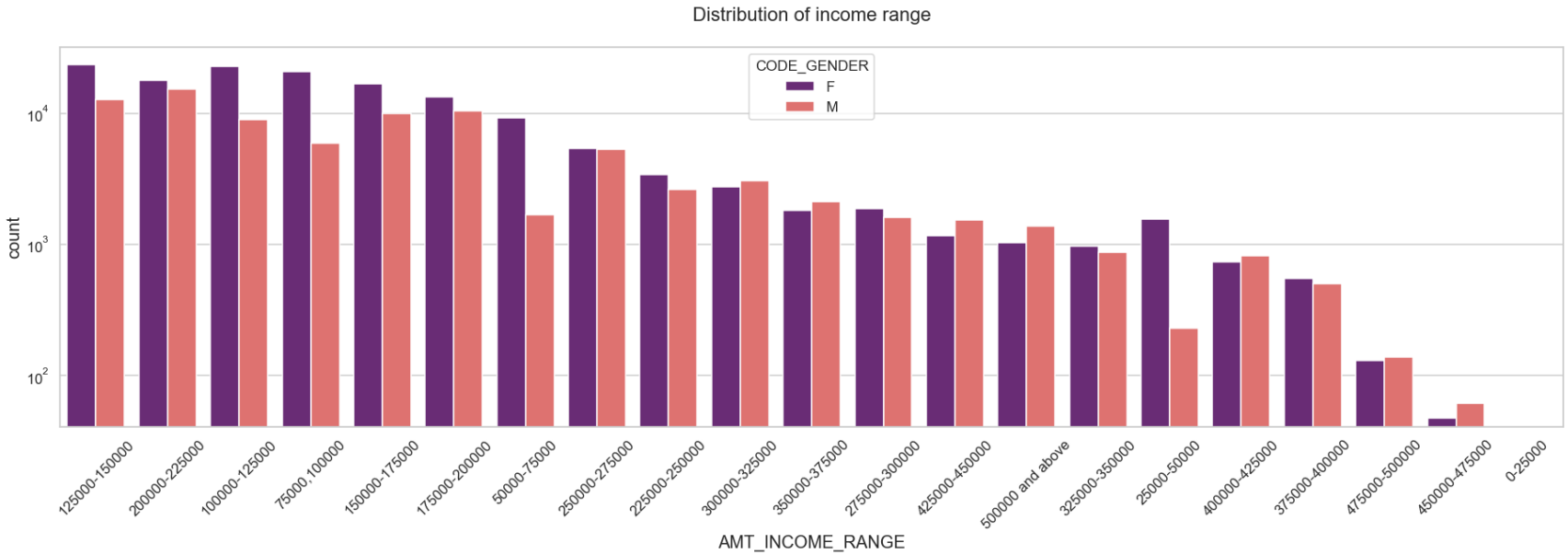
CREDIT EDA CASE STUDY

SUBMITTED BY-
RISHABH KUMAR SHARMA

CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET 0

DISTRIBUTION OF INCOME RANGE



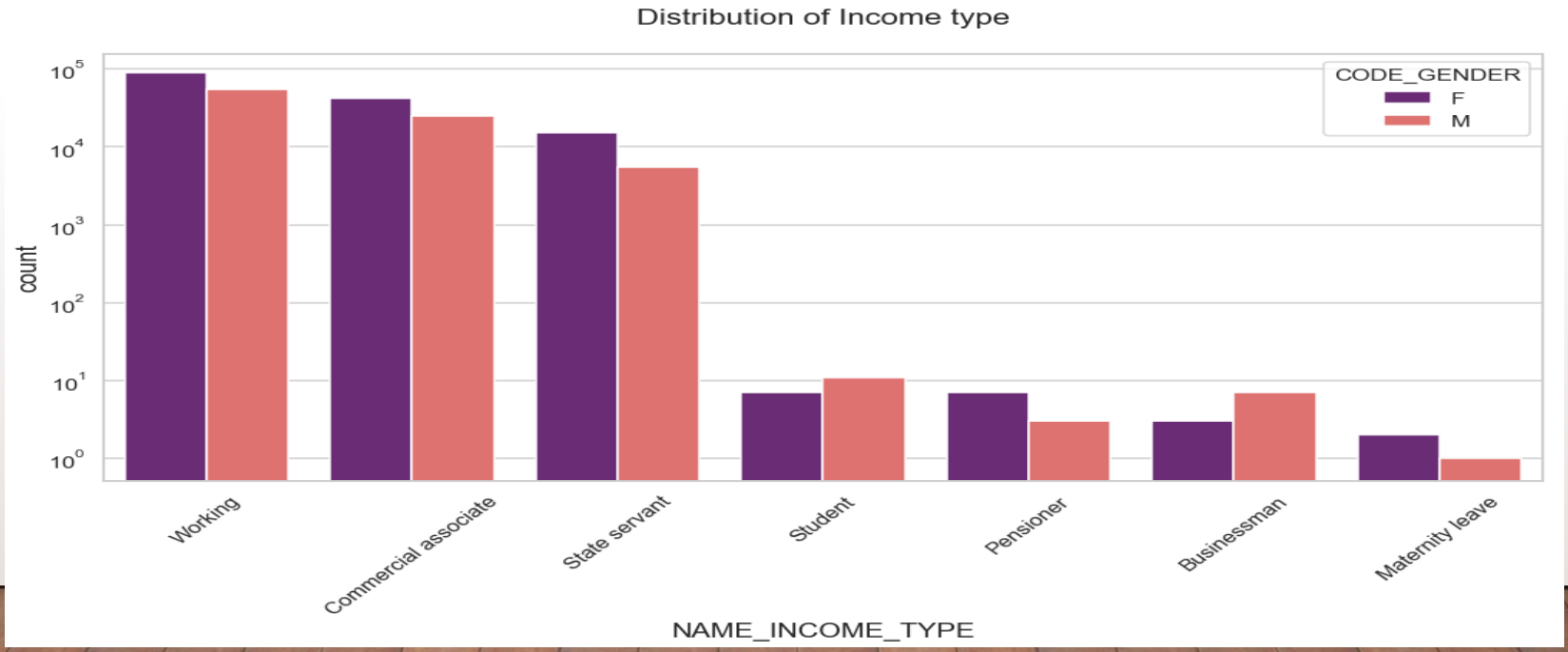
Conclusion from the graph -

- Female counts are higher than male.
- Income range from 100000 to 200000 is having more 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

Conclusion we get from the graph:-

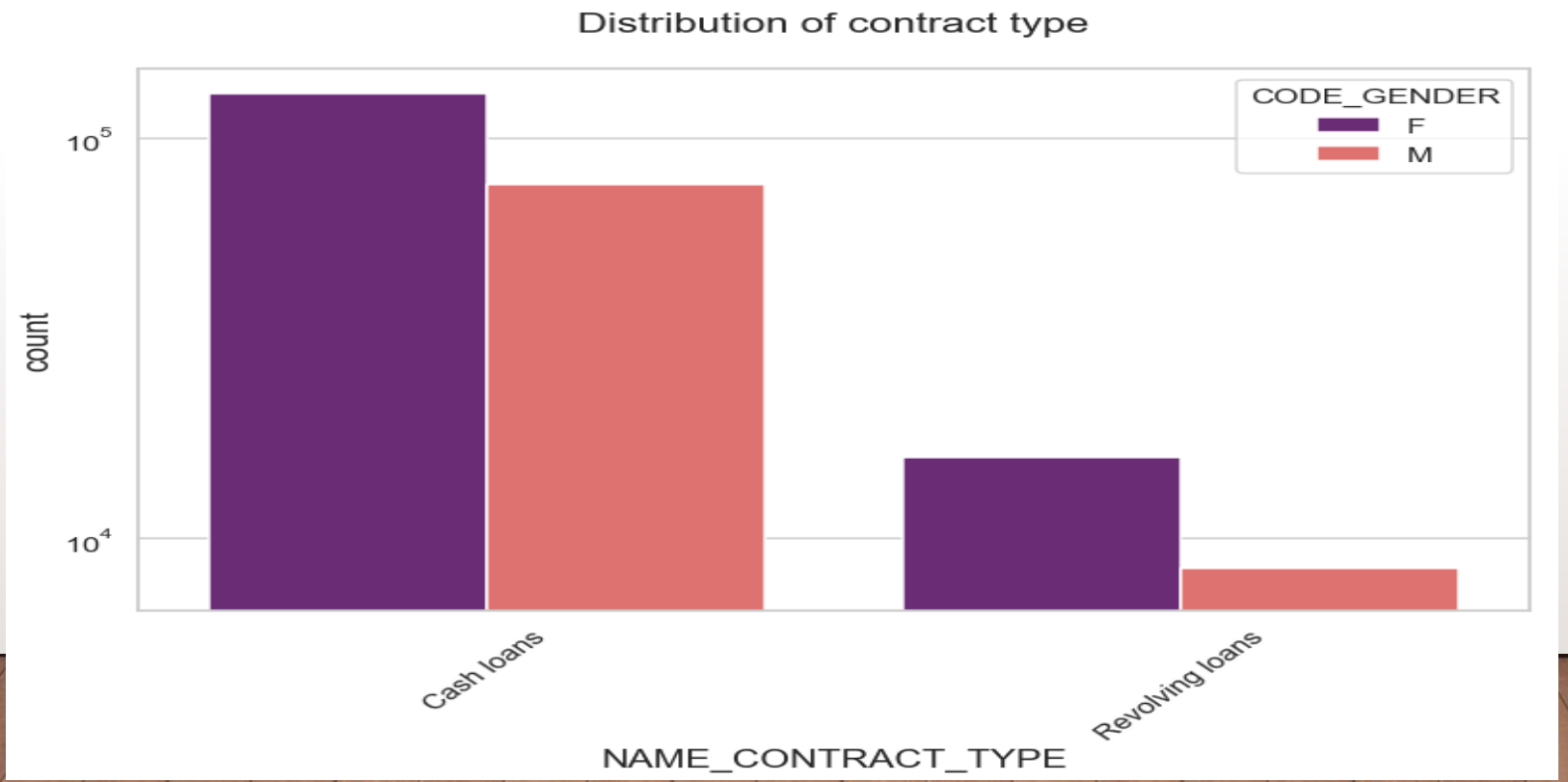
- For income type ‘working’, ’commercial associate’, and ‘State Servant’ the number of credits are higher than others.
- For this Females are having more number of credits than male.
- Less number of credits for income type ‘student’ ,’pensioner’, ‘Businessman’ and ‘Maternity leave’.



