Lending Club Case Study

DHARMAGNA DHYANESH VYAS
MURALI KRISHNA KONDAPALLY

Objective

Implementing the EDA technique on a real-world problem, comprehending the insights, and presenting in a business-first manner are the goals of this case study.

- Benefits of the case study:
  - Provides insight into the application of EDA to actual business issues.
  - Additionally, a foundational understanding of risk analytics in banking and financial services is developed.
  - The data's application to reduce financial loss when lending money to customers.
  - Our comprehension of visualization and the appropriate charts for real-world data is enhanced.

Dataset Understanding

This dataset comprises details regarding loan applicants, encompassing their credit history, income, loan amounts, payment habits, and loan status. It evaluates elements such as creditworthiness, delinquencies, loan utilization, and employment status to forecast payback probability and associated lending risks. This information is crucial for making informed judgments regarding loan approvals and managing financial portfolios.

EDA: Step-by-Step Guide

01 Importing Necessary Libraries

Importing required libraries for data manipulation, visualization, and analysis

02 Initial Data Exploration

Checking the first few rows, data types, and basic statistics of the dataset

03 Loading the Dataset

Load the loan dataset and display its structure to understand its contents

04 Data Cleaning

The goal is to clean and structure the dataset for analysis.

05 Analysis

Apply statistical and visual methods to comprehend patterns and trends in order to derive insightful information.

