```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
from scipy import stats

prev = pd.read_csv('previous_application.csv')
post = pd.read_csv('application_data.csv')

prev.head()
```

{"type":"dataframe","variable_name":"prev"}

```python
post.head()
```

{"type":"dataframe","variable_name":"post"}

```python
prev.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1670214 entries, 0 to 1670213
Data columns (total 37 columns):
 #   Column                       Non-Null Count    Dtype
---  ------                       --------------    -----
 0   SK_ID_PREV                   1670214 non-null  int64
 1   SK_ID_CURR                   1670214 non-null  int64
 2   NAME_CONTRACT_TYPE           1670214 non-null  object
 3   AMT_ANNUITY                  1297979 non-null  float64
 4   AMT_APPLICATION              1670214 non-null  float64
 5   AMT_CREDIT                   1670213 non-null  float64
 6   AMT_DOWN_PAYMENT             774370 non-null   float64
 7   AMT_GOODS_PRICE              1284699 non-null  float64
 8   WEEKDAY_APPR_PROCESS_START   1670214 non-null  object
 9   HOUR_APPR_PROCESS_START      1670214 non-null  int64
 10  FLAG_LAST_APPL_PER_CONTRACT  1670214 non-null  object
 11  NFLAG_LAST_APPL_IN_DAY       1670214 non-null  int64
 12  RATE_DOWN_PAYMENT            774370 non-null   float64
 13  RATE_INTEREST_PRIMARY        5951 non-null     float64
 14  RATE_INTEREST_PRIVILEGED     5951 non-null     float64
 15  NAME_CASH_LOAN_PURPOSE       1670214 non-null  object
 16  NAME_CONTRACT_STATUS         1670214 non-null  object
 17  DAYS_DECISION                1670214 non-null  int64
 18  NAME_PAYMENT_TYPE            1670214 non-null  object
 19  CODE_REJECT_REASON           1670214 non-null  object
 20  NAME_TYPE_SUITE              849809 non-null   object
 21  NAME_CLIENT_TYPE             1670214 non-null  object
 22  NAME_GOODS_CATEGORY          1670214 non-null  object
 23  NAME_PORTFOLIO               1670214 non-null  object
 24  NAME_PRODUCT_TYPE            1670214 non-null  object
 25  CHANNEL_TYPE                 1670214 non-null  object
```

```
 26  SELLERPLACE_AREA             1670214 non-null  int64
 27  NAME_SELLER_INDUSTRY         1670214 non-null  object
 28  CNT_PAYMENT                  1297984 non-null  float64
 29  NAME_YIELD_GROUP             1670214 non-null  object
 30  PRODUCT_COMBINATION          1669868 non-null  object
 31  DAYS_FIRST_DRAWING           997149 non-null   float64
 32  DAYS_FIRST_DUE               997149 non-null   float64
 33  DAYS_LAST_DUE_1ST_VERSION    997149 non-null   float64
 34  DAYS_LAST_DUE                997149 non-null   float64
 35  DAYS_TERMINATION             997149 non-null   float64
 36  NFLAG_INSURED_ON_APPROVAL    997149 non-null   float64
dtypes: float64(15), int64(6), object(16)
memory usage: 471.5+ MB
```

```
post.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 307511 entries, 0 to 307510
Columns: 122 entries, SK_ID_CURR to AMT_REQ_CREDIT_BUREAU_YEAR
dtypes: float64(65), int64(41), object(16)
memory usage: 286.2+ MB
```

```
prev.describe()
```

{"type":"dataframe"}

```
post.describe()
```

{"type":"dataframe"}

```
prev.dtypes
```

```
SK_ID_PREV                        int64
SK_ID_CURR                        int64
NAME_CONTRACT_TYPE               object
AMT_ANNUITY                     float64
AMT_APPLICATION                 float64
AMT_CREDIT                      float64
AMT_DOWN_PAYMENT                float64
AMT_GOODS_PRICE                 float64
WEEKDAY_APPR_PROCESS_START       object
HOUR_APPR_PROCESS_START           int64
FLAG_LAST_APPL_PER_CONTRACT      object
NFLAG_LAST_APPL_IN_DAY            int64
RATE_DOWN_PAYMENT               float64
RATE_INTEREST_PRIMARY           float64
RATE_INTEREST_PRIVILEGED        float64
NAME_CASH_LOAN_PURPOSE           object
NAME_CONTRACT_STATUS             object
DAYS_DECISION                     int64
NAME_PAYMENT_TYPE                object
```

```
CODE_REJECT_REASON               object
NAME_TYPE_SUITE                  object
NAME_CLIENT_TYPE                 object
NAME_GOODS_CATEGORY              object
NAME_PORTFOLIO                   object
NAME_PRODUCT_TYPE                object
CHANNEL_TYPE                     object
SELLERPLACE_AREA                  int64
NAME_SELLER_INDUSTRY             object
CNT_PAYMENT                     float64
NAME_YIELD_GROUP                 object
PRODUCT_COMBINATION              object
DAYS_FIRST_DRAWING              float64
DAYS_FIRST_DUE                  float64
DAYS_LAST_DUE_1ST_VERSION       float64
DAYS_LAST_DUE                   float64
DAYS_TERMINATION                float64
NFLAG_INSURED_ON_APPROVAL       float64
dtype: object
```

```
post.dtypes
```

```
SK_ID_CURR                       int64
TARGET                           int64
NAME_CONTRACT_TYPE              object
CODE_GENDER                     object
FLAG_OWN_CAR                    object
                                 ...
AMT_REQ_CREDIT_BUREAU_DAY      float64
AMT_REQ_CREDIT_BUREAU_WEEK     float64
AMT_REQ_CREDIT_BUREAU_MON      float64
AMT_REQ_CREDIT_BUREAU_QRT      float64
AMT_REQ_CREDIT_BUREAU_YEAR     float64
Length: 122, dtype: object
```

```
prev.columns
```

```
Index(['SK_ID_PREV', 'SK_ID_CURR', 'NAME_CONTRACT_TYPE',
'AMT_ANNUITY',
       'AMT_APPLICATION', 'AMT_CREDIT', 'AMT_DOWN_PAYMENT',
'AMT_GOODS_PRICE',
       'WEEKDAY_APPR_PROCESS_START', 'HOUR_APPR_PROCESS_START',
       'FLAG_LAST_APPL_PER_CONTRACT', 'NFLAG_LAST_APPL_IN_DAY',
       'RATE_DOWN_PAYMENT', 'RATE_INTEREST_PRIMARY',
       'RATE_INTEREST_PRIVILEGED', 'NAME_CASH_LOAN_PURPOSE',
       'NAME_CONTRACT_STATUS', 'DAYS_DECISION', 'NAME_PAYMENT_TYPE',
       'CODE_REJECT_REASON', 'NAME_TYPE_SUITE', 'NAME_CLIENT_TYPE',
       'NAME_GOODS_CATEGORY', 'NAME_PORTFOLIO', 'NAME_PRODUCT_TYPE',
       'CHANNEL_TYPE', 'SELLERPLACE_AREA', 'NAME_SELLER_INDUSTRY',
       'CNT_PAYMENT', 'NAME_YIELD_GROUP', 'PRODUCT_COMBINATION',
```

```
       'DAYS_FIRST_DRAWING', 'DAYS_FIRST_DUE',
'DAYS_LAST_DUE_1ST_VERSION',
       'DAYS_LAST_DUE', 'DAYS_TERMINATION',
'NFLAG_INSURED_ON_APPROVAL'],
      dtype='object')
```

```
post.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)
```

```
# check for missing data in each column
print(prev.isnull().sum())
```

```
SK_ID_PREV                           0
SK_ID_CURR                           0
NAME_CONTRACT_TYPE                   0
AMT_ANNUITY                     372235
AMT_APPLICATION                      0
AMT_CREDIT                           1
AMT_DOWN_PAYMENT                895844
AMT_GOODS_PRICE                 385515
WEEKDAY_APPR_PROCESS_START           0
HOUR_APPR_PROCESS_START              0
FLAG_LAST_APPL_PER_CONTRACT          0
NFLAG_LAST_APPL_IN_DAY               0
RATE_DOWN_PAYMENT               895844
RATE_INTEREST_PRIMARY          1664263
RATE_INTEREST_PRIVILEGED       1664263
NAME_CASH_LOAN_PURPOSE               0
NAME_CONTRACT_STATUS                 0
DAYS_DECISION                        0
NAME_PAYMENT_TYPE                    0
CODE_REJECT_REASON                   0
NAME_TYPE_SUITE                 820405
NAME_CLIENT_TYPE                     0
NAME_GOODS_CATEGORY                  0
NAME_PORTFOLIO                       0
NAME_PRODUCT_TYPE                    0
CHANNEL_TYPE                         0
SELLERPLACE_AREA                     0
```

```
NAME_SELLER_INDUSTRY                      0
CNT_PAYMENT                          372230
NAME_YIELD_GROUP                          0
PRODUCT_COMBINATION                     346
DAYS_FIRST_DRAWING                   673065
DAYS_FIRST_DUE                       673065
DAYS_LAST_DUE_1ST_VERSION            673065
DAYS_LAST_DUE                        673065
DAYS_TERMINATION                     673065
NFLAG_INSURED_ON_APPROVAL            673065
dtype: int64
```

```
print(post.isnull().sum())
```

```
SK_ID_CURR                            0
TARGET                                0
NAME_CONTRACT_TYPE                    0
CODE_GENDER                           0
FLAG_OWN_CAR                          0
                                  ...
AMT_REQ_CREDIT_BUREAU_DAY         41519
AMT_REQ_CREDIT_BUREAU_WEEK        41519
AMT_REQ_CREDIT_BUREAU_MON         41519
AMT_REQ_CREDIT_BUREAU_QRT         41519
AMT_REQ_CREDIT_BUREAU_YEAR        41519
Length: 122, dtype: int64
```

```
missing_val_per_1 = (prev.isnull().sum() / len(prev)) * 100
missing_val_per_1
```

```
SK_ID_PREV                         0.000000
SK_ID_CURR                         0.000000
NAME_CONTRACT_TYPE                 0.000000
AMT_ANNUITY                       22.286665
AMT_APPLICATION                    0.000000
AMT_CREDIT                         0.000060
AMT_DOWN_PAYMENT                  53.636480
AMT_GOODS_PRICE                   23.081773
WEEKDAY_APPR_PROCESS_START         0.000000
HOUR_APPR_PROCESS_START            0.000000
FLAG_LAST_APPL_PER_CONTRACT        0.000000
NFLAG_LAST_APPL_IN_DAY             0.000000
RATE_DOWN_PAYMENT                 53.636480
RATE_INTEREST_PRIMARY             99.643698
RATE_INTEREST_PRIVILEGED          99.643698
NAME_CASH_LOAN_PURPOSE             0.000000
NAME_CONTRACT_STATUS               0.000000
DAYS_DECISION                      0.000000
NAME_PAYMENT_TYPE                  0.000000
CODE_REJECT_REASON                 0.000000
```

```
NAME_TYPE_SUITE                 49.119754
NAME_CLIENT_TYPE                 0.000000
NAME_GOODS_CATEGORY              0.000000
NAME_PORTFOLIO                   0.000000
NAME_PRODUCT_TYPE                0.000000
CHANNEL_TYPE                     0.000000
SELLERPLACE_AREA                 0.000000
NAME_SELLER_INDUSTRY             0.000000
CNT_PAYMENT                     22.286366
NAME_YIELD_GROUP                 0.000000
PRODUCT_COMBINATION              0.020716
DAYS_FIRST_DRAWING              40.298129
DAYS_FIRST_DUE                  40.298129
DAYS_LAST_DUE_1ST_VERSION       40.298129
DAYS_LAST_DUE                   40.298129
DAYS_TERMINATION                40.298129
NFLAG_INSURED_ON_APPROVAL       40.298129
dtype: float64
```

```
missing_val_per_2 = (post.isnull().sum() / len(post)) * 100
missing_val_per_2
```

```
SK_ID_CURR                      0.000000
TARGET                          0.000000
NAME_CONTRACT_TYPE              0.000000
CODE_GENDER                     0.000000
FLAG_OWN_CAR                    0.000000
                                  ...
AMT_REQ_CREDIT_BUREAU_DAY      13.501631
AMT_REQ_CREDIT_BUREAU_WEEK     13.501631
AMT_REQ_CREDIT_BUREAU_MON      13.501631
AMT_REQ_CREDIT_BUREAU_QRT      13.501631
AMT_REQ_CREDIT_BUREAU_YEAR     13.501631
Length: 122, dtype: float64
```

```
prev = prev.dropna()
prev.head()
```

{"type":"dataframe","variable_name":"prev"}

```
prev.isnull().sum()
```

```
SK_ID_PREV                      0
SK_ID_CURR                      0
NAME_CONTRACT_TYPE              0
AMT_ANNUITY                     0
AMT_APPLICATION                 0
AMT_CREDIT                      0
AMT_DOWN_PAYMENT                0
AMT_GOODS_PRICE                 0
WEEKDAY_APPR_PROCESS_START      0
```

```
HOUR_APPR_PROCESS_START        0
FLAG_LAST_APPL_PER_CONTRACT    0
NFLAG_LAST_APPL_IN_DAY         0
RATE_DOWN_PAYMENT              0
RATE_INTEREST_PRIMARY          0
RATE_INTEREST_PRIVILEGED       0
NAME_CASH_LOAN_PURPOSE         0
NAME_CONTRACT_STATUS           0
DAYS_DECISION                  0
NAME_PAYMENT_TYPE              0
CODE_REJECT_REASON             0
NAME_TYPE_SUITE                0
NAME_CLIENT_TYPE               0
NAME_GOODS_CATEGORY            0
NAME_PORTFOLIO                 0
NAME_PRODUCT_TYPE              0
CHANNEL_TYPE                   0
SELLERPLACE_AREA               0
NAME_SELLER_INDUSTRY           0
CNT_PAYMENT                    0
NAME_YIELD_GROUP               0
PRODUCT_COMBINATION            0
DAYS_FIRST_DRAWING             0
DAYS_FIRST_DUE                 0
DAYS_LAST_DUE_1ST_VERSION      0
DAYS_LAST_DUE                  0
DAYS_TERMINATION               0
NFLAG_INSURED_ON_APPROVAL      0
dtype: int64
```

```
missing1_per = (prev.isnull().sum() / len(prev)) * 100
missing1_per
```

```
SK_ID_PREV                     0.0
SK_ID_CURR                     0.0
NAME_CONTRACT_TYPE             0.0
AMT_ANNUITY                    0.0
AMT_APPLICATION                0.0
AMT_CREDIT                     0.0
AMT_DOWN_PAYMENT               0.0
AMT_GOODS_PRICE                0.0
WEEKDAY_APPR_PROCESS_START     0.0
HOUR_APPR_PROCESS_START        0.0
FLAG_LAST_APPL_PER_CONTRACT    0.0
NFLAG_LAST_APPL_IN_DAY         0.0
RATE_DOWN_PAYMENT              0.0
RATE_INTEREST_PRIMARY          0.0
RATE_INTEREST_PRIVILEGED       0.0
NAME_CASH_LOAN_PURPOSE         0.0
NAME_CONTRACT_STATUS           0.0
```

```
DAYS_DECISION                         0.0
NAME_PAYMENT_TYPE                     0.0
CODE_REJECT_REASON                    0.0
NAME_TYPE_SUITE                       0.0
NAME_CLIENT_TYPE                      0.0
NAME_GOODS_CATEGORY                   0.0
NAME_PORTFOLIO                        0.0
NAME_PRODUCT_TYPE                     0.0
CHANNEL_TYPE                          0.0
SELLERPLACE_AREA                      0.0
NAME_SELLER_INDUSTRY                  0.0
CNT_PAYMENT                           0.0
NAME_YIELD_GROUP                      0.0
PRODUCT_COMBINATION                   0.0
DAYS_FIRST_DRAWING                    0.0
DAYS_FIRST_DUE                        0.0
DAYS_LAST_DUE_1ST_VERSION             0.0
DAYS_LAST_DUE                         0.0
DAYS_TERMINATION                      0.0
NFLAG_INSURED_ON_APPROVAL             0.0
dtype: float64
```

```
post = post.dropna()
post.head()
```

{"type":"dataframe","variable_name":"post"}

```
missing2_per = (post.isnull().sum() / len(post)) * 100
missing2_per
```

```
SK_ID_CURR                           0.0
TARGET                               0.0
NAME_CONTRACT_TYPE                   0.0
CODE_GENDER                          0.0
FLAG_OWN_CAR                         0.0
                                    ...
AMT_REQ_CREDIT_BUREAU_DAY            0.0
AMT_REQ_CREDIT_BUREAU_WEEK           0.0
AMT_REQ_CREDIT_BUREAU_MON            0.0
AMT_REQ_CREDIT_BUREAU_QRT            0.0
AMT_REQ_CREDIT_BUREAU_YEAR           0.0
Length: 122, dtype: float64
```

```
post.fillna(method='ffill', inplace=True)
post.head()
```

```
post['CODE_GENDER'].fillna('XNA', inplace=True)
post.head()
```

{"type":"dataframe","variable_name":"post"}

DIVISION OF DATA SET INTO TWO PARTS

```
default = post[post['TARGET'] == 1] # had difficulties
non_default = post[post["TARGET"] == 0]

default.head()
```


{"type":"dataframe","variable_name":"default"}


```
default.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 526 entries, 255 to 307407
Columns: 122 entries, SK_ID_CURR to AMT_REQ_CREDIT_BUREAU_YEAR
dtypes: float64(65), int64(41), object(16)
memory usage: 505.5+ KB
```

```
default.describe()
```


{"type":"dataframe"}


```
non_default.head()
```


{"type":"dataframe","variable_name":"non_default"}


```
non_default.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 8076 entries, 71 to 307482
Columns: 122 entries, SK_ID_CURR to AMT_REQ_CREDIT_BUREAU_YEAR
dtypes: float64(65), int64(41), object(16)
memory usage: 7.6+ MB
```

```
default.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)
```

DATA ANALYSIS AND VISUALISATION OF "POST"

