Lending Club Case Study

Sonipriya Singh

Anuja Kudumula

Problem Statement

- Our objective is to assist a finance company in making smarter lending decisions by predicting loan defaults from past data. The challenge is to avoid approving loans to borrowers who are more likely to default, which leads to financial losses, and to ensure that reliable borrowers are not rejected, which results in lost business opportunities. By identifying patterns in loan defaults, we aim to improve the company's decision-making process.

Business Objective

- The objective is to use Exploratory Data Analysis to identify the key factors driving loan defaults. By understanding these factors, the finance company can minimize credit losses from bad loans and improve risk assessment. This will help the company make more informed lending decisions, reduce credit loss, and ensure sustainable business growth.

Approach

- Data cleaning and manipulation

- Handling missing values & outliers

- Univariate analysis

- Segmented univariate analysis
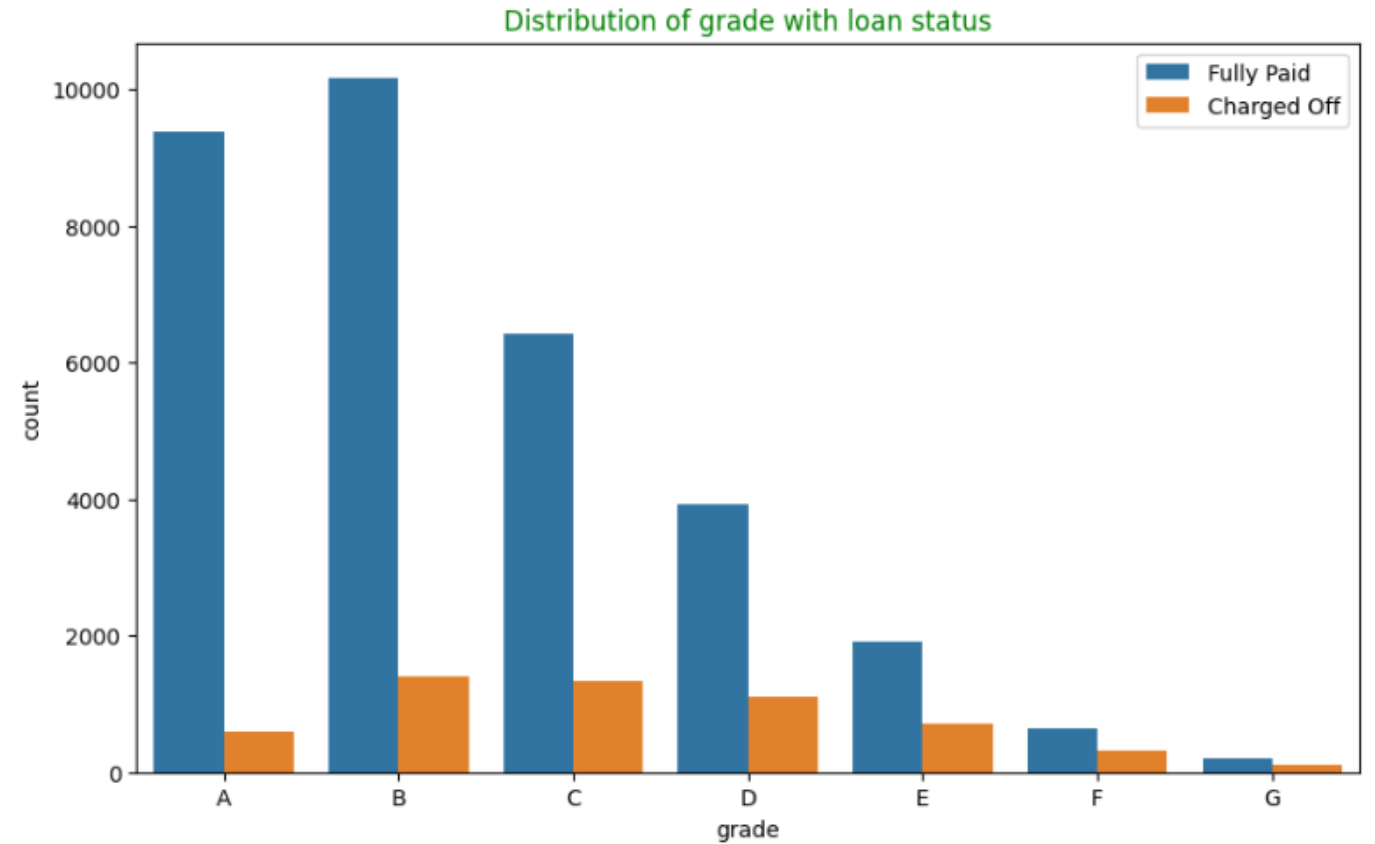
- Bivariate analysis

- Multivariate analysis

Analysis

- **Grade and Interest rate**

- **Grade and loan status**
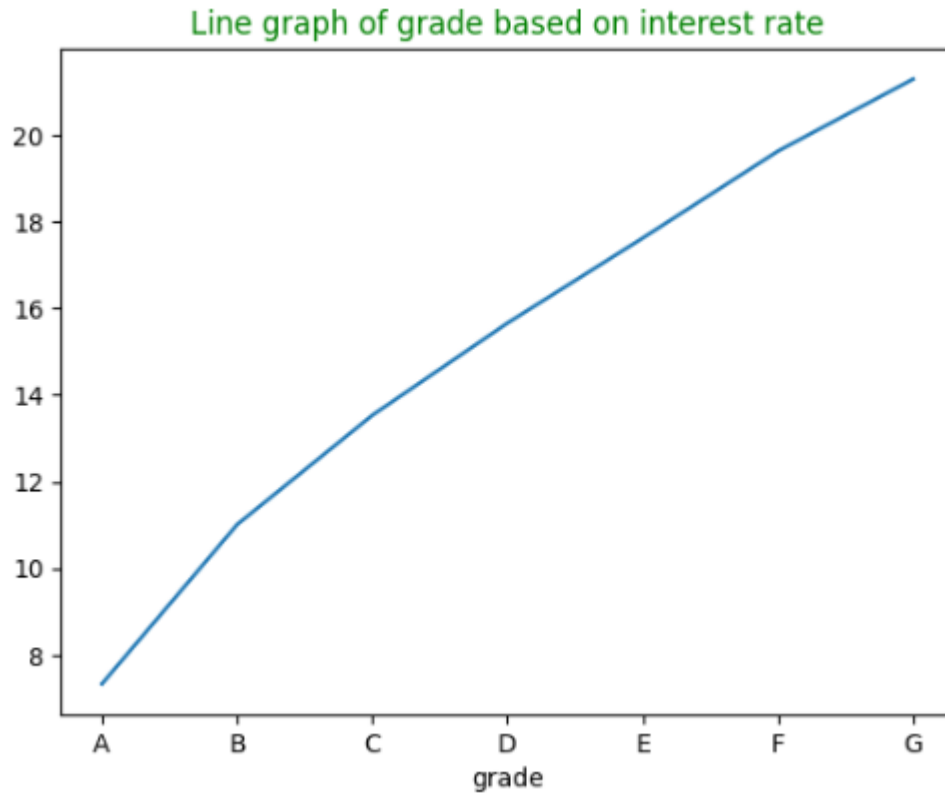
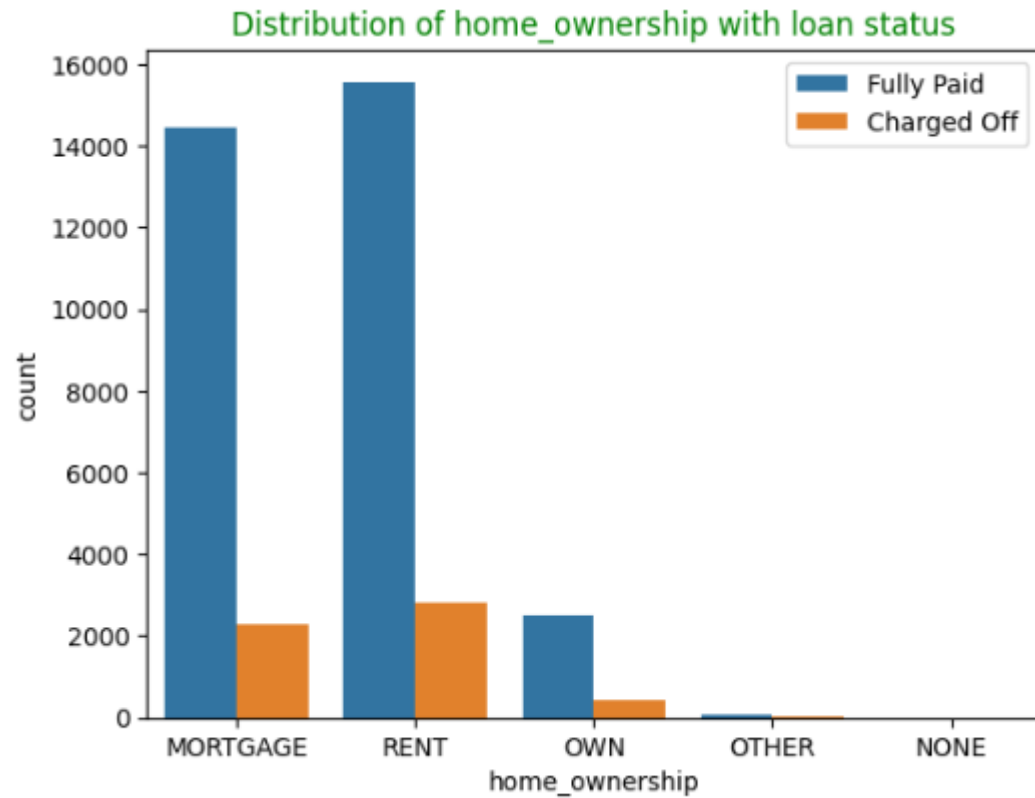As grade increases(meaning from A to G), interest rate has also increased

