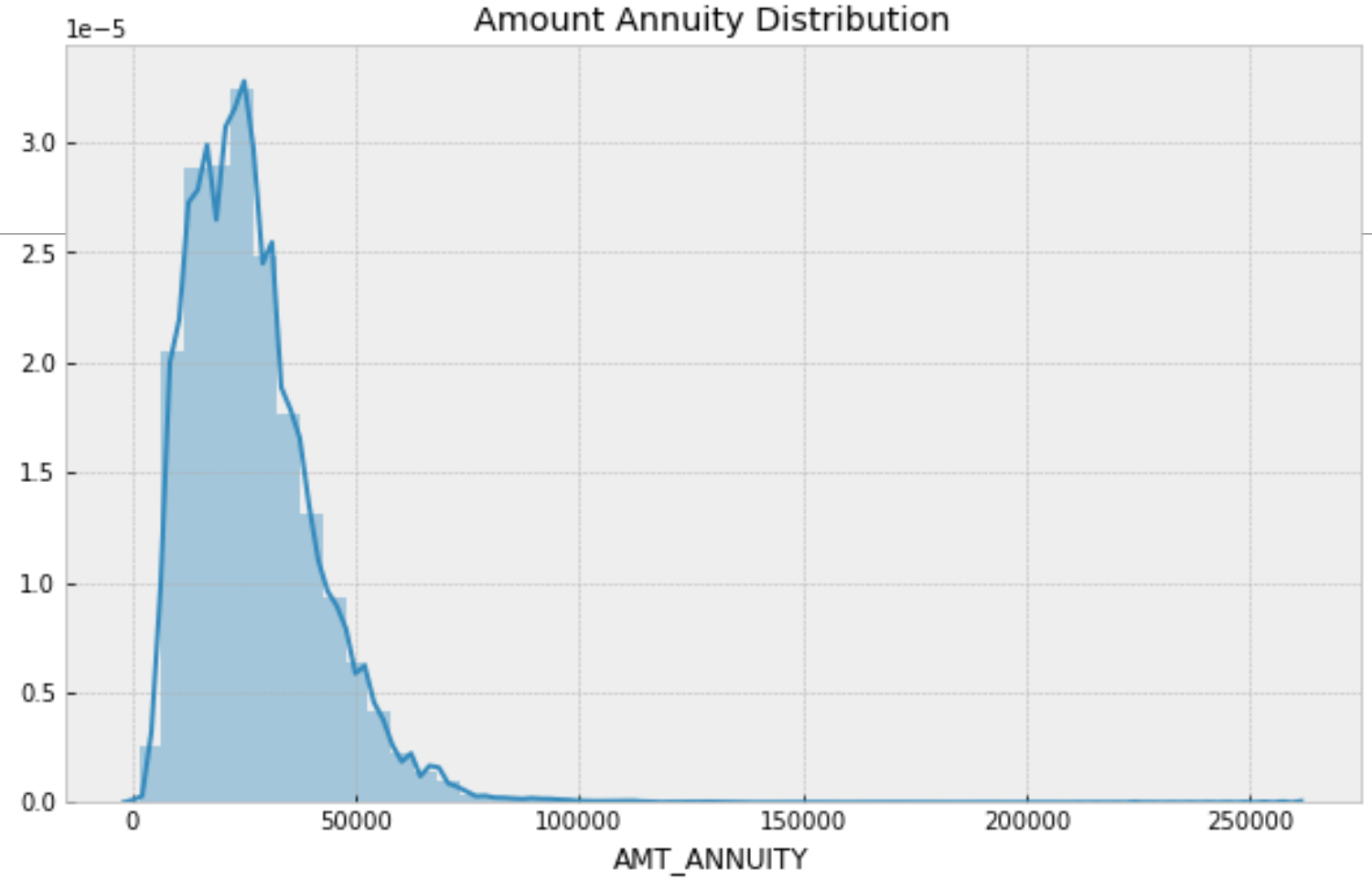
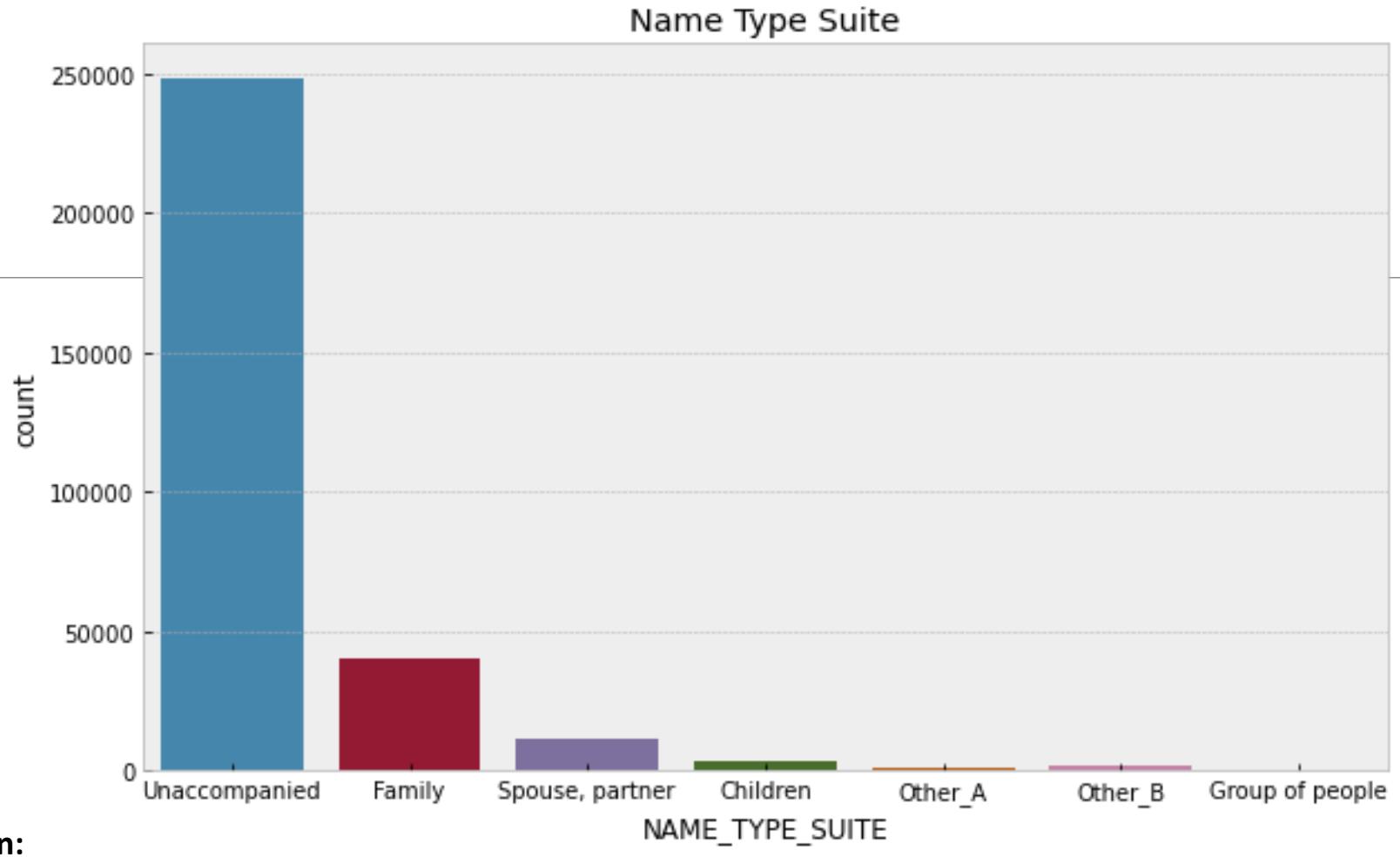
CREDIT EDA CASE STUDY

BY E.S.V.SRIRAM AND CH.SAI KIRAN

Data Cleaning

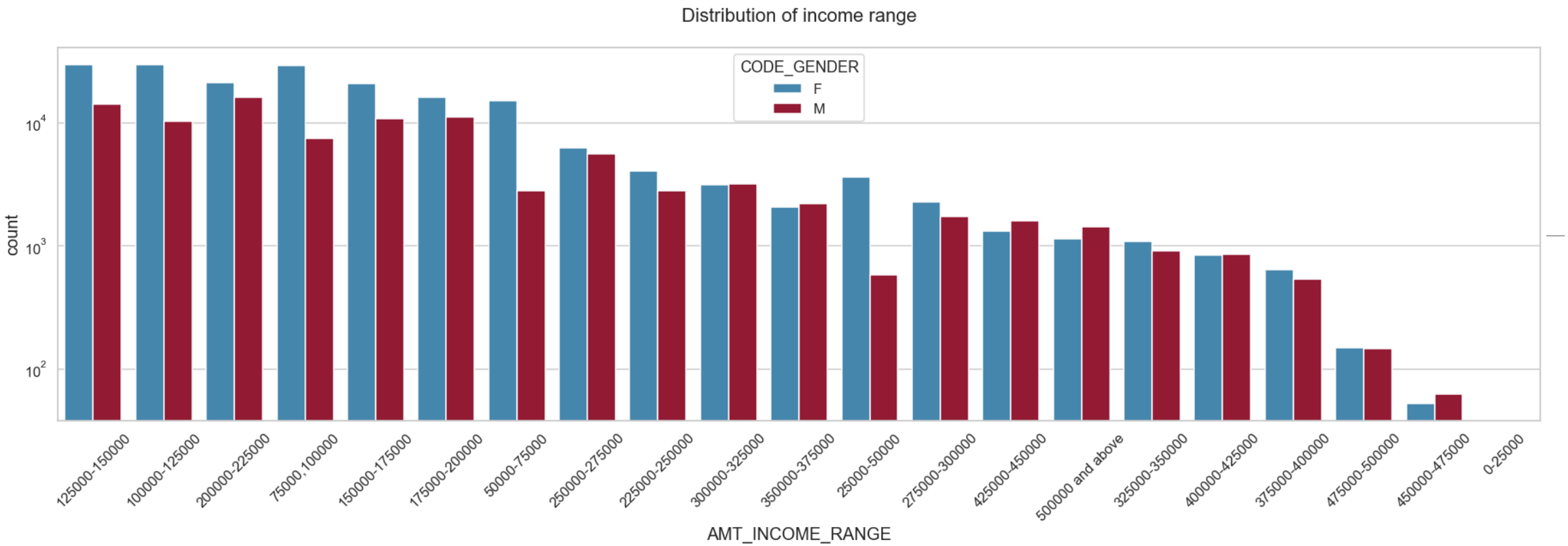


- **Observation :**
- AMT_ANNUITY is right skewed since it has outliers. So imputing the missing values with median seems like an ideal choice.



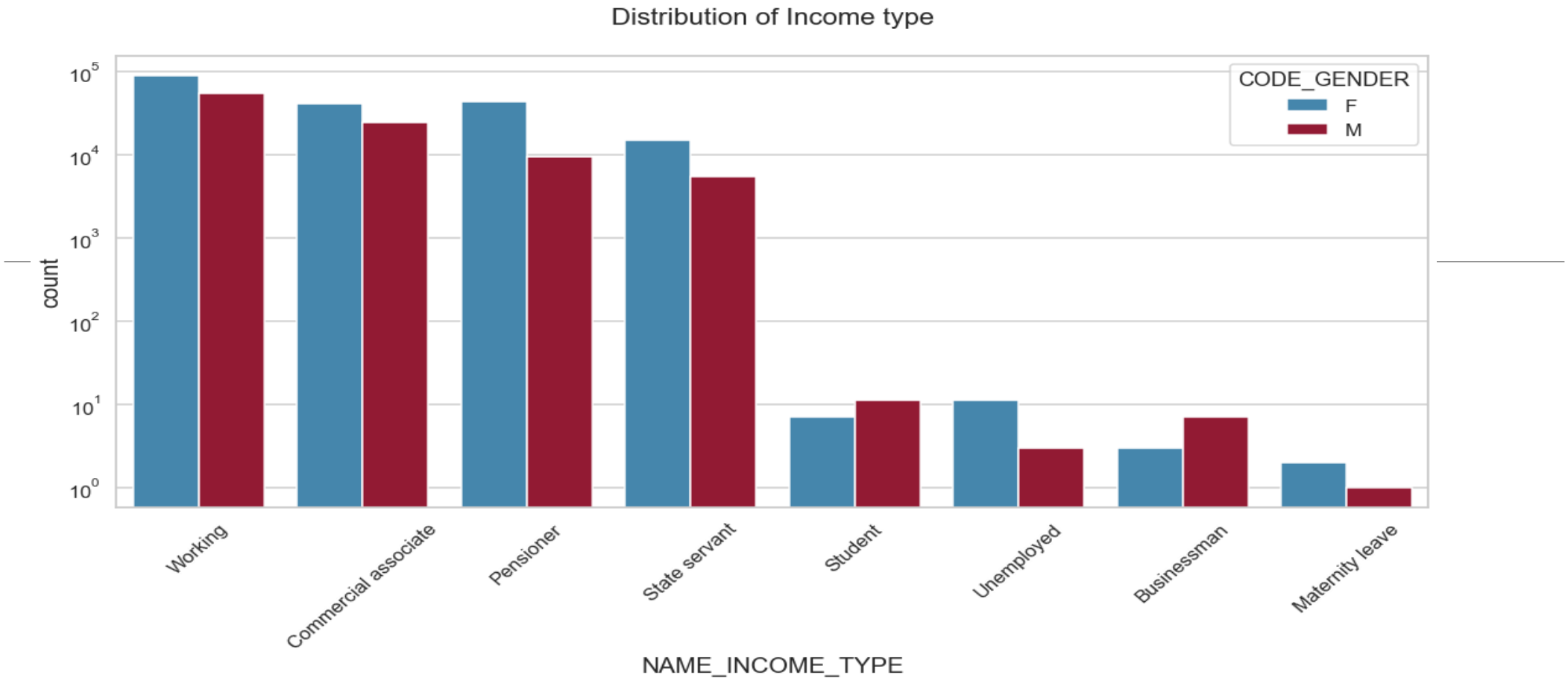
- **Observation:**
- NAME_TYPE_SUITE is categorical column, and it has missing values. So imputiing the missing values with mode seems like an ideal choice

