Lending Club Case Study

August 9

2023

Exploratory Data Analysis of the Loan data set of a lending business , to analyze the risk of the bad loans and help business understand different variants that indicates if an applicant 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

EDA of Loans

Presented by Goutham and Swapna

Lending Case Study

Problem Statement

This case study is to understand the data and analyse the patterns of defaulters (risky applicants) to help the business in loan processing and avoid the financial losses in way of bad loans.

Analysis Approach

The loan data given for analysis has 39717 rows with 111 columns related to the individuals who has taken Loans form the business. They fall under different categories like Loan Fully Paid, Currently still Paying and Defaulters. Below steps we have followed to do the analysis.

Understanding The Data

Loan Data which is given in the csv format file. It has 39717 rows and 111 columns. It has data of different applicants whose loan have been processed and having different loan statuses. There are three different loan statuses and they are

Loans Fully Paid - The customer has taken the loan and repaid it in full.

Current - The customer who is still paying the loan

Charged Off - The customer who has defaulted his loan payments.

This dataset has all the information of its customers loan and different attributes like id, loan amount, loan status, employment details, term , issue date ,purpose of loan, interest rate of the loan etc.

Data Cleaning

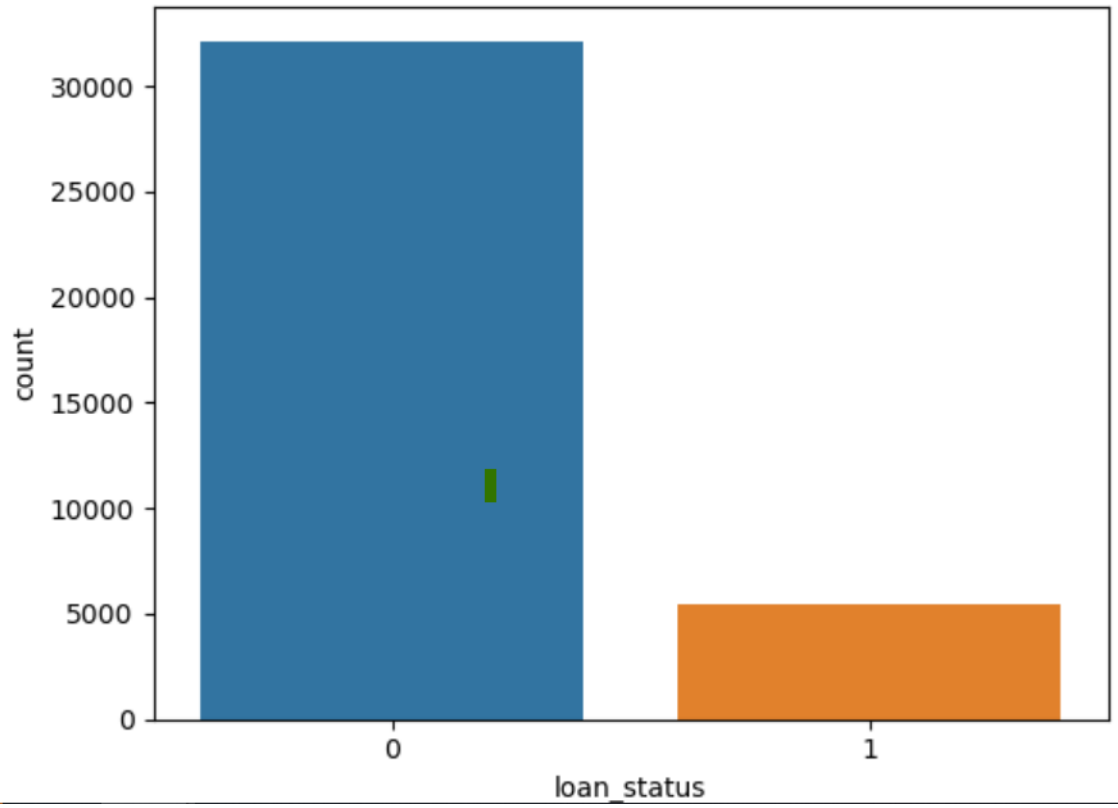
As we are analysing the risk of defaulters, we considered only the customers who are fully paid and defaulter's data as the current loan applicant will not give the required data to analyse risk of bad loan. While analysing the data we have noticed there are many columns which are having missing data and also some columns which are irrelevant for our analysis like 'pymnt_plan', 'initial_list_status', 'collections_12_mths_ex_med', 'policy_code',

'application_type', 'acc_now_delinq', 'chargeoff_within_12_mths', 'delinq_amnt', 'tax_liens' . *Hence we dropped such columns from our dataset. There are some columns like interest rate which is in object format we converted it to numeric format to have better analysis.*

