

A decorative graphic on the left side of the slide, consisting of white lines and circles on a blue gradient background, resembling a circuit board or a stylized tree structure.

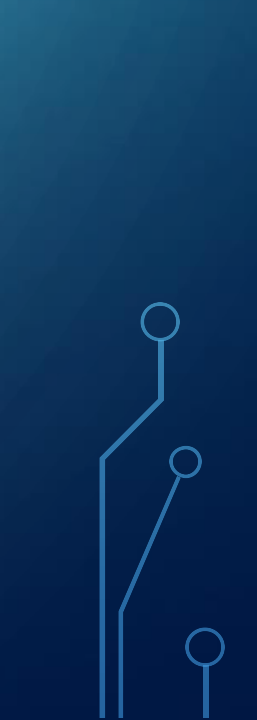


ANALYSIS OF LOAN DEFAULT RISK: LENDING CLUB CASE STUDY

BY RAVIKIRAN KRISHNAPRASAD AND AMIT KUMAR



ABSTRACT

This report presents a detailed analysis of the Lending Club's loan dataset with the objective of identifying key predictors of loan default. Through rigorous data cleaning, exploratory data analysis (EDA), and multivariate analysis, we uncover patterns and relationships within the data that signal potential default risk. Our findings aim to assist the Lending Club in minimizing credit losses by improving loan approval criteria and risk assessment strategies.



