

LENDING CLUB CASE STUDY

By Aafreen Khan

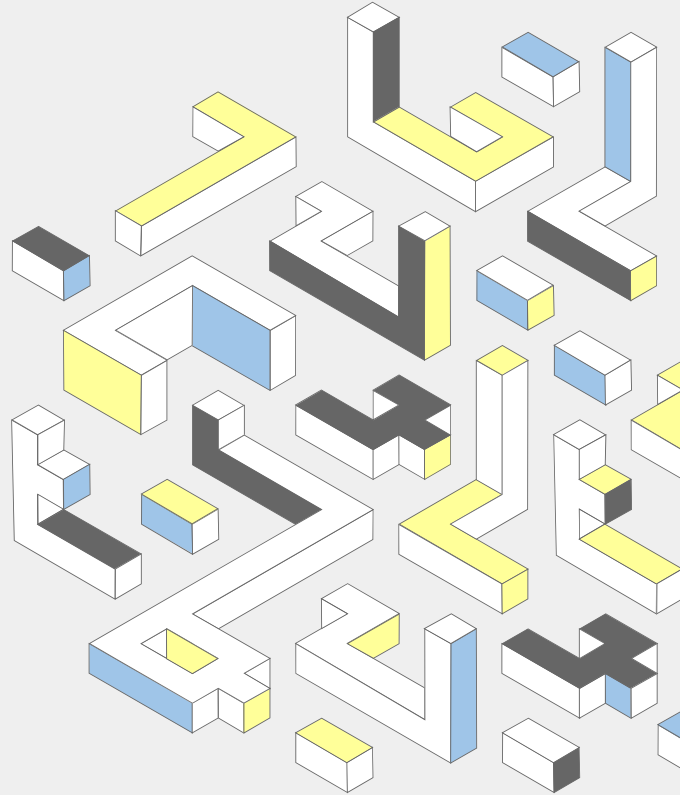


Table of contents

01

Business Objective

02

Data Cleaning and
Creating Derived
Metrics

03

Univariate Analysis

04

Bivariate Analysis

05

Final Analysis

06

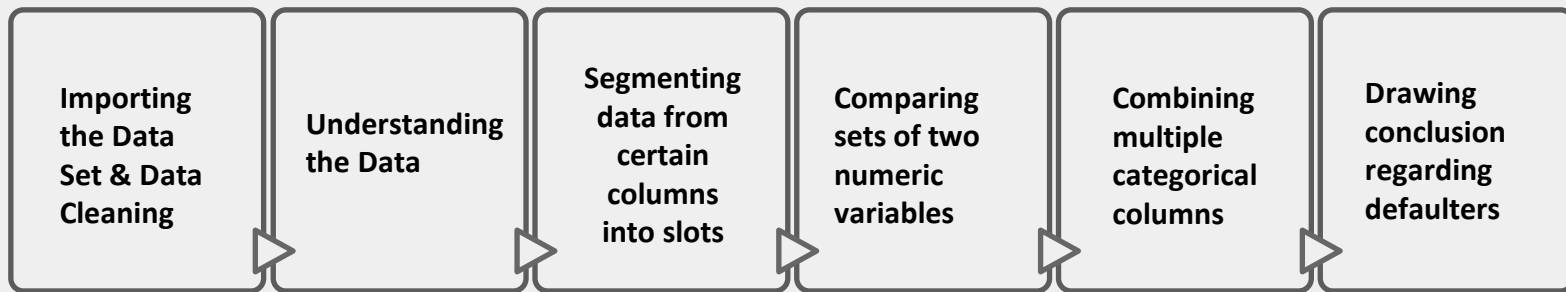
Conclusion

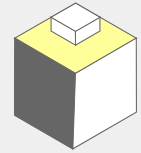
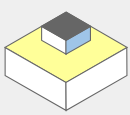
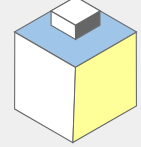




Business Objective

The objective is to identify patterns which indicate if a applicant of a loan is likely to default, which may be used by a consumer finance company for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc. The goal is to understand and highlight the consumer attributes and loan attributes that influence the tendency of default.





Data Cleaning and Creating Derived Metrics

The following data cleansing processes were applied to make the data dependable so that it can provide significant business value by improving Decision Making Process:

Removing columns with Unique Values

- Some of the columns from the loan data frame have only one value. Deleting these columns with one unique value from the data frame.

Removing 'Current Loans'

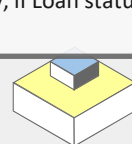
- For the current analysis, we are interested in only FULLY PAID and CHARGED OFF loan statuses. Removing all loans with statuses = "CURRENT" because we have no way of knowing if they will turn out to be fully_paid or charged_off loans.

Dropping Null values in Rows and Columns

- The null values in some columns that aren't required for analysis have been dropped from the loan_dataframe
- High Percentage of Null values in some columns doesn't make any sense or help in analysis. Hence all rows containing null values for these columns were dropped from master_frame.

Creating new columns for Analysis

- Some columns like 'int_rate%' & 'revol_util%' have been broken into categorical slots for segmented analysis.
- Creating 2 new columns 'charged_off' and 'fully_paid' with values 1 for True and 0 for False. If Loan status = Charged Off, CHARGED_OFF column will have the value 1, else 0. Similarly, if Loan status = 'Fully Paid', FULLY_PAID column will have the value 1, else 0.