Univariate Analysis

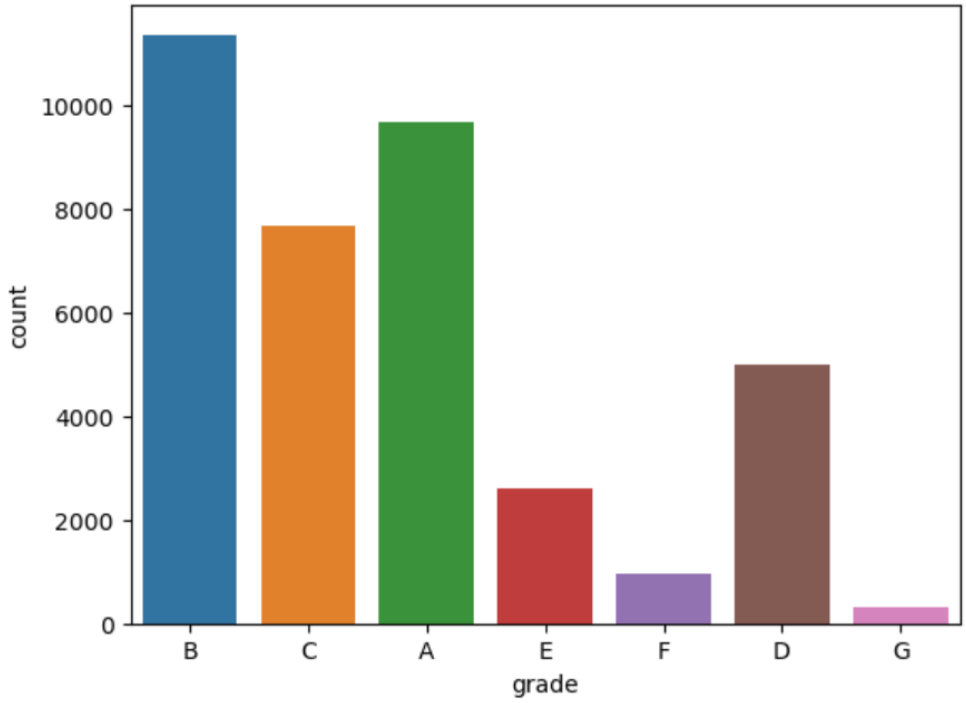
1.Loan Status

Out of all the records there are 32145 records as ***fully paid*** *and 5399 as* ***Charged Off i.e defaulted loans .*** We can say there are 16.79 % of defaulters



2. Grades

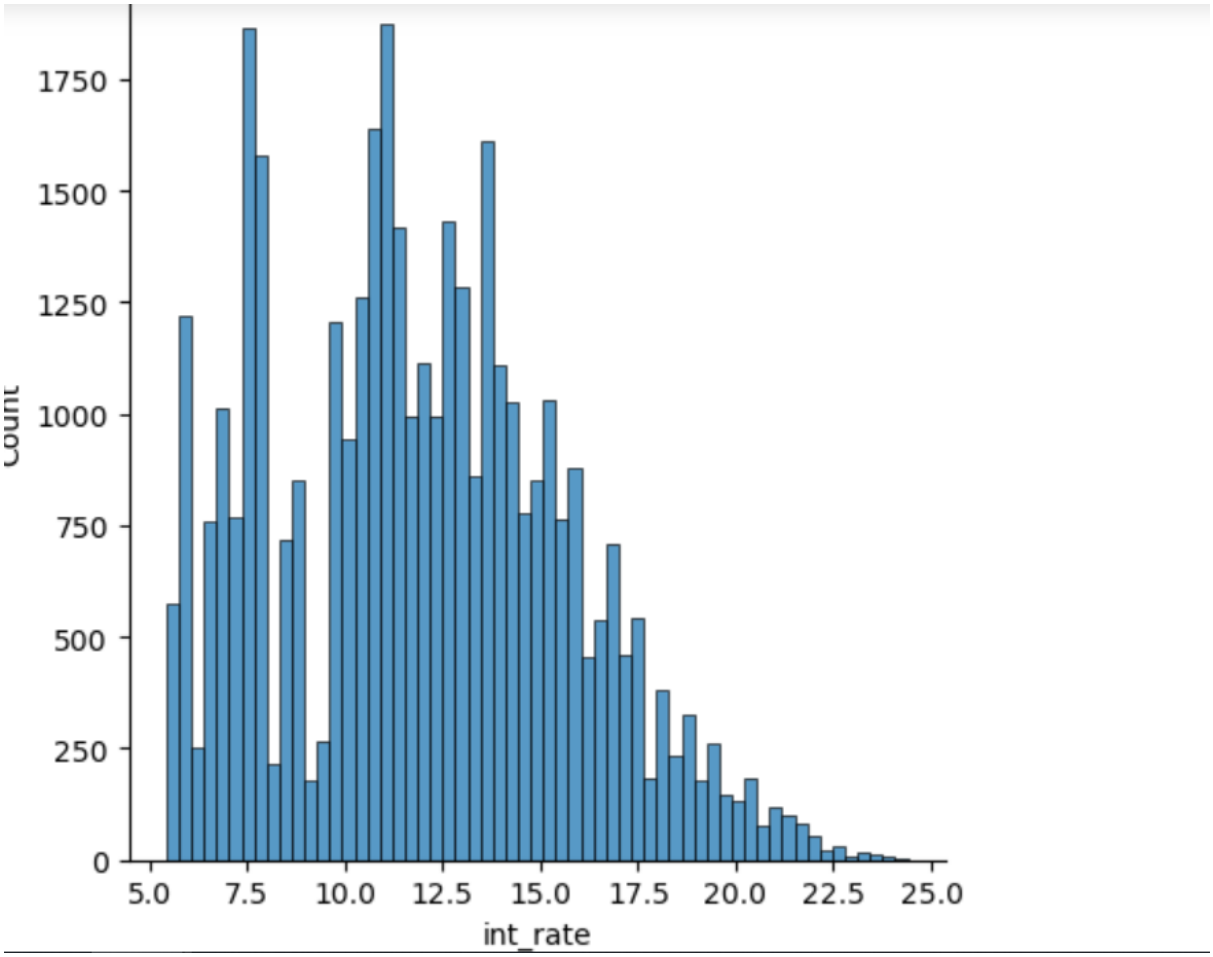
There are different Grades are associated with the applicants based on their credit scores . It ranges from A to G .



