LENDING CLUB CASE STUDY

CONTRIBUTORS:

RAMYA M
RAJASHEKAR P

Business Understanding:

You work for a consumer finance company which specializes 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. Two types of risks are associated with the bank’s decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

Objective:

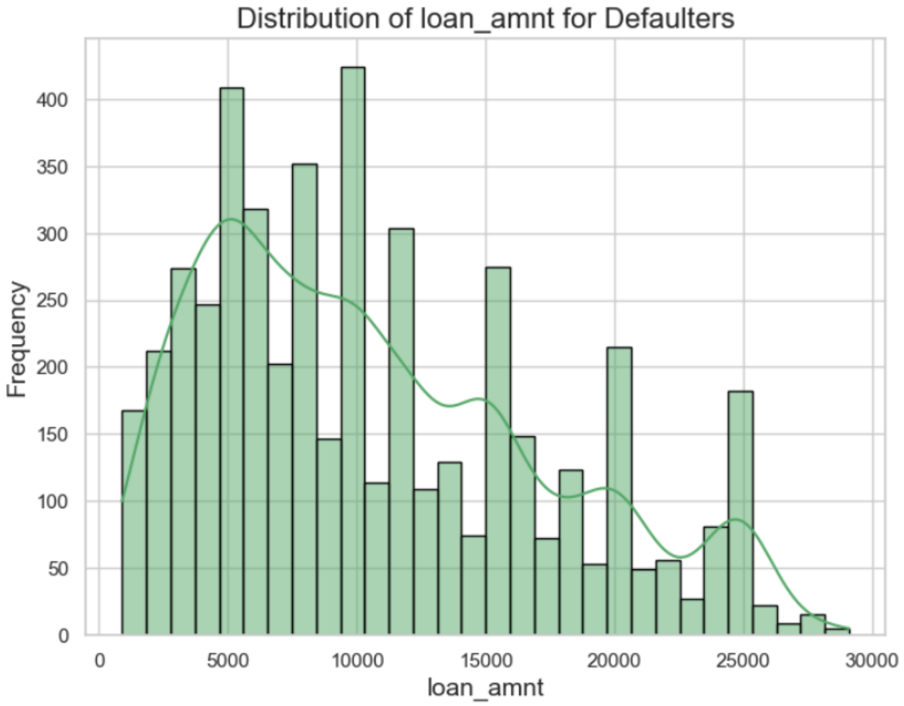
Identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

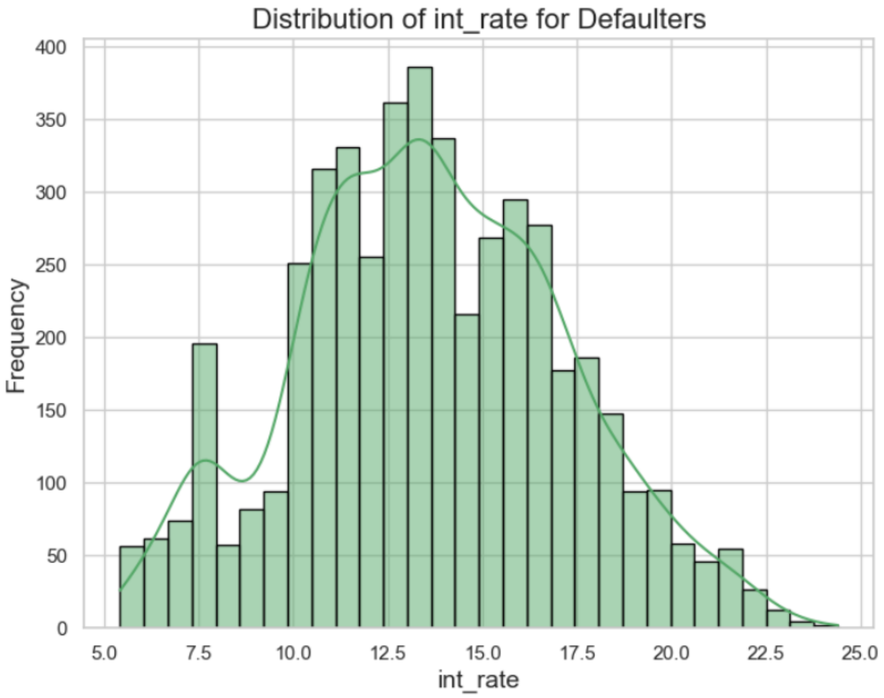
Analysis Approach:

- **Understand and Clean the Dataset:**
  - **Review Data Dictionary:** Analyze column names and descriptions to understand data structure.
  - **Data Quality Checks:**
    - **Null Values:** Identify and address missing values in columns and rows.
    - **Outliers:** Detect and handle outliers to ensure data integrity.
    - **Duplicates:** Identify and remove duplicate entries.
    - **Categorical Data Validation:** Verify accuracy and consistency in categorical columns.
  - **Select relevant columns:** Identify and choose relevant columns for focused analysis
- **Perform EDA Analysis by Univariate, Bivariate and Multivariate:**
    - **Numerical Columns:** Analyze distributions, correlations, and patterns.
    - **Categorical Columns:** Assess frequency distributions, relationships, and trends.

Data Analysis : Univariate and Segmented Univariate

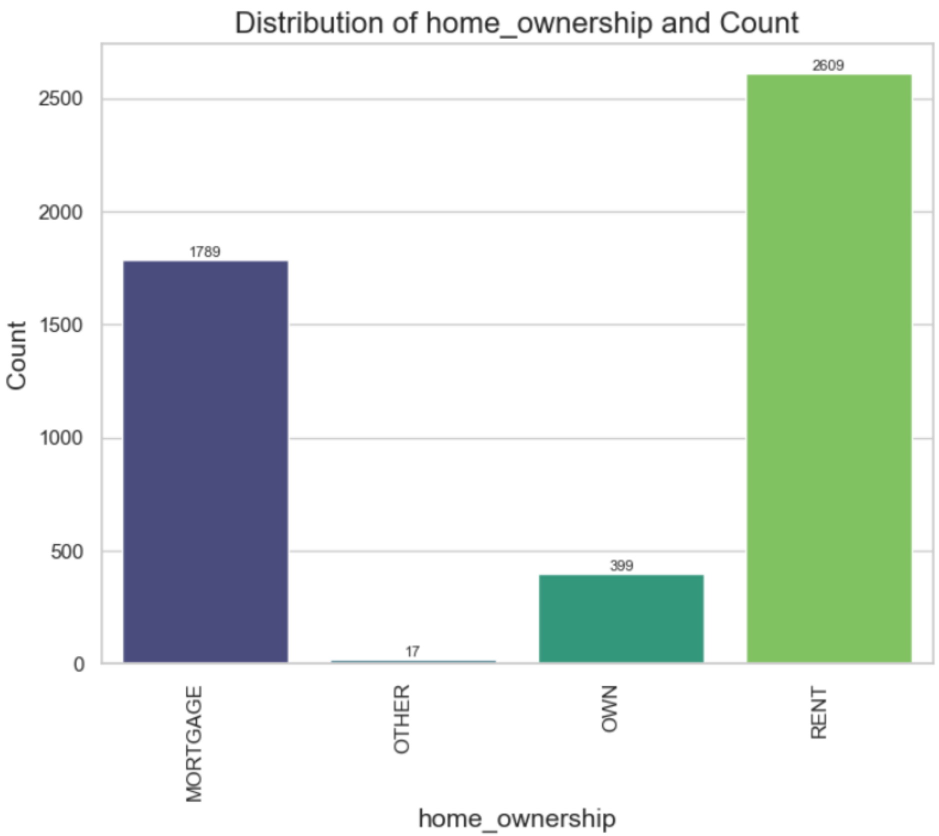
Loan Amounts: More defaulters are observed with loan amounts between 3k and 12k.
Interest Rates: Defaulters are more common with interest rates between 10% and 18%.