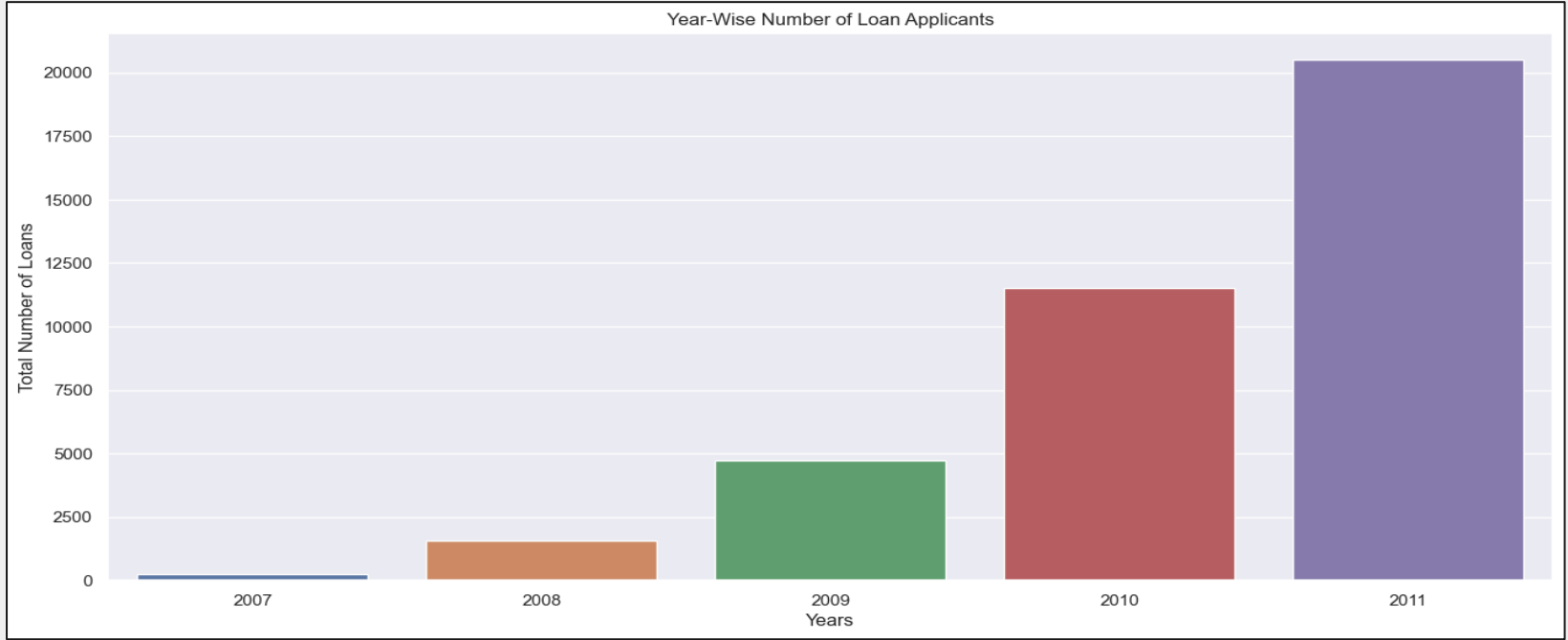
UNIVARIATE ANALYSIS

Univariate Analysis is an analysis of a single variable.

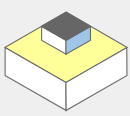
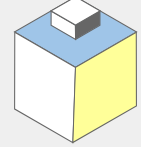
Because Univariate Analysis does not involve relationships between two or more variables, its purpose is more toward descriptive rather than explanatory.



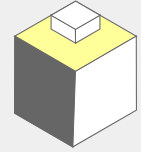
Increase in loan amounts year by year in Lending Club



During the period from 2007 to 2011, there was a significant increase in the number of loan applicants, indicating substantial growth in the applicant pool over those five years

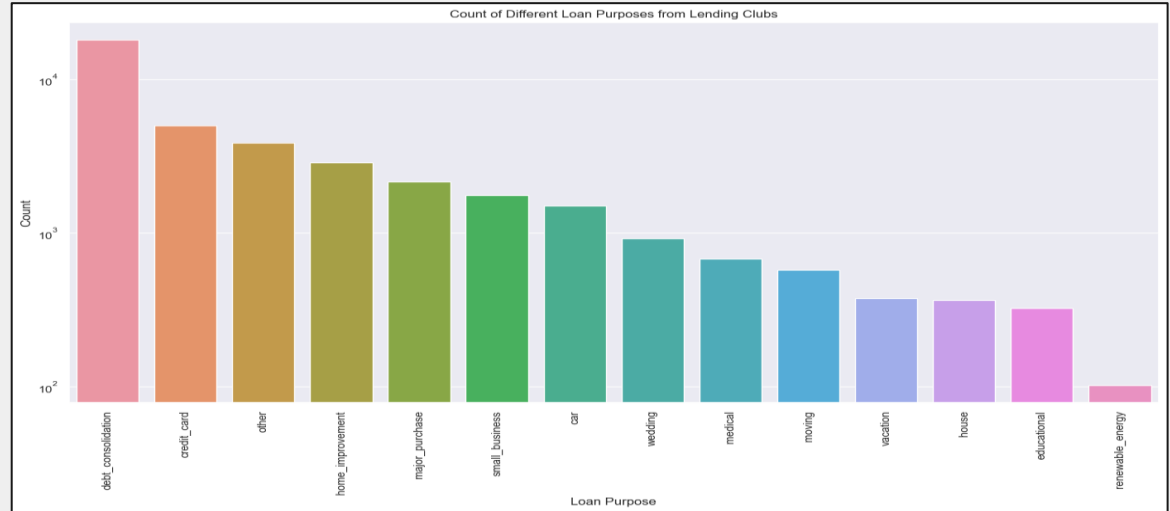


Major purposes for Loan Application

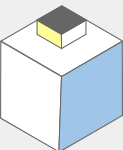


In our analysis, we identified five major purposes for loan applications, with 'debt_consolidation' being the most common loan purpose. The top five loan purposes, in descending order of frequency, are as follows:

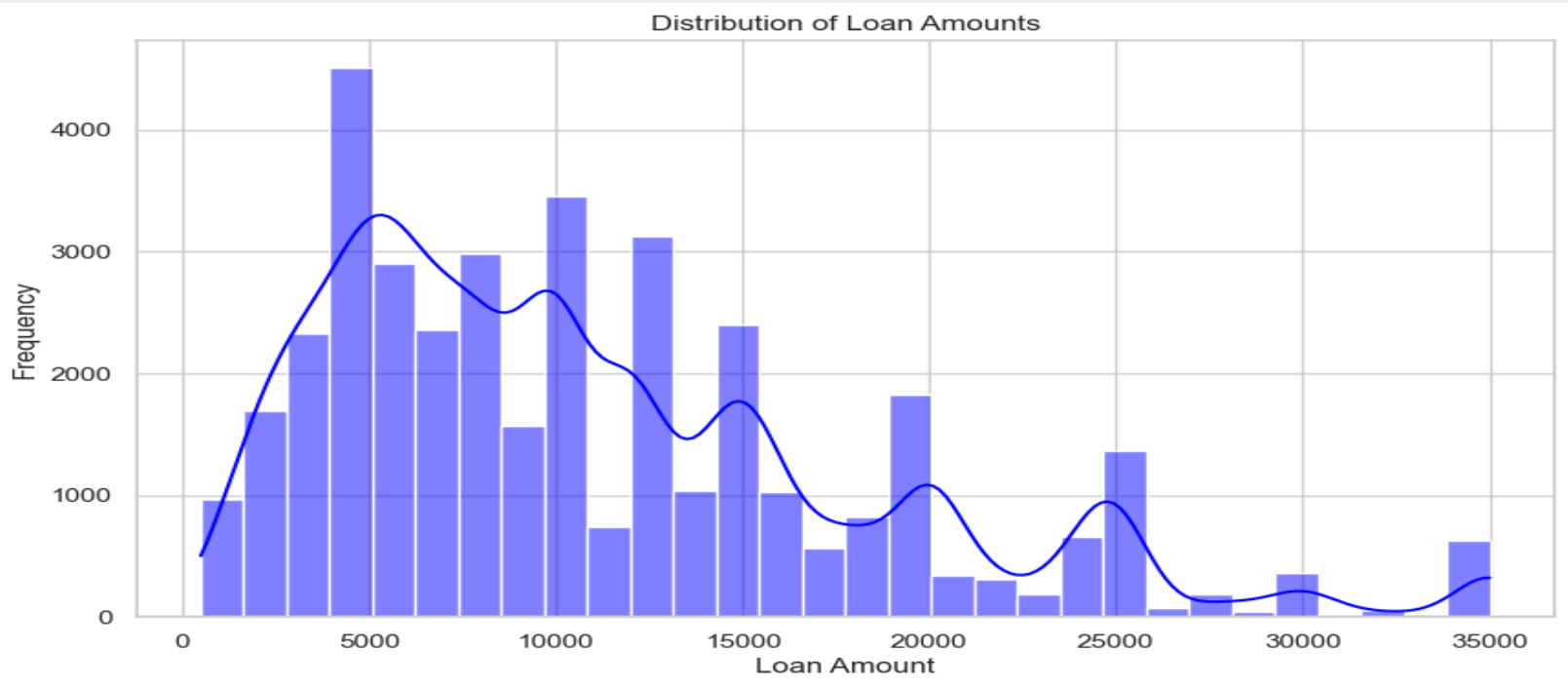
1. 'debt_consolidation'
2. 'credit_card'
3. 'home_improvement'
4. 'other'
5. 'major_purchase'



