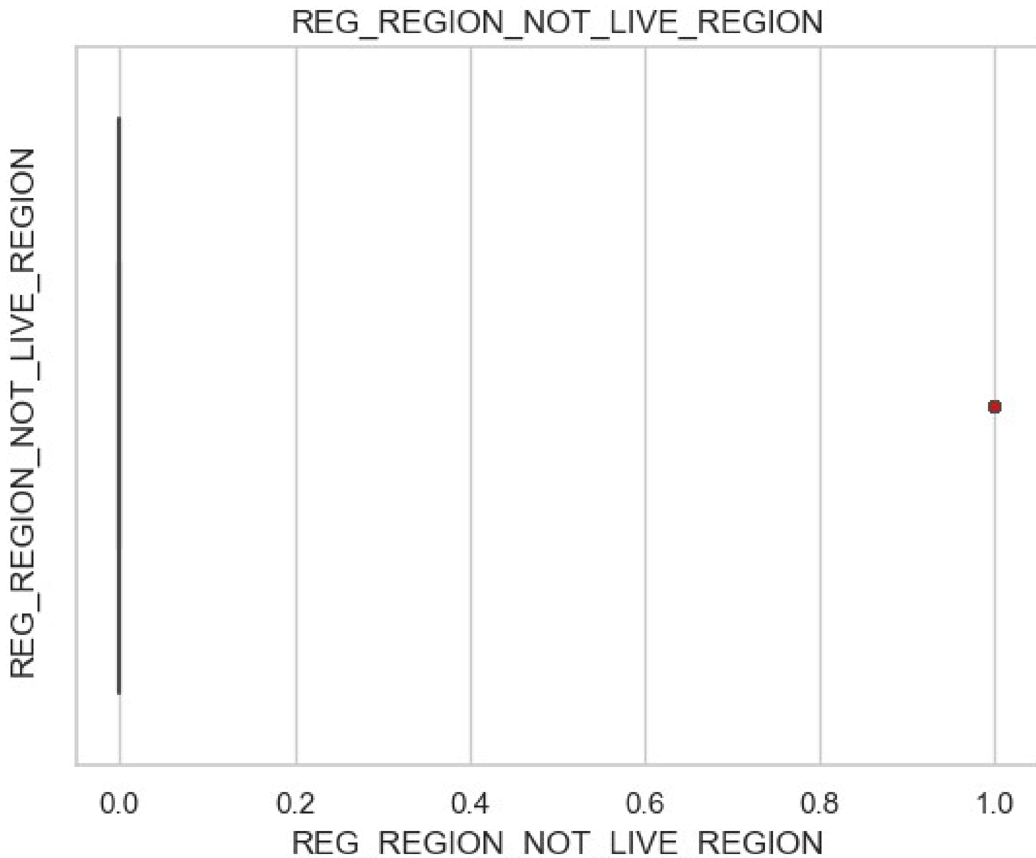

FLAG_EMAIL
FLAG_EMAIL
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_EMAIL

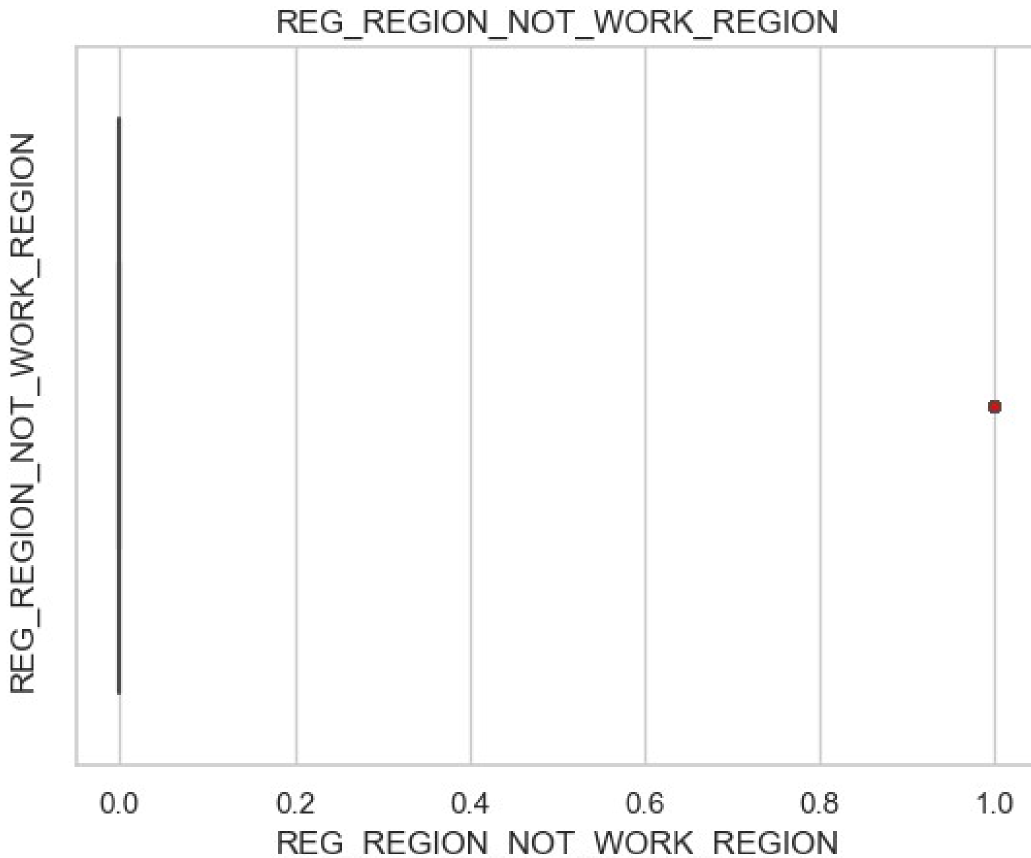

CNT_FAM_MEMBERS
CNT_FAM_MEMBERS
0
2
4
6
8
10
12
CNT_FAM_MEMBERS

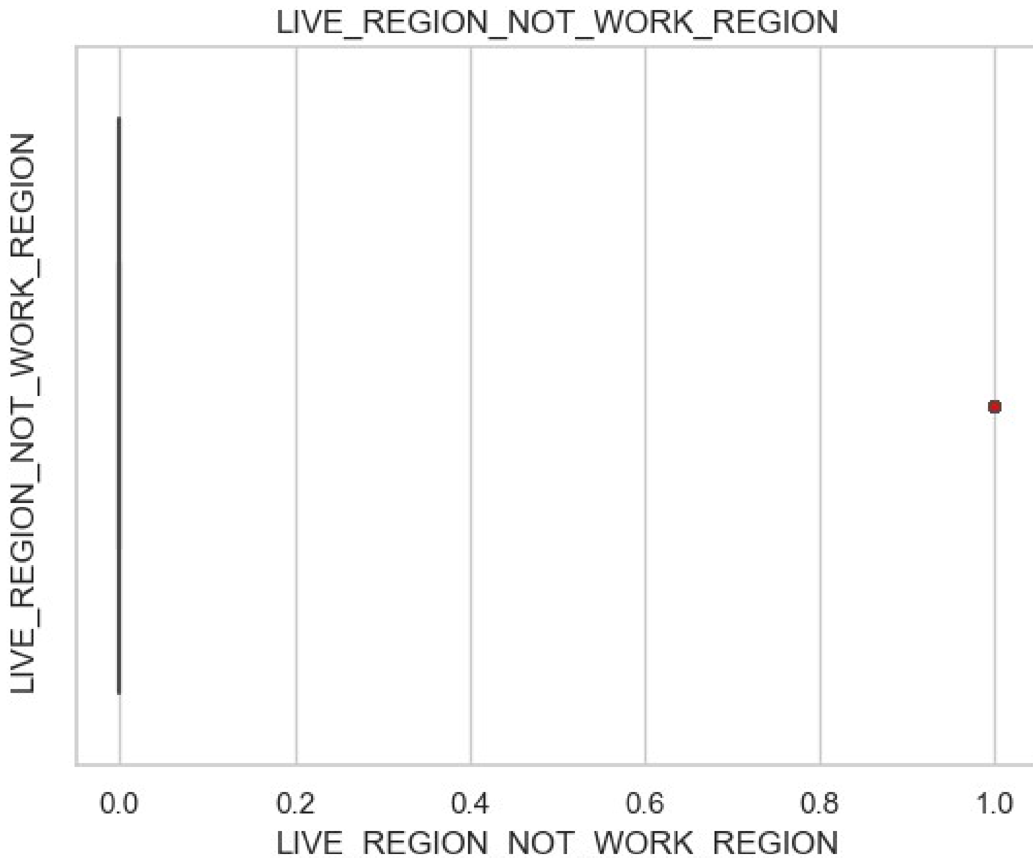

REGION_RATING_CLIENT
REGION_RATING_CLIENT
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
REGION_RATING_CLIENT

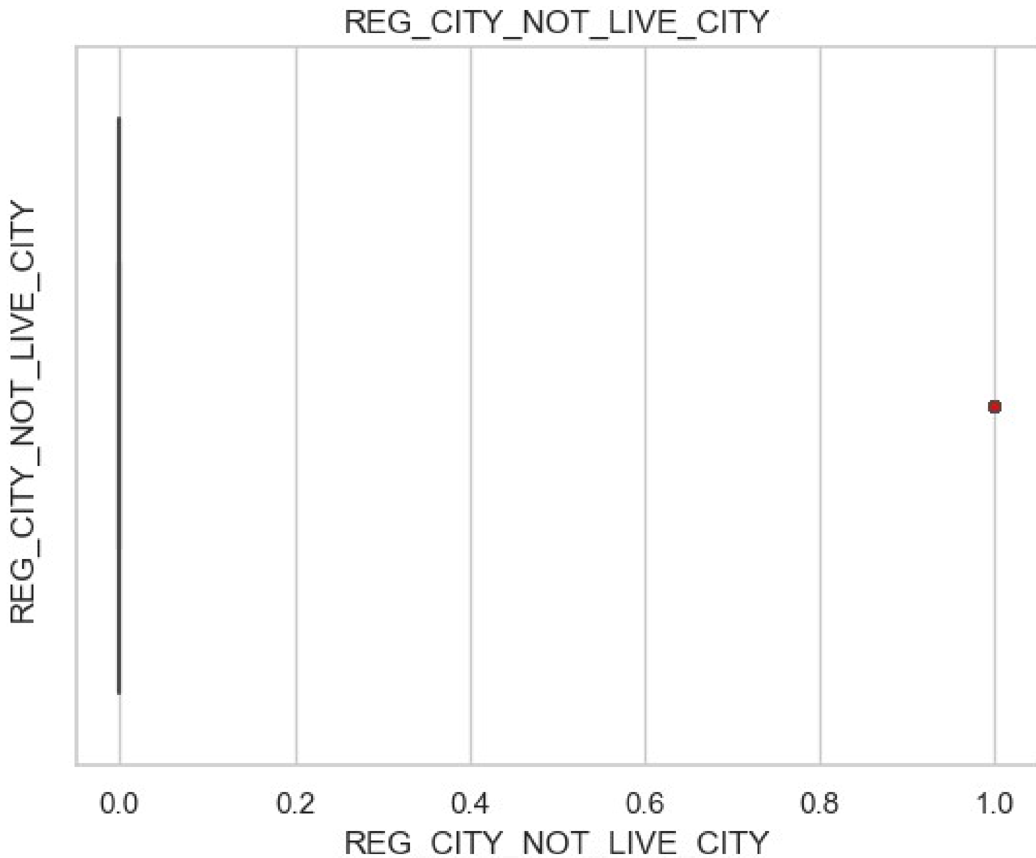

REGION_RATING_CLIENT_W_CITY
REGION_RATING_CLIENT_W_CITY
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
REGION_RATING_CLIENT_W_CITY

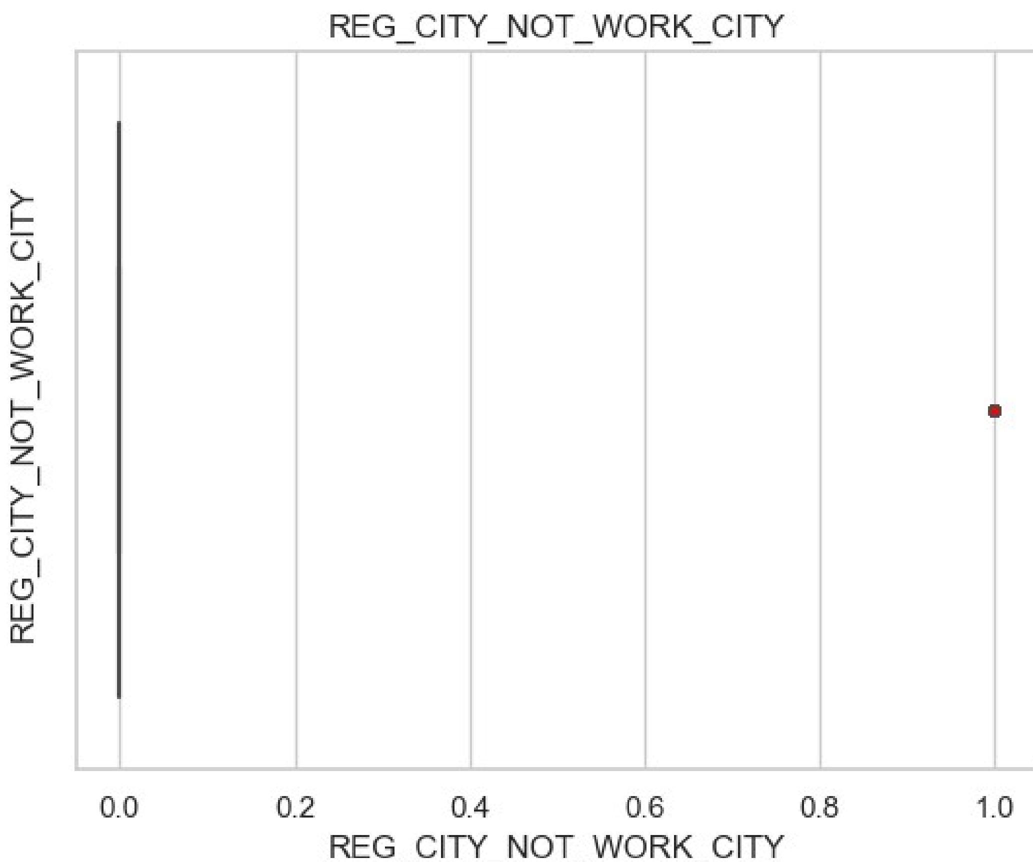

HOUR_APPR_PROCESS_START
HOUR_APPR_PROCESS_START
0
5
10
15
20
HOUR_APPR_PROCESS_START

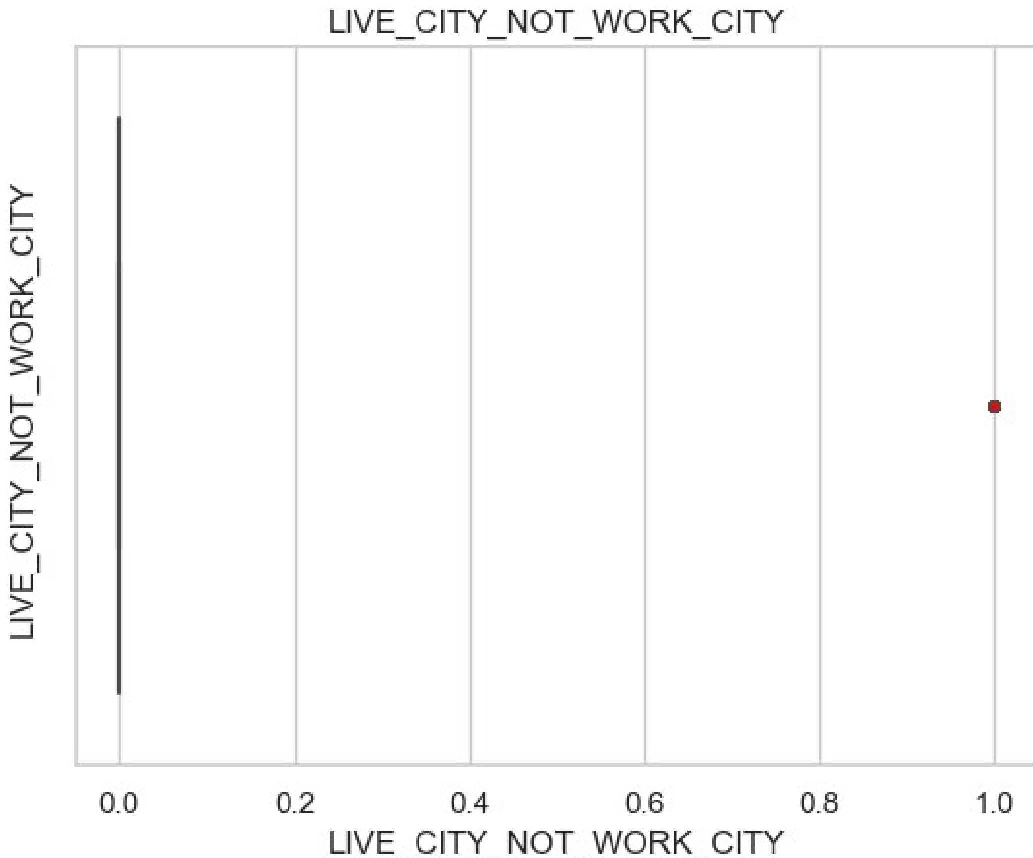

REG_REGION_NOT_LIVE_REGION
REG_REGION_NOT_LIVE_REGION
0.0
0.2
0.4
0.6
0.8
1.0
REG_REGION_NOT_LIVE_REGION

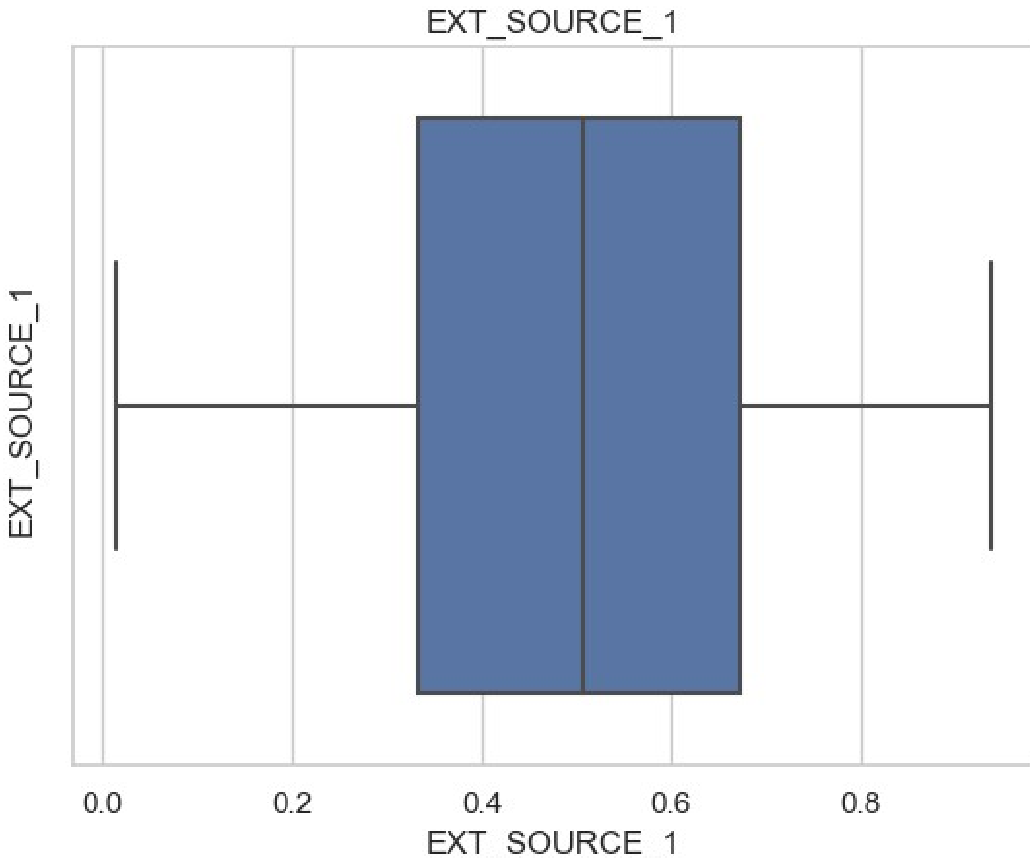

REG_REGION_NOT_WORK_REGION
REG_REGION_NOT_WORK_REGION
0.0
0.2
0.4
0.6
0.8
1.0
REG_REGION_NOT_WORK_REGION

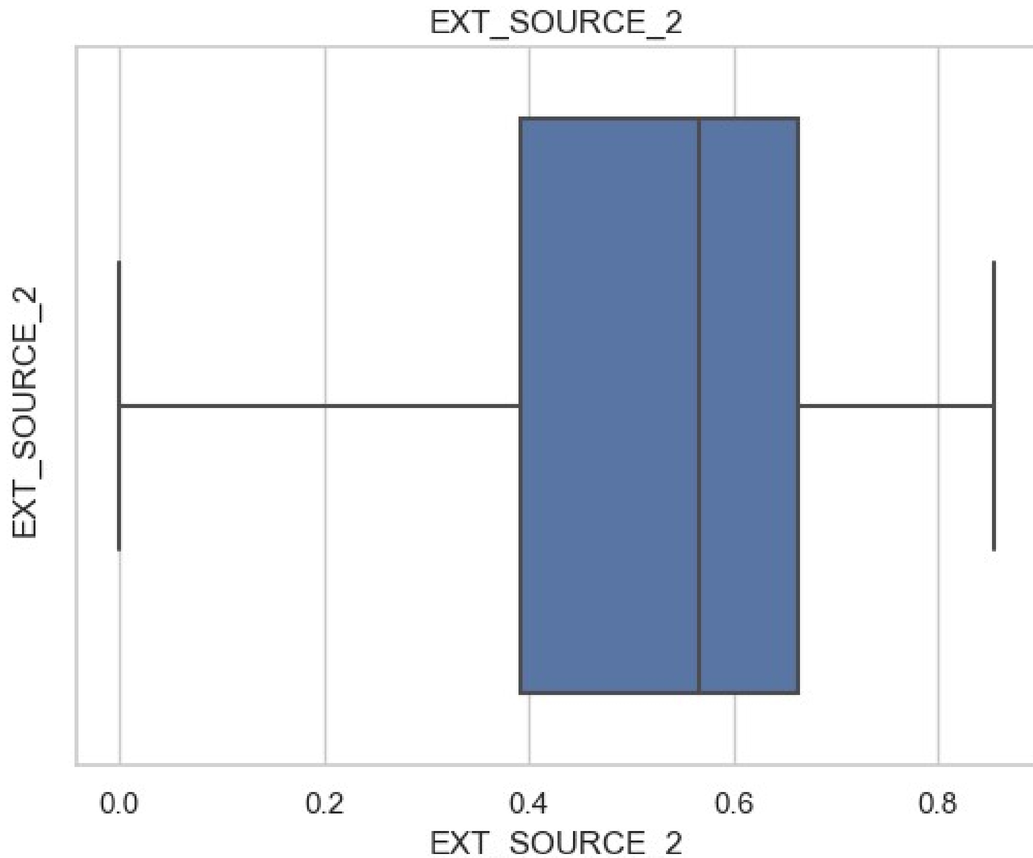

LIVE_REGION_NOT_WORK_REGION
LIVE_REGION_NOT_WORK_REGION
0.0
0.2
0.4
0.6
0.8
1.0
LIVE_REGION_NOT_WORK_REGION

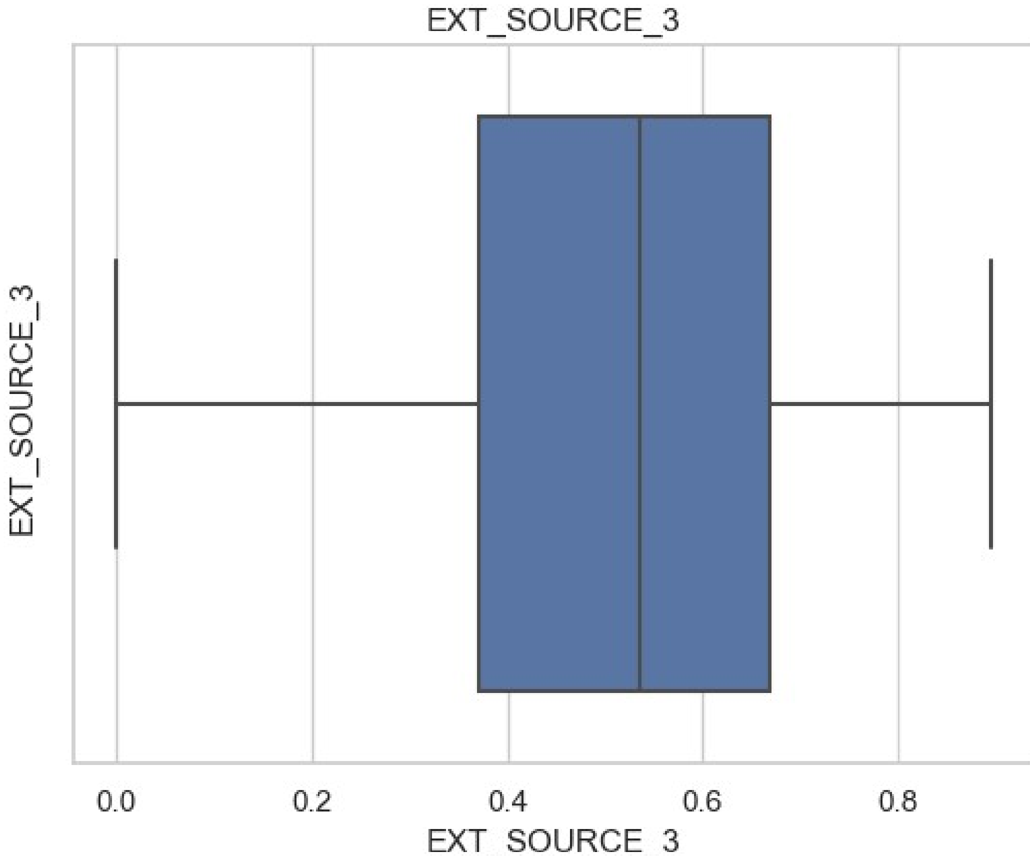

REG_CITY_NOT_LIVE_CITY
REG_CITY_NOT_LIVE_CITY
0.0
0.2
0.4
0.6
0.8
1.0
REG_CITY_NOT_LIVE_CITY

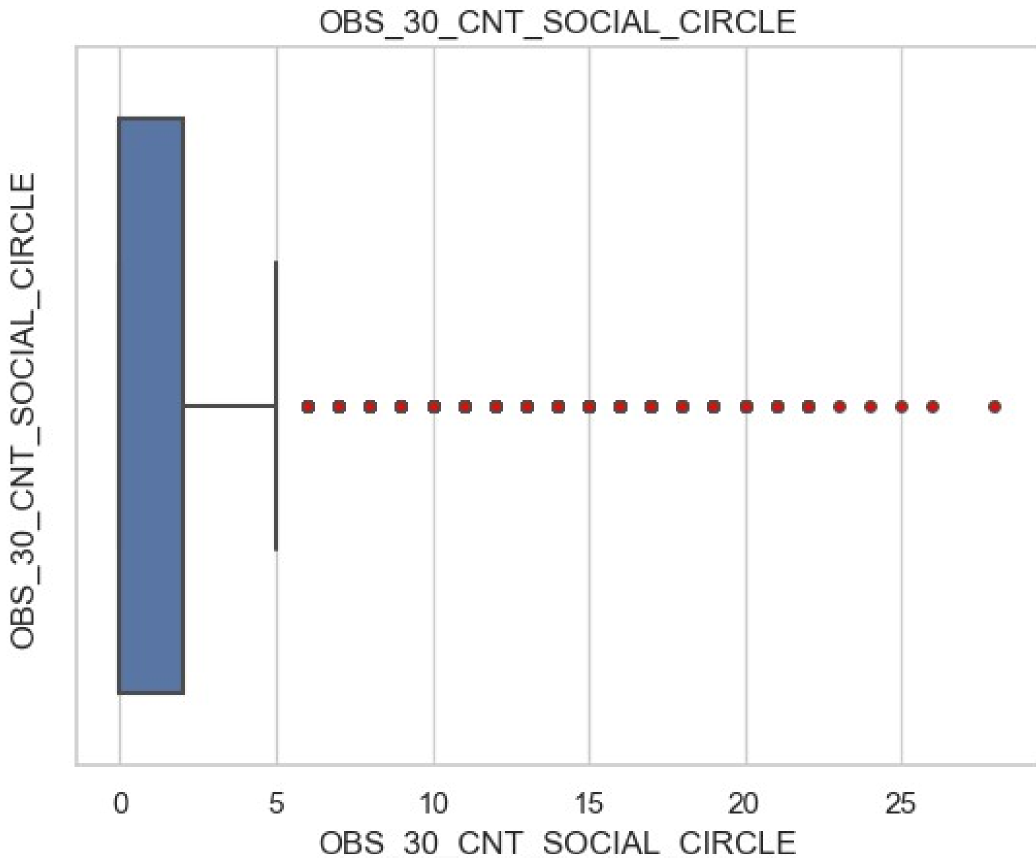

REG_CITY_NOT_WORK_CITY
REG_CITY_NOT_WORK_CITY
0.0
0.2
0.4
0.6
0.8
1.0
REG_CITY_NOT_WORK_CITY

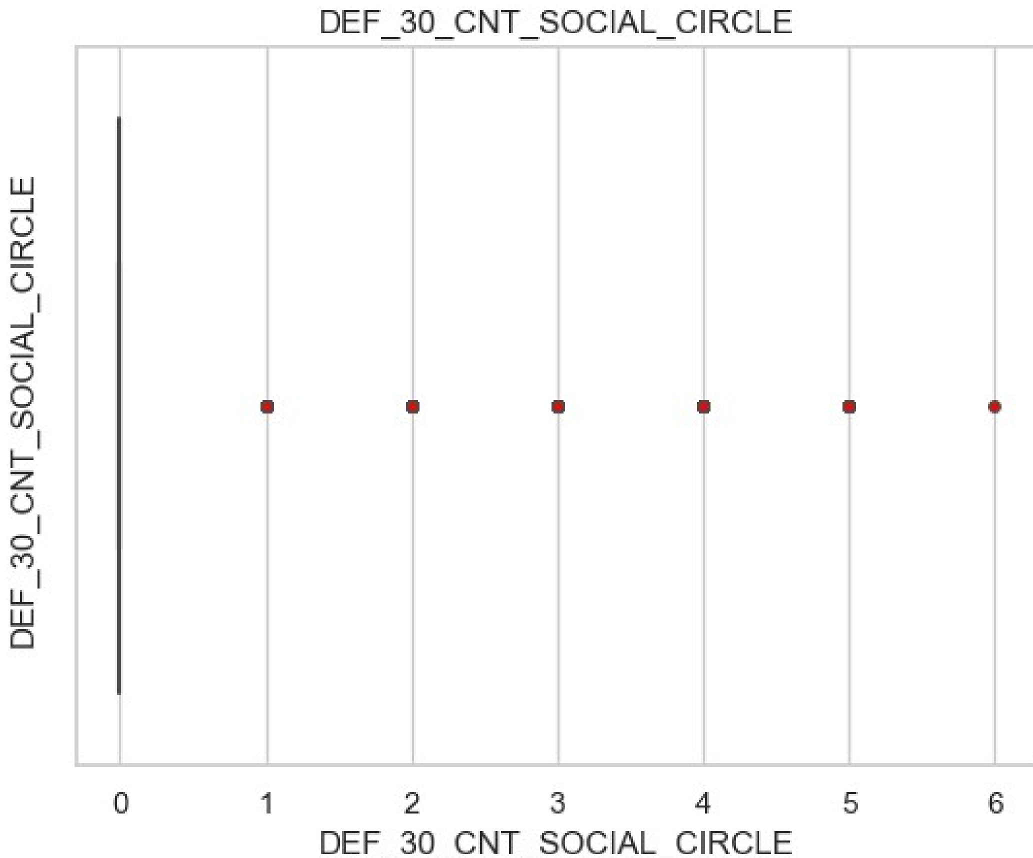

LIVE_CITY_NOT_WORK_CITY
LIVE_CITY_NOT_WORK_CITY
0.0
0.2
0.4
0.6
0.8
1.0
LIVE_CITY_NOT_WORK_CITY

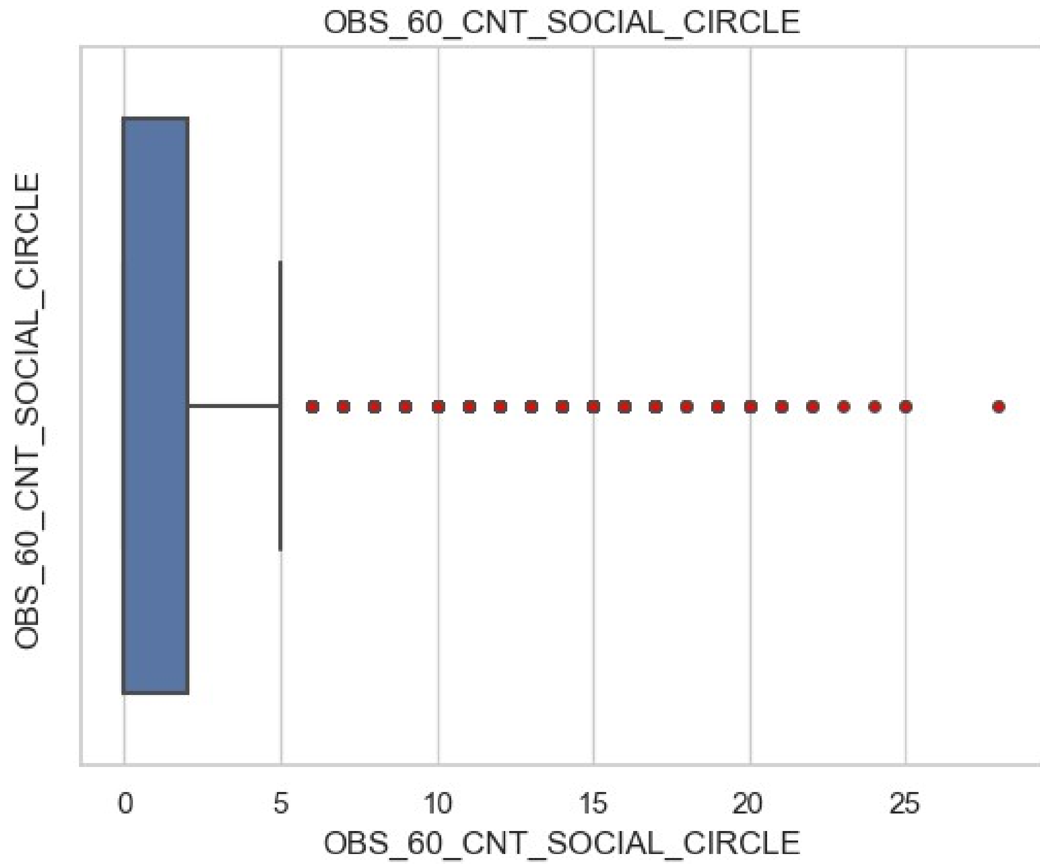

EXT_SOURCE_1
EXT_SOURCE_1
0.0
0.2
0.4
0.6
0.8
EXT_SOURCE_1

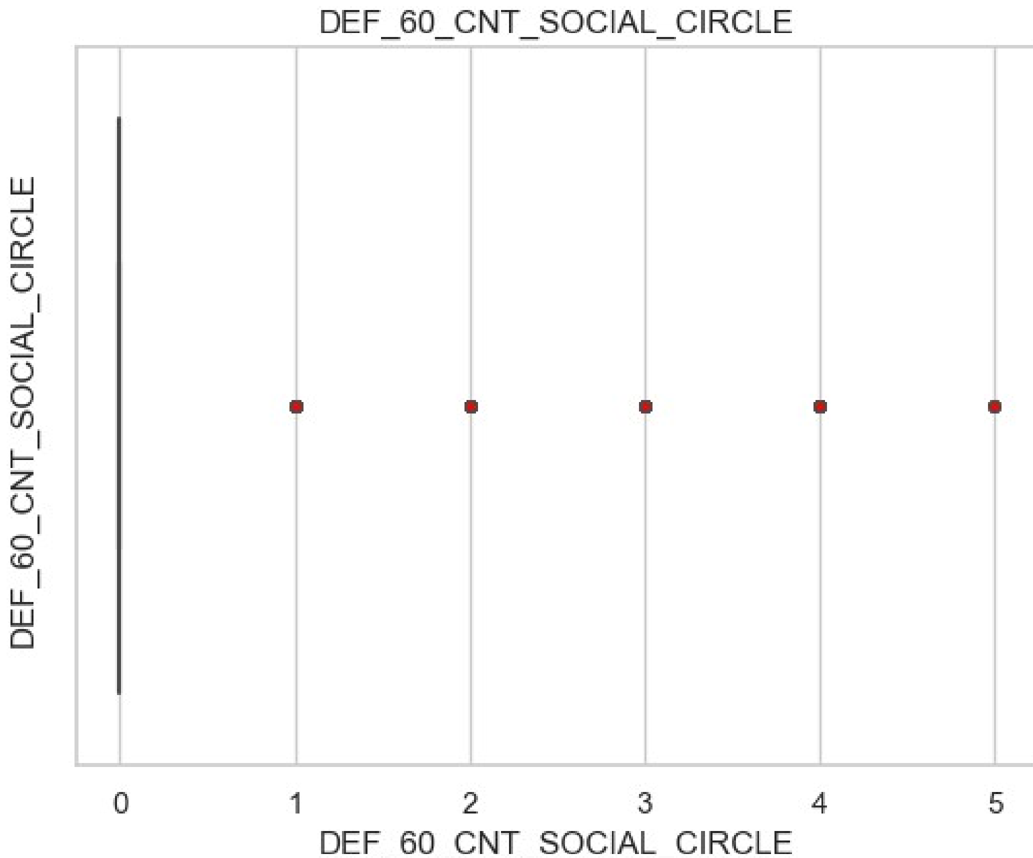

EXT_SOURCE_2
EXT_SOURCE_2
0.0
0.2
0.4
0.6
0.8
EXT_SOURCE_2

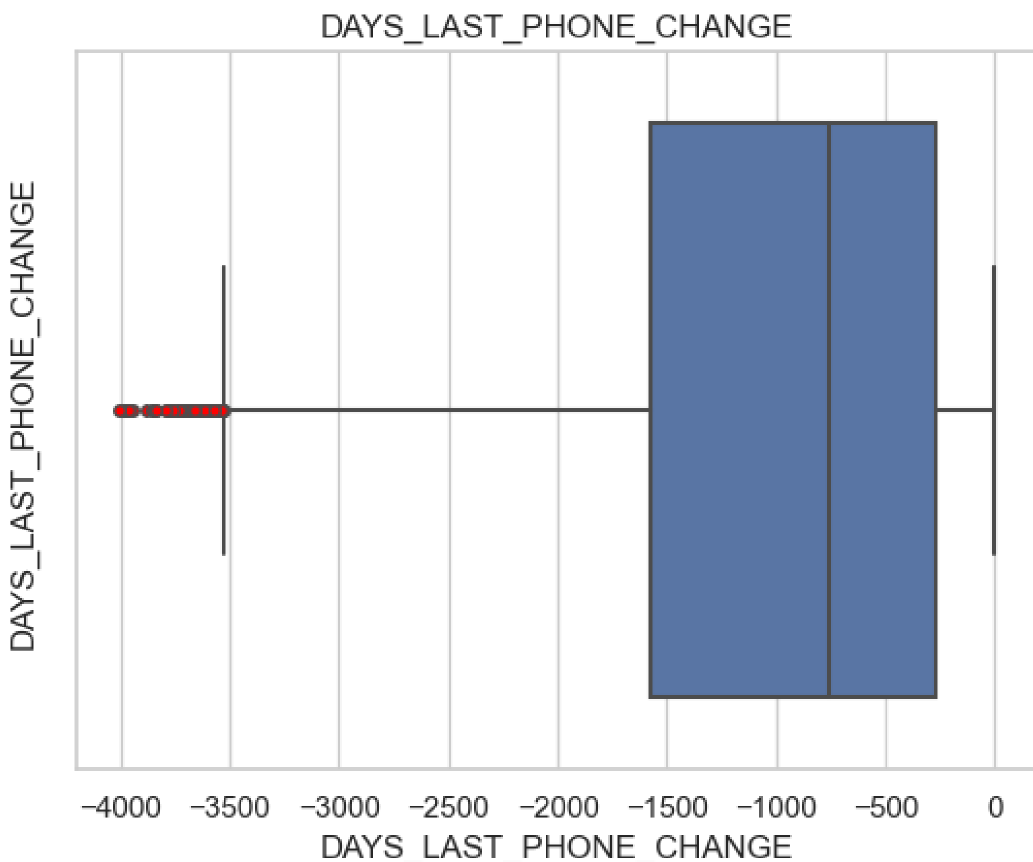

EXT_SOURCE_3
EXT_SOURCE_3
0.0
0.2
0.4
0.6
0.8
EXT_SOURCE_3

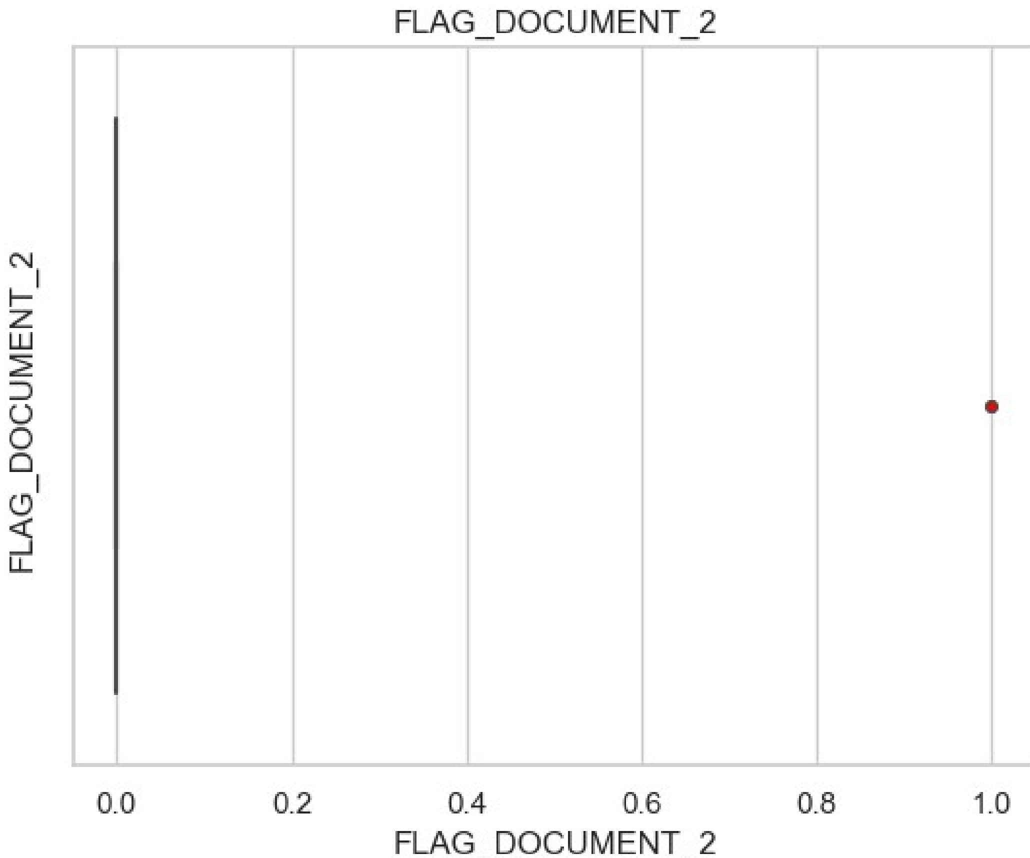

OBS_30_CNT_SOCIAL_CIRCLE
OBS_30_CNT_SOCIAL_CIRCLE
0
5
10
15
20
25
OBS_30_CNT_SOCIAL_CIRCLE

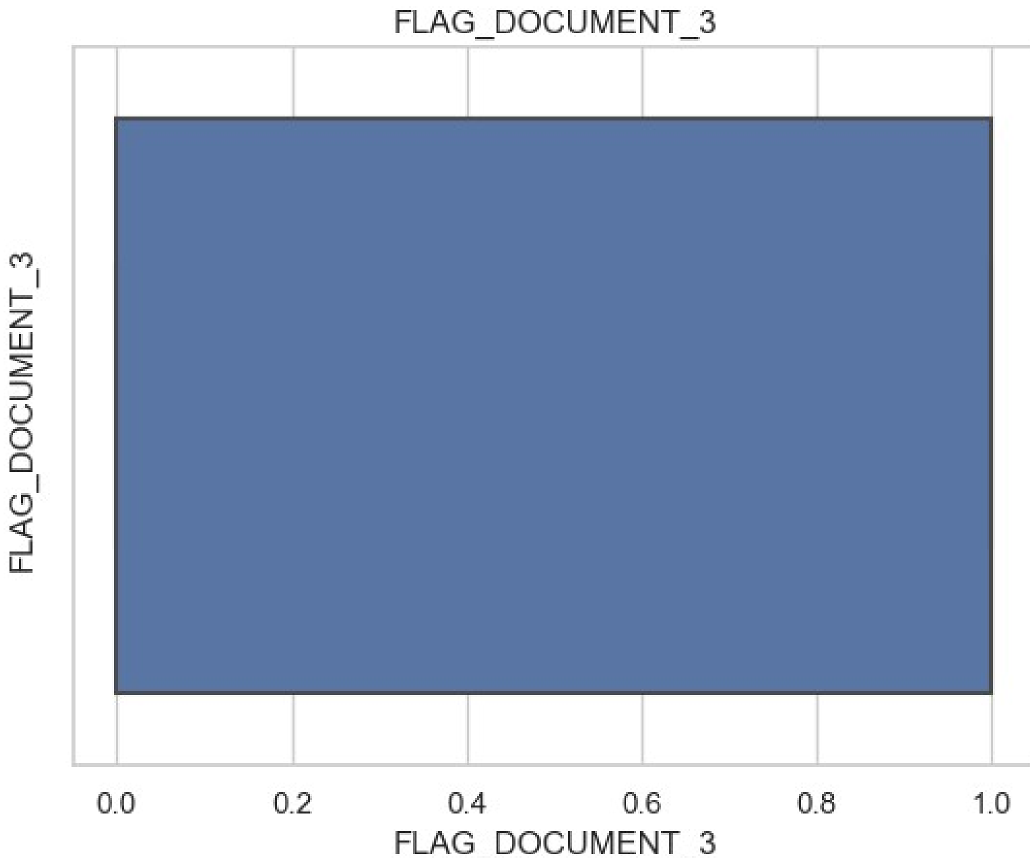

DEF_30_CNT_SOCIAL_CIRCLE
DEF_30_CNT_SOCIAL_CIRCLE
0
1
2
3
4
5
6
DEF_30_CNT_SOCIAL_CIRCLE

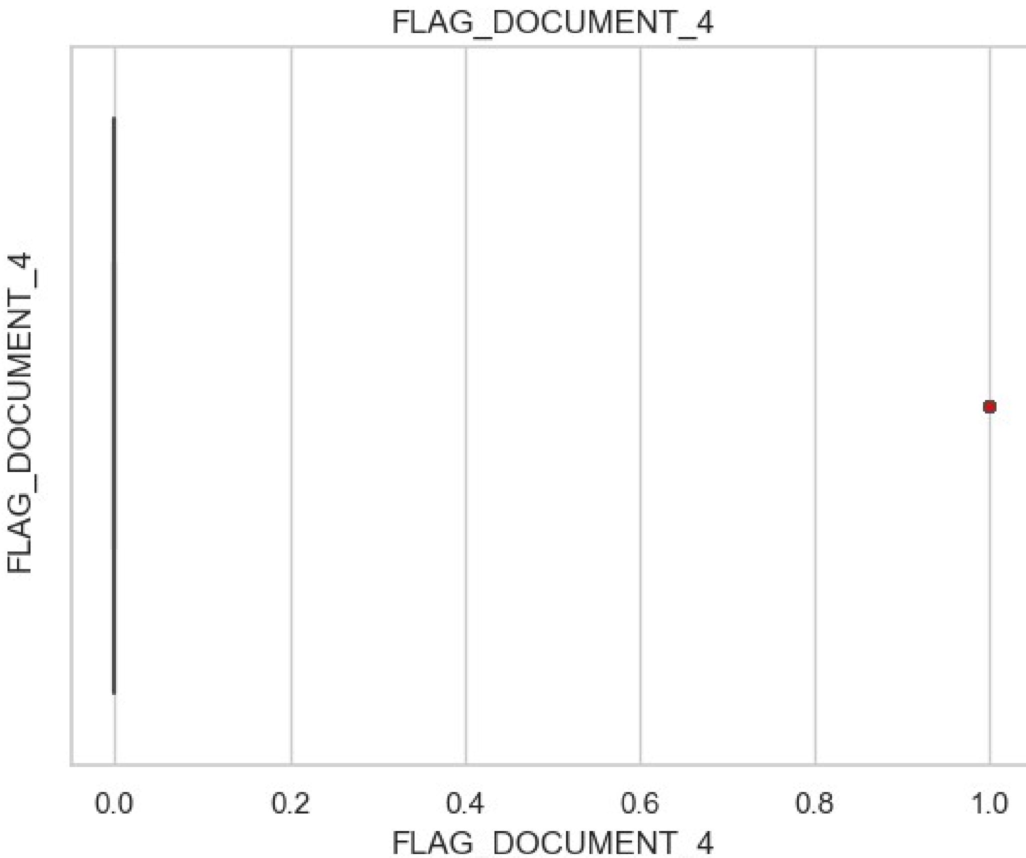

OBS_60_CNT_SOCIAL_CIRCLE
OBS_60_CNT_SOCIAL_CIRCLE
0
5
10
15
20
25
OBS_60_CNT_SOCIAL_CIRCLE

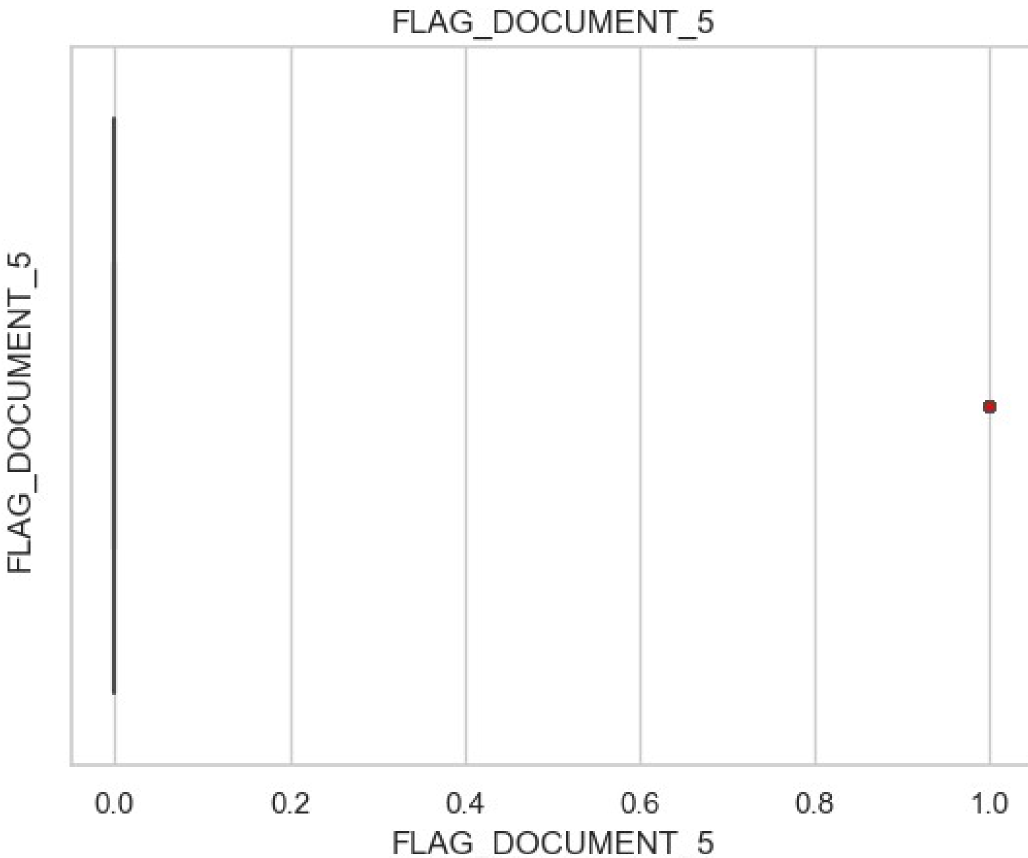

DEF_60_CNT_SOCIAL_CIRCLE
DEF_60_CNT_SOCIAL_CIRCLE
0
1
2
3
4
5
DEF_60_CNT_SOCIAL_CIRCLE

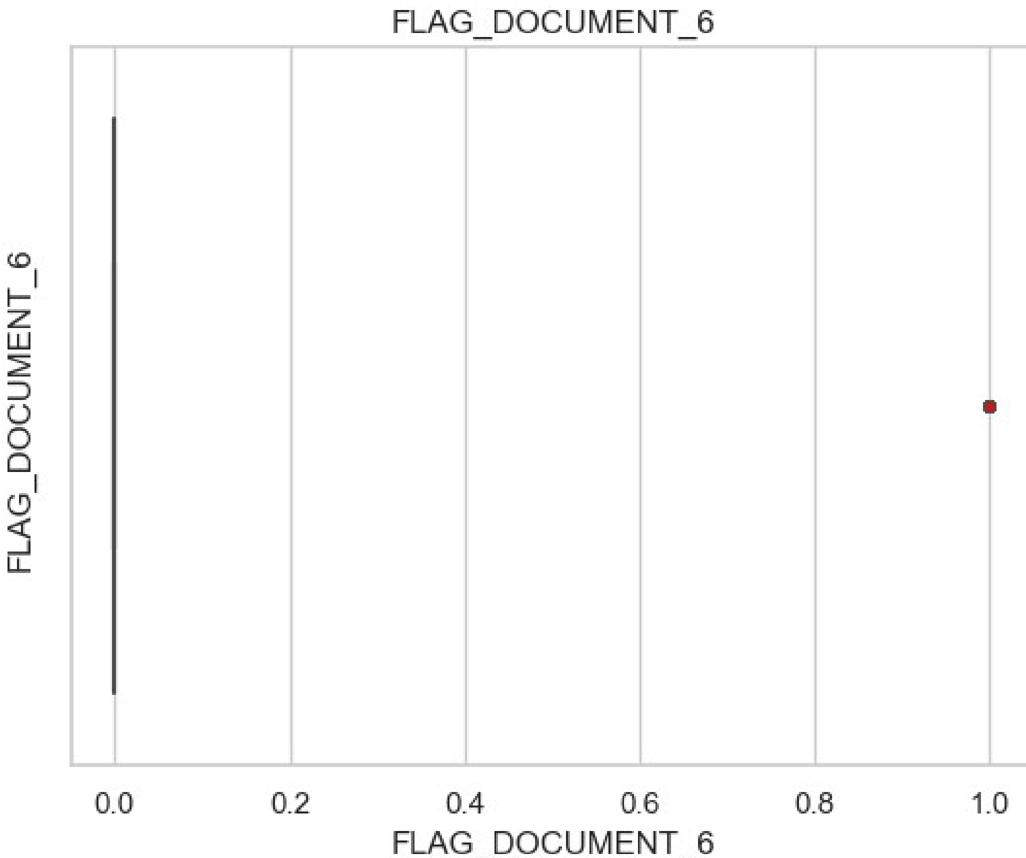

DAYS_LAST_PHONE_CHANGE
DAYS_LAST_PHONE_CHANGE
−4000
−3500
−3000
−2500
−2000
−1500
−1000
−500
0
DAYS_LAST_PHONE_CHANGE

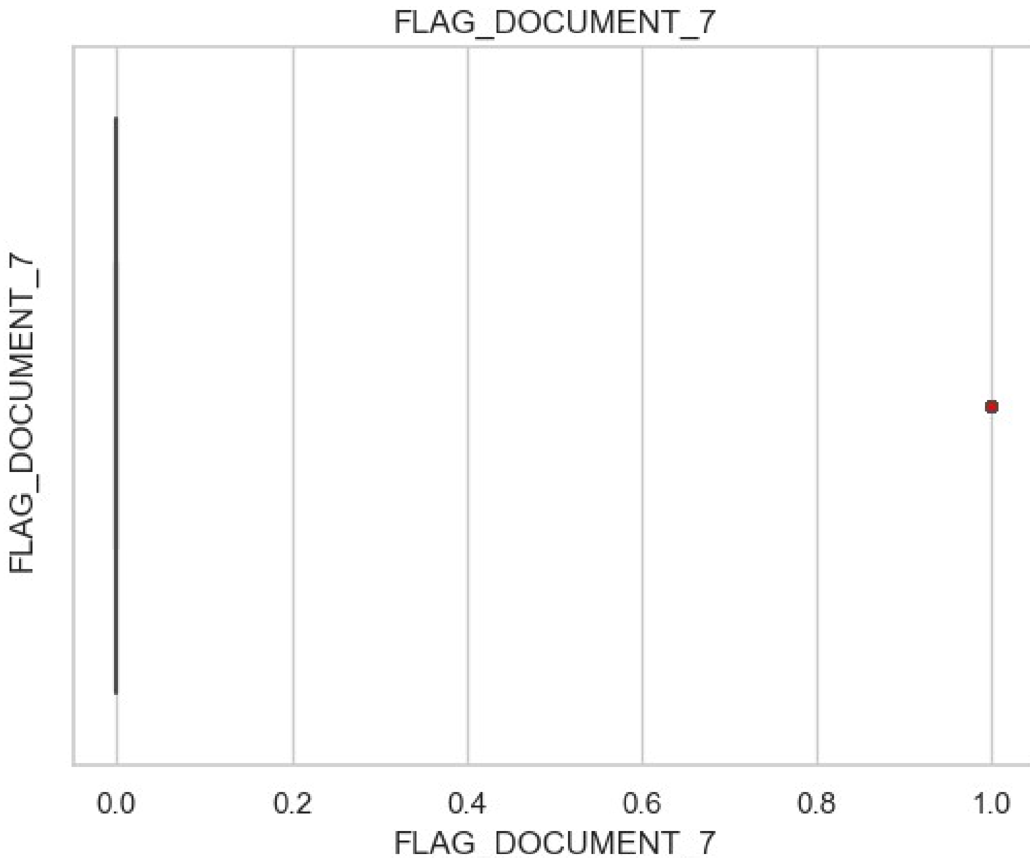

FLAG_DOCUMENT_2
FLAG_DOCUMENT_2
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_2

