Gramener
Case Study

Problem Statement

Objective

Abstract

Gramener wants to **improve its portfolio & risk assessment for new loan** applications - Two types of risks associated with its 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

Case Study Objective

- Analyse the provided loan data set and identify patterns in the loan applicant's attributes which indicate if a person is likely to default on the loan payments
- Identify the most important driving factors (variables) which strongly indicate a tendency to default

1.

Data Cleaning

Data

Source – Loan.csv

contains the complete loan data for all loans issued through the time period 2007 to 2011)

There are 111 columns in which 50 columns have valid data.

Data Cleaning

Ensure no duplicate data exists

Remove columns where for ALL rows the value is NA, 1, 0, “”

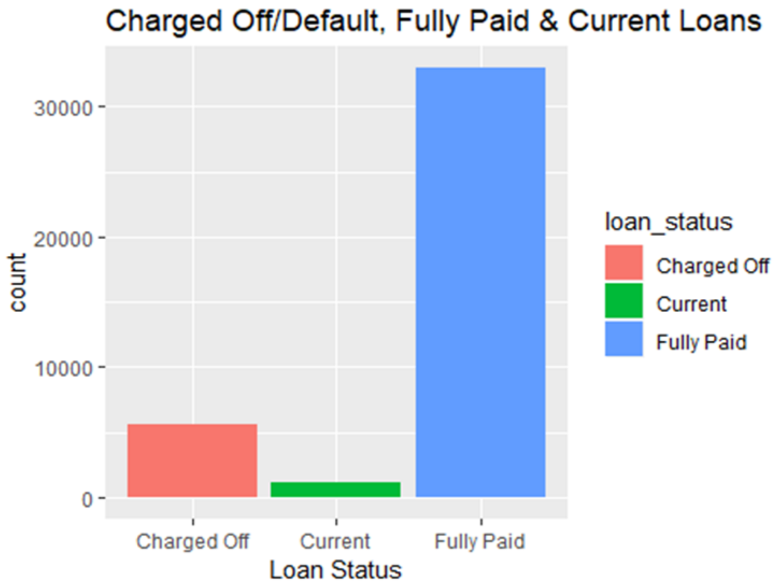
Convert numeric fields required for analysis like rate and term to numeric values

Change Date and time to proper format, required for analysis

Convert to factors.

2.

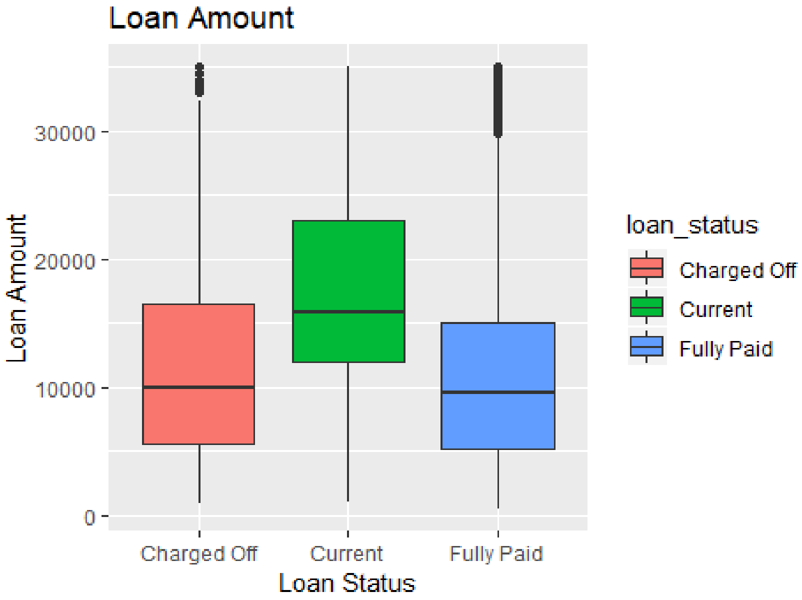
Overall Data Understanding



Overall Default Rate is 15%

10 % of principle amount is lost due to bad loans.

45788065 USD principal amount is lost because of bad loans.



Default/ charged off loan amount median is around 10000 USD

For Default there is considerable loan above 30K.

Important Observations – Driving Factors (Plots follow)

Grade & Subgrade

Loan Grades **“E”, “F” and “G”** show higher percentage of loan defaults

Loan Subgrade **“F5”** shows highest default percentage

Interest Rate

Higher the interest rate, more the probability of defaulting a loan

Purpose

“Small Business” is the purpose of loan for most defaulting borrowers

State

State **“NE”** – **Nebraska** shows highest defaulters on loan payment, but this is small loan amounts

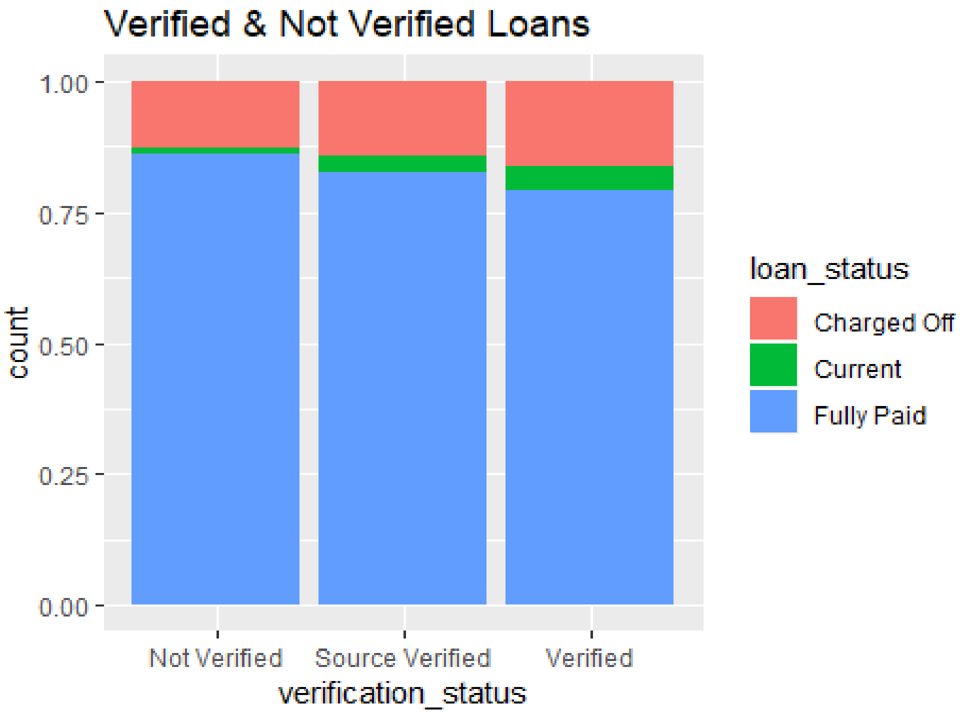
“AK” – Alaska, “UT” – Utah have high loan amounts & have considerable loan defaulters

Term

“60” months term has almost twice as much defaulters as the “36” month term

3.

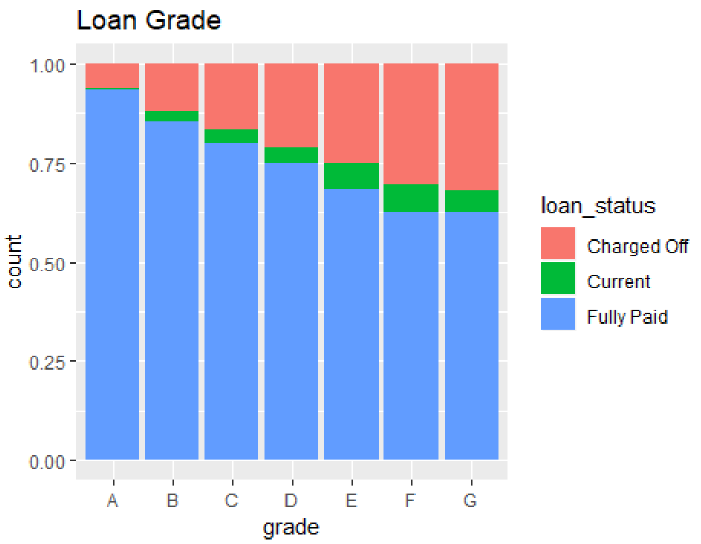
Important Observations

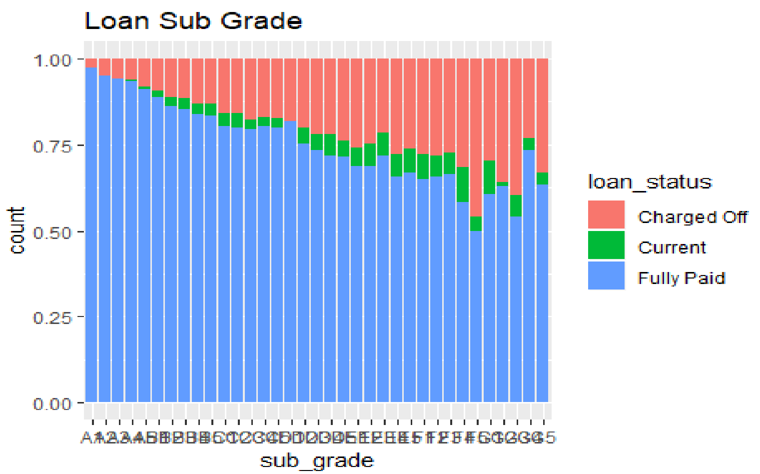


Its is noted, that the loan defaults in case of Verified and Non Verified incomes of applicants is similar

In fact Non Verified incomes show lesser defaulters compared to LC verified incomes

This indicates current verification process is not effective enough and may require improvement





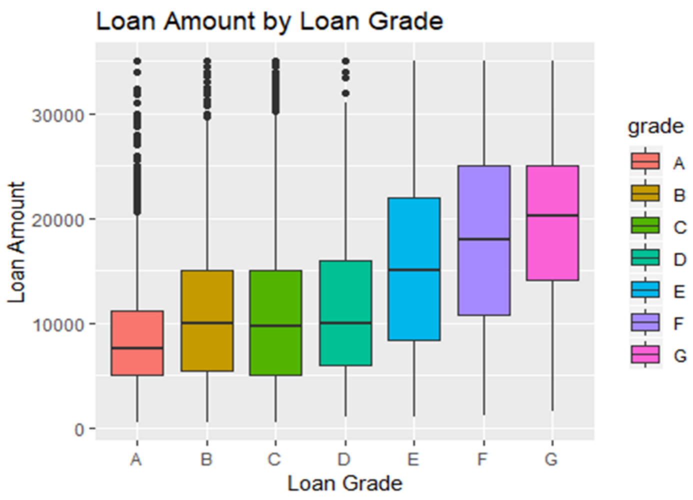
As the loan grade moves from “A” to “G”, the percentage of loan default/charge off increases.

“F” and “G” grades show highest loan defaults

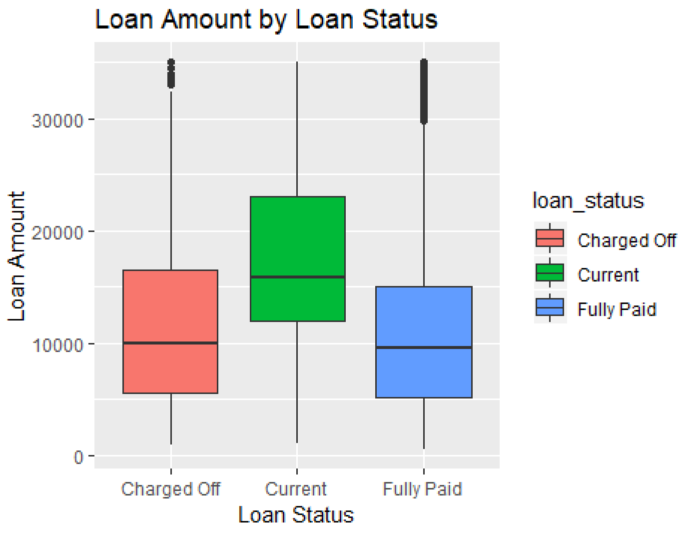
Loan Subgrade “F5” shows highest default percentage

4.

Analysis



We observe as grade moves from “A” to “G”, the median loan amount increases.



We observe for default/charged off loans, the median loan amount is more than fully paid

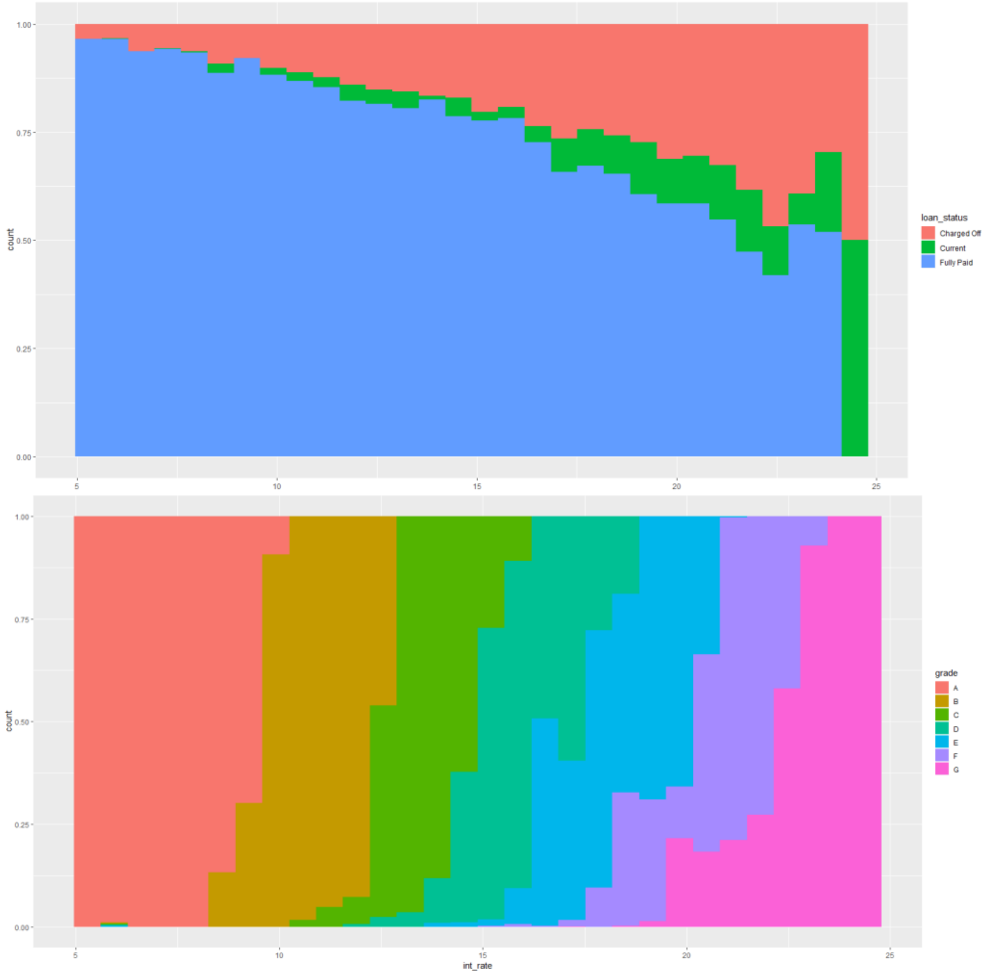
Also major of loan is in higher quartile.

Loan Interest Rate & Bad Loans

Higher the interest rate, more the probability of defaulting a loan

Also loan grade decreases when interest rate increases

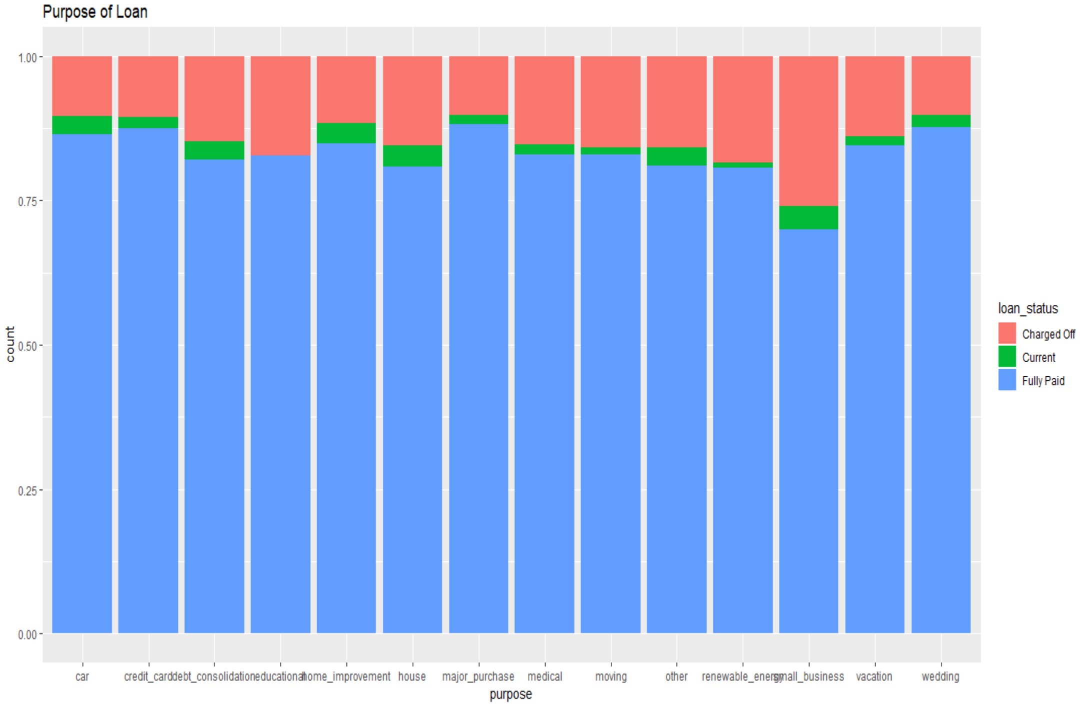
So higher interest rate could be one of reason for defaulting on Loan payments



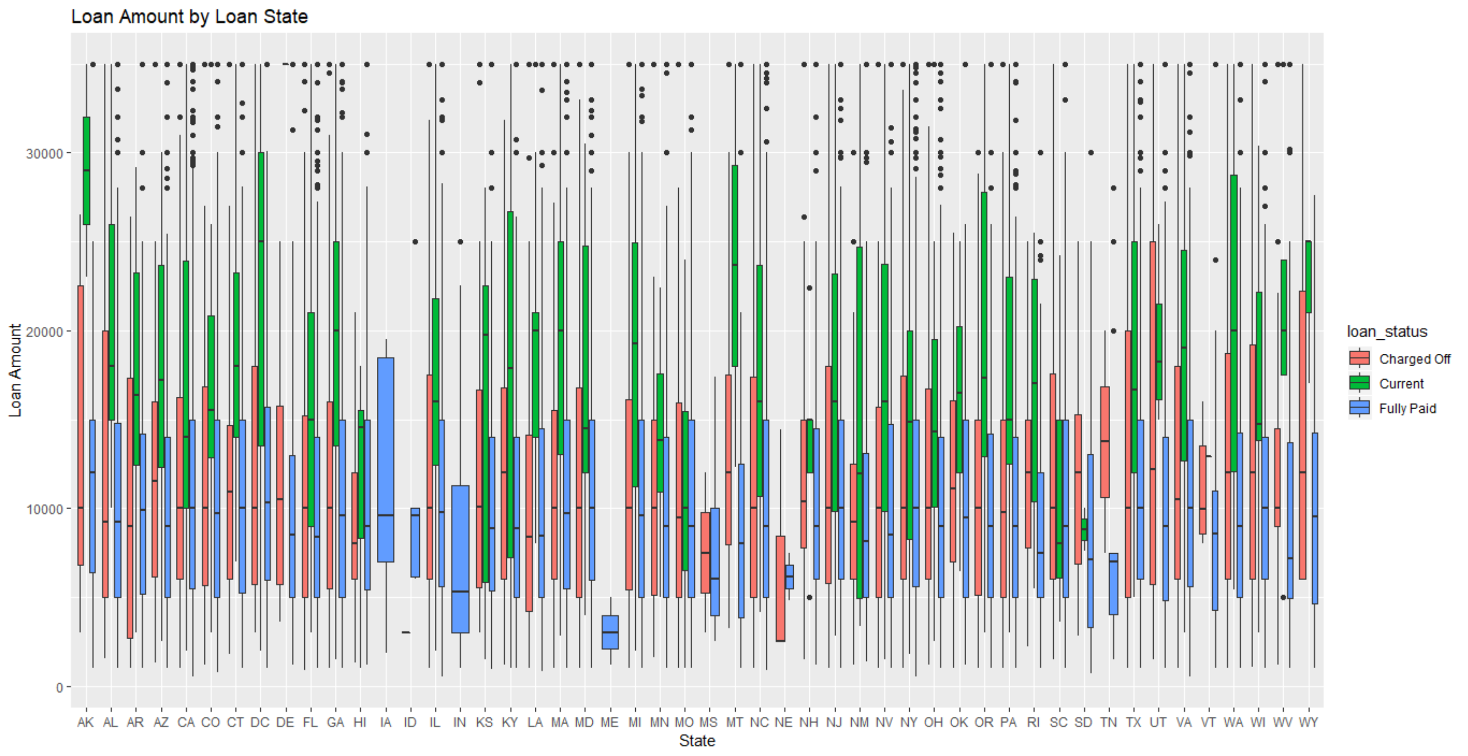
Loan Purpose Observation

“Small Business” is the purpose of loan for most defaulting borrowers

Loan purpose of “debt consolidation” and “housing” also have high chance of turning into bad loans.



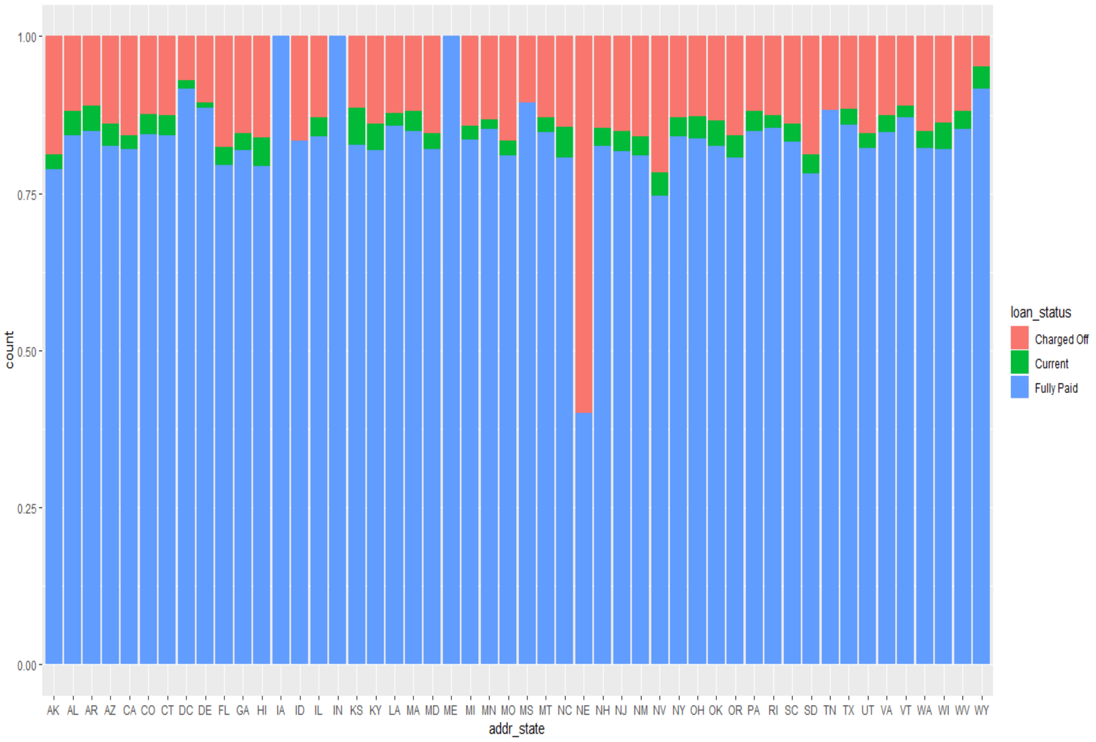
Loan Applicants State and loan amount

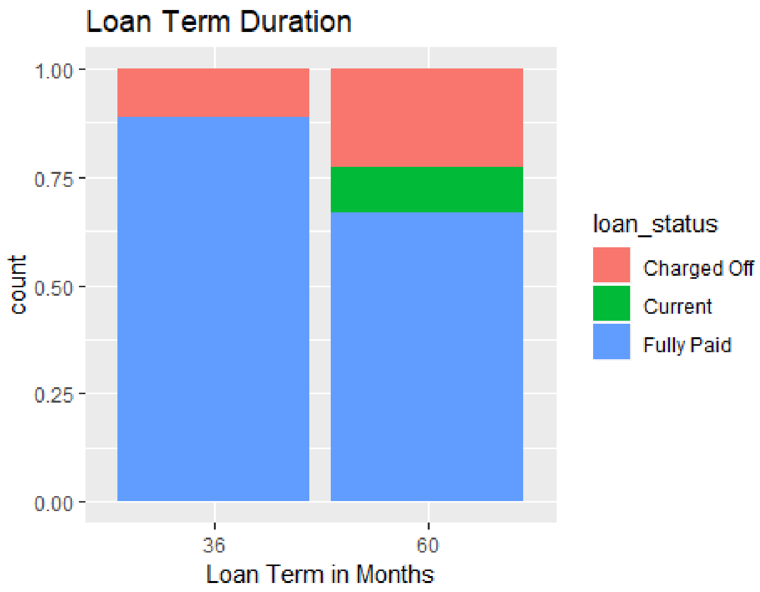


State of Loan Applicants Observation

State “NE” – Nebraska shows highest defaulters on loan payment, but this is small loan amounts

“AK” – Alaska, “UT” – Utah have high loan amounts & have considerable loan defaulters



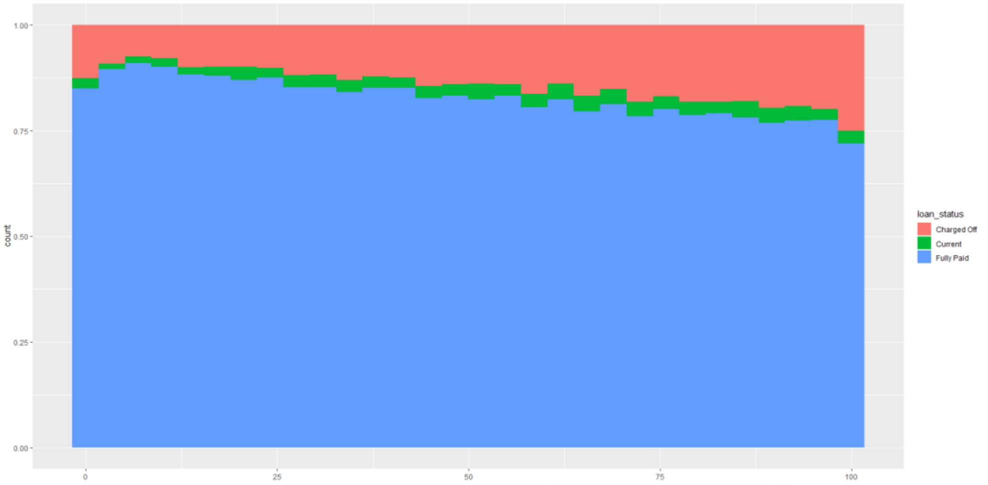


The higher the loan term duration, more the defaulters on loan payment

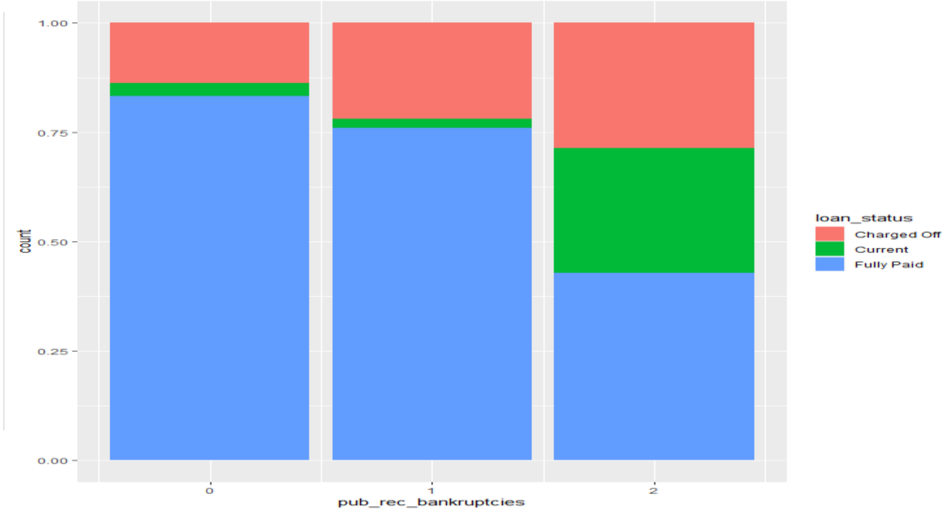
“60” months term has almost twice as much defaulters as the “36” month term

5.

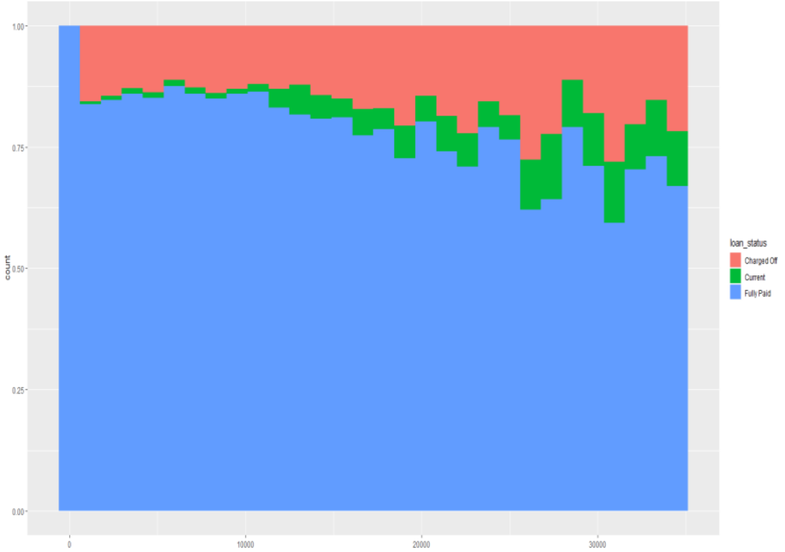
Data which doesn’t have much correlation

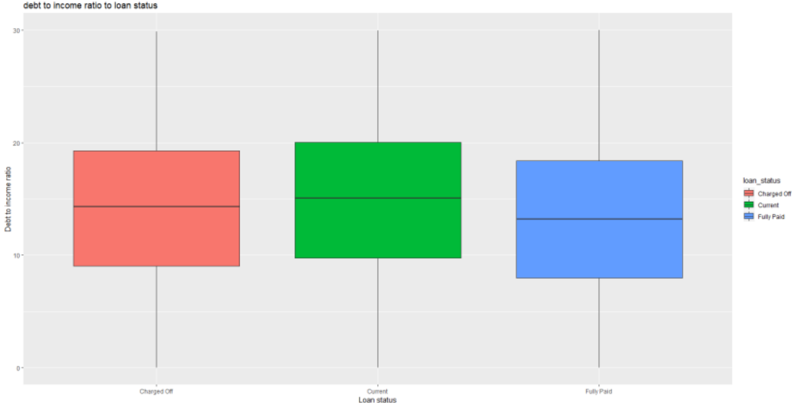


Revol_ util: Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit. We can observe that as this increases bad loan increases, but can’t decide much on this.



pub_ rec_ bankruptcies: Number of public record bankruptcies. We can observe that for value 2 default rate is more





Loan amount: We can observe that as loan amount increases, the probability of bad loan increases. But there is no set pattern.

Debt to income ratio: We can observe that there not much difference debt to income ratio for default and other type of loans.

6.

Recommendation

Major driving factor for loan defaults are,

Loan grade and sub grade. As grade decreases, loan default rate decreases.

Loan term: It is observed that loan with 60 months have higher default rate.

Loan purpose: It is observed that small business and debt reconstruction has higher default rate.

Applicant state: It is observed that state from which applicant like Alaska, Nebraska have higher default rate.

Interest rate: If interest rate is more than 15 % then default rate is more.

Factor's which doesn't have much correlation are,

annual_inc, verification_status, emp_length, application_type, installment amount.

High value loan: Even loan which is more than 25000 follows same pattern.