Lending Club Case Study

Group Members:

Himanshu Dongre

Introduction

- Lending Club 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.
- Like most other lending companies, 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'.
- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.
- In other words, 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.

Methodology

Data Cleaning

- Remove null value columns
- Remove columns which are not significant to analysis
- Remove redundant columns
- Remove current loan records
- Standardize values
- Fix rows and columns to have appropriate datatype
- Create preliminary derived variables

Analysis

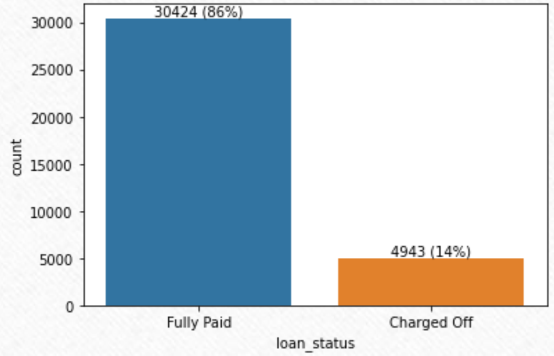
- Univariate Analysis: Analyze each variable separately
- Segmented Univariate Analysis: Segment the categorical variables and then conduct univariate analysis across its categories
- Bivariate Analysis: Analyze 2 variables against each other

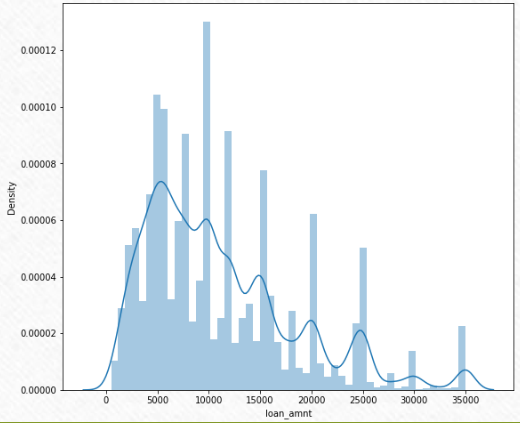
Recommendations

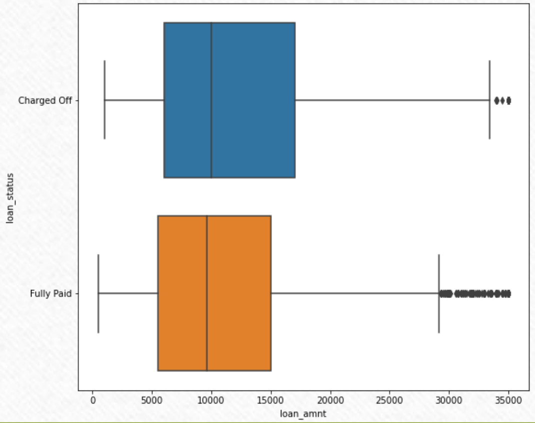
- Summarize Results
- Publish Recommendations

Analysis

- Around 86% loans are fully paid and the current default rate is around 14%.
- Most people seem to have taken a loan of around 10000 and very few people have taken more than 30000
- Charged off loans seems to have a higher average loan amount request indicating that the higher the loan amount higher the chances of default







