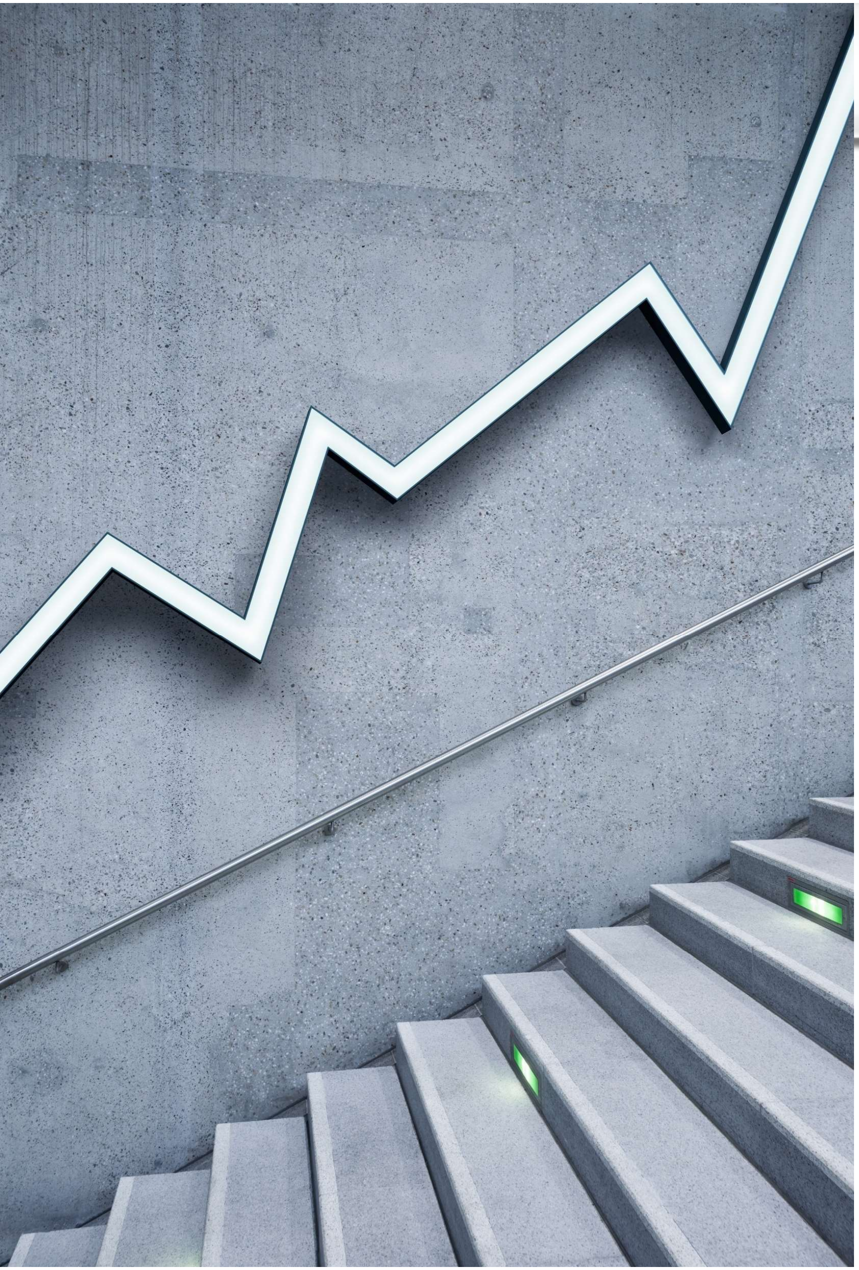


LOAN DEFAULT PREDICTION USING EXPLORATORY DATA ANALYSIS

IDENTIFYING KEY FACTORS LEADING TO LOAN DEFAULT

BY PRATHAM MHATRE AND PRAVIN GAIKAR

INTRODUCTION

PROBLEM STATEMENT

- The finance company needs to decide on loan approvals by predicting the likelihood of applicants defaulting.
- Incorrect decisions lead to either loss of business (rejecting safe applicants) or financial loss (approving risky applicants).

OBJECTIVE

- Identify key factors influencing loan defaults using Exploratory Data Analysis (EDA).
- Minimize credit loss by understanding patterns associated with defaulting borrowers.

BUSINESS UNDERSTANDING

COMPANY OVERVIEW

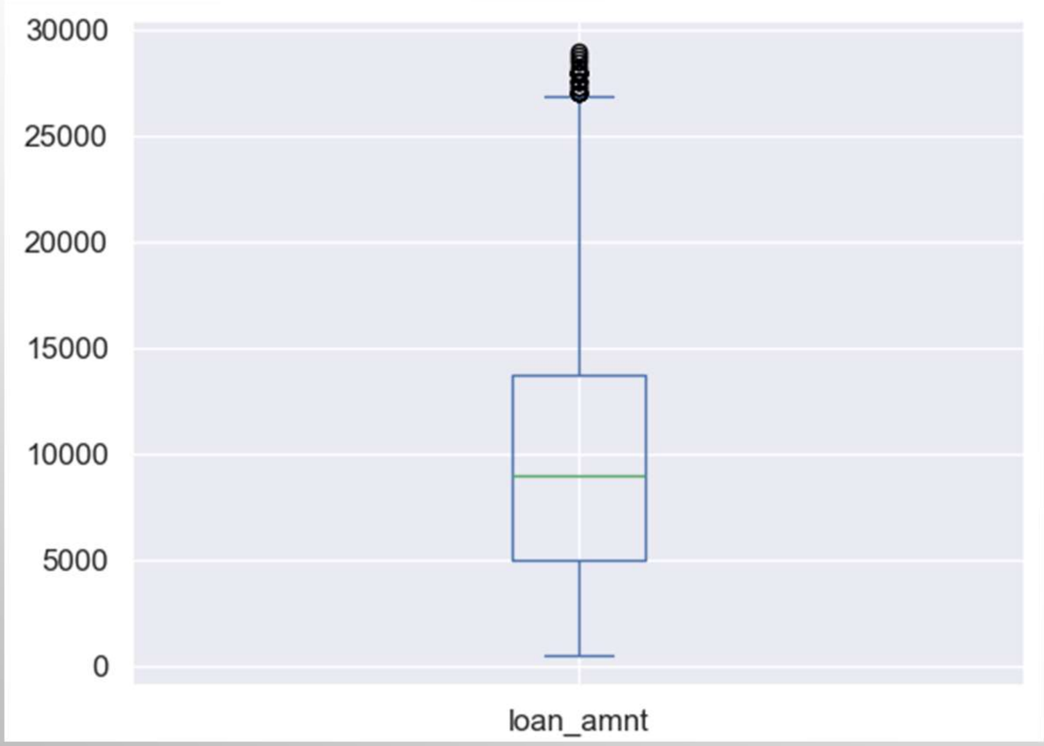
- A consumer finance company specializing in lending various types of loans to urban customers.
- When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

LOAN APPROVAL DECISION

- Repayment Likely: Not approving the loan results in business loss.
- Default Likely: Approving the loan may lead to financial loss due to customer defaulting.

