CREDIT EDA CASE STUDY

Shital Jadhav

Problem Statement – I :

- This case study aims to give an idea of applying EDA in a real business scenario., apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimise the risk of losing money while lending to customers.

Business Objectives :

- This case study aims to identify patterns which indicate if a client has difficulty paying their instalments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.

DATA VISUALISATION OF APPLICATIONS DATA

UNIVARIANT ANALYSIS OF CATEGORICAL VARIABLES FOR target=0(all other)

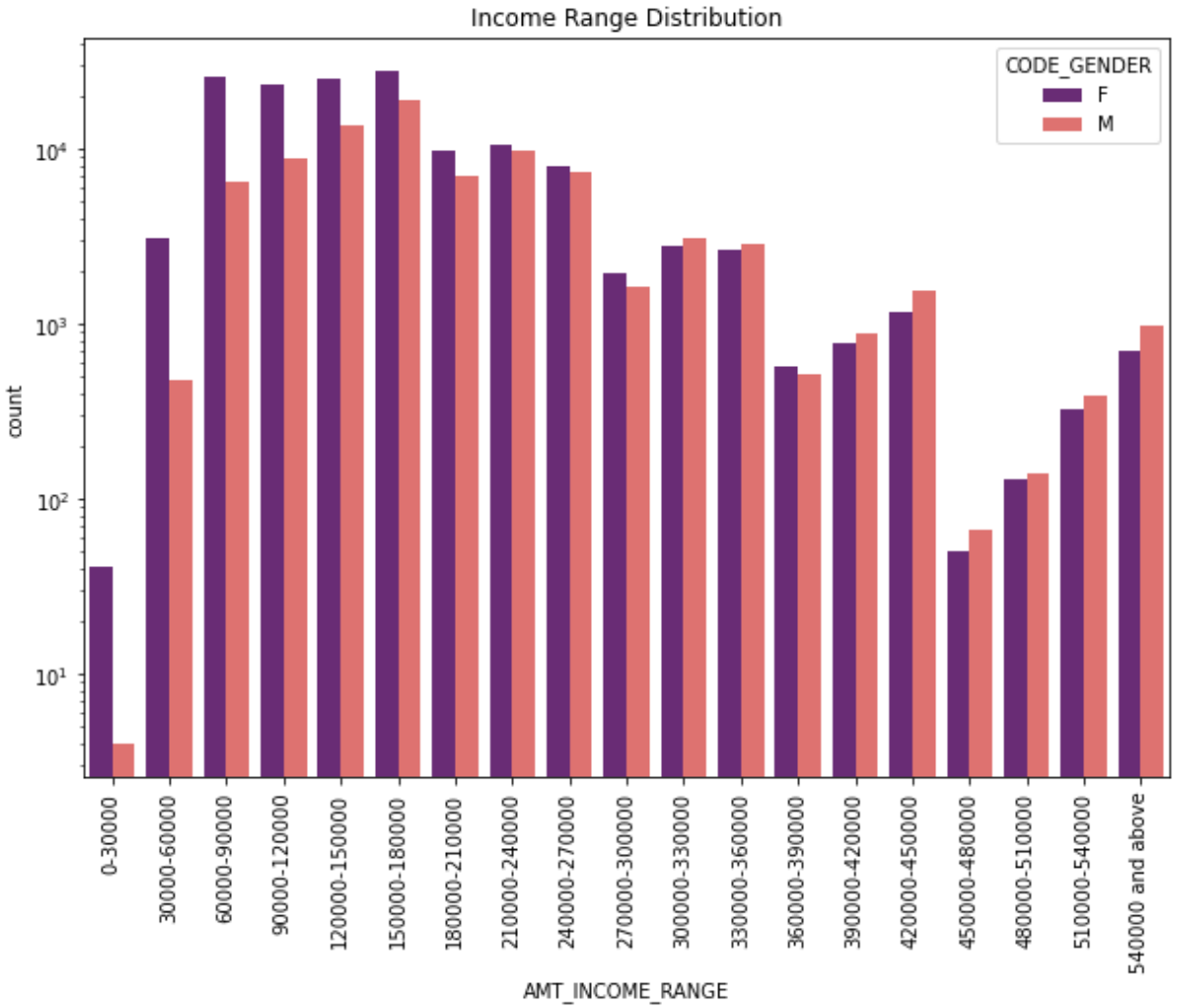
Inference from Income Range Distribution graph:

1.Female counts higher than male count.

2.This graph show that females are more than male for higher income range

3.Income range from 150000 to 180000 is having more number of credits.

4.Very less count for income range 450000 and above.



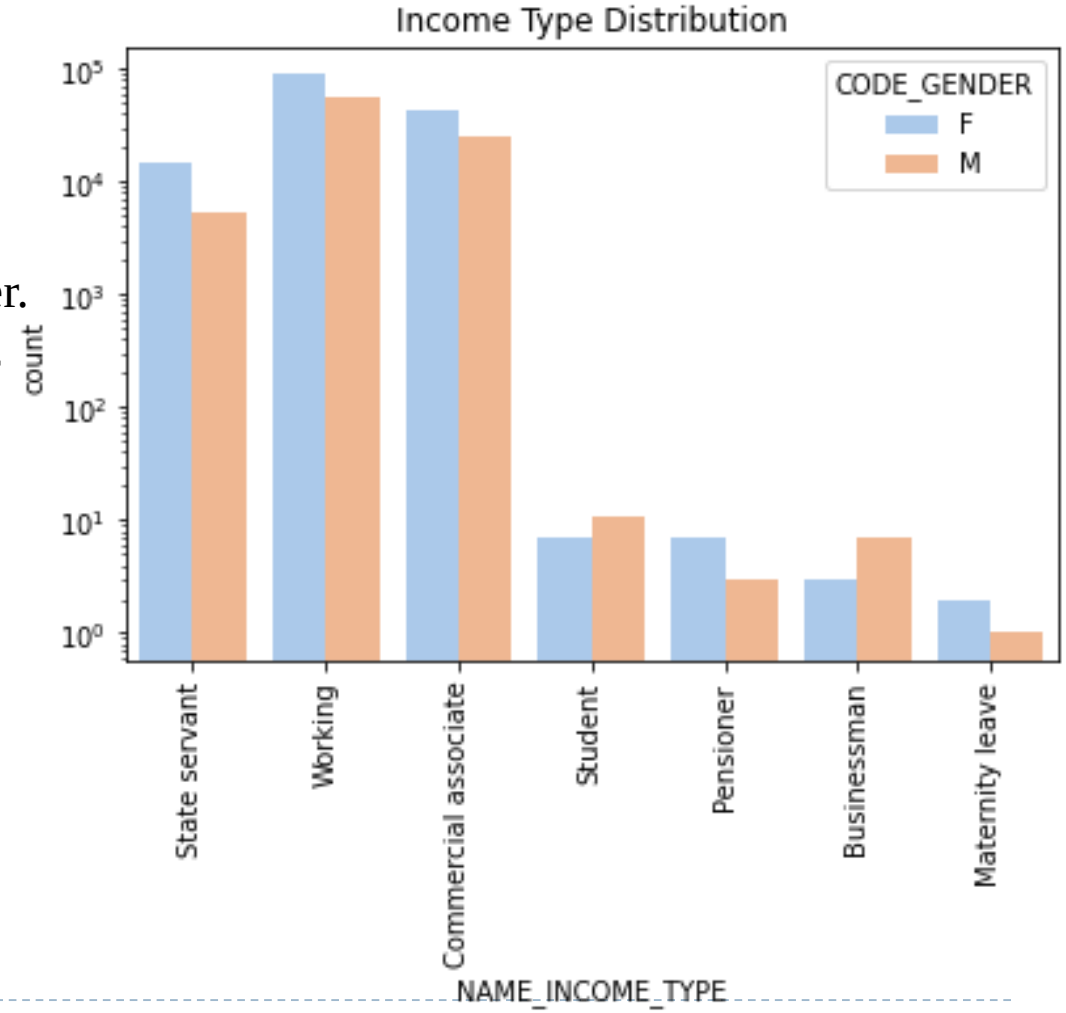
Plotting for income type:

Inference for income type:

1.State servant , Working and commercial associate are having the number of credits are higher than other.

2.Females are having more number of credits than male.

3.Less number of credits for income type "Maternity leave".