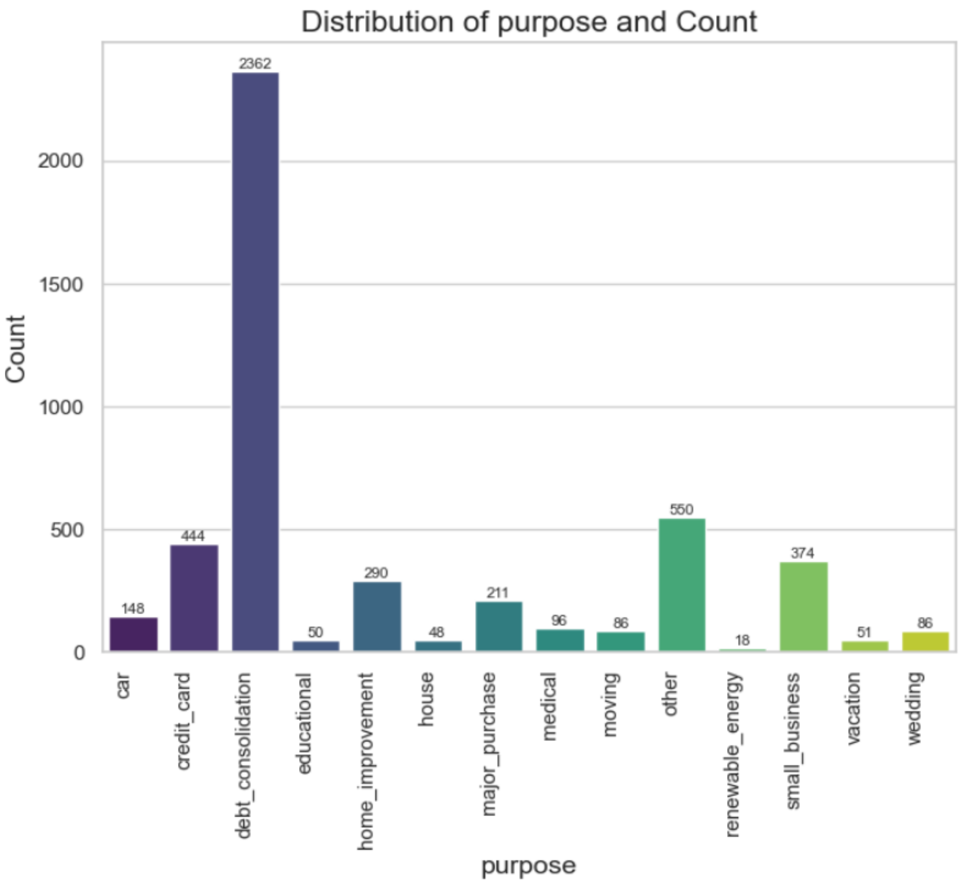




Data Analysis : Univariate and Segmented Univariate (contd.)

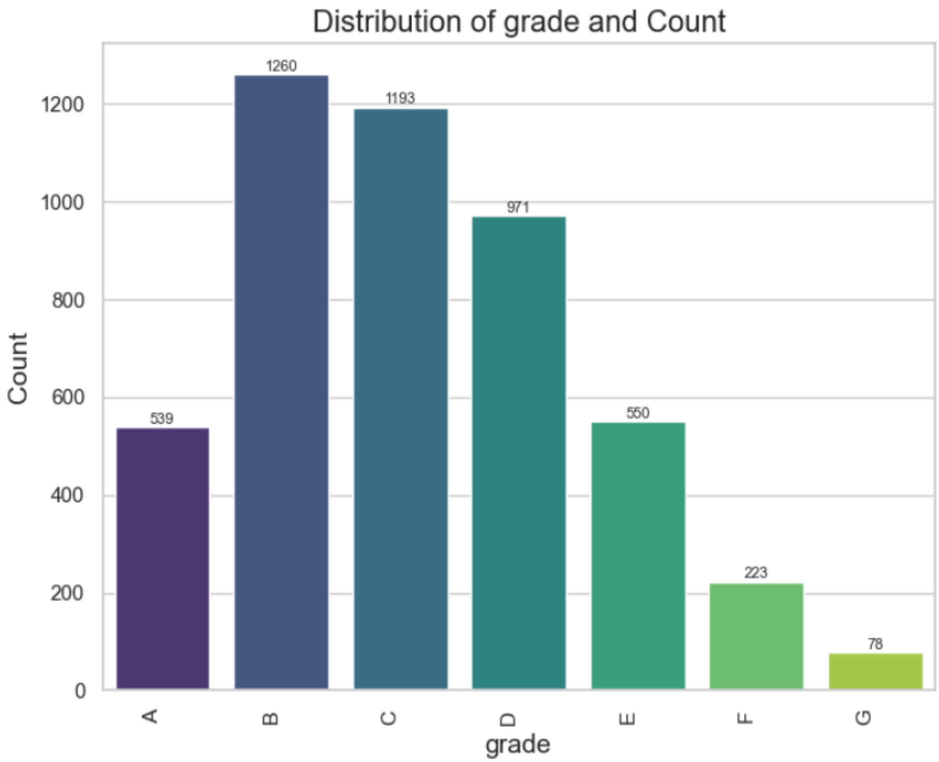
Homeownership: Defaulters are ranked as RENT > MORTGAGE > OWN > OTHER.
Purpose: Debt consolidation is the most common purpose among defaulters.

