

Lending Club Case Study

- Prepared by
- Rahul Gupta, Kiran Kumar Rao K
- Submitted Date: 29th Nov 2024
- GitHub repository link -
<https://github.com/KiranKRao/LendingClubCaseStudy>

Objectives

- Define the issue and the method of analysis.
- Describe the findings from both univariate and bivariate analyses.
- Incorporate visual aids and summarize the key results in the presentation.
- Identify the key factors contributing to loan defaults and provide recommendations.
- Summarize the overall conclusions.

Project Overview

The project focuses on using data analysis to identify high-risk loan applicants in an online lending marketplace.

1. It examines various factors, including
 - a) Demographics
 - b) Credit history
 - c) Loan characteristics
 - d) Economic indicators
2. Exploratory Data Analysis (EDA) will be utilized to uncover patterns related to loan defaults.
3. Insights gained will help the company to:
 - a) Reduce credit losses
 - b) Refine lending criteria
 - c) Improve the overall loan approval process
4. The initiative aims to enhance risk management practices. Ultimately, it seeks to create a more profitable lending environment.

Problem Statement

High Default Rates: among borrowers labeled as 'charged-off' result in significant financial losses for lenders. It is crucial to understand the characteristics and behaviors that indicate a higher likelihood of default.

Inefficient Loan Approval Processes: The current lending practices may not adequately consider the risk factors associated with borrowers, leading to misguided approvals and increased exposure to defaults.

Lack of Data-Driven Insights: Without thorough data analysis, it is challenging to establish effective criteria for evaluating borrower risk. This lack of insight hampers the company's ability to optimize its loan offerings and customer segmentation.

Impact on Profitability: Credit losses directly impact the bottom line. Reducing the number of high-risk loans can enhance overall profitability and sustainability for the business.

Project Objective: The project aims to enhance risk assessment strategies, reduce credit loss, and ultimately improve the company's financial performance and customer experience.

Analysis Strategy

- Outline the problem and the analytical approach taken.
- Present the results from both single-variable and two-variable analyses.
- Use visual representations and highlight the main outcomes in the presentation.
- Determine the primary elements leading to loan defaults and suggest improvements.
- Conclude with a summary of the findings.

Comprehending the Data

Loan Amount and Funding: Fields like loan_amnt, funded_amnt, and funded_amnt_inv offer details on the requested loan amount, the amount that was funded, and the actual investment made.

Loan Terms and Interest Rate: The term indicates the length of the loan (either 36 or 60 months), while nt_rate represents the interest rate applied to the loan (for example, 10.65% or 15.27%).

Borrower Risk Assessment: The grade and sub_grade evaluate the risk level of the borrower (for instance, grade B and sub-grade B2).

Payment Information: The installment indicates the monthly payment for each loan, and last_pymnt_amnt shows the amount paid in the most recent payment cycle.

Credit History Insights: Fields such as last_credit_pull_d and pub_rec_bankruptcies provide information about the borrower's credit history, including the date of the last credit inquiry and any bankruptcy records.

Data Cleansing and Conversion

Note: Prior to the analysis, the Loan Table contained 39,717 rows and 111 columns.

Data Cleansing:

1. Conducted a data analysis.
2. Eliminated unnecessary columns, including headers and footers.
3. Ensured there were no duplicate entries in the table.
4. Removed all NULL values.
5. Deleted the column with a unique value of 1.
6. Noted that fields related to loan applications, such as id, member_id, and url, vary for each application, and the #zip code field lacks complete data.

Data Conversion:

1. Created derived columns for 'issue_year' and 'issue_month' from 'issue_d' for future analysis.
2. Developed derived columns such as 'loan_amnt_b', 'annual_inc_b', 'int_rate_b', and 'dti_b' to categorize continuous data for improved analysis.
3. Removed the percentage sign from the interest column.
4. Eliminated text from the term column.
5. Excluded data where the loan status is marked as Current.
6. Converted the loan amount and funded amount to float data type.

