

CREDIT EDA ASSIGNMENT

By: Ria Singh



Approach to Solve the Problem Statement



DATA CLEANING



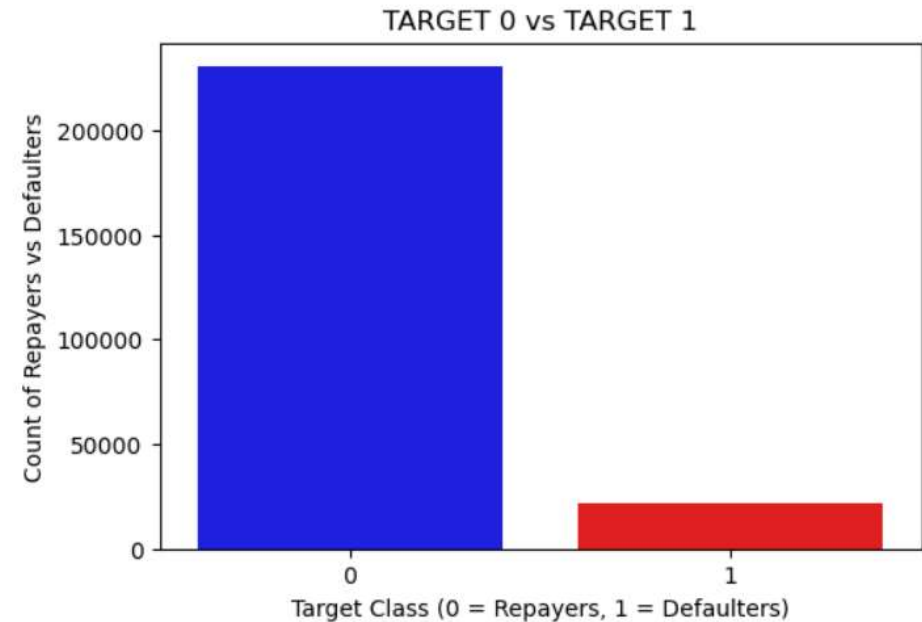
DATA ANALYSIS



DATA VISUALIZATION

Dividing the Target Variable

- The imbalance ratio in the dataset for clients that can repay vs clients that have repayment difficulties is **10.55:1**

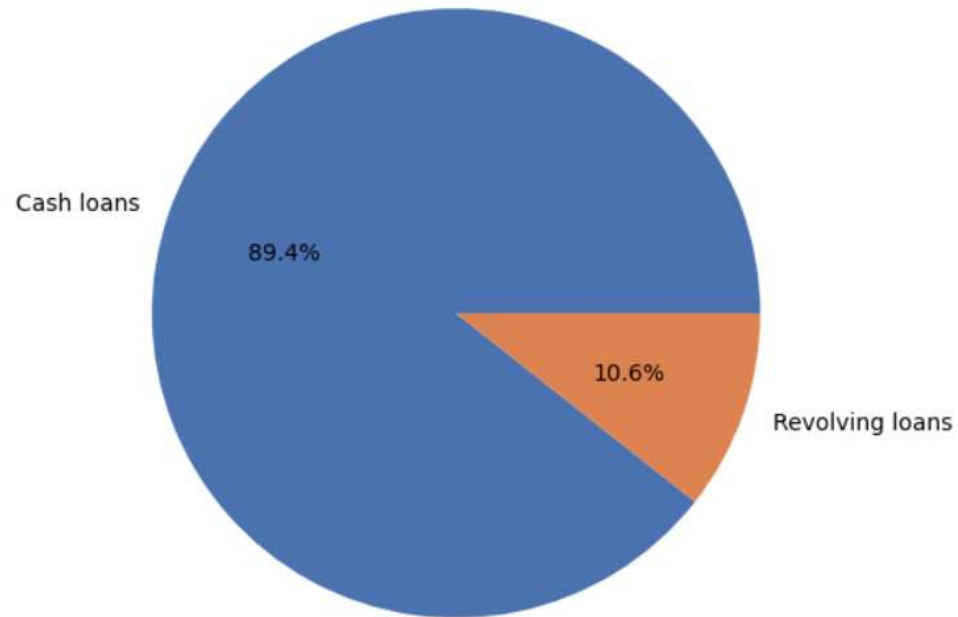


UNIVARIATE ANALYSIS FOR TARGET 0 VS TARGET 1

- Type of Contract
- Gender of the Applicant
- Range of Applicant's Income
- Family Status of Applicants
- Income Type of Applicants
- Education Type

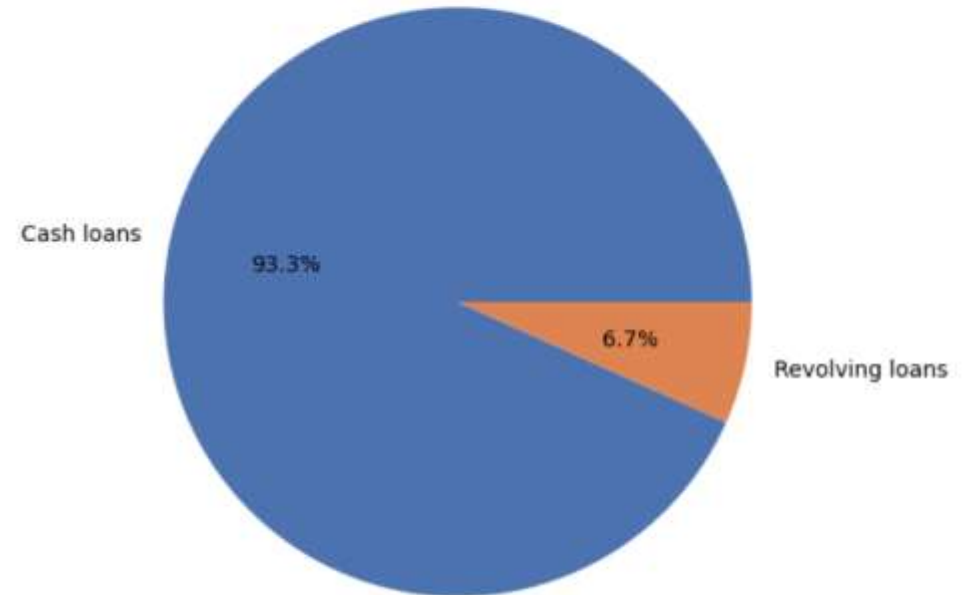
The vast majority of applicants prefer cash loans over revolving loans for both target variables.

Distribution of NAME_CONTRACT_TYPE



Target 0

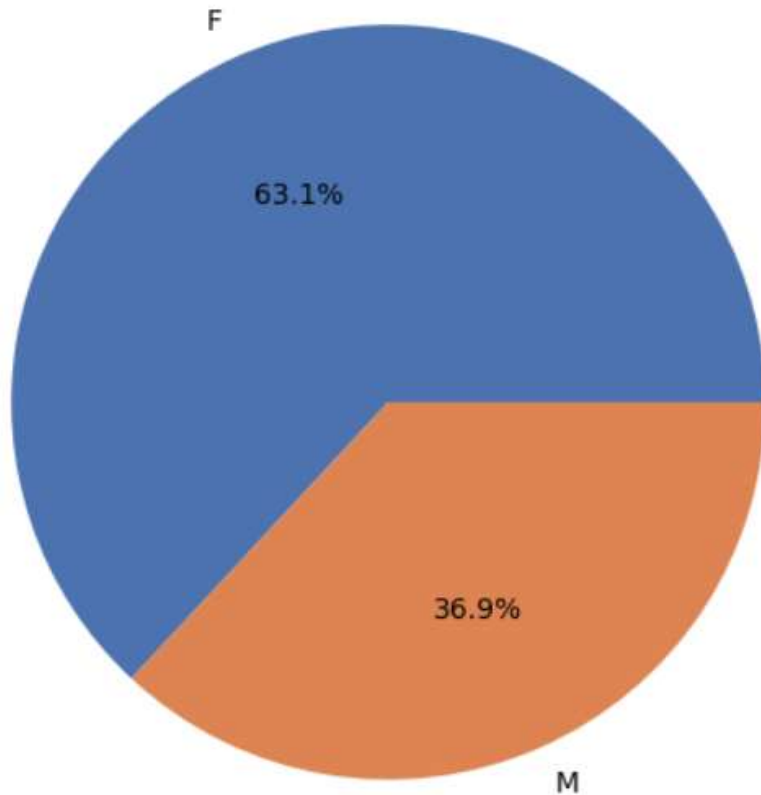
Distribution of NAME_CONTRACT_TYPE



Target 1

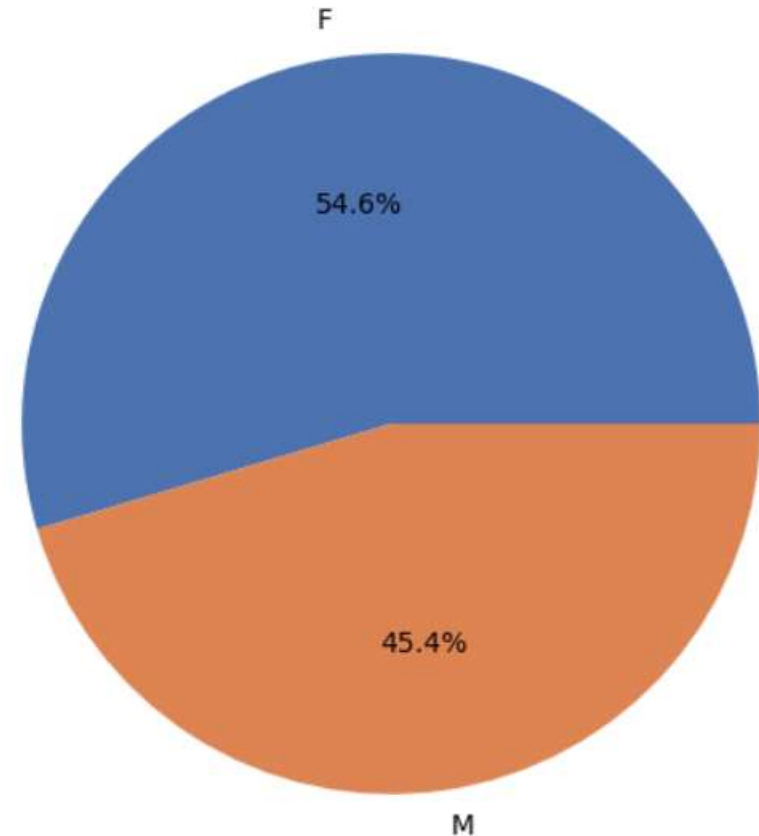
**The number of female applicants are higher than male applicants.
Female applicants are more likely to face difficulties in repaying loan
compared to male applicants.**

Distribution of CODE_GENDER



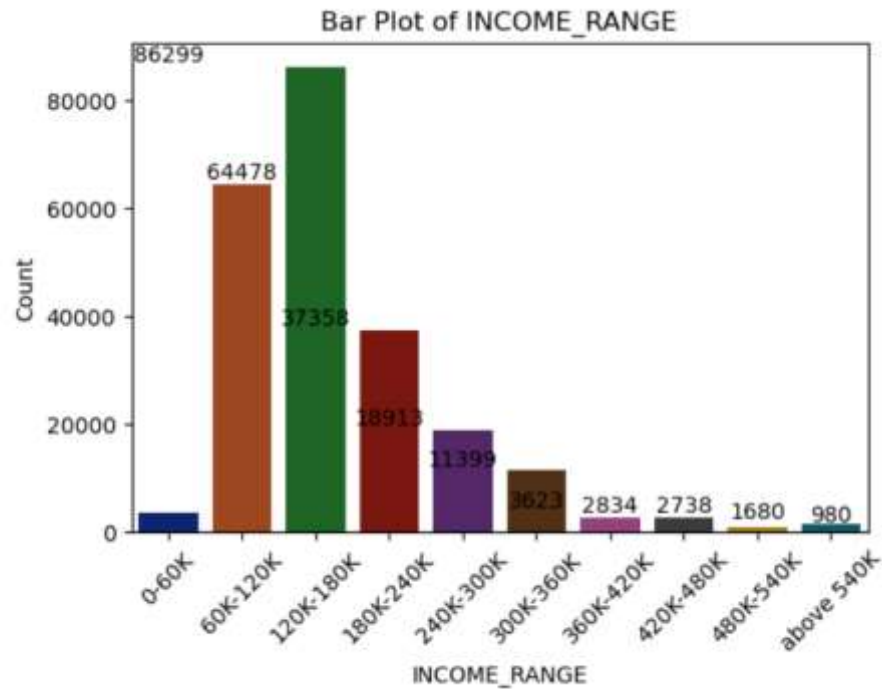
Target 0

Distribution of CODE_GENDER

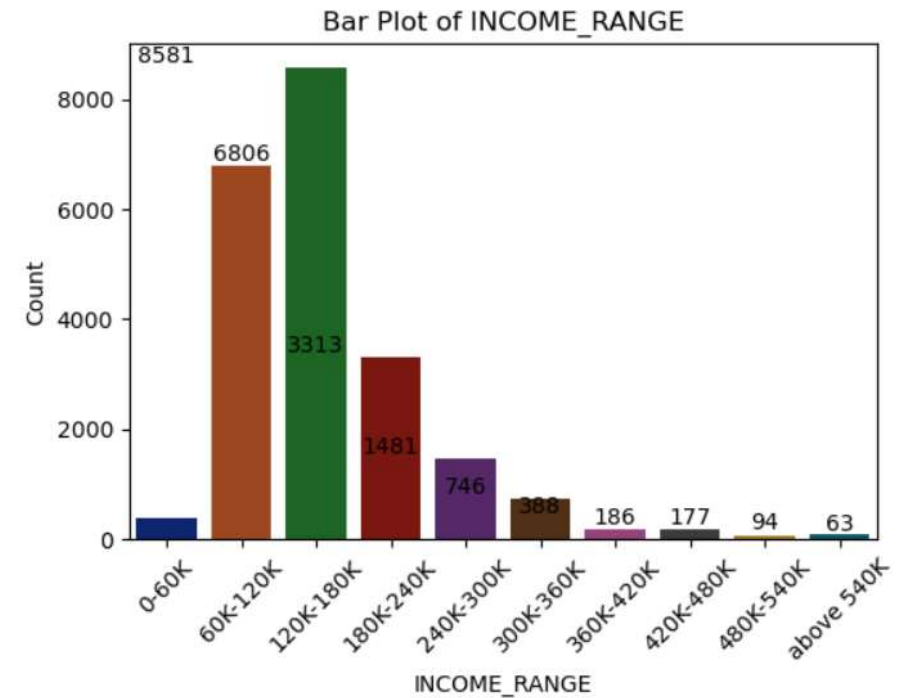


Target 1

- The maximum number of applicants lie in the income range of 60k to 180k.
- The applicants in the income range between 120k to 180k are most likely to be defaulters.

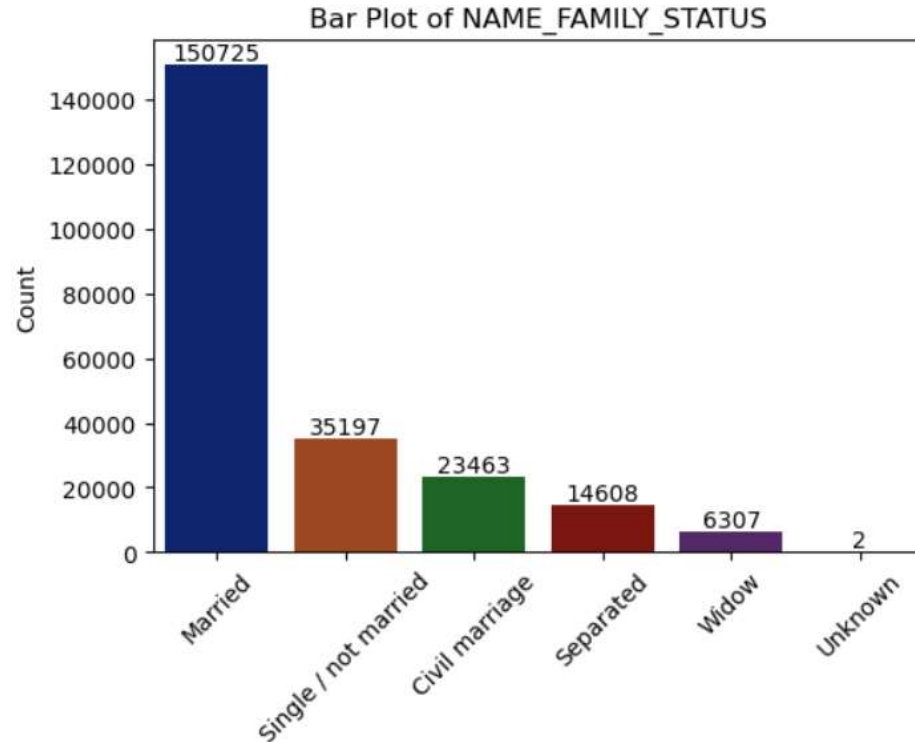


Target 0

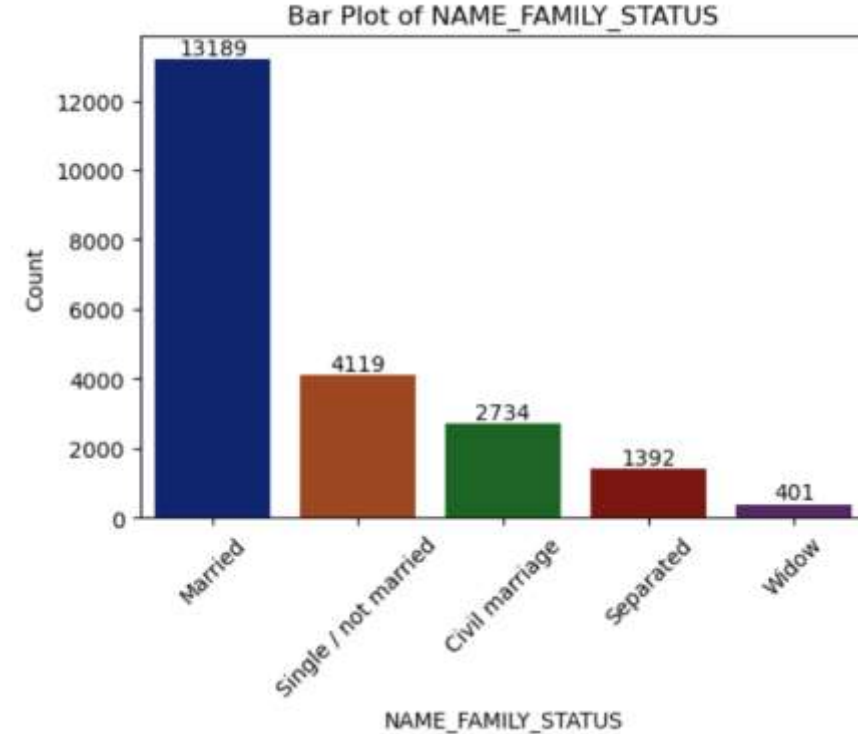


Target 1

Most of the applicants are married followed by single/married.



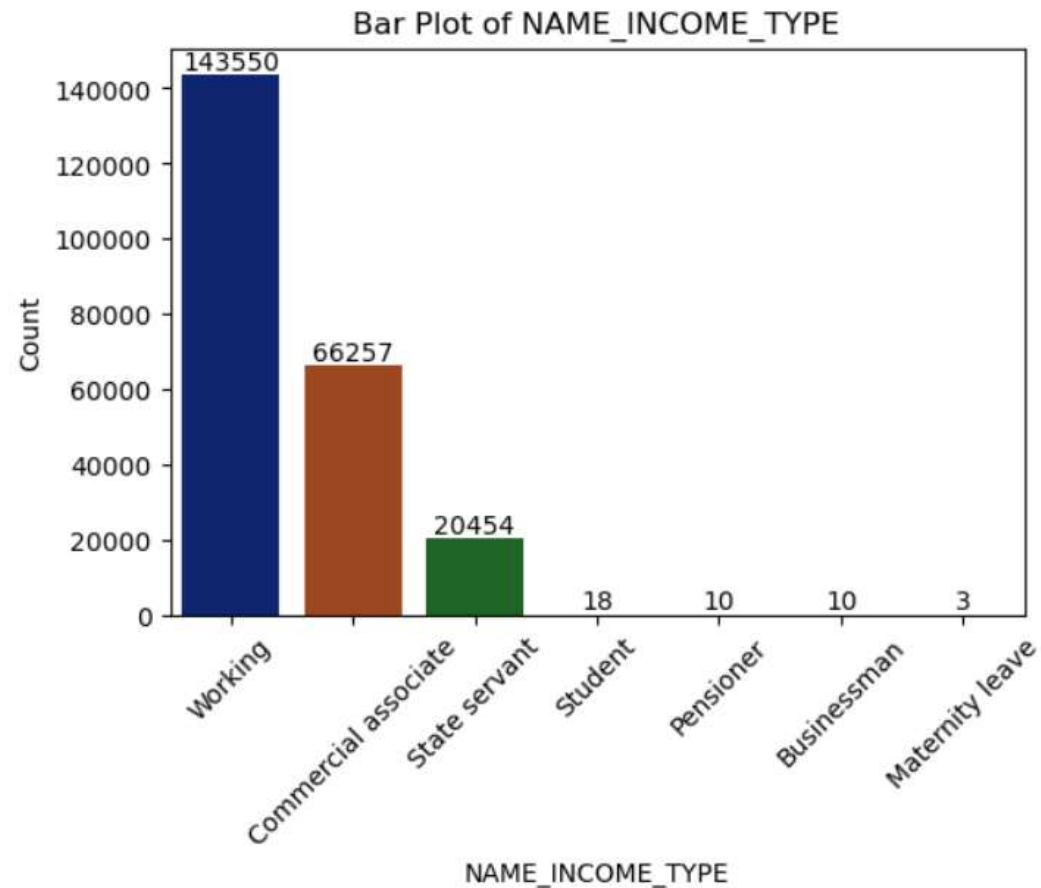
Target 0



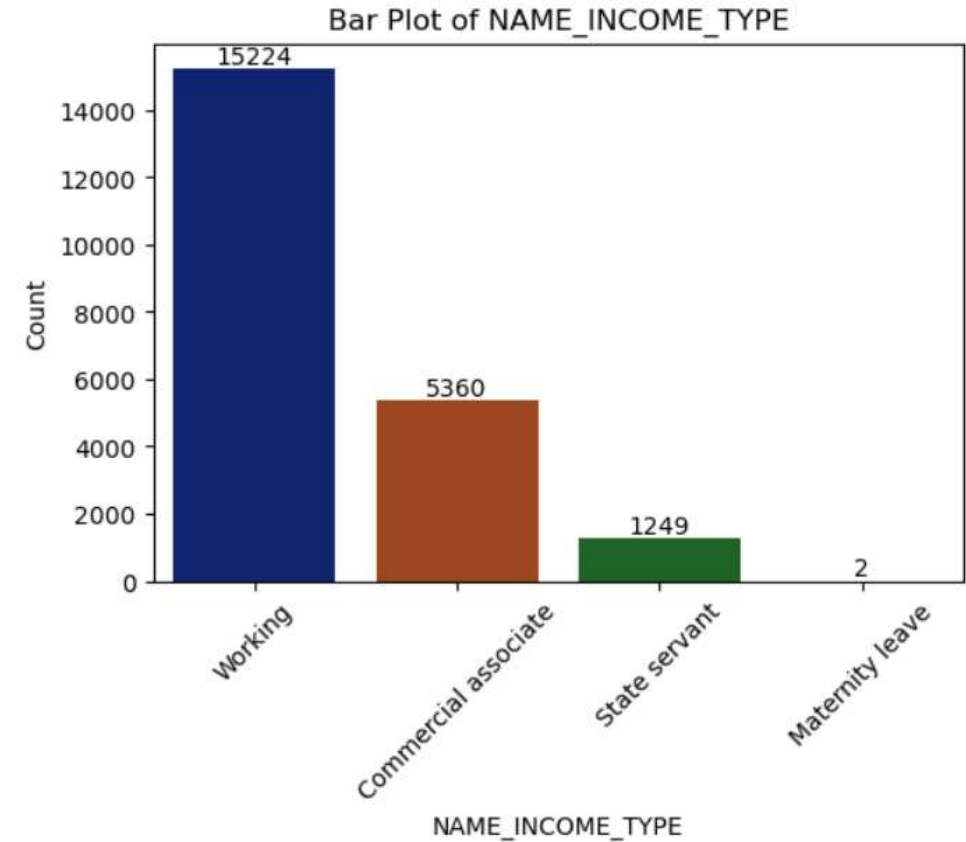
Target 1

Most of the applicants are working.

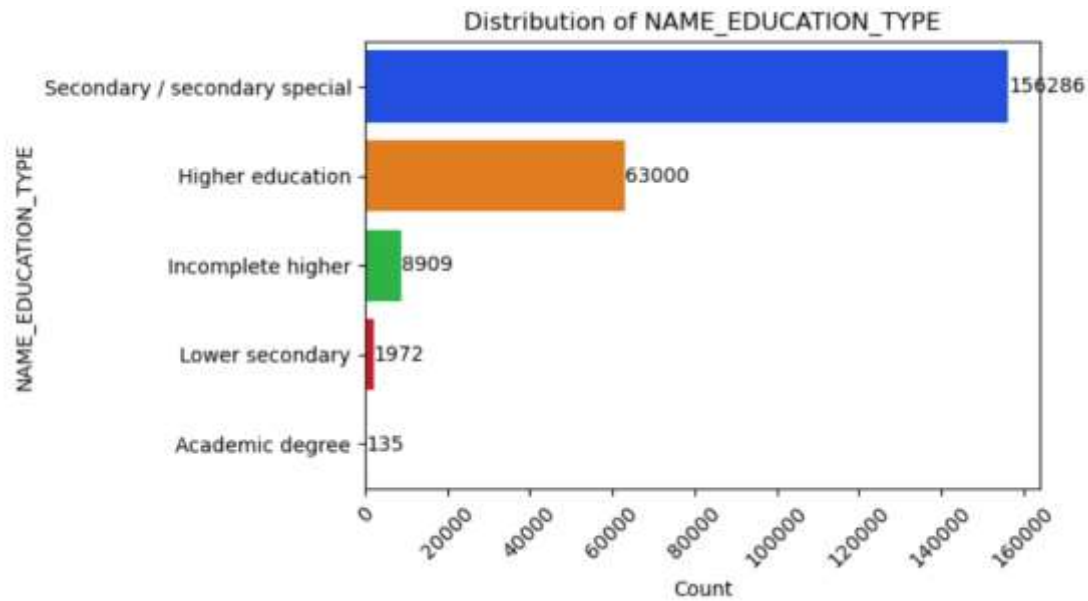
Student, Businessman, and pensioner are not present in the target 1, hence even if their number is lesser, they are safer applicants compared to the others.



Target 0

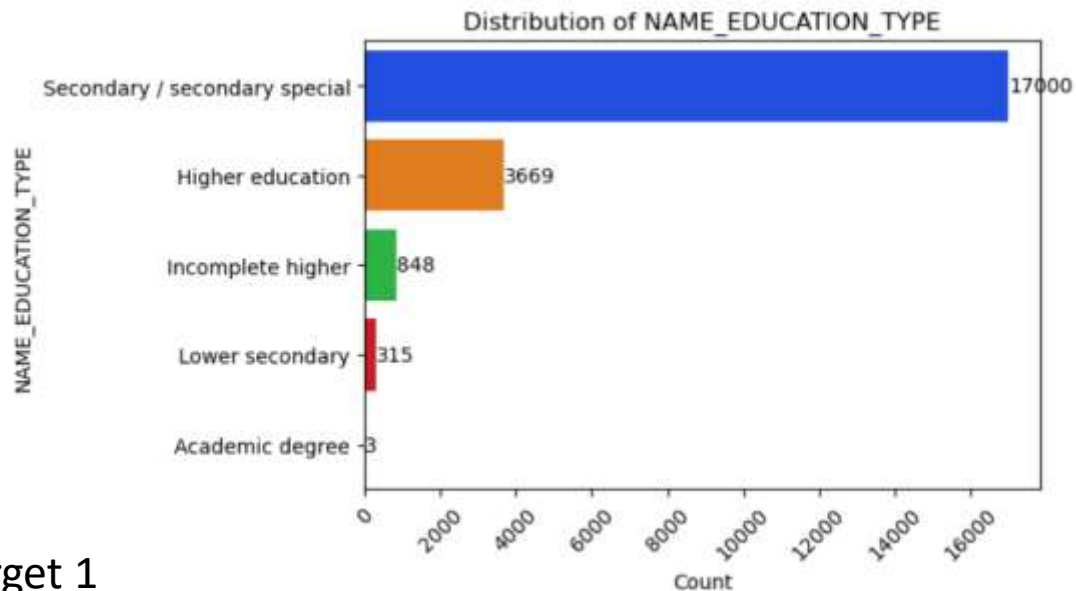


Target 1

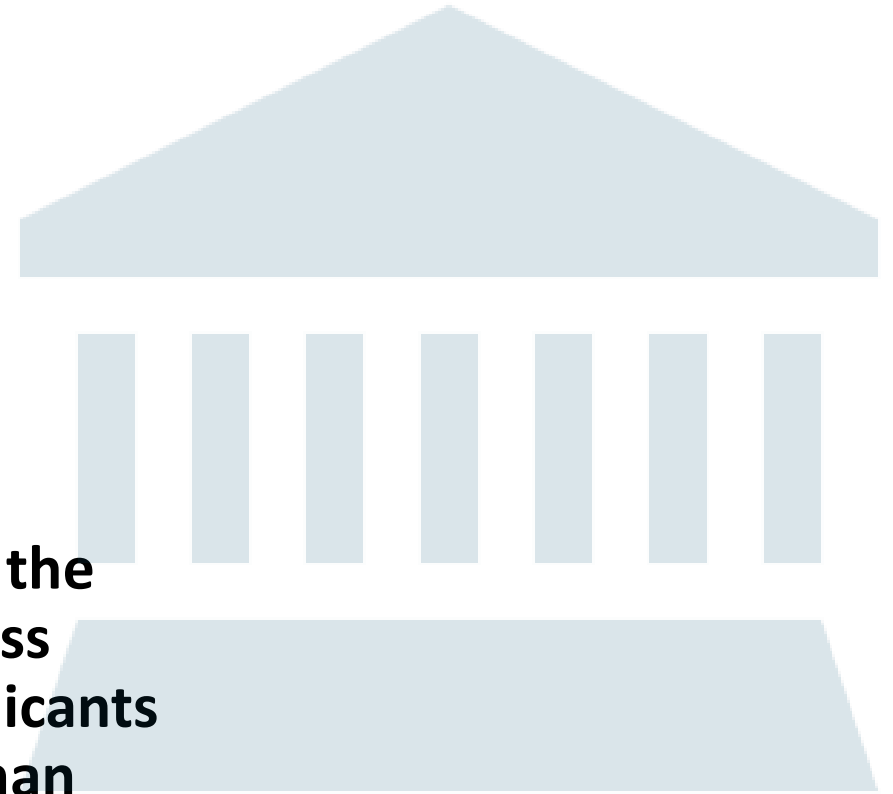


- Most applicants have higher education.
- Applicants with academic degree are least likely to be defaulters.

Target 0



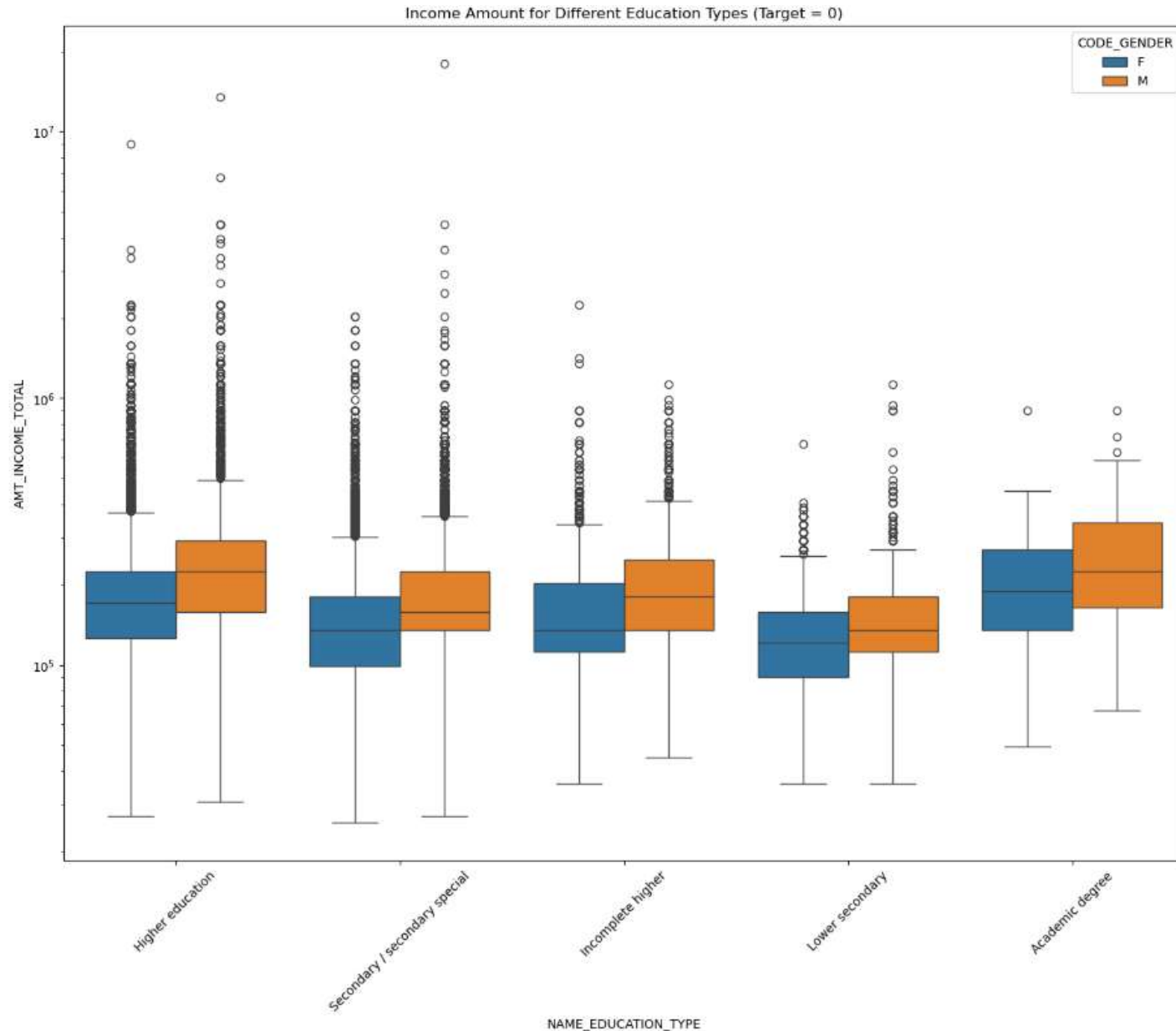
Target 1



It is also observed that around 83% of the overall loan applicants have income less than 240k and around 17% of the applicants have taken loan amounting to more than 1M.

BIVARIATE ANALYSIS W.R.T TARGET VARIABLE

- Total Income vs Education Type
- Credit Amount vs Income Type
- Credit Amount vs Income Range

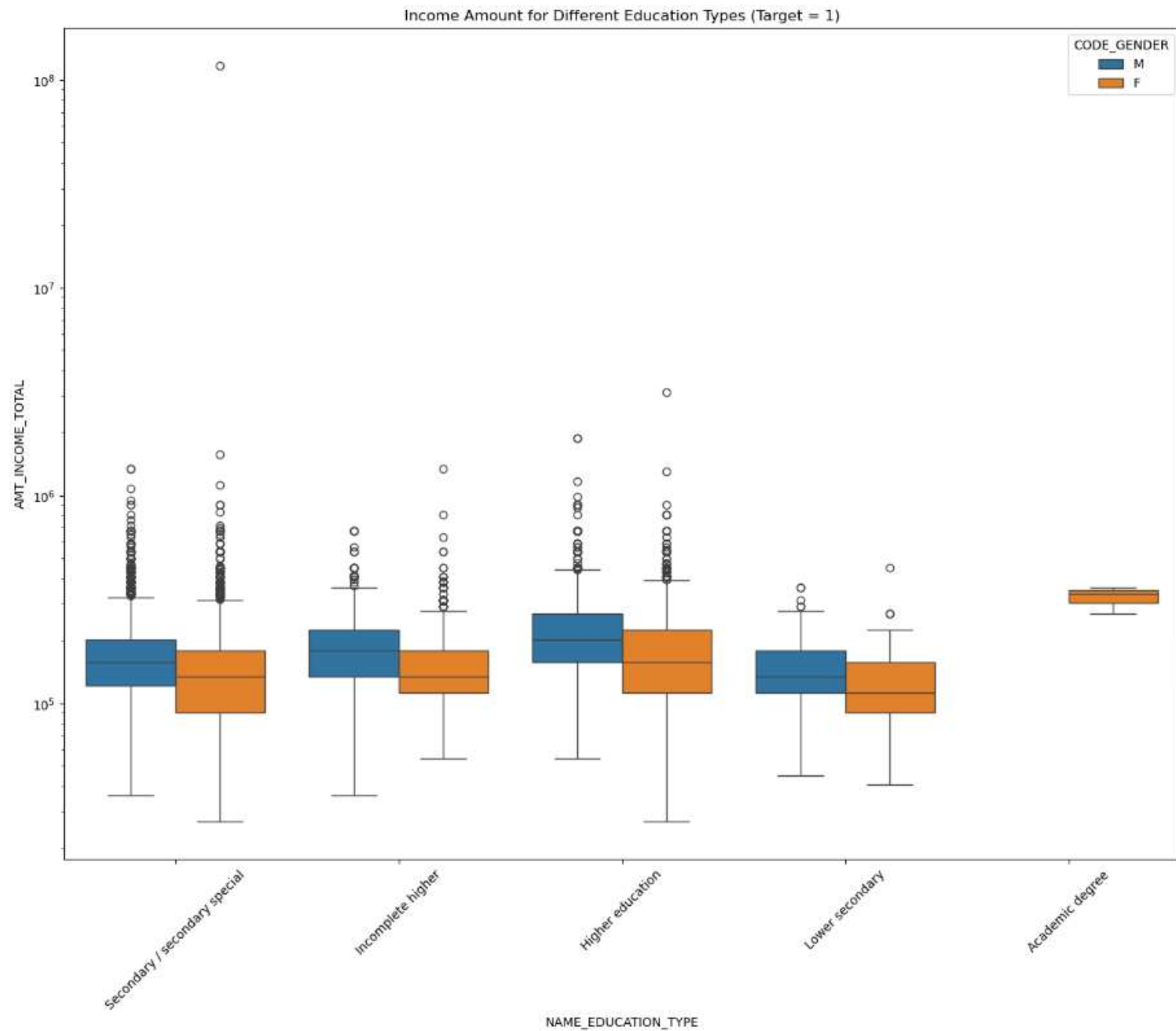


We can observe that most of the education levels appear to have similar median income levels.

Higher Education and Secondary/secondary special categories show high number of outliers compared to other education level.

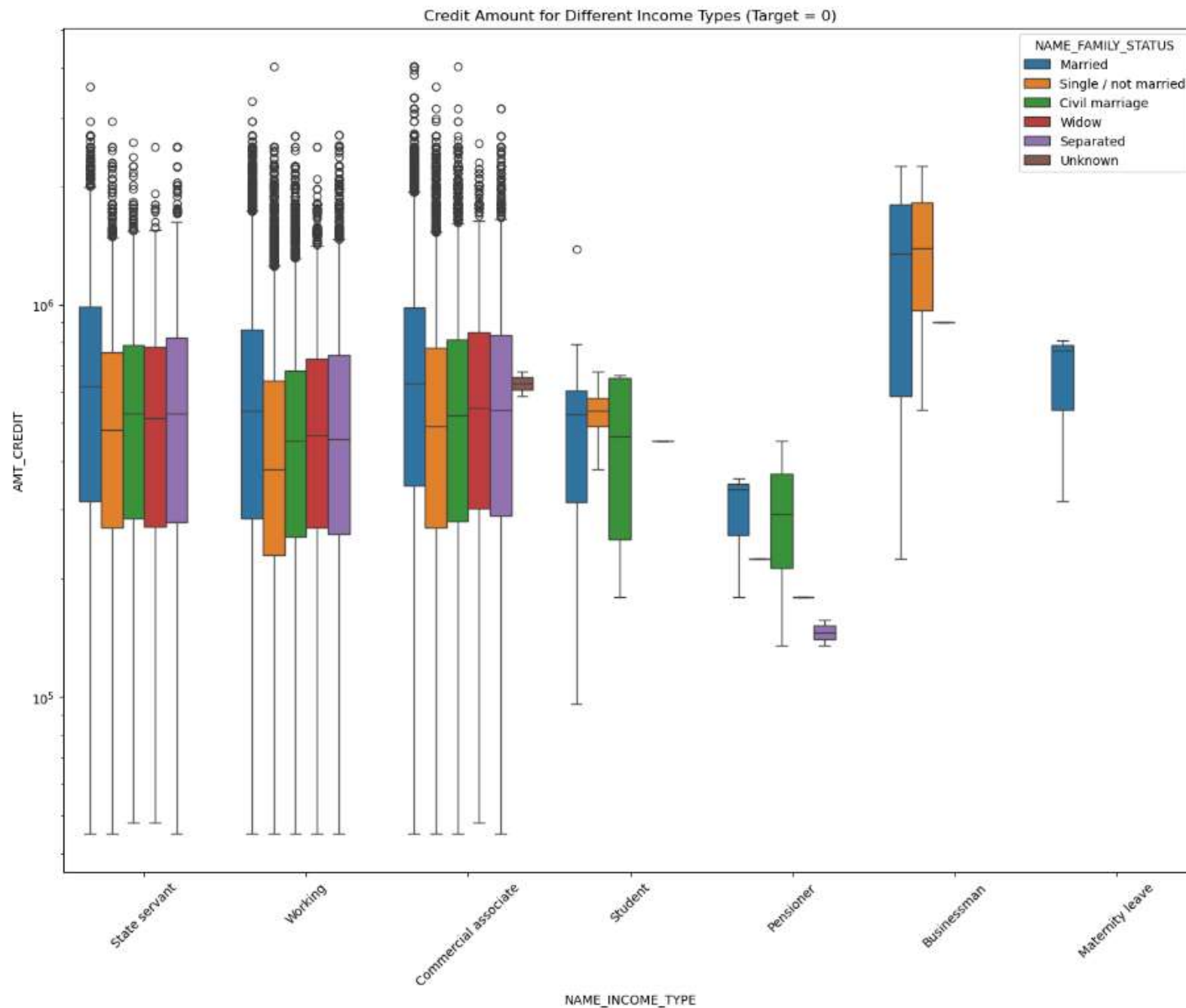
Male applicants across all education levels appear to have higher income.

Male applicants with academic degree have higher median income whereas Female applicants with lower secondary education appear to have lower median income compared to others.



Applicants with lower secondary education appear to have lower income compared to others.

There are large number of outliers especially among male applicants.

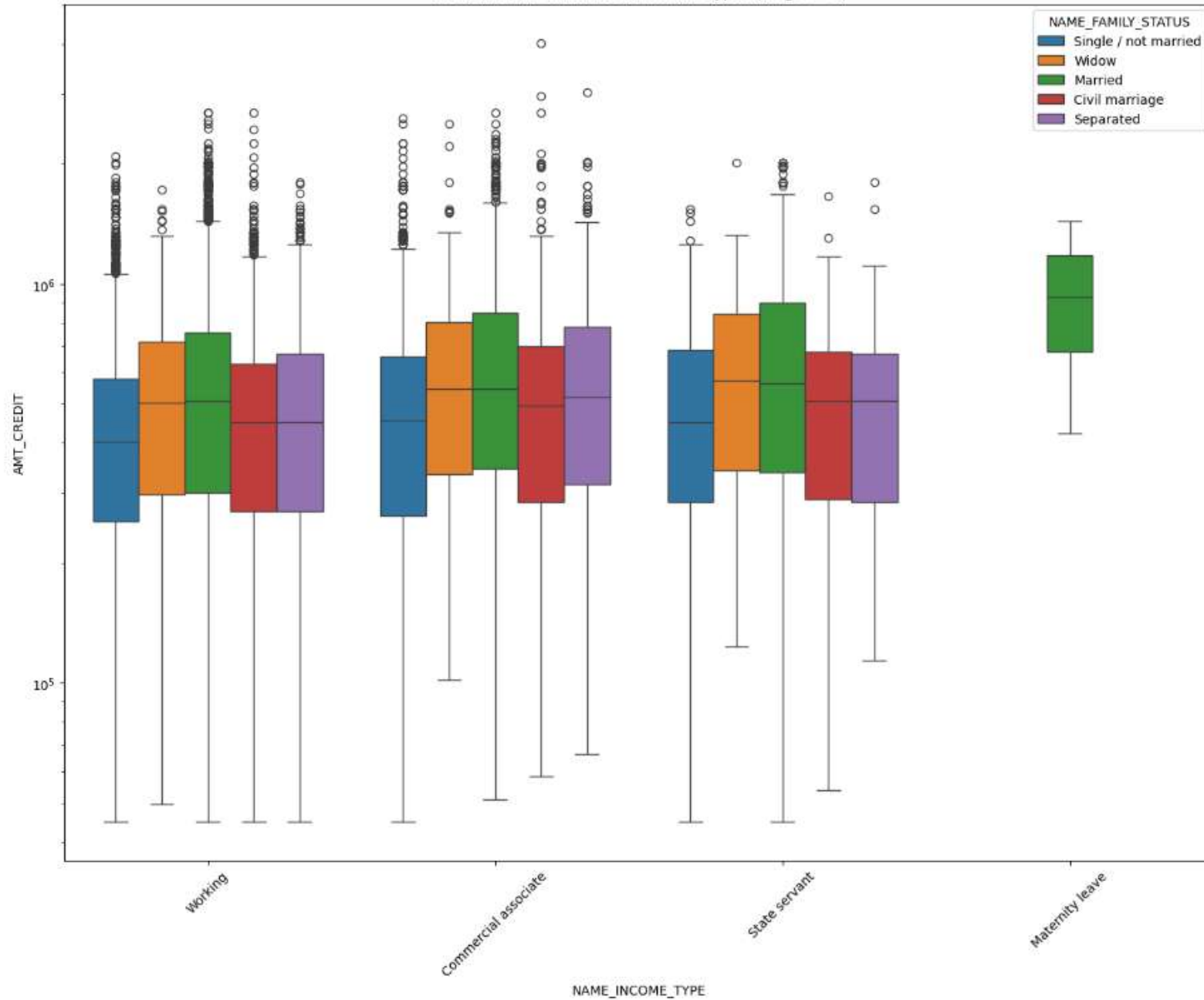


Most of the outliers lie in commercial associate income type. The credit amounts for Working and Commercial associate income types are relatively stable across different family statuses.

Businessmen tend to apply for credit amounts compared to other income types.

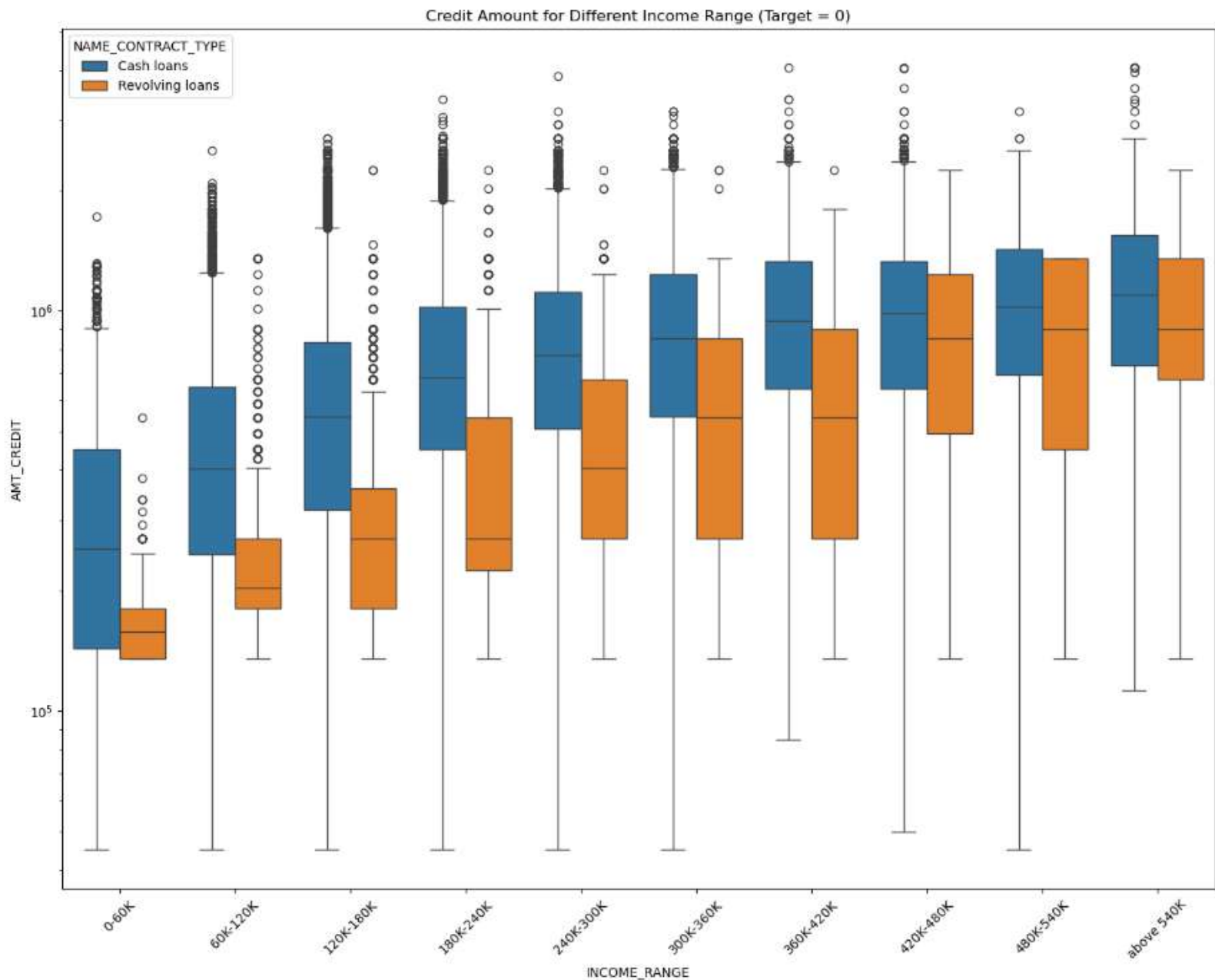
Pensioners and people on maternity leave tend to apply for lower credit amounts.

Credit Amount for Different Income Types (Target = 1)



State servants tend to have higher credit amounts compared to other income types.

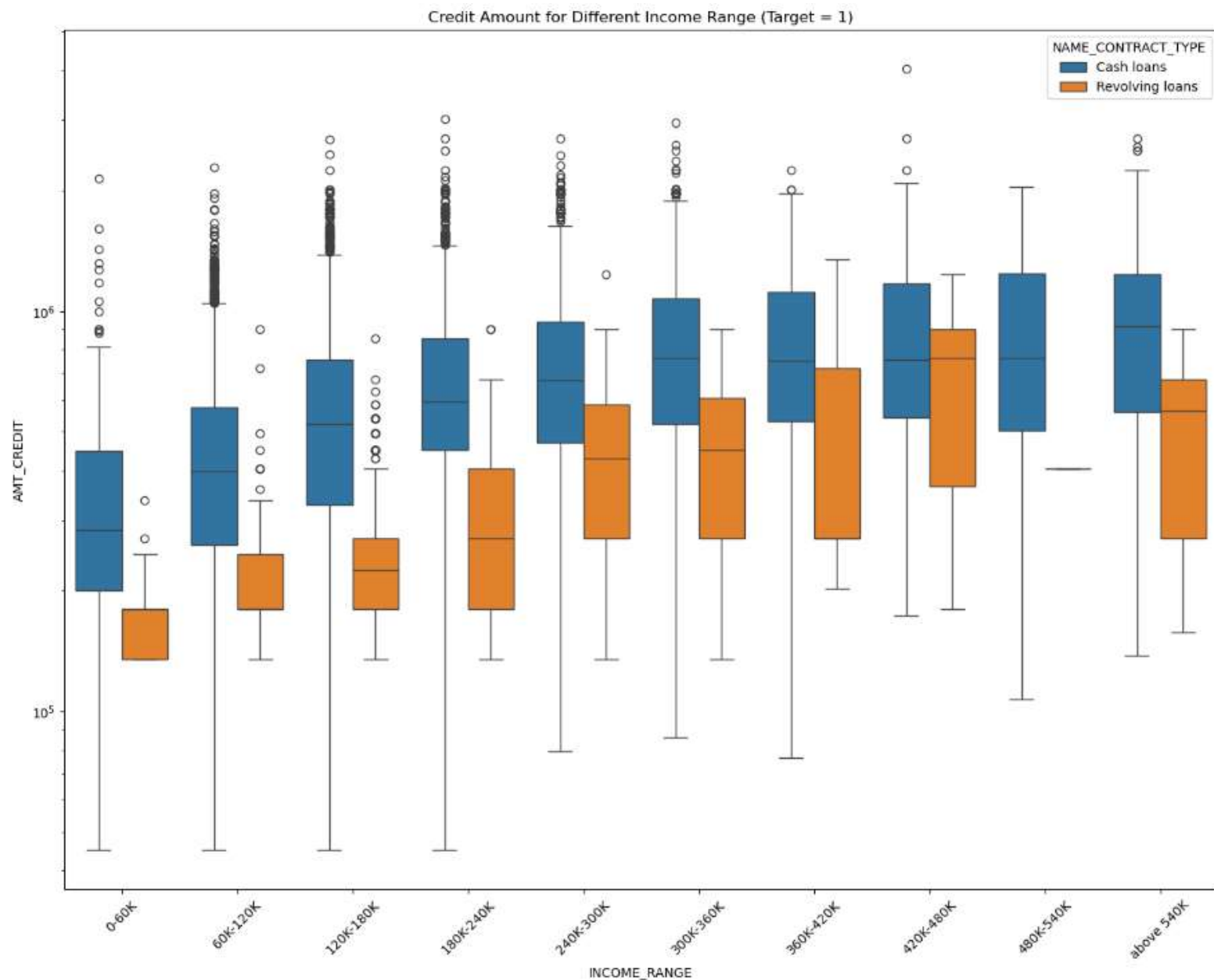
Widows and separated individuals tend to apply for lower credit amounts across all income types.



Revolving loans show relatively stable credit amounts across different income ranges, making them less risky compared to Cash loans.

Clients with higher incomes tend to apply for larger Cash loans.

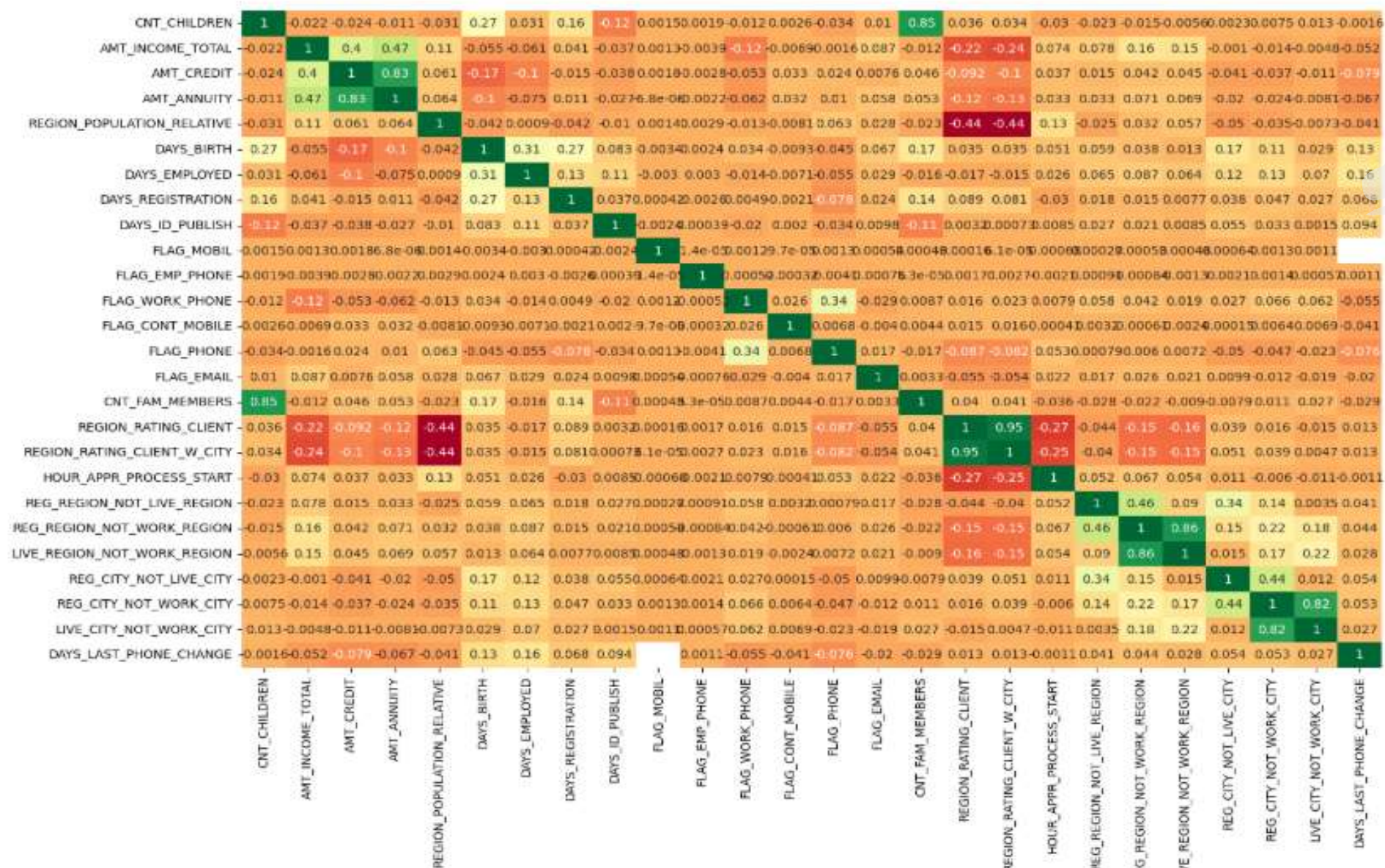
The income ranges of 0-60K and 60K-120K have the lowest credit amounts for both Cash and Revolving loans.



The median credit amounts for Cash loans increase significantly for higher income ranges (e.g., above 360K), but they might still face payment difficulties.

Cash loans generally have higher credit amounts compared to Revolving loans across all income ranges.

Correlation for target 0





Amount of credit is inversely proportional to the date of birth of applicant.



Total income amount is higher in areas with large population whereas amount of annuity is higher for high credit amount .



Region rating for client is inversely proportional to the relative population of a region. It means higher the rating lower the population.

Correlation for target 1





Total income is inversely proportional to the region rating of client taking city into consideration.

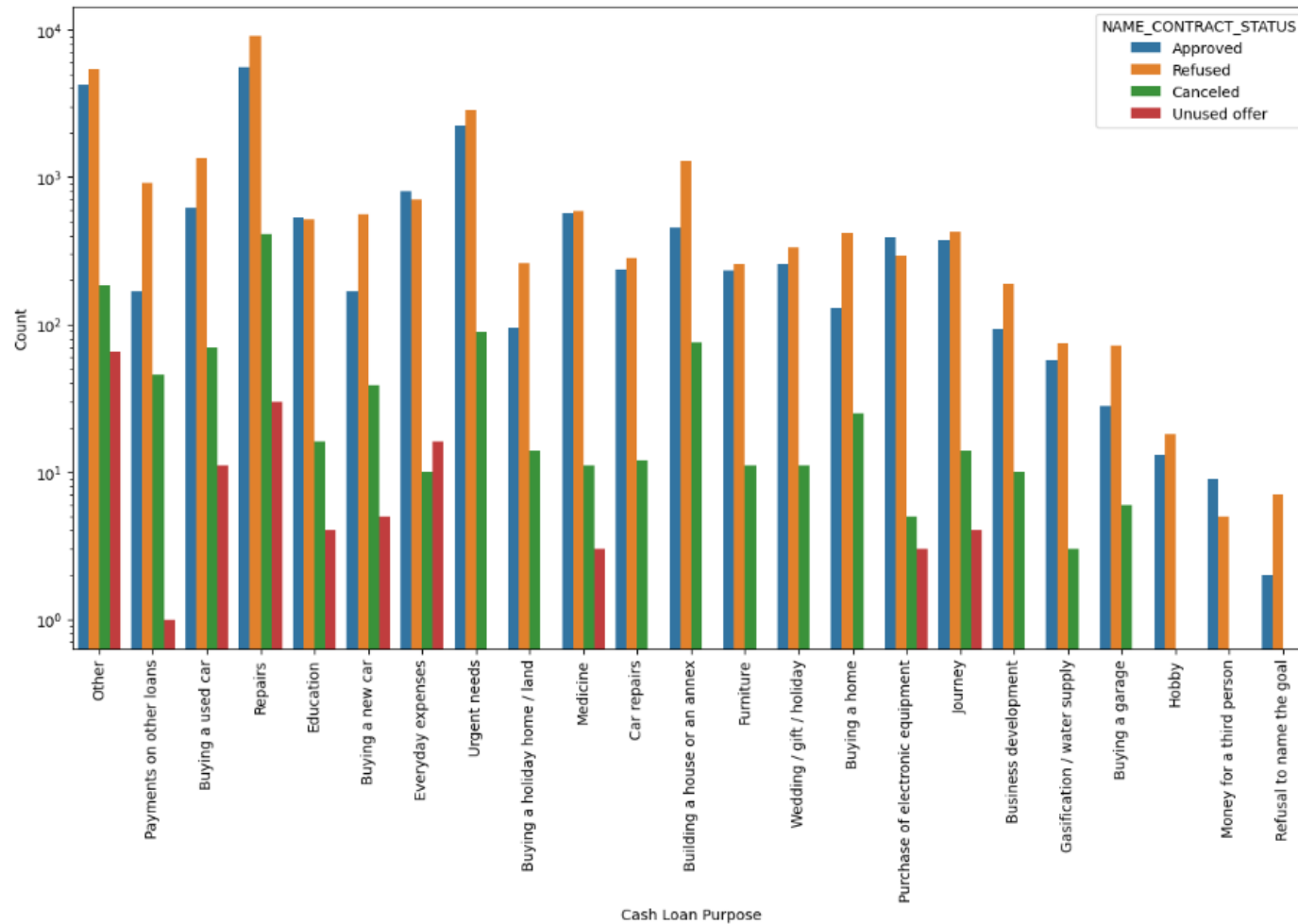


Credit amount is directly proportional to the total amount of annuity and inversely proportional to the date of birth.



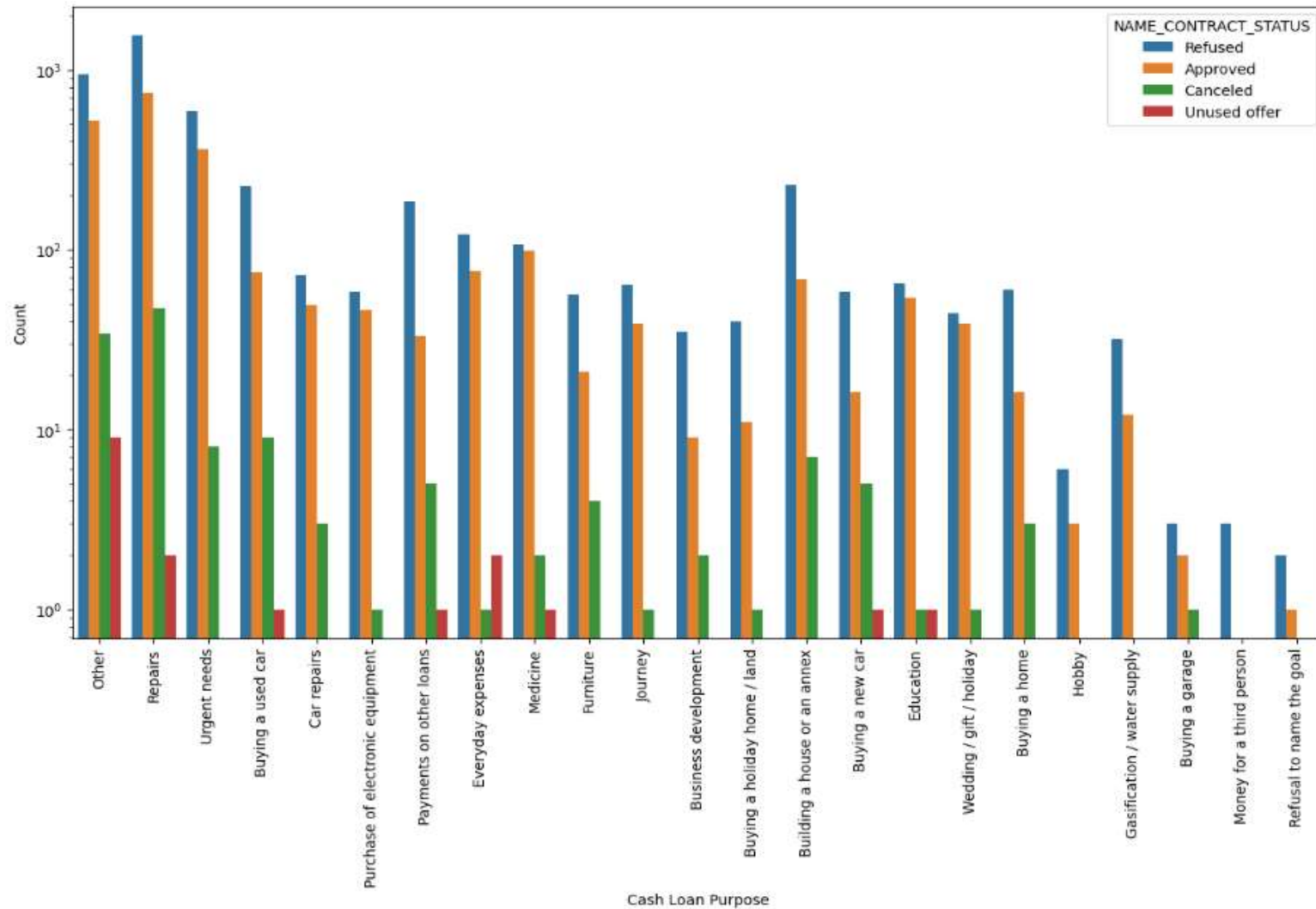
Region rating of a client is inversely proportional to the approx. hour since the process started .

Cash Loan Purpose vs Contract Status for Target 0



- The category 'repairs' has a high count of refused loans whereas the least amount of approved loans are for 'refusal to name the goal' category.
- 'Payment on other loans' category have highest difference between refused vs approved.

Cash Loan Purpose vs Contract Status for Target 1

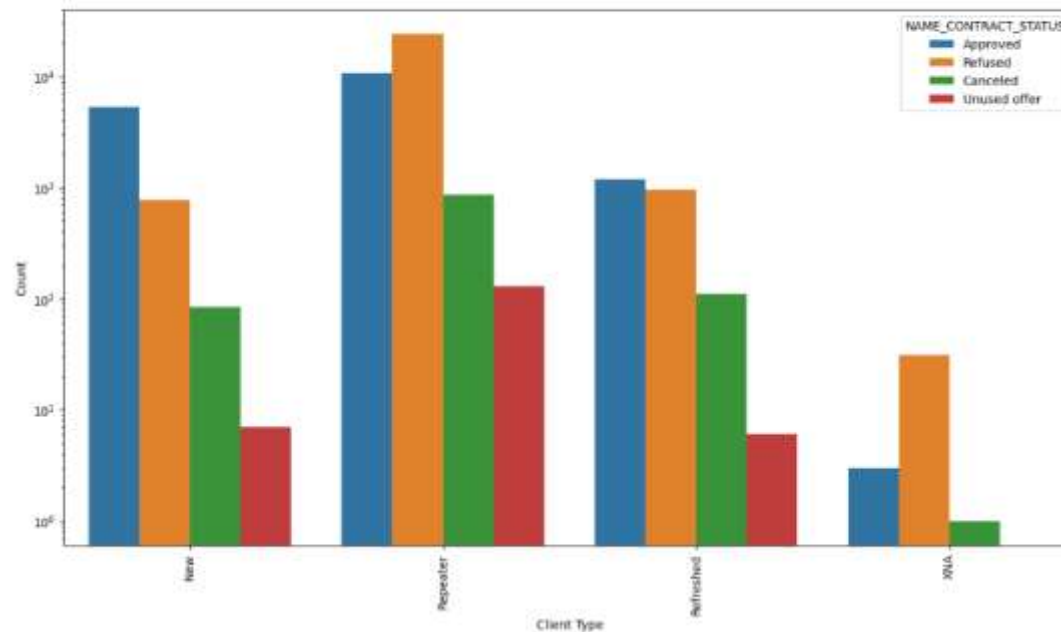


- The number of potential defaulters are highest for repair.
- A high number of payment of other loans and buying a holiday home/land seems to have been rejected by bank.

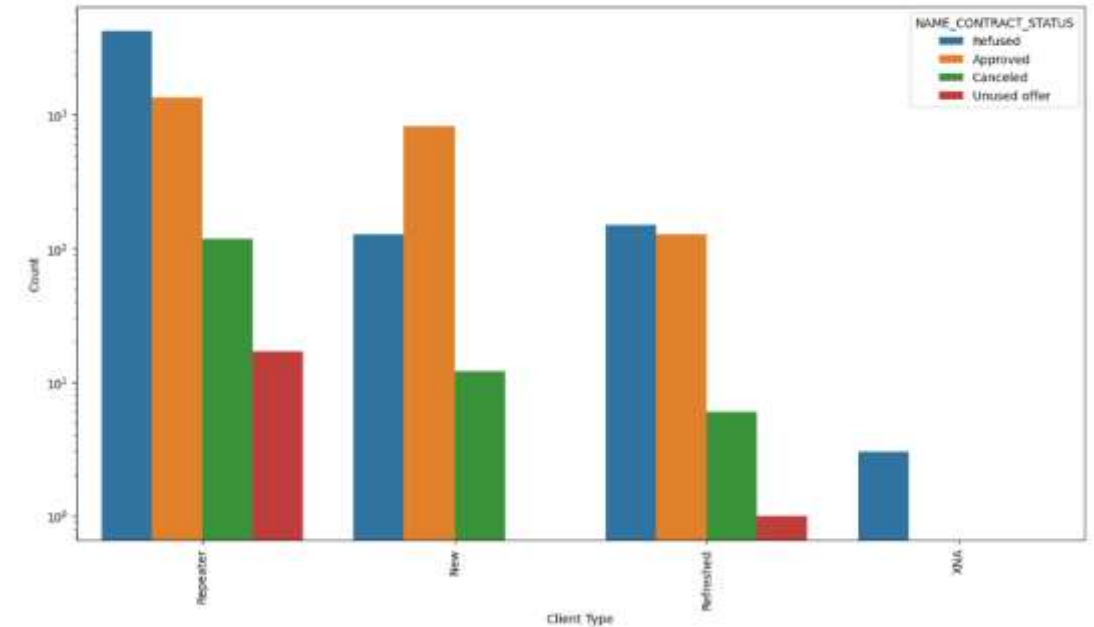
New clients have high amount of approved loans compared to repeaters.
There is significant difference between no of approved vs refused loans for new clients for both variables.
There are large number of unused offers for repeaters for both target variables.



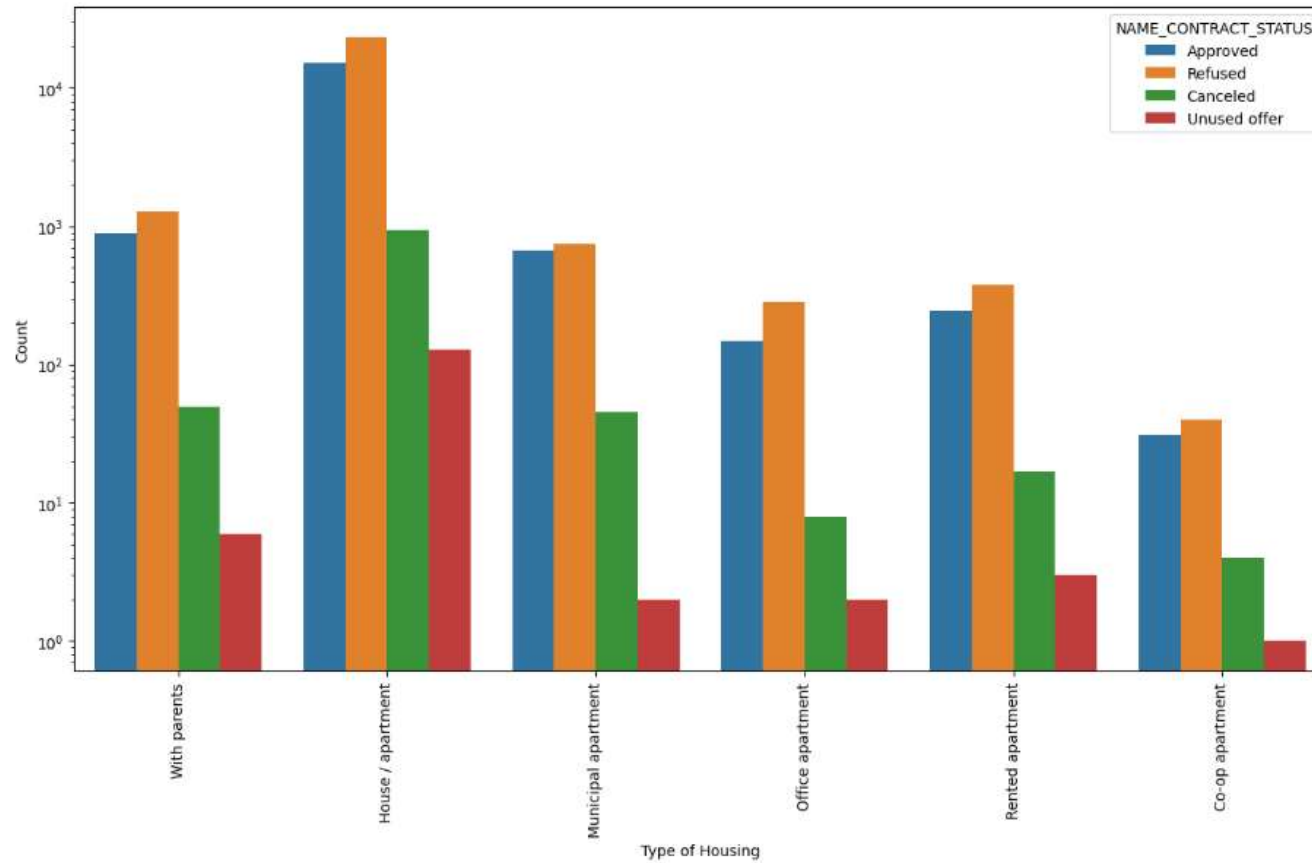
Client Type vs Contract Status for Target 0



Client Type vs Contract Status for Target 1

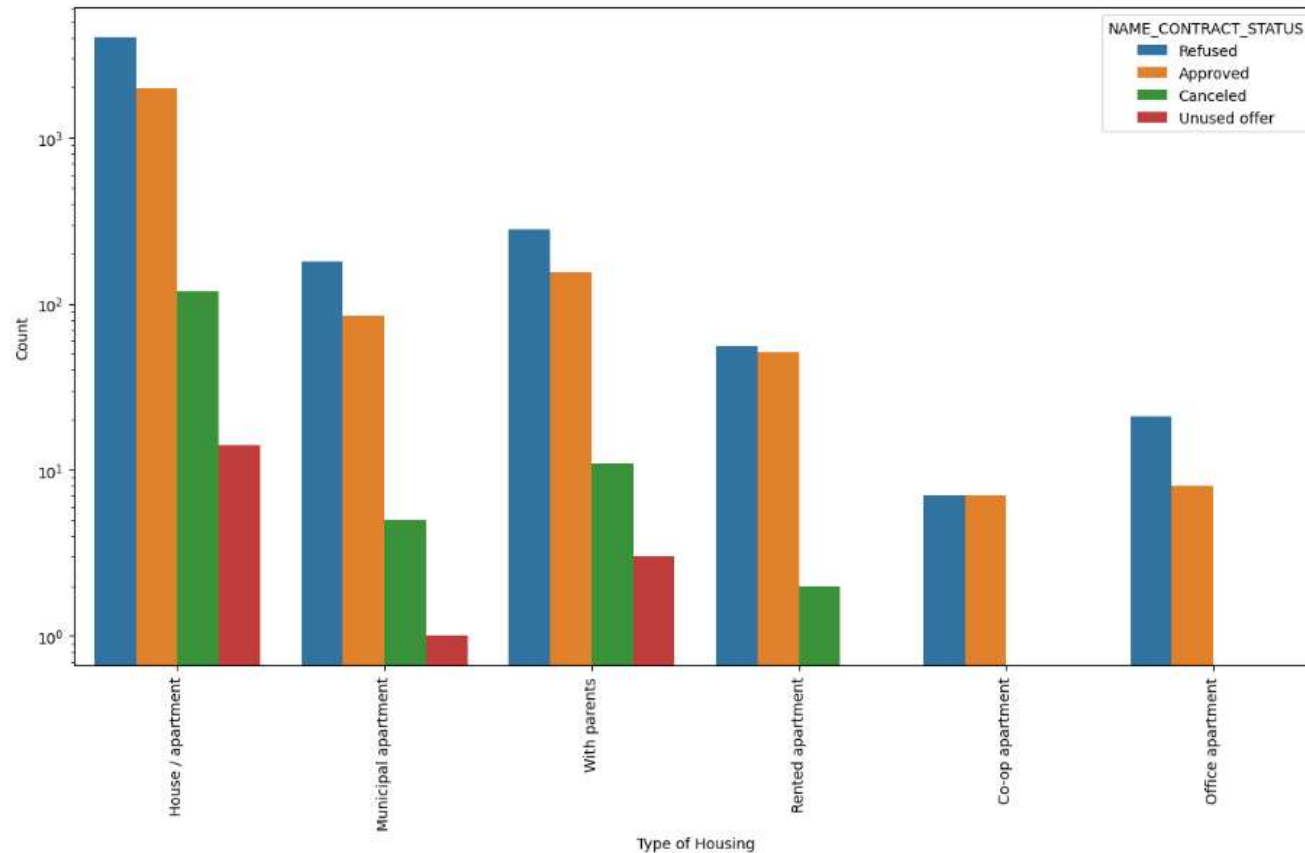


Housing type vs Contract Status for Target 0



- The house/apartment category has both highest number of approved loans as well as refused loans.
- Municipal apartment has the least difference between approved vs refused loans.

Housing type vs Contract Status for Target 1



The house/apartment category has both highest number of approved loans as well as refused loans.

The co-op apartment category has equal amount of approved vs refused loans.

The house/apartment for both target variables have large amounts of unused offers compared to other housing type.

IMPORTANT INFERENCES:

Students, Pensioners, and businessman can be considered as good applicants as they are not likely to face payment difficulties.

New clients can be targeted as they have higher no. of approved loans .

Municipal apartment has least no. difference between approved and refused loans for clients with strong repaying capacity and can be considered by the bank.

Bank should give less focus to loan purpose 'repair' as it has high count of refused loans as well as applications by potential defaulters.

Bank should also give less priority to payment on other loans and buying a holiday home/land which seems to have high number of rejections.

THE END