House Loan Data Analysis

Description

For safe and secure lending experience, it's important to analyze the past data. In this project, you have to build a deep learning model to predict the chance of default for future loans using the historical data. As you will see, this dataset is highly imbalanced and includes a lot of features that make this problem more challenging.

Objective: Create a model that predicts whether or not an applicant will be able to repay a loan using historical data.

Domain: Finance

Analysis to be done: Perform data preprocessing and build a deep learning prediction model.

Steps to be done:

- Load the dataset that is given to you
- Check for null values in the dataset
- Print percentage of default to payer of the dataset for the TARGET column
- Balance the dataset if the data is imbalanced
- Plot the balanced data or imbalanced data
- Encode the columns that is required for the model
- Calculate Sensitivity as a metrice
- Calculate area under receiver operating characteristics curve

```
#import necessary libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import recall_score, roc_auc_score
from sklearn.metrics import accuracy_score, precision_score, 
recall_score, f1_score, roc_auc_score
from imblearn.over_sampling import RandomOverSampler
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import Dropout
pd.options.display.max_columns = None
import warnings
warnings.filterwarnings('ignore')

# Load the dataset
df = pd.read_csv('/content/loan_data (1).csv')
```

```
df.shape
```

```
(307511, 122)
```

```
df.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

   AMT_GOODS_PRICE NAME_TYPE_SUITE NAME_INCOME_TYPE  \
0         351000.0   Unaccompanied          Working
1        1129500.0          Family    State servant
2         135000.0   Unaccompanied          Working
3         297000.0   Unaccompanied          Working
4         513000.0   Unaccompanied          Working

             NAME_EDUCATION_TYPE    NAME_FAMILY_STATUS
NAME_HOUSING_TYPE  \
0  Secondary / secondary special  Single / not married  House /
apartment
1               Higher education               Married  House /
apartment
2  Secondary / secondary special  Single / not married  House /
apartment
3  Secondary / secondary special        Civil marriage  House /
apartment
4  Secondary / secondary special  Single / not married  House /
apartment

   REGION_POPULATION_RELATIVE  DAYS_BIRTH  DAYS_EMPLOYED
DAYS_REGISTRATION  \
0                    0.018801       -9461           -637              -
3648.0
```

```
1                    0.003541      -16765          -1188             -
1186.0
2                    0.010032      -19046           -225             -
4260.0
3                    0.008019      -19005          -3039             -
9833.0
4                    0.028663      -19932          -3038             -
4311.0

   DAYS_ID_PUBLISH  OWN_CAR_AGE  FLAG_MOBIL  FLAG_EMP_PHONE
FLAG_WORK_PHONE  \
0            -2120          NaN           1               1
0
1             -291          NaN           1               1
0
2            -2531         26.0           1               1
1
3            -2437          NaN           1               1
0
4            -3458          NaN           1               1
0

   FLAG_CONT_MOBILE  FLAG_PHONE  FLAG_EMAIL OCCUPATION_TYPE
CNT_FAM_MEMBERS  \
0                 1           1           0        Laborers
1.0
1                 1           1           0      Core staff
2.0
2                 1           1           0        Laborers
1.0
3                 1           0           0        Laborers
2.0
4                 1           0           0      Core staff
1.0

   REGION_RATING_CLIENT  REGION_RATING_CLIENT_W_CITY  \
0                     2                            2
1                     1                            1
2                     2                            2
3                     2                            2
4                     2                            2

  WEEKDAY_APPR_PROCESS_START  HOUR_APPR_PROCESS_START  \
0                  WEDNESDAY                       10
1                     MONDAY                       11
2                     MONDAY                        9
3                  WEDNESDAY                       17
4                   THURSDAY                       11

   REG_REGION_NOT_LIVE_REGION  REG_REGION_NOT_WORK_REGION  \
```

```
0                            0                            0
1                            0                            0
2                            0                            0
3                            0                            0
4                            0                            0

   LIVE_REGION_NOT_WORK_REGION  REG_CITY_NOT_LIVE_CITY  \
0                            0                       0
1                            0                       0
2                            0                       0
3                            0                       0
4                            0                       0

   REG_CITY_NOT_WORK_CITY  LIVE_CITY_NOT_WORK_CITY
ORGANIZATION_TYPE  \
0                       0                        0  Business Entity
Type 3
1                       0                        0
School
2                       0                        0
Government
3                       0                        0  Business Entity
Type 3
4                       1                        1
Religion

   EXT_SOURCE_1  EXT_SOURCE_2  EXT_SOURCE_3  APARTMENTS_AVG
BASEMENTAREA_AVG  \
0      0.083037      0.262949      0.139376          0.0247
0.0369
1      0.311267      0.622246           NaN          0.0959
0.0529
2           NaN      0.555912      0.729567             NaN
NaN
3           NaN      0.650442           NaN             NaN
NaN
4           NaN      0.322738           NaN             NaN
NaN

   YEARS_BEGINEXPLUATATION_AVG  YEARS_BUILD_AVG  COMMONAREA_AVG  \
0                       0.9722           0.6192          0.0143
1                       0.9851           0.7960          0.0605
2                          NaN              NaN             NaN
3                          NaN              NaN             NaN
4                          NaN              NaN             NaN

   ELEVATORS_AVG  ENTRANCES_AVG  FLOORSMAX_AVG  FLOORSMIN_AVG
LANDAREA_AVG  \
0           0.00         0.0690         0.0833         0.1250
0.0369
```

```
1          0.08         0.0345         0.2917         0.3333
0.0130
2           NaN            NaN            NaN            NaN
NaN
3           NaN            NaN            NaN            NaN
NaN
4           NaN            NaN            NaN            NaN
NaN

   LIVINGAPARTMENTS_AVG  LIVINGAREA_AVG  NONLIVINGAPARTMENTS_AVG  \
0                0.0202          0.0190                   0.0000
1                0.0773          0.0549                   0.0039
2                   NaN             NaN                      NaN
3                   NaN             NaN                      NaN
4                   NaN             NaN                      NaN

   NONLIVINGAREA_AVG  APARTMENTS_MODE  BASEMENTAREA_MODE  \
0             0.0000           0.0252             0.0383
1             0.0098           0.0924             0.0538
2                NaN              NaN                NaN
3                NaN              NaN                NaN
4                NaN              NaN                NaN

   YEARS_BEGINEXPLUATATION_MODE  YEARS_BUILD_MODE  COMMONAREA_MODE  \
0                        0.9722            0.6341           0.0144
1                        0.9851            0.8040           0.0497
2                           NaN               NaN              NaN
3                           NaN               NaN              NaN
4                           NaN               NaN              NaN

   ELEVATORS_MODE  ENTRANCES_MODE  FLOORSMAX_MODE  FLOORSMIN_MODE  \
0          0.0000          0.0690          0.0833          0.1250
1          0.0806          0.0345          0.2917          0.3333
2             NaN             NaN             NaN             NaN
3             NaN             NaN             NaN             NaN
4             NaN             NaN             NaN             NaN

   LANDAREA_MODE  LIVINGAPARTMENTS_MODE  LIVINGAREA_MODE  \
0         0.0377                  0.022           0.0198
1         0.0128                  0.079           0.0554
2            NaN                    NaN              NaN
3            NaN                    NaN              NaN
4            NaN                    NaN              NaN

   NONLIVINGAPARTMENTS_MODE  NONLIVINGAREA_MODE  APARTMENTS_MEDI  \
0                       0.0                 0.0           0.0250
1                       0.0                 0.0           0.0968
2                       NaN                 NaN              NaN
3                       NaN                 NaN              NaN
4                       NaN                 NaN              NaN
```

```
   BASEMENTAREA_MEDI  YEARS_BEGINEXPLUATATION_MEDI
YEARS_BUILD_MEDI  \
0             0.0369                        0.9722            0.6243

1             0.0529                        0.9851            0.7987

2                NaN                           NaN               NaN

3                NaN                           NaN               NaN

4                NaN                           NaN               NaN


   COMMONAREA_MEDI  ELEVATORS_MEDI  ENTRANCES_MEDI  FLOORSMAX_MEDI  \
0           0.0144            0.00          0.0690          0.0833
1           0.0608            0.08          0.0345          0.2917
2              NaN             NaN             NaN             NaN
3              NaN             NaN             NaN             NaN
4              NaN             NaN             NaN             NaN

   FLOORSMIN_MEDI  LANDAREA_MEDI  LIVINGAPARTMENTS_MEDI
LIVINGAREA_MEDI  \
0          0.1250         0.0375                 0.0205
0.0193
1          0.3333         0.0132                 0.0787
0.0558
2             NaN            NaN                    NaN
NaN
3             NaN            NaN                    NaN
NaN
4             NaN            NaN                    NaN
NaN

   NONLIVINGAPARTMENTS_MEDI  NONLIVINGAREA_MEDI FONDKAPREMONT_MODE  \
0                    0.0000                0.00   reg oper account
1                    0.0039                0.01   reg oper account
2                       NaN                 NaN                NaN
3                       NaN                 NaN                NaN
4                       NaN                 NaN                NaN

   HOUSETYPE_MODE  TOTALAREA_MODE WALLSMATERIAL_MODE
EMERGENCYSTATE_MODE  \
0  block of flats          0.0149       Stone, brick
No
1  block of flats          0.0714              Block
No
2             NaN             NaN                NaN
NaN
3             NaN             NaN                NaN
```

```
NaN
4              NaN              NaN                NaN
NaN

   OBS_30_CNT_SOCIAL_CIRCLE  DEF_30_CNT_SOCIAL_CIRCLE  \
0                       2.0                       2.0
1                       1.0                       0.0
2                       0.0                       0.0
3                       2.0                       0.0
4                       0.0                       0.0

   OBS_60_CNT_SOCIAL_CIRCLE  DEF_60_CNT_SOCIAL_CIRCLE
DAYS_LAST_PHONE_CHANGE  \
0                       2.0                       2.0
-1134.0
1                       1.0                       0.0
-828.0
2                       0.0                       0.0
-815.0
3                       2.0                       0.0
-617.0
4                       0.0                       0.0
-1106.0

   FLAG_DOCUMENT_2  FLAG_DOCUMENT_3  FLAG_DOCUMENT_4  FLAG_DOCUMENT_5
\
0                0                1                0                0

1                0                1                0                0

2                0                0                0                0

3                0                1                0                0

4                0                0                0                0


   FLAG_DOCUMENT_6  FLAG_DOCUMENT_7  FLAG_DOCUMENT_8  FLAG_DOCUMENT_9
\
0                0                0                0                0

1                0                0                0                0

2                0                0                0                0

3                0                0                0                0

4                0                0                1                0


   FLAG_DOCUMENT_10  FLAG_DOCUMENT_11  FLAG_DOCUMENT_12
```

```
FLAG_DOCUMENT_13  \
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

   FLAG_DOCUMENT_14  FLAG_DOCUMENT_15  FLAG_DOCUMENT_16
FLAG_DOCUMENT_17  \
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

   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
```

```
df.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
```

```
df.describe() #summary of whole data
```

```
          SK_ID_CURR         TARGET   CNT_CHILDREN
AMT_INCOME_TOTAL  \
count  307511.000000  307511.000000  307511.000000      3.075110e+05

mean   278180.518577       0.080729       0.417052      1.687979e+05

std    102790.175348       0.272419       0.722121      2.371231e+05

min    100002.000000       0.000000       0.000000      2.565000e+04

25%    189145.500000       0.000000       0.000000      1.125000e+05

50%    278202.000000       0.000000       0.000000      1.471500e+05

75%    367142.500000       0.000000       1.000000      2.025000e+05

max    456255.000000       1.000000      19.000000      1.170000e+08


         AMT_CREDIT    AMT_ANNUITY  AMT_GOODS_PRICE  \
count  3.075110e+05  307499.000000     3.072330e+05
mean   5.990260e+05   27108.573909     5.383962e+05
std    4.024908e+05   14493.737315     3.694465e+05
min    4.500000e+04    1615.500000     4.050000e+04
25%    2.700000e+05   16524.000000     2.385000e+05
50%    5.135310e+05   24903.000000     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               307511.000000  307511.000000  307511.000000
mean                     0.020868  -16036.995067   63815.045904
```

```
std                         0.013831    4363.988632  141275.766519
min                         0.000290  -25229.000000  -17912.000000
25%                         0.010006  -19682.000000   -2760.000000
50%                         0.018850  -15750.000000   -1213.000000
75%                         0.028663  -12413.000000    -289.000000
max                         0.072508   -7489.000000  365243.000000

       DAYS_REGISTRATION  DAYS_ID_PUBLISH    OWN_CAR_AGE
FLAG_MOBIL  \
count      307511.000000    307511.000000  104582.000000
307511.000000
mean        -4986.120328     -2994.202373      12.061091
0.999997
std          3522.886321      1509.450419      11.944812
0.001803
min        -24672.000000     -7197.000000       0.000000
0.000000
25%         -7479.500000     -4299.000000       5.000000
1.000000
50%         -4504.000000     -3254.000000       9.000000
1.000000
75%         -2010.000000     -1720.000000      15.000000
1.000000
max             0.000000         0.000000      91.000000
1.000000

       FLAG_EMP_PHONE  FLAG_WORK_PHONE  FLAG_CONT_MOBILE
FLAG_PHONE  \
count   307511.000000    307511.000000     307511.000000
307511.000000
mean         0.819889         0.199368          0.998133
0.281066
std          0.384280         0.399526          0.043164
0.449521
min          0.000000         0.000000          0.000000
0.000000
25%          1.000000         0.000000          1.000000
0.000000
50%          1.000000         0.000000          1.000000
0.000000
75%          1.000000         0.000000          1.000000
1.000000
max          1.000000         1.000000          1.000000
1.000000

          FLAG_EMAIL  CNT_FAM_MEMBERS  REGION_RATING_CLIENT  \
count  307511.000000    307509.000000         307511.000000
mean        0.056720         2.152665              2.052463
std         0.231307         0.910682              0.509034
min         0.000000         1.000000              1.000000
```

```
25%         0.000000         2.000000                     2.000000
50%         0.000000         2.000000                     2.000000
75%         0.000000         3.000000                     2.000000
max         1.000000        20.000000                     3.000000

       REGION_RATING_CLIENT_W_CITY  HOUR_APPR_PROCESS_START  \
count                307511.000000            307511.000000
mean                      2.031521                12.063419
std                       0.502737                 3.265832
min                       1.000000                 0.000000
25%                       2.000000                10.000000
50%                       2.000000                12.000000
75%                       2.000000                14.000000
max                       3.000000                23.000000

       REG_REGION_NOT_LIVE_REGION  REG_REGION_NOT_WORK_REGION  \
count               307511.000000               307511.000000
mean                     0.015144                    0.050769
std                      0.122126                    0.219526
min                      0.000000                    0.000000
25%                      0.000000                    0.000000
50%                      0.000000                    0.000000
75%                      0.000000                    0.000000