DISTRIBUTION FOR CONTRACT TYPE

Conclusion from the graph:-

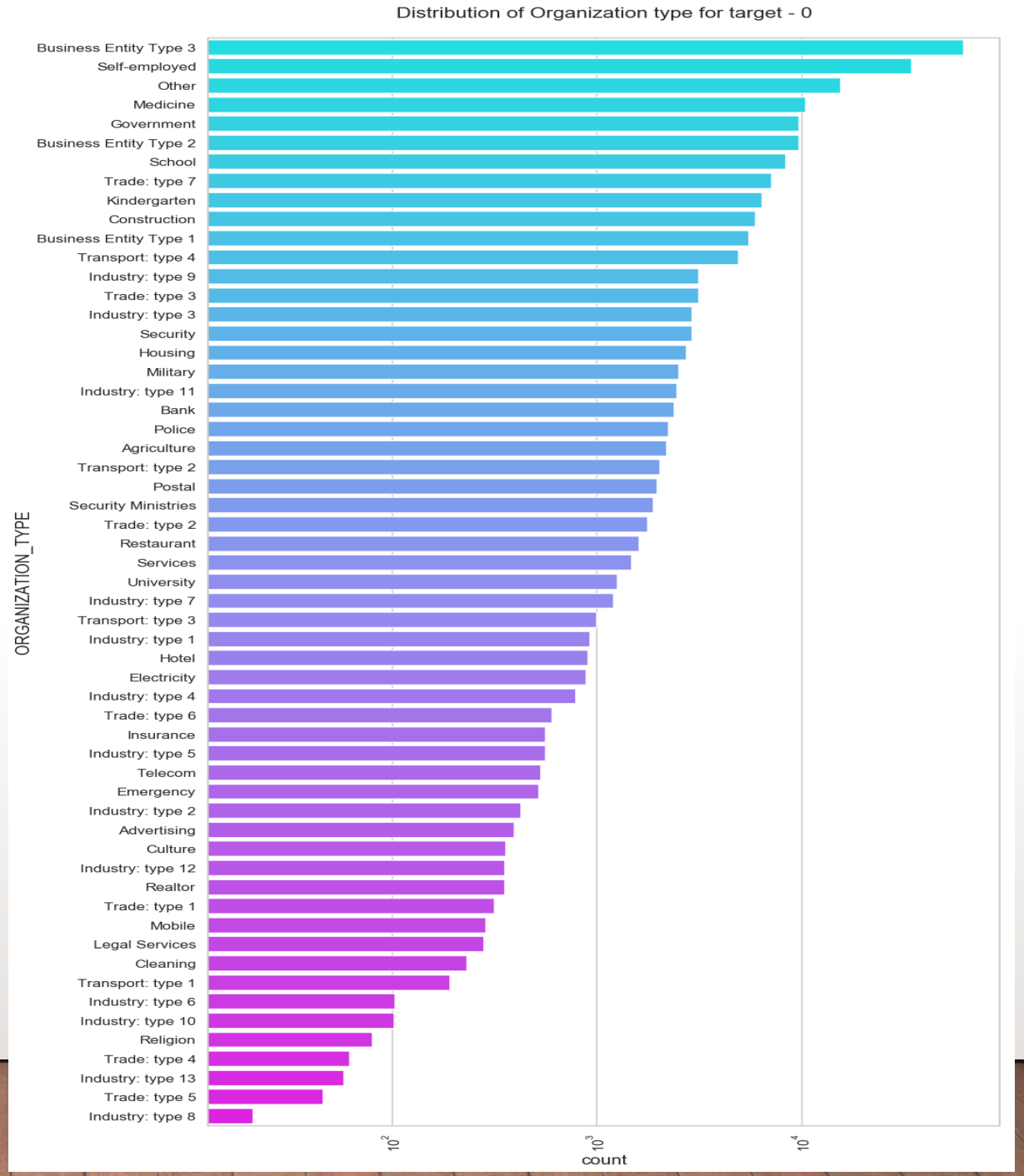
- For contract type ‘cash loans’ is having higher number of credits than ‘Revolving loans’ contract type.
- For this also Female is leading for applying credits.



ORGANIZATION

Conclusion from the graph:-

- 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.

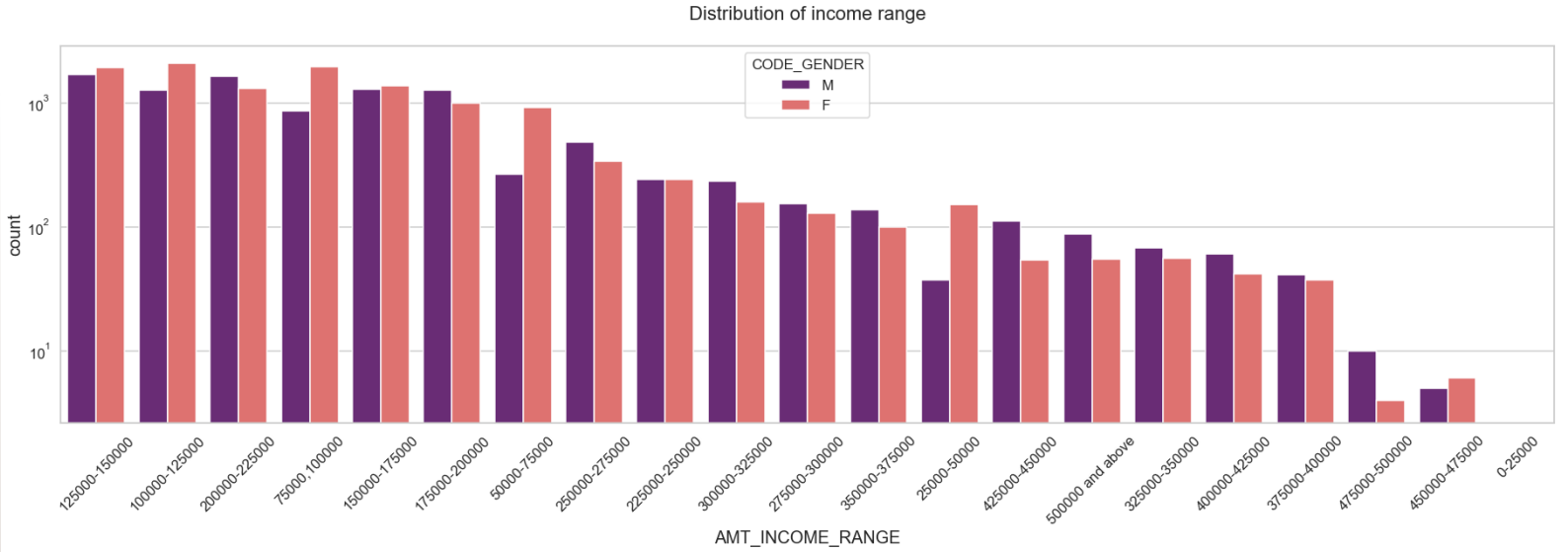


CATEGORICAL UNIVARIATE ANALYSIS FOR TARGET 1

RANGE

Conclusion from the graph:-

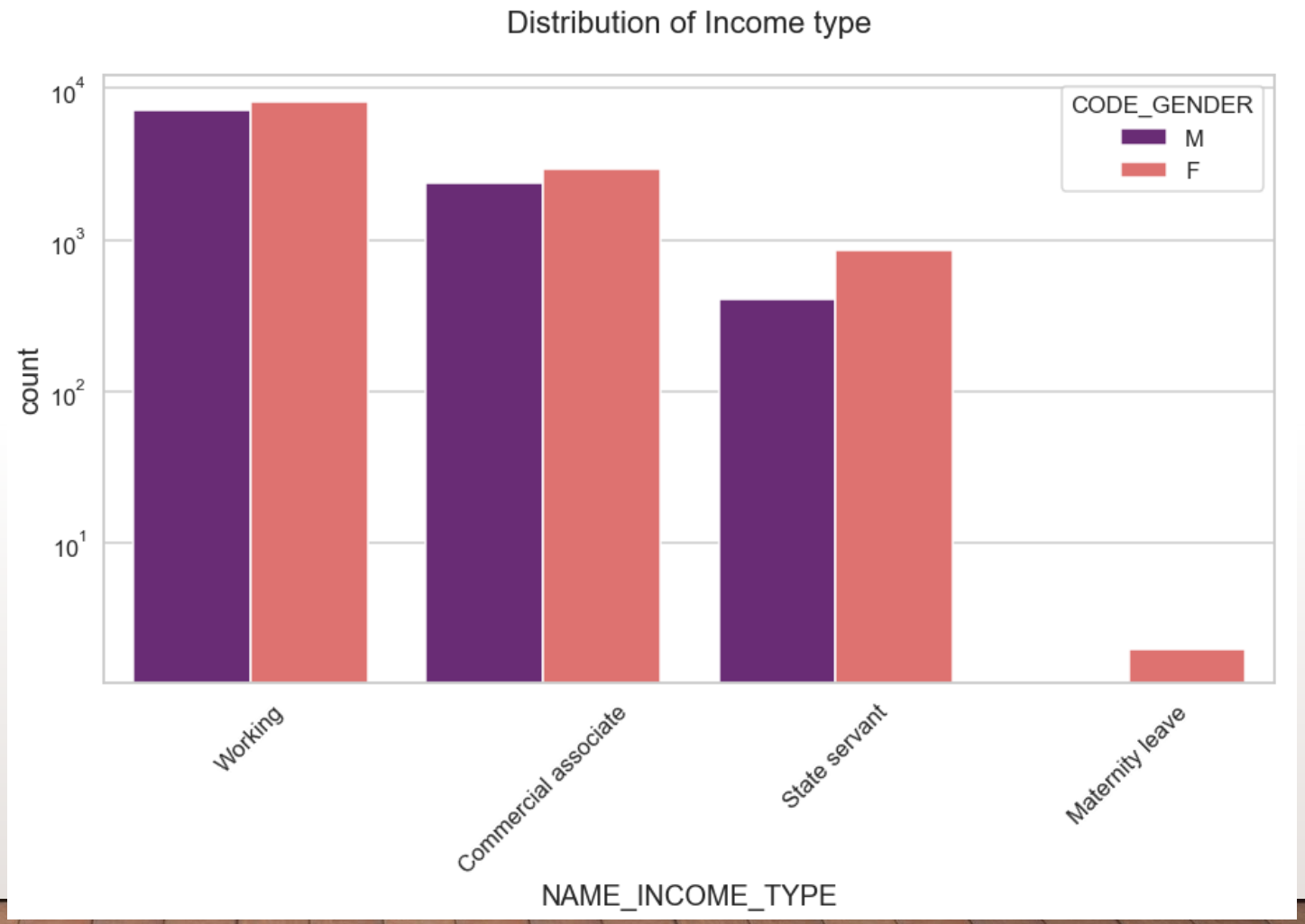
- Male counts are higher than female.
- Income range from 100000 to 200000 is having more number of credits.
- This graph show that males are more than female in having credits for that range.
- Very less count for income range 400000 and above.