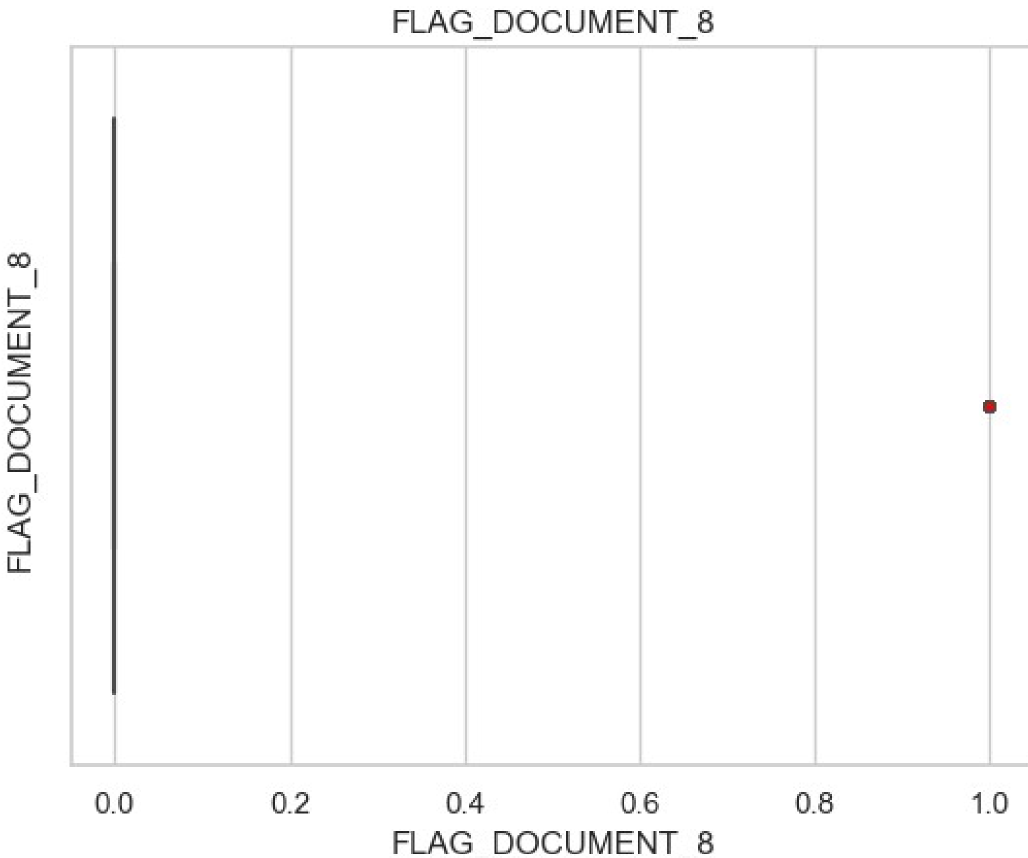

FLAG_DOCUMENT_3
FLAG_DOCUMENT_3
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_3

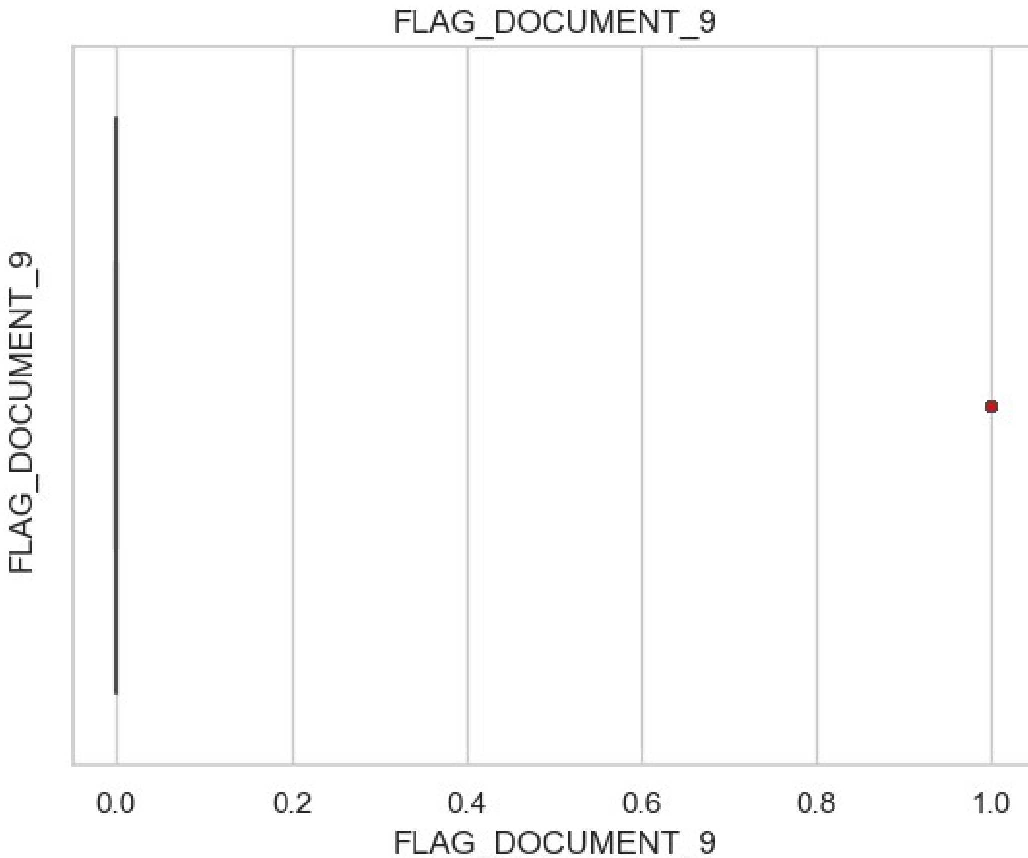

FLAG_DOCUMENT_4
FLAG_DOCUMENT_4
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_4

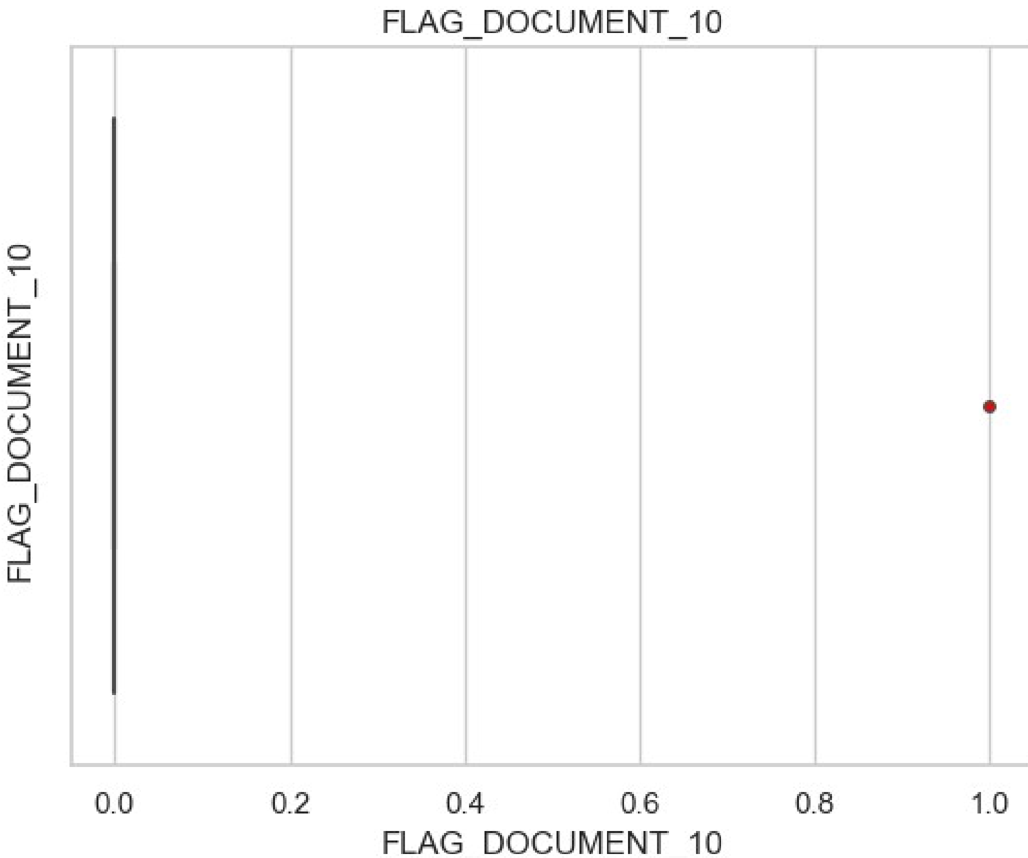

FLAG_DOCUMENT_5
FLAG_DOCUMENT_5
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_5

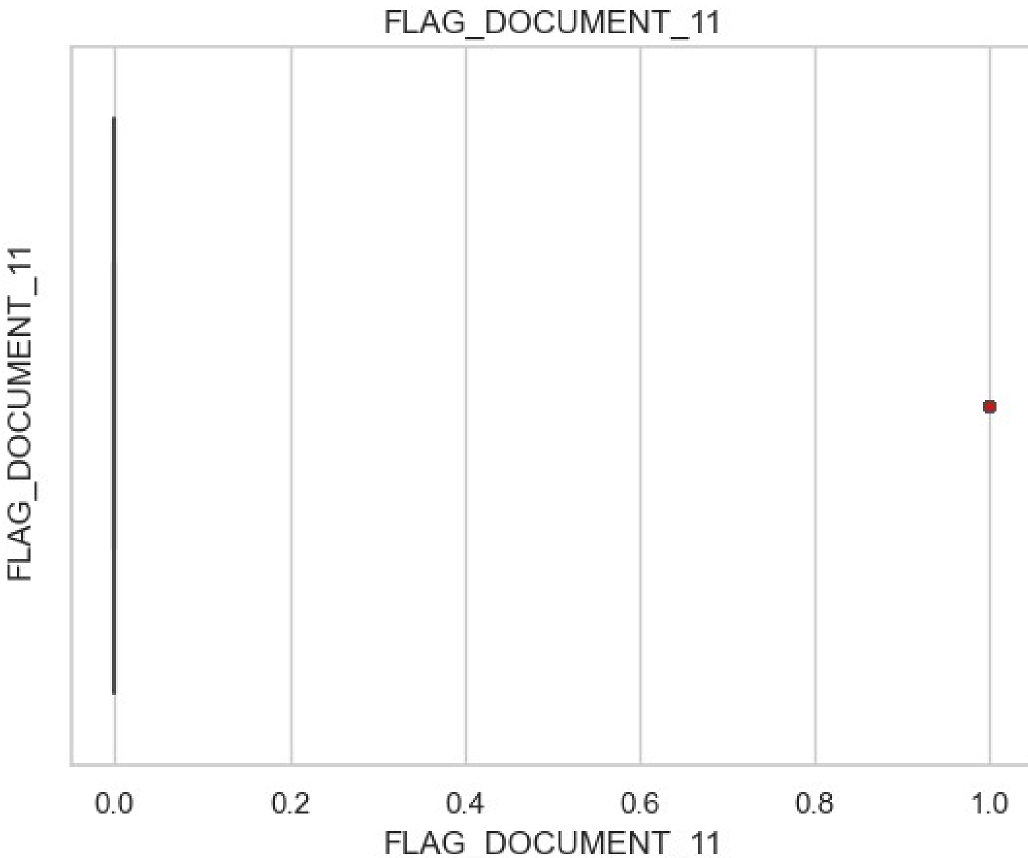

FLAG_DOCUMENT_6
FLAG_DOCUMENT_6
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_6

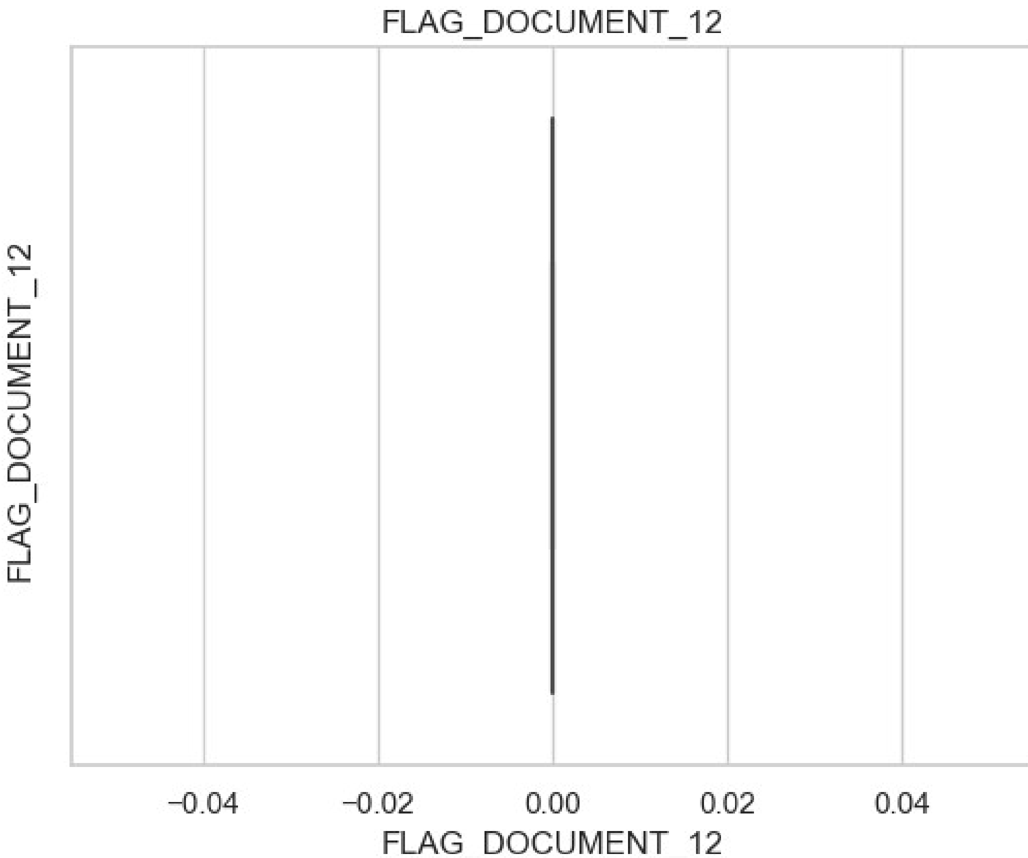

FLAG_DOCUMENT_7
FLAG_DOCUMENT_7
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_7

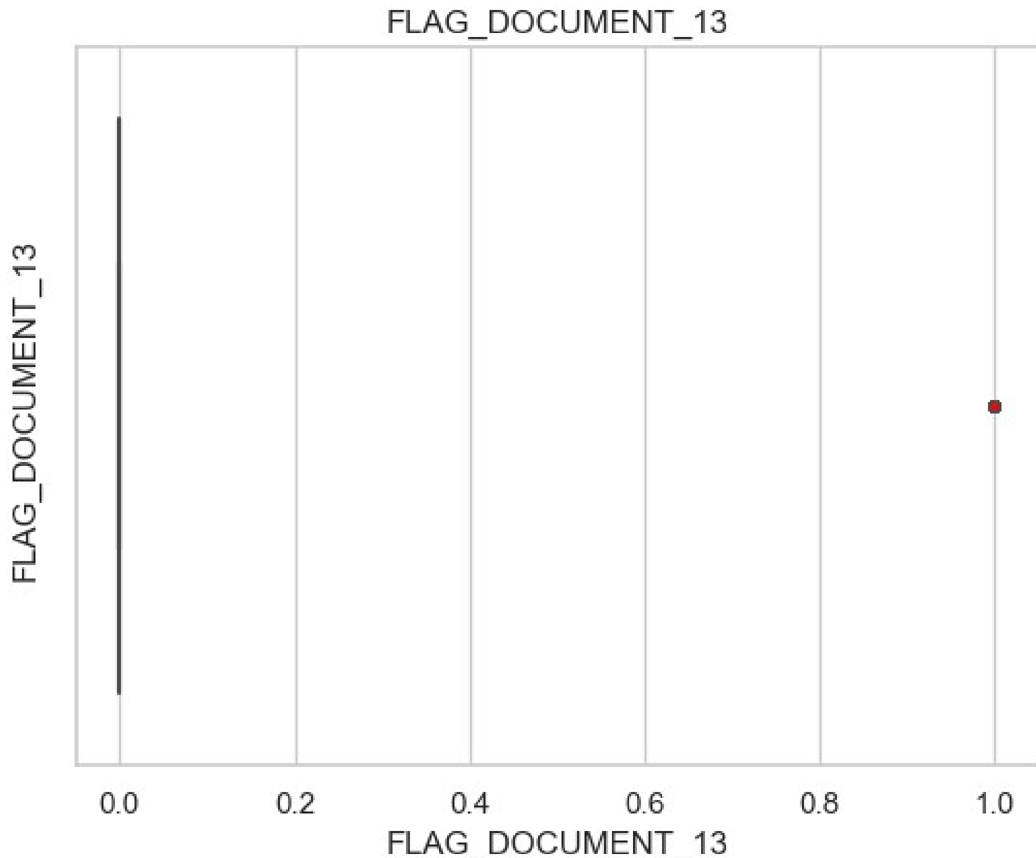

FLAG_DOCUMENT_8
FLAG_DOCUMENT_8
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_8

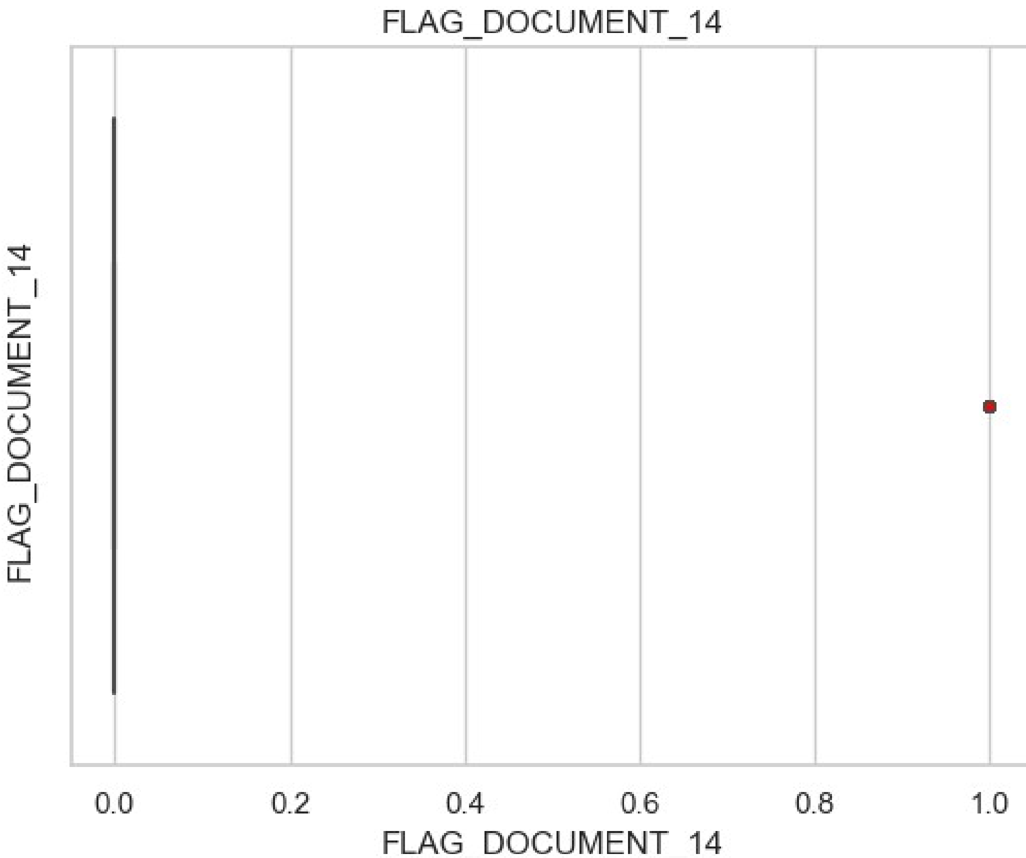

FLAG_DOCUMENT_9
FLAG_DOCUMENT_9
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_9

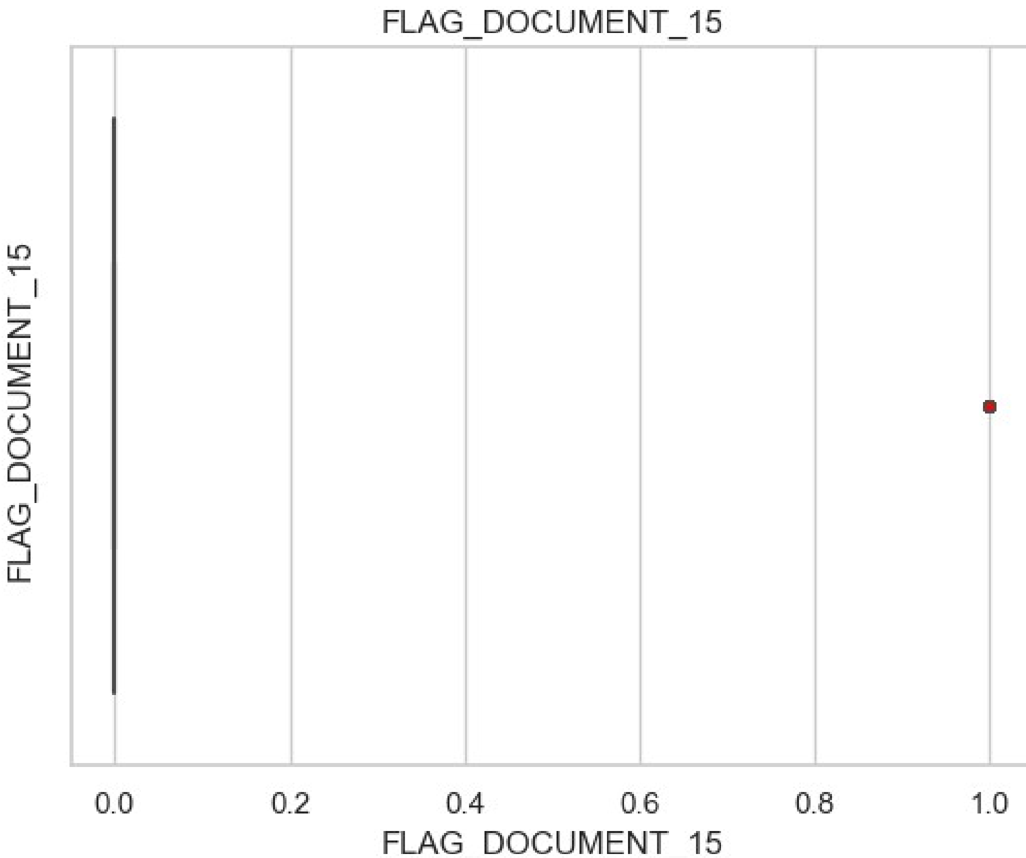

FLAG_DOCUMENT_10
FLAG_DOCUMENT_10
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_10

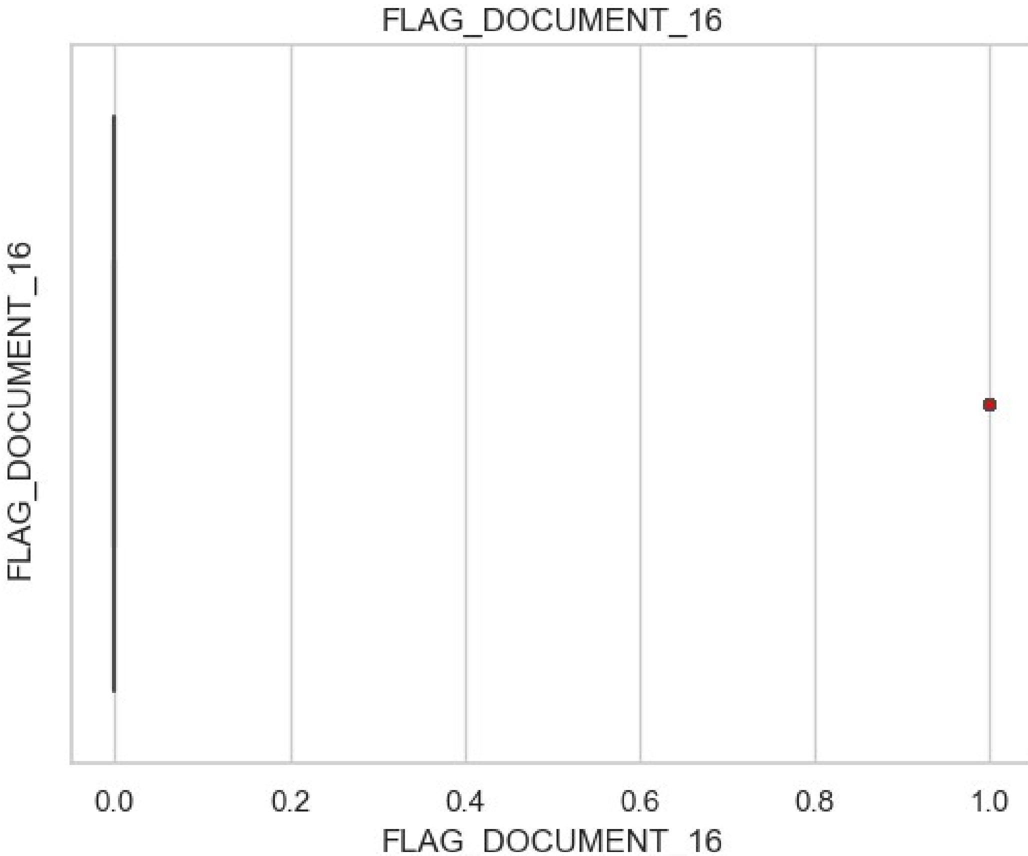

FLAG_DOCUMENT_11
FLAG_DOCUMENT_11
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_11

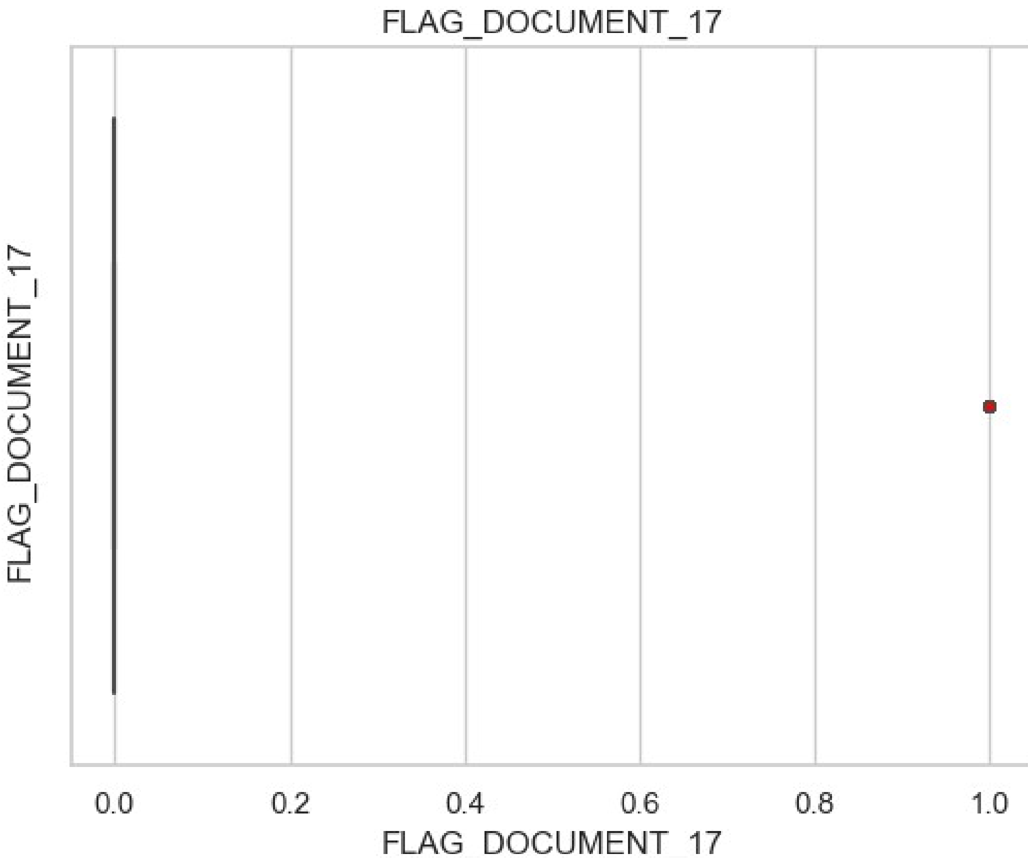

FLAG_DOCUMENT_12
FLAG_DOCUMENT_12
−0.04
−0.02
0.00
0.02
0.04
FLAG_DOCUMENT_12

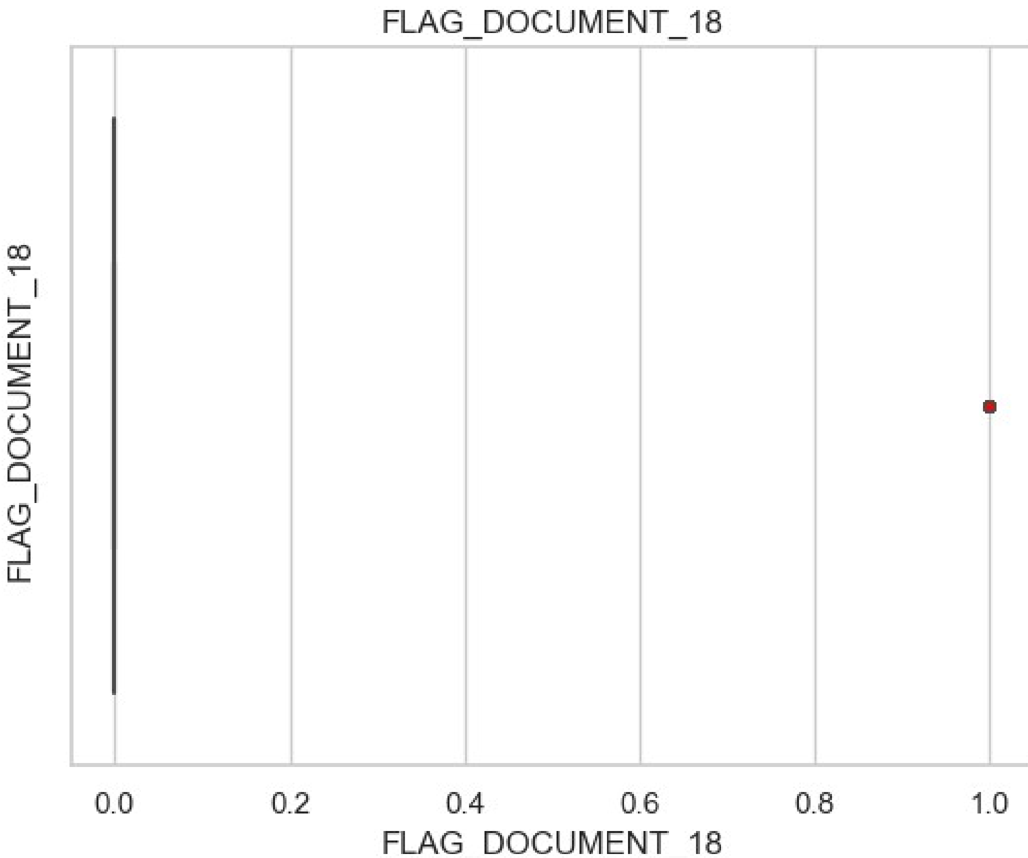

FLAG_DOCUMENT_13
FLAG_DOCUMENT_13
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_13

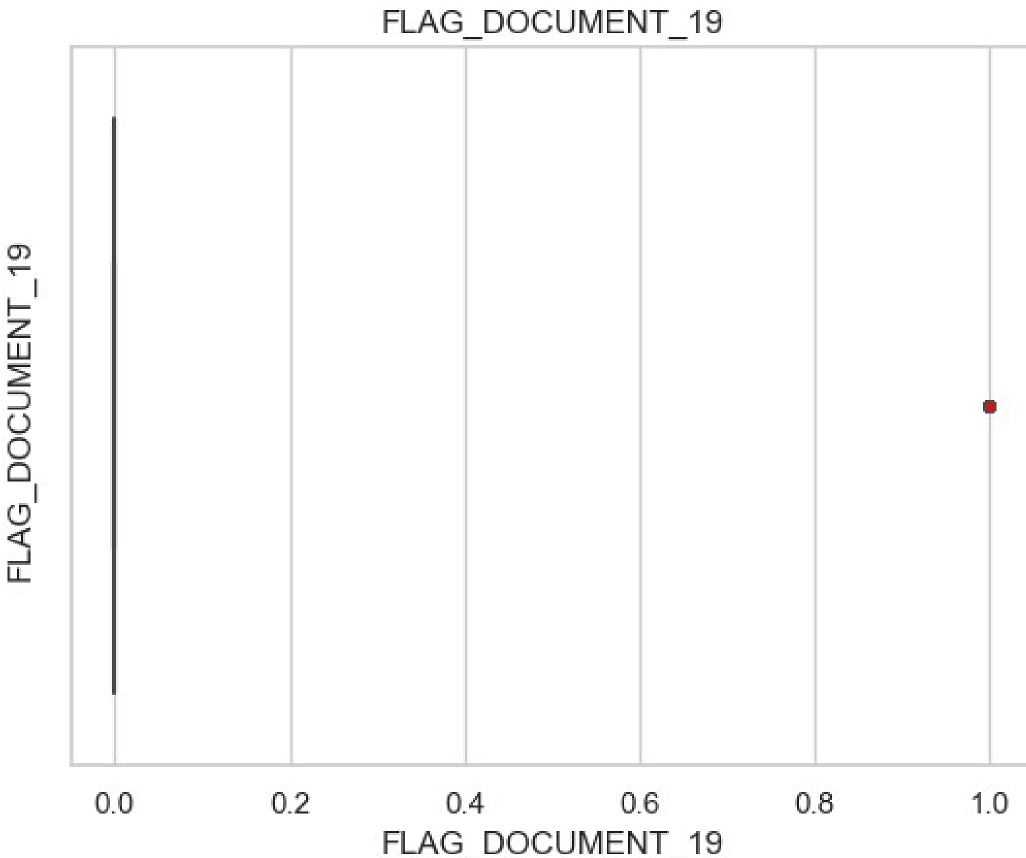

FLAG_DOCUMENT_14
FLAG_DOCUMENT_14
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_14

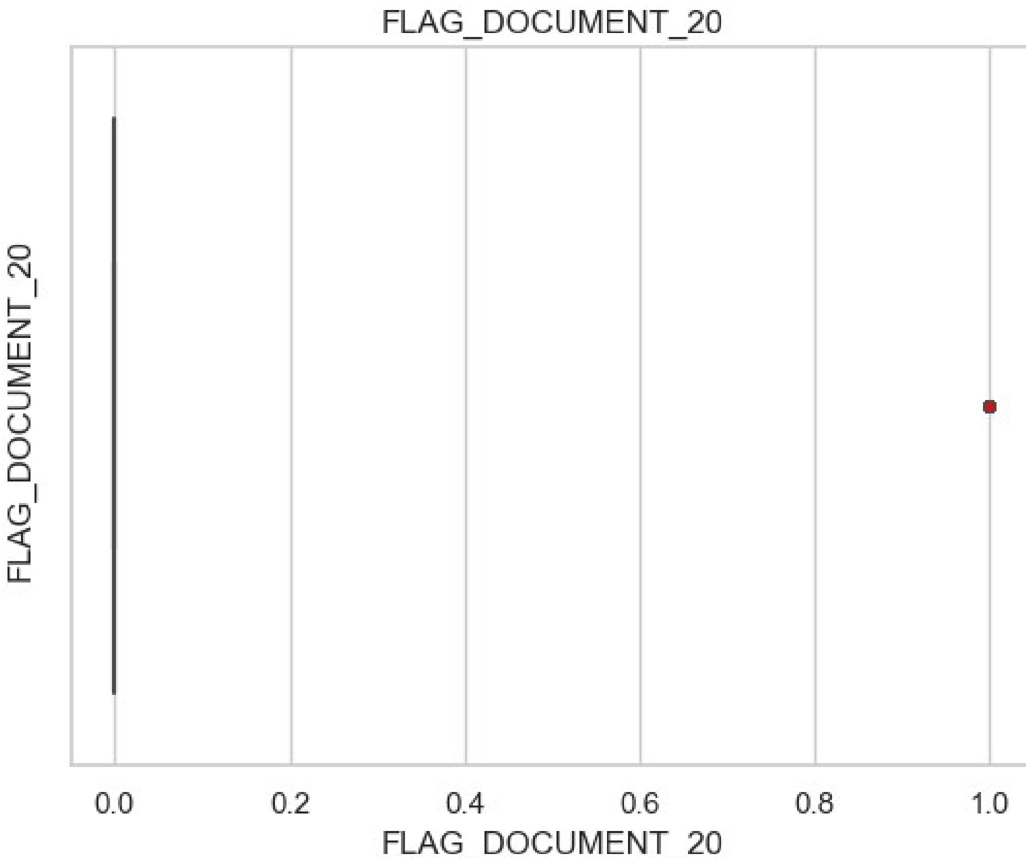

FLAG_DOCUMENT_15
FLAG_DOCUMENT_15
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_15

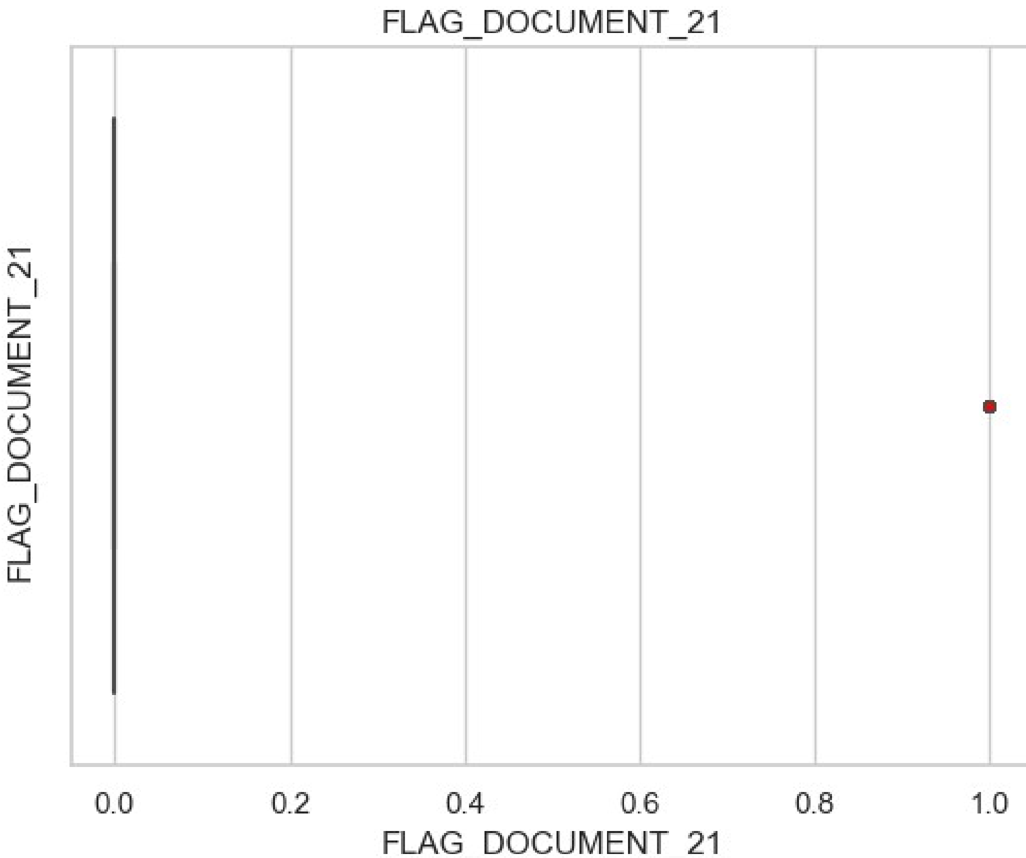

FLAG_DOCUMENT_16
FLAG_DOCUMENT_16
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_16

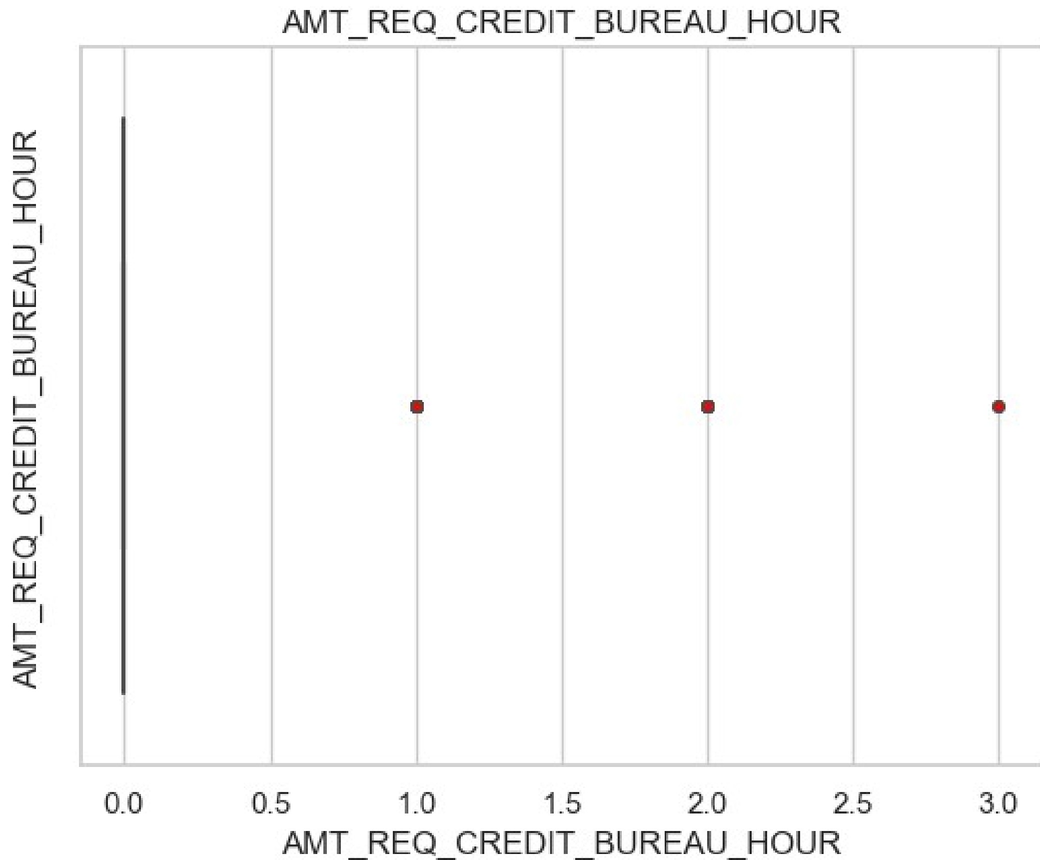

FLAG_DOCUMENT_17
FLAG_DOCUMENT_17
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_17

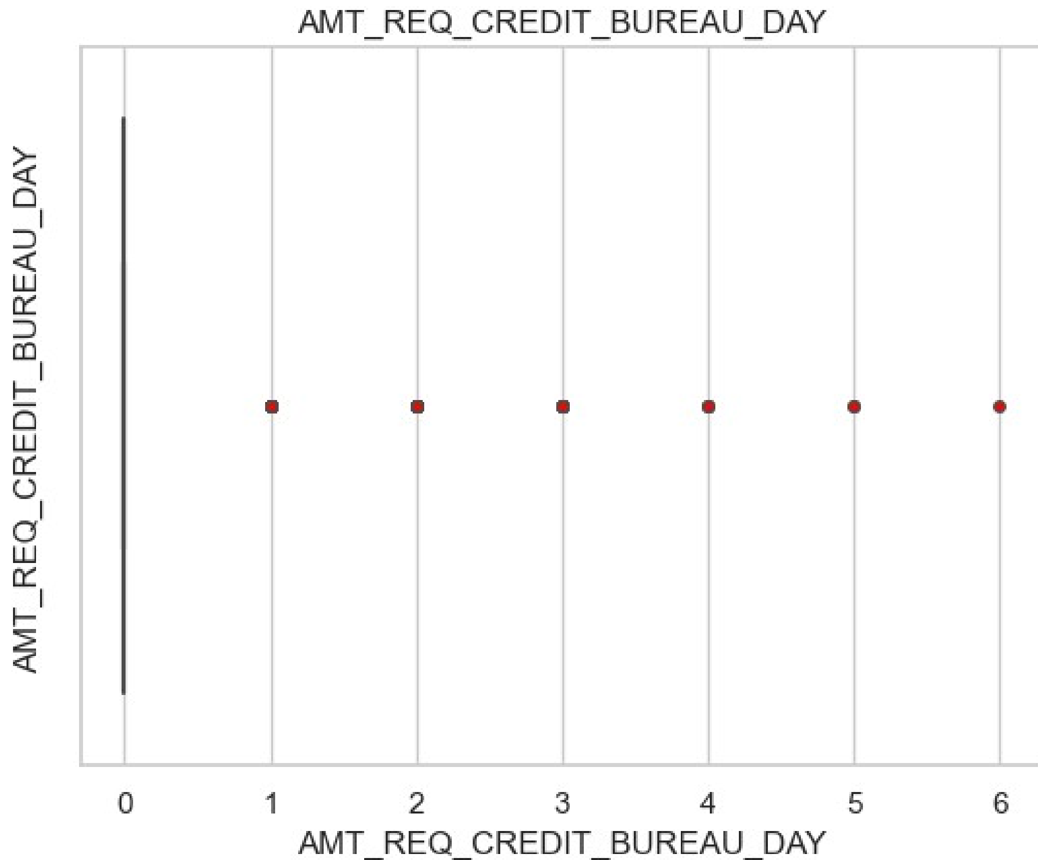

FLAG_DOCUMENT_18
FLAG_DOCUMENT_18
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_18

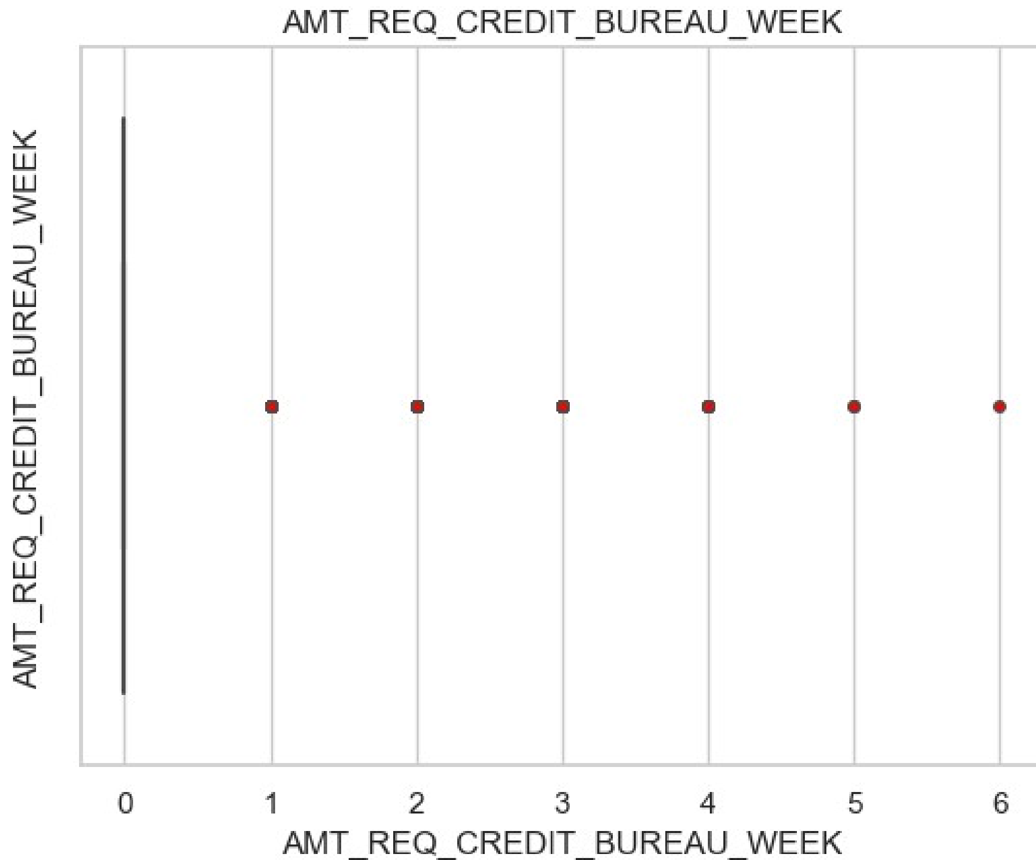

FLAG_DOCUMENT_19
FLAG_DOCUMENT_19
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_19

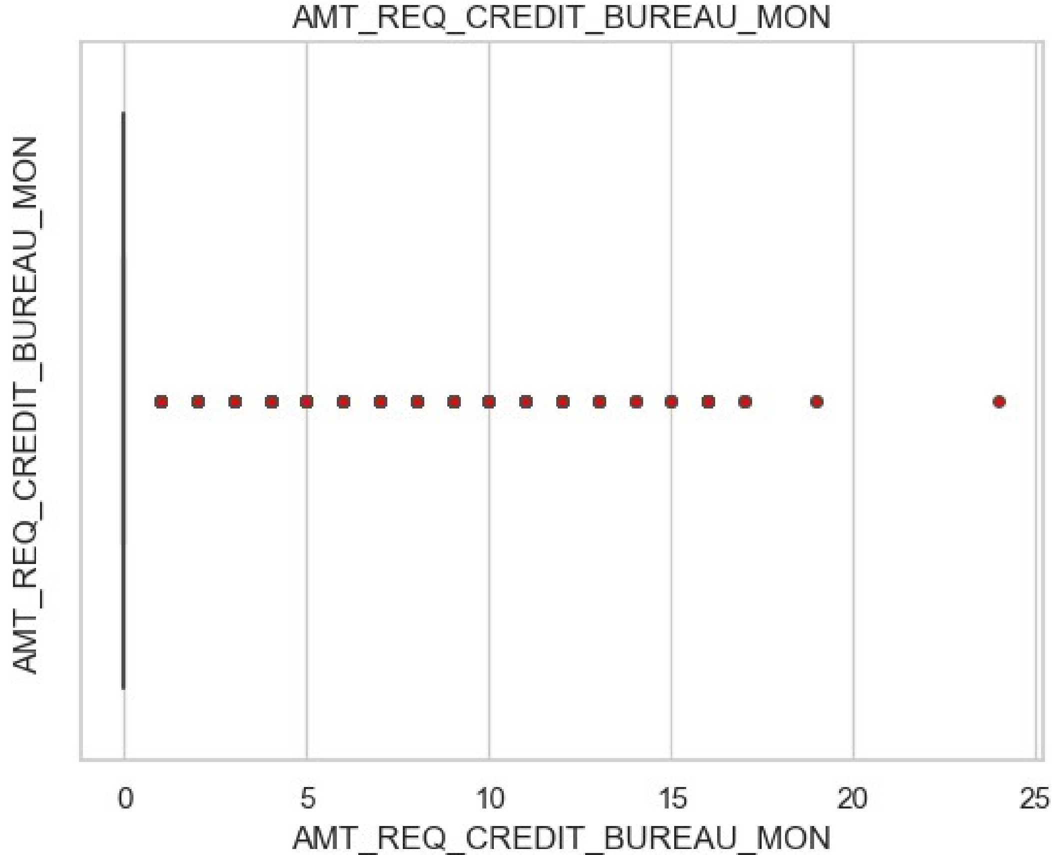

FLAG_DOCUMENT_20
FLAG_DOCUMENT_20
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_20

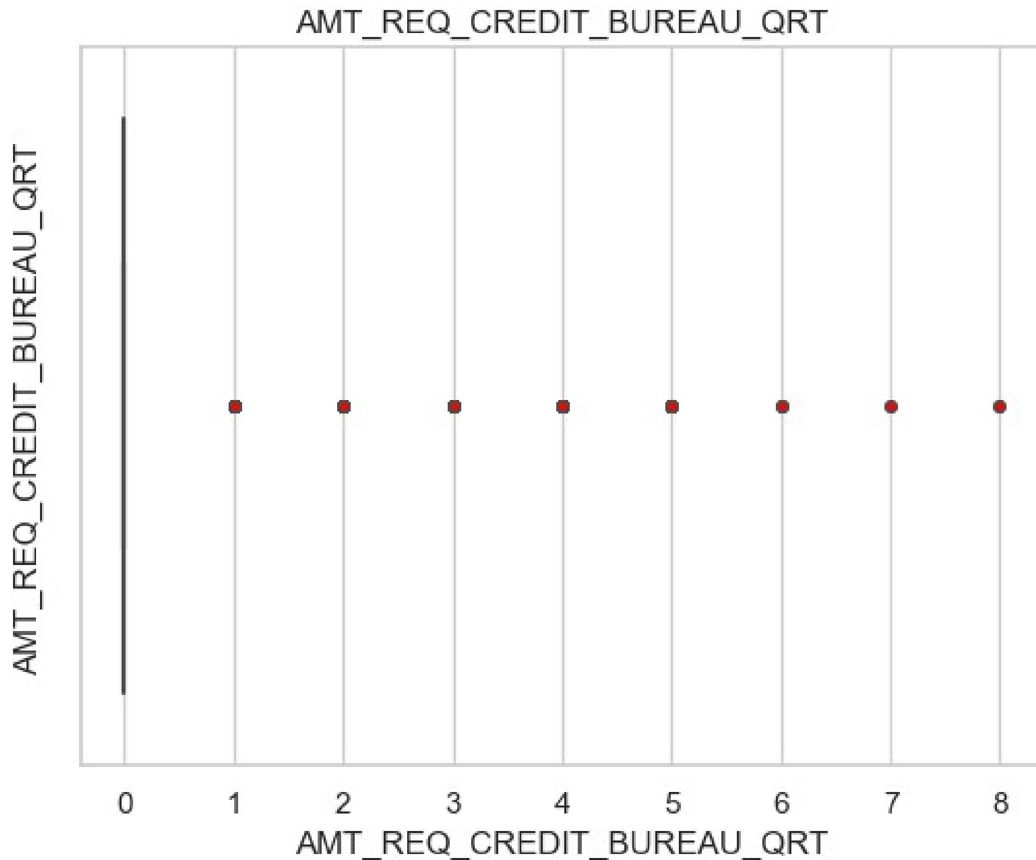

FLAG_DOCUMENT_21
FLAG_DOCUMENT_21
0.0
0.2
0.4
0.6
0.8
1.0
FLAG_DOCUMENT_21

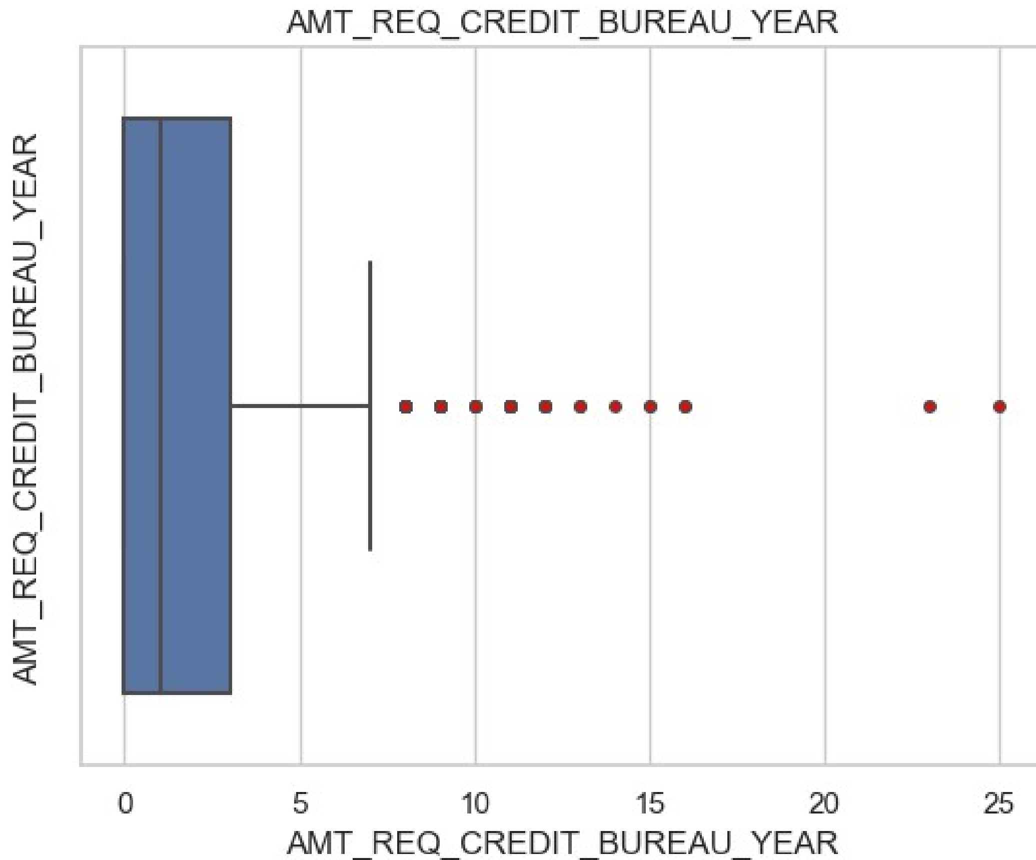

AMT_REQ_CREDIT_BUREAU_HOUR
AMT_REQ_CREDIT_BUREAU_HOUR
0.0
0.5
1.0
1.5
2.0
2.5
3.0
AMT_REQ_CREDIT_BUREAU_HOUR


AMT_REQ_CREDIT_BUREAU_DAY
AMT_REQ_CREDIT_BUREAU_DAY
0
1
2
3
4
5
6
AMT_REQ_CREDIT_BUREAU_DAY


AMT_REQ_CREDIT_BUREAU_WEEK
AMT_REQ_CREDIT_BUREAU_WEEK
0
1
2
3
4
5
6
AMT_REQ_CREDIT_BUREAU_WEEK


AMT_REQ_CREDIT_BUREAU_MON
AMT_REQ_CREDIT_BUREAU_MON
0
5
10
15
20
25
AMT_REQ_CREDIT_BUREAU_MON


AMT_REQ_CREDIT_BUREAU_QRT
AMT_REQ_CREDIT_BUREAU_QRT
0
1
2
3
4
5
6
7
8
AMT_REQ_CREDIT_BUREAU_QRT




```python
# let's see how many outliers each variable have

for key, value in dictoutlier.items():
    if len(value) > 0:
        print(f"Length of {key}: {len(value)}")
```

```
Length of TARGET: 4026
Length of CNT_CHILDREN: 723
Length of AMT_INCOME_TOTAL: 2295
Length of AMT_CREDIT: 1063
Length of REGION_POPULATION_RELATIVE: 1329
Length of DAYS_EMPLOYED: 9082
Length of DAYS_REGISTRATION: 96
Length of FLAG_MOBIL: 1
Length of FLAG_EMP_PHONE: 8926
Length of FLAG_WORK_PHONE: 9963
Length of FLAG_CONT_MOBILE: 101
Length of FLAG_EMAIL: 2783
Length of CNT_FAM_MEMBERS: 684
Length of REGION_RATING_CLIENT: 13035
Length of REGION_RATING_CLIENT_W_CITY: 12658
Length of HOUR_APPR_PROCESS_START: 353
Length of REG_REGION_NOT_LIVE_REGION: 750
```

```
Length of REG_REGION_NOT_WORK_REGION: 2496
Length of LIVE_REGION_NOT_WORK_REGION: 1982
Length of REG_CITY_NOT_LIVE_CITY: 3998
Length of REG_CITY_NOT_WORK_CITY: 11608
Length of LIVE_CITY_NOT_WORK_CITY: 8985
Length of FLAG_DOCUMENT_2: 2
Length of FLAG_DOCUMENT_4: 9
Length of FLAG_DOCUMENT_5: 785
Length of FLAG_DOCUMENT_6: 4335
Length of FLAG_DOCUMENT_7: 11
Length of FLAG_DOCUMENT_8: 4038
Length of FLAG_DOCUMENT_9: 184
Length of FLAG_DOCUMENT_10: 1
Length of FLAG_DOCUMENT_11: 213
Length of FLAG_DOCUMENT_13: 161
Length of FLAG_DOCUMENT_14: 158
Length of FLAG_DOCUMENT_15: 41
Length of FLAG_DOCUMENT_16: 501
Length of FLAG_DOCUMENT_17: 15
Length of FLAG_DOCUMENT_18: 425
Length of FLAG_DOCUMENT_19: 35
Length of FLAG_DOCUMENT_20: 26
Length of FLAG_DOCUMENT_21: 19
Length of AMT_REQ_CREDIT_BUREAU_HOUR: 295
Length of AMT_REQ_CREDIT_BUREAU_DAY: 272
Length of AMT_REQ_CREDIT_BUREAU_WEEK: 1314
Length of AMT_REQ_CREDIT_BUREAU_MON: 7140
Length of AMT_REQ_CREDIT_BUREAU_QRT: 8134
Length of AMT_REQ_CREDIT_BUREAU_YEAR: 552
```

```python
# let's see how many unique outliers each variable have

for key, value in dictoutlier.items():
    if len(value) > 0:
        print(f"Length of {key}: {len(set(value))}")
```

```
Length of TARGET: 1
Length of CNT_CHILDREN: 8
Length of AMT_INCOME_TOTAL: 120
Length of AMT_CREDIT: 262
Length of REGION_POPULATION_RELATIVE: 1
Length of DAYS_EMPLOYED: 151
Length of DAYS_REGISTRATION: 96
Length of FLAG_MOBIL: 1
Length of FLAG_EMP_PHONE: 1
Length of FLAG_WORK_PHONE: 1
Length of FLAG_CONT_MOBILE: 1
Length of FLAG_EMAIL: 1
Length of CNT_FAM_MEMBERS: 8
Length of REGION_RATING_CLIENT: 2
```

```
Length of REGION_RATING_CLIENT_W_CITY: 2
Length of HOUR_APPR_PROCESS_START: 7
Length of REG_REGION_NOT_LIVE_REGION: 1
Length of REG_REGION_NOT_WORK_REGION: 1
Length of LIVE_REGION_NOT_WORK_REGION: 1
Length of REG_CITY_NOT_LIVE_CITY: 1
Length of REG_CITY_NOT_WORK_CITY: 1
Length of LIVE_CITY_NOT_WORK_CITY: 1
Length of FLAG_DOCUMENT_2: 1
Length of FLAG_DOCUMENT_4: 1
Length of FLAG_DOCUMENT_5: 1
Length of FLAG_DOCUMENT_6: 1
Length of FLAG_DOCUMENT_7: 1
Length of FLAG_DOCUMENT_8: 1
Length of FLAG_DOCUMENT_9: 1
Length of FLAG_DOCUMENT_10: 1
Length of FLAG_DOCUMENT_11: 1
Length of FLAG_DOCUMENT_13: 1
Length of FLAG_DOCUMENT_14: 1
Length of FLAG_DOCUMENT_15: 1
Length of FLAG_DOCUMENT_16: 1
Length of FLAG_DOCUMENT_17: 1
Length of FLAG_DOCUMENT_18: 1
Length of FLAG_DOCUMENT_19: 1
Length of FLAG_DOCUMENT_20: 1
Length of FLAG_DOCUMENT_21: 1
Length of AMT_REQ_CREDIT_BUREAU_HOUR: 3
Length of AMT_REQ_CREDIT_BUREAU_DAY: 6
Length of AMT_REQ_CREDIT_BUREAU_WEEK: 6
Length of AMT_REQ_CREDIT_BUREAU_MON: 19
Length of AMT_REQ_CREDIT_BUREAU_QRT: 8
Length of AMT_REQ_CREDIT_BUREAU_YEAR: 11
```

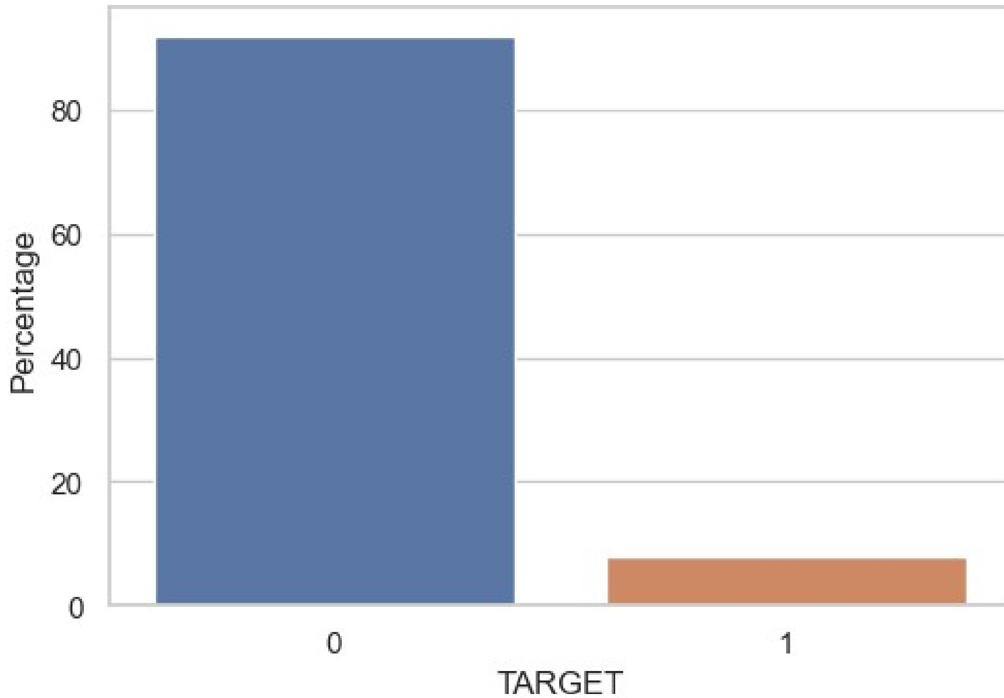
So here, outliers for each variable is found and stored in a dictionary named dictoutliers. We can use this result in our futhur study.

