# home_loan

June 1, 2023

## 0.1 House Loan Data Analysis

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

[1]:
```
#importing libraries

import pandas as pd
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
%matplotlib inline

import warnings
warnings.filterwarnings("ignore")

# to visualise all the columns in the dataframe
pd.pandas.set_option('display.max_columns', None)
```

[2]:
```
#loading the data
df=pd.read_csv('loan_data (1).csv')
```

[3]:
```
#checking first five records
df.head()
```

[3]:

| | 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
```

[4]:
```
#checking of the dataset
df.shape
```

[4]:
```
(307511, 122)
```

[5]:
```
#The dataset contains 307,511 rows and 122 columns
```

[6]:
```
#checking datatypeof the dataset
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
```

[7]:
```
#Let's generate Descriptive Statistics
df.describe()
```

[7]:
```
          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
```

[8]:
```
#checking null values in data
df.isnull().sum()
```

[8]:
```
SK_ID_CURR                    0
TARGET                        0
NAME_CONTRACT_TYPE            0
CODE_GENDER                   0
FLAG_OWN_CAR                  0
                             ..
```

```
AMT_REQ_CREDIT_BUREAU_DAY     41519
AMT_REQ_CREDIT_BUREAU_WEEK    41519
AMT_REQ_CREDIT_BUREAU_MON     41519
AMT_REQ_CREDIT_BUREAU_QRT     41519
AMT_REQ_CREDIT_BUREAU_YEAR    41519
Length: 122, dtype: int64
```

[9]:
```
def missing_values_table(df):
        # Total missing values
        mis_val = df.isnull().sum()

        # Percentage of missing values
        mis_val_percent = 100 * df.isnull().sum() / len(df)

        # Make a table with the results
        mis_val_table = pd.concat([mis_val, mis_val_percent], axis=1)

        # Rename the columns
        mis_val_table_ren_columns = mis_val_table.rename(
        columns = {0 : 'Missing Values', 1 : '% of Total Values'})

        # Sort the table by percentage of missing descending
        mis_val_table_ren_columns = mis_val_table_ren_columns[
            mis_val_table_ren_columns.iloc[:,1] != 0].sort_values(
        '% of Total Values', ascending=False).round(1)

        # Print some summary information
        print ("Your selected dataframe has " + str(df.shape[1]) + " columns.\n"
            "There are " + str(mis_val_table_ren_columns.shape[0]) +
              " columns that have missing values.")


        # Return the dataframe with missing information
        return mis_val_table_ren_columns
```

[10]:
```
missing_values_table(df).head(10)
```

```
Your selected dataframe has 122 columns.
There are 67 columns that have missing values.
```

[10]:

| | Missing Values | % of Total Values |
|---|---|---|
| COMMONAREA_MEDI | 214865 | 69.9 |
| COMMONAREA_AVG | 214865 | 69.9 |
| COMMONAREA_MODE | 214865 | 69.9 |
| NONLIVINGAPARTMENTS_MEDI | 213514 | 69.4 |
| NONLIVINGAPARTMENTS_MODE | 213514 | 69.4 |
| NONLIVINGAPARTMENTS_AVG | 213514 | 69.4 |

```
FONDKAPREMONT_MODE          210295          68.4
LIVINGAPARTMENTS_MODE       210199          68.4
LIVINGAPARTMENTS_MEDI       210199          68.4
LIVINGAPARTMENTS_AVG        210199          68.4
```

[11]:
```
#its evident that our dataset has huge missing values, so we are dropping␣
↪columns which has missing values above 60%
```

[12]:
```
def missing_preprocess_data(df):

    # Identify columns with more than 60% missing values
    missing_cols = df.columns[df.isnull().mean() > 0.6]

    # Drop columns with more than 60% missing values
    num_cols_dropped = len(missing_cols)
    df.drop(columns=missing_cols, inplace=True)
    print(f'Dropped {num_cols_dropped} columns due to missing value threshold')

    # Print out the original and preprocessed datasets
    #print('Preprocessed dataset:')
    #print(df)
```

[13]:
```
missing_preprocess_data(df)
```

```
Dropped 17 columns due to missing value threshold
```

[14]:
```
#check data imbalance
df["TARGET"].value_counts()
```

[14]:
```
0    282686
1     24825
Name: TARGET, dtype: int64
```

[15]:
```
#The TARGET is the binary variable that we are trying to predict with 2 values:
#0: Loan was repaid
#1: Loan was not repaid
```

[16]:
```
# Percentage calculation
print("Percentage: ")
(df["TARGET"].value_counts()/df["TARGET"].count())*100
```

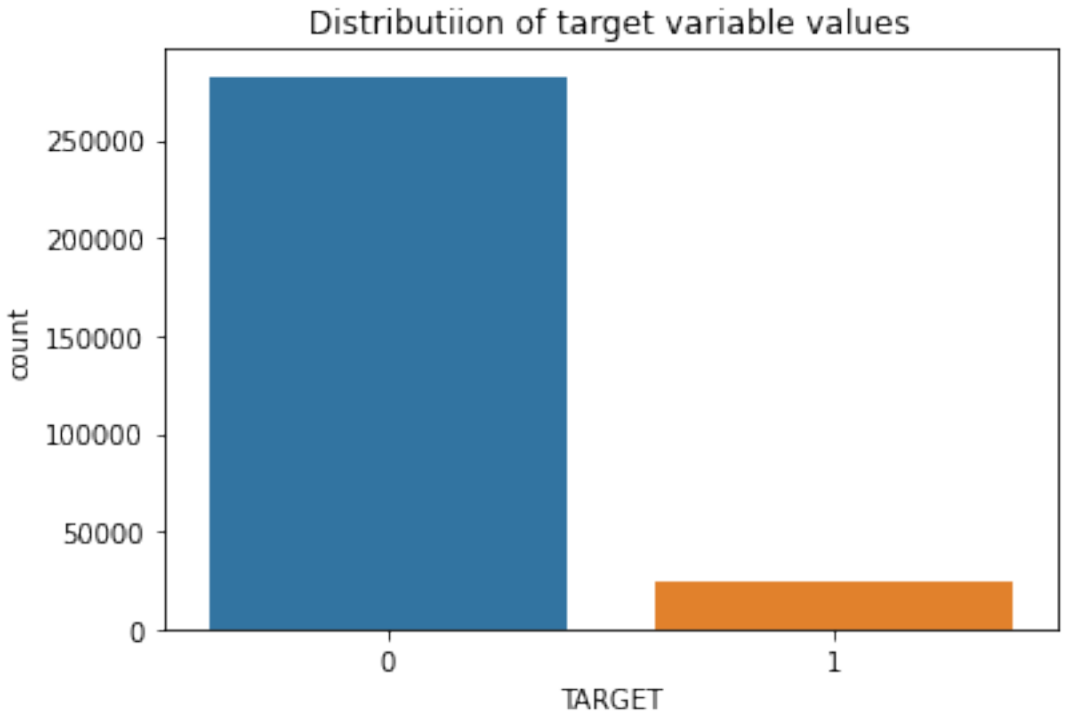
```
Percentage:
```

[16]:
```
0    91.927118
1     8.072882
```

```
Name: TARGET, dtype: float64
```

[17]:
```
sns.countplot(data=df,x='TARGET')
plt.title('Distributiion of target variable values')
```

[17]:
```
Text(0.5, 1.0, 'Distributiion of target variable values')
```




[18]:
```
#observatin:As evident from the distribution above, there is an imbalance␣
 ↪distribution of TARGET variable
#that leads to an imbalanced class problem
```

[19]:
```
#splitting data into categorical and numerical for encoding

char=df.select_dtypes(exclude='number')
numeric=df.select_dtypes(include='number')
```

[20]:
```
char.head()
```

[20]:

| | NAME_CONTRACT_TYPE | CODE_GENDER | FLAG_OWN_CAR | FLAG_OWN_REALTY | NAME_TYPE_SUITE | \ |
|---|---|---|---|---|---|---|
| 0 | Cash loans | M | N | Y | Unaccompanied | |
| 1 | Cash loans | F | N | N | Family | |
| 2 | Revolving loans | M | Y | Y | Unaccompanied | |
| 3 | Cash loans | F | N | Y | Unaccompanied | |

```
4         Cash loans           M            N               Y   Unaccompanied

  NAME_INCOME_TYPE            NAME_EDUCATION_TYPE    NAME_FAMILY_STATUS  \
0          Working  Secondary / secondary special  Single / not married
1    State servant               Higher education             Married
2          Working  Secondary / secondary special  Single / not married
3          Working  Secondary / secondary special       Civil marriage
4          Working  Secondary / secondary special  Single / not married

