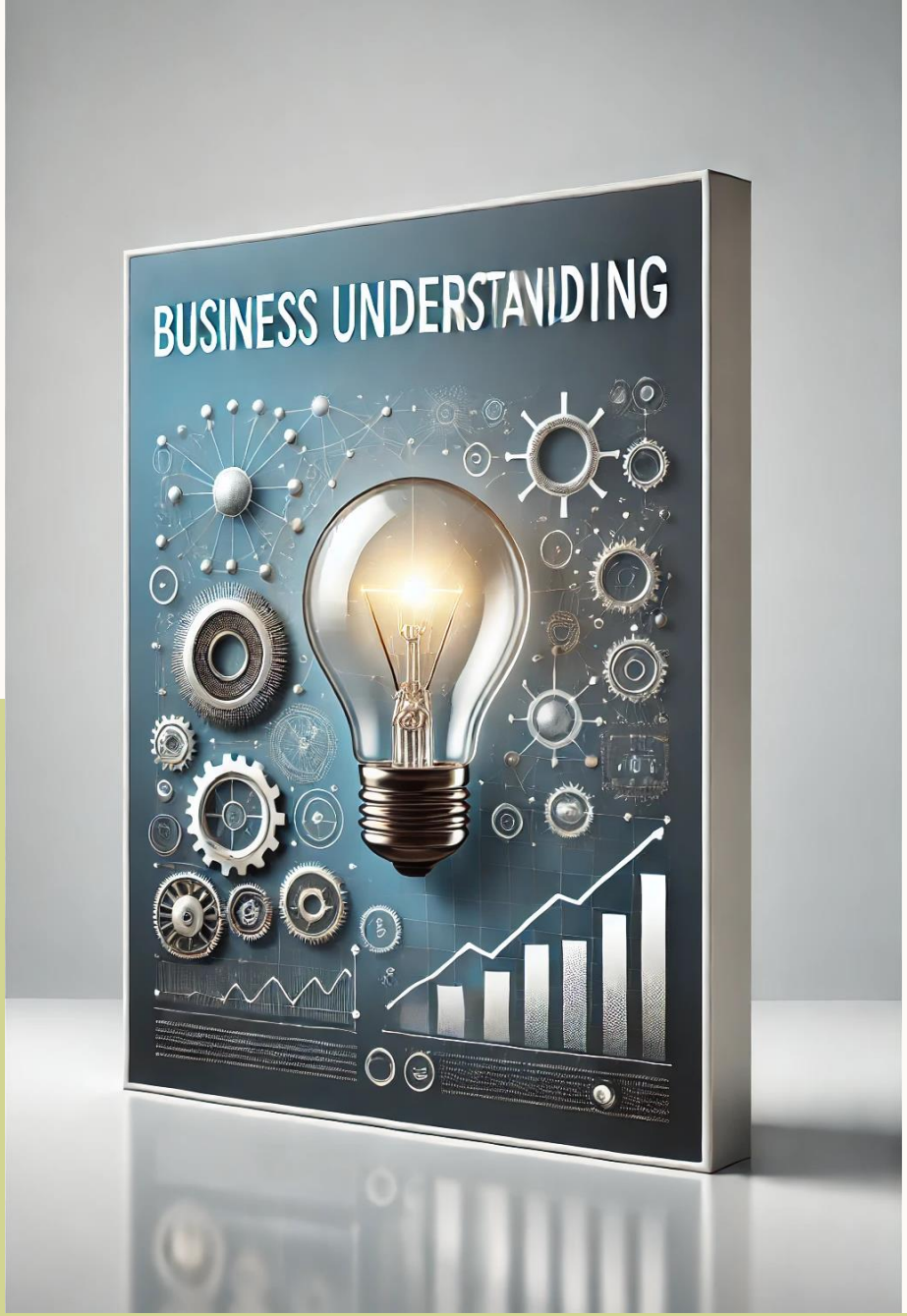
LENDING MONEY
CASE STUDY

NIKHIL LUNAWAT (FACILITATOR)
NIKHIL UGALE
ML – C67

INTRODUCTION

Risk analytics plays a critical role in the banking and financial services sector, especially when it comes to lending decisions. This case study delves into how data is leveraged to minimize the risks associated with loan defaults. By exploring patterns in consumer and loan attributes, the goal is to identify the key drivers behind loan default, allowing the finance company to make informed decisions, optimize lending strategies, and reduce credit loss.



BUSINESS UNDERSTANDING

The finance company specializes in providing various types of loans to urban customers. As with any lending institution, there are two primary risks associated with every loan approval decision:

1. Credit loss risk: if a risky applicant defaults, the company suffers a financial loss.
2. Business loss risk: if a loan-worthy applicant is denied, the company loses potential business.

The objective of this analysis is to identify patterns in historical loan data that highlight risky applicants, thus helping the company reduce its overall credit losses. By understanding the factors that drive loan defaults, the company can either reject high-risk applicants or adjust loan terms and interest rates accordingly.

DATA UNDERSTANDING

The dataset comprises historical data on loan applicants over a period from 2007 to 2011. Each loan issued is characterized by multiple attributes such as loan amount, interest rate, loan purpose, annual income, and loan status (whether the loan was fully paid, currently being repaid, or charged-off).

Some key attributes in the dataset include:

1. Loan Amount: The amount requested by the applicant.

2. Interest Rate: The rate of interest on the loan.

3. Term: The duration of the loan (in months).

4. Purpose: The stated purpose of the loan (e.g., debt consolidation, small business).

5. Annual Income: The applicant's annual income at the time of applying for the loan.

6. Loan Status: The final status of the loan (Fully Paid, Charged-off, or Current). This dataset allows us to explore how various borrower characteristics and loan attributes influence the likelihood of default.