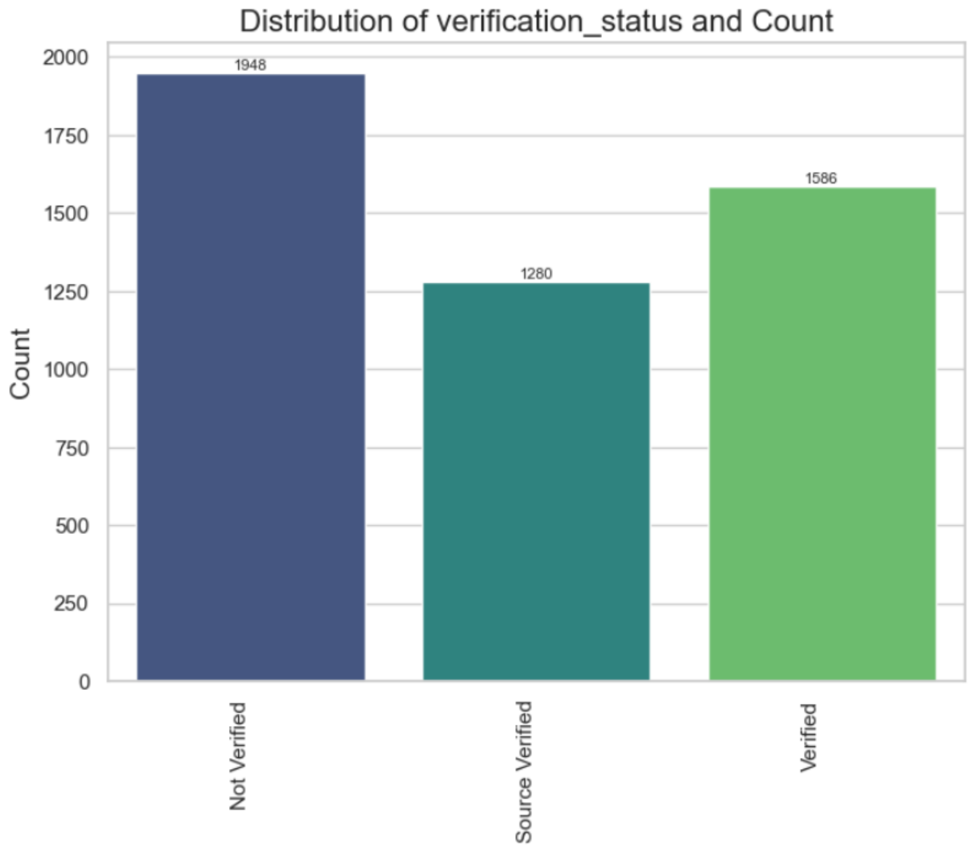




Data Analysis : Univariate Analysis and Segmented Univariate (contd.)

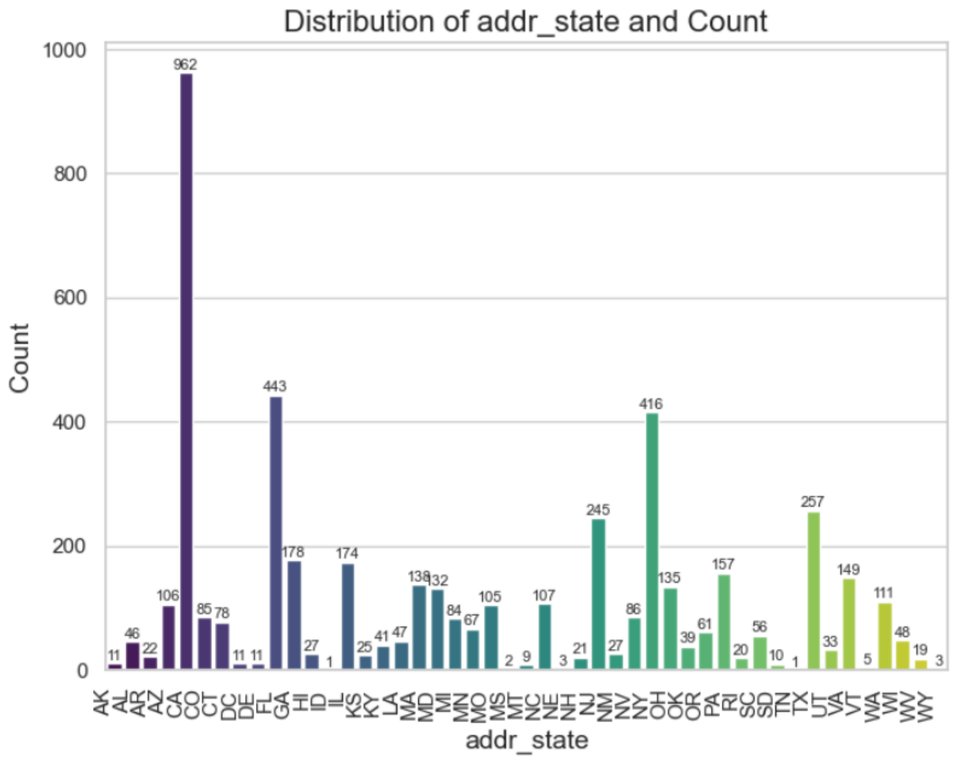
Grade: Defaulters are ranked as B > C > D > E > A > F > G.
Verification Status: Defaulters are ranked as Not Verified > Verified > Source Verified.

