Lending Club Case Study

Prepared by
Patitapaban Das, Deepak Rana
Submitted Date: 22nd September 2024

Objectives

1. Problem statement and the analysis approach
2. Explain the results of univariate, bivariate analysis
3. Include visualizations and summaries the most important results in the presentation
4. Driving factors (or driver variables) behind loan default and Recomendation
5. Conclusion

Brief description about the project.

This project aims to leverage data analysis to identify risky loan applicants within an online loan marketplace.

By examining various factors—including demographics, credit history, loan characteristics, and economic indicators

We will utilize Exploratory Data Analysis (EDA) to reveal patterns associated with loan defaults.

The insights gained will enable the company to reduce credit loss, refine lending criteria, and improve the overall loan approval process. Ultimately, this initiative will enhance risk management practices and foster a more profitable lending environment.

Problem statement

1. **High Default Rates:** Many borrowers labelled as 'charged-off' lead to substantial financial losses for lenders. Understanding the characteristics and behaviours that indicate a higher likelihood of default is essential.
2. **Inefficient Loan Approval Processes:** Current lending practices may not adequately account for the risk factors associated with borrowers, leading to misguided approvals and increased exposure to defaults.
3. **Lack of Data-Driven Insights:** Without thorough data analysis, it is challenging to establish effective criteria for evaluating borrower risk. This lack of insight can hinder the company's ability to optimize its loan offerings and customer segmentation.
4. **Impact on Profitability:** Credit losses directly affect the bottom line. Reducing the number of high-risk loans can improve overall profitability and sustainability for the business.

By solving this problem, the project aims to enhance risk assessment strategies, reduce credit loss, and ultimately improve the company's financial performance and customer experience

Analysis approach

Analysing Loan/Credit Data

Data Cleansing and Normalization

EDA- Univariate, Bivariate analysis

Data Visualization

Data Understanding

Loan Amount & Funding: Columns such as loan_amnt, funded_amnt, and funded_amnt_inv provide information on the requested loan amount, the amount funded, and the amount actually invested.

Loan Terms & Rate of Interest: Term refers to the duration of the loan (36 or 60 months).
nt_rate indicates the interest rate charged on the loan (e.g., 10.65%, 15.27%).

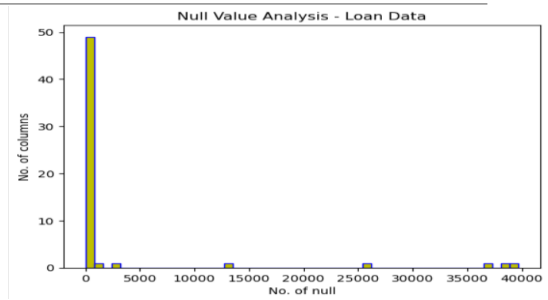
Borrower Risk: grade and sub_grade assess the risk associated with the borrower (e.g., grade B, sub-grade B2).

Payment Details: installment shows the monthly payment amount for each loan. last_pymnt_amnt refers to the amount paid during the last payment cycle.

Credit History: Columns like last_credit_pull_d and pub_rec_bankruptcies reflect aspects of the borrower's credit history, such as the date of the last credit pull and records of bankruptcies.

Data Cleansing and Conversion

Before Analysis, there 39717 row and 111 columns in the Loan Table.



Data Cleansing

1. Analysis of Data
2. Removing the unnecessary Columns(Header and Footer)
3. Make sure no duplicate values are in the table
4. Removed all NULL Values
5. Removed the column where unique value is 1.
6. Fields after loan application approved and fields like id, member_id & url are different for each application, #zip code doesn't have full data

Final analysis performed on 39717 rows and 51 columns

Data Conversion

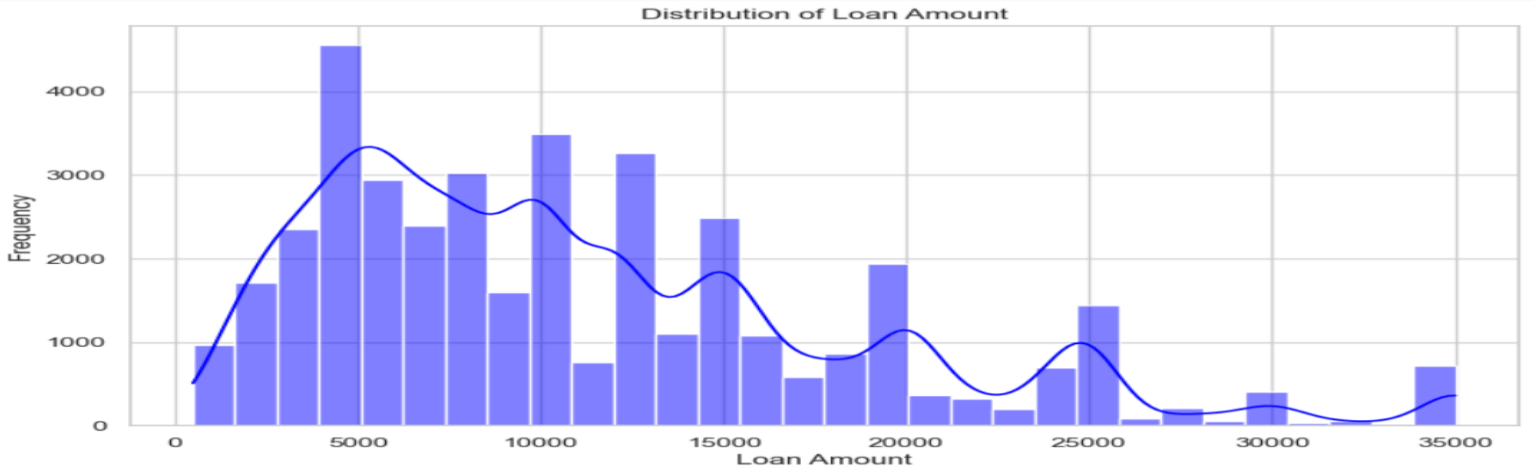
1. Creating a derived columns for ‘issue_year’ and ‘issue_month ’ from ‘issue_d’ which will be using for further analysis.
2. ‘loan_amnt_b’, ‘annual_inc_b’, ‘int_rate_b, and ‘dti_b’ derived columns(multiple bucket kind of data from continuous data) has been created for better analysis.
3. Removed % sign from interest column
4. Removed String from term column
5. Drop the data where loan status = Current
6. Converted loan amount and funded amount to float