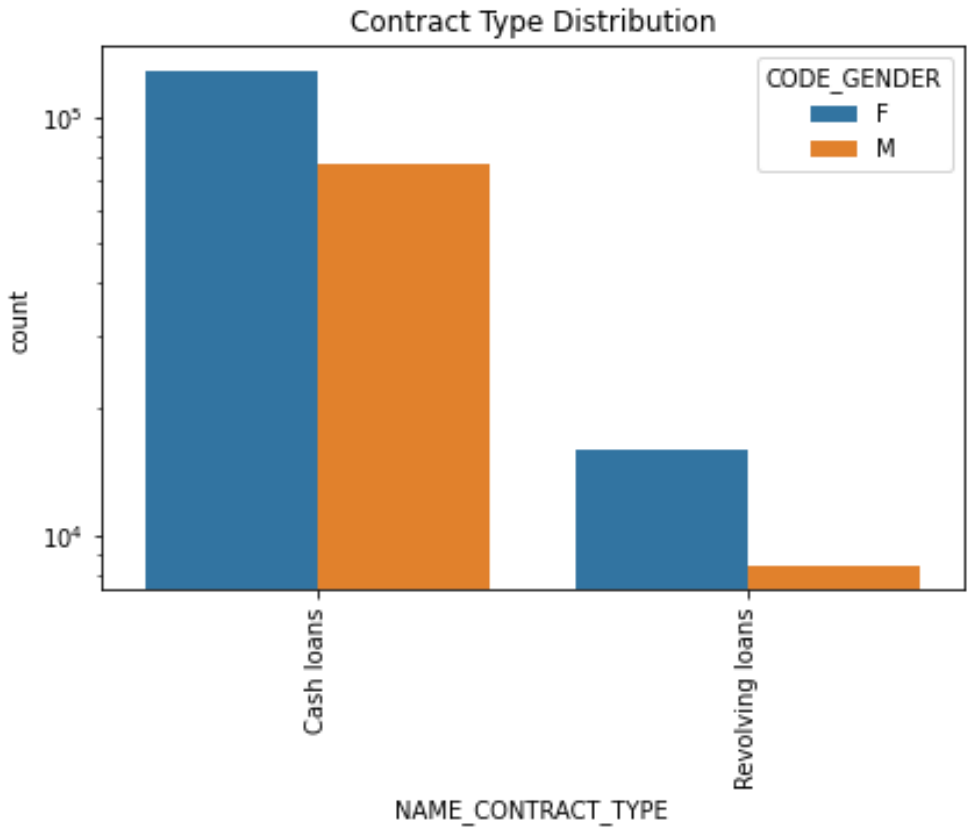


Plotting for contract type:

Inference from Contract Type Distribution:

1.'Cash loans' contract type is having higher number of credits than ‘Revolving loans’ contract type.

2. Female is leading for applying credits.



UNIVARIANT ANALYSIS OF CATEGORICAL VARIABLES FOR target=1(Client with payment difficulties)

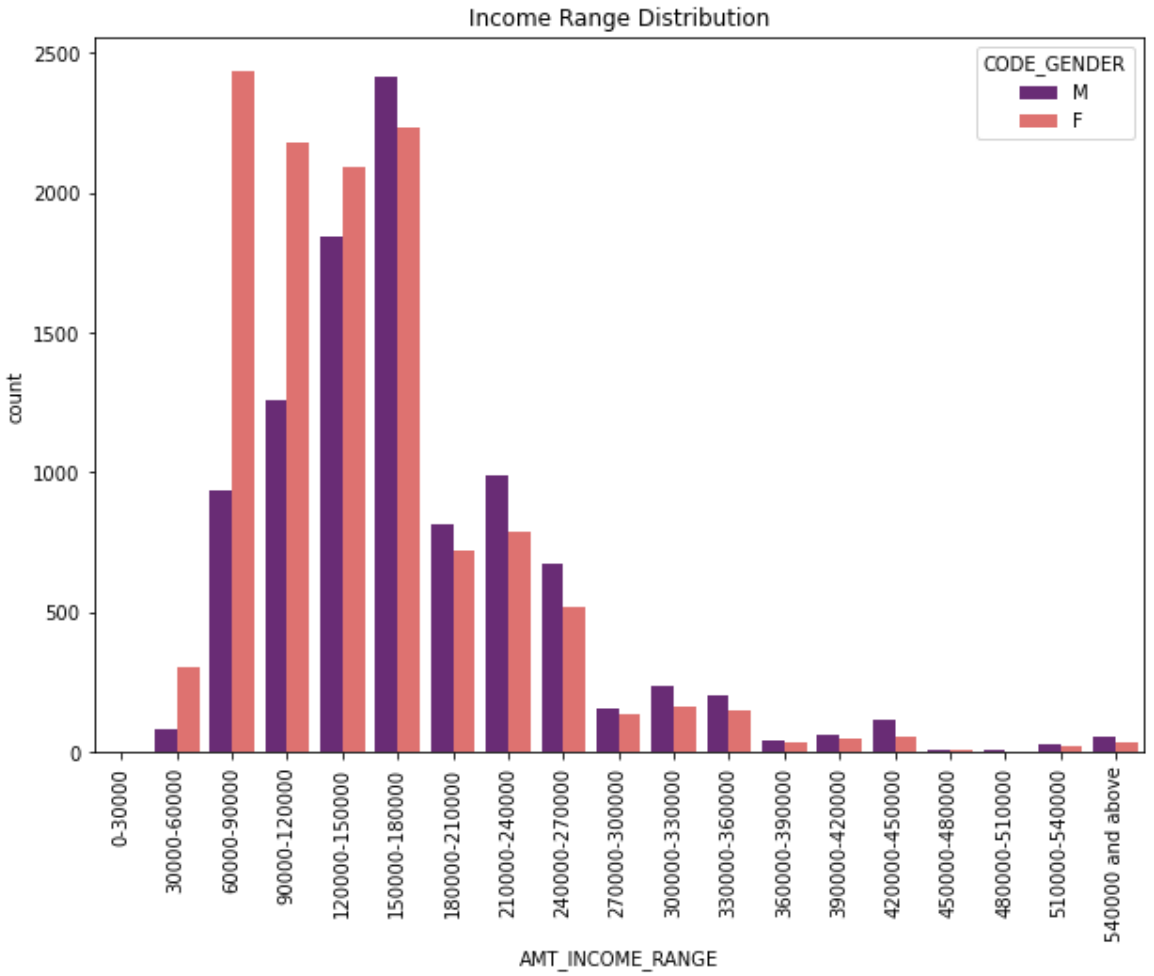
Inference from Income Range Distribution graph:

1.Male counts higher than female count.

2.This graph show that females are more than male for higher income range

3.Income range from 150000 to 180000 is having more number of credits.

4.Very less count for income range 450000 and above.



Plotting for income type:

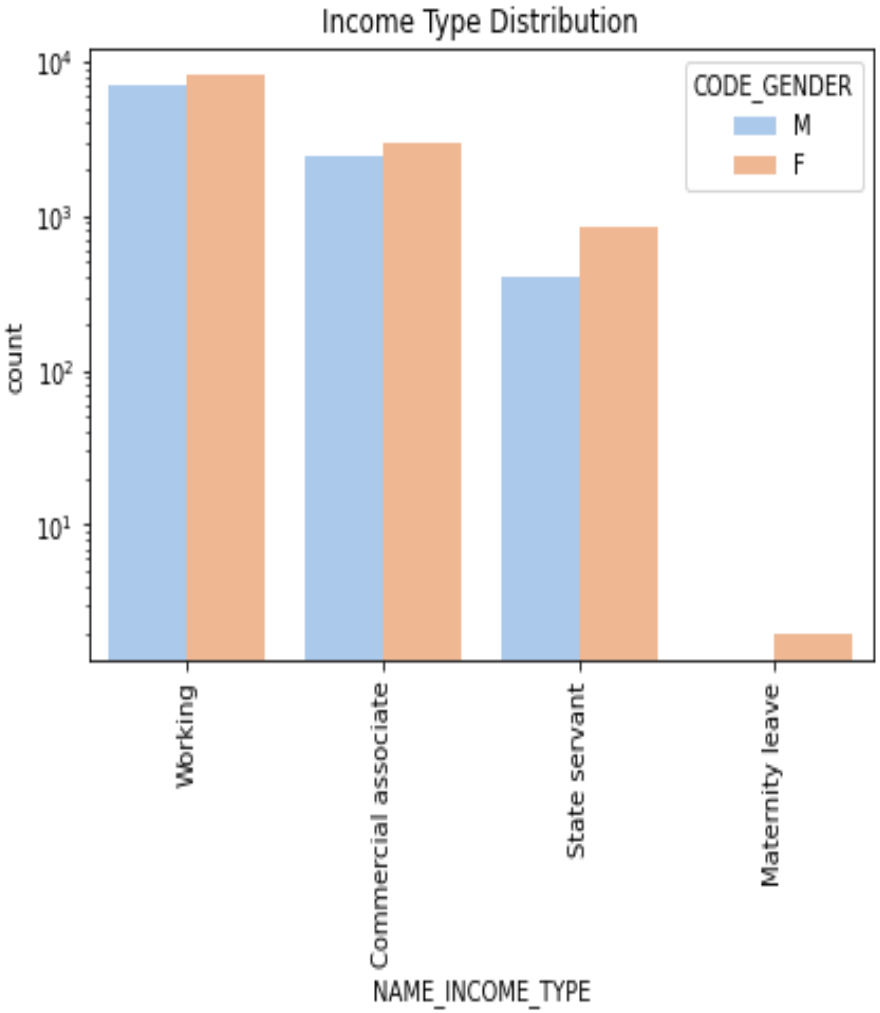
Inference for income type:

1.State servant, Working and commercial associate are having the number of credits are higher than other.

