Lending Club Case Study

Rashmi Salankar

Purpose of the case study

The lending club has provided a dataset about its customers. The purpose of this study is to analyze this data and identify 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.

Understanding the Data

Variables dropped from the study:

- Variables which hold Nan value across the entire dataset

  e.g. open_acc_6m, open_il_6m etc.

- Irrelevant data which has no bearing on loan application

  e.g. desc (contains descriptive text about the loan)

Understanding the Data

- Customer character variables : Variables whose values cannot be predicted at the time of application approval. This is data related to the current but not defaulted loans.

  e.g. last_pymnt_amt, recoveries, delinq_2yrs etc.

Since these variables are not available at the time of loan approval, they cannot be used as predictors for credit approval. Hence, we can safely drop them.
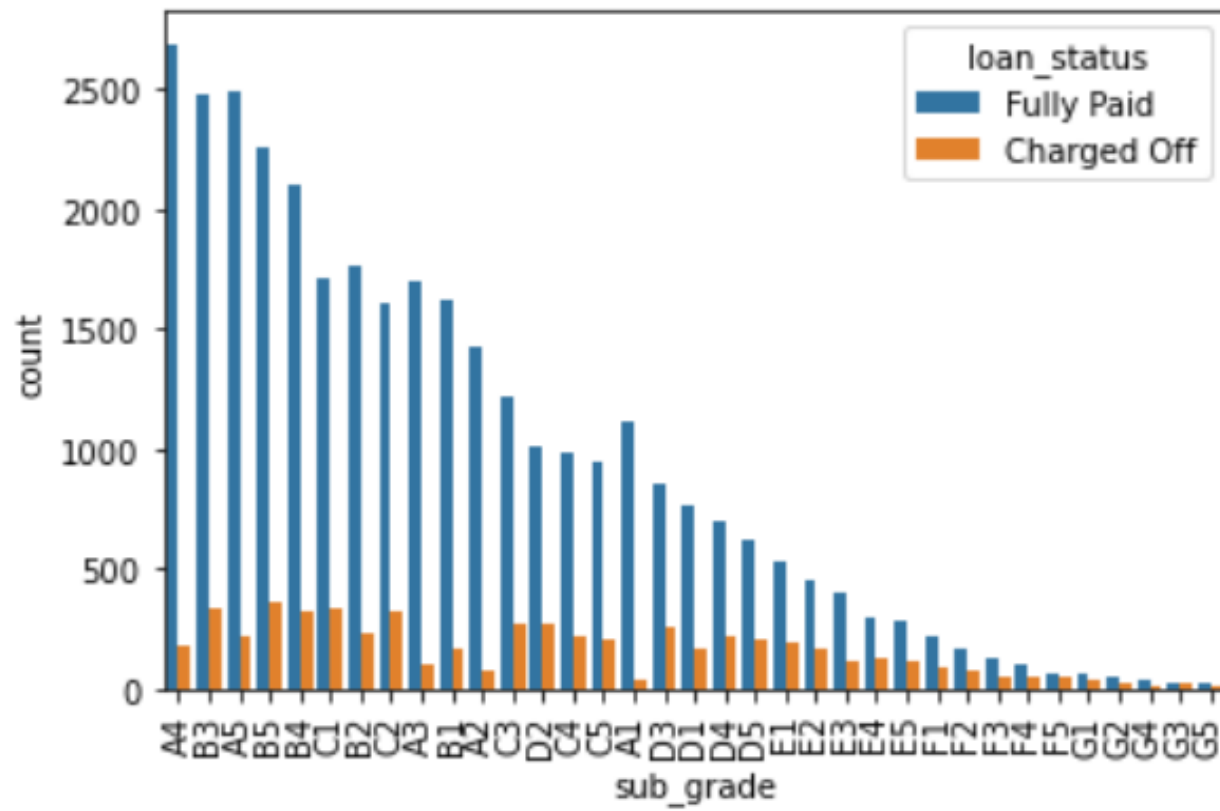
Variables considered for Analysis

- Grade
- Sub-Grade
- Debt to Income ratio
- Purpose of loan
- Home ownership
- Verification Status
- Employment Length