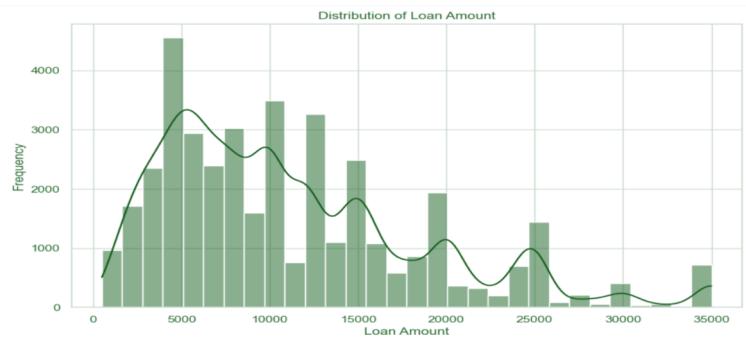
Univariate, Bivariate analysis

Univariate Analysis: Analyze distributions of key numeric variables like loan_amnt, int_rate, etc. Explore categorical variables like grade

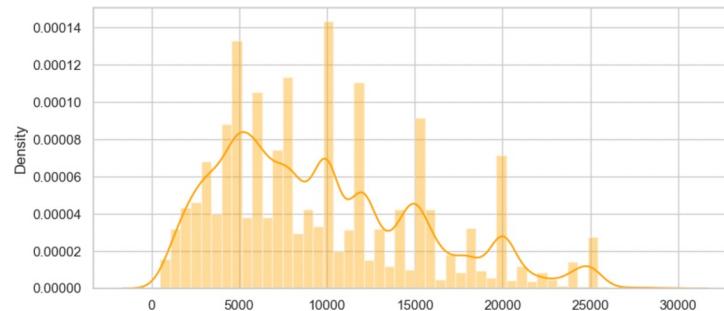
Bivariate Analysis: Correlation between loan_amnt and int_rate. Relationship between grade and loan_amnt. Analyze the impact of term on int_rate.

Co-Relations- Discussed About various Co-relations.

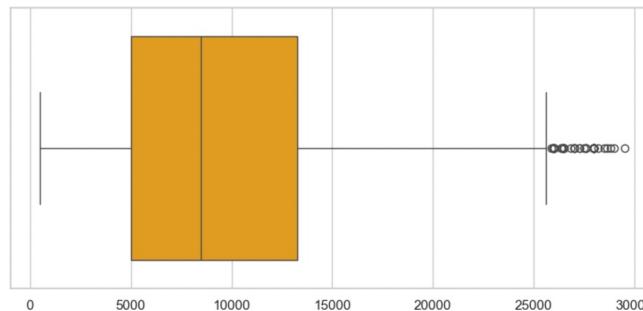
Univariate Analysis: Loan Amount



Loan Amount Distribution- Most loans are concentrated around smaller values, with a few higher amounts

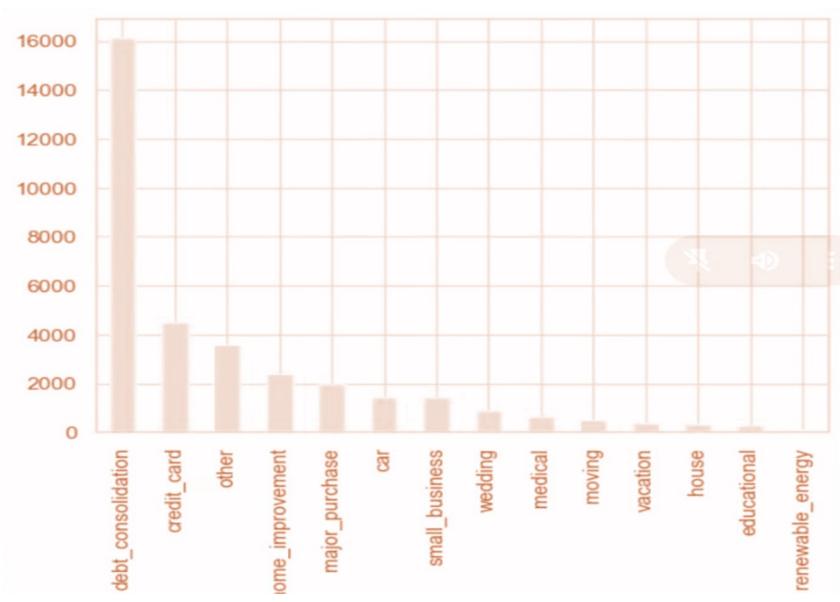
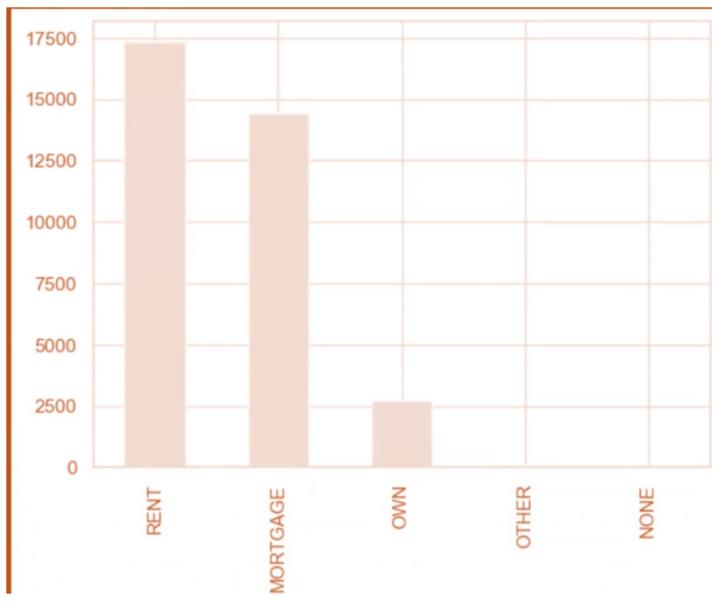


