

Lending Club Case Study

By: Prachi Shivekar

Lending Club

- It is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

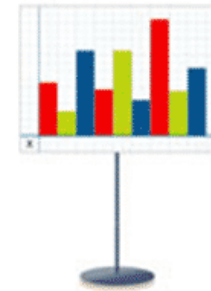
How Lending Club Works



Borrowers apply for loans.
Investors open an account.



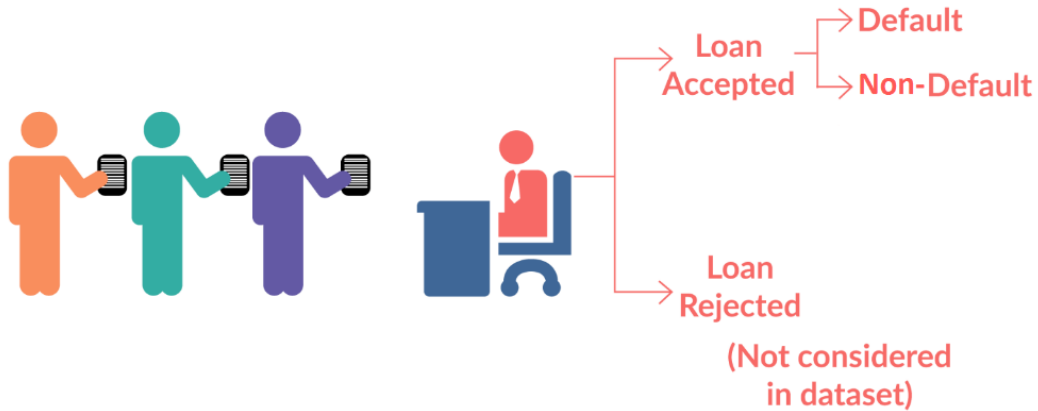
Borrowers get funded.
Investors build a portfolio.



Borrowers repay automatically.
Investors earn & reinvest.

Lending Club

LOAN DATASET



- **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

Problem Statement

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

Objective: 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.

Problem Solving steps

Data understanding

Data Cleaning and Manipulation

Data Analysis

Presentation and Recommendations



Data understanding

- The given loan dataset contains 39717 rows and 111 columns.
- There are no duplicate rows.
- There are no rows with all null values. However there are 54 columns which contain all null values.

Data Cleaning and Manipulation

- Remove 54 columns will all null values.
- Remove all those with very high missing percentage.
- Delete the columns will all zero or 'NA' values.
- Delete columns which have one value for all rows.
- Find and correct the column datatypes.
- Standardize the column values.
- Remove outliers

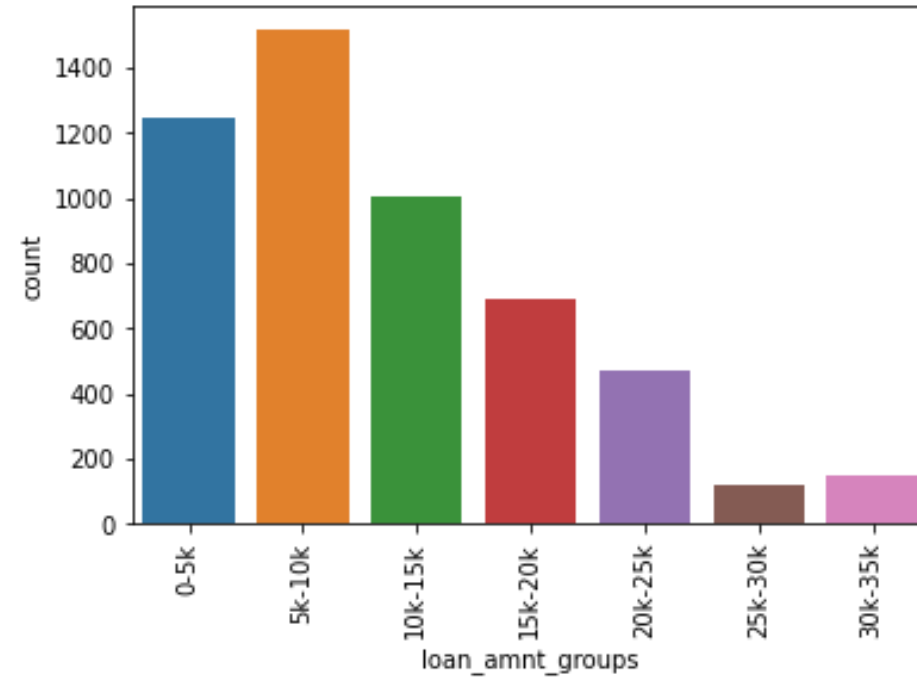
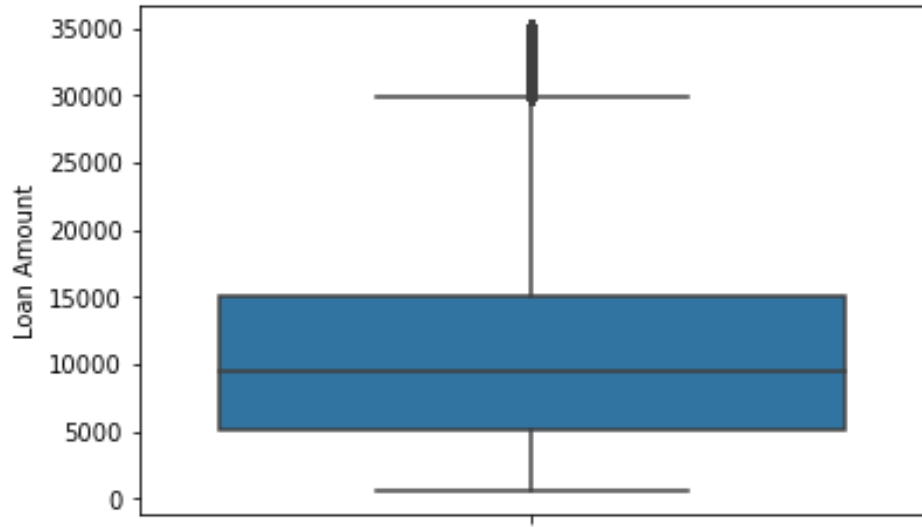
Data Analysis

- The objective is to identify predictors of default
- It is important to consider the "Dimensions" and "Measures" so that at the time of loan application, we can use those variables for approval/rejection of the loan.
- Dimensions are qualitative values and they are termed as categorical values. In given dataset age, occupation, employment details etc. are dimensions which help in data analysis.
- Measures are quantitative values. In given dataset amount of loan, interest rate, purpose of loan etc. are the measures which help in data analysis.
- There are some values which will be generated when loan is approved like delinquent 2 years, revolving balance, next payment date etc. These values are not present at the time of loan approval, hence they are not useful for loan approval process.
- So keep dimensions and measures and remove all other columns.

