Lending Club Case Study

Name - Aditya Anil Regundwar

Abstract

- In the last decade, the number of retail borrowers, SMEs, and commercial borrowers has increased dramatically.

- Though most financial institutions have seen an increase in revenue and profit due to this rising trend, not everything is green.

- In recent years, there has been an increase in loan defaults, which has already begun to affect the bottom lines of several financial institutions.

- An organization wants to forecast who would default on a consumer lending product. Based on what they've seen, they have data on previous client behavior.

- As a result, when they gain new consumers, they want to know who is riskier and who isn't.

Steps to be followed in problem solving :

1. Data Understanding :
    1. Understanding the size of dataset
    2. Information about the columns and its datatype

2. Data Cleaning :
    1. Dealing with the Null values
    2. Removing unnecessary columns
    3. Removing the columns having high percentage of missing values

3. Data Analysis :
    1. Univariate Analysis - Analysing each column, plotting the distributions of each column.
    2. Bivariate Analysis - Analysing the continuous data columns with respect to the categorical column
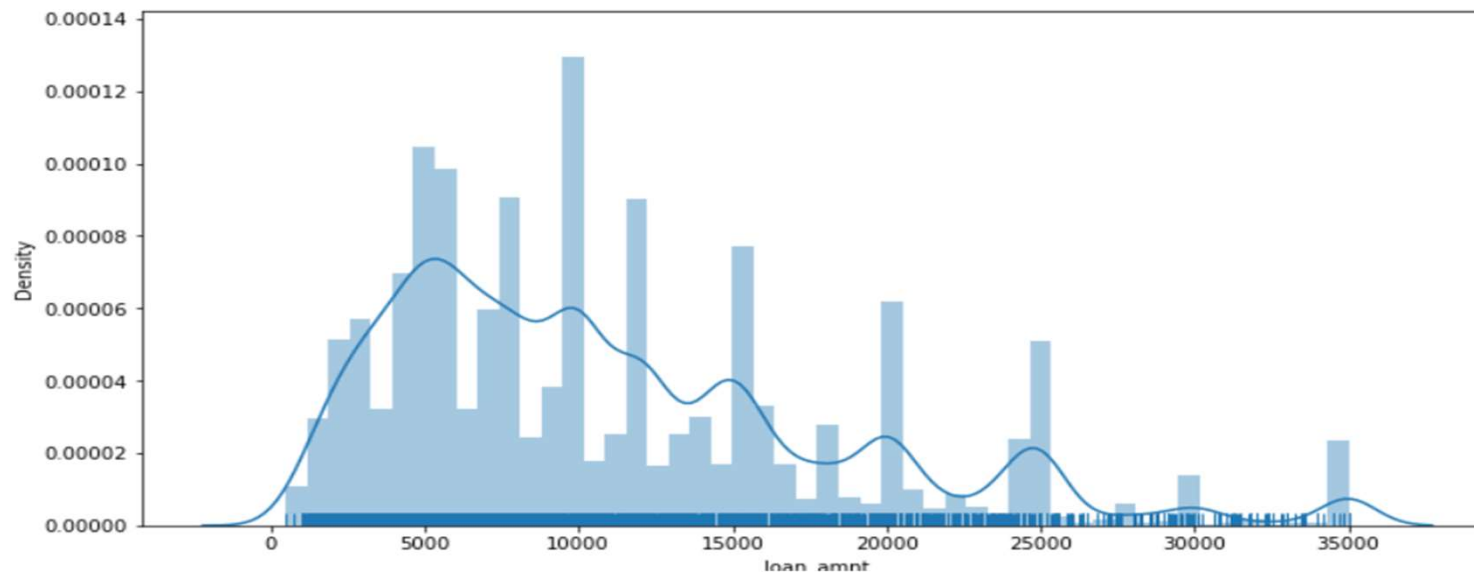
4. Recommendation
    Analysing all plots and recommendations for reducing the loss of business by detecting columns best which contribute to loan defaulters.
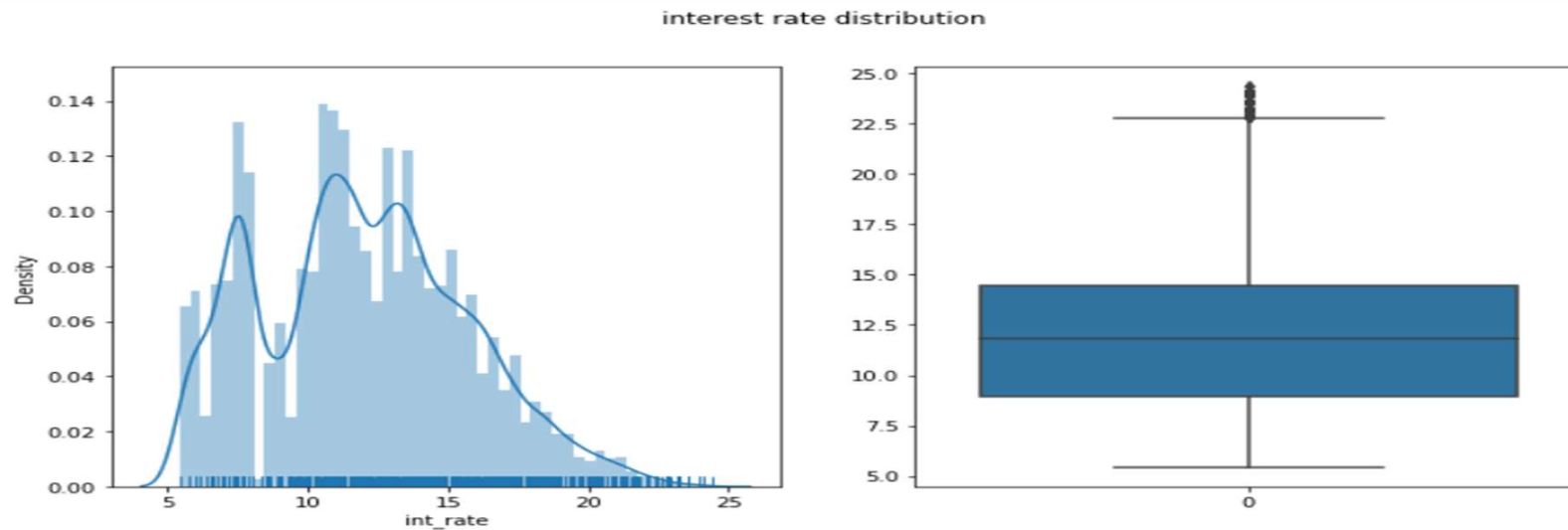
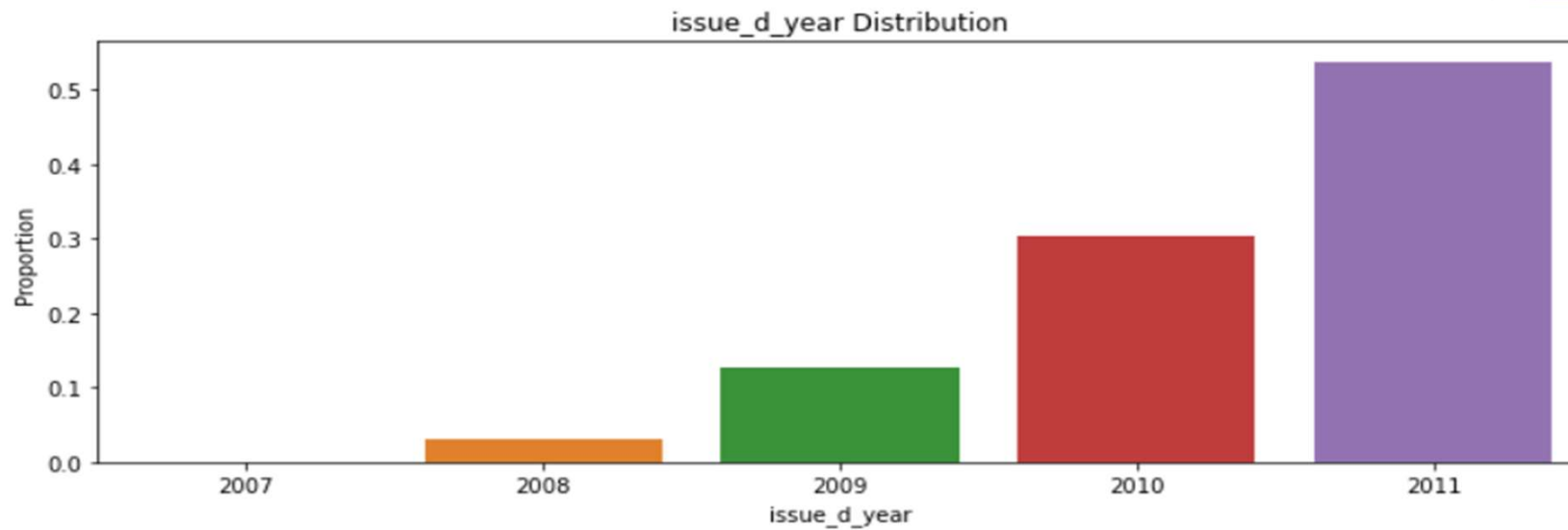# Univariate Analysis

