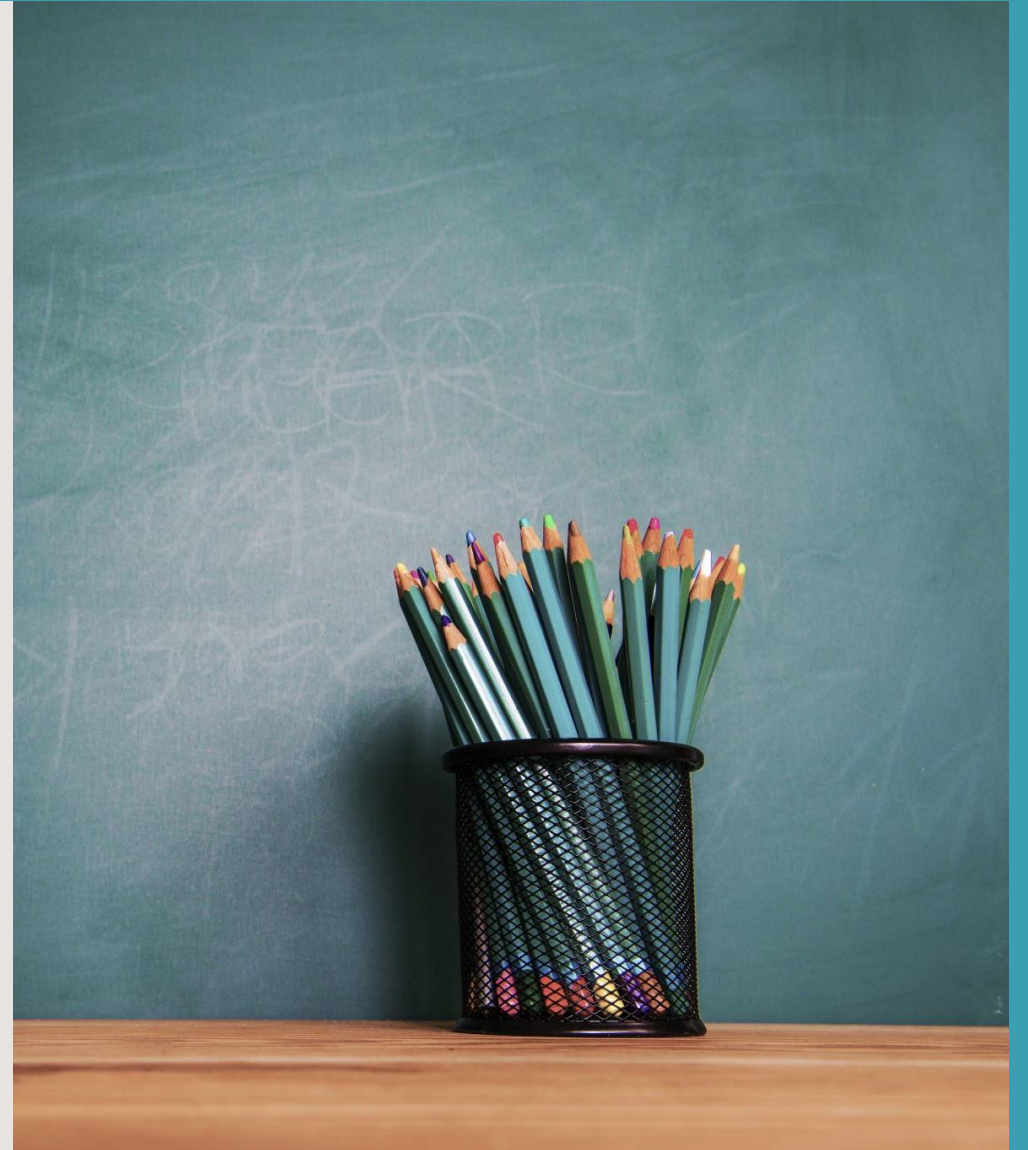


LENDING CLUB CASE STUDY

TEAM MEMBERS:

UMA MAHESWARI KUDDIRILLA

THOMAS PURACKAL



Contents



Problem Statement

Identify factors influencing loan defaults for better risk management



Analysis Approach

Used univariate, bivariate, and multivariate methods for comprehensive insights.



Data Understanding

Explored loan attributes crucial for default prediction.



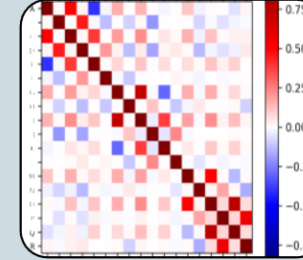
Univariate Analysis

Identified high-risk factors like 13%-17% interest rates and Grade B, Sub-Grade B5 loans.



Bivariate Analysis

Explored correlations between interest rates, employment length, and loan grades.



Multivariate Analysis

Analyzed interactions among key variables impacting default rates.



Conclusions

Summarized findings and recommendations for improving loan risk assessment.

Problem statement

Problem Statement:

Urban Consumer Finance Company faces dual risks: approving loans to potential defaulters and rejecting viable applicants, posing financial loss and missed business opportunities. We aim to utilize data analysis to mitigate defaults and optimize loan decisions.