Data Analysis

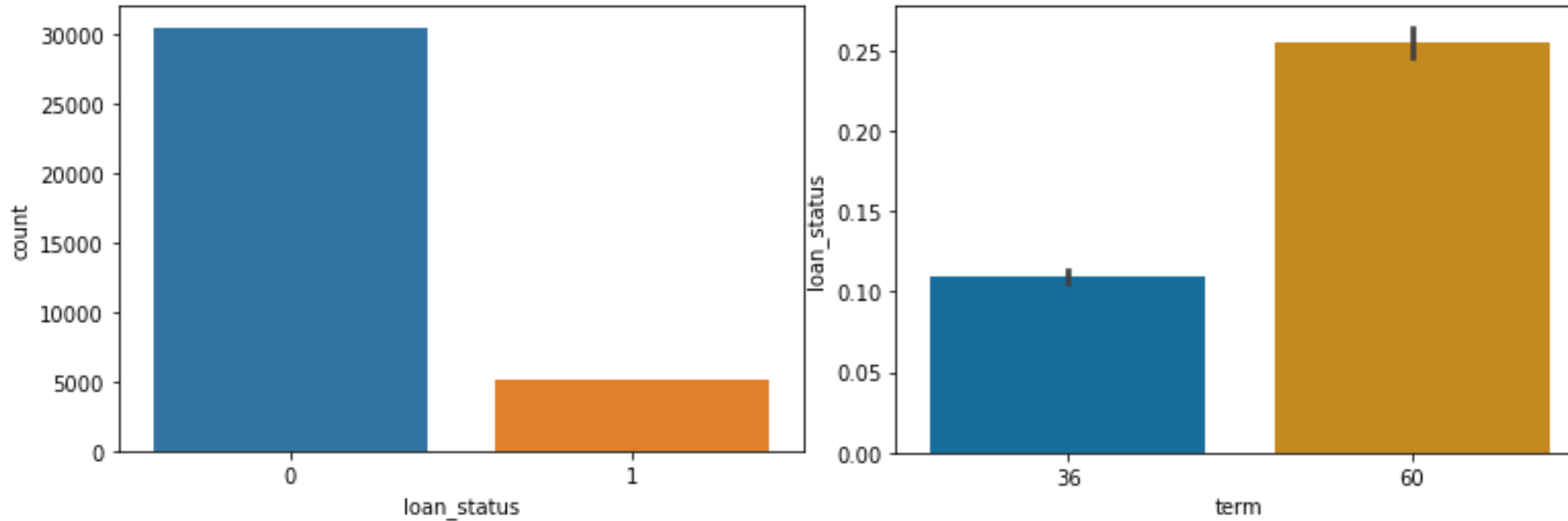
- The objective is to identify predictors of default
- It is important to consider the "Dimensions" and "Measures" so that at the time of loan application, we can use those variables for approval/rejection of the loan.
- Dimensions are qualitative values and they are termed as categorical values. In given dataset age, occupation, employment details etc. are dimensions which help in data analysis.
- Measures are quantitative values. In given dataset amount of loan, interest rate, purpose of loan etc. are the measures which help in data analysis.
- There are some values which will be generated when loan is approved like delinquent 2 years, revolving balance, next payment date etc. These values are not present at the time of loan approval, hence they are not useful for loan approval process.
- So keep dimensions and measures and remove all other columns.

Loan Amount Analysis



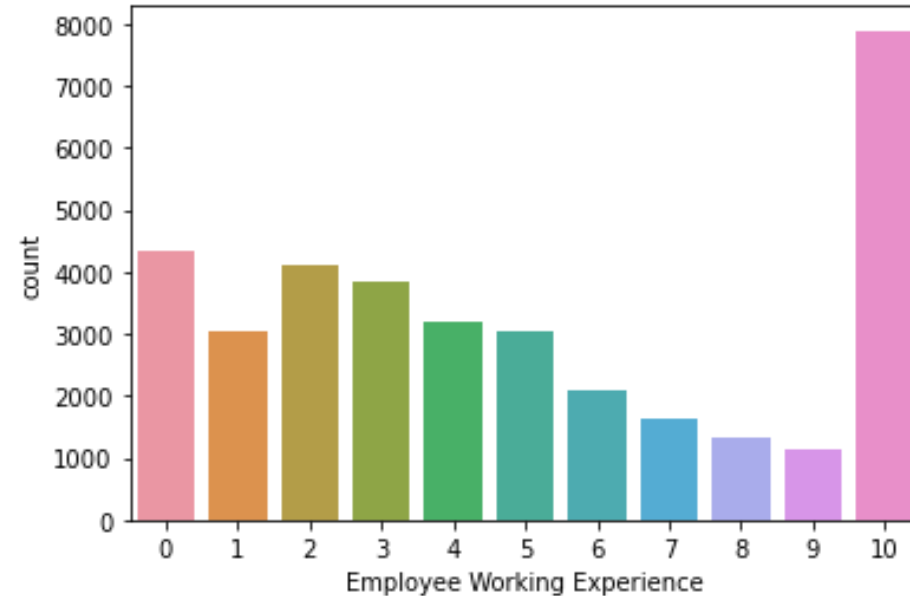
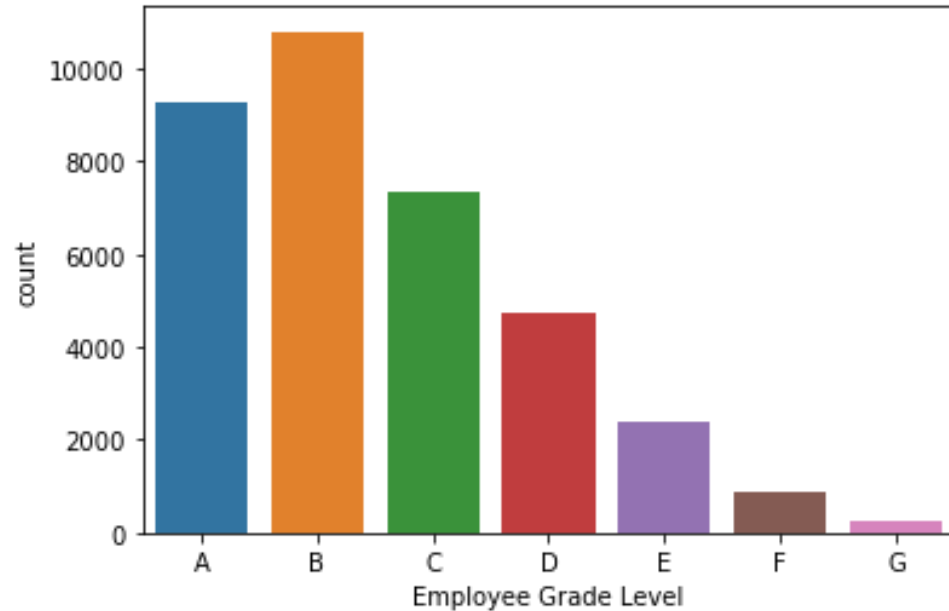
- The loan amount varies from 0 to 35,000 having mean of 10,000
- Mostly the Loan amount is between 5-10K.