max                      1.000000                    1.000000

       LIVE_REGION_NOT_WORK_REGION  REG_CITY_NOT_LIVE_CITY  \
count                307511.000000           307511.000000
mean                      0.040659                0.078173
std                       0.197499                0.268444
min                       0.000000                0.000000
25%                       0.000000                0.000000
50%                       0.000000                0.000000
75%                       0.000000                0.000000
max                       1.000000                1.000000

       REG_CITY_NOT_WORK_CITY  LIVE_CITY_NOT_WORK_CITY   EXT_SOURCE_1
\
count           307511.000000            307511.000000  134133.000000

mean                 0.230454                 0.179555       0.502130

std                  0.421124                 0.383817       0.211062

min                  0.000000                 0.000000       0.014568

25%                  0.000000                 0.000000       0.334007

50%                  0.000000                 0.000000       0.505998

75%                  0.000000                 0.000000       0.675053
```

```
max                     1.000000                  1.000000       0.962693


       EXT_SOURCE_2   EXT_SOURCE_3  APARTMENTS_AVG
BASEMENTAREA_AVG  \
count  3.068510e+05  246546.000000    151450.00000     127568.000000

mean   5.143927e-01       0.510853         0.11744          0.088442

std    1.910602e-01       0.194844         0.10824          0.082438

min    8.173617e-08       0.000527         0.00000          0.000000

25%    3.924574e-01       0.370650         0.05770          0.044200

50%    5.659614e-01       0.535276         0.08760          0.076300

75%    6.636171e-01       0.669057         0.14850          0.112200

max    8.549997e-01       0.896010         1.00000          1.000000


       YEARS_BEGINEXPLUATATION_AVG  YEARS_BUILD_AVG  COMMONAREA_AVG  \
count                157504.000000    103023.000000    92646.000000
mean                      0.977735         0.752471        0.044621
std                       0.059223         0.113280        0.076036
min                       0.000000         0.000000        0.000000
25%                       0.976700         0.687200        0.007800
50%                       0.981600         0.755200        0.021100
75%                       0.986600         0.823200        0.051500
max                       1.000000         1.000000        1.000000

       ELEVATORS_AVG  ENTRANCES_AVG  FLOORSMAX_AVG  FLOORSMIN_AVG  \
count  143620.000000  152683.000000  154491.000000   98869.000000
mean        0.078942       0.149725       0.226282       0.231894
std         0.134576       0.100049       0.144641       0.161380
min         0.000000       0.000000       0.000000       0.000000
25%         0.000000       0.069000       0.166700       0.083300
50%         0.000000       0.137900       0.166700       0.208300
75%         0.120000       0.206900       0.333300       0.375000
max         1.000000       1.000000       1.000000       1.000000

        LANDAREA_AVG  LIVINGAPARTMENTS_AVG  LIVINGAREA_AVG  \
count  124921.000000          97312.000000   153161.000000
mean        0.066333              0.100775        0.107399
std         0.081184              0.092576        0.110565
min         0.000000              0.000000        0.000000
25%         0.018700              0.050400        0.045300
50%         0.048100              0.075600        0.074500
75%         0.085600              0.121000        0.129900
```

```
max        1.000000                1.000000       1.000000

       NONLIVINGAPARTMENTS_AVG  NONLIVINGAREA_AVG  APARTMENTS_MODE  \
count             93997.000000      137829.000000    151450.000000
mean                  0.008809           0.028358         0.114231
std                   0.047732           0.069523         0.107936
min                   0.000000           0.000000         0.000000
25%                   0.000000           0.000000         0.052500
50%                   0.000000           0.003600         0.084000
75%                   0.003900           0.027700         0.143900
max                   1.000000           1.000000         1.000000

       BASEMENTAREA_MODE  YEARS_BEGINEXPLUATATION_MODE
YEARS_BUILD_MODE  \
count      127568.000000                 157504.000000
103023.000000
mean            0.087543                      0.977065
0.759637
std             0.084307                      0.064575
0.110111
min             0.000000                      0.000000
0.000000
25%             0.040700                      0.976700
0.699400
50%             0.074600                      0.981600
0.764800
75%             0.112400                      0.986600
0.823600
max             1.000000                      1.000000
1.000000

       COMMONAREA_MODE  ELEVATORS_MODE  ENTRANCES_MODE  FLOORSMAX_MODE
\
count     92646.000000   143620.000000   152683.000000   154491.000000

mean          0.042553        0.074490        0.145193        0.222315

std           0.074445        0.132256        0.100977        0.143709

min           0.000000        0.000000        0.000000        0.000000

25%           0.007200        0.000000        0.069000        0.166700

50%           0.019000        0.000000        0.137900        0.166700

75%           0.049000        0.120800        0.206900        0.333300

max           1.000000        1.000000        1.000000        1.000000
```

```
       FLOORSMIN_MODE  LANDAREA_MODE  LIVINGAPARTMENTS_MODE
LIVINGAREA_MODE  \
count    98869.000000  124921.000000           97312.000000
153161.000000
mean         0.228058       0.064958               0.105645
0.105975
std          0.161160       0.081750               0.097880
0.111845
min          0.000000       0.000000               0.000000
0.000000
25%          0.083300       0.016600               0.054200
0.042700
50%          0.208300       0.045800               0.077100
0.073100
75%          0.375000       0.084100               0.131300
0.125200
max          1.000000       1.000000               1.000000
1.000000

       NONLIVINGAPARTMENTS_MODE  NONLIVINGAREA_MODE
APARTMENTS_MEDI  \
count              93997.000000       137829.000000    151450.000000

mean                   0.008076            0.027022         0.117850

std                    0.046276            0.070254         0.109076

min                    0.000000            0.000000         0.000000

25%                    0.000000            0.000000         0.058300

50%                    0.000000            0.001100         0.086400

75%                    0.003900            0.023100         0.148900

max                    1.000000            1.000000         1.000000


       BASEMENTAREA_MEDI  YEARS_BEGINEXPLUATATION_MEDI
YEARS_BUILD_MEDI  \
count      127568.000000                 157504.000000
103023.000000
mean            0.087955                      0.977752
0.755746
std             0.082179                      0.059897
0.112066
min             0.000000                      0.000000
0.000000
25%             0.043700                      0.976700
0.691400
```

```
50%              0.075800                            0.981600
0.758500
75%              0.111600                            0.986600
0.825600
max              1.000000                            1.000000
1.000000

       COMMONAREA_MEDI  ELEVATORS_MEDI  ENTRANCES_MEDI  FLOORSMAX_MEDI
\
count     92646.000000   143620.000000   152683.000000   154491.000000

mean          0.044595        0.078078        0.149213        0.225897

std           0.076144        0.134467        0.100368        0.145067

min           0.000000        0.000000        0.000000        0.000000

25%           0.007900        0.000000        0.069000        0.166700

50%           0.020800        0.000000        0.137900        0.166700

75%           0.051300        0.120000        0.206900        0.333300

max           1.000000        1.000000        1.000000        1.000000


       FLOORSMIN_MEDI  LANDAREA_MEDI  LIVINGAPARTMENTS_MEDI
LIVINGAREA_MEDI  \
count    98869.000000  124921.000000           97312.000000
153161.000000
mean         0.231625       0.067169               0.101954
0.108607
std          0.161934       0.082167               0.093642
0.112260
min          0.000000       0.000000               0.000000
0.000000
25%          0.083300       0.018700               0.051300
0.045700
50%          0.208300       0.048700               0.076100
0.074900
75%          0.375000       0.086800               0.123100
0.130300
max          1.000000       1.000000               1.000000
1.000000

       NONLIVINGAPARTMENTS_MEDI  NONLIVINGAREA_MEDI  TOTALAREA_MODE  \
count              93997.000000       137829.000000   159080.000000
mean                   0.008651            0.028236        0.102547
std                    0.047415            0.070166        0.107462
min                    0.000000            0.000000        0.000000
```

```
25%                     0.000000             0.000000        0.041200
50%                     0.000000             0.003100        0.068800
75%                     0.003900             0.026600        0.127600
max                     1.000000             1.000000        1.000000

       OBS_30_CNT_SOCIAL_CIRCLE  DEF_30_CNT_SOCIAL_CIRCLE  \
count             306490.000000             306490.000000
mean                   1.422245                  0.143421
std                    2.400989                  0.446698
min                    0.000000                  0.000000
25%                    0.000000                  0.000000
50%                    0.000000                  0.000000
75%                    2.000000                  0.000000
max                  348.000000                 34.000000

       OBS_60_CNT_SOCIAL_CIRCLE  DEF_60_CNT_SOCIAL_CIRCLE  \
count             306490.000000             306490.000000
mean                   1.405292                  0.100049
std                    2.379803                  0.362291
min                    0.000000                  0.000000
25%                    0.000000                  0.000000
50%                    0.000000                  0.000000
75%                    2.000000                  0.000000
max                  344.000000                 24.000000

       DAYS_LAST_PHONE_CHANGE  FLAG_DOCUMENT_2  FLAG_DOCUMENT_3  \
count           307510.000000    307511.000000    307511.000000
mean              -962.858788         0.000042         0.710023
std                826.808487         0.006502         0.453752
min              -4292.000000         0.000000         0.000000
25%              -1570.000000         0.000000         0.000000
50%               -757.000000         0.000000         1.000000
75%               -274.000000         0.000000         1.000000
max                  0.000000         1.000000         1.000000

       FLAG_DOCUMENT_4  FLAG_DOCUMENT_5  FLAG_DOCUMENT_6
FLAG_DOCUMENT_7  \
count    307511.000000    307511.000000    307511.000000
307511.000000
mean          0.000081         0.015115         0.088055
0.000192
std           0.009016         0.122010         0.283376
0.013850
min           0.000000         0.000000         0.000000
0.000000
25%           0.000000         0.000000         0.000000
0.000000
50%           0.000000         0.000000         0.000000
0.000000
75%           0.000000         0.000000         0.000000
```

```
0.000000
max              1.000000         1.000000         1.000000
1.000000

       FLAG_DOCUMENT_8  FLAG_DOCUMENT_9  FLAG_DOCUMENT_10
FLAG_DOCUMENT_11  \
count    307511.000000    307511.000000     307511.000000
307511.000000
mean          0.081376         0.003896          0.000023
0.003912
std           0.273412         0.062295          0.004771
0.062424
min           0.000000         0.000000          0.000000
0.000000
25%           0.000000         0.000000          0.000000
0.000000
50%           0.000000         0.000000          0.000000
0.000000
75%           0.000000         0.000000          0.000000
0.000000
max           1.000000         1.000000          1.000000
1.000000

       FLAG_DOCUMENT_12  FLAG_DOCUMENT_13  FLAG_DOCUMENT_14
FLAG_DOCUMENT_15  \
count     307511.000000     307511.000000     307511.000000
307511.00000
mean           0.000007          0.003525          0.002936
0.00121
std            0.002550          0.059268          0.054110
0.03476
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

       FLAG_DOCUMENT_16  FLAG_DOCUMENT_17  FLAG_DOCUMENT_18
FLAG_DOCUMENT_19  \
count     307511.000000     307511.000000     307511.000000
307511.000000
mean           0.009928          0.000267          0.008130
0.000595
std            0.099144          0.016327          0.089798
0.024387
```

```
min              0.000000          0.000000          0.000000
0.000000
25%              0.000000          0.000000          0.000000
0.000000
50%              0.000000          0.000000          0.000000
0.000000
75%              0.000000          0.000000          0.000000
0.000000
max              1.000000          1.000000          1.000000
1.000000

       FLAG_DOCUMENT_20  FLAG_DOCUMENT_21  AMT_REQ_CREDIT_BUREAU_HOUR
\
count     307511.000000     307511.000000               265992.000000

mean           0.000507          0.000335                    0.006402

std            0.022518          0.018299                    0.083849

min            0.000000          0.000000                    0.000000

25%            0.000000          0.000000                    0.000000

50%            0.000000          0.000000                    0.000000

75%            0.000000          0.000000                    0.000000

max            1.000000          1.000000                    4.000000


       AMT_REQ_CREDIT_BUREAU_DAY  AMT_REQ_CREDIT_BUREAU_WEEK  \
count              265992.000000               265992.000000
mean                    0.007000                    0.034362
std                     0.110757                    0.204685
min                     0.000000                    0.000000
25%                     0.000000                    0.000000
50%                     0.000000                    0.000000
75%                     0.000000                    0.000000
max                     9.000000                    8.000000

       AMT_REQ_CREDIT_BUREAU_MON  AMT_REQ_CREDIT_BUREAU_QRT  \
count              265992.000000              265992.000000
mean                    0.267395                   0.265474
std                     0.916002                   0.794056
min                     0.000000                   0.000000
25%                     0.000000                   0.000000
50%                     0.000000                   0.000000
75%                     0.000000                   0.000000
max                    27.000000                 261.000000
```

```
       AMT_REQ_CREDIT_BUREAU_YEAR
count               265992.000000
mean                     1.899974
std                      1.869295
min                      0.000000
25%                      0.000000
50%                      1.000000
75%                      3.000000
max                     25.000000
```

```python
df.describe(include='O') # summary of object columns
```

```
       NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR FLAG_OWN_REALTY  \
count              307511      307511       307511          307511
unique                  2           3            2               2
top            Cash loans           F            N               Y
freq               278232      202448       202924          213312

       NAME_TYPE_SUITE NAME_INCOME_TYPE            NAME_EDUCATION_TYPE
\
count           306219           307511                         307511

unique               7                8                              5

top       Unaccompanied          Working  Secondary / secondary special

freq            248526           158774                         218391


       NAME_FAMILY_STATUS  NAME_HOUSING_TYPE OCCUPATION_TYPE  \
count              307511             307511          211120
unique                  6                  6              18
top               Married  House / apartment        Laborers
freq               196432             272868           55186

       WEEKDAY_APPR_PROCESS_START       ORGANIZATION_TYPE
FONDKAPREMONT_MODE  \
count                      307511                  307511
97216
unique                          7                      58
4
top                       TUESDAY  Business Entity Type 3   reg oper
account
freq                        53901                   67992
73830

        HOUSETYPE_MODE WALLSMATERIAL_MODE EMERGENCYSTATE_MODE
count           153214             151170              161756
unique               3                  7                   2
```

```
top     block of flats              Panel                  No
freq            150503              66040              159428
```

```python
pd.options.display.max_rows = None
df.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)
```

```python
df.set_index(keys=['SK_ID_CURR'],inplace=True) # set column
'SK_ID_CURR' as index

# Check percentage of defaults
default_percentage = df['TARGET'].mean() * 100
print(f"Percentage of defaults: {default_percentage:.2f}%")
```

```
Percentage of defaults: 8.07%
```