This insight highlights the prevalence of 'debt_consolidation' as the primary motivation for loan applications within our dataset

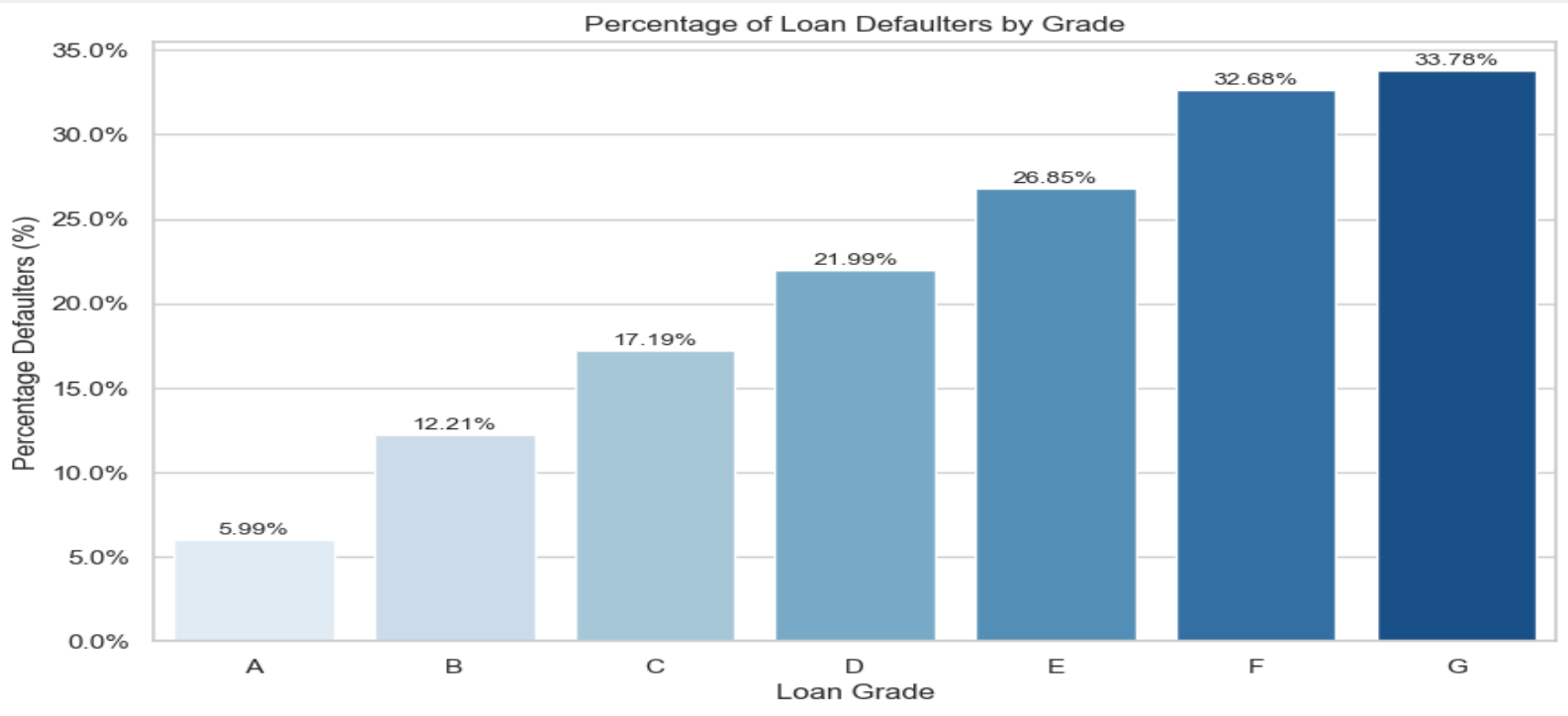


The Distribution Of Loan Amount



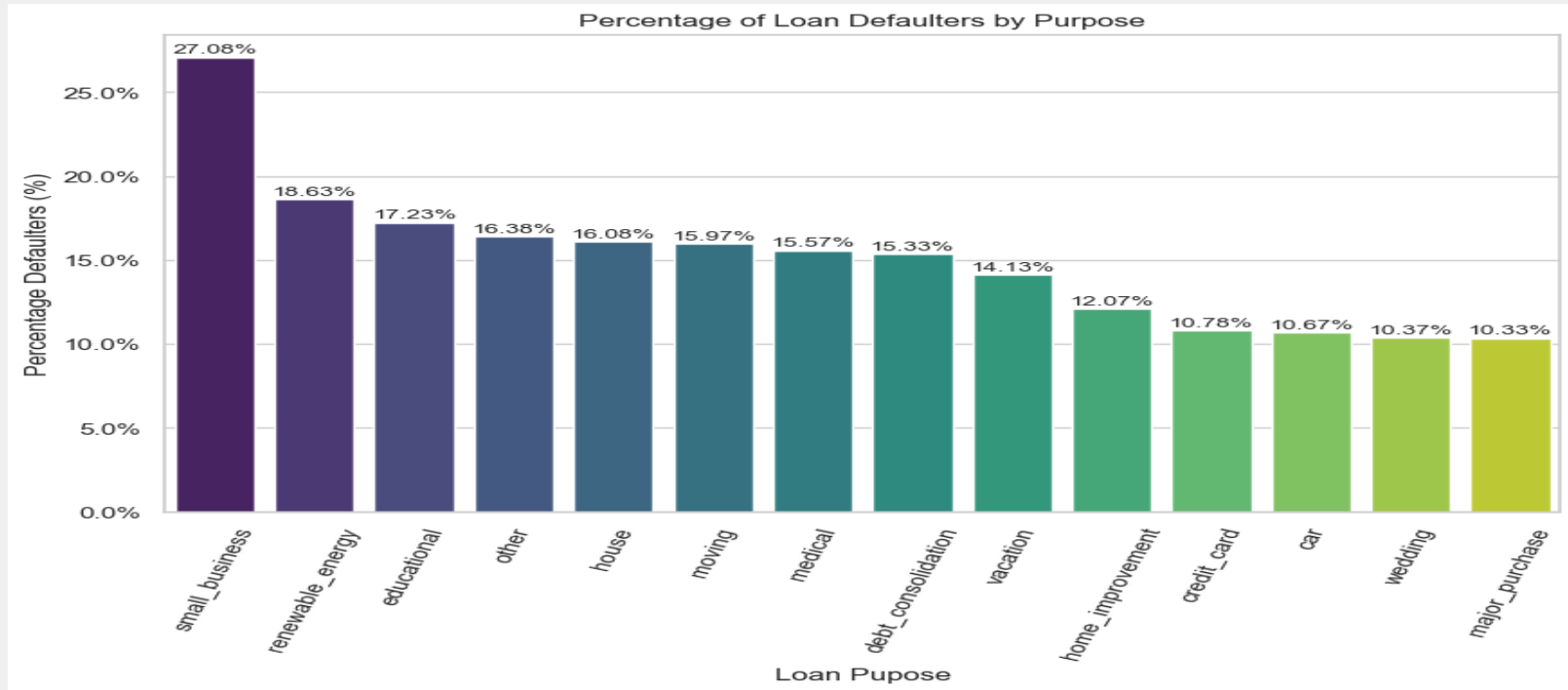
From the plot above, we can visualize the distribution of loan amounts applied by borrowers. It's evident that the majority of loan applicants fall within the loan amount range of 5000 to 10000, indicating that this range is the most common among applicants.

