Exploratory Data Analysis –Case Study

Overview

Lending club is 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

This company 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.

Business Objective

Lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who **default** cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

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.

Data Sourcing

Loan data set

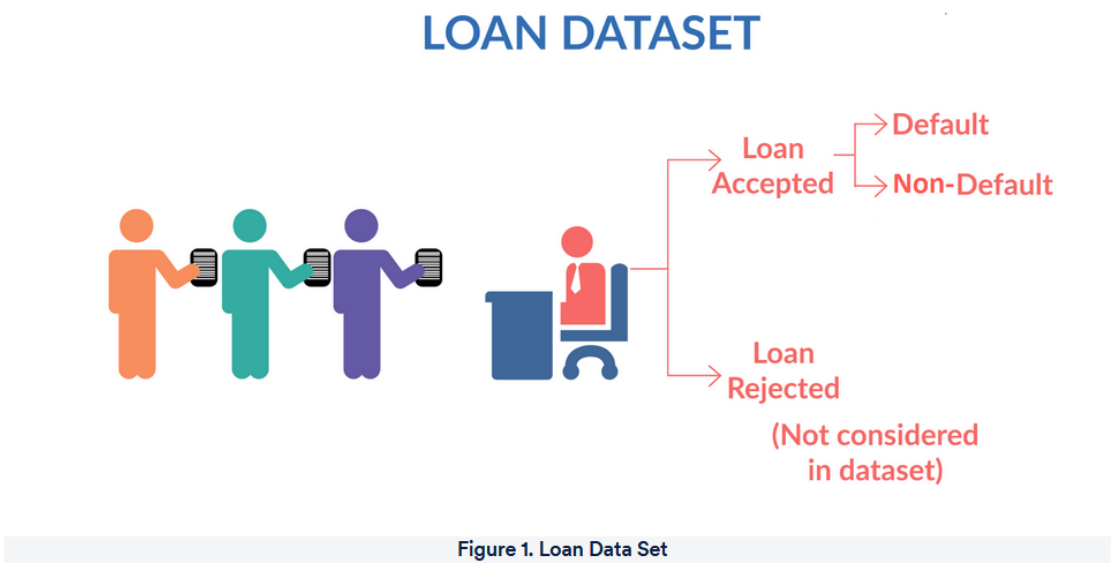


Figure 1. Loan Data Set

Data Understanding

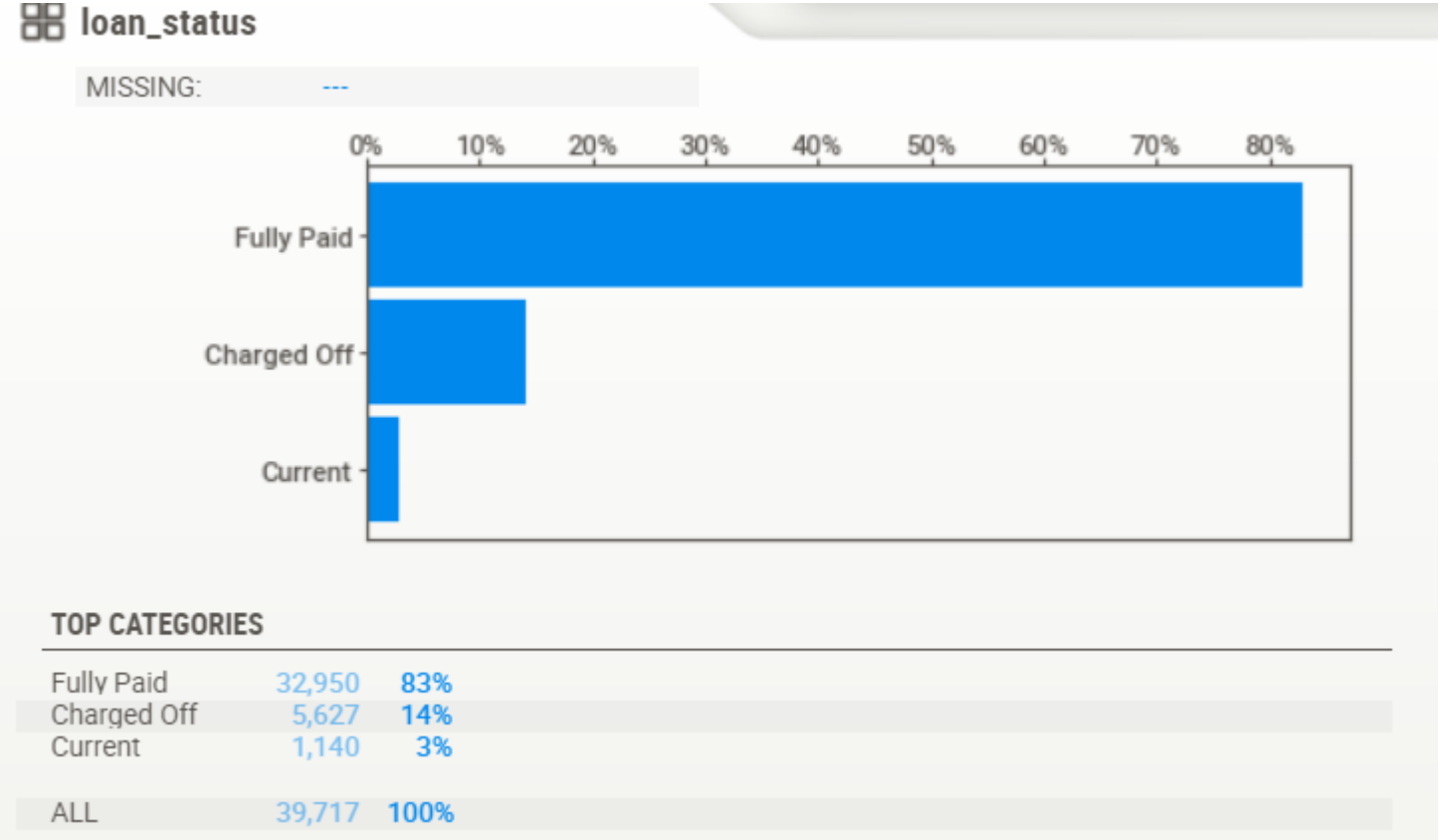
Types of attributes found in the Loan Dataset

1. Customer attributes
   - - emp title, emp_length,home_ownership , annual_inc , zip_code
2. Loan attributes
   - -member_id , loan_amnt , funded_amnt, term , int_rate
3. Behavior attributes
   - delinq_2yrs , inq_last_6mths , revol_bal , out_prncp , total_pymnt

