```python
# Importing all the necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
import random
pd.set_option("max_rows", None)
pd.set_option('display.max_columns', 500)

# Supress Warnings
import warnings
warnings.filterwarnings('ignore')

# Read data from file 'application_data.csv'
df= pd.read_csv("application_data.csv")

# Preview the first 5 lines of the loaded data
df.head()
```

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

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

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

             NAME_EDUCATION_TYPE    NAME_FAMILY_STATUS
NAME_HOUSING_TYPE  \
0  Secondary / secondary special  Single / not married  House /
```

```
apartment
1               Higher education               Married  House /
apartment
2  Secondary / secondary special  Single / not married  House /
apartment
3  Secondary / secondary special        Civil marriage  House /
apartment
4  Secondary / secondary special  Single / not married  House /
apartment

   REGION_POPULATION_RELATIVE  DAYS_BIRTH  DAYS_EMPLOYED
DAYS_REGISTRATION  \
0                    0.018801       -9461           -637            -
3648.0
1                    0.003541      -16765          -1188            -
1186.0
2                    0.010032      -19046           -225            -
4260.0
3                    0.008019      -19005          -3039            -
9833.0
4                    0.028663      -19932          -3038            -
4311.0

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

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

   REGION_RATING_CLIENT  REGION_RATING_CLIENT_W_CITY  \
```

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

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

   REG_REGION_NOT_LIVE_REGION  REG_REGION_NOT_WORK_REGION  \
0                           0                           0
1                           0                           0
2                           0                           0
3                           0                           0
4                           0                           0

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

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

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

```
4           NaN      0.322738           NaN             NaN
NaN

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

   ELEVATORS_AVG  ENTRANCES_AVG  FLOORSMAX_AVG  FLOORSMIN_AVG
LANDAREA_AVG  \
0           0.00         0.0690         0.0833         0.1250
0.0369
1           0.08         0.0345         0.2917         0.3333
0.0130
2            NaN            NaN            NaN            NaN
NaN
3            NaN            NaN            NaN            NaN
NaN
4            NaN            NaN            NaN            NaN
NaN

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

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

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

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

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

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

   BASEMENTAREA_MEDI  YEARS_BEGINEXPLUATATION_MEDI
YEARS_BUILD_MEDI  \
0             0.0369                        0.9722           0.6243

1             0.0529                        0.9851           0.7987

2                NaN                           NaN              NaN

3                NaN                           NaN              NaN

4                NaN                           NaN              NaN


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

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

   NONLIVINGAPARTMENTS_MEDI  NONLIVINGAREA_MEDI FONDKAPREMONT_MODE  \
```

```
0                     0.0000                0.00   reg oper account
1                     0.0039                0.01   reg oper account
2                        NaN                 NaN                NaN
3                        NaN                 NaN                NaN
4                        NaN                 NaN                NaN

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

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

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

   FLAG_DOCUMENT_2  FLAG_DOCUMENT_3  FLAG_DOCUMENT_4  FLAG_DOCUMENT_5
\
0                0                1                0                0

1                0                1                0                0

2                0                0                0                0

3                0                1                0                0
```

```
4                0                0                0                0


   FLAG_DOCUMENT_6  FLAG_DOCUMENT_7  FLAG_DOCUMENT_8  FLAG_DOCUMENT_9
\
0                0                0                0                0

1                0                0                0                0

2                0                0                0                0

3                0                0                0                0

4                0                0                1                0


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

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

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

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

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

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

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

```
#Determining the number of rows and columns
df.shape
```

```
(307511, 122)
```

```
#summary of all the numeric columns in the dataset
df.describe()
```

```
          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
```

```
#Datatypes of each column
df.info(verbose=True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 307511 entries, 0 to 307510
Data columns (total 122 columns):
 #    Column                          Dtype
---   ------                          -----
 0    SK_ID_CURR                      int64
 1    TARGET                          int64
 2    NAME_CONTRACT_TYPE              object
 3    CODE_GENDER                     object
 4    FLAG_OWN_CAR                    object
 5    FLAG_OWN_REALTY                 object
 6    CNT_CHILDREN                    int64
 7    AMT_INCOME_TOTAL                float64
 8    AMT_CREDIT                      float64
```

9	AMT_ANNUITY	float64
10	AMT_GOODS_PRICE	float64
11	NAME_TYPE_SUITE	object
12	NAME_INCOME_TYPE	object
13	NAME_EDUCATION_TYPE	object
14	NAME_FAMILY_STATUS	object
15	NAME_HOUSING_TYPE	object
16	REGION_POPULATION_RELATIVE	float64
17	DAYS_BIRTH	int64
18	DAYS_EMPLOYED	int64
19	DAYS_REGISTRATION	float64
20	DAYS_ID_PUBLISH	int64
21	OWN_CAR_AGE	float64
22	FLAG_MOBIL	int64
23	FLAG_EMP_PHONE	int64
24	FLAG_WORK_PHONE	int64
25	FLAG_CONT_MOBILE	int64
26	FLAG_PHONE	int64
27	FLAG_EMAIL	int64
28	OCCUPATION_TYPE	object
29	CNT_FAM_MEMBERS	float64
30	REGION_RATING_CLIENT	int64
31	REGION_RATING_CLIENT_W_CITY	int64
32	WEEKDAY_APPR_PROCESS_START	object
33	HOUR_APPR_PROCESS_START	int64
34	REG_REGION_NOT_LIVE_REGION	int64
35	REG_REGION_NOT_WORK_REGION	int64
36	LIVE_REGION_NOT_WORK_REGION	int64
37	REG_CITY_NOT_LIVE_CITY	int64
38	REG_CITY_NOT_WORK_CITY	int64
39	LIVE_CITY_NOT_WORK_CITY	int64
40	ORGANIZATION_TYPE	object
41	EXT_SOURCE_1	float64
42	EXT_SOURCE_2	float64
43	EXT_SOURCE_3	float64
44	APARTMENTS_AVG	float64
45	BASEMENTAREA_AVG	float64
46	YEARS_BEGINEXPLUATATION_AVG	float64
47	YEARS_BUILD_AVG	float64
48	COMMONAREA_AVG	float64
49	ELEVATORS_AVG	float64
50	ENTRANCES_AVG	float64
51	FLOORSMAX_AVG	float64
52	FLOORSMIN_AVG	float64
53	LANDAREA_AVG	float64
54	LIVINGAPARTMENTS_AVG	float64
55	LIVINGAREA_AVG	float64
56	NONLIVINGAPARTMENTS_AVG	float64
57	NONLIVINGAREA_AVG	float64
58	APARTMENTS_MODE	float64

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

```
 109  FLAG_DOCUMENT_15              int64
 110  FLAG_DOCUMENT_16              int64
 111  FLAG_DOCUMENT_17              int64
 112  FLAG_DOCUMENT_18              int64
 113  FLAG_DOCUMENT_19              int64
 114  FLAG_DOCUMENT_20              int64
 115  FLAG_DOCUMENT_21              int64
 116  AMT_REQ_CREDIT_BUREAU_HOUR    float64
 117  AMT_REQ_CREDIT_BUREAU_DAY     float64
 118  AMT_REQ_CREDIT_BUREAU_WEEK    float64
 119  AMT_REQ_CREDIT_BUREAU_MON     float64
 120  AMT_REQ_CREDIT_BUREAU_QRT     float64
 121  AMT_REQ_CREDIT_BUREAU_YEAR    float64
dtypes: float64(65), int64(41), object(16)
memory usage: 286.2+ MB
```

Checking missing values

```
#Column-wise null count
df.isnull().sum()
```

```
SK_ID_CURR                         0
TARGET                             0
NAME_CONTRACT_TYPE                 0
CODE_GENDER                        0
FLAG_OWN_CAR                       0
FLAG_OWN_REALTY                    0
CNT_CHILDREN                       0
AMT_INCOME_TOTAL                   0
AMT_CREDIT                         0
AMT_ANNUITY                       12
AMT_GOODS_PRICE                  278
NAME_TYPE_SUITE                 1292
NAME_INCOME_TYPE                   0
NAME_EDUCATION_TYPE                0
NAME_FAMILY_STATUS                 0
NAME_HOUSING_TYPE                  0
REGION_POPULATION_RELATIVE         0
DAYS_BIRTH                         0
DAYS_EMPLOYED                      0
DAYS_REGISTRATION                  0
DAYS_ID_PUBLISH                    0
OWN_CAR_AGE                   202929
FLAG_MOBIL                         0
FLAG_EMP_PHONE                     0
FLAG_WORK_PHONE                    0
FLAG_CONT_MOBILE                   0
FLAG_PHONE                         0
FLAG_EMAIL                         0
OCCUPATION_TYPE                96391
```

CNT_FAM_MEMBERS	2
REGION_RATING_CLIENT	0
REGION_RATING_CLIENT_W_CITY	0
WEEKDAY_APPR_PROCESS_START	0
HOUR_APPR_PROCESS_START	0
REG_REGION_NOT_LIVE_REGION	0
REG_REGION_NOT_WORK_REGION	0
LIVE_REGION_NOT_WORK_REGION	0
REG_CITY_NOT_LIVE_CITY	0
REG_CITY_NOT_WORK_CITY	0
LIVE_CITY_NOT_WORK_CITY	0
ORGANIZATION_TYPE	0
EXT_SOURCE_1	173378
EXT_SOURCE_2	660
EXT_SOURCE_3	60965
APARTMENTS_AVG	156061
BASEMENTAREA_AVG	179943
YEARS_BEGINEXPLUATATION_AVG	150007
YEARS_BUILD_AVG	204488
COMMONAREA_AVG	214865
ELEVATORS_AVG	163891
ENTRANCES_AVG	154828
FLOORSMAX_AVG	153020
FLOORSMIN_AVG	208642
LANDAREA_AVG	182590
LIVINGAPARTMENTS_AVG	210199
LIVINGAREA_AVG	154350
NONLIVINGAPARTMENTS_AVG	213514
NONLIVINGAREA_AVG	169682
APARTMENTS_MODE	156061
BASEMENTAREA_MODE	179943
YEARS_BEGINEXPLUATATION_MODE	150007
YEARS_BUILD_MODE	204488
COMMONAREA_MODE	214865
ELEVATORS_MODE	163891
ENTRANCES_MODE	154828
FLOORSMAX_MODE	153020
FLOORSMIN_MODE	208642
LANDAREA_MODE	182590
LIVINGAPARTMENTS_MODE	210199
LIVINGAREA_MODE	154350
NONLIVINGAPARTMENTS_MODE	213514
NONLIVINGAREA_MODE	169682
APARTMENTS_MEDI	156061
BASEMENTAREA_MEDI	179943
YEARS_BEGINEXPLUATATION_MEDI	150007
YEARS_BUILD_MEDI	204488
COMMONAREA_MEDI	214865
ELEVATORS_MEDI	163891
ENTRANCES_MEDI	154828

```
FLOORSMAX_MEDI                 153020
FLOORSMIN_MEDI                 208642
LANDAREA_MEDI                  182590
LIVINGAPARTMENTS_MEDI          210199
LIVINGAREA_MEDI                154350
NONLIVINGAPARTMENTS_MEDI       213514
NONLIVINGAREA_MEDI             169682
FONDKAPREMONT_MODE             210295
HOUSETYPE_MODE                 154297
TOTALAREA_MODE                 148431
WALLSMATERIAL_MODE             156341
EMERGENCYSTATE_MODE            145755
OBS_30_CNT_SOCIAL_CIRCLE         1021
DEF_30_CNT_SOCIAL_CIRCLE         1021
OBS_60_CNT_SOCIAL_CIRCLE         1021
DEF_60_CNT_SOCIAL_CIRCLE         1021
DAYS_LAST_PHONE_CHANGE              1
FLAG_DOCUMENT_2                     0
FLAG_DOCUMENT_3                     0
FLAG_DOCUMENT_4                     0
FLAG_DOCUMENT_5                     0
FLAG_DOCUMENT_6                     0
FLAG_DOCUMENT_7                     0
FLAG_DOCUMENT_8                     0
FLAG_DOCUMENT_9                     0
FLAG_DOCUMENT_10                    0
FLAG_DOCUMENT_11                    0
FLAG_DOCUMENT_12                    0
FLAG_DOCUMENT_13                    0
FLAG_DOCUMENT_14                    0
FLAG_DOCUMENT_15                    0
FLAG_DOCUMENT_16                    0
FLAG_DOCUMENT_17                    0
FLAG_DOCUMENT_18                    0
FLAG_DOCUMENT_19                    0
FLAG_DOCUMENT_20                    0
FLAG_DOCUMENT_21                    0
AMT_REQ_CREDIT_BUREAU_HOUR      41519
AMT_REQ_CREDIT_BUREAU_DAY       41519
AMT_REQ_CREDIT_BUREAU_WEEK      41519
AMT_REQ_CREDIT_BUREAU_MON       41519
AMT_REQ_CREDIT_BUREAU_QRT       41519
AMT_REQ_CREDIT_BUREAU_YEAR      41519
dtype: int64
```

```python
#Percentage of missing values for all columns
null_count=round(100*(df.isnull().sum()/len(df.index)),2)
null_count
```

```
SK_ID_CURR                       0.00
TARGET                           0.00
```

NAME_CONTRACT_TYPE	0.00
CODE_GENDER	0.00
FLAG_OWN_CAR	0.00
FLAG_OWN_REALTY	0.00
CNT_CHILDREN	0.00
AMT_INCOME_TOTAL	0.00
AMT_CREDIT	0.00
AMT_ANNUITY	0.00
AMT_GOODS_PRICE	0.09
NAME_TYPE_SUITE	0.42
NAME_INCOME_TYPE	0.00
NAME_EDUCATION_TYPE	0.00
NAME_FAMILY_STATUS	0.00
NAME_HOUSING_TYPE	0.00
REGION_POPULATION_RELATIVE	0.00
DAYS_BIRTH	0.00
DAYS_EMPLOYED	0.00
DAYS_REGISTRATION	0.00
DAYS_ID_PUBLISH	0.00
OWN_CAR_AGE	65.99
FLAG_MOBIL	0.00
FLAG_EMP_PHONE	0.00
FLAG_WORK_PHONE	0.00
FLAG_CONT_MOBILE	0.00
FLAG_PHONE	0.00
FLAG_EMAIL	0.00
OCCUPATION_TYPE	31.35
CNT_FAM_MEMBERS	0.00
REGION_RATING_CLIENT	0.00
REGION_RATING_CLIENT_W_CITY	0.00
WEEKDAY_APPR_PROCESS_START	0.00
HOUR_APPR_PROCESS_START	0.00
REG_REGION_NOT_LIVE_REGION	0.00
REG_REGION_NOT_WORK_REGION	0.00
LIVE_REGION_NOT_WORK_REGION	0.00
REG_CITY_NOT_LIVE_CITY	0.00
REG_CITY_NOT_WORK_CITY	0.00
LIVE_CITY_NOT_WORK_CITY	0.00
ORGANIZATION_TYPE	0.00
EXT_SOURCE_1	56.38
EXT_SOURCE_2	0.21
EXT_SOURCE_3	19.83
APARTMENTS_AVG	50.75
BASEMENTAREA_AVG	58.52
YEARS_BEGINEXPLUATATION_AVG	48.78
YEARS_BUILD_AVG	66.50
COMMONAREA_AVG	69.87
ELEVATORS_AVG	53.30
ENTRANCES_AVG	50.35
FLOORSMAX_AVG	49.76

FLOORSMIN_AVG	67.85
LANDAREA_AVG	59.38
LIVINGAPARTMENTS_AVG	68.35
LIVINGAREA_AVG	50.19
NONLIVINGAPARTMENTS_AVG	69.43
NONLIVINGAREA_AVG	55.18
APARTMENTS_MODE	50.75
BASEMENTAREA_MODE	58.52
YEARS_BEGINEXPLUATATION_MODE	48.78
YEARS_BUILD_MODE	66.50
COMMONAREA_MODE	69.87
ELEVATORS_MODE	53.30
ENTRANCES_MODE	50.35
FLOORSMAX_MODE	49.76
FLOORSMIN_MODE	67.85
LANDAREA_MODE	59.38
LIVINGAPARTMENTS_MODE	68.35
LIVINGAREA_MODE	50.19
NONLIVINGAPARTMENTS_MODE	69.43
NONLIVINGAREA_MODE	55.18
APARTMENTS_MEDI	50.75
BASEMENTAREA_MEDI	58.52
YEARS_BEGINEXPLUATATION_MEDI	48.78
YEARS_BUILD_MEDI	66.50
COMMONAREA_MEDI	69.87
ELEVATORS_MEDI	53.30
ENTRANCES_MEDI	50.35
FLOORSMAX_MEDI	49.76
FLOORSMIN_MEDI	67.85
LANDAREA_MEDI	59.38
LIVINGAPARTMENTS_MEDI	68.35
LIVINGAREA_MEDI	50.19
NONLIVINGAPARTMENTS_MEDI	69.43
NONLIVINGAREA_MEDI	55.18
FONDKAPREMONT_MODE	68.39
HOUSETYPE_MODE	50.18
TOTALAREA_MODE	48.27
WALLSMATERIAL_MODE	50.84
EMERGENCYSTATE_MODE	47.40
OBS_30_CNT_SOCIAL_CIRCLE	0.33
DEF_30_CNT_SOCIAL_CIRCLE	0.33
OBS_60_CNT_SOCIAL_CIRCLE	0.33
DEF_60_CNT_SOCIAL_CIRCLE	0.33
DAYS_LAST_PHONE_CHANGE	0.00
FLAG_DOCUMENT_2	0.00
FLAG_DOCUMENT_3	0.00
FLAG_DOCUMENT_4	0.00
FLAG_DOCUMENT_5	0.00
FLAG_DOCUMENT_6	0.00
FLAG_DOCUMENT_7	0.00

```
FLAG_DOCUMENT_8                  0.00
FLAG_DOCUMENT_9                  0.00
FLAG_DOCUMENT_10                 0.00
FLAG_DOCUMENT_11                 0.00
FLAG_DOCUMENT_12                 0.00
FLAG_DOCUMENT_13                 0.00
FLAG_DOCUMENT_14                 0.00
FLAG_DOCUMENT_15                 0.00
FLAG_DOCUMENT_16                 0.00
FLAG_DOCUMENT_17                 0.00
FLAG_DOCUMENT_18                 0.00
FLAG_DOCUMENT_19                 0.00
FLAG_DOCUMENT_20                 0.00
FLAG_DOCUMENT_21                 0.00
AMT_REQ_CREDIT_BUREAU_HOUR      13.50
AMT_REQ_CREDIT_BUREAU_DAY       13.50
AMT_REQ_CREDIT_BUREAU_WEEK      13.50
AMT_REQ_CREDIT_BUREAU_MON       13.50
AMT_REQ_CREDIT_BUREAU_QRT       13.50
AMT_REQ_CREDIT_BUREAU_YEAR      13.50
dtype: float64
```

```
#Columns with high missing percentage >=50%
null_count=null_count[null_count>=50]
null_count
```

```
OWN_CAR_AGE                 65.99
EXT_SOURCE_1                56.38
APARTMENTS_AVG              50.75
BASEMENTAREA_AVG            58.52
YEARS_BUILD_AVG             66.50
COMMONAREA_AVG              69.87
ELEVATORS_AVG               53.30
ENTRANCES_AVG               50.35
FLOORSMIN_AVG               67.85
LANDAREA_AVG                59.38
LIVINGAPARTMENTS_AVG        68.35
LIVINGAREA_AVG              50.19
NONLIVINGAPARTMENTS_AVG     69.43
NONLIVINGAREA_AVG           55.18
APARTMENTS_MODE             50.75
BASEMENTAREA_MODE           58.52
YEARS_BUILD_MODE            66.50
COMMONAREA_MODE             69.87
ELEVATORS_MODE              53.30
ENTRANCES_MODE              50.35
FLOORSMIN_MODE              67.85
LANDAREA_MODE               59.38
LIVINGAPARTMENTS_MODE       68.35
LIVINGAREA_MODE             50.19
NONLIVINGAPARTMENTS_MODE    69.43
```

```
NONLIVINGAREA_MODE          55.18
APARTMENTS_MEDI             50.75
BASEMENTAREA_MEDI           58.52
YEARS_BUILD_MEDI            66.50
COMMONAREA_MEDI             69.87
ELEVATORS_MEDI              53.30
ENTRANCES_MEDI              50.35
FLOORSMIN_MEDI              67.85
LANDAREA_MEDI               59.38
LIVINGAPARTMENTS_MEDI       68.35
LIVINGAREA_MEDI             50.19
NONLIVINGAPARTMENTS_MEDI    69.43
NONLIVINGAREA_MEDI          55.18
FONDKAPREMONT_MODE          68.39
HOUSETYPE_MODE              50.18
WALLSMATERIAL_MODE          50.84
dtype: float64
```

```
len(null_count)
```

```
41
```

Thus, here we can see that there are 41 columns having the highest missing percentage(>=50%) So here we remove those columns

Remove missing values

```
#Removing those 41 columns which have missing values more than 50%
cols_dropped=([null_count[null_count >=50.00].index])
df.drop(labels=cols_dropped[0],axis=1,inplace=True)
```

```
#Rows and Columns after removing columns with high missing percentage
df.shape
```

```
(307511, 81)
```

```
# Checking the columns after removing high missing values(>=50%)
100*(df.isnull().sum()/len(df.index))
```

```
SK_ID_CURR                        0.000000
TARGET                            0.000000
NAME_CONTRACT_TYPE                0.000000
CODE_GENDER                       0.000000
FLAG_OWN_CAR                      0.000000
FLAG_OWN_REALTY                   0.000000
CNT_CHILDREN                      0.000000
AMT_INCOME_TOTAL                  0.000000
AMT_CREDIT                        0.000000
AMT_ANNUITY                       0.003902
AMT_GOODS_PRICE                   0.090403
NAME_TYPE_SUITE                   0.420148
```

```
NAME_INCOME_TYPE                  0.000000
NAME_EDUCATION_TYPE               0.000000
NAME_FAMILY_STATUS                0.000000
NAME_HOUSING_TYPE                 0.000000
REGION_POPULATION_RELATIVE        0.000000
DAYS_BIRTH                        0.000000
DAYS_EMPLOYED                     0.000000
DAYS_REGISTRATION                 0.000000
DAYS_ID_PUBLISH                   0.000000
FLAG_MOBIL                        0.000000
FLAG_EMP_PHONE                    0.000000
FLAG_WORK_PHONE                   0.000000
FLAG_CONT_MOBILE                  0.000000
FLAG_PHONE                        0.000000
FLAG_EMAIL                        0.000000
OCCUPATION_TYPE                  31.345545
CNT_FAM_MEMBERS                   0.000650
REGION_RATING_CLIENT              0.000000
REGION_RATING_CLIENT_W_CITY       0.000000
WEEKDAY_APPR_PROCESS_START        0.000000
HOUR_APPR_PROCESS_START           0.000000
REG_REGION_NOT_LIVE_REGION        0.000000
REG_REGION_NOT_WORK_REGION        0.000000
LIVE_REGION_NOT_WORK_REGION       0.000000
REG_CITY_NOT_LIVE_CITY            0.000000
REG_CITY_NOT_WORK_CITY            0.000000
LIVE_CITY_NOT_WORK_CITY           0.000000
ORGANIZATION_TYPE                 0.000000
EXT_SOURCE_2                      0.214626
EXT_SOURCE_3                     19.825307
YEARS_BEGINEXPLUATATION_AVG      48.781019
FLOORSMAX_AVG                    49.760822
YEARS_BEGINEXPLUATATION_MODE     48.781019
FLOORSMAX_MODE                   49.760822
YEARS_BEGINEXPLUATATION_MEDI     48.781019
FLOORSMAX_MEDI                   49.760822
TOTALAREA_MODE                   48.268517
EMERGENCYSTATE_MODE              47.398304
OBS_30_CNT_SOCIAL_CIRCLE          0.332021
DEF_30_CNT_SOCIAL_CIRCLE          0.332021
OBS_60_CNT_SOCIAL_CIRCLE          0.332021
DEF_60_CNT_SOCIAL_CIRCLE          0.332021
DAYS_LAST_PHONE_CHANGE            0.000325
FLAG_DOCUMENT_2                   0.000000
FLAG_DOCUMENT_3                   0.000000
FLAG_DOCUMENT_4                   0.000000
FLAG_DOCUMENT_5                   0.000000
FLAG_DOCUMENT_6                   0.000000
FLAG_DOCUMENT_7                   0.000000
FLAG_DOCUMENT_8                   0.000000
```

```
FLAG_DOCUMENT_9                  0.000000
FLAG_DOCUMENT_10                 0.000000
FLAG_DOCUMENT_11                 0.000000
FLAG_DOCUMENT_12                 0.000000
FLAG_DOCUMENT_13                 0.000000
FLAG_DOCUMENT_14                 0.000000
FLAG_DOCUMENT_15                 0.000000
FLAG_DOCUMENT_16                 0.000000
FLAG_DOCUMENT_17                 0.000000
FLAG_DOCUMENT_18                 0.000000
FLAG_DOCUMENT_19                 0.000000
FLAG_DOCUMENT_20                 0.000000
FLAG_DOCUMENT_21                 0.000000
AMT_REQ_CREDIT_BUREAU_HOUR      13.501631
AMT_REQ_CREDIT_BUREAU_DAY       13.501631
AMT_REQ_CREDIT_BUREAU_WEEK      13.501631
AMT_REQ_CREDIT_BUREAU_MON       13.501631
AMT_REQ_CREDIT_BUREAU_QRT       13.501631
AMT_REQ_CREDIT_BUREAU_YEAR      13.501631
dtype: float64
```

```
# Drop unnecessary columns from the dataset

Drop_col=['FLAG_MOBIL', 'FLAG_EMP_PHONE', 'FLAG_WORK_PHONE',
'FLAG_CONT_MOBILE','YEARS_BEGINEXPLUATATION_MODE','FLOORSMAX_MODE','TO
TALAREA_MODE','EMERGENCYSTATE_MODE',
       'FLAG_PHONE',
'FLAG_EMAIL','REGION_RATING_CLIENT','REGION_RATING_CLIENT_W_CITY','FLA
G_EMAIL','CNT_FAM_MEMBERS','REGION_RATING_CLIENT',
       'REGION_RATING_CLIENT_W_CITY','DAYS_LAST_PHONE_CHANGE',
'FLAG_DOCUMENT_2', 'FLAG_DOCUMENT_3','FLAG_DOCUMENT_4',
'FLAG_DOCUMENT_5', 'FLAG_DOCUMENT_6',
       'FLAG_DOCUMENT_7', 'FLAG_DOCUMENT_8',
'FLAG_DOCUMENT_9','FLAG_DOCUMENT_10', 'FLAG_DOCUMENT_11',
'FLAG_DOCUMENT_12',
       'FLAG_DOCUMENT_13', 'FLAG_DOCUMENT_14',
'FLAG_DOCUMENT_15','FLAG_DOCUMENT_16', 'FLAG_DOCUMENT_17',
'FLAG_DOCUMENT_18',
       'FLAG_DOCUMENT_19', 'FLAG_DOCUMENT_20',
'FLAG_DOCUMENT_21','EXT_SOURCE_3','YEARS_BEGINEXPLUATATION_AVG','FLOOR
SMAX_AVG','YEARS_BEGINEXPLUATATION_MEDI','FLOORSMAX_MEDI']

df.drop(labels=Drop_col,axis=1,inplace=True)

# Checking the columns after dropping unncessary columns

100*(df.isnull().sum()/len(df.index))
```

```
SK_ID_CURR                      0.000000
TARGET                          0.000000
```

```
NAME_CONTRACT_TYPE              0.000000
CODE_GENDER                     0.000000
FLAG_OWN_CAR                    0.000000
FLAG_OWN_REALTY                 0.000000
CNT_CHILDREN                    0.000000
AMT_INCOME_TOTAL                0.000000
AMT_CREDIT                      0.000000
AMT_ANNUITY                     0.003902
AMT_GOODS_PRICE                 0.090403
NAME_TYPE_SUITE                 0.420148
NAME_INCOME_TYPE                0.000000
NAME_EDUCATION_TYPE             0.000000
NAME_FAMILY_STATUS              0.000000
NAME_HOUSING_TYPE               0.000000
REGION_POPULATION_RELATIVE      0.000000
DAYS_BIRTH                      0.000000
DAYS_EMPLOYED                   0.000000
DAYS_REGISTRATION               0.000000
DAYS_ID_PUBLISH                 0.000000
OCCUPATION_TYPE                31.345545
WEEKDAY_APPR_PROCESS_START      0.000000
HOUR_APPR_PROCESS_START         0.000000
REG_REGION_NOT_LIVE_REGION      0.000000
REG_REGION_NOT_WORK_REGION      0.000000
LIVE_REGION_NOT_WORK_REGION     0.000000
REG_CITY_NOT_LIVE_CITY          0.000000
REG_CITY_NOT_WORK_CITY          0.000000
LIVE_CITY_NOT_WORK_CITY         0.000000
ORGANIZATION_TYPE               0.000000
EXT_SOURCE_2                    0.214626
OBS_30_CNT_SOCIAL_CIRCLE        0.332021
DEF_30_CNT_SOCIAL_CIRCLE        0.332021
OBS_60_CNT_SOCIAL_CIRCLE        0.332021
DEF_60_CNT_SOCIAL_CIRCLE        0.332021
AMT_REQ_CREDIT_BUREAU_HOUR     13.501631
AMT_REQ_CREDIT_BUREAU_DAY      13.501631
AMT_REQ_CREDIT_BUREAU_WEEK     13.501631
AMT_REQ_CREDIT_BUREAU_MON      13.501631
AMT_REQ_CREDIT_BUREAU_QRT      13.501631
AMT_REQ_CREDIT_BUREAU_YEAR     13.501631
dtype: float64
```

AMT_ANNUITY Variable

```
#Missing values in "AMT_ANNUITY" column
df.AMT_ANNUITY.isnull().sum()
```

```
12
```

```
#Percentage of missing values in "AMT_ANNUITY" column
float(100*(12/307511))
```

```
0.003902299429939092
```

We can see that 'AMT_ANNUITY' column is having very less percentage of null values and is also a numeric data. These values can be imputed by the mean of the complete cases of the variable. Since this column is also having an outlier which is very large we can impute missing values with Median.

