CREDIT EDA CASE STUDY

BY

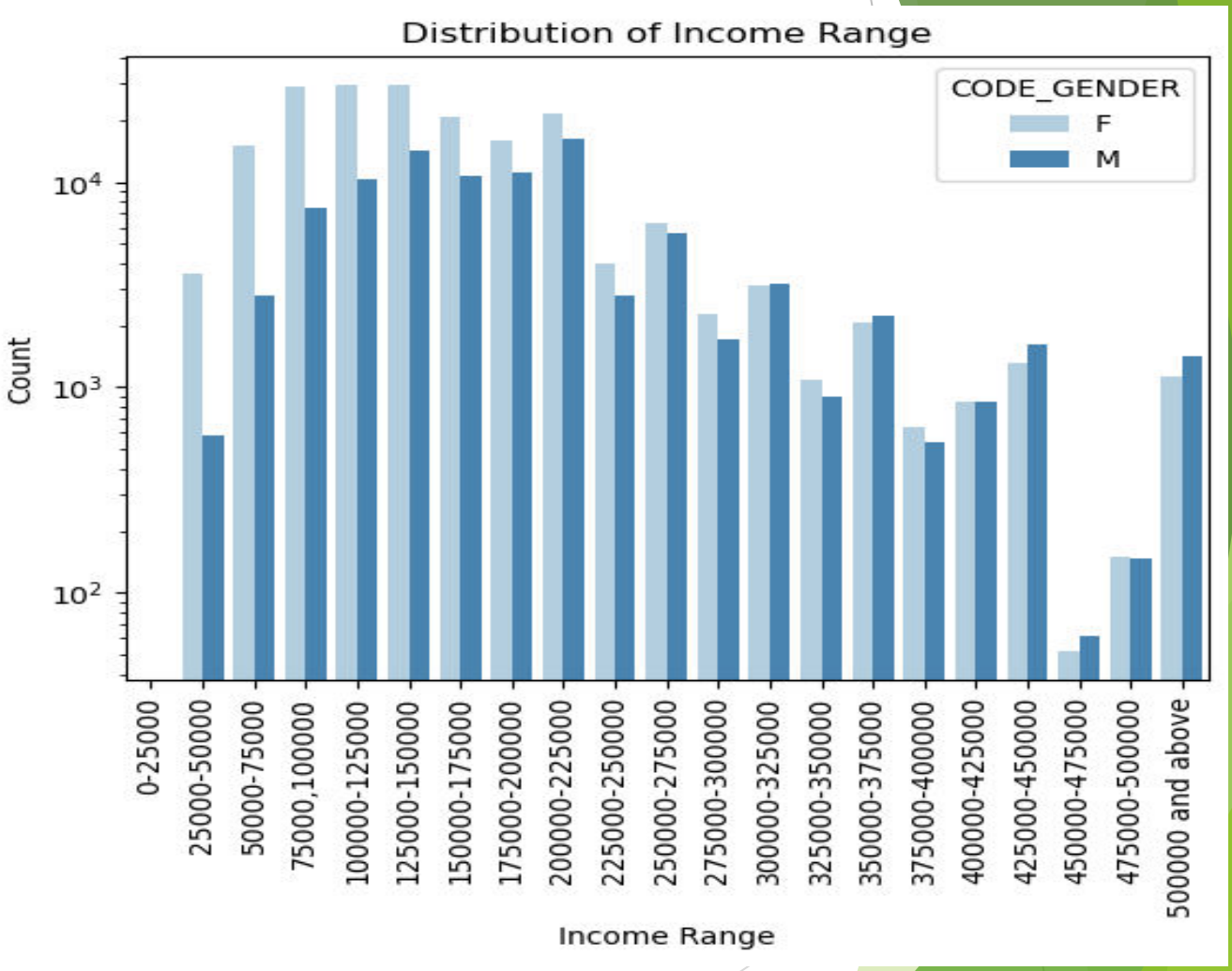
POOJA R SHETTY

CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET 0

Distribution of Income Range

Points need to be concluded from the graph:

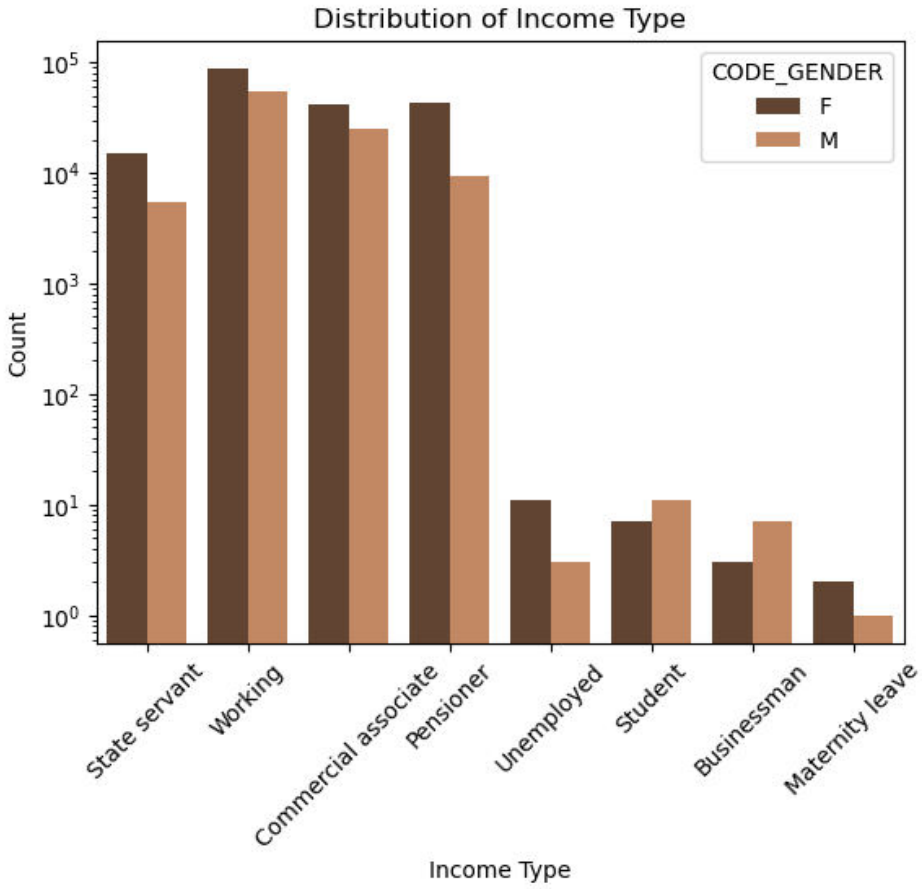
1. Female counts are higher than male.
2. Income range from 100000 to 200000 is having more number od credits.
3. This graph show that females are more than male in having credits for that range.
4. Very less count for income range 400000 and above.



Distribution of Income Type

Points need to be concluded from the graph:

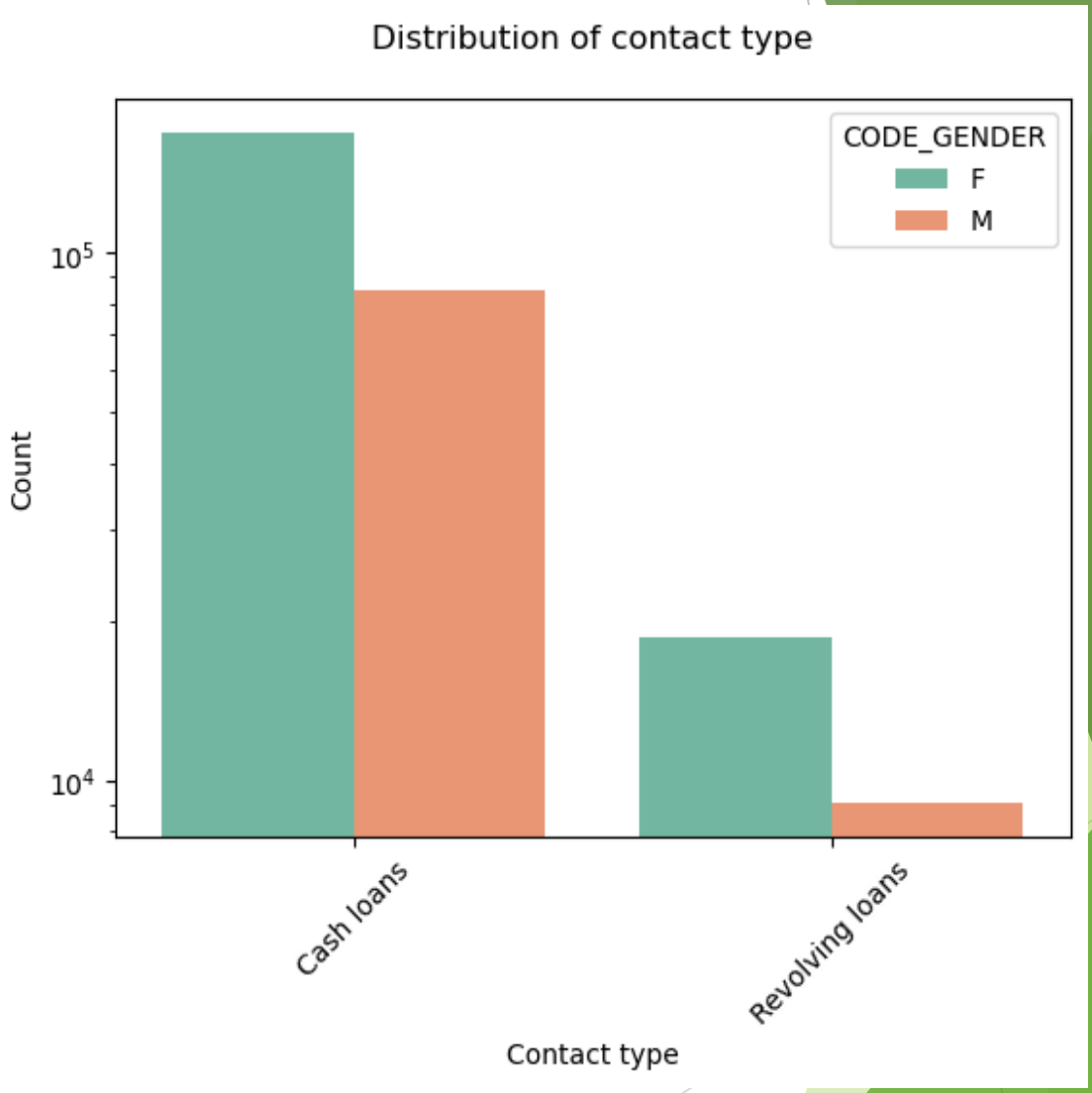
1. For income type “working”, “Commercial associate”, “pensioner” and “State servant” the number of credits are higher than others.
2. For this females are having more number of credits than male.
3. Less number of credits for income type “student”, “Businessman”, and “Maternity leave”.