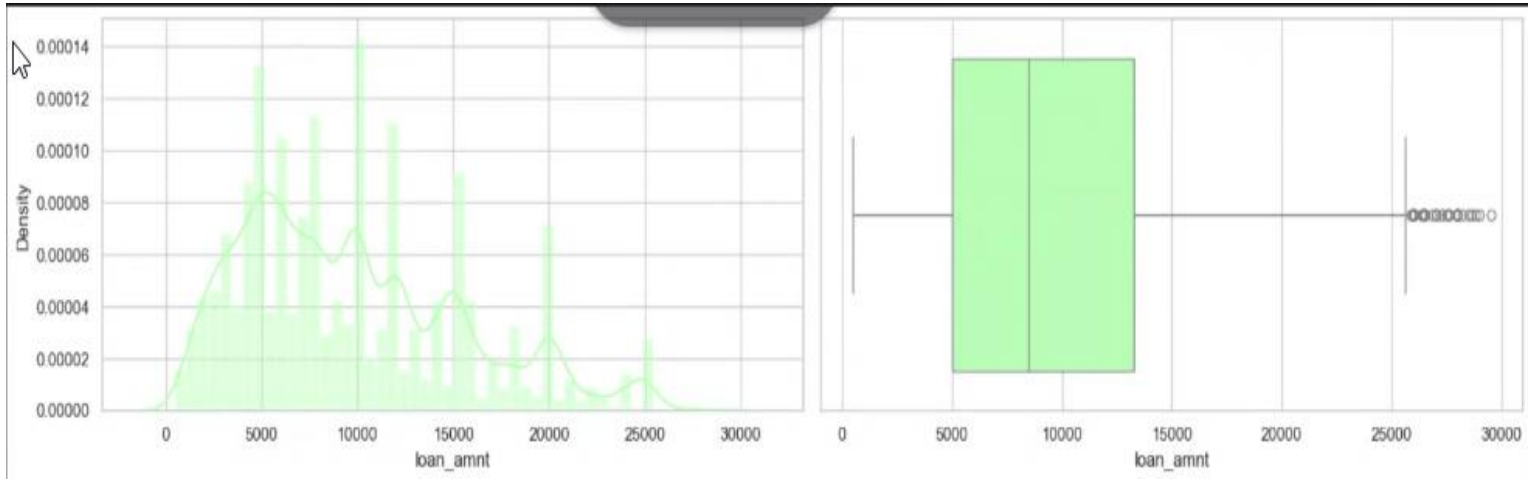
Univariate, Bivariate analysis

- **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





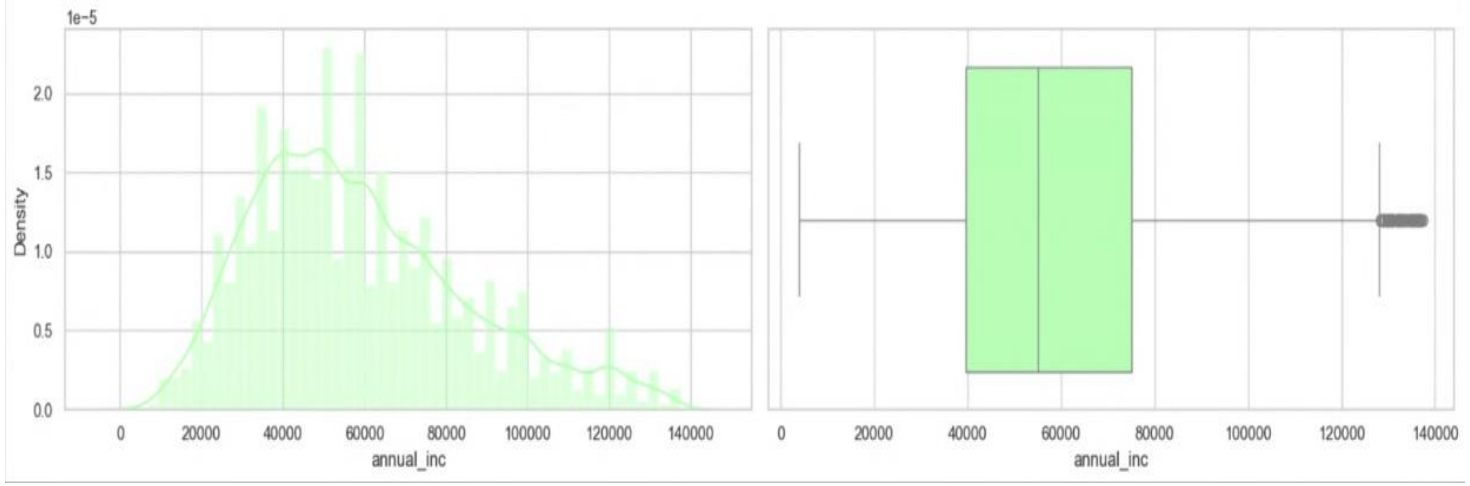
Loan Amount Distribution-

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: Annual income

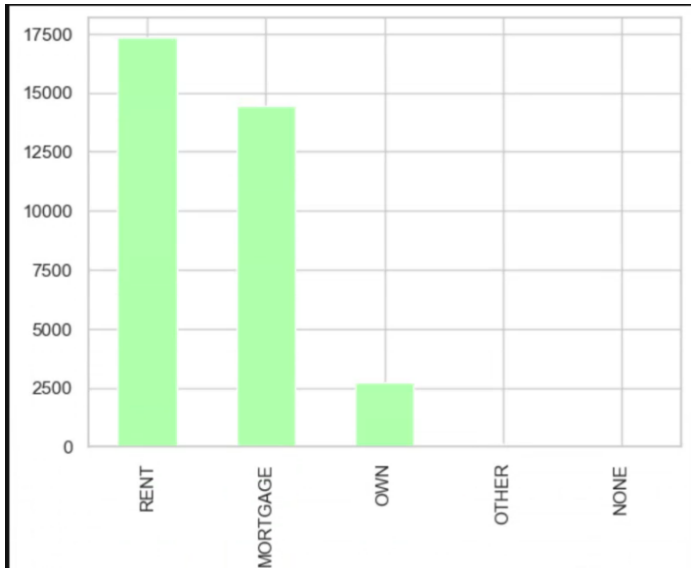


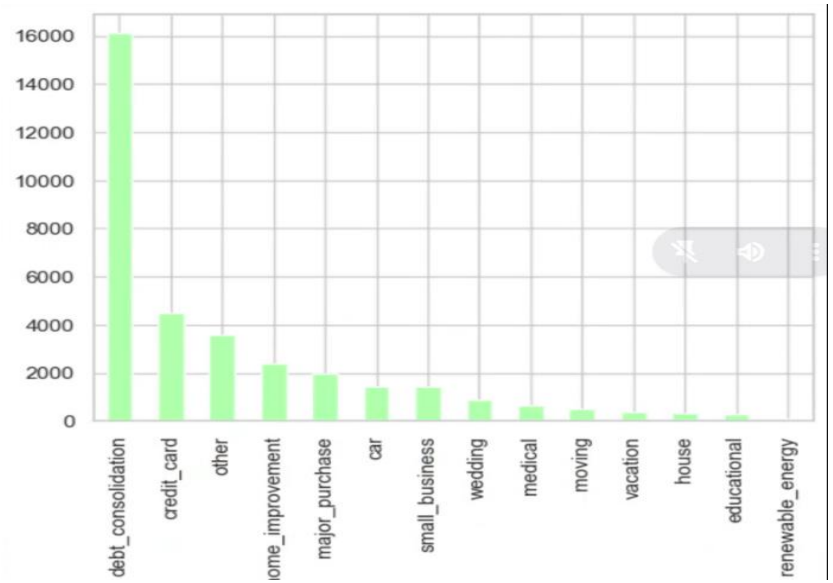
Annual Income Distribution

- Annual income of Loan Applicants are between ~40K to ~75K
- Average Annual Income is ~59K

.

