CREDIT EDA CASE STUDY

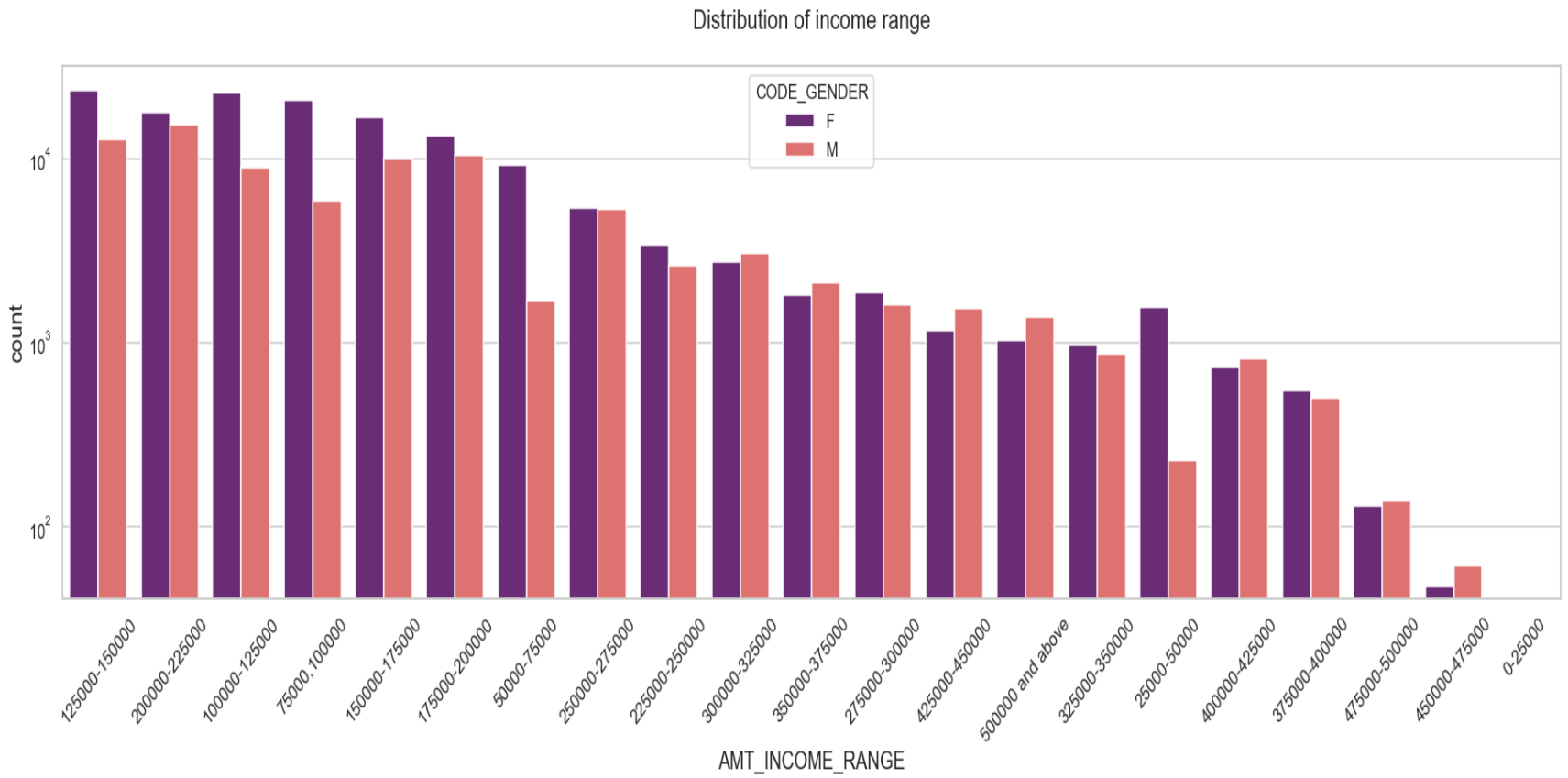
BY
SANJAY KUMAR MISHRA
GAURAV GILALKAR

Categorical Univariate analysis for Target = 0

Distribution of Income range

Conclusion

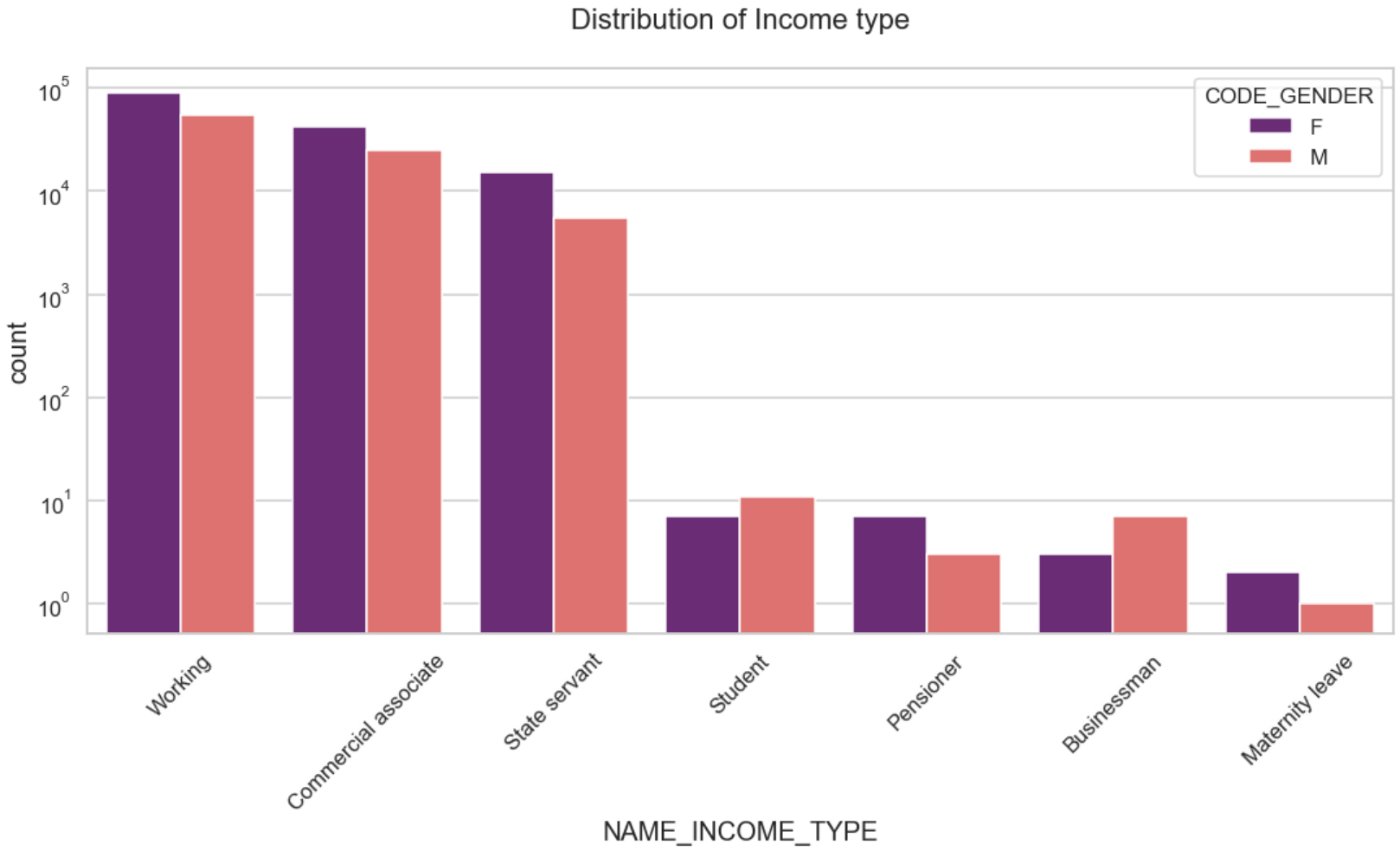
- Female counts are higher than male.
- Income ranges from 100000 to 200000 is having a greater number of credits.
- This graph show that females are more than male in having credits for that range.
- Very less count for income range 400000 and above.



Distribution of income type

Conclusion

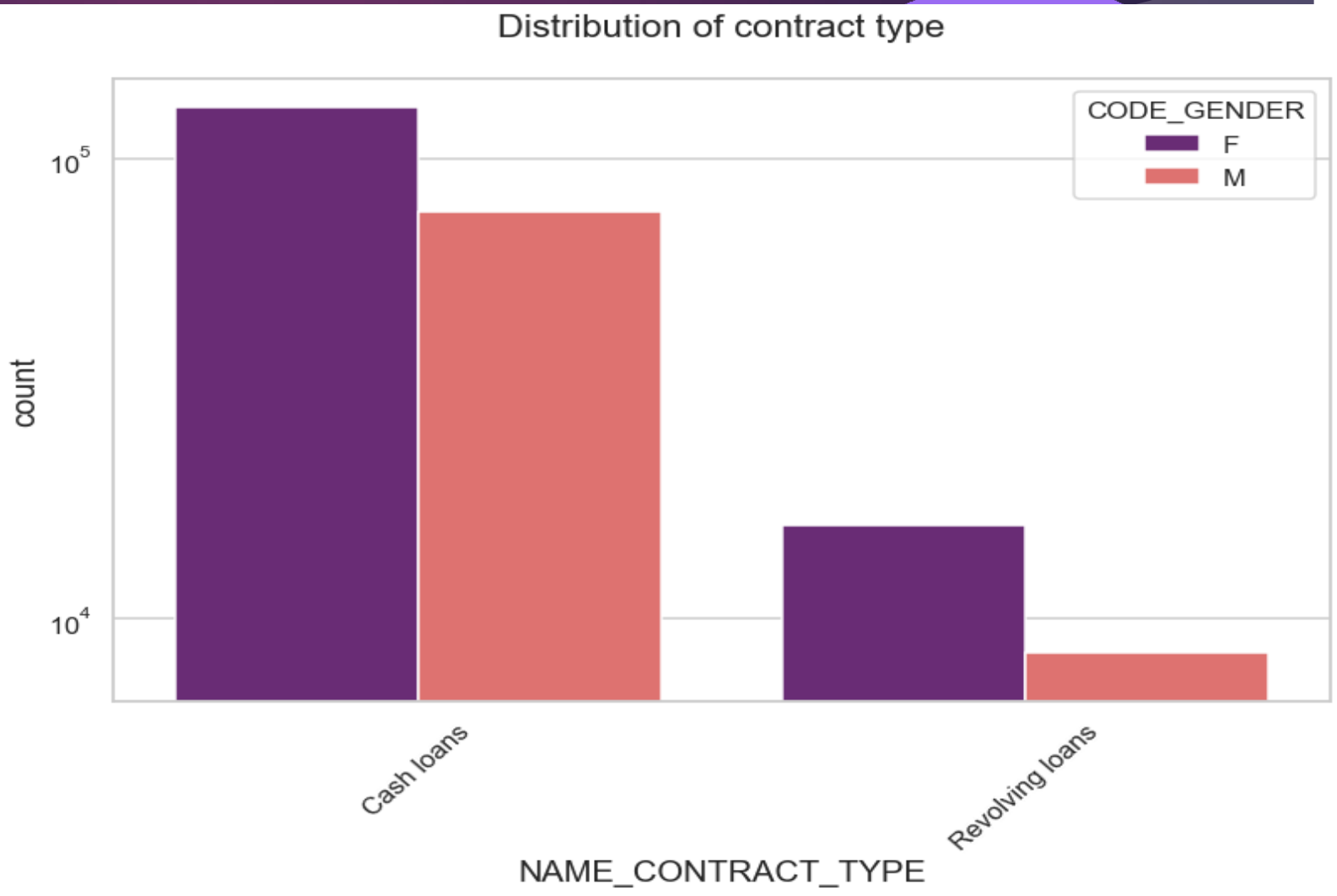
- For income type ‘working’, ’commercial associate’, and ‘State Servant’ the number of credits is higher than others.
  - For this Females are having a greater number of credits than male.
- Less number of credits for income type ‘student’ ,’pensioner’, ‘Businessman’ and ‘Maternity leave’.



