# CREDIT EDA CASE STUDY

BY SHANTANU SINGH & ARINDAM

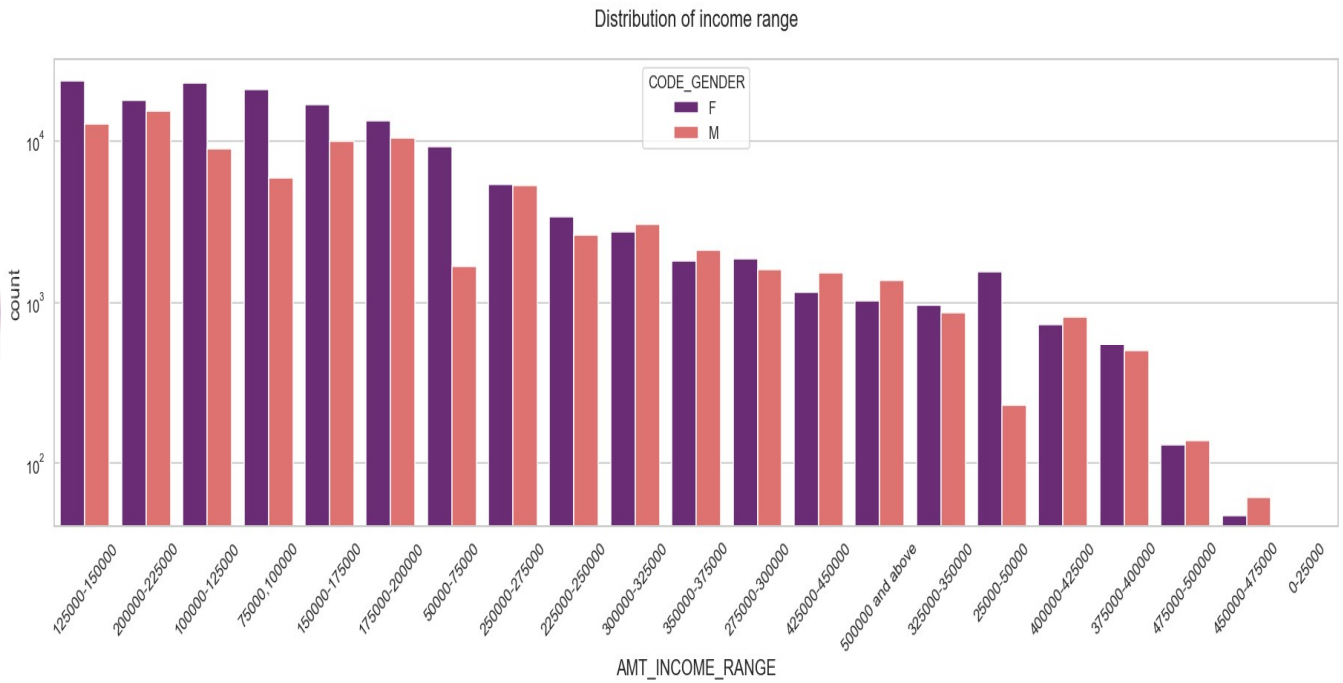
# Categorical Univariate analysis for Target 0

**[Applicants with no payment difficulty]**

# Distribution of Income

Conclusions

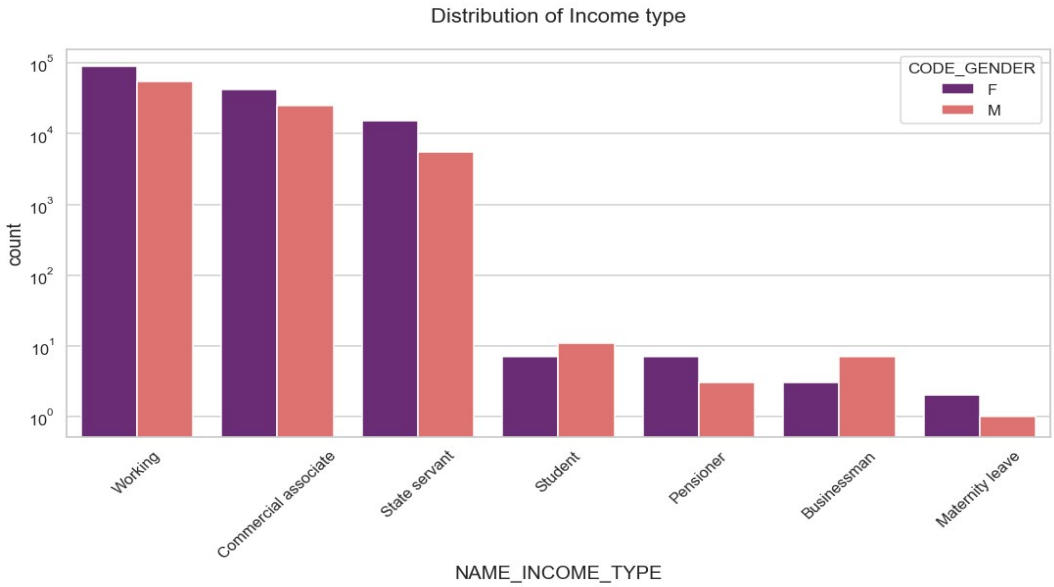
- More Females than Males.
- Income range from 1,00,000 to 2,00,000 has more applicants.
- Females have more applicants than men for that range.
- Very few applicants have income 4,00,000 and above.



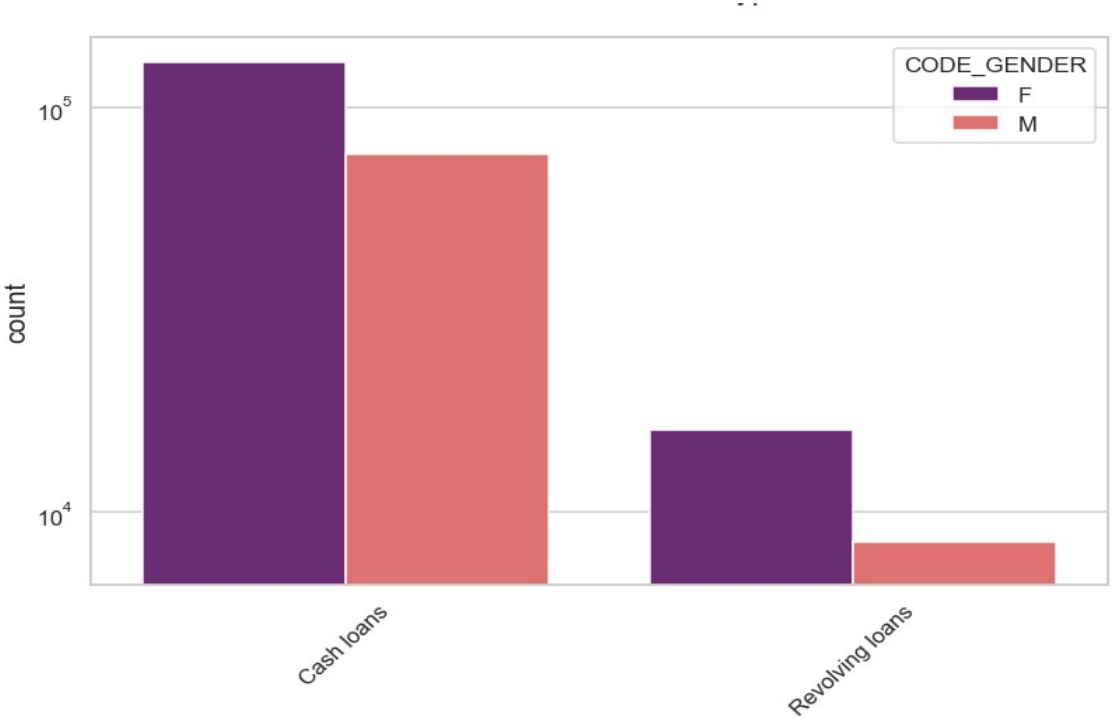
# Distribution of Income Type

Conclusions

- For income type ‘working’, ’commercial associate’, and ‘State Servant’ the number of applicants are higher than others.
- For those income type Females have more applications than Males.
- Less applicants for income type ‘student’ ,’pensioner’, ‘Businessman’ and ‘Maternity leave’.



