Credit EDA Case Study

By

Rohit Birajdar

# Problem Statement And The Approach

- **<u>Problem Statement</u>**

In this EDA Case Study we have to use the given datasets and find the insights from the data that which customer or client segment is best for the bank to focus on. We have to also find that to whom bank should avoid giving loan.

- **<u>Approach</u>**

Our approach to solve this problem is first we have to perform EDA steps like cleaning the data set, plotting the graph and analysing the insights from those graphs. We have to find the indicator for the defaulters and non defaulters and then look for the best segment to which it is safer to give loans

# Categorial Univariate Analysis for Target_1 i.e people having difficulty to pay loan

Distribution Of Income Range
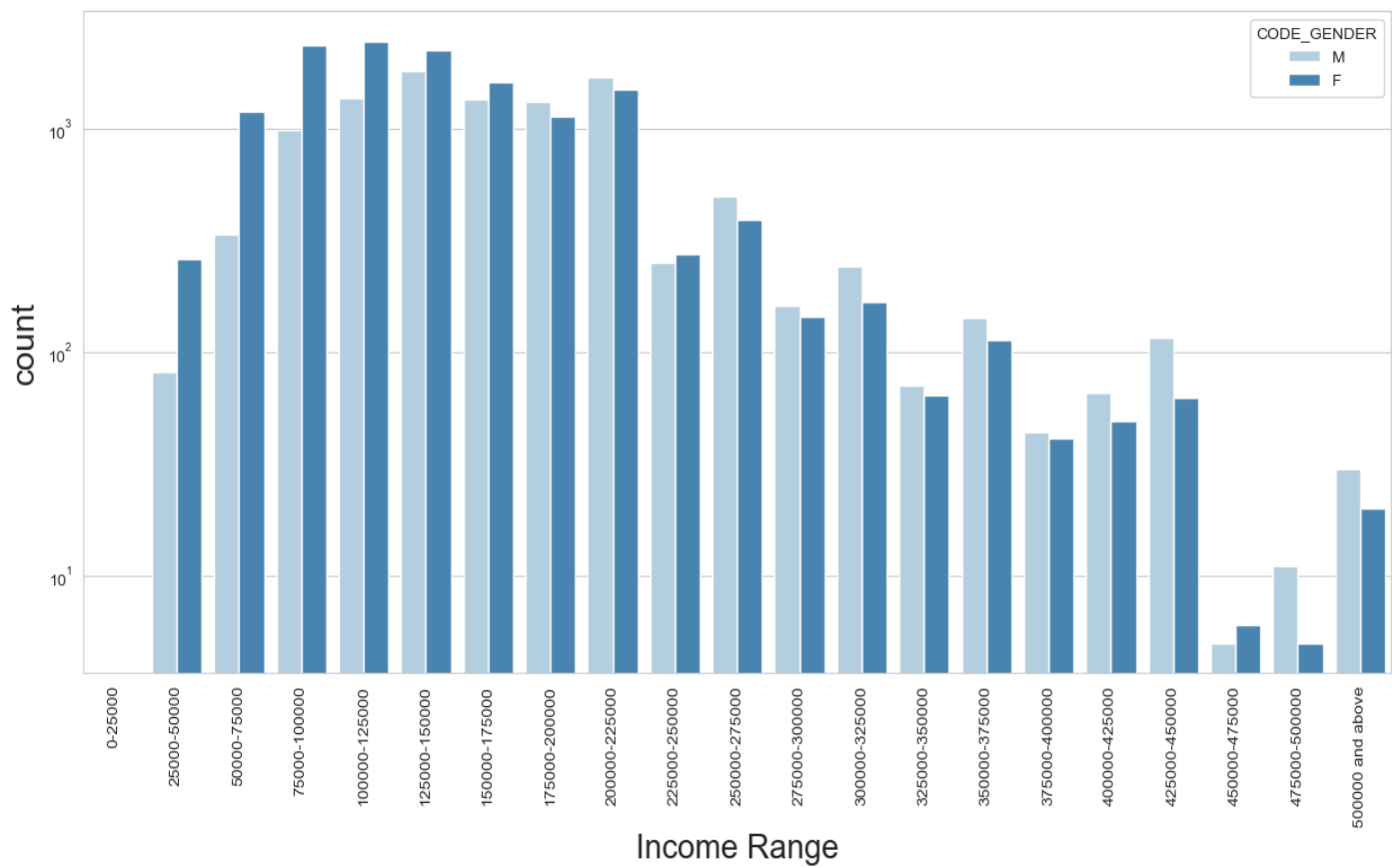
Distribution of income range

- On an average, Male client's income is higher then female's income

- Female in income range 450000-475000 is least likely to default the loan when compared to others

- Male in income range 475000-500000 is least likely to default the loan when compared to others

- In Income ranges 75000-100000, 100000-125000,1250000-150000, 150000-175000. Males are less likely to default on loan, when compared to other ranges these are the ranges where difference is significant

Distribution Of Income Type
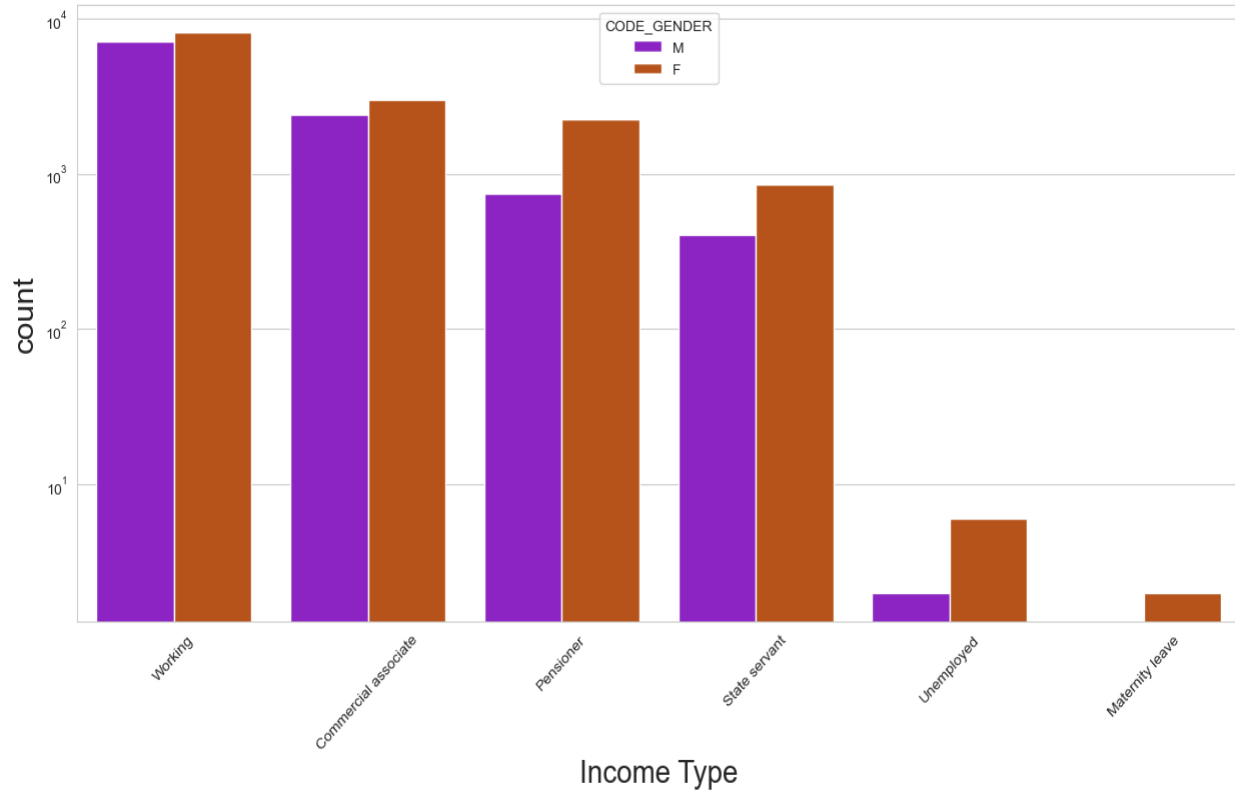
Distribution of income type

- Income of type 'working', 'commercial associate', 'pensioners' and 'State Servant' the number of defaults are higher than other i.e. 'Maternity leave & Unemployed.

- Less number of defaults of income type 'Maternity leave'

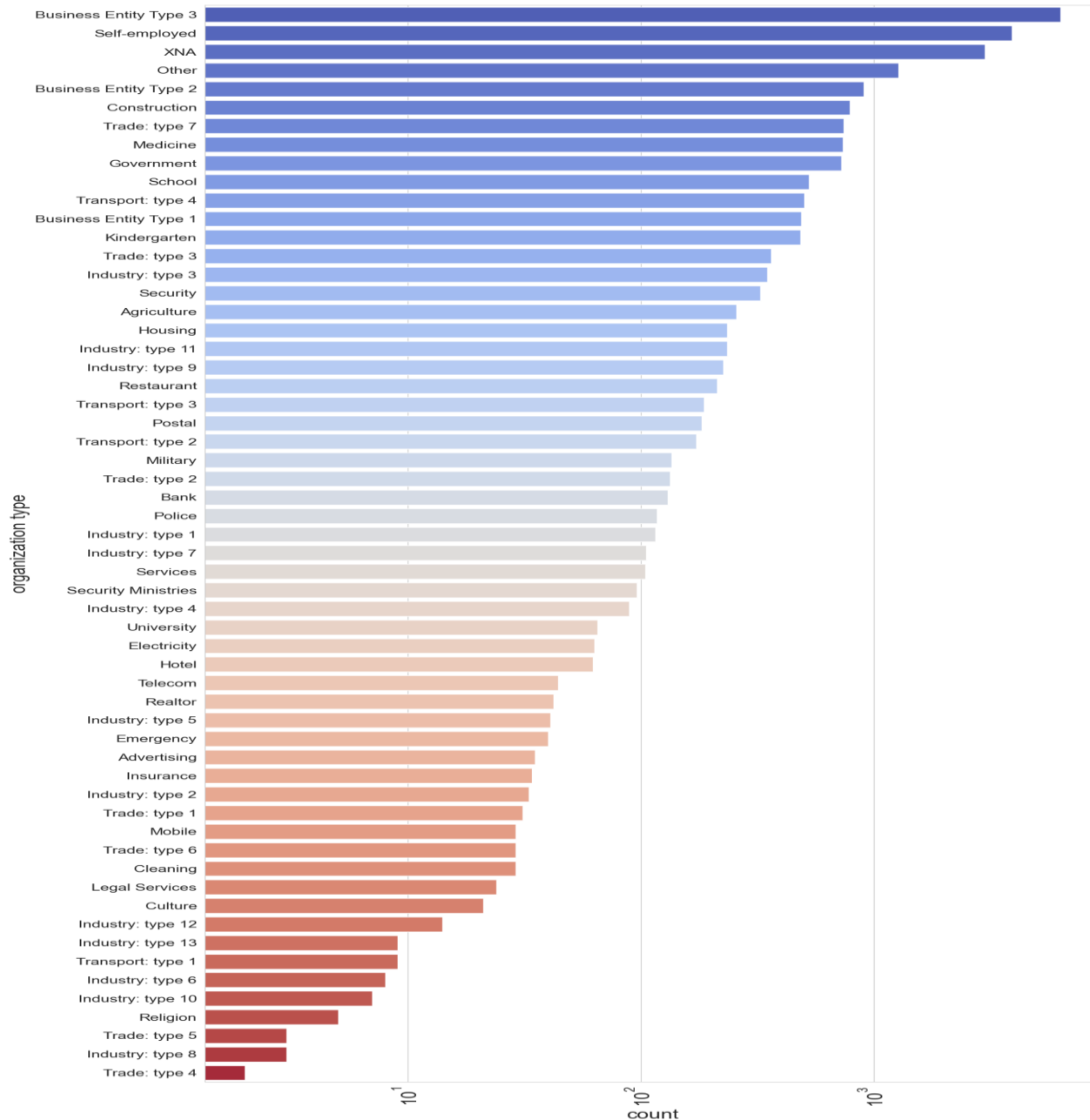- There is no income type for students it means students are not defaulter

Distribution Of Contract Type
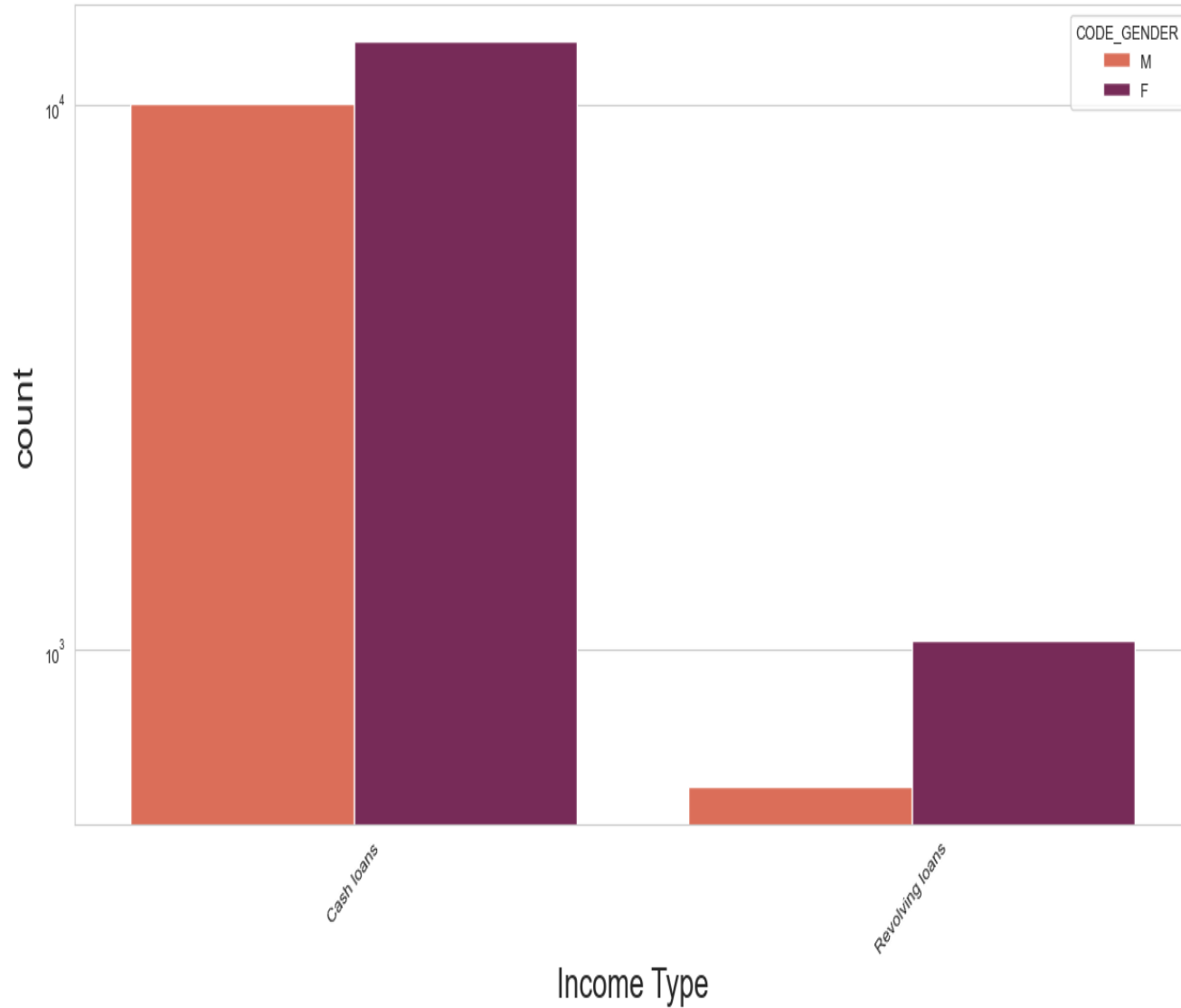
Conclusion

- It seems that "cash loan" is having higher number of credit than "Revolving loan"

- Female applies for more number of credit than male i.e means female are more likely to default the loans

Distribution Of various organization type

Distribution Of Various Organization Type

- Clients of the organization type 'Business entity Type 3,self employed, other, Medicine and government are less likely to pay the loan on time.

- Clients from industry Type 8, Type6,Type 10 ,Religion and Trade Type 5 are more like to default

Correlation for Target_0

correlation for Target=0
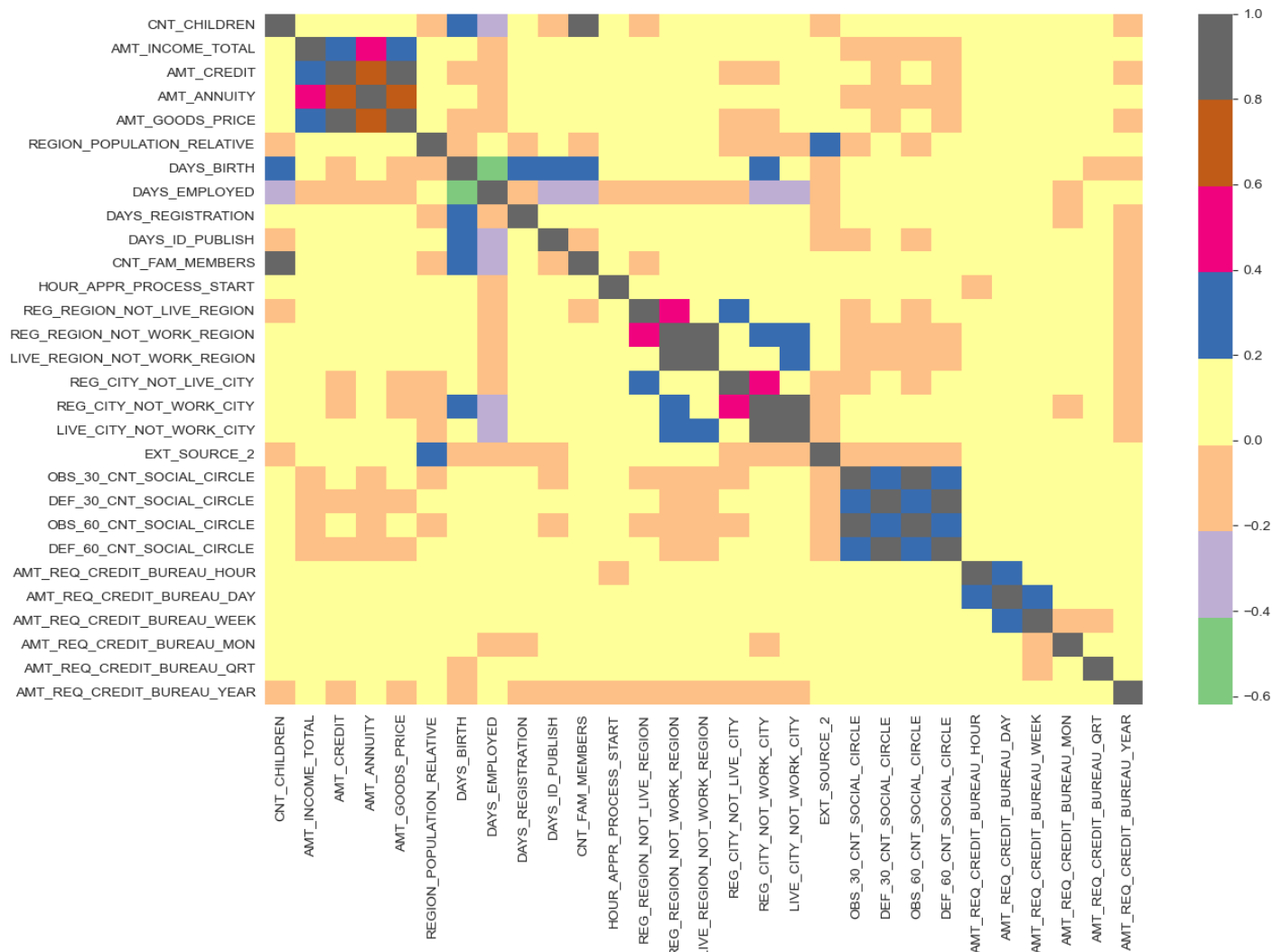
Conclusion

- Credit amount is inversely proportional to the date of birth, which means Credit amount is higher for low age and lower for high age.

- Credit amount is inversely proportional to the number of children client have, means Credit amount is less for higher children count client have and vice-versa.

- lesser children client have in densely populated area. • Credit amount is higher in densely populated area.

- The income is also higher in densely populated area

Corelation for Target_1


corelation for Target_1
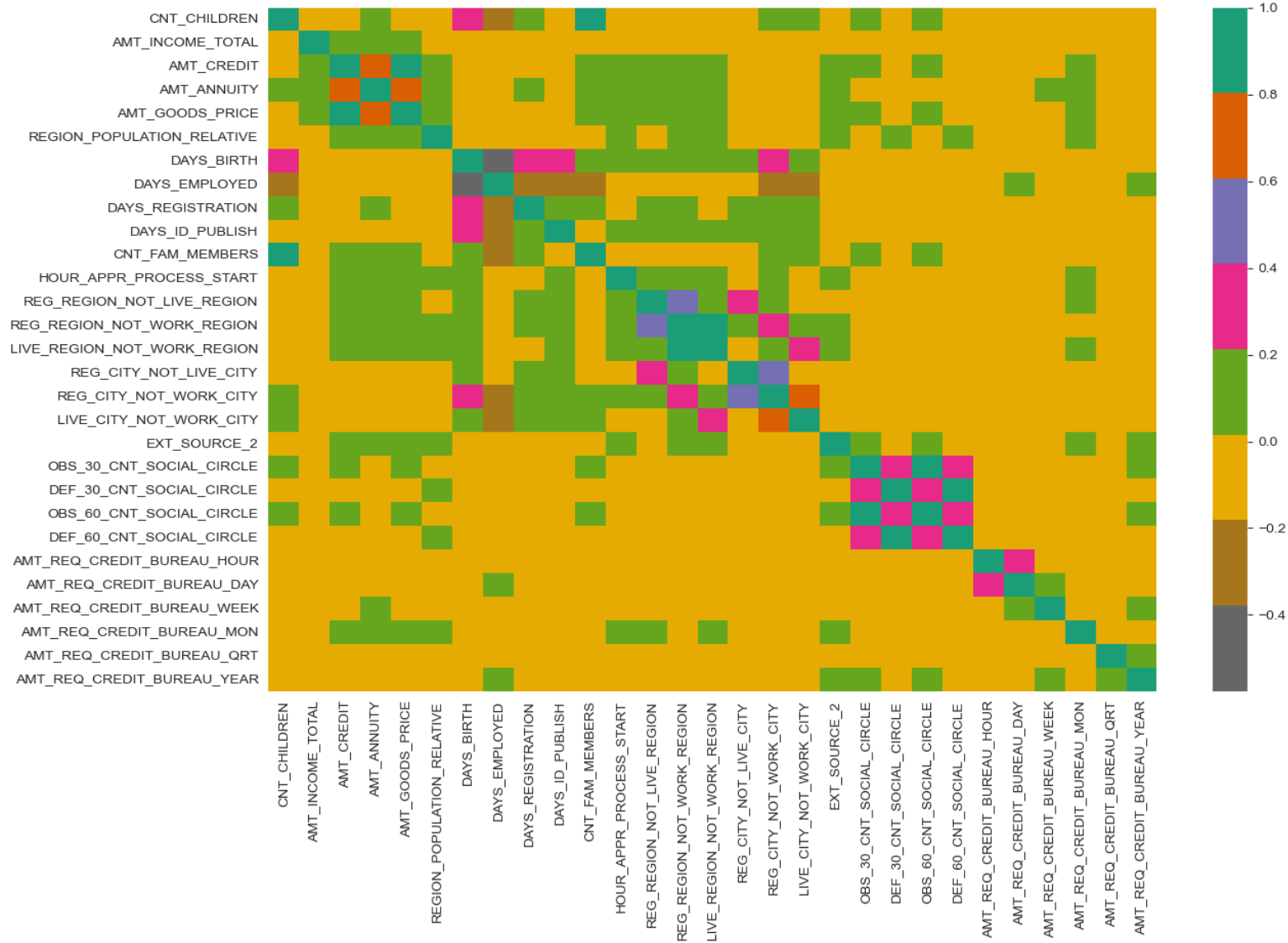
Conclusion

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
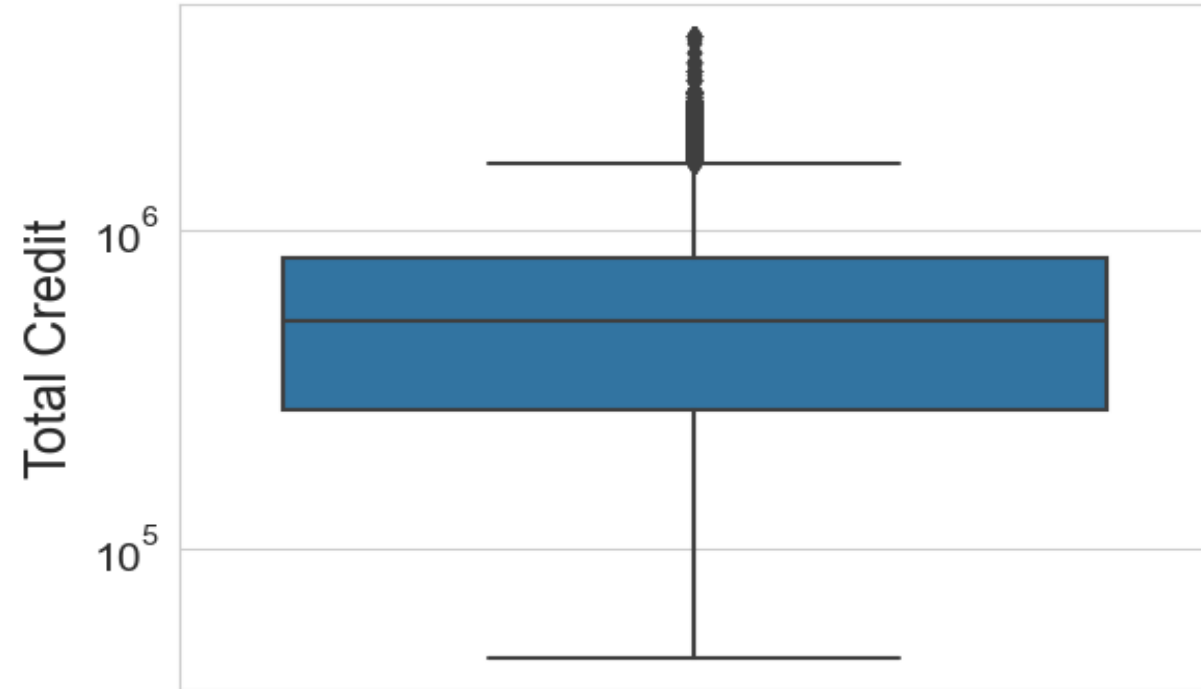
Distribution of Income Amount
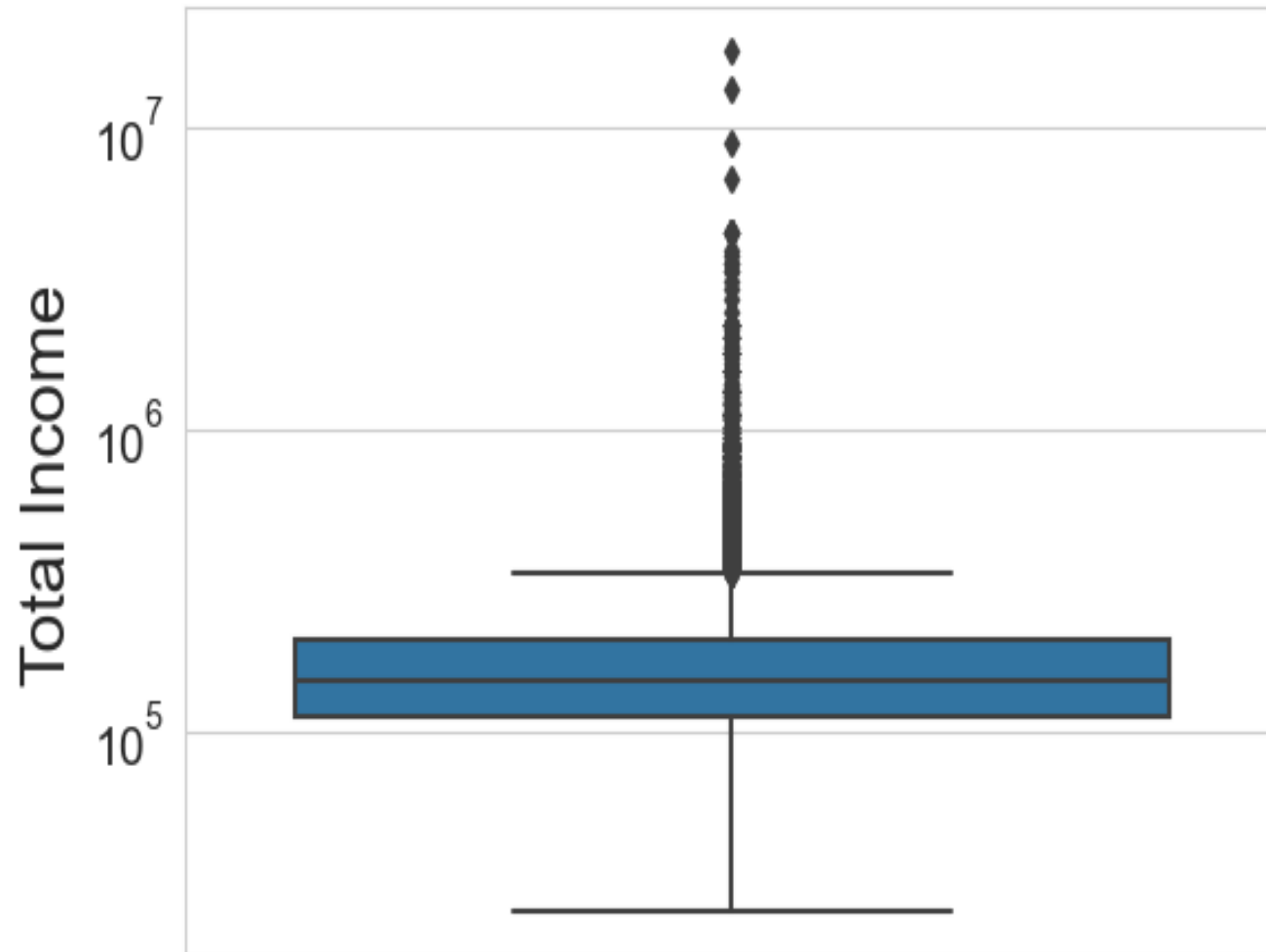
Conclusion

- There seems to be equal distribution of income amount of the clients

- Some irregularities are noticed in income amount.

- The third quartiles is very thin for income amount.
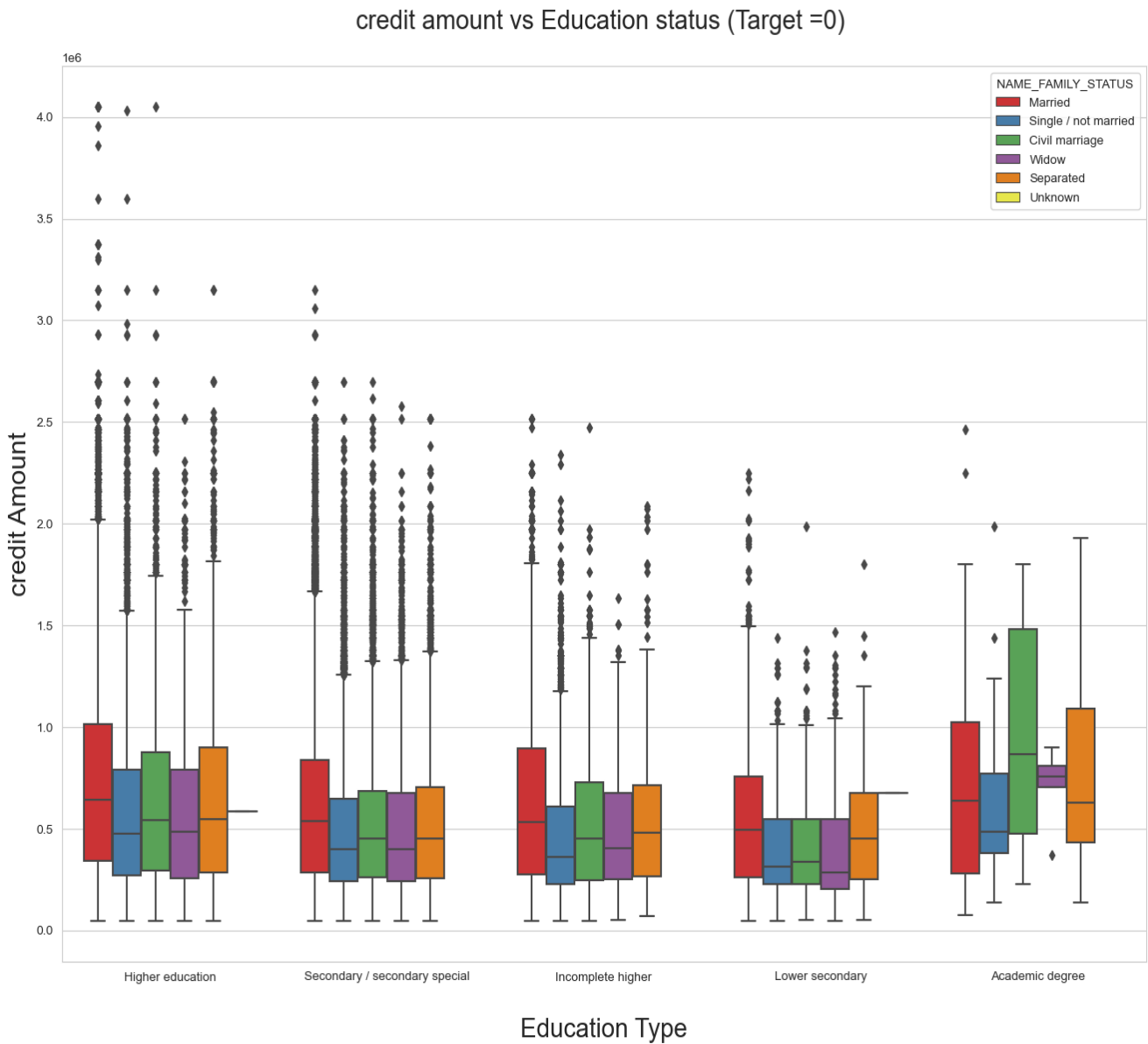
Box Plot For Credit Amount

Distribution Of Credit Amount

Conclusion

- No 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.
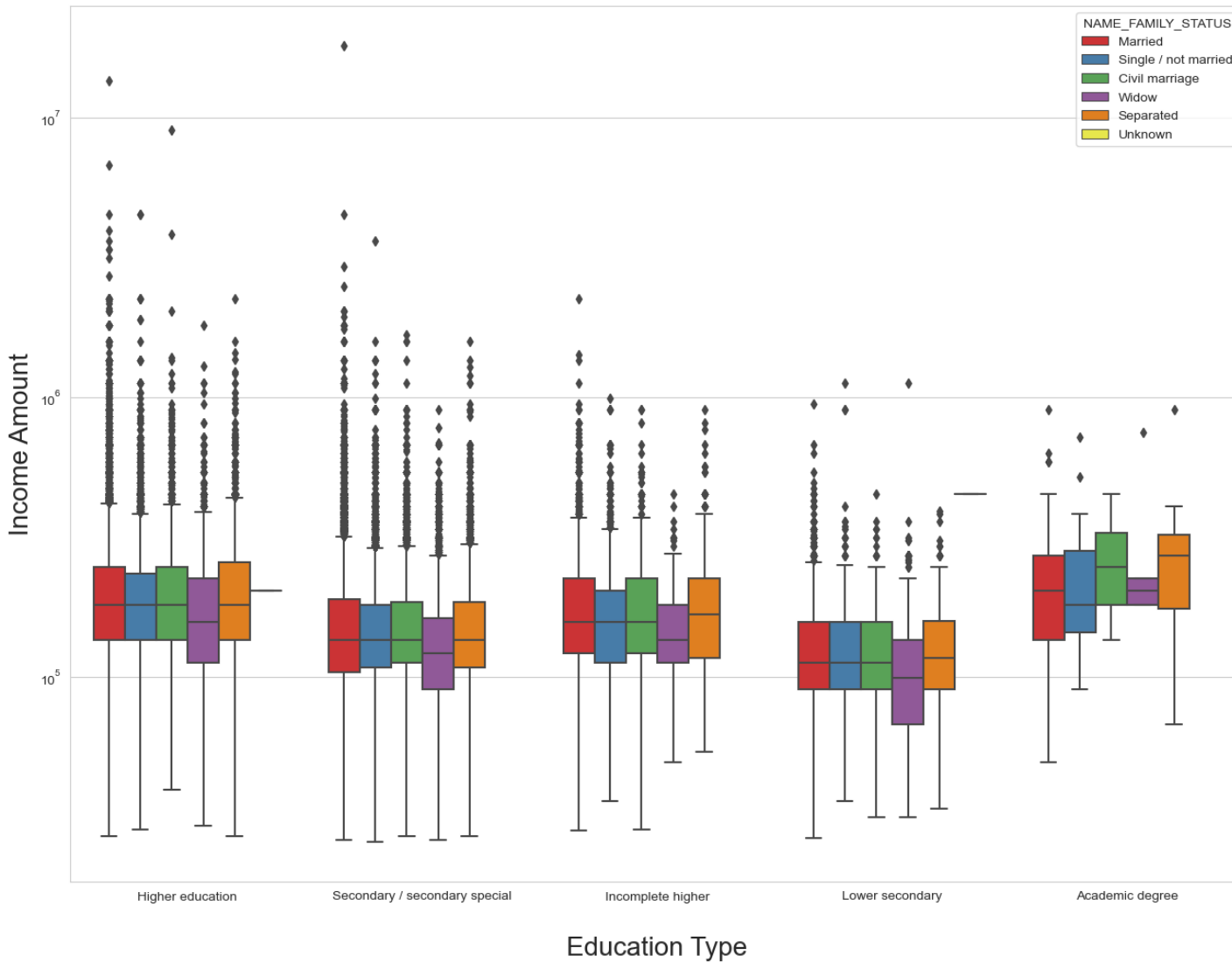
Multivariate Analysis for Target_0

credit amount vs Education status (Target =0)

Credit amount vs Education Status

- Family status of 'civil marriage','marriage' and separated' of Academic degree education are having higher number of credits than others.

- Higher education of family status of 'marriage', 'single' and 'civil marriage' are having more outliers.

- Civil marriage for Academic degree is having highest of the credits in the third quartile.
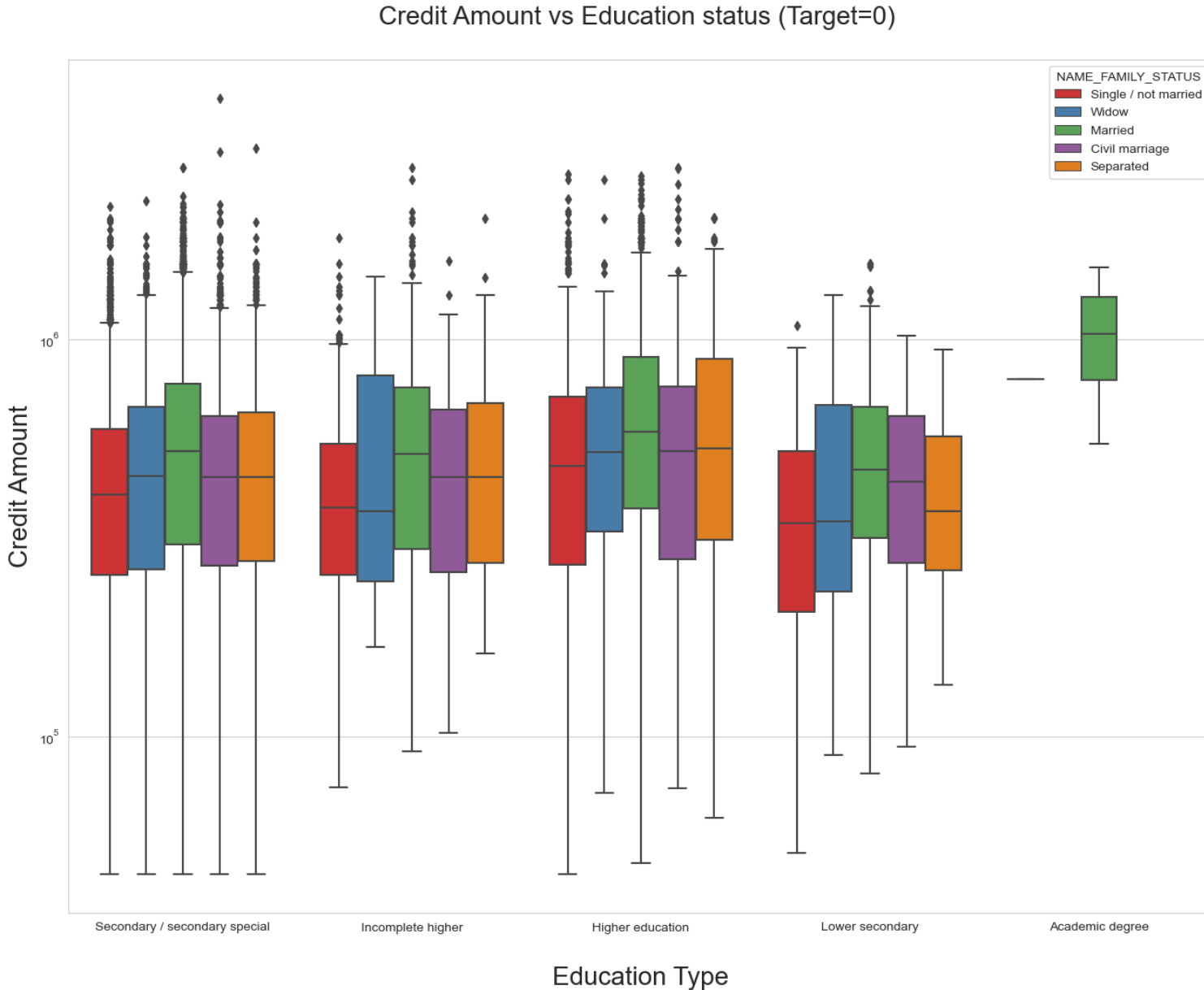
Income Amount vs Education status (Target=0)

Income amount vs Education Status

- For Education '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 their income amount is l higher than that of 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 (Target=0)

Credit amount vs Education Status

- 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 'Higher education' and 'Secondary'.

- married for Academic degree is having most of the credits in the third quartile

Income Amount vs Education Status

Income Amount vs Education status (Target=1)
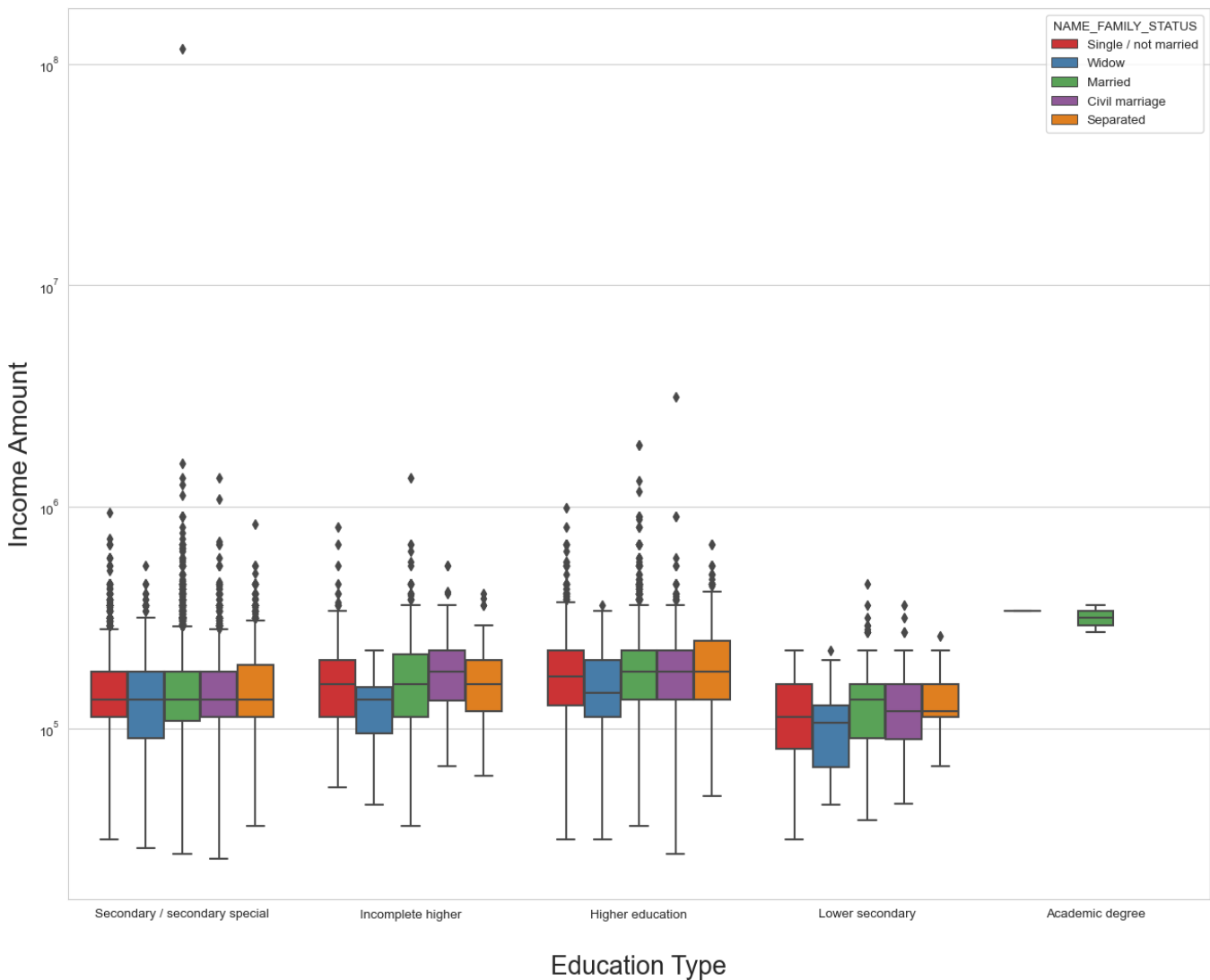
Conclusion

- some similarity with Target, From the boxplot for Education type 'Higher education' the income amount is mostly equal with family status.

- Less outlier are having for Academic degree but their income amount is little higher than Higher education.

- Lower secondary have less income amount than others

Univariate analysis after merging previous data

Distribution of contract status with purposse
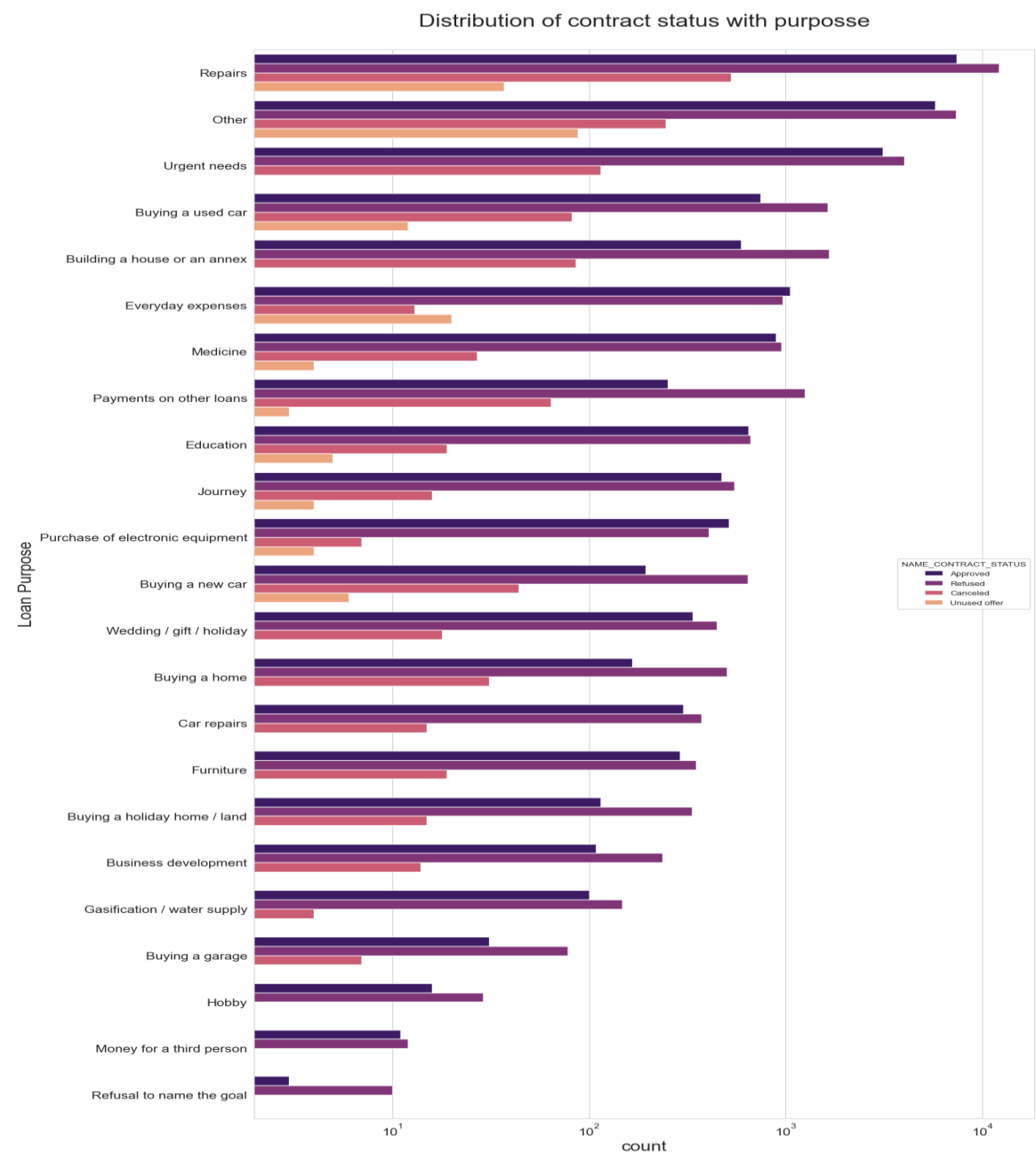
Conclusion
1) Most loan rejection came from 'Repair' purpose.
2) There are equal approves and rejection for ' education' purpose.
3) Paying other loans and buying new car have higher number of rejection

Distribution of contract status with Target's
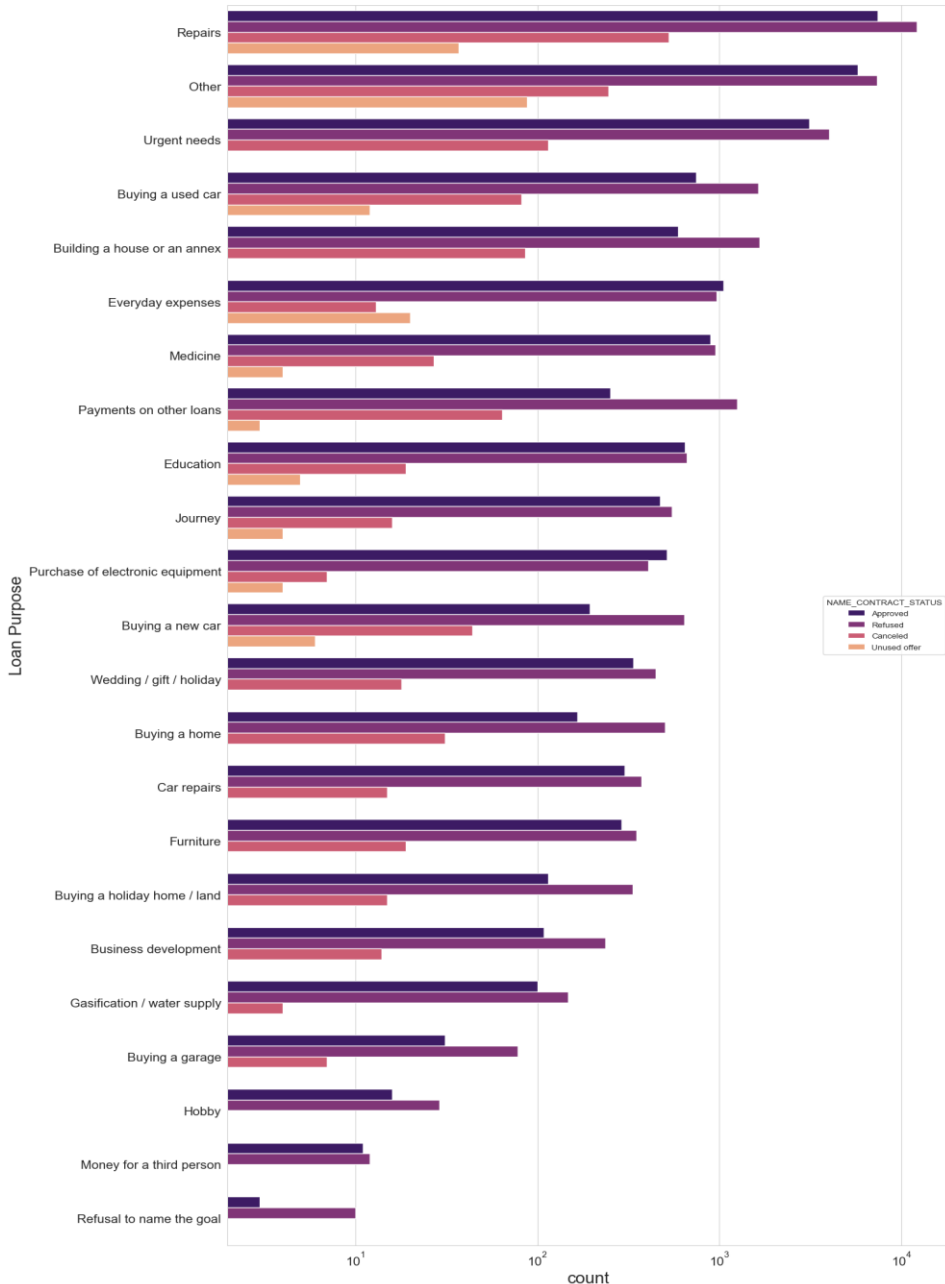
Distribution of contract status with Target's