Distribution for contract type

Points need to be concluded from the graph:

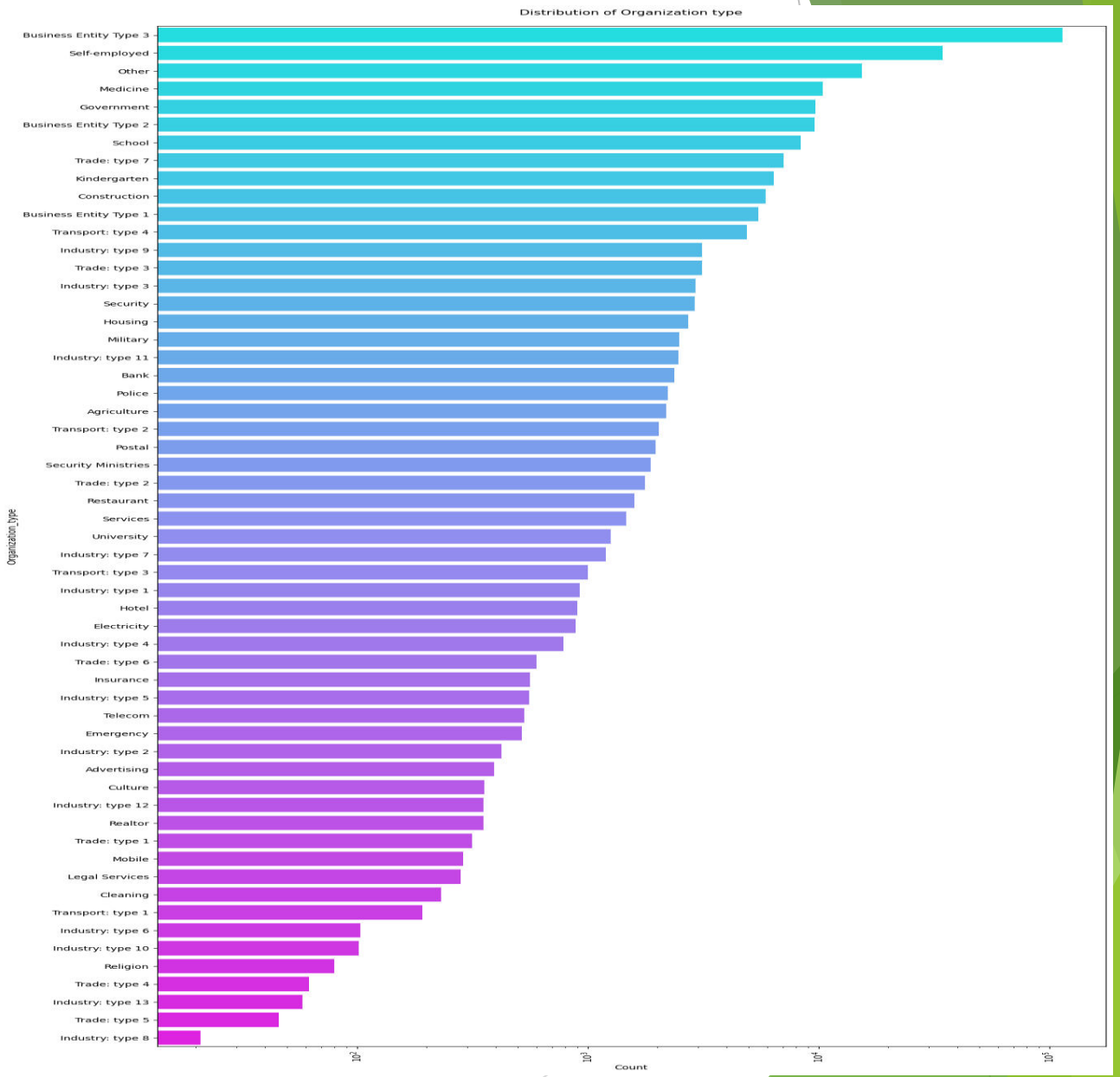
1. For the contract type “cash loans” is having number of credits than “ Revolving loans” contract type.
2. For this also Female is leading for applying credits.



Distribution for Organization Type

Points to be concluded from the graph .

1. Clients which have applied for credits are from most of the organization type “Business entity type3”, “Self employed”, “Other”, “Medicine” and “Government”.
2. Less clients are from Industry type 8, type6, type 10, religion and trade 5, type 4.

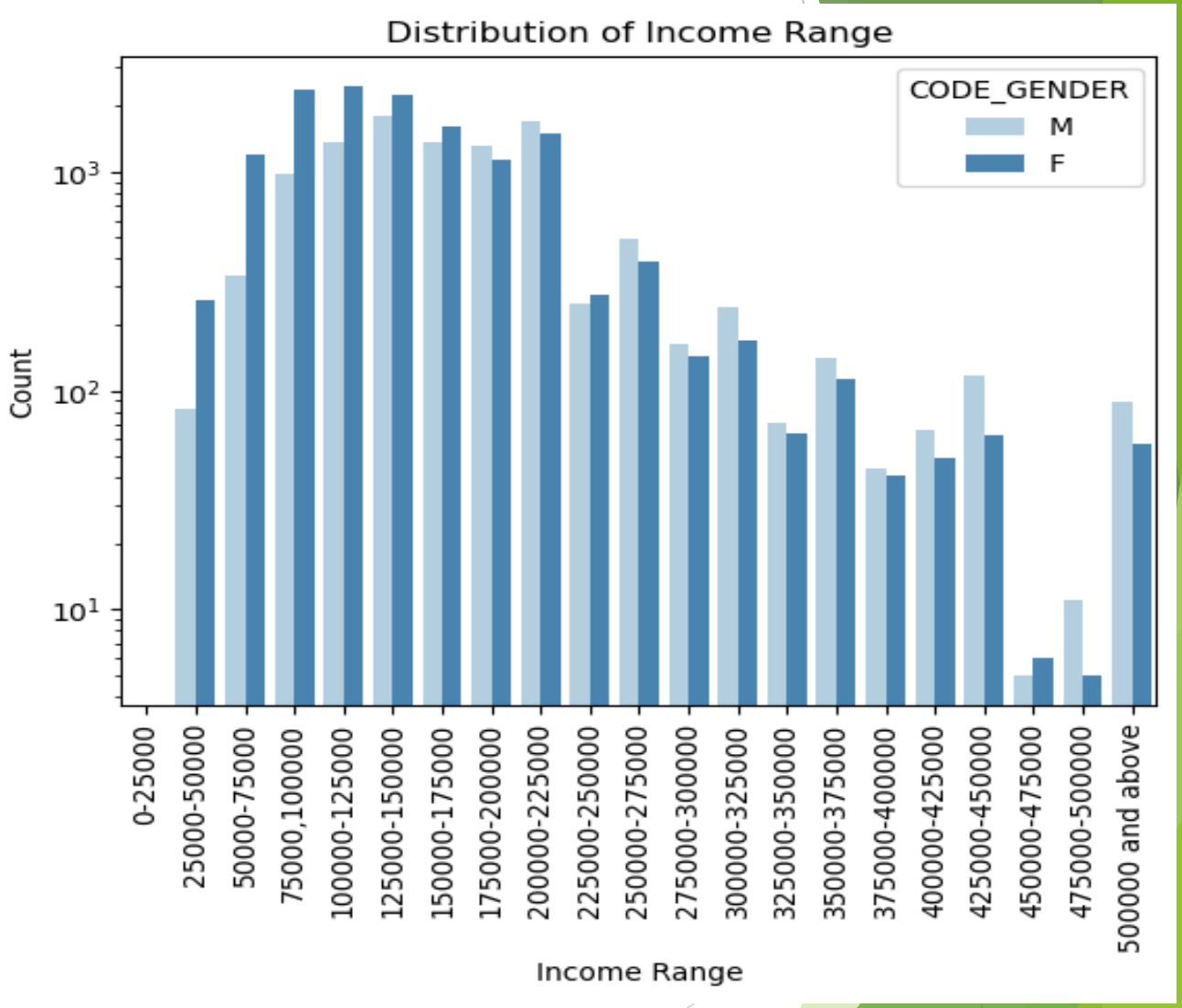


CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET 1

DISTRIBUTION OF INCOME RANGE

Points to be concluded from the graph .

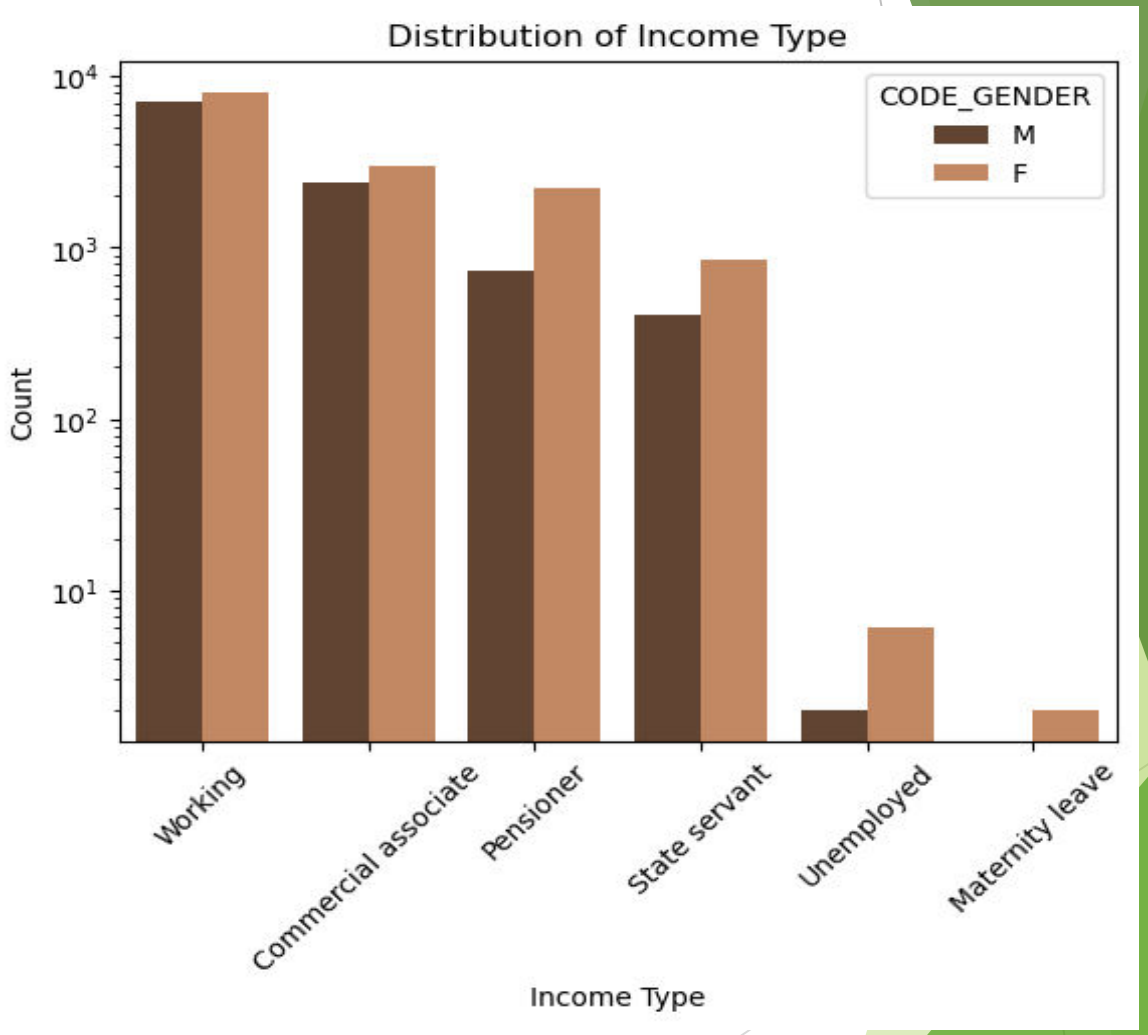
1. Male counts are higher than female.
2. Income range from 100000 to 200000 is having more number of credits.
3. This graph show that males are more than females in having credits for the range.
4. Very less count for income range 400000 and above.



Distribution of Income Type

Points to be concluded from the graph.