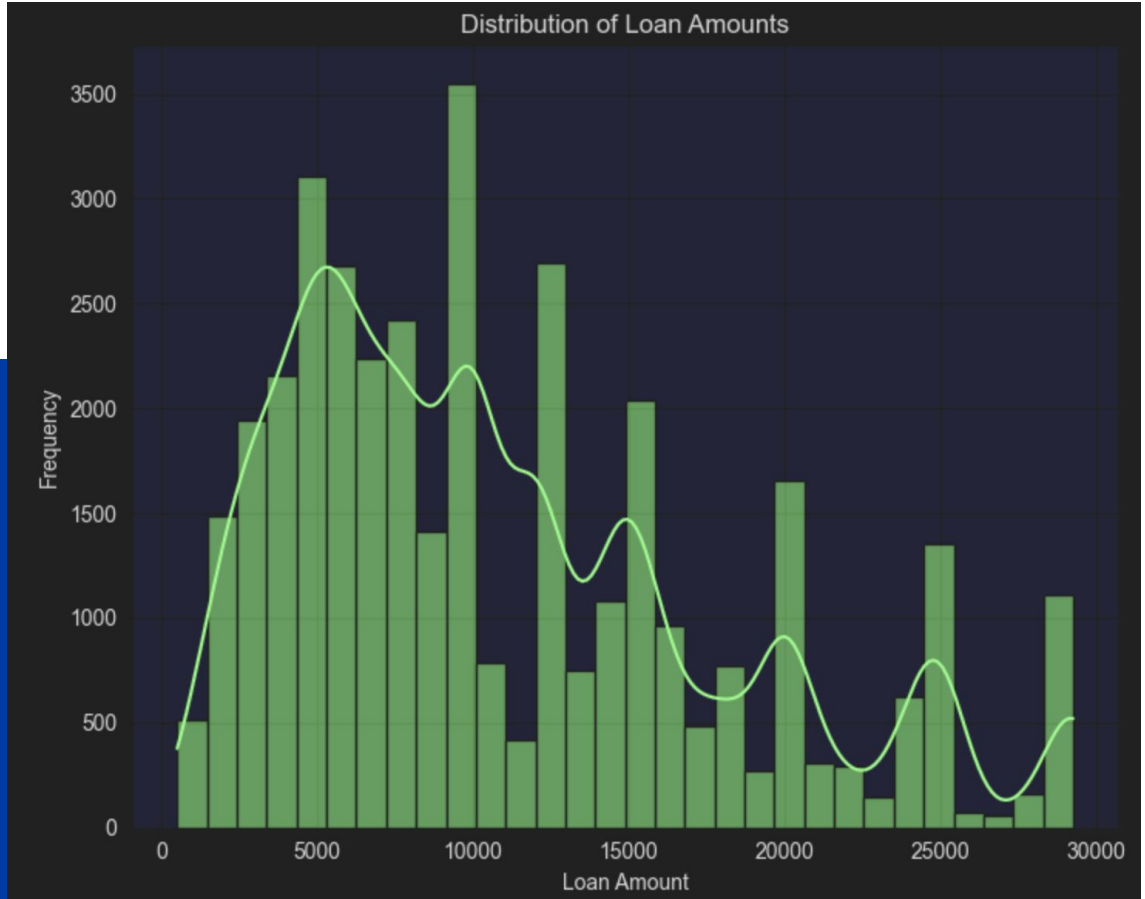
06 Conclusion

Based on the different analysis determine the factors for defaulting loans

Univariant Analysis

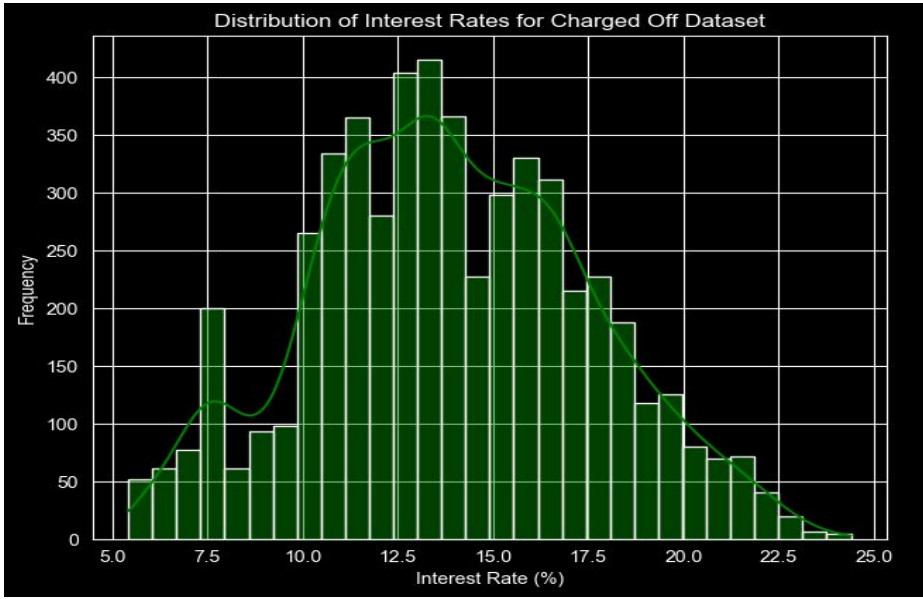
To examine the distribution of individual variables like loan amounts, interest rates, and employment length