DISTRIBUTION OF INCOME TYPE

Conclusion from the graph :

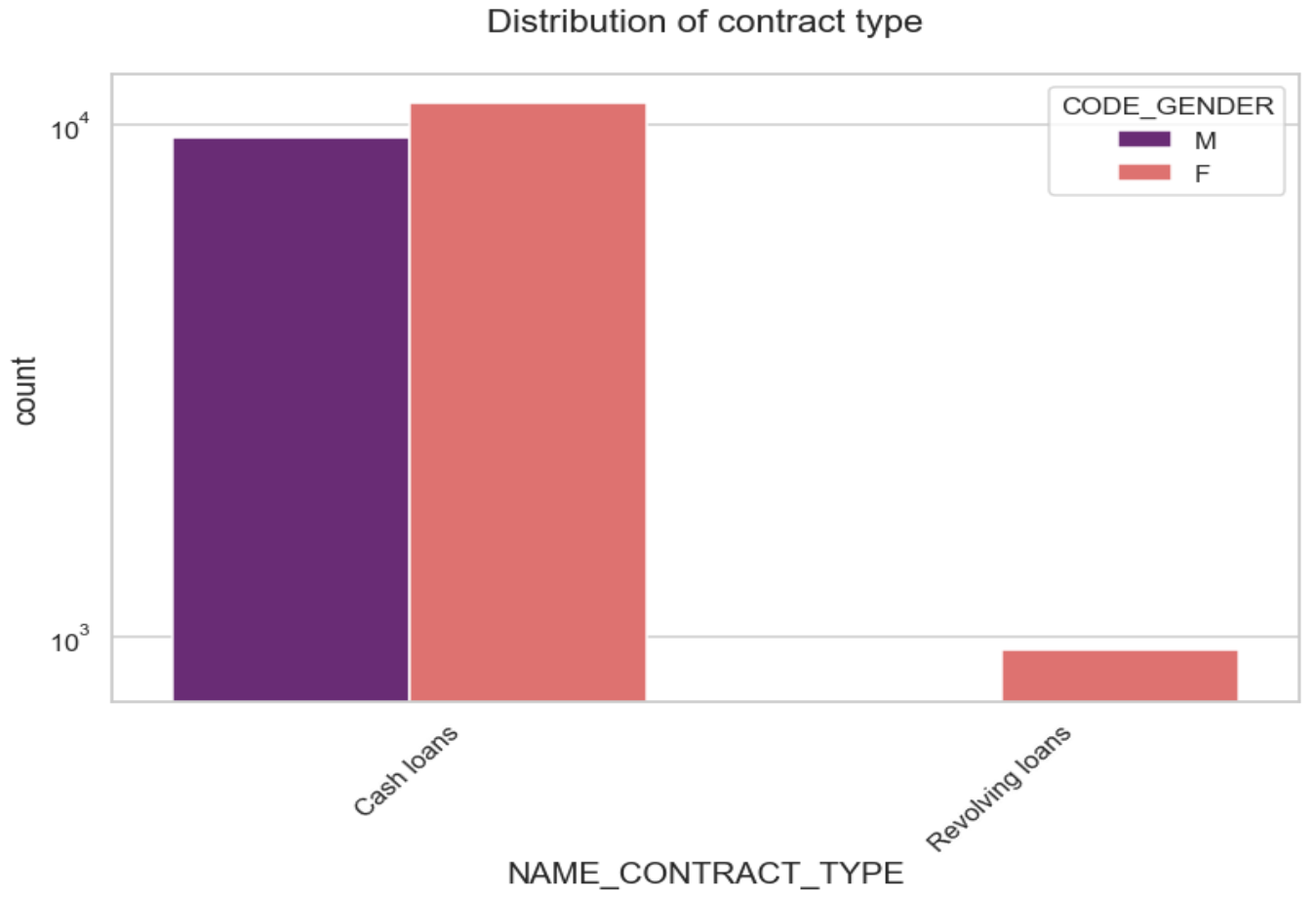
- For income type ‘working’, ’commercial associate’, and ‘State Servant’ the number of credits are higher than other i.e. ‘Maternity leave.
- For this Females are having more number of credits than male.
- Less number of credits for income type ‘Maternity leave’.
- For type 1: There is no income type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



DISTRIBUTION FOR CONTRACT TYPE

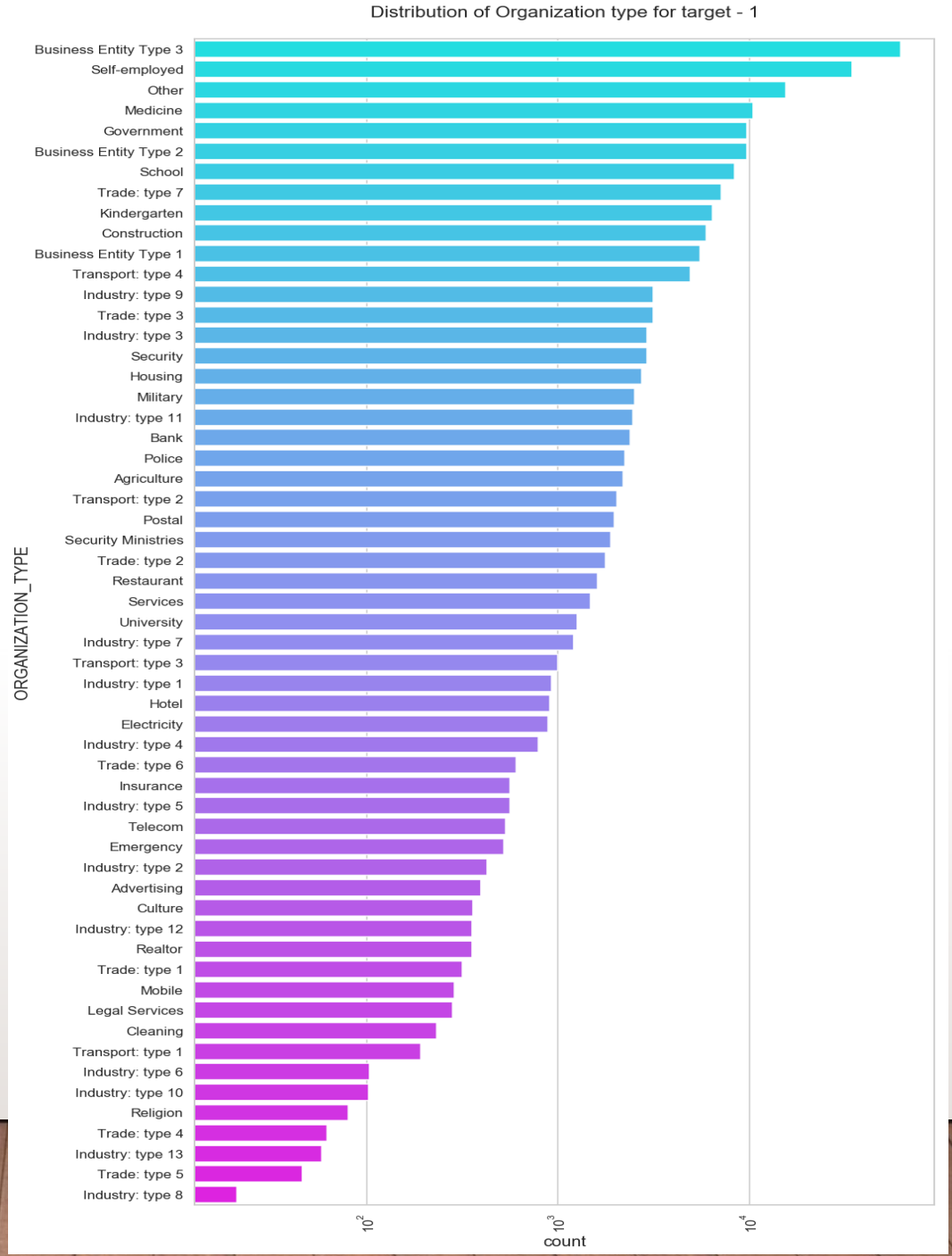
Conclusion from the graph:.

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

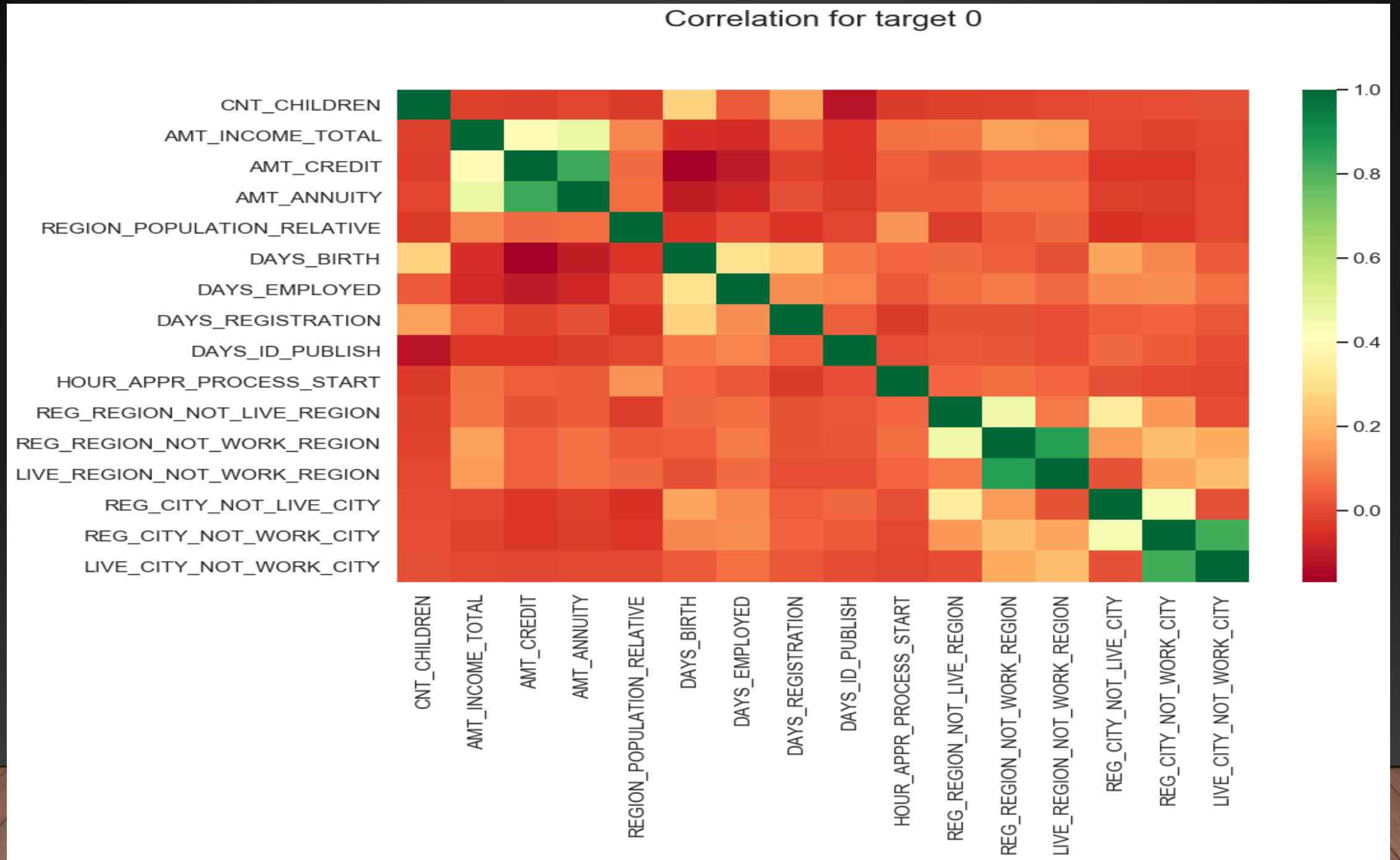


DISTRIBUTION OF ORGANIZATION TYPE

- 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.