Other attributes which doesn't impact

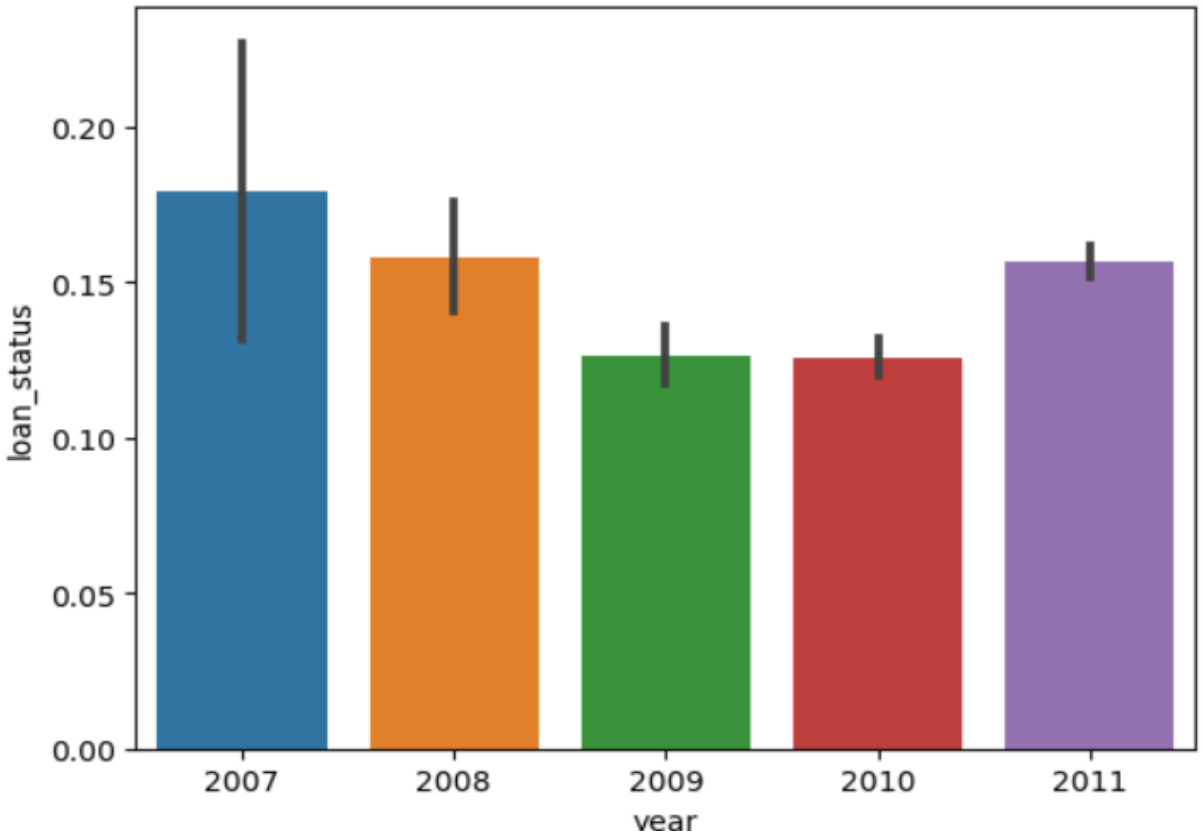
1 Unique values - tax_liens, delinq_amnt, chargeoff_within_12_mths,acc_now_delinq,application_type,policy_code collections_12_mths_ex_med, initial_list_status,pymnt_plan