Loan Status Analysis



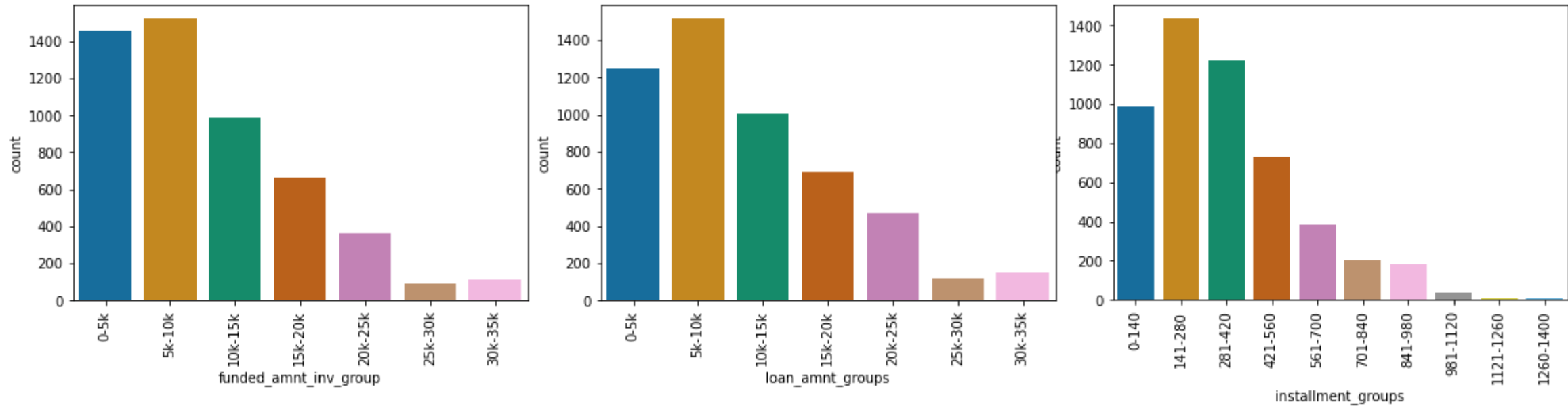
- Most of the loans are Fully Paid.
- 60 month loan default count is more than 36 months loan default

Grade and experience Analysis



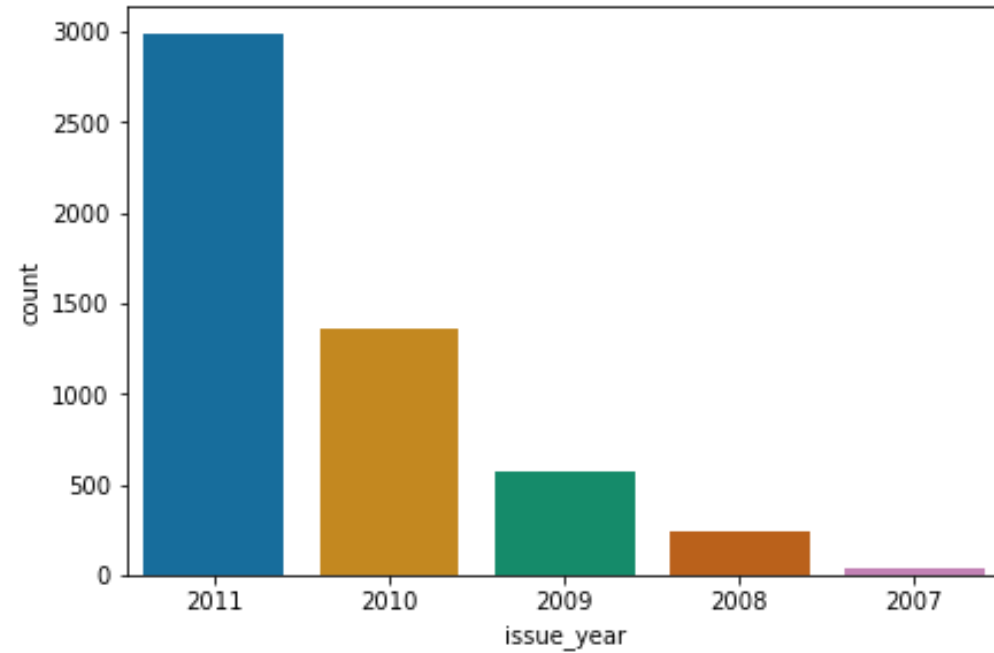
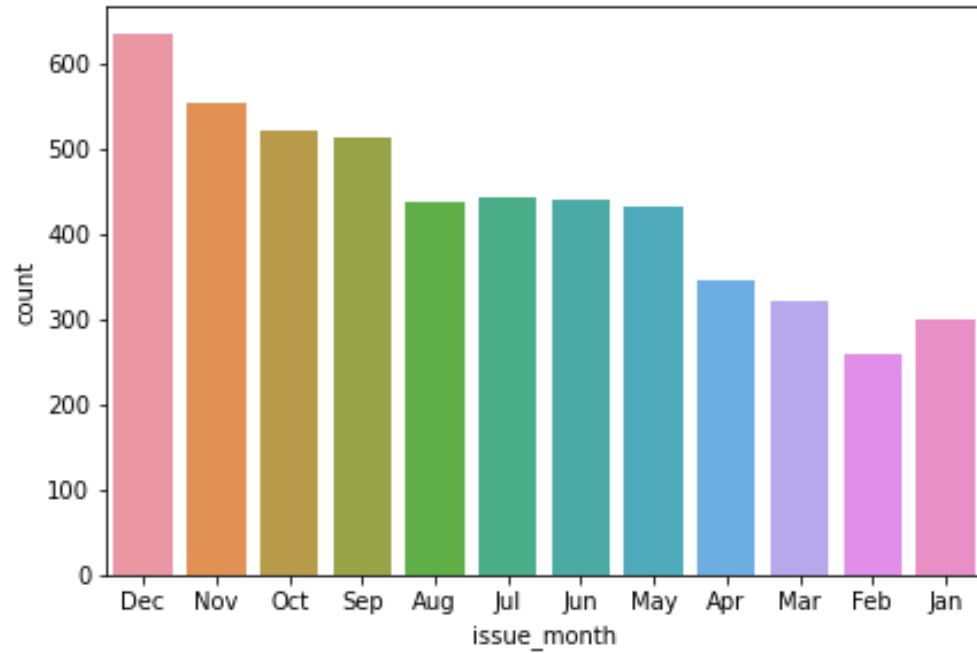
- Most of the loans have grade of A and B.
- Majority of employees applying for the loan have more than 10 years of experience.

