Lending Club Case Study

Group Members:

1. Aniket Tote
2. Ashish Kumar Rajpurohit
3. Vikas Kumar Pandey
4. Siddeswara Basireddy

Introduction

The consumer finance company(Lending Club) which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan**,** i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

Problem Statement

- Borrowers who default cause the largest amount of loss to the lenders labelled as 'charged-off' i.e. defaulters.
- Identification of risky loan applicants using Exploratory Data Analysis(EDA).
- The company wants to understand the driving factors behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Strategy To Solve Problem

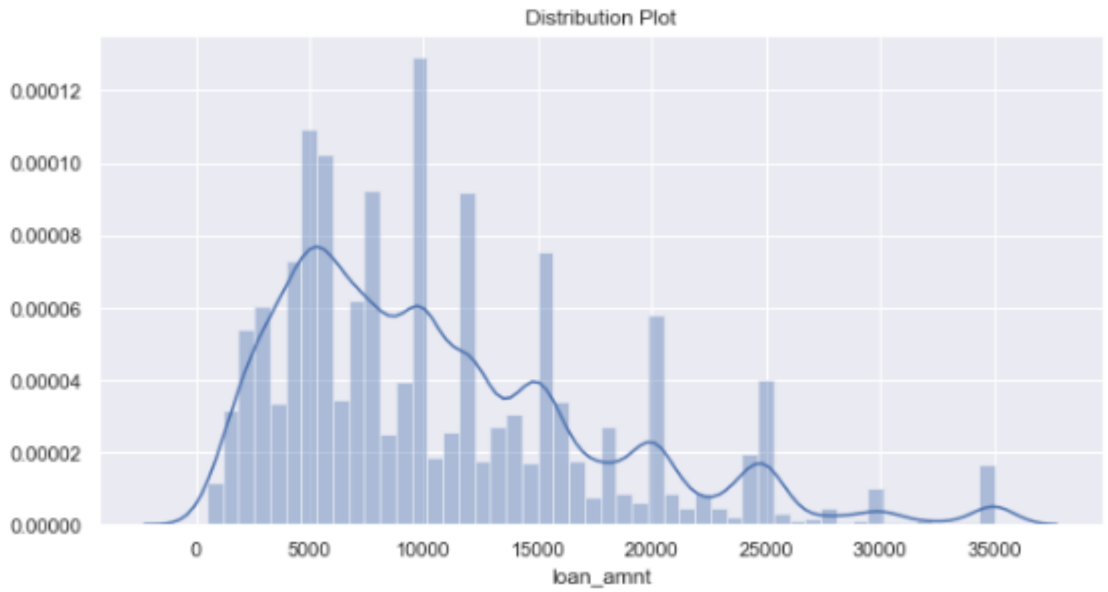
Strategy

1. Univariate Analysis
2. Bivariate Analysis
3. Correlation Matrix
4. Hypothesis Testing

Univariate Analysis

Loan Amount

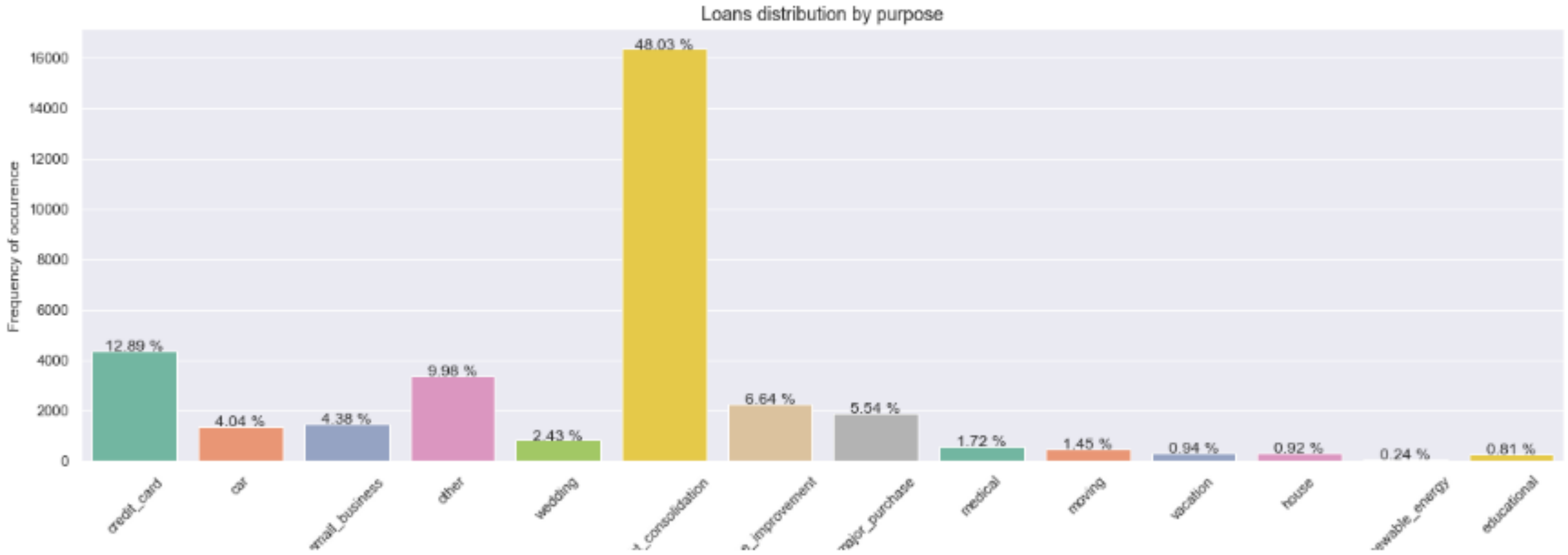
- Most of the loan amounts are distributed between 5000 to 20000.



Univariate Analysis

Purpose Of Loan

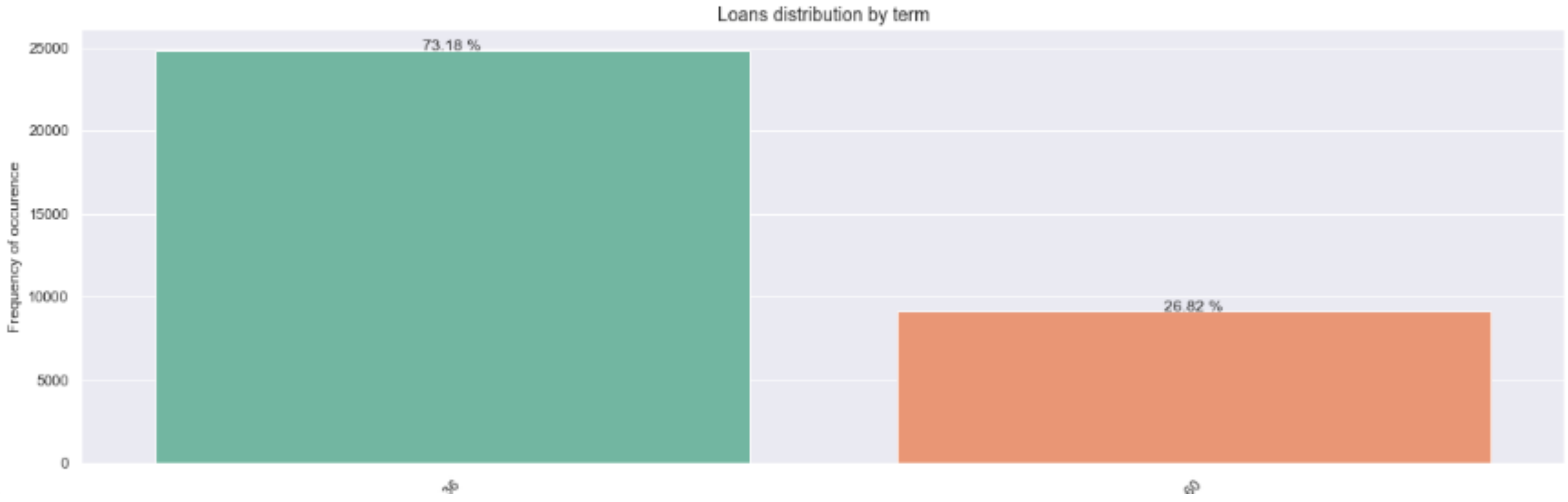
- Approx 48% of the applicants applied loan for paying their other loans(Debt Consolidation).



