

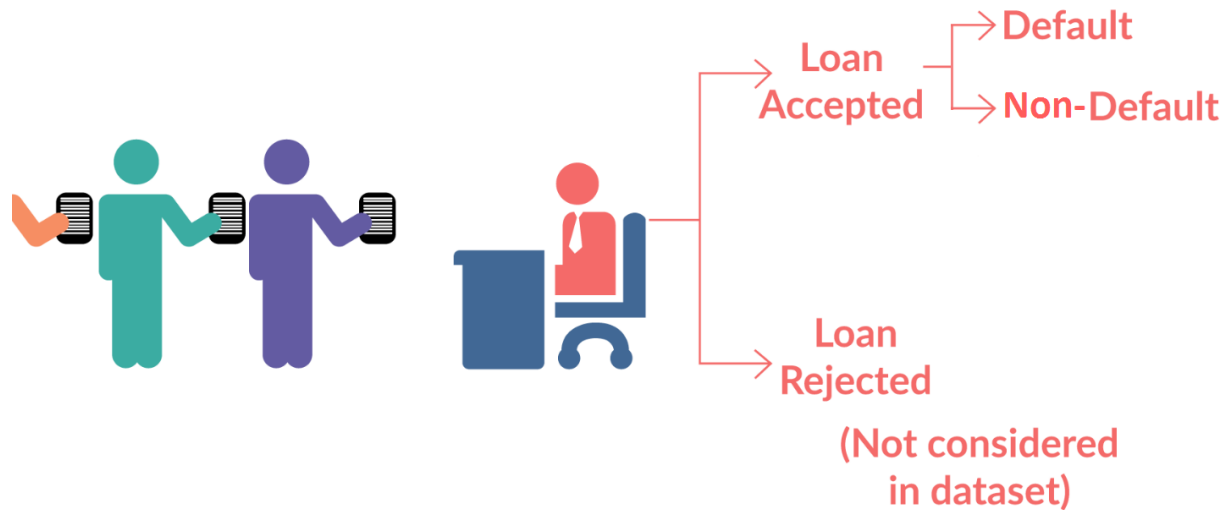
LENDING CLUB CASE STUDY

Namrata Bafna

Pooja Trivedi

01/18/2021

LOAN DATASET



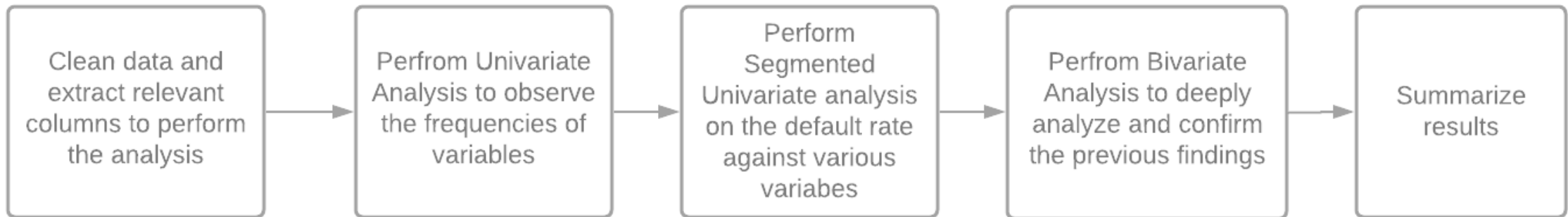
Business Understanding

- You work for a **consumer finance company** which specializes 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 data given below contains the information about past loan applicants and whether they 'defaulted' or not. The aim is to identify patterns 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.
- In this case study, you will use EDA to understand how **consumer attributes** and **loan attributes** influence the tendency of default.

Business Objective:

- The company wants to understand the driving factors (or driver variables behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Flow Chart



Variables
considered
for Analysis

LOAN CHARACTERISTICS

DEMOGRAPHICS
CHARACTERISTICS

Loan_amnt

Emp_length

Grade

Home_ownership

Term

annual_inc

Installment

verification_status

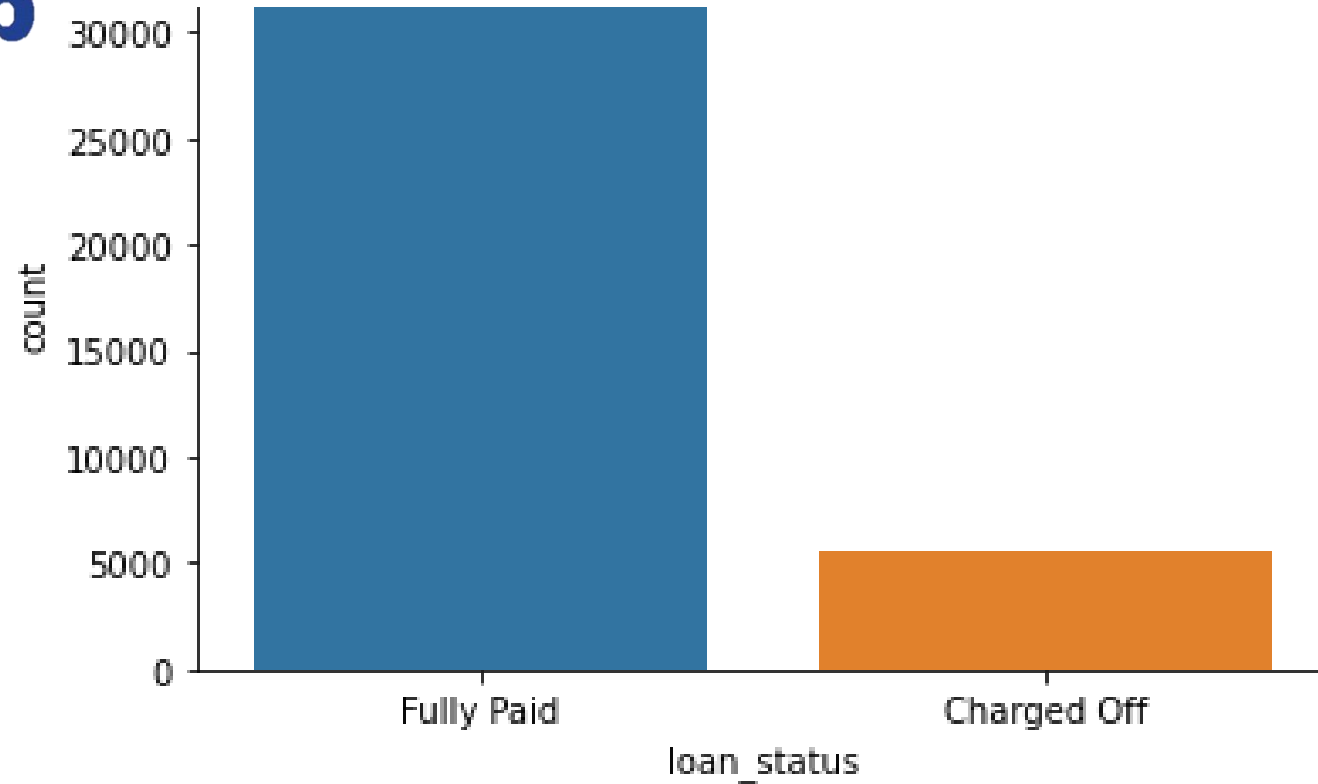
Int_rate

purpose

issue_d

pub_rec_bankruptcies

Univariate Analysis



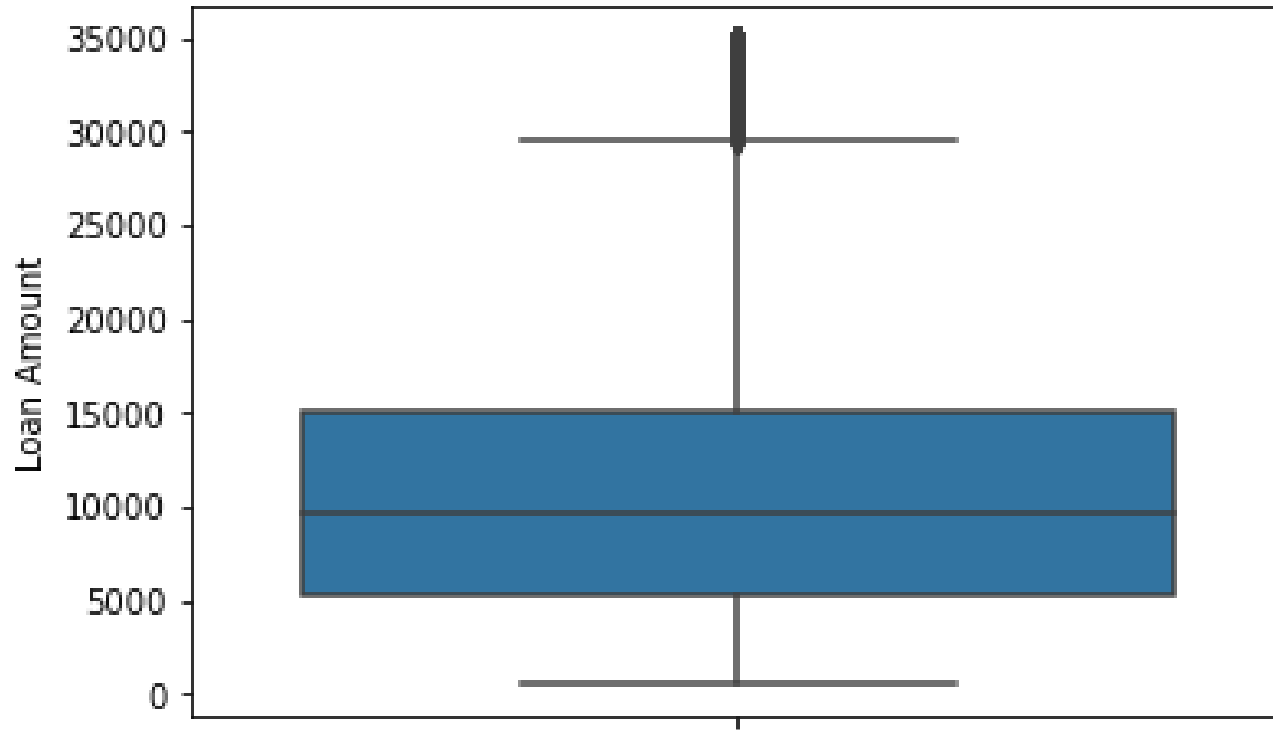
Observation:

From the above plot we can infer that most of the loans have been repaid fully.

