upGrad

Credit EDA Case Study

Shreshtha Doharey

Introduction

upGrad

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

Objective

upGrad

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.

Overview

upGrad

Context

- ✓ Loan companies struggle to approve loans due to applicants' insufficient or non-existent credit history.
- ✓ Some consumers exploit this by defaulting on loans.
- ✓ Your task is to help a consumer finance company specializing in urban loans analyze data to ensure capable applicants are approved.

Risks

False Rejection

Not approving a loan for an applicant likely to repay results in lost business.

False Approval

Approving a loan for an applicant likely to default results in financial loss.

Data Scenarios

Clients with Payment Difficulties

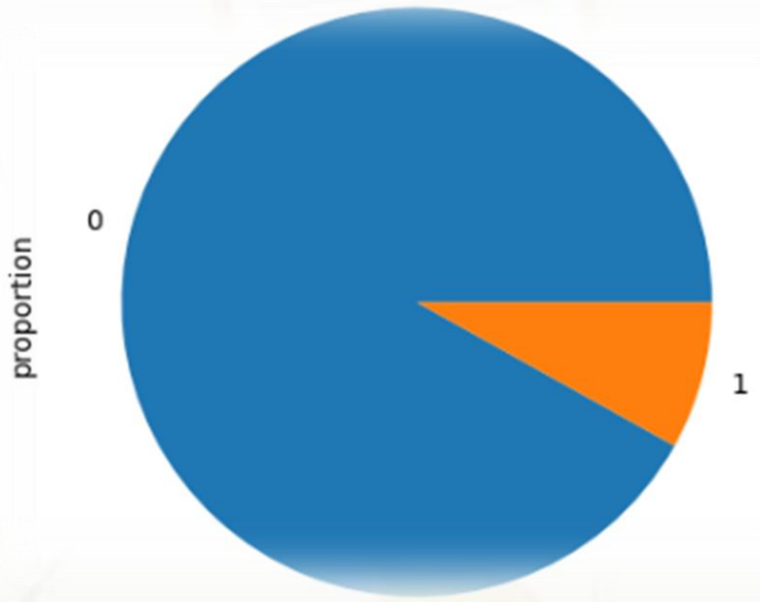
Late payment more than X days on at least one of the first Y installments.

Clients with Timely Payments

Payments made on time in all other cases.

Imbalance

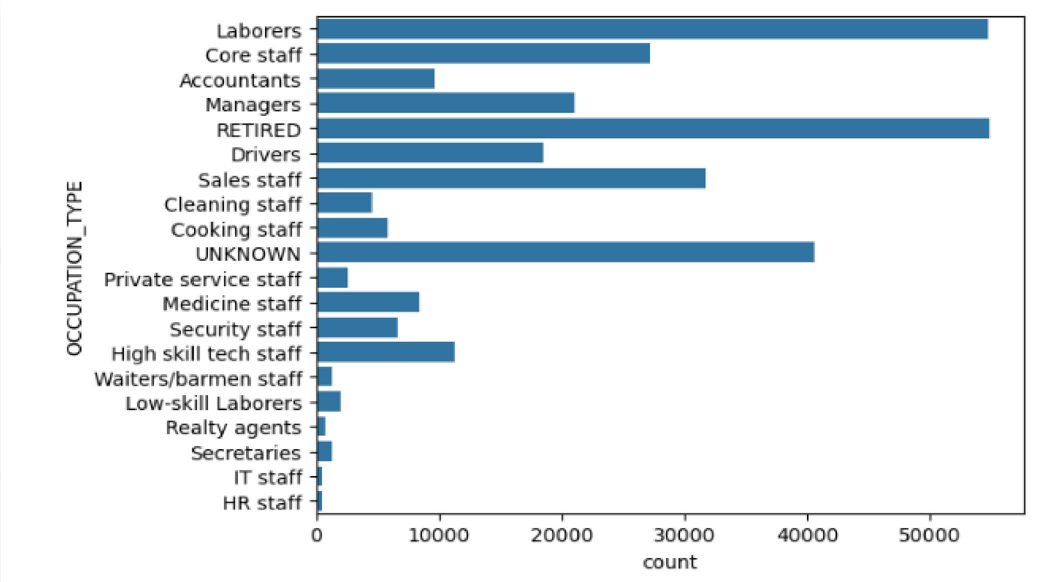
upGrad

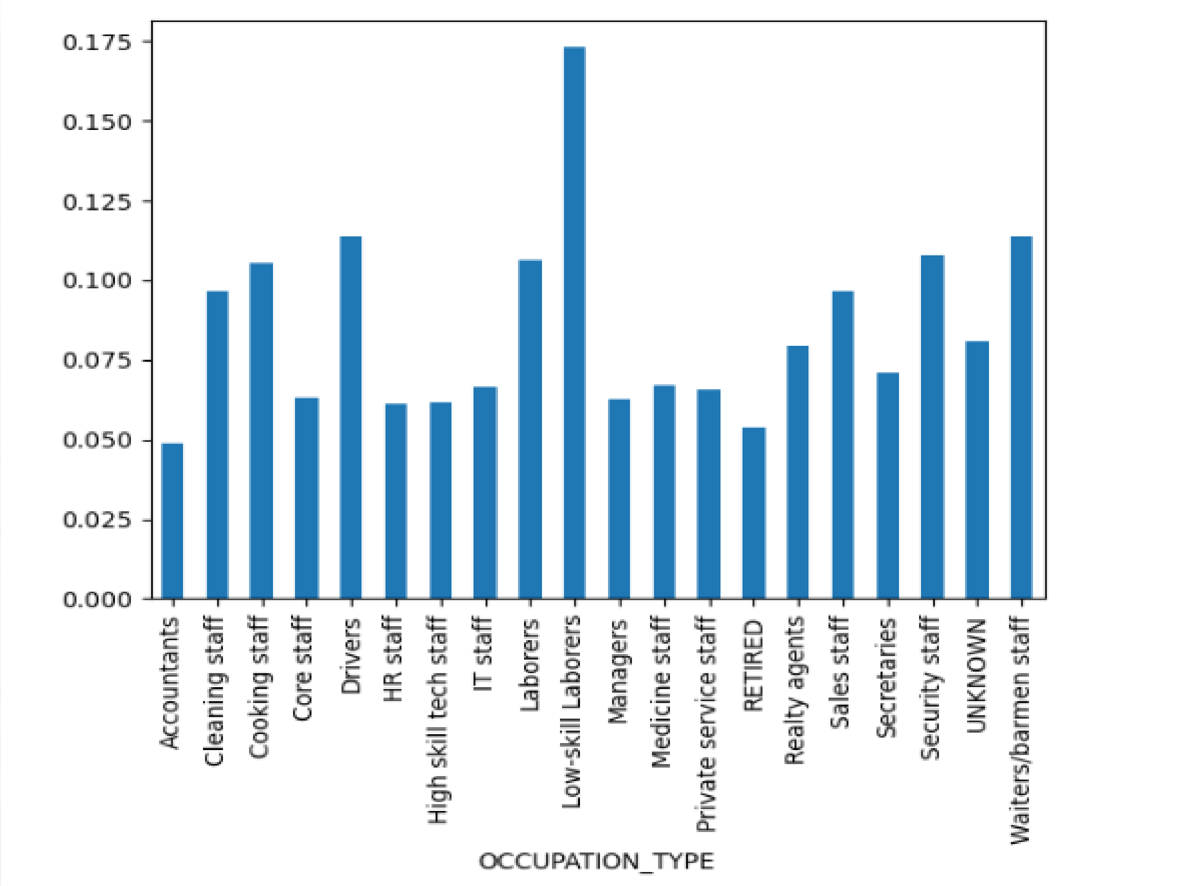


Data imbalance ratio is approximately 20:1 which means we have 20 sample of customer doing payment on time in comparison of 1 sample of customer facing payment difficulty.