Funded, Loan amount and Dti Analysis



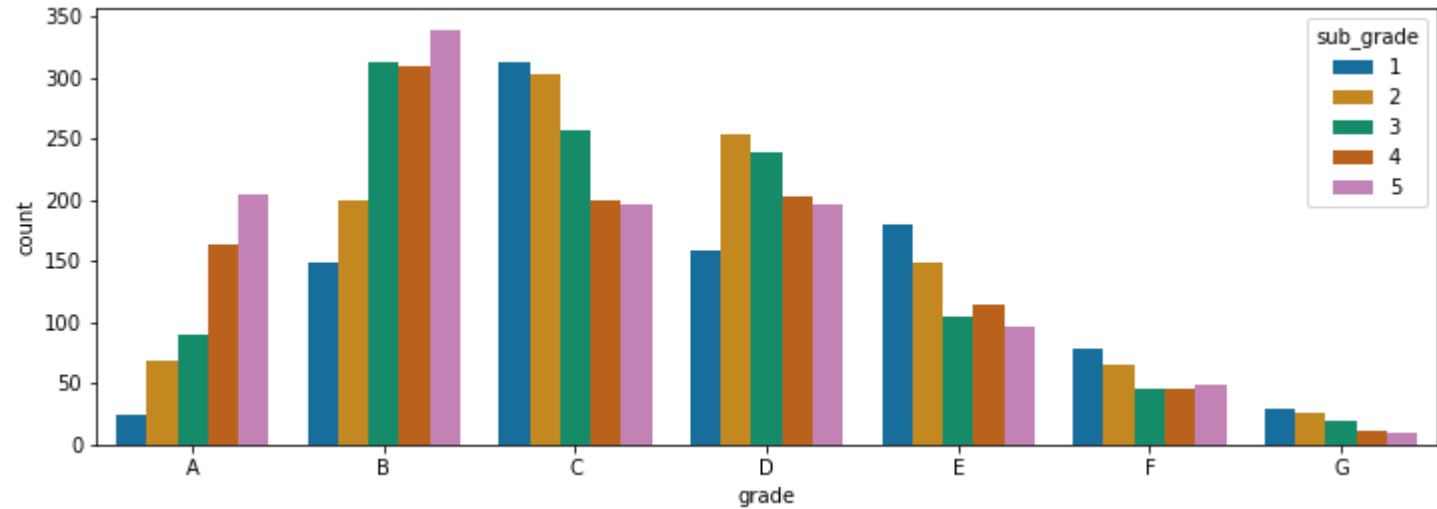
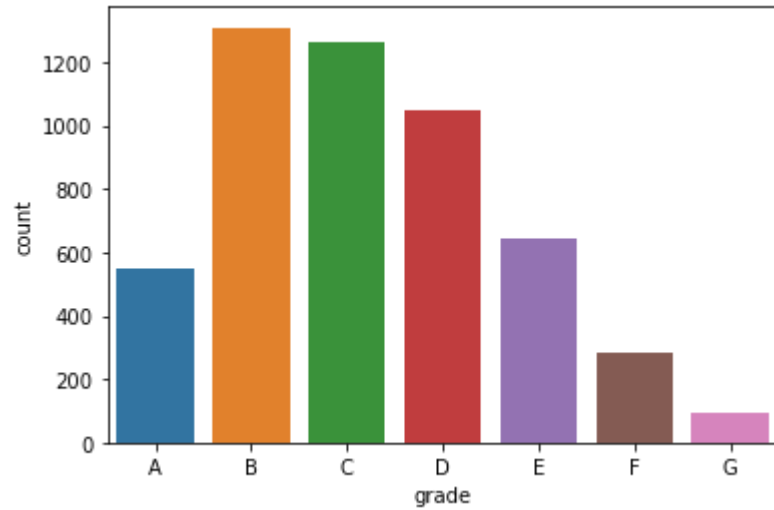
- Mostly funded amount is between 5000-10000.
- Mostly the Loan amount is between 5429 – 10357
- Most of the monthly installments are between 140-240

Loan default Analysis



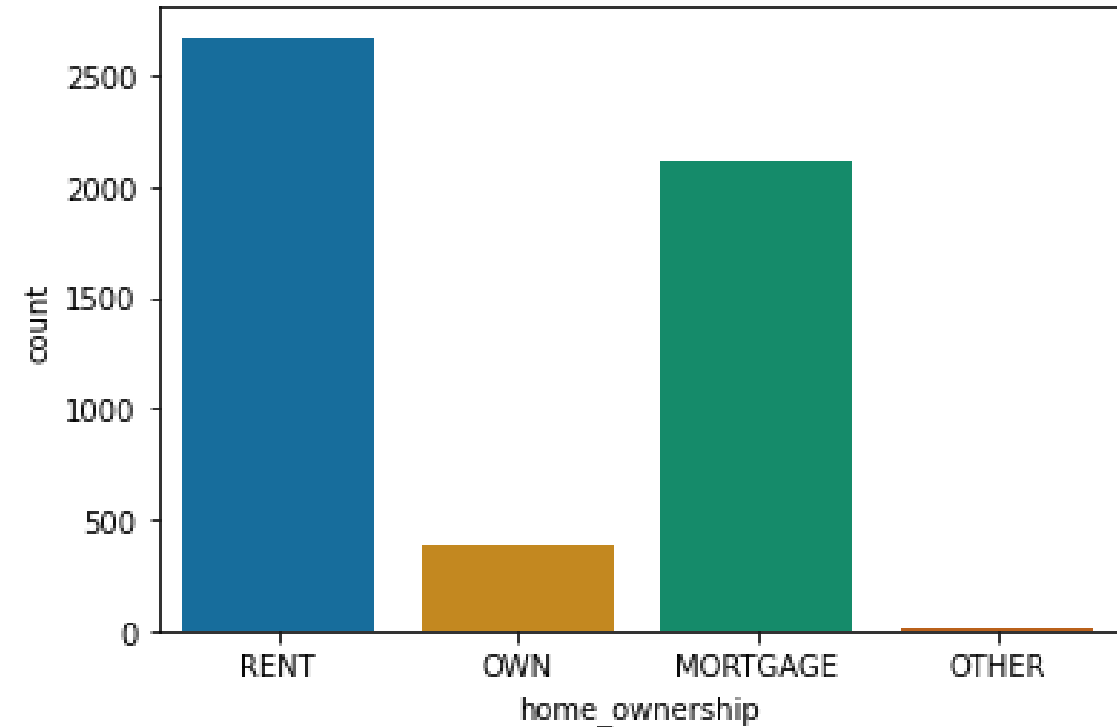
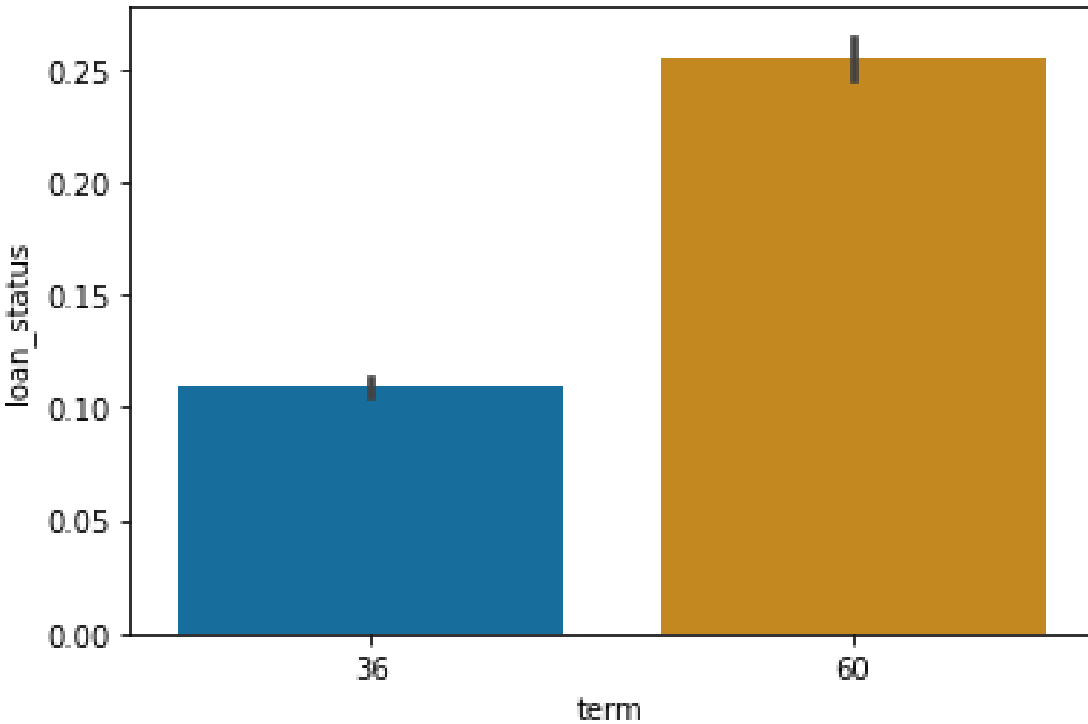
- Maximum number of defaults occurred when the loan was sanctioned/issued in Dec.
- Loan issued in the year 2011 were also as compared to other years.