The percentage of defaulters is 14.17%

Univariate Analysis

Loan Status – Defaulters vs Non-Defaulters

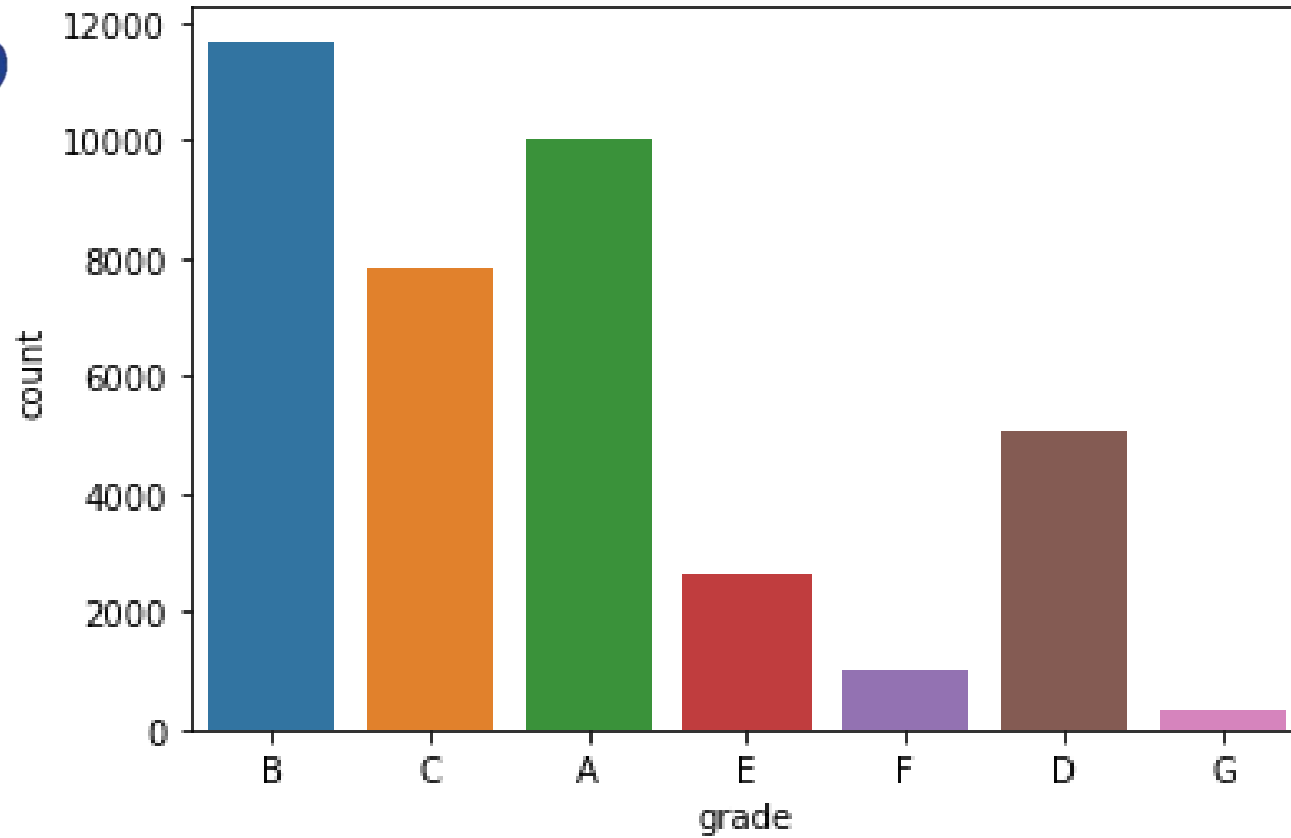


Observation:

On an average the sanctioned loan amount varies between USD 5300 and USD 9600 with some outliers values having amount sanctioned as USD 35000.

Univariate Analysis

Loan Amount

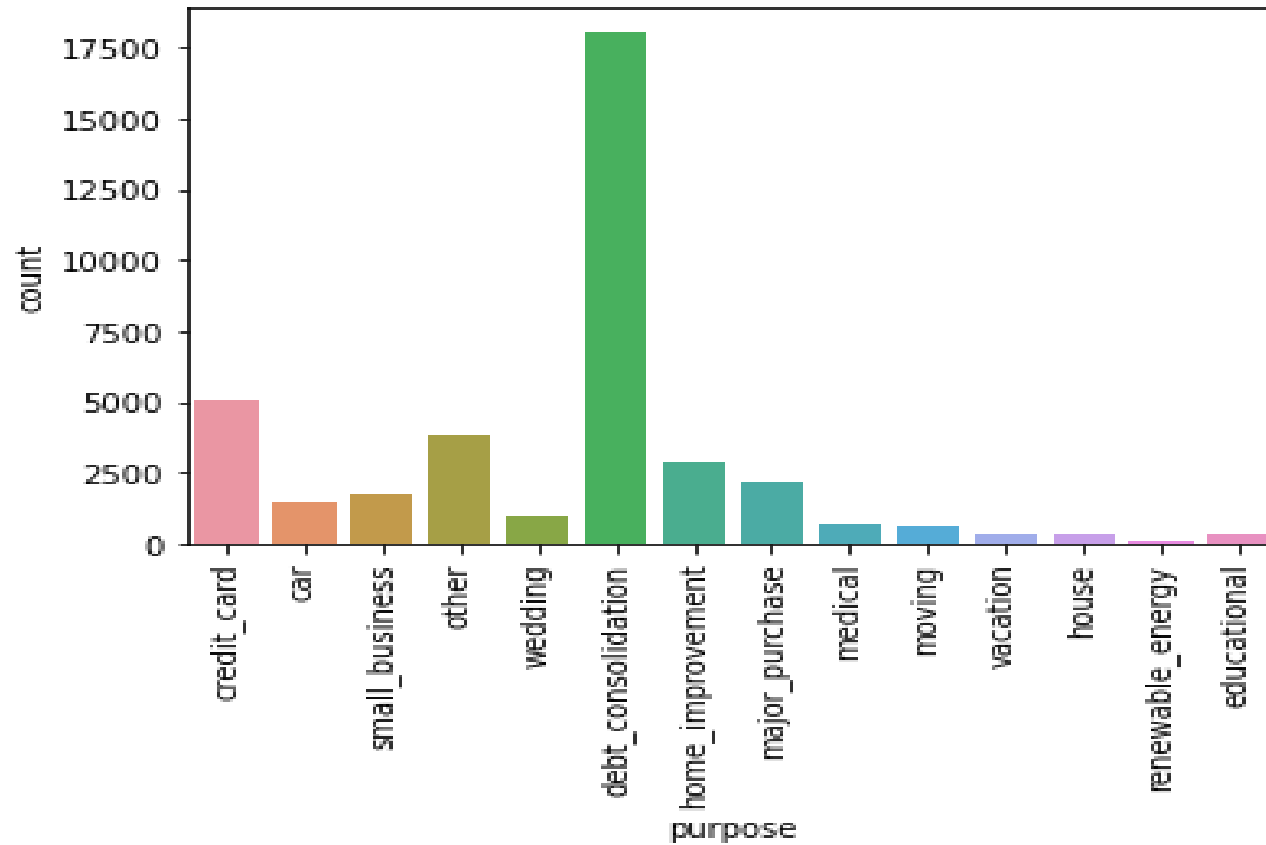


Observation:

Most of the loans have grade of A and B.

Univariate Analysis

Grade

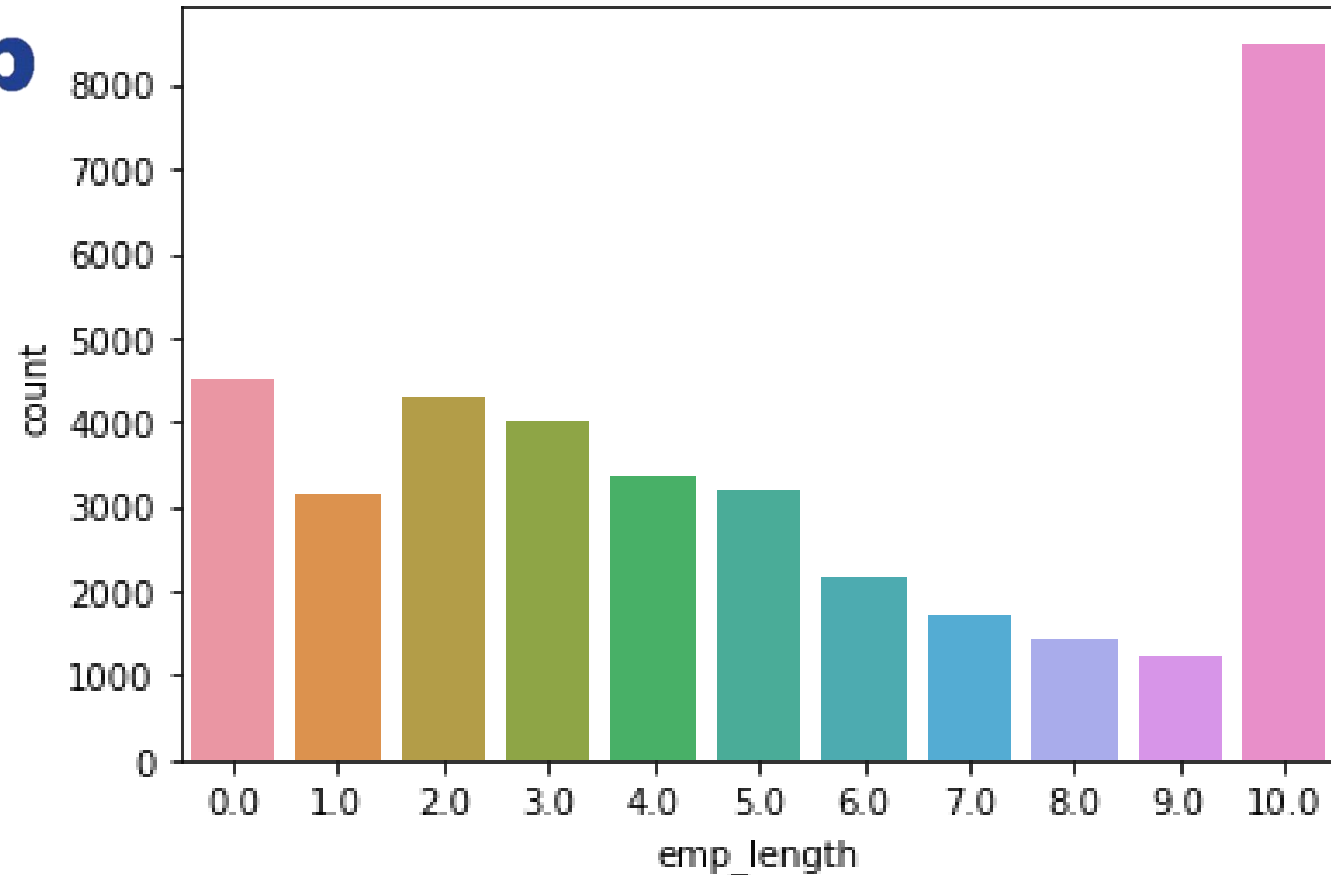


Observation:

Maximum number of loans have been for the category of Debt Consolidation followed by credit cards and others.

