Lending Club analysis

By Sandeep Kumar

Project Overview

Lending Club:

- Lending Club Offer loan to customers. Interest is variable based of factors like term, amount, Grade etc.
- Reducing financial loss by analysing past data of defaulter vs normal customer for trends and factors affecting defaulters.

Data Source : Lendingclub.com

Business Constrains

- 1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- 2. 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

Business Objective

The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Data Cleaning

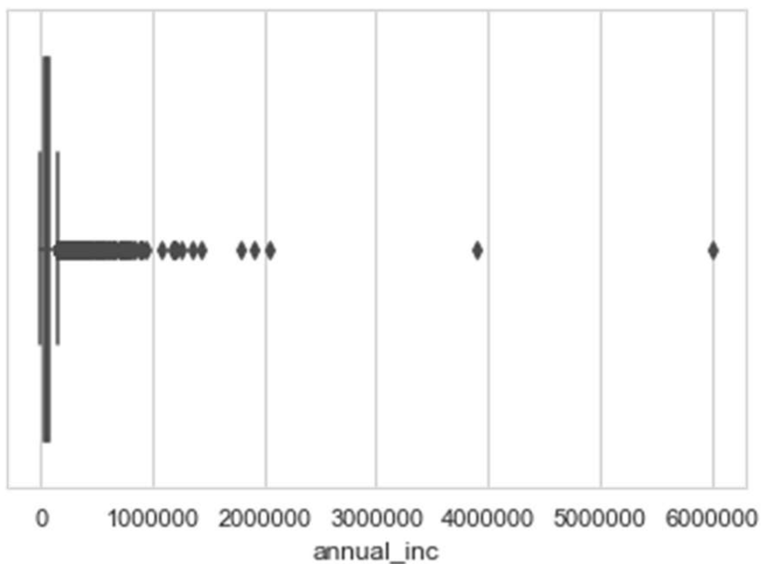
Data Cleaning and Preparation	
Deleting null or missing values or column with same values	outlier treatment, standardizing data, date formatting etc

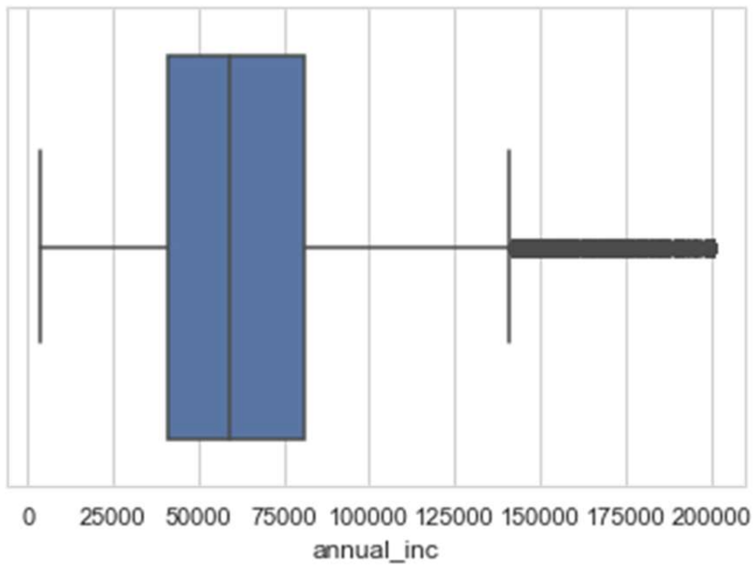
Univariate Data Analysis for loan defaulters vs fully paid w.r.t all columns features	
Plotting graph and stats for each column one at time	Estimating effects of features on loan status by observing

Segmented univariate analysis w.r.t charged off or fully paid loan status
Inferring risks of defaulters by feature like date, annual income, enquires, state, interested rate etc.

Bivariate analysis	
Plotting correlation and scatter w.r.t loan status and income	Deducing conclusions based on observed trends

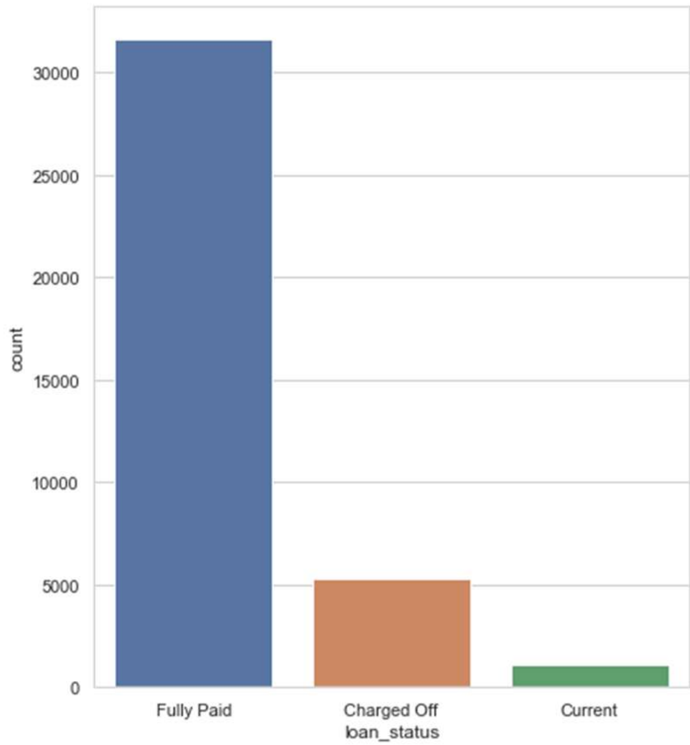
Lets look at annual income



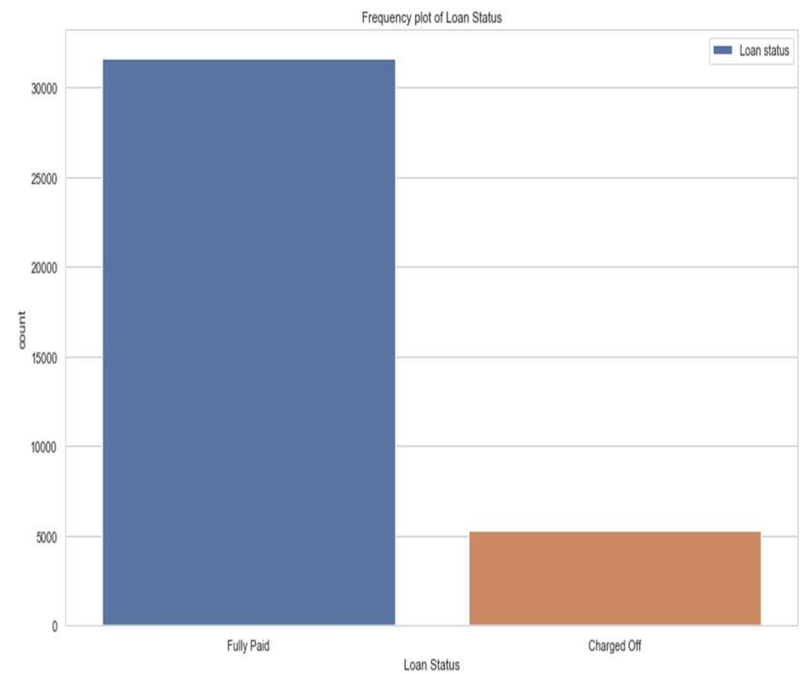


Before outlier treatment **After outlier treatment**

Loan status preparation

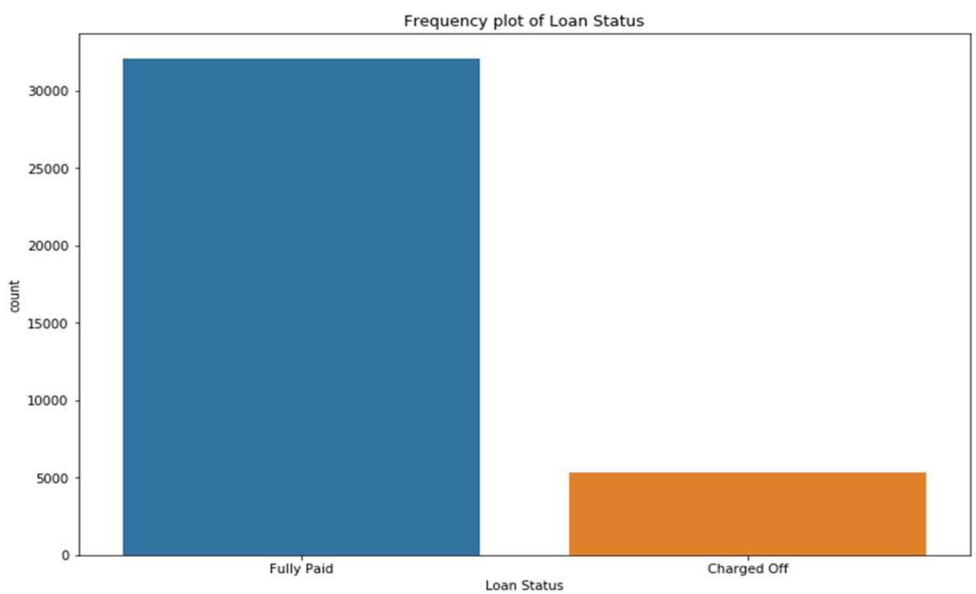


Raw data

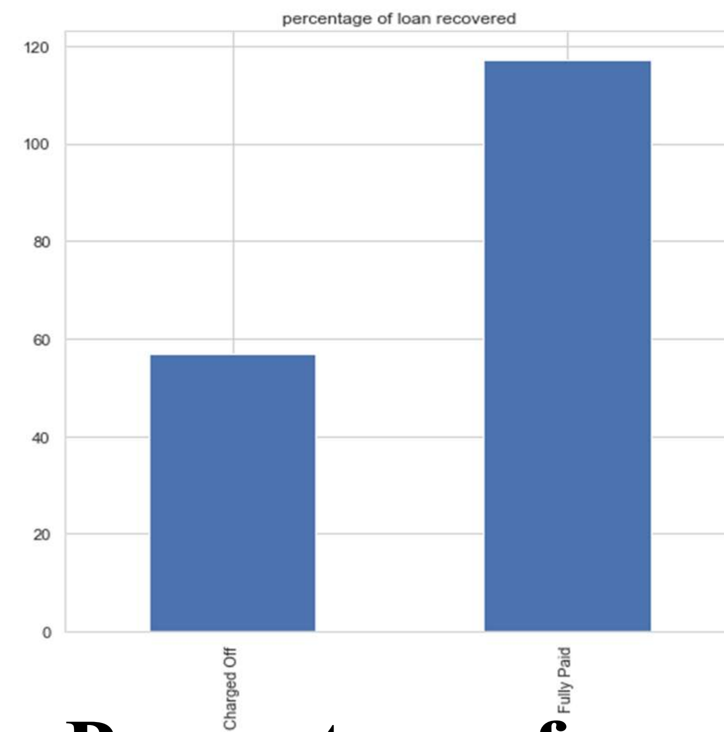


excluding current loan status

14.2% of people were charged off with 8.9% of total loaned amount

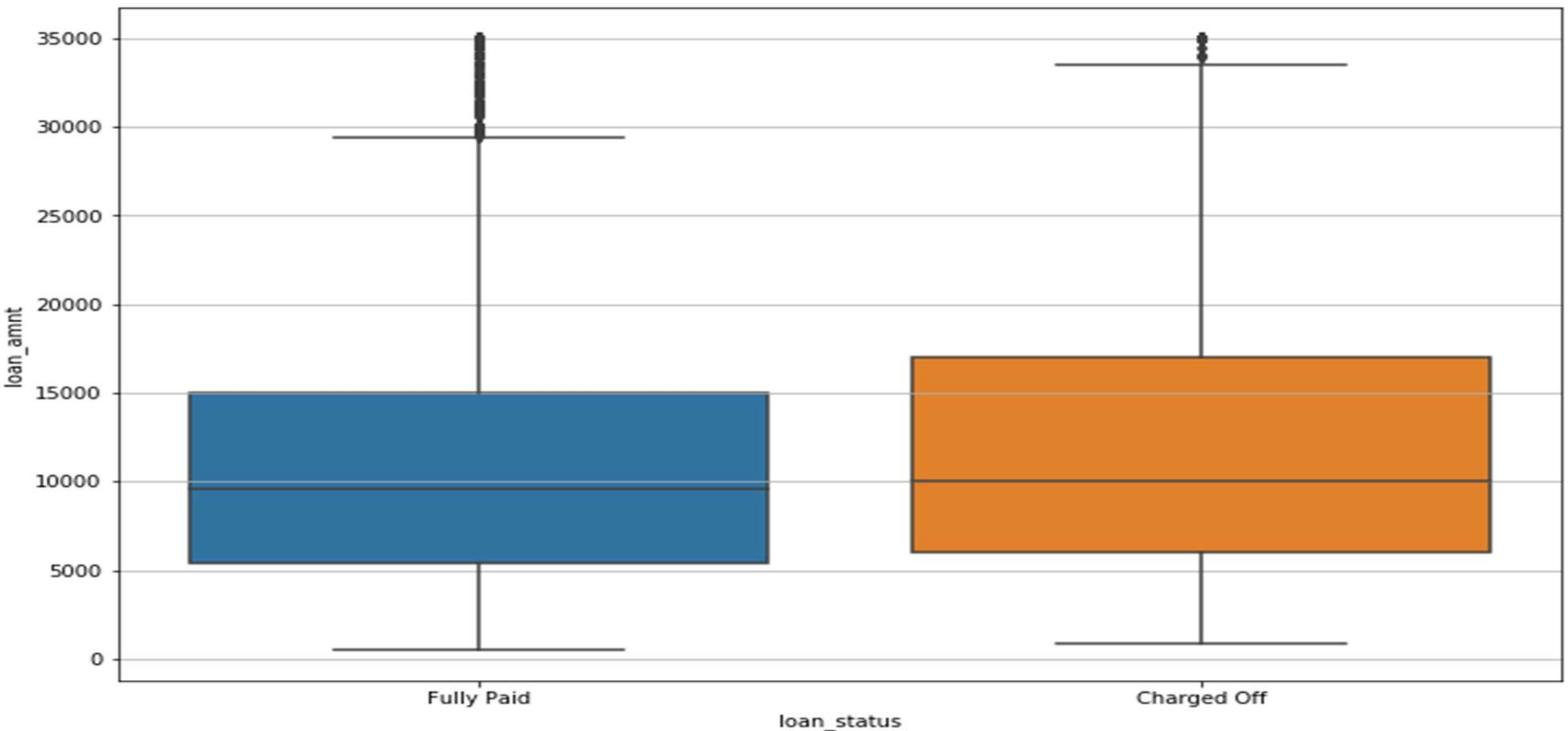


Frequency

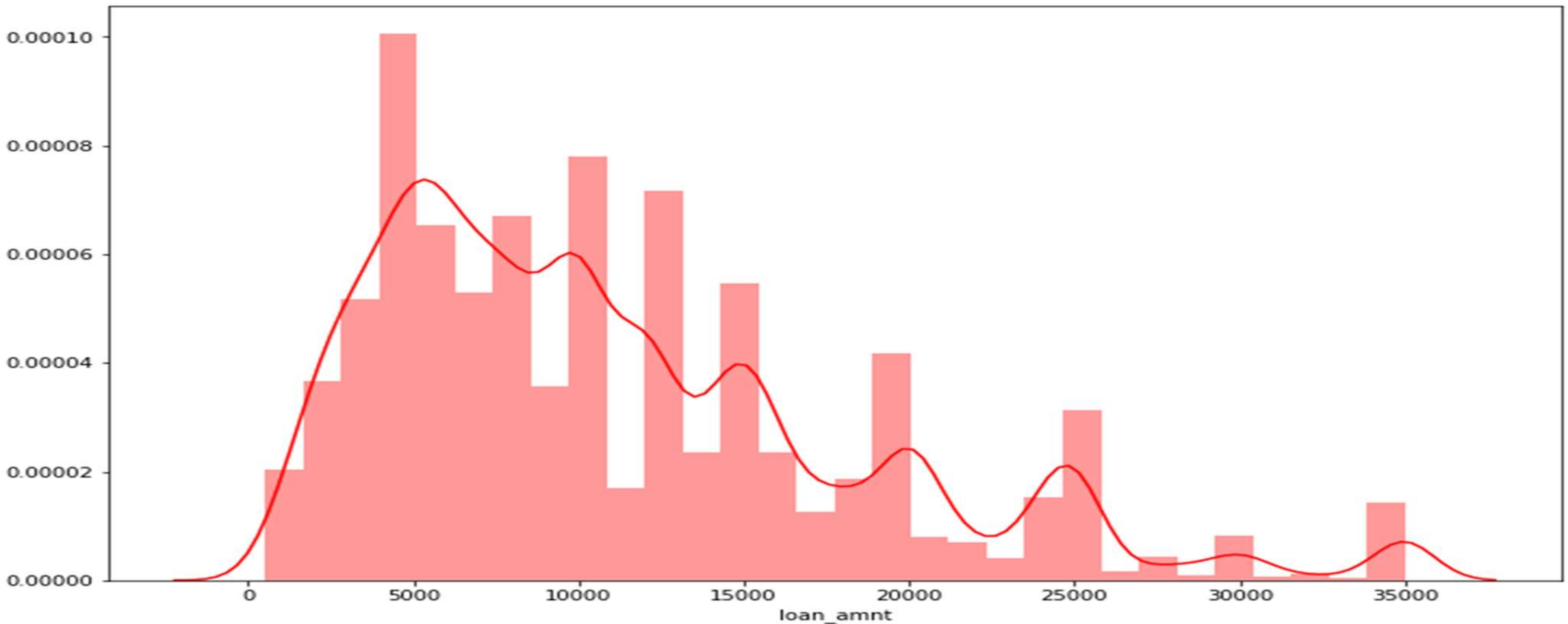


Percentage of amount

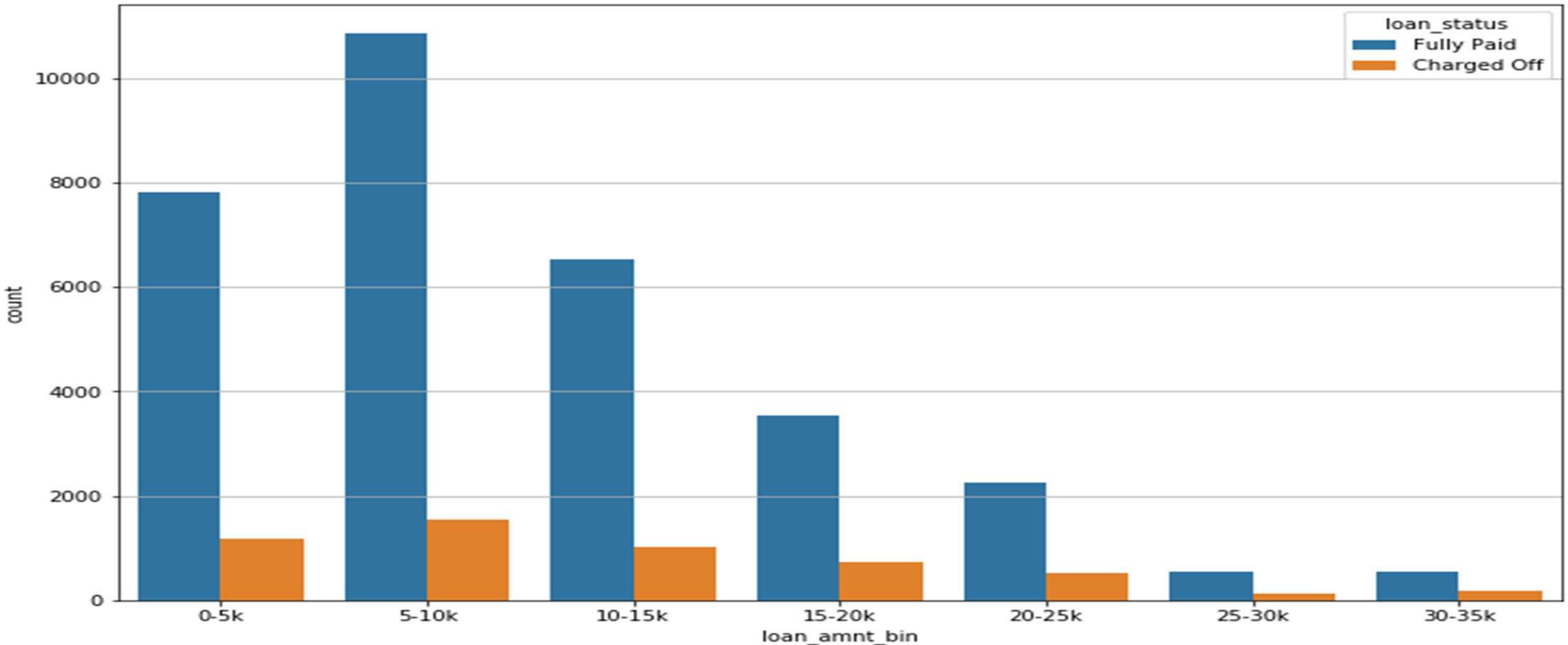
10k loan amount requested more it seems there



