

Credit EDA Case Study

Soumya Roy



Introduction

Scenario

- ✓ Real-world application of Exploratory Data Analysis (EDA) in the banking and financial services sector.
- ✓ Objective: Utilize EDA techniques to identify and mitigate lending risks.

Focus

- ✓ Case study focus: Applying EDA to inform lending decisions and minimize default risks.
- ✓ Highlight the importance of data-driven decision-making in responsible lending practices.
- ✓ EDA serves as a critical tool for efficient risk mitigation, ensuring both financial institution stability and customer protection.

RISK ANALYSIS IN BANK AND FINANCIAL SERVICES

This case study focuses on risk analytics in banking and financial services, particularly in the context of loan approval decisions. The problem statement outlines the two main risks associated with loan approval:

- ✓ Risk of not approving a loan to a client who would repay it, resulting in a loss of business.
- ✓ Risk of approving a loan to a client who would default, resulting in financial losses for the company.

The data provided contains information about loan applications, categorized into two scenarios:

- ✓ Clients with payment difficulties: Those who had late payments on at least one of the first few instalments.
- ✓ All other cases: Applications where payments were made on time.
- ✓ When a client applies for a loan, there are four possible decisions:

Approved: Loan application approved.

Cancelled: Client cancels the application during approval.

Refused: Loan application rejected by the company.

Unused offer: Loan cancelled by the client at different stages of the process.

Objective

Use EDA to analyze loan application data and identify patterns to help decide loan approvals.

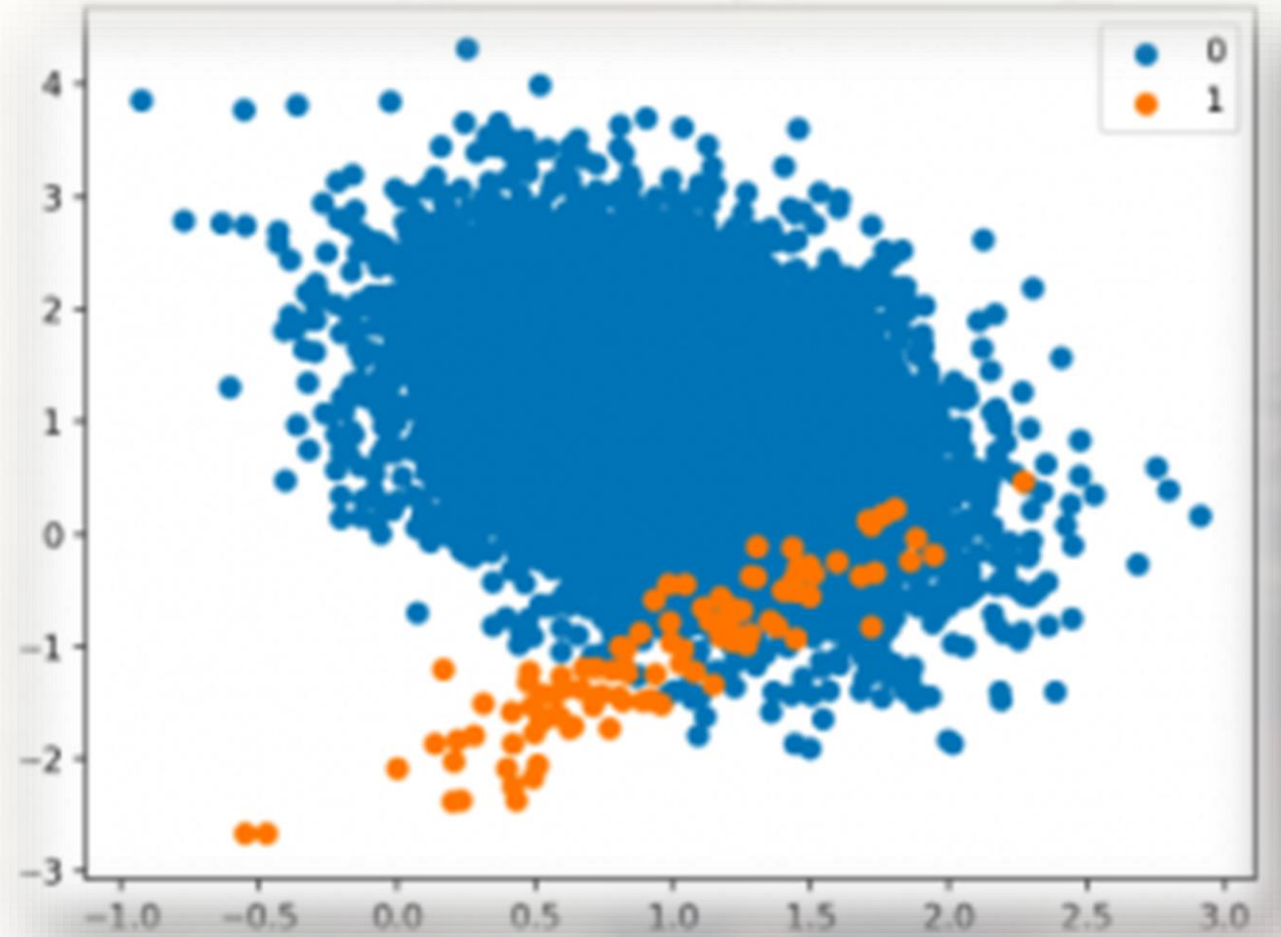
Actions

- ✓ Deny the loan
- ✓ Reduce the loan amount
- ✓ Offer higher interest rates to risky applicants

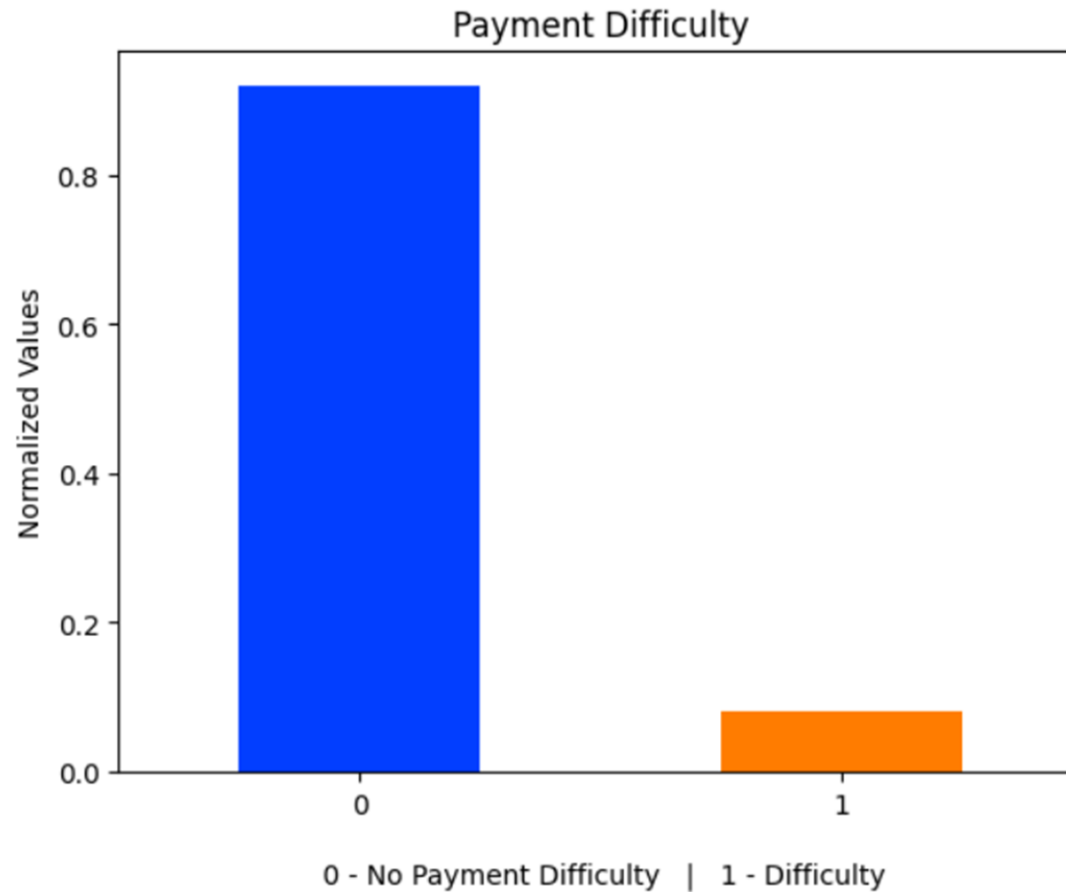
Goals

- ✓ Ensure capable consumers are approved
- ✓ Identify strong indicators of loan default (driver variables)
- ✓ Utilize findings for portfolio and risk assessment
- ✓ Recommendation: Research risk analytics to understand variable types and their significance.