# Distribution for contract type



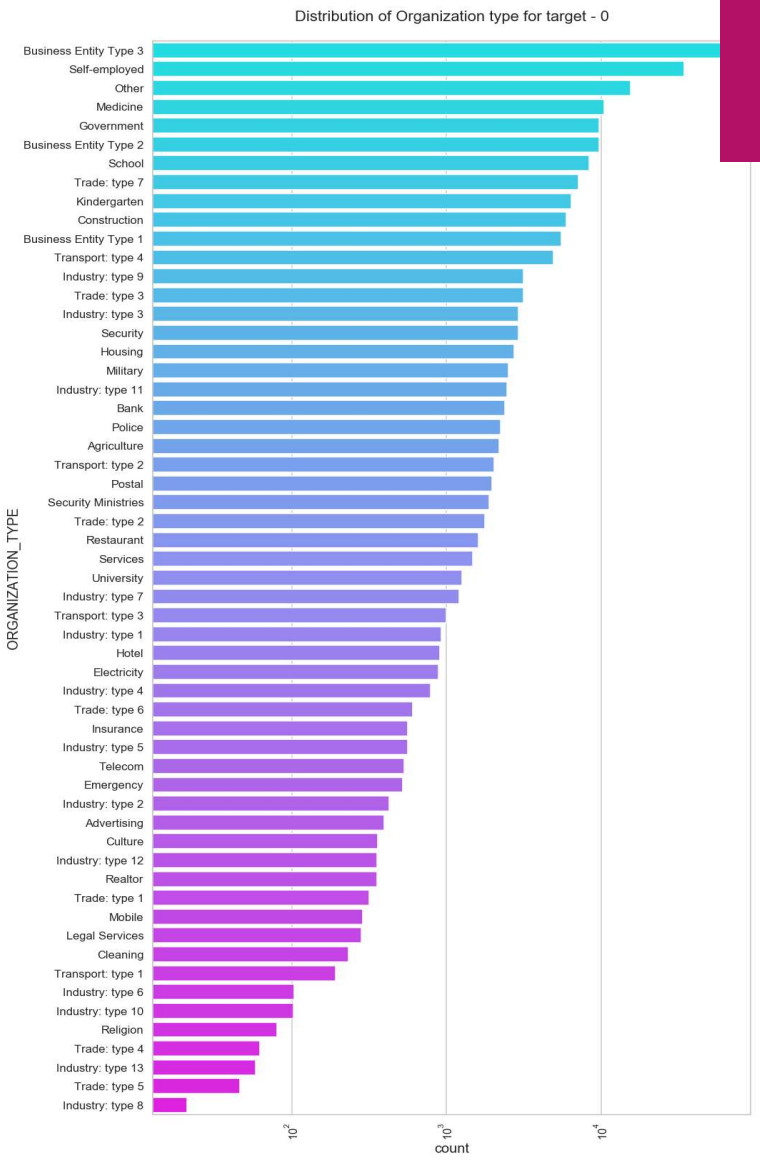

Conclusions

- More ‘cash loans’ than ‘Revolving loans’
- More Female applicants

# Distribution of organization type

Conclusions

- Most clients that have applied for loans are from 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.



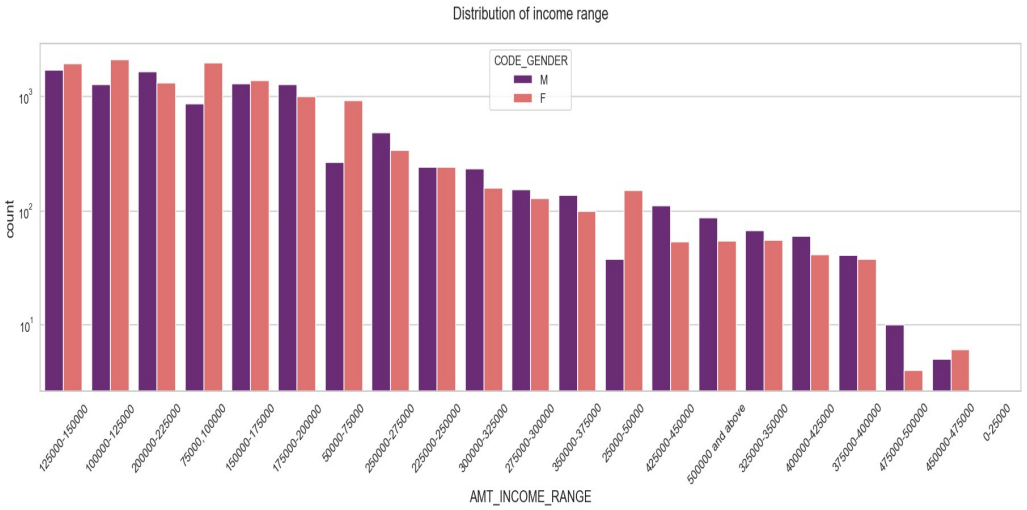
# Categorical Univariate analysis for Target 1

**[Applicants with payment difficulty]**

# Distribution of Income range

Conclusions

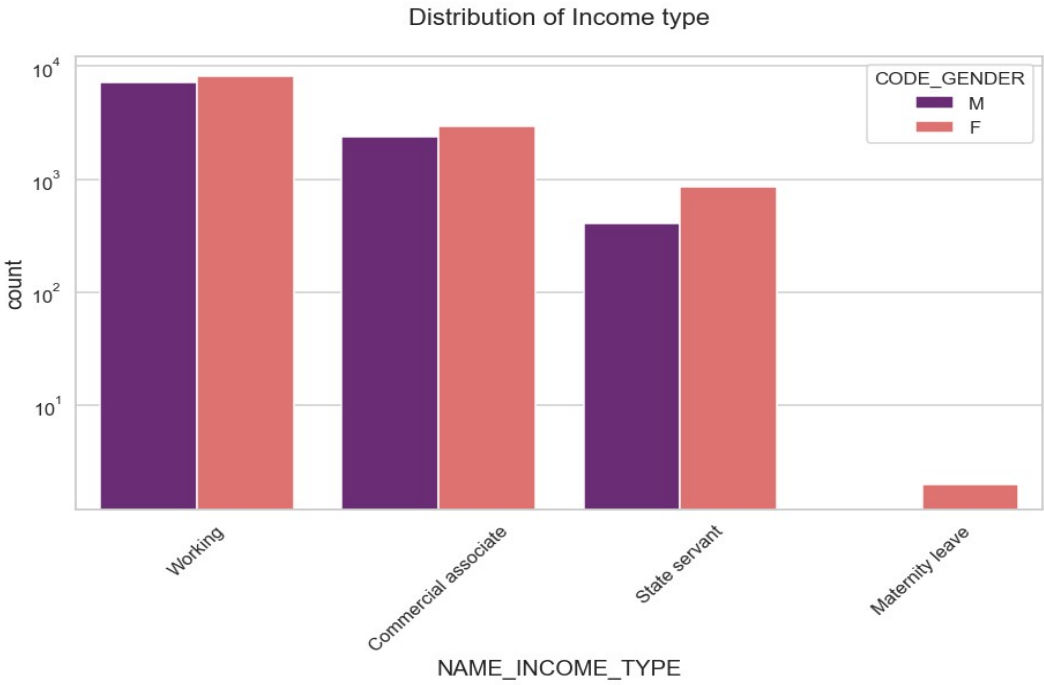
- More males than female.
- Income range 1,00,000 to 2,00,000 has more applicants.
- More males than females in that range.
- Very less applicants for income range 4,00,000 and above.



# Distribution of Income type

Conclusions

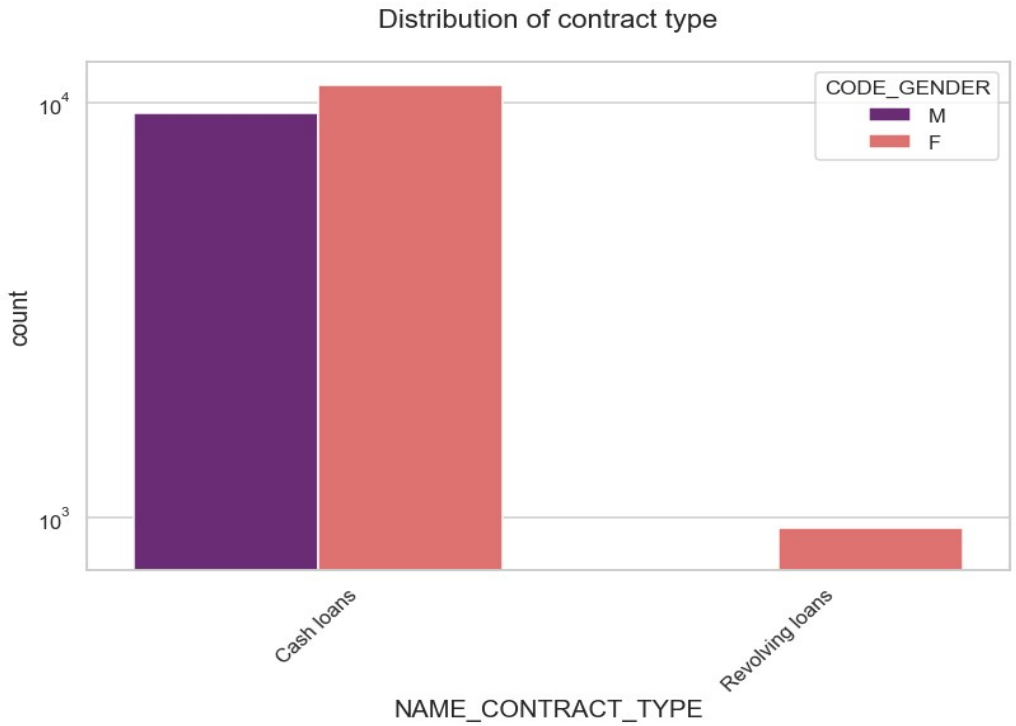
- For income type ‘working’, ’commercial associate’, and ‘State Servant’ the number of applicants are higher than other.
- Females have more number of applicants than male.
- Less number of credits for income type ‘Maternity leave’.
- For type 1: There is no income type for ‘student’ , ’pensioner’ and ‘Businessman’ which means they don’t do any late payments.