Loan default Analysis



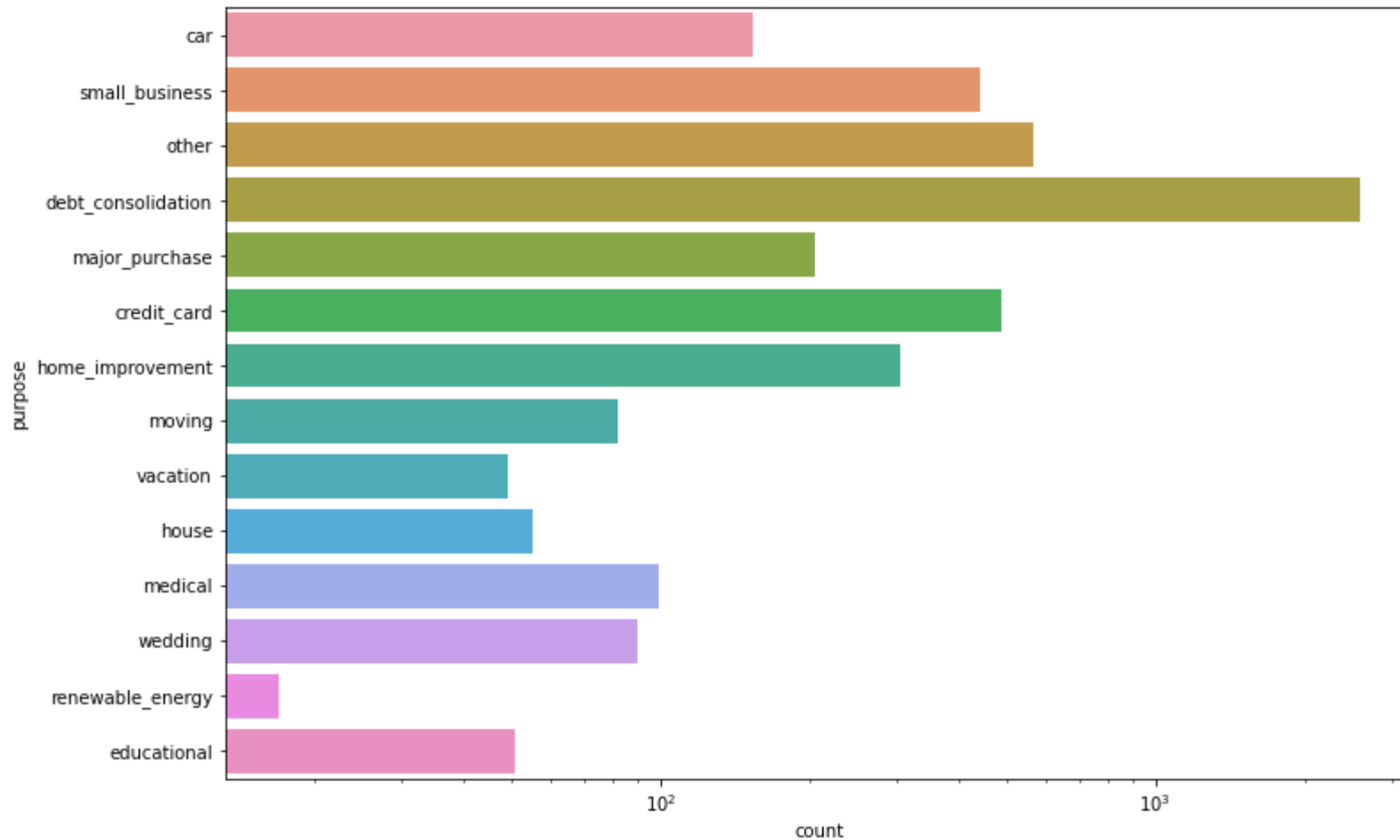
- Most of the defaulters are at grade B level.
- Most defaults are at 'B5' level

Loan default Analysis



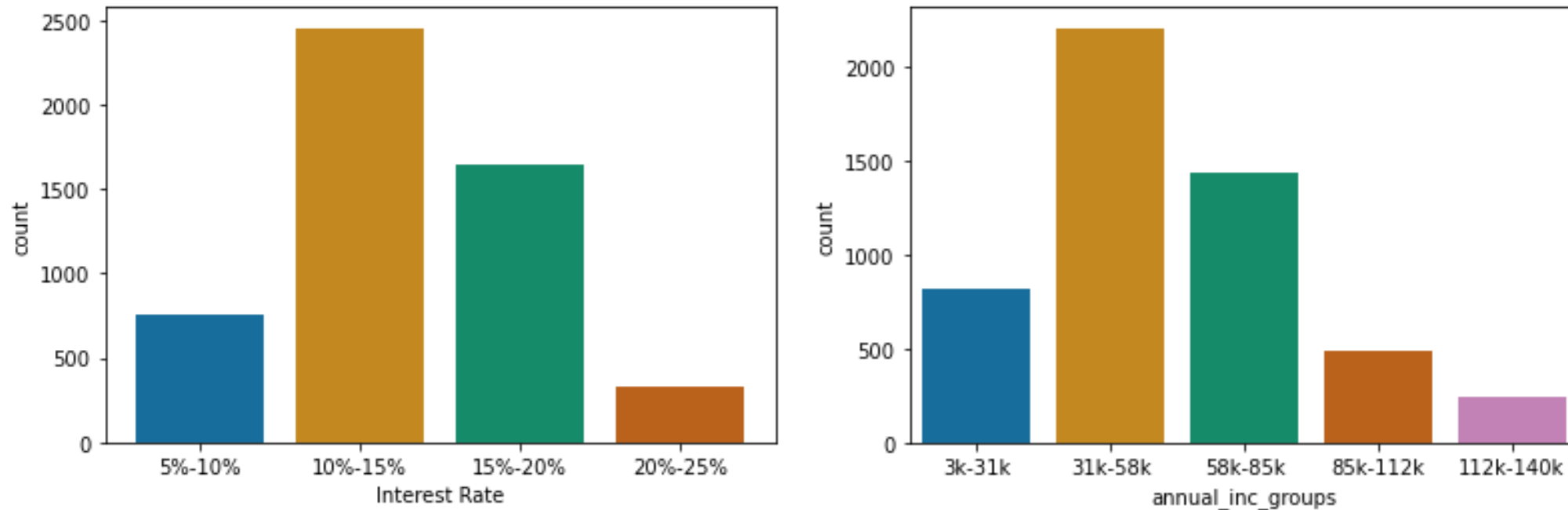
- 60 month loan default count is more than 36 months loan defaults.
- Maximum number of default have "Rental" home ownership