C. Analyze Data Imbalance:

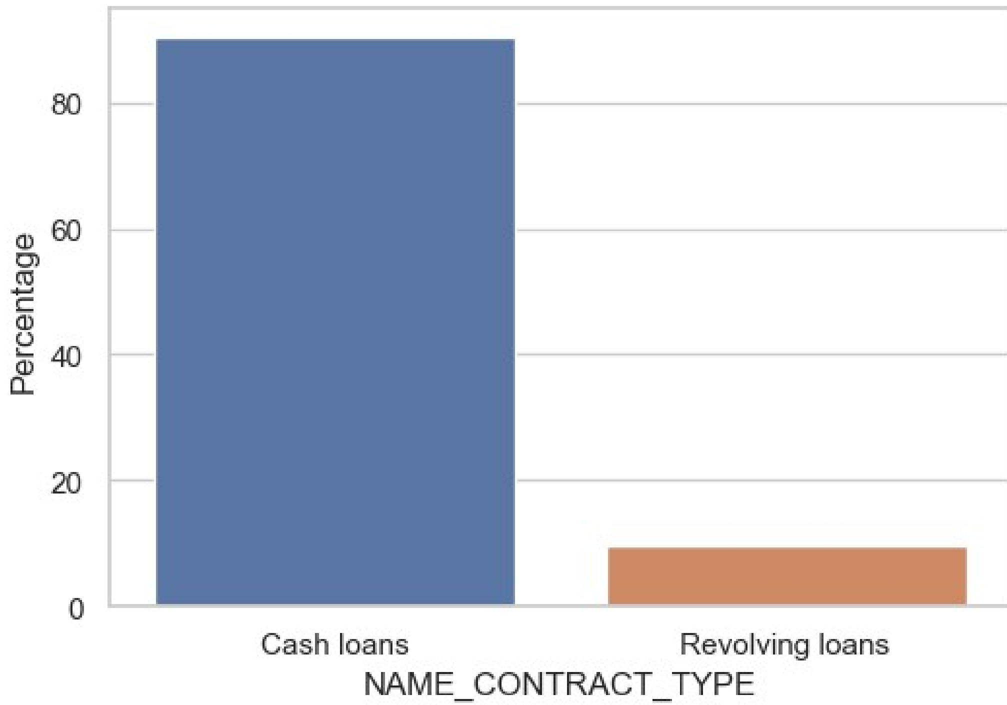
Top Correlations for Different Scenarios

```
class_frequencies =
application_data["NAME_CONTRACT_TYPE"].value_counts()

print(class_frequencies)
type(class_frequencies)
class_frequencies

Cash loans          45276
Revolving loans      4723
Name: NAME_CONTRACT_TYPE, dtype: int64

Cash loans          45276
Revolving loans      4723
Name: NAME_CONTRACT_TYPE, dtype: int64

for i in application_data.columns:
    # Calculate the class frequencies
    class_frequencies = application_data[i].value_counts()

    # Create a bar chart to visualize Data imbalance.
    # Using if condition to give only those we are categorical
variables.

    if len(class_frequencies)<=3:

        # printing the percentage of each value_count in list
        percentage =
np.around((class_frequencies.values)/len(application_data[i])*100)
        print(percentage)

        # Create a bar chart to visualize Data imbalance.
        plt.figure(figsize=(6, 4))
        sns.barplot(x = class_frequencies.index, y =
(class_frequencies.values)/len(application_data[i])*100)
        plt.xlabel(i)
        plt.ylabel("Percentage")
        plt.show()
```

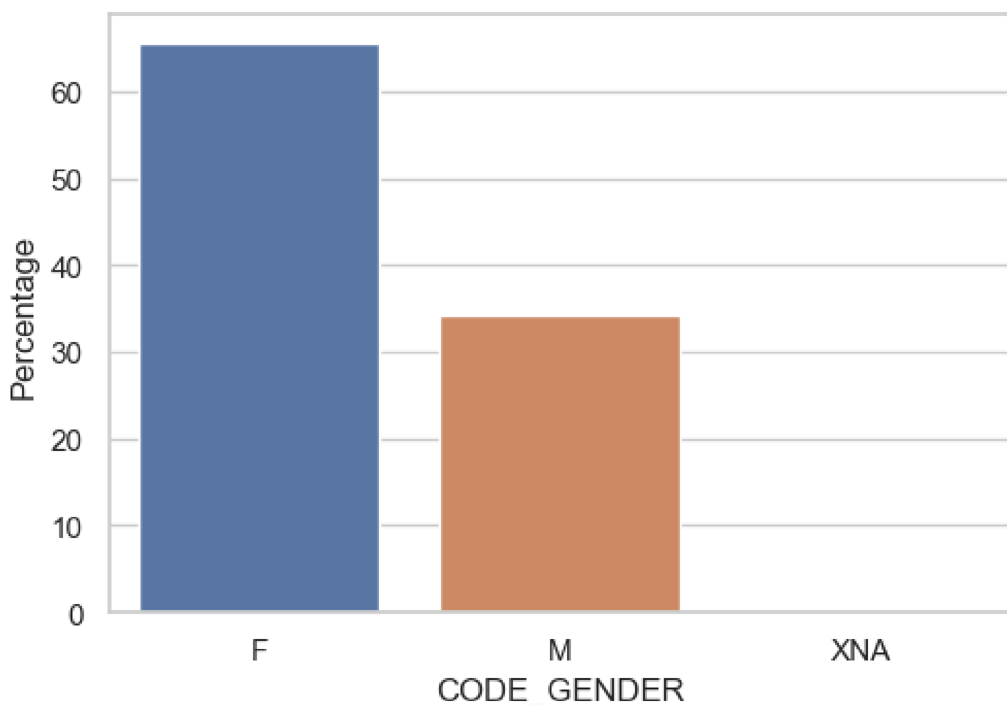
```
[92.  8.]
```



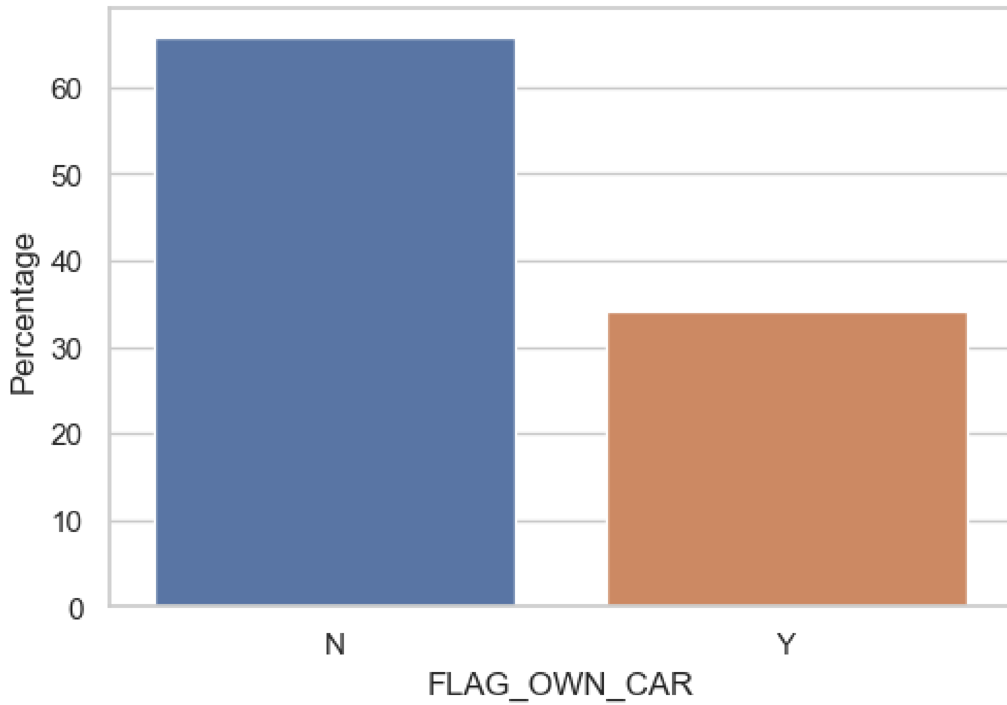

```
[91.  9.]
```


80
60
40
20
0
Percentage
Cash loans
Revolving loans
NAME_CONTRACT_TYPE

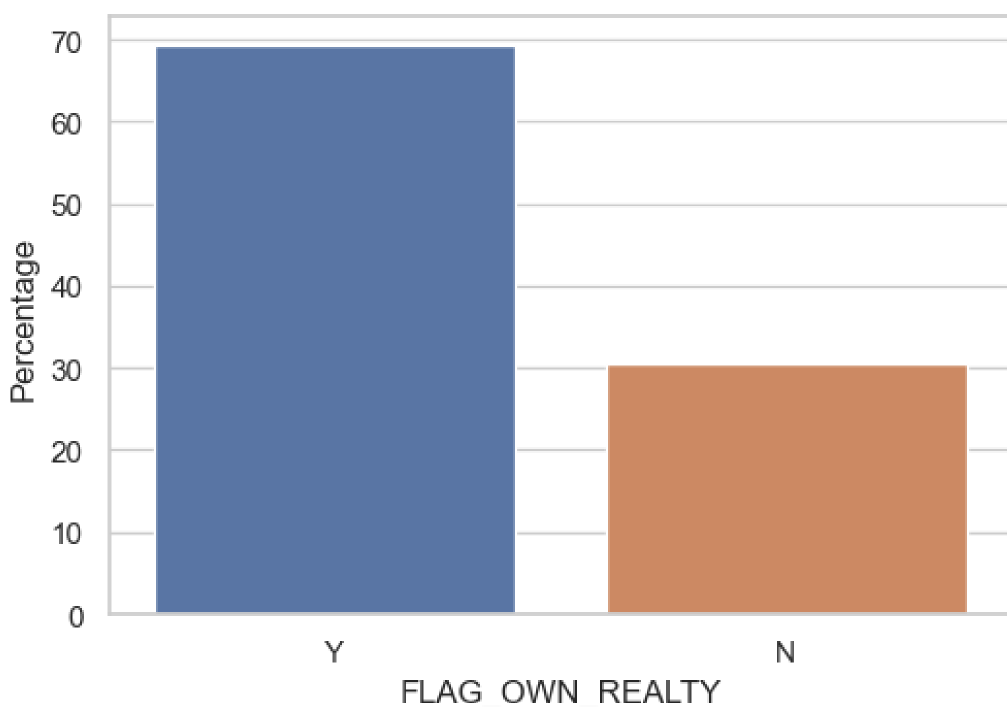
```
[66. 34.  0.]
```


60
50
40
30
20
10
0
Percentage
F
M
XNA
CODE_GENDER

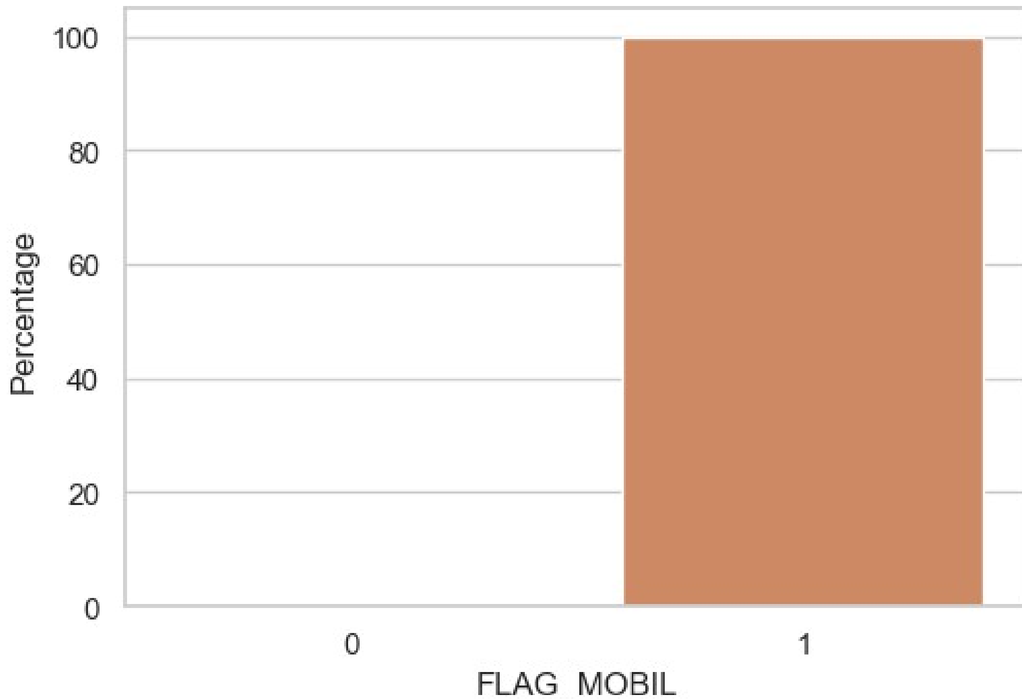
```
[66. 34.]
```


Percentage
60
50
40
30
20
10
0
N
Y
FLAG_OWN_CAR

```
[69. 31.]
```

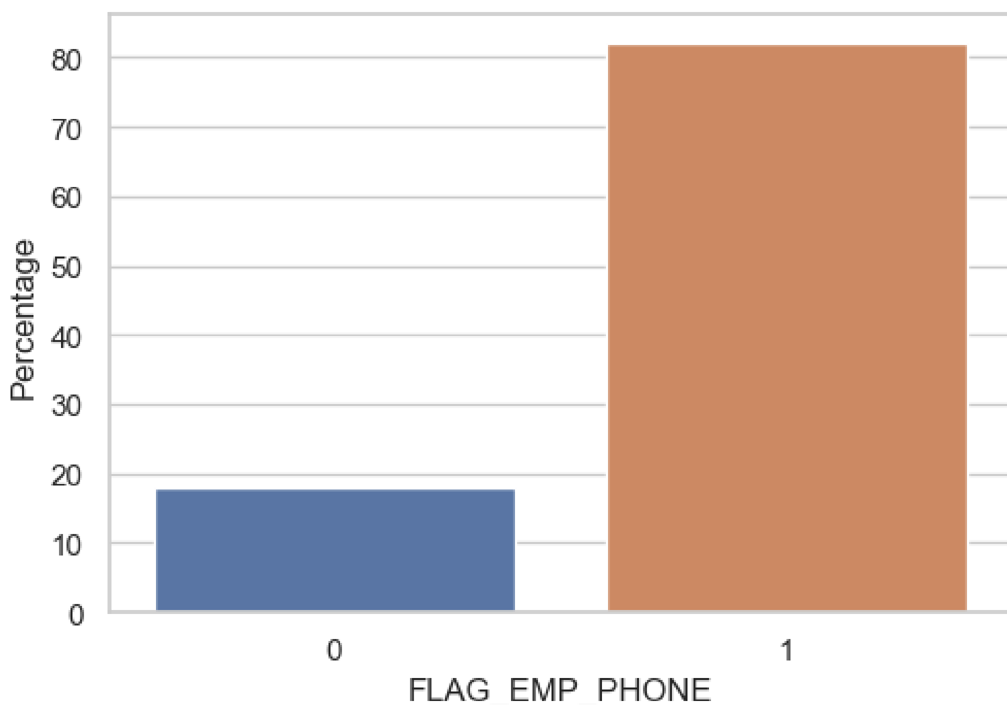

Percentage
70
60
50
40
30
20
10
0
Y
N
FLAG_OWN_REALTY

```
[100.   0.]
```



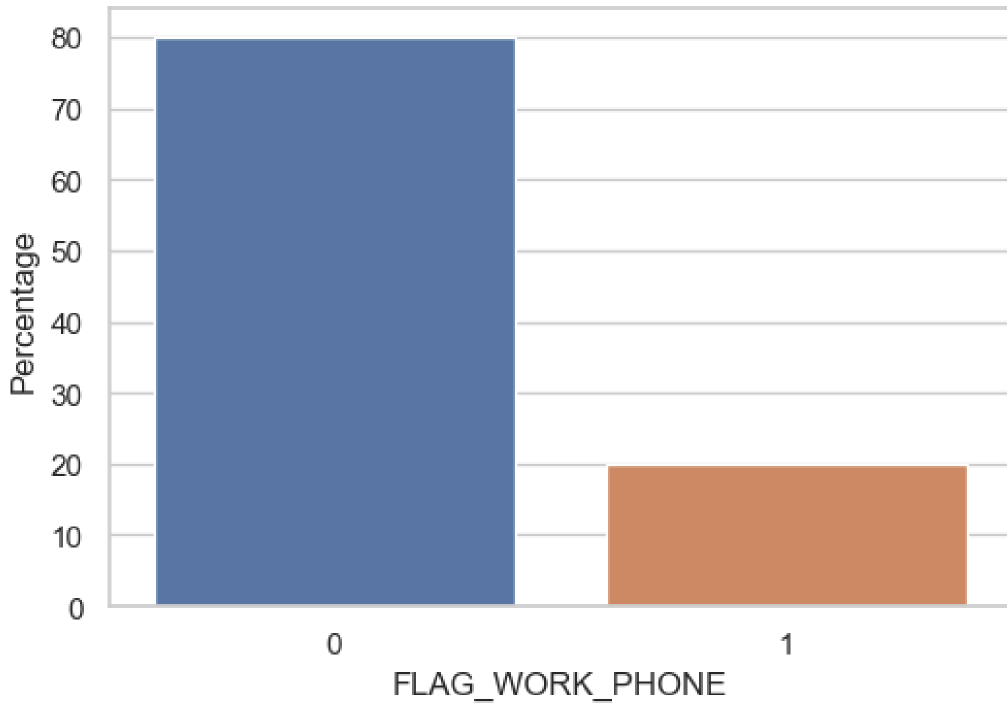
100
80
60
40
20
0
Percentage
0
1
FLAG_MOBIL


```
[82. 18.]
```



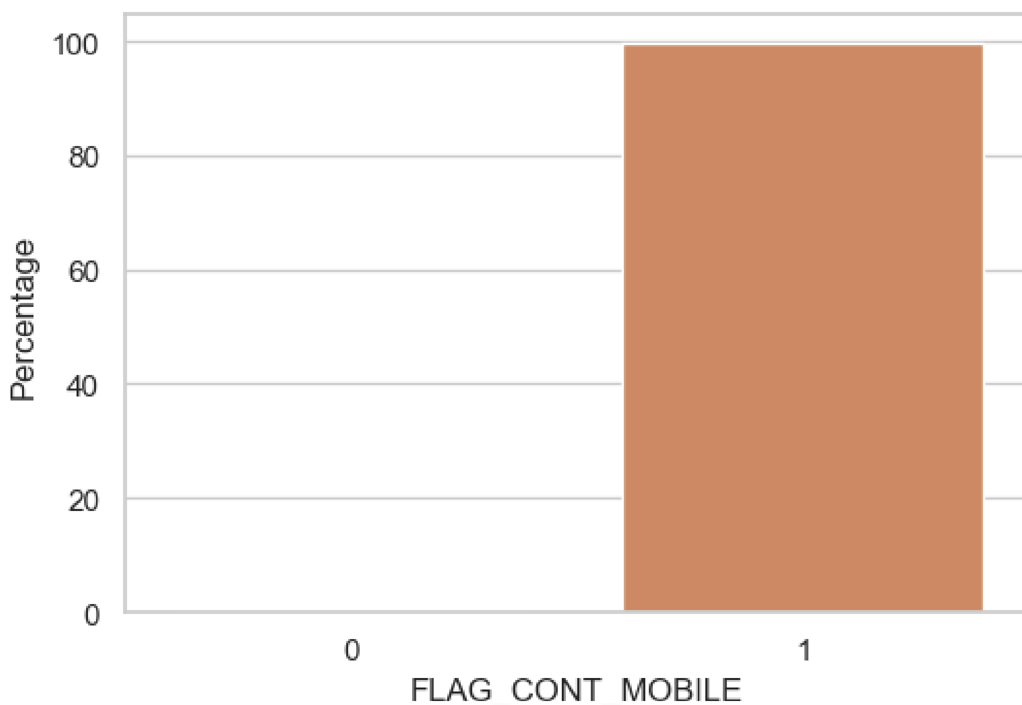
80
70
60
50
40
30
20
10
0
Percentage
0
1
FLAG_EMP_PHONE


```
[80. 20.]
```



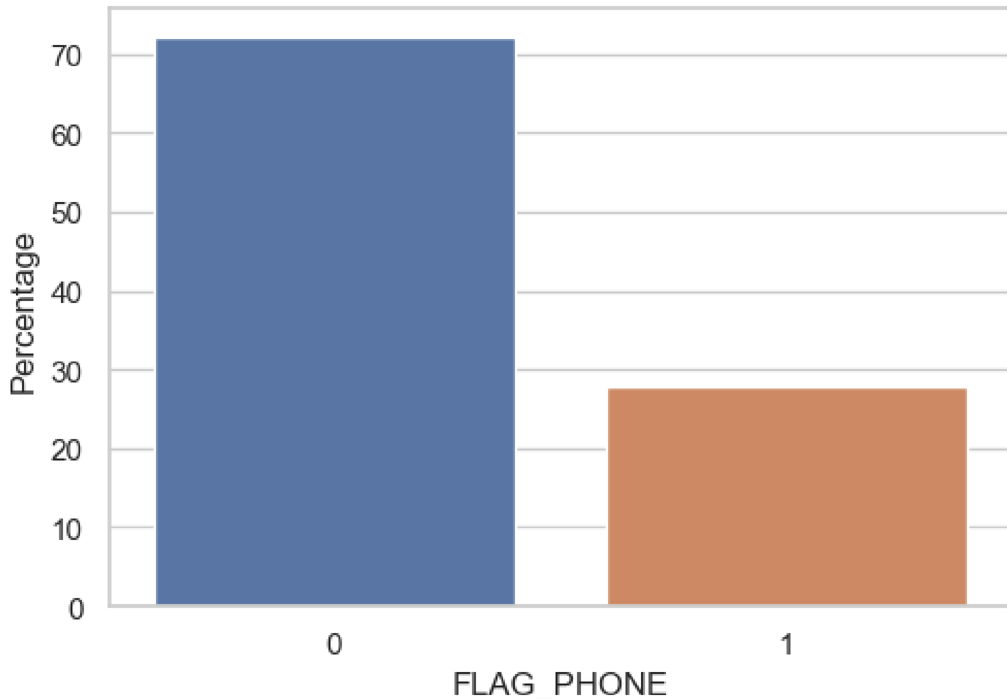
80
70
60
50
40
30
20
10
0
Percentage
0
1
FLAG_WORK_PHONE


```
[100.   0.]
```



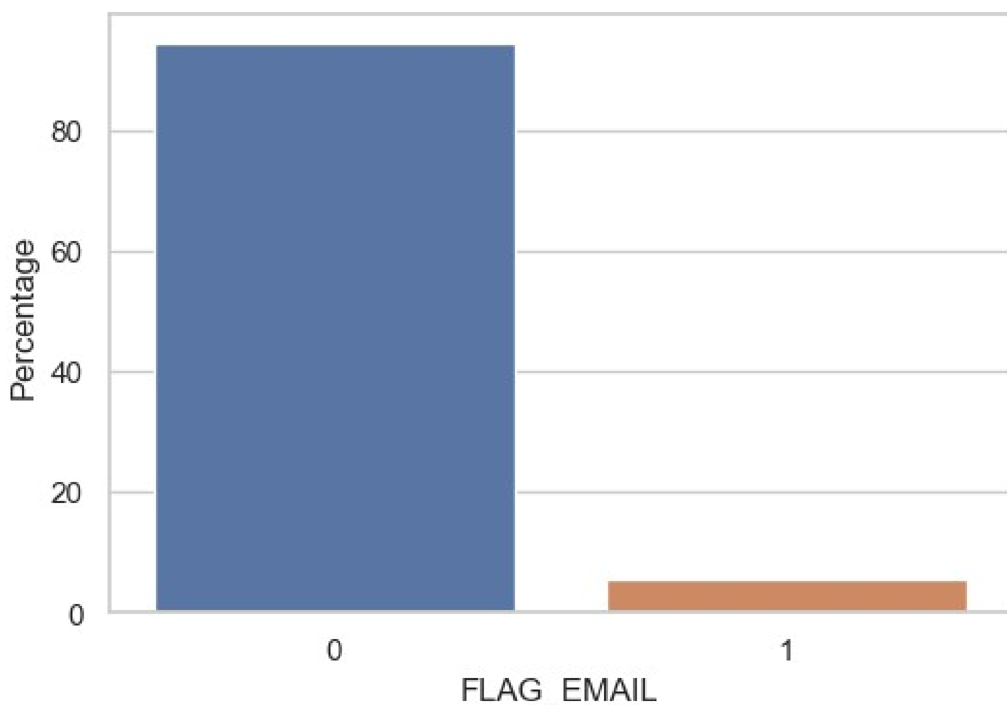
100
80
60
40
20
0
Percentage
0
1
FLAG_CONT_MOBILE


```
[72. 28.]
```



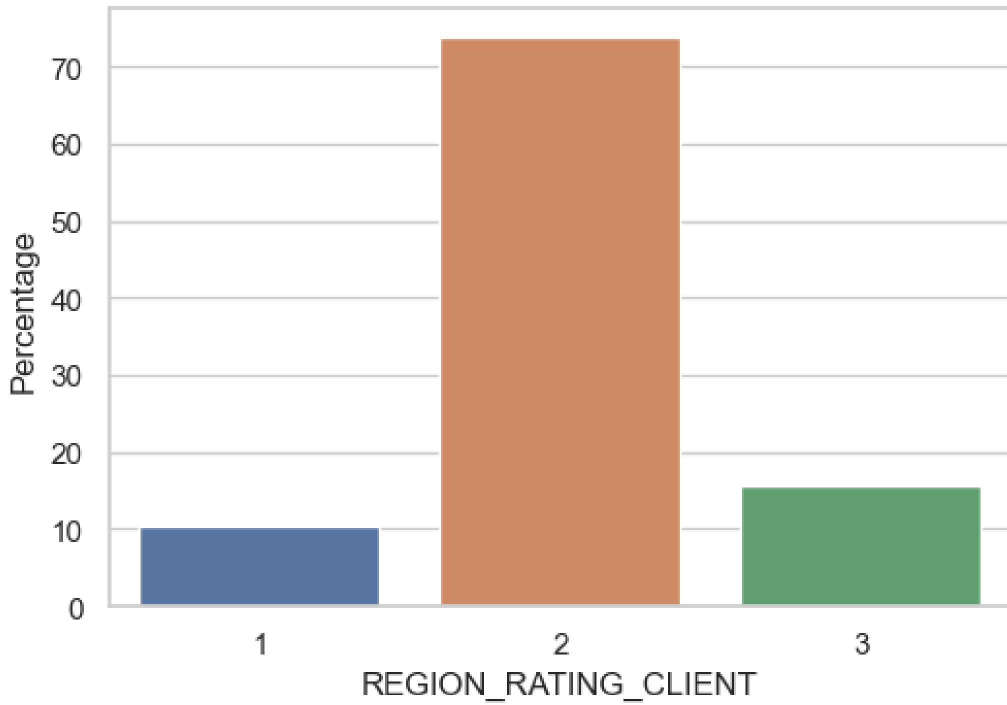
70
60
50
40
30
20
10
0
Percentage
0
1
FLAG_PHONE


```
[94.  6.]
```



80
60
40
20
0
Percentage
0
1
FLAG_EMAIL


```
[74. 16. 10.]
```

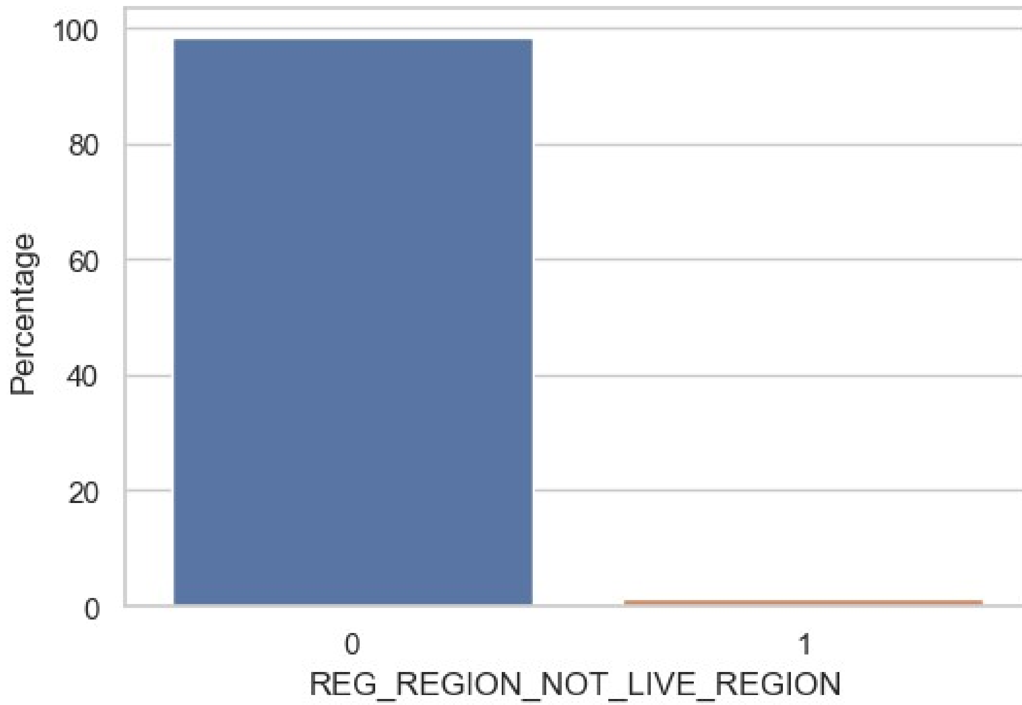


Percentage
0
10
20
30
40
50
60
70
1
2
3
REGION_RATING_CLIENT


```
[75. 14. 11.]
```

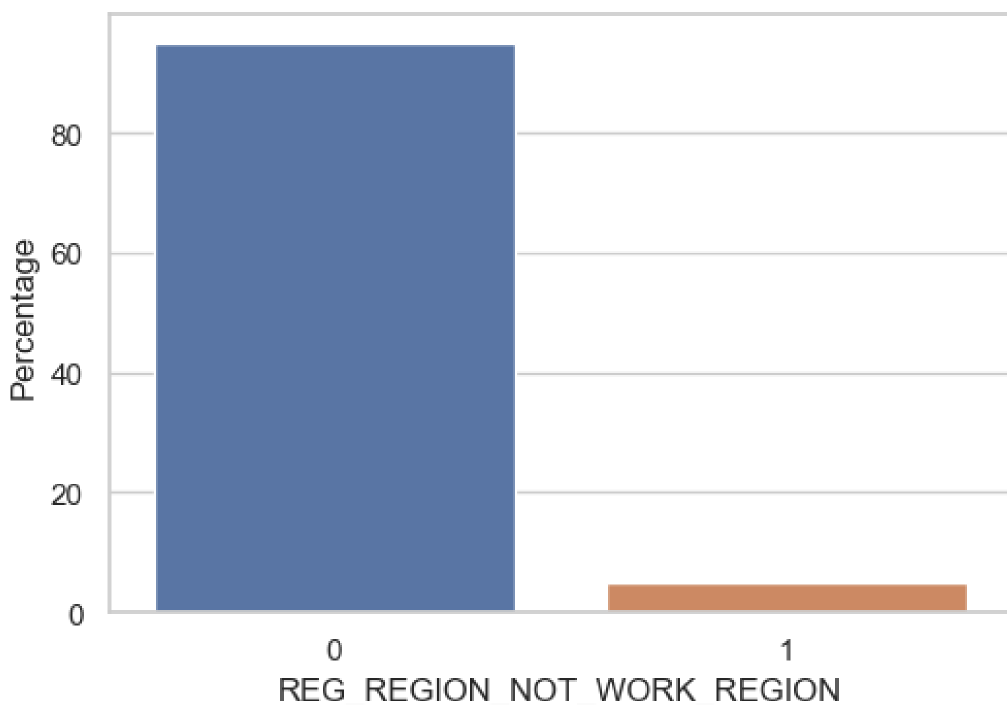
Percentage
0
10
20
30
40
50
60
70
1
2
3
REGION_RATING_CLIENT_W_CITY


```
[98.  2.]
```



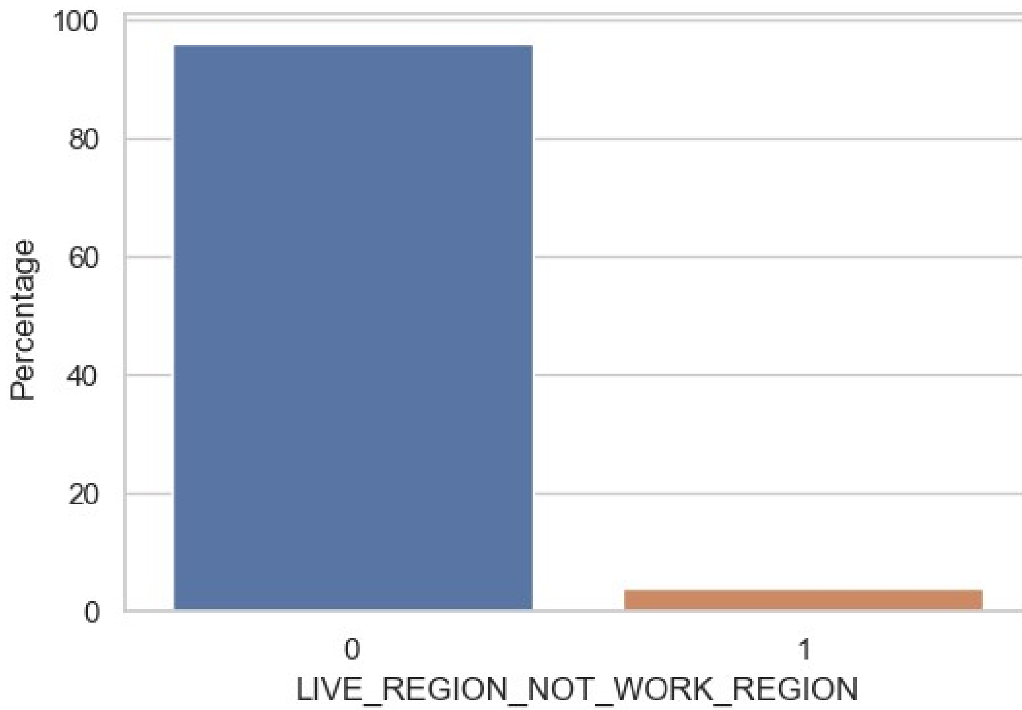
100
80
60
40
20
0
Percentage
0
1
REG_REGION_NOT_LIVE_REGION


```
[95.  5.]
```

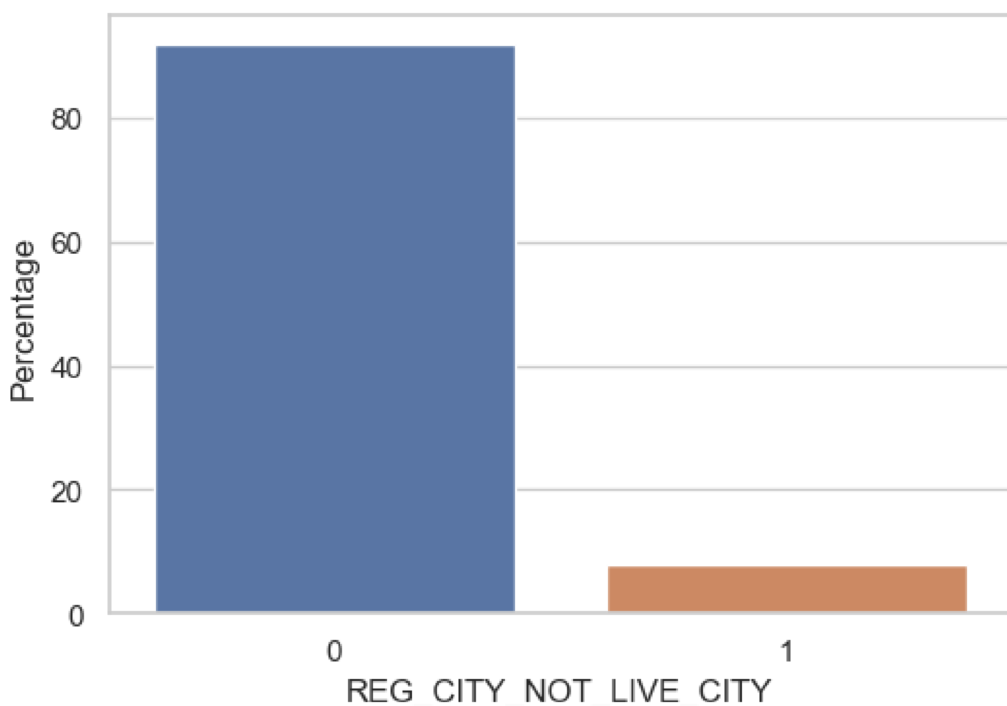


80
60
40
20
0
Percentage
0
1
REG_REGION_NOT_WORK_REGION

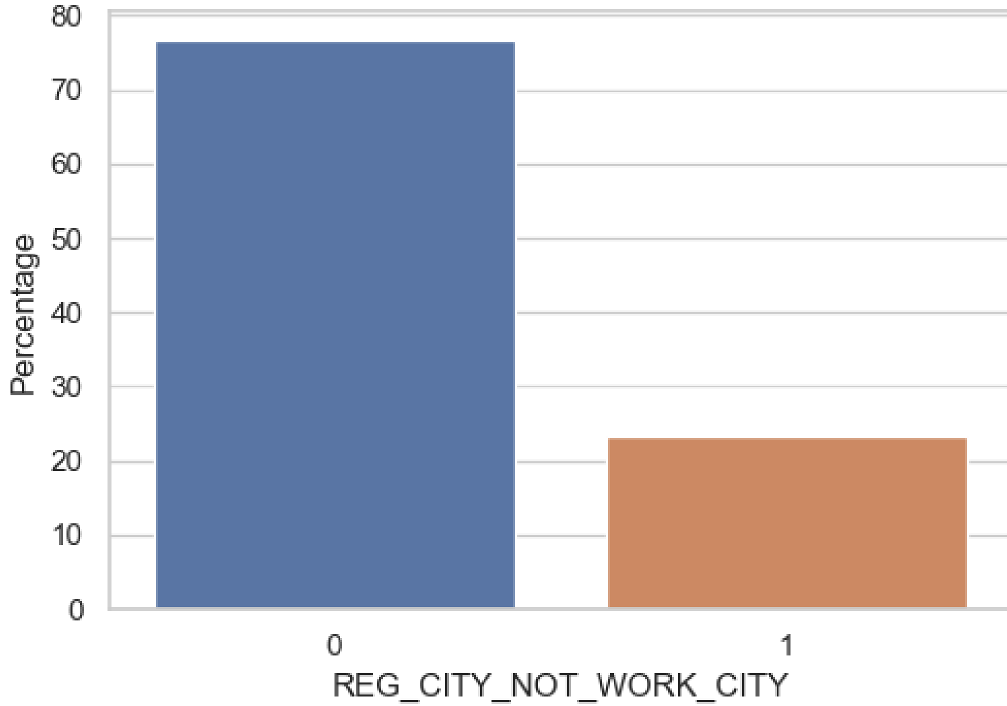

```
[96.  4.]
```


100
80
60
40
20
0
Percentage
0
1
LIVE_REGION_NOT_WORK_REGION

```
[92.  8.]
```


80
60
40
20
0
Percentage
0
1
REG_CITY_NOT_LIVE_CITY

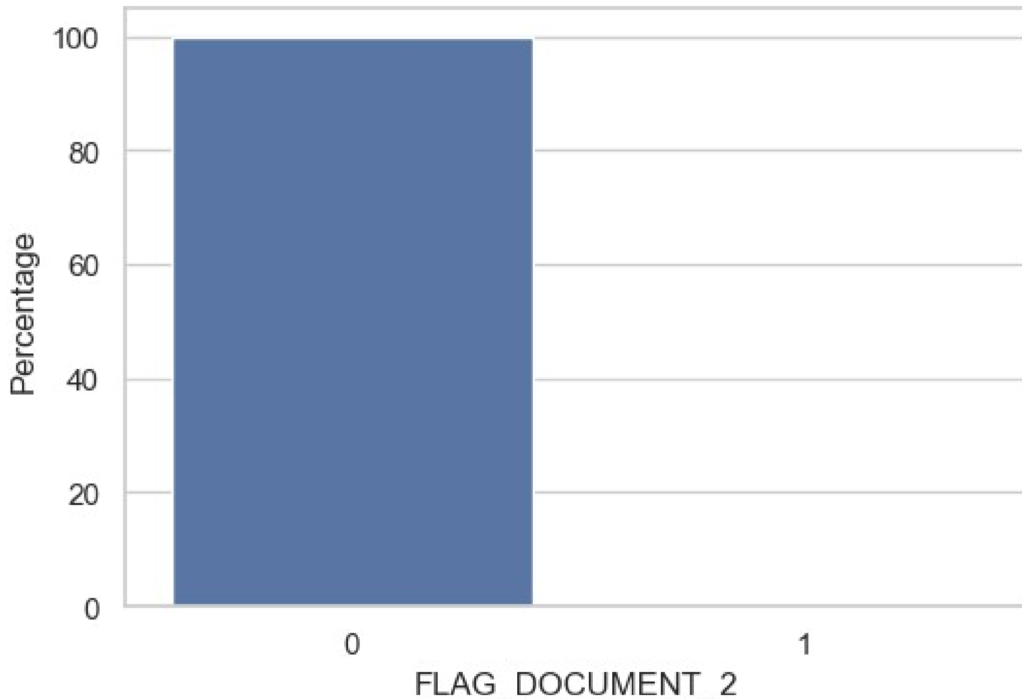
```
[77. 23.]
```


Percentage
80
70
60
50
40
30
20
10
0
0
1
REG_CITY_NOT_WORK_CITY

```
[82. 18.]
```

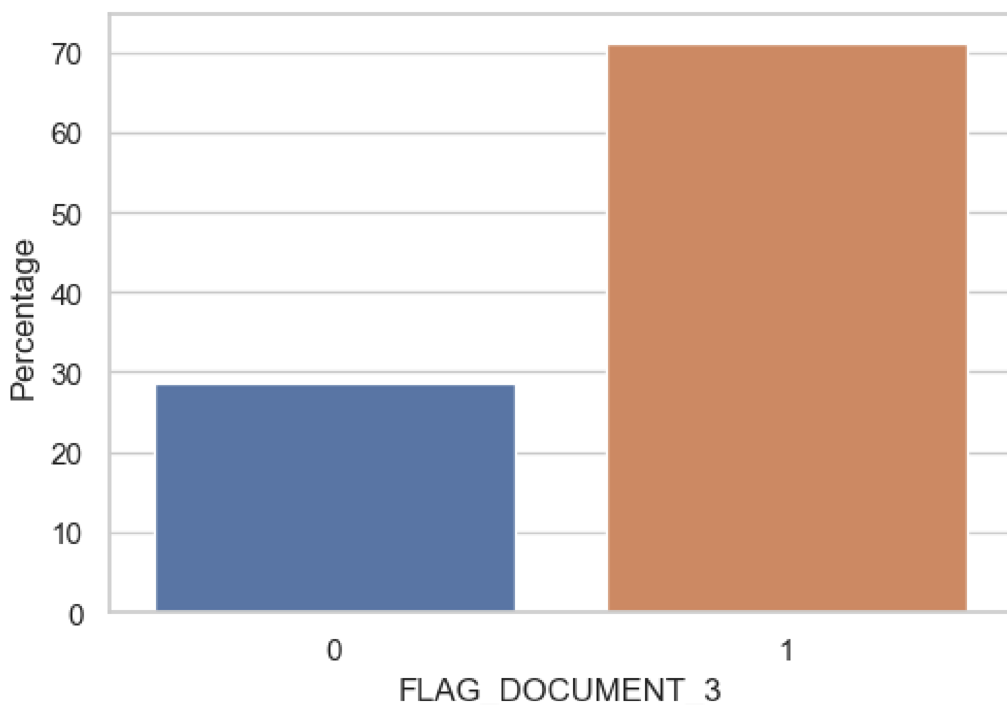

Percentage
80
70
60
50
40
30
20
10
0
0
1
LIVE_CITY_NOT_WORK_CITY

```
[100.   0.]
```



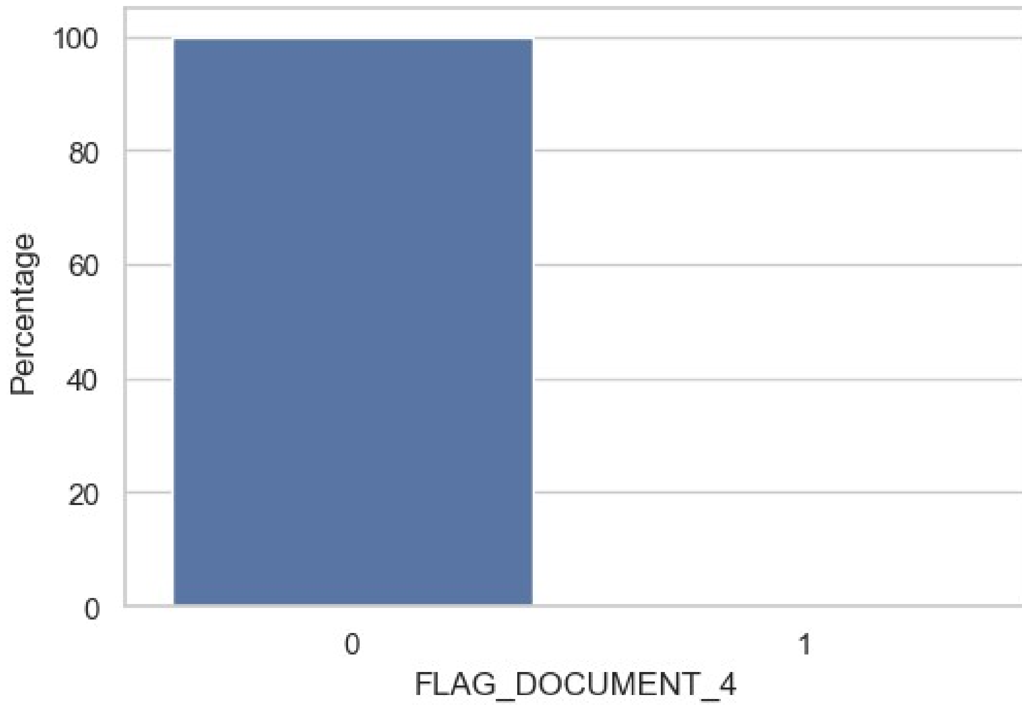
100
80
60
40
20
0
Percentage
0
1
FLAG_DOCUMENT_2


```
[71. 29.]
```

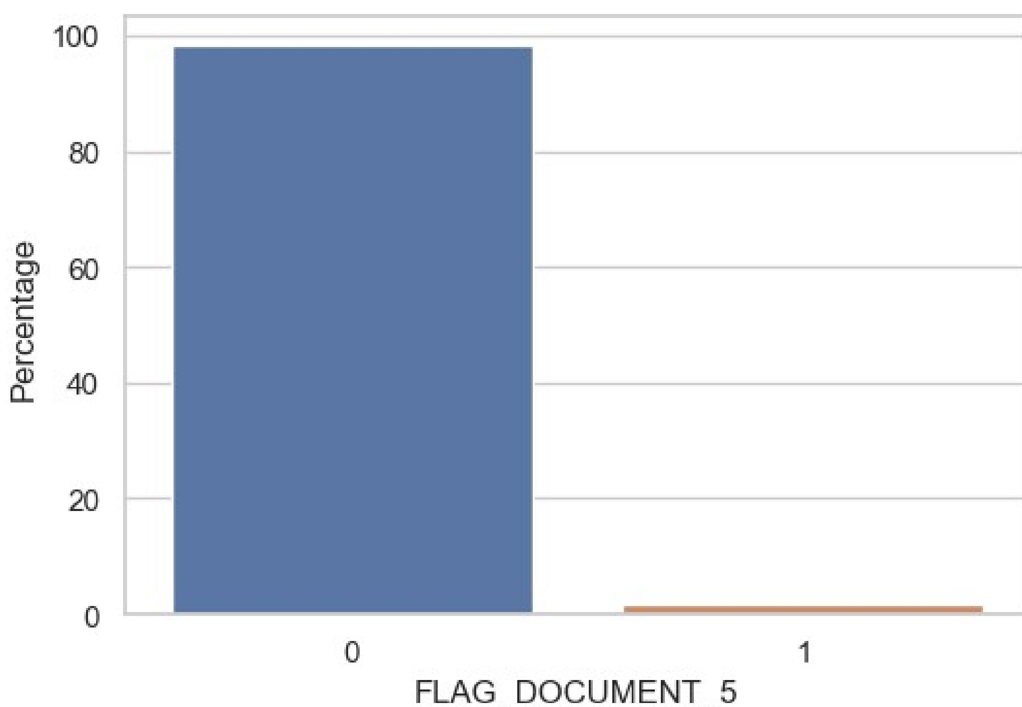


70
60
50
40
30
20
10
0
Percentage
0
1
FLAG_DOCUMENT_3

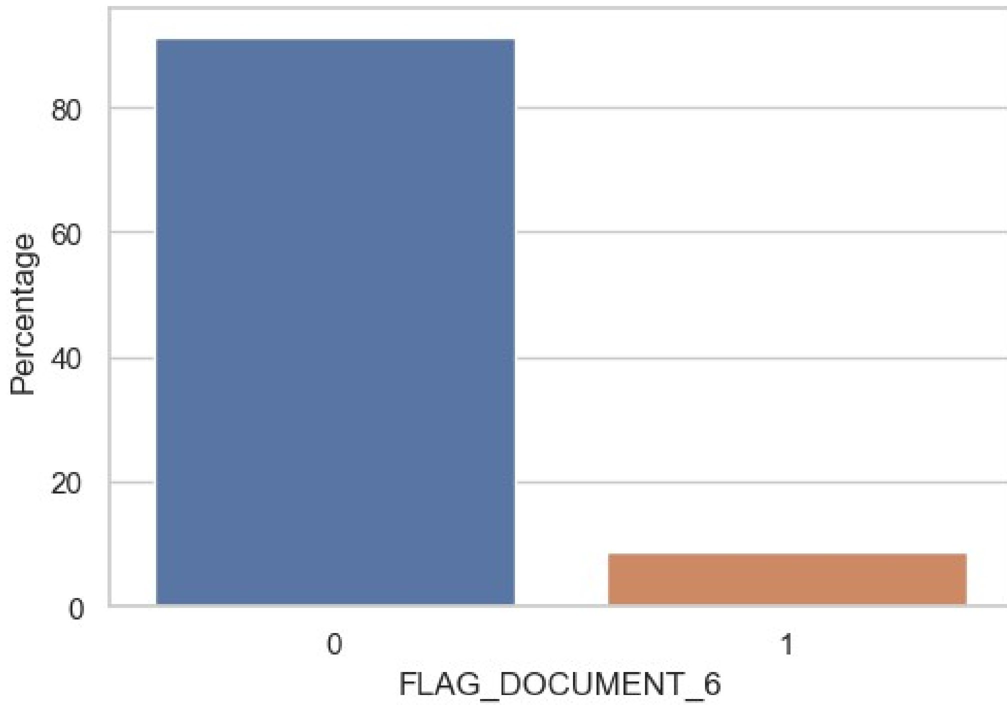

```
[100.   0.]
```


100
80
60
40
20
0
Percentage
0
1
FLAG_DOCUMENT_4

```
[98.  2.]
```

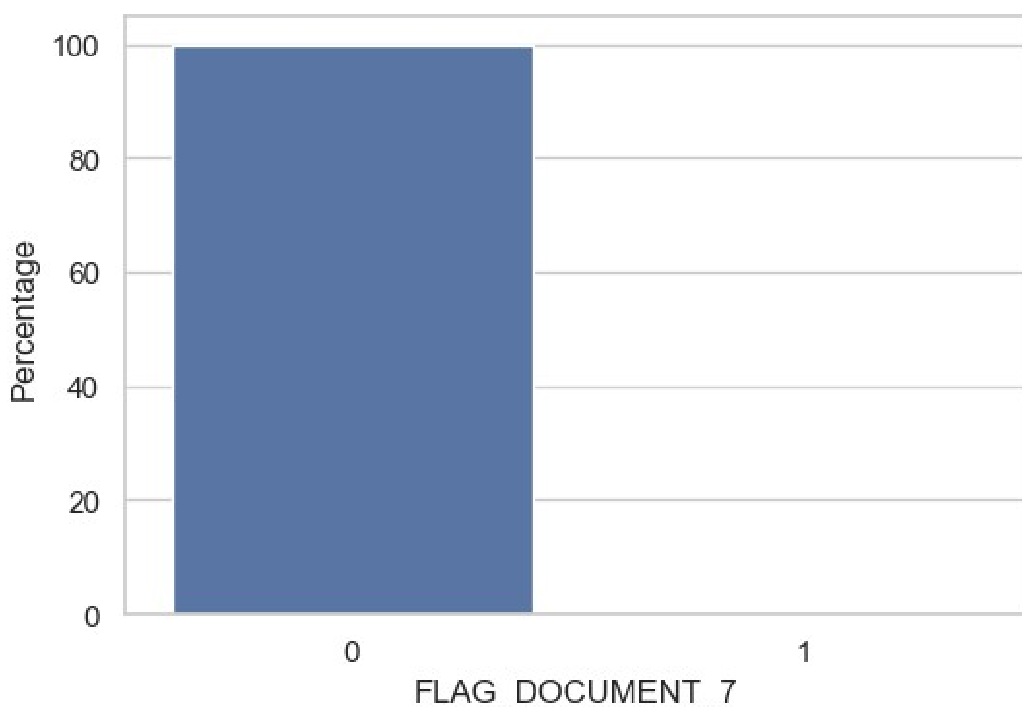

100
80
60
40
20
0
Percentage
0
1
FLAG_DOCUMENT_5

```
[91.  9.]
```



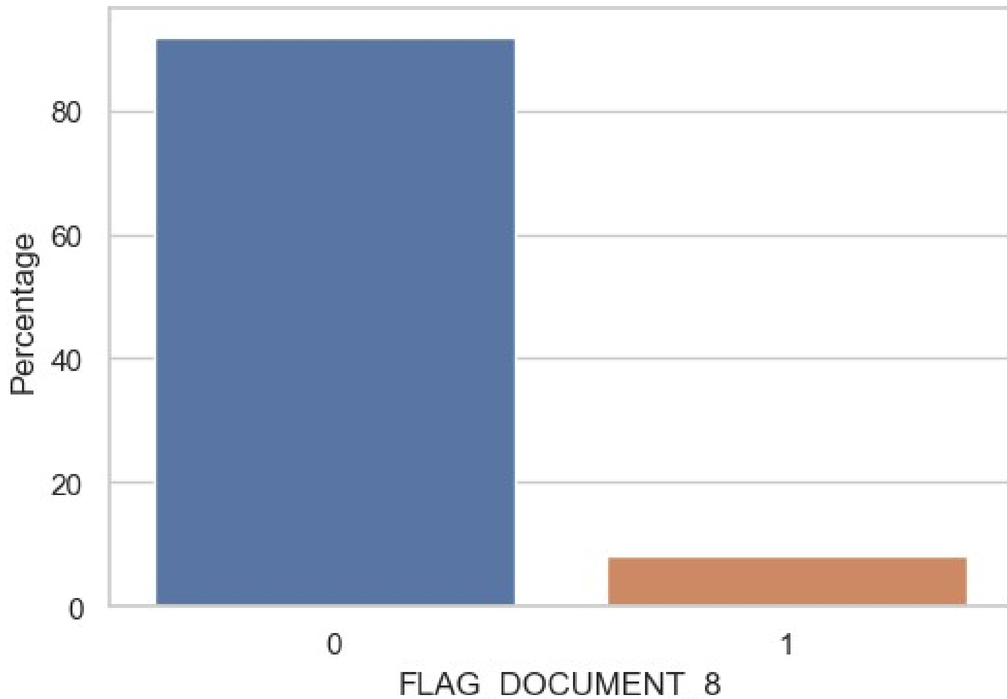
80
60
40
20
0
Percentage
0
1
FLAG_DOCUMENT_6


```
[100.   0.]
```