Univariate Analysis

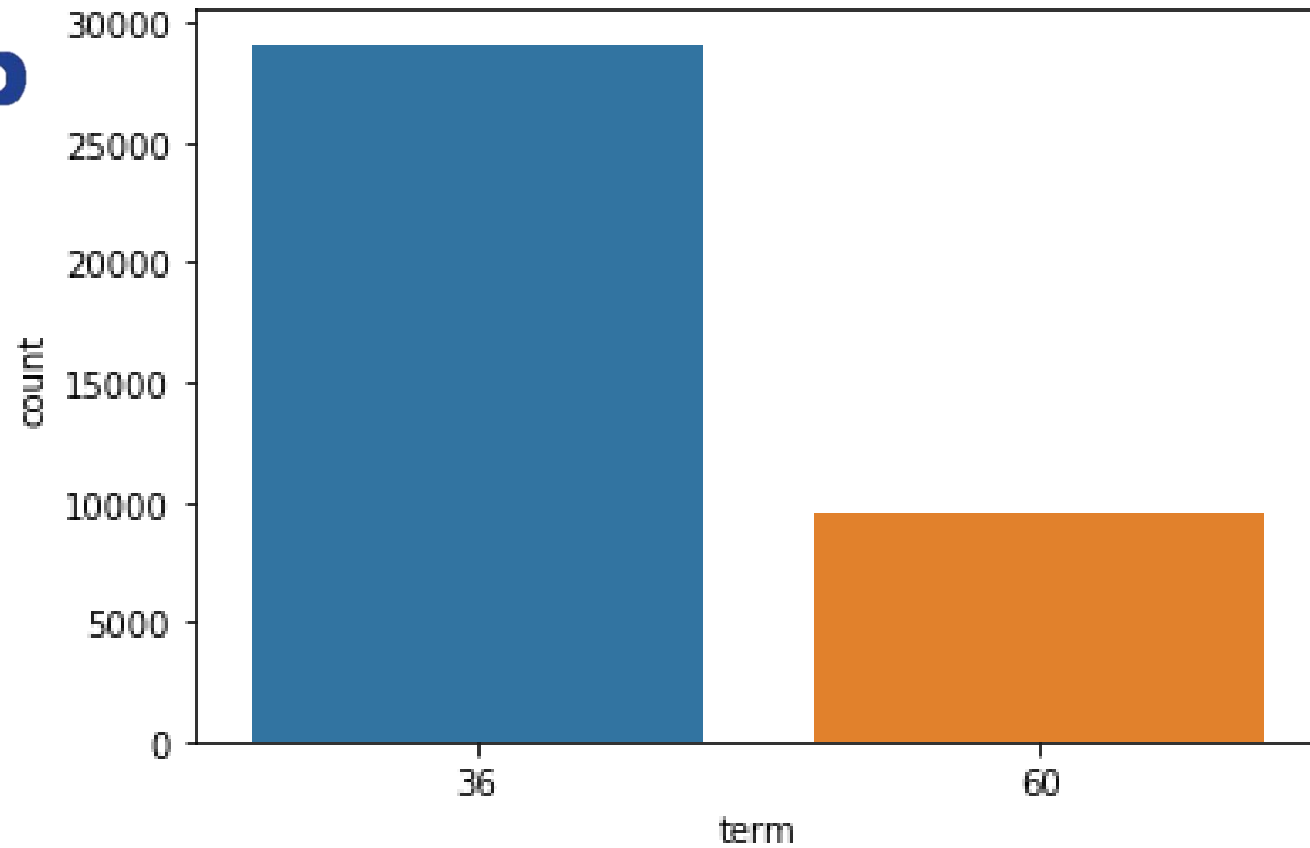
Purpose

**Observation:**

Maximum loans are sanctioned to loan takers having around 10 years of work experience.

Univariate Analysis

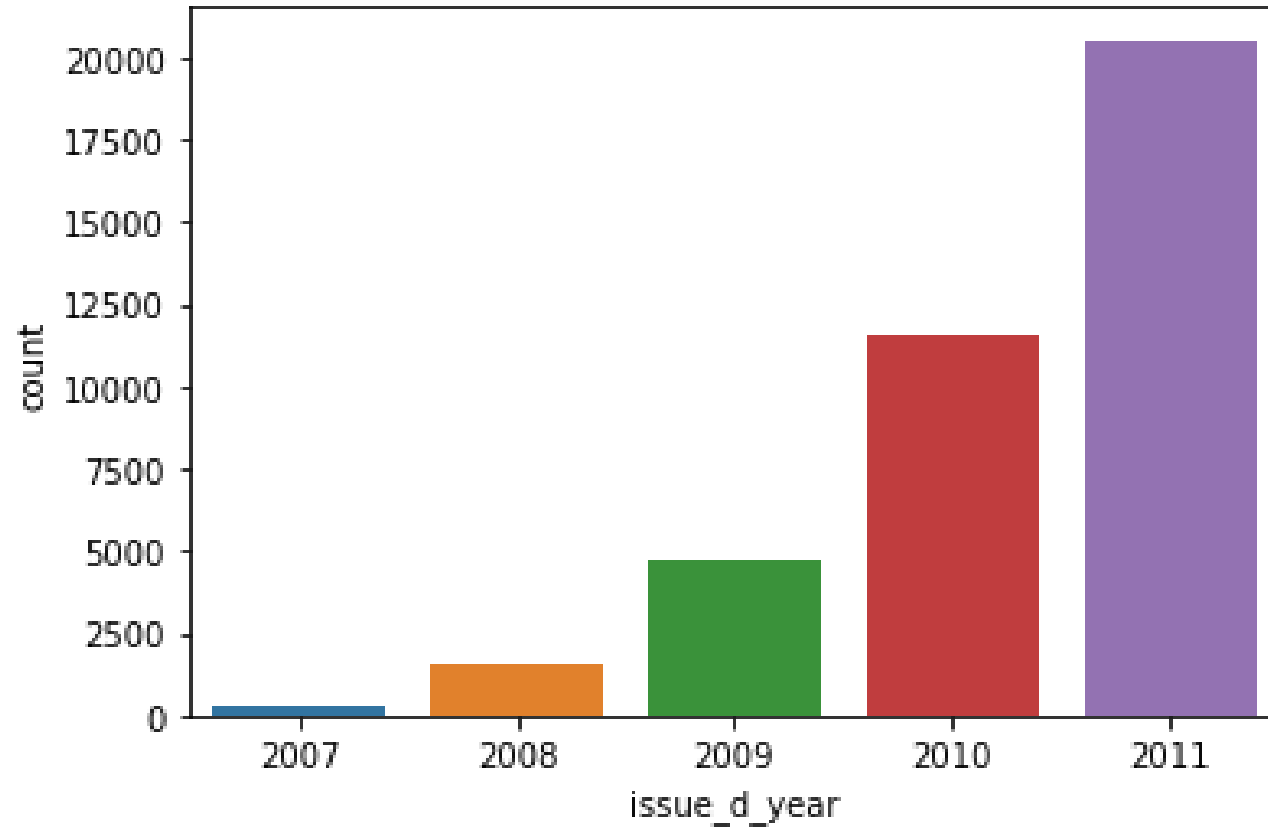
Employee Length

**Observation:**

More number of loans have been sanctioned for the duration of 36 months as compared to 60 months.