ANALYSIS APPROACH

To uncover insights from the dataset, the following exploratory data analysis (EDA) techniques were used:

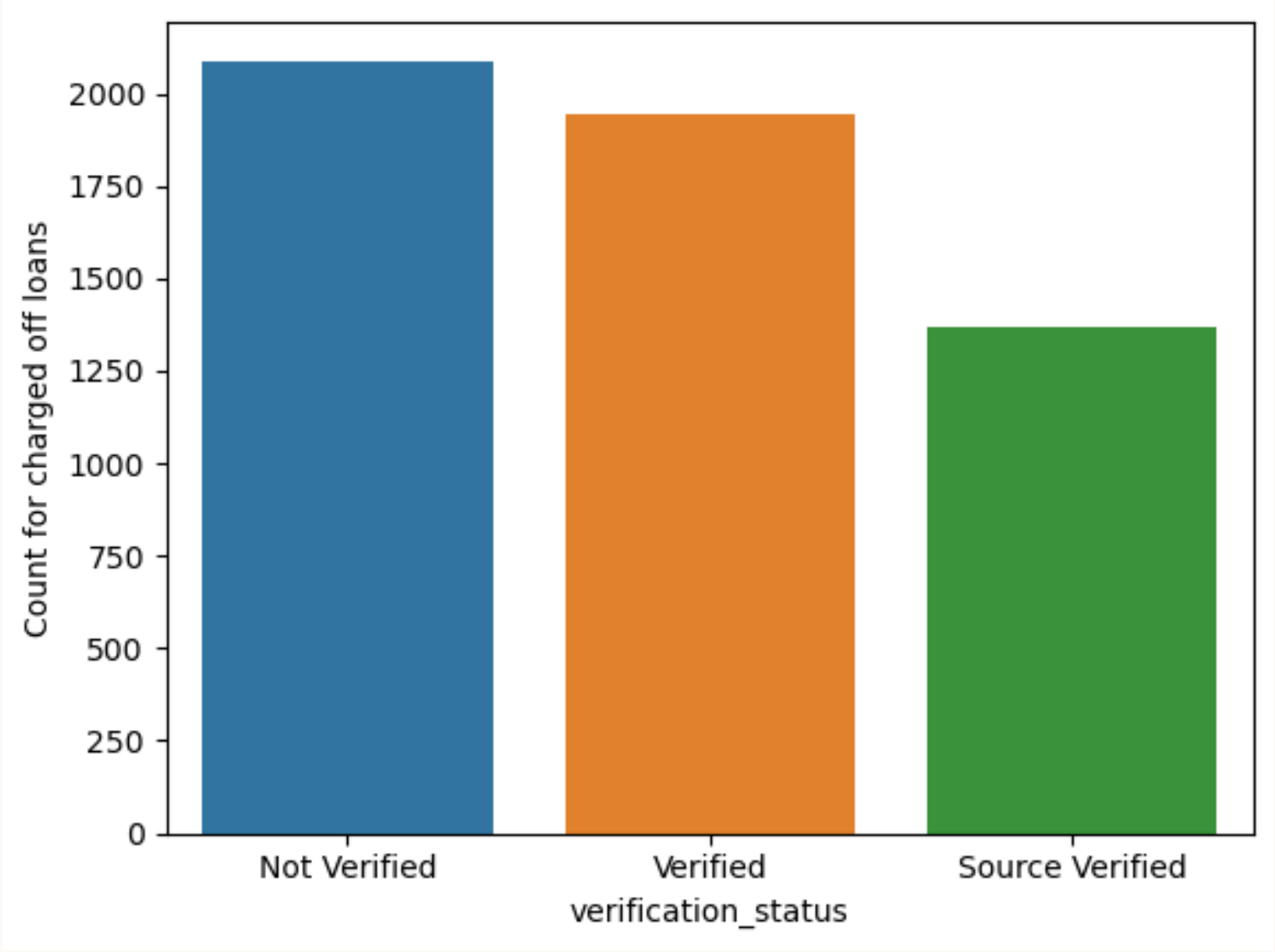
1. Univariate Analysis: This technique was used to explore individual attributes in the dataset. For example, the distribution of loan amounts or interest rates among applicants provides insight into typical loan characteristics.

2. Bivariate Analysis: Bivariate analysis was used to explore relationships between two variables. For example, the relationship between loan grade and loan default provides insight into whether riskier loan grades are associated with higher default rates

