PROJECT 1

August 1, 2023

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

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

[3]:
```
pd.set_option('display.max_columns', None)
data.head()
```

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

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

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

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

```
4  Secondary / secondary special  Single / not married  House / apartment

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

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

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

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

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

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

   LIVE_REGION_NOT_WORK_REGION  REG_CITY_NOT_LIVE_CITY  \
0                            0                       0
1                            0                       0
```

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

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

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

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

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

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

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

   YEARS_BEGINEXPLUATATION_MODE  YEARS_BUILD_MODE  COMMONAREA_MODE  \
```

```
0                        0.9722             0.6341           0.0144
1                        0.9851             0.8040           0.0497
2                           NaN                NaN              NaN
3                           NaN                NaN              NaN
4                           NaN                NaN              NaN

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

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

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

   BASEMENTAREA_MEDI  YEARS_BEGINEXPLUATATION_MEDI  YEARS_BUILD_MEDI  \
0             0.0369                        0.9722            0.6243
1             0.0529                        0.9851            0.7987
2                NaN                           NaN               NaN
3                NaN                           NaN               NaN
4                NaN                           NaN               NaN

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

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

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

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

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

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

   FLAG_DOCUMENT_2  FLAG_DOCUMENT_3  FLAG_DOCUMENT_4  FLAG_DOCUMENT_5  \
0                0                1                0                0
1                0                1                0                0
2                0                0                0                0
3                0                1                0                0
4                0                0                0                0

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

   FLAG_DOCUMENT_10  FLAG_DOCUMENT_11  FLAG_DOCUMENT_12  FLAG_DOCUMENT_13  \
0                 0                 0                 0                 0
1                 0                 0                 0                 0
2                 0                 0                 0                 0
```

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

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

   FLAG_DOCUMENT_18  FLAG_DOCUMENT_19  FLAG_DOCUMENT_20  FLAG_DOCUMENT_21  \
0                0                0                0                0
1                0                0                0                0
2                0                0                0                0
3                0                0                0                0
4                0                0                0                0

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

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

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

[4]:
```
# this a binary classification problem
```

[5]:
```
data.info()
```

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

[6]:
```
data.describe()
```

[6]:

	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
```

[7]:
```
# have both variable categorical& numerical here. so , going separate the␣
↪variable into 2.
```

[8]:
```
categorical_columns = data.select_dtypes(include=['object'])
```

[9]:
```
num_categorical_columns = categorical_columns.shape[1]
num_categorical_columns
```

[9]:
```
16
```

[10]:
```
categorical_variable_names = categorical_columns.columns.tolist()
```

[11]:
```
categorical_variable_names
```

[11]:
```
['NAME_CONTRACT_TYPE',
 'CODE_GENDER',
 'FLAG_OWN_CAR',
 'FLAG_OWN_REALTY',
 'NAME_TYPE_SUITE',
 'NAME_INCOME_TYPE',
 'NAME_EDUCATION_TYPE',
 'NAME_FAMILY_STATUS',
 'NAME_HOUSING_TYPE',
 'OCCUPATION_TYPE',
 'WEEKDAY_APPR_PROCESS_START',
 'ORGANIZATION_TYPE',
 'FONDKAPREMONT_MODE',
 'HOUSETYPE_MODE',
 'WALLSMATERIAL_MODE',
 'EMERGENCYSTATE_MODE']
```

[12]:
```
print("Number of categorical columns:", num_categorical_columns)
```

```
Number of categorical columns: 16
```

[13]:
```
from sklearn import preprocessing
lb = preprocessing.LabelEncoder()
data['NAME_CONTRACT_TYPE']= lb.fit_transform(data['NAME_CONTRACT_TYPE'])
data['CODE_GENDER']= lb.fit_transform(data['CODE_GENDER'])
data['FLAG_OWN_REALTY']= lb.fit_transform(data['FLAG_OWN_REALTY'])
data['FLAG_OWN_CAR']= lb.fit_transform(data['FLAG_OWN_CAR'])
data['NAME_TYPE_SUITE']= lb.fit_transform(data['NAME_TYPE_SUITE'])
data['NAME_INCOME_TYPE']= lb.fit_transform(data['NAME_INCOME_TYPE'])
data['NAME_EDUCATION_TYPE']= lb.fit_transform(data['NAME_EDUCATION_TYPE'])
data['NAME_FAMILY_STATUS']= lb.fit_transform(data['NAME_FAMILY_STATUS'])
data['NAME_HOUSING_TYPE']= lb.fit_transform(data['NAME_HOUSING_TYPE'])
data['OCCUPATION_TYPE']= lb.fit_transform(data['OCCUPATION_TYPE'])
data['WEEKDAY_APPR_PROCESS_START']= lb.
 ↪fit_transform(data['WEEKDAY_APPR_PROCESS_START'])
data['ORGANIZATION_TYPE']= lb.fit_transform(data['ORGANIZATION_TYPE'])
data['FONDKAPREMONT_MODE']= lb.fit_transform(data['FONDKAPREMONT_MODE'])
data['HOUSETYPE_MODE']= lb.fit_transform(data['HOUSETYPE_MODE'])
data['WALLSMATERIAL_MODE']= lb.fit_transform(data['WALLSMATERIAL_MODE'])
data['EMERGENCYSTATE_MODE']= lb.fit_transform(data['EMERGENCYSTATE_MODE'])
```

[14]:
```
pd.set_option('display.max_rows', None)
data.isnull().sum()
```

[14]:
```
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                     0
NAME_INCOME_TYPE                    0
NAME_EDUCATION_TYPE                 0
NAME_FAMILY_STATUS                  0
NAME_HOUSING_TYPE                   0
REGION_POPULATION_RELATIVE          0
DAYS_BIRTH                          0
DAYS_EMPLOYED                       0
DAYS_REGISTRATION                   0
DAYS_ID_PUBLISH                     0
OWN_CAR_AGE                    202929
FLAG_MOBIL                          0
FLAG_EMP_PHONE                      0
```

