

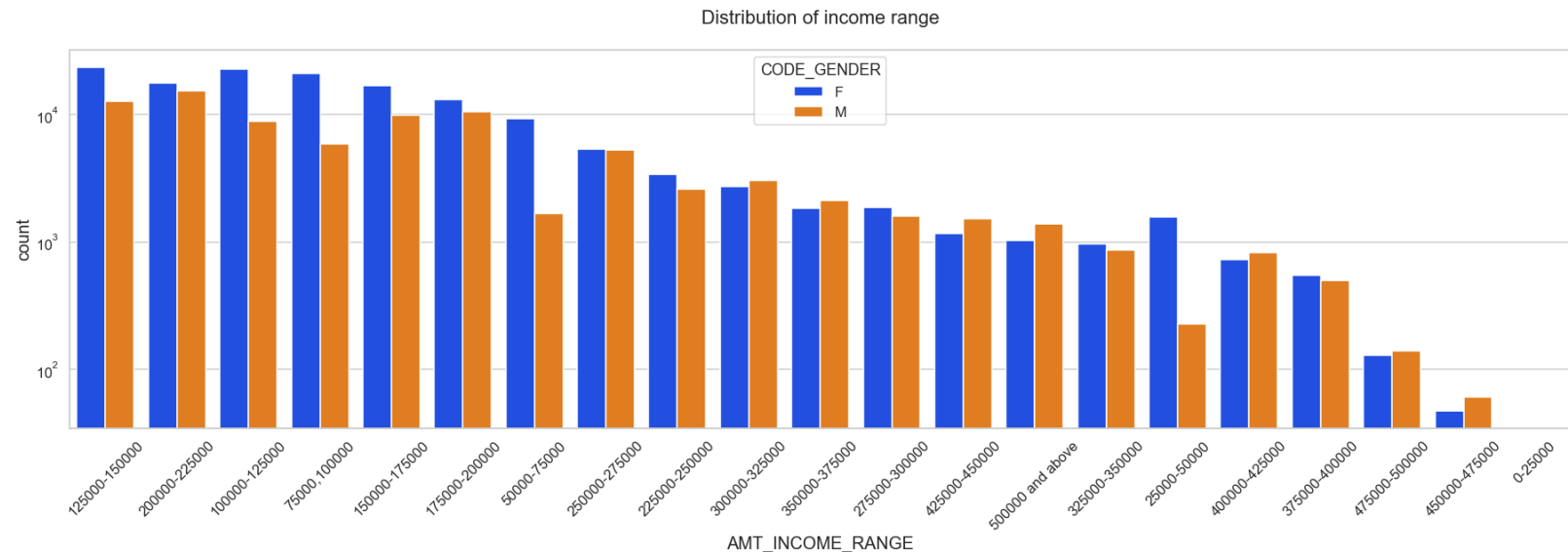
# CREDIT EDA CASE STUDY

By AYUSHYA SIDHNATH and  
PRATIK KUMAR

# Distribution of Income range

Points to be concluded from the graph on the right side.

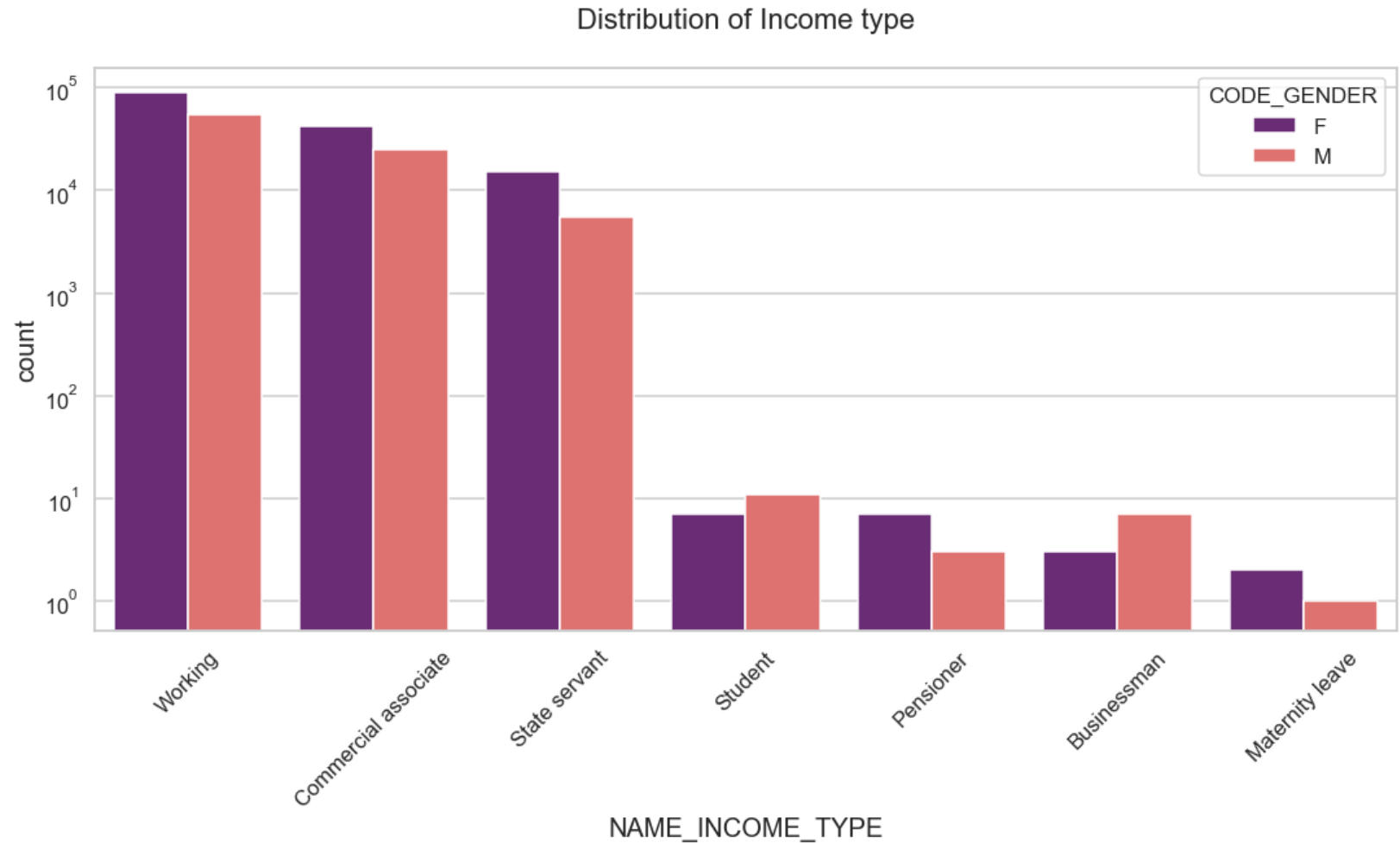
- ▣ Female counts are higher than male.
- ▣ Income range from 100000 to 200000 is having more number of credits.
- ▣ This graph show that females are more than male in having credits for that range.
- ▣ Very less count for income range 400000 and above.



# Distribution of income type

Points to be concluded from the graph on the right.

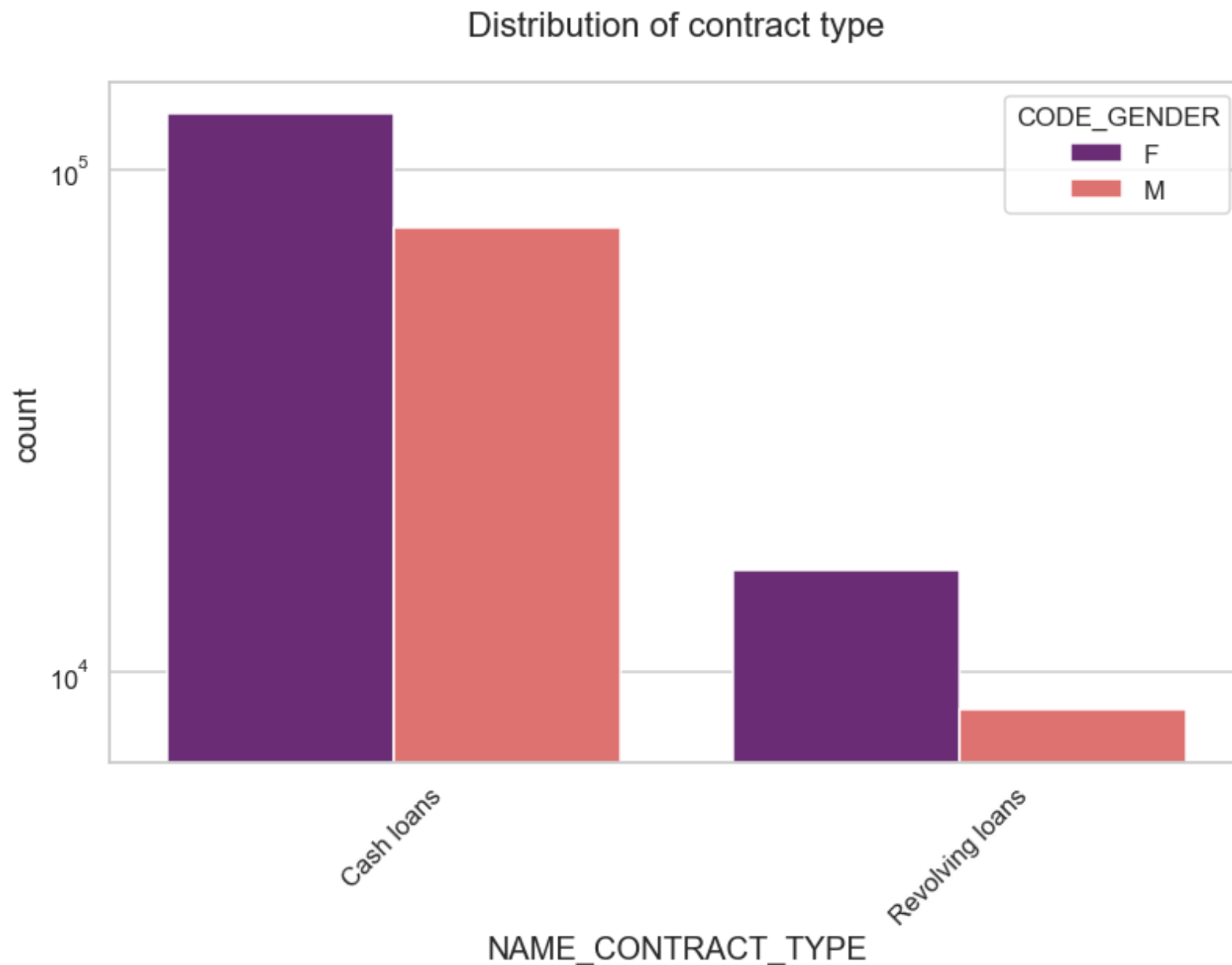
- ▣ For income type 'working', 'commercial associate', and 'State Servant' the number of credits are higher than others.
- ▣ For this Females are having more number of credits than male.
- ▣ Less number of credits for income type 'student', 'pensioner', 'Businessman' and 'Maternity leave'.