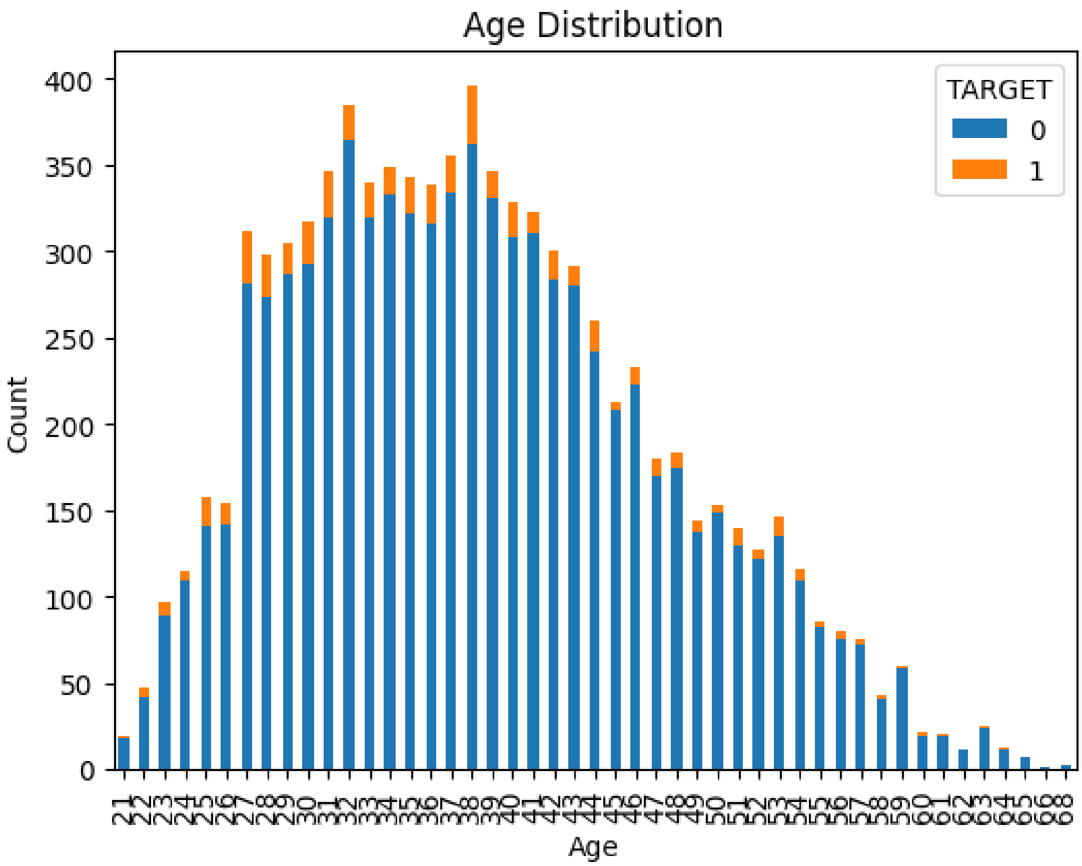
```
# calculating age // feature engineering
post['AGE'] = -post['DAYS_BIRTH'] // 365
print(post[['DAYS_BIRTH', 'AGE']])
```

```
        DAYS_BIRTH  AGE
71          -15406   42
124         -16282   44
152         -11375   31
161         -13972   38
255         -11356   31
...            ...  ...
307358      -15006   41
307359      -14007   38
307407      -11407   31
307456      -20246   55
307482      -14106   38

[8602 rows x 2 columns]
```

```
# AGE

age = post.groupby('AGE')['TARGET'].value_counts().unstack()
print(age)

plt.figure(figsize=(20,15))
age.plot(kind='bar', stacked=True)
plt.title('Age Distribution')
plt.xlabel('Age')
plt.ylabel('Count')
plt.show()
```

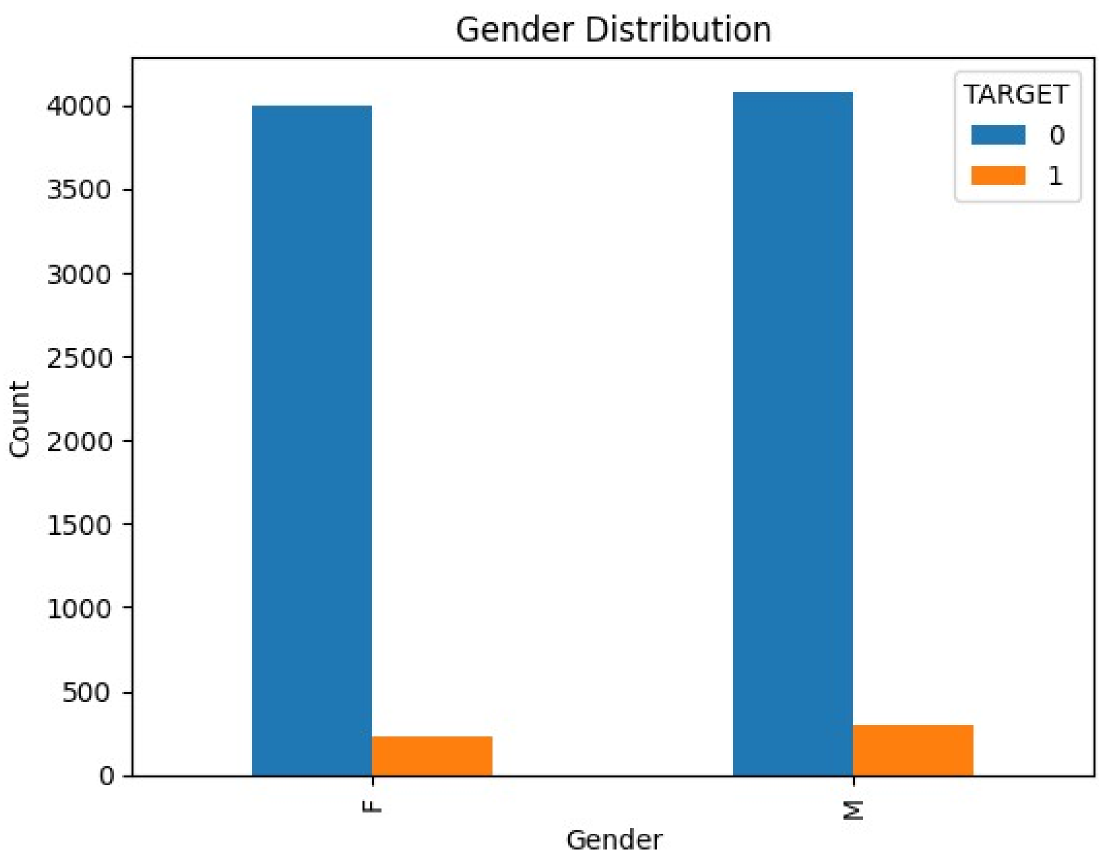
```
TARGET      0     1
AGE
21       18.0   1.0
22       42.0   5.0
23       89.0   8.0
24      109.0   6.0
25      141.0  17.0
26      142.0  12.0
27      281.0  31.0
28      273.0  25.0
29      287.0  18.0
30      293.0  24.0
31      319.0  28.0
32      365.0  20.0
33      320.0  20.0
34      333.0  16.0
35      322.0  21.0
36      316.0  23.0
37      334.0  22.0
38      362.0  34.0
39      331.0  16.0
40      308.0  20.0
41      310.0  13.0
```

```
42      284.0  16.0
43      280.0  11.0
44      242.0  18.0
45      208.0   5.0
46      223.0  10.0
47      170.0  10.0
48      174.0   9.0
49      137.0   7.0
50      148.0   5.0
51      129.0  10.0
52      121.0   6.0
53      135.0  11.0
54      109.0   7.0
55       82.0   3.0
56       75.0   5.0
57       72.0   3.0
58       40.0   3.0
59       58.0   2.0
60       19.0   2.0
61       19.0   1.0
62       11.0   NaN
63       24.0   1.0
64       11.0   1.0
65        7.0   NaN
66        1.0   NaN
68        2.0   NaN

<Figure size 2000x1500 with 0 Axes>
```



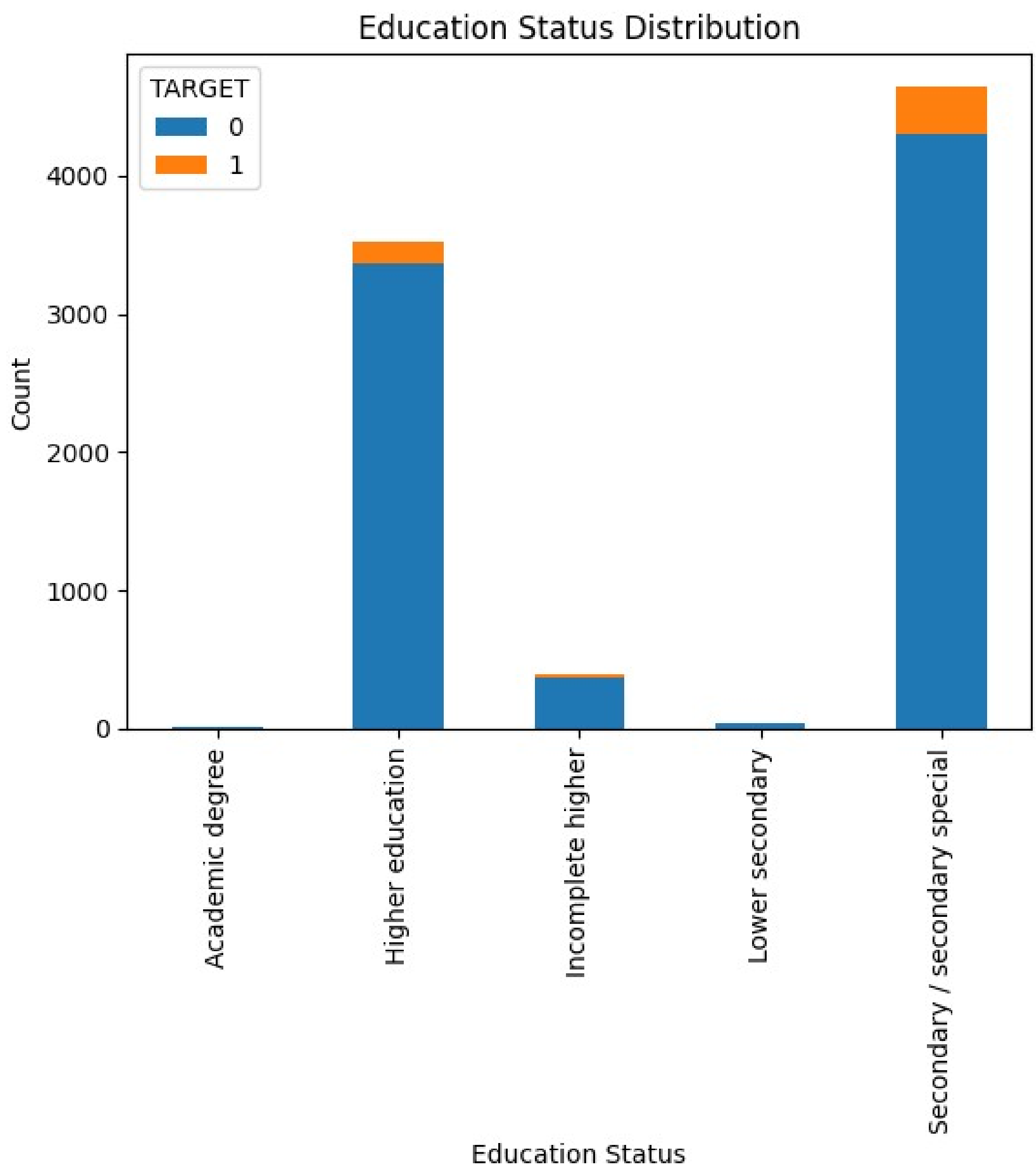

```
# GENDER
post = post[post['CODE_GENDER'] != 'XNA']
gender = post.groupby('CODE_GENDER')
['TARGET'].value_counts().unstack()
print(gender)

gender.plot(kind='bar', stacked=False)
plt.title('Gender Distribution')
plt.xlabel('Gender')
plt.ylabel('Count')
```

```
TARGET          0    1
CODE_GENDER
F            3997  224
M            4079  302

Text(0, 0.5, 'Count')
```



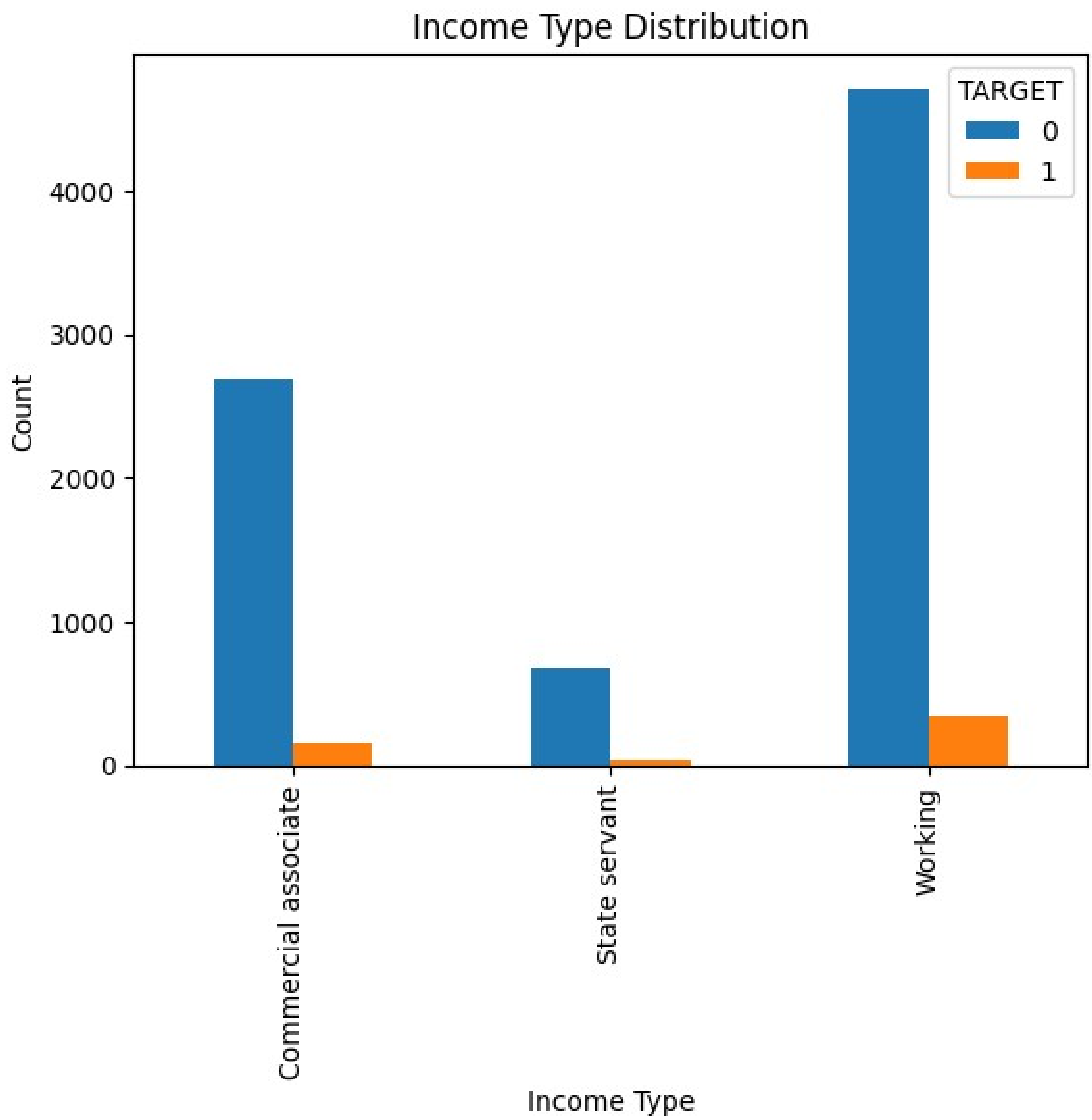

```
# EDUCATION STATUS

edu =  post.groupby('NAME_EDUCATION_TYPE')
['TARGET'].value_counts().unstack()
print(edu)

edu.plot(kind='bar', stacked=True)
plt.title('Education Status Distribution')
plt.xlabel('Education Status')
plt.ylabel('Count')
```

```
TARGET                              0      1
NAME_EDUCATION_TYPE
Academic degree                   6.0    NaN
Higher education               3364.0  159.0
Incomplete higher               370.0   22.0
Lower secondary                  32.0    3.0
Secondary / secondary special  4304.0  342.0

Text(0, 0.5, 'Count')
```




```
# INCOME TYPE // bu boyle gosterilmez
income_type = post.groupby('NAME_INCOME_TYPE')
['TARGET'].value_counts().unstack()
print(income_type)

income_type.plot(kind='bar', stacked=False)
plt.title('Income Type Distribution')
plt.xlabel('Income Type')
plt.ylabel('Count')

TARGET                        0     1
NAME_INCOME_TYPE
```

