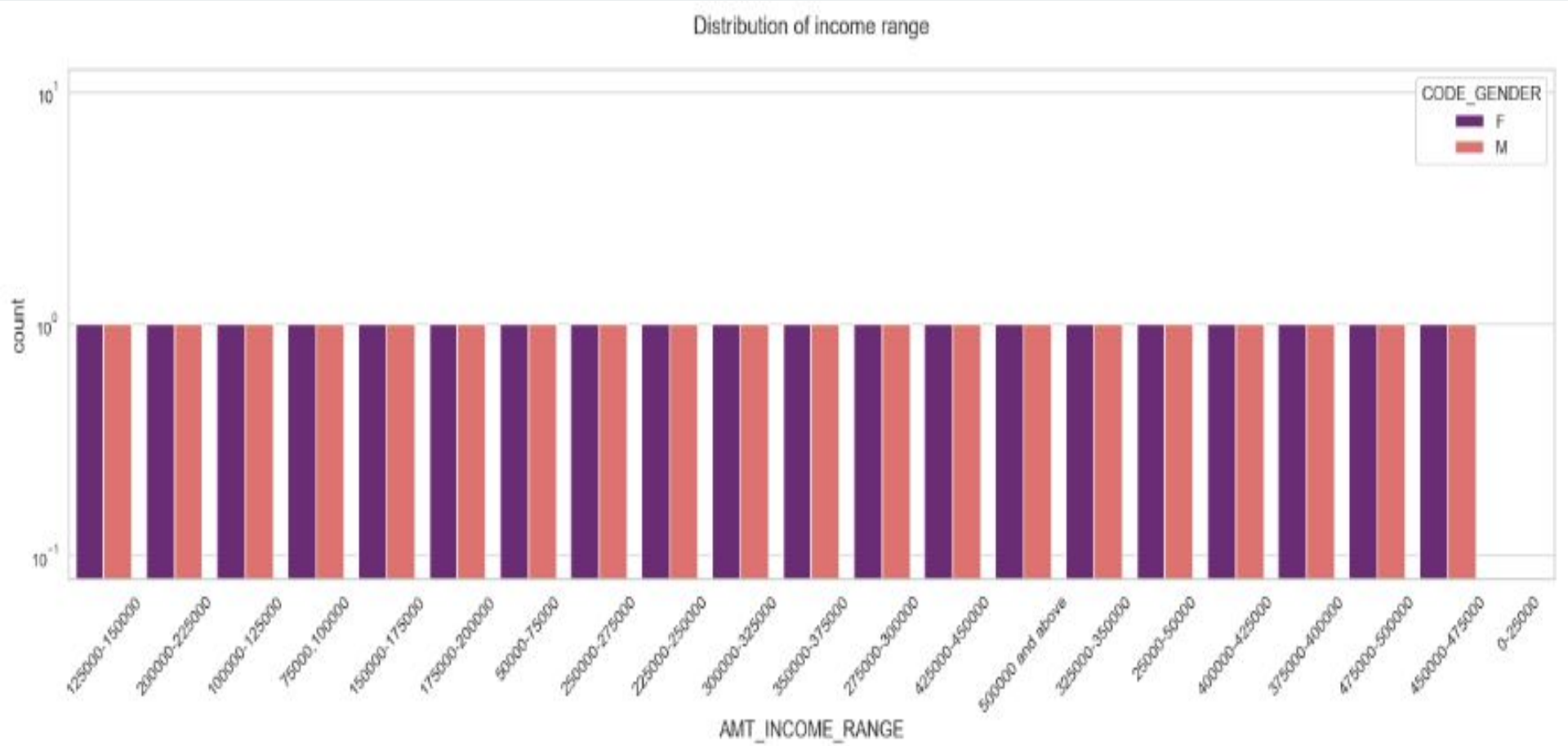
CREDIT EDA CASE STUDY

BY SHAYAMON BASTAKOTI

Categorical Univariate analysis for Target = 0

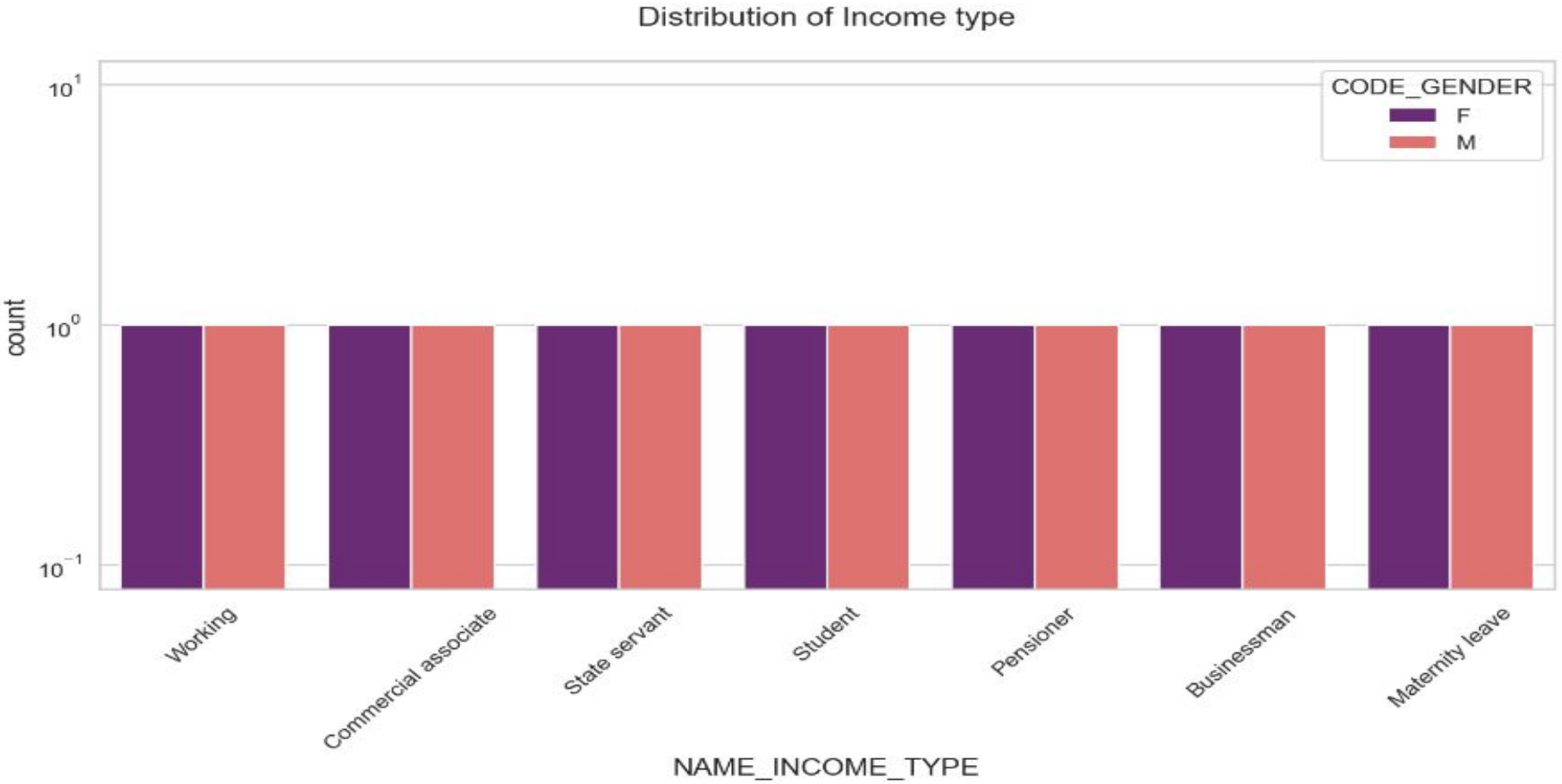
Plotting for income range



Conclusion

- 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.