Borrowers with higher Grade E,F,G have higher chance of defaulting the loan.
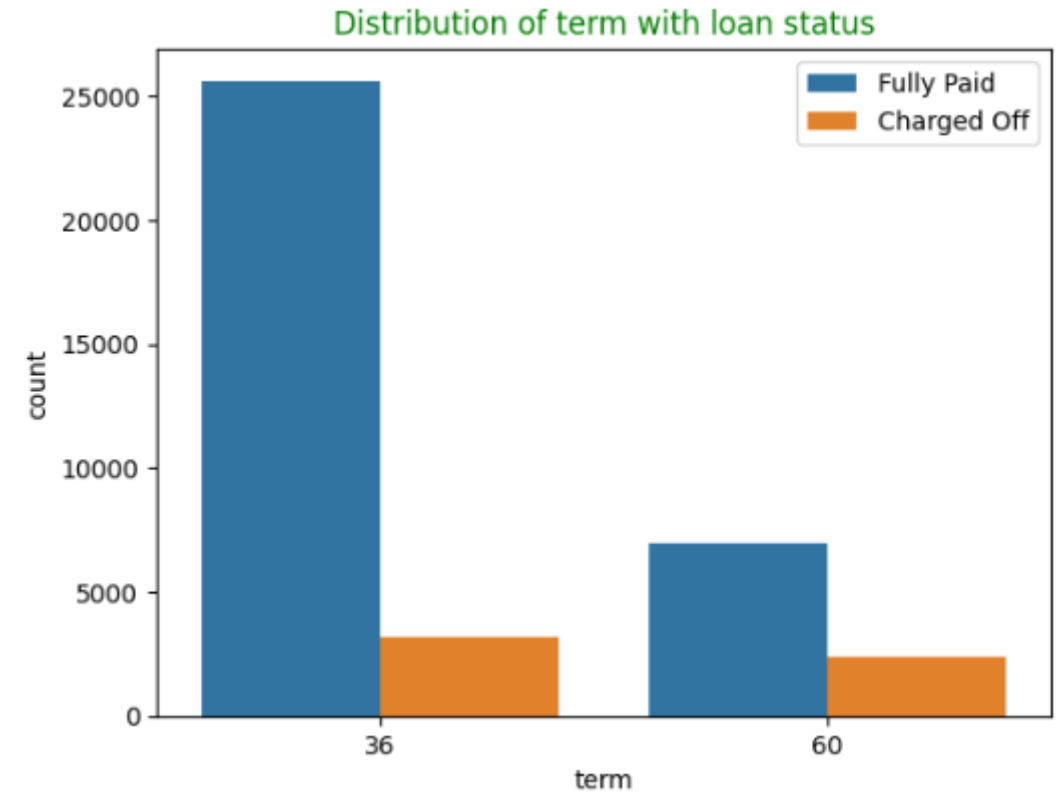
Analysis

- **Home Ownership**

Applicants who don't own a property and on are rent/mortgage are more likely to default the loan.
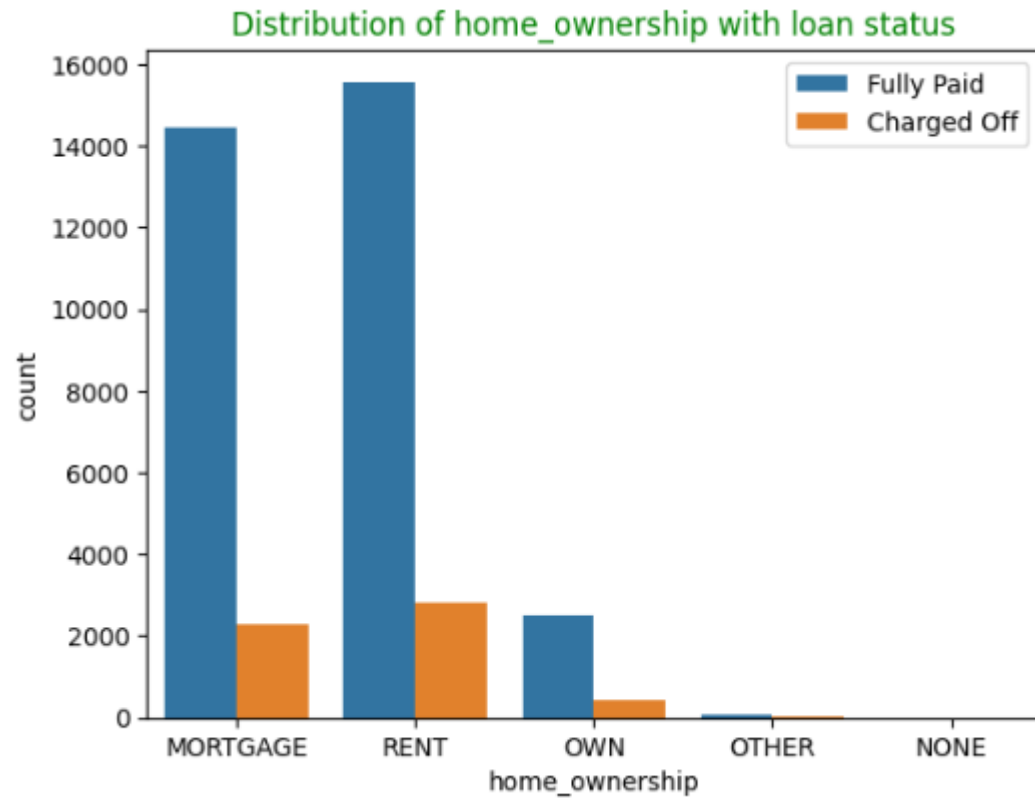
- **Term**