- **Target:**

Conduct extensive data exploration on loan data to uncover trends, patterns, and risk factors influencing defaults. Provide actionable insights to enhance decision-making and minimize financial risks.

- **Risks:**

Financial Loss due to Defaults: Understand default patterns to reduce credit losses.

Missed Business Opportunities: Optimize loan decisions to maximize profitability and minimize missed lending opportunities.

Data Understanding

Import
data

Remove
headers,
footers,
summary
rows

Remove
Null
values

Remove
duplicates

Remove
columns
with
unique
values

Remove
rows
with
loan
status as
Current

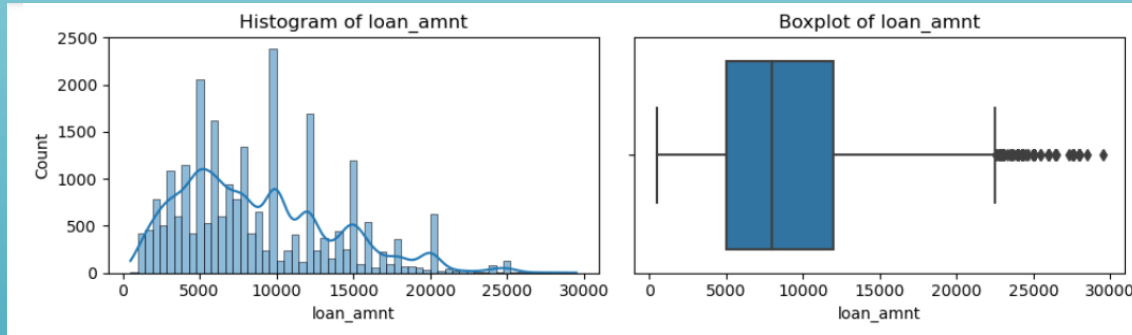
Delete
unnecessary
columns

Handling
missing
values

Data
type
conversions

Derived
columns

Univariate Analysis - Numerical

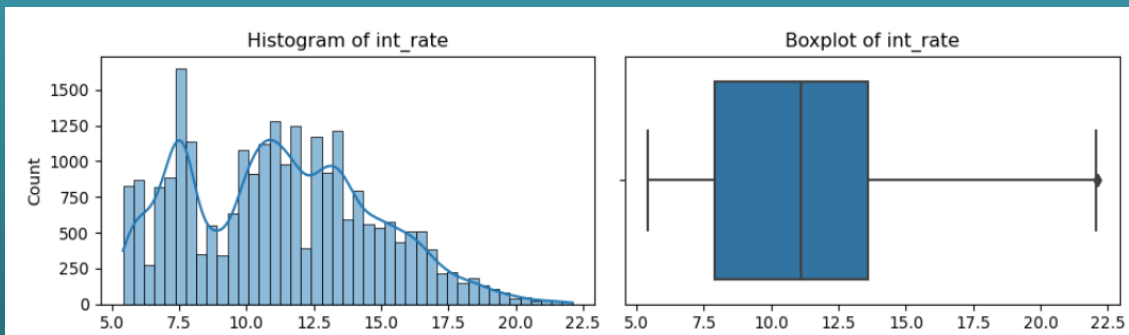
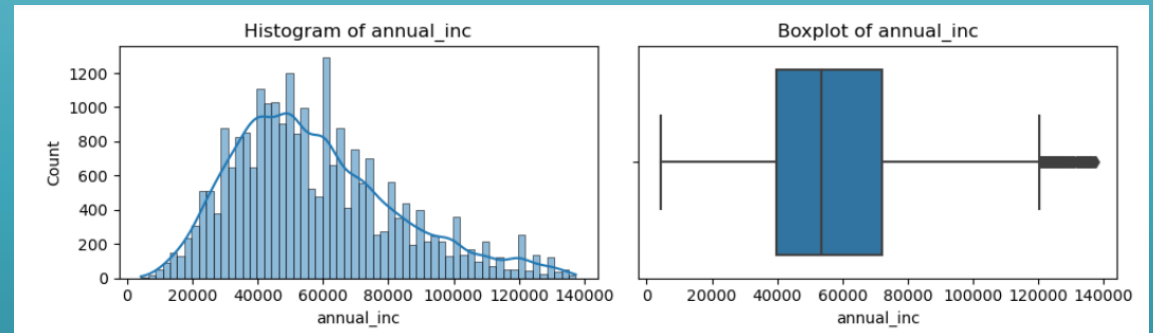


Loan Amount

- ❑ The distribution is positively skewed, indicating a longer tail on the right side.
- ❑ Most loan amounts are between \$5,000 and \$15,000.
- ❑ Outliers are predominantly on the higher end, reflecting a few significantly larger loan amounts.