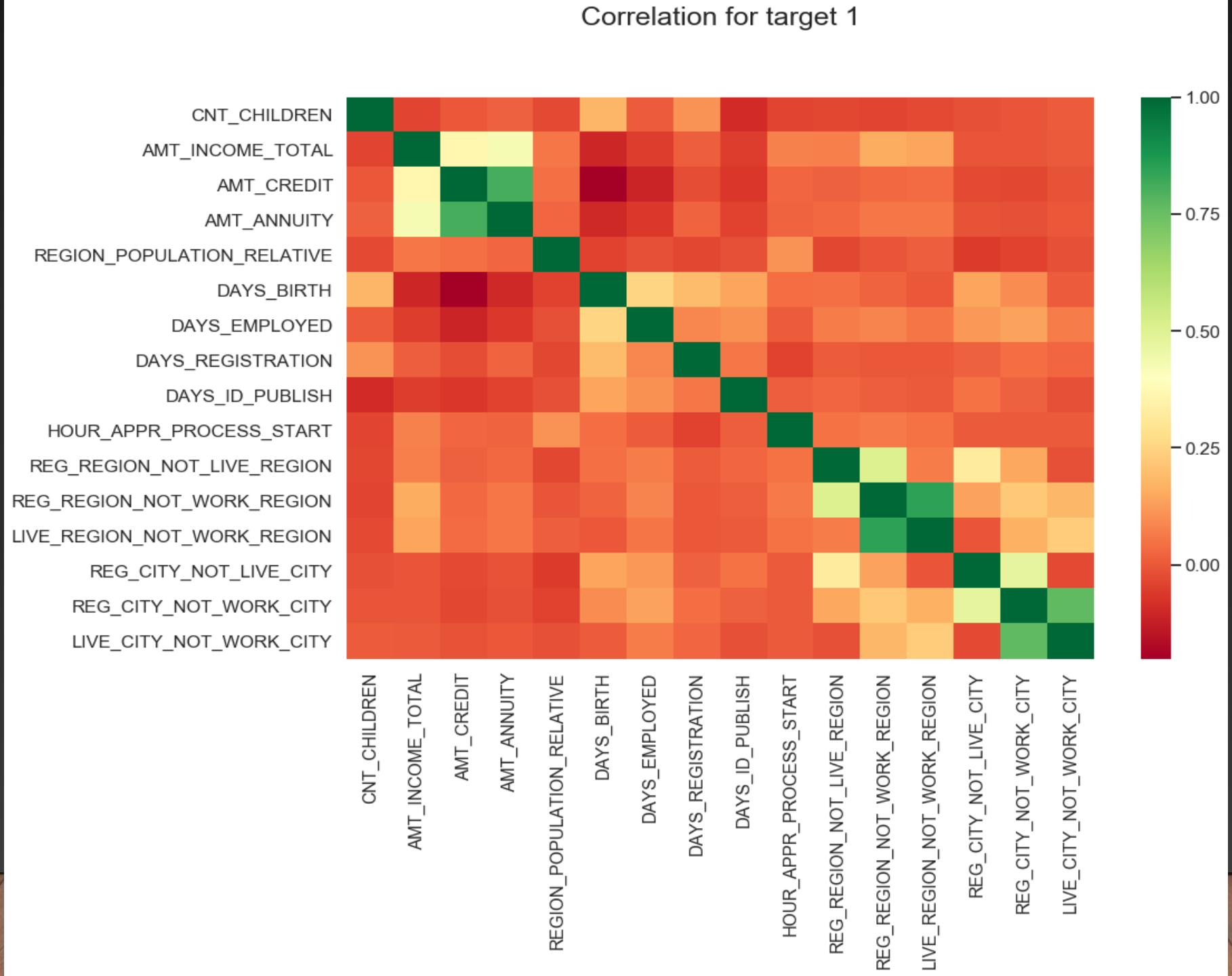


CORRELATION OF TARGET 0


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

- 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 the client has, which means Credit amount is higher for fewer children count the client has and vice-versa.
- The income amount is inversely proportional to the number of children the client has, which means more income for fewer children clients and vice-versa.
- fewer children clients have in a densely populated area.
- The credit amount is higher in densely populated areas.
- The income is also higher in a densely populated areas.


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 TYPE 1

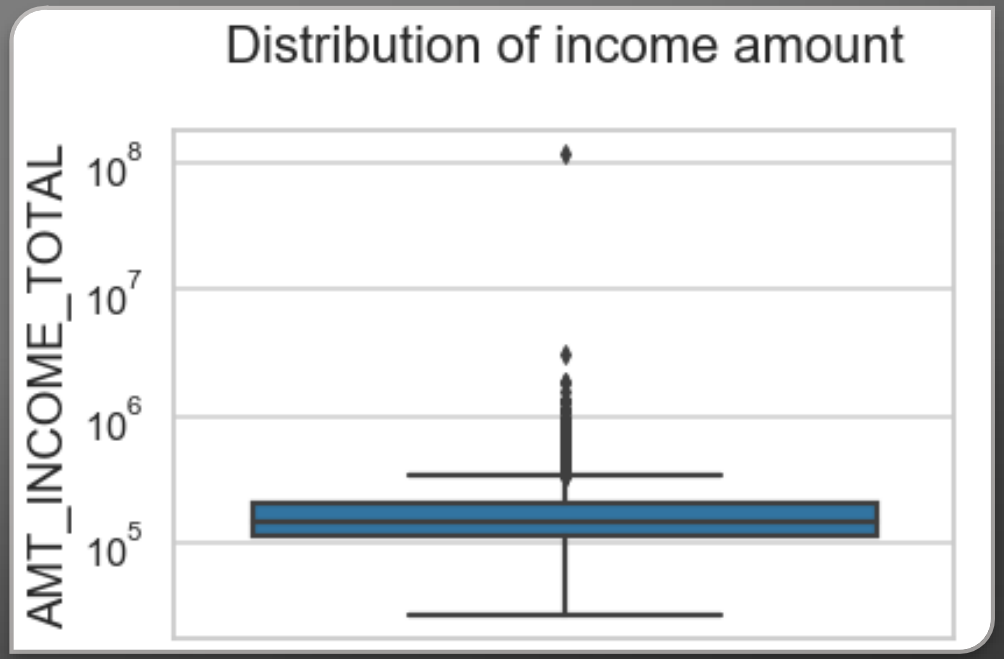
- This heat map for Target 1 is also having quite the same observation just like Target 0. But for few points are different. They are listed below.
  - The client's permanent address does not match the contact address are having fewer children and vice-versa
  - The client's permanent address does not match the work address are having fewer children and vice-versa

CATEGORICAL UNIVARIATE ANALYSIS FOR VARIABLES TARGET 0

BOXPLOT FOR INCOME AMOUNT

Few points can be concluded from the graph.

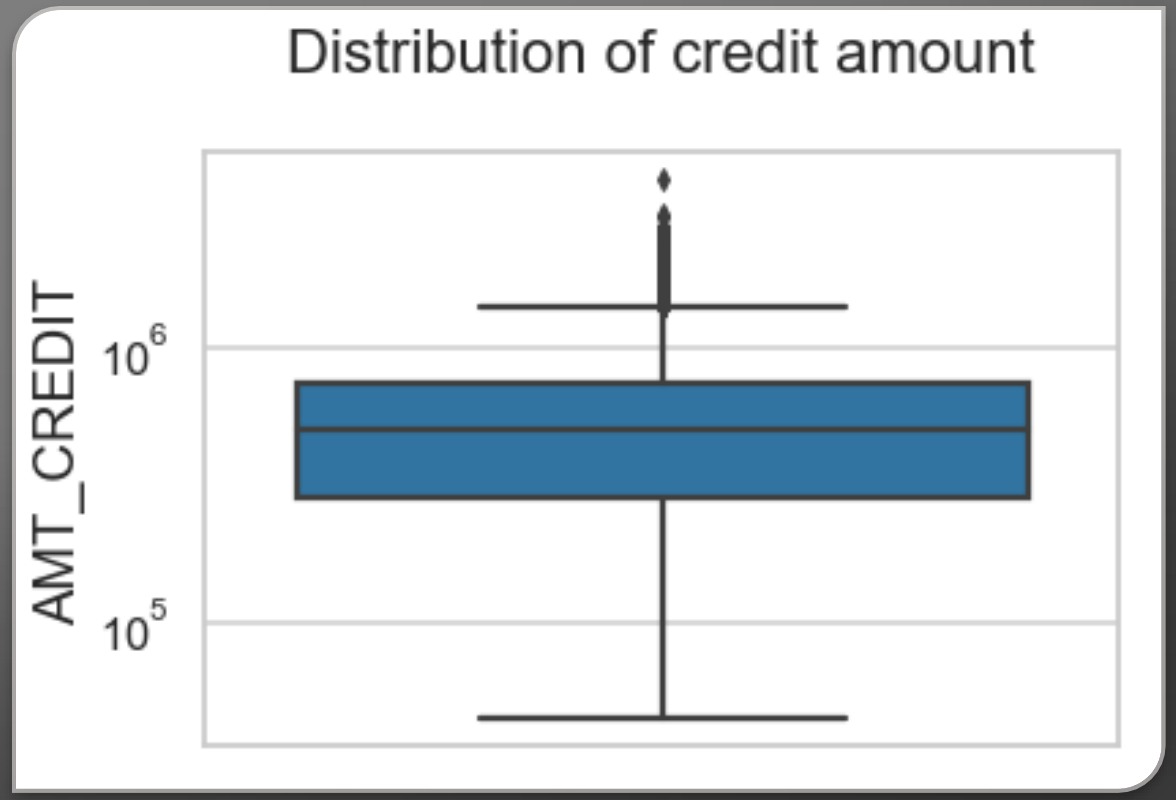
- Some outliers are noticed in income amount.
- The third quartiles is very slim for income amount.



BOXPLOT FOR CREDIT AMOUNT

Few points can be concluded from the graph.