The Percentage of Defaulters by Grade



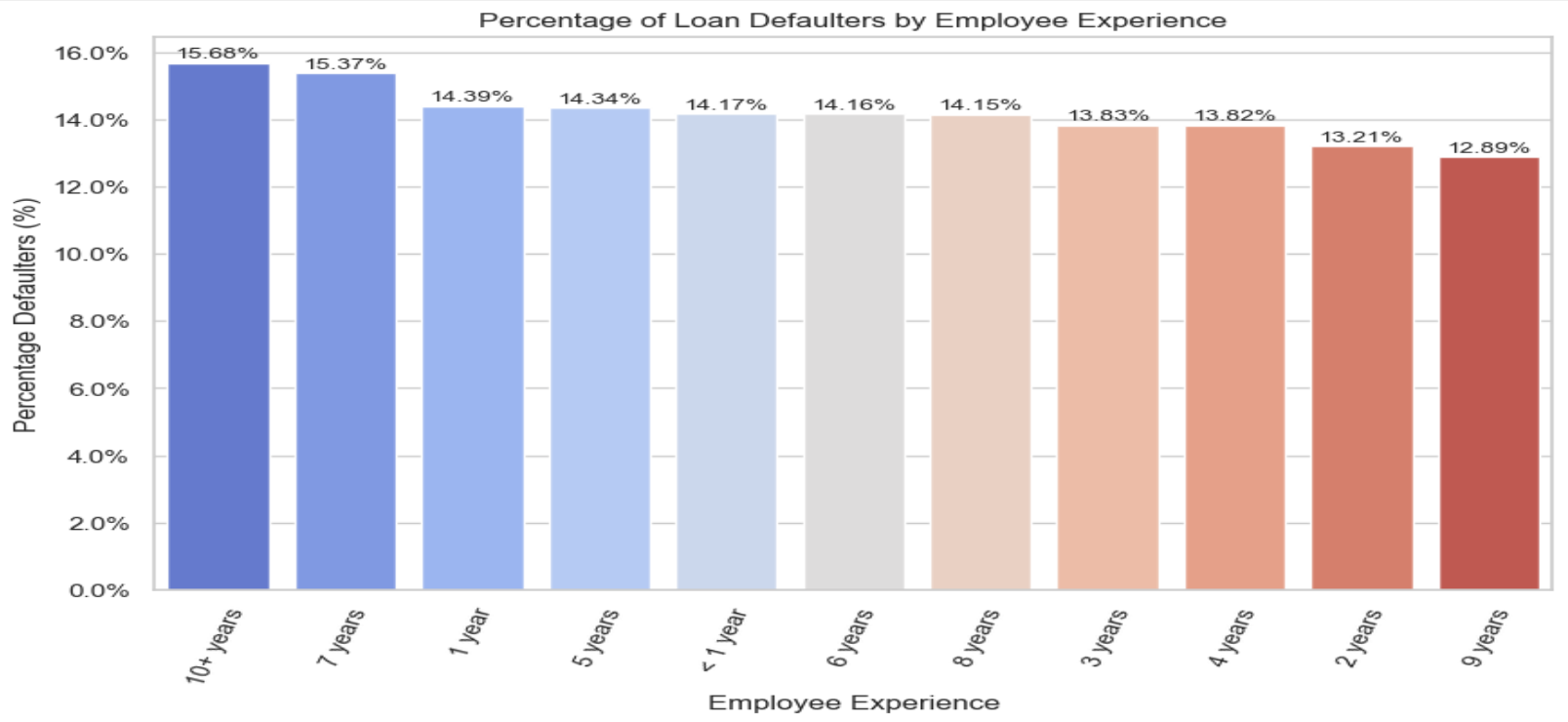
It's notable that Grade G exhibits the highest percentage of defaulters among all loan grades. This indicates that borrowers assigned to Grade G are more likely to default on their loans compared to other grade categories.

The Percentage of Defaulters by Purpose



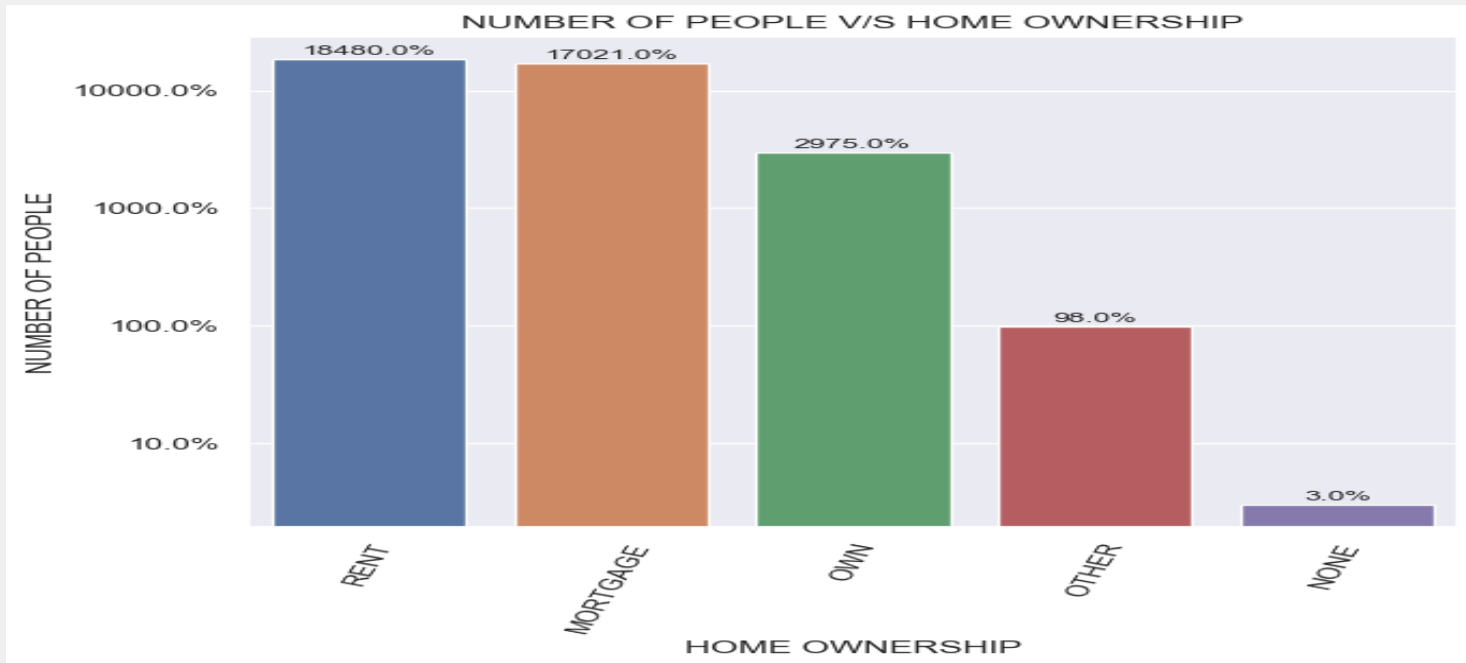
It's worth noting that the loan purpose 'small_business' exhibits the highest percentage of risk of defaulters among all loan purposes. This is followed by 'renewable_energy' and 'educational' purposes, indicating a higher likelihood of default in these categories

