



International
Institute of Information
Technology Bangalore

upGrad

EDA on Bank Loan Applications

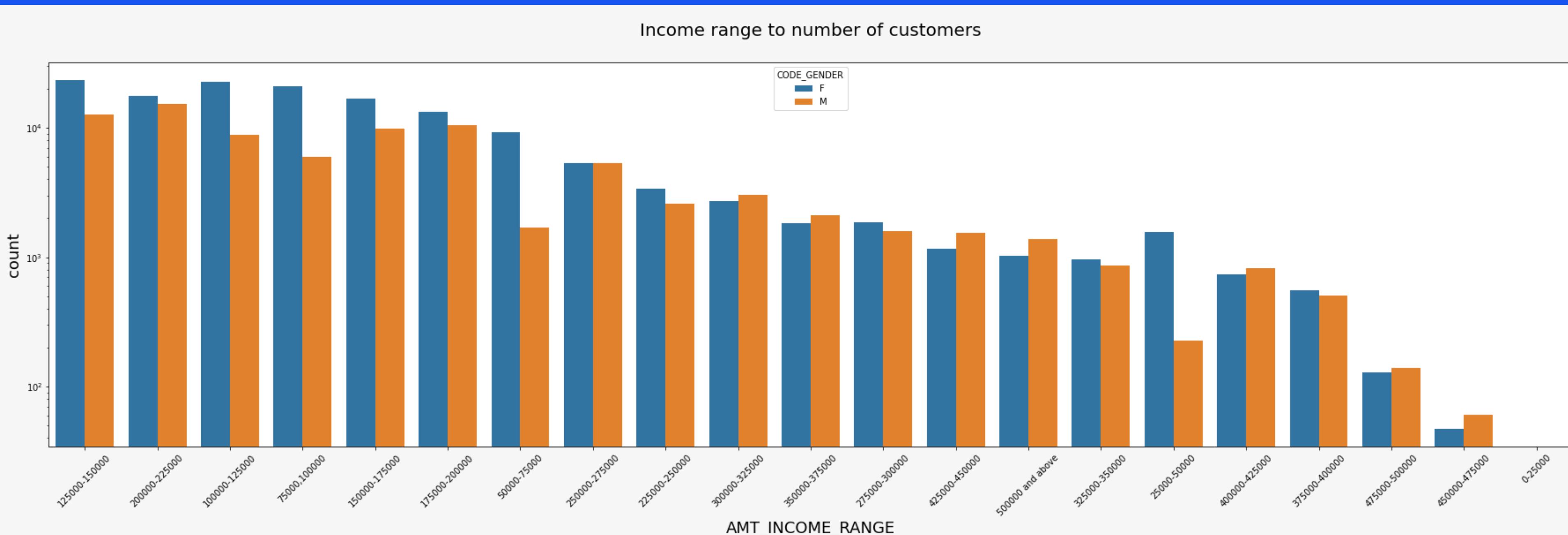
By Mayur Jangale



Categorical Univariate Analysis for Target 0

Distribution of Income Range

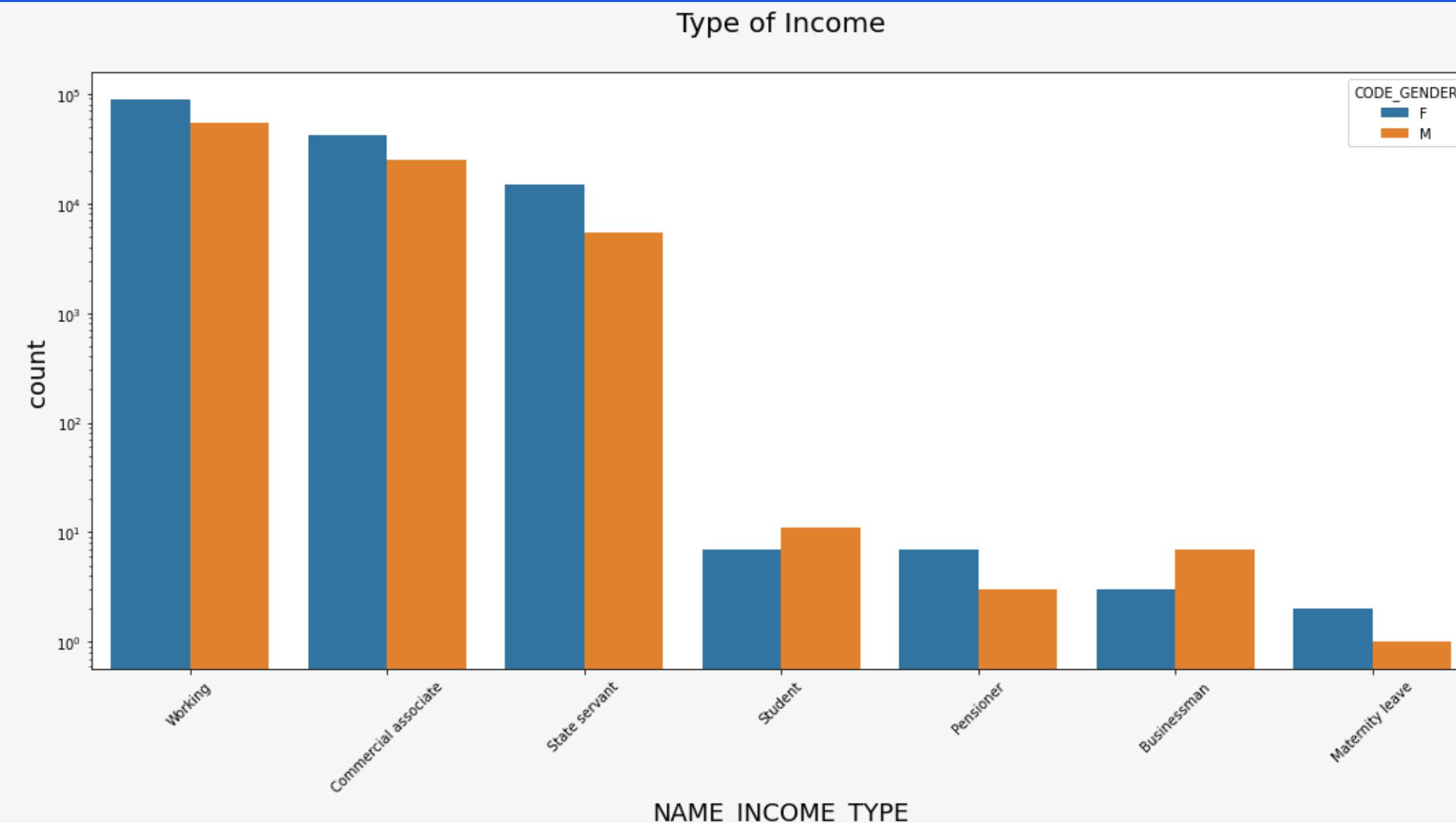
1. Count of female is more than that of males
2. Income range decreases when income count is decreased. However, this looks like coincident.
3. 100000 to 200000 income range holds majority of counts.
4. Females are having more credits than men.
5. Very less count for income range 500000 to 0.



Distribution of Income Type

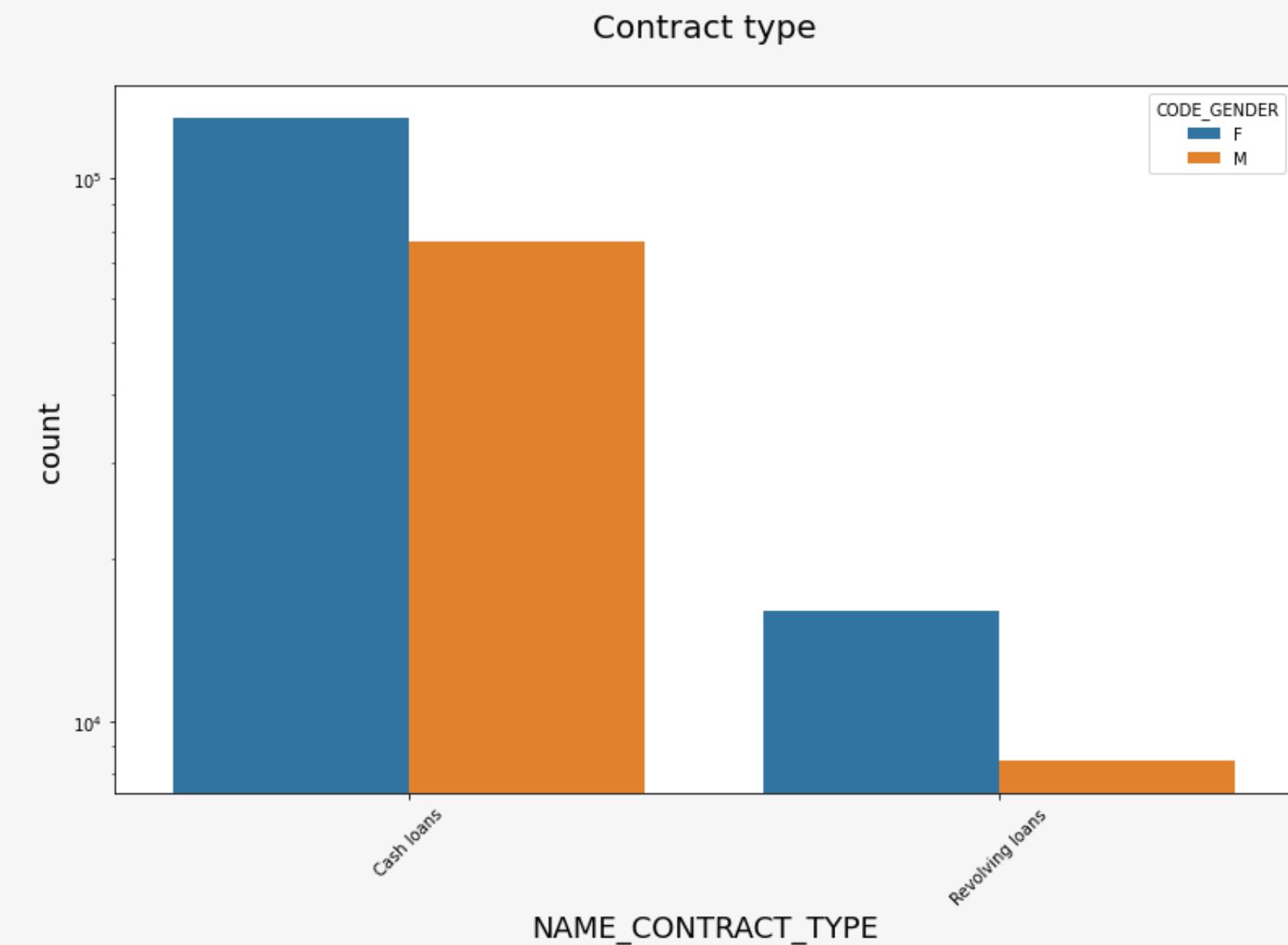
1. 'working', 'commercial associate', and 'State Servant' have more credits than others
2. For above 3 and 'maternity leave', pensioner females are having more credits for rest males are having more credits.
3. For 'Students', 'Pensioner', 'Businessman', 'Maternity Leave' credit count is low

Data Defects: Men count should be 0 for maternity leave/there should be separate value as paternity leave.

