RISK ANALYSIS OF LENDING LOANS

By
Parvathy Jayaram
Venu Kulala

Problem Statement

- Identify the patterns in the given data to identify the risk of a person being defaulted
- Identify patterns to deny the loan based for defaulters, reduce the loan amount,
- Identify patterns to Approve the loan for the risky profiles with high interest rates or reduce the loan amount
- Patterns to approve the loan for non-defaulters and reduce the risk of losing the business

Objective

- Objective is to **identify the risky loan applicants at the time of loan application** so that such loans can be reduced thereby cutting down the amount of credit loss. Identification of such risky profiles using **Exploratory Data Analysis** is the aim of this case study.
- In other words, **to understand the driving factors (or driver variables) behind loan default, i.e., the variables which are strong indicators of default.** The company can utilise this knowledge for its portfolio and risk assessment. And thus minimise the risk of losing money while lending to customers.

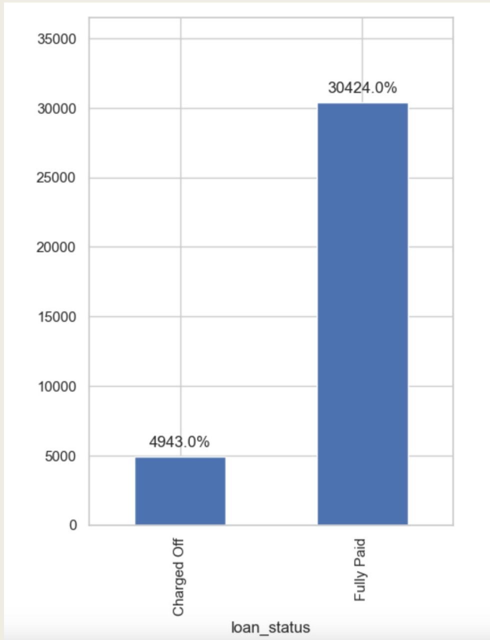
Data Cleaning

- There are columns which has **100% missing values.** Selecting those columns with more than 50% because they add noise into the data in large propotion.
- There were **57 columns** with more than 50% values as missing. Removing such columns as these columns clearly add noise and wont help in analysis.
- We see that there are 50 records with the earliest credit line date greater than 2011(where 2011 is the max range of the data we should have).
Lets check and fix the issue

Analysis – Loan Status

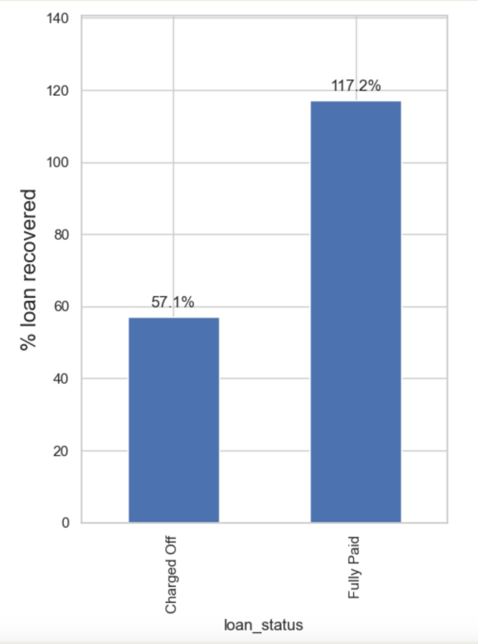
- Approximately **14%** of loans in the dataset are **defaulted.**

- Any variable that increases percentage of default to higher than 16.5% should be considered a business risk. (16.5 is 18% higher than 13.97 - a large enough increase)



Analysis – Loan recovery rate

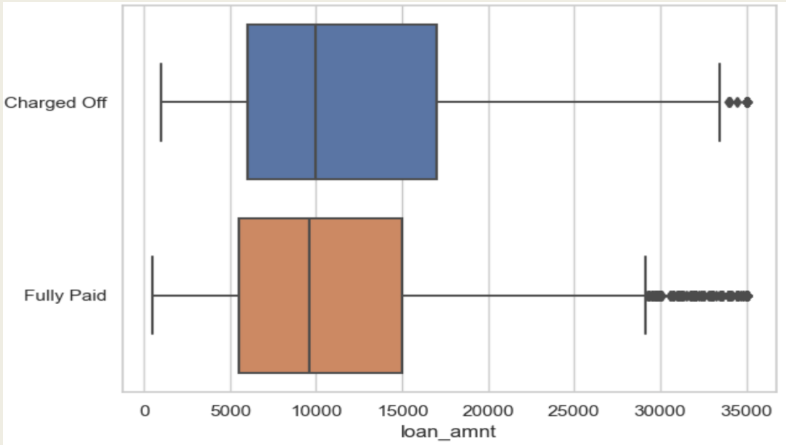
- Lending Club only recovers 57% of the loan amount when loans are defaulted.
- On fully paid up loans, the company makes 17% profit.