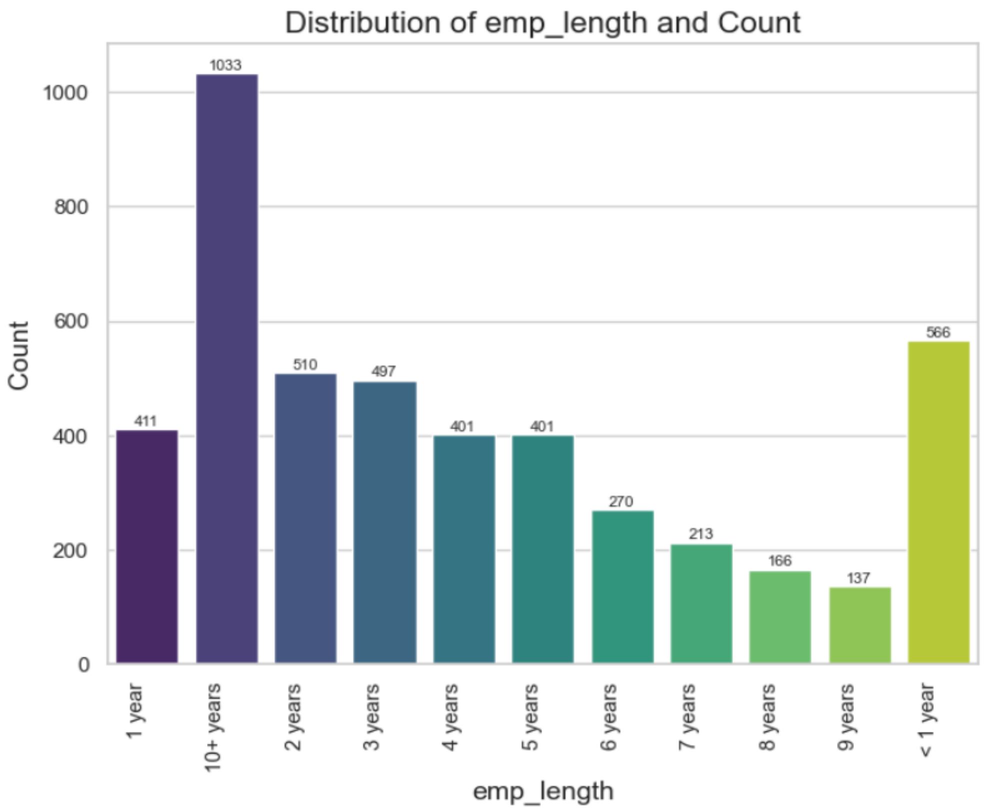




Data Analysis : Univariate and Segmented Univariate (contd.)

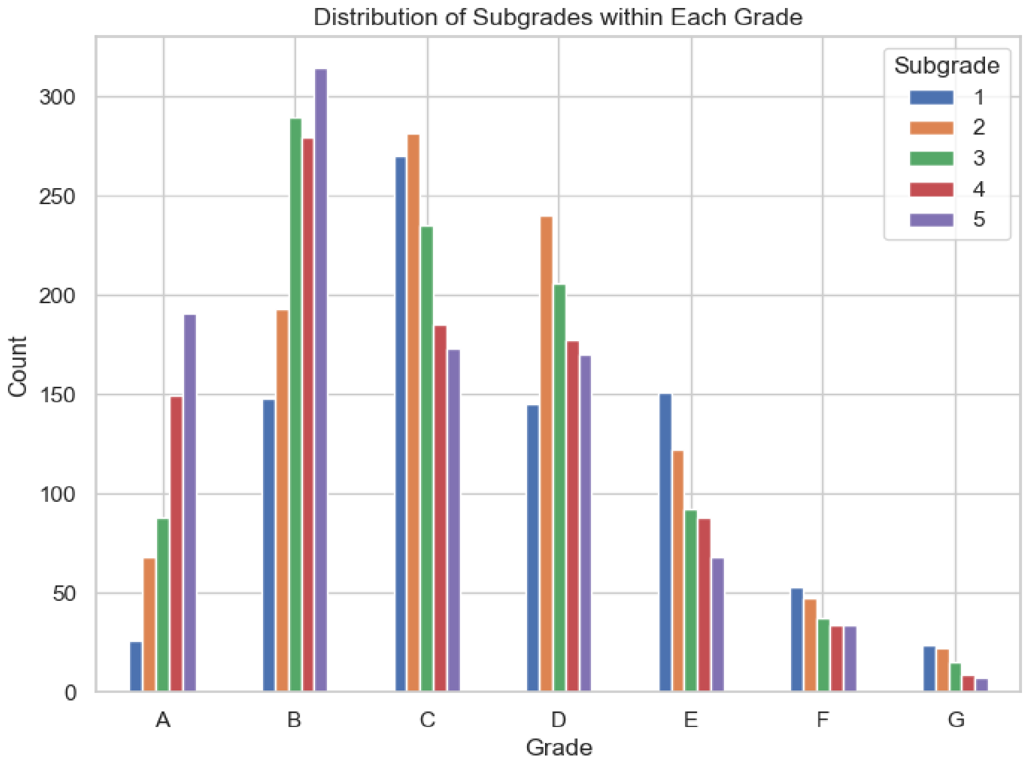
Address State: CA has the highest number of defaulters, followed by FL, NY, TX, and NJ.
Employment Length: Defaulters with 10+ years of employment are notably high.

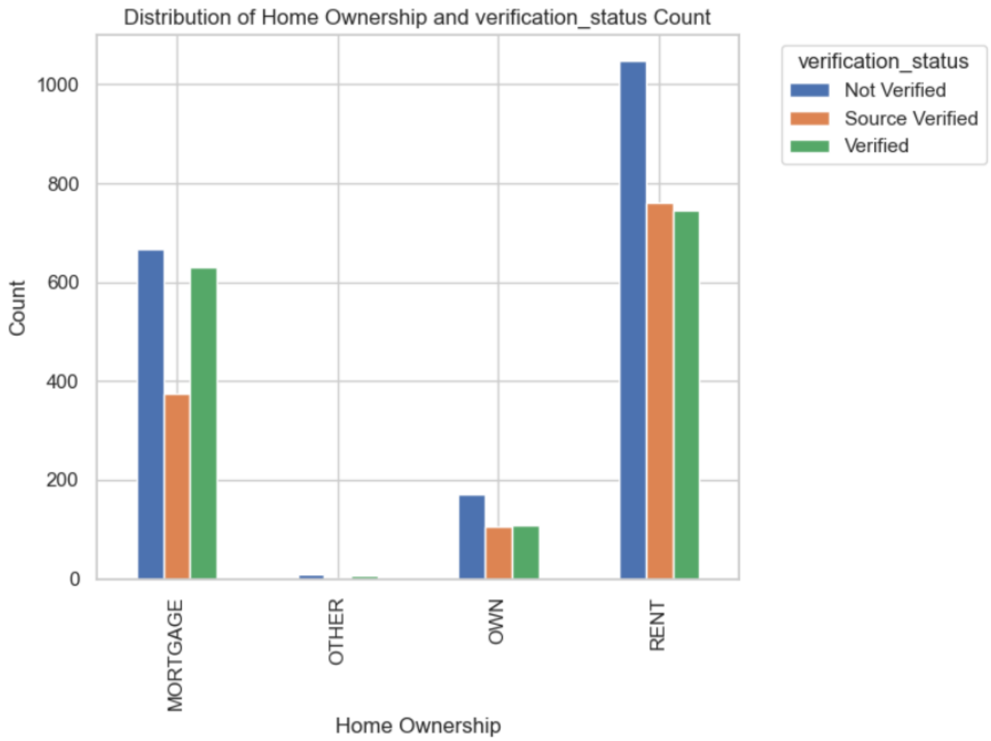




Data Analysis : Bivariate (contd.)