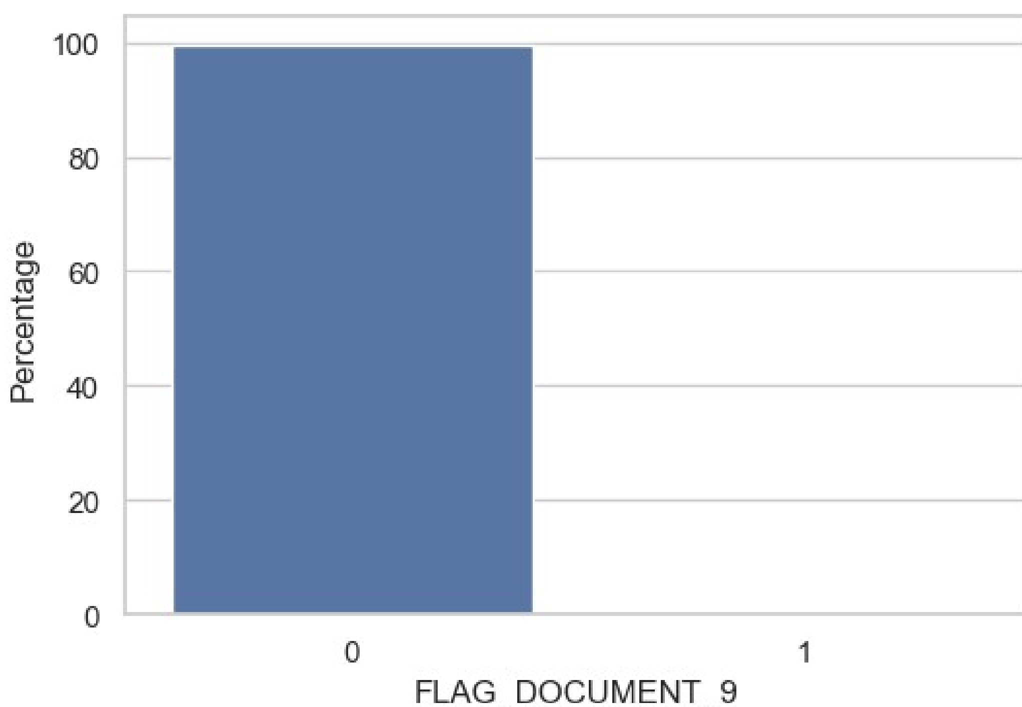


100
80
60
40
20
0
Percentage
0
1
FLAG_DOCUMENT_7

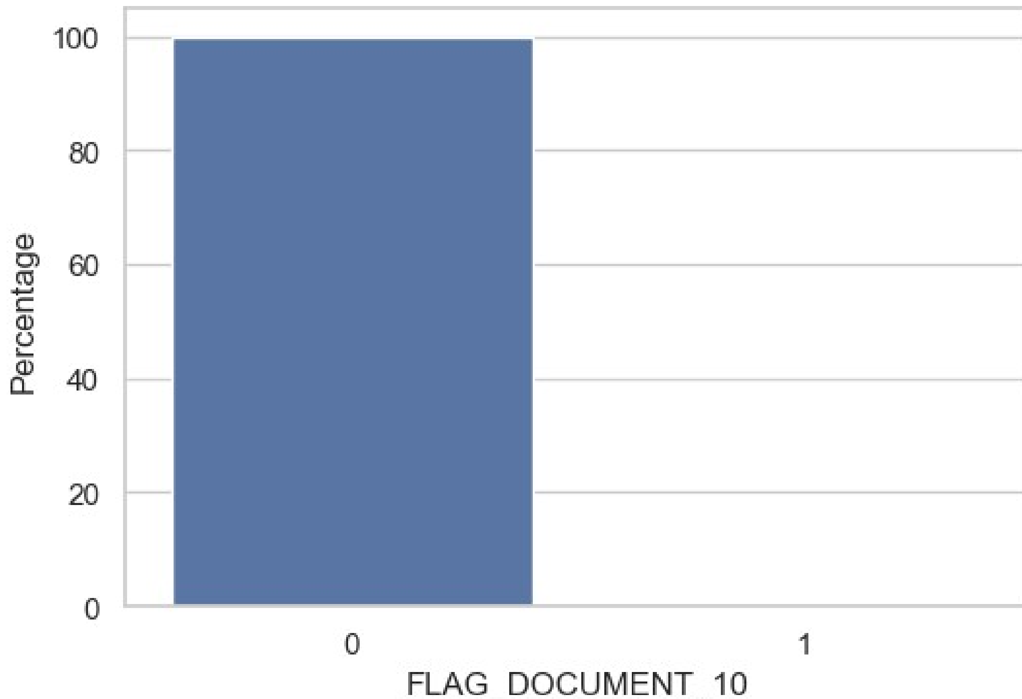

```
[92.  8.]
```


80
60
40
20
0
Percentage
0
1
FLAG_DOCUMENT_8

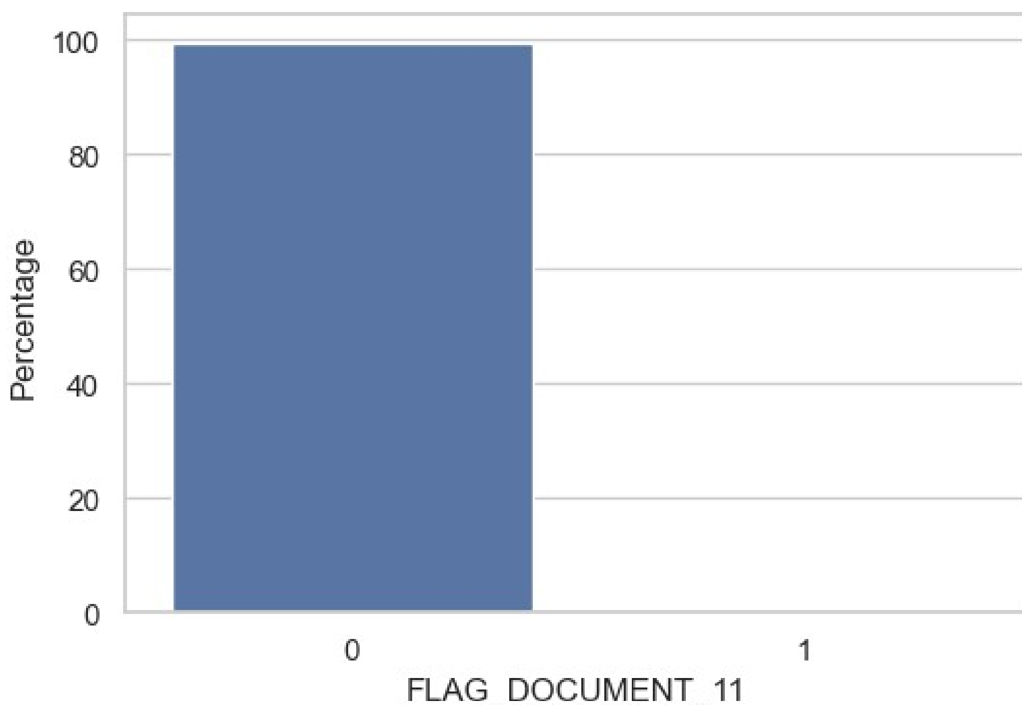
```
[100.   0.]
```


100
80
60
40
20
0
Percentage
0
1
FLAG_DOCUMENT_9

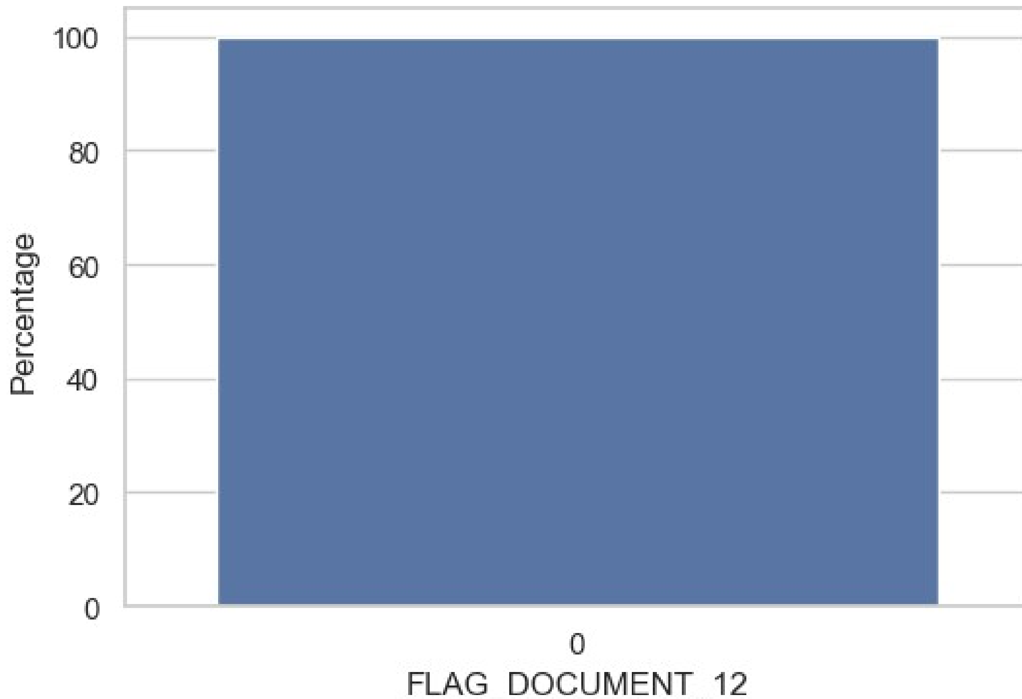
```
[100.   0.]
```


100
80
60
40
20
0
Percentage
0
1
FLAG_DOCUMENT_10

```
[100.   0.]
```

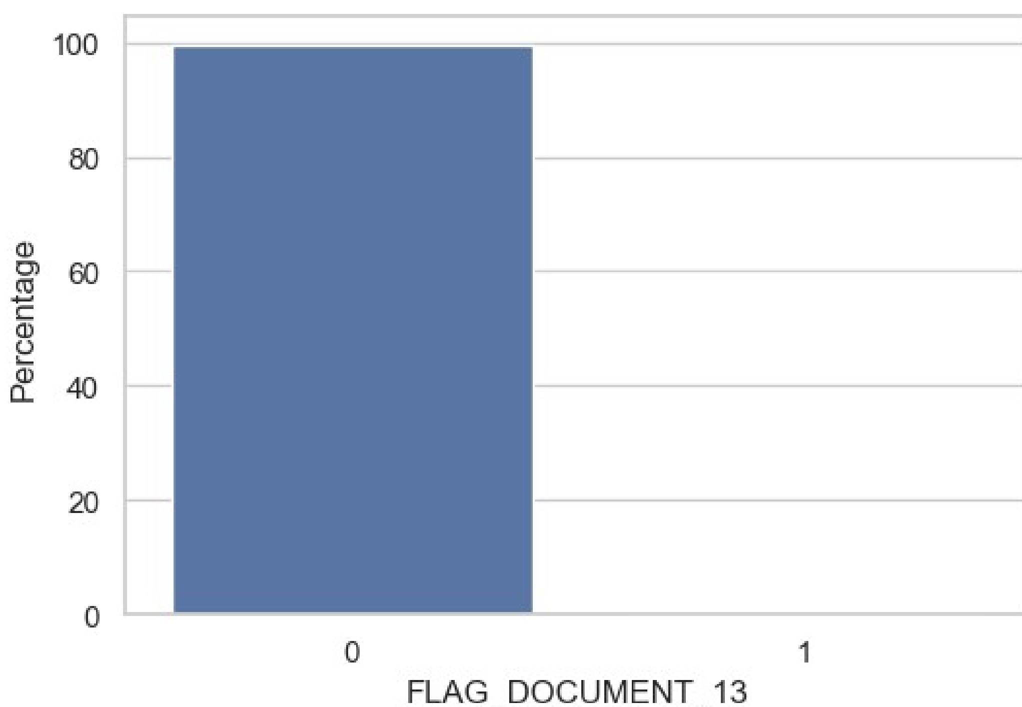

100
80
60
40
20
0
Percentage
0
1
FLAG_DOCUMENT_11

```
[100.]
```



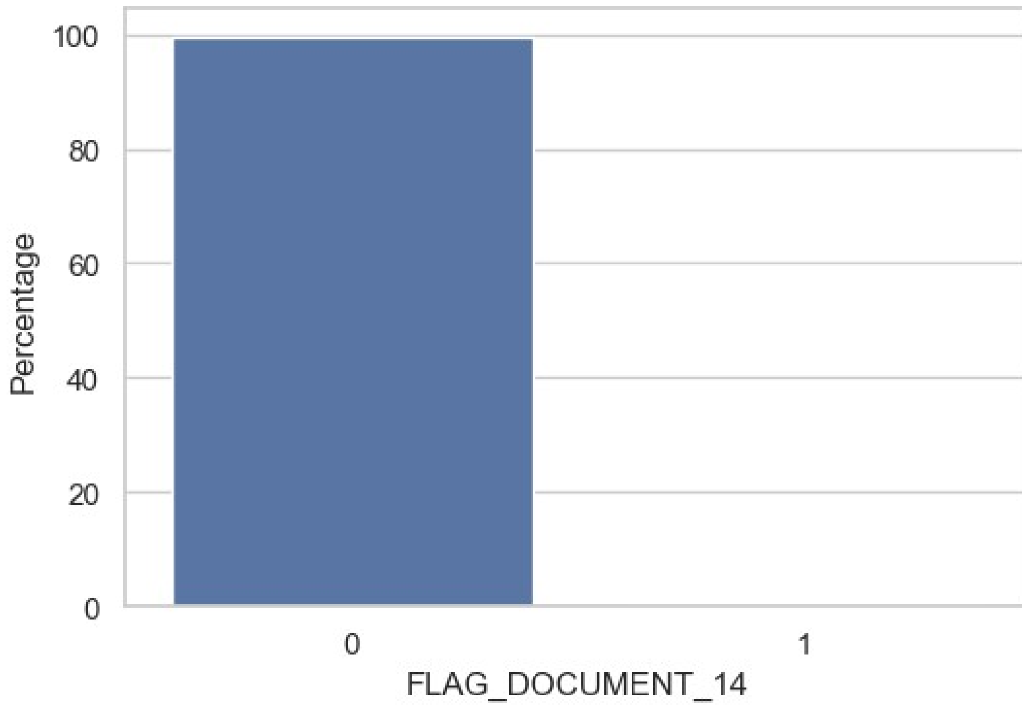
100
80
60
40
20
0
Percentage
0
FLAG_DOCUMENT_12


```
[100.   0.]
```

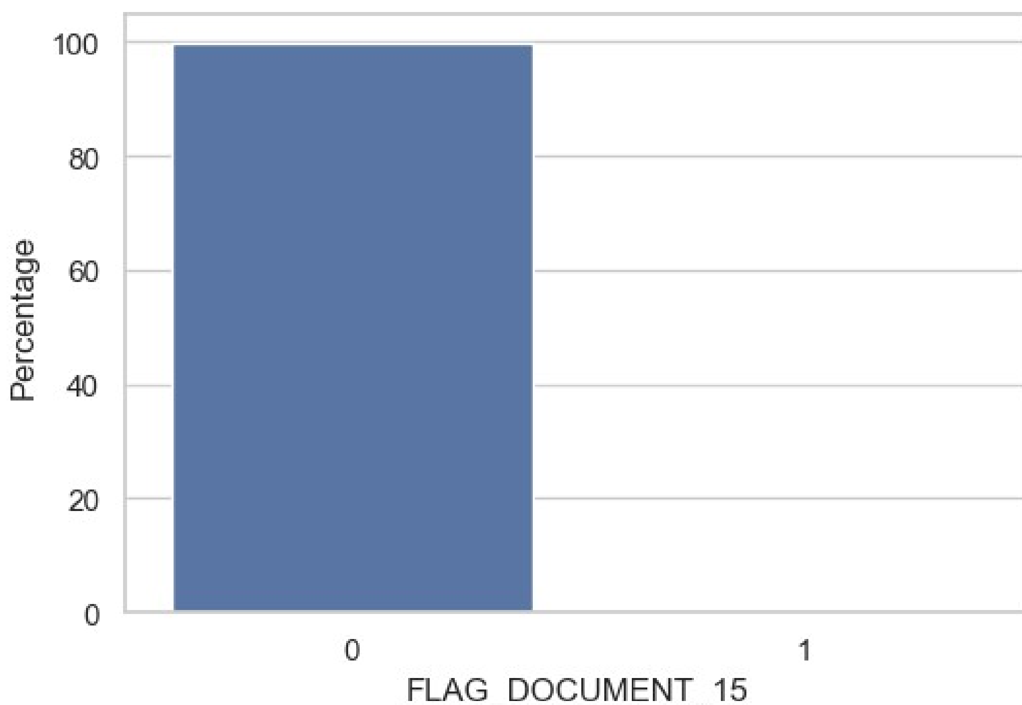


100
80
60
40
20
0
Percentage
0
1
FLAG_DOCUMENT_13

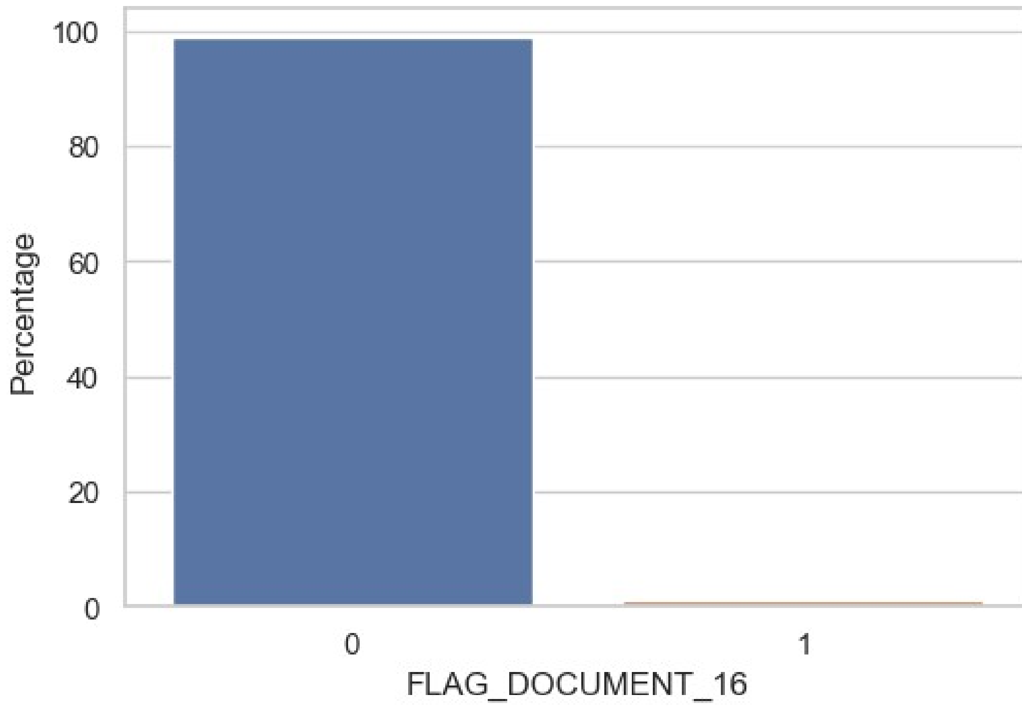

```
[100.   0.]
```


100
80
60
40
20
0
Percentage
0
1
FLAG_DOCUMENT_14

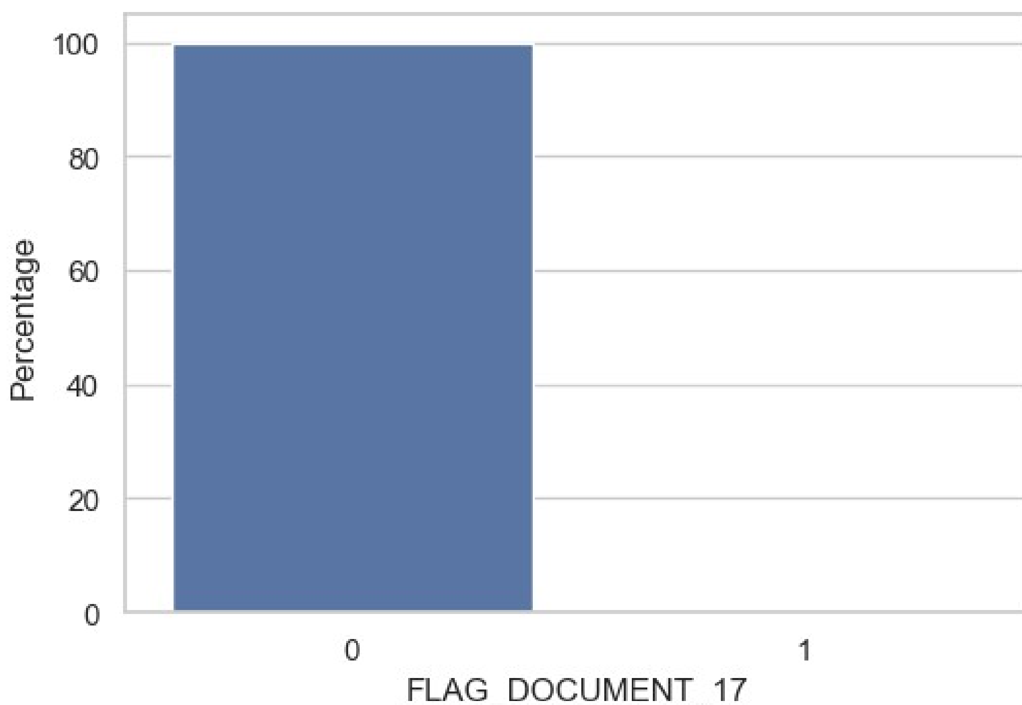
```
[100.   0.]
```


100
80
60
40
20
0
Percentage
0
1
FLAG_DOCUMENT_15

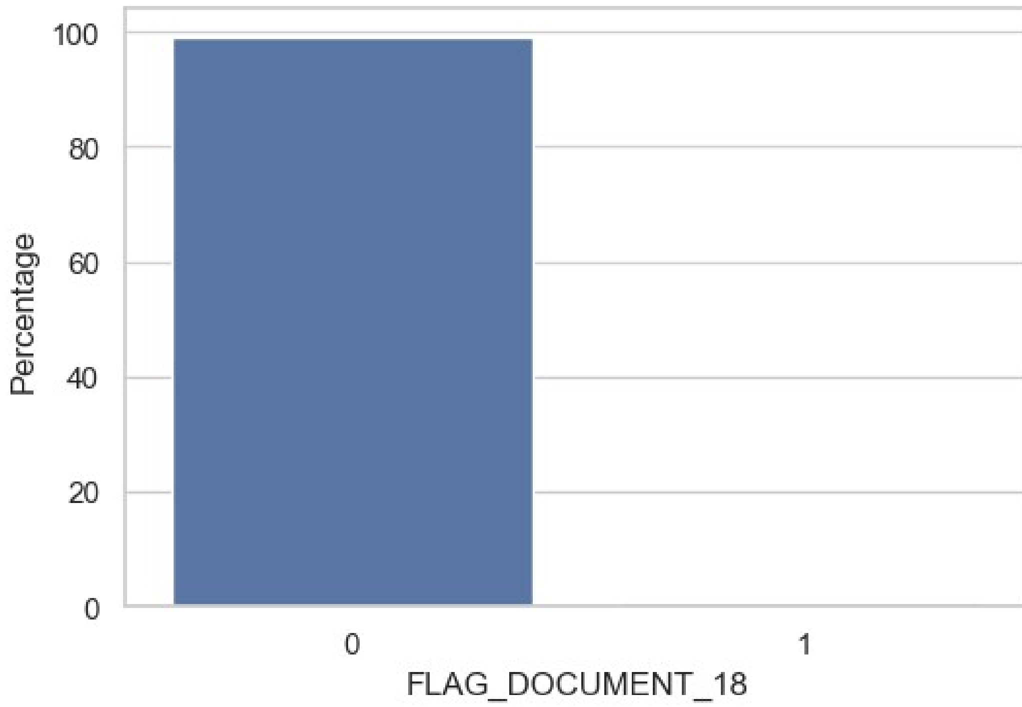
```
[99.  1.]
```


100
80
60
40
20
0
Percentage
0
1
FLAG_DOCUMENT_16

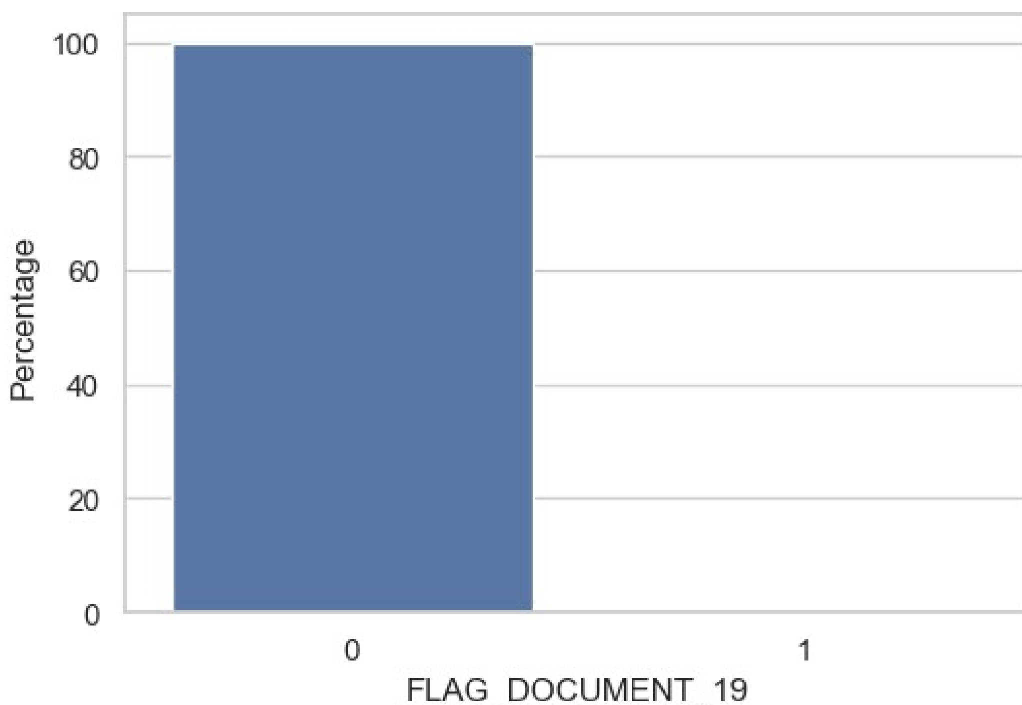
```
[100.   0.]
```


100
80
60
40
20
0
Percentage
0
1
FLAG_DOCUMENT_17

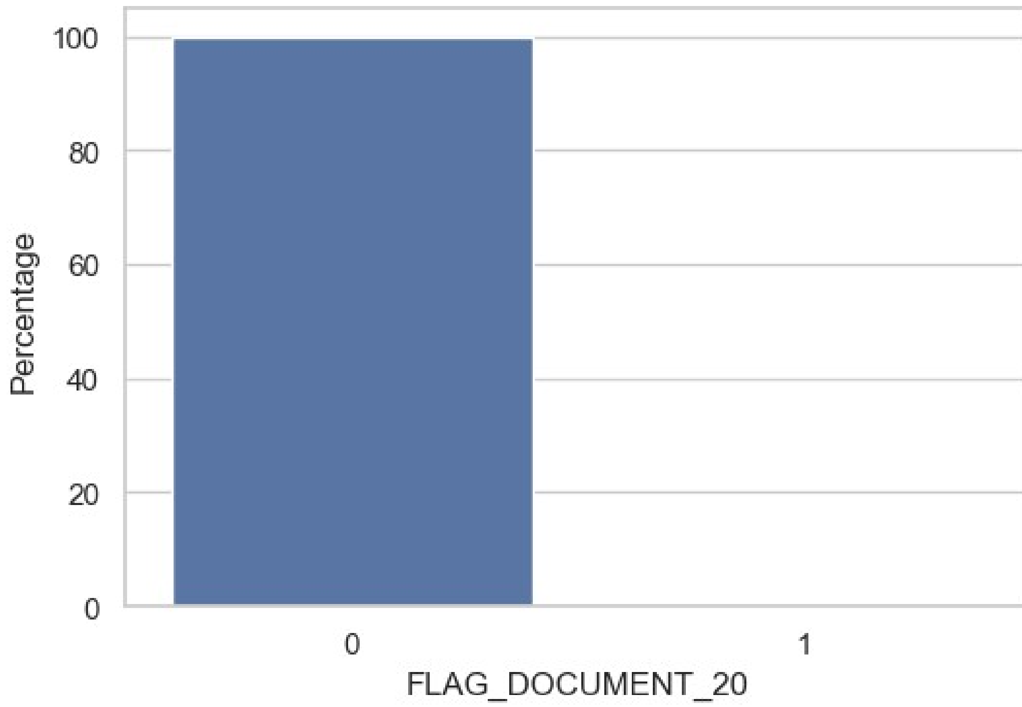
```
[99.  1.]
```


100
80
60
40
20
0
Percentage
0
1
FLAG_DOCUMENT_18

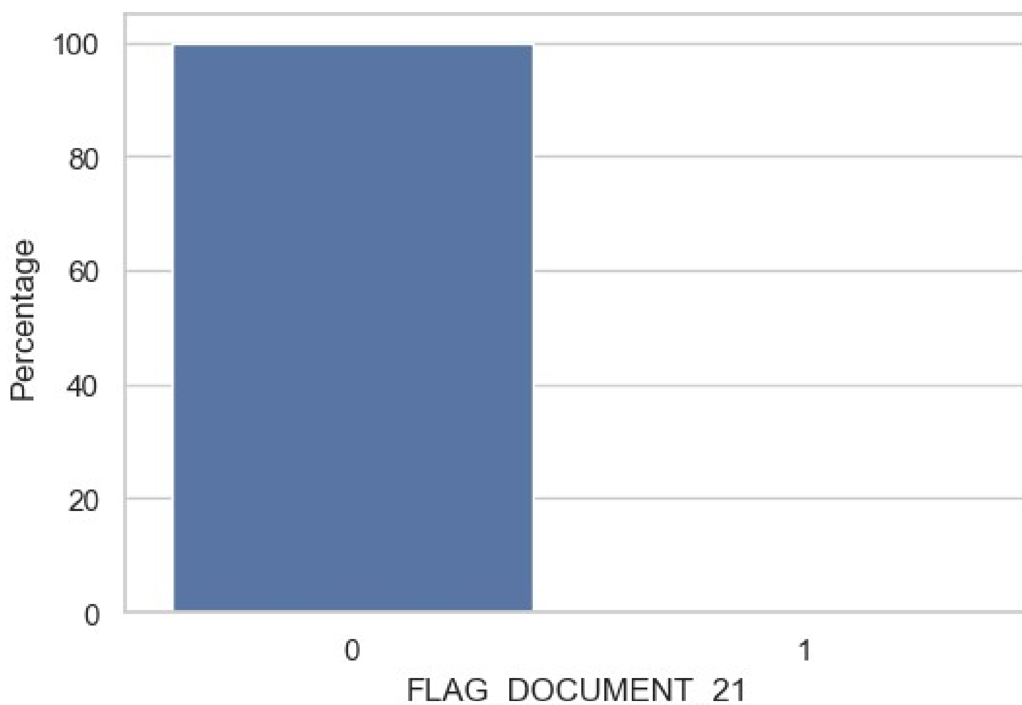
```
[100.   0.]
```


100
80
60
40
20
0
Percentage
0
1
FLAG_DOCUMENT_19

```
[100.   0.]
```



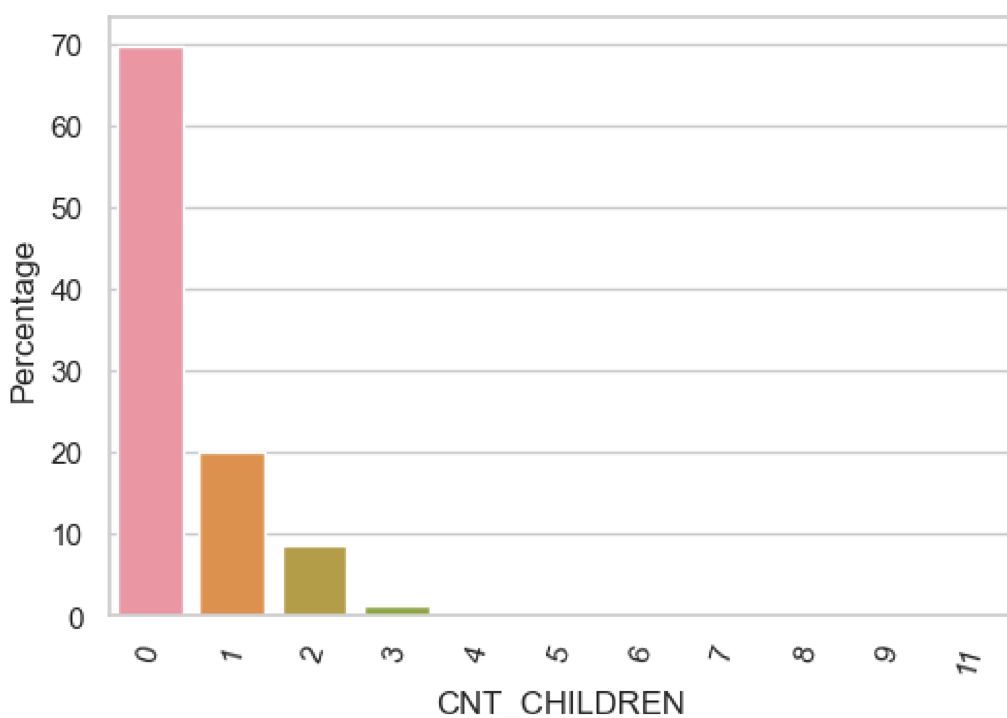

```
[100.   0.]
```



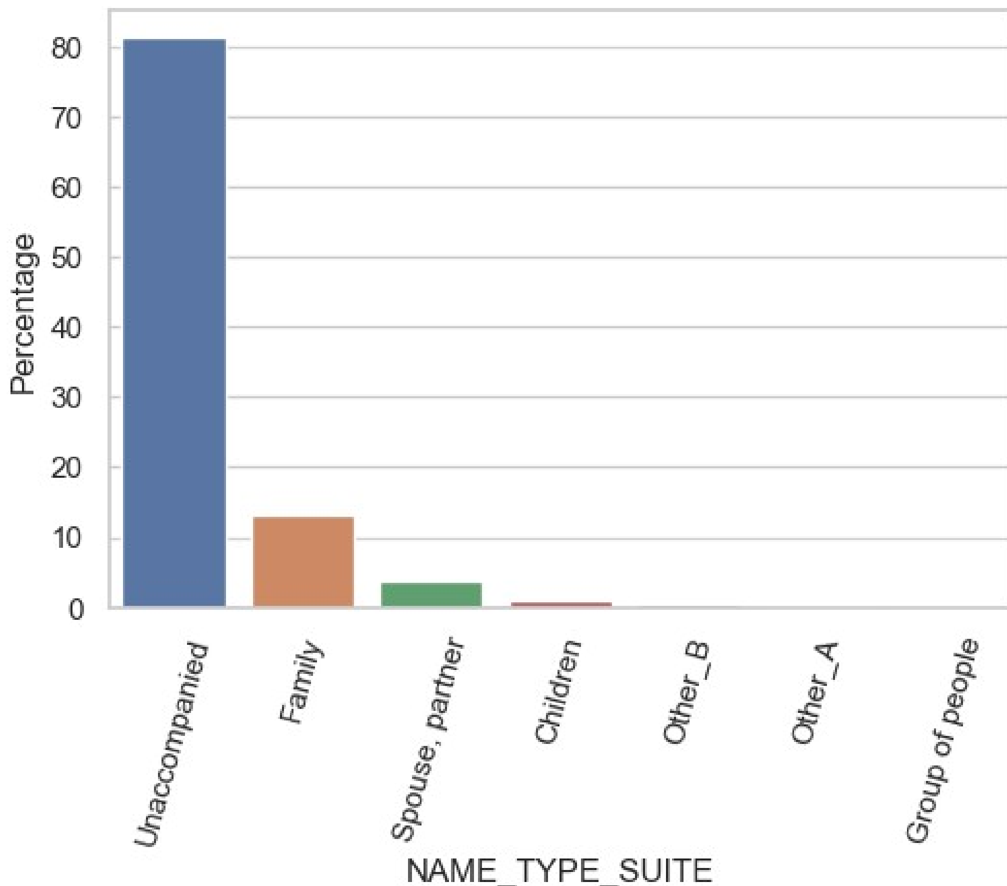

```
for i in application_data.columns:
    # Calculate the class frequencies
    class_frequencies = application_data[i].value_counts()
```

```
    # Create a bar chart to visualize Data imbalance.
    # Using if condition to give only those we are categorical
variables.

    if len(class_frequencies)<=15 and len(class_frequencies)>3:

        # printing the percentage of each value_count in list
        percentage =
np.around((class_frequencies.values)/len(application_data[i])*100)
        print(percentage)

        # Create a bar chart to visualize Data imbalance.
        plt.figure(figsize=(6, 4))
        sns.barplot(x = class_frequencies.index, y =
(class_frequencies.values)/len(application_data[i])*100)
        plt.xlabel(i)
        plt.ylabel("Percentage")
        plt.xticks(rotation=75)
        plt.show()
```

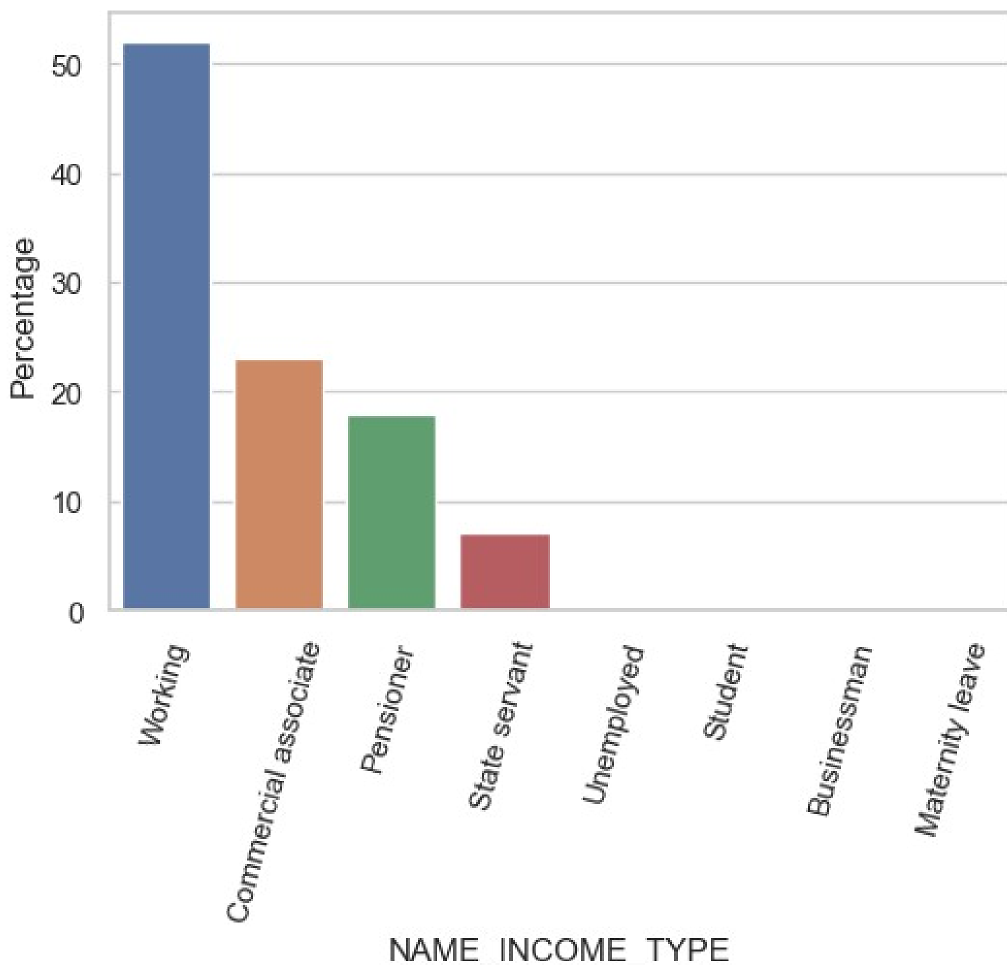
```
[70. 20.  9.  1.  0.  0.  0.  0.  0.  0.  0.]
```



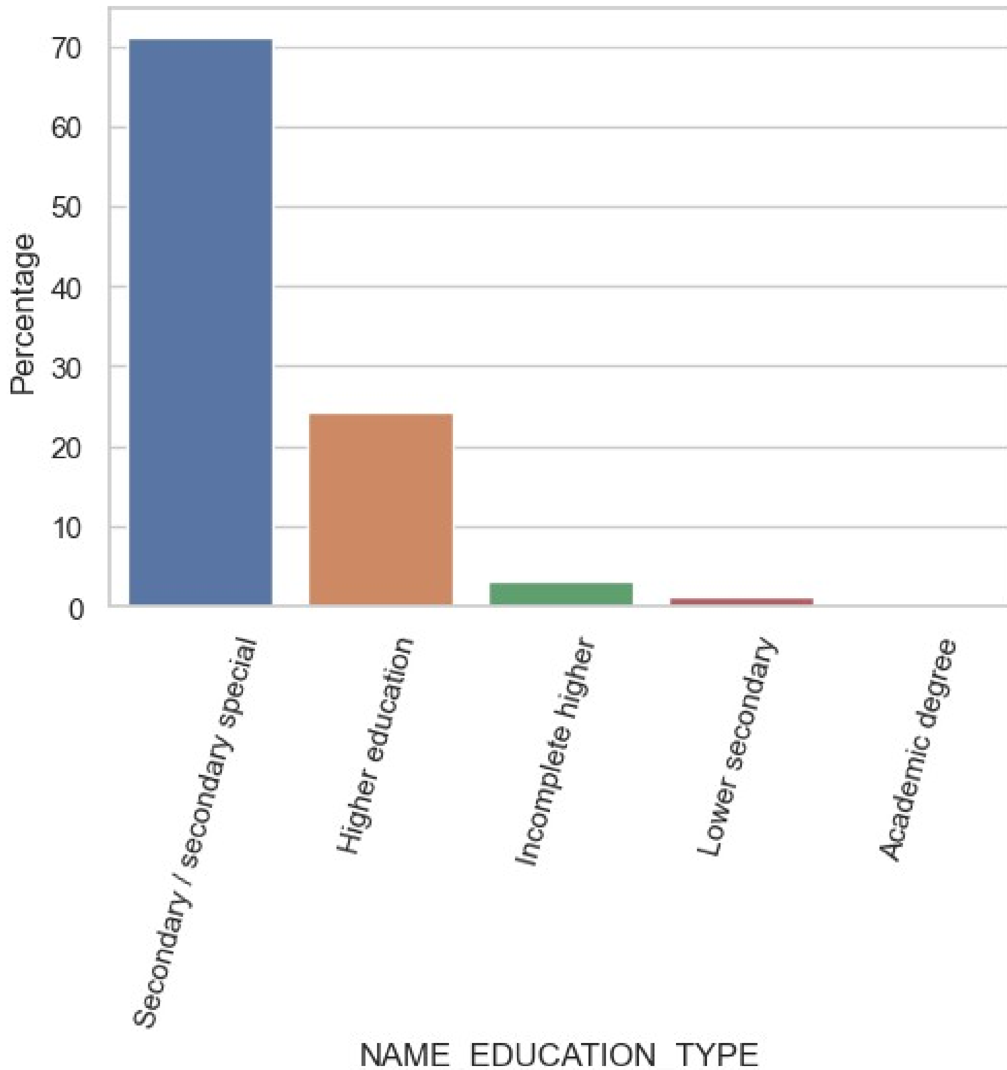

```
[81. 13.  4.  1.  1.  0.  0.]
```



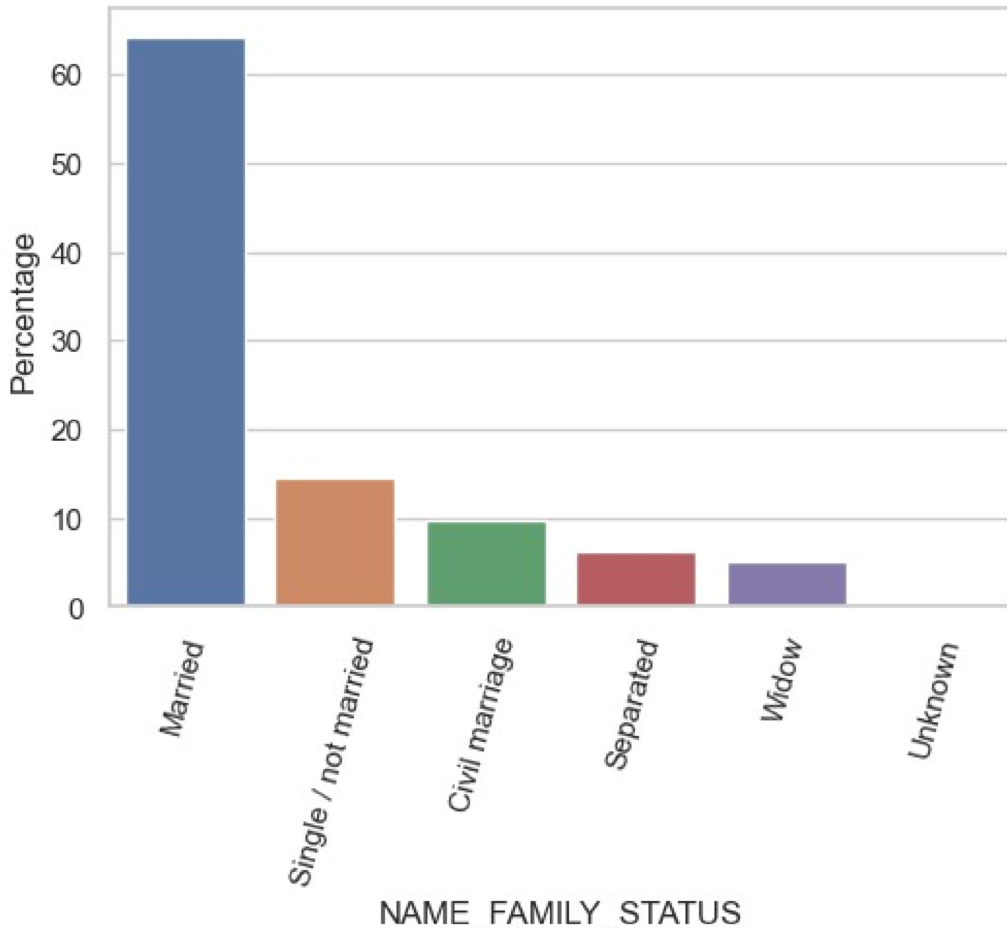

```
[52. 23. 18.  7.  0.  0.  0.  0.]
```


Percentage
50
40
30
20
10
0
Working
Commercial associate
Pensioner
State servant
Unemployed
Student
Businessman
Maternity leave
NAME_INCOME_TYPE

```
[71. 24.  3.  1.  0.]
```

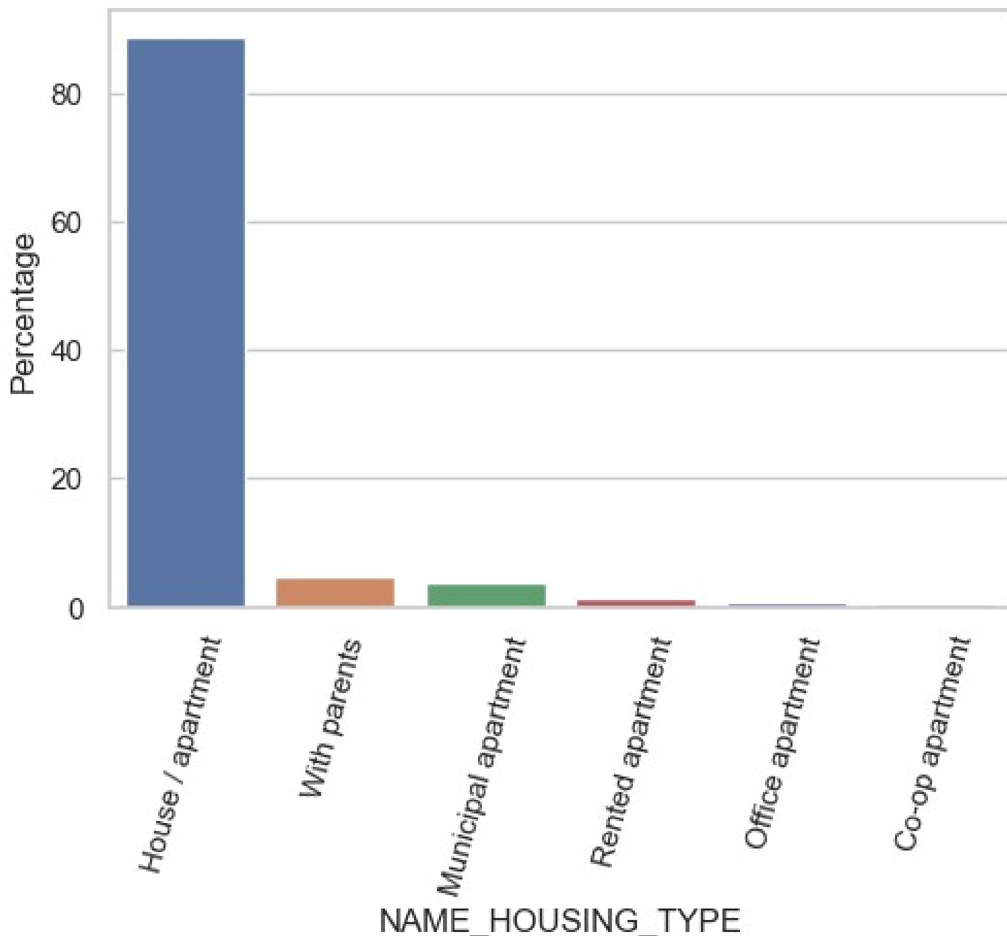

70
60
50
40
30
20
10
0
Percentage
Secondary / secondary special
Higher education
Incomplete higher
Lower secondary
Academic degree
NAME_EDUCATION_TYPE

```
[64. 15. 10.  6.  5.  0.]
```

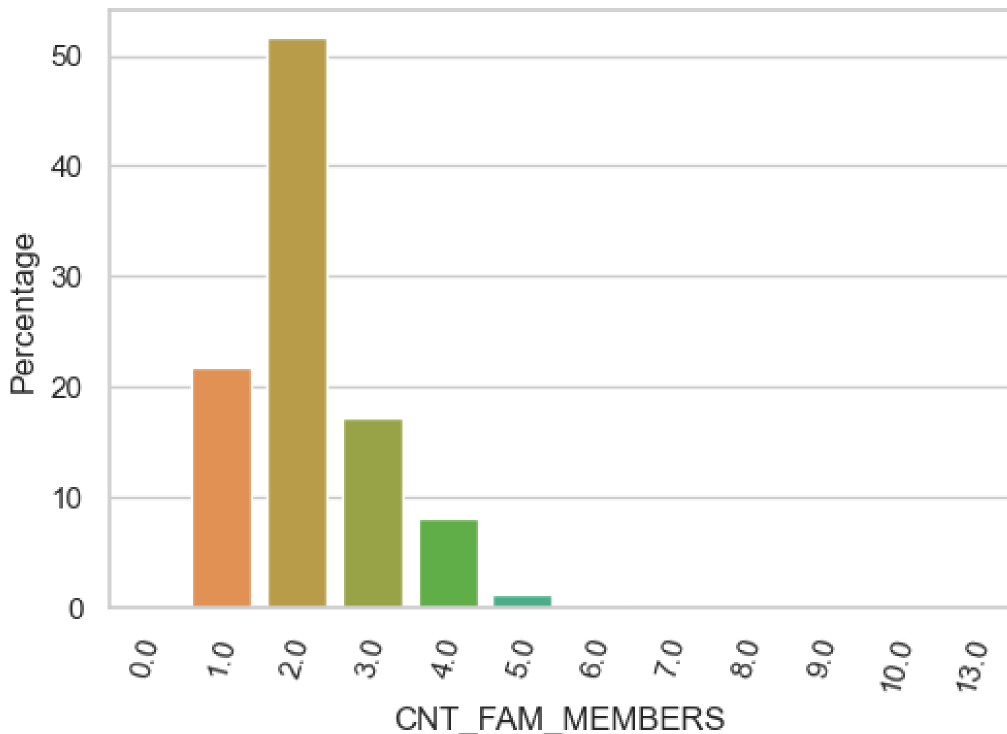


60
50
40
30
20
10
0
Percentage
Married
Single / not married
Civil marriage
Separated
Widow
Unknown
NAME_FAMILY_STATUS