Univariate Analysis

Term

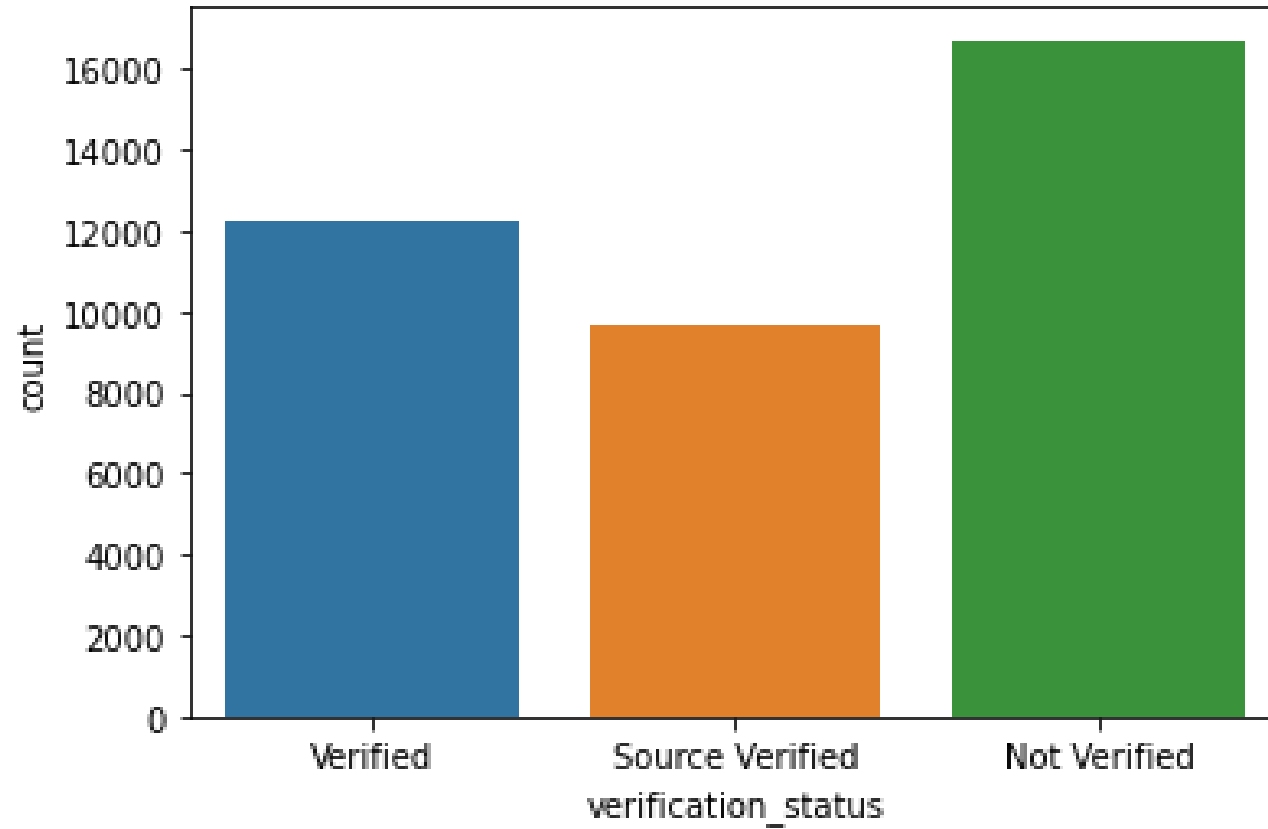


Observation:

Over the span of 4 years (2007 - 2011) the number of loans approved by LC has witnessed a considerable rise.

Univariate Analysis

Issue Date



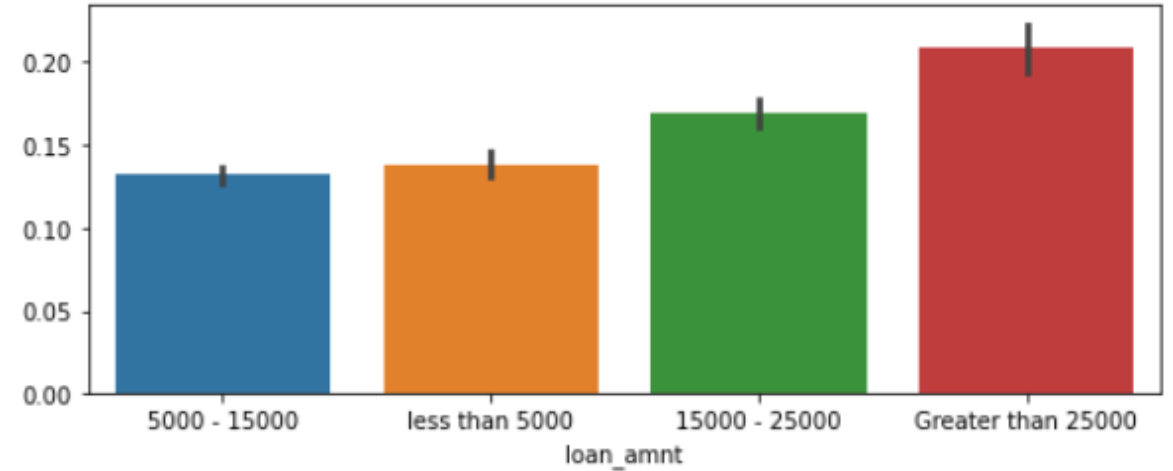
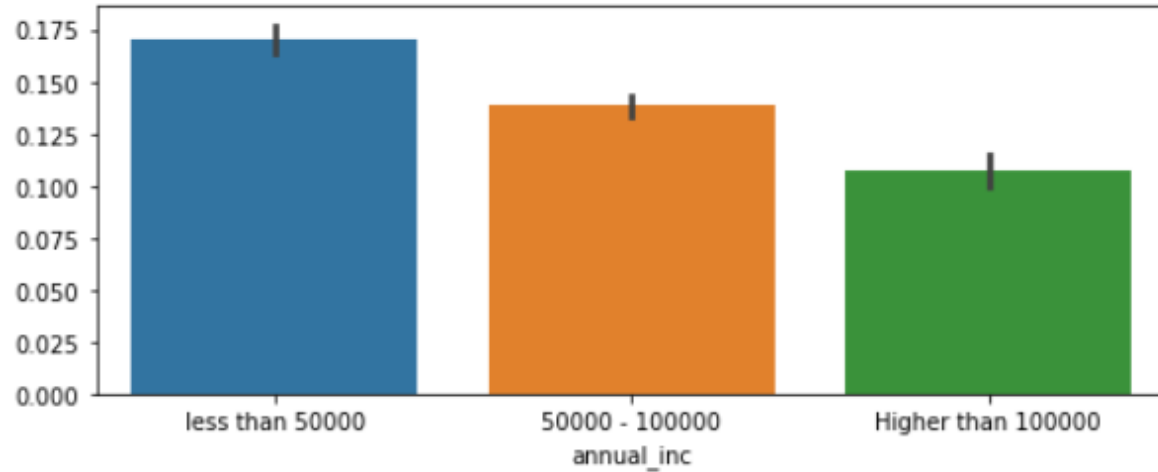
Observation:

For majority of the loans approved, the income of borrower was unverified by LC. LC needs to improve in its verification process.

Univariate Analysis

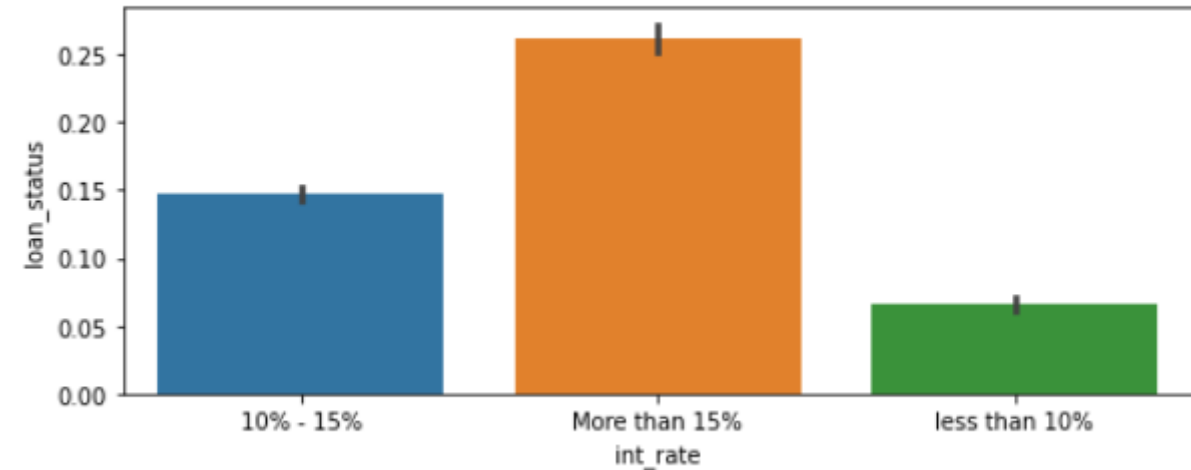
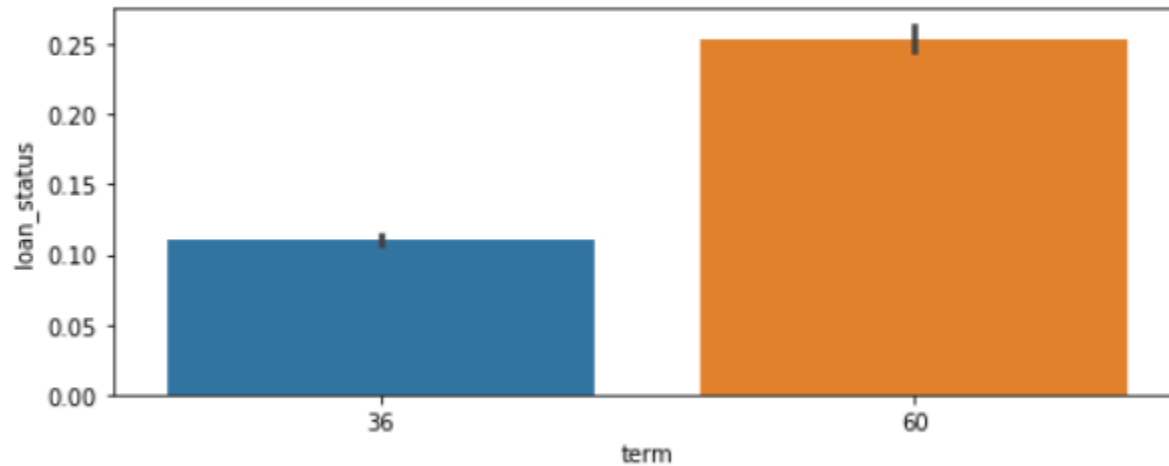
Verification Status