- loan purpose with 'Repairs' is facing more difficulties in payment on time

- There are some places where loan payment is facing higher than significant difficulties these are 'Buying a garage', 'Business Development ' ,'Buying land ', 'Buying new car ',and 'Education' hence we can focus on these purpose where clients are facing difficulties in repayment
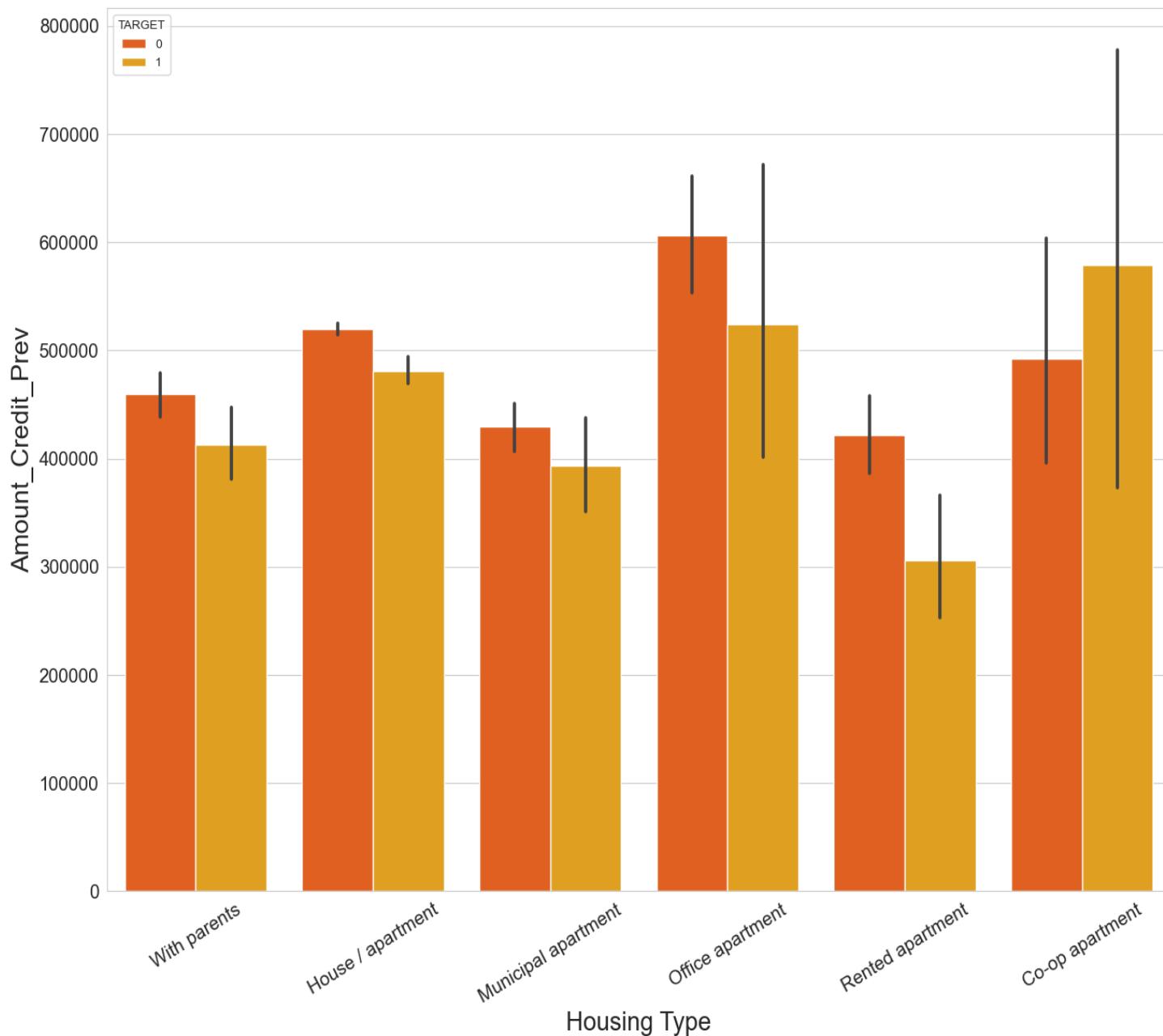
Distribution of contract status with Target's
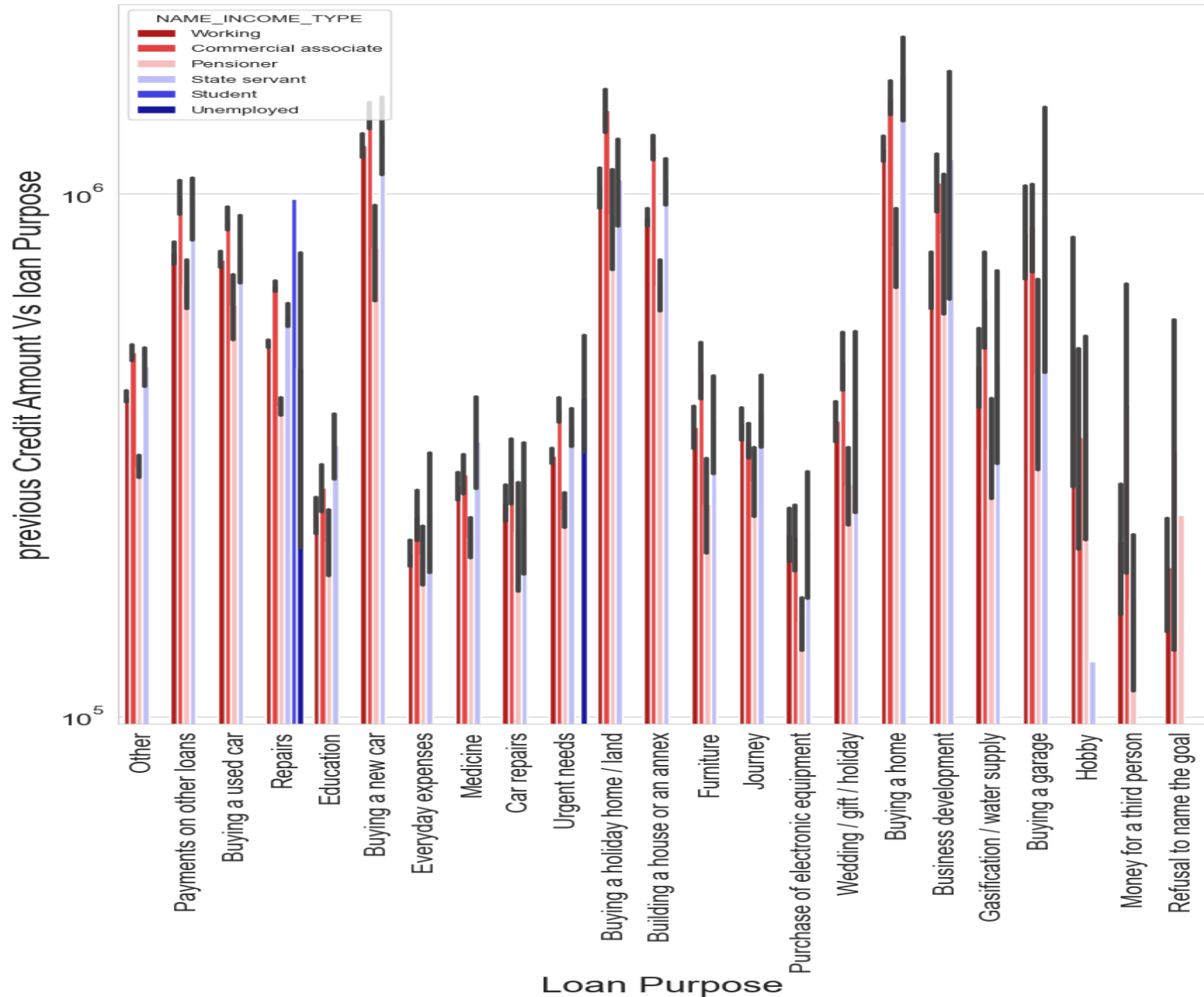
Conclusion
- Credit Amount of Loan purpose like Buying land', 'Buying new car, and 'Building a house' is higher

- Income type of state servant have more credit applied

- Hobby and money for third person have less credit amount

Prev Credit amount Vs Housing Type

Prev Credit amount vs Housing type

- 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 the housing type of co-op apartment as they are having failure in payment.

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

Conclusion Of EDA Loan Case Study

1. Bank should approve more loans to the office apartments, co-operative apartment housing type they have less payment difficulties.

4. Female in income range 450000-475000 is least likely to default the loan

5. Male in income range 475000-500000 is least likely to default the loan

6. In Revolving loans category. male is not very likely to default. Bank can attract with specific offers

7. Clients from Industry type 8,type 6, type 10, religion and trade type 5, type 4 are more likely to default.

8. People with Income type 'Working' have most number of unsuccessful payments.

9. Loan purpose type 'Repair' is having higher number of late payments

Thank you