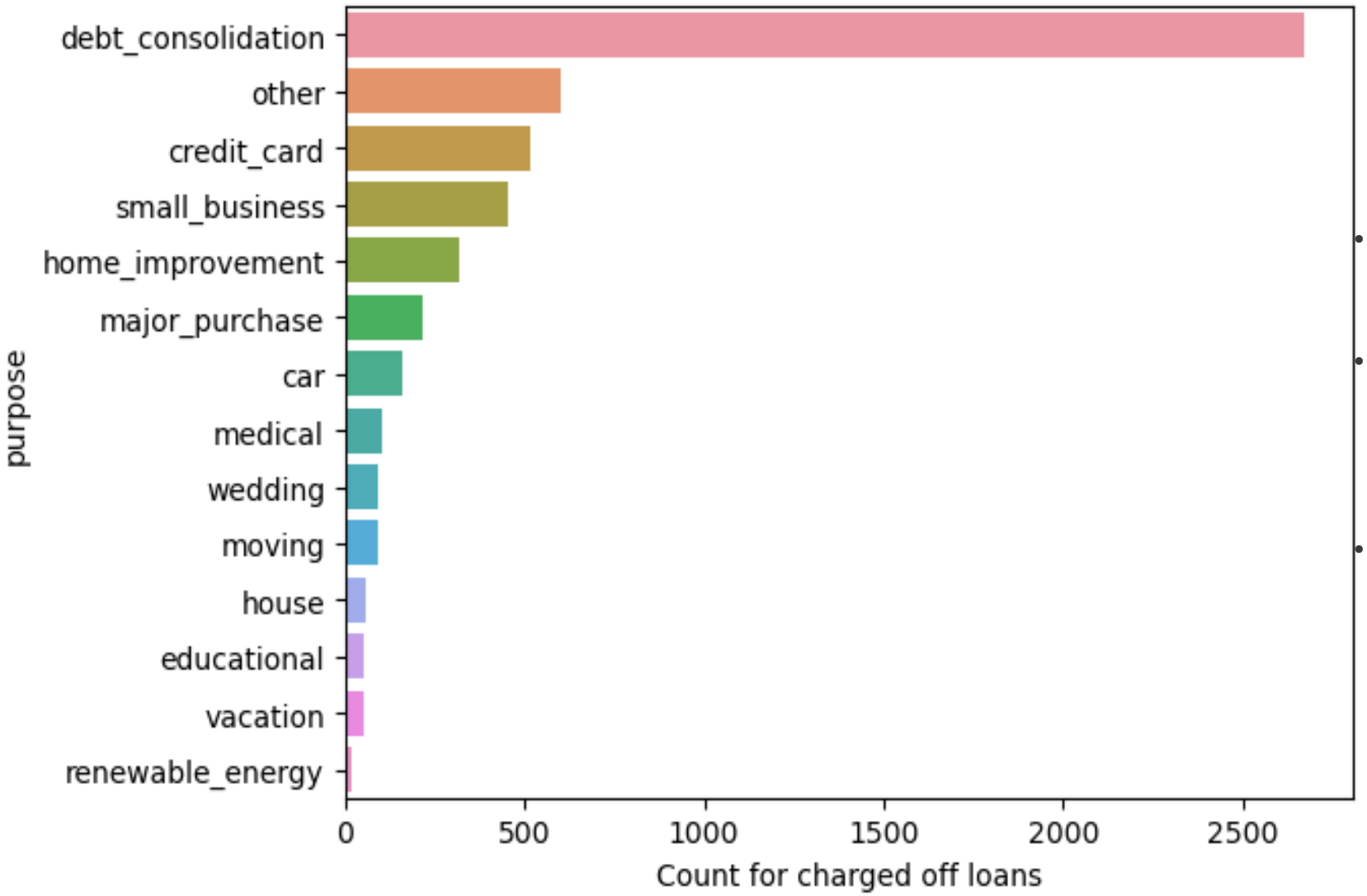
. 3. Multivariate Analysis: This analysis helped identify more complex interactions between multiple variables. For instance, the relationship between loan purpose, interest rate, and annual income can highlight nuanced trends related to loan default

UNIVARIATE ANALYSIS

This bar chart displays the count of charged-off loans across three verification statuses: "Not Verified" (highest count), "Verified," and "Source Verified" (lowest count), indicating loan performance based on verification.