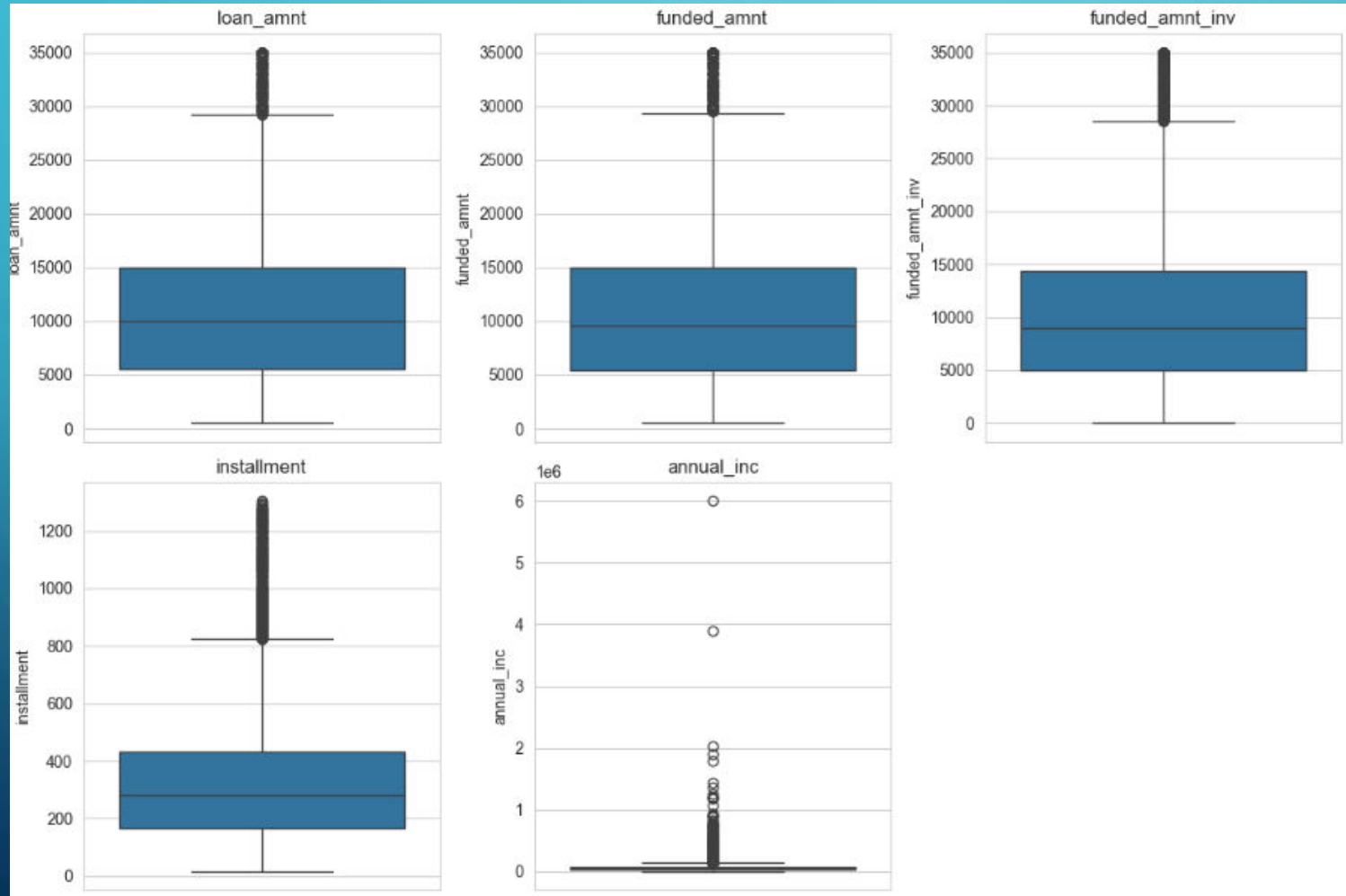
INTRODUCTION

Lending Club operates a consumer finance platform that matches borrowers seeking personal loans with investors. However, lending money to "risky" applicants poses a significant threat of credit loss. This study leverages Exploratory Data Analysis (EDA) to identify risky loan applicants, thereby helping in reducing credit loss. The primary goal is to discern the driving factors behind loan defaults, enabling informed decision-making for loan approvals.

DATA PREPARATION AND CLEANING

- **Initial Data Understanding:** The dataset initially contained 39,717 entries across 111 variables, including loan amount, interest rates, employment details, and loan status.