Handling Missing Values

```python
pd.options.display.max_rows = None
# Check for null values
null_values = df.isnull().sum()
null_values
```

```
TARGET                            0
NAME_CONTRACT_TYPE                0
CODE_GENDER                       0
FLAG_OWN_CAR                      0
FLAG_OWN_REALTY                   0
CNT_CHILDREN                      0
AMT_INCOME_TOTAL                  0
AMT_CREDIT                        0
AMT_ANNUITY                      12
AMT_GOODS_PRICE                 278
NAME_TYPE_SUITE                1292
NAME_INCOME_TYPE                  0
NAME_EDUCATION_TYPE               0
NAME_FAMILY_STATUS                0
NAME_HOUSING_TYPE                 0
REGION_POPULATION_RELATIVE        0
```

```
DAYS_BIRTH                           0
DAYS_EMPLOYED                        0
DAYS_REGISTRATION                    0
DAYS_ID_PUBLISH                      0
OWN_CAR_AGE                     202929
FLAG_MOBIL                           0
FLAG_EMP_PHONE                       0
FLAG_WORK_PHONE                      0
FLAG_CONT_MOBILE                     0
FLAG_PHONE                           0
FLAG_EMAIL                           0
OCCUPATION_TYPE                  96391
CNT_FAM_MEMBERS                      2
REGION_RATING_CLIENT                 0
REGION_RATING_CLIENT_W_CITY          0
WEEKDAY_APPR_PROCESS_START           0
HOUR_APPR_PROCESS_START              0
REG_REGION_NOT_LIVE_REGION           0
REG_REGION_NOT_WORK_REGION           0
LIVE_REGION_NOT_WORK_REGION          0
REG_CITY_NOT_LIVE_CITY               0
REG_CITY_NOT_WORK_CITY               0
LIVE_CITY_NOT_WORK_CITY              0
ORGANIZATION_TYPE                    0
EXT_SOURCE_1                    173378
EXT_SOURCE_2                       660
EXT_SOURCE_3                     60965
APARTMENTS_AVG                  156061
BASEMENTAREA_AVG                179943
YEARS_BEGINEXPLUATATION_AVG     150007
YEARS_BUILD_AVG                 204488
COMMONAREA_AVG                  214865
ELEVATORS_AVG                   163891
ENTRANCES_AVG                   154828
FLOORSMAX_AVG                   153020
FLOORSMIN_AVG                   208642
LANDAREA_AVG                    182590
LIVINGAPARTMENTS_AVG            210199
LIVINGAREA_AVG                  154350
NONLIVINGAPARTMENTS_AVG         213514
NONLIVINGAREA_AVG               169682
APARTMENTS_MODE                 156061
BASEMENTAREA_MODE               179943
YEARS_BEGINEXPLUATATION_MODE    150007
YEARS_BUILD_MODE                204488
COMMONAREA_MODE                 214865
ELEVATORS_MODE                  163891
ENTRANCES_MODE                  154828
FLOORSMAX_MODE                  153020
```

