Loan Default Analysis

A Comprehensive Analysis of Factors Influencing Loan Defaults

-*Prepared by Anche Sujana & Sai Krishna*

Introduction

- Objective: To conduct a comprehensive analysis and gain insights into the key factors influencing loan defaults. This analysis will examine various loan attributes in order to identify patterns and predictors that contribute to the likelihood of default. goal is to explore how data-driven insights can be used to minimize the risk of financial loss when lending to customers
- Dataset Overview:
  The dataset comprises 39,717 loan records with 111 variables, capturing a wide range of factors related to each loan. Key attributes include financial details such as loan amount and interest rate, as well as categorical data on loan status, which will serve as a focal point for understanding default behavior.

Data Understanding

- Identified missing values, outliers, and inconsistent formats: Detected gaps in data, extreme values, and inconsistencies in formatting across key variables.
- Key variables analyzed: Focused on loan amount, interest rate, loan status, grade, and Debt-to-Income (DTI) ratio to assess risk factors.
- Initial assessment methods: Utilized ‘data.describe()’, ‘data.info()’, ‘data.isnull.sum()’ to gain an overview of data distribution and detect potential issues.

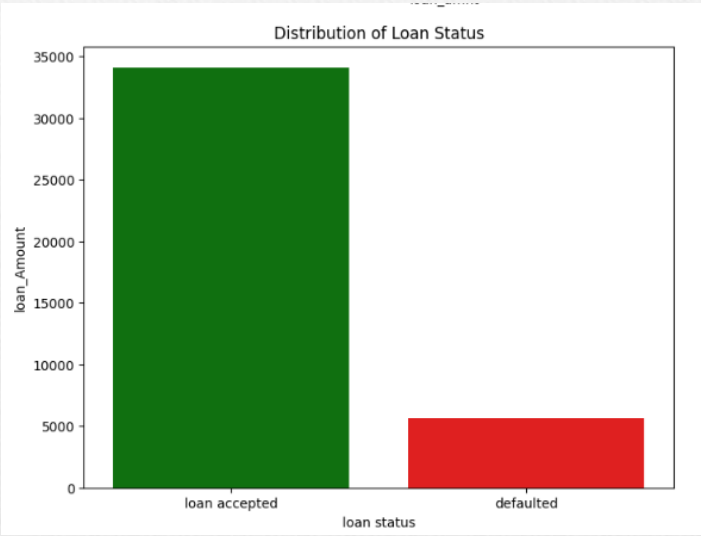
Data Cleaning and Manipulation

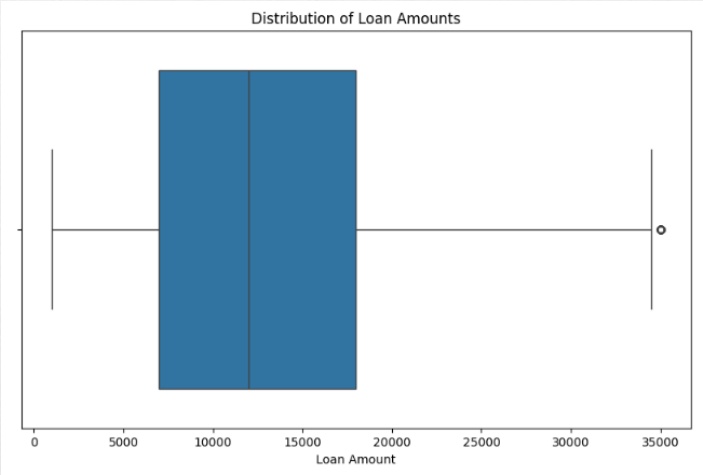
- **Dropped columns with >30% missing values:** Removed columns with excessive missing data (>30%) to improve dataset quality.
- **Imputed missing numeric values with medians**: Filled missing numeric values using median imputation to preserve data integrity.
- **Standardized/formatted columns like** emp_length and int_rate: Standardized and formatted columns such as employment work experience and interest rate for consistency.
- **Created derived metrics, e.g., Loan-to-Income Ratio**: Generated new metrics like Loan-to-Income (LTI) ratio to capture additional insights for risk assessment.

Exploratory Data Analysis: Univariate

Loan Status: Around 34,000 loans were accepted, while approximately 5,000 defaulted.