DATA IMBALANCE



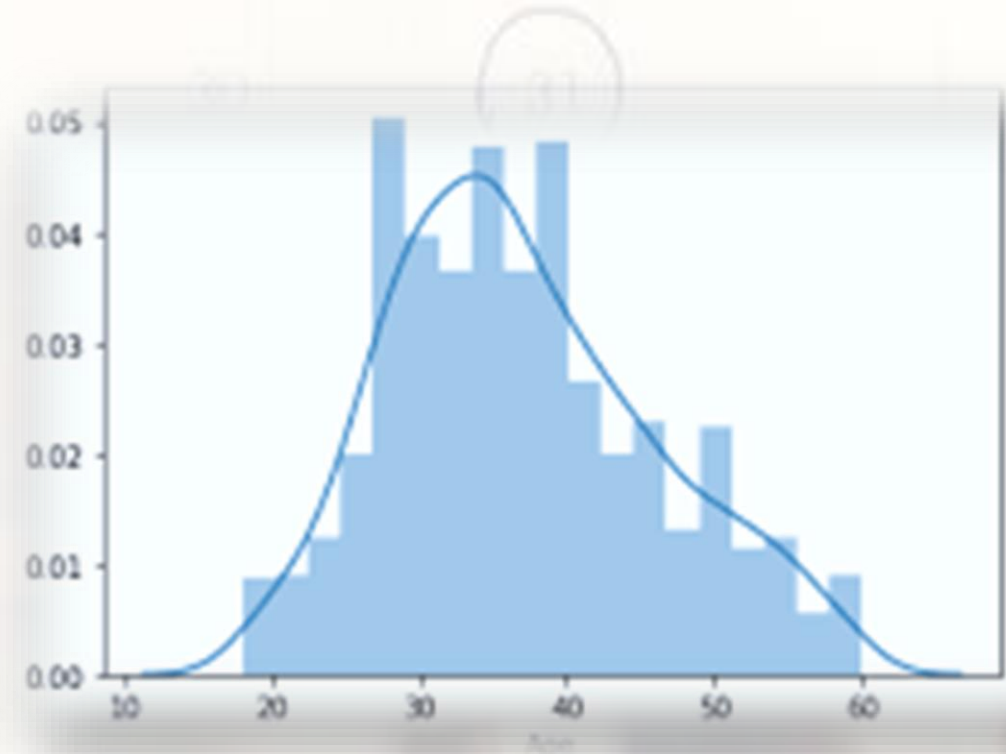
Imbalance



DATA IMBALANCE

There is a high imbalance observed in the data, as depicted in the plot, where out of 11 individuals, only 1 person is facing payment difficulties