UNIVARIATE ANALYSIS

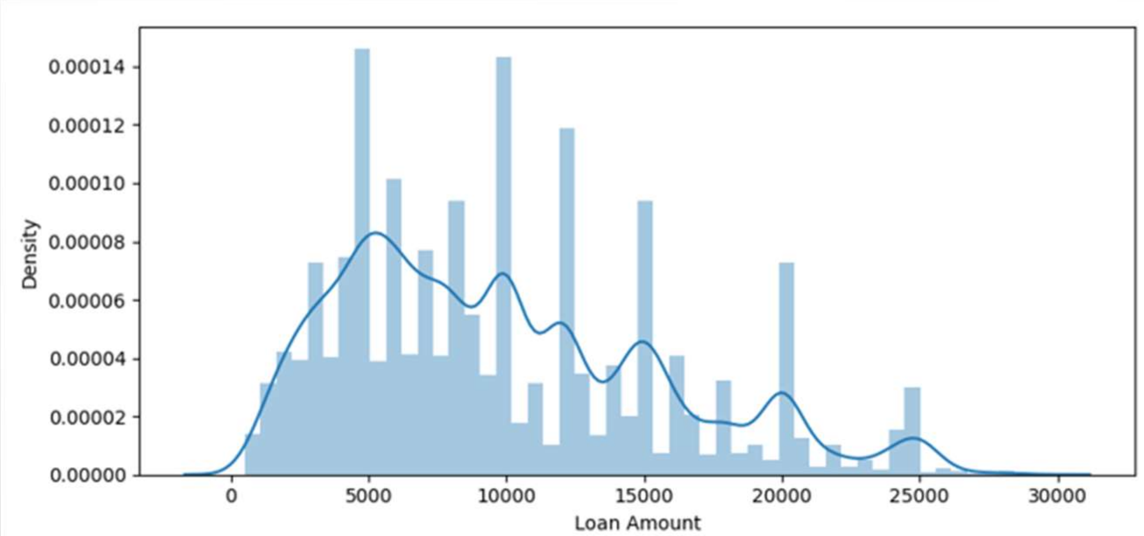
LOAN_AMNT



OBSERVATIONS

- MOST LOANS FALL BETWEEN $5,000 AND $15,000
- WITH THE AVERAGE LOAN IS AROUND $11,000.
- THERE ARE A FEW EXCEPTIONS, WHERE SOME LOANS ARE MUCH LARGER THAN USUAL.

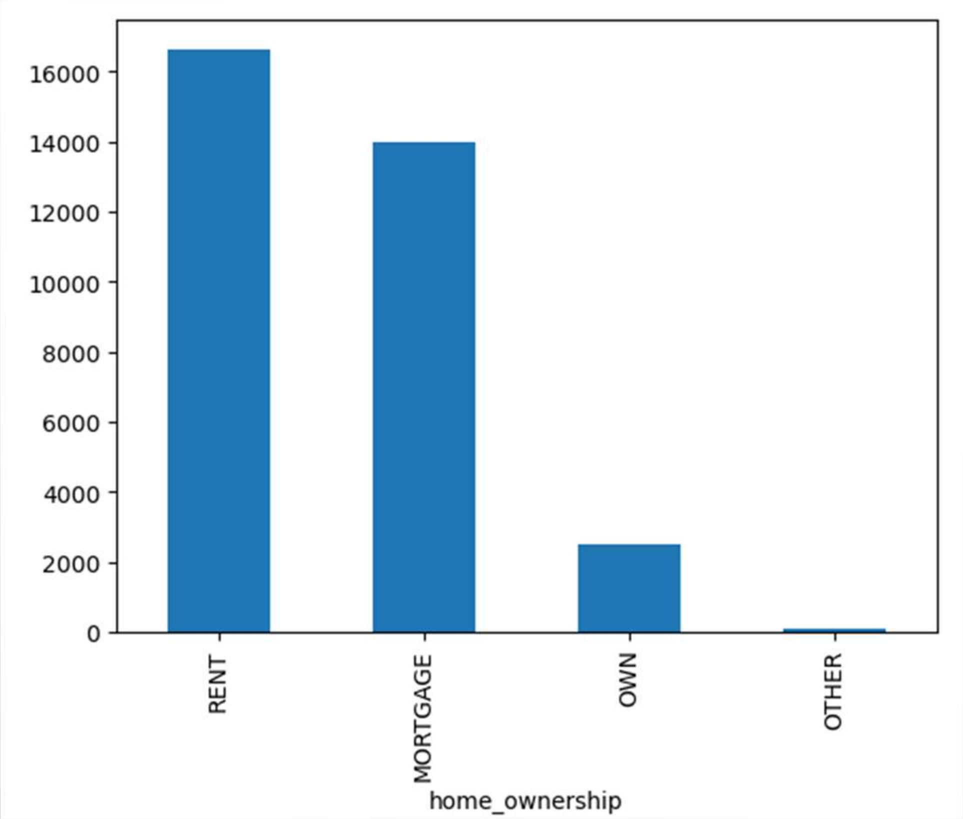
LOAN_AMNT



OBSERVATIONS

- MOST LOANS FALL BETWEEN $5,000 AND $15,000
- WITH THE AVERAGE LOAN IS AROUND $11,000.
- THERE ARE A FEW EXCEPTIONS, WHERE SOME LOANS ARE MUCH LARGER THAN USUAL.