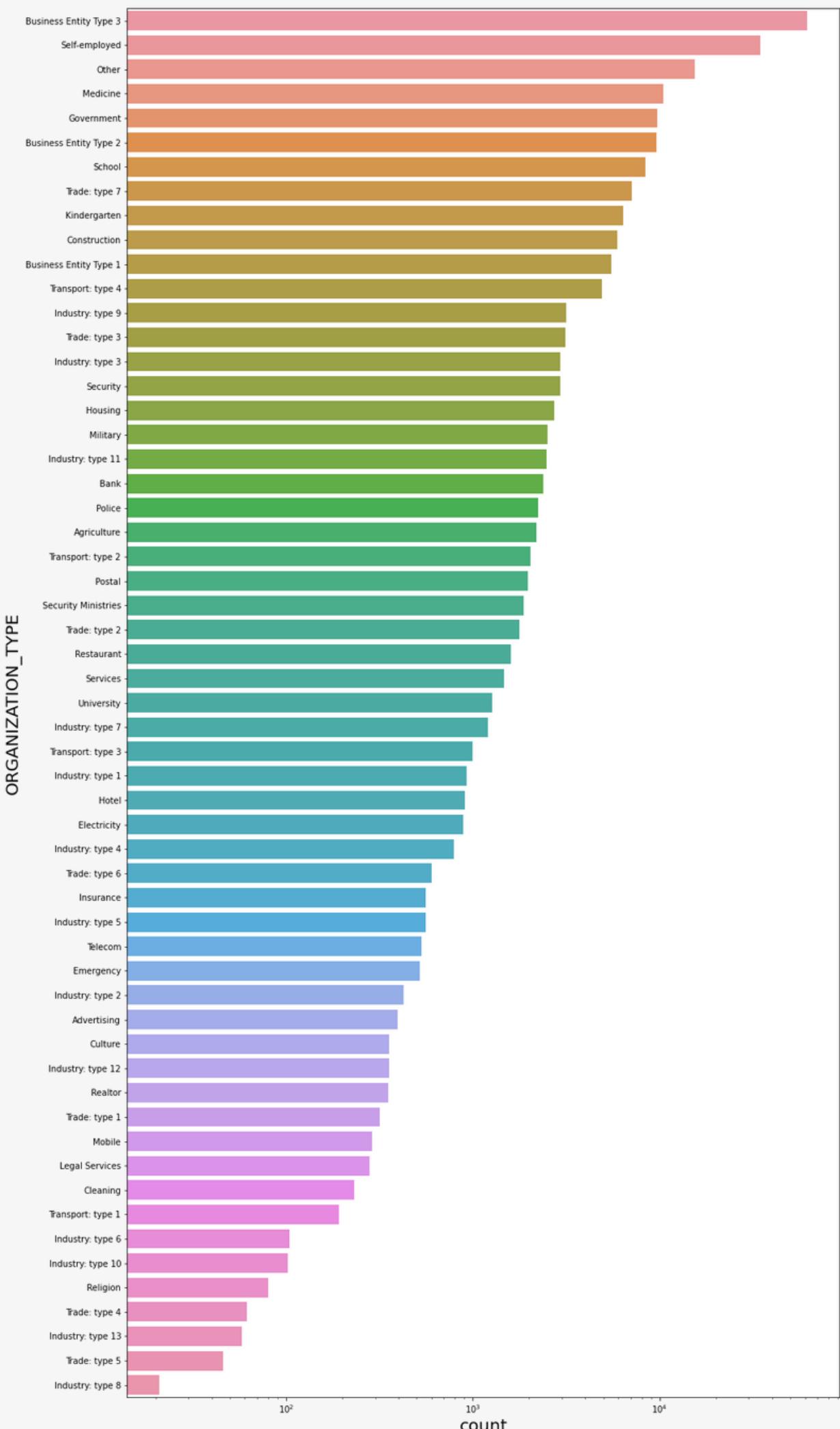


Distribution of Contract Type

1. Cash loan contracts are having more counts than revolving loans
2. Females have majority in both catagories.



Organization type for df0



Distribution of Organization Type

1. Organisations from which most of the users have applied for loan:

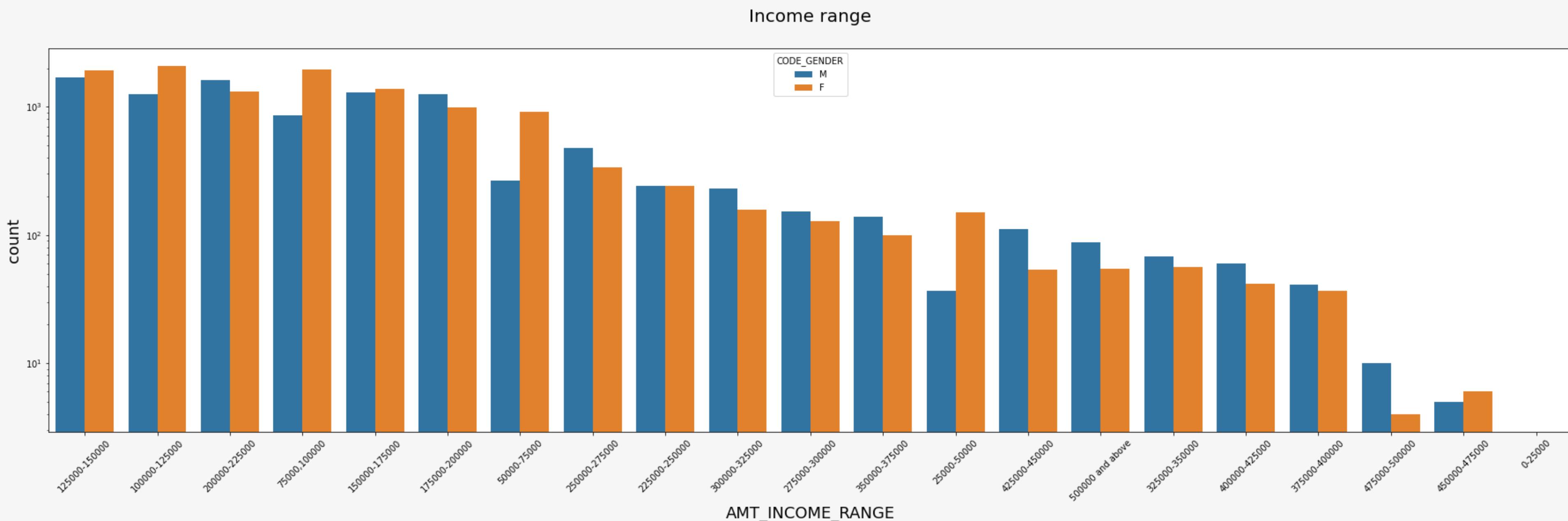
'Business entity Type 3' , 'Self employed', 'Other' ,
'Medicine', 'Government', Business entity type 2.

2. Organisations from which least users have applied for loan: Industry type 8, type 6, type 10, religion and trade type 5, type 4.

Categorical Univariate Analysis for Target 1

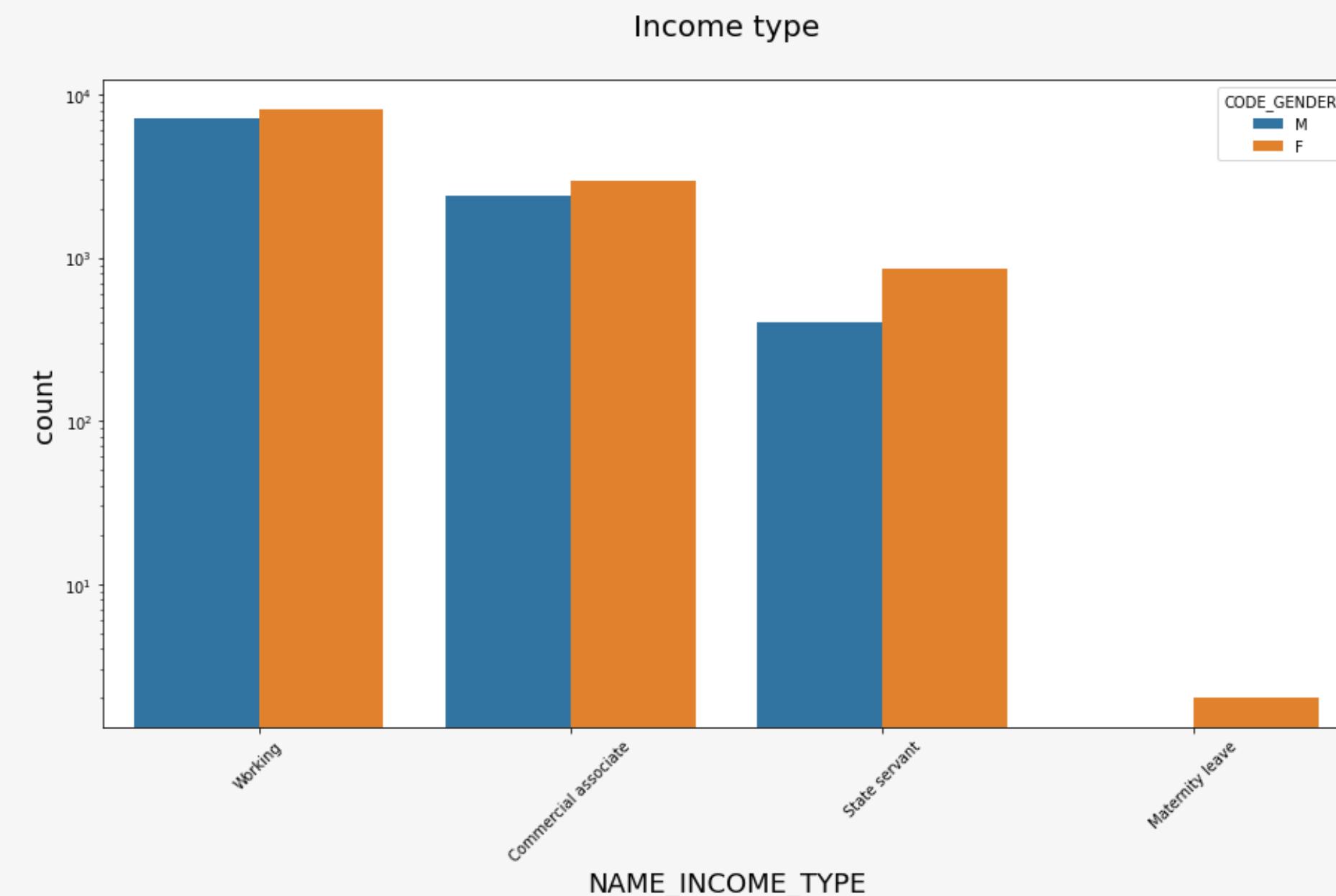
Distribution of Income Range

1. Male counts are lesser than female.
2. very less count for income range 400000 to 0
3. 100000 to 200000 income range is having more number of credits.



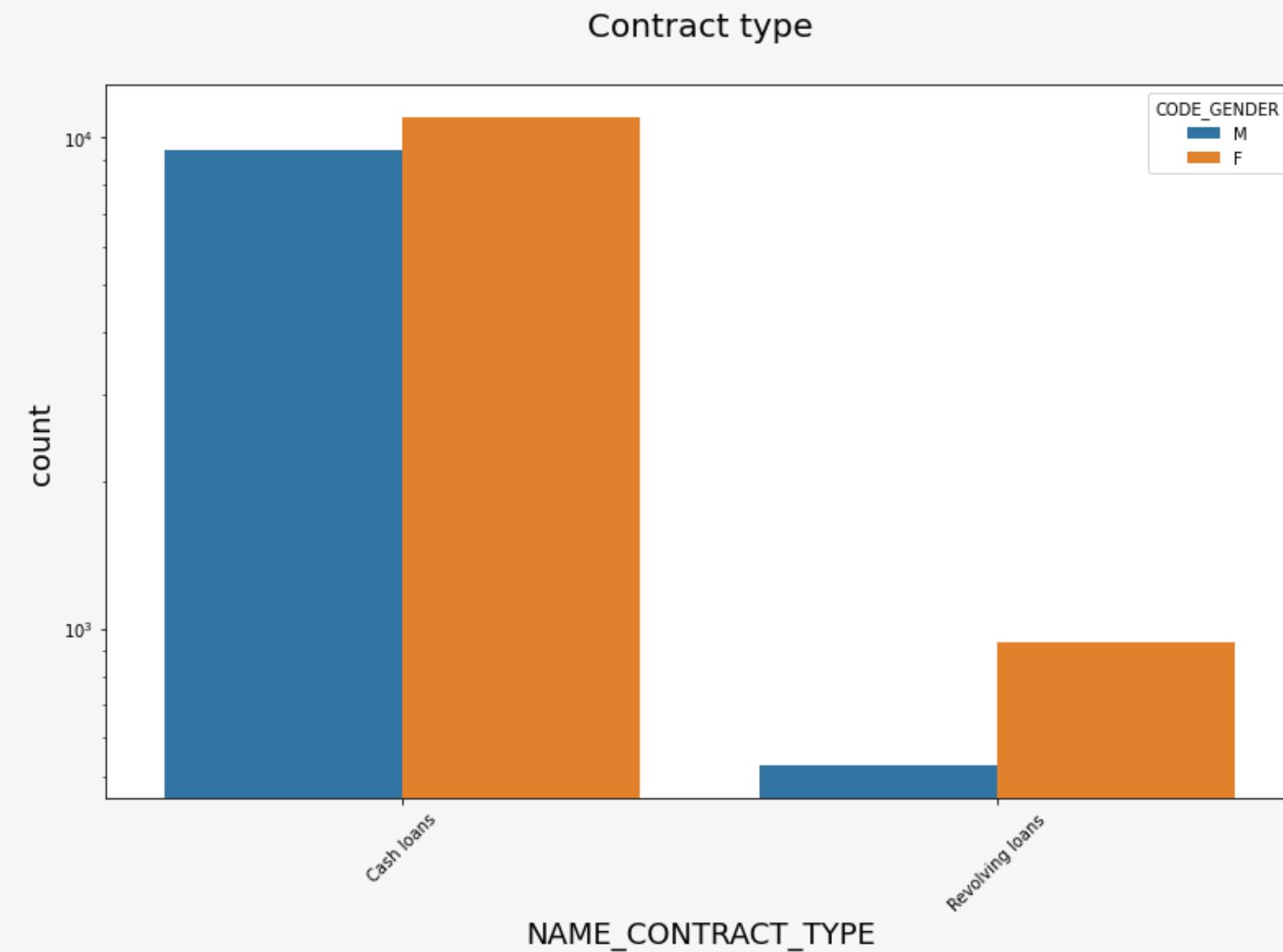
Distribution of Income Type

1. For 'working', 'commercial associate', and 'State Servant' Count is higher than that of Maternity leave.
2. Females are having slightly more credits than Males.
3. 'Maternity leave' carries least counts.
4. 'student' , 'pensioner' and 'Businessman' are missing from type 1.