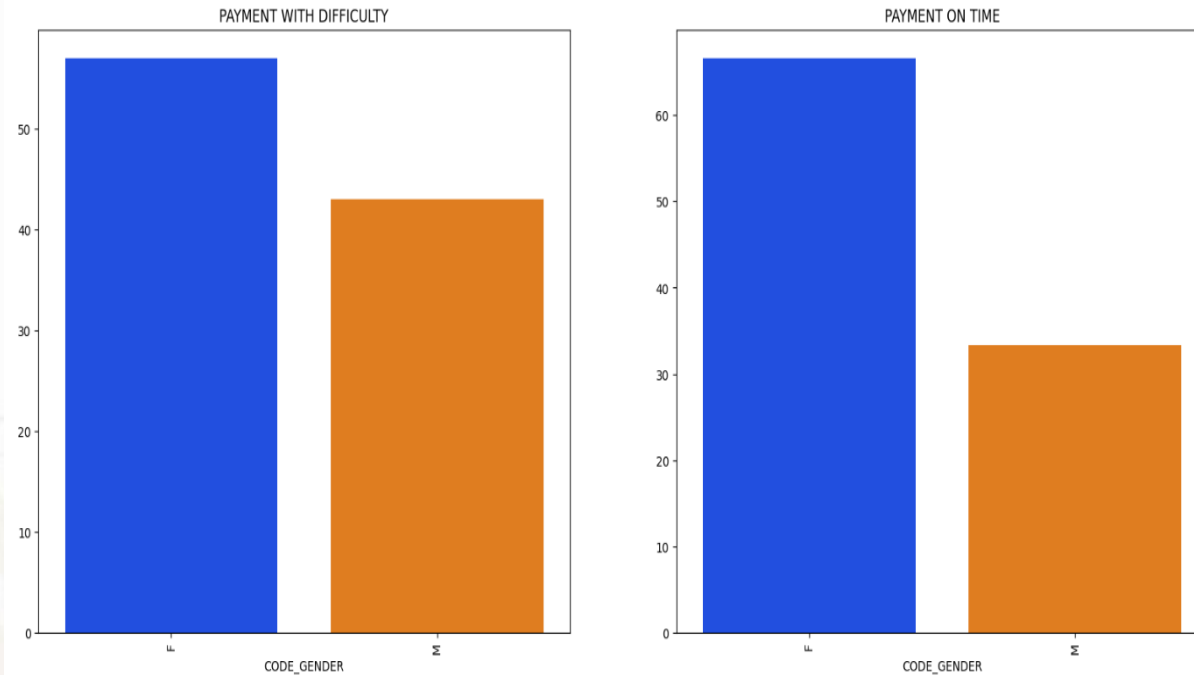
UNIVARIATE ANALYSIS



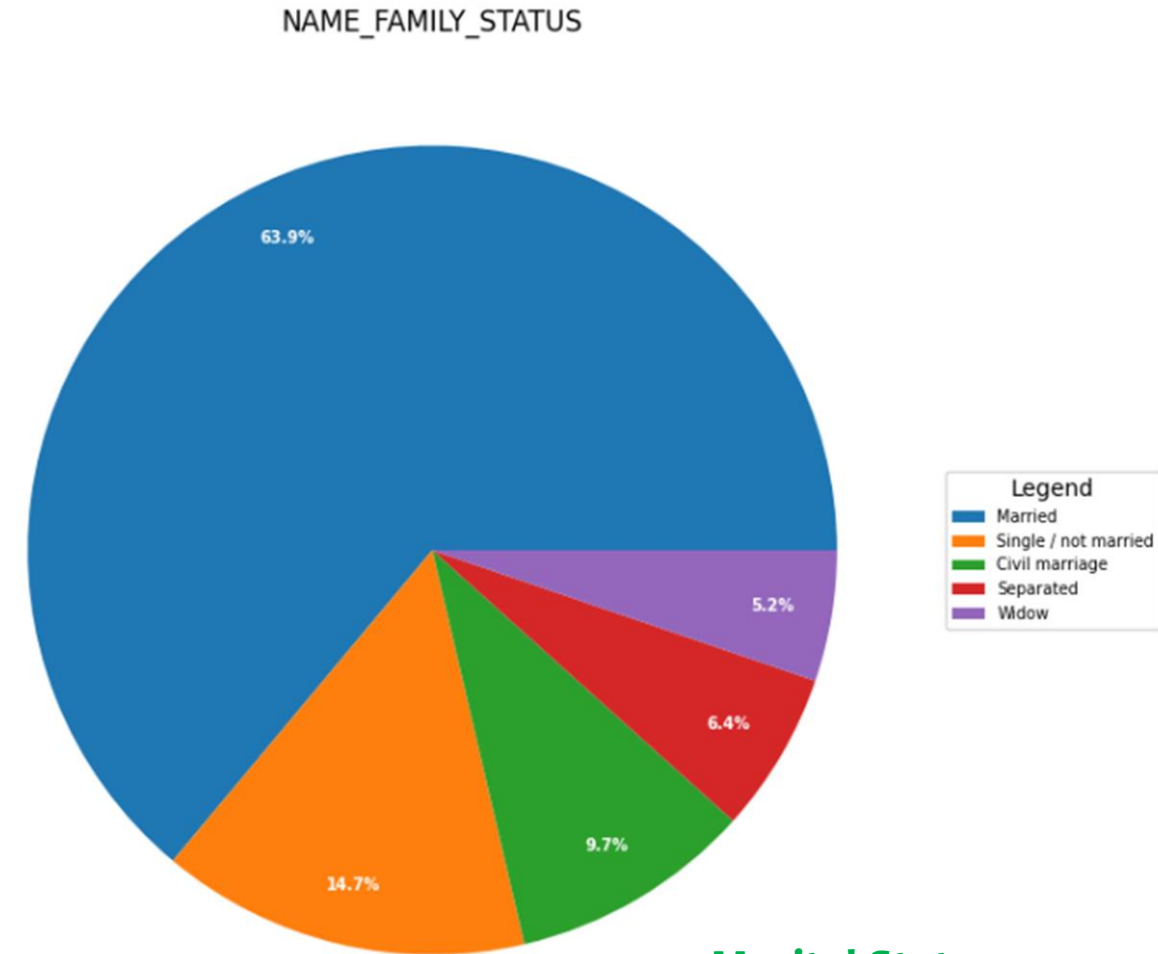
Univariate Analysis

CODE_GENDER

Female applicants outnumber **male applicants**. The percentage of male applicants experiencing payment difficulties is about 10%, compared to 7% for female applicants. The difference in payment difficulty percentages between genders is minimal; it is nearly the same.



NAME_FAMILY_STATUS



Marital Status

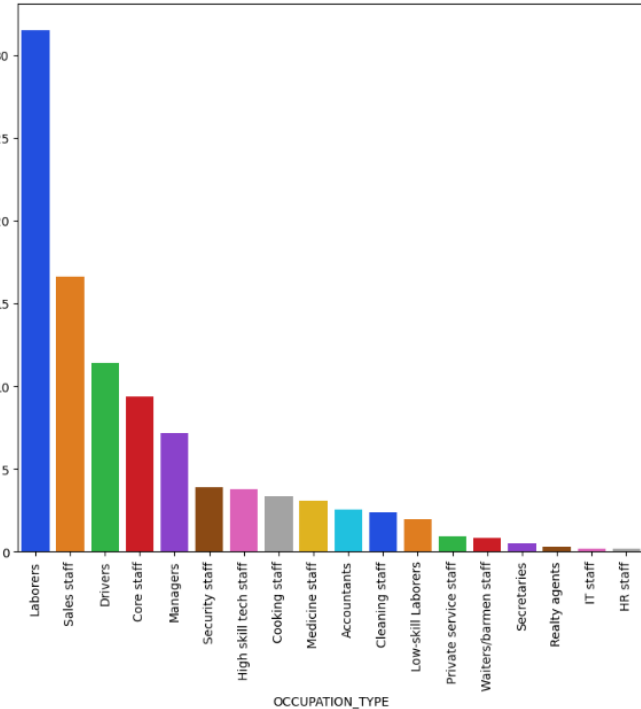
The majority, 63%, of applicants are married

Univariate Analysis

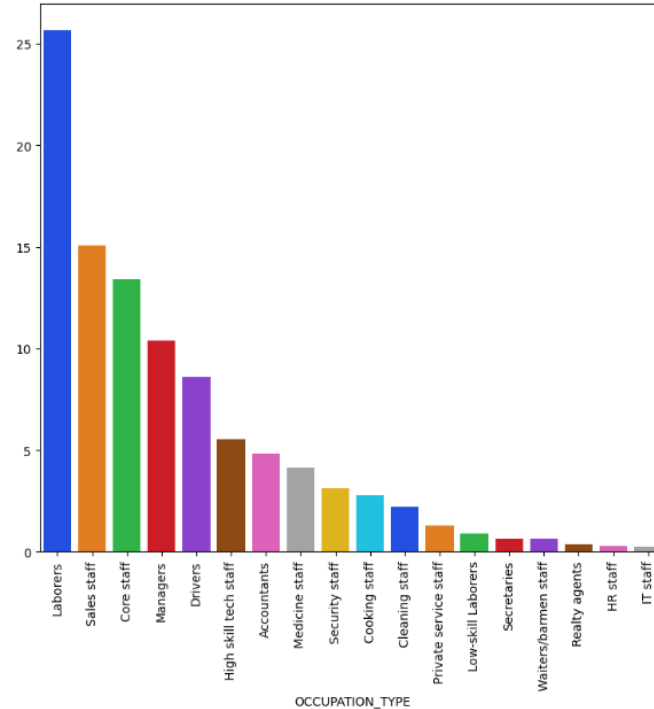
OCCUPATION_TYPE

Laborers are in maximum number followed by **sales stuff applicants**

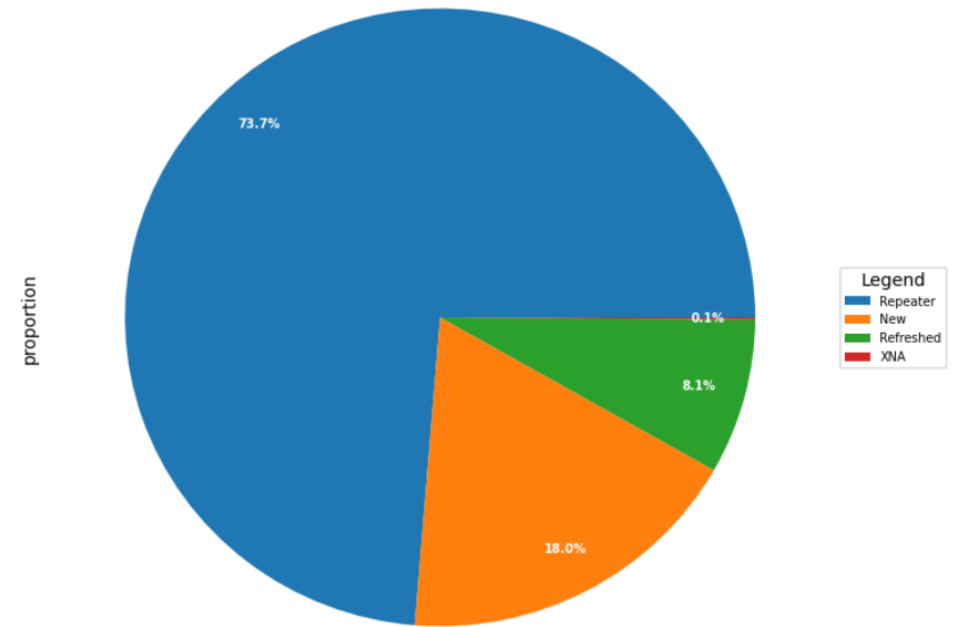
PAYMENT WITH DIFFICULTY



PAYMENT ON TIME



NAME_CLIENT_TYPE Value Counts



Considering the previous applicant's Record , it has been observed , the majority of the applicants are **Repeaters**