Grade
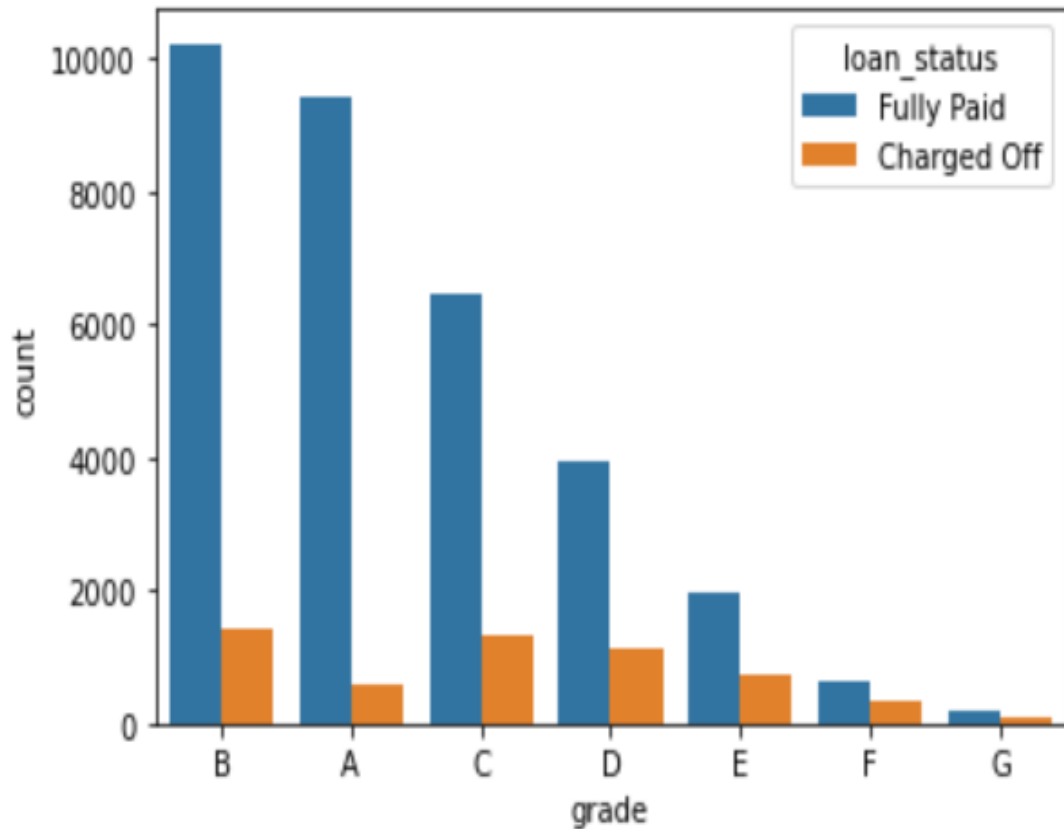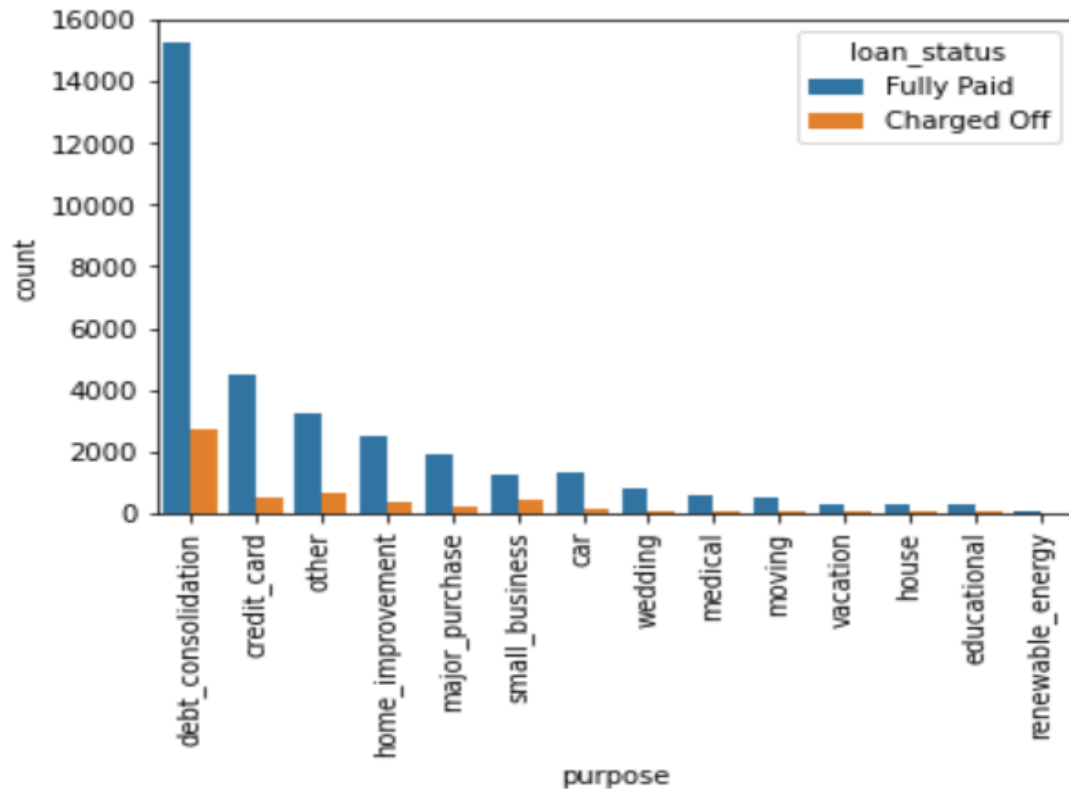