Loan default Analysis



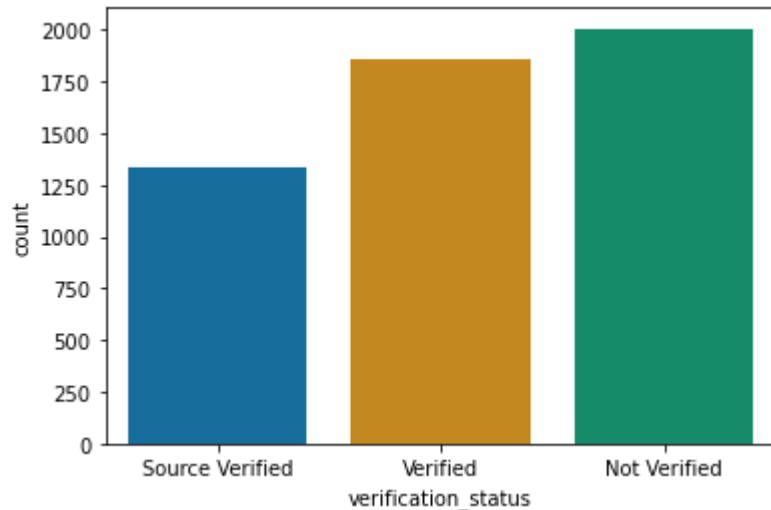
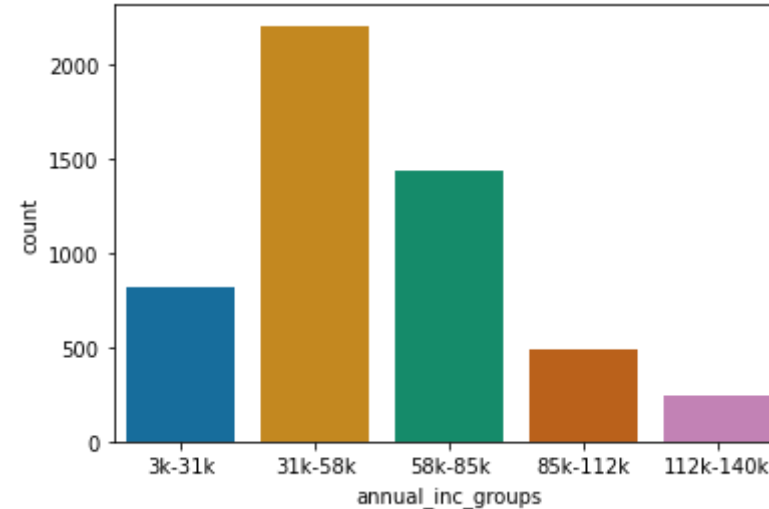
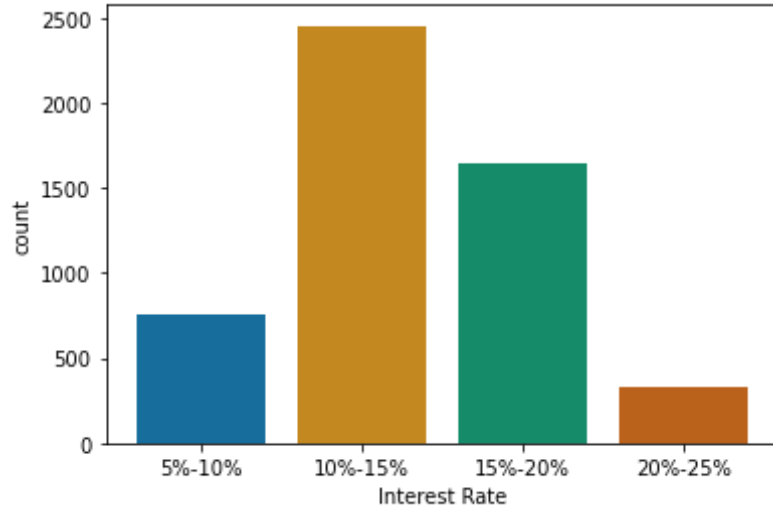
- Maximum number of defaults are with the purpose of 'debt_consolidation'.

Loan default Analysis



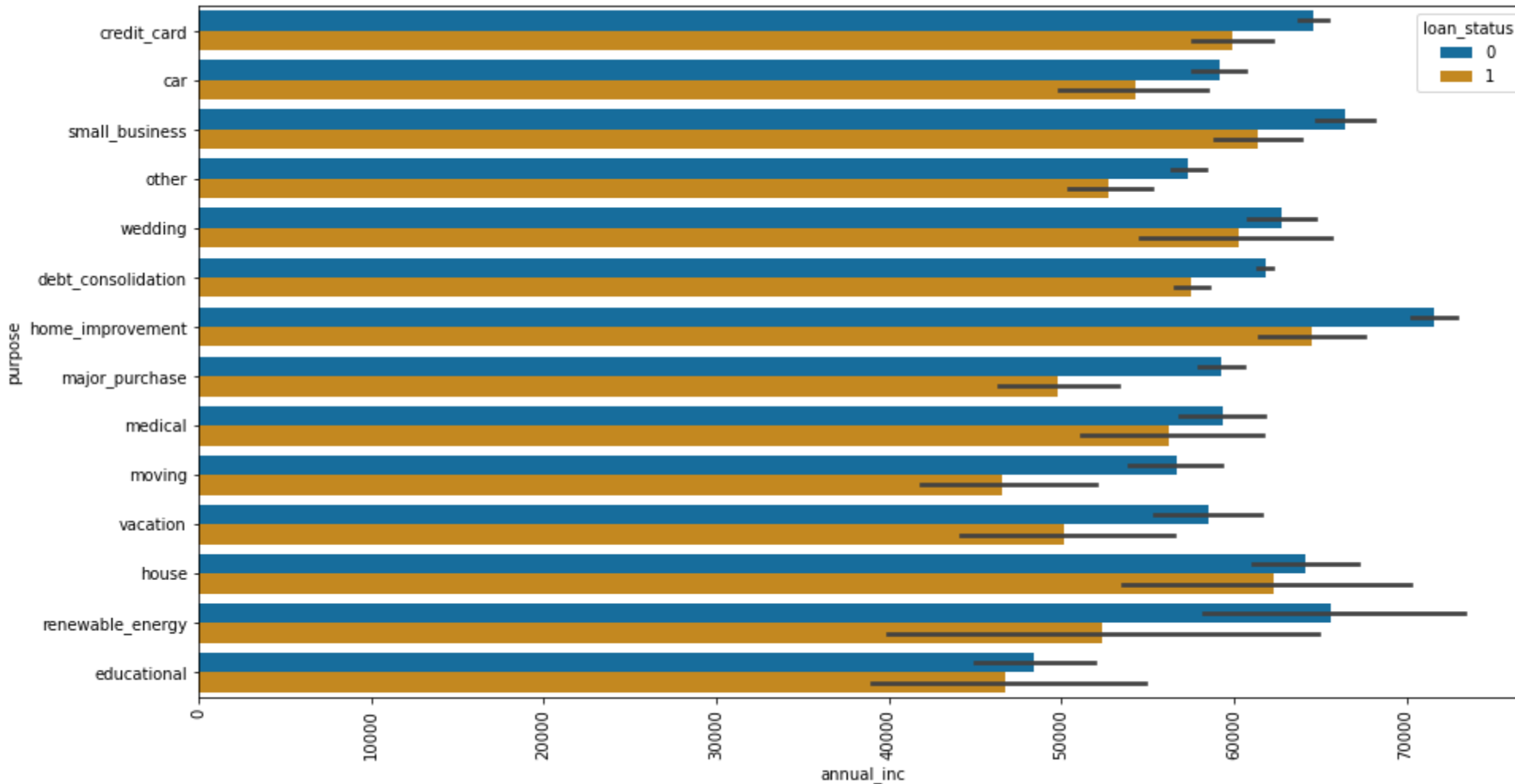
- Most number of defaults are the applicants who receive interest at the rate of 10-15%
- Most number of defaults are the applicants whose annual income is in between 31 to 58K.