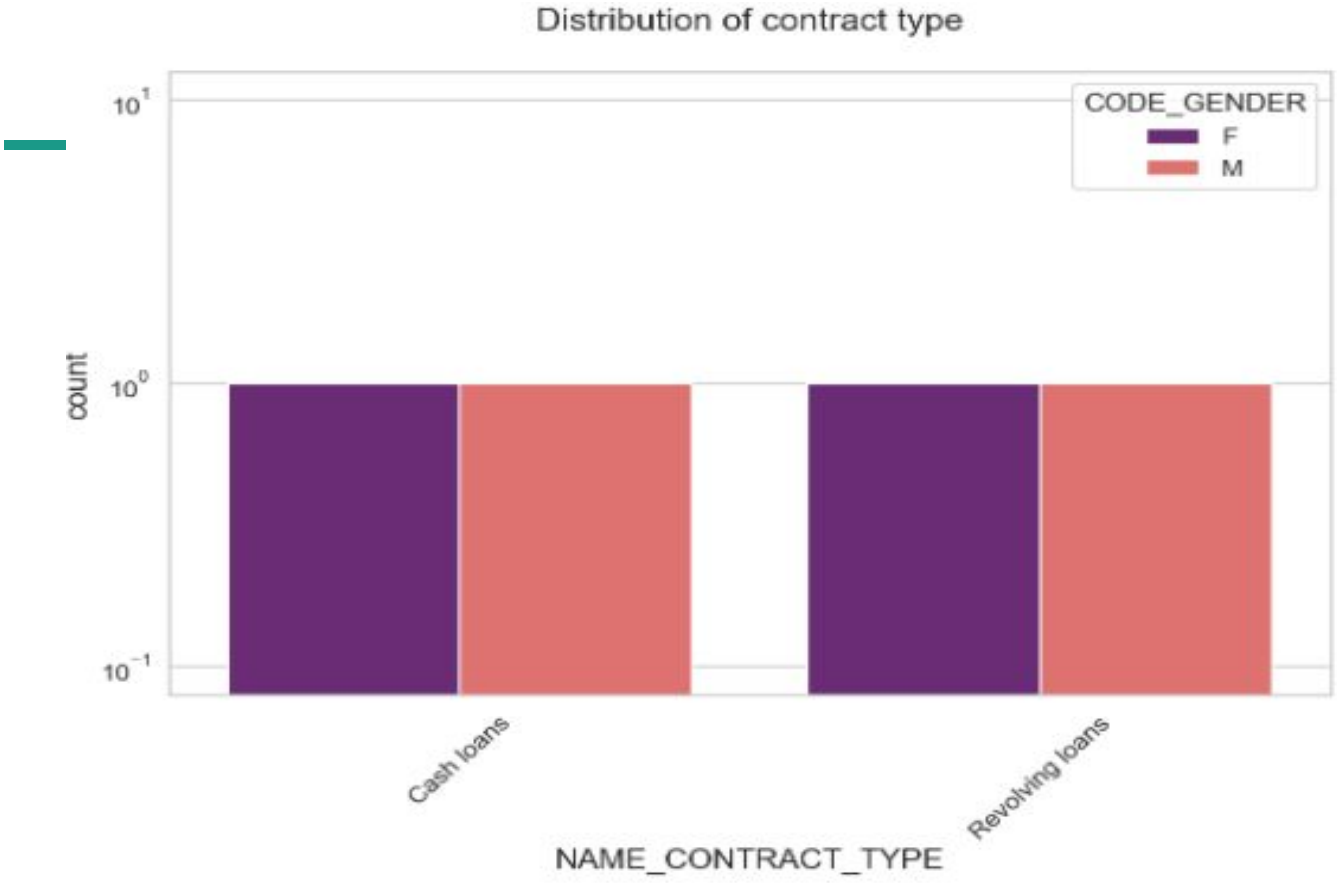
Plotting for Income type



Conclusion

- 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’.

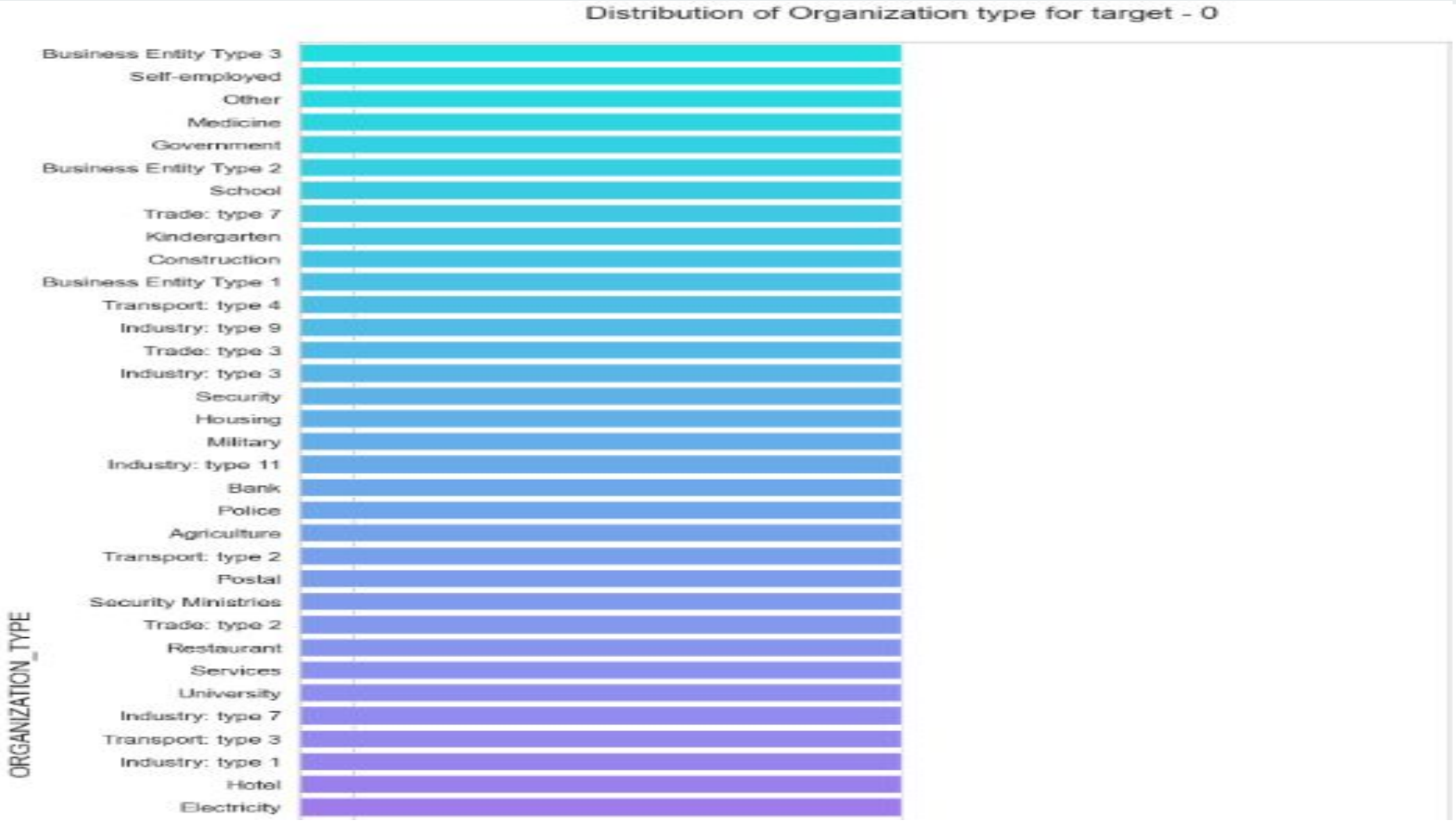
Distribution for Contract type



Points to be concluded from the above 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.

Distribution for Organization type for target



Conclusion

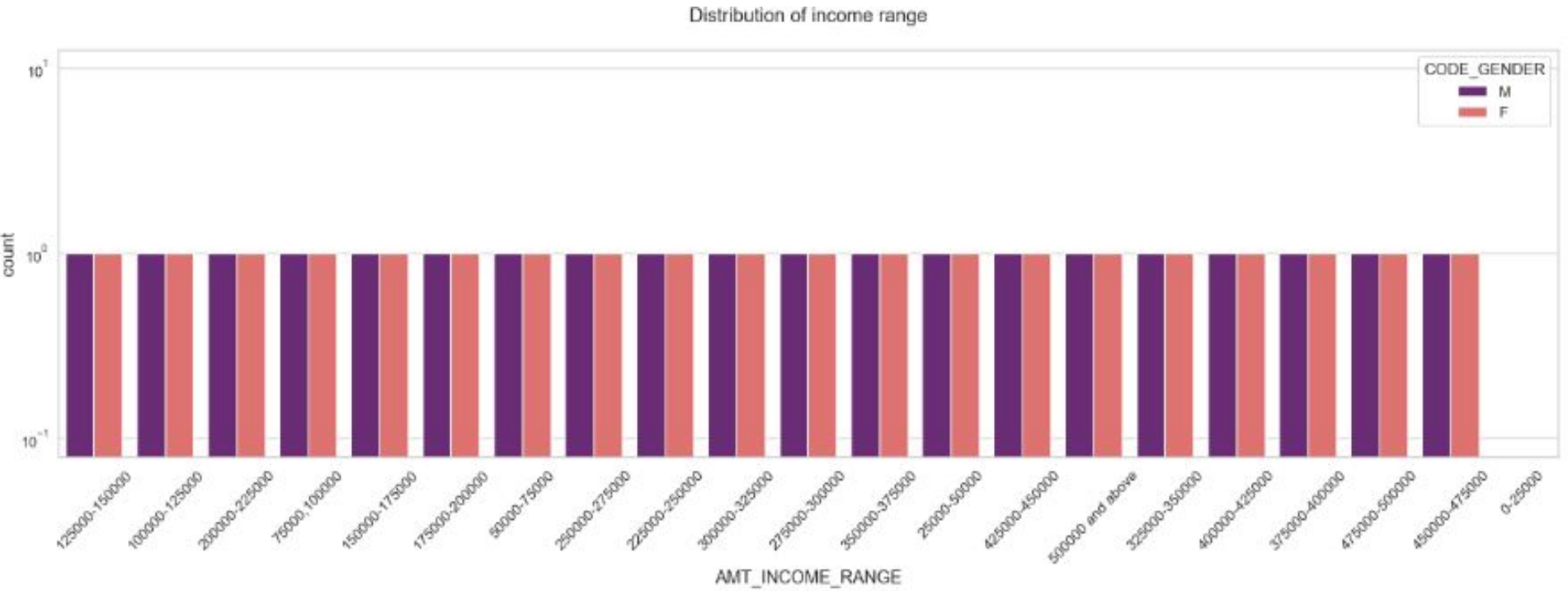
- Clients which have applied for credits are from most of the organization type ‘Business entity Type 3’ , ‘Self employed’,