In this we can see many of the applicants are in Grade B and lowest count of grades is F and G but in later section we can see that F and G has more influence on default rate.

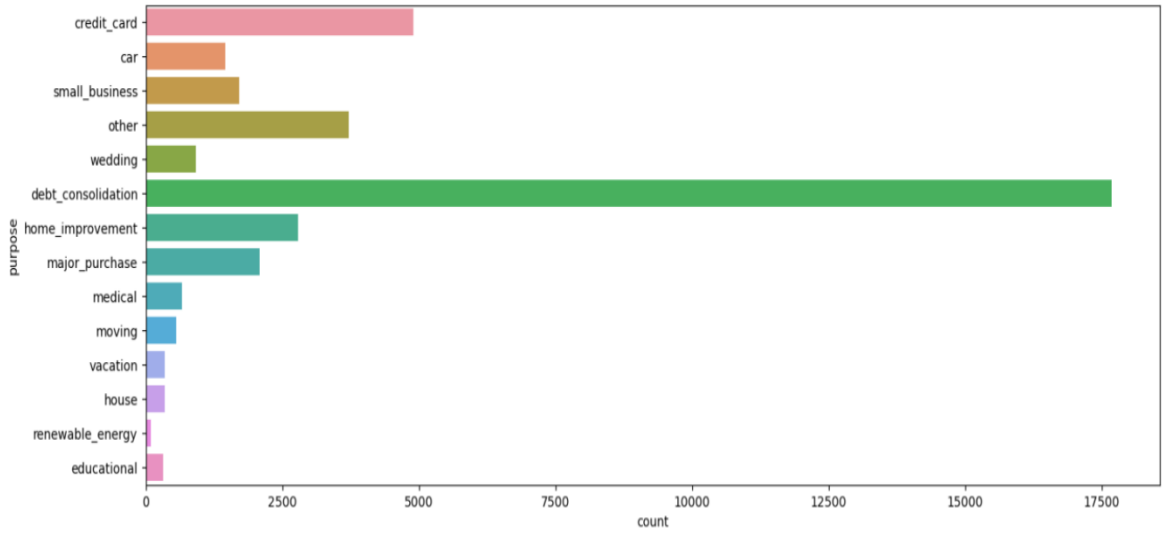
3. Interest Rate

There are different Interest Rates are given on bases the loan applicants attributes like his Grade , employment, amount of loan etc .It has range from 5.4% to 24.59% .