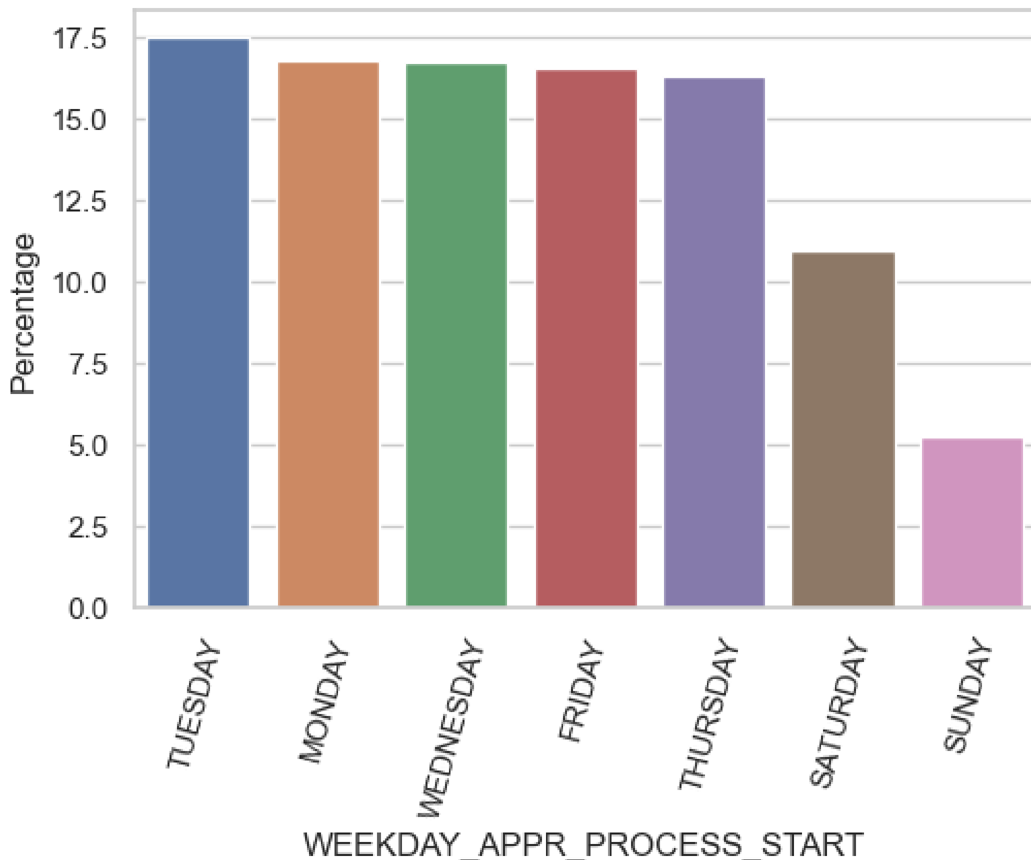

```
[89.  5.  4.  2.  1.  0.]
```



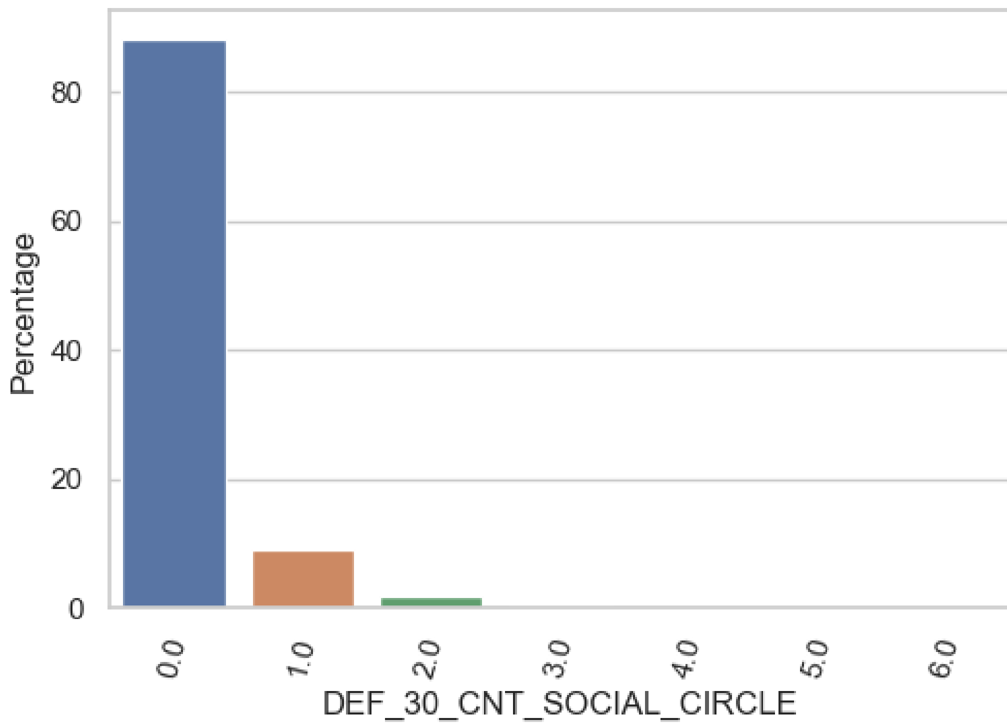

```
[52. 22. 17.  8.  1.  0.  0.  0.  0.  0.  0.  0.]
```



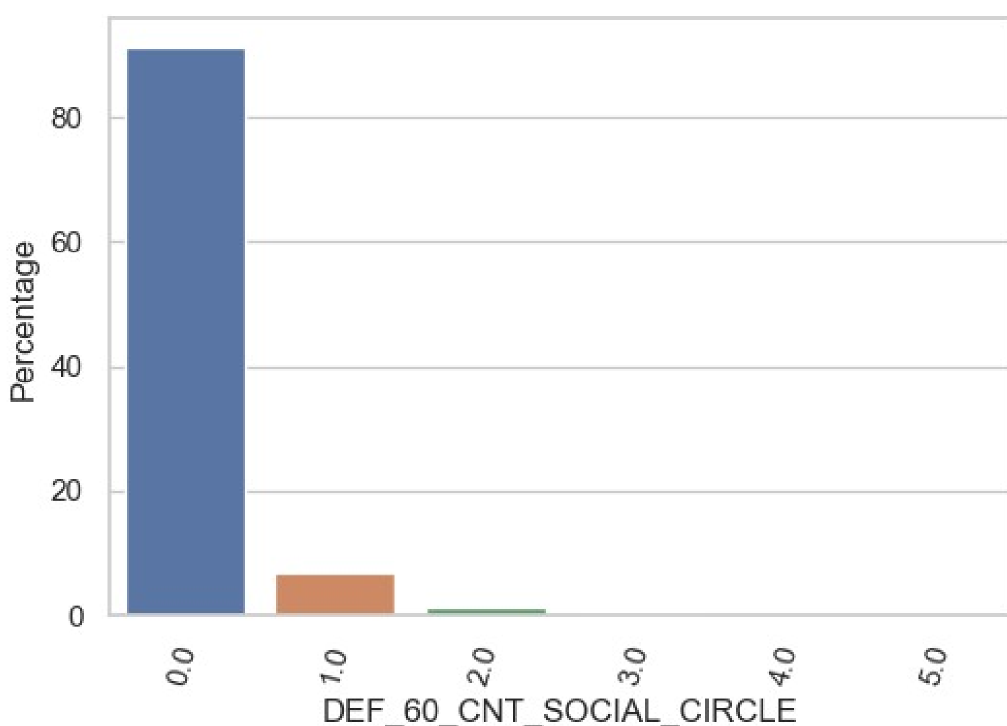

```
[17. 17. 17. 17. 16. 11.  5.]
```


17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Percentage
TUESDAY
MONDAY
WEDNESDAY
FRIDAY
THURSDAY
SATURDAY
SUNDAY
WEEKDAY_APPR_PROCESS_START

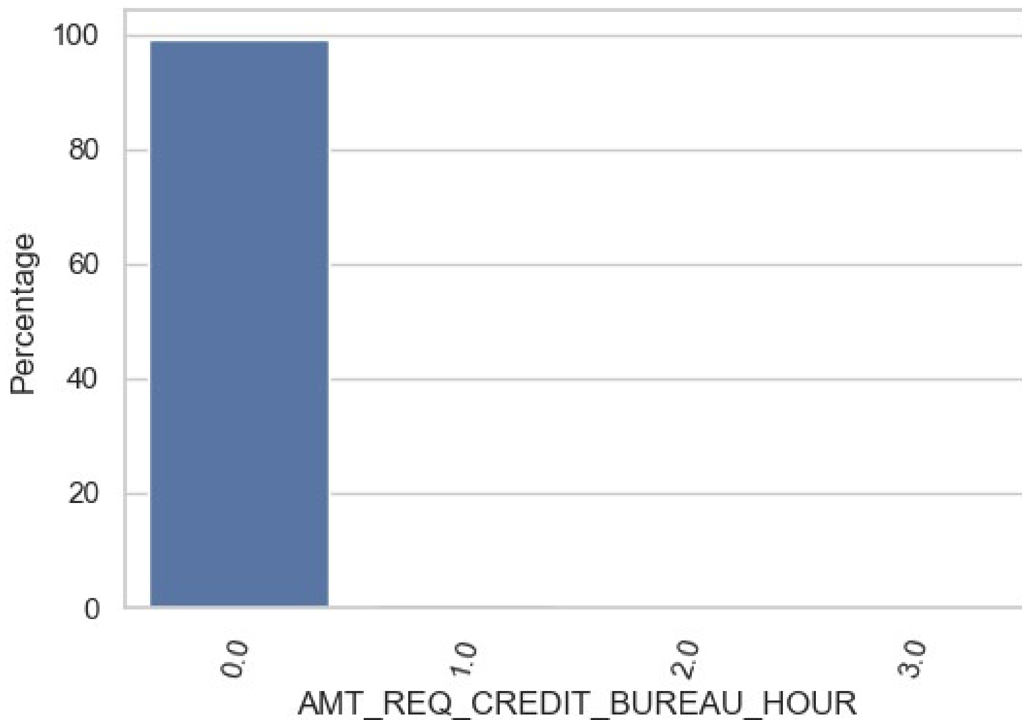
```
[88.  9.  2.  0.  0.  0.  0.]
```



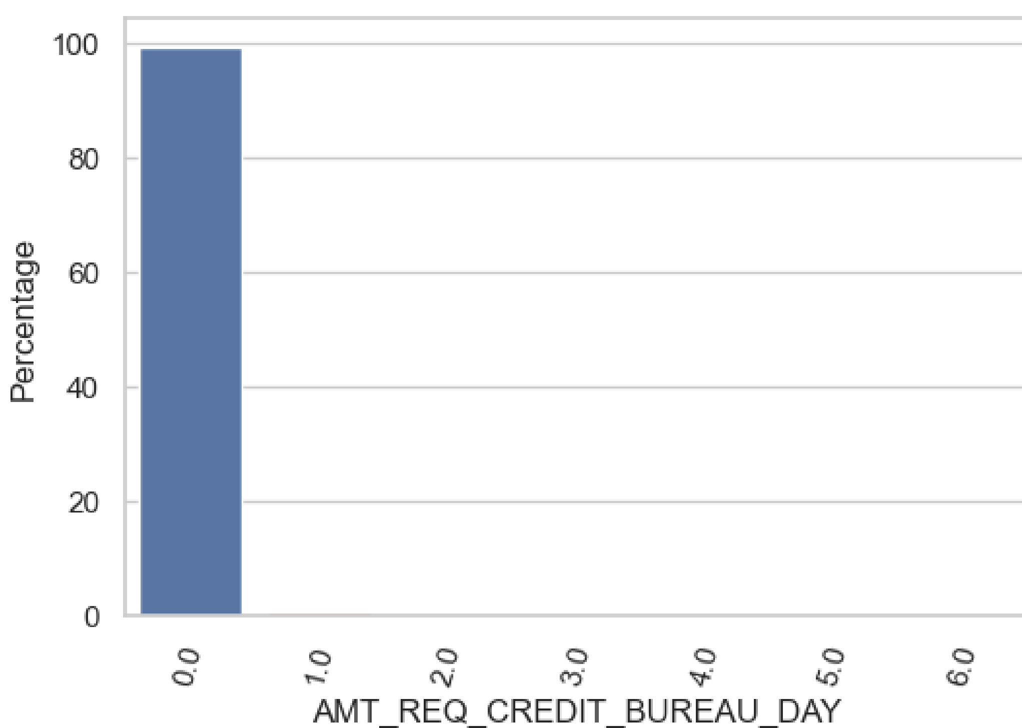

```
[91.  7.  1.  0.  0.  0.]
```



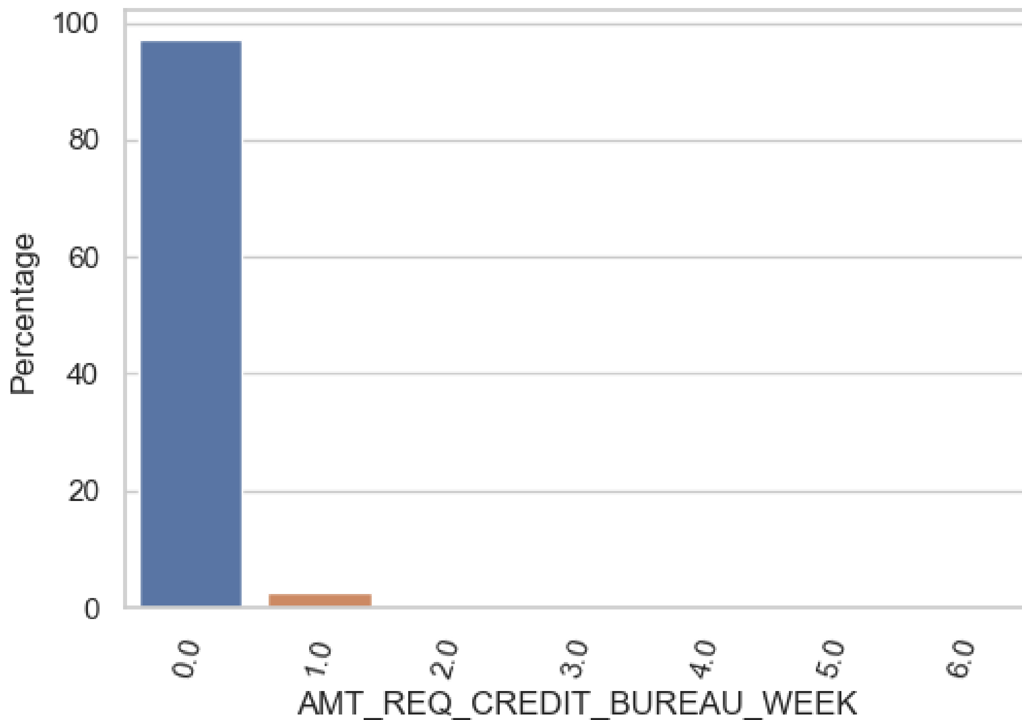

```
[99.  1.  0.  0.]
```



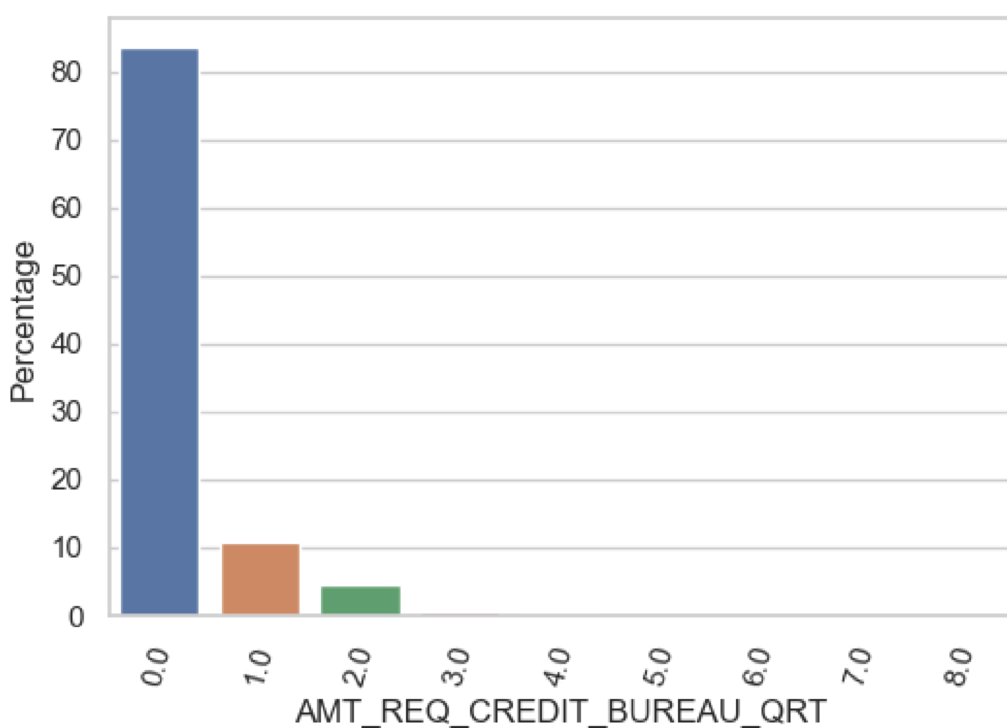

```
[99.  0.  0.  0.  0.  0.  0.]
```




```
[97.  3.  0.  0.  0.  0.  0.]
```




```
[84. 11.  5.  1.  0.  0.  0.  0.  0.]
```



Inference/Observation from the analysis of Data Imbalance

```
# target variable shows us that 92% percent people have the value 1
i.e. client with payment difficulties
# NAME_CONTRACT_TYPE has the 91% of loans of Cash Loans type.
# CODE_GENDER data imbalance shows us that there are 66% female
applicatns and 34% male applicants
# FLAG_OWN_CAR data imbalance shows us that 66% applicants don't own
car.
# FLAG_OWN_REALTY data imbalance shows us that 69% applicants own
house.
# CNT_CHILDREN data imbalance shows us that 70% applicants don't have
childrens.
# NAME_TYPE_SUITE data imbalance shows us that 81% of the applicants
are unaccompined
# NAME_INCOME_TYPE data imbalance shows us that 52% applicants are
working.
# NAME_FAMILY_STATUS data imbalance shows us that 64% applicants are
married.
# NAME_HOUSINGIN_TYPE data imbalance shows us that 52% applicants are
live in House/Apartment.
# CNT_FAM_MEMBERES data imbalance shows us that 52% applicants have 2
family members
# and 22%, 17% applicants have 1 & 3 family members respectively
```

D. Perform Univariate, Segmented Univariate, and Bivariate Analysis

Top Correlations for Different Scenarios

```
application_data.columns

Index(['SK_ID_CURR', 'TARGET', 'NAME_CONTRACT_TYPE', 'CODE_GENDER',
       'FLAG_OWN_CAR', 'FLAG_OWN_REALTY', 'CNT_CHILDREN',
'AMT_INCOME_TOTAL',
       'AMT_CREDIT', 'AMT_ANNUITY', 'AMT_GOODS_PRICE',
'NAME_TYPE_SUITE',
       'NAME_INCOME_TYPE', 'NAME_EDUCATION_TYPE',
'NAME_FAMILY_STATUS',
       'NAME_HOUSING_TYPE', 'REGION_POPULATION_RELATIVE',
'DAYS_BIRTH',
       'DAYS_EMPLOYED', 'DAYS_REGISTRATION', 'DAYS_ID_PUBLISH',
'OWN_CAR_AGE',
       'FLAG_MOBIL', 'FLAG_EMP_PHONE', 'FLAG_WORK_PHONE',
'FLAG_CONT_MOBILE',
       'FLAG_PHONE', 'FLAG_EMAIL', 'OCCUPATION_TYPE',
'CNT_FAM_MEMBERS',
       'REGION_RATING_CLIENT', 'REGION_RATING_CLIENT_W_CITY',
       'WEEKDAY_APPR_PROCESS_START', 'HOUR_APPR_PROCESS_START',
       'REG_REGION_NOT_LIVE_REGION', 'REG_REGION_NOT_WORK_REGION',
       'LIVE_REGION_NOT_WORK_REGION', 'REG_CITY_NOT_LIVE_CITY',
       'REG_CITY_NOT_WORK_CITY', 'LIVE_CITY_NOT_WORK_CITY',
       'ORGANIZATION_TYPE', 'EXT_SOURCE_1', 'EXT_SOURCE_2',
'EXT_SOURCE_3',
       'OBS_30_CNT_SOCIAL_CIRCLE', 'DEF_30_CNT_SOCIAL_CIRCLE',
       'OBS_60_CNT_SOCIAL_CIRCLE', 'DEF_60_CNT_SOCIAL_CIRCLE',
       'DAYS_LAST_PHONE_CHANGE', 'FLAG_DOCUMENT_2', 'FLAG_DOCUMENT_3',
       'FLAG_DOCUMENT_4', 'FLAG_DOCUMENT_5', 'FLAG_DOCUMENT_6',
       'FLAG_DOCUMENT_7', 'FLAG_DOCUMENT_8', 'FLAG_DOCUMENT_9',
       'FLAG_DOCUMENT_10', 'FLAG_DOCUMENT_11', 'FLAG_DOCUMENT_12',
       'FLAG_DOCUMENT_13', 'FLAG_DOCUMENT_14', 'FLAG_DOCUMENT_15',
       'FLAG_DOCUMENT_16', 'FLAG_DOCUMENT_17', 'FLAG_DOCUMENT_18',
       'FLAG_DOCUMENT_19', 'FLAG_DOCUMENT_20', 'FLAG_DOCUMENT_21',
       'AMT_REQ_CREDIT_BUREAU_HOUR', 'AMT_REQ_CREDIT_BUREAU_DAY',
```

```
       'AMT_REQ_CREDIT_BUREAU_WEEK', 'AMT_REQ_CREDIT_BUREAU_MON',
       'AMT_REQ_CREDIT_BUREAU_QRT', 'AMT_REQ_CREDIT_BUREAU_YEAR'],
      dtype='object')
```

Univariate Analysis

Univariate analysis focuses on examining the distribution and summary statistics of individual variables.

```
# Calculated Summary Stats for all numeric variables using the
describe function.

summary_stats = application_data.describe()

# createing the seperate table using the summary_stats for each
variable

from tabulate import tabulate
# create separate tables for each variable
for col in summary_stats.columns:
    col_table = summary_stats[[col]].transpose()

    # Converted the table to a string using tabulate
    table_str = tabulate(col_table, tablefmt='grid')
    print(f"Summary Statistics for {col}:\n{table_str}\n")
    plt.show()

Summary Statistics for SK_ID_CURR:
+------------+-------+--------+---------+--------+--------+--------
+--------+--------+
| SK_ID_CURR | 49999 | 129013 | 16690.5 | 100002 | 114570 | 129076 |
143438 | 157875 |
+------------+-------+--------+---------+--------+--------+--------
+--------+--------+

Summary Statistics for TARGET:
+--------+-------+-----------+----------+---+---+---+---+---+
| TARGET | 49999 | 0.0805216 | 0.272102 | 0 | 0 | 0 | 0 | 1 |
+--------+-------+-----------+----------+---+---+---+---+---+

Summary Statistics for CNT_CHILDREN:
+--------------+-------+----------+----------+---+---+---+---+----+
| CNT_CHILDREN | 49999 | 0.419848 | 0.724039 | 0 | 0 | 0 | 1 | 11 |
+--------------+-------+----------+----------+---+---+---+---+----+

Summary Statistics for AMT_INCOME_TOTAL:
+------------------+-------+--------+--------+-------+--------
+--------+--------+----------+
| AMT_INCOME_TOTAL | 49999 | 170768 | 531819 | 25650 | 112500 | 145800
```

```
| 202500 | 1.17e+08 |
+------------------+-------+--------+--------+-------+--------
+--------+--------+----------+

Summary Statistics for AMT_CREDIT:
+------------+-------+--------+--------+-------+--------+--------
+--------+----------+
| AMT_CREDIT | 49999 | 599701 | 402415 | 45000 | 270000 | 514778 |
808650 | 4.05e+06 |
+------------+-------+--------+--------+-------+--------+--------
+--------+----------+

Summary Statistics for AMT_ANNUITY:
+-------------+-------+---------+---------+------+---------+-------
+-------+--------+
| AMT_ANNUITY | 49998 | 27107.4 | 14562.9 | 2052 | 16456.5 | 24939 |
34596 | 258026 |
+-------------+-------+---------+---------+------+---------+-------
+-------+--------+

Summary Statistics for AMT_GOODS_PRICE:
+-----------------+-------+--------+--------+-------+--------+--------
+--------+----------+
| AMT_GOODS_PRICE | 49961 | 539060 | 369853 | 45000 | 238500 | 450000
| 679500 | 4.05e+06 |
+-----------------+-------+--------+--------+-------+--------+--------
+--------+----------+

Summary Statistics for REGION_POPULATION_RELATIVE:
+----------------------------+-------+-----------+-----------
+----------+----------+---------+----------+----------+
| REGION_POPULATION_RELATIVE | 49999 | 0.0207983 | 0.0137606 |
0.000533 | 0.010006 | 0.01885 | 0.028663 | 0.072508 |
+----------------------------+-------+-----------+-----------
+----------+----------+---------+----------+----------+

Summary Statistics for DAYS_BIRTH:
+------------+-------+-------+--------+------+---------+-------
+-------+-------+
| DAYS_BIRTH | 49999 | 16022 | 4361.4 | 7680 | 12378.5 | 15731 | 19644
| 25184 |
+------------+-------+-------+--------+------+---------+-------
+-------+-------+

Summary Statistics for DAYS_EMPLOYED:
+---------------+-------+---------+--------+---+-----+------+------
+--------+
| DAYS_EMPLOYED | 49999 | 67160.3 | 138958 | 0 | 933 | 2216 | 5718 |
365243 |
+---------------+-------+---------+--------+---+-----+------+------
```

```
+--------+

Summary Statistics for DAYS_REGISTRATION:
+-------------------+-------+---------+---------+---+------+------
+--------+-------+
| DAYS_REGISTRATION | 49999 | 4977.28 | 3525.55 | 0 | 1998 | 4490 |
7463.5 | 22392 |
+-------------------+-------+---------+---------+---+------+------
+--------+-------+

Summary Statistics for DAYS_ID_PUBLISH:
+-----------------+-------+--------+---------+---+------+------+------
+------+
| DAYS_ID_PUBLISH | 49999 | 2996.8 | 1509.24 | 0 | 1722 | 3261 | 4297
| 6232 |
+-----------------+-------+--------+---------+---+------+------+------
+------+

Summary Statistics for OWN_CAR_AGE:
+-------------+-------+---------+---------+---+---+---+----+----+
| OWN_CAR_AGE | 17049 | 12.0257 | 11.8875 | 0 | 5 | 9 | 15 | 65 |
+-------------+-------+---------+---------+---+---+---+----+----+

Summary Statistics for FLAG_MOBIL:
+------------+-------+---------+------------+---+---+---+---+---+
| FLAG_MOBIL | 49999 | 0.99998 | 0.00447218 | 0 | 1 | 1 | 1 | 1 |
+------------+-------+---------+------------+---+---+---+---+---+

Summary Statistics for FLAG_EMP_PHONE:
+----------------+-------+----------+----------+---+---+---+---+---+
| FLAG_EMP_PHONE | 49999 | 0.821476 | 0.382957 | 0 | 1 | 1 | 1 | 1 |
+----------------+-------+----------+----------+---+---+---+---+---+

Summary Statistics for FLAG_WORK_PHONE:
+-----------------+-------+----------+----------+---+---+---+---+---+
| FLAG_WORK_PHONE | 49999 | 0.199264 | 0.399451 | 0 | 0 | 0 | 0 | 1 |
+-----------------+-------+----------+----------+---+---+---+---+---+

Summary Statistics for FLAG_CONT_MOBILE:
+------------------+-------+---------+-----------+---+---+---+---+---+
| FLAG_CONT_MOBILE | 49999 | 0.99798 | 0.0448999 | 0 | 1 | 1 | 1 | 1 |
+------------------+-------+---------+-----------+---+---+---+---+---+

Summary Statistics for FLAG_PHONE:
+------------+-------+----------+----------+---+---+---+---+---+
| FLAG_PHONE | 49999 | 0.277726 | 0.447882 | 0 | 0 | 0 | 1 | 1 |
+------------+-------+----------+----------+---+---+---+---+---+

Summary Statistics for FLAG_EMAIL:
+------------+-------+-----------+----------+---+---+---+---+---+
```

```
| FLAG_EMAIL | 49999 | 0.0556611 | 0.229268 | 0 | 0 | 0 | 0 | 1 |
+------------+-------+-----------+----------+---+---+---+---+---+

Summary Statistics for CNT_FAM_MEMBERS:
+-----------------+-------+--------+----------+---+---+---+---+----+
| CNT_FAM_MEMBERS | 49999 | 2.1589 | 0.911375 | 0 | 2 | 2 | 3 | 13 |
+-----------------+-------+--------+----------+---+---+---+---+----+

Summary Statistics for REGION_RATING_CLIENT:
+----------------------+-------+---------+----------+---+---+---+---
+---+
| REGION_RATING_CLIENT | 49999 | 2.05166 | 0.507978 | 1 | 2 | 2 | 2 |
3 |
+----------------------+-------+---------+----------+---+---+---+---
+---+

Summary Statistics for REGION_RATING_CLIENT_W_CITY:
+-----------------------------+-------+---------+----------+---+---
+---+---+---+
| REGION_RATING_CLIENT_W_CITY | 49999 | 2.03072 | 0.502221 | 1 | 2 | 2
| 2 | 3 |
+-----------------------------+-------+---------+----------+---+---
+---+---+---+

Summary Statistics for HOUR_APPR_PROCESS_START:
+-------------------------+-------+---------+---------+---+----+----
+----+----+
| HOUR_APPR_PROCESS_START | 49999 | 12.0526 | 3.25258 | 0 | 10 | 12 |
14 | 23 |
+-------------------------+-------+---------+---------+---+----+----
+----+----+

Summary Statistics for REG_REGION_NOT_LIVE_REGION:
+----------------------------+-------+-----------+----------+---+---
+---+---+---+
| REG_REGION_NOT_LIVE_REGION | 49999 | 0.0150003 | 0.121555 | 0 | 0 |
0 | 0 | 1 |
+----------------------------+-------+-----------+----------+---+---
+---+---+---+

Summary Statistics for REG_REGION_NOT_WORK_REGION:
+----------------------------+-------+----------+----------+---+---
+---+---+---+
| REG_REGION_NOT_WORK_REGION | 49999 | 0.049921 | 0.217784 | 0 | 0 | 0
| 0 | 1 |
+----------------------------+-------+----------+----------+---+---
+---+---+---+

Summary Statistics for LIVE_REGION_NOT_WORK_REGION:
+-----------------------------+-------+-----------+----------+---+---
```

```
+---+---+---+
| LIVE_REGION_NOT_WORK_REGION | 49999 | 0.0396408 | 0.195116 | 0 | 0 |
0 | 0 | 1 |
+-----------------------------+-------+-----------+----------+---+---
+---+---+---+

Summary Statistics for REG_CITY_NOT_LIVE_CITY:
+-------------------------+-------+-----------+----------+---+---+---
+---+---+
| REG_CITY_NOT_LIVE_CITY | 49999 | 0.0799616 | 0.271236 | 0 | 0 | 0 |
0 | 1 |
+-------------------------+-------+-----------+----------+---+---+---
+---+---+

Summary Statistics for REG_CITY_NOT_WORK_CITY:
+-------------------------+-------+----------+----------+---+---+---
+---+---+
| REG_CITY_NOT_WORK_CITY | 49999 | 0.232165 | 0.422218 | 0 | 0 | 0 | 0
| 1 |
+-------------------------+-------+----------+----------+---+---+---
+---+---+

Summary Statistics for LIVE_CITY_NOT_WORK_CITY:
+--------------------------+-------+----------+----------+---+---+---
+---+---+
| LIVE_CITY_NOT_WORK_CITY | 49999 | 0.179704 | 0.383944 | 0 | 0 | 0 |
0 | 1 |
+--------------------------+-------+----------+----------+---+---+---
+---+---+

Summary Statistics for EXT_SOURCE_1:
+---------------+-------+----------+----------+-----------+----------
+----------+----------+----------+
| EXT_SOURCE_1 | 21827 | 0.502257 | 0.211017 | 0.0145681 | 0.333476 |
0.506884 | 0.673923 | 0.93825 |
+---------------+-------+----------+----------+-----------+----------
+----------+----------+----------+

Summary Statistics for EXT_SOURCE_2:
+---------------+-------+----------+----------+----------+----------
+----------+----------+-------+
| EXT_SOURCE_2 | 49873 | 0.513824 | 0.191165 | 8.17e-08 | 0.391722 |
0.565585 | 0.663402 | 0.855 |
+---------------+-------+----------+----------+----------+----------
+----------+----------+-------+

Summary Statistics for EXT_SOURCE_3:
+---------------+-------+----------+--------+-------------+---------
+----------+----------+----------+
| EXT_SOURCE_3 | 40055 | 0.511881 | 0.1947 | 0.000527265 | 0.37065 |
```

```
0.535276 | 0.669057 | 0.89601 |
+--------------+-------+---------+-------+------------+---------
+----------+---------+---------+

Summary Statistics for OBS_30_CNT_SOCIAL_CIRCLE:
+--------------------------+-------+---------+---------+---+---+---
+---+----+
| OBS_30_CNT_SOCIAL_CIRCLE | 49831 | 1.42078 | 2.30209 | 0 | 0 | 0 | 2
| 28 |
+--------------------------+-------+---------+---------+---+---+---
+---+----+

Summary Statistics for DEF_30_CNT_SOCIAL_CIRCLE:
+--------------------------+-------+----------+---------+---+---+---
+---+---+
| DEF_30_CNT_SOCIAL_CIRCLE | 49831 | 0.141819 | 0.44054 | 0 | 0 | 0 |
0 | 6 |
+--------------------------+-------+----------+---------+---+---+---
+---+---+

Summary Statistics for OBS_60_CNT_SOCIAL_CIRCLE:
+--------------------------+-------+---------+---------+---+---+---
+---+----+
| OBS_60_CNT_SOCIAL_CIRCLE | 49831 | 1.40366 | 2.28178 | 0 | 0 | 0 | 2
| 28 |
+--------------------------+-------+---------+---------+---+---+---
+---+----+

Summary Statistics for DEF_60_CNT_SOCIAL_CIRCLE:
+--------------------------+-------+-----------+----------+---+---+---
+---+---+
| DEF_60_CNT_SOCIAL_CIRCLE | 49831 | 0.0983324 | 0.357264 | 0 | 0 | 0
| 0 | 5 |
+--------------------------+-------+-----------+----------+---+---+---
+---+---+

Summary Statistics for DAYS_LAST_PHONE_CHANGE:
+------------------------+-------+----------+---------+-------+-------
+------+------+---+
| DAYS_LAST_PHONE_CHANGE | 49998 | -964.296 | 829.486 | -4002 | -1573
| -755 | -270 | 0 |
+------------------------+-------+----------+---------+-------+-------
+------+------+---+

Summary Statistics for FLAG_DOCUMENT_2:
+-----------------+-------+-------------+------------+---+---+---+---
+---+
| FLAG_DOCUMENT_2 | 49999 | 4.00008e-05 | 0.00632456 | 0 | 0 | 0 | 0 |
1 |
+-----------------+-------+-------------+------------+---+---+---+---
```

```
+---+

Summary Statistics for FLAG_DOCUMENT_3:
+-----------------+-------+----------+----------+---+---+---+---+---+
| FLAG_DOCUMENT_3 | 49999 | 0.712254 | 0.452717 | 0 | 0 | 1 | 1 | 1 |
+-----------------+-------+----------+----------+---+---+---+---+---+

Summary Statistics for FLAG_DOCUMENT_4:
+-----------------+-------+-------------+-----------+---+---+---+---
+---+
| FLAG_DOCUMENT_4 | 49999 | 0.000180004 | 0.0134155 | 0 | 0 | 0 | 0 |
1 |
+-----------------+-------+-------------+-----------+---+---+---+---
+---+

Summary Statistics for FLAG_DOCUMENT_5:
+-----------------+-------+-----------+----------+---+---+---+---+---+
| FLAG_DOCUMENT_5 | 49999 | 0.0157003 | 0.124315 | 0 | 0 | 0 | 0 | 1 |
+-----------------+-------+-----------+----------+---+---+---+---+---+

Summary Statistics for FLAG_DOCUMENT_6:
+-----------------+-------+-----------+--------+---+---+---+---+---+
| FLAG_DOCUMENT_6 | 49999 | 0.0867017 | 0.2814 | 0 | 0 | 0 | 0 | 1 |
+-----------------+-------+-----------+--------+---+---+---+---+---+

Summary Statistics for FLAG_DOCUMENT_7:
+-----------------+-------+-------------+-----------+---+---+---+---
+---+
| FLAG_DOCUMENT_7 | 49999 | 0.000220004 | 0.0148311 | 0 | 0 | 0 | 0 |
1 |
+-----------------+-------+-------------+-----------+---+---+---+---
+---+

Summary Statistics for FLAG_DOCUMENT_8:
+-----------------+-------+-----------+----------+---+---+---+---+---+
| FLAG_DOCUMENT_8 | 49999 | 0.0807616 | 0.272471 | 0 | 0 | 0 | 0 | 1 |
+-----------------+-------+-----------+----------+---+---+---+---+---+

Summary Statistics for FLAG_DOCUMENT_9:
+-----------------+-------+------------+-----------+---+---+---+---
+---+
| FLAG_DOCUMENT_9 | 49999 | 0.00368007 | 0.0605525 | 0 | 0 | 0 | 0 | 1
|
+-----------------+-------+------------+-----------+---+---+---+---
+---+

Summary Statistics for FLAG_DOCUMENT_10:
+------------------+-------+-------------+------------+---+---+---+---
+---+
| FLAG_DOCUMENT_10 | 49999 | 2.00004e-05 | 0.00447218 | 0 | 0 | 0 | 0
```

```
| 1 |
+------------------+-------+-------------+------------+---+---+---+---
+---+

Summary Statistics for FLAG_DOCUMENT_11:
+------------------+-------+------------+-----------+---+---+---+---
+---+
| FLAG_DOCUMENT_11 | 49999 | 0.00426009 | 0.0651308 | 0 | 0 | 0 | 0 |
1 |
+------------------+-------+------------+-----------+---+---+---+---
+---+

Summary Statistics for FLAG_DOCUMENT_12:
+------------------+-------+---+---+---+---+---+---+---+
| FLAG_DOCUMENT_12 | 49999 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
+------------------+-------+---+---+---+---+---+---+---+

Summary Statistics for FLAG_DOCUMENT_13:
+------------------+-------+------------+-----------+---+---+---+---
+---+
| FLAG_DOCUMENT_13 | 49999 | 0.00322006 | 0.0566547 | 0 | 0 | 0 | 0 |
1 |
+------------------+-------+------------+-----------+---+---+---+---
+---+

Summary Statistics for FLAG_DOCUMENT_14:
+------------------+-------+------------+-----------+---+---+---+---
+---+
| FLAG_DOCUMENT_14 | 49999 | 0.00316006 | 0.0561261 | 0 | 0 | 0 | 0 |
1 |
+------------------+-------+------------+-----------+---+---+---+---
+---+

Summary Statistics for FLAG_DOCUMENT_15:
+------------------+-------+-------------+-----------+---+---+---+---
+---+
| FLAG_DOCUMENT_15 | 49999 | 0.000820016 | 0.0286245 | 0 | 0 | 0 | 0 |
1 |
+------------------+-------+-------------+-----------+---+---+---+---
+---+

Summary Statistics for FLAG_DOCUMENT_16:
+------------------+-------+-----------+-----------+---+---+---+---
+---+
| FLAG_DOCUMENT_16 | 49999 | 0.0100202 | 0.0995992 | 0 | 0 | 0 | 0 | 1
|
+------------------+-------+-----------+-----------+---+---+---+---
+---+

Summary Statistics for FLAG_DOCUMENT_17:
```

```
+------------------+-------+-------------+-----------+---+---+---+---
+---+
| FLAG_DOCUMENT_17 | 49999 | 0.000300006 | 0.0173183 | 0 | 0 | 0 | 0 |
1 |
+------------------+-------+-------------+-----------+---+---+---+---
+---+

Summary Statistics for FLAG_DOCUMENT_18:
+------------------+-------+------------+-----------+---+---+---+---
+---+
| FLAG_DOCUMENT_18 | 49999 | 0.00850017 | 0.0918046 | 0 | 0 | 0 | 0 |
1 |
+------------------+-------+------------+-----------+---+---+---+---
+---+

Summary Statistics for FLAG_DOCUMENT_19:
+------------------+-------+-------------+-----------+---+---+---+---
+---+
| FLAG_DOCUMENT_19 | 49999 | 0.000700014 | 0.0264488 | 0 | 0 | 0 | 0 |
1 |
+------------------+-------+-------------+-----------+---+---+---+---
+---+

Summary Statistics for FLAG_DOCUMENT_20:
+------------------+-------+------------+----------+---+---+---+---
+---+
| FLAG_DOCUMENT_20 | 49999 | 0.00052001 | 0.022798 | 0 | 0 | 0 | 0 | 1
|
+------------------+-------+------------+----------+---+---+---+---
+---+

Summary Statistics for FLAG_DOCUMENT_21:
+------------------+-------+-------------+-----------+---+---+---+---
+---+
| FLAG_DOCUMENT_21 | 49999 | 0.000380008 | 0.0194903 | 0 | 0 | 0 | 0 |
1 |
+------------------+-------+-------------+-----------+---+---+---+---
+---+

Summary Statistics for AMT_REQ_CREDIT_BUREAU_HOUR:
+----------------------------+-------+------------+-----------+---+---
+---+---+---+
| AMT_REQ_CREDIT_BUREAU_HOUR | 49999 | 0.00614012 | 0.0816245 | 0 | 0
| 0 | 0 | 3 |
+----------------------------+-------+------------+-----------+---+---
+---+---+---+

Summary Statistics for AMT_REQ_CREDIT_BUREAU_DAY:
+---------------------------+-------+------------+---------+---+---
+---+---+---+
```

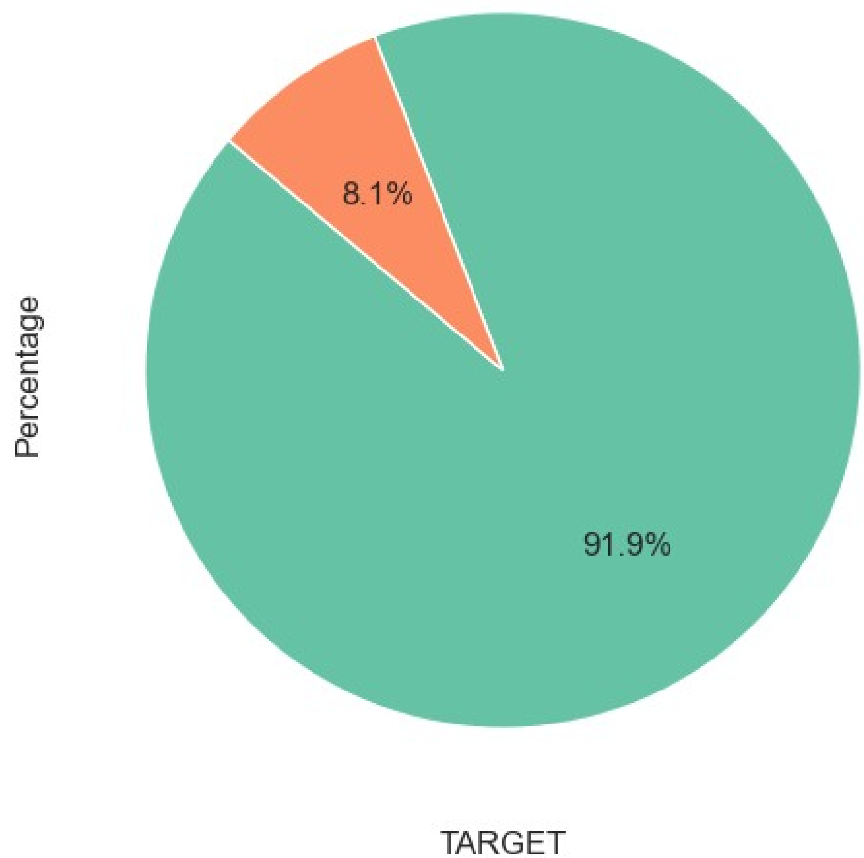
```
| AMT_REQ_CREDIT_BUREAU_DAY | 49999 | 0.00650013 | 0.10049 | 0 | 0 | 0
| 0 | 6 |
+---------------------------+-------+------------+---------+---+---
+---+---+---+

Summary Statistics for AMT_REQ_CREDIT_BUREAU_WEEK:
+----------------------------+-------+-----------+----------+---+---
+---+---+---+
| AMT_REQ_CREDIT_BUREAU_WEEK | 49999 | 0.0280206 | 0.180876 | 0 | 0 |
0 | 0 | 6 |
+----------------------------+-------+-----------+----------+---+---
+---+---+---+

Summary Statistics for AMT_REQ_CREDIT_BUREAU_MON:
+---------------------------+-------+----------+----------+---+---+---
+---+----+
| AMT_REQ_CREDIT_BUREAU_MON | 49999 | 0.233885 | 0.868682 | 0 | 0 | 0
| 0 | 24 |
+---------------------------+-------+----------+----------+---+---+---
+---+----+

Summary Statistics for AMT_REQ_CREDIT_BUREAU_QRT:
+---------------------------+-------+----------+----------+---+---+---
+---+---+
| AMT_REQ_CREDIT_BUREAU_QRT | 49999 | 0.225825 | 0.571627 | 0 | 0 | 0
| 0 | 8 |
+---------------------------+-------+----------+----------+---+---+---
+---+---+

Summary Statistics for AMT_REQ_CREDIT_BUREAU_YEAR:
+----------------------------+-------+---------+---------+---+---+---
+---+----+
| AMT_REQ_CREDIT_BUREAU_YEAR | 49999 | 1.62769 | 1.84995 | 0 | 0 | 1 |
3 | 25 |
+----------------------------+-------+---------+---------+---+---+---
+---+----+
```

Segmented Univariate Analysis:

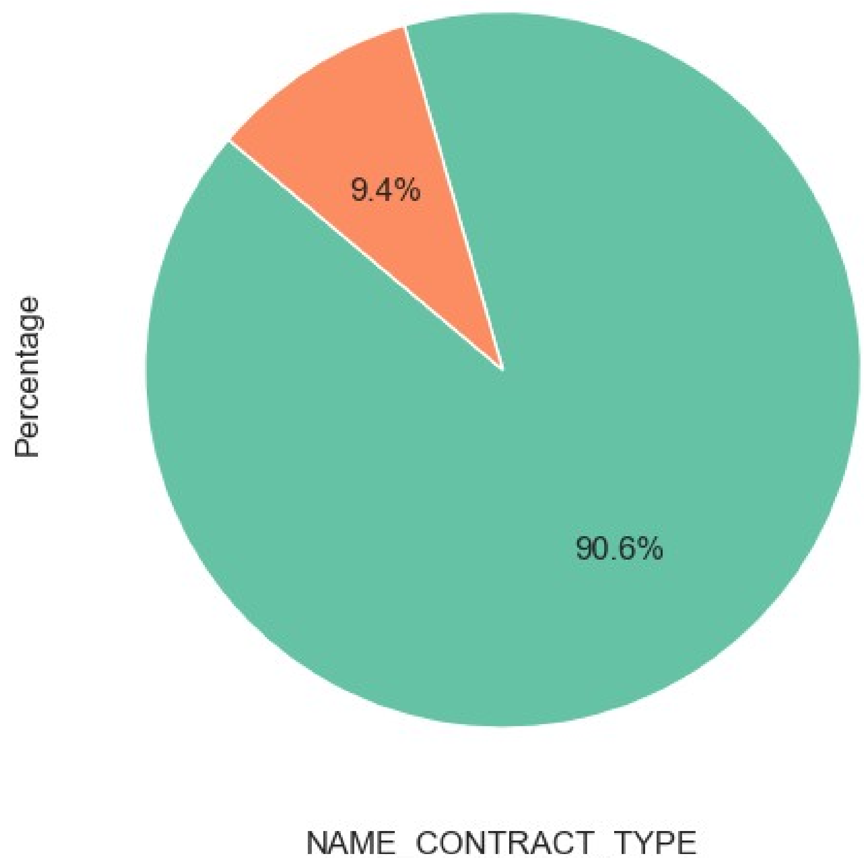
```
for i in application_data.columns:
    # Calculate the class frequencies
    class_frequencies = application_data[i].value_counts()

    # Create a pie chart to performt Univariate Analysis.
    # Using if condition to give only those we are categorical
variables.
```

```
    if len(class_frequencies)<=10:

        print(class_frequencies)
        # printing the percentage of each value_count in list
        percentage =
np.around((class_frequencies.values)/len(application_data[i])*100)
        print(percentage)

        # Create a bar chart to visualize Data imbalance.
        plt.figure(figsize=(6, 6))
        #sns.pie(x = class_frequencies.index, y =
(class_frequencies.values)/len(application_data[i])*100)
        sns.set_palette("Set2")  # You can choose a different color
palette

plt.pie((class_frequencies.values)/len(application_data[i])*100,
autopct='%1.1f%%', startangle=140)
        plt.xlabel(i)
        plt.ylabel("Percentage")
        plt.show()
```

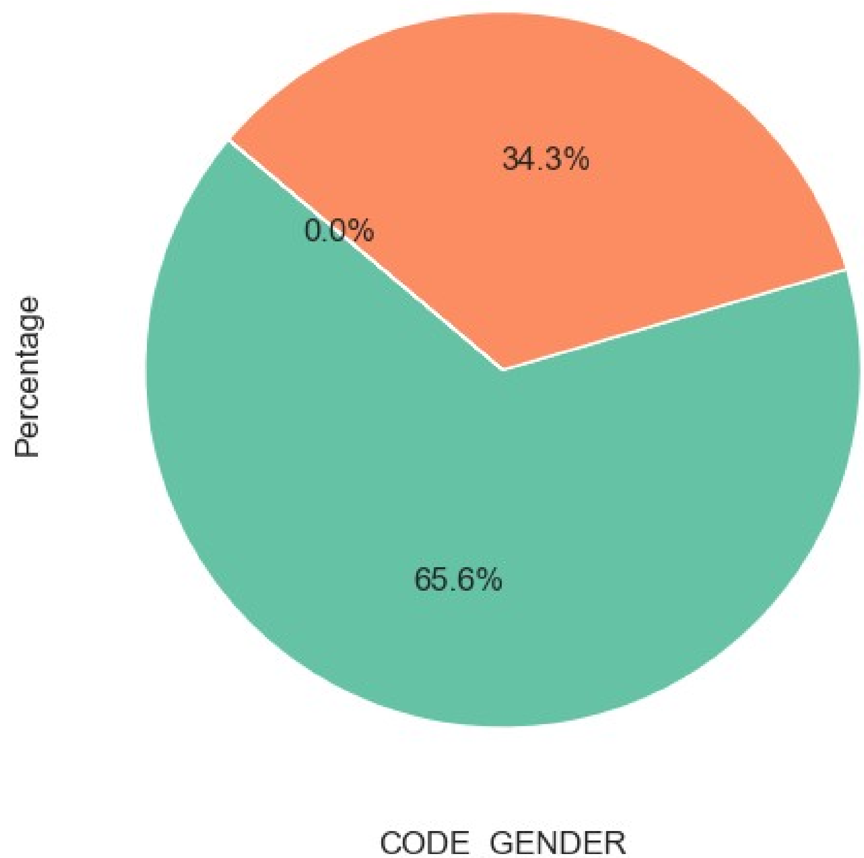
```
0    45973
1     4026
Name: TARGET, dtype: int64
[92.  8.]
```



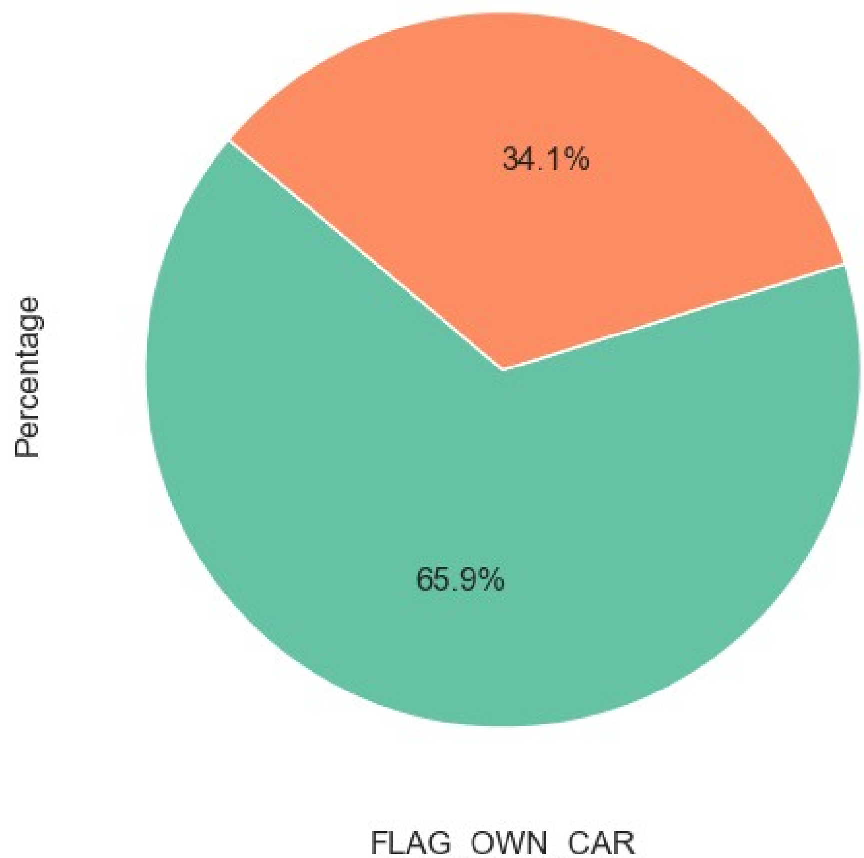

```
Cash loans         45276
Revolving loans     4723
Name: NAME_CONTRACT_TYPE, dtype: int64
[91.  9.]
```



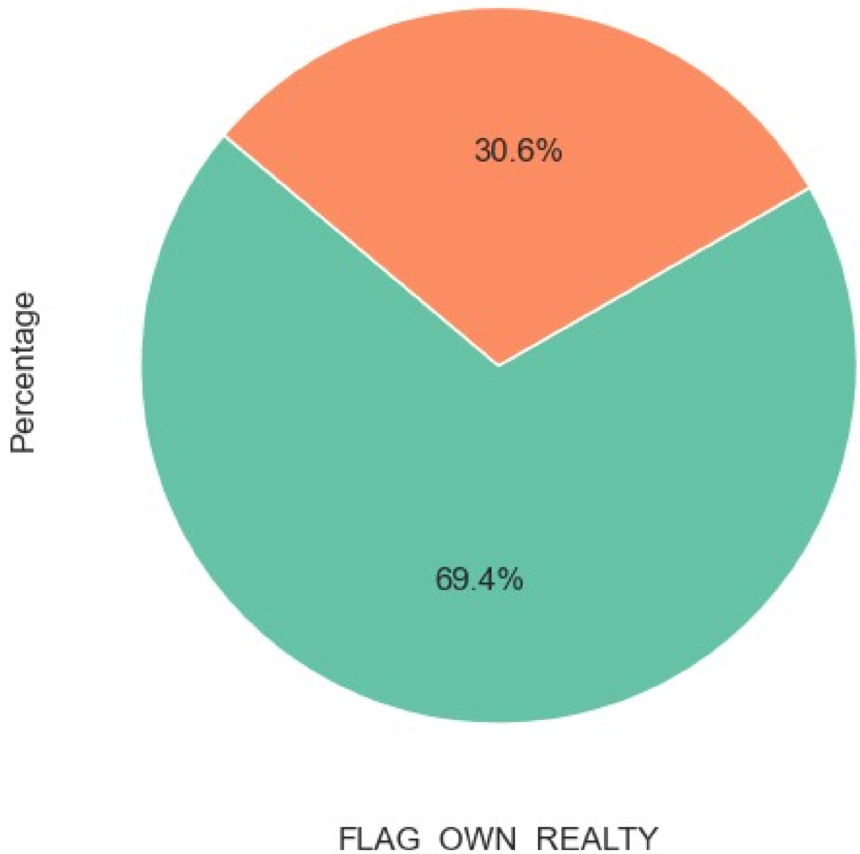

```
F      32823
M      17174
XNA        2
Name: CODE_GENDER, dtype: int64
[66. 34.  0.]
```



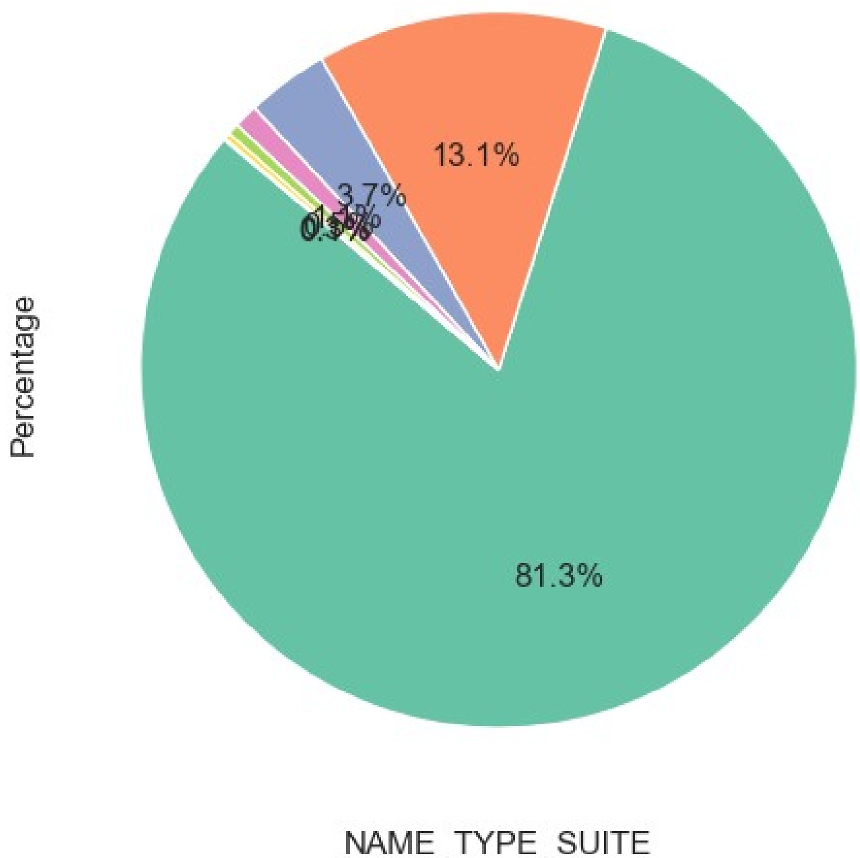

```
N    32949
Y    17050
Name: FLAG_OWN_CAR, dtype: int64
[66. 34.]
```



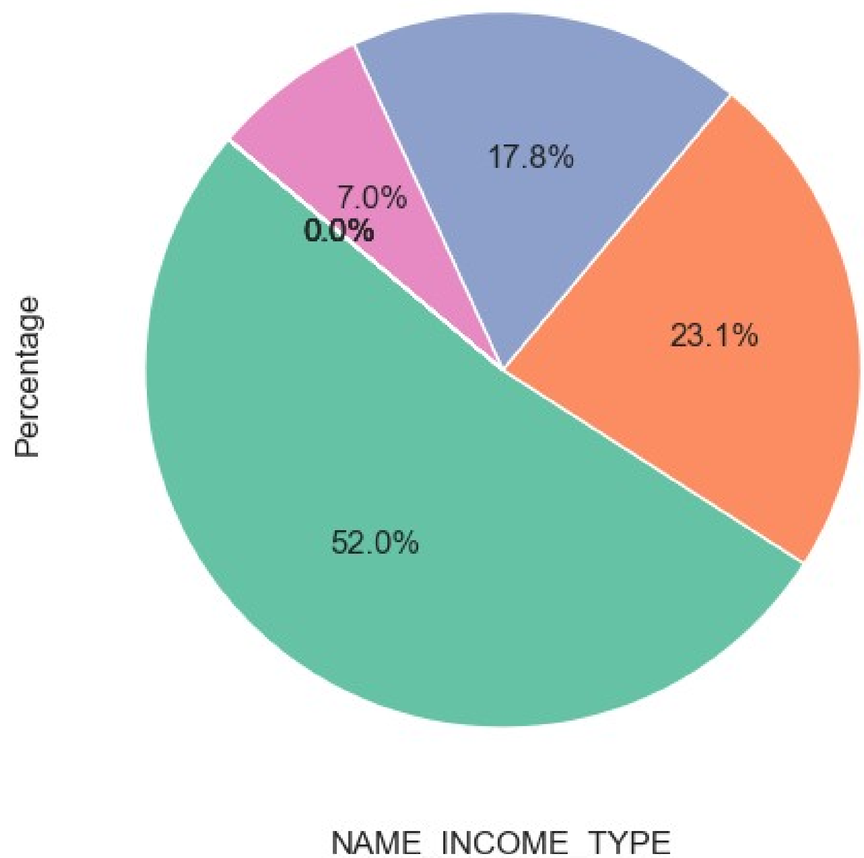

```
Y    34691
N    15308
Name: FLAG_OWN_REALTY, dtype: int64
[69. 31.]
```