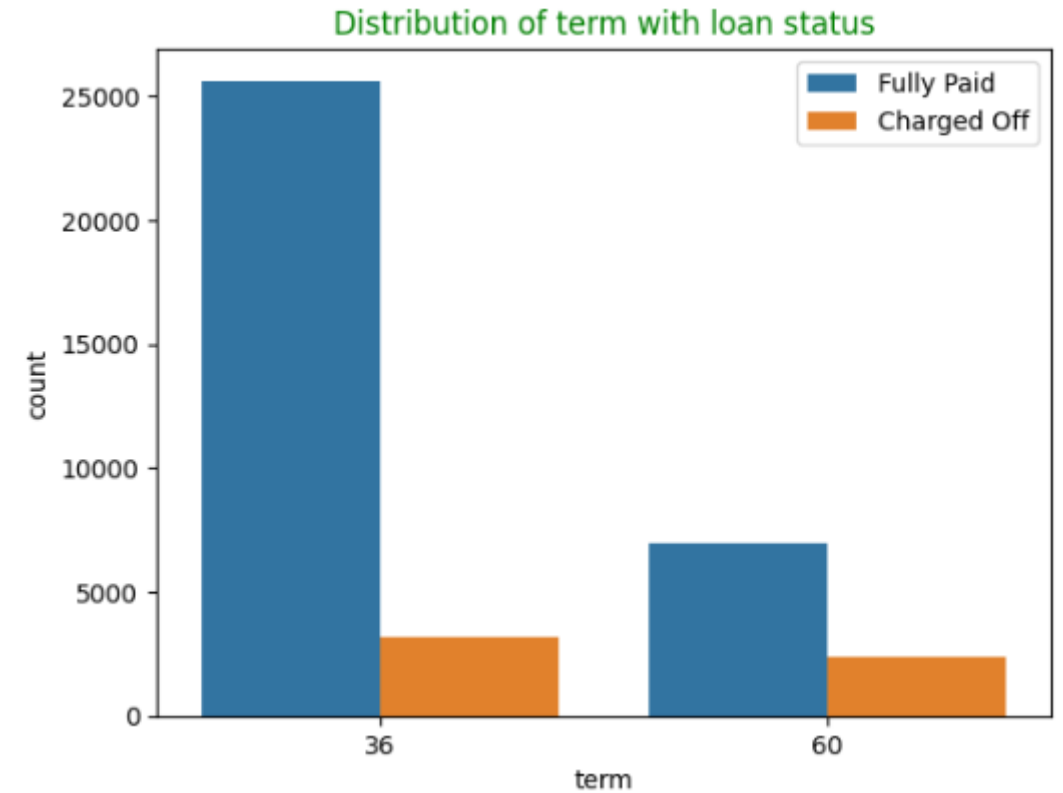Majority of loan given to 36 months tenure and are less likely to default

Analysis

- **Home Ownership**

Distribution of home_ownership with loan status

Applicants who don't own a property and on are rent/mortgage are more likely to default the loan.

- **Term**

Distribution of term with loan status

Majority of loan given to 36 months tenure and are less likely to default

Analysis

- **Employment Length**

Distribution of emp_length with loan status
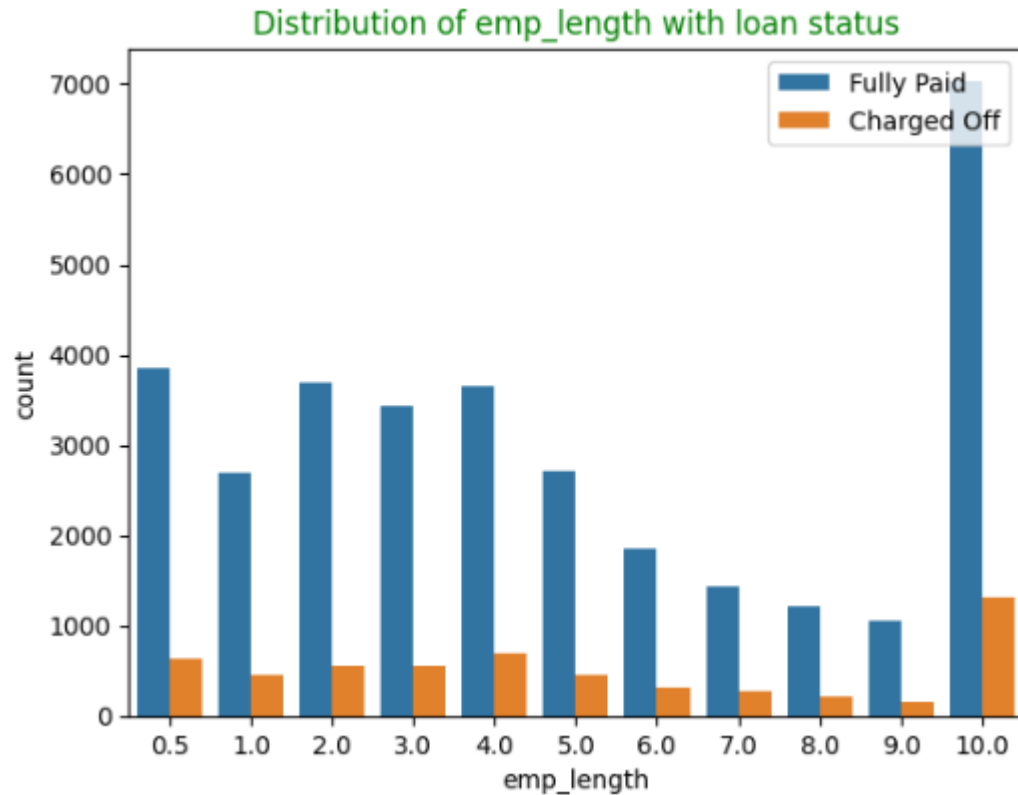
Majority of the borrowers have 10+ years of experience and have highest number of defaulters.