- Grade G has the highest percentage of defaulters (34%), closely followed by grade F (33%)

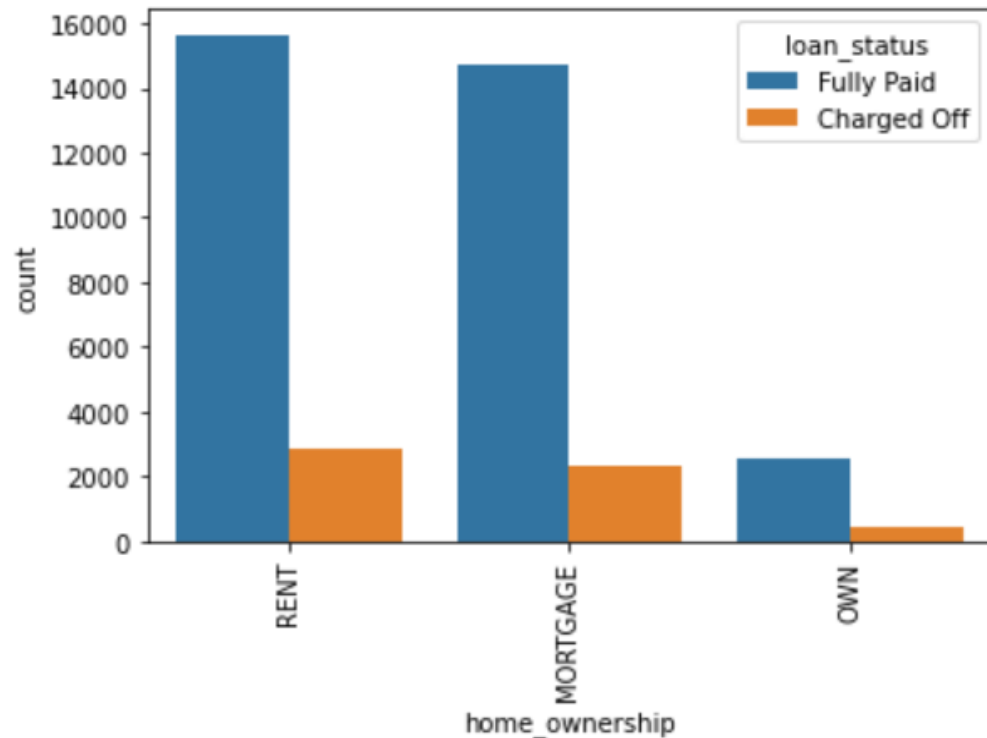- Grade A has the lowest percentage of defaulters (6%)

Sub-Grade

- sub-grade F5 has the highest default rate of 48%
- sub-grade A1 has the lowest default rate of 3%

Purpose

- Debt Consolidation is the most common purpose in defaulted loans
- However, loans for the purposes of Small Business are most likely to default, having default rate of 27%

Home Ownership

- Even though, the Renters constitute the majority of the defaulters (over 50%), there isn't much difference in the percentage of defaulters among borrowers who rent, own or have a mortgage.