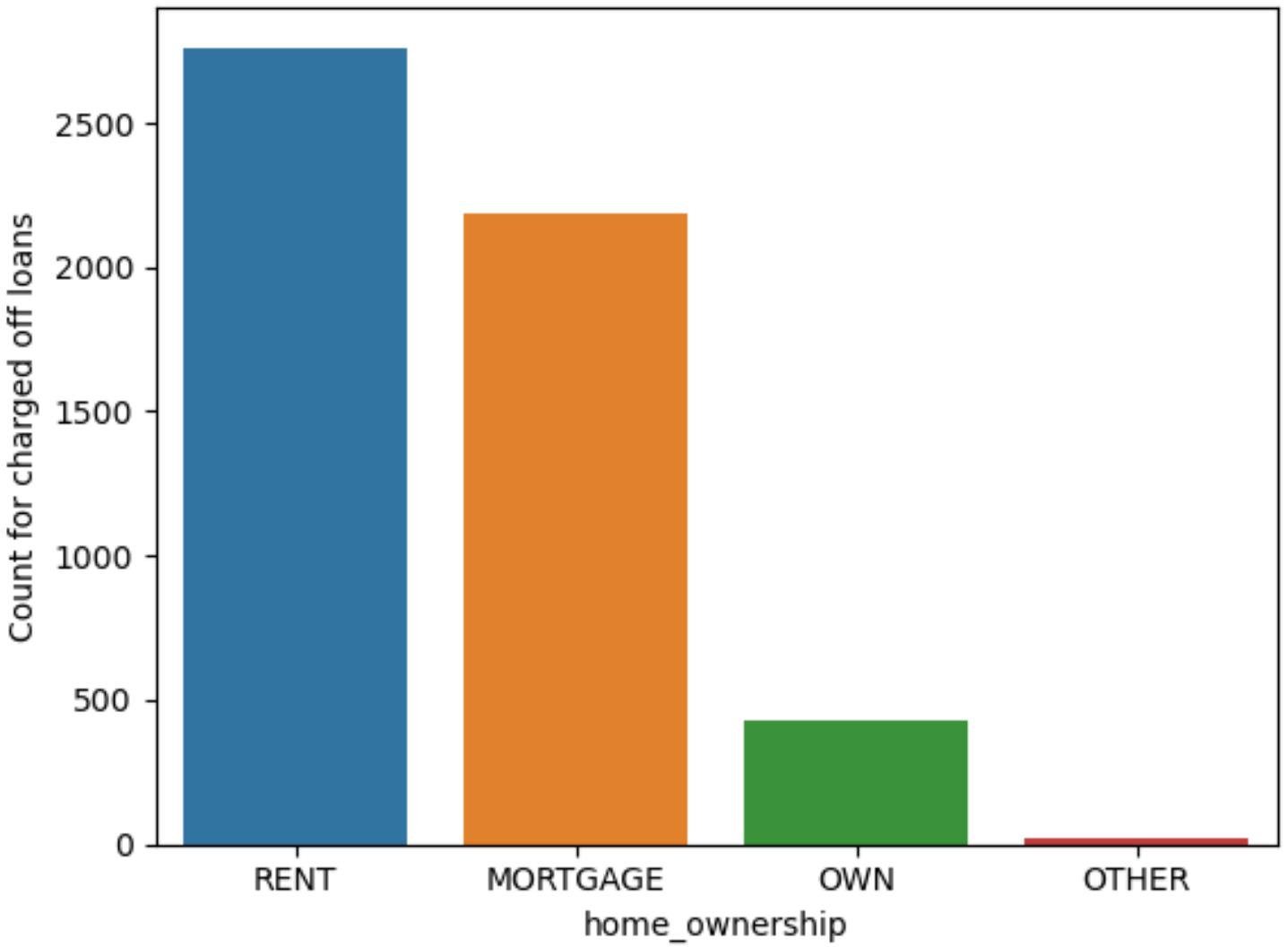


PURPOSE OF LOAN



- This bar chart illustrates the distribution of loans based on their purpose.
- "Debt consolidation" is the leading loan purpose, representing the majority of loans, with over 2,500 applicants, indicating that many borrowers seek to consolidate their existing debts.
- Other notable loan purposes include "credit card," "small business," and "home improvement," each showing a significant number of applicants. Less frequent loan purposes include "wedding," "moving," and "renewable energy," showing that fewer people apply for loans in these areas. The variety of loan purposes reflects the diverse financial needs of borrowers.

