LENDING CLUB CASE STUDY

Problem Statement

- In the context of consumer finance, effectively managing credit risk is crucial for minimizing financial losses. The task is to analyze historical loan application data to identify patterns and factors associated with loan defaults.

- The goal of this analysis is to apply Exploratory Data Analysis (EDA) techniques to uncover key indicators of loan defaults. By identifying these risk factors, the company aims to enhance its decision-making process, reduce credit loss, and optimize its loan approval criteria.

Objective

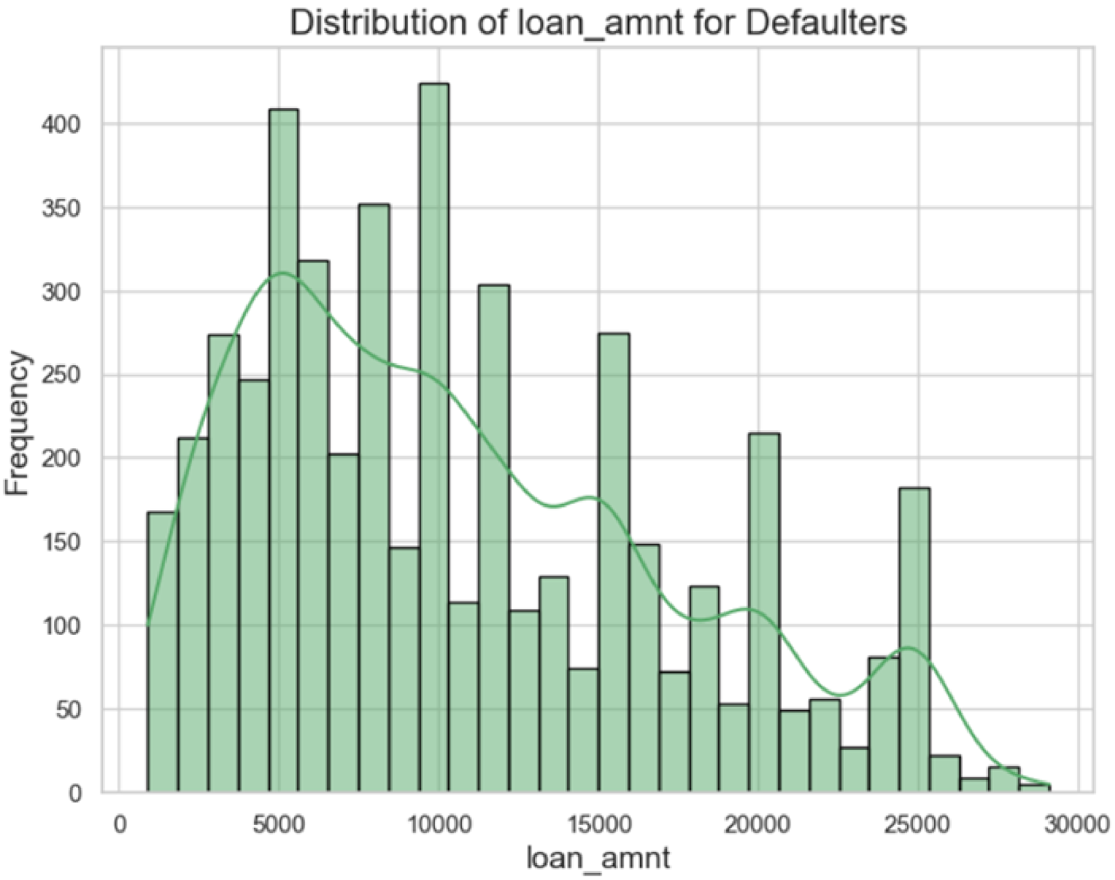
- Determine which attributes are significant predictors of loan default.
- Analyze consumer and loan attributes to uncover patterns and trends that indicate higher risk of default.
- Develop insights into the characteristics of high-risk loan applicants.
- Provide actionable recommendations to improve the company’s risk assessment strategies and reduce the likelihood of default.

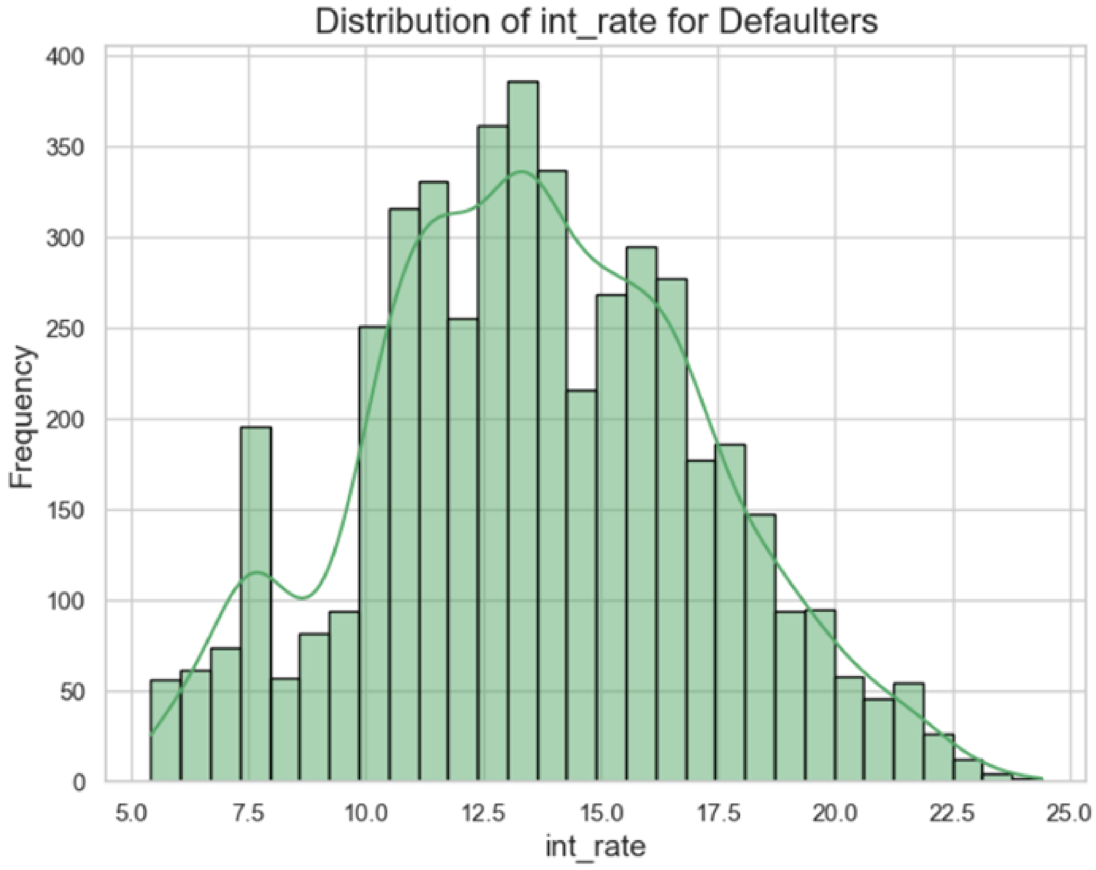
Analysis Approach

- **Data Understanding and Cleaning**
  - Review Data Dictionary: Examine column names and their descriptions to comprehend the data structure.
  - Data Quality Assessment: Handling Missing Values, Outlier Detection, Duplicate Records and Categorical Data Consistency
  - Column Selection: Identify and select the most relevant columns for a focused analysis.
- **Exploratory Data Analysis (EDA)**
  - Univariate Analysis: Numerical and categorical variable analysis
  - Bivariate Analysis: Explore the relationships and correlations between pairs of variables.
  - Multivariate Analysis: Investigate interactions among multiple variables

