Lending Club Case Study

Submitted by:
Ambily M
Subhash G

Problem Statement

- A consumer finance company 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
- The aim is to identify patterns from the given dataset which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Process Overview

Data Understanding

- Analysing the data set with the help of data dictionary
- Understanding the domain

Data Pre-Processing

- Missing value analysis and treatment
- Column Data type alignment
- Removing irrelevant rows and columns
- Outlier treatment

Derived Metrics

- Derive date metrices
- Derive new columns from existing attributes
- Convert the existing column to more meaningful with a derivation

Univariate Analysis

- Analysing the frequency and distribution of continuous and categorical variables
- Segmented univariate analysis wrt. Loan Status

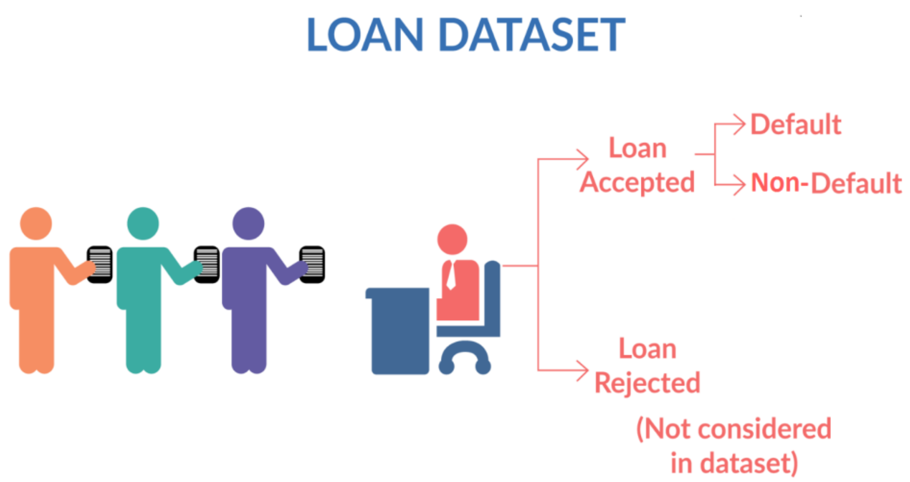
Bivariate Analysis

- Analysing the loan status with two or more features
- Analysis for both continuous and categorical variables

Conclusion

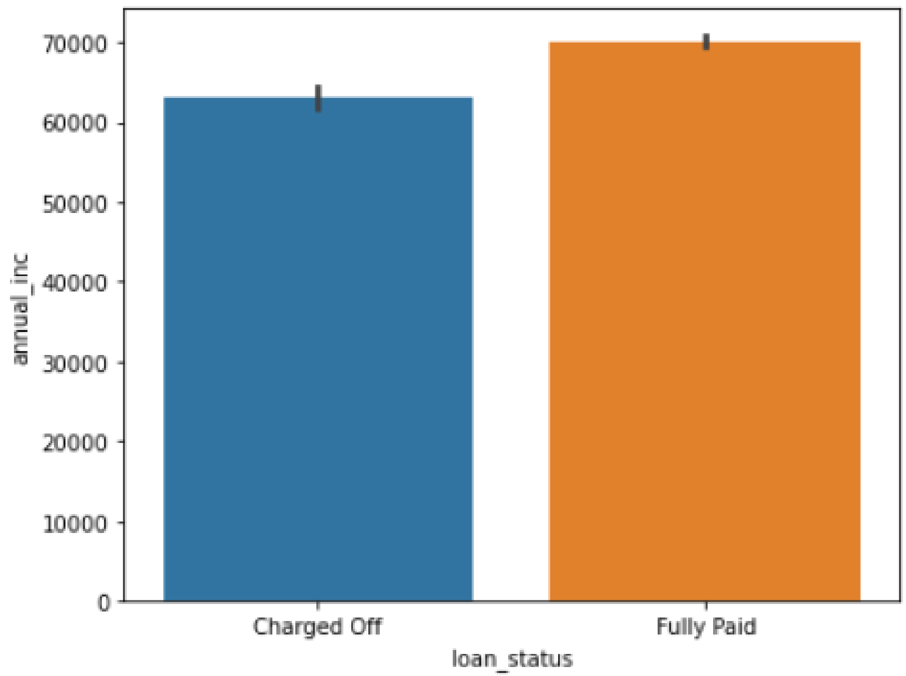
- Outcome of analysis of various plots
- Pointers to loan defaulters