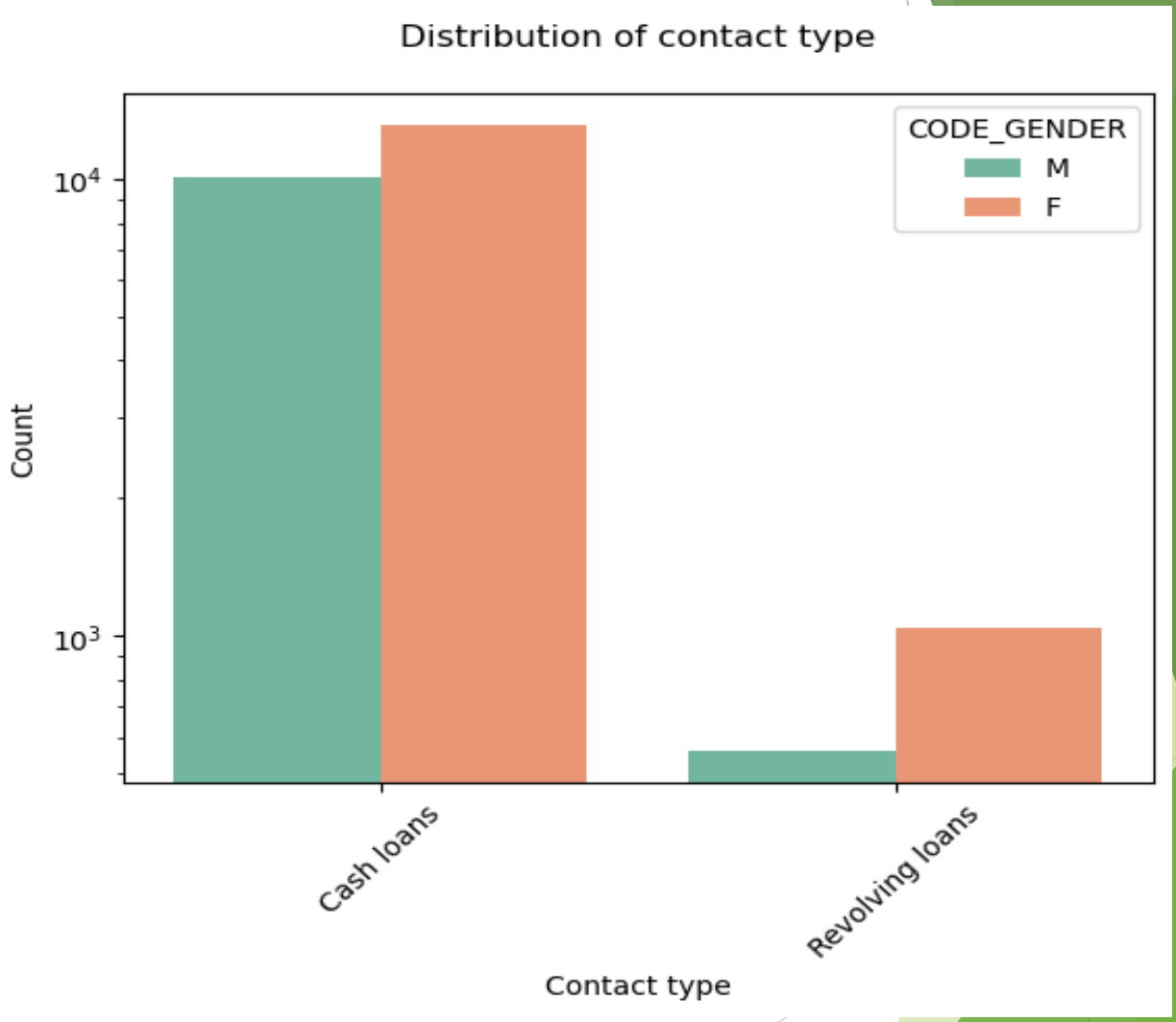
1. For income type “Working”, “Commercial associate”, and “state servant” the number of credits are higher than other i. e “Maternity leave”.
2. For this females are having more number of credits than male.
3. Less number of credits for income type “Maternity leave”.
4. For type1 there is no income type for “unemployed “ and “maternity leave”.



Distribution for contract type

Points to be considered.

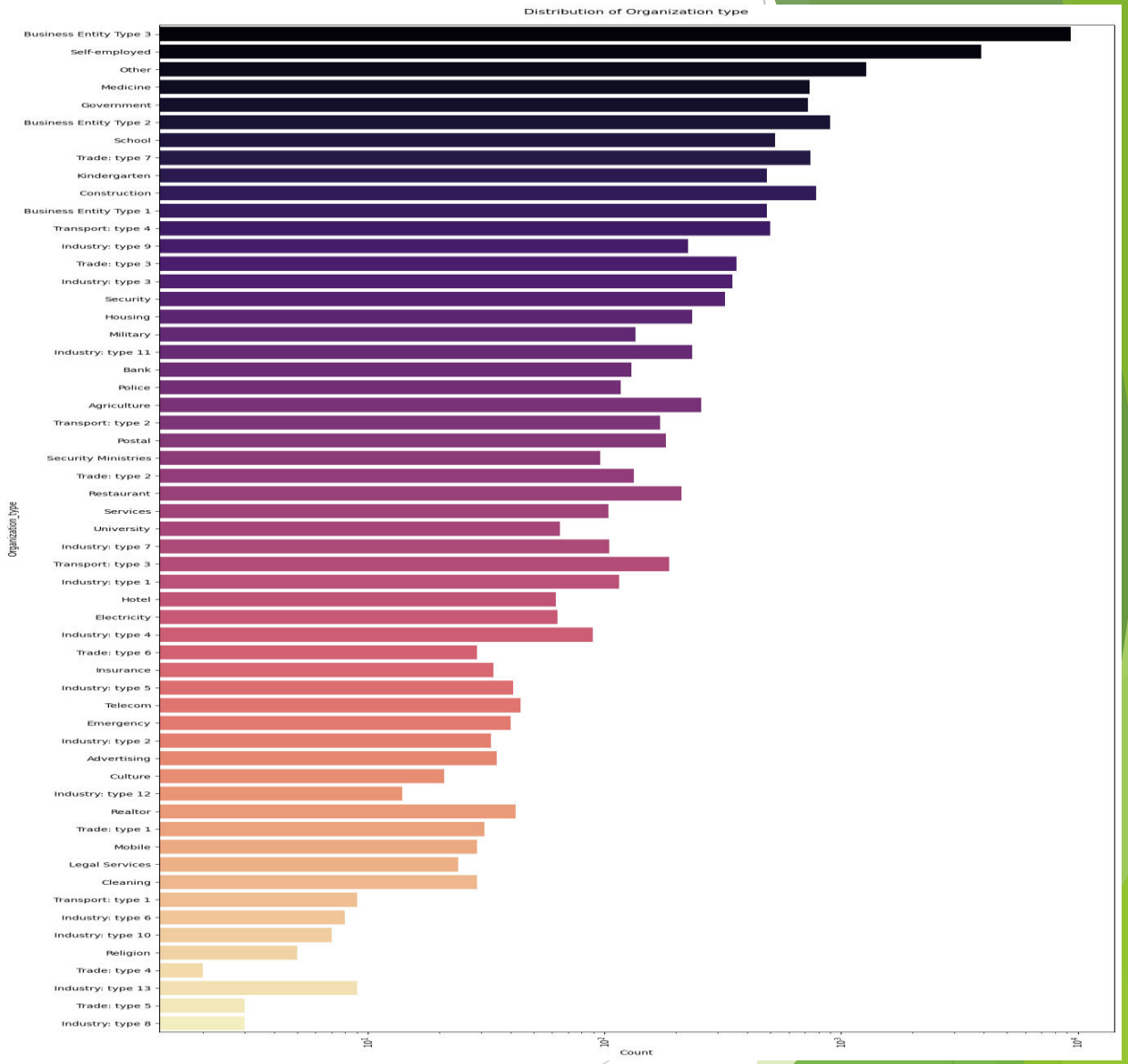
1. For contract type “Cash loans” is having higher number of credits than “Revolving loans” contract type.
2. For this also Female is leading for applying credits.
3. For type1: there is only Female revolving loans.



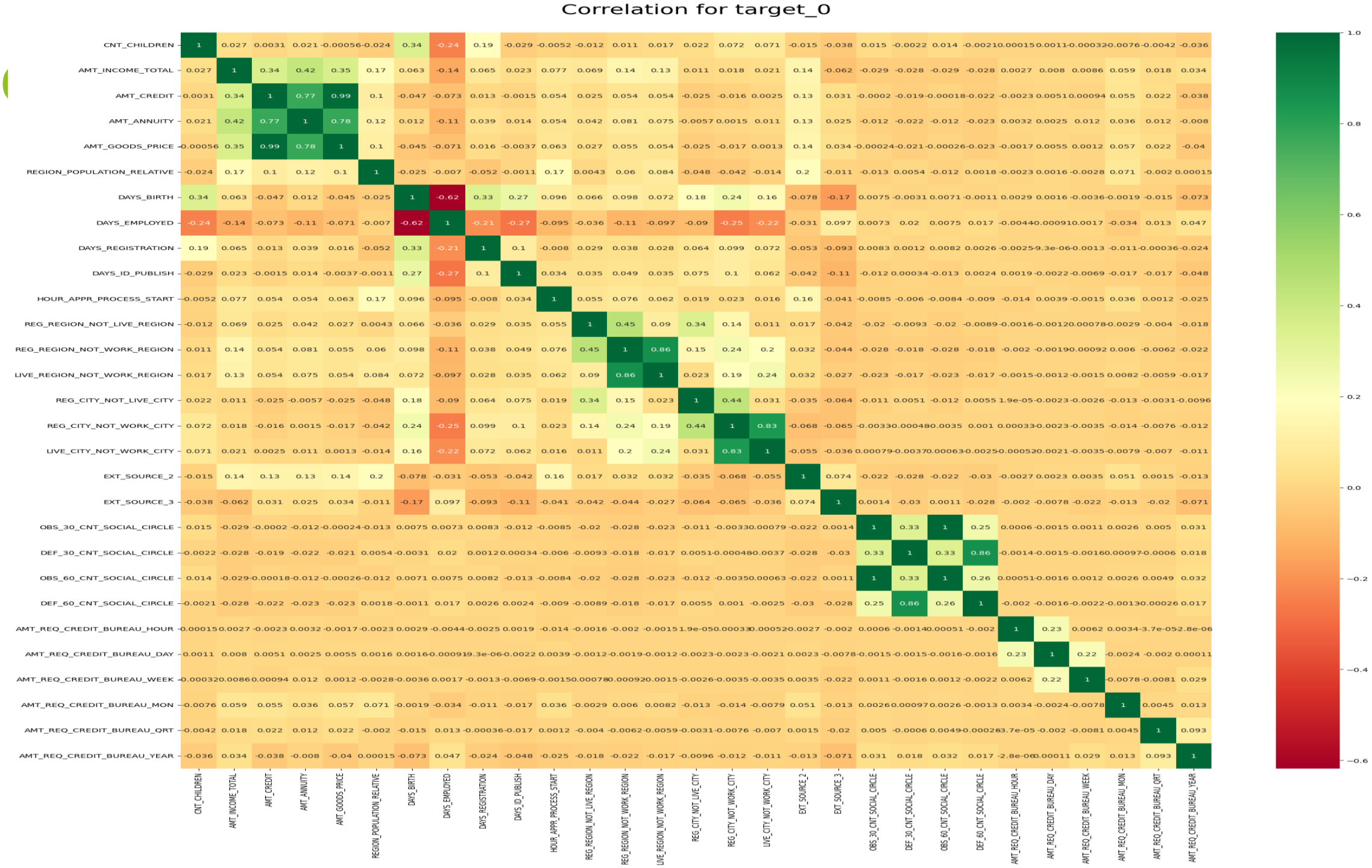
Distribution of organization type

Points to be concluded.

1. Clients which applied for credits are from most of the organization type “Business entity type 3”, “Self employed”, “other”, “medicine” and “Government”.
2. Less clients are from industry type 8, type 6, type 10, religion and trade type 5, type 4.
3. Same as type 0 in distribution of organization type.



