CREDIT EDA Case Study

By
Jitu Moni Das

Business Understanding & Objective

Introduction

This case study aims to give us an idea of applying EDA in a real business scenario. In this case study, apart from applying the techniques that us have learnt in the EDA module, we will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.

Business Understanding -1

- The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specializes in lending various types of loans to urban customers. You have to use EDA to analyze the patterns present in the data. This will ensure that the applicants capable of repaying the loan are not rejected.
- •When the company receives a loan application, the company has to decide for loan approval based on the applicant’s profile. Two types of risks are associated with the bank’s decision:
- •If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- •If the applicant is not likely to repay the loan, i.e., he/she is likely to default, then approving the loan may lead to a financial loss for the company.

Business Understanding -2

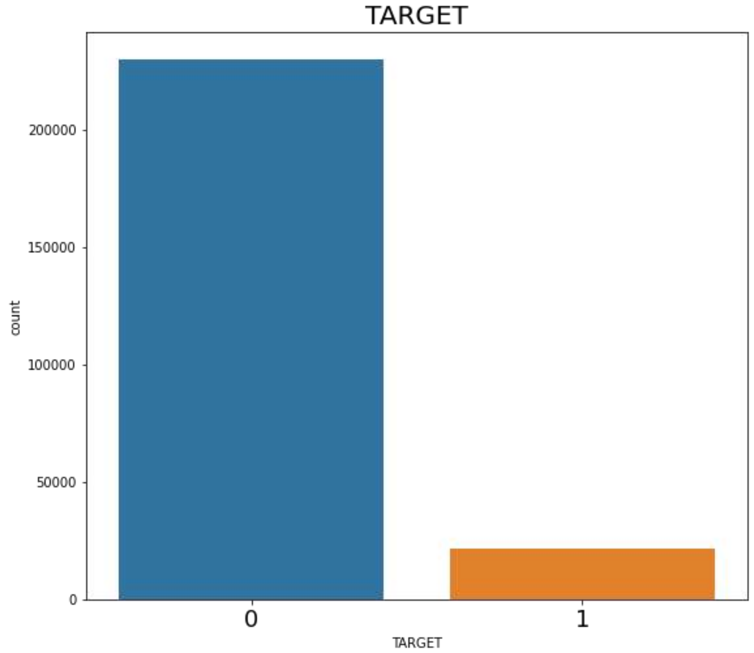
- The data given below contains the information about the loan application at the time of applying for the loan. It contains two types of scenarios:
- •**The client with payment difficulties:** he/she had late payment more than X days on at least one of the first Y instalments of the loan in our sample,
- •**All other cases:** All other cases when the payment is paid on time.
- When a client applies for a loan, there are four types of decisions that could be taken by the client/company):
- **1.Approved:**The Company has approved loan Application
- **2.Cancelled:**The client cancelled the application sometime during approval. Either the client changed her/his mind about the loan or in some cases due to a higher risk of the client he received worse pricing which he did not want.
- **3.Refused:**The company had rejected the loan (because the client does not meet their requirements etc.).
- **4.Unused offer:**Loan has been cancelled by the client but on different stages of the process.
- In this case study, we will use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

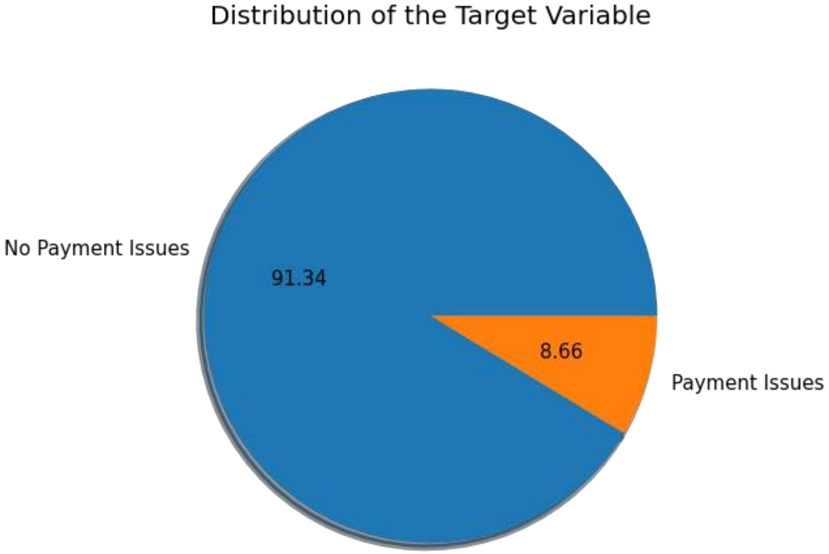
Business Objectives

- This case study aims to identify patterns which indicate if a client has difficulty paying their installments which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. This will ensure that the consumers capable of repaying the loan are not rejected. Identification of such applicants using EDA is the aim of this case study.
- •In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e., the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.
- •To develop your understanding of the domain, you are advised to independently research a little about risk analytics -understanding the types of variables and their significance should be enough).

Analysis- Data Imbalance, Uni/Bi Variate and Correlation

Data Imbalance Check

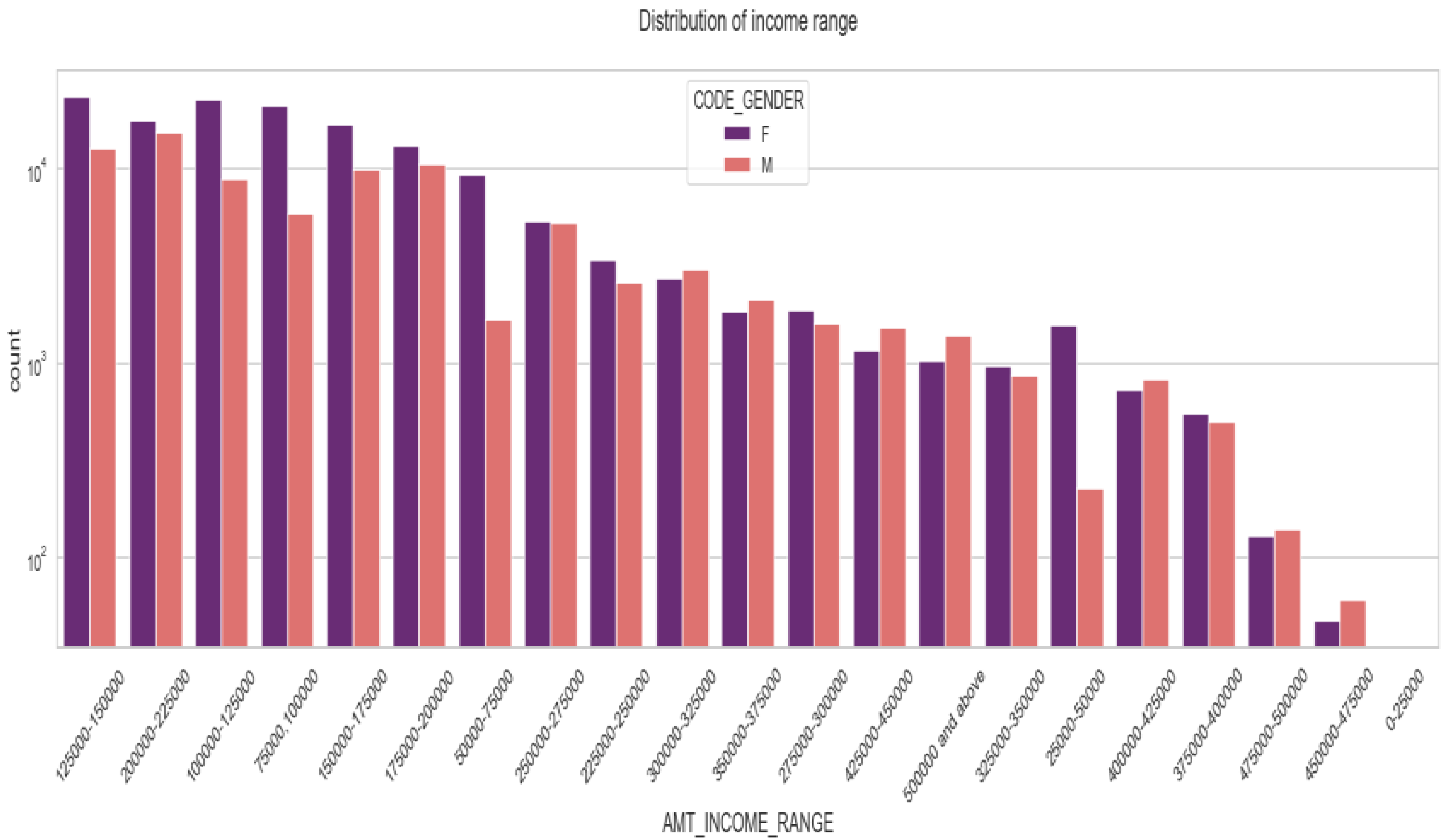




Imbalance Percentage is : 10.55

Univariate Analysis on Target 0

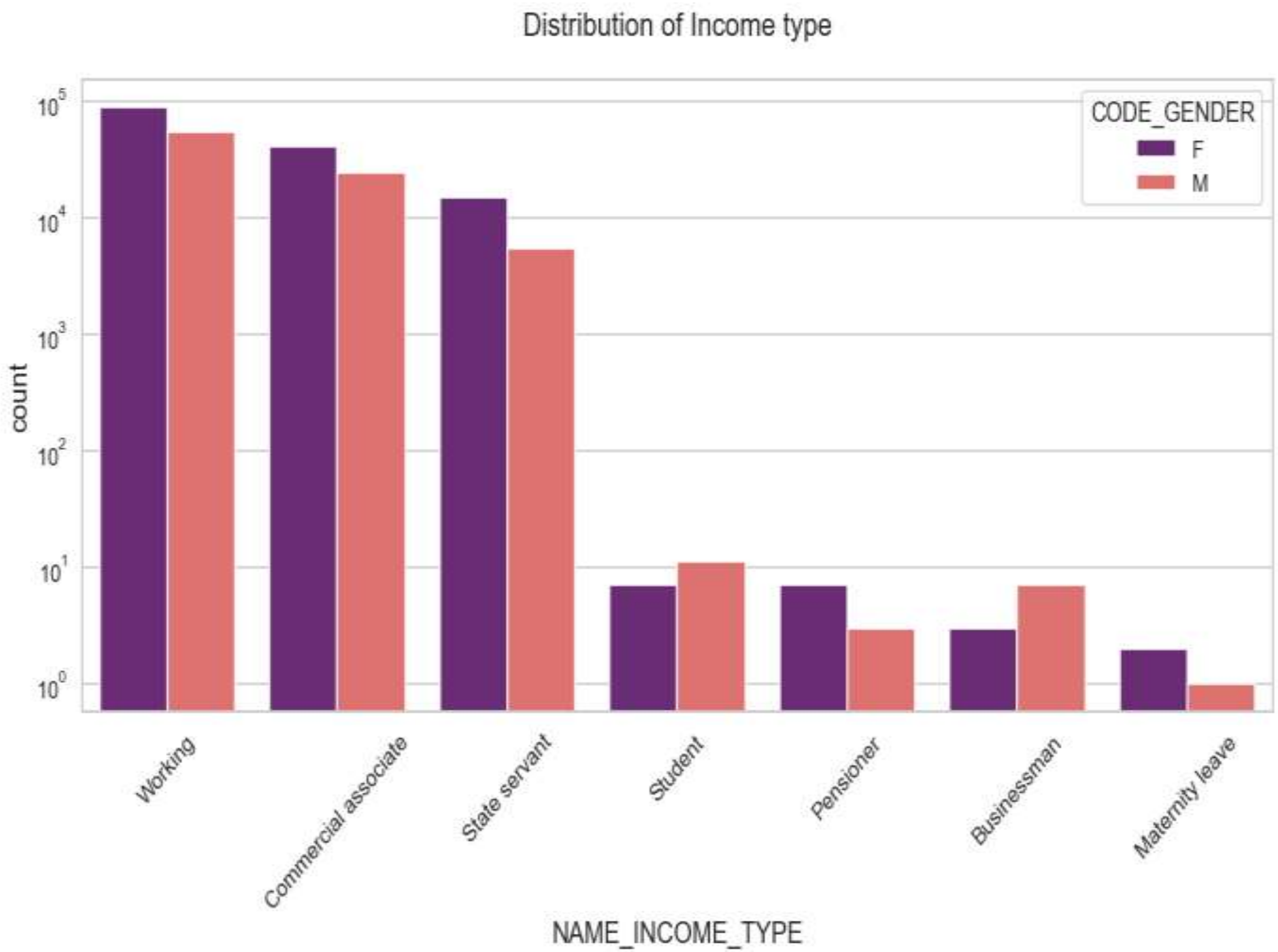
Distribution of Income Range



Observations:-

- 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

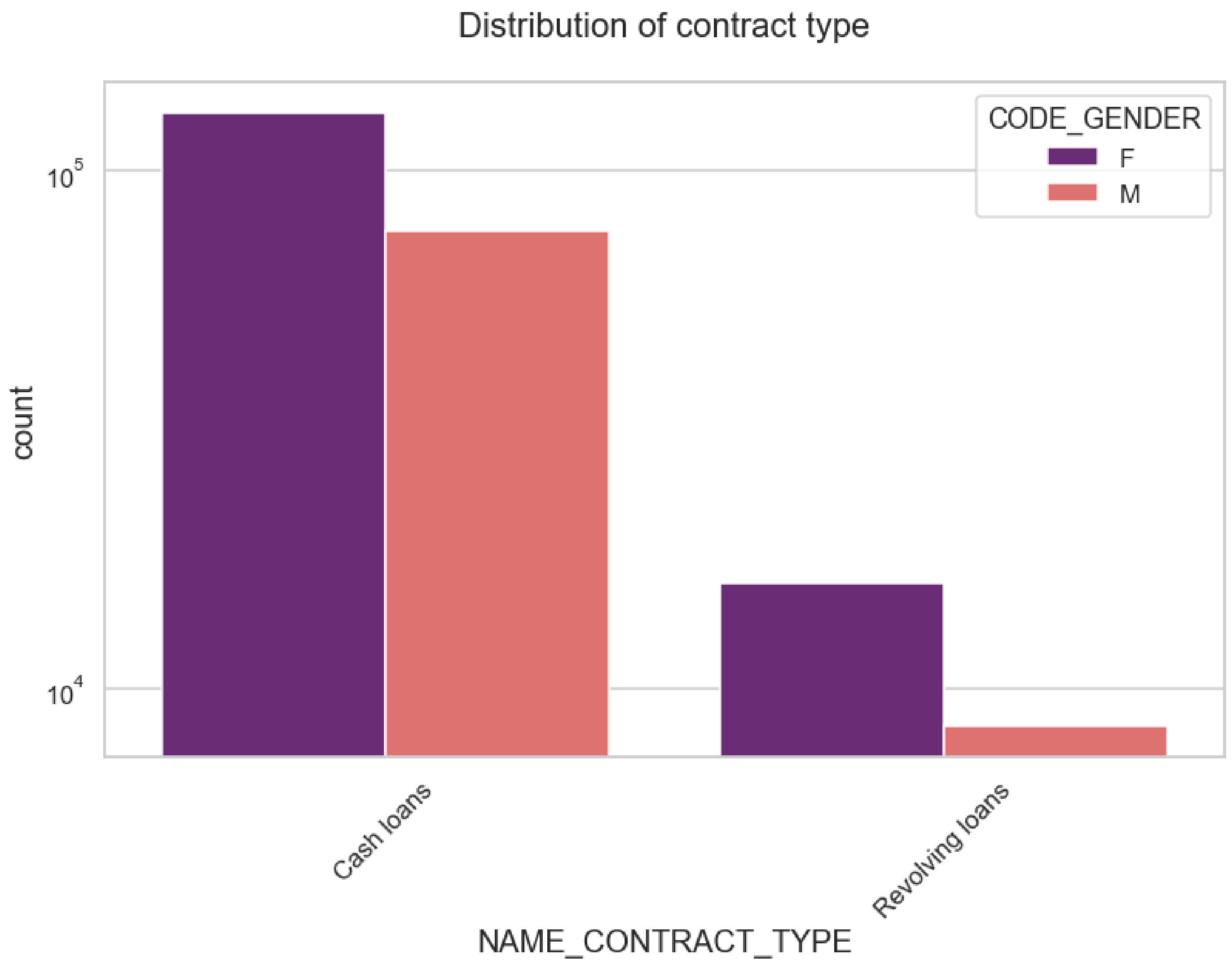
Distribution of Income Type



Observations:-

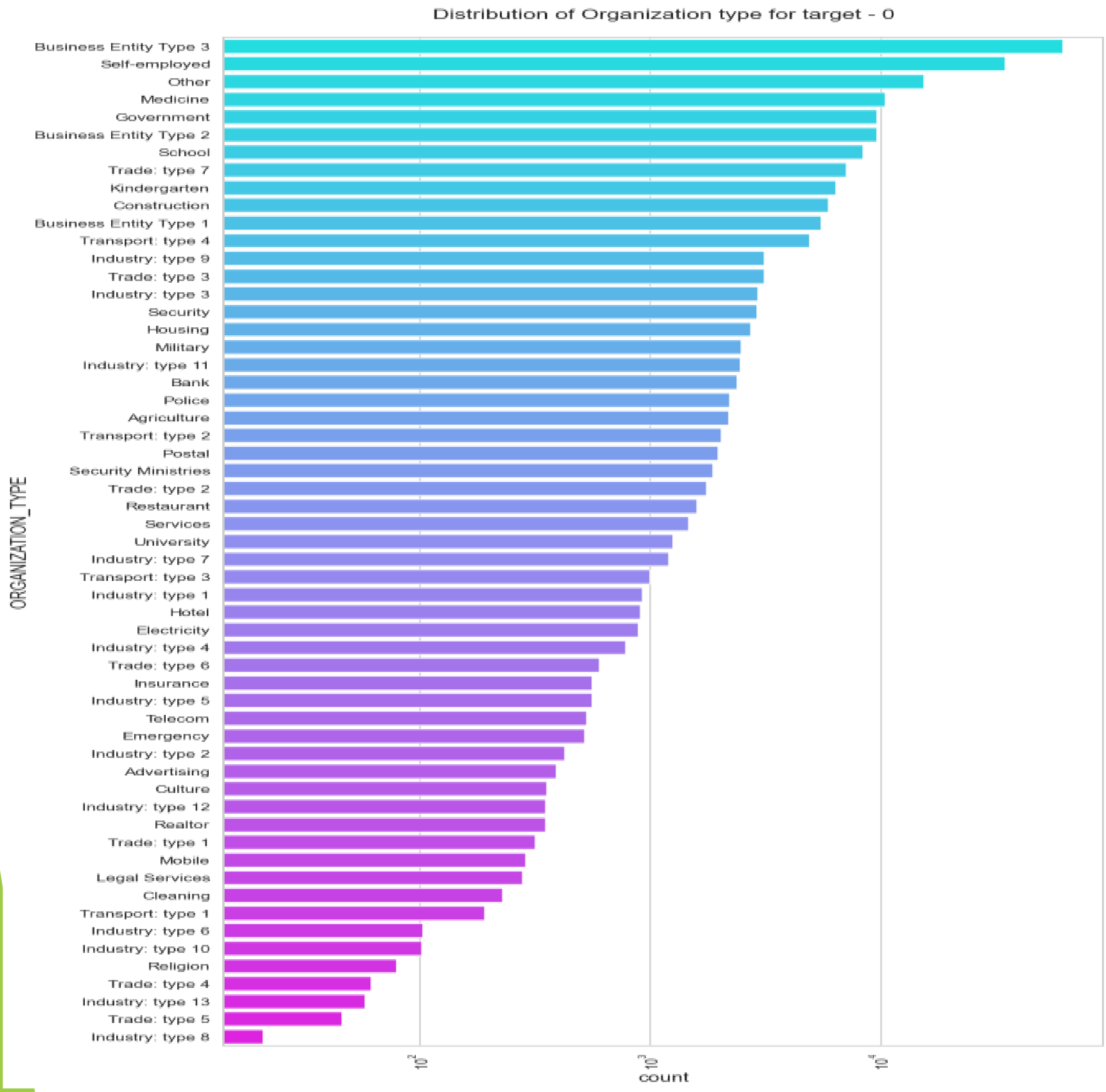
- For income type ‘working’, ’commercial associate’, and ‘State Servant’ the number of credits are higher than others
- For this Females are having more number of credits than male
- Less number of credits for income type ‘student’ ,’pensioner’, ‘Businessman’ and ‘Maternity leave’.

Distribution of Contract Type



Observations:-

- ❑ 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

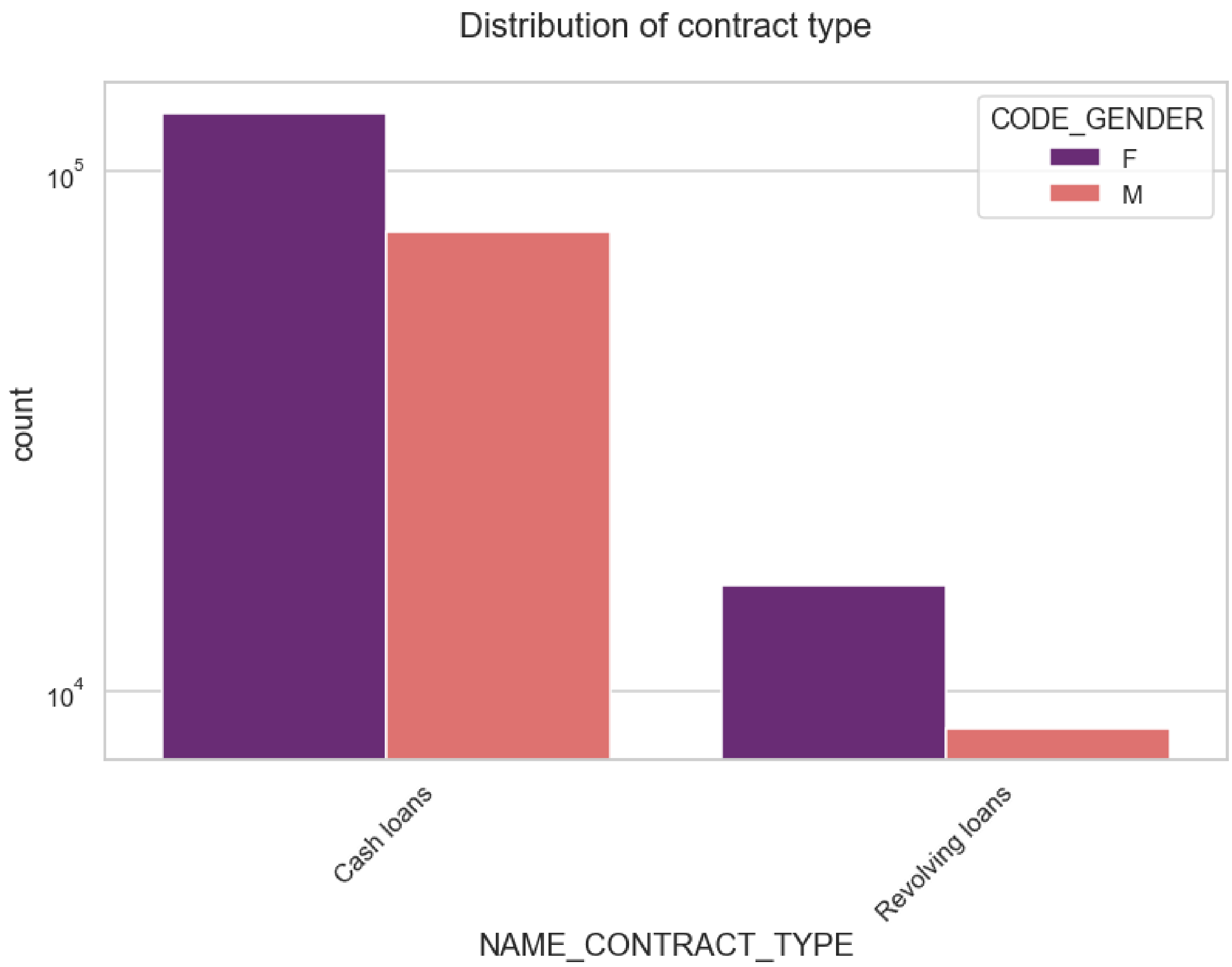


Distribution of Organization Type

Observations:-

- ❑ 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 Contract Type

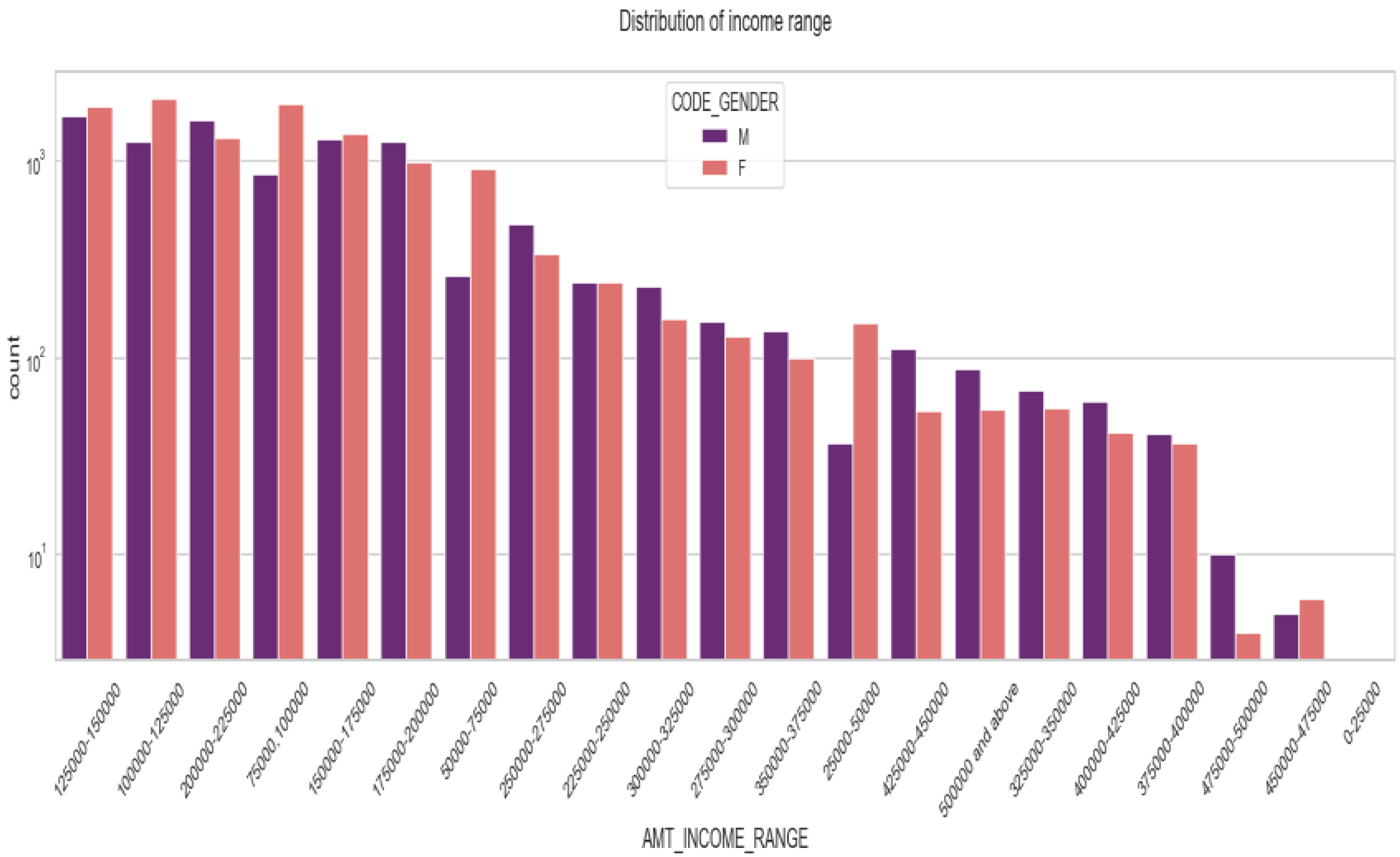


Observations:-

- ❑ 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

Univariate Analysis on Target 1

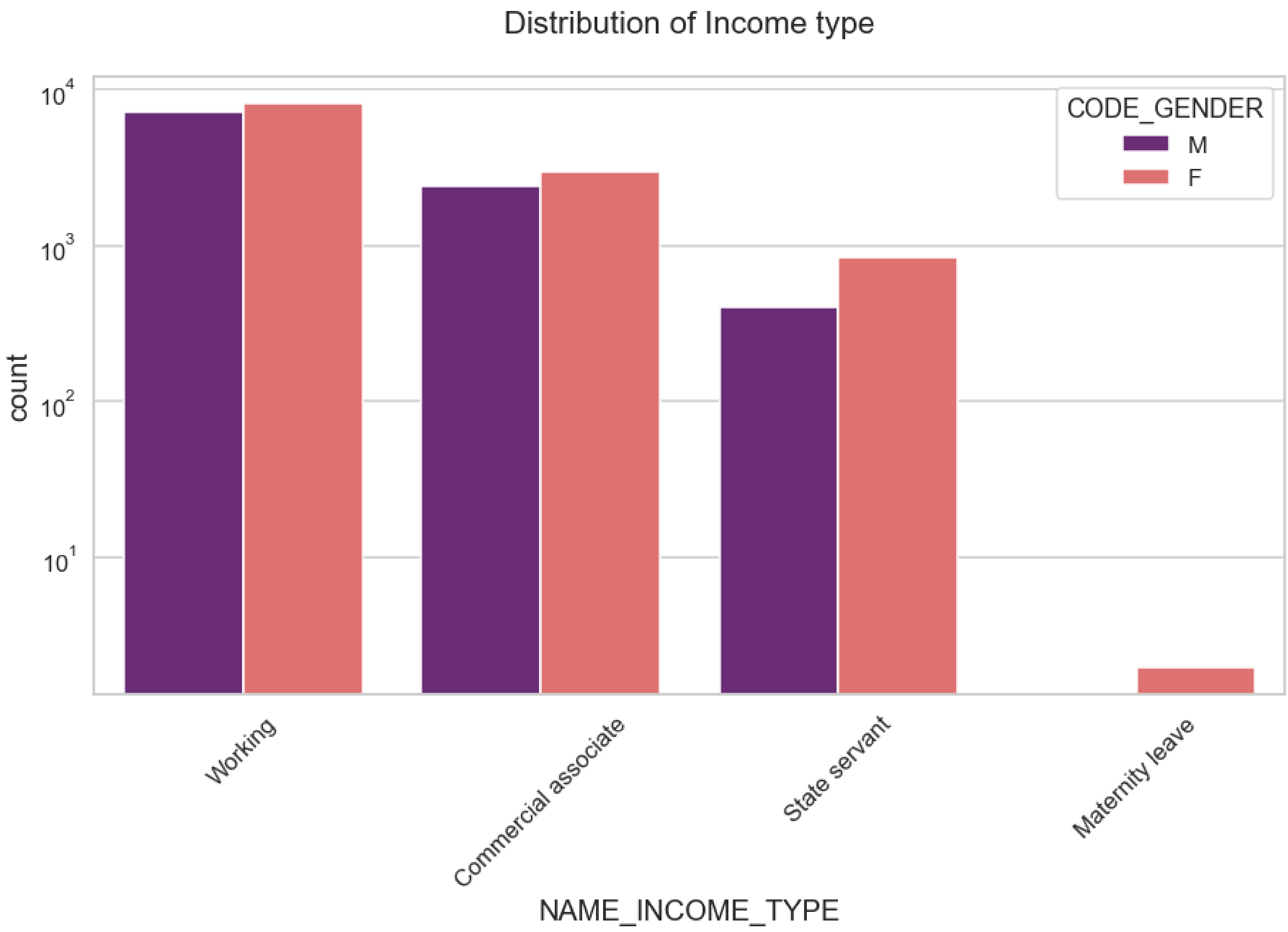
Distribution of Income Range



Observations:-

- ❑ 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

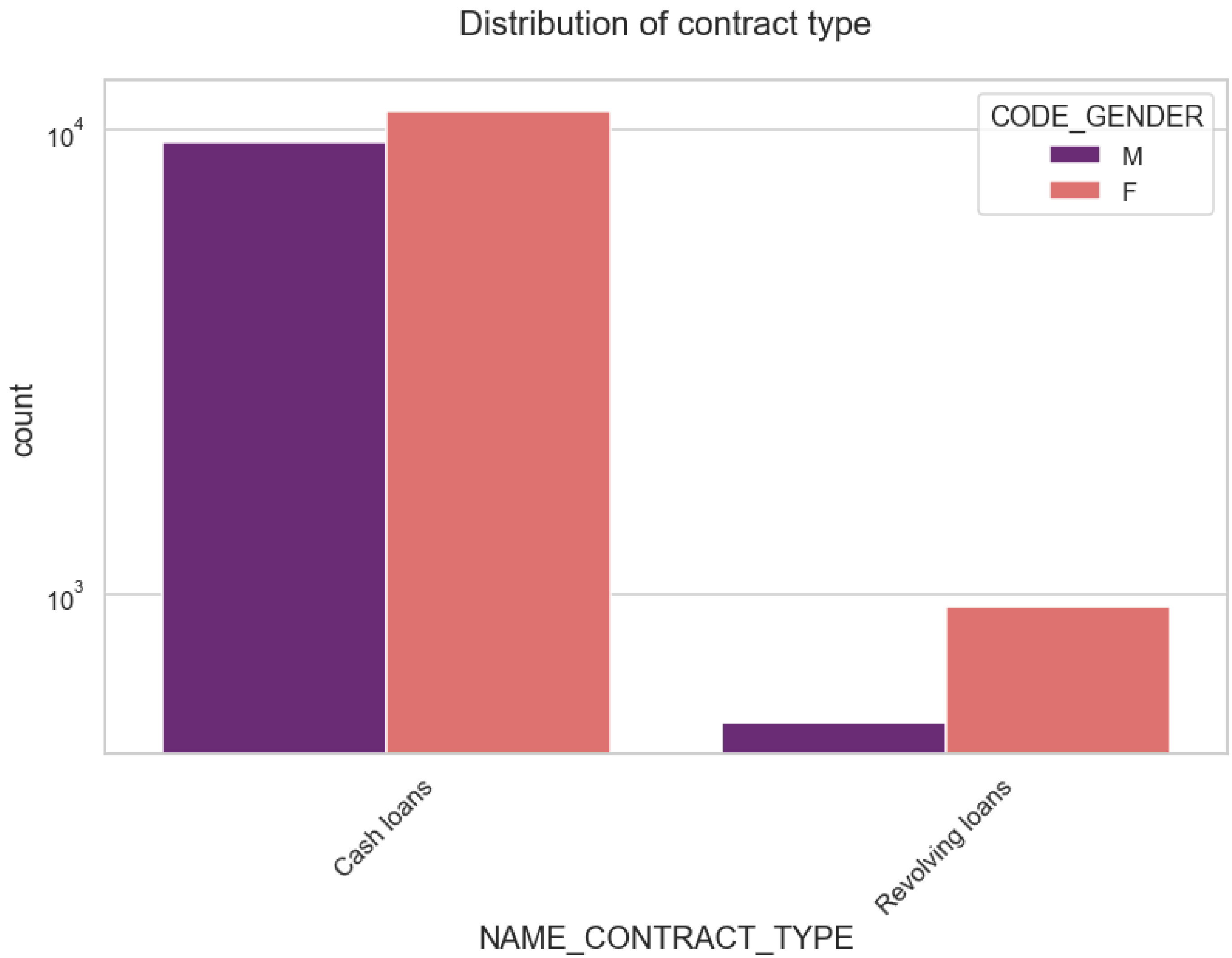
Distribution of Income Type



Observations:-

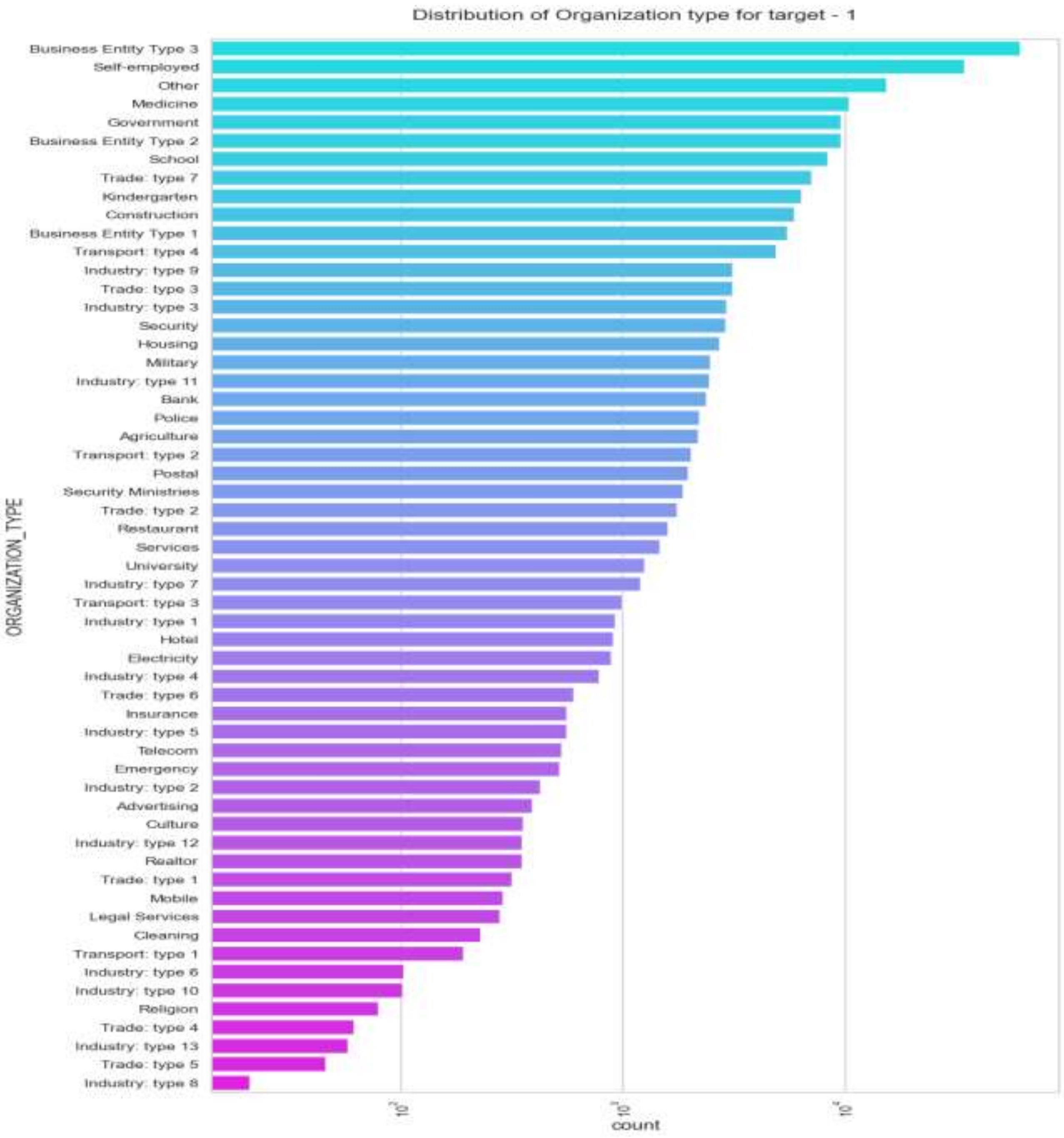
- For income type ‘working’, ’commercial associate’, and ‘State Servant’ the number of credits are higher than other i.e.Maternity leave
- For this Females are having more number of credits than male
- Less number of credits for income type ‘Maternity leave’

Distribution of Contract Type



Observations:-

- 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



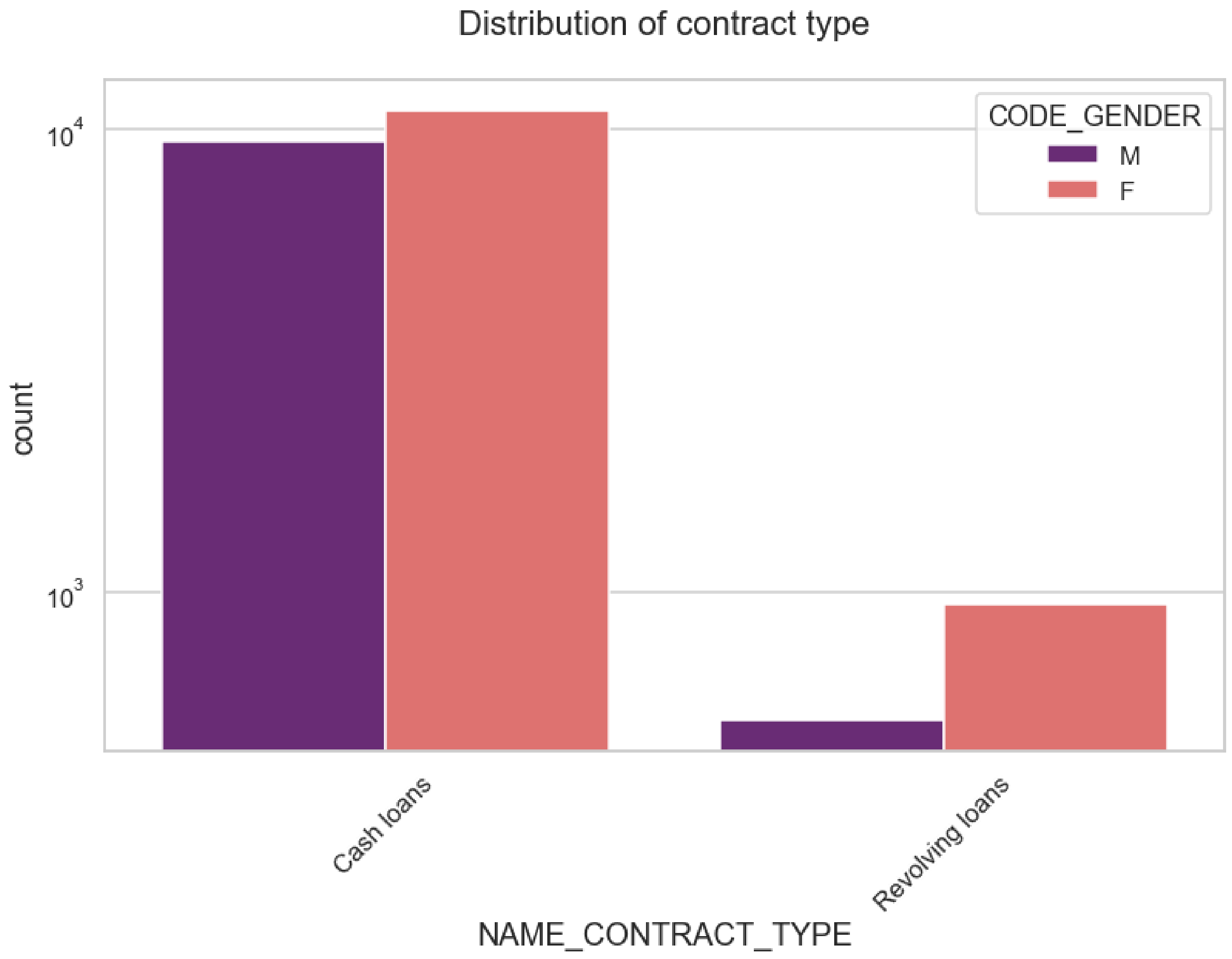
Distribution of Organization Type

Observations:-

- ❑ 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 Contract Type

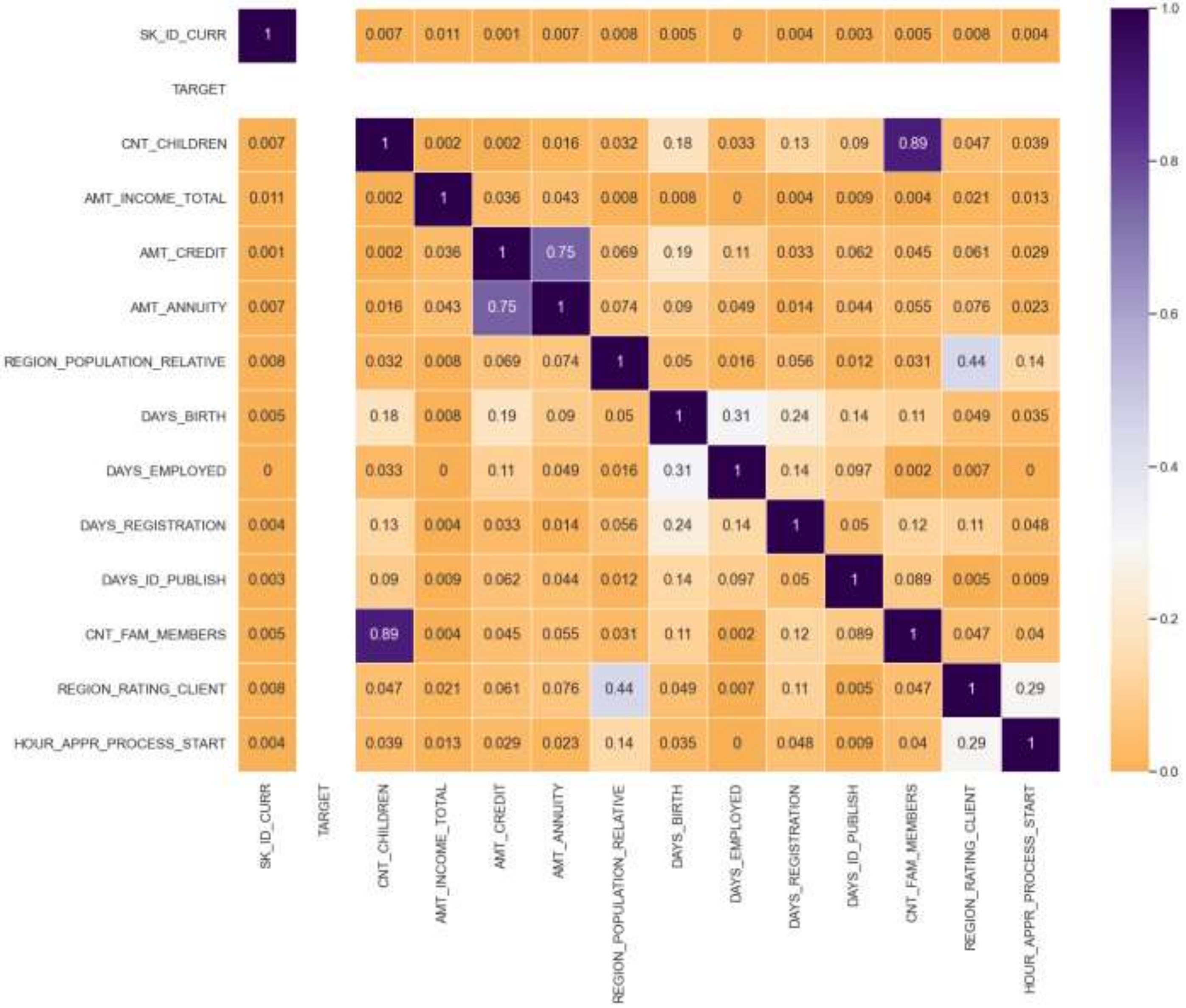


Observations:-

- 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

Correlation of Target 1

Correlation



Insights from Correlation

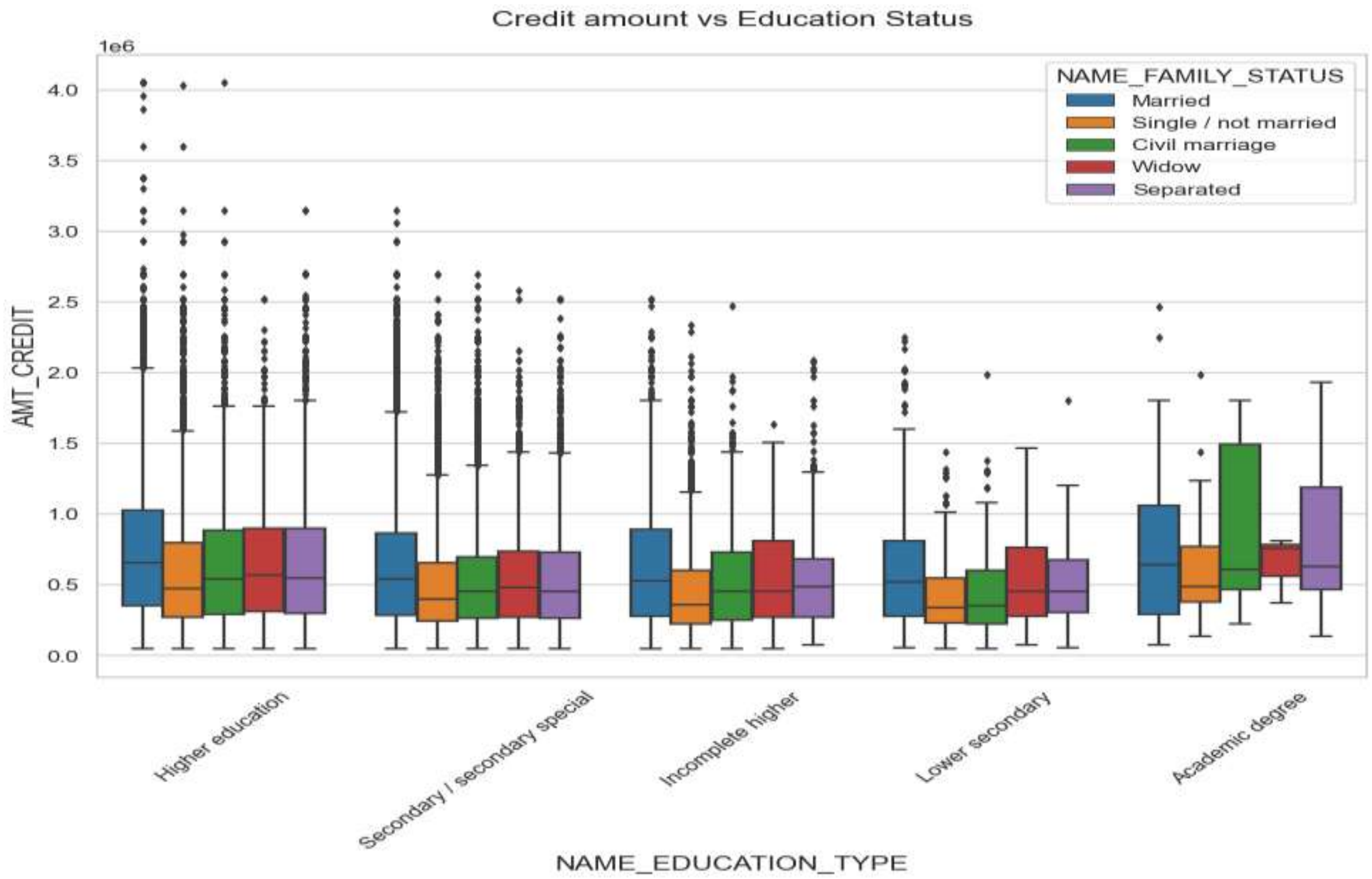
1. Credit amount is highly correlated with amount of goods price for the Client with payment difficulties (Target 1)

2. The correlation is strong between family member and children counts with payment difficulties (Target 1)

3. Days_birth and number of children also has a correlation.

Bi-variate Analysis on Target 0

Credit Amount Vs Education Status



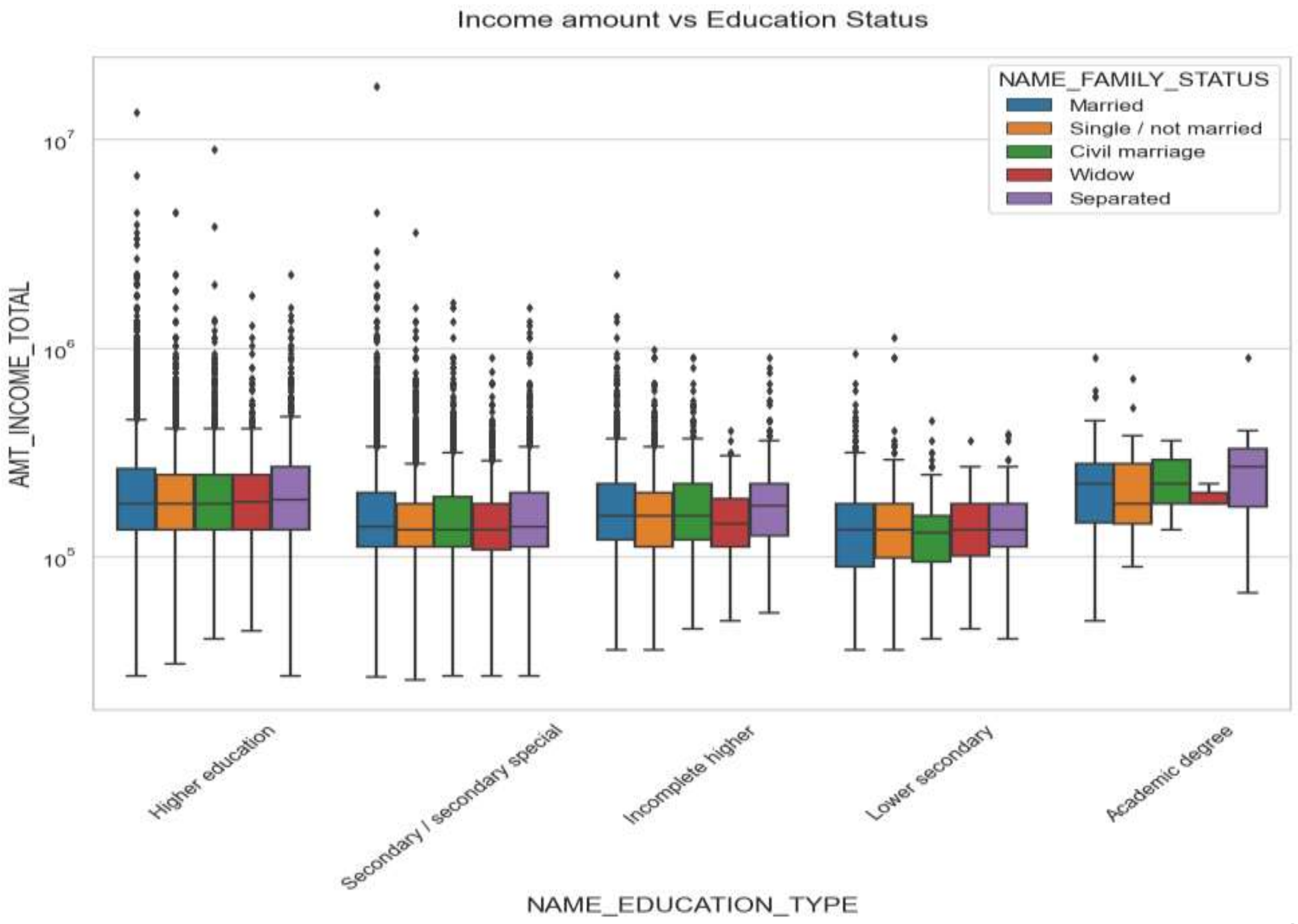
Observations:-

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

2. higher education of family status of 'marriage', 'single' and 'civil marriage' are having more outliers

3. Civil marriage for Academic degree is having most of the credits in the third quartile.

Income Amount Vs Education Status



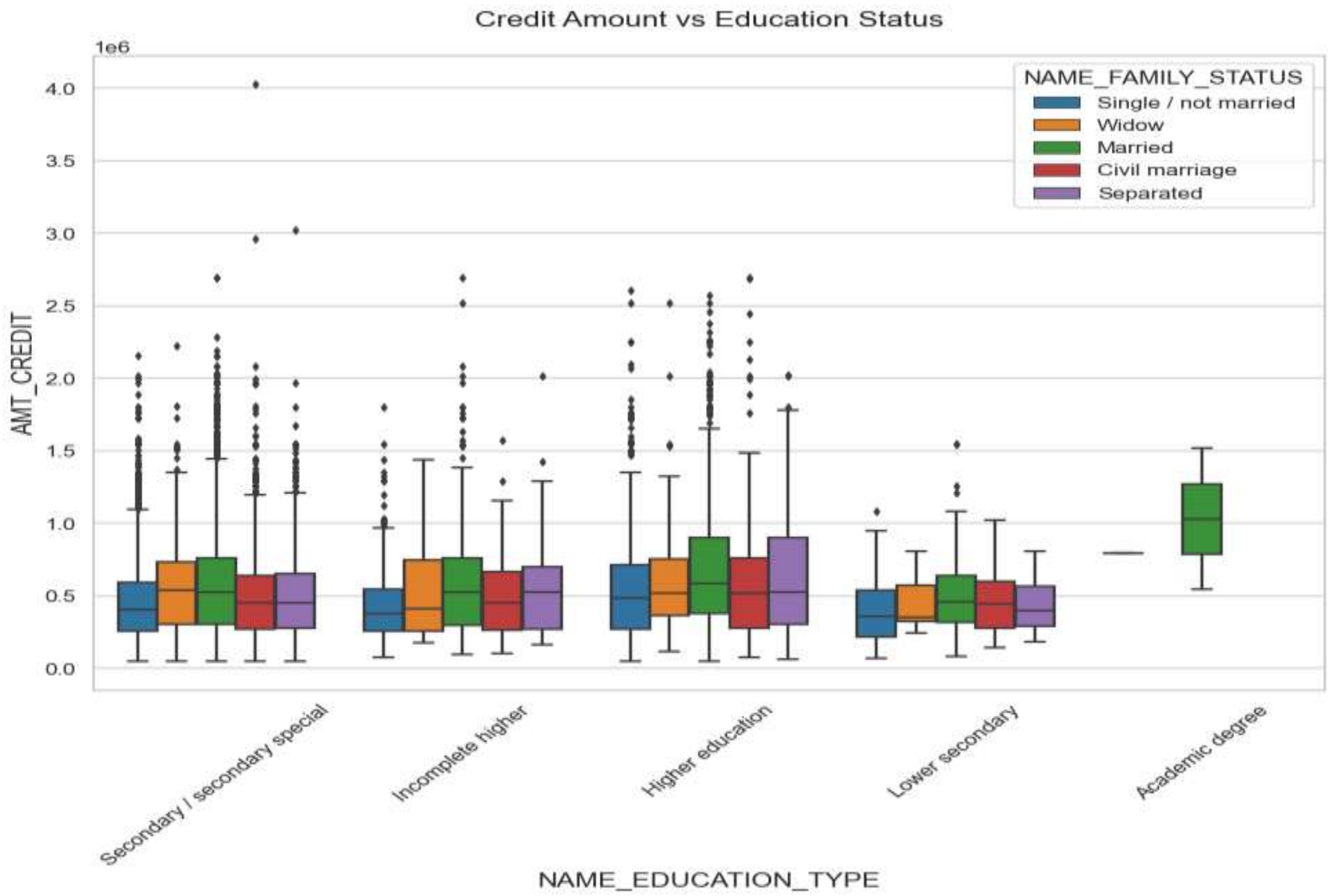
Observations:-

1. For Education type 'Higher education' the income amount is mostly equal with family status

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

Bi-variate Analysis on Target 1

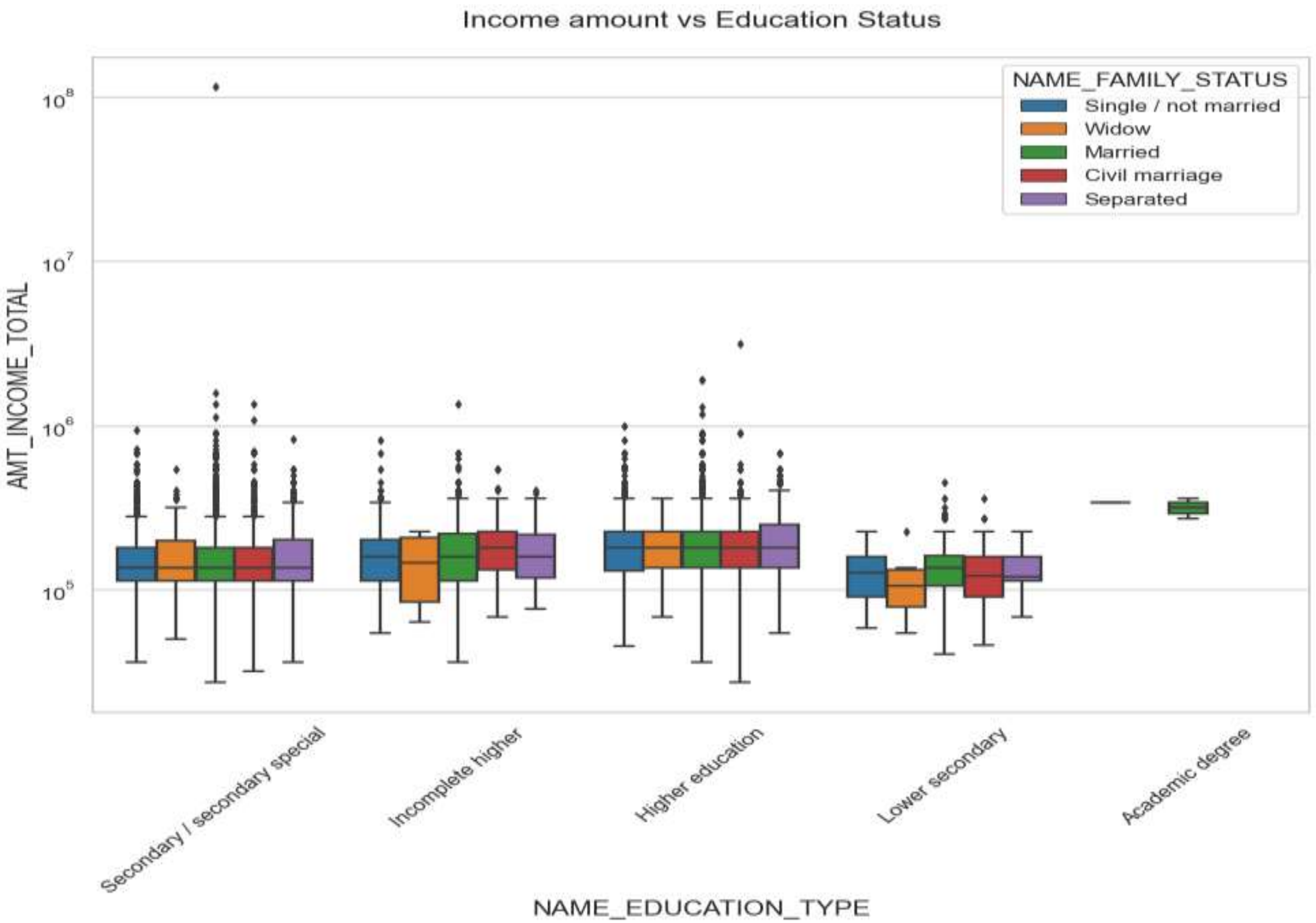
Credit Amount Vs Education Status



Observations:-

1. Family status of 'civil marriage', 'marriage' and 'separated' of Academic degree education are having higher number of credits than others.
2. Civil marriage for Academic degree is having most of the credits in the third quartile.