Univariate Analysis: Categorical Variable Analysis- 1



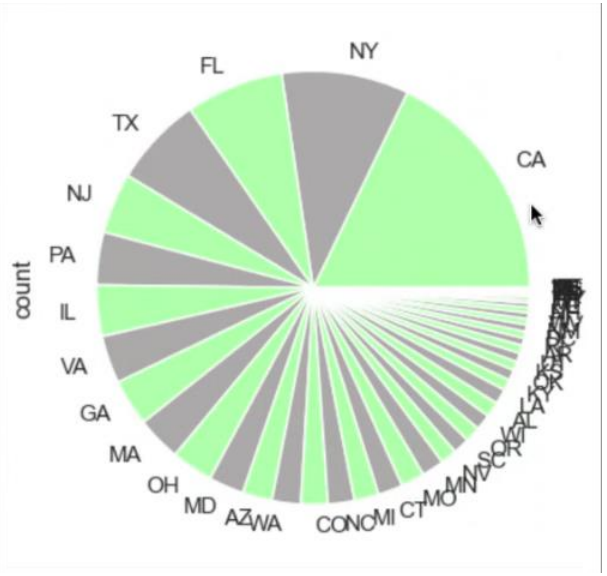


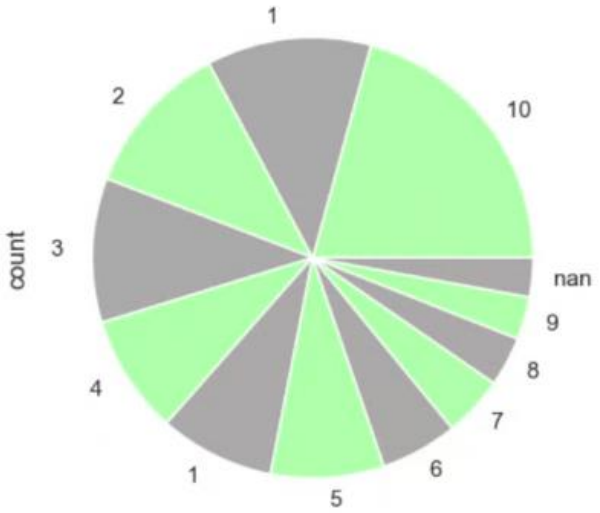
Categorical Variable Analysis-

- 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



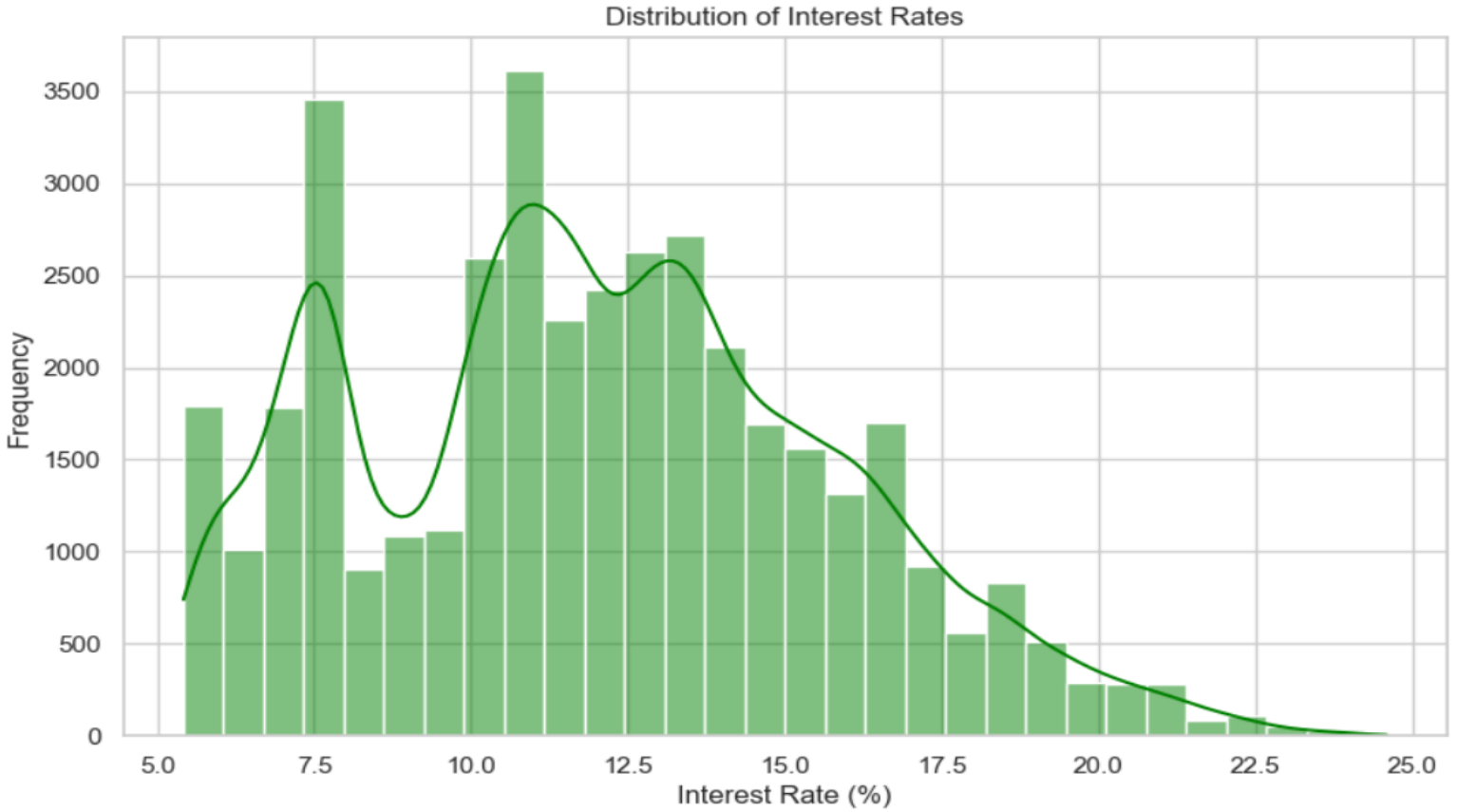


Categorical Variable Analysis- 2