# Distribution for contract type

Points to be concluded from the graph on the right.

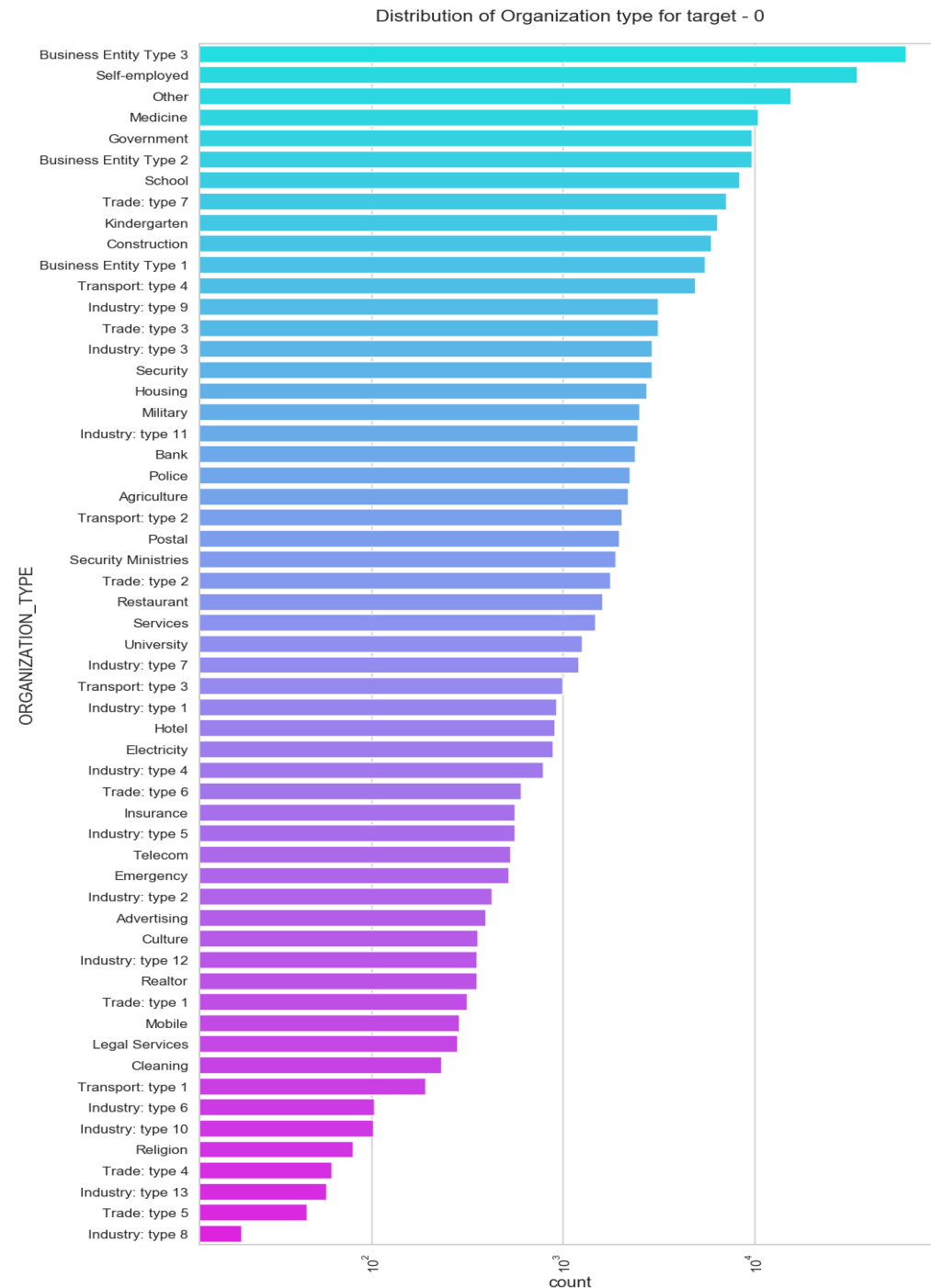
- ▣ For contract type 'cash loans' is having higher number of credits than 'Revolving loans' contract type.
- ▣ For this also Female is leading for applying credits.



# Distribution of organization type

Points to be concluded from the graph on the right.

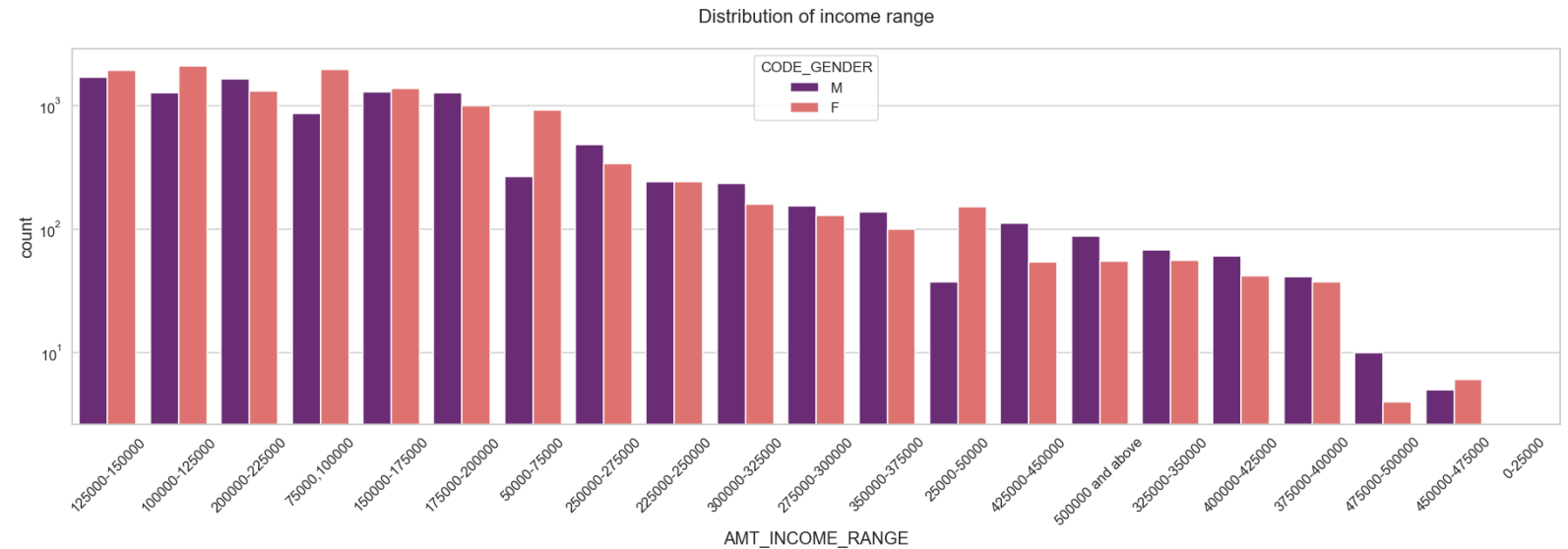
- ▣ Clients which have applied for credits are from most of the organization type 'Business entity Type 3', 'Self employed', 'Other', 'Medicine' and 'Government'.
- ▣ Less clients are from Industry type 8, type 6, type 10, religion and trade type 5, type 4.



# Distribution of Income range

Points to be concluded from the graph on the right side.