Segmented Univariate



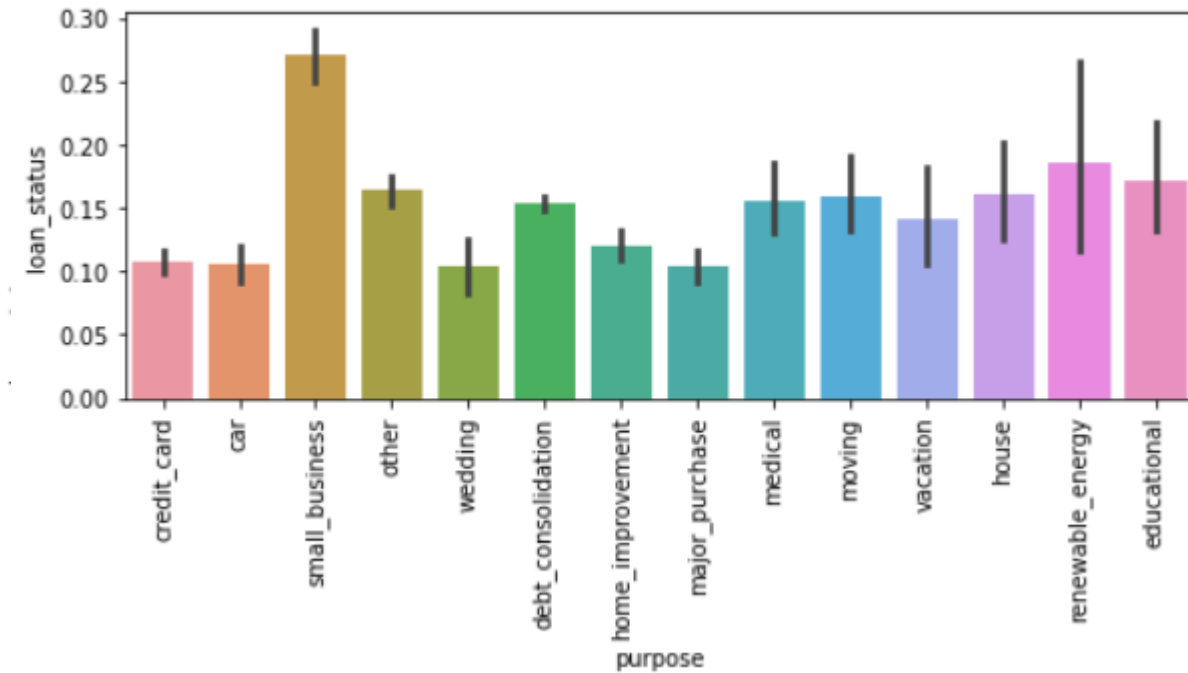
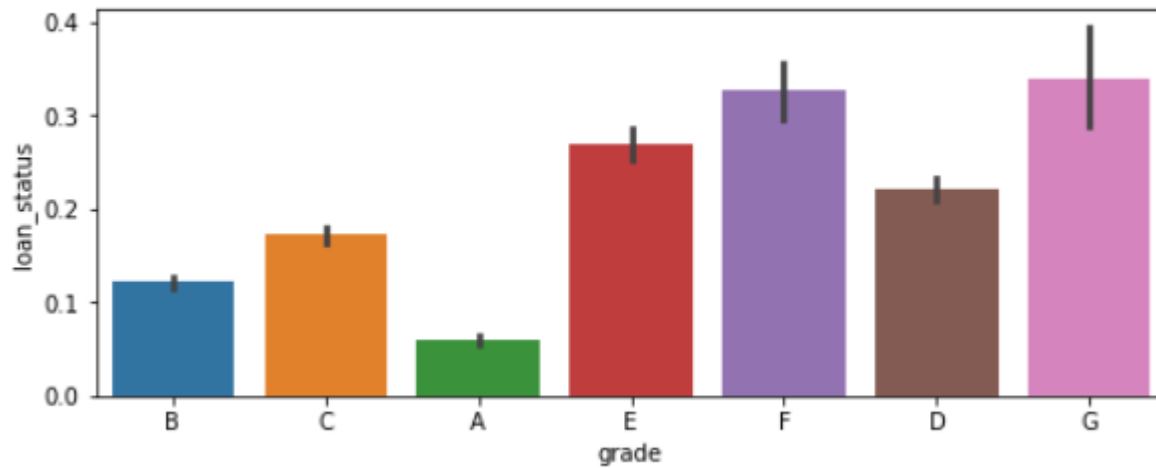
Default rate vs annual_inc: The probability to default is higher in lower income groups and lower in higher income groups

Default rate vs loan_amnt: If the loan amount is increasing, the probability of defaulting also increases. Higher the loan amounts, higher the probability to default.



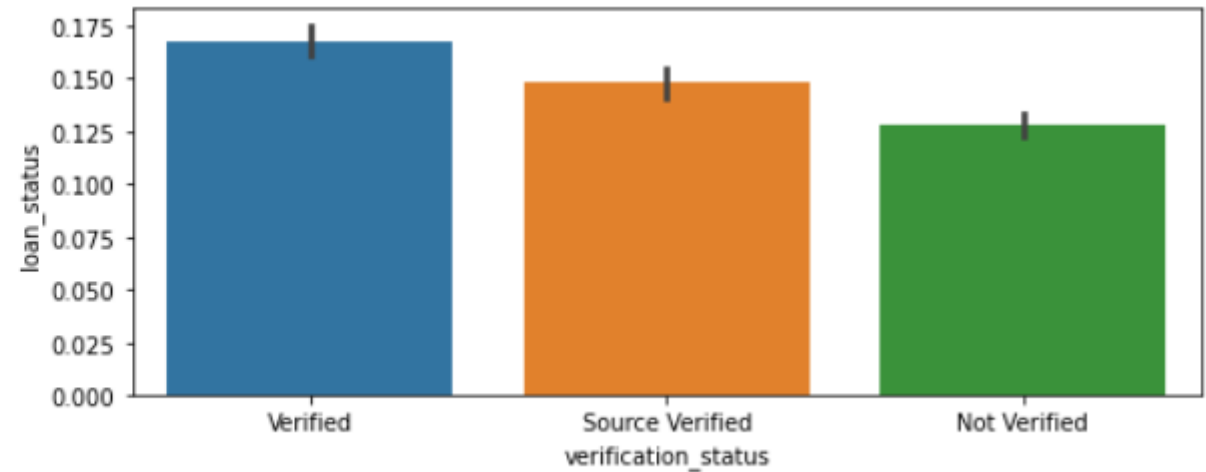
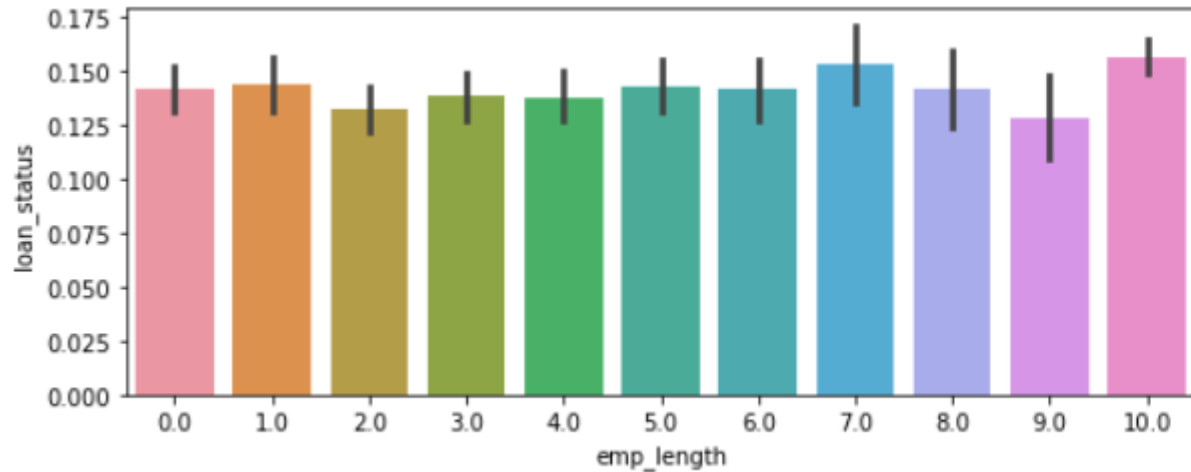
Default rate vs term: The longer the duration of loan, higher is the probability of defaulting

Default rate vs int_rate: The probability to default rises proportionately with the rate of interest. For loans with higher interest rate, the default rate observed is higher



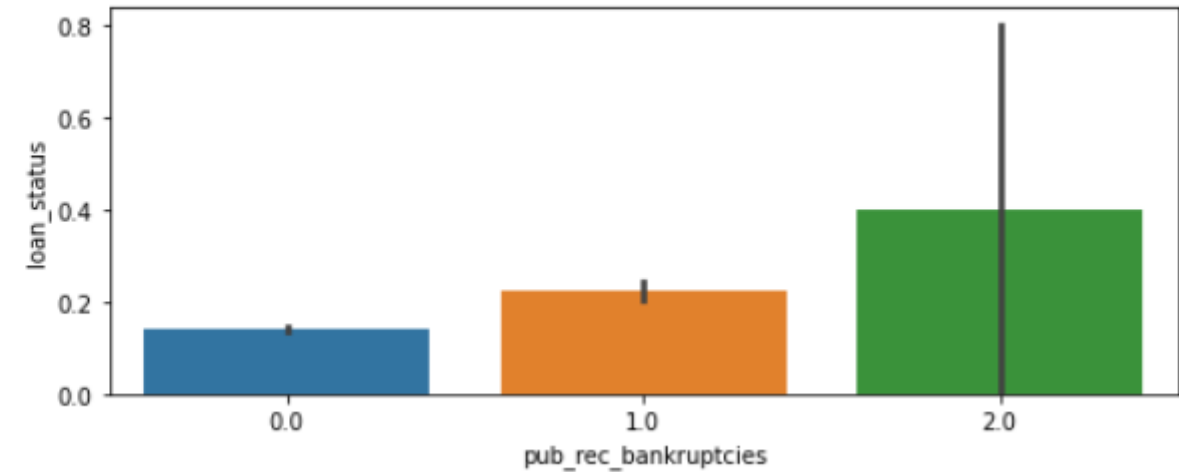
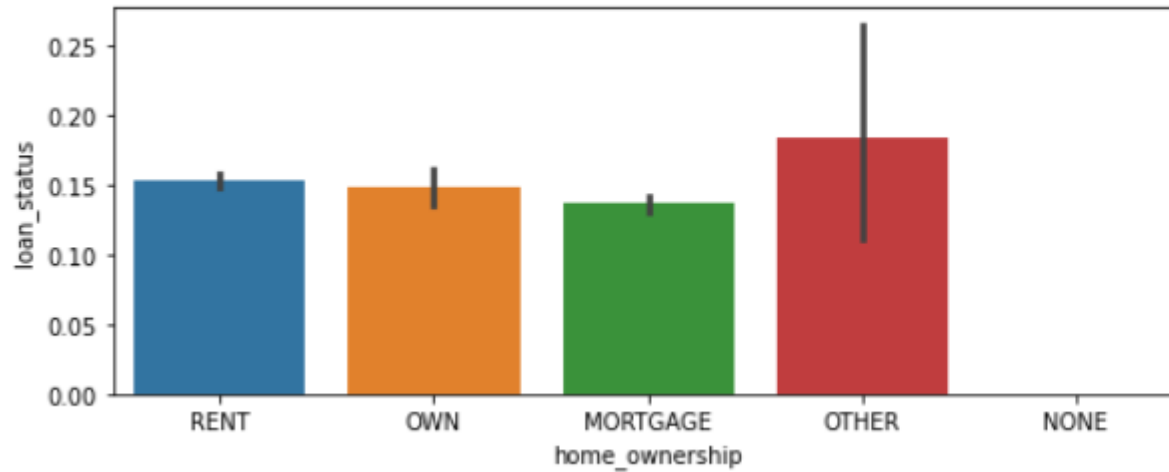
Default rate vs grade: For loans with Grade 'A', the default rate is least, and it increases proportionately with maximum value for grade 'G'. So lower grade quality loans are more prone to default as compared to those of higher grade.