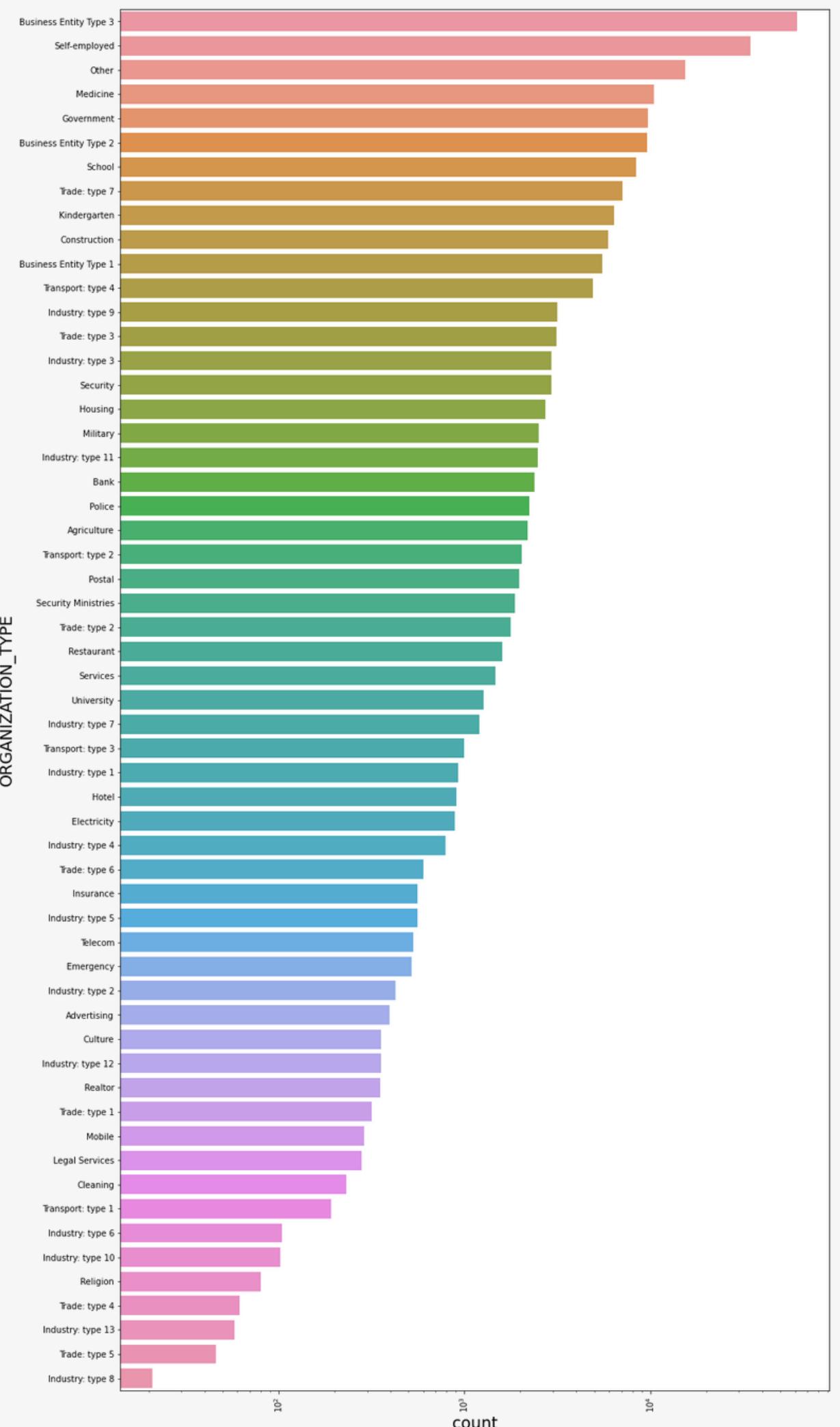


Distribution of Contract Type

1. 'cash loans' contract type is having more credits than 'Revolving loans'.
2. Females are in majority again.
3. For target 1 : Females counts are to high as compared to mens.



Type 1 - Organization type



Distribution of Organization Type

1. 'Business entity Type 3' , 'Self employed' , 'Other' , 'Medicine' and 'Government' has highest counts.
2. Industry type 8,type 6, type 10, religion and trade type 5, type 4 has lowest count.
3. Results are same as per target 0.

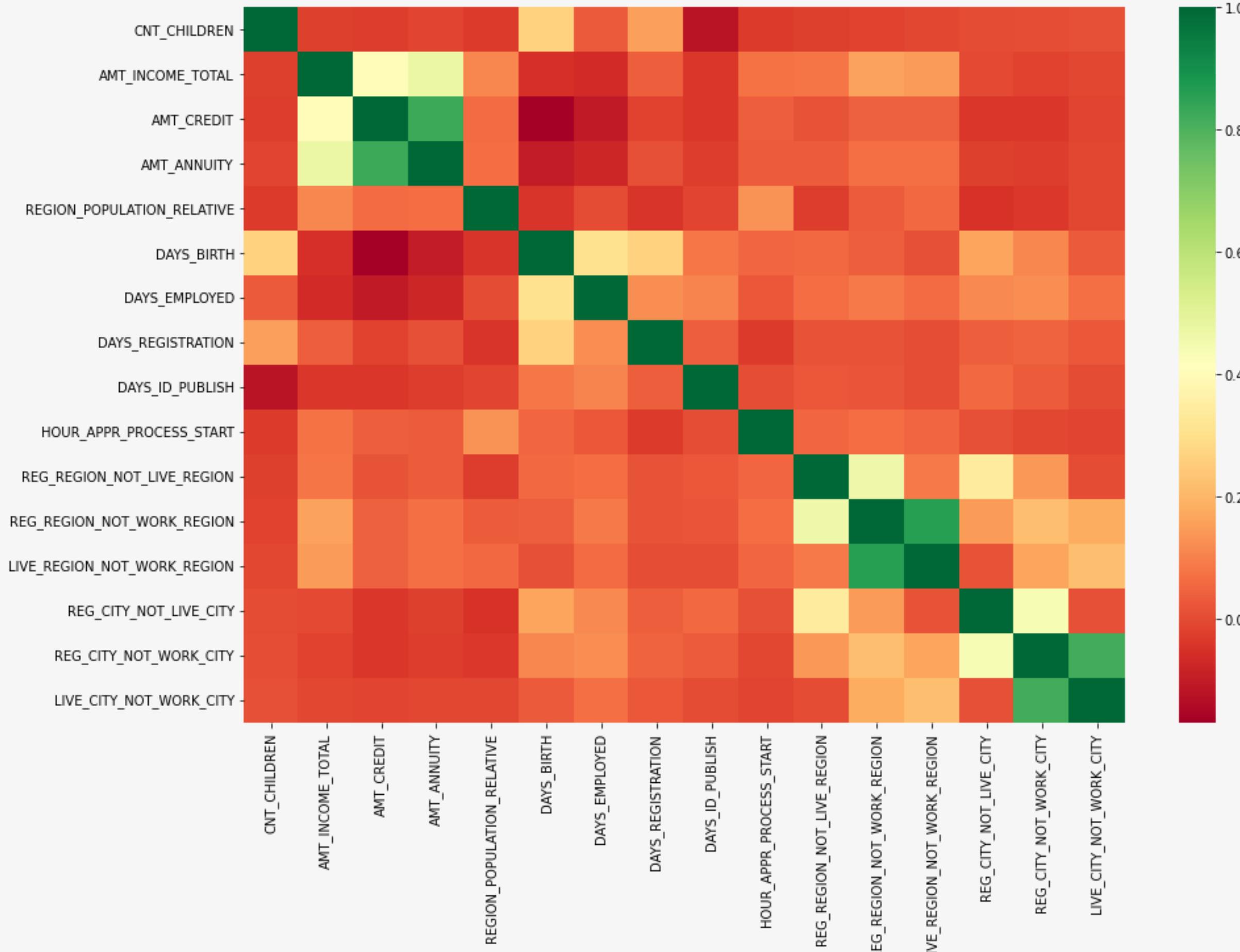
Correlation of Target

0

Correlation in target 0

1. Number of children client have is inversely proportional to the credit amount, means Credit amount is lesser for high children count client have and vice-versa.
2. Credit amount is inversely proportional to the date of birth, which means Credit amount is lower for high age and vice-versa.
3. Income amount is inversely proportional to the number of children client have, means less income for more 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.