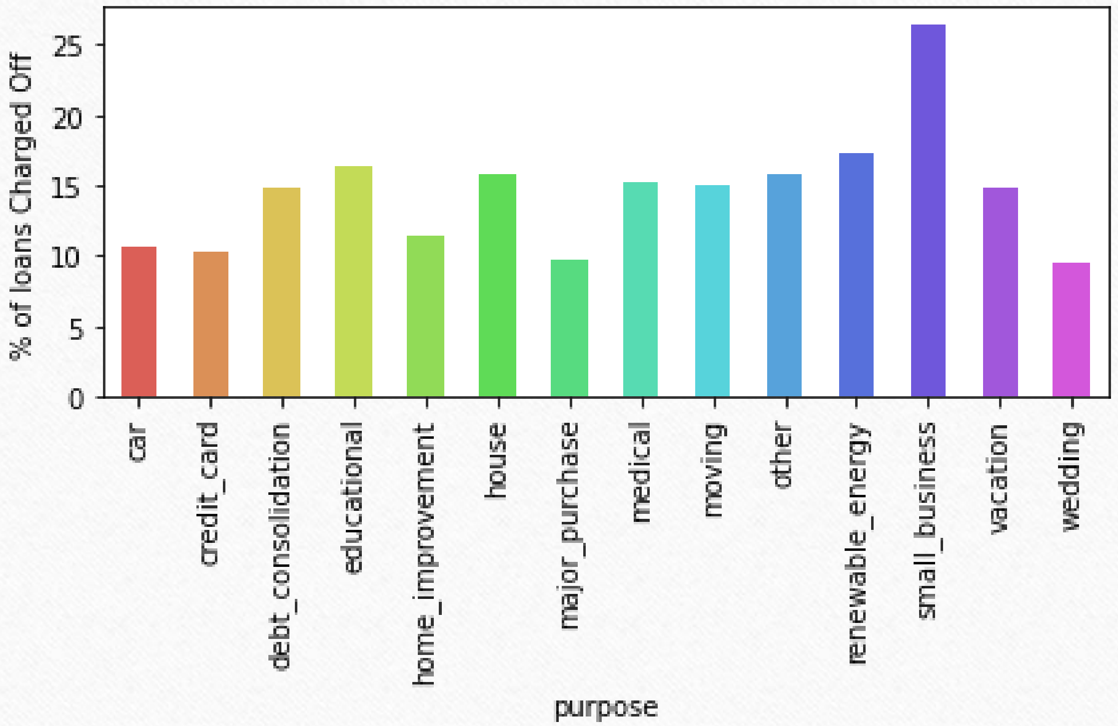
Analysis- Overall Insights

- 75% of loans are for 36 months
- Chances of getting a loan are more for high grade customers
- Most people applying for loan either live on rent or have existing mortgages
- Surprisingly, lot of non-verified people have received loans.
- 48% of customers' main purpose of buying a loan is debt consolidation.
- Most loans are from CA. This is expected as the company's is based out of CA.
- 25% loan applications have 10+ years of employment.
- Every year the number of loans approved has increased.
- Analyzing monthly issue data, company seems to approve more loans in the second half of the year

Analysis

- Loans taken for small business purpose are more prone to default

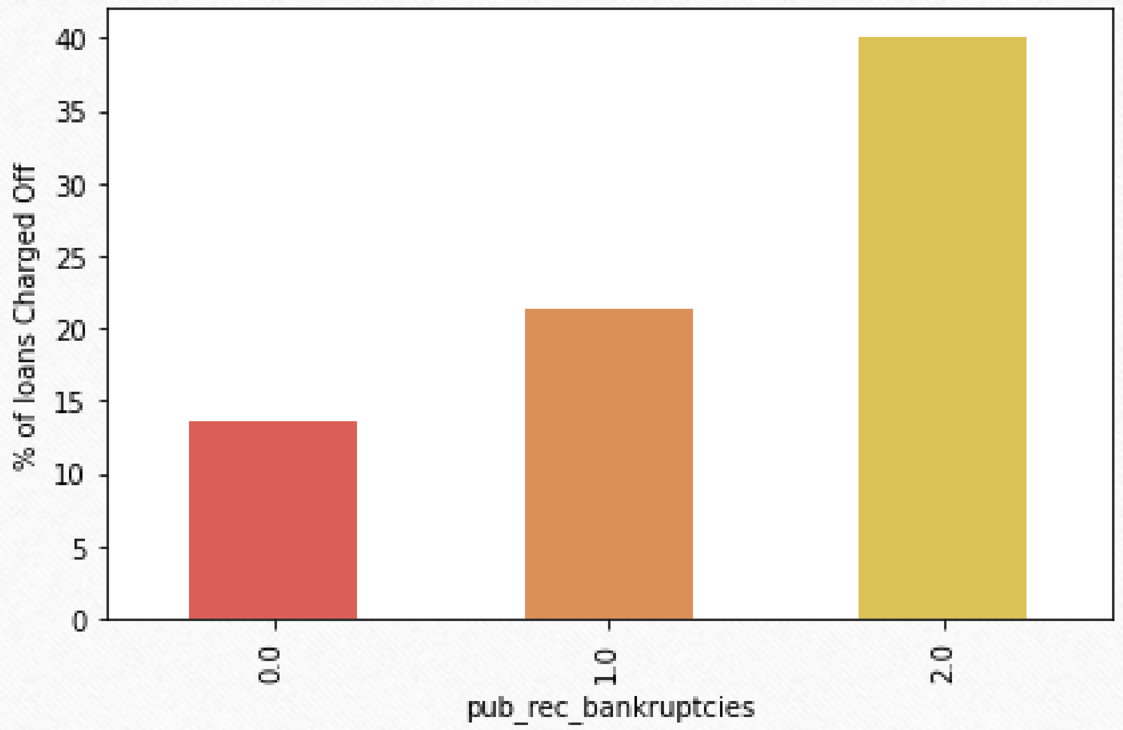
- Suggestion: Reduce number of loan approvals to small businesses in general.