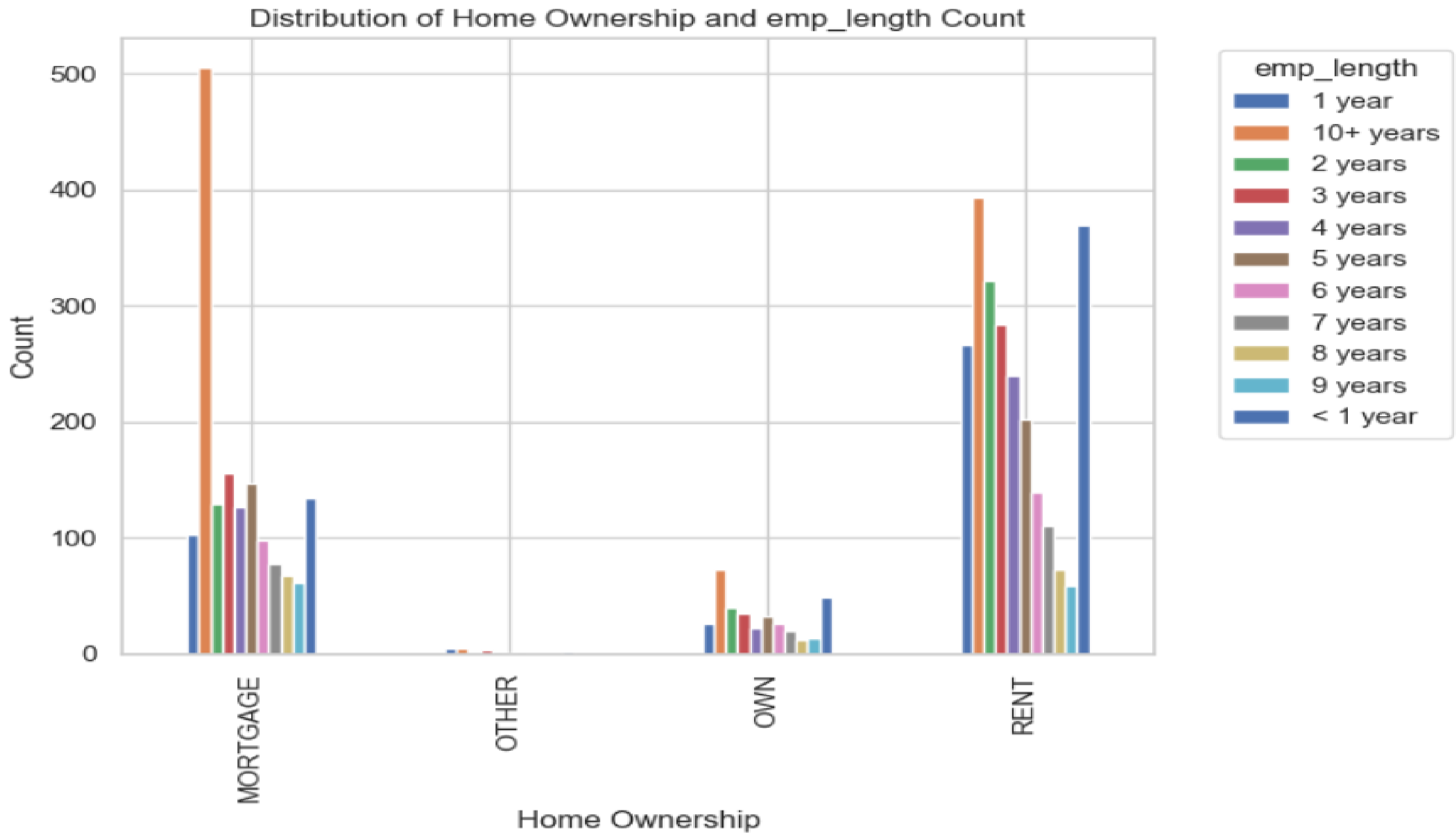
Grades and Sub-Grades: Defaulters are predominantly in grades B and C. Subgrades B3, B4, B5, C1, C2, and C3 have the highest number of defaulters.
Home Ownership and Verification Status: Defaulters are primarily those who rent and do not have verified status.





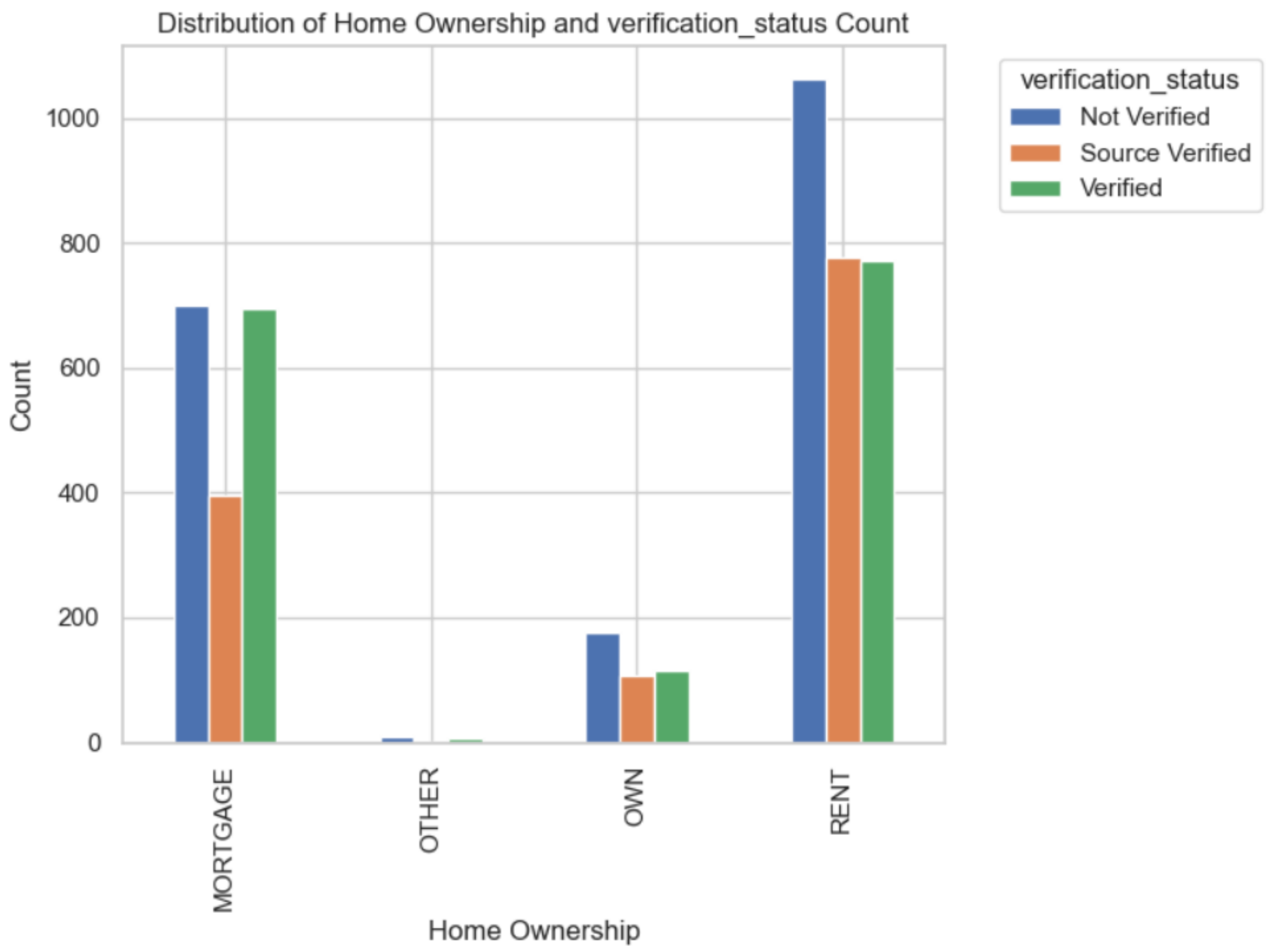
Data Analysis : Bivariate (contd.)