   NAME_HOUSING_TYPE OCCUPATION_TYPE WEEKDAY_APPR_PROCESS_START  \
0  House / apartment       Laborers                  WEDNESDAY
1  House / apartment     Core staff                     MONDAY
2  House / apartment       Laborers                     MONDAY
3  House / apartment       Laborers                  WEDNESDAY
4  House / apartment     Core staff                   THURSDAY

        ORGANIZATION_TYPE  HOUSETYPE_MODE WALLSMATERIAL_MODE  \
0  Business Entity Type 3  block of flats      Stone, brick
1                  School  block of flats             Block
2              Government             NaN               NaN
3  Business Entity Type 3             NaN               NaN
4                Religion             NaN               NaN

  EMERGENCYSTATE_MODE
0                  No
1                  No
2                 NaN
3                 NaN
4                 NaN
```

[21]:
```
char.nunique()
```

[21]:
```
NAME_CONTRACT_TYPE            2
CODE_GENDER                   3
FLAG_OWN_CAR                  2
FLAG_OWN_REALTY               2
NAME_TYPE_SUITE               7
NAME_INCOME_TYPE              8
NAME_EDUCATION_TYPE           5
NAME_FAMILY_STATUS            6
NAME_HOUSING_TYPE             6
OCCUPATION_TYPE              18
WEEKDAY_APPR_PROCESS_START    7
ORGANIZATION_TYPE            58
HOUSETYPE_MODE                3
WALLSMATERIAL_MODE            7
EMERGENCYSTATE_MODE           2
```

```
dtype: int64
```

[22]:
```
char.isna().sum()
```

[22]:
```
NAME_CONTRACT_TYPE                0
CODE_GENDER                       0
FLAG_OWN_CAR                      0
FLAG_OWN_REALTY                   0
NAME_TYPE_SUITE                1292
NAME_INCOME_TYPE                  0
NAME_EDUCATION_TYPE               0
NAME_FAMILY_STATUS                0
NAME_HOUSING_TYPE                 0
OCCUPATION_TYPE               96391
WEEKDAY_APPR_PROCESS_START        0
ORGANIZATION_TYPE                 0
HOUSETYPE_MODE               154297
WALLSMATERIAL_MODE           156341
EMERGENCYSTATE_MODE          145755
dtype: int64
```

[ ]:
```
```

[23]:
```
# Filling missing values using median imputation for categoricaal columns
from sklearn.impute import SimpleImputer
imputer = SimpleImputer(strategy='most_frequent')
df1 = pd.DataFrame(imputer.fit_transform(char),columns=char.columns)
```

[24]:
```
#now it has no null values after imputing
df1.isnull().sum()
```

[24]:
```
NAME_CONTRACT_TYPE            0
CODE_GENDER                   0
FLAG_OWN_CAR                  0
FLAG_OWN_REALTY               0
NAME_TYPE_SUITE               0
NAME_INCOME_TYPE              0
NAME_EDUCATION_TYPE           0
NAME_FAMILY_STATUS            0
NAME_HOUSING_TYPE             0
OCCUPATION_TYPE               0
WEEKDAY_APPR_PROCESS_START    0
ORGANIZATION_TYPE             0
HOUSETYPE_MODE                0
WALLSMATERIAL_MODE            0
EMERGENCYSTATE_MODE           0
dtype: int64
```

[ ]:

[25]:
```
# Filling missing values using mean imputation for numeric columns
imputer = SimpleImputer(strategy='mean')
df2 = pd.DataFrame(imputer.fit_transform(numeric),columns=numeric.columns)
```

[26]:
```
sum(df2.isnull().sum())
```

[26]: 0

[27]:
```
#creating dummy var for categorical variables like one hot encoder
```

[28]:
```
dum_var=pd.get_dummies(df1,drop_first=True)
```

[29]:
```
dum_var
```

[29]:

| | NAME_CONTRACT_TYPE_Revolving loans | CODE_GENDER_M | CODE_GENDER_XNA | \ |
|---|---|---|---|---|
| 0 | 0 | 1 | 0 | |
| 1 | 0 | 0 | 0 | |
| 2 | 1 | 1 | 0 | |
| 3 | 0 | 0 | 0 | |
| 4 | 0 | 1 | 0 | |
| … | … | … | … | |
| 307506 | 0 | 1 | 0 | |
| 307507 | 0 | 0 | 0 | |
| 307508 | 0 | 0 | 0 | |
| 307509 | 0 | 0 | 0 | |
| 307510 | 0 | 0 | 0 | |

| | FLAG_OWN_CAR_Y | FLAG_OWN_REALTY_Y | NAME_TYPE_SUITE_Family | \ |
|---|---|---|---|---|
| 0 | 0 | 1 | 0 | |
| 1 | 0 | 0 | 1 | |
| 2 | 1 | 1 | 0 | |
| 3 | 0 | 1 | 0 | |
| 4 | 0 | 1 | 0 | |
| … | … | … | … | |
| 307506 | 0 | 0 | 0 | |
| 307507 | 0 | 1 | 0 | |
| 307508 | 0 | 1 | 0 | |
| 307509 | 0 | 1 | 0 | |
| 307510 | 0 | 0 | 0 | |

| | NAME_TYPE_SUITE_Group of people | NAME_TYPE_SUITE_Other_A | \ |
|---|---|---|---|
| 0 | 0 | 0 | |
| 1 | 0 | 0 | |
| 2 | 0 | 0 | |
| 3 | 0 | 0 | |

```
4                                    0                          0
…                                  …                          …
307506                               0                          0
307507                               0                          0
307508                               0                          0
307509                               0                          0
307510                               0                          0

        NAME_TYPE_SUITE_Other_B  NAME_TYPE_SUITE_Spouse, partner  \
0                             0                                 0
1                             0                                 0
2                             0                                 0
3                             0                                 0
4                             0                                 0
…                           …                                 …
307506                        0                                 0
307507                        0                                 0
307508                        0                                 0
307509                        0                                 0
307510                        0                                 0

        NAME_TYPE_SUITE_Unaccompanied  NAME_INCOME_TYPE_Commercial associate  \
0                                   1                                      0
1                                   0                                      0
2                                   1                                      0
3                                   1                                      0
4                                   1                                      0
…                                 …                                      …
307506                              1                                      0
307507                              1                                      0
307508                              1                                      0
307509                              1                                      1
307510                              1                                      1

        NAME_INCOME_TYPE_Maternity leave  NAME_INCOME_TYPE_Pensioner  \
0                                      0                           0
1                                      0                           0
2                                      0                           0
3                                      0                           0
4                                      0                           0
…                                    …                           …
307506                                 0                           0
307507                                 0                           1
307508                                 0                           0
307509                                 0                           0
307510                                 0                           0
```

```
        NAME_INCOME_TYPE_State servant  NAME_INCOME_TYPE_Student  \
0                                    0                         0
1                                    1                         0
2                                    0                         0
3                                    0                         0
4                                    0                         0
…                                  …                       …
307506                               0                         0
307507                               0                         0
307508                               0                         0
307509                               0                         0
307510                               0                         0

        NAME_INCOME_TYPE_Unemployed  NAME_INCOME_TYPE_Working  \
0                                 0                         1
1                                 0                         0
2                                 0                         1
3                                 0                         1
4                                 0                         1
…                               …                       …
307506                            0                         1
307507                            0                         0
307508                            0                         1
307509                            0                         0
307510                            0                         0

        NAME_EDUCATION_TYPE_Higher education  \
0                                          0
1                                          1
2                                          0
3                                          0
4                                          0
…                                        …
307506                                     0
307507                                     0
307508                                     1
307509                                     0
307510                                     1

        NAME_EDUCATION_TYPE_Incomplete higher  \
0                                           0
1                                           0
2                                           0
3                                           0
4                                           0
…                                         …
307506                                      0
```

```
307507                                       0
307508                                       0
307509                                       0
307510                                       0

        NAME_EDUCATION_TYPE_Lower secondary  \
0                                         0
1                                         0
2                                         0
3                                         0
4                                         0
…                                       …
307506                                    0
307507                                    0
307508                                    0
307509                                    0
307510                                    0

        NAME_EDUCATION_TYPE_Secondary / secondary special  \
0                                                       1
1                                                       0
2                                                       1
3                                                       1
4                                                       1
…                                                     …
307506                                                  1
307507                                                  1
307508                                                  0
307509                                                  1
307510                                                  0

        NAME_FAMILY_STATUS_Married  NAME_FAMILY_STATUS_Separated  \
0                                0                             0
1                                1                             0
2                                0                             0
3                                0                             0
4                                0                             0
…                              …                             …
307506                           0                             1
307507                           0                             0
307508                           0                             1
307509                           1                             0
307510                           1                             0

        NAME_FAMILY_STATUS_Single / not married  NAME_FAMILY_STATUS_Unknown  \
0                                             1                           0
1                                             0                           0
```