Most of the loan amount distributed between ~5k to ~13K



Highest loan amount is applied is ~29K

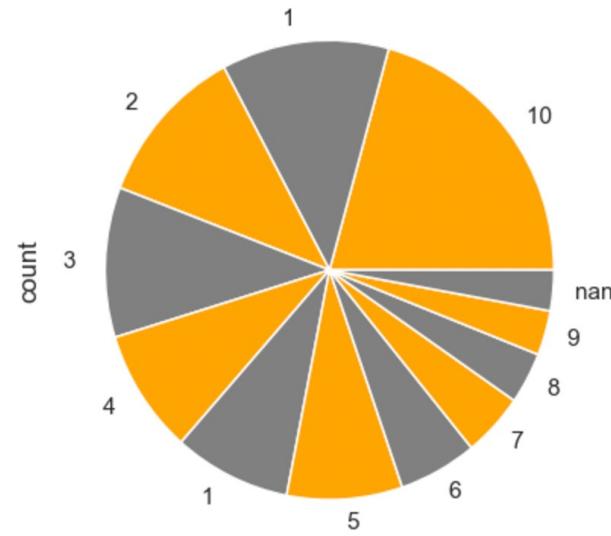
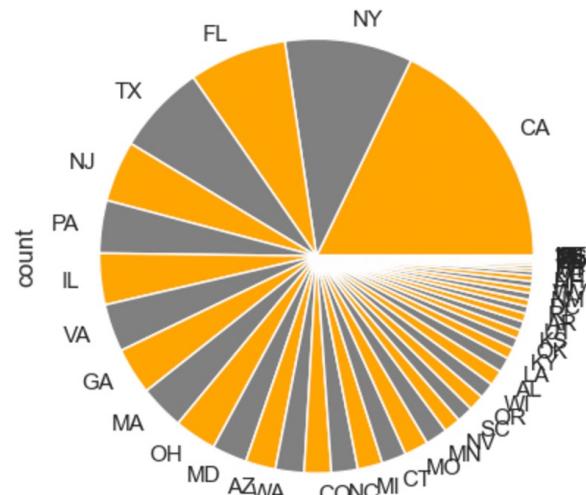
Univariate Analysis: Categorical Variable Analysis- 1



Categorical Variable Analysis-

- Most of the loan applicants are for Debt consolidation
- Most of the loan applicants are staying in Rental house followed by mortgage

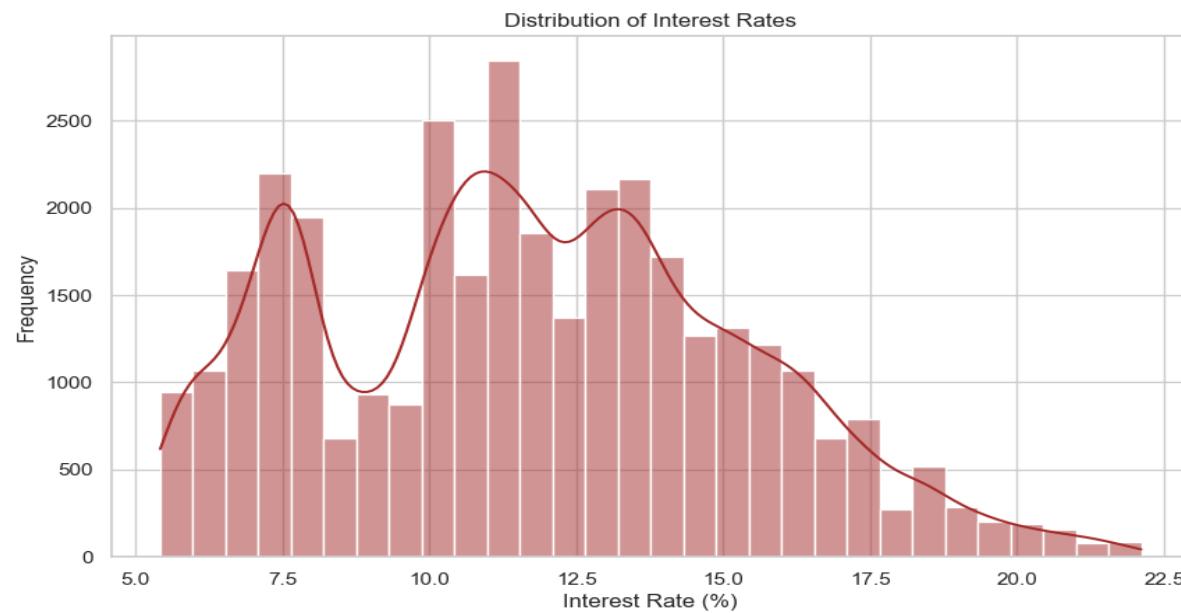
Univariate Analysis: Categorical Variable Analysis- 2