Categorical Univariate analysis for target 0



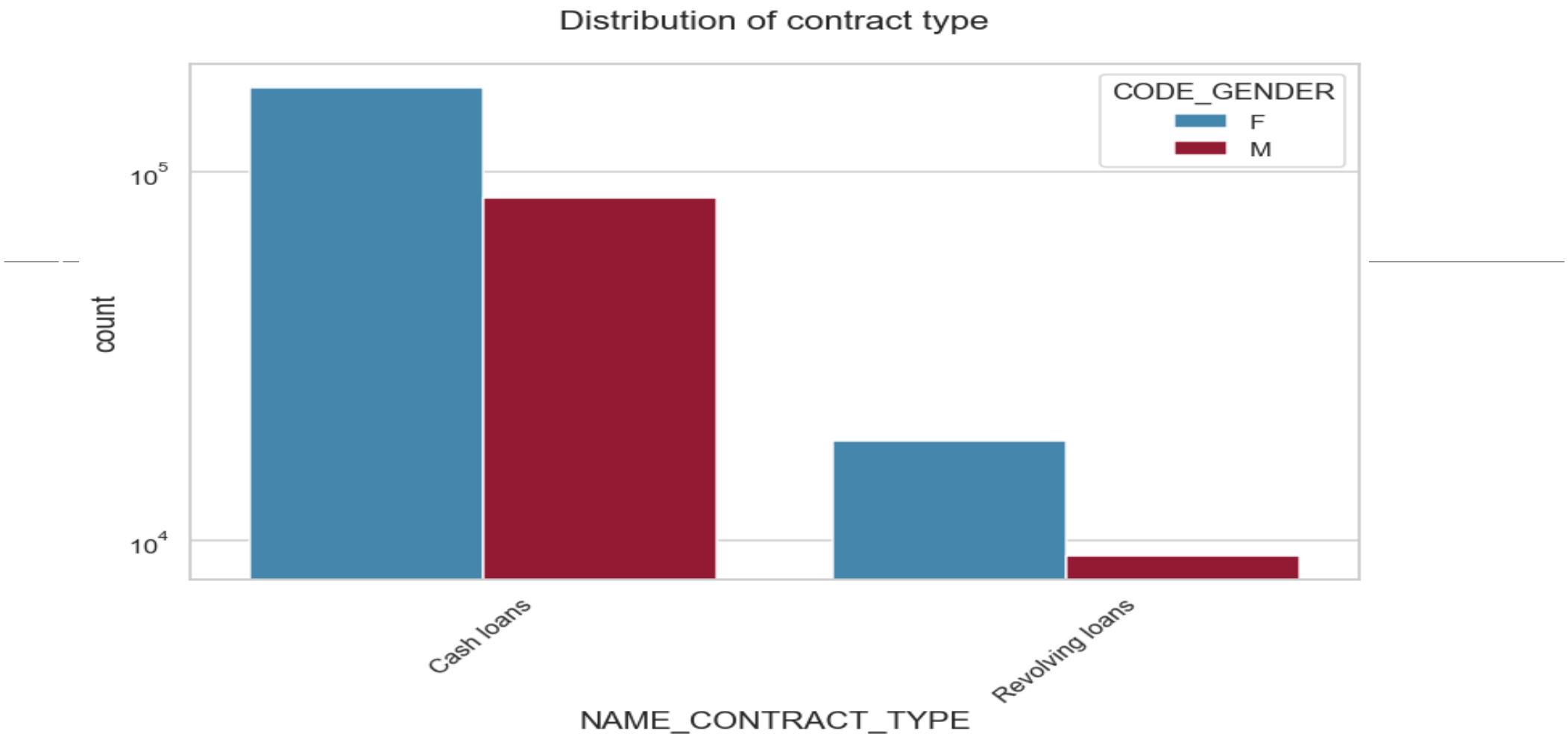
Observations:

1. On an average Female frequency is higher than male.
2. 0-25000 range has almost zero credits.
3. Income range from 75000 to 200000 has most number of credits.
4. Very less count/credits for the income range 400000 and above.



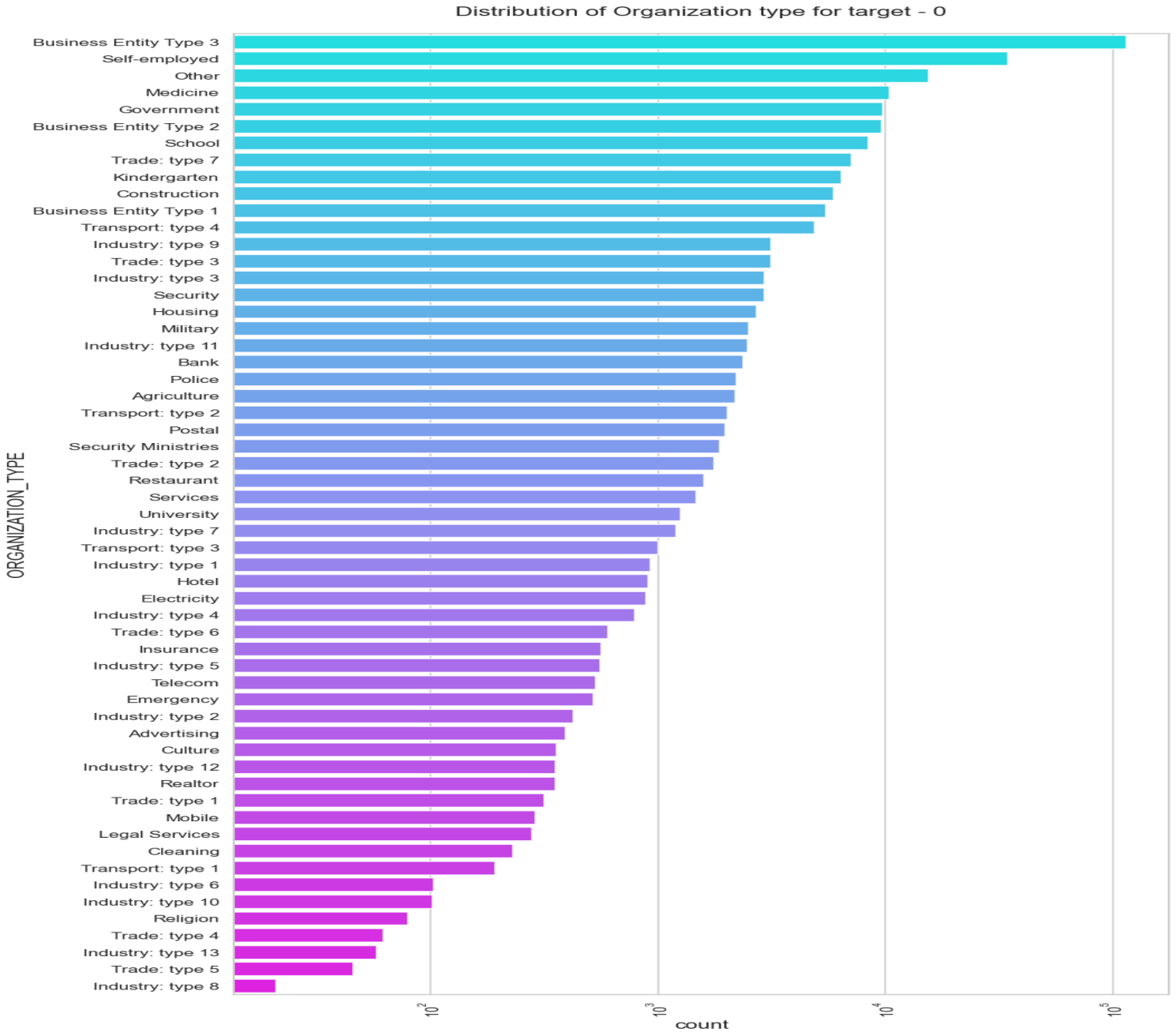
Observations:

1. For income type ‘working’, ’commercial associate’, 'pensioner' and ‘State Servant’ the number of credits are higher than others.
2. On an average the frequency of female is more than male.
3. Less number of credits for income type ‘Maternity leave’, ‘Businessman’, 'Student' and ‘Unemployed'.



Observations:

1. ‘Cash loans’ have higher number of credits than ‘Revolving loans’.
2. Female frequency is more than male.

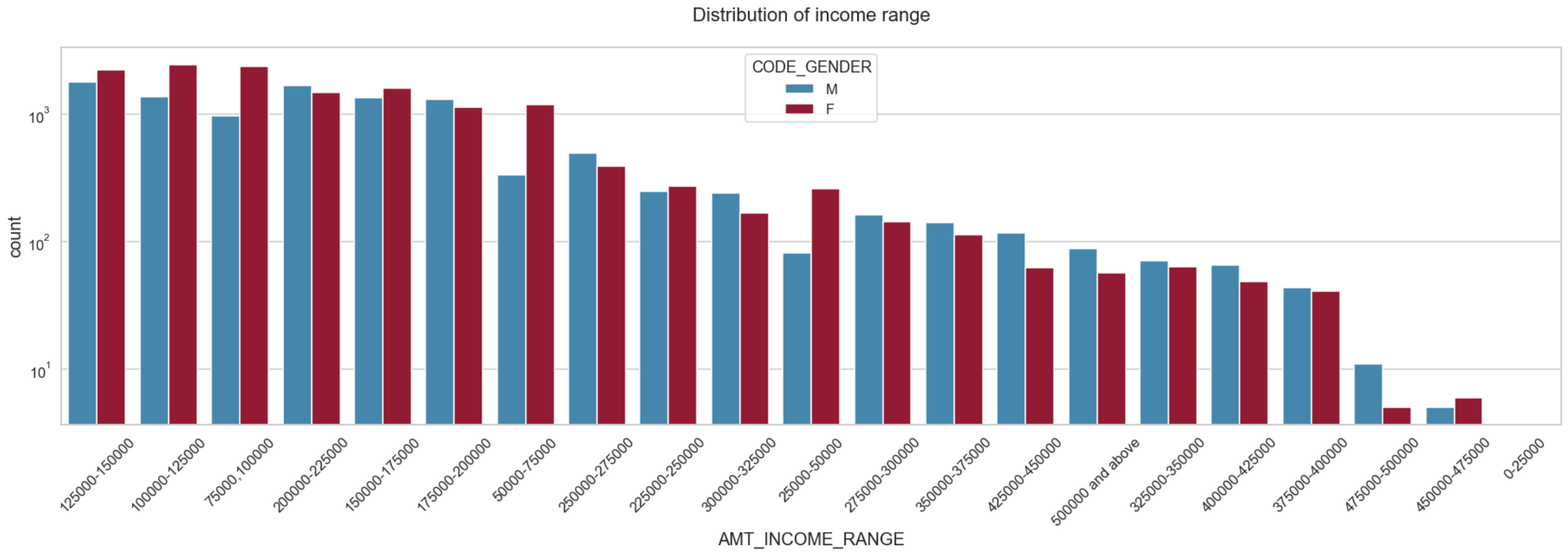


Observations:

1. Clients which have applied for credits are from most of the organization type ‘Business entity Type 3’ , ‘Self employed’, ‘Other’ , ‘Medicine’ and ‘Government’.