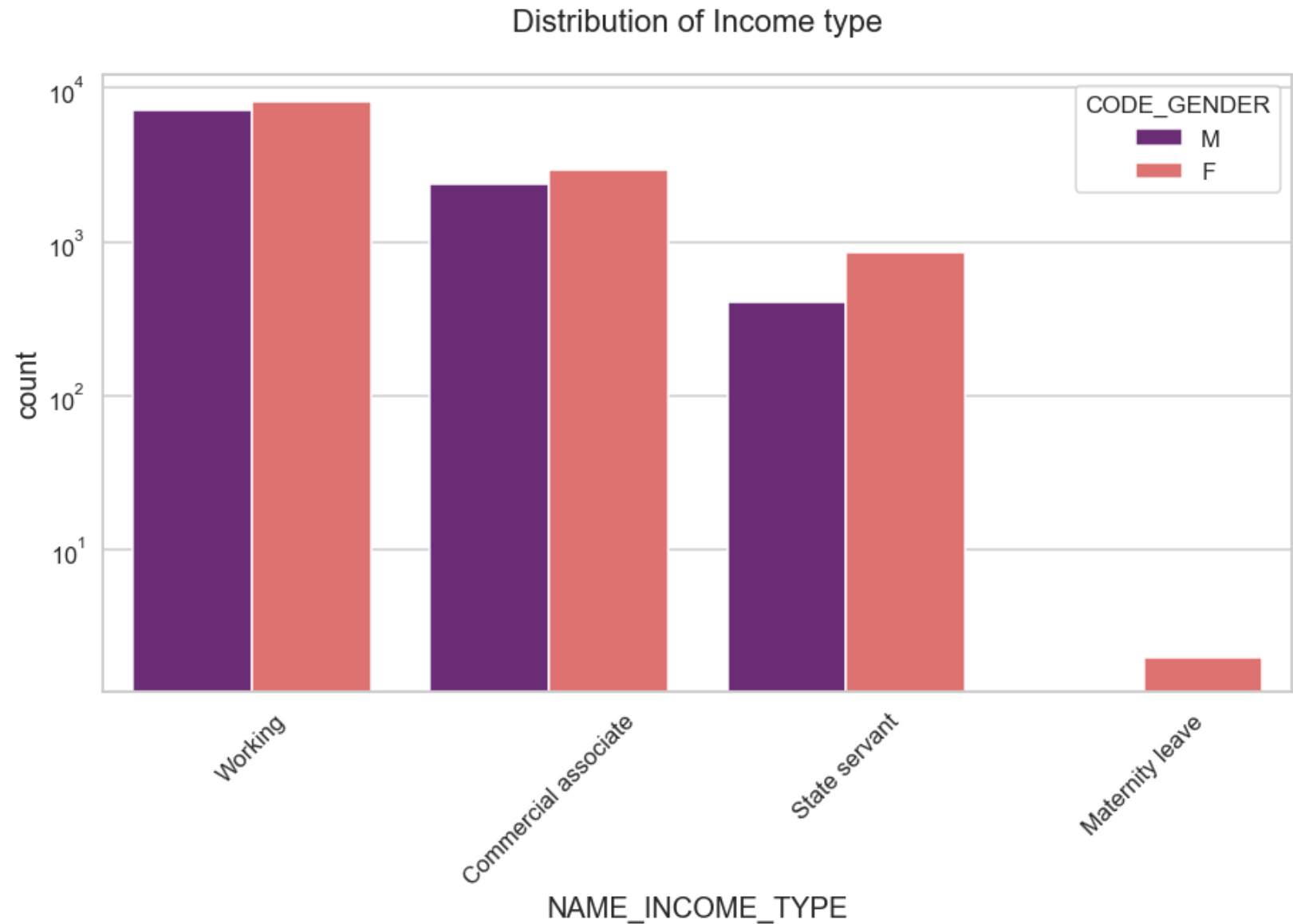
- ▣ Male counts are higher than female.
- ▣ Income range from 100000 to 200000 is having more number of credits.
- ▣ This graph show that males are more than female in having credits for that range.
- ▣ Very less count for income range 400000 and above.



# Distribution of income type

Points to be concluded from the graph on the right side.

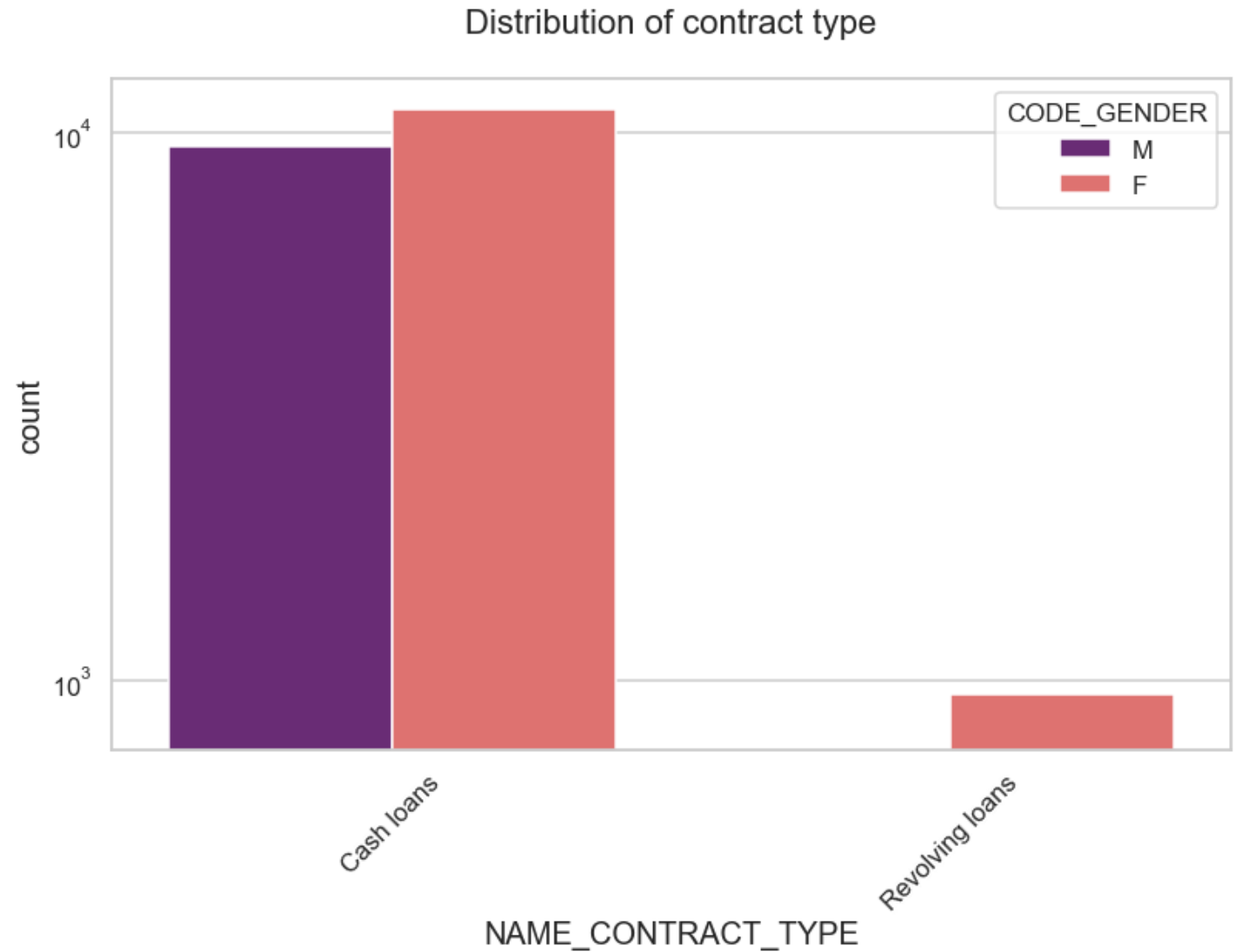
- ▣ For income type 'working', 'commercial associate', and 'State Servant' the number of credits are higher than other i.e. 'Maternity leave'.
- ▣ For this Females are having more number of credits than male.
- ▣ Less number of credits for income type 'Maternity leave'.
- ▣ For type 1: There is no income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



# Distribution for contract type

Points to be concluded from the graph on the right.

- ▣ 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.

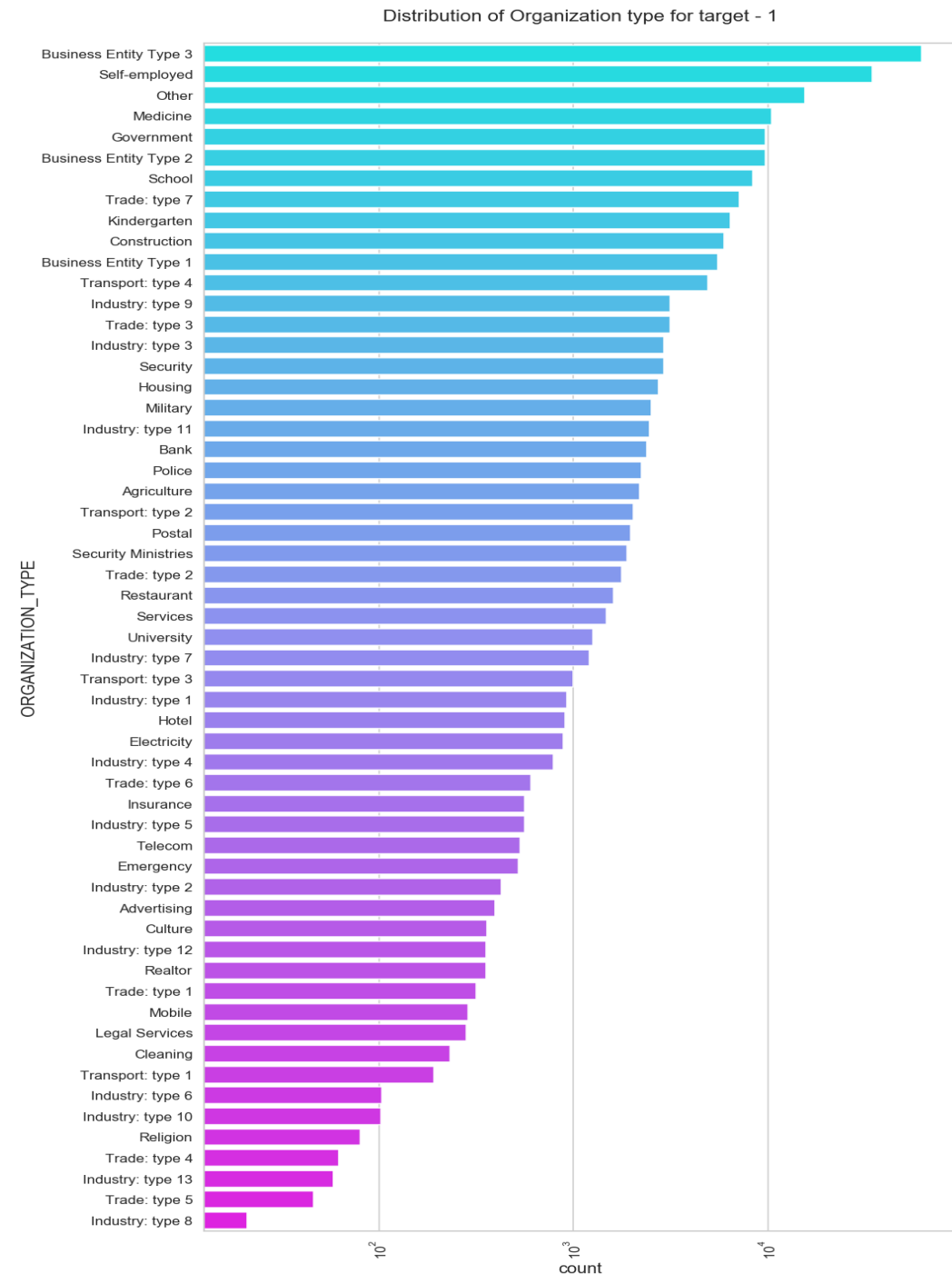




# Distribution of organization type

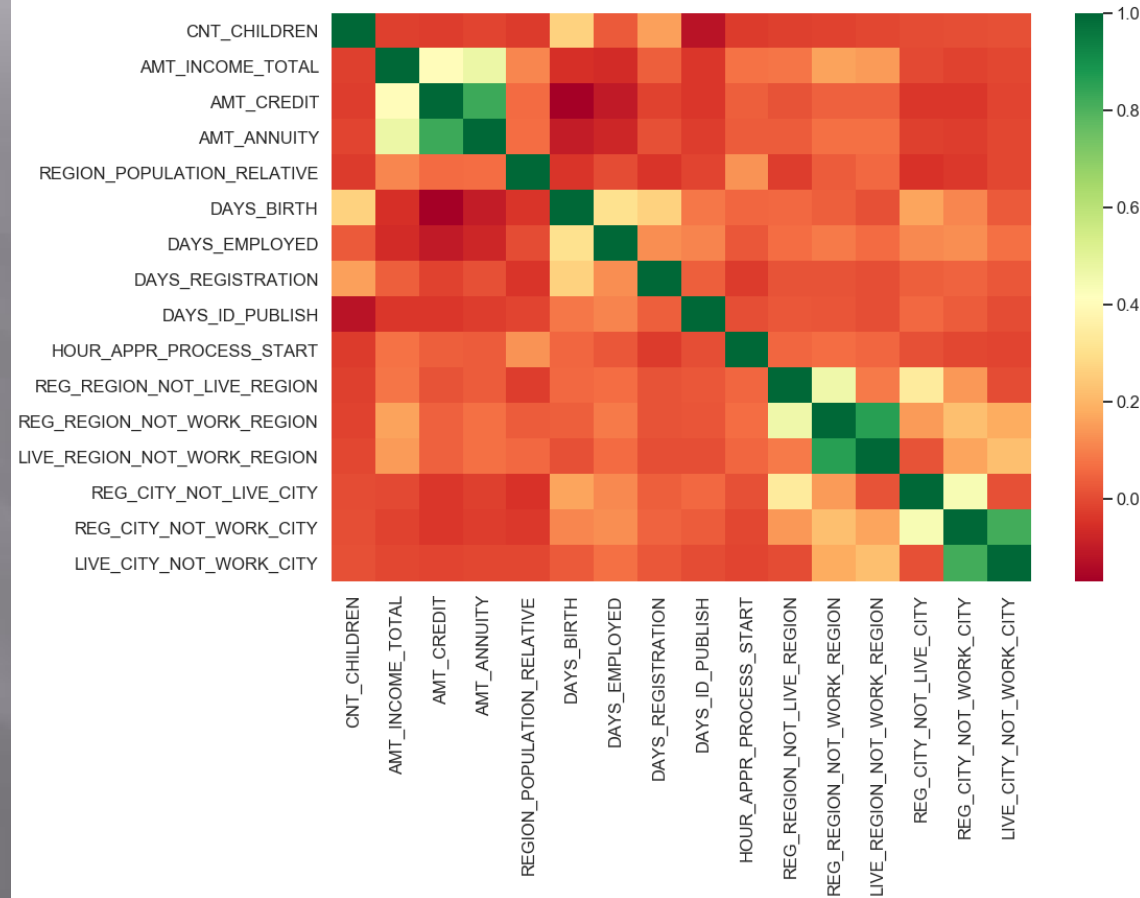
Points to be concluded from the graph on the right.

- ▣ 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



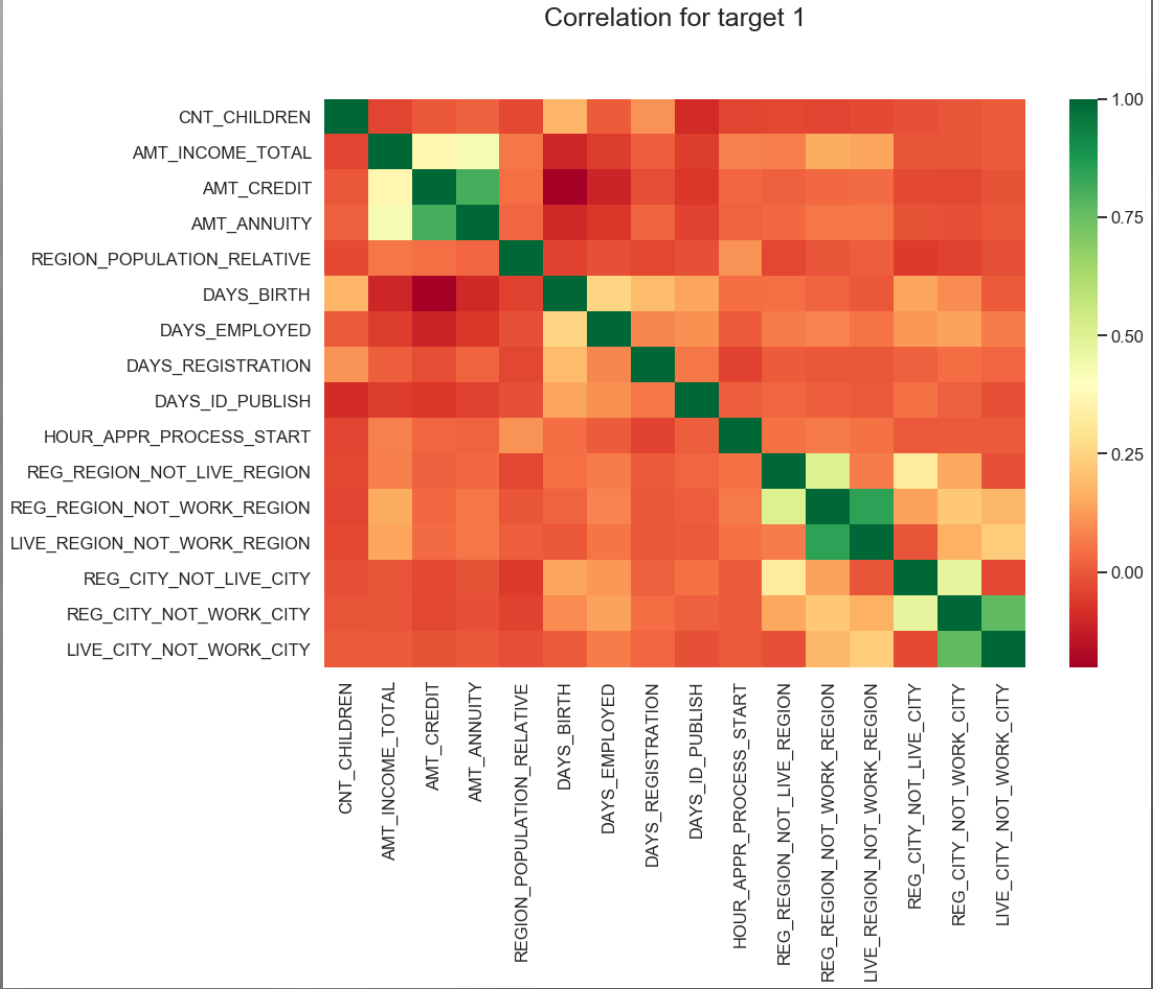
# Correlation For target 0

Points to be concluded from the graph presented before.

- ▣ 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

Heatmap showing the correlation matrix for target 1. The color scale ranges from 0.00 (dark red) to 1.00 (dark green). The diagonal elements are all 1.00. The off-diagonal elements show varying degrees of correlation, with the highest correlations (darkest green) appearing between 'AMT\_CREDIT' and 'AMT\_ANNUITY', and between 'REG\_CITY\_NOT\_WORK\_CITY' and 'LIVE\_CITY\_NOT\_WORK\_CITY'.



# Correlation for type 1

This heat map for Target 1 is also having quite a same observation just like Target 0. But for few points are different. They are listed below.

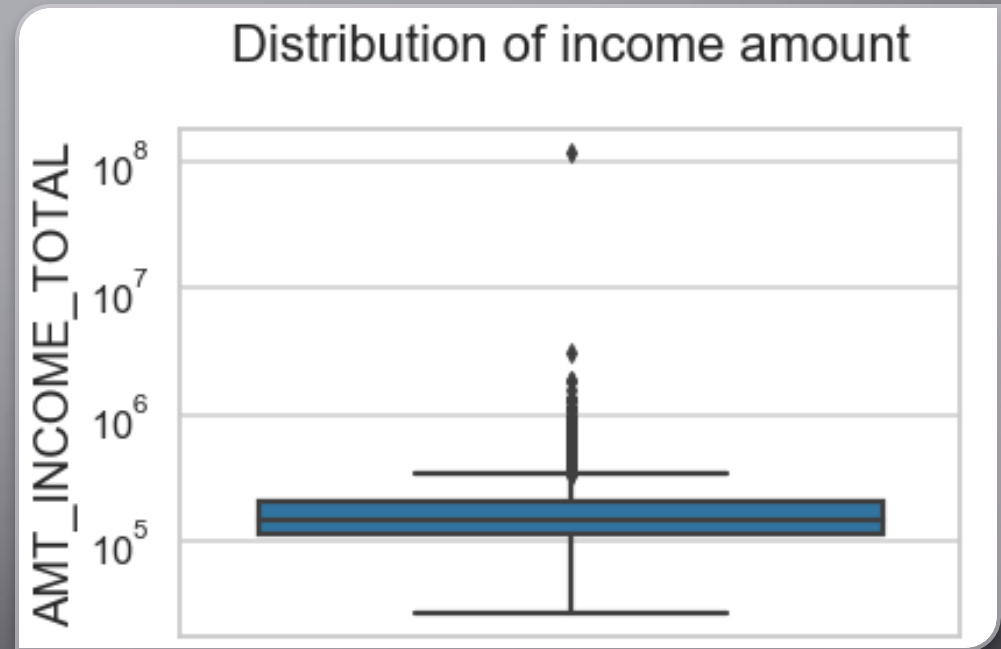
- The client's permanent address does not match contact address are having less children and vice-versa
- The client's permanent address does not match work address are having less children and vice-versa

# Categorical Univariate analysis for variables 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.

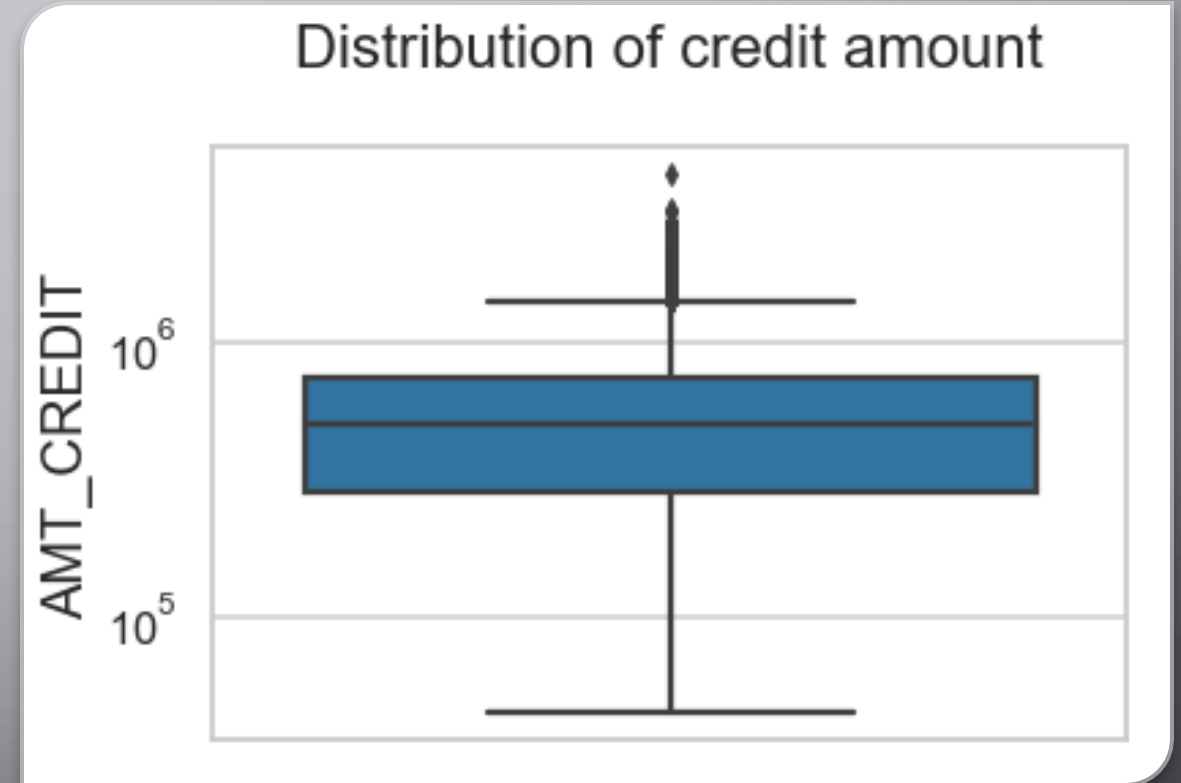




# Boxplot for credit amount

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.

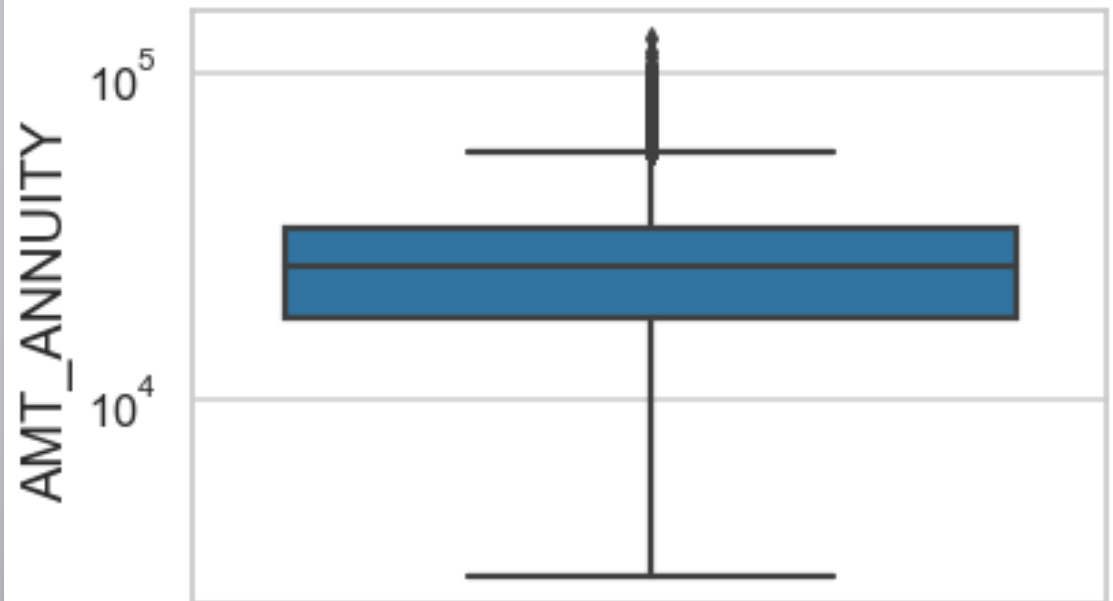


# Boxplot for annuity amount

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.

Distribution of Annuity amount



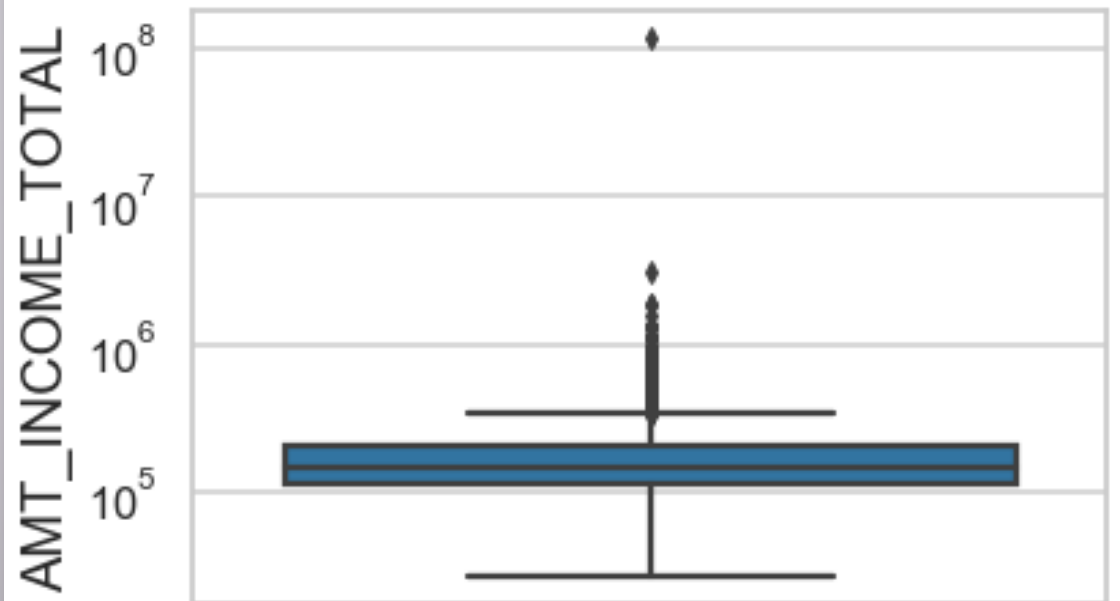
# Categorical Univariate analysis for variables target 1

# 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.
- ▣ Most of the clients of income are present in first quartile.

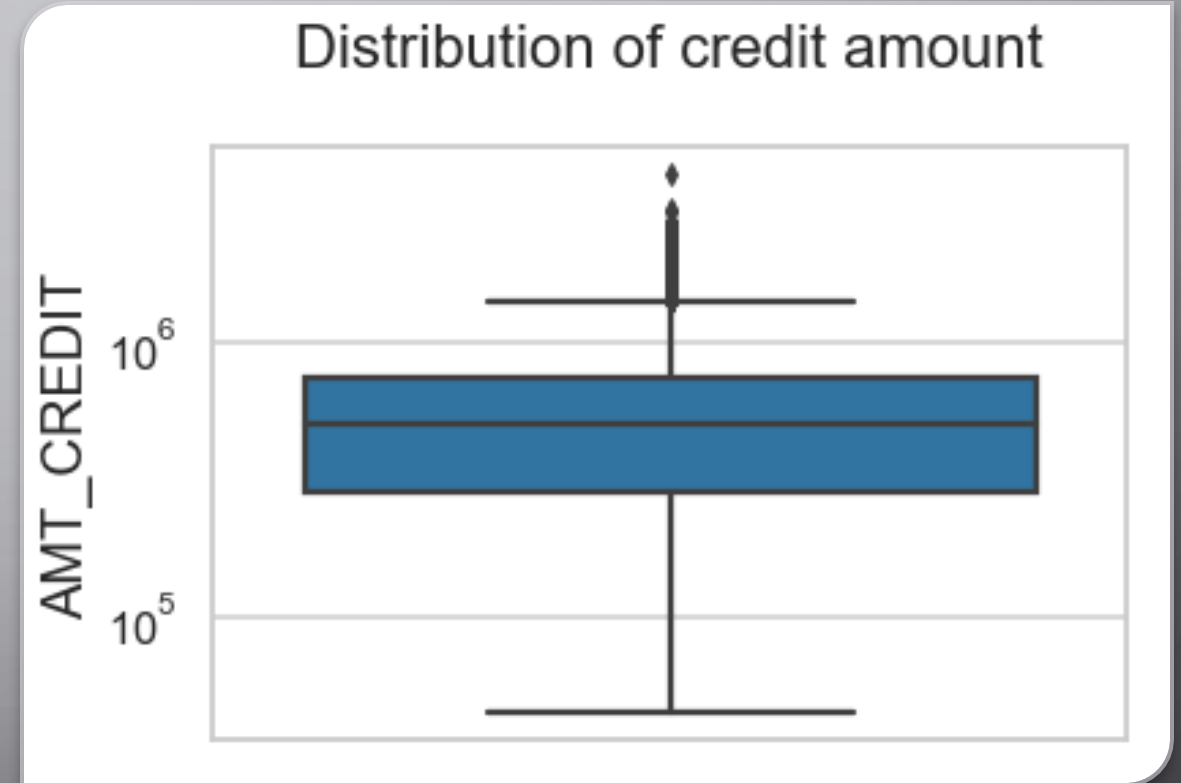
Distribution of income amount



# 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.

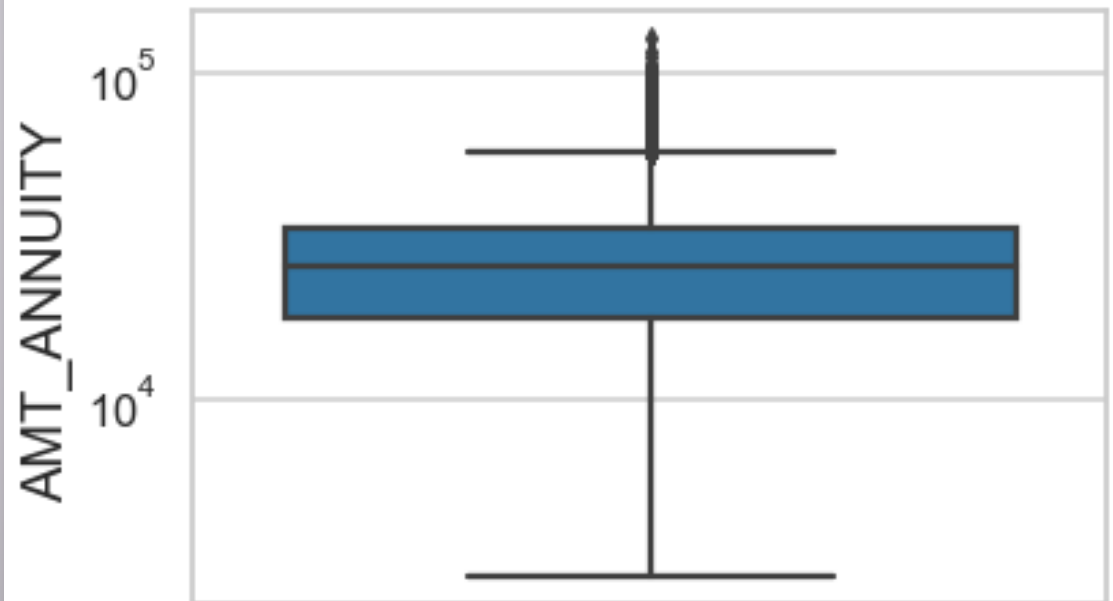


# Boxplot for annuity amount

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.

Distribution of Annuity amount



# Bivariate analysis for type 0

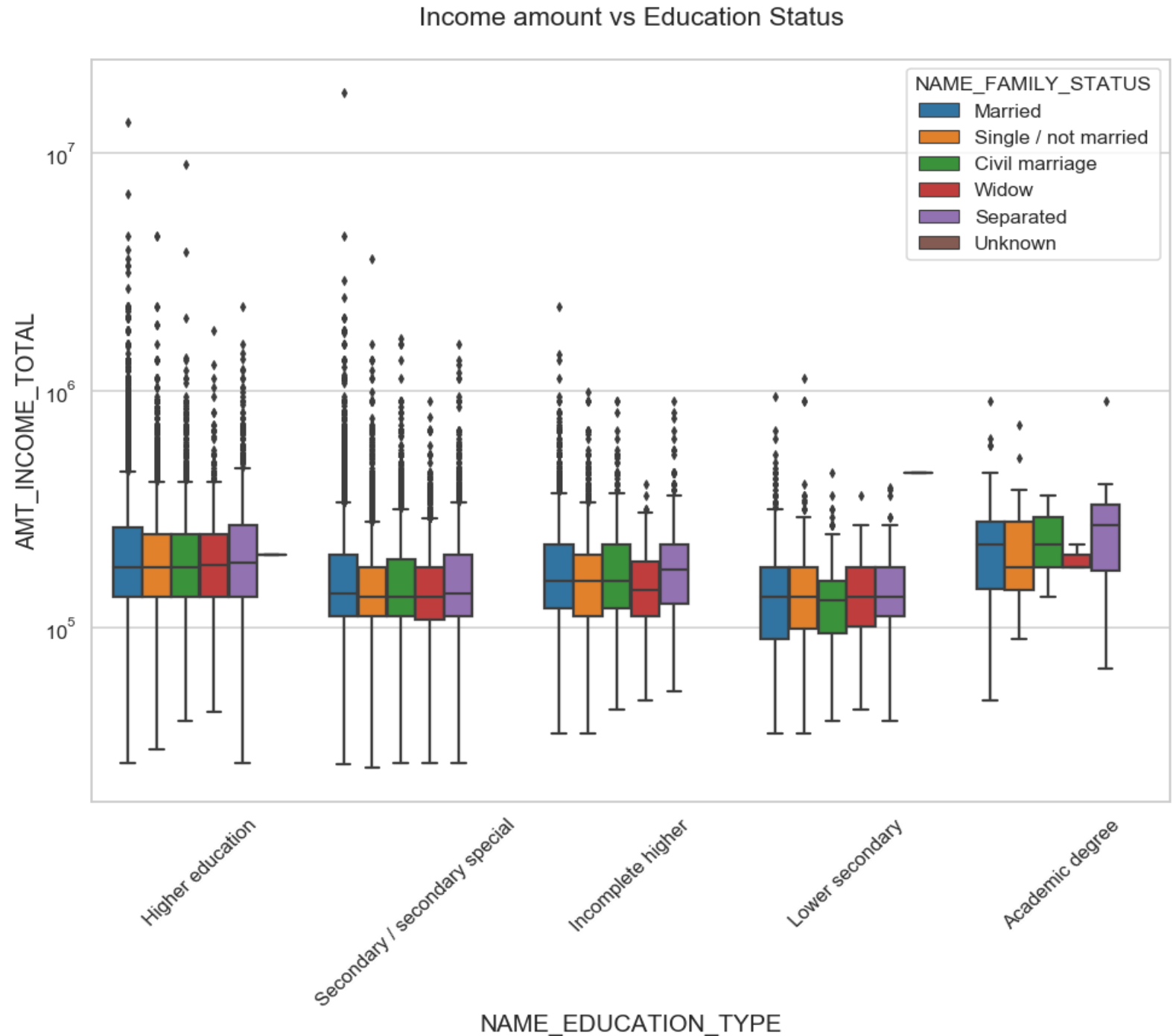




# Income amount vs Education Status

Few points can be concluded from the graph.

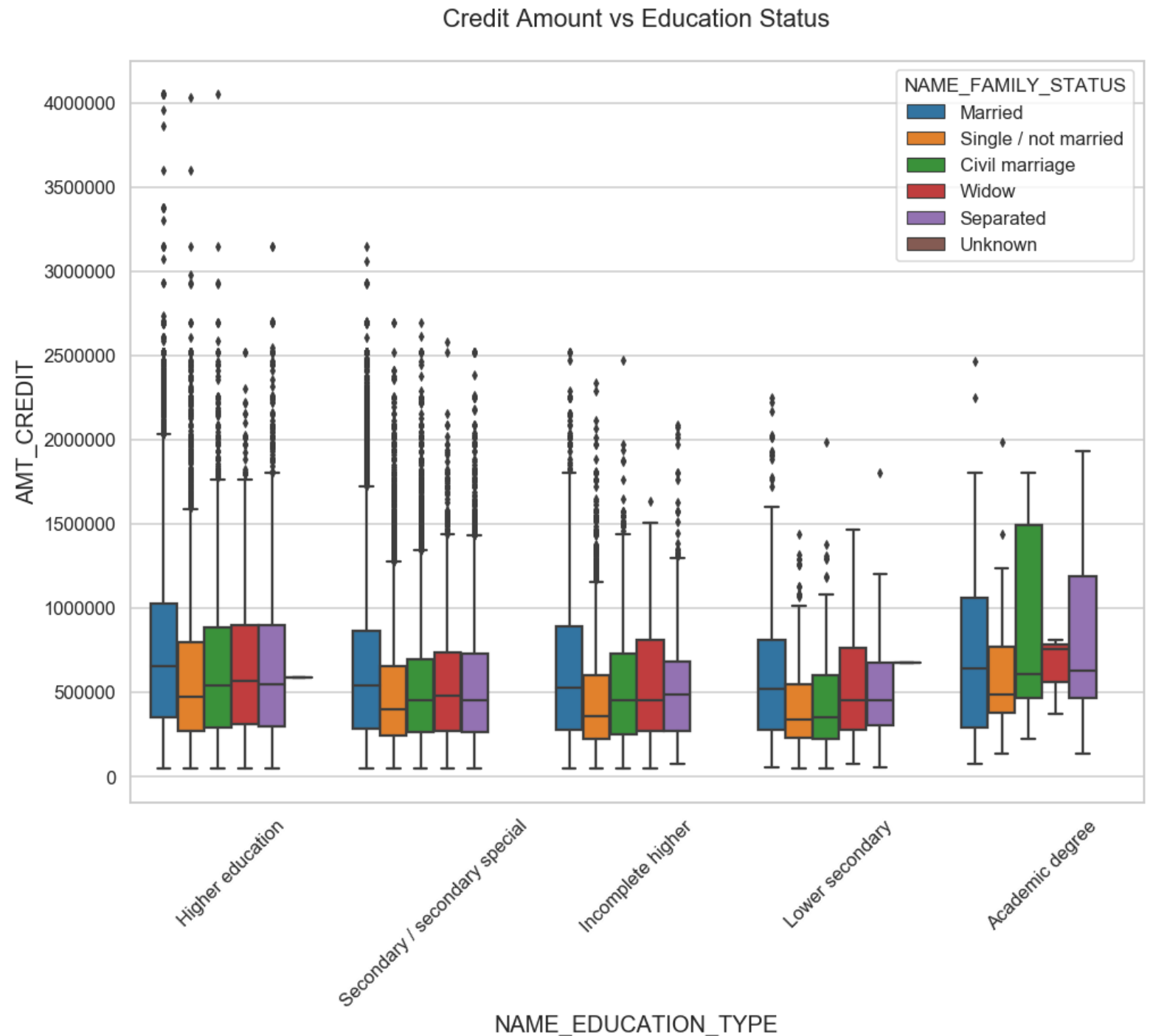
- ▣ For Education type 'Higher education' the income amount mean is mostly equal with family status. It does contain many outliers.
- ▣ Less outlier are having for Academic degree but they are having the income amount is little higher than Higher education.
- ▣ Lower secondary of civil marriage family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



# Credit amount vs Education Status

Few points can be concluded from the graph.