- **Null Value Treatment:** Columns with 100% missing values were removed, significantly reducing the dimensionality of the dataset to more manageable levels.
- **Irrelevant Column Elimination:** Columns with single unique values or those not contributing to loan approval (like customer IDs and descriptions) were dropped.
- **Data Type Conversion:** Percentage strings in 'int_rate' and 'revol_util' were converted to floats for numerical analysis.
- **Outlier Identification:** Using box plots, outliers, especially in the 'annual_inc' column, were identified for further treatment.

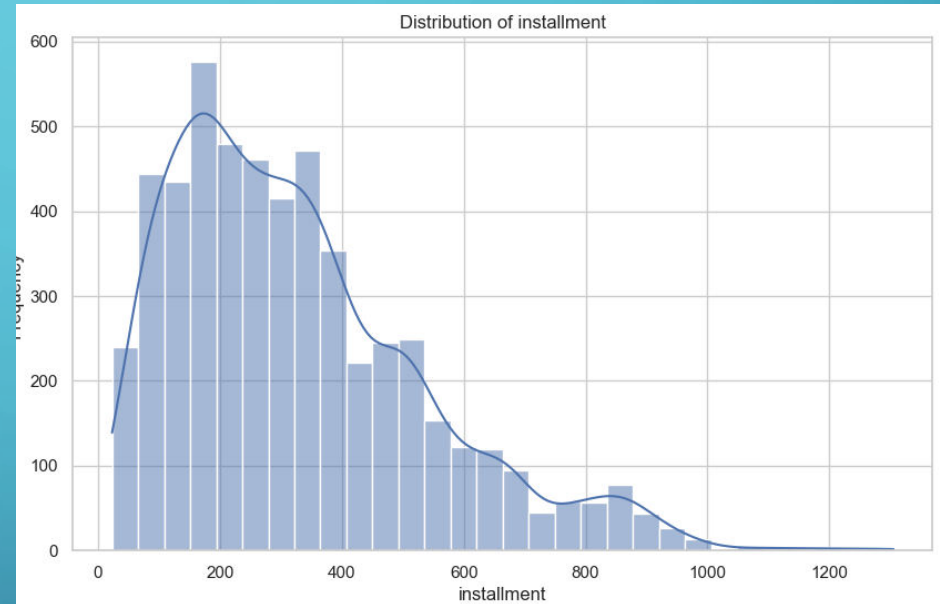
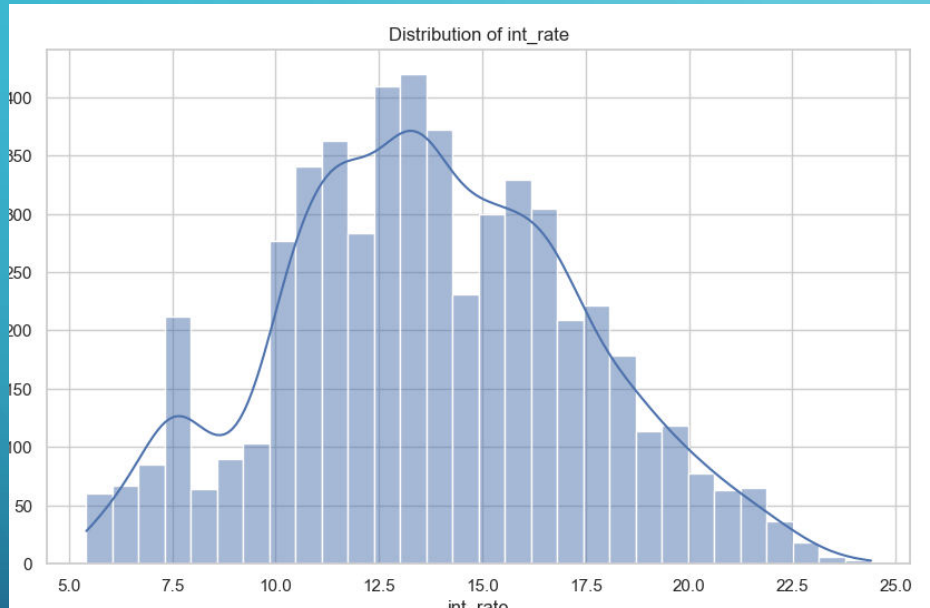
DATA PREPARATION AND CLEANING



