RISK ANALYTICS FOR LOAN DEFAULT PREDICTION

INTRODUCTION

1. Objective:

   Identify and mitigate credit loss through better risk assessment using EDA.

2. Problem Statement:

- As a consumer finance company, we provide various loans to urban customers.
- We face the challenge of making informed decisions to approve or reject loan applications.
- Balancing between approving loans to reliable applicants and avoiding potential defaults is crucial.

Key Questions

- How do consumer and loan attributes influence the tendency of loan default?
- Can we uncover significant driver variables behind loan defaults using EDA?

Business Understanding

1. Company Profile:

- We are a consumer finance company specializing in various loan types for urban customers.
- Our success hinges on effective loan approval decisions that manage risks and drive profitability.

2. Loan Approval Dilemma:

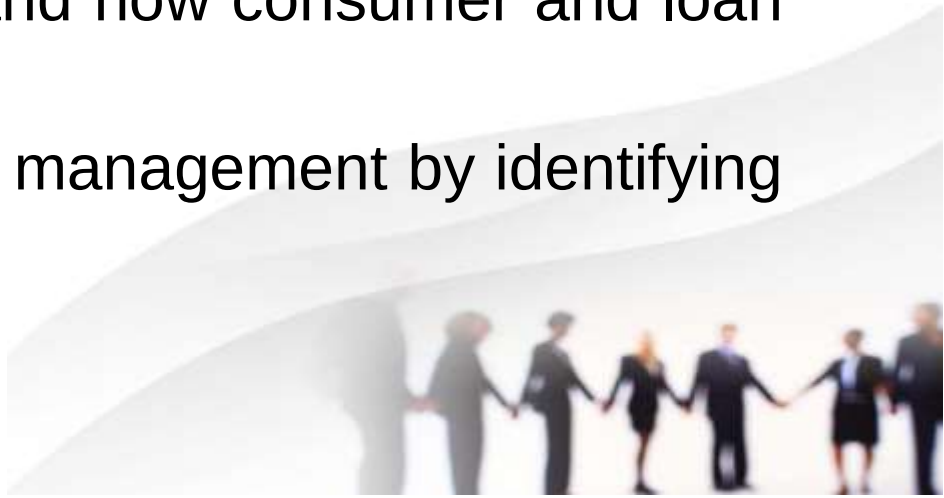
- Approve loans: Potential revenue, but risk of defaults.
- Reject loans: Loss of business opportunity if reliable applicants are rejected.

3. Risk Management:

- Two types of risks: Losing business due to rejected loans, and financial loss from defaults.
- Need to strike a balance between profit and risk through data-driven decisions.

4. Objective:

- Use Exploratory Data Analysis (EDA) to understand how consumer and loan attributes influence loan default.
- Enhance risk assessment strategies and portfolio management by identifying key factors linked to defaults.



Data Overview and Preparation

1. Data Import and Initial Inspection:

- Imported dataset containing loan application information.
- Explored dataset to understand its structure and attributes.

2. Data Cleaning and Formatting:

- Handled missing values by dropping columns with over 30% missing data.
- Focused on relevant columns for analysis.

3. Data Type Correction:

- Corrected data types for "int_rate" and "revol_util" columns.
- Replaced state abbreviations with full state names.

4. Outlier Handling:

- Used box plots to visualize and manage outliers in numerical columns.
- Addressed outliers in "loan_amnt," "int_rate," "total_pymnt," and more.

