Lending Club Case Study

Sri Chakra Raj Pyaraka

Purpose

- The primary cause of financial depletion, termed as credit loss, is attributed to loaning money to high-risk borrowers.
- The provided data offers insights into previous loan applicants and their default status. The primary goal is to pinpoint these high-risk loan applicants, thereby lessening the possibility of credit losses.
- Identifying such candidates utilizing EDA can help the firm in understanding portfolio management and risk evaluation.

EDA application on the Dataset

Exploration

- Understanding the dataset
- Mapping relevant variables which contribute to the risks of loan payments

Cleaning

- Handling duplicate, null values
- Removing Columns which do not contribute to analysis after valuation

Transformations

- Creating required derived columns to support analysis
- Converting few column values to desired data types

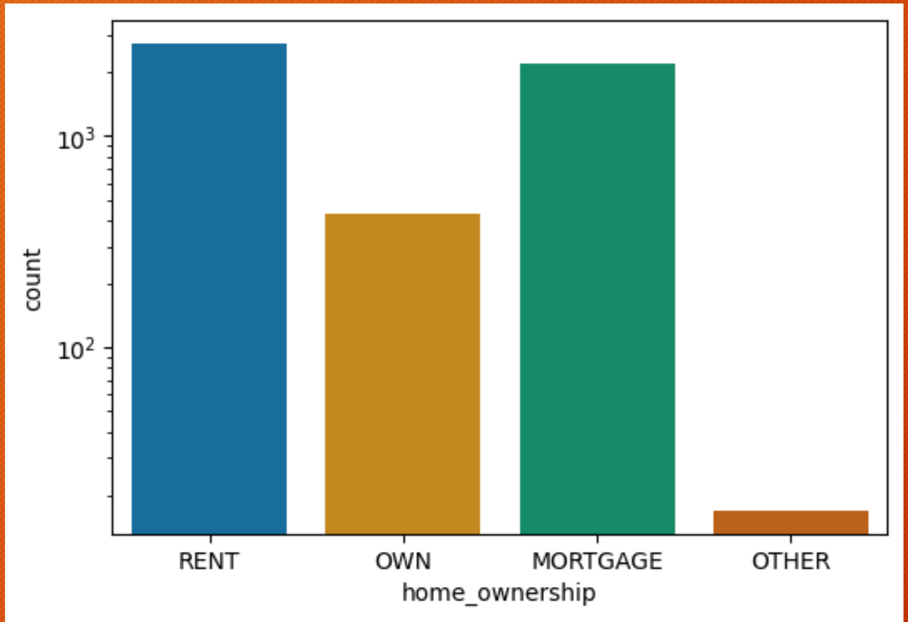
Univariate and Bivariate Analysis

- Deriving impact of various continuous and categorical variables on ‘loan_status’
- Identifying Outliers across dataset