# Distribution for contract type

Conclusions

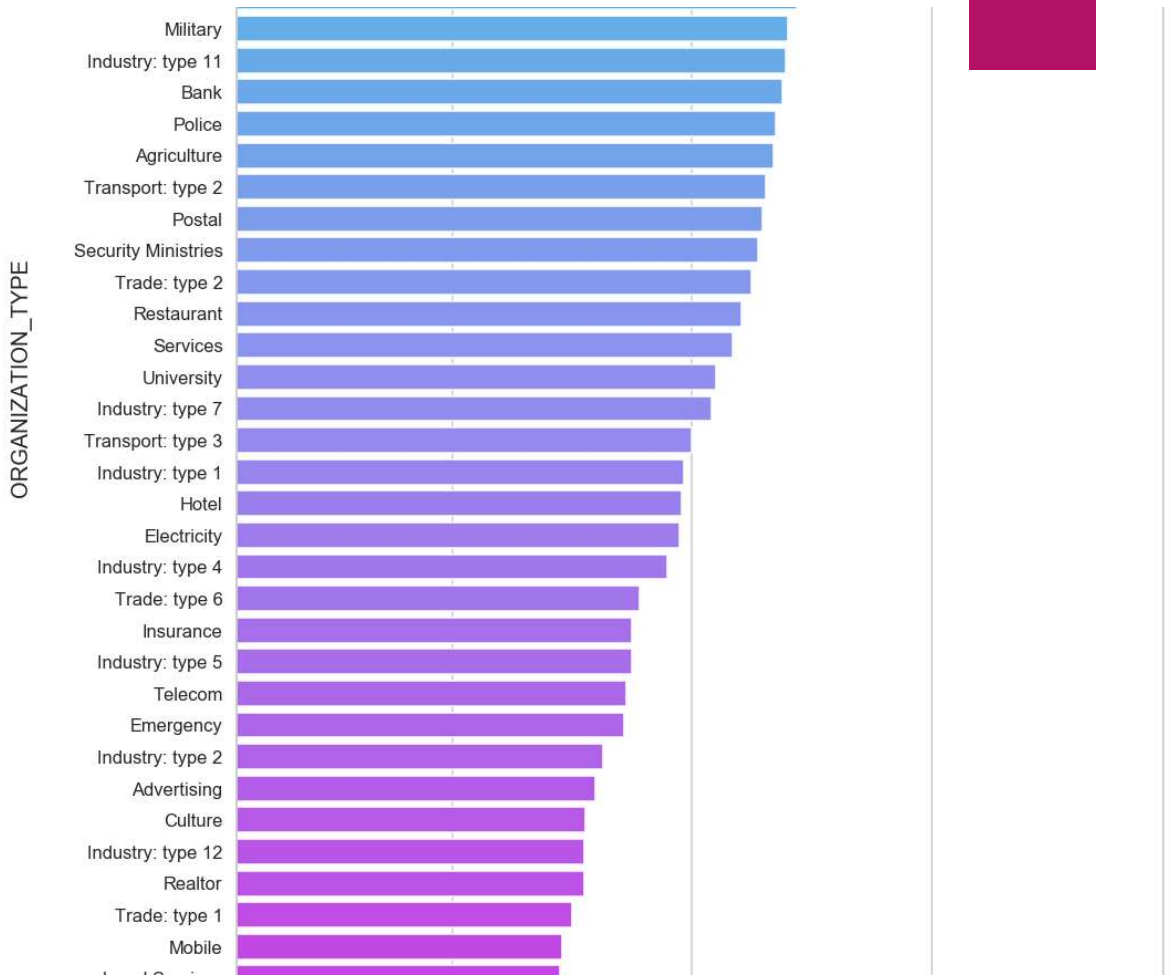
- For contract type ‘cash loans’ is having higher number of applicants than ‘Revolving loans’ contract type.
- Females is leading for applying credits.
- For type 1 : there is only Female Revolving loans.



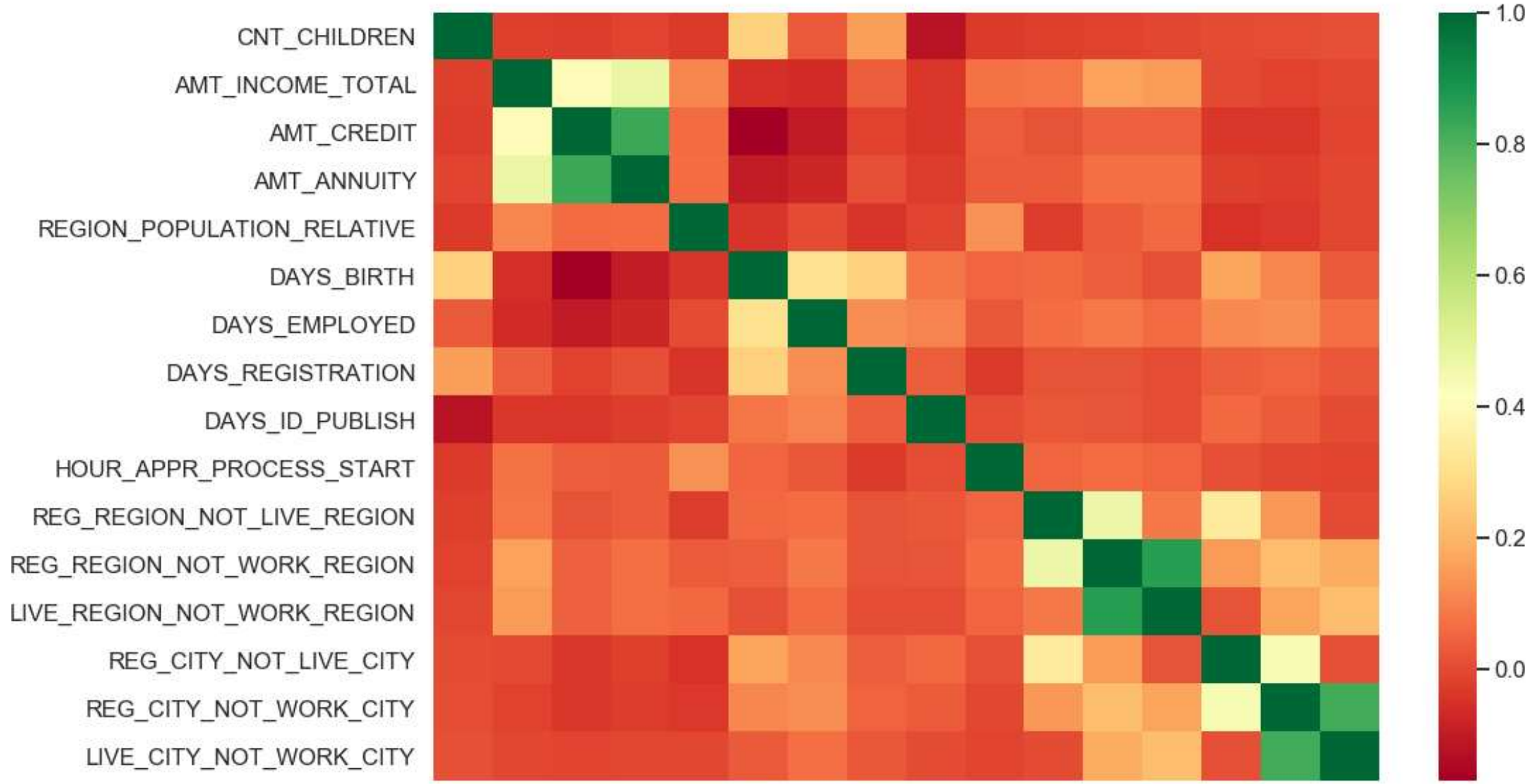
# Distribution of organization type

Conclusions

- 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 as type 0 in distribution of organization type.

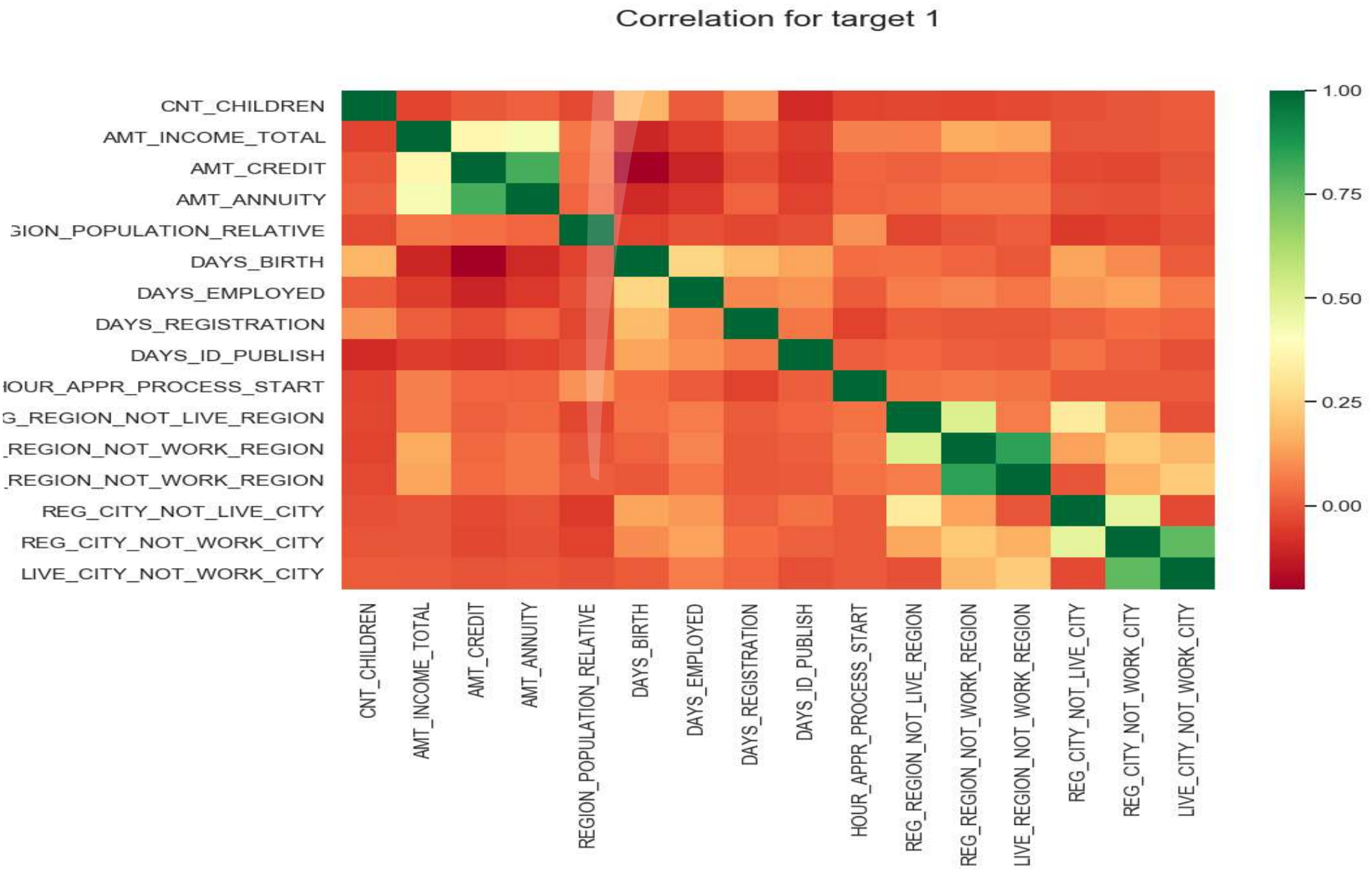


# Correlation of Target 0


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

| |
|---|
| Loan amount is inversely proportional to the date of birth - loan amount is higher for lesser age. |
| Loan amount is inversely proportional to the number of children client have – loan amount is higher for people with less children |
| Income is inversely proportional to the number of children client have - more income if people have less children. |
| Loan amount is higher for densely populated area. |
| Income is also higher in densely populated area. |