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

.

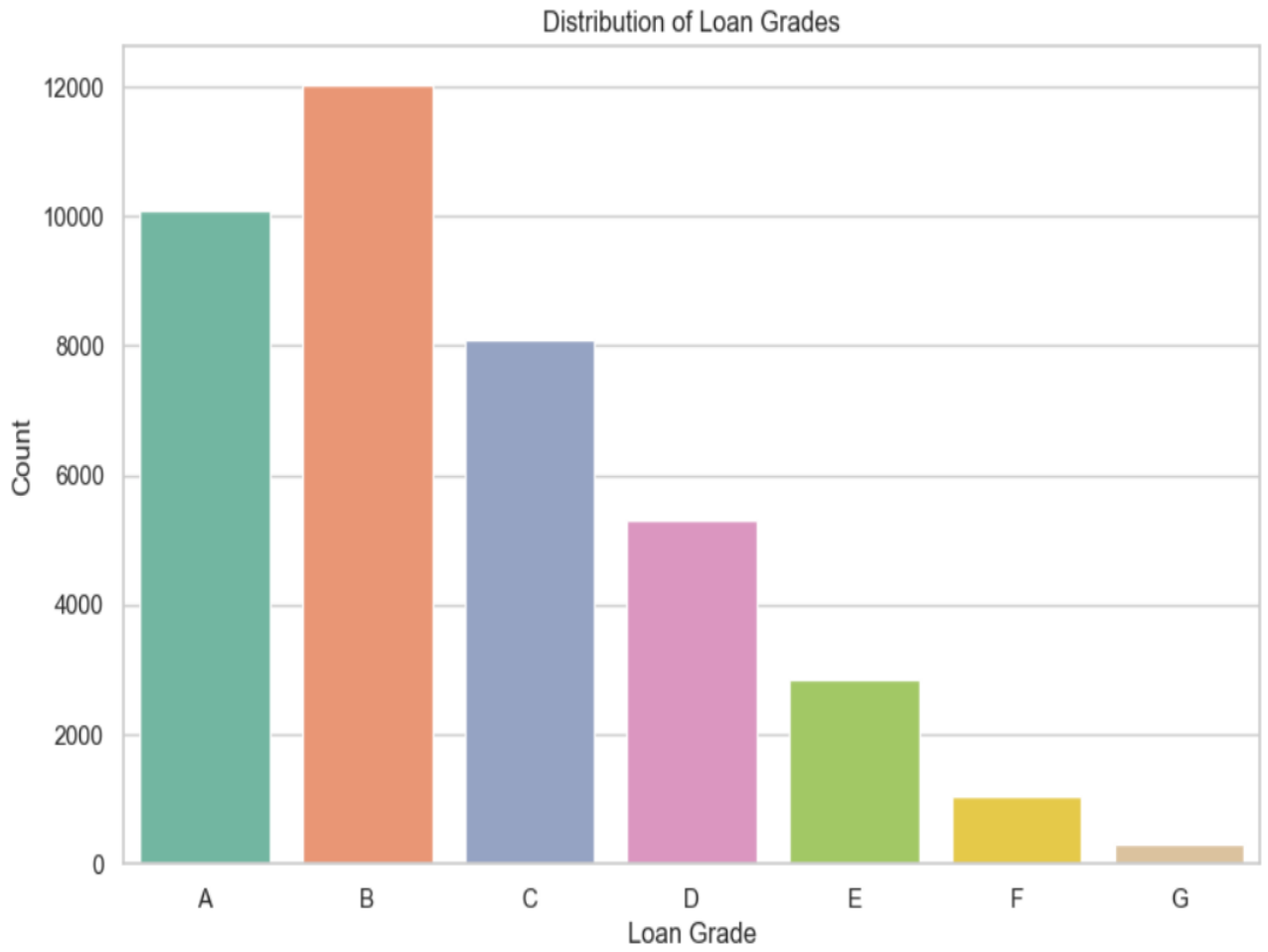
Univariate Analysis:



Interest Rate Distribution-

- Interest 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%

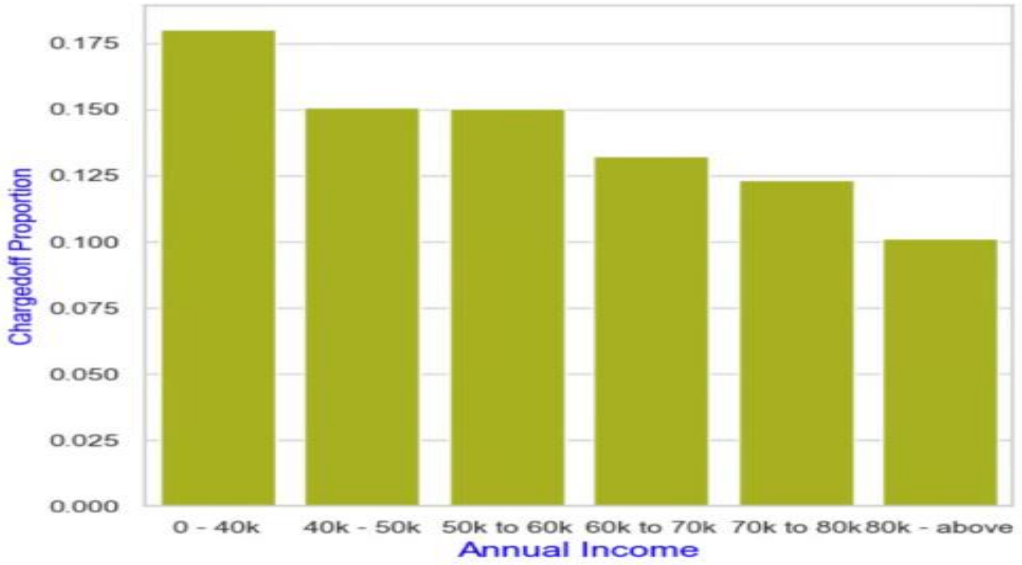
Univariate Analysis



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.