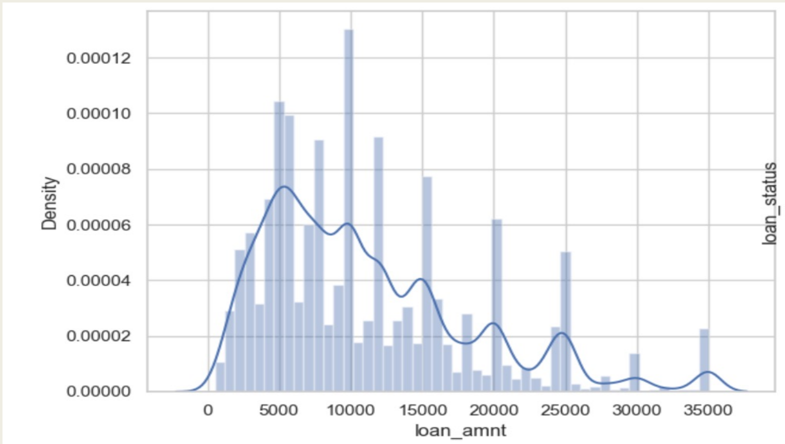


Analysis – Loan Amount

- Overall, the applied loan amount distribution is slightly **right-skewed** with mean greater than the median.

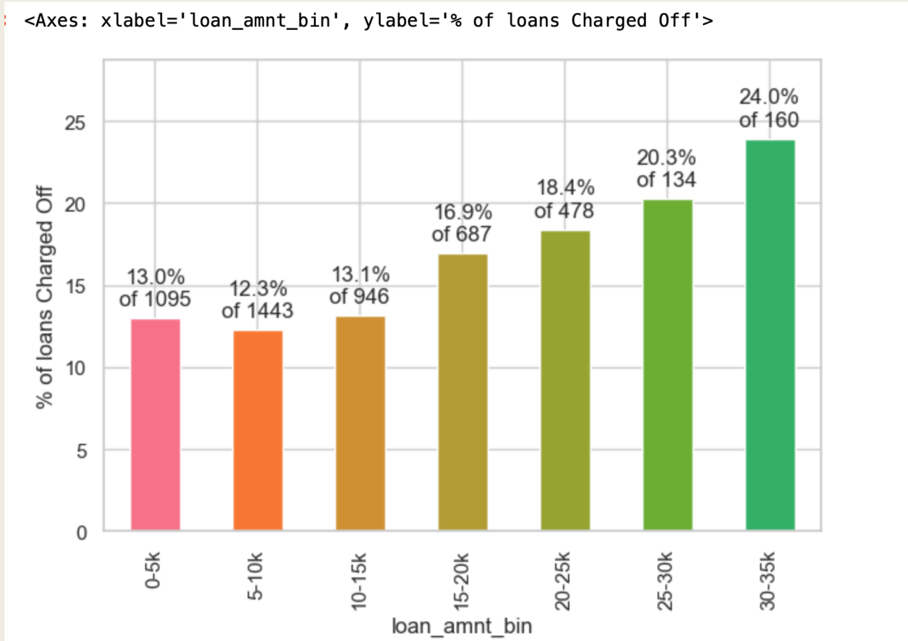
- Most of the loans granted are below 15000 (75 percentile value)
- Funding amounts see a spike around each 5000 boundary. We will use 5000 as bucket size for later analysis.
- Charged off loans are shifted towards higher average loan amount request.