Analysis

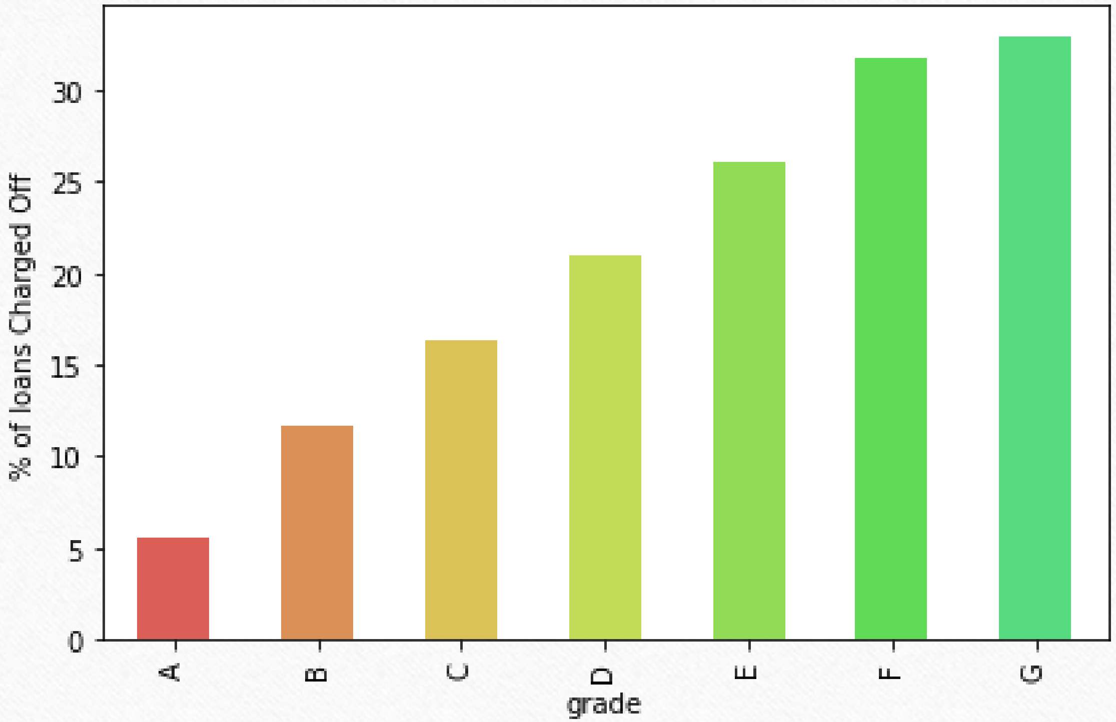
- Loans taken by applicants with public record of bankruptcies have a high default rate

- Suggestion: Stop approving loans to applicants with existing public record bankruptcies.