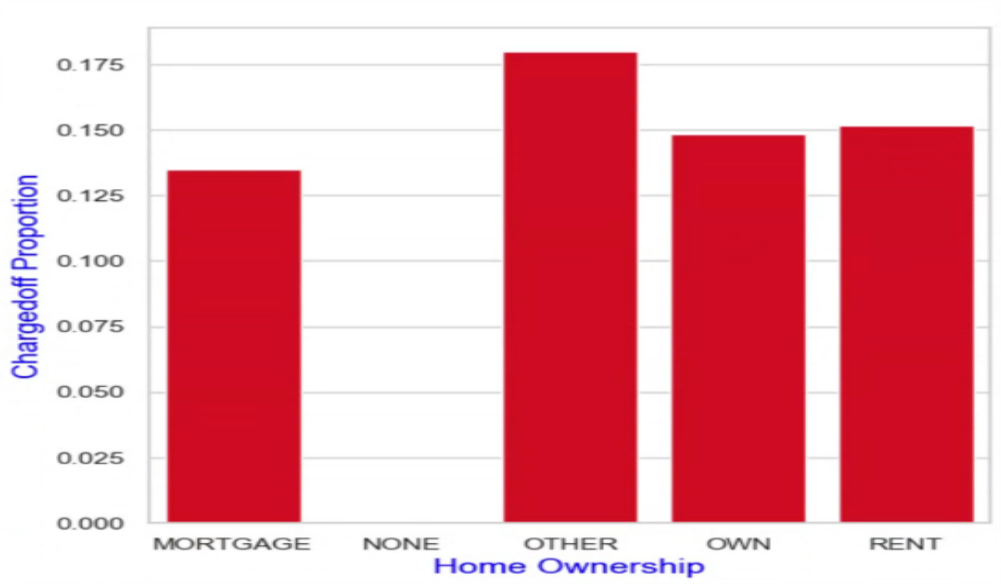
Bivariate Analysis

Comparison of 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.

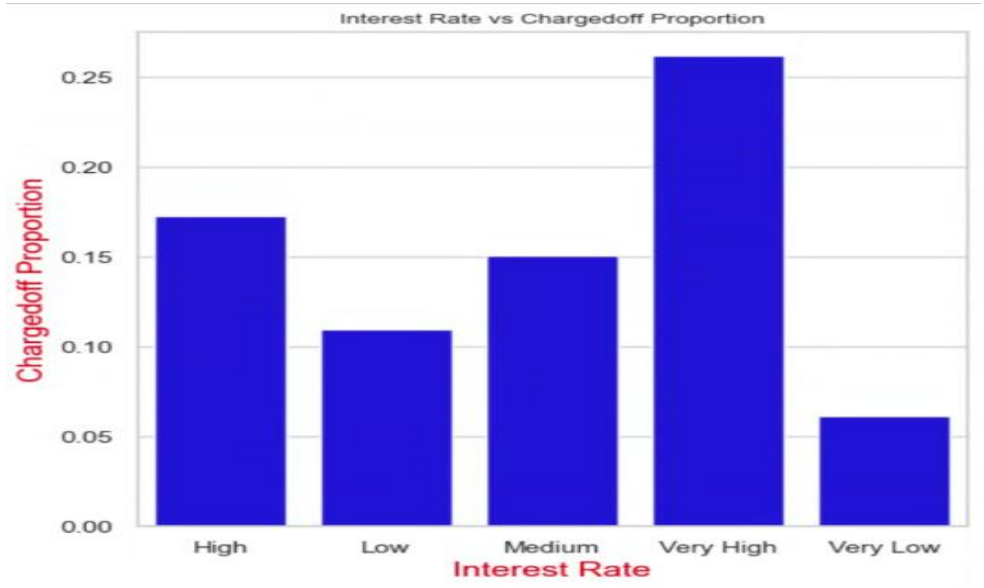
Comparison of Annual Income against Charged-off