FLOORSMIN_MODE	208642
LANDAREA_MODE	182590
LIVINGAPARTMENTS_MODE	210199
LIVINGAREA_MODE	154350
NONLIVINGAPARTMENTS_MODE	213514
NONLIVINGAREA_MODE	169682
APARTMENTS_MEDI	156061
BASEMENTAREA_MEDI	179943
YEARS_BEGINEXPLUATATION_MEDI	150007
YEARS_BUILD_MEDI	204488
COMMONAREA_MEDI	214865
ELEVATORS_MEDI	163891
ENTRANCES_MEDI	154828
FLOORSMAX_MEDI	153020
FLOORSMIN_MEDI	208642
LANDAREA_MEDI	182590
LIVINGAPARTMENTS_MEDI	210199
LIVINGAREA_MEDI	154350
NONLIVINGAPARTMENTS_MEDI	213514
NONLIVINGAREA_MEDI	169682
FONDKAPREMONT_MODE	210295
HOUSETYPE_MODE	154297
TOTALAREA_MODE	148431
WALLSMATERIAL_MODE	156341
EMERGENCYSTATE_MODE	145755
OBS_30_CNT_SOCIAL_CIRCLE	1021
DEF_30_CNT_SOCIAL_CIRCLE	1021
OBS_60_CNT_SOCIAL_CIRCLE	1021
DEF_60_CNT_SOCIAL_CIRCLE	1021
DAYS_LAST_PHONE_CHANGE	1
FLAG_DOCUMENT_2	0
FLAG_DOCUMENT_3	0
FLAG_DOCUMENT_4	0
FLAG_DOCUMENT_5	0
FLAG_DOCUMENT_6	0
FLAG_DOCUMENT_7	0
FLAG_DOCUMENT_8	0
FLAG_DOCUMENT_9	0
FLAG_DOCUMENT_10	0
FLAG_DOCUMENT_11	0
FLAG_DOCUMENT_12	0
FLAG_DOCUMENT_13	0
FLAG_DOCUMENT_14	0
FLAG_DOCUMENT_15	0
FLAG_DOCUMENT_16	0
FLAG_DOCUMENT_17	0
FLAG_DOCUMENT_18	0
FLAG_DOCUMENT_19	0
FLAG_DOCUMENT_20	0

```
FLAG_DOCUMENT_21                    0
AMT_REQ_CREDIT_BUREAU_HOUR      41519
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
dtype: int64
```

```python
# Separate numeric and categorical columns
numeric_columns = df.select_dtypes(include=['int64',
'float64']).columns
categorical_columns = df.select_dtypes(include=['object']).columns

# Impute missing values with mean for numeric columns
for col in numeric_columns:
    df[col].fillna(df[col].mean(), inplace=True)

# Impute missing values with mode for categorical columns
for col in categorical_columns:
    df[col].fillna(df[col].mode()[0], inplace=True)

# Check for null values after handling missing values
null_values_after = df.isnull().sum()
null_values_after
```

```
TARGET                          0
NAME_CONTRACT_TYPE              0
CODE_GENDER                     0
FLAG_OWN_CAR                    0
FLAG_OWN_REALTY                 0
CNT_CHILDREN                    0
AMT_INCOME_TOTAL                0
AMT_CREDIT                      0
AMT_ANNUITY                     0
AMT_GOODS_PRICE                 0
NAME_TYPE_SUITE                 0
NAME_INCOME_TYPE                0
NAME_EDUCATION_TYPE             0
NAME_FAMILY_STATUS              0
NAME_HOUSING_TYPE               0
REGION_POPULATION_RELATIVE      0
DAYS_BIRTH                      0
DAYS_EMPLOYED                   0
DAYS_REGISTRATION               0
DAYS_ID_PUBLISH                 0
OWN_CAR_AGE                     0
FLAG_MOBIL                      0
FLAG_EMP_PHONE                  0
FLAG_WORK_PHONE                 0
```

```
FLAG_CONT_MOBILE                0
FLAG_PHONE                      0
FLAG_EMAIL                      0
OCCUPATION_TYPE                 0
CNT_FAM_MEMBERS                 0
REGION_RATING_CLIENT            0
REGION_RATING_CLIENT_W_CITY     0
WEEKDAY_APPR_PROCESS_START      0
HOUR_APPR_PROCESS_START         0
REG_REGION_NOT_LIVE_REGION      0
REG_REGION_NOT_WORK_REGION      0
LIVE_REGION_NOT_WORK_REGION     0
REG_CITY_NOT_LIVE_CITY          0
REG_CITY_NOT_WORK_CITY          0
LIVE_CITY_NOT_WORK_CITY         0
ORGANIZATION_TYPE               0
EXT_SOURCE_1                    0
EXT_SOURCE_2                    0
EXT_SOURCE_3                    0
APARTMENTS_AVG                  0
BASEMENTAREA_AVG                0
YEARS_BEGINEXPLUATATION_AVG     0
YEARS_BUILD_AVG                 0
COMMONAREA_AVG                  0
ELEVATORS_AVG                   0
ENTRANCES_AVG                   0
FLOORSMAX_AVG                   0
FLOORSMIN_AVG                   0
LANDAREA_AVG                    0
LIVINGAPARTMENTS_AVG            0
LIVINGAREA_AVG                  0
NONLIVINGAPARTMENTS_AVG         0
NONLIVINGAREA_AVG               0
APARTMENTS_MODE                 0
BASEMENTAREA_MODE               0
YEARS_BEGINEXPLUATATION_MODE    0
YEARS_BUILD_MODE                0
COMMONAREA_MODE                 0
ELEVATORS_MODE                  0
ENTRANCES_MODE                  0
FLOORSMAX_MODE                  0
FLOORSMIN_MODE                  0
LANDAREA_MODE                   0
LIVINGAPARTMENTS_MODE           0
LIVINGAREA_MODE                 0
NONLIVINGAPARTMENTS_MODE        0
NONLIVINGAREA_MODE              0
APARTMENTS_MEDI                 0
BASEMENTAREA_MEDI               0
```

```
YEARS_BEGINEXPLUATATION_MEDI    0
YEARS_BUILD_MEDI                0
COMMONAREA_MEDI                 0
ELEVATORS_MEDI                  0
ENTRANCES_MEDI                  0
FLOORSMAX_MEDI                  0
FLOORSMIN_MEDI                  0
LANDAREA_MEDI                   0
LIVINGAPARTMENTS_MEDI           0
LIVINGAREA_MEDI                 0
NONLIVINGAPARTMENTS_MEDI        0
NONLIVINGAREA_MEDI              0
FONDKAPREMONT_MODE              0
HOUSETYPE_MODE                  0
TOTALAREA_MODE                  0
WALLSMATERIAL_MODE              0
EMERGENCYSTATE_MODE             0
OBS_30_CNT_SOCIAL_CIRCLE        0
DEF_30_CNT_SOCIAL_CIRCLE        0
OBS_60_CNT_SOCIAL_CIRCLE        0
DEF_60_CNT_SOCIAL_CIRCLE        0
DAYS_LAST_PHONE_CHANGE          0
FLAG_DOCUMENT_2                 0
FLAG_DOCUMENT_3                 0
FLAG_DOCUMENT_4                 0
FLAG_DOCUMENT_5                 0
FLAG_DOCUMENT_6                 0
FLAG_DOCUMENT_7                 0
FLAG_DOCUMENT_8                 0
FLAG_DOCUMENT_9                 0
FLAG_DOCUMENT_10                0
FLAG_DOCUMENT_11                0
FLAG_DOCUMENT_12                0
FLAG_DOCUMENT_13                0
FLAG_DOCUMENT_14                0
FLAG_DOCUMENT_15                0
FLAG_DOCUMENT_16                0
FLAG_DOCUMENT_17                0
FLAG_DOCUMENT_18                0
FLAG_DOCUMENT_19                0
FLAG_DOCUMENT_20                0
FLAG_DOCUMENT_21                0
AMT_REQ_CREDIT_BUREAU_HOUR      0
AMT_REQ_CREDIT_BUREAU_DAY       0
AMT_REQ_CREDIT_BUREAU_WEEK      0
AMT_REQ_CREDIT_BUREAU_MON       0
AMT_REQ_CREDIT_BUREAU_QRT       0
AMT_REQ_CREDIT_BUREAU_YEAR      0
dtype: int64
```

```
df.head()

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

           FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT
\
SK_ID_CURR

100002                   Y             0          202500.0    406597.5

100003                   N             0          270000.0   1293502.5

100004                   Y             0           67500.0    135000.0

100006                   Y             0          135000.0    312682.5

100007                   Y             0          121500.0    513000.0


            AMT_ANNUITY  AMT_GOODS_PRICE NAME_TYPE_SUITE
NAME_INCOME_TYPE  \
SK_ID_CURR

100002          24700.5         351000.0   Unaccompanied
Working
100003          35698.5        1129500.0          Family    State
servant
100004           6750.0         135000.0   Unaccompanied
Working
100006          29686.5         297000.0   Unaccompanied
Working
100007          21865.5         513000.0   Unaccompanied
Working

                      NAME_EDUCATION_TYPE    NAME_FAMILY_STATUS  \
SK_ID_CURR
100002      Secondary / secondary special  Single / not married
100003                   Higher education               Married
100004      Secondary / secondary special  Single / not married
100006      Secondary / secondary special        Civil marriage
100007      Secondary / secondary special  Single / not married

            NAME_HOUSING_TYPE  REGION_POPULATION_RELATIVE  DAYS_BIRTH
\
SK_ID_CURR
```

```
100002      House / apartment                    0.018801       -9461
100003      House / apartment                    0.003541      -16765
100004      House / apartment                    0.010032      -19046
100006      House / apartment                    0.008019      -19005
100007      House / apartment                    0.028663      -19932

            DAYS_EMPLOYED  DAYS_REGISTRATION  DAYS_ID_PUBLISH
OWN_CAR_AGE  \
SK_ID_CURR

100002               -637            -3648.0            -2120
12.061091
100003              -1188            -1186.0             -291
12.061091
100004               -225            -4260.0            -2531
26.000000
100006              -3039            -9833.0            -2437
12.061091
100007              -3038            -4311.0            -3458
12.061091

            FLAG_MOBIL  FLAG_EMP_PHONE  FLAG_WORK_PHONE
FLAG_CONT_MOBILE  \
SK_ID_CURR

100002               1               1                0
1
100003               1               1                0
1
100004               1               1                1
1
100006               1               1                0
1
100007               1               1                0
1

            FLAG_PHONE  FLAG_EMAIL OCCUPATION_TYPE  CNT_FAM_MEMBERS  \
SK_ID_CURR
100002               1           0        Laborers              1.0
100003               1           0      Core staff              2.0
100004               1           0        Laborers              1.0
100006               0           0        Laborers              2.0
100007               0           0      Core staff              1.0
```

```
            REGION_RATING_CLIENT  REGION_RATING_CLIENT_W_CITY  \
SK_ID_CURR
100002                         2                            2
100003                         1                            1
100004                         2                            2
100006                         2                            2
100007                         2                            2

           WEEKDAY_APPR_PROCESS_START  HOUR_APPR_PROCESS_START  \
SK_ID_CURR
100002                      WEDNESDAY                       10
100003                         MONDAY                       11
100004                         MONDAY                        9
100006                      WEDNESDAY                       17
100007                       THURSDAY                       11

            REG_REGION_NOT_LIVE_REGION  REG_REGION_NOT_WORK_REGION  \
SK_ID_CURR
100002                               0                           0
100003                               0                           0
100004                               0                           0
100006                               0                           0
100007                               0                           0

            LIVE_REGION_NOT_WORK_REGION  REG_CITY_NOT_LIVE_CITY  \
SK_ID_CURR
100002                                0                       0
100003                                0                       0
100004                                0                       0
100006                                0                       0
100007                                0                       0

            REG_CITY_NOT_WORK_CITY  LIVE_CITY_NOT_WORK_CITY  \
SK_ID_CURR
100002                           0                        0
100003                           0                        0
100004                           0                        0
100006                           0                        0
100007                           1                        1

                 ORGANIZATION_TYPE  EXT_SOURCE_1  EXT_SOURCE_2
EXT_SOURCE_3  \
SK_ID_CURR

100002      Business Entity Type 3      0.083037      0.262949
0.139376
100003                      School      0.311267      0.622246
0.510853
100004                  Government      0.502130      0.555912
0.729567
```

```
100006      Business Entity Type 3      0.502130      0.650442
0.510853
100007                    Religion      0.502130      0.322738
0.510853

            APARTMENTS_AVG  BASEMENTAREA_AVG
YEARS_BEGINEXPLUATATION_AVG  \
SK_ID_CURR

100002             0.02470          0.036900
0.972200
100003             0.09590          0.052900
0.985100
100004             0.11744          0.088442
0.977735
100006             0.11744          0.088442
0.977735
100007             0.11744          0.088442
0.977735

            YEARS_BUILD_AVG  COMMONAREA_AVG  ELEVATORS_AVG
ENTRANCES_AVG  \
SK_ID_CURR

100002             0.619200        0.014300       0.000000
0.069000
100003             0.796000        0.060500       0.080000
0.034500
100004             0.752471        0.044621       0.078942
0.149725
100006             0.752471        0.044621       0.078942
0.149725
100007             0.752471        0.044621       0.078942
0.149725

            FLOORSMAX_AVG  FLOORSMIN_AVG  LANDAREA_AVG
LIVINGAPARTMENTS_AVG  \
SK_ID_CURR

100002           0.083300       0.125000      0.036900
0.020200
100003           0.291700       0.333300      0.013000
0.077300
100004           0.226282       0.231894      0.066333
0.100775
100006           0.226282       0.231894      0.066333
0.100775
100007           0.226282       0.231894      0.066333
0.100775
```

```
            LIVINGAREA_AVG  NONLIVINGAPARTMENTS_AVG  NONLIVINGAREA_AVG
\
SK_ID_CURR

100002            0.019000                 0.000000           0.000000

100003            0.054900                 0.003900           0.009800

100004            0.107399                 0.008809           0.028358

100006            0.107399                 0.008809           0.028358

100007            0.107399                 0.008809           0.028358


            APARTMENTS_MODE  BASEMENTAREA_MODE
YEARS_BEGINEXPLUATATION_MODE  \
SK_ID_CURR

100002             0.025200           0.038300
0.972200
100003             0.092400           0.053800
0.985100
100004             0.114231           0.087543
0.977065
100006             0.114231           0.087543
0.977065
100007             0.114231           0.087543
0.977065

            YEARS_BUILD_MODE  COMMONAREA_MODE  ELEVATORS_MODE
ENTRANCES_MODE  \
SK_ID_CURR

100002              0.634100         0.014400         0.00000
0.069000
100003              0.804000         0.049700         0.08060
0.034500
100004              0.759637         0.042553         0.07449
0.145193
100006              0.759637         0.042553         0.07449
0.145193
100007              0.759637         0.042553         0.07449
0.145193

            FLOORSMAX_MODE  FLOORSMIN_MODE  LANDAREA_MODE  \
SK_ID_CURR
100002            0.083300        0.125000       0.037700
100003            0.291700        0.333300       0.012800
100004            0.222315        0.228058       0.064958
```