2.Females are having more number of credits than male.

3.Less number of credits for income type 'Maternity leave'.

4.No income type for ‘student’ , ’pensioner’ and ‘Businessman’ this indicates that they don’t do any late payments or they do not have payment difficulties.

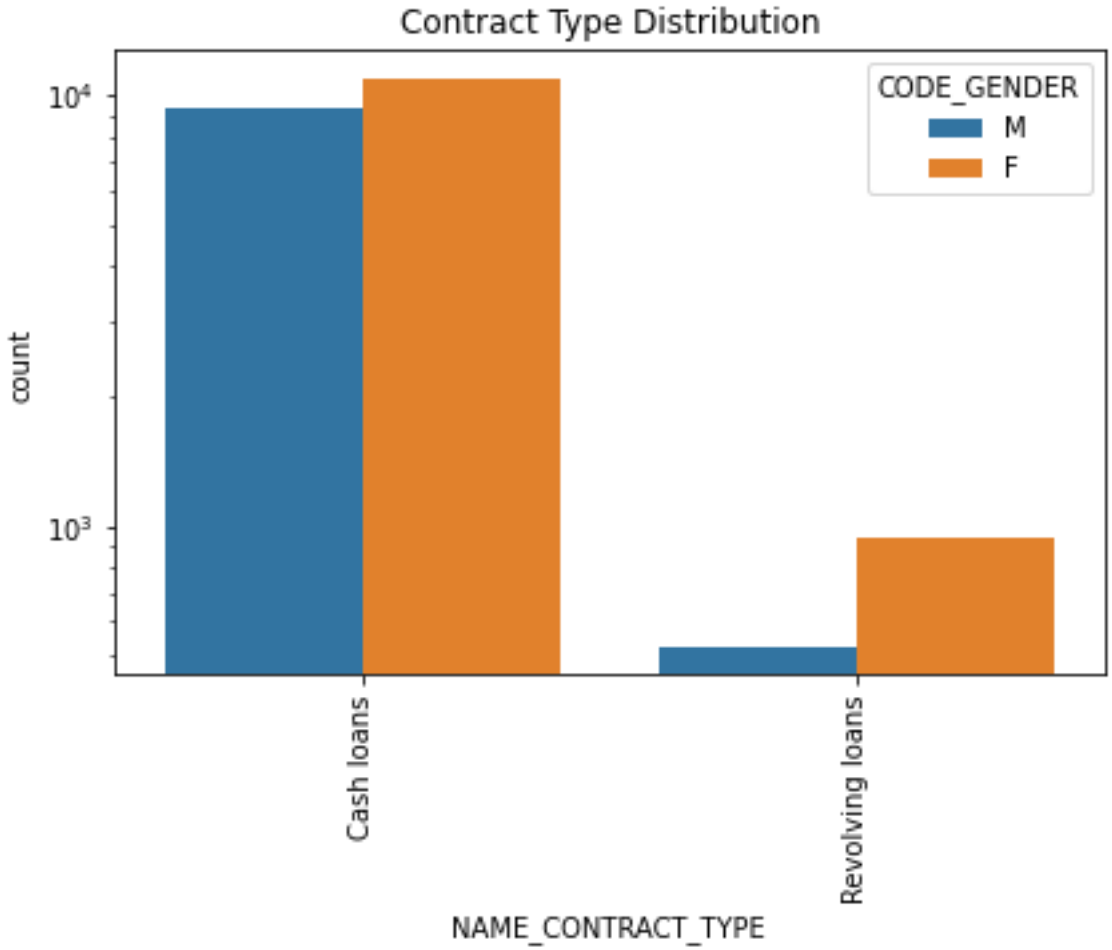


Plotting for contract type:

Inference from Contract Type Distribution:

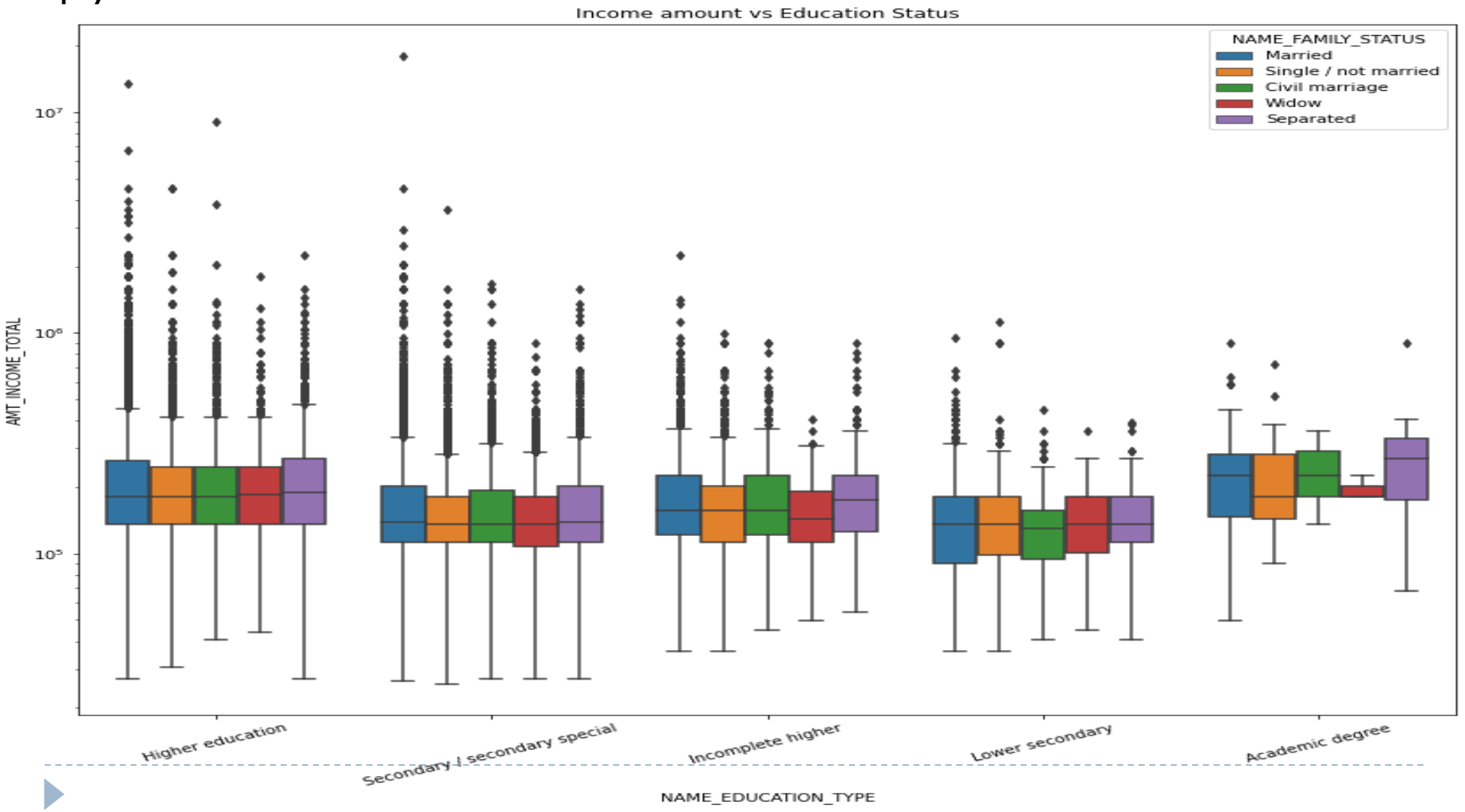
1.'Cash loans' contract type is having higher number of credits than ‘Revolving loans contract type.