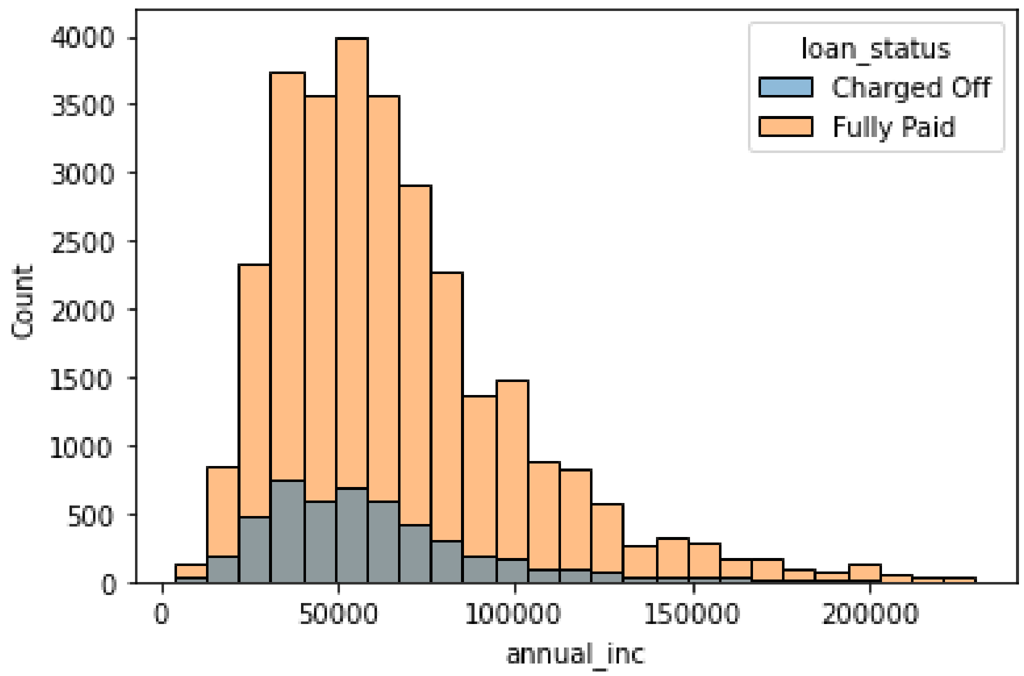
Data Set



- Loan dataset consists of 5 year credit history of borrowers of lending club for all the approved loans.
- The rejected loan data is not provided.
- Approved borrowers are categorized as below:
  - **Fully paid**: Applicant has fully paid the loan (the principal and the interest rate)
  - **Current**: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.
  - **Charged-off**: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has **defaulted** on the loan
- Current applicants are removed from further analysis, as their attributes doesn’t give indication of default
- Data set consists of attributes categorising the borrower as well as the sanctioned loan.

Analysis – Annual Income

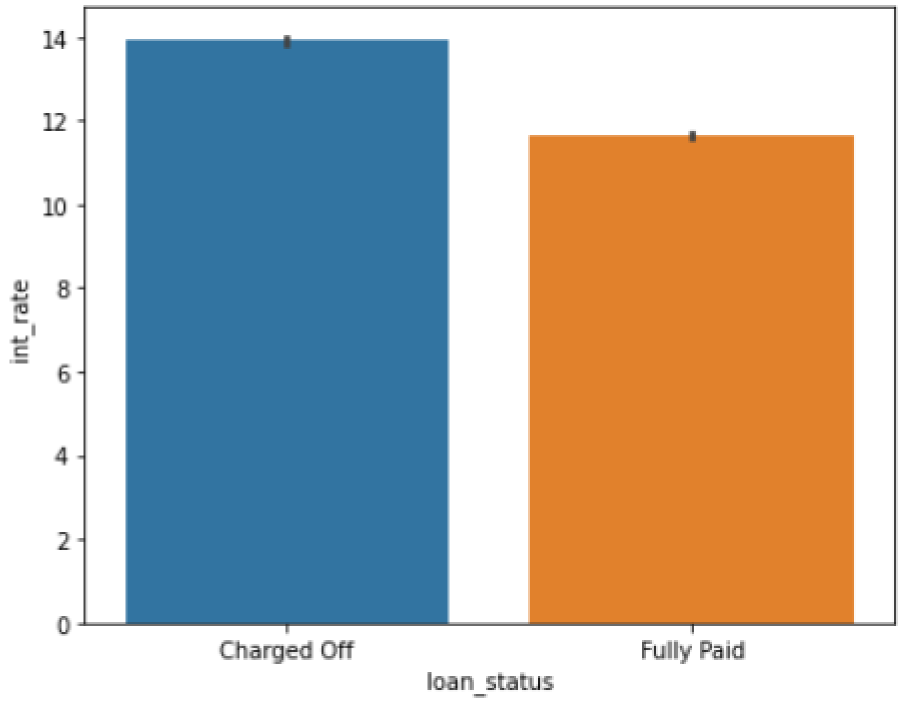


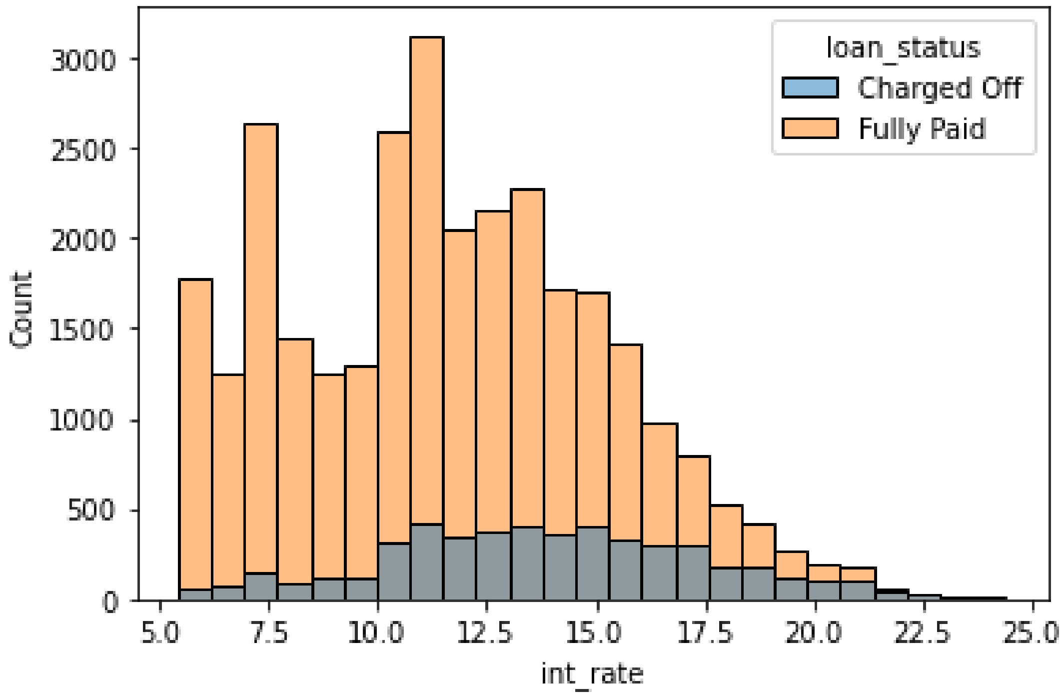


Observations:

- More borrowers are having lesser annual income and are prone to default.