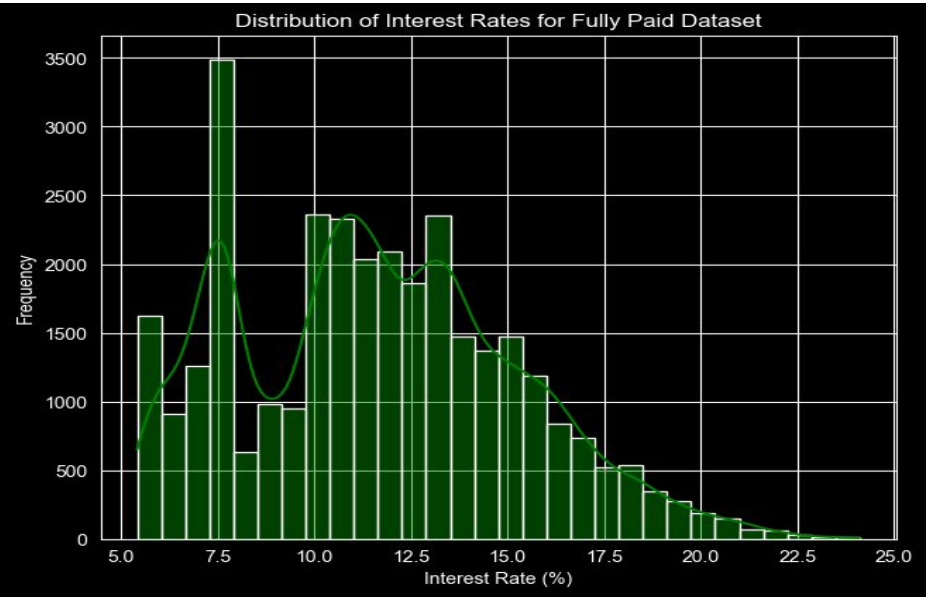
Distribution of Loan Amounts



- Loan amounts are concentrated between 5,000 and 15,000, with fewer loans exceeding 20,000.
- The apparent clustering around certain amounts like 10k, 20k, and 25k may reflect common borrowing patterns or lending policies.