Default rate vs purpose: The probability of defaulting varies considerably with the purpose of applying for a loan.



Default rate vs emp_length: There is a marginal difference in the probability of defaulting with respect to the employment length of the loan taker. However, it can be observed that those with 10 years of employment length tend to default more as compared to those with lesser years of employment length.

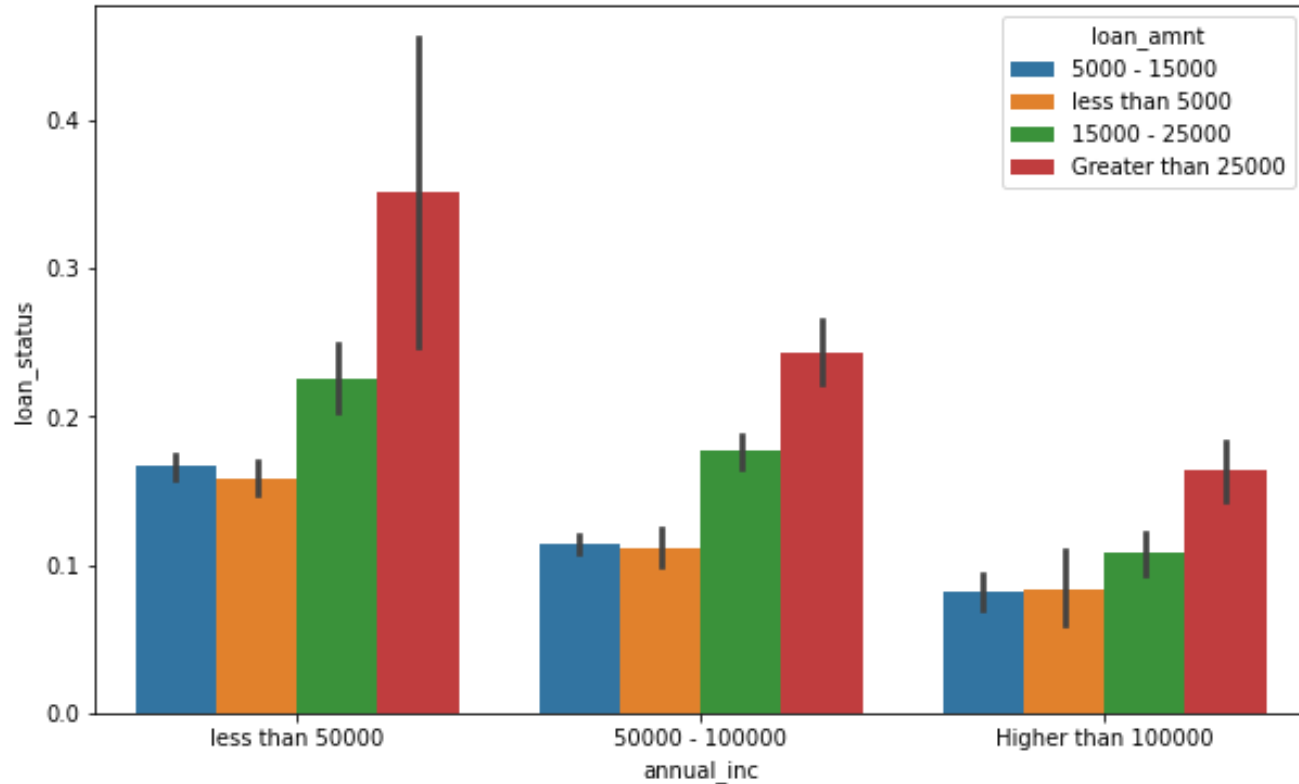
Default rate vs verification_status: It appears that for Verified loans the probability of defaulting is higher as compared to Not verified loans.



Default rate vs home_ownership: The probability of defaulting is highest for 'Other'. However, there is just a marginal impact of various home ownership statuses on the default rate

Default rate vs pub_rec_bankruptcies: It can be observed that for loan takers with higher record of bankruptcies in the past, the probability of defaulting increases. Those with no bankruptcy records have a very small chance of defaulting.

Bivariate Analysis

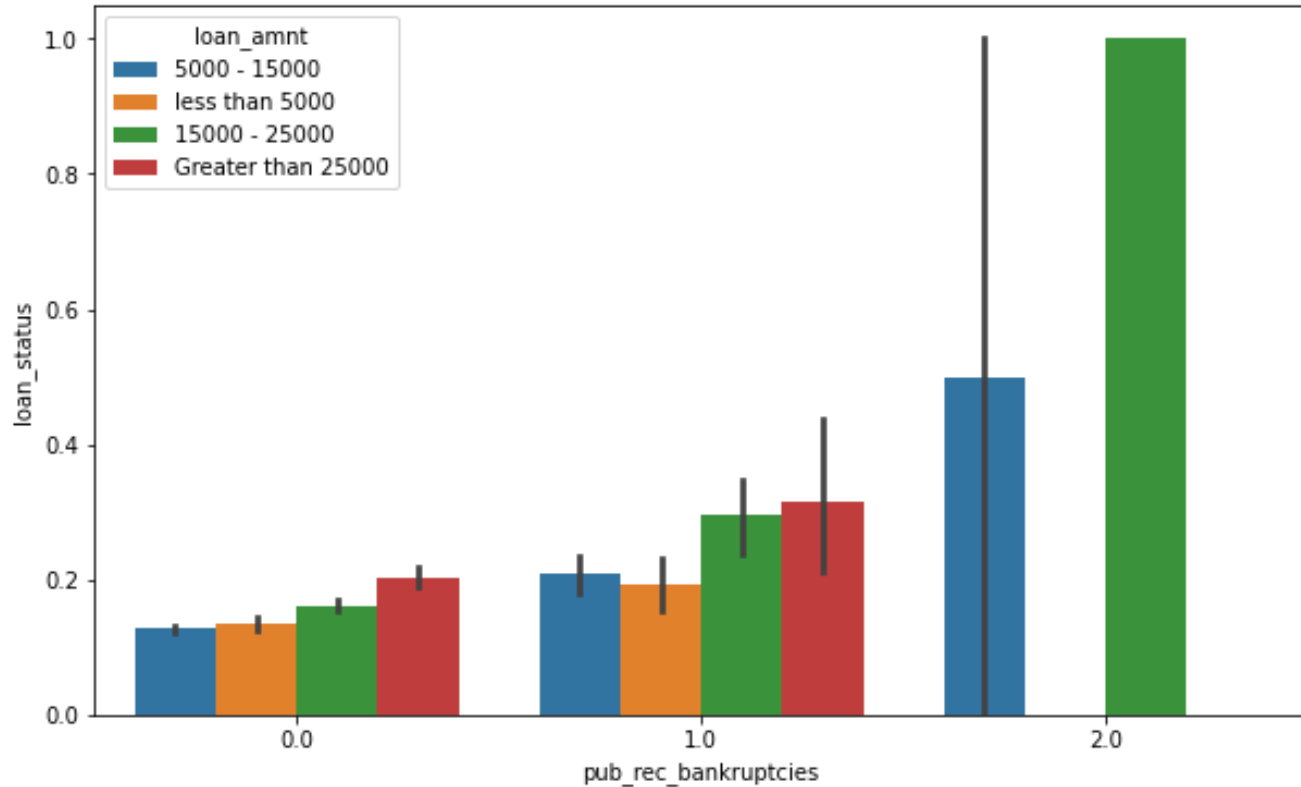


BiVariate Analysis

annual_inc and loan_amnt vs
Loan Status

Observation:

The probability to default is higher with higher loan amounts. However, loan takers with higher annual incomes, even if they have applied for higher loan amounts have lesser probability of defaulting as compared to the ones with medium or lower salaries.

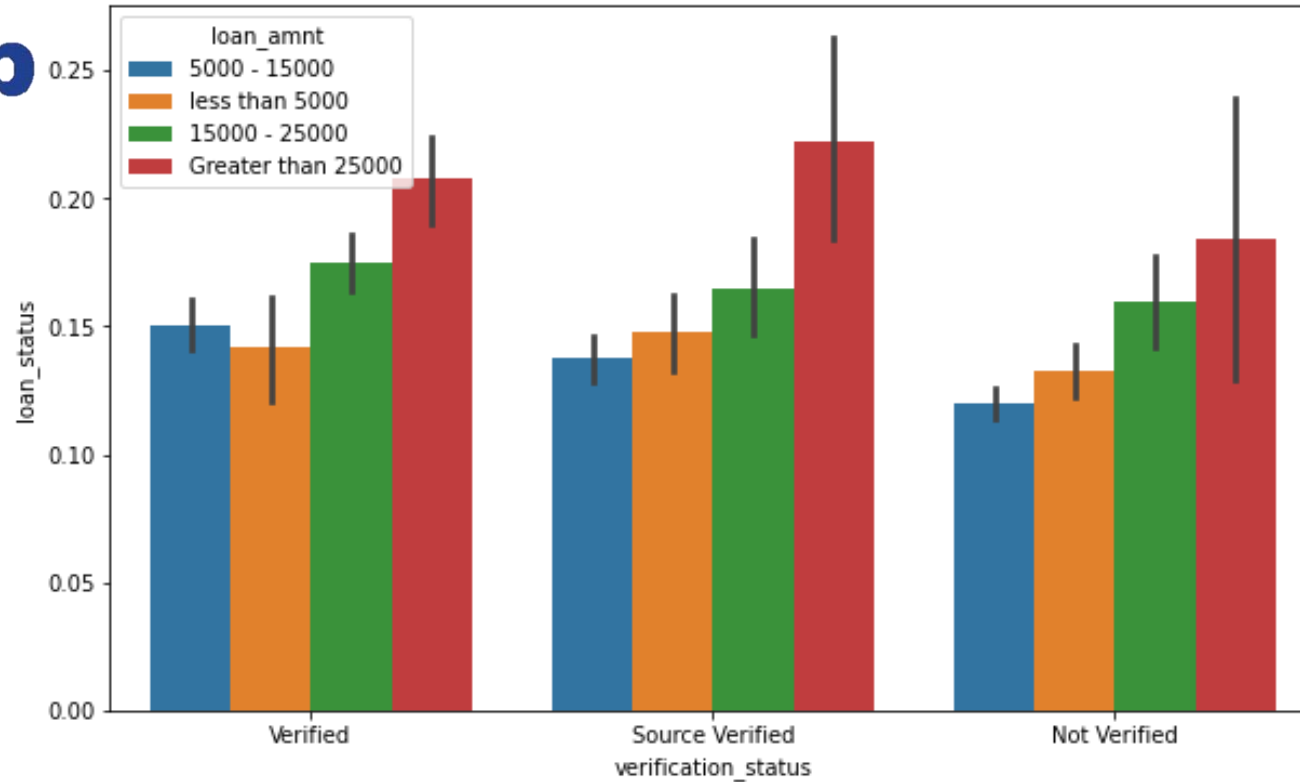


BiVariate Analysis

pub_rec_bankruptcies and
loan_amnt vs Loan Status

Observation:

Loan takers with loan amounts less than 5000 have lesser chances of defaulting even if they have been bankrupt before. The default rate is directly proportional to the loan amount and bankruptcy values. Increase in either of the two will increase the tendency to default.

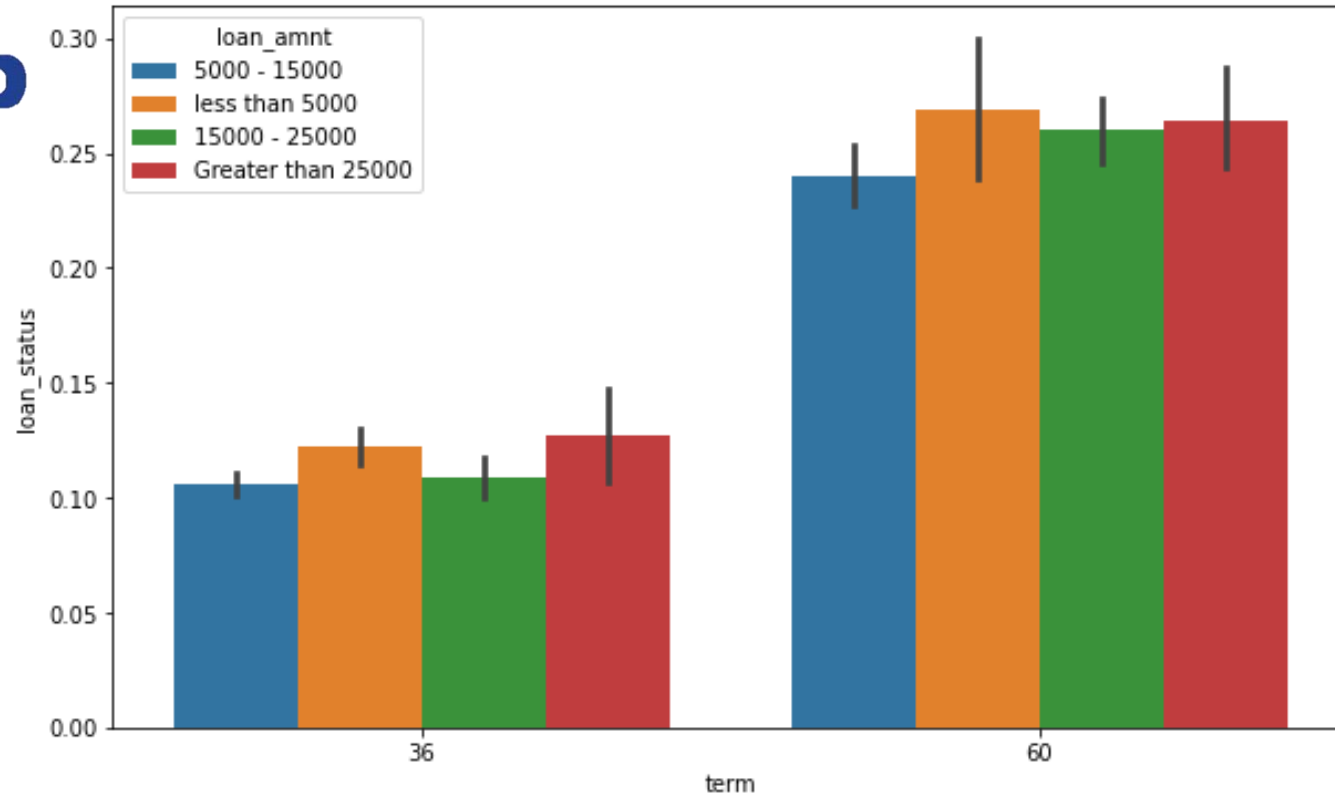


Observation:

Most of the loans of higher values were verified by Lending Club. However, the default rate is higher for verified and source verified loans. From the previous observations, it can be said that the probability to default depends more on the loan amount and annual income and verification status doesn't contribute much to the driving factors.

BiVariate Analysis

verification_status and
loan_amnt vs Loan Status

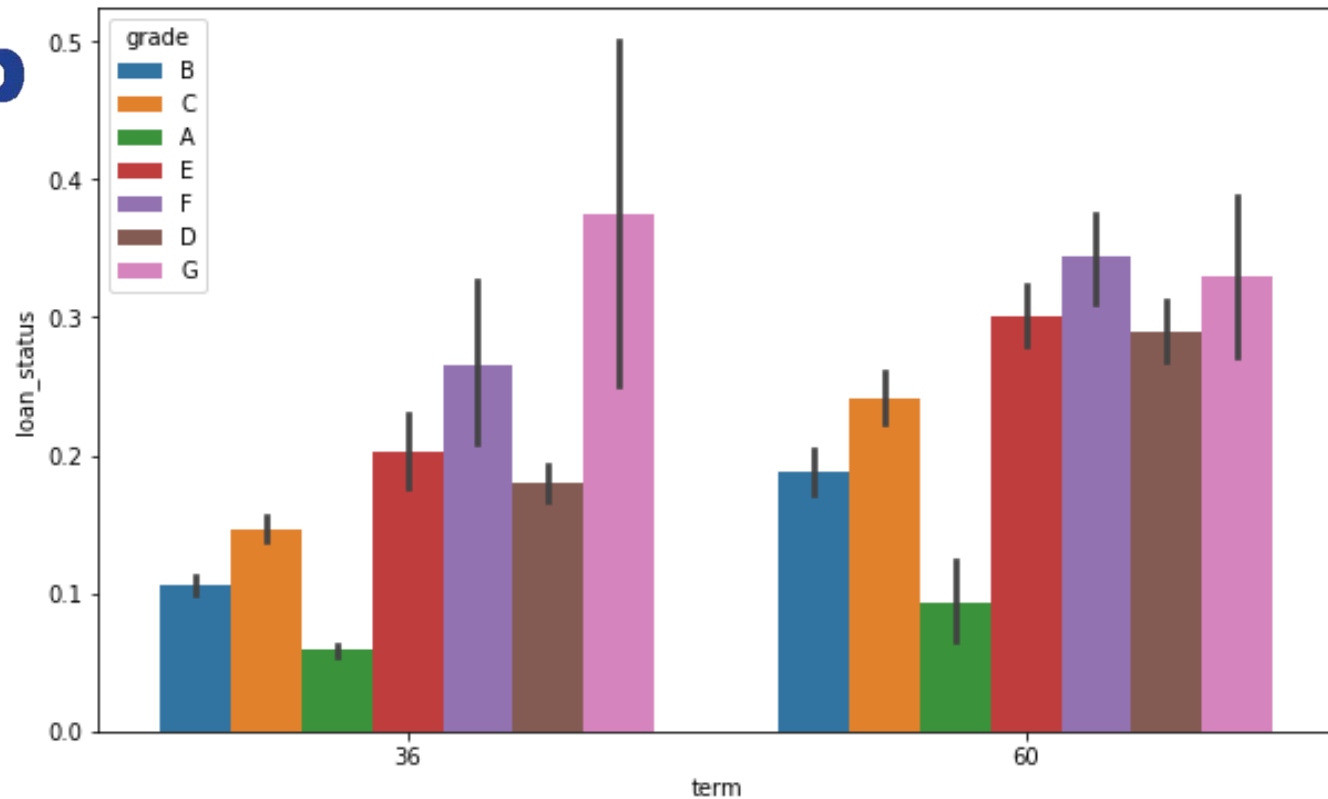


BiVariate Analysis

term and loan_amnt vs Loan Status

Observation:

Most of the loans defaulted have been of a higher duration and belongs to income level less than USD 5000.