Correlation for target 1
CNT_CHILDREN
AMT_INCOME_TOTAL
AMT_CREDIT
AMT_ANNUITY
GION_POPULATION_RELATIVE
DAYS_BIRTH
DAYS_EMPLOYED
DAYS_REGISTRATION
DAYS_ID_PUBLISH
HOUR_APPR_PROCESS_START
G_REGION_NOT_LIVE_REGION
_REGION_NOT_WORK_REGION
_REGION_NOT_WORK_REGION
REG_CITY_NOT_LIVE_CITY
REG_CITY_NOT_WORK_CITY
LIVE_CITY_NOT_WORK_CITY
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 type 1

This heat map for Target 1 is also having quite a same observation just like Target 0. Other Conclusions.

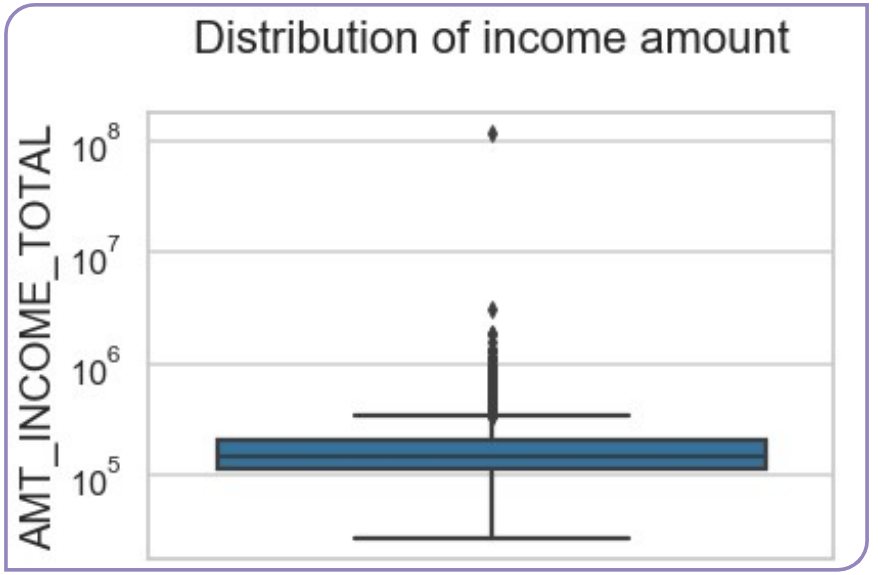
- The client's permanent address does not match contact address are having less children and vice-versa
- The client's permanent address does not match work address are having less children and vice-versa

# Categorical Univariate analysis for Variables with Target 0

# Boxplot for income amount

Conclusions

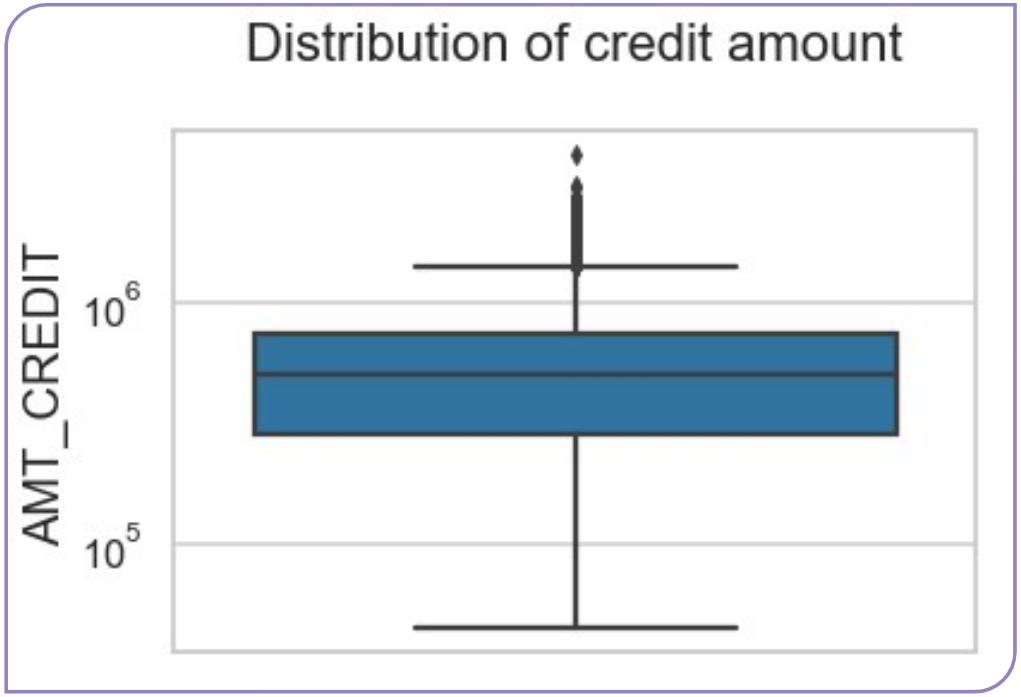
- Some outliers are noticed in income amount.
- The third quartiles is very slim for income amount.



# Boxplot for loan amount

Conclusions

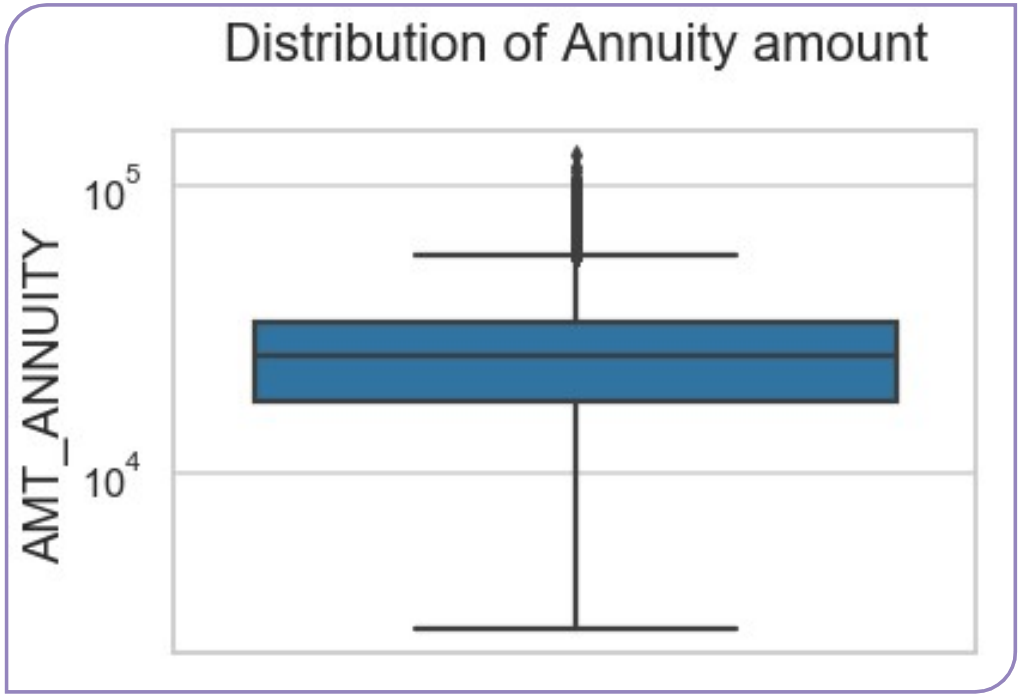
- Some outliers are noticed in loan amount.
- The first quartile is bigger than third quartile for loan amount which means most of the loan amounts are present in the first quartile.



# Boxplot for Annuity amount

Conclusions

- 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.