Distribution of Interest Rates

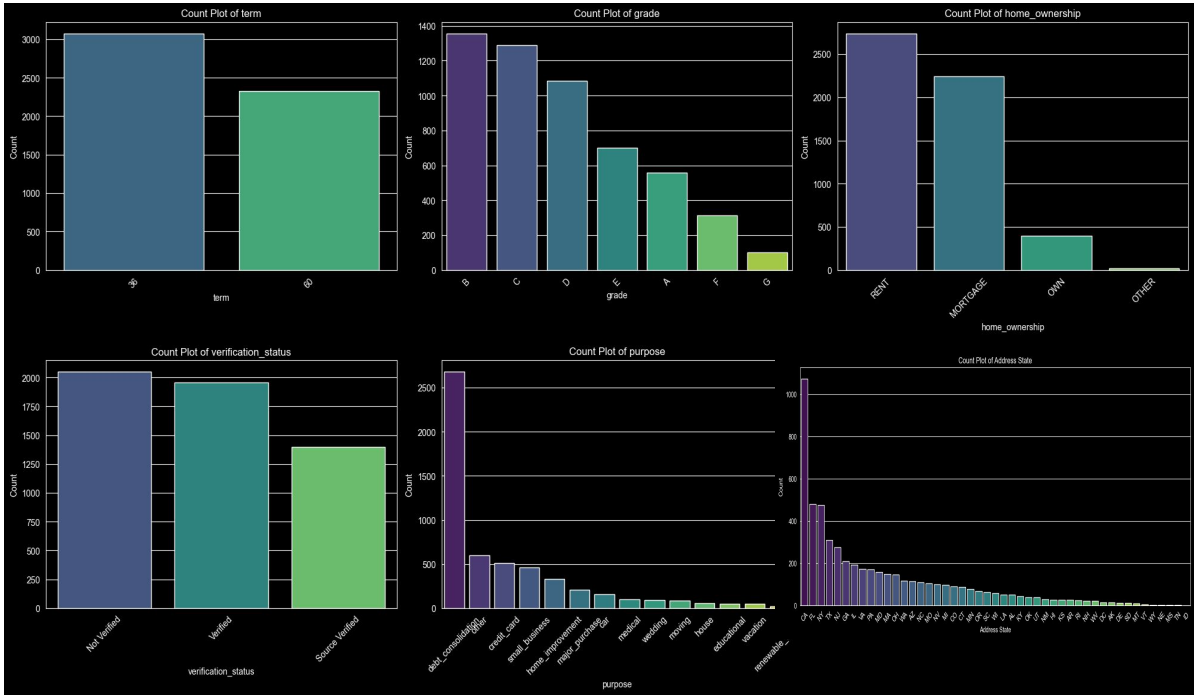




- The distribution of interest rates is concentrated between 10% and 15% and peak near 7.5%
- The majority of loans fall within the range of 10% to 15%
- Outliers are present beyond 22% which represents loans with significantly higher interest rates.
- For interest rates for charged off are mostly between 10%-17% when compared to fully paid it is between 10% - 15%

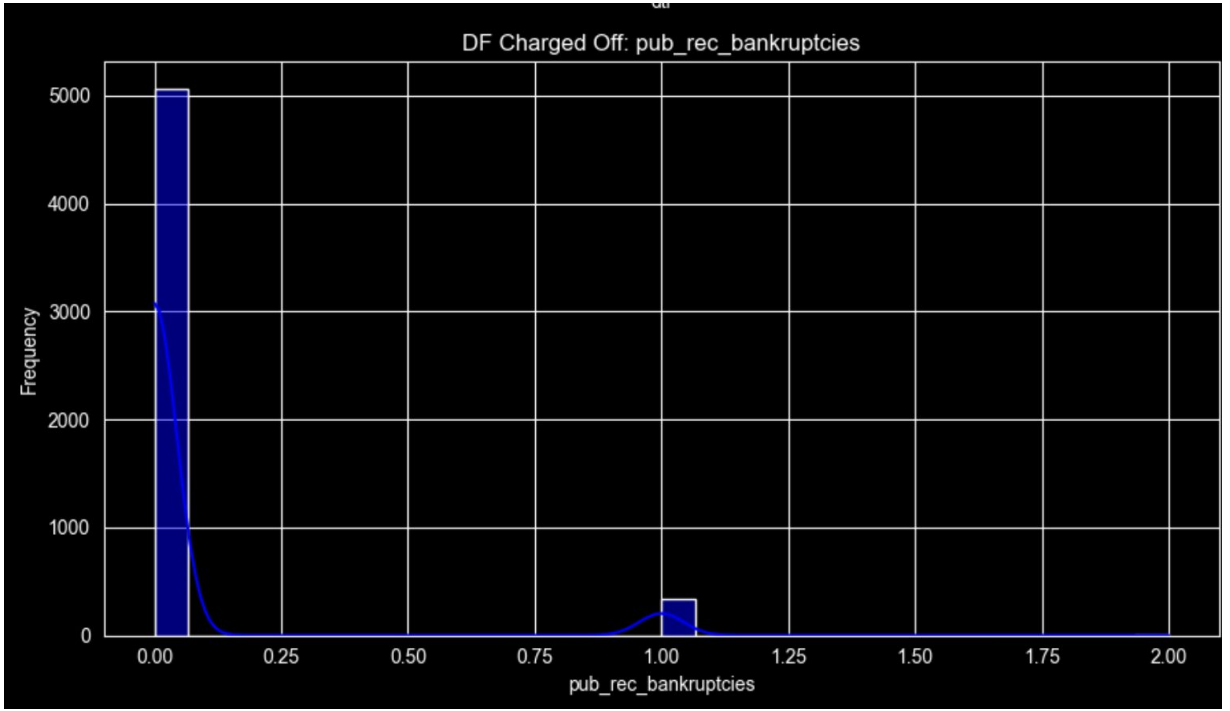
Analysis of Categorical Values for Charged off dataset