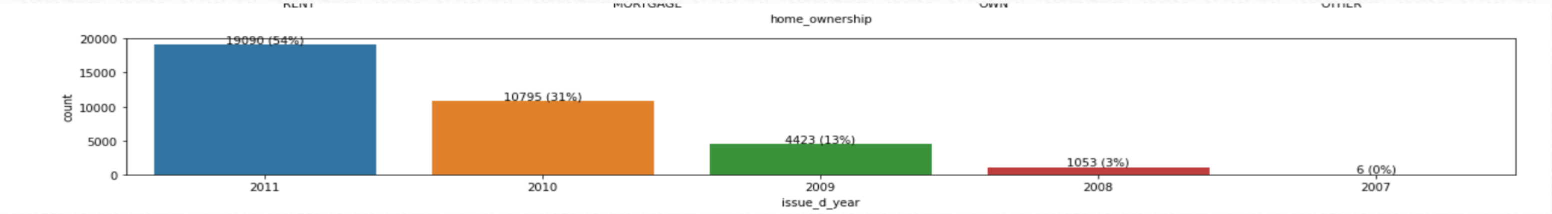
Analysis

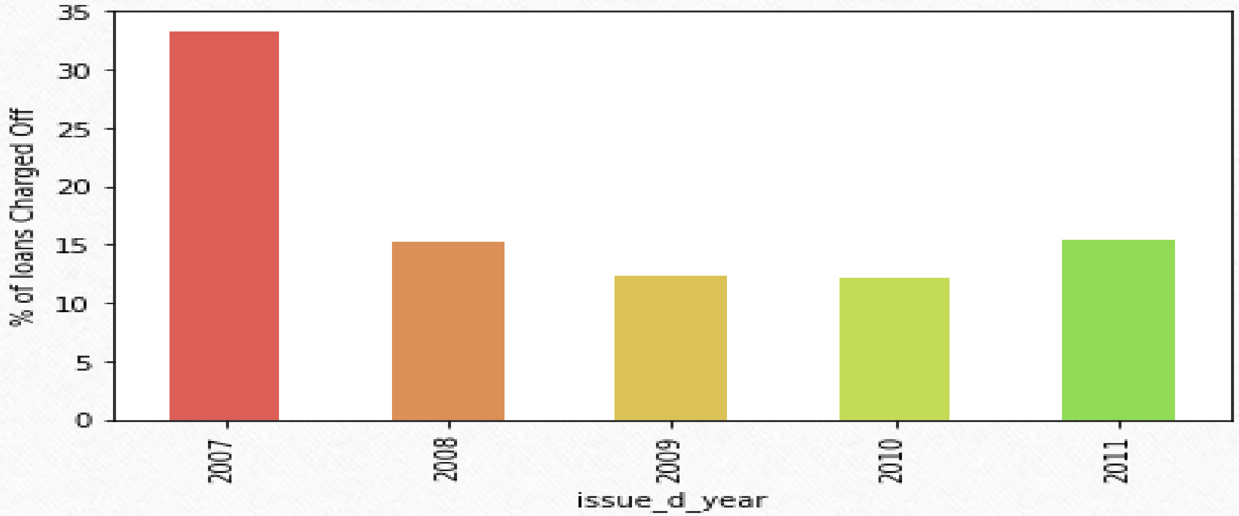
- As expected, % of defaults are higher for lower grade loans
- Grade policy of the company seems to be working correctly.
- Suggestion: Reduce approvals to applicants with grade D and below



Analysis

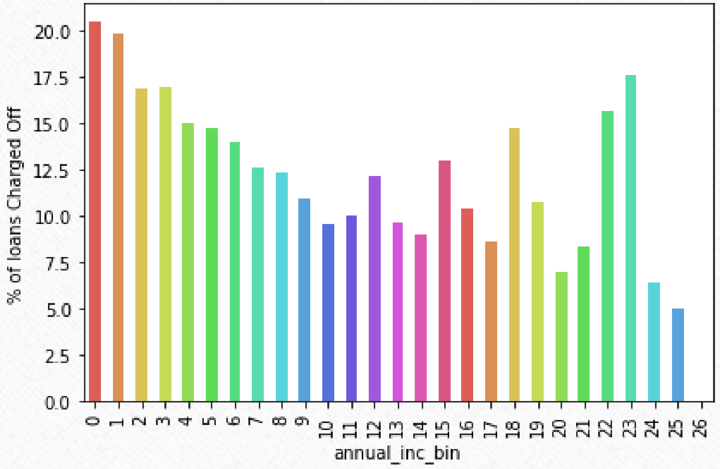
- While the number of loans approved have increased each year, the default rate has come down significantly. This indicates that loan company has significantly improved the review process for loan approval over the years.