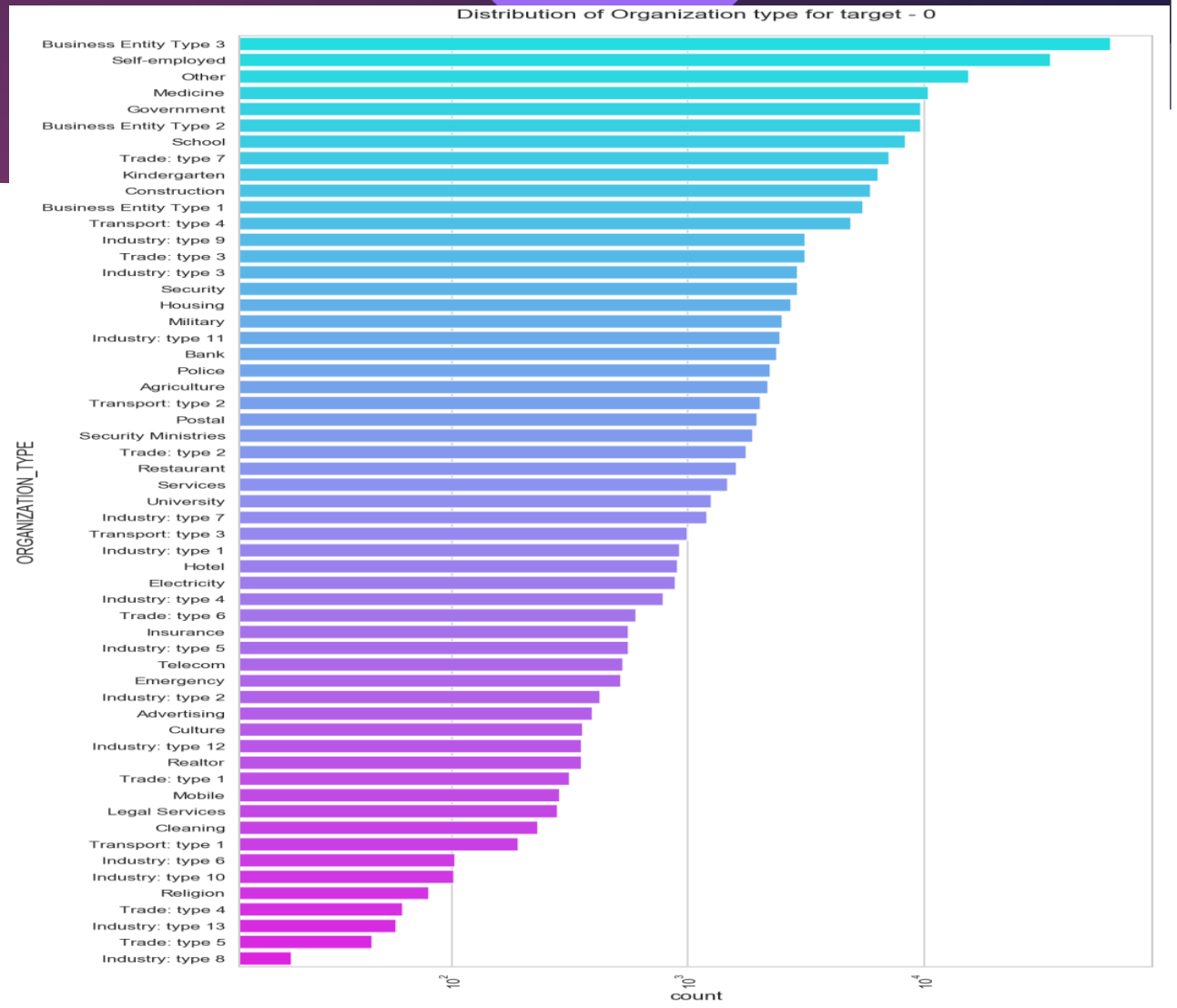
Distribution for contract type

Conclusion:

- Contract type ‘cash loans’ is having higher number of credits than ‘Revolving loans’.
  - For 'cash loans' Females lead in applying credits.



Distribution of organization type



Conclusion

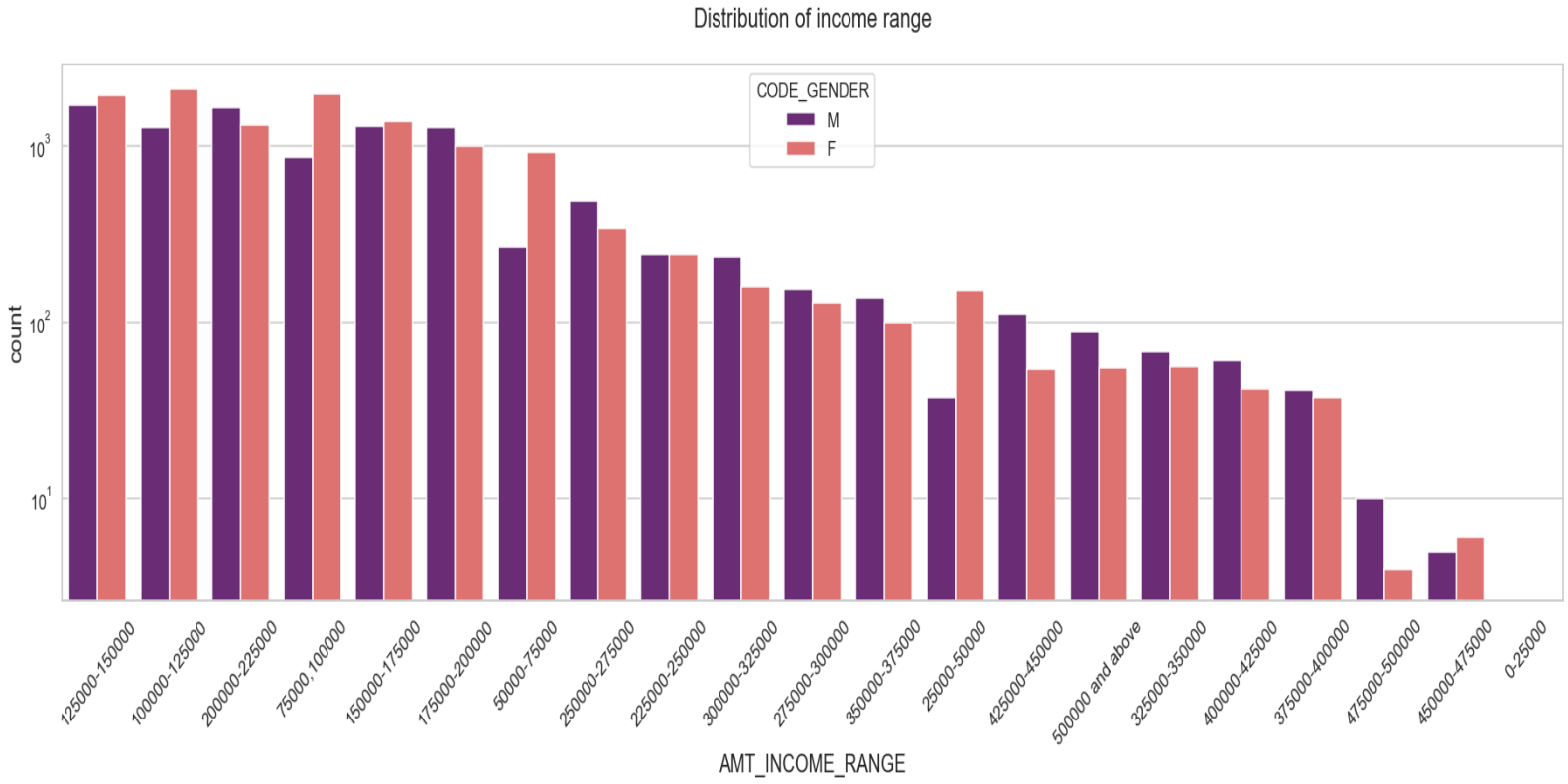
- Clients which have applied for credits are from most of the organization type ‘Business entity Type 3’ , ‘Self employed’, ‘Other’ , ‘Medicine’ and ‘Government’.
- Less clients are from Industry type 8,type 6, type 10, religion and trade type 5, type 4.