Categorical Variable Analysis- 2

- Most of the loan applicants are based out of California(CA)
- Maximum loan applicants are having 10+ year of experience

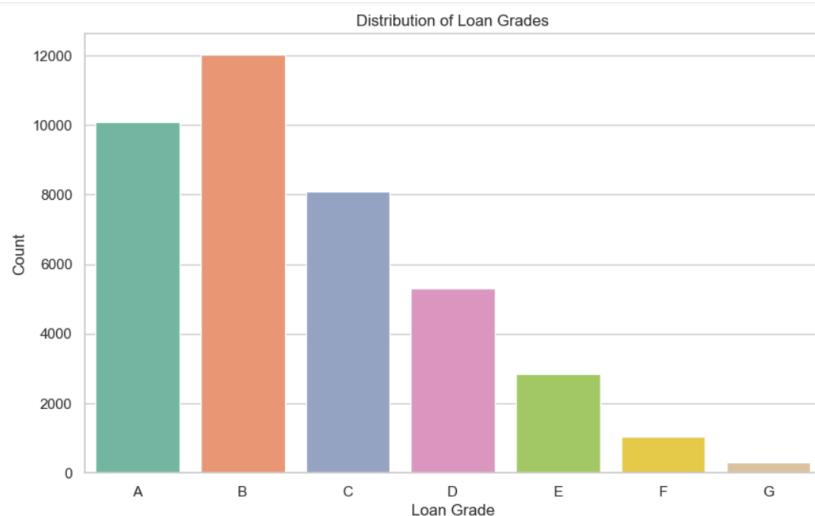
Univariate Analysis: Interest Rate Distribution



Interest Rate Distribution-

- Interest rates are generally skewed towards lower percentages, with a significant number of loans between 10% and 20%.
- Rate of Interest for most of the applicants is between ~8%-~14%

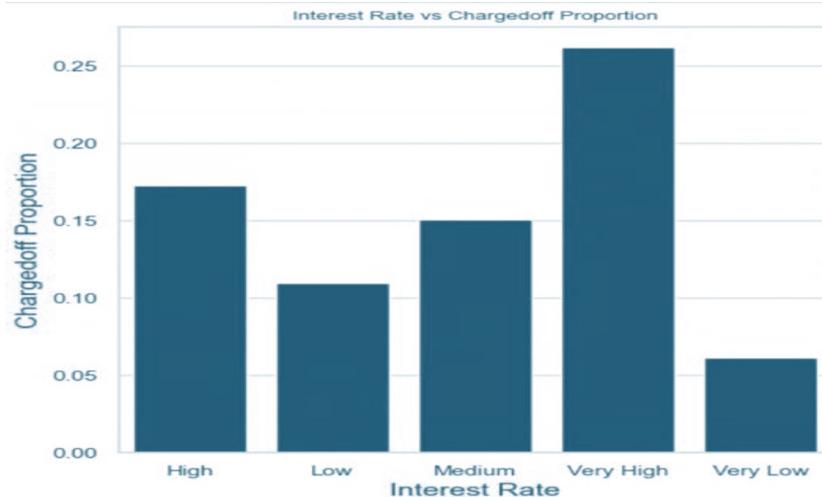
Univariate Analysis: Interest Rate Distribution



- **Loan Grade Distribution:** Grade B has
 - the highest number of loans, followed by
 - Grade C. Grades A and F have fewer loans,
 - suggesting a concentration around medium-risk borrowers.