Target Attribute – Loan Status



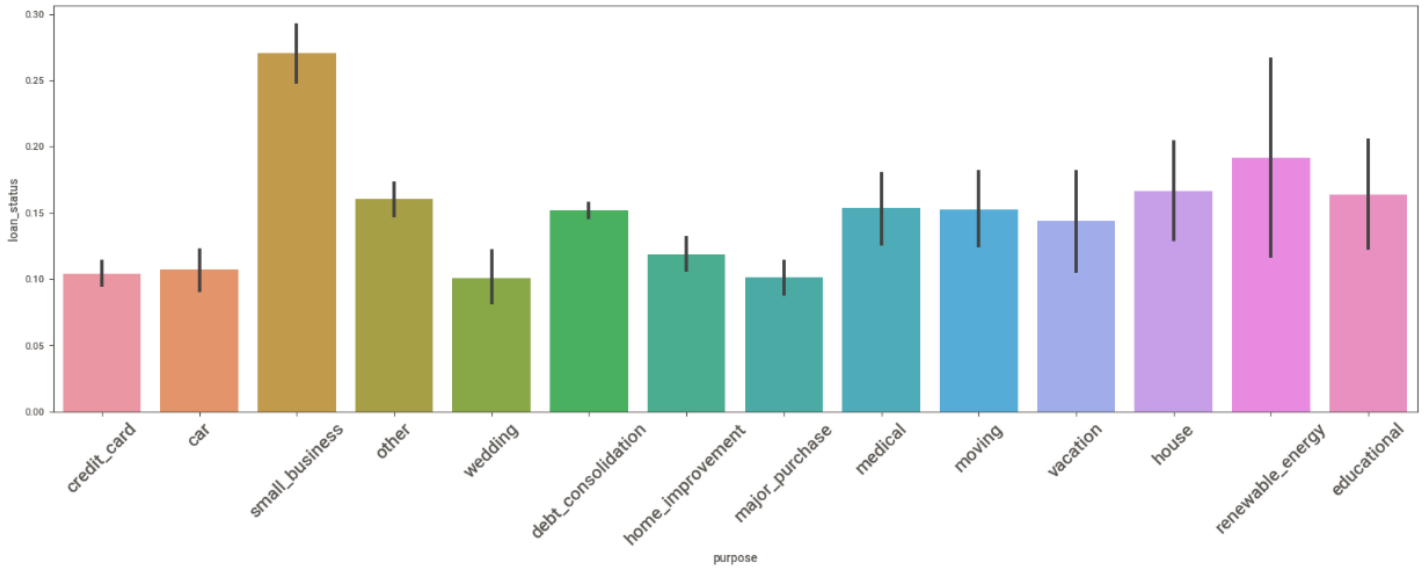
As per the given records default is 14%

Loan status distribution for the years



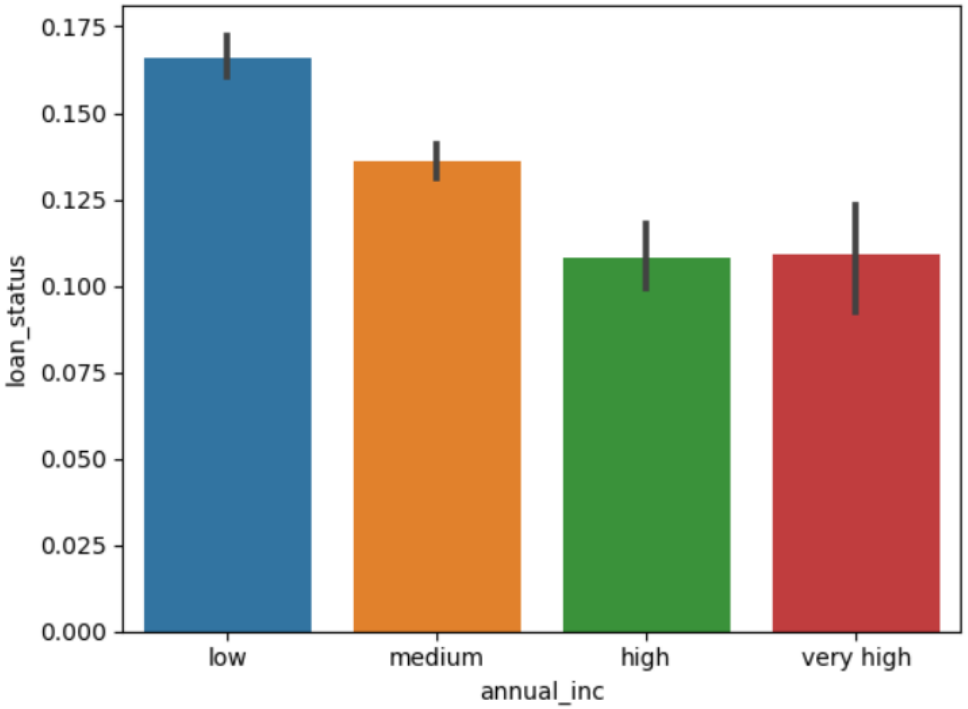
Higher default during the year of 2007 compared to 2009 and 2010