Univariate Analysis

Term

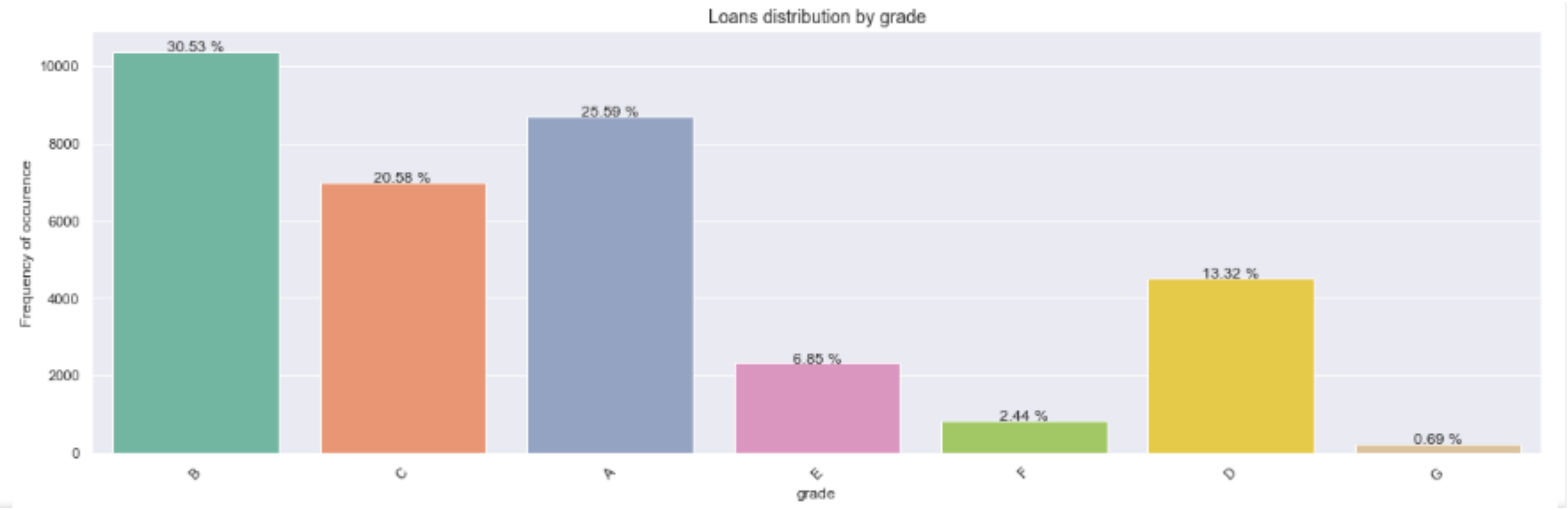
- 73.18% of applicants applied loan for 36 months term period.