```
2                                          1                                0
3                                          0                                0
4                                          1                                0
…                                        …                                …
307506                                     0                                0
307507                                     0                                0
307508                                     0                                0
307509                                     0                                0
307510                                     0                                0

        NAME_FAMILY_STATUS_Widow  NAME_HOUSING_TYPE_House / apartment  \
0                              0                                    1
1                              0                                    1
2                              0                                    1
3                              0                                    1
4                              0                                    1
…                            …                                    …
307506                         0                                    0
307507                         1                                    1
307508                         0                                    1
307509                         0                                    1
307510                         0                                    1

        NAME_HOUSING_TYPE_Municipal apartment  \
0                                           0
1                                           0
2                                           0
3                                           0
4                                           0
…                                         …
307506                                      0
307507                                      0
307508                                      0
307509                                      0
307510                                      0

        NAME_HOUSING_TYPE_Office apartment  \
0                                        0
1                                        0
2                                        0
3                                        0
4                                        0
…                                      …
307506                                   0
307507                                   0
307508                                   0
307509                                   0
```

```
307510                                    0

        NAME_HOUSING_TYPE_Rented apartment  NAME_HOUSING_TYPE_With parents  \
0                                        0                               0
1                                        0                               0
2                                        0                               0
3                                        0                               0
4                                        0                               0
…                                      …                               …
307506                                   0                               1
307507                                   0                               0
307508                                   0                               0
307509                                   0                               0
307510                                   0                               0

        OCCUPATION_TYPE_Cleaning staff  OCCUPATION_TYPE_Cooking staff  \
0                                    0                             0
1                                    0                             0
2                                    0                             0
3                                    0                             0
4                                    0                             0
…                                  …                             …
307506                               0                             0
307507                               0                             0
307508                               0                             0
307509                               0                             0
307510                               0                             0

        OCCUPATION_TYPE_Core staff  OCCUPATION_TYPE_Drivers  \
0                                0                        0
1                                1                        0
2                                0                        0
3                                0                        0
4                                1                        0
…                              …                        …
307506                           0                        0
307507                           0                        0
307508                           0                        0
307509                           0                        0
307510                           0                        0

        OCCUPATION_TYPE_HR staff  OCCUPATION_TYPE_High skill tech staff  \
0                              0                                      0
1                              0                                      0
2                              0                                      0
3                              0                                      0
4                              0                                      0
```

```
…                          …                                       …
307506                         0                                       0
307507                         0                                       0
307508                         0                                       0
307509                         0                                       0
307510                         0                                       0

        OCCUPATION_TYPE_IT staff  OCCUPATION_TYPE_Laborers  \
0                               0                         1
1                               0                         0
2                               0                         1
3                               0                         1
4                               0                         0
…                           …                         …
307506                          0                         0
307507                          0                         1
307508                          0                         0
307509                          0                         1
307510                          0                         1

        OCCUPATION_TYPE_Low-skill Laborers  OCCUPATION_TYPE_Managers  \
0                                        0                         0
1                                        0                         0
2                                        0                         0
3                                        0                         0
4                                        0                         0
…                                    …                         …
307506                                   0                         0
307507                                   0                         0
307508                                   0                         1
307509                                   0                         0
307510                                   0                         0

        OCCUPATION_TYPE_Medicine staff  OCCUPATION_TYPE_Private service staff  \
0                                    0                                       0
1                                    0                                       0
2                                    0                                       0
3                                    0                                       0
4                                    0                                       0
…                                …                                       …
307506                               0                                       0
307507                               0                                       0
307508                               0                                       0
307509                               0                                       0
307510                               0                                       0

        OCCUPATION_TYPE_Realty agents  OCCUPATION_TYPE_Sales staff  \
```

```
0                                    0                                    0
1                                    0                                    0
2                                    0                                    0
3                                    0                                    0
4                                    0                                    0
…                                  …                                    …
307506                               0                                    1
307507                               0                                    0
307508                               0                                    0
307509                               0                                    0
307510                               0                                    0

        OCCUPATION_TYPE_Secretaries  OCCUPATION_TYPE_Security staff  \
0                                    0                                    0
1                                    0                                    0
2                                    0                                    0
3                                    0                                    0
4                                    0                                    0
…                                  …                                    …
307506                               0                                    0
307507                               0                                    0
307508                               0                                    0
307509                               0                                    0
307510                               0                                    0

        OCCUPATION_TYPE_Waiters/barmen staff  \
0                                          0
1                                          0
2                                          0
3                                          0
4                                          0
…                                        …
307506                                     0
307507                                     0
307508                                     0
307509                                     0
307510                                     0

        WEEKDAY_APPR_PROCESS_START_MONDAY  \
0                                       0
1                                       1
2                                       1
3                                       0
4                                       0
…                                     …
307506                                  0
307507                                  1
```

```
307508                                    0
307509                                    0
307510                                    0

        WEEKDAY_APPR_PROCESS_START_SATURDAY  \
0                                         0
1                                         0
2                                         0
3                                         0
4                                         0
…                                         …
307506                                    0
307507                                    0
307508                                    0
307509                                    0
307510                                    0

        WEEKDAY_APPR_PROCESS_START_SUNDAY  \
0                                       0
1                                       0
2                                       0
3                                       0
4                                       0
…                                       …
307506                                  0
307507                                  0
307508                                  0
307509                                  0
307510                                  0

        WEEKDAY_APPR_PROCESS_START_THURSDAY  \
0                                         0
1                                         0
2                                         0
3                                         0
4                                         1
…                                         …
307506                                    1
307507                                    0
307508                                    1
307509                                    0
307510                                    1

        WEEKDAY_APPR_PROCESS_START_TUESDAY  \
0                                        0
1                                        0
2                                        0
```

```
3                                       0
4                                       0
…                                       …
307506                                  0
307507                                  0
307508                                  0
307509                                  0
307510                                  0

        WEEKDAY_APPR_PROCESS_START_WEDNESDAY  ORGANIZATION_TYPE_Agriculture  \
0                                          1                              0
1                                          0                              0
2                                          0                              0
3                                          1                              0
4                                          0                              0
…                                          …                              …
307506                                     0                              0
307507                                     0                              0
307508                                     0                              0
307509                                     1                              0
307510                                     0                              0

        ORGANIZATION_TYPE_Bank  ORGANIZATION_TYPE_Business Entity Type 1  \
0                            0                                         0
1                            0                                         0
2                            0                                         0
3                            0                                         0
4                            0                                         0
…                            …                                         …
307506                       0                                         0
307507                       0                                         0
307508                       0                                         0
307509                       0                                         1
307510                       0                                         0

        ORGANIZATION_TYPE_Business Entity Type 2  \
0                                               0
1                                               0
2                                               0
3                                               0
4                                               0
…                                               …
307506                                          0
307507                                          0
307508                                          0
307509                                          0
307510                                          0
```

```
        ORGANIZATION_TYPE_Business Entity Type 3  ORGANIZATION_TYPE_Cleaning  \
0                                                1                          0
1                                                0                          0
2                                                0                          0
3                                                1                          0
4                                                0                          0
…                                              …                          …
307506                                           0                          0
307507                                           0                          0
307508                                           0                          0
307509                                           0                          0
307510                                           1                          0

        ORGANIZATION_TYPE_Construction  ORGANIZATION_TYPE_Culture  \
0                                    0                         0
1                                    0                         0
2                                    0                         0
3                                    0                         0
4                                    0                         0
…                                  …                         …
307506                               0                         0
307507                               0                         0
307508                               0                         0
307509                               0                         0
307510                               0                         0

        ORGANIZATION_TYPE_Electricity  ORGANIZATION_TYPE_Emergency  \
0                                   0                           0
1                                   0                           0
2                                   0                           0
3                                   0                           0
4                                   0                           0
…                                 …                           …
307506                              0                           0
307507                              0                           0
307508                              0                           0
307509                              0                           0
307510                              0                           0

        ORGANIZATION_TYPE_Government  ORGANIZATION_TYPE_Hotel  \
0                                  0                        0
1                                  0                        0
2                                  1                        0
3                                  0                        0
4                                  0                        0
…                                …                        …
```

```
307506                              0                          0
307507                              0                          0
307508                              0                          0
307509                              0                          0
307510                              0                          0

        ORGANIZATION_TYPE_Housing  ORGANIZATION_TYPE_Industry: type 1  \
0                               0                                    0
1                               0                                    0
2                               0                                    0
3                               0                                    0
4                               0                                    0
…                               …                                    …
307506                          0                                    0
307507                          0                                    0
307508                          0                                    0
307509                          0                                    0
307510                          0                                    0

        ORGANIZATION_TYPE_Industry: type 10  \
0                                         0
1                                         0
2                                         0
3                                         0
4                                         0
…                                         …
307506                                    0
307507                                    0
307508                                    0
307509                                    0
307510                                    0