5. Standardization and Formatting:

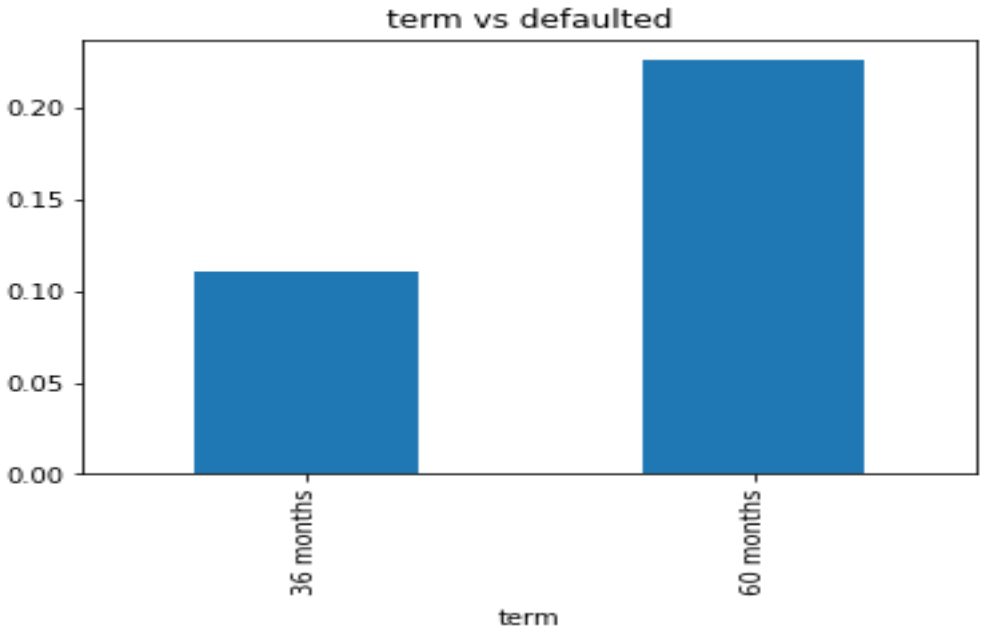
- Ensured that numerical columns were on the same scale.
- Rounded "total_pymnt" values to two decimal places for consistency.

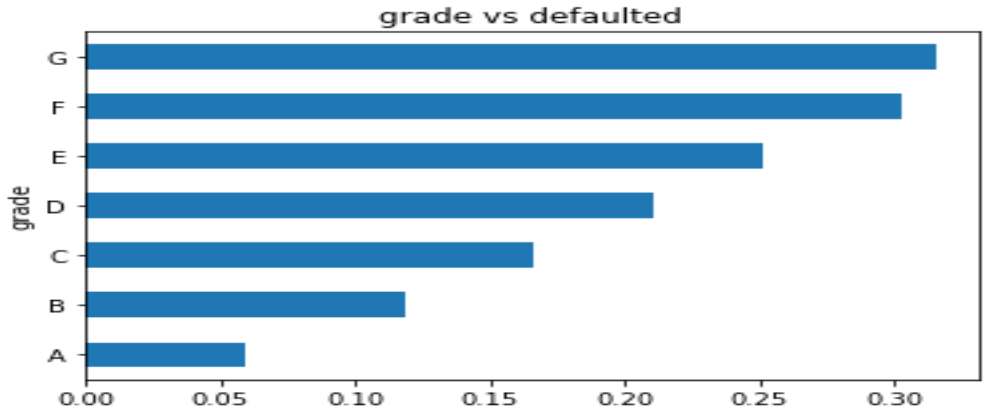
6. Exploratory Data Analysis (EDA) Techniques Applied:

- Univariate Analysis: Explored individual attributes to understand their distribution and patterns.

- Bivariate Analysis: Investigated relationships between attributes and their impact on loan defaults.
- Multivariate Analysis: Analyzed interactions between multiple attributes to gain deeper insights.