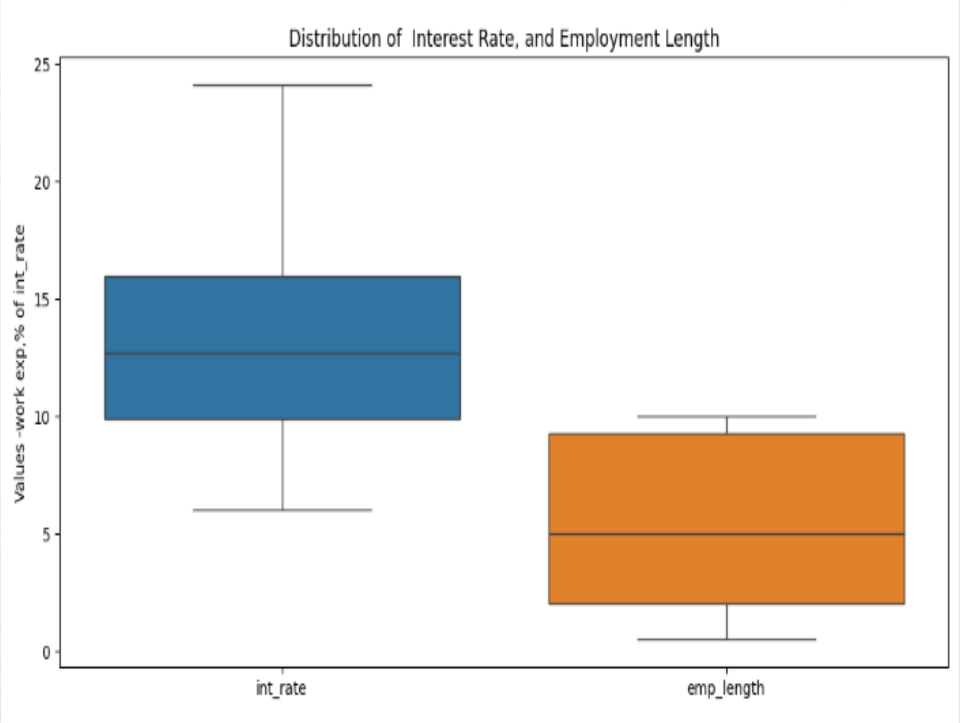
Loan Amount Distribution: The majority of loans fall between $5,500 and $15,000.





Exploratory Data Analysis: Univariate

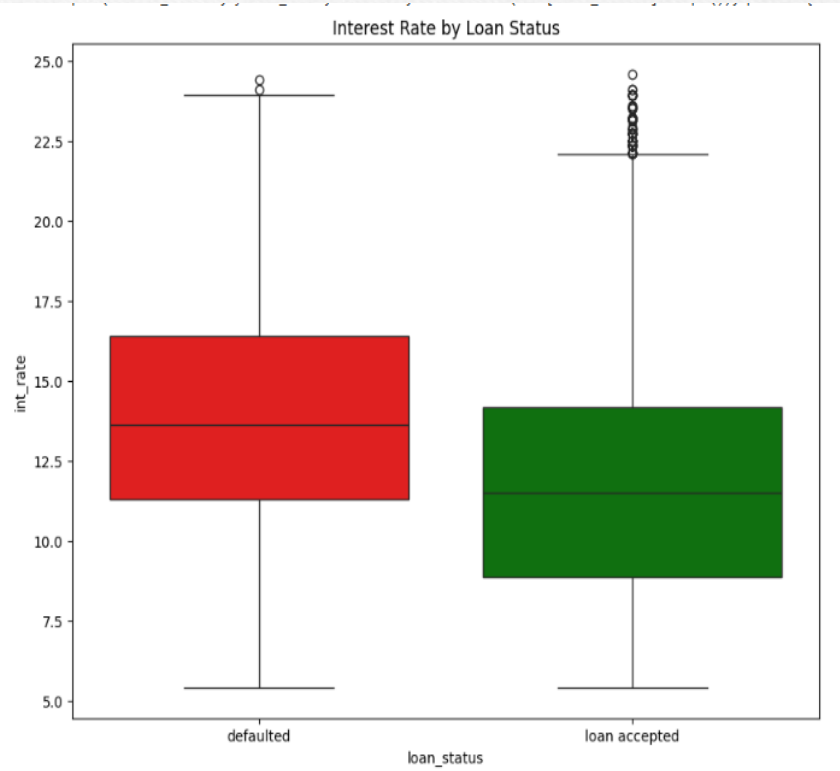
- **Interest Rate Distribution**: The average interest rate across loans is approximately 12%.
- **Employment Experience**: Applicants typically having an average of 5 years of work experience are applying for loans.