Univariate Analysis

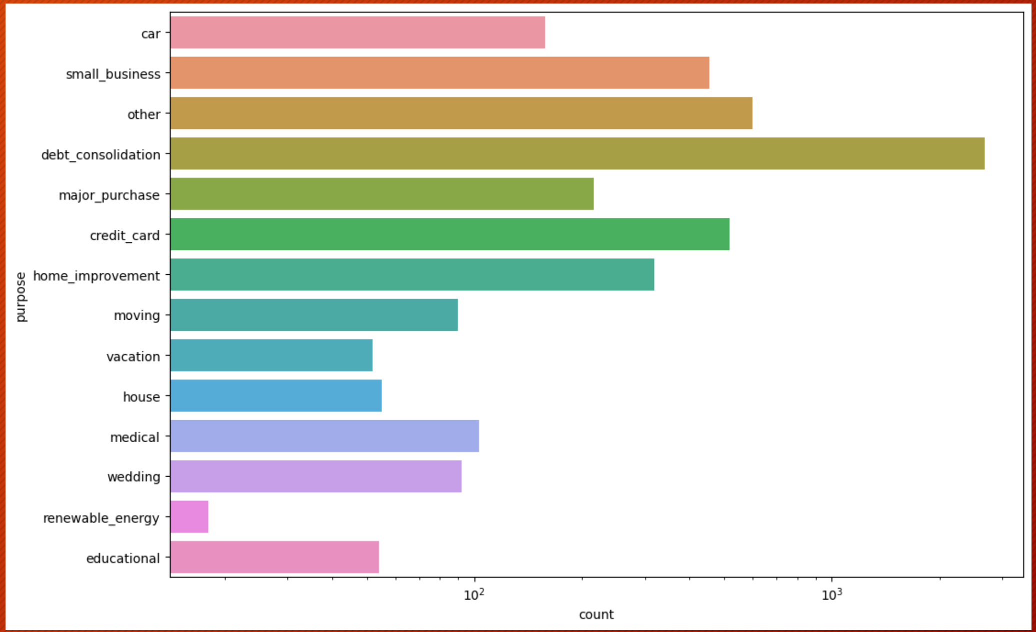
Home Ownership

- Majority of the loan applicants either live in rent or have a house mortgage



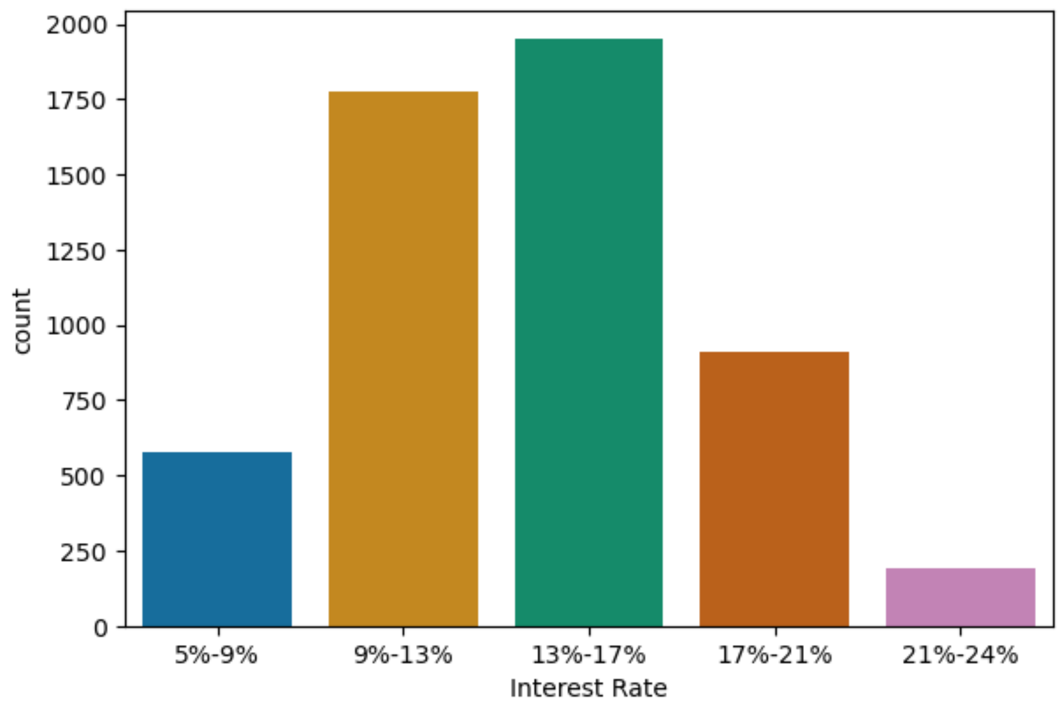
Purpose

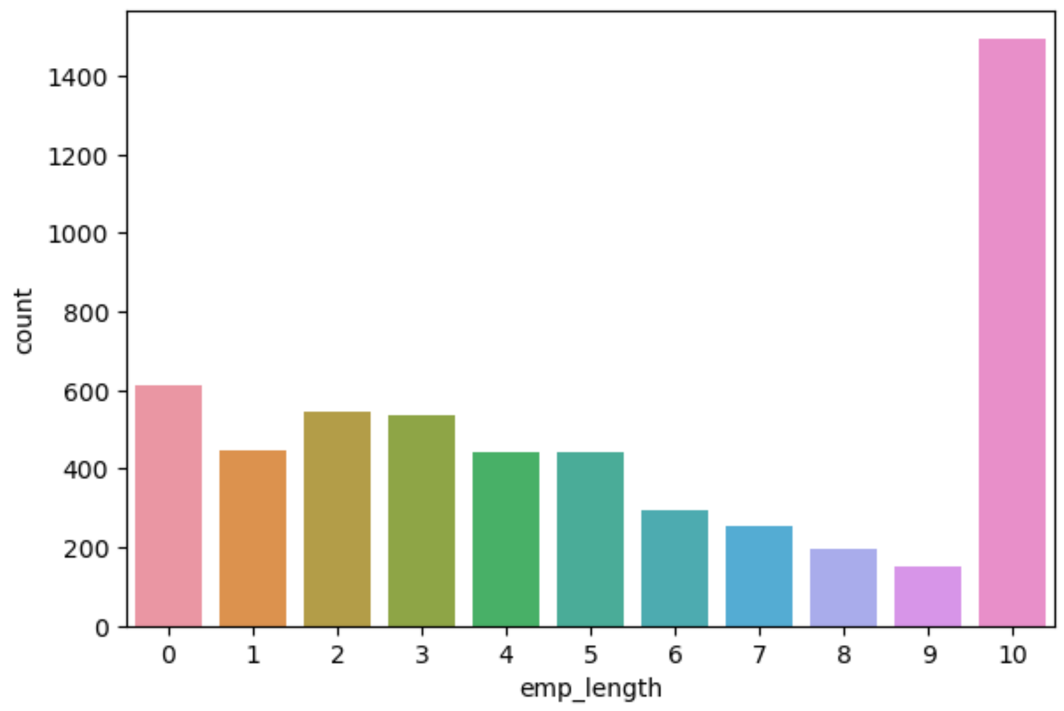
- Based on the plot, it is evident that majority of the defaults fall under debt consolidation, credit card, home improvement and others categories.