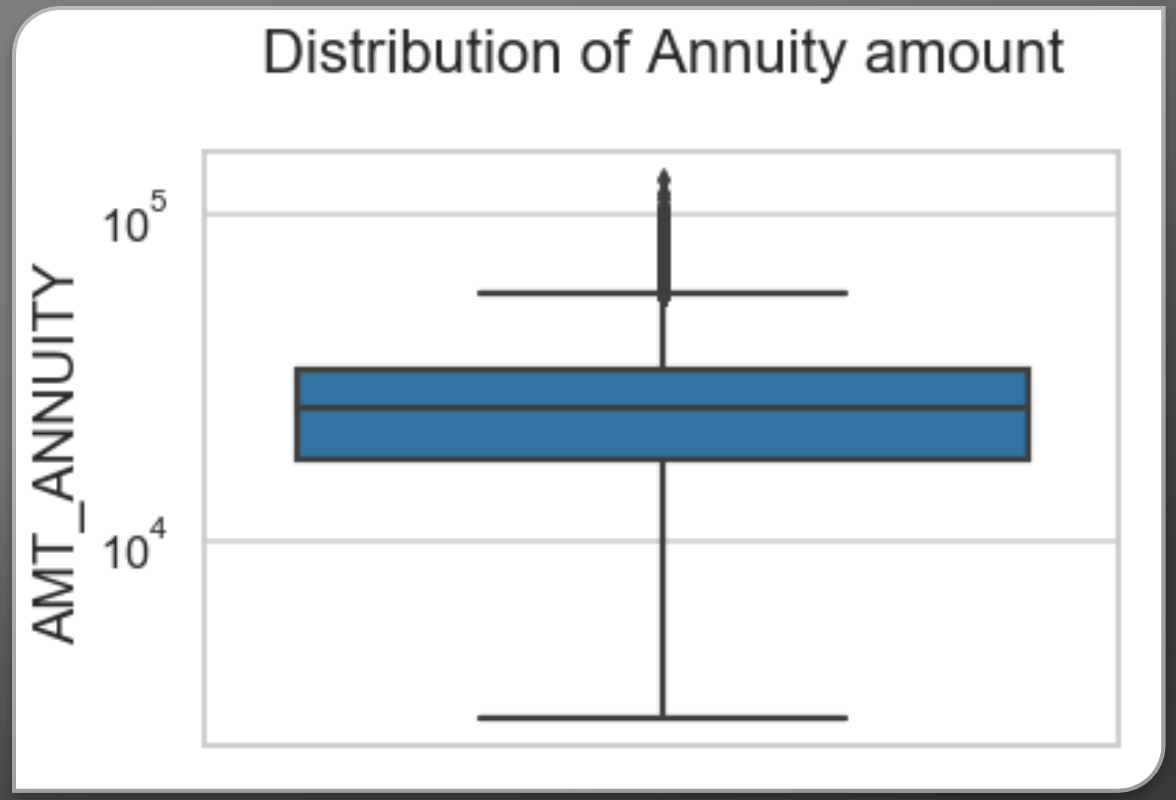
- Some outliers are noticed in credit amount.
- The first quartile is bigger than third quartile for credit amount which means most of the credits of clients are present in the first quartile.



BOXPLOT FOR ANNUITY AMOUNT

Few points can be concluded from the graph.

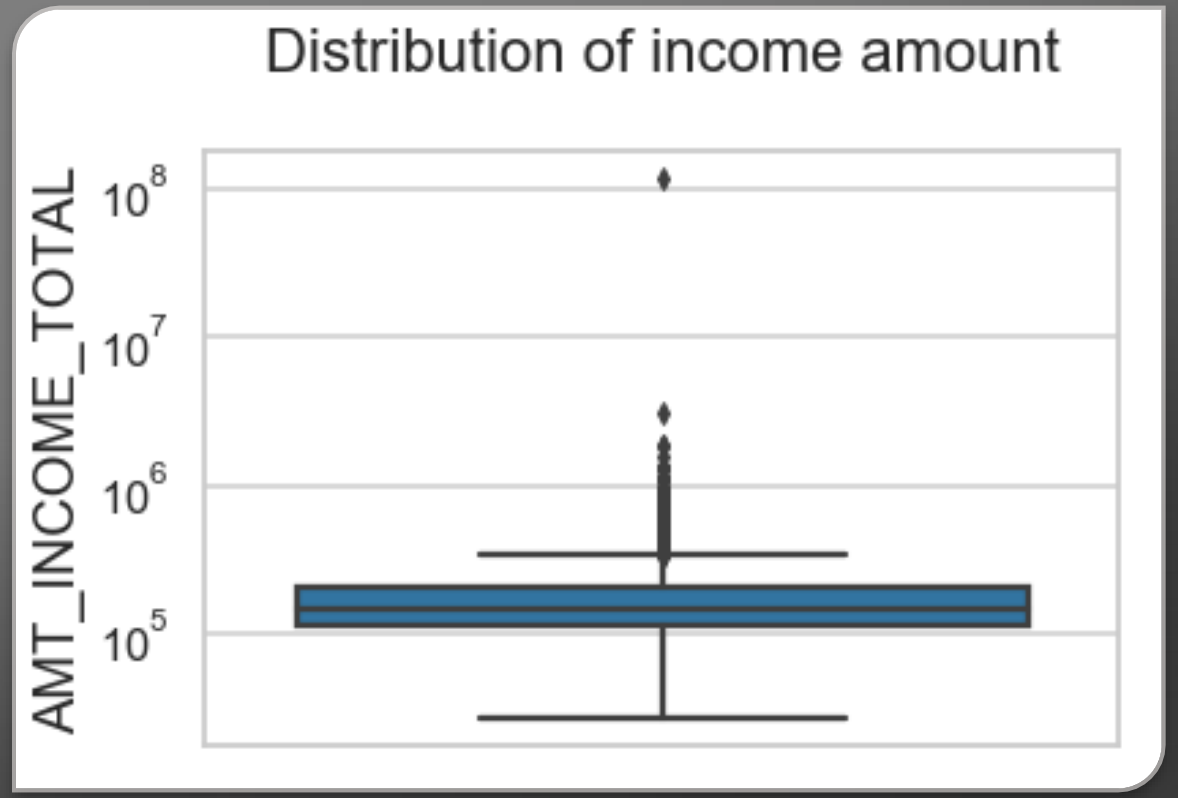
- Some outliers are noticed in annuity amount.
- The first quartile is bigger than third quartile for annuity amount which means most of the annuity clients are from first quartile.