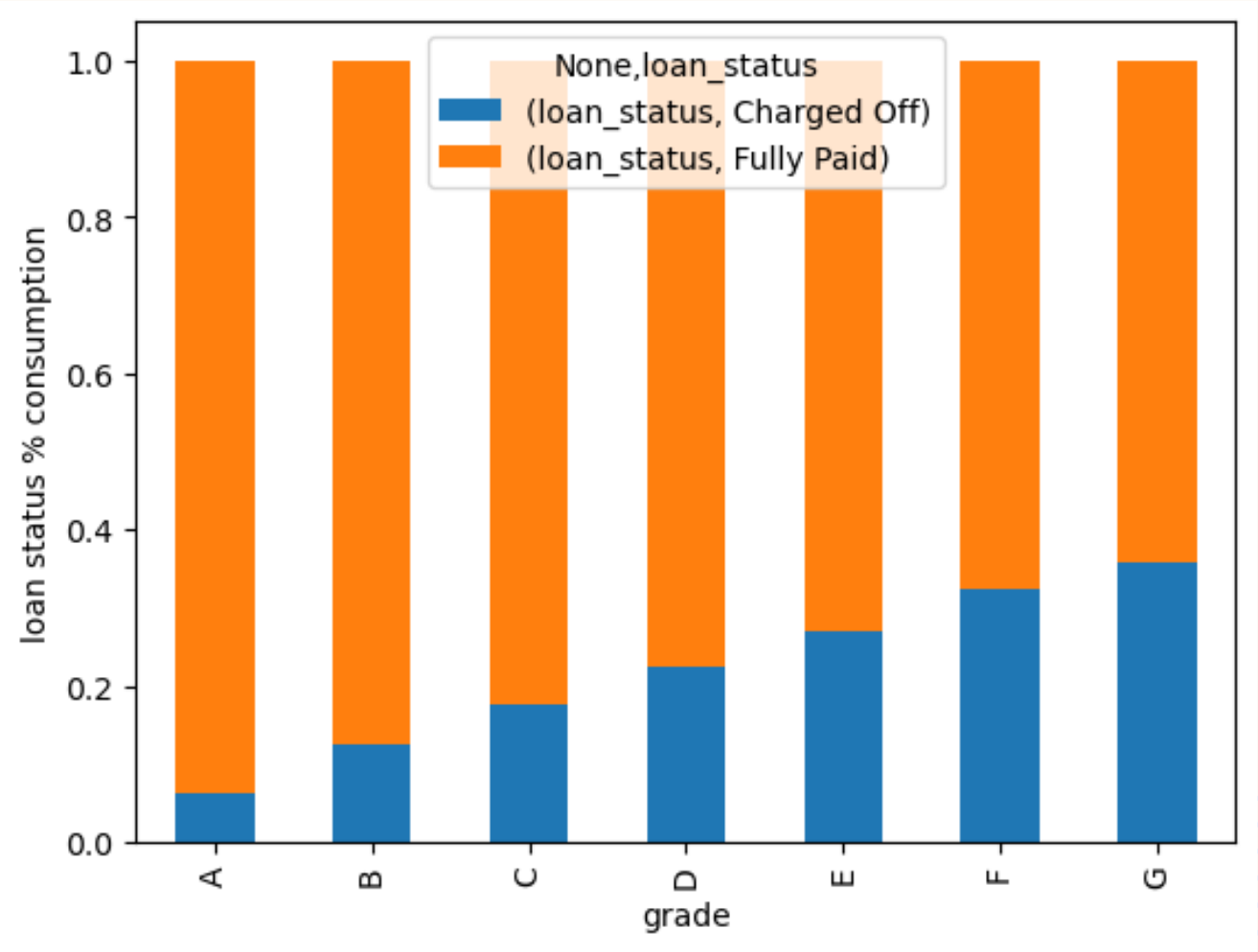
HOME OWNERSHIP DISTRIBUTION

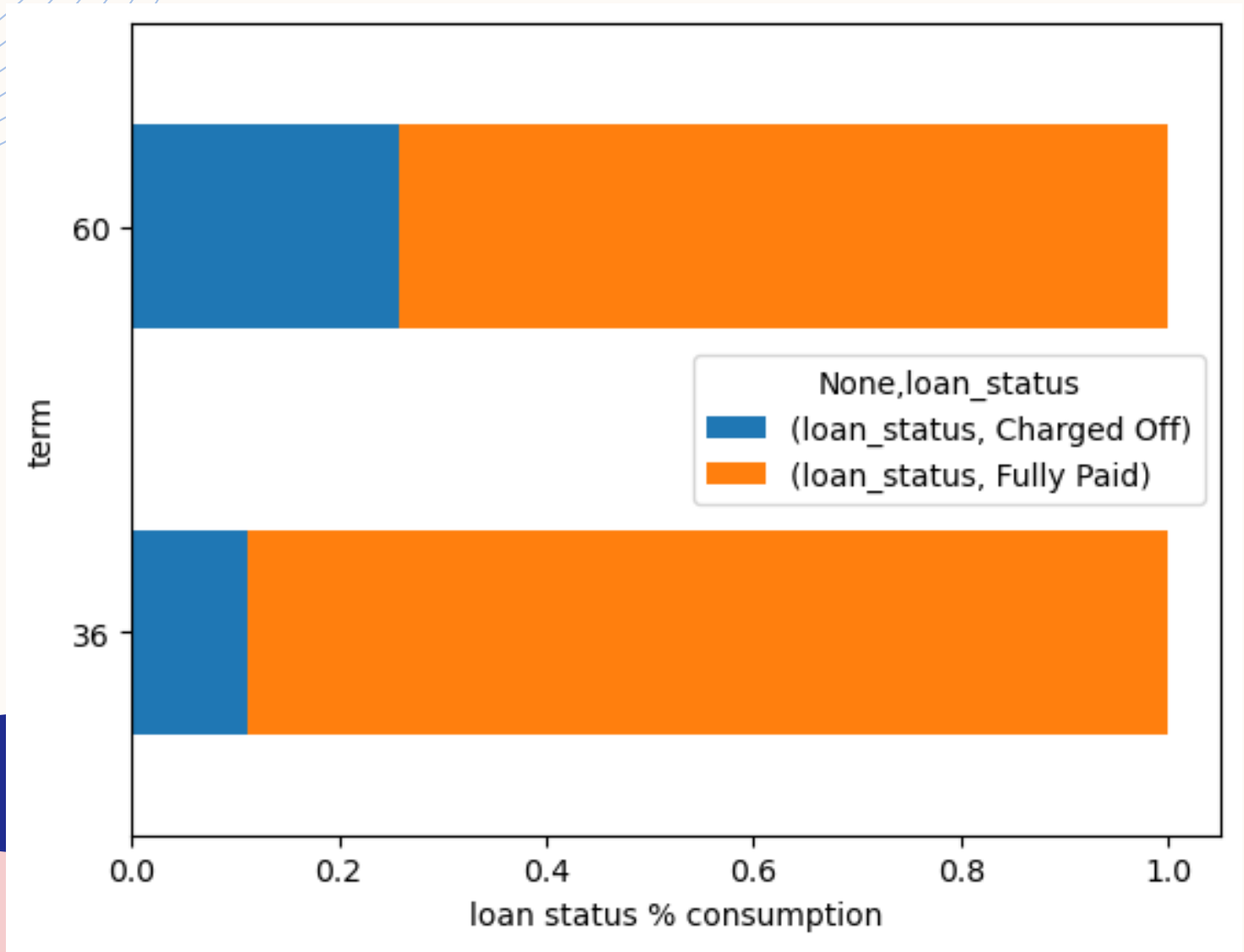


This bar chart illustrates the count of charged-off loans based on home ownership status. Borrowers who rent have the highest number of charged-off loans, followed by those with mortgages. Individuals who own their homes and those with "Other" ownership status have significantly fewer charged-off loans.