Income Amount Vs Education Status



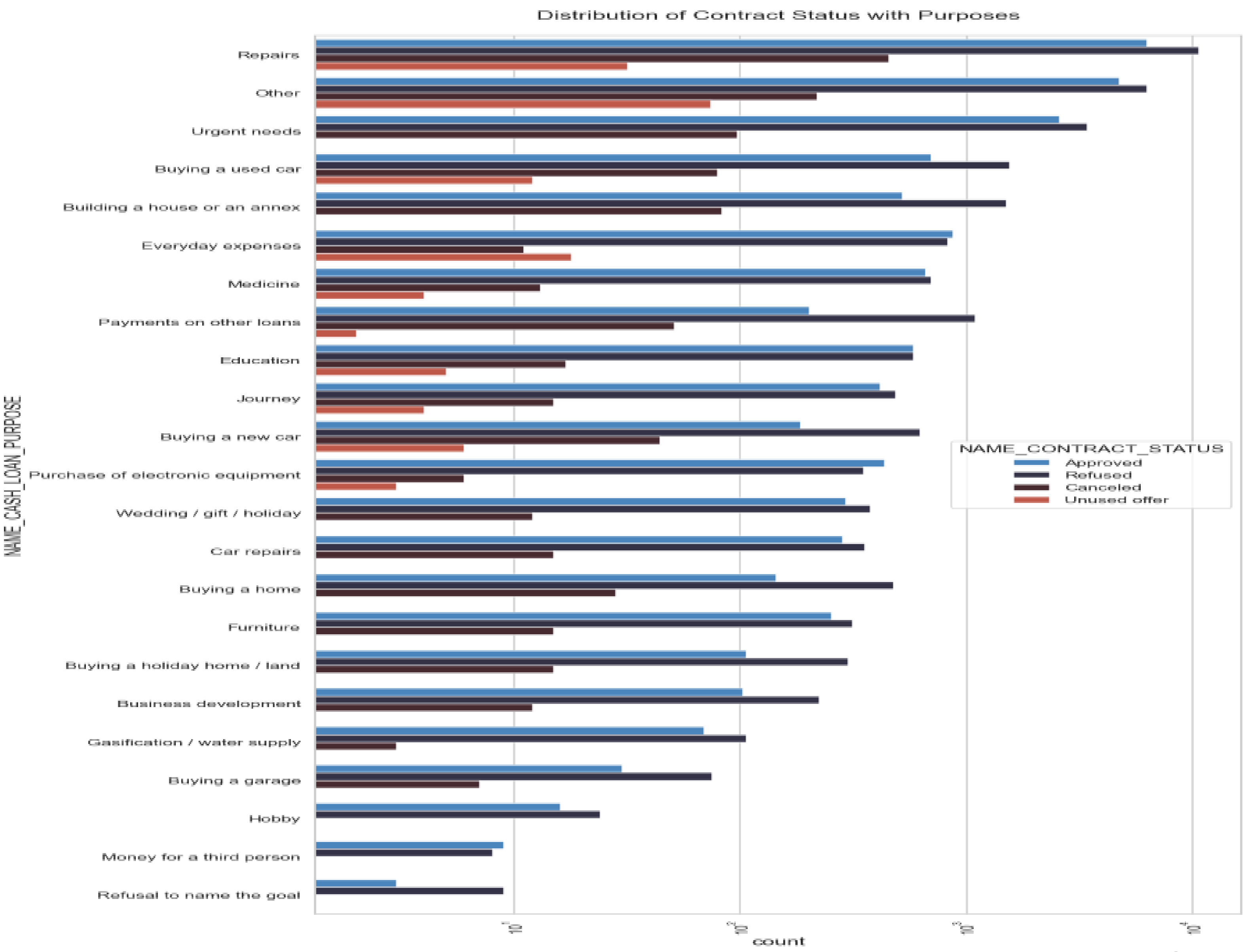
Observations:-

1. 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 that Higher education

2. Lower secondary are have less income amount than others.

Univariate Analysis after Merging Data

Income Amount Vs Education Status

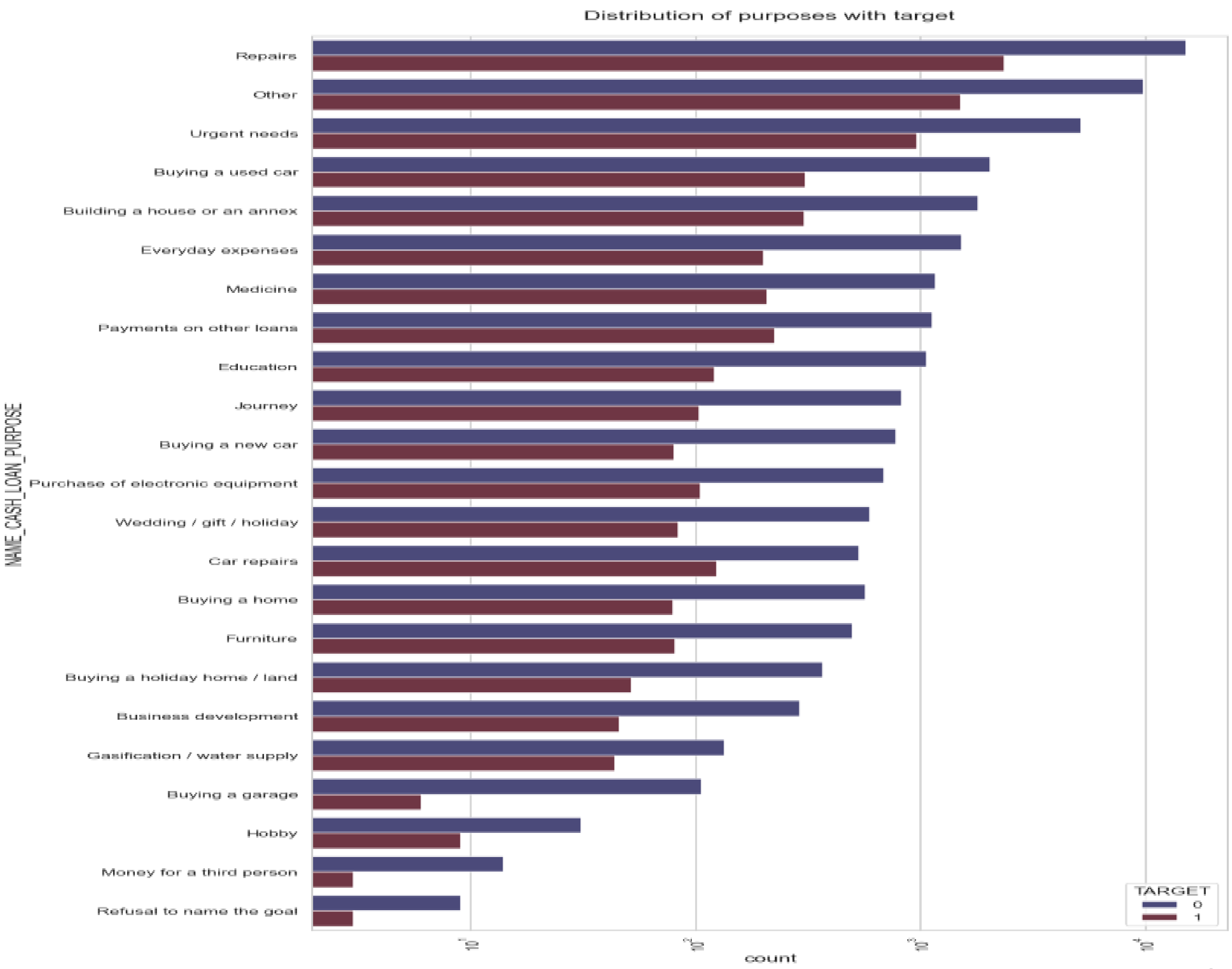


Observations:-

1. Repairs got most refused loans
Education has similar outcomes for approval and rejection of loans

2. Paying other loans and buying a new car is having significant higher rejection than approves.

Distribution of Purpose with Target

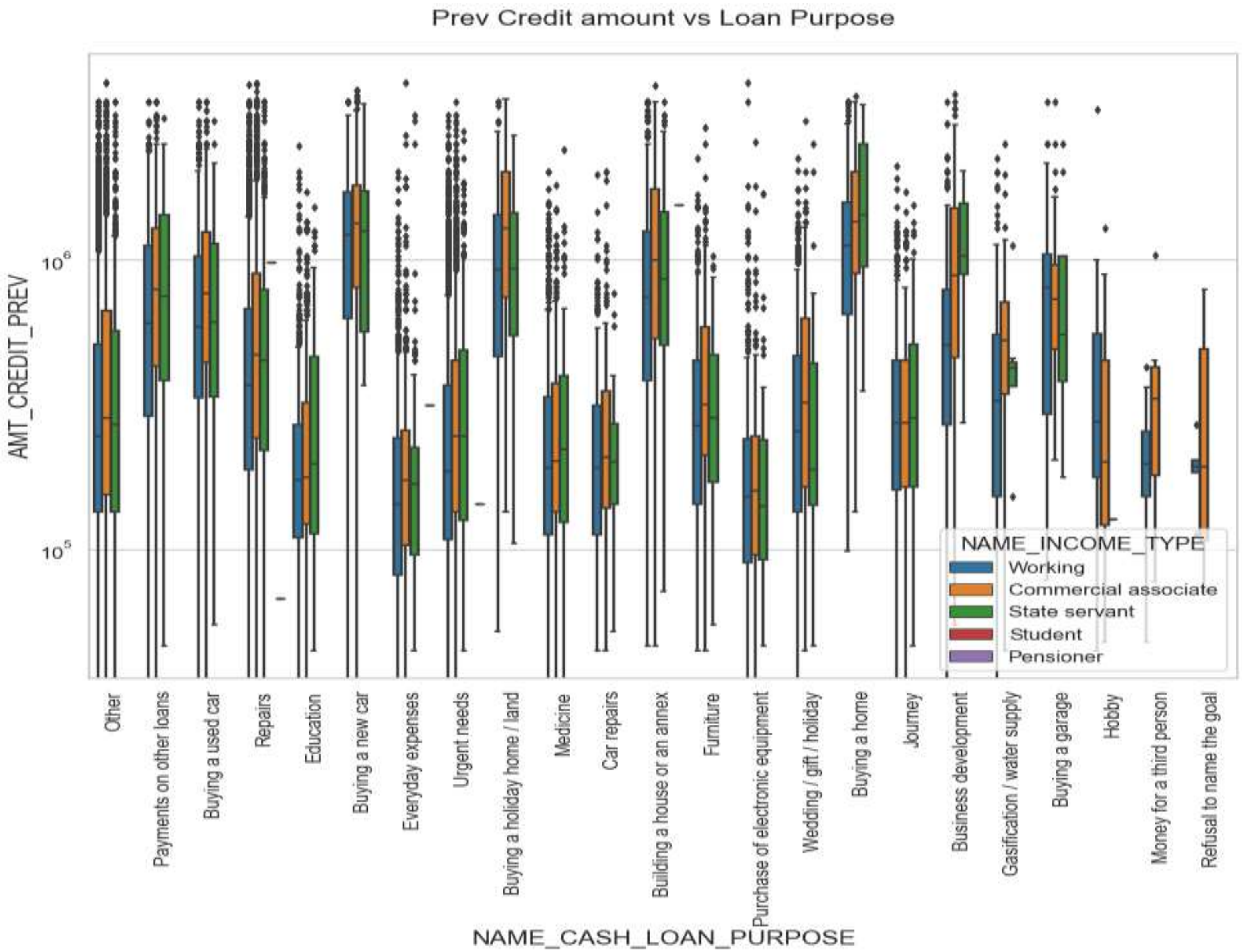


Observations:-

1 Repairs are dealing with more difficulties in payment on time

2. Buying a garage, Business development, Buying land, Buying a new car and Education having basically higher loan payment

Previous Credit Amount vs Loan Purpose



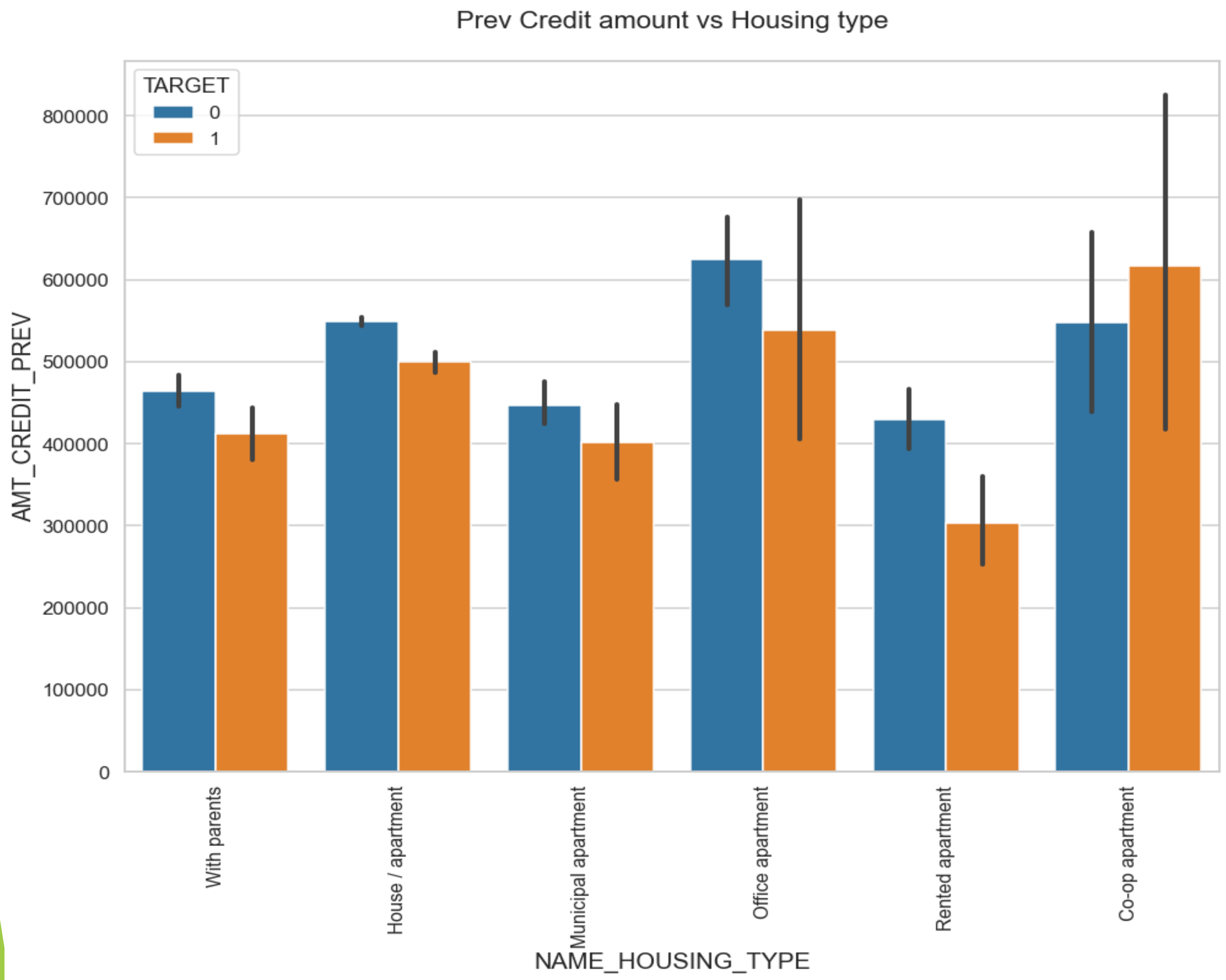
Observations:-

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 Housing Type



Observations:-

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 appartment for successful payments.

Conclusion

- Conclusion:-
- Bank should ive more loans from Housing Type 'With parents' as they less number of defaulters. Co-OP Apartment has the highest defaulters.
- Bank should avoid people on income type as 'working' as they less successful payments.
- Also, Loan Purpose 'Repairs' have the highest unsuccessful payments.

Thank You