The Percentage of Defaulters by Employee Experience



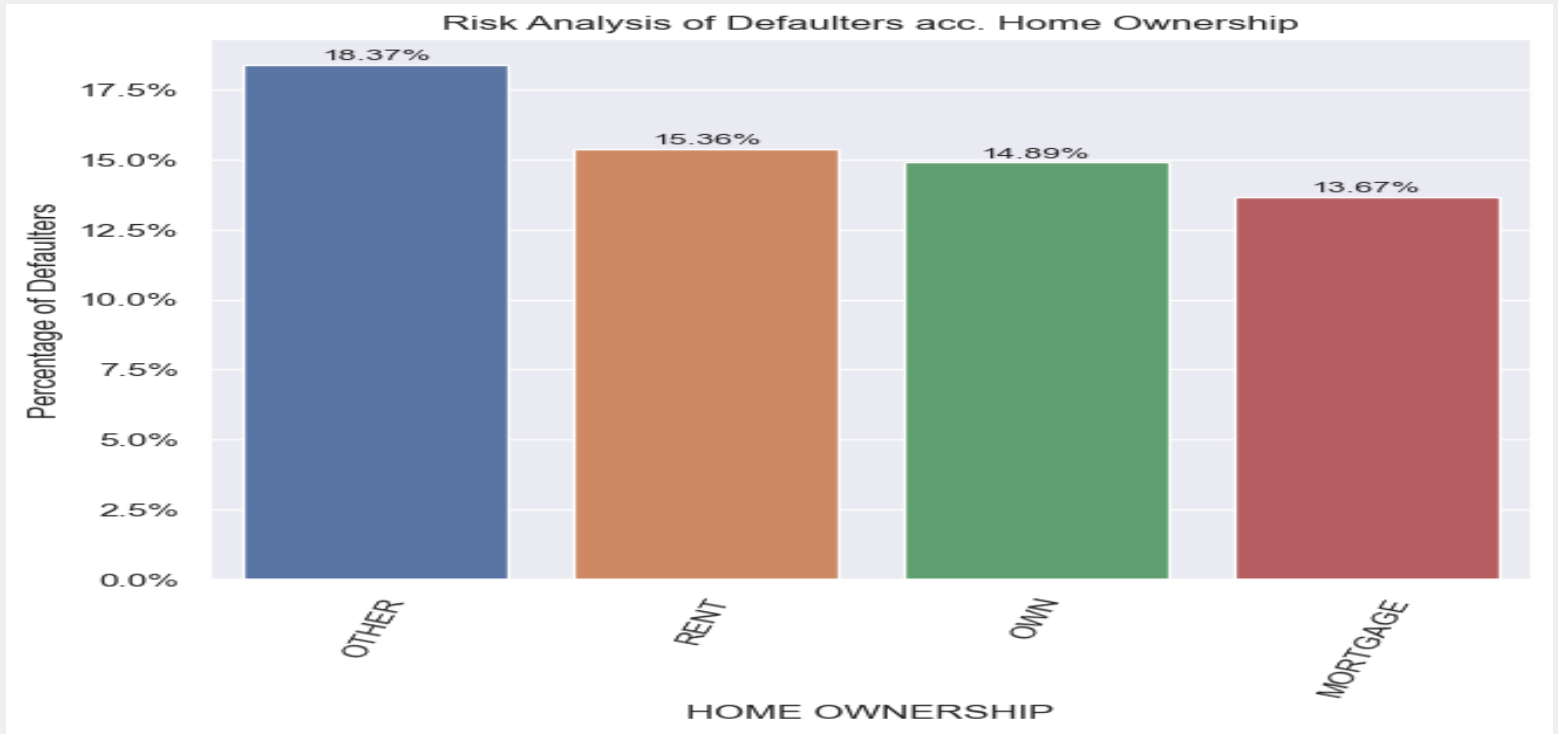
It's noteworthy that the highest percentage of risk of defaulters is associated with customers having 10+ years of experience. On the other hand, the lowest percentage of risk is observed among customers with 9 years of experience. Customers with other years of experience fall in between these extremes, indicating varying levels of risk.

Visualization about People's Home Ownership



The analysis reveals that the number of people who have home ownership categorized as 'Rent' is the highest among all the home ownership categories in the dataset.

The Percentage of Defaulters based on Home Ownership

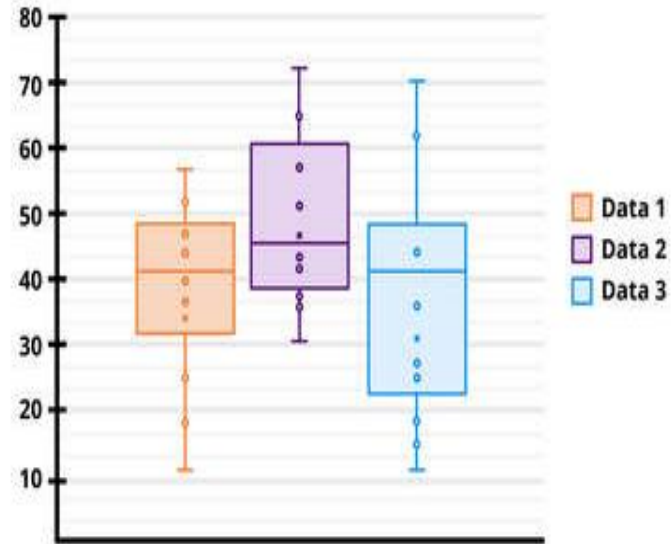


When considering the percentage of defaulters based on home ownership, the highest percentage is observed among those categorized as 'Other,' followed by 'Rent,' 'Own,' and 'Mortgage,' in descending order.