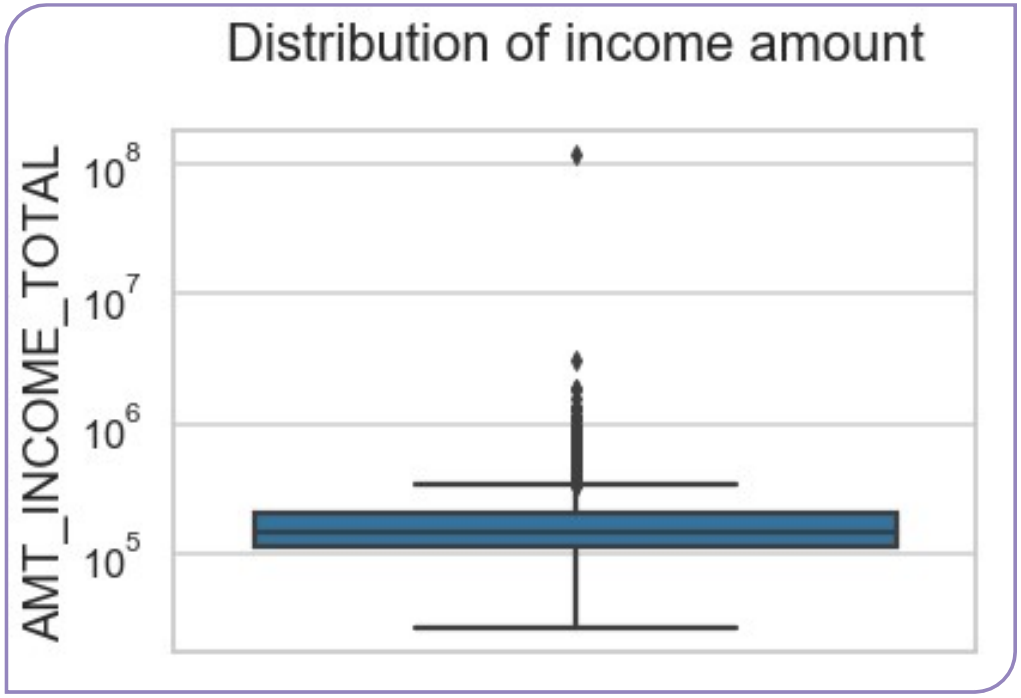


# Categorical Univariate analysis for Variables with Target 1

# Boxplot for Income amount

Conclusions

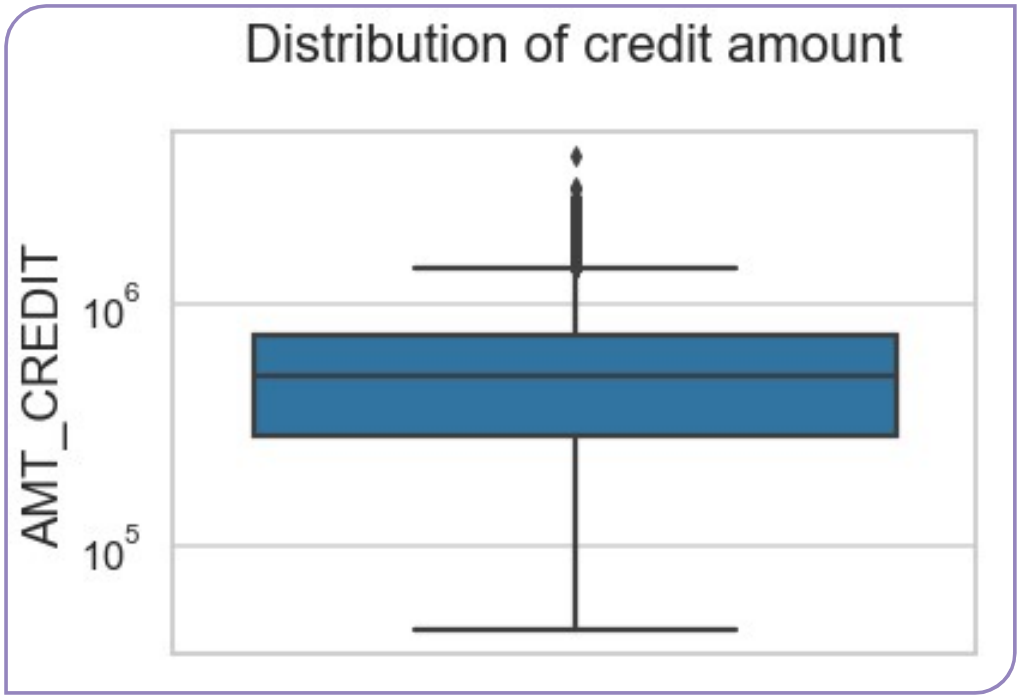
- Some outliers are noticed in income amount.
- The third quartiles is very slim for income amount.
- Most of the clients of income are present in first quartile.



# Boxplot for loan amount

Conclusions

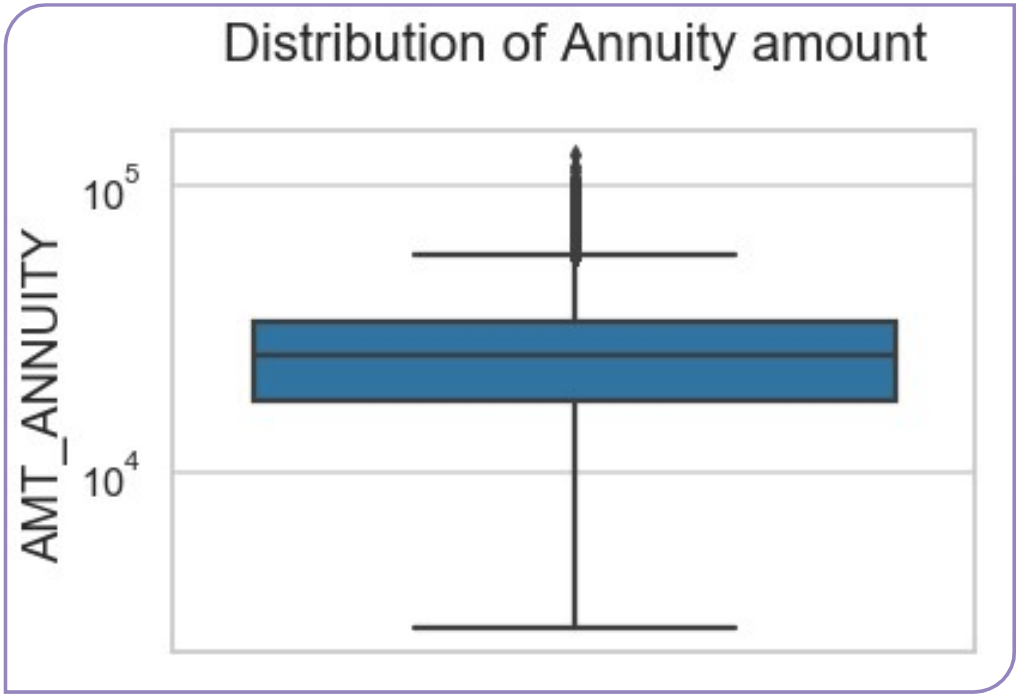
- Some outliers are noticed in credit amount.
- The first quartile is bigger than third quartile for credit amount which means most of the credits of clients are present in the first quartile.



# Boxplot for annuity amount

Conclusions

- 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.