CORRELATION OF TARGET 0


Correlation for target_0
CNT_CHILDREN 1 0.027 0.0031 0.021 -0.00056 -0.024 0.34 -0.24 0.19 -0.029 -0.0052 -0.012 0.011 0.017 0.022 0.072 0.071 -0.015 -0.038 0.015 -0.0022 0.014 -0.0021 0.00015 0.0011 -0.00032 0.0076 -0.0042 -0.036
AMT_INCOME_TOTAL 0.027 1 0.34 0.42 0.35 0.17 0.063 -0.14 0.065 0.023 0.077 0.069 0.14 0.13 0.011 0.018 0.021 0.14 -0.062 -0.029 -0.028 -0.029 -0.028 0.0027 0.008 0.0086 0.059 0.018 0.034
AMT_CREDIT 0.0031 0.34 1 0.77 0.99 0.1 -0.047 -0.073 0.013 -0.0015 0.054 0.025 0.054 0.054 -0.025 -0.016 0.0025 0.13 0.031 -0.0002 -0.019 -0.00018 -0.022 -0.0023 0.0051 0.00094 0.055 0.022 -0.038
AMT_ANNUITY 0.021 0.42 0.77 1 0.78 0.12 0.012 -0.11 0.039 0.014 0.054 0.042 0.081 0.075 -0.0057 0.0015 0.011 0.13 0.025 -0.012 -0.022 -0.012 -0.023 0.0032 0.0025 0.012 0.036 0.012 -0.008
AMT_GOODS_PRICE -0.00056 0.35 0.99 0.78 1 0.1 -0.045 -0.071 0.016 -0.0037 0.063 0.027 0.055 0.054 -0.025 -0.017 0.0013 0.14 0.034 -0.00024 -0.021 -0.00026 -0.023 -0.0017 0.0055 0.0012 0.057 0.022 -0.04
REGION_POPULATION_RELATIVE -0.024 0.17 0.1 0.12 0.1 1 -0.025 -0.007 -0.052 -0.0011 0.17 0.0043 0.06 0.084 -0.048 -0.042 -0.014 0.2 -0.011 -0.013 0.0054 -0.012 0.0018 -0.0023 0.0016 -0.0028 0.071 -0.002 0.00015
DAYS_BIRTH 0.34 0.063 -0.047 0.012 -0.045 -0.025 1 -0.62 0.33 0.27 0.096 0.066 0.098 0.072 0.18 0.24 0.16 -0.078 -0.17 0.0075 -0.0031 0.0071 -0.0011 0.0029 0.0016 -0.0036 -0.0019 -0.015 -0.073
DAYS_EMPLOYED -0.24 -0.14 -0.073 -0.11 -0.071 -0.007 -0.62 1 -0.21 -0.27 -0.095 -0.036 -0.11 -0.097 -0.09 -0.25 -0.22 -0.031 0.097 0.0073 0.02 0.0075 0.017 -0.0044 0.00091 0.0017 -0.034 0.013 0.047
DAYS_REGISTRATION 0.19 0.065 0.013 0.039 0.016 -0.052 0.33 -0.21 1 0.1 -0.008 0.029 0.038 0.028 0.064 0.099 0.072 -0.053 -0.093 0.0083 0.0012 0.0082 0.0026 -0.0025 -9.3e-06 -0.0013 -0.011 -0.00036 -0.024
DAYS_ID_PUBLISH -0.029 0.023 -0.0015 0.014 -0.0037 -0.0011 0.27 -0.27 0.1 1 0.034 0.035 0.049 0.035 0.075 0.1 0.062 -0.042 -0.11 -0.012 0.00034 -0.013 0.0024 0.0019 -0.0022 -0.0069 -0.017 -0.017 -0.048
HOUR_APPR_PROCESS_START -0.0052 0.077 0.054 0.054 0.063 0.17 0.096 -0.095 -0.008 0.034 1 0.055 0.076 0.062 0.019 0.023 0.016 0.16 -0.041 -0.0085 -0.006 -0.0084 -0.009 -0.014 0.0039 -0.0015 0.036 0.0012 -0.025
REG_REGION_NOT_LIVE_REGION -0.012 0.069 0.025 0.042 0.027 0.0043 0.066 -0.036 0.029 0.035 0.055 1 0.45 0.09 0.34 0.14 0.011 0.017 -0.042 -0.02 -0.0093 -0.02 -0.0089 -0.0016 -0.0012 0.00078 -0.0029 -0.004 -0.018
REG_REGION_NOT_WORK_REGION 0.011 0.14 0.054 0.081 0.055 0.06 0.098 -0.11 0.038 0.049 0.076 0.45 1 0.86 0.15 0.24 0.2 0.032 -0.044 -0.028 -0.018 -0.028 -0.018 -0.002 -0.0019 0.00092 0.006 -0.0062 -0.022
LIVE_REGION_NOT_WORK_REGION 0.017 0.13 0.054 0.075 0.054 0.084 0.072 -0.097 0.028 0.035 0.062 0.09 0.86 1 0.023 0.19 0.24 0.032 -0.027 -0.023 -0.017 -0.023 -0.017 -0.0015 -0.0012 -0.0015 0.0082 -0.0059 -0.017
REG_CITY_NOT_LIVE_CITY 0.022 0.011 -0.025 -0.0057 -0.025 -0.048 0.18 -0.09 0.064 0.075 0.019 0.34 0.15 0.023 1 0.44 0.031 -0.035 -0.064 -0.011 0.0051 -0.012 0.0055 1.9e-05 -0.0023 -0.0026 -0.013 -0.0031 -0.0096
REG_CITY_NOT_WORK_CITY 0.072 0.018 -0.016 0.0015 -0.017 -0.042 0.24 -0.25 0.099 0.1 0.023 0.14 0.24 0.19 0.44 1 0.83 -0.068 -0.065 -0.0033 0.00048 0.0035 0.001 0.00033 -0.0023 -0.0035 -0.014 -0.0076 -0.012
LIVE_CITY_NOT_WORK_CITY 0.071 0.021 0.0025 0.011 0.0013 -0.014 0.16 -0.22 0.072 0.062 0.016 0.011 0.2 0.24 0.031 0.83 1 -0.055 -0.036 0.00079 -0.0037 0.00063 -0.0025 -0.00052 0.0021 -0.0035 -0.0079 -0.007 -0.011
EXT_SOURCE_2 -0.015 0.14 0.13 0.13 0.14 0.2 -0.078 -0.031 -0.053 -0.042 0.16 0.017 0.032 0.032 -0.035 -0.068 -0.055 1 0.074 -0.022 -0.028 -0.022 -0.03 -0.0027 0.0023 0.0035 0.051 0.0015 -0.013
EXT_SOURCE_3 -0.038 -0.062 0.031 0.025 0.034 -0.011 -0.17 0.097 -0.093 -0.11 -0.041 -0.042 -0.044 -0.027 -0.064 -0.065 -0.036 0.074 1 0.0014 -0.03 0.0011 -0.028 -0.002 -0.0078 -0.022 -0.013 -0.02 -0.071
OBS_30_CNT_SOCIAL_CIRCLE 0.015 -0.029 -0.0002 -0.012 -0.00024 -0.013 0.0075 0.0073 0.0083 -0.012 -0.0085 -0.02 -0.028 -0.023 -0.011 -0.0033 0.00079 -0.022 0.0014 1 0.33 1 0.25 0.0006 -0.0015 0.0011 0.0026 0.005 0.031
DEF_30_CNT_SOCIAL_CIRCLE -0.0022 -0.028 -0.019 -0.022 -0.021 0.0054 -0.0031 0.02 0.0012 0.00034 -0.006 -0.0093 -0.018 -0.017 0.0051 -0.00048 0.0037 -0.028 -0.03 0.33 1 0.33 0.86 -0.0014 -0.0015 -0.0016 0.00097 -0.0006 0.018
OBS_60_CNT_SOCIAL_CIRCLE 0.014 -0.029 -0.00018 -0.012 -0.00026 -0.012 0.0071 0.0075 0.0082 -0.013 -0.0084 -0.02 -0.028 -0.023 -0.012 -0.0035 0.00063 -0.022 0.0011 1 0.33 1 0.26 0.00051 -0.0016 0.0012 0.0026 0.0049 0.032
DEF_60_CNT_SOCIAL_CIRCLE -0.0021 -0.028 -0.022 -0.023 -0.023 0.0018 -0.0011 0.017 0.0026 0.0024 -0.009 -0.0089 -0.018 -0.017 0.0055 0.001 -0.0025 -0.03 -0.028 0.25 0.86 0.26 1 -0.002 -0.0016 -0.0022 -0.0013 0.00026 0.017
AMT_REQ_CREDIT_BUREAU_HOUR -0.00015 0.0027 -0.0023 0.0032 -0.0017 -0.0023 0.0029 -0.0044 -0.0025 0.0019 -0.014 -0.0016 -0.002 -0.0015 1.9e-05 0.00033 0.00052 0.0027 -0.002 0.0006 -0.0014 0.00051 -0.002 1 0.23 0.0062 0.0034 -3.7e-05 2.8e-06
AMT_REQ_CREDIT_BUREAU_DAY 0.0011 0.008 0.0051 0.0025 0.0055 0.0016 0.0016 0.00091 9.3e-06 0.0022 0.0039 0.0012 0.0019 0.0012 0.0023 0.0023 0.0021 0.0023 0.0078 0.0015 0.0015 0.0016 0.0016 0.23 1 0.22 0.0024 0.002 0.00011
AMT_REQ_CREDIT_BUREAU_WEEK 0.00032 0.0086 0.00094 0.012 0.0012 0.0028 0.0036 0.0017 0.0013 0.0069 0.0015 0.00078 0.00092 0.0015 0.0026 0.0035 0.0035 0.0035 0.022 0.0011 0.0016 0.0012 0.0022 0.0062 0.22 1 0.0078 0.0081 0.029
AMT_REQ_CREDIT_BUREAU_MON 0.0076 0.059 0.055 0.036 0.057 0.071 0.0019 0.034 0.011 0.017 0.036 0.0029 0.006 0.0082 0.013 0.014 0.0079 0.051 0.013 0.0026 0.00097 0.0026 0.0013 0.0034 0.0024 0.0078 1 0.0045 0.013
AMT_REQ_CREDIT_BUREAU_QRT 0.0042 0.018 0.022 0.012 0.022 0.002 0.015 0.013 0.00036 0.017 0.0012 0.004 0.0062 0.0059 0.0031 0.0076 0.007 0.0015 0.02 0.005 0.0006 0.0049 0.00026 3.7e-05 0.002 0.0081 0.0045 1 0.093
AMT_REQ_CREDIT_BUREAU_YEAR 0.036 0.034 0.038 0.008 0.04 0.00015 0.073 0.047 0.024 0.048 0.025 0.018 0.022 0.017 0.0096 0.012 0.011 0.013 0.071 0.031 0.018 0.032 0.017 2.8e-06 0.00011 0.029 0.013 0.093 1
CNT_CHILDREN
AMT_INCOME_TOTAL
AMT_CREDIT
AMT_ANNUITY
AMT_GOODS_PRICE
REGION_POPULATION_RELATIVE
DAYS_BIRTH
DAYS_EMPLOYED
DAYS_REGISTRATION
DAYS_ID_PUBLISH
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
EXT_SOURCE_2
EXT_SOURCE_3
OBS_30_CNT_SOCIAL_CIRCLE
DEF_30_CNT_SOCIAL_CIRCLE
OBS_60_CNT_SOCIAL_CIRCLE
DEF_60_CNT_SOCIAL_CIRCLE
AMT_REQ_CREDIT_BUREAU_HOUR
AMT_REQ_CREDIT_BUREAU_DAY
AMT_REQ_CREDIT_BUREAU_WEEK
AMT_REQ_CREDIT_BUREAU_MON
AMT_REQ_CREDIT_BUREAU_QRT
AMT_REQ_CREDIT_BUREAU_YEAR
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

Correlation for target 0

Points to be concluded.

1.Credit amount is inversely proportional to the date of birth, which means Credit amount is higher for low age and vice-versa.

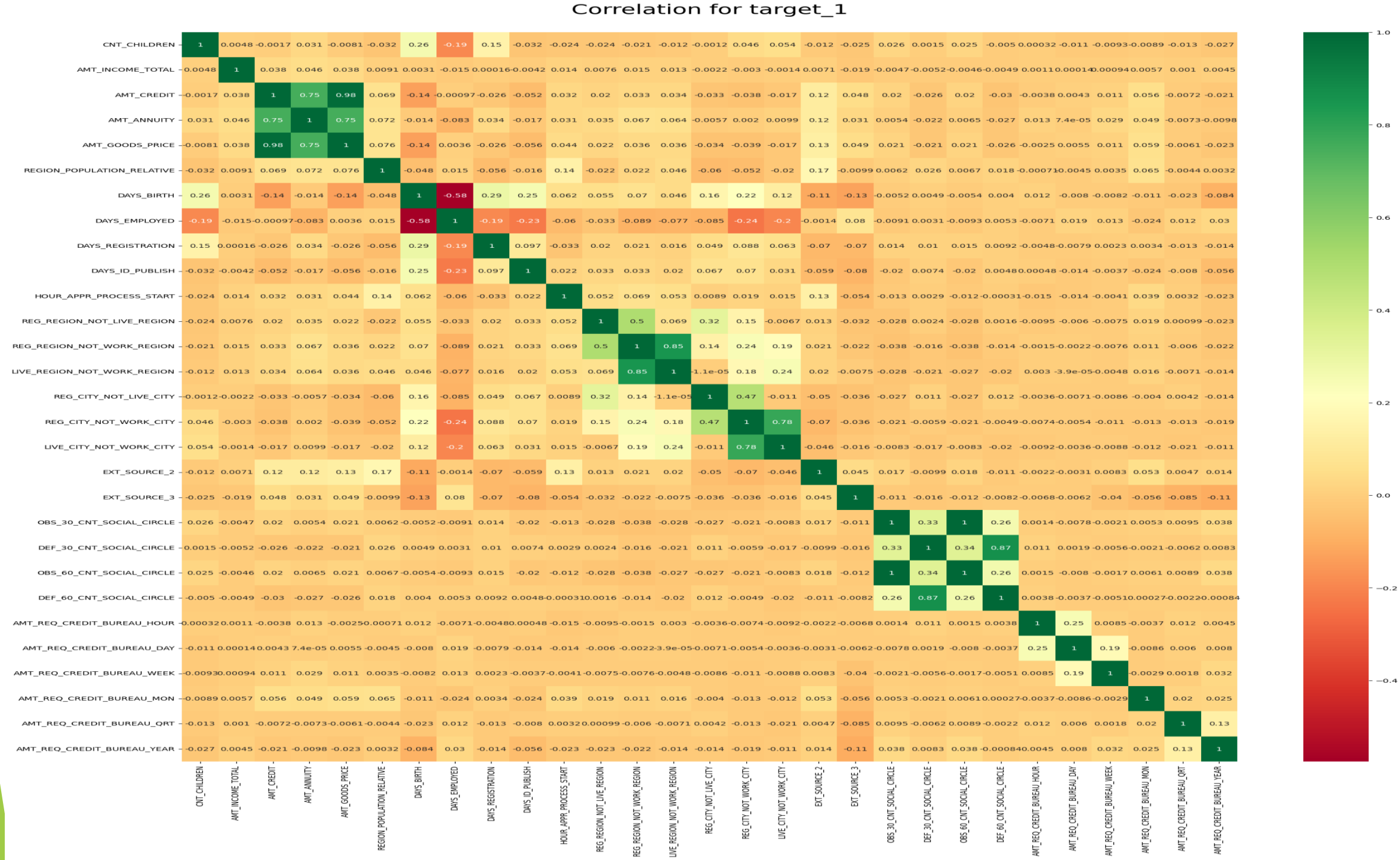
2. Credit amount is inversely proportional to the number of children client have, means credit amount is higher for less children count have and vice-versa.