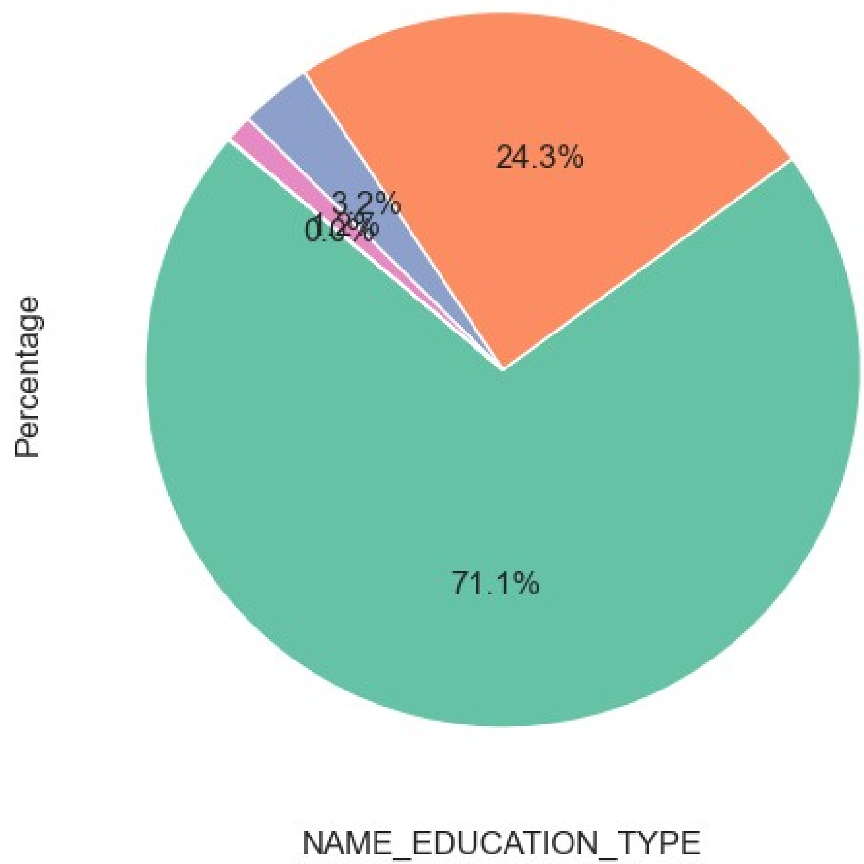

```
Unaccompanied      40627
Family              6549
Spouse, partner     1849
Children             542
Other_B              259
Other_A              137
Group of people       36
Name: NAME_TYPE_SUITE, dtype: int64
[81. 13.  4.  1.  1.  0.  0.]
```



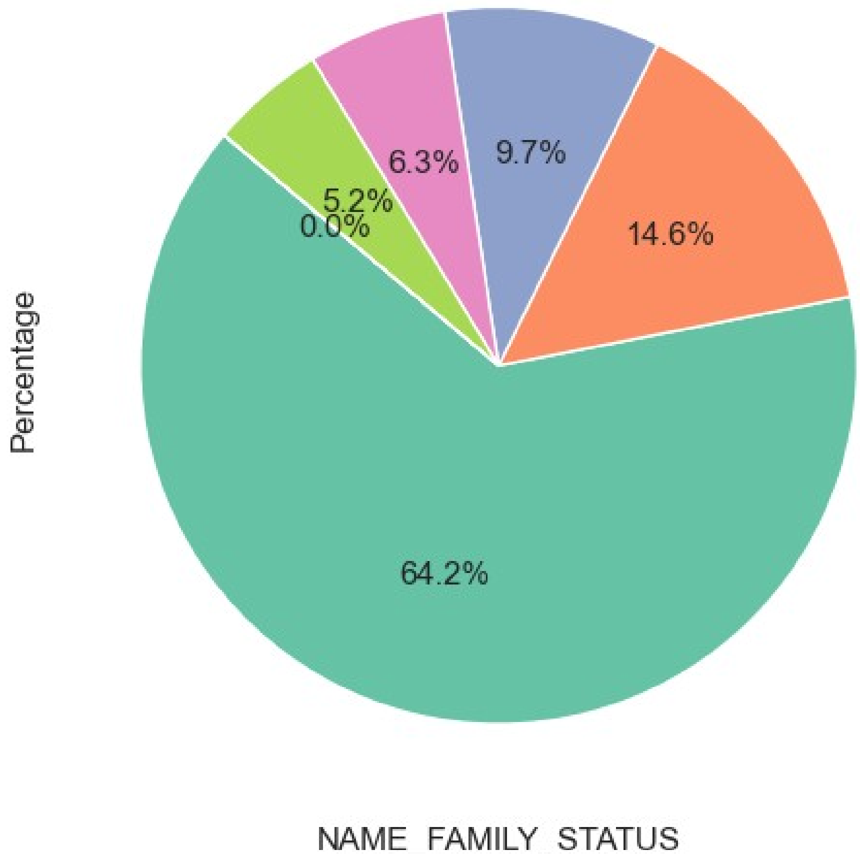

```
Working                  26010
Commercial associate     11543
Pensioner                 8920
State servant             3512
Unemployed                   6
Student                      5
Businessman                  2
Maternity leave              1
Name: NAME_INCOME_TYPE, dtype: int64
[52. 23. 18.  7.  0.  0.  0.  0.]
```



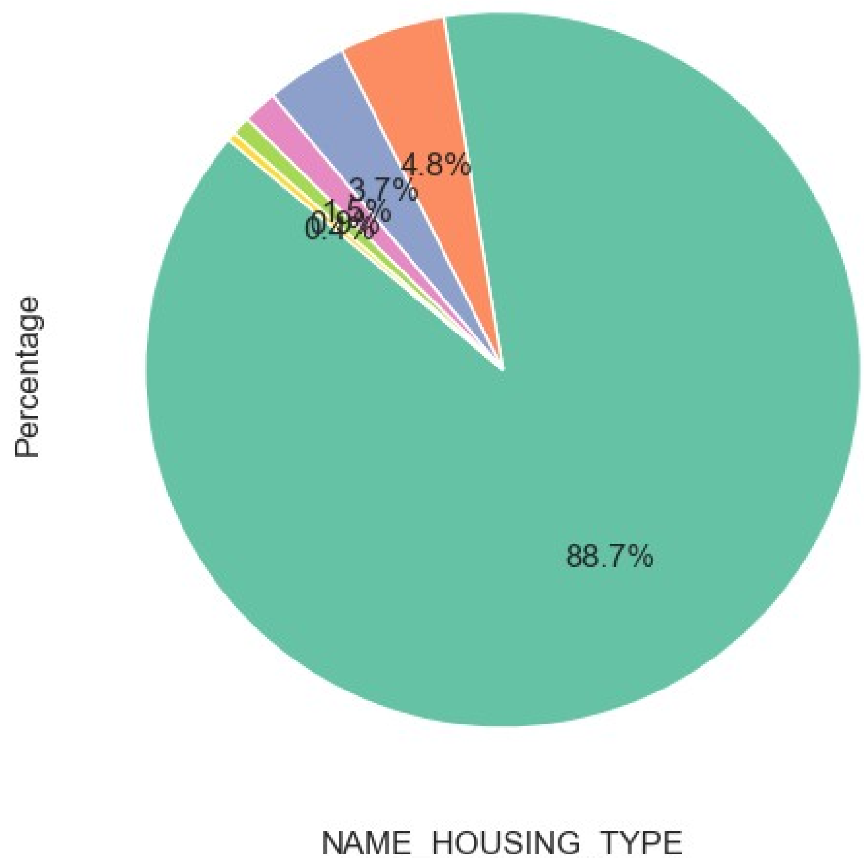

```
Secondary / secondary special    35572
Higher education                 12167
Incomplete higher                 1620
Lower secondary                    620
Academic degree                     20
Name: NAME_EDUCATION_TYPE, dtype: int64
[71. 24.  3.  1.  0.]
```



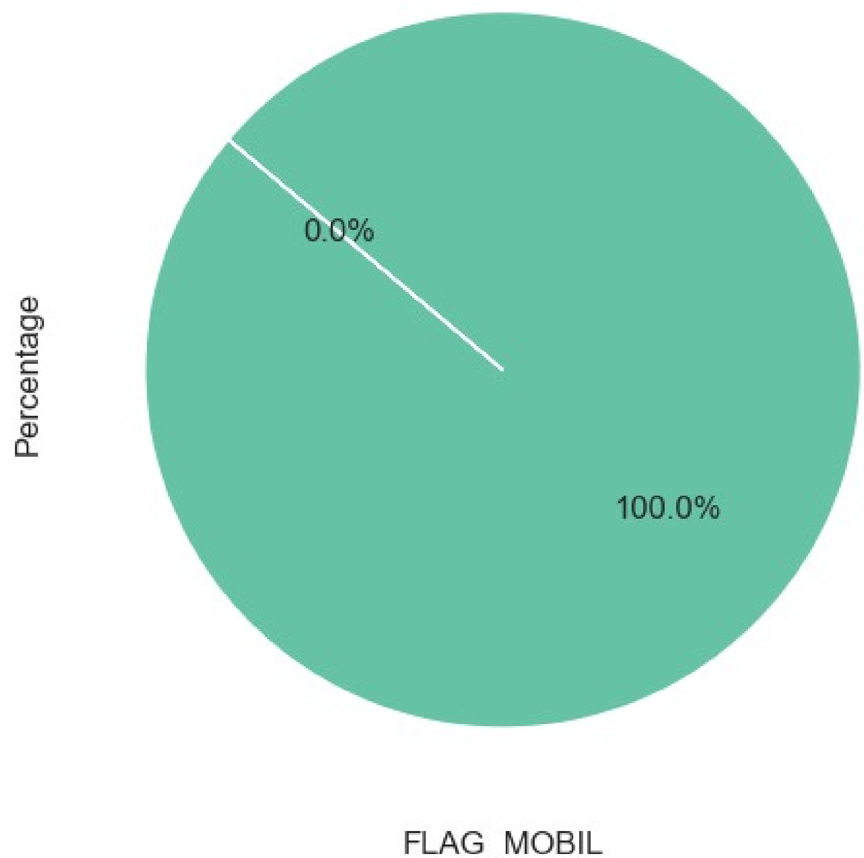

```
Married                   32094
Single / not married       7306
Civil marriage             4859
Separated                  3142
Widow                      2597
Unknown                       1
Name: NAME_FAMILY_STATUS, dtype: int64
[64. 15. 10.  6.  5.  0.]
```



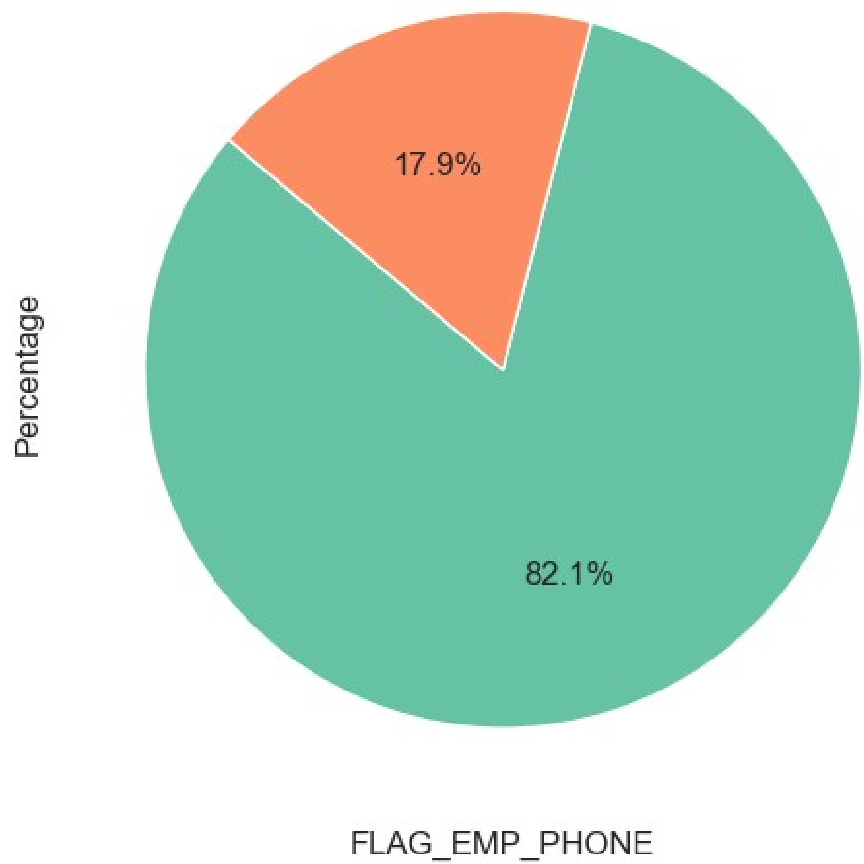

```
House / apartment      44368
With parents            2399
Municipal apartment     1845
Rented apartment         769
Office apartment         427
Co-op apartment          191
Name: NAME_HOUSING_TYPE, dtype: int64
[89.  5.  4.  2.  1.  0.]
```



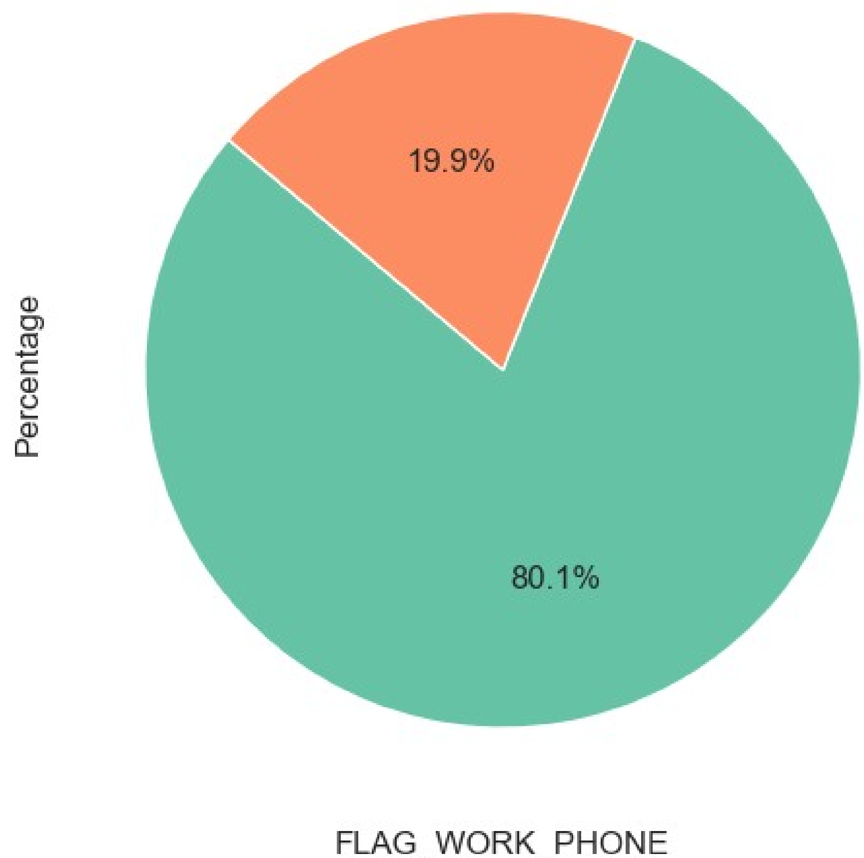

```
1    49998
0        1
Name: FLAG_MOBIL, dtype: int64
[100.   0.]
```



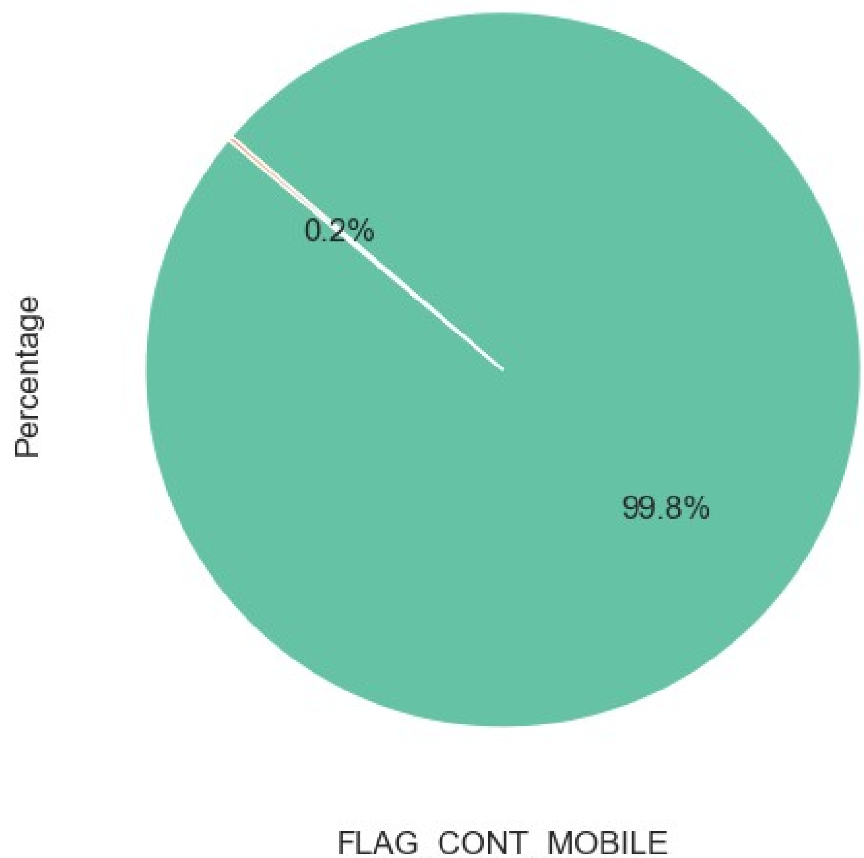

```
1    41073
0     8926
Name: FLAG_EMP_PHONE, dtype: int64
[82. 18.]
```



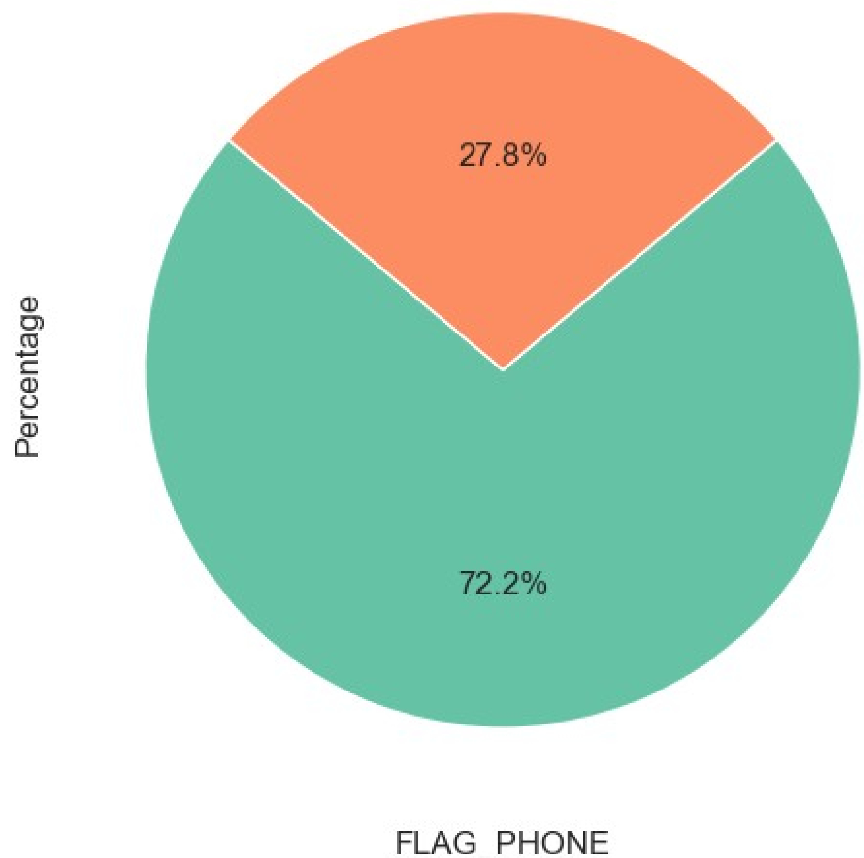

```
0    40036
1     9963
Name: FLAG_WORK_PHONE, dtype: int64
[80. 20.]
```



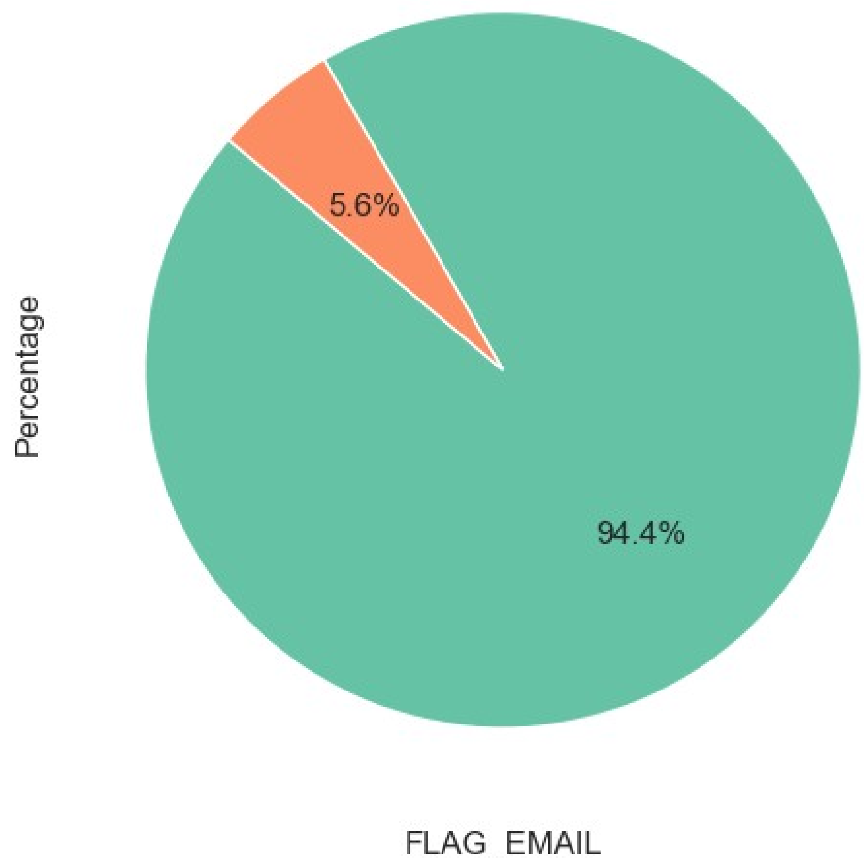

```
1    49898
0      101
Name: FLAG_CONT_MOBILE, dtype: int64
[100.   0.]
```



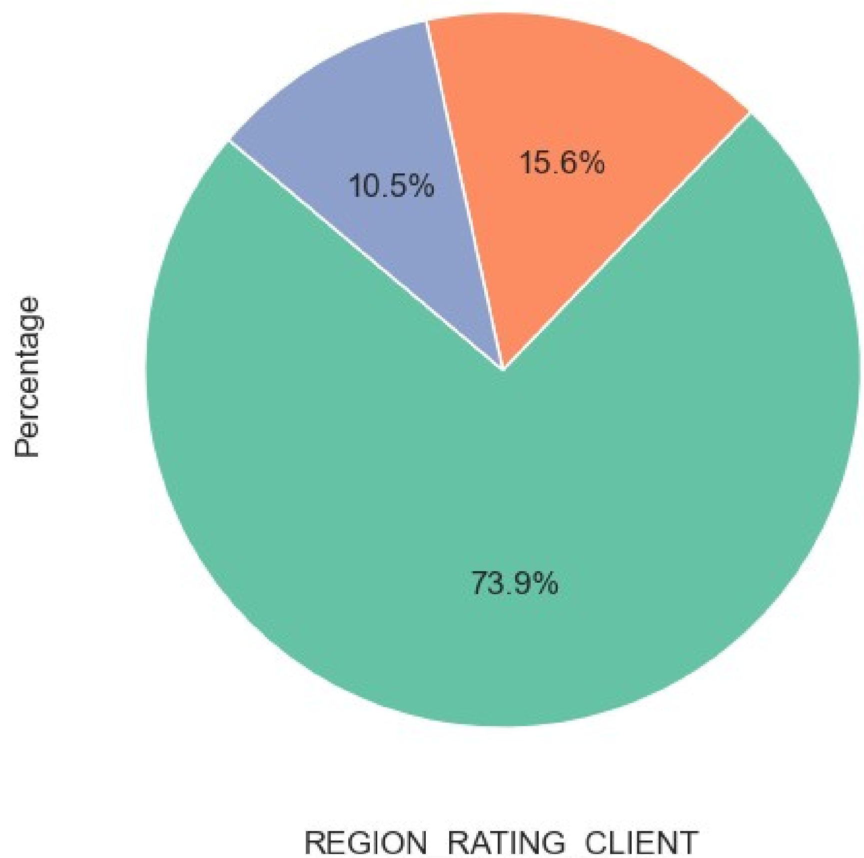

```
0    36113
1    13886
Name: FLAG_PHONE, dtype: int64
[72. 28.]
```



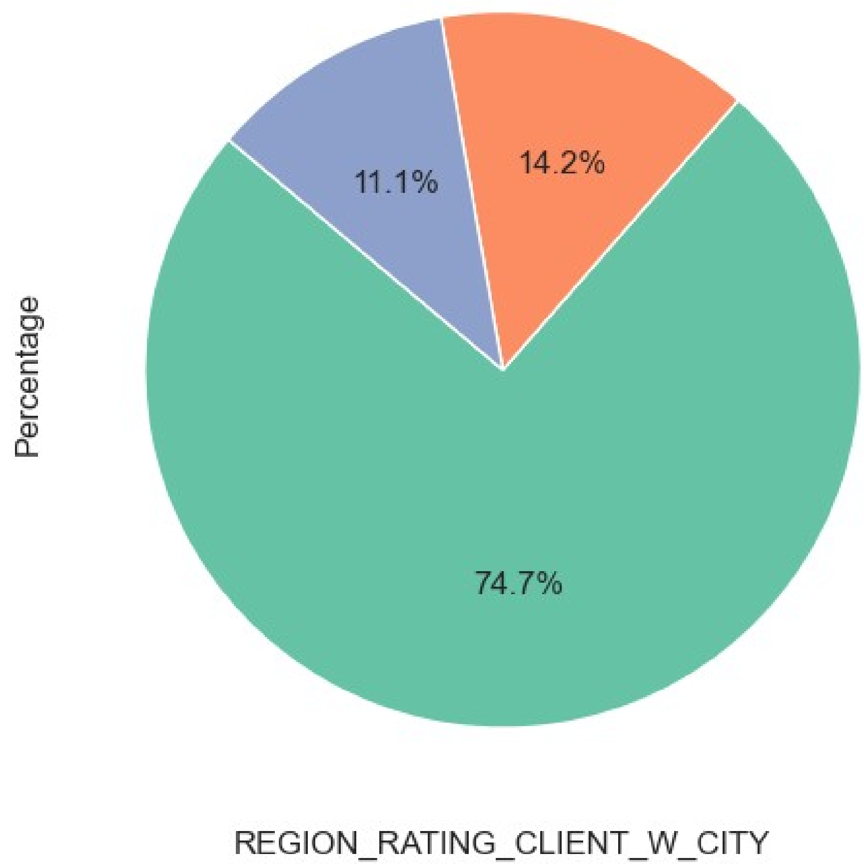

```
0    47216
1     2783
Name: FLAG_EMAIL, dtype: int64
[94.  6.]
```



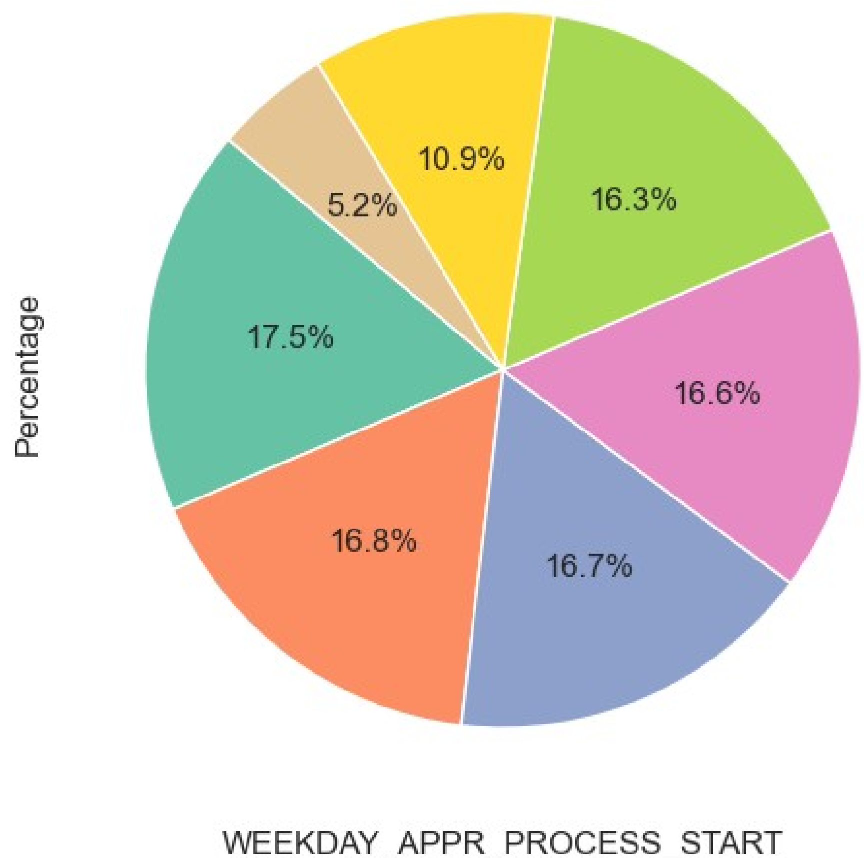

```
2    36964
3     7809
1     5226
Name: REGION_RATING_CLIENT, dtype: int64
[74. 16. 10.]
```



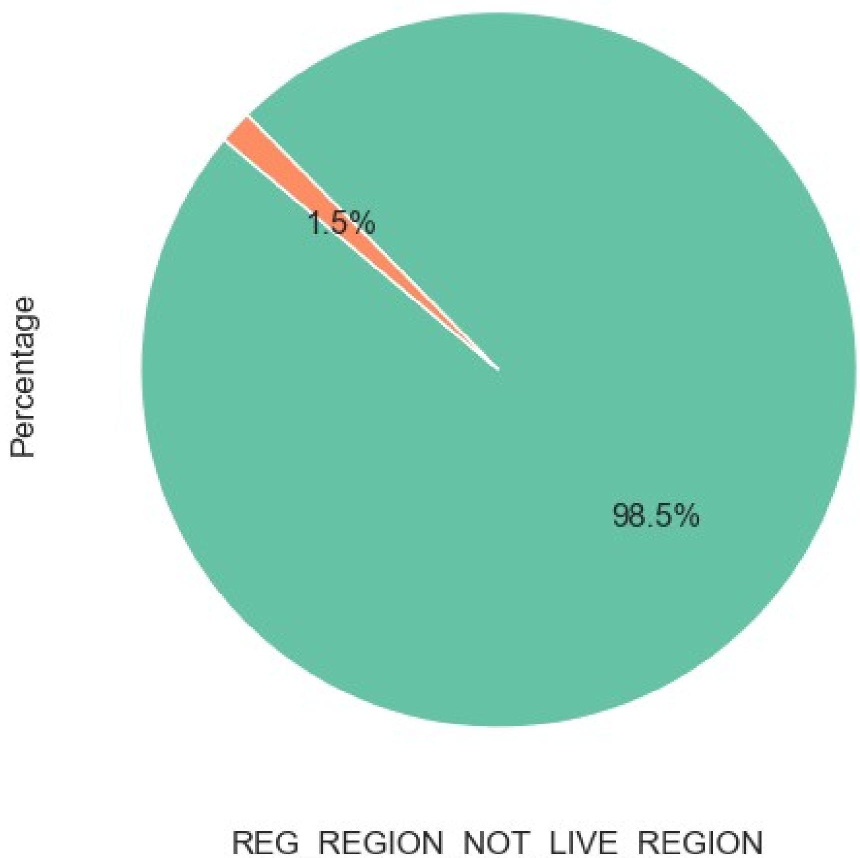

```
2    37341
3     7097
1     5561
Name: REGION_RATING_CLIENT_W_CITY, dtype: int64
[75. 14. 11.]
```



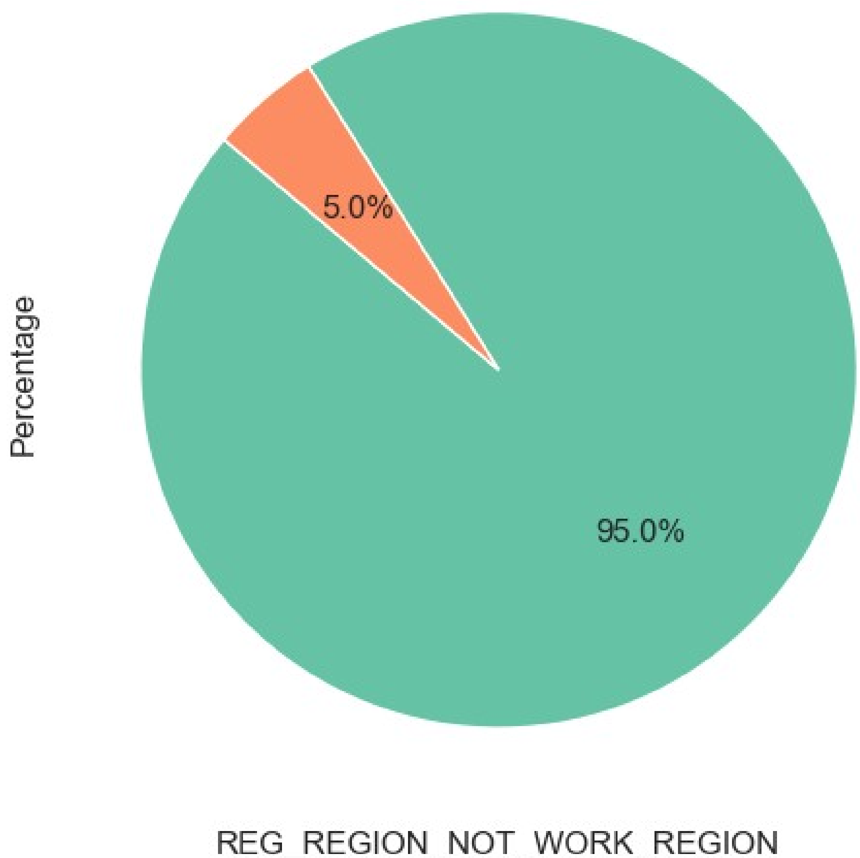

```
TUESDAY      8741
MONDAY       8385
WEDNESDAY    8355
FRIDAY       8286
THURSDAY     8149
SATURDAY     5467
SUNDAY       2616
Name: WEEKDAY_APPR_PROCESS_START, dtype: int64
[17. 17. 17. 17. 16. 11.  5.]
```



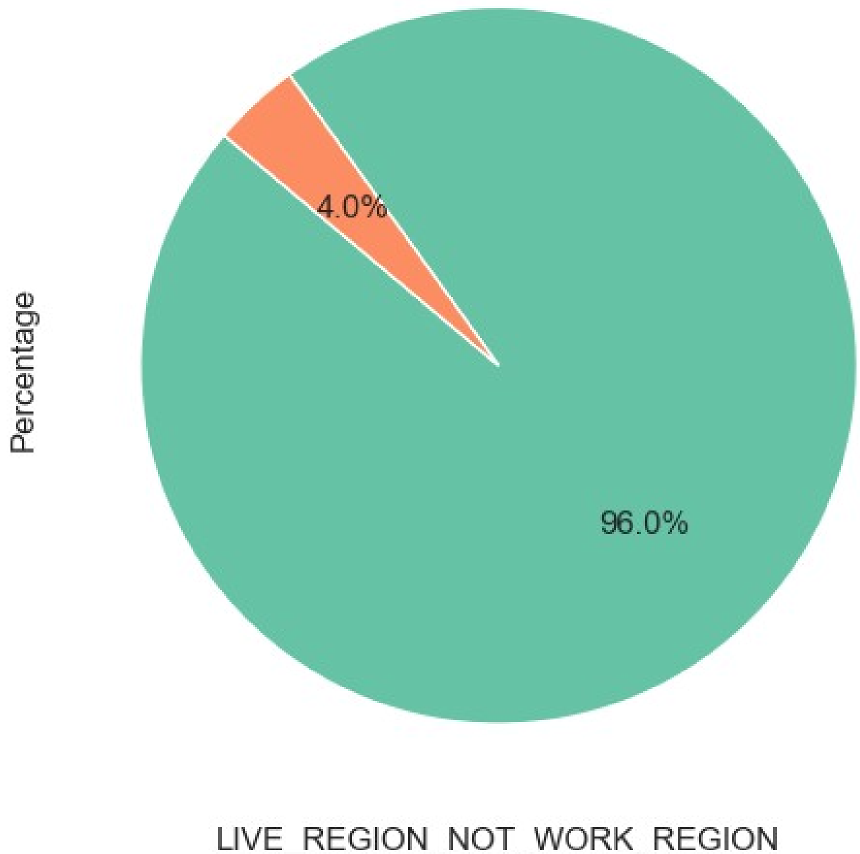

```
0    49249
1      750
Name: REG_REGION_NOT_LIVE_REGION, dtype: int64
[98.  2.]
```



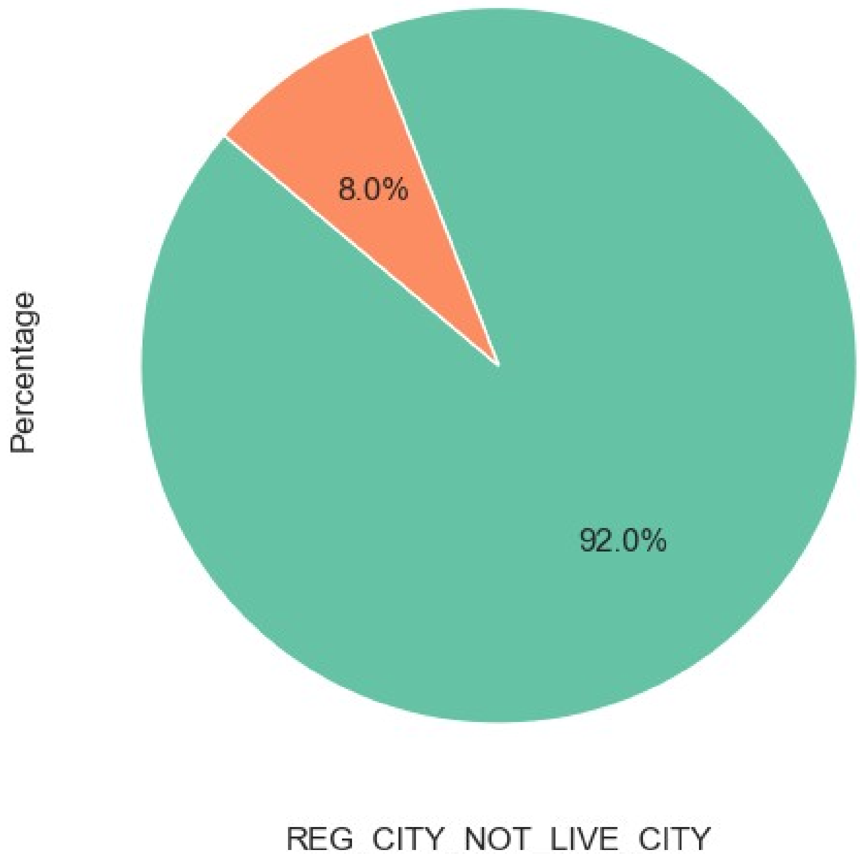

```
0    47503
1     2496
Name: REG_REGION_NOT_WORK_REGION, dtype: int64
[95.  5.]
```



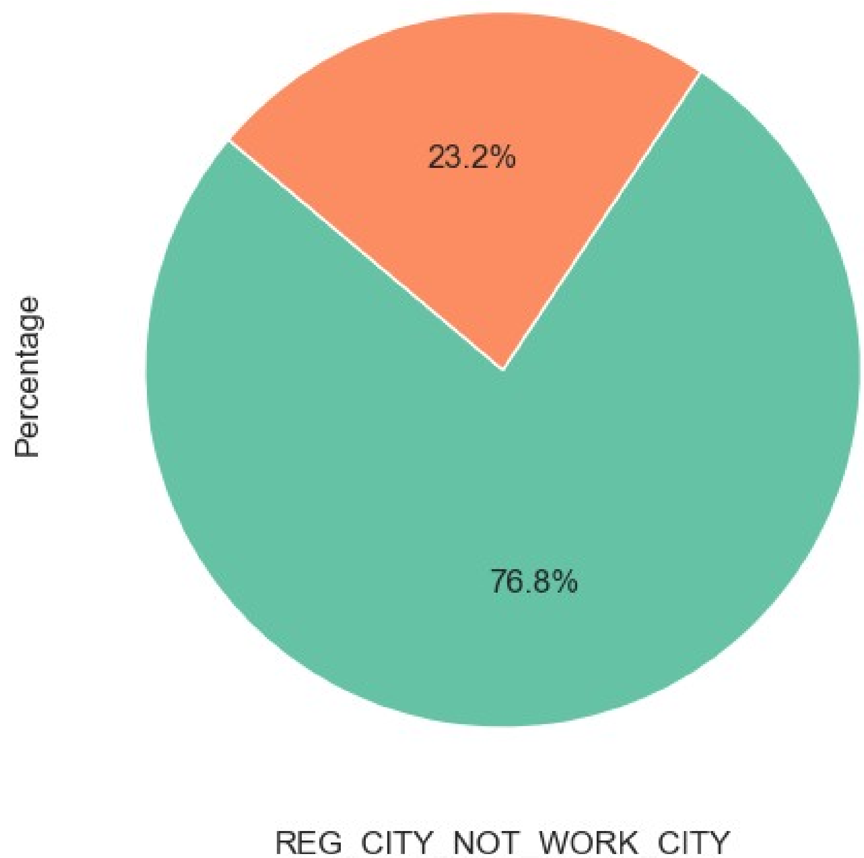

```
0    48017
1     1982
Name: LIVE_REGION_NOT_WORK_REGION, dtype: int64
[96.  4.]
```



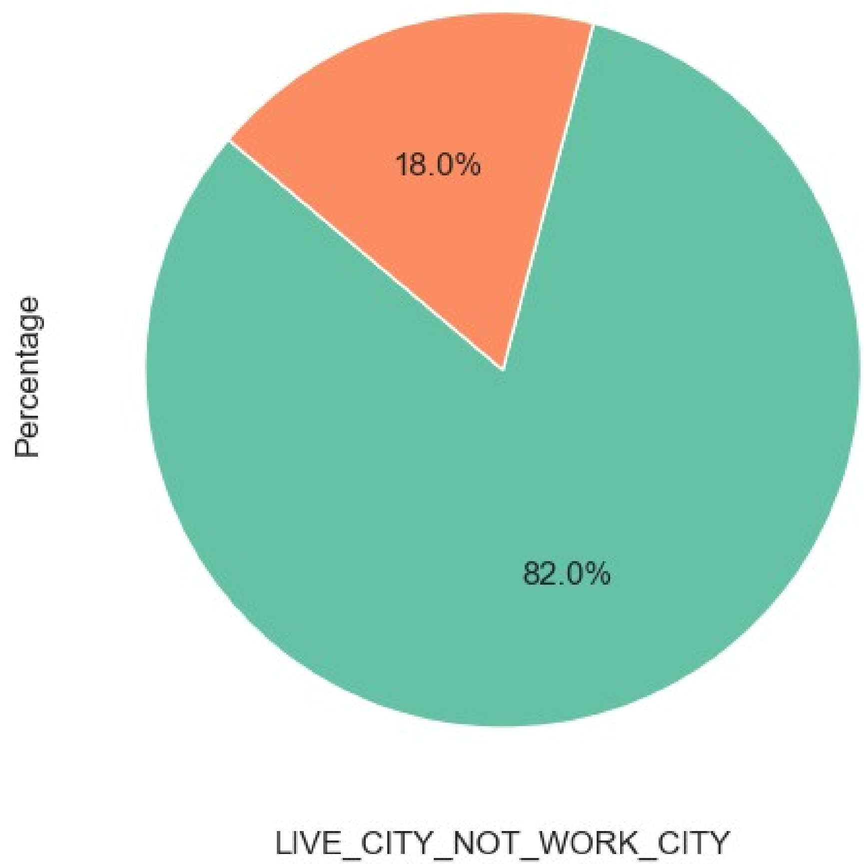

```
0    46001
1     3998
Name: REG_CITY_NOT_LIVE_CITY, dtype: int64
[92.  8.]
```



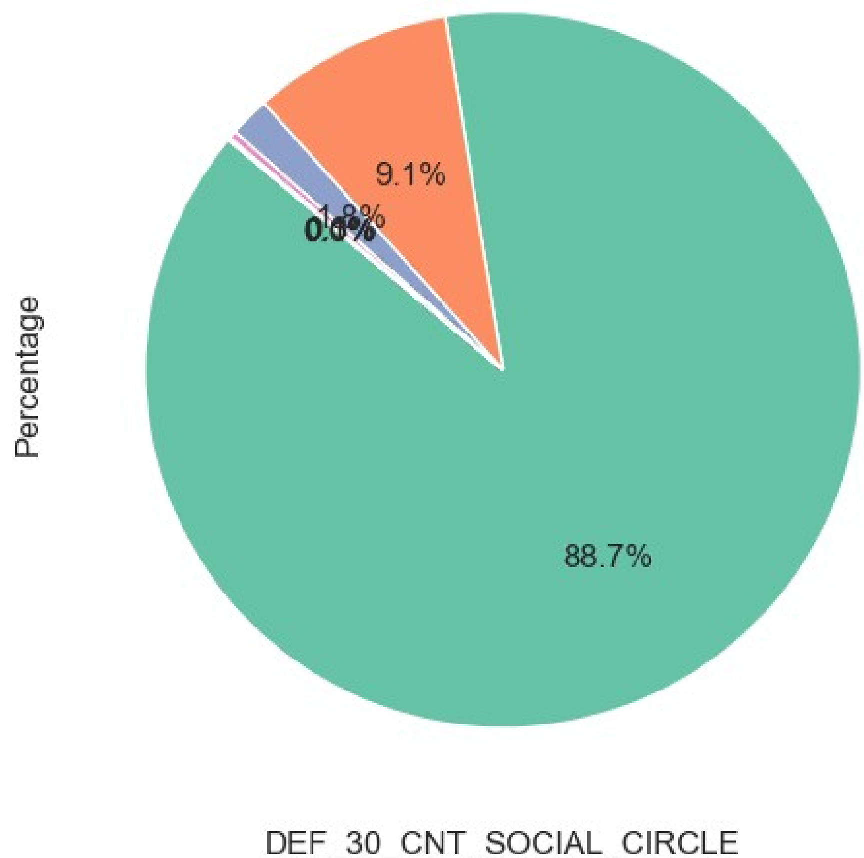

```
0    38391
1    11608
Name: REG_CITY_NOT_WORK_CITY, dtype: int64
[77. 23.]
```



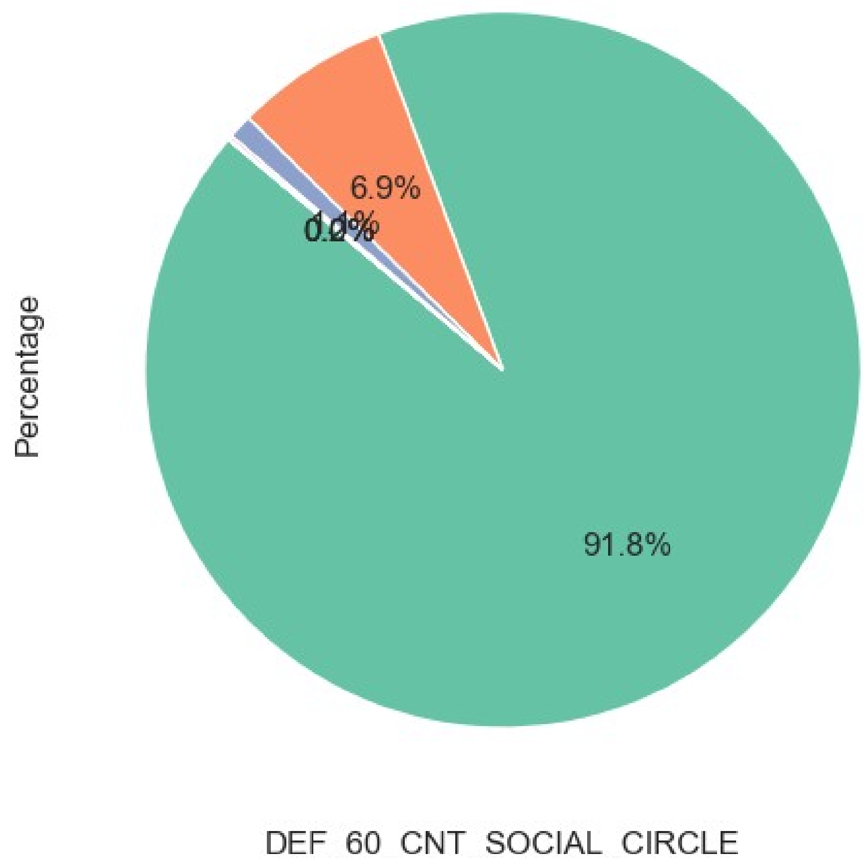

```
0    41014
1     8985
Name: LIVE_CITY_NOT_WORK_CITY, dtype: int64
[82. 18.]
```



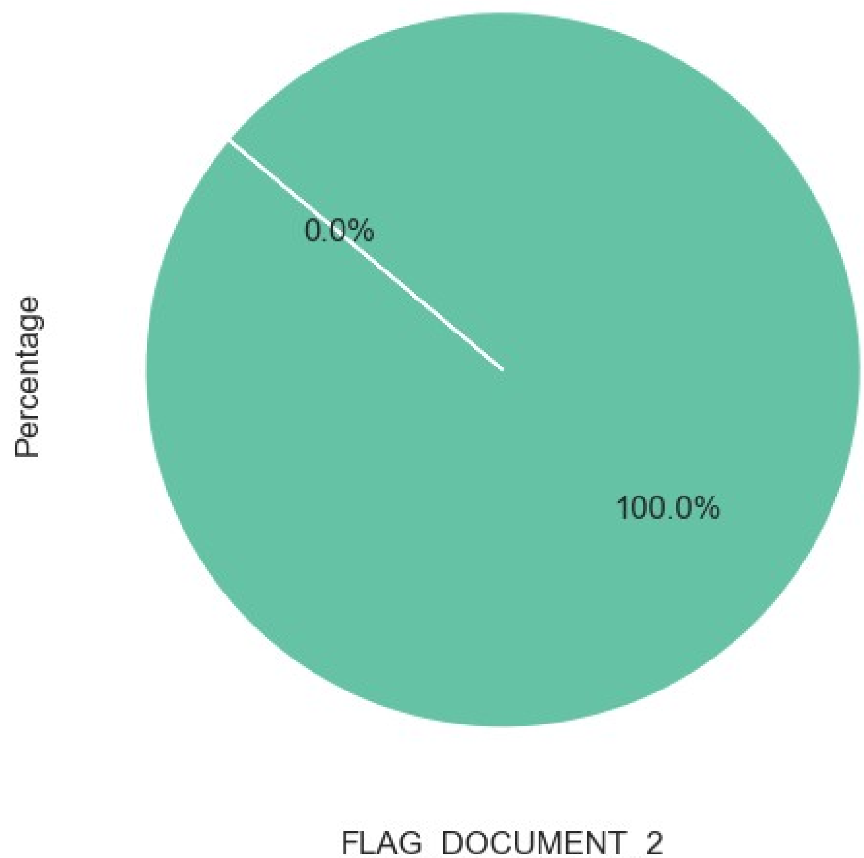

```
0.0    44189
1.0     4514
2.0      899
3.0      173
4.0       45
5.0       10
6.0        1
Name: DEF_30_CNT_SOCIAL_CIRCLE, dtype: int64
[88.  9.  2.  0.  0.  0.  0.]
```



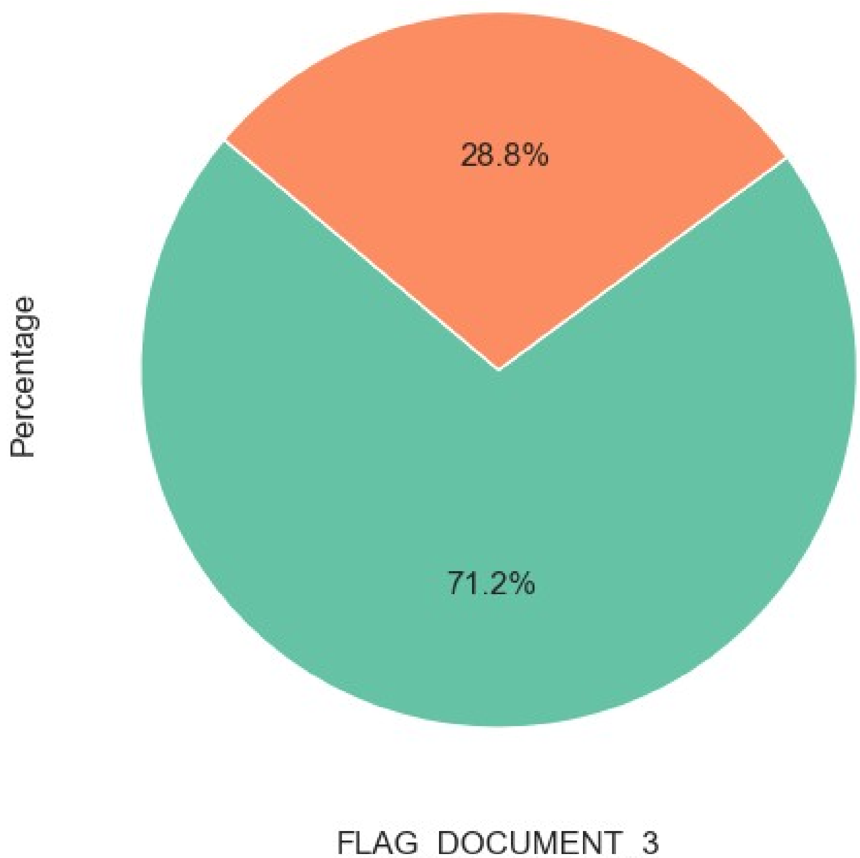

```
0.0    45723
1.0     3457
2.0      543
3.0       80
4.0       23
5.0        5
Name: DEF_60_CNT_SOCIAL_CIRCLE, dtype: int64
[91.  7.  1.  0.  0.  0.]
```



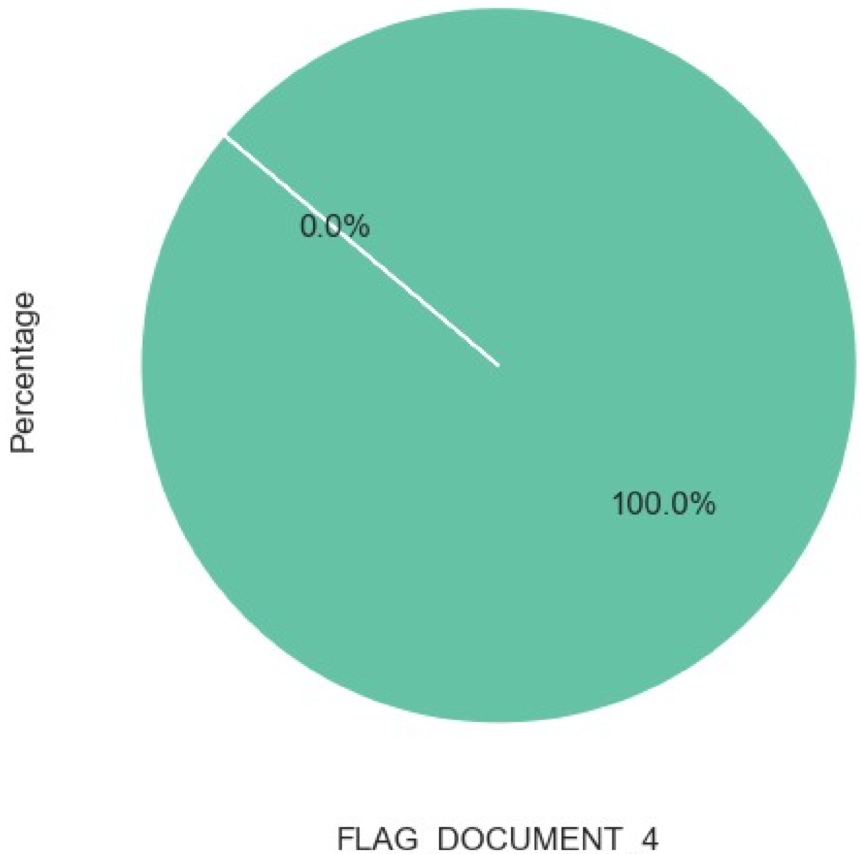

```
0    49997
1        2
Name: FLAG_DOCUMENT_2, dtype: int64
[100.   0.]
```



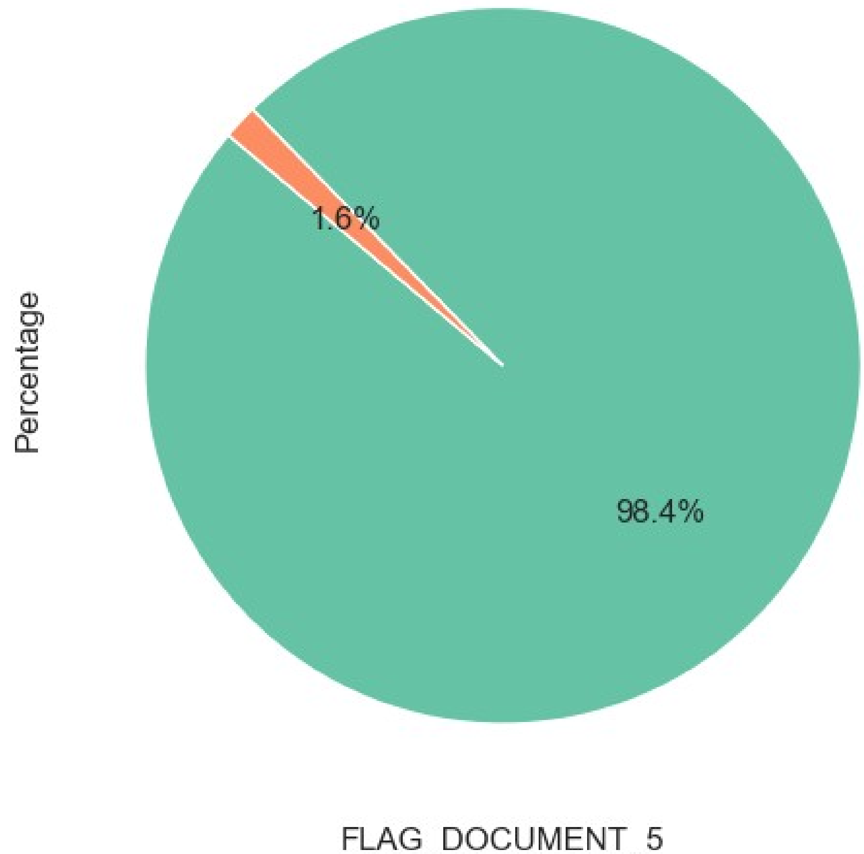

```
1    35612
0    14387
Name: FLAG_DOCUMENT_3, dtype: int64
[71. 29.]
```