- **Public record bankruptcies**

Distribution of pub_rec_bankruptcies with loan status
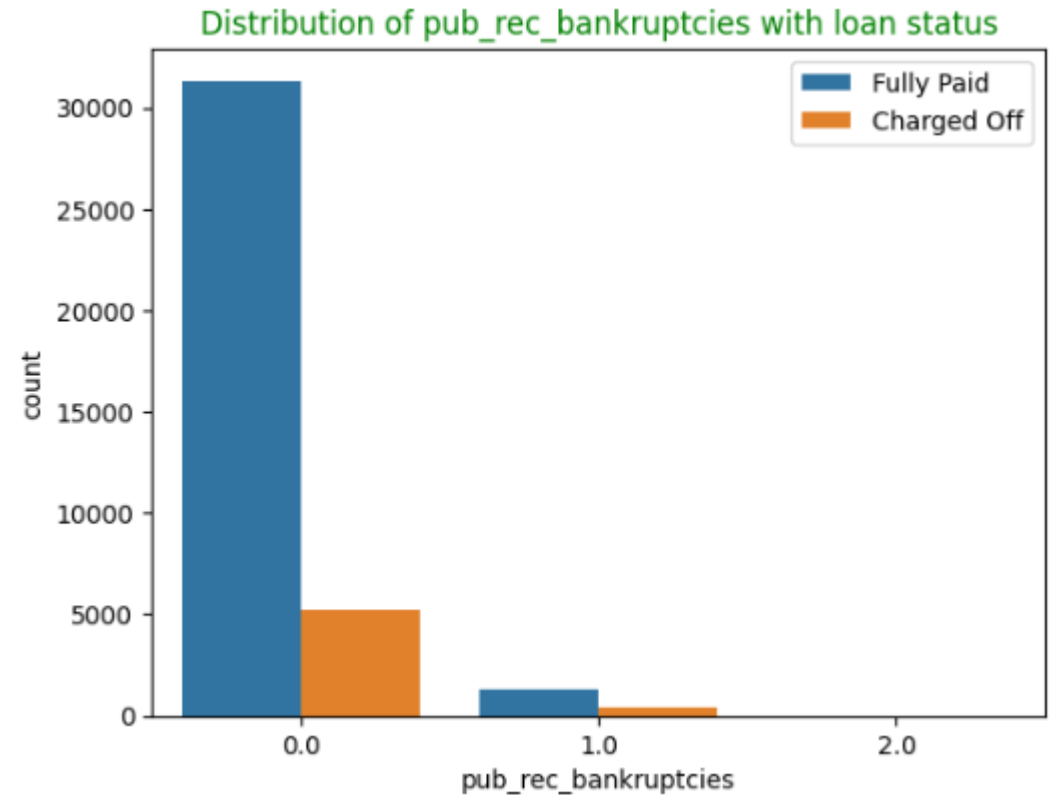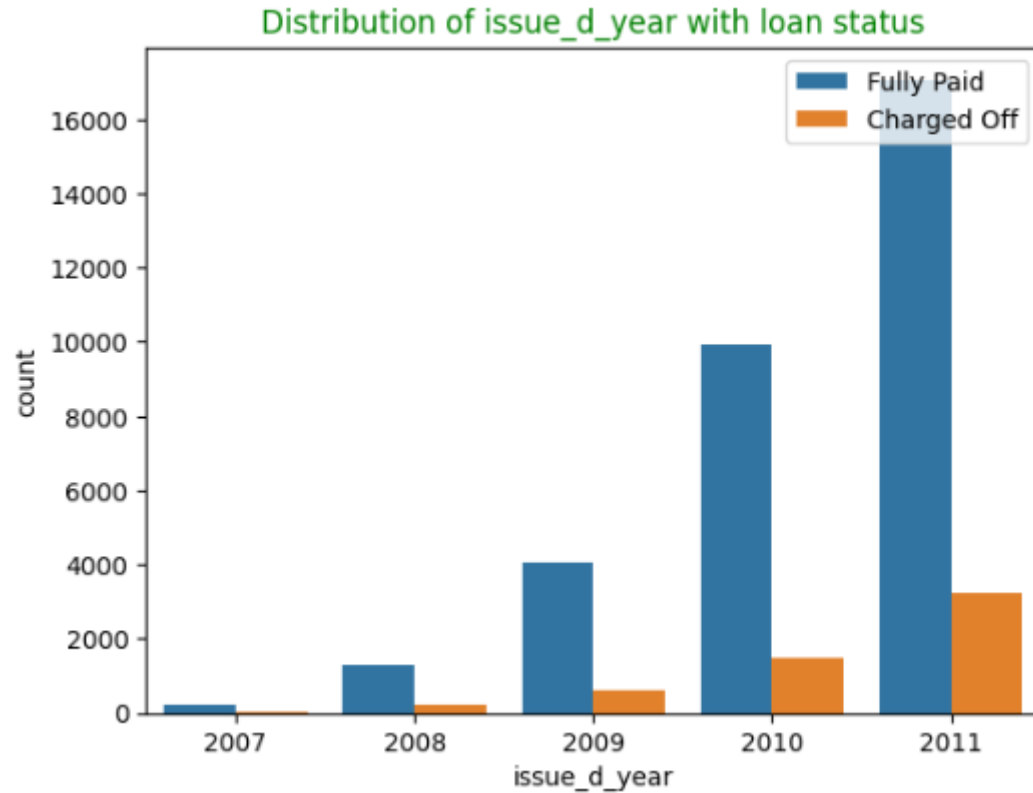
Majority of the borrowers have no history of bankruptcy. Borrowers with large number of public records bankruptcies are more likely to default.
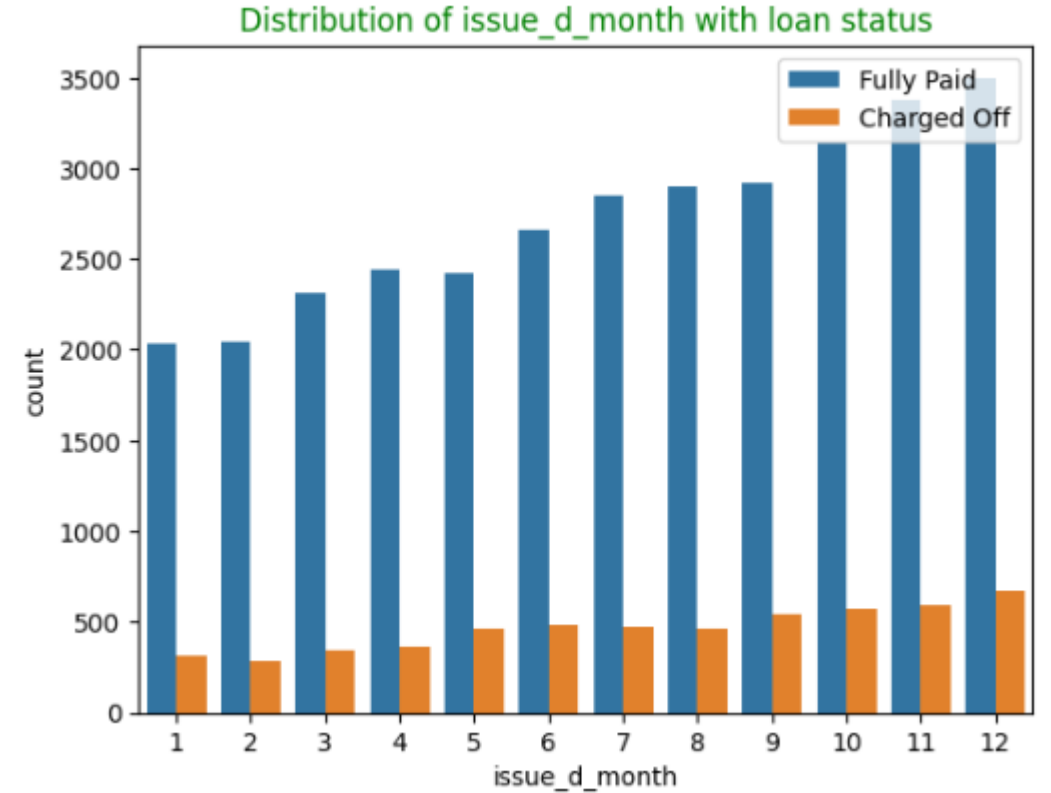
Analysis

- **Loan trend over years**

Distribution of issue_d_year with loan status

Number of loans issued have increased from 2007 to 2011.