2. Female is leading for applying credits.



BIVARIANT ANALYSIS OF NUMERICAL VARIABLES FOR target=0(All others)

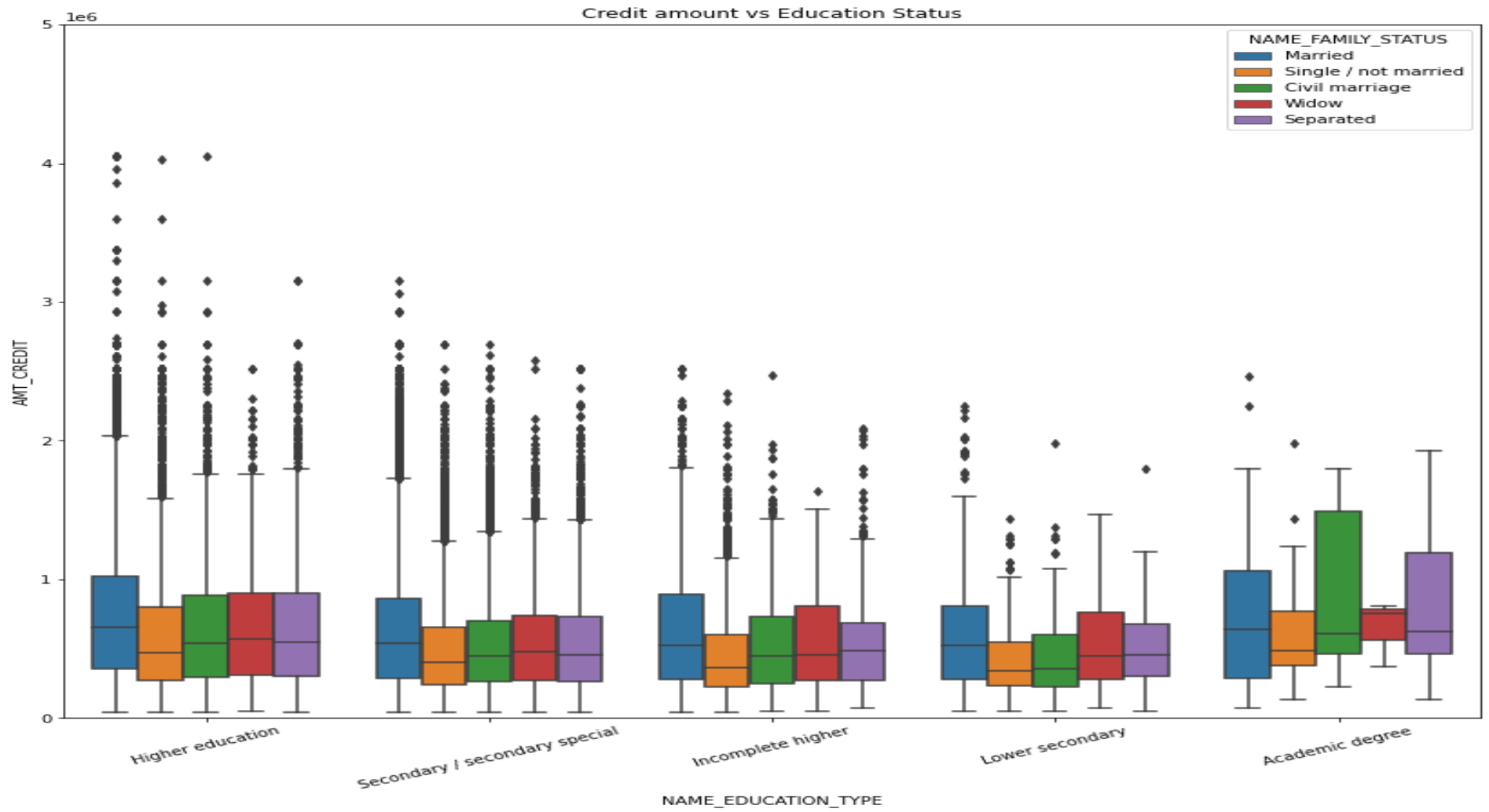
Income amount Vs Education Status in different types of family of clients with no payments difficulties:



Inferences:

1. More outliers are present in family status of 'marriage', 'single' and 'civil marriage' with higher education
2. Civil marriage for academic degree is having most of the credits in the third quartile.

Amount Vs Education Status in different types of family of clients with no payments difficulties:



Inferences:

1.'Higher education‘ category have income amount mostly equal with family status. Also many outliers are present .

2.'Academic degree' is having least outliers and income amount is higher among the all.

3.'Lower secondary' of single/not married family status are having less income amount than others.

BIVARIANT ANALYSIS OF NUMERICAL VARIABLES FOR target=1(Client with payment difficulties)

Income amount Verses Education Status in different types of family of Client with payment difficulties)

Inferences:

1. 'Higher education‘ and 'Secondary/Secondary special' education categories have income amount mostly equal with family status. Also many outliers are present .

2. Married in 'Academic degree' is having no outliers and income amount is higher among the all.

3. 'Lower secondary' of widow family status are having less income amount than others.

DATA VISUALISATION OF PREVIOUS APPLICATIONS DATA

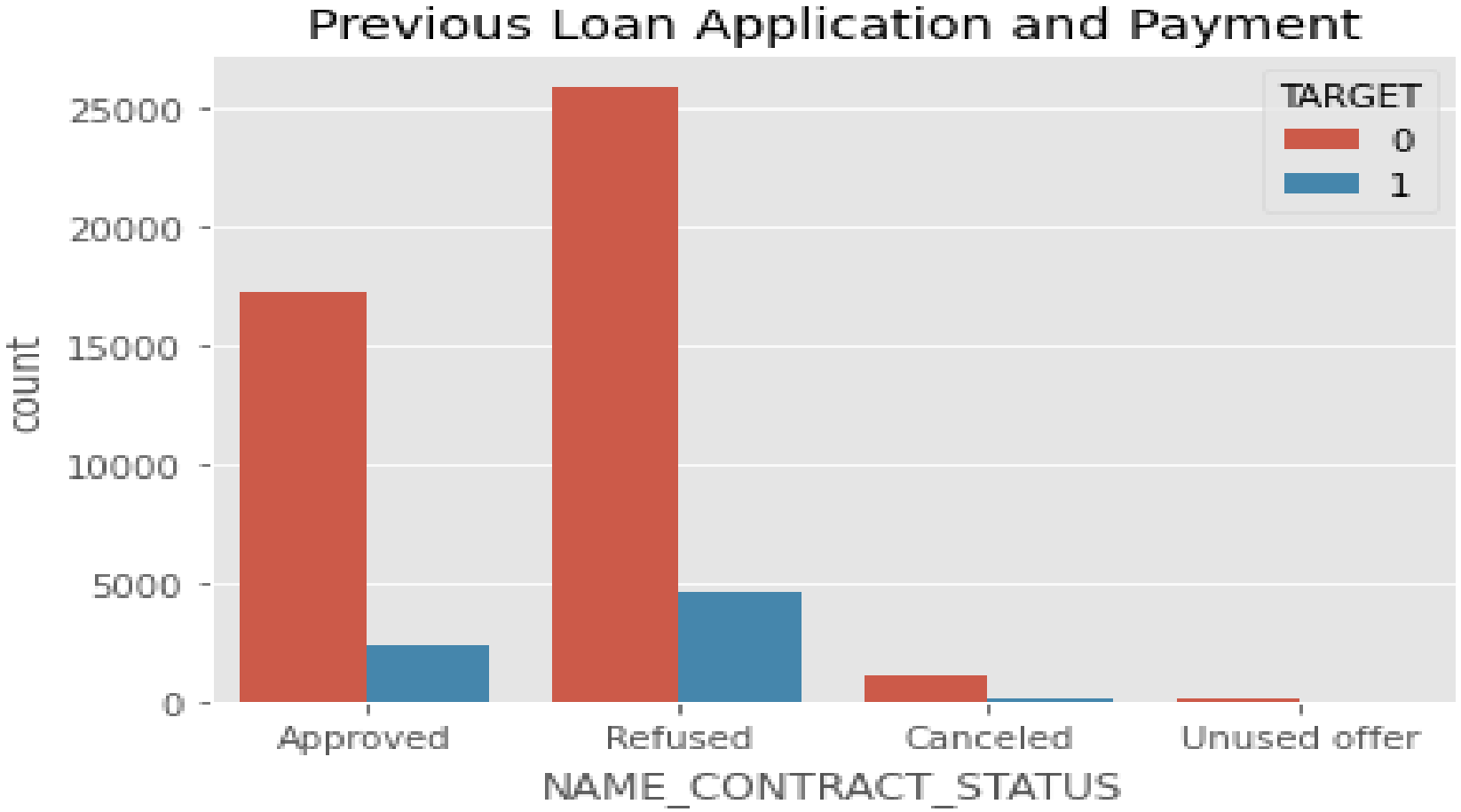
Correlation between numeric features of previous application data:

Inferences:

1. 'DAYS_LAST_DUE' and 'DAYS_TERMINATION' are highly correlated

2.'AMT_ANNUITY','AMT_APPLICATION','AMT_CREDIT','AMT_GOODS_PRICE ' are highly correlated.

Data Imbalance in Previous Application Data:



Inferences:

1. Than the applicants whose previous loans were rejected, the applicants whose previous loans were approved are more in number to pay current loan in time.