Univariate Analysis

Grade

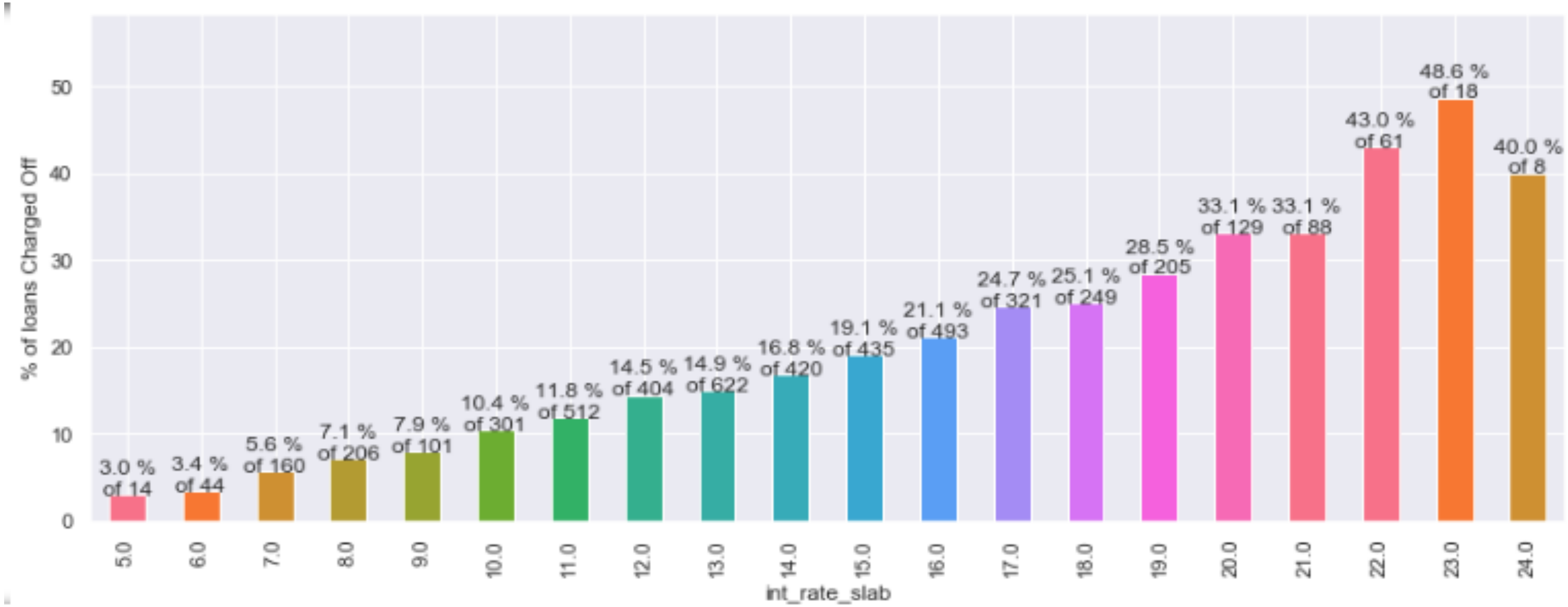
- Around 31% loans come under B grade and 26% comes under A grade .



Bivariate Analysis

Interest Rate vs Percentage Of Loans Charged Off

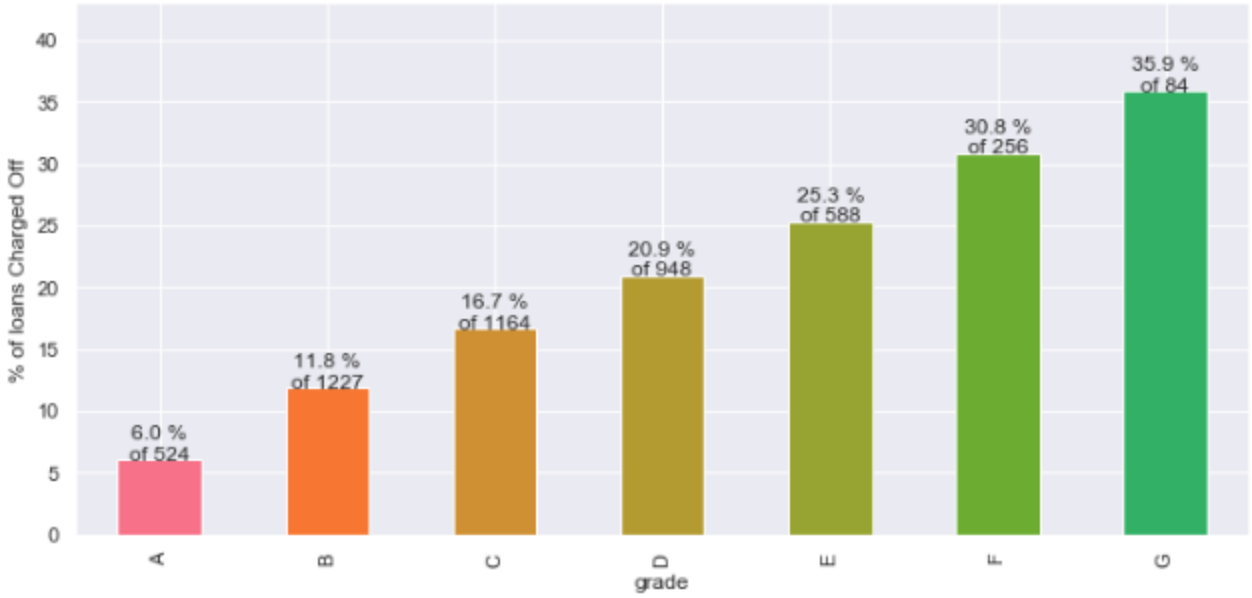
- As Interest Rate increases the default rate also increase.