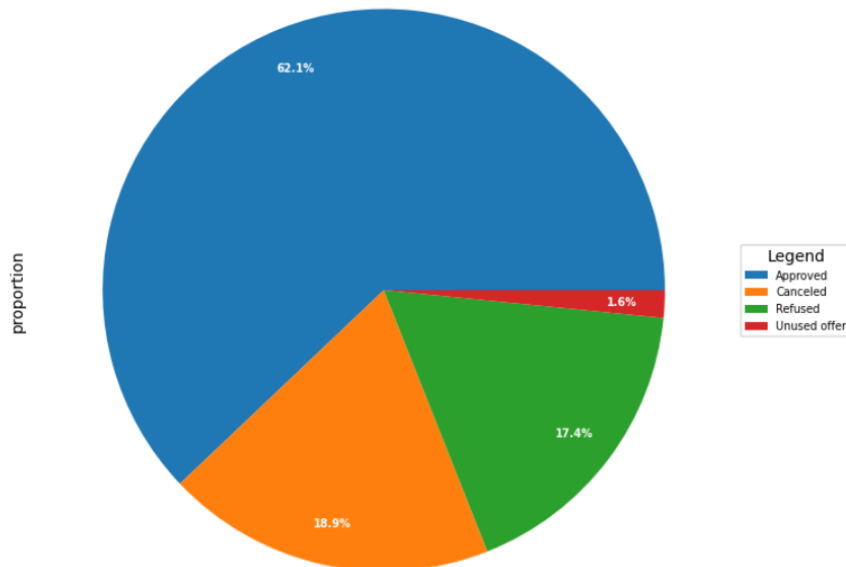
NAME_CLIENT_TYPE

Univariate Analysis

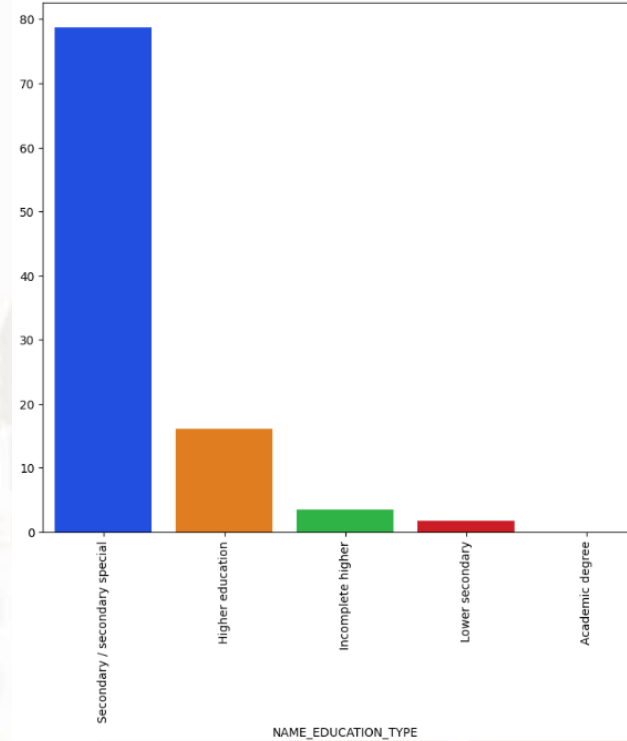
NAME_CONTRACT_STATUS

Majority of the loan applicants nearly 62% has been considered for **loan approval**, however the trends of **Refusal and Cancellation** are similar to each other

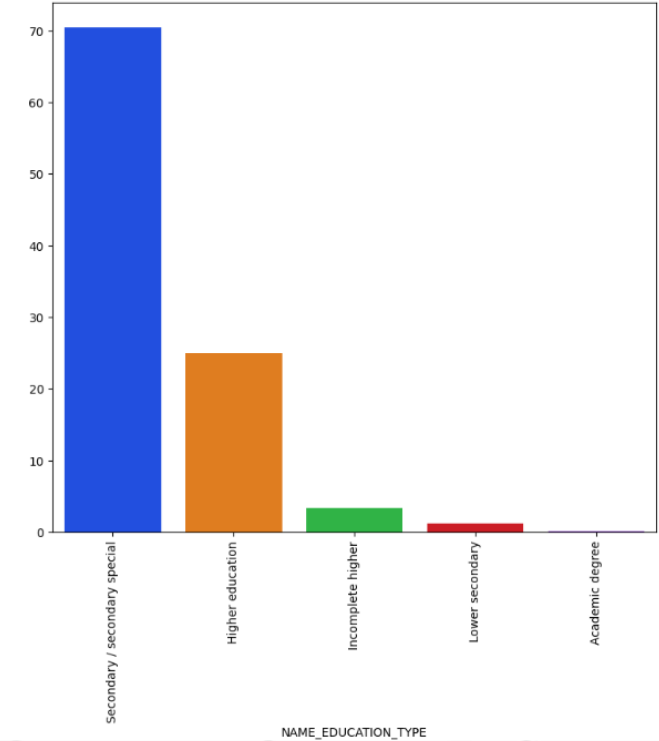
NAME_CONTRACT_STATUS Value Counts



PAYMENT WITH DIFFICULTY



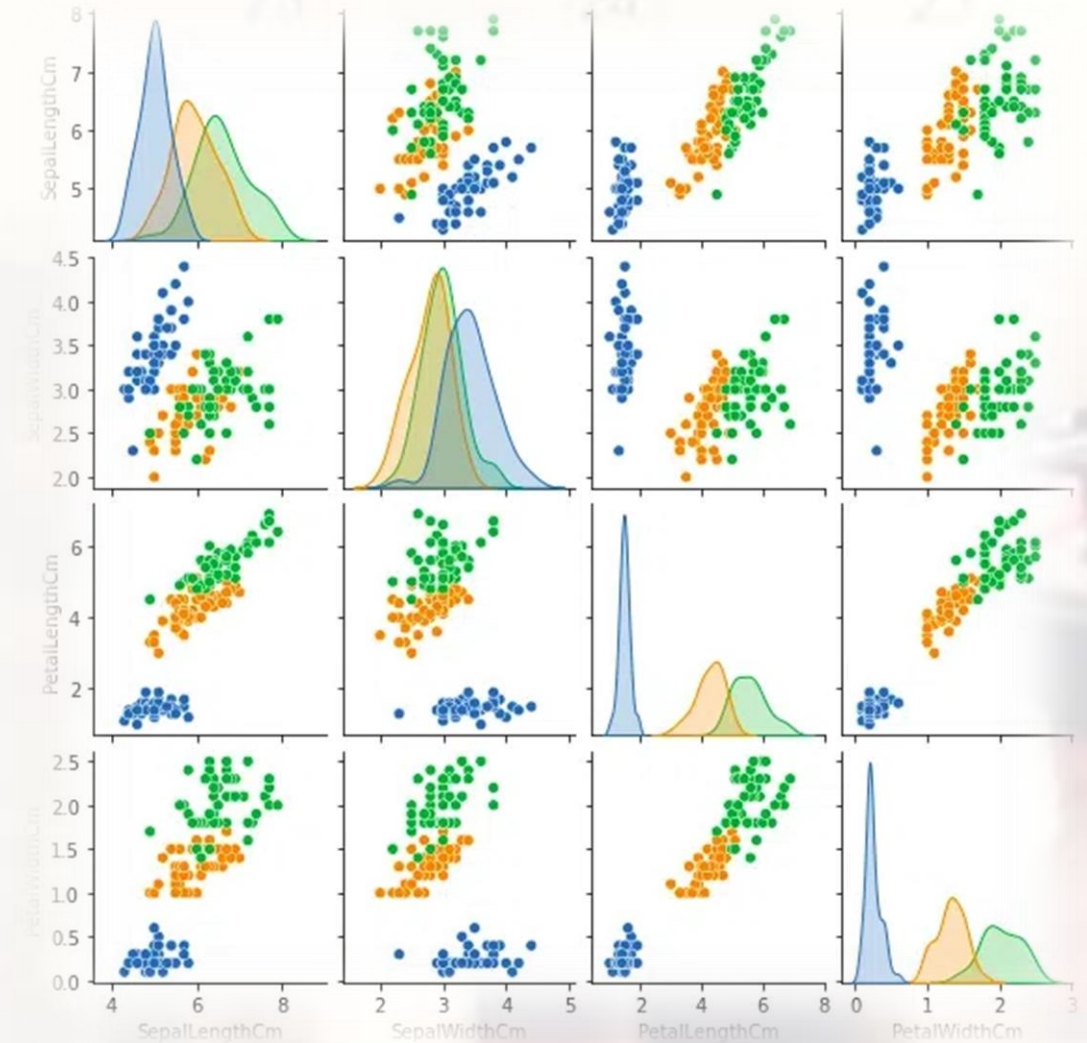
PAYMENT ON TIME



NAME_EDUCATION_TYPE

Secondary / Secondary Special facing payment difficulties whereas on the other hand **Higher education people** are doing payment on time