Univariate Analysis

upGrad





✓Laborers are highest in the applicants list followed by Retired applicants.

✓Graph shows mean payment difficulty by occupation type

✓Low-skill laborers face the highest payment difficulty (~17.5%).

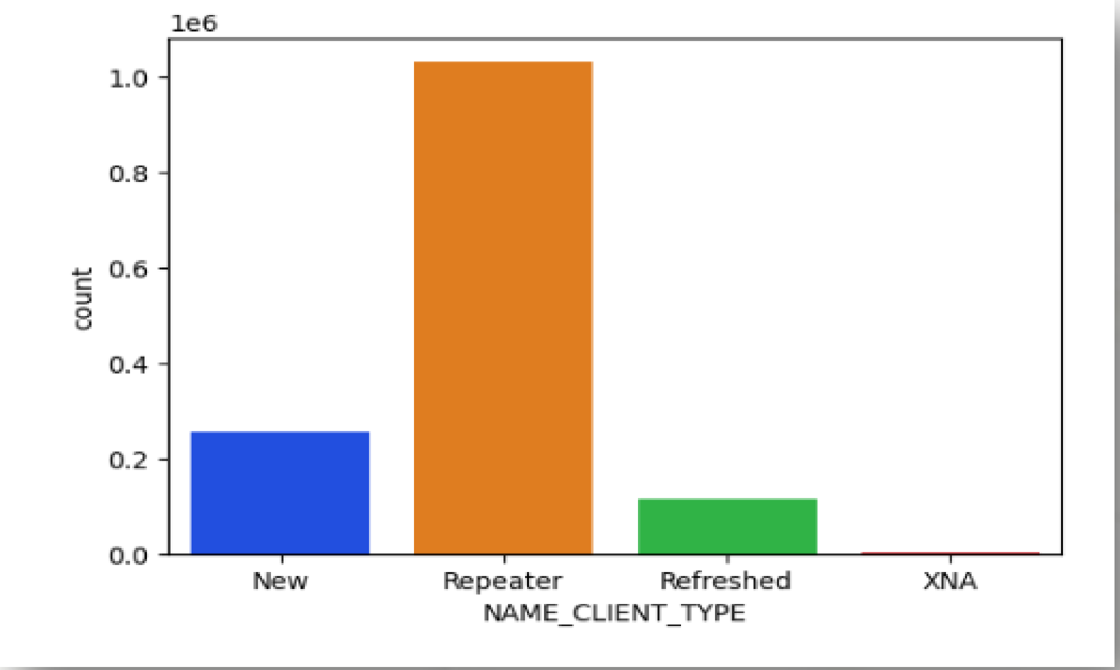
✓Accountants and retired people mostly pay on time.

OCCUPATION_TYPE

Univariate Analysis

NAME_CLIENT_TYPE

✓Repeater are more as compared to the new customers



FLAG_OWN_CAR	FLAG_OWN_REALTY	TARGET	
0	0	0	90.946859
		1	9.053141
	1	0	91.702857
		1	8.297143
1	0	0	92.923058
		1	7.076942
	1	0	92.642271
		1	7.357729

Name: proportion, dtype: float64

✓Image shows payment difficulty percentages based on car or house ownership.

✓Having a car or house does not significantly affect payment difficulties.

✓In all cases, the percentage is around 8%.