Analysis – Rate of Interest

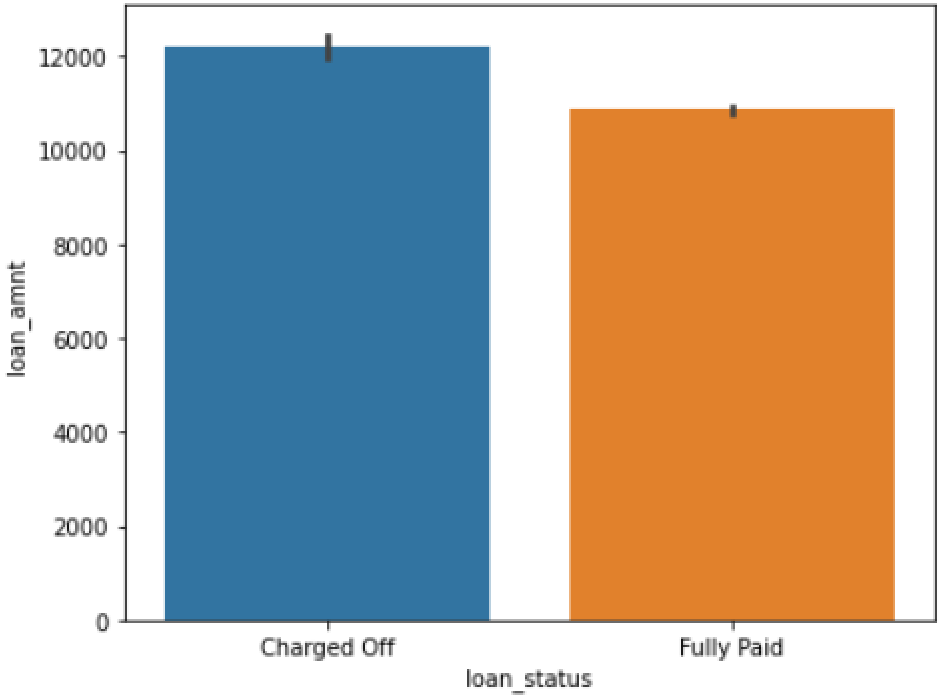




Observations:

- Borrowers taking loan at less interest are repaying the loan.
- As interest rate increases (>= 11%) the borrowers are prone to default.

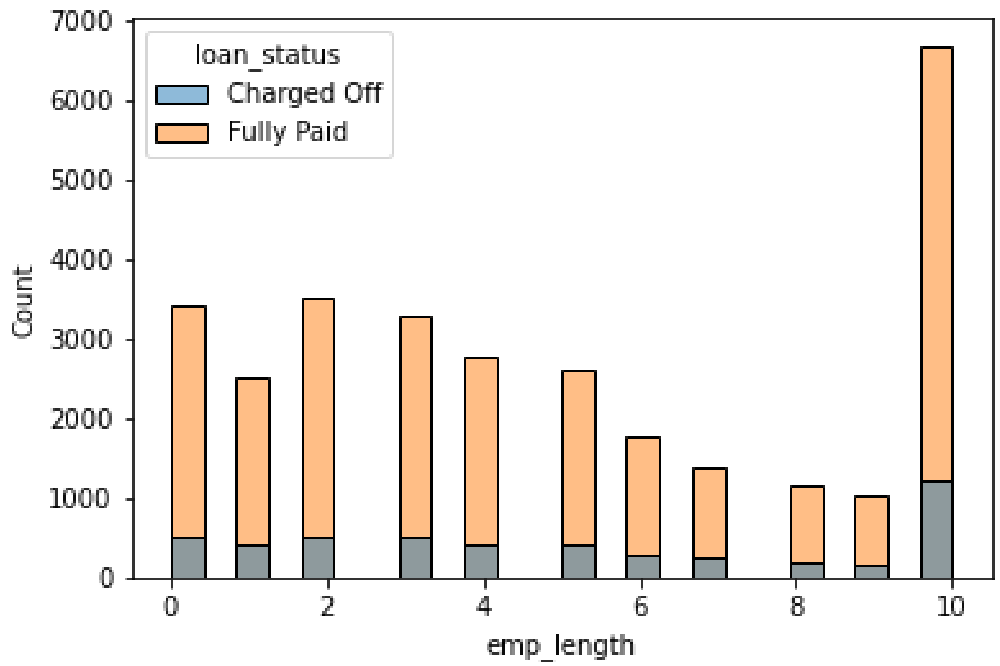
Analysis – Loan Amount

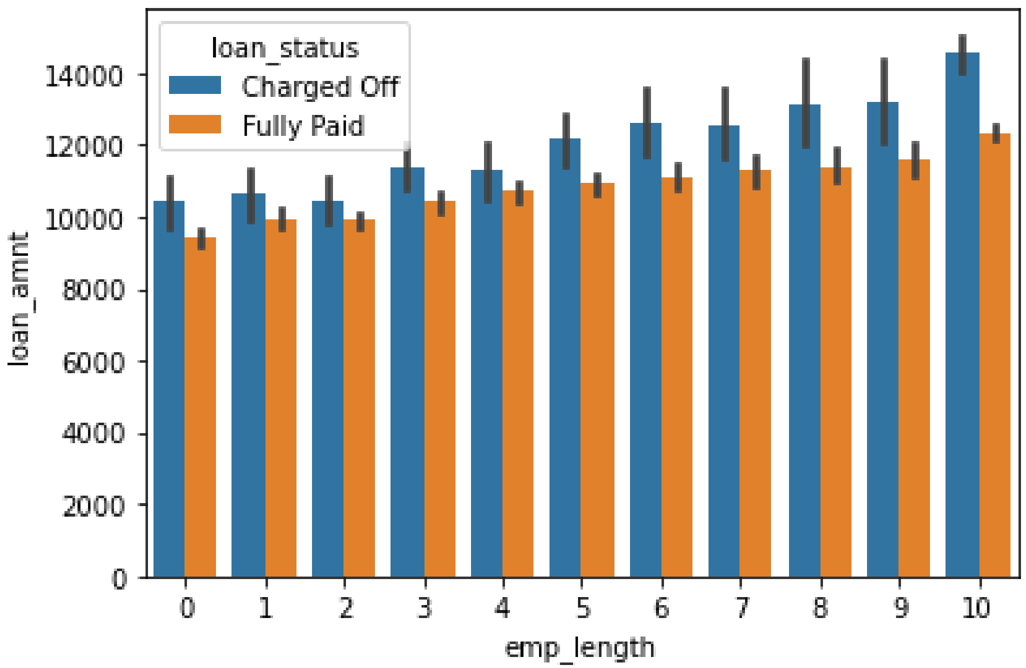


Observations:

- Higher the loan amount, borrowers are prone to default

Analysis – Experience Level

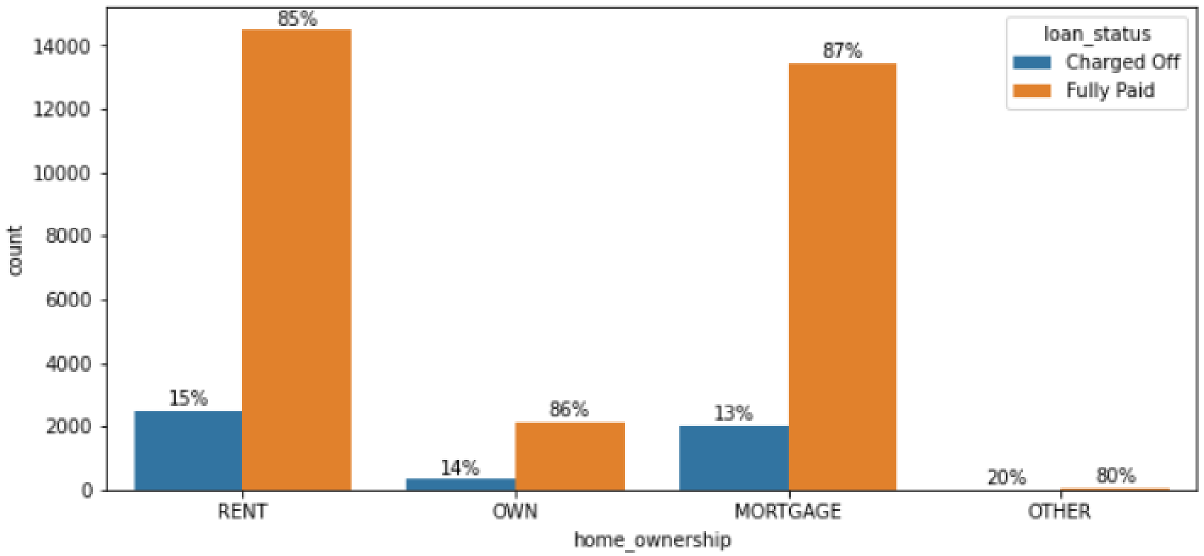


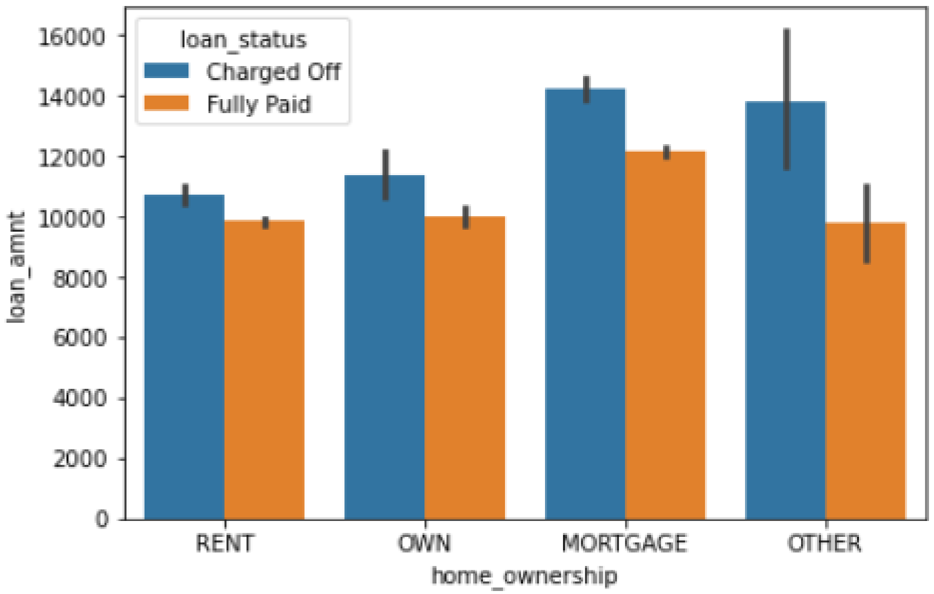


Observations:

- Most of the borrowers have experience of >=10 years and are prone to default.
- Borrowers with higher experience are given higher loan amount, hence are prone to default.