Annual Income

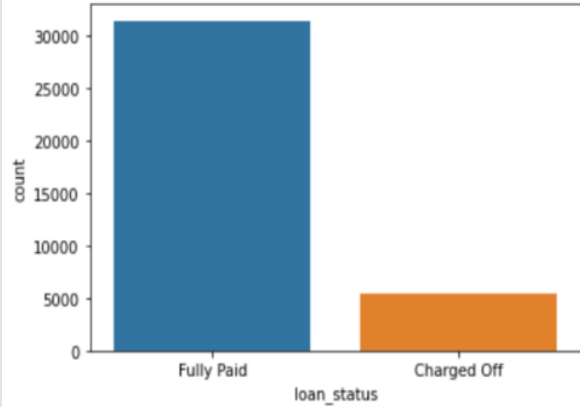
- ❑ The distribution is positively skewed, suggesting most incomes are on the lower end with a few higher values.
- ❑ Most annual incomes range between \$40,000 and \$78,000.
- ❑ Some annual income values extend up to \$145,000, indicating significant outliers.



Interest Rate

- ❑ The distribution appears almost normal, suggesting a bell-shaped curve.
- ❑ Many interest rates fall between 8% and 14%.
- ❑ This indicates that most loans are offered at moderate interest rates, with fewer loans at extremely high or low rates.

Univariate Analysis - Categorical



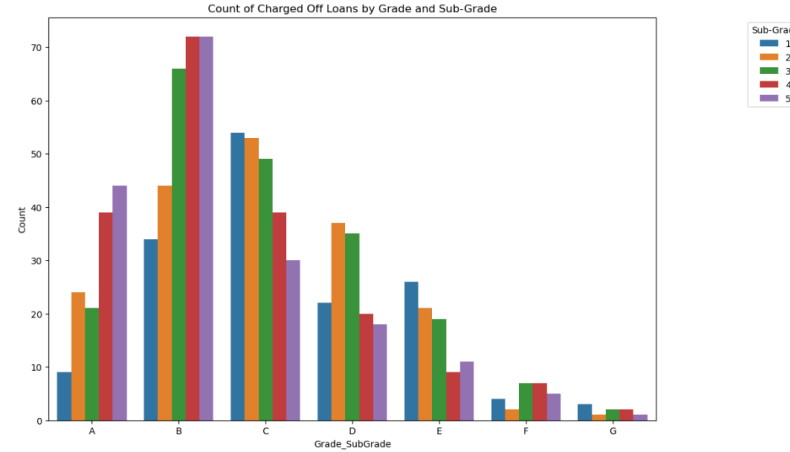
❑ Charged Off Borrowers:

15% of borrowers have their loans charged off.

❑ Fully Paid Borrowers:

85% of borrowers repay their loans in full.

Loan Status



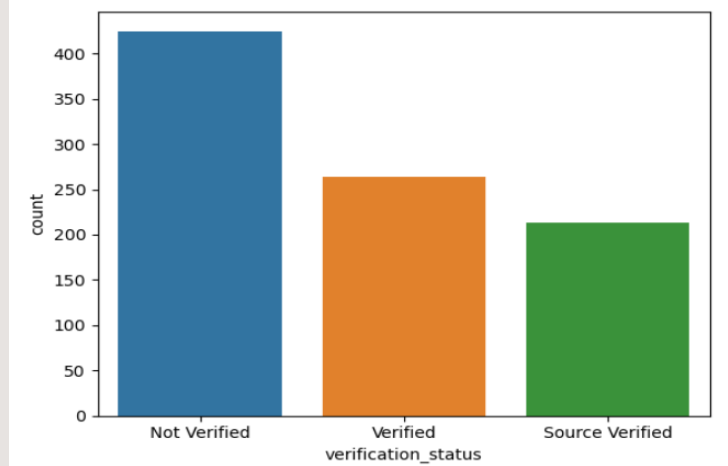
❑ Higher Default Probability with Grade 'B':

Loans under Grade 'B' show higher default rates.

❑ Increased Risk at Sub-Grade 'B5':

Sub-grade 'B5' within Grade 'B' has an even higher default probability

Grades and Subgrades

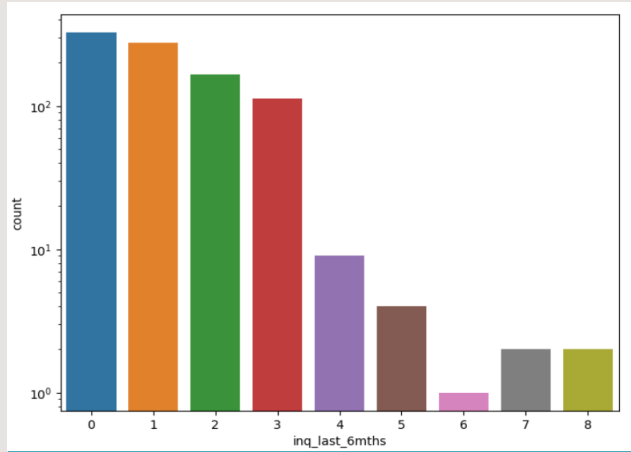


❑ Higher Default Probability with "Not Verified" Status

❑ This underscores the importance of thorough verification processes to ensure the reliability of borrower information and reduce default rates.

