Credit Risk Analysis

Exploratory Data Analysis Case Study

by

Rajat Kawalkar

Problem Statement

The primary objective of this case study is to detect patterns that serve as indicators of a client's potential challenges in meeting their installment payments. These patterns can subsequently inform decision-making processes, including loan denial, reducing loan amounts, and offering loans to higher-risk applicants at elevated interest rates, among other actions.

By adopting these measures, the aim is to safeguard the interests of loan applicants who have the financial capacity to repay their loans, ensuring they are not wrongfully declined. Thus, the central focus of this case study is the identification of such applicants through Exploratory Data Analysis (EDA).

Fig: Taken from available open source

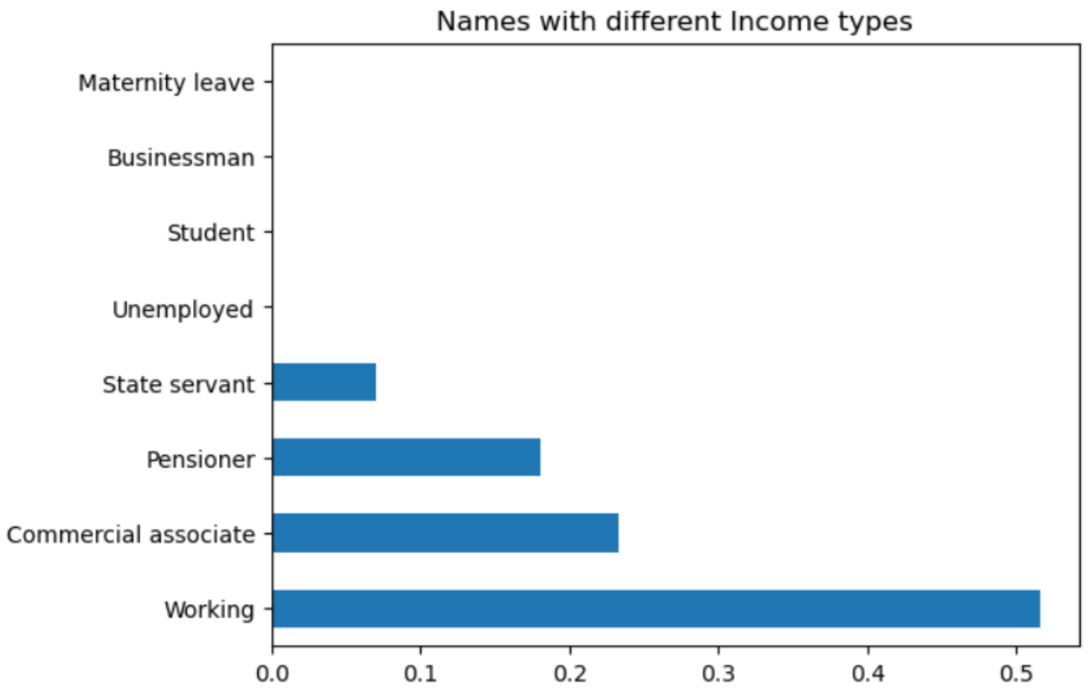
Steps to be followed

1. Begin by loading the data files and importing them using the necessary libraries.
2. Examine the dataset for missing values and consider employing imputation techniques, such as replacing them with mean, median, or mode values.
3. Assess the data quality for any anomalies and organize the data into groups to facilitate subsequent analysis.
4. Investigate the presence of data imbalance in both univariate and bivariate analyses, and explore correlation between variables.
5. Merge the application data with the previous dataset to create a comprehensive dataset.
6. Perform a thorough data analysis, including univariate and bivariate analyses, as well as correlation analysis.
7. Draw meaningful insights from the data to inform decision-making.
8. Categorize loan applicants based on eligibility and associated risks.

Categorical Unordered Univariate Analysis

Conclusion:

(a) Working associates make the most income while those on maternity leave make the least.
(b) There are more female clients than male clients, the income type of "maternity leave" is still associated with the fewest clients. Thus, it is conclusive that families do not ask for loans when they are expecting since they have other financial obligations for the upcoming addition to the family.

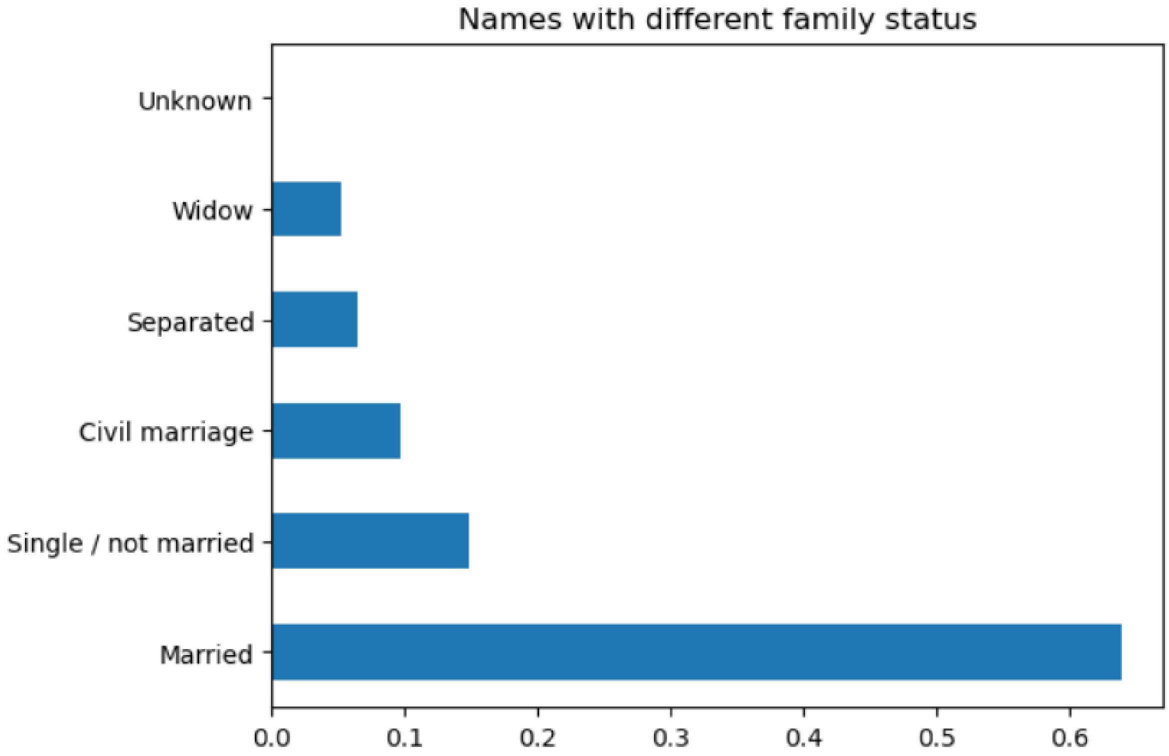



```
F        0.658344
M        0.341643
XNA      0.000013
Name: CODE_GENDER
```

Bar Graph for Family Status Categories

Conclusion:

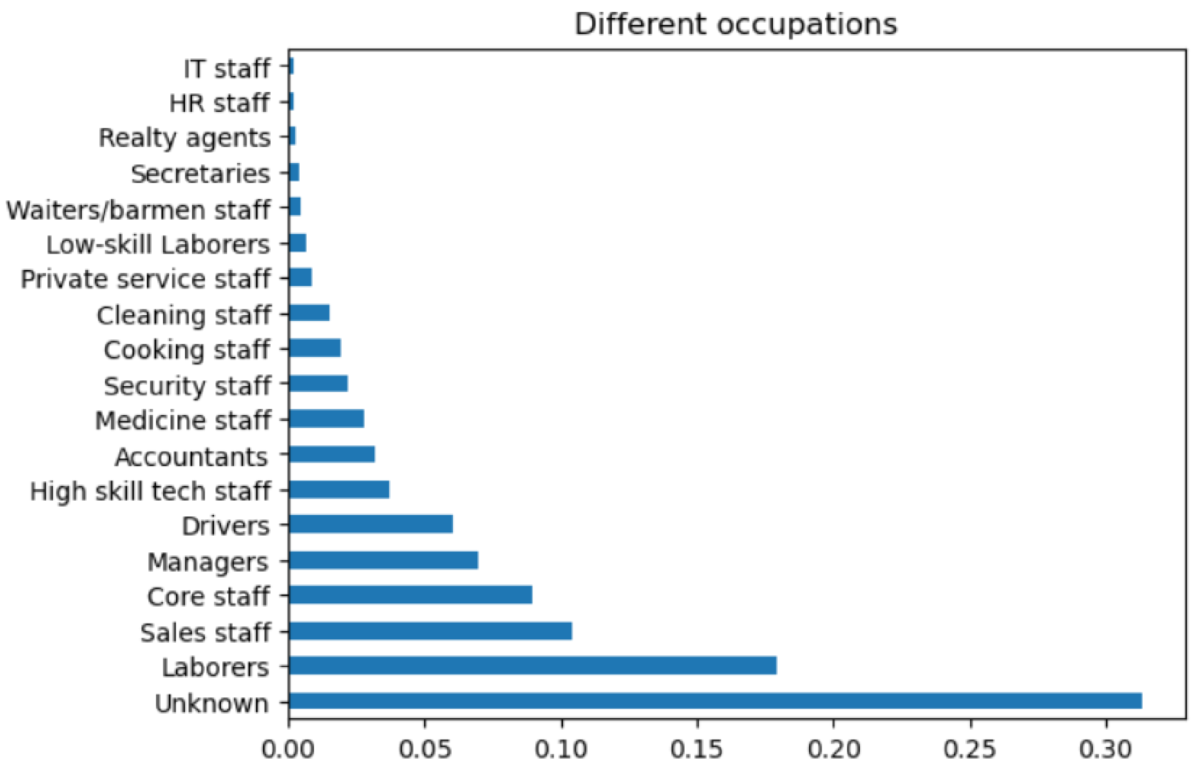
It is observed from above graph that people having "Married" status apply for loan the most.



Bar Graph for different Occupation Type Categories

Conclusion:

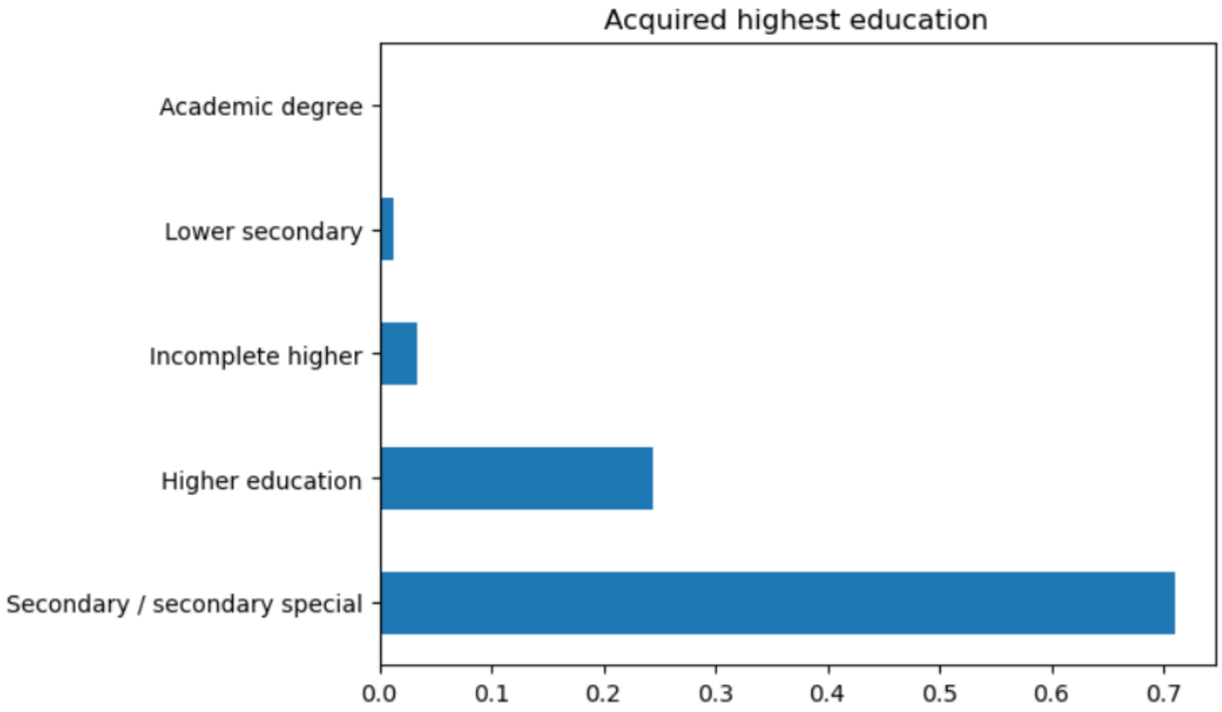
From all the clients, mostly are of unknown profession followed by laborers and sales staff. The least are from IT.



Bar Graph for different Name Education Type Categories

Conclusion:

Most of the clients have acquired secondary education as highest qualification.

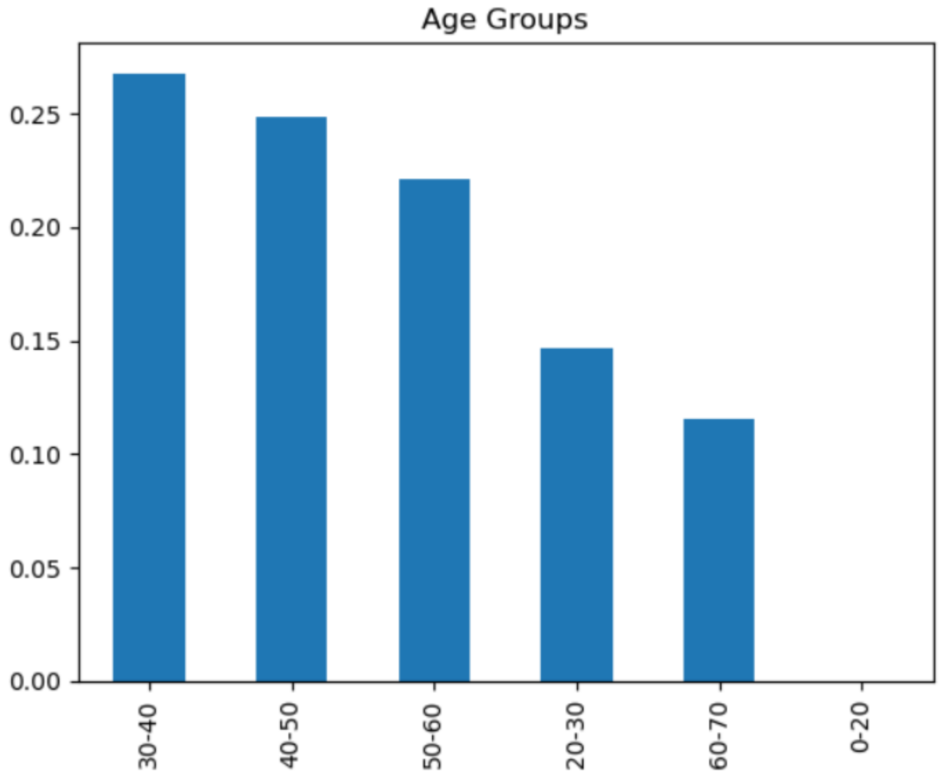


Numerical Univariate Analysis

Bar Graph for different Age Group Categories

Conclusion:

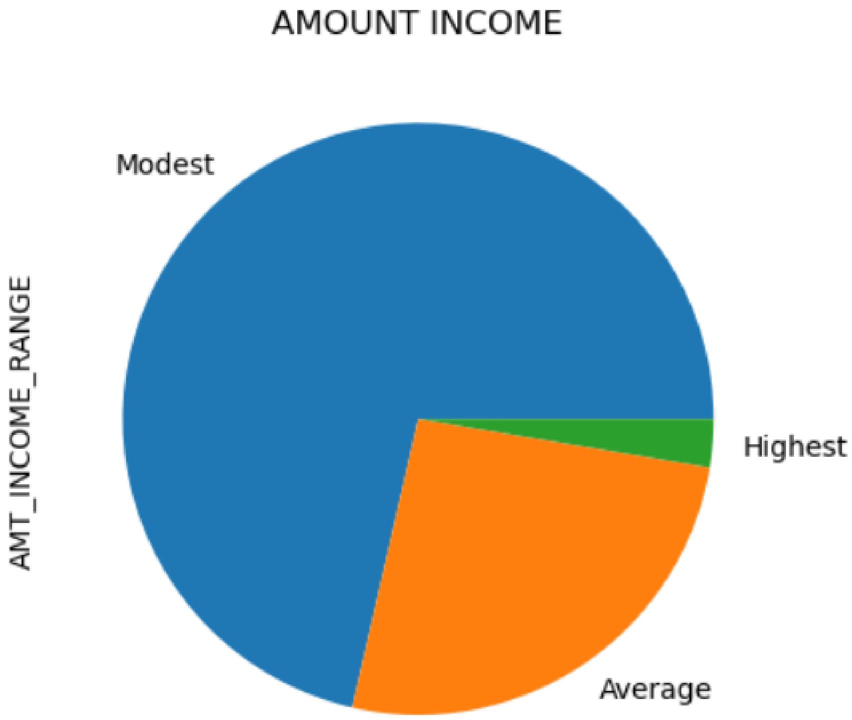
Most of the people belong to the age category of 30-40.



Pie Chart for Total Amount Income Range Categories

Conclusion:

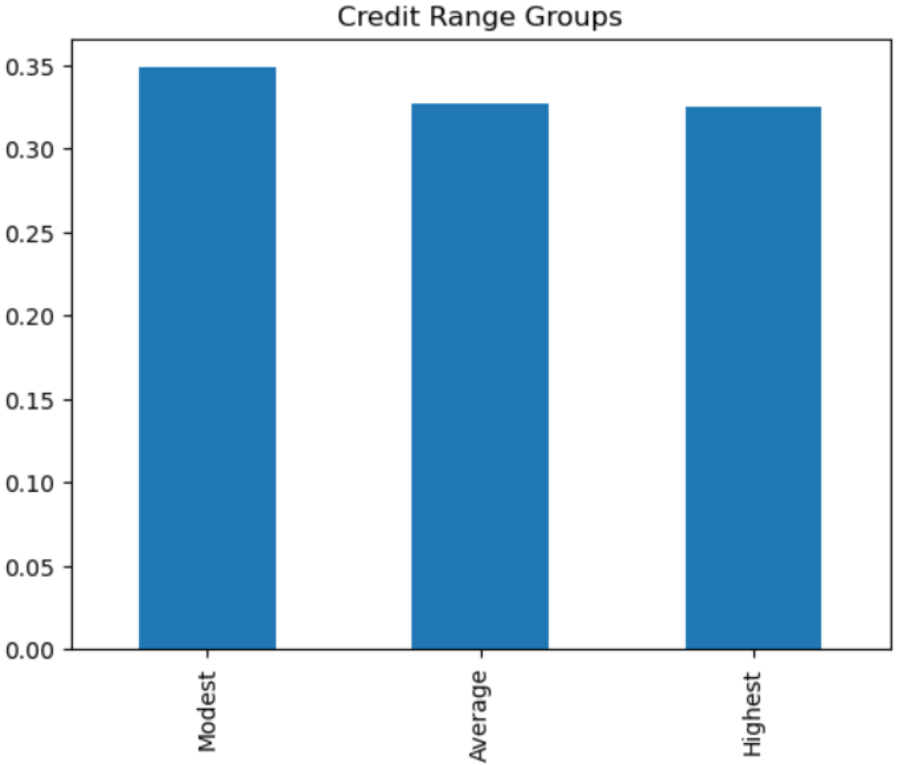
As can be seen from pie chart, majority of the clients are modest income holders. Hence, they are the ones who need loans more often in contrast to the high earning ones.



Bar Graph for different Amount Credit Range Categories

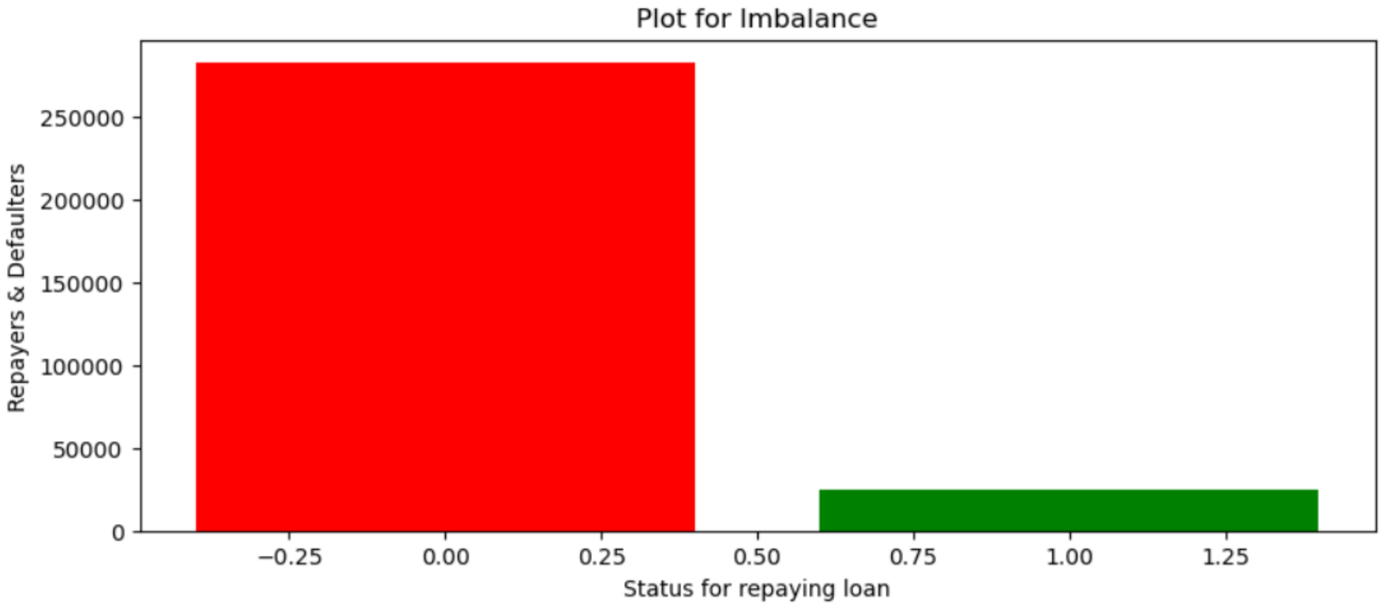
Conclusion:

From the above graph it is clear that most clients apply with modest amount of credit for loan.



Distribution of Data in Target Column

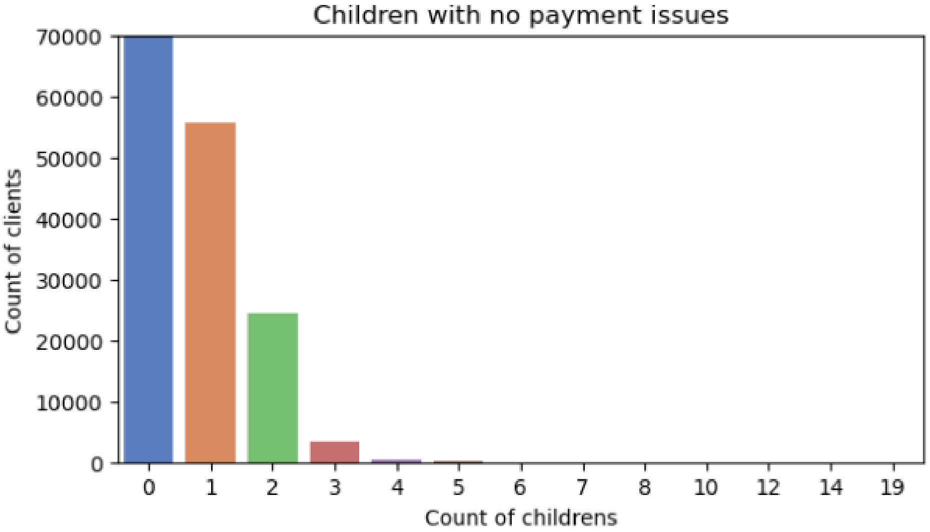
- The TARGET column consists of 8.07% '1's (ones), signifying that 8% of clients encounter payment difficulties.
- The remaining 91.93% are '0's (zeroes), indicating that 91.93% of clients do not face any payment issues.

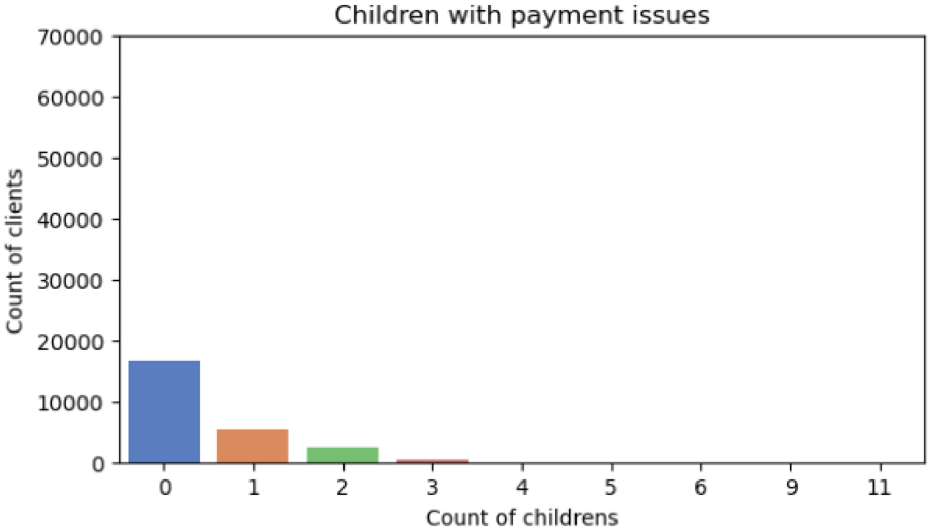



```
0    91.927118
1     8.072882
Name: TARGET,
```

Univariate Analysis for Target Variable

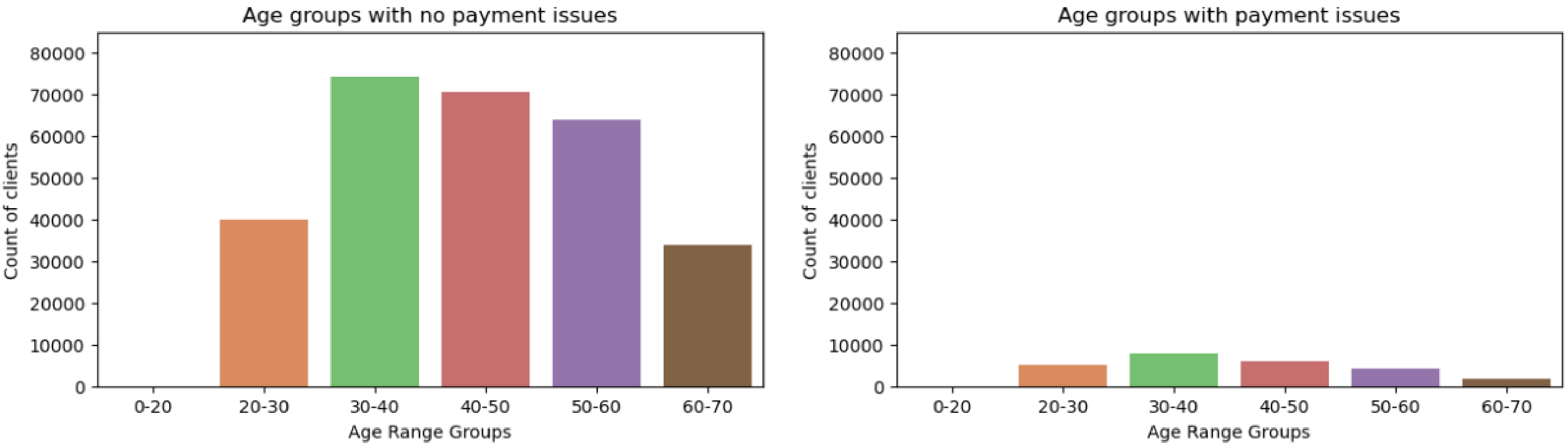
Conclusion: From above plots, it is evident that bank can consider to give away loans for client having no children.





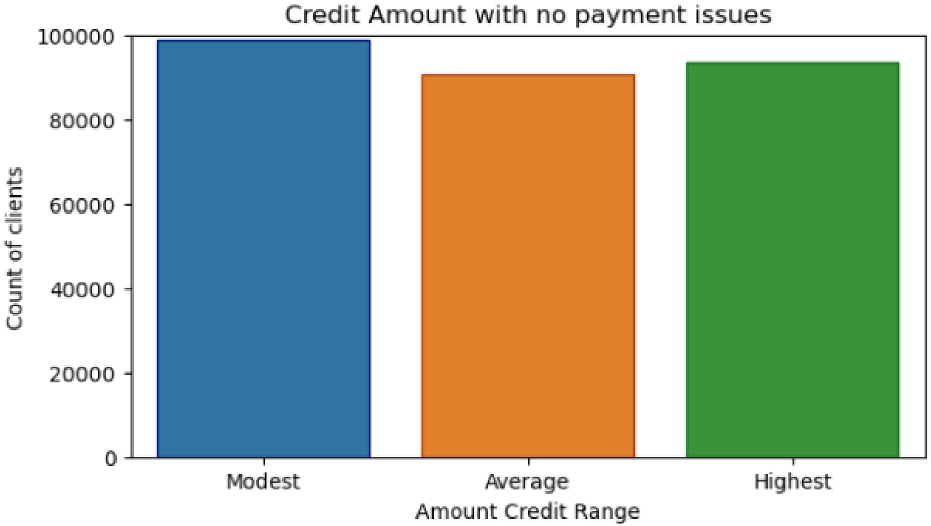
Age Analysis for defaulters(ones) and non-defaulters(zeroes)

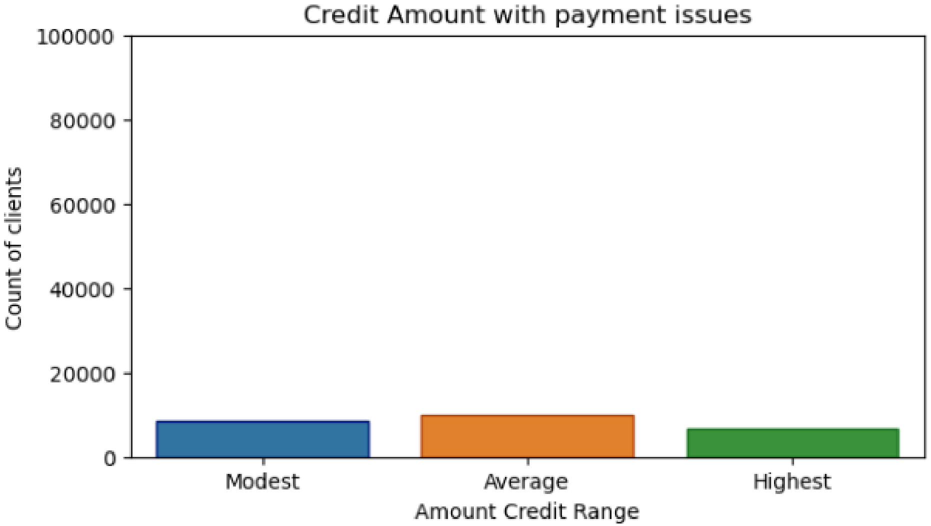
Conclusion: As most of the clients lie in the age group between 30 and 40, banks should lend out loans to this particular groups since they are also able to repay back. Also, low credit amounts are relatively easily returned.



Age Analysis for defaulters(ones) and non-defaulters(zeroes)

Conclusion: Modest credit amounts are relatively easily returned.

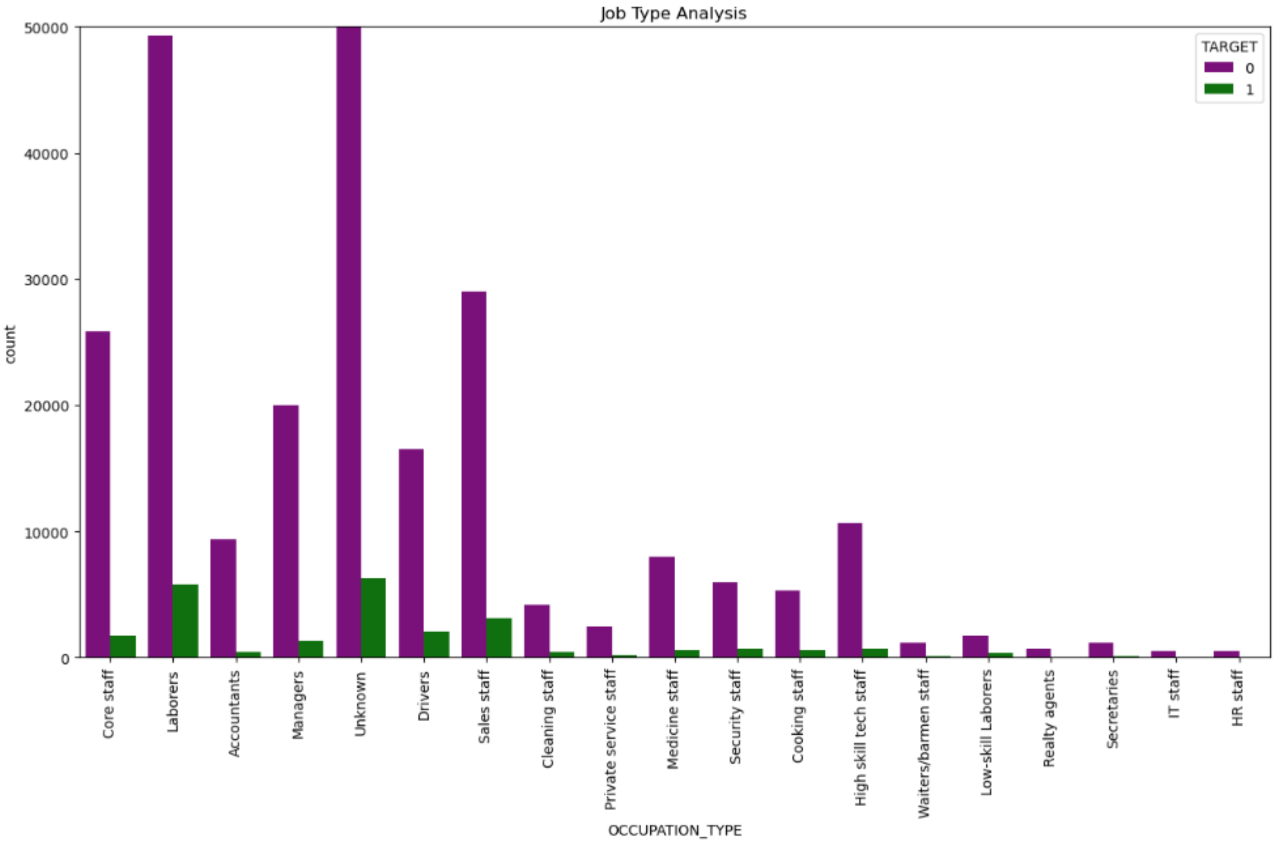




Occupation type categorical variable analysis for defaulters(ones) and non-defaulters(zeroes) data frame

Conclusion:

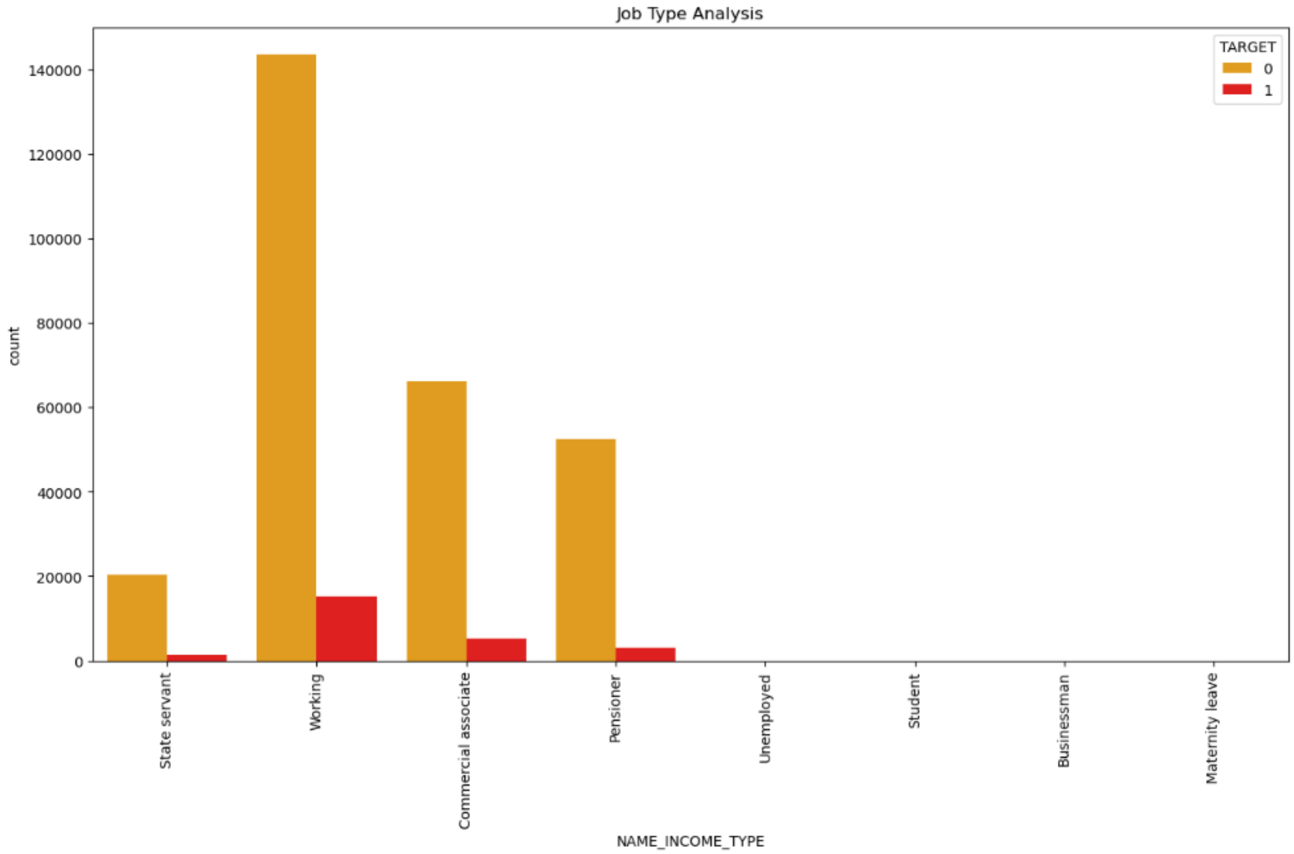
From above plots, (a) it is not conclusive to decide whether to lend loans to laborers or not since they have highest count in both target groups. However, they can be lended small amount of loans to this particular client group while (b) sales, core, IT and HR staff can be lended loans easily due to their repaying ability.



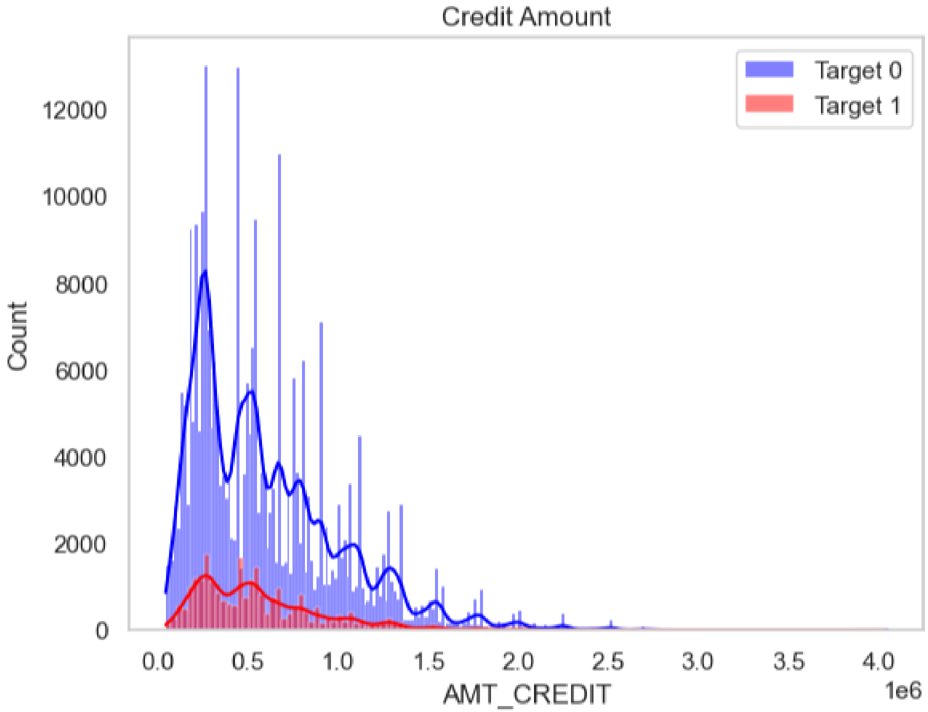
Income type categorical variable analysis for defaulters(ones) and non-defaulters(zeroes) data frame

Conclusion:

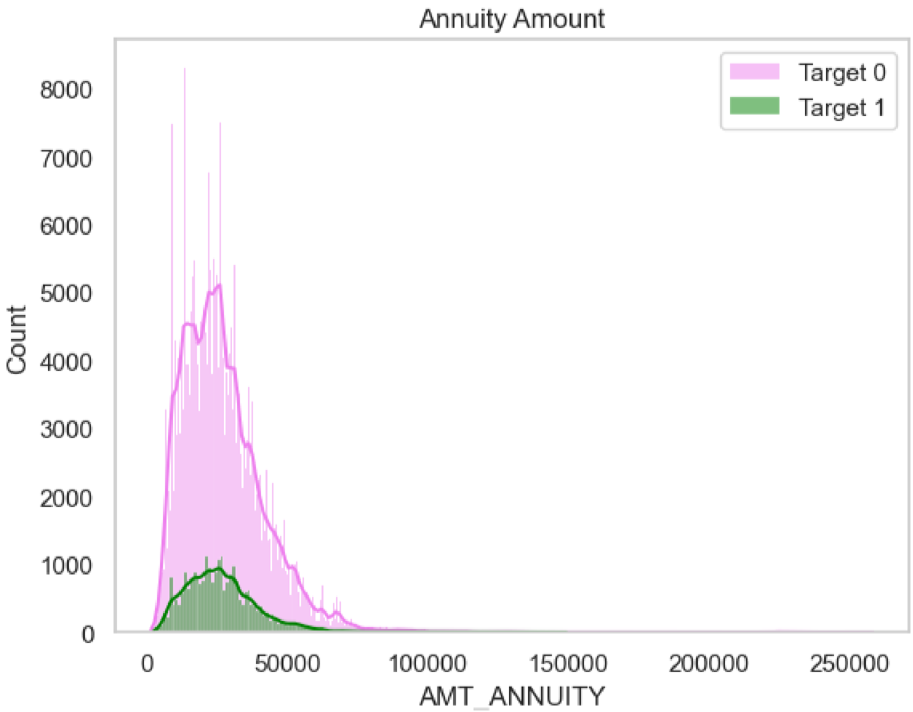
It can be seen that working and commercial associates can be lended loan as they repay it most likely among rest group of clients.



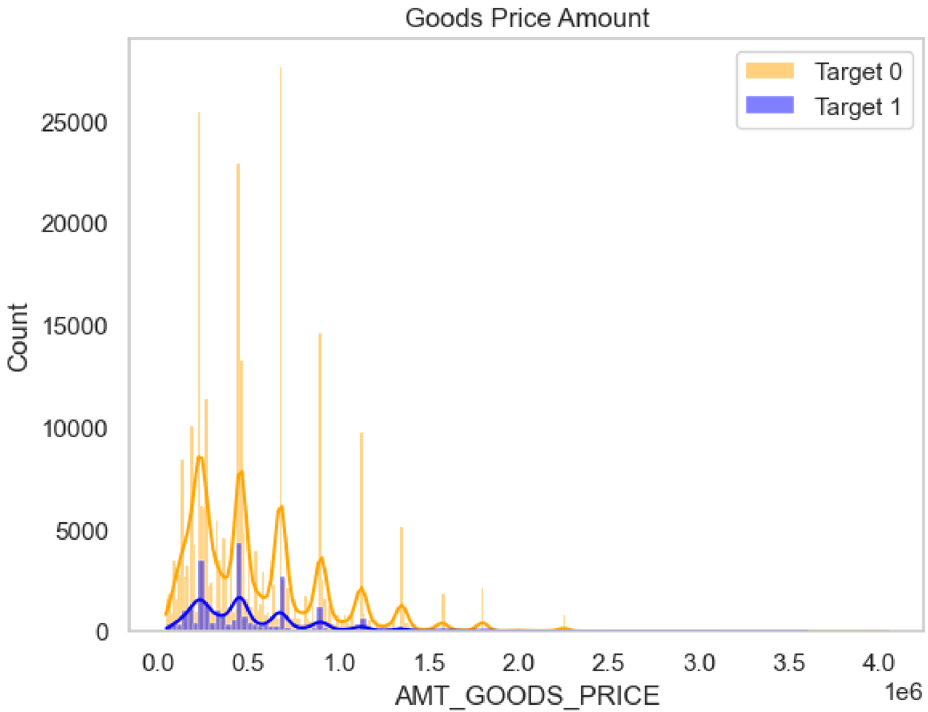
Analysing continuous column with target column for credit amount



Analysing continuous column with target column for annuity amount

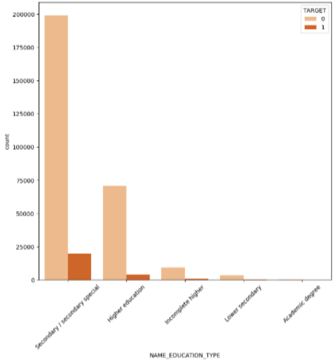


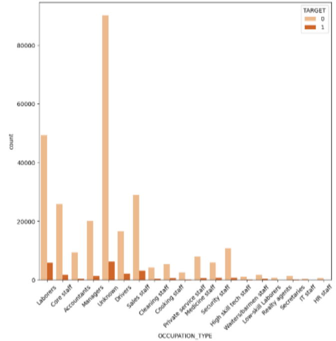
Analysing continuous column with target column for goods price amount

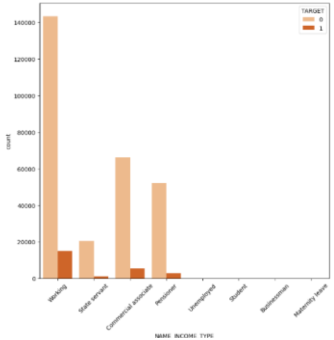


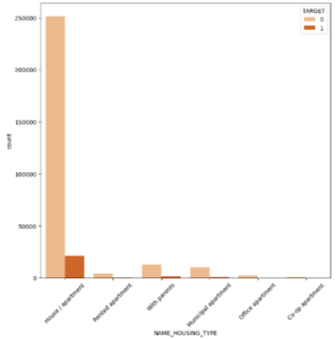
Conclusions:

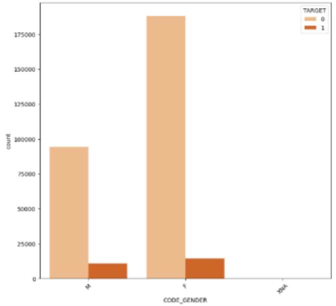
1.Based on the first graph, it is evident that banks may want to consider targeting more female customers for loan services, as they demonstrate a higher tendency to make timely repayments.
2.The second graph suggests that lending institutions might find working clients to be a promising target for loans, as they exhibit a notably higher rate of on-time payments.
3.The third graph highlights that customers with secondary education qualifications are more likely to meet their payment deadlines compared to those with academic degrees.
4.In the fourth graph, it becomes apparent that married customers tend to have a better track record of making loan payments on schedule, in contrast to widowed and separated individuals.
5.The fifth graph reveals that customers who own their houses or apartments are the most reliable in terms of making timely payments, outperforming other customer categories.
6.The sixth graph underscores that both sales staff and core staff members have a significantly higher percentage of on-time payments, making them attractive candidates for loan services.

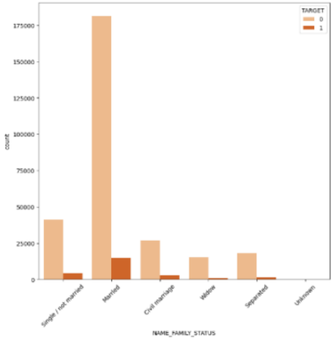




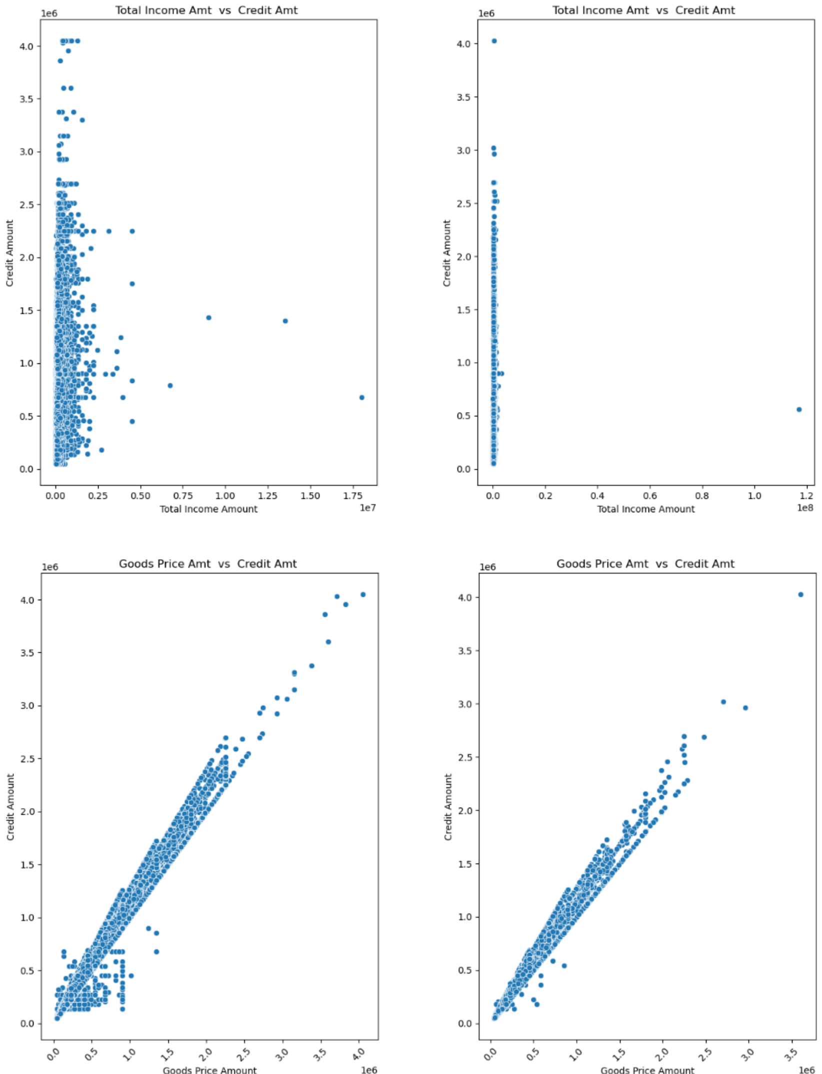








Bivariate analysis for defaulters(ones) and non-defaulters(zeroes) data frame

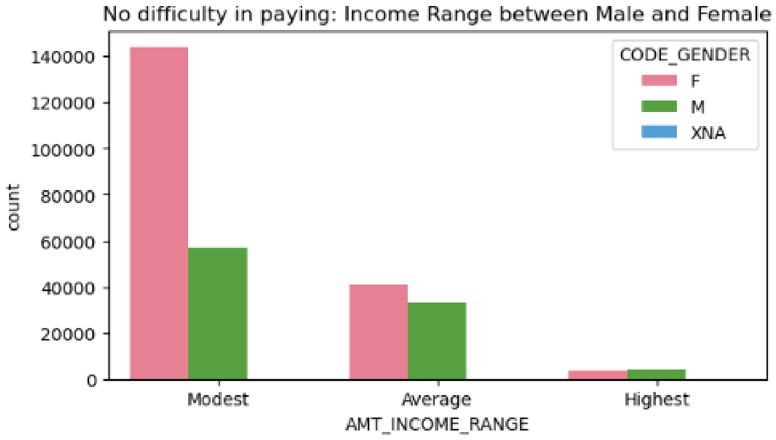


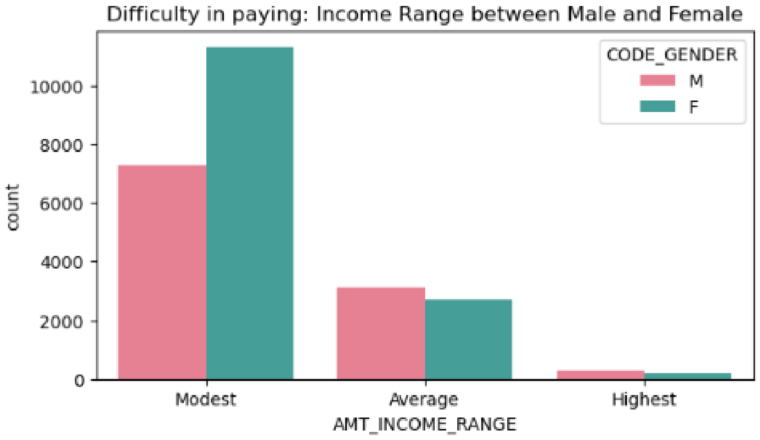
Conclusion:

1.Timely loan repayments often lead to improved credit scores, while individuals with lower incomes may encounter challenges meeting their installment obligations, resulting in lower credit ratings.
2.There exists a linear relationship between the price of goods and the credit amount. Clients who consistently make punctual repayments and possess higher-priced goods tend to enjoy higher credit scores

Numerical categorical analysis for defaulters(ones) and non-defaulters(zeroes) data frame

Conclusion: It can be seen that females are better in repaying the amount back as compared to males.

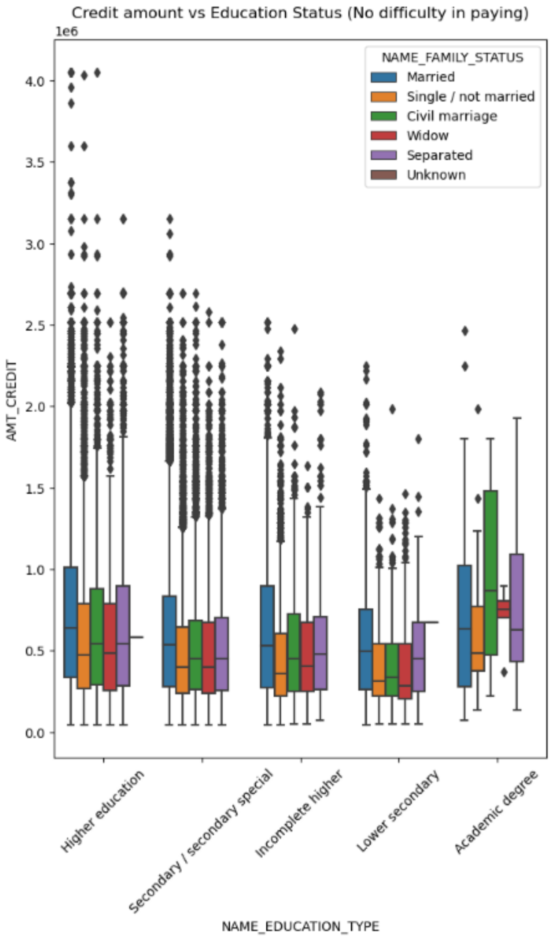


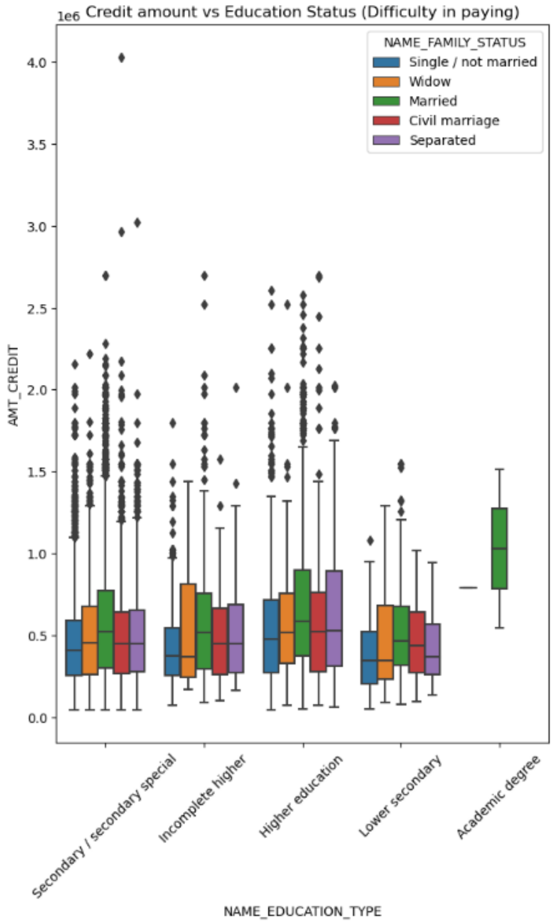


Numerical categorical analysis for defaulters(ones) and non-defaulters(zeroes) data frame for credit amount

Conclusions:

1.Payment difficulties are primarily observed among married individuals with academic degrees.
2.Clients with higher education levels tend to have better credit profiles and encounter fewer payment difficulties.
3.Clients with secondary special education qualifications are more likely to experience payment difficulties despite having higher credit amounts.
4.Individuals with lower secondary education have notably low credit scores.

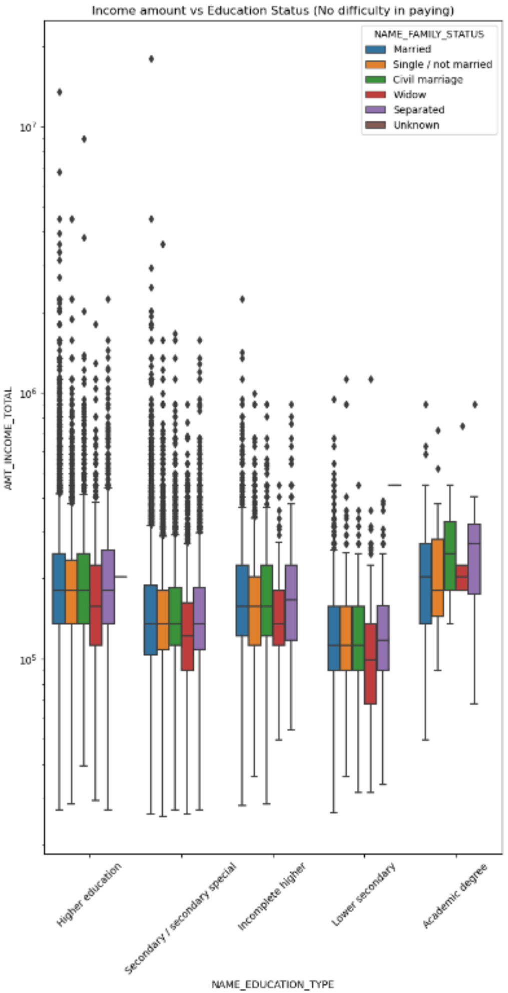


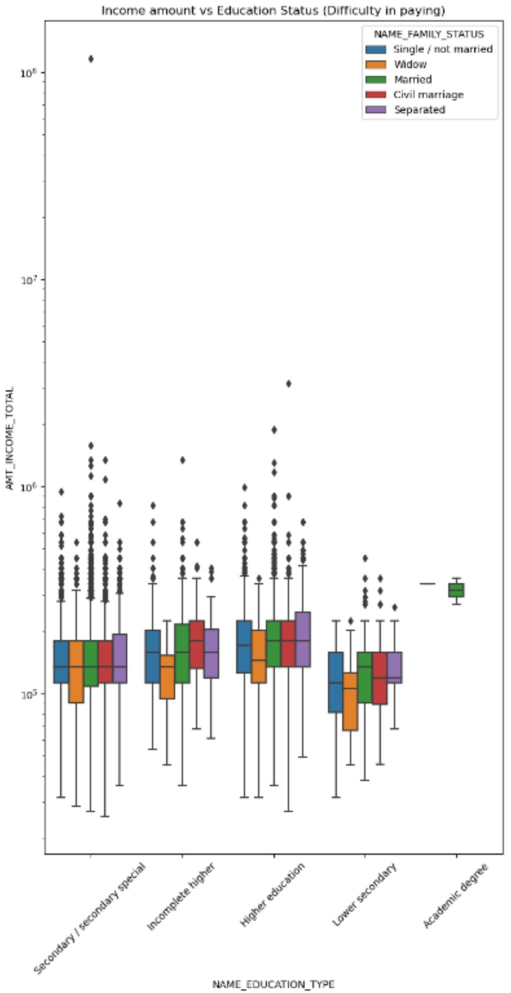


Numerical categorical analysis for defaulters(ones) and non-defaulters(zeroes) data frame for income amount in log scale

Conclusion:

1.Clients with advanced education tend to enjoy higher incomes and exhibit fewer payment difficulties.
2.Individuals with lower income levels often face challenges in making payments, particularly within the secondary education category.
3.The category of higher education tends to exhibit a notable number of outliers in income distribution.
4.The lower secondary education category typically corresponds to lower income levels.
In summary, it can be inferred that clients with higher educational qualifications and better income prospects are more likely to meet their loan repayment obligations on time.

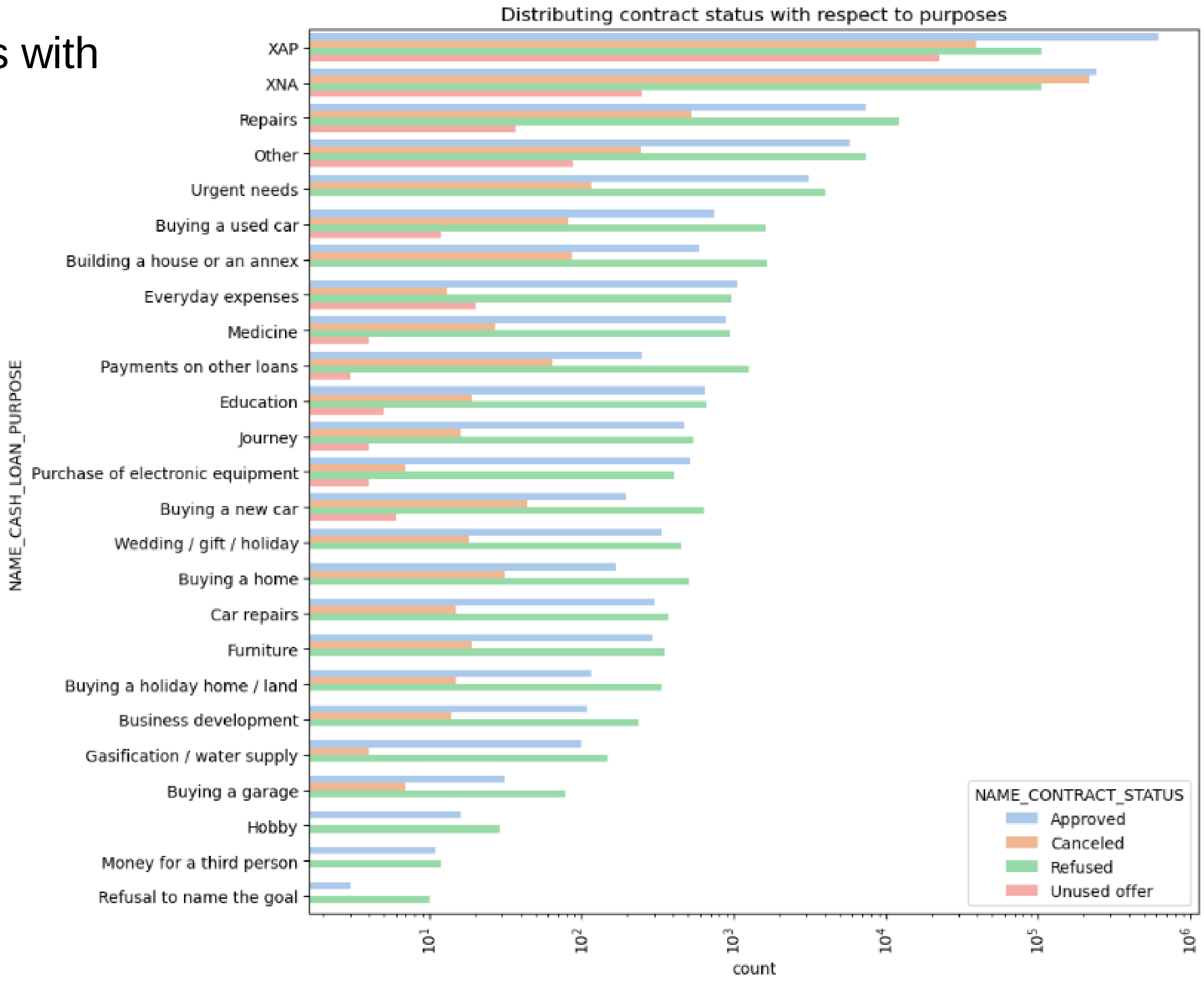




Distributing contract status with respect to purposes.

Conclusions:

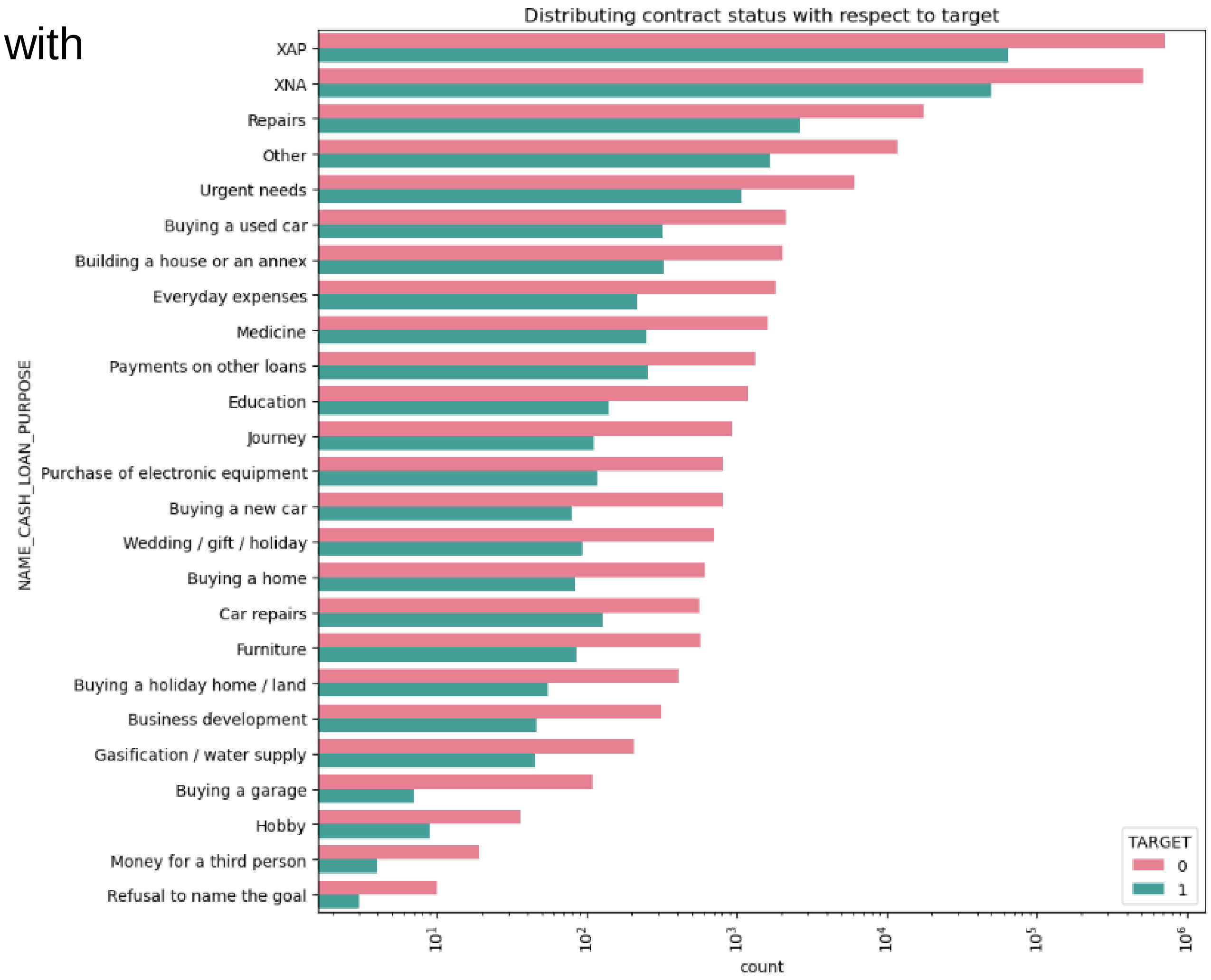
(a) Medicine and Education have an equal count of offers, with an identical number of approvals and refusals.
(b) Repairs exhibits a higher number of refusals than approvals, indicating a notable default rate associated with this category.
(c) The category of Buying a car also demonstrates a greater count of refusals compared to approvals, signifying that financial institutions perceive it as a higher-risk application.
(d) This category has a substantial quantity of missing values.



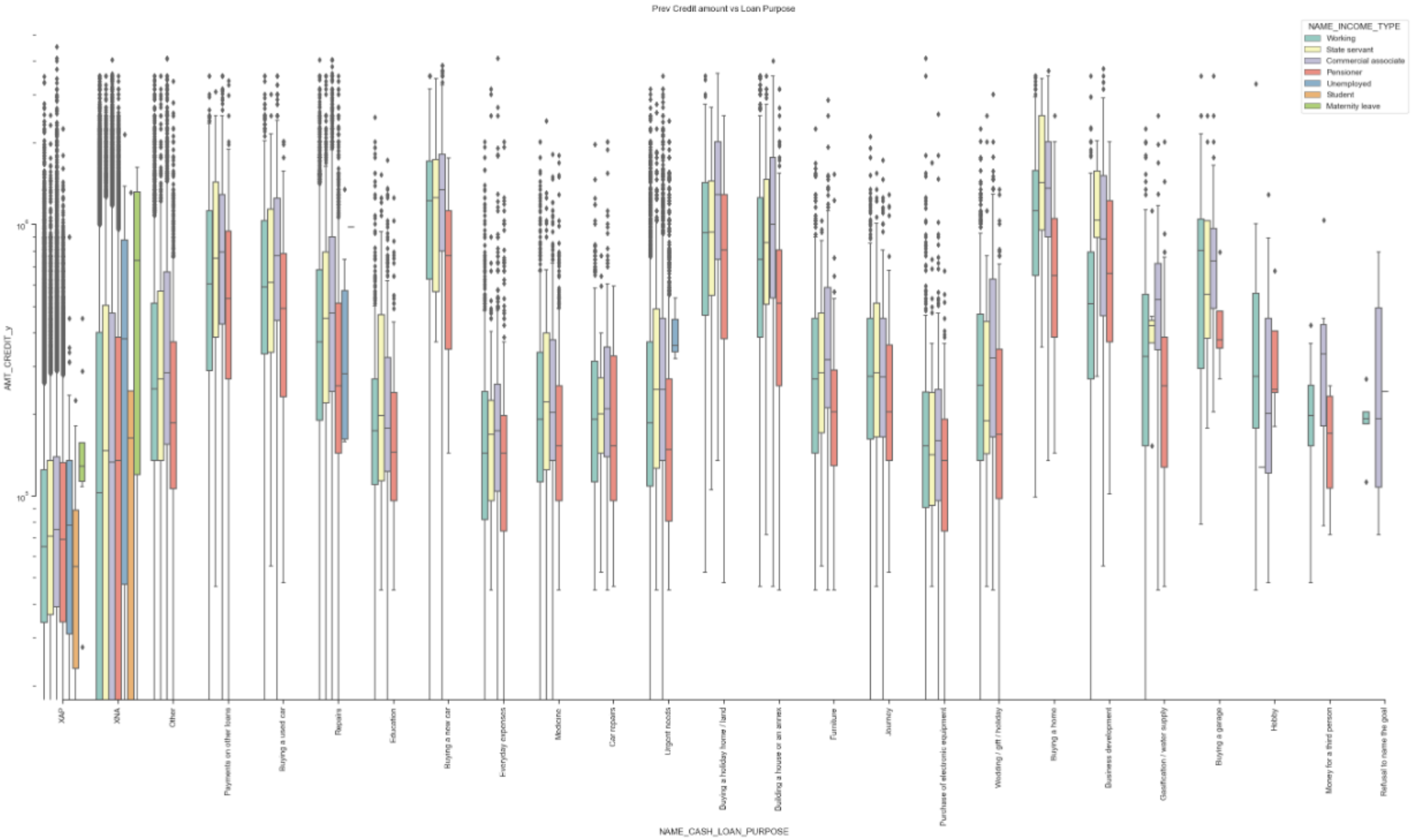
Distributing contract status with respect to target.

Conclusion:

In this case, clients who applied for loans with the purpose of "repairs" appear to experience the highest challenges when it comes to loan repayment. However, there are instances, such as "education," "Buying a Home," and "Business development," where borrowers encounter fewer difficulties in meeting their payment obligations. Hence, these specific categories may be deemed as more favorable for loan approvals.



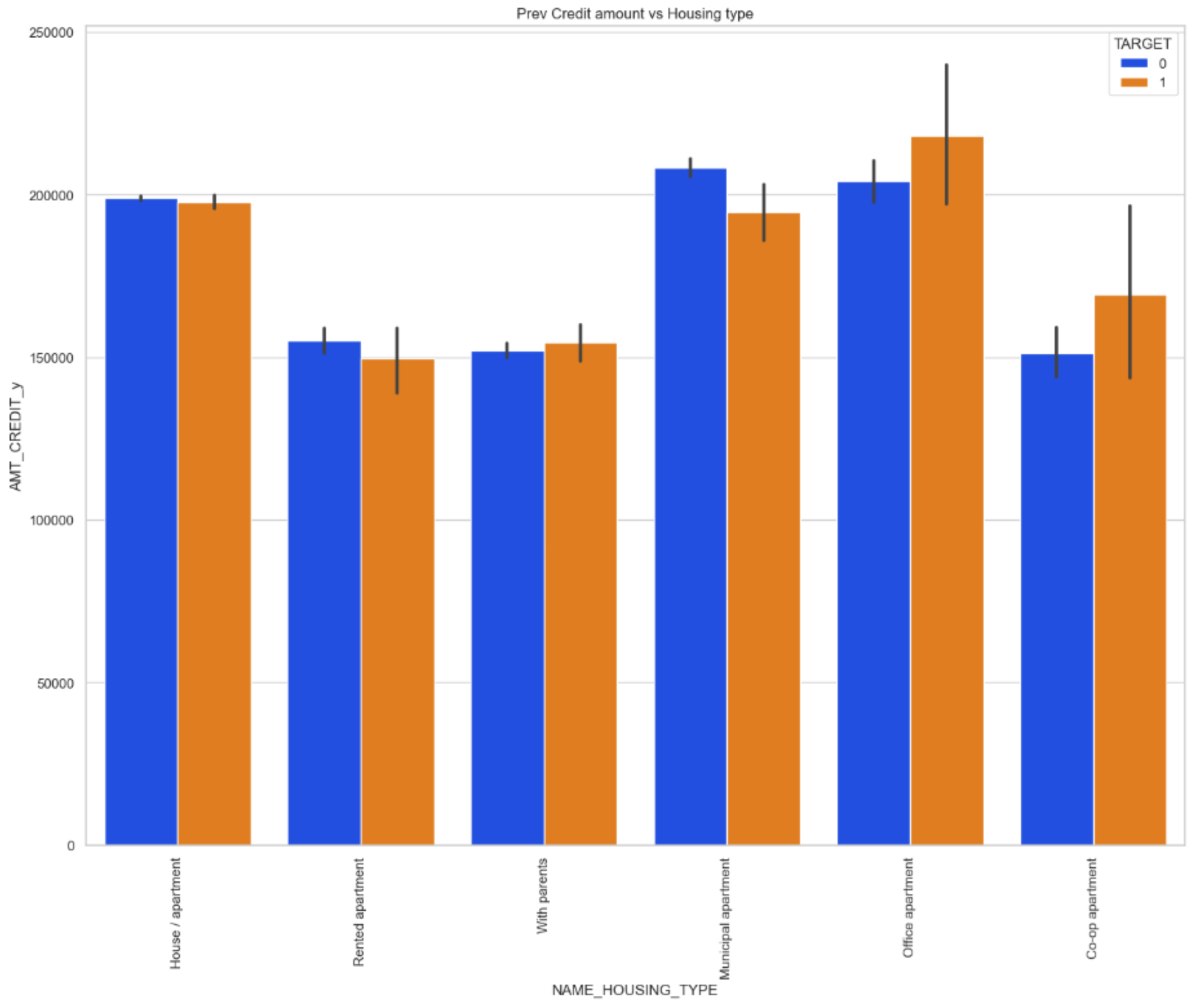
Previous credit amount vs Loan purpose using log scale.



Conclusion:

The loan purposes, such as 'Buying a home,' 'Buying a land,' 'Buying a new car,' and 'Building a house,' typically have higher credit amounts. Among those who are state servants, purchasing a new home tends to result in a notably higher credit amount, while the 'Hobby' category generally has lower credit amounts. On the other hand, individuals with a 'Working' income type often secure the highest credit amounts when purchasing a new home.

Previous credit amount vs Housing type using log scale.



Conclusion:

Residents of office apartments tend to face significant challenges in repaying loans, while those residing in municipal apartments generally experience fewer difficulties in loan repayment. Therefore, it would be prudent for banks to prioritize clients with municipal apartment housing when considering loan applications.

Final Conclusions

Criteria for predicting whether an applicant will be a responsible borrower include:

- **NAME_EDUCATION_TYPE:** Clients with academic degrees and higher education levels tend to have lower default rates.

- **NAME_INCOME_TYPE:** Students and business professionals typically have a low default rate.

- **DAYS_BIRTH:** Individuals in the age group of 50-60 demonstrate a lower likelihood of defaulting on loans.

- **AMT_INCOME_TOTAL:** Applicants with incomes exceeding 700,000 are less prone to default.

- **CNT_CHILDREN:** Individuals with zero to two children are more likely to repay their loans.

- **CREDIT_SCORE:** Those with average to high credit scores are less inclined to default.

- **AMT_CREDIT_RANGE:** Loans of smaller amounts are generally repaid more easily.

- **OCCUPATION_TYPE:** Sales and core staff members are less likely to default on loans.

- **NAME_INCOME_TYPE:** State servants and working professionals exhibit a lower propensity to default.

Criteria for Predicting Loan Defaulters:

- **CODE_GENDER:** Male applicants exhibit a higher likelihood of defaulting.

- **NAME_FAMILY_STATUS:** Individuals who are single or in civil marriages have a significantly higher default rate.

- **NAME_EDUCATION_TYPE:** Those with Lower Secondary education tend to experience a higher default rate.

- **NAME_INCOME_TYPE:** Unemployed clients are more prone to defaulting.

- **OCCUPATION_TYPE:** It's advisable to be cautious with applicants in the occupations of Drivers, Waiters/Barmen staff, Security staff, and Cooking staff, as they exhibit a notably higher default rate.

- **DAYS_BIRTH:** It is recommended to exercise caution with both young and older applicants, as they have a heightened probability of defaulting.