Univariate and Segmented Univariate

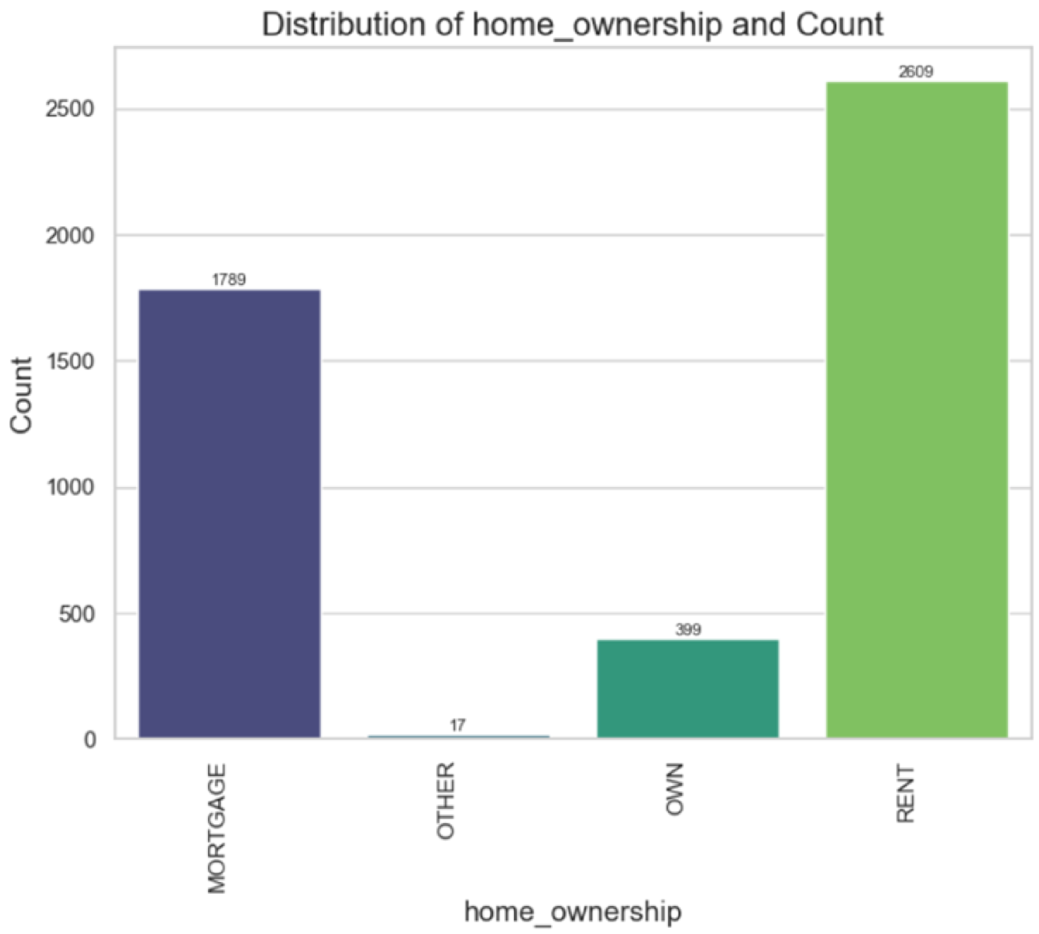
- Loan Amounts: Higher default rates are observed for loan amounts ranging from $3,000 to $12,000.
- Interest Rates: Default rates are elevated for interest rates between 10% and 18%.

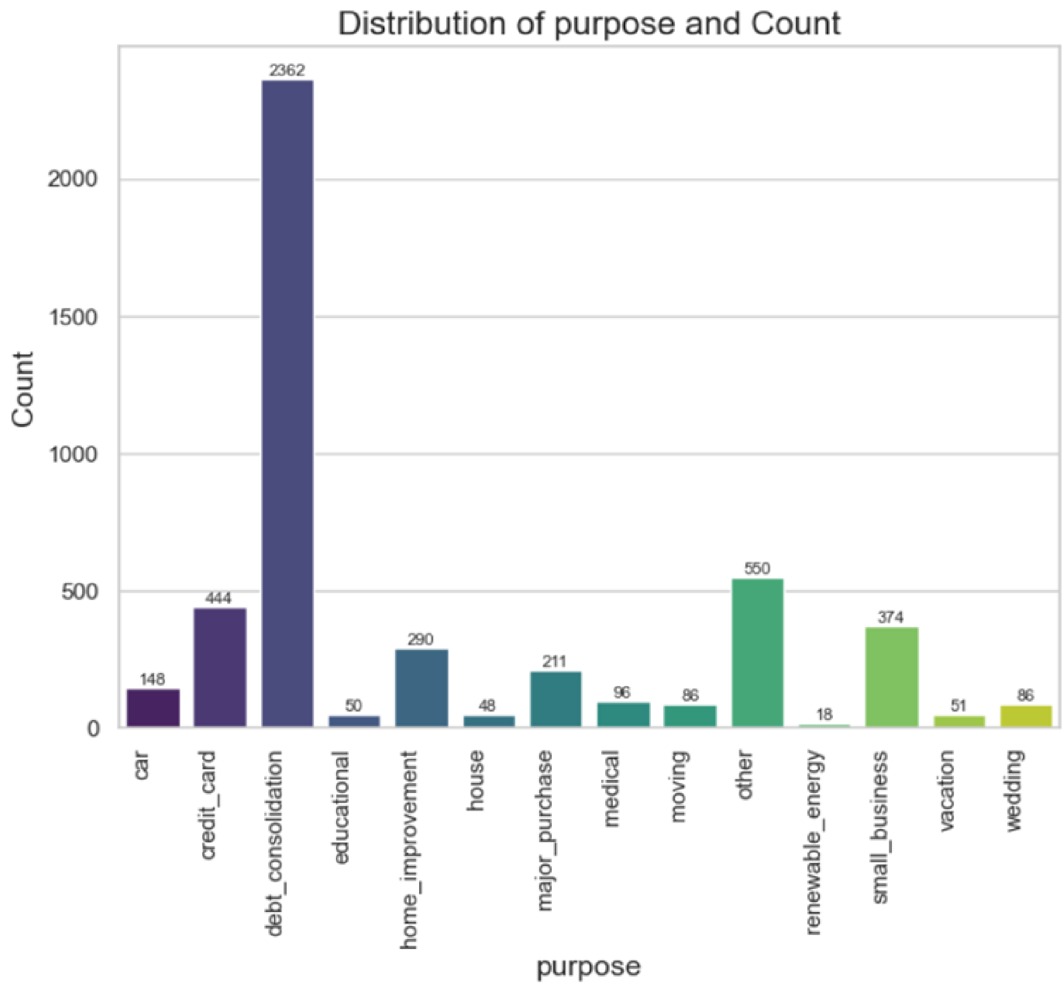




Univariate and Segmented Univariate(contd.)