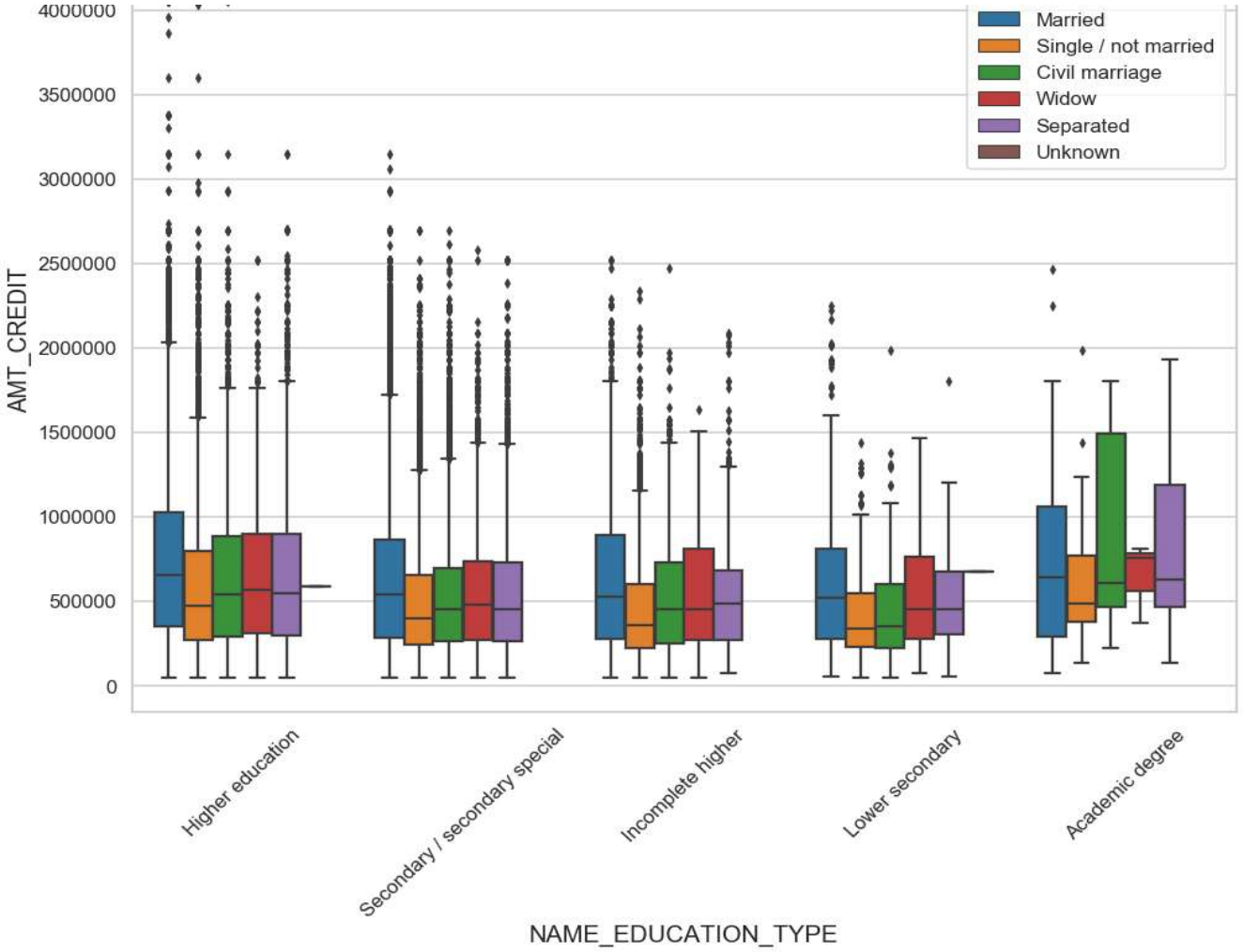


# Bivariate Analysis for Type 0

# Credit amount vs Education Status

Conclusions

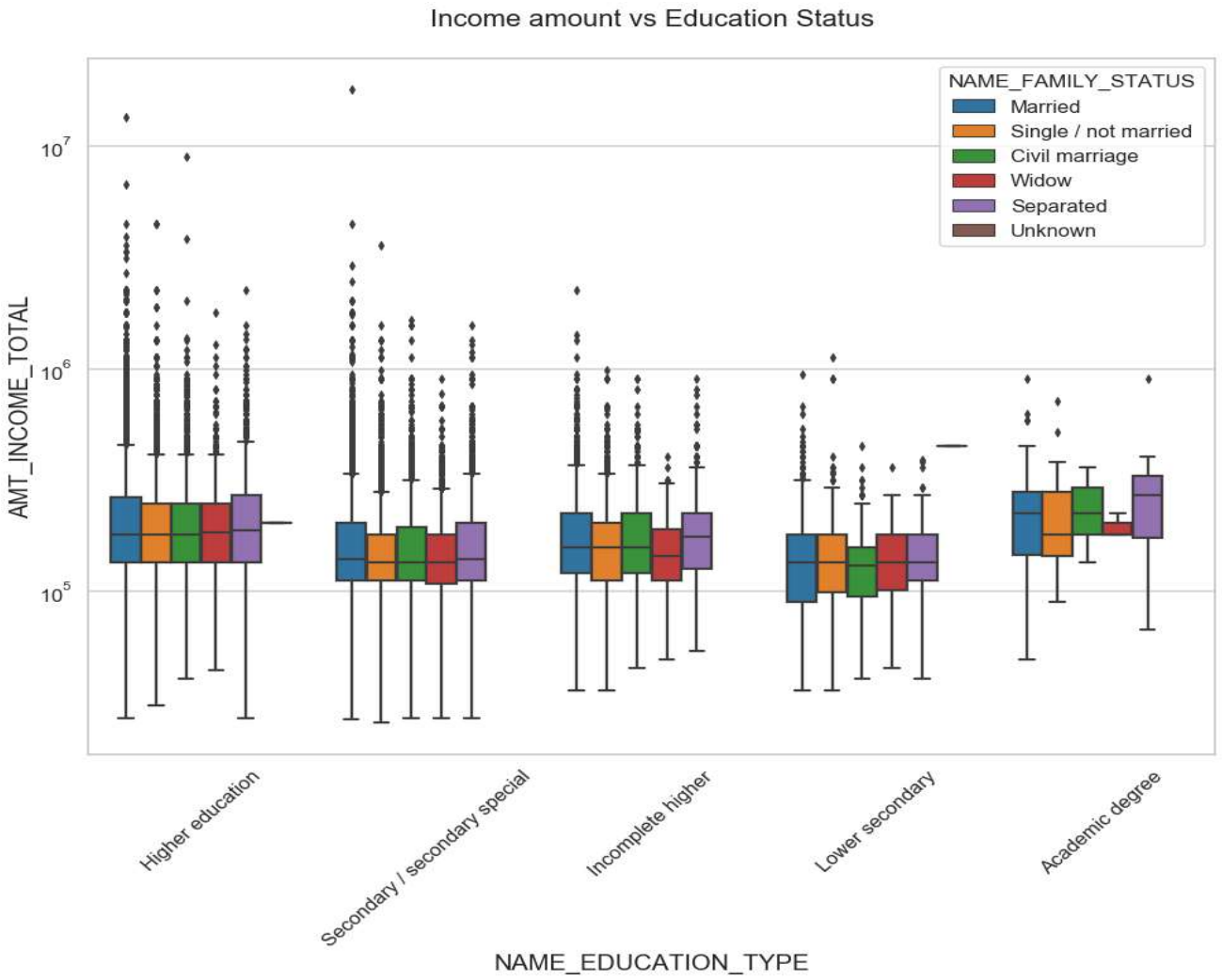
- Family status of 'civil marriage', 'marriage' and 'separated' with Academic degree education have higher number of credits than others.
- Higher education of family status of 'marriage', 'single' and 'civil marriage' are having more outliers.
- Civil marriage for Academic degree is having most of the credits in the third quartile.



# Income amount vs Education Status

Conclusions

- For Education type 'Higher education' the income amount mean is mostly equal with family status. It does contain many outliers.
- Less outlier are having for Academic degree but they are having the income amount is little higher that Higher education.
- Lower secondary of civil marriage family status are have less income amount than others.