  ‘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

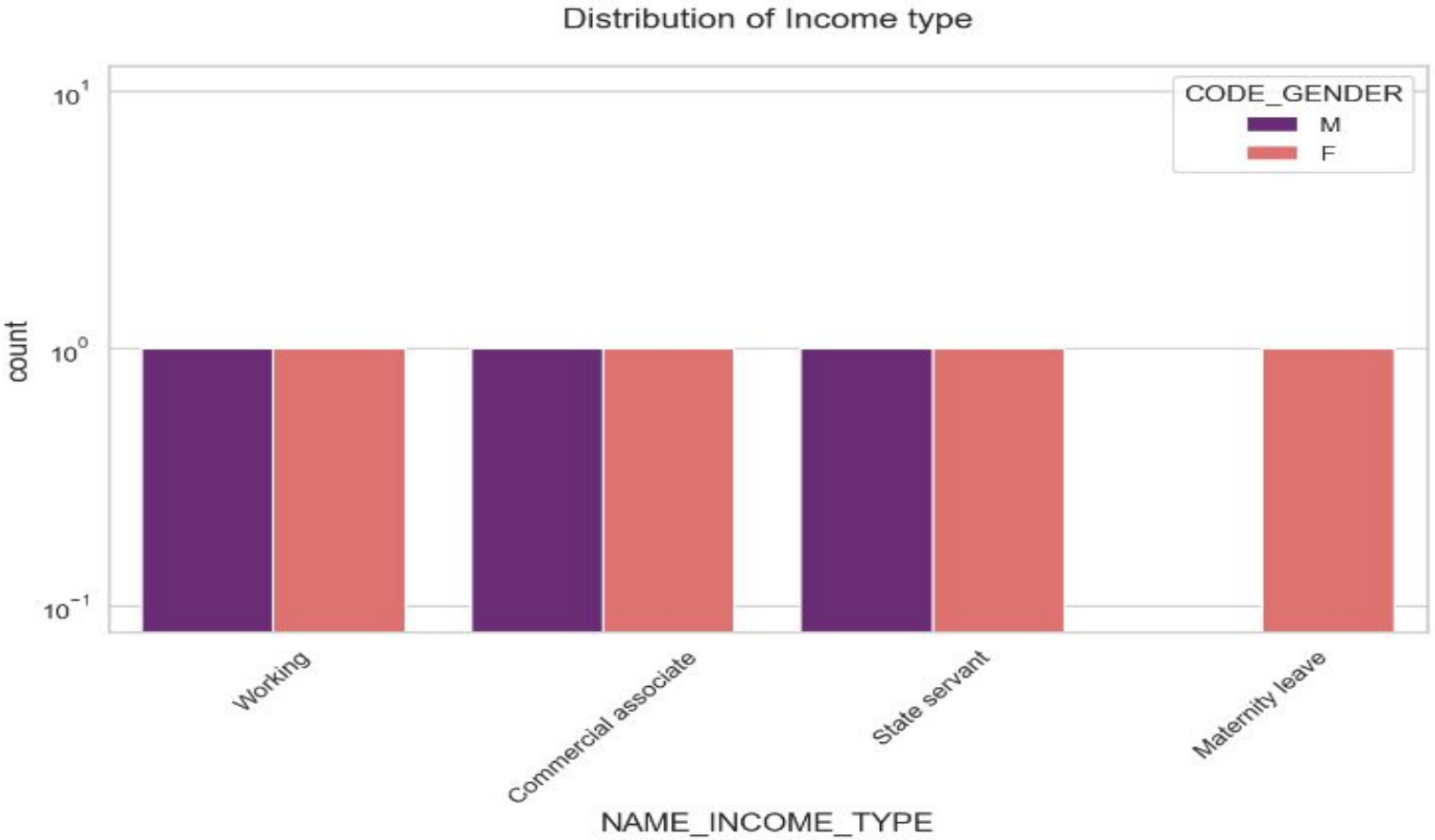
Plotting for income range



Points to be concluded from the above 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.

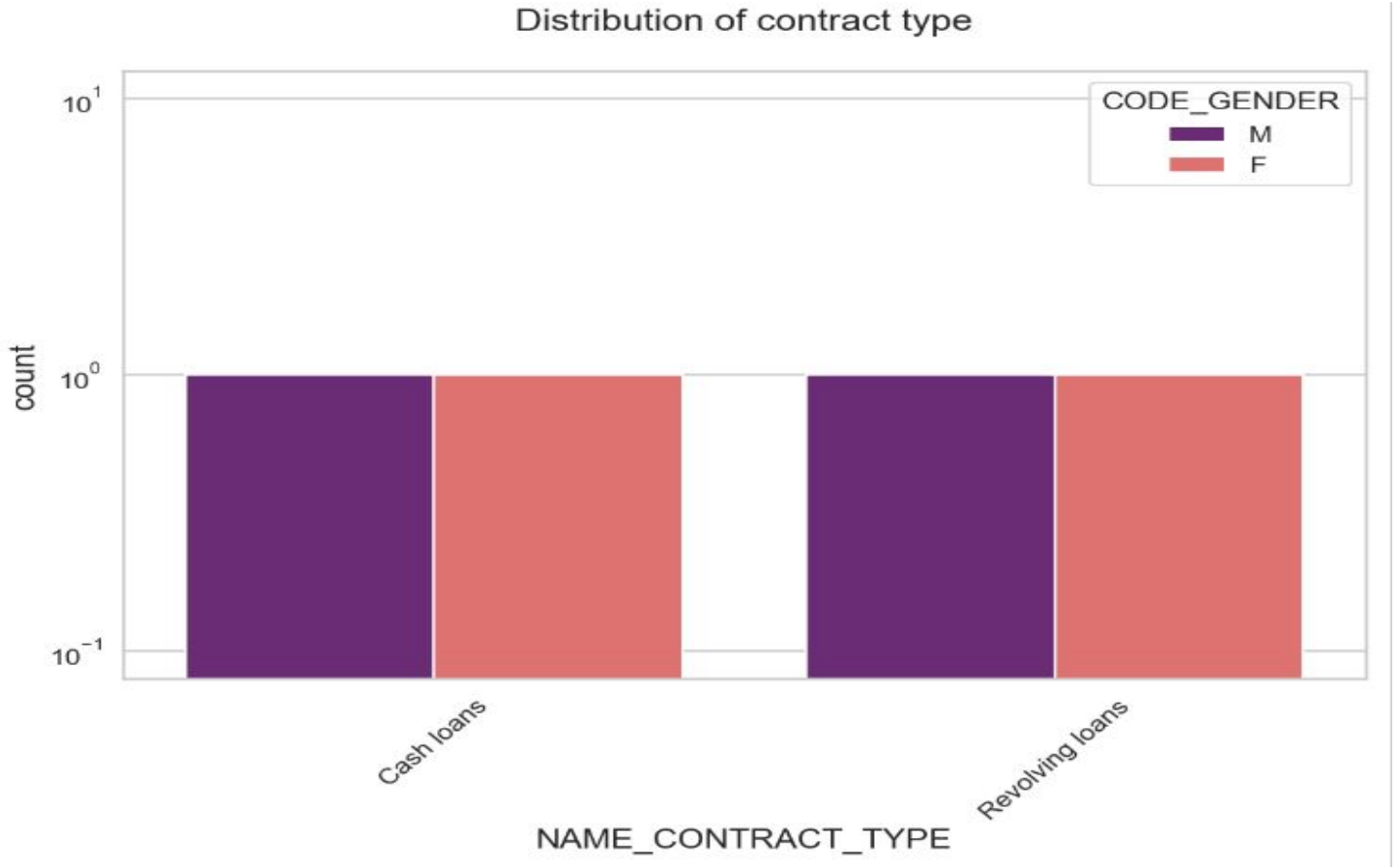
Plotting for Income type



Points to be concluded from the above 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.