- Borrowers who have their home status as renting or mortgage have defaulted more when compared to own house
- Debt Consolidation and Credit Card are two primary reasons for defaulting the loans
- States such as California (CA), Florida (FL), New York (NY) and Texas (TX) have higher defaulters as this could be related population size of these states
- The defaulters are trending down from Grade **B - G**
- Not Verified & Verified have defaulted more compared to Source Verified



Bankruptcies filed for Charged off Loans

- Most of the people who defaulted have not filed for bankrupt

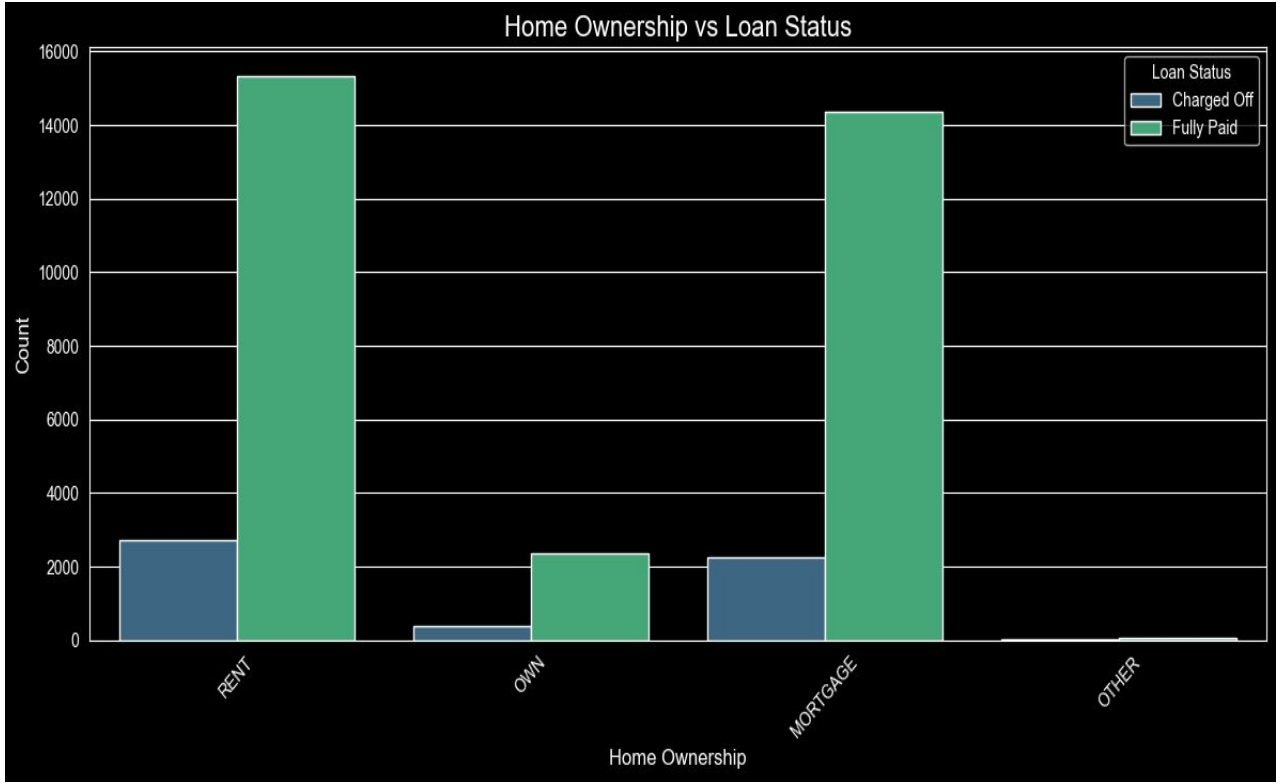


Bivariant Analysis

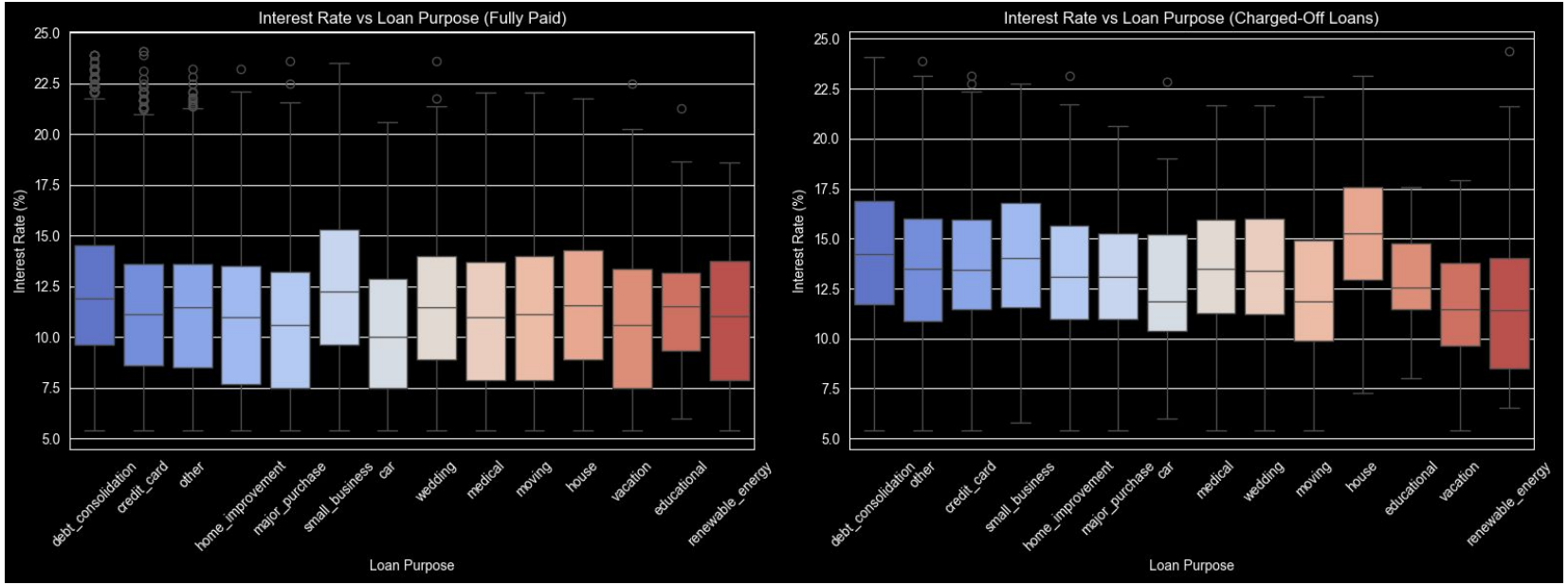
To determine the correlation of more than one variables like home ownership & loan status, interest rates & purpose

Home Ownership vs Loan Status

- Most loans are given to borrowers who are either mortgaged or renters.
- Defaulters are mostly who in category of Rent and Mortgaged house