Loan default Analysis



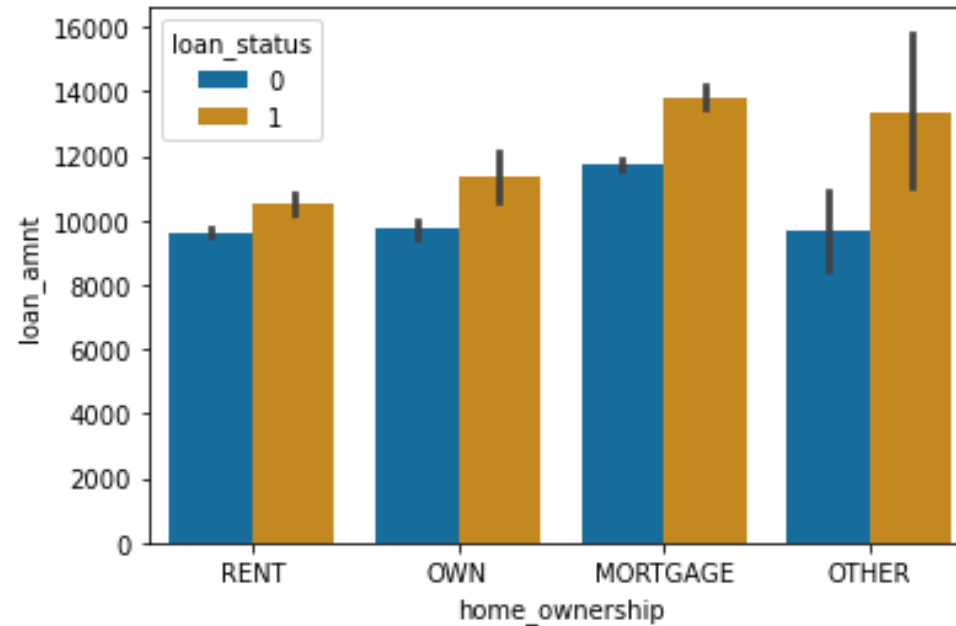
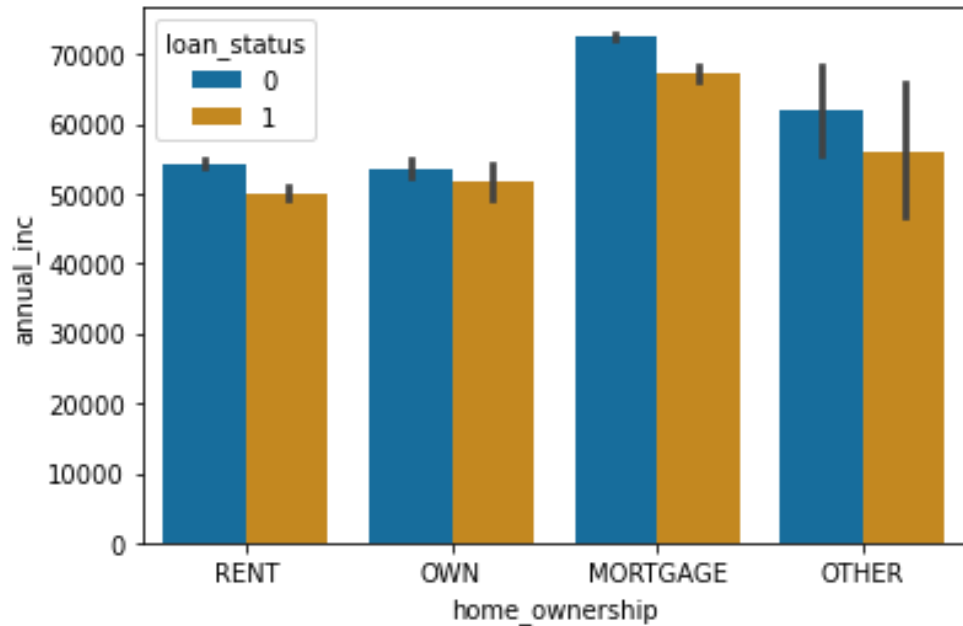
- Most number of defaults are the applicants who receive interest at the rate of 10-15%
- Most number of defaults are the applicants whose annual income is in between 31 to 58K.
- Most number of defaults are the applicants whose verification status is " Not verified"