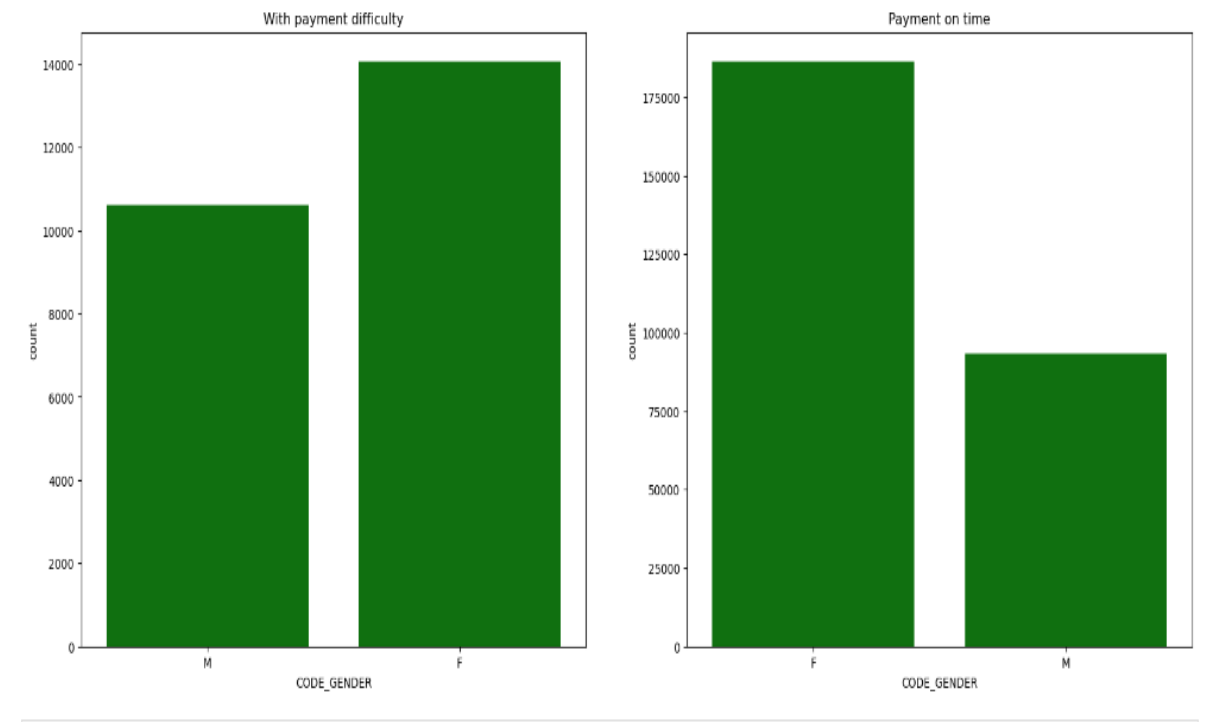
FLAG_OWN_CAR

FLAG_OWN_R EALTY

Univariate Analysis

CODE_GENDER

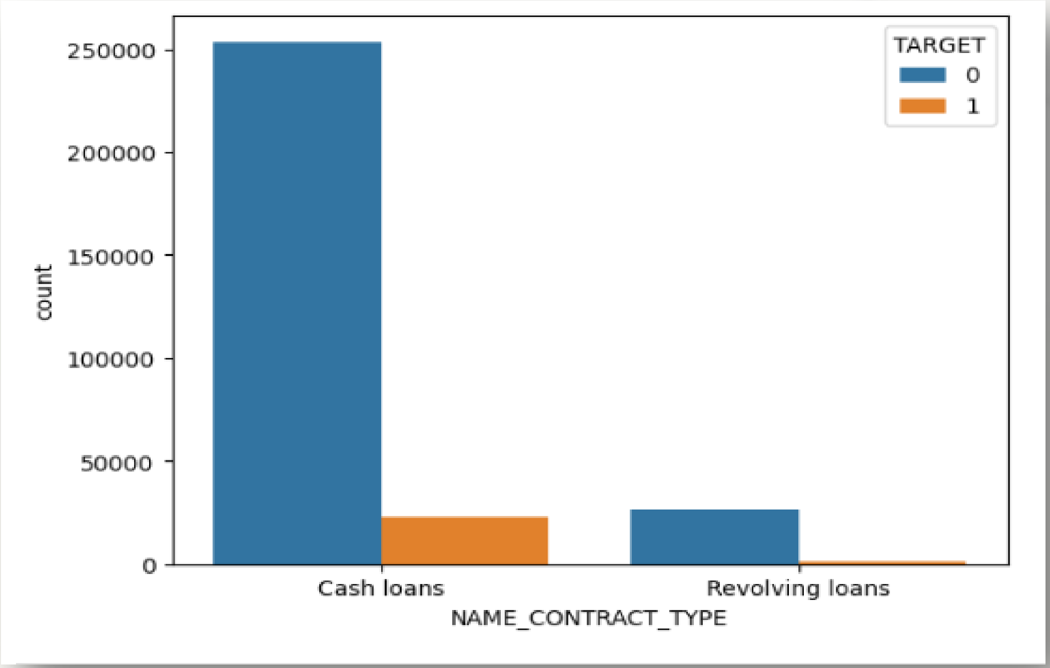
CODE_GENDER	TARGET	
F	0	92.985545
	1	7.014455
M	0	89.808256
	1	10.191744



✓Female applicants outnumber male applicants.

✓Payment difficulty: ~10% for males, ~7% for females.

✓Payment difficulties are similar across genders, with minimal difference.



✓There is not much percentage difference between cash loans and revolving loans regarding payment difficulties.

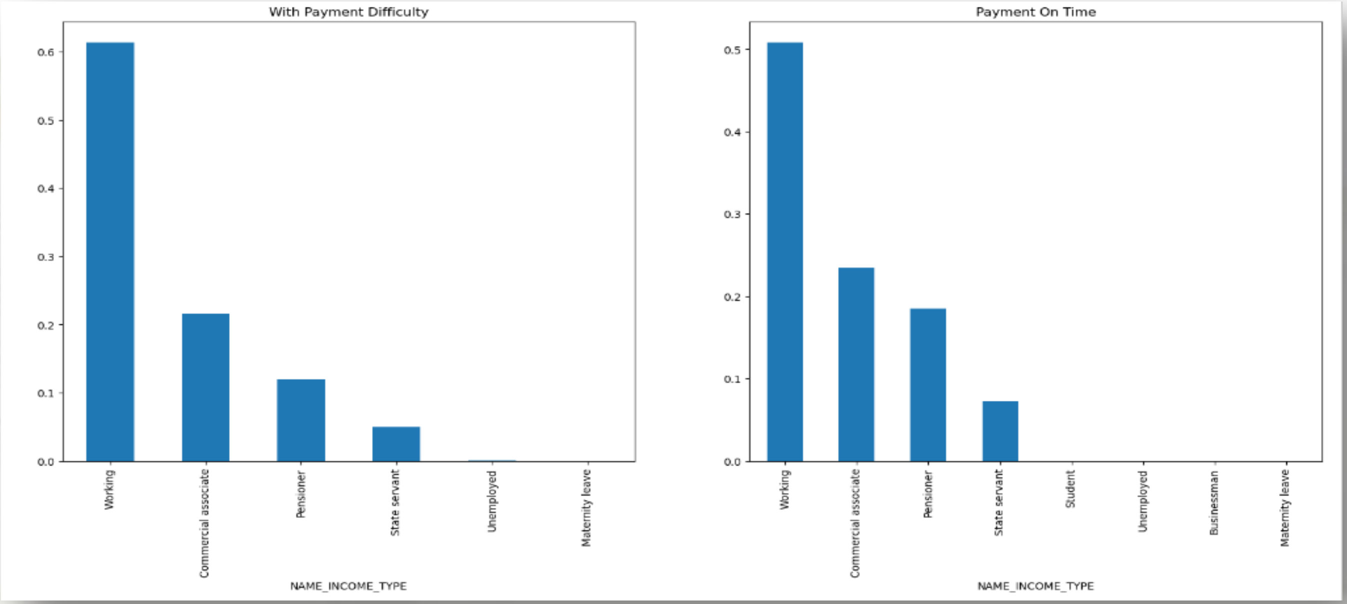
✓In cash loans, 8% of customers experience payment difficulties, while in revolving loans, there are 5%.