Analysis – Home Ownership

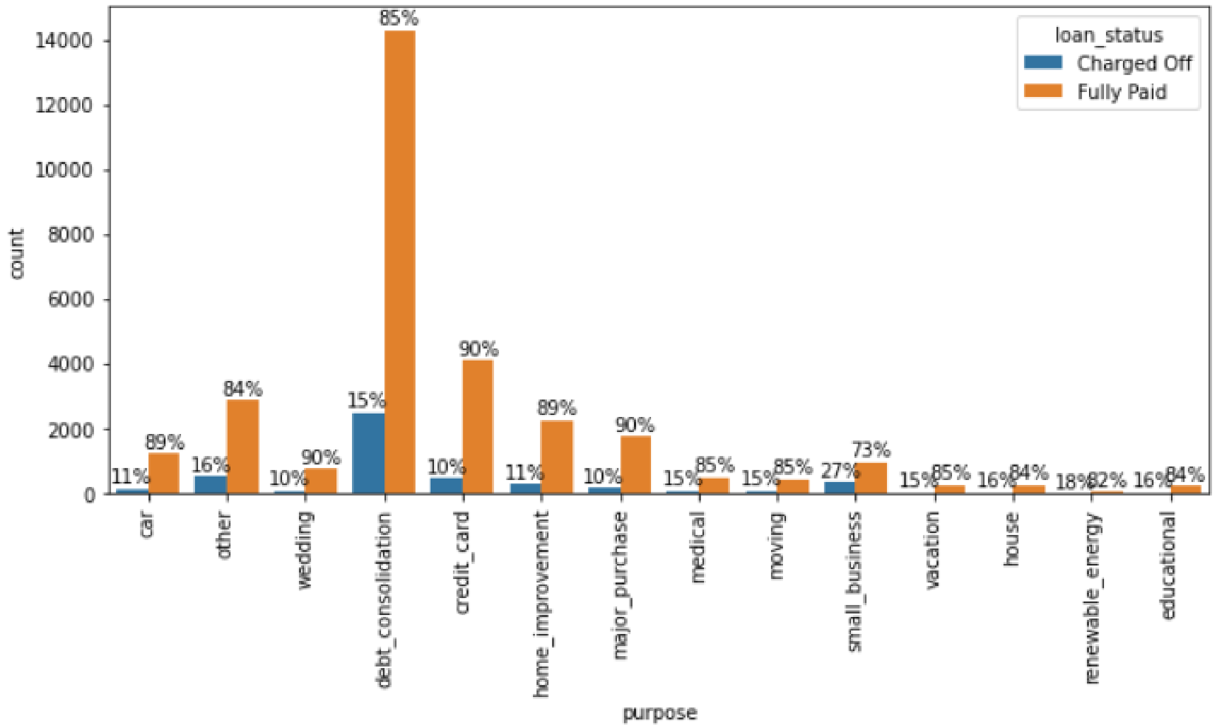


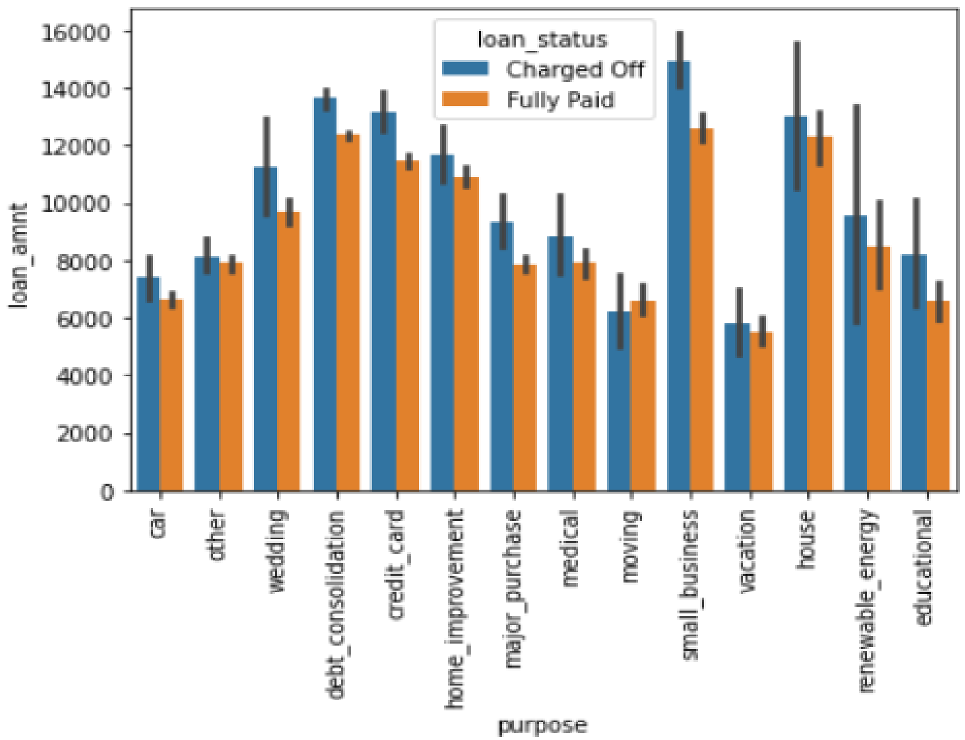


Observations:

- Most of the borrowers have rented house or house which is mortgaged.
- Higher loan amount sanctioned for borrowers with Mortgaged houses or houses in Other category are prone to default, compared to the borrowers with own or rented house