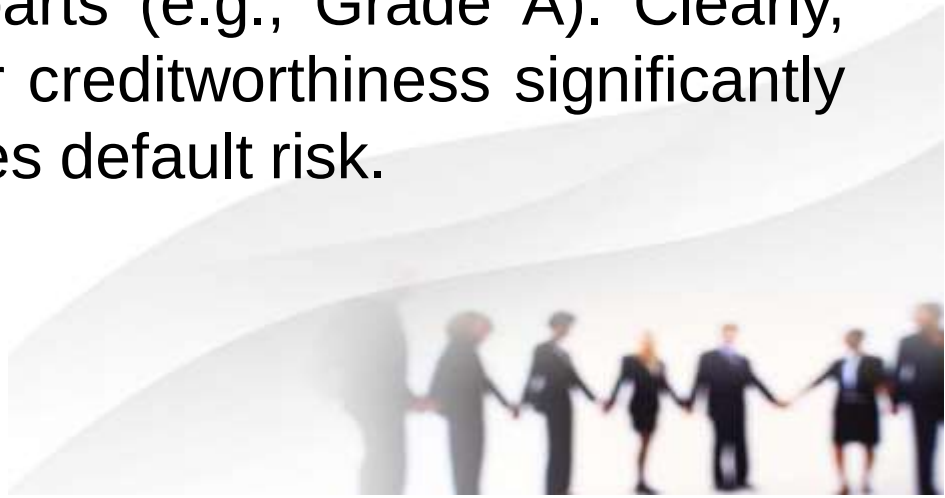
Insights



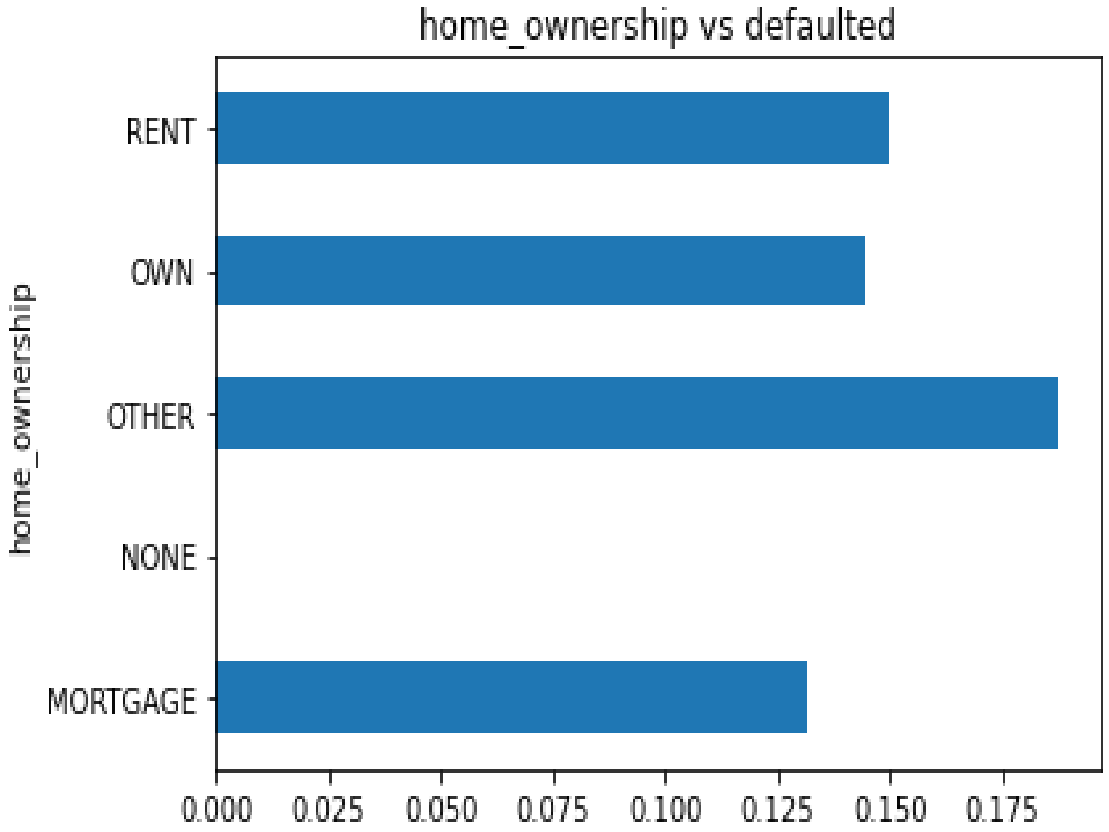


Loan Term and Grade:

Loans with extended terms, particularly 60 months, exhibit heightened default probabilities. Moreover, loans graded lower (e.g., Grade G) present higher default rates compared to more creditworthy counterparts (e.g., Grade A). Clearly, borrower creditworthiness significantly influences default risk.