3. Income amount is inversely proportional to the number of children client have, means more income for less children client have and vice-versa.

4. Less children client have in densely populated area.

5. Credit amount is higher to densely populated area.

6. The income is also higher in densely populated area.


Correlation for target_1
CNT_CHILDREN
AMT_INCOME_TOTAL
AMT_CREDIT
AMT_ANNUITY
AMT_GOODS_PRICE
REGION_POPULATION_RELATIVE
DAYS_BIRTH
DAYS_EMPLOYED
DAYS_REGISTRATION
DAYS_ID_PUBLISH
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
EXT_SOURCE_2
EXT_SOURCE_3
OBS_30_CNT_SOCIAL_CIRCLE
DEF_30_CNT_SOCIAL_CIRCLE
OBS_60_CNT_SOCIAL_CIRCLE
DEF_60_CNT_SOCIAL_CIRCLE
AMT_REQ_CREDIT_BUREAU_HOUR
AMT_REQ_CREDIT_BUREAU_DAY
AMT_REQ_CREDIT_BUREAU_WEEK
AMT_REQ_CREDIT_BUREAU_MON
AMT_REQ_CREDIT_BUREAU_QRT
AMT_REQ_CREDIT_BUREAU_YEAR
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

Correlation for type1

This heat map for Target 1 is also having quite a same observation just like Target 0. But for few points are different. They are listed below.

1. The client's permanent address does not match contract address are having less children and vice-versa.
2. The client's permanent address does not match work address are having less children and vice_versa.

CATEGORICAL UNIVARIATE ANALYSIS FOR VARIABLES TARGET 0

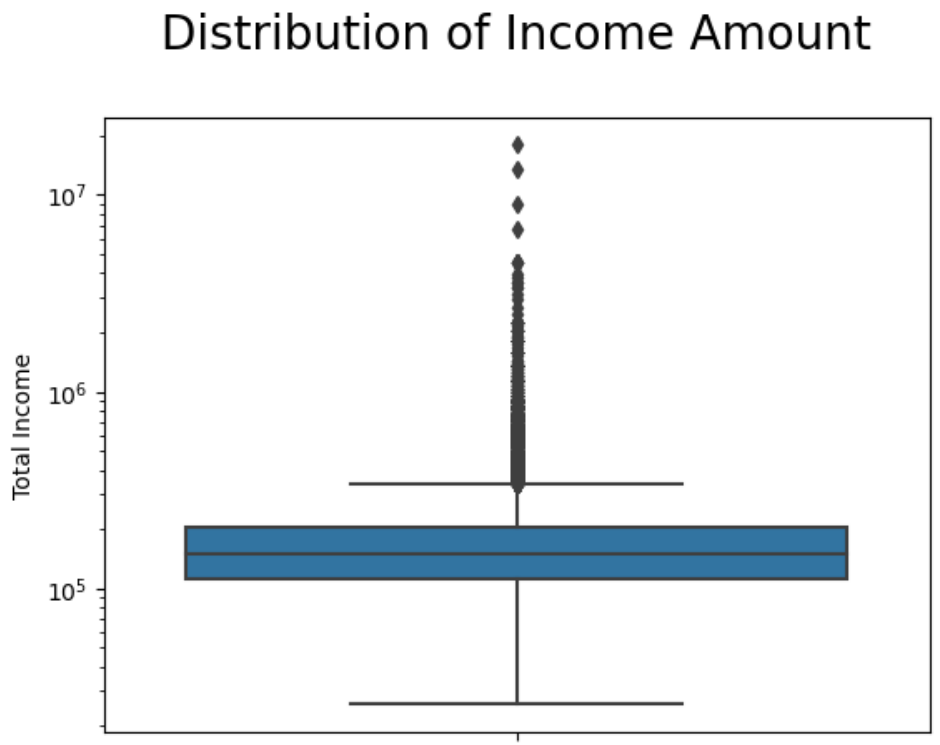
BOXPLOT FOR INCOME AMOUNT

Few points can be concluded.

1.Some outliers are noticed in income amount.

2. The third quartiles is very slim for income

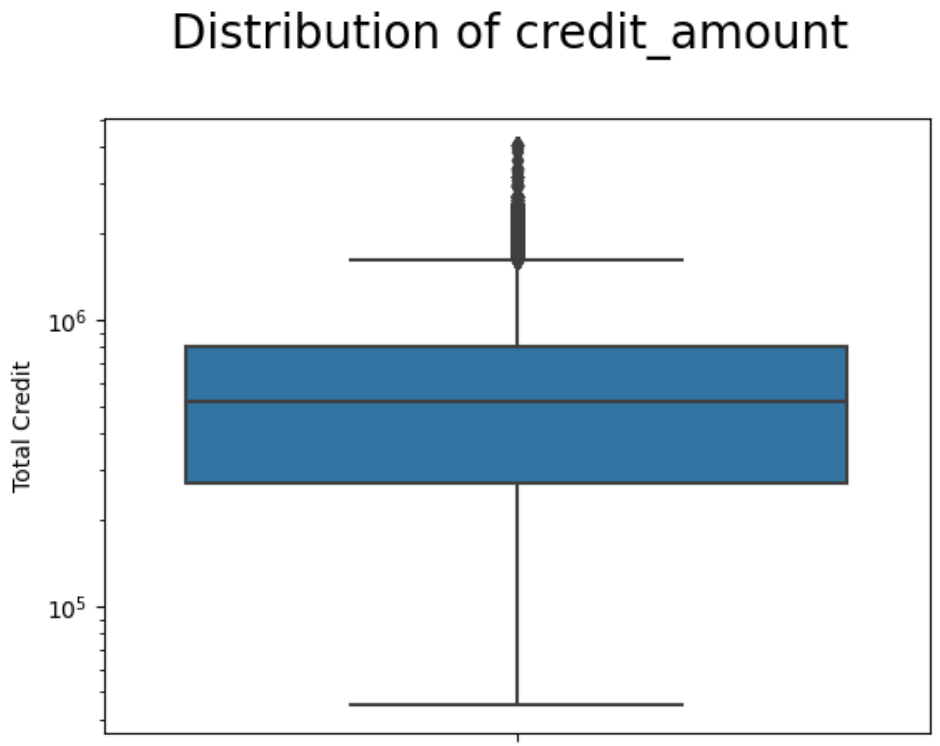
Amount.



BOX PLOT FOR CREDIT AMOUNT

Few Points can be concluded:

1. Some outliers are noticed in credit amount.
2. The first quartile is bigger than third quartile for credit amount which means most of the credits of clients are present in the first quartile.

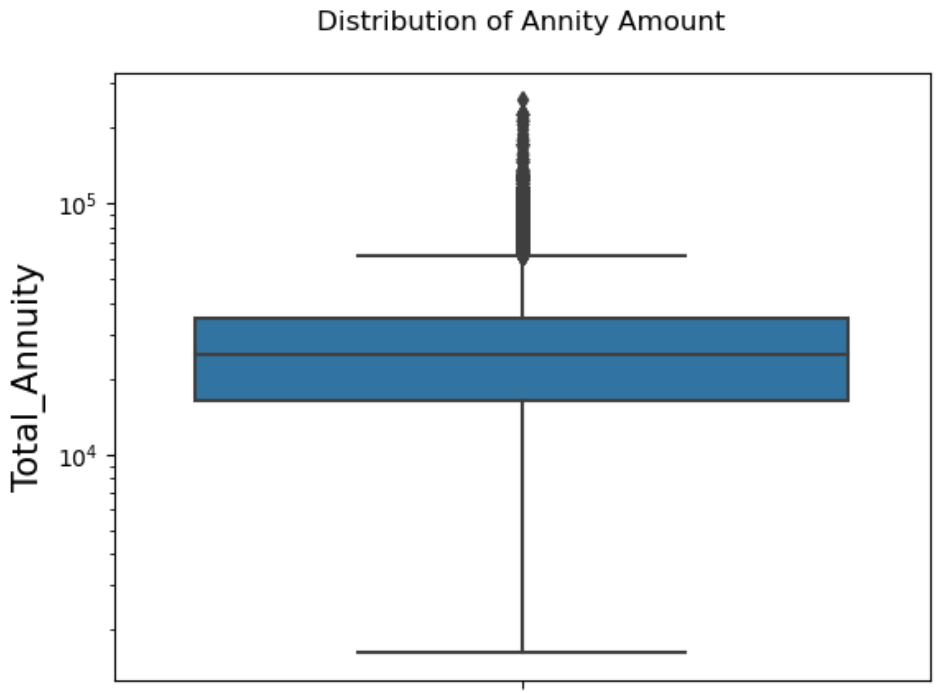


BOXPLOT FOR ANNUITY AMOUNT

Few points can be concluded.

1.Some outliers are noticed in annuity amount.

2. The first quartile is bigger than third quartile for annuity amount which means most of the annuity clients are from first quartile.

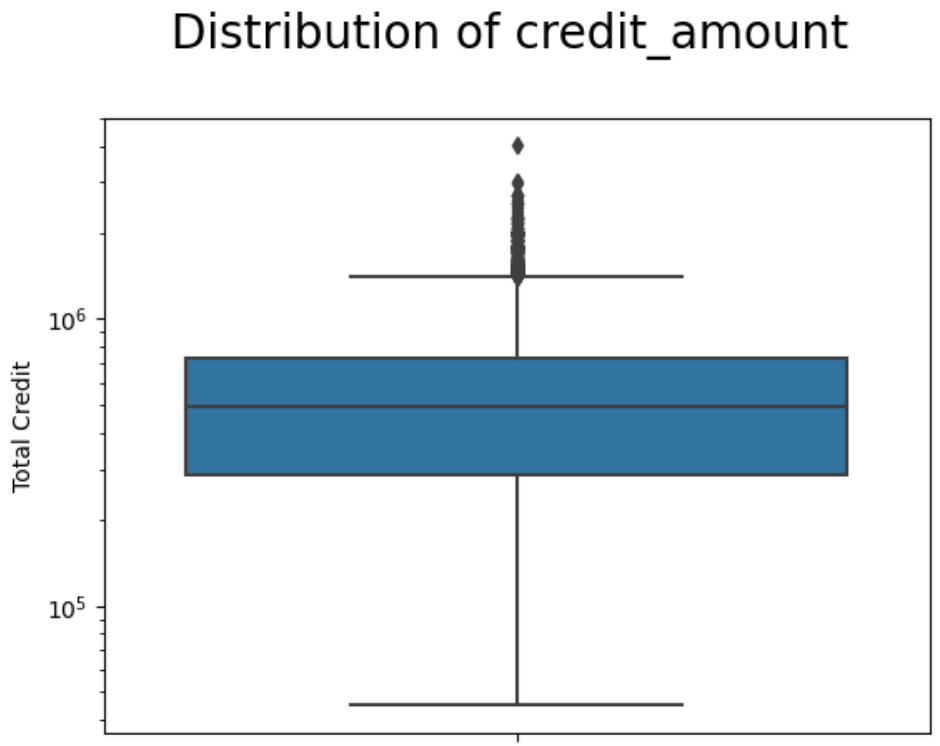


CATEGORICAL UNIVARIATE ANALYSIS FOR VARIABLES TARGET1

BOXPLOT FOR INCOME AMOUNT

Few points can be concluded.

1. Some outliers are noticed in income amount.
2. The third quartile is very slim for income amount.
3. Most of the clients of income are present in first quartile.