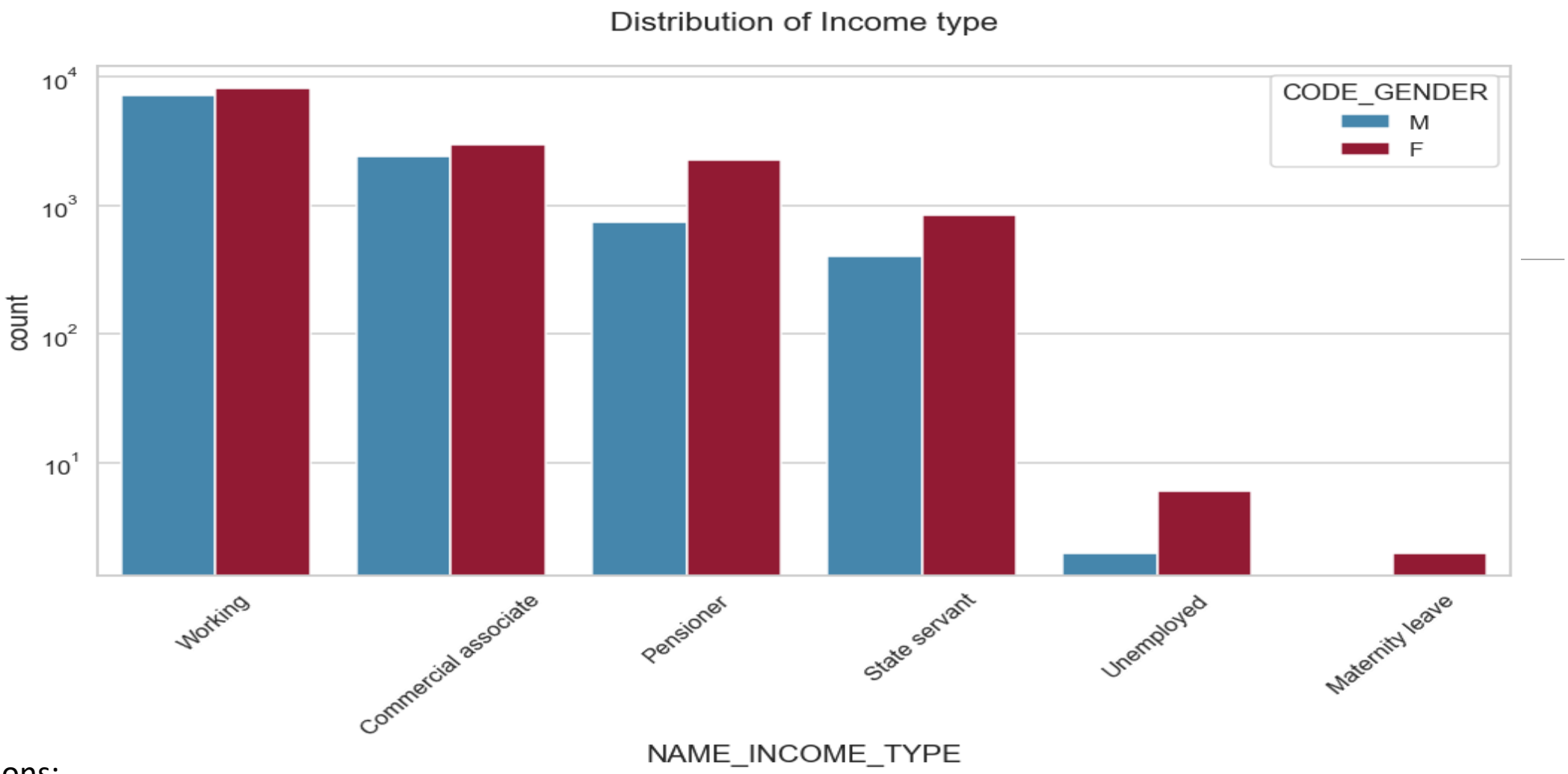
2. Clients for less credits are from Industry type 8,type 6, type 10, religion and trade type 5, type 4.

Categorical Univariate analysis for target 1



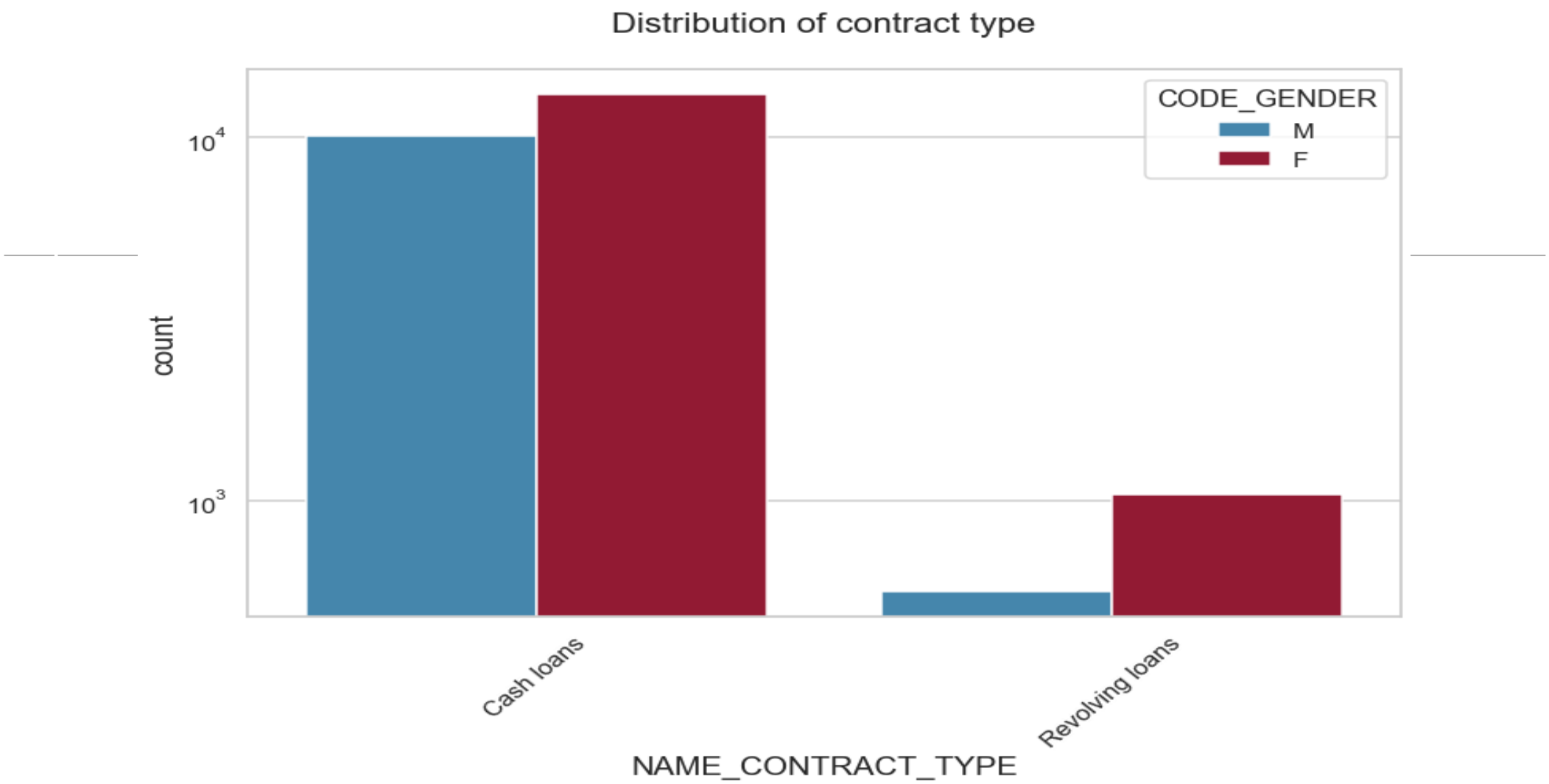
Observations:

1. Male defaulters are higher than female.
2. Income range from 75000 to 125000 is having more number of defaulters.
3. Very less number of defaulters for income range 400000 and above.



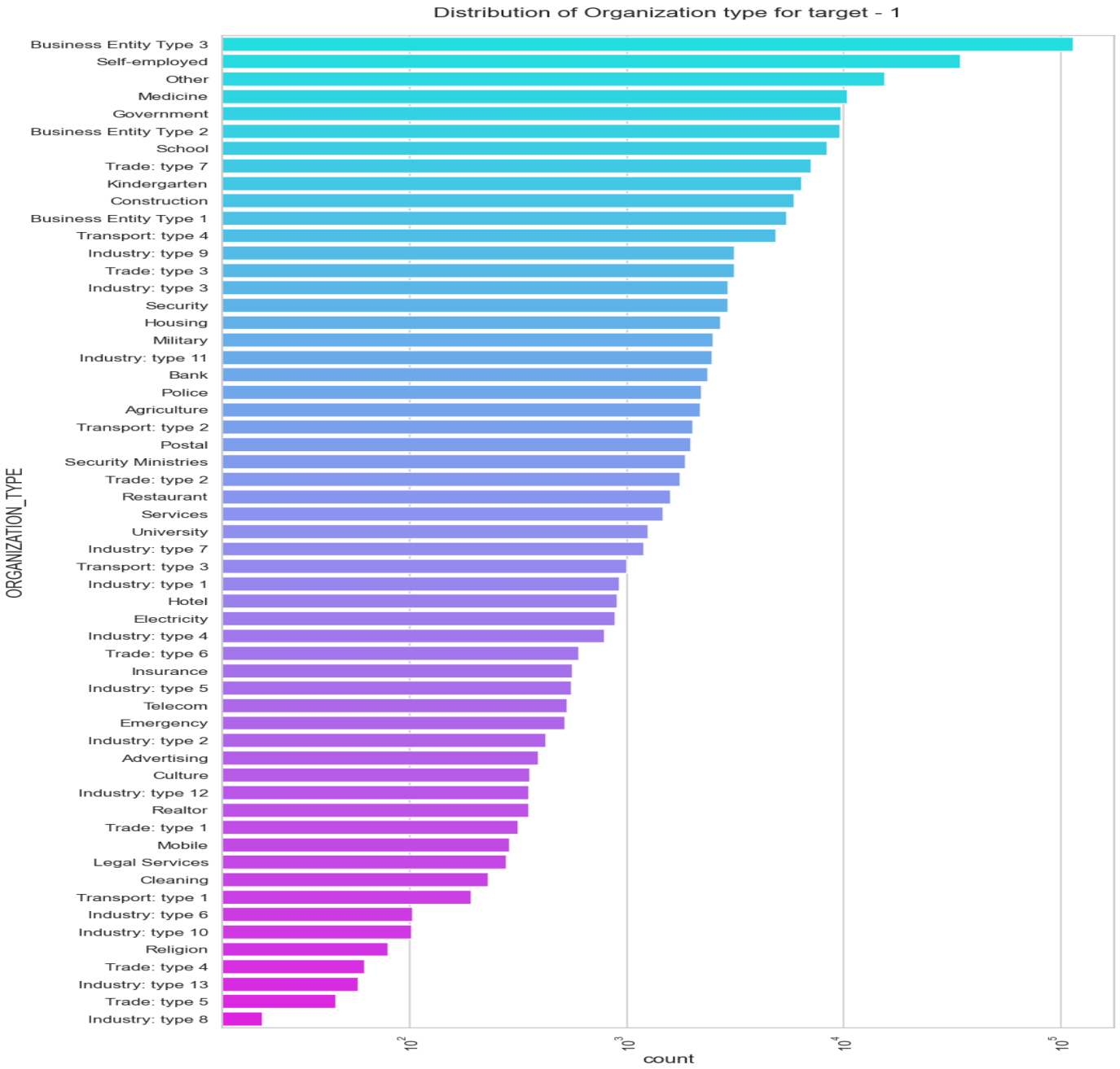
Observations:

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



Observations:

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 type 1 : there is only Female Revolving loans.