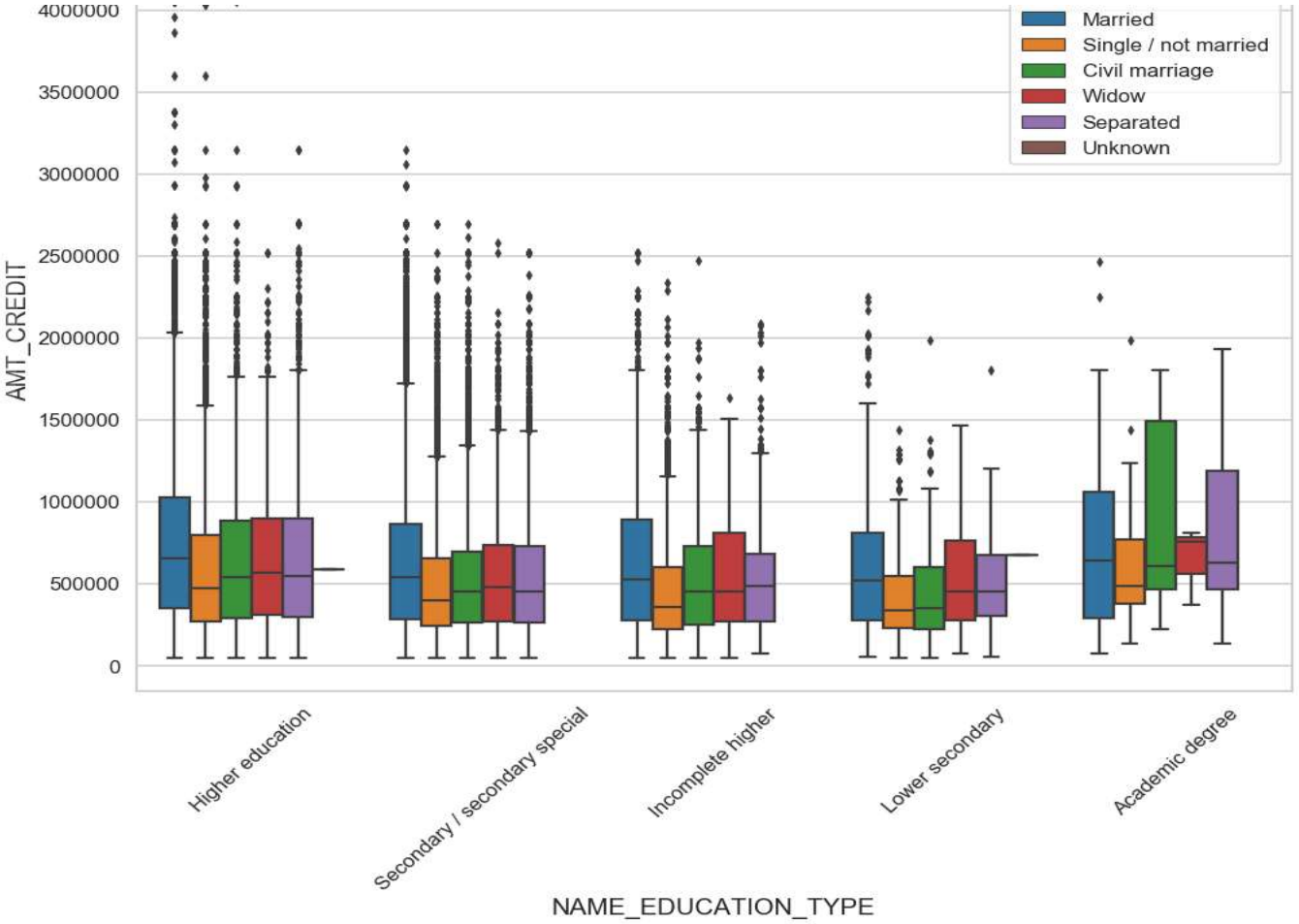


# Bivariate Analysis for Type 1

# Credit amount vs Education Status

Conclusions

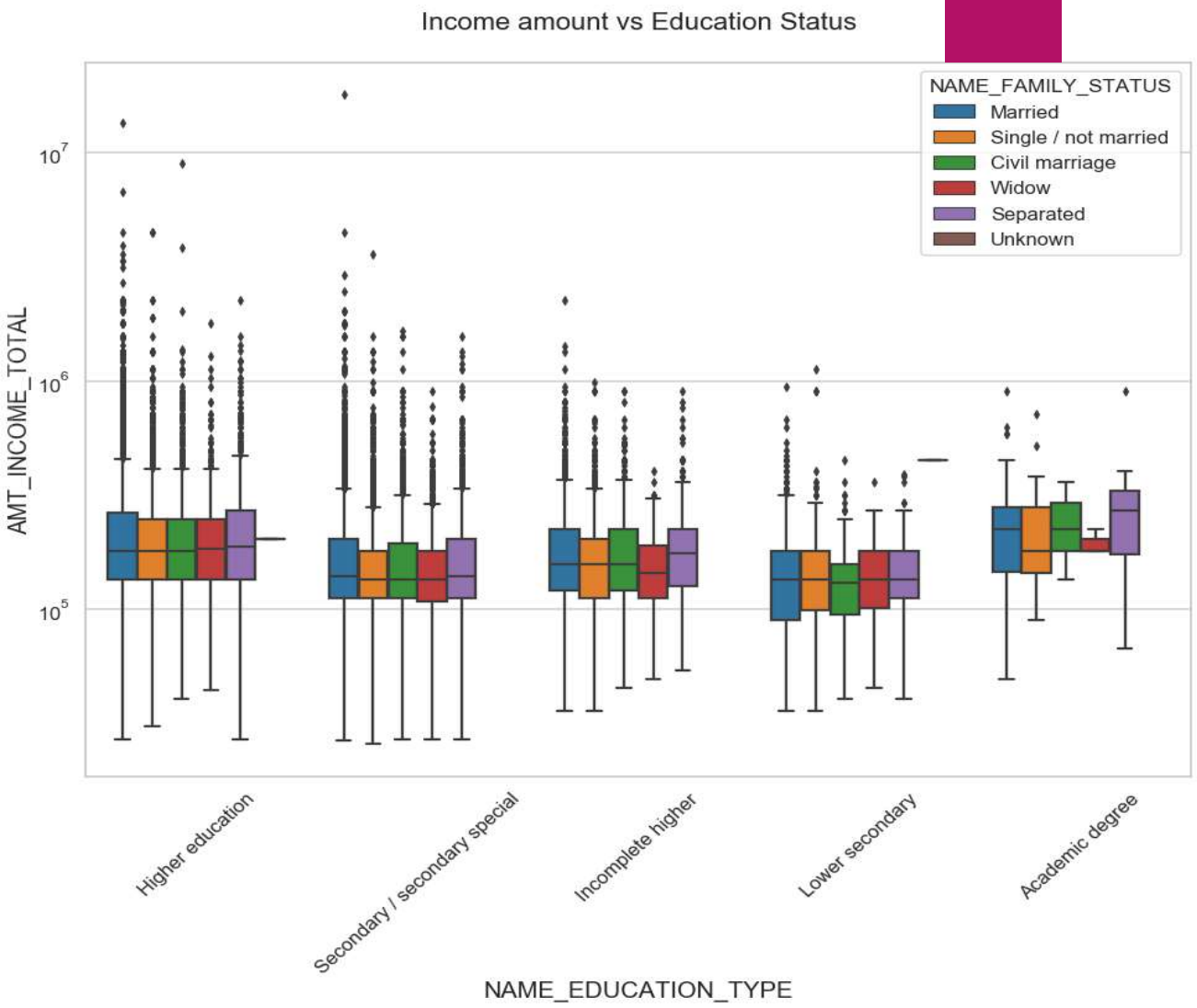
- Quite similar from Target 0, we can say that Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others.
- Most of the outliers are from Education type 'Higher education' and 'Secondary'.
- Civil marriage for Academic degree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



# Income amount vs Education Status

Conclusions

- Have some similarity with Target0, From above boxplot for Education type 'Higher education' the income amount is mostly equal with family status.
- Less outlier are having for Academic degree but there income amount is little higher that Higher education.
- Lower secondary are have less income amount than others.

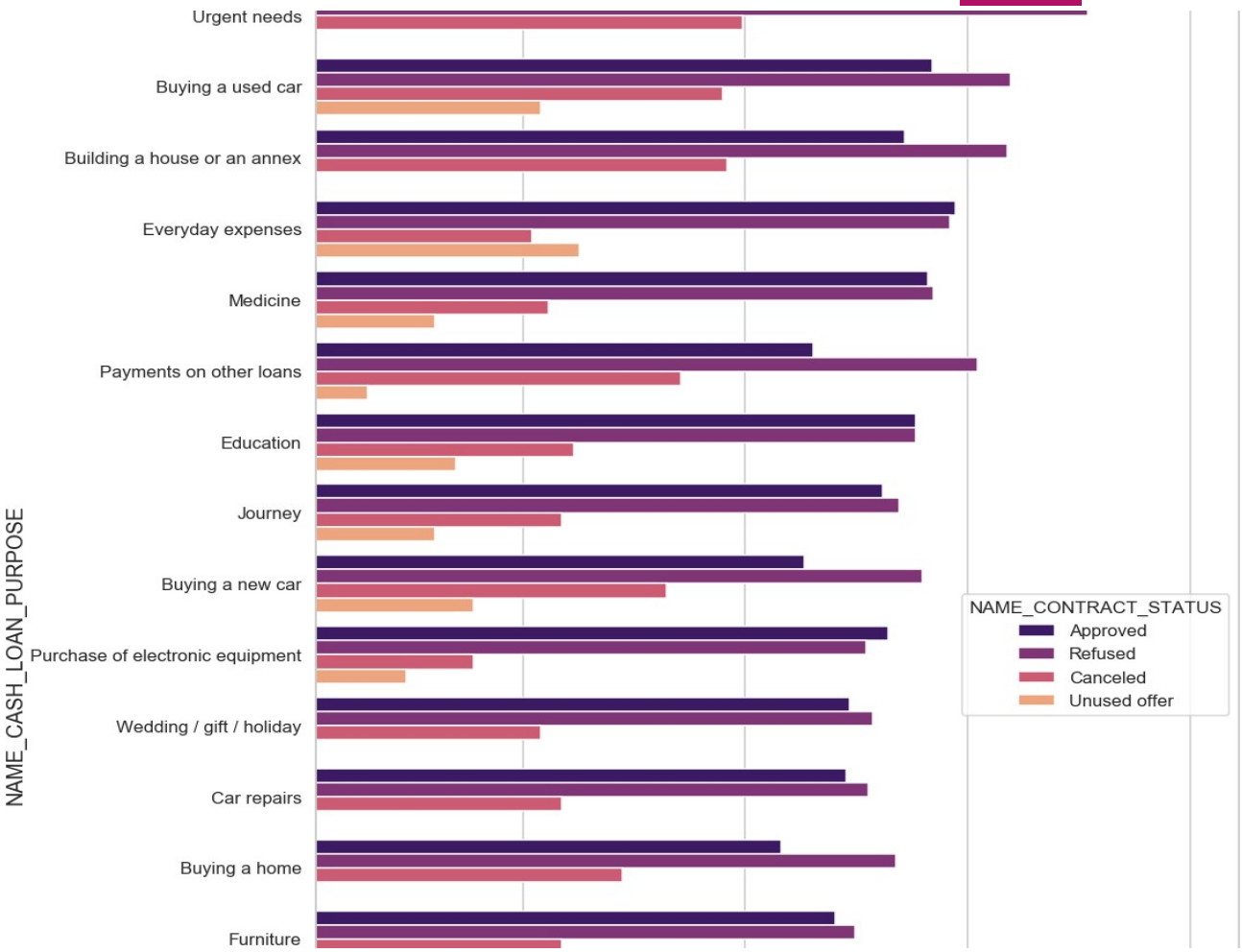


# Univariate analysis after merging previous data

## Distribution of contract status with purposes

Conclusions

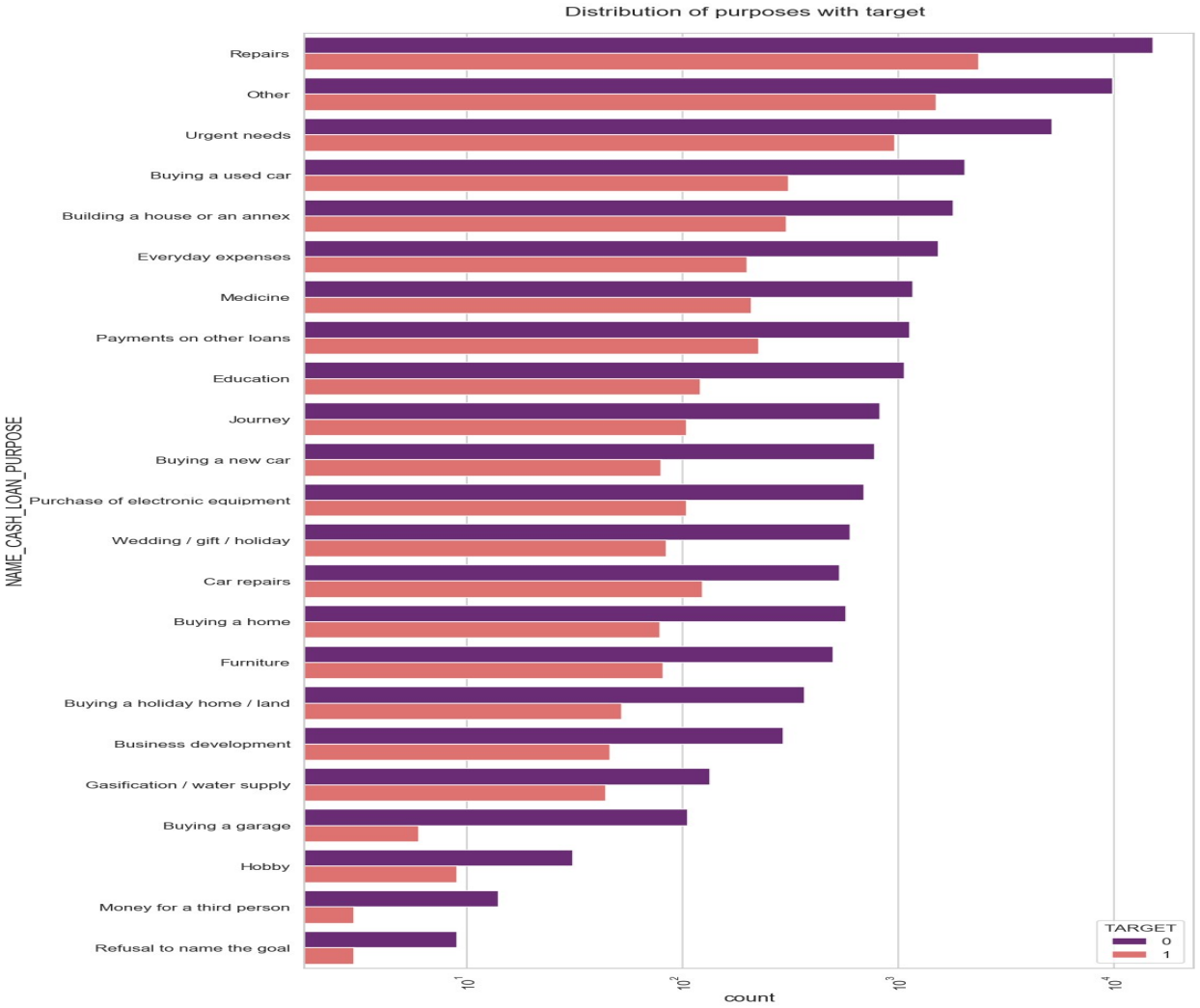
- Most rejection of loans came from purpose 'repairs'.
- For education purposes we have equal number of approves and rejection
- Paying other loans and buying a new car is having significant higher rejection than approves.