Univariate Analysis

- More number of people took loan amount of 10000, and also median of distribution is 10000.

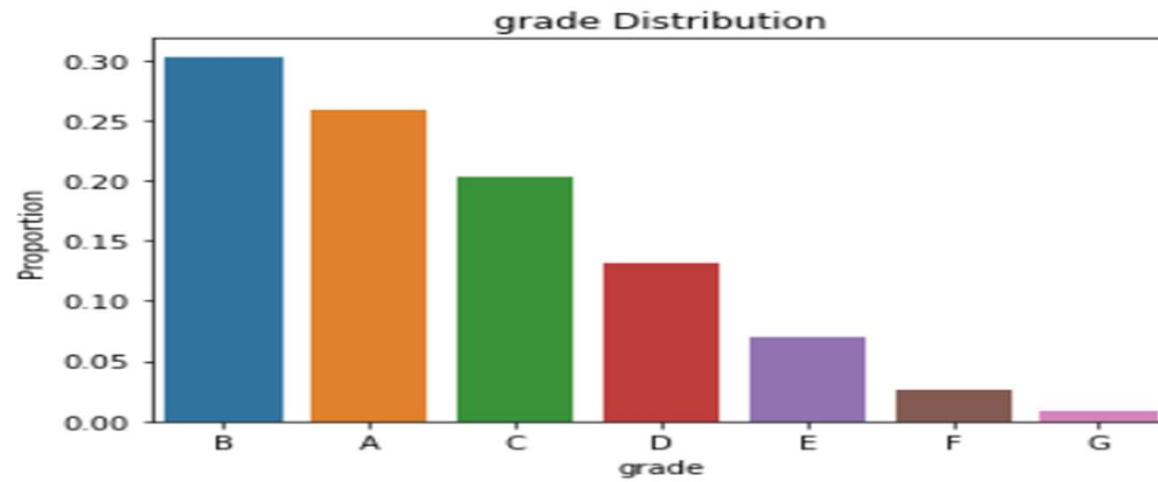- Very few people(25%) took more than 15000 loan amount.