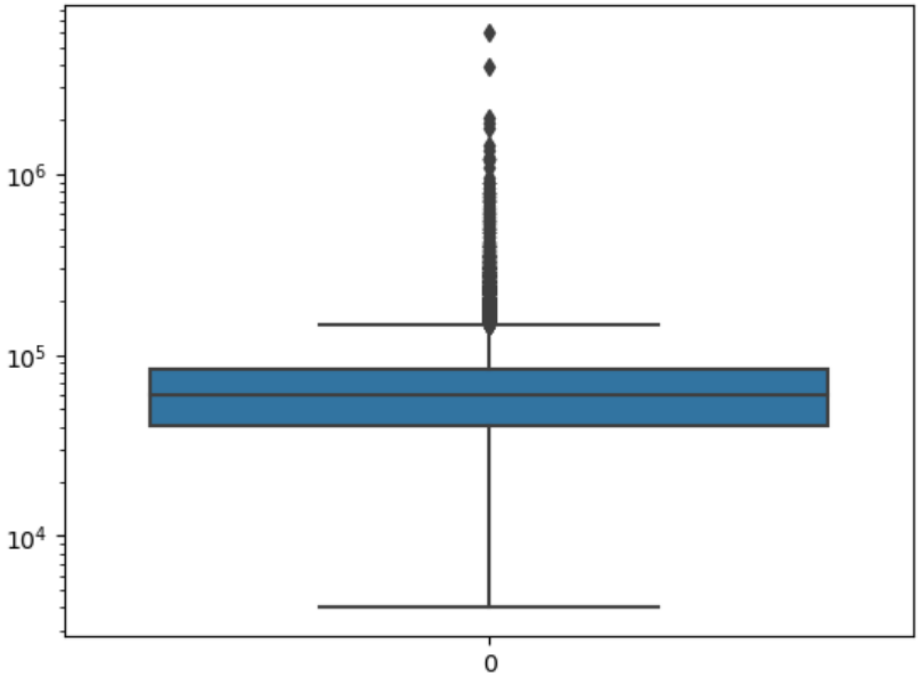
4. Purpose

The loans are processed for different purposes like Credit Cards, Small Business Loans ,Debt Consolidation etc.



5. Annual Income

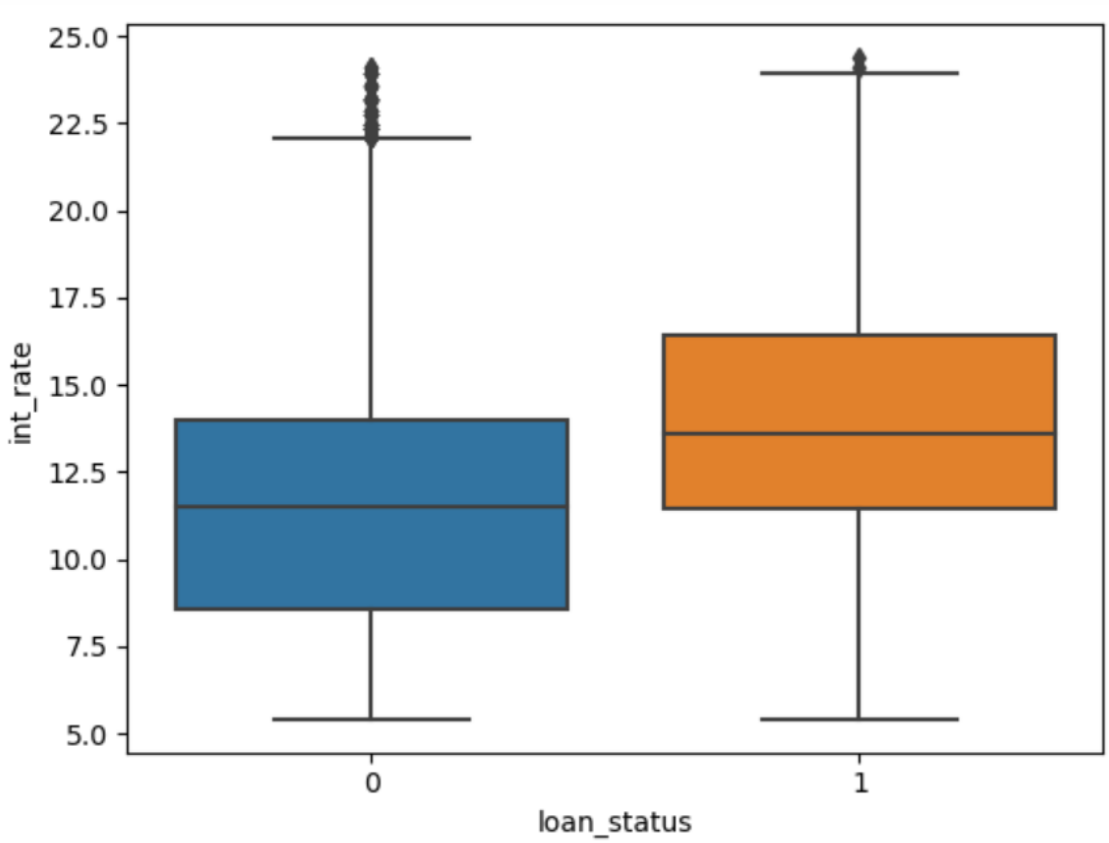
Annual Income of the applicants is given for all the applicants.