Analysis – Purpose of Loan

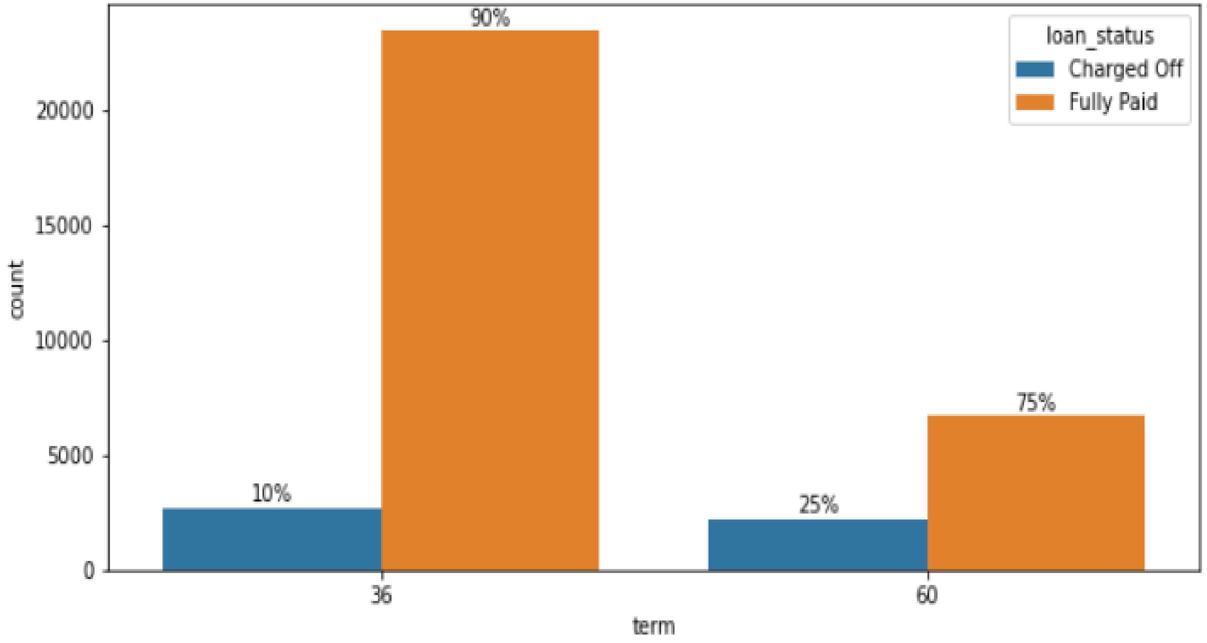


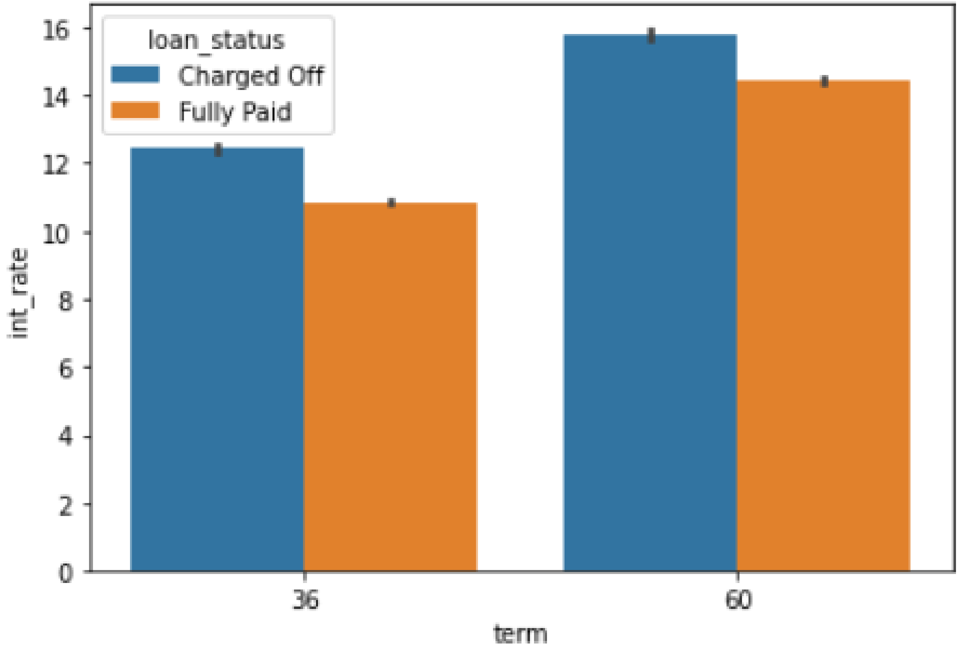


Observations:

- Most of the borrowers apply loan for debt consolidation and has higher number of defaulters.
- Higher loan amount given for small business purpose are less likely to repay loan.