Bivariate Analysis

Grade vs Percentage Of Loans Charged Off

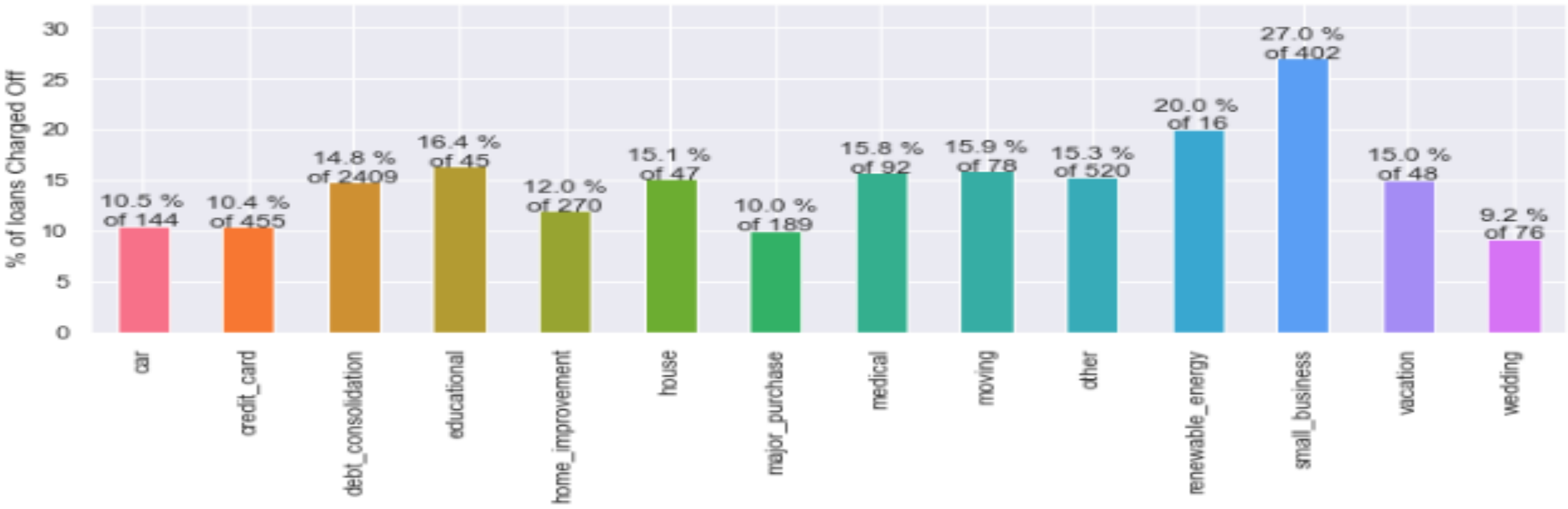
- We can see that , as Grade increases the default rate also increases.