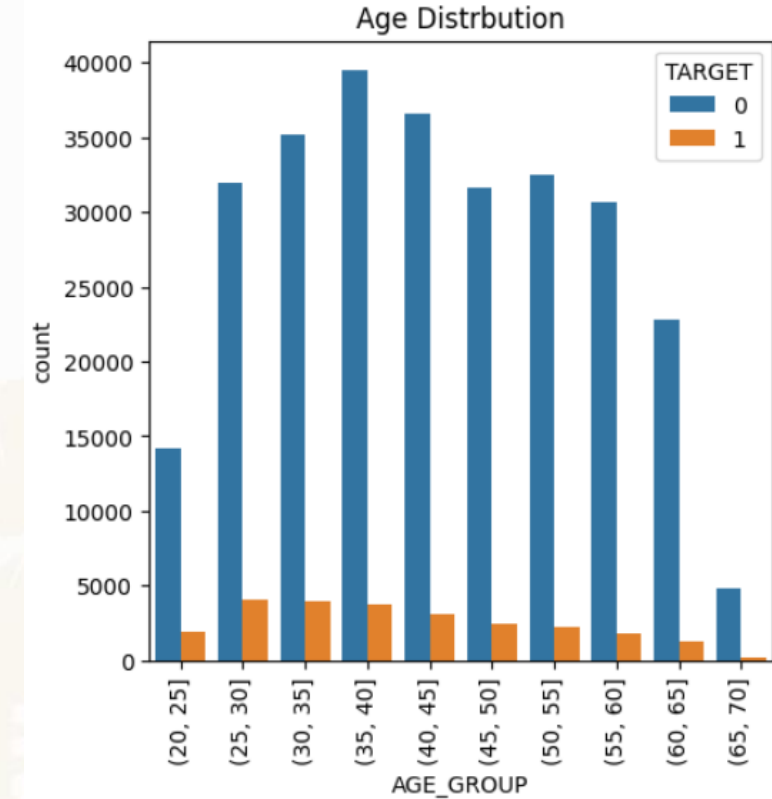
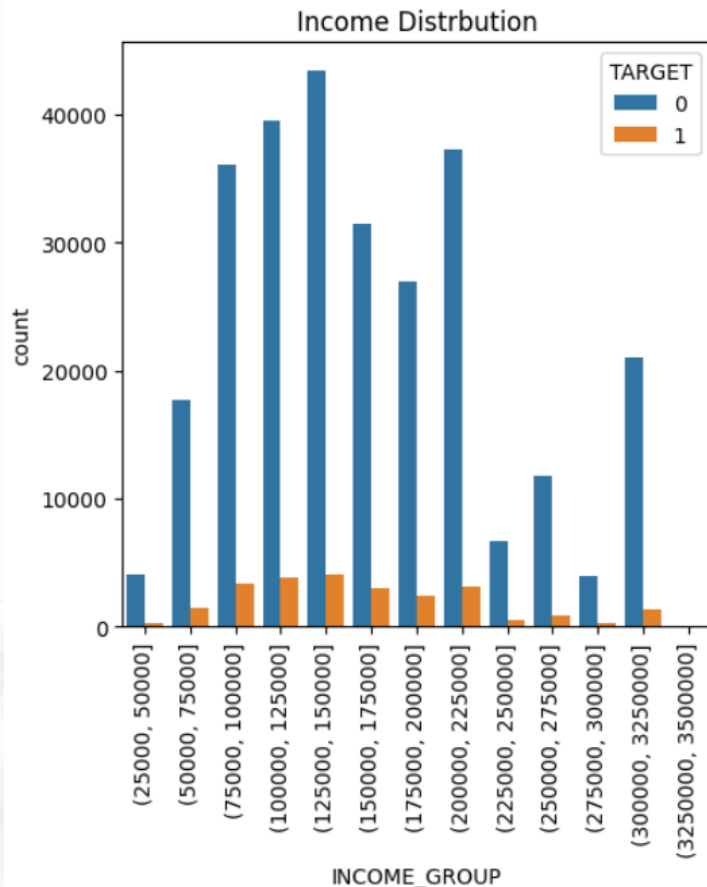
BIVARIATE ANALYSIS



Bivariate Analysis

INCOME_DISTRIBUTION

As the customers income increases percentage of difficulty in payments decreases. Applicants having income range of **1.25 l to 1.50 l** are doing payment on time maximum count in compared to others



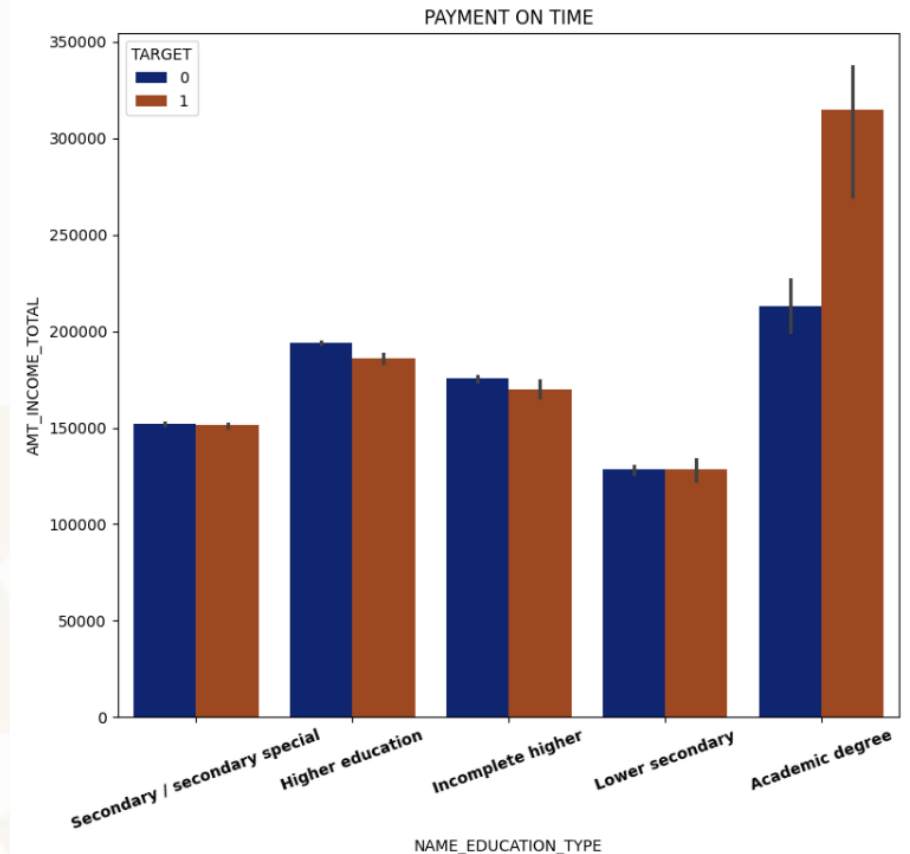
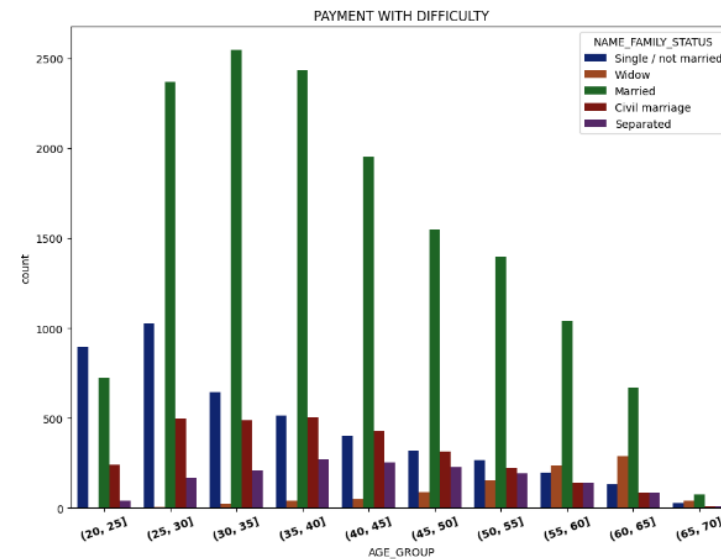
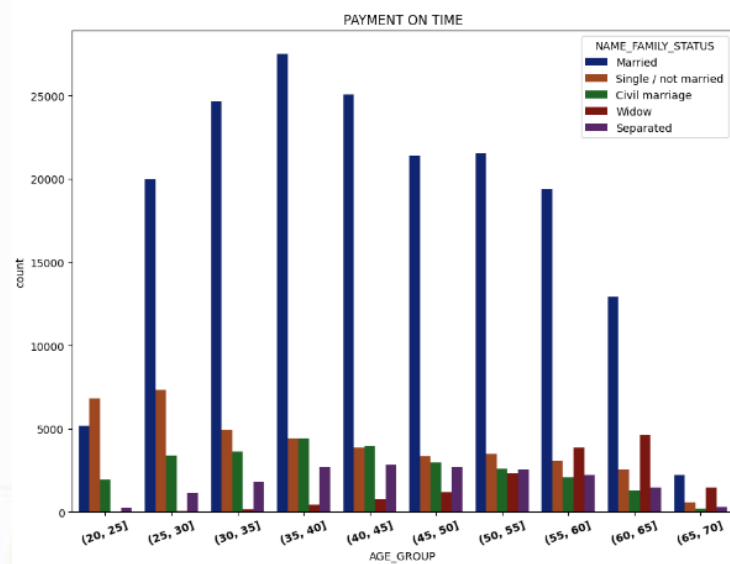
AGE_DISTRIBUTION