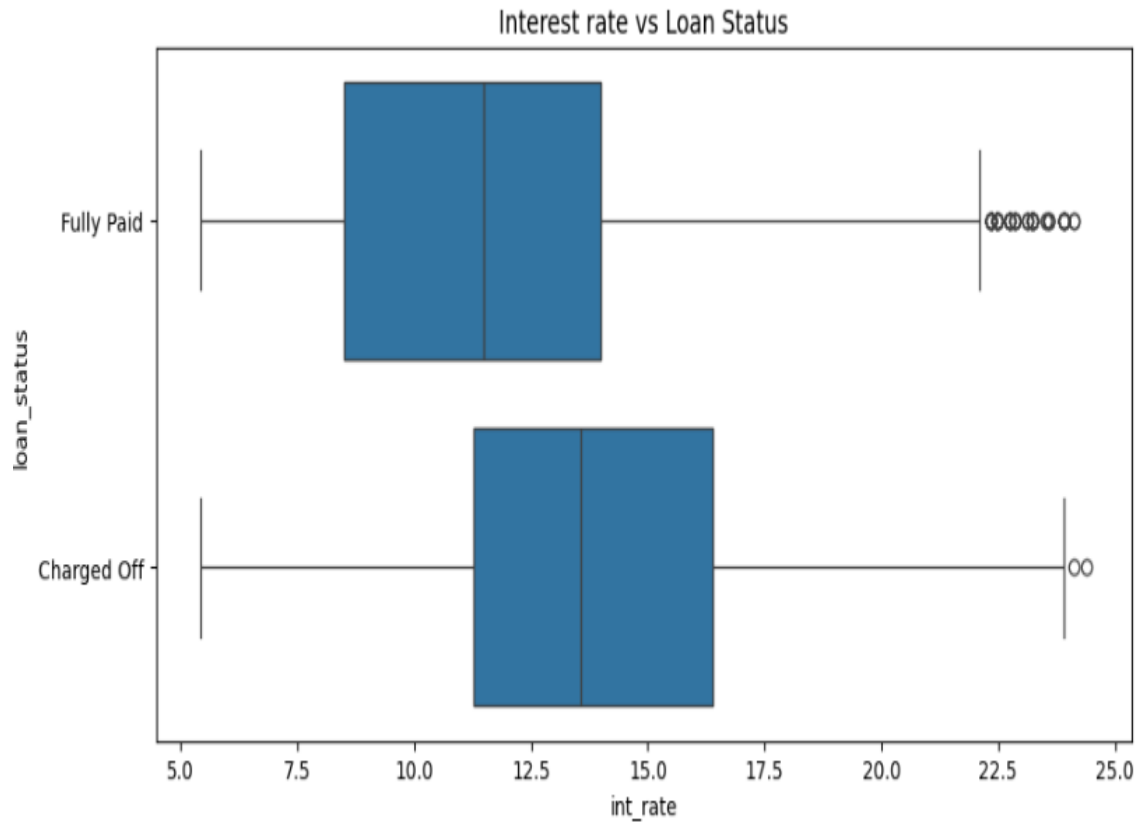- **Loan trend over months**

Distribution of issue_d_month with loan status
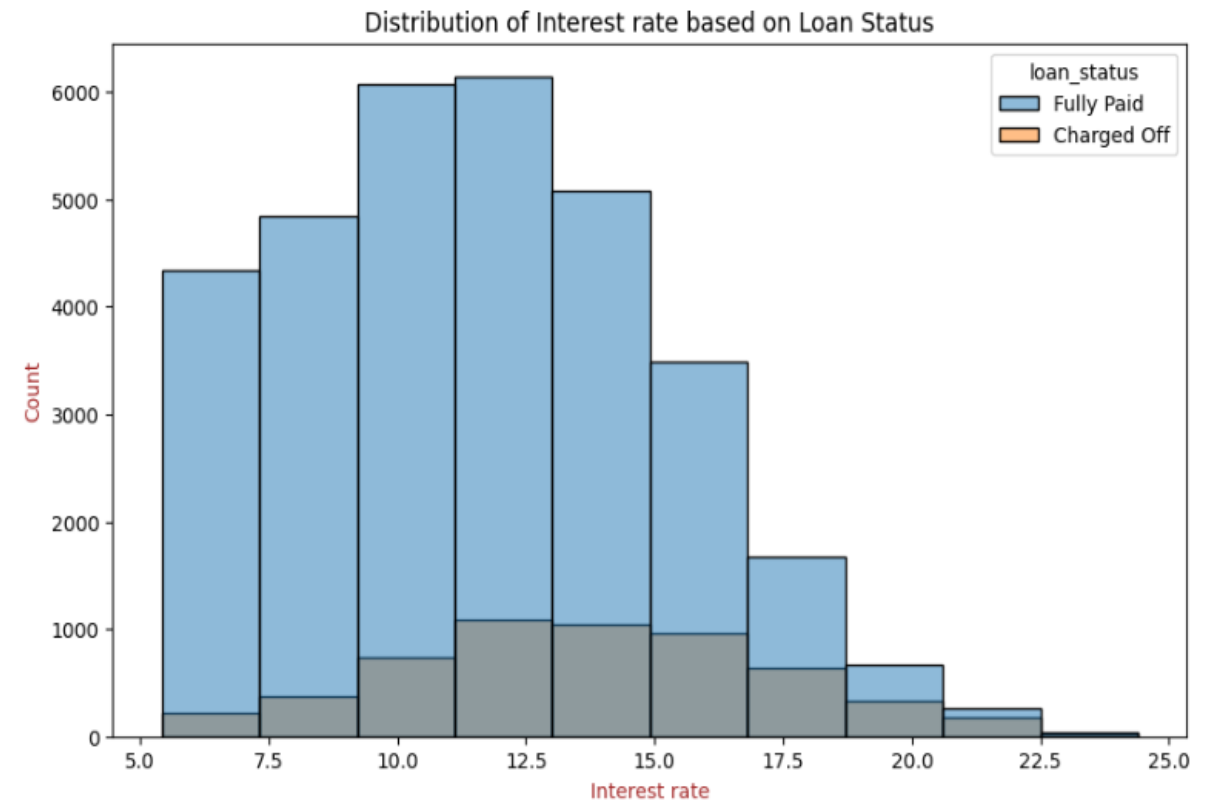
Number of the loans issued have increased in the last quarter.