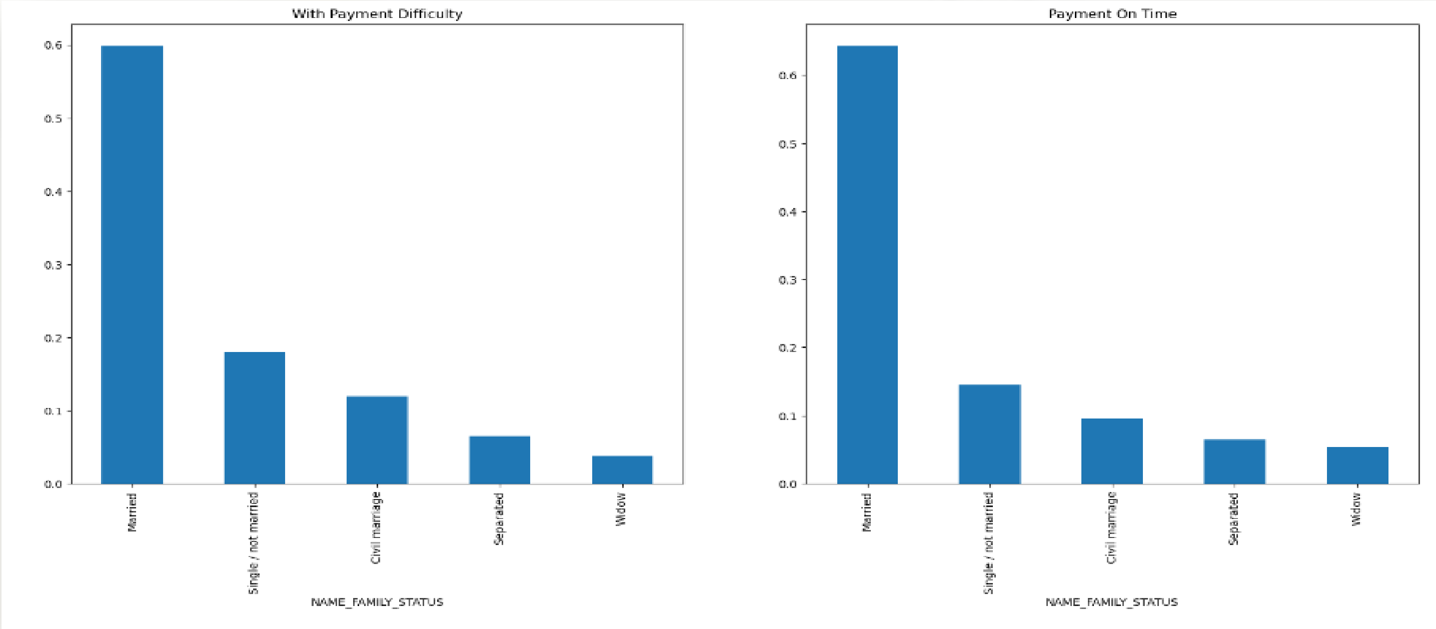
NAME_CONTRACT_TYPE

Bivariate Analysis

NAME_INCOME_TYPE

- ✓ Pensioners and commercial associates mostly make payments on time.
- ✓ Working class customers face the highest payment difficulties.
- ✓ Businessmen, unemployed individuals, students, and those on maternity leave have zero instances of timely payments, but their numbers are few, showing weak correlation.



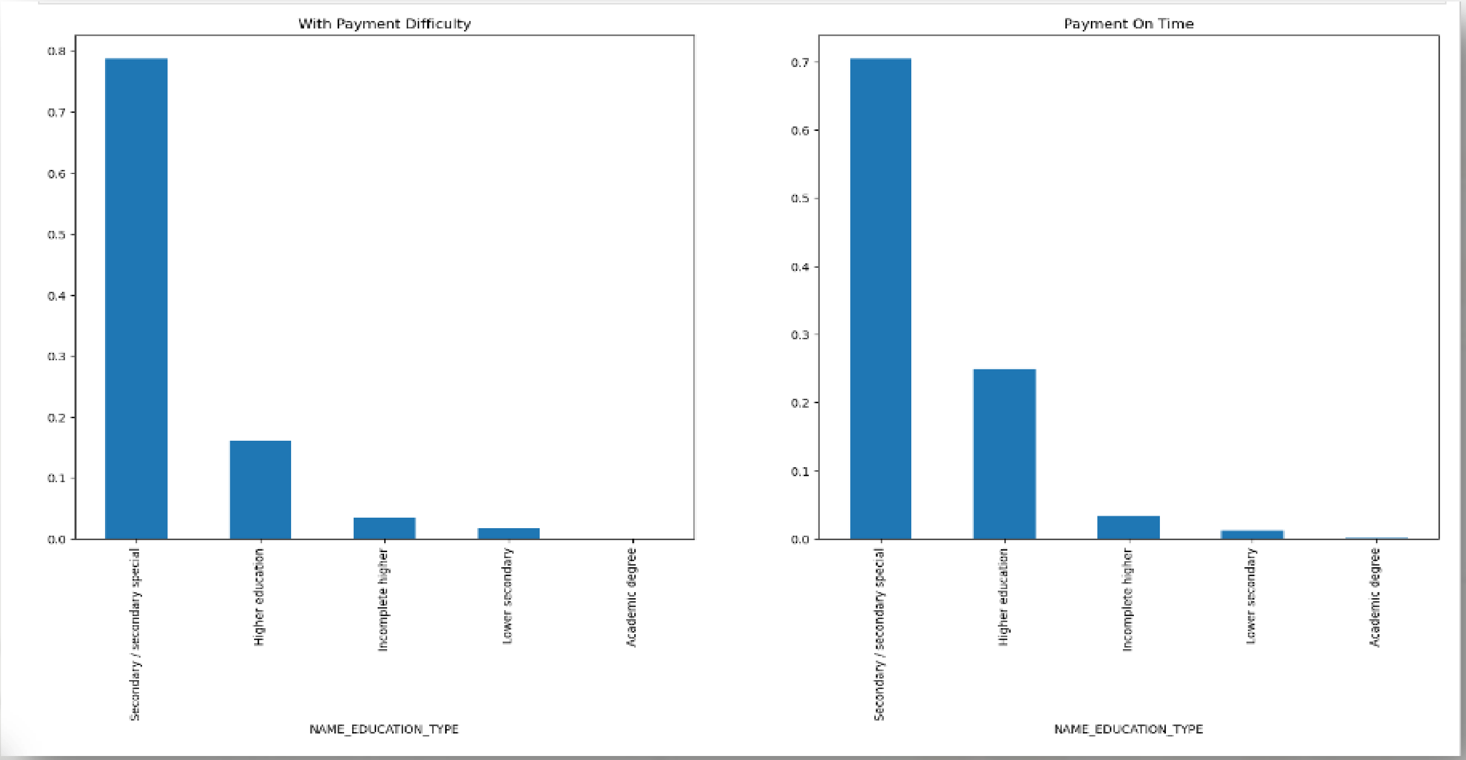


- ✓ Singles or those not married are facing payment difficulties.
- ✓ Among married individuals:
  - 59% are facing payment difficulties.
  - 63% are making payments on time.