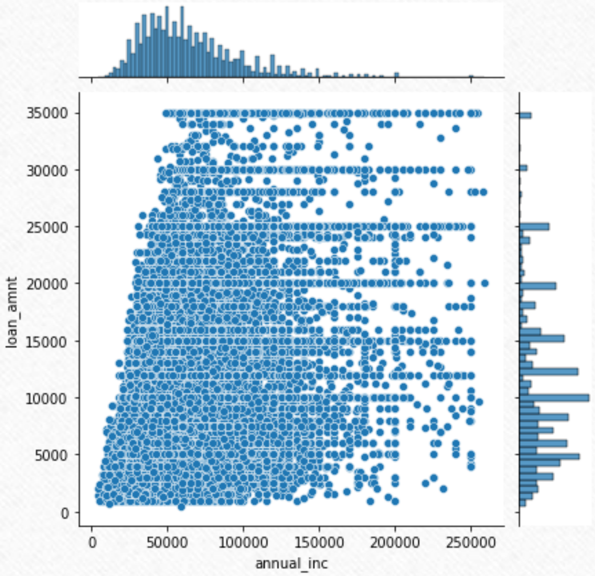




Analysis

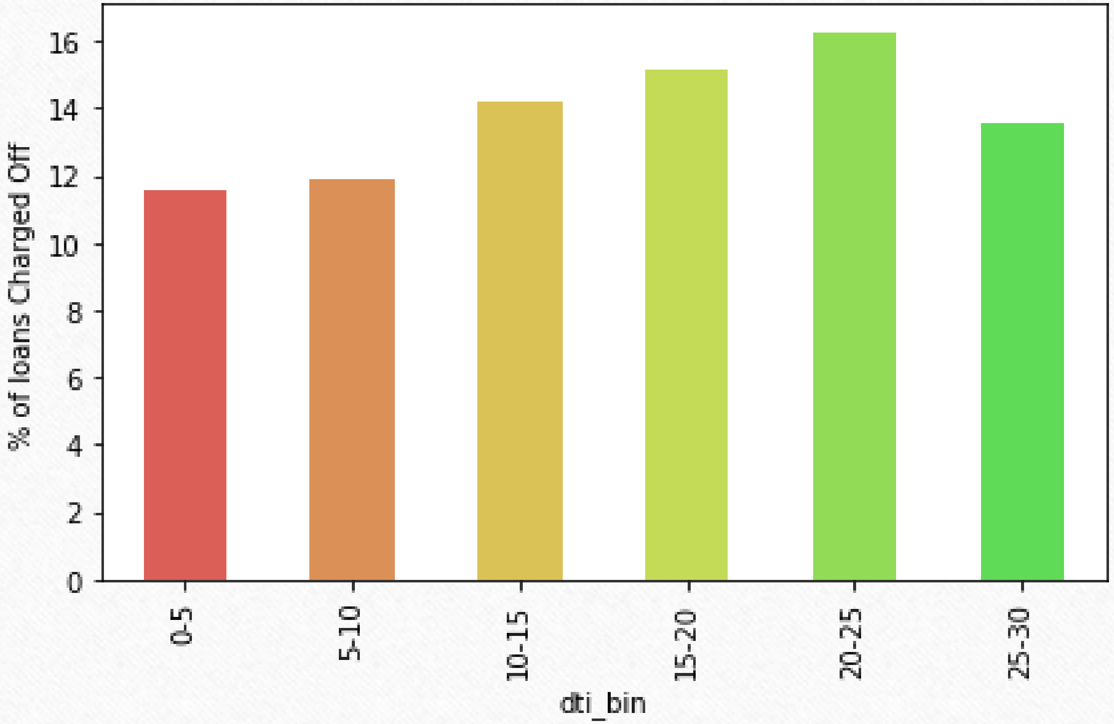
- Loan defaults are higher for applicants from lower income group.
- People with lower annual income also seems to be taking high value loans which has high risk of default
- Suggestion: Annual income in the lower 25% quartile have more chances of defaults. Consider reducing such loan approvals.