        ORGANIZATION_TYPE_Industry: type 11  \
0                                         0
1                                         0
2                                         0
3                                         0
4                                         0
…                                         …
307506                                    0
307507                                    0
307508                                    0
307509                                    0
307510                                    0

        ORGANIZATION_TYPE_Industry: type 12  \
0                                         0
```

```
1                                          0
2                                          0
3                                          0
4                                          0
…                                          …
307506                                     0
307507                                     0
307508                                     0
307509                                     0
307510                                     0

        ORGANIZATION_TYPE_Industry: type 13  \
0                                          0
1                                          0
2                                          0
3                                          0
4                                          0
…                                          …
307506                                     0
307507                                     0
307508                                     0
307509                                     0
307510                                     0

        ORGANIZATION_TYPE_Industry: type 2  \
0                                         0
1                                         0
2                                         0
3                                         0
4                                         0
…                                         …
307506                                    0
307507                                    0
307508                                    0
307509                                    0
307510                                    0

        ORGANIZATION_TYPE_Industry: type 3  \
0                                         0
1                                         0
2                                         0
3                                         0
4                                         0
…                                         …
307506                                    0
307507                                    0
307508                                    0
```

```
307509                                    0
307510                                    0

        ORGANIZATION_TYPE_Industry: type 4  \
0                                         0
1                                         0
2                                         0
3                                         0
4                                         0
…                                       …
307506                                    0
307507                                    0
307508                                    0
307509                                    0
307510                                    0

        ORGANIZATION_TYPE_Industry: type 5  \
0                                         0
1                                         0
2                                         0
3                                         0
4                                         0
…                                       …
307506                                    0
307507                                    0
307508                                    0
307509                                    0
307510                                    0

        ORGANIZATION_TYPE_Industry: type 6  \
0                                         0
1                                         0
2                                         0
3                                         0
4                                         0
…                                       …
307506                                    0
307507                                    0
307508                                    0
307509                                    0
307510                                    0

        ORGANIZATION_TYPE_Industry: type 7  \
0                                         0
1                                         0
2                                         0
3                                         0
```

```
4                                        0
…                                      …
307506                                   0
307507                                   0
307508                                   0
307509                                   0
307510                                   0

        ORGANIZATION_TYPE_Industry: type 8  \
0                                        0
1                                        0
2                                        0
3                                        0
4                                        0
…                                      …
307506                                   0
307507                                   0
307508                                   0
307509                                   0
307510                                   0

        ORGANIZATION_TYPE_Industry: type 9  ORGANIZATION_TYPE_Insurance  \
0                                        0                             0
1                                        0                             0
2                                        0                             0
3                                        0                             0
4                                        0                             0
…                                      …                           …
307506                                   0                             0
307507                                   0                             0
307508                                   0                             0
307509                                   0                             0
307510                                   0                             0

        ORGANIZATION_TYPE_Kindergarten  ORGANIZATION_TYPE_Legal Services  \
0                                    0                                 0
1                                    0                                 0
2                                    0                                 0
3                                    0                                 0
4                                    0                                 0
…                                  …                                 …
307506                               0                                 0
307507                               0                                 0
307508                               0                                 0
307509                               0                                 0
307510                               0                                 0
```

```
        ORGANIZATION_TYPE_Medicine  ORGANIZATION_TYPE_Military  \
0                                0                           0
1                                0                           0
2                                0                           0
3                                0                           0
4                                0                           0
…                              …                           …
307506                           0                           0
307507                           0                           0
307508                           0                           0
307509                           0                           0
307510                           0                           0

        ORGANIZATION_TYPE_Mobile  ORGANIZATION_TYPE_Other  \
0                              0                        0
1                              0                        0
2                              0                        0
3                              0                        0
4                              0                        0
…                            …                        …
307506                         0                        0
307507                         0                        0
307508                         0                        0
307509                         0                        0
307510                         0                        0

        ORGANIZATION_TYPE_Police  ORGANIZATION_TYPE_Postal  \
0                              0                         0
1                              0                         0
2                              0                         0
3                              0                         0
4                              0                         0
…                            …                         …
307506                         0                         0
307507                         0                         0
307508                         0                         0
307509                         0                         0
307510                         0                         0

        ORGANIZATION_TYPE_Realtor  ORGANIZATION_TYPE_Religion  \
0                               0                           0
1                               0                           0
2                               0                           0
3                               0                           0
4                               0                           1
…                             …                           …
307506                          0                           0
```

```
307507                           0                           0
307508                           0                           0
307509                           0                           0
307510                           0                           0

        ORGANIZATION_TYPE_Restaurant  ORGANIZATION_TYPE_School  \
0                                  0                         0
1                                  0                         1
2                                  0                         0
3                                  0                         0
4                                  0                         0
…                                …                         …
307506                             0                         0
307507                             0                         0
307508                             0                         1
307509                             0                         0
307510                             0                         0

        ORGANIZATION_TYPE_Security  ORGANIZATION_TYPE_Security Ministries  \
0                                0                                       0
1                                0                                       0
2                                0                                       0
3                                0                                       0
4                                0                                       0
…                              …                                       …
307506                           0                                       0
307507                           0                                       0
307508                           0                                       0
307509                           0                                       0
307510                           0                                       0

        ORGANIZATION_TYPE_Self-employed  ORGANIZATION_TYPE_Services  \
0                                     0                           0
1                                     0                           0
2                                     0                           0
3                                     0                           0
4                                     0                           0
…                                   …                           …
307506                                0                           1
307507                                0                           0
307508                                0                           0
307509                                0                           0
307510                                0                           0

        ORGANIZATION_TYPE_Telecom  ORGANIZATION_TYPE_Trade: type 1  \
0                               0                                0
1                               0                                0
```

```
2                                   0                                   0
3                                   0                                   0
4                                   0                                   0
…                                   …                                   …
307506                              0                                   0
307507                              0                                   0
307508                              0                                   0
307509                              0                                   0
307510                              0                                   0

        ORGANIZATION_TYPE_Trade: type 2  ORGANIZATION_TYPE_Trade: type 3  \
0                                     0                                   0
1                                     0                                   0
2                                     0                                   0
3                                     0                                   0
4                                     0                                   0
…                                     …                                   …
307506                                0                                   0
307507                                0                                   0
307508                                0                                   0
307509                                0                                   0
307510                                0                                   0

        ORGANIZATION_TYPE_Trade: type 4  ORGANIZATION_TYPE_Trade: type 5  \
0                                     0                                   0
1                                     0                                   0
2                                     0                                   0
3                                     0                                   0
4                                     0                                   0
…                                     …                                   …
307506                                0                                   0
307507                                0                                   0
307508                                0                                   0
307509                                0                                   0
307510                                0                                   0

        ORGANIZATION_TYPE_Trade: type 6  ORGANIZATION_TYPE_Trade: type 7  \
0                                     0                                   0
1                                     0                                   0
2                                     0                                   0
3                                     0                                   0
4                                     0                                   0
…                                     …                                   …
307506                                0                                   0
307507                                0                                   0
307508                                0                                   0
307509                                0                                   0
```

```
307510                                  0                                  0

        ORGANIZATION_TYPE_Transport: type 1  \
0                                       0
1                                       0
2                                       0
3                                       0
4                                       0
…                                       …
307506                                  0
307507                                  0
307508                                  0
307509                                  0
307510                                  0

        ORGANIZATION_TYPE_Transport: type 2  \
0                                       0
1                                       0
2                                       0
3                                       0
4                                       0
…                                       …
307506                                  0
307507                                  0
307508                                  0
307509                                  0
307510                                  0

        ORGANIZATION_TYPE_Transport: type 3  \
0                                       0
1                                       0
2                                       0
3                                       0
4                                       0
…                                       …
307506                                  0
307507                                  0
307508                                  0
307509                                  0
307510                                  0

        ORGANIZATION_TYPE_Transport: type 4  ORGANIZATION_TYPE_University  \
0                                       0                                  0
1                                       0                                  0
2                                       0                                  0
3                                       0                                  0
4                                       0                                  0
```

```
…                                    …                                …
307506                                   0                                0
307507                                   0                                0
307508                                   0                                0
307509                                   0                                0
307510                                   0                                0

        ORGANIZATION_TYPE_XNA  HOUSETYPE_MODE_specific housing  \
0                           0                                 0
1                           0                                 0
2                           0                                 0
3                           0                                 0
4                           0                                 0
…                         …                                 …
307506                      0                                 0
307507                      1                                 0
307508                      0                                 0
307509                      0                                 0
307510                      0                                 0

        HOUSETYPE_MODE_terraced house  WALLSMATERIAL_MODE_Mixed  \
0                                   0                          0
1                                   0                          0
2                                   0                          0
3                                   0                          0
4                                   0                          0
…                                 …                          …
307506                              0                          0
307507                              0                          0
307508                              0                          0
307509                              0                          0
307510                              0                          0