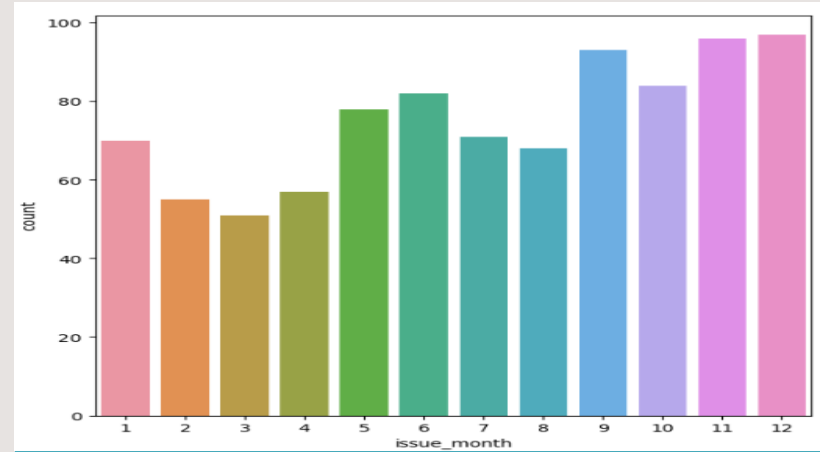
Verification status

Univariate Analysis - Categorical



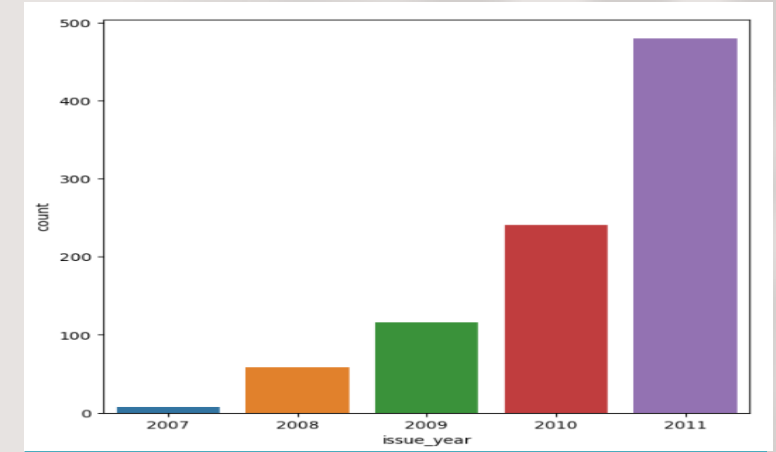
- ❑ **Higher Default Probability with Zero Inquiries:**
- ❑ Loans where the number of inquiries in the past 6 months is zero show a higher probability of defaulting.

Inquires in last 6 months



- ❑ **Issue Month Observations:**
- ❑ Maximum defaults occur in December (Issue Month 12).

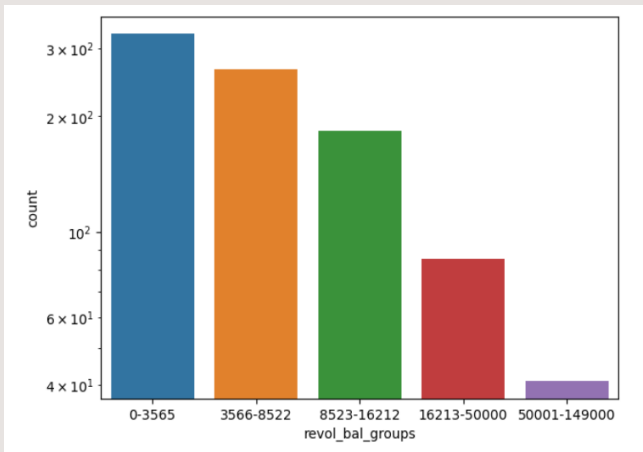
Issue Month



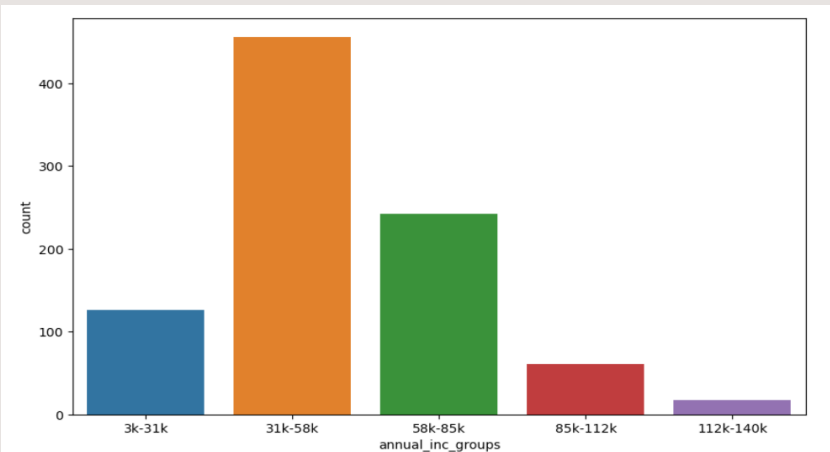
- ❑ **Issue Year Observations**
- ❑ :Loans issued in the year 2011 have the highest default rates