Employment Length and Ownership: Defaulters with 10+ years of employment are most common across all homeownership types. For RENT homeownership, the second highest category is "< 1 year" and the third is "2 years," whereas for MORTGAGE homeownership, "3 years" ranks second and "5 years" third.



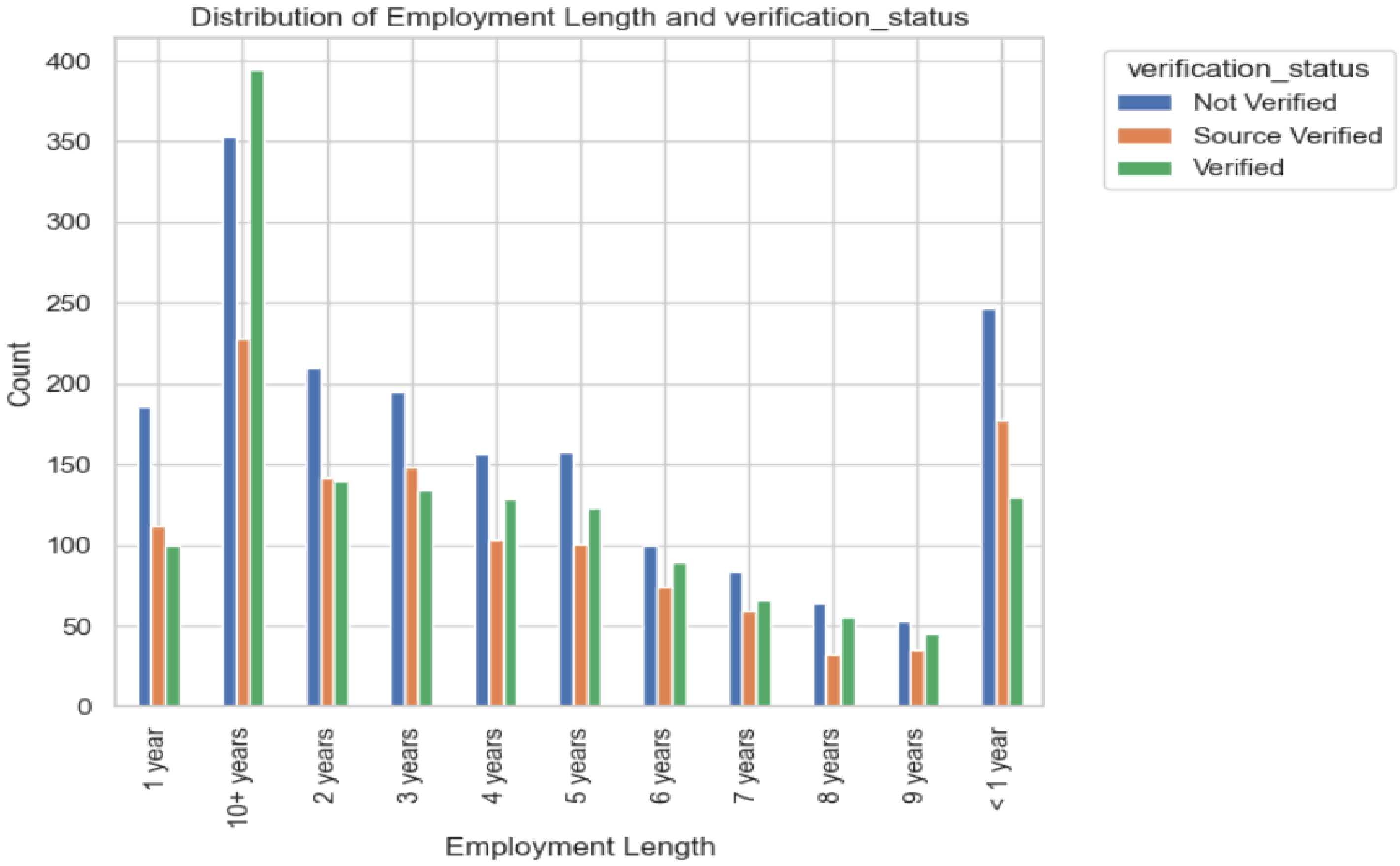
Data Analysis : Bivariate (contd.)

Verification Status and Ownership: Across all homeownership types, the "Not Verified" status is most common among defaulters. For RENT homeownership, defaulters are equally distributed between "Source Verified" and "Verified" statuses.



Data Analysis : Bivariate (contd.)