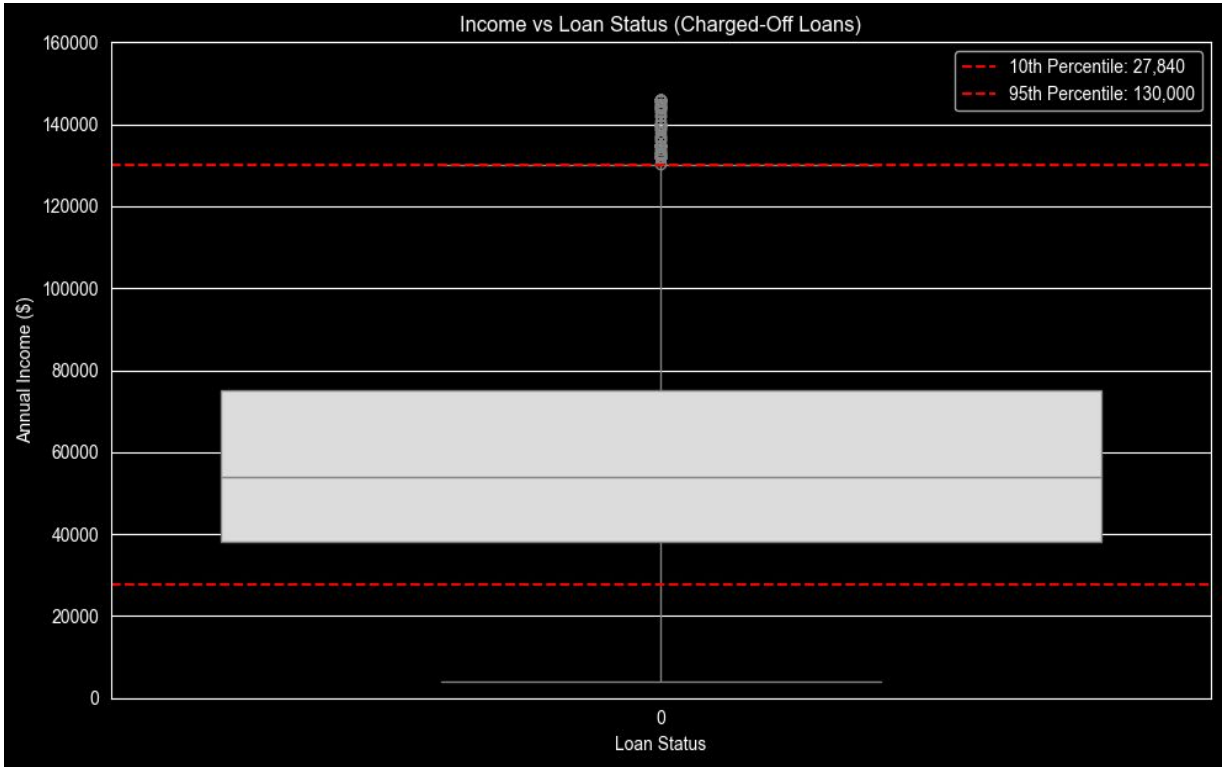
Interest Rate vs Loan Purpose



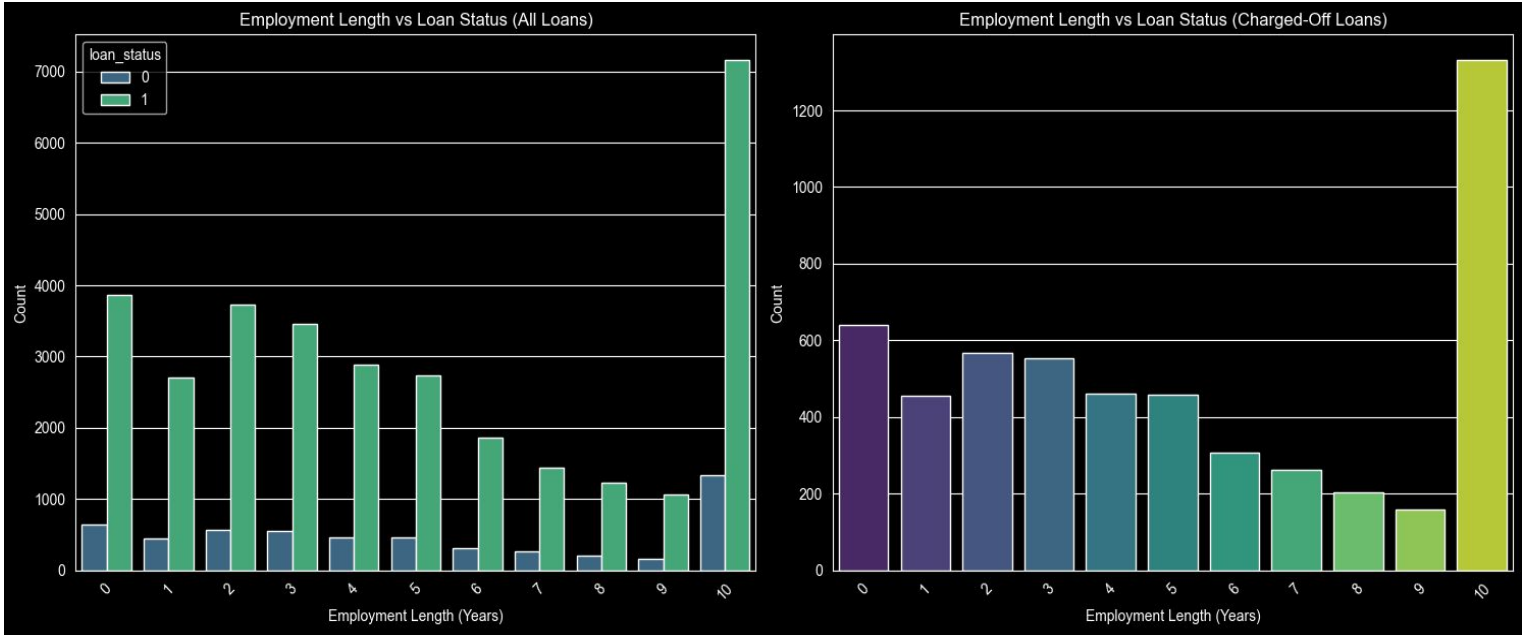
- In General, Charged off loans have higher interest rates which might be one of the reason for defaulting
- House Category loans have much higher interest rate ~17.5% which could be another reason for defaulting
- Small Business loans have relatively higher interest rates which is between 15% - 17% again adding to cause for defulting loans
- Debt Consolidation and Credit Card are top 2 contributors for loan defaulting

Income vs Loan Status (Charged-Off Loans)

- Income range between 40k to 75k have defaulted more
- Borrowers with higher income seem more likely to default their loans.
- Higher income people also have defaulted more when compared to fully paid. There are lot of outliers w.r.t Income range (ex: 130k - 150k)