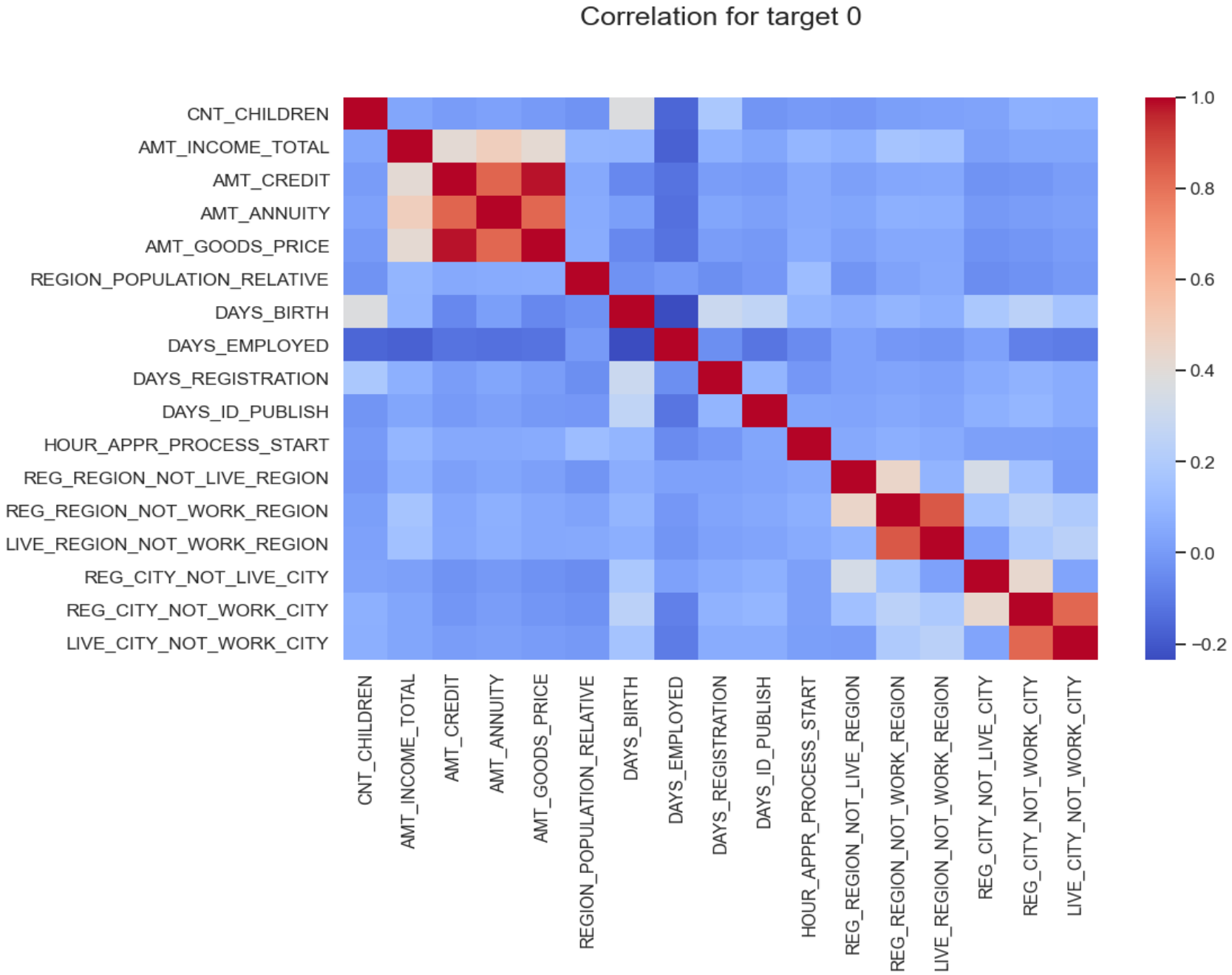
<u>Observations:</u>

1. Clients which have 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.

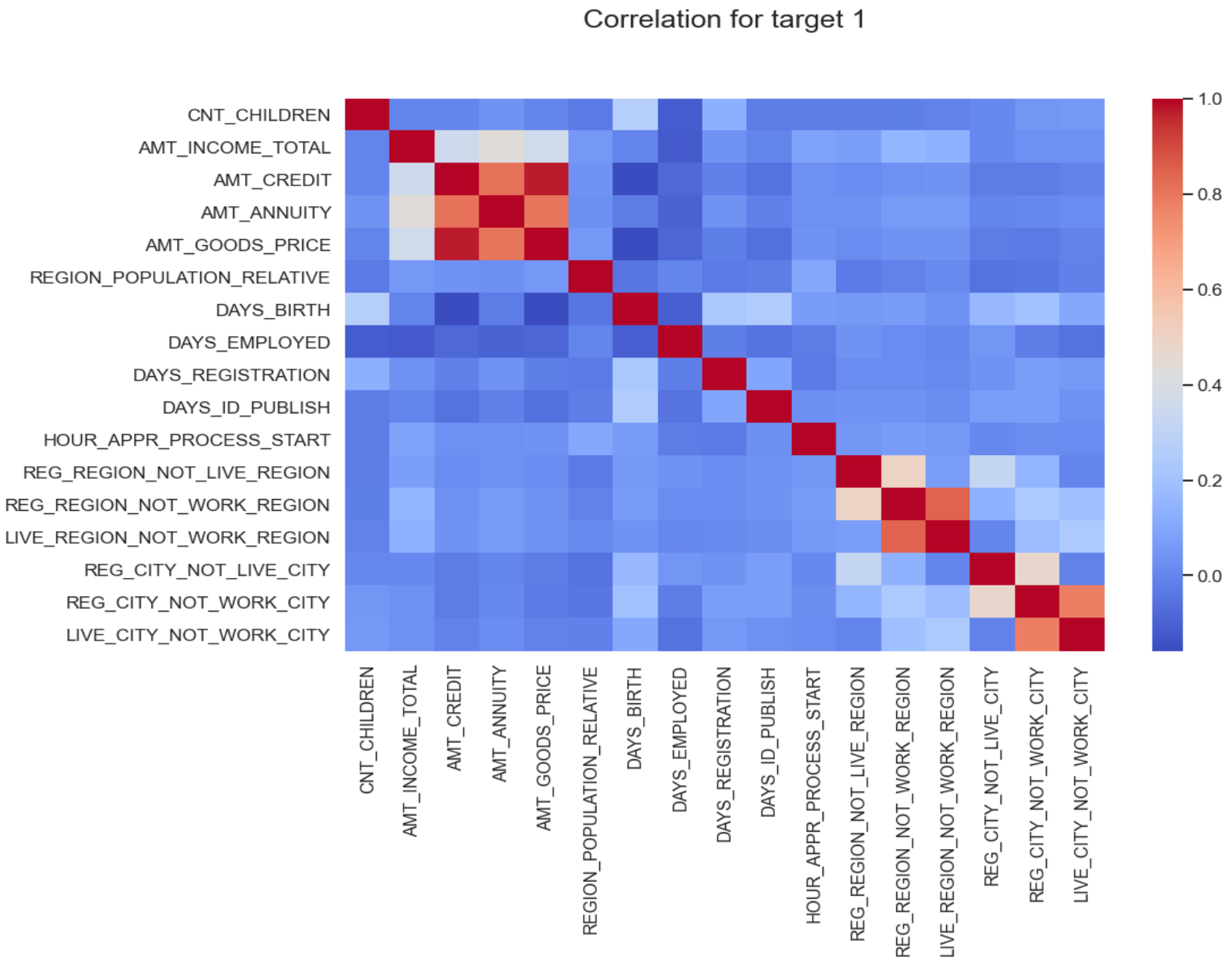
Correlation of target 0

Observations:

1) Credit amount is inversely proportional to the days 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 a client has which means Credit amount is higher for less children count client have and vice-versa.
3) Income amount is inversely proportional to the number of children the client has.
4) Clients with less children are more in densely populated area.
5) Credit amount is higher in densely populated area.
6) Income is also higher in densely populated area.



Correlation For target 1



Observations:

1. Amount credit and amount goods price are directly proportional
2. The client's permanent address does not match with the contact address are having less children and vice-versa
3. The client's permanent address does not match work address are having less children and vice-versa

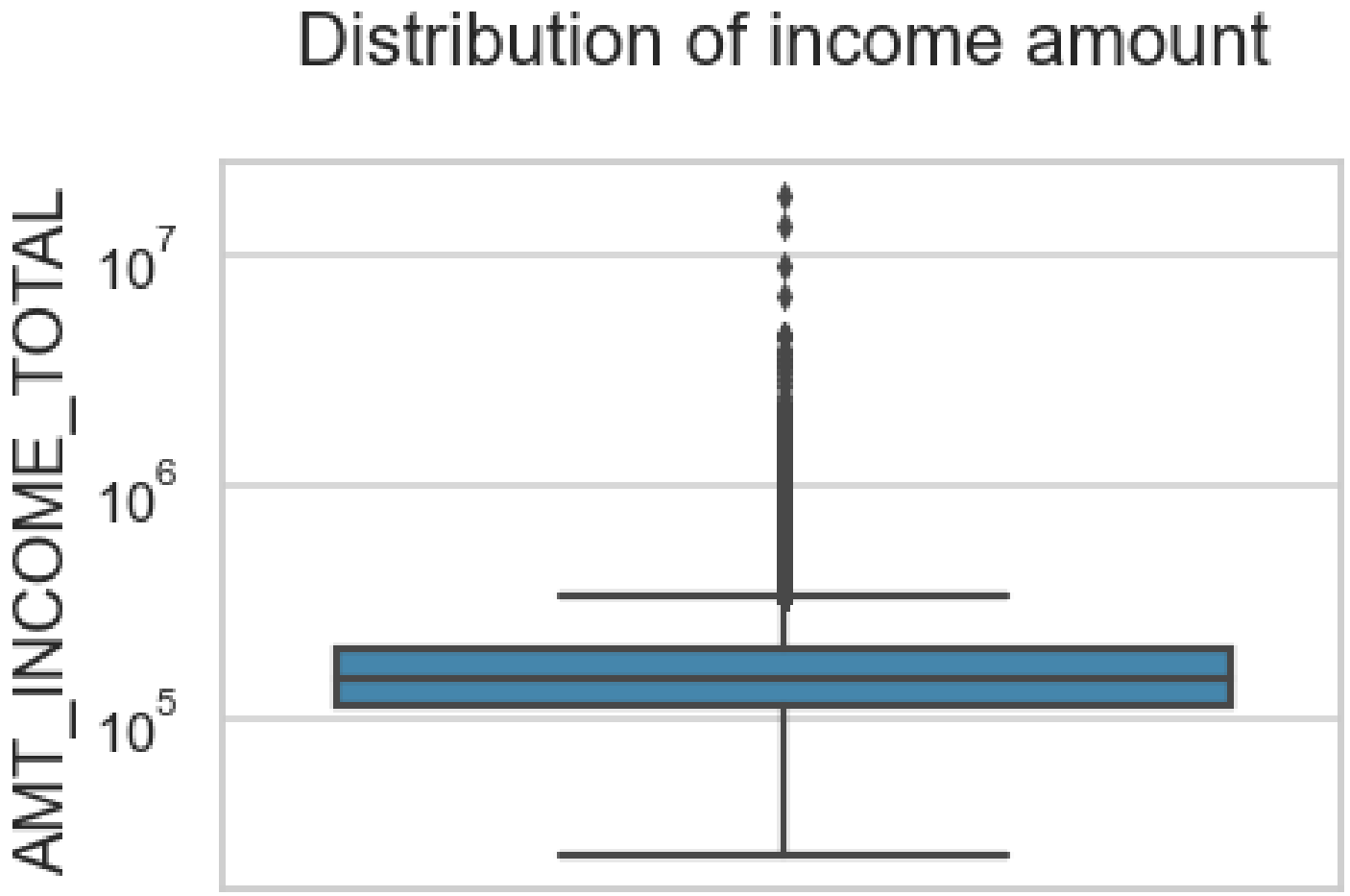
Categorical Univariate analysis for variables target 0

Boxplot for income amount

Observations:

1. Some outliers are noticed in income amount.

2. The third quartile range is very narrow.

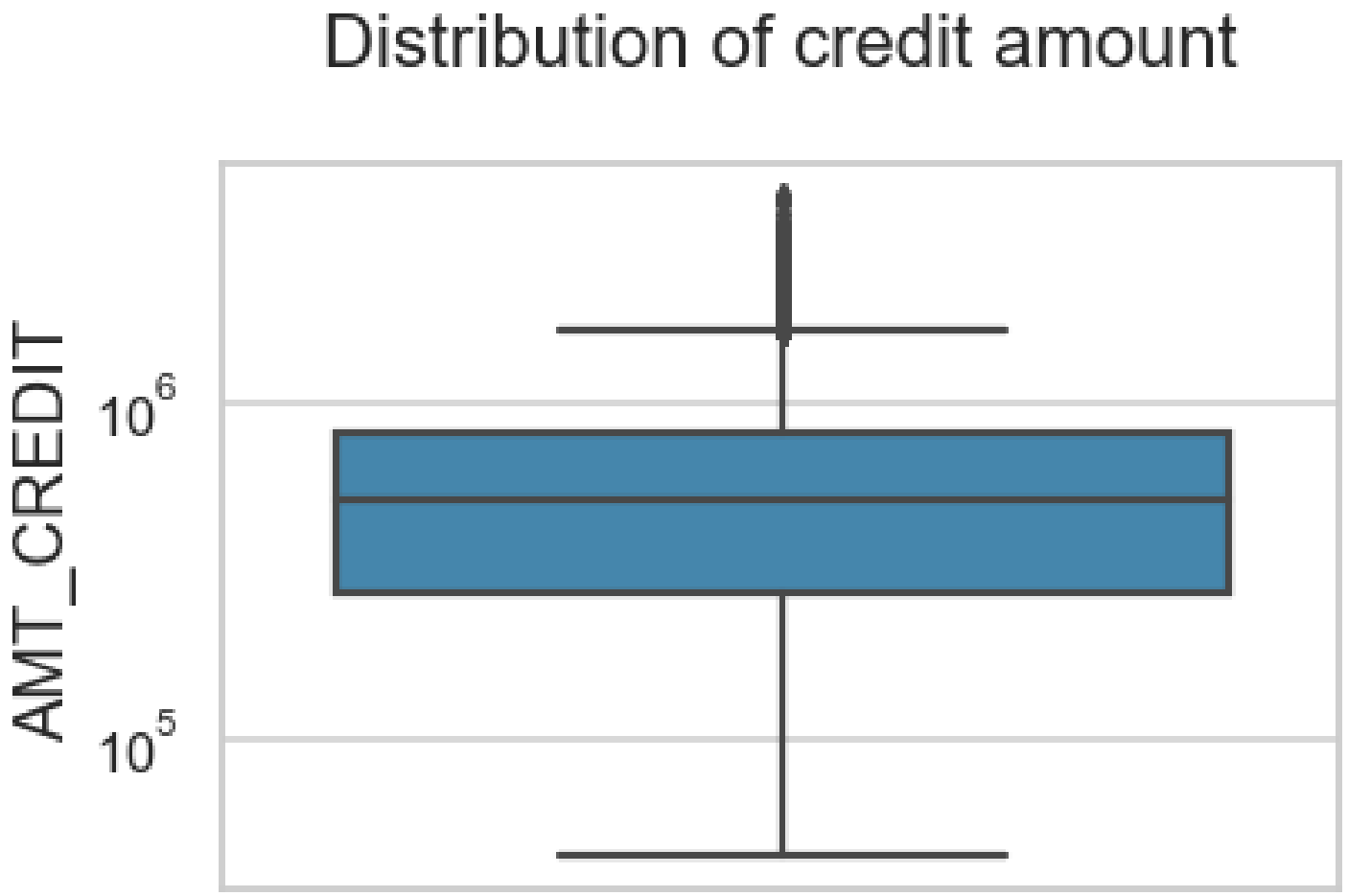


Boxplot for credit amount

Observations:

1. Some outliers are noticed in credit amount.

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

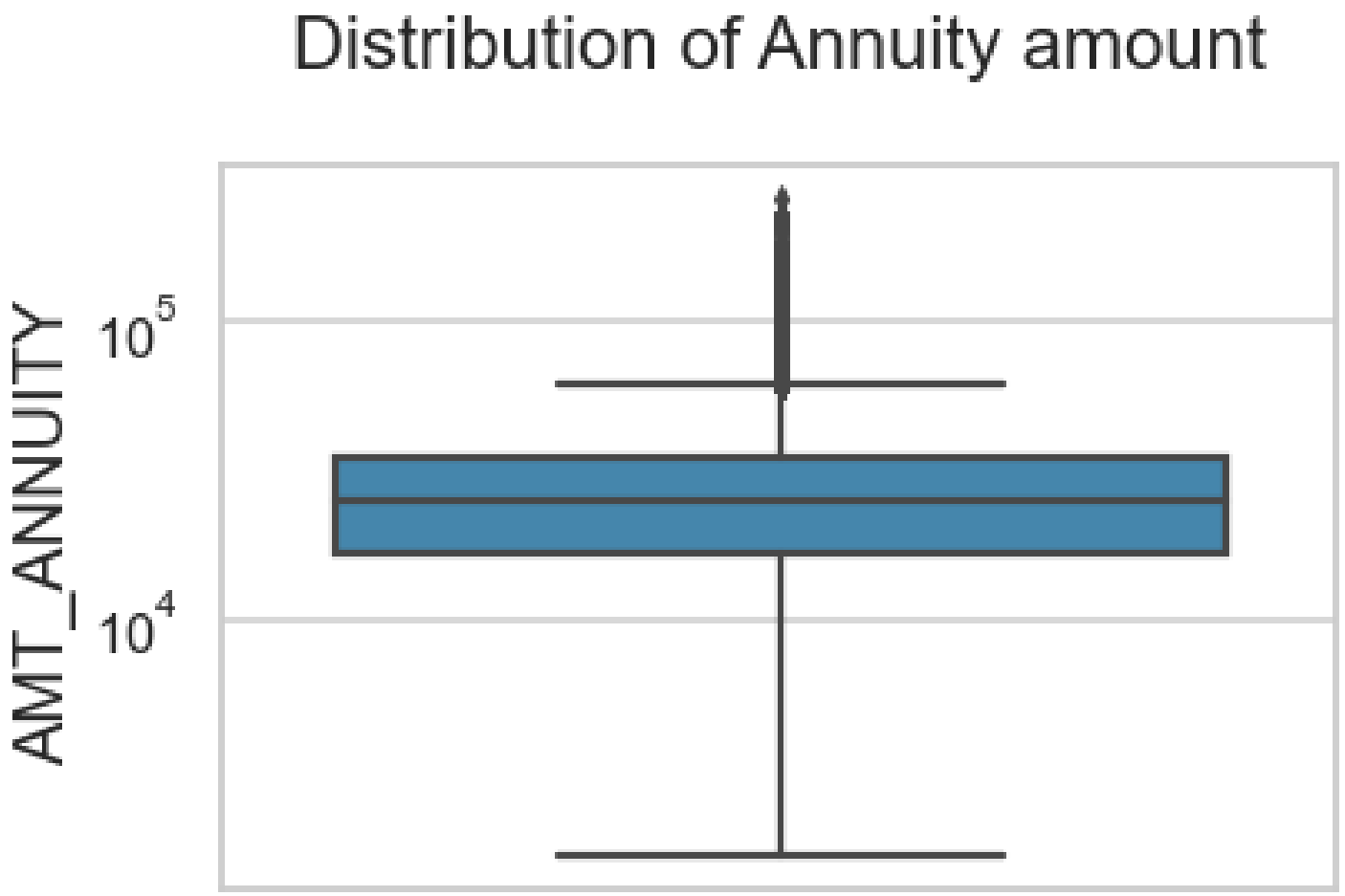


Boxplot for annuity amount

Observations:

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 target 1

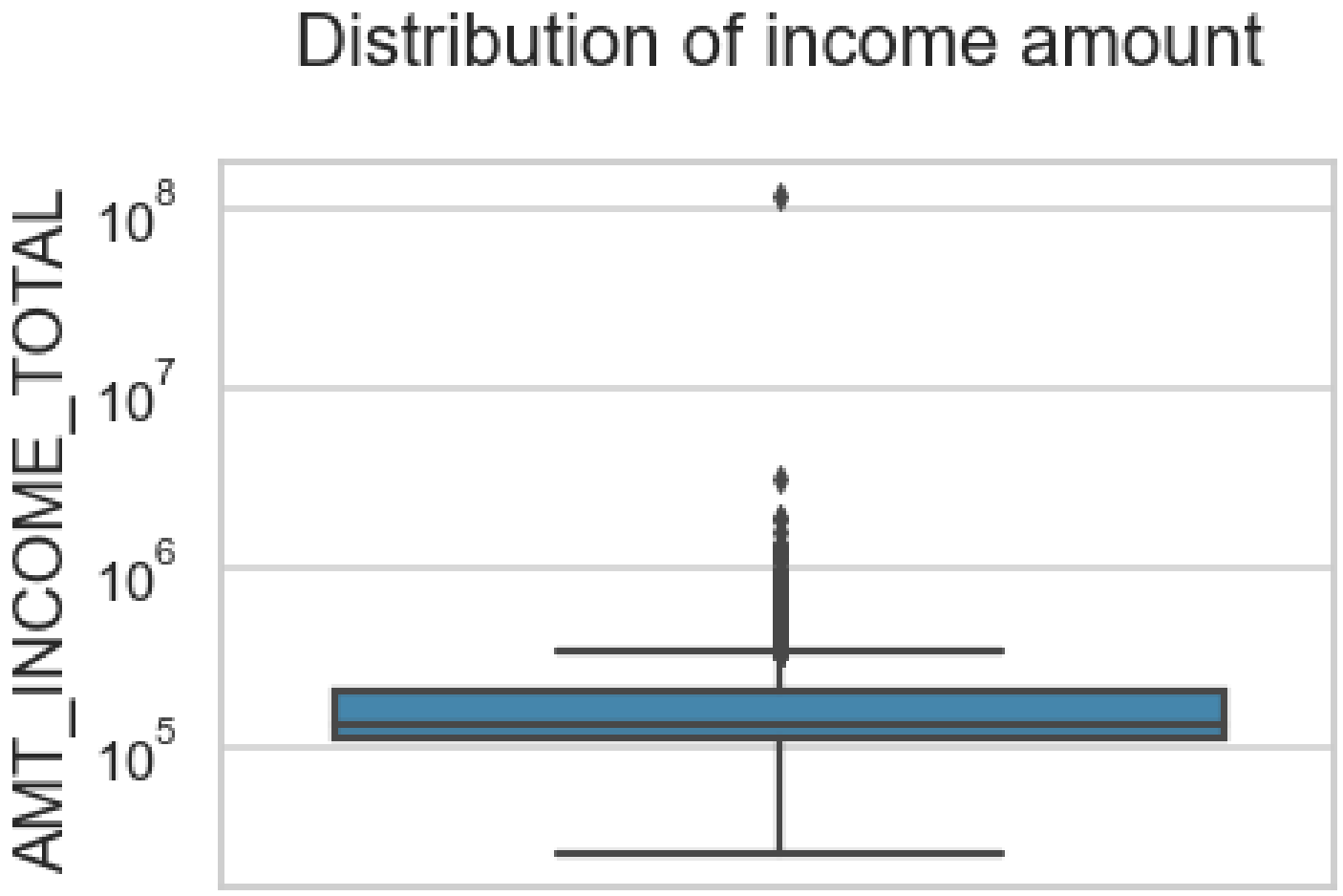
Boxplot for income amount

Observations:

1. Some outliers are noticed in income amount.

2. The third quartile is very narrow for income amount.

3. Most of the clients of income are present in first quartile.

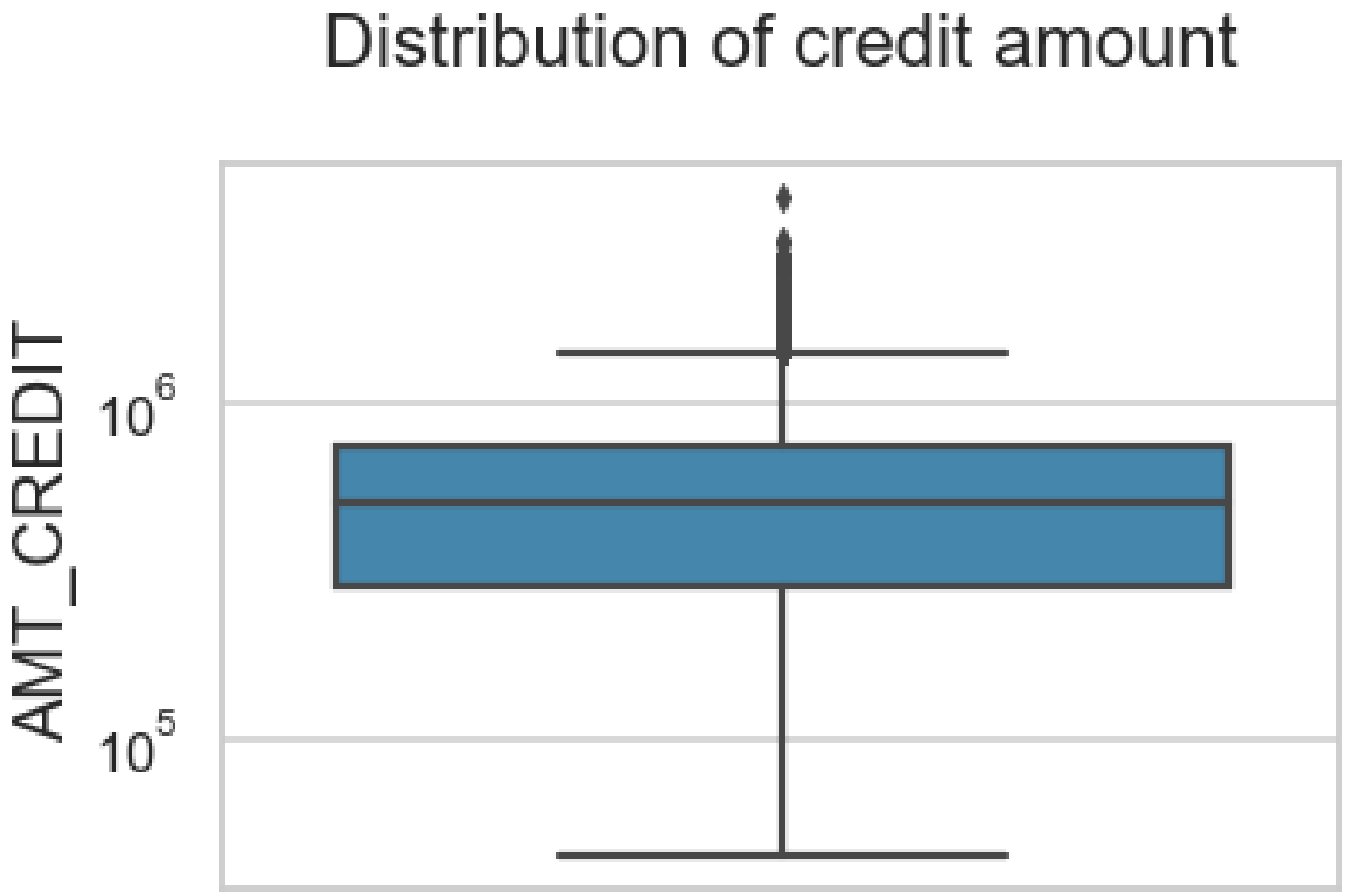


Boxplot for credit amount

Observations:

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.