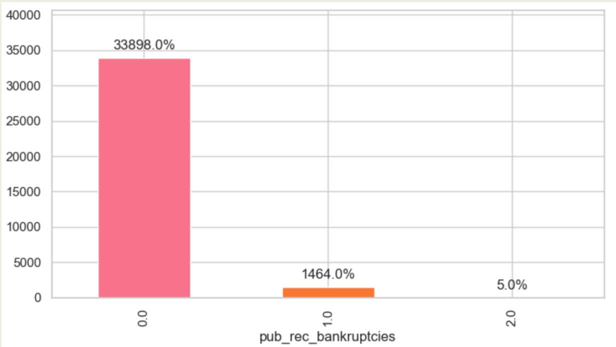


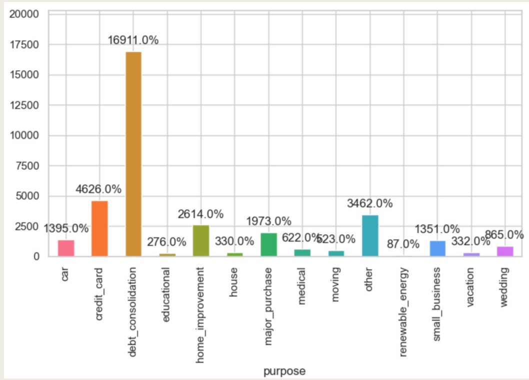
Analysis – Derived Categorical Loan Amount Variable

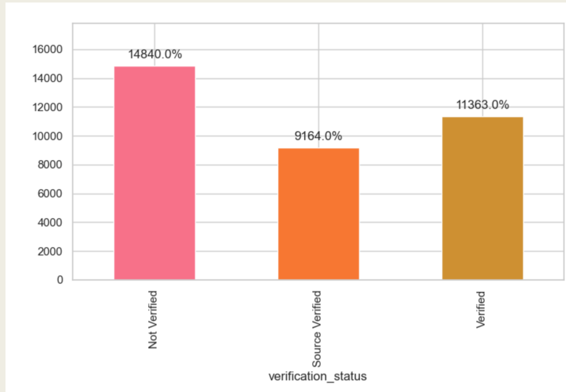
- The % of charged off loans increases substantially as we go up the loan amount buckets.
- Most loans are below 20000 amount. The higher loans, though lesser in number, carry a substantially higher risk of default.



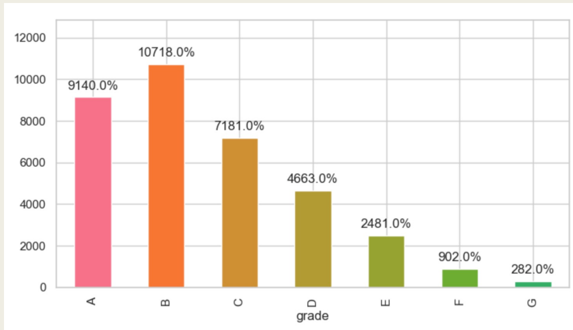
Other Categorical Variables

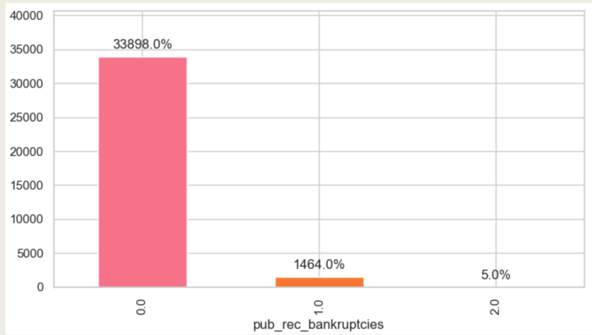






- Most loans are for 36 month term
- Very few loans are extended to people with prior record of bankruptcy
- debt_consolidation is by far the largest 'purpose' for seeking a loan
- Most approved loans are high grade