CATEGORICAL UNIVARIATE ANALYSIS FOR VARIABLES TARGET 1

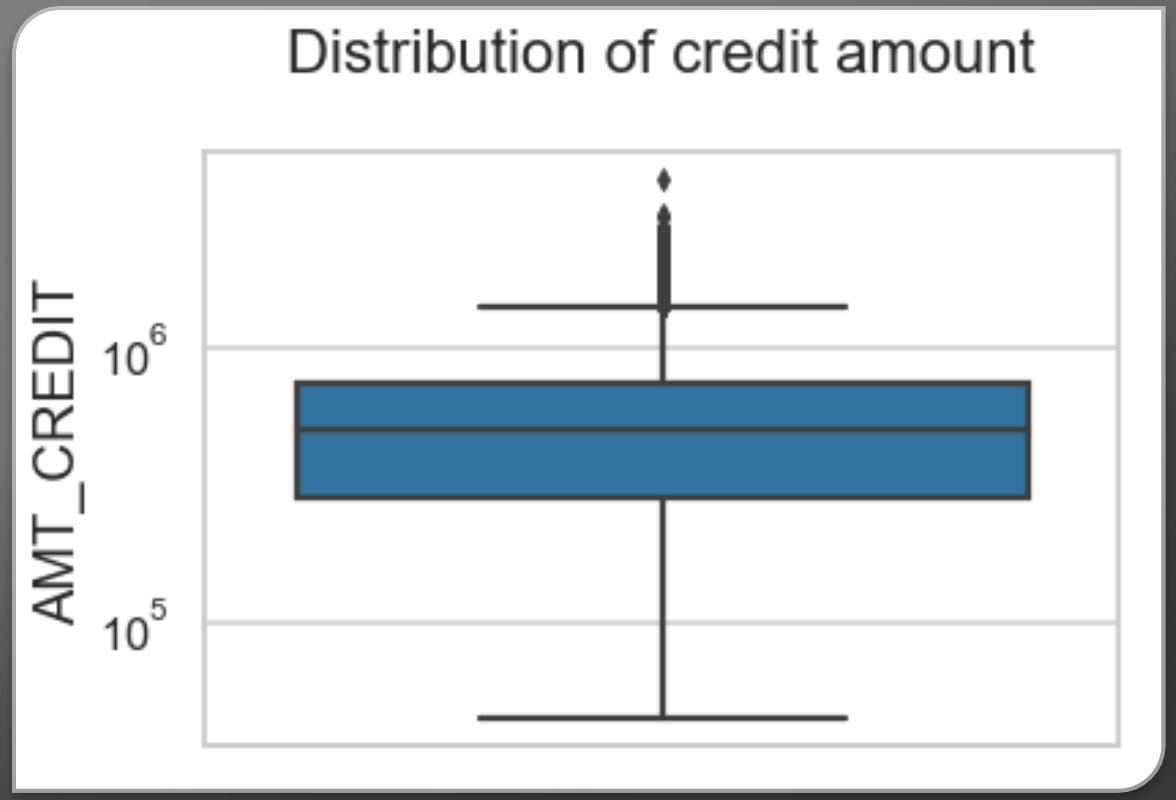
BOXPLOT FOR INCOME AMOUNT

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



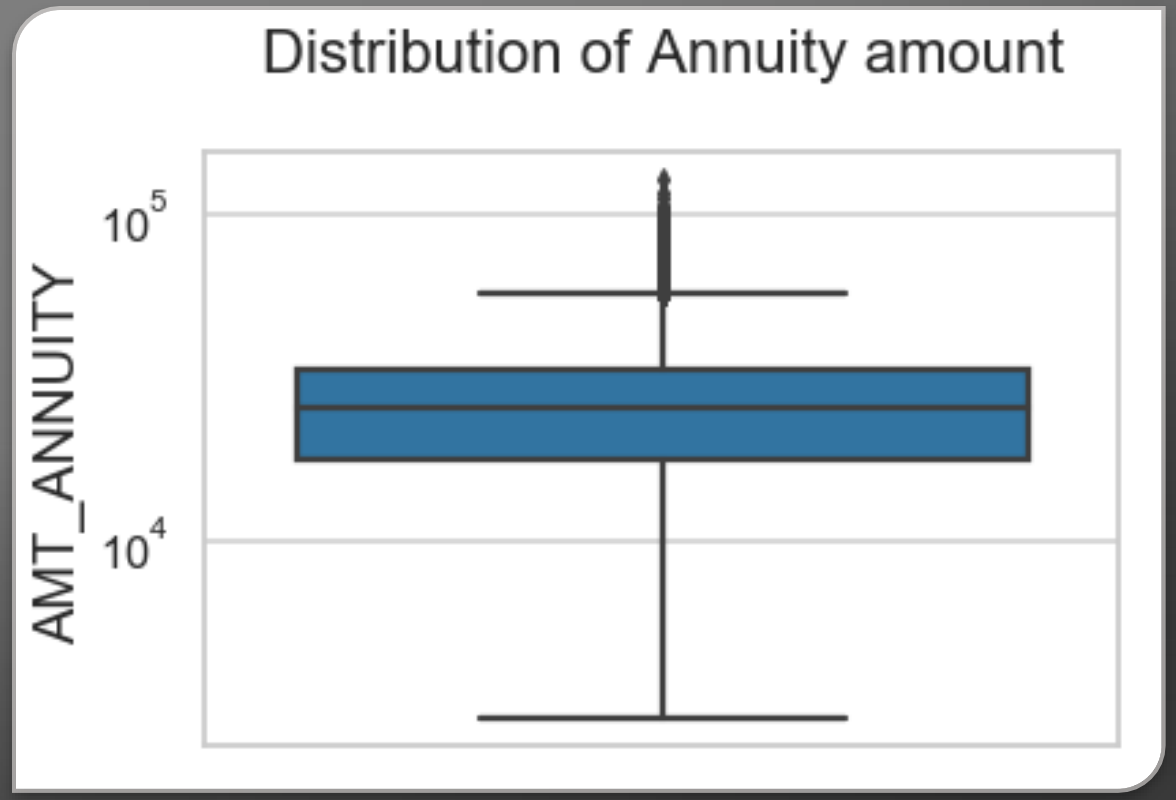
BOXPLOT FOR CREDIT AMOUNT

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



BOXPLOT FOR ANNUITY AMOUNT

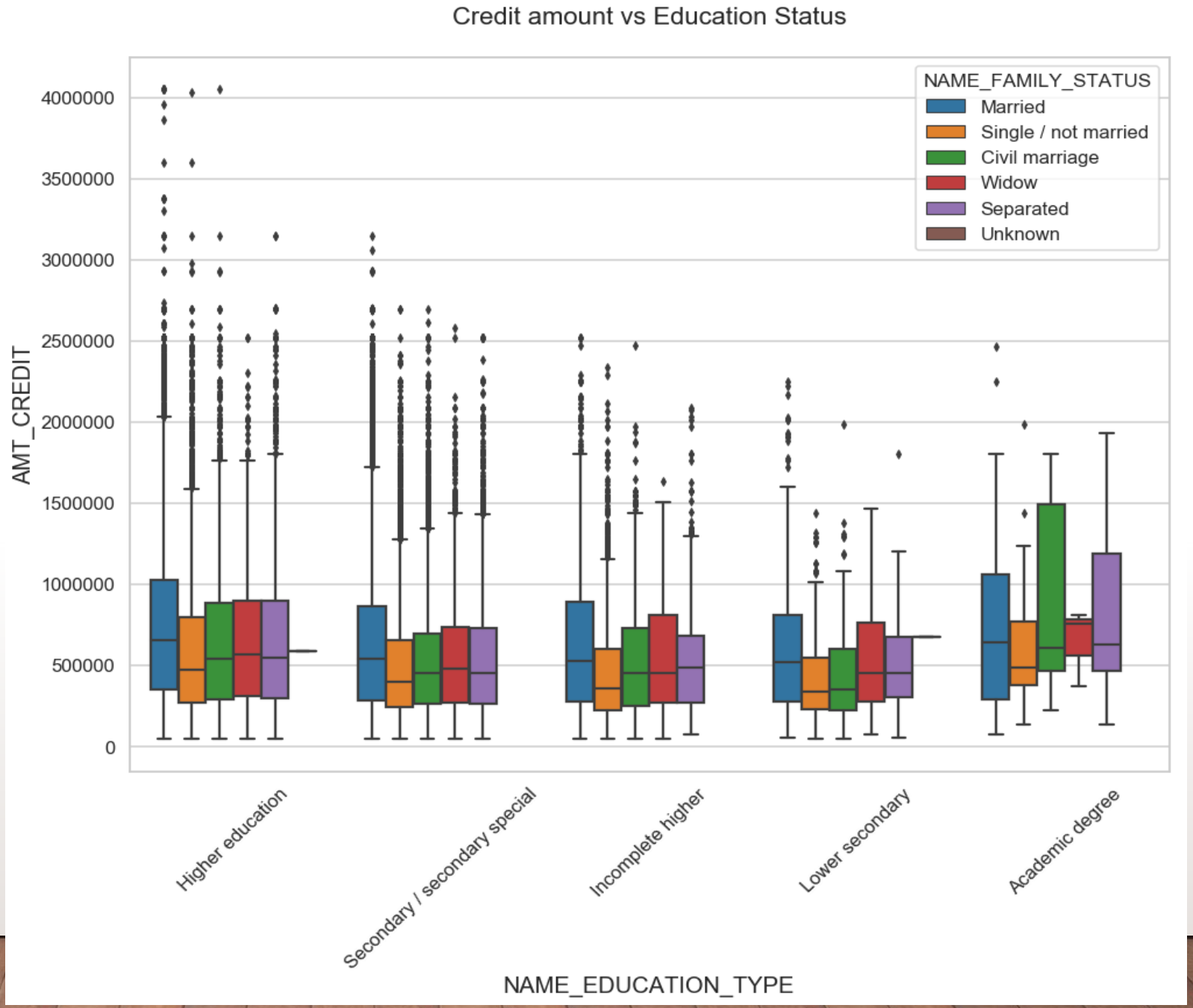
- Some outliers are noticed in the annuity amount.
- The first quartile is bigger than the third quartile for annuity amount which means most of the annuity clients are from the first quartile



BIVARIATE ANALYSIS FOR TYPE 0

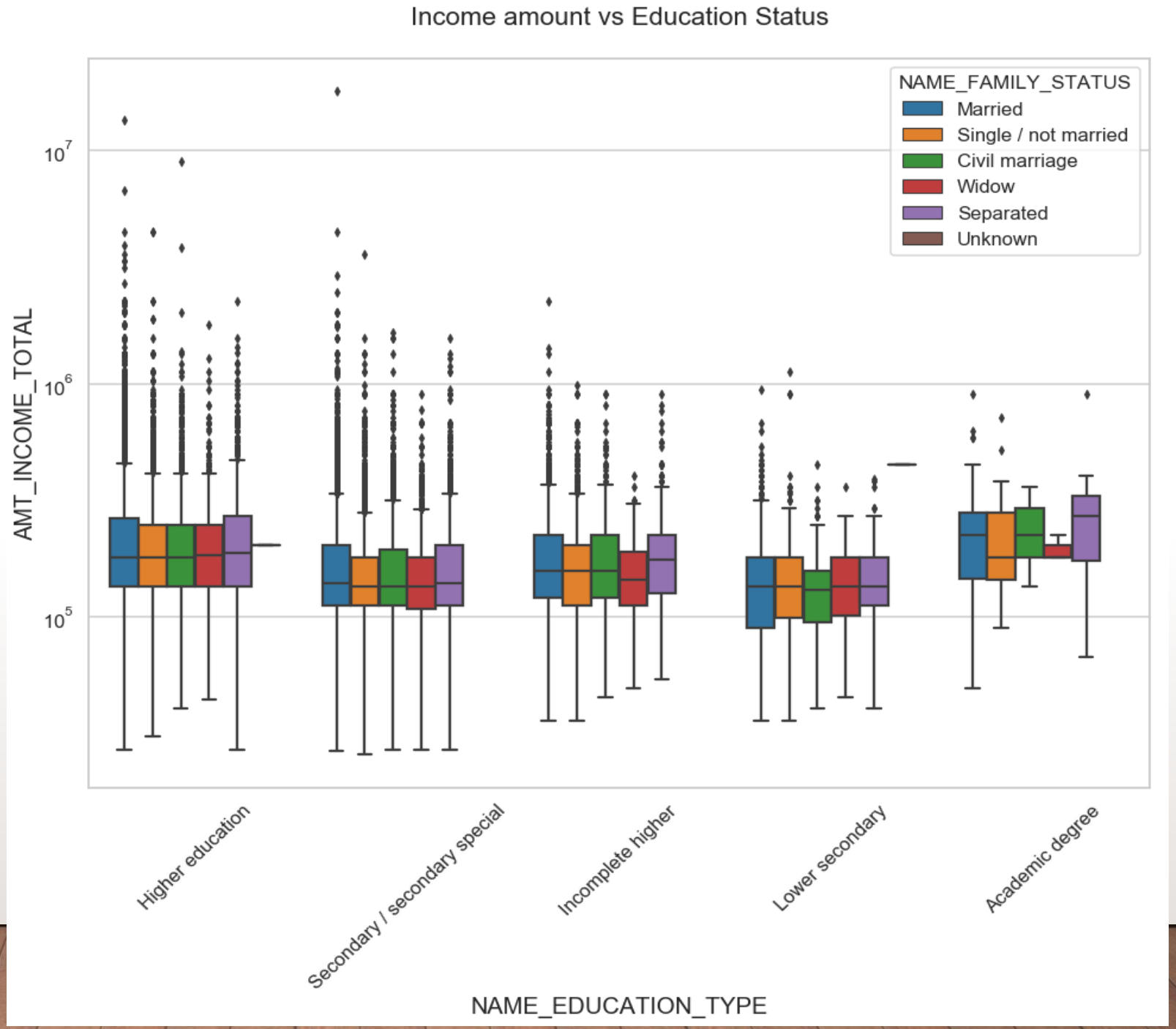
CREDIT AMOUNT VS EDUCATION STATUS

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



INCOME AMOUNT VS EDUCATION STATUS

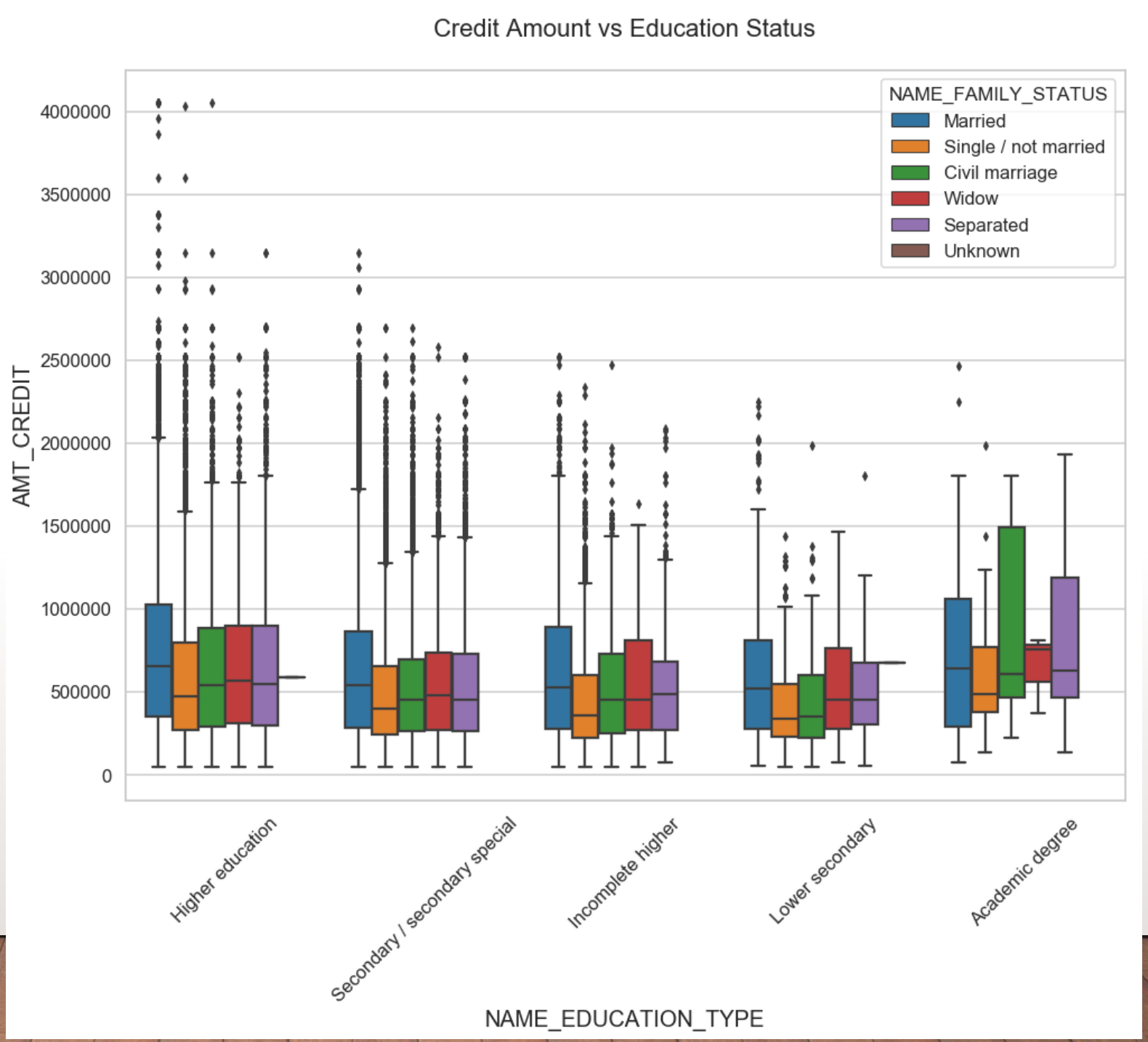
- For the Education type 'Higher education' the income amount mean is mostly equal with family status. It does contain many outliers.
- Fewer outliers are having Academic degrees but they are having income amount is a little higher than Higher education.
- Lower secondary of civil marriage family status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