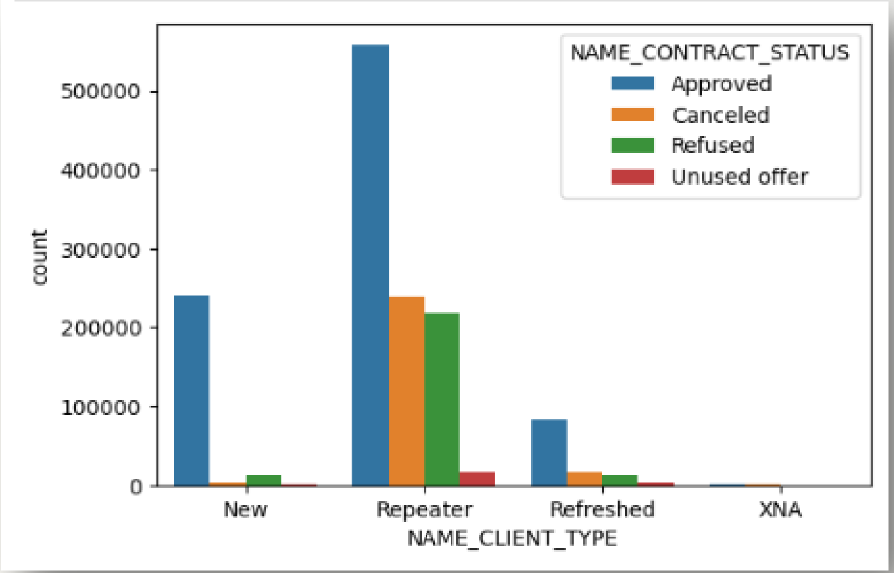
NAME_FAMILY_STATUS

Bivariate Analysis

NAME_EDUCATION_TYPE



✓ Applicants with the Higher education have done their payments mostly on time



✓ Repeaters got maximum number of approvals followed by new customers

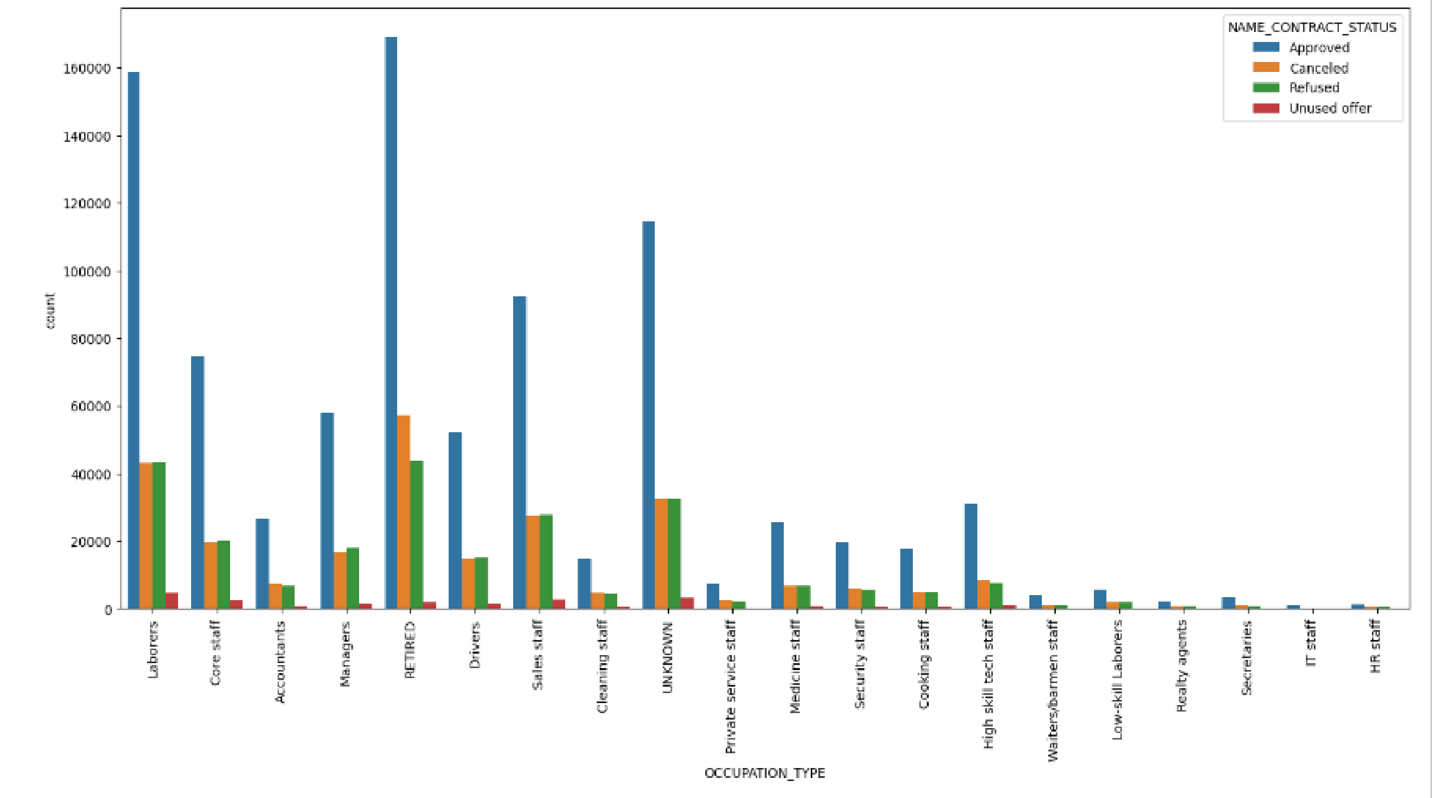
NAME_CLIENT_TYPE

vs

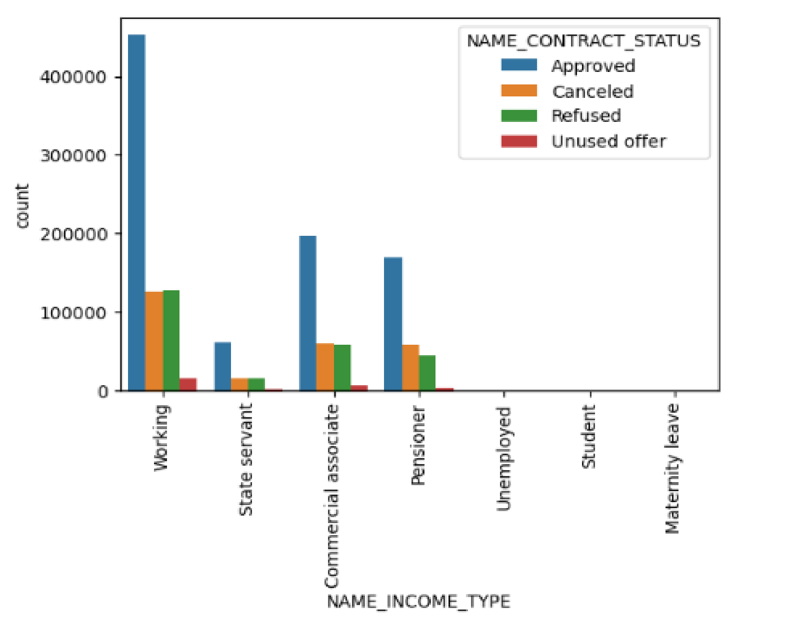
NAME_CONTRACT_STATUS

Bivariate Analysis

OCCUPATION_TYPE vs NAME_CONTRACT_STATUS



- ✓ Retired customers have got maximum number of approvals followed by labourers



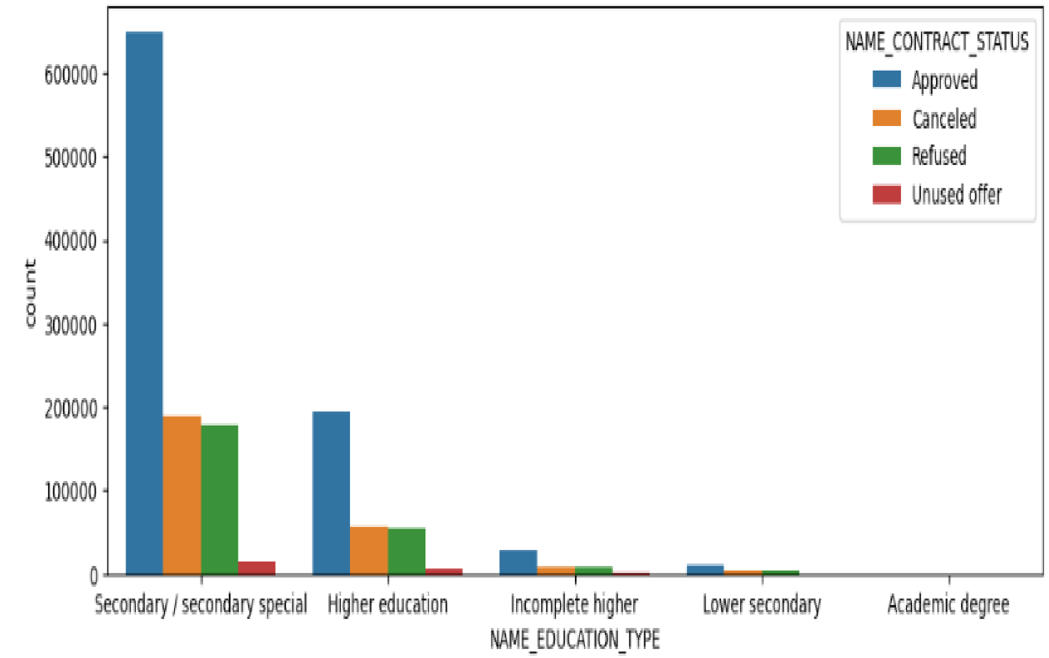
- ✓ Student approval rate is highest at 83%, followed by civil servants at 65%.
- ✓ The highest refusal rate is among unemployed individuals at 30%.