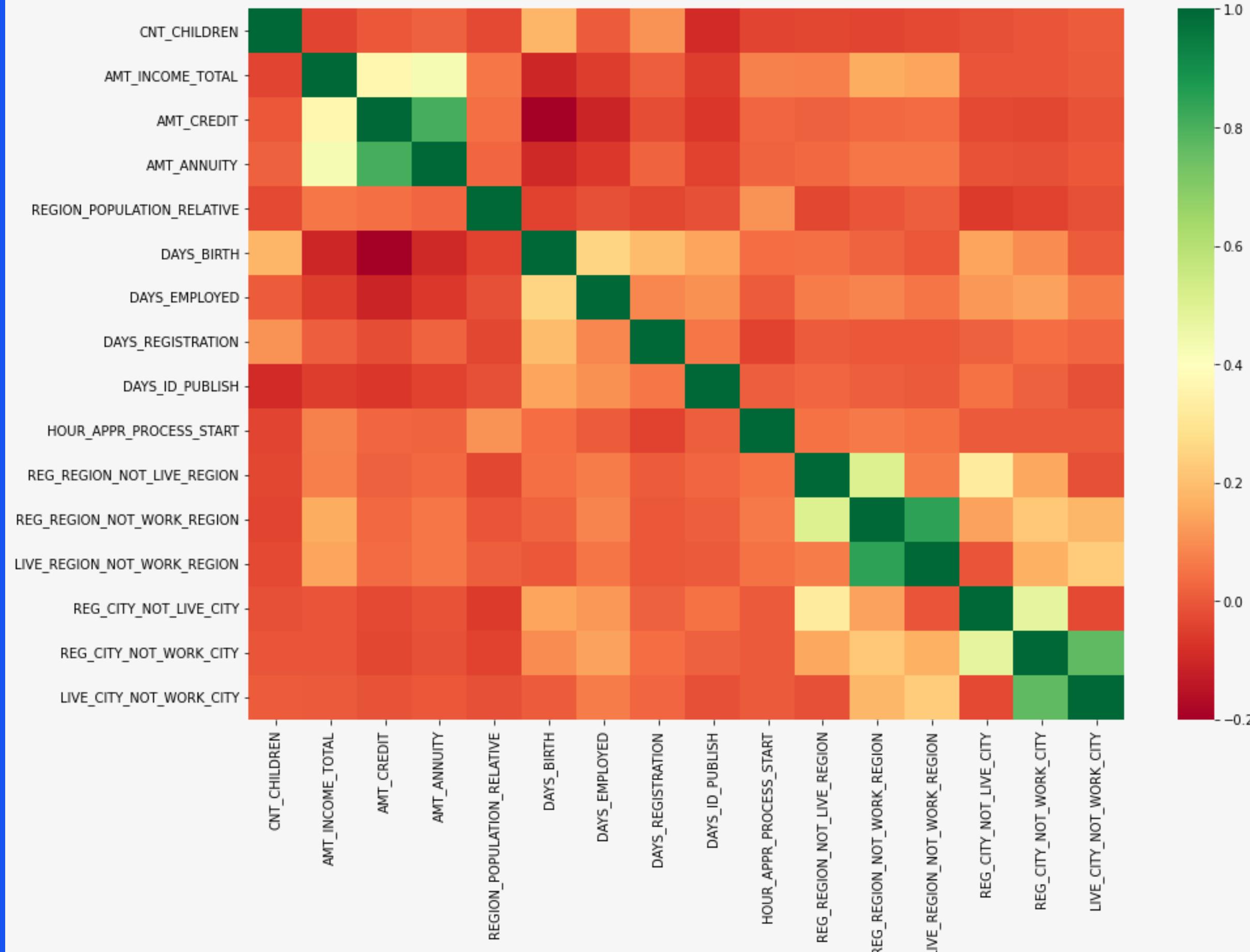
Correlation in target 0



Correlation in target 1

Correlation in target 1

1. Target 1 has very simillar results as target 0.



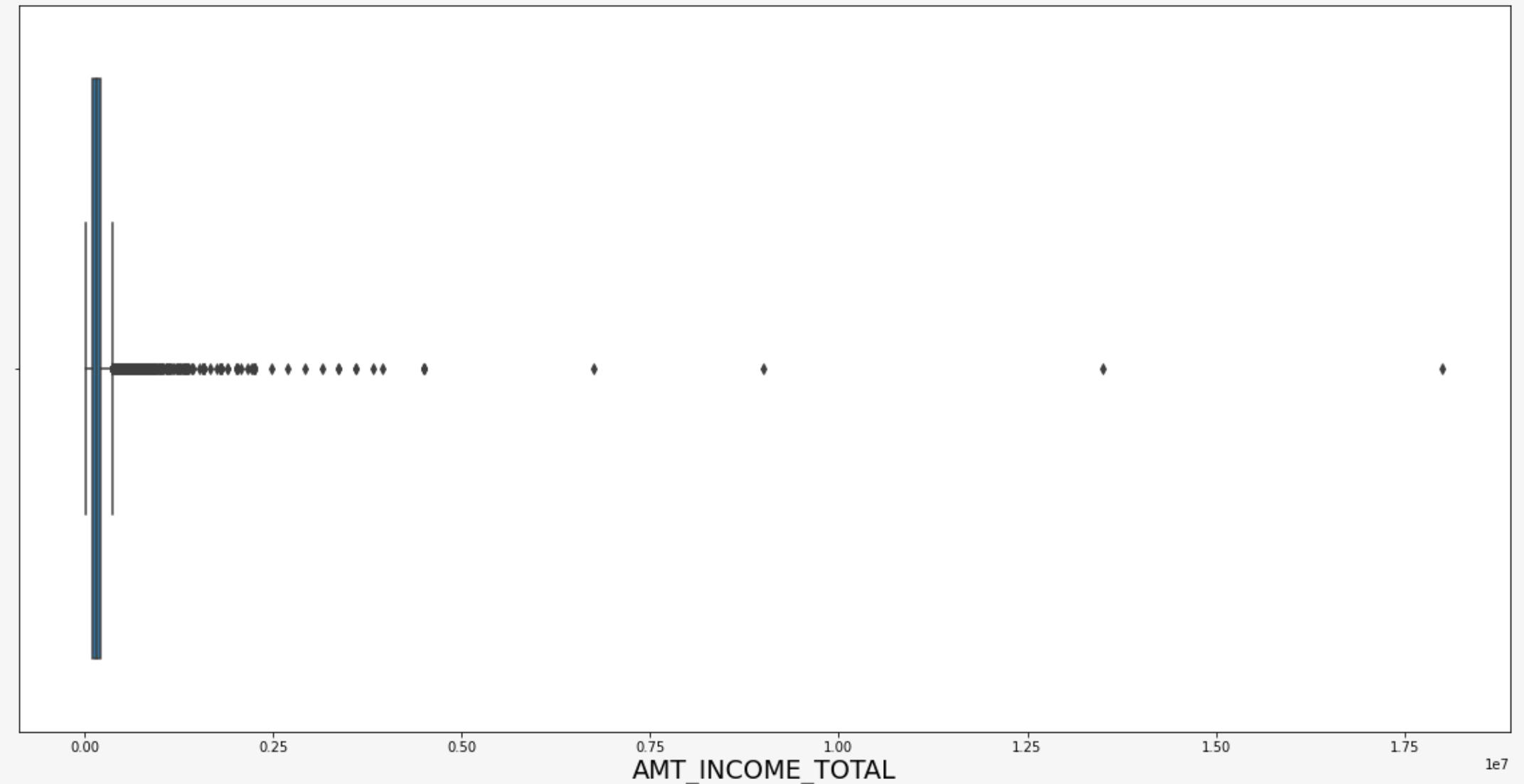
Categorical Univariate Analysis For Variables - Target 0

Box Plot For Income Amount

Points to note:

1. Outliers are present.
2. IQR is too small.

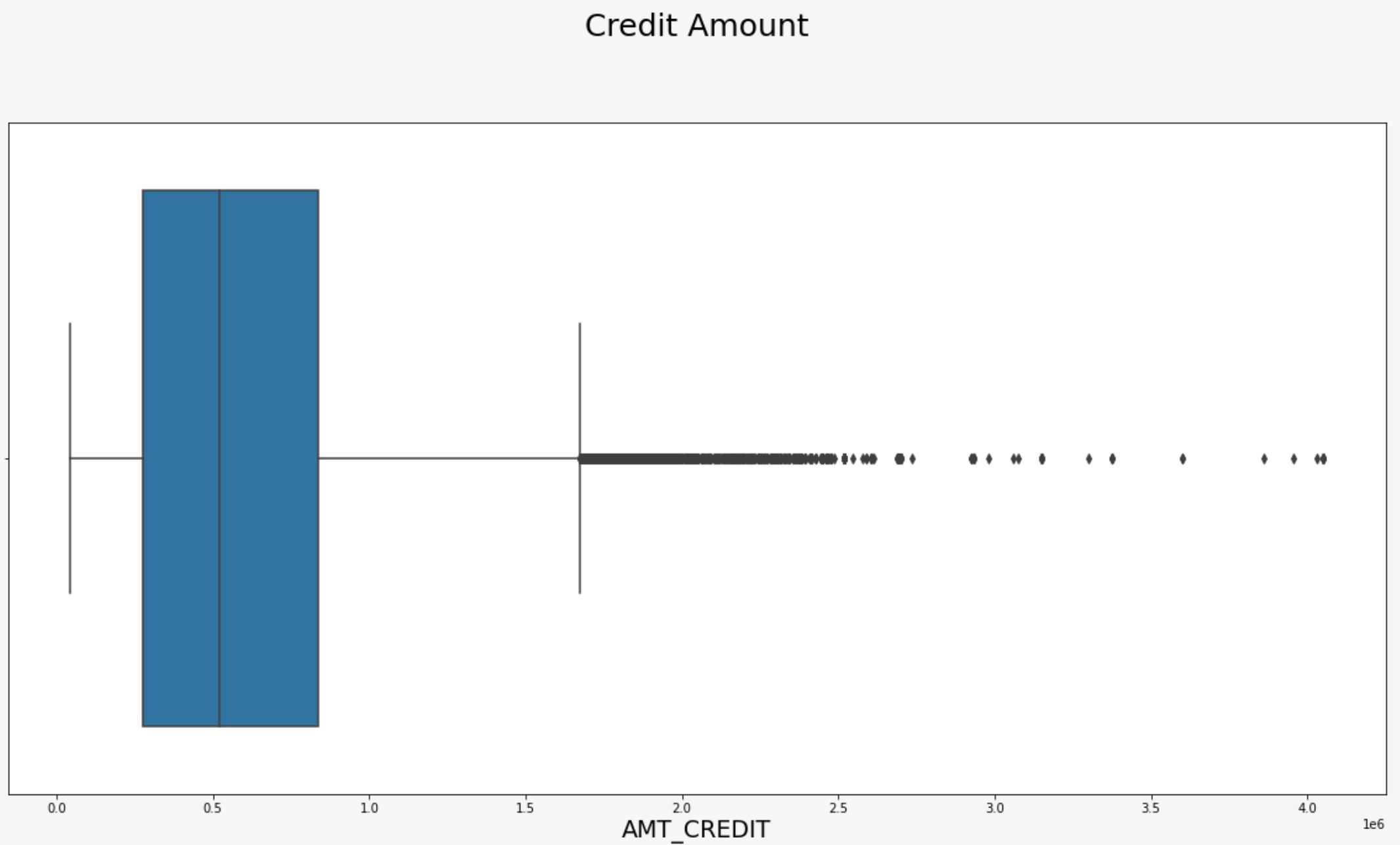
Income amount



Box Plot For Credit Amount

Points to note:

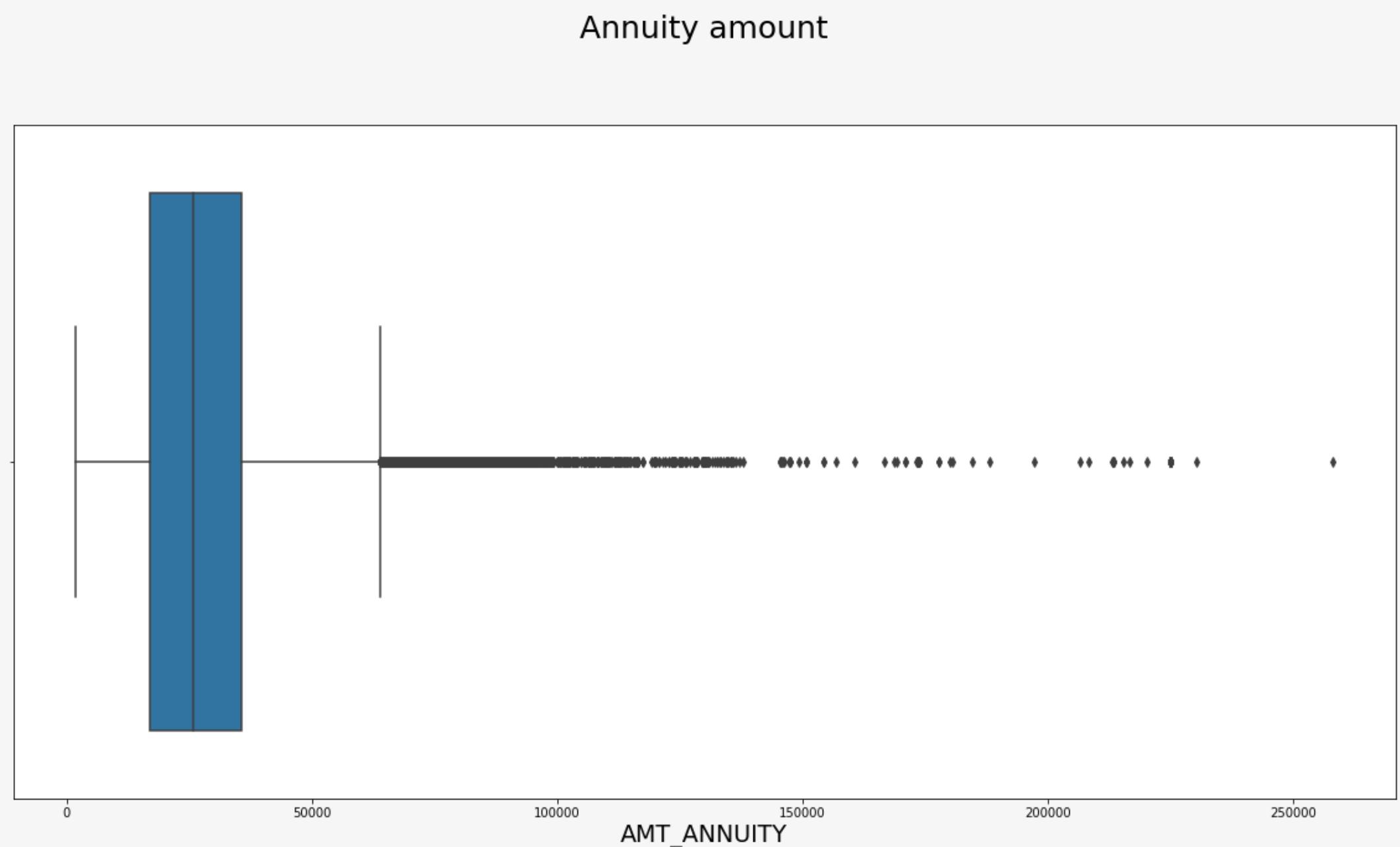
1. Outliers are present.
2. More credits of clients are present in the median to third quartile side.