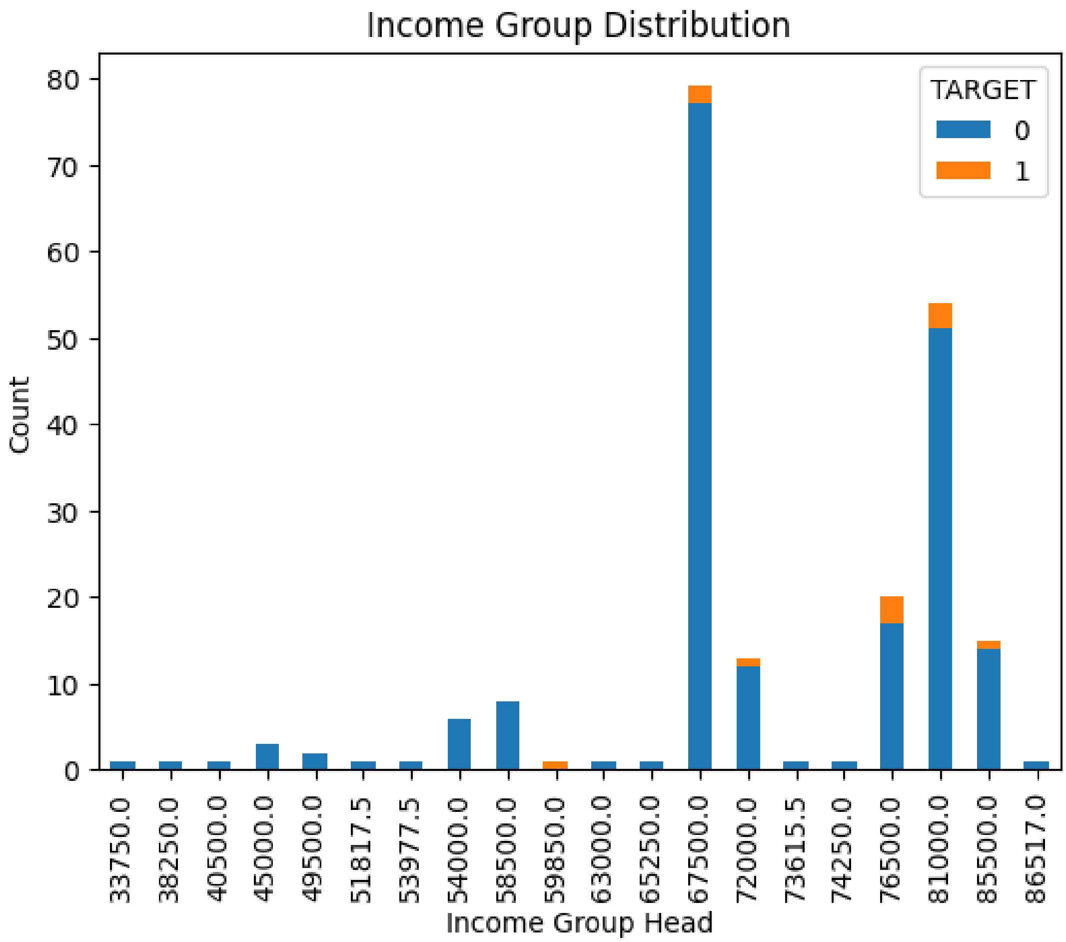
```
Commercial associate  2683  153
State servant          680   36
Working               4713  337

Text(0, 0.5, 'Count')
```



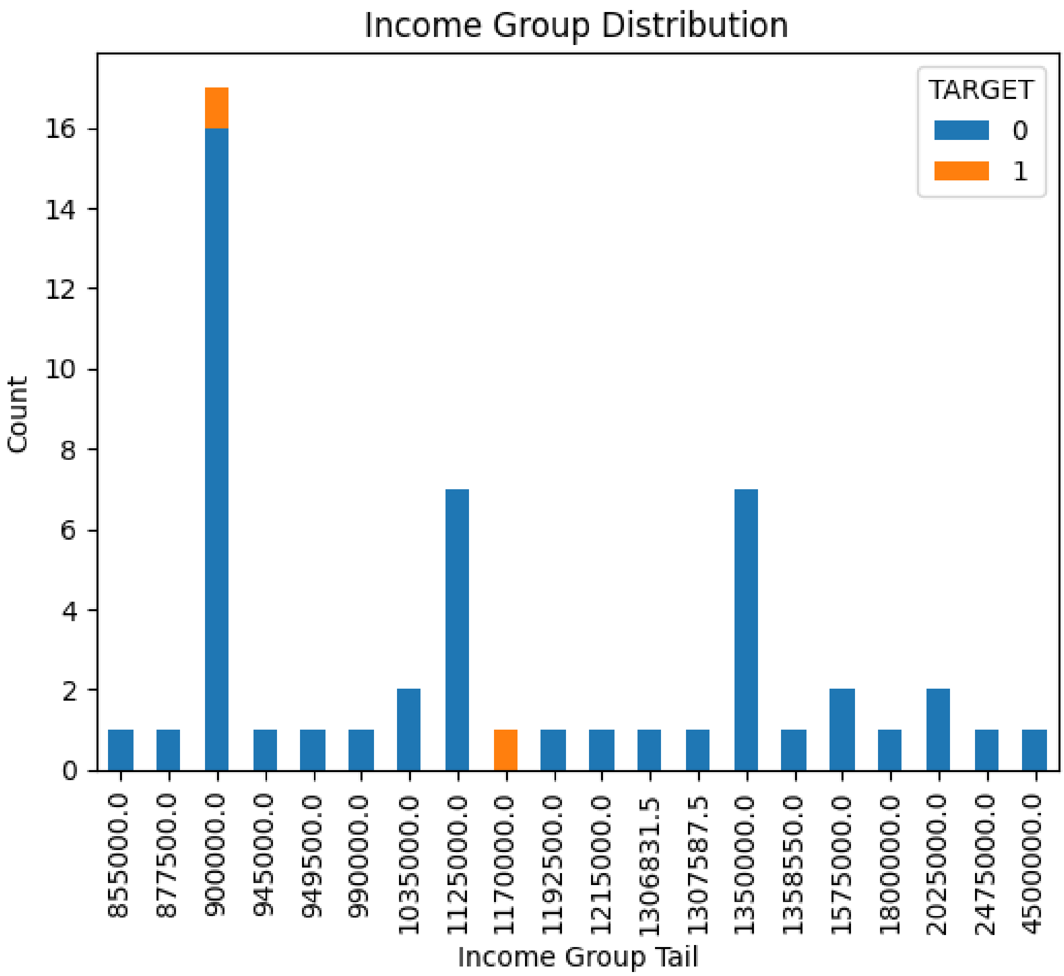

```
# income group
income_group = post.groupby('AMT_INCOME_TOTAL')
['TARGET'].value_counts().unstack()
print(income_group)


income_group.head(20).plot(kind='bar', stacked=True)
plt.title('Income Group Distribution')
plt.xlabel('Income Group Head')
```

```
plt.ylabel('Count')

income_group.tail(20).plot(kind='bar', stacked=True)
plt.title('Income Group Distribution')
plt.xlabel('Income Group Tail')
plt.ylabel('Count')
plt.show()
```

```
TARGET                0   1
AMT_INCOME_TOTAL
33750.0             1.0 NaN
38250.0             1.0 NaN
40500.0             1.0 NaN
45000.0             3.0 NaN
49500.0             2.0 NaN
...                 ...  ..
1575000.0           2.0 NaN
1800000.0           1.0 NaN
2025000.0           2.0 NaN
2475000.0           1.0 NaN
4500000.0           1.0 NaN

[266 rows x 2 columns]
```

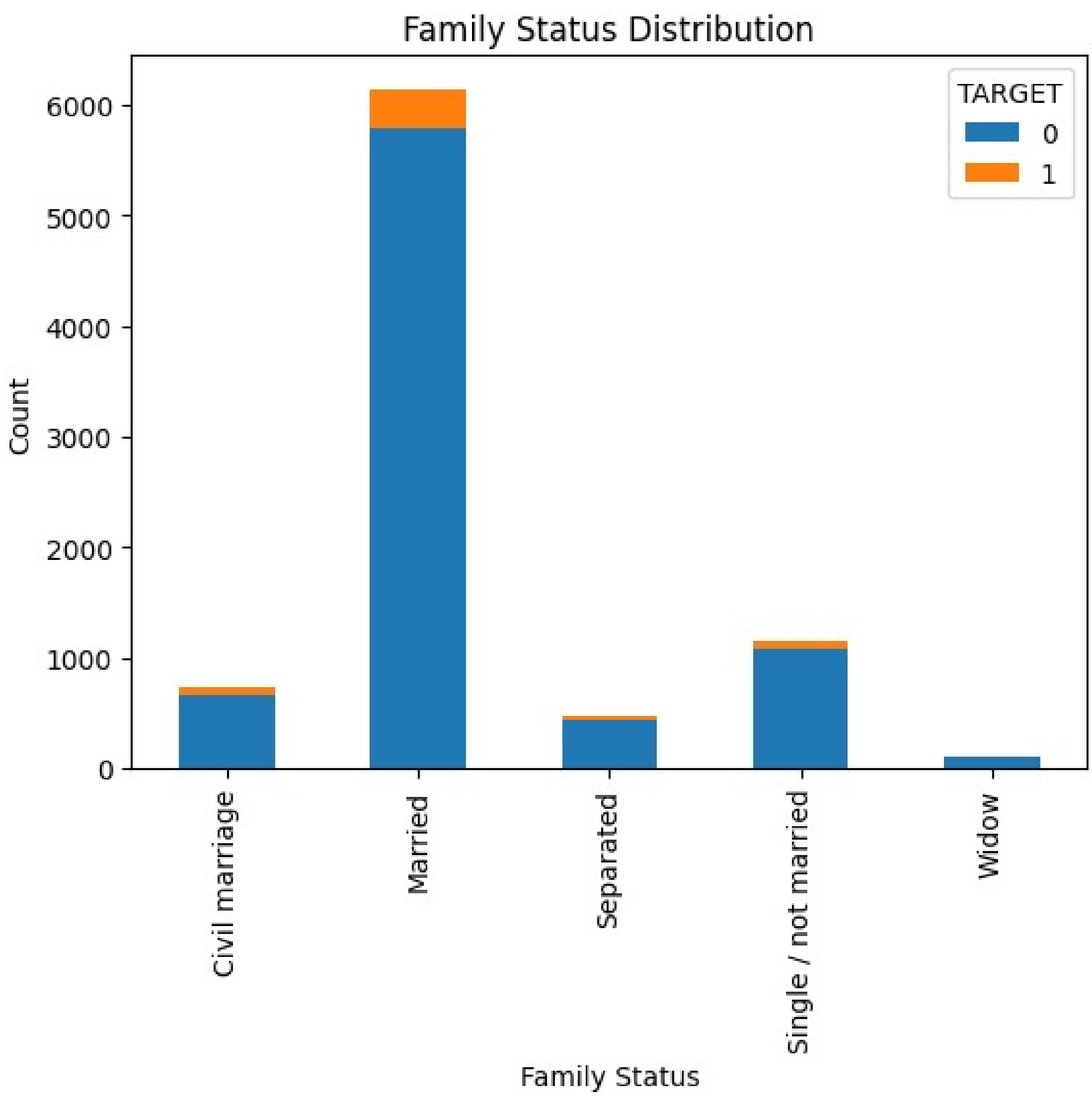


Income Group Distribution
TARGET
0
1
Count
0
10
20
30
40
50
60
70
80
33750.0
38250.0
40500.0
45000.0
49500.0
51817.5
53977.5
54000.0
58500.0
59850.0
63000.0
65250.0
67500.0
72000.0
73615.5
74250.0
76500.0
81000.0
85500.0
86517.0
Income Group Head




```
# FAMILY STATUS

family_status = post.groupby('NAME_FAMILY_STATUS')
['TARGET'].value_counts().unstack()
print(family_status)

family_status.plot(kind='bar', stacked=True)
plt.title('Family Status Distribution')
plt.xlabel('Family Status')
plt.ylabel('Count')
plt.show()
```

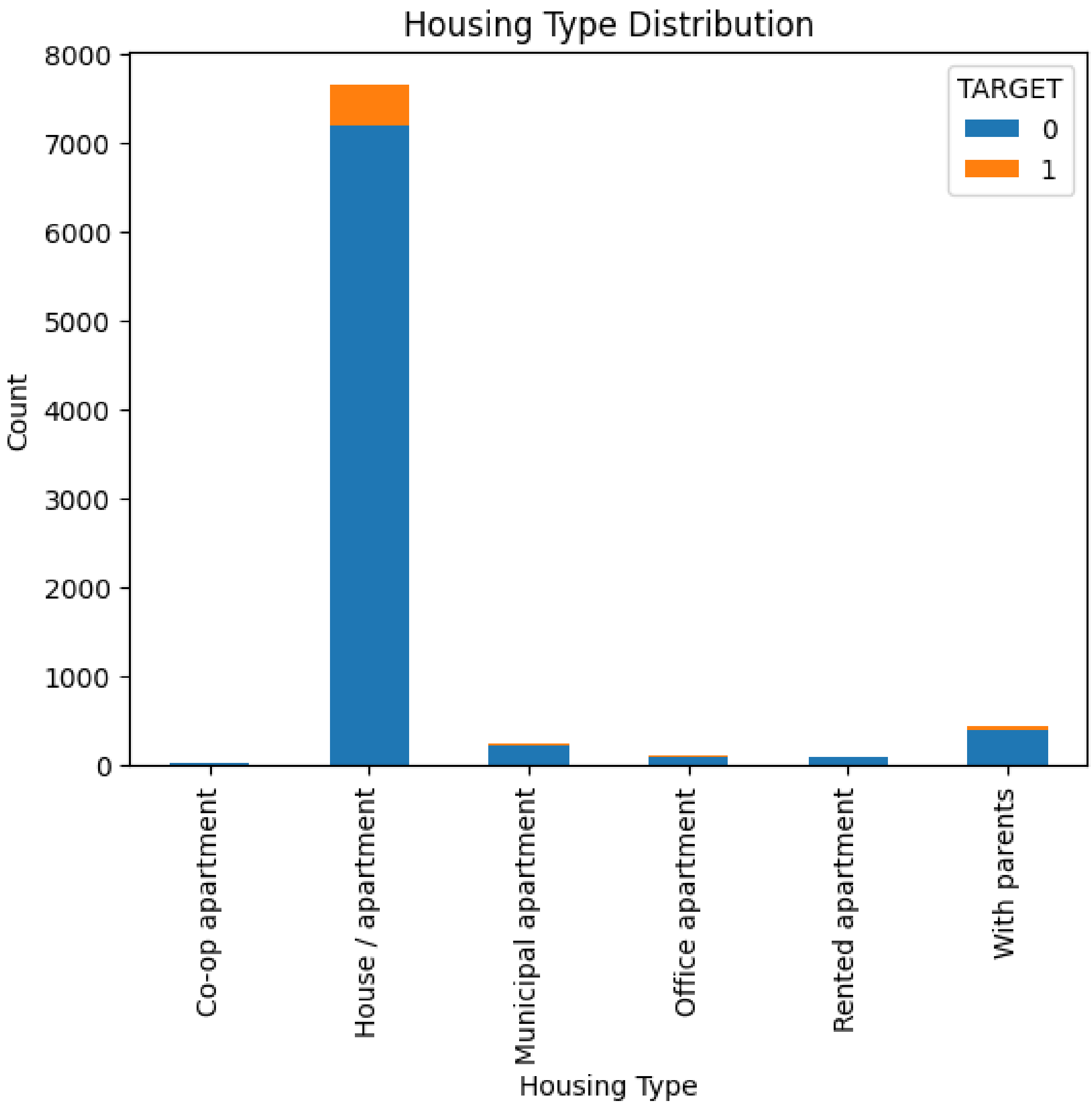
```
TARGET                    0    1
NAME_FAMILY_STATUS
Civil marriage          669   64
Married                5796  343
Separated               437   33
Single / not married   1077   82
Widow                    97    4
```




```
# HOUSING TYPE

housing_type = post.groupby('NAME_HOUSING_TYPE')
['TARGET'].value_counts().unstack()
print(housing_type)

housing_type.plot(kind='bar', stacked=True)
plt.title('Housing Type Distribution')
plt.xlabel('Housing Type')
plt.ylabel('Count')
plt.show()
```

```
TARGET                  0    1
NAME_HOUSING_TYPE
Co-op apartment        31    4
House / apartment    7201  447
```

```
Municipal apartment    233   23
Office apartment       105    5
Rented apartment        95    3
With parents           411   44
```



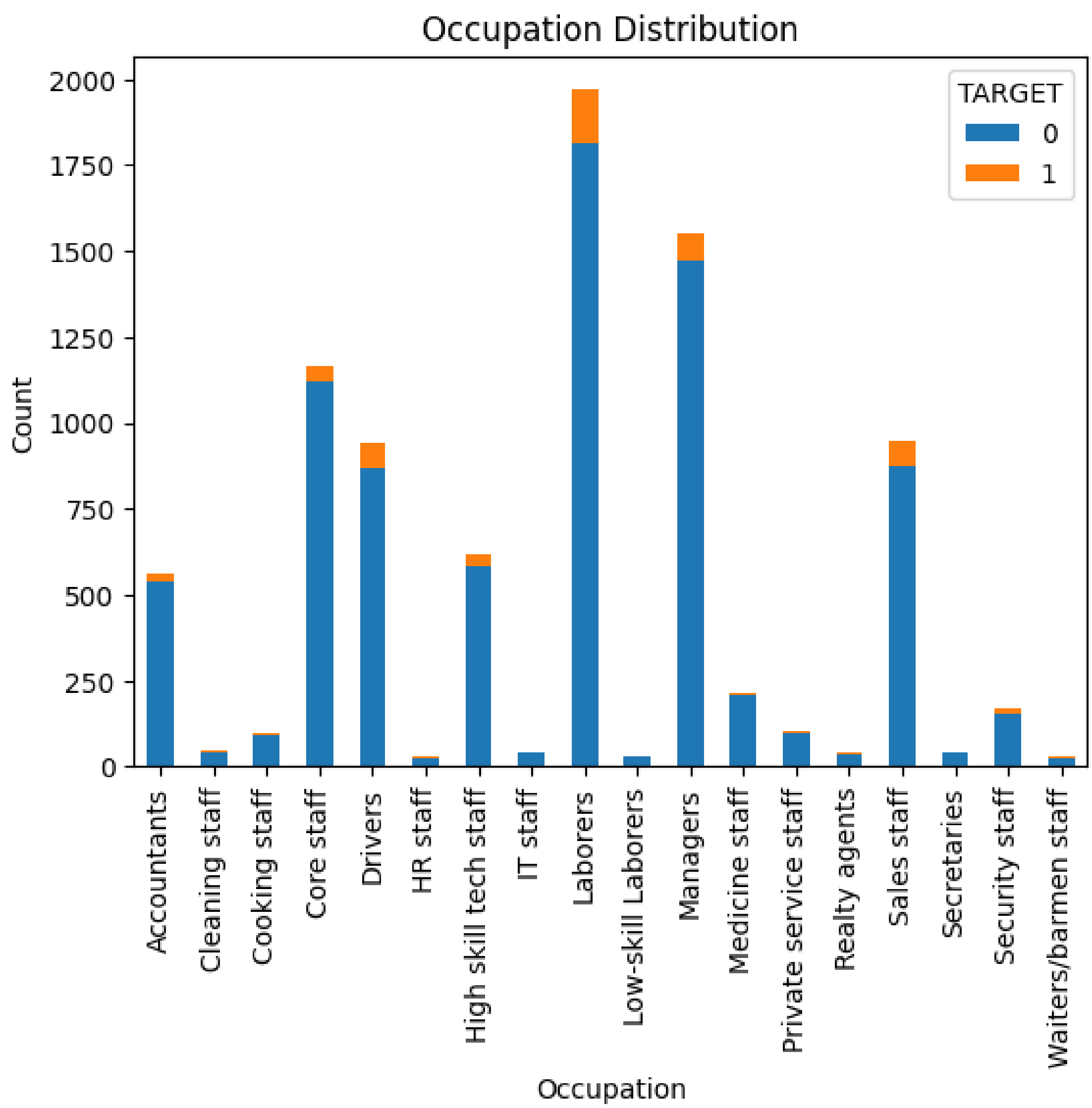

```
post['HOUSING_TYPE'] = post['NAME_HOUSING_TYPE']
post['HOUSING_TYPE#'] =
post['NAME_HOUSING_TYPE'].astype('category').cat.codes
post.head()
```