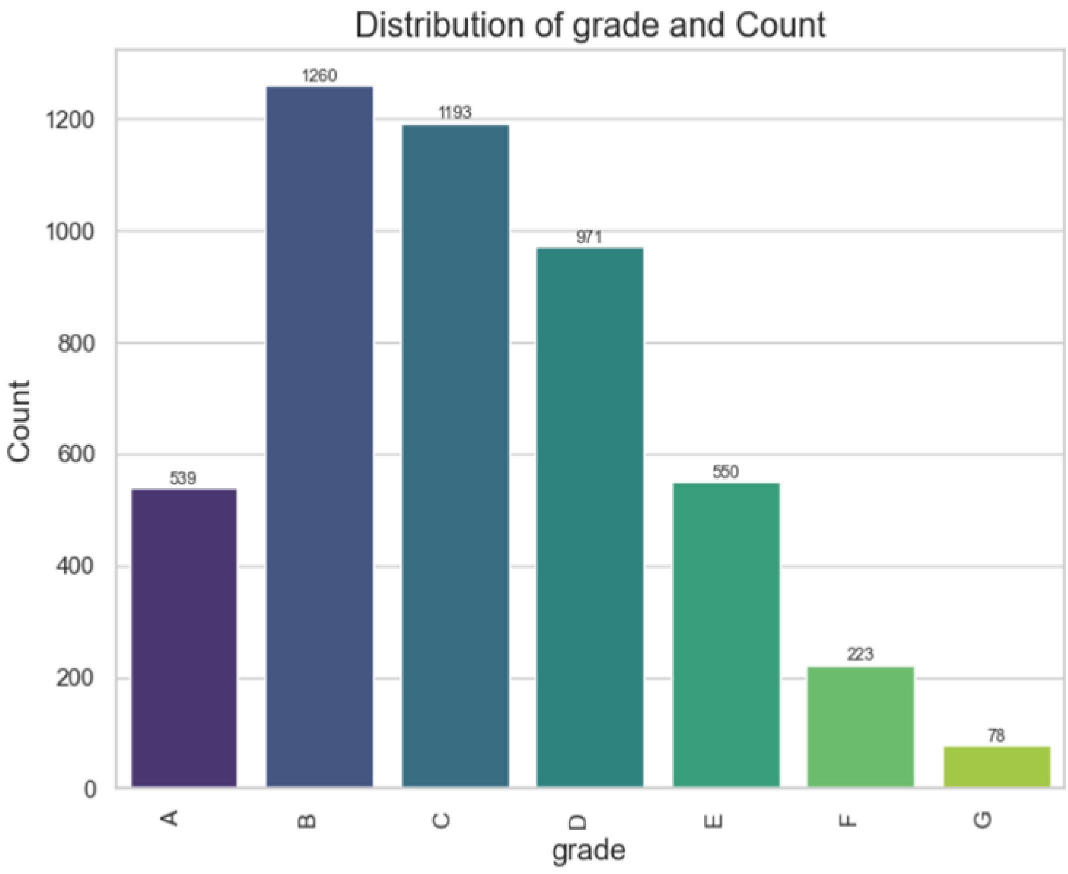
- Homeownership Status: Ranking of Default Rates as RENT > MORTGAGE > OWN > OTHER
- Loan Purpose: Debt consolidation is the leading reason among defaulters

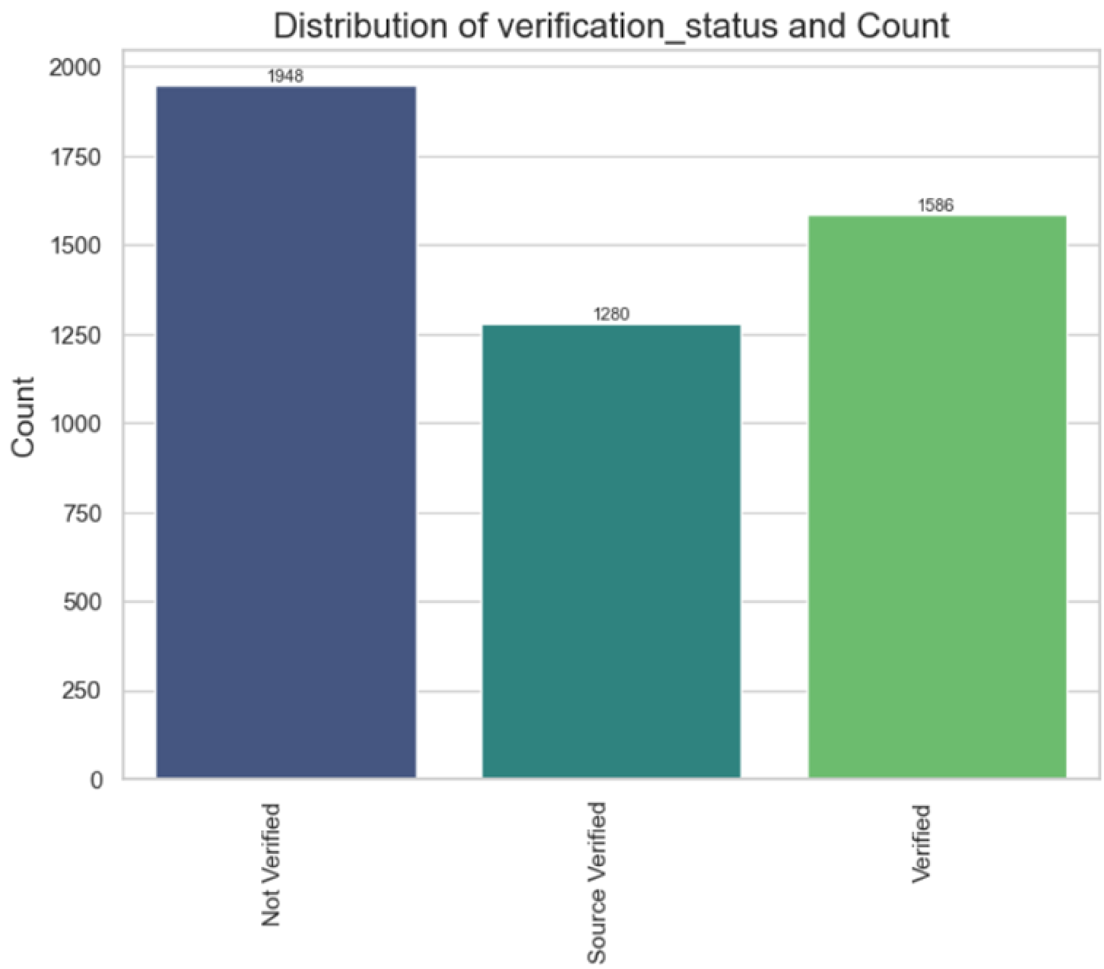




Univariate and Segmented Univariate(contd.)

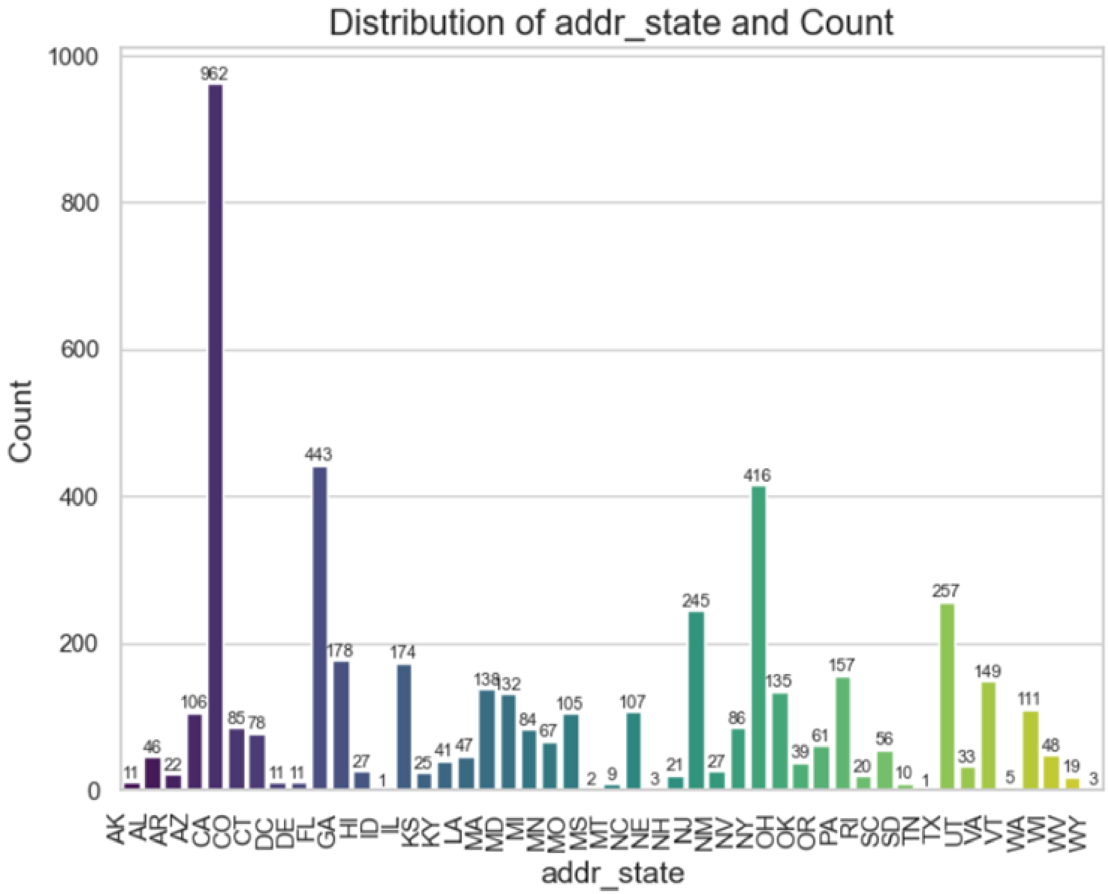
- Grade: Ranking of Default Rates as B > C > D > E > A > F > G
- Verification Status: Ranking of Default Rates as Not Verified > Verified > Source Verified