        WALLSMATERIAL_MODE_Monolithic  WALLSMATERIAL_MODE_Others  \
0                                   0                           0
1                                   0                           0
2                                   0                           0
3                                   0                           0
4                                   0                           0
…                                 …                           …
307506                              0                           0
307507                              0                           0
307508                              0                           0
307509                              0                           0
307510                              0                           0

        WALLSMATERIAL_MODE_Panel  WALLSMATERIAL_MODE_Stone, brick  \
```

```
0                                0                                1
1                                0                                0
2                                1                                0
3                                1                                0
4                                1                                0
…                              …                                …
307506                           0                                1
307507                           0                                1
307508                           1                                0
307509                           0                                1
307510                           1                                0

        WALLSMATERIAL_MODE_Wooden  EMERGENCYSTATE_MODE_Yes
0                               0                        0
1                               0                        0
2                               0                        0
3                               0                        0
4                               0                        0
…                             …                        …
307506                          0                        0
307507                          0                        0
307508                          0                        0
307509                          0                        0
307510                          0                        0

[307511 rows x 121 columns]
```

[30]:
```
#concatinating both numeric and categorical coluns after imputing and encoding
df_new=pd.concat([df2,dum_var],axis=1)
```

[31]:
```
df_new.head()
```

[31]:
```
   SK_ID_CURR  TARGET  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT  \
0    100002.0     1.0           0.0          202500.0    406597.5
1    100003.0     0.0           0.0          270000.0   1293502.5
2    100004.0     0.0           0.0           67500.0    135000.0
3    100006.0     0.0           0.0          135000.0    312682.5
4    100007.0     0.0           0.0          121500.0    513000.0

   AMT_ANNUITY  AMT_GOODS_PRICE  REGION_POPULATION_RELATIVE  DAYS_BIRTH  \
0      24700.5         351000.0                    0.018801     -9461.0
1      35698.5        1129500.0                    0.003541    -16765.0
2       6750.0         135000.0                    0.010032    -19046.0
3      29686.5         297000.0                    0.008019    -19005.0
4      21865.5         513000.0                    0.028663    -19932.0

   DAYS_EMPLOYED  DAYS_REGISTRATION  DAYS_ID_PUBLISH  FLAG_MOBIL  \
```

```
0         -637.0           -3648.0          -2120.0         1.0
1        -1188.0           -1186.0           -291.0         1.0
2         -225.0           -4260.0          -2531.0         1.0
3        -3039.0           -9833.0          -2437.0         1.0
4        -3038.0           -4311.0          -3458.0         1.0

   FLAG_EMP_PHONE  FLAG_WORK_PHONE  FLAG_CONT_MOBILE  FLAG_PHONE  FLAG_EMAIL  \
0             1.0              0.0               1.0         1.0         0.0
1             1.0              0.0               1.0         1.0         0.0
2             1.0              1.0               1.0         1.0         0.0
3             1.0              0.0               1.0         0.0         0.0
4             1.0              0.0               1.0         0.0         0.0

   CNT_FAM_MEMBERS  REGION_RATING_CLIENT  REGION_RATING_CLIENT_W_CITY  \
0              1.0                   2.0                          2.0
1              2.0                   1.0                          1.0
2              1.0                   2.0                          2.0
3              2.0                   2.0                          2.0
4              1.0                   2.0                          2.0

   HOUR_APPR_PROCESS_START  REG_REGION_NOT_LIVE_REGION  \
0                     10.0                         0.0
1                     11.0                         0.0
2                      9.0                         0.0
3                     17.0                         0.0
4                     11.0                         0.0

   REG_REGION_NOT_WORK_REGION  LIVE_REGION_NOT_WORK_REGION  \
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

   REG_CITY_NOT_LIVE_CITY  REG_CITY_NOT_WORK_CITY  LIVE_CITY_NOT_WORK_CITY  \
0                     0.0                     0.0                      0.0
1                     0.0                     0.0                      0.0
2                     0.0                     0.0                      0.0
3                     0.0                     0.0                      0.0
4                     0.0                     1.0                      1.0

   EXT_SOURCE_1  EXT_SOURCE_2  EXT_SOURCE_3  APARTMENTS_AVG  BASEMENTAREA_AVG  \
0      0.083037      0.262949      0.139376         0.02470          0.036900
1      0.311267      0.622246      0.510853         0.09590          0.052900
2      0.502130      0.555912      0.729567         0.11744          0.088442
3      0.502130      0.650442      0.510853         0.11744          0.088442
4      0.502130      0.322738      0.510853         0.11744          0.088442
```

```
   YEARS_BEGINEXPLUATATION_AVG  ELEVATORS_AVG  ENTRANCES_AVG  FLOORSMAX_AVG  \
0                     0.972200       0.000000       0.069000       0.083300
1                     0.985100       0.080000       0.034500       0.291700
2                     0.977735       0.078942       0.149725       0.226282
3                     0.977735       0.078942       0.149725       0.226282
4                     0.977735       0.078942       0.149725       0.226282

   LANDAREA_AVG  LIVINGAREA_AVG  NONLIVINGAREA_AVG  APARTMENTS_MODE  \
0      0.036900        0.019000           0.000000         0.025200
1      0.013000        0.054900           0.009800         0.092400
2      0.066333        0.107399           0.028358         0.114231
3      0.066333        0.107399           0.028358         0.114231
4      0.066333        0.107399           0.028358         0.114231

   BASEMENTAREA_MODE  YEARS_BEGINEXPLUATATION_MODE  ELEVATORS_MODE  \
0           0.038300                      0.972200         0.00000
1           0.053800                      0.985100         0.08060
2           0.087543                      0.977065         0.07449
3           0.087543                      0.977065         0.07449
4           0.087543                      0.977065         0.07449

   ENTRANCES_MODE  FLOORSMAX_MODE  LANDAREA_MODE  LIVINGAREA_MODE  \
0        0.069000        0.083300       0.037700         0.019800
1        0.034500        0.291700       0.012800         0.055400
2        0.145193        0.222315       0.064958         0.105975
3        0.145193        0.222315       0.064958         0.105975
4        0.145193        0.222315       0.064958         0.105975

   NONLIVINGAREA_MODE  APARTMENTS_MEDI  BASEMENTAREA_MEDI  \
0            0.000000          0.02500           0.036900
1            0.000000          0.09680           0.052900
2            0.027022          0.11785           0.087955
3            0.027022          0.11785           0.087955
4            0.027022          0.11785           0.087955

   YEARS_BEGINEXPLUATATION_MEDI  ELEVATORS_MEDI  ENTRANCES_MEDI  \
0                      0.972200        0.000000        0.069000
1                      0.985100        0.080000        0.034500
2                      0.977752        0.078078        0.149213
3                      0.977752        0.078078        0.149213
4                      0.977752        0.078078        0.149213

   FLOORSMAX_MEDI  LANDAREA_MEDI  LIVINGAREA_MEDI  NONLIVINGAREA_MEDI  \
0        0.083300       0.037500         0.019300            0.000000
1        0.291700       0.013200         0.055800            0.010000
2        0.225897       0.067169         0.108607            0.028236
```

```
3       0.225897       0.067169         0.108607            0.028236
4       0.225897       0.067169         0.108607            0.028236

   TOTALAREA_MODE  OBS_30_CNT_SOCIAL_CIRCLE  DEF_30_CNT_SOCIAL_CIRCLE  \
0        0.014900                       2.0                       2.0
1        0.071400                       1.0                       0.0
2        0.102547                       0.0                       0.0
3        0.102547                       2.0                       0.0
4        0.102547                       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              0.0              1.0              0.0              0.0
1              0.0              1.0              0.0              0.0
2              0.0              0.0              0.0              0.0
3              0.0              1.0              0.0              0.0
4              0.0              0.0              0.0              0.0

   FLAG_DOCUMENT_6  FLAG_DOCUMENT_7  FLAG_DOCUMENT_8  FLAG_DOCUMENT_9  \
0              0.0              0.0              0.0              0.0
1              0.0              0.0              0.0              0.0
2              0.0              0.0              0.0              0.0
3              0.0              0.0              0.0              0.0
4              0.0              0.0              1.0              0.0

   FLAG_DOCUMENT_10  FLAG_DOCUMENT_11  FLAG_DOCUMENT_12  FLAG_DOCUMENT_13  \
0               0.0               0.0               0.0               0.0
1               0.0               0.0               0.0               0.0
2               0.0               0.0               0.0               0.0
3               0.0               0.0               0.0               0.0
4               0.0               0.0               0.0               0.0