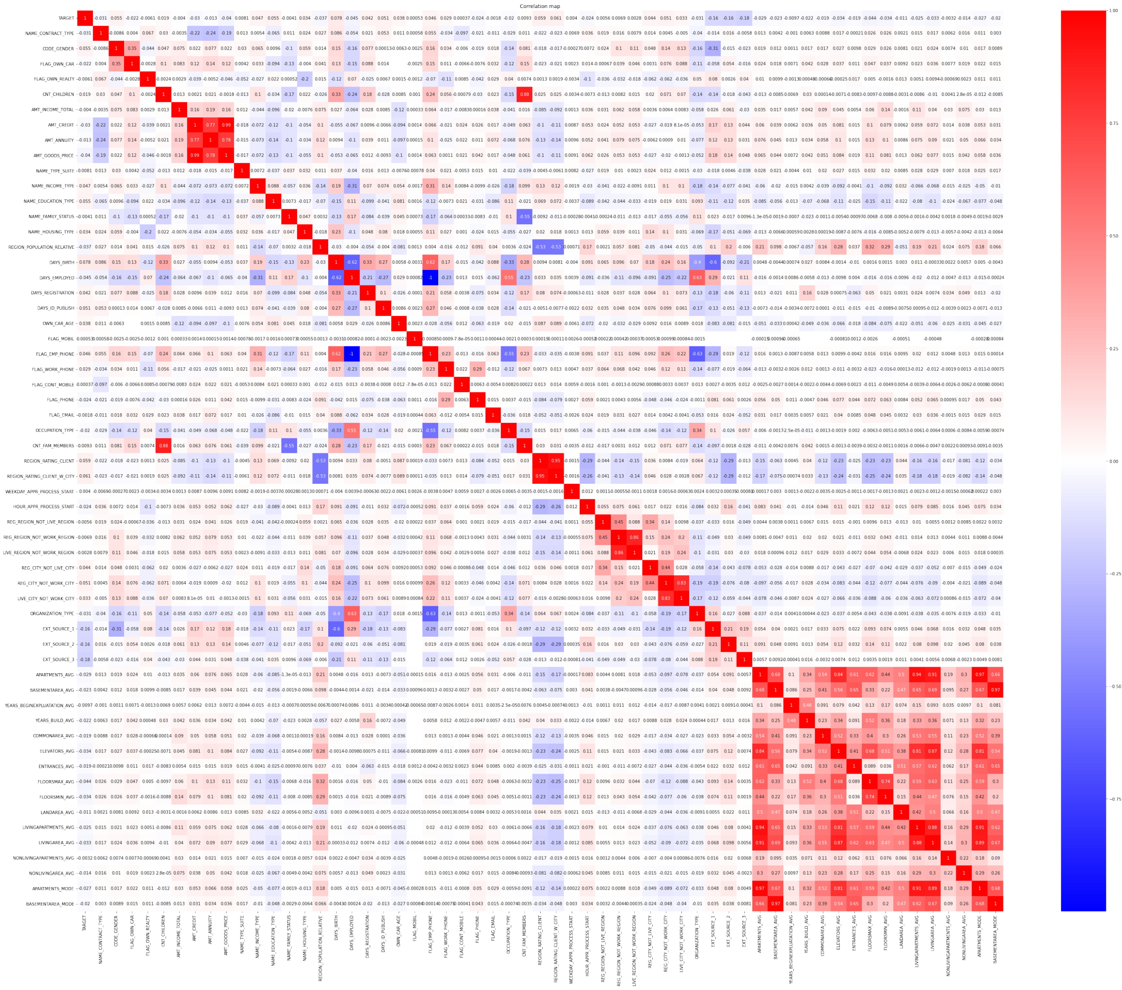
FLAG_WORK_PHONE	0
FLAG_CONT_MOBILE	0
FLAG_PHONE	0
FLAG_EMAIL	0
OCCUPATION_TYPE	0
CNT_FAM_MEMBERS	2
REGION_RATING_CLIENT	0
REGION_RATING_CLIENT_W_CITY	0
WEEKDAY_APPR_PROCESS_START	0
HOUR_APPR_PROCESS_START	0
REG_REGION_NOT_LIVE_REGION	0
REG_REGION_NOT_WORK_REGION	0
LIVE_REGION_NOT_WORK_REGION	0
REG_CITY_NOT_LIVE_CITY	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	0
HOUSETYPE_MODE	0
TOTALAREA_MODE	148431
WALLSMATERIAL_MODE	0
EMERGENCYSTATE_MODE	0
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
```

[15]:
```
# plethora of missing values in it. after check the correlation... plan to drop␣
↪the variables which have the least correlation with target...&
#balance variable replace with 0 , mean, mode
```

[16]:
```
# dataset contains 121 columns...so its difficult to check the correlation in a␣
↪heatmap...going to split the into two part. each part
# contains 60 columns
data1=data.iloc[:,1:60]




plt.figure(figsize=(50,40));
sns.heatmap(data1.corr(),annot=True,cmap='bwr');
plt.title("Correlation map")
plt.savefig('h1.jpg')
plt.show()
```



[17]:
```
data2 = data.iloc[:, -60:]

data2['TARGET'] = data['TARGET']

# Reorder the columns to include the 'TARGET' column as the first column
columns_order = ['TARGET'] + data2.columns[:-1].tolist()