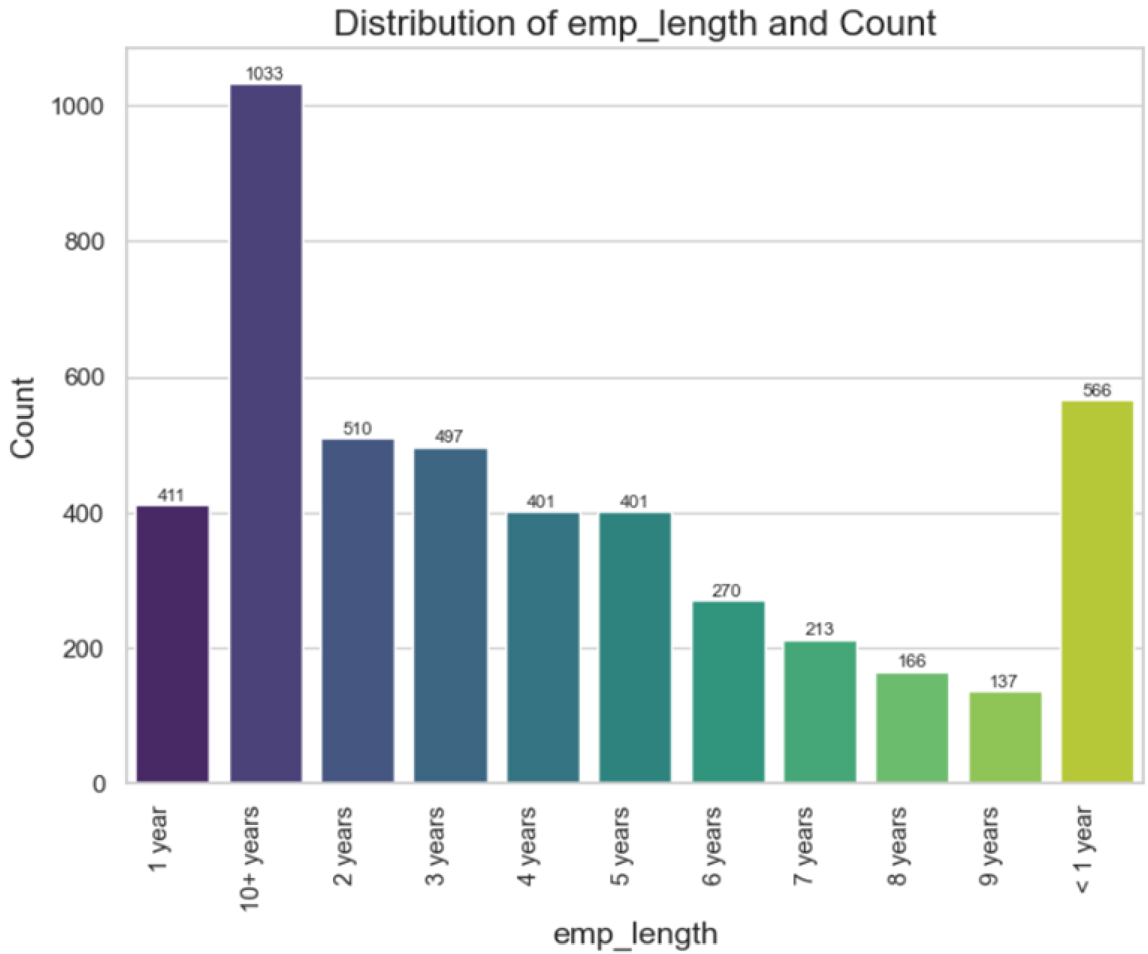




Univariate and Segmented Univariate(contd.)

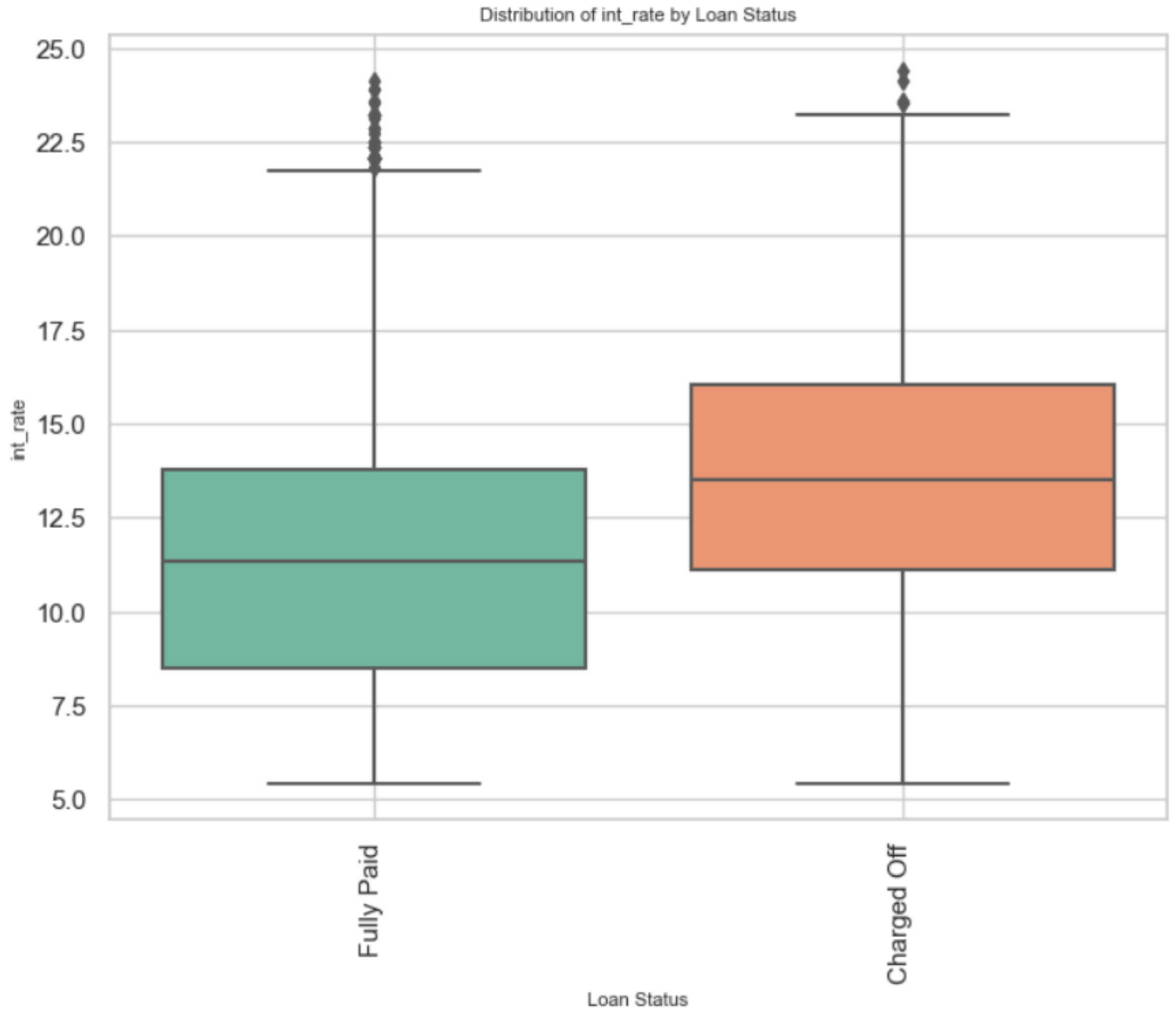
- Address State: Top States by Default Rate as CA > FL > NY > TX > NJ
- Employment Length: A significant proportion of defaulters have over 10 years of employment.