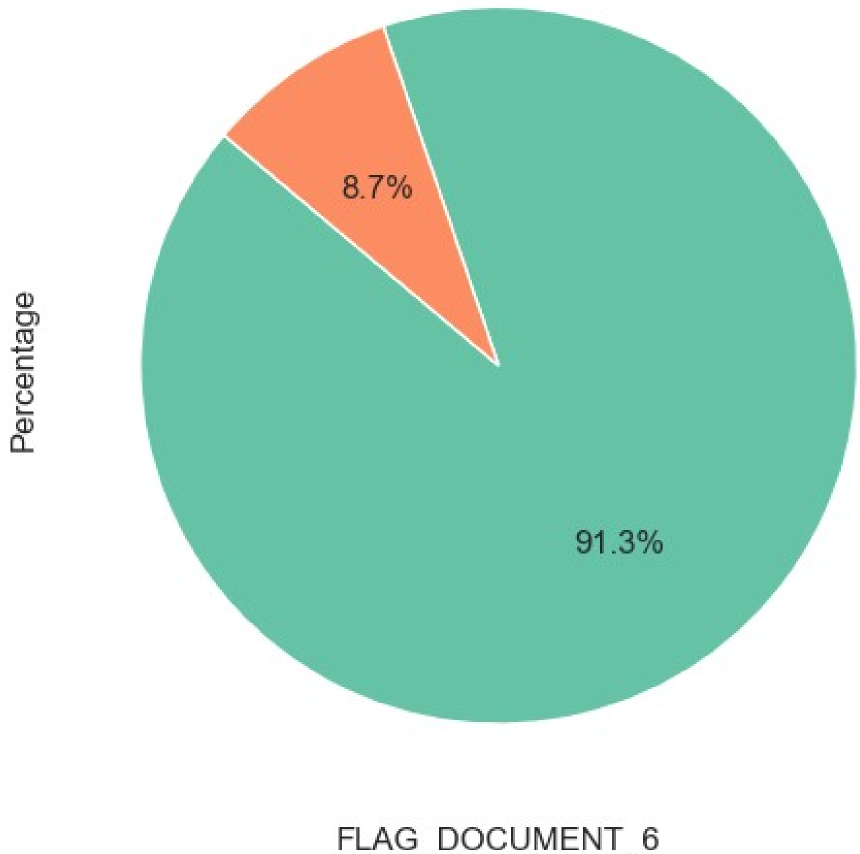

```
0    49990
1        9
Name: FLAG_DOCUMENT_4, dtype: int64
[100.   0.]
```



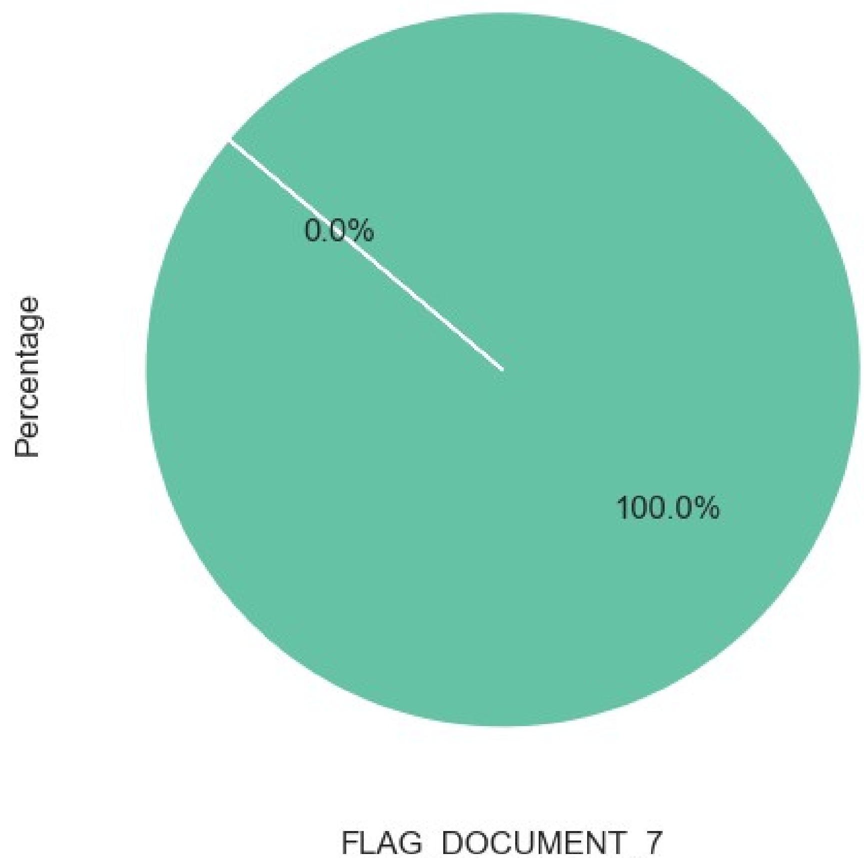

```
0    49214
1      785
Name: FLAG_DOCUMENT_5, dtype: int64
[98.  2.]
```



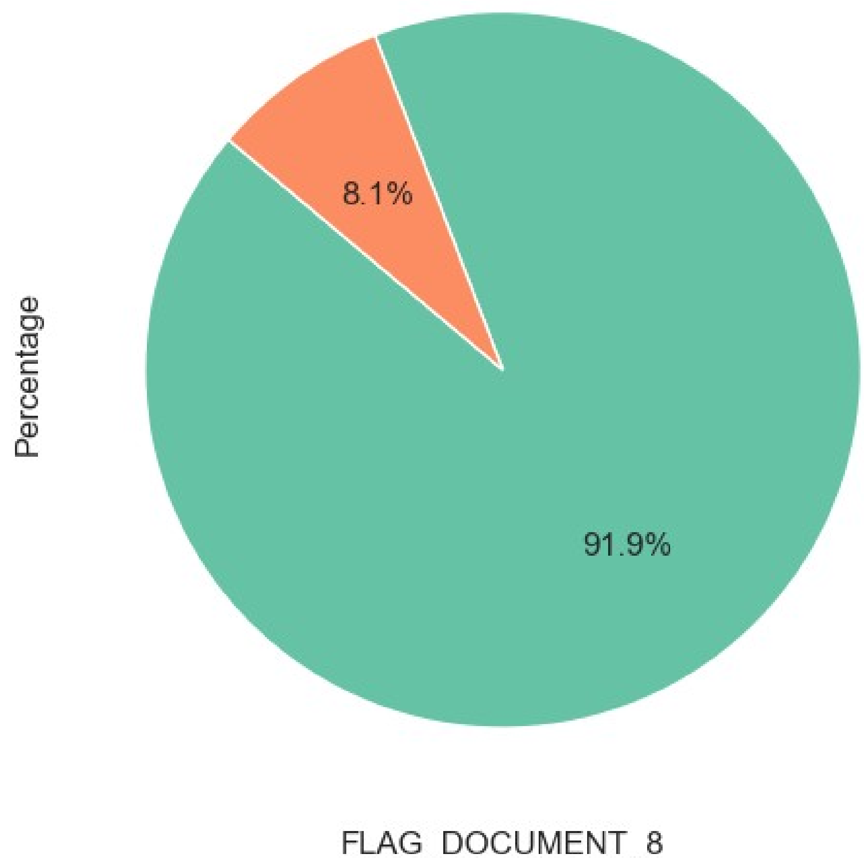

```
0    45664
1     4335
Name: FLAG_DOCUMENT_6, dtype: int64
[91.  9.]
```



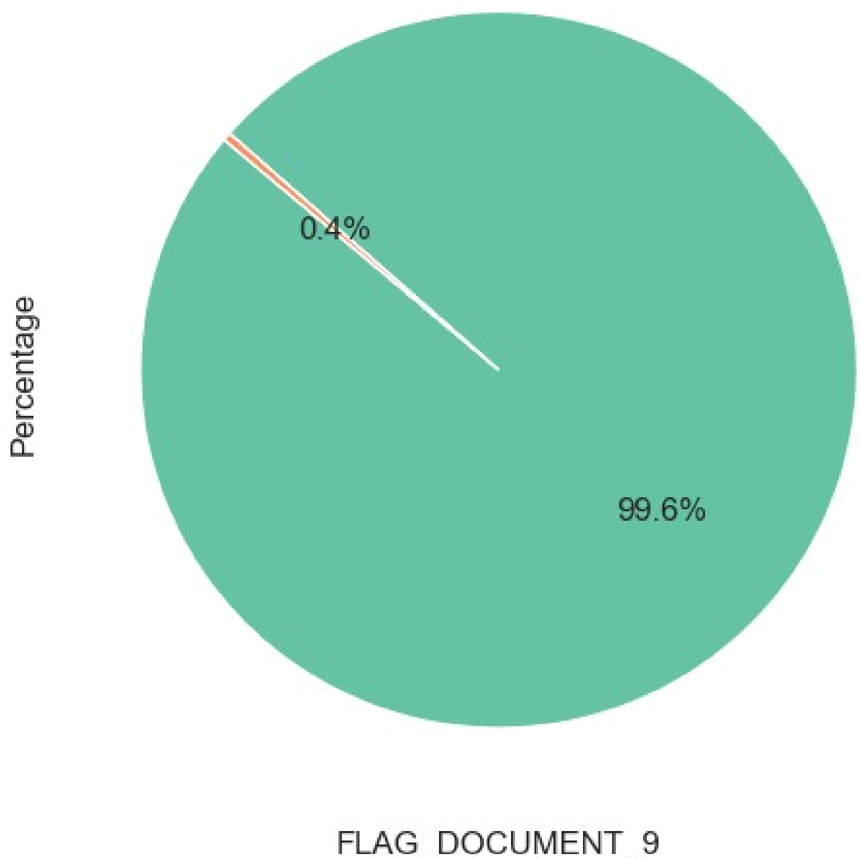

```
0    49988
1       11
Name: FLAG_DOCUMENT_7, dtype: int64
[100.   0.]
```



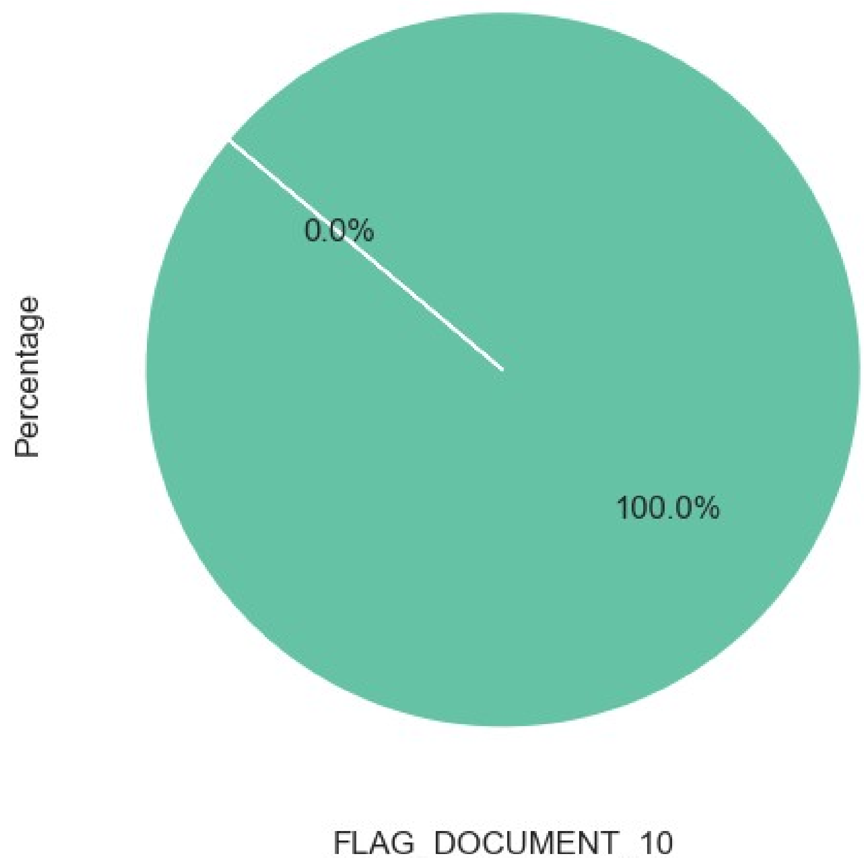

```
0    45961
1     4038
Name: FLAG_DOCUMENT_8, dtype: int64
[92.  8.]
```



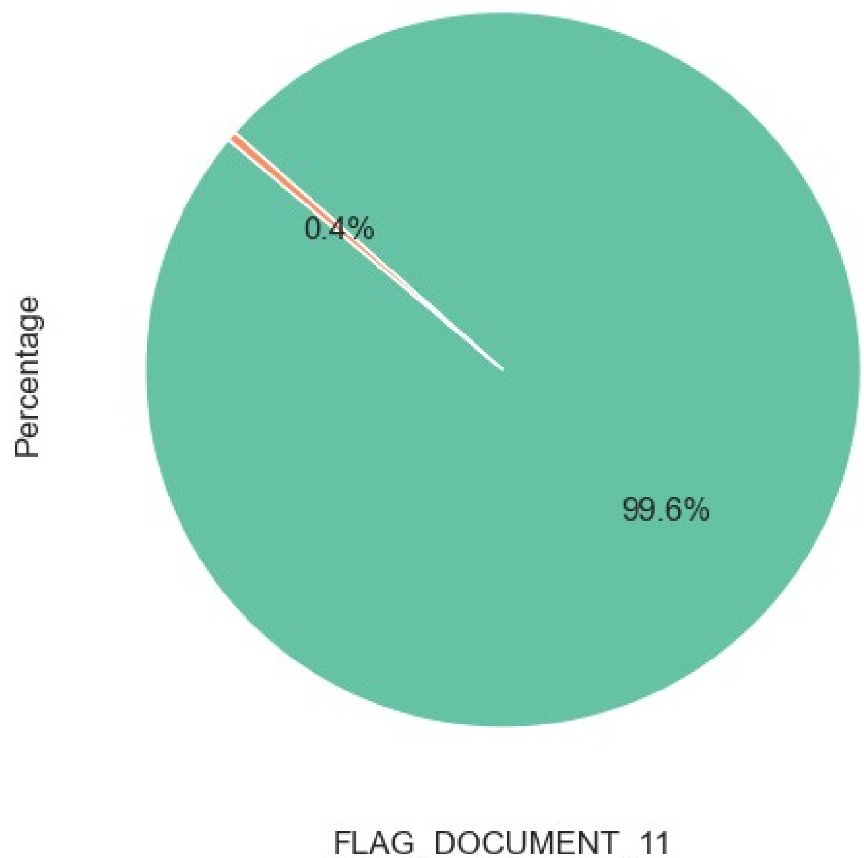

```
0    49815
1      184
Name: FLAG_DOCUMENT_9, dtype: int64
[100.   0.]
```



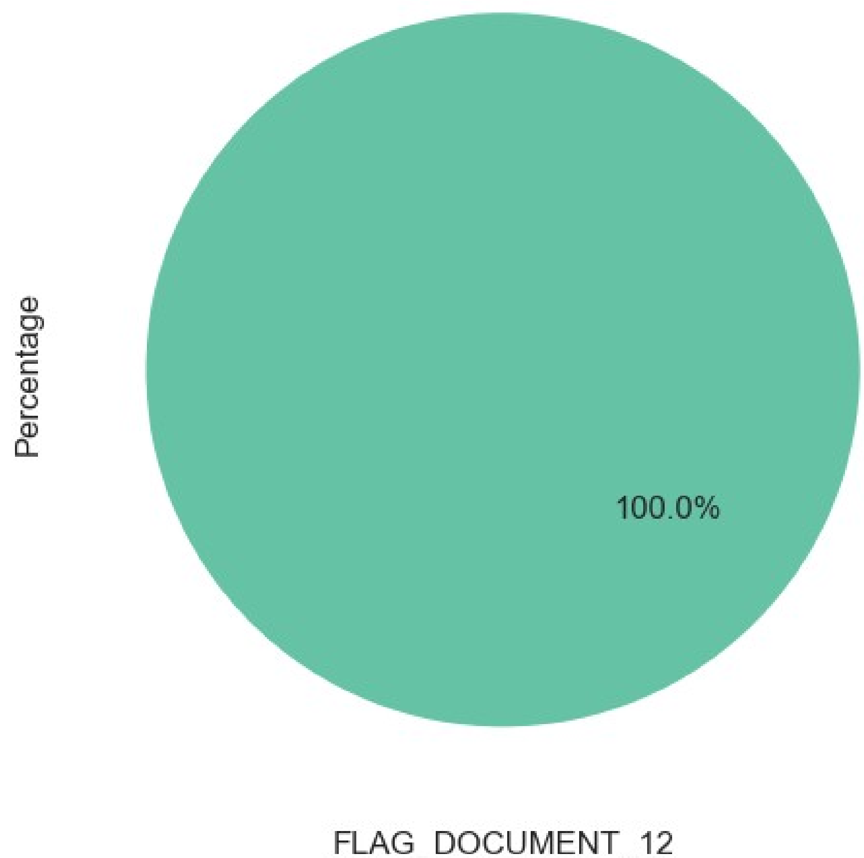

```
0    49998
1        1
Name: FLAG_DOCUMENT_10, dtype: int64
[100.   0.]
```



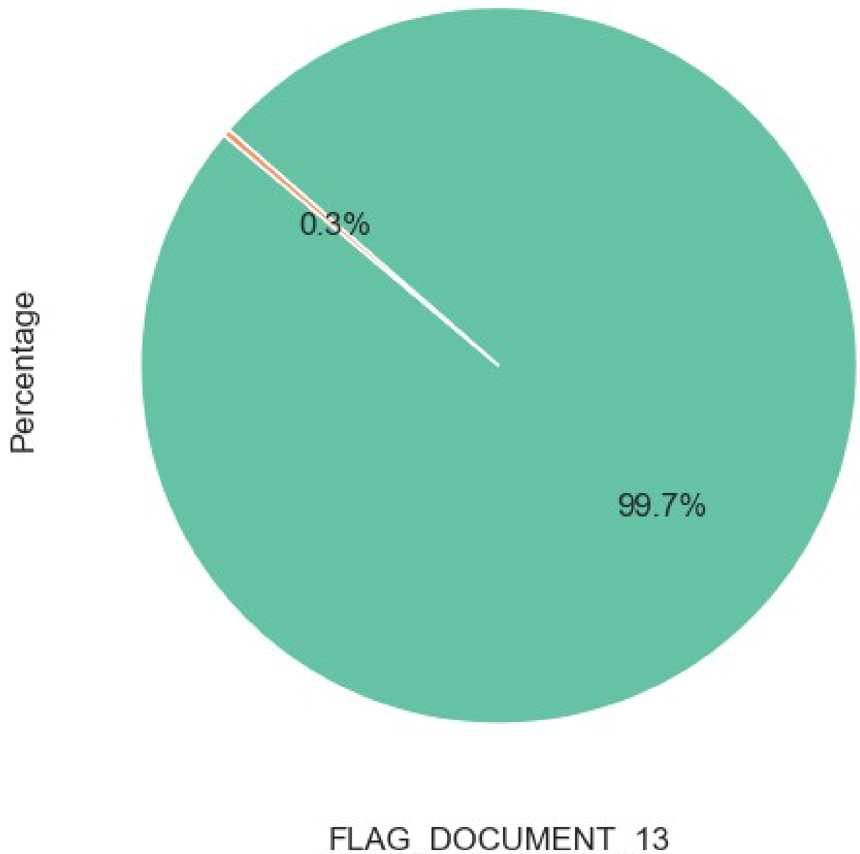

```
0    49786
1      213
Name: FLAG_DOCUMENT_11, dtype: int64
[100.   0.]
```



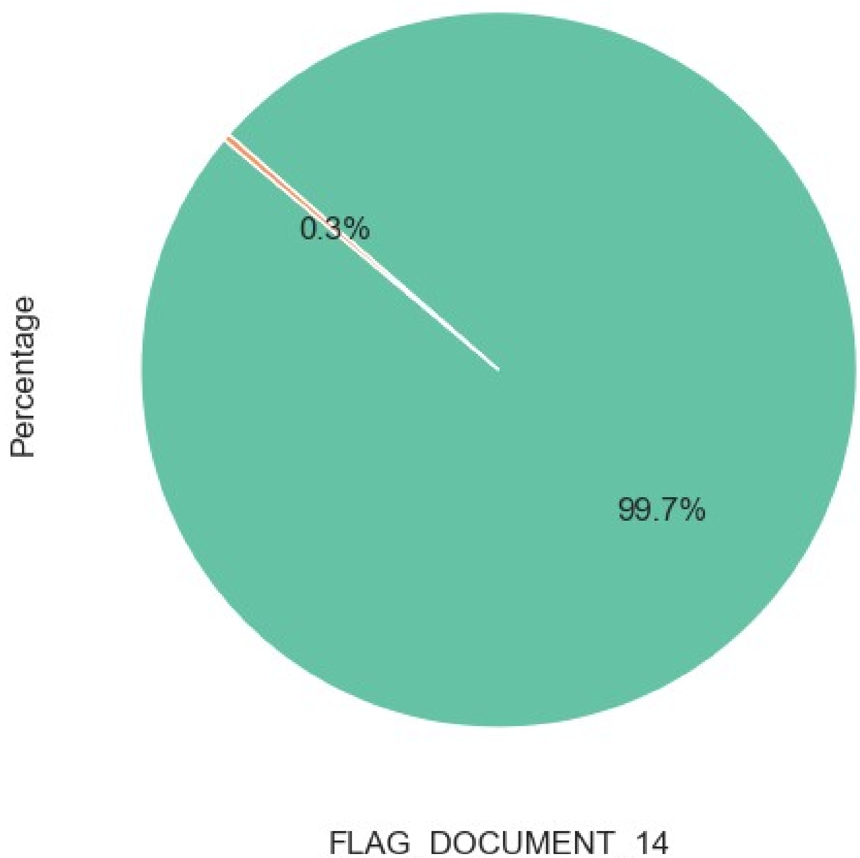

```
0    49999
Name: FLAG_DOCUMENT_12, dtype: int64
[100.]
```



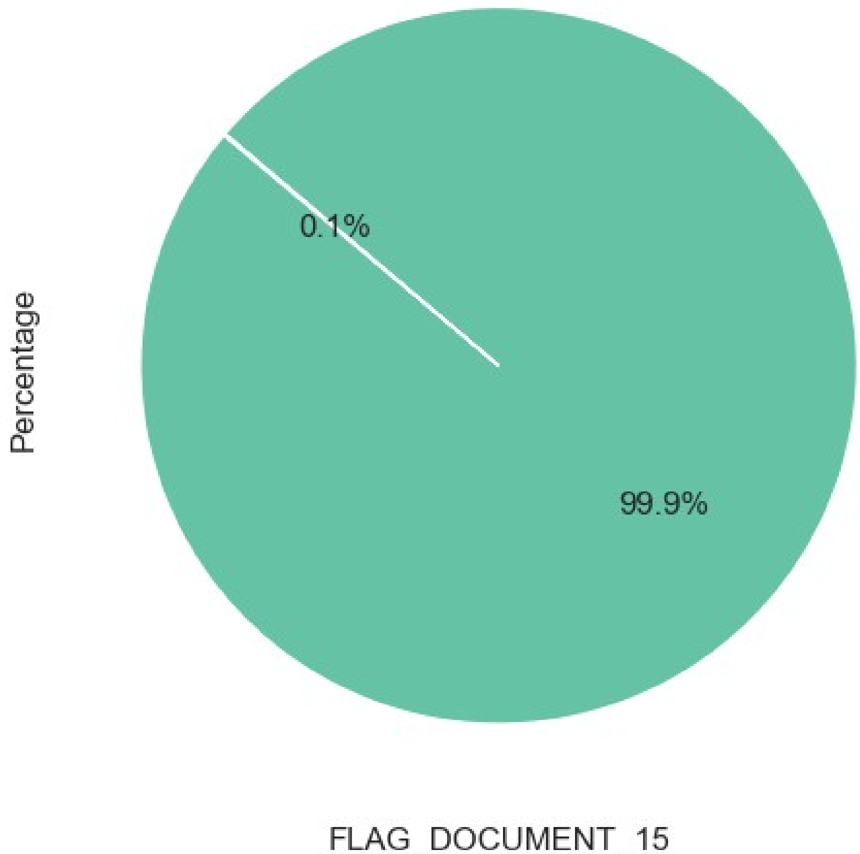

```
0    49838
1      161
Name: FLAG_DOCUMENT_13, dtype: int64
[100.   0.]
```



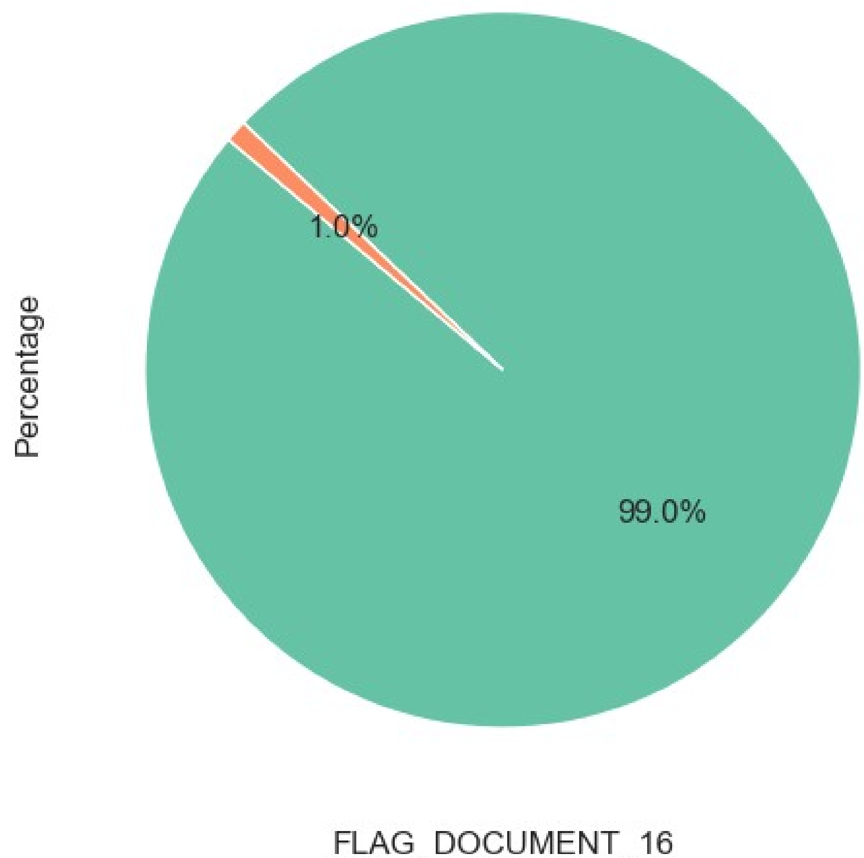

```
0    49841
1      158
Name: FLAG_DOCUMENT_14, dtype: int64
[100.   0.]
```



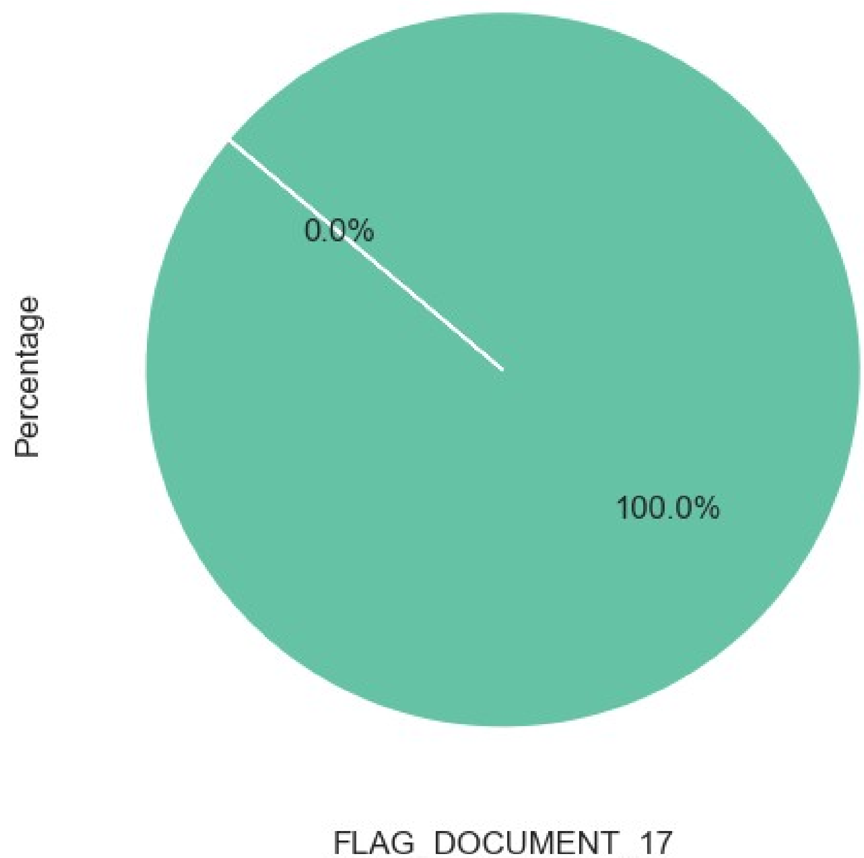

```
0    49958
1       41
Name: FLAG_DOCUMENT_15, dtype: int64
[100.   0.]
```



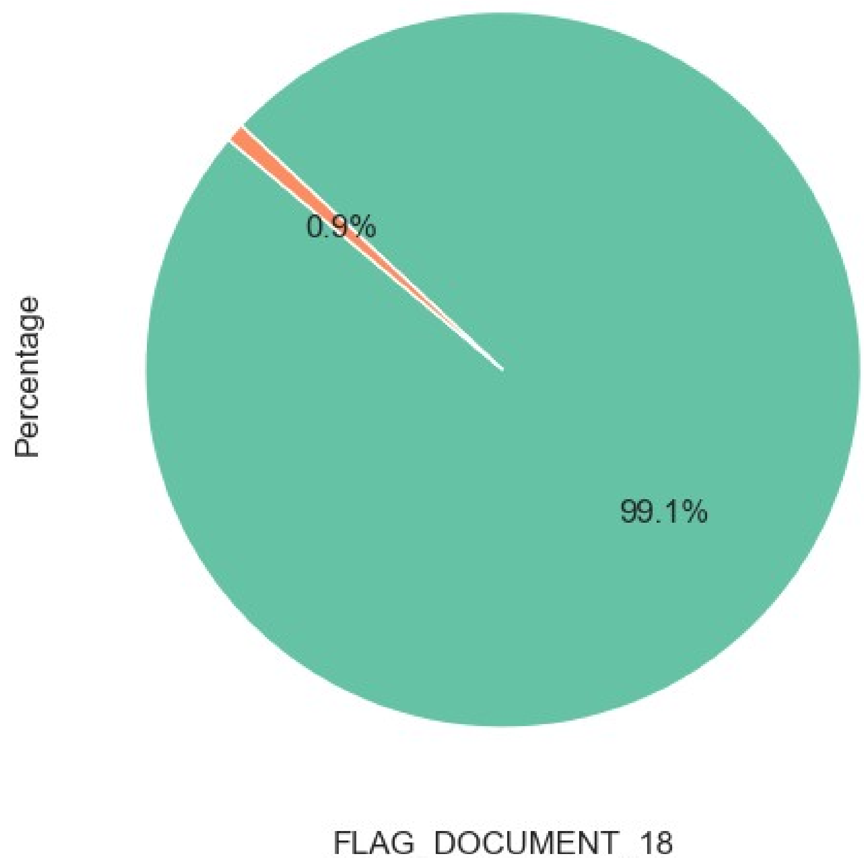

```
0    49498
1      501
Name: FLAG_DOCUMENT_16, dtype: int64
[99.  1.]
```



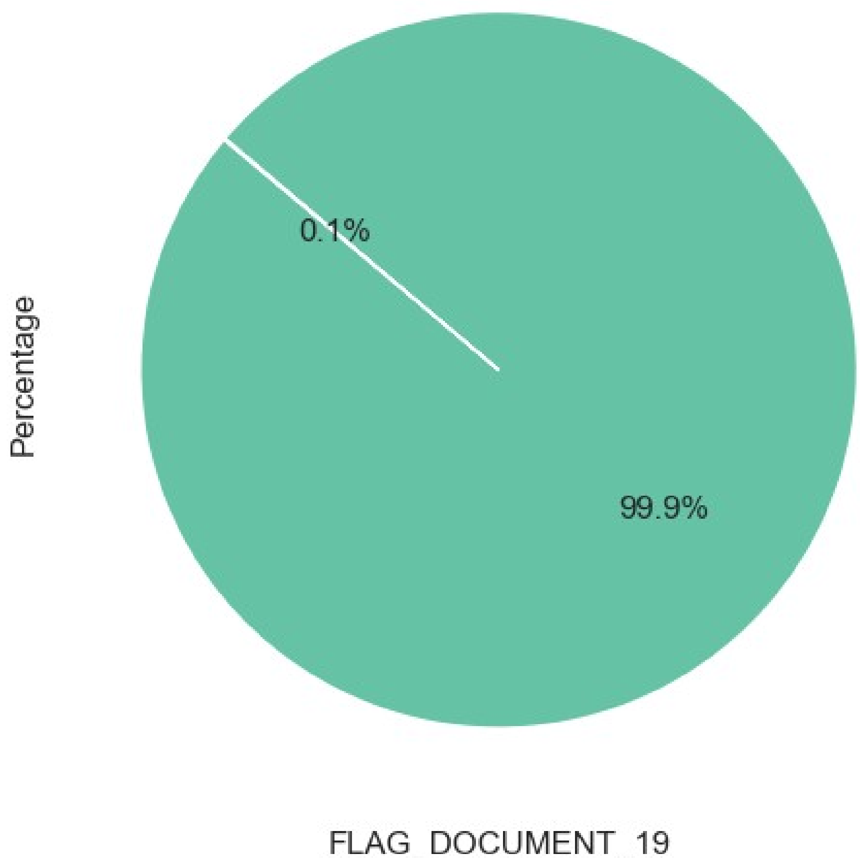

```
0    49984
1       15
Name: FLAG_DOCUMENT_17, dtype: int64
[100.   0.]
```



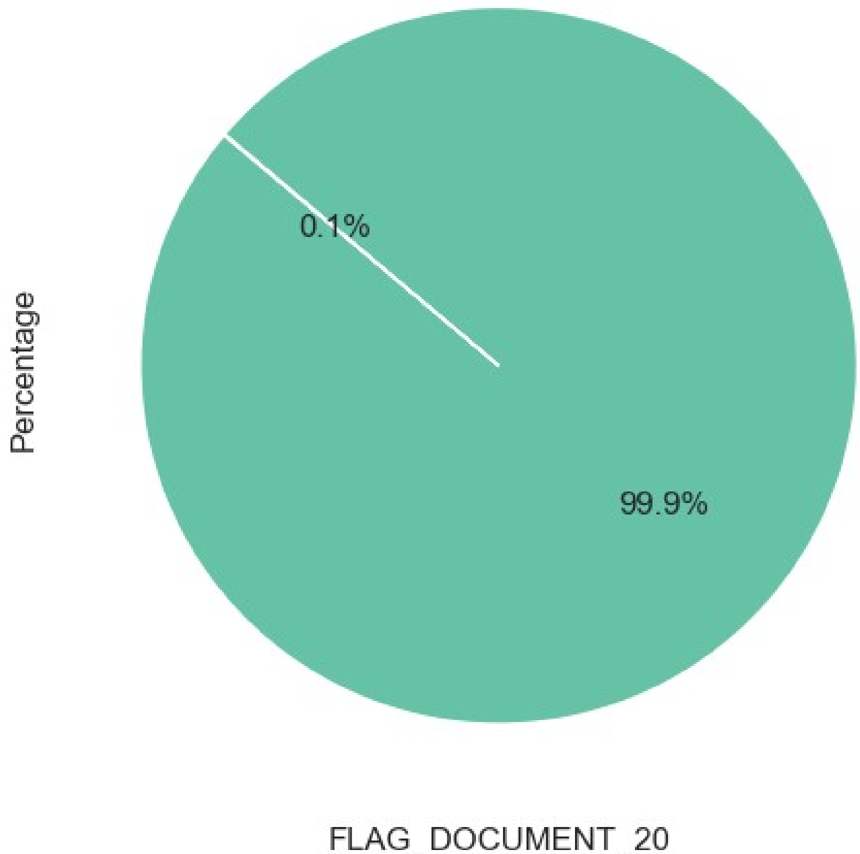

```
0    49574
1      425
Name: FLAG_DOCUMENT_18, dtype: int64
[99.  1.]
```



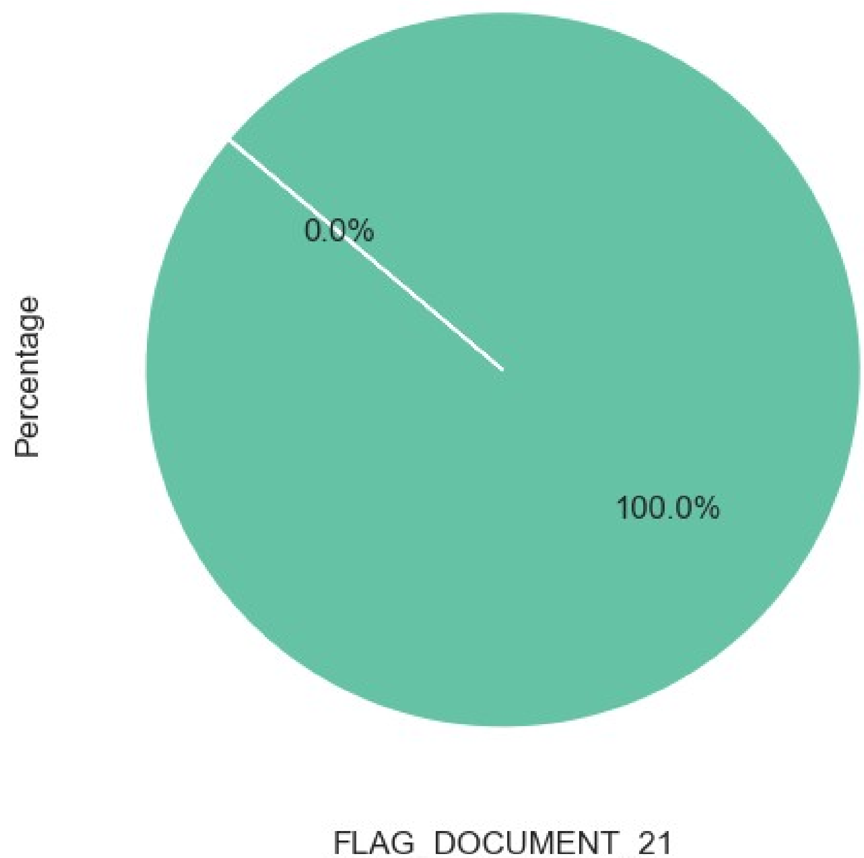

```
0    49964
1       35
Name: FLAG_DOCUMENT_19, dtype: int64
[100.   0.]
```



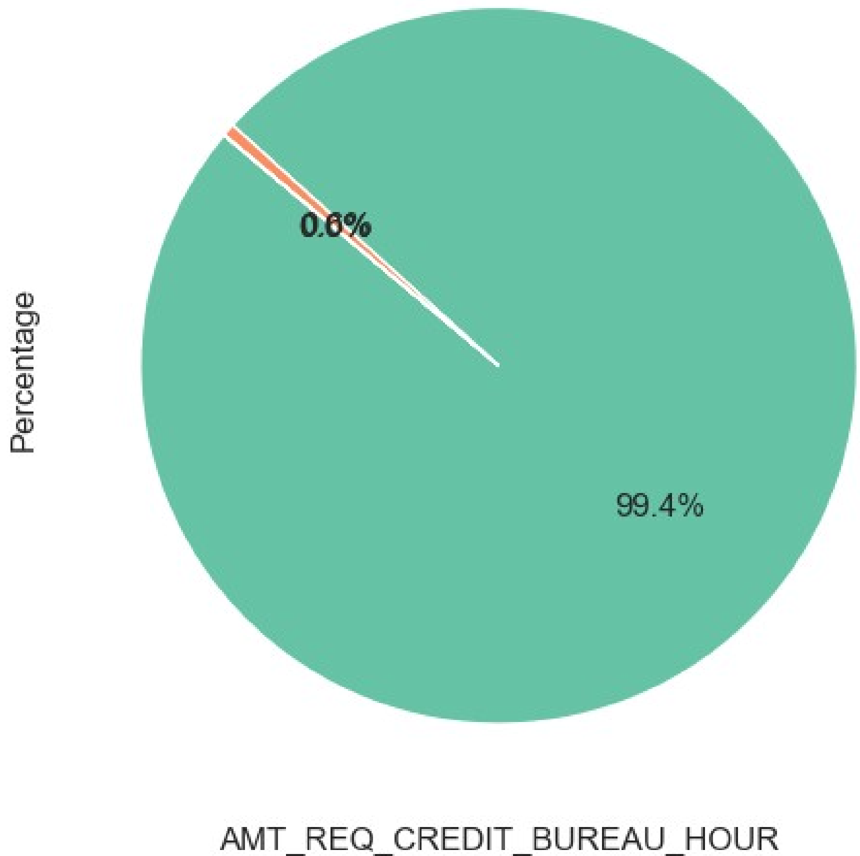

```
0    49973
1       26
Name: FLAG_DOCUMENT_20, dtype: int64
[100.   0.]
```



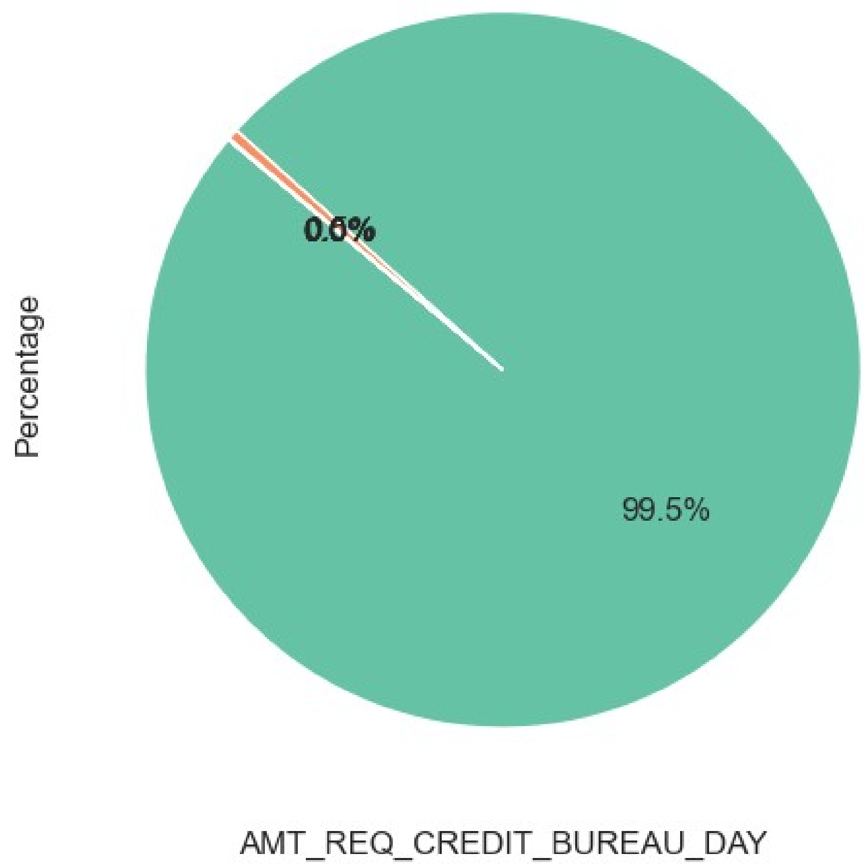

```
0    49980
1       19
Name: FLAG_DOCUMENT_21, dtype: int64
[100.   0.]
```



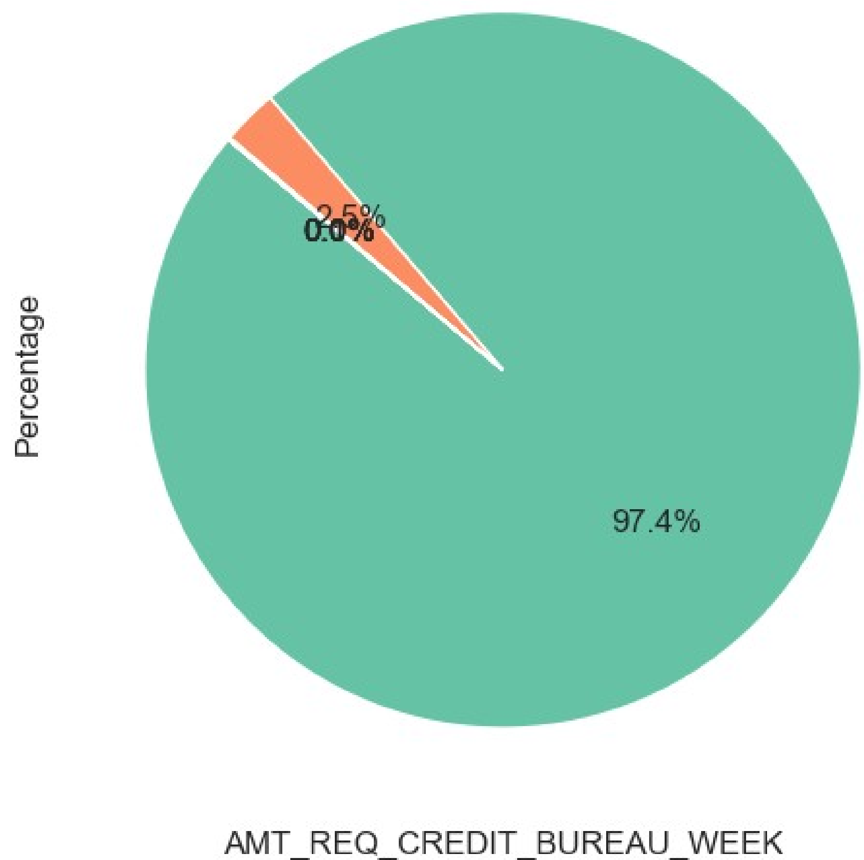

```
0.0    49704
1.0      285
2.0        8
3.0        2
Name: AMT_REQ_CREDIT_BUREAU_HOUR, dtype: int64
[99.  1.  0.  0.]
```



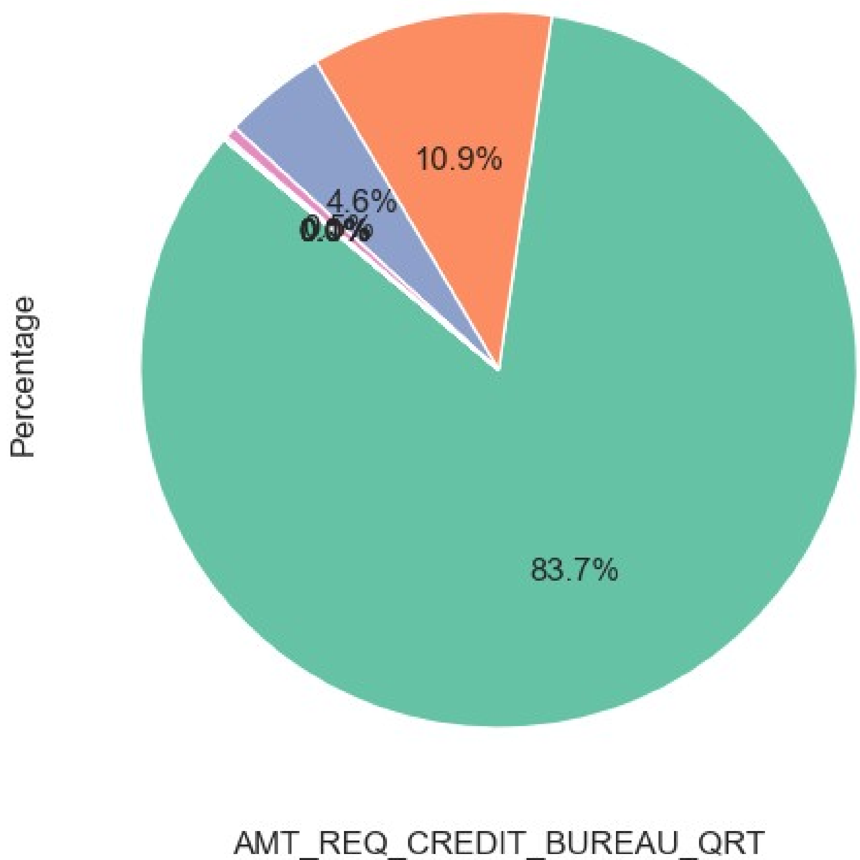

```
0.0    49727
1.0      242
2.0       17
3.0        7
4.0        3
5.0        2
6.0        1
Name: AMT_REQ_CREDIT_BUREAU_DAY, dtype: int64
[99.  0.  0.  0.  0.  0.  0.]
```




```
0.0    48685
1.0     1259
2.0       36
3.0       10
4.0        6
5.0        2
6.0        1
Name: AMT_REQ_CREDIT_BUREAU_WEEK, dtype: int64
[97.  3.  0.  0.  0.  0.  0.]
```




```
0.0    41865
1.0     5452
2.0     2324
3.0      269
4.0       74
5.0        8
6.0        3
8.0        2
7.0        2
Name: AMT_REQ_CREDIT_BUREAU_QRT, dtype: int64
[84. 11.  5.  1.  0.  0.  0.  0.  0.]
```


10.9%
4.6%
83.7%
Percentage
AMT_REQ_CREDIT_BUREAU_QRT

Inferences

target variable shows us that 92% percent people have the value 1 i.e. client with payment difficulties

NAME_CONTRACT_TYPE has the 91% of loans of Cash Loans type.

CODE_GENDER univariate analysis shows us that there are 66% female applicatns and 34% male applicants

FLAG_OWN_CAR univariate analysis shows us that 66% applicants don't own car.

FLAG_OWN_REALTY univariate analysis shows us that 69% applicants own house.

CNT_CHILDREN univariate analysis shows us that 70% applicants don't have childrens.

NAME_TYPE_SUITE univariate analysis shows us that 81% of the applicants are unaccompined

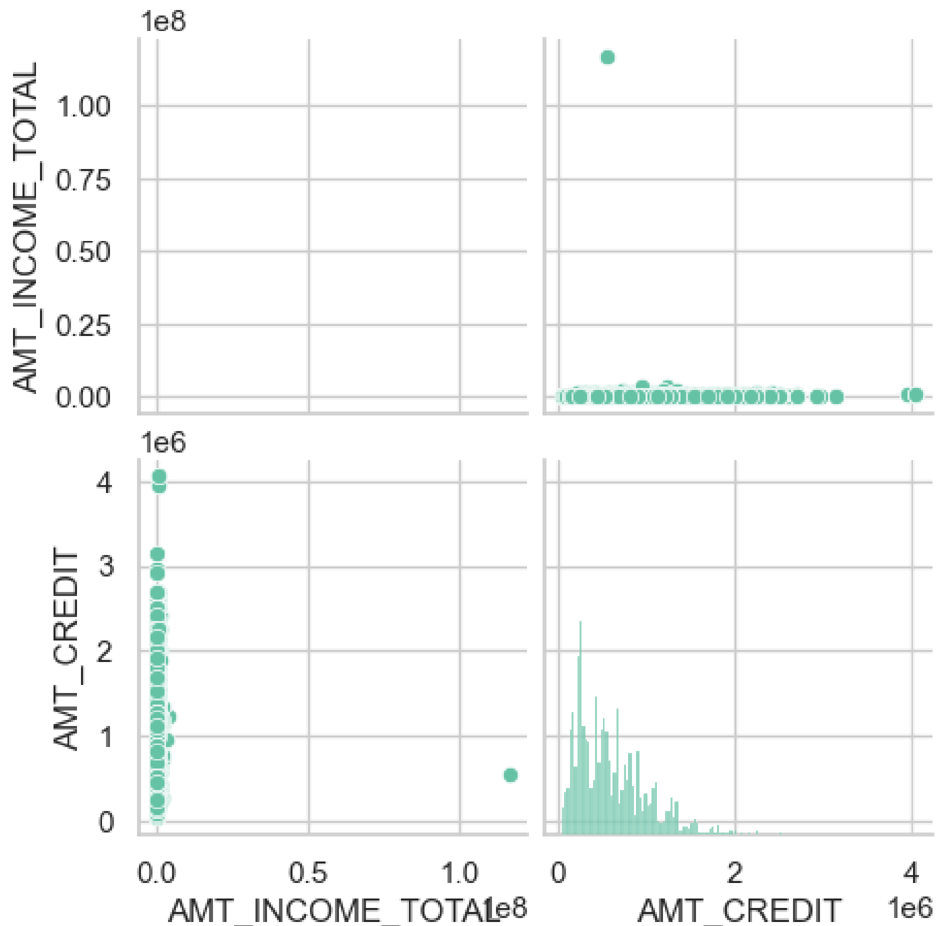
NAME_INCOME_TYPE univariate analysis shows us that 52% applicants are working.

NAME_FAMILY_STATUS univariate analysis shows us that 64% applicants are married.

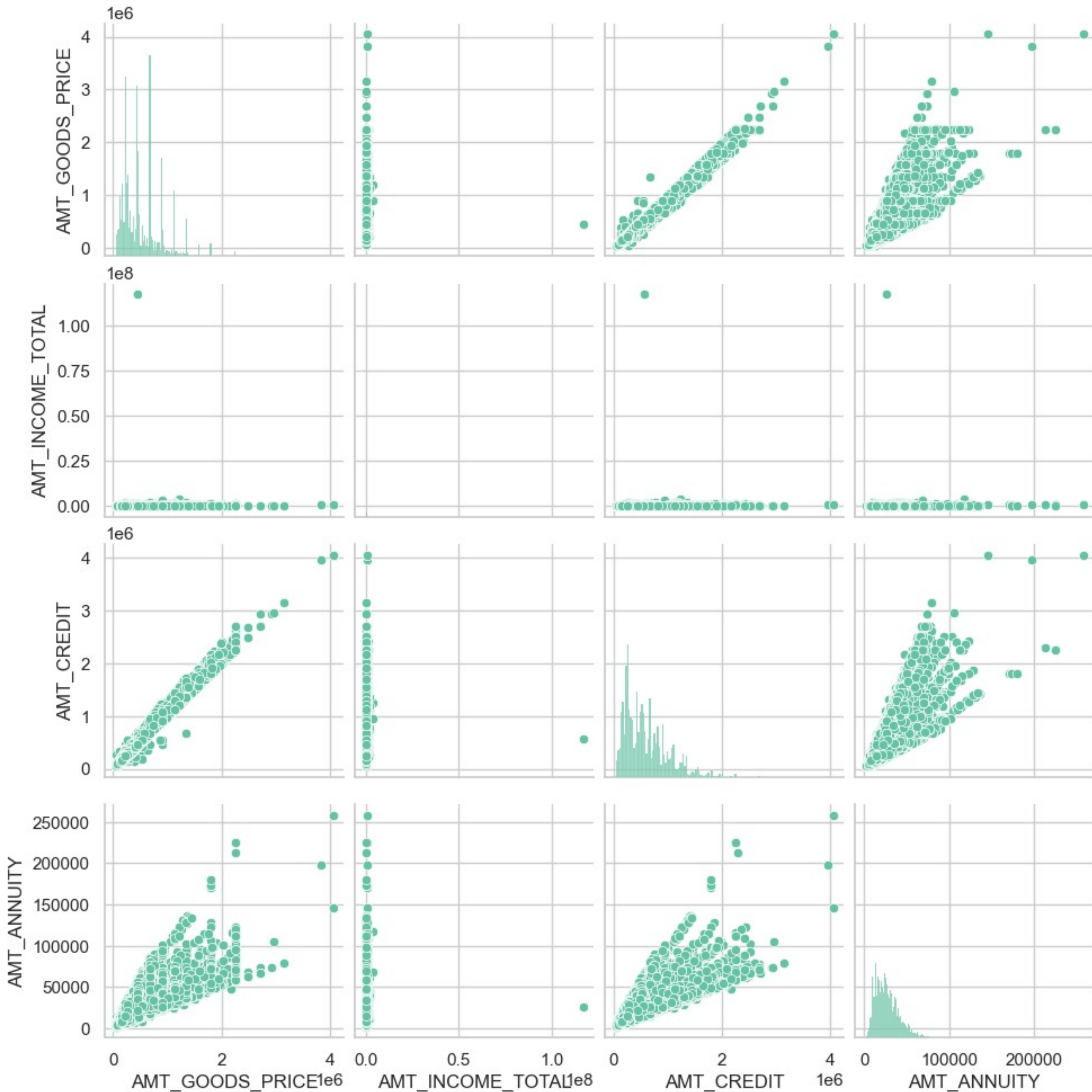
NAME_HOUSINGIN_TYPE univariate analysis shows us that 52% applicants are live in House/Apartment.