Univariate Analysis (Interest Rate vs Employment Period)

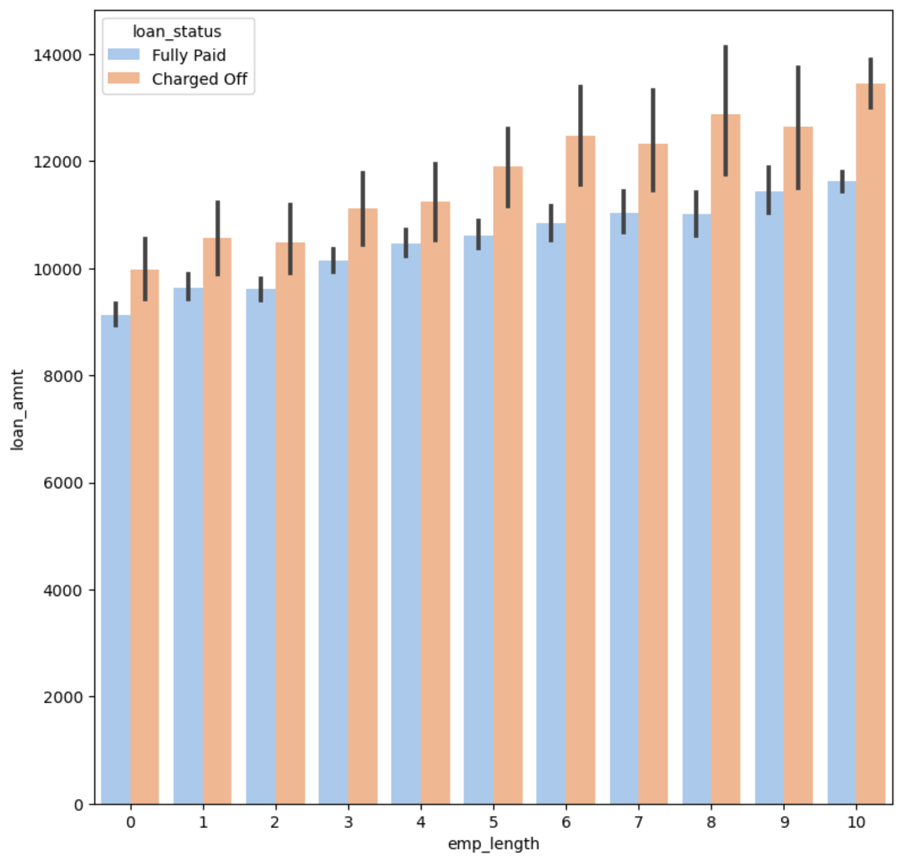
- It is observed that people with higher employment length (>=10 years) are likely to default, along with few having less than 3 years
- Higher Interest rates eventually get inclined to higher defaults, observed the same from the data