AMT GOODS PRICE variable

```
#Missing values in "AMT_GOODS_PRICE" column
df.AMT_GOODS_PRICE.isnull().sum()
```

```
278
```

```
#Percentage of missing values in "AMT_GOODS_PRICE" column
float(100*(278/307511))
```

```
0.09040327012692229
```

AMT_GOODS_PRICE has less percentage of missing values. And also has outliers. These missing values can be imputed with median

EXT SOURCE 2 variable

```
df.EXT_SOURCE_2.isnull().sum()
```

```
660
```

```
float(100*(660/307511))
```

```
0.21462646864665005
```

EXT_SOURCE_2 column has less percentage of missing values. Here we can impute the missing values with 0

NAME TYPE SUITE Variable

```
#Missing values in "NAME_TYPE_SUITE" column
df.NAME_TYPE_SUITE.isnull().sum()
```

```
1292
```

```
#Percentage of missing values in "NAME_TYPE_SUITE" column
100*(1292/307511)
```

```
0.42014757195677555
```

```
#print the percentage of each NAME_TYPE_SUITE in the data frame df.
df.NAME_TYPE_SUITE.value_counts(normalize= True)
```

```
Unaccompanied      0.811596
Family             0.131112
Spouse, partner    0.037130
Children           0.010669
Other_B            0.005780
Other_A            0.002828
Group of people    0.000885
Name: NAME_TYPE_SUITE, dtype: float64
```

```
#find the mode of NAME_TYPE_SUITE in df, check which category is most
repeated
name_mode=df.NAME_TYPE_SUITE.mode()[0]
name_mode
```

```
'Unaccompanied'
```

Hence here 'NAME_TYPE_SUITE' being a categorical variable, has about 0.42% of missing values. So here we can impute the missing values with the most popular category which is "Unaccompanied"

OCCUPATION TYPE variable

```
#Missing values in "OCCUPATION_TYPE" column
df.OCCUPATION_TYPE.isnull().sum()
```

```
96391
```

```
#Percentage of missing values in "OCCUPATION_TYPE" column
100*(96391/307511)
```

```
31.345545362604916
```

```
#print the percentage of each OCCUPATION TYPES in the data frame df
df.OCCUPATION_TYPE .value_counts(normalize=True)
```

```
Laborers                 0.261396
Sales staff              0.152056
Core staff               0.130589
Managers                 0.101227
Drivers                  0.088116
High skill tech staff    0.053903
Accountants              0.046481
Medicine staff           0.040437
Security staff           0.031835
Cooking staff            0.028164
Cleaning staff           0.022040
Private service staff    0.012562
Low-skill Laborers       0.009914
Waiters/barmen staff     0.006385
Secretaries              0.006181
Realty agents            0.003557
```

```
HR staff                 0.002667
IT staff                 0.002491
Name: OCCUPATION_TYPE, dtype: float64
```

```
#find the mode of NAME_TYPE_SUITE in df, check which category is most
repeated
occ_mode=df.OCCUPATION_TYPE.mode()[0]
occ_mode
```

```
'Laborers'
```

Here 'OCCUPATION_TYPE' is also a categorical variable, has about 31.3% of missing values. So here we can impute the missing values with the most popular category which is "Laborers"

Checking Datatypes

```
#Preview the first 5 lines of the data
df.head()
```

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

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

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

             NAME_EDUCATION_TYPE    NAME_FAMILY_STATUS
NAME_HOUSING_TYPE  \
0  Secondary / secondary special  Single / not married  House /
```

```
apartment
1                Higher education               Married  House /
apartment
2  Secondary / secondary special  Single / not married  House /
apartment
3  Secondary / secondary special        Civil marriage  House /
apartment
4  Secondary / secondary special  Single / not married  House /
apartment

   REGION_POPULATION_RELATIVE  DAYS_BIRTH  DAYS_EMPLOYED
DAYS_REGISTRATION  \
0                    0.018801       -9461           -637            -
3648.0
1                    0.003541      -16765          -1188            -
1186.0
2                    0.010032      -19046           -225            -
4260.0
3                    0.008019      -19005          -3039            -
9833.0
4                    0.028663      -19932          -3038            -
4311.0

   DAYS_ID_PUBLISH OCCUPATION_TYPE WEEKDAY_APPR_PROCESS_START  \
0            -2120        Laborers                  WEDNESDAY
1             -291      Core staff                     MONDAY
2            -2531        Laborers                     MONDAY
3            -2437        Laborers                  WEDNESDAY
4            -3458      Core staff                   THURSDAY

   HOUR_APPR_PROCESS_START  REG_REGION_NOT_LIVE_REGION  \
0                       10                           0
1                       11                           0
2                        9                           0
3                       17                           0
4                       11                           0

   REG_REGION_NOT_WORK_REGION  LIVE_REGION_NOT_WORK_REGION  \
0                           0                            0
1                           0                            0
2                           0                            0
3                           0                            0
4                           0                            0

   REG_CITY_NOT_LIVE_CITY  REG_CITY_NOT_WORK_CITY
LIVE_CITY_NOT_WORK_CITY  \
0                       0                       0
0
1                       0                       0
0
```

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

       ORGANIZATION_TYPE  EXT_SOURCE_2  OBS_30_CNT_SOCIAL_CIRCLE  \
0  Business Entity Type 3      0.262949                       2.0
1                  School      0.622246                       1.0
2              Government      0.555912                       0.0
3  Business Entity Type 3      0.650442                       2.0
4                Religion      0.322738                       0.0

   DEF_30_CNT_SOCIAL_CIRCLE  OBS_60_CNT_SOCIAL_CIRCLE  \
0                       2.0                       2.0
1                       0.0                       1.0
2                       0.0                       0.0
3                       0.0                       2.0
4                       0.0                       0.0

   DEF_60_CNT_SOCIAL_CIRCLE  AMT_REQ_CREDIT_BUREAU_HOUR  \
0                       2.0                         0.0
1                       0.0                         0.0
2                       0.0                         0.0
3                       0.0                         NaN
4                       0.0                         0.0

   AMT_REQ_CREDIT_BUREAU_DAY  AMT_REQ_CREDIT_BUREAU_WEEK  \
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_MON  AMT_REQ_CREDIT_BUREAU_QRT  \
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_YEAR
0                         1.0
1                         0.0
2                         0.0
3                         NaN
4                         0.0
```

```python
# I can check the number of unique values is a column
# If the number of unique values <=40: Categorical column
# If the number of unique values in a columns> 50: Continuous

df.nunique().sort_values()
```

```
LIVE_REGION_NOT_WORK_REGION          2
TARGET                               2
NAME_CONTRACT_TYPE                   2
FLAG_OWN_CAR                         2
FLAG_OWN_REALTY                      2
REG_REGION_NOT_LIVE_REGION           2
LIVE_CITY_NOT_WORK_CITY              2
REG_CITY_NOT_LIVE_CITY               2
REG_CITY_NOT_WORK_CITY               2
REG_REGION_NOT_WORK_REGION           2
CODE_GENDER                          3
AMT_REQ_CREDIT_BUREAU_HOUR           5
NAME_EDUCATION_TYPE                  5
NAME_FAMILY_STATUS                   6
NAME_HOUSING_TYPE                    6
NAME_TYPE_SUITE                      7
WEEKDAY_APPR_PROCESS_START           7
NAME_INCOME_TYPE                     8
AMT_REQ_CREDIT_BUREAU_DAY            9
DEF_60_CNT_SOCIAL_CIRCLE             9
AMT_REQ_CREDIT_BUREAU_WEEK           9
DEF_30_CNT_SOCIAL_CIRCLE            10
AMT_REQ_CREDIT_BUREAU_QRT           11
CNT_CHILDREN                        15
OCCUPATION_TYPE                     18
AMT_REQ_CREDIT_BUREAU_MON           24
HOUR_APPR_PROCESS_START             24
AMT_REQ_CREDIT_BUREAU_YEAR          25
OBS_30_CNT_SOCIAL_CIRCLE            33
OBS_60_CNT_SOCIAL_CIRCLE            33
ORGANIZATION_TYPE                   58
REGION_POPULATION_RELATIVE          81
AMT_GOODS_PRICE                   1002
AMT_INCOME_TOTAL                  2548
AMT_CREDIT                        5603
DAYS_ID_PUBLISH                   6168
DAYS_EMPLOYED                    12574
AMT_ANNUITY                      13672
DAYS_REGISTRATION                15688
DAYS_BIRTH                       17460
EXT_SOURCE_2                    119831
SK_ID_CURR                      307511
dtype: int64
```

Now it is clear that which are Continuous and Categorical variables in the dataset givenn. Now we can consider some continuous and categorical variables and change the datatypes if needed

```
#changing negative ages to positive ages.
df['DAYS_BIRTH']=abs(df['DAYS_BIRTH'])
df['DAYS_BIRTH'].describe()
```

```
count    307511.000000
mean      16036.995067
std        4363.988632
min        7489.000000
25%       12413.000000
50%       15750.000000
75%       19682.000000
max       25229.000000
Name: DAYS_BIRTH, dtype: float64
```

```
#changing negative values in days to positive days
df['DAYS_EMPLOYED']=abs(df['DAYS_EMPLOYED'])
df['DAYS_EMPLOYED'].describe()
```

```
count    307511.000000
mean      67724.742149
std      139443.751806
min           0.000000
25%         933.000000
50%        2219.000000
75%        5707.000000
max      365243.000000
Name: DAYS_EMPLOYED, dtype: float64
```

```
#changing negative days to positive days.
df['DAYS_REGISTRATION']=abs(df['DAYS_REGISTRATION'])
df['DAYS_REGISTRATION'].describe()
```

```
count    307511.000000
mean       4986.120328
std        3522.886321
min           0.000000
25%        2010.000000
50%        4504.000000
75%        7479.500000
max       24672.000000
Name: DAYS_REGISTRATION, dtype: float64
```

```
#changing negative days to positive
df['DAYS_ID_PUBLISH']=abs(df['DAYS_ID_PUBLISH'])
df['DAYS_ID_PUBLISH'].describe()
```

```
count    307511.000000
mean       2994.202373
```

```
std        1509.450419
min           0.000000
25%        1720.000000
50%        3254.000000
75%        4299.000000
max        7197.000000
Name: DAYS_ID_PUBLISH, dtype: float64
```

```
#converting the data type of categorical column
df['REG_REGION_NOT_LIVE_REGION'] =
df['REG_REGION_NOT_LIVE_REGION'].astype(object)
df.dtypes
```

```
SK_ID_CURR                        int64
TARGET                            int64
NAME_CONTRACT_TYPE               object
CODE_GENDER                      object
FLAG_OWN_CAR                     object
FLAG_OWN_REALTY                  object
CNT_CHILDREN                      int64
AMT_INCOME_TOTAL                float64
AMT_CREDIT                      float64
AMT_ANNUITY                     float64
AMT_GOODS_PRICE                 float64
NAME_TYPE_SUITE                  object
NAME_INCOME_TYPE                 object
NAME_EDUCATION_TYPE              object
NAME_FAMILY_STATUS               object
NAME_HOUSING_TYPE                object
REGION_POPULATION_RELATIVE      float64
DAYS_BIRTH                        int64
DAYS_EMPLOYED                     int64
DAYS_REGISTRATION               float64
DAYS_ID_PUBLISH                   int64
OCCUPATION_TYPE                  object
WEEKDAY_APPR_PROCESS_START       object
HOUR_APPR_PROCESS_START           int64
REG_REGION_NOT_LIVE_REGION       object
REG_REGION_NOT_WORK_REGION        int64
LIVE_REGION_NOT_WORK_REGION       int64
REG_CITY_NOT_LIVE_CITY            int64
REG_CITY_NOT_WORK_CITY            int64
LIVE_CITY_NOT_WORK_CITY           int64
ORGANIZATION_TYPE                object
EXT_SOURCE_2                    float64
OBS_30_CNT_SOCIAL_CIRCLE        float64
DEF_30_CNT_SOCIAL_CIRCLE        float64
OBS_60_CNT_SOCIAL_CIRCLE        float64
DEF_60_CNT_SOCIAL_CIRCLE        float64
AMT_REQ_CREDIT_BUREAU_HOUR      float64
AMT_REQ_CREDIT_BUREAU_DAY       float64
```

```
AMT_REQ_CREDIT_BUREAU_WEEK     float64
AMT_REQ_CREDIT_BUREAU_MON      float64
AMT_REQ_CREDIT_BUREAU_QRT      float64
AMT_REQ_CREDIT_BUREAU_YEAR     float64
dtype: object
```

```
#Changing region from int to object
df['REG_REGION_NOT_WORK_REGION'] =
df['REG_REGION_NOT_WORK_REGION'].astype(object)

#Changing region from int to object
df['LIVE_REGION_NOT_WORK_REGION'] =
df['LIVE_REGION_NOT_WORK_REGION'].astype(object)

#Changing city from int to object
df['REG_CITY_NOT_LIVE_CITY'] =
df['REG_CITY_NOT_LIVE_CITY'].astype(object)

#Changing city from int to object
df['REG_CITY_NOT_WORK_CITY'] =
df['REG_CITY_NOT_WORK_CITY'].astype(object)

#Changing city from int to object
df['LIVE_CITY_NOT_WORK_CITY']=df['LIVE_CITY_NOT_WORK_CITY'].astype(obj
ect)

df.head()
```

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

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

   AMT_GOODS_PRICE NAME_TYPE_SUITE NAME_INCOME_TYPE  \
0         351000.0   Unaccompanied          Working
1        1129500.0          Family    State servant
```

```
2         135000.0   Unaccompanied          Working
3         297000.0   Unaccompanied          Working
4         513000.0   Unaccompanied          Working

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

   REGION_POPULATION_RELATIVE  DAYS_BIRTH  DAYS_EMPLOYED
DAYS_REGISTRATION  \
0                    0.018801        9461            637
3648.0
1                    0.003541       16765           1188
1186.0
2                    0.010032       19046            225
4260.0
3                    0.008019       19005           3039
9833.0
4                    0.028663       19932           3038
4311.0

   DAYS_ID_PUBLISH OCCUPATION_TYPE WEEKDAY_APPR_PROCESS_START  \
0             2120        Laborers                  WEDNESDAY
1              291      Core staff                     MONDAY
2             2531        Laborers                     MONDAY
3             2437        Laborers                  WEDNESDAY
4             3458      Core staff                   THURSDAY

   HOUR_APPR_PROCESS_START REG_REGION_NOT_LIVE_REGION  \
0                       10                          0
1                       11                          0
2                        9                          0
3                       17                          0
4                       11                          0

  REG_REGION_NOT_WORK_REGION LIVE_REGION_NOT_WORK_REGION  \
0                          0                           0
1                          0                           0
2                          0                           0
3                          0                           0
4                          0                           0
```

```
  REG_CITY_NOT_LIVE_CITY REG_CITY_NOT_WORK_CITY
LIVE_CITY_NOT_WORK_CITY  \
0                      0                      0
0
1                      0                      0
0
2                      0                      0
0
3                      0                      0
0
4                      0                      1
1

        ORGANIZATION_TYPE  EXT_SOURCE_2  OBS_30_CNT_SOCIAL_CIRCLE  \
0  Business Entity Type 3      0.262949                       2.0
1                  School      0.622246                       1.0
2              Government      0.555912                       0.0
3  Business Entity Type 3      0.650442                       2.0
4                Religion      0.322738                       0.0

   DEF_30_CNT_SOCIAL_CIRCLE  OBS_60_CNT_SOCIAL_CIRCLE  \
0                       2.0                       2.0
1                       0.0                       1.0
2                       0.0                       0.0
3                       0.0                       2.0
4                       0.0                       0.0

   DEF_60_CNT_SOCIAL_CIRCLE  AMT_REQ_CREDIT_BUREAU_HOUR  \
0                       2.0                         0.0
1                       0.0                         0.0
2                       0.0                         0.0
3                       0.0                         NaN
4                       0.0                         0.0

   AMT_REQ_CREDIT_BUREAU_DAY  AMT_REQ_CREDIT_BUREAU_WEEK  \
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_MON  AMT_REQ_CREDIT_BUREAU_QRT  \
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_YEAR
0                         1.0
1                         0.0
2                         0.0
3                         NaN
4                         0.0
```

Handling Outliers

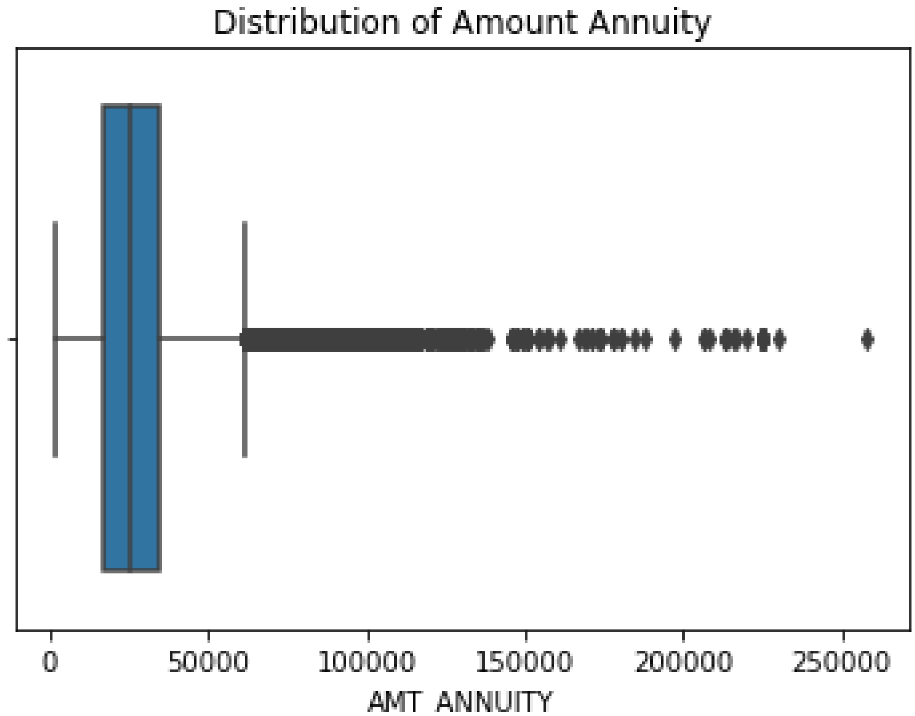
Major approaches to the treat outliers:

a)Imputation b)Deletion of outliers c)Binning of values d)Cap the outlier

```
#describe the AMT_ANNUITY variable of df

df.AMT_ANNUITY.describe()
```

```
count    307499.000000
mean      27108.573909
std       14493.737315
min        1615.500000
25%       16524.000000
50%       24903.000000
75%       34596.000000
max      258025.500000
Name: AMT_ANNUITY, dtype: float64
```

```
sns.boxplot(df.AMT_ANNUITY)
plt.title('Distribution of Amount Annuity')
plt.show()
```

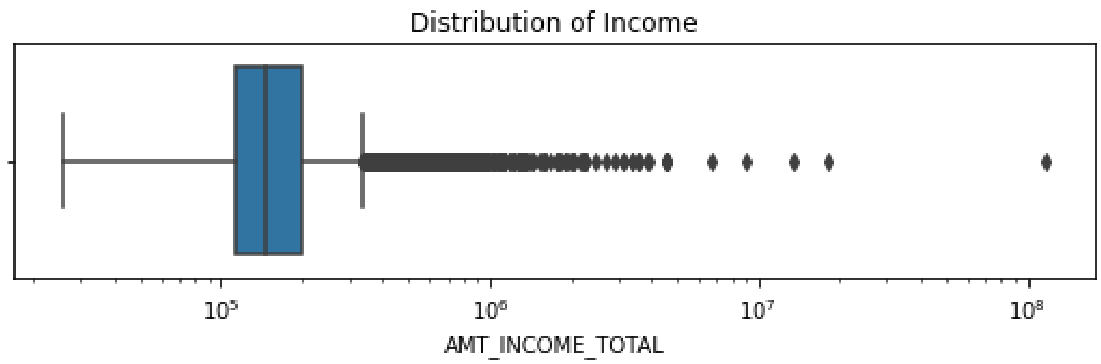


As we take a look at AMT_ANNUITY column we can see that there are outliers at 258025. But there is no much differece between the mean and median, We can impute the outliers with Median here

```
df.AMT_INCOME_TOTAL.describe()
```

```
count    3.075110e+05
mean     1.687979e+05
std      2.371231e+05
min      2.565000e+04
25%      1.125000e+05
50%      1.471500e+05
75%      2.025000e+05
max      1.170000e+08
Name: AMT_INCOME_TOTAL, dtype: float64
```

```
plt.figure(figsize=(9,2))
sns.boxplot(df.AMT_INCOME_TOTAL)
plt.xscale('log')
plt.title('Distribution of Income')
plt.show()
```




```
df.AMT_INCOME_TOTAL.quantile([0.5, 0.7, 0.9,0.95,0.99])
```

```
0.50    147150.0
0.70    180000.0
0.90    270000.0
0.95    337500.0
0.99    472500.0
Name: AMT_INCOME_TOTAL, dtype: float64
```

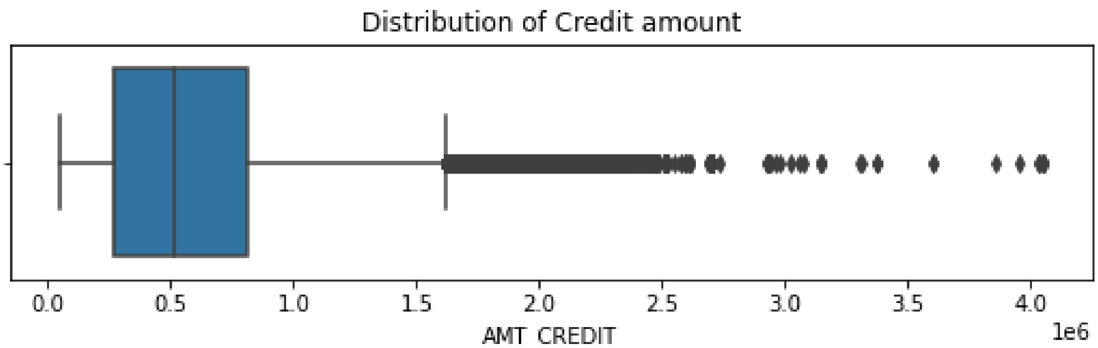
In 'AMT_INCOME_TOTAL' column, We can see that there are outlier values at 1.17x10^8. Sometimes, it is beneficial to look into the quantiles instead of the box plot, mean or median. Quantile may give you a fair idea about the outliers. If there is a huge difference between the maximum value and the 95th or 99th quantiles, then there are outliers in the data set.

Total income will definitely vary from person to person. We can cap the outliers here

```
df.AMT_CREDIT.describe()
```

```
count    3.075110e+05
mean     5.990260e+05
std      4.024908e+05
min      4.500000e+04
25%      2.700000e+05
50%      5.135310e+05
75%      8.086500e+05
max      4.050000e+06
Name: AMT_CREDIT, dtype: float64
```

```
plt.figure(figsize=(9,2))
sns.boxplot(df.AMT_CREDIT)
plt.title('Distribution of Credit amount')
plt.show()
```