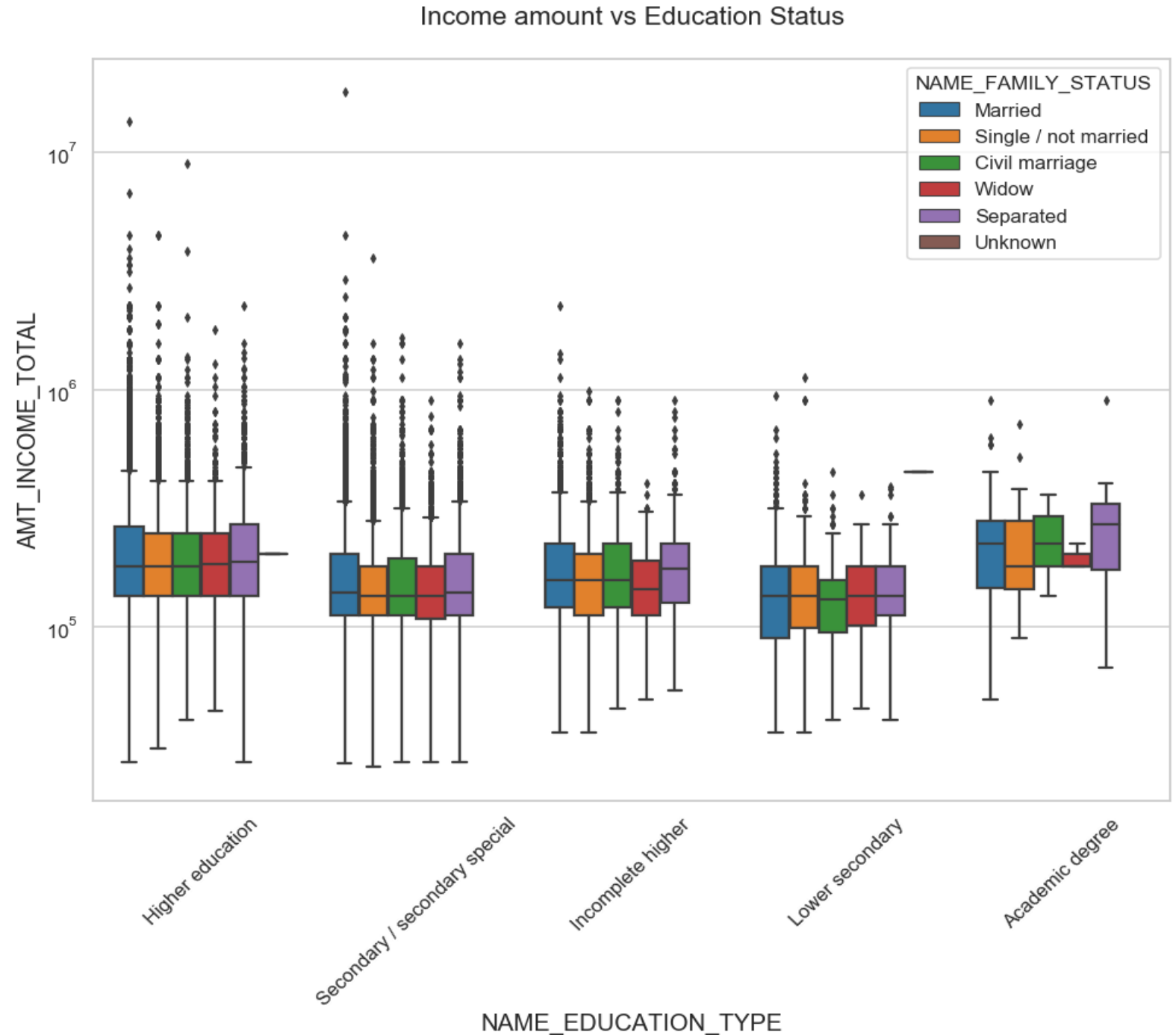
- ▣ Quite similar from Target 0, we can say that Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others.
- ▣ Most of the outliers are from Education type 'Higher education' and 'Secondary'.
- ▣ Civil marriage for Academic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



# Income amount vs Education Status

Few points can be concluded from the graph.

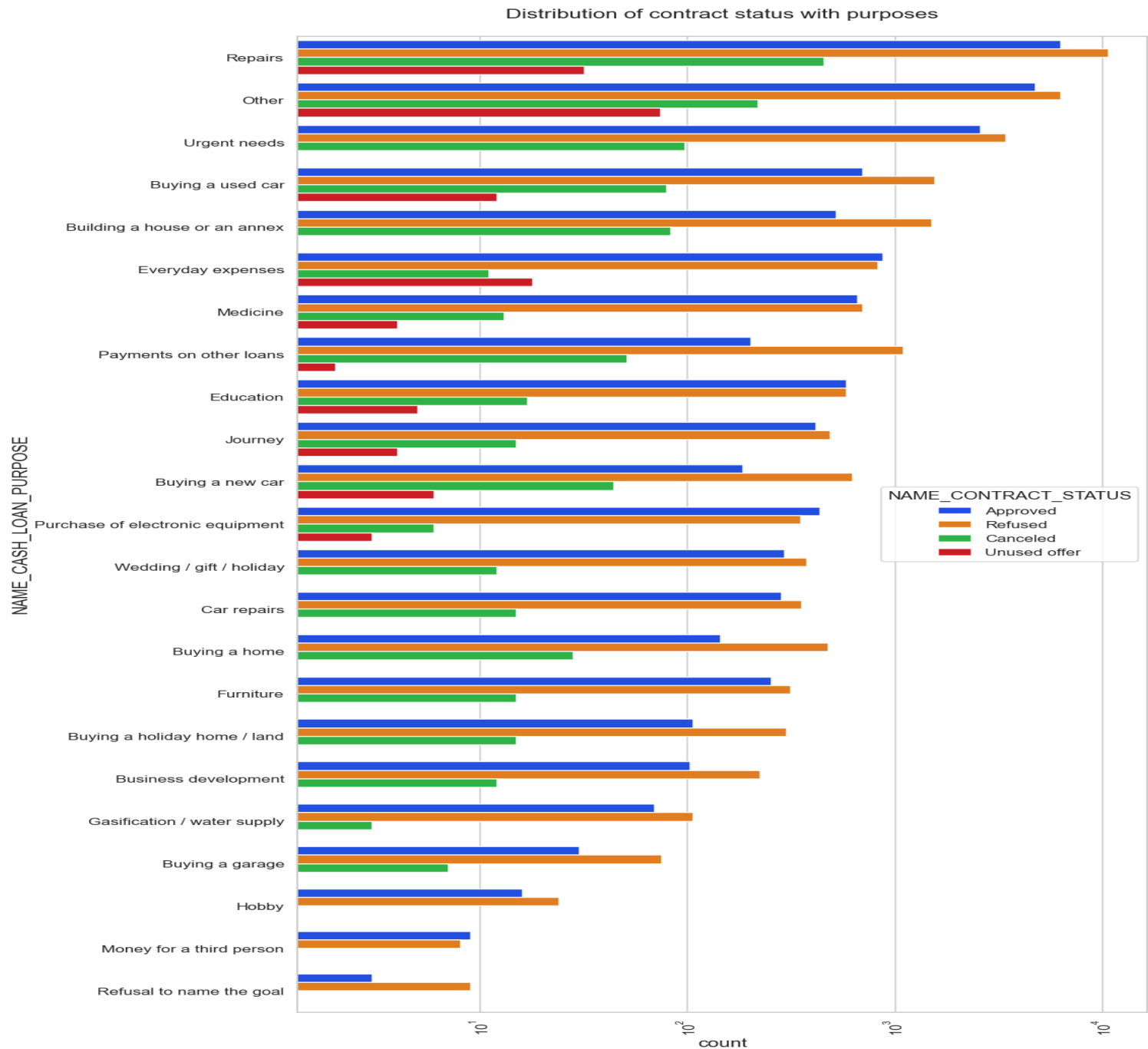
- ▣ Have some similarity with Target0, From above boxplot for Education type 'Higher education' the income amount is mostly equal with family status.
- ▣ Less outlier are having for Academic degree but there income amount is little higher than Higher education.
- ▣ Lower secondary are have less income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



# Distribution of contract status with purposes

Few points can be concluded from the graph.