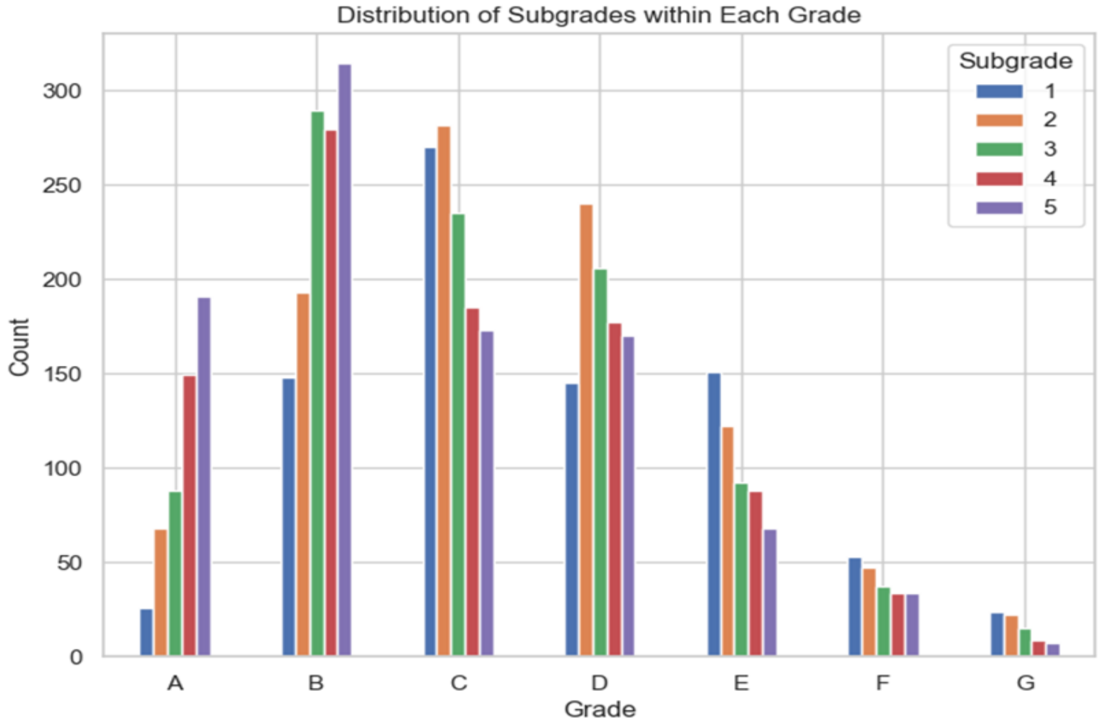
Univariate and Segmented Univariate(contd.)

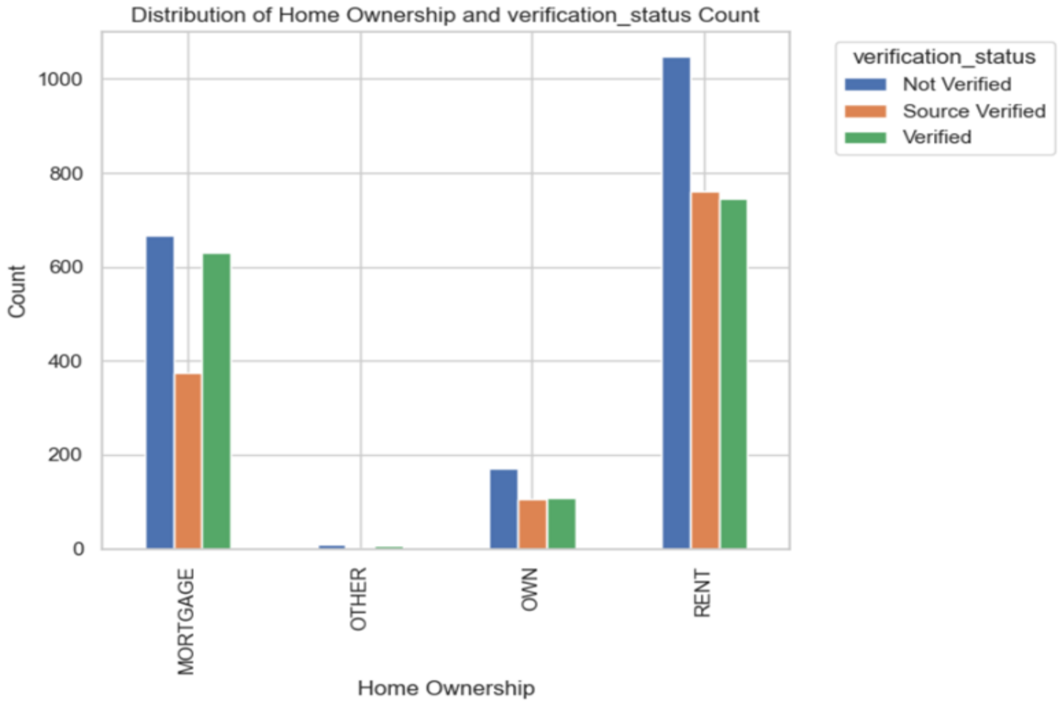
Plot of the spread of Interest Rates for fully paid and defaulters. Interest rates for charged off applicants is bit higher compared to fully paid applicants.