Univariate Analysis

interest rate distribution

- Most of the interest rates lies between 9% to 14.5%

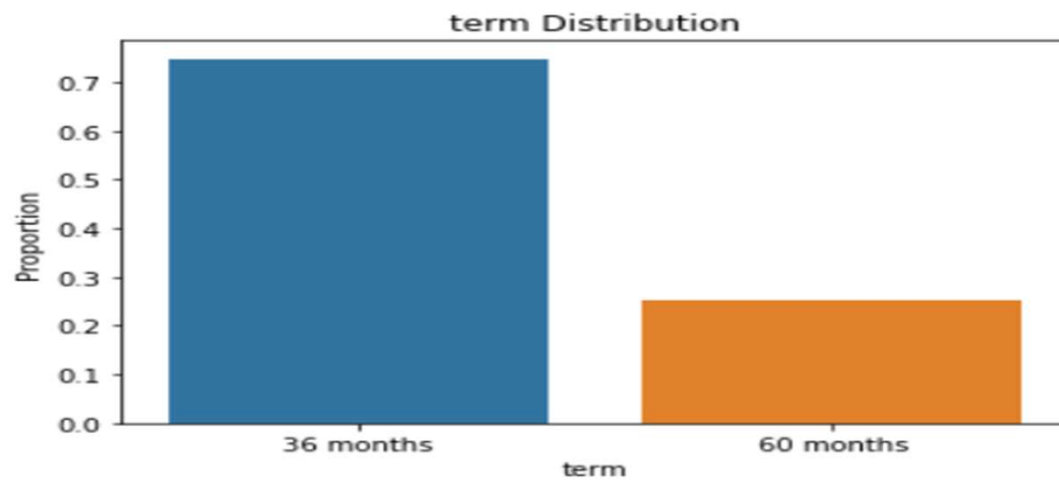- Some people took loan at higher rates of interest that is above 22.5%

Univariate Analysis

issue_d_year Distribution

- The lending club(LC) has doubling loan issues every year
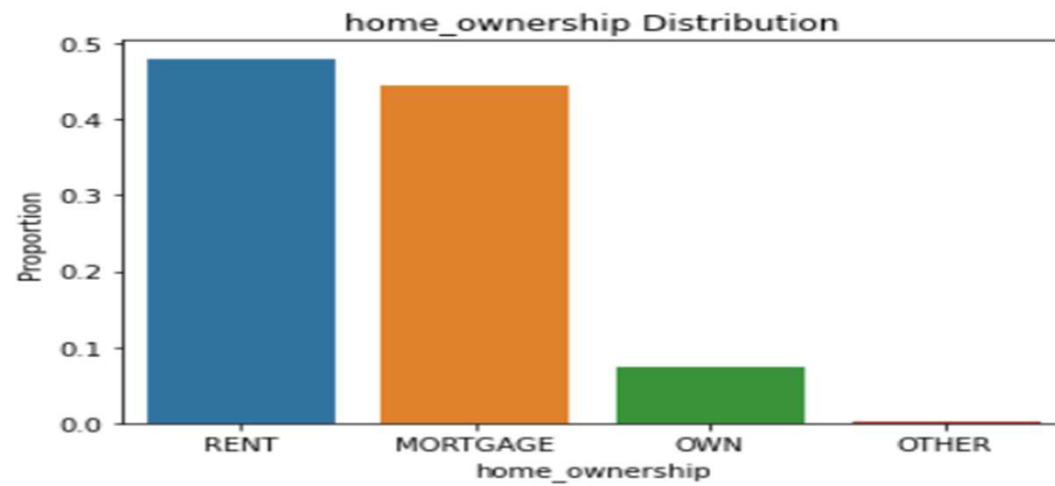
Univariate Analysis

grade Distribution

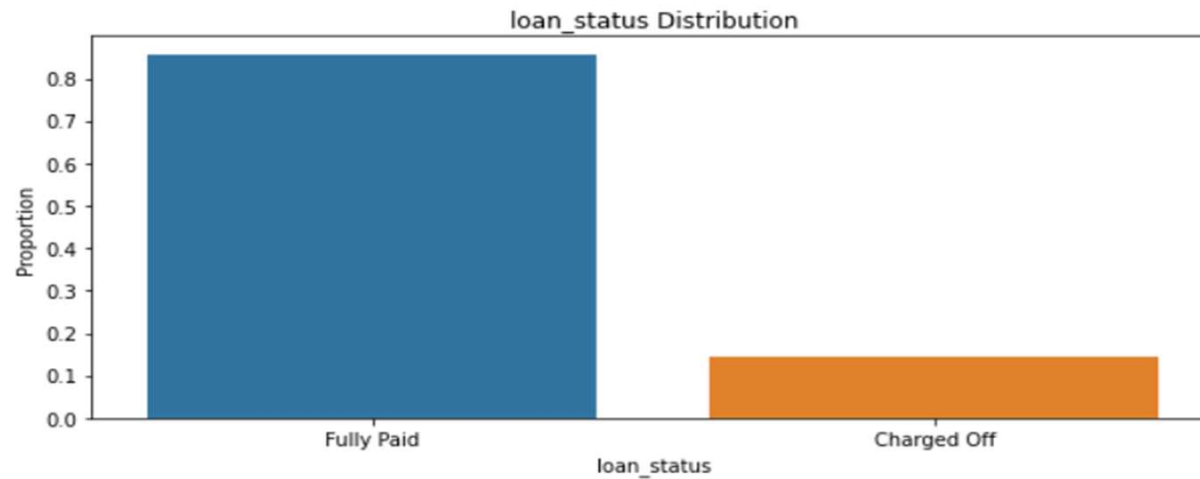- Most borrowers fall under A and B grades then other grades