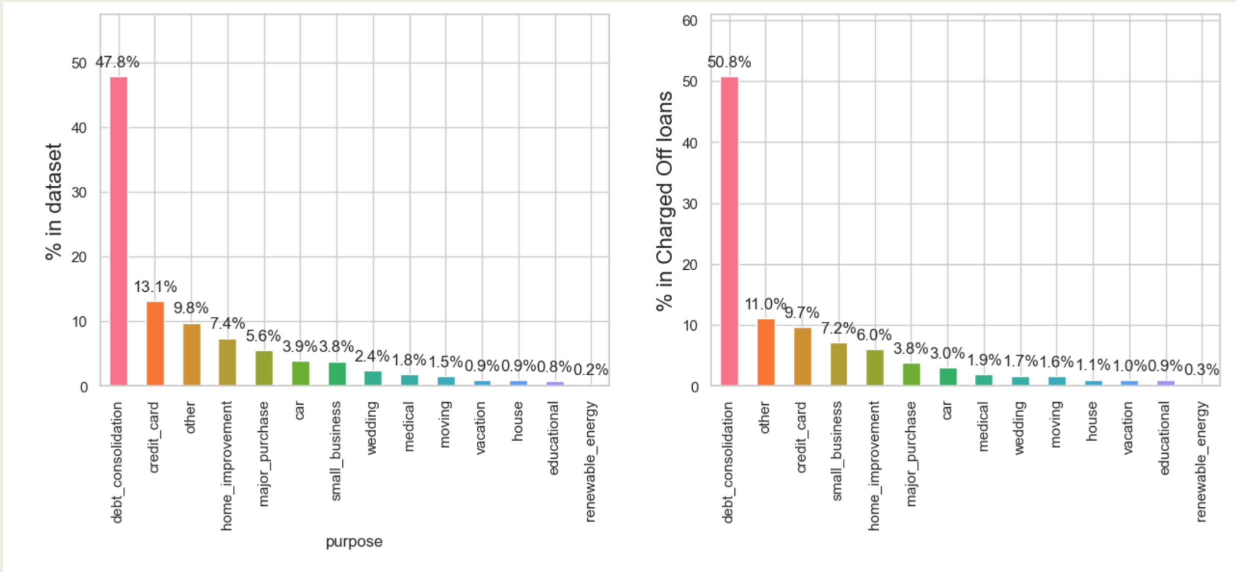




We should investigate **frequencies of defaults** by term, purpose and prior record of bankruptcy

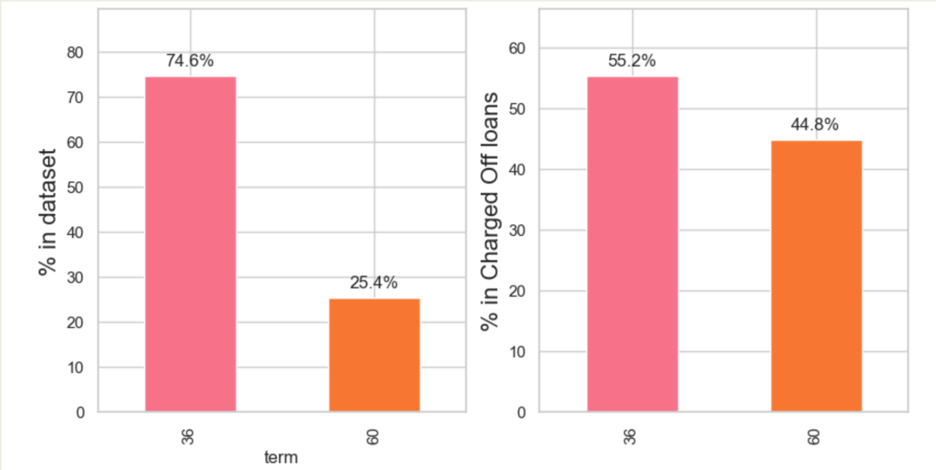
Analysis –Loan Status

The category 'small_business' percentage **doubles from 3.8 to 7.2** for Charged Off loans. Let's see how the categories of this variable behave.