NAME_INCOME_TYPE vs NAME_CONTRACT_STATUS

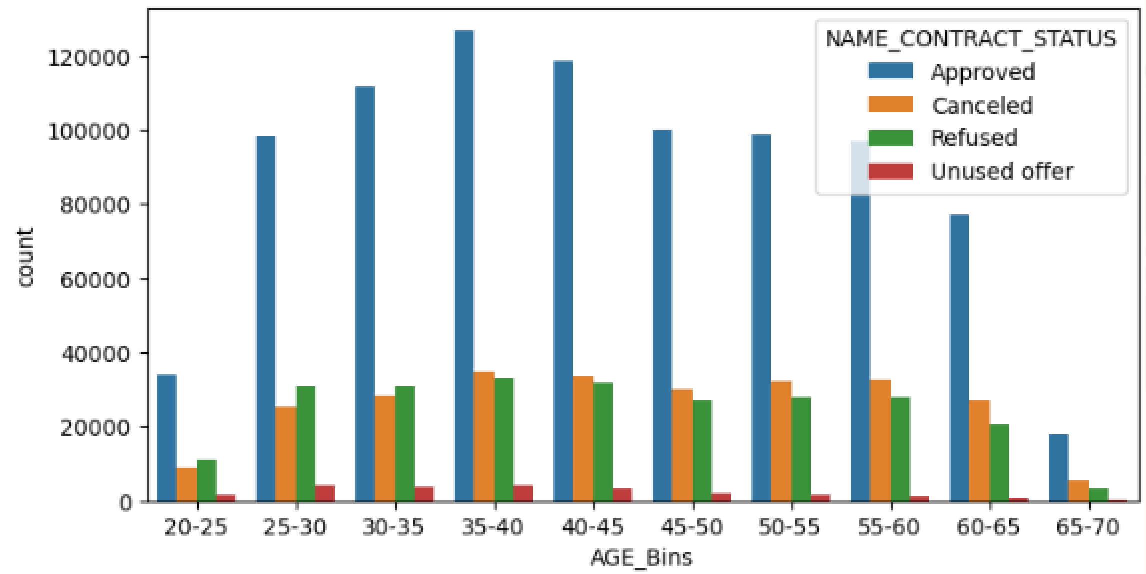
Bivariate Analysis

upGrad

NAME_EDUCATION_TYPE
vs
NAME_CONTRACT_STATUS



- ✓ Applicants with academic degree has maximum percentage of approval.



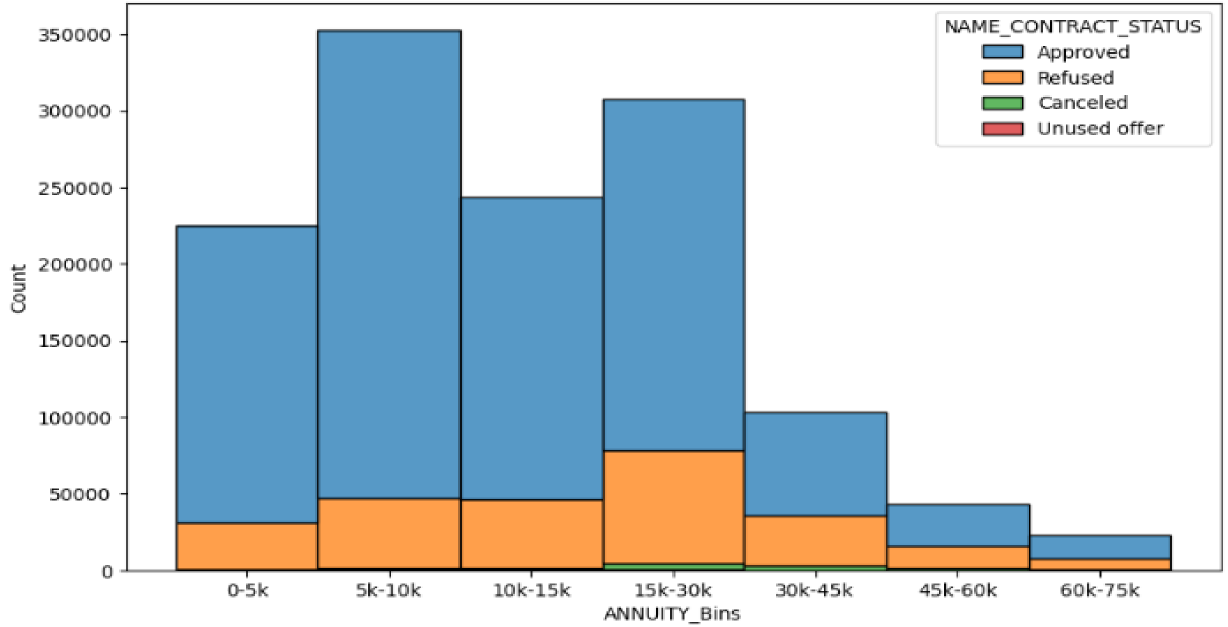
✓Percentage of approvals are maximum in the age range of between 30-45 and also surprisingly between 65-70 age range.