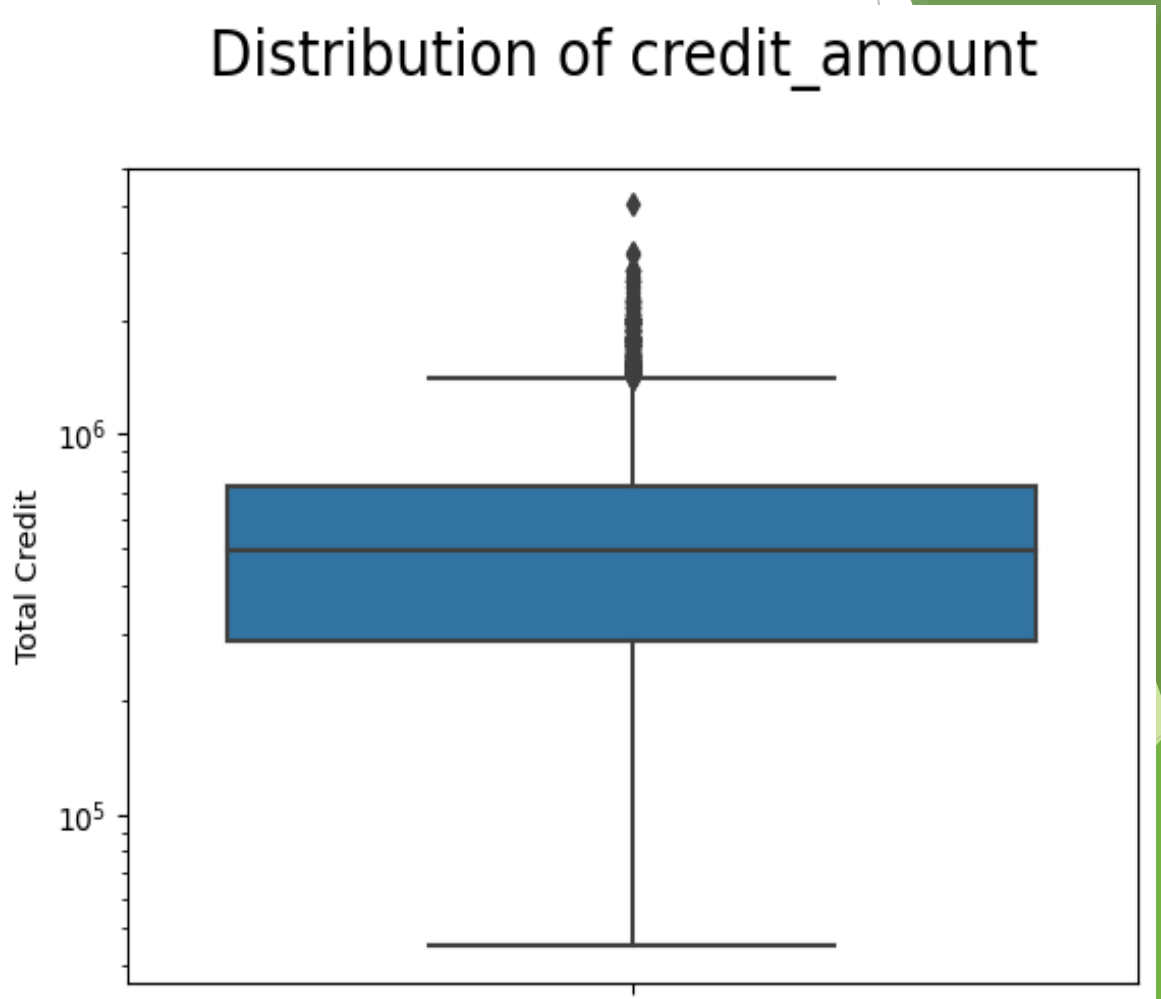


BOXPLOT FOR CREDIT AMOUNT

Few points can be concluded from the graph.

1.Some outliers are noticed in credit amount.

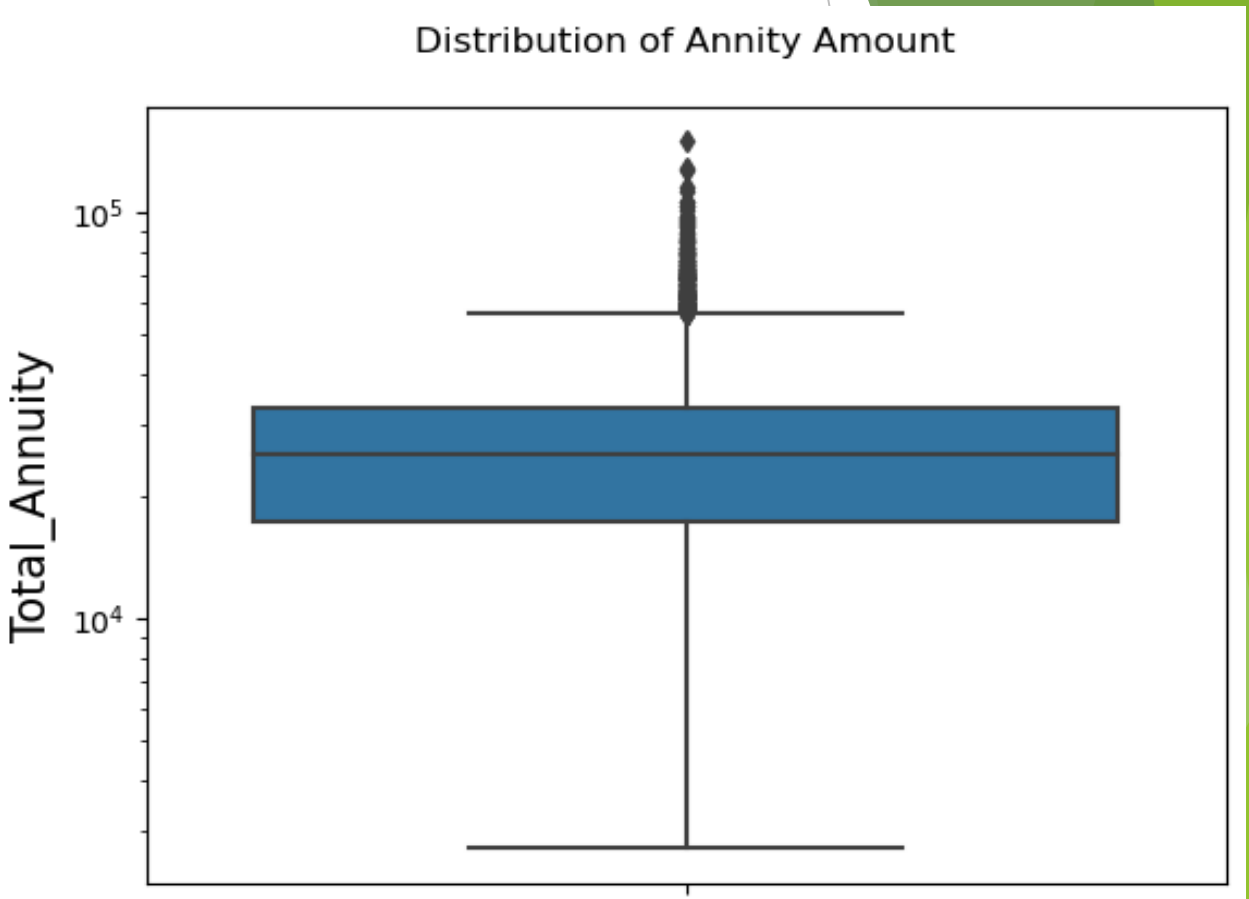
2.The first quartile is bigger than third quartile for credit amount which means most of the credits of clients are present in the first quartile.



Boxplot for Annuity amount

Few Points can be concluded from the graph.

1. Some outliers are noticed in annuity amount.
2. The first quartile is bigger than third quartile for annuity amount which means most of the annuity clients are from first quartile.



BIVARIATE ANALYSIS FOR TYPE 0

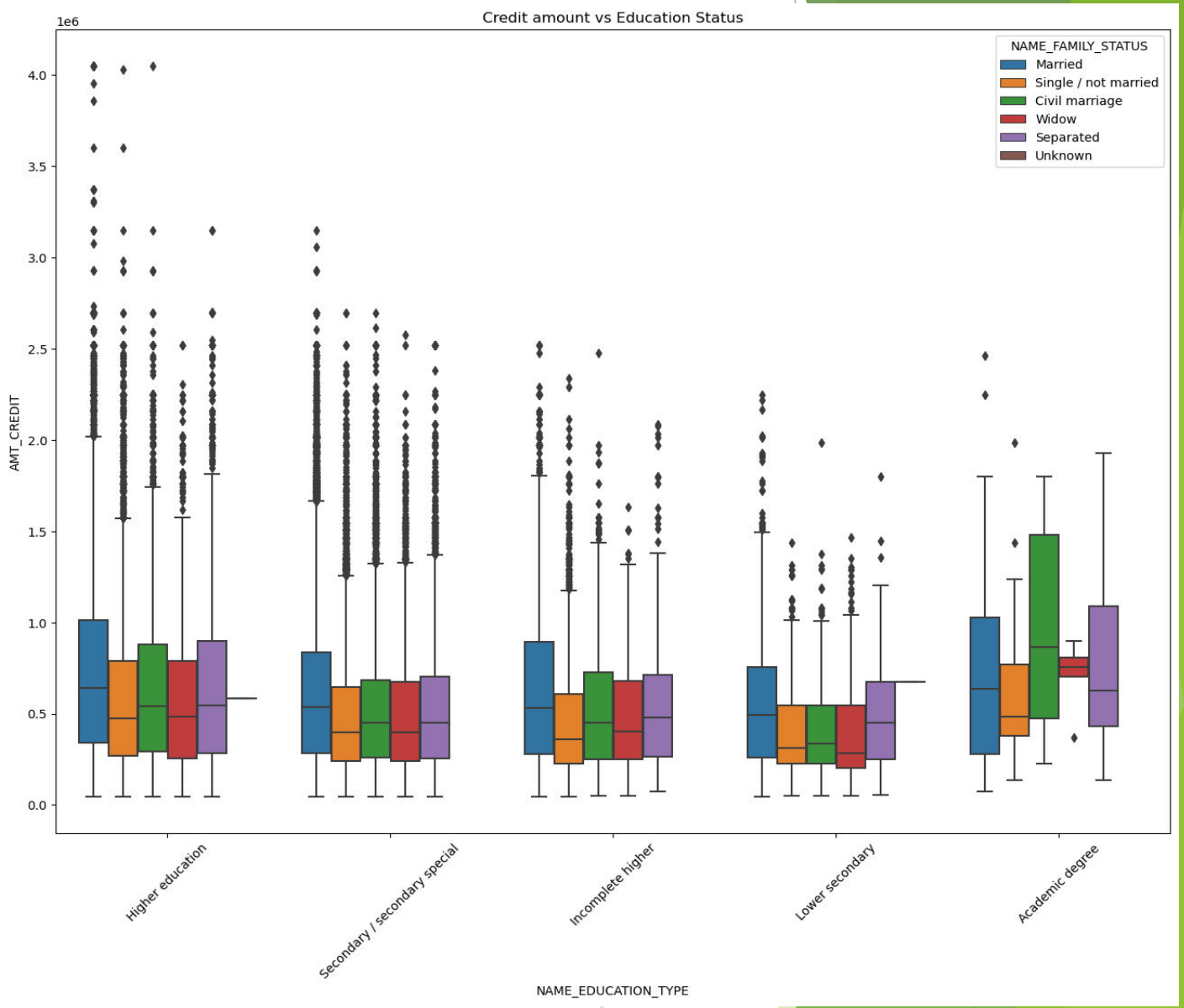
CREDIT AMOUNT VS EDUCATION STATUS

Few points can be concluded from the graph:

1.Family status of “civil marriage”, “marriage” and separation of academic degree education are having higher number of credits than others.