Univariate Analysis

- Borrowers have taken 36 months tenure more than 60 months.

Univariate Analysis

home_ownership Distribution

- The loan borrowers are mostly having rented and mortgage houses

Univariate Analysis

- 85% of borrowers has paid the loan fully. where are 14% are defaulted the loan
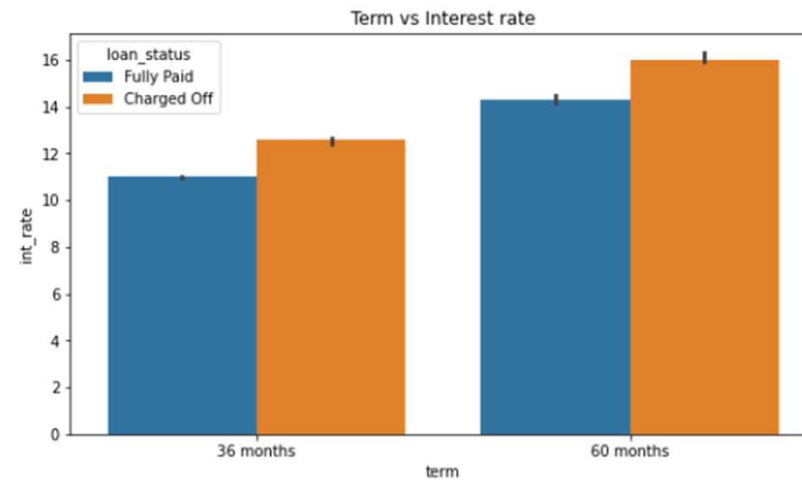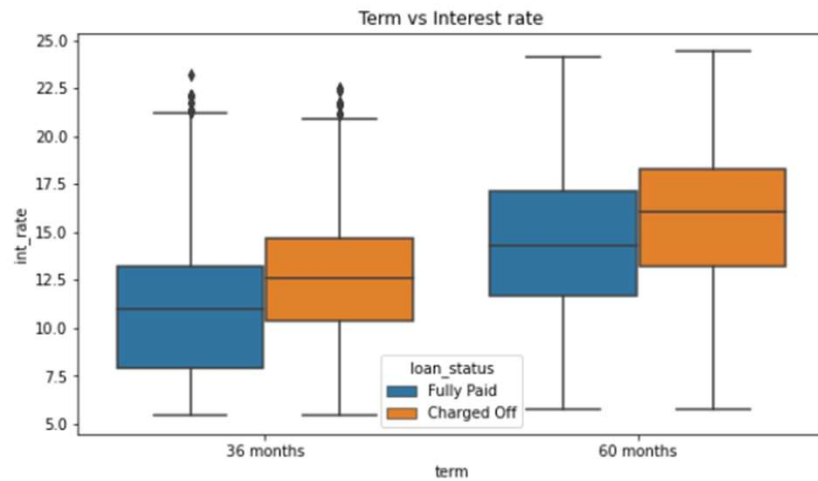
Bi-variate Analysis
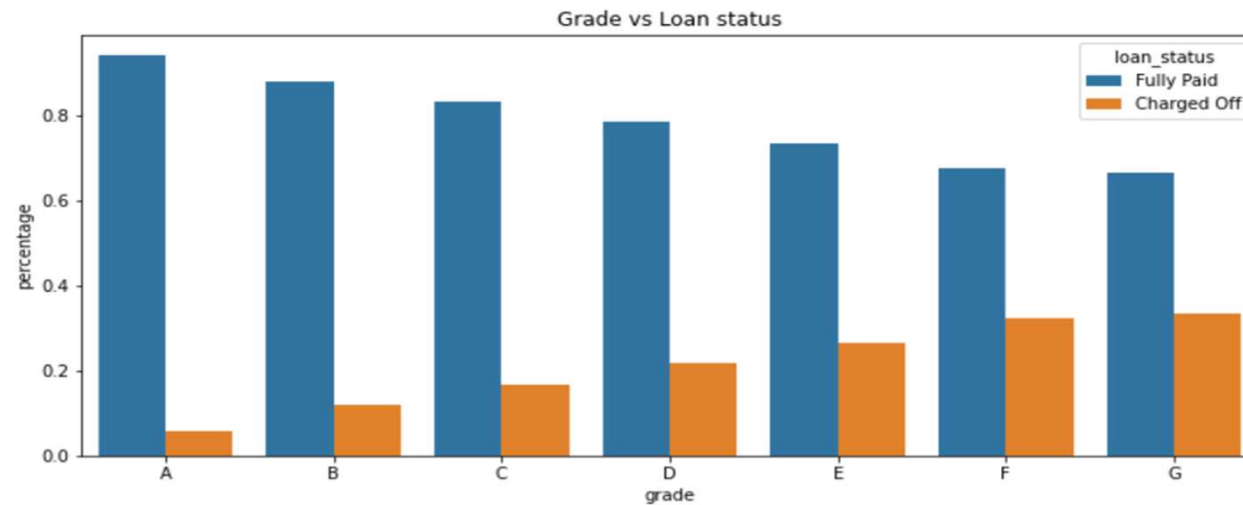
Bivariate Analysis