Analysis

- **Interest rate over loan status**

High interest rate borrowers are more likely to default.
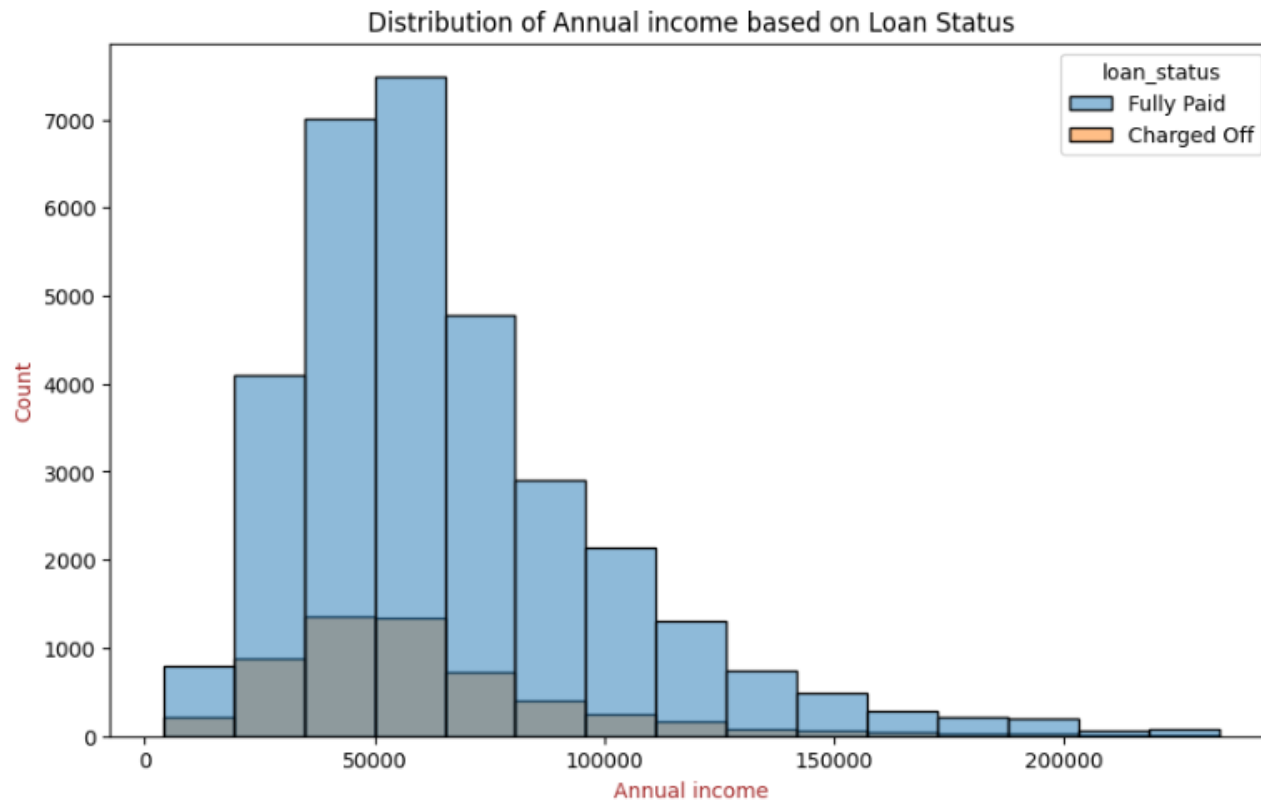
- **Interest rate over loan status**

Majority of borrowers are in the interest range 10-15.

Analysis

- **Annual income**

Distribution of Annual income based on Loan Status
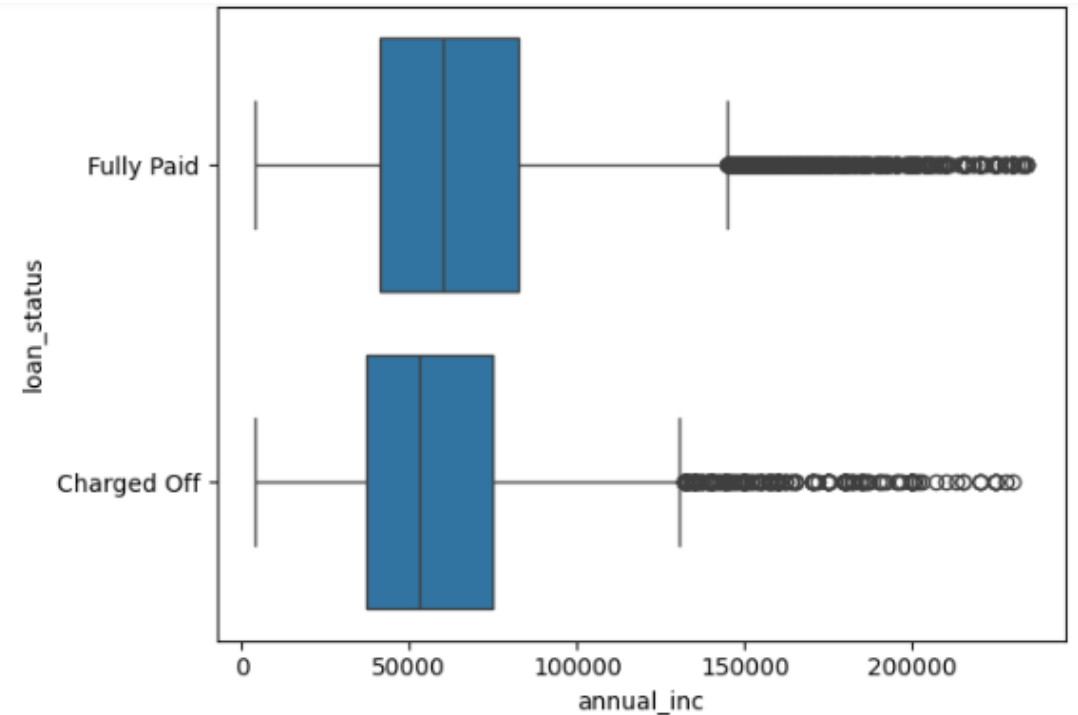
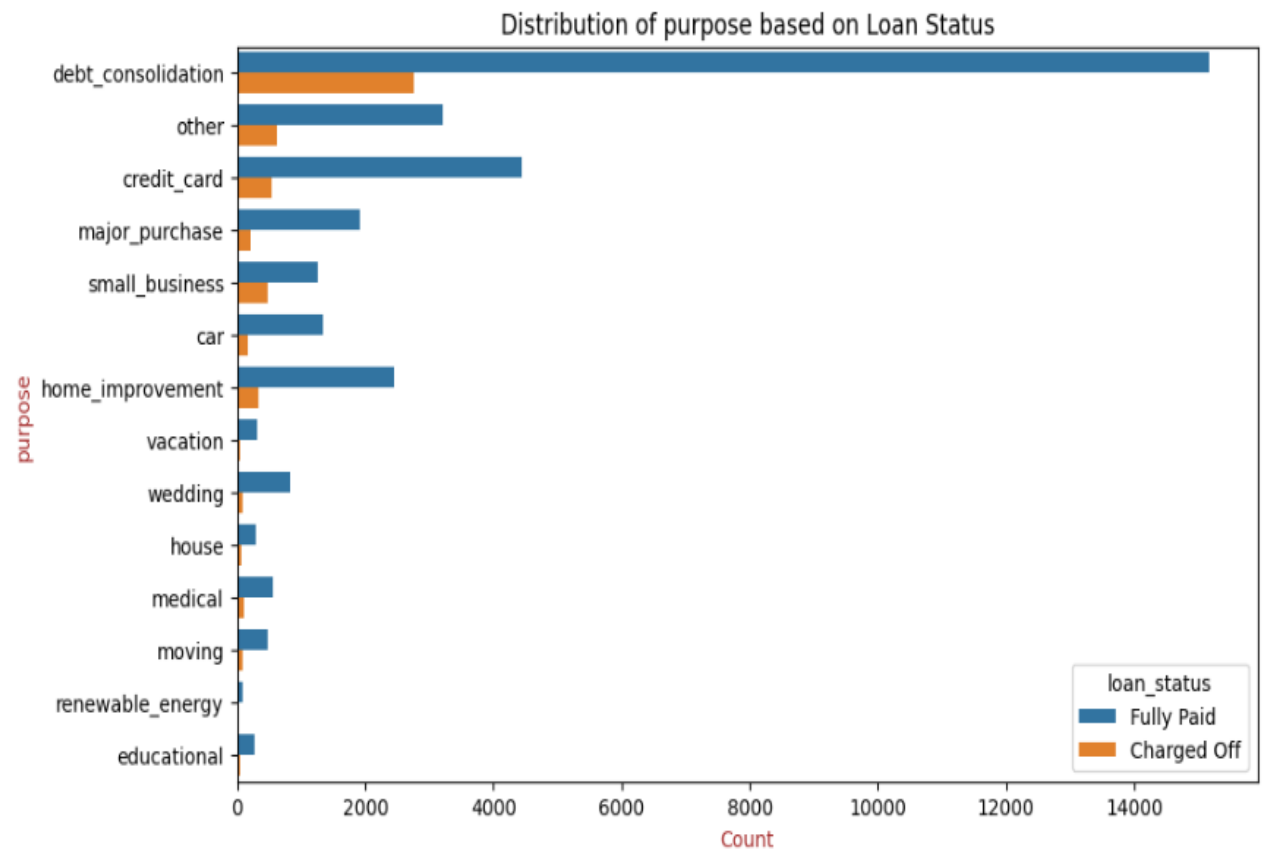Majority of Borrowers are with annual income around 50000.