Loan status distribution onpurpose

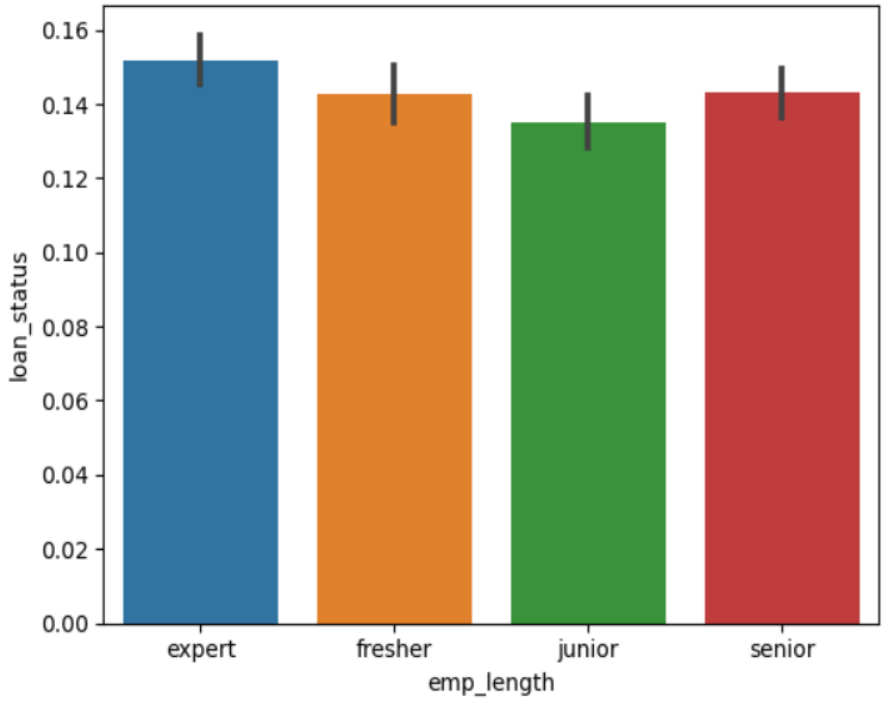


1. Small_business – loan purpose has the highest defaulter
2. Top 3 loan purpose defaulter includes small business, renewableenergy, house