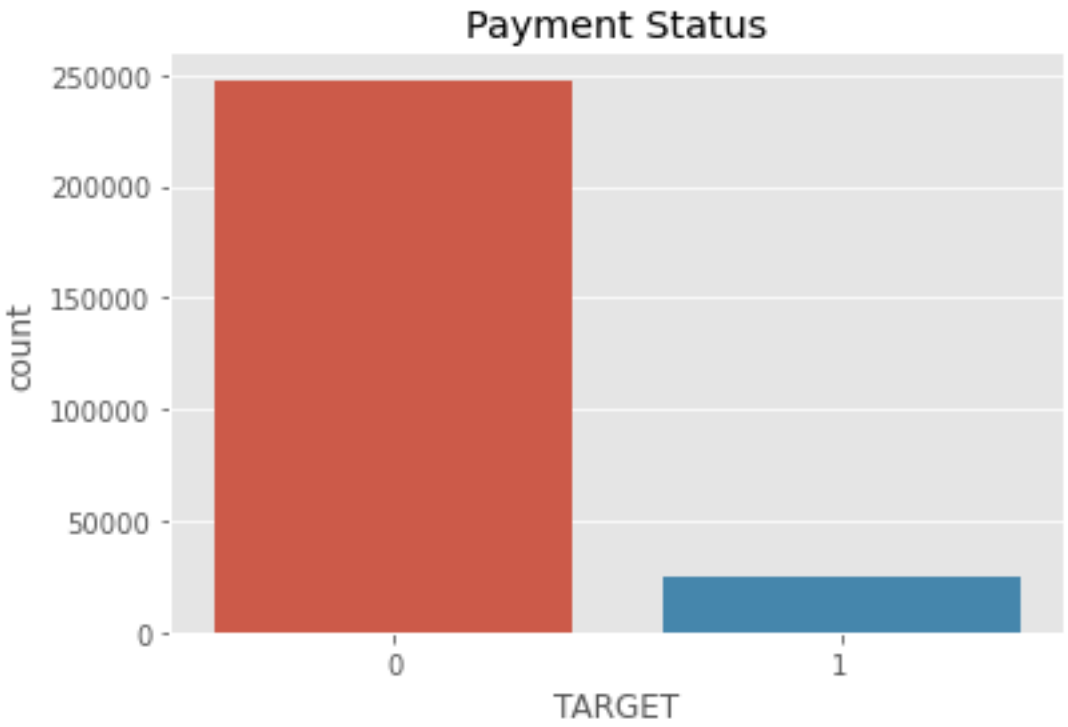
2. 12% of the previously approved loan applicants that defaulted in current loan.

3. 85% of the previously non approved loan applicants that are able to pay current loan.

Data imbalance in previous application Payment Status:

Inference:

Data is highly imbalanced since no. of defaulter is very less in total population

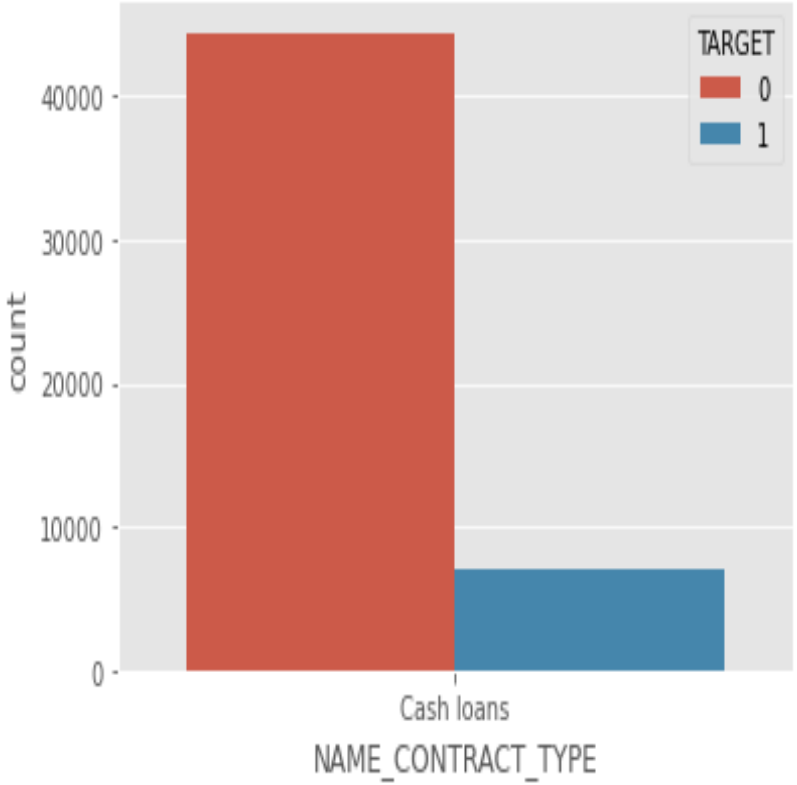


Univariant, Bivariant and Multivariant analysis:

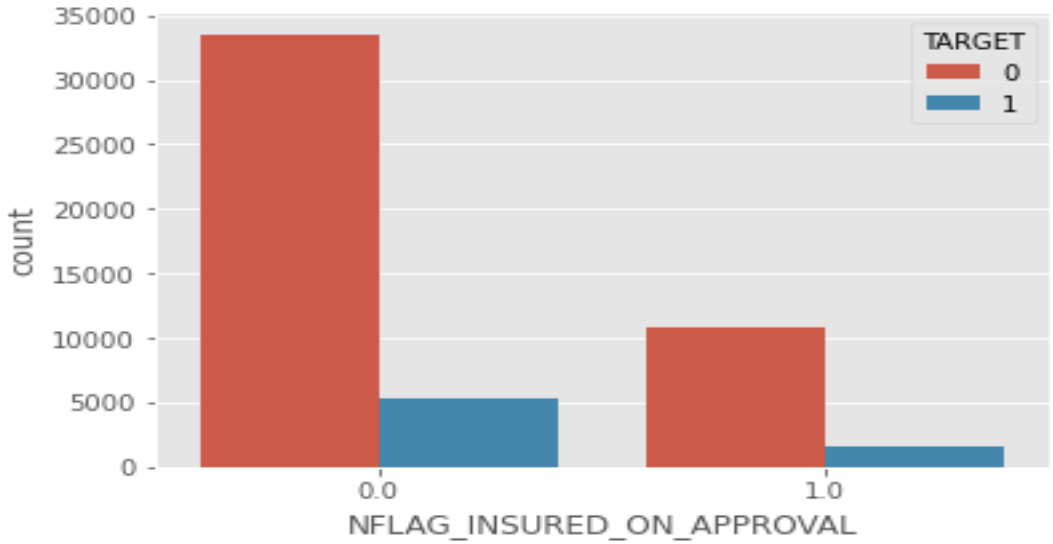
NAME_CONTRACT_TYPE:

Inference:

Highest number of loans are applied for Cash Loans by Non-Defaulters in previous application



'NFLAG_INSURED_ON_APPROVAL:

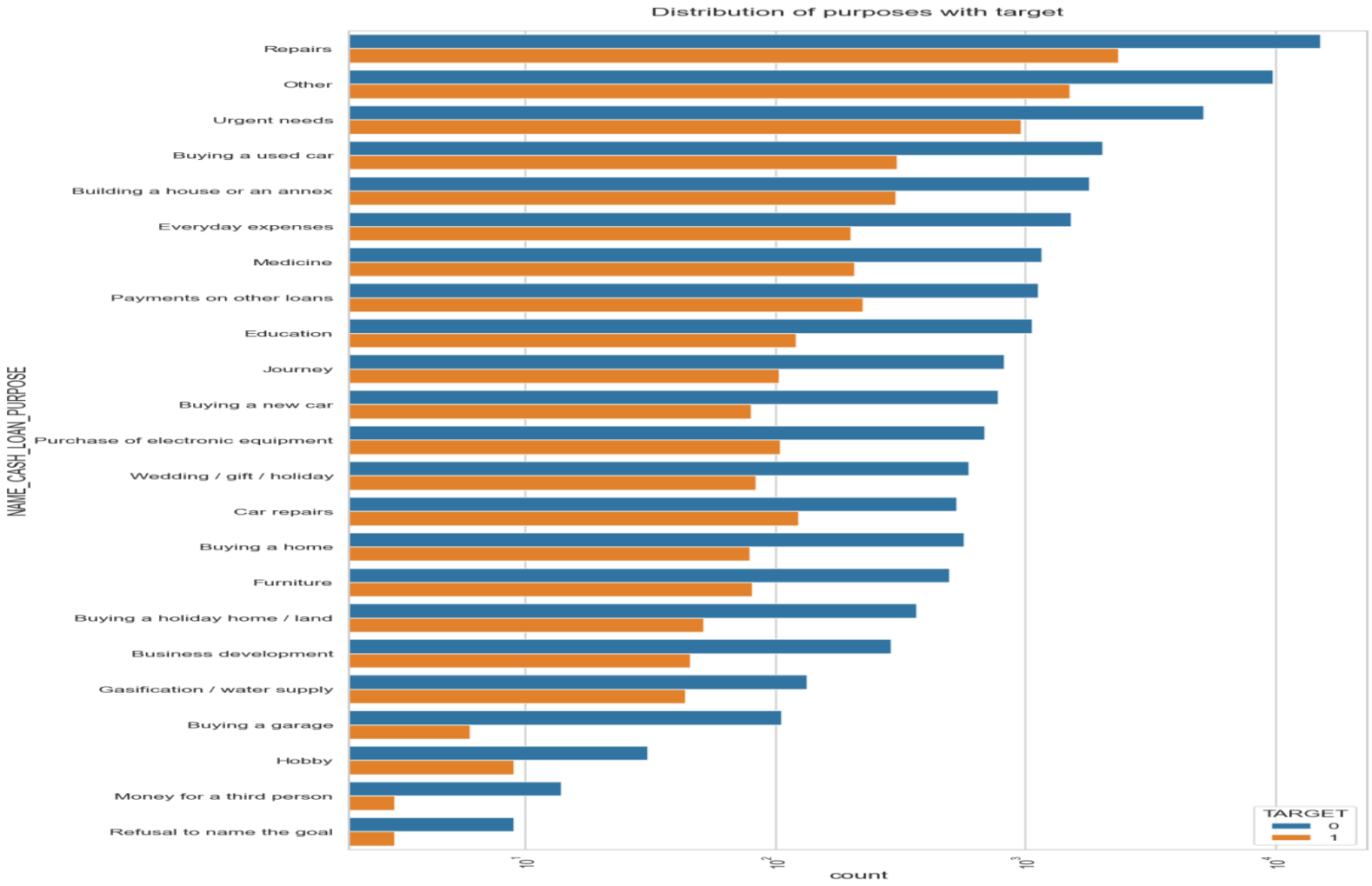


Inference:

Majority of the people didn't requested insurance during previous loan application.

Concatenating application Data and previous application Data by inner join:

Univariant analysis on “Distribution of contract status”



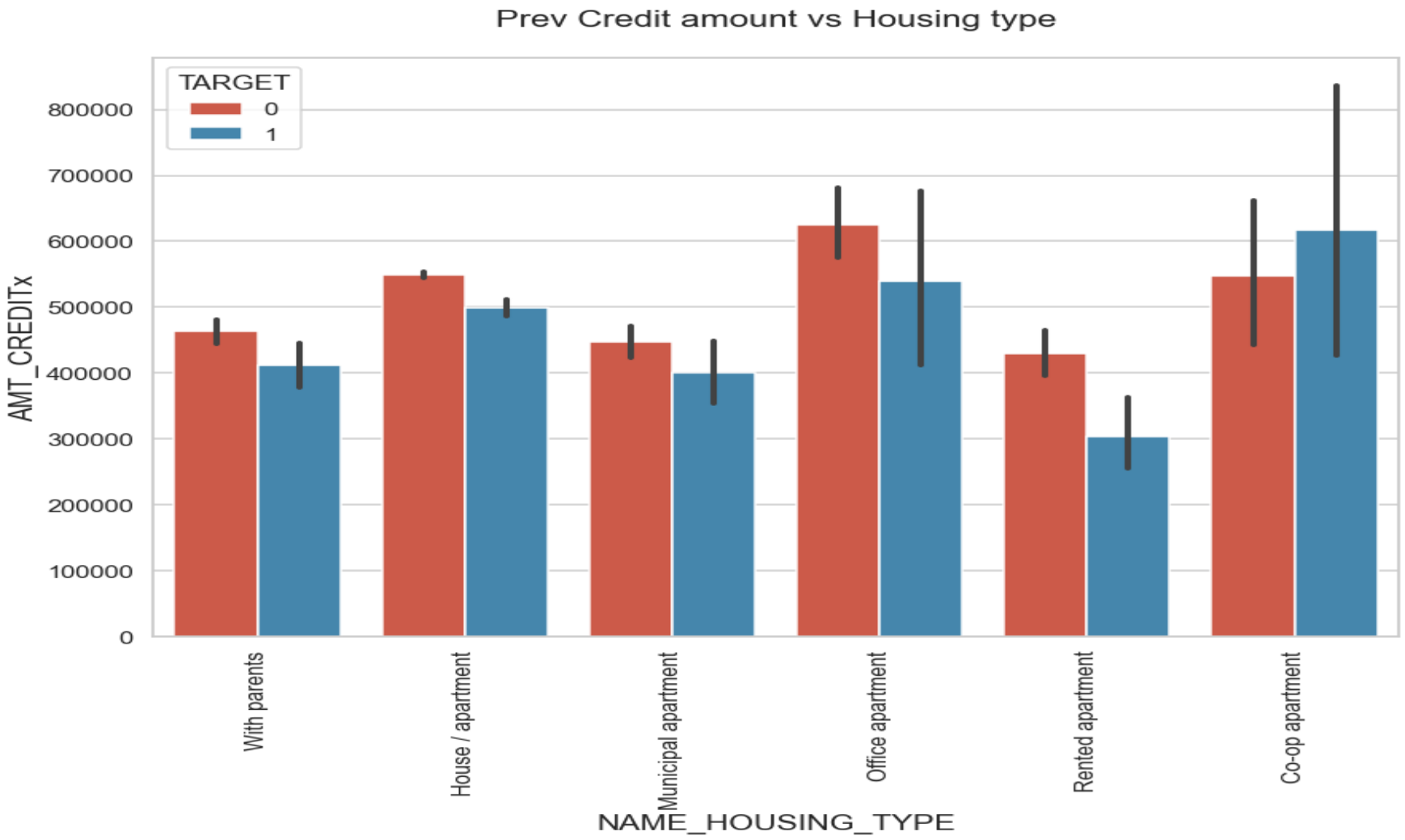
Inferences:

1.Loan purposes with 'Repairs' are facing more difficulties in payment on time.

2.There are few places where loan payment is significant higher than facing difficulties.

3.Clients are taking loan for 'Buying a garage', 'Business development', '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.

Bivariant Analysis:



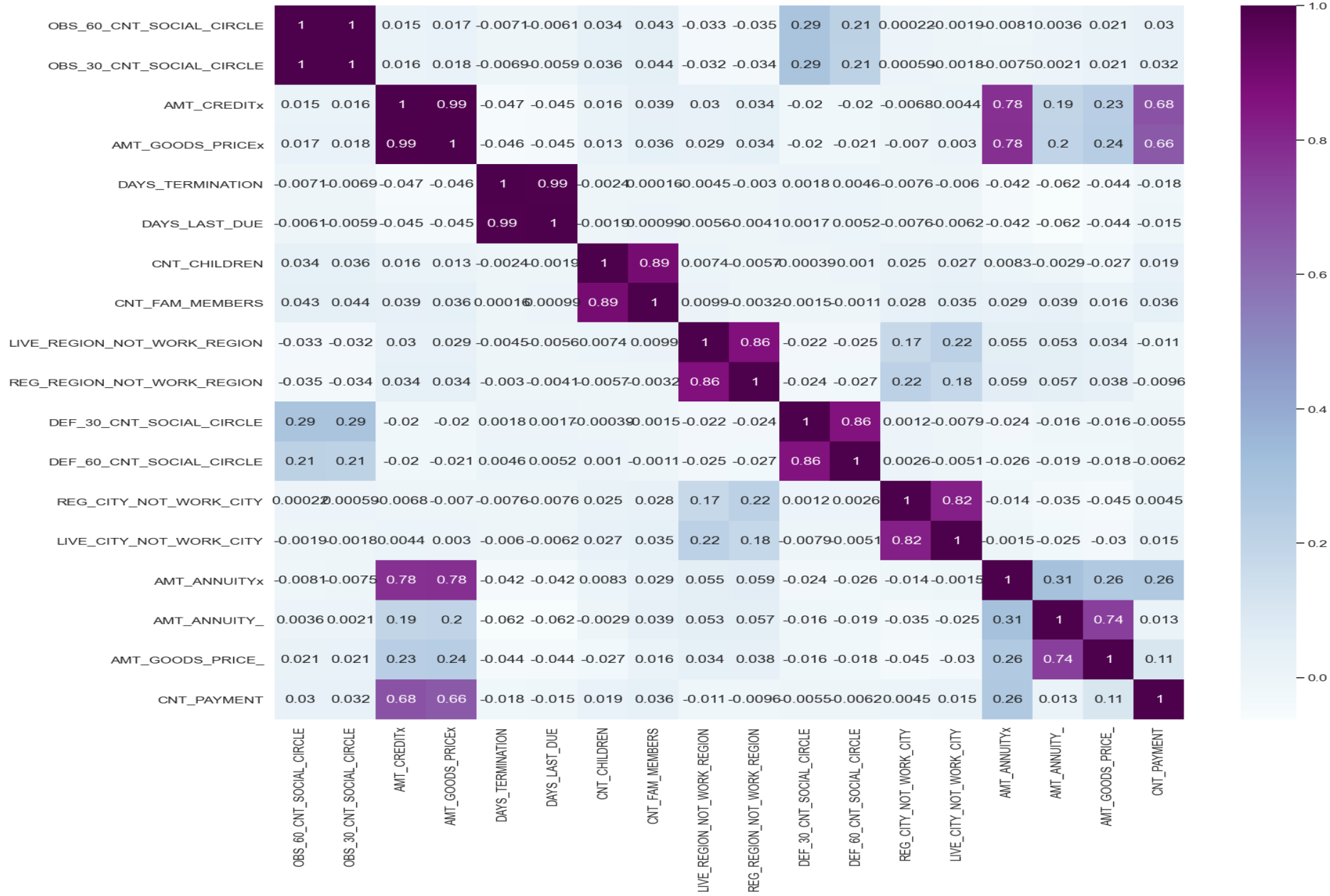
Inference:

1.Here for co-op apartment is having higher credit of target 1 and Housing type, office apartment is having higher credit of target 0. So, we can conclude that bank should take care or think before making decision of giving loans to the housing type of co-op apartment as they are having difficulties in payment.