{"type":"dataframe","variable_name":"post"}

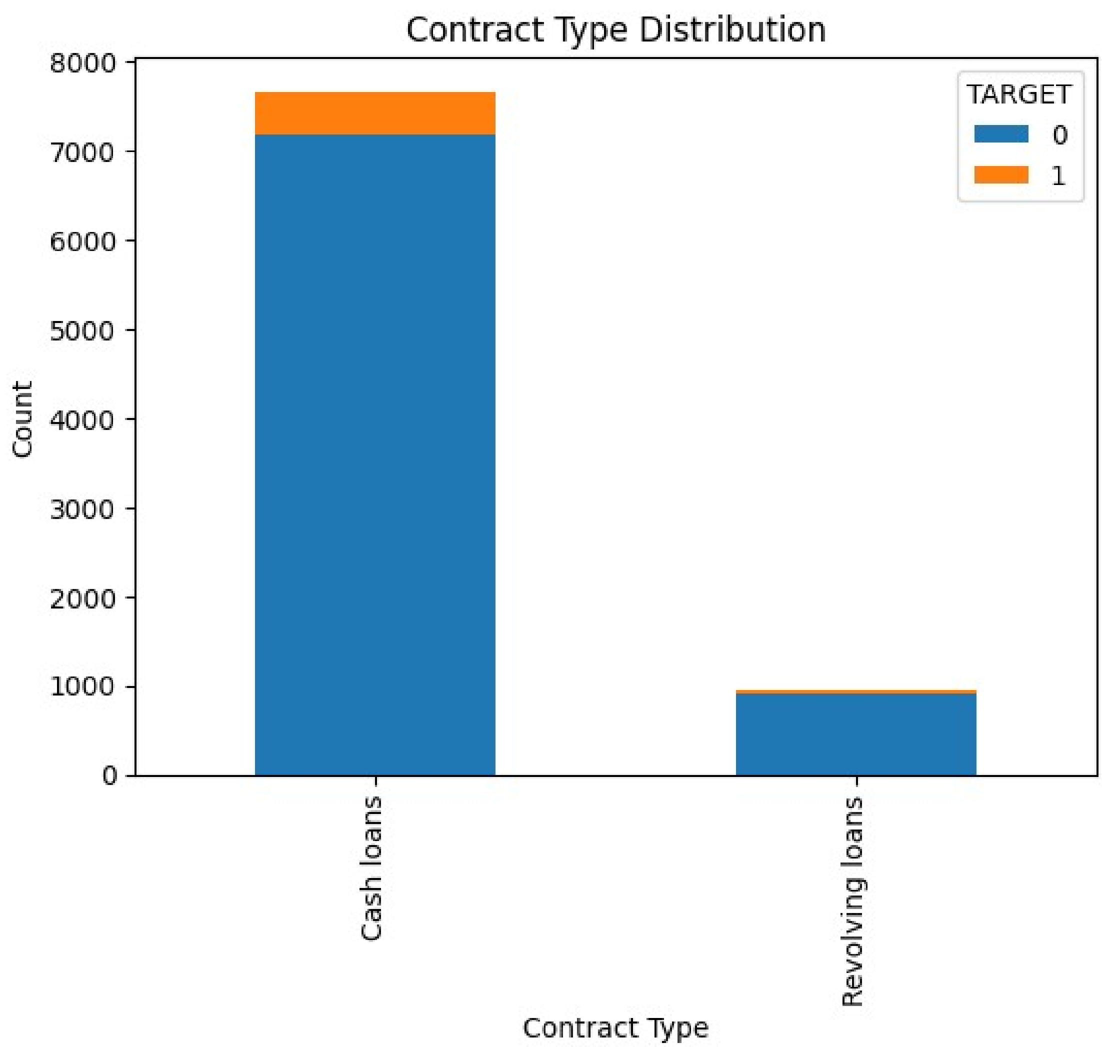
```
# occupation

occupation = post.groupby('OCCUPATION_TYPE')
```

```
['TARGET'].value_counts().unstack()
print(occupation)

occupation.plot(kind='bar', stacked=True)
plt.title('Occupation Distribution')
plt.xlabel('Occupation')
plt.ylabel('Count')
plt.show()
```

```
TARGET                     0    1
OCCUPATION_TYPE
Accountants              541   20
Cleaning staff            43    3
Cooking staff             92    6
Core staff              1119   47
Drivers                  868   72
HR staff                  27    1
High skill tech staff    586   32
IT staff                  41    3
Laborers                1812  155
Low-skill Laborers        30    3
Managers                1474   76
Medicine staff           207   10
Private service staff     98    4
Realty agents             38    3
Sales staff              875   72
Secretaries               43    1
Security staff           156   15
Waiters/barmen staff      26    3
```



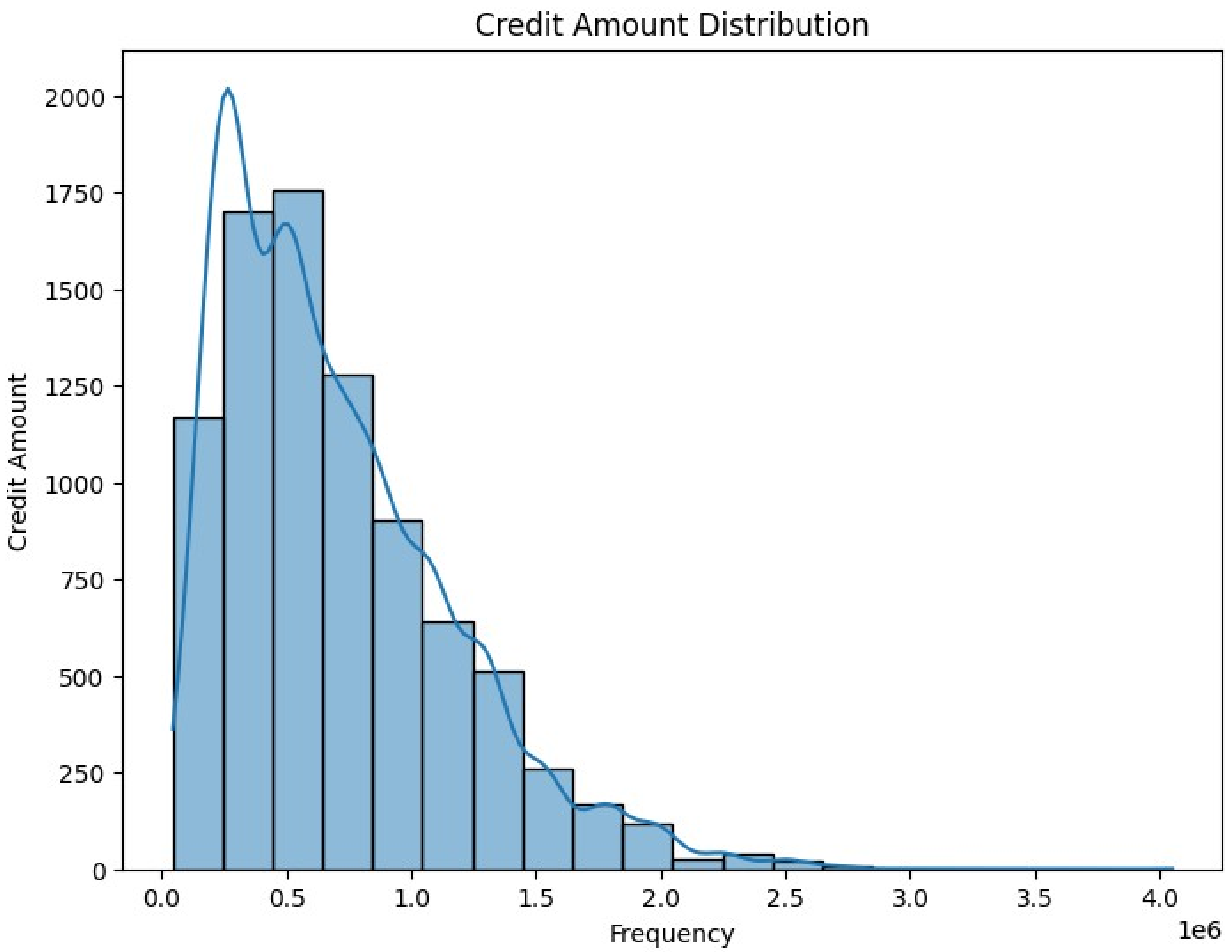

```
# contract type

contract_type = post.groupby('NAME_CONTRACT_TYPE')
['TARGET'].value_counts().unstack()
print(contract_type)

contract_type.plot(kind='bar', stacked=True)
plt.title('Contract Type Distribution')
plt.xlabel('Contract Type')
plt.ylabel('Count')
plt.show()

TARGET                 0    1
NAME_CONTRACT_TYPE
Cash loans          7162  498
Revolving loans      914   28
```



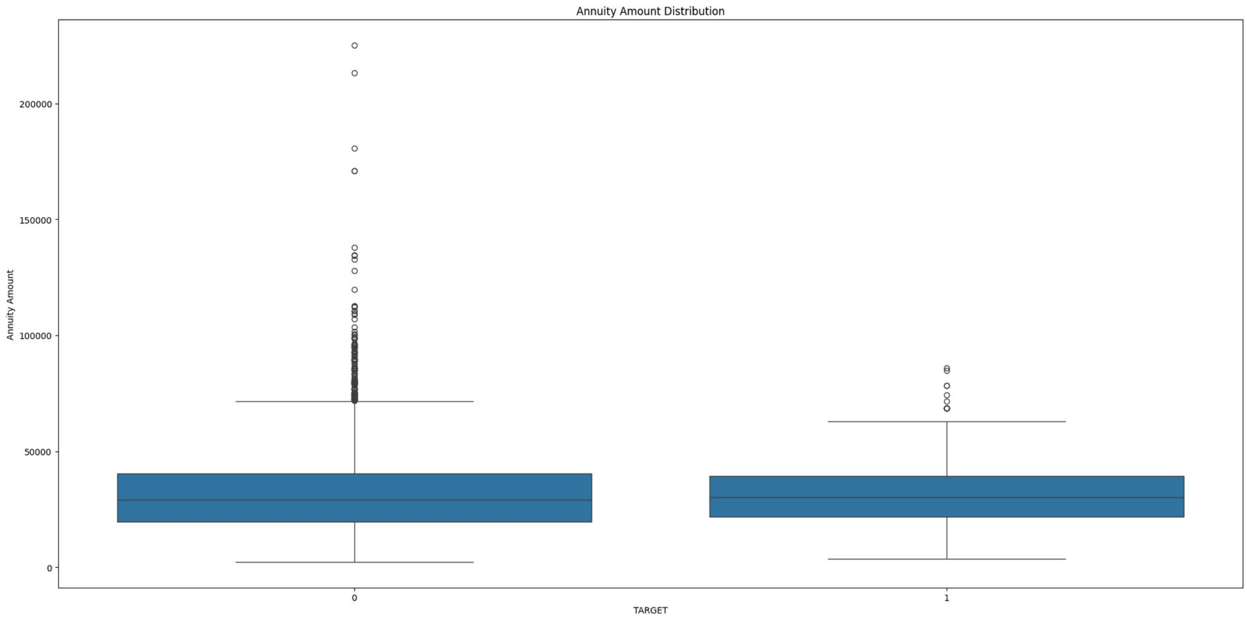

```
credit_amount = post.groupby('AMT_CREDIT')
['TARGET'].value_counts().unstack()
print(credit_amount)
plt.figure(figsize=(8, 6))
sns.histplot(post['AMT_CREDIT'], bins=20, kde=True)
plt.title('Credit Amount Distribution')
plt.ylabel('Credit Amount')
plt.xlabel('Frequency')
plt.show()
```

```
TARGET         0    1
AMT_CREDIT
45000.0      5.0  NaN
47970.0      1.0  NaN
49752.0      1.0  NaN
50940.0      7.0  1.0
52128.0      3.0  NaN
...          ...  ...
```

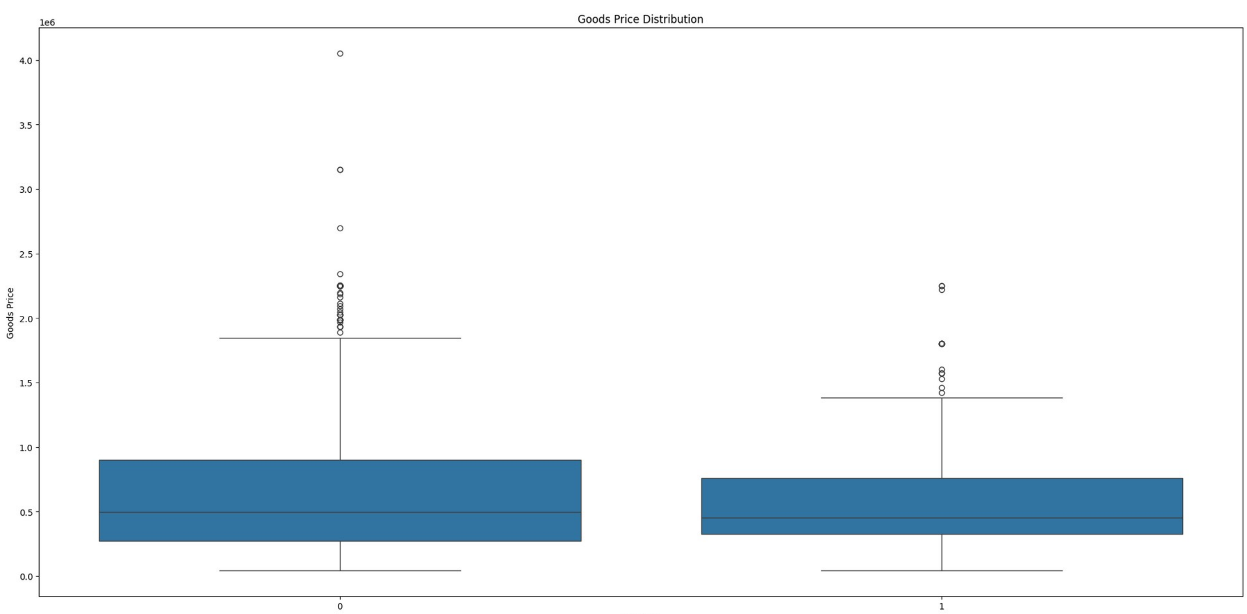
```
2695500.0   6.0  1.0
2931660.0   1.0  NaN
3150000.0   1.0  NaN
3299688.0   1.0  NaN
4050000.0   1.0  NaN

[1806 rows x 2 columns]
```