- Home ownership parameter does NOT have an effect of the number of defaulters

Verification Status
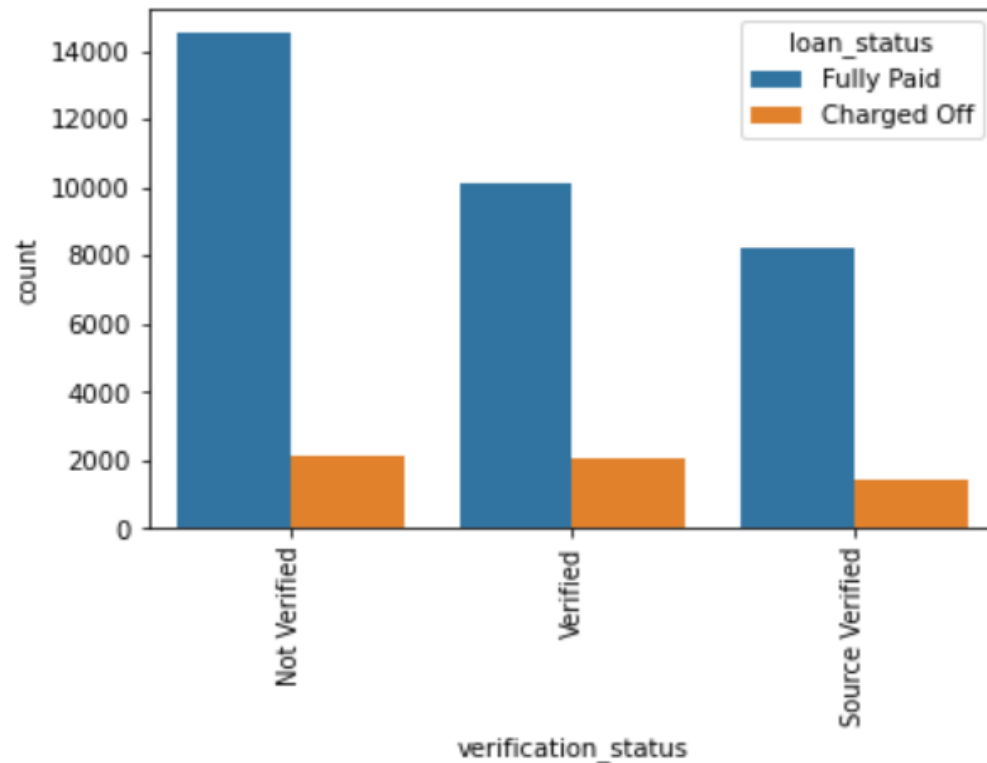
- 'Not verified' customers show a lower percentage of defaulting than the 'Verified' and 'Source Verified' customers.

- This parameter should be IGNORED.

Debt-to-Income Ratio
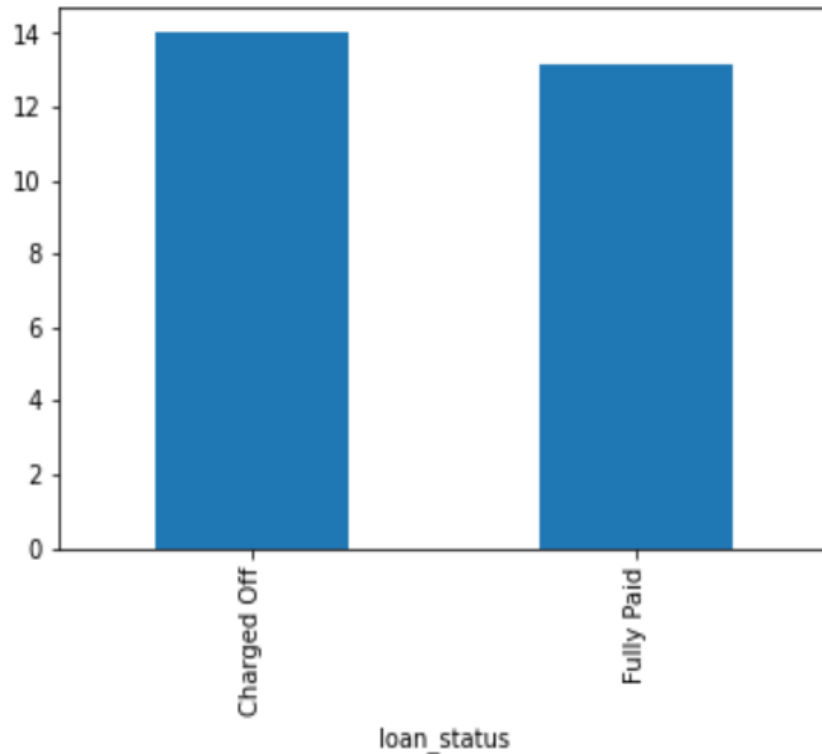
- Defaulted customers have a higher debt-to-income ratio than those who fully paid off.
- This is an IMPORTANT parameter to be considered during a loan application.