```
df.AMT_CREDIT.quantile([0.5, 0.7, 0.9,0.95,0.99])
```

```
0.50     513531.0
0.70     755190.0
0.90    1133748.0
0.95    1350000.0
0.99    1854000.0
Name: AMT_CREDIT, dtype: float64
```

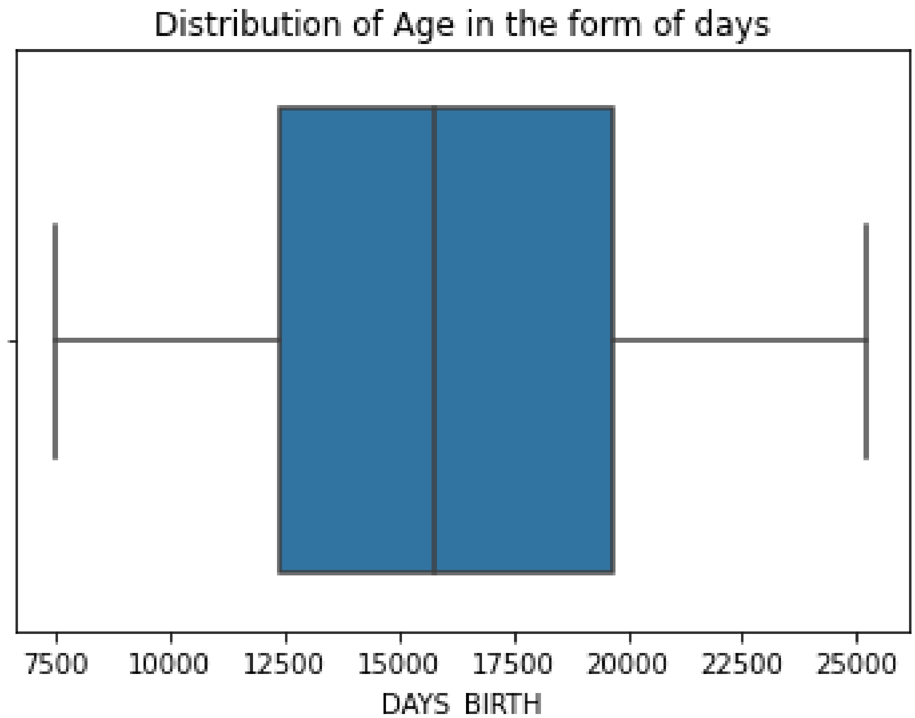
In this AMT_CREDIT column we can see the outliers after 99th quantile at 4.05x10^6 Amount credited also varies from person to person.

DAYS BIRTH variable

```
df.DAYS_BIRTH.describe()
```

```
count    307511.000000
mean      16036.995067
std        4363.988632
min        7489.000000
25%       12413.000000
50%       15750.000000
75%       19682.000000
max       25229.000000
Name: DAYS_BIRTH, dtype: float64
```

```
sns.boxplot(df.DAYS_BIRTH)
plt.title('Distribution of Age in the form of days')
plt.show()
```



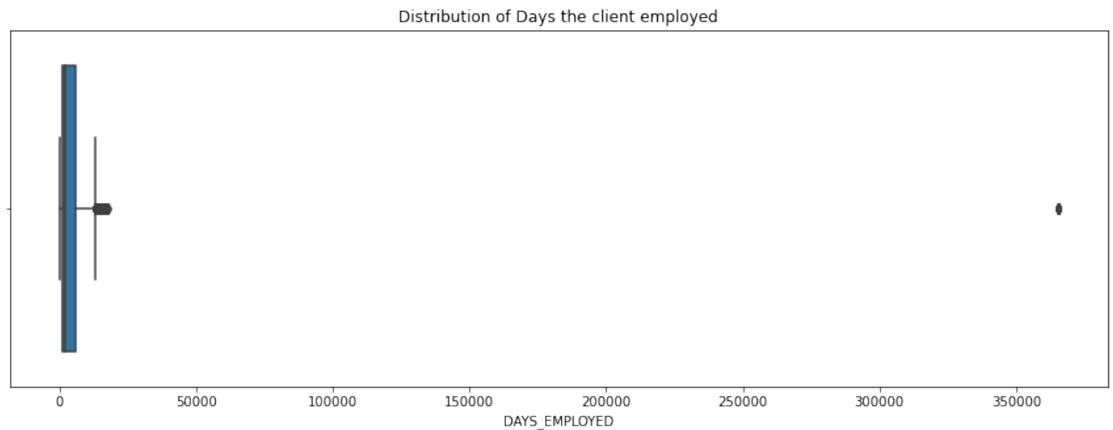
DAYS_BIRTH column we can see from box plot that there are no outliers. There is no much difference between mean and median. Which means that all the applications received from the customers are of almost same age.

DAYS EMPLOYED variable

```
df.DAYS_EMPLOYED.describe()
```

```
count    307511.000000
mean      67724.742149
std      139443.751806
min           0.000000
25%         933.000000
50%        2219.000000
75%        5707.000000
max      365243.000000
Name: DAYS_EMPLOYED, dtype: float64
```

```
plt.figure(figsize=(15,5))
sns.boxplot(df.DAYS_EMPLOYED)
#plt.yscale('log')
plt.title('Distribution of Days the client employed')

plt.show()
```



DAYS_EMPLOYED column has outliers at 365243. Number of days the person was employed varies from person to person

Binning Continuous Variable

```
#Creating bins for Credit amount

bins = [0,350000,700000,1000000000]
slots = ['Low','Medium','High']

df['AMT_CREDIT_RANGE']=pd.cut(df['AMT_CREDIT'],bins=bins,labels=slots)

# Creating bins for income amount

bins = [0,200000,400000,10000000000]
slot = ['Low','Medium','High']

df['AMT_INCOME_RANGE']=pd.cut(df['AMT_INCOME_TOTAL'],bins,labels=slot)

# Creating bins for days_birth

bins = [0,7300,10950,14600,18250,21900,25500]
slot = ['0-20','20-30','30-40','40-50','50-60','60-70']

df['AGE_RANGE']=pd.cut(df['DAYS_BIRTH'],bins,labels=slot)

#Checking bin columns created in df.
df.head()

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

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

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

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

   REGION_POPULATION_RELATIVE  DAYS_BIRTH  DAYS_EMPLOYED
DAYS_REGISTRATION  \
0                    0.018801        9461            637
3648.0
1                    0.003541       16765           1188
1186.0
2                    0.010032       19046            225
4260.0
3                    0.008019       19005           3039
9833.0
4                    0.028663       19932           3038
4311.0

   DAYS_ID_PUBLISH OCCUPATION_TYPE WEEKDAY_APPR_PROCESS_START  \
0             2120        Laborers                  WEDNESDAY
1              291      Core staff                     MONDAY
2             2531        Laborers                     MONDAY
```

```
3             2437        Laborers                  WEDNESDAY
4             3458      Core staff                   THURSDAY

   HOUR_APPR_PROCESS_START REG_REGION_NOT_LIVE_REGION  \
0                       10                          0
1                       11                          0
2                        9                          0
3                       17                          0
4                       11                          0

  REG_REGION_NOT_WORK_REGION LIVE_REGION_NOT_WORK_REGION  \
0                          0                           0
1                          0                           0
2                          0                           0
3                          0                           0
4                          0                           0

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

        ORGANIZATION_TYPE  EXT_SOURCE_2  OBS_30_CNT_SOCIAL_CIRCLE  \
0  Business Entity Type 3      0.262949                       2.0
1                  School      0.622246                       1.0
2              Government      0.555912                       0.0
3  Business Entity Type 3      0.650442                       2.0
4                Religion      0.322738                       0.0

   DEF_30_CNT_SOCIAL_CIRCLE  OBS_60_CNT_SOCIAL_CIRCLE  \
0                       2.0                       2.0
1                       0.0                       1.0
2                       0.0                       0.0
3                       0.0                       2.0
4                       0.0                       0.0

   DEF_60_CNT_SOCIAL_CIRCLE  AMT_REQ_CREDIT_BUREAU_HOUR  \
0                       2.0                         0.0
1                       0.0                         0.0
2                       0.0                         0.0
3                       0.0                         NaN
4                       0.0                         0.0
```

```
   AMT_REQ_CREDIT_BUREAU_DAY  AMT_REQ_CREDIT_BUREAU_WEEK  \
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_MON  AMT_REQ_CREDIT_BUREAU_QRT  \
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_YEAR AMT_CREDIT_RANGE AMT_INCOME_RANGE
AGE_RANGE
0                         1.0           Medium           Medium
20-30
1                         0.0             High           Medium
40-50
2                         0.0              Low              Low
50-60
3                         NaN              Low              Low
50-60
4                         0.0           Medium              Low
50-60
```

Analysis

```
# Calculating Imbalance percentage
100*(df.TARGET.value_counts())/ (len(df))
```

```
0    91.927118
1     8.072882
Name: TARGET, dtype: float64
```

So TARGET column has 8.07% of 1's which means 8% clients have payment difficulties and 91.92% are having no difficulties

```
# Dividing the dataset into two dataset of  target=1(client with
payment difficulties) and target=0(all other)
target_1 = df[df['TARGET']==1]
target_0 = df[df['TARGET']==0]
```

```
#Dataframe having target values 0
target_0.head()
```

```
   SK_ID_CURR  TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR  \
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
5      100008       0         Cash loans           M            N

  FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT
AMT_ANNUITY  \
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
5               Y             0           99000.0    490495.5
27517.5

   AMT_GOODS_PRICE  NAME_TYPE_SUITE NAME_INCOME_TYPE  \
1        1129500.0           Family    State servant
2         135000.0    Unaccompanied          Working
3         297000.0    Unaccompanied          Working
4         513000.0    Unaccompanied          Working
5         454500.0  Spouse, partner    State servant

             NAME_EDUCATION_TYPE    NAME_FAMILY_STATUS
NAME_HOUSING_TYPE  \
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
5  Secondary / secondary special               Married  House /
apartment

   REGION_POPULATION_RELATIVE  DAYS_BIRTH  DAYS_EMPLOYED
DAYS_REGISTRATION  \
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
5                    0.035792       16941           1588
4970.0
```

```
   DAYS_ID_PUBLISH OCCUPATION_TYPE WEEKDAY_APPR_PROCESS_START  \
1              291      Core staff                     MONDAY
2             2531        Laborers                     MONDAY
3             2437        Laborers                  WEDNESDAY
4             3458      Core staff                   THURSDAY
5              477        Laborers                  WEDNESDAY

   HOUR_APPR_PROCESS_START REG_REGION_NOT_LIVE_REGION  \
1                       11                          0
2                        9                          0
3                       17                          0
4                       11                          0
5                       16                          0

  REG_REGION_NOT_WORK_REGION LIVE_REGION_NOT_WORK_REGION  \
1                          0                           0
2                          0                           0
3                          0                           0
4                          0                           0
5                          0                           0

  REG_CITY_NOT_LIVE_CITY REG_CITY_NOT_WORK_CITY
LIVE_CITY_NOT_WORK_CITY  \
1                      0                      0
0
2                      0                      0
0
3                      0                      0
0
4                      0                      1
1
5                      0                      0
0

        ORGANIZATION_TYPE  EXT_SOURCE_2  OBS_30_CNT_SOCIAL_CIRCLE  \
1                  School      0.622246                       1.0
2              Government      0.555912                       0.0
3  Business Entity Type 3      0.650442                       2.0
4                Religion      0.322738                       0.0
5                   Other      0.354225                       0.0

   DEF_30_CNT_SOCIAL_CIRCLE  OBS_60_CNT_SOCIAL_CIRCLE  \
1                       0.0                       1.0
2                       0.0                       0.0
3                       0.0                       2.0
4                       0.0                       0.0
5                       0.0                       0.0
```

```
   DEF_60_CNT_SOCIAL_CIRCLE  AMT_REQ_CREDIT_BUREAU_HOUR  \
1                       0.0                         0.0
2                       0.0                         0.0
3                       0.0                         NaN
4                       0.0                         0.0
5                       0.0                         0.0

   AMT_REQ_CREDIT_BUREAU_DAY  AMT_REQ_CREDIT_BUREAU_WEEK  \
1                        0.0                         0.0
2                        0.0                         0.0
3                        NaN                         NaN
4                        0.0                         0.0
5                        0.0                         0.0

   AMT_REQ_CREDIT_BUREAU_MON  AMT_REQ_CREDIT_BUREAU_QRT  \
1                        0.0                        0.0
2                        0.0                        0.0
3                        NaN                        NaN
4                        0.0                        0.0
5                        0.0                        1.0

   AMT_REQ_CREDIT_BUREAU_YEAR AMT_CREDIT_RANGE AMT_INCOME_RANGE
AGE_RANGE
1                         0.0             High           Medium
40-50
2                         0.0              Low              Low
50-60
3                         NaN              Low              Low
50-60
4                         0.0           Medium              Low
50-60
5                         1.0           Medium              Low
40-50
```

```
#Dataframe having target values 1
target_1.head()
```

```
    SK_ID_CURR  TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR  \
0       100002       1         Cash loans           M            N
26      100031       1         Cash loans           F            N
40      100047       1         Cash loans           M            N
42      100049       1         Cash loans           F            N
81      100096       1         Cash loans           F            N

   FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT
AMT_ANNUITY  \
0                Y             0          202500.0    406597.5
24700.5
26               Y             0          112500.0    979992.0
27076.5
```

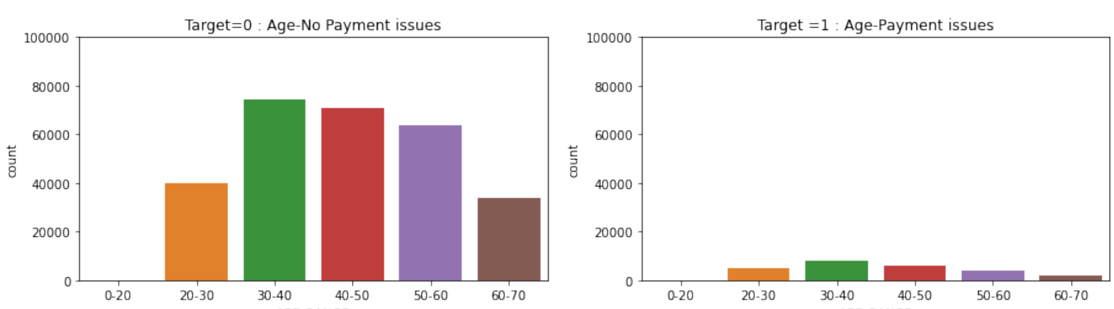
```
40                Y             0          202500.0   1193580.0
35028.0
42                N             0          135000.0    288873.0
16258.5
81                Y             0           81000.0    252000.0
14593.5

    AMT_GOODS_PRICE NAME_TYPE_SUITE      NAME_INCOME_TYPE  \
0          351000.0   Unaccompanied               Working
26         702000.0   Unaccompanied               Working
40         855000.0   Unaccompanied  Commercial associate
42         238500.0   Unaccompanied               Working
81         252000.0   Unaccompanied             Pensioner

              NAME_EDUCATION_TYPE    NAME_FAMILY_STATUS
NAME_HOUSING_TYPE  \
0   Secondary / secondary special  Single / not married  House /
apartment
26  Secondary / secondary special                 Widow  House /
apartment
40  Secondary / secondary special               Married  House /
apartment
42  Secondary / secondary special        Civil marriage  House /
apartment
81  Secondary / secondary special               Married  House /
apartment

    REGION_POPULATION_RELATIVE  DAYS_BIRTH  DAYS_EMPLOYED
DAYS_REGISTRATION  \
0                     0.018801        9461            637
3648.0
26                    0.018029       18724           2628
6573.0
40                    0.025164       17482           1262
1182.0
42                    0.007305       13384           3597
45.0
81                    0.028663       24794         365243
5391.0

    DAYS_ID_PUBLISH OCCUPATION_TYPE WEEKDAY_APPR_PROCESS_START  \
0              2120        Laborers                  WEDNESDAY
26             1827   Cooking staff                     MONDAY
40             1029        Laborers                    TUESDAY
42             4409     Sales staff                   THURSDAY
81             4199             NaN                   THURSDAY

    HOUR_APPR_PROCESS_START REG_REGION_NOT_LIVE_REGION  \
0                        10                          0
26                        9                          0
```

```
40                        9                          0
42                       11                          0
81                       10                          0

   REG_REGION_NOT_WORK_REGION LIVE_REGION_NOT_WORK_REGION  \
0                           0                           0
26                          0                           0
40                          0                           0
42                          0                           0
81                          0                           0

   REG_CITY_NOT_LIVE_CITY REG_CITY_NOT_WORK_CITY
LIVE_CITY_NOT_WORK_CITY  \
0                       0                      0
0
26                      0                      0
0
40                      0                      0
0
42                      0                      0
0
81                      0                      0
0

         ORGANIZATION_TYPE  EXT_SOURCE_2  OBS_30_CNT_SOCIAL_CIRCLE  \
0   Business Entity Type 3      0.262949                        2.0
26  Business Entity Type 3      0.548477                       10.0
40  Business Entity Type 3      0.306841                        0.0
42           Self-employed      0.674203                        1.0
81                     XNA      0.023952                        1.0

    DEF_30_CNT_SOCIAL_CIRCLE  OBS_60_CNT_SOCIAL_CIRCLE  \
0                        2.0                       2.0
26                       1.0                      10.0
40                       0.0                       0.0
42                       0.0                       1.0
81                       1.0                       1.0

    DEF_60_CNT_SOCIAL_CIRCLE  AMT_REQ_CREDIT_BUREAU_HOUR  \
0                        2.0                         0.0
26                       0.0                         0.0
40                       0.0                         0.0
42                       0.0                         0.0
81                       1.0                         0.0

    AMT_REQ_CREDIT_BUREAU_DAY  AMT_REQ_CREDIT_BUREAU_WEEK  \
0                         0.0                         0.0
26                        0.0                         0.0
40                        0.0                         0.0
42                        0.0                         0.0
```

```
81                          0.0                        0.0

    AMT_REQ_CREDIT_BUREAU_MON  AMT_REQ_CREDIT_BUREAU_QRT  \
0                         0.0                        0.0
26                        0.0                        2.0
40                        2.0                        0.0
42                        0.0                        0.0
81                        0.0                        0.0

    AMT_REQ_CREDIT_BUREAU_YEAR AMT_CREDIT_RANGE AMT_INCOME_RANGE
AGE_RANGE
0                          1.0           Medium           Medium
20-30
26                         2.0             High              Low
50-60
40                         4.0             High           Medium
40-50
42                         2.0              Low              Low
30-40
81                         0.0              Low              Low
60-70
```

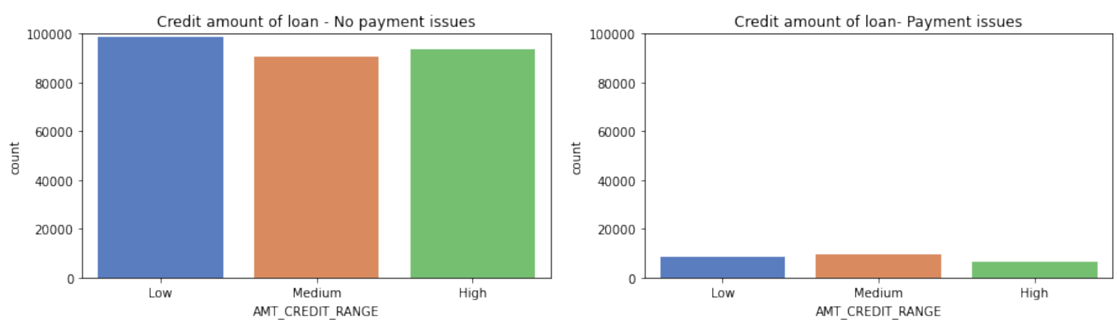
Univariate Analysis for target =0 and target=1

```
# Numeric variable analysis for target_0 & target_1 dataframe
plt.figure(figsize = (15, 8))
plt.subplot(2, 2, 1)
plt.ylim(0,100000)
plt.title('Target=0 : Age-No Payment issues')
sns.countplot(target_0['AGE_RANGE'])

# subplot 2
plt.subplot(2, 2, 2)
plt.title('Target =1 : Age-Payment issues')
plt.ylim(0,100000)
sns.countplot(target_1['AGE_RANGE'])
plt.show()
```



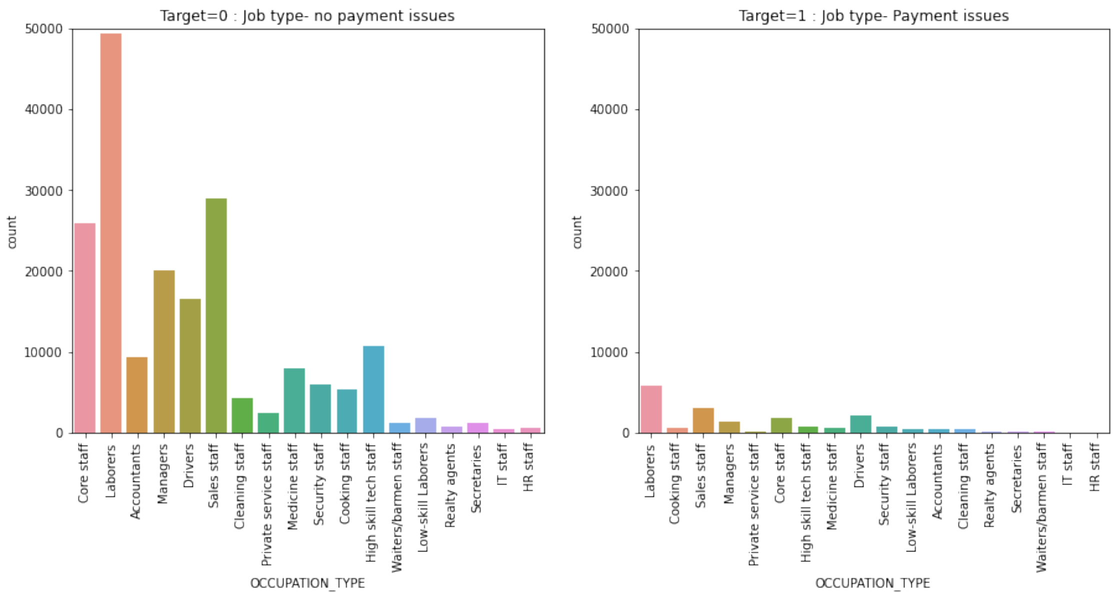
We can observe that customers belonging to age group 30-40 are able to make payment on time and can be considered while lending loan. The customers from 40 to 60 age are also can be considered.

```
# Numeric variable analysis for target_0 & target_1 dataframe
plt.figure(figsize = (15, 8))
plt.subplot(2, 2, 1)
plt.ylim(0,100000)
plt.title('Credit amount of loan - No payment issues')
sns.countplot(target_0['AMT_CREDIT_RANGE'],palette='muted')

# subplot 2
plt.subplot(2, 2, 2)
plt.title('Credit amount of loan- Payment issues')
plt.ylim(0,100000)
sns.countplot(target_1['AMT_CREDIT_RANGE'], palette='muted')
plt.show()
```



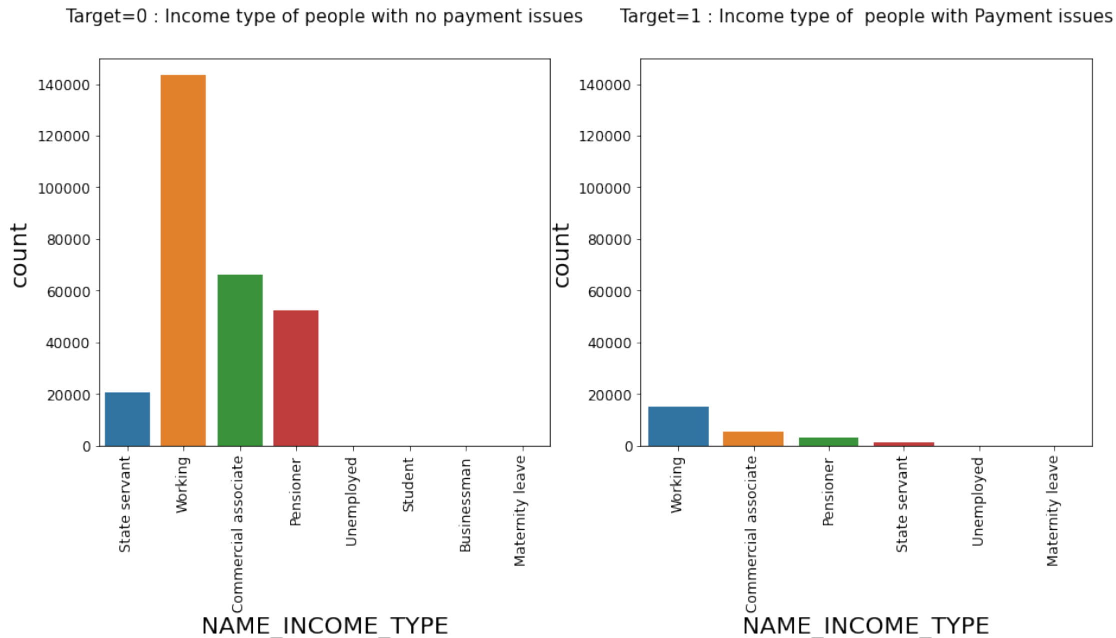
Customers with less credit and most likely to make payment. Customers having medium and high credit can also be considered while lending the loan

```
# Categorical variable analysis for target_0 & target_1 dataframe
plt.figure(figsize = (15,6))
plt.subplot(1, 2, 1)
#plt.subplots_adjust(wspace=0.5)


sns.countplot(target_0['OCCUPATION_TYPE'])
plt.title('Target=0 : Job type- no payment issues')
plt.ylim(0,50000)
plt.xticks(rotation = 90)

# subplot 2
plt.subplot(1, 2, 2)

sns.countplot(target_1['OCCUPATION_TYPE'])
plt.title('Target=1 : Job type- Payment issues')
plt.ylim(0,50000)
plt.xticks(rotation = 90)
plt.show()
```



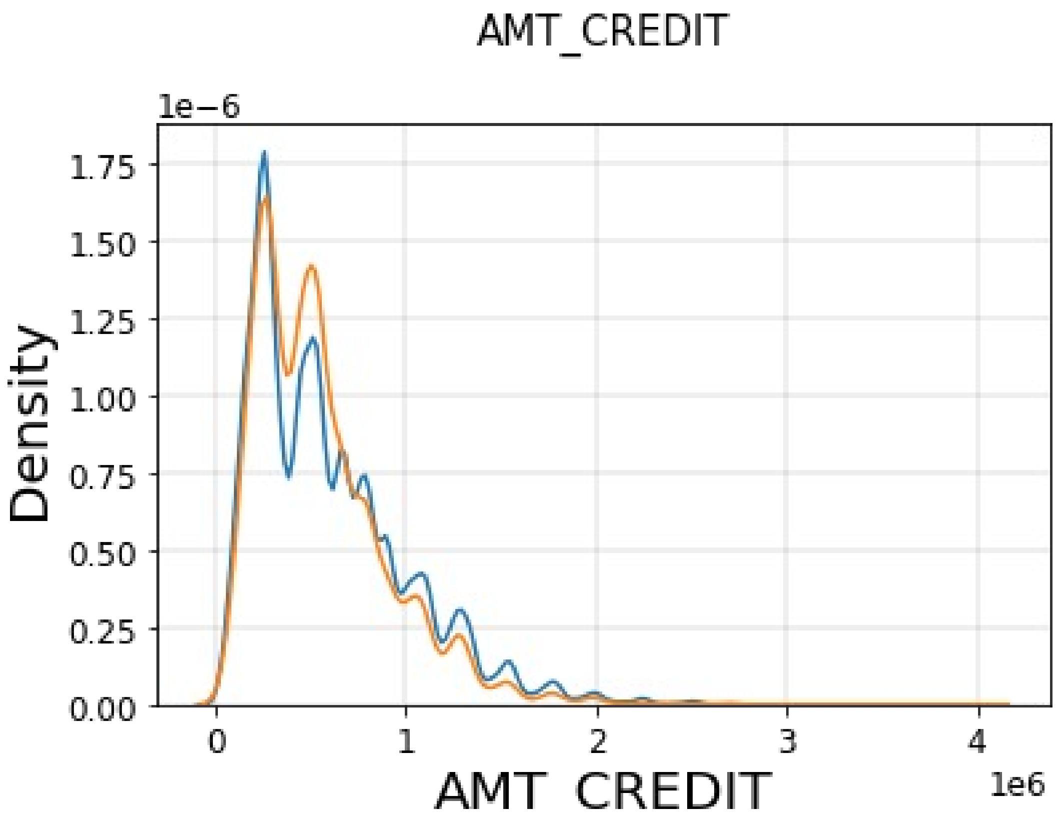
The plot clearly shows that labourers are most likely to make payment on time whereas HR staff are less likely to make payment on time