Loan default Analysis



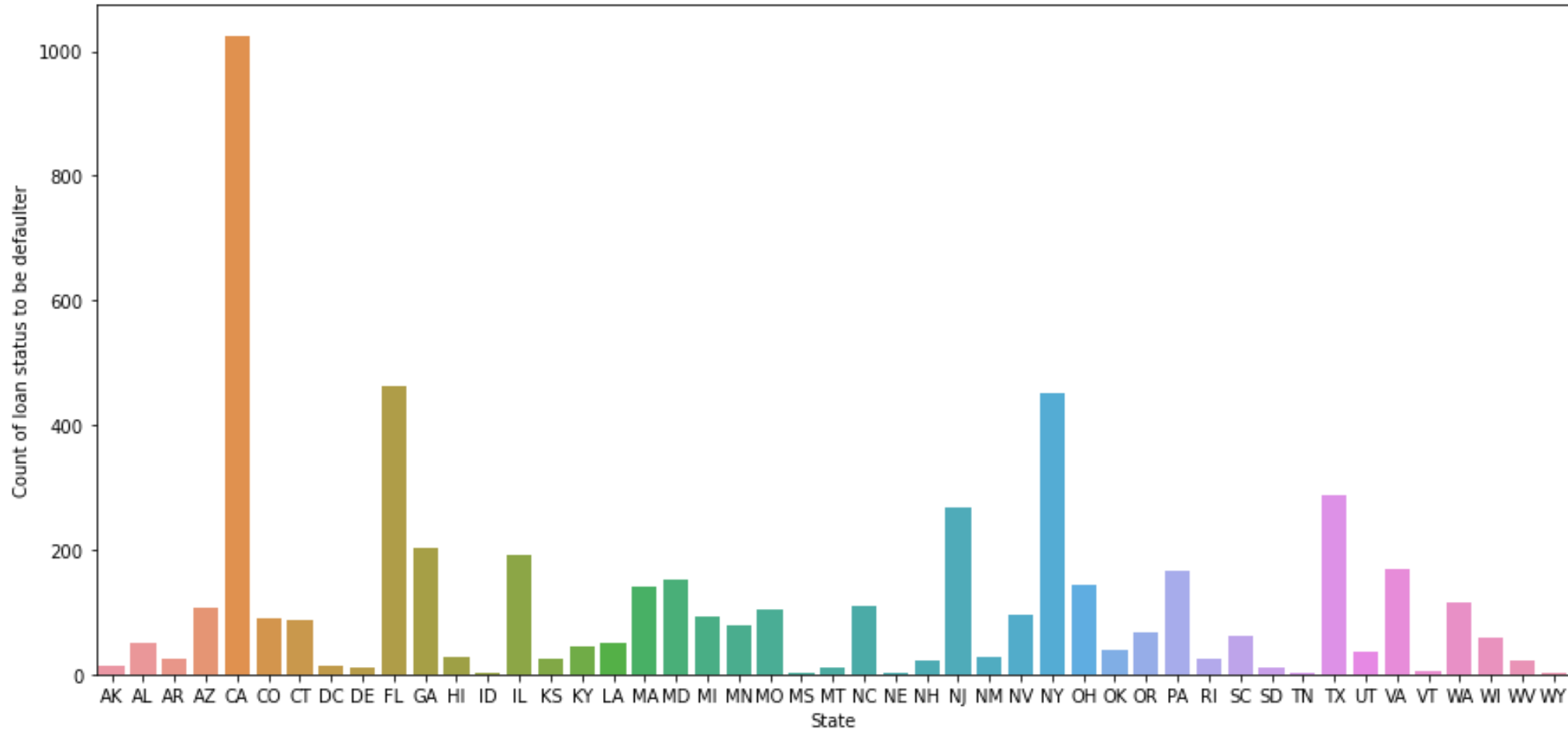
- Applicants with higher salary mostly applied loans Home improvement, house, renewable energy and small businesses.
- Applicants taking loan for 'home improvement' and have income of 60k -70k.

Loan default Analysis



- Applicants whose home ownership is 'MORTGAGE' and have income of 60-70k
- Applicants whose home ownership is 'MORTGAGE' and have loan of 14-16k

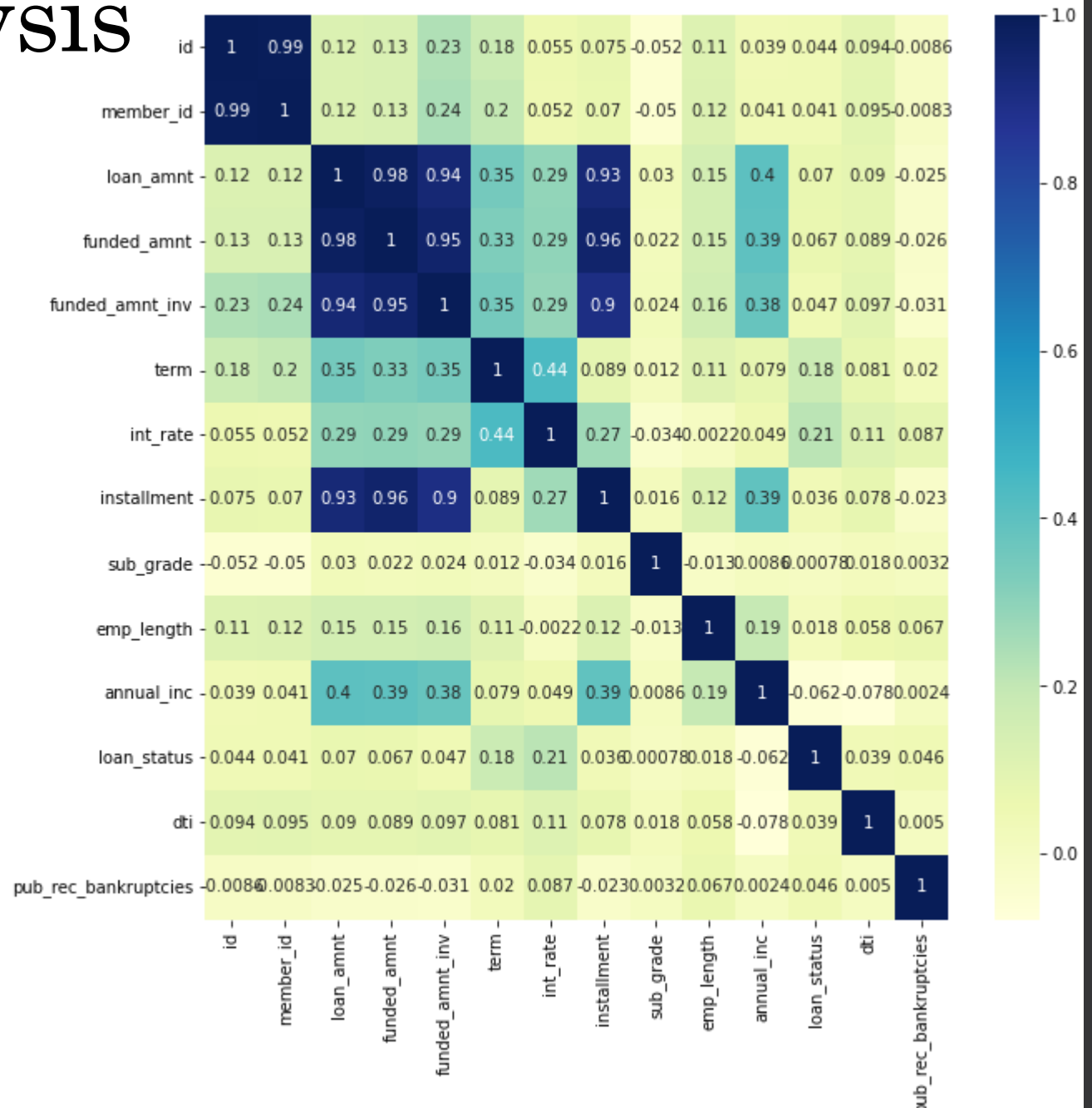
Loan default Analysis



➤ Applicants from the state CA are having high probability to be default

Loan default Analysis

- Mean, 25% and 75% Loan amount of Fully paid and charged off is exactly same.



Recommendations

- Grades are good metric for detecting defaulters. Lending club should examine more information from borrowers before issuing loans to Low grade (G to A).
- Lending Club should control their number of loan issues to borrowers who are from CA, FL and NY to make profits.
- People with rental home ownership are highest defaulter. Check the home ownership.
- Borrowers with mortgage home ownership are taking higher loans and defaulting the approved loans. Lending club should stop giving loans to this category when loan amount requested is more.
- Home improvement, house, renewable energy and small businesses loans are defaulted more. Lending club should stop/reduce issuing the loans to them.
- Loans having higher interest rate have more defaulters. Check the background of applicant thoroughly if interest rate is high.