Bivariate Analysis

Purpose Rate vs Percentage Of Loans Charged Off

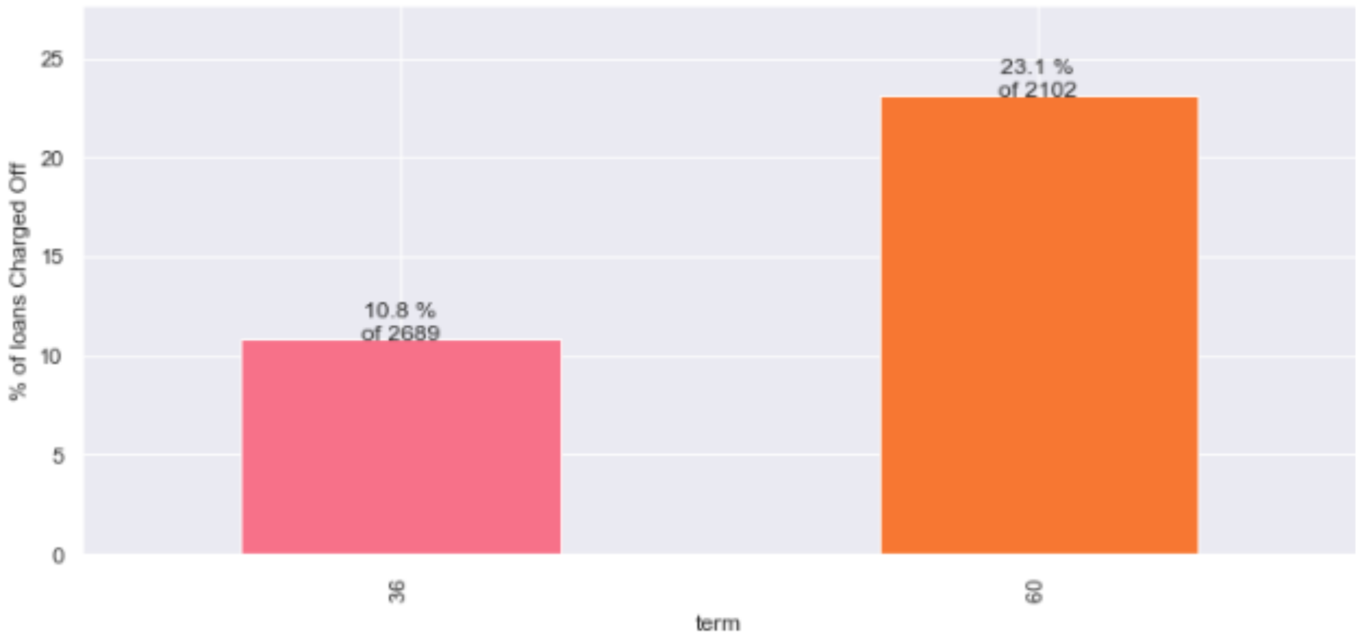
- 27% of loans for small business are Charged Off. Making them the most risky purpose.
- Approximately ~48% of the loans are issued for the purpose of dept consolidation.
- 20% of the loans for renewable_enrgy are charged Off, but the number is too less to be of significance.



Bivariate Analysis

Term vs Percentage Of Loans Charged Off

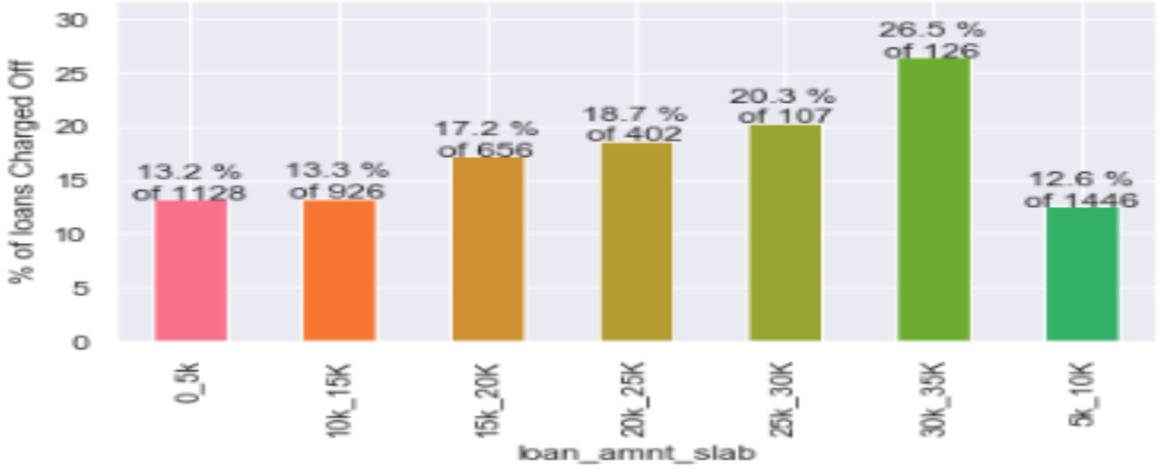
- The 60month term is having more chances of **Charged Off** as compare to 36month term.