Categorical Univariate analysis for Target = 1

Distribution of Income range

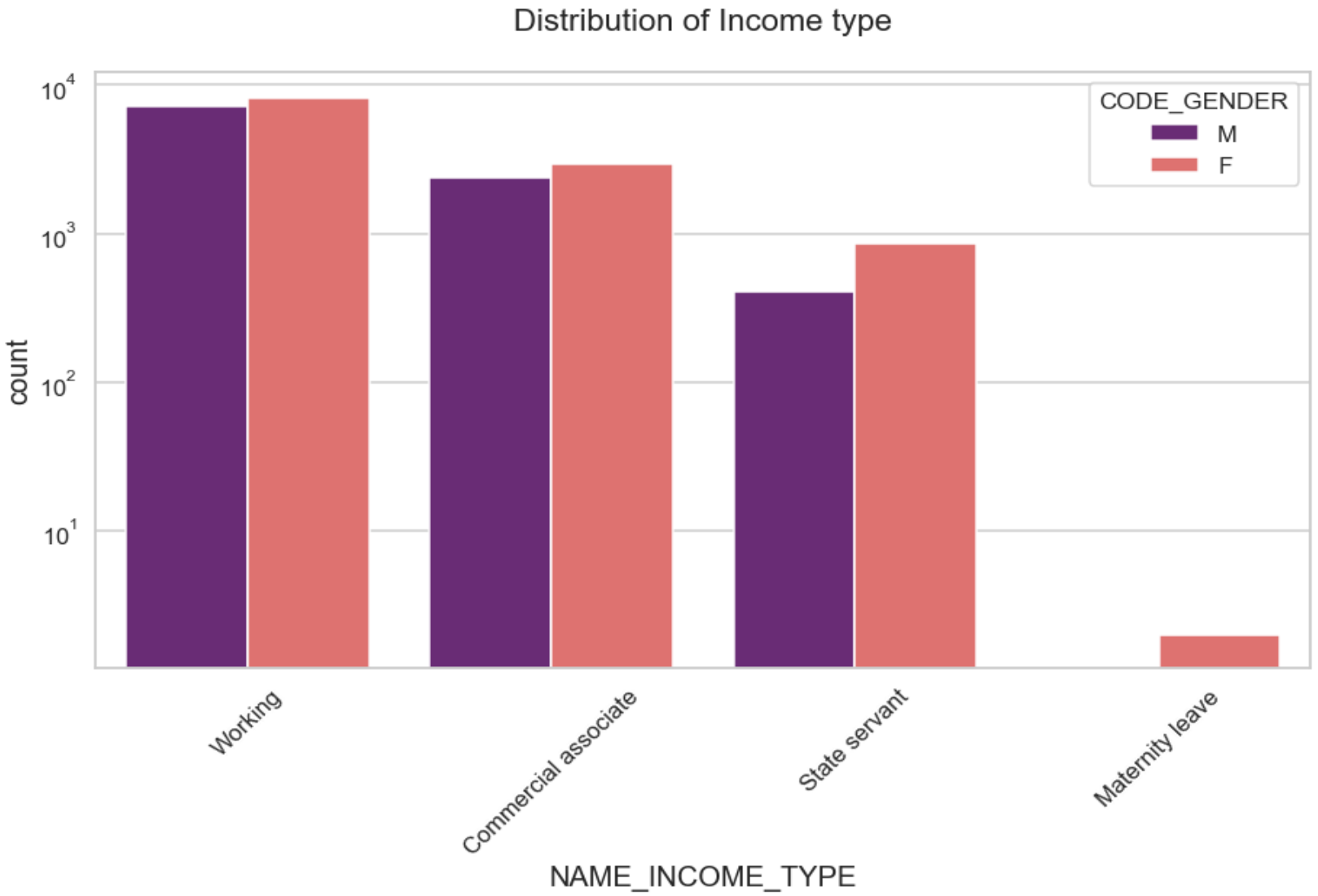
Conclusion

- Overall, Male count is higher than female.
- Income ranges from 100000 to 200000 is having a greater number of credits.
- Very less count for income range 400000 and above.



Distribution of income type

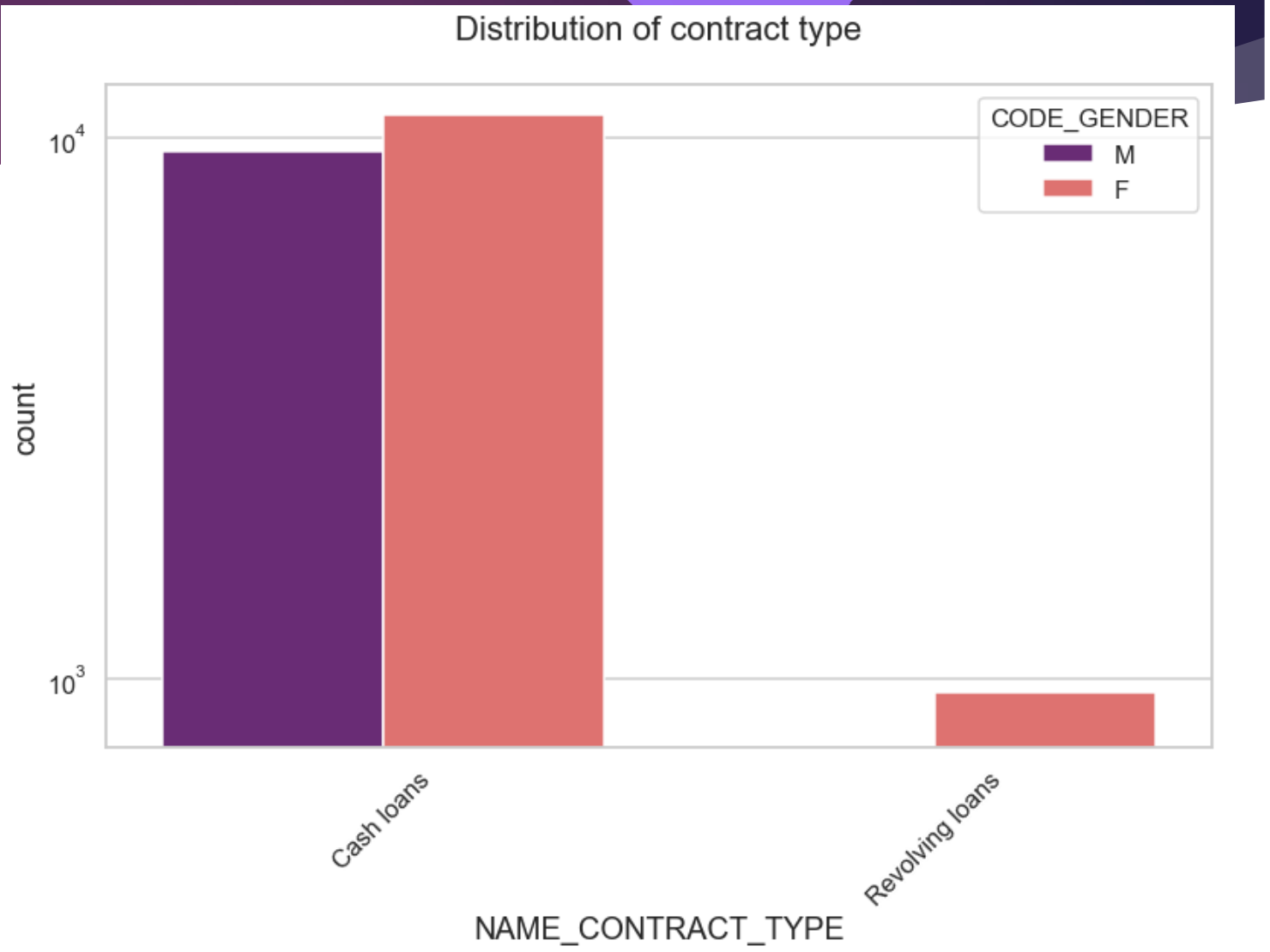
- **Conclusion**
- For income type ‘working’, ’commercial associate’, and ‘State Servant’ the number of credits is higher than other i.e. ‘Maternity leave.
- For this Females are having a greater number of credits than males.
- Less number of credits for income type ‘Maternity leave’.
- For target = 1: There is no income type for ‘student’ , ’pensioner’ and ‘Businessman’ which means they don’t do any late payments.



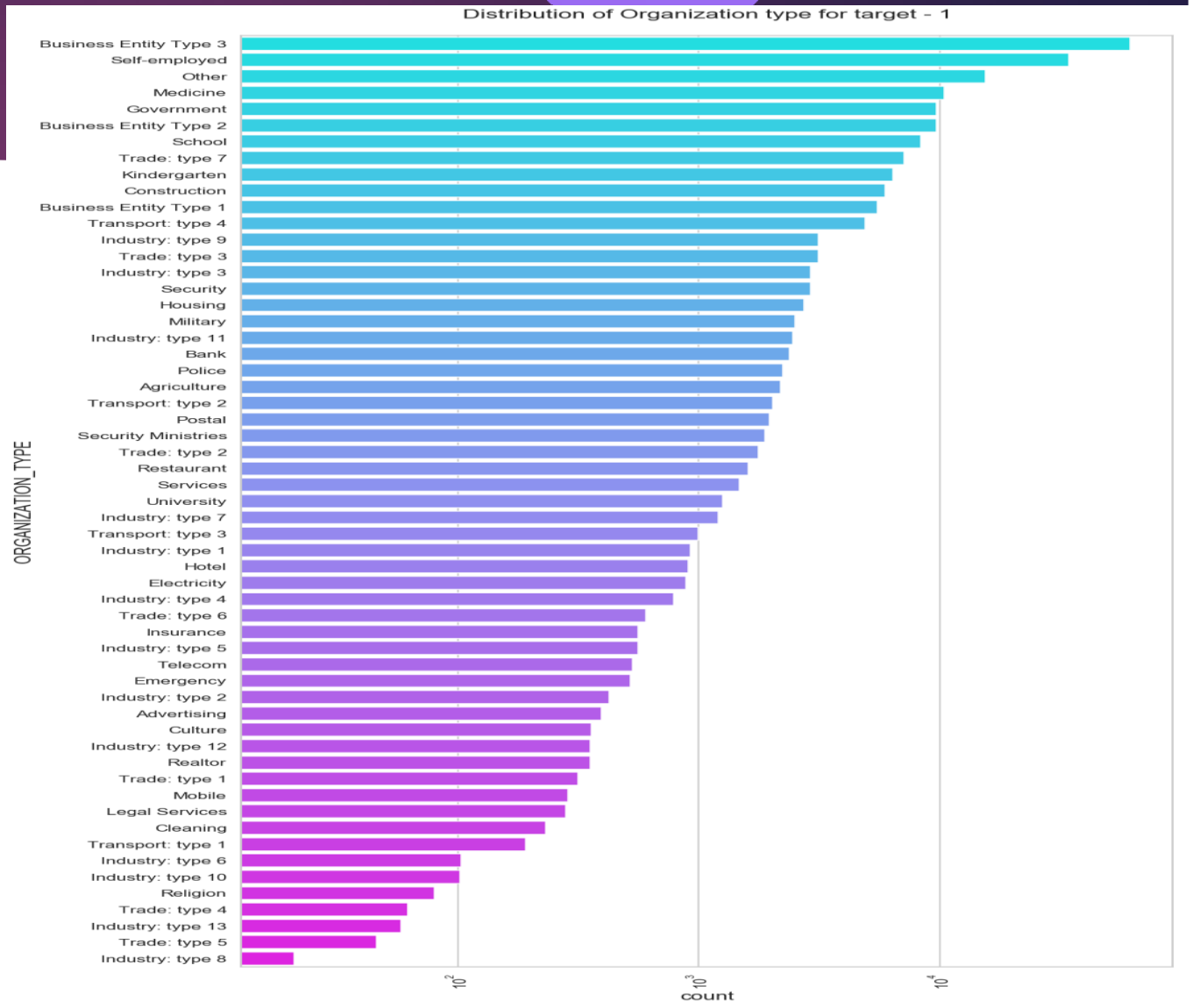
Distribution for contract type

Conclusion

- ‘cash loans’ are more than ‘Revolving loans’ .
- Females are leading for applying credits here as well.
- For target = 1 : there is only Female who applied for Revolving loan.