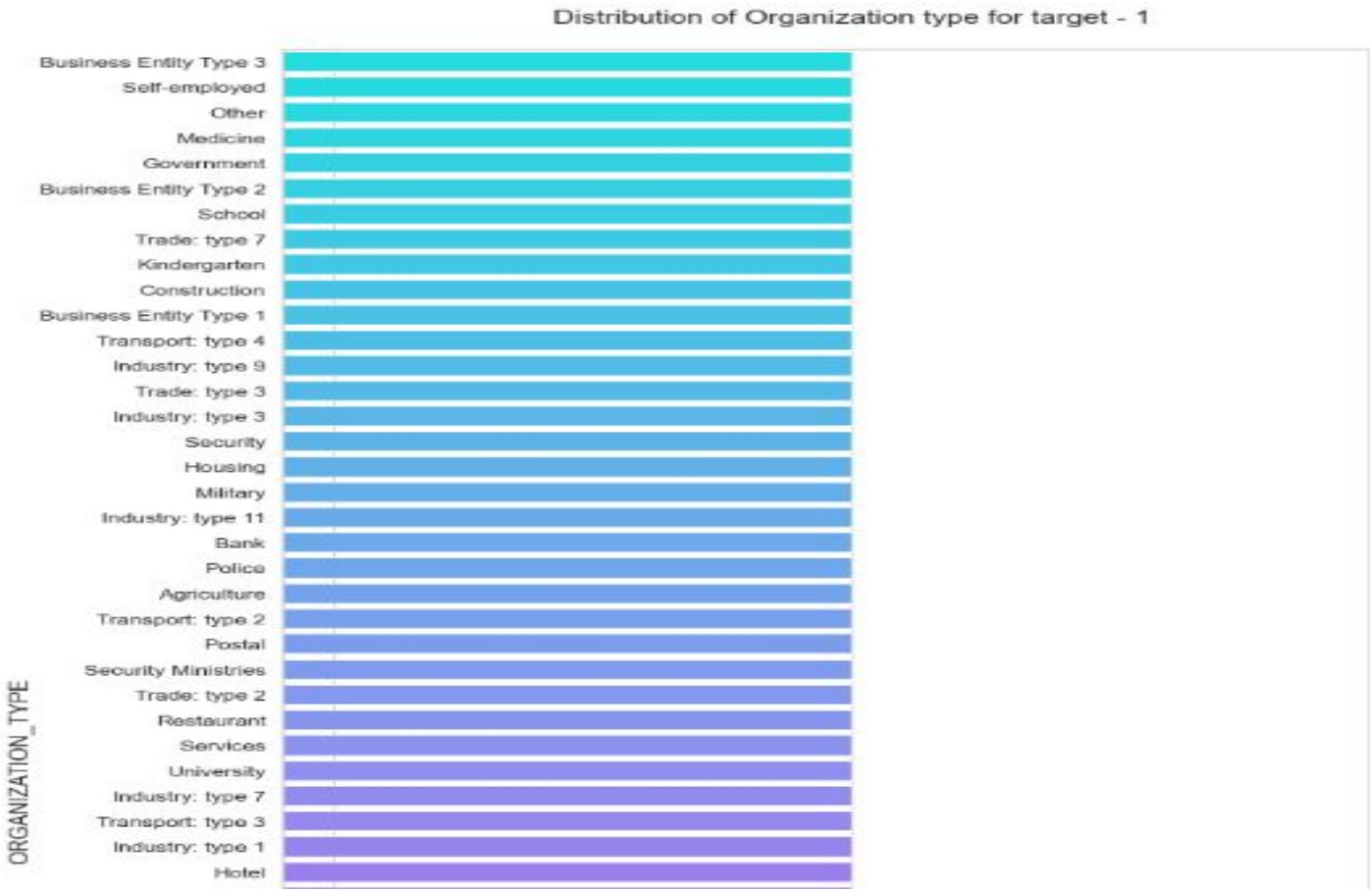
Plotting for Contract type



Points to be concluded from the above 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.

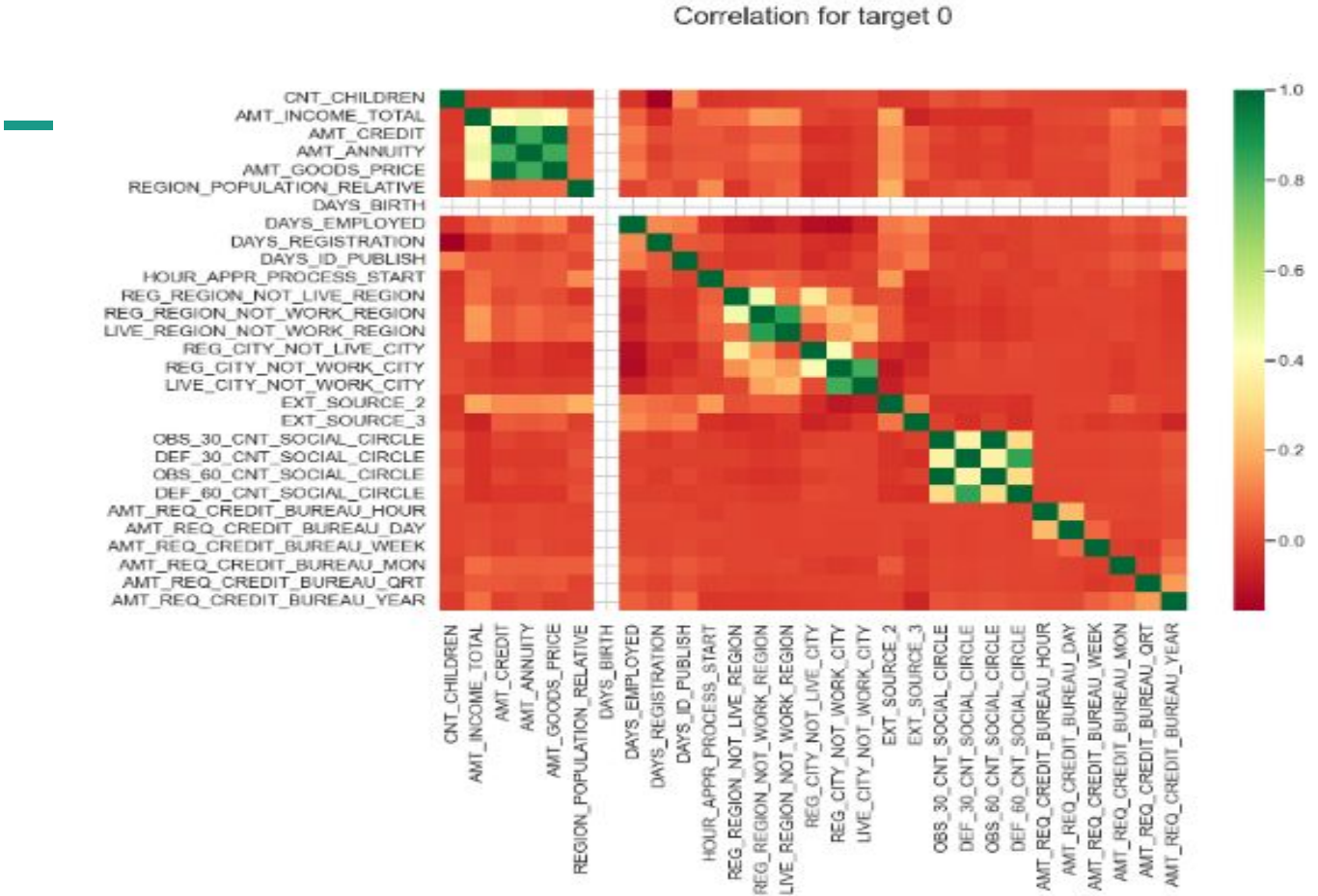
Distribution for Organization type



Points to be concluded from the above 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.
- Same as type 0 in distribution of organization type.