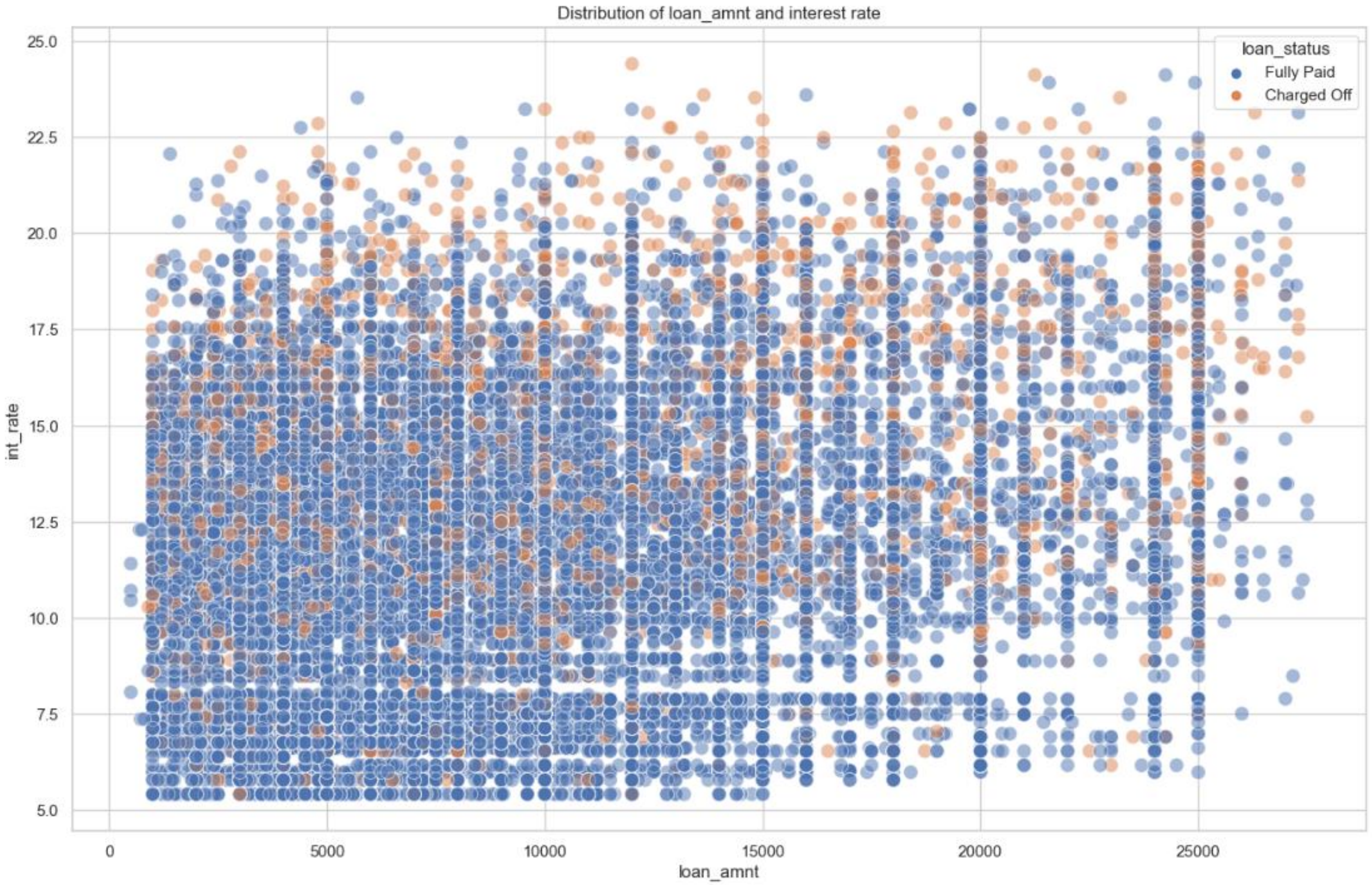
Verification Status and Employment Length: Across all employment lengths, the "Not Verified" status is more prevalent. However, for employees with 10+ years of experience, the "Verified" status is more common.



Data Analysis : Bivariate (contd.)

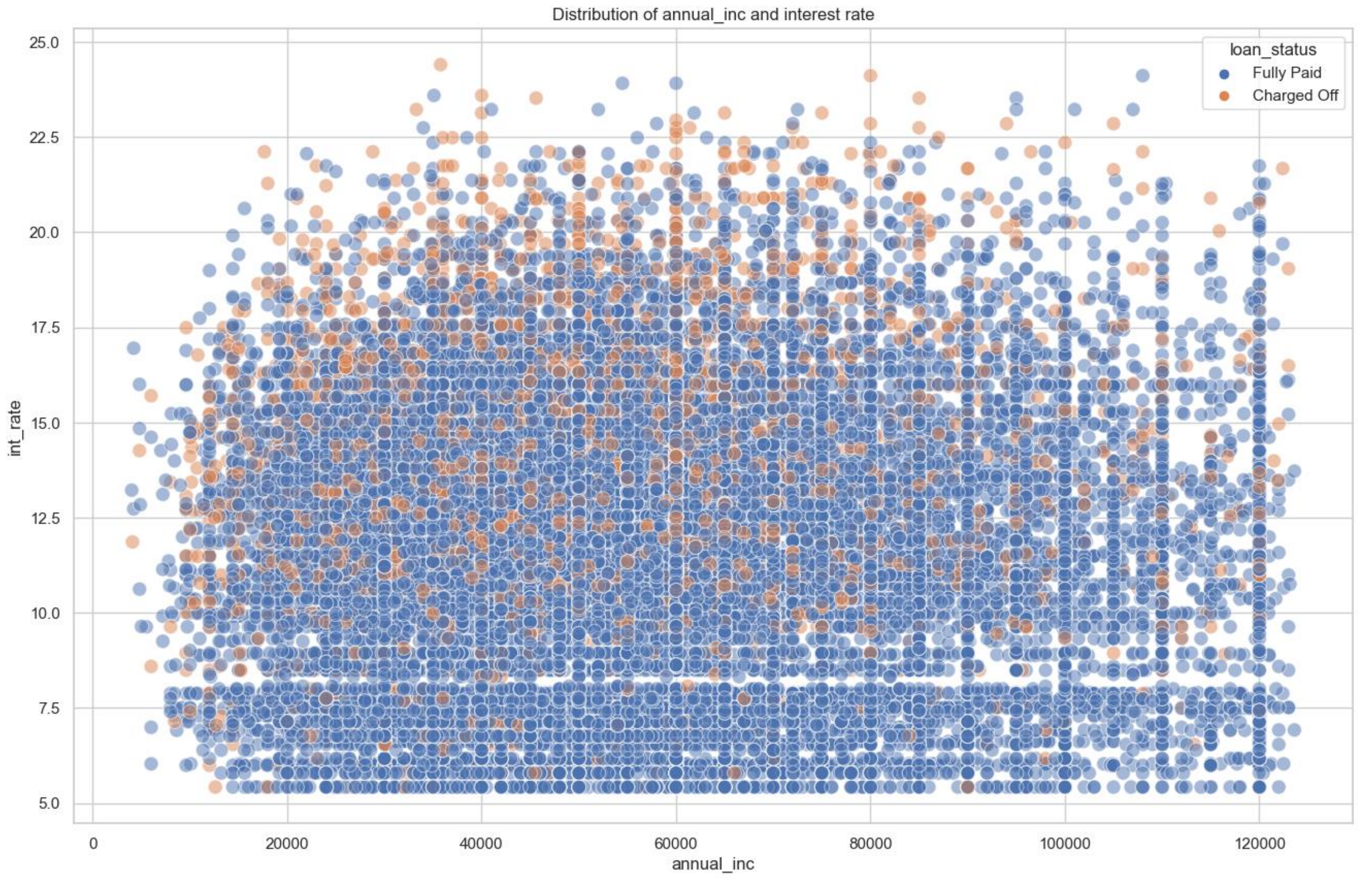
Interest Rate and Loan Amount

- The number of defaulters rises with increasing interest rates, regardless of annual income or loan amount.
- Defaulters tend to increase significantly when interest rates are 12.5% or higher irrespective of loan_amnt