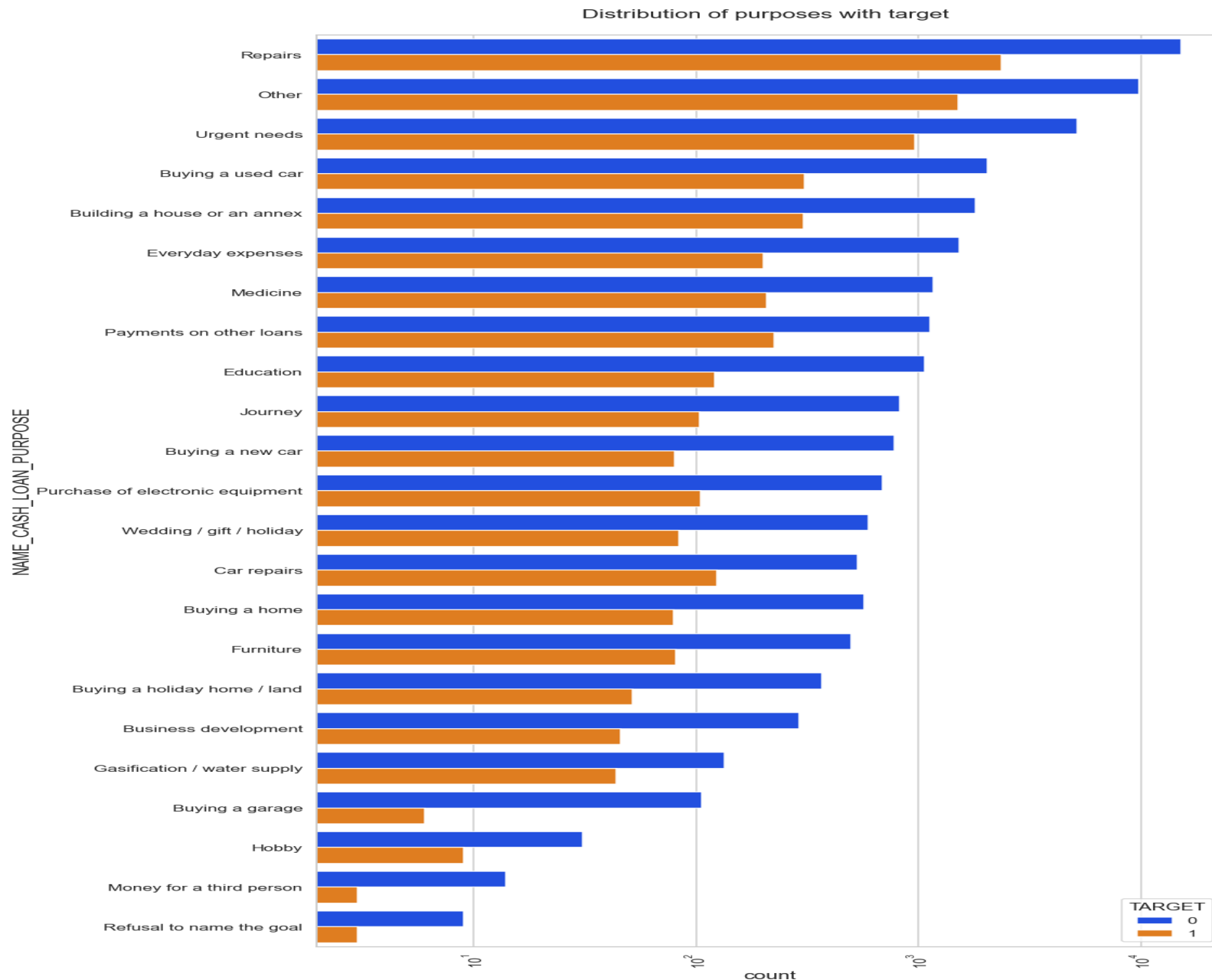
- ▣ Most rejection of loans came from purpose 'repairs'.
- ▣ For education purposes we have equal number of approves and rejection
- ▣ Paying other loans and buying a new car is having significant higher rejection than approves.



# Distribution of purposes with target

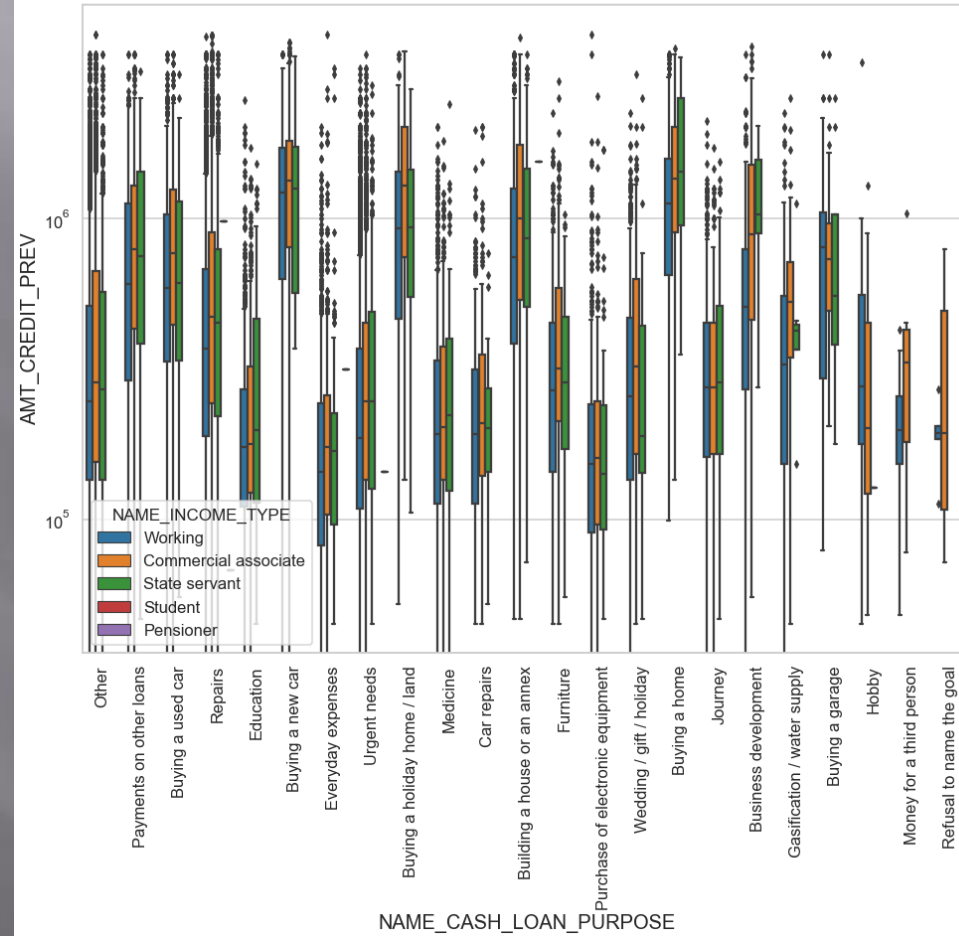
Few points can be concluded from the graph.

- ▣ 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.



# Performing bivariate analysis

Prev Credit amount vs Loan Purpose





# Prev Credit amount vs Loan Purpose

From the previous graph we can conclude the below points:

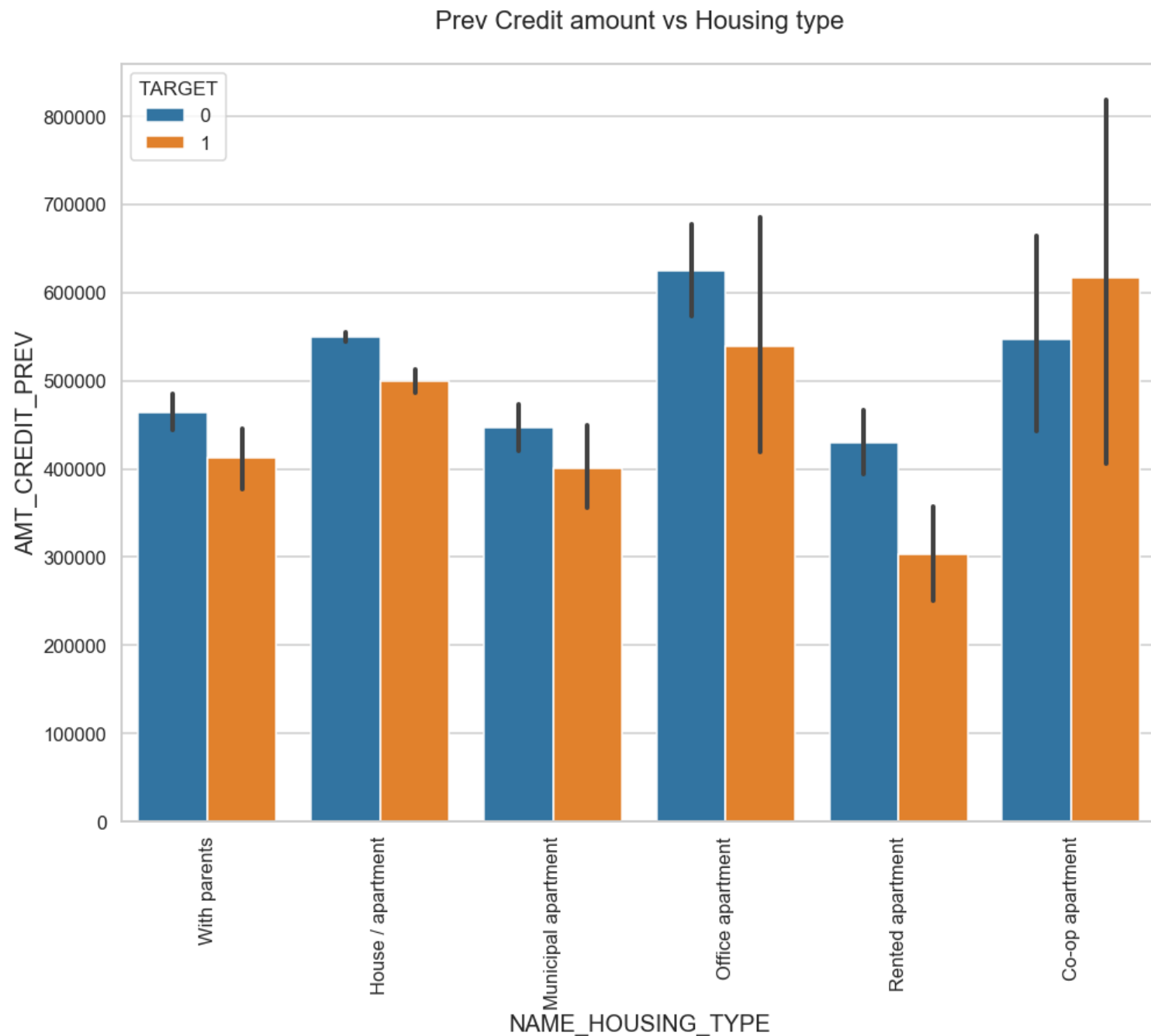
- ▣ The credit amount of Loan purposes like 'Buying a home', 'Buying a land', 'Buying a new car' and 'Building a house' is higher.
- ▣ Income type of state servants have a significant amount of credit applied
- ▣ Money for third person or a Hobby is having less credits applied for.



# Prev Credit amount vs Housing type

Few points can be concluded from the graph.

- ▣ 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 most number of unsuccessful payments.
- ▣ Also with loan purpose 'Repair' is having higher number of unsuccessful payments on time.
- ▣ Get as much as clients from housing type 'With parents' as they are having least number of unsuccessful payments.