data2 = data2[columns_order]
data2.head()
```

[17]:

	TARGET	COMMONAREA_MODE	ELEVATORS_MODE	ENTRANCES_MODE	FLOORSMAX_MODE	\
0	1	0.0144	0.0000	0.0690	0.0833	
1	0	0.0497	0.0806	0.0345	0.2917	
2	0	NaN	NaN	NaN	NaN	
3	0	NaN	NaN	NaN	NaN	
4	0	NaN	NaN	NaN	NaN	

	FLOORSMIN_MODE	LANDAREA_MODE	LIVINGAPARTMENTS_MODE	LIVINGAREA_MODE	\
0	0.1250	0.0377	0.022	0.0198	
1	0.3333	0.0128	0.079	0.0554	
2	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	
4	NaN	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	

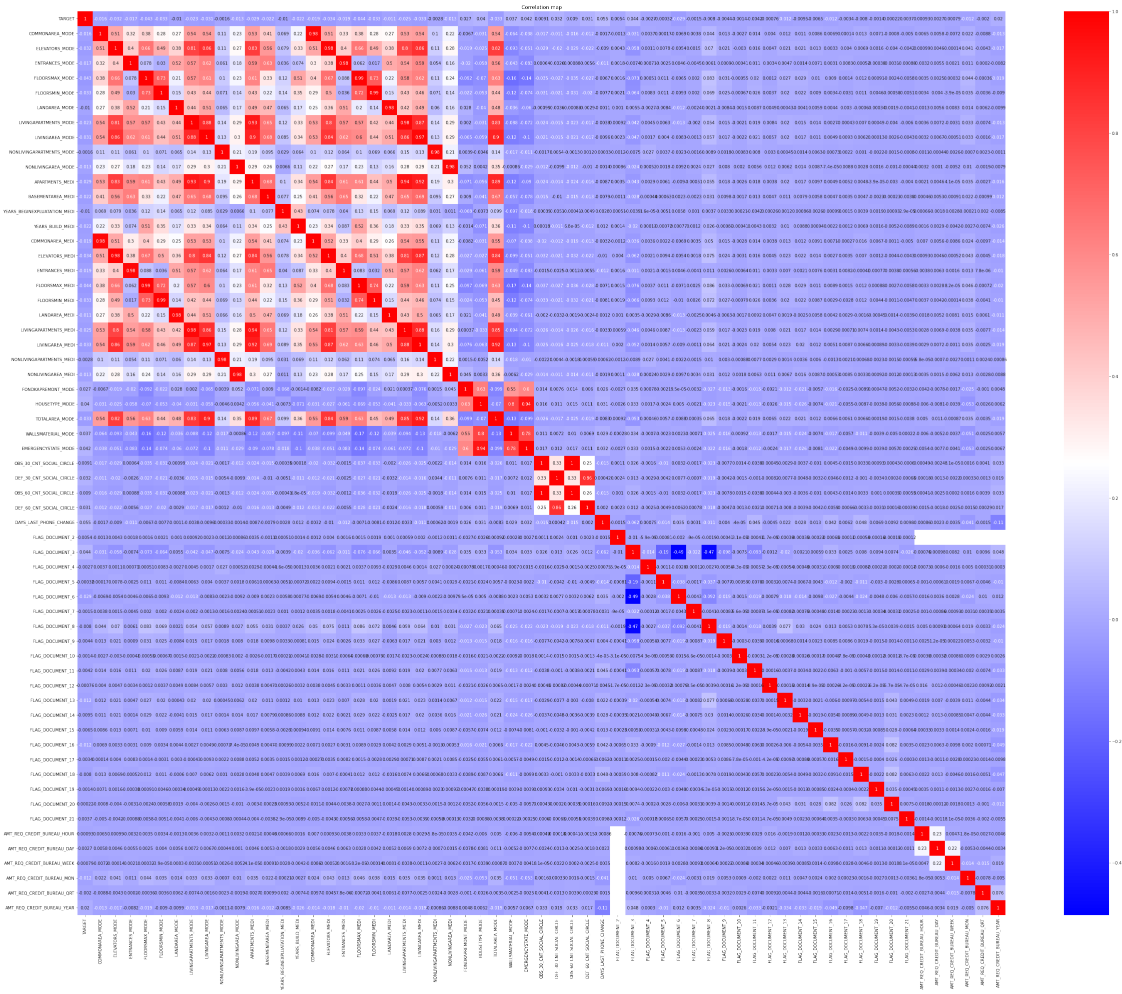
	NONLIVINGAPARTMENTS_MEDI	NONLIVINGAREA_MEDI	FONDKAPREMONT_MODE	\
0	0.0000	0.00	2	
1	0.0039	0.01	2	
2	NaN	NaN	4	
3	NaN	NaN	4	
4	NaN	NaN	4	

	HOUSETYPE_MODE	TOTALAREA_MODE	WALLSMATERIAL_MODE	EMERGENCYSTATE_MODE	\
0	0	0.0149	5	0	
1	0	0.0714	0	0	
2	3	NaN	7	2	
3	3	NaN	7	2	

```
4             3            NaN                  7                    2

   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
```

[18]:
```
plt.figure(figsize=(50, 40))
sns.heatmap(data2.corr(), annot=True, cmap='bwr')
plt.title("Correlation map")
plt.savefig('h1.jpg')
plt.show()
```



[18]:
```python
correlation = data.corr()['TARGET']
correlation_sorted = correlation.sort_values(ascending=True)
```

[19]:
```python
missing_values=data.isnull().sum()