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

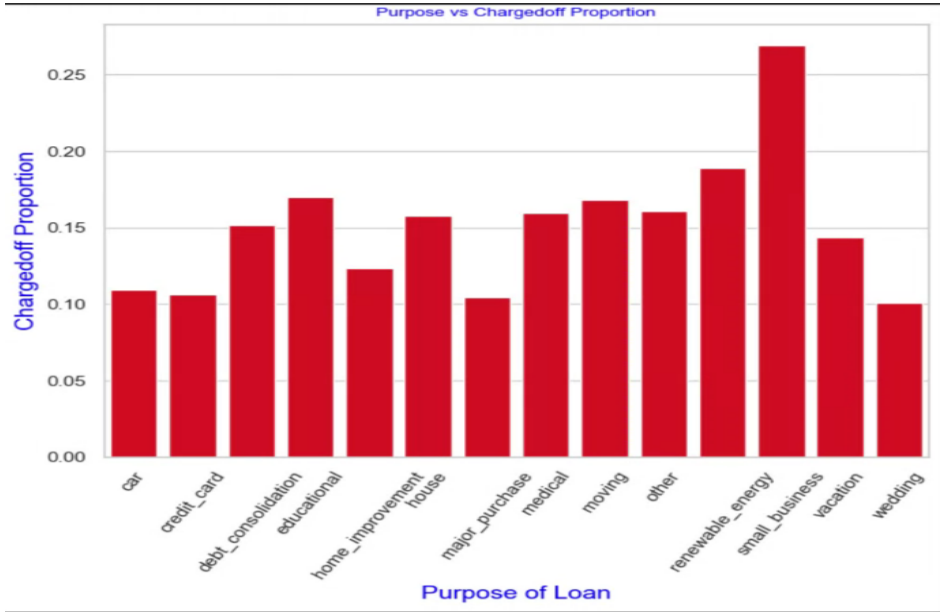
Bivariate Analysis

Comparison of Interest Rate against Charged-off



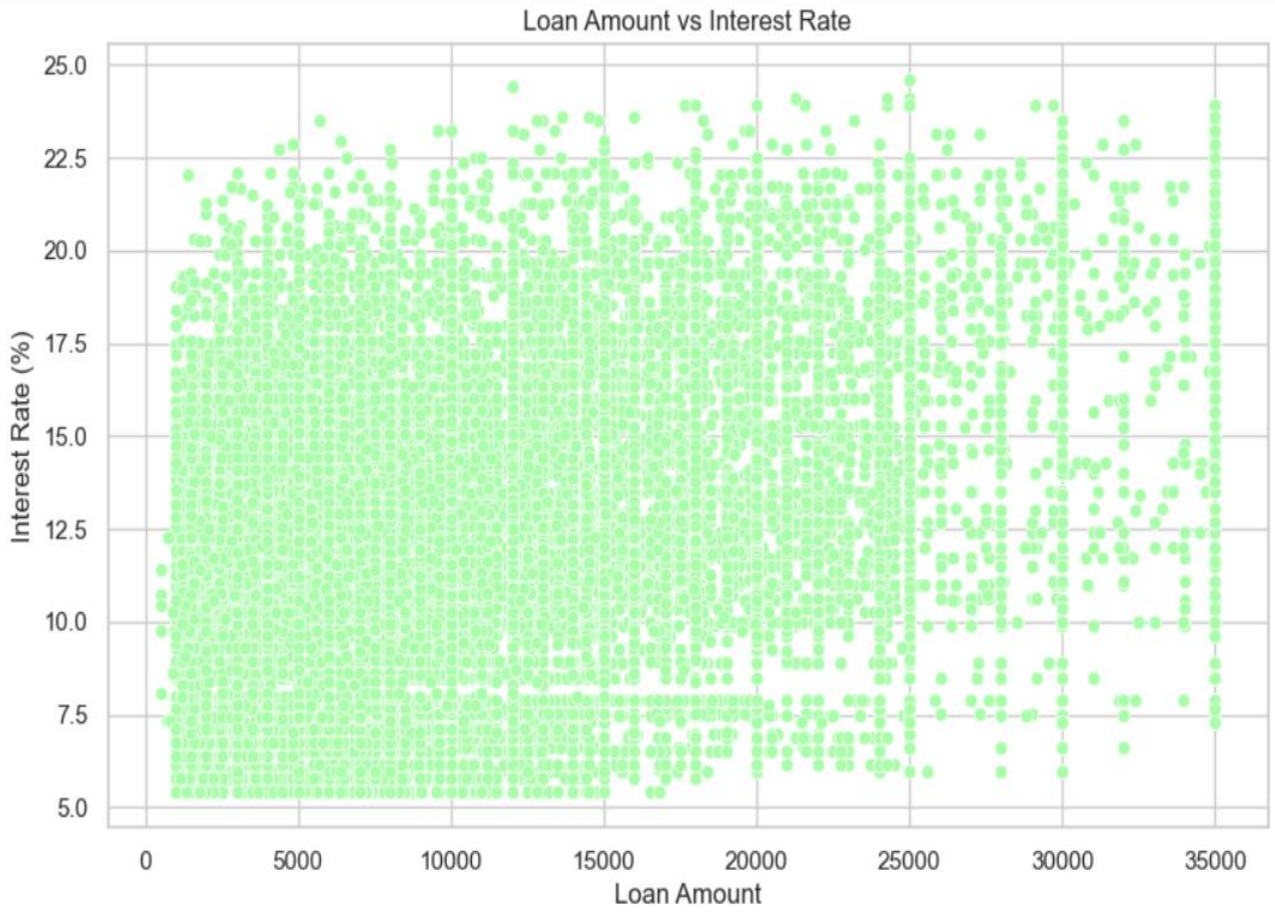
- Loan applicants are having ‘very high’ interest are more chances of maximum charged off.
- Loan applicants are having ‘very low’ interest are less chances of charged-off.

Comparison of Purpose of Loans against Charged-off



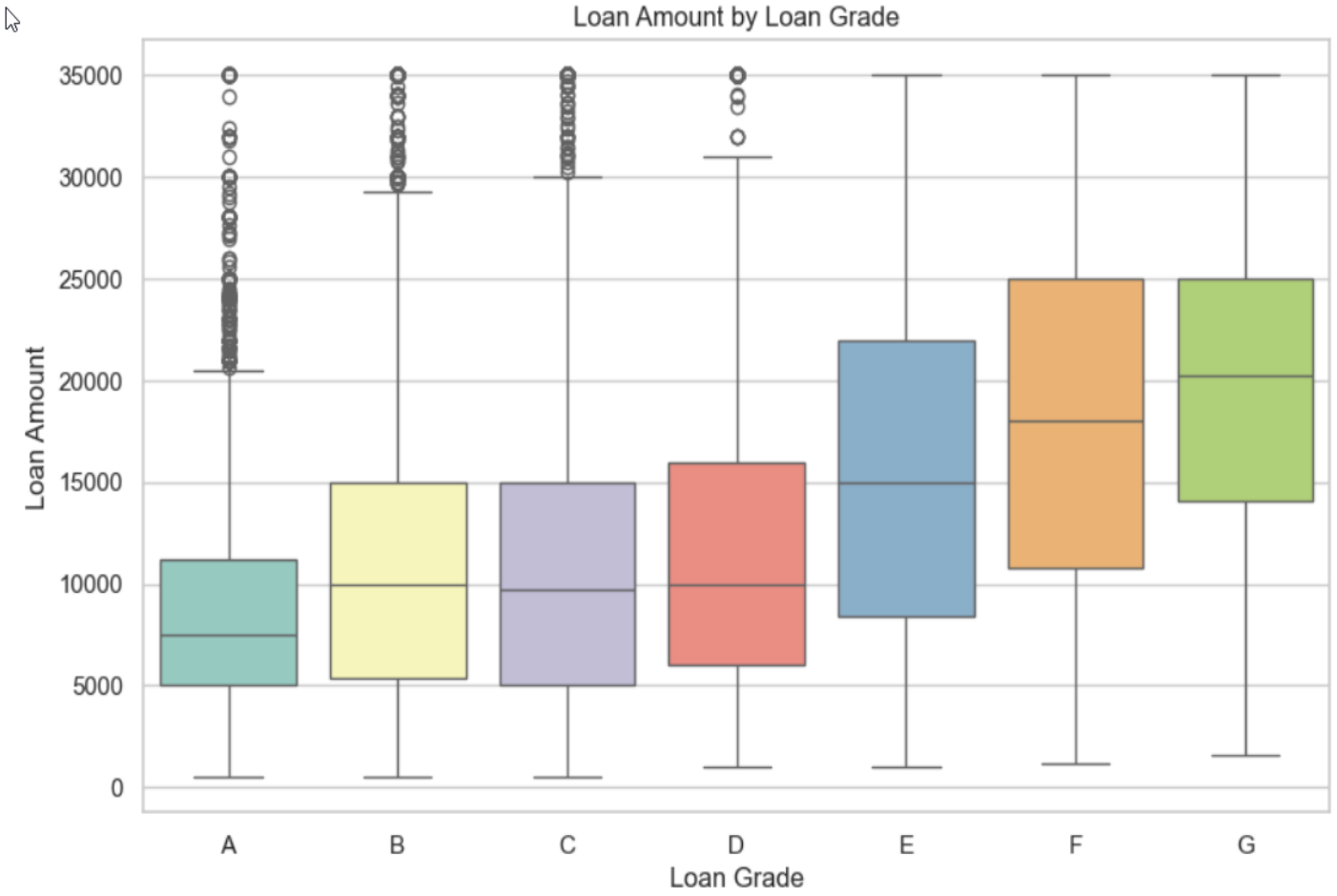
- Applicants who applied for ‘Renewable Energy’ or ‘small business’ are having more chances to be charged-off.