AGE_Bins
vs
NAME_CONTRACT_STATUS

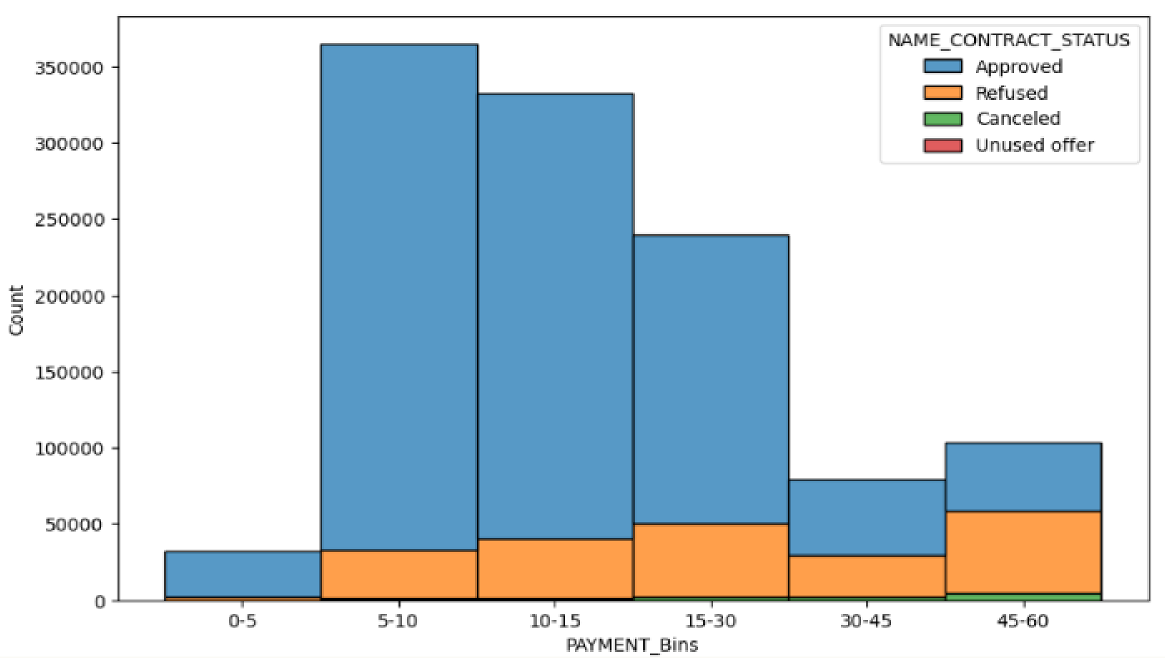
Bivariate Analysis

upGrad

ANNUITY_Bins
vs
NAME_CONTRACT_STATUS



- ✓ **AMT_ANNUITY** represents the monthly EMI for loan repayment.
- ✓ Customers with EMI less than 15,000 receive more loan approvals.
- ✓ Customers with EMI in the range of 15,000-30,000 face the highest number of refusals.



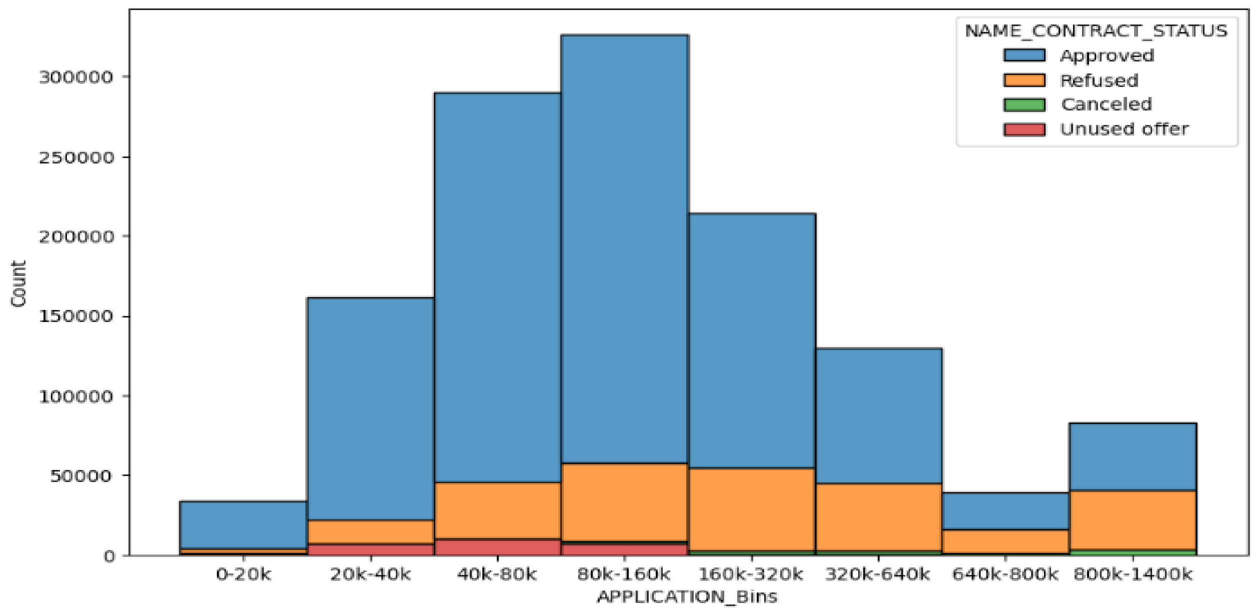
- ✓**CNT_PAYMENT** represents loan repayment duration.
- ✓Longer repayment durations have higher refusal rates.
- ✓Loans with repayment durations less than 10 years have higher approval rates.
- ✓Shorter repayment periods increase the likelihood of loan approval.

PAYMENT_Bins
vs
NAME_CONTRACT_STATUS

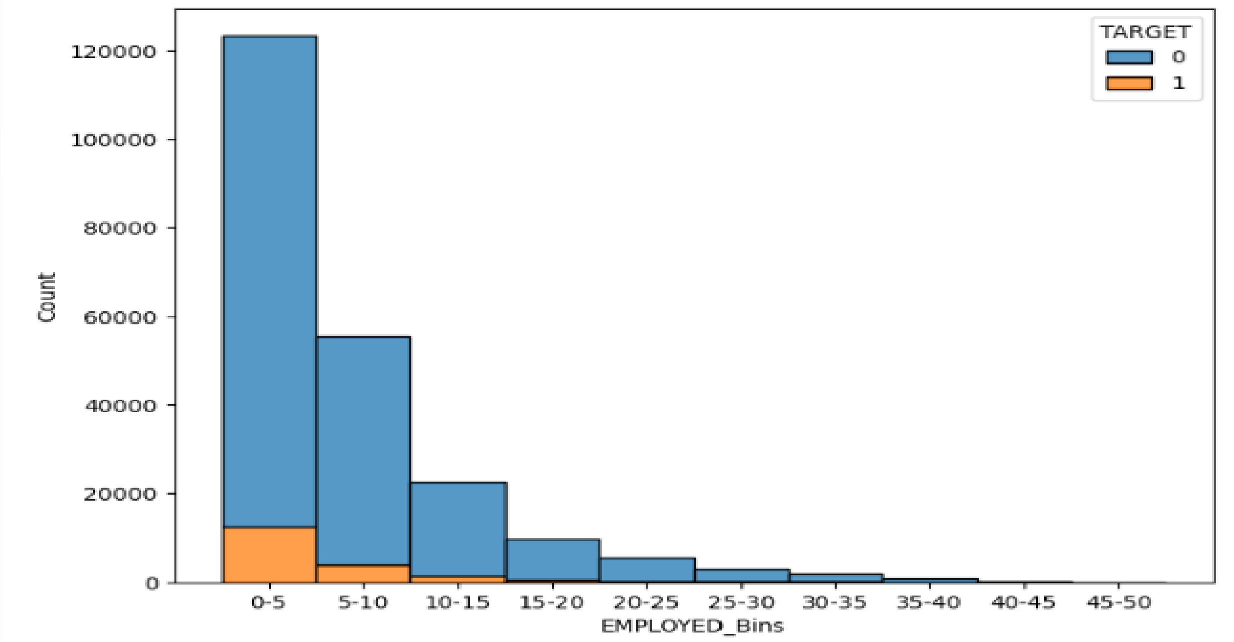
Bivariate Analysis

upGrad

APPLICATION_Bins
vs
NAME_CONTRACT_STATUS



- ✓ AMT_APPLICATION represents the loan amount requested by the customer.
- ✓ Refusals are more common for loan amounts greater than 160k.