Box Plot For Annuity Amount

Points to note:

1. Outliers are present.
2. Both quartiles are equal for annuity amount which means, annuity clients are equal in both quartiles.

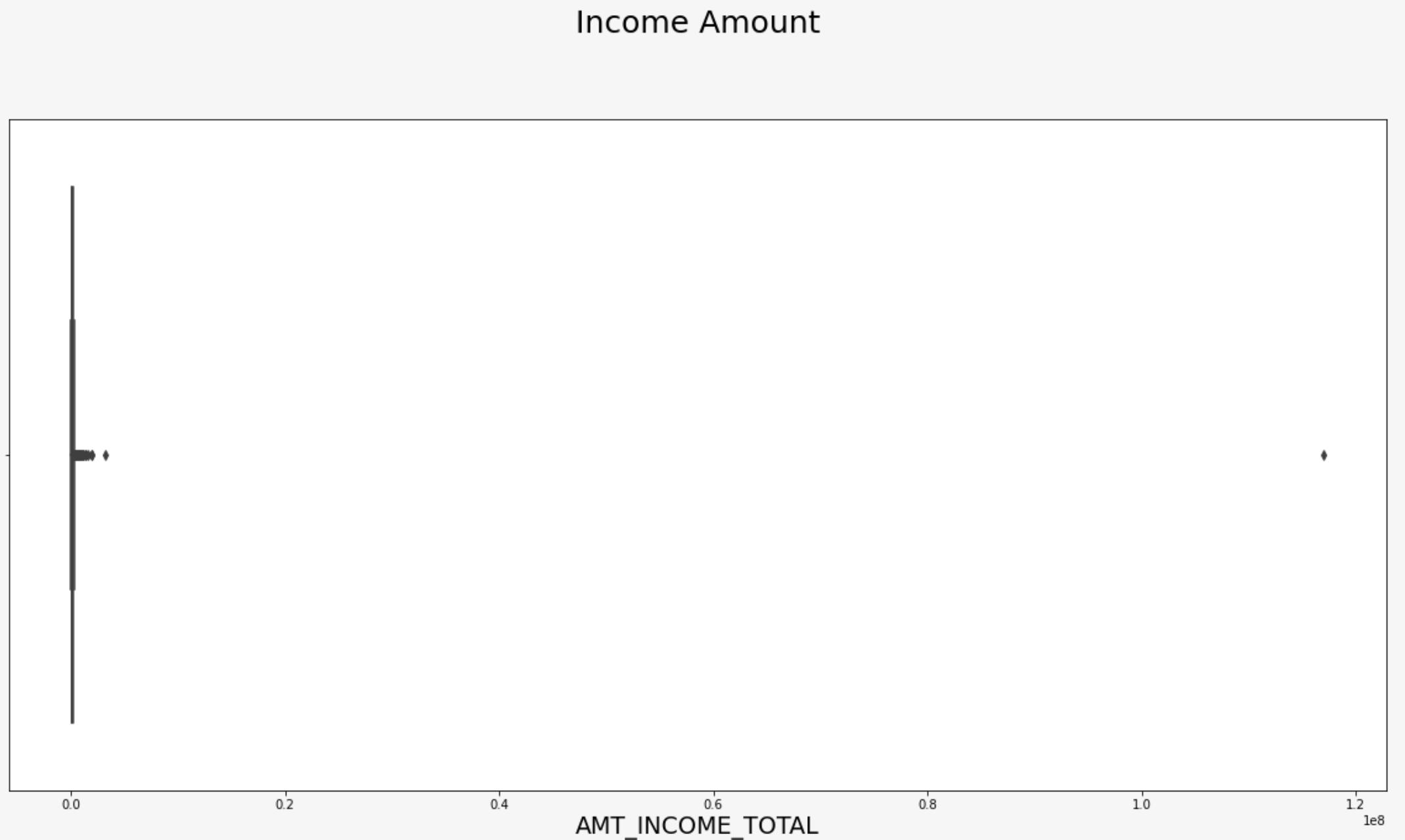


Categorical Univariate Analysis For Variables - Target 1

Box Plot For Income Amount

Points to note:

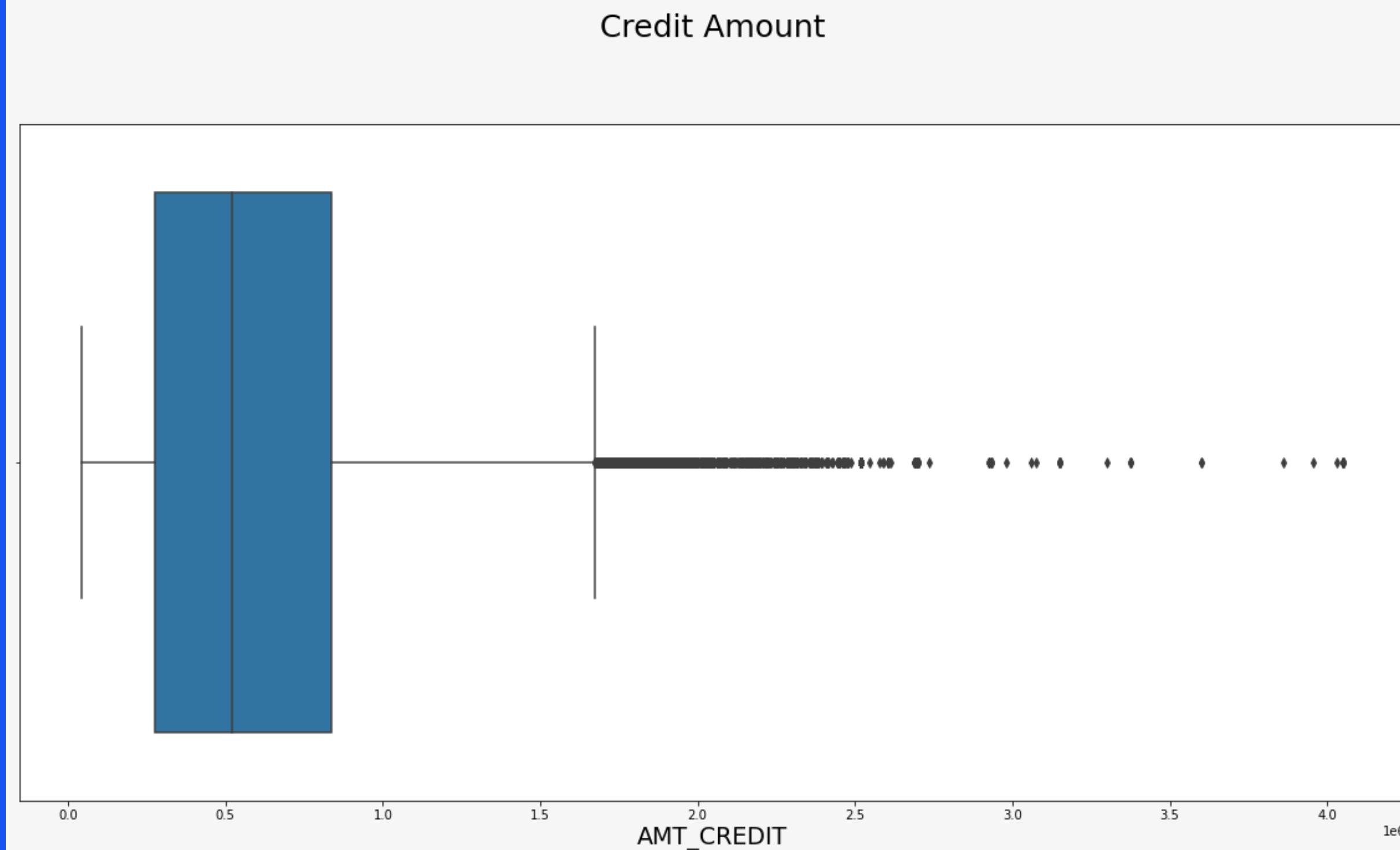
1. Outliers are present.
2. IQR is too small.



Box Plot For Credit Amount

Points to note:

1. Outliers are present
2. The first quartile is smaller than third quartile hence third quartile has most of the credits.

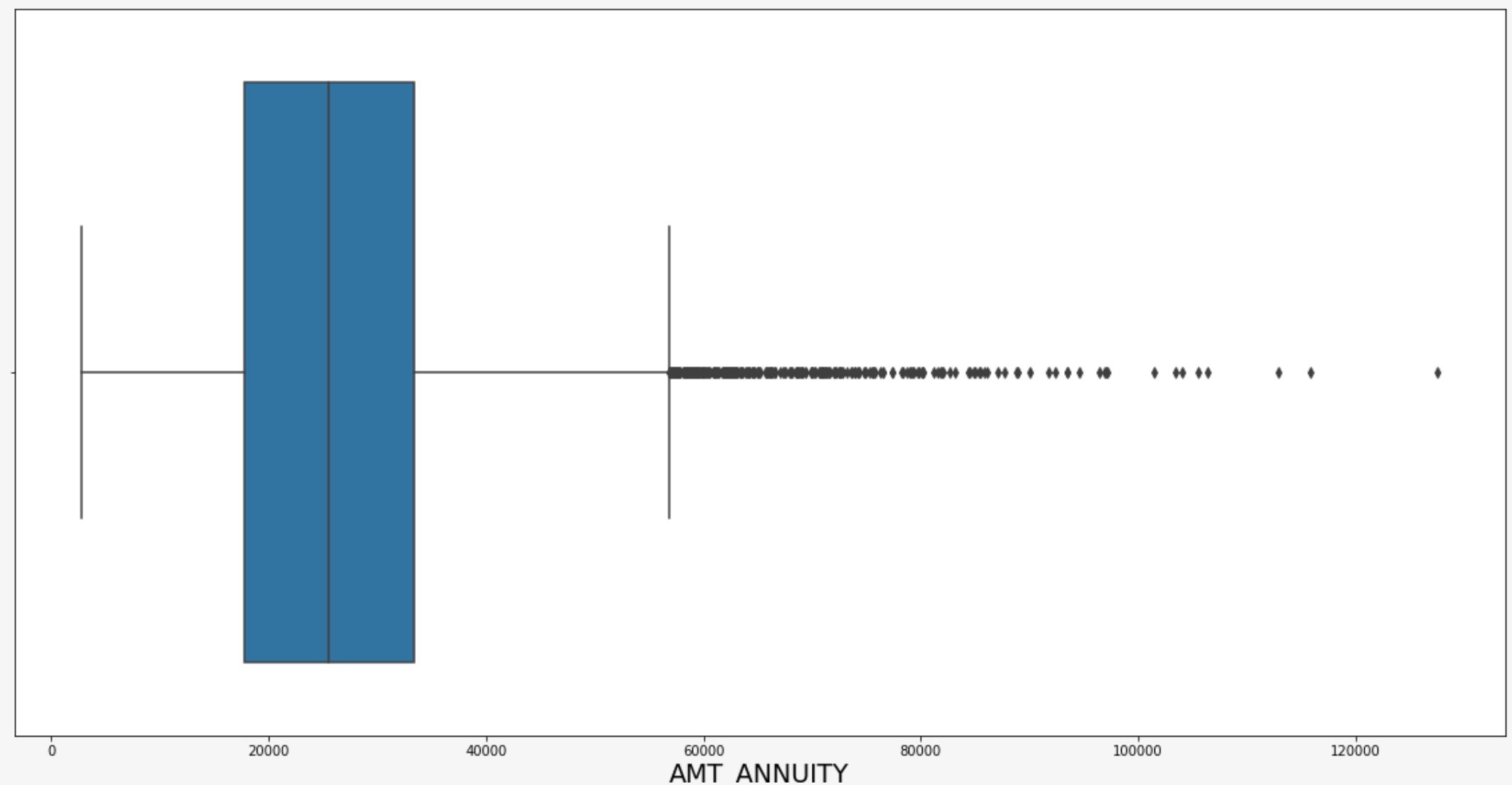


Box Plot For Annuity Amount

Points to note:

1. Outliers are present.
2. Both quartiles are equal.

Annuity amount



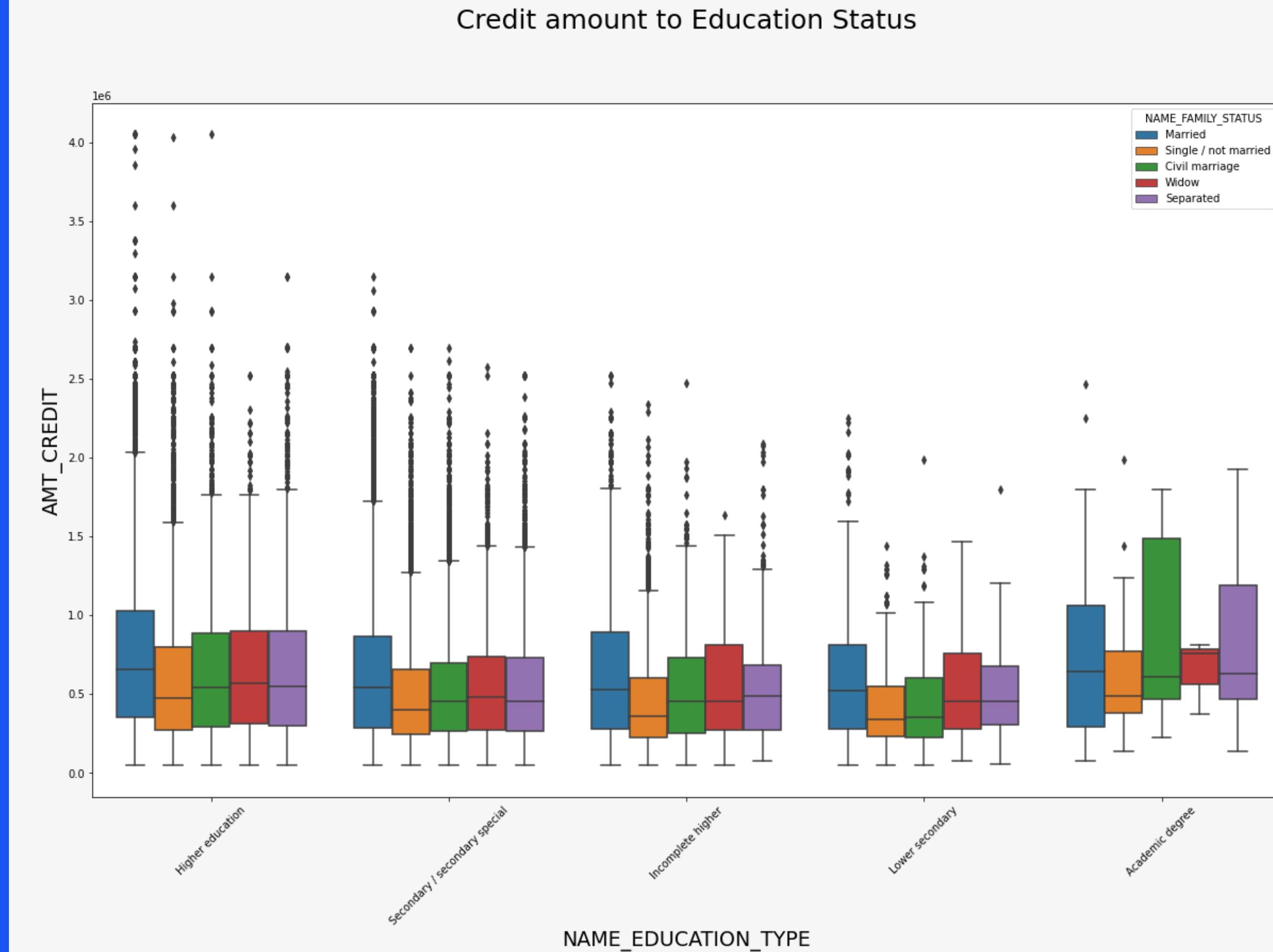
Bivariate Analysis For Type 0



Credit amount to Education Status

Points to note:

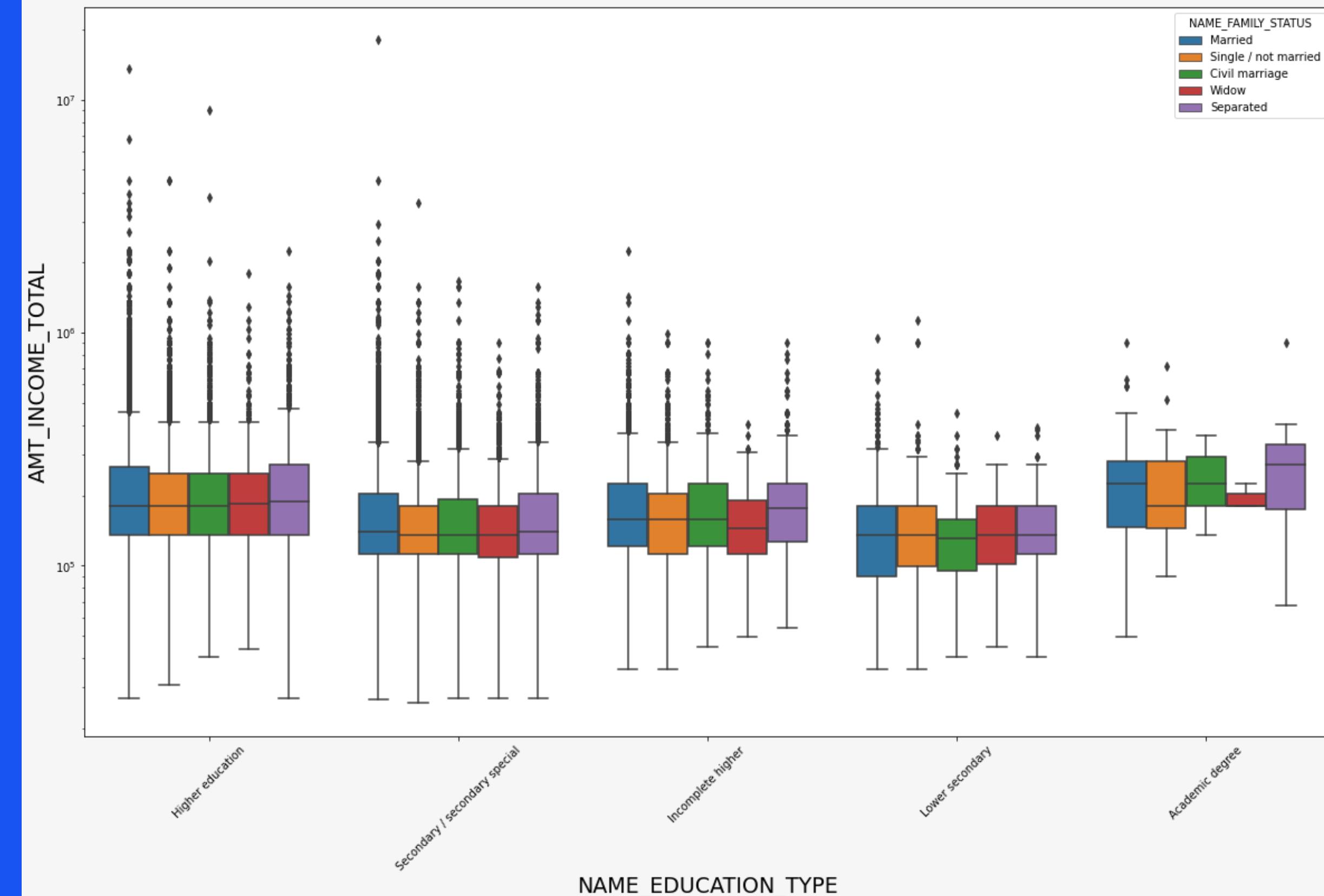
1. 'civil marriage', 'marriage' and 'separated' are having higher credits
2. 'marriage', 'single' and 'civil marriage' have more outliers
3. Civil marriage for Academic degree are having more credits in third quartile.



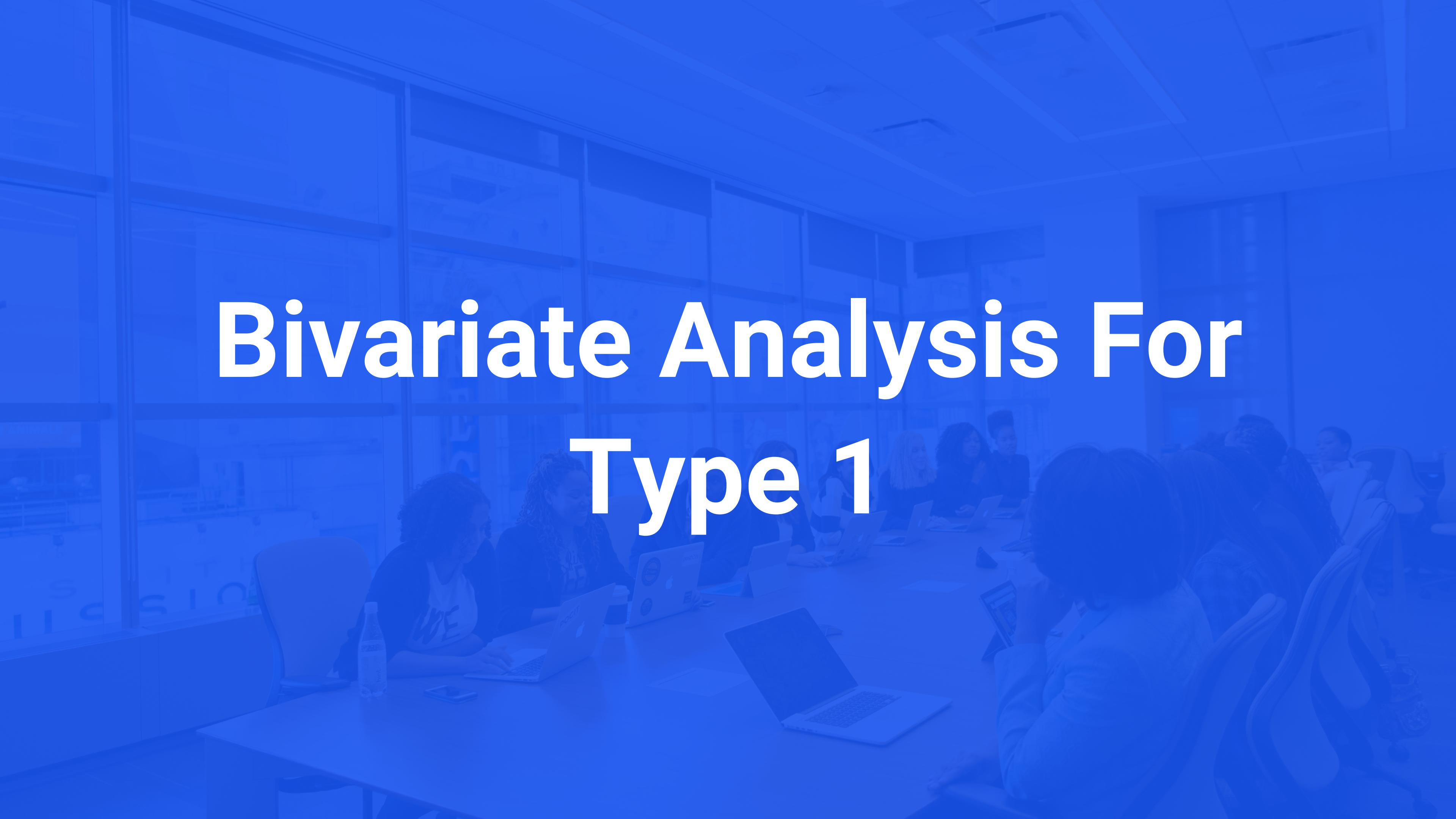
Income amount to Education Status

1. For 'Higher education' income amount is mostly equal to family status. It contains many outliers.
2. Lower secondary of civil marriage family status are having less income amount compared to others.