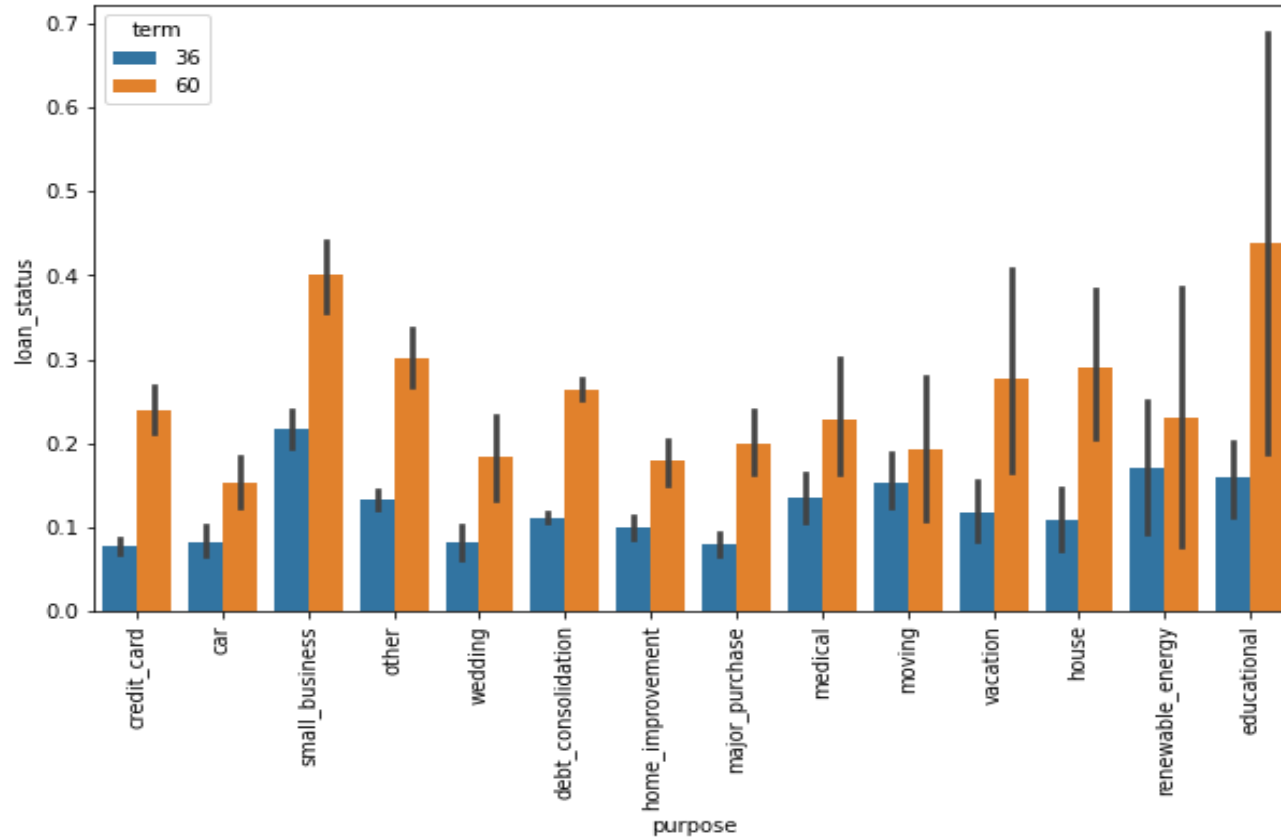


BiVariate Analysis

term and grade vs Loan Status

Observation:

Default rates are higher for loans of lower grades of longer durations.

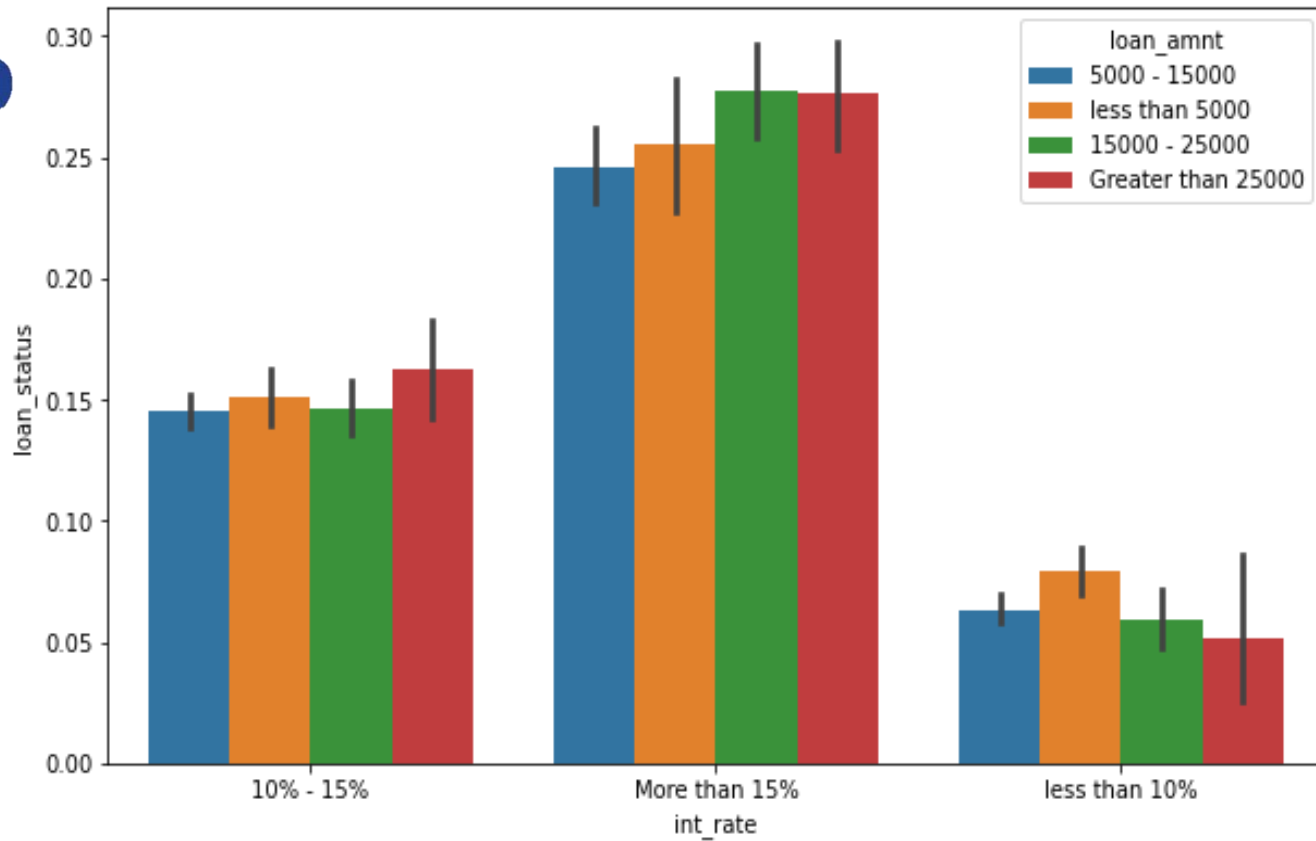


Observation:

Small businesses have been the riskiest loan category. Maximum default rate is observed for longer as well as shorter durations. For all other categories, longer term has defaulted more as compared to shorter duration ones. Educational loans of longer duration have also resulted in large defaults followed by house, vacation and debt-consolidation loans

BiVariate Analysis

term and purpose vs Loan Status

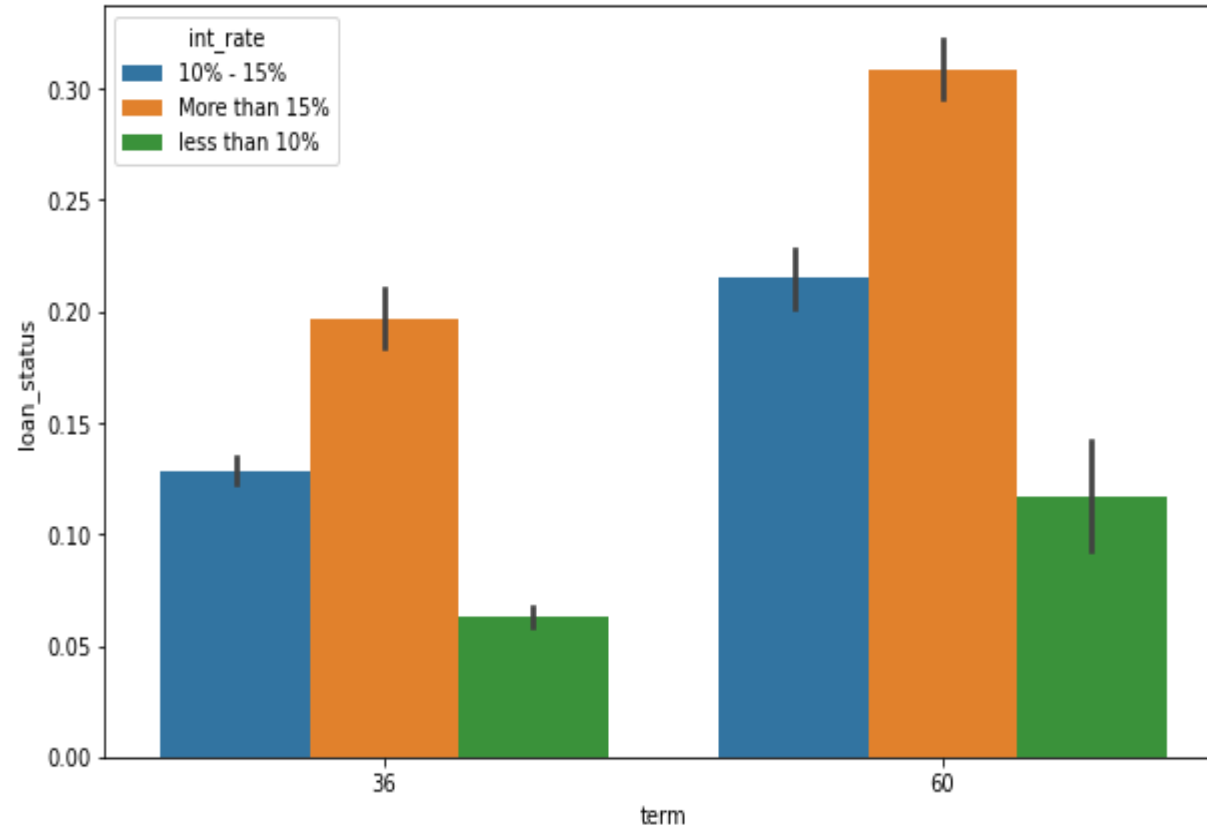


BiVariate Analysis

int_rate and loan_amnt vs Loan Status

Observation:

Loans of any values of higher interest rates have defaulted most



Observation:

Loans of longer as well as shorter durations of higher interest rates have defaulted more as compared to those with lesser interest rates

BiVariate Analysis

int_rate and term vs Loan Status

Conclusion

Driving Factors for Loan Default:

1. Loan Amount
 2. Annual Income
 3. Term
 4. Interest Rate
 5. Grade
 6. Purpose
 7. Public Recorded Bankruptcies
- Higher value loans should only be approved for higher income groups.
 - Higher value loans should be approved for a shorter term.
 - Reduce the rate of interest for loans of all amounts.
 - Most of the loans approved are high grade loans and its a good practice.
 - Stop giving loans to for the Small Business purposes.
 - Stop approving loans to people with higher bankruptcy records.