   FLAG_DOCUMENT_14  FLAG_DOCUMENT_15  FLAG_DOCUMENT_16  FLAG_DOCUMENT_17  \
0               0.0               0.0               0.0               0.0
1               0.0               0.0               0.0               0.0
2               0.0               0.0               0.0               0.0
3               0.0               0.0               0.0               0.0
4               0.0               0.0               0.0               0.0

   FLAG_DOCUMENT_18  FLAG_DOCUMENT_19  FLAG_DOCUMENT_20  FLAG_DOCUMENT_21  \
0               0.0               0.0               0.0               0.0
```

```
1                0.0                0.0                0.0                0.0
2                0.0                0.0                0.0                0.0
3                0.0                0.0                0.0                0.0
4                0.0                0.0                0.0                0.0

   AMT_REQ_CREDIT_BUREAU_HOUR  AMT_REQ_CREDIT_BUREAU_DAY  \
0                    0.000000                      0.000
1                    0.000000                      0.000
2                    0.000000                      0.000
3                    0.006402                      0.007
4                    0.000000                      0.000

   AMT_REQ_CREDIT_BUREAU_WEEK  AMT_REQ_CREDIT_BUREAU_MON  \
0                    0.000000                   0.000000
1                    0.000000                   0.000000
2                    0.000000                   0.000000
3                    0.034362                   0.267395
4                    0.000000                   0.000000

   AMT_REQ_CREDIT_BUREAU_QRT  AMT_REQ_CREDIT_BUREAU_YEAR  \
0                   0.000000                   1.000000
1                   0.000000                   0.000000
2                   0.000000                   0.000000
3                   0.265474                   1.899974
4                   0.000000                   0.000000

   NAME_CONTRACT_TYPE_Revolving loans  CODE_GENDER_M  CODE_GENDER_XNA  \
0                                   0              1                0
1                                   0              0                0
2                                   1              1                0
3                                   0              0                0
4                                   0              1                0

   FLAG_OWN_CAR_Y  FLAG_OWN_REALTY_Y  NAME_TYPE_SUITE_Family  \
0               0                  1                       0
1               0                  0                       1
2               1                  1                       0
3               0                  1                       0
4               0                  1                       0

   NAME_TYPE_SUITE_Group of people  NAME_TYPE_SUITE_Other_A  \
0                                0                       0
1                                0                       0
2                                0                       0
3                                0                       0
4                                0                       0
```

```
   NAME_TYPE_SUITE_Other_B  NAME_TYPE_SUITE_Spouse, partner  \
0                        0                                0
1                        0                                0
2                        0                                0
3                        0                                0
4                        0                                0

   NAME_TYPE_SUITE_Unaccompanied  NAME_INCOME_TYPE_Commercial associate  \
0                              1                                     0
1                              0                                     0
2                              1                                     0
3                              1                                     0
4                              1                                     0

   NAME_INCOME_TYPE_Maternity leave  NAME_INCOME_TYPE_Pensioner  \
0                                 0                           0
1                                 0                           0
2                                 0                           0
3                                 0                           0
4                                 0                           0

   NAME_INCOME_TYPE_State servant  NAME_INCOME_TYPE_Student  \
0                               0                         0
1                               1                         0
2                               0                         0
3                               0                         0
4                               0                         0

   NAME_INCOME_TYPE_Unemployed  NAME_INCOME_TYPE_Working  \
0                            0                         1
1                            0                         0
2                            0                         1
3                            0                         1
4                            0                         1

   NAME_EDUCATION_TYPE_Higher education  \
0                                     0
1                                     1
2                                     0
3                                     0
4                                     0

   NAME_EDUCATION_TYPE_Incomplete higher  NAME_EDUCATION_TYPE_Lower secondary  \
0                                      0                                   0
1                                      0                                   0
2                                      0                                   0
3                                      0                                   0
```

```
4                                     0                                     0

   NAME_EDUCATION_TYPE_Secondary / secondary special  \
0                                                  1
1                                                  0
2                                                  1
3                                                  1
4                                                  1

   NAME_FAMILY_STATUS_Married  NAME_FAMILY_STATUS_Separated  \
0                           0                             0
1                           1                             0
2                           0                             0
3                           0                             0
4                           0                             0

   NAME_FAMILY_STATUS_Single / not married  NAME_FAMILY_STATUS_Unknown  \
0                                        1                           0
1                                        0                           0
2                                        1                           0
3                                        0                           0
4                                        1                           0

   NAME_FAMILY_STATUS_Widow  NAME_HOUSING_TYPE_House / apartment  \
0                         0                                   1
1                         0                                   1
2                         0                                   1
3                         0                                   1
4                         0                                   1

   NAME_HOUSING_TYPE_Municipal apartment  NAME_HOUSING_TYPE_Office apartment  \
0                                      0                                   0
1                                      0                                   0
2                                      0                                   0
3                                      0                                   0
4                                      0                                   0

   NAME_HOUSING_TYPE_Rented apartment  NAME_HOUSING_TYPE_With parents  \
0                                   0                               0
1                                   0                               0
2                                   0                               0
3                                   0                               0
4                                   0                               0

   OCCUPATION_TYPE_Cleaning staff  OCCUPATION_TYPE_Cooking staff  \
0                               0                              0
1                               0                              0
```

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

   OCCUPATION_TYPE_Core staff  OCCUPATION_TYPE_Drivers  \
0                           0                        0
1                           1                        0
2                           0                        0
3                           0                        0
4                           1                        0

   OCCUPATION_TYPE_HR staff  OCCUPATION_TYPE_High skill tech staff  \
0                         0                                      0
1                         0                                      0
2                         0                                      0
3                         0                                      0
4                         0                                      0

   OCCUPATION_TYPE_IT staff  OCCUPATION_TYPE_Laborers  \
0                         0                        1
1                         0                        0
2                         0                        1
3                         0                        1
4                         0                        0

   OCCUPATION_TYPE_Low-skill Laborers  OCCUPATION_TYPE_Managers  \
0                                   0                        0
1                                   0                        0
2                                   0                        0
3                                   0                        0
4                                   0                        0

   OCCUPATION_TYPE_Medicine staff  OCCUPATION_TYPE_Private service staff  \
0                               0                                      0
1                               0                                      0
2                               0                                      0
3                               0                                      0
4                               0                                      0

   OCCUPATION_TYPE_Realty agents  OCCUPATION_TYPE_Sales staff  \
0                              0                           0
1                              0                           0
2                              0                           0
3                              0                           0
4                              0                           0

   OCCUPATION_TYPE_Secretaries  OCCUPATION_TYPE_Security staff  \
```

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

   OCCUPATION_TYPE_Waiters/barmen staff  WEEKDAY_APPR_PROCESS_START_MONDAY  \
0                                     0                                  0
1                                     0                                  1
2                                     0                                  1
3                                     0                                  0
4                                     0                                  0

   WEEKDAY_APPR_PROCESS_START_SATURDAY  WEEKDAY_APPR_PROCESS_START_SUNDAY  \
0                                    0                                  0
1                                    0                                  0
2                                    0                                  0
3                                    0                                  0
4                                    0                                  0

   WEEKDAY_APPR_PROCESS_START_THURSDAY  WEEKDAY_APPR_PROCESS_START_TUESDAY  \
0                                    0                                  0
1                                    0                                  0
2                                    0                                  0
3                                    0                                  0
4                                    1                                  0

   WEEKDAY_APPR_PROCESS_START_WEDNESDAY  ORGANIZATION_TYPE_Agriculture  \
0                                     1                             0
1                                     0                             0
2                                     0                             0
3                                     1                             0
4                                     0                             0

   ORGANIZATION_TYPE_Bank  ORGANIZATION_TYPE_Business Entity Type 1  \
0                       0                                         0
1                       0                                         0
2                       0                                         0
3                       0                                         0
4                       0                                         0

   ORGANIZATION_TYPE_Business Entity Type 2  \
0                                         0
1                                         0
2                                         0
3                                         0
4                                         0
```

```
   ORGANIZATION_TYPE_Business Entity Type 3  ORGANIZATION_TYPE_Cleaning  \
0                                         1                          0
1                                         0                          0
2                                         0                          0
3                                         1                          0
4                                         0                          0

   ORGANIZATION_TYPE_Construction  ORGANIZATION_TYPE_Culture  \
0                               0                         0
1                               0                         0
2                               0                         0
3                               0                         0
4                               0                         0

   ORGANIZATION_TYPE_Electricity  ORGANIZATION_TYPE_Emergency  \
0                              0                           0
1                              0                           0
2                              0                           0
3                              0                           0
4                              0                           0

   ORGANIZATION_TYPE_Government  ORGANIZATION_TYPE_Hotel  \
0                             0                        0
1                             0                        0
2                             1                        0
3                             0                        0
4                             0                        0