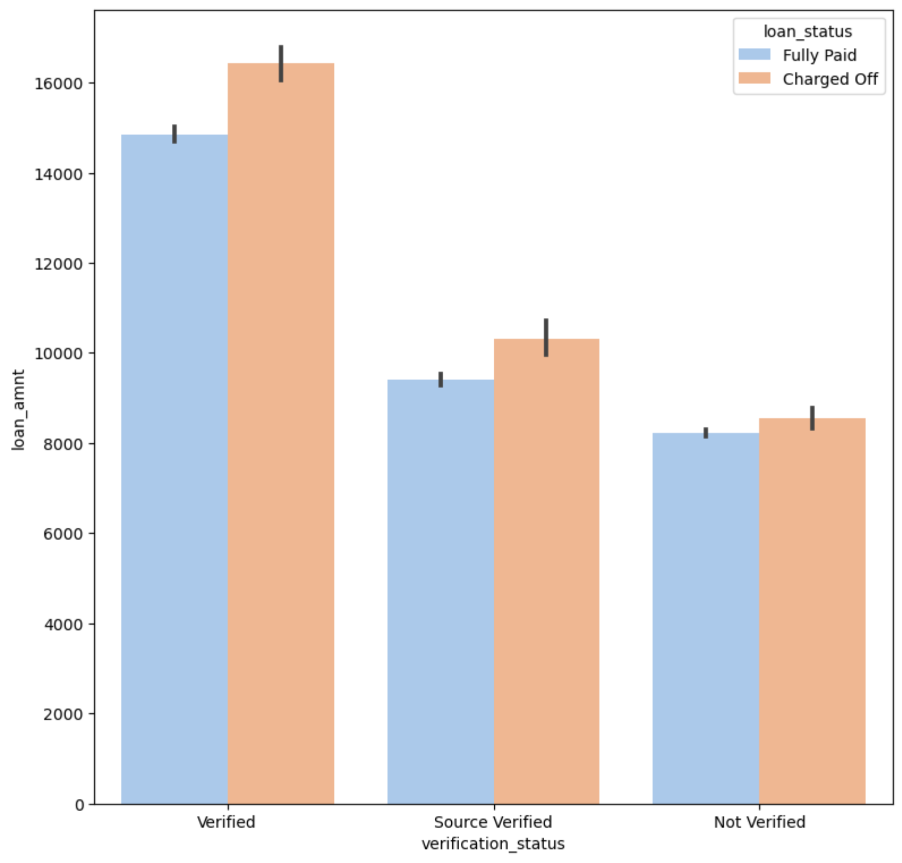




Bivariate Analysis (Loan Amount vs Employment Period and Verification Status)

The higher the loan_amount and the employment_period, the more changes loan getting defaulted. When corelating it with verification status, the firm should ensure when lending out high amounts of loan, they properly validate the employee’s length and employer