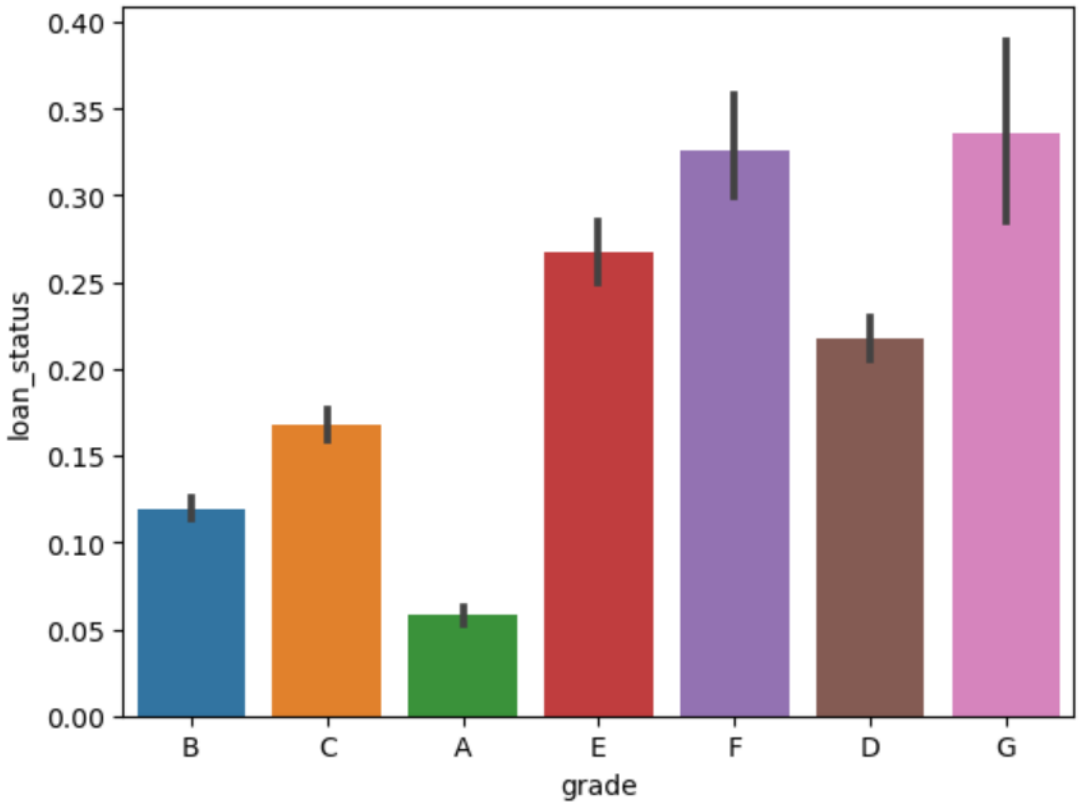
Bivariate Analysis

To analyse the data we used two different columns of data and try to understand is the impact of the data on each other.