Bivariate Analysis

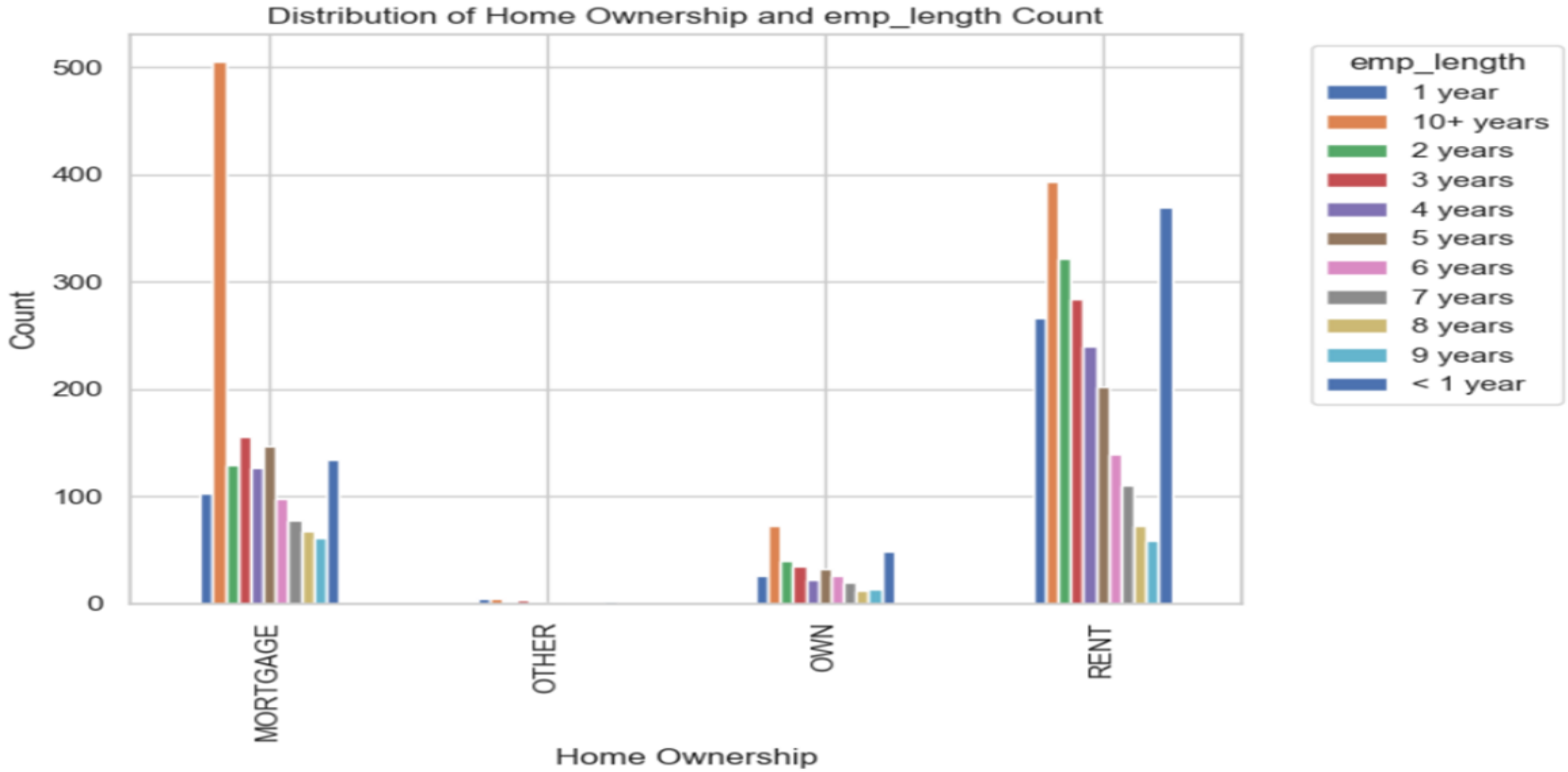
- Grades and Sub-Grades
  - **Overall Trend:** Defaulters are predominantly in grades B and C.
  - **High Default Sub-Grades:**
    - **Grade B:** Subgrades B3, B4, B5
    - **Grade C:** Subgrades C1, C2, C3
- Home Ownership and Verification Status
  - Defaulters are mainly renters.
  - High incidence of defaults among those with "Not Verified" status.





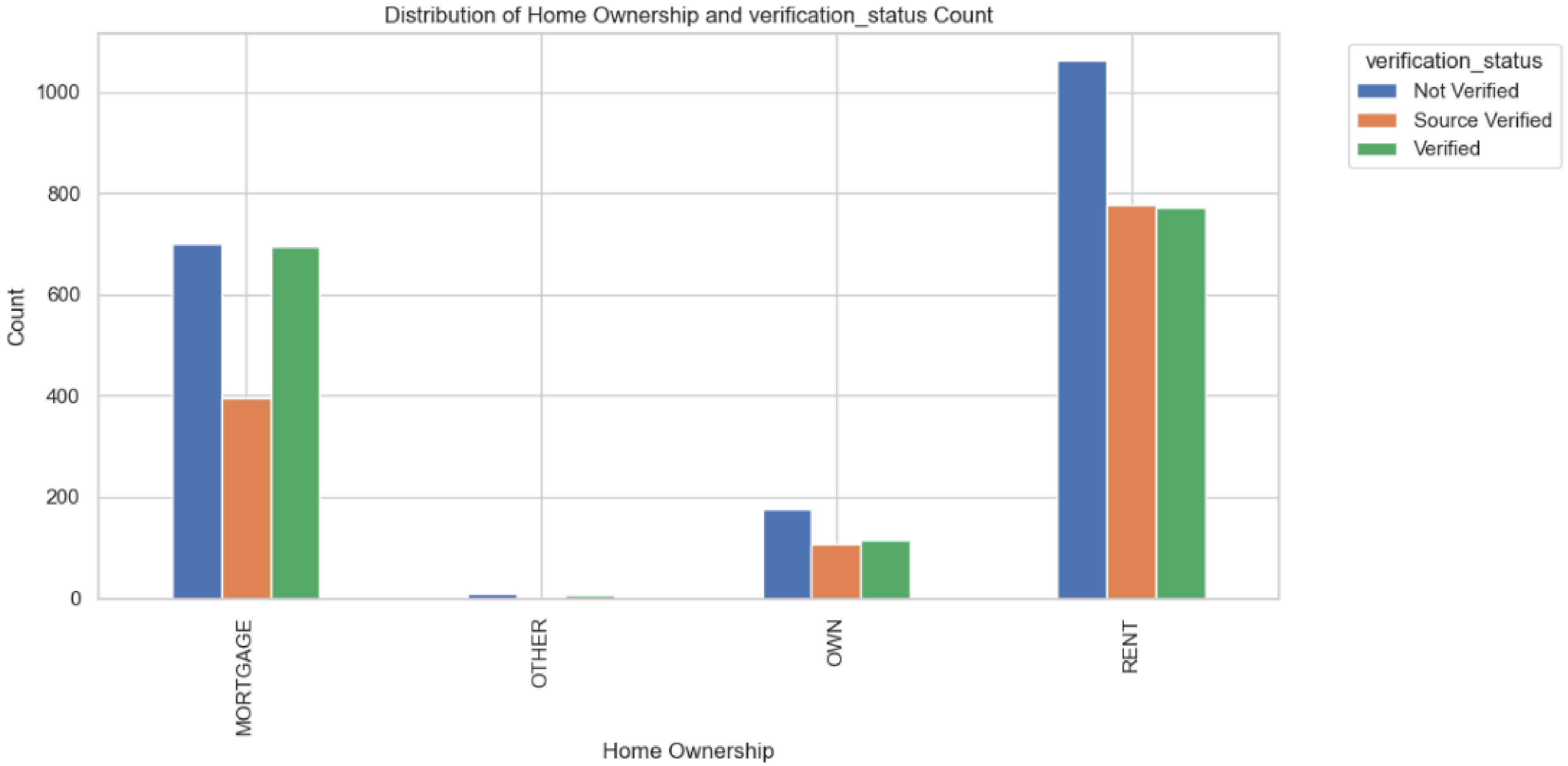
Bivariate Analysis(contd.)