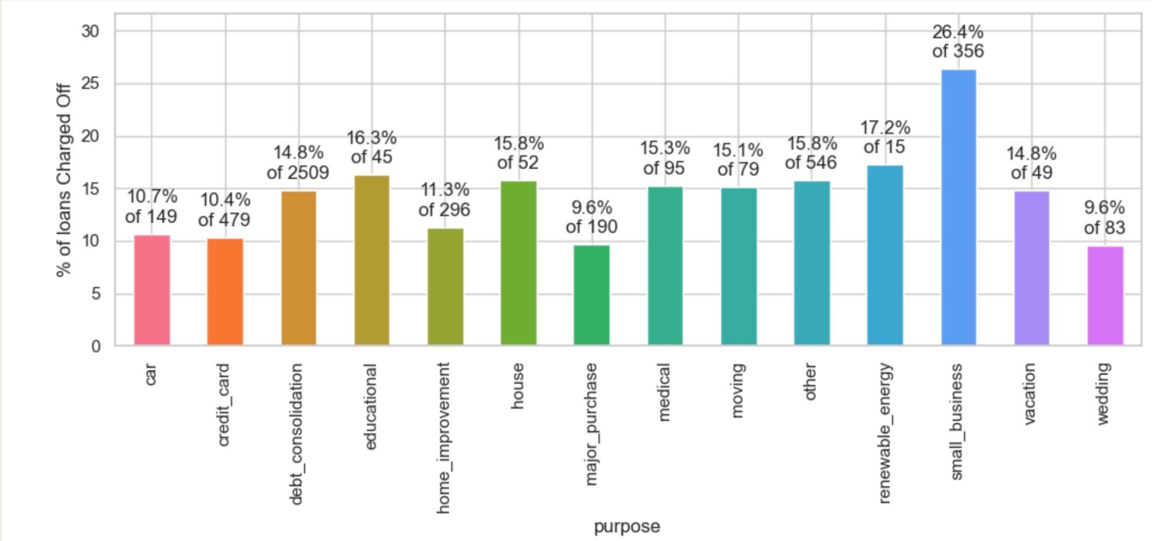
Analysis – Loan Status vs Term

- Around 75% of the total loans are given for duration of 3 years. while just 25% of the loans are those given for 5 years.
- Among Charged Off loans, percentage of term 60 months rises to 45%. The higher term loans have a higher chance of default.



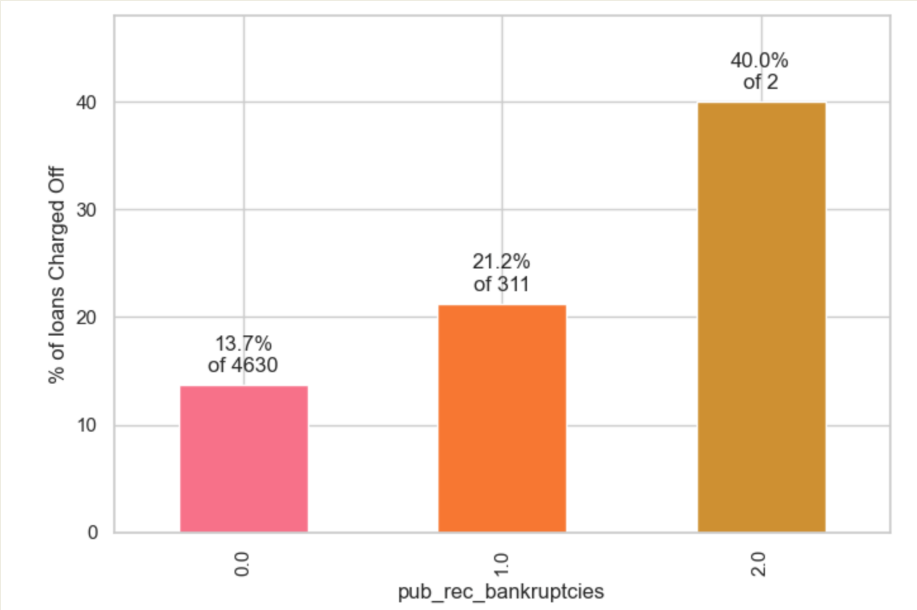
Analysis - Loan Status Vs Purpose

- 26% of loans for small business are Charged Off. Making them the most risky purpose.
- Approximately ~14.9% of the loans are issued for the purpose of dept consolidation.
- 17% of the loans for renewable_enrgy are charged Off, but the number is too less to be of significance.