- **Annual income based on loan status**

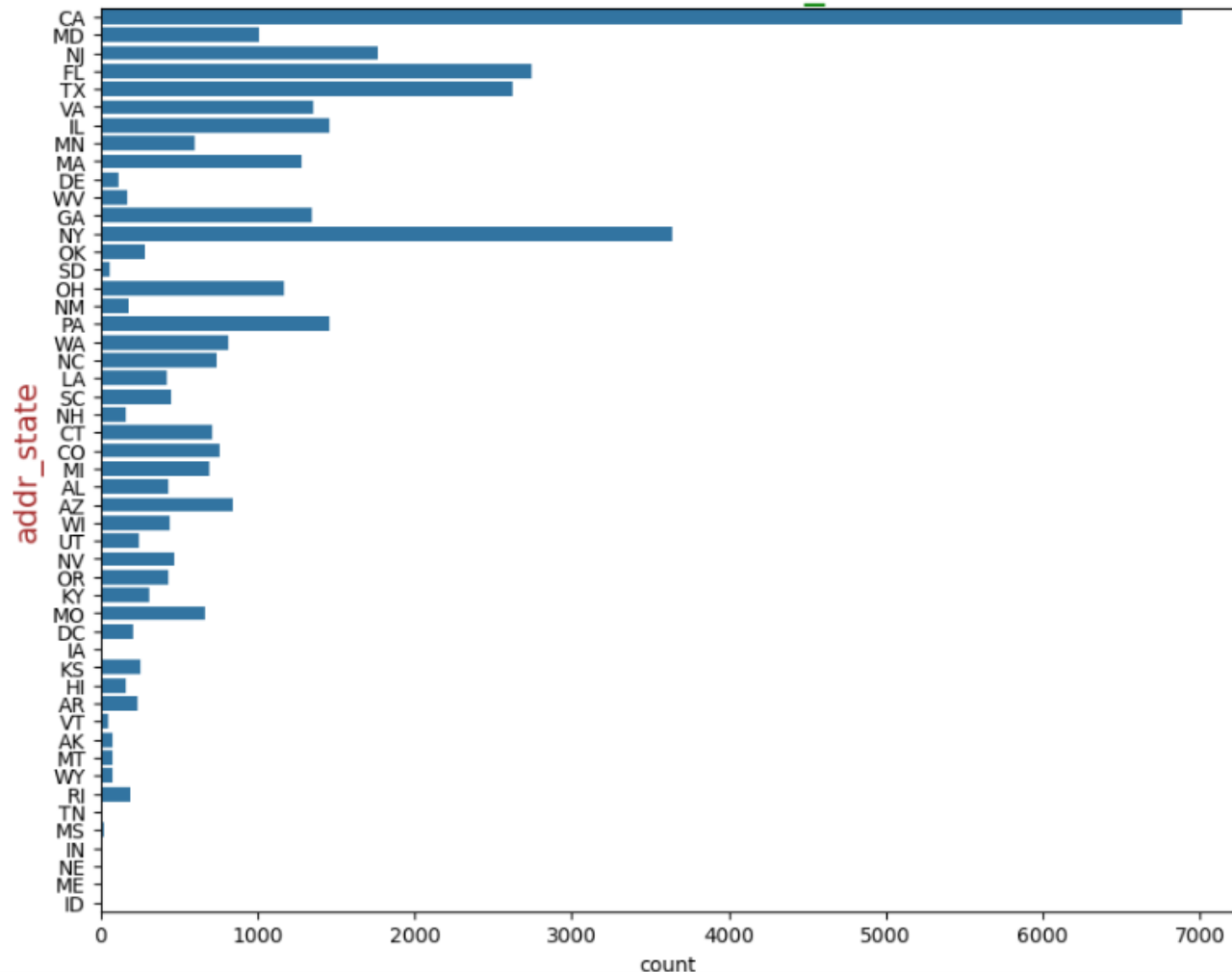People with low annual income are more likely to default.

Analysis

- **Loan Purpose**

Majority of loans were taken for the purpose of debt_consolidation followed by credit_card.
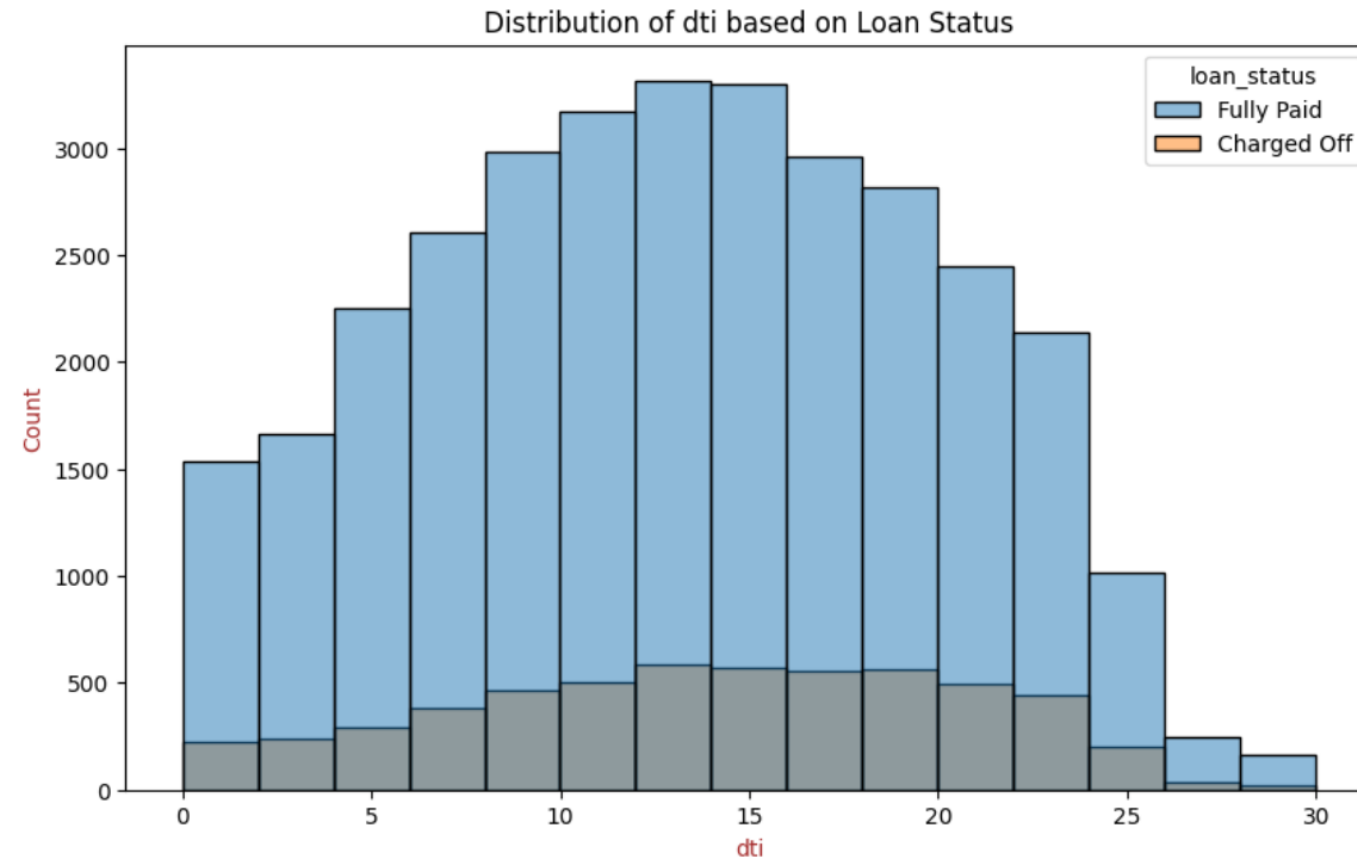
- **State of the borrower**

Majority of borrowers are from CA followed by NY, FL, TX.