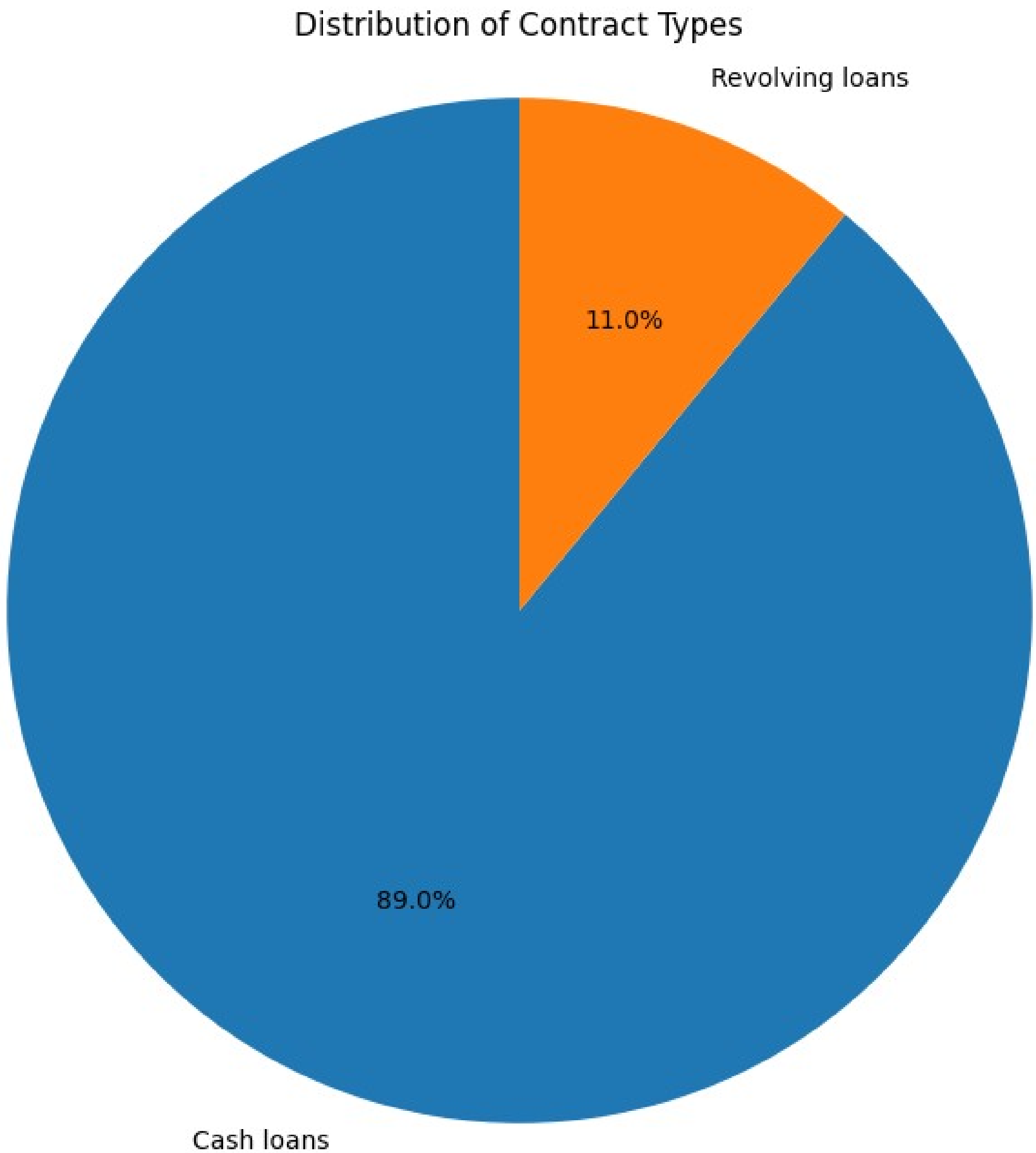

```
# annuity amount

plt.figure(figsize=[20,10])

sns.boxplot(x='TARGET', y='AMT_ANNUITY', data=post)
plt.title('Annuity Amount Distribution')
plt.xlabel('TARGET')
plt.ylabel('Annuity Amount')

plt.tight_layout()
plt.show()
```




```python
# goods price distribution
plt.figure(figsize=[20,10])

sns.boxplot(x='TARGET', y='AMT_GOODS_PRICE', data=post)
plt.title('Goods Price Distribution')
plt.xlabel('TARGET')
plt.ylabel('Goods Price')

plt.tight_layout()
plt.show()
```



```
contract_counts = post['NAME_CONTRACT_TYPE'].value_counts()

plt.figure(figsize=(8, 8))
plt.pie(contract_counts, labels=contract_counts.index, autopct='%1.1f%
%', startangle=90)
plt.title('Distribution of Contract Types')
plt.axis('equal')
plt.show()
```



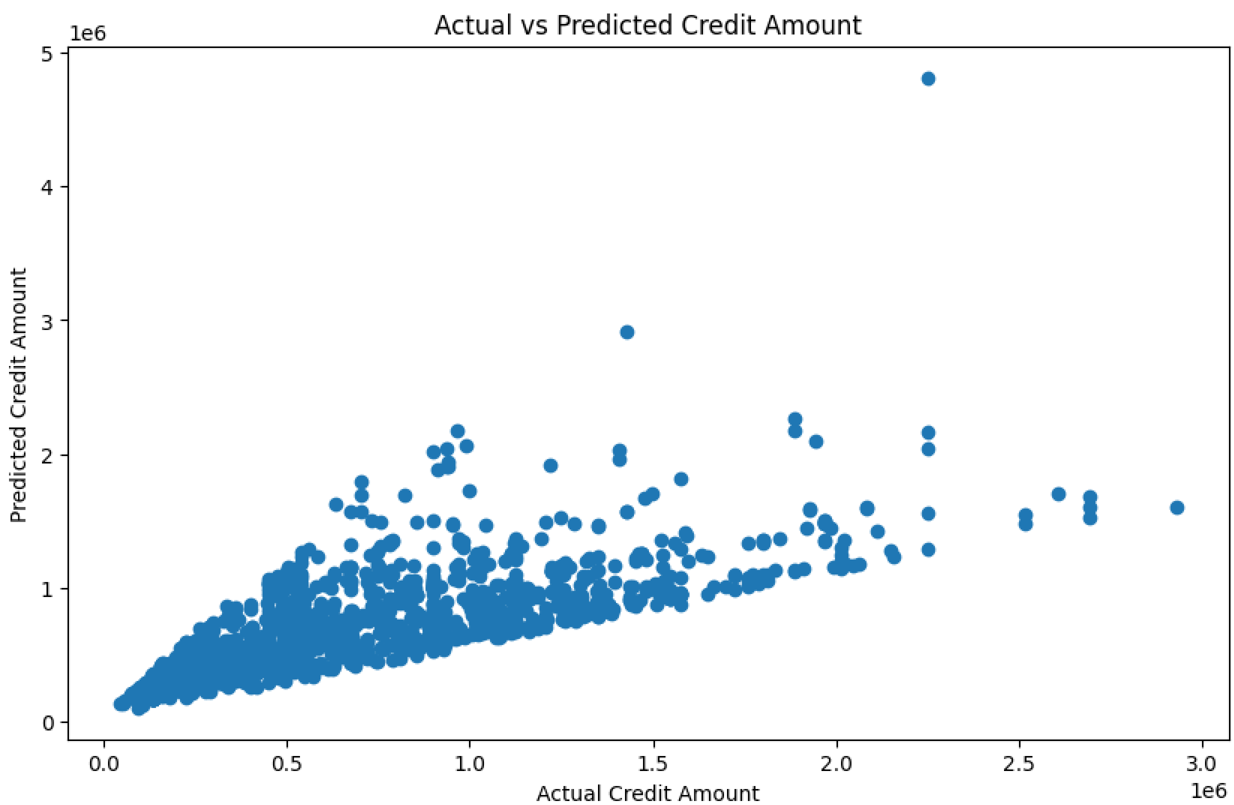
```
post.columns
if 'AMT_APPLICATTON' in post.columns:
    print("Column exists")
else:
  print("Does not exist")

Does not exist

annual_income = post.groupby("AMT_INCOME_TOTAL")
["TARGET"].value_counts().unstack()
print(annual_income)

TARGET                0   1
AMT_INCOME_TOTAL
33750.0             1.0 NaN
38250.0             1.0 NaN
40500.0             1.0 NaN
45000.0             3.0 NaN
49500.0             2.0 NaN
...                 ...  ..
1575000.0           2.0 NaN
1800000.0           1.0 NaN
2025000.0           2.0 NaN
2475000.0           1.0 NaN
4500000.0           1.0 NaN

[266 rows x 2 columns]

numeric_data = post.select_dtypes(include=np.number)
numeric_data.head()

{"type":"dataframe","variable_name":"numeric_data"}

post.isnull().sum()
post

{"type":"dataframe","variable_name":"post"}

post_dropped = post.dropna()
post_dropped

{"type":"dataframe","variable_name":"post_dropped"}
```

A simple linear regression model

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
import pandas as pd

features = ['AMT_INCOME_TOTAL', 'AMT_ANNUITY', 'CNT_CHILDREN']
```

```
x = post_dropped[features]
y = post_dropped["AMT_CREDIT"]

x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.2,
random_state = 42)
model = LinearRegression()
model.fit(x_train, y_train)
y_pred = model.predict(x_test)

mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print(f"Mean squared error: {mse}")
print(f"R-squared: {r2}")

plt.figure(figsize=(10,6))
plt.scatter(y_test, y_pred)
plt.xlabel("Actual Credit Amount")
plt.ylabel("Predicted Credit Amount")
plt.title("Actual vs Predicted Credit Amount")
plt.show()
```

```
Mean squared error: 94038873532.07123
R-squared: 0.5610474029402006
```



```
post.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_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', 'AGE', 'HOUSING_TYPE',
'HOUSING_TYPE#'],
      dtype='object', length=125)
```

STATISTICAL ANALYSIS

```
from scipy import stats

# two sample t- test, credit amounts of default and non_default
loaners
group1 = post[post['TARGET'] == 0]['AMT_CREDIT']
group2 = post[post['TARGET'] == 1]['AMT_CREDIT']

t_stat, p_value = stats.ttest_ind(group1, group2)

print(f"T.statistic: {t_stat}, P-value: {p_value}")
```

```
T.statistic: 1.3572008503729496, P-value: 0.17475301721429964
```

```
prev.columns
```

```
Index(['SK_ID_PREV', 'SK_ID_CURR', 'NAME_CONTRACT_TYPE',
'AMT_ANNUITY',
       'AMT_APPLICATION', 'AMT_CREDIT', 'AMT_DOWN_PAYMENT',
'AMT_GOODS_PRICE',
       'WEEKDAY_APPR_PROCESS_START', 'HOUR_APPR_PROCESS_START',
       'FLAG_LAST_APPL_PER_CONTRACT', 'NFLAG_LAST_APPL_IN_DAY',
       'RATE_DOWN_PAYMENT', 'RATE_INTEREST_PRIMARY',
       'RATE_INTEREST_PRIVILEGED', 'NAME_CASH_LOAN_PURPOSE',
       'NAME_CONTRACT_STATUS', 'DAYS_DECISION', 'NAME_PAYMENT_TYPE',
       'CODE_REJECT_REASON', 'NAME_TYPE_SUITE', 'NAME_CLIENT_TYPE',
       'NAME_GOODS_CATEGORY', 'NAME_PORTFOLIO', 'NAME_PRODUCT_TYPE',
       'CHANNEL_TYPE', 'SELLERPLACE_AREA', 'NAME_SELLER_INDUSTRY',
       'CNT_PAYMENT', 'NAME_YIELD_GROUP', 'PRODUCT_COMBINATION',
       'DAYS_FIRST_DRAWING', 'DAYS_FIRST_DUE',
'DAYS_LAST_DUE_1ST_VERSION',
       'DAYS_LAST_DUE', 'DAYS_TERMINATION',
'NFLAG_INSURED_ON_APPROVAL'],
      dtype='object')
```

```
prev.head()
```

{"type":"dataframe","variable_name":"prev"}

```
prev.duplicated().sum()
```

```
0
```

```
post.duplicated().sum()
```

```
0
```

```
if 'AMT_APPLICATION' in post.columns:
  print("exists")
else:
  print("does not exist")

post = pd.read_csv("application_data.csv")
print(post.head())
```

```
does not exist
   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]
```

```
# ANOVA

group1 = post[post['TARGET'] == 0]['AMT_CREDIT']
group2 = post[post['TARGET'] == 1]['AMT_CREDIT']

F_statistic, p_value = stats.f_oneway(group1, group2)

print(f"F-statistic: {F_statistic}")
print(f"P-value: {p_value}")
```

```
F-statistic: 283.8753868374507
P-value: 1.1474602724260586e-63
```

```
post['AMT_CREDIT'].corr(post['AMT_GOODS_PRICE'])
```

```
0.9869683054221501
```

```
post['AMT_CREDIT'].corr(post['AMT_ANNUITY'])
```

```
0.7701380033118824
```

```
post2 = post.select_dtypes(include=np.number).corr()
post2
```

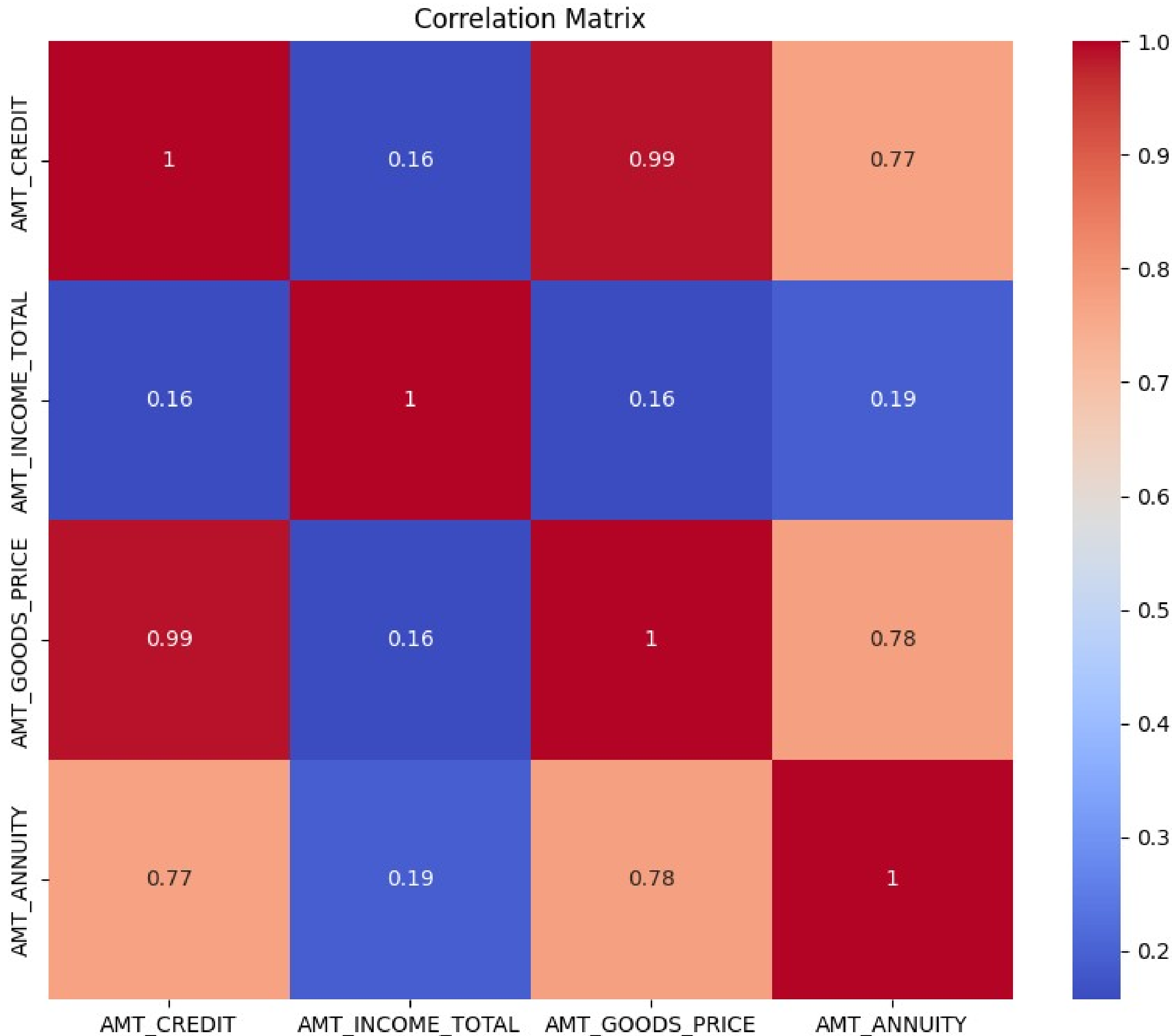
{"type":"dataframe","variable_name":"post2"}

```
variables = post[['AMT_CREDIT', 'AMT_INCOME_TOTAL', 'CNT_CHILDREN',]]
matrix = variables.corr()
print(matrix)
```

```
                  AMT_CREDIT  AMT_INCOME_TOTAL  CNT_CHILDREN
AMT_CREDIT          1.000000          0.156870      0.002145
AMT_INCOME_TOTAL    0.156870          1.000000      0.012882
CNT_CHILDREN        0.002145          0.012882      1.000000
```

```
variables2 = post[['AMT_CREDIT','AMT_INCOME_TOTAL', 'AMT_GOODS_PRICE',
'AMT_ANNUITY']]
matrix2 = variables2.corr()
print(matrix2)

plt.figure(figsize=(10,8))
sns.heatmap(matrix2, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
```

```
                  AMT_CREDIT  AMT_INCOME_TOTAL  AMT_GOODS_PRICE
AMT_ANNUITY
AMT_CREDIT          1.000000          0.156870         0.986968
0.770138
AMT_INCOME_TOTAL    0.156870          1.000000         0.159610
0.191657
AMT_GOODS_PRICE     0.986968          0.159610         1.000000
0.775109
AMT_ANNUITY         0.770138          0.191657         0.775109
1.000000
```




```
prev_numerical = prev.select_dtypes(include=np.number)
corr_matrix1 = prev_numerical.corr()
print(corr_matrix1)
```

```
                           SK_ID_PREV  SK_ID_CURR  AMT_ANNUITY  \
SK_ID_PREV                   1.000000   -0.002514    -0.062103
SK_ID_CURR                  -0.002514    1.000000     0.085439
AMT_ANNUITY                 -0.062103    0.085439     1.000000
AMT_APPLICATION             -0.065285    0.113620     0.986229
AMT_CREDIT                  -0.061630    0.086436     0.999939
AMT_DOWN_PAYMENT            -0.017276    0.175086     0.775581
AMT_GOODS_PRICE             -0.065285    0.113620     0.986229
HOUR_APPR_PROCESS_START      0.098916   -0.112946    -0.048648
NFLAG_LAST_APPL_IN_DAY            NaN         NaN          NaN
RATE_DOWN_PAYMENT           -0.022402    0.147016     0.072657
RATE_INTEREST_PRIMARY        0.096463    0.105772    -0.220981
```

```
RATE_INTEREST_PRIVILEGED     0.101816    0.154202    -0.146728
DAYS_DECISION               -0.057981    0.208837    -0.101702
SELLERPLACE_AREA            -0.183640   -0.052637     0.113311
CNT_PAYMENT                       NaN         NaN          NaN
DAYS_FIRST_DRAWING                NaN         NaN          NaN
DAYS_FIRST_DUE              -0.171873   -0.048295    -0.103207
DAYS_LAST_DUE_1ST_VERSION   -0.060763    0.206990    -0.100969
DAYS_LAST_DUE               -0.215247   -0.092463    -0.005796
DAYS_TERMINATION            -0.176348   -0.040080     0.051414
NFLAG_INSURED_ON_APPROVAL    0.115358   -0.047376    -0.060163