Employment Length
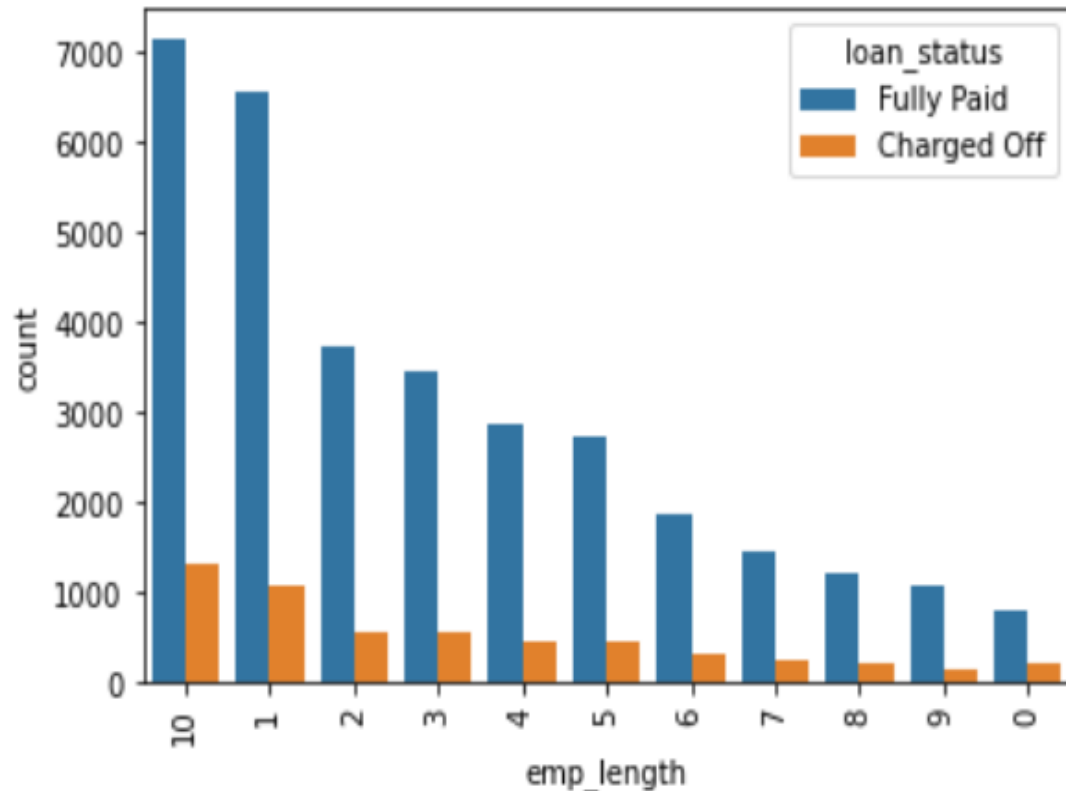
- The default rates for customers with employment history are comparable (approximately 12 - 15 %)
- Customers with no Employment history have a much higher rate of defaulting at 22%
- This is a IMPORTANT parameter to be considered.

Recommendations

Based on the analysis, we make following recommendations:

- Grades/Sub-grades : Loans applications with grades G & F should be scrutinized more carefully.

- Employment Length: Loans to customers with no employment history should be given considering their future prospects to repay.

- Purpose: Background and financials of Small Businesses should be verified more stringently before approval

- Debt-to-income ratio: A threshold should be established for debt-to-income ratio, above which loans should not be granted.

Conclusion

It is possible to reduce the number of defaulters in the future if steps are taken to re-analyse the predictive variables at the time of loan application.