Loan status distribution on customer skills

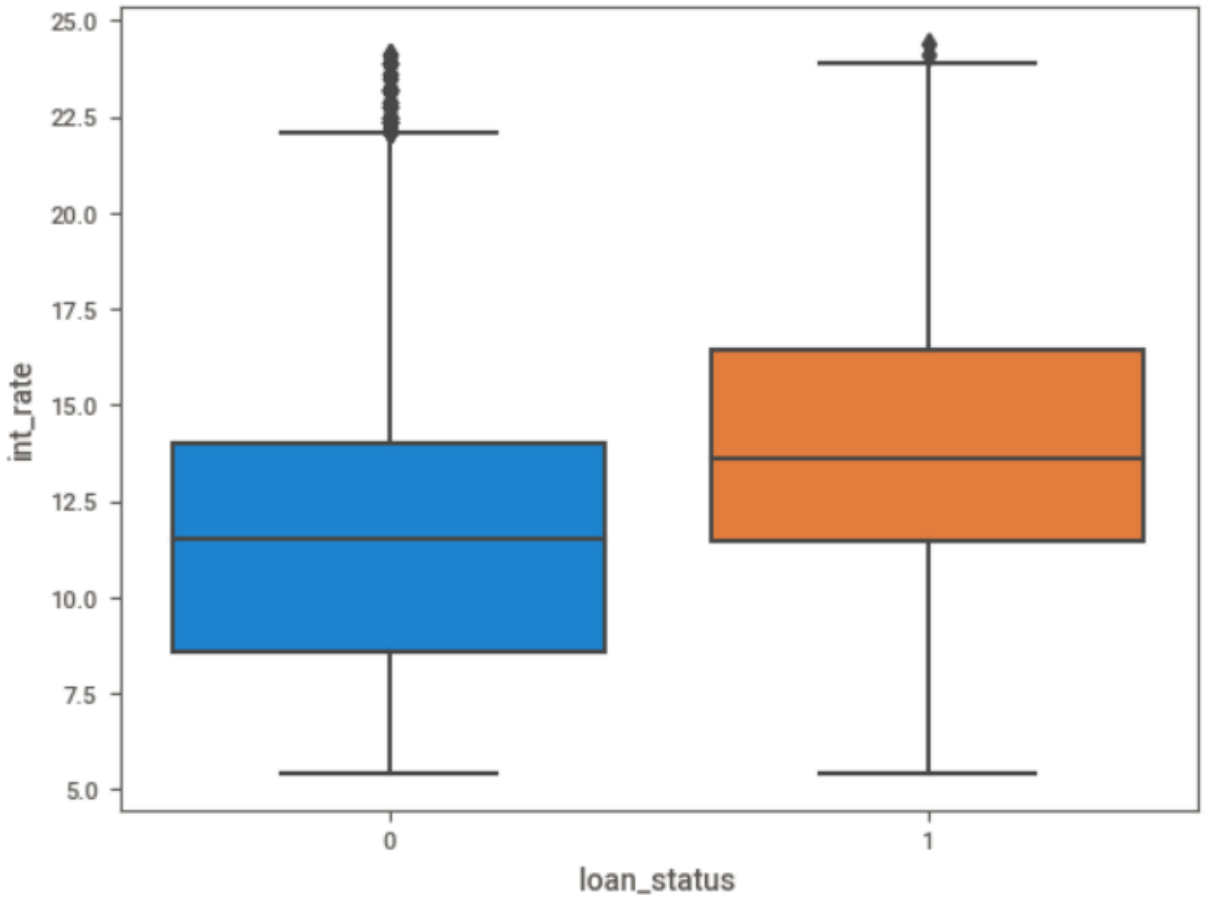


Customers with low salary has highest default than customers with high salary