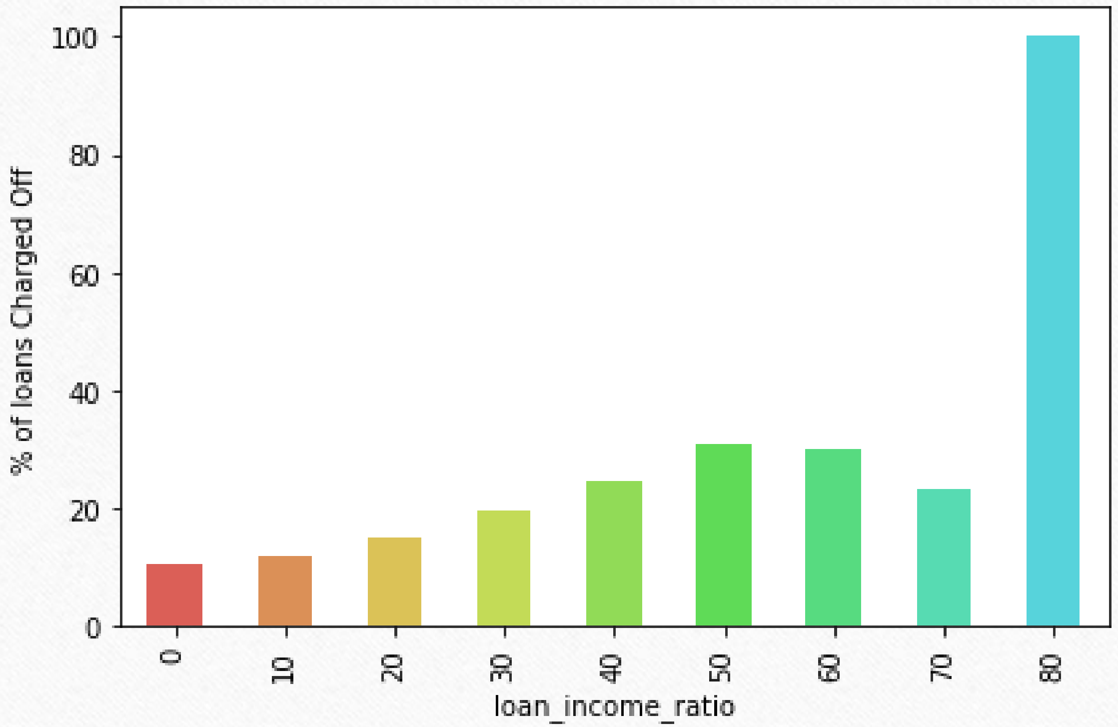
Analysis

- Applicants with debt to income ratio greater than 20 have higher chances of defaults
- Suggestion: Consider reducing such loan approvals.



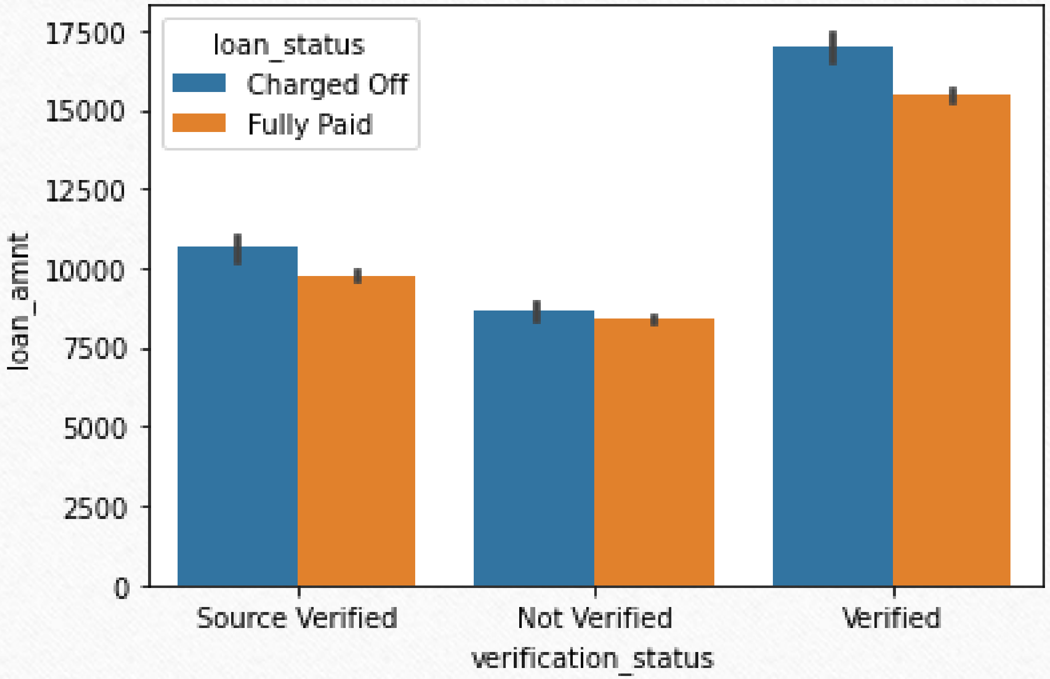
Analysis

- Loan to income ratio above 20 have high chances of default
- Suggestion: Stop approving loans where loan amount to income ratio is greater than 20