BIVARIATE ANALYSIS FOR TYPE 1

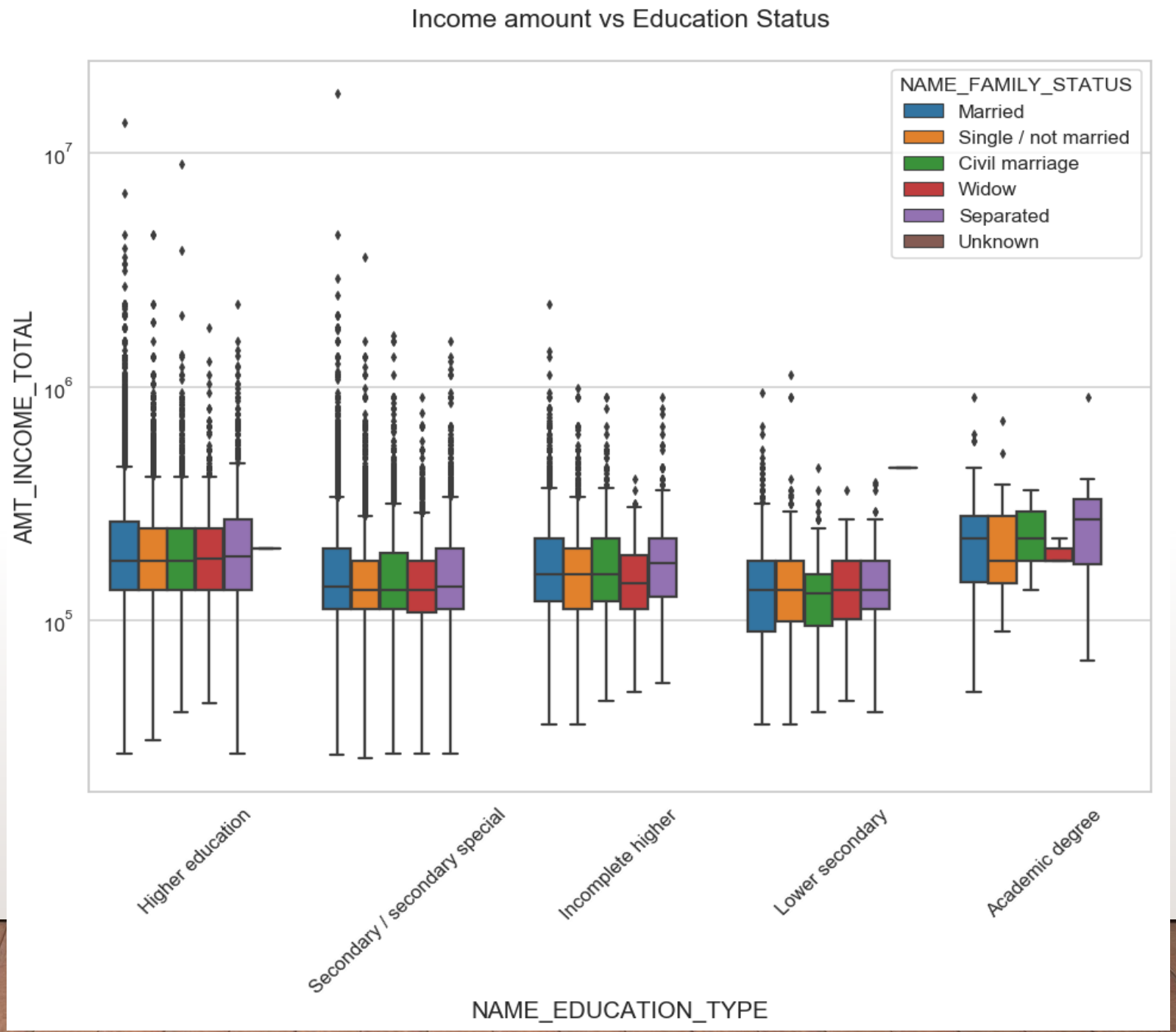
CREDIT AMOUNT VS EDUCATION STATUS

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



INCOME AMOUNT VS EDUCATION STATUS

- Have some similarities with Target0, From the above boxplot for Education type 'Higher education' the income amount is mostly equal with family status.
- Fewer outliers is having for Academic degree but their income amount is a little higher than Higher education.
- Lower secondary are have less income amount than others.
-