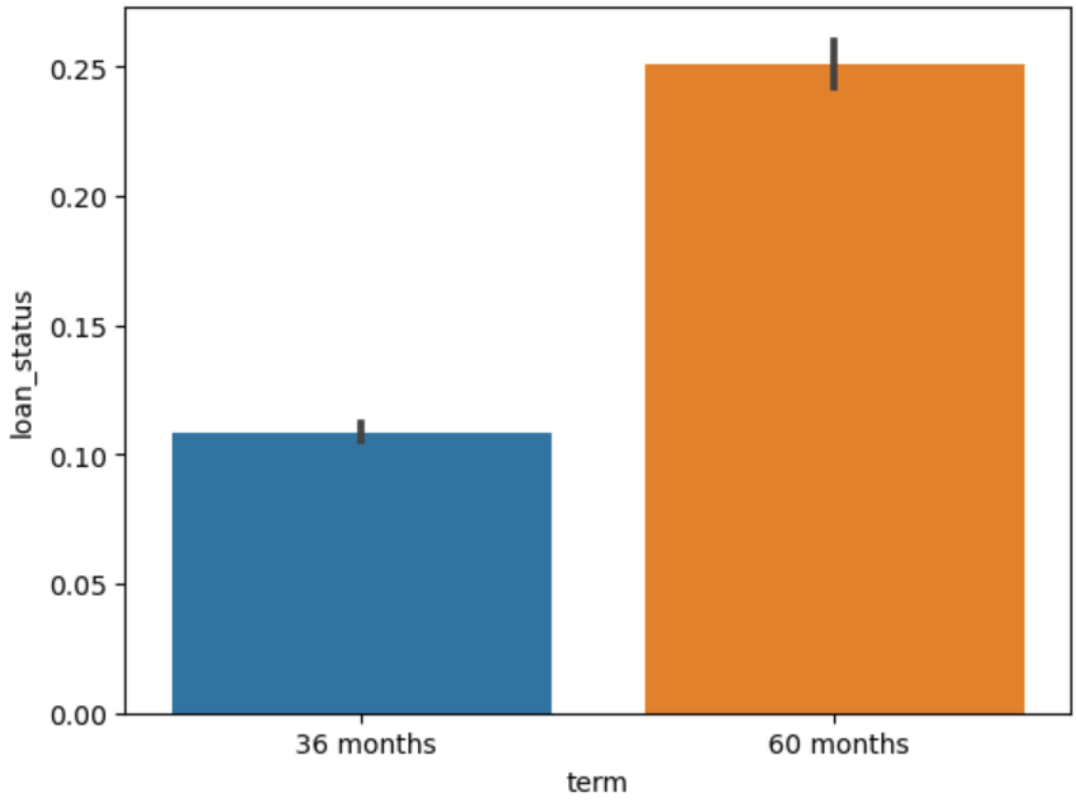
1. Loan Status and Interest Rate - When the interest rate is more than 13% then there are more defaulters. In the below image Loan Status 0 indicates – Fully Paid and 1 indicates – Charged Off i.e Defaulters.



2. Loan Status and Grade - The grade F and G has more defaulters compared to other grades.



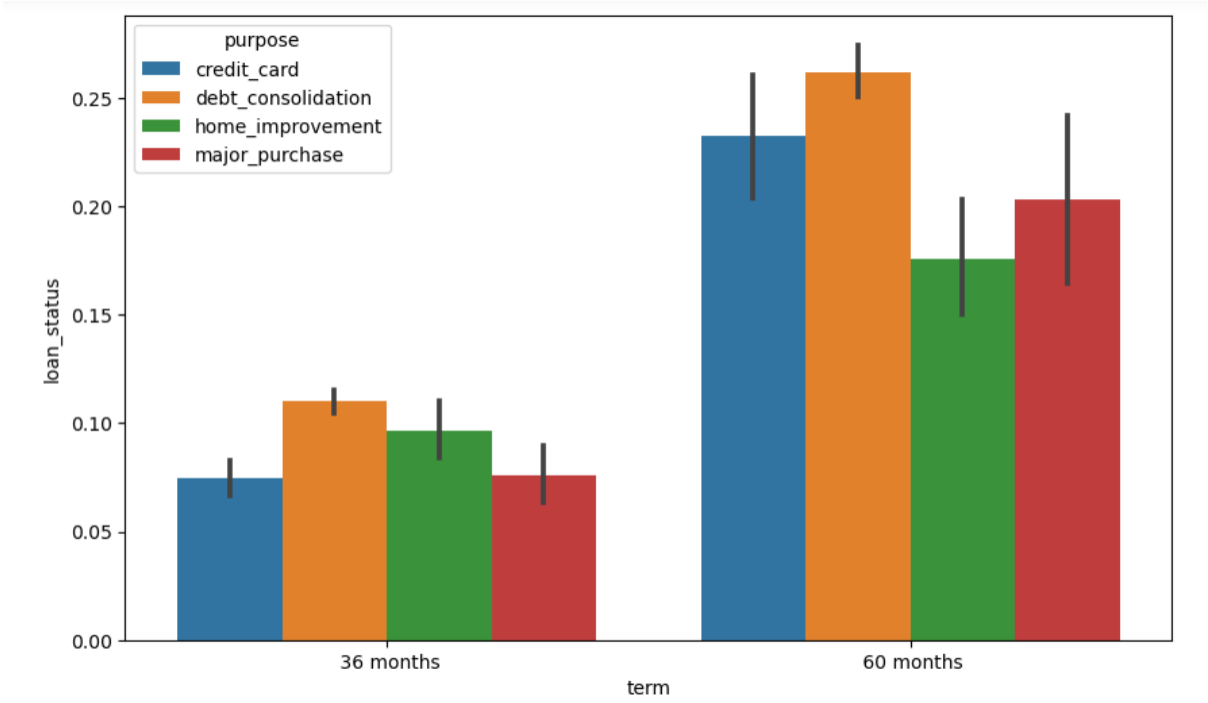
3. Loan Status and Term of Loan – There 2 different term of loan 36 months and 60 months. We noticed 60 months term has more defaulters than 36 months.