Income amount to Education Status



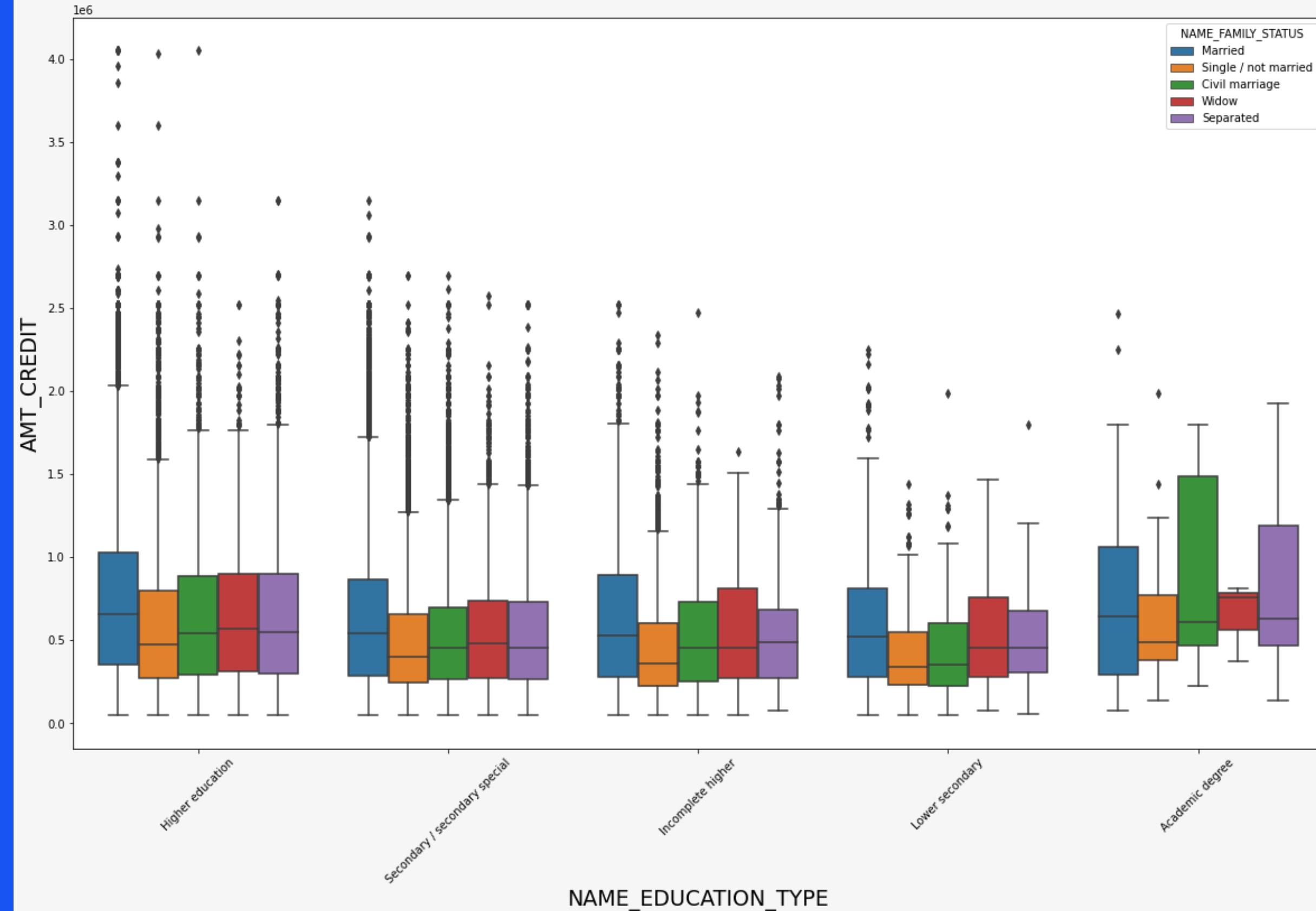
Bivariate Analysis For Type 1



Credit Amount to Education Status

1. Target 0 and target 1 both are having same results.
2. 'civil marriage', 'marriage' and 'separated' have higher credits than others
3. 'Higher education' and 'Secondary' have most of the outliers.
4. Civil marriage for Academic degree has most of credits in third quartile.

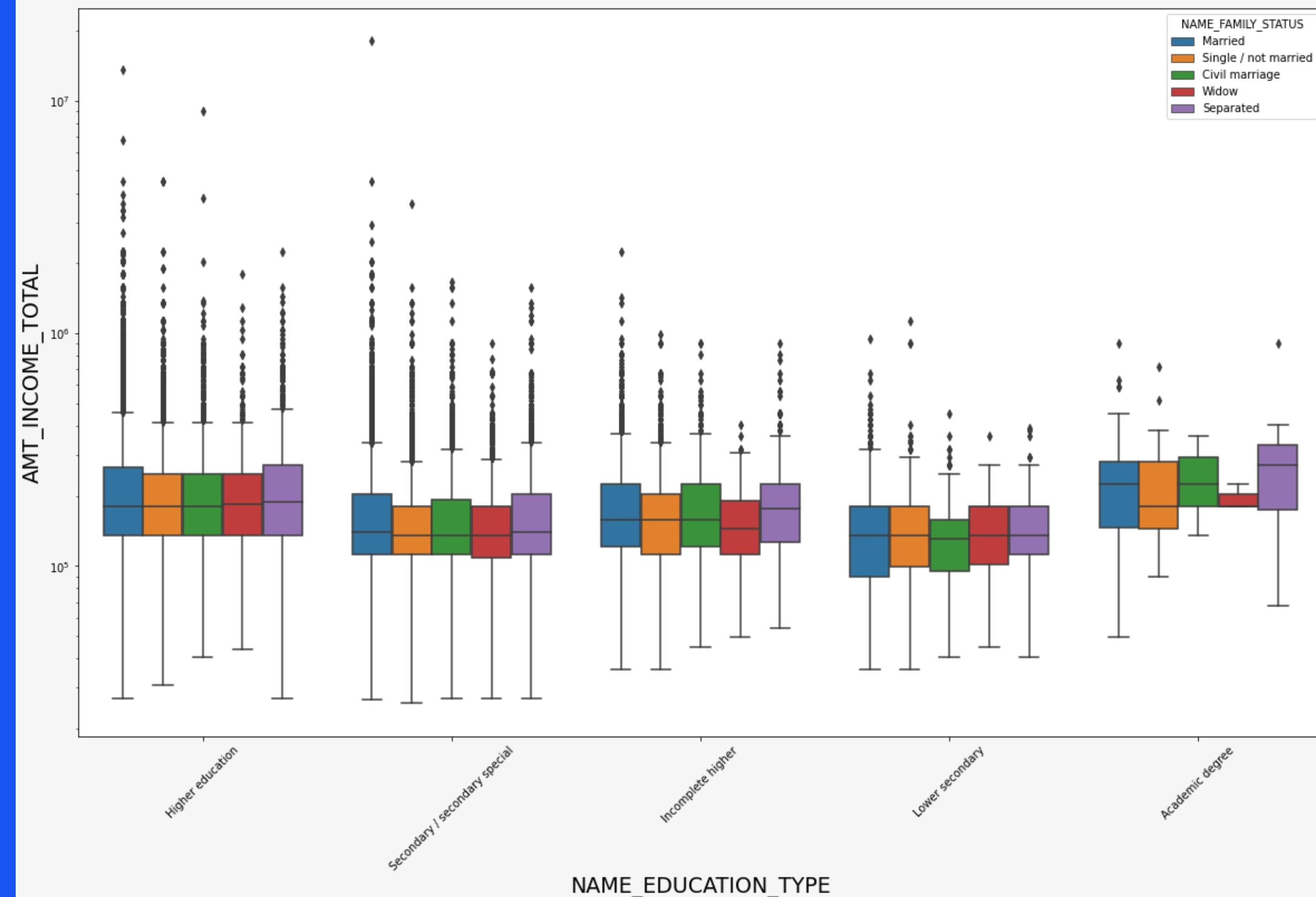
Credit Amount to Education Status



Income amount to Education Status

1. Target0 and target 1 are mostly same.
2. For 'Higher education' the income amount is mostly equal with family status.
2. Less outlier for Academic degree but income amount is little higher than Higher education.
3. Lower secondary are having less income amount than others.