Issue year

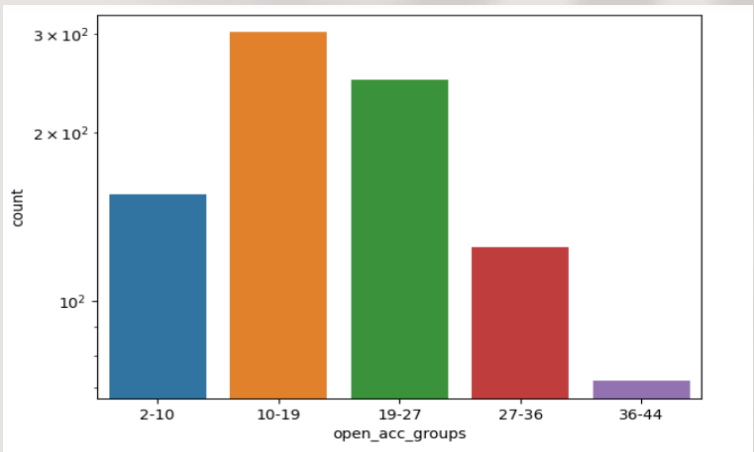
Categorizing Columns Using Binning for Enhanced Analysis



- ❑ **Revolving Balance Group:**
- ❑ Higher Default Risk:
- ❑ Applicants with revolving balances in the range of 0-3565 show a higher likelihood of defaulting on loans.



- ❑ **Annual Income Group:**
- ❑ Higher Default Risk:
- ❑ Applicants with annual incomes in the range of 31k-58k are more prone to defaulting on loans.



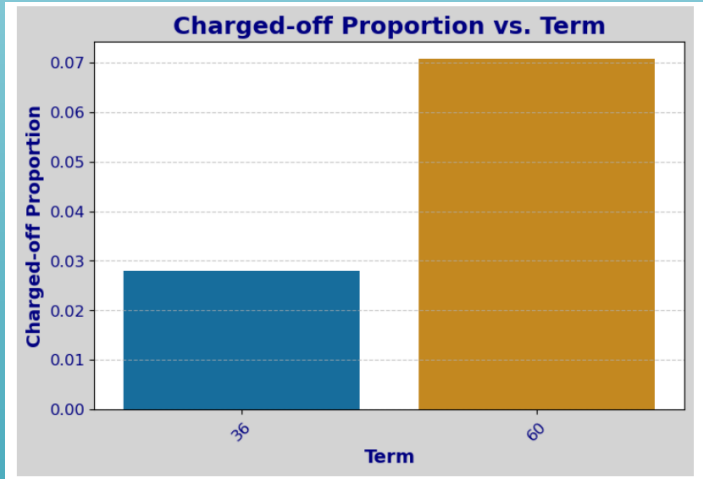
- ❑ **Open Account Group:**
- ❑ Higher Default Risk:
- ❑ Applicants with 2-10 open accounts are observed to have a higher probability of defaulting.

Revol_bal_groups

Annual_inc_groups

Open account groups

Bivariate Analysis

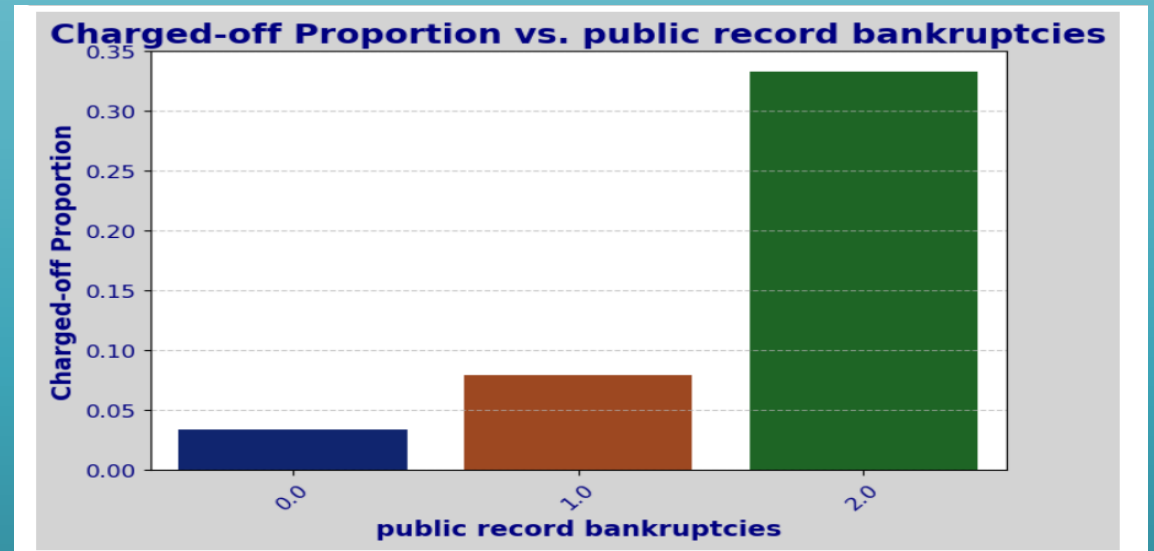


Loan Term:

- 36-month term: Higher count of charged-off loans.
- 60-month term: Higher proportion of charged-off loans.