BIVARIATE ANALYSIS

Bivariate analysis refers to the analysis of two variables to determine relationships between them.

Bivariate analyses are conducted to determine whether a statistical association exists between two variables, the degree of association if one does exist, and whether one variable may be predicted from another.

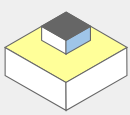
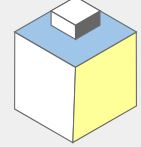


A Boxplot of Debt To Income versus Employee Experience

After conducting a bivariate analysis of debt to income (DTI) versus employee experience, a notable pattern emerges. Employees with 10+ years of experience exhibit a significantly wider range of DTI values, indicating a greater variability in their debt to income ratios. In contrast, employees with fewer years of experience generally fall within a more consistent DTI range.

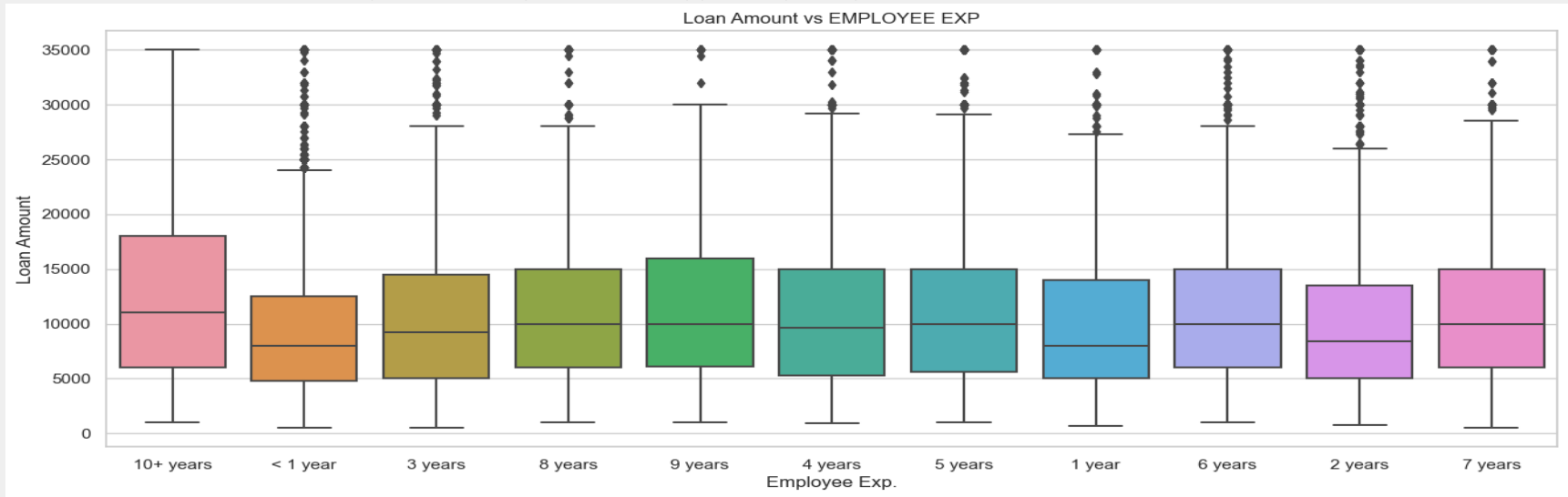
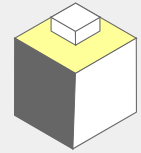


This suggests that employee experience, particularly in the '10+ years' category, may have a significant influence on DTI, potentially reflecting variations in financial circumstances and obligations.

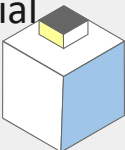
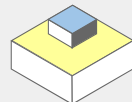


A Boxplot of Loan Amount versus Employee Experience

The bivariate analysis reveals a clear trend in the relationship between employee experience and loan amount. Borrowers with 10+ years of experience tend to request higher loan amounts, while those with less than 1 year of experience typically request lower loan amounts.

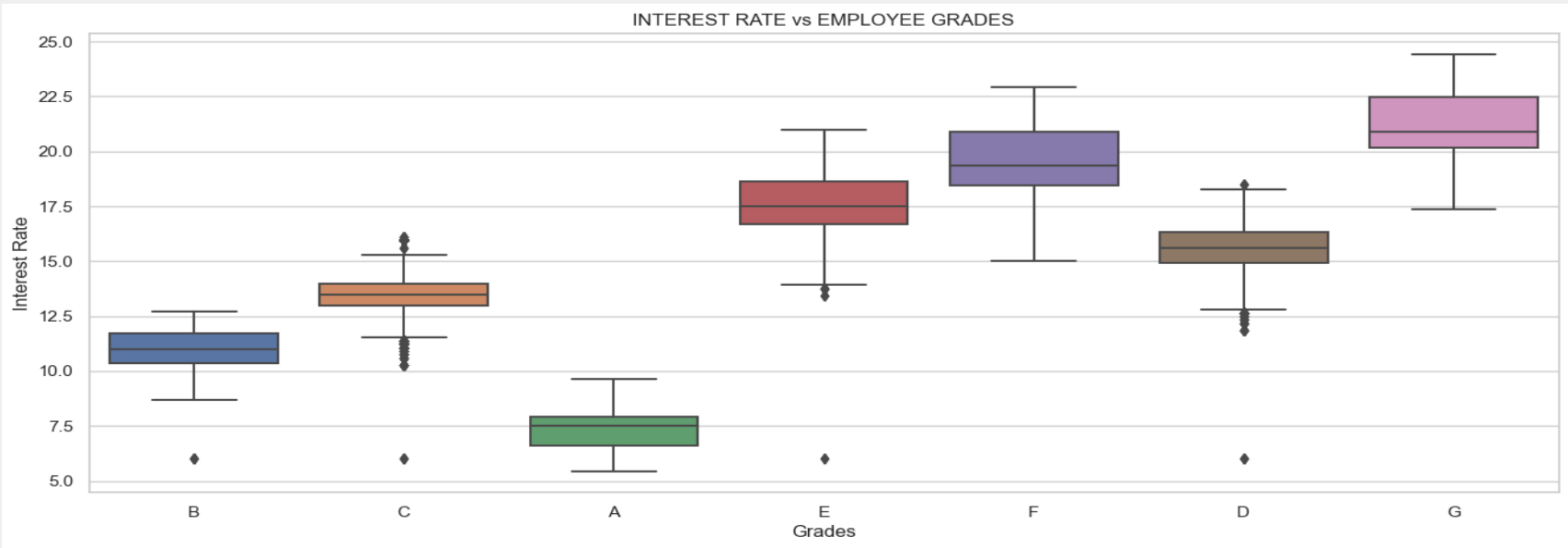


This suggests that employee experience plays a significant role in determining the loan amount requested, with more experienced individuals seeking larger loans, possibly due to their financial stability and creditworthiness.