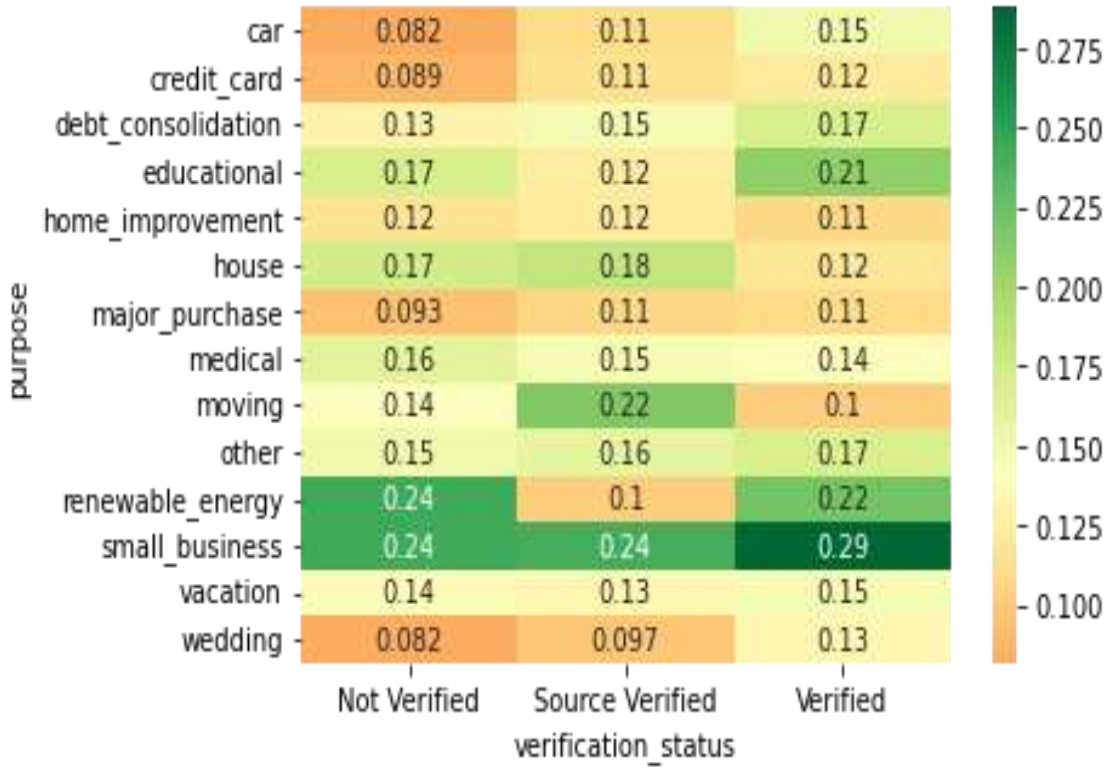


Home Ownership:



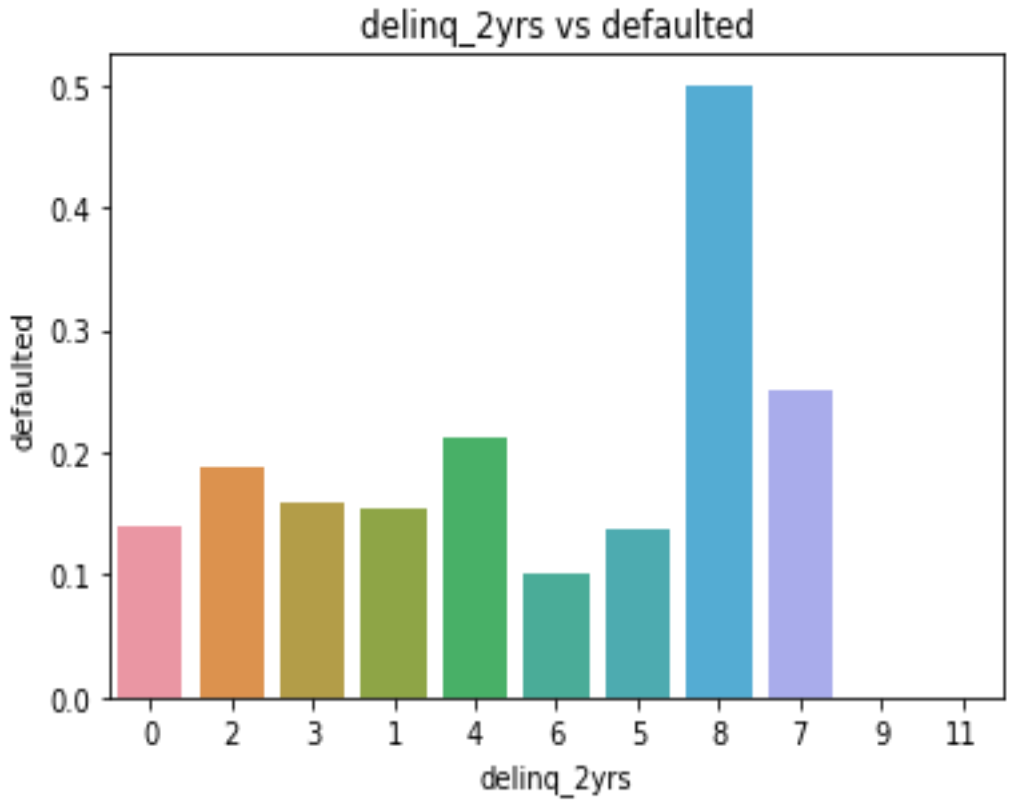
Borrowers categorized as "Other" in homeownership status manifest the highest default rates. In contrast, those with "None" as their status show the lowest default rates. This underscores the significance of stability in homeownership for risk assessment.

Loan Purpose



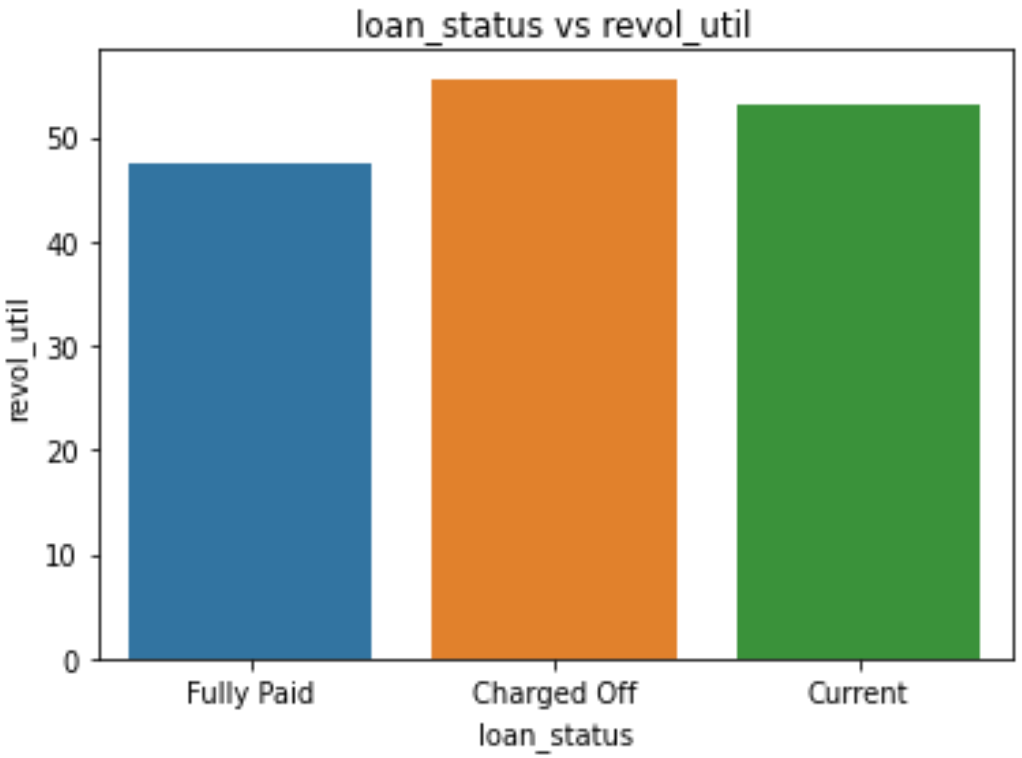
"Small Business" loans, especially when verified, display elevated default rates. Similarly, "Educational" and "Moving" loans carry a higher likelihood of default, particularly when combined with specific verification statuses.Borrowers with "Verified" status generally demonstrate a greater tendency to default, especially in conjunction with specific loan purposes.