- Employment Patterns:
  - General Trend**:** Defaulters with over 10 years of employment are prevalent across all homeownership types.
  - By Homeownership Type:
    - RENT: Second highest is "< 1 year" and Third highest is "2 years"
    - MORTGAGE: Second highest is "< 1 year" and Third highest is "2 years"



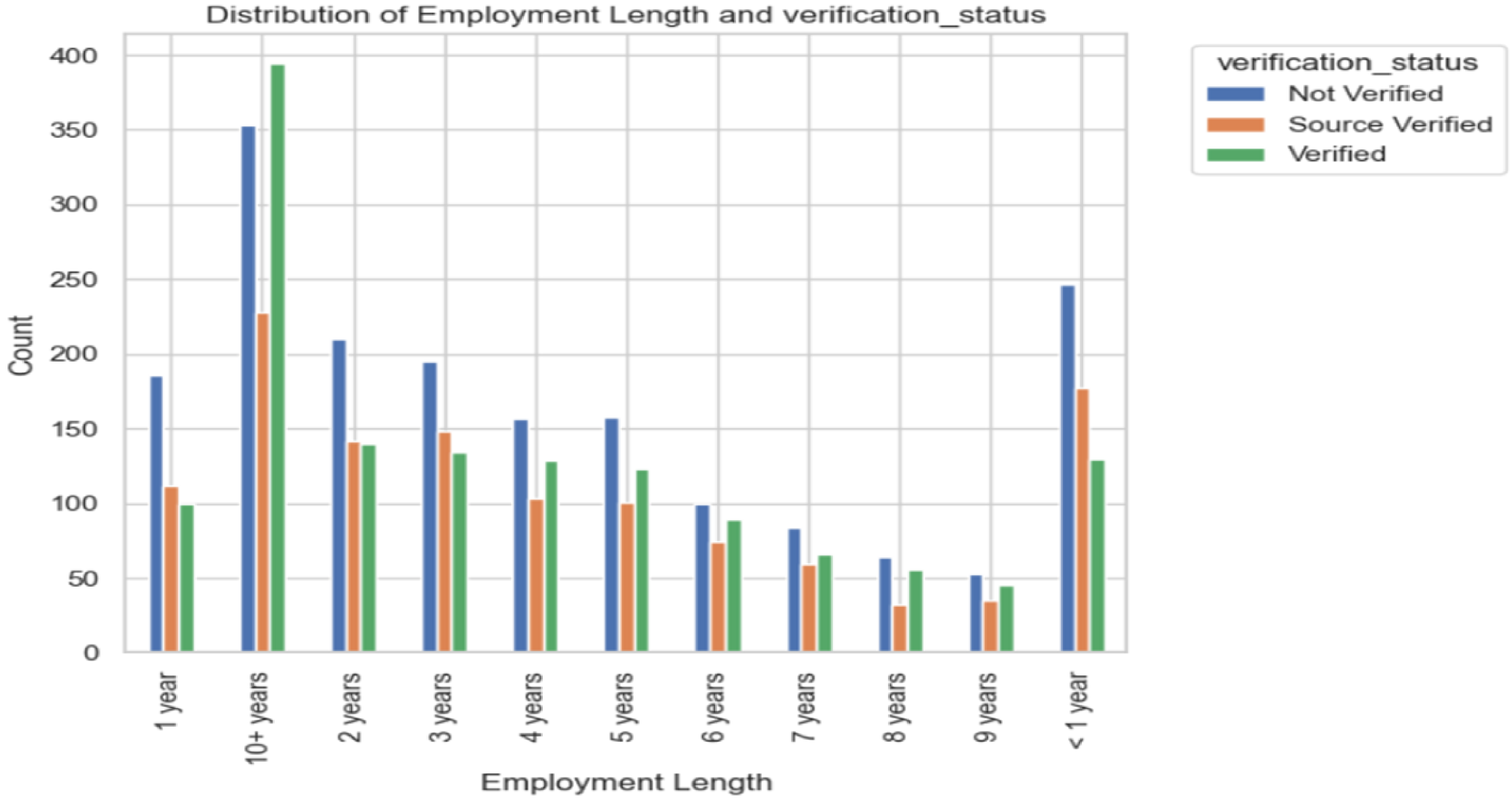
Bivariate Analysis(contd.)

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



Bivariate Analysis(contd.)

- Verification Status and Employee Length:
  - Verification Trends:
    - Across all employee lengths, "Not Verified" status is most common among defaulters.
    - For those with over 10+ years of employment, the "Verified" status is more prevalent among defaulters.

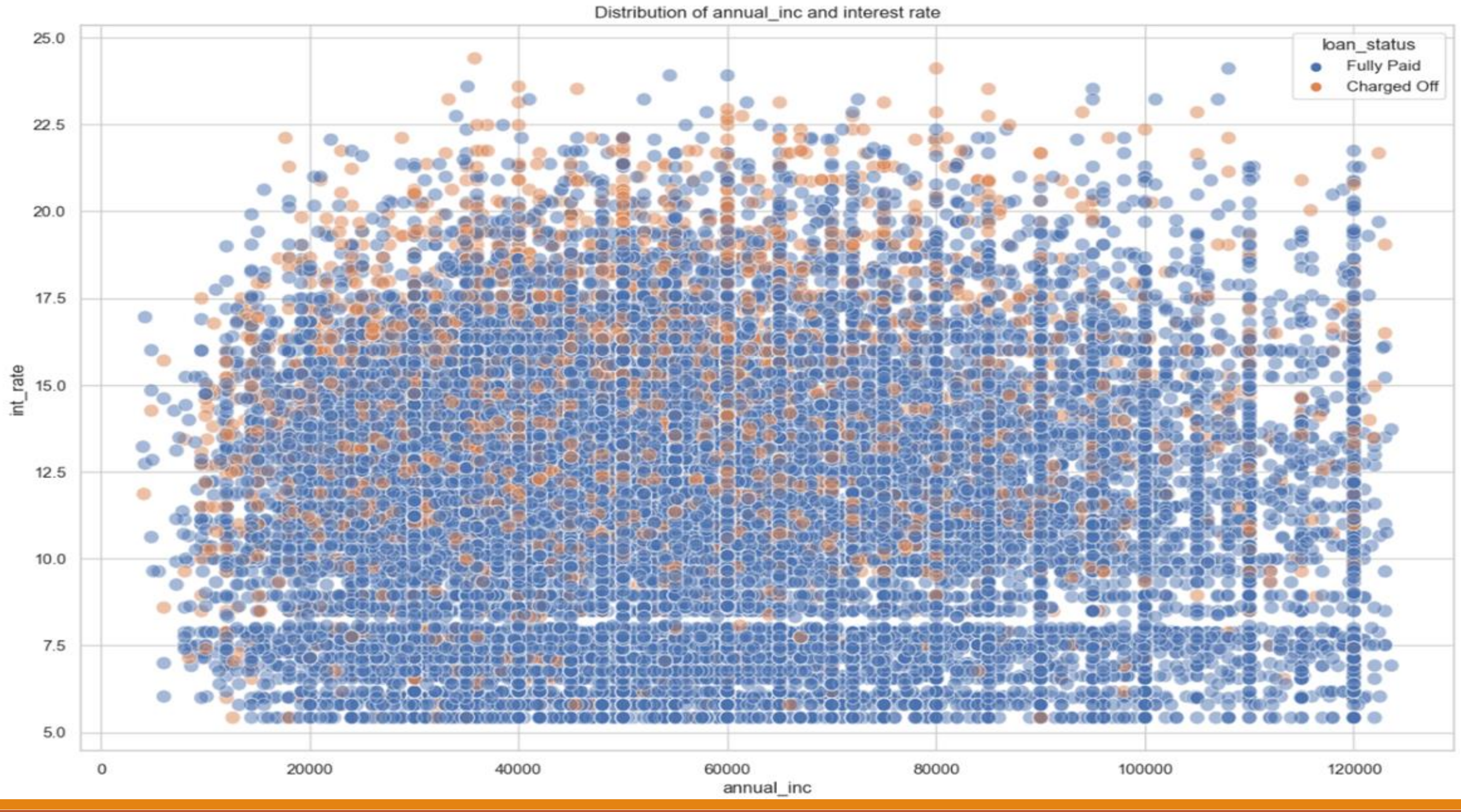


Bivariate Analysis(contd.)

- Interest Rate and Loan Amount:
    - Interest Rate Trends:
        - Number of defaulters increases with higher interest rates.
        - Significant rise in defaults when interest rates exceed 12.5%, regardless of loan amount.