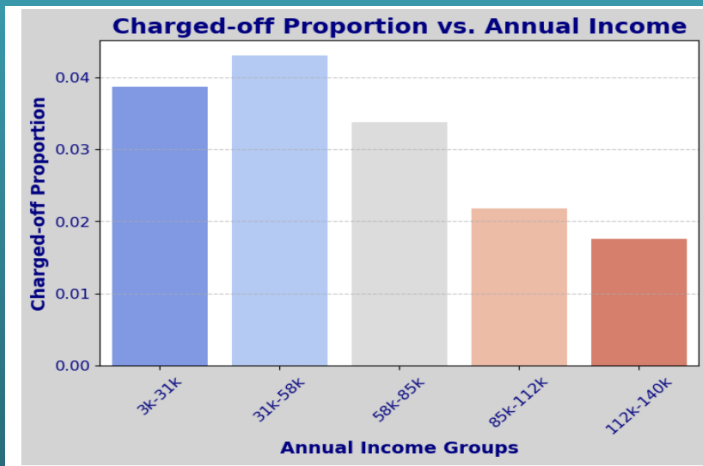
Public Record Bankruptcies:

- Two bankruptcies: Highest proportion of charged-off loans.
- One bankruptcy: Higher likelihood of default.
- No bankruptcies: Lowest proportion of charged-off loans.

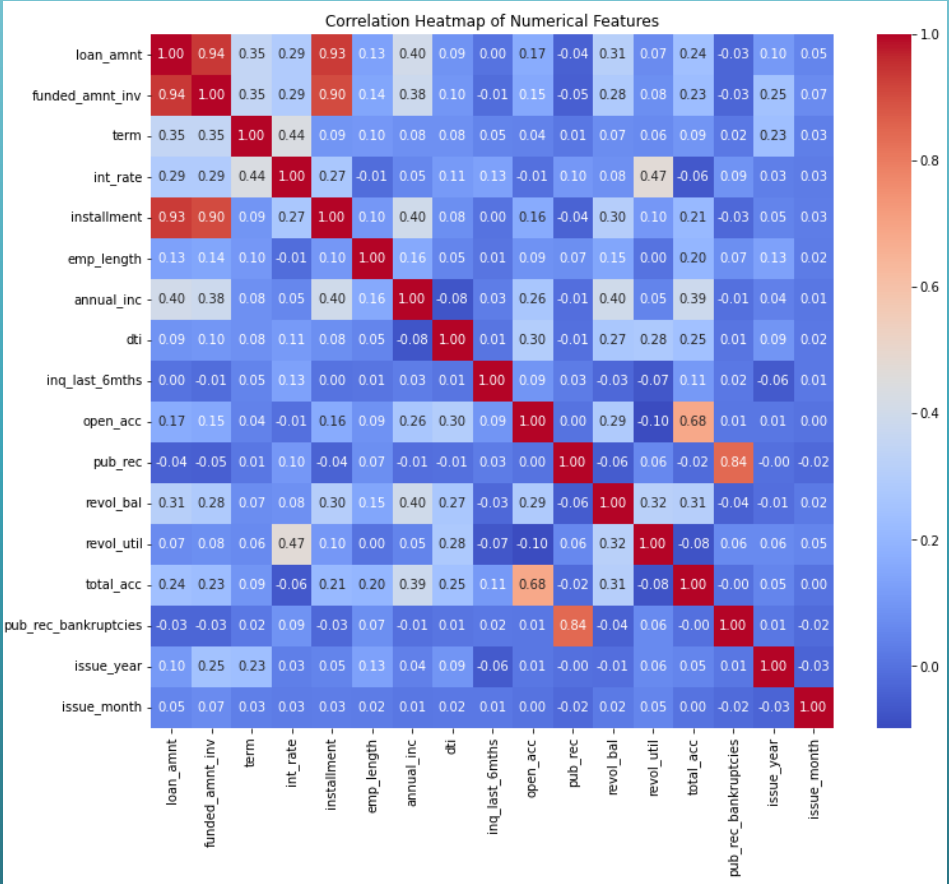


Annual Income:

- Income > 85k: Lower default probability.
- Income 3k-31k: Higher default probability.
- Higher income: Decreased charged-off loan proportion.