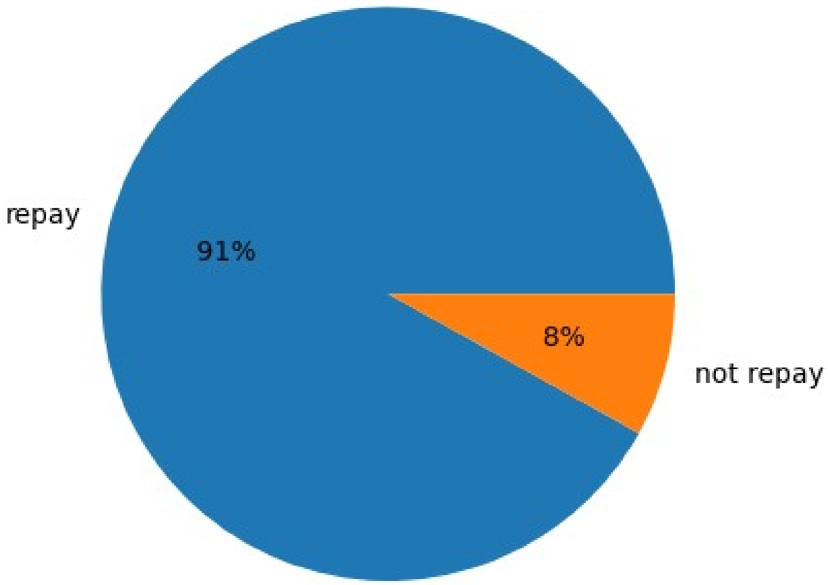
```
100006             0.222315        0.228058       0.064958
100007             0.222315        0.228058       0.064958

            LIVINGAPARTMENTS_MODE  LIVINGAREA_MODE
NONLIVINGAPARTMENTS_MODE  \
SK_ID_CURR

100002                      0.022000         0.019800
0.000000
100003                      0.079000         0.055400
0.000000
100004                      0.105645         0.105975
0.008076
100006                      0.105645         0.105975
0.008076
100007                      0.105645         0.105975
0.008076

            NONLIVINGAREA_MODE  APARTMENTS_MEDI  BASEMENTAREA_MEDI  \
SK_ID_CURR
100002                0.000000          0.02500           0.036900
100003                0.000000          0.09680           0.052900
100004                0.027022          0.11785           0.087955
100006                0.027022          0.11785           0.087955
100007                0.027022          0.11785           0.087955

            YEARS_BEGINEXPLUATATION_MEDI  YEARS_BUILD_MEDI
COMMONAREA_MEDI  \
SK_ID_CURR

100002                          0.972200          0.624300
0.014400
100003                          0.985100          0.798700
0.060800
100004                          0.977752          0.755746
0.044595
100006                          0.977752          0.755746
0.044595
100007                          0.977752          0.755746
0.044595

            ELEVATORS_MEDI  ENTRANCES_MEDI  FLOORSMAX_MEDI
FLOORSMIN_MEDI  \
SK_ID_CURR

100002             0.000000        0.069000        0.083300
0.125000
100003             0.080000        0.034500        0.291700
0.333300
100004             0.078078        0.149213        0.225897
```

```
0.231625
100006            0.078078        0.149213        0.225897
0.231625
100007            0.078078        0.149213        0.225897
0.231625

            LANDAREA_MEDI  LIVINGAPARTMENTS_MEDI  LIVINGAREA_MEDI  \
SK_ID_CURR
100002           0.037500               0.020500         0.019300
100003           0.013200               0.078700         0.055800
100004           0.067169               0.101954         0.108607
100006           0.067169               0.101954         0.108607
100007           0.067169               0.101954         0.108607

            NONLIVINGAPARTMENTS_MEDI  NONLIVINGAREA_MEDI
FONDKAPREMONT_MODE  \
SK_ID_CURR

100002                      0.000000            0.000000   reg oper
account
100003                      0.003900            0.010000   reg oper
account
100004                      0.008651            0.028236   reg oper
account
100006                      0.008651            0.028236   reg oper
account
100007                      0.008651            0.028236   reg oper
account

            HOUSETYPE_MODE  TOTALAREA_MODE WALLSMATERIAL_MODE  \
SK_ID_CURR
100002      block of flats        0.014900       Stone, brick
100003      block of flats        0.071400              Block
100004      block of flats        0.102547              Panel
100006      block of flats        0.102547              Panel
100007      block of flats        0.102547              Panel

           EMERGENCYSTATE_MODE  OBS_30_CNT_SOCIAL_CIRCLE  \
SK_ID_CURR
100002                      No                       2.0
100003                      No                       1.0
100004                      No                       0.0
100006                      No                       2.0
100007                      No                       0.0

            DEF_30_CNT_SOCIAL_CIRCLE  OBS_60_CNT_SOCIAL_CIRCLE  \
SK_ID_CURR
100002                           2.0                       2.0
100003                           0.0                       1.0
100004                           0.0                       0.0
```

```
100006                             0.0                         2.0
100007                             0.0                         0.0

            DEF_60_CNT_SOCIAL_CIRCLE  DAYS_LAST_PHONE_CHANGE
FLAG_DOCUMENT_2  \
SK_ID_CURR

100002                             2.0                   -1134.0
0
100003                             0.0                    -828.0
0
100004                             0.0                    -815.0
0
100006                             0.0                    -617.0
0
100007                             0.0                   -1106.0
0

            FLAG_DOCUMENT_3  FLAG_DOCUMENT_4  FLAG_DOCUMENT_5  \
SK_ID_CURR
100002                    1                0                0
100003                    1                0                0
100004                    0                0                0
100006                    1                0                0
100007                    0                0                0

            FLAG_DOCUMENT_6  FLAG_DOCUMENT_7  FLAG_DOCUMENT_8  \
SK_ID_CURR
100002                    0                0                0
100003                    0                0                0
100004                    0                0                0
100006                    0                0                0
100007                    0                0                1

            FLAG_DOCUMENT_9  FLAG_DOCUMENT_10  FLAG_DOCUMENT_11  \
SK_ID_CURR
100002                    0                 0                 0
100003                    0                 0                 0
100004                    0                 0                 0
100006                    0                 0                 0
100007                    0                 0                 0

            FLAG_DOCUMENT_12  FLAG_DOCUMENT_13  FLAG_DOCUMENT_14  \
SK_ID_CURR
100002                     0                 0                 0
100003                     0                 0                 0
100004                     0                 0                 0
100006                     0                 0                 0
100007                     0                 0                 0
```

```
            FLAG_DOCUMENT_15  FLAG_DOCUMENT_16  FLAG_DOCUMENT_17  \
SK_ID_CURR
100002                     0                 0                 0
100003                     0                 0                 0
100004                     0                 0                 0
100006                     0                 0                 0
100007                     0                 0                 0

            FLAG_DOCUMENT_18  FLAG_DOCUMENT_19  FLAG_DOCUMENT_20  \
SK_ID_CURR
100002                     0                 0                 0
100003                     0                 0                 0
100004                     0                 0                 0
100006                     0                 0                 0
100007                     0                 0                 0

            FLAG_DOCUMENT_21  AMT_REQ_CREDIT_BUREAU_HOUR  \
SK_ID_CURR
100002                     0                    0.000000
100003                     0                    0.000000
100004                     0                    0.000000
100006                     0                    0.006402
100007                     0                    0.000000

            AMT_REQ_CREDIT_BUREAU_DAY  AMT_REQ_CREDIT_BUREAU_WEEK  \
SK_ID_CURR
100002                          0.000                    0.000000
100003                          0.000                    0.000000
100004                          0.000                    0.000000
100006                          0.007                    0.034362
100007                          0.000                    0.000000

            AMT_REQ_CREDIT_BUREAU_MON  AMT_REQ_CREDIT_BUREAU_QRT  \
SK_ID_CURR
100002                       0.000000                   0.000000
100003                       0.000000                   0.000000
100004                       0.000000                   0.000000
100006                       0.267395                   0.265474
100007                       0.000000                   0.000000

            AMT_REQ_CREDIT_BUREAU_YEAR
SK_ID_CURR
100002                        1.000000
100003                        0.000000
100004                        0.000000
100006                        1.899974
100007                        0.000000

# Pie plot of target column
plt.pie(df['TARGET'].value_counts(),autopct="%3d%
```

```
%",labels=['repay','not repay'])
plt.show()
```



Applying Label Encoder

```
# Label encode categorical columns
label_encoder = LabelEncoder()
for col in categorical_columns:
    if col != 'TARGET':  # Exclude the target column
        df[col] = label_encoder.fit_transform(df[col])

df.head()
```

```
            TARGET  NAME_CONTRACT_TYPE  CODE_GENDER  FLAG_OWN_CAR  \
SK_ID_CURR
100002           1                   0            1             0
100003           0                   0            0             0
100004           0                   1            1             1
100006           0                   0            0             0
100007           0                   0            1             0

            FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL
AMT_CREDIT  \
SK_ID_CURR

100002                    1             0          202500.0
406597.5
```

```
100003                         0              0         270000.0
1293502.5
100004                         1              0          67500.0
135000.0
100006                         1              0         135000.0
312682.5
100007                         1              0         121500.0
513000.0

            AMT_ANNUITY  AMT_GOODS_PRICE  NAME_TYPE_SUITE
NAME_INCOME_TYPE  \
SK_ID_CURR

100002          24700.5         351000.0                6
7
100003          35698.5        1129500.0                1
4
100004           6750.0         135000.0                6
7
100006          29686.5         297000.0                6
7
100007          21865.5         513000.0                6
7

            NAME_EDUCATION_TYPE  NAME_FAMILY_STATUS  NAME_HOUSING_TYPE
\
SK_ID_CURR

100002                        4                   3                  1

100003                        1                   1                  1

100004                        4                   3                  1

100006                        4                   0                  1

100007                        4                   3                  1


            REGION_POPULATION_RELATIVE  DAYS_BIRTH  DAYS_EMPLOYED  \
SK_ID_CURR
100002                        0.018801       -9461           -637
100003                        0.003541      -16765          -1188
100004                        0.010032      -19046           -225
100006                        0.008019      -19005          -3039
100007                        0.028663      -19932          -3038

            DAYS_REGISTRATION  DAYS_ID_PUBLISH  OWN_CAR_AGE
FLAG_MOBIL  \
SK_ID_CURR
```

```
100002                -3648.0            -2120    12.061091
1
100003                -1186.0             -291    12.061091
1
100004                -4260.0            -2531    26.000000
1
100006                -9833.0            -2437    12.061091
1
100007                -4311.0            -3458    12.061091
1

            FLAG_EMP_PHONE  FLAG_WORK_PHONE  FLAG_CONT_MOBILE
FLAG_PHONE  \
SK_ID_CURR

100002                   1                0                 1
1
100003                   1                0                 1
1
100004                   1                1                 1
1
100006                   1                0                 1
0
100007                   1                0                 1
0

            FLAG_EMAIL  OCCUPATION_TYPE  CNT_FAM_MEMBERS  \
SK_ID_CURR
100002               0                8              1.0
100003               0                3              2.0
100004               0                8              1.0
100006               0                8              2.0
100007               0                3              1.0

            REGION_RATING_CLIENT  REGION_RATING_CLIENT_W_CITY  \
SK_ID_CURR
100002                         2                            2
100003                         1                            1
100004                         2                            2
100006                         2                            2
100007                         2                            2

            WEEKDAY_APPR_PROCESS_START  HOUR_APPR_PROCESS_START  \
SK_ID_CURR
100002                               6                       10
100003                               1                       11
100004                               1                        9
100006                               6                       17
100007                               4                       11
```

```
            REG_REGION_NOT_LIVE_REGION  REG_REGION_NOT_WORK_REGION  \
SK_ID_CURR
100002                               0                           0
100003                               0                           0
100004                               0                           0
100006                               0                           0
100007                               0                           0

            LIVE_REGION_NOT_WORK_REGION  REG_CITY_NOT_LIVE_CITY  \
SK_ID_CURR
100002                                0                       0
100003                                0                       0
100004                                0                       0
100006                                0                       0
100007                                0                       0

            REG_CITY_NOT_WORK_CITY  LIVE_CITY_NOT_WORK_CITY  \
SK_ID_CURR
100002                           0                        0
100003                           0                        0
100004                           0                        0
100006                           0                        0
100007                           1                        1

            ORGANIZATION_TYPE  EXT_SOURCE_1  EXT_SOURCE_2
EXT_SOURCE_3  \
SK_ID_CURR

100002                      5      0.083037      0.262949
0.139376
100003                     39      0.311267      0.622246
0.510853
100004                     11      0.502130      0.555912
0.729567
100006                      5      0.502130      0.650442
0.510853
100007                     37      0.502130      0.322738
0.510853

            APARTMENTS_AVG  BASEMENTAREA_AVG
YEARS_BEGINEXPLUATATION_AVG  \
SK_ID_CURR

100002             0.02470          0.036900
0.972200
100003             0.09590          0.052900
0.985100
100004             0.11744          0.088442
0.977735
```

```
100006             0.11744         0.088442
0.977735
100007             0.11744         0.088442
0.977735

            YEARS_BUILD_AVG  COMMONAREA_AVG  ELEVATORS_AVG
ENTRANCES_AVG  \
SK_ID_CURR

100002             0.619200        0.014300       0.000000
0.069000
100003             0.796000        0.060500       0.080000
0.034500
100004             0.752471        0.044621       0.078942
0.149725
100006             0.752471        0.044621       0.078942
0.149725
100007             0.752471        0.044621       0.078942
0.149725

            FLOORSMAX_AVG  FLOORSMIN_AVG  LANDAREA_AVG
LIVINGAPARTMENTS_AVG  \
SK_ID_CURR

100002           0.083300       0.125000      0.036900
0.020200
100003           0.291700       0.333300      0.013000
0.077300
100004           0.226282       0.231894      0.066333
0.100775
100006           0.226282       0.231894      0.066333
0.100775
100007           0.226282       0.231894      0.066333
0.100775

            LIVINGAREA_AVG  NONLIVINGAPARTMENTS_AVG  NONLIVINGAREA_AVG
\
SK_ID_CURR

100002            0.019000                 0.000000           0.000000

100003            0.054900                 0.003900           0.009800

100004            0.107399                 0.008809           0.028358

100006            0.107399                 0.008809           0.028358

100007            0.107399                 0.008809           0.028358
```