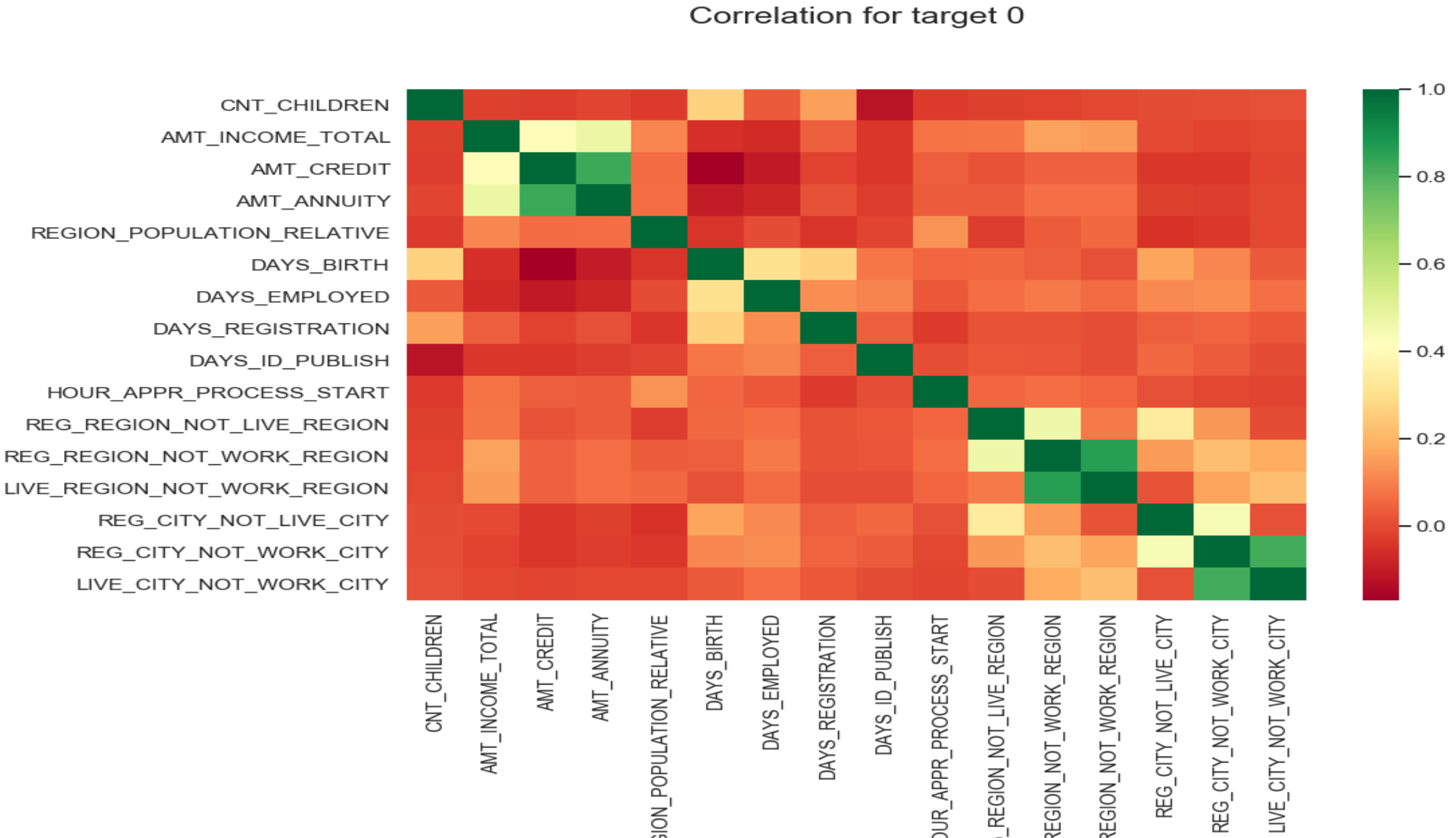
Distribution of organization type



Conclusion

- Clients which have applied for credits are from most of the organization type ‘Business entity Type 3’ , ‘Self employed’ , ‘Other’ , ‘Medicine’ and ‘Government’.
- Less clients are from Industry type 8,type 6, type 10, religion and trade type 5, type 4.
- Same distribution as target = 0.

Correlation for Target = 0

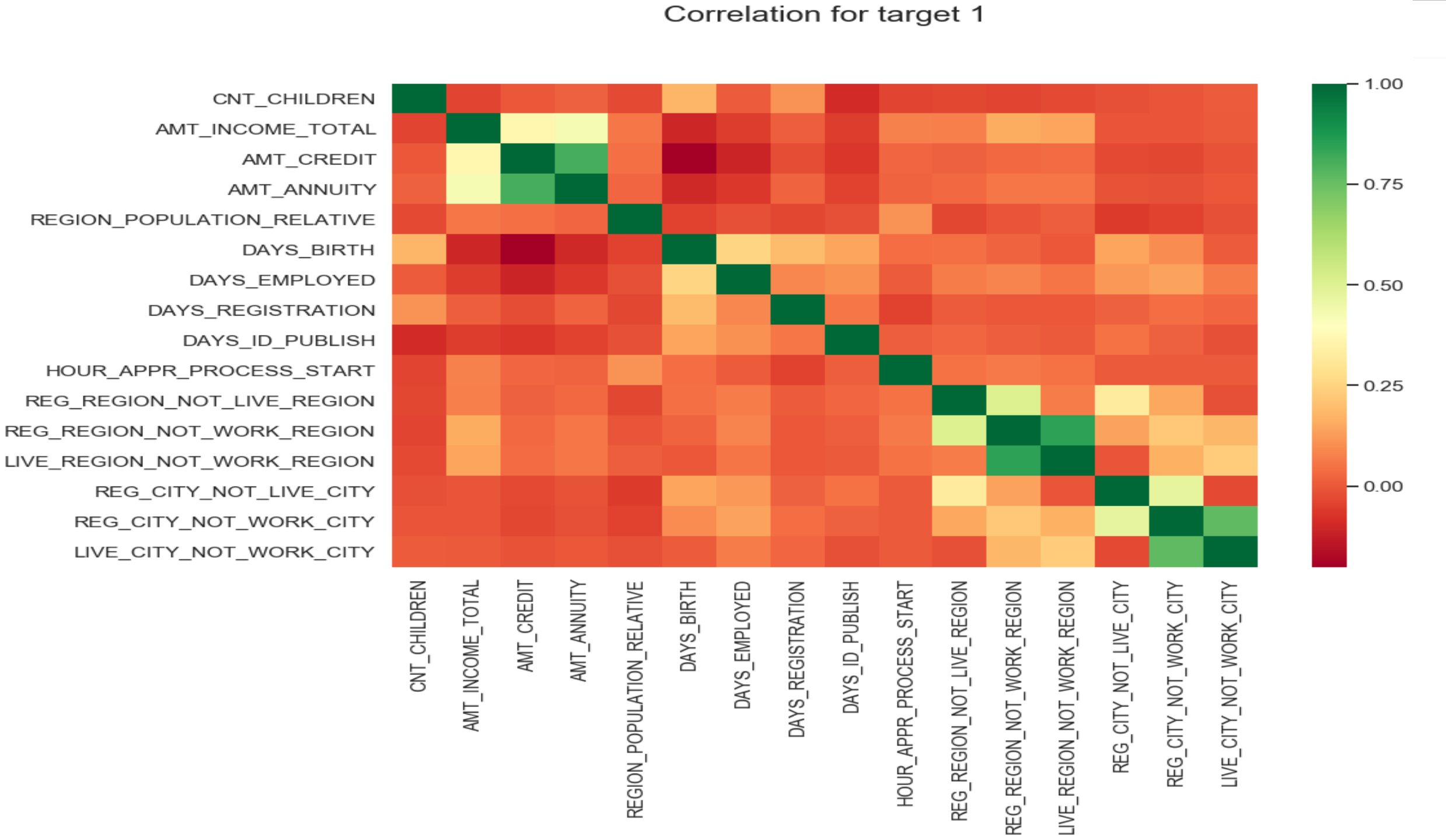

Correlation for target 0
CNT_CHILDREN
AMT_INCOME_TOTAL
AMT_CREDIT
AMT_ANNUITY
REGION_POPULATION_RELATIVE
DAYS_BIRTH
DAYS_EMPLOYED
DAYS_REGISTRATION
DAYS_ID_PUBLISH
HOUR_APPR_PROCESS_START
REG_REGION_NOT_LIVE_REGION
REG_REGION_NOT_WORK_REGION
LIVE_REGION_NOT_WORK_REGION
REG_CITY_NOT_LIVE_CITY
REG_CITY_NOT_WORK_CITY
LIVE_CITY_NOT_WORK_CITY
1.0
0.8
0.6
0.4
0.2
0.0

Correlation For Target = 0

Conclusion

- Credit amount is inversely proportional to the date of birth, which means Credit amount is higher for low age and vice-versa.
    - Credit amount is inversely proportional to the number of children client have, means Credit amount is higher for less children count client have and vice-versa.
- Income amount is inversely proportional to the number of children client have, means more income for less children client have and vice-versa.
- People have less children in densely populated area.
- Credit and annuity amount is higher in densely populated area.
- The income is also higher in densely populated area.

Correlation for Target = 1


Correlation for target 1
CNT_CHILDREN
AMT_INCOME_TOTAL
AMT_CREDIT
AMT_ANNUITY
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
1.00
0.75
0.50
0.25
0.00

Correlation for Target = 1

This heat map for Target 1 is quite similar to Target 0 but with few differences.

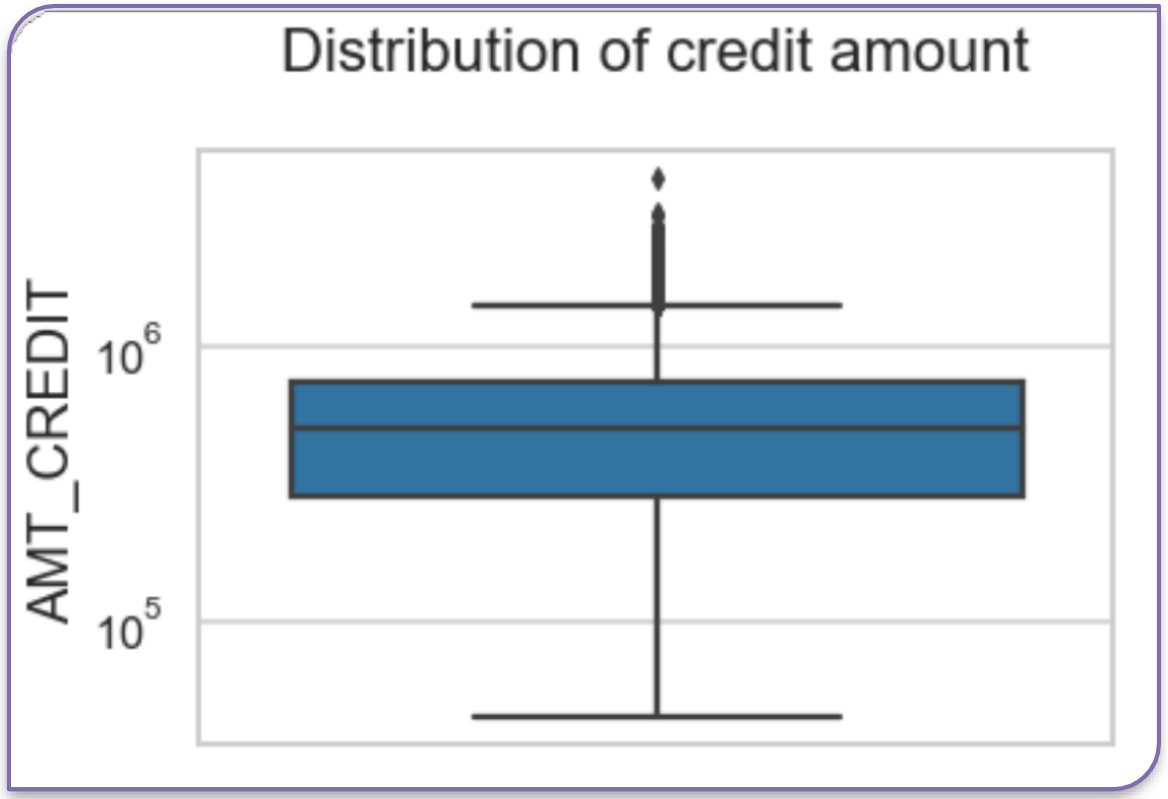
- People with permanent address not matching contact address are having less children (inverse proportion)
- People with permanent address not matching work address are having less children (inverse proportion)