Term vs Loan status
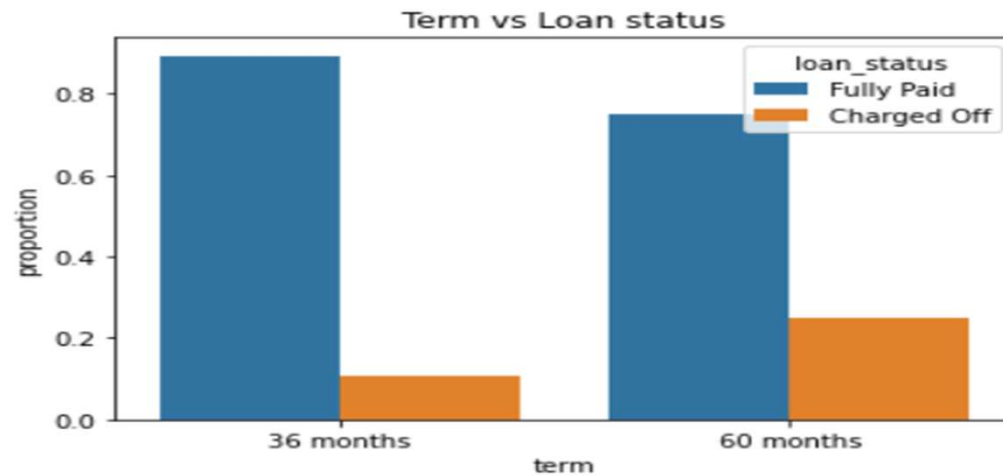
- More proportion of borrowers defaulted loan in 60 months term then 36 months.
- Also the Fully Paid rate is higher in 36 months tenure.
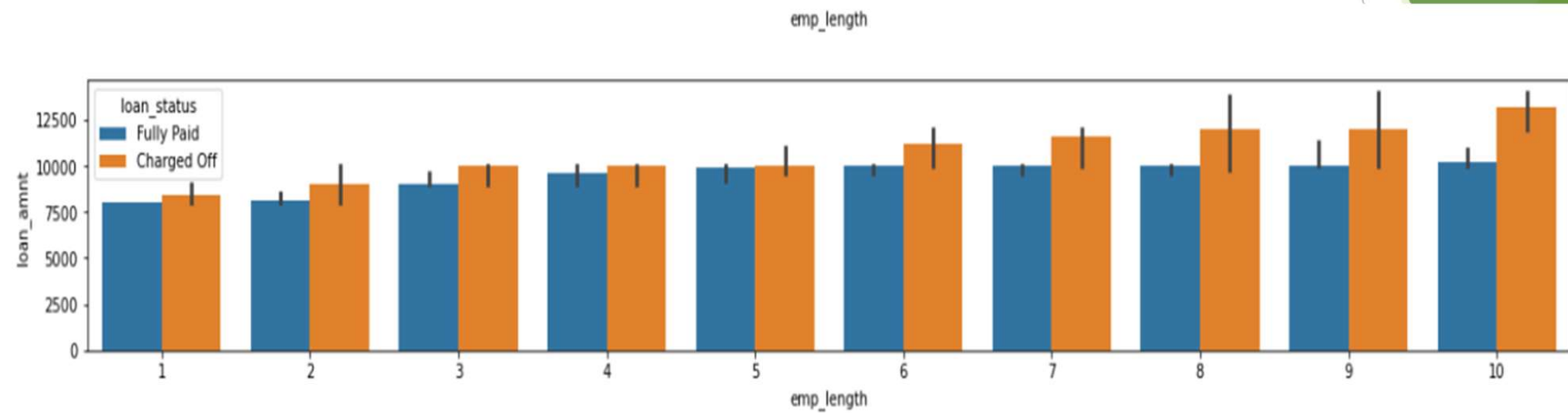
Bivariate Analysis