   ORGANIZATION_TYPE_Housing  ORGANIZATION_TYPE_Industry: type 1  \
0                          0                                   0
1                          0                                   0
2                          0                                   0
3                          0                                   0
4                          0                                   0

   ORGANIZATION_TYPE_Industry: type 10  ORGANIZATION_TYPE_Industry: type 11  \
0                                    0                                    0
1                                    0                                    0
2                                    0                                    0
3                                    0                                    0
4                                    0                                    0

   ORGANIZATION_TYPE_Industry: type 12  ORGANIZATION_TYPE_Industry: type 13  \
0                                    0                                    0
1                                    0                                    0
2                                    0                                    0
```

```
3                                    0                                    0
4                                    0                                    0

   ORGANIZATION_TYPE_Industry: type 2  ORGANIZATION_TYPE_Industry: type 3  \
0                                    0                                    0
1                                    0                                    0
2                                    0                                    0
3                                    0                                    0
4                                    0                                    0

   ORGANIZATION_TYPE_Industry: type 4  ORGANIZATION_TYPE_Industry: type 5  \
0                                    0                                    0
1                                    0                                    0
2                                    0                                    0
3                                    0                                    0
4                                    0                                    0

   ORGANIZATION_TYPE_Industry: type 6  ORGANIZATION_TYPE_Industry: type 7  \
0                                    0                                    0
1                                    0                                    0
2                                    0                                    0
3                                    0                                    0
4                                    0                                    0

   ORGANIZATION_TYPE_Industry: type 8  ORGANIZATION_TYPE_Industry: type 9  \
0                                    0                                    0
1                                    0                                    0
2                                    0                                    0
3                                    0                                    0
4                                    0                                    0

   ORGANIZATION_TYPE_Insurance  ORGANIZATION_TYPE_Kindergarten  \
0                            0                               0
1                            0                               0
2                            0                               0
3                            0                               0
4                            0                               0

   ORGANIZATION_TYPE_Legal Services  ORGANIZATION_TYPE_Medicine  \
0                                 0                           0
1                                 0                           0
2                                 0                           0
3                                 0                           0
4                                 0                           0

   ORGANIZATION_TYPE_Military  ORGANIZATION_TYPE_Mobile  \
0                           0                         0
```

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

   ORGANIZATION_TYPE_Other  ORGANIZATION_TYPE_Police  \
0                        0                         0
1                        0                         0
2                        0                         0
3                        0                         0
4                        0                         0

   ORGANIZATION_TYPE_Postal  ORGANIZATION_TYPE_Realtor  \
0                         0                          0
1                         0                          0
2                         0                          0
3                         0                          0
4                         0                          0

   ORGANIZATION_TYPE_Religion  ORGANIZATION_TYPE_Restaurant  \
0                           0                             0
1                           0                             0
2                           0                             0
3                           0                             0
4                           1                             0

   ORGANIZATION_TYPE_School  ORGANIZATION_TYPE_Security  \
0                         0                           0
1                         1                           0
2                         0                           0
3                         0                           0
4                         0                           0

   ORGANIZATION_TYPE_Security Ministries  ORGANIZATION_TYPE_Self-employed  \
0                                      0                                0
1                                      0                                0
2                                      0                                0
3                                      0                                0
4                                      0                                0

   ORGANIZATION_TYPE_Services  ORGANIZATION_TYPE_Telecom  \
0                           0                          0
1                           0                          0
2                           0                          0
3                           0                          0
4                           0                          0
```

```
   ORGANIZATION_TYPE_Trade: type 1  ORGANIZATION_TYPE_Trade: type 2  \
0                                0                                0
1                                0                                0
2                                0                                0
3                                0                                0
4                                0                                0

   ORGANIZATION_TYPE_Trade: type 3  ORGANIZATION_TYPE_Trade: type 4  \
0                                0                                0
1                                0                                0
2                                0                                0
3                                0                                0
4                                0                                0

   ORGANIZATION_TYPE_Trade: type 5  ORGANIZATION_TYPE_Trade: type 6  \
0                                0                                0
1                                0                                0
2                                0                                0
3                                0                                0
4                                0                                0

   ORGANIZATION_TYPE_Trade: type 7  ORGANIZATION_TYPE_Transport: type 1  \
0                                0                                   0
1                                0                                   0
2                                0                                   0
3                                0                                   0
4                                0                                   0

   ORGANIZATION_TYPE_Transport: type 2  ORGANIZATION_TYPE_Transport: type 3  \
0                                    0                                   0
1                                    0                                   0
2                                    0                                   0
3                                    0                                   0
4                                    0                                   0

   ORGANIZATION_TYPE_Transport: type 4  ORGANIZATION_TYPE_University  \
0                                    0                             0
1                                    0                             0
2                                    0                             0
3                                    0                             0
4                                    0                             0

   ORGANIZATION_TYPE_XNA  HOUSETYPE_MODE_specific housing  \
0                      0                                0
1                      0                                0
2                      0                                0
3                      0                                0
```

```
4                       0                                   0

   HOUSETYPE_MODE_terraced house  WALLSMATERIAL_MODE_Mixed  \
0                              0                         0
1                              0                         0
2                              0                         0
3                              0                         0
4                              0                         0

   WALLSMATERIAL_MODE_Monolithic  WALLSMATERIAL_MODE_Others  \
0                              0                         0
1                              0                         0
2                              0                         0
3                              0                         0
4                              0                         0

   WALLSMATERIAL_MODE_Panel  WALLSMATERIAL_MODE_Stone, brick  \
0                         0                               1
1                         0                               0
2                         1                               0
3                         1                               0
4                         1                               0

   WALLSMATERIAL_MODE_Wooden  EMERGENCYSTATE_MODE_Yes
0                          0                       0
1                          0                       0
2                          0                       0
3                          0                       0
4                          0                       0
```

[32]:
```
#now the dataset cleaned from missing values and categorical encode, now we can␣
 ↪split the dataset for training
```

[33]:
```
df_new.shape
```

[33]:
```
(307511, 211)
```

[34]:
```
#splitting dependent and independent variables
#we are remmoving SK_ID_CURR and TARGET column for independent var

y=df_new['TARGET']
x=df_new.drop(['SK_ID_CURR', 'TARGET'],1)
```

[35]:
```
#the target column has imbalance data, we are doing oversampling technique to␣
 ↪make data balance using smote
```

[36]:
```
from imblearn.over_sampling import SMOTE
smk = SMOTE()
x_train_smote,y_train_smote=smk.fit_resample(x,y)
```

[37]:
```
from collections import Counter
print('Original dataset shape {}'.format(Counter(y)))
print('Resampled dataset shape {}'.format(Counter(y_train_smote)))
```

```
Original dataset shape Counter({0.0: 282686, 1.0: 24825})
Resampled dataset shape Counter({1.0: 282686, 0.0: 282686})
```

[38]:
```
#checking shape of the final data
```

[39]:
```
print(x_train_smote.shape)
print(y_train_smote.shape)
```

```
(565372, 209)
(565372,)
```

[ ]:
```

```

[40]:
```
# Split the data set into training and testing
x_train, x_test, y_train, y_test = train_test_split(
   x_train_smote, y_train_smote, test_size=0.25, random_state=2)
```

[41]:
```
#scaling values between 0 and 1
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
x_train_std = scaler.fit_transform(x_train)
x_test_std=scaler.transform(x_test)
```

[42]:
```
#shape of training data
x_train_std.shape
```

[42]: (424029, 209)