Categorical Univariate analysis of variables for Target = 0

Boxplot for credit amount

Conclusion

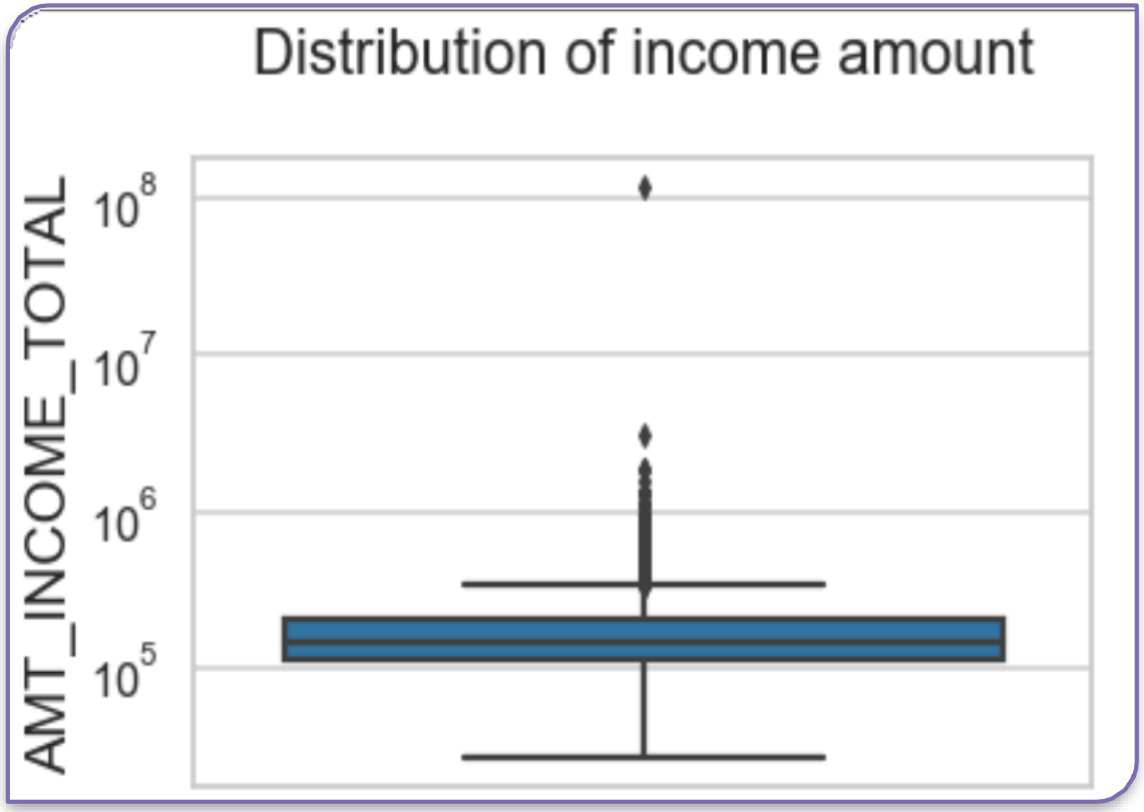
- Some outliers are noticed in credit amount.
- The first quartile is bigger than third quartile for credit amount which means most of the credits of clients are present in the first quartile.



Boxplot for income amount

Conclusion

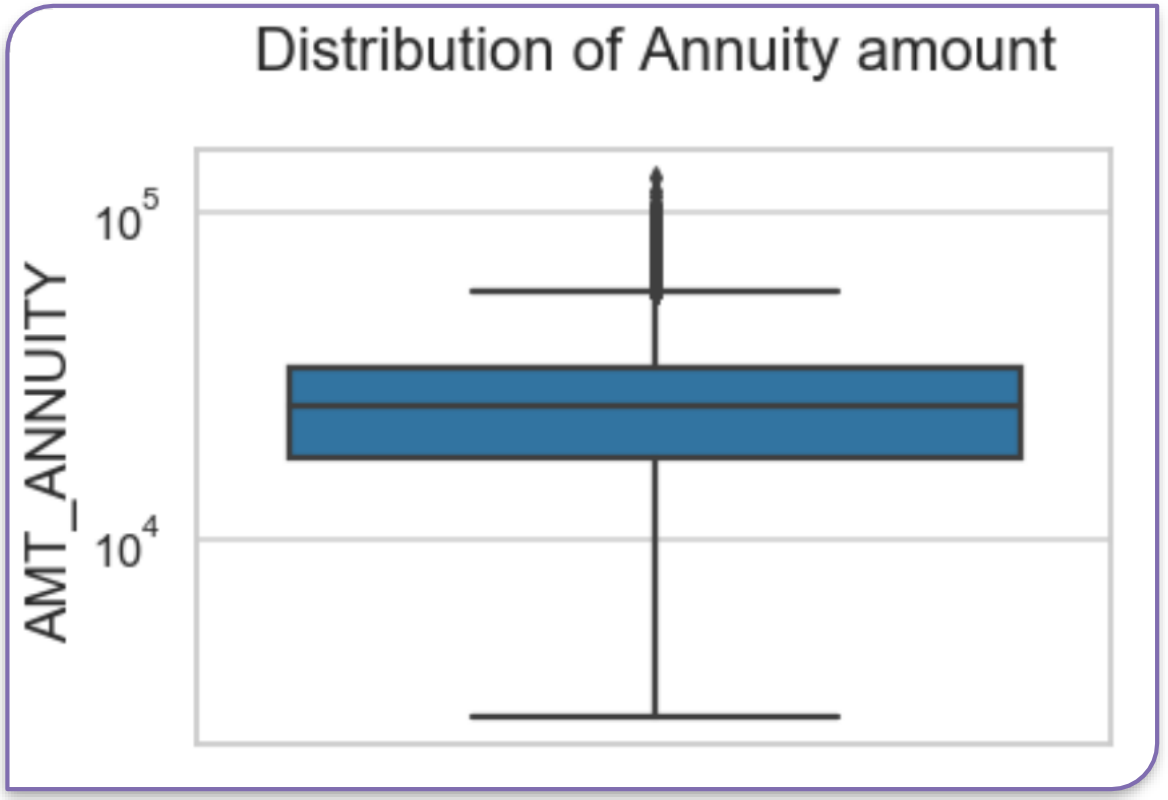
- Some outliers are noticed in income amount.
- The third quartile is very slim for income amount.



Boxplot for annuity amount

Conclusion

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

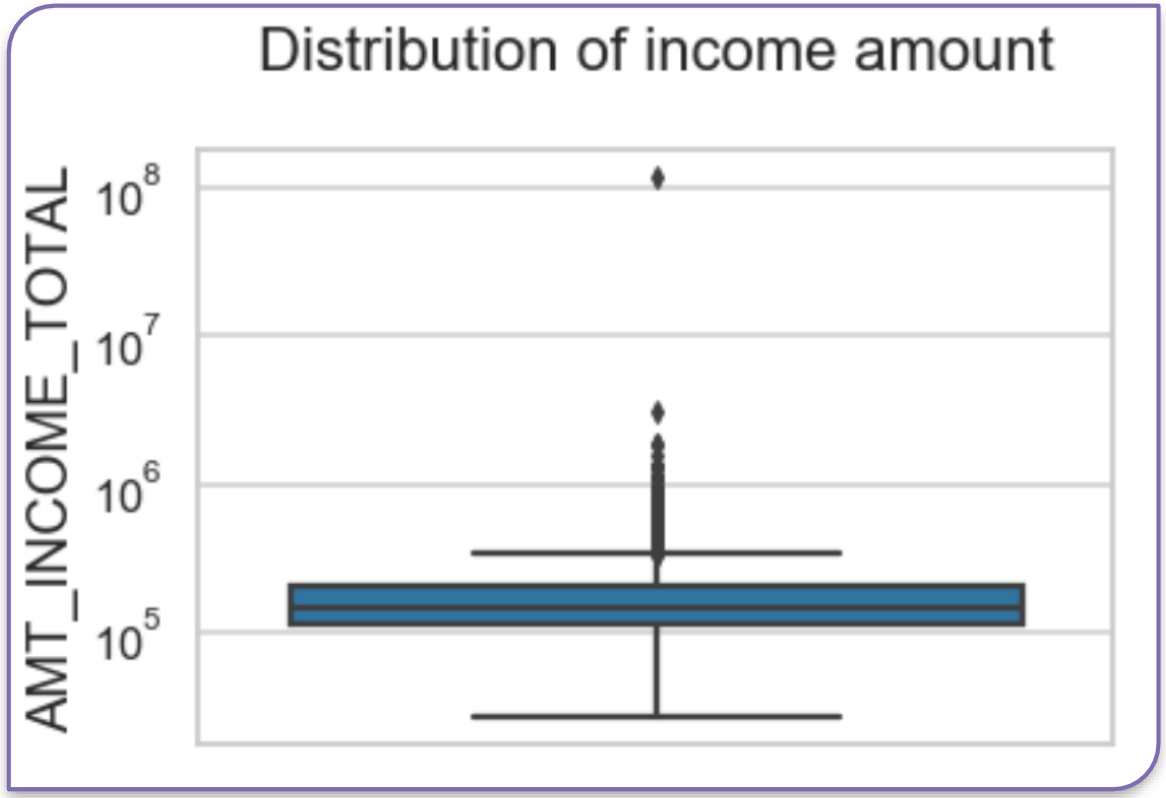


Categorical Univariate analysis of variables for Target = 1

Boxplot for income amount

Conclusion

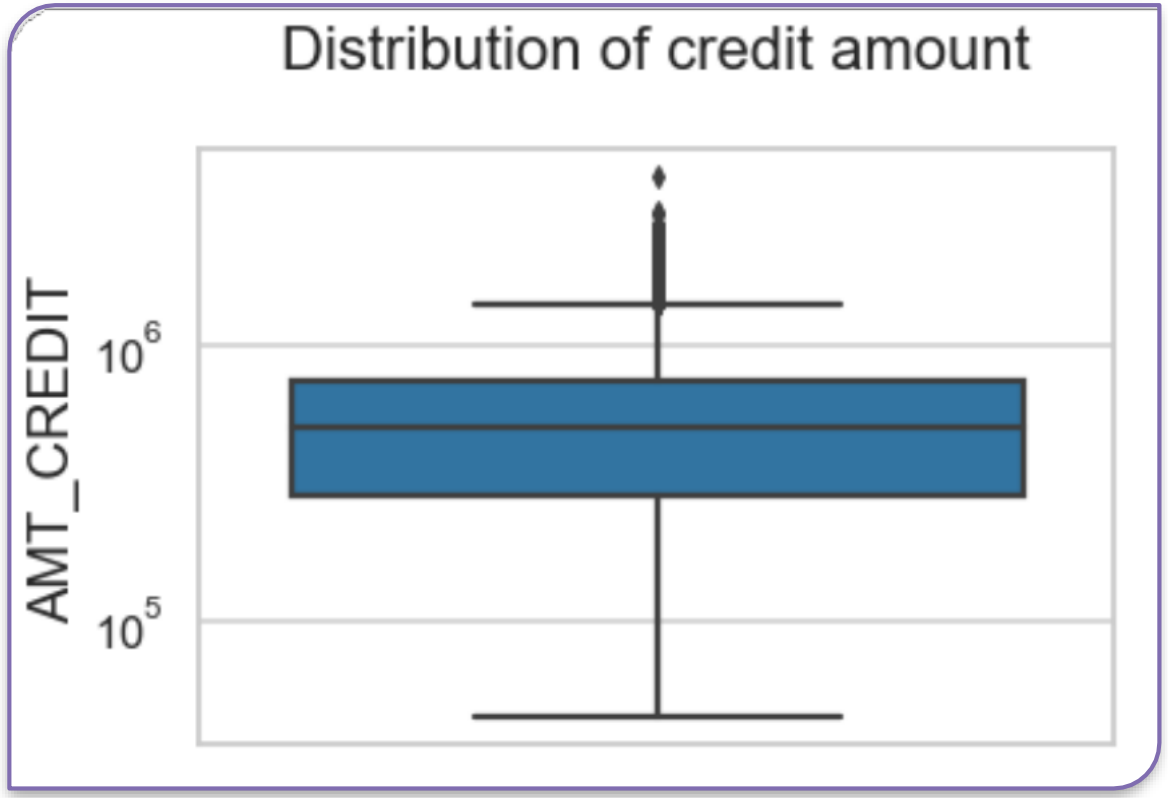
- Some outliers are noticed in income amount.
- The third quartile is very slim for income amount.
- Most of the clients of income are present in first quartile.