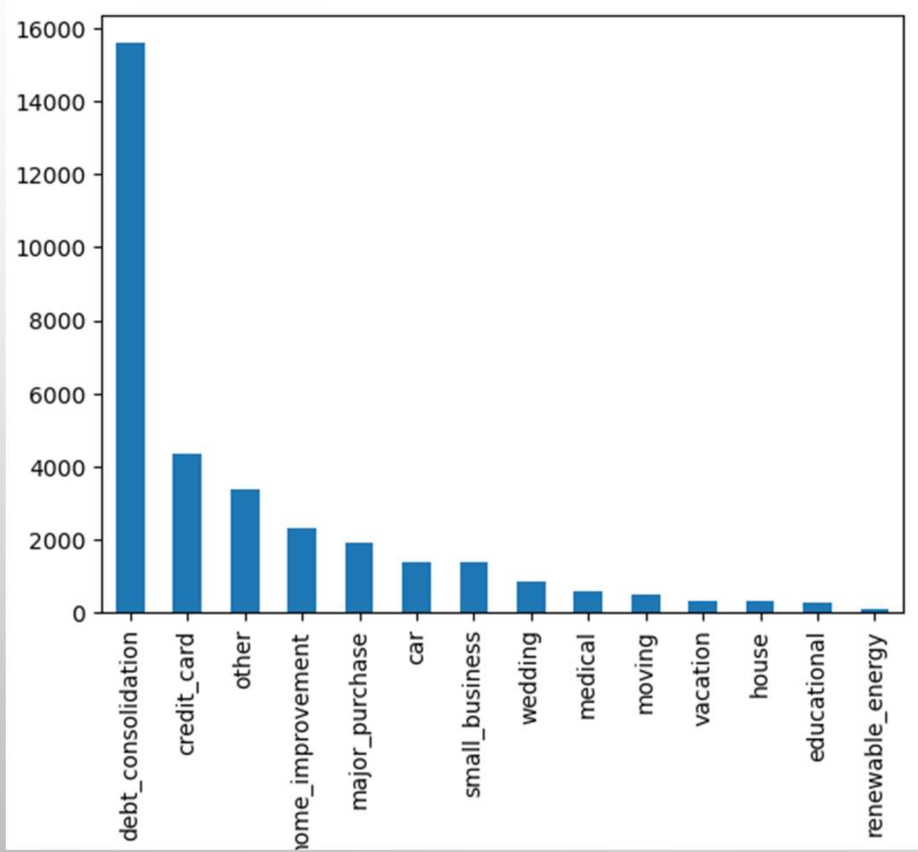
HOME_OWNERSHIP



OBSERVATIONS

- MOST LOAN APPLICANTS ARE EITHER RENTING THEIR HOMES OR HAVE A MORTGAGE

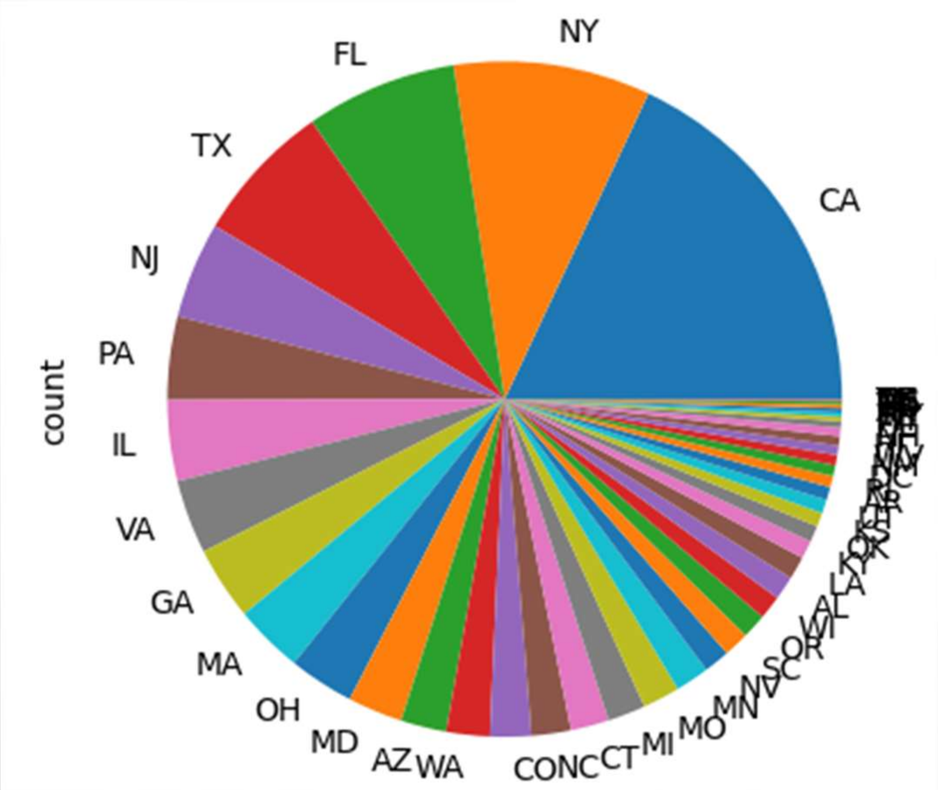
PURPOSE



OBSERVATIONS

- MOST LOAN APPLICANTS ARE SEEKING LOANS FOR DEBT CONSOLIDATION

ADDR_STATE

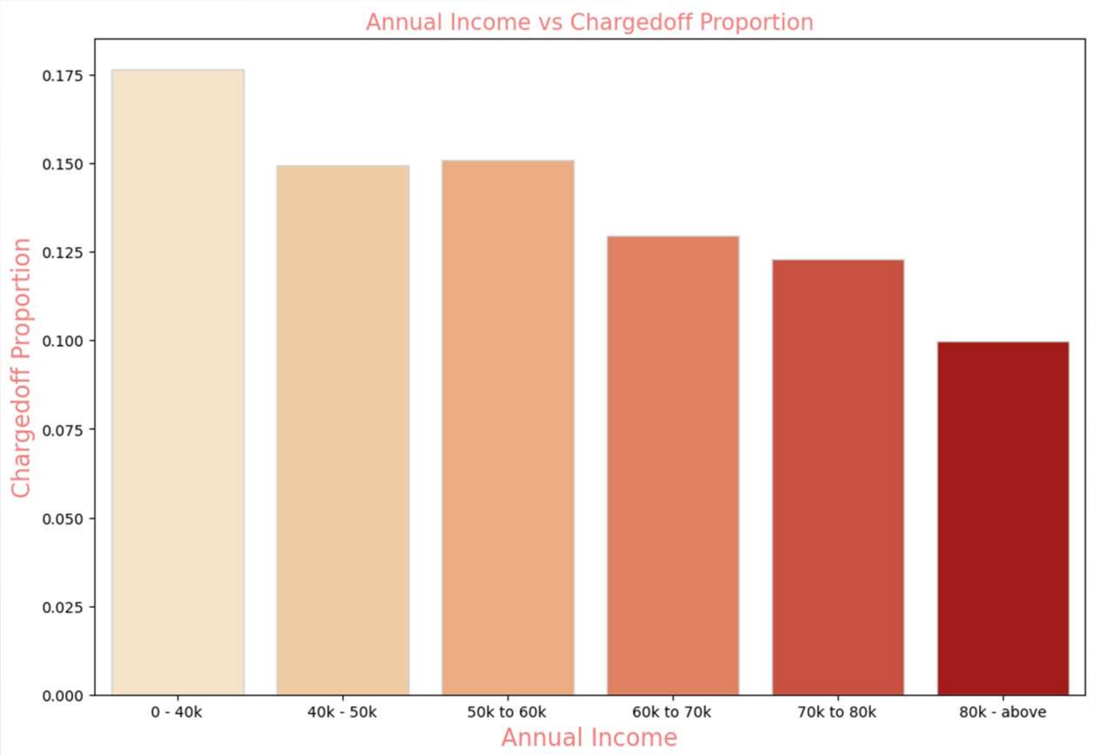


OBSERVATIONS

- THE MAJORITY OF LOAN APPLICANTS ARE FROM CALIFORNIA (CA) STATE.