Number of Delinquencies



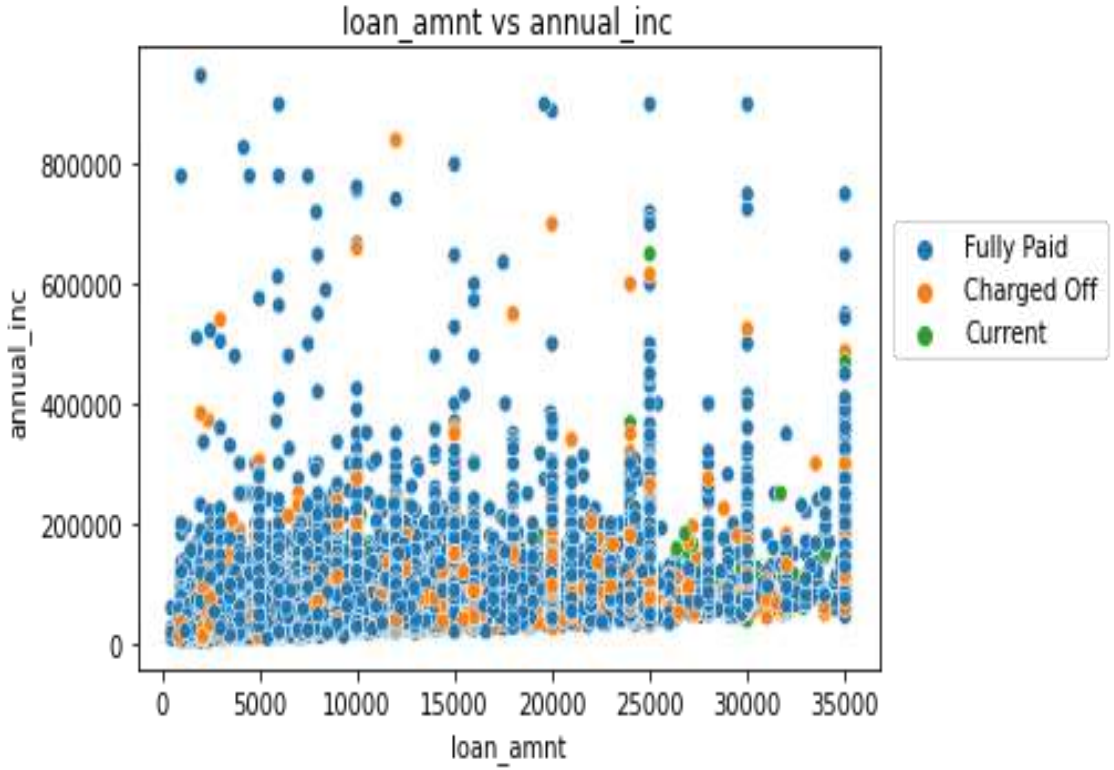
A history of multiple delinquencies over the past two years correlates with an increased default risk. As delinquency instances rise, so does the likelihood of default.