Boxplot for credit amount

Conclusion

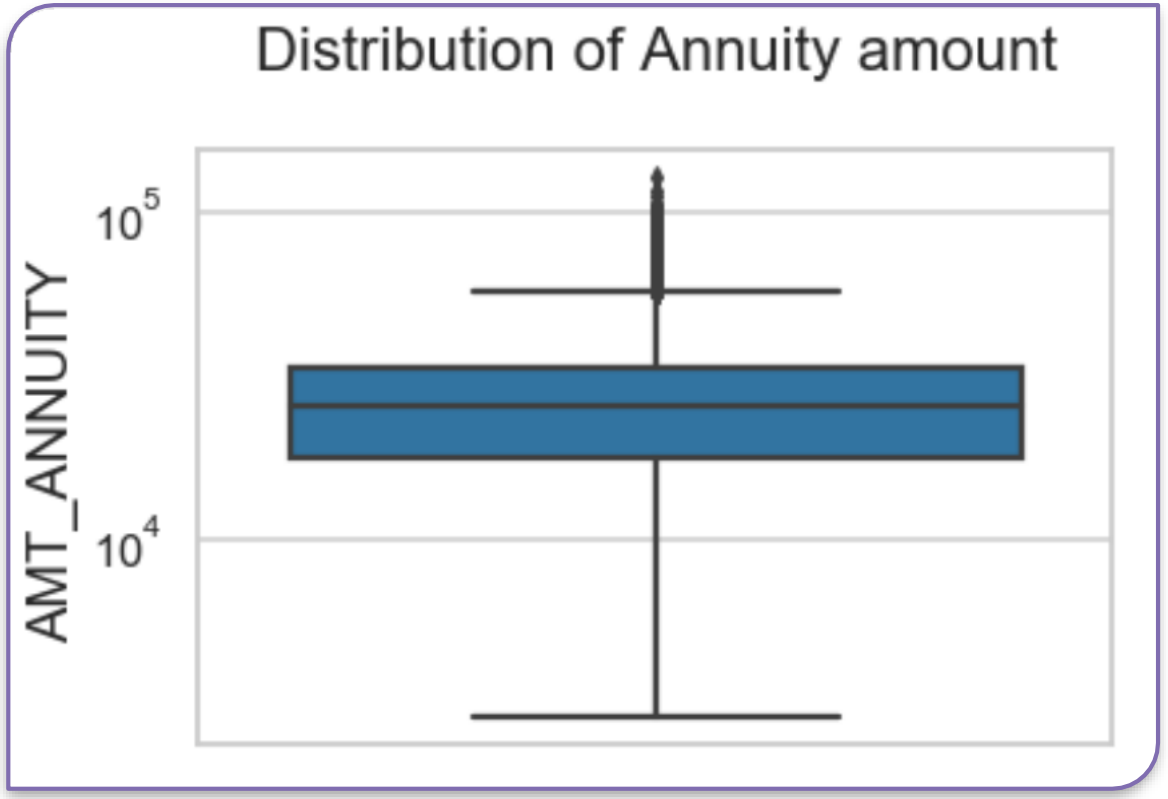
- Some outliers are noticed in credit amount.
- The first quartile is bigger than third quartile for credit amount which means most of the credits of clients are present in the first quartile.



Boxplot for annuity amount

Conclusion

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

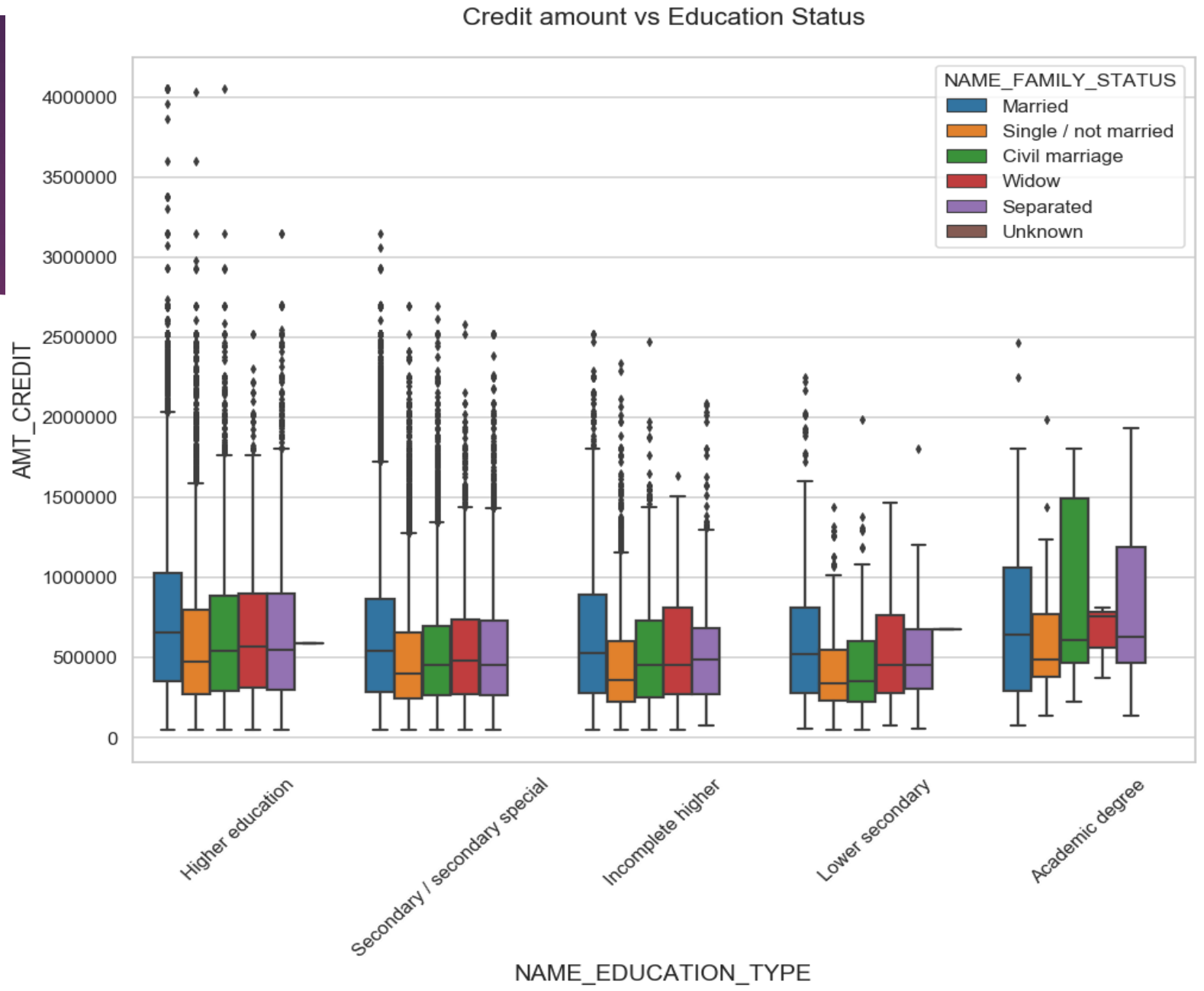


Bivariate analysis for Target = 0

Credit amount vs Education Status

Conclusion

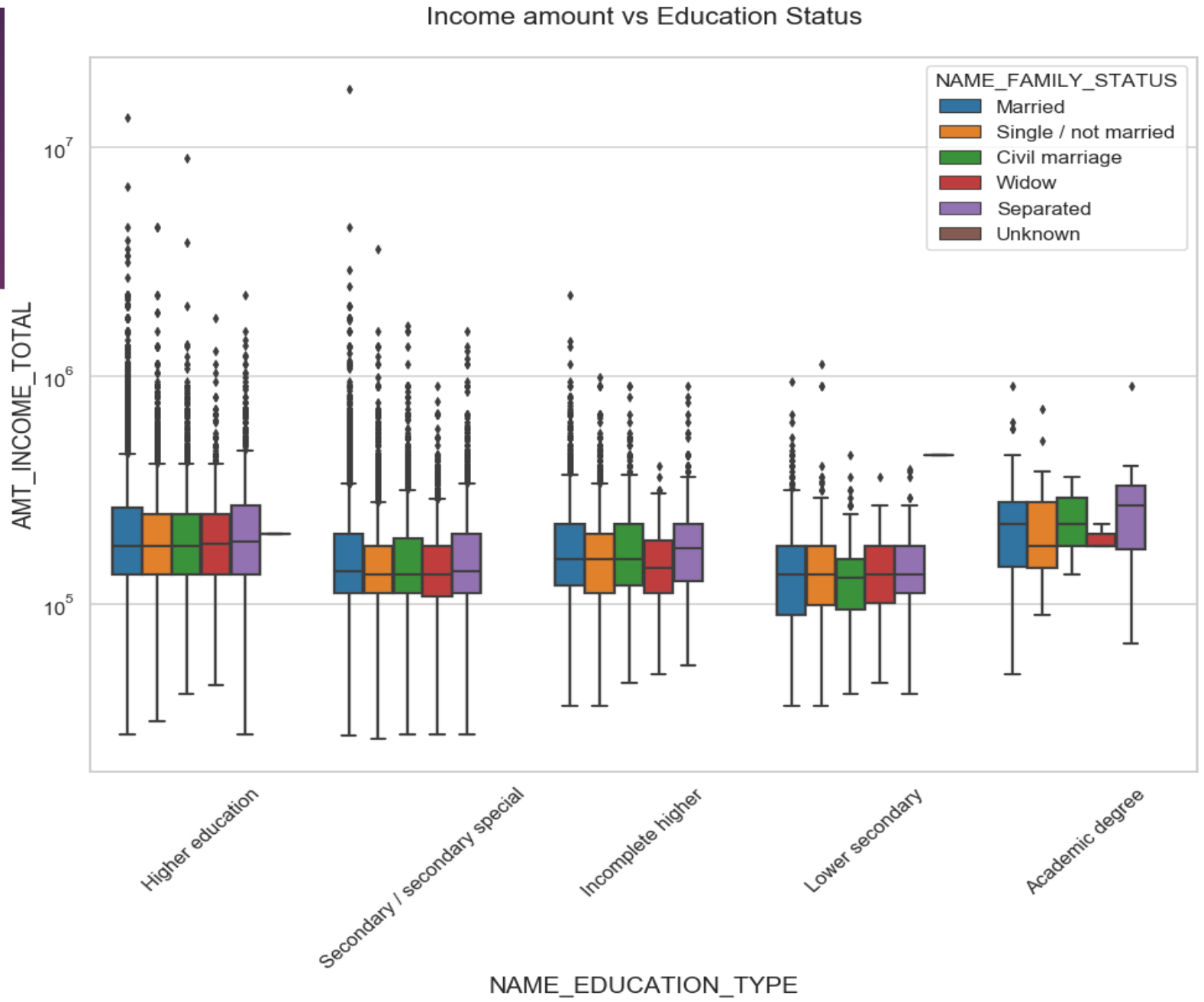
- From the above box plot we can conclude that people with Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others.
- Also, higher education of people with family status of 'marriage', 'single' and 'civil marriage' has more outliers.
- Civil marriage for Academic degree is having most of the credits in the third quartile.