Employment Length vs Loan Status



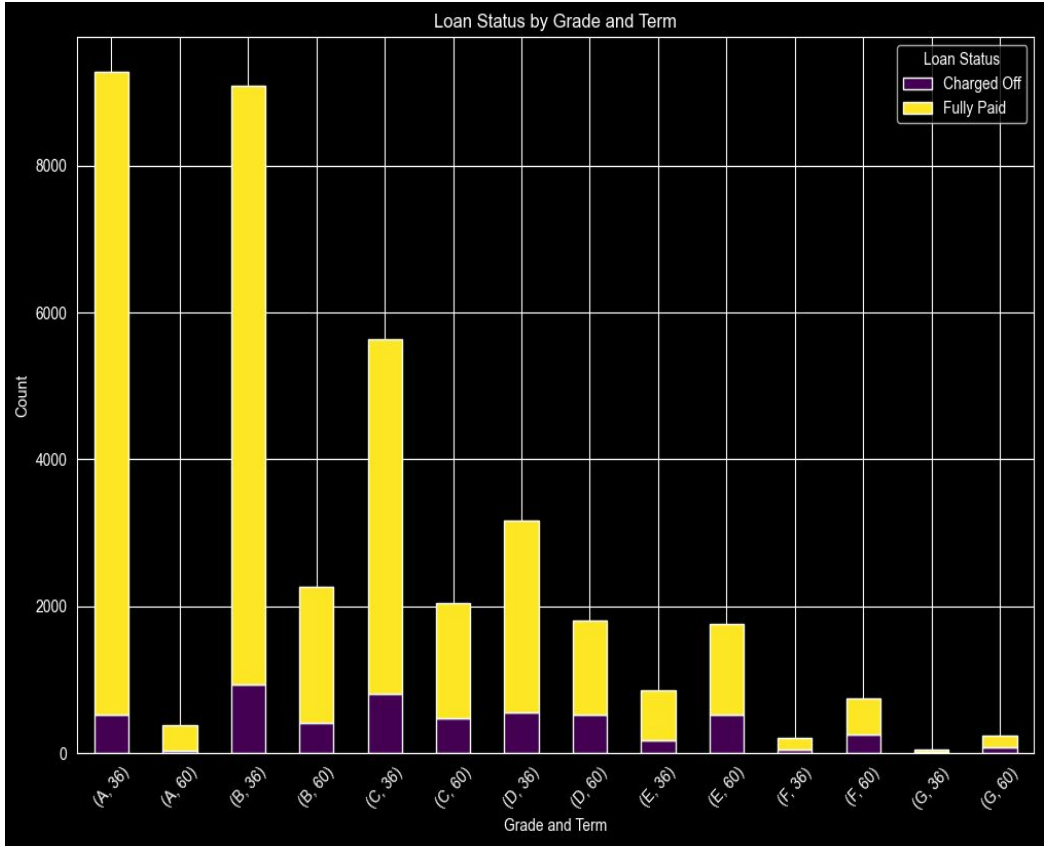
- Most of the defaulters have tenure between 0-1 Years or greater than 10 years
- Majority of Fully paid loans have higher employment tenure which is +10 years

Multivariant Analysis

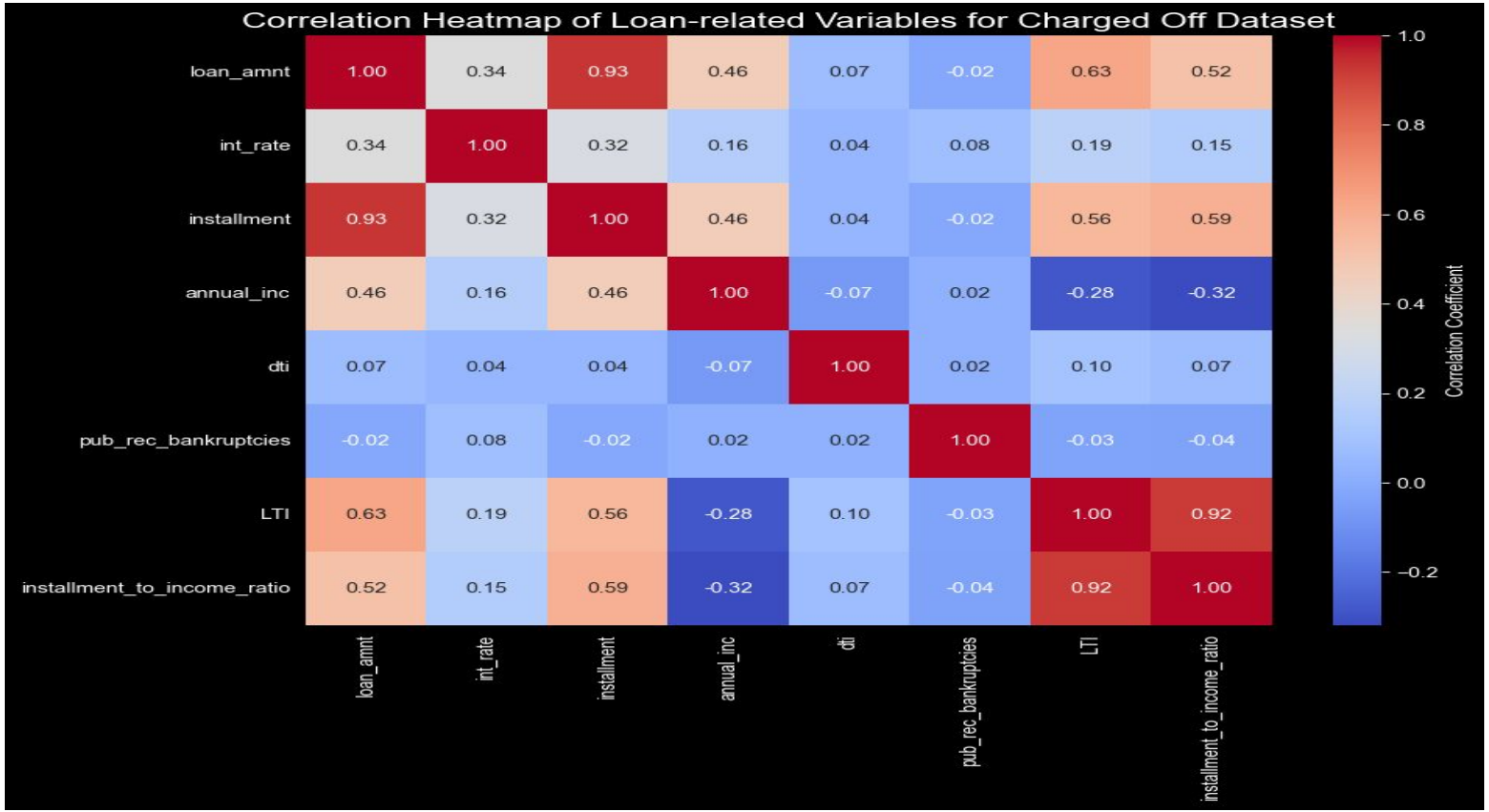
To understand the correlation of multiple variables find connections that affect loan results, eventually offering useful information for evaluating risk and making decisions.