EXPLORATORY DATA ANALYSIS (EDA)- NUMERICAL FEATURE ANALYSIS

- Analyzed distributions of key numerical features such as loan amount, funded amount, interest rate, and annual income.
- Identified ranges of values where the likelihood of default significantly increases. For example, loans with interest rates higher than 10% showed a higher default rate.

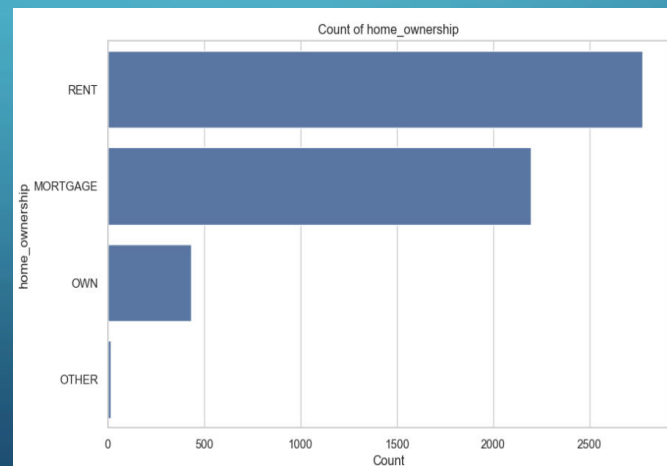
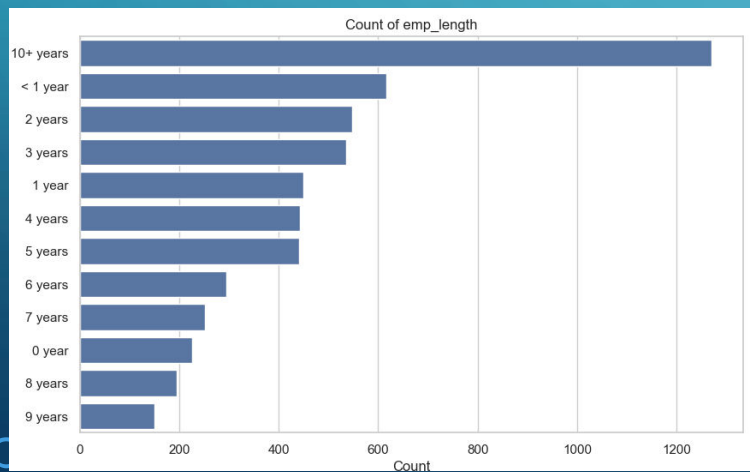
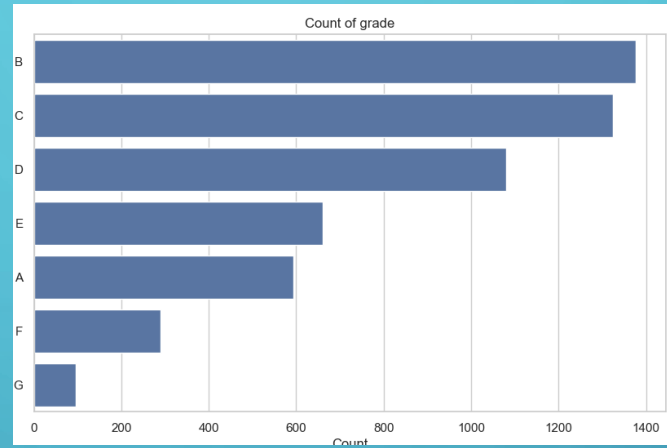
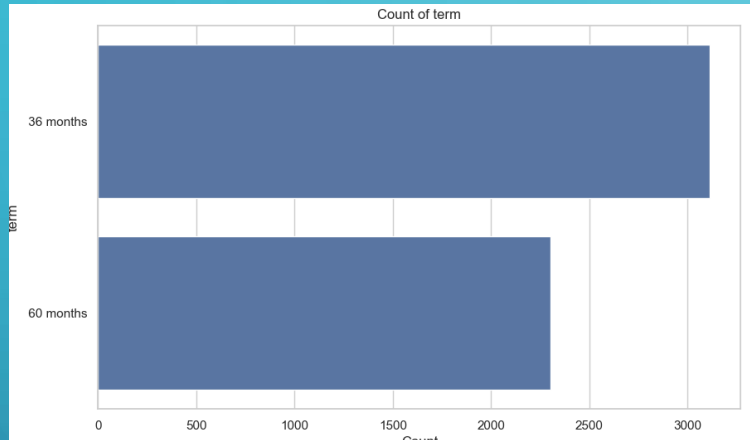
EXPLORATORY DATA ANALYSIS (EDA)- NUMERICAL FEATURE ANALYSIS



EXPLORATORY DATA ANALYSIS (EDA)- CATEGORICAL FEATURE ANALYSIS

- Evaluated the impact of categorical variables like term, grade, employment length, and home ownership on loan default rates.
- Found that loans with terms of 36 months, grades B, C, and D, and purposes such as debt consolidation showed higher default rates.