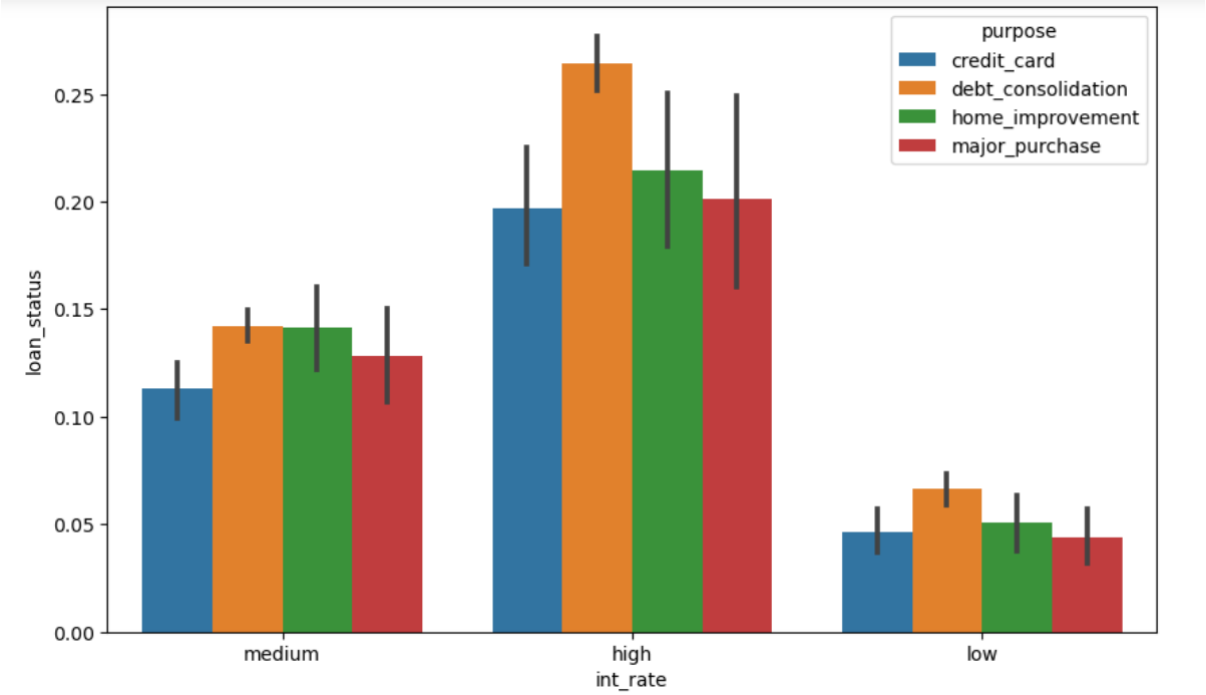
Segmented Analysis

We compared different subsets of data and tried to understand the impact across different segments.