Income vs Loan Status (Charged-Off Loans)

- Loan Grades A and B (36-Month Term) have the highest number of loans, with a majority being fully paid.
- Charged-off loans are significantly lower in comparison. ~ 14% of people have defaulted out of complete dataset
- B, C grades with 36 months tenure have defaulted more when compared to other grades
- Higher Grades (D, E, F, G) have increasing proportion of charged-off loans relative to fully paid loans.
- 60-month term loans show a higher proportion of charged-off loans compared to 36-month term loans. This also indicate longer loan durations have bigger risk.



Correlation Chart Analysis



Correlation Chart Analysis (Contd..)

- The somewhat positive connection (0.63) suggests defaulters had larger debt amounts than income. This may indicate overborrowing risk.
- The moderate connection (0.46) between annual income and loan size shows that high-income borrowers may have defaulted on significant debts.
- Income alone isn't a good predictor of repayment behavior; higher income defaulters had better LTI ratios but defaulted, according to a negative correlation with LTI (-0.28).
- The small positive correlation (0.34) between interest rate and loan size suggests that defaulters with larger debts did not always get considerably higher interest rates.
- The data suggests that DTI is not a significant predictor of default behavior due to the poor correlations across variables (interest rate: 0.04; loan amount: 0.07).
- A small correlation with installment (0.59) indicated that defaulters frequently had high installments relative to their income.
- A negative correlation with annual income (-0.32) suggests that lower-income borrowers were more likely to take on disproportionately high installment burdens, which may have contributed to defaults.
- For defaulters, bankruptcies did not significantly differentiate them.

Conclusion

Conclusion

Summary of the Study

The analysis gives us important information about the trends and causes of loan defaults:

Amount of the loan and changes in interest rates:

- The majority of loan amounts are between $5,000 and $15,000. There is also a noticeable cluster between $10,000 and $20,000, which is likely due to popular borrowing or lending patterns.
- The interest rates on charged-off loans are slightly higher (10% to 17%) than those on fully paid loans (10% to 15%), with some loans having rates higher than 22%, which could mean they are riskier.

Home Ownership and Default Behavior:

- People who rent or mortgage their homes are more likely to not pay back their loans than people who own their own houses.
- A lot more people don't pay back loans that were taken out to consolidate debt or pay off credit cards.

Location and patterns of employment:

- The failure rates are higher in states like California, Florida, New York, and Texas, which may be because they have more people.
- Most people who don't pay their loans either have been employed for a short time (0–1 years) or a long time (10 years or more). Borrowers who have been employed for 10 years or more usually pay their loans in full.

Conclusion (Contd..)

- Types of Loans and Terms: Loans with terms of 36 months and Grades B and C have the highest default rates. Loans with terms of 60 months or more have a higher risk of failure.
- Higher Grades (D, E, F, and G) have higher percentages of failures compared to fully paid loans.
- Borrowers with annual incomes between $40,000 and $75,000 are the most likely to not pay back their loans.
- Even people with higher incomes sometimes don't pay back bigger loans, showing that income alone isn't a good way to predict how people will return their debts.

Important numbers:

- A lot of loans compared to income and a lot of payments compared to income Ratios are popular among people who don't pay their debts, which shows that too much borrowing or big monthly payments are major risk factors.
- People with lower incomes often have to pay a lot of money for their monthly payments, which can lead to failures.

Few thoughts:

- The interest rates on debt consolidation and small business loans are higher (15–17%), which makes the risk of failure even higher.
- Bankruptcies don't make a difference between defaulters, which suggests that things like income & spending habits are important.

Conclusion (Contd..)

Important Points:

- High interest rates, large loan amounts compared to income, and heavy monthly payments are all strong indicators of defaulters.
- Policies that focus on stable income, reasonable payment loads, and stricter risk ratings for bigger loans could help lower the number of defaults.
- Loans for things like debt consolidation, small businesses, and credit card payments should be closely looked.