BIVARIATE ANALYSIS

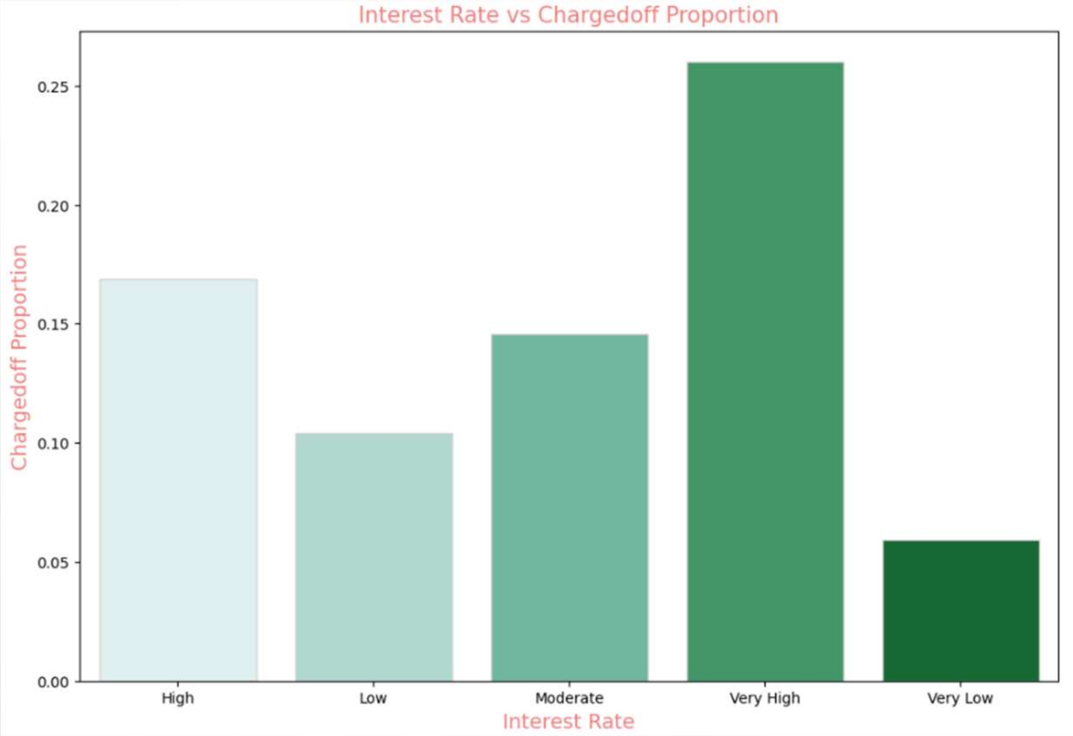
ANNUAL INCOME VS CHARGEDOFF



OBSERVATIONS

- APPLICANTS WITH LOWER ANNUAL INCOMES ARE MORE LIKELY TO DEFAULT ON LOANS THAN THOSE WITH HIGHER INCOMES.
- THIS INDICATES THAT INCOME LEVEL PLAYS A CRUCIAL ROLE IN EVALUATING CREDIT RISK

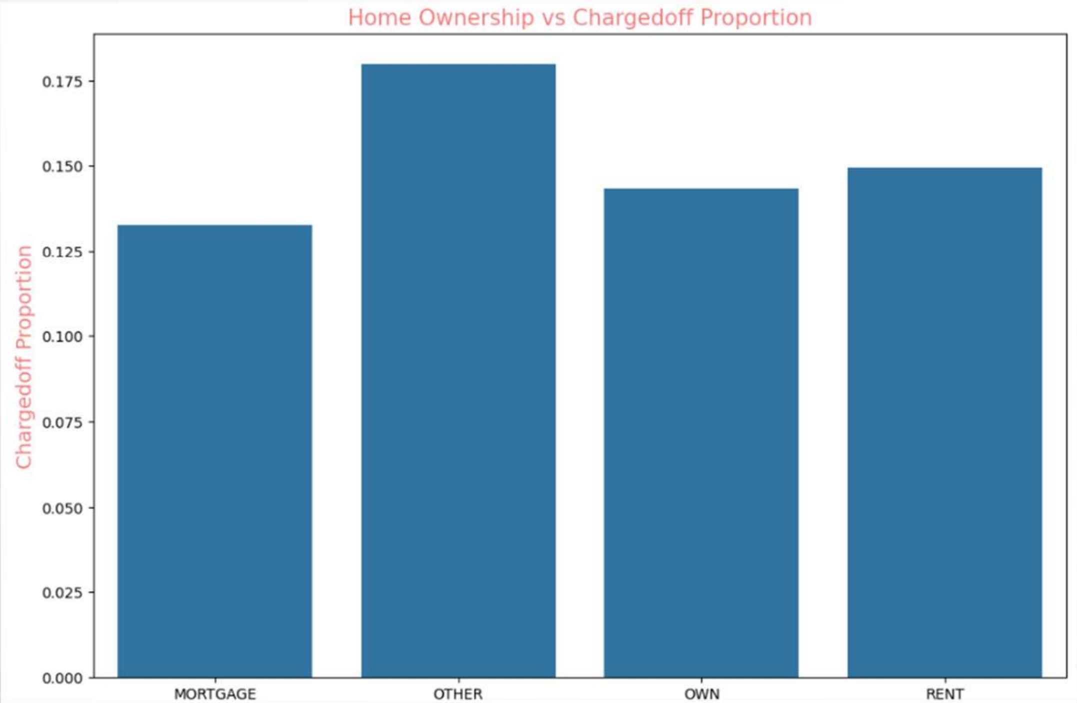
INTEREST RATE VS CHARGEDOFF



OBSERVATIONS

- LOANS WITH HIGHER INTEREST RATES HAVE A NOTICEABLY HIGHER RATE OF CHARGE-OFFS, SUGGESTING THAT INTEREST RATES ARE A STRONG INDICATOR OF DEFAULT RISK.
- HIGHER RATES ARE LINKED TO A GREATER LIKELIHOOD OF LOAN DEFAULTS.

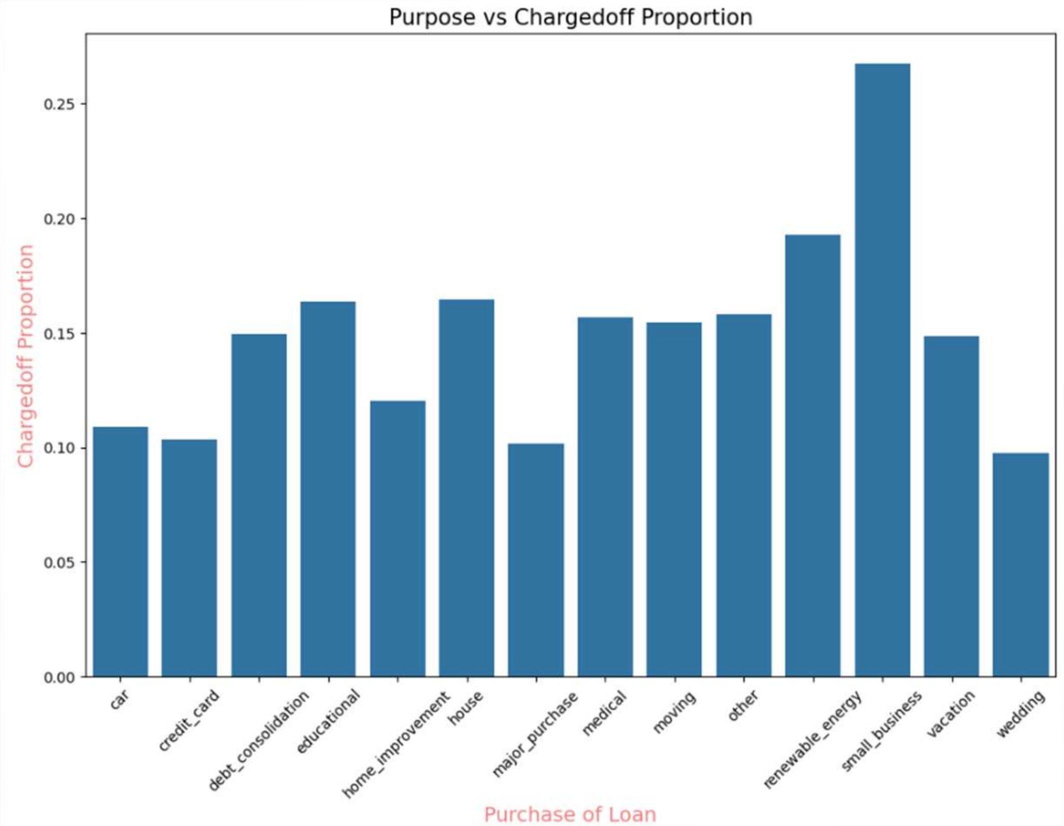
HOME OWNERSHIP VS CHARGEDOFF



OBSERVATIONS

- APPLICANTS WHO DO NOT OWN A HOME HAVE A HIGHER LIKELIHOOD OF DEFAULTING ON THEIR LOANS.