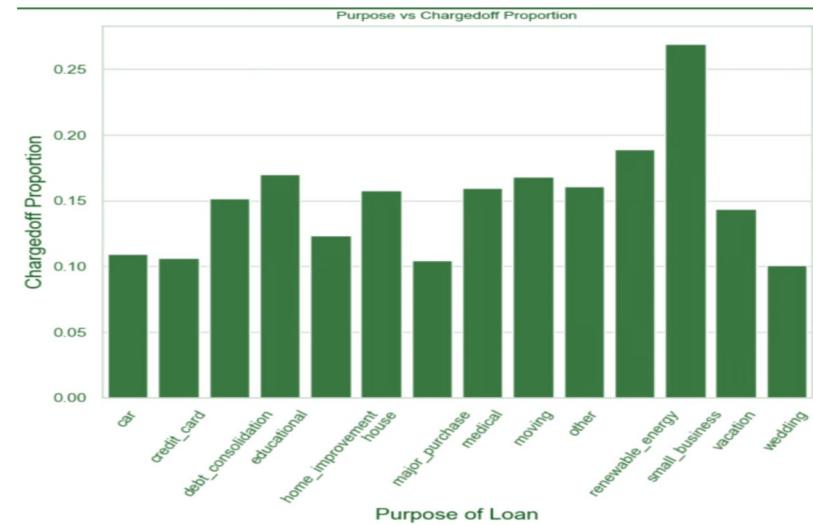
Bivariate Analysis

Comparison of Interest Rate against Charged-off



- Loan applicants are having 'very high' interest are more chances of maximum charged off.
- Loan applicants are having 'very low' interest are less chances of charged-off.

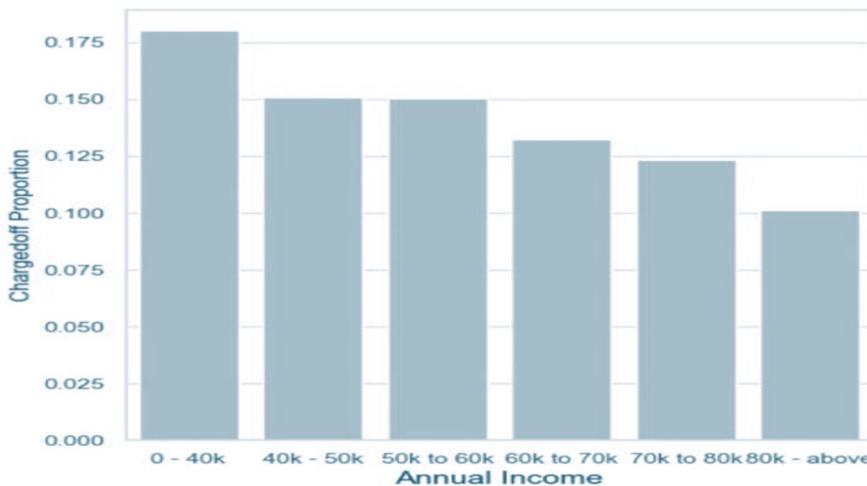
Comparison of Purpose of Loans against Charged-off



- Applicants who applied for 'Renewable Energy' or 'small business' are having more chances to be charged-off.