                           AMT_APPLICATION  AMT_CREDIT
AMT_DOWN_PAYMENT  \
SK_ID_PREV                       -0.065285   -0.061630          -
0.017276
SK_ID_CURR                        0.113620    0.086436
0.175086
AMT_ANNUITY                       0.986229    0.999939
0.775581
AMT_APPLICATION                   1.000000    0.986439
0.866017
AMT_CREDIT                        0.986439    1.000000
0.776133
AMT_DOWN_PAYMENT                  0.866017    0.776133
1.000000
AMT_GOODS_PRICE                   1.000000    0.986439
0.866017
HOUR_APPR_PROCESS_START          -0.037915   -0.047202
0.003292
NFLAG_LAST_APPL_IN_DAY                 NaN         NaN
NaN
RATE_DOWN_PAYMENT                 0.181470    0.074077
0.496795
RATE_INTEREST_PRIMARY            -0.199475   -0.221591          -
0.064177
RATE_INTEREST_PRIVILEGED         -0.105586   -0.140401
0.025169
DAYS_DECISION                    -0.065138   -0.099303
0.044893
SELLERPLACE_AREA                  0.084914    0.111770          -
0.001798
CNT_PAYMENT                            NaN         NaN
NaN
DAYS_FIRST_DRAWING                     NaN         NaN
NaN
DAYS_FIRST_DUE                   -0.100672   -0.104360          -
0.076833
DAYS_LAST_DUE_1ST_VERSION        -0.064507   -0.098585
0.044578
```

```
DAYS_LAST_DUE                        -0.010928   -0.005403          -
0.052079
DAYS_TERMINATION                      0.025092    0.052355          -
0.075942
NFLAG_INSURED_ON_APPROVAL            -0.091162   -0.058894          -
0.088352

                           AMT_GOODS_PRICE  HOUR_APPR_PROCESS_START  \
SK_ID_PREV                       -0.065285                 0.098916
SK_ID_CURR                        0.113620                -0.112946
AMT_ANNUITY                       0.986229                -0.048648
AMT_APPLICATION                   1.000000                -0.037915
AMT_CREDIT                        0.986439                -0.047202
AMT_DOWN_PAYMENT                  0.866017                 0.003292
AMT_GOODS_PRICE                   1.000000                -0.037915
HOUR_APPR_PROCESS_START          -0.037915                 1.000000
NFLAG_LAST_APPL_IN_DAY                 NaN                      NaN
RATE_DOWN_PAYMENT                 0.181470                 0.168831
RATE_INTEREST_PRIMARY            -0.199475                -0.011590
RATE_INTEREST_PRIVILEGED         -0.105586                 0.304157
DAYS_DECISION                    -0.065138                 0.013474
SELLERPLACE_AREA                  0.084914                 0.117393
CNT_PAYMENT                            NaN                      NaN
DAYS_FIRST_DRAWING                     NaN                      NaN
DAYS_FIRST_DUE                   -0.100672                -0.105091
DAYS_LAST_DUE_1ST_VERSION        -0.064507                 0.014113
DAYS_LAST_DUE                    -0.010928                 0.114230
DAYS_TERMINATION                  0.025092                 0.153627
NFLAG_INSURED_ON_APPROVAL        -0.091162                -0.146447

                           NFLAG_LAST_APPL_IN_DAY
RATE_DOWN_PAYMENT  ...  \
SK_ID_PREV                                    NaN           -
0.022402  ...
SK_ID_CURR                                    NaN
0.147016  ...
AMT_ANNUITY                                   NaN
0.072657  ...
AMT_APPLICATION                               NaN
0.181470  ...
AMT_CREDIT                                    NaN
0.074077  ...
AMT_DOWN_PAYMENT                              NaN
0.496795  ...
AMT_GOODS_PRICE                               NaN
0.181470  ...
HOUR_APPR_PROCESS_START                       NaN
0.168831  ...
NFLAG_LAST_APPL_IN_DAY                        NaN
```

```
NaN  ...
RATE_DOWN_PAYMENT                                  NaN
1.000000  ...
RATE_INTEREST_PRIMARY                              NaN
0.058842  ...
RATE_INTEREST_PRIVILEGED                           NaN
0.204840  ...
DAYS_DECISION                                      NaN
0.162642  ...
SELLERPLACE_AREA                                   NaN          -
0.099124  ...
CNT_PAYMENT                                        NaN
NaN  ...
DAYS_FIRST_DRAWING                                 NaN
NaN  ...
DAYS_FIRST_DUE                                     NaN          -
0.247103  ...
DAYS_LAST_DUE_1ST_VERSION                          NaN
0.163324  ...
DAYS_LAST_DUE                                      NaN          -
0.070213  ...
DAYS_TERMINATION                                   NaN          -
0.151233  ...
NFLAG_INSURED_ON_APPROVAL                          NaN          -
0.039446  ...

                           RATE_INTEREST_PRIVILEGED  DAYS_DECISION  \
SK_ID_PREV                                 0.101816      -0.057981
SK_ID_CURR                                 0.154202       0.208837
AMT_ANNUITY                               -0.146728      -0.101702
AMT_APPLICATION                           -0.105586      -0.065138
AMT_CREDIT                                -0.140401      -0.099303
AMT_DOWN_PAYMENT                           0.025169       0.044893
AMT_GOODS_PRICE                           -0.105586      -0.065138
HOUR_APPR_PROCESS_START                    0.304157       0.013474
NFLAG_LAST_APPL_IN_DAY                          NaN            NaN
RATE_DOWN_PAYMENT                          0.204840       0.162642
RATE_INTEREST_PRIMARY                     -0.035909       0.086303
RATE_INTEREST_PRIVILEGED                   1.000000       0.608370
DAYS_DECISION                              0.608370       1.000000
SELLERPLACE_AREA                          -0.159818      -0.000628
CNT_PAYMENT                                     NaN            NaN
DAYS_FIRST_DRAWING                              NaN            NaN
DAYS_FIRST_DUE                             0.177689       0.192757
DAYS_LAST_DUE_1ST_VERSION                  0.607948       0.999875
DAYS_LAST_DUE                              0.215970       0.374097
DAYS_TERMINATION                           0.265466       0.404214
NFLAG_INSURED_ON_APPROVAL                  0.036203       0.250287
```

```
                           SELLERPLACE_AREA  CNT_PAYMENT
DAYS_FIRST_DRAWING  \
SK_ID_PREV                        -0.183640          NaN
NaN
SK_ID_CURR                        -0.052637          NaN
NaN
AMT_ANNUITY                        0.113311          NaN
NaN
AMT_APPLICATION                    0.084914          NaN
NaN
AMT_CREDIT                         0.111770          NaN
NaN
AMT_DOWN_PAYMENT                  -0.001798          NaN
NaN
AMT_GOODS_PRICE                    0.084914          NaN
NaN
HOUR_APPR_PROCESS_START            0.117393          NaN
NaN
NFLAG_LAST_APPL_IN_DAY                  NaN          NaN
NaN
RATE_DOWN_PAYMENT                 -0.099124          NaN
NaN
RATE_INTEREST_PRIMARY              0.181596          NaN
NaN
RATE_INTEREST_PRIVILEGED          -0.159818          NaN
NaN
DAYS_DECISION                     -0.000628          NaN
NaN
SELLERPLACE_AREA                   1.000000          NaN
NaN
CNT_PAYMENT                             NaN          NaN
NaN
DAYS_FIRST_DRAWING                      NaN          NaN
NaN
DAYS_FIRST_DUE                    -0.031593          NaN
NaN
DAYS_LAST_DUE_1ST_VERSION         -0.002432          NaN
NaN
DAYS_LAST_DUE                     -0.011935          NaN
NaN
DAYS_TERMINATION                  -0.022329          NaN
NaN
NFLAG_INSURED_ON_APPROVAL          0.241941          NaN
NaN

                           DAYS_FIRST_DUE
DAYS_LAST_DUE_1ST_VERSION  \
SK_ID_PREV                      -0.171873                  -0.060763
```

SK_ID_CURR	-0.048295	0.206990
AMT_ANNUITY	-0.103207	-0.100969
AMT_APPLICATION	-0.100672	-0.064507
AMT_CREDIT	-0.104360	-0.098585
AMT_DOWN_PAYMENT	-0.076833	0.044578
AMT_GOODS_PRICE	-0.100672	-0.064507
HOUR_APPR_PROCESS_START	-0.105091	0.014113
NFLAG_LAST_APPL_IN_DAY	NaN	NaN
RATE_DOWN_PAYMENT	-0.247103	0.163324
RATE_INTEREST_PRIMARY	0.006460	0.075696
RATE_INTEREST_PRIVILEGED	0.177689	0.607948
DAYS_DECISION	0.192757	0.999875
SELLERPLACE_AREA	-0.031593	-0.002432
CNT_PAYMENT	NaN	NaN
DAYS_FIRST_DRAWING	NaN	NaN
DAYS_FIRST_DUE	1.000000	0.192068
DAYS_LAST_DUE_1ST_VERSION	0.192068	1.000000
DAYS_LAST_DUE	0.315712	0.375886
DAYS_TERMINATION	0.297415	0.405664
NFLAG_INSURED_ON_APPROVAL	-0.052273	0.246943

```
                           DAYS_LAST_DUE  DAYS_TERMINATION  \
SK_ID_PREV                     -0.215247         -0.176348
SK_ID_CURR                     -0.092463         -0.040080
AMT_ANNUITY                    -0.005796          0.051414
AMT_APPLICATION                -0.010928          0.025092
AMT_CREDIT                     -0.005403          0.052355
AMT_DOWN_PAYMENT               -0.052079         -0.075942
AMT_GOODS_PRICE                -0.010928          0.025092
HOUR_APPR_PROCESS_START         0.114230          0.153627
NFLAG_LAST_APPL_IN_DAY               NaN               NaN
```

```
RATE_DOWN_PAYMENT                  -0.070213        -0.151233
RATE_INTEREST_PRIMARY              -0.063183        -0.039266
RATE_INTEREST_PRIVILEGED            0.215970         0.265466
DAYS_DECISION                       0.374097         0.404214
SELLERPLACE_AREA                   -0.011935        -0.022329
CNT_PAYMENT                              NaN              NaN
DAYS_FIRST_DRAWING                       NaN              NaN
DAYS_FIRST_DUE                      0.315712         0.297415
DAYS_LAST_DUE_1ST_VERSION           0.375886         0.405664
DAYS_LAST_DUE                       1.000000         0.941107
DAYS_TERMINATION                    0.941107         1.000000
NFLAG_INSURED_ON_APPROVAL          -0.058250         0.033985

                           NFLAG_INSURED_ON_APPROVAL
SK_ID_PREV                                  0.115358
SK_ID_CURR                                 -0.047376
AMT_ANNUITY                                -0.060163
AMT_APPLICATION                            -0.091162
AMT_CREDIT                                 -0.058894
AMT_DOWN_PAYMENT                           -0.088352
AMT_GOODS_PRICE                            -0.091162
HOUR_APPR_PROCESS_START                    -0.146447
NFLAG_LAST_APPL_IN_DAY                           NaN
RATE_DOWN_PAYMENT                          -0.039446
RATE_INTEREST_PRIMARY                       0.307477
RATE_INTEREST_PRIVILEGED                    0.036203
DAYS_DECISION                               0.250287
SELLERPLACE_AREA                            0.241941
CNT_PAYMENT                                      NaN
DAYS_FIRST_DRAWING                               NaN
DAYS_FIRST_DUE                             -0.052273
DAYS_LAST_DUE_1ST_VERSION                   0.246943
DAYS_LAST_DUE                              -0.058250
DAYS_TERMINATION                            0.033985
NFLAG_INSURED_ON_APPROVAL                   1.000000

[21 rows x 21 columns]
```

```
# t- test
group1 = post[post['TARGET'] == 0]['AMT_CREDIT']
group2 = post[post['TARGET'] == 1]['AMT_CREDIT']

t_stat, p_value = stats.ttest_ind(group1, group2)

print(f"T.statistic: {t_stat}, P-value: {p_value}")
```

```
T.statistic: 16.848601925306788, P-value: 1.1474602724788813e-63
```

```
prev_numerical = prev.select_dtypes(include=np.number)
corr_matrix1 = prev_numerical.corr()
```

```
print(corr_matrix1)

plt.figure(figsize=(10,8))
sns.heatmap(corr_matrix1, annot=True, cmap='coolwarm')
plt.title('Correlation Matrix')
plt.show()
```

```
                            SK_ID_PREV  SK_ID_CURR  AMT_ANNUITY  \
SK_ID_PREV                    1.000000   -0.002514    -0.062103
SK_ID_CURR                   -0.002514    1.000000     0.085439
AMT_ANNUITY                  -0.062103    0.085439     1.000000
AMT_APPLICATION              -0.065285    0.113620     0.986229
AMT_CREDIT                   -0.061630    0.086436     0.999939
AMT_DOWN_PAYMENT             -0.017276    0.175086     0.775581
AMT_GOODS_PRICE              -0.065285    0.113620     0.986229
HOUR_APPR_PROCESS_START       0.098916   -0.112946    -0.048648
NFLAG_LAST_APPL_IN_DAY             NaN         NaN          NaN
RATE_DOWN_PAYMENT            -0.022402    0.147016     0.072657
RATE_INTEREST_PRIMARY         0.096463    0.105772    -0.220981
RATE_INTEREST_PRIVILEGED      0.101816    0.154202    -0.146728
DAYS_DECISION                -0.057981    0.208837    -0.101702
SELLERPLACE_AREA             -0.183640   -0.052637     0.113311
CNT_PAYMENT                        NaN         NaN          NaN
DAYS_FIRST_DRAWING                 NaN         NaN          NaN
DAYS_FIRST_DUE               -0.171873   -0.048295    -0.103207
DAYS_LAST_DUE_1ST_VERSION    -0.060763    0.206990    -0.100969
DAYS_LAST_DUE                -0.215247   -0.092463    -0.005796
DAYS_TERMINATION             -0.176348   -0.040080     0.051414
NFLAG_INSURED_ON_APPROVAL     0.115358   -0.047376    -0.060163

                            AMT_APPLICATION  AMT_CREDIT
AMT_DOWN_PAYMENT  \
SK_ID_PREV                        -0.065285   -0.061630         -
0.017276
SK_ID_CURR                         0.113620    0.086436
0.175086
AMT_ANNUITY                        0.986229    0.999939
0.775581
AMT_APPLICATION                    1.000000    0.986439
0.866017
AMT_CREDIT                         0.986439    1.000000
0.776133
AMT_DOWN_PAYMENT                   0.866017    0.776133
1.000000
AMT_GOODS_PRICE                    1.000000    0.986439
0.866017
HOUR_APPR_PROCESS_START           -0.037915   -0.047202
0.003292
NFLAG_LAST_APPL_IN_DAY                  NaN         NaN
NaN
```

```
RATE_DOWN_PAYMENT                    0.181470    0.074077
0.496795
RATE_INTEREST_PRIMARY               -0.199475   -0.221591          -
0.064177
RATE_INTEREST_PRIVILEGED            -0.105586   -0.140401
0.025169
DAYS_DECISION                       -0.065138   -0.099303
0.044893
SELLERPLACE_AREA                     0.084914    0.111770          -
0.001798
CNT_PAYMENT                               NaN         NaN
NaN
DAYS_FIRST_DRAWING                        NaN         NaN
NaN
DAYS_FIRST_DUE                      -0.100672   -0.104360          -
0.076833
DAYS_LAST_DUE_1ST_VERSION           -0.064507   -0.098585
0.044578
DAYS_LAST_DUE                       -0.010928   -0.005403          -
0.052079
DAYS_TERMINATION                     0.025092    0.052355          -
0.075942
NFLAG_INSURED_ON_APPROVAL           -0.091162   -0.058894          -
0.088352

                           AMT_GOODS_PRICE  HOUR_APPR_PROCESS_START  \
SK_ID_PREV                       -0.065285                 0.098916
SK_ID_CURR                        0.113620                -0.112946
AMT_ANNUITY                       0.986229                -0.048648
AMT_APPLICATION                   1.000000                -0.037915
AMT_CREDIT                        0.986439                -0.047202
AMT_DOWN_PAYMENT                  0.866017                 0.003292
AMT_GOODS_PRICE                   1.000000                -0.037915
HOUR_APPR_PROCESS_START          -0.037915                 1.000000
NFLAG_LAST_APPL_IN_DAY                 NaN                      NaN
RATE_DOWN_PAYMENT                 0.181470                 0.168831
RATE_INTEREST_PRIMARY            -0.199475                -0.011590
RATE_INTEREST_PRIVILEGED         -0.105586                 0.304157
DAYS_DECISION                    -0.065138                 0.013474
SELLERPLACE_AREA                  0.084914                 0.117393
CNT_PAYMENT                            NaN                      NaN
DAYS_FIRST_DRAWING                     NaN                      NaN
DAYS_FIRST_DUE                   -0.100672                -0.105091
DAYS_LAST_DUE_1ST_VERSION        -0.064507                 0.014113
DAYS_LAST_DUE                    -0.010928                 0.114230
DAYS_TERMINATION                  0.025092                 0.153627
NFLAG_INSURED_ON_APPROVAL        -0.091162                -0.146447

                           NFLAG_LAST_APPL_IN_DAY
```

```
RATE_DOWN_PAYMENT  ...  \
SK_ID_PREV                                          NaN           -
0.022402  ...
SK_ID_CURR                                          NaN
0.147016  ...
AMT_ANNUITY                                         NaN
0.072657  ...
AMT_APPLICATION                                     NaN
0.181470  ...
AMT_CREDIT                                          NaN
0.074077  ...
AMT_DOWN_PAYMENT                                    NaN
0.496795  ...
AMT_GOODS_PRICE                                     NaN
0.181470  ...
HOUR_APPR_PROCESS_START                             NaN
0.168831  ...
NFLAG_LAST_APPL_IN_DAY                              NaN
NaN  ...
RATE_DOWN_PAYMENT                                   NaN
1.000000  ...
RATE_INTEREST_PRIMARY                               NaN
0.058842  ...
RATE_INTEREST_PRIVILEGED                            NaN
0.204840  ...
DAYS_DECISION                                       NaN
0.162642  ...
SELLERPLACE_AREA                                    NaN           -
0.099124  ...
CNT_PAYMENT                                         NaN
NaN  ...
DAYS_FIRST_DRAWING                                  NaN
NaN  ...
DAYS_FIRST_DUE                                      NaN           -
0.247103  ...
DAYS_LAST_DUE_1ST_VERSION                           NaN
0.163324  ...
DAYS_LAST_DUE                                       NaN           -
0.070213  ...
DAYS_TERMINATION                                    NaN           -
0.151233  ...
NFLAG_INSURED_ON_APPROVAL                           NaN           -
0.039446  ...