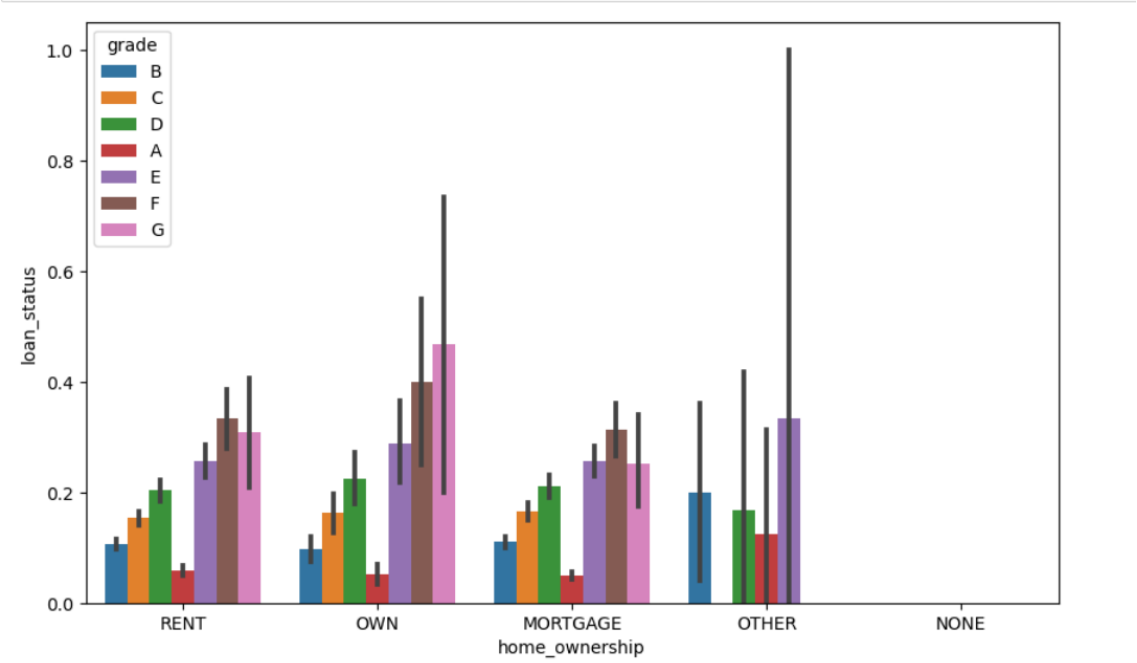
1. *Loan Status, Term and Purpose – When the term is 60 months and Purpose is Debt Consolidation or Credit Card there are more defaulters compared to other purposes and term being 36 months.*



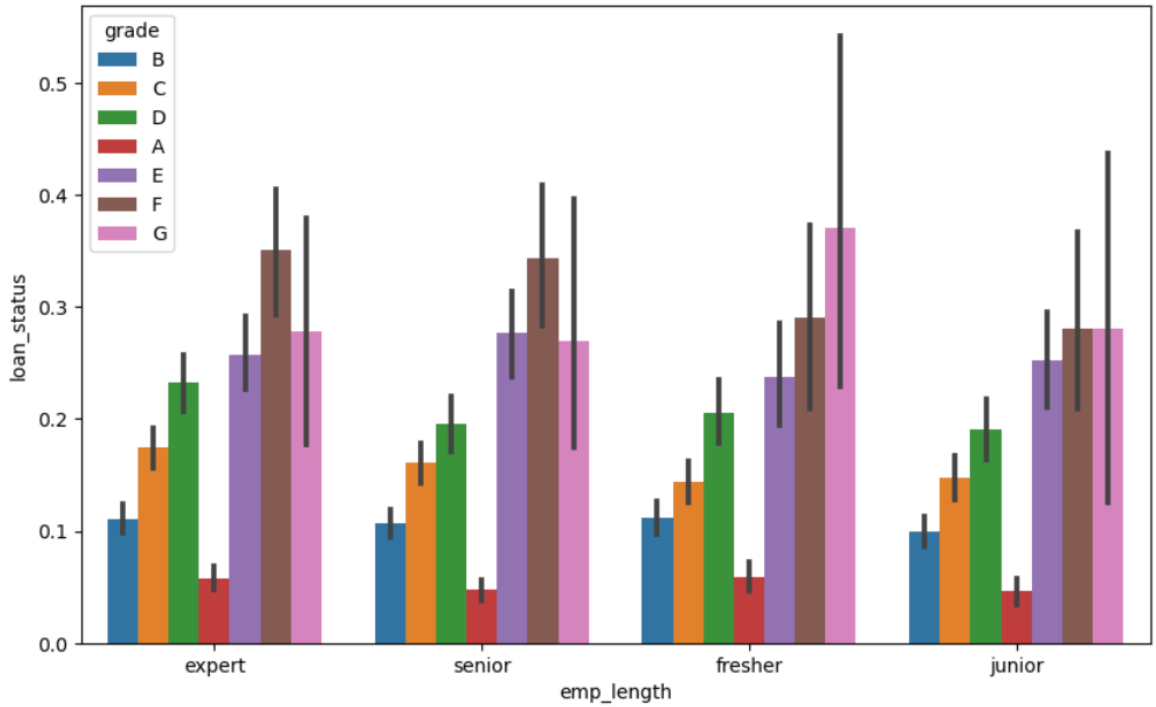
2. *Loan Status , Interest Rate and Purpose – When the interest rate is high and purpose is debt consolidation we have noticed there are more defaulters.*



Below are few more analysis which has high the impact on the defaulters rate.



When the house ownership is Own and the grade is G then default rate is more than 0.4



We categorised the applicants work experience as Expert, Senior ,Junior and Fresher based on their length of employment. When applicant is a Fresher and his grade is G then defaulter rate is 0.35 and When the grade is F and he is either Expert or Senior their defaulter rate is more than 0.3