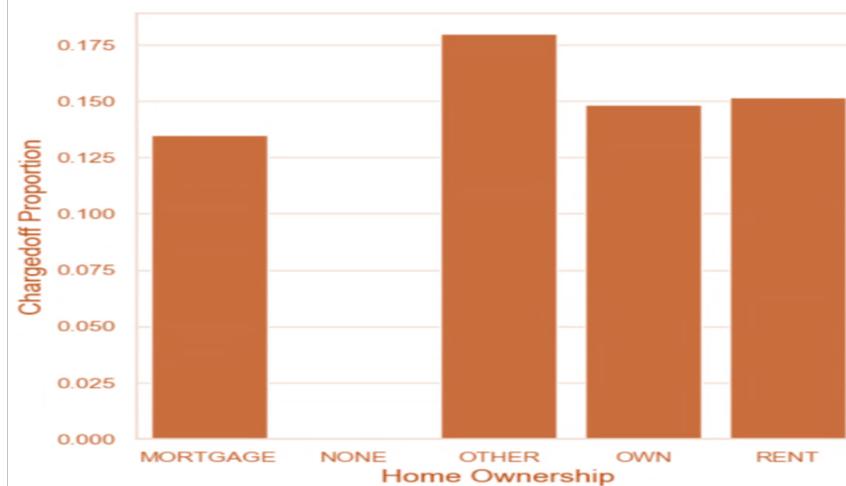
Bivariate Analysis

Comparison of Annual Income against Charged-off



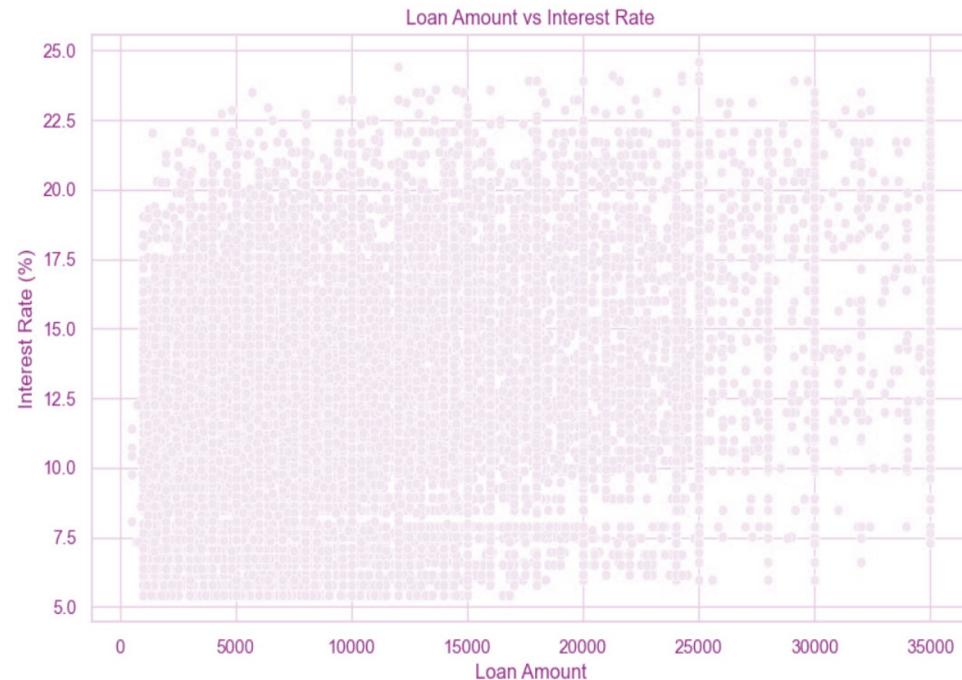
- Loan applicants are having salary of ~0-20K are more likely to be charged off.
- Loan applicants are having salary of 80K+ are less likely to be charged off.

Comparison of Annual Income against Charged-off



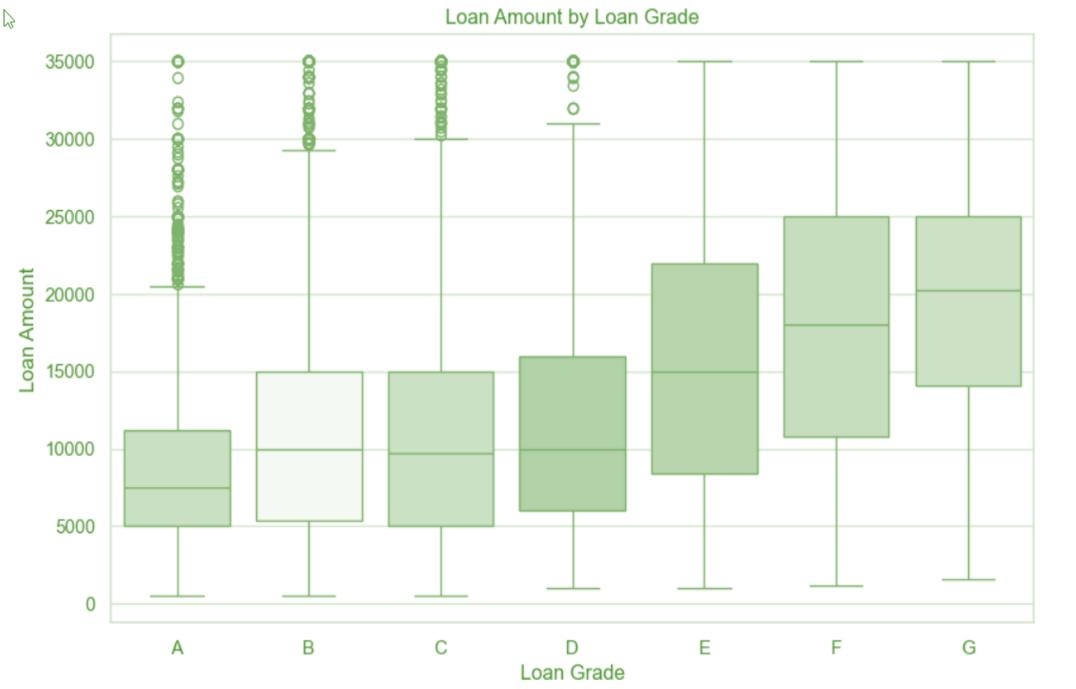
- Applicants who belongs to either category or does not having a home more likely to be charged-off.