Bivariate Analysis



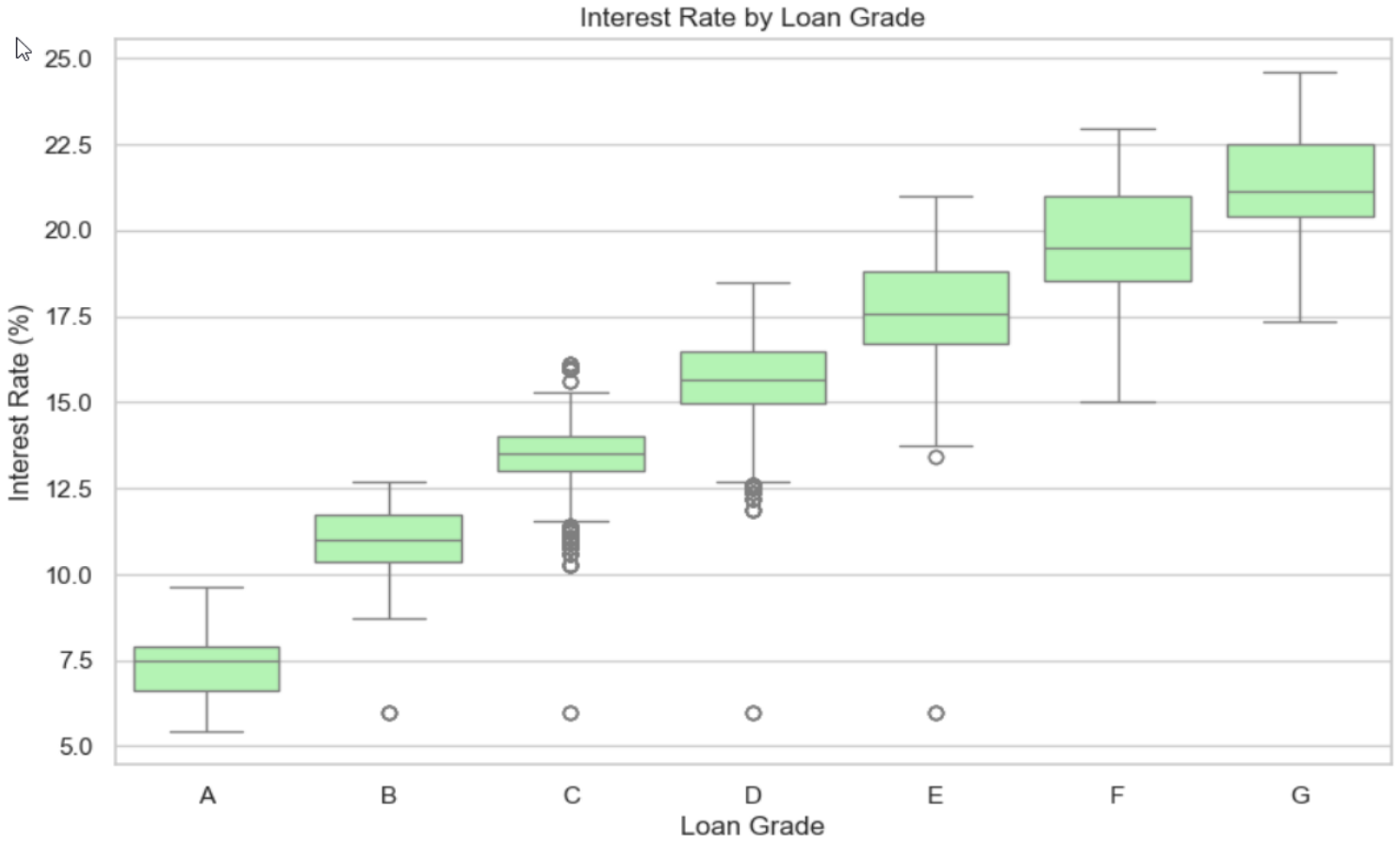
Loan Amount vs Interest Rate: There's no clear linear relationship, but higher loan amounts generally tend to have slightly lower interest rates.