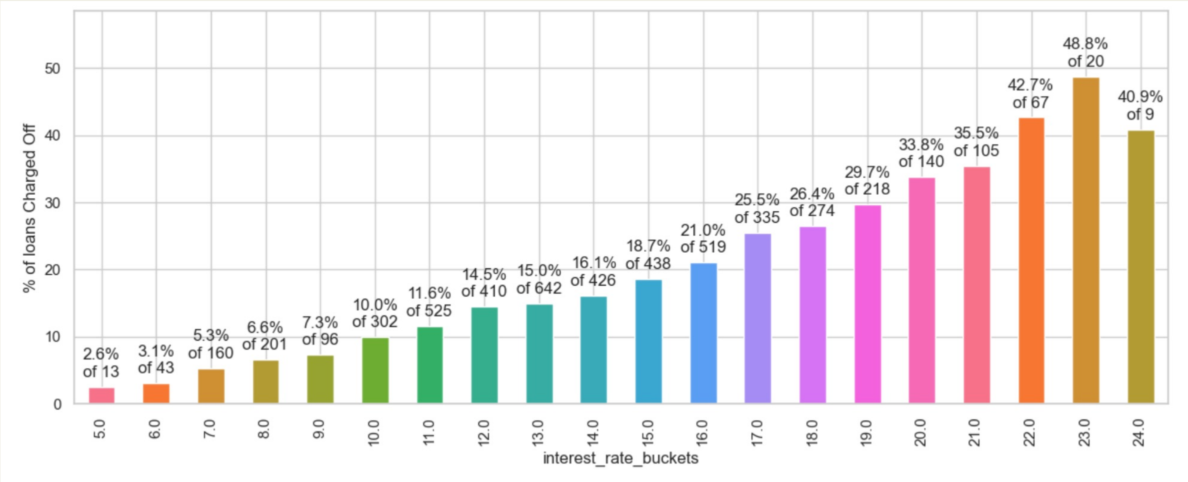
Analysis – Default vs History of Bankruptcy

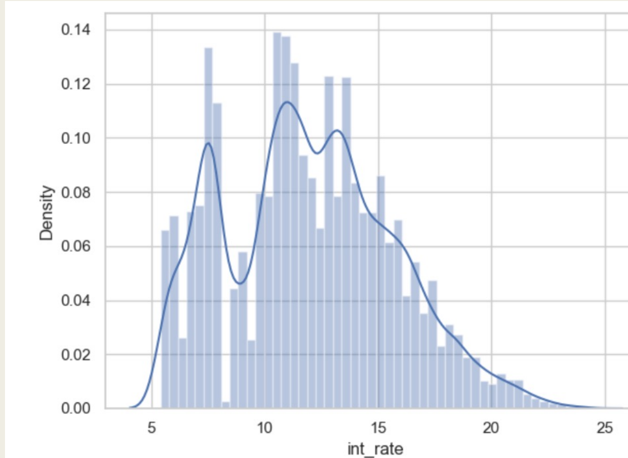
The percentage of Charged Off loans is markedly higher when the borrower has a prior record of bankruptcy

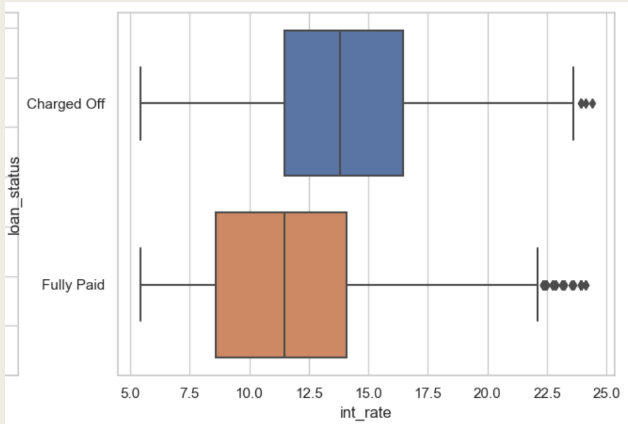


Analysis – Loan Status vs Interest rate

- Overall, the intrest rate varies from 5.42% to 24.4% with average intrest rate of 11.8%.
- The interest rate for Charged Off loans appear to be higher than for Fully paid. This is naturally expected. As, the risk increases the rate of interest imposed on the loan also increases.





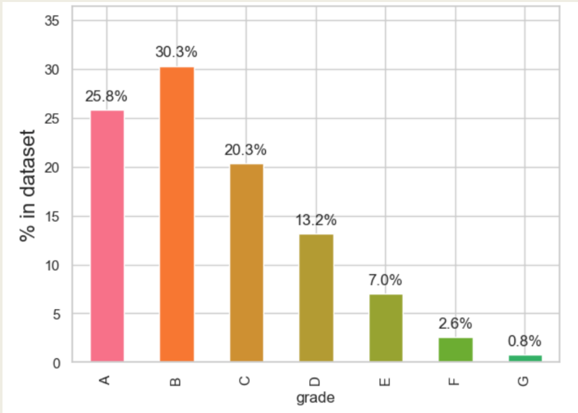


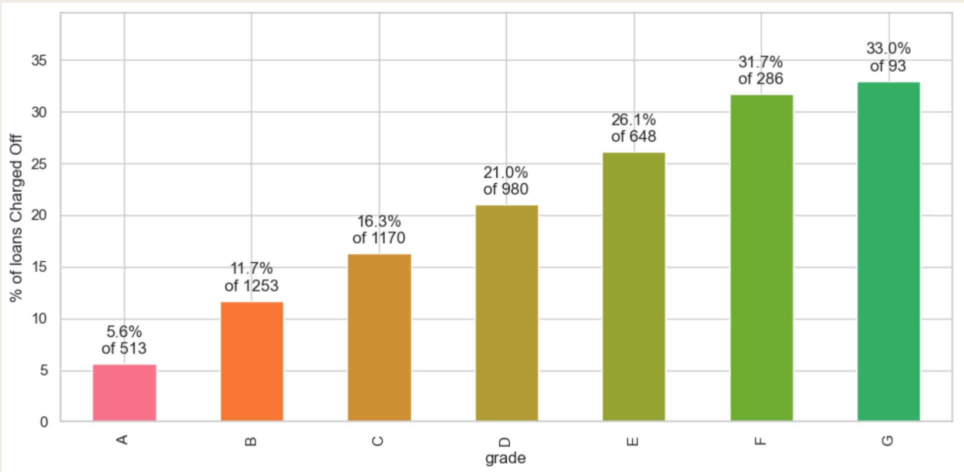
Analysis – Loan grade

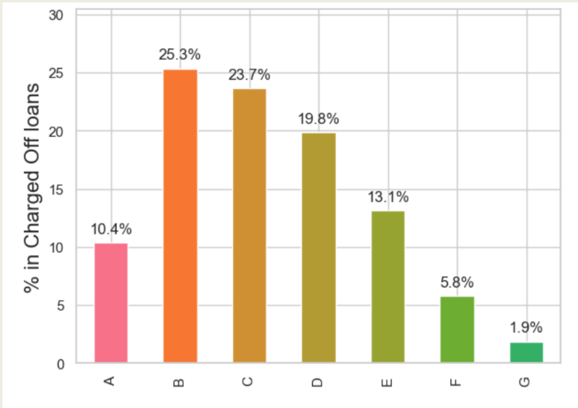
- Grade A and B loans are safe. The percentages in full dataset are much higher than percentages in Charged Off loans.

- Grade D, E, F, G loans are less safe. We should plot grade by percentage Charged Off by category

- Lending Clubs grading system is working well.
- Nearly 30% of all loans in Grades F and G see a default.
- Grade E onwards are risky, and less numerous. Lending Club should either refuse loans or charge high interest rates.

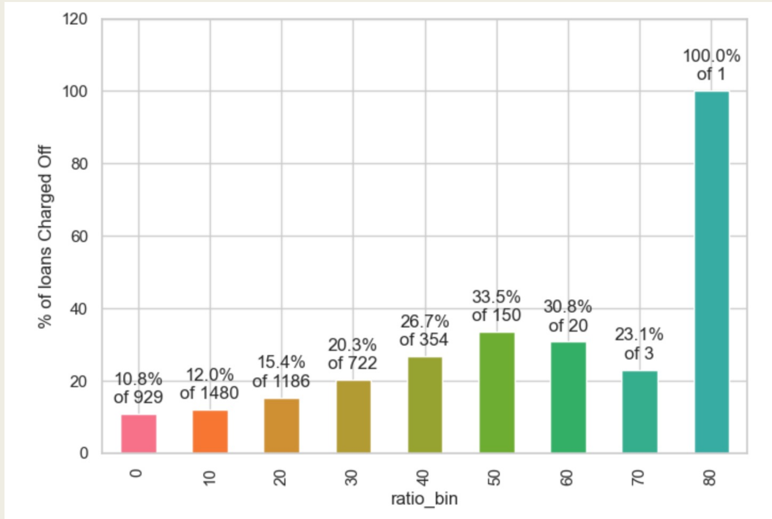




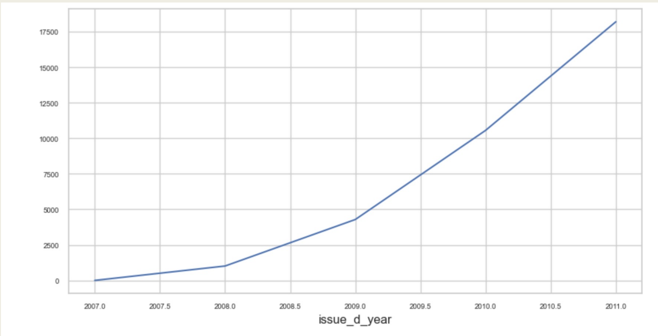


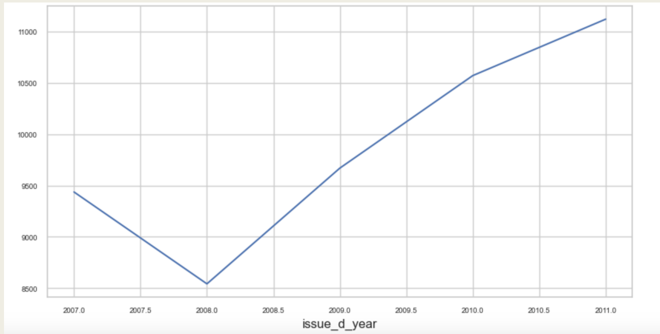
Analysis – Loan Amount vs Income

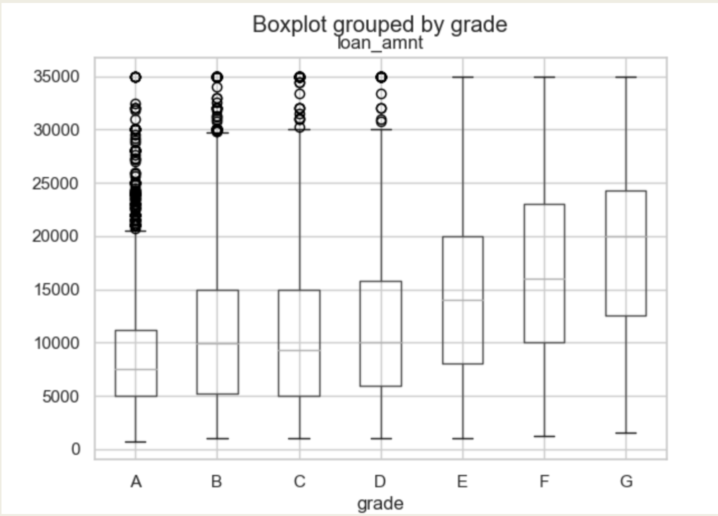
- As long as loan amount is less than 20% of annual income, defaults are low.
- Loan amounts of 30% of annual income or higher see a high rate of default.

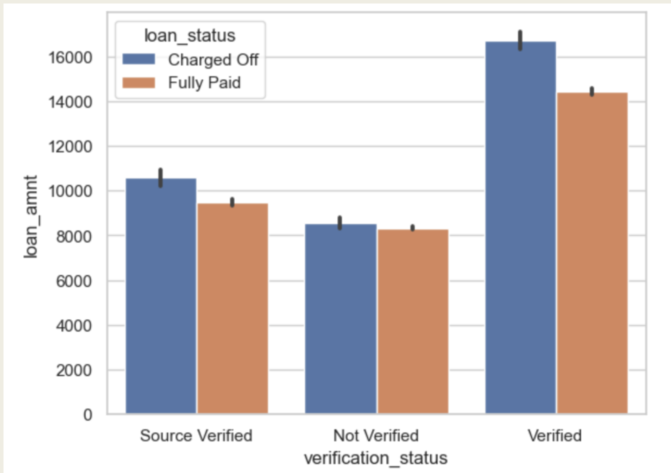


Bivariate Analysis









- Larger loans generally appear to be given a lower grade, with the median loan amount for a grade G loan being almost 10000 higher than that of a grade A, B, or C loan

- Higher loan amounts are Verified more often.
- We already know that larger loans are less in number, but see a higher charge off rate.

- This, combined with previous observation, explains why verfied loans see a higher rate of default. It's not the verified status per se, it's the fact that higher loan amounts are riskier and are also verified more often by Lending Club.

Correlation

- loan_amnt is correlated to last_payment_amount with r factor 0.44, as expected
- int_rate is correlated to revol_util with r factor of .47 - This is good, as company is charging higher interest from riskier loan.
- loan_amnt revol_bal are correlated with r factor .35 - This is not good as it suggests that higher loan amount is being approved to riskier borrowers.
- delinq_2yrs is totally un-correlated with public record of bankruptcy. Therefore they represent distinct features with individual predictive value.