2. Higher education of the family status of “marriage”, “single” and “civil marriage” are having more outliers.

3. Civil marriage for academic degree is having most of the credits in the third quartile.



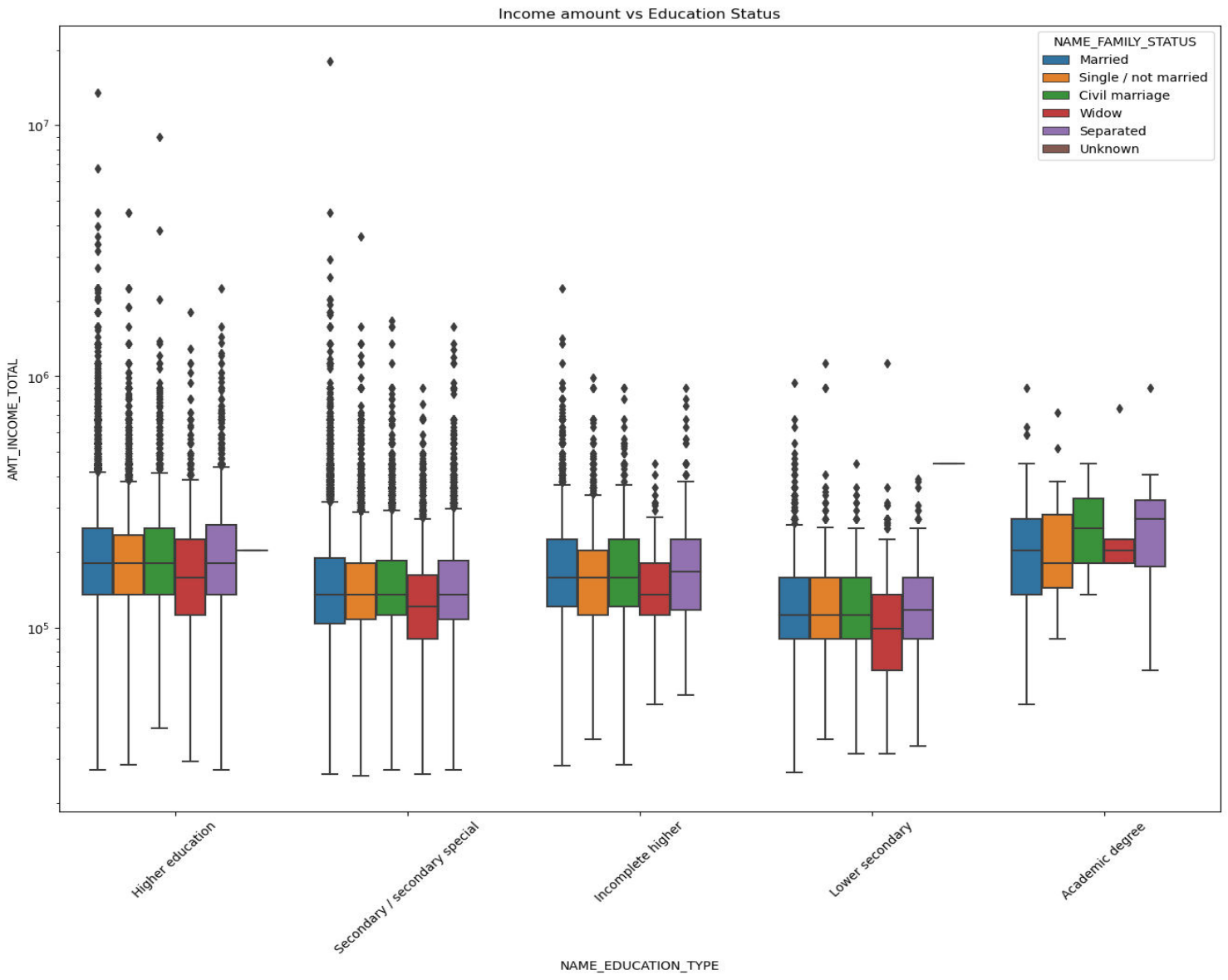
INCOME AMOUNT VS EDUCATION STATUS

Few points can be concluded:

1. for education type “Higher education” contains most outliers.

2. Less outliers are having for academic degree but they are having the income amount is little higher that higher education.

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



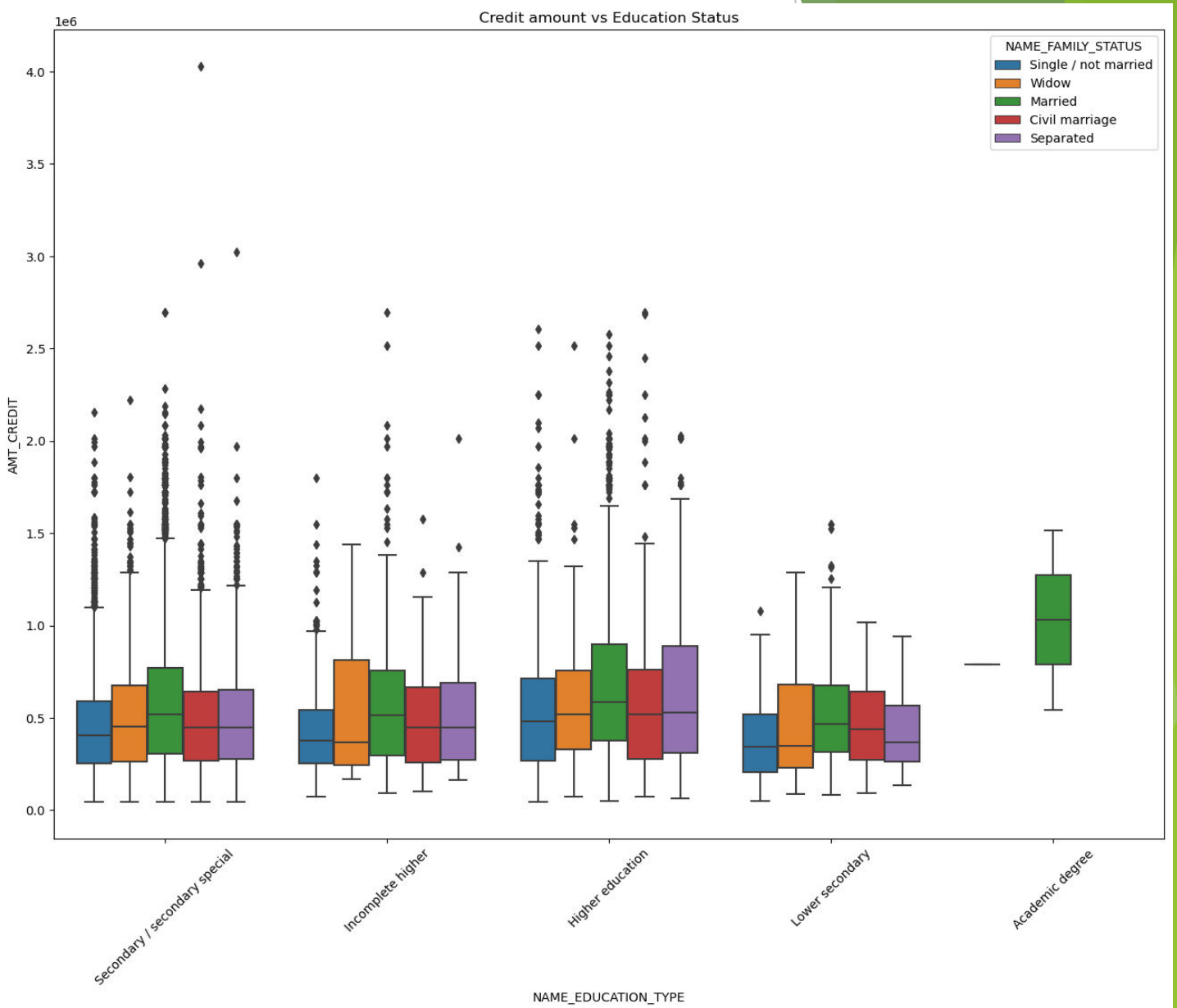
BIVARIATE ANALYSIS FOR TYPE 1

CREDIT AMOUNT VS EDUCATION STATUS

Few points can be concluded.

1. Quite similar from Target 0, we can say that Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others.

2. Most of the outliers are from Education type 'Higher education' and 'Secondary'.

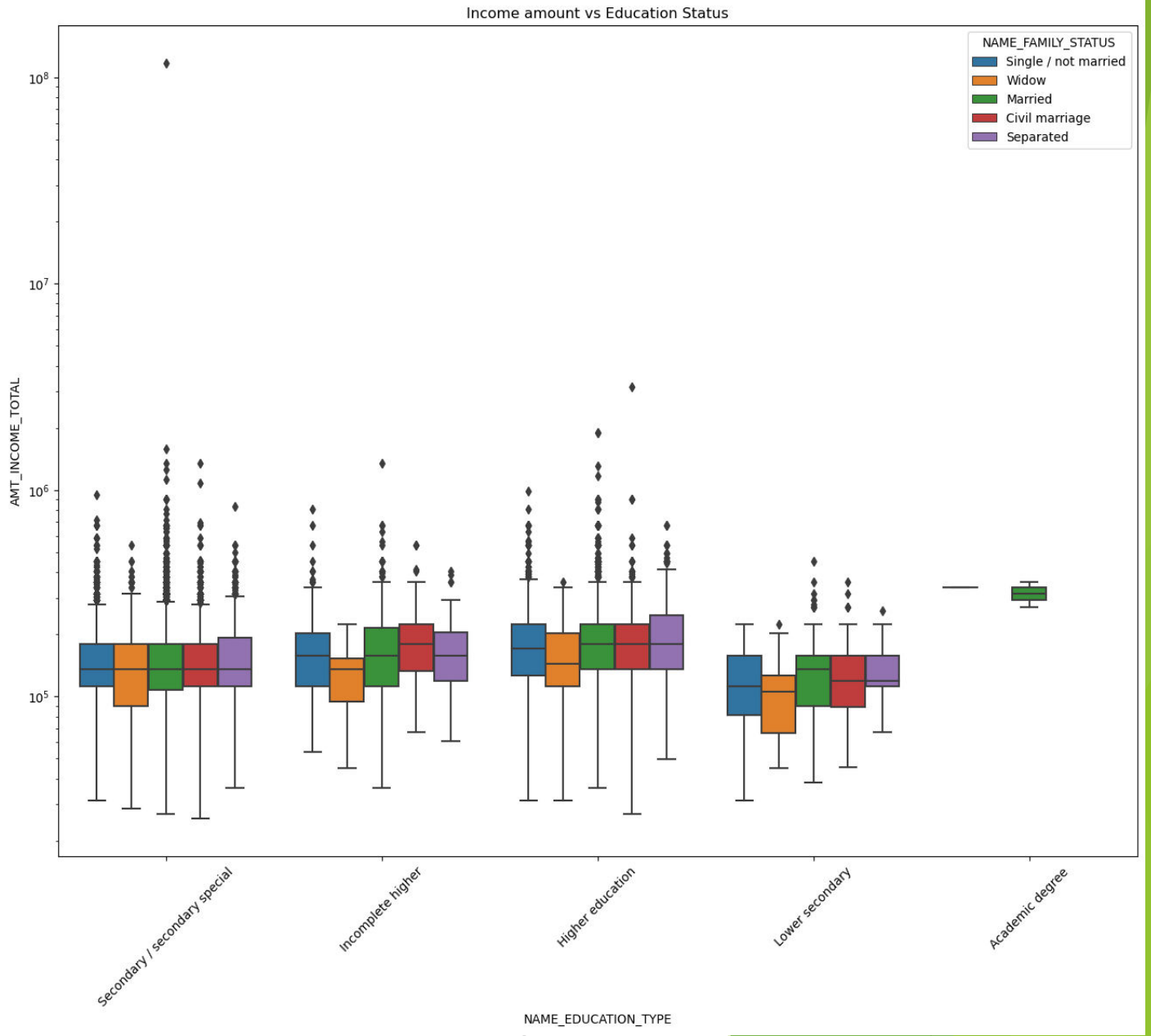


INCOME AMOUNT VS EDUCATION STATUS.

Few points can be concluded from the graph.

1. Have some similarity with Target0, From above boxplot for Education type 'Higher education' the income amount is mostly equal with family status.