CNT_FAM_MEMBERES univariate analysis shows us that 52% applicants have 2 family membersand 22%, 17% applicants have 1 & 3 family members respectively

```
from scipy.stats import chi2_contingency

# Bivariate Analysis: Numeric vs. Numeric Variables
numeric_vars = ['AMT_INCOME_TOTAL', 'AMT_CREDIT']
numeric_df = application_data[numeric_vars]

# Calculate correlation coefficient
correlation_matrix = numeric_df.corr()

# Create scatter plots
sns.pairplot(numeric_df)
plt.show()
```




```
plt.figure(figsize=[20,8])
sns.pairplot(application_data[['AMT_GOODS_PRICE','AMT_INCOME_TOTAL','A
MT_CREDIT','AMT_ANNUITY']])
plt.show()

<Figure size 2000x800 with 0 Axes>
```

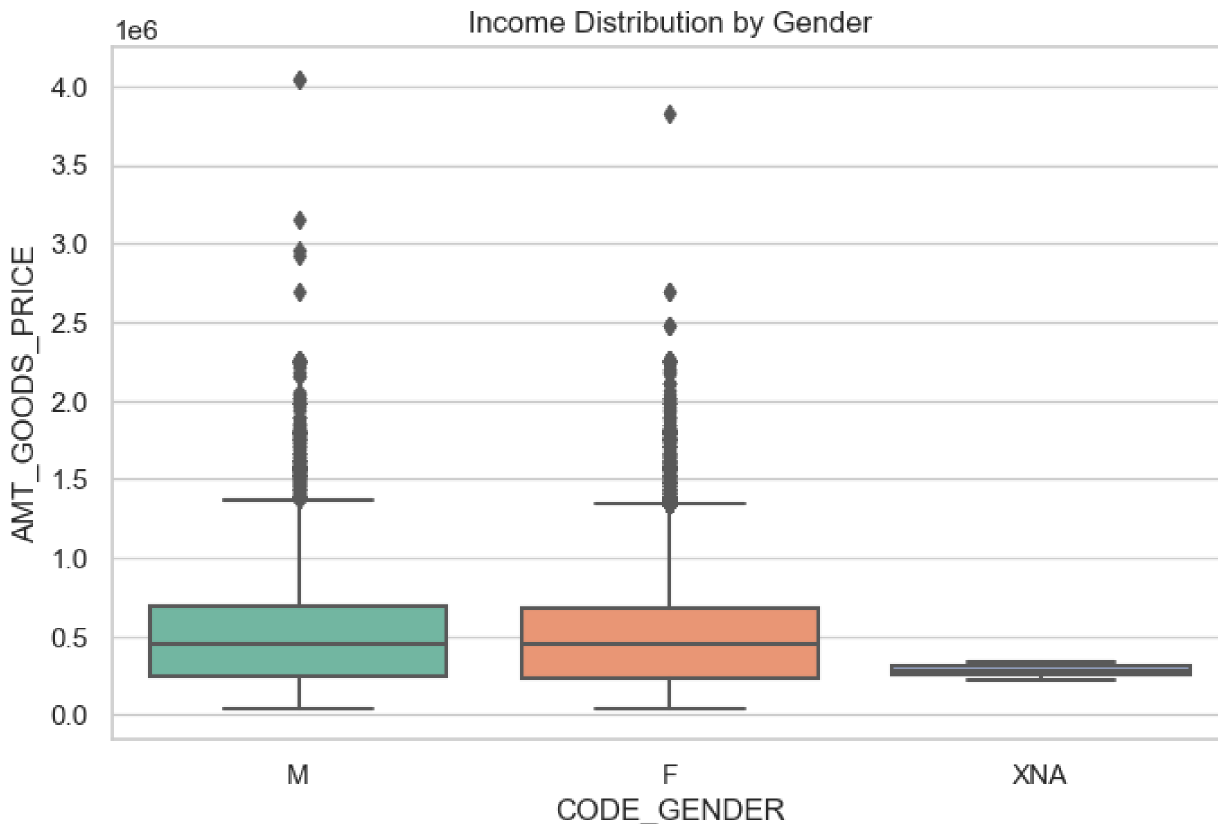


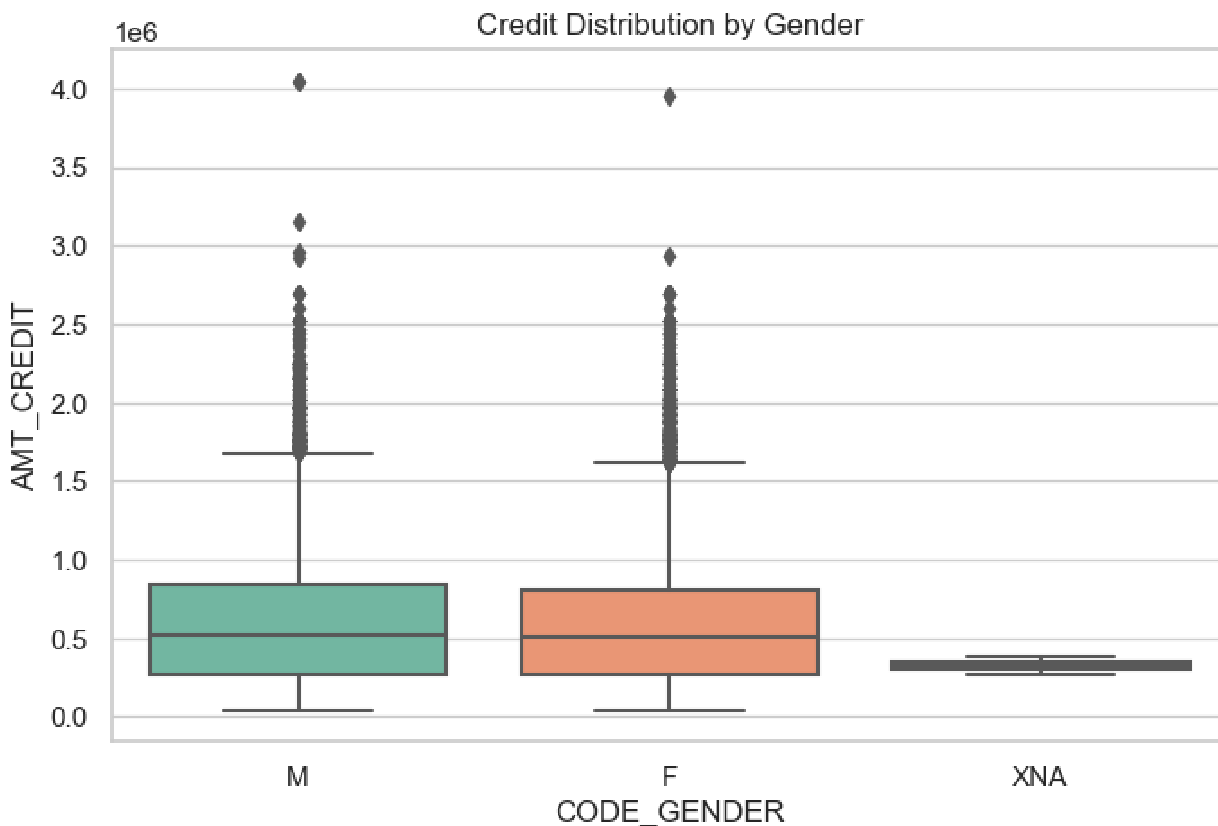
Inference

The variables AMT_GOODS_PRICE, AMT_ANNUITY, and AMT_CREDIT show a good positive correlation, which is expected due to the higher cost of goods leading to larger loan amounts and subsequent annuity payments.

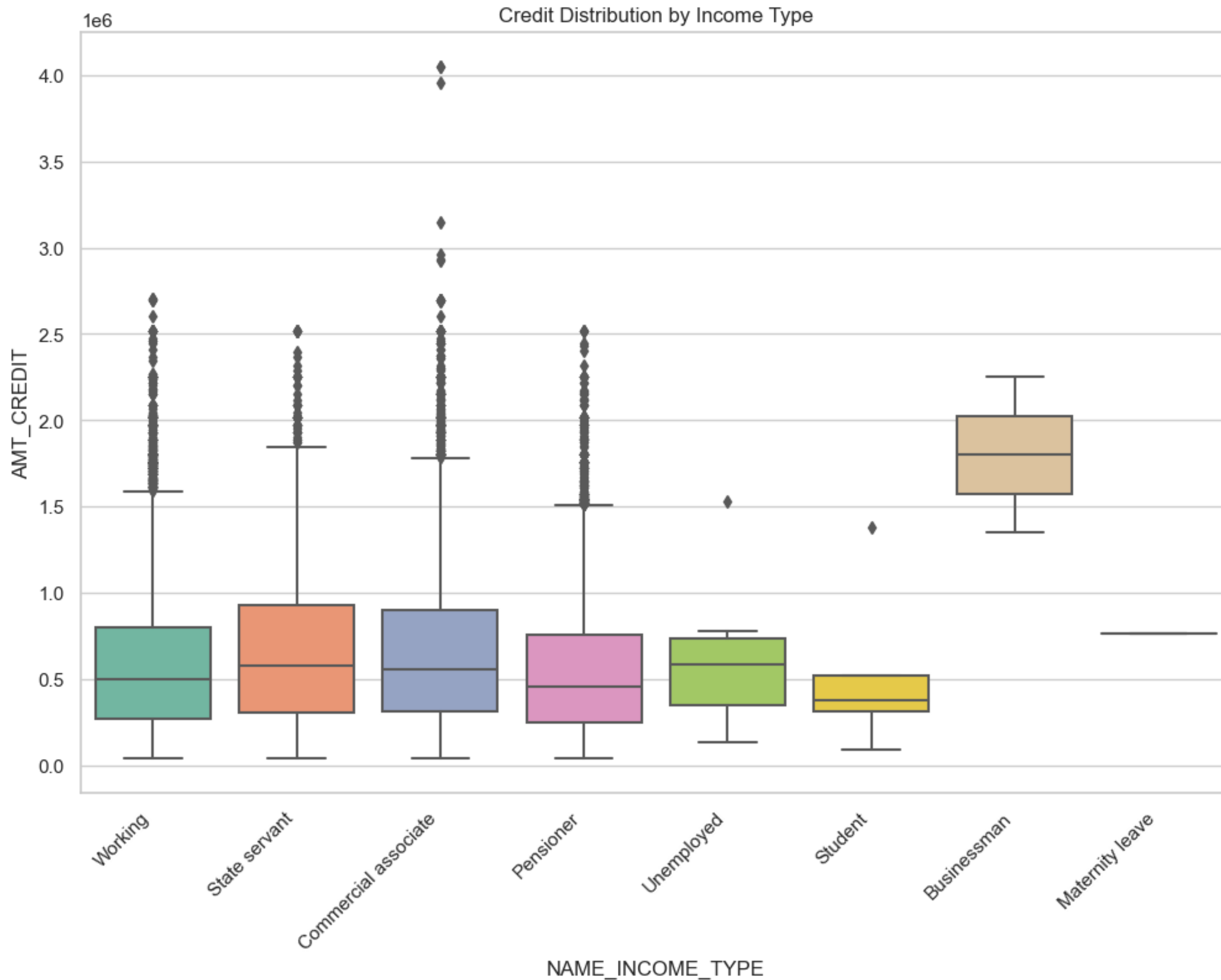
```
# To find Income Distribution by Gender
plt.figure(figsize=(8, 5))
sns.boxplot(x='CODE_GENDER', y='AMT_GOODS_PRICE',
data=application_data)
plt.title('Income Distribution by Gender')
plt.show()
```

```python
# To find Credit Distribution by Gender
plt.figure(figsize=(8, 5))
sns.boxplot(x='CODE_GENDER', y='AMT_CREDIT', data=application_data)
plt.title('Credit Distribution by Gender')
plt.show()
```






```
# To find Credit Distribution by Income Type

plt.figure(figsize=(12, 8))
sns.boxplot(x='NAME_INCOME_TYPE', y='AMT_CREDIT',
data=application_data)
plt.xticks(rotation=45, ha='right')
plt.title('Credit Distribution by Income Type')
plt.show()
```



Inference

_Bivariate analysis between AMT_GOODS_PRICE and AMT_CREDIT through box plots segmented by CODE_GENDER shows that both male and female applicants tend to receive similar loan amounts.

_Bivariate analysis of AMT_CREDIT through box plots segmented by NAME_INCOME_TYPE shows that Businessman get more Credit amount than that of other income types.

E. Identify Top Correlations for Different Scenarios:

Top Correlations for Different Scenarios

```
# Select numerical columns for Univariate Analysis

numerical_columns =
application_data.select_dtypes(include=[np.number])

numerical_columns =
numerical_columns.drop(columns=["FLAG_DOCUMENT_3","FLAG_DOCUMENT_5","F
LAG_DOCUMENT_7","FLAG_DOCUMENT_9","FLAG_DOCUMENT_11","FLAG_DOCUMENT_13
","FLAG_DOCUMENT_15","FLAG_DOCUMENT_17","FLAG_DOCUMENT_19","FLAG_DOCUM
ENT_21","FLAG_DOCUMENT_4","FLAG_DOCUMENT_2","FLAG_DOCUMENT_6","FLAG_DO
CUMENT_8","FLAG_DOCUMENT_10","FLAG_DOCUMENT_12","FLAG_DOCUMENT_14","FL
AG_DOCUMENT_16","FLAG_DOCUMENT_18","FLAG_DOCUMENT_20"])

print(numerical_columns.shape)
numerical_columns.columns
```

```
(49999, 43)
```

```
Index(['SK_ID_CURR', 'TARGET', 'CNT_CHILDREN', 'AMT_INCOME_TOTAL',
       'AMT_CREDIT', 'AMT_ANNUITY', 'AMT_GOODS_PRICE',
       'REGION_POPULATION_RELATIVE', 'DAYS_BIRTH', 'DAYS_EMPLOYED',
       'DAYS_REGISTRATION', 'DAYS_ID_PUBLISH', 'OWN_CAR_AGE',
'FLAG_MOBIL',
       'FLAG_EMP_PHONE', 'FLAG_WORK_PHONE', 'FLAG_CONT_MOBILE',
'FLAG_PHONE',
       'FLAG_EMAIL', 'CNT_FAM_MEMBERS', 'REGION_RATING_CLIENT',
       'REGION_RATING_CLIENT_W_CITY', 'HOUR_APPR_PROCESS_START',
       'REG_REGION_NOT_LIVE_REGION', 'REG_REGION_NOT_WORK_REGION',
       'LIVE_REGION_NOT_WORK_REGION', 'REG_CITY_NOT_LIVE_CITY',
       'REG_CITY_NOT_WORK_CITY', 'LIVE_CITY_NOT_WORK_CITY',
'EXT_SOURCE_1',
       'EXT_SOURCE_2', 'EXT_SOURCE_3', 'OBS_30_CNT_SOCIAL_CIRCLE',
       'DEF_30_CNT_SOCIAL_CIRCLE', 'OBS_60_CNT_SOCIAL_CIRCLE',
       'DEF_60_CNT_SOCIAL_CIRCLE', 'DAYS_LAST_PHONE_CHANGE',
       'AMT_REQ_CREDIT_BUREAU_HOUR', 'AMT_REQ_CREDIT_BUREAU_DAY',
```

```
       'AMT_REQ_CREDIT_BUREAU_WEEK', 'AMT_REQ_CREDIT_BUREAU_MON',
       'AMT_REQ_CREDIT_BUREAU_QRT', 'AMT_REQ_CREDIT_BUREAU_YEAR'],
      dtype='object')

# Generate heatmap

correlation = numerical_columns.corr()
print(correlation)
correlation.to_clipboard(index=True)
```

```
                             SK_ID_CURR    TARGET  CNT_CHILDREN  \
SK_ID_CURR                     1.000000  0.003295      0.005538
TARGET                         0.003295  1.000000      0.026364
CNT_CHILDREN                   0.005538  0.026364      1.000000
AMT_INCOME_TOTAL              -0.003014  0.010894      0.009589
AMT_CREDIT                    -0.000732 -0.032428      0.004972
AMT_ANNUITY                   -0.002084 -0.012399      0.026179
AMT_GOODS_PRICE               -0.000743 -0.041307      0.000253
REGION_POPULATION_RELATIVE     0.001979 -0.040799     -0.025556
DAYS_BIRTH                    -0.001324 -0.076788     -0.329264
DAYS_EMPLOYED                 -0.004393 -0.042472     -0.241540
DAYS_REGISTRATION              0.003709 -0.042343     -0.181217
DAYS_ID_PUBLISH                0.008738 -0.046927      0.032116
OWN_CAR_AGE                    0.002485  0.039534      0.017437
FLAG_MOBIL                     0.002863  0.001323      0.002593
FLAG_EMP_PHONE                 0.004449  0.041408      0.240678
FLAG_WORK_PHONE               -0.003127  0.021302      0.055881
FLAG_CONT_MOBILE               0.001012  0.006766     -0.002827
FLAG_PHONE                    -0.004005 -0.032679     -0.030654
FLAG_EMAIL                     0.002436 -0.001312      0.026816
CNT_FAM_MEMBERS                0.001898  0.013006      0.880430
REGION_RATING_CLIENT           0.004509  0.066130      0.025914
REGION_RATING_CLIENT_W_CITY    0.003209  0.067079      0.022778
HOUR_APPR_PROCESS_START       -0.007562 -0.032036     -0.006254
REG_REGION_NOT_LIVE_REGION    -0.013077  0.009439     -0.010655
REG_REGION_NOT_WORK_REGION    -0.001866 -0.001006      0.012057
LIVE_REGION_NOT_WORK_REGION    0.003249 -0.005498      0.019659
REG_CITY_NOT_LIVE_CITY        -0.005295  0.038773      0.019192
REG_CITY_NOT_WORK_CITY        -0.003629  0.048451      0.070032
LIVE_CITY_NOT_WORK_CITY        0.000143  0.032261      0.067751
EXT_SOURCE_1                  -0.006882 -0.156806     -0.146199
EXT_SOURCE_2                  -0.003865 -0.158424     -0.017641
EXT_SOURCE_3                   0.000215 -0.181276     -0.043791
OBS_30_CNT_SOCIAL_CIRCLE      -0.003183  0.014180      0.016616
DEF_30_CNT_SOCIAL_CIRCLE      -0.008408  0.041603     -0.002965
OBS_60_CNT_SOCIAL_CIRCLE      -0.003520  0.013945      0.016500
DEF_60_CNT_SOCIAL_CIRCLE      -0.007563  0.044260     -0.003955
DAYS_LAST_PHONE_CHANGE        -0.003461  0.056137     -0.002023
AMT_REQ_CREDIT_BUREAU_HOUR    -0.003376  0.002053      0.002405
AMT_REQ_CREDIT_BUREAU_DAY     -0.003983  0.010117     -0.001498
```

```
AMT_REQ_CREDIT_BUREAU_WEEK    -0.005486  0.003328      0.001343
AMT_REQ_CREDIT_BUREAU_MON     -0.001118 -0.013591     -0.010548
AMT_REQ_CREDIT_BUREAU_QRT     -0.003482 -0.005551     -0.005485
AMT_REQ_CREDIT_BUREAU_YEAR    -0.004058  0.009493     -0.029619

                             AMT_INCOME_TOTAL  AMT_CREDIT  AMT_ANNUITY  \
SK_ID_CURR                          -0.003014   -0.000732    -0.002084
TARGET                               0.010894   -0.032428    -0.012399
CNT_CHILDREN                         0.009589    0.004972     0.026179
AMT_INCOME_TOTAL                     1.000000    0.069316     0.083009
AMT_CREDIT                           0.069316    1.000000     0.769499
AMT_ANNUITY                          0.083009    0.769499     1.000000
AMT_GOODS_PRICE                      0.069886    0.986944     0.774434
REGION_POPULATION_RELATIVE           0.029841    0.095111     0.115112
DAYS_BIRTH                          -0.016003    0.059343    -0.007712
DAYS_EMPLOYED                       -0.031510   -0.067739    -0.108710
DAYS_REGISTRATION                   -0.009952   -0.003449    -0.033219
DAYS_ID_PUBLISH                     -0.003507    0.012229    -0.006716
OWN_CAR_AGE                         -0.142108   -0.092520    -0.096866
FLAG_MOBIL                           0.000376    0.003568     0.000367
FLAG_EMP_PHONE                       0.031568    0.069063     0.109570
FLAG_WORK_PHONE                     -0.009181   -0.015115    -0.020965
FLAG_CONT_MOBILE                    -0.003246    0.024481     0.023111
FLAG_PHONE                          -0.002044    0.019460     0.005127
FLAG_EMAIL                           0.015215    0.010812     0.065896
CNT_FAM_MEMBERS                      0.011202    0.063985     0.077353
REGION_RATING_CLIENT                -0.038189   -0.100507    -0.125803
REGION_RATING_CLIENT_W_CITY         -0.040719   -0.109487    -0.139322
```

```
HOUR_APPR_PROCESS_START                     0.018464     0.056677      0.053275
REG_REGION_NOT_LIVE_REGION                  0.013773     0.025772      0.044803
REG_REGION_NOT_WORK_REGION                  0.027597     0.053846      0.081295
LIVE_REGION_NOT_WORK_REGION                 0.026152     0.053262      0.074849
REG_CITY_NOT_LIVE_CITY                      0.000136    -0.024978     -0.006721
REG_CITY_NOT_WORK_CITY                     -0.000021    -0.017377      0.001061
LIVE_CITY_NOT_WORK_CITY                     0.001295     0.002136      0.010940
EXT_SOURCE_1                                0.010034     0.170622      0.115437
EXT_SOURCE_2                                0.019518     0.138125      0.128928
EXT_SOURCE_3                               -0.021566     0.041746      0.023602
OBS_30_CNT_SOCIAL_CIRCLE                   -0.008622     0.001806     -0.009325
DEF_30_CNT_SOCIAL_CIRCLE                   -0.007629    -0.016197     -0.021819
OBS_60_CNT_SOCIAL_CIRCLE                   -0.008591     0.002163     -0.009012
DEF_60_CNT_SOCIAL_CIRCLE                   -0.007344    -0.021152     -0.025276
DAYS_LAST_PHONE_CHANGE                     -0.004804    -0.076179     -0.067257
AMT_REQ_CREDIT_BUREAU_HOUR                  0.001102     0.001224      0.012014
AMT_REQ_CREDIT_BUREAU_DAY                   0.001311     0.011537      0.007011
AMT_REQ_CREDIT_BUREAU_WEEK                  0.001057     0.004902      0.019918
AMT_REQ_CREDIT_BUREAU_MON                   0.012808     0.065406      0.039915
AMT_REQ_CREDIT_BUREAU_QRT                   0.001589     0.023708      0.009306
AMT_REQ_CREDIT_BUREAU_YEAR                  0.006281    -0.019420      0.000027

                             AMT_GOODS_PRICE
REGION_POPULATION_RELATIVE  \
SK_ID_CURR                         -0.000743
0.001979
TARGET                             -0.041307                    -
0.040799
CNT_CHILDREN                        0.000253                    -
0.025556
```

```
AMT_INCOME_TOTAL                     0.069886
0.029841
AMT_CREDIT                           0.986944
0.095111
AMT_ANNUITY                          0.774434
0.115112
AMT_GOODS_PRICE                      1.000000
0.099190
REGION_POPULATION_RELATIVE           0.099190
1.000000
DAYS_BIRTH                           0.057611
0.032514
DAYS_EMPLOYED                       -0.065059                   -
0.004158
DAYS_REGISTRATION                   -0.006101
0.059322
DAYS_ID_PUBLISH                      0.013968
0.004345
OWN_CAR_AGE                         -0.100465                   -
0.083575
FLAG_MOBIL                           0.003472
0.003301
FLAG_EMP_PHONE                       0.066428
0.004186
FLAG_WORK_PHONE                      0.006491                   -
0.016787
FLAG_CONT_MOBILE                     0.022094                   -
0.005050
FLAG_PHONE                           0.035532
0.094068
FLAG_EMAIL                           0.010534
0.039508
CNT_FAM_MEMBERS                      0.061624                   -
0.023018
REGION_RATING_CLIENT                -0.103722                   -
0.532667
REGION_RATING_CLIENT_W_CITY         -0.111796                   -
0.530439
HOUR_APPR_PROCESS_START              0.066006
0.167725
REG_REGION_NOT_LIVE_REGION           0.028023                   -
0.003552
REG_REGION_NOT_WORK_REGION           0.055274
0.060070
LIVE_REGION_NOT_WORK_REGION          0.053566
0.085697
REG_CITY_NOT_LIVE_CITY              -0.024507                   -
0.046482
REG_CITY_NOT_WORK_CITY              -0.018651                   -
```

```
0.040443
LIVE_CITY_NOT_WORK_CITY                0.000333                    -
0.013596
EXT_SOURCE_1                           0.176724
0.102956
EXT_SOURCE_2                           0.146936
0.201242
EXT_SOURCE_3                           0.045647                    -
0.009622
OBS_30_CNT_SOCIAL_CIRCLE               0.001504                    -
0.018011
DEF_30_CNT_SOCIAL_CIRCLE              -0.017494
0.009276
OBS_60_CNT_SOCIAL_CIRCLE               0.001794                    -
0.016897
DEF_60_CNT_SOCIAL_CIRCLE              -0.021743
0.004000
DAYS_LAST_PHONE_CHANGE                -0.079750                    -
0.047752
AMT_REQ_CREDIT_BUREAU_HOUR             0.001717                    -
0.002315
AMT_REQ_CREDIT_BUREAU_DAY              0.011749                    -
0.001006
AMT_REQ_CREDIT_BUREAU_WEEK             0.005279
0.003154
AMT_REQ_CREDIT_BUREAU_MON              0.066869
0.071462
AMT_REQ_CREDIT_BUREAU_QRT              0.024387                    -
0.007684
AMT_REQ_CREDIT_BUREAU_YEAR            -0.022805
0.004505

                             DAYS_BIRTH  DAYS_EMPLOYED  ...  \
SK_ID_CURR                    -0.001324      -0.004393  ...
TARGET                        -0.076788      -0.042472  ...
CNT_CHILDREN                  -0.329264      -0.241540  ...
AMT_INCOME_TOTAL              -0.016003      -0.031510  ...
AMT_CREDIT                     0.059343      -0.067739  ...
AMT_ANNUITY                   -0.007712      -0.108710  ...
AMT_GOODS_PRICE                0.057611      -0.065059  ...
REGION_POPULATION_RELATIVE     0.032514      -0.004158  ...
DAYS_BIRTH                     1.000000       0.621728  ...
DAYS_EMPLOYED                  0.621728       1.000000  ...
DAYS_REGISTRATION              0.333633       0.209172  ...
DAYS_ID_PUBLISH                0.270825       0.272767  ...
OWN_CAR_AGE                   -0.003714       0.026944  ...
FLAG_MOBIL                     0.007637       0.002152  ...
FLAG_EMP_PHONE                -0.617703      -0.999746  ...
FLAG_WORK_PHONE               -0.175690      -0.232311  ...
```

```
FLAG_CONT_MOBILE                  0.011900       0.015164  ...
FLAG_PHONE                        0.044219       0.024336  ...
FLAG_EMAIL                       -0.092150      -0.067357  ...
CNT_FAM_MEMBERS                  -0.277189      -0.230727  ...
REGION_RATING_CLIENT             -0.016779       0.034559  ...
REGION_RATING_CLIENT_W_CITY      -0.014552       0.036973  ...
HOUR_APPR_PROCESS_START          -0.090589      -0.088523  ...
REG_REGION_NOT_LIVE_REGION       -0.059105      -0.038080  ...
REG_REGION_NOT_WORK_REGION       -0.093985      -0.108099  ...
LIVE_REGION_NOT_WORK_REGION      -0.068120      -0.095692  ...
REG_CITY_NOT_LIVE_CITY           -0.182109      -0.096498  ...
REG_CITY_NOT_WORK_CITY           -0.237897      -0.258209  ...
LIVE_CITY_NOT_WORK_CITY          -0.150552      -0.219286  ...
EXT_SOURCE_1                      0.600884       0.290723  ...
EXT_SOURCE_2                      0.093882      -0.023593  ...
EXT_SOURCE_3                      0.212215       0.118108  ...
OBS_30_CNT_SOCIAL_CIRCLE         -0.011682       0.004930  ...
DEF_30_CNT_SOCIAL_CIRCLE         -0.001899       0.015937  ...
OBS_60_CNT_SOCIAL_CIRCLE         -0.011565       0.004869  ...
DEF_60_CNT_SOCIAL_CIRCLE         -0.002842       0.014810  ...
DAYS_LAST_PHONE_CHANGE           -0.080190       0.023612  ...
AMT_REQ_CREDIT_BUREAU_HOUR       -0.003514      -0.004418  ...
AMT_REQ_CREDIT_BUREAU_DAY        -0.000736       0.004622  ...
AMT_REQ_CREDIT_BUREAU_WEEK        0.002699      -0.004528  ...
AMT_REQ_CREDIT_BUREAU_MON         0.003778      -0.031611  ...
AMT_REQ_CREDIT_BUREAU_QRT         0.020875       0.015096  ...
AMT_REQ_CREDIT_BUREAU_YEAR        0.074000       0.037550  ...

                             DEF_30_CNT_SOCIAL_CIRCLE  \
SK_ID_CURR                                  -0.008408
TARGET                                       0.041603
CNT_CHILDREN                                -0.002965
AMT_INCOME_TOTAL                            -0.007629
AMT_CREDIT                                  -0.016197
AMT_ANNUITY                                 -0.021819
AMT_GOODS_PRICE                             -0.017494
REGION_POPULATION_RELATIVE                   0.009276
DAYS_BIRTH                                  -0.001899
DAYS_EMPLOYED                                0.015937
DAYS_REGISTRATION                           -0.004954
DAYS_ID_PUBLISH                             -0.001267
OWN_CAR_AGE                                  0.014954
FLAG_MOBIL                                   0.001442
FLAG_EMP_PHONE                              -0.016121
FLAG_WORK_PHONE                             -0.014769
FLAG_CONT_MOBILE                             0.000328
FLAG_PHONE                                  -0.028320
FLAG_EMAIL                                  -0.004021
CNT_FAM_MEMBERS                             -0.002709
```

```
REGION_RATING_CLIENT                                       0.010547
REGION_RATING_CLIENT_W_CITY                                0.008884
HOUR_APPR_PROCESS_START                                   -0.001399
REG_REGION_NOT_LIVE_REGION                                -0.005975
REG_REGION_NOT_WORK_REGION                                -0.007859
LIVE_REGION_NOT_WORK_REGION                               -0.006901
REG_CITY_NOT_LIVE_CITY                                     0.006860
REG_CITY_NOT_WORK_CITY                                     0.001111
LIVE_CITY_NOT_WORK_CITY                                   -0.003794
EXT_SOURCE_1                                              -0.026880
EXT_SOURCE_2                                              -0.032851
EXT_SOURCE_3                                              -0.038033
OBS_30_CNT_SOCIAL_CIRCLE                                   0.311411
DEF_30_CNT_SOCIAL_CIRCLE                                   1.000000
OBS_60_CNT_SOCIAL_CIRCLE                                   0.313822
DEF_60_CNT_SOCIAL_CIRCLE                                   0.856223
DAYS_LAST_PHONE_CHANGE                                     0.005240
AMT_REQ_CREDIT_BUREAU_HOUR                                -0.003421
AMT_REQ_CREDIT_BUREAU_DAY                                  0.004937
AMT_REQ_CREDIT_BUREAU_WEEK                                -0.005671
AMT_REQ_CREDIT_BUREAU_MON                                  0.006967
AMT_REQ_CREDIT_BUREAU_QRT                                  0.006531
AMT_REQ_CREDIT_BUREAU_YEAR                                 0.014308

                             OBS_60_CNT_SOCIAL_CIRCLE  \
SK_ID_CURR                                                -0.003520
TARGET                                                     0.013945
CNT_CHILDREN                                               0.016500
AMT_INCOME_TOTAL                                          -0.008591
AMT_CREDIT                                                 0.002163
AMT_ANNUITY                                               -0.009012
AMT_GOODS_PRICE                                            0.001794
REGION_POPULATION_RELATIVE                                -0.016897
DAYS_BIRTH                                                -0.011565
DAYS_EMPLOYED                                              0.004869
DAYS_REGISTRATION                                         -0.010552
DAYS_ID_PUBLISH                                            0.012464
OWN_CAR_AGE                                                0.004866
FLAG_MOBIL                                                 0.002756
FLAG_EMP_PHONE                                            -0.004832
FLAG_WORK_PHONE                                           -0.020758
FLAG_CONT_MOBILE                                           0.006604
FLAG_PHONE                                                -0.034807
FLAG_EMAIL                                                -0.003188
CNT_FAM_MEMBERS                                            0.025812
REGION_RATING_CLIENT                                       0.034581
REGION_RATING_CLIENT_W_CITY                                0.032125
HOUR_APPR_PROCESS_START                                   -0.008885
REG_REGION_NOT_LIVE_REGION                                -0.016541
```

```
REG_REGION_NOT_WORK_REGION                       -0.025874
LIVE_REGION_NOT_WORK_REGION                      -0.020243
REG_CITY_NOT_LIVE_CITY                           -0.009358
REG_CITY_NOT_WORK_CITY                           -0.008722
LIVE_CITY_NOT_WORK_CITY                          -0.006472
EXT_SOURCE_1                                     -0.024871
EXT_SOURCE_2                                     -0.017695
EXT_SOURCE_3                                     -0.000598
OBS_30_CNT_SOCIAL_CIRCLE                          0.998331
DEF_30_CNT_SOCIAL_CIRCLE                          0.313822
OBS_60_CNT_SOCIAL_CIRCLE                          1.000000
DEF_60_CNT_SOCIAL_CIRCLE                          0.237930
DAYS_LAST_PHONE_CHANGE                           -0.014415
AMT_REQ_CREDIT_BUREAU_HOUR                        0.001544
AMT_REQ_CREDIT_BUREAU_DAY                        -0.000716
AMT_REQ_CREDIT_BUREAU_WEEK                       -0.004231
AMT_REQ_CREDIT_BUREAU_MON                         0.008277
AMT_REQ_CREDIT_BUREAU_QRT                         0.010772
AMT_REQ_CREDIT_BUREAU_YEAR                        0.036337

                             DEF_60_CNT_SOCIAL_CIRCLE
DAYS_LAST_PHONE_CHANGE  \
SK_ID_CURR                                       -0.007563                -
0.003461
TARGET                                            0.044260
0.056137
CNT_CHILDREN                                     -0.003955                -
0.002023
AMT_INCOME_TOTAL                                 -0.007344                -
0.004804
AMT_CREDIT                                       -0.021152                -
0.076179
AMT_ANNUITY                                      -0.025276                -
0.067257
AMT_GOODS_PRICE                                  -0.021743                -
0.079750
REGION_POPULATION_RELATIVE                        0.004000                -
0.047752
DAYS_BIRTH                                       -0.002842                -
0.080190
DAYS_EMPLOYED                                     0.014810
0.023612
DAYS_REGISTRATION                                -0.006863                -
0.052146
DAYS_ID_PUBLISH                                  -0.001560                -
0.091375
OWN_CAR_AGE                                       0.014948                -
0.000053
FLAG_MOBIL                                        0.001233
```

```
NaN
FLAG_EMP_PHONE                                  -0.015011               -
0.025556
FLAG_WORK_PHONE                                 -0.012288               -
0.041372
FLAG_CONT_MOBILE                                 0.002412               -
0.024574
FLAG_PHONE                                      -0.027020               -
0.067253
FLAG_EMAIL                                      -0.003760               -
0.018054
CNT_FAM_MEMBERS                                 -0.004528               -
0.022708
REGION_RATING_CLIENT                             0.012628
0.027327
REGION_RATING_CLIENT_W_CITY                      0.010966
0.026789
HOUR_APPR_PROCESS_START                         -0.004749               -
0.017873
REG_REGION_NOT_LIVE_REGION                      -0.007144
0.031606
REG_REGION_NOT_WORK_REGION                      -0.011727
0.034668
LIVE_REGION_NOT_WORK_REGION                     -0.010898
0.024238
REG_CITY_NOT_LIVE_CITY                           0.006864
0.053764
REG_CITY_NOT_WORK_CITY                           0.003419
0.046866
LIVE_CITY_NOT_WORK_CITY                         -0.001646
0.021903
EXT_SOURCE_1                                    -0.028560               -
0.134245
EXT_SOURCE_2                                    -0.036537               -
0.192592
EXT_SOURCE_3                                    -0.035986               -
0.078136
OBS_30_CNT_SOCIAL_CIRCLE                         0.235737               -
0.013601
DEF_30_CNT_SOCIAL_CIRCLE                         0.856223
0.005240
OBS_60_CNT_SOCIAL_CIRCLE                         0.237930               -
0.014415
DEF_60_CNT_SOCIAL_CIRCLE                         1.000000
0.006231
DAYS_LAST_PHONE_CHANGE                           0.006231
1.000000
AMT_REQ_CREDIT_BUREAU_HOUR                      -0.004067               -
0.000114
```

```
AMT_REQ_CREDIT_BUREAU_DAY                             0.001698                -
0.000380
AMT_REQ_CREDIT_BUREAU_WEEK                           -0.005479                -
0.005433
AMT_REQ_CREDIT_BUREAU_MON                             0.003971                -
0.048532
AMT_REQ_CREDIT_BUREAU_QRT                             0.009792                -
0.012134
AMT_REQ_CREDIT_BUREAU_YEAR                            0.014640                -
0.121967

                             AMT_REQ_CREDIT_BUREAU_HOUR  \
SK_ID_CURR                                    -0.003376
TARGET                                         0.002053
CNT_CHILDREN                                   0.002405
AMT_INCOME_TOTAL                               0.001102
AMT_CREDIT                                     0.001224
AMT_ANNUITY                                    0.012014
AMT_GOODS_PRICE                                0.001717
REGION_POPULATION_RELATIVE                    -0.002315
DAYS_BIRTH                                    -0.003514
DAYS_EMPLOYED                                 -0.004418
DAYS_REGISTRATION                              0.002817
DAYS_ID_PUBLISH                               -0.003866
OWN_CAR_AGE                                    0.007001
FLAG_MOBIL                                     0.000336
FLAG_EMP_PHONE                                 0.003715
FLAG_WORK_PHONE                               -0.008695
FLAG_CONT_MOBILE                              -0.002073
FLAG_PHONE                                    -0.008897
FLAG_EMAIL                                     0.006319
CNT_FAM_MEMBERS                                0.003822
REGION_RATING_CLIENT                           0.006821
REGION_RATING_CLIENT_W_CITY                    0.005644
HOUR_APPR_PROCESS_START                       -0.009653
REG_REGION_NOT_LIVE_REGION                    -0.003236
REG_REGION_NOT_WORK_REGION                     0.001884
LIVE_REGION_NOT_WORK_REGION                    0.004810
REG_CITY_NOT_LIVE_CITY                         0.000408
REG_CITY_NOT_WORK_CITY                         0.005644
LIVE_CITY_NOT_WORK_CITY                        0.004998
EXT_SOURCE_1                                   0.001420
EXT_SOURCE_2                                  -0.001662
EXT_SOURCE_3                                   0.000435
OBS_30_CNT_SOCIAL_CIRCLE                       0.001294
DEF_30_CNT_SOCIAL_CIRCLE                      -0.003421
OBS_60_CNT_SOCIAL_CIRCLE                       0.001544
DEF_60_CNT_SOCIAL_CIRCLE                      -0.004067
DAYS_LAST_PHONE_CHANGE                        -0.000114
```

```
AMT_REQ_CREDIT_BUREAU_HOUR                           1.000000
AMT_REQ_CREDIT_BUREAU_DAY                            0.241413
AMT_REQ_CREDIT_BUREAU_WEEK                           0.012731
AMT_REQ_CREDIT_BUREAU_MON                            0.008800
AMT_REQ_CREDIT_BUREAU_QRT                            0.005861
AMT_REQ_CREDIT_BUREAU_YEAR                           0.009576

                             AMT_REQ_CREDIT_BUREAU_DAY  \
SK_ID_CURR                                  -0.003983
TARGET                                       0.010117
CNT_CHILDREN                                -0.001498
AMT_INCOME_TOTAL                             0.001311
AMT_CREDIT                                   0.011537
AMT_ANNUITY                                  0.007011
AMT_GOODS_PRICE                              0.011749
REGION_POPULATION_RELATIVE                  -0.001006
DAYS_BIRTH                                  -0.000736
DAYS_EMPLOYED                                0.004622
DAYS_REGISTRATION                            0.002799
DAYS_ID_PUBLISH                             -0.003129
OWN_CAR_AGE                                  0.004615
FLAG_MOBIL                                   0.000289
FLAG_EMP_PHONE                              -0.004667
FLAG_WORK_PHONE                             -0.005362
FLAG_CONT_MOBILE                            -0.014821
FLAG_PHONE                                   0.000773
FLAG_EMAIL                                   0.003394
CNT_FAM_MEMBERS                             -0.002324
REGION_RATING_CLIENT                         0.004392
REGION_RATING_CLIENT_W_CITY                  0.003573
HOUR_APPR_PROCESS_START                      0.009235
REG_REGION_NOT_LIVE_REGION                  -0.004708
REG_REGION_NOT_WORK_REGION                   0.001623
LIVE_REGION_NOT_WORK_REGION                  0.003179
REG_CITY_NOT_LIVE_CITY                      -0.001458
REG_CITY_NOT_WORK_CITY                      -0.000214
LIVE_CITY_NOT_WORK_CITY                     -0.000728
EXT_SOURCE_1                                 0.000222
EXT_SOURCE_2                                -0.003756
EXT_SOURCE_3                                -0.004208
OBS_30_CNT_SOCIAL_CIRCLE                    -0.000585
DEF_30_CNT_SOCIAL_CIRCLE                     0.004937
OBS_60_CNT_SOCIAL_CIRCLE                    -0.000716
DEF_60_CNT_SOCIAL_CIRCLE                     0.001698
DAYS_LAST_PHONE_CHANGE                      -0.000380
AMT_REQ_CREDIT_BUREAU_HOUR                   0.241413
AMT_REQ_CREDIT_BUREAU_DAY                    1.000000
AMT_REQ_CREDIT_BUREAU_WEEK                   0.233164
AMT_REQ_CREDIT_BUREAU_MON                   -0.001836
AMT_REQ_CREDIT_BUREAU_QRT                   -0.005011
```

```
AMT_REQ_CREDIT_BUREAU_YEAR                        0.006455

                             AMT_REQ_CREDIT_BUREAU_WEEK  \
SK_ID_CURR                                    -0.005486
TARGET                                         0.003328
CNT_CHILDREN                                   0.001343
AMT_INCOME_TOTAL                               0.001057
AMT_CREDIT                                     0.004902
AMT_ANNUITY                                    0.019918
AMT_GOODS_PRICE                                0.005279
REGION_POPULATION_RELATIVE                     0.003154
DAYS_BIRTH                                     0.002699
DAYS_EMPLOYED                                 -0.004528
DAYS_REGISTRATION                              0.000606
DAYS_ID_PUBLISH                                0.002539
OWN_CAR_AGE                                    0.010364
FLAG_MOBIL                                     0.000693
FLAG_EMP_PHONE                                 0.004652
FLAG_WORK_PHONE                               -0.004199
FLAG_CONT_MOBILE                              -0.020121
FLAG_PHONE                                     0.002940
FLAG_EMAIL                                     0.024124
CNT_FAM_MEMBERS                                0.003200
REGION_RATING_CLIENT                          -0.000517
REGION_RATING_CLIENT_W_CITY                   -0.003532
HOUR_APPR_PROCESS_START                       -0.007025
REG_REGION_NOT_LIVE_REGION                    -0.000924
REG_REGION_NOT_WORK_REGION                     0.001554
LIVE_REGION_NOT_WORK_REGION                    0.002529
REG_CITY_NOT_LIVE_CITY                        -0.001234
REG_CITY_NOT_WORK_CITY                         0.002550
LIVE_CITY_NOT_WORK_CITY                        0.003236
EXT_SOURCE_1                                  -0.006674
EXT_SOURCE_2                                   0.006190
EXT_SOURCE_3                                  -0.023144
OBS_30_CNT_SOCIAL_CIRCLE                      -0.003662
DEF_30_CNT_SOCIAL_CIRCLE                      -0.005671
OBS_60_CNT_SOCIAL_CIRCLE                      -0.004231
DEF_60_CNT_SOCIAL_CIRCLE                      -0.005479
DAYS_LAST_PHONE_CHANGE                        -0.005433
AMT_REQ_CREDIT_BUREAU_HOUR                     0.012731
AMT_REQ_CREDIT_BUREAU_DAY                      0.233164
AMT_REQ_CREDIT_BUREAU_WEEK                     1.000000
AMT_REQ_CREDIT_BUREAU_MON                     -0.010014
AMT_REQ_CREDIT_BUREAU_QRT                     -0.012454
AMT_REQ_CREDIT_BUREAU_YEAR                     0.035362

                             AMT_REQ_CREDIT_BUREAU_MON  \
SK_ID_CURR                                   -0.001118
```

```
TARGET                                              -0.013591
CNT_CHILDREN                                        -0.010548
AMT_INCOME_TOTAL                                     0.012808
AMT_CREDIT                                           0.065406
AMT_ANNUITY                                          0.039915
AMT_GOODS_PRICE                                      0.066869
REGION_POPULATION_RELATIVE                           0.071462
DAYS_BIRTH                                           0.003778
DAYS_EMPLOYED                                       -0.031611
DAYS_REGISTRATION                                    0.010740
DAYS_ID_PUBLISH                                      0.015401
OWN_CAR_AGE                                         -0.014377
FLAG_MOBIL                                           0.001204
FLAG_EMP_PHONE                                       0.032024
FLAG_WORK_PHONE                                     -0.008081
FLAG_CONT_MOBILE                                     0.005447
FLAG_PHONE                                           0.042271
FLAG_EMAIL                                           0.020898
CNT_FAM_MEMBERS                                     -0.003239
REGION_RATING_CLIENT                                -0.065410
REGION_RATING_CLIENT_W_CITY                         -0.063231
HOUR_APPR_PROCESS_START                              0.031588
REG_REGION_NOT_LIVE_REGION                          -0.004435
REG_REGION_NOT_WORK_REGION                           0.002984
LIVE_REGION_NOT_WORK_REGION                          0.011852
REG_CITY_NOT_LIVE_CITY                              -0.015625
REG_CITY_NOT_WORK_CITY                              -0.015211
LIVE_CITY_NOT_WORK_CITY                             -0.006084
EXT_SOURCE_1                                         0.032069
EXT_SOURCE_2                                         0.052304
EXT_SOURCE_3                                        -0.005081
OBS_30_CNT_SOCIAL_CIRCLE                             0.008257
DEF_30_CNT_SOCIAL_CIRCLE                             0.006967
OBS_60_CNT_SOCIAL_CIRCLE                             0.008277
DEF_60_CNT_SOCIAL_CIRCLE                             0.003971
DAYS_LAST_PHONE_CHANGE                              -0.048532
AMT_REQ_CREDIT_BUREAU_HOUR                           0.008800
AMT_REQ_CREDIT_BUREAU_DAY                           -0.001836
AMT_REQ_CREDIT_BUREAU_WEEK                          -0.010014
AMT_REQ_CREDIT_BUREAU_MON                            1.000000
AMT_REQ_CREDIT_BUREAU_QRT                            0.012454
AMT_REQ_CREDIT_BUREAU_YEAR                           0.038940

                             AMT_REQ_CREDIT_BUREAU_QRT  \
SK_ID_CURR                                          -0.003482
TARGET                                              -0.005551
CNT_CHILDREN                                        -0.005485
AMT_INCOME_TOTAL                                     0.001589
AMT_CREDIT                                           0.023708
```

AMT_ANNUITY	0.009306
AMT_GOODS_PRICE	0.024387
REGION_POPULATION_RELATIVE	-0.007684
DAYS_BIRTH	0.020875
DAYS_EMPLOYED	0.015096
DAYS_REGISTRATION	-0.002168
DAYS_ID_PUBLISH	0.025493
OWN_CAR_AGE	-0.006915
FLAG_MOBIL	0.001767
FLAG_EMP_PHONE	-0.015193
FLAG_WORK_PHONE	-0.028020
FLAG_CONT_MOBILE	0.000630
FLAG_PHONE	-0.011156
FLAG_EMAIL	0.016410
CNT_FAM_MEMBERS	-0.003731
REGION_RATING_CLIENT	0.010586
REGION_RATING_CLIENT_W_CITY	0.008997
HOUR_APPR_PROCESS_START	-0.001253
REG_REGION_NOT_LIVE_REGION	-0.001257
REG_REGION_NOT_WORK_REGION	-0.008942
LIVE_REGION_NOT_WORK_REGION	-0.012478
REG_CITY_NOT_LIVE_CITY	-0.000238
REG_CITY_NOT_WORK_CITY	-0.008235
LIVE_CITY_NOT_WORK_CITY	-0.008296
EXT_SOURCE_1	0.004251
EXT_SOURCE_2	0.000290
EXT_SOURCE_3	-0.026090
OBS_30_CNT_SOCIAL_CIRCLE	0.010803
DEF_30_CNT_SOCIAL_CIRCLE	0.006531
OBS_60_CNT_SOCIAL_CIRCLE	0.010772
DEF_60_CNT_SOCIAL_CIRCLE	0.009792
DAYS_LAST_PHONE_CHANGE	-0.012134
AMT_REQ_CREDIT_BUREAU_HOUR	0.005861
AMT_REQ_CREDIT_BUREAU_DAY	-0.005011
AMT_REQ_CREDIT_BUREAU_WEEK	-0.012454
AMT_REQ_CREDIT_BUREAU_MON	0.012454
AMT_REQ_CREDIT_BUREAU_QRT	1.000000
AMT_REQ_CREDIT_BUREAU_YEAR	0.143058

	AMT_REQ_CREDIT_BUREAU_YEAR
SK_ID_CURR	-0.004058
TARGET	0.009493
CNT_CHILDREN	-0.029619
AMT_INCOME_TOTAL	0.006281
AMT_CREDIT	-0.019420
AMT_ANNUITY	0.000027
AMT_GOODS_PRICE	-0.022805
REGION_POPULATION_RELATIVE	0.004505
DAYS_BIRTH	0.074000

```
DAYS_EMPLOYED                                               0.037550
DAYS_REGISTRATION                                           0.022053
DAYS_ID_PUBLISH                                             0.061546
OWN_CAR_AGE                                                -0.021589
FLAG_MOBIL                                                  0.003935
FLAG_EMP_PHONE                                             -0.037329
FLAG_WORK_PHONE                                            -0.066926
FLAG_CONT_MOBILE                                            0.021767
FLAG_PHONE                                                 -0.012145
FLAG_EMAIL                                                  0.045041
CNT_FAM_MEMBERS                                            -0.014805
REGION_RATING_CLIENT                                        0.008251
REGION_RATING_CLIENT_W_CITY                                 0.006391
HOUR_APPR_PROCESS_START                                    -0.022741
REG_REGION_NOT_LIVE_REGION                                 -0.022749
REG_REGION_NOT_WORK_REGION                                 -0.028481
LIVE_REGION_NOT_WORK_REGION                                -0.022003
REG_CITY_NOT_LIVE_CITY                                     -0.012058
REG_CITY_NOT_WORK_CITY                                     -0.016600
LIVE_CITY_NOT_WORK_CITY                                    -0.013596
EXT_SOURCE_1                                                0.032711
EXT_SOURCE_2                                                0.000103
EXT_SOURCE_3                                               -0.070507
OBS_30_CNT_SOCIAL_CIRCLE                                    0.035988
DEF_30_CNT_SOCIAL_CIRCLE                                    0.014308
OBS_60_CNT_SOCIAL_CIRCLE                                    0.036337
DEF_60_CNT_SOCIAL_CIRCLE                                    0.014640
DAYS_LAST_PHONE_CHANGE                                     -0.121967
AMT_REQ_CREDIT_BUREAU_HOUR                                  0.009576
AMT_REQ_CREDIT_BUREAU_DAY                                   0.006455
AMT_REQ_CREDIT_BUREAU_WEEK                                  0.035362
AMT_REQ_CREDIT_BUREAU_MON                                   0.038940
AMT_REQ_CREDIT_BUREAU_QRT                                   0.143058
AMT_REQ_CREDIT_BUREAU_YEAR                                  1.000000

[43 rows x 43 columns]
```

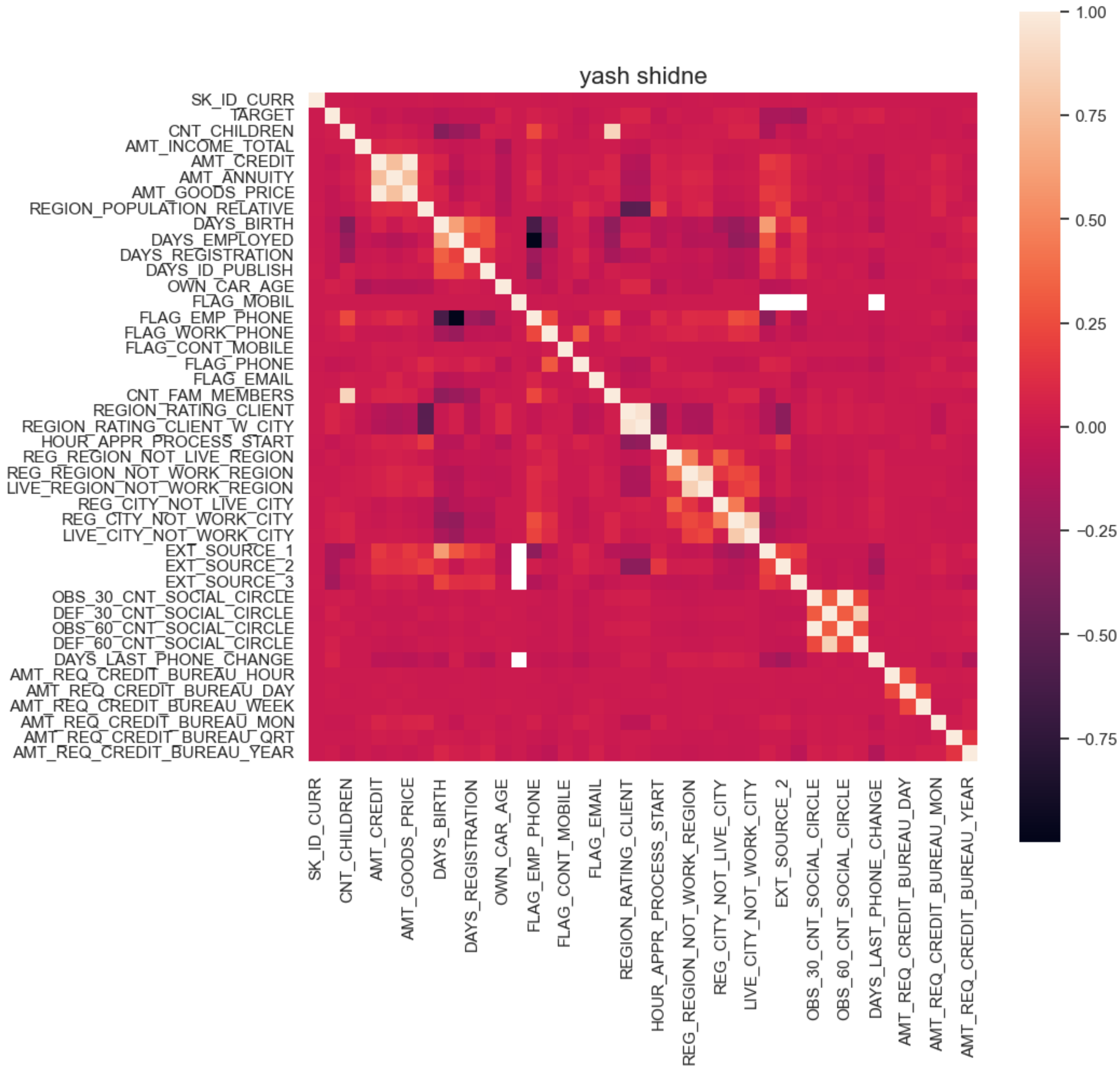
```
print(correlation["AMT_ANNUITY"].sort_values(ascending = False), "\n")
```

```
AMT_ANNUITY                        1.000000
AMT_GOODS_PRICE                    0.774434
AMT_CREDIT                         0.769499
EXT_SOURCE_2                       0.128928
EXT_SOURCE_1                       0.115437
REGION_POPULATION_RELATIVE         0.115112
FLAG_EMP_PHONE                     0.109570
AMT_INCOME_TOTAL                   0.083009
REG_REGION_NOT_WORK_REGION         0.081295
CNT_FAM_MEMBERS                    0.077353
LIVE_REGION_NOT_WORK_REGION        0.074849
```

```
FLAG_EMAIL                      0.065896
HOUR_APPR_PROCESS_START         0.053275
REG_REGION_NOT_LIVE_REGION      0.044803
AMT_REQ_CREDIT_BUREAU_MON       0.039915
CNT_CHILDREN                    0.026179
EXT_SOURCE_3                    0.023602
FLAG_CONT_MOBILE                0.023111
AMT_REQ_CREDIT_BUREAU_WEEK      0.019918
AMT_REQ_CREDIT_BUREAU_HOUR      0.012014
LIVE_CITY_NOT_WORK_CITY         0.010940
AMT_REQ_CREDIT_BUREAU_QRT       0.009306
AMT_REQ_CREDIT_BUREAU_DAY       0.007011
FLAG_PHONE                      0.005127
REG_CITY_NOT_WORK_CITY          0.001061
FLAG_MOBIL                      0.000367
AMT_REQ_CREDIT_BUREAU_YEAR      0.000027
SK_ID_CURR                     -0.002084
DAYS_ID_PUBLISH                -0.006716
REG_CITY_NOT_LIVE_CITY         -0.006721
DAYS_BIRTH                     -0.007712
OBS_60_CNT_SOCIAL_CIRCLE       -0.009012
OBS_30_CNT_SOCIAL_CIRCLE       -0.009325
TARGET                         -0.012399
FLAG_WORK_PHONE                -0.020965
DEF_30_CNT_SOCIAL_CIRCLE       -0.021819
DEF_60_CNT_SOCIAL_CIRCLE       -0.025276
DAYS_REGISTRATION              -0.033219
DAYS_LAST_PHONE_CHANGE         -0.067257
OWN_CAR_AGE                    -0.096866
DAYS_EMPLOYED                  -0.108710
REGION_RATING_CLIENT           -0.125803
REGION_RATING_CLIENT_W_CITY    -0.139322
Name: AMT_ANNUITY, dtype: float64
```

```
f, ax = plt.subplots(figsize=(10, 10))
plt.title("yash shidne", y =1, size = 16)

sns.heatmap(correlation, square = True, vmax = 1, linewidths =
0.000001)
```

```
<AxesSubplot:title={'center':'yash shidne'}>
```




```
# Get the upper triangle of the correlation matrix so that there is no
same correlation values counted twice.
upper_triangle_values =
correlation.where(np.triu(np.ones(correlation.shape),
k=1).astype(np.bool))

# Find the top 5 highest correlations
top_positive_correlations =
upper_triangle_values.unstack().sort_values(ascending=False).head(5)

# Find the top 5 lowest correlations
top_negative_correlations =
upper_triangle_values.unstack().sort_values(ascending=True).head(5)
```

```python
print("Top 5 highest correlations:")
print(top_positive_correlations)

print("\nTop 5 lowest correlations:")
print(top_negative_correlations)
```

```
Top 5 highest correlations:
OBS_60_CNT_SOCIAL_CIRCLE     OBS_30_CNT_SOCIAL_CIRCLE      0.998331
AMT_GOODS_PRICE              AMT_CREDIT                    0.986944
REGION_RATING_CLIENT_W_CITY  REGION_RATING_CLIENT          0.950710
CNT_FAM_MEMBERS              CNT_CHILDREN                  0.880430
LIVE_REGION_NOT_WORK_REGION  REG_REGION_NOT_WORK_REGION    0.857142
dtype: float64

Top 5 lowest correlations:
FLAG_EMP_PHONE               DAYS_EMPLOYED                -0.999746
                             DAYS_BIRTH                   -0.617703
REGION_RATING_CLIENT         REGION_POPULATION_RELATIVE   -0.532667
REGION_RATING_CLIENT_W_CITY  REGION_POPULATION_RELATIVE   -0.530439
DAYS_BIRTH                   CNT_CHILDREN                 -0.329264
dtype: float64

C:\Users\Yash\AppData\Local\Temp\ipykernel_8388\2900829348.py:2:
DeprecationWarning: `np.bool` is a deprecated alias for the builtin
`bool`. To silence this warning, use `bool` by itself. Doing this will
not modify any behavior and is safe. If you specifically wanted the
numpy scalar type, use `np.bool_` here.
Deprecated in NumPy 1.20; for more details and guidance:
https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations
  upper_triangle_values =
correlation.where(np.triu(np.ones(correlation.shape),
k=1).astype(np.bool))
```