A Boxplot of Interest Rate versus Employee Grades

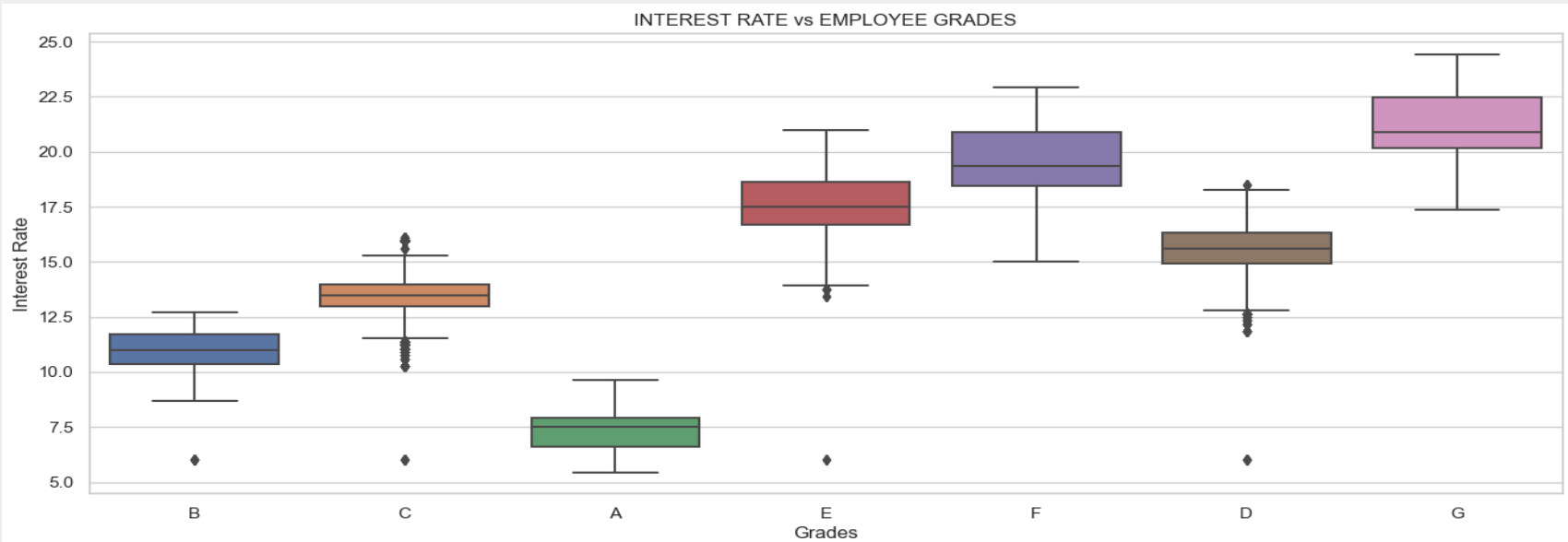
Based on the visual analysis of the boxplot, it becomes evident that interest rates tend to vary significantly across different loan grades. Specifically, loans categorized under 'A' grade exhibit the lowest interest rates, signifying a lower cost of borrowing. On the other hand, loans categorized under 'B' grade tend to have higher interest rates, suggesting a relatively higher cost of borrowing for borrowers in this category.



This visualization highlights the role of loan grades in determining interest rates, with lower-grade loans incurring higher interest costs.

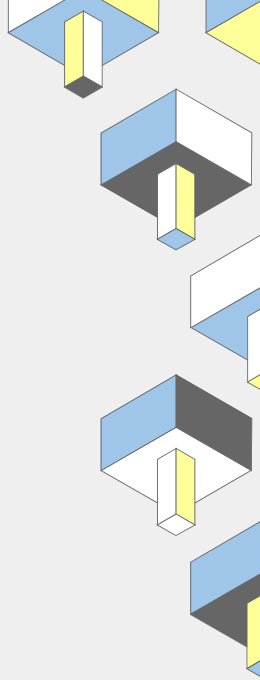
A Boxplot of Interest Rate versus Employee Grades

Based on the visual analysis of the boxplot, it becomes evident that interest rates tend to vary significantly across different loan grades. Specifically, loans categorized under 'A' grade exhibit the lowest interest rates, signifying a lower cost of borrowing. On the other hand, loans categorized under 'B' grade tend to have higher interest rates, suggesting a relatively higher cost of borrowing for borrowers in this category.

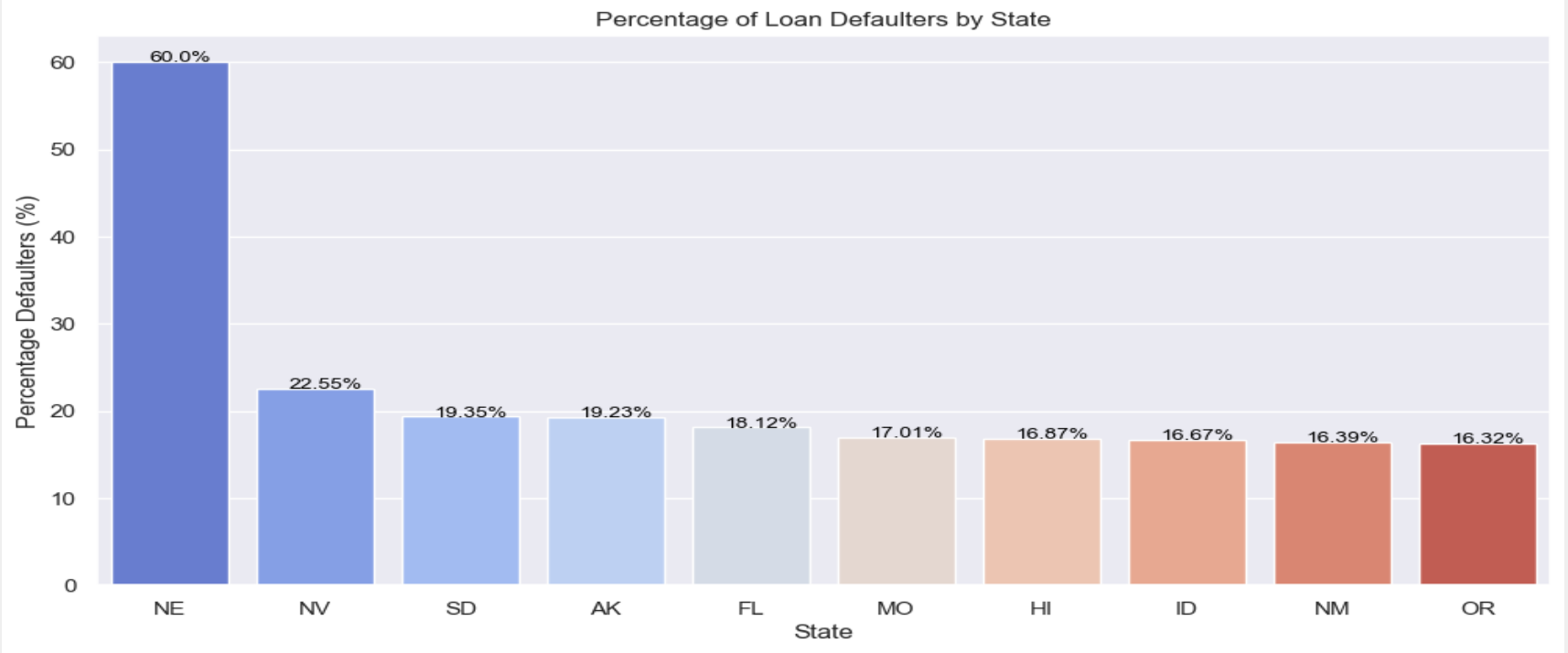


This visualization highlights the role of loan grades in determining interest rates, with lower-grade loans incurring higher interest costs.