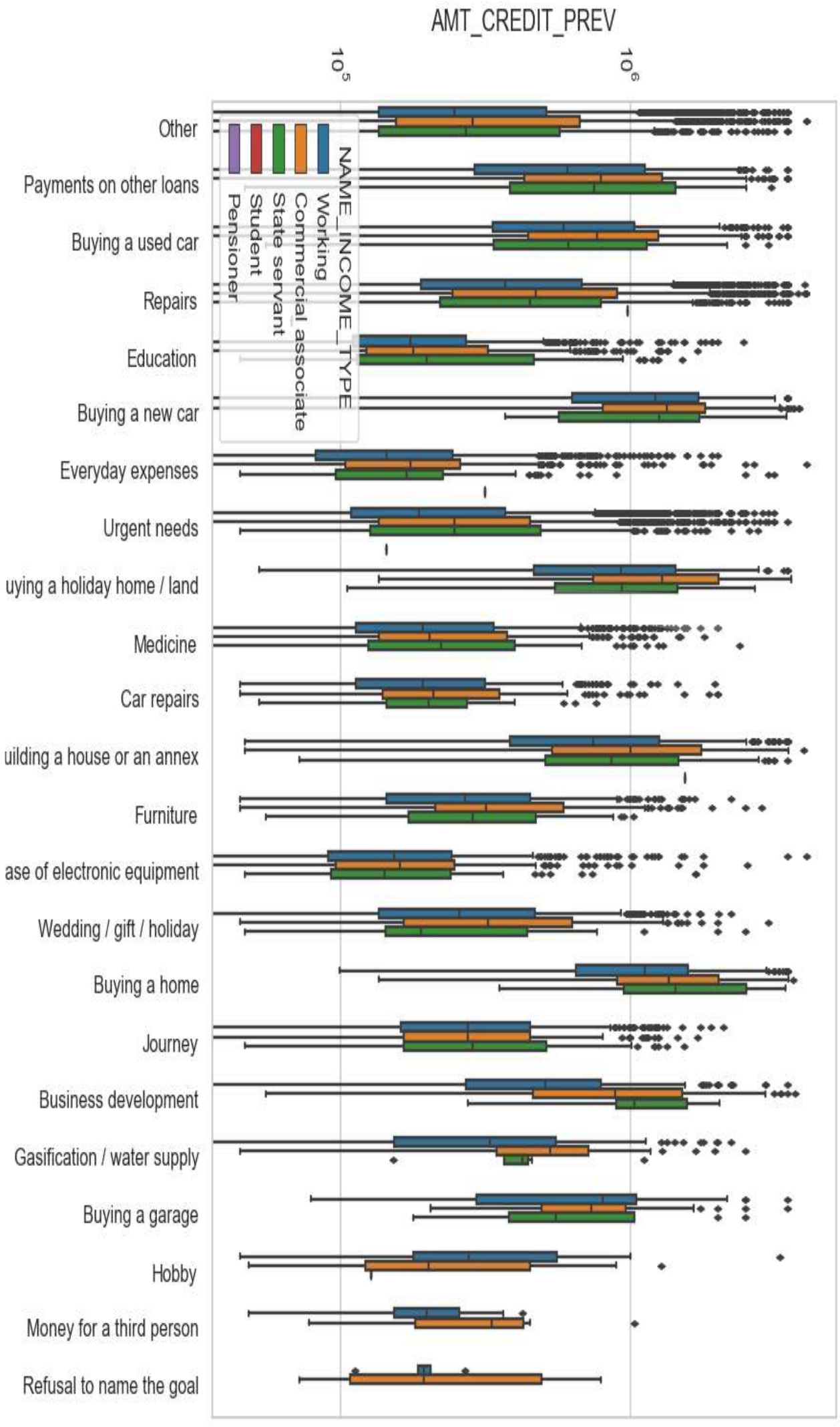
# Distribution of purposes with target

Conclusions

- Loan purposes with 'Repairs' are facing more difficulties in payment on time.
- 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 applicant is having for minimal payment difficulties.



# Performing Bivariate Analysis


AMT_CREDIT_PREV
$10^5$
$10^6$
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
uying a holiday home / land
Medicine
Car repairs
uilding a house or an annex
Furniture
ase 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

# Previous Credit amount vs Loan Purpose

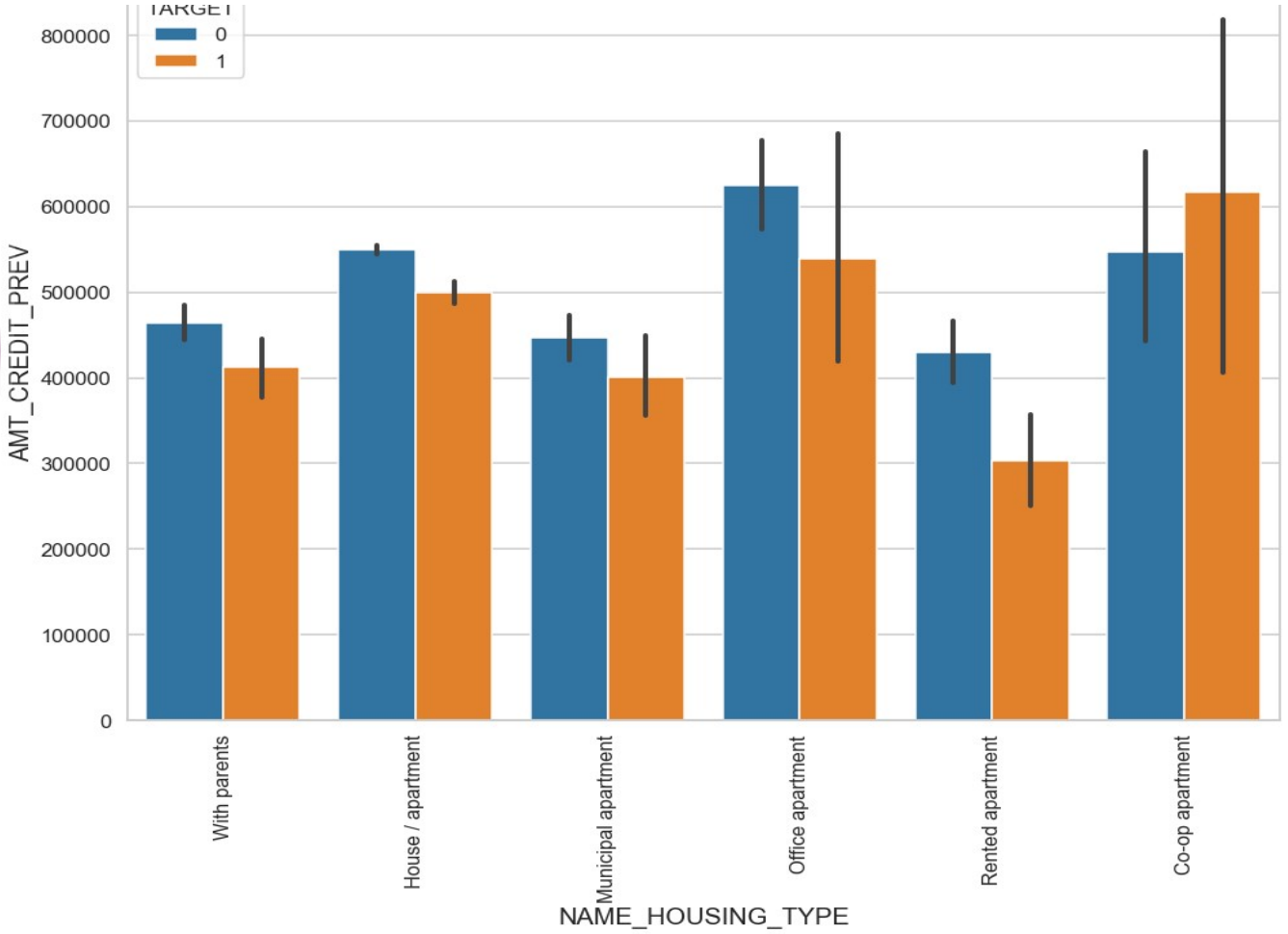
Conclusions:

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

# Previous Credit amount vs Housing type

Conlusions:.

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



# conclusion

- Banks should **focus more on - Contract type 'Student' ,'pensioner' and 'Businessman' with housing 'type other than 'Co-op apartment' for successful payments.**
- Banks should **focus less on - income type 'Working' as they are having the highest number of unsuccessful payments.**
- Loan purpose 'Repair' has highest number of unsuccessful payments
- Applicants with housing type 'With parents' have least number of unsuccessful payments.