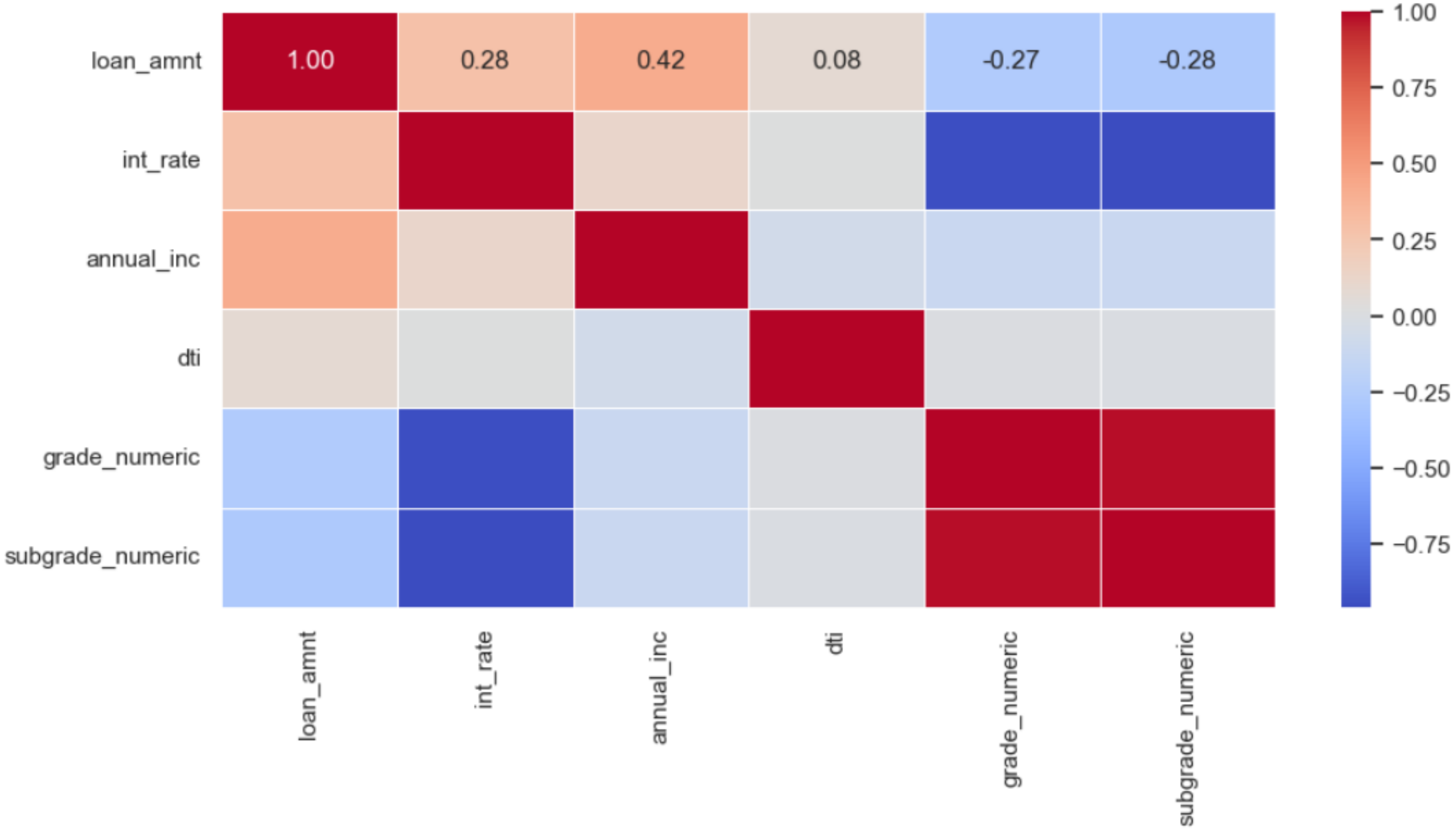
Data Analysis : Bivariate (contd.)

Interest Rate and Annual income: Applicants with lower annual incomes are more likely to default compared to those with higher annual incomes.



Data Analysis : Multivariate

- Lower loan amounts correlate with better grades among defaulters.
- A moderate positive correlation (0.37) exists between loan amount and annual income, indicating higher incomes are linked to larger loans for defaulters.
- Higher loan grades are closely associated with lower interest rates among defaulters.



Conclusion based on Risk factors

- **Loan Amount:** Defaulters are more common with loan amounts between $3,000 and $12,000.
  Lower loan amounts are generally associated with better loan grades, indicating less risk.

- **Annual Income:** Defaulters typically have annual incomes between $2,000 and $8,000.
  Lower annual incomes are linked to a higher likelihood of default. Income is moderately correlated with loan amounts, suggesting that higher incomes might lead to larger,
  less risky loans.

- **Debt-to-Income (DTI) Ratio:** Higher DTI values are associated with an increased number of defaulters.
  Applicants with higher DTI ratios pose a higher risk.

- **Interest Rates**: Defaulters are more prevalent with interest rates between 10% and 18%.
  Higher interest rates correlate strongly with higher grades and default risk.

- **Homeownership**: The risk of default is highest among renters (RENT), followed by those with a mortgage (MORTGAGE), and lowest among owners (OWN).
  Homeownership status is a strong indicator of default risk.

Conclusion based on Risk factors (contd.)

- **Verification Status**: Defaulters are most common among applicants who are "Not Verified." Verification status is critical; "Not Verified" applicants present a higher risk.
- **Purpose of Loan**: Debt consolidation is the most common purpose among defaulters. Loans for debt consolidation should be scrutinized more rigorously.
- **Employment Length**: Defaulters with over 10 years of employment are notably high. Defaulters are also significant in shorter employment durations. Employment length should be considered along with other factors.
- **Loan Grades**: Defaulters are predominantly in grades B and C. Specific subgrades like B3, B4, B5, C1, C2, and C3 show the highest default rates.
- **Term of Loan**: A higher number of defaulters are associated with 36-month terms compared to 60-month terms. Shorter-term loans (36 months) have a higher risk of default.
- **Geographic Location**: States like California (CA), Florida (FL), New York (NY), Texas (TX), and New Jersey (NJ) have the highest number of defaulters.

Recommendations

Enhanced Verification:
Implement stricter verification processes to reduce the risk associated with "Not Verified" applicants.

Targeted Risk Assessment:
Focus on mid-range loan amounts, lower annual incomes, higher DTI ratios, and higher interest rates as key risk factors.

Geographic and Purpose-Based Policies:
Develop policies that account for geographic risk and the purpose of loans, especially for debt consolidation.

Holistic Evaluation:
Combine employment length, loan grade, and term with other factors for a comprehensive risk evaluation.
By leveraging these insights, Lending Club can make more informed decisions, reducing the risk of defaults and enhancing overall loan portfolio quality.

Thank You