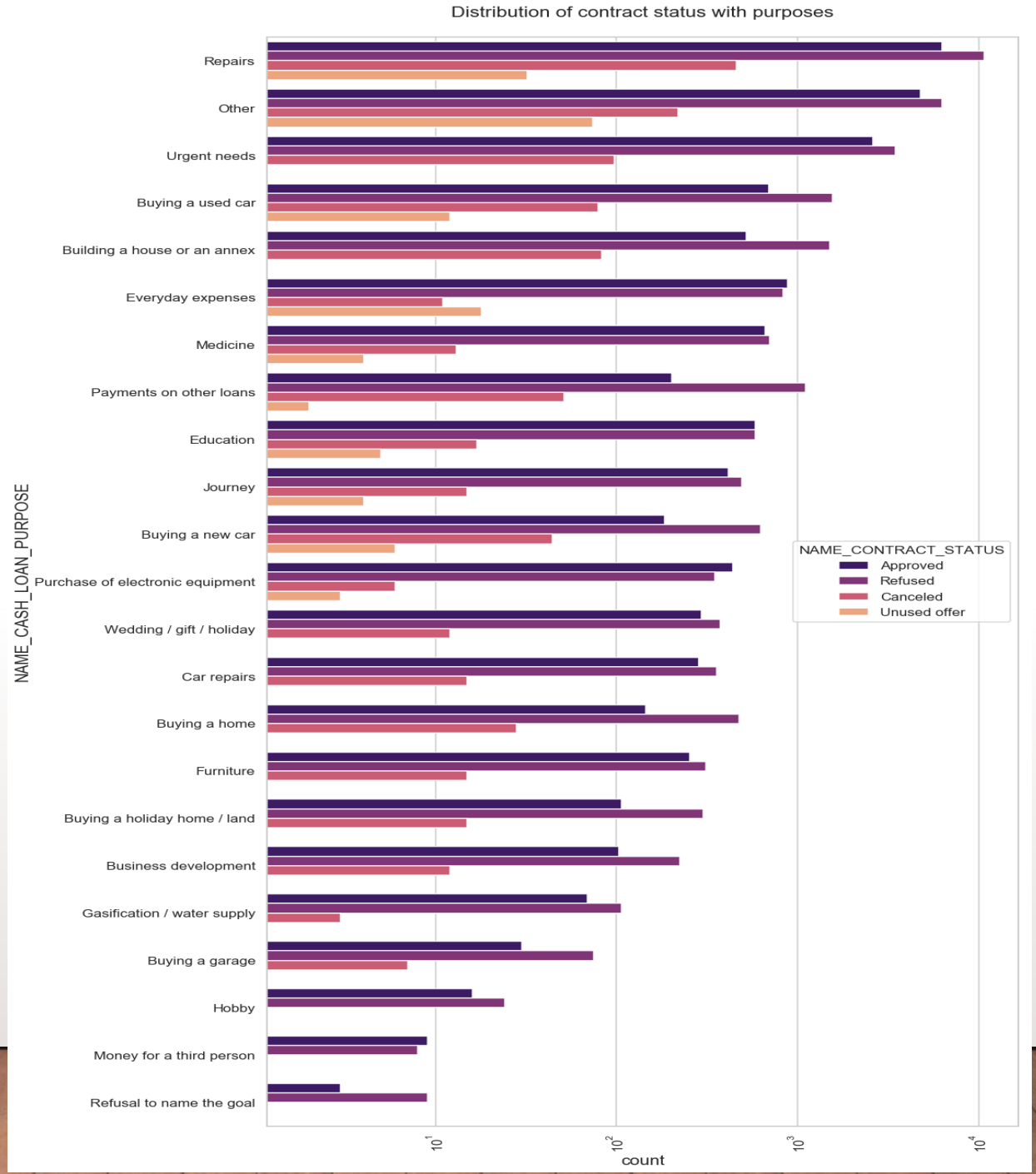


UNIVARIATE ANALYSIS AFTER MERGING PREVIOUS DATA

DISTRIBUTION OF CONTRACT STATUS WITH PURPOSES

Some points extracted from the graph;-

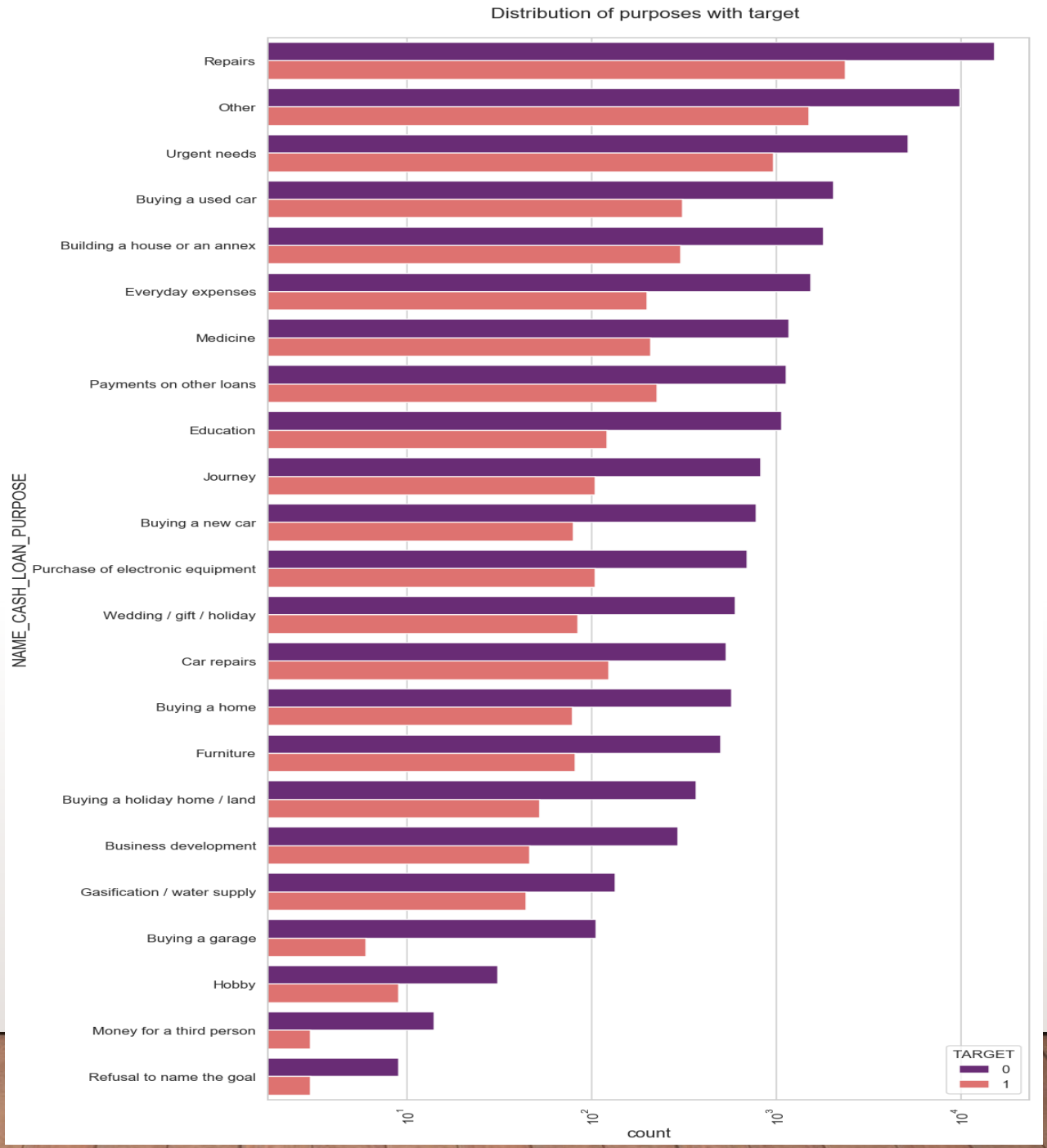
- Most rejections of loans came from purpose 'repairs'.
- For education purposes, we have an equal number of approved and rejection
- Paying other loans and buying a new car is having significantly higher rejections than approves.



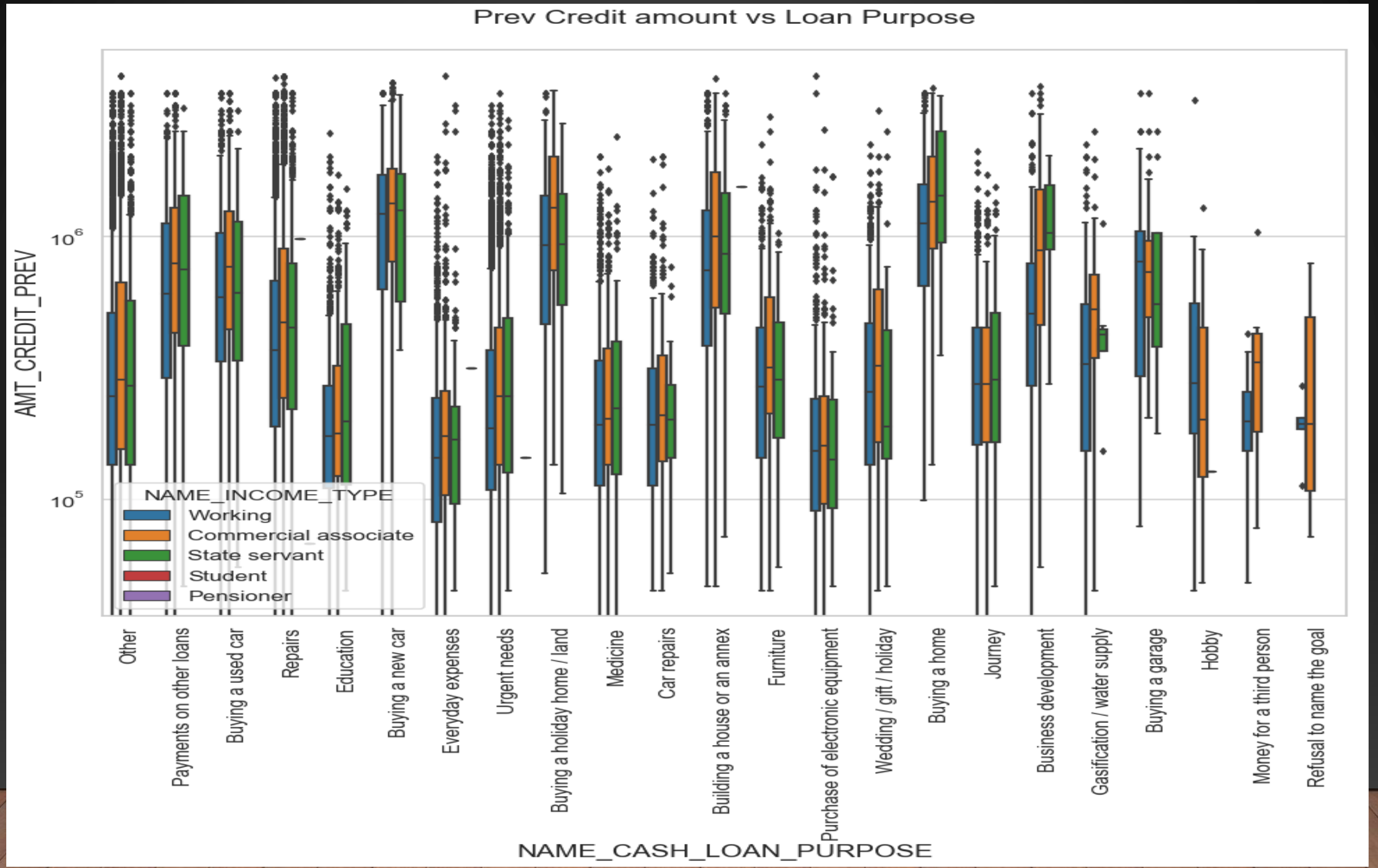
DISTRIBUTION OF PURPOSES WITH TARGET

Some points extracted from the graph;-

- 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 client is having minimal payment difficulties.



PERFORMING 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

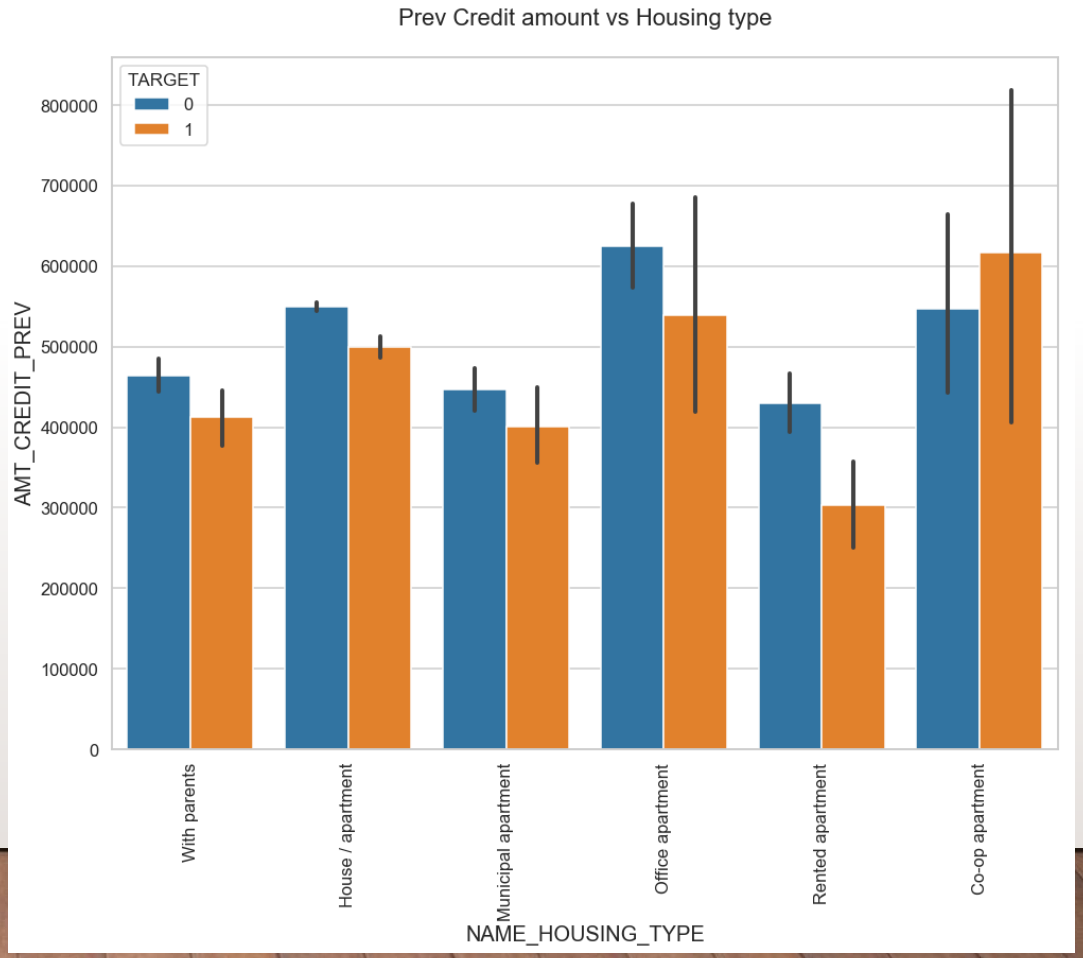
From the previous graph we can conclude the below points:

- The credit amount for Loan purposes like 'Buying a home’, ’Buying 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 the third person or a Hobby is having fewer credits applied for.

PREVIOUS CREDIT AMOUNT VS HOUSING TYPE

Some Concluded points from the plotted graph-

- Here for the Housing type, office apartments is having higher credit of target 0 and co-op apartments is having higher credit of target 1.
- So, we can conclude that banks should avoid giving loans to the housing type of co-op apartments as they are having payment difficulties.
- 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 types ‘Student’,’ pensioner’, and ‘Businessman’ with housing ‘types other than ‘Co-op apartment’ for successful payments.
- Banks should focus less on income type ‘Working’ as they are having the most number of unsuccessful payments.
- Also with loan purposes ‘Repair’ is having a higher number of unsuccessful payments on time.
- Get as many clients from housing type ‘With parents’ as they are having the least number of unsuccessful payments.

THANK YOU