SEGMENTED ANALYSIS

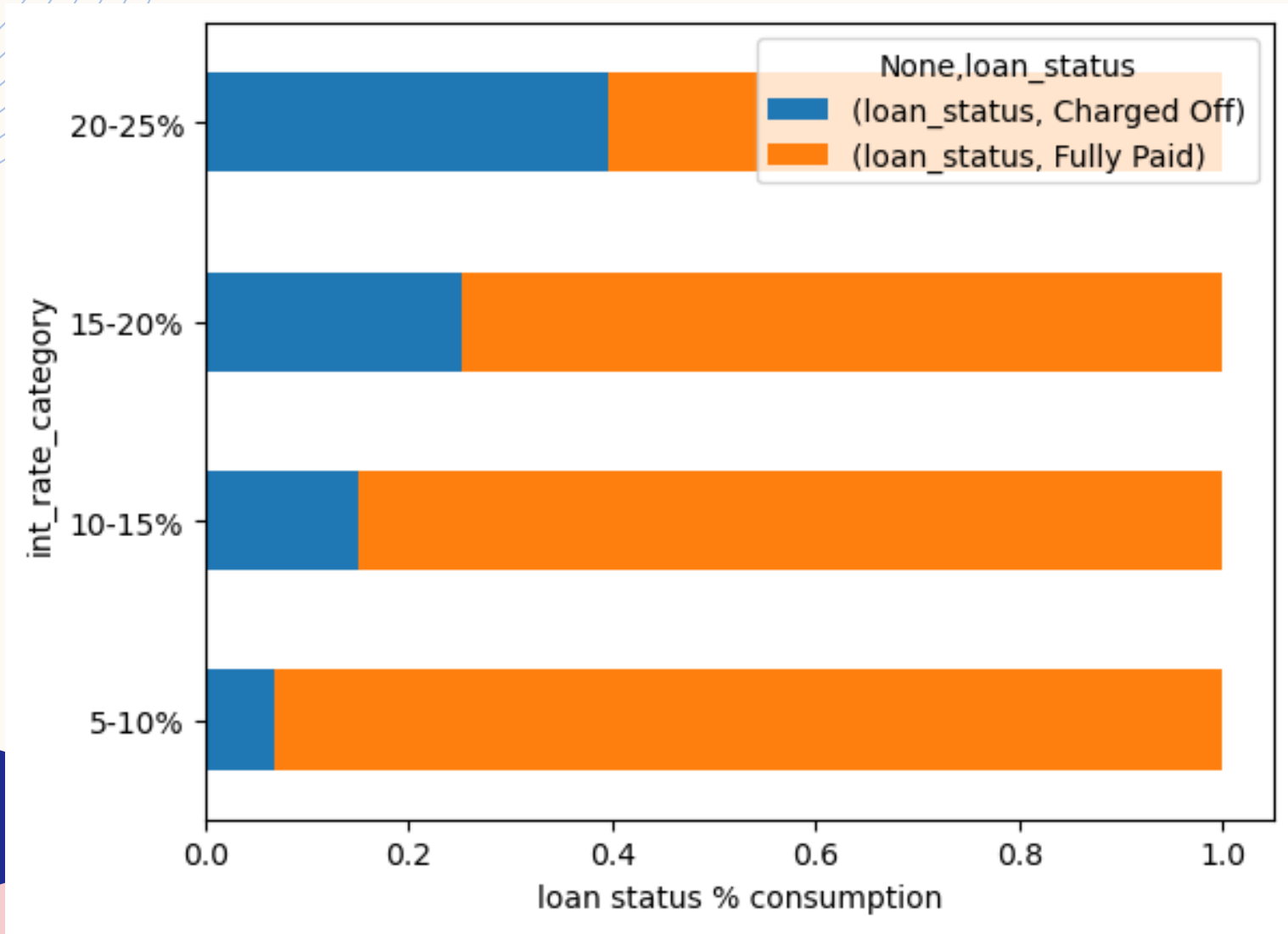
This stacked bar chart displays the percentage distribution of loan statuses (charged off vs. fully paid) across different loan grades (A to G). Grades A and B have the smallest percentage of charged-off loans, while grades E, F, and G show a higher proportion of charged-off loans compared to fully paid ones.