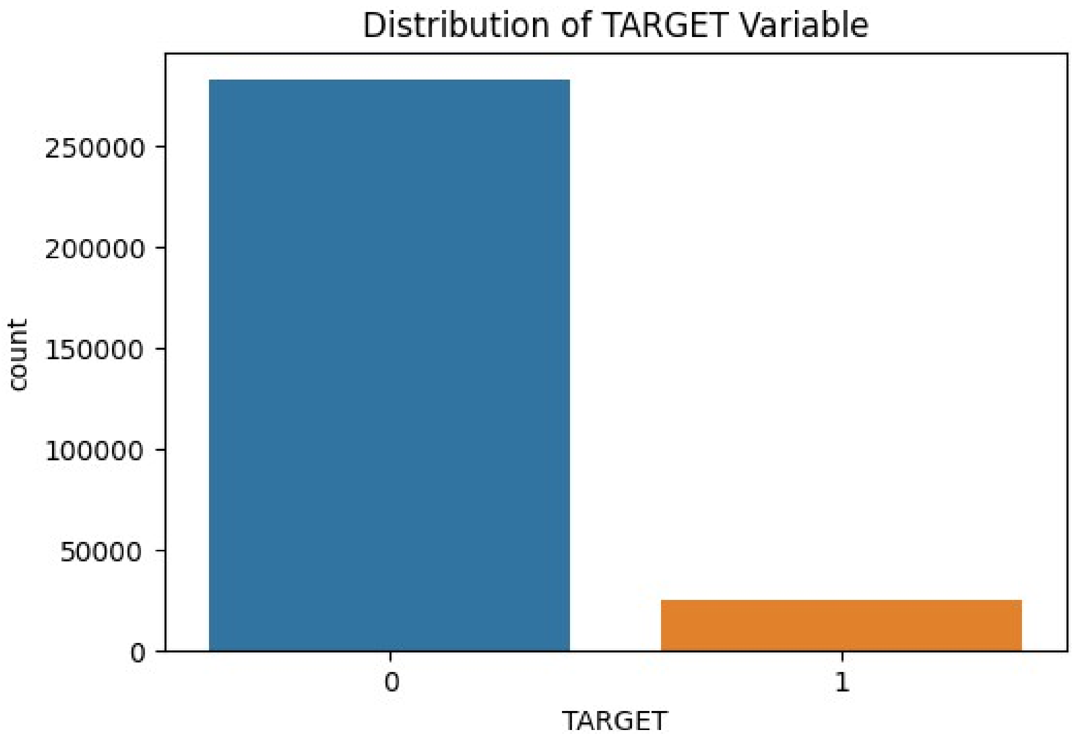
```
            APARTMENTS_MODE  BASEMENTAREA_MODE
YEARS_BEGINEXPLUATATION_MODE  \
SK_ID_CURR

100002                 0.025200           0.038300
0.972200
100003                 0.092400           0.053800
0.985100
100004                 0.114231           0.087543
0.977065
100006                 0.114231           0.087543
0.977065
100007                 0.114231           0.087543
0.977065

            YEARS_BUILD_MODE  COMMONAREA_MODE  ELEVATORS_MODE
ENTRANCES_MODE  \
SK_ID_CURR

100002                  0.634100         0.014400         0.00000
0.069000
100003                  0.804000         0.049700         0.08060
0.034500
100004                  0.759637         0.042553         0.07449
0.145193
100006                  0.759637         0.042553         0.07449
0.145193
100007                  0.759637         0.042553         0.07449
0.145193

            FLOORSMAX_MODE  FLOORSMIN_MODE  LANDAREA_MODE  \
SK_ID_CURR
100002               0.083300        0.125000       0.037700
100003               0.291700        0.333300       0.012800
100004               0.222315        0.228058       0.064958
100006               0.222315        0.228058       0.064958
100007               0.222315        0.228058       0.064958

            LIVINGAPARTMENTS_MODE  LIVINGAREA_MODE
NONLIVINGAPARTMENTS_MODE  \
SK_ID_CURR

100002                        0.022000         0.019800
0.000000
100003                        0.079000         0.055400
0.000000
100004                        0.105645         0.105975
0.008076
100006                        0.105645         0.105975
0.008076
```

```
100007                   0.105645        0.105975
0.008076

            NONLIVINGAREA_MODE  APARTMENTS_MEDI  BASEMENTAREA_MEDI  \
SK_ID_CURR
100002                0.000000          0.02500           0.036900
100003                0.000000          0.09680           0.052900
100004                0.027022          0.11785           0.087955
100006                0.027022          0.11785           0.087955
100007                0.027022          0.11785           0.087955

            YEARS_BEGINEXPLUATATION_MEDI  YEARS_BUILD_MEDI
COMMONAREA_MEDI  \
SK_ID_CURR

100002                          0.972200          0.624300
0.014400
100003                          0.985100          0.798700
0.060800
100004                          0.977752          0.755746
0.044595
100006                          0.977752          0.755746
0.044595
100007                          0.977752          0.755746
0.044595

            ELEVATORS_MEDI  ENTRANCES_MEDI  FLOORSMAX_MEDI
FLOORSMIN_MEDI  \
SK_ID_CURR

100002            0.000000        0.069000        0.083300
0.125000
100003            0.080000        0.034500        0.291700
0.333300
100004            0.078078        0.149213        0.225897
0.231625
100006            0.078078        0.149213        0.225897
0.231625
100007            0.078078        0.149213        0.225897
0.231625

            LANDAREA_MEDI  LIVINGAPARTMENTS_MEDI  LIVINGAREA_MEDI  \
SK_ID_CURR
100002           0.037500               0.020500         0.019300
100003           0.013200               0.078700         0.055800
100004           0.067169               0.101954         0.108607
100006           0.067169               0.101954         0.108607
100007           0.067169               0.101954         0.108607

            NONLIVINGAPARTMENTS_MEDI  NONLIVINGAREA_MEDI
```

```
FONDKAPREMONT_MODE  \
SK_ID_CURR

100002                            0.000000            0.000000
2
100003                            0.003900            0.010000
2
100004                            0.008651            0.028236
2
100006                            0.008651            0.028236
2
100007                            0.008651            0.028236
2

            HOUSETYPE_MODE  TOTALAREA_MODE  WALLSMATERIAL_MODE  \
SK_ID_CURR
100002                   0        0.014900                   5
100003                   0        0.071400                   0
100004                   0        0.102547                   4
100006                   0        0.102547                   4
100007                   0        0.102547                   4

            EMERGENCYSTATE_MODE  OBS_30_CNT_SOCIAL_CIRCLE  \
SK_ID_CURR
100002                        0                       2.0
100003                        0                       1.0
100004                        0                       0.0
100006                        0                       2.0
100007                        0                       0.0

            DEF_30_CNT_SOCIAL_CIRCLE  OBS_60_CNT_SOCIAL_CIRCLE  \
SK_ID_CURR
100002                           2.0                       2.0
100003                           0.0                       1.0
100004                           0.0                       0.0
100006                           0.0                       2.0
100007                           0.0                       0.0

            DEF_60_CNT_SOCIAL_CIRCLE  DAYS_LAST_PHONE_CHANGE
FLAG_DOCUMENT_2  \
SK_ID_CURR

100002                           2.0                 -1134.0
0
100003                           0.0                  -828.0
0
100004                           0.0                  -815.0
0
100006                           0.0                  -617.0
0
```

```
100007                                 0.0                   -1106.0
0

                FLAG_DOCUMENT_3  FLAG_DOCUMENT_4  FLAG_DOCUMENT_5  \
SK_ID_CURR
100002                        1                0                0
100003                        1                0                0
100004                        0                0                0
100006                        1                0                0
100007                        0                0                0

                FLAG_DOCUMENT_6  FLAG_DOCUMENT_7  FLAG_DOCUMENT_8  \
SK_ID_CURR
100002                        0                0                0
100003                        0                0                0
100004                        0                0                0
100006                        0                0                0
100007                        0                0                1

                FLAG_DOCUMENT_9  FLAG_DOCUMENT_10  FLAG_DOCUMENT_11  \
SK_ID_CURR
100002                        0                 0                 0
100003                        0                 0                 0
100004                        0                 0                 0
100006                        0                 0                 0
100007                        0                 0                 0

                FLAG_DOCUMENT_12  FLAG_DOCUMENT_13  FLAG_DOCUMENT_14  \
SK_ID_CURR
100002                         0                 0                 0
100003                         0                 0                 0
100004                         0                 0                 0
100006                         0                 0                 0
100007                         0                 0                 0

                FLAG_DOCUMENT_15  FLAG_DOCUMENT_16  FLAG_DOCUMENT_17  \
SK_ID_CURR
100002                         0                 0                 0
100003                         0                 0                 0
100004                         0                 0                 0
100006                         0                 0                 0
100007                         0                 0                 0

                FLAG_DOCUMENT_18  FLAG_DOCUMENT_19  FLAG_DOCUMENT_20  \
SK_ID_CURR
100002                         0                 0                 0
100003                         0                 0                 0
100004                         0                 0                 0
100006                         0                 0                 0
100007                         0                 0                 0
```

```
                FLAG_DOCUMENT_21  AMT_REQ_CREDIT_BUREAU_HOUR  \
SK_ID_CURR
100002                         0                    0.000000
100003                         0                    0.000000
100004                         0                    0.000000
100006                         0                    0.006402
100007                         0                    0.000000

                AMT_REQ_CREDIT_BUREAU_DAY  AMT_REQ_CREDIT_BUREAU_WEEK  \
SK_ID_CURR
100002                              0.000                    0.000000
100003                              0.000                    0.000000
100004                              0.000                    0.000000
100006                              0.007                    0.034362
100007                              0.000                    0.000000

                AMT_REQ_CREDIT_BUREAU_MON  AMT_REQ_CREDIT_BUREAU_QRT  \
SK_ID_CURR
100002                           0.000000                   0.000000
100003                           0.000000                   0.000000
100004                           0.000000                   0.000000
100006                           0.267395                   0.265474
100007                           0.000000                   0.000000

                AMT_REQ_CREDIT_BUREAU_YEAR
SK_ID_CURR
100002                            1.000000
100003                            0.000000
100004                            0.000000
100006                            1.899974
100007                            0.000000

# Histogram of 'TARGET' variable
import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(6, 4))
sns.countplot(data=df, x='TARGET')
plt.title('Distribution of TARGET Variable')
plt.show()
```



Upon analyzing the target variable, it's evident that there's a considerable imbalance between the classes. The countplot above illustrates a significant discrepancy between the frequencies of the classes within the dataset. This highly imbalanced distribution might pose challenges during model training and evaluation, particularly affecting the minority class's predictive performance.

To address this issue, we will be applying sampling techniques, specifically random undersampling, to create a more balanced representation of both classes in the dataset. This approach will ensure that the model's training isn't biased towards the majority class, thus improving its ability to learn from both classes equally.

Balancing the dataset

```
from imblearn.under_sampling import RandomUnderSampler
undersampler = RandomUnderSampler(random_state=0)

# Separate input and output variables.
X = df.drop('TARGET', axis=1)
y = df['TARGET']

# Split data into train and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

print(X_train.shape)
print(X_test.shape)

(246008, 120)
(61503, 120)
```

```
X_train_bal, y_train_bal = undersampler.fit_resample(X_train, y_train)

# Shape of resampled data.
print(X_train_bal.shape)
print(y_train_bal.shape)

(39752, 120)
(39752,)

# Scale the data
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train_bal)
X_test_scaled = scaler.transform(X_test)
```