- For higher interest rates the default rate is higher in both 36 and 60 months tenure.
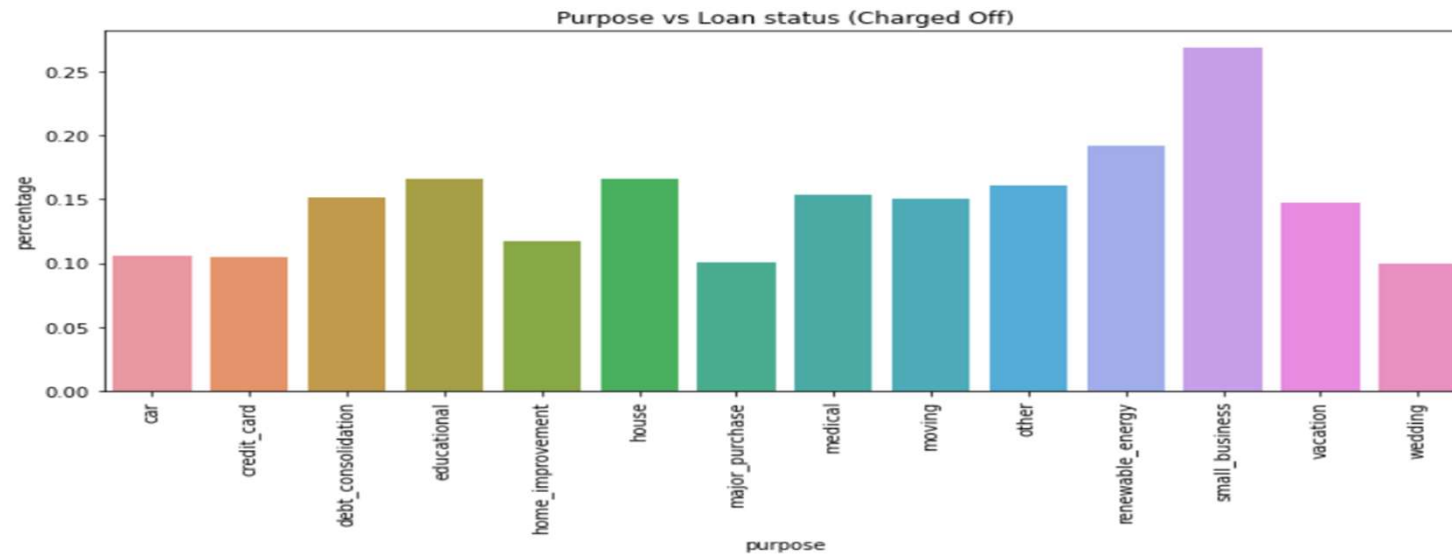
Bivariate Analysis

Grade vs Loan status

- Charged off increases as grades decreases..