Bivariate Analysis



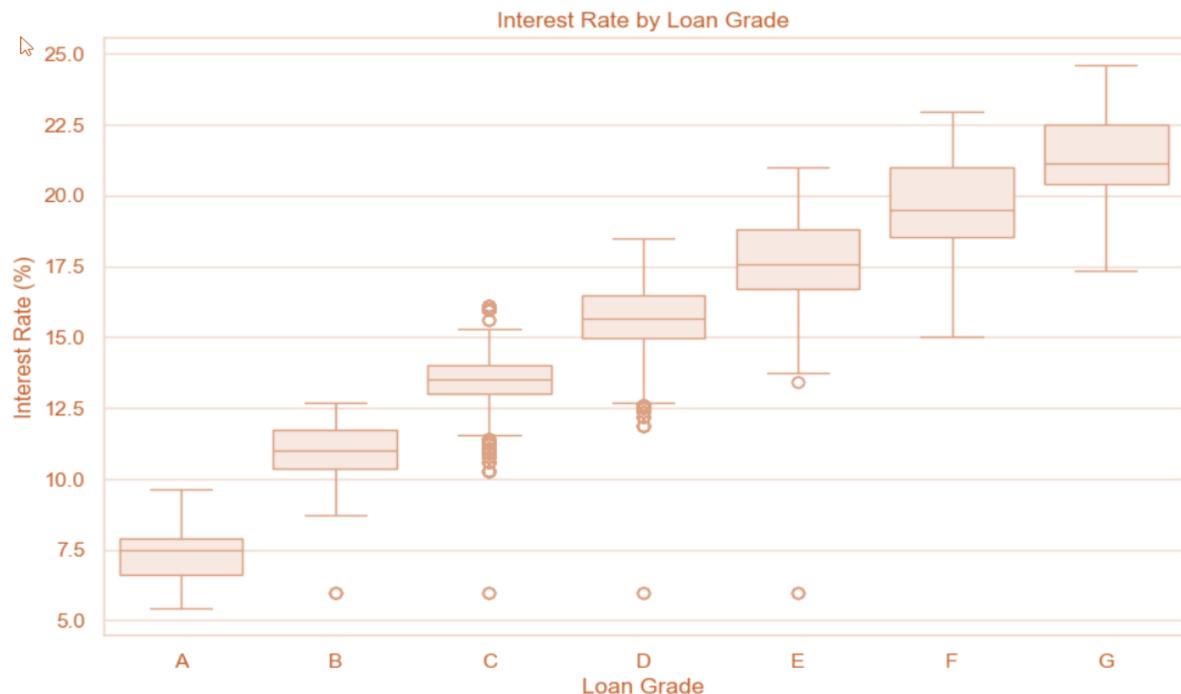
Loan Amount vs Interest Rate: There's no clear linear relationship, but higher loan amounts generally tend to have slightly lower interest rates.

Bivariate Analysis



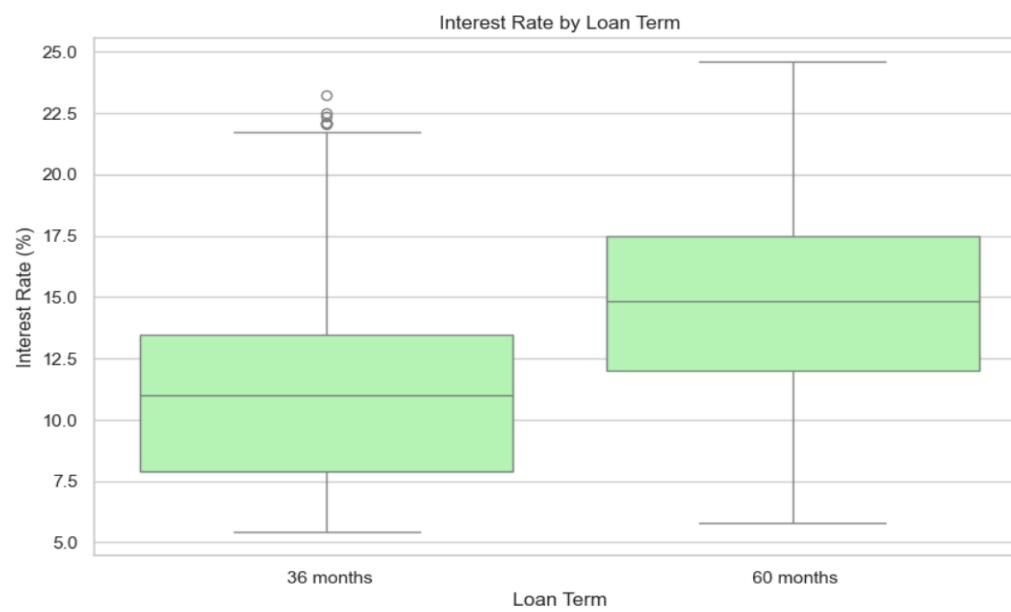
Loan Amount by Grade: Higher-grade loans (A, B) have relatively smaller loan amounts compared to lower grades (C, D, E), indicating that higher-risk borrowers tend to request larger loans.