Experienced customers has high default rate than the entry level customers in their respective job

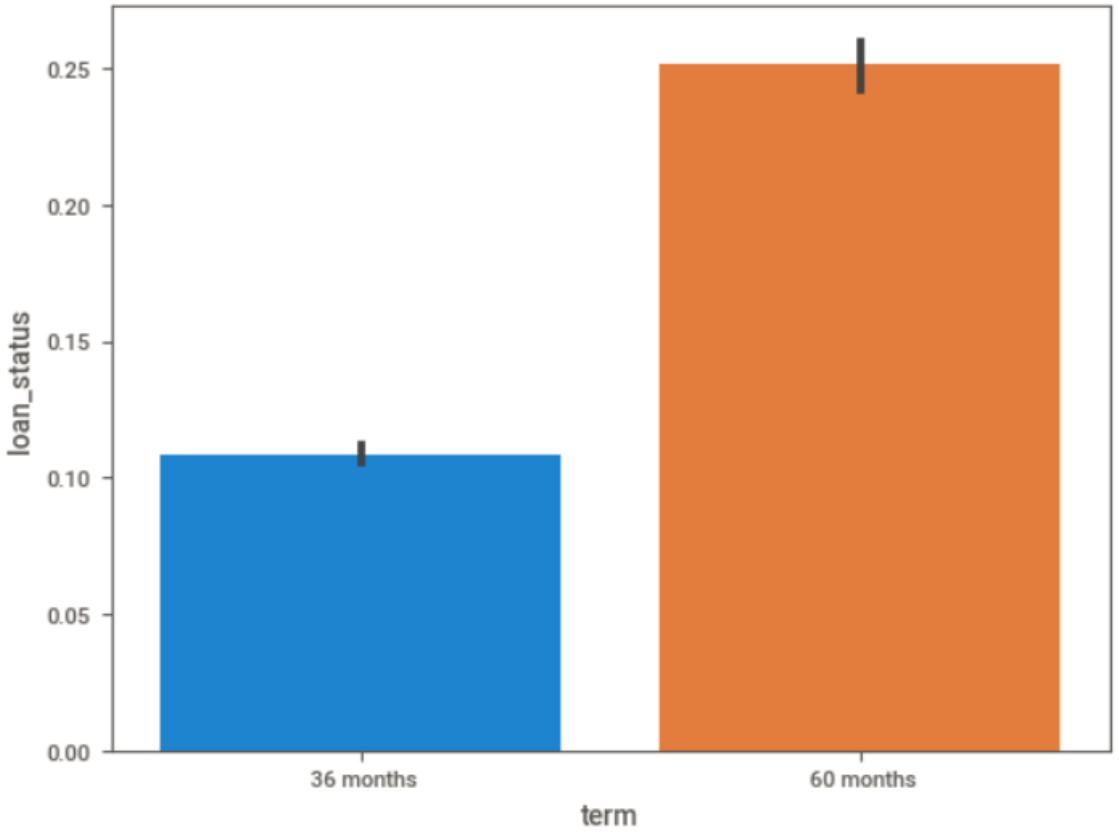
Loan status impact based on interest rate