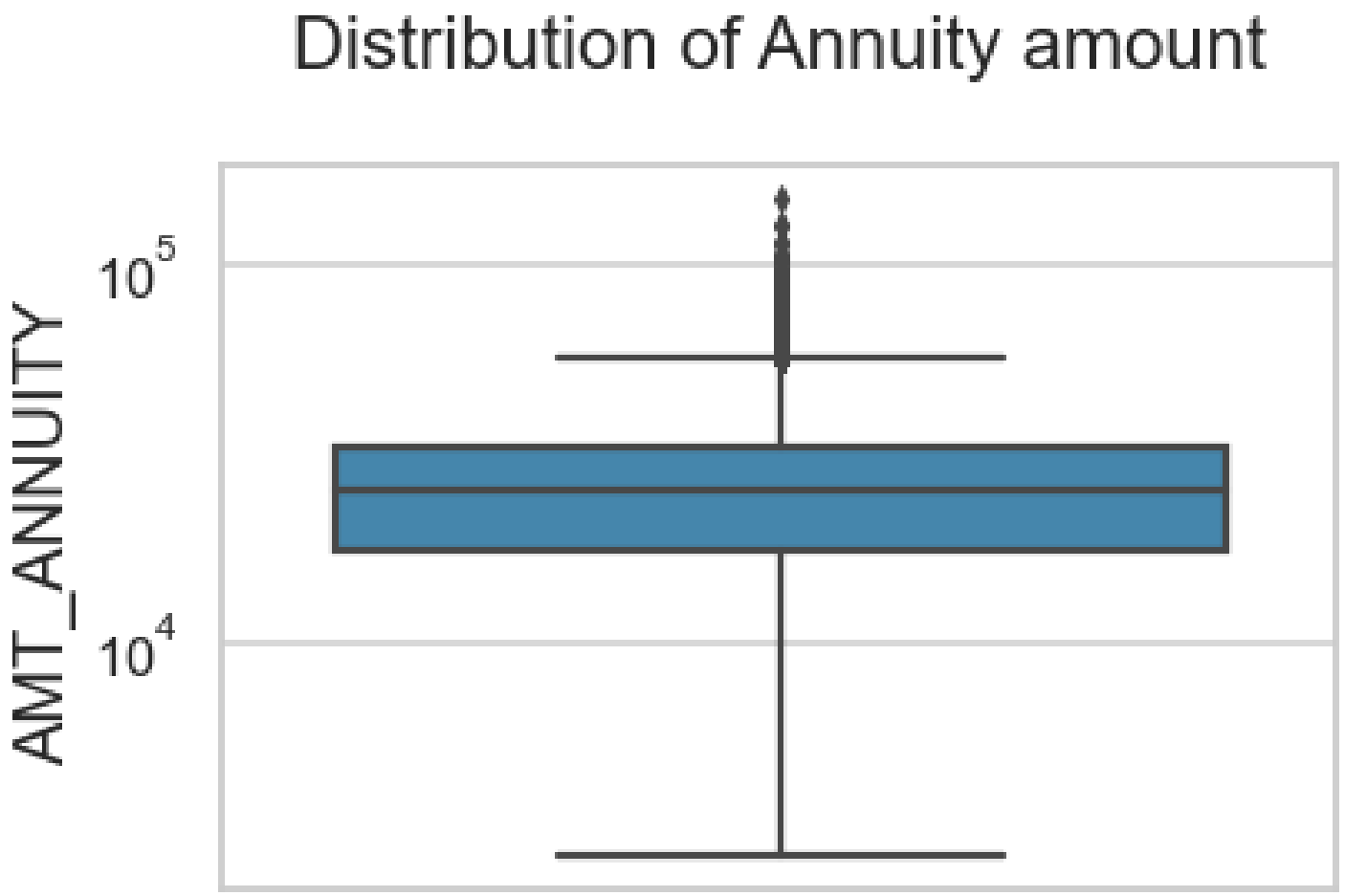


Boxplot for annuity amount

Observations:

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.



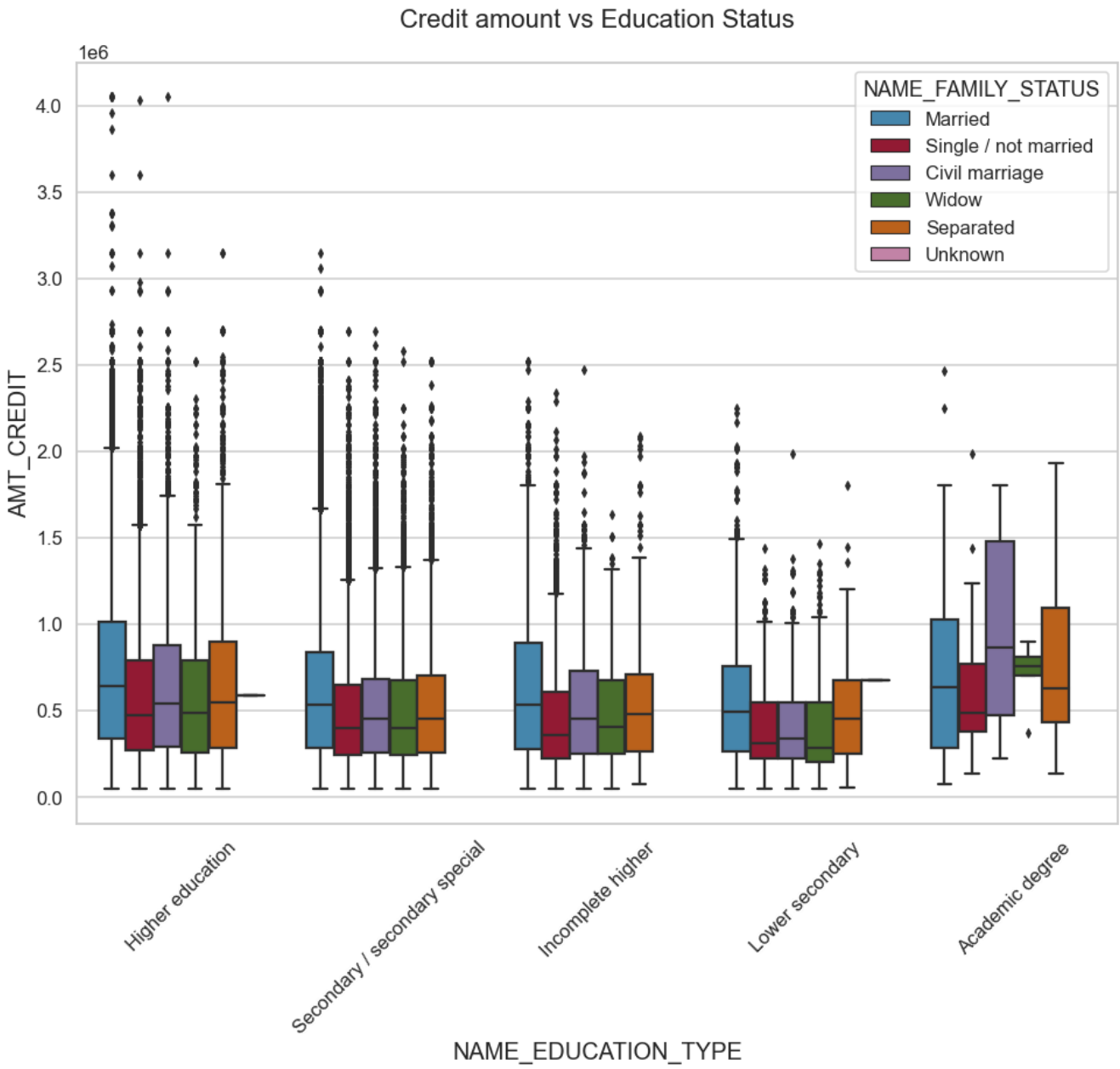
Bivariate analysis for target 0

Credit amount vs Education Status

Observations:

1. From the above box plot we can conclude that Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others.

2. Also, 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..

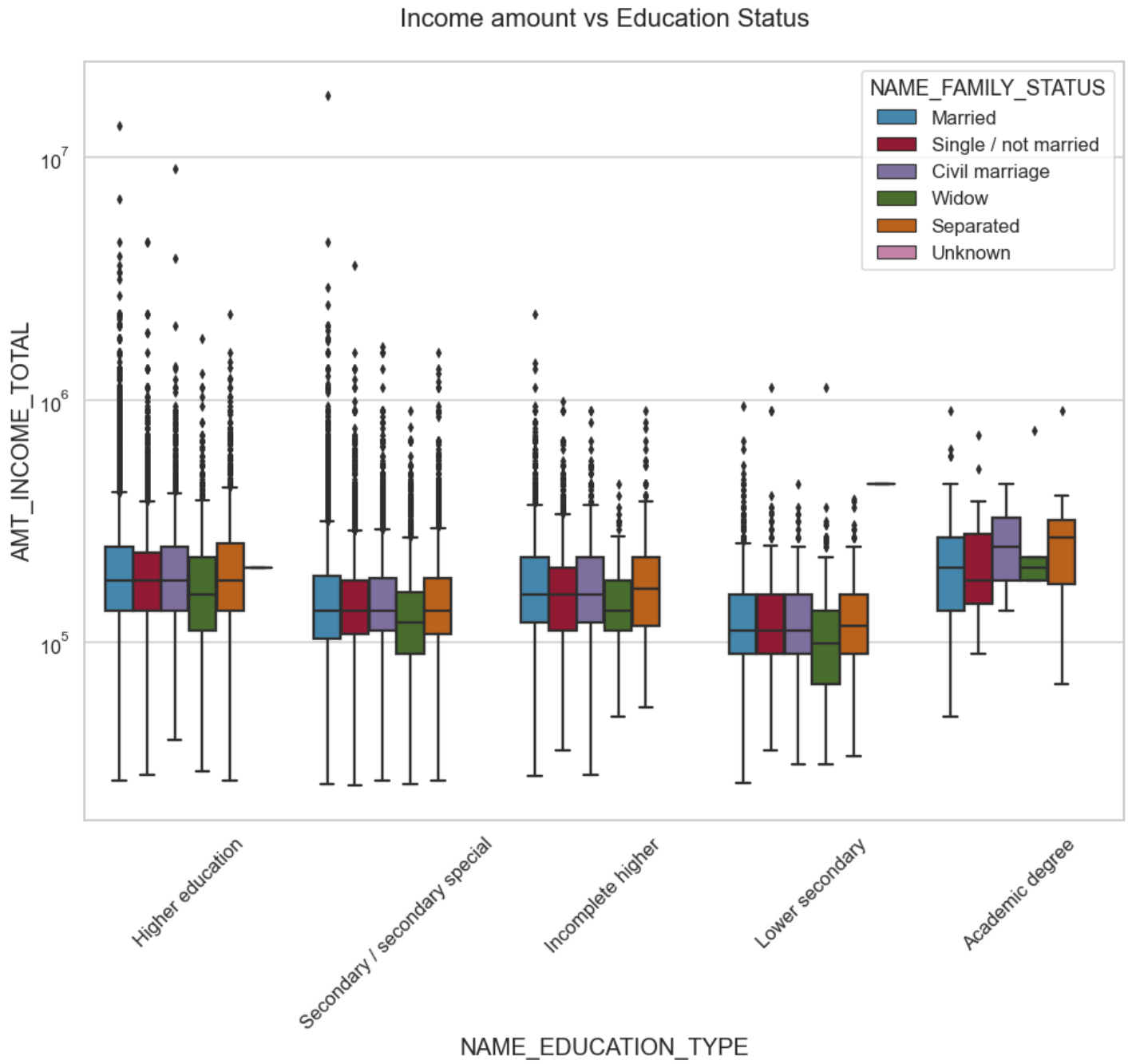


Income amount vs Education Status

<u>Observations:</u>

1. From above boxplot for Education type 'Higher education' the income amount is mostly equal with family status. It does contain many outliers.