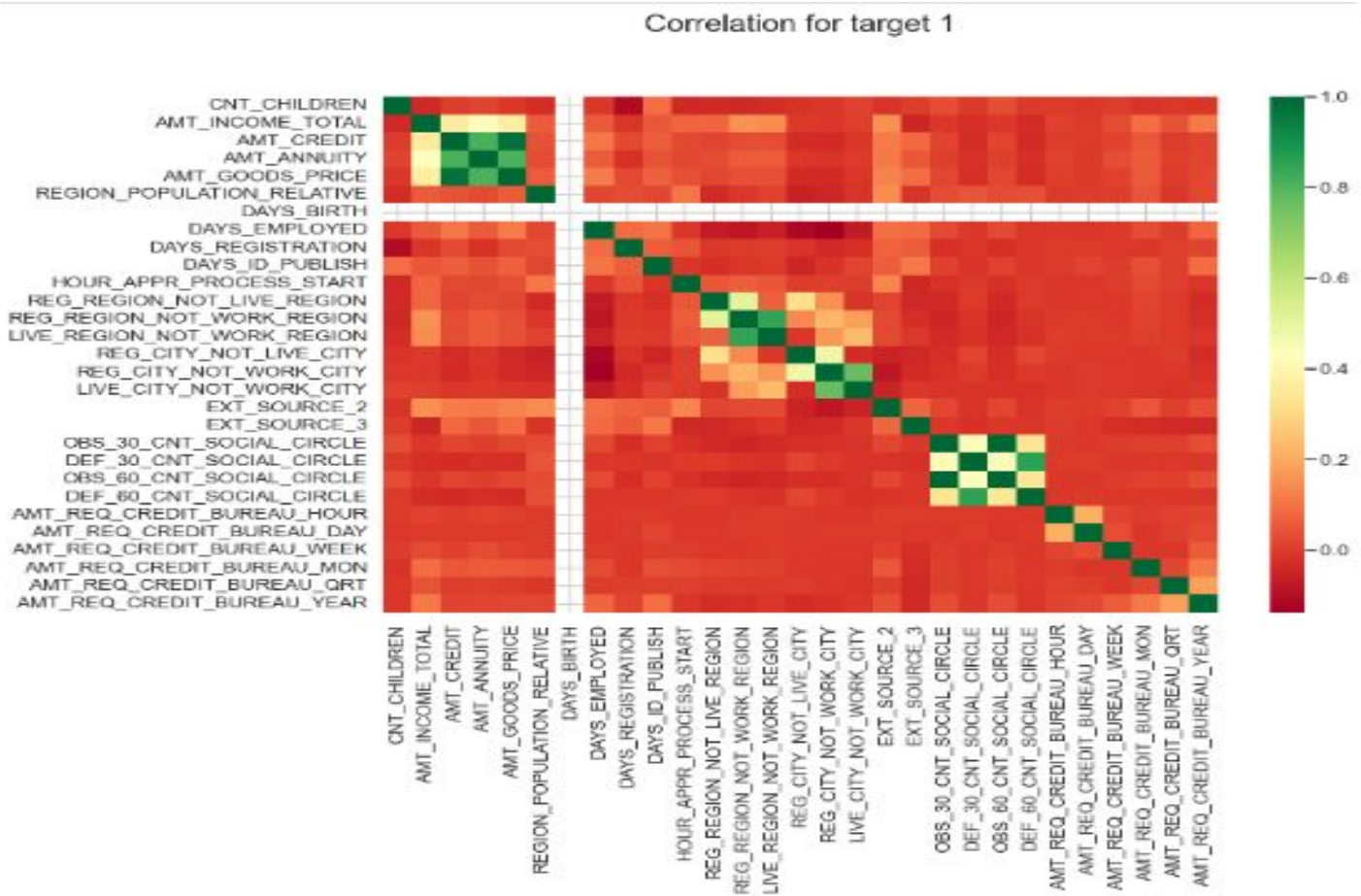
Plotting correlation for target-0



Observation we can point out

- 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.
- less children client have in densely populated area.
- Credit amount is higher to densely populated area.
- The income is also higher in densely populated area.

Correlation for target-1

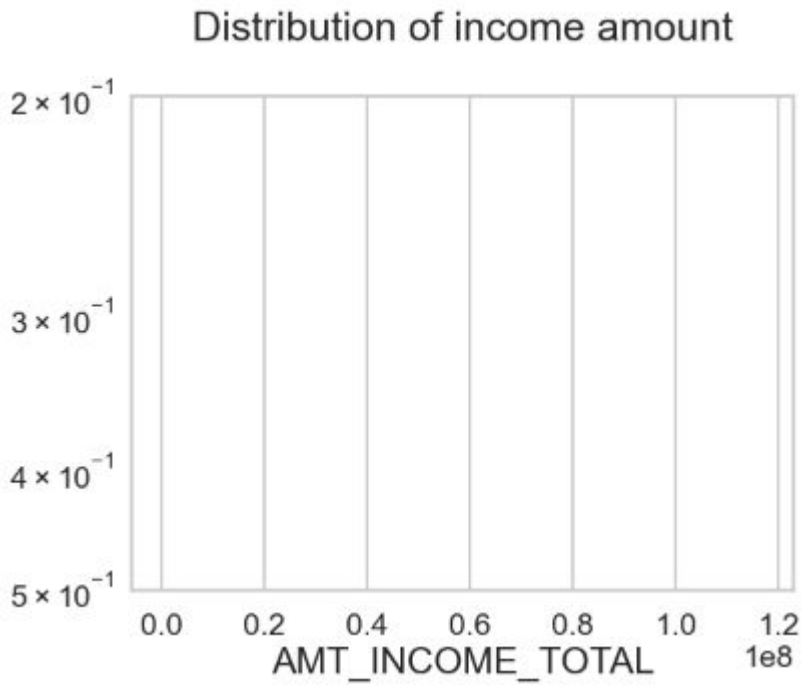


This heat map for Target 1 is also having quite a same observation just like Target 0.

- 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

For Target 0 - Finding any outliers

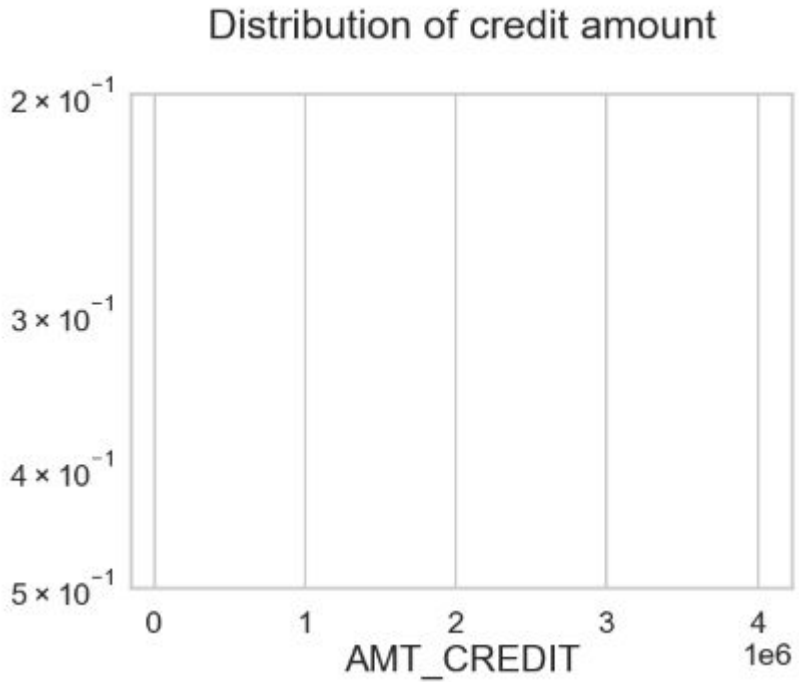
Distribution of income amount



Few points can be concluded from the graph above.

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

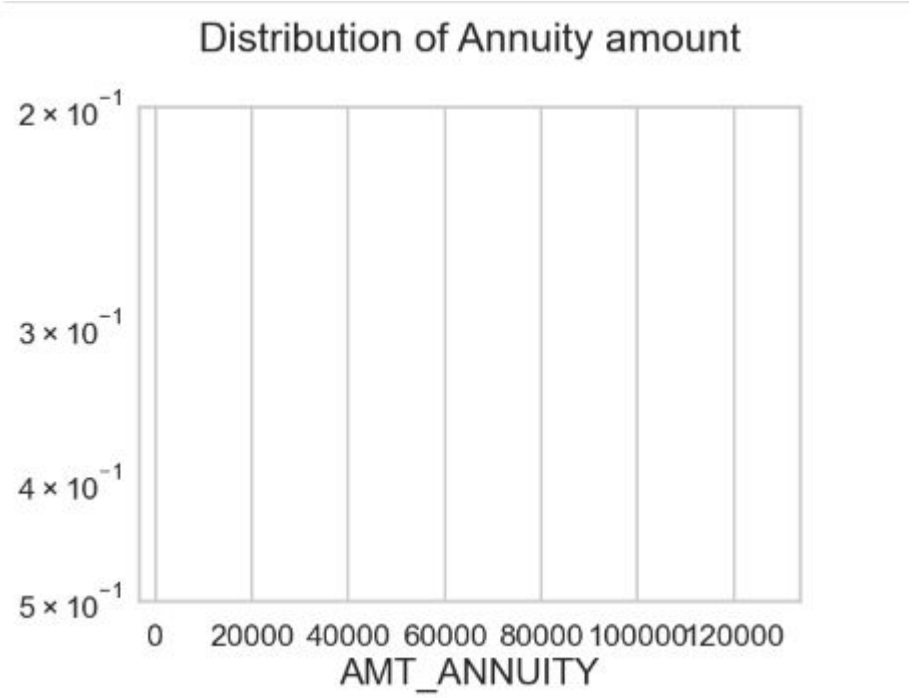
Distribution of credit amount



Few points can be concluded from the graph above.

- 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.

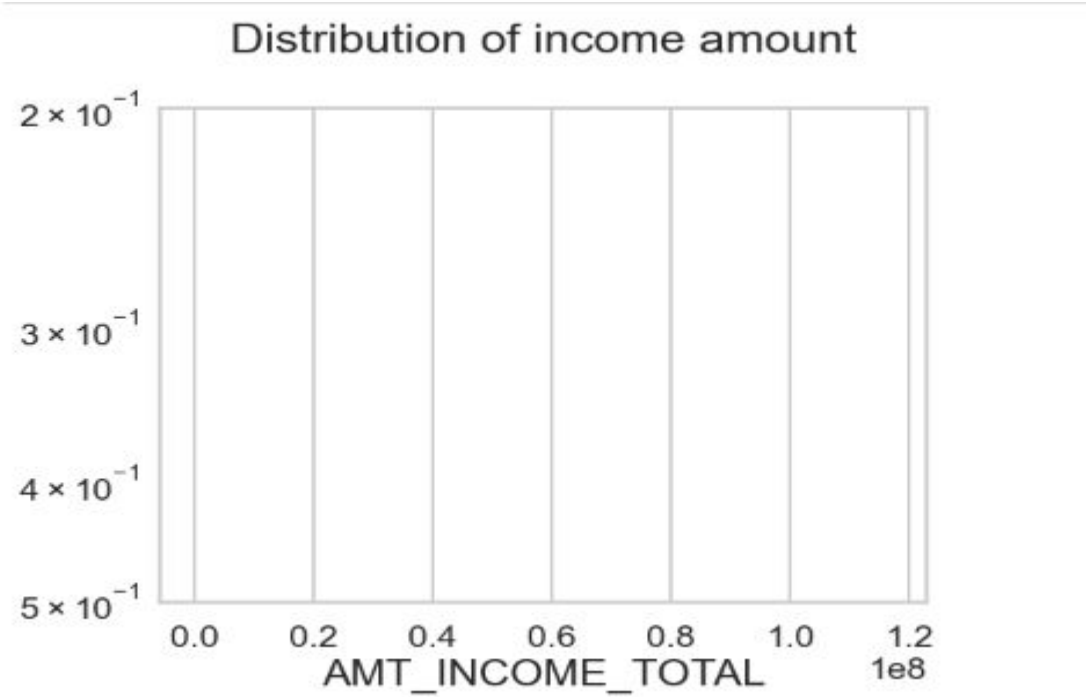
Distribution of annuity amount



Few points can be concluded from the graph above.