Correlation



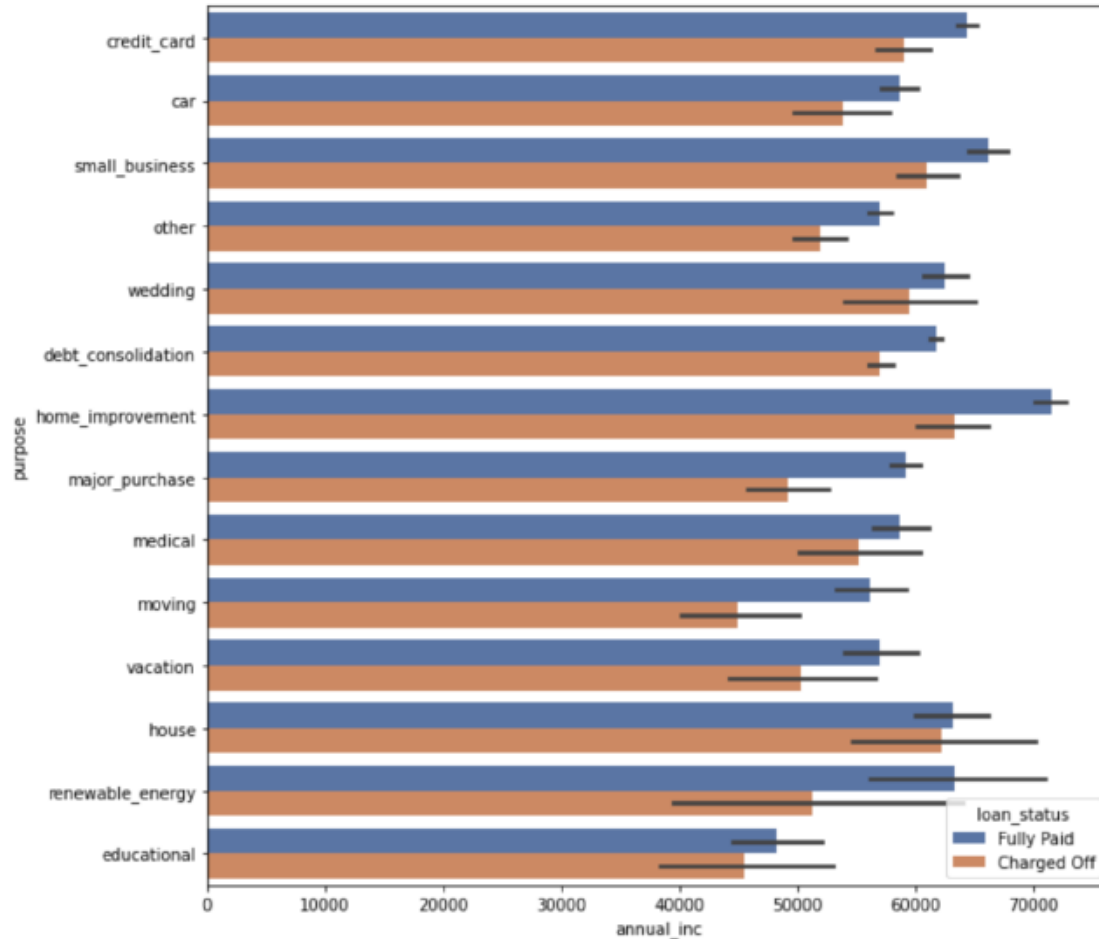
Negative Correlation:

- Loan Amount with Public Record Bankruptcies.
- Annual Income with Debt-to-Income Ratio

Strong Correlation:

- Term has a strong correlation with loan amount
- Term has a strong correlation with interest rate
- Annual income has a strong correlation with loan_amount