2. We see less outlier are present for Academic degree but the income amount is little higher that Higher education. Lower secondary of civil marriage family status are have less income amount than others.



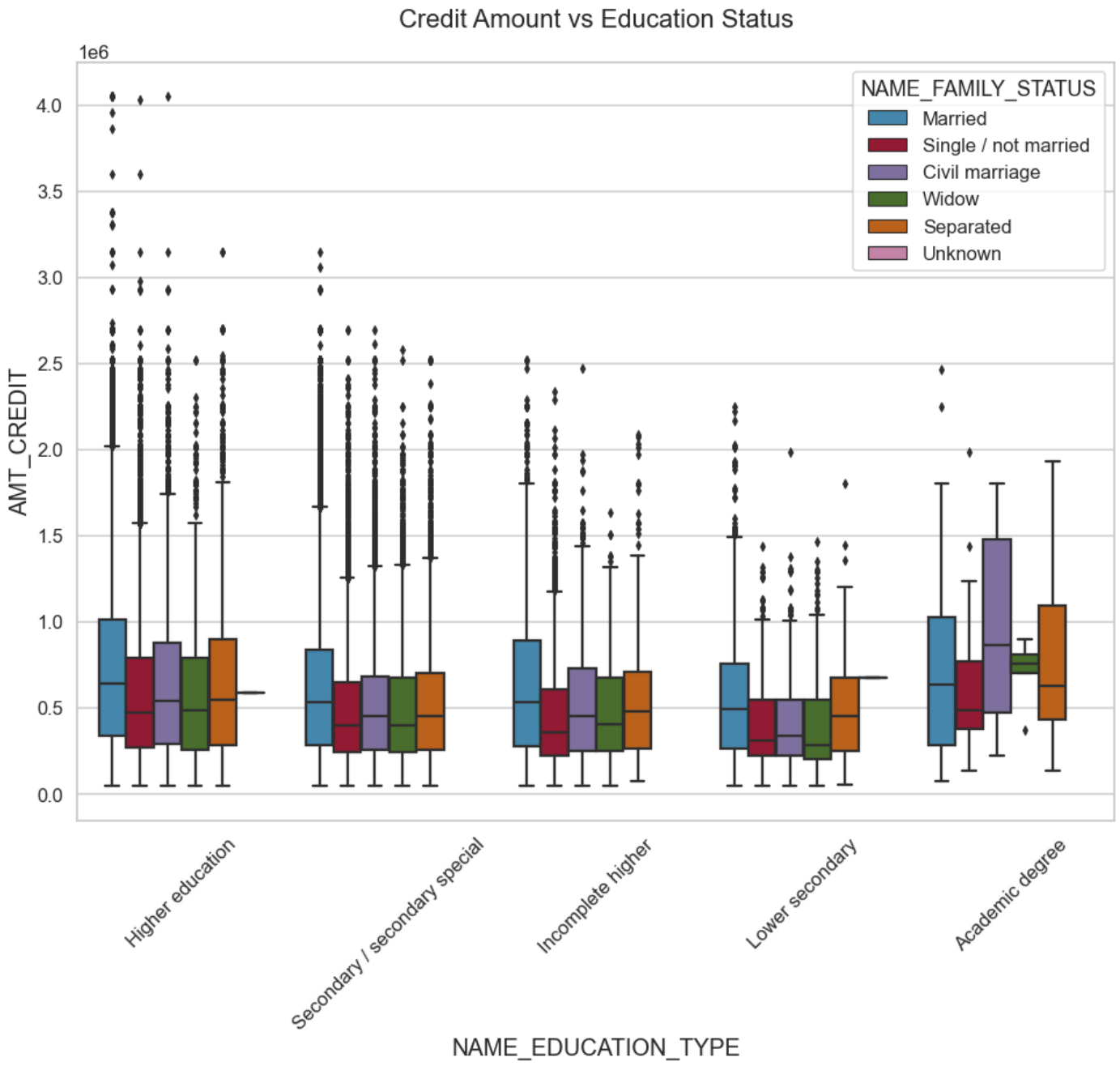
Bivariate analysis for target 1

Credit amount vs Education Status

Observations:

1. Quite similar with Target 0 From the above box plot 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'. Civil marriage for Academic degree is having most of the credits in the third quartile.

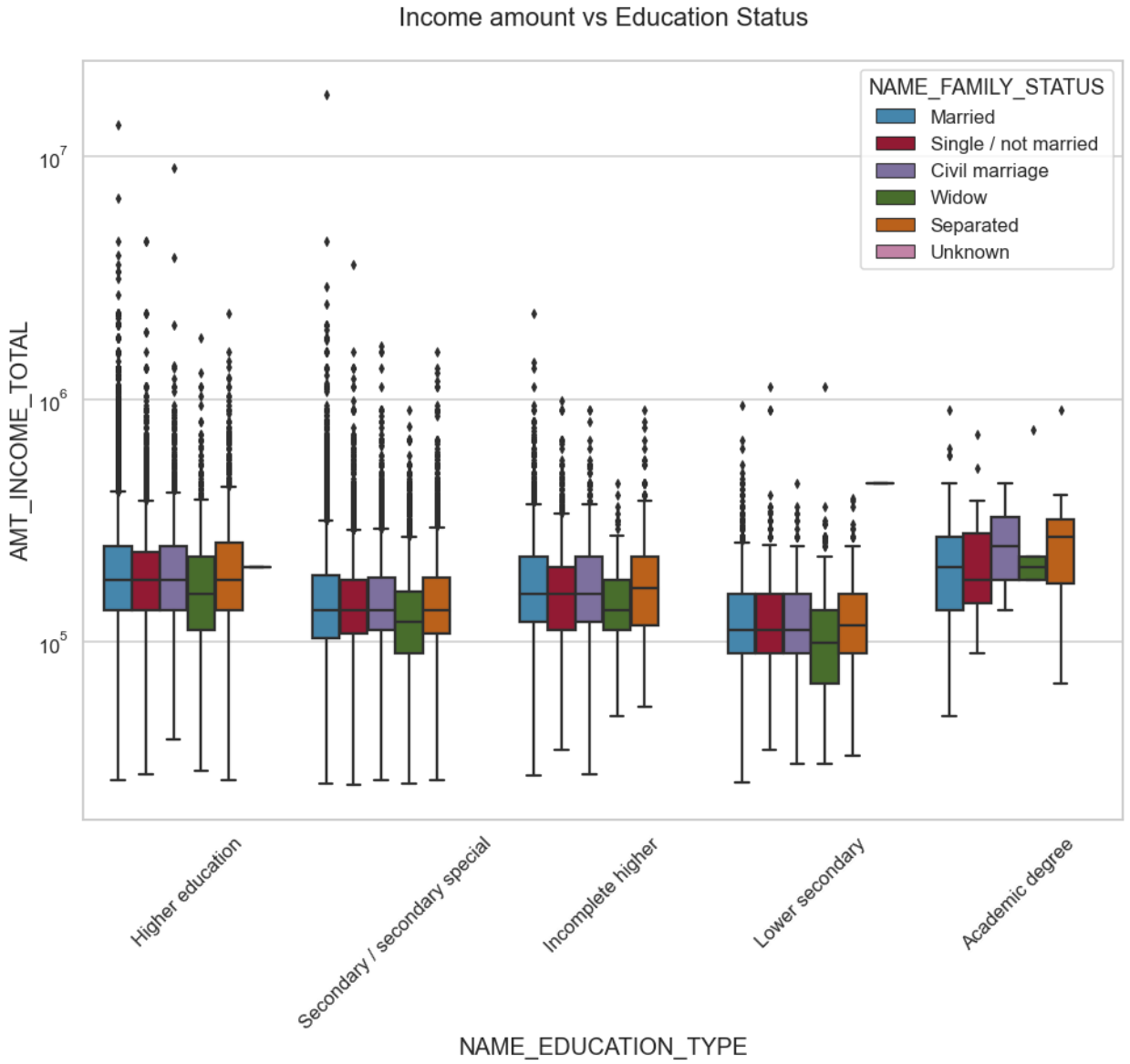


Income amount vs Education Status

Observations:

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

2. Less outlier are present for Academic degree but their income amount is little higher that Higher education. Lower secondary has less income amount than others.



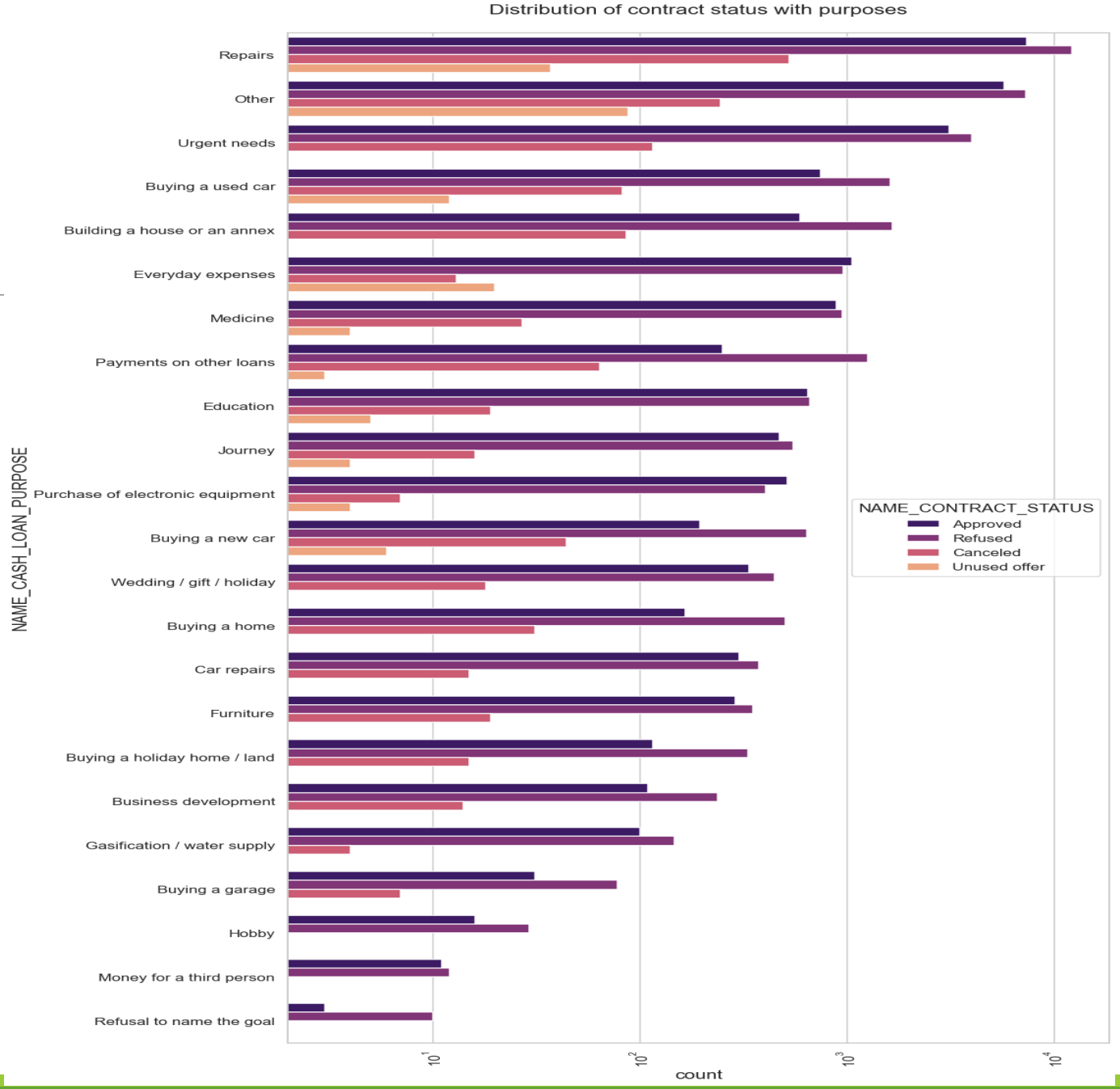
Univariate analysis after merging previous data