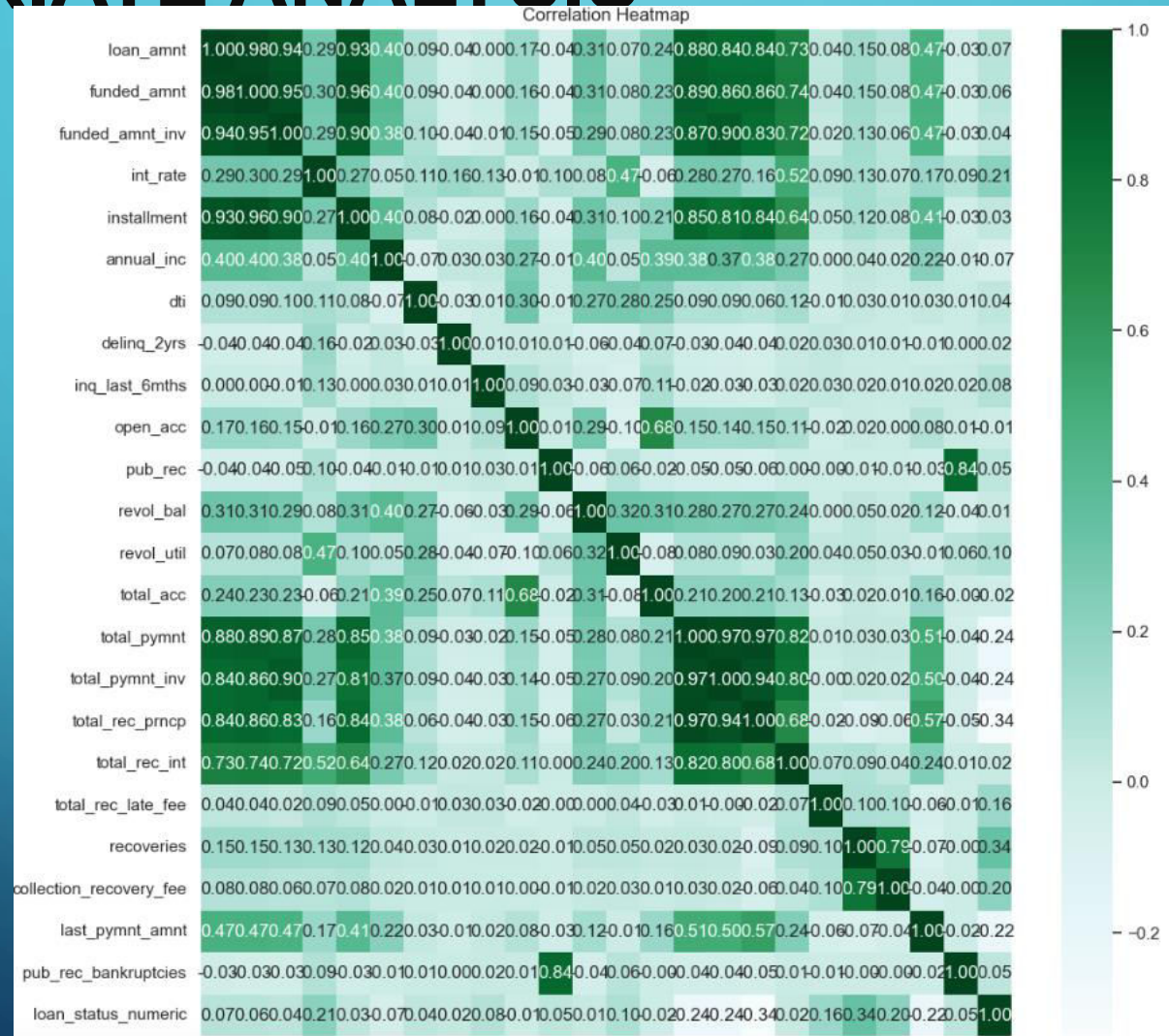
EXPLORATORY DATA ANALYSIS (EDA)- CATEGORICAL FEATURE ANALYSIS



EXPLORATORY DATA ANALYSIS (EDA)- MULTIVARIATE ANALYSIS

- Correlation analysis revealed that interest rates, recoveries, and collection recovery fees are positively correlated with loan default risk.
- Additionally, a higher funded amount and a higher installment were found to be associated with an increased likelihood of default.

EXPLORATORY DATA ANALYSIS (EDA)- MULTIVARIATE ANALYSIS



KEY FINDINGS

- 1. Interest Rates:** Loans with higher interest rates tend to have higher default rates, suggesting the burden of higher costs on borrowers.
- 2. Loan Purpose:** Debt consolidation emerged as a significant purpose for which default rates were high, indicating financial distress among these borrowers.
- 3. Employment Length:** Contrary to expectations, employment length did not significantly affect default rates, suggesting that job stability alone is not a reliable predictor.
- 4. Home Ownership:** Borrowers who rent or have a mortgage are slightly more likely to default, potentially due to higher fixed monthly expenses.

RECOMMENDATIONS

- Based on our findings, we recommend the Lending Club:
- Refine loan approval criteria with a heightened focus on interest rates and loan purposes.
- Consider additional screening for debt consolidation loan applicants.
- Incorporate more nuanced risk assessment models that account for the multifaceted nature of loan defaults.

CONCLUSION

This analysis underscores the complexity of predicting loan defaults. By combining insights from numerical and categorical data, we provide a more nuanced understanding of default risk. The recommendations offered aim to strike a balance between minimizing credit loss and fostering healthy loan issuance.

ACKNOWLEDGMENTS

We extend our gratitude to the IIITB and UPGRAD communities for their invaluable resources and discussions, which have significantly contributed to this analysis.