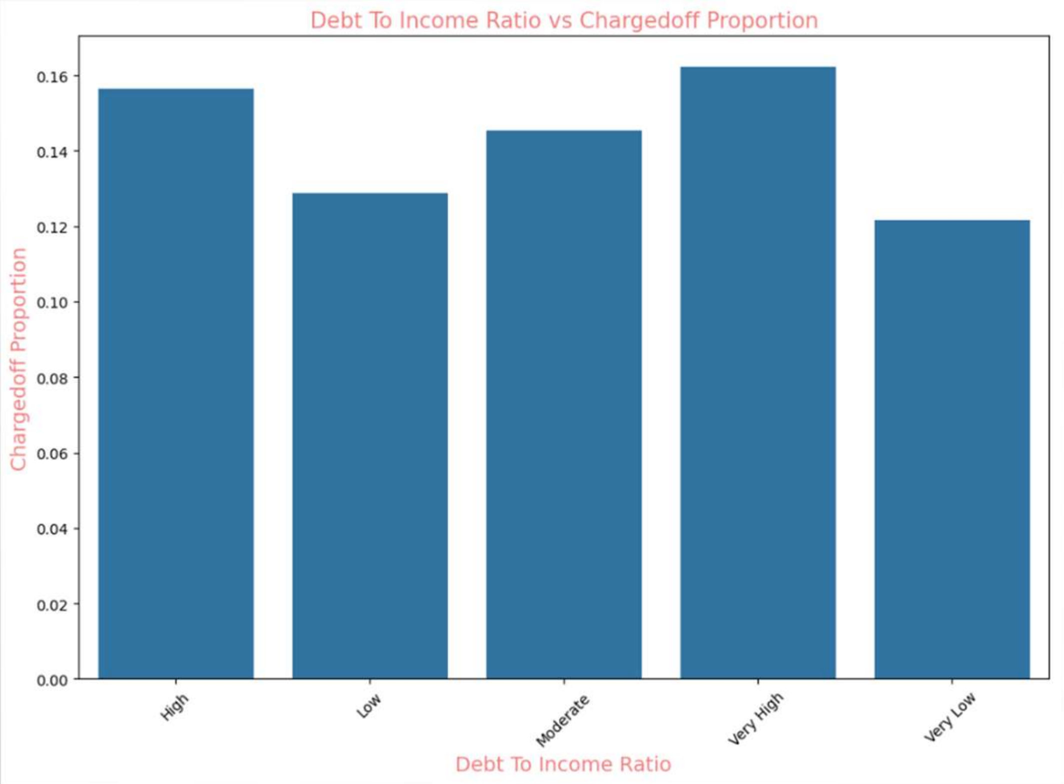
PURPOSE VS CHARGEDOFF



OBSERVATIONS

- APPLICANTS WITH A HOME LOAN ARE LESS LIKELY TO DEFAULT ON THEIR LOANS, WHILE THOSE WITH LOANS FOR SMALL BUSINESSES HAVE A HIGHER RISK OF DEFAULTING.

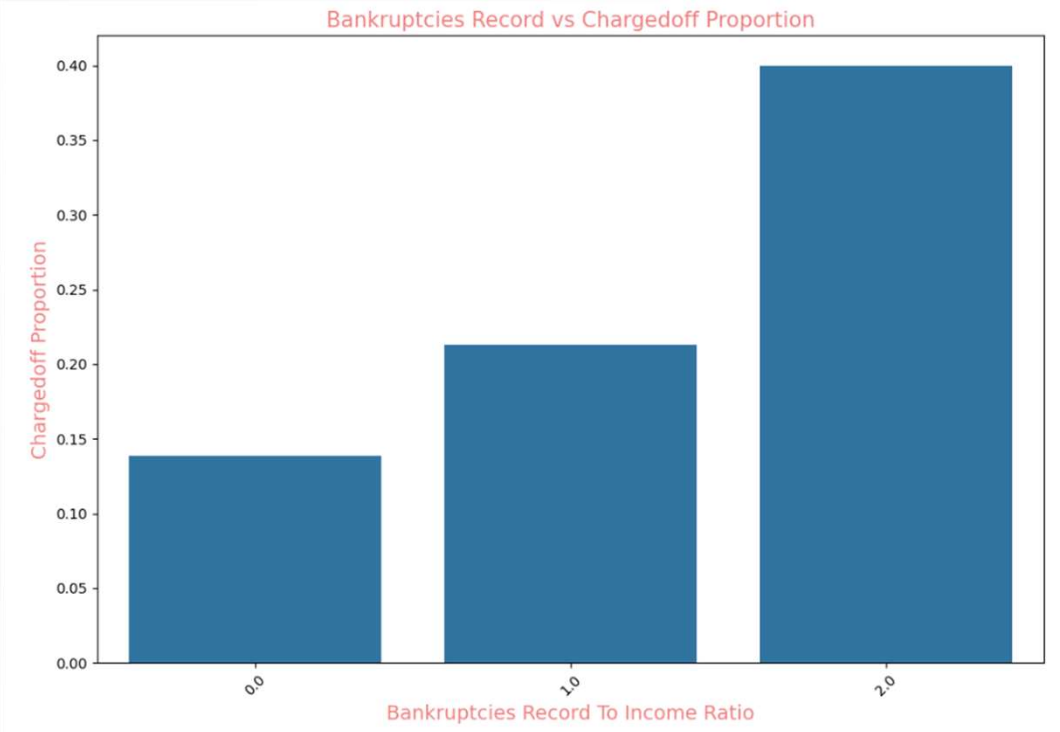
DTI RATIO VS CHARGEDOFF



OBSERVATIONS

- AS THE DEBT-TO-INCOME (DTI) RATIO RISES, THE PROPORTION OF CHARGED-OFF LOANS ALSO INCREASES.
- THIS IMPLIES THAT BORROWERS WITH HIGHER DTI RATIOS ARE MORE PRONE TO DEFAULT.