Bivariate Analysis
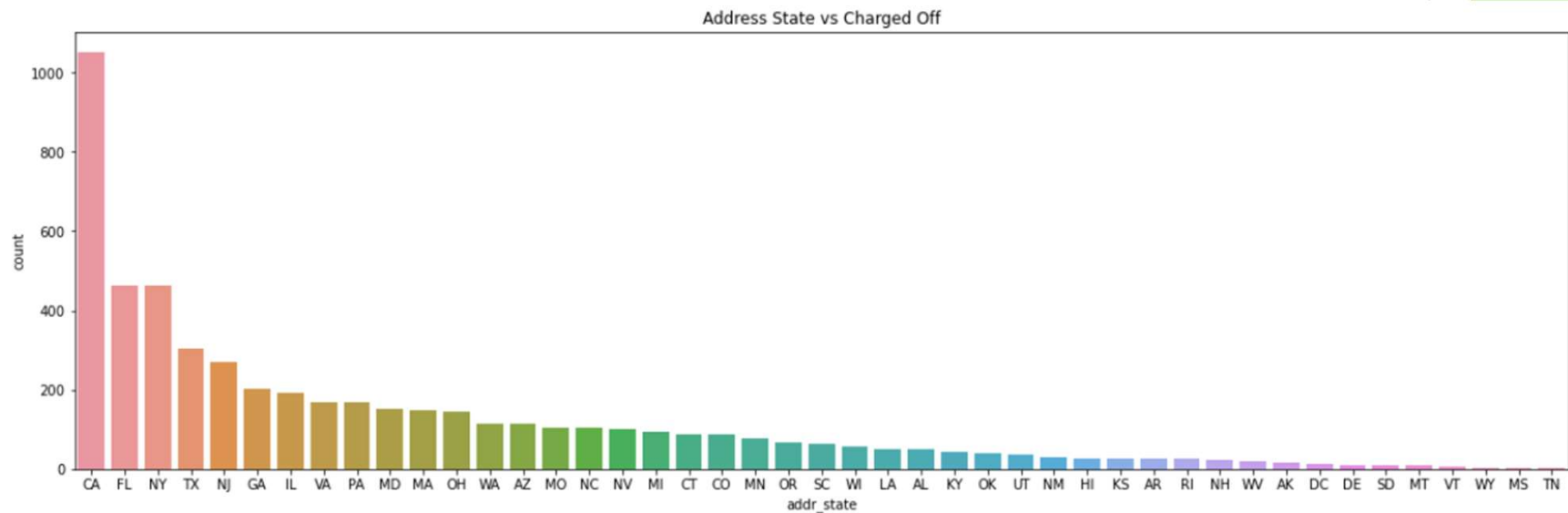

- Borrowers with higher employment lengths and took more loan amounts got more default rate

Bivariate Analysis

Purpose vs Loan status (Charged Off)

- Charged-off are higher for small business comparatively.

Bivariate Analysis

Address State vs Charged Off

- More number of borrowers defaulted in CA , FL and NY states

Recommendations

Recommendations

- Lending Club should avoid to give loan for longer term that is for 60 months

- To avoid more defaulter the interest should not be high, High interest rate have more tendency of getting loan charged off

- Lending Club should resist to provide a loan to lower grades as it have higher chances of default

- The employee who have higher length of employment are more prone to default loan, Lending club should consider the this parameter as it gave more loan to them

- Lending Club should avoid to give loan to the smaller business as they defaulted most

- More number of borrowers defaulted in CA , FL and NY states, there should be more scrutiny to the borrower of that state