Income amount vs Education Status

Conclusion

- From above boxplot for Education type 'Higher education' the income amount is mostly equal with family status. It does contain many outliers.
- Less outlier exist for Academic degree but their income amount is little higher that of Higher education.
- People with Lower secondary education and civil marriage have less income than others.

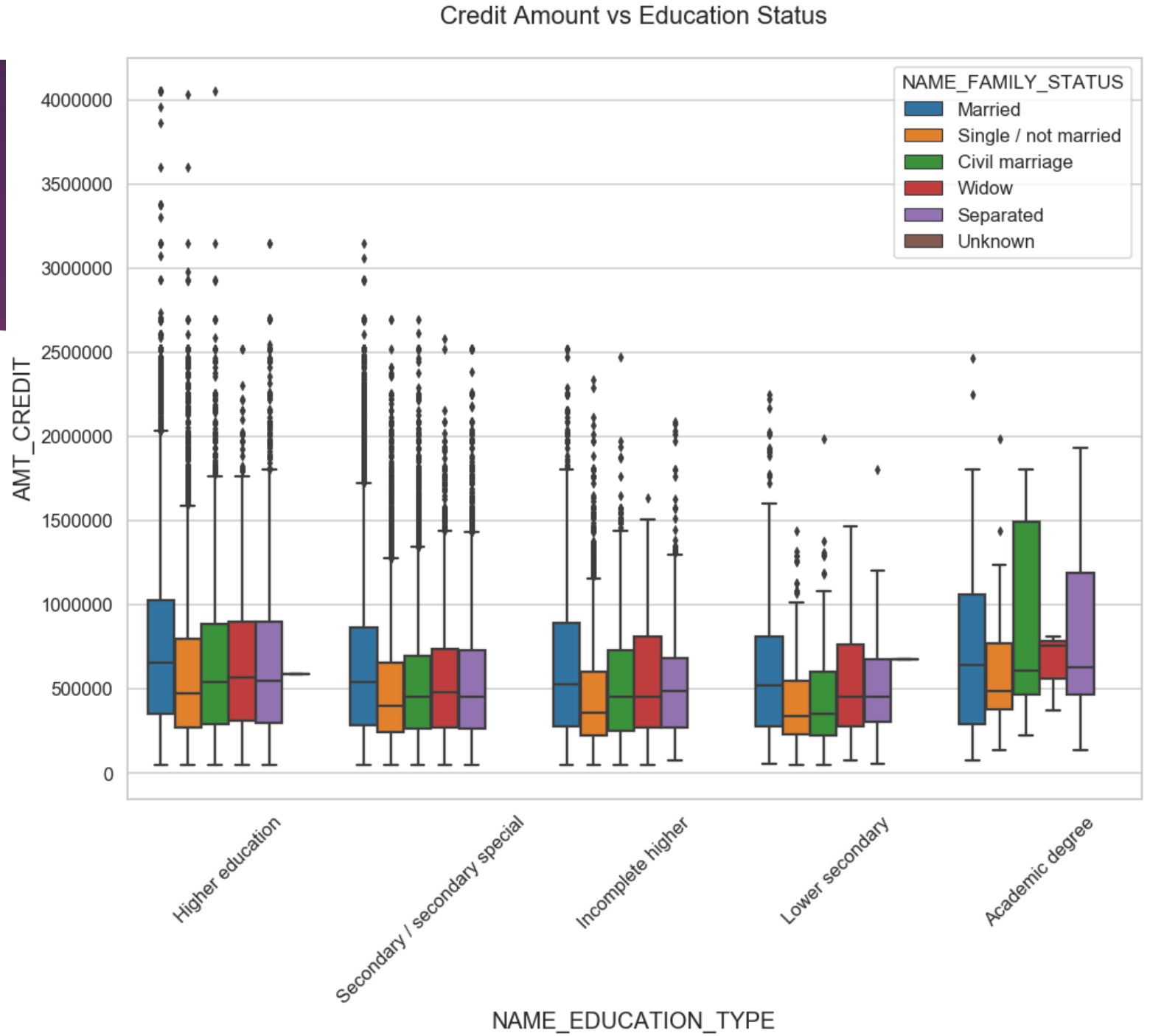


Bivariate analysis for Target = 1

Credit amount vs Education Status

Conclusion

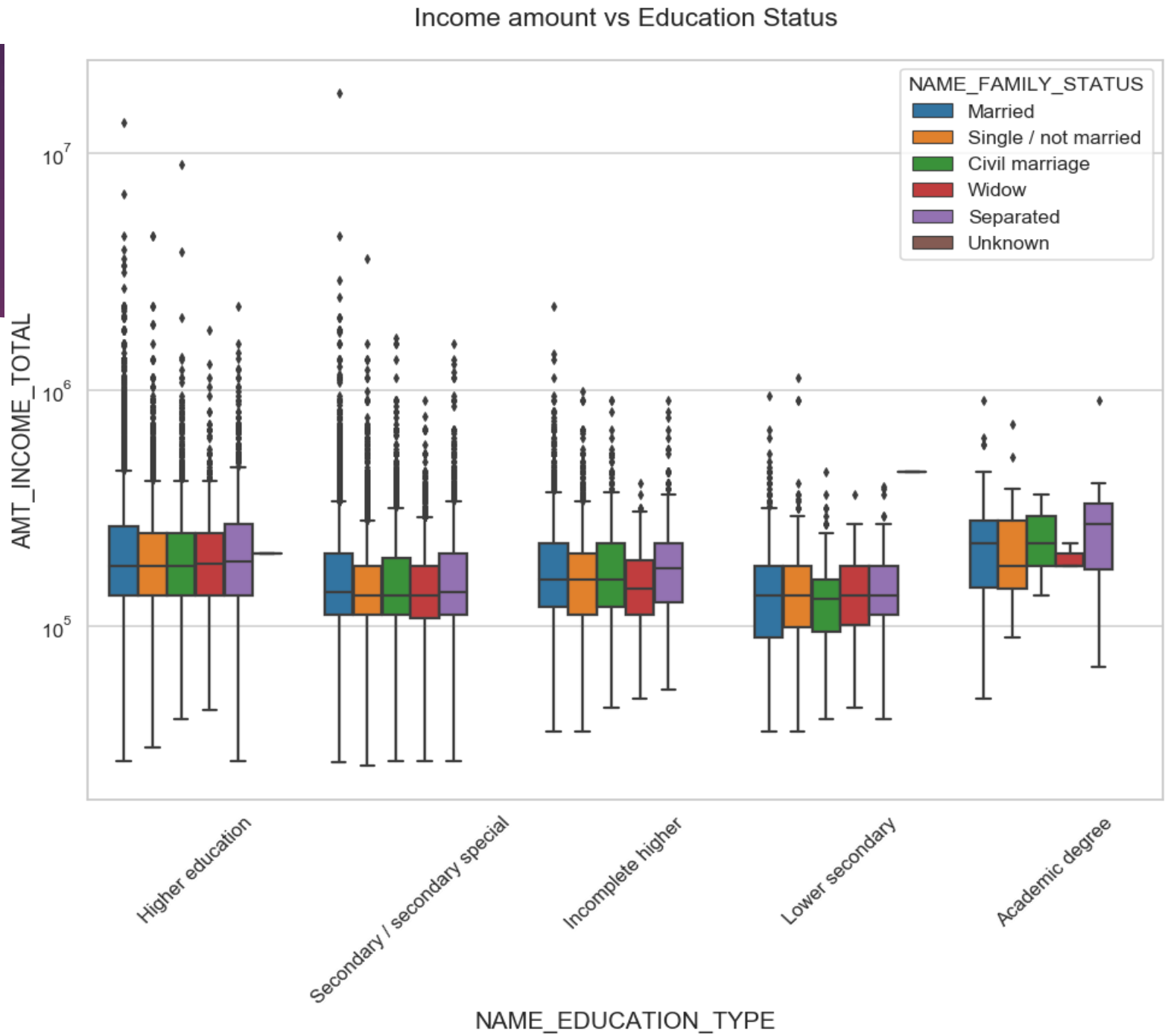
- Quite similar to Target = 0
- From the above box plot we can say that people with Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others. Most of the outliers are from Education type 'Higher education' and 'Secondary'.
- Civil marriage for Academic degree is having most of the credits in the third quartile.



Income amount vs Education Status

Conclusion

- From above boxplot for Education type 'Higher education' the income amount is mostly equal for almost all family status. Less outliers exist for Academic degree and their income amount is a bit higher that of Higher education.
- Lower secondary educated people are having less income amount than others.