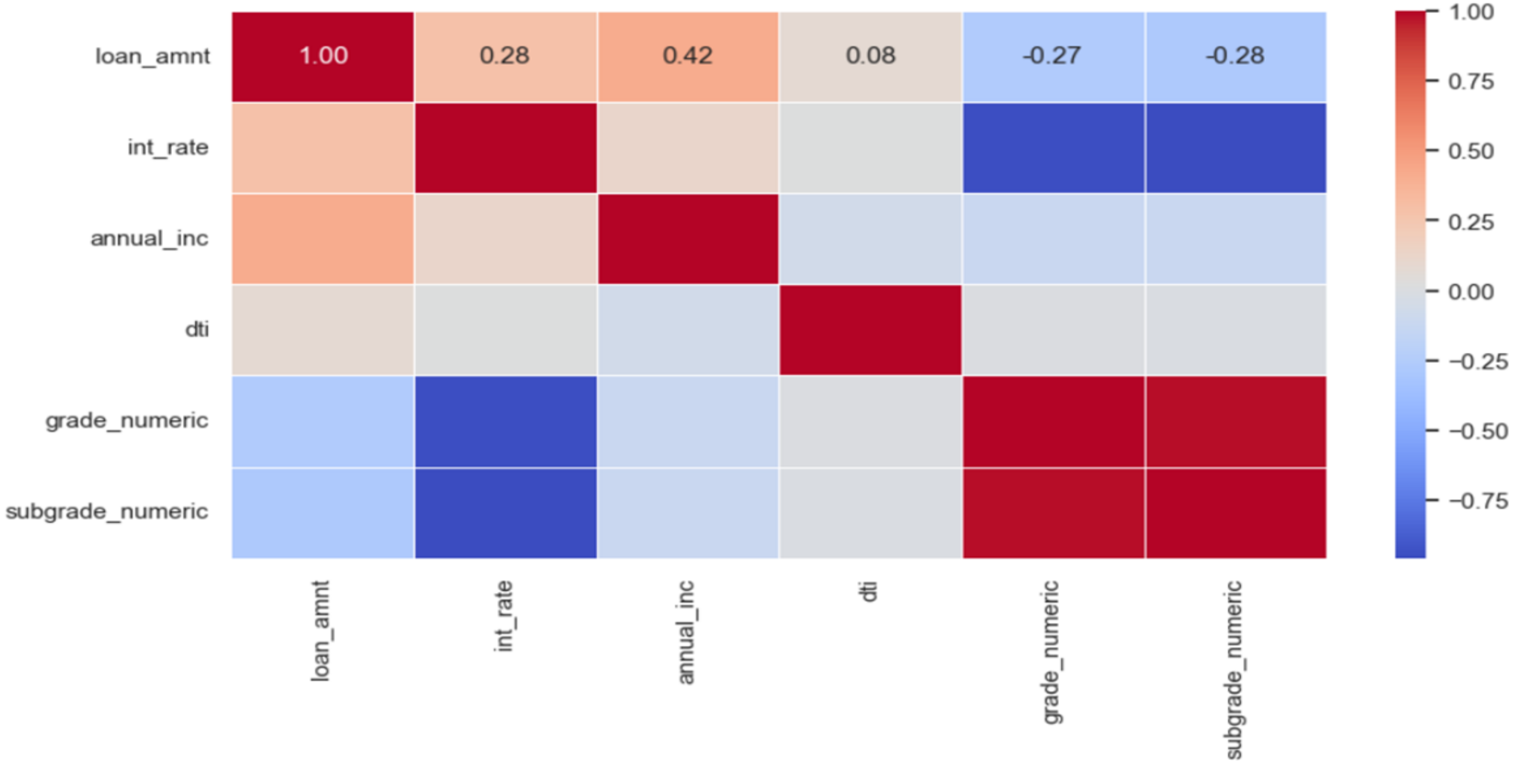
Bivariate Analysis(contd.)

- Interest Rate and Annual Income:
  - Income Trends:
    - Lower annual incomes are associated with higher default rates compared to higher annual incomes.



Multivariate Analysis

- Loan Amounts and Credit Grades
  - Observation: Lower loan amounts are associated with better credit grades among defaulters.
- Loan Amount and Annual Income
  - Correlation: A moderate positive correlation (0.37) exists between loan amount and annual income, indicating that higher incomes are linked to larger loan amounts among defaulters.
- Loan Grades and Interest Rates
  - Association: Higher loan grades are closely associated with lower interest rates among defaulters.



Conclusion

- **Loan Amount**:
  - **Insight:** Loans between $3,000 and $12,000 are at higher risk of default.
  - **Analysis:** Smaller loan amounts often correlate with higher risk grades. Increased vigilance is required for these loans.

- **Annual Income**:
  - **Insight:** Default risk is notably higher for applicants with annual incomes between $10,000 and $30,000.
  - **Analysis:** Lower income levels are a significant predictor of default. Higher incomes are generally associated with larger, less risky loans.

- **Debt-to-Income (DTI) Ratio**:
  - **Insight:** An elevated DTI ratio strongly indicates potential default.
  - **Analysis:** Applicants with higher DTI ratios face more challenges in managing debt, increasing the risk of default.

- **Interest Rates**:
  - **Insight:** Higher interest rates (10%-18%) are associated with increased default rates.
  - **Analysis:** Elevated interest rates correlate with higher-risk loans, necessitating careful assessment.

Conclusion(contd.)

- **Verification Status**:
  - **Insight:** Applicants who are "Not Verified" are more likely to default.
  - **Analysis:** Thorough verification processes can significantly reduce default risk.

- **Purpose of Loan**:
  - **Insight:** Loans for debt consolidation have the highest default rates.
  - **Analysis:** These loans require stringent scrutiny to mitigate associated risks.

- **Employment Length**:
  - **Insight:** Default rates are high among applicants with over 10 years of employment and those with very short employment durations.
  - **Analysis:** Employment length should be evaluated along with other risk factors for a comprehensive assessment.

- **Loan Grades**:
  - **Insight:** Grades B and C, especially subgrades B3, B4, B5, C1, C2, and C3, exhibit the highest default rates.
  - **Analysis:** Loans in these grades necessitate more rigorous evaluation due to their higher default risk.

Conclusion(contd.)

- **Homeownership**:
  - **Insight:** Renters have the highest default rates, followed by mortgage holders, with owners being the least.
  - **Analysis:** Homeownership status is a critical factor in assessing default risk, with renters posing the greatest risk.

- **Loan Term**:
  - **Insight:** Loans with a 36-month term have a higher default rate compared to those with a 60-month term.
  - **Analysis:** The increased monthly payment burden of shorter-term loans contributes to a higher default risk.

- **Geographic Location**:
  - **Insight:** States like California, Florida, New York, Texas, and New Jersey show the highest default rates.
  - **Analysis:** Regional economic conditions and living costs are significant factors influencing default rates.

- **Trends Over Time**:
  - **Insight:** Default rates have been rising steadily on both a monthly and yearly basis.
  - **Analysis:** Continuous refinement of risk assessment strategies is crucial to adapt to these evolving trends.

Recommendations

- **Strengthen Verification Processes**:
  - **Action:** Enhance verification procedures for loan applicants.
  - **Rationale:** By reducing the number of "Not Verified" applicants, default risks can be significantly mitigated.
- **Focused Risk Assessment**:
  - **Action:** Pay special attention to mid-range loan amounts, lower annual incomes, higher DTI ratios, and elevated interest rates.
  - **Rationale:** These factors are critical indicators of increased default risk and warrant closer scrutiny.
- **Develop State and Purpose-Based Restrictions**:
  - **Action:** Implement targeted policies for high-risk geographic regions and specific loan purposes, especially debt consolidation.
  - **Rationale:** Tailored policies address the unique risks associated with different geographic areas and loan purposes.
- **Comprehensive Risk Evaluation**:
  - **Action:** Integrate various risk factors, such as employment length, loan grade, and term duration, into the evaluation process.
  - **Rationale:** A holistic evaluation approach provides a more accurate assessment of default risk.

Thank you

UDAY KIRAN PUJAARI
VAIBHAV AGARWAL