- 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.
- For Target 1 - Finding any outliers

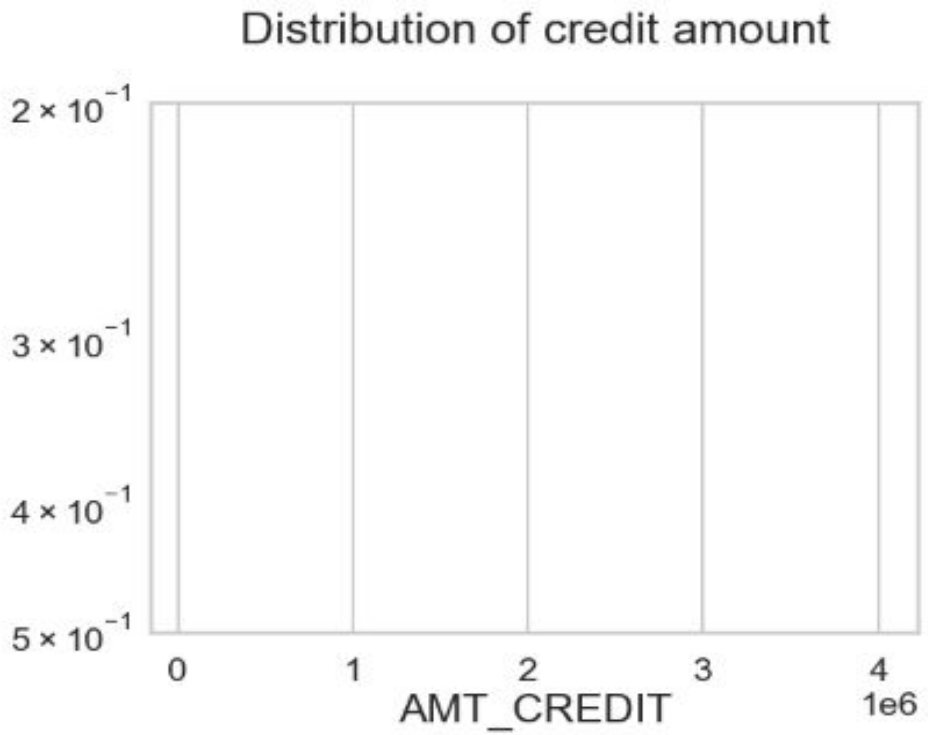
Distribution of income amount



Few points can be concluded from the graph above.

- 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.

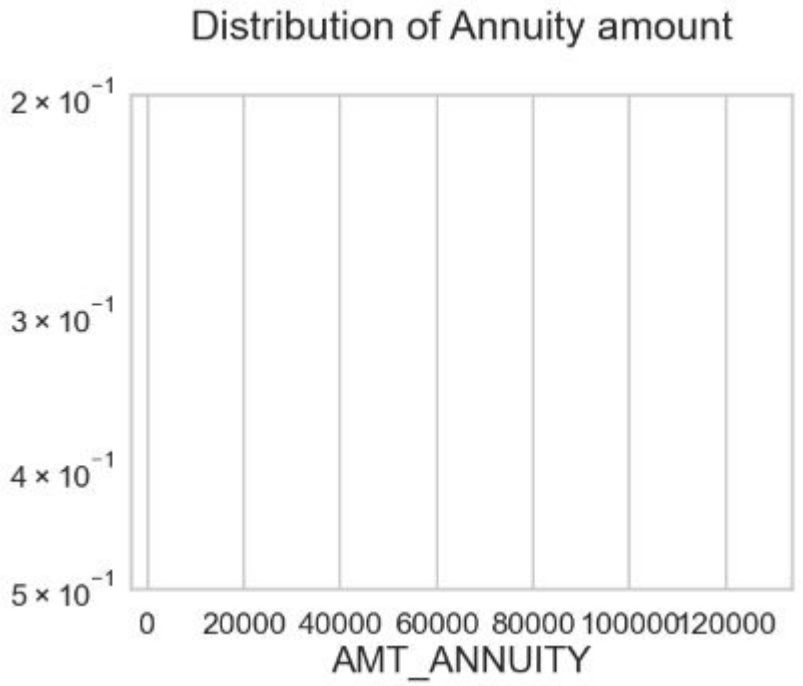
Distribution of credit amount



Few points can be concluded from the graph above.

- 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.

Distribution of Annuity amount

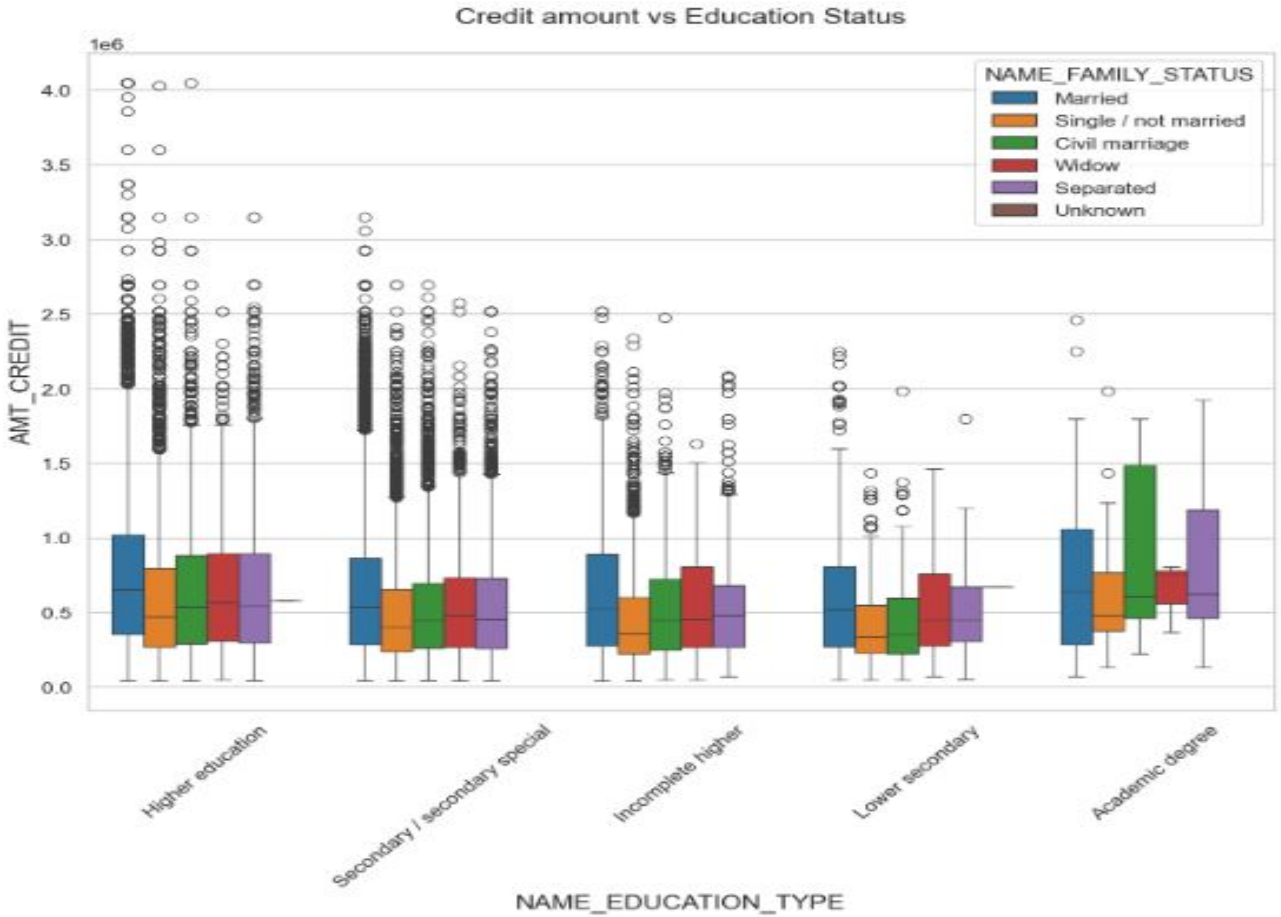


Few points can be concluded from the graph above.