Highest interest rate above 14% causes more defaulters

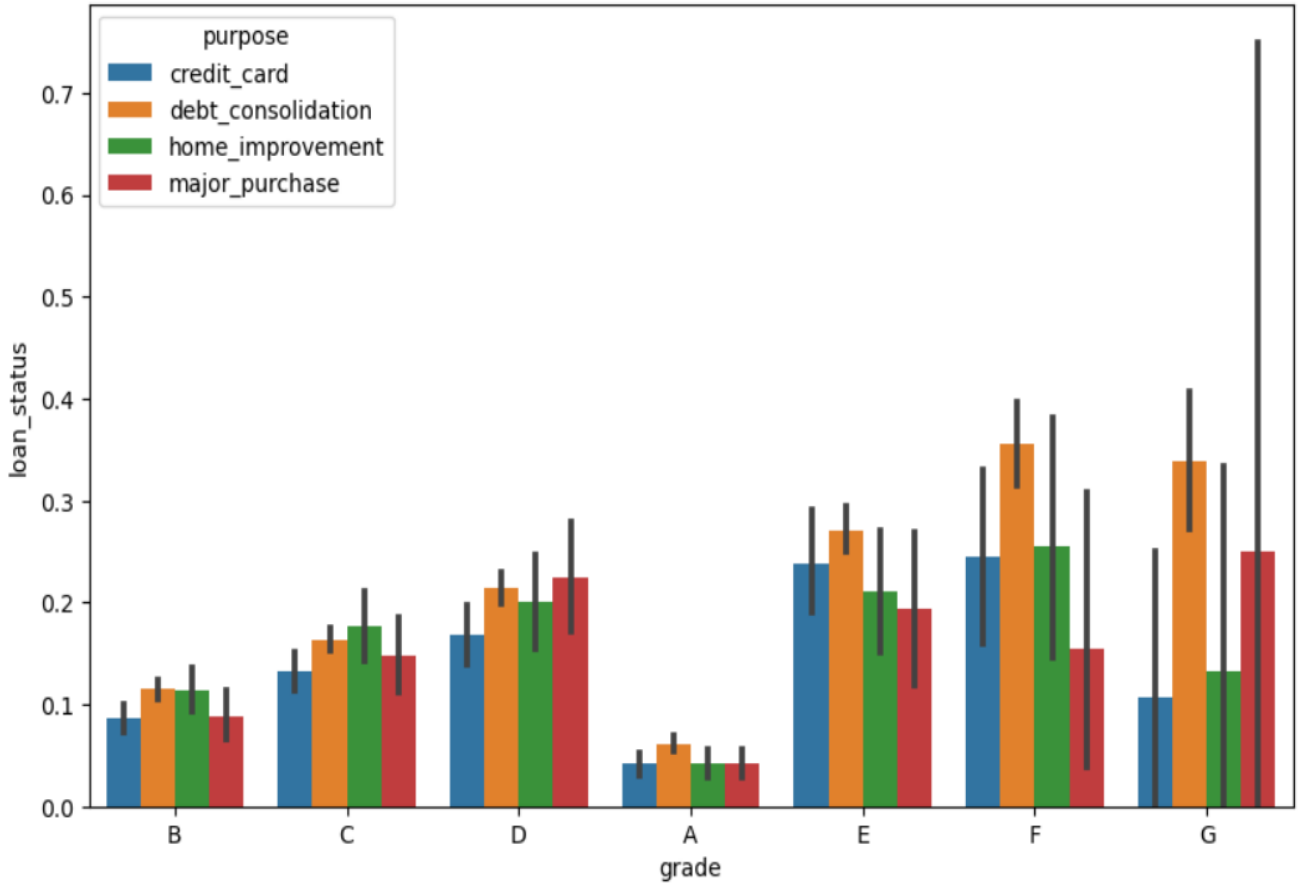
Increase in interest rate increases the customers to be charged off

Tenure impact on Loan status



Increase in tenure increases the defaulters by ~60%

Loan status impact by grade and purpose



Grade G with purpose major purchase has the highest defaulter compared top grade A major purchase