Analysis

- **dti**
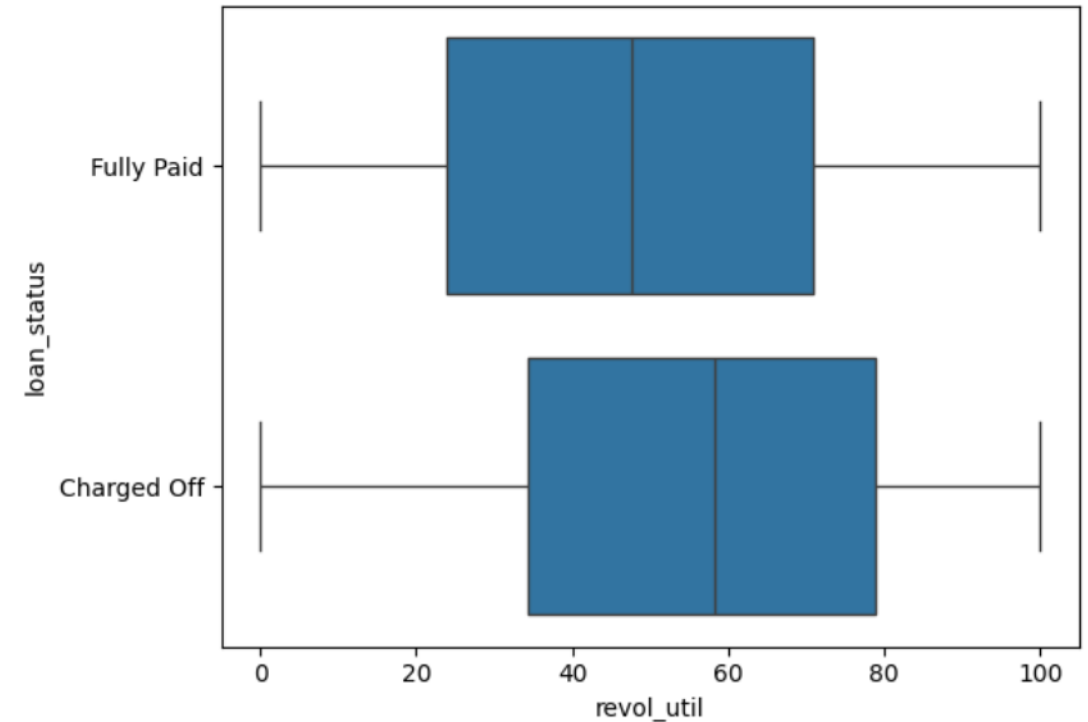
- **Revolving line utilization**

Distribution of dti based on Loan Status

Majority of the borrowers are with dti range 10 -15, the defaulters are also more.

Borrowers having high revolving utilization are more likely to default.

Analysis

- **Correlation matrix**

correlation/heatmap between all the numerical variables
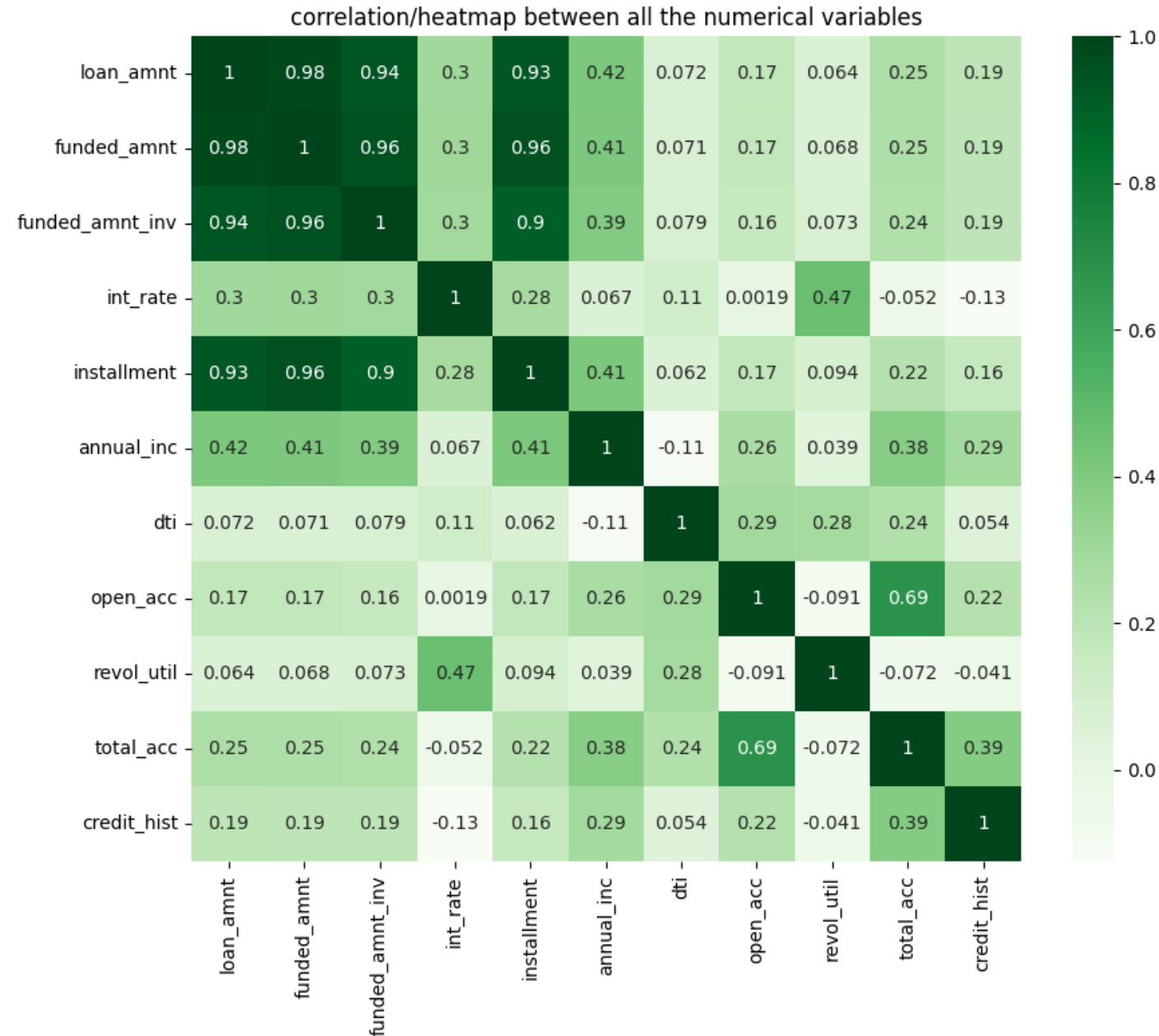
There is strong correlation between installment and loan_amnt, funded_amnt and funded_amnt_inv. These parameters are directly proportional to one another.

Conclusion

Below are the Major Driving factors which can be used to predict the loan deafulters.

- home_ownership

- Grade

- revolving_util

- int_rate

- Pub_rec

- annual_income