- 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 numerical variables
For Target 0

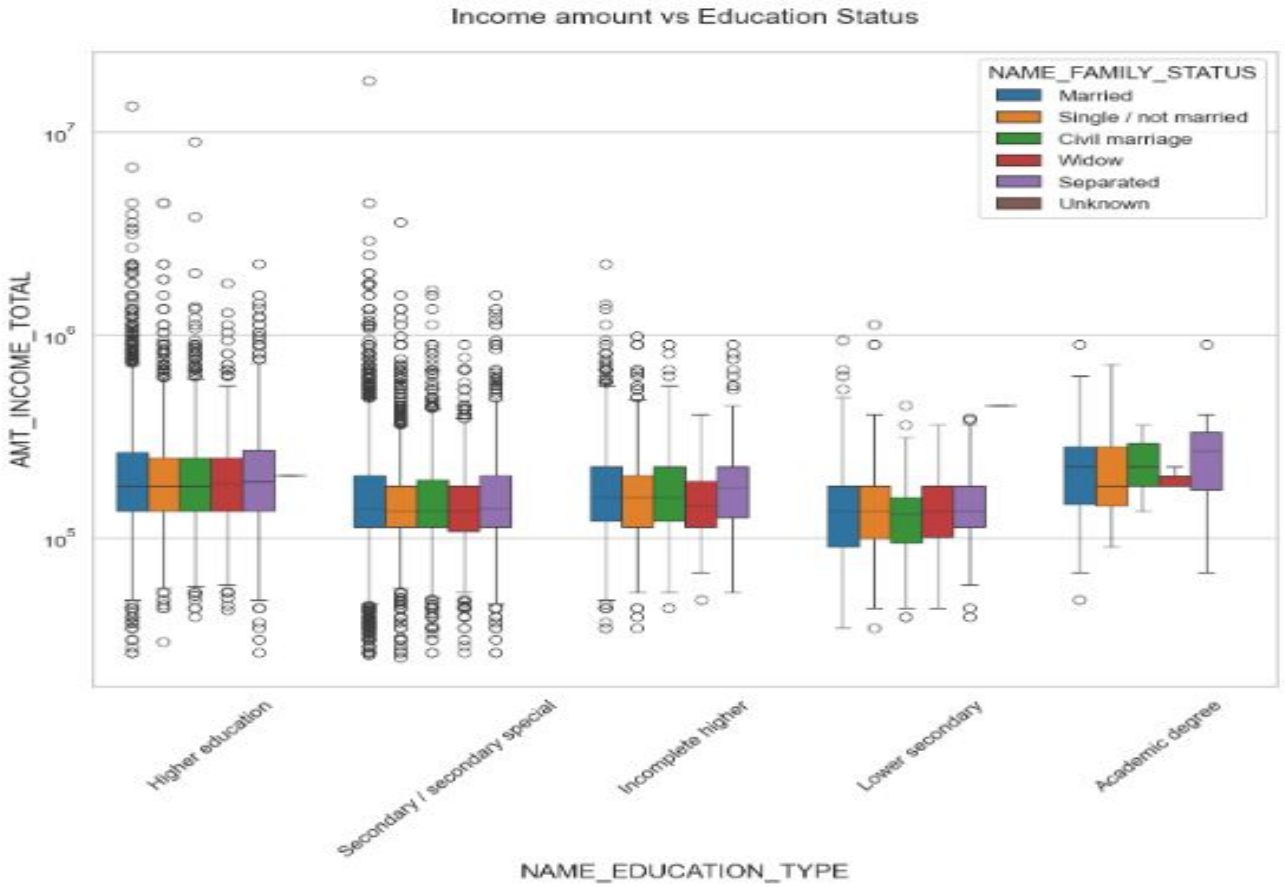
Box plotting for Credit amount



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. 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.

Box plotting for Income amount in logarithmic scale

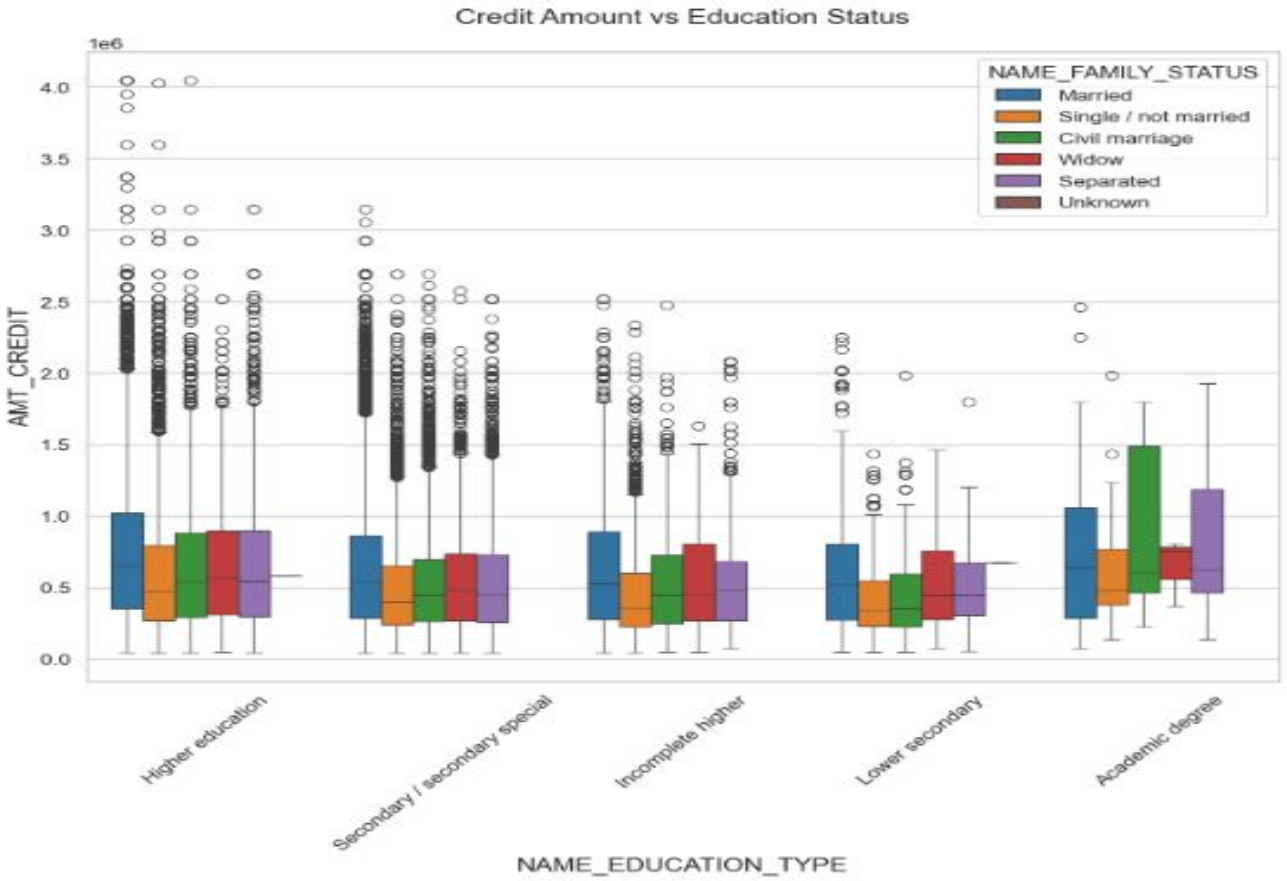


From above boxplot for Education type 'Higher education' the income amount is mostly equal with family status. It does contain many outliers. Less outlier are having for Academic degree but there income amount is little higher that Higher education.

Lower secondary of civil marriage family status are have less income amount than others.