Bivariate Analysis



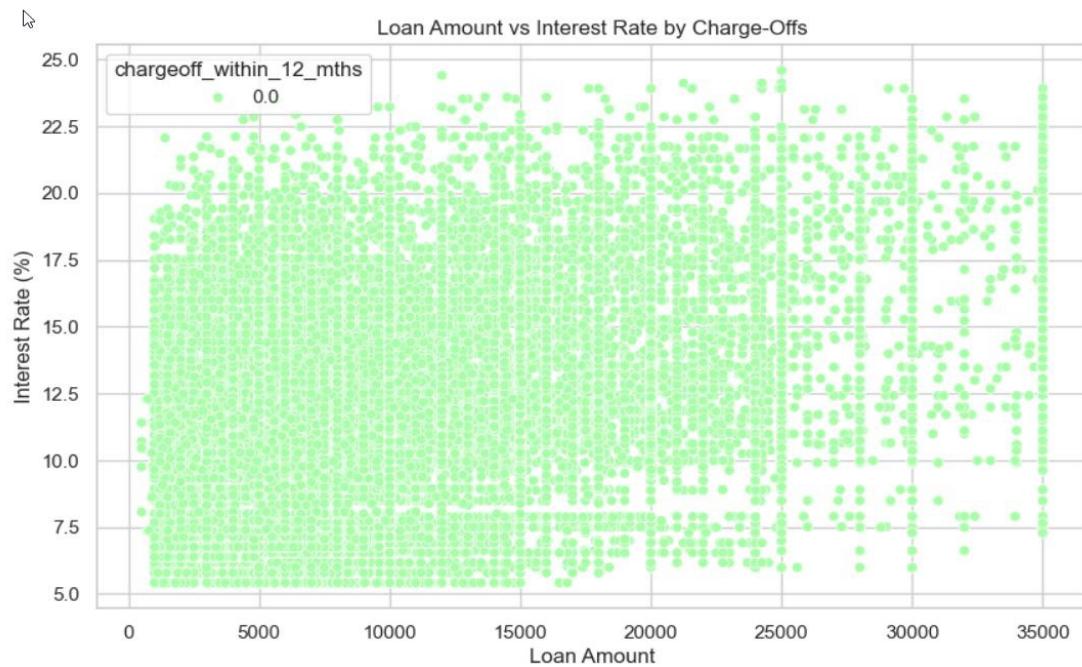
Interest Rate by Grade: Risk heatmap based on grade and interest rate. Higher grades (A, B) are typically less risky, lower grades (D, E) are higher risk. Higher interest rates generally indicate higher risk loans

Bivariate Analysis



Interest Rate by Loan Term: 60-month loans generally have higher interest rates compared to 36-month loans, reflecting the increased risk for longer-term loans.

Further Analysis



Risk area based on funded amount and default likelihood (charge-offs)

We are using 'loan_amnt' and 'chargeoff_within_12_mths' to highlight potential risk areas

Conclusion

Loan applicants are having salary of ~0-20K are more likely to be charged off.

- Applicants with salaries between \$0-\$20K are more likely to experience loan charge-offs.
- Applicants earning \$80K or more are less likely to face charge-offs.
- Applicants without a home or those in other categories are more prone to being charged off.
- Loans with very high interest rates have a higher probability of charge-offs, while loans with very low interest rates are less likely to result in charge-offs.
- Applicants applying for loans for "Renewable Energy" or "Small Business" purposes have a higher chance of charge-offs.
- Larger loan amounts tend to have slightly lower interest rates, though there's no clear linear relationship.
- Loans with higher grades (A, B) tend to have smaller loan amounts, whereas lower-grade loans (C, D, E) are often associated with larger loan amounts, suggesting that higher-risk borrowers request bigger loans.
- A risk heatmap has been created based on loan grade and interest rate. 60-month loans generally carry higher interest rates than 36-month loans, reflecting the increased risk of longer-term loans.
- A risk assessment has been conducted based on the funded loan amount and the likelihood of charge-offs