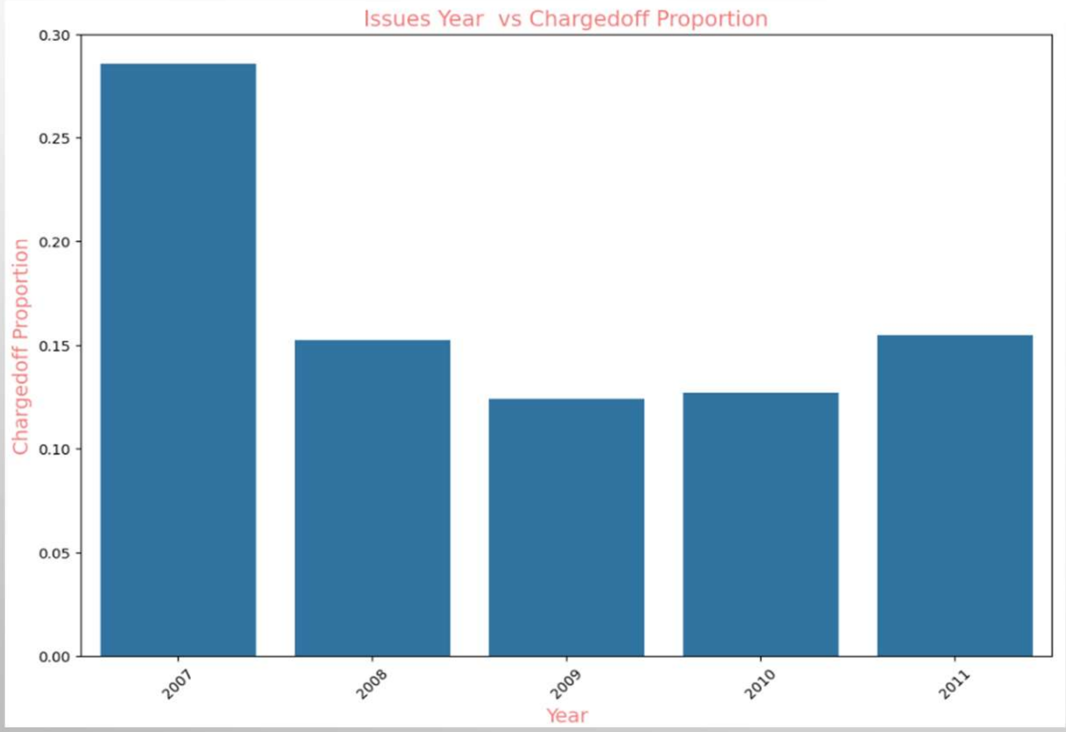
BANKRUPTCIES RECORD VS CHARGEDOFF



OBSERVATIONS

- AS THE DEBT-TO-INCOME (DTI) RATIO RISES, THE PROPORTION OF CHARGED-OFF LOANS ALSO INCREASES.
- THIS IMPLIES THAT BORROWERS WITH HIGHER DTI RATIOS ARE MORE PRONE TO DEFAULT.

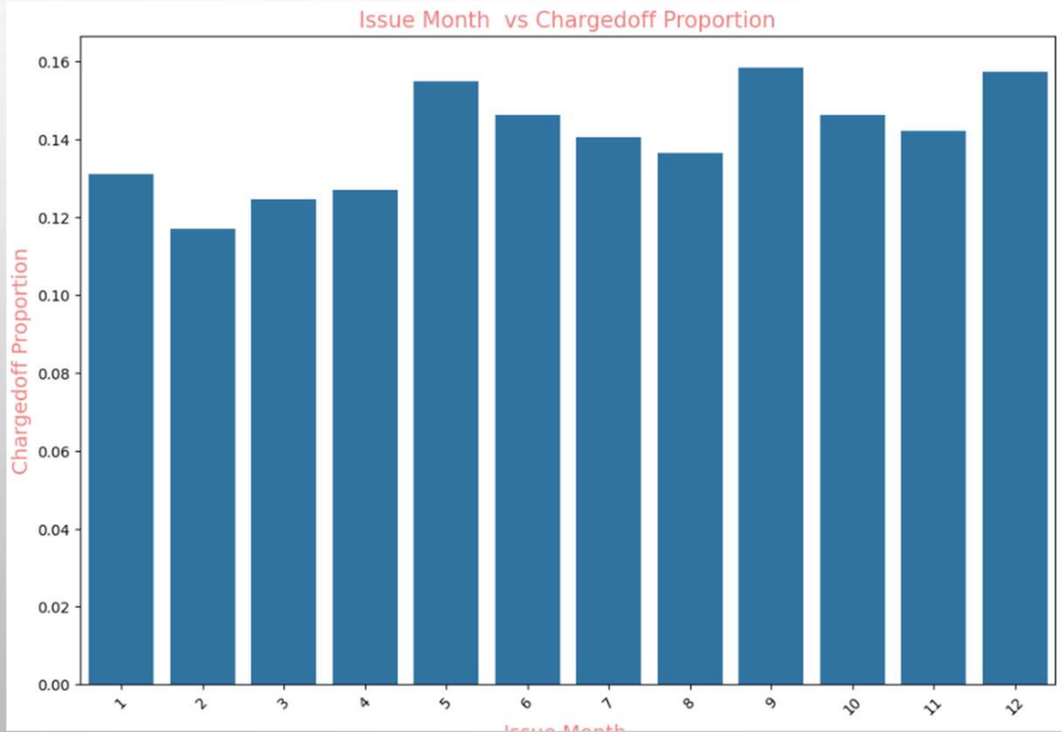
ISSUES YEAR VS CHARGEDOFF



OBSERVATIONS

- LOANS ISSUED IN 2007 AND 2008 HAVE MORE CHARGE-OFFS COMPARED TO THOSE FROM RECENT YEARS.
- THIS COULD BE DUE TO THE ECONOMIC SITUATION AT THE TIME OR HOW LENDING PRACTICES HAVE EVOLVED SINCE THEN.

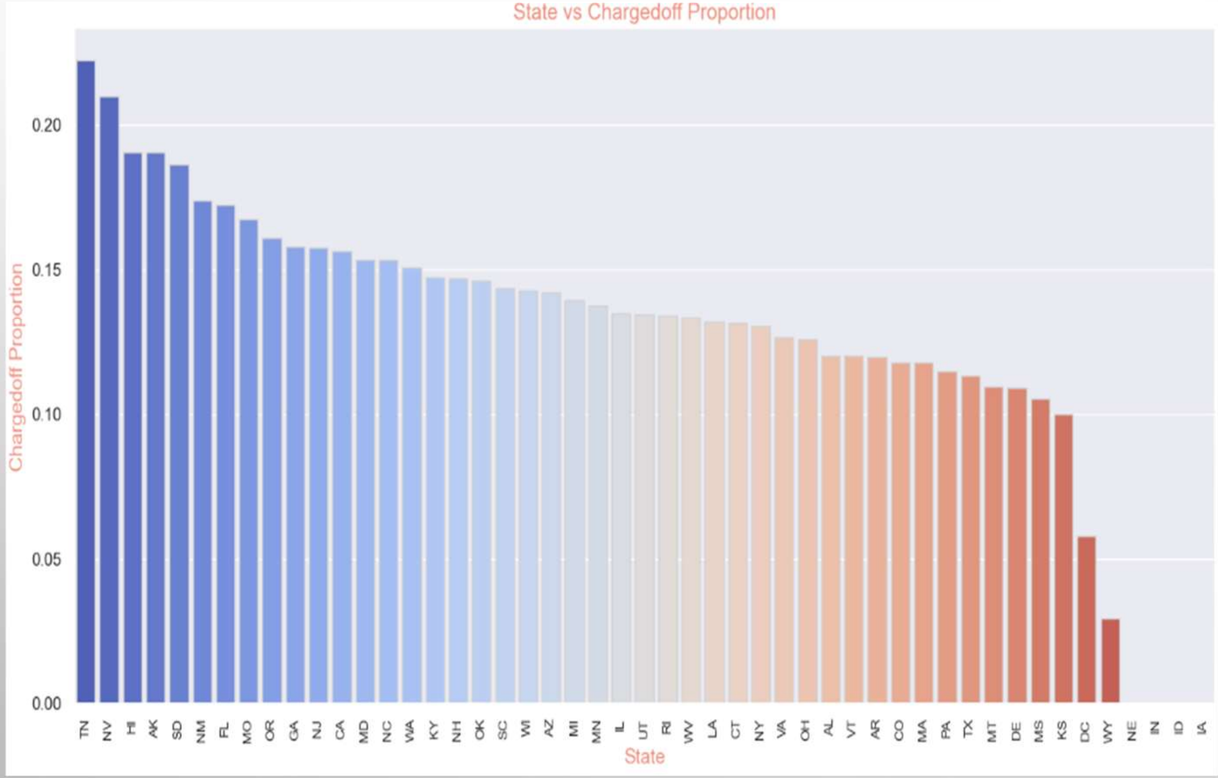
ISSUE MONTH VS CHARGEDOFF



OBSERVATIONS

- THERE DOESN'T SEEM TO BE A NOTICEABLE PATTERN OR BIG DIFFERENCES IN CHARGE-OFF RATES DEPENDING ON THE MONTH A LOAN WAS ISSUED.
- THIS IMPLIES THAT THE MONTH OF ISSUANCE MAY NOT BE A KEY FACTOR IN PREDICTING DEFAULT RISK.