Model Building

```
# Define and compile the model
model = Sequential()
model.add(Dense(units=128, activation='relu',
input_shape=(X_train.shape[1],)))
model.add(Dropout(0.2))
model.add(Dense(units=64, activation='relu'))
model.add(Dropout(0.2))
model.add(Dense(units=1, activation='sigmoid'))

model.summary()

Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 128)               15488

 dropout (Dropout)           (None, 128)               0

 dense_1 (Dense)             (None, 64)                8256

 dropout_1 (Dropout)         (None, 64)                0

 dense_2 (Dense)             (None, 1)                 65

=================================================================
Total params: 23809 (93.00 KB)
Trainable params: 23809 (93.00 KB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________

model.compile(optimizer='adam', loss='binary_crossentropy',
metrics=['accuracy'])
```

```
from tensorflow.keras.callbacks import EarlyStopping

# Train the model
history = model.fit(X_train_scaled, y_train_bal, epochs=100,
verbose=1, batch_size=32, validation_split=0.2,
                    callbacks=[EarlyStopping(patience=10)])

Epoch 1/100
994/994 [==============================] - 6s 5ms/step - loss: 0.6017
- accuracy: 0.6841 - val_loss: 0.8313 - val_accuracy: 0.4332
Epoch 2/100
994/994 [==============================] - 3s 3ms/step - loss: 0.5803
- accuracy: 0.6982 - val_loss: 0.8622 - val_accuracy: 0.4431
Epoch 3/100
994/994 [==============================] - 4s 4ms/step - loss: 0.5749
- accuracy: 0.7021 - val_loss: 0.8264 - val_accuracy: 0.4630
Epoch 4/100
994/994 [==============================] - 3s 3ms/step - loss: 0.5708
- accuracy: 0.7043 - val_loss: 0.9204 - val_accuracy: 0.4367
Epoch 5/100
994/994 [==============================] - 4s 4ms/step - loss: 0.5657
- accuracy: 0.7079 - val_loss: 0.8485 - val_accuracy: 0.4320
Epoch 6/100
994/994 [==============================] - 3s 3ms/step - loss: 0.5644
- accuracy: 0.7077 - val_loss: 0.8267 - val_accuracy: 0.4335
Epoch 7/100
994/994 [==============================] - 3s 3ms/step - loss: 0.5613
- accuracy: 0.7119 - val_loss: 0.8655 - val_accuracy: 0.4186
Epoch 8/100
994/994 [==============================] - 3s 3ms/step - loss: 0.5579
- accuracy: 0.7135 - val_loss: 0.8823 - val_accuracy: 0.4172
Epoch 9/100
994/994 [==============================] - 4s 4ms/step - loss: 0.5539
- accuracy: 0.7173 - val_loss: 0.8566 - val_accuracy: 0.4353
Epoch 10/100
994/994 [==============================] - 3s 3ms/step - loss: 0.5519
- accuracy: 0.7166 - val_loss: 0.8402 - val_accuracy: 0.4032
Epoch 11/100
994/994 [==============================] - 3s 3ms/step - loss: 0.5488
- accuracy: 0.7175 - val_loss: 0.8580 - val_accuracy: 0.4187
Epoch 12/100
994/994 [==============================] - 3s 3ms/step - loss: 0.5459
- accuracy: 0.7203 - val_loss: 0.8860 - val_accuracy: 0.4134
Epoch 13/100
994/994 [==============================] - 4s 4ms/step - loss: 0.5437
- accuracy: 0.7237 - val_loss: 0.8383 - val_accuracy: 0.4572

# Evaluate the model
y_pred_prob_final = model.predict(X_test_scaled)
y_pred_final = (y_pred_prob_final > 0.5).astype(int)
```

```python
# Calculate evaluation metrics for the final model
accuracy_final = accuracy_score(y_test, y_pred_final)
precision_final = precision_score(y_test, y_pred_final)
recall_final = recall_score(y_test, y_pred_final)
f1_final = f1_score(y_test, y_pred_final)
roc_auc_final = roc_auc_score(y_test, y_pred_prob_final)
```

```
1922/1922 [==============================] - 5s 2ms/step
```

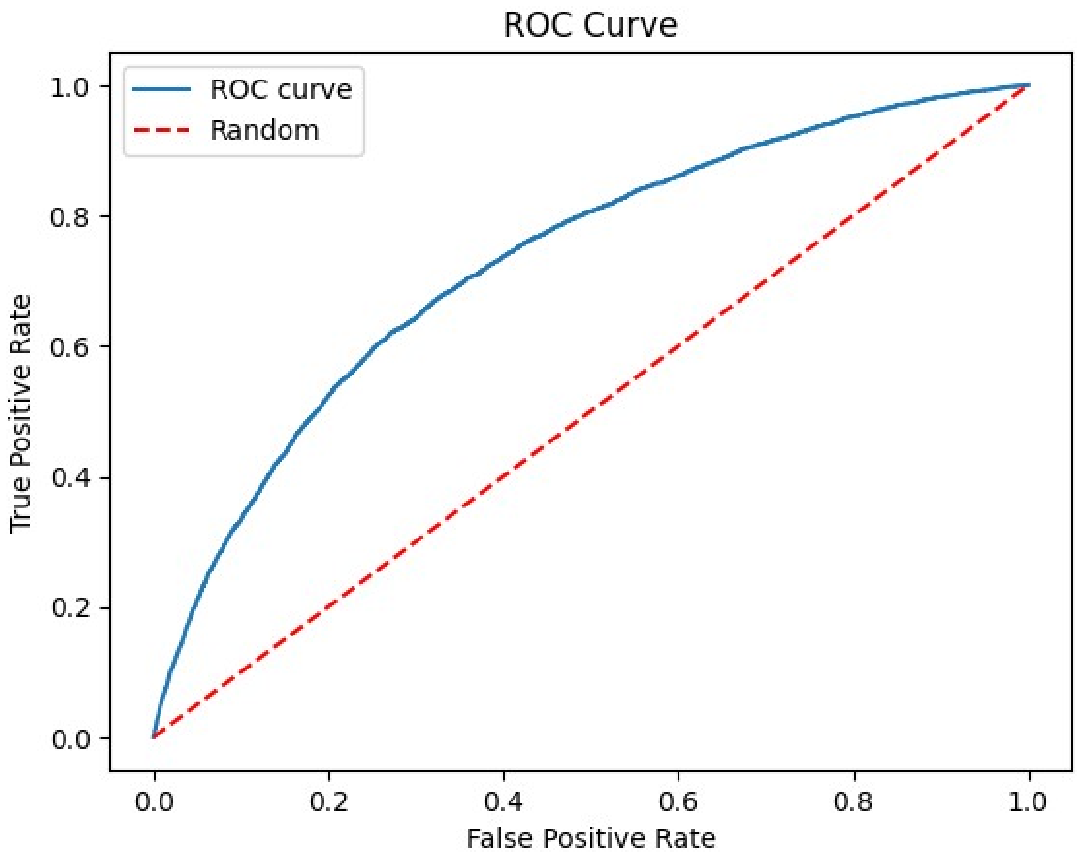
```python
# Display evaluation metrics for the final model
print(f"Final Model Evaluation:")
print(f"Accuracy: {accuracy_final:.4f}")
print(f"Precision: {precision_final:.4f}")
print(f"Recall: {recall_final:.4f}")
print(f"F1-score: {f1_final:.4f}")
print(f"ROC-AUC: {roc_auc_final:.4f}")
```

```
Final Model Evaluation:
Accuracy: 0.8090
Precision: 0.1991
Recall: 0.4544
F1-score: 0.2769
ROC-AUC: 0.7301
```

```python
from sklearn.metrics import roc_curve, confusion_matrix

# Predict probabilities for the test set
y_pred_prob_final = model.predict(X_test_scaled)
```

```
1922/1922 [==============================] - 4s 2ms/step
```

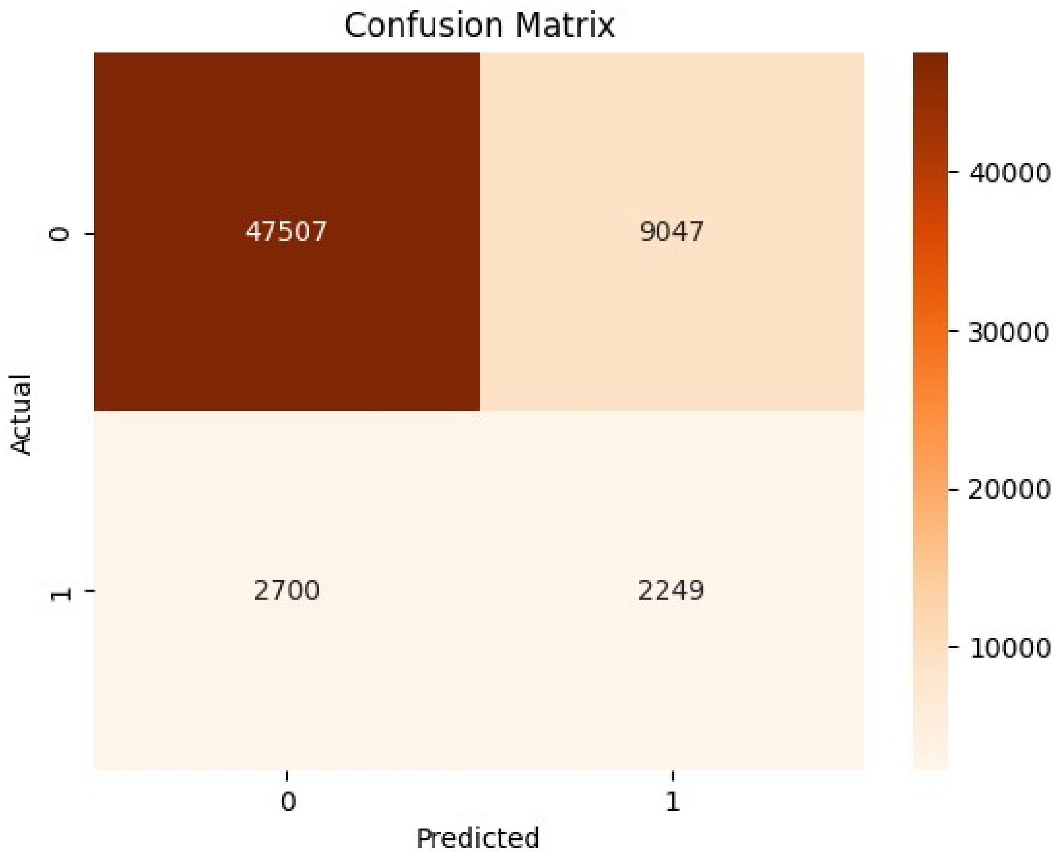
```python
# Convert probabilities to binary predictions
y_pred_final = (y_pred_prob_final > 0.5).astype(int)

# Plot ROC curve
fpr, tpr, thresholds = roc_curve(y_test, y_pred_prob_final)
plt.plot(fpr, tpr, label='ROC curve')
plt.plot([0, 1], [0, 1], linestyle='--', color='red', label='Random')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curve')
plt.legend()
plt.show()
```




```
from sklearn.metrics import roc_auc_score

# Calculate AUC-ROC
roc_auc = roc_auc_score(y_test, y_pred_prob_final)
print(f"AUC-ROC: {roc_auc:.4f}")
```

```
AUC-ROC: 0.7301
```

```
# Plot confusion matrix
cm = confusion_matrix(y_test, y_pred_final)
sns.heatmap(cm, annot=True, fmt='d', cmap='Oranges')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
```



Hyperparameter tuning and model refinement

```
# Perform further error analysis, hyperparameter tuning, and model
refinement
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout

def build_model():
    model = Sequential()
    model.add(Dense(units=128, activation='relu',
input_shape=(X_train.shape[1],)))
    model.add(Dropout(0.2))
    model.add(Dense(units=64, activation='relu'))
    model.add(Dropout(0.2))
    model.add(Dense(units=1, activation='sigmoid'))
    model.compile(optimizer='adam', loss='binary_crossentropy',
metrics=['accuracy'])
    return model


from sklearn.metrics import accuracy_score, precision_score,
recall_score, roc_auc_score
```

```
# Initialize best score and parameters for multiple metrics
best_score = {
    'accuracy': 0,
    'precision': 0,
    'recall': 0,
    'roc_auc': 0
}
best_params = {'epochs': None, 'batch_size': None}

# Define the parameter grid
params = {'epochs': [10, 20, 30], 'batch_size': [32, 64]}

from sklearn.model_selection import StratifiedKFold
# Initialize StratifiedKFold
skf = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)


# Iterate through different combinations of epochs and batch sizes
for epoch in params['epochs']:
    for batch_size in params['batch_size']:
        cv_metrics = {
            'accuracy': [],
            'precision': [],
            'recall': [],
            'roc_auc': []
        }

        # Perform cross-validation
        for train_index, val_index in skf.split(X_train_scaled,
y_train_bal):
            X_train_cv, X_val = X_train_scaled[train_index],
X_train_scaled[val_index]
            y_train_cv, y_val = y_train_bal[train_index],
y_train_bal[val_index]

            # Build and train the model
            model = build_model()
            history = model.fit(X_train_cv, y_train_cv, epochs=epoch,
batch_size=batch_size,
                                verbose=0, validation_data=(X_val,
y_val), callbacks=[EarlyStopping(patience=5)])

            # Get predictions on validation set
            y_pred = (model.predict(X_val) > 0.5).astype(int)

            # Calculate metrics
            cv_metrics['accuracy'].append(accuracy_score(y_val,
y_pred))
            cv_metrics['precision'].append(precision_score(y_val,
y_pred))
```

```
            cv_metrics['recall'].append(recall_score(y_val, y_pred))
            cv_metrics['roc_auc'].append(roc_auc_score(y_val, y_pred))

        # Calculate average metrics
        avg_metrics = {metric: np.mean(scores) for metric, scores in
cv_metrics.items()}

        # Check if this combination of parameters gives better scores
for all metrics
        improvement = all(avg_metrics[metric] > best_score[metric] for
metric in best_score)
        if improvement:
            best_score = avg_metrics
            best_params['epochs'] = epoch
            best_params['batch_size'] = batch_size

print(f"Best parameters: {best_params}")
print(f"Best scores: {best_score}")
```

```
249/249 [==============================] - 0s 2ms/step
249/249 [==============================] - 1s 2ms/step
249/249 [==============================] - 0s 2ms/step
249/249 [==============================] - 1s 2ms/step
249/249 [==============================] - 0s 2ms/step
249/249 [==============================] - 1s 2ms/step
249/249 [==============================] - 0s 2ms/step
249/249 [==============================] - 0s 2ms/step
249/249 [==============================] - 0s 2ms/step
249/249 [==============================] - 1s 2ms/step
249/249 [==============================] - 1s 2ms/step
249/249 [==============================] - 1s 2ms/step
249/249 [==============================] - 1s 2ms/step
249/249 [==============================] - 1s 2ms/step
249/249 [==============================] - 1s 2ms/step
249/249 [==============================] - 0s 2ms/step
249/249 [==============================] - 0s 2ms/step
249/249 [==============================] - 0s 2ms/step
249/249 [==============================] - 1s 2ms/step
249/249 [==============================] - 1s 2ms/step
249/249 [==============================] - 1s 2ms/step
249/249 [==============================] - 1s 2ms/step
249/249 [==============================] - 1s 2ms/step
249/249 [==============================] - 0s 2ms/step
249/249 [==============================] - 1s 2ms/step
249/249 [==============================] - 0s 2ms/step
249/249 [==============================] - 0s 2ms/step
249/249 [==============================] - 0s 2ms/step
249/249 [==============================] - 1s 2ms/step
249/249 [==============================] - 0s 2ms/step
Best parameters: {'epochs': 10, 'batch_size': 32}
```

```
Best scores: {'accuracy': 0.6768718241999543, 'precision':
0.6780189503906972, 'recall': 0.674029333232097, 'roc_auc':
0.6768719233640839}
```