Analysis – Loan Tenure

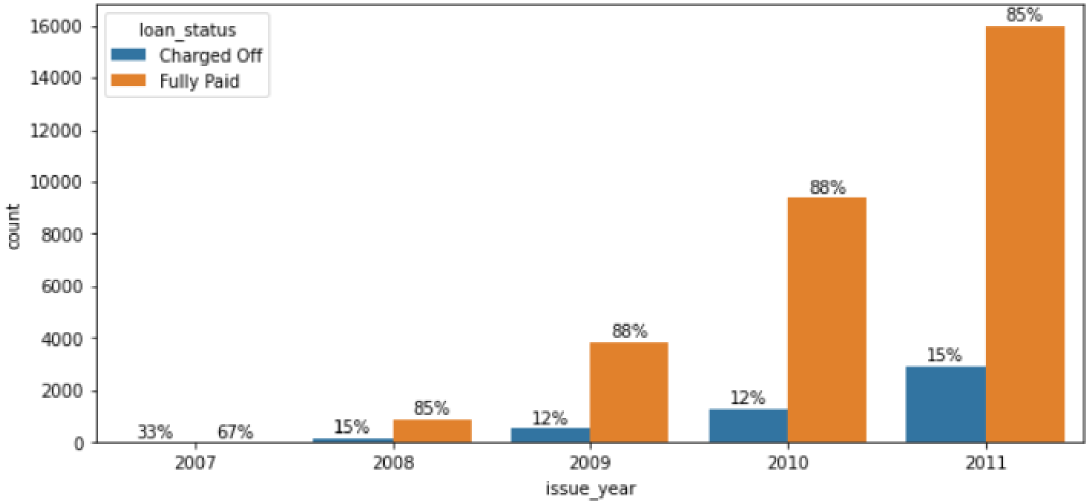


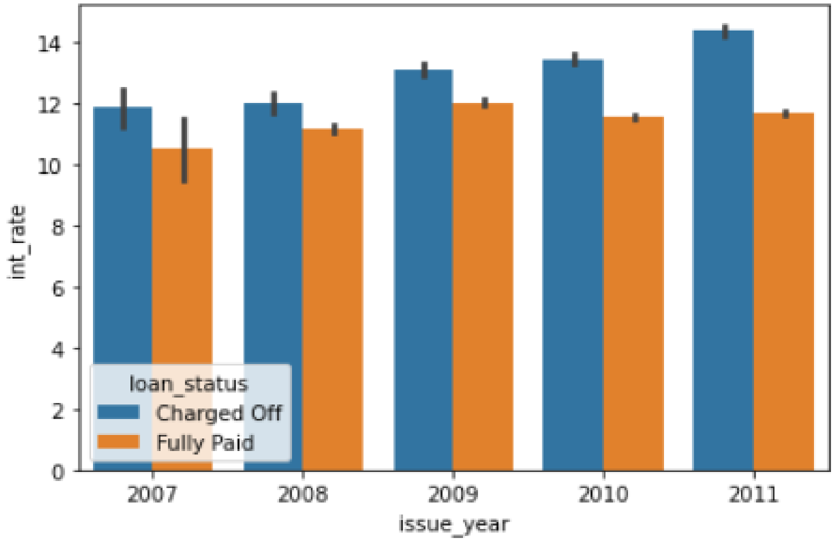


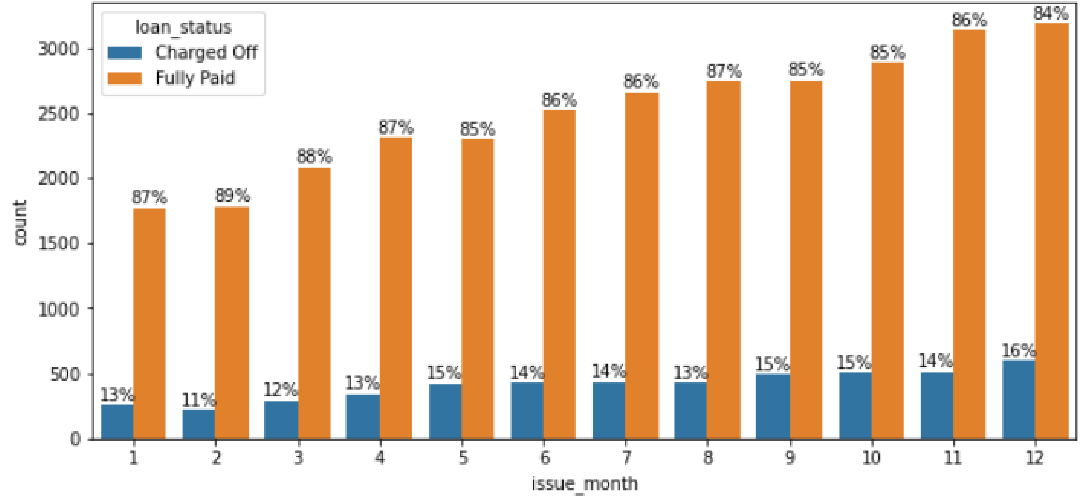
Observations:

- Borrowers with less loan tenure are fully repaying the loan.
- Longer loan tenure increases the chance of more defaulters.
- Interest rate increases for long tenure loan and is prone to default.

Analysis – Loan Issue Date



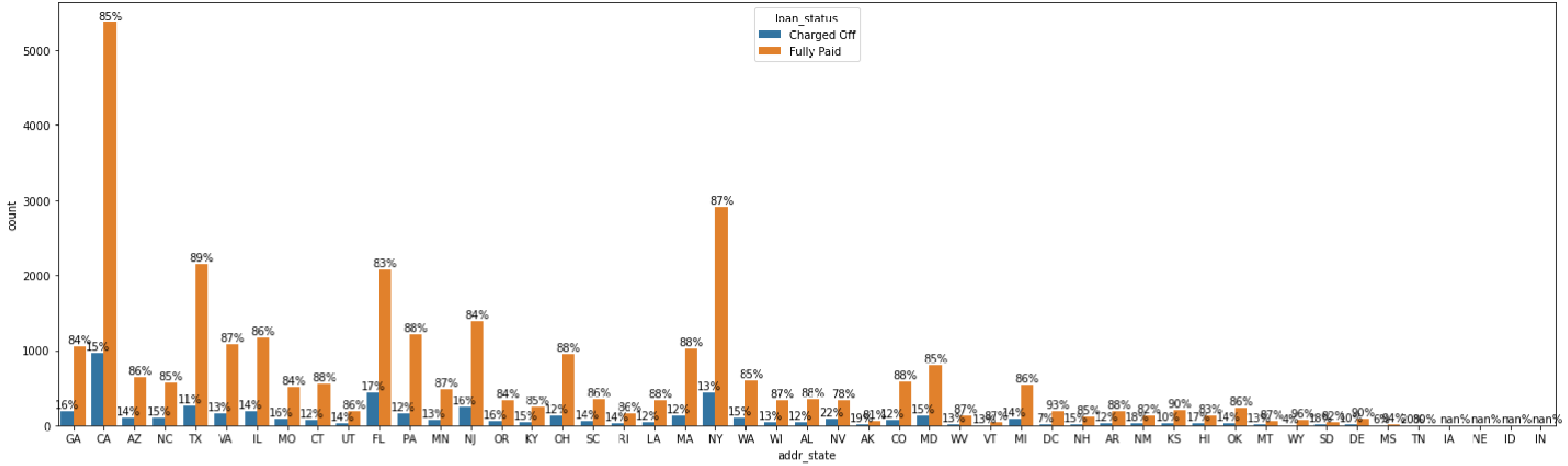




Observations:

- Number of borrowers are increasing Year-on-Year and Month-on-Month, with- increase in defaulters.
- Increase in rate of interest YoY contributing to loan defaulters.