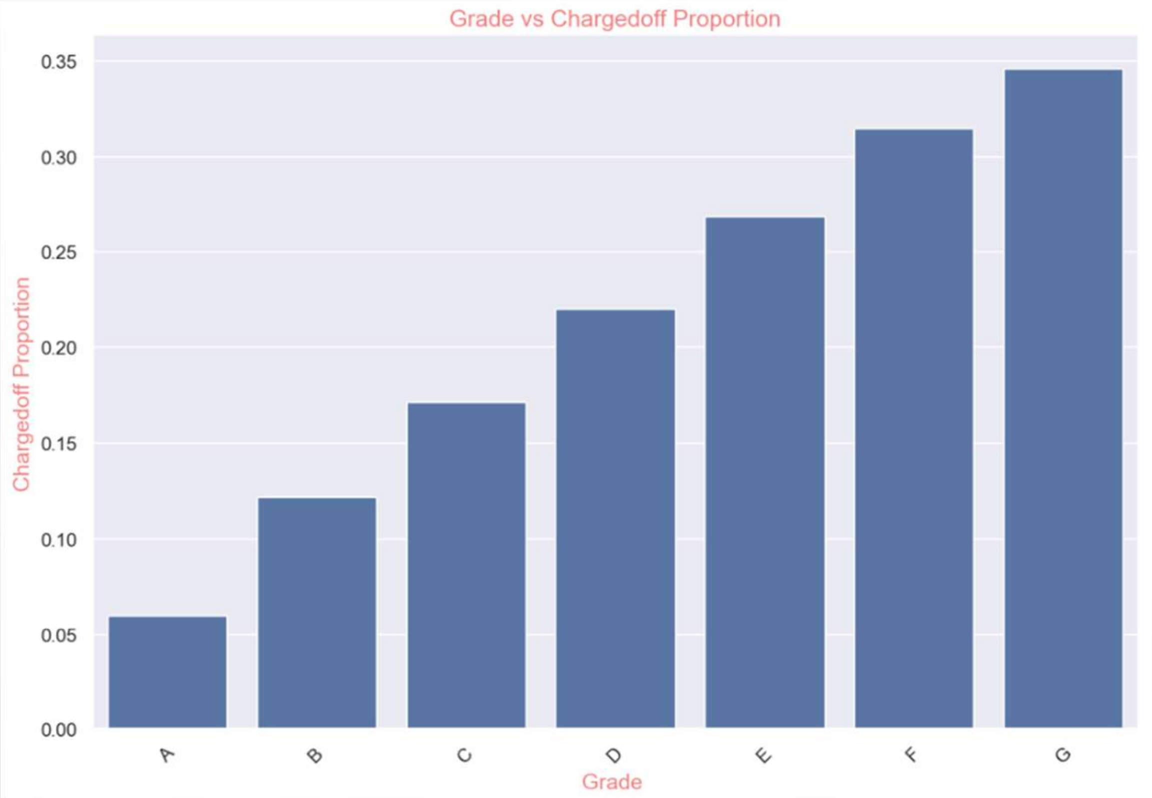
STATE VS CHARGEDOFF



OBSERVATIONS

- STATES LIKE NV, AK, FL ARE HAVING HIGH CHANCES OF LOAN DEFAULTS.
- STATES LIKE IA, ME, ID ARE HAVING LOW CHANCES OF LOAN DEFAULTS.

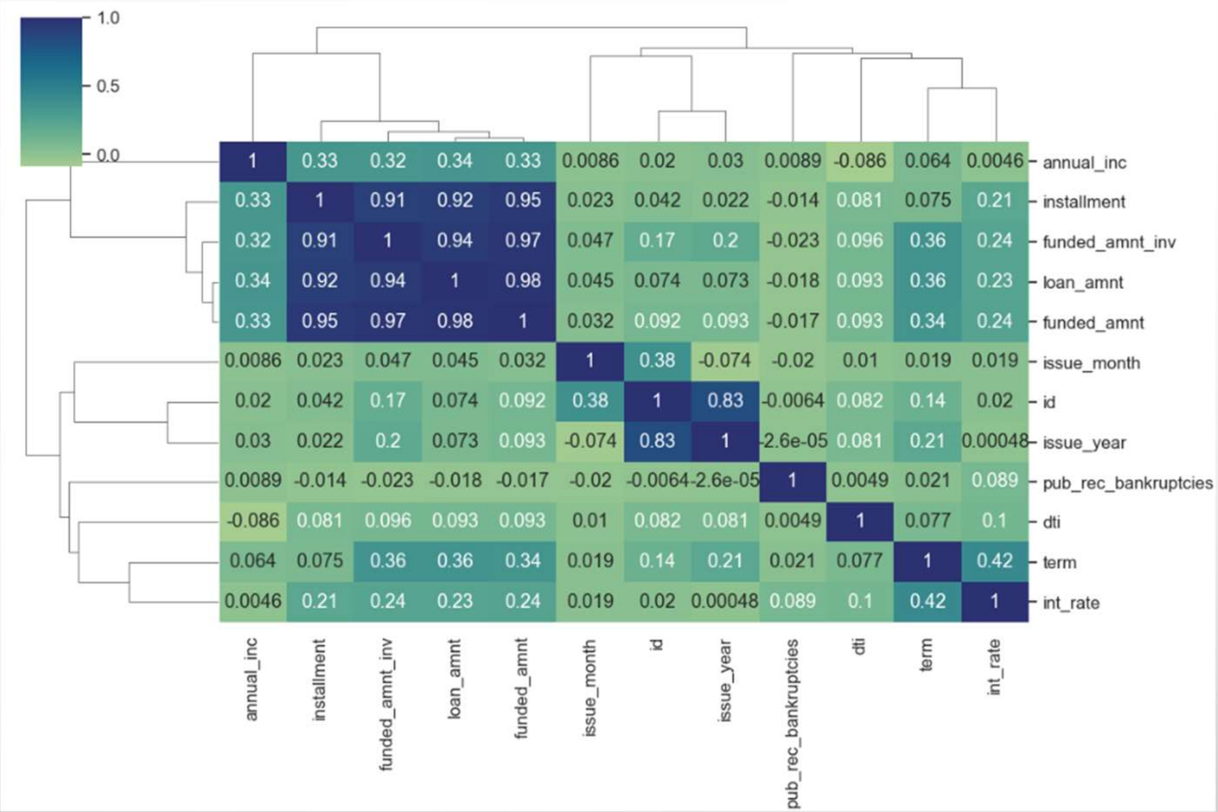
GRADE VS CHARGEDOFF



OBSERVATIONS

- 1. LOAN APPLICANTS WITH A GRADE G LOAN HAVE THE HIGHEST DEFAULT RATES.
- 2. LOAN APPLICANTS WITH A GRADE A LOAN HAVE THE LOWEST DEFAULT RATES.