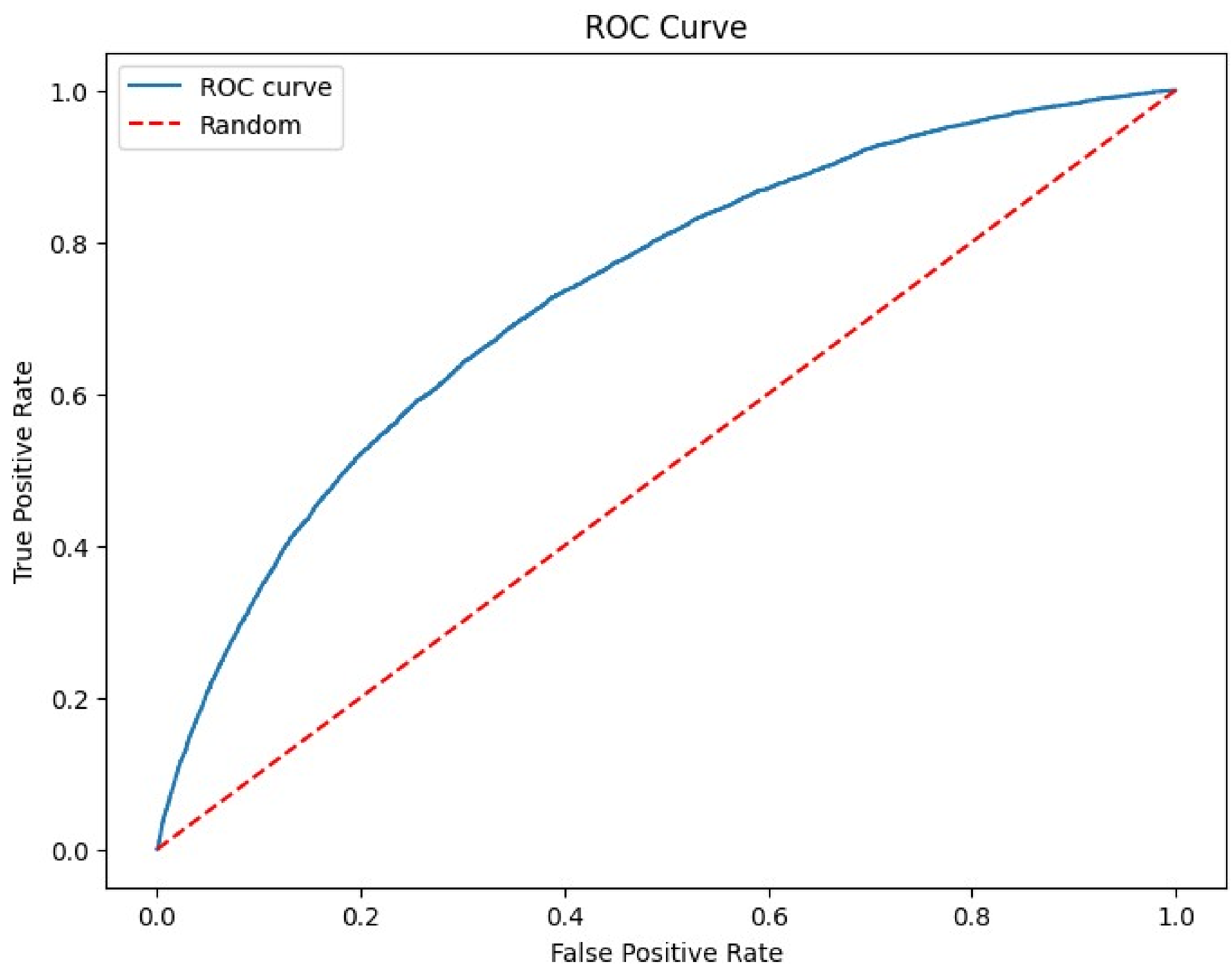
```python
# Predict probabilities for test set
y_pred_prob_final_mod = model.predict(X_test_scaled)
```

```
1922/1922 [==============================] - 3s 2ms/step
```

```python
# Convert probabilities to binary predictions
y_pred_final_mod = (y_pred_prob_final_mod > 0.5).astype(int)


# Calculate evaluation metrics for the final model
accuracy_final_mod = accuracy_score(y_test, y_pred_final_mod)
precision_final_mod = precision_score(y_test, y_pred_final_mod)
recall_final_mod = recall_score(y_test, y_pred_final_mod)
f1_final_mod = f1_score(y_test, y_pred_final_mod)
roc_auc_final_mod = roc_auc_score(y_test, y_pred_prob_final_mod)

# Display evaluation metrics for the final model
print(f"Final Model Evaluation:")
print(f"Accuracy: {accuracy_final_mod:.4f}")
print(f"Precision: {precision_final_mod:.4f}")
print(f"Recall: {recall_final_mod:.4f}")
print(f"F1-score: {f1_final_mod:.4f}")
print(f"ROC-AUC: {roc_auc_final_mod:.4f}")
```

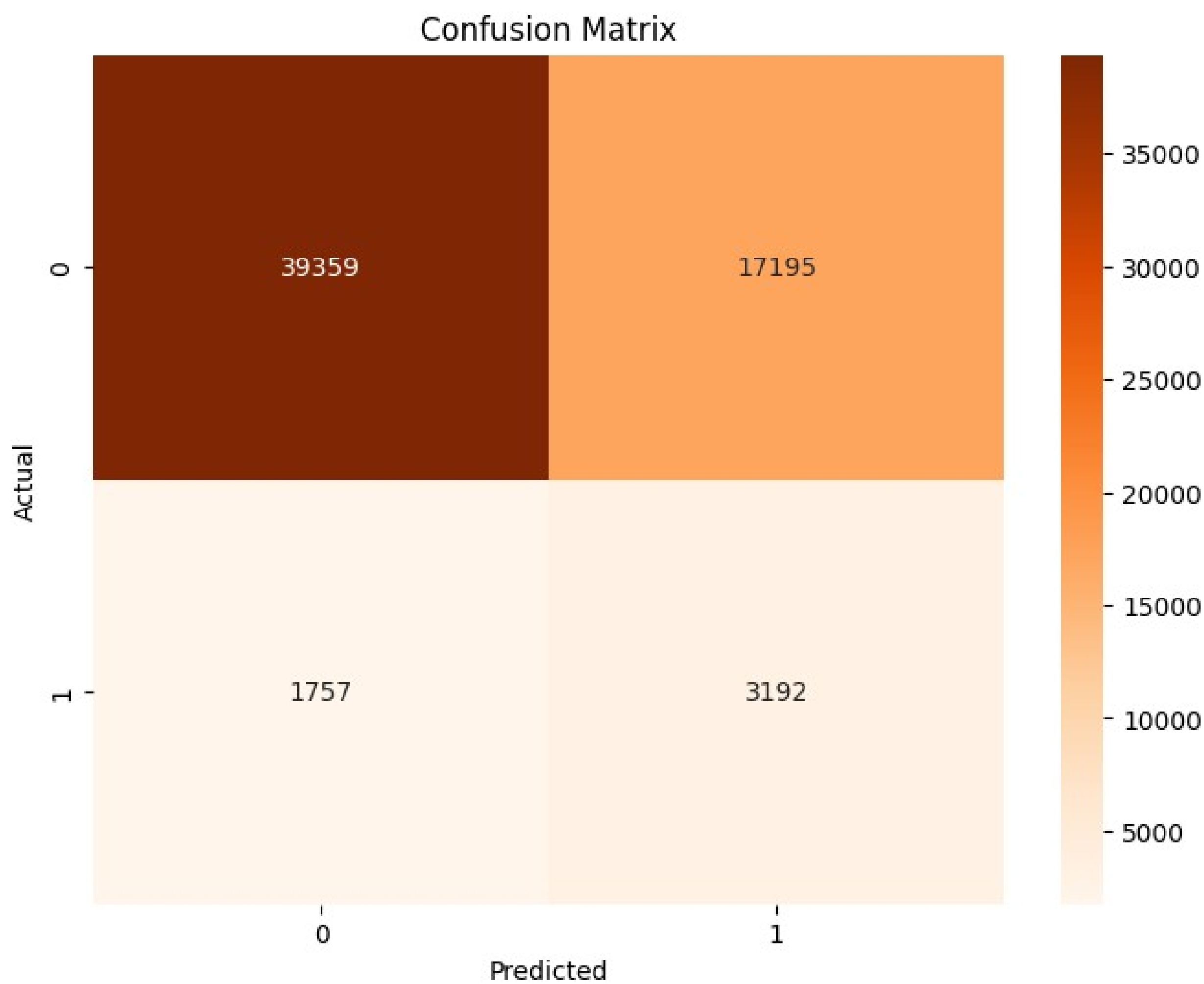
```
Final Model Evaluation:
Accuracy: 0.6919
Precision: 0.1566
Recall: 0.6450
F1-score: 0.2520
ROC-AUC: 0.7320
```

```python
# Calculate ROC curve
fpr, tpr, thresholds = roc_curve(y_test, y_pred_prob_final_mod)

# Plot ROC curve
plt.figure(figsize=(8, 6))
plt.plot(fpr, tpr, label='ROC curve')
plt.plot([0, 1], [0, 1], linestyle='--', color='red', label='Random')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC Curve')
plt.legend()
plt.show()
```




```
from sklearn.metrics import classification_report
# Calculate classification report for the final model
classification_rep = classification_report(y_test, y_pred_final_mod)
print("Classification Report for Final Model:")
print(classification_rep)

Classification Report for Final Model:
              precision    recall  f1-score   support

           0       0.96      0.70      0.81     56554
           1       0.16      0.64      0.25      4949

    accuracy                           0.69     61503
   macro avg       0.56      0.67      0.53     61503
weighted avg       0.89      0.69      0.76     61503


# Generate Confusion Matrix
cm = confusion_matrix(y_test, y_pred_final_mod)
```

```
# Plot Confusion Matrix
plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt='d', cmap='Oranges')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
```




```
# Calculate AUC-ROC
roc_auc_mod = roc_auc_score(y_test, y_pred_prob_final_mod)
print(f"AUC-ROC: {roc_auc_mod:.4f}")

AUC-ROC: 0.7320
```

Conclusion: The ROC curve graphically represents the model's ability to differentiate between defaulters and non-defaulters across various threshold values. The AUC-ROC score improved marginally from 0.7301 to 0.7320 post hyperparameter tuning, reflecting a slight enhancement in the model's capacity to distinguish between loan default and non-default instances. This improvement signifies refined predictive capabilities, crucial for robust risk assessment in lending scenarios.