2.Bank should focus more on housing type with parents or House\apartment or muncipal apartment for no payment difficulties by client.

Top 10 correlations are as follow (Repayer /Target=0)

1. OBS_60_CNT_SOCIAL_CIRCLE & OBS_30_CNT_SOCIAL_CIRCLE	0.99
2. AMT_CREDITx & AMT_GOODS_PRICEx	0.99
3. DAYS_TERMINATION & DAYS_LAST_DUE	0.98
4. CNT_CHILDREN & CNT_FAM_MEMBERS	0.89
5. LIVE_REGION_NOT_WORK_REGION & REG_REGION_NOT_WORK_REGION	0.86
6. DEF_30_CNT_SOCIAL_CIRCLE & DEF_60_CNT_SOCIAL_CIRCLE	0.85
7. REG_CITY_NOT_WORK_CITY & LIVE_CITY_NOT_WORK_CITY	0.82
8. AMT_GOODS_PRICEx & AMT_ANNUITYx	0.78
9. AMT_ANNUITY_ & AMT_GOODS_PRICE_	0.73
10. CNT_PAYMENT & AMT_CREDITx	0.67


OBS_60_CNT_SOCIAL_CIRCLE 1 1 0.015 0.017 -0.0071 -0.0061 0.034 0.043 -0.033 -0.035 0.29 0.21 0.00022 0.0019 -0.0081 0.0036 0.021 0.03
OBS_30_CNT_SOCIAL_CIRCLE 1 1 0.016 0.018 -0.0069 -0.0059 0.036 0.044 -0.032 -0.034 0.29 0.21 0.00059 0.0018 -0.0075 0.0021 0.021 0.032
AMT_CREDITx 0.015 0.016 1 0.99 -0.047 -0.045 0.016 0.039 0.03 0.034 -0.02 -0.02 -0.0068 0.0044 0.78 0.19 0.23 0.68
AMT_GOODS_PRICEx 0.017 0.018 0.99 1 -0.046 -0.045 0.013 0.036 0.029 0.034 -0.02 -0.021 -0.007 0.003 0.78 0.2 0.24 0.66
DAYS_TERMINATION -0.0071 -0.0069 -0.047 -0.046 1 0.99 -0.0024 0.00016 0.0045 -0.003 0.0018 0.0046 -0.0076 -0.006 -0.042 -0.062 -0.044 -0.018
DAYS_LAST_DUE -0.0061 -0.0059 -0.045 -0.045 0.99 1 -0.0019 0.00099 0.0056 -0.0041 0.0017 0.0052 -0.0076 -0.0062 -0.042 -0.062 -0.044 -0.015
CNT_CHILDREN 0.034 0.036 0.016 0.013 -0.0024 -0.0019 1 0.89 0.0074 -0.0057 0.00039 0.001 0.025 0.027 0.0083 -0.0029 -0.027 0.019
CNT_FAM_MEMBERS 0.043 0.044 0.039 0.036 0.00016 0.00099 0.89 1 0.0099 -0.0032 -0.0015 -0.0011 0.028 0.035 0.029 0.039 0.016 0.036
LIVE_REGION_NOT_WORK_REGION -0.033 -0.032 0.03 0.029 -0.0045 -0.0056 0.0074 0.0099 1 0.86 -0.022 -0.025 0.17 0.22 0.055 0.053 0.034 -0.011
REG_REGION_NOT_WORK_REGION -0.035 -0.034 0.034 0.034 -0.003 -0.0041 -0.0057 -0.0032 0.86 1 -0.024 -0.027 0.22 0.18 0.059 0.057 0.038 -0.0096
DEF_30_CNT_SOCIAL_CIRCLE 0.29 0.29 -0.02 -0.02 0.0018 0.0017 -0.00039 0.0015 -0.022 -0.024 1 0.86 0.0012 -0.0079 -0.024 -0.016 -0.016 -0.0055
DEF_60_CNT_SOCIAL_CIRCLE 0.21 0.21 -0.02 -0.021 0.0046 0.0052 0.001 -0.0011 -0.025 -0.027 0.86 1 0.0026 -0.0051 -0.026 -0.019 -0.018 -0.0062
REG_CITY_NOT_WORK_CITY 0.00022 0.00059 0.0068 -0.007 -0.0076 -0.0076 0.025 0.028 0.17 0.22 0.0012 0.0026 1 0.82 -0.014 -0.035 -0.045 0.0045
LIVE_CITY_NOT_WORK_CITY -0.0019 -0.0018 0.0044 0.003 -0.006 -0.0062 0.027 0.035 0.22 0.18 -0.0079 -0.0051 0.82 1 -0.0015 -0.025 -0.03 0.015
AMT_ANNUITYx -0.0081 -0.0075 0.78 0.78 -0.042 -0.042 0.0083 0.029 0.055 0.059 -0.024 -0.026 -0.014 -0.0015 1 0.31 0.26 0.26
AMT_ANNUITY_ 0.0036 0.0021 0.19 0.2 -0.062 -0.062 -0.0029 0.039 0.053 0.057 -0.016 -0.019 -0.035 -0.025 0.31 1 0.74 0.013
AMT_GOODS_PRICE_ 0.021 0.021 0.23 0.24 -0.044 -0.044 -0.027 0.016 0.034 0.038 -0.016 -0.018 -0.045 -0.03 0.26 0.74 1 0.11
CNT_PAYMENT 0.03 0.032 0.68 0.66 -0.018 -0.015 0.019 0.036 -0.011 -0.0096 -0.0055 -0.0062 0.0045 0.015 0.26 0.013 0.11 1
OBS_60_CNT_SOCIAL_CIRCLE
OBS_30_CNT_SOCIAL_CIRCLE
AMT_CREDITx
AMT_GOODS_PRICEx
DAYS_TERMINATION
DAYS_LAST_DUE
CNT_CHILDREN
CNT_FAM_MEMBERS
LIVE_REGION_NOT_WORK_REGION
REG_REGION_NOT_WORK_REGION
DEF_30_CNT_SOCIAL_CIRCLE
DEF_60_CNT_SOCIAL_CIRCLE
REG_CITY_NOT_WORK_CITY
LIVE_CITY_NOT_WORK_CITY
AMT_ANNUITYx
AMT_ANNUITY_
AMT_GOODS_PRICE_
CNT_PAYMENT
1.0
0.8
0.6
0.4
0.2
0.0

Inferences:

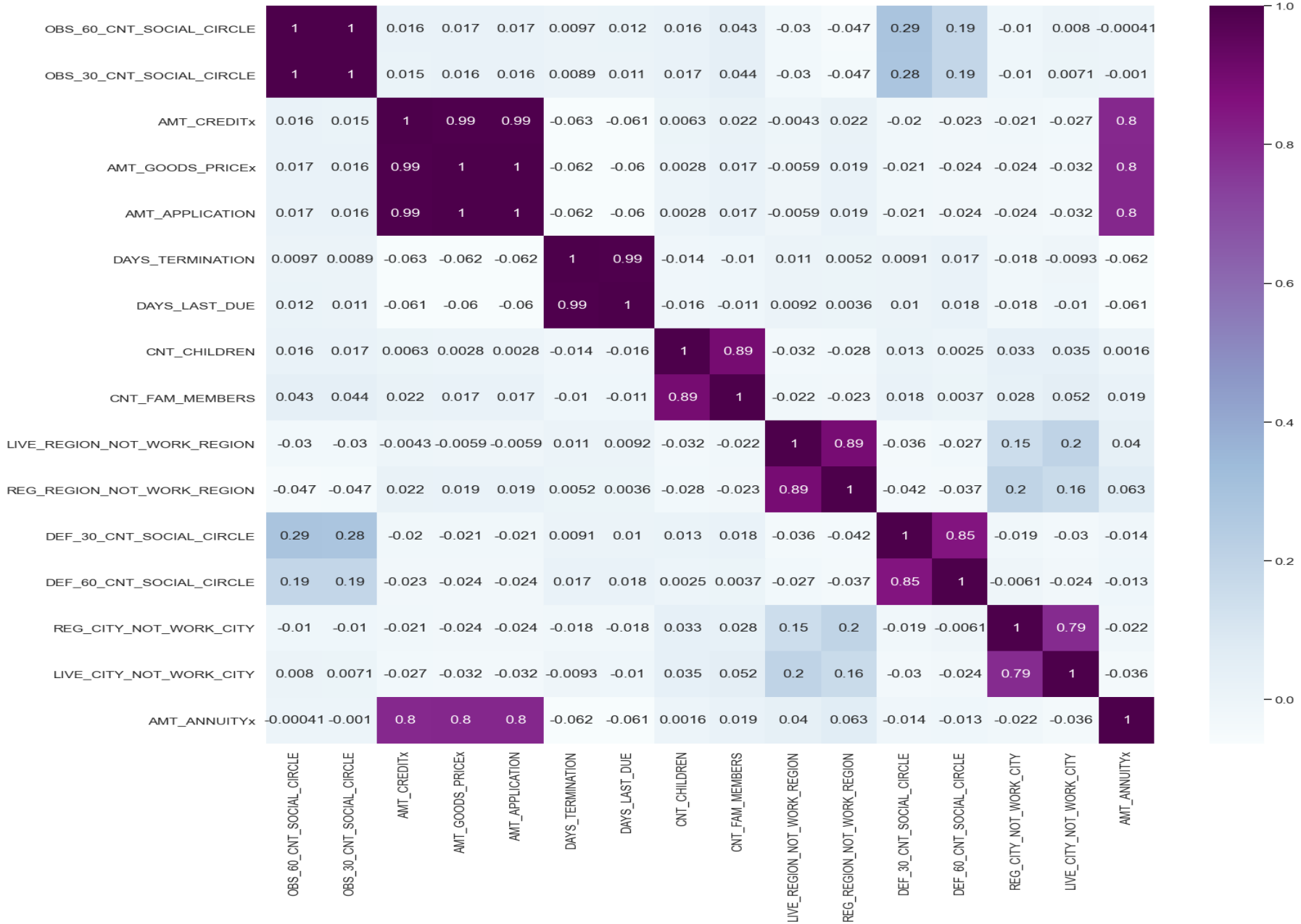
1. Higher the goods price, the credit by the client also will be high.

2. If the client's contact address does not match the address where he works , then there is more chances that the client's permanent address also does not match the address where he works

3.Client’s children and clients family members are highly correlated which means a client with children have high chances to have family members as well.

Top 10 correlations are as follow (Defaulters /Target=1)

1. OBS_60_CNT_SOCIAL_CIRCLE & OBS_30_CNT_SOCIAL_CIRCLE	1.00
2. AMT_CREDITx & AMT_GOODS_PRICEx	0.99
3. AMT_APPLICATION & AMT_CREDITx	0.99
4. DAYS_TERMINATION & DAYS_LAST_DUE	0.99
5. CNT_CHILDREN & CNT_FAM_MEMBERS	0.89
6. LIVE_REGION_NOT_WORK_REGION & REG_REGION_NOT_WORK_REGION	0.88
7. DEF_30_CNT_SOCIAL_CIRCLE & DEF_60_CNT_SOCIAL_CIRCLE	0.85
8. AMT_GOODS_PRICEx & AMT_ANNUITYx	0.80
9. AMT_CREDITx & AMT_ANNUITYx	0.79
10. REG_CITY_NOT_WORK_CITY & LIVE_CITY_NOT_WORK_CITY	0.78


OBS_60_CNT_SOCIAL_CIRCLE 1 1 0.016 0.017 0.017 0.0097 0.012 0.016 0.043 -0.03 -0.047 0.29 0.19 -0.01 0.008 -0.00041
OBS_30_CNT_SOCIAL_CIRCLE 1 1 0.015 0.016 0.016 0.0089 0.011 0.017 0.044 -0.03 -0.047 0.28 0.19 -0.01 0.0071 -0.001
AMT_CREDITx 0.016 0.015 1 0.99 0.99 -0.063 -0.061 0.0063 0.022 -0.0043 0.022 -0.02 -0.023 -0.021 -0.027 0.8
AMT_GOODS_PRICEx 0.017 0.016 0.99 1 1 -0.062 -0.06 0.0028 0.017 -0.0059 0.019 -0.021 -0.024 -0.024 -0.032 0.8
AMT_APPLICATION 0.017 0.016 0.99 1 1 -0.062 -0.06 0.0028 0.017 -0.0059 0.019 -0.021 -0.024 -0.024 -0.032 0.8
DAYS_TERMINATION 0.0097 0.0089 -0.063 -0.062 -0.062 1 0.99 -0.014 -0.01 0.011 0.0052 0.0091 0.017 -0.018 -0.0093 -0.062
DAYS_LAST_DUE 0.012 0.011 -0.061 -0.06 -0.06 0.99 1 -0.016 -0.011 0.0092 0.0036 0.01 0.018 -0.018 -0.01 -0.061
CNT_CHILDREN 0.016 0.017 0.0063 0.0028 0.0028 -0.014 -0.016 1 0.89 -0.032 -0.028 0.013 0.0025 0.033 0.035 0.0016
CNT_FAM_MEMBERS 0.043 0.044 0.022 0.017 0.017 -0.01 -0.011 0.89 1 -0.022 -0.023 0.018 0.0037 0.028 0.052 0.019
LIVE_REGION_NOT_WORK_REGION -0.03 -0.03 -0.0043 -0.0059 -0.0059 0.011 0.0092 -0.032 -0.022 1 0.89 -0.036 -0.027 0.15 0.2 0.04
REG_REGION_NOT_WORK_REGION -0.047 -0.047 0.022 0.019 0.019 0.0052 0.0036 -0.028 -0.023 0.89 1 -0.042 -0.037 0.2 0.16 0.063
DEF_30_CNT_SOCIAL_CIRCLE 0.29 0.28 -0.02 -0.021 -0.021 0.0091 0.01 0.013 0.018 -0.036 -0.042 1 0.85 -0.019 -0.03 -0.014
DEF_60_CNT_SOCIAL_CIRCLE 0.19 0.19 -0.023 -0.024 -0.024 0.017 0.018 0.0025 0.0037 -0.027 -0.037 0.85 1 -0.0061 -0.024 -0.013
REG_CITY_NOT_WORK_CITY -0.01 -0.01 -0.021 -0.024 -0.024 -0.018 -0.018 0.033 0.028 0.15 0.2 -0.019 -0.0061 1 0.79 -0.022
LIVE_CITY_NOT_WORK_CITY 0.008 0.0071 -0.027 -0.032 -0.032 -0.0093 -0.01 0.035 0.052 0.2 0.16 -0.03 -0.024 0.79 1 -0.036
AMT_ANNUITYx -0.00041 -0.001 0.8 0.8 0.8 -0.062 -0.061 0.0016 0.019 0.04 0.063 -0.014 -0.013 -0.022 -0.036 1
OBS_60_CNT_SOCIAL_CIRCLE
OBS_30_CNT_SOCIAL_CIRCLE
AMT_CREDITx
AMT_GOODS_PRICEx
AMT_APPLICATION
DAYS_TERMINATION
DAYS_LAST_DUE
CNT_CHILDREN
CNT_FAM_MEMBERS
LIVE_REGION_NOT_WORK_REGION
REG_REGION_NOT_WORK_REGION
DEF_30_CNT_SOCIAL_CIRCLE
DEF_60_CNT_SOCIAL_CIRCLE
REG_CITY_NOT_WORK_CITY
LIVE_CITY_NOT_WORK_CITY
AMT_ANNUITYx
1.0
0.8
0.6
0.4
0.2
0.0

Inferences:

1. AMT_GOODS_PRICE and AMT_APPLICATION have a high correlation here as well, it means that more credit the client asked for previously is proportional to the goods price that the client asked for previously.
2. Higher the goods price, the credit by the client also will be high.
3. AMT_ANNUITY and AMT_APPLICATION have a high correlation, which means the higher the loan annuity issued, the higher the goods price that the client asked for previously.
4. If the client's contact address does not match the work address, then there's a high chance that the client's permanent address also does not match the work address.
5. Clients children and clients family members are highly correlated which means a client with children have high chances to have family members as well.

Conclusion:

1. State servant, Working and commercial associate are having the number of credits are higher than other.
2. Females are having more number of credits than male.
3. No income type for ‘student’ , ’pensioner’ and ‘Businessman’ this indicates that they don’t do any late payments or they do not have payment difficulties.
4. Loan purposes with 'Repairs' are facing more difficulties in payment on time.
5. Bank should focus more on housing type with parents or House\apartment or municipal apartment for no payment difficulties by client.
6. Bank should take care or think before making decision of giving loans to the housing type of co-op apartment as they are having difficulties in payment.

Thank You