Customer with age in the range of **35-40** are doing their payments on time, where as customers in the of **25-30** are facing difficulty doing payments on time. Age greater than 40, percentage of payment with difficulty decreases

Bivariate Analysis

AGE_GROUP vs NAME_CONTRACT_STATUS

Married people of age group (35-40) are safe in terms of repayment single people of age group (25 -30) having tendency of **payment difficulties**

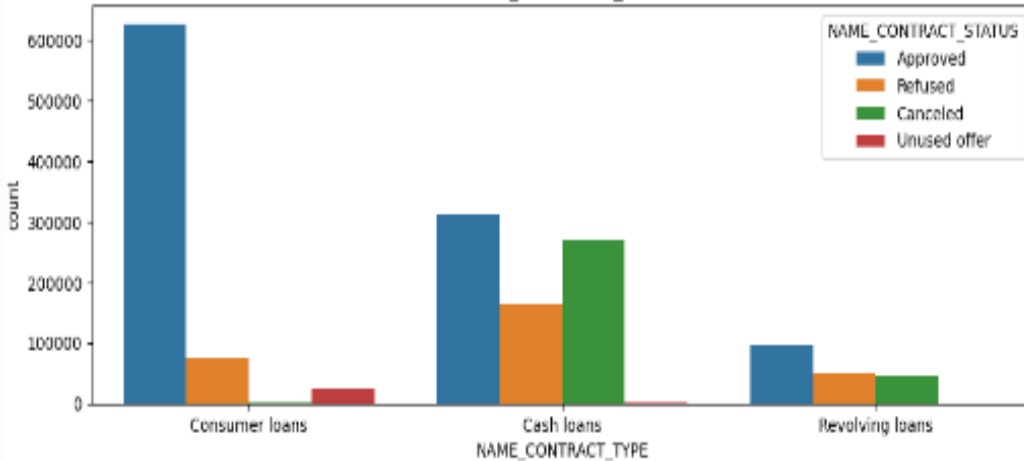


NAME_EDUCATION_TYPE vs AMT_INCOME_TOTAL

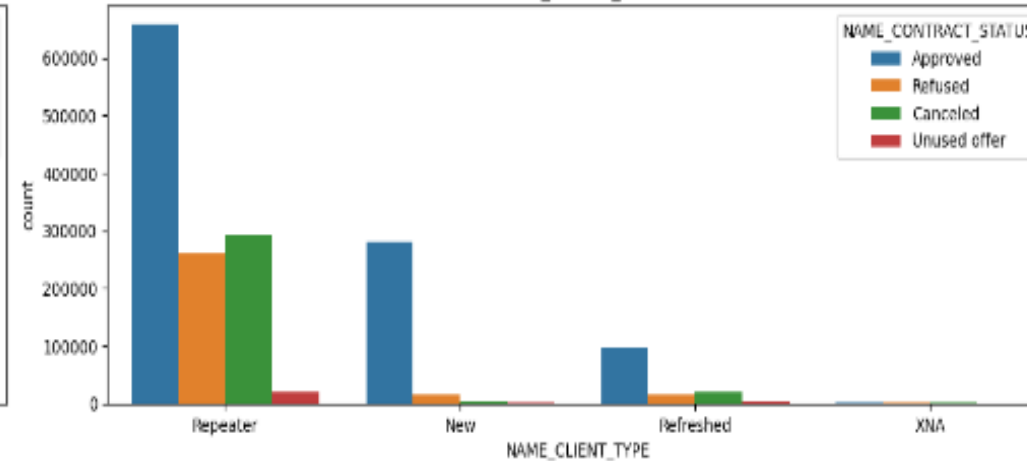
- Academic Degree holders** having income above 30K coming under payment difficult with higher risk
- Lower secondary** with having income **below 15k** are coming under lower risk payment difficulties category

Bivariate Analysis

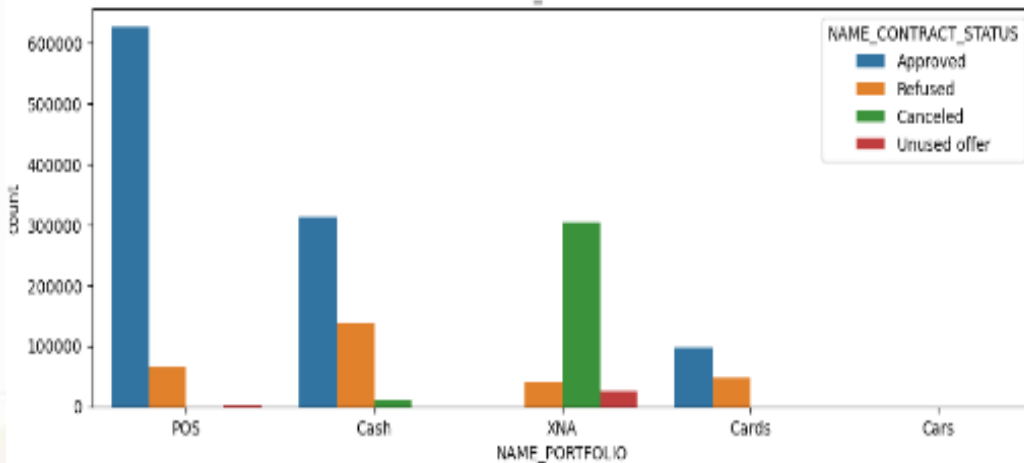
NAME_CONTRACT_TYPE



NAME_CLIENT_TYPE



NAME_PORTFOLIO



Consumer Loan Approval Prevalence

- Majority of approved applicants are for consumer loans.

No Cancelled Cash Loans

- Absence of cancelled loans in the cash loan category, unlike consumer loans.

Refusal Rate Discrepancy

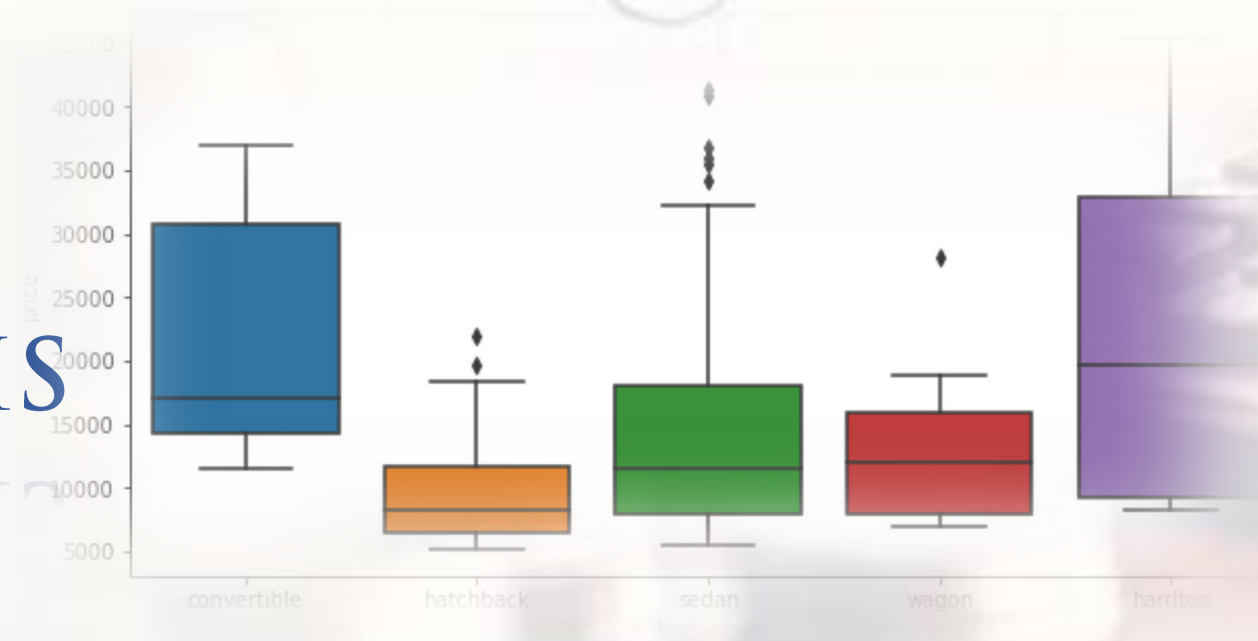
- Cash loans experience higher refusal rates compared to consumer loans.