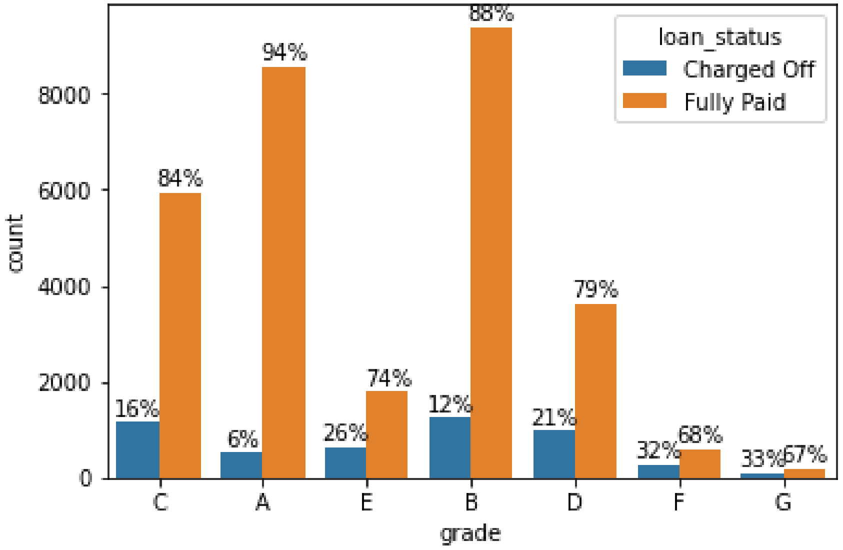
Analysis – State

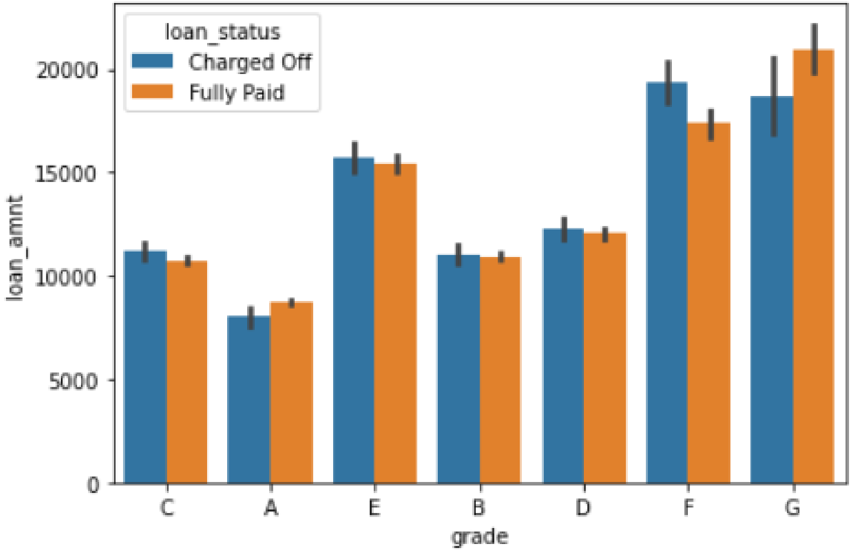


Observations:

- More borrowers are from State “CA” and hence contributes to more number of defaulters.
- Percentage wise every state has similar distribution of defaulters.

Analysis - Grade

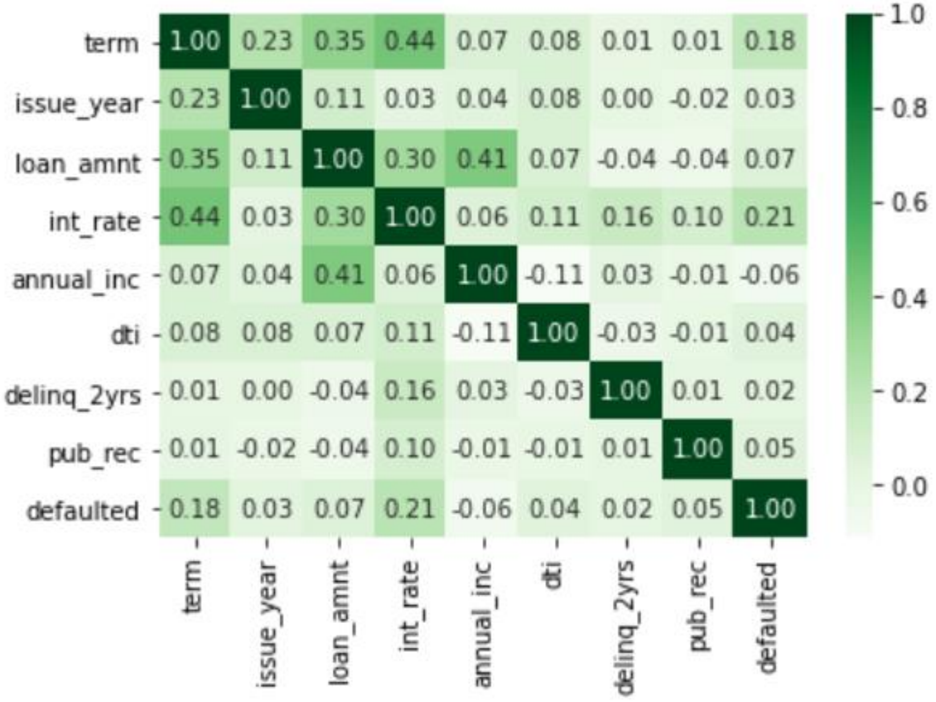




Observations:

- More borrowers are from Grades A,B and C, and their loan amount is less, hence are repaying the loan.
- Higher loan amount given to Grade F borrowers are prone to default.

Analysis – Correlation Matrix



Observations:

- Higher the loan tenure and Higher the Interest rate, chances of default
- Higher the annual income, high loan amount sanctioned
- Higher the loan tenure, the interest rate and loan amount are higher
- Annual Income increases debt-to-income ratio reduces

Conclusion

- **Higher Interest Rate :**

✓ Rate of interest increases year-on-year

✓ Borrowers with higher Interest Rate (>=11%) are prone to default, irrespective of the category/segment they belong to.

- **Higher Loan Amount :**

✓ Highly experienced borrowers are given loan with higher amount, hence are prone to default.

✓ Higher loan amount sanctioned for borrowers with Mortgaged houses or houses in Other category are prone to default, compared to the borrowers with own house or rented house.

✓ Higher loan amount given for small business purpose are less likely to repay loan.

✓ Higher loan amount given to Income Grade 'F' borrowers are prone to default.

- **Higher Loan Tenure :**

✓ Higher interest rate for long tenure loan have more chances to default.

✓ Higher loan amount for long tenure loan have more chances to default.

Thank You