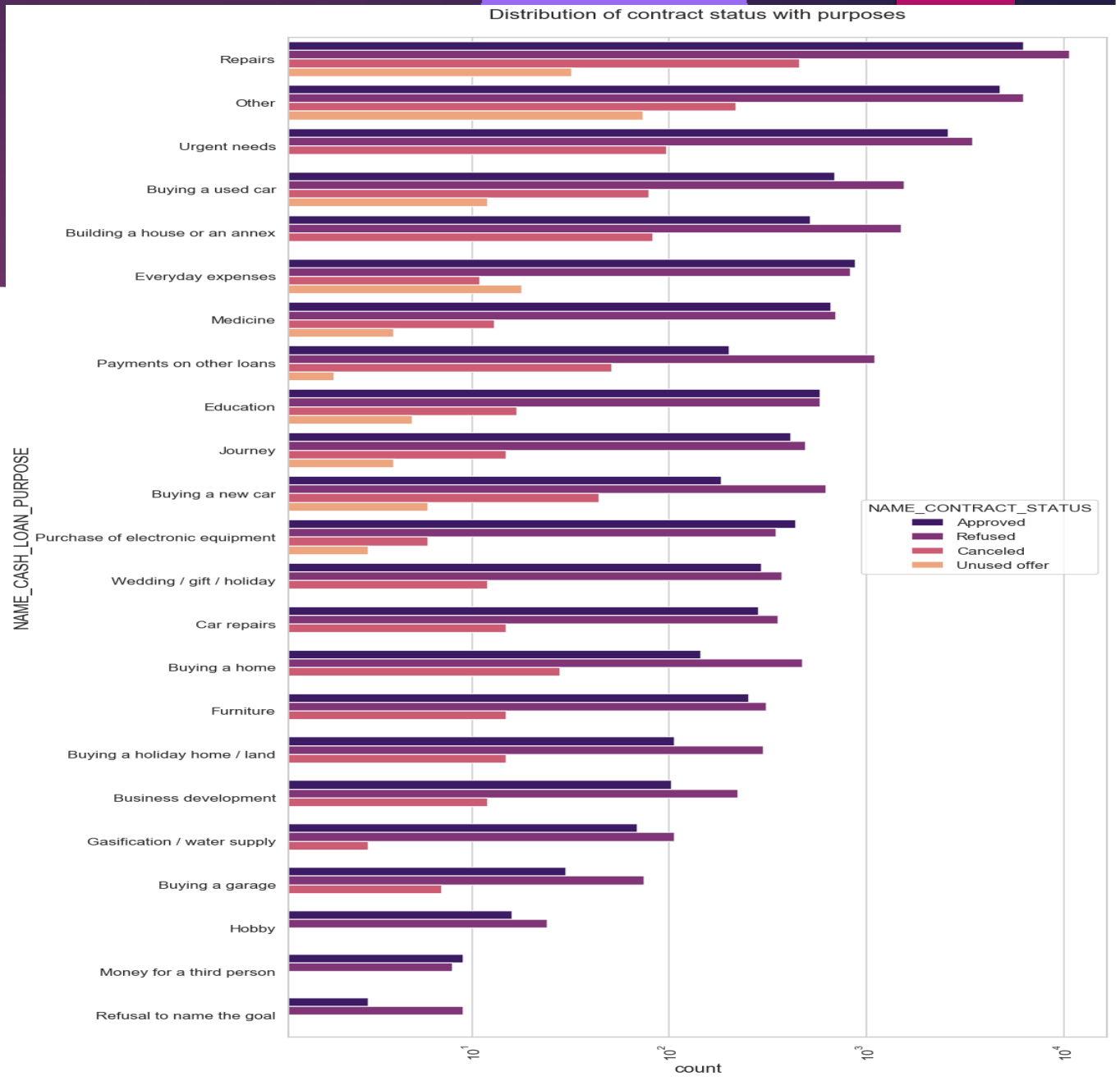


Univariate analysis (after merging data from previous application)

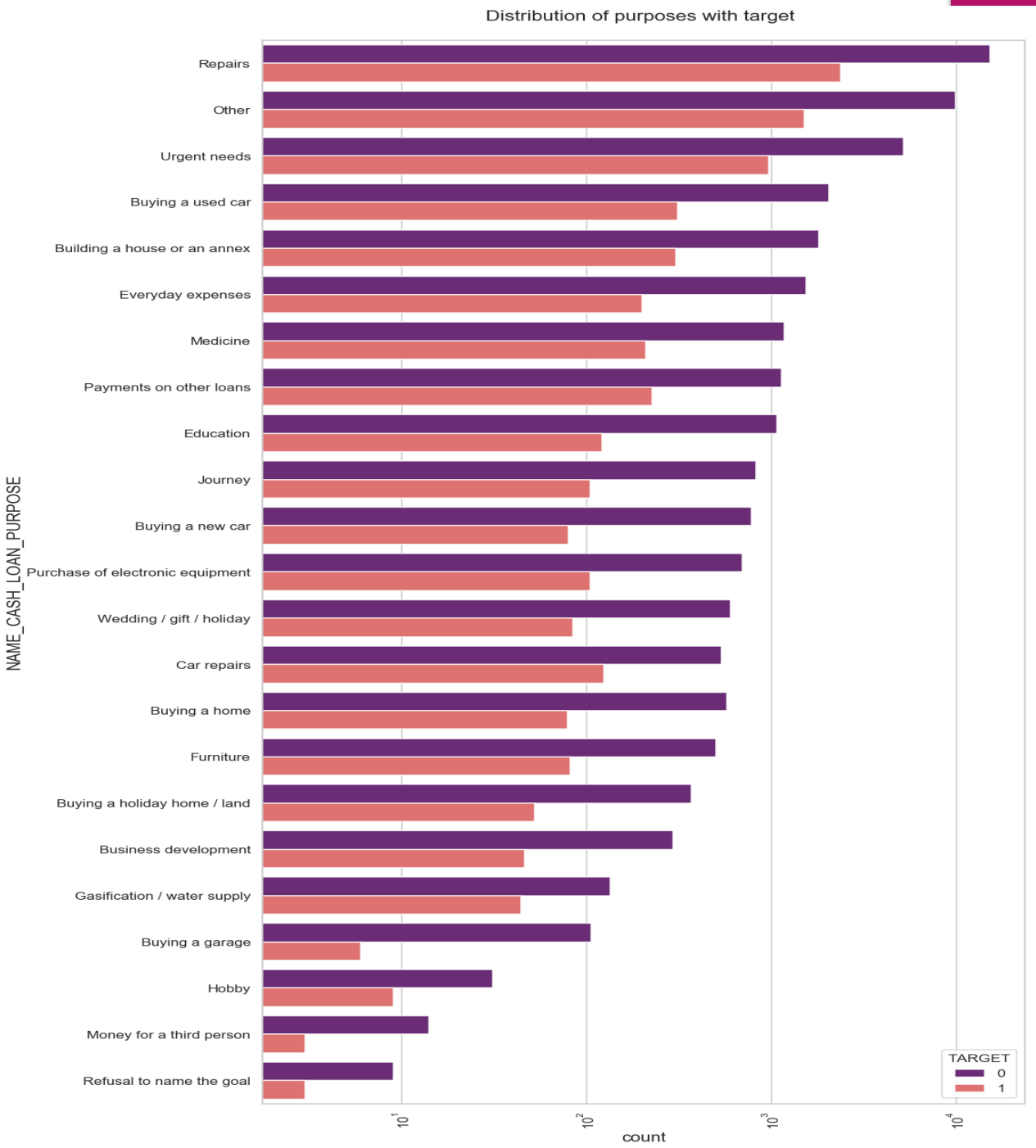
Distribution of contract status with purposes

Conclusion

- 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



Conclusion

- Loan purposes with 'Repairs' are facing most difficulties in payment on time (among all purposes).
- There are few places where loan payment is significantly 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.

Bivariate analysis

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

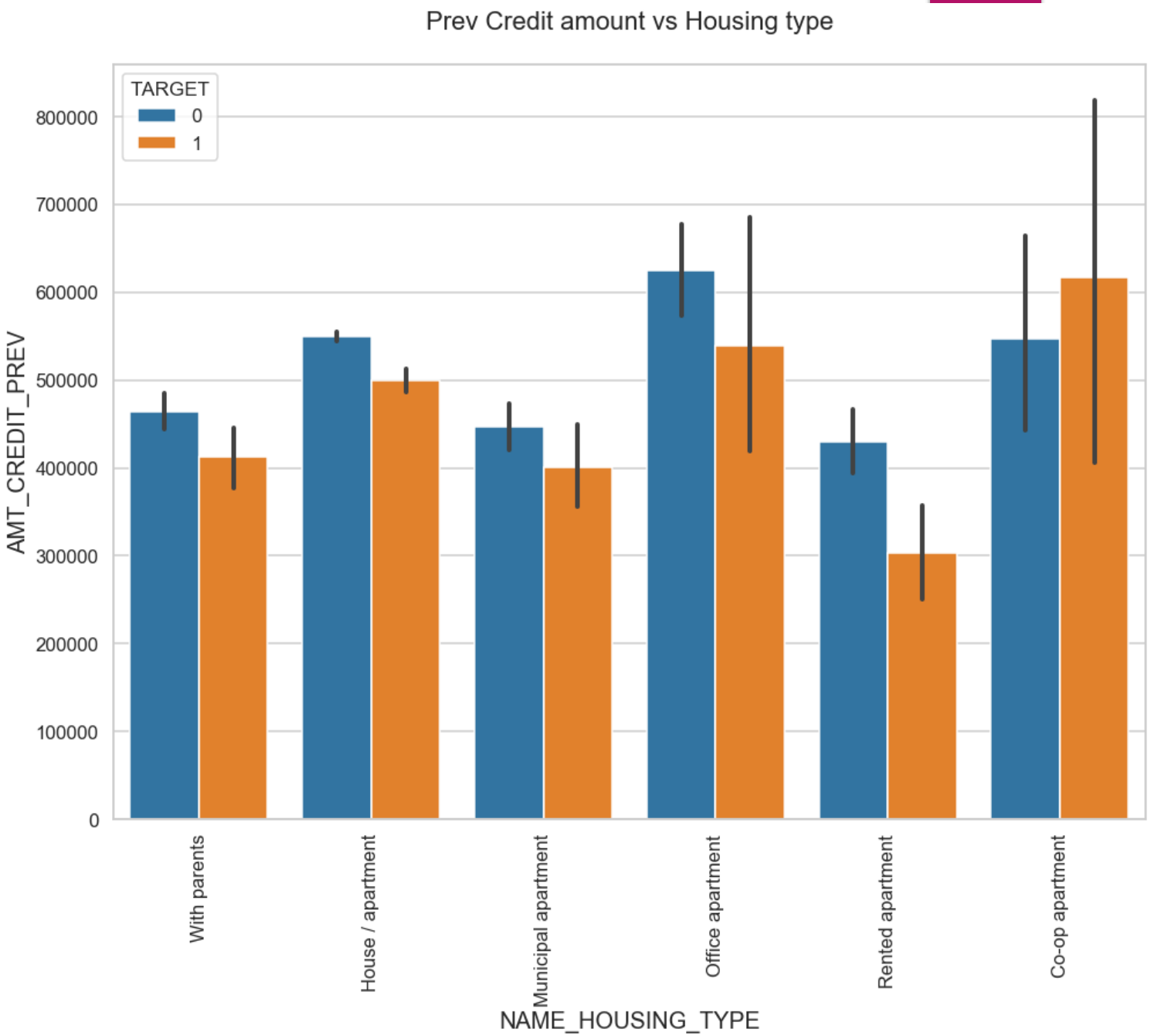
Previous Credit amount vs Loan Purpose

- **Conclusion**
- The credit amount of Loan purposes like 'Buying a home','Buying a land','Buying a new car' and'Building a house' is higher.
- State servants have a significant amount of credit applied
- Money for third person or a Hobby is having less credits applied for.

Prev. Credit amount vs Housing type

Conclusion

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



Final Remarks:

- Focus shall be on people with income type ‘Student’ ,’pensioner’ and ‘Businessman’ with housing type other than ‘Co-op apartment’ for successful payments.
- People with housing type ‘With parents’ can be targeted as they are having least number of unsuccessful payments.

- Banks should focus less on income type ‘Working’ as they are having the greatest number of unsuccessful payments.
- Also, with loan purpose ‘Repair’ is having higher number of unsuccessful payments on time.

Thank you