Repeat Applicants Across Categories

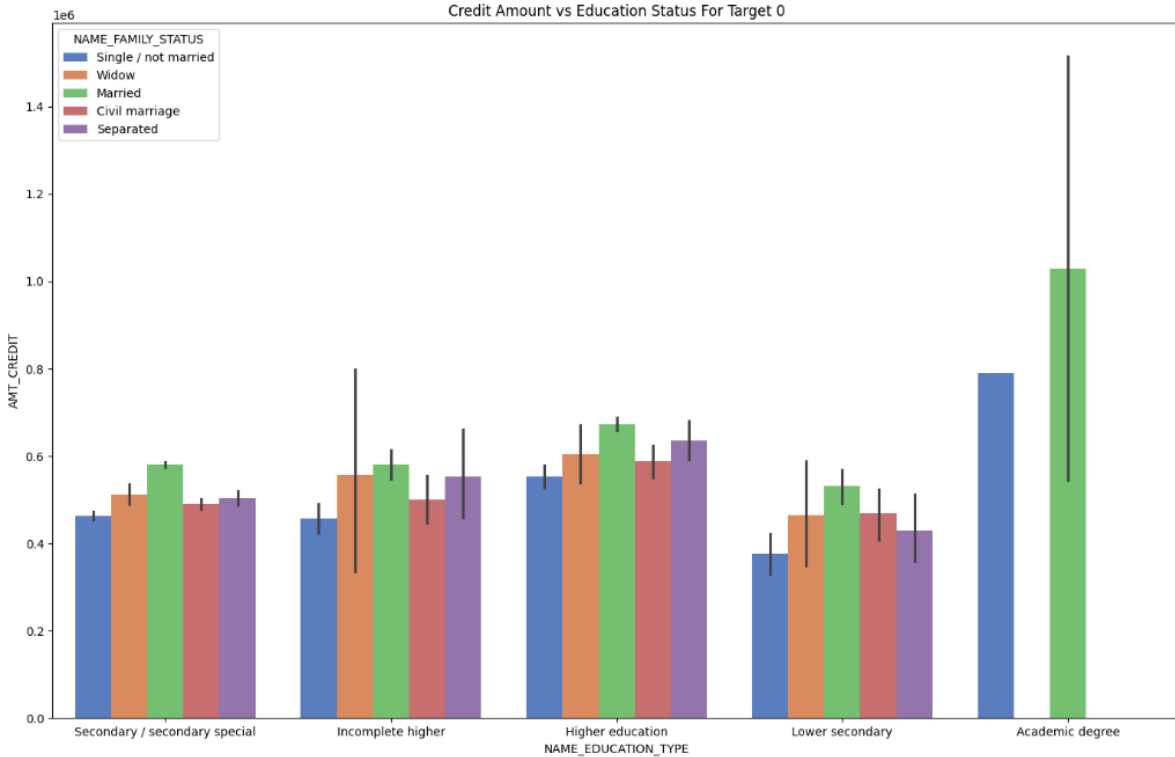
- Higher number of repeat applicants observed across all categories.
- POS Transactions Influence - POS transactions are associated with consumer loans, contributing to a higher refusal rate in cash loans.

NAME CONTRACT STATUS vs NAME_CONTRACT_TYPE,
NAME_PORTFOLIO, NAME_CLIENT_TYPE

MULTIVARIATE ANALYSIS



Multivariate Analysis

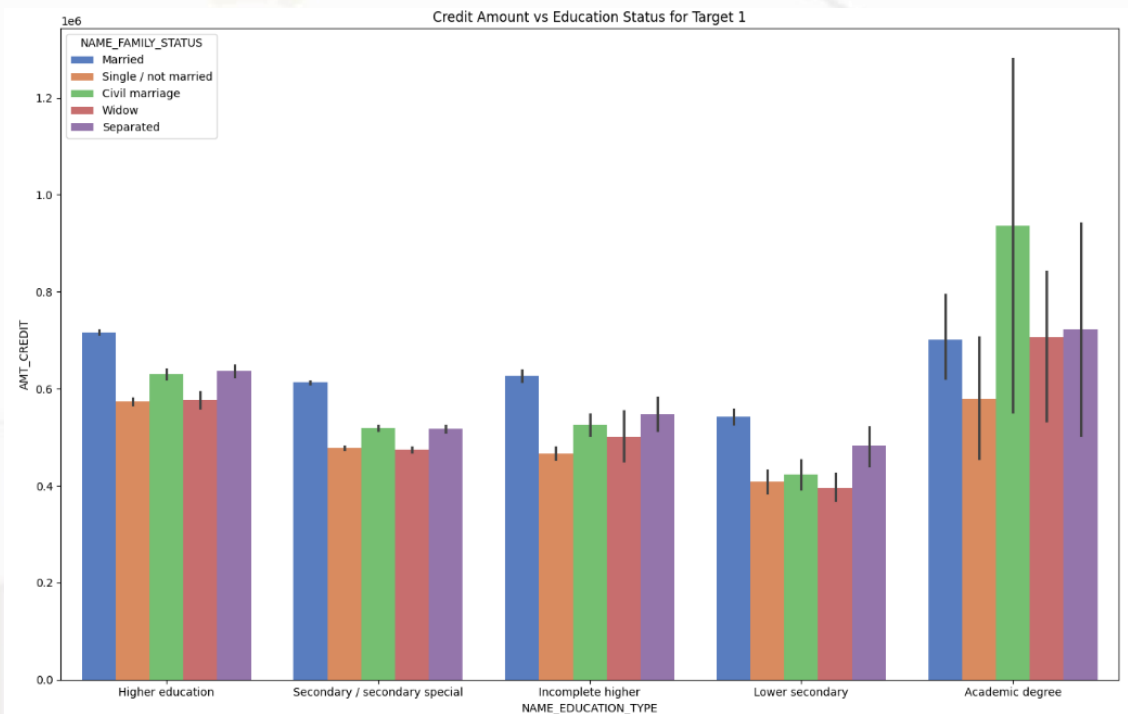


Non-Defaulters

- ✓ Academically educated, civilly married customers tend to have higher credit amounts.
- ✓ Lower education levels are associated with lower credit amounts, except for certain marital statuses.

Defaulters

- ✓ Married, academically educated customers often have higher credit amounts.
- ✓ Lower education levels typically correlate with lower credit amounts and limited family status diversity.