FINAL ANALYSIS

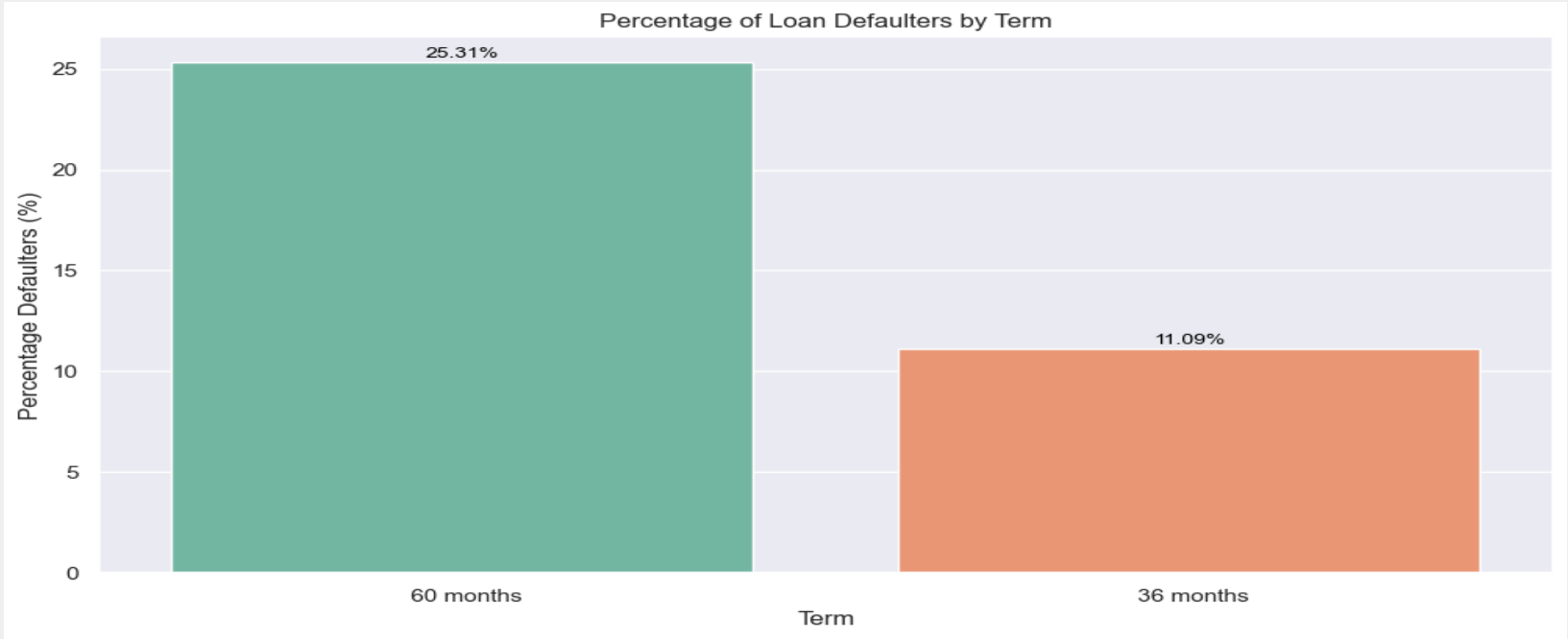


The Percentage of Defaulters by State



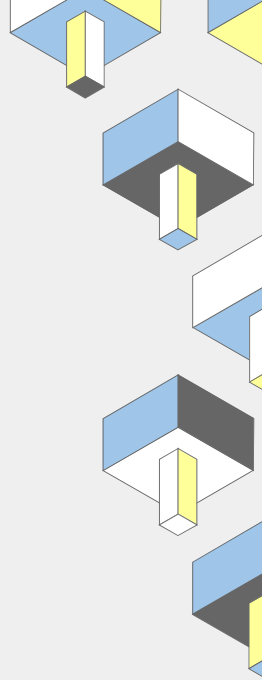
It's important to note that the highest percentage of defaulters is observed in the Northeast (address) region, accounting for 60%. This information can be crucial for making informed decisions and risk assessments in the lending business, as it highlights the region with the highest default risk.

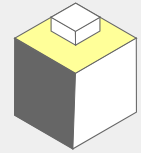
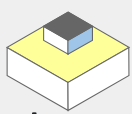
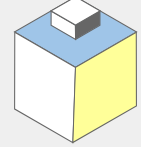
The Percentage of Defaulters by Term



Analysis of loan terms reveals that a higher percentage of defaulters, approximately 25.31%, are observed in the 60-month loan term compared to the 36-month loan term, which has a lower default rate of approximately 11.09%. This suggests that longer loan terms may carry a higher risk of default.

CONCLUSION

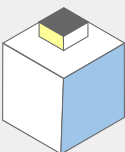
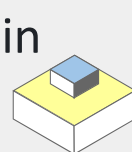


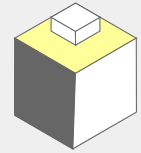
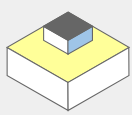
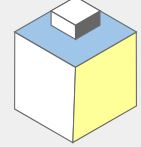


- It's notable that Grade G exhibits the highest percentage of defaulters among all loan grades. This indicates that borrowers assigned to Grade G are more likely to default on their loans compared to other grade categories.

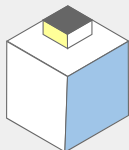
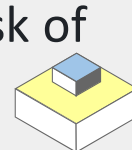
- It's worth noting that the loan purpose 'small_business' exhibits the highest percentage of risk of defaulters among all loan purposes. This is followed by 'renewable_energy' and 'educational' purposes, indicating a higher likelihood of default in these categories.

- It's noteworthy that the highest percentage of risk of defaulters is associated with customers having 10+ years of experience. On the other hand, the lowest percentage of risk is observed among customers with 9 years of experience. Customers with other years of experience fall in between these extremes, indicating varying levels of risk.





- When considering the percentage of defaulters based on home ownership, the highest percentage is observed among those categorized as 'Other,' followed by 'Rent,' 'Own,' and 'Mortgage,' in descending order.
- It's important to note that the highest percentage of defaulters is observed in the Northeast (address) region, accounting for 60%. This information can be crucial for making informed decisions and risk assessments in the lending business, as it highlights the region with the highest default risk.
- Analysis of loan terms reveals that a higher percentage of defaulters, approximately 25.31%, are observed in the 60-month loan term compared to the 36-month loan term, which has a lower default rate of approximately 11.09%. This suggests that longer loan terms may carry a higher risk of default.



THANK YOU