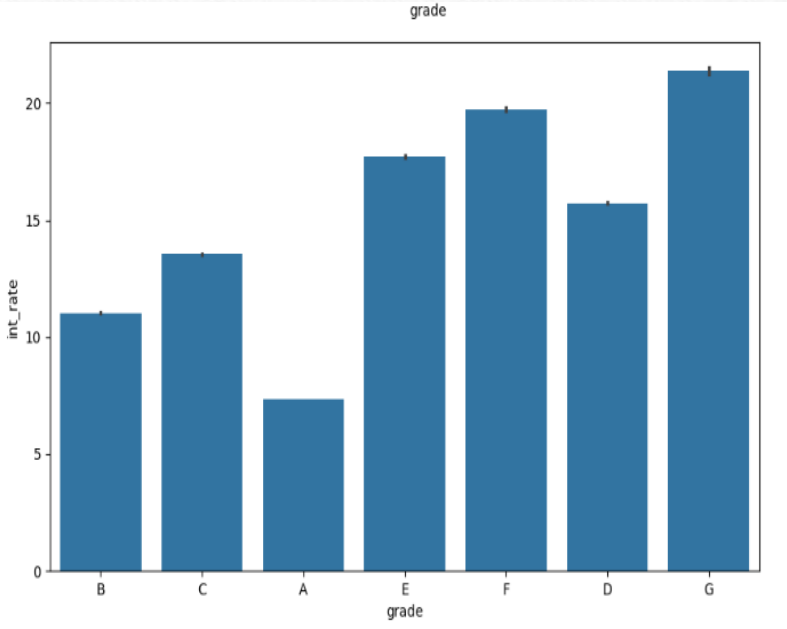
Exploratory Data Analysis: Bivariate

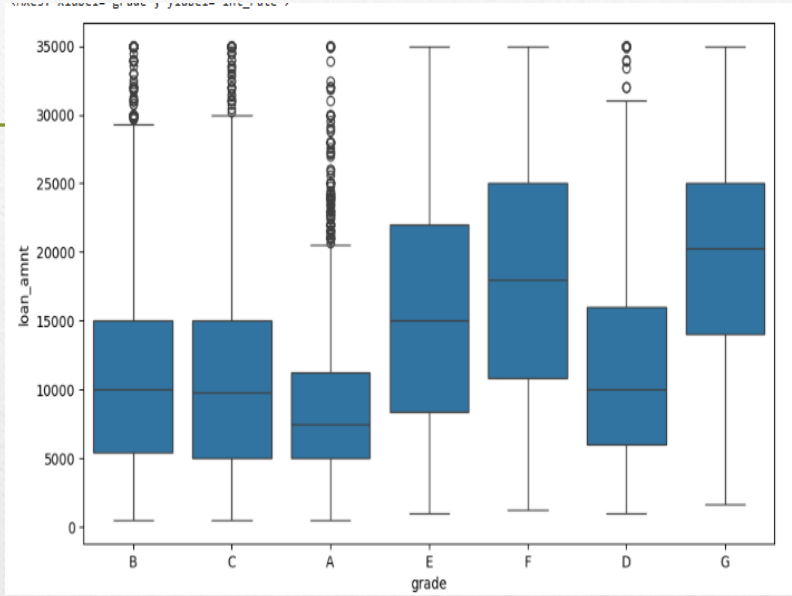
Interest Rate by Loan status:

- Indicating that higher interest rates are typically associated with defaulted customers
- By observing the horizontal line within each box (the median), we can compare the typical interest rates for each loan status. The median interest rate for defaulted loans is noticeably higher than the median for loans in good standing.



Exploratory Data Analysis: Bivariate





Grade vs Interest rate:

- Strong positive correlation between loan grade and interest rate. As the grade decreases (from A to G), the average interest rate increases.
- This confirms that interest rates are largely determined by the perceived risk associated with a borrower

Grades Vs Loan amounts:

- Higher grades (like A and B) tend to have lower loan amounts, while lower grades (like F and G) have higher loan amounts.

Multivariate Analysis

- - Correlation heatmap reveals strong relationships between numeric features.
- - Pair plots highlight patterns between loan amount, interest rate, and status.
- - Focused on key drivers of loan defaults for deeper insights.

Segmentation Analysis

- - Analyzed loan performance across states (`addr_state`).
- - Identified states with higher default rates using bar plots.
- - Observed trends linking default rates to state-specific economic conditions.

Driver Analysis

- - Key drivers of defaults:
- - Loan Grade: Grades F and G show higher default rates.
- - DTI: Higher DTI ratios increase the risk of default.
- - Annual Income: Lower income levels correlate with higher defaults.

Recommendations

- Can adjust approval criteria based on specific attributes (e.g., grades, DTI thresholds).Can periodically monitor loan portfolio performance by tracking key metrics like loan status, interest rates, and default rates.
- Adjust approval criteria or pricing models based on above insights to minimize risk and to gain profits.
- Further segment your borrowers to identify subgroups with unique risk profiles and can customize lending strategies accordingly.

Conclusion

We observed a strong correlation between loan grade, interest rate, and loan status, highlighting that higher-risk borrowers are typically assigned higher interest rates and potentially larger loan amounts. Borrower profile is also essential for minimizing financial losses.

By understanding these relationships, businesses can refine their lending practices, improve risk assessment models