Income amount to Education Status

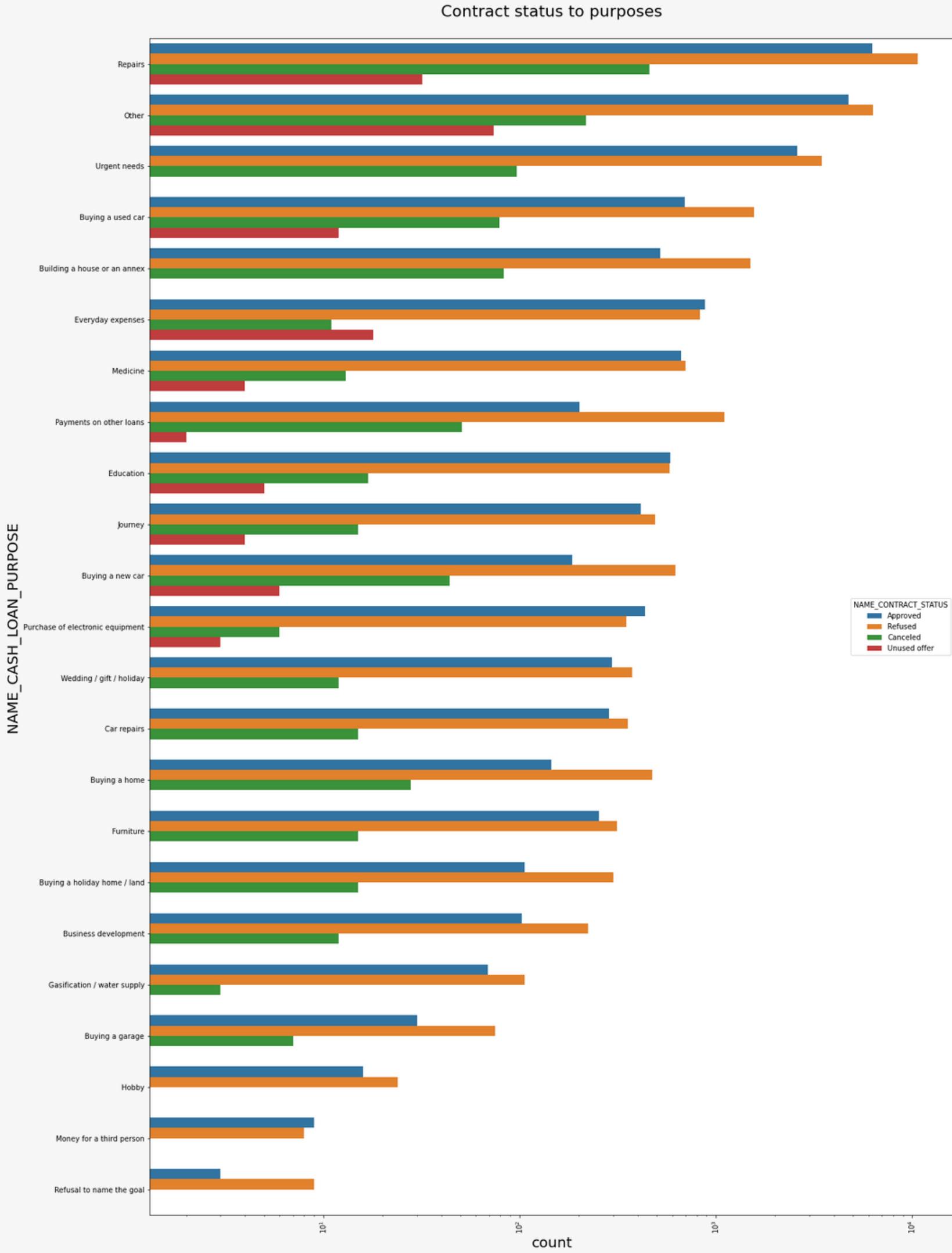




Univariate Analysis After Merging Previous data

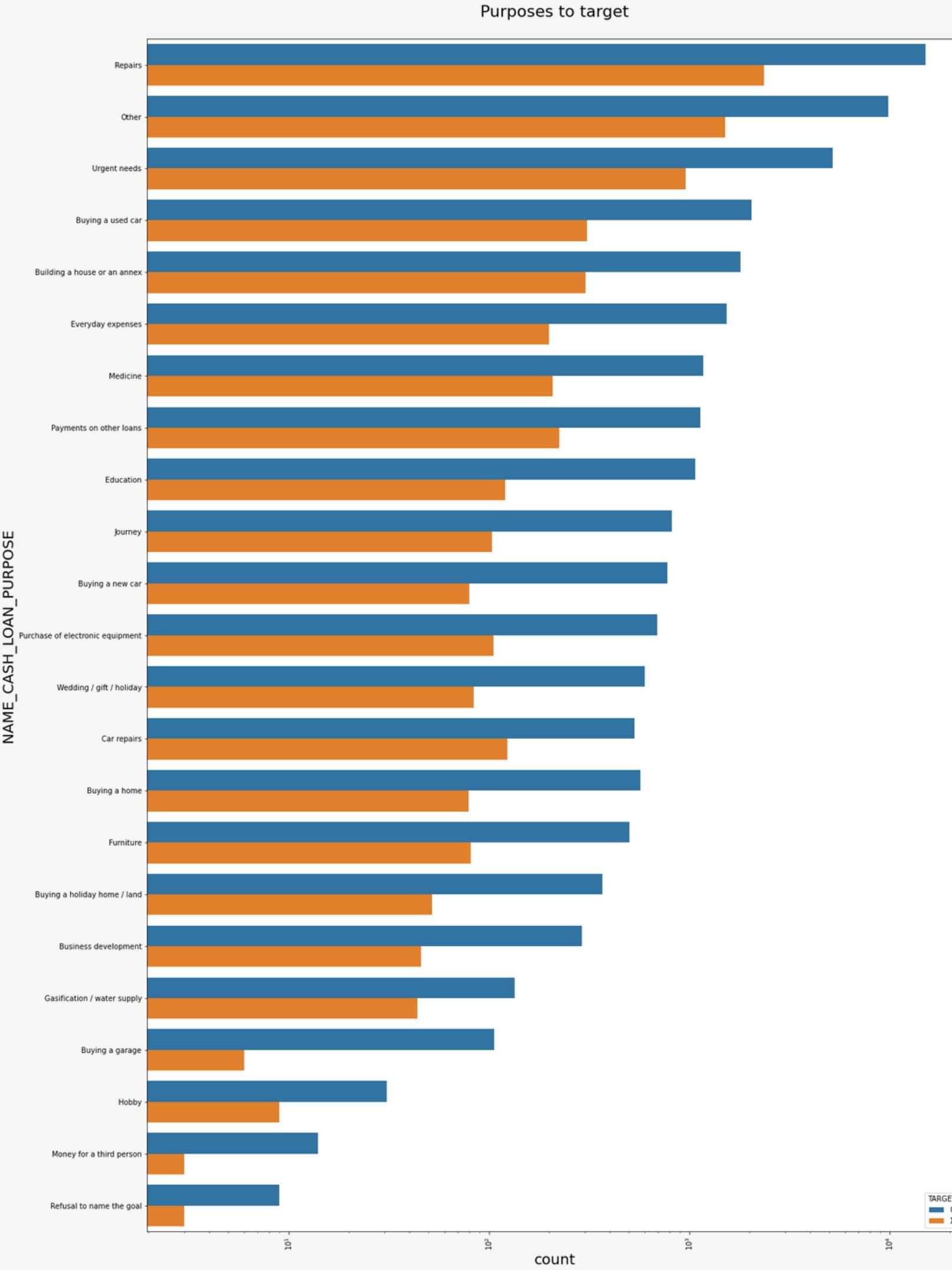
Contract status to purposes

1. 'repairs' has most of the rejections
2. education purposes there are equal approves and rejection
3. Paying other loans and buying a new car is having higher rejection than that of approves.



Purposes to target

1. 'Repairs' are facing difficulties in payment on time.
2. 'Buying a garage', 'Business developemt', 'Buying land','Buying a new car' and 'Education' are paying loans at high rate and they are not facing any difficulties.
Hence loan can be given for above mentioned reasons as it will benifit the organisation.

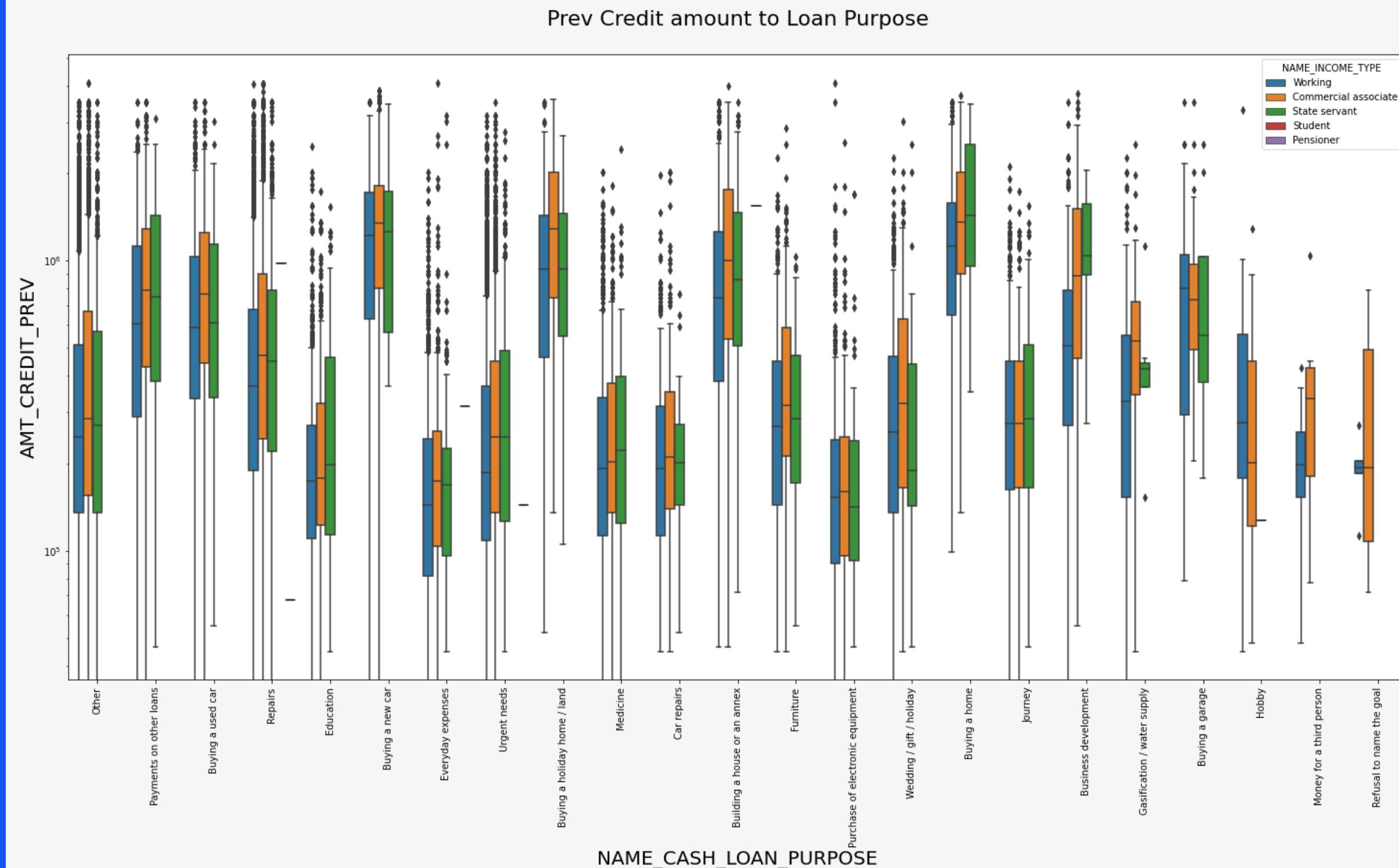




Univariate Analysis After Merging Previous data

Prev Credit amount to Loan Purpose

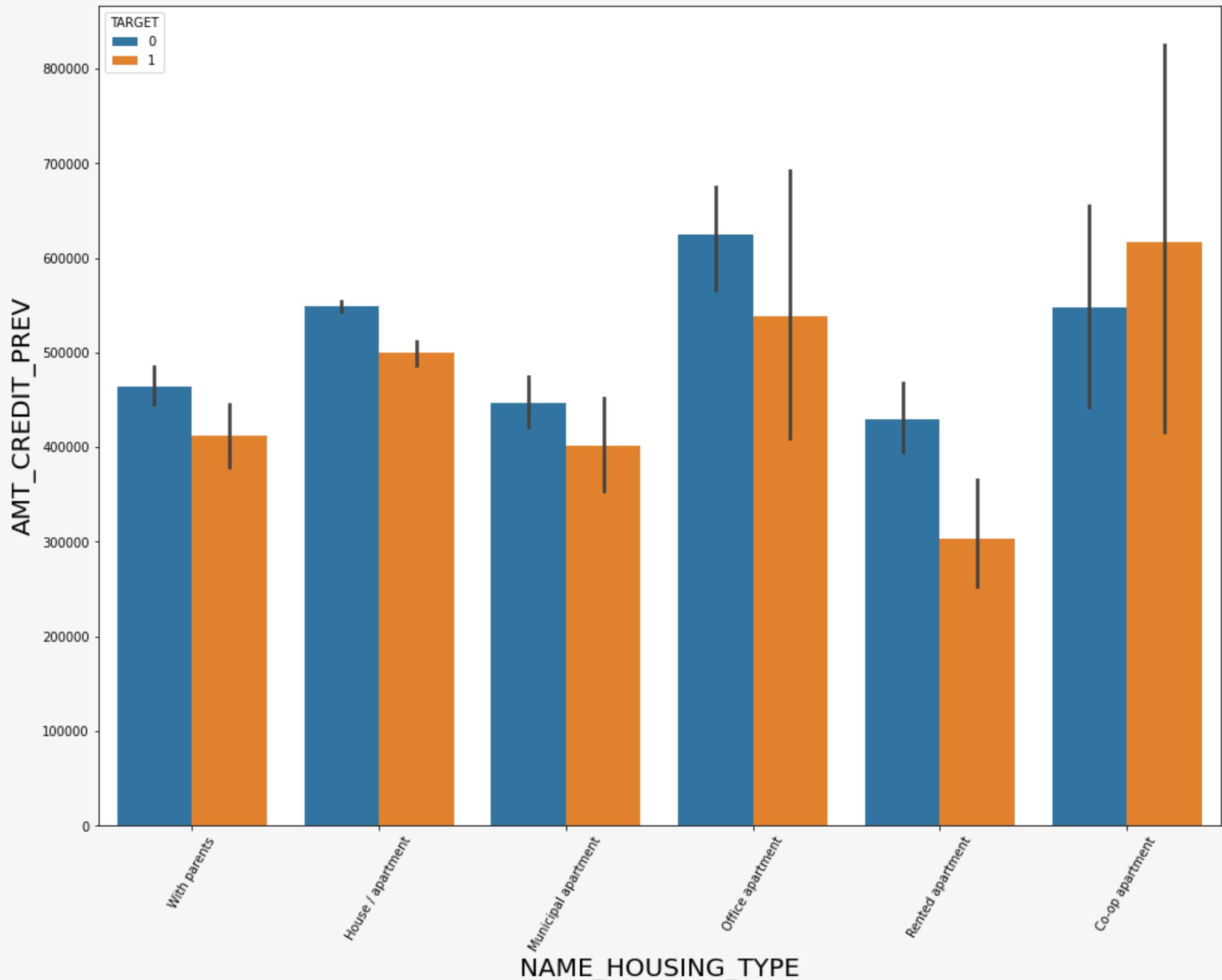
1. 'Buying a home','Buying a land','Buying a new car' and 'Building a house' are having higher credit ammount.
2. State servants have high applicants rate.
3. Money for third person or a Hobby have less applicants rate.



Prev Credit amount to Housing type

1. Housing type, office apartment are having higher credit of target 0 and co-op apartment is having higher credit of target 1.
2. Bank should avoid giving loans to housing type of co-op apartment because of difficulties in payment.
3. Housing type with parents or House\apartment or miunicipal apartment are the areas where banks should pay their attention for successful payments.

Prev Credit amount to Housing type



Conclusion

1. Housing type 'With parents' should be the ideal customer and banks should focus on getting customers from this area as they have very less payment failures.
2. Contract type 'Student', 'pensioner' and 'Businessman' with housing type other than 'Co-op apartment' is the area where bank should put more attention for successful payments.
3. Income type 'Working' should not be focused much as for gaining new customers from this area as they have most of unsuccessful payments.
4. Loan purpose 'Repair' is having higher number of unsuccessful payments this category should also be avoided for giving loans.



Thank you!