                           RATE_INTEREST_PRIVILEGED  DAYS_DECISION  \
SK_ID_PREV                                 0.101816      -0.057981
SK_ID_CURR                                 0.154202       0.208837
AMT_ANNUITY                               -0.146728      -0.101702
AMT_APPLICATION                           -0.105586      -0.065138
```

```
AMT_CREDIT                                    -0.140401      -0.099303
AMT_DOWN_PAYMENT                               0.025169       0.044893
AMT_GOODS_PRICE                               -0.105586      -0.065138
HOUR_APPR_PROCESS_START                        0.304157       0.013474
NFLAG_LAST_APPL_IN_DAY                              NaN            NaN
RATE_DOWN_PAYMENT                              0.204840       0.162642
RATE_INTEREST_PRIMARY                         -0.035909       0.086303
RATE_INTEREST_PRIVILEGED                       1.000000       0.608370
DAYS_DECISION                                  0.608370       1.000000
SELLERPLACE_AREA                              -0.159818      -0.000628
CNT_PAYMENT                                         NaN            NaN
DAYS_FIRST_DRAWING                                  NaN            NaN
DAYS_FIRST_DUE                                 0.177689       0.192757
DAYS_LAST_DUE_1ST_VERSION                      0.607948       0.999875
DAYS_LAST_DUE                                  0.215970       0.374097
DAYS_TERMINATION                               0.265466       0.404214
NFLAG_INSURED_ON_APPROVAL                      0.036203       0.250287

                           SELLERPLACE_AREA  CNT_PAYMENT
DAYS_FIRST_DRAWING  \
SK_ID_PREV                        -0.183640          NaN
NaN
SK_ID_CURR                        -0.052637          NaN
NaN
AMT_ANNUITY                        0.113311          NaN
NaN
AMT_APPLICATION                    0.084914          NaN
NaN
AMT_CREDIT                         0.111770          NaN
NaN
AMT_DOWN_PAYMENT                  -0.001798          NaN
NaN
AMT_GOODS_PRICE                    0.084914          NaN
NaN
HOUR_APPR_PROCESS_START            0.117393          NaN
NaN
NFLAG_LAST_APPL_IN_DAY                  NaN          NaN
NaN
RATE_DOWN_PAYMENT                 -0.099124          NaN
NaN
RATE_INTEREST_PRIMARY              0.181596          NaN
NaN
RATE_INTEREST_PRIVILEGED          -0.159818          NaN
NaN
DAYS_DECISION                     -0.000628          NaN
NaN
SELLERPLACE_AREA                   1.000000          NaN
NaN
CNT_PAYMENT                             NaN          NaN
```

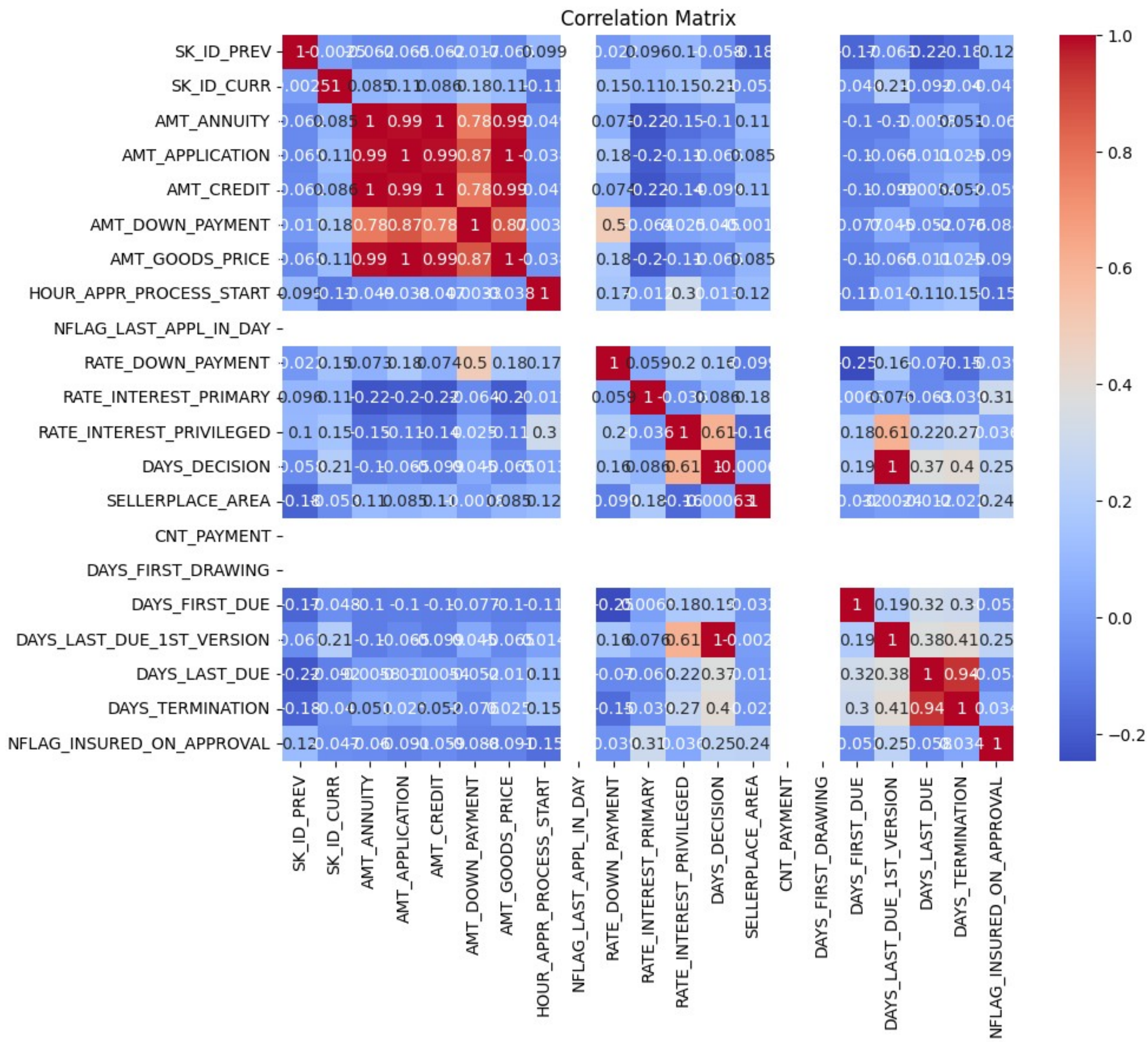
```
NaN
DAYS_FIRST_DRAWING                      NaN          NaN
NaN
DAYS_FIRST_DUE                    -0.031593          NaN
NaN
DAYS_LAST_DUE_1ST_VERSION         -0.002432          NaN
NaN
DAYS_LAST_DUE                     -0.011935          NaN
NaN
DAYS_TERMINATION                  -0.022329          NaN
NaN
NFLAG_INSURED_ON_APPROVAL          0.241941          NaN
NaN

                           DAYS_FIRST_DUE
DAYS_LAST_DUE_1ST_VERSION  \
SK_ID_PREV                      -0.171873                  -0.060763

SK_ID_CURR                      -0.048295                   0.206990

AMT_ANNUITY                     -0.103207                  -0.100969

AMT_APPLICATION                 -0.100672                  -0.064507

AMT_CREDIT                      -0.104360                  -0.098585

AMT_DOWN_PAYMENT                -0.076833                   0.044578

AMT_GOODS_PRICE                 -0.100672                  -0.064507

HOUR_APPR_PROCESS_START         -0.105091                   0.014113

NFLAG_LAST_APPL_IN_DAY                NaN                        NaN

RATE_DOWN_PAYMENT               -0.247103                   0.163324

RATE_INTEREST_PRIMARY            0.006460                   0.075696

RATE_INTEREST_PRIVILEGED         0.177689                   0.607948

DAYS_DECISION                    0.192757                   0.999875

SELLERPLACE_AREA                -0.031593                  -0.002432

CNT_PAYMENT                           NaN                        NaN

DAYS_FIRST_DRAWING                    NaN                        NaN

DAYS_FIRST_DUE                   1.000000                   0.192068

DAYS_LAST_DUE_1ST_VERSION        0.192068                   1.000000
```

DAYS_LAST_DUE	0.315712	0.375886
DAYS_TERMINATION	0.297415	0.405664
NFLAG_INSURED_ON_APPROVAL	-0.052273	0.246943

	DAYS_LAST_DUE	DAYS_TERMINATION \
SK_ID_PREV	-0.215247	-0.176348
SK_ID_CURR	-0.092463	-0.040080
AMT_ANNUITY	-0.005796	0.051414
AMT_APPLICATION	-0.010928	0.025092
AMT_CREDIT	-0.005403	0.052355
AMT_DOWN_PAYMENT	-0.052079	-0.075942
AMT_GOODS_PRICE	-0.010928	0.025092
HOUR_APPR_PROCESS_START	0.114230	0.153627
NFLAG_LAST_APPL_IN_DAY	NaN	NaN
RATE_DOWN_PAYMENT	-0.070213	-0.151233
RATE_INTEREST_PRIMARY	-0.063183	-0.039266
RATE_INTEREST_PRIVILEGED	0.215970	0.265466
DAYS_DECISION	0.374097	0.404214
SELLERPLACE_AREA	-0.011935	-0.022329
CNT_PAYMENT	NaN	NaN
DAYS_FIRST_DRAWING	NaN	NaN
DAYS_FIRST_DUE	0.315712	0.297415
DAYS_LAST_DUE_1ST_VERSION	0.375886	0.405664
DAYS_LAST_DUE	1.000000	0.941107
DAYS_TERMINATION	0.941107	1.000000
NFLAG_INSURED_ON_APPROVAL	-0.058250	0.033985

	NFLAG_INSURED_ON_APPROVAL
SK_ID_PREV	0.115358
SK_ID_CURR	-0.047376
AMT_ANNUITY	-0.060163
AMT_APPLICATION	-0.091162
AMT_CREDIT	-0.058894
AMT_DOWN_PAYMENT	-0.088352
AMT_GOODS_PRICE	-0.091162
HOUR_APPR_PROCESS_START	-0.146447
NFLAG_LAST_APPL_IN_DAY	NaN
RATE_DOWN_PAYMENT	-0.039446
RATE_INTEREST_PRIMARY	0.307477
RATE_INTEREST_PRIVILEGED	0.036203
DAYS_DECISION	0.250287
SELLERPLACE_AREA	0.241941
CNT_PAYMENT	NaN
DAYS_FIRST_DRAWING	NaN
DAYS_FIRST_DUE	-0.052273
DAYS_LAST_DUE_1ST_VERSION	0.246943

```
DAYS_LAST_DUE                                     -0.058250
DAYS_TERMINATION                                   0.033985
NFLAG_INSURED_ON_APPROVAL                          1.000000

[21 rows x 21 columns]
```

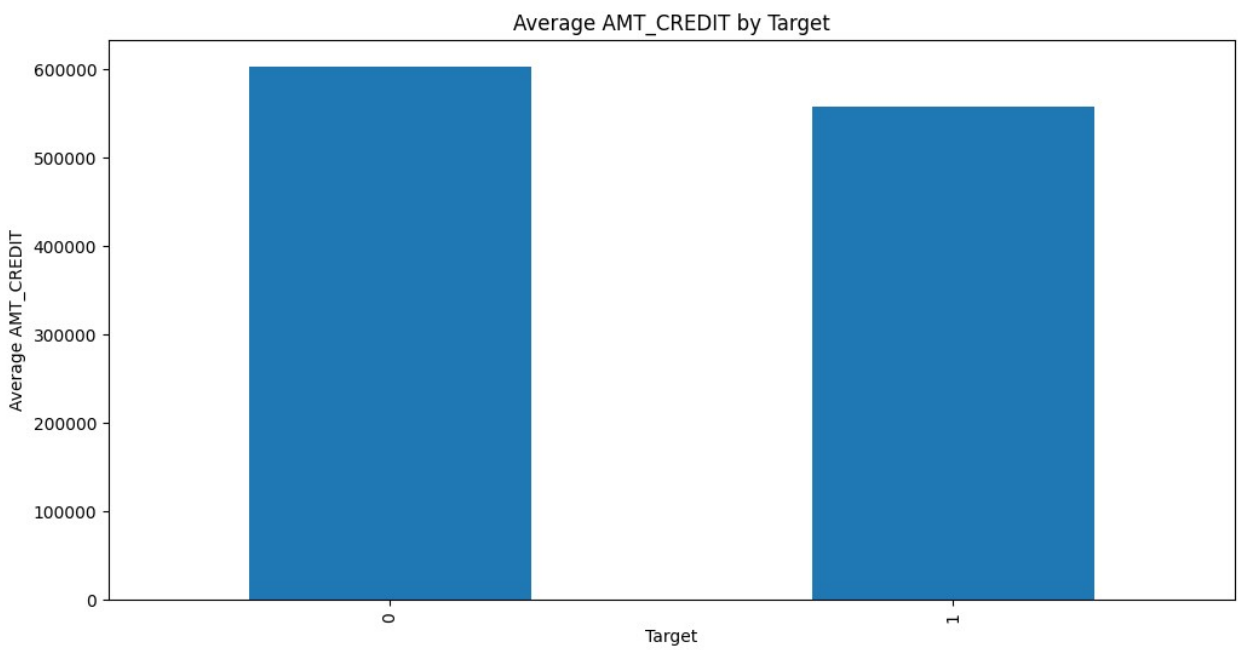


PLAYGROUND

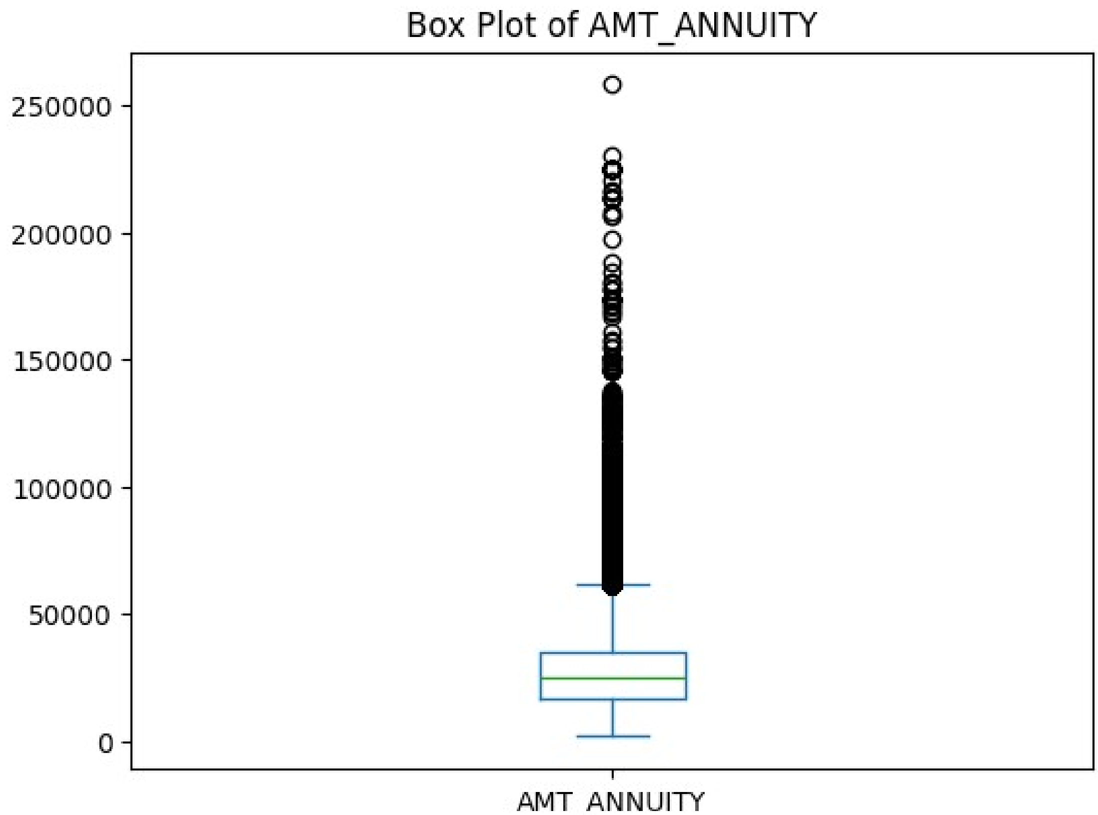
```
amt_income = post.groupby('AMT_INCOME_TOTAL')
['TARGET'].value_counts().unstack()
print(amt_income)

TARGET                   0    1
AMT_INCOME_TOTAL
25650.0                1.0  1.0
26100.0                3.0  NaN
```

```
26460.0          1.0  NaN
26550.0          2.0  NaN
27000.0         63.0  3.0
...              ...  ...
6750000.0        1.0  NaN
9000000.0        1.0  NaN
13500000.0       1.0  NaN
18000090.0       1.0  NaN
117000000.0      NaN  1.0