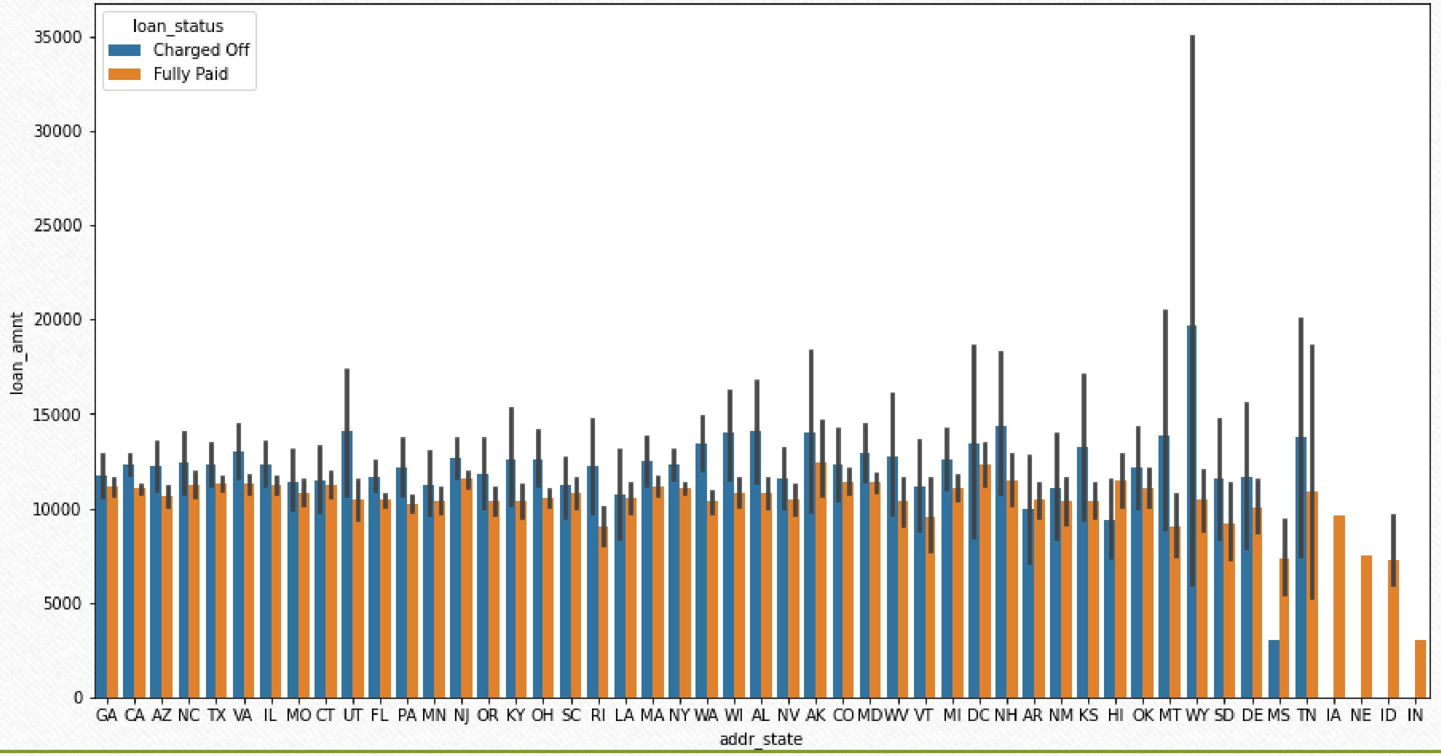
Analysis

- It was surprising to see more number of verified loans were charged off than non verified loans. However, it seems to be the case that since higher loan amounts are risky they are more likely to be verified. Therefore the high charged off rate.
- Tip: It looks like the defaults in case of non-verified loans can be avoided if the verification process is simplified and more loan applications are subjected to verification.