Bivariate analysis for numerical variables
For Target 1

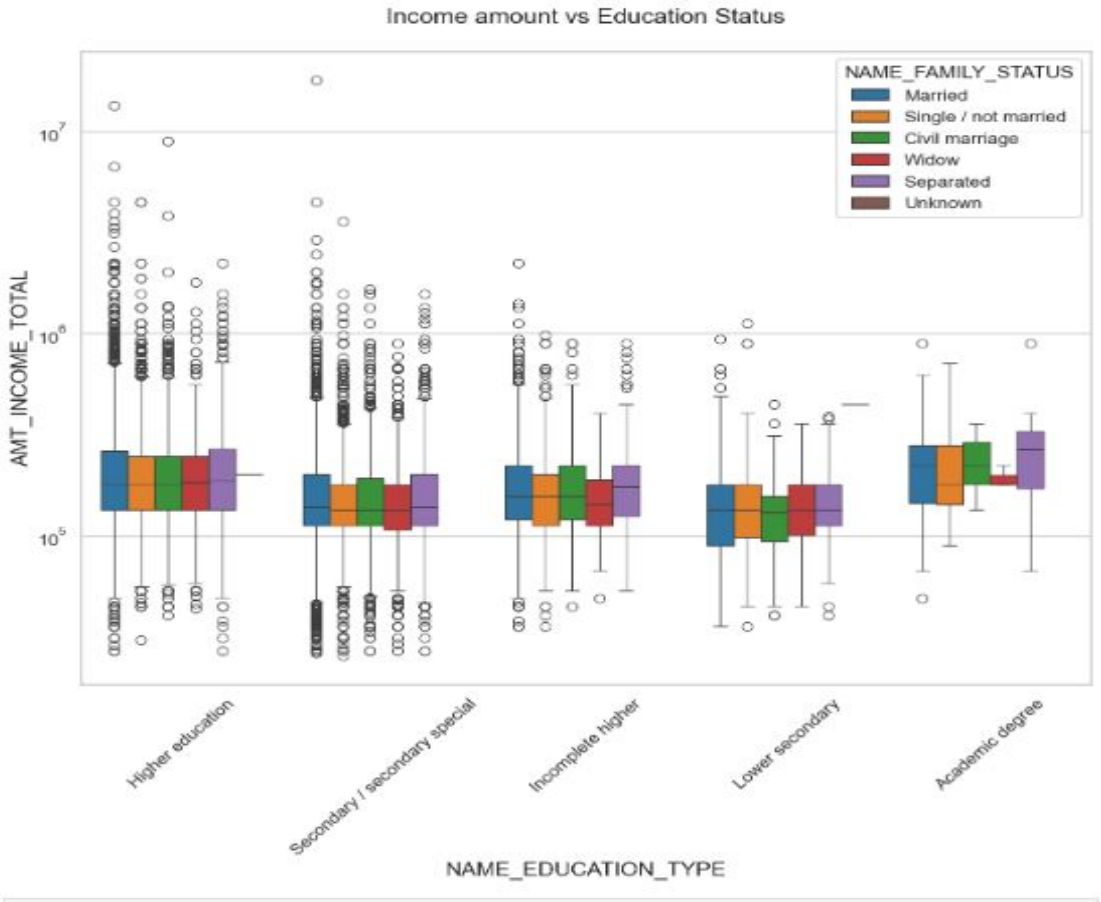
Box plotting for credit amount



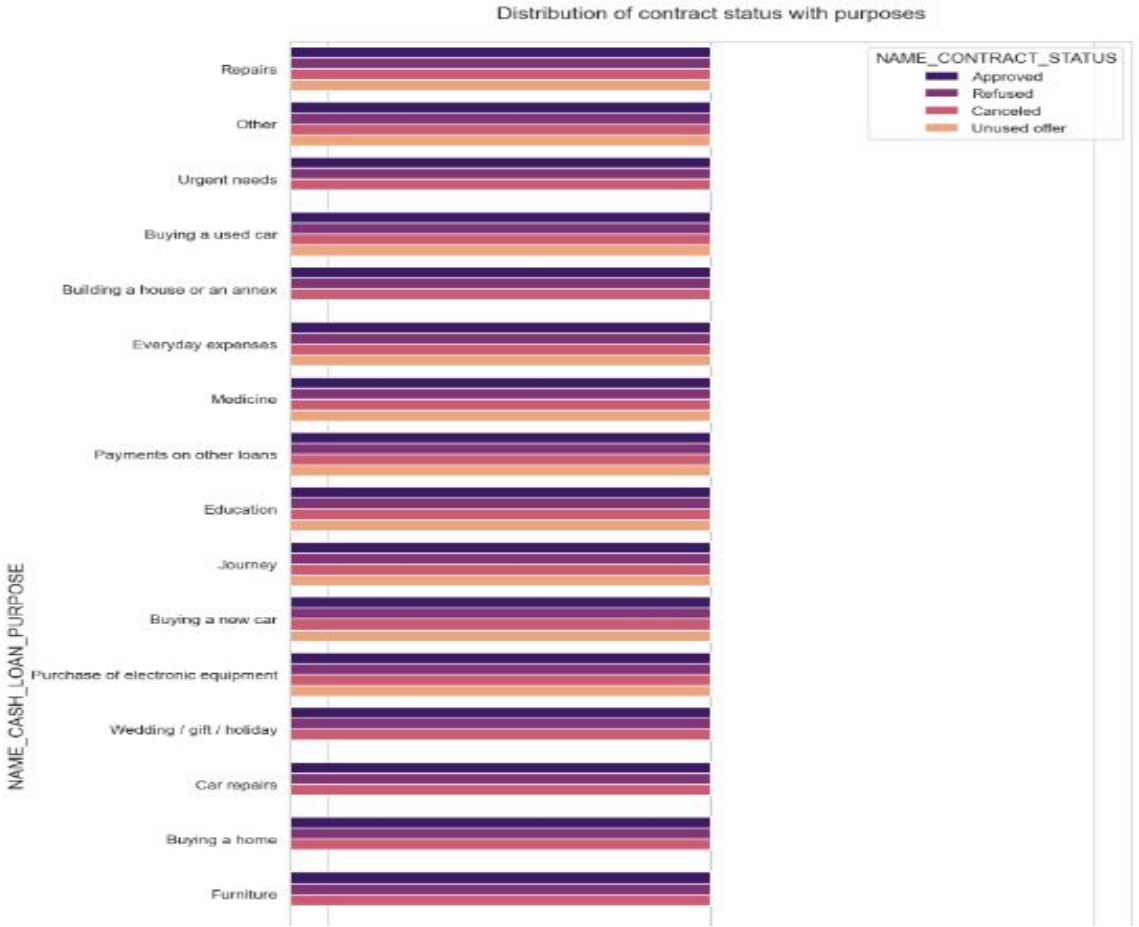
Quite similar with Target 0 From the above box plot we can say that Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree

education are having higher number of credits than others. 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.

Box plotting for Income amount in logarithmic scale



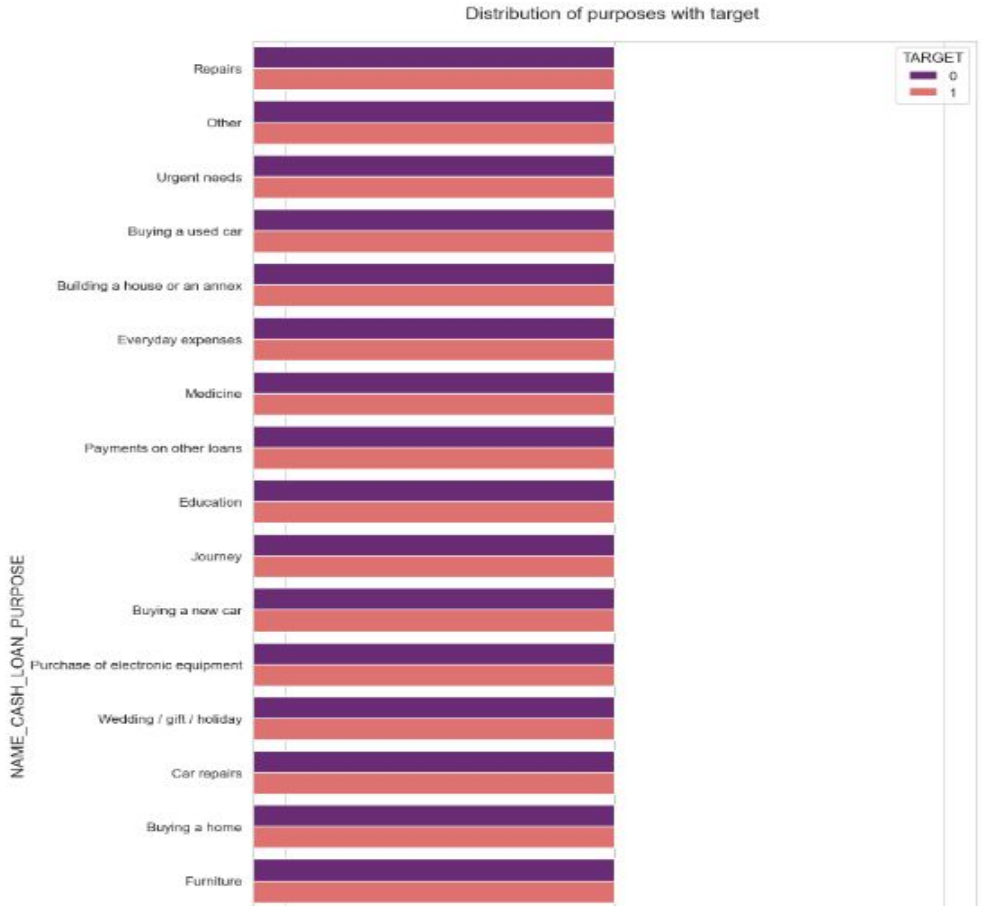
Distribution of contract status in logarithmic scale



Points to be concluded from above plot:

- 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 contract status



Few points we can conclude from above plot:

- Loan purposes with 'Repairs' are facing more difficulties in payment on time.
- There are few places where loan payment is significant 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.

CONCLUSION

- Banks should prioritize clients with contract types such as ‘Student’, ‘Pensioner’, and ‘Businessman’ who reside in housing types other than ‘Co-op apartment’ for higher chances of successful payments.
- Bank should reduce focus on clients with the income type ‘Working’, as they have the highest rate of unsuccessful payments.
- Additionally, loans for the purpose of ‘Repair’ tend to have a greater number of missed payments.
- Banks should aim to attract more clients from the housing type ‘With parents’, as this group has the lowest rate of missed payments.