Bivariate Analysis

Loan Amount vs Percentage Of Loans Charged Off

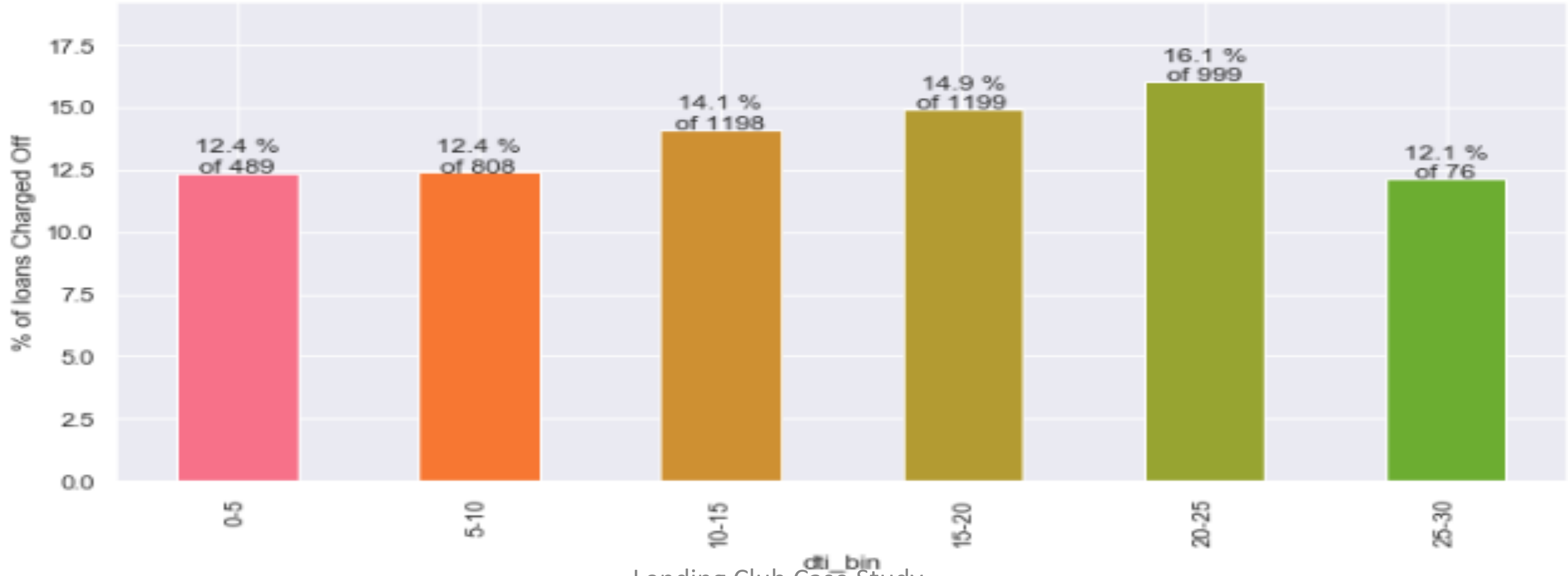
- If the loan amount increases, default rate also increases



Bivariate Analysis

Dti vs Percentage Of Loans Charged Off

- When the dti or dept payment to income ratio is higher than 20, higher percentage of loans are Charged Off
- Higher the dti higher the chances of loan being Charged Off



Hypothesis Testing

Hypotheses testing on income of customer using P - Value method.

Null and Alternate hypotheses of income of customer.

- H0 : There is NO correlation between income of customer and loan_status.
- H1:There is correlation between income of customer and loan_status.

The Calculated p – Value is zero that means Null hypothesis is rejected i.e. There is correlation between income of customer and loan_status.

Hypothesis Testing

Hypotheses testing on loan Term using p - Value method.

Null and Alternate hypotheses of income of customer.

- H0 : There is NO correlation between loan term and loan_status.
- H1:There is correlation between loan term and loan_status.

The Calculated p – Value is nan that means Null hypothesis is rejected i.e. There is correlation between loan term and loan_status.

Hypothesis Testing

Hypotheses testing on Interest rate using p - Value method.

Null and Alternate hypotheses of income of customer.

- H0 : There is NO correlation between Interest rate of loan and loan_status.
- H1: There is correlation between Interest rate of loan and loan_status

The Calculated p – Value is zero that means Null hypothesis is rejected i.e. There is correlation between Interest rate of loan and loan_status.

Conclusion

From the Univariate and Bivariate analysis we can conclude, the below mention column to be the key factor to decide customer is going to be defaulter or not.

Target Variable

Loan Status

Major variables to consider for loan prediction:

1) Purpose of Loan
2) Grade
3) Interest Rate
4) Term
5) dti
6) Loan Amount