Bivariate Analysis



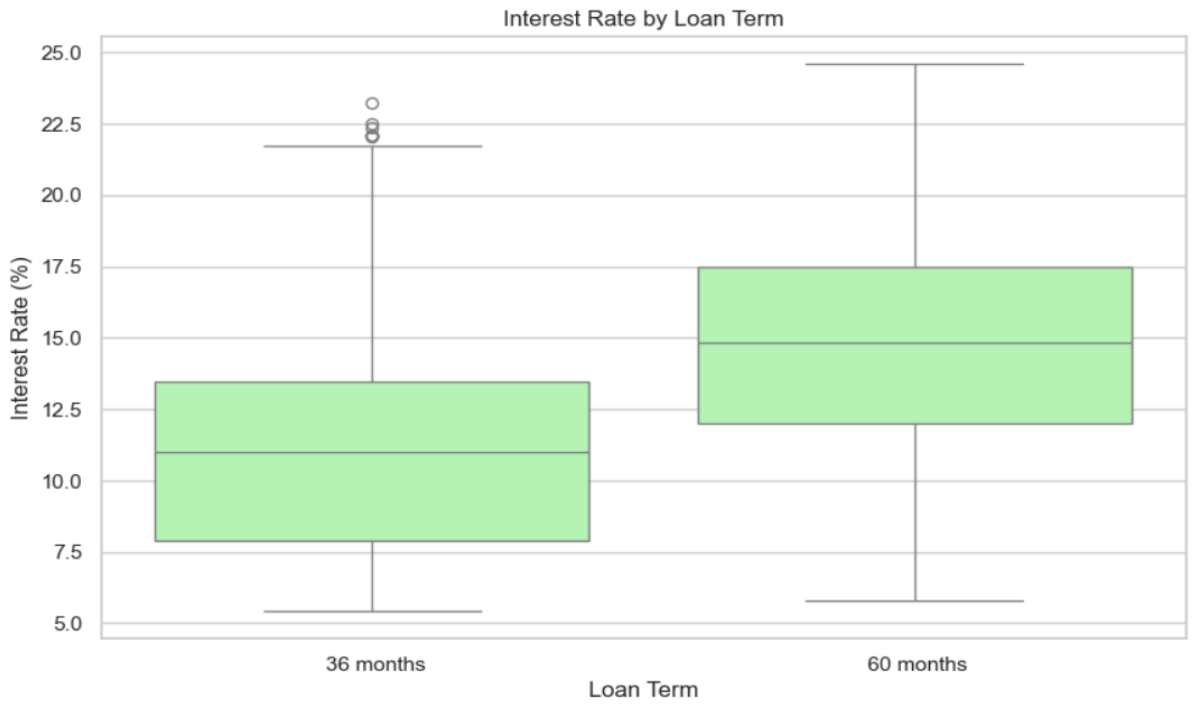
Loan Amount by Grade: Higher-grade loans (A, B) have relatively smaller loan amounts compared to lower grades (C, D, E), indicating that higher-risk borrowers tend to request larger loans.

Bivariate Analysis



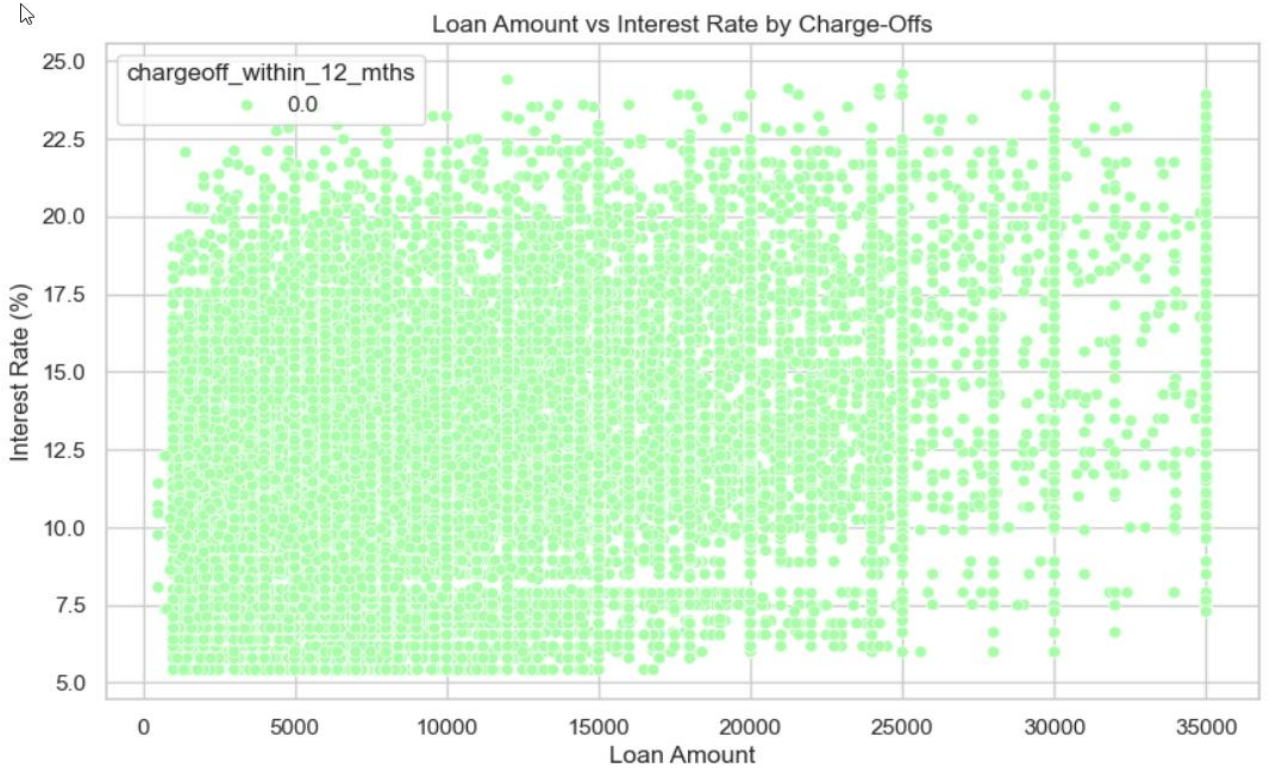
<u>Interest Rate by Grade</u>: Risk heatmap based on grade and interest rate
Higher grades (A, B) are typically less risky, lower grades (D, E) are higher risk. Higher interest rates generally indicate higher risk loans

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.
- Loan applicants are having salary of 80K+ are less likely to be charged off.
- Applicants who belongs to ither category or does not having a home more likely to be 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.
- Applicants who applied for ‘Renewable Energy’ or ‘small business’ are having more chances to be charged-off.
- There's no clear linear relationship, but higher loan amounts generally tend to have slightly lower interest rates.
- 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.
- 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
- 60-month loans generally have higher interest rates compared to 36-month loans, reflecting the increased risk for longer-term loans.
- 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