[2548 rows x 2 columns]
```

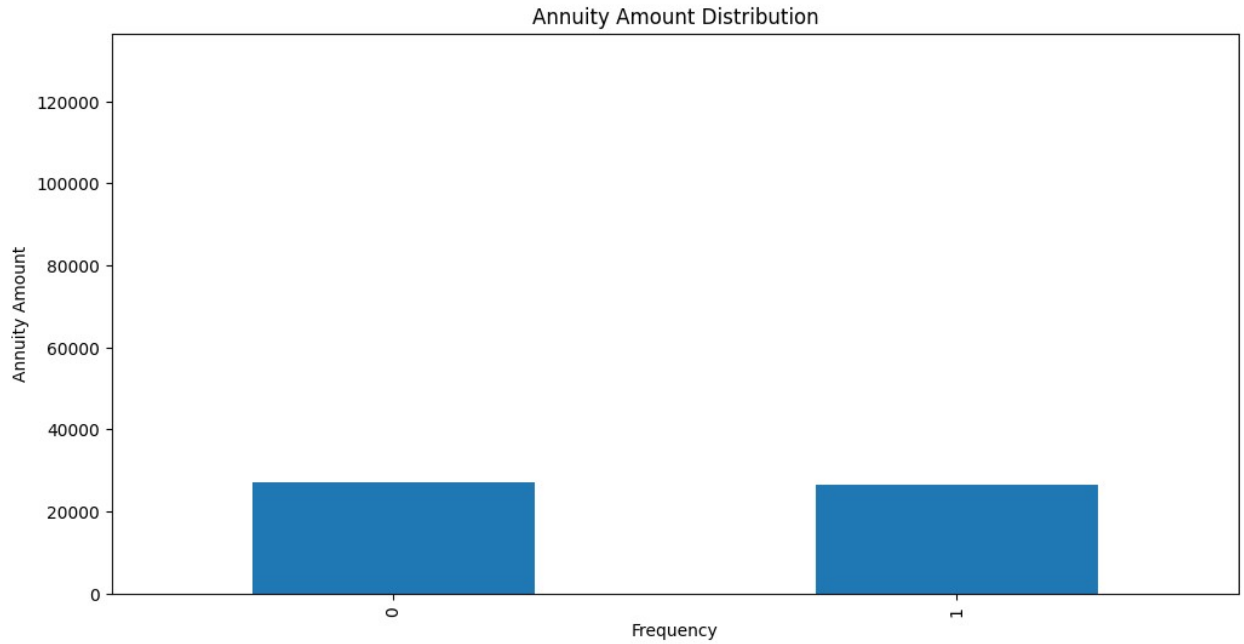
```
plt.figure(figsize=(12, 6))
post.groupby('TARGET')['AMT_CREDIT'].mean().plot(kind='bar')
plt.title('Average AMT_CREDIT by Target')
plt.xlabel('Target')
plt.ylabel('Average AMT_CREDIT')
plt.show()
```



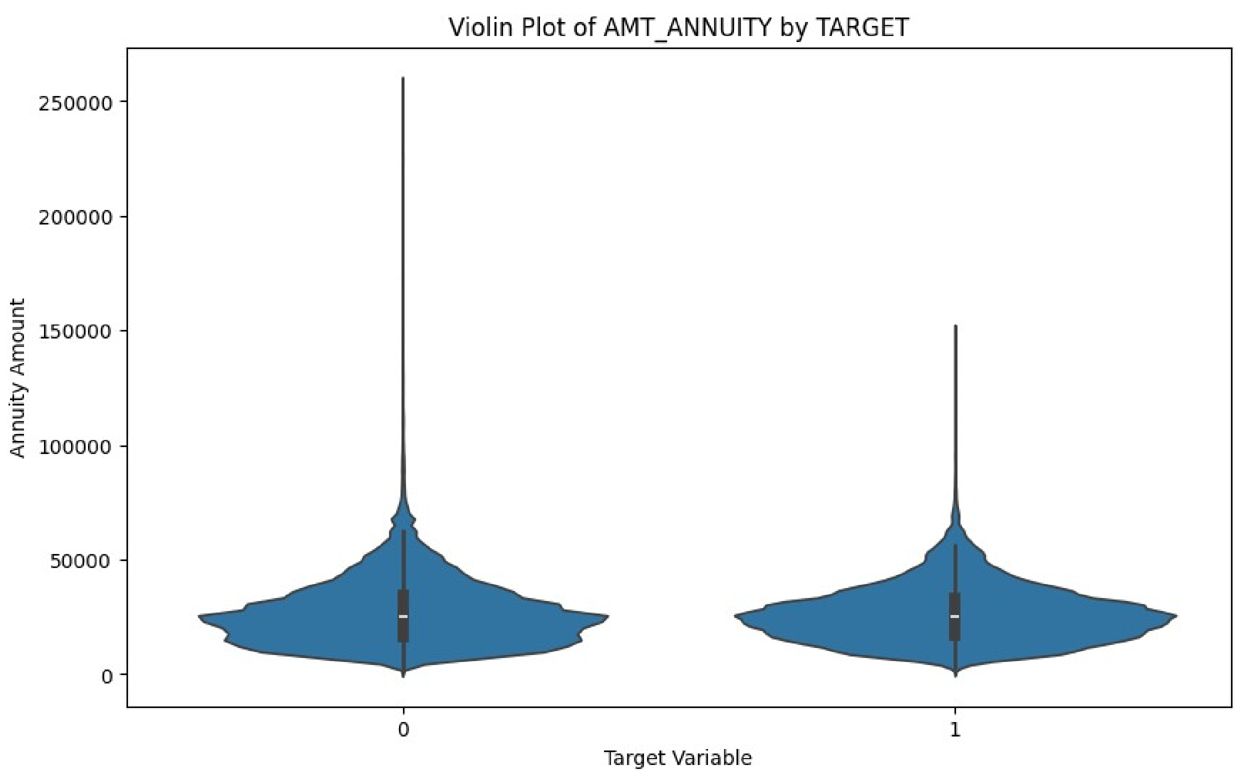

```
post['AMT_ANNUITY'].plot(kind='box')
plt.title('Box Plot of AMT_ANNUITY')
plt.show()
```




```
plt.figure(figsize=(12, 6))
post.groupby('TARGET')['AMT_ANNUITY'].mean().plot(kind='bar')

sns.histplot(post['AMT_ANNUITY'], bins=20, kde=True)
plt.title('Annuity Amount Distribution')
plt.ylabel('Annuity Amount')
plt.xlabel('Frequency')
plt.show()
```




```
plt.figure(figsize=(10, 6))
sns.violinplot(x='TARGET', y='AMT_ANNUITY', data=post)
plt.title('Violin Plot of AMT_ANNUITY by TARGET')
plt.xlabel('Target Variable')
plt.ylabel('Annuity Amount')
plt.show()
```



```python
# t-tests for AMT_CREDIT against other relevant variables
variables_for_ttest = ['AMT_INCOME_TOTAL', 'AMT_ANNUITY',
'AMT_GOODS_PRICE', 'CNT_CHILDREN', 'DAYS_EMPLOYED']

for var in variables_for_ttest:
    group1 = post[post['TARGET'] == 0][var]
    group2 = post[post['TARGET'] == 1][var]
    t_stat, p_value = stats.ttest_ind(group1, group2)
    print(f"T-test for AMT_CREDIT vs {var}:")
    print(f"  T-statistic: {t_stat}")
    print(f"  P-value: {p_value}")
    print("-" * 20)
```

```
T-test for AMT_CREDIT vs AMT_INCOME_TOTAL:
  T-statistic: 2.2081011084695983
  P-value: 0.027237960879677118
--------------------
T-test for AMT_CREDIT vs AMT_ANNUITY:
  T-statistic: nan
  P-value: nan
--------------------
T-test for AMT_CREDIT vs AMT_GOODS_PRICE:
  T-statistic: nan
  P-value: nan
--------------------
T-test for AMT_CREDIT vs CNT_CHILDREN:
  T-statistic: -10.64189504948037
  P-value: 1.9224915502038074e-26
--------------------
T-test for AMT_CREDIT vs DAYS_EMPLOYED:
  T-statistic: 24.94136608089604
  P-value: 3.6311730828848897e-137
--------------------
```

```python
# correlation coefficients between 'AMT_CREDIT' and other numerical
variables
numerical_cols = post.select_dtypes(include=np.number).columns
for col in numerical_cols:
    if col != 'AMT_CREDIT' and col != 'TARGET':  # Exclude
'AMT_CREDIT' and 'TARGET' itself
        correlation = post['AMT_CREDIT'].corr(post[col])
        print(f"Correlation between AMT_CREDIT and {col}:
{correlation}")
```

```
Correlation between AMT_CREDIT and SK_ID_CURR: -0.00034316120624785715
Correlation between AMT_CREDIT and CNT_CHILDREN: 0.0021454434616763416
Correlation between AMT_CREDIT and AMT_INCOME_TOTAL:
0.15687027185193095
Correlation between AMT_CREDIT and AMT_ANNUITY: 0.7701380033118824
```

```
Correlation between AMT_CREDIT and AMT_GOODS_PRICE: 0.9869683054221501
Correlation between AMT_CREDIT and REGION_POPULATION_RELATIVE:
0.09973787576226029
Correlation between AMT_CREDIT and DAYS_BIRTH: -0.055435947014526185
Correlation between AMT_CREDIT and DAYS_EMPLOYED: -0.06683834093092188
Correlation between AMT_CREDIT and DAYS_REGISTRATION:
0.009621325617372911
Correlation between AMT_CREDIT and DAYS_ID_PUBLISH: -
0.006574774427157878
Correlation between AMT_CREDIT and OWN_CAR_AGE: -0.09419071925702023
Correlation between AMT_CREDIT and FLAG_MOBIL: 0.001435572992417004
Correlation between AMT_CREDIT and FLAG_EMP_PHONE: 0.0655185936551337
Correlation between AMT_CREDIT and FLAG_WORK_PHONE: -
0.02108459932914736
Correlation between AMT_CREDIT and FLAG_CONT_MOBILE:
0.02365284569336862
Correlation between AMT_CREDIT and FLAG_PHONE: 0.026212789103157257
Correlation between AMT_CREDIT and FLAG_EMAIL: 0.016631912352195515
Correlation between AMT_CREDIT and CNT_FAM_MEMBERS:
0.06315981300414085
Correlation between AMT_CREDIT and REGION_RATING_CLIENT: -
0.10177638603517988
Correlation between AMT_CREDIT and REGION_RATING_CLIENT_W_CITY: -
0.11091525790094263
Correlation between AMT_CREDIT and HOUR_APPR_PROCESS_START:
0.05273764467838875
Correlation between AMT_CREDIT and REG_REGION_NOT_LIVE_REGION:
0.024009863420362577
Correlation between AMT_CREDIT and REG_REGION_NOT_WORK_REGION:
0.05192888904376901
Correlation between AMT_CREDIT and LIVE_REGION_NOT_WORK_REGION:
0.05260898463510743
Correlation between AMT_CREDIT and REG_CITY_NOT_LIVE_CITY: -
0.02688639203217113
Correlation between AMT_CREDIT and REG_CITY_NOT_WORK_CITY: -
0.018856162160153324
Correlation between AMT_CREDIT and LIVE_CITY_NOT_WORK_CITY:
8.081559967033301e-05
Correlation between AMT_CREDIT and EXT_SOURCE_1: 0.1684288997732408
Correlation between AMT_CREDIT and EXT_SOURCE_2: 0.13122793587324422
Correlation between AMT_CREDIT and EXT_SOURCE_3: 0.043516255467307655
Correlation between AMT_CREDIT and APARTMENTS_AVG:
0.060439225827111996
Correlation between AMT_CREDIT and BASEMENTAREA_AVG:
0.03922616454895798
Correlation between AMT_CREDIT and YEARS_BEGINEXPLUATATION_AVG:
0.006248757791297846
Correlation between AMT_CREDIT and YEARS_BUILD_AVG:
0.03587469910926304
```

```
Correlation between AMT_CREDIT and COMMONAREA_AVG: 0.04953654103383191
Correlation between AMT_CREDIT and ELEVATORS_AVG: 0.08063501985507392
Correlation between AMT_CREDIT and ENTRANCES_AVG: 0.014929275958488674
Correlation between AMT_CREDIT and FLOORSMAX_AVG: 0.10329576974789065
Correlation between AMT_CREDIT and FLOORSMIN_AVG: 0.07883228907176901
Correlation between AMT_CREDIT and LANDAREA_AVG: 0.006217970472558238
Correlation between AMT_CREDIT and LIVINGAPARTMENTS_AVG:
0.05878801295091067
Correlation between AMT_CREDIT and LIVINGAREA_AVG: 0.07214592307116335
Correlation between AMT_CREDIT and NONLIVINGAPARTMENTS_AVG:
0.014362257440574809
Correlation between AMT_CREDIT and NONLIVINGAREA_AVG:
0.037884524864310834
Correlation between AMT_CREDIT and APARTMENTS_MODE:
0.05307204666917764
Correlation between AMT_CREDIT and BASEMENTAREA_MODE:
0.031212687499085057
Correlation between AMT_CREDIT and YEARS_BEGINEXPLUATATION_MODE:
0.004803849937521007
Correlation between AMT_CREDIT and YEARS_BUILD_MODE:
0.03347769090588901
Correlation between AMT_CREDIT and COMMONAREA_MODE:
0.04234141331656315
Correlation between AMT_CREDIT and ELEVATORS_MODE: 0.07473989060992746
Correlation between AMT_CREDIT and ENTRANCES_MODE:
0.009361223501575467
Correlation between AMT_CREDIT and FLOORSMAX_MODE: 0.10041808945378886
Correlation between AMT_CREDIT and FLOORSMIN_MODE: 0.07548461793785806
Correlation between AMT_CREDIT and LANDAREA_MODE: 0.002532117987502865
Correlation between AMT_CREDIT and LIVINGAPARTMENTS_MODE:
0.05120766109472908
Correlation between AMT_CREDIT and LIVINGAREA_MODE: 0.0641417397480827
Correlation between AMT_CREDIT and NONLIVINGAPARTMENTS_MODE:
0.011106103032482563
Correlation between AMT_CREDIT and NONLIVINGAREA_MODE:
0.03239040703696545
Correlation between AMT_CREDIT and APARTMENTS_MEDI:
0.05868221353891732
Correlation between AMT_CREDIT and BASEMENTAREA_MEDI:
0.03728099335515085
Correlation between AMT_CREDIT and YEARS_BEGINEXPLUATATION_MEDI:
0.005765094016371986
Correlation between AMT_CREDIT and YEARS_BUILD_MEDI:
0.035589199461384814
Correlation between AMT_CREDIT and COMMONAREA_MEDI:
0.04856543458806807
Correlation between AMT_CREDIT and ELEVATORS_MEDI: 0.07909438259849577
Correlation between AMT_CREDIT and ENTRANCES_MEDI:
0.013692074339138558
```

```
Correlation between AMT_CREDIT and FLOORSMAX_MEDI: 0.10277029009650318
Correlation between AMT_CREDIT and FLOORSMIN_MEDI: 0.07837505102056644
Correlation between AMT_CREDIT and LANDAREA_MEDI: 0.005414971790057232
Correlation between AMT_CREDIT and LIVINGAPARTMENTS_MEDI:
0.05705784887270478‌5
Correlation between AMT_CREDIT and LIVINGAREA_MEDI:
0.07085961905209327
Correlation between AMT_CREDIT and NONLIVINGAPARTMENTS_MEDI:
0.01340168191209257
Correlation between AMT_CREDIT and NONLIVINGAREA_MEDI:
0.035829431332536736
Correlation between AMT_CREDIT and TOTALAREA_MODE: 0.07281803102925878
Correlation between AMT_CREDIT and OBS_30_CNT_SOCIAL_CIRCLE:
0.00018955131512404333
Correlation between AMT_CREDIT and DEF_30_CNT_SOCIAL_CIRCLE: -
0.021229440364687283
Correlation between AMT_CREDIT and OBS_60_CNT_SOCIAL_CIRCLE:
0.00023856508904488325
Correlation between AMT_CREDIT and DEF_60_CNT_SOCIAL_CIRCLE: -
0.023766984908497674
Correlation between AMT_CREDIT and DAYS_LAST_PHONE_CHANGE: -
0.07370111227659121
Correlation between AMT_CREDIT and FLAG_DOCUMENT_2:
0.00890509529288391
Correlation between AMT_CREDIT and FLAG_DOCUMENT_3:
0.09636501231135611
Correlation between AMT_CREDIT and FLAG_DOCUMENT_4:
0.0006302630686693493
Correlation between AMT_CREDIT and FLAG_DOCUMENT_5: -
0.011756414180512613
Correlation between AMT_CREDIT and FLAG_DOCUMENT_6: -
0.046716688894504005
Correlation between AMT_CREDIT and FLAG_DOCUMENT_7: -
0.004039726981010674
Correlation between AMT_CREDIT and FLAG_DOCUMENT_8: 0.082819300915849
Correlation between AMT_CREDIT and FLAG_DOCUMENT_9:
0.022602178970474274
Correlation between AMT_CREDIT and FLAG_DOCUMENT_10: -
0.003100154081162714
Correlation between AMT_CREDIT and FLAG_DOCUMENT_11:
0.02898564195854198
Correlation between AMT_CREDIT and FLAG_DOCUMENT_12:
0.003856746874381347
Correlation between AMT_CREDIT and FLAG_DOCUMENT_13:
0.05242887397921347
Correlation between AMT_CREDIT and FLAG_DOCUMENT_14:
0.048828421582429685
Correlation between AMT_CREDIT and FLAG_DOCUMENT_15:
0.03225196285195589
```

```
Correlation between AMT_CREDIT and FLAG_DOCUMENT_16:
0.061925372673529044
Correlation between AMT_CREDIT and FLAG_DOCUMENT_17:
0.01174335141930548
Correlation between AMT_CREDIT and FLAG_DOCUMENT_18:
0.034329099084095344
Correlation between AMT_CREDIT and FLAG_DOCUMENT_19:
0.021081610105608346
Correlation between AMT_CREDIT and FLAG_DOCUMENT_20:
0.031023062059498268
Correlation between AMT_CREDIT and FLAG_DOCUMENT_21: -
0.016148280699585547
Correlation between AMT_CREDIT and AMT_REQ_CREDIT_BUREAU_HOUR: -
0.0039059183670565292
Correlation between AMT_CREDIT and AMT_REQ_CREDIT_BUREAU_DAY:
0.004237703562044031
Correlation between AMT_CREDIT and AMT_REQ_CREDIT_BUREAU_WEEK: -
0.0012748106379535452
Correlation between AMT_CREDIT and AMT_REQ_CREDIT_BUREAU_MON:
0.054451451253393644
Correlation between AMT_CREDIT and AMT_REQ_CREDIT_BUREAU_QRT:
0.015925231396336725
Correlation between AMT_CREDIT and AMT_REQ_CREDIT_BUREAU_YEAR: -
0.048447739061690975
```