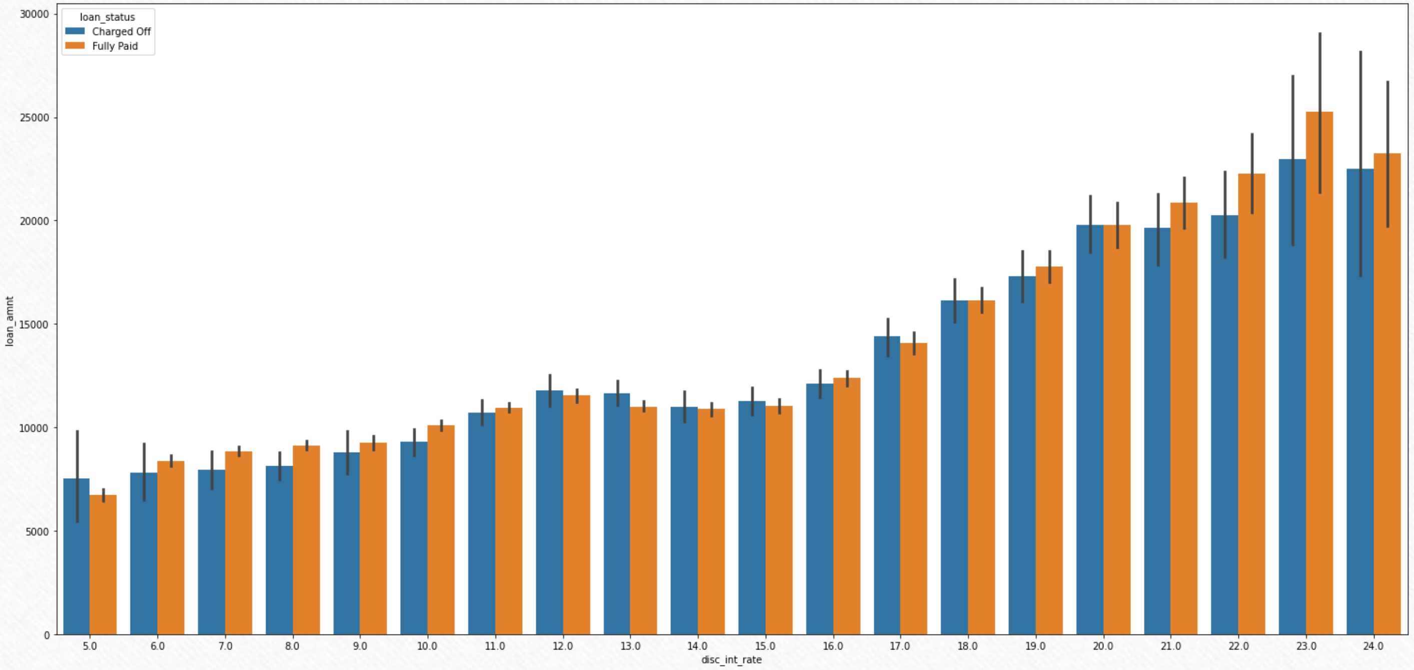
Analysis

- Applicants from certain states like WY seem to have high default rates
- If we stop accepting applications from such states, it may lead to a lot of loss of business. It would be better to make the approval process in such states more stringent.
- Suggestion: Reduce approvals to applicants with grade C and below for high risk states



Analysis

- Chances of default increases as the interest rate and loan amount increases
- Suggestion: Consider only approving loans where loan amount and interest rate are below median .



Summary

- Based on the above analysis below is the list of primary features which should be used to predict default for loan applications:

Features	Description	Effect on Default Rate
loan_amnt	The listed amount of the loan applied for by the borrower.	loan amount above median leads to more default
pub_rec_bankruptcies	Number of public record bankruptcies	if any public record bankrupcies are present then there is a strong chance of default
grade	LC assigned loan grade	grade C and below have higher chances of defaults
annual_inc	The self-reported annual income provided by the borrower during registration.	annual income in the lower 25% quartile have more chances of defaults
loan_income_ratio	Loan to income Ratio	loan income ratio above 20 have high chances of default
int_rate	Interest Rate on the loan	int rate above median have higher chances of default

Summary

Apart from these primary features below are few more features which have some co-relation with default:

1. People taking loan for **small business** purpose have a high tendency of defaulting.
2. Applicants from **certain states** like **WY** seem to have high default rates.
3. Applicants with **debt to income greater than 20** have higher chances of defaults.

Recommendations

1. Stop approving loans where loan amount to income ratio is greater than 20.
2. Reduce number of loan approvals to small businesses in general.
3. Reduce approvals to applicants with grade C and below for high risk states like WY and grade D and below for everyone else.
4. Stop approving loans to applicants with existing public record bankruptcies.
5. Start charging more interest rate to applicants with dti greater than 20.