2. Lower secondary are have less income amount than others.



UNIVARIATE ANALYSIS AFTER MERGING PREVIOUS DATA

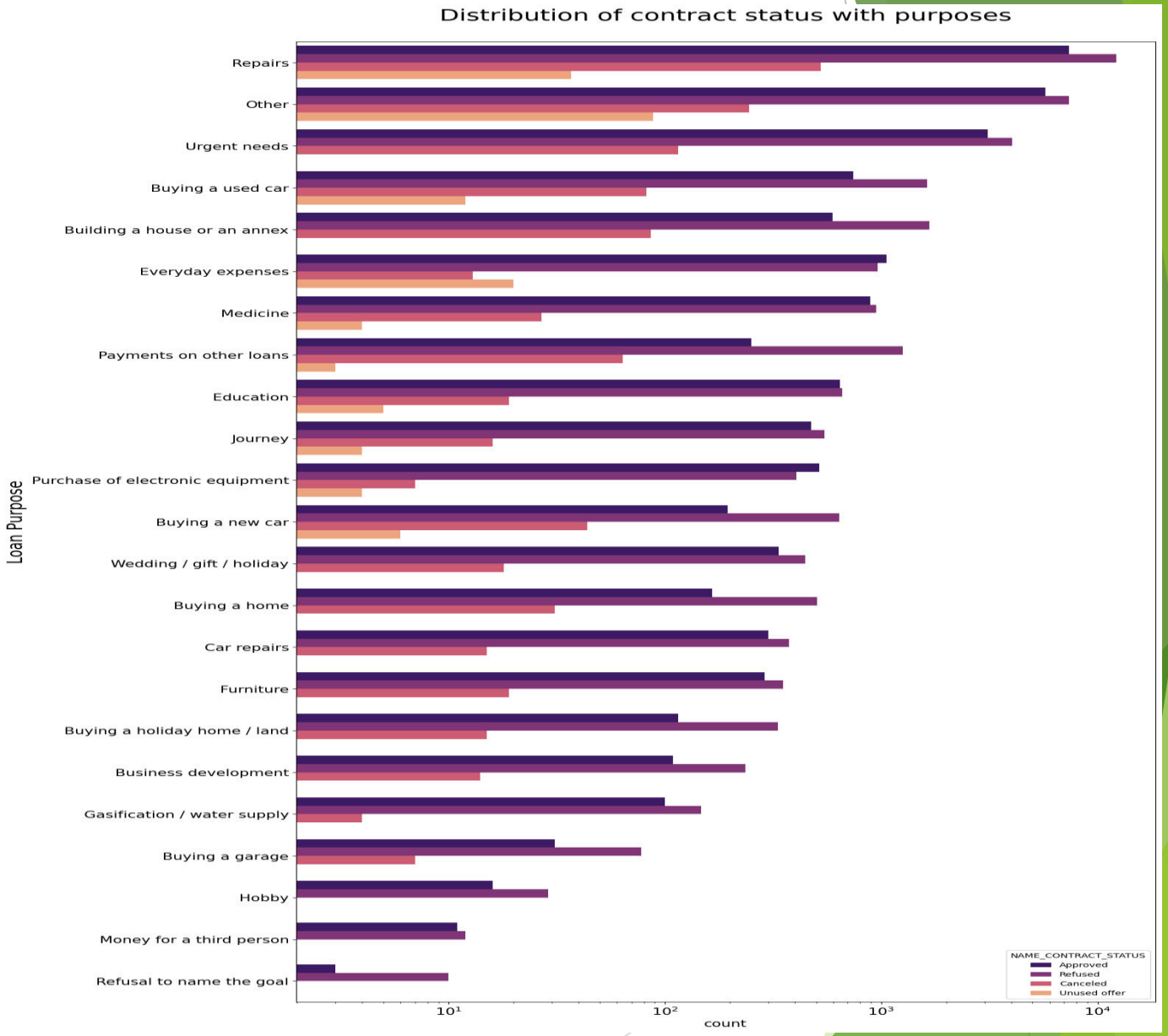
DISTRIBUTION OF CONTRACT STATUS WITH PURPOSES

Few points can be concluded from the graph.

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.

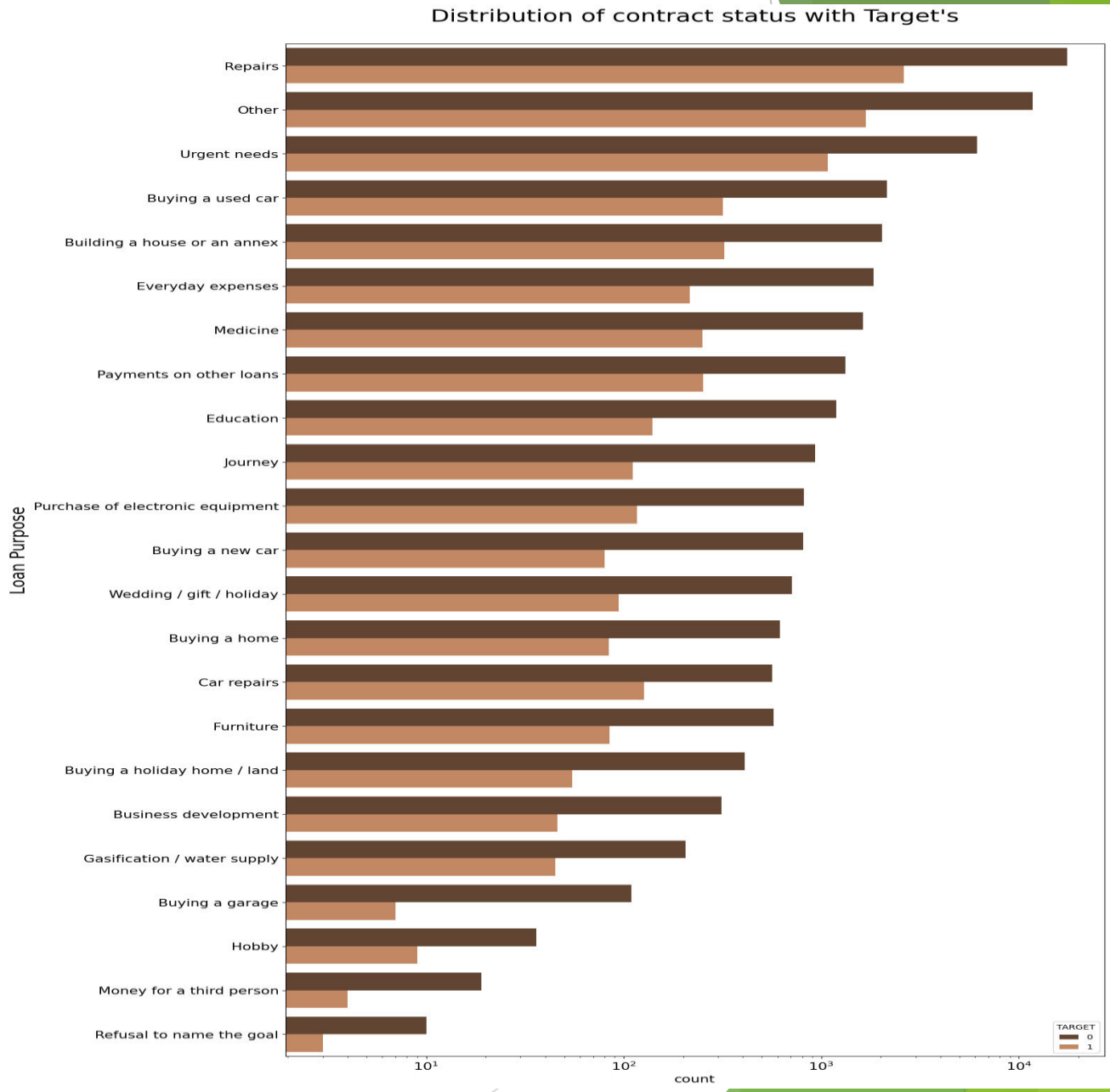


DISTRIBUTION OF PURPOSES WITH TARGET

Few points can be concluded from the graph.

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. They are '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.



PERFORMING BIVARIATE ANALYSIS

Prev Credit amount vs Loan Purpose
NAME_INCOME_TYPE
Working
Commercial associate
Pensioner
State servant
Student
Unemployed
AMT_CREDIT_PREV
10^6
10^5
Other
Payments on other loans
Buying a used car
Repairs
Education
Buying a new car
Everyday expenses
Medicine
Car repairs
Urgent needs
Buying a holiday home / land
Building a house or an annex
Furniture
Journey
Purchase of electronic equipment
Wedding / gift / holiday
Buying a home
Business development
Gasification / water supply
Buying a garage
Hobby
Money for a third person
Refusal to name the goal
NAME_CASH_LOAN_PURPOSE

Prev Credit amount vs Loan Purpose

From the previous graph we can conclude the below points:

1. The credit amount of loan purposes like “Buying a home”, “buying a land”, “Buying a new car” and “building a house” is higher.
2. Income type of state servants have a significant amount of credit applied.
3. Money for third person or a hobby is having less credits applied for.

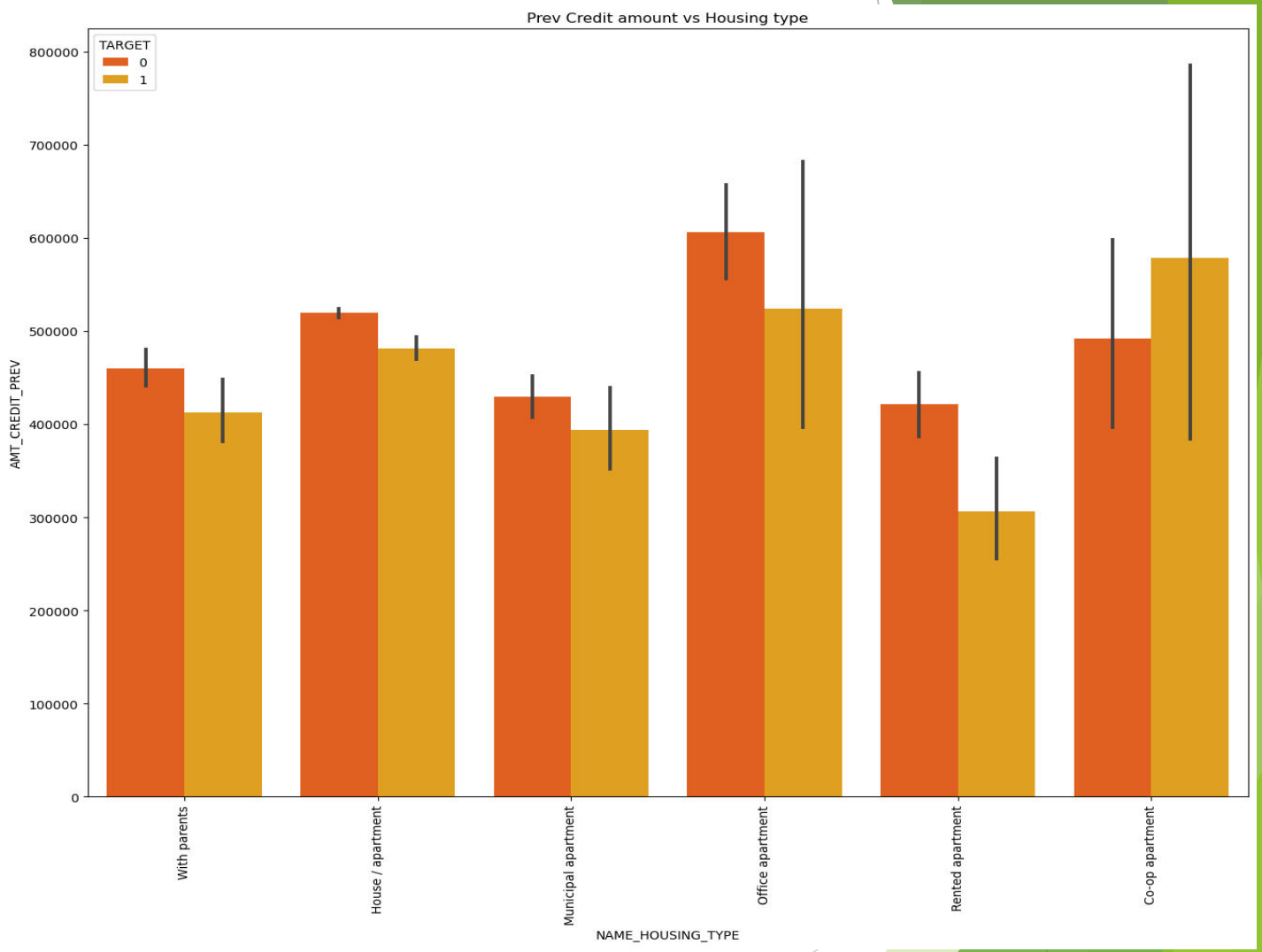
PREV CREDIT AMOUNT VS HOUSING TYPE

Few points can be concluded from the graph.

1.Here for Housing type, office apartment is having higher credit of target 0 and co-op apartment is having higher credit of target 1.

2.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.

3.Bank can focus mostly on housing type with parents or House\apartment or municipal apartment for successful payments.



Conclusion:

- **1.Bank should approve loans more for the office-apartment , housing type as there are less payment difficulties**
- **2.Banks should provide loans to "Repairs" and "Others purposes"**
- **3.Banks should provide loans to the "Business entity type 3" and "self-employed persons"**
- **4."Working" people are specially female employee are the best to target to the loans.¶**

THANK YOU