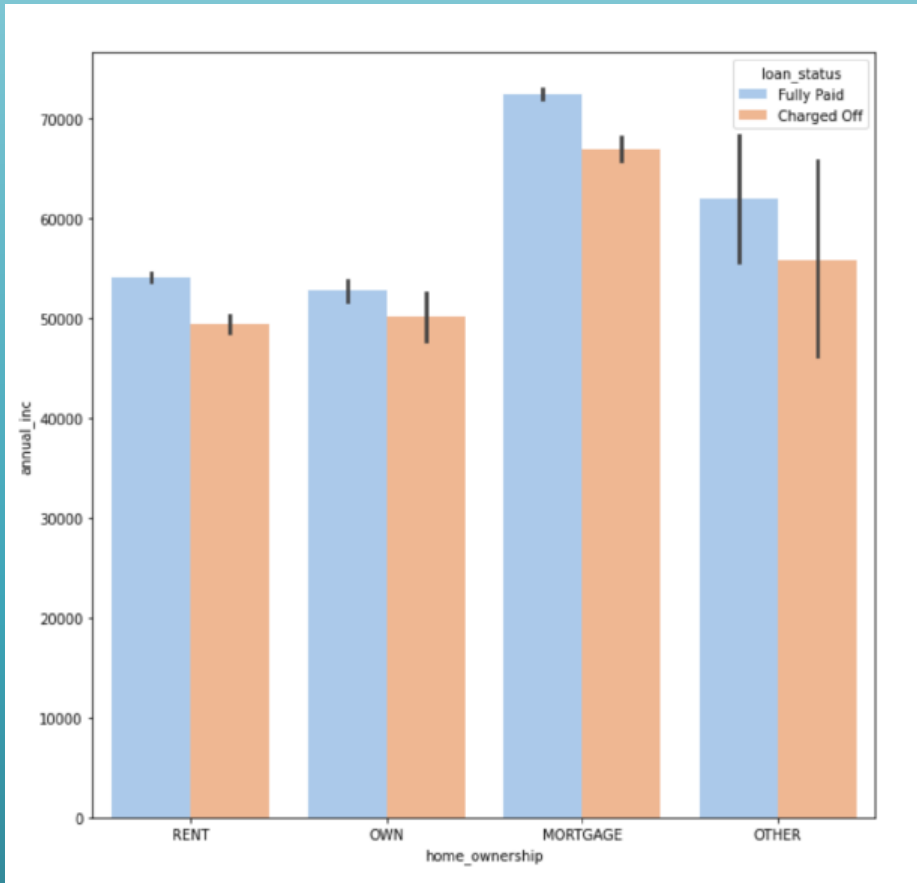
Multivariate Analysis



□ ANNUAL INCOME, PURPOSE AND LOAN STATUS

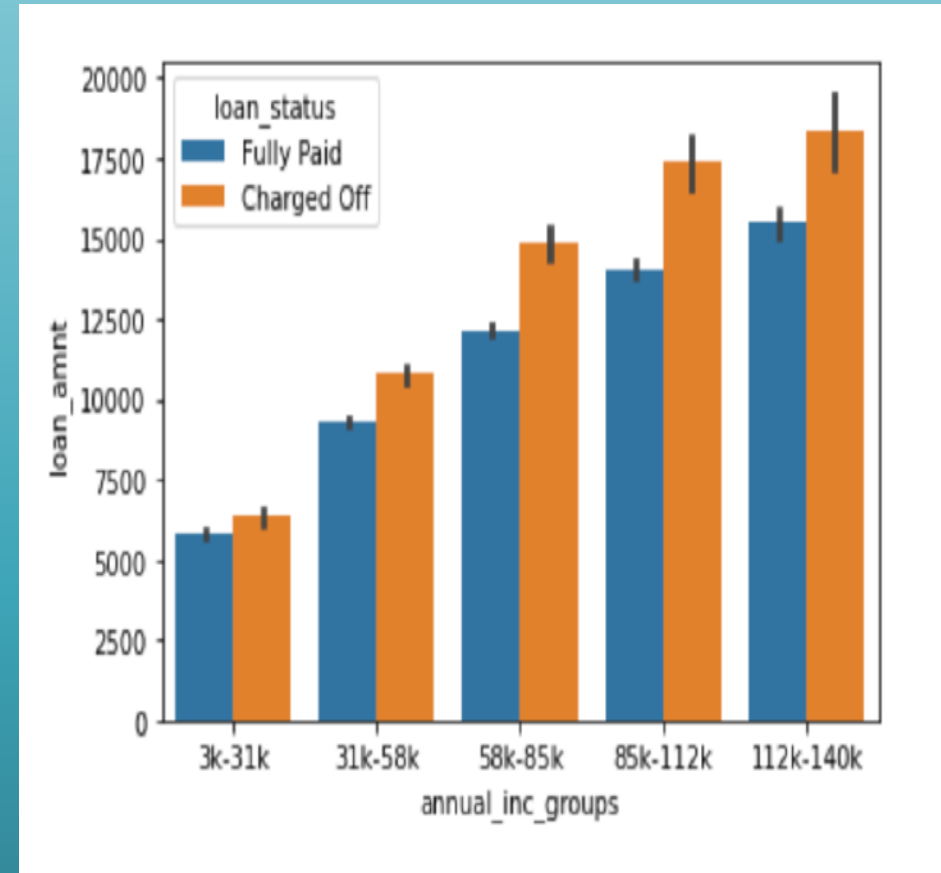
- I. **Loan Purpose:** Highest number of loans applied and defaulted are for "debt consolidation."
- II. **Annual Income:** Applicants with higher salaries mostly apply for loans for "home improvement," "house," "renewable energy," "credit card," and "small businesses."
- III. **Default Likelihood:** Applicants taking loans for "home improvement" with an income of 60k-70k are more likely to default.

Multivariate Analysis



ANNUAL INCOME HOME OWNERSHIP AND LOAN STATUS

- Default Risk: More likely to default are applicants whose home ownership is 'MORTGAGE' and have an income of 60-70k.



ANNUAL INCOME LOAN AMOUNT AND LOAN STATUS:

- Across all income groups, higher loan amounts are associated with higher default risk.

Exercise caution for applicants with interest rates of 13%-17% and 10 years of employment.

Income Range and Purpose

Lower incomes (3k-31k) and small business loans correlate with higher default rates.

Interest Rate and Employment Length

Grade and Sub-Grade

Scrutinize Grade B and Sub-Grade B5 loans more closely due to higher default likelihood

Home Ownership

Renters are more likely to default; assess additional risk factors for renters.

Loan Purpose:

Debt consolidation loans pose higher default risk; evaluate financial stability thoroughly

Geographical Considerations

Focus on California, Nebraska, Nevada, and South Dakota for regional economic impacts on defaults

Revolving Line Utilization Rate and Balance

High utilization rates (60-80%) and balances increase default risk.

Derogatory Public Records and Installment Amount

Derogatory records and higher monthly installments (145-274) increase default likelihood.

Debt-to-Income Ratio and Funded Amount

Higher DTI ratios (12-18) and funded amounts (5k-10k) correlate with higher defaults.

Total Account and Credit Lines

Fewer credit lines (2-20) correlate with higher default rates

Loan Term

Longer terms (60 months) have higher default rates; assess repayment capacity

Verification Status and Inquiries

Not Verified" status and fewer recent inquiries correlate with higher defaults

RECOMMENDATIONS

THANK YOU