CREDIT EDA ASSIGNMENT

Presented by – Ayush Das
Batch – DS-C62

Problem Statement

Business Understanding

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 to their advantage by becoming a defaulter. Suppose you work for a consumer finance company which specialises in lending various types of loans to urban customers. You have to use EDA to analyse 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.

Problem Statement

Business Objectives

This case study aims to identify patterns which indicate if a client has difficulty paying their instalments 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 utilise 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.

Assumptions and Limitations

- Assumes data used is accurate and complete
- Variables used can predict creditworthiness
- Assumes statistical properties of data remain constant over time
- Sometimes the information isn’t about everyone, just some people
- Might not have enough old data, especially for new people

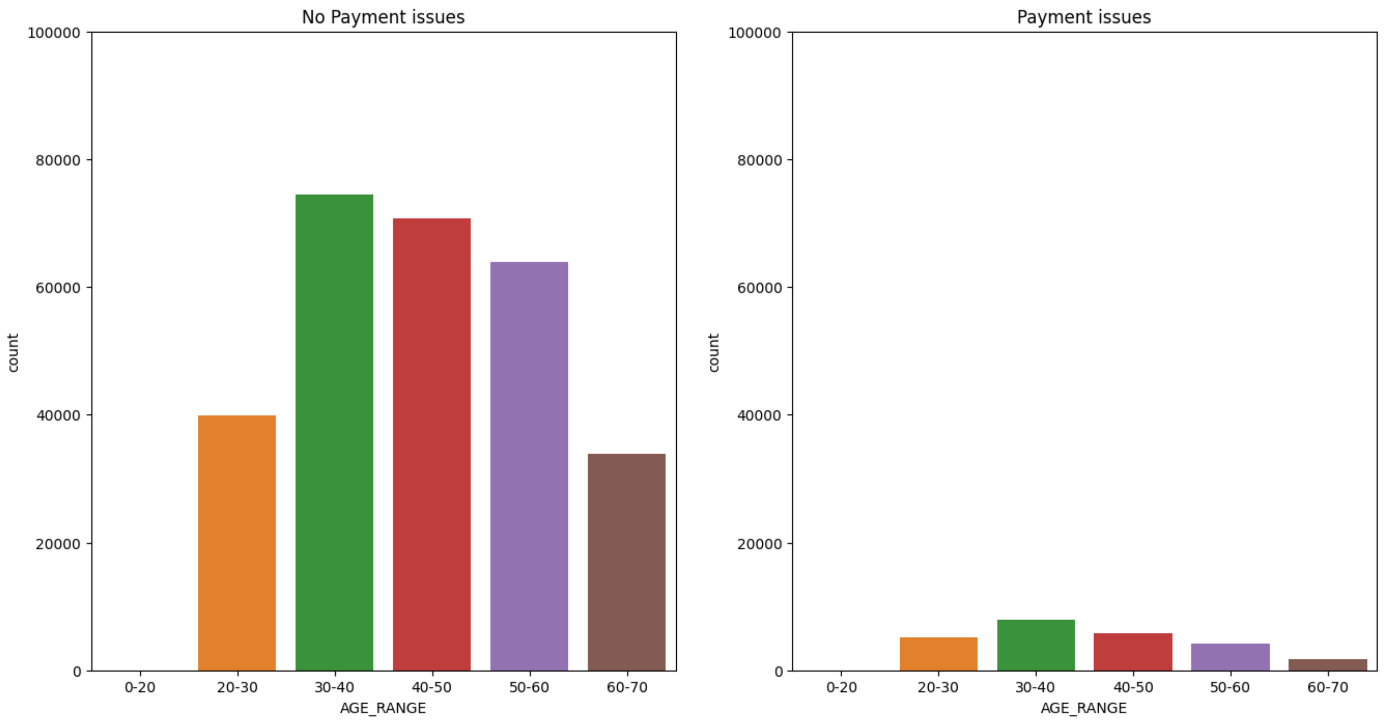
Approach and Methodologies

- Started off by importing various libraries for data analysis as well as data visualization
- Loaded "application_data.csv" data set and created a DataFrame out of it for further analysis
- Various pandas in-built function was used to have an inspection of the DataFrame
- Columns with null percentage value less than 40% were dropped as it would be irrelevant for those columns to study upon
- Treatment of missing values were done as follows:
  1. For numerical columns, null values were replaced by median as mean is affected by the outliers
  2. For categorical columns, null values were replaced by mode

Approach and Methodologies

- Outliers were identified and dealt accordingly
- Univariate/Bivariate analysis was done to find the insights
- “previous_application.csv” was loaded and same tasks were performed in this dataset too
- Finally, both the datasets were merged to get detailed insights.

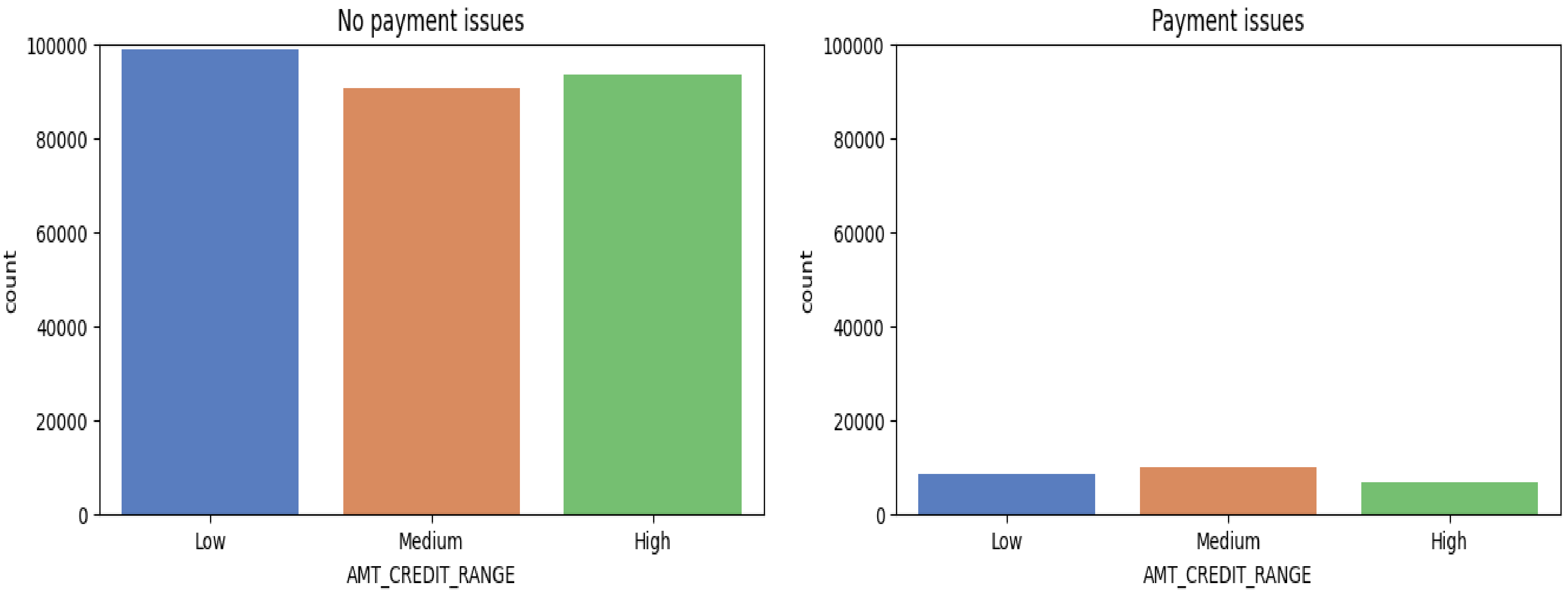
Graphs and Insights



Conclusion –

- We can observe that customers belonging to age group 30-40 are able to make payment on time and can be considered while lending loan!
- The customers from 40 to 60 age are also can be considered

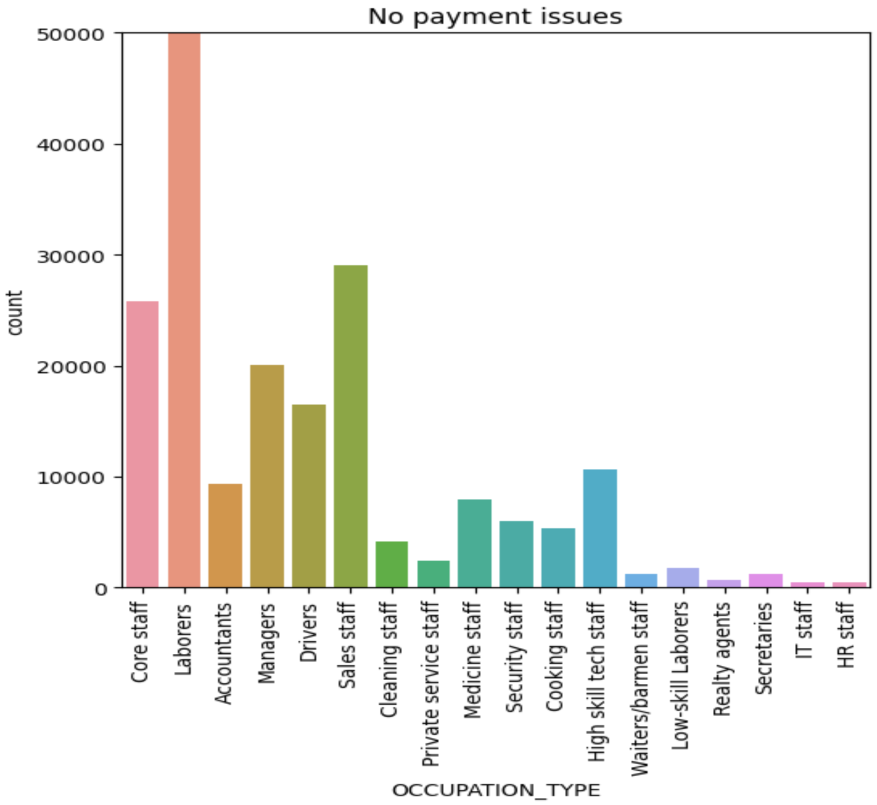
Graphs and Insights

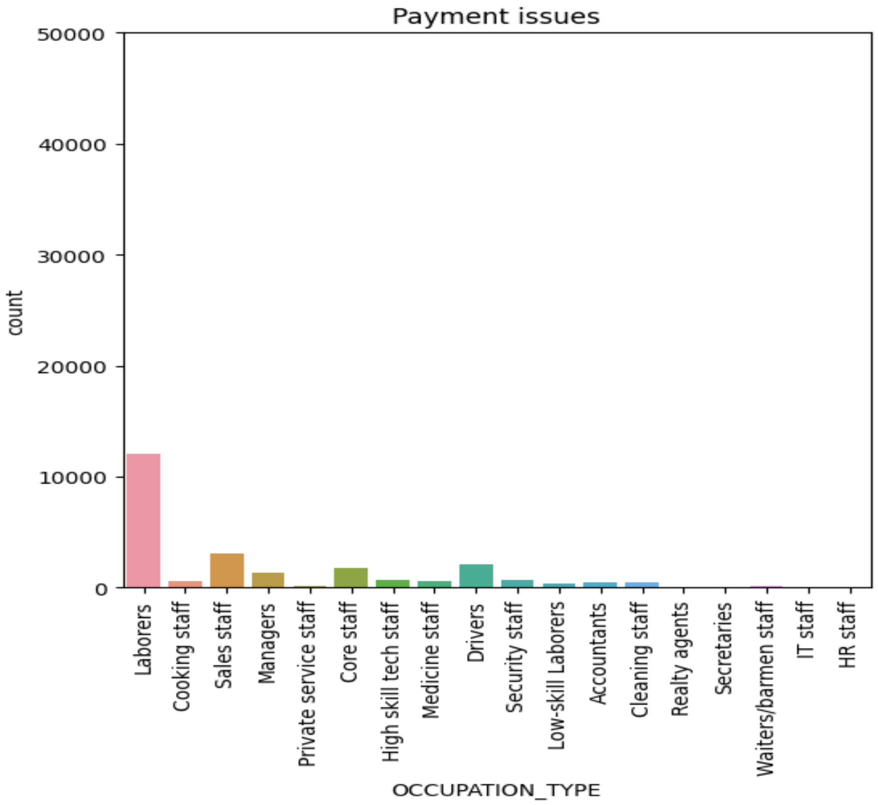


Conclusion –

- Customers with less credit and most likely to make payment
- Customers having medium and high credit can also be considered while lending the loan

Graphs and Insights

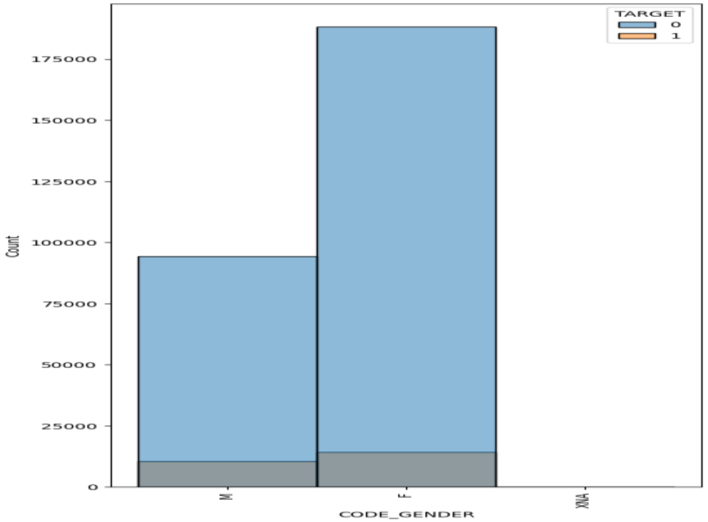


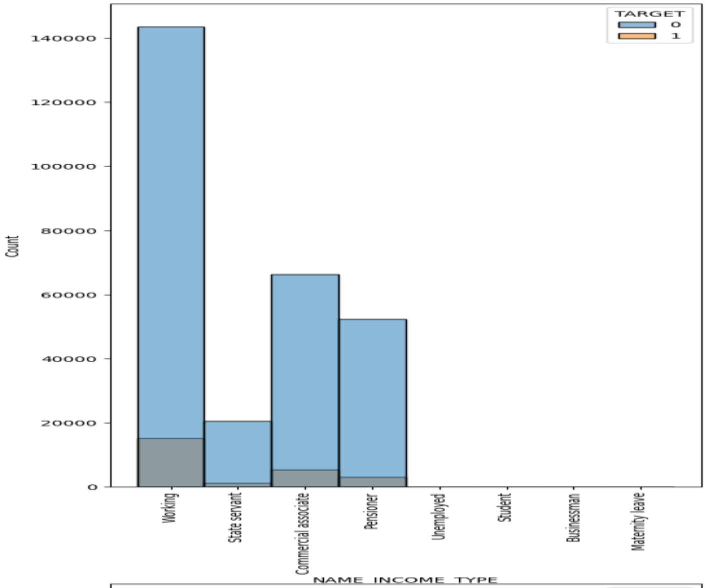


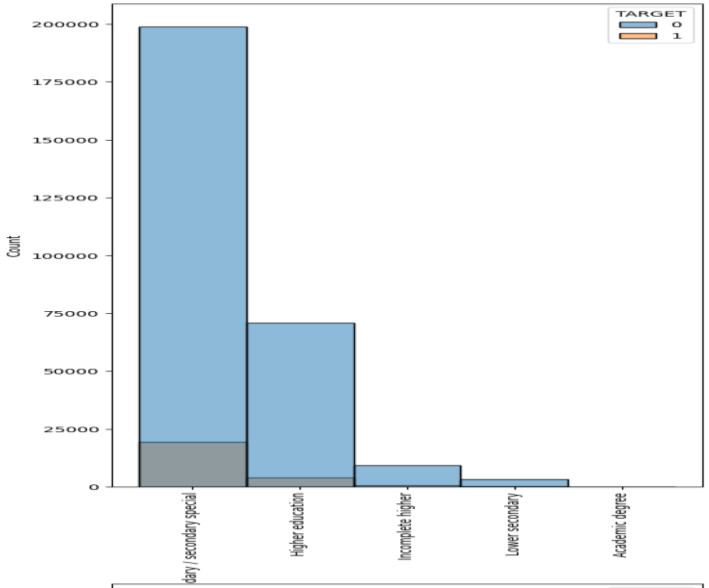
Conclusion:

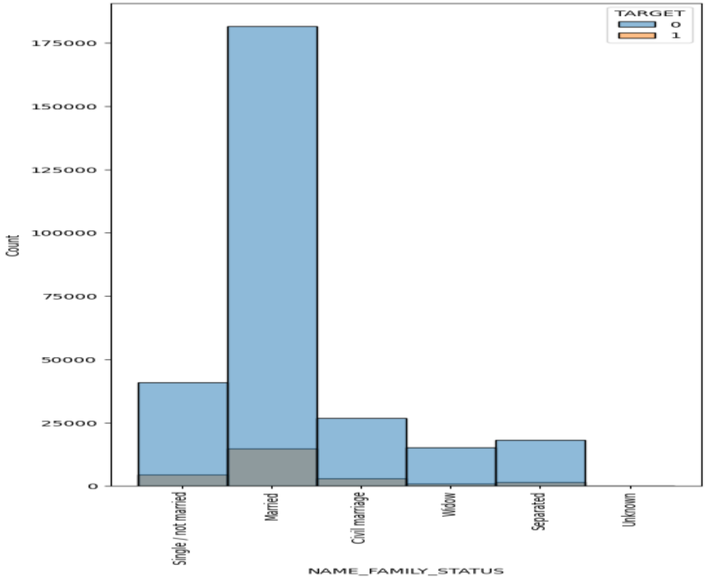
- The plot clearly shows that labourers are most likely to make payment on time whereas HR staff are less likely to make payment on time

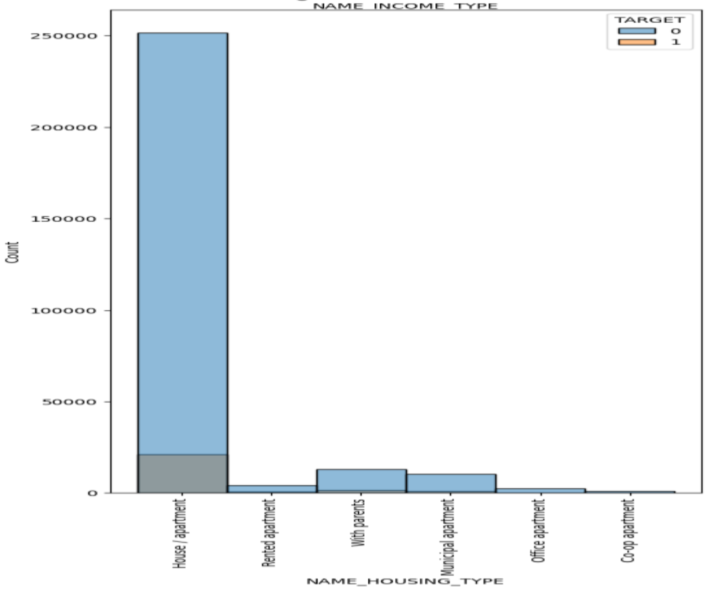
Graphs and Insights

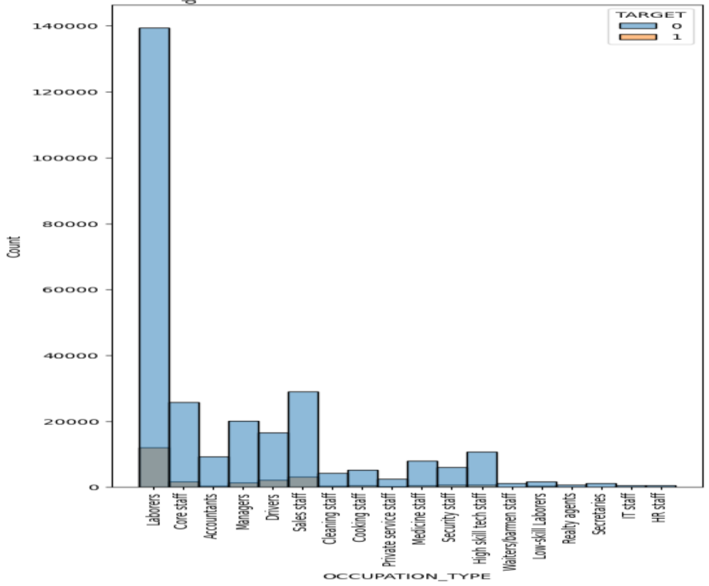










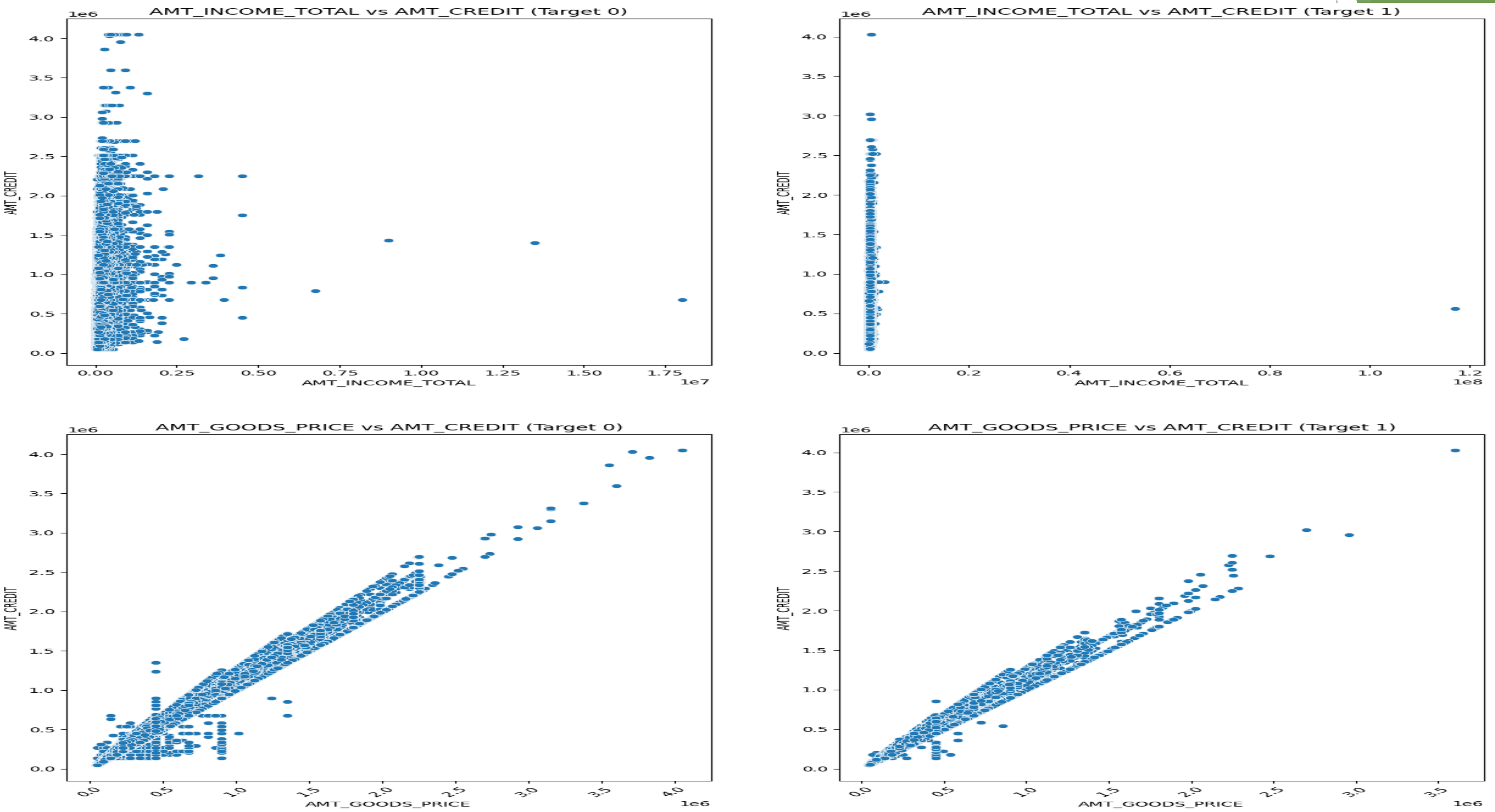


Graphs and Insights

Conclusion From The Above Plots:

- Female customers pay loan amount on time and banks can target more female customers for lending loan
- Working customers can be targeted to lend loans as they have higher percentage of making payments on time
- Customers with secondary education are most likely to make payments when compared to customers with academic degree
- Married customers have paid loan amount on time when compared to widows
- Customers owning House/apartment are most likely to make payments on time compared to those living in CO-OP apartment
- Labourers have high repayment percentage. Hence banks can think of lending small amount loans to them

Graphs and Insights

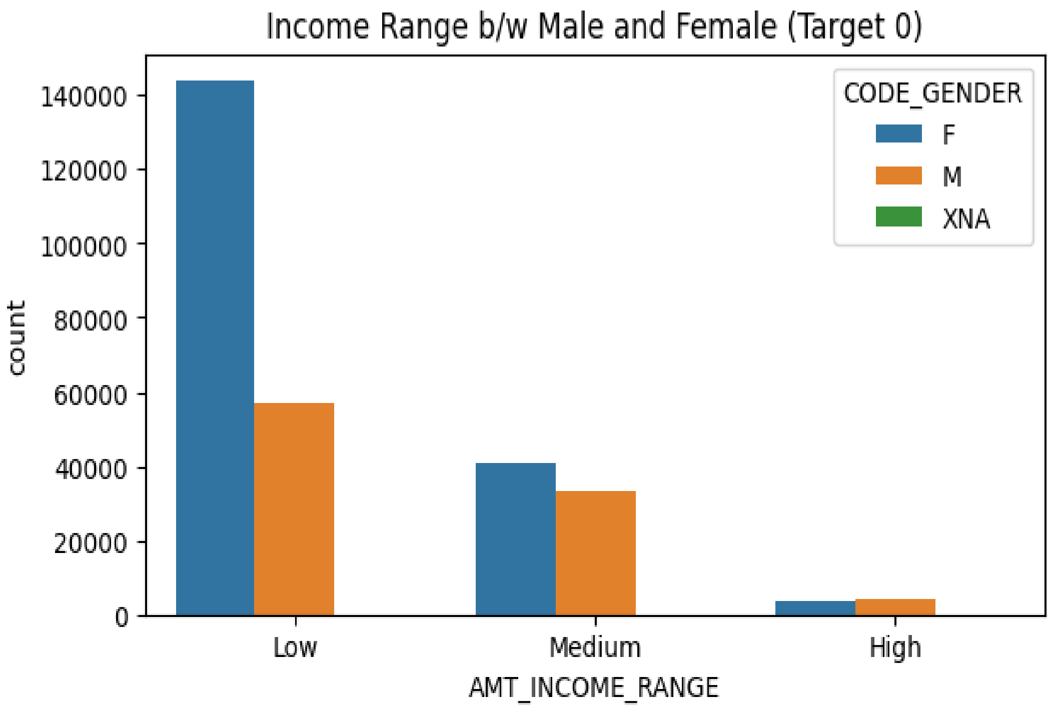


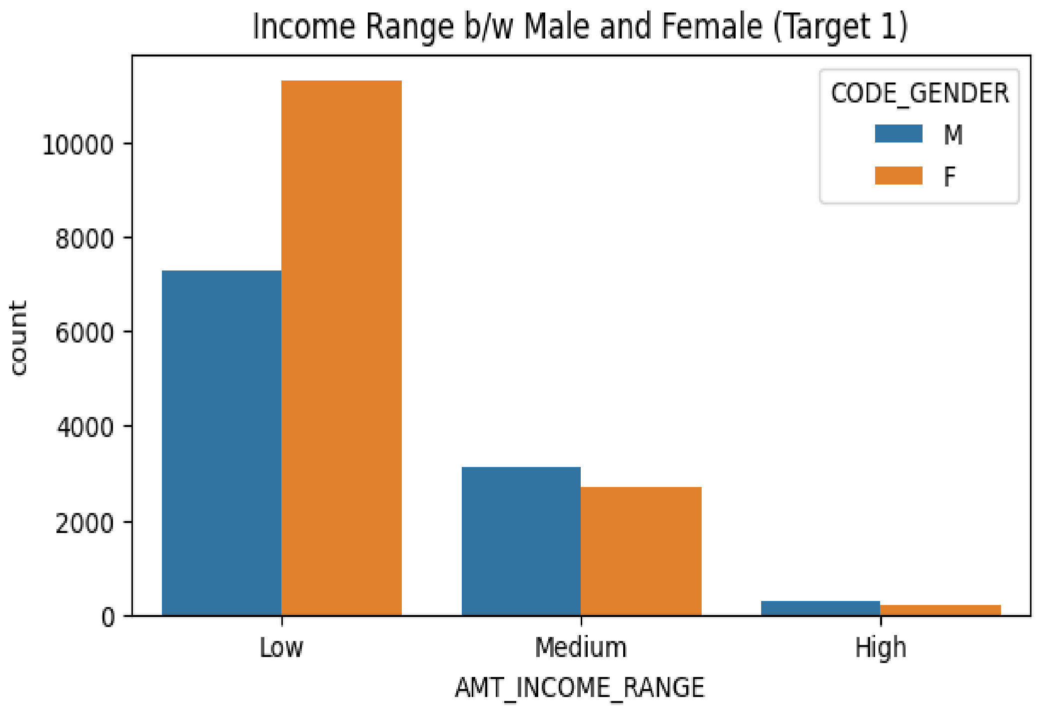
Graphs and Insights

Conclusion From The Above Plots:

- Those who have paid the loan amount on/within time are more likely to get higher credits than those who didn’t pay/did late payments
- People who have higher goods price and have made payments on time have higher credits than those with higher goods price but didn’t pay loan

Graphs and Insights

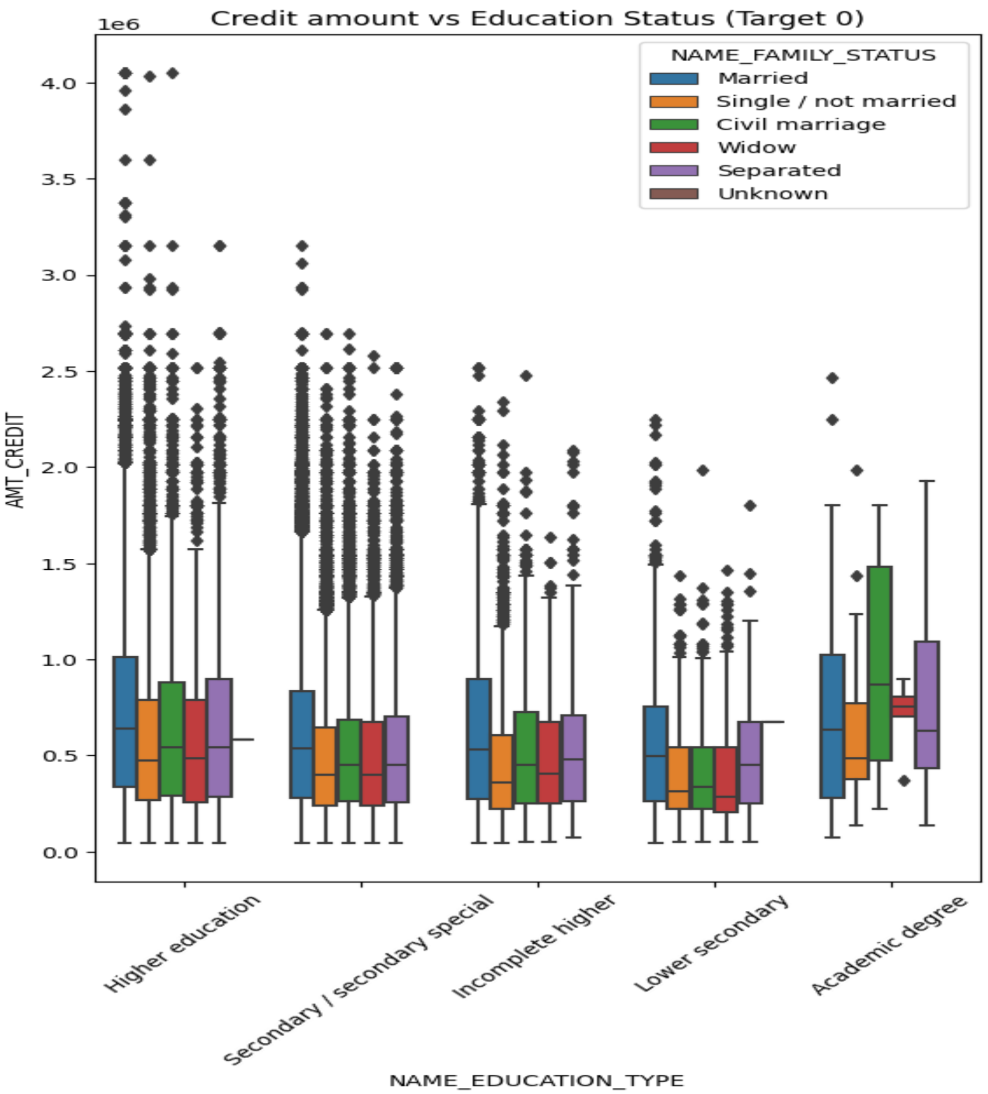


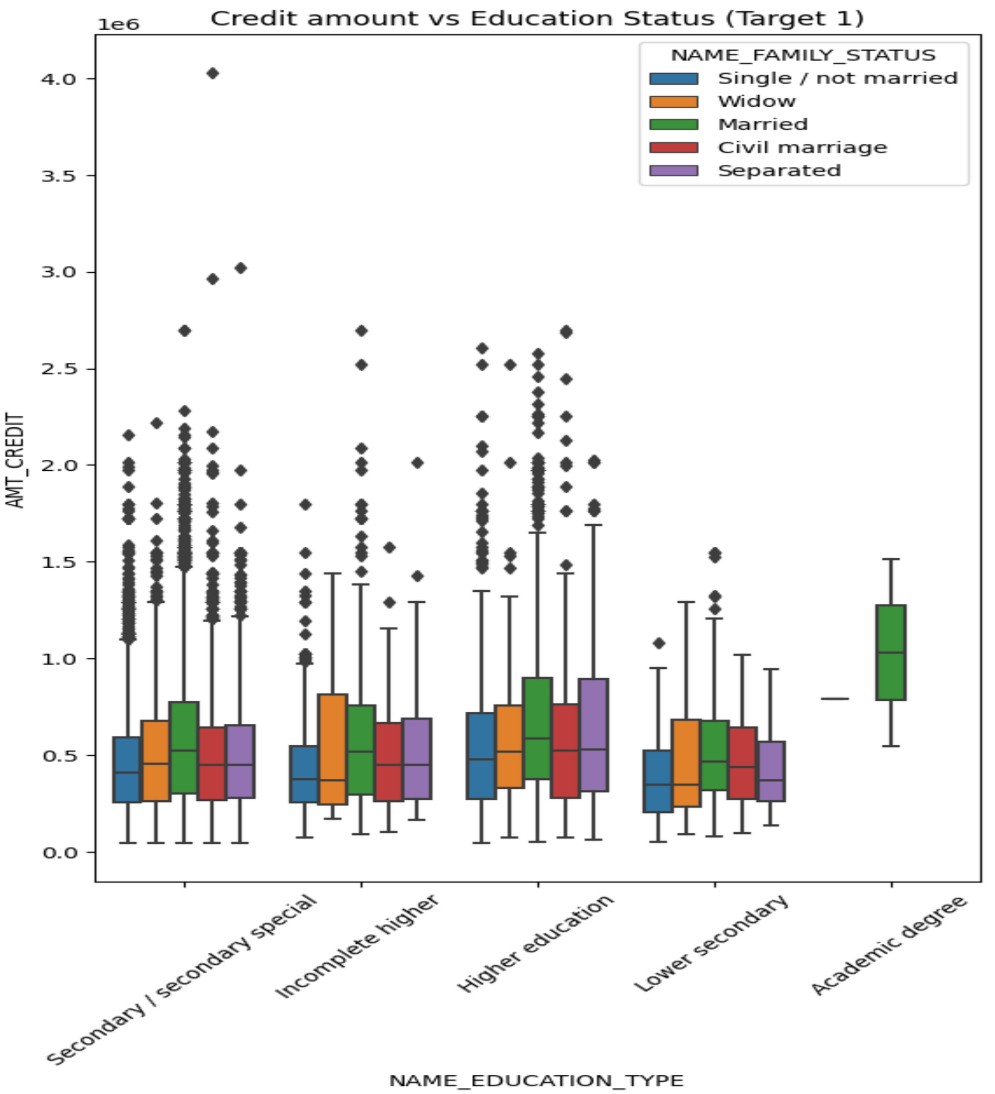


Conclusion -

- We can see that Females with low income don’t have any payment issues

Graphs and Insights





Graphs and Insights

Conclusion From The Above Plots:

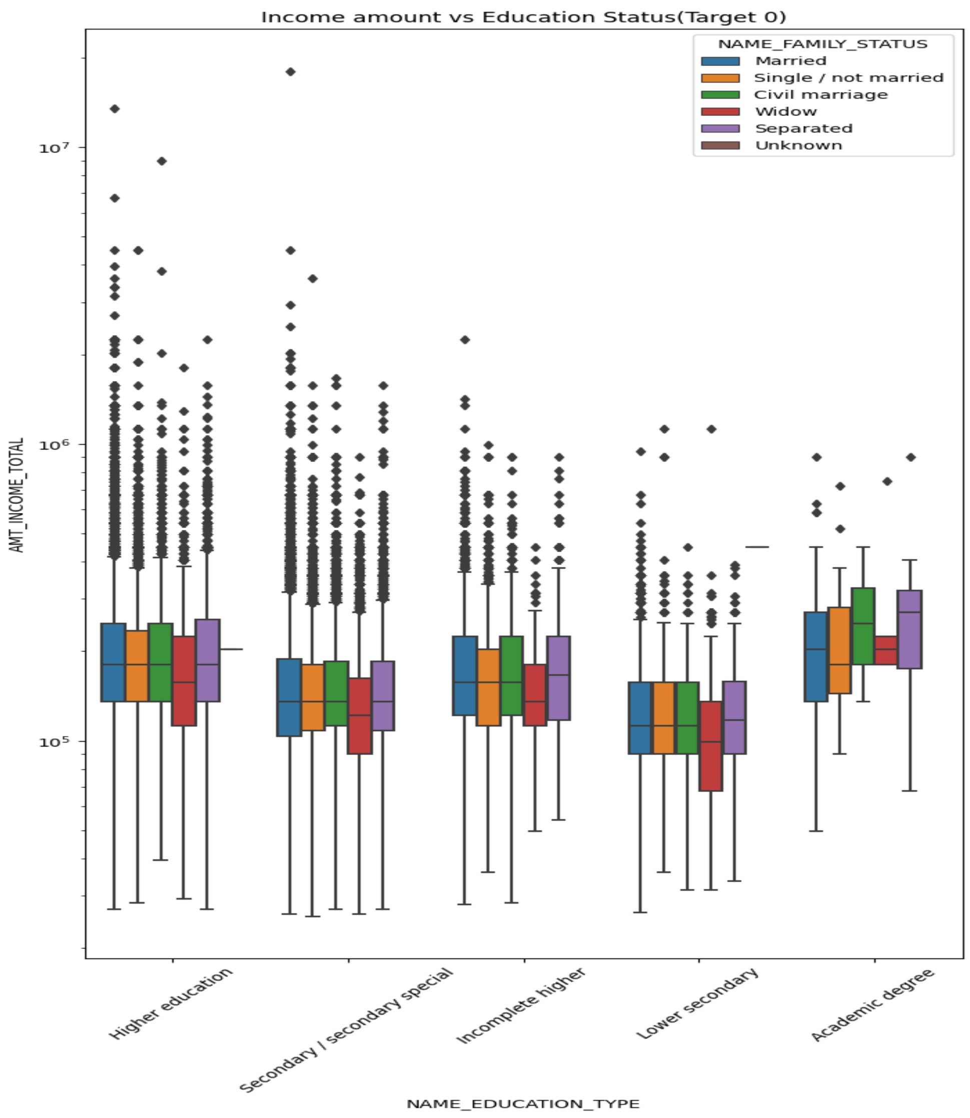
- Some of the highly educated, married person are having credits higher than those who have done lower secondary education

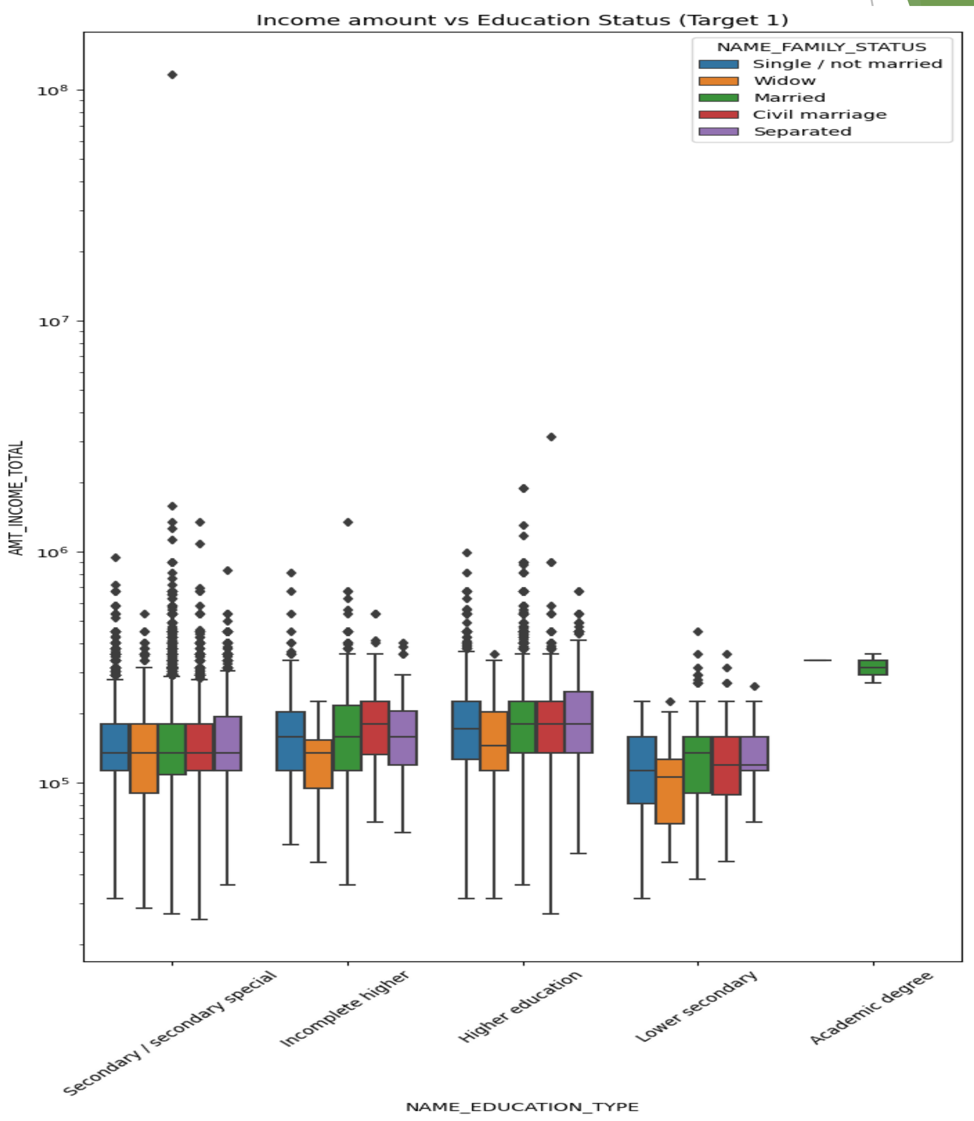
- Those with higher education have higher credits and are more likely to make payments on time

- More number of outliers are seen in higher education

- The people with secondary and secondary special education are less likely to make payments on time

Graphs and Insights



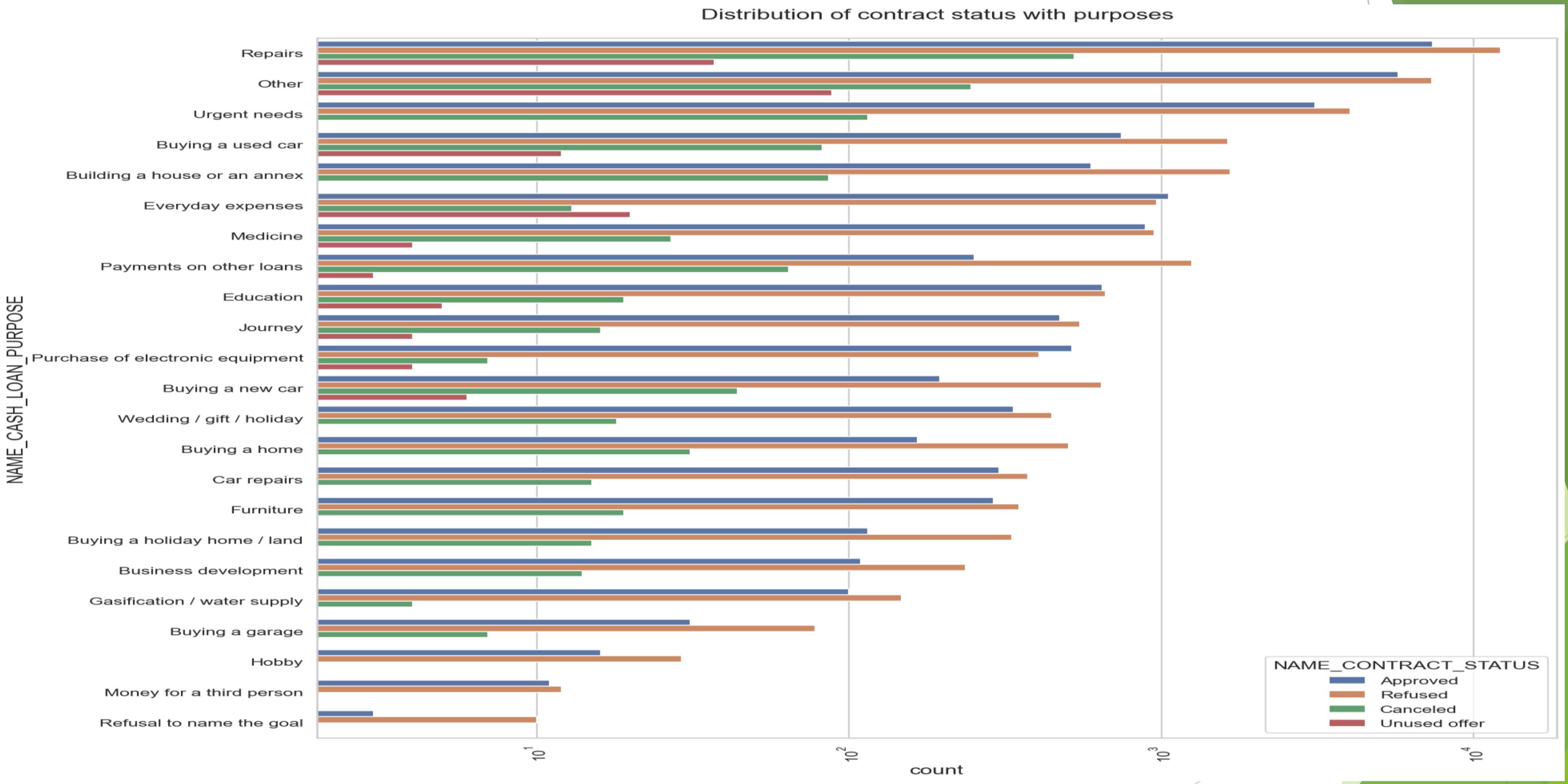


Graphs and Insights

Conclusion From The Above Plots:

- We can see that Higher education has many outliers
- People with higher education have higher income and dont have difficulties in making loan payment
- People with higher education who ave lesser income are unable to pay the loan
- Hence we can conclude that, people with higher income are most likely to make payments.

Graphs and Insights



Graphs and Insights

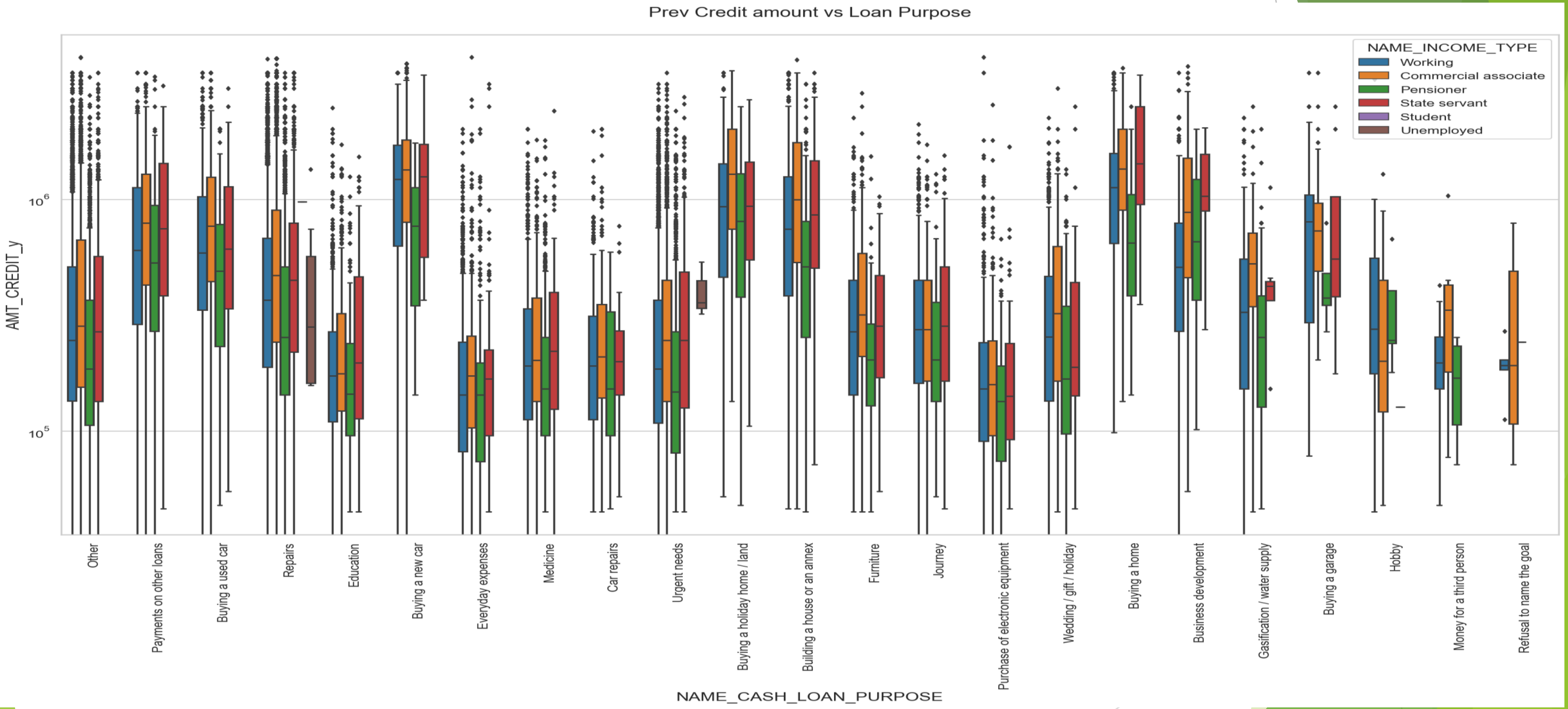
Conclusion From The Above Plots:

- 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

Graphs and Insights

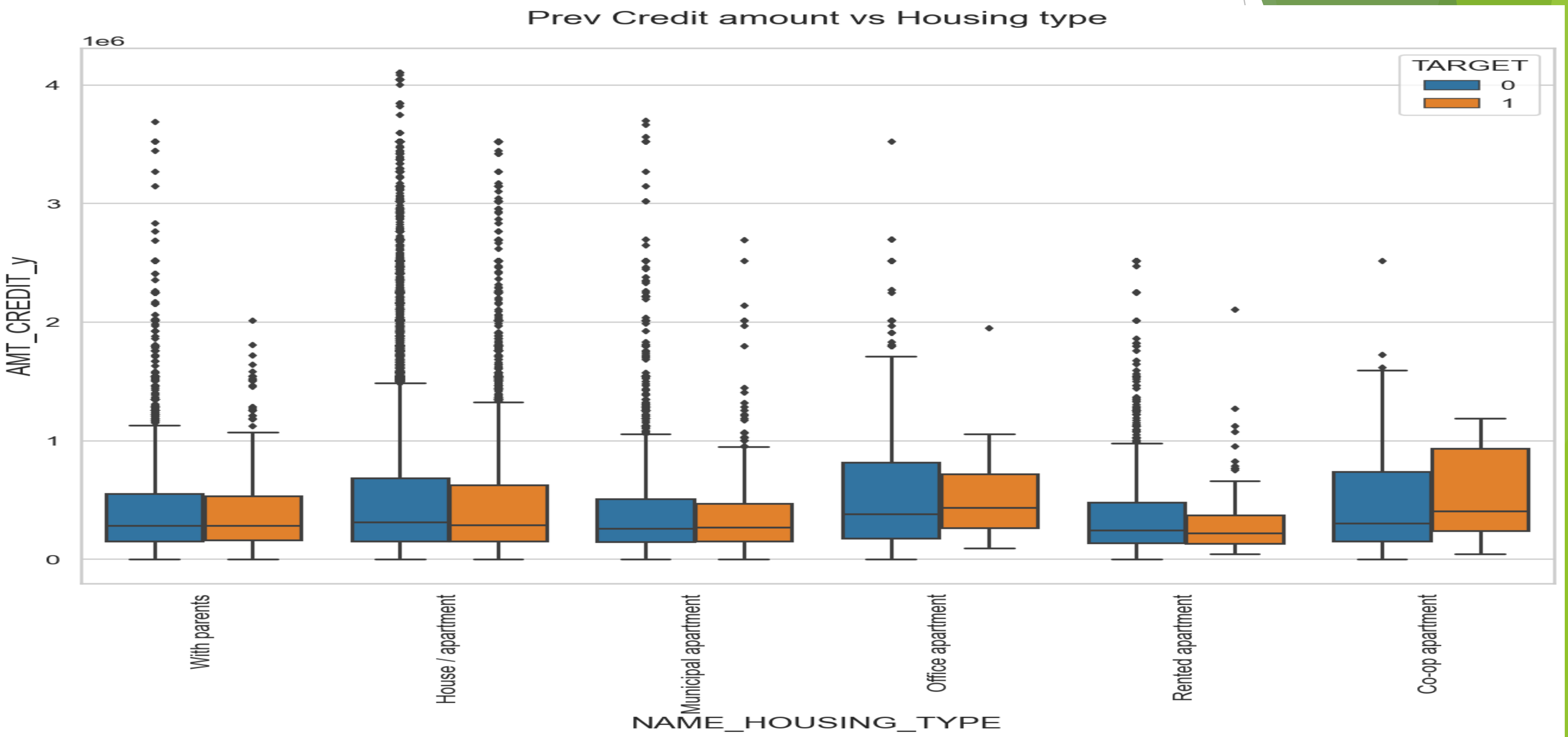


Graphs and Insights

Conclusion From The Above Plots:

- 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

Graphs and Insights



Graphs and Insights

Conclusion From The Above Plots:

- 1) 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

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

Final Summary

- 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

Thank You!