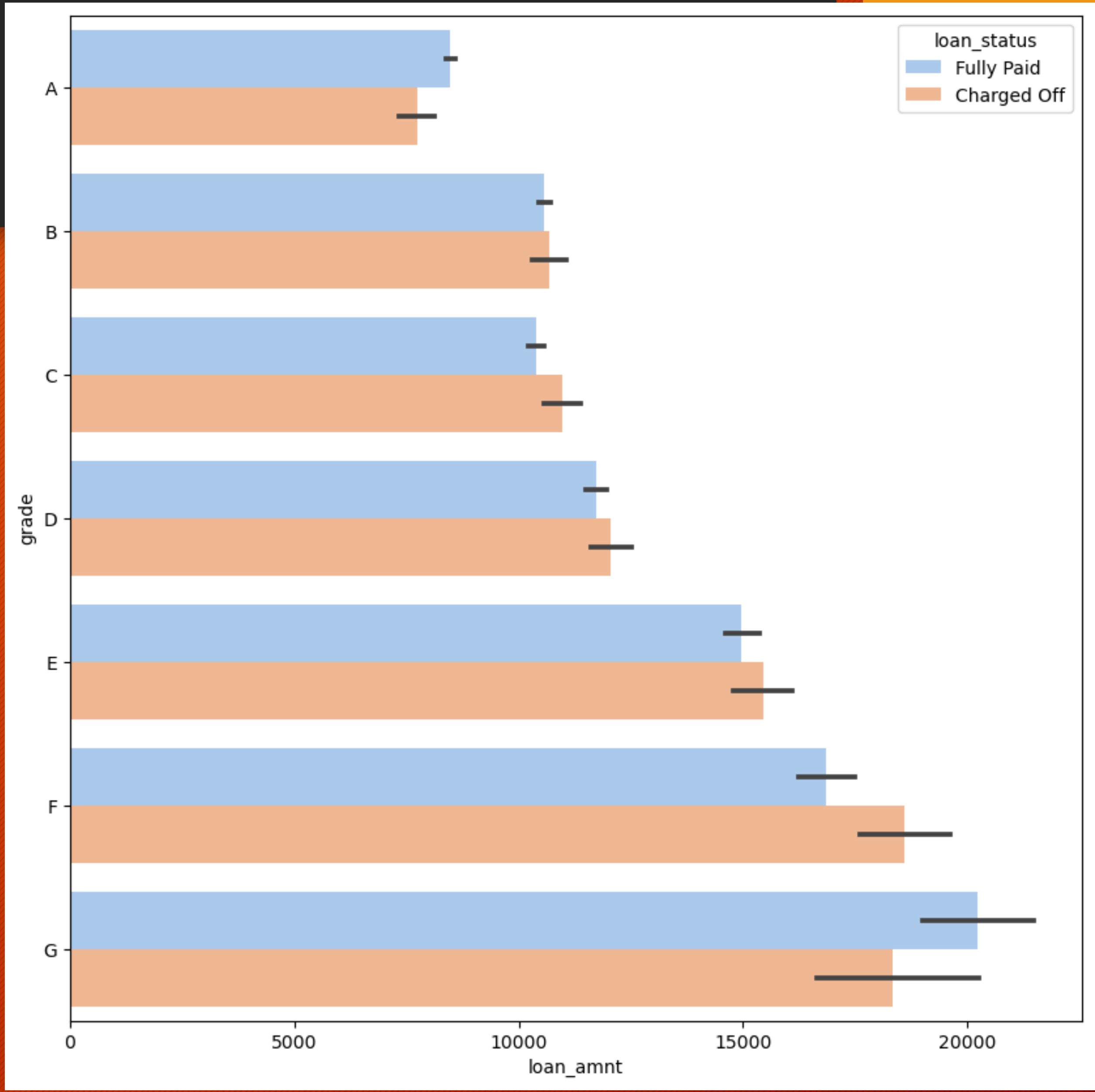




Bivariate Analysis

Grade vs Loan Amount

- As the loan grade increases, the chances of getting defaulted increases. Higher Grade ~ Higher Credit Risk



Conclusion

- There is a greater volume of fully completed loans compared to defaulted ones.
- The most common type of loan is for debt consolidation.
- Loans with larger sums are more prone to default.
- Individuals with lower annual earnings are more prone to default on their loans.
- An increase in interest risk leads to an increase in defaulted loans.
- The better the rating of the loan, the less likely it is to be defaulted.
- Despite debt consolidation being the most common type of loan, it's the small business loans that contribute the most to defaults.
- The proportion of defaulted loans is highest in New York.
- A higher debt-to-income (dti) ratio leads to an increase in defaulted loans.
- An elevated revolving utilization ratio results in more defaulted loans.
- A higher number of loan inquiries is correlated with an increased number of defaulted loans.
- As the total number of accounts rises, the number of defaulted loans decreases.