✓Customers employed for less than 5 years face the most payment difficulties

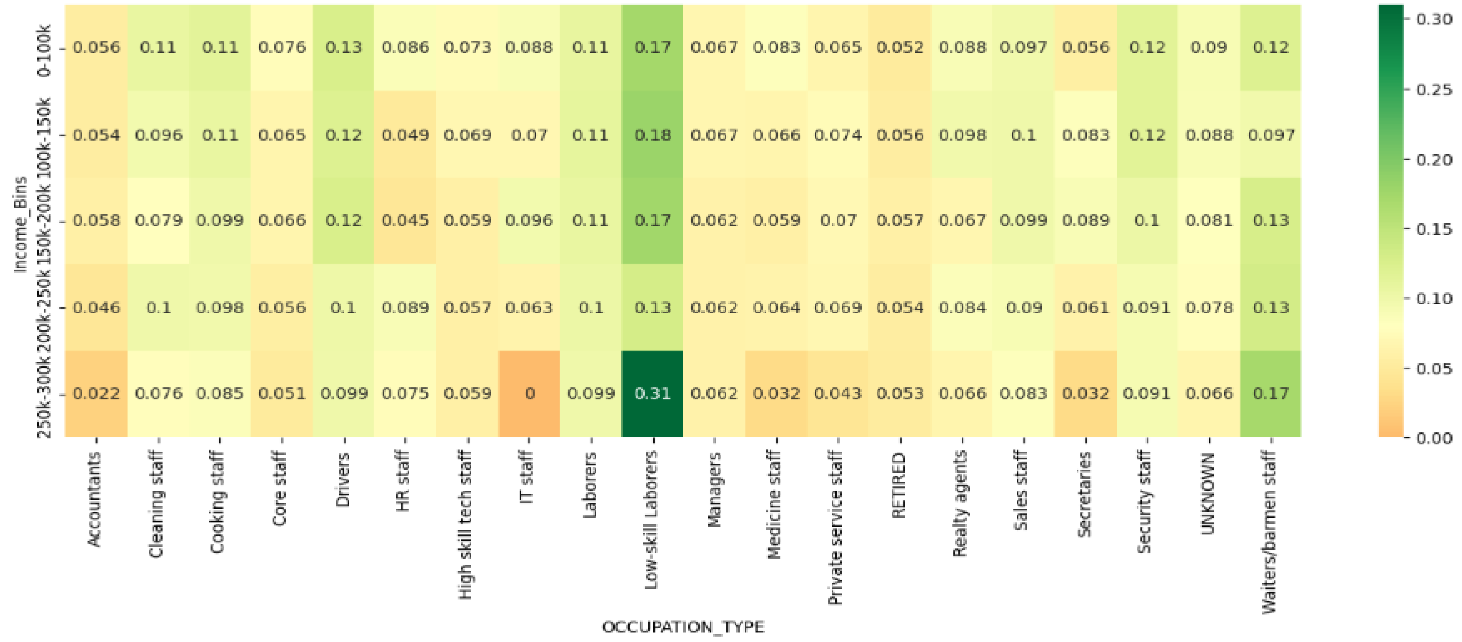
✓Customers employed for more than 10 years have a higher chance of repaying loans on time.

EMPOLYED_Bins
vs
NAME_CONTRACT_STATUS

Multivariate Analysis

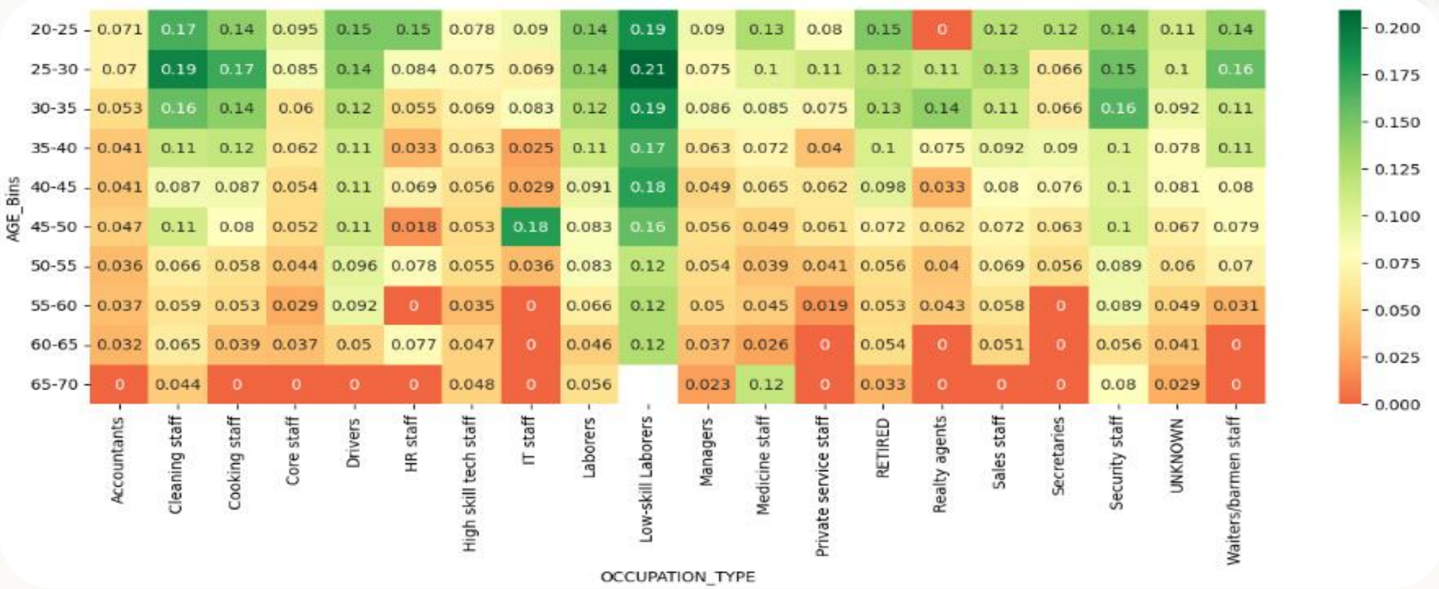
upGrad

INCOME_Bins
vs
NAME_CONTRACT_STATUS
VS
OCCUPATION_TYPE



✓Low-skill laborers and waiters/barmen staff face maximum payment difficulty.

✓Income range of 250k-300k seems incorrect for Low-skill-Laborers observation.



✓ Customers over 40 mostly make payments on time across all occupations.

✓Customers under 40 struggle with timely payments in specific occupations: cleaning staff, cooking staff, drivers, Laborers, low-skill laborers, security staff, waiters, sales staff

AGE_Bins
VS
NAME_CONTRACT_STATUS
VS
OCCUPATION_TYPE

Conclusion

upGrad

Customers to be targeted for Providing Loan:

- ✓ Customers in occupations such as Accountant, retired, and students tend to make payments on time.
- ✓Customers with higher education or academic degrees are favorable targets for lenders.
- ✓Customers who have been employed for more than 10 years should be considered.
- ✓Customers aged over 40 are considered reliable borrowers.

Customers to be avoided from Providing Loan:

- ✓ Customers in occupations such as drivers and low-skill laborers often face difficulties in making payments on time.
- ✓Customers with lower secondary education tend not to make payments on time.
- ✓It is advisable to avoid working-class and unemployed customers.
- ✓Customers who have been employed for less than 10 years should not be considered for loans.
- ✓Customers under the age of 30 frequently encounter difficulties in making payments