Revolving Credit Utilization

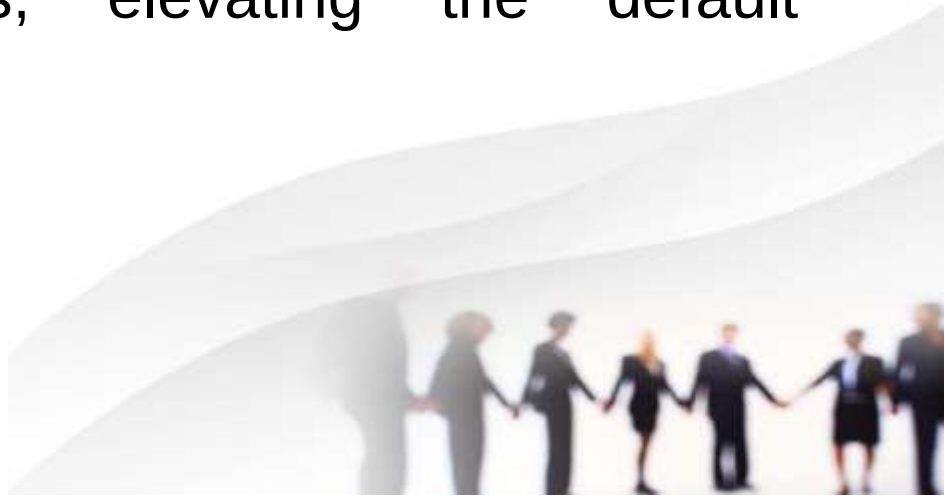


Higher utilization of revolving credit lines connects with elevated default risk. Borrowers with substantial credit utilization might struggle with managing debt, leading to a heightened probability of default.

Annual Income

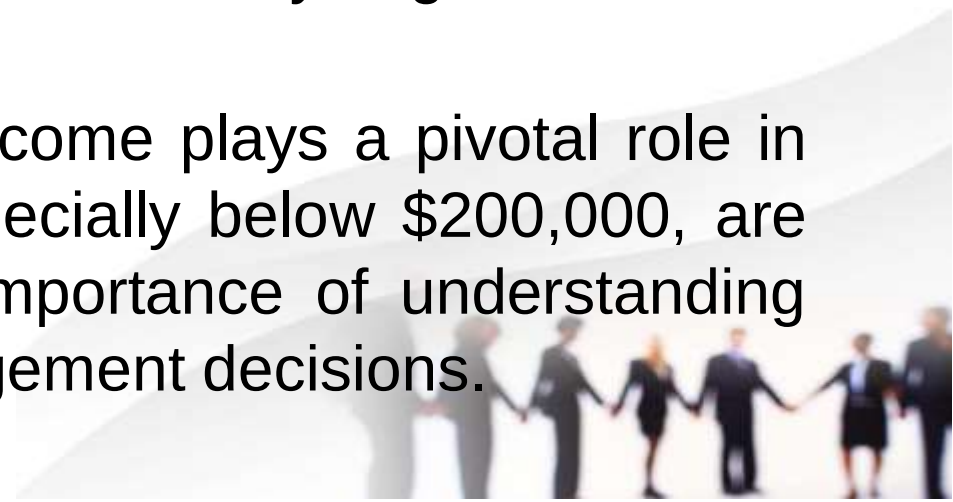


Annual income emerges as a pivotal factor impacting loan default rates. The analysis reveals that borrowers with lower annual incomes, especially below $200,000, exhibit elevated default risk. Lower income levels might translate to repayment challenges, elevating the default likelihood.



Summary of Key Takeaways for Risk Management

- **Loan Term and Grade:** Longer terms and lower grades correlate with higher default rates, urging vigilance in assessing loan terms and creditworthiness.
- **Home Ownership:** "Other" homeownership status signifies elevated default risk, emphasizing stability in homeownership.
- **Loan Purpose:** Caution is warranted for "Small Business," "Educational," and "Moving" loans, particularly with certain verification statuses.
- **Verification Status:** "Verified" borrowers may still carry significant risk, warranting thorough evaluation.
- **Income and Repayment Capability:** Annual income plays a pivotal role in default risk. Borrowers with lower incomes, especially below $200,000, are more prone to default. This underscores the importance of understanding borrowers' income levels for informed risk management decisions.



Conclusion

- By decoding these insights through EDA, we can refine risk assessment practices and optimize portfolio management strategies.
- Visualizations facilitate intuitive comprehension of intricate relationships, guiding informed decisions.
- Incorporating these findings into risk assessment protocols can foster more effective loan default mitigation and overall risk management.

THANK YOU