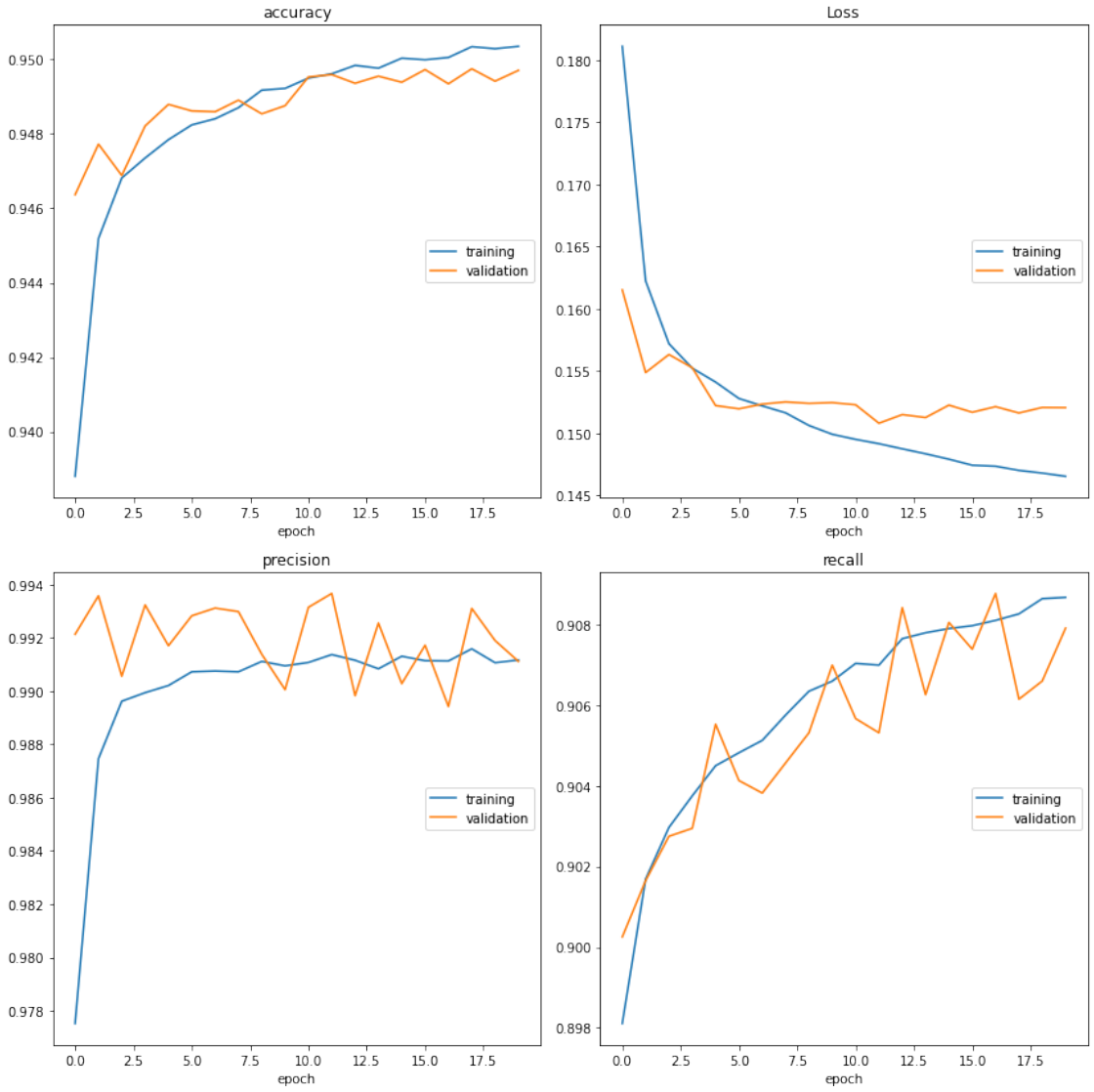
[43]:
```
#shape of testing data
x_test_std.shape
```

[43]: (141343, 209)

[44]:
```
#cearting deep learning model to make predictions
```

[45]:
```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, BatchNormalization,Input
from tensorflow.keras.metrics import Precision, Recall
from livelossplot import PlotLossesKerasTF
```

[46]:
```
model=Sequential()

model.add(Input(shape=(209,)))

model.add(Dense(32,activation='relu'))
model.add(BatchNormalization())

model.add(Dense(64,activation='relu'))
model.add(BatchNormalization())

model.add(Dense(1,activation='sigmoid'))

model.
 ↪compile(loss='binary_crossentropy',optimizer='adam',metrics=['accuracy',Precision(),Recall(

model.fit(x_train_std,y_train,epochs=20,␣
 ↪batch_size=64,validation_data=(x_test_std,y_test),callbacks=[PlotLossesKerasTF()])
```




```
accuracy
        training                (min:    0.939, max:    0.950, cur:    0.950)
        validation              (min:    0.946, max:    0.950, cur:    0.950)
Loss
        training                (min:    0.147, max:    0.181, cur:    0.147)
        validation              (min:    0.151, max:    0.162, cur:    0.152)
precision
        training                (min:    0.978, max:    0.992, cur:    0.991)
        validation              (min:    0.989, max:    0.994, cur:    0.991)
recall
        training                (min:    0.898, max:    0.909, cur:    0.909)
        validation              (min:    0.900, max:    0.909, cur:    0.908)
6626/6626 [==============================] - 16s 2ms/step - loss: 0.1465 -
```

```
accuracy: 0.9504 - precision: 0.9912 - recall: 0.9087 - val_loss: 0.1521 -
val_accuracy: 0.9497 - val_precision: 0.9911 - val_recall: 0.9079
```

[46]:
```
<keras.callbacks.History at 0x7f408c6bcc90>
```

[47]:
```
# Calculate Sensitivity or recall as a metric
# recallor senstivity has 90% for both training and testing
```

[48]:
```
# Calculate area under receiver operating characteristics curve
```

[49]:
```
from sklearn.metrics import roc_curve,roc_auc_score
```

[50]:
```
pred=model.predict(x_test_std)
```

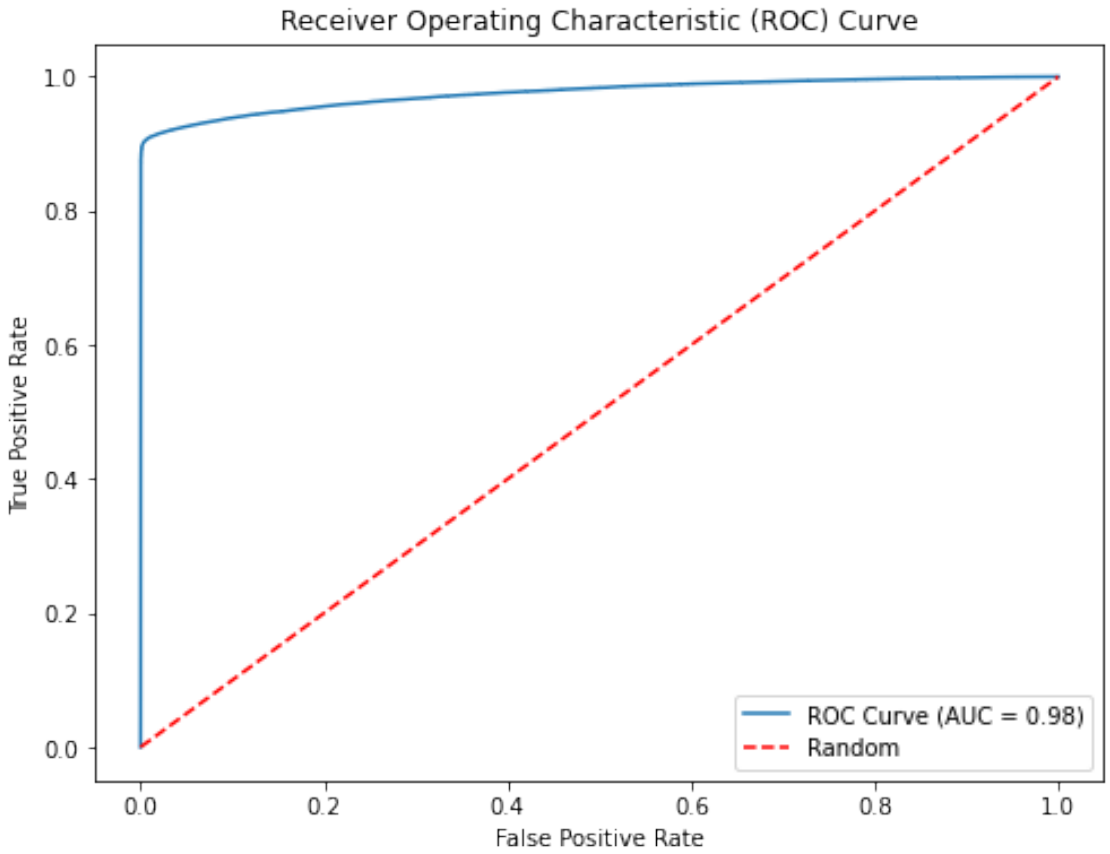
[51]:
```
pred
```

[51]:
```
array([[0.99976355],
       [0.9999459 ],
       [0.99998665],
       …,
       [0.08996144],
       [0.00360951],
       [0.99991405]], dtype=float32)
```

[52]:
```
fpr,tpr,thresholds=roc_curve(y_test,pred)
```

[53]:
```
auc=roc_auc_score(y_test,pred)
```

[54]:
```
# Plot the ROC curve
plt.figure(figsize=(8, 6))
plt.plot(fpr, tpr, label='ROC Curve (AUC = {:.2f})'.format(auc))
plt.plot([0, 1], [0, 1], linestyle='--', color='r', label='Random')
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver Operating Characteristic (ROC) Curve')
plt.legend()
plt.show()
```




[55]:
```
#from the graph ,greater the auc, better is the performance of the model
#simply means that when auc is equal to 1, the classifier is prefectly␣
↪distinguish bw positive and neg classes
```

[56]:
```
#this model has predicted with accuracy of 95% and an AUC_ROC of 0.98 on this␣
↪data
```

[ ]:
```

```