CORRELATION ANALYSIS



OBSERVATIONS

POSITIVE CORRELATIONS:

- LOAN AMOUNT, FUNDED AMOUNT, AND INVESTOR-FUNDED AMOUNT ARE CLOSELY LINKED, MEANING THAT LARGER LOAN REQUESTS GENERALLY GET MORE FUNDING.
- INTEREST RATES HAVE A MODERATE POSITIVE CORRELATION WITH LOAN AMOUNTS, SUGGESTING THAT BIGGER LOANS OFTEN COME WITH HIGHER INTEREST RATES.
- THE INSTALLMENT AMOUNT IS STRONGLY RELATED TO BOTH THE LOAN AMOUNT AND THE FUNDED AMOUNT, SHOWING THAT LARGER LOANS LEAD TO HIGHER MONTHLY PAYMENTS.

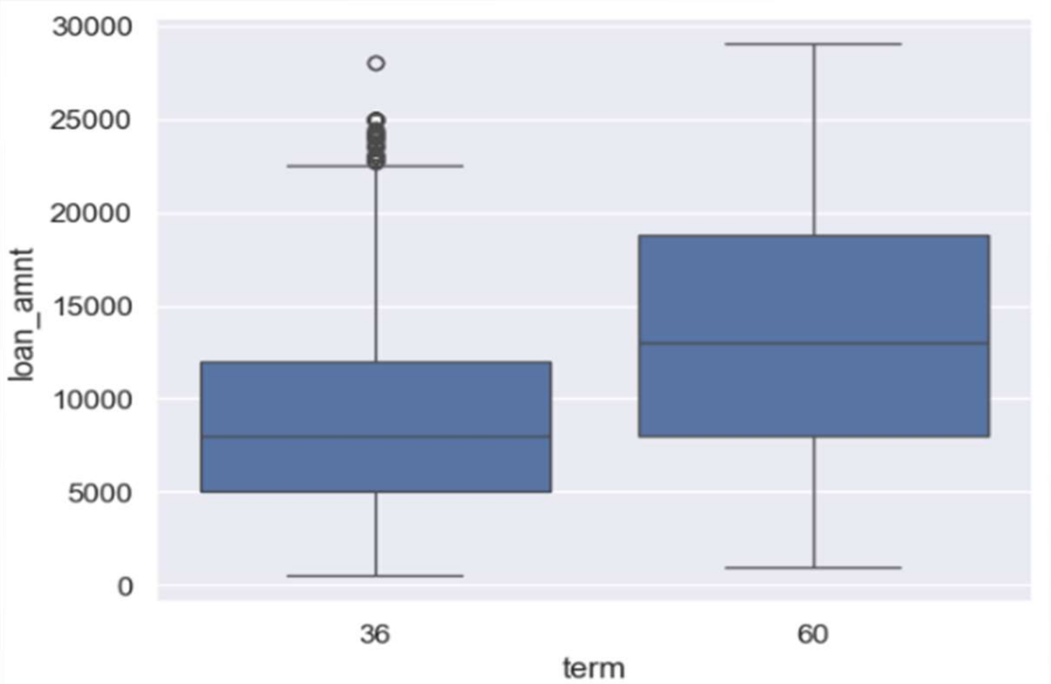
NEGATIVE CORRELATIONS:

- THERE'S A WEAK NEGATIVE CORRELATION BETWEEN PUBLIC RECORD BANKRUPTCIES AND ANNUAL INCOME, INDICATING THAT PEOPLE WITH HIGHER INCOMES ARE LESS LIKELY TO HAVE BANKRUPTCY RECORDS.

OVERALL:

- THE HEATMAP HIGHLIGHTS HOW DIFFERENT NUMERICAL VARIABLES IN THE LOAN DATASET ARE RELATED.
- THESE CORRELATIONS CAN HELP US UNDERSTAND FACTORS THAT AFFECT LOAN APPROVAL, INTEREST RATES, AND THE CHANCES OF DEFAULT.

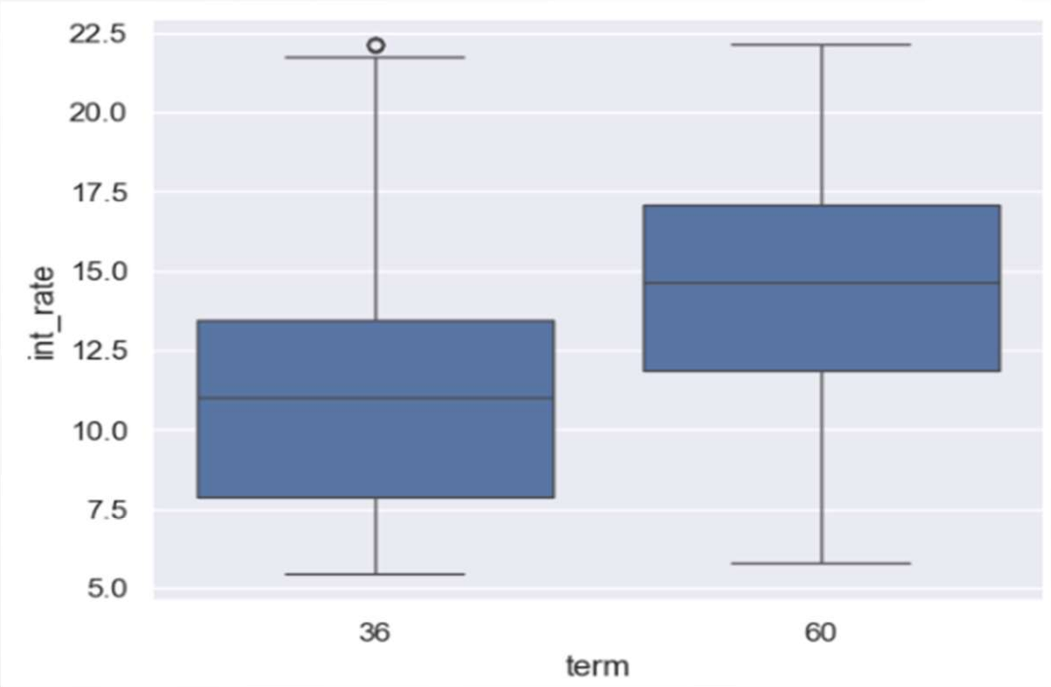
LOAN_AMNT VS TERM



OBSERVATIONS

- IF THE TERM OF THE LOAN IS INCREASING, THE LOAN AMOUNT IS ALSO INCREASING.
- THIS SUGGESTS A TREND WHERE LONGER LOAN TERMS ARE ASSOCIATED WITH HIGHER LOAN AMOUNTS.

INT_RATE VS TERM



OBSERVATIONS

- LOANS WITH LONGER TERMS (60 MONTHS) TYPICALLY COME WITH HIGHER INTEREST RATES COMPARED TO SHORTER TERMS (36 MONTHS).
- THIS TREND REFLECTS THE GREATER RISK ASSOCIATED WITH LONGER-TERM LENDING, AS LENDERS FACE THE POSSIBILITY OF DEFAULTS OVER A MORE EXTENDED PERIOD.

SUMMARY

- LOAN APPLICANTS WITH LOWER INCOMES AND HIGHER INTEREST RATES ARE MORE LIKELY TO DEFAULT.
- THOSE WHO DON'T OWN HOMES AND HAVE A HISTORY OF BANKRUPTCIES ARE ALSO AT HIGHER RISK.
- THE PURPOSE OF THE LOAN AFFECTS DEFAULT RATES, WITH SMALL BUSINESS LOANS BEING RISKIER.
- ADDITIONALLY, LOANS ISSUED IN EARLIER YEARS (2007-2008) HAVE HIGHER CHARGE-OFF RATES. STATES LIKE NEVADA, ALASKA, AND FLORIDA SHOW HIGHER DEFAULT RATES COMPARED TO IOWA, MAINE, AND IDAHO.