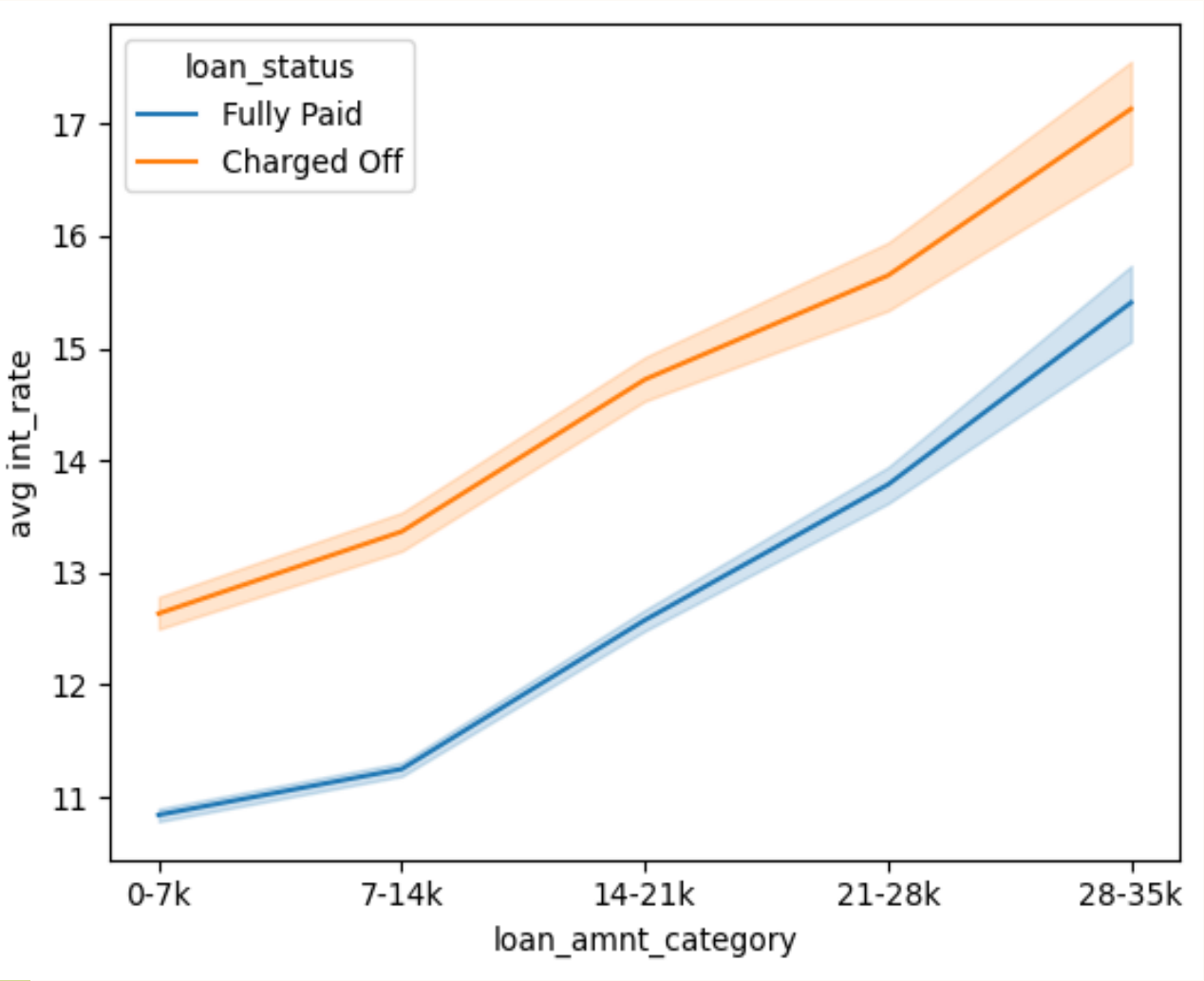
This horizontal stacked bar chart shows the percentage distribution of loan statuses (charged off vs. fully paid) across two loan terms: 36 months and 60 months.