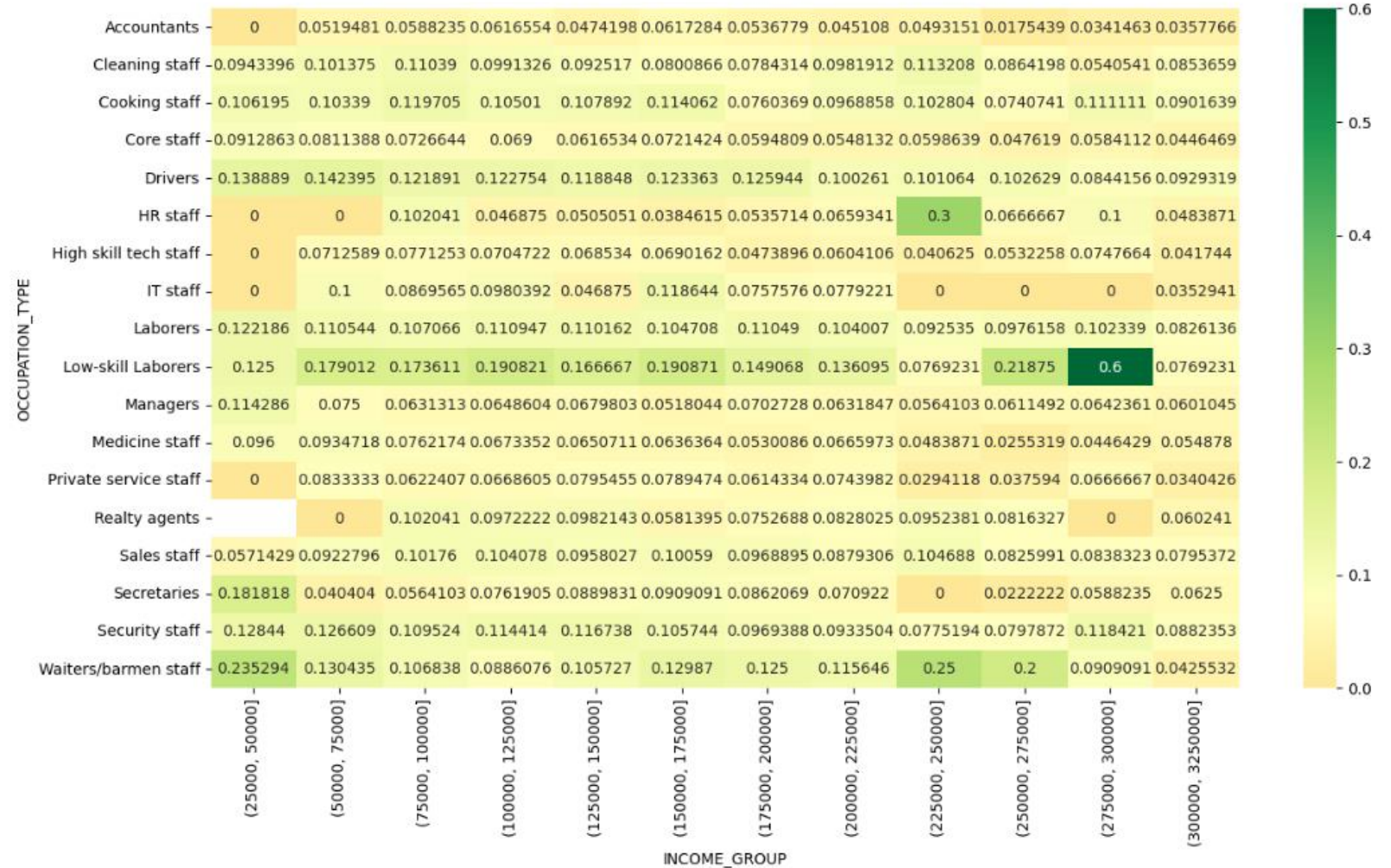


CREDIT_AMOUNT vs EDUCATION_STATUS vs TARGET

Multivariate Analysis

Low-skill labourers and **waiters/barmen staff** with income in **275k -300k** facing maximum percentage of payment difficulty seems not true income range provided is incorrect.

OCCUPATION_TYPE vs INCOME_GROUP vs TARGET



Multivariate Analysis

NAME_EDUCATIONAL_TYPE vs AGE_GROUP vs TARGET

Lower secondary people

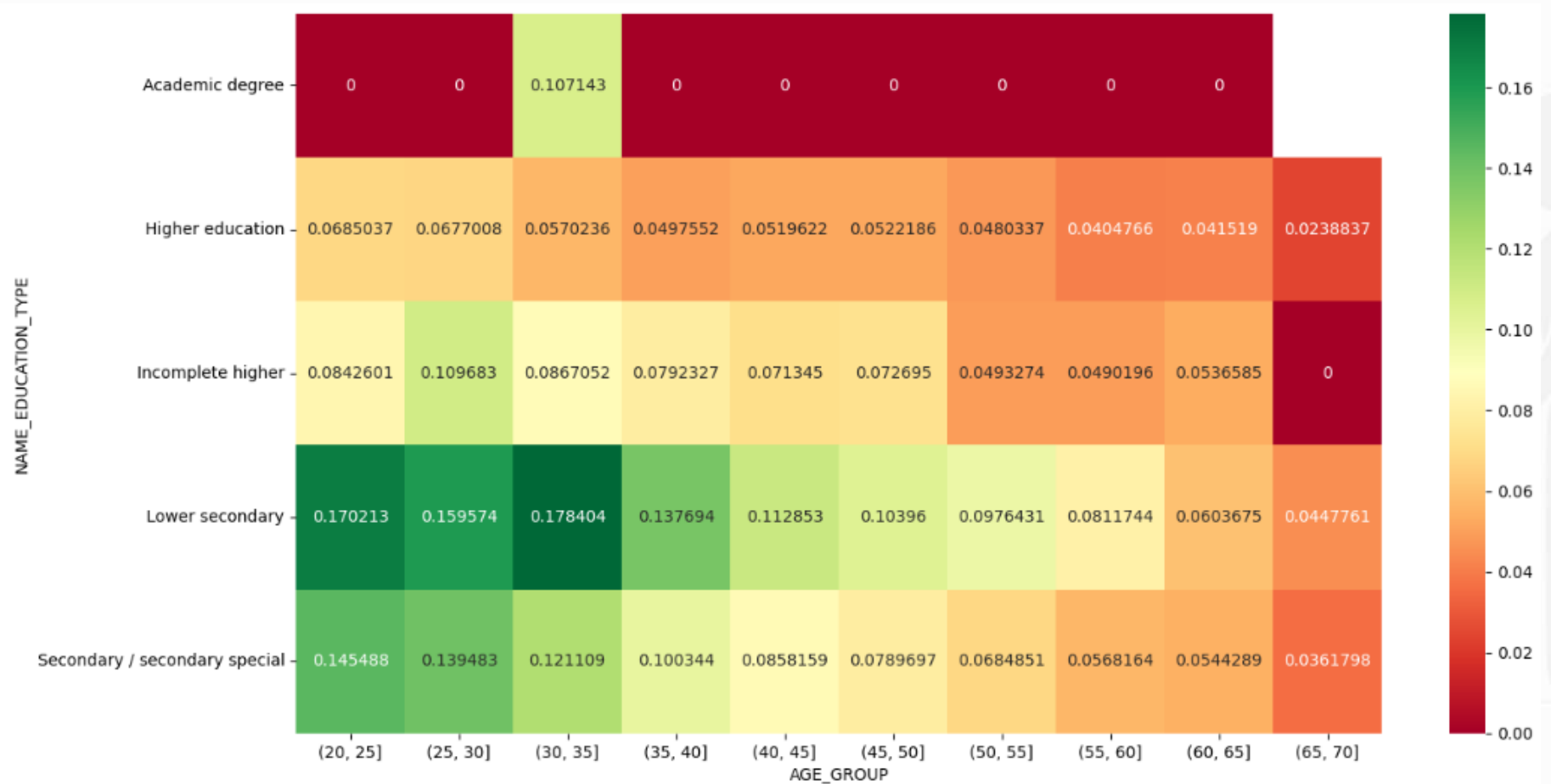
being in age group of 20 - 40 is

facing challenge with payment

where as having **academic**

degree are much more safer in

terms of timely payment



Conclusion

Reliable Customers

- ✓ Accountants, retired individuals, and students tend to make payments on time.
- ✓ Higher education or academic degree holders are favourable targets for lenders.
- ✓ Customers employed for over 10 years and those aged over 40 are considered reliable borrowers.

Less Reliable Customers

- ✓ Drivers, low-skill labourers, and those employed for less than 10 years often face payment difficulties.
- ✓ Customers under 30 and those with lower secondary education may encounter payment challenges.
- ✓ Single or unmarried individuals are less reliable borrowers.

Not Recommended

- ✓ Working-class and unemployed customers should be avoided due to their payment difficulties.