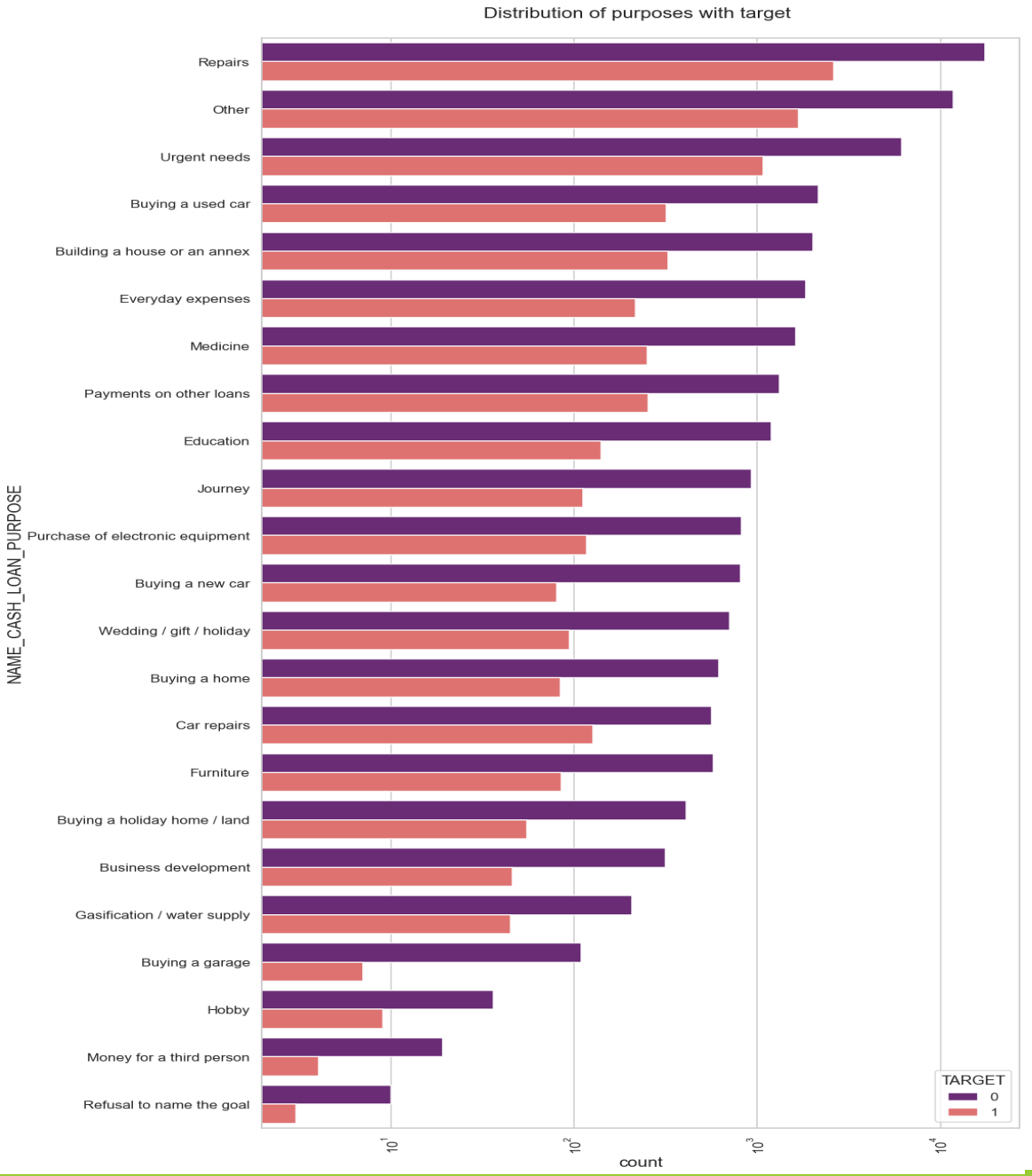
Observations :

1. Most rejection of loans came from purpose 'repairs'.

2. For education purposes we have equal number of approves and rejection

3. Paying other loans and buying a new car is having significant higher rejection than approvals.



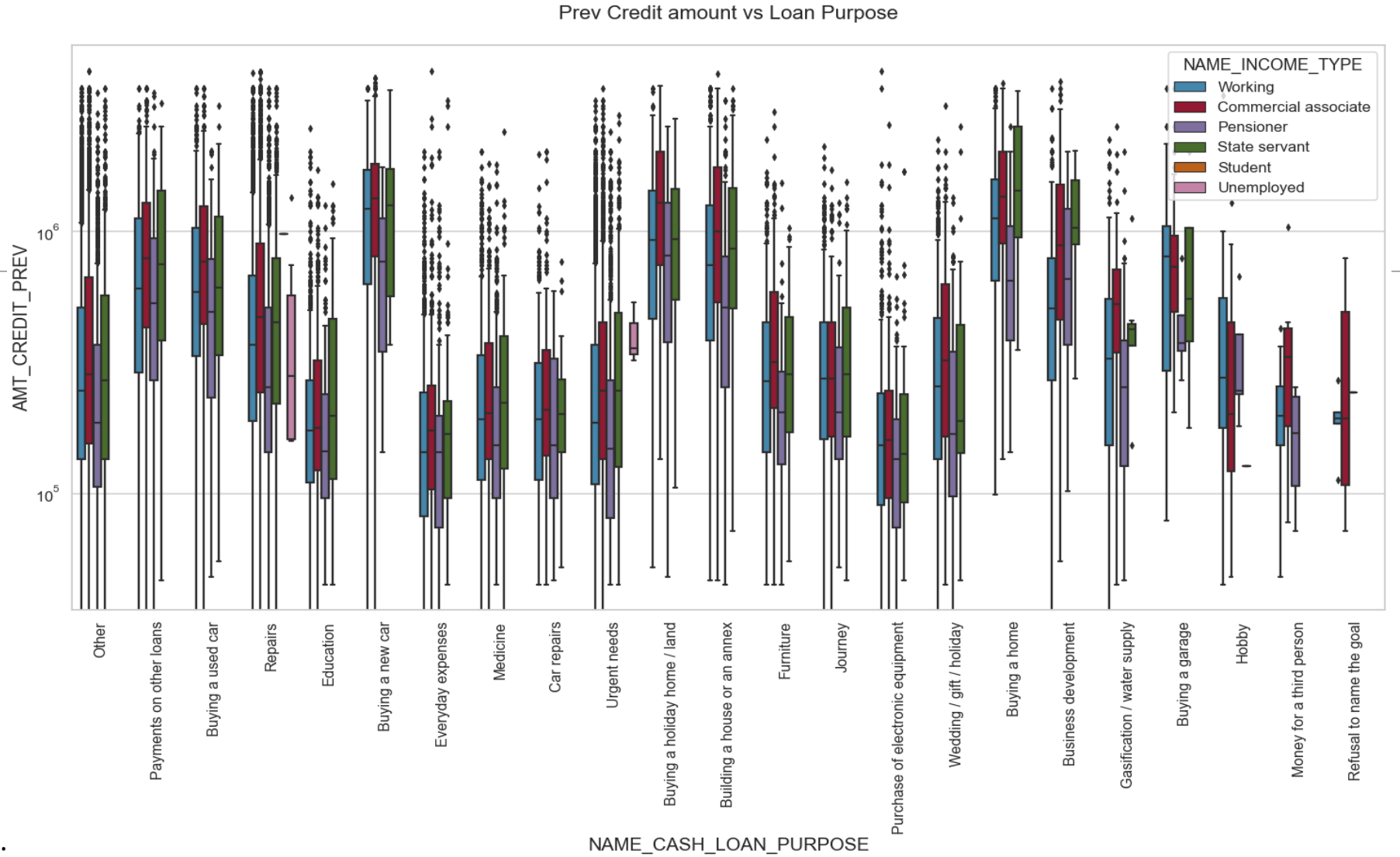


Observations:

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

2. 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.

Performing bivariate analysis



Observations:

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.

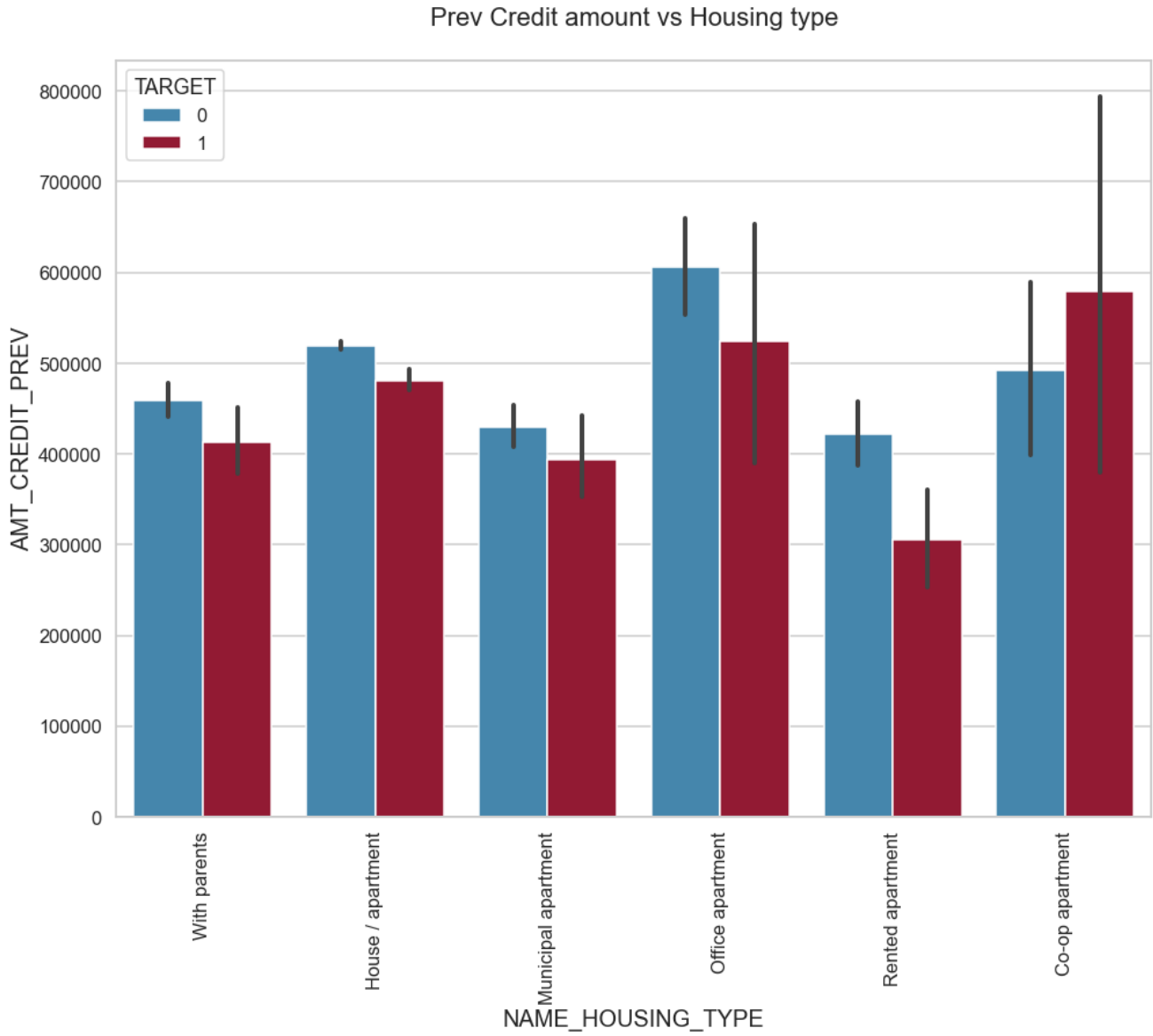
Previous Credit amount vs Loan Purpose

Prev Credit amount vs Housing type

<u>Observations:</u>

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

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



conclusion

1. Banks should focus on densely populated areas as they are less likely to be defaulted.

2. Banks should focus more on contract type ‘Student’ ,’pensioner’ and ‘Businessman’ with housing ‘type other than ‘Co-op apartment’ for successful payments.

3. Banks should focus less on income type ‘Working’ as they are having most number of unsuccessful payments.

4. Also with loan purpose ‘Repair’ is having higher number of unsuccessful payments on time.

5. Get as much as clients from housing type ‘With parents’ as they are having least number of unsuccessful payments.

Thank you