```
# Categorical variable analysis for target_0 & target_1 dataframe
plt.figure(figsize = (15,6))
plt.rcParams['axes.titlesize'] = 15
plt.rcParams["axes.labelsize"] = 20
plt.rcParams['axes.titlepad'] = 30
plt.rc('xtick',labelsize=12)
plt.rc('ytick',labelsize=12)
plt.subplot(1, 2, 1)
#plt.title()
sns.countplot(target_0['NAME_INCOME_TYPE'].dropna())
plt.title('Target=0 : Income type of people with no payment issues')
plt.ylim(0,150000)
plt.xticks(rotation = 90)

# subplot 2
plt.subplot(1, 2, 2)

sns.countplot(target_1['NAME_INCOME_TYPE'].dropna())
plt.title('Target=1 : Income type of  people with Payment issues')
plt.ylim(0,150000)
plt.xticks(rotation = 90)
plt.show()
```

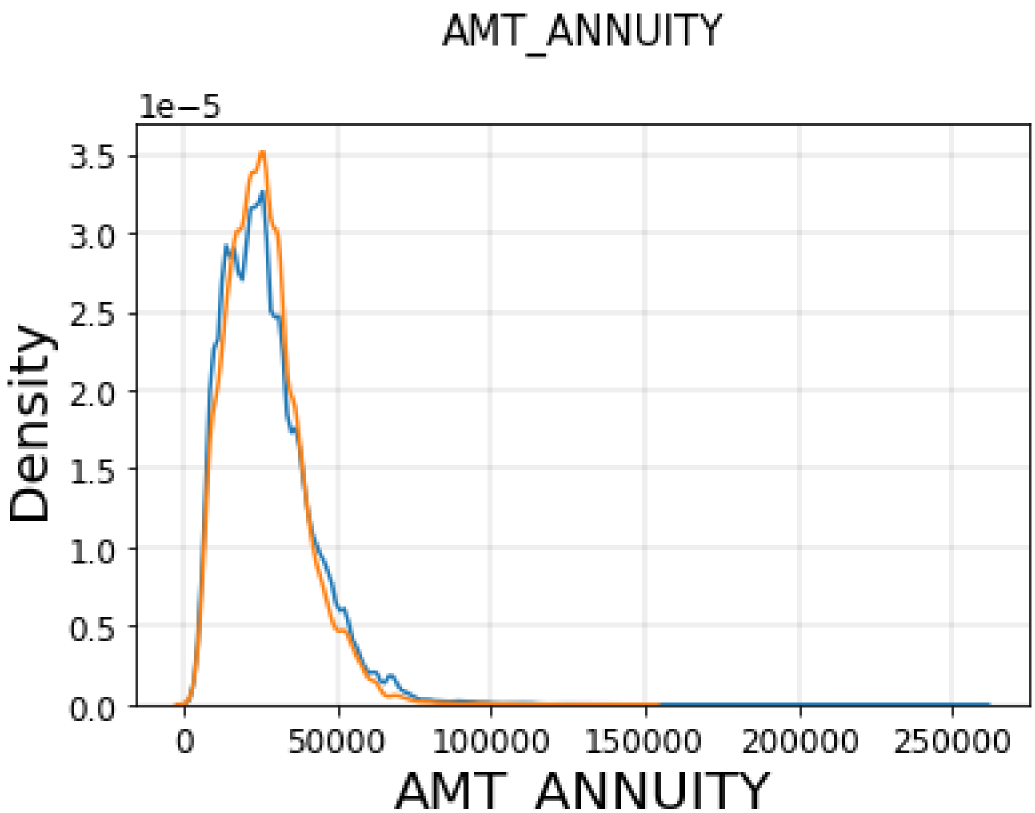


The plot clearly shows that labourers are most likely to make payment on time whereas HR staff are less likely to make payment on time

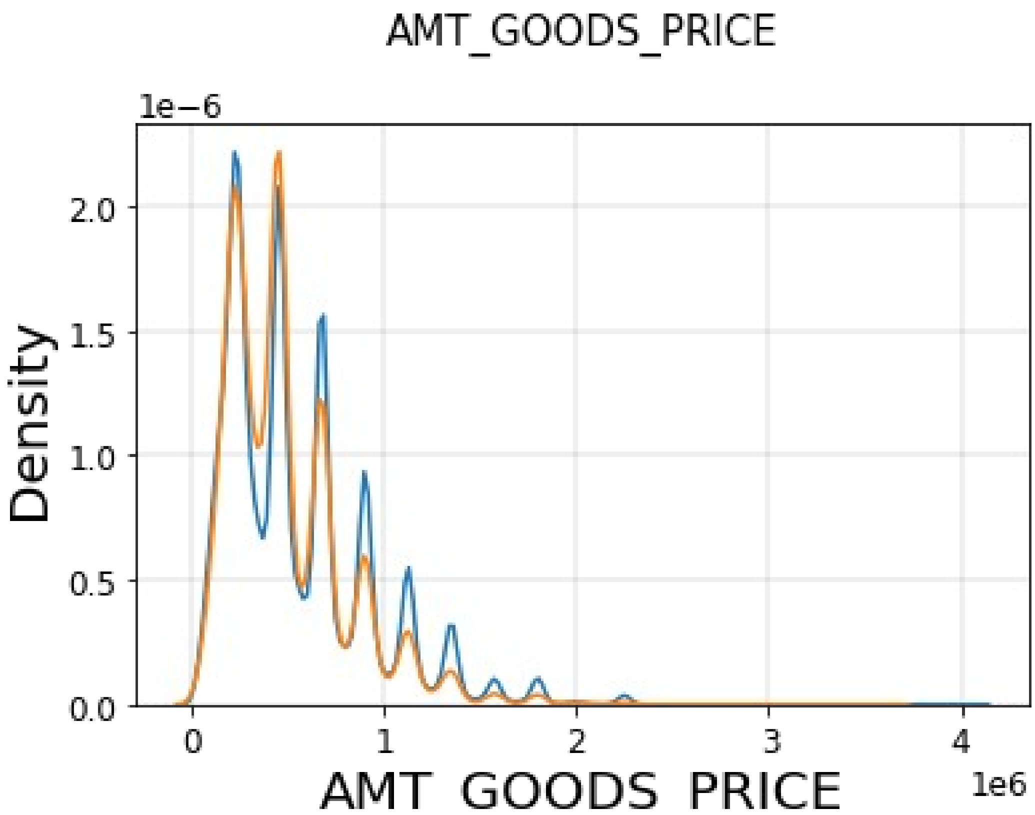
```
#Analyse continuous column with respect to the target column
sns.distplot(target_0['AMT_CREDIT'], hist = False, label="Good")# Target = 0
sns.distplot(target_1['AMT_CREDIT'], hist = False, label='Bad')# Taget = 1
plt.title('AMT_CREDIT')
plt.grid(color='black', linestyle='-', linewidth=0.25, alpha=0.5)
plt.show()
```




```
#Analyse continuous column with respect to the target column
sns.distplot(target_0['AMT_ANNUITY'], hist = False, label="Good")#
Target = 0
sns.distplot(target_1['AMT_ANNUITY'], hist = False, label="Bad")#
Taget = 1
plt.title('AMT_ANNUITY')
plt.grid(color='black', linestyle='-', linewidth=0.25, alpha=0.5)
plt.show()
```



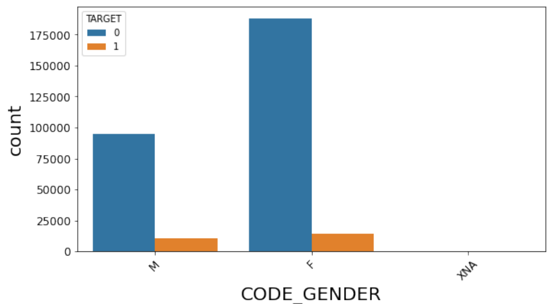

```
#Analyse continuous column with respect to the target column
sns.distplot(target_0['AMT_GOODS_PRICE'], hist = False,label= "good")# Target = 0
sns.distplot(target_1['AMT_GOODS_PRICE'], hist = False, label="bad")# Taget = 1
plt.title('AMT_GOODS_PRICE')
plt.grid(color='black', linestyle='-', linewidth=0.25, alpha=0.5)
plt.show()
```

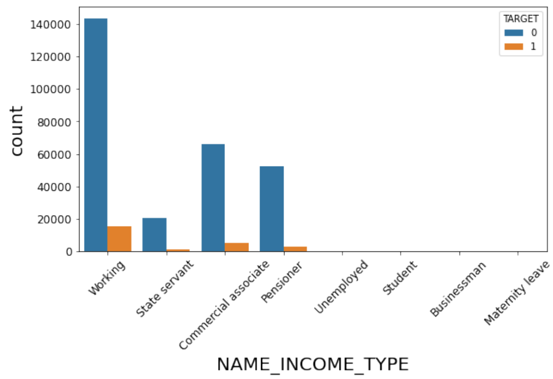


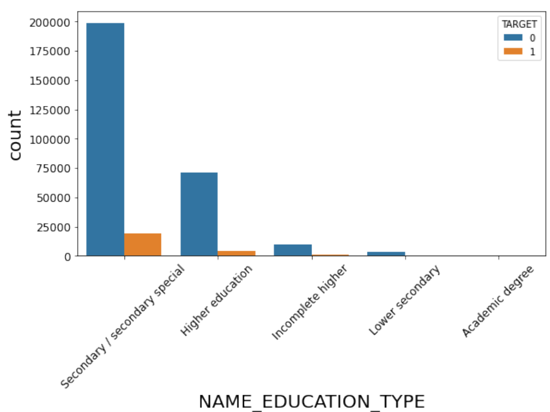

```
#Plot mutiple categorical columns with respect to Target column: 
Subplot
features = ['CODE_GENDER', 'NAME_INCOME_TYPE', 
'NAME_EDUCATION_TYPE','NAME_FAMILY_STATUS','NAME_HOUSING_TYPE','OCCUPA
TION_TYPE']
list(enumerate(features))
```

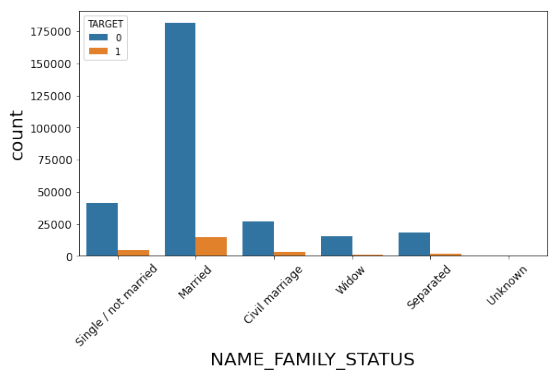
```
[(0, 'CODE_GENDER'),
 (1, 'NAME_INCOME_TYPE'),
 (2, 'NAME_EDUCATION_TYPE'),
 (3, 'NAME_FAMILY_STATUS'),
 (4, 'NAME_HOUSING_TYPE'),
 (5, 'OCCUPATION_TYPE')]
```

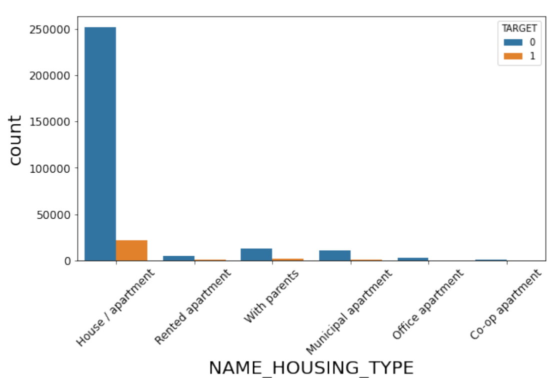
```
features = ['CODE_GENDER', 'NAME_INCOME_TYPE', 
'NAME_EDUCATION_TYPE','NAME_FAMILY_STATUS','NAME_HOUSING_TYPE','OCCUPA
TION_TYPE']
plt.figure(figsize = (20, 40))

plt.subplots_adjust(hspace=0.8)
for i in enumerate(features):
    plt.subplot(5, 2, i[0]+1)
    sns.countplot(x = i[1], hue = 'TARGET', data = df)
    plt.xticks(rotation = 45)
```

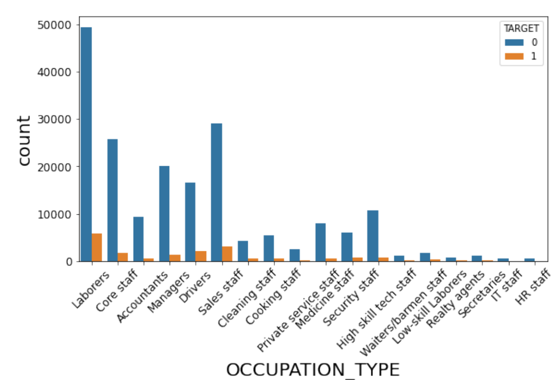












From the above plot we can see that,

a) Female customers pay loan amount on time and banks can target more female cusytomers for lending loan.

b) Working customers can be targetted to lend loans as they have higher percentage of making payments on time.

c) Customers with secondary education are most likely to make payments when compared to customers with academic degree.

d) Married customers have paid loan amount on time when compared to widows.

e) Customers owning House/apartment are most likely to make payments on time compared to those living in CO-OP apartment.

f) Labourers have high repayement percentage. Hence baks can think of lending small amount loans to them.

Correlation Matrix

```
#correlation matrix for all numerical columns
corr=target_0.corr()
corr
```

```
                            SK_ID_CURR  TARGET  CNT_CHILDREN  \
SK_ID_CURR                    1.000000     NaN     -0.000716
TARGET                             NaN     NaN           NaN
CNT_CHILDREN                 -0.000716     NaN      1.000000
AMT_INCOME_TOTAL              0.001739     NaN      0.027397
AMT_CREDIT                   -0.000342     NaN      0.003081
AMT_ANNUITY                   0.000068     NaN      0.020905
AMT_GOODS_PRICE              -0.000205     NaN     -0.000525
REGION_POPULATION_RELATIVE    0.000360     NaN     -0.024363
DAYS_BIRTH                    0.001346     NaN     -0.336966
DAYS_EMPLOYED                 0.001744     NaN     -0.245174
DAYS_REGISTRATION             0.001475     NaN     -0.185792
DAYS_ID_PUBLISH               0.000077     NaN      0.028751
HOUR_APPR_PROCESS_START      -0.000115     NaN     -0.005244
EXT_SOURCE_2                  0.001589     NaN     -0.015455
OBS_30_CNT_SOCIAL_CIRCLE     -0.000695     NaN      0.014471
DEF_30_CNT_SOCIAL_CIRCLE      0.000556     NaN     -0.002246
OBS_60_CNT_SOCIAL_CIRCLE     -0.000741     NaN      0.014137
DEF_60_CNT_SOCIAL_CIRCLE      0.002382     NaN     -0.002172
AMT_REQ_CREDIT_BUREAU_HOUR   -0.001879     NaN     -0.000432
AMT_REQ_CREDIT_BUREAU_DAY    -0.001725     NaN      0.000648
AMT_REQ_CREDIT_BUREAU_WEEK    0.002524     NaN     -0.001632
AMT_REQ_CREDIT_BUREAU_MON     0.000054     NaN     -0.010455
AMT_REQ_CREDIT_BUREAU_QRT     0.001148     NaN     -0.007087
AMT_REQ_CREDIT_BUREAU_YEAR    0.004349     NaN     -0.042547

                            AMT_INCOME_TOTAL  AMT_CREDIT  AMT_ANNUITY
\
SK_ID_CURR                          0.001739   -0.000342     0.000068

TARGET                                   NaN         NaN          NaN

CNT_CHILDREN                        0.027397    0.003081     0.020905

AMT_INCOME_TOTAL                    1.000000    0.342799     0.418953

AMT_CREDIT                          0.342799    1.000000     0.771309
```

AMT_ANNUITY	0.418953	0.771309	1.000000
AMT_GOODS_PRICE	0.349462	0.987250	0.776686
REGION_POPULATION_RELATIVE	0.167851	0.100604	0.120988
DAYS_BIRTH	-0.062609	0.047378	-0.012263
DAYS_EMPLOYED	-0.140392	-0.070104	-0.104978
DAYS_REGISTRATION	-0.064937	-0.013477	-0.039436
DAYS_ID_PUBLISH	-0.022896	0.001464	-0.014113
HOUR_APPR_PROCESS_START	0.076743	0.053619	0.053589
EXT_SOURCE_2	0.139598	0.129140	0.126615
OBS_30_CNT_SOCIAL_CIRCLE	-0.027828	-0.000914	-0.013091
DEF_30_CNT_SOCIAL_CIRCLE	-0.027621	-0.019851	-0.022535
OBS_60_CNT_SOCIAL_CIRCLE	-0.027690	-0.000892	-0.012893
DEF_60_CNT_SOCIAL_CIRCLE	-0.027593	-0.022225	-0.023385
AMT_REQ_CREDIT_BUREAU_HOUR	0.001417	-0.003734	0.003148
AMT_REQ_CREDIT_BUREAU_DAY	0.007862	0.004409	0.002392
AMT_REQ_CREDIT_BUREAU_WEEK	0.006234	-0.001883	0.012681
AMT_REQ_CREDIT_BUREAU_MON	0.061470	0.054071	0.038181
AMT_REQ_CREDIT_BUREAU_QRT	0.013128	0.017767	0.011248
AMT_REQ_CREDIT_BUREAU_YEAR	0.029536	-0.048866	-0.010882

```
                            AMT_GOODS_PRICE
REGION_POPULATION_RELATIVE  \
SK_ID_CURR                        -0.000205
0.000360
TARGET                                  NaN
NaN
CNT_CHILDREN                      -0.000525                   -
0.024363
```

```
AMT_INCOME_TOTAL                   0.349462                   
0.167851  
AMT_CREDIT                         0.987250                   
0.100604  
AMT_ANNUITY                        0.776686                   
0.120988  
AMT_GOODS_PRICE                    1.000000                   
0.103827  
REGION_POPULATION_RELATIVE         0.103827                   
1.000000  
DAYS_BIRTH                         0.044565                   
0.025244  
DAYS_EMPLOYED                     -0.068609                  -
0.007198  
DAYS_REGISTRATION                 -0.015916                   
0.052083  
DAYS_ID_PUBLISH                    0.003649                   
0.001071  
HOUR_APPR_PROCESS_START            0.062766                   
0.172814  
EXT_SOURCE_2                       0.135694                   
0.198493  
OBS_30_CNT_SOCIAL_CIRCLE          -0.000707                  -
0.012107  
DEF_30_CNT_SOCIAL_CIRCLE          -0.021025                   
0.005941  
OBS_60_CNT_SOCIAL_CIRCLE          -0.000723                  -
0.011591  
DEF_60_CNT_SOCIAL_CIRCLE          -0.023171                   
0.002255  
AMT_REQ_CREDIT_BUREAU_HOUR        -0.003116                  -
0.002265  
AMT_REQ_CREDIT_BUREAU_DAY          0.004820                   
0.001969  
AMT_REQ_CREDIT_BUREAU_WEEK        -0.001597                  -
0.002480  
AMT_REQ_CREDIT_BUREAU_MON          0.055850                   
0.078629  
AMT_REQ_CREDIT_BUREAU_QRT          0.018163                  -
0.001013  
AMT_REQ_CREDIT_BUREAU_YEAR        -0.051266                   
0.001775  

                            DAYS_BIRTH  DAYS_EMPLOYED  
DAYS_REGISTRATION  \
SK_ID_CURR                    0.001346       0.001744           
0.001475   
TARGET                             NaN            NaN           
NaN   
CNT_CHILDREN                 -0.336966      -0.245174          -
```

```
0.185792
AMT_INCOME_TOTAL             -0.062609      -0.140392          -
0.064937
AMT_CREDIT                    0.047378      -0.070104          -
0.013477
AMT_ANNUITY                  -0.012263      -0.104978          -
0.039436
AMT_GOODS_PRICE               0.044565      -0.068609          -
0.015916
REGION_POPULATION_RELATIVE    0.025244      -0.007198
0.052083
DAYS_BIRTH                    1.000000       0.626114
0.333151
DAYS_EMPLOYED                 0.626114       1.000000
0.214511
DAYS_REGISTRATION             0.333151       0.214511
1.000000
DAYS_ID_PUBLISH               0.271314       0.276663
0.100236
HOUR_APPR_PROCESS_START      -0.095916      -0.095455
0.008044
EXT_SOURCE_2                  0.078158      -0.028530
0.052503
OBS_30_CNT_SOCIAL_CIRCLE     -0.007726       0.007272          -
0.008315
DEF_30_CNT_SOCIAL_CIRCLE      0.003057       0.019379          -
0.001213
OBS_60_CNT_SOCIAL_CIRCLE     -0.007316       0.007453          -
0.008247
DEF_60_CNT_SOCIAL_CIRCLE      0.000990       0.016383          -
0.002652
AMT_REQ_CREDIT_BUREAU_HOUR   -0.004461      -0.004460
0.002730
AMT_REQ_CREDIT_BUREAU_DAY    -0.002772      -0.000874
0.000035
AMT_REQ_CREDIT_BUREAU_WEEK    0.001069       0.002223
0.001431
AMT_REQ_CREDIT_BUREAU_MON    -0.002855      -0.034890
0.011857
AMT_REQ_CREDIT_BUREAU_QRT     0.011335       0.015116
0.000527
AMT_REQ_CREDIT_BUREAU_YEAR    0.072886       0.052169
0.026639

                            DAYS_ID_PUBLISH
HOUR_APPR_PROCESS_START  \
SK_ID_CURR                         0.000077                -0.000115

TARGET                                  NaN                      NaN
```

CNT_CHILDREN	0.028751	-0.005244
AMT_INCOME_TOTAL	-0.022896	0.076743
AMT_CREDIT	0.001464	0.053619
AMT_ANNUITY	-0.014113	0.053589
AMT_GOODS_PRICE	0.003649	0.062766
REGION_POPULATION_RELATIVE	0.001071	0.172814
DAYS_BIRTH	0.271314	-0.095916
DAYS_EMPLOYED	0.276663	-0.095455
DAYS_REGISTRATION	0.100236	0.008044
DAYS_ID_PUBLISH	1.000000	-0.033980
HOUR_APPR_PROCESS_START	-0.033980	1.000000
EXT_SOURCE_2	0.041703	0.157221
OBS_30_CNT_SOCIAL_CIRCLE	0.012326	-0.008013
DEF_30_CNT_SOCIAL_CIRCLE	-0.000403	-0.005718
OBS_60_CNT_SOCIAL_CIRCLE	0.012810	-0.007956
DEF_60_CNT_SOCIAL_CIRCLE	-0.002491	-0.008826
AMT_REQ_CREDIT_BUREAU_HOUR	-0.004797	-0.015832
AMT_REQ_CREDIT_BUREAU_DAY	0.000198	0.003847
AMT_REQ_CREDIT_BUREAU_WEEK	0.001655	-0.002652
AMT_REQ_CREDIT_BUREAU_MON	0.008728	0.037332
AMT_REQ_CREDIT_BUREAU_QRT	0.007071	-0.000671
AMT_REQ_CREDIT_BUREAU_YEAR	0.035639	-0.030330

	EXT_SOURCE_2	OBS_30_CNT_SOCIAL_CIRCLE \
SK_ID_CURR	0.001589	-0.000695
TARGET	NaN	NaN

```
CNT_CHILDREN                    -0.015455              0.014471
AMT_INCOME_TOTAL                 0.139598             -0.027828
AMT_CREDIT                       0.129140             -0.000914
AMT_ANNUITY                      0.126615             -0.013091
AMT_GOODS_PRICE                  0.135694             -0.000707
REGION_POPULATION_RELATIVE       0.198493             -0.012107
DAYS_BIRTH                       0.078158             -0.007726
DAYS_EMPLOYED                   -0.028530              0.007272
DAYS_REGISTRATION                0.052503             -0.008315
DAYS_ID_PUBLISH                  0.041703              0.012326
HOUR_APPR_PROCESS_START          0.157221             -0.008013
EXT_SOURCE_2                     1.000000             -0.021569
OBS_30_CNT_SOCIAL_CIRCLE        -0.021569              1.000000
DEF_30_CNT_SOCIAL_CIRCLE        -0.027427              0.329206
OBS_60_CNT_SOCIAL_CIRCLE        -0.021224              0.998508
DEF_60_CNT_SOCIAL_CIRCLE        -0.029722              0.253000
AMT_REQ_CREDIT_BUREAU_HOUR      -0.003918              0.000420
AMT_REQ_CREDIT_BUREAU_DAY        0.001597             -0.001940
AMT_REQ_CREDIT_BUREAU_WEEK       0.001420              0.000383
AMT_REQ_CREDIT_BUREAU_MON        0.050993              0.001574
AMT_REQ_CREDIT_BUREAU_QRT       -0.002961              0.003966
AMT_REQ_CREDIT_BUREAU_YEAR      -0.021168              0.031884

                            DEF_30_CNT_SOCIAL_CIRCLE  \
SK_ID_CURR                                  0.000556
TARGET                                           NaN
CNT_CHILDREN                               -0.002246
AMT_INCOME_TOTAL                           -0.027621
AMT_CREDIT                                 -0.019851
AMT_ANNUITY                                -0.022535
AMT_GOODS_PRICE                            -0.021025
REGION_POPULATION_RELATIVE                  0.005941
DAYS_BIRTH                                  0.003057
DAYS_EMPLOYED                               0.019379
DAYS_REGISTRATION                          -0.001213
DAYS_ID_PUBLISH                            -0.000403
HOUR_APPR_PROCESS_START                    -0.005718
EXT_SOURCE_2                               -0.027427
OBS_30_CNT_SOCIAL_CIRCLE                    0.329206
DEF_30_CNT_SOCIAL_CIRCLE                    1.000000
OBS_60_CNT_SOCIAL_CIRCLE                    0.331336
DEF_60_CNT_SOCIAL_CIRCLE                    0.859332
AMT_REQ_CREDIT_BUREAU_HOUR                 -0.001472
AMT_REQ_CREDIT_BUREAU_DAY                  -0.001712
AMT_REQ_CREDIT_BUREAU_WEEK                 -0.001882
AMT_REQ_CREDIT_BUREAU_MON                   0.000817
AMT_REQ_CREDIT_BUREAU_QRT                  -0.000960
AMT_REQ_CREDIT_BUREAU_YEAR                  0.018606

                            OBS_60_CNT_SOCIAL_CIRCLE  \
```

```
SK_ID_CURR                                    -0.000741
TARGET                                              NaN
CNT_CHILDREN                                   0.014137
AMT_INCOME_TOTAL                              -0.027690
AMT_CREDIT                                    -0.000892
AMT_ANNUITY                                   -0.012893
AMT_GOODS_PRICE                               -0.000723
REGION_POPULATION_RELATIVE                    -0.011591
DAYS_BIRTH                                    -0.007316
DAYS_EMPLOYED                                  0.007453
DAYS_REGISTRATION                             -0.008247
DAYS_ID_PUBLISH                                0.012810
HOUR_APPR_PROCESS_START                       -0.007956
EXT_SOURCE_2                                  -0.021224
OBS_30_CNT_SOCIAL_CIRCLE                       0.998508
DEF_30_CNT_SOCIAL_CIRCLE                       0.331336
OBS_60_CNT_SOCIAL_CIRCLE                       1.000000
DEF_60_CNT_SOCIAL_CIRCLE                       0.254970
AMT_REQ_CREDIT_BUREAU_HOUR                     0.000328
AMT_REQ_CREDIT_BUREAU_DAY                     -0.002024
AMT_REQ_CREDIT_BUREAU_WEEK                     0.000468
AMT_REQ_CREDIT_BUREAU_MON                      0.001598
AMT_REQ_CREDIT_BUREAU_QRT                      0.003793
AMT_REQ_CREDIT_BUREAU_YEAR                     0.032291

                            DEF_60_CNT_SOCIAL_CIRCLE  \
SK_ID_CURR                                     0.002382
TARGET                                              NaN
CNT_CHILDREN                                  -0.002172
AMT_INCOME_TOTAL                              -0.027593
AMT_CREDIT                                    -0.022225
AMT_ANNUITY                                   -0.023385
AMT_GOODS_PRICE                               -0.023171
REGION_POPULATION_RELATIVE                     0.002255
DAYS_BIRTH                                     0.000990
DAYS_EMPLOYED                                  0.016383
DAYS_REGISTRATION                             -0.002652
DAYS_ID_PUBLISH                               -0.002491
HOUR_APPR_PROCESS_START                       -0.008826
EXT_SOURCE_2                                  -0.029722
OBS_30_CNT_SOCIAL_CIRCLE                       0.253000
DEF_30_CNT_SOCIAL_CIRCLE                       0.859332
OBS_60_CNT_SOCIAL_CIRCLE                       0.254970
DEF_60_CNT_SOCIAL_CIRCLE                       1.000000
AMT_REQ_CREDIT_BUREAU_HOUR                    -0.002132
AMT_REQ_CREDIT_BUREAU_DAY                     -0.001742
AMT_REQ_CREDIT_BUREAU_WEEK                    -0.002374
AMT_REQ_CREDIT_BUREAU_MON                     -0.001396
AMT_REQ_CREDIT_BUREAU_QRT                     -0.000274
AMT_REQ_CREDIT_BUREAU_YEAR                     0.018088
```

```
                             AMT_REQ_CREDIT_BUREAU_HOUR  \
SK_ID_CURR                                    -0.001879
TARGET                                              NaN
CNT_CHILDREN                                  -0.000432
AMT_INCOME_TOTAL                               0.001417
AMT_CREDIT                                    -0.003734
AMT_ANNUITY                                    0.003148
AMT_GOODS_PRICE                               -0.003116
REGION_POPULATION_RELATIVE                    -0.002265
DAYS_BIRTH                                    -0.004461
DAYS_EMPLOYED                                 -0.004460
DAYS_REGISTRATION                              0.002730
DAYS_ID_PUBLISH                               -0.004797
HOUR_APPR_PROCESS_START                       -0.015832
EXT_SOURCE_2                                  -0.003918
OBS_30_CNT_SOCIAL_CIRCLE                       0.000420
DEF_30_CNT_SOCIAL_CIRCLE                      -0.001472
OBS_60_CNT_SOCIAL_CIRCLE                       0.000328
DEF_60_CNT_SOCIAL_CIRCLE                      -0.002132
AMT_REQ_CREDIT_BUREAU_HOUR                     1.000000
AMT_REQ_CREDIT_BUREAU_DAY                      0.229065
AMT_REQ_CREDIT_BUREAU_WEEK                     0.004576
AMT_REQ_CREDIT_BUREAU_MON                      0.000528
AMT_REQ_CREDIT_BUREAU_QRT                     -0.003350
AMT_REQ_CREDIT_BUREAU_YEAR                    -0.004790

                             AMT_REQ_CREDIT_BUREAU_DAY  \
SK_ID_CURR                                   -0.001725
TARGET                                             NaN
CNT_CHILDREN                                  0.000648
AMT_INCOME_TOTAL                              0.007862
AMT_CREDIT                                    0.004409
AMT_ANNUITY                                   0.002392
AMT_GOODS_PRICE                               0.004820
REGION_POPULATION_RELATIVE                    0.001969
DAYS_BIRTH                                   -0.002772
DAYS_EMPLOYED                                -0.000874
DAYS_REGISTRATION                             0.000035
DAYS_ID_PUBLISH                               0.000198
HOUR_APPR_PROCESS_START                       0.003847
EXT_SOURCE_2                                  0.001597
OBS_30_CNT_SOCIAL_CIRCLE                     -0.001940
DEF_30_CNT_SOCIAL_CIRCLE                     -0.001712
OBS_60_CNT_SOCIAL_CIRCLE                     -0.002024
DEF_60_CNT_SOCIAL_CIRCLE                     -0.001742
AMT_REQ_CREDIT_BUREAU_HOUR                    0.229065
AMT_REQ_CREDIT_BUREAU_DAY                     1.000000
AMT_REQ_CREDIT_BUREAU_WEEK                    0.220087
AMT_REQ_CREDIT_BUREAU_MON                    -0.004753
```

```
AMT_REQ_CREDIT_BUREAU_QRT                     -0.004749
AMT_REQ_CREDIT_BUREAU_YEAR                    -0.003798

                            AMT_REQ_CREDIT_BUREAU_WEEK  \
SK_ID_CURR                                    0.002524
TARGET                                             NaN
CNT_CHILDREN                                 -0.001632
AMT_INCOME_TOTAL                              0.006234
AMT_CREDIT                                   -0.001883
AMT_ANNUITY                                   0.012681
AMT_GOODS_PRICE                              -0.001597
REGION_POPULATION_RELATIVE                   -0.002480
DAYS_BIRTH                                    0.001069
DAYS_EMPLOYED                                 0.002223
DAYS_REGISTRATION                             0.001431
DAYS_ID_PUBLISH                               0.001655
HOUR_APPR_PROCESS_START                      -0.002652
EXT_SOURCE_2                                  0.001420
OBS_30_CNT_SOCIAL_CIRCLE                      0.000383
DEF_30_CNT_SOCIAL_CIRCLE                     -0.001882
OBS_60_CNT_SOCIAL_CIRCLE                      0.000468
DEF_60_CNT_SOCIAL_CIRCLE                     -0.002374
AMT_REQ_CREDIT_BUREAU_HOUR                    0.004576
AMT_REQ_CREDIT_BUREAU_DAY                     0.220087
AMT_REQ_CREDIT_BUREAU_WEEK                    1.000000
AMT_REQ_CREDIT_BUREAU_MON                    -0.014248
AMT_REQ_CREDIT_BUREAU_QRT                    -0.015466
AMT_REQ_CREDIT_BUREAU_YEAR                    0.019085

                            AMT_REQ_CREDIT_BUREAU_MON  \
SK_ID_CURR                                   0.000054
TARGET                                            NaN
CNT_CHILDREN                                -0.010455
AMT_INCOME_TOTAL                             0.061470
AMT_CREDIT                                   0.054071
AMT_ANNUITY                                  0.038181
AMT_GOODS_PRICE                              0.055850
REGION_POPULATION_RELATIVE                   0.078629
DAYS_BIRTH                                  -0.002855
DAYS_EMPLOYED                               -0.034890
DAYS_REGISTRATION                            0.011857
DAYS_ID_PUBLISH                              0.008728
HOUR_APPR_PROCESS_START                      0.037332
EXT_SOURCE_2                                 0.050993
OBS_30_CNT_SOCIAL_CIRCLE                     0.001574
DEF_30_CNT_SOCIAL_CIRCLE                     0.000817
OBS_60_CNT_SOCIAL_CIRCLE                     0.001598
DEF_60_CNT_SOCIAL_CIRCLE                    -0.001396
AMT_REQ_CREDIT_BUREAU_HOUR                   0.000528
AMT_REQ_CREDIT_BUREAU_DAY                   -0.004753
```

```
AMT_REQ_CREDIT_BUREAU_WEEK                    -0.014248
AMT_REQ_CREDIT_BUREAU_MON                      1.000000
AMT_REQ_CREDIT_BUREAU_QRT                     -0.008160
AMT_REQ_CREDIT_BUREAU_YEAR                    -0.004889

                            AMT_REQ_CREDIT_BUREAU_QRT  \
SK_ID_CURR                                     0.001148
TARGET                                              NaN
CNT_CHILDREN                                  -0.007087
AMT_INCOME_TOTAL                               0.013128
AMT_CREDIT                                     0.017767
AMT_ANNUITY                                    0.011248
AMT_GOODS_PRICE                                0.018163
REGION_POPULATION_RELATIVE                    -0.001013
DAYS_BIRTH                                     0.011335
DAYS_EMPLOYED                                  0.015116
DAYS_REGISTRATION                              0.000527
DAYS_ID_PUBLISH                                0.007071
HOUR_APPR_PROCESS_START                       -0.000671
EXT_SOURCE_2                                  -0.002961
OBS_30_CNT_SOCIAL_CIRCLE                       0.003966
DEF_30_CNT_SOCIAL_CIRCLE                      -0.000960
OBS_60_CNT_SOCIAL_CIRCLE                       0.003793
DEF_60_CNT_SOCIAL_CIRCLE                      -0.000274
AMT_REQ_CREDIT_BUREAU_HOUR                    -0.003350
AMT_REQ_CREDIT_BUREAU_DAY                     -0.004749
AMT_REQ_CREDIT_BUREAU_WEEK                    -0.015466
AMT_REQ_CREDIT_BUREAU_MON                     -0.008160
AMT_REQ_CREDIT_BUREAU_QRT                      1.000000
AMT_REQ_CREDIT_BUREAU_YEAR                     0.074664

                            AMT_REQ_CREDIT_BUREAU_YEAR
SK_ID_CURR                                    0.004349
TARGET                                             NaN
CNT_CHILDREN                                 -0.042547
AMT_INCOME_TOTAL                              0.029536
AMT_CREDIT                                   -0.048866
AMT_ANNUITY                                  -0.010882
AMT_GOODS_PRICE                              -0.051266
REGION_POPULATION_RELATIVE                    0.001775
DAYS_BIRTH                                    0.072886
DAYS_EMPLOYED                                 0.052169
DAYS_REGISTRATION                             0.026639
DAYS_ID_PUBLISH                               0.035639
HOUR_APPR_PROCESS_START                      -0.030330
EXT_SOURCE_2                                 -0.021168
OBS_30_CNT_SOCIAL_CIRCLE                      0.031884
DEF_30_CNT_SOCIAL_CIRCLE                      0.018606
OBS_60_CNT_SOCIAL_CIRCLE                      0.032291
DEF_60_CNT_SOCIAL_CIRCLE                      0.018088
```

```
AMT_REQ_CREDIT_BUREAU_HOUR                    -0.004790
AMT_REQ_CREDIT_BUREAU_DAY                     -0.003798
AMT_REQ_CREDIT_BUREAU_WEEK                     0.019085
AMT_REQ_CREDIT_BUREAU_MON                     -0.004889
AMT_REQ_CREDIT_BUREAU_QRT                      0.074664
AMT_REQ_CREDIT_BUREAU_YEAR                     1.000000
```

```
#Convert the diagonal and below diagonal values of matrix to False,
Whereever False is there is replaced with NaN on execution
corr=corr.where(np.triu(np.ones(corr.shape), k=1).astype(np.bool))
corr
```

```
                            SK_ID_CURR  TARGET  CNT_CHILDREN  \
SK_ID_CURR                         NaN     NaN     -0.000716
TARGET                             NaN     NaN           NaN
CNT_CHILDREN                       NaN     NaN           NaN
AMT_INCOME_TOTAL                   NaN     NaN           NaN
AMT_CREDIT                         NaN     NaN           NaN
AMT_ANNUITY                        NaN     NaN           NaN
AMT_GOODS_PRICE                    NaN     NaN           NaN
REGION_POPULATION_RELATIVE         NaN     NaN           NaN
DAYS_BIRTH                         NaN     NaN           NaN
DAYS_EMPLOYED                      NaN     NaN           NaN
DAYS_REGISTRATION                  NaN     NaN           NaN
DAYS_ID_PUBLISH                    NaN     NaN           NaN
HOUR_APPR_PROCESS_START            NaN     NaN           NaN
EXT_SOURCE_2                       NaN     NaN           NaN
OBS_30_CNT_SOCIAL_CIRCLE           NaN     NaN           NaN
DEF_30_CNT_SOCIAL_CIRCLE           NaN     NaN           NaN
OBS_60_CNT_SOCIAL_CIRCLE           NaN     NaN           NaN
DEF_60_CNT_SOCIAL_CIRCLE           NaN     NaN           NaN
AMT_REQ_CREDIT_BUREAU_HOUR         NaN     NaN           NaN
AMT_REQ_CREDIT_BUREAU_DAY          NaN     NaN           NaN
AMT_REQ_CREDIT_BUREAU_WEEK         NaN     NaN           NaN
AMT_REQ_CREDIT_BUREAU_MON          NaN     NaN           NaN
AMT_REQ_CREDIT_BUREAU_QRT          NaN     NaN           NaN
AMT_REQ_CREDIT_BUREAU_YEAR         NaN     NaN           NaN

                            AMT_INCOME_TOTAL  AMT_CREDIT  AMT_ANNUITY
\
SK_ID_CURR                          0.001739   -0.000342     0.000068

TARGET                                   NaN         NaN          NaN

CNT_CHILDREN                        0.027397    0.003081     0.020905

AMT_INCOME_TOTAL                         NaN    0.342799     0.418953

AMT_CREDIT                               NaN         NaN     0.771309
```

```
AMT_ANNUITY                                    NaN          NaN           NaN
AMT_GOODS_PRICE                                NaN          NaN           NaN
REGION_POPULATION_RELATIVE                     NaN          NaN           NaN
DAYS_BIRTH                                     NaN          NaN           NaN
DAYS_EMPLOYED                                  NaN          NaN           NaN
DAYS_REGISTRATION                              NaN          NaN           NaN
DAYS_ID_PUBLISH                                NaN          NaN           NaN
HOUR_APPR_PROCESS_START                        NaN          NaN           NaN
EXT_SOURCE_2                                   NaN          NaN           NaN
OBS_30_CNT_SOCIAL_CIRCLE                       NaN          NaN           NaN
DEF_30_CNT_SOCIAL_CIRCLE                       NaN          NaN           NaN
OBS_60_CNT_SOCIAL_CIRCLE                       NaN          NaN           NaN
DEF_60_CNT_SOCIAL_CIRCLE                       NaN          NaN           NaN
AMT_REQ_CREDIT_BUREAU_HOUR                     NaN          NaN           NaN
AMT_REQ_CREDIT_BUREAU_DAY                      NaN          NaN           NaN
AMT_REQ_CREDIT_BUREAU_WEEK                     NaN          NaN           NaN
AMT_REQ_CREDIT_BUREAU_MON                      NaN          NaN           NaN
AMT_REQ_CREDIT_BUREAU_QRT                      NaN          NaN           NaN
AMT_REQ_CREDIT_BUREAU_YEAR                     NaN          NaN           NaN

                            AMT_GOODS_PRICE
REGION_POPULATION_RELATIVE  \
SK_ID_CURR                        -0.000205
0.000360
TARGET                                  NaN
NaN
CNT_CHILDREN                      -0.000525                    -
0.024363
AMT_INCOME_TOTAL                   0.349462
```

```
0.167851
AMT_CREDIT                         0.987250
0.100604
AMT_ANNUITY                        0.776686
0.120988
AMT_GOODS_PRICE                         NaN
0.103827
REGION_POPULATION_RELATIVE              NaN
NaN
DAYS_BIRTH                              NaN
NaN
DAYS_EMPLOYED                           NaN
NaN
DAYS_REGISTRATION                       NaN
NaN
DAYS_ID_PUBLISH                         NaN
NaN
HOUR_APPR_PROCESS_START                 NaN
NaN
EXT_SOURCE_2                            NaN
NaN
OBS_30_CNT_SOCIAL_CIRCLE                NaN
NaN
DEF_30_CNT_SOCIAL_CIRCLE                NaN
NaN
OBS_60_CNT_SOCIAL_CIRCLE                NaN
NaN
DEF_60_CNT_SOCIAL_CIRCLE                NaN
NaN
AMT_REQ_CREDIT_BUREAU_HOUR              NaN
NaN
AMT_REQ_CREDIT_BUREAU_DAY               NaN
NaN
AMT_REQ_CREDIT_BUREAU_WEEK              NaN
NaN
AMT_REQ_CREDIT_BUREAU_MON               NaN
NaN
AMT_REQ_CREDIT_BUREAU_QRT               NaN
NaN
AMT_REQ_CREDIT_BUREAU_YEAR              NaN
NaN

                            DAYS_BIRTH  DAYS_EMPLOYED
DAYS_REGISTRATION  \
SK_ID_CURR                    0.001346       0.001744
0.001475
TARGET                             NaN            NaN
NaN
CNT_CHILDREN                 -0.336966      -0.245174          -
0.185792
```

```
AMT_INCOME_TOTAL              -0.062609      -0.140392           -
0.064937
AMT_CREDIT                     0.047378      -0.070104           -
0.013477
AMT_ANNUITY                   -0.012263      -0.104978           -
0.039436
AMT_GOODS_PRICE                0.044565      -0.068609           -
0.015916
REGION_POPULATION_RELATIVE     0.025244      -0.007198
0.052083
DAYS_BIRTH                          NaN       0.626114
0.333151
DAYS_EMPLOYED                       NaN            NaN
0.214511
DAYS_REGISTRATION                   NaN            NaN
NaN
DAYS_ID_PUBLISH                     NaN            NaN
NaN
HOUR_APPR_PROCESS_START             NaN            NaN
NaN
EXT_SOURCE_2                        NaN            NaN
NaN
OBS_30_CNT_SOCIAL_CIRCLE            NaN            NaN
NaN
DEF_30_CNT_SOCIAL_CIRCLE            NaN            NaN
NaN
OBS_60_CNT_SOCIAL_CIRCLE            NaN            NaN
NaN
DEF_60_CNT_SOCIAL_CIRCLE            NaN            NaN
NaN
AMT_REQ_CREDIT_BUREAU_HOUR          NaN            NaN
NaN
AMT_REQ_CREDIT_BUREAU_DAY           NaN            NaN
NaN
AMT_REQ_CREDIT_BUREAU_WEEK          NaN            NaN
NaN
AMT_REQ_CREDIT_BUREAU_MON           NaN            NaN
NaN
AMT_REQ_CREDIT_BUREAU_QRT           NaN            NaN
NaN
AMT_REQ_CREDIT_BUREAU_YEAR          NaN            NaN
NaN

                            DAYS_ID_PUBLISH
HOUR_APPR_PROCESS_START  \
SK_ID_CURR                         0.000077                   -0.000115

TARGET                                  NaN                         NaN

CNT_CHILDREN                       0.028751                   -0.005244
```

AMT_INCOME_TOTAL	-0.022896	0.076743
AMT_CREDIT	0.001464	0.053619
AMT_ANNUITY	-0.014113	0.053589
AMT_GOODS_PRICE	0.003649	0.062766
REGION_POPULATION_RELATIVE	0.001071	0.172814
DAYS_BIRTH	0.271314	-0.095916
DAYS_EMPLOYED	0.276663	-0.095455
DAYS_REGISTRATION	0.100236	0.008044
DAYS_ID_PUBLISH	NaN	-0.033980
HOUR_APPR_PROCESS_START	NaN	NaN
EXT_SOURCE_2	NaN	NaN
OBS_30_CNT_SOCIAL_CIRCLE	NaN	NaN
DEF_30_CNT_SOCIAL_CIRCLE	NaN	NaN
OBS_60_CNT_SOCIAL_CIRCLE	NaN	NaN
DEF_60_CNT_SOCIAL_CIRCLE	NaN	NaN
AMT_REQ_CREDIT_BUREAU_HOUR	NaN	NaN
AMT_REQ_CREDIT_BUREAU_DAY	NaN	NaN
AMT_REQ_CREDIT_BUREAU_WEEK	NaN	NaN
AMT_REQ_CREDIT_BUREAU_MON	NaN	NaN
AMT_REQ_CREDIT_BUREAU_QRT	NaN	NaN
AMT_REQ_CREDIT_BUREAU_YEAR	NaN	NaN

	EXT_SOURCE_2	OBS_30_CNT_SOCIAL_CIRCLE \
SK_ID_CURR	0.001589	-0.000695
TARGET	NaN	NaN
CNT_CHILDREN	-0.015455	0.014471

```
AMT_INCOME_TOTAL              0.139598                 -0.027828
AMT_CREDIT                    0.129140                 -0.000914
AMT_ANNUITY                   0.126615                 -0.013091
AMT_GOODS_PRICE               0.135694                 -0.000707
REGION_POPULATION_RELATIVE    0.198493                 -0.012107
DAYS_BIRTH                    0.078158                 -0.007726
DAYS_EMPLOYED                -0.028530                  0.007272
DAYS_REGISTRATION             0.052503                 -0.008315
DAYS_ID_PUBLISH               0.041703                  0.012326
HOUR_APPR_PROCESS_START       0.157221                 -0.008013
EXT_SOURCE_2                       NaN                 -0.021569
OBS_30_CNT_SOCIAL_CIRCLE           NaN                       NaN
DEF_30_CNT_SOCIAL_CIRCLE           NaN                       NaN
OBS_60_CNT_SOCIAL_CIRCLE           NaN                       NaN
DEF_60_CNT_SOCIAL_CIRCLE           NaN                       NaN
AMT_REQ_CREDIT_BUREAU_HOUR         NaN                       NaN
AMT_REQ_CREDIT_BUREAU_DAY          NaN                       NaN
AMT_REQ_CREDIT_BUREAU_WEEK         NaN                       NaN
AMT_REQ_CREDIT_BUREAU_MON          NaN                       NaN
AMT_REQ_CREDIT_BUREAU_QRT          NaN                       NaN
AMT_REQ_CREDIT_BUREAU_YEAR         NaN                       NaN

                            DEF_30_CNT_SOCIAL_CIRCLE  \
SK_ID_CURR                                  0.000556
TARGET                                           NaN
CNT_CHILDREN                               -0.002246
AMT_INCOME_TOTAL                           -0.027621
AMT_CREDIT                                 -0.019851
AMT_ANNUITY                                -0.022535
AMT_GOODS_PRICE                            -0.021025
REGION_POPULATION_RELATIVE                  0.005941
DAYS_BIRTH                                  0.003057
DAYS_EMPLOYED                               0.019379
DAYS_REGISTRATION                          -0.001213
DAYS_ID_PUBLISH                            -0.000403
HOUR_APPR_PROCESS_START                    -0.005718
EXT_SOURCE_2                               -0.027427
OBS_30_CNT_SOCIAL_CIRCLE                    0.329206
DEF_30_CNT_SOCIAL_CIRCLE                         NaN
OBS_60_CNT_SOCIAL_CIRCLE                         NaN
DEF_60_CNT_SOCIAL_CIRCLE                         NaN
AMT_REQ_CREDIT_BUREAU_HOUR                       NaN
AMT_REQ_CREDIT_BUREAU_DAY                        NaN
AMT_REQ_CREDIT_BUREAU_WEEK                       NaN
AMT_REQ_CREDIT_BUREAU_MON                        NaN
AMT_REQ_CREDIT_BUREAU_QRT                        NaN
AMT_REQ_CREDIT_BUREAU_YEAR                       NaN

                            OBS_60_CNT_SOCIAL_CIRCLE  \
SK_ID_CURR                                 -0.000741
```

```
TARGET                                           NaN
CNT_CHILDREN                                0.014137
AMT_INCOME_TOTAL                           -0.027690
AMT_CREDIT                                 -0.000892
AMT_ANNUITY                                -0.012893
AMT_GOODS_PRICE                            -0.000723
REGION_POPULATION_RELATIVE                 -0.011591
DAYS_BIRTH                                 -0.007316
DAYS_EMPLOYED                               0.007453
DAYS_REGISTRATION                          -0.008247
DAYS_ID_PUBLISH                             0.012810
HOUR_APPR_PROCESS_START                    -0.007956
EXT_SOURCE_2                               -0.021224
OBS_30_CNT_SOCIAL_CIRCLE                    0.998508
DEF_30_CNT_SOCIAL_CIRCLE                    0.331336
OBS_60_CNT_SOCIAL_CIRCLE                         NaN
DEF_60_CNT_SOCIAL_CIRCLE                         NaN
AMT_REQ_CREDIT_BUREAU_HOUR                       NaN
AMT_REQ_CREDIT_BUREAU_DAY                        NaN
AMT_REQ_CREDIT_BUREAU_WEEK                       NaN
AMT_REQ_CREDIT_BUREAU_MON                        NaN
AMT_REQ_CREDIT_BUREAU_QRT                        NaN
AMT_REQ_CREDIT_BUREAU_YEAR                       NaN

                            DEF_60_CNT_SOCIAL_CIRCLE  \
SK_ID_CURR                                  0.002382
TARGET                                           NaN
CNT_CHILDREN                               -0.002172
AMT_INCOME_TOTAL                           -0.027593
AMT_CREDIT                                 -0.022225
AMT_ANNUITY                                -0.023385
AMT_GOODS_PRICE                            -0.023171
REGION_POPULATION_RELATIVE                  0.002255
DAYS_BIRTH                                  0.000990
DAYS_EMPLOYED                               0.016383
DAYS_REGISTRATION                          -0.002652
DAYS_ID_PUBLISH                            -0.002491
HOUR_APPR_PROCESS_START                    -0.008826
EXT_SOURCE_2                               -0.029722
OBS_30_CNT_SOCIAL_CIRCLE                    0.253000
DEF_30_CNT_SOCIAL_CIRCLE                    0.859332
OBS_60_CNT_SOCIAL_CIRCLE                    0.254970
DEF_60_CNT_SOCIAL_CIRCLE                         NaN
AMT_REQ_CREDIT_BUREAU_HOUR                       NaN
AMT_REQ_CREDIT_BUREAU_DAY                        NaN
AMT_REQ_CREDIT_BUREAU_WEEK                       NaN
AMT_REQ_CREDIT_BUREAU_MON                        NaN
AMT_REQ_CREDIT_BUREAU_QRT                        NaN
AMT_REQ_CREDIT_BUREAU_YEAR                       NaN
```

	AMT_REQ_CREDIT_BUREAU_HOUR \
SK_ID_CURR	-0.001879
TARGET	NaN
CNT_CHILDREN	-0.000432
AMT_INCOME_TOTAL	0.001417
AMT_CREDIT	-0.003734
AMT_ANNUITY	0.003148
AMT_GOODS_PRICE	-0.003116
REGION_POPULATION_RELATIVE	-0.002265
DAYS_BIRTH	-0.004461
DAYS_EMPLOYED	-0.004460
DAYS_REGISTRATION	0.002730
DAYS_ID_PUBLISH	-0.004797
HOUR_APPR_PROCESS_START	-0.015832
EXT_SOURCE_2	-0.003918
OBS_30_CNT_SOCIAL_CIRCLE	0.000420
DEF_30_CNT_SOCIAL_CIRCLE	-0.001472
OBS_60_CNT_SOCIAL_CIRCLE	0.000328
DEF_60_CNT_SOCIAL_CIRCLE	-0.002132
AMT_REQ_CREDIT_BUREAU_HOUR	NaN
AMT_REQ_CREDIT_BUREAU_DAY	NaN
AMT_REQ_CREDIT_BUREAU_WEEK	NaN
AMT_REQ_CREDIT_BUREAU_MON	NaN
AMT_REQ_CREDIT_BUREAU_QRT	NaN
AMT_REQ_CREDIT_BUREAU_YEAR	NaN

	AMT_REQ_CREDIT_BUREAU_DAY \
SK_ID_CURR	-0.001725
TARGET	NaN
CNT_CHILDREN	0.000648
AMT_INCOME_TOTAL	0.007862
AMT_CREDIT	0.004409
AMT_ANNUITY	0.002392
AMT_GOODS_PRICE	0.004820
REGION_POPULATION_RELATIVE	0.001969
DAYS_BIRTH	-0.002772
DAYS_EMPLOYED	-0.000874
DAYS_REGISTRATION	0.000035
DAYS_ID_PUBLISH	0.000198
HOUR_APPR_PROCESS_START	0.003847
EXT_SOURCE_2	0.001597
OBS_30_CNT_SOCIAL_CIRCLE	-0.001940
DEF_30_CNT_SOCIAL_CIRCLE	-0.001712
OBS_60_CNT_SOCIAL_CIRCLE	-0.002024
DEF_60_CNT_SOCIAL_CIRCLE	-0.001742
AMT_REQ_CREDIT_BUREAU_HOUR	0.229065
AMT_REQ_CREDIT_BUREAU_DAY	NaN
AMT_REQ_CREDIT_BUREAU_WEEK	NaN
AMT_REQ_CREDIT_BUREAU_MON	NaN
AMT_REQ_CREDIT_BUREAU_QRT	NaN

```
AMT_REQ_CREDIT_BUREAU_YEAR                               NaN

                            AMT_REQ_CREDIT_BUREAU_WEEK  \
SK_ID_CURR                                    0.002524
TARGET                                             NaN
CNT_CHILDREN                                 -0.001632
AMT_INCOME_TOTAL                              0.006234
AMT_CREDIT                                   -0.001883
AMT_ANNUITY                                   0.012681
AMT_GOODS_PRICE                              -0.001597
REGION_POPULATION_RELATIVE                   -0.002480
DAYS_BIRTH                                    0.001069
DAYS_EMPLOYED                                 0.002223
DAYS_REGISTRATION                             0.001431
DAYS_ID_PUBLISH                               0.001655
HOUR_APPR_PROCESS_START                      -0.002652
EXT_SOURCE_2                                  0.001420
OBS_30_CNT_SOCIAL_CIRCLE                      0.000383
DEF_30_CNT_SOCIAL_CIRCLE                     -0.001882
OBS_60_CNT_SOCIAL_CIRCLE                      0.000468
DEF_60_CNT_SOCIAL_CIRCLE                     -0.002374
AMT_REQ_CREDIT_BUREAU_HOUR                    0.004576
AMT_REQ_CREDIT_BUREAU_DAY                     0.220087
AMT_REQ_CREDIT_BUREAU_WEEK                         NaN
AMT_REQ_CREDIT_BUREAU_MON                          NaN
AMT_REQ_CREDIT_BUREAU_QRT                          NaN
AMT_REQ_CREDIT_BUREAU_YEAR                         NaN

                            AMT_REQ_CREDIT_BUREAU_MON  \
SK_ID_CURR                                   0.000054
TARGET                                            NaN
CNT_CHILDREN                                -0.010455
AMT_INCOME_TOTAL                             0.061470
AMT_CREDIT                                   0.054071
AMT_ANNUITY                                  0.038181
AMT_GOODS_PRICE                              0.055850
REGION_POPULATION_RELATIVE                   0.078629
DAYS_BIRTH                                  -0.002855
DAYS_EMPLOYED                               -0.034890
DAYS_REGISTRATION                            0.011857
DAYS_ID_PUBLISH                              0.008728
HOUR_APPR_PROCESS_START                      0.037332
EXT_SOURCE_2                                 0.050993
OBS_30_CNT_SOCIAL_CIRCLE                     0.001574
DEF_30_CNT_SOCIAL_CIRCLE                     0.000817
OBS_60_CNT_SOCIAL_CIRCLE                     0.001598
DEF_60_CNT_SOCIAL_CIRCLE                    -0.001396
AMT_REQ_CREDIT_BUREAU_HOUR                   0.000528
AMT_REQ_CREDIT_BUREAU_DAY                   -0.004753
AMT_REQ_CREDIT_BUREAU_WEEK                  -0.014248
```

```
AMT_REQ_CREDIT_BUREAU_MON                              NaN
AMT_REQ_CREDIT_BUREAU_QRT                              NaN
AMT_REQ_CREDIT_BUREAU_YEAR                             NaN

                            AMT_REQ_CREDIT_BUREAU_QRT  \
SK_ID_CURR                                   0.001148
TARGET                                            NaN
CNT_CHILDREN                                -0.007087
AMT_INCOME_TOTAL                             0.013128
AMT_CREDIT                                   0.017767
AMT_ANNUITY                                  0.011248
AMT_GOODS_PRICE                              0.018163
REGION_POPULATION_RELATIVE                  -0.001013
DAYS_BIRTH                                   0.011335
DAYS_EMPLOYED                                0.015116
DAYS_REGISTRATION                            0.000527
DAYS_ID_PUBLISH                              0.007071
HOUR_APPR_PROCESS_START                     -0.000671
EXT_SOURCE_2                                -0.002961
OBS_30_CNT_SOCIAL_CIRCLE                     0.003966
DEF_30_CNT_SOCIAL_CIRCLE                    -0.000960
OBS_60_CNT_SOCIAL_CIRCLE                     0.003793
DEF_60_CNT_SOCIAL_CIRCLE                    -0.000274
AMT_REQ_CREDIT_BUREAU_HOUR                  -0.003350
AMT_REQ_CREDIT_BUREAU_DAY                   -0.004749
AMT_REQ_CREDIT_BUREAU_WEEK                  -0.015466
AMT_REQ_CREDIT_BUREAU_MON                   -0.008160
AMT_REQ_CREDIT_BUREAU_QRT                         NaN
AMT_REQ_CREDIT_BUREAU_YEAR                        NaN

                            AMT_REQ_CREDIT_BUREAU_YEAR
SK_ID_CURR                                    0.004349
TARGET                                             NaN
CNT_CHILDREN                                 -0.042547
AMT_INCOME_TOTAL                              0.029536
AMT_CREDIT                                   -0.048866
AMT_ANNUITY                                  -0.010882
AMT_GOODS_PRICE                              -0.051266
REGION_POPULATION_RELATIVE                    0.001775
DAYS_BIRTH                                    0.072886
DAYS_EMPLOYED                                 0.052169
DAYS_REGISTRATION                             0.026639
DAYS_ID_PUBLISH                               0.035639
HOUR_APPR_PROCESS_START                      -0.030330
EXT_SOURCE_2                                 -0.021168
OBS_30_CNT_SOCIAL_CIRCLE                      0.031884
DEF_30_CNT_SOCIAL_CIRCLE                      0.018606
OBS_60_CNT_SOCIAL_CIRCLE                      0.032291
DEF_60_CNT_SOCIAL_CIRCLE                      0.018088
AMT_REQ_CREDIT_BUREAU_HOUR                   -0.004790
```

```
AMT_REQ_CREDIT_BUREAU_DAY                       -0.003798
AMT_REQ_CREDIT_BUREAU_WEEK                       0.019085
AMT_REQ_CREDIT_BUREAU_MON                       -0.004889
AMT_REQ_CREDIT_BUREAU_QRT                        0.074664
AMT_REQ_CREDIT_BUREAU_YEAR                            NaN
```

```python
#convert it to dataframe
corrdf = corr.unstack().reset_index()
corrdf.head()
```

```
      level_0           level_1   0
0  SK_ID_CURR        SK_ID_CURR NaN
1  SK_ID_CURR            TARGET NaN
2  SK_ID_CURR      CNT_CHILDREN NaN
3  SK_ID_CURR  AMT_INCOME_TOTAL NaN
4  SK_ID_CURR        AMT_CREDIT NaN
```

```python
#Changing the names of solumns
corrdf.columns=['VAR1','VAR2','Correlation']
corrdf.head()
```

```
         VAR1              VAR2  Correlation
0  SK_ID_CURR        SK_ID_CURR          NaN
1  SK_ID_CURR            TARGET          NaN
2  SK_ID_CURR      CNT_CHILDREN          NaN
3  SK_ID_CURR  AMT_INCOME_TOTAL          NaN
4  SK_ID_CURR        AMT_CREDIT          NaN
```

```python
#Drop the columns having the missing data
corrdf.dropna(subset = ['Correlation'], inplace = True)
corrdf.head()
```

```
                VAR1          VAR2  Correlation
48      CNT_CHILDREN    SK_ID_CURR    -0.000716
72  AMT_INCOME_TOTAL    SK_ID_CURR     0.001739
74  AMT_INCOME_TOTAL  CNT_CHILDREN     0.027397
96        AMT_CREDIT    SK_ID_CURR    -0.000342
98        AMT_CREDIT  CNT_CHILDREN     0.003081
```

```python
#Rounding off the values
corrdf['Correlation'] = round(corrdf['Correlation'], 2)
corrdf.head()
```

```
                VAR1          VAR2  Correlation
48      CNT_CHILDREN    SK_ID_CURR        -0.00
72  AMT_INCOME_TOTAL    SK_ID_CURR         0.00
74  AMT_INCOME_TOTAL  CNT_CHILDREN         0.03
96        AMT_CREDIT    SK_ID_CURR        -0.00
98        AMT_CREDIT  CNT_CHILDREN         0.00
```

```python
# Since we see correlation as an absolute value, we are converting it
into absolute value
```

```
corrdf['Correlation'] = corrdf['Correlation'].abs()
corrdf.head()
```

```
                VAR1          VAR2  Correlation
48      CNT_CHILDREN    SK_ID_CURR         0.00
72  AMT_INCOME_TOTAL    SK_ID_CURR         0.00
74  AMT_INCOME_TOTAL  CNT_CHILDREN         0.03
96        AMT_CREDIT    SK_ID_CURR         0.00
98        AMT_CREDIT  CNT_CHILDREN         0.00
```

```
#Sorting the correlation values
corrdf.sort_values(by = 'Correlation', ascending = False).head(10)
```

```
                         VAR1                       VAR2  Correlation
398  OBS_60_CNT_SOCIAL_CIRCLE  OBS_30_CNT_SOCIAL_CIRCLE         1.00
148           AMT_GOODS_PRICE                 AMT_CREDIT         0.99
423  DEF_60_CNT_SOCIAL_CIRCLE  DEF_30_CNT_SOCIAL_CIRCLE         0.86
149           AMT_GOODS_PRICE                AMT_ANNUITY         0.78
124               AMT_ANNUITY                 AMT_CREDIT         0.77
224             DAYS_EMPLOYED                 DAYS_BIRTH         0.63
123               AMT_ANNUITY           AMT_INCOME_TOTAL         0.42
147           AMT_GOODS_PRICE           AMT_INCOME_TOTAL         0.35
194                DAYS_BIRTH               CNT_CHILDREN         0.34
99                 AMT_CREDIT           AMT_INCOME_TOTAL         0.34
```

We can see that for Target_0 dataframe, Social circle for 30 days and 60 days are most correlated and Goods price and Loan amount credit are highly correlated.Then we have Goods price and amount annuity on 4th place

```
#For target_1 dataframe we perform the same operations to find
correlation
corr = target_1.corr()

corr = corr.where(np.triu(np.ones(corr.shape), k=1).astype(np.bool))
corrdf = corr.unstack().reset_index()
corrdf.columns = ['VAR1', 'VAR2', 'Correlation']
corrdf.dropna(subset = ['Correlation'], inplace = True)
corrdf['Correlation'] = round(corrdf['Correlation'], 2)
# Since we see correlation as an absolute value, we are converting it
into absolute valucorrdf['Correlation_abs'] =
corrdf0['Correlation'].abs()
corrdf.sort_values(by = 'Correlation', ascending = False).head(10)
```

```
                         VAR1                       VAR2  Correlation
398  OBS_60_CNT_SOCIAL_CIRCLE  OBS_30_CNT_SOCIAL_CIRCLE         1.00
148           AMT_GOODS_PRICE                 AMT_CREDIT         0.98
423  DEF_60_CNT_SOCIAL_CIRCLE  DEF_30_CNT_SOCIAL_CIRCLE         0.87
124               AMT_ANNUITY                 AMT_CREDIT         0.75
149           AMT_GOODS_PRICE                AMT_ANNUITY         0.75
224             DAYS_EMPLOYED                 DAYS_BIRTH         0.58
399  OBS_60_CNT_SOCIAL_CIRCLE  DEF_30_CNT_SOCIAL_CIRCLE         0.34
```

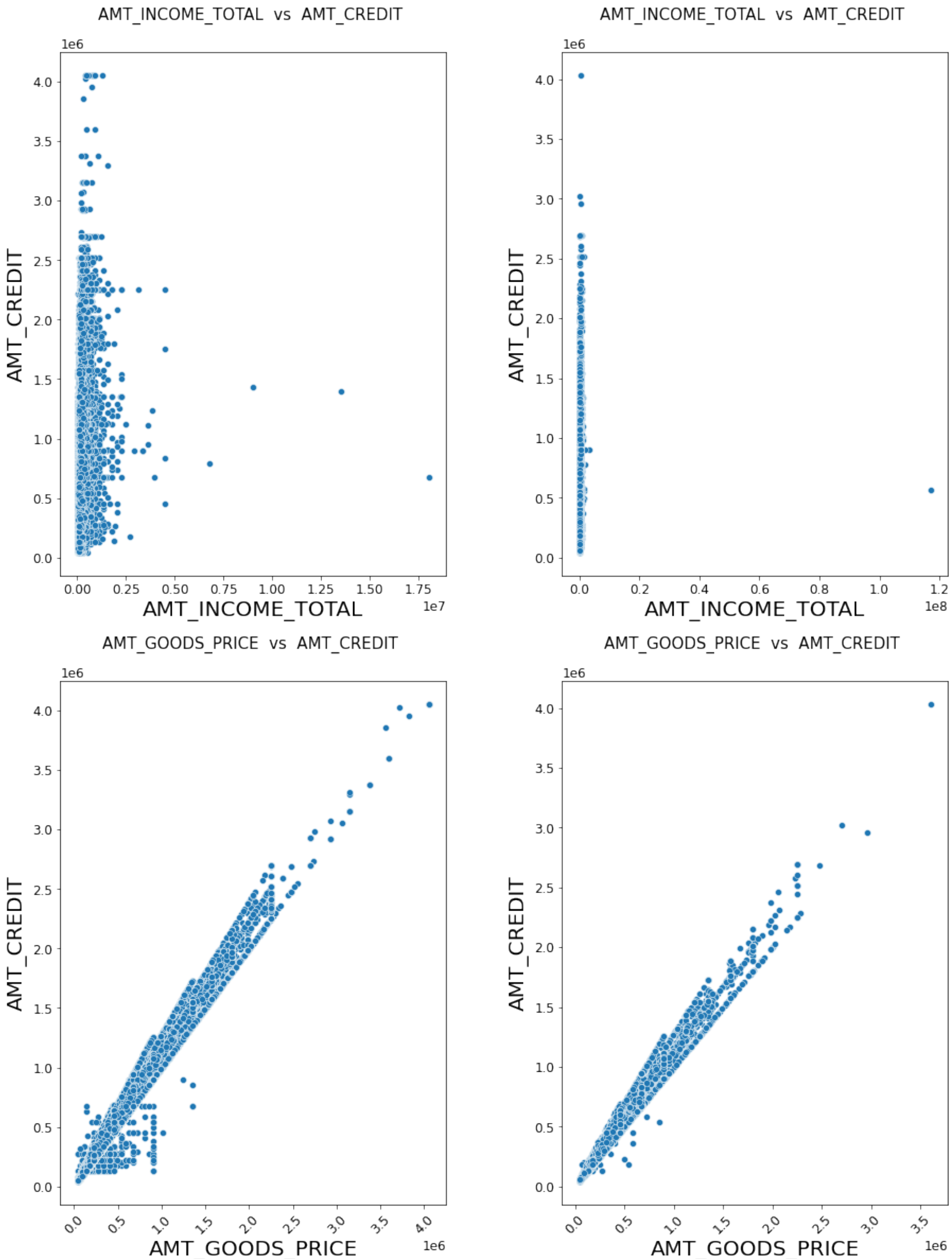
```
374  DEF_30_CNT_SOCIAL_CIRCLE  OBS_30_CNT_SOCIAL_CIRCLE         0.33
248         DAYS_REGISTRATION                DAYS_BIRTH         0.29
422  DEF_60_CNT_SOCIAL_CIRCLE  OBS_30_CNT_SOCIAL_CIRCLE         0.26
```

From the observation above we can say that for target_1 dataframe Goods price and loan credit amount are most correlated next to social circle observations for different days. So the variables correlated in target_0 dataframe and target_1 dataframe are same with slightly verying correlation values

Bivariate Analysis for target 0 and target 1

Income vs Credit, Goods price vs Credit

```
#Scatter plot for numeric columns

plt.figure(figsize = (15, 20))
plt.subplots_adjust(wspace=0.3)


plt.subplot(2,2,1)
sns.scatterplot(target_0.AMT_INCOME_TOTAL,target_0.AMT_CREDIT)
plt.xlabel('AMT_INCOME_TOTAL')
plt.ylabel('AMT_CREDIT')
plt.title('AMT_INCOME_TOTAL  vs  AMT_CREDIT ')

plt.subplot(2,2,2)
sns.scatterplot(target_1.AMT_INCOME_TOTAL,target_1.AMT_CREDIT)
plt.xlabel('AMT_INCOME_TOTAL')
plt.ylabel('AMT_CREDIT')
plt.title('AMT_INCOME_TOTAL  vs  AMT_CREDIT ')


plt.subplot(2,2,3)
sns.scatterplot(target_0.AMT_GOODS_PRICE,target_0.AMT_CREDIT)
plt.xlabel('AMT_GOODS_PRICE')
plt.ylabel('AMT_CREDIT')
plt.title('AMT_GOODS_PRICE  vs  AMT_CREDIT ')
plt.xticks(rotation = 45)

plt.subplot(2,2,4)
sns.scatterplot(target_1.AMT_GOODS_PRICE,target_1.AMT_CREDIT)
plt.xlabel('AMT_GOODS_PRICE')
plt.ylabel('AMT_CREDIT')
plt.title('AMT_GOODS_PRICE  vs  AMT_CREDIT ')
plt.xticks(rotation = 45)
plt.show()
```



Those who have paid the loan amount on/within time are more likely to get higher credits than those who didnt pay/did late payments.People who have higher goods price and have made payments on time have higher credits than those with higher goods price but didnt pay loan.

Numerical categorical analysis

Income range- Gender

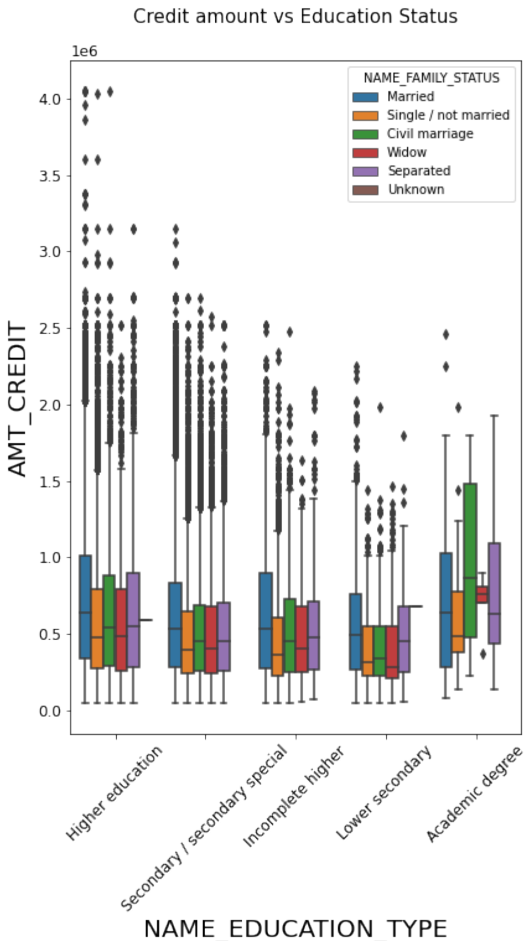
```
# Numeric variable analysis for target_0 & target_1 dataframe
plt.figure(figsize = (15, 8))
plt.subplot(2, 2, 1)
plt.title('Target_0:Income Range b/w Male and Female')

sns.countplot(x='AMT_INCOME_RANGE', hue='CODE_GENDER', data=target_0,
palette='rocket')

# subplot 2
plt.subplot(2, 2, 2)
plt.title('Target_1:Income Range b/w Male and Female')

sns.countplot(x='AMT_INCOME_RANGE', hue='CODE_GENDER',
data=target_1,palette='rocket')
plt.show()
```

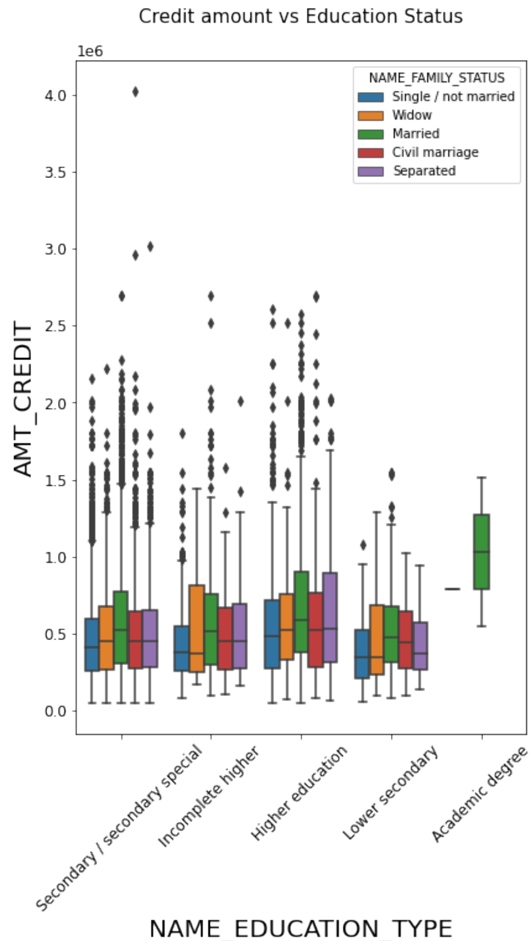
We can see that Females with low income don’t have any payment issues.

Credit amount vs Education Status

```
# Box plotting for Credit amount

plt.figure(figsize=(15,10))
plt.subplots_adjust(wspace=0.3)

plt.subplot(121)
sns.boxplot(data =target_0, x='NAME_EDUCATION_TYPE',y='AMT_CREDIT',
hue ='NAME_FAMILY_STATUS',orient='v]')
plt.title('Credit amount vs Education Status')
plt.xticks(rotation=45)

plt.subplot(122)
sns.boxplot(data =target_1, x='NAME_EDUCATION_TYPE',y='AMT_CREDIT',
hue ='NAME_FAMILY_STATUS',orient='v]')
```

```
plt.title('Credit amount vs Education Status')
plt.xticks(rotation=45)
plt.show()
```





From the above plot,we can see that

1.Some of the highly educated, married person are having credits higher than those who have done lower secondary eduction.

2.Those with higher eduction have higher credits and are more likely to make payments on time.

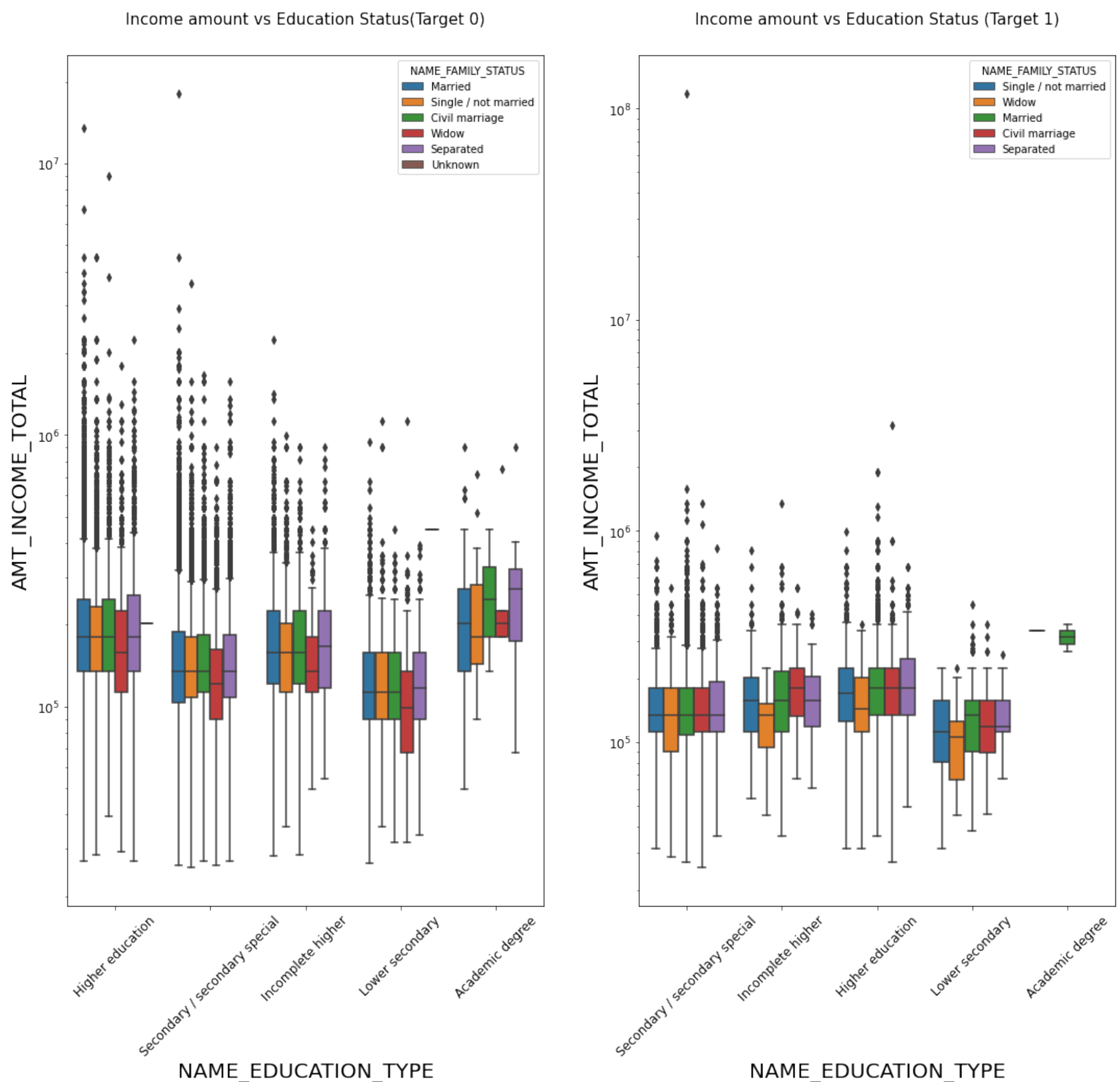
3.More number of outliers are seen in higher education.

4.The people with secondary and secndary special eduction are less likely to make payments on time.

Income vs Education Status

```
# Box plotting for Income amount in logarithmic scale

plt.figure(figsize=(18,15))
```

```
plt.subplot(1,2,1)
plt.yscale('log')
sns.boxplot(data =target_0,
x='NAME_EDUCATION_TYPE',y='AMT_INCOME_TOTAL', hue
='NAME_FAMILY_STATUS',orient='v')
plt.title('Income amount vs Education Status(Target 0)')
plt.xticks(rotation=45)



plt.subplot(1,2,2)
plt.yscale('log')
sns.boxplot(data =target_1,
x='NAME_EDUCATION_TYPE',y='AMT_INCOME_TOTAL', hue
='NAME_FAMILY_STATUS',orient='v')
plt.title('Income amount vs Education Status (Target 1)')
plt.xticks(rotation=45)
plt.show()
```



From the above plots,

1.we can see that Higher education has many outliers.

2.People with higher eductaion have higher income and dont have difficulties in making loan payment.

3.People with higher education who ave lesser income are unable to pay the loan.

Hence we can conclude that,people with Higher income are most likely to make payments.

Reading the previous application

```
#Reading the data from file previous_application
df1=pd.read_csv("previous_application.csv")
df1.head()
```

```
   SK_ID_PREV  SK_ID_CURR NAME_CONTRACT_TYPE  AMT_ANNUITY
AMT_APPLICATION  \
0     2030495      271877     Consumer loans     1730.430
17145.0
1     2802425      108129         Cash loans    25188.615
607500.0
2     2523466      122040         Cash loans    15060.735
112500.0
3     2819243      176158         Cash loans    47041.335
450000.0
4     1784265      202054         Cash loans    31924.395
337500.0

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

   HOUR_APPR_PROCESS_START FLAG_LAST_APPL_PER_CONTRACT  \
0                       15                           Y
1                       11                           Y
2                       11                           Y
3                        7                           Y
4                        9                           Y

   NFLAG_LAST_APPL_IN_DAY  RATE_DOWN_PAYMENT  RATE_INTEREST_PRIMARY  \
0                       1                0.0               0.182832
1                       1                NaN                    NaN
2                       1                NaN                    NaN
3                       1                NaN                    NaN
4                       1                NaN                    NaN

   RATE_INTEREST_PRIVILEGED NAME_CASH_LOAN_PURPOSE
NAME_CONTRACT_STATUS  \
0                  0.867336                    XAP
Approved
1                       NaN                    XNA
Approved
2                       NaN                    XNA
Approved
3                       NaN                    XNA
Approved
```

```
4                       NaN                Repairs
Refused

   DAYS_DECISION      NAME_PAYMENT_TYPE CODE_REJECT_REASON
NAME_TYPE_SUITE  \
0            -73  Cash through the bank                XAP
NaN
1           -164                    XNA                XAP
Unaccompanied
2           -301  Cash through the bank                XAP  Spouse,
partner
3           -512  Cash through the bank                XAP
NaN
4           -781  Cash through the bank                 HC
NaN

  NAME_CLIENT_TYPE NAME_GOODS_CATEGORY NAME_PORTFOLIO
NAME_PRODUCT_TYPE  \
0         Repeater              Mobile            POS
XNA
1         Repeater                 XNA           Cash              x-
sell
2         Repeater                 XNA           Cash              x-
sell
3         Repeater                 XNA           Cash              x-
sell
4         Repeater                 XNA           Cash           walk-
in

              CHANNEL_TYPE  SELLERPLACE_AREA NAME_SELLER_INDUSTRY  \
0             Country-wide                35         Connectivity
1           Contact center                -1                  XNA
2  Credit and cash offices                -1                  XNA
3  Credit and cash offices                -1                  XNA
4  Credit and cash offices                -1                  XNA

   CNT_PAYMENT NAME_YIELD_GROUP       PRODUCT_COMBINATION
DAYS_FIRST_DRAWING  \
0         12.0           middle  POS mobile with interest
365243.0
1         36.0       low_action          Cash X-Sell: low
365243.0
2         12.0             high         Cash X-Sell: high
365243.0
3         12.0           middle       Cash X-Sell: middle
365243.0
4         24.0             high         Cash Street: high
NaN

   DAYS_FIRST_DUE  DAYS_LAST_DUE_1ST_VERSION  DAYS_LAST_DUE
```

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

   NFLAG_INSURED_ON_APPROVAL
0                        0.0
1                        1.0
2                        1.0
3                        1.0
4                        NaN
```

```
# Removing the column values of 'XNA' and 'XAP'

df1=df1.drop(df1[df1['NAME_CASH_LOAN_PURPOSE']=='XNA'].index)
df1=df1.drop(df1[df1['NAME_CASH_LOAN_PURPOSE']=='XNA'].index)
df1=df1.drop(df1[df1['NAME_CASH_LOAN_PURPOSE']=='XAP'].index)

#Merge the previous application with the current application data file
merged_df= pd.merge(df, df1, how='inner',
on='SK_ID_CURR',suffixes='_x')
merged_df.head()
```

```
   SK_ID_CURR  TARGET NAME_CONTRACT_TYPE_ CODE_GENDER FLAG_OWN_CAR  \
0      100034       0     Revolving loans           M            N
1      100035       0          Cash loans           F            N
2      100039       0          Cash loans           M            Y
3      100046       0     Revolving loans           M            Y
4      100046       0     Revolving loans           M            Y

  FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT_
AMT_ANNUITY_  \
0               Y             0           90000.0     180000.0
9000.0
1               Y             0          292500.0     665892.0
24592.5
2               N             1          360000.0     733315.5
39069.0
3               Y             0          180000.0     540000.0
27000.0
4               Y             0          180000.0     540000.0
27000.0
```

```
   AMT_GOODS_PRICE_ NAME_TYPE_SUITE_      NAME_INCOME_TYPE  \
0          180000.0    Unaccompanied               Working   
1          477000.0    Unaccompanied  Commercial associate   
2          679500.0    Unaccompanied  Commercial associate   
3          540000.0    Unaccompanied               Working   
4          540000.0    Unaccompanied               Working   

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

   REGION_POPULATION_RELATIVE  DAYS_BIRTH  DAYS_EMPLOYED 
DAYS_REGISTRATION  \
0                    0.030755       10341           1010 
4799.0   
1                    0.025164       15280           2668 
5266.0   
2                    0.015221       11694           2060 
3557.0   
3                    0.025164       16126           1761 
8236.0   
4                    0.025164       16126           1761 
8236.0   

   DAYS_ID_PUBLISH OCCUPATION_TYPE WEEKDAY_APPR_PROCESS_START_  \
0             3015        Laborers                     TUESDAY   
1             3787             NaN                   WEDNESDAY   
2             3557         Drivers                    THURSDAY   
3             4292        Managers                     TUESDAY   
4             4292        Managers                     TUESDAY   

   HOUR_APPR_PROCESS_START_ REG_REGION_NOT_LIVE_REGION  \
0                        16                          0   
1                        13                          0   
2                        10                          0   
3                         8                          0   
4                         8                          0   

  REG_REGION_NOT_WORK_REGION LIVE_REGION_NOT_WORK_REGION  \
0                          0                           0   
1                          0                           0   
```

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

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

        ORGANIZATION_TYPE  EXT_SOURCE_2  OBS_30_CNT_SOCIAL_CIRCLE  \
0  Business Entity Type 3      0.502779                       0.0
1  Business Entity Type 3      0.479987                       0.0
2           Self-employed      0.321745                       2.0
3  Business Entity Type 3      0.738053                       1.0
4  Business Entity Type 3      0.738053                       1.0

   DEF_30_CNT_SOCIAL_CIRCLE  OBS_60_CNT_SOCIAL_CIRCLE  \
0                       0.0                       0.0
1                       0.0                       0.0
2                       0.0                       2.0
3                       0.0                       1.0
4                       0.0                       1.0

   DEF_60_CNT_SOCIAL_CIRCLE  AMT_REQ_CREDIT_BUREAU_HOUR  \
0                       0.0                         NaN
1                       0.0                         0.0
2                       0.0                         0.0
3                       0.0                         0.0
4                       0.0                         0.0

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

   AMT_REQ_CREDIT_BUREAU_MON  AMT_REQ_CREDIT_BUREAU_QRT  \
0                        NaN                        NaN
1                        1.0                        0.0
2                        0.0                        1.0
3                        1.0                        0.0
```

```
4                         1.0                          0.0

   AMT_REQ_CREDIT_BUREAU_YEAR AMT_CREDIT_RANGE AMT_INCOME_RANGE
AGE_RANGE  \
0                          NaN              Low              Low
20-30
1                          5.0           Medium           Medium
40-50
2                          1.0             High           Medium
30-40
3                          1.0           Medium              Low
40-50
4                          1.0           Medium              Low
40-50

   SK_ID_PREV NAME_CONTRACT_TYPEx  AMT_ANNUITYx  AMT_APPLICATION
AMT_CREDITx  \
0     1390369          Cash loans     22430.430          94500.0
109971.0
1     1344613          Cash loans     33238.800        1260000.0
1260000.0
2     1077565          Cash loans     52513.515        1350000.0
1487214.0
3     1223113          Cash loans     28390.635         337500.0
407911.5
4     1529558          Cash loans     29053.215         450000.0
555723.0

   AMT_DOWN_PAYMENT  AMT_GOODS_PRICEx WEEKDAY_APPR_PROCESS_STARTx  \
0               NaN           94500.0                      FRIDAY
1               NaN         1260000.0                   WEDNESDAY
2               NaN         1350000.0                      MONDAY
3               NaN          337500.0                      MONDAY
4               NaN          450000.0                      MONDAY

   HOUR_APPR_PROCESS_STARTx FLAG_LAST_APPL_PER_CONTRACT  \
0                        11                           Y
1                        14                           Y
2                        14                           Y
3                        13                           Y
4                        15                           Y

   NFLAG_LAST_APPL_IN_DAY  RATE_DOWN_PAYMENT  RATE_INTEREST_PRIMARY  \
0                       1                NaN                    NaN
1                       1                NaN                    NaN
2                       1                NaN                    NaN
3                       1                NaN                    NaN
4                       1                NaN                    NaN

   RATE_INTEREST_PRIVILEGED   NAME_CASH_LOAN_PURPOSE
```

```
NAME_CONTRACT_STATUS  \
0                     NaN                    Other
Approved
1                     NaN  Payments on other loans
Refused
2                     NaN        Buying a used car
Approved
3                     NaN                  Repairs
Approved
4                     NaN                  Repairs
Refused

   DAYS_DECISION      NAME_PAYMENT_TYPE CODE_REJECT_REASON
NAME_TYPE_SUITEx  \
0           -599  Cash through the bank                XAP
NaN
1           -119  Cash through the bank                 HC
Unaccompanied
2           -695  Cash through the bank                XAP
Unaccompanied
3           -539  Cash through the bank                XAP
Unaccompanied
4           -449  Cash through the bank              LIMIT
NaN

  NAME_CLIENT_TYPE NAME_GOODS_CATEGORY NAME_PORTFOLIO
NAME_PRODUCT_TYPE  \
0              New                 XNA           Cash           walk-
in
1         Repeater                 XNA           Cash           walk-
in
2        Refreshed                 XNA           Cash           walk-
in
3              New                 XNA           Cash           walk-
in
4         Repeater                 XNA           Cash           walk-
in

                 CHANNEL_TYPE  SELLERPLACE_AREA
NAME_SELLER_INDUSTRY  \
0     Credit and cash offices                -1                  XNA

1     Credit and cash offices                -1                  XNA

2  Channel of corporate sales                -1                  XNA

3     Credit and cash offices                -1                  XNA

4     Credit and cash offices                -1                  XNA
```

```
   CNT_PAYMENT NAME_YIELD_GROUP PRODUCT_COMBINATION
DAYS_FIRST_DRAWING  \
0          6.0             high   Cash Street: high
365243.0
1         60.0       low_action    Cash Street: low
NaN
2         42.0       low_normal    Cash Street: low
365243.0
3         18.0       low_normal    Cash Street: low
365243.0
4         24.0       low_normal    Cash Street: low
NaN

   DAYS_FIRST_DUE  DAYS_LAST_DUE_1ST_VERSION  DAYS_LAST_DUE
DAYS_TERMINATION  \
0          -569.0                     -419.0         -449.0
-443.0
1             NaN                        NaN            NaN
NaN
2          -665.0                      565.0         -455.0
-446.0
3          -509.0                        1.0       365243.0
365243.0
4             NaN                        NaN            NaN
NaN

   NFLAG_INSURED_ON_APPROVAL
0                        1.0
1                        NaN
2                        1.0
3                        1.0
4                        NaN
```

```
# Renaming the column names after merging

new_df = merged_df.rename({'NAME_CONTRACT_TYPE_' :
'NAME_CONTRACT_TYPE','AMT_CREDIT_':'AMT_CREDIT','AMT_ANNUITY_':'AMT_AN
NUITY',
                         'WEEKDAY_APPR_PROCESS_START_' :
'WEEKDAY_APPR_PROCESS_START',

'HOUR_APPR_PROCESS_START_':'HOUR_APPR_PROCESS_START','NAME_CONTRACT_TY
PEx':'NAME_CONTRACT_TYPE_PREV',

'AMT_CREDITx':'AMT_CREDIT_PREV','AMT_ANNUITYx':'AMT_ANNUITY_PREV',

'WEEKDAY_APPR_PROCESS_STARTx':'WEEKDAY_APPR_PROCESS_START_PREV',
```

```
'HOUR_APPR_PROCESS_STARTx':'HOUR_APPR_PROCESS_START_PREV'}, axis=1)
new_df.head()

   SK_ID_CURR  TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR  \
0      100034       0    Revolving loans           M            N
1      100035       0         Cash loans           F            N
2      100039       0         Cash loans           M            Y
3      100046       0    Revolving loans           M            Y
4      100046       0    Revolving loans           M            Y

  FLAG_OWN_REALTY  CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT
AMT_ANNUITY  \
0               Y             0           90000.0    180000.0
9000.0
1               Y             0          292500.0    665892.0
24592.5
2               N             1          360000.0    733315.5
39069.0
3               Y             0          180000.0    540000.0
27000.0
4               Y             0          180000.0    540000.0
27000.0

   AMT_GOODS_PRICE_ NAME_TYPE_SUITE_      NAME_INCOME_TYPE  \
0          180000.0    Unaccompanied               Working
1          477000.0    Unaccompanied  Commercial associate
2          679500.0    Unaccompanied  Commercial associate
3          540000.0    Unaccompanied               Working
4          540000.0    Unaccompanied               Working

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

   REGION_POPULATION_RELATIVE  DAYS_BIRTH  DAYS_EMPLOYED
DAYS_REGISTRATION  \
0                    0.030755       10341           1010
4799.0
1                    0.025164       15280           2668
5266.0
```

```
2                    0.015221       11694           2060
3557.0
3                    0.025164       16126           1761
8236.0
4                    0.025164       16126           1761
8236.0

   DAYS_ID_PUBLISH OCCUPATION_TYPE WEEKDAY_APPR_PROCESS_START  \
0             3015        Laborers                    TUESDAY
1             3787             NaN                  WEDNESDAY
2             3557         Drivers                   THURSDAY
3             4292        Managers                    TUESDAY
4             4292        Managers                    TUESDAY

   HOUR_APPR_PROCESS_START REG_REGION_NOT_LIVE_REGION  \
0                       16                          0
1                       13                          0
2                       10                          0
3                        8                          0
4                        8                          0

  REG_REGION_NOT_WORK_REGION LIVE_REGION_NOT_WORK_REGION  \
0                          0                           0
1                          0                           0
2                          0                           0
3                          0                           0
4                          0                           0

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

        ORGANIZATION_TYPE  EXT_SOURCE_2  OBS_30_CNT_SOCIAL_CIRCLE  \
0  Business Entity Type 3      0.502779                       0.0
1  Business Entity Type 3      0.479987                       0.0
2           Self-employed      0.321745                       2.0
3  Business Entity Type 3      0.738053                       1.0
4  Business Entity Type 3      0.738053                       1.0

   DEF_30_CNT_SOCIAL_CIRCLE  OBS_60_CNT_SOCIAL_CIRCLE  \
0                       0.0                       0.0
```

```
1                        0.0                        0.0
2                        0.0                        2.0
3                        0.0                        1.0
4                        0.0                        1.0

   DEF_60_CNT_SOCIAL_CIRCLE  AMT_REQ_CREDIT_BUREAU_HOUR  \
0                        0.0                        NaN
1                        0.0                        0.0
2                        0.0                        0.0
3                        0.0                        0.0
4                        0.0                        0.0

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

   AMT_REQ_CREDIT_BUREAU_MON  AMT_REQ_CREDIT_BUREAU_QRT  \
0                        NaN                        NaN
1                        1.0                        0.0
2                        0.0                        1.0
3                        1.0                        0.0
4                        1.0                        0.0

   AMT_REQ_CREDIT_BUREAU_YEAR AMT_CREDIT_RANGE AMT_INCOME_RANGE
AGE_RANGE  \
0                         NaN              Low              Low
20-30
1                         5.0           Medium           Medium
40-50
2                         1.0             High           Medium
30-40
3                         1.0           Medium              Low
40-50
4                         1.0           Medium              Low
40-50

   SK_ID_PREV NAME_CONTRACT_TYPE_PREV  AMT_ANNUITY_PREV
AMT_APPLICATION  \
0     1390369              Cash loans         22430.430
94500.0
1     1344613              Cash loans         33238.800
1260000.0
2     1077565              Cash loans         52513.515
1350000.0
3     1223113              Cash loans         28390.635
337500.0
4     1529558              Cash loans         29053.215
```

```
450000.0

   AMT_CREDIT_PREV  AMT_DOWN_PAYMENT  AMT_GOODS_PRICEx  \
0         109971.0               NaN           94500.0
1        1260000.0               NaN         1260000.0
2        1487214.0               NaN         1350000.0
3         407911.5               NaN          337500.0
4         555723.0               NaN          450000.0

  WEEKDAY_APPR_PROCESS_START_PREV  HOUR_APPR_PROCESS_START_PREV  \
0                          FRIDAY                            11
1                       WEDNESDAY                            14
2                          MONDAY                            14
3                          MONDAY                            13
4                          MONDAY                            15

  FLAG_LAST_APPL_PER_CONTRACT  NFLAG_LAST_APPL_IN_DAY
RATE_DOWN_PAYMENT  \
0                           Y                       1
NaN
1                           Y                       1
NaN
2                           Y                       1
NaN
3                           Y                       1
NaN
4                           Y                       1
NaN

   RATE_INTEREST_PRIMARY  RATE_INTEREST_PRIVILEGED
NAME_CASH_LOAN_PURPOSE  \
0                    NaN                       NaN
Other
1                    NaN                       NaN  Payments on other
loans
2                    NaN                       NaN        Buying a
used car
3                    NaN                       NaN
Repairs
4                    NaN                       NaN
Repairs

  NAME_CONTRACT_STATUS  DAYS_DECISION      NAME_PAYMENT_TYPE  \
0             Approved           -599  Cash through the bank
1              Refused           -119  Cash through the bank
2             Approved           -695  Cash through the bank
3             Approved           -539  Cash through the bank
4              Refused           -449  Cash through the bank

  CODE_REJECT_REASON NAME_TYPE_SUITEx NAME_CLIENT_TYPE
```

```
NAME_GOODS_CATEGORY  \
0                XAP                NaN              New
XNA
1                 HC    Unaccompanied         Repeater
XNA
2                XAP    Unaccompanied        Refreshed
XNA
3                XAP    Unaccompanied              New
XNA
4              LIMIT                NaN         Repeater
XNA

  NAME_PORTFOLIO NAME_PRODUCT_TYPE                CHANNEL_TYPE  \
0           Cash           walk-in     Credit and cash offices
1           Cash           walk-in     Credit and cash offices
2           Cash           walk-in  Channel of corporate sales
3           Cash           walk-in     Credit and cash offices
4           Cash           walk-in     Credit and cash offices

   SELLERPLACE_AREA NAME_SELLER_INDUSTRY  CNT_PAYMENT NAME_YIELD_GROUP
\
0                -1                  XNA          6.0             high

1                -1                  XNA         60.0       low_action

2                -1                  XNA         42.0       low_normal

3                -1                  XNA         18.0       low_normal

4                -1                  XNA         24.0       low_normal


  PRODUCT_COMBINATION  DAYS_FIRST_DRAWING  DAYS_FIRST_DUE  \
0   Cash Street: high            365243.0          -569.0
1    Cash Street: low                 NaN             NaN
2    Cash Street: low            365243.0          -665.0
3    Cash Street: low            365243.0          -509.0
4    Cash Street: low                 NaN             NaN

   DAYS_LAST_DUE_1ST_VERSION  DAYS_LAST_DUE  DAYS_TERMINATION  \
0                     -419.0         -449.0            -443.0
1                        NaN            NaN               NaN
2                      565.0         -455.0            -446.0
3                        1.0       365243.0          365243.0
4                        NaN            NaN               NaN

   NFLAG_INSURED_ON_APPROVAL
0                        1.0
1                        NaN
```

```
2                         1.0
3                         1.0
4                         NaN
```

```
# Removing unwanted columns for analysis

new_df.drop(['SK_ID_CURR','WEEKDAY_APPR_PROCESS_START',
'HOUR_APPR_PROCESS_START','REG_REGION_NOT_LIVE_REGION',

'REG_REGION_NOT_WORK_REGION','LIVE_REGION_NOT_WORK_REGION',
'REG_CITY_NOT_LIVE_CITY',
              'REG_CITY_NOT_WORK_CITY',
'LIVE_CITY_NOT_WORK_CITY','WEEKDAY_APPR_PROCESS_START_PREV',
              'HOUR_APPR_PROCESS_START_PREV',
'FLAG_LAST_APPL_PER_CONTRACT','NFLAG_LAST_APPL_IN_DAY'],axis=1,inplace
=True)

new_df.head()
```

```
   TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR FLAG_OWN_REALTY
\
0       0    Revolving loans           M            N               Y

1       0         Cash loans           F            N               Y

2       0         Cash loans           M            Y               N

3       0    Revolving loans           M            Y               Y

4       0    Revolving loans           M            Y               Y


   CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT  AMT_ANNUITY
AMT_GOODS_PRICE_  \
0             0           90000.0    180000.0       9000.0
180000.0
1             0          292500.0    665892.0      24592.5
477000.0
2             1          360000.0    733315.5      39069.0
679500.0
3             0          180000.0    540000.0      27000.0
540000.0
4             0          180000.0    540000.0      27000.0
540000.0

  NAME_TYPE_SUITE_      NAME_INCOME_TYPE
NAME_EDUCATION_TYPE  \
0    Unaccompanied               Working               Higher
education
1    Unaccompanied  Commercial associate  Secondary / secondary
```

```
special
2    Unaccompanied  Commercial associate  Secondary / secondary
special
3    Unaccompanied               Working              Higher
education
4    Unaccompanied               Working              Higher
education

     NAME_FAMILY_STATUS  NAME_HOUSING_TYPE  REGION_POPULATION_RELATIVE
\
0  Single / not married       With parents                    0.030755

1        Civil marriage  House / apartment                    0.025164

2               Married  House / apartment                    0.015221

3               Married  House / apartment                    0.025164

4               Married  House / apartment                    0.025164


   DAYS_BIRTH  DAYS_EMPLOYED  DAYS_REGISTRATION  DAYS_ID_PUBLISH  \
0       10341           1010             4799.0             3015
1       15280           2668             5266.0             3787
2       11694           2060             3557.0             3557
3       16126           1761             8236.0             4292
4       16126           1761             8236.0             4292

  OCCUPATION_TYPE       ORGANIZATION_TYPE  EXT_SOURCE_2  \
0        Laborers  Business Entity Type 3      0.502779
1             NaN  Business Entity Type 3      0.479987
2         Drivers           Self-employed      0.321745
3        Managers  Business Entity Type 3      0.738053
4        Managers  Business Entity Type 3      0.738053

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

   OBS_60_CNT_SOCIAL_CIRCLE  DEF_60_CNT_SOCIAL_CIRCLE  \
0                       0.0                       0.0
1                       0.0                       0.0
2                       2.0                       0.0
3                       1.0                       0.0
4                       1.0                       0.0
```

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

   AMT_REQ_CREDIT_BUREAU_WEEK  AMT_REQ_CREDIT_BUREAU_MON  \
0                         NaN                        NaN   
1                         0.0                        1.0   
2                         0.0                        0.0   
3                         0.0                        1.0   
4                         0.0                        1.0   

   AMT_REQ_CREDIT_BUREAU_QRT  AMT_REQ_CREDIT_BUREAU_YEAR 
AMT_CREDIT_RANGE  \
0                        NaN                        NaN 
Low   
1                        0.0                        5.0 
Medium   
2                        1.0                        1.0 
High   
3                        0.0                        1.0 
Medium   
4                        0.0                        1.0 
Medium   

  AMT_INCOME_RANGE AGE_RANGE  SK_ID_PREV NAME_CONTRACT_TYPE_PREV  \
0              Low     20-30     1390369              Cash loans   
1           Medium     40-50     1344613              Cash loans   
2           Medium     30-40     1077565              Cash loans   
3              Low     40-50     1223113              Cash loans   
4              Low     40-50     1529558              Cash loans   

   AMT_ANNUITY_PREV  AMT_APPLICATION  AMT_CREDIT_PREV 
AMT_DOWN_PAYMENT  \
0         22430.430          94500.0         109971.0 
NaN   
1         33238.800        1260000.0        1260000.0 
NaN   
2         52513.515        1350000.0        1487214.0 
NaN   
3         28390.635         337500.0         407911.5 
NaN   
4         29053.215         450000.0         555723.0 
NaN   

   AMT_GOODS_PRICEx  RATE_DOWN_PAYMENT  RATE_INTEREST_PRIMARY  \
0           94500.0                NaN                    NaN   
1         1260000.0                NaN                    NaN   
```

```
2         1350000.0                NaN                    NaN
3          337500.0                NaN                    NaN
4          450000.0                NaN                    NaN

   RATE_INTEREST_PRIVILEGED   NAME_CASH_LOAN_PURPOSE
NAME_CONTRACT_STATUS  \
0                       NaN                    Other
Approved
1                       NaN  Payments on other loans
Refused
2                       NaN        Buying a used car
Approved
3                       NaN                  Repairs
Approved
4                       NaN                  Repairs
Refused

   DAYS_DECISION      NAME_PAYMENT_TYPE CODE_REJECT_REASON
NAME_TYPE_SUITEx  \
0           -599  Cash through the bank                XAP
NaN
1           -119  Cash through the bank                 HC
Unaccompanied
2           -695  Cash through the bank                XAP
Unaccompanied
3           -539  Cash through the bank                XAP
Unaccompanied
4           -449  Cash through the bank              LIMIT
NaN

  NAME_CLIENT_TYPE NAME_GOODS_CATEGORY NAME_PORTFOLIO
NAME_PRODUCT_TYPE  \
0              New                 XNA           Cash           walk-
in
1         Repeater                 XNA           Cash           walk-
in
2        Refreshed                 XNA           Cash           walk-
in
3              New                 XNA           Cash           walk-
in
4         Repeater                 XNA           Cash           walk-
in

                  CHANNEL_TYPE  SELLERPLACE_AREA
NAME_SELLER_INDUSTRY  \
0     Credit and cash offices                -1                  XNA

1     Credit and cash offices                -1                  XNA

2  Channel of corporate sales                -1                  XNA
```

```
3     Credit and cash offices                -1                   XNA

4     Credit and cash offices                -1                   XNA


   CNT_PAYMENT NAME_YIELD_GROUP PRODUCT_COMBINATION
DAYS_FIRST_DRAWING  \
0          6.0             high   Cash Street: high
365243.0
1         60.0       low_action    Cash Street: low
NaN
2         42.0       low_normal    Cash Street: low
365243.0
3         18.0       low_normal    Cash Street: low
365243.0
4         24.0       low_normal    Cash Street: low
NaN

   DAYS_FIRST_DUE  DAYS_LAST_DUE_1ST_VERSION  DAYS_LAST_DUE
DAYS_TERMINATION  \
0          -569.0                     -419.0         -449.0
-443.0
1             NaN                        NaN            NaN
NaN
2          -665.0                      565.0         -455.0
-446.0
3          -509.0                        1.0       365243.0
365243.0
4             NaN                        NaN            NaN
NaN

   NFLAG_INSURED_ON_APPROVAL
0                        1.0
1                        NaN
2                        1.0
3                        1.0
4                        NaN
```

```
new_df.head()
```

```
   TARGET NAME_CONTRACT_TYPE CODE_GENDER FLAG_OWN_CAR FLAG_OWN_REALTY
\
0       0    Revolving loans           M            N               Y

1       0         Cash loans           F            N               Y

2       0         Cash loans           M            Y               N

3       0    Revolving loans           M            Y               Y
```

```
4       0    Revolving loans           M            Y               Y


   CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT  AMT_ANNUITY
AMT_GOODS_PRICE_  \
0             0           90000.0    180000.0       9000.0
180000.0
1             0          292500.0    665892.0      24592.5
477000.0
2             1          360000.0    733315.5      39069.0
679500.0
3             0          180000.0    540000.0      27000.0
540000.0
4             0          180000.0    540000.0      27000.0
540000.0

  NAME_TYPE_SUITE_      NAME_INCOME_TYPE
NAME_EDUCATION_TYPE  \
0    Unaccompanied               Working               Higher
education
1    Unaccompanied  Commercial associate  Secondary / secondary
special
2    Unaccompanied  Commercial associate  Secondary / secondary
special
3    Unaccompanied               Working               Higher
education
4    Unaccompanied               Working               Higher
education

     NAME_FAMILY_STATUS  NAME_HOUSING_TYPE  REGION_POPULATION_RELATIVE
\
0  Single / not married       With parents                    0.030755

1        Civil marriage  House / apartment                    0.025164

2               Married  House / apartment                    0.015221

3               Married  House / apartment                    0.025164

4               Married  House / apartment                    0.025164


   DAYS_BIRTH  DAYS_EMPLOYED  DAYS_REGISTRATION  DAYS_ID_PUBLISH  \
0       10341           1010             4799.0             3015
1       15280           2668             5266.0             3787
2       11694           2060             3557.0             3557
3       16126           1761             8236.0             4292
4       16126           1761             8236.0             4292
```

```
  OCCUPATION_TYPE       ORGANIZATION_TYPE  EXT_SOURCE_2  \
0        Laborers  Business Entity Type 3      0.502779
1             NaN  Business Entity Type 3      0.479987
2         Drivers           Self-employed      0.321745
3        Managers  Business Entity Type 3      0.738053
4        Managers  Business Entity Type 3      0.738053

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

   OBS_60_CNT_SOCIAL_CIRCLE  DEF_60_CNT_SOCIAL_CIRCLE  \
0                       0.0                       0.0
1                       0.0                       0.0
2                       2.0                       0.0
3                       1.0                       0.0
4                       1.0                       0.0

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

   AMT_REQ_CREDIT_BUREAU_WEEK  AMT_REQ_CREDIT_BUREAU_MON  \
0                         NaN                        NaN
1                         0.0                        1.0
2                         0.0                        0.0
3                         0.0                        1.0
4                         0.0                        1.0

   AMT_REQ_CREDIT_BUREAU_QRT  AMT_REQ_CREDIT_BUREAU_YEAR
AMT_CREDIT_RANGE  \
0                        NaN                         NaN
Low
1                        0.0                         5.0
Medium
2                        1.0                         1.0
High
3                        0.0                         1.0
Medium
4                        0.0                         1.0
Medium
```

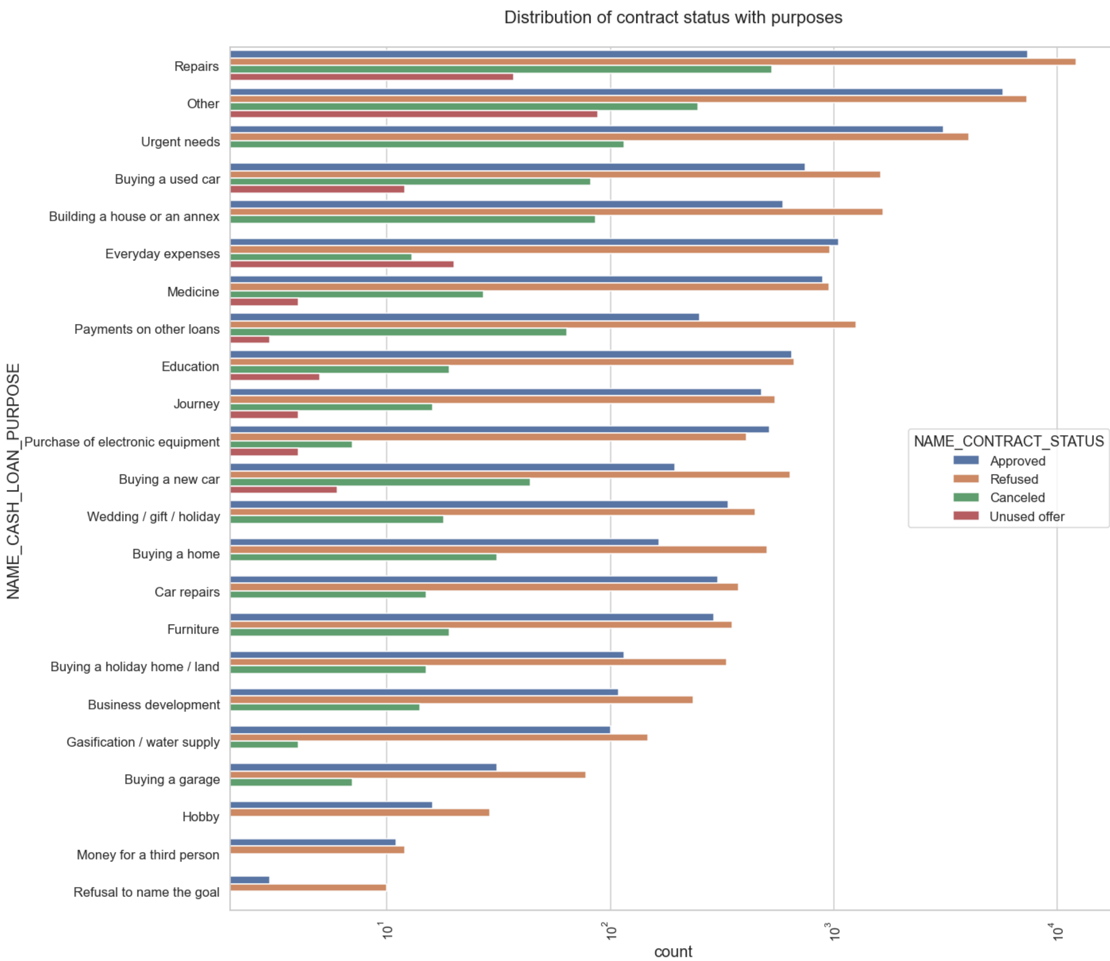
```
  AMT_INCOME_RANGE AGE_RANGE  SK_ID_PREV NAME_CONTRACT_TYPE_PREV  \
0              Low     20-30     1390369              Cash loans
1           Medium     40-50     1344613              Cash loans
2           Medium     30-40     1077565              Cash loans
3              Low     40-50     1223113              Cash loans
4              Low     40-50     1529558              Cash loans

   AMT_ANNUITY_PREV  AMT_APPLICATION  AMT_CREDIT_PREV
AMT_DOWN_PAYMENT  \
0         22430.430          94500.0         109971.0
NaN
1         33238.800        1260000.0        1260000.0
NaN
2         52513.515        1350000.0        1487214.0
NaN
3         28390.635         337500.0         407911.5
NaN
4         29053.215         450000.0         555723.0
NaN

   AMT_GOODS_PRICEx  RATE_DOWN_PAYMENT  RATE_INTEREST_PRIMARY  \
0           94500.0                NaN                    NaN
1         1260000.0                NaN                    NaN
2         1350000.0                NaN                    NaN
3          337500.0                NaN                    NaN
4          450000.0                NaN                    NaN

   RATE_INTEREST_PRIVILEGED   NAME_CASH_LOAN_PURPOSE
NAME_CONTRACT_STATUS  \
0                       NaN                    Other
Approved
1                       NaN  Payments on other loans
Refused
2                       NaN        Buying a used car
Approved
3                       NaN                  Repairs
Approved
4                       NaN                  Repairs
Refused

   DAYS_DECISION      NAME_PAYMENT_TYPE CODE_REJECT_REASON
NAME_TYPE_SUITEx  \
0           -599  Cash through the bank                XAP
NaN
1           -119  Cash through the bank                 HC
Unaccompanied
2           -695  Cash through the bank                XAP
Unaccompanied
3           -539  Cash through the bank                XAP
Unaccompanied
```

```
4           -449  Cash through the bank              LIMIT
NaN

  NAME_CLIENT_TYPE NAME_GOODS_CATEGORY NAME_PORTFOLIO
NAME_PRODUCT_TYPE  \
0              New                 XNA           Cash           walk-
in
1         Repeater                 XNA           Cash           walk-
in
2        Refreshed                 XNA           Cash           walk-
in
3              New                 XNA           Cash           walk-
in
4         Repeater                 XNA           Cash           walk-
in

                 CHANNEL_TYPE  SELLERPLACE_AREA
NAME_SELLER_INDUSTRY  \
0     Credit and cash offices                -1                  XNA

1     Credit and cash offices                -1                  XNA

2  Channel of corporate sales                -1                  XNA

3     Credit and cash offices                -1                  XNA

4     Credit and cash offices                -1                  XNA


   CNT_PAYMENT NAME_YIELD_GROUP PRODUCT_COMBINATION
DAYS_FIRST_DRAWING  \
0          6.0             high   Cash Street: high
365243.0
1         60.0       low_action    Cash Street: low
NaN
2         42.0       low_normal    Cash Street: low
365243.0
3         18.0       low_normal    Cash Street: low
365243.0
4         24.0       low_normal    Cash Street: low
NaN

   DAYS_FIRST_DUE  DAYS_LAST_DUE_1ST_VERSION  DAYS_LAST_DUE
DAYS_TERMINATION  \
0          -569.0                     -419.0         -449.0
-443.0
1             NaN                        NaN            NaN
NaN
2          -665.0                      565.0         -455.0
```

```
-446.0
3          -509.0                       1.0       365243.0
365243.0
4             NaN                       NaN            NaN
NaN

   NFLAG_INSURED_ON_APPROVAL
0                        1.0
1                        NaN
2                        1.0
3                        1.0
4                        NaN
```

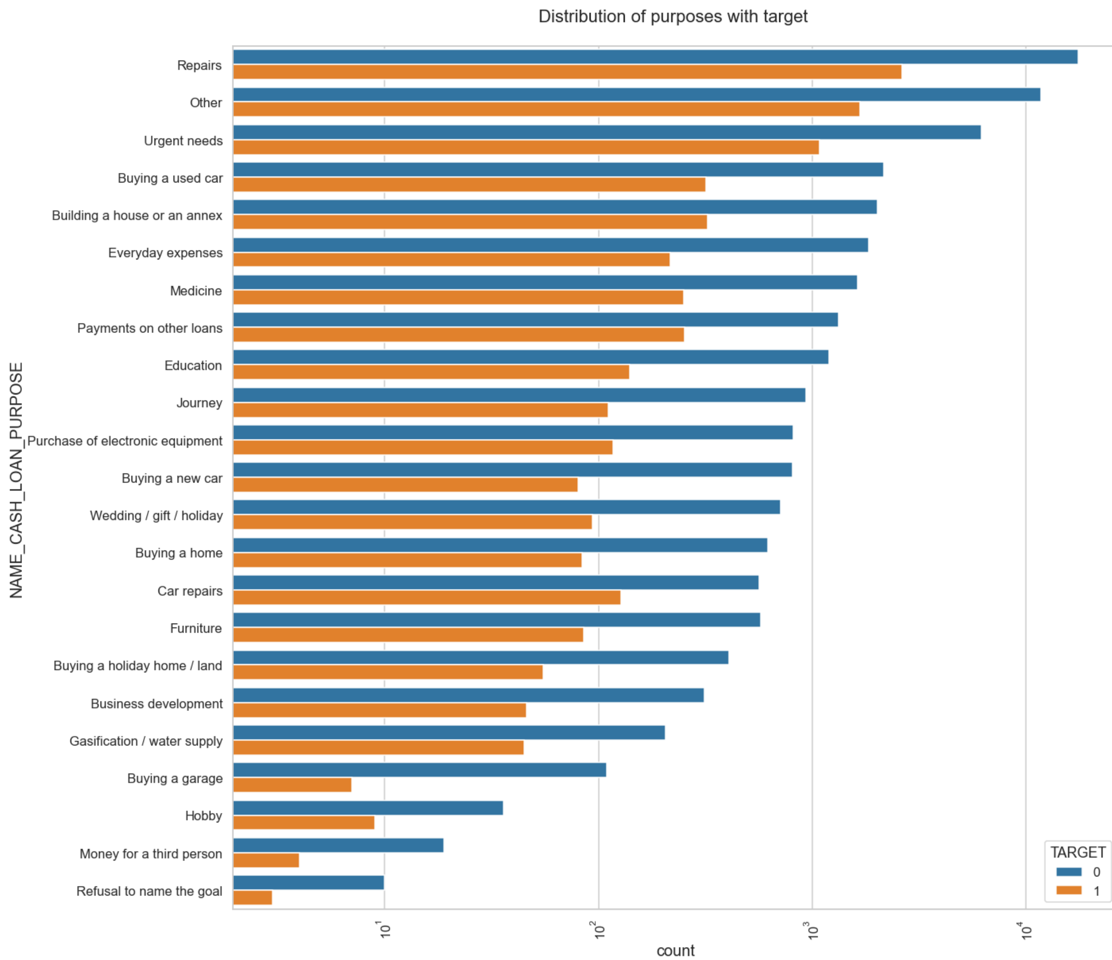
Univariate Analysis

```
# Distribution of contract status in logarithmic scale

sns.set_style('whitegrid')
sns.set_context('talk')

plt.figure(figsize=(20,20))
plt.rcParams["axes.labelsize"] = 20
plt.rcParams['axes.titlesize'] = 22
plt.rcParams['axes.titlepad'] = 30
plt.xticks(rotation=90)
plt.xscale('log')
plt.title('Distribution of contract status with purposes')
ax=sns.countplot(data = new_df, y='NAME_CASH_LOAN_PURPOSE',
order=new_df['NAME_CASH_LOAN_PURPOSE'].value_counts().index,hue =
'NAME_CONTRACT_STATUS',palette='deep')
```



Points to be concluded from above plot: Most rejection of loans came from purpose 'Repairs'. For education purposes we have equal number of approves and rejection PayinG other loans and buying a new car is having significant higher rejection than approves.

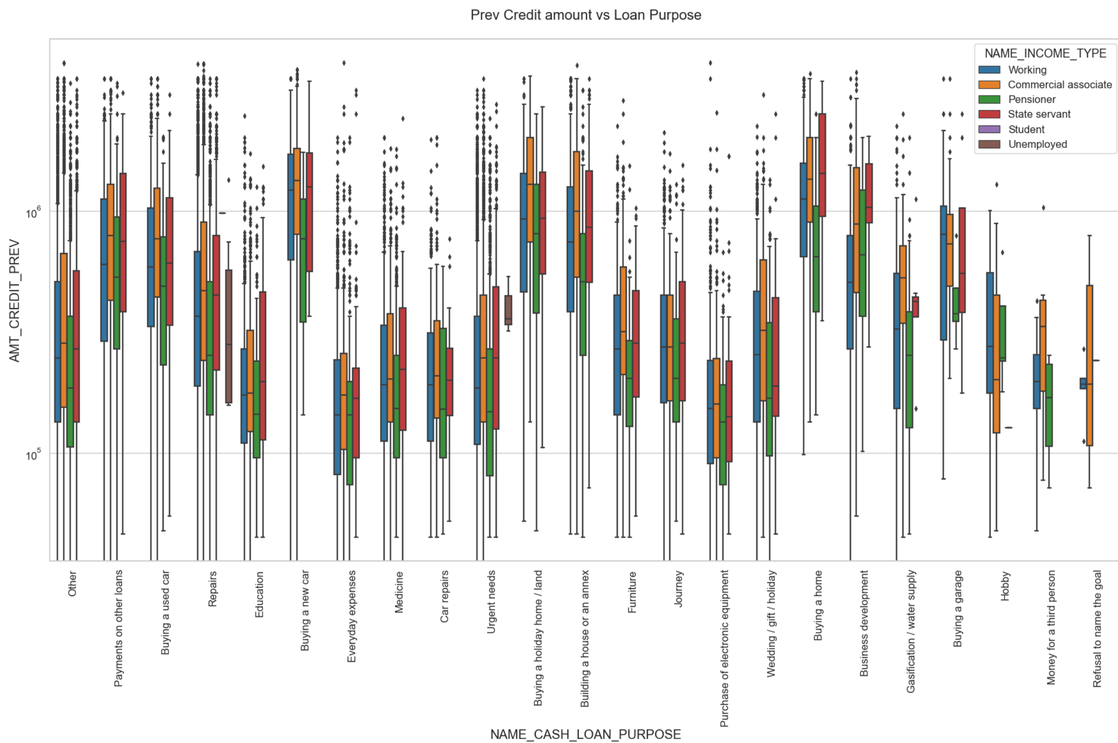
```
# Distribution of contract status

sns.set_style('whitegrid')
sns.set_context('talk')

plt.figure(figsize=(20,20))
plt.rcParams["axes.labelsize"] = 20
plt.rcParams['axes.titlesize'] = 22
plt.rcParams['axes.titlepad'] = 30
plt.xticks(rotation=90)
plt.xscale('log')
plt.title('Distribution of purposes with target ')
ax = sns.countplot(data = new_df, y= 'NAME_CASH_LOAN_PURPOSE',
order=new_df['NAME_CASH_LOAN_PURPOSE'].value_counts().index,hue =
'TARGET')
```



Few points we can conclude from above plot: Loan purposes with 'Repairs' are facing more difficulites in payment on time. There are few places where loan payment is significant higher than facing difficulties. They are 'Buying a garage', 'Business developemt', 'Buying land','Buying a new car' and 'Education' Hence we can focus on these purposes for which the client is having for minimal payment difficulties

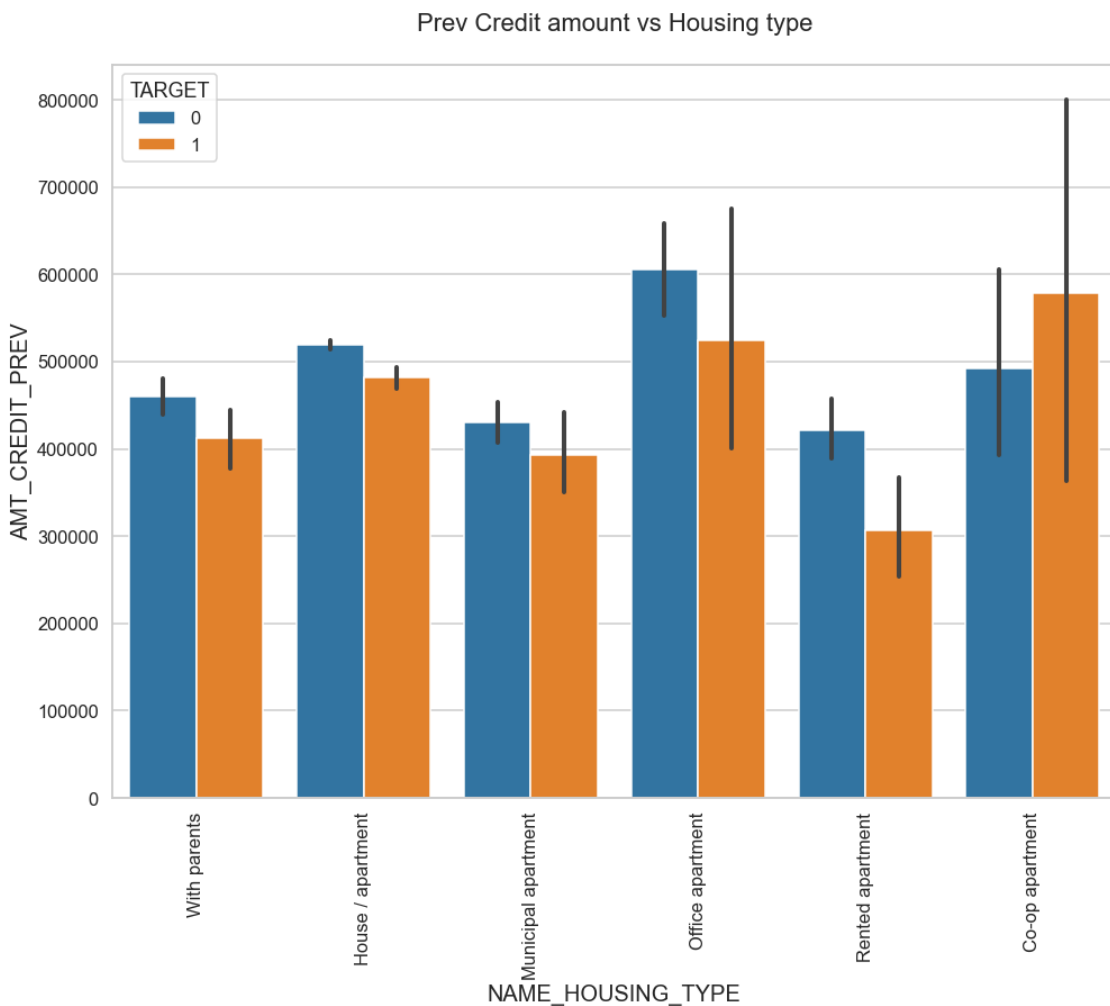
Bivariate Analysis

```
# Box plotting for Credit amount in logarithmic scale

plt.figure(figsize=(30,15))
plt.xticks(rotation=90)
plt.yscale('log')

sns.boxplot(data =new_df,
x='NAME_CASH_LOAN_PURPOSE',hue='NAME_INCOME_TYPE',y='AMT_CREDIT_PREV',
orient='v')
plt.title('Prev Credit amount vs Loan Purpose')
plt.show()
```



From the above we can conclude some points- The credit amount of Loan purposes like 'Buying a home','Buying a land','Buying a new car' and'Building a house' is higher. Income type of state servants have a significant amount of credit applied Money for third person or a Hobby is having less credits applied for.

```
# Box plotting for Credit amount prev vs Housing type in logarithmic
scale

plt.figure(figsize=(16,12))
plt.xticks(rotation=90)
sns.barplot(data =new_df,
y='AMT_CREDIT_PREV',hue='TARGET',x='NAME_HOUSING_TYPE',)
plt.title('Prev Credit amount vs Housing type')
plt.show()
```



Here for Housing type, office appartment is having higher credit of target 0 and co-op apartment is having higher credit of target 1. So, we can conclude that bank should avoid giving loans to the housing type of co-op apartment as they are having difficulties in payment. Bank can focus mostly on housing type with parents or House\appartment or miuncipal appartment for successful payments.

CONCLUSION

1. Banks should focus more on contract type ‘Student’ ,’pensioner’ and ‘Businessman’ with housing ‘type other than ‘Co-op apartment’ for successful payments.
2. Banks should focus less on income type ‘Working’ as they are having most number of unsuccessful payments.
3. Also with loan purpose ‘Repair’ is having higher number of unsuccessful payments on time.
4. Get as much as clients from housing type ‘With parents’ as they are having least number of unsuccessful payments.