Loans with a 60-month term have a higher percentage of charged-off loans compared to those with a 36-month term, where fully paid loans dominate.



This horizontal stacked bar chart shows the percentage distribution of loan statuses (charged off vs. fully paid) across different interest rate categories.

Higher interest rate categories (20-25% and 15-20%) have a larger percentage of charged-off loans compared to lower interest rate categories (5-10% and 10-15%), where fully paid loans dominate.

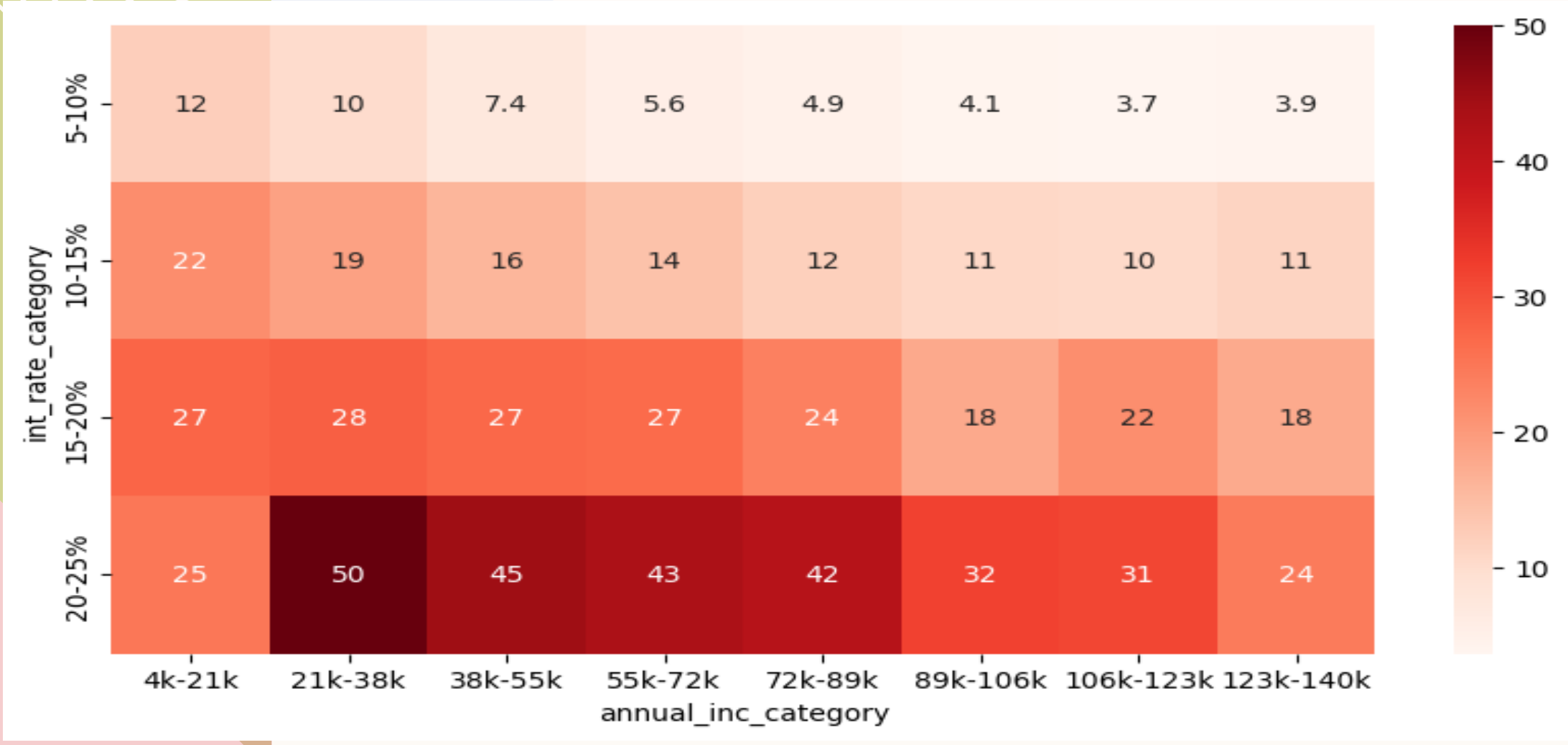


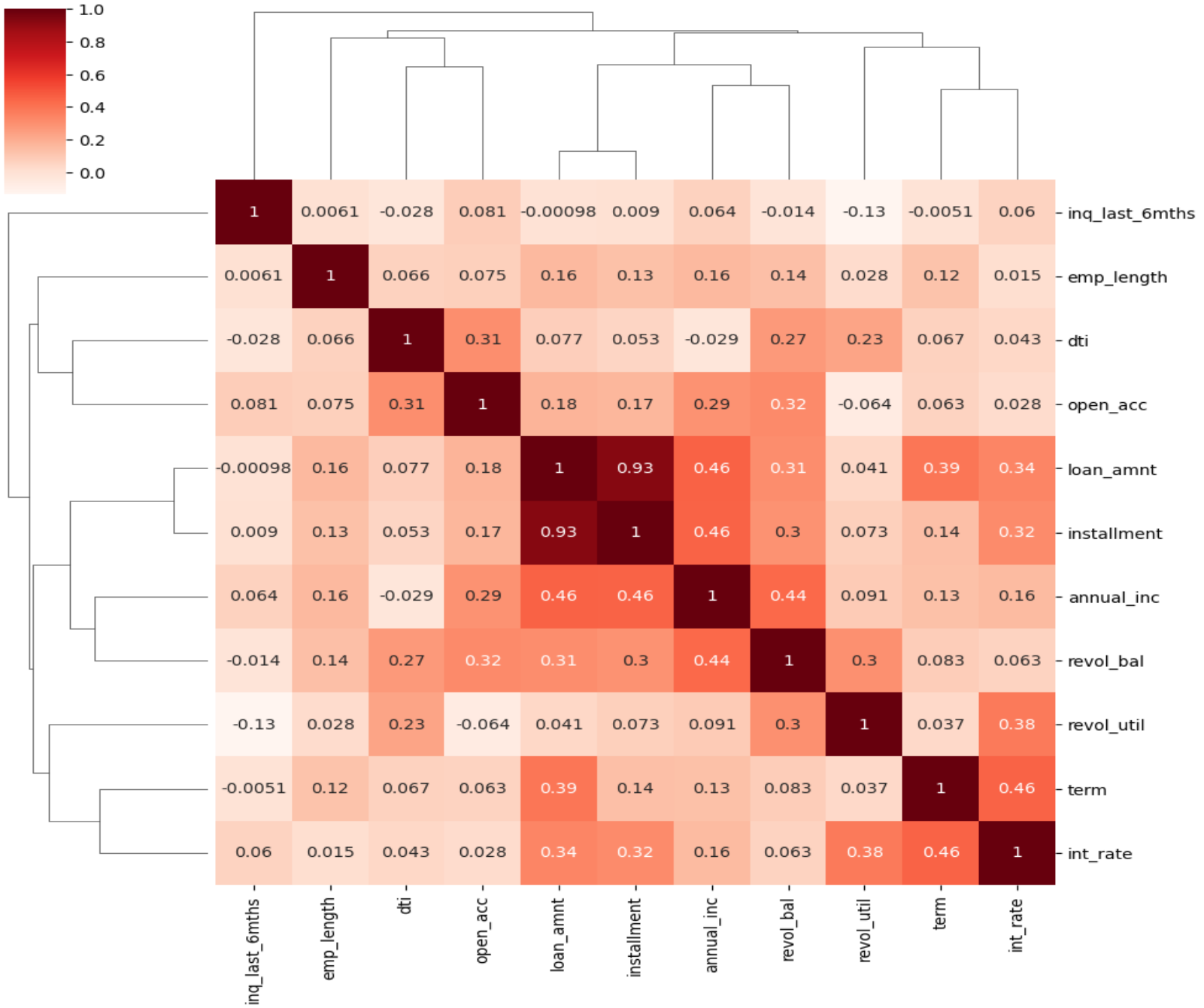
BIVARIATE ANALYSIS

This graph illustrates the relationship between interest rates and loan amounts, segmented by loan status (fully paid vs. Charged off). As loan amounts increase, interest rates rise for both fully paid and charged-off loans.
However, charged-off loans consistently exhibit higher interest rates across all loan amount categories. For example, for loans between 28,000 and 35,000, charged-off loans have an interest rate nearing 17%, while fully paid loans are closer to 15%.

This trend suggests that loans with higher interest rates are more likely to be charged off, indicating a higher default risk.

This heatmap illustrates the correlation between annual income categories and interest rate categories for charged-off loan takers.
Borrowers with lower incomes (21k-38k) tend to have higher interest rates (15-25%) and a higher likelihood of charged-off loans, while borrowers with higher incomes (89k+) generally have lower interest rates and fewer charged-off loans.





- The correlation matrix with hierarchical clustering highlights relationships between loan variables.

- Loan amount : strongly correlated with installments (0.93) and moderately with annual income (0.30) and interest rates (0.46).

- Installments : correlate similarly with income and interest, reflecting loan size and repayment structure.

- Loan term & interest rate : positively correlated (0.44), indicating longer loans have higher rates.

- Revolving utilization (revol_util) : correlates with interest rate (0.46) and loan amount (0.34).

- Debt-to-income ratio : mild correlations with installments and loan amount.

- Hierarchical clustering groups related variables, aiding risk assessment and lending strategies.

CONCLUSION.

Univariate Analysis:

- Higher default frequency is observed when:

1. Loan grade is "B“.
2. Loan status is "Not verified“.
3. Loan purpose is "Debt consolidation“.
4. Loan term is "36 months“.
5. Home ownership is "Rent“.
6. Loan amount is between 5,000 and 10,000.
7. Interest rate is 10-15%.
8. Annual income is 38,000-55,000.
9. Open credit accounts are 5-10.
10. 0 inquiries in the last 6 months.
11. Loan issued in December.
12. Applicant resides in California.

Segmented Analysis:

- Defaults are likely when:

1. Loan status is "Verified“.
2. Loan grade is “G”.
3. Loan term is "60 months”.
4. Loan purpose is "Small business”.

Correlation Analysis:

- Loan defaults positively correlate with:

1. Loan amount, installment, and annual income.
2. Loan term and interest rate, where longer loans carry higher rates, increasing default risk.

Bivariate/Multivariate Analysis:

- High default risk is seen when:

1. Interest rates increase by 2%.
2. Loan purpose is "Small business" with amounts over 14,000.
3. Homeownership is "Mortgage" with loan amounts above 13,000.
4. Loan grade is "G" with interest rates above 20%.
5. Interest rates of 20-25% with incomes between 21,000-38,000.

Final Recommendations:

Based on the analysis, it is recommended that the finance company:

1. **Adjust lending criteria** for applicants requesting loans with high-risk characteristics, particularly those with higher interest rates, longer loan terms, and lower loan grades.

2. **Increase scrutiny for specific loan purposes**, such as small businesses and debt consolidation, as these have shown higher default rates.

3. **Fine-tune interest rate policies** to reflect the risk profiles of different loan grades and purposes.

4. **Implement dynamic credit risk models** that account for multivariate interactions between factors such as loan amount, income, and interest rate, allowing for more precise risk predictions.

By incorporating these recommendations, the finance company can effectively mitigate credit losses, improve profitability, and make more informed, data-driven lending decisions.

THANK YOU

NIKHIL LUNAWAT - Facilitator
nikhilonly12@gmail.com

NIKHIL UGALE
nikhilugale6955@gmail.com