result = pd.concat([correlation_sorted, missing_values], axis=1)
result.columns = ['Correlation', 'Missing Values']
result.head(123)
```

[19]:

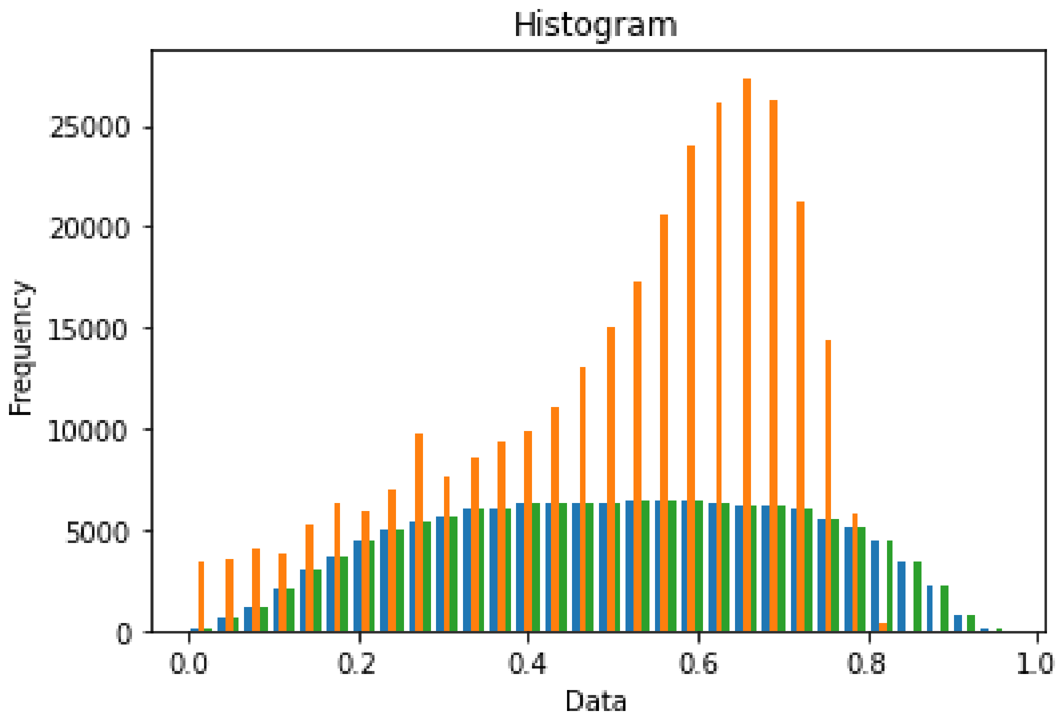
	Correlation	Missing Values
EXT_SOURCE_3	-0.178919	60965
EXT_SOURCE_2	-0.160472	660
EXT_SOURCE_1	-0.155317	173378
DAYS_EMPLOYED	-0.044932	0
FLOORSMAX_AVG	-0.044003	153020
FLOORSMAX_MEDI	-0.043768	153020

FLOORSMAX_MODE	-0.043226	153020
AMT_GOODS_PRICE	-0.039645	278
REGION_POPULATION_RELATIVE	-0.037227	0
ELEVATORS_AVG	-0.034199	163891
ELEVATORS_MEDI	-0.033863	163891
FLOORSMIN_AVG	-0.033614	208642
FLOORSMIN_MEDI	-0.033394	208642
LIVINGAREA_AVG	-0.032997	154350
LIVINGAREA_MEDI	-0.032739	154350
FLOORSMIN_MODE	-0.032698	208642
TOTALAREA_MODE	-0.032596	148431
ELEVATORS_MODE	-0.032131	163891
NAME_CONTRACT_TYPE	-0.030896	0
ORGANIZATION_TYPE	-0.030765	0
LIVINGAREA_MODE	-0.030685	154350
AMT_CREDIT	-0.030369	0
APARTMENTS_AVG	-0.029498	156061
APARTMENTS_MEDI	-0.029184	156061
FLAG_DOCUMENT_6	-0.028602	0
APARTMENTS_MODE	-0.027284	156061
LIVINGAPARTMENTS_AVG	-0.025031	210199
LIVINGAPARTMENTS_MEDI	-0.024621	210199
HOUR_APPR_PROCESS_START	-0.024166	0
FLAG_PHONE	-0.023806	0
LIVINGAPARTMENTS_MODE	-0.023393	210199
BASEMENTAREA_AVG	-0.022746	179943
YEARS_BUILD_MEDI	-0.022326	204488
YEARS_BUILD_AVG	-0.022149	204488
BASEMENTAREA_MEDI	-0.022081	179943
YEARS_BUILD_MODE	-0.022068	204488
FLAG_OWN_CAR	-0.021851	0
BASEMENTAREA_MODE	-0.019952	179943
OCCUPATION_TYPE	-0.019510	0
ENTRANCES_AVG	-0.019172	154828
ENTRANCES_MEDI	-0.019025	154828
COMMONAREA_MEDI	-0.018573	214865
COMMONAREA_AVG	-0.018550	214865
ENTRANCES_MODE	-0.017387	154828
COMMONAREA_MODE	-0.016340	214865
NONLIVINGAREA_AVG	-0.013578	169682
NONLIVINGAREA_MEDI	-0.013337	169682
AMT_ANNUITY	-0.012817	12
NONLIVINGAREA_MODE	-0.012711	169682
AMT_REQ_CREDIT_BUREAU_MON	-0.012462	41519
FLAG_DOCUMENT_16	-0.011615	0
FLAG_DOCUMENT_13	-0.011583	0
LANDAREA_MEDI	-0.011256	182590

LANDAREA_AVG	-0.010885	182590
LANDAREA_MODE	-0.010174	182590
YEARS_BEGINEXPLUATATION_MEDI	-0.009993	150007
YEARS_BEGINEXPLUATATION_AVG	-0.009728	150007
FLAG_DOCUMENT_14	-0.009464	0
YEARS_BEGINEXPLUATATION_MODE	-0.009036	150007
FLAG_DOCUMENT_8	-0.008040	0
FLAG_DOCUMENT_18	-0.007952	0
FLAG_DOCUMENT_15	-0.006536	0
FLAG_OWN_REALTY	-0.006148	0
FLAG_DOCUMENT_9	-0.004352	0
FLAG_DOCUMENT_11	-0.004229	0
NAME_FAMILY_STATUS	-0.004127	0
AMT_INCOME_TOTAL	-0.003982	0
FLAG_DOCUMENT_17	-0.003378	0
NONLIVINGAPARTMENTS_AVG	-0.003176	213514
NONLIVINGAPARTMENTS_MEDI	-0.002757	213514
FLAG_DOCUMENT_4	-0.002672	0
SK_ID_CURR	-0.002108	0
AMT_REQ_CREDIT_BUREAU_QRT	-0.002022	41519
FLAG_EMAIL	-0.001758	0
NONLIVINGAPARTMENTS_MODE	-0.001557	213514
FLAG_DOCUMENT_7	-0.001520	0
FLAG_DOCUMENT_10	-0.001414	0
FLAG_DOCUMENT_19	-0.001358	0
FLAG_DOCUMENT_12	-0.000756	0
FLAG_DOCUMENT_5	-0.000316	0
FLAG_DOCUMENT_20	0.000215	0
FLAG_CONT_MOBILE	0.000370	0
FLAG_MOBIL	0.000534	0
AMT_REQ_CREDIT_BUREAU_WEEK	0.000788	41519
AMT_REQ_CREDIT_BUREAU_HOUR	0.000930	41519
AMT_REQ_CREDIT_BUREAU_DAY	0.002704	41519
LIVE_REGION_NOT_WORK_REGION	0.002819	0
FLAG_DOCUMENT_21	0.003709	0
WEEKDAY_APPR_PROCESS_START	0.004002	0
FLAG_DOCUMENT_2	0.005417	0
REG_REGION_NOT_LIVE_REGION	0.005576	0
REG_REGION_NOT_WORK_REGION	0.006942	0
NAME_TYPE_SUITE	0.008074	0
OBS_60_CNT_SOCIAL_CIRCLE	0.009022	1021
OBS_30_CNT_SOCIAL_CIRCLE	0.009131	1021
CNT_FAM_MEMBERS	0.009308	2
CNT_CHILDREN	0.019187	0
AMT_REQ_CREDIT_BUREAU_YEAR	0.019930	41519
FONDKAPREMONT_MODE	0.026924	0
FLAG_WORK_PHONE	0.028524	0

```
DEF_60_CNT_SOCIAL_CIRCLE          0.031276        1021
DEF_30_CNT_SOCIAL_CIRCLE          0.032248        1021
LIVE_CITY_NOT_WORK_CITY           0.032518           0
NAME_HOUSING_TYPE                 0.034489           0
WALLSMATERIAL_MODE                0.037076           0
OWN_CAR_AGE                       0.037612      202929
HOUSETYPE_MODE                    0.040211           0
EMERGENCYSTATE_MODE               0.041955           0
DAYS_REGISTRATION                 0.041975           0
FLAG_DOCUMENT_3                   0.044346           0
REG_CITY_NOT_LIVE_CITY            0.044395           0
FLAG_EMP_PHONE                    0.045982           0
NAME_INCOME_TYPE                  0.046829           0
REG_CITY_NOT_WORK_CITY            0.050994           0
DAYS_ID_PUBLISH                   0.051457           0
CODE_GENDER                       0.054692           0
NAME_EDUCATION_TYPE               0.054699           0
DAYS_LAST_PHONE_CHANGE            0.055218           1
REGION_RATING_CLIENT              0.058899           0
REGION_RATING_CLIENT_W_CITY       0.060893           0
DAYS_BIRTH                        0.078239           0
TARGET                            1.000000           0
```

[20]:
```
import matplotlib.pyplot as plt

# Plot a histogram
plt.hist(data[['EXT_SOURCE_1','EXT_SOURCE_2','EXT_SOURCE_1']], bins=30)
plt.xlabel('Data')
plt.ylabel('Frequency')
plt.title('Histogram')
plt.show()
```




```
[21]: #im not planning to drop the independent variable...will do feature feature␣
       ↪selection.
      # missing values replace with mode, mean, 0
      data['EXT_SOURCE_2'].fillna(0, inplace=True)
```

```
[21]: data['AMT_REQ_CREDIT_BUREAU_YEAR'].replace([np.nan],␣
       ↪data['AMT_REQ_CREDIT_BUREAU_YEAR'].mean(), inplace=True)
      data['EXT_SOURCE_1'].replace([np.nan], data['EXT_SOURCE_1'].mean(),␣
       ↪inplace=True)
      data['EXT_SOURCE_2'].replace([np.nan], data['EXT_SOURCE_2'].mean(),␣
       ↪inplace=True)
      data['EXT_SOURCE_3'].replace([np.nan], data['EXT_SOURCE_3'].mean(),␣
       ↪inplace=True)
      data['YEARS_BEGINEXPLUATATION_AVG'].replace([np.nan],␣
       ↪data['YEARS_BEGINEXPLUATATION_AVG'].mean(), inplace=True)
      data['YEARS_BUILD_AVG'].replace([np.nan], data['YEARS_BUILD_AVG'].mean(),␣
       ↪inplace=True)
      data['NAME_CONTRACT_TYPE'].replace([np.nan], data['NAME_CONTRACT_TYPE'].mean(),␣
       ↪inplace=True)
      data['COMMONAREA_MODE'].replace([np.nan], data['COMMONAREA_MODE'].mean(),␣
       ↪inplace=True)
      data['ELEVATORS_MODE'].replace([np.nan], data['ELEVATORS_MODE'].mean(),␣
       ↪inplace=True)
```

```python
data['ENTRANCES_MODE'].replace([np.nan], data['ENTRANCES_MODE'].mean(), inplace=True)
data['FLOORSMAX_MODE'].replace([np.nan], data['FLOORSMAX_MODE'].mean(), inplace=True)
```

[22]:

```python
data['FLOORSMIN_MODE'].replace([np.nan], data['FLOORSMIN_MODE'].mean(), inplace=True)
data['LANDAREA_MODE'].replace([np.nan], data['LANDAREA_MODE'].mean(), inplace=True)
data['LIVINGAPARTMENTS_MODE'].replace([np.nan], data['LIVINGAPARTMENTS_MODE'].mean(), inplace=True)
data['LIVINGAREA_MODE'].replace([np.nan], data['LIVINGAREA_MODE'].mean(), inplace=True)
data['NONLIVINGAPARTMENTS_MODE'].replace([np.nan], data['NONLIVINGAPARTMENTS_MODE'].mean(), inplace=True)
data['NONLIVINGAREA_MODE'].replace([np.nan], data['NONLIVINGAREA_MODE'].mean(), inplace=True)
data['APARTMENTS_MEDI'].replace([np.nan], data['APARTMENTS_MEDI'].mean(), inplace=True)
data['BASEMENTAREA_MEDI'].replace([np.nan], data['BASEMENTAREA_MEDI'].mean(), inplace=True)
data['YEARS_BEGINEXPLUATATION_MEDI'].replace([np.nan], data['YEARS_BEGINEXPLUATATION_MEDI'].mean(), inplace=True)
data['YEARS_BUILD_MEDI'].replace([np.nan], data['YEARS_BUILD_MEDI'].mean(), inplace=True)
```

[23]:

```python
data['COMMONAREA_MEDI'].replace([np.nan], data['COMMONAREA_MEDI'].mean(), inplace=True)
data['ELEVATORS_MEDI'].replace([np.nan], data['ELEVATORS_MEDI'].mean(), inplace=True)
data['ENTRANCES_MEDI'].replace([np.nan], data['ENTRANCES_MEDI'].mean(), inplace=True)
data['FLOORSMAX_MEDI'].replace([np.nan], data['FLOORSMAX_MEDI'].mean(), inplace=True)
data['FLOORSMIN_MEDI'].replace([np.nan], data['FLOORSMIN_MEDI'].mean(), inplace=True)
data['LANDAREA_MEDI'].replace([np.nan], data['LANDAREA_MEDI'].mean(), inplace=True)
data['LIVINGAPARTMENTS_MEDI'].replace([np.nan], data['LIVINGAPARTMENTS_MEDI'].mean(), inplace=True)
data['LIVINGAREA_MEDI'].replace([np.nan], data['LIVINGAREA_MEDI'].mean(), inplace=True)
data['NONLIVINGAREA_MEDI'].replace([np.nan], data['NONLIVINGAREA_MEDI'].mean(), inplace=True)
```

```
data['NONLIVINGAPARTMENTS_MEDI'].replace([np.nan],␣
↪data['NONLIVINGAPARTMENTS_MEDI'].mean(), inplace=True)
data['FONDKAPREMONT_MODE'].replace([np.nan], data['FONDKAPREMONT_MODE'].mean(),␣
↪inplace=True)
data['HOUSETYPE_MODE'].replace([np.nan], data['HOUSETYPE_MODE'].mean(),␣
↪inplace=True)
data['TOTALAREA_MODE'].replace([np.nan], data['TOTALAREA_MODE'].mean(),␣
↪inplace=True)
data['WALLSMATERIAL_MODE'].replace([np.nan], data['WALLSMATERIAL_MODE'].mean(),␣
↪inplace=True)
data['EMERGENCYSTATE_MODE'].replace([np.nan], data['EMERGENCYSTATE_MODE'].
↪mean(), inplace=True)
data['OWN_CAR_AGE'].replace([np.nan], data['OWN_CAR_AGE'].mean(), inplace=True)
data['APARTMENTS_AVG'].replace([np.nan], data['APARTMENTS_AVG'].mean(),␣
↪inplace=True)
data['BASEMENTAREA_AVG'].replace([np.nan], data['BASEMENTAREA_AVG'].mean(),␣
↪inplace=True)
data['COMMONAREA_MEDI'].replace([np.nan], data['COMMONAREA_MEDI'].mean(),␣
↪inplace=True)
```

[24]:

```
data['COMMONAREA_AVG'].replace([np.nan], data['COMMONAREA_AVG'].mean(),␣
↪inplace=True)
data['ELEVATORS_AVG'].replace([np.nan], data['ELEVATORS_AVG'].mean(),␣
↪inplace=True)
data['ENTRANCES_AVG'].replace([np.nan], data['ENTRANCES_AVG'].mean(),␣
↪inplace=True)
data['FLOORSMAX_AVG'].replace([np.nan], data['FLOORSMAX_AVG'].mean(),␣
↪inplace=True)
data['FLOORSMIN_AVG'].replace([np.nan], data['FLOORSMIN_AVG'].mean(),␣
↪inplace=True)
data['LANDAREA_AVG'].replace([np.nan], data['LANDAREA_AVG'].mean(),␣
↪inplace=True)
data['LIVINGAPARTMENTS_AVG'].replace([np.nan], data['LIVINGAPARTMENTS_AVG'].
↪mean(), inplace=True)
data['LIVINGAREA_AVG'].replace([np.nan], data['LIVINGAREA_AVG'].mean(),␣
↪inplace=True)
data['NONLIVINGAPARTMENTS_AVG'].replace([np.nan],␣
↪data['NONLIVINGAPARTMENTS_AVG'].mean(), inplace=True)
data['NONLIVINGAREA_AVG'].replace([np.nan], data['NONLIVINGAREA_AVG'].mean(),␣
↪inplace=True)
data['BASEMENTAREA_MODE'].replace([np.nan], data['BASEMENTAREA_MODE'].mean(),␣
↪inplace=True)
data['YEARS_BEGINEXPLUATATION_MODE'].replace([np.nan],␣
↪data['YEARS_BEGINEXPLUATATION_MODE'].mean(), inplace=True)
```

```
data['YEARS_BUILD_MODE'].replace([np.nan], data['YEARS_BUILD_MODE'].mean(),␣
↪inplace=True)
```

[25]:
```
data['AMT_REQ_CREDIT_BUREAU_HOUR'].replace([np.nan],␣
↪data['AMT_REQ_CREDIT_BUREAU_HOUR'].mean(), inplace=True)
data['AMT_REQ_CREDIT_BUREAU_DAY'].replace([np.nan],␣
↪data['AMT_REQ_CREDIT_BUREAU_DAY'].mean(), inplace=True)
data['AMT_REQ_CREDIT_BUREAU_WEEK'].replace([np.nan],␣
↪data['AMT_REQ_CREDIT_BUREAU_WEEK'].mean(), inplace=True)
data['AMT_REQ_CREDIT_BUREAU_MON'].replace([np.nan],␣
↪data['AMT_REQ_CREDIT_BUREAU_MON'].mean(), inplace=True)
data['AMT_REQ_CREDIT_BUREAU_QRT'].replace([np.nan],␣
↪data['AMT_REQ_CREDIT_BUREAU_QRT'].mean(), inplace=True)
data['APARTMENTS_MODE'].replace([np.nan], data['APARTMENTS_MODE'].mean(),␣
↪inplace=True)
```

[27]:
```
# splitting the data
```

[26]:
```
X=data.iloc[:,2:]
X.head()
```

[26]:
```
   NAME_CONTRACT_TYPE  CODE_GENDER  FLAG_OWN_CAR  FLAG_OWN_REALTY  \
0                   0            1             0                1
1                   0            0             0                0
2                   1            1             1                1
3                   0            0             0                1
4                   0            1             0                1

   CNT_CHILDREN  AMT_INCOME_TOTAL  AMT_CREDIT  AMT_ANNUITY  AMT_GOODS_PRICE  \
0             0          202500.0    406597.5      24700.5         351000.0
1             0          270000.0   1293502.5      35698.5        1129500.0
2             0           67500.0    135000.0       6750.0         135000.0
3             0          135000.0    312682.5      29686.5         297000.0
4             0          121500.0    513000.0      21865.5         513000.0

   NAME_TYPE_SUITE  NAME_INCOME_TYPE  NAME_EDUCATION_TYPE  NAME_FAMILY_STATUS  \
0                6                 7                    4                   3
1                1                 4                    1                   1
2                6                 7                    4                   3
3                6                 7                    4                   0
4                6                 7                    4                   3

   NAME_HOUSING_TYPE  REGION_POPULATION_RELATIVE  DAYS_BIRTH  DAYS_EMPLOYED  \
0                  1                    0.018801       -9461           -637
1                  1                    0.003541      -16765          -1188
2                  1                    0.010032      -19046           -225
```

```
3                  1                    0.008019      -19005           -3039
4                  1                    0.028663      -19932           -3038

   DAYS_REGISTRATION  DAYS_ID_PUBLISH  OWN_CAR_AGE  FLAG_MOBIL  \
0            -3648.0            -2120    12.061091           1
1            -1186.0             -291    12.061091           1
2            -4260.0            -2531    26.000000           1
3            -9833.0            -2437    12.061091           1
4            -4311.0            -3458    12.061091           1

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

   OCCUPATION_TYPE  CNT_FAM_MEMBERS  REGION_RATING_CLIENT  \
0                8              1.0                     2
1                3              2.0                     1
2                8              1.0                     2
3                8              2.0                     2
4                3              1.0                     2

   REGION_RATING_CLIENT_W_CITY  WEEKDAY_APPR_PROCESS_START  \
0                            2                           6
1                            1                           1
2                            2                           1
3                            2                           6
4                            2                           4

   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_1  EXT_SOURCE_2  EXT_SOURCE_3  \
0                  5      0.083037      0.262949      0.139376
1                 39      0.311267      0.622246      0.510853
2                 11      0.502130      0.555912      0.729567
3                  5      0.502130      0.650442      0.510853
4                 37      0.502130      0.322738      0.510853

   APARTMENTS_AVG  BASEMENTAREA_AVG  YEARS_BEGINEXPLUATATION_AVG  \
0         0.02470          0.036900                     0.972200
1         0.09590          0.052900                     0.985100
2         0.11744          0.088442                     0.977735
3         0.11744          0.088442                     0.977735
4         0.11744          0.088442                     0.977735

   YEARS_BUILD_AVG  COMMONAREA_AVG  ELEVATORS_AVG  ENTRANCES_AVG  \
0         0.619200        0.014300       0.000000       0.069000
1         0.796000        0.060500       0.080000       0.034500
2         0.752471        0.044621       0.078942       0.149725
3         0.752471        0.044621       0.078942       0.149725
4         0.752471        0.044621       0.078942       0.149725

   FLOORSMAX_AVG  FLOORSMIN_AVG  LANDAREA_AVG  LIVINGAPARTMENTS_AVG  \
0       0.083300       0.125000      0.036900              0.020200
1       0.291700       0.333300      0.013000              0.077300
2       0.226282       0.231894      0.066333              0.100775
3       0.226282       0.231894      0.066333              0.100775
4       0.226282       0.231894      0.066333              0.100775

   LIVINGAREA_AVG  NONLIVINGAPARTMENTS_AVG  NONLIVINGAREA_AVG  \
0        0.019000                 0.000000           0.000000
1        0.054900                 0.003900           0.009800
2        0.107399                 0.008809           0.028358
3        0.107399                 0.008809           0.028358
4        0.107399                 0.008809           0.028358

   APARTMENTS_MODE  BASEMENTAREA_MODE  YEARS_BEGINEXPLUATATION_MODE  \
0         0.025200           0.038300                      0.972200
1         0.092400           0.053800                      0.985100
2         0.114231           0.087543                      0.977065
3         0.114231           0.087543                      0.977065
4         0.114231           0.087543                      0.977065
```

```
   YEARS_BUILD_MODE  COMMONAREA_MODE  ELEVATORS_MODE  ENTRANCES_MODE  \
0          0.634100         0.014400         0.00000        0.069000   
1          0.804000         0.049700         0.08060        0.034500   
2          0.759637         0.042553         0.07449        0.145193   
3          0.759637         0.042553         0.07449        0.145193   
4          0.759637         0.042553         0.07449        0.145193   

   FLOORSMAX_MODE  FLOORSMIN_MODE  LANDAREA_MODE  LIVINGAPARTMENTS_MODE  \
0        0.083300        0.125000       0.037700               0.022000   
1        0.291700        0.333300       0.012800               0.079000   
2        0.222315        0.228058       0.064958               0.105645   
3        0.222315        0.228058       0.064958               0.105645   
4        0.222315        0.228058       0.064958               0.105645   

   LIVINGAREA_MODE  NONLIVINGAPARTMENTS_MODE  NONLIVINGAREA_MODE  \
0         0.019800                  0.000000            0.000000   
1         0.055400                  0.000000            0.000000   
2         0.105975                  0.008076            0.027022   
3         0.105975                  0.008076            0.027022   
4         0.105975                  0.008076            0.027022   

   APARTMENTS_MEDI  BASEMENTAREA_MEDI  YEARS_BEGINEXPLUATATION_MEDI  \
0          0.02500           0.036900                      0.972200   
1          0.09680           0.052900                      0.985100   
2          0.11785           0.087955                      0.977752   
3          0.11785           0.087955                      0.977752   
4          0.11785           0.087955                      0.977752   

   YEARS_BUILD_MEDI  COMMONAREA_MEDI  ELEVATORS_MEDI  ENTRANCES_MEDI  \
0          0.624300         0.014400        0.000000        0.069000   
1          0.798700         0.060800        0.080000        0.034500   
2          0.755746         0.044595        0.078078        0.149213   
3          0.755746         0.044595        0.078078        0.149213   
4          0.755746         0.044595        0.078078        0.149213   

   FLOORSMAX_MEDI  FLOORSMIN_MEDI  LANDAREA_MEDI  LIVINGAPARTMENTS_MEDI  \
0        0.083300        0.125000       0.037500               0.020500   
1        0.291700        0.333300       0.013200               0.078700   
2        0.225897        0.231625       0.067169               0.101954   
3        0.225897        0.231625       0.067169               0.101954   
4        0.225897        0.231625       0.067169               0.101954   

   LIVINGAREA_MEDI  NONLIVINGAPARTMENTS_MEDI  NONLIVINGAREA_MEDI  \
0         0.019300                  0.000000            0.000000   
1         0.055800                  0.003900            0.010000   
2         0.108607                  0.008651            0.028236   
3         0.108607                  0.008651            0.028236   
```

```
4        0.108607                  0.008651           0.028236

   FONDKAPREMONT_MODE  HOUSETYPE_MODE  TOTALAREA_MODE  WALLSMATERIAL_MODE  \
0                   2               0        0.014900                   5
1                   2               0        0.071400                   0
2                   4               3        0.102547                   7
3                   4               3        0.102547                   7
4                   4               3        0.102547                   7

   EMERGENCYSTATE_MODE  OBS_30_CNT_SOCIAL_CIRCLE  DEF_30_CNT_SOCIAL_CIRCLE  \
0                    0                       2.0                       2.0
1                    0                       1.0                       0.0
2                    2                       0.0                       0.0
3                    2                       2.0                       0.0
4                    2                       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                    0.000000                      0.000
1                    0.000000                      0.000
2                    0.000000                      0.000
3                    0.006402                      0.007
4                    0.000000                      0.000

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

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

[27]:
```
y=data.TARGET
```

[28]:
```
X.shape
```

[28]:
```
(307511, 120)
```

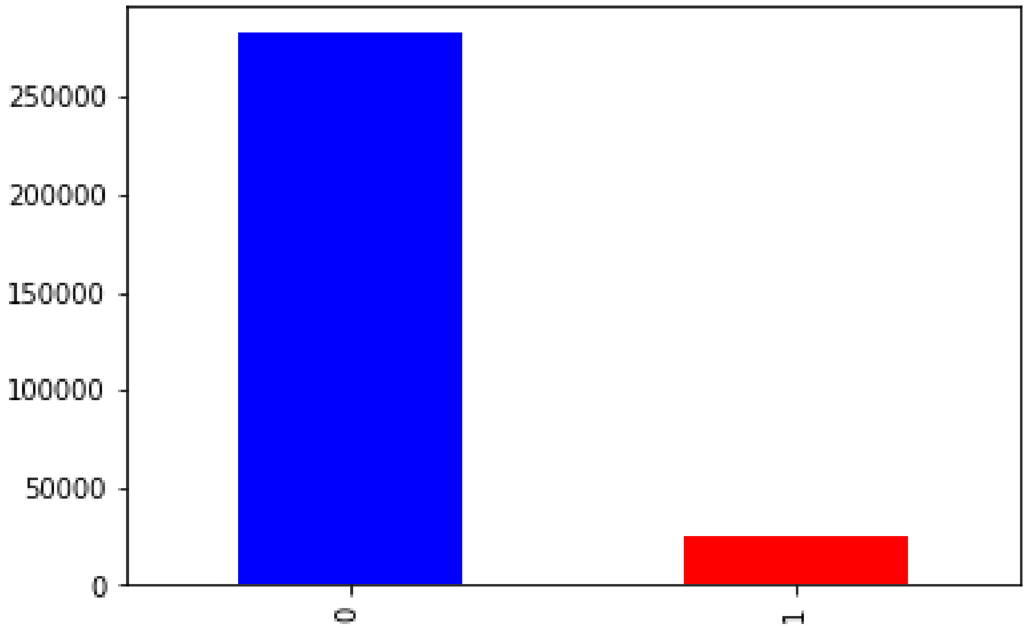
[29]:
```
y.shape
```

[29]:
```
(307511,)
```

[30]:
```
y.value_counts()*100/data.shape[0]
```

[30]:
```
0    91.927118
1     8.072882
Name: TARGET, dtype: float64
```

[31]:
```
y.value_counts().plot(kind='bar', color=["blue","red"])
plt.show()
```



[32]:
```
# target is imbalance..
```

[37]:
```
# inorder to overcome imbalance im using the SMOTE technique
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.
 ↪25,random_state=0)
st= StandardScaler()

X_train_std=st.fit_transform(X_train)
X_test_std= st.fit_transform(X_test)
```

[40]:
```
from imblearn.over_sampling import SMOTE
from sklearn.impute import SimpleImputer
```

[41]:
```
imputer = SimpleImputer(strategy='constant')
X = pd.DataFrame(imputer.fit_transform(X), columns=X.columns)

smk = SMOTE()
X_train_smote, y_train_smote = smk.fit_resample(X, y)
```

[42]:
```
X_train, X_test, y_train, y_test =␣
↪train_test_split(X_train_smote,y_train_smote, test_size=0.25, random_state=0)
```

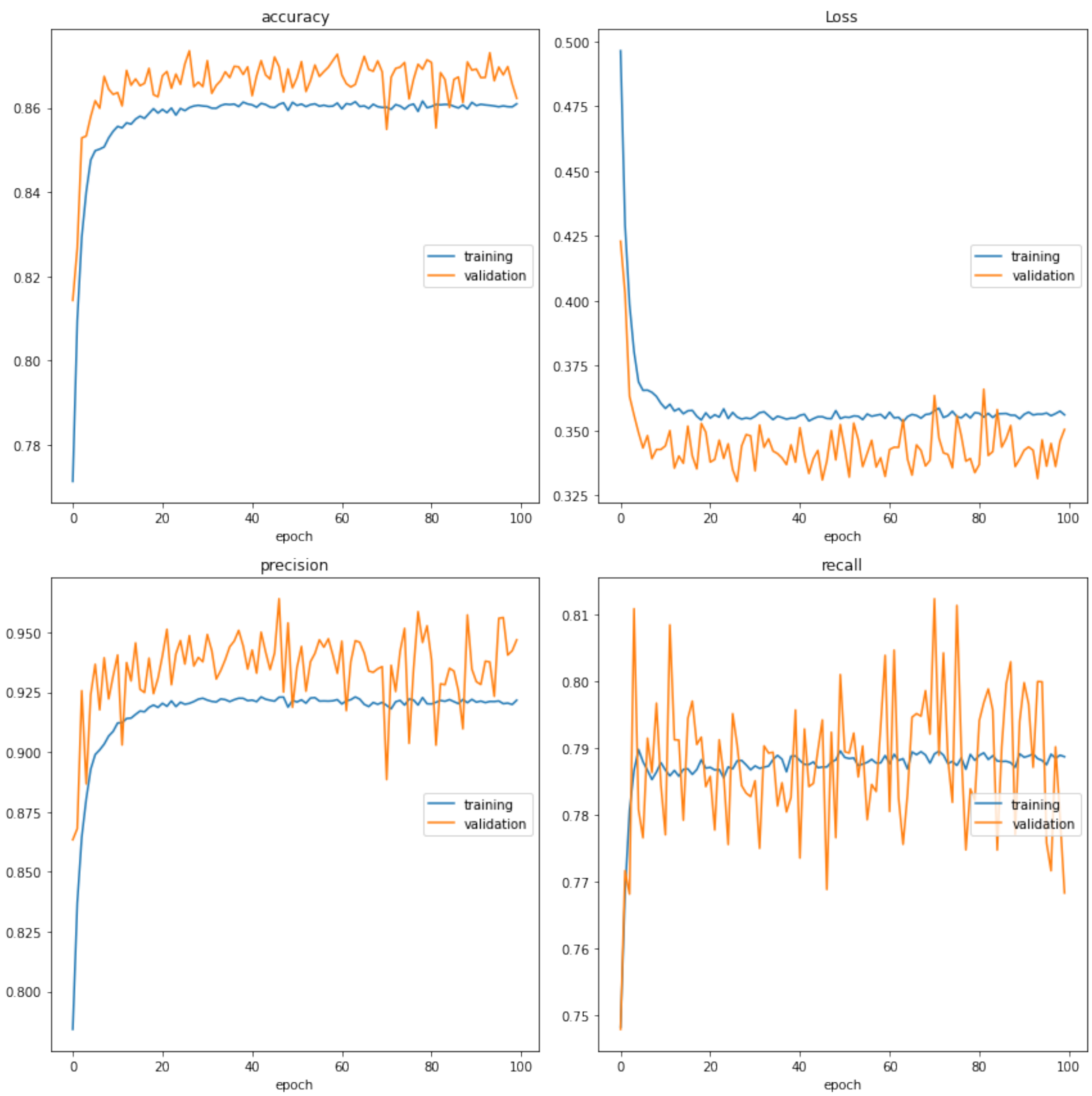
[43]:
```
scaler = StandardScaler()
X_train_std = scaler.fit_transform(X_train)
X_test_std=scaler.transform(X_test)
```

[44]:
```
from tensorflow.keras.layers import Dense,BatchNormalization,Input
from tensorflow.keras.models import Sequential
from livelossplot import PlotLossesKerasTF
from tensorflow.keras.metrics import Precision,Recall
from tensorflow.keras.regularizers import l2
```

[45]:
```
model=Sequential()
model.add(Input(shape=(120,)))

model.add(Dense(20,activation='relu',kernel_regularizer=l2(l2=0.001)))
model.add(Dense(20,activation='relu'))
model.add(BatchNormalization())

model.add(Dense(20,activation='relu',kernel_regularizer=l2(l2=0.001)))
model.add(BatchNormalization())

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

[46]:
```
model.compile(loss='binary_crossentropy',optimizer='adam',
              metrics=['accuracy',Precision(),Recall()])
```

[47]:
```
model.
↪fit(X_train_std,y_train,epochs=100,batch_size=32,callbacks=[PlotLossesKerasTF()],validation
```




```
accuracy
	training              (min:    0.771, max:    0.862, cur:    0.861)
	validation            (min:    0.814, max:    0.874, cur:    0.862)
Loss
	training              (min:    0.353, max:    0.496, cur:    0.356)
	validation            (min:    0.330, max:    0.423, cur:    0.350)
precision
	training              (min:    0.784, max:    0.923, cur:    0.922)
	validation            (min:    0.863, max:    0.964, cur:    0.947)
recall
	training              (min:    0.748, max:    0.790, cur:    0.789)
	validation            (min:    0.748, max:    0.812, cur:    0.768)
13251/13251 [==============================] - 98s 7ms/step - loss: 0.3560 -
```

```
accuracy: 0.8610 - precision: 0.9217 - recall: 0.7887 - val_loss: 0.3504 -
val_accuracy: 0.8623 - val_precision: 0.9469 - val_recall: 0.7683
```

[47]: <keras.callbacks.History at 0x7f71cd2a9210>

The deep learning model for loan default prediction has demonstrated strong performance based on the training and validation results. Throughout the training process, the model achieved accuracy ranging from 77.1% to 86.2%, while validation accuracy ranged from 81.4% to 87.4%. This indicates that the model generalizes well to unseen data, reducing the risk of overfitting. The model's precision during training ranged from 78.4% to 92.3% and during validation from 86.3% to 96.4%, demonstrating its ability to make accurate positive predictions. Additionally, the recall ranged from 74.8% to 79.0% during training and from 74.8% to 81.2% during validation, indicating the model's capacity to correctly identify positive cases.

In the final evaluation on the validation set, the deep learning model achieved an accuracy of 86.23%, precision of 94.69%, and recall of 76.83%. These metrics highlight the model's effectiveness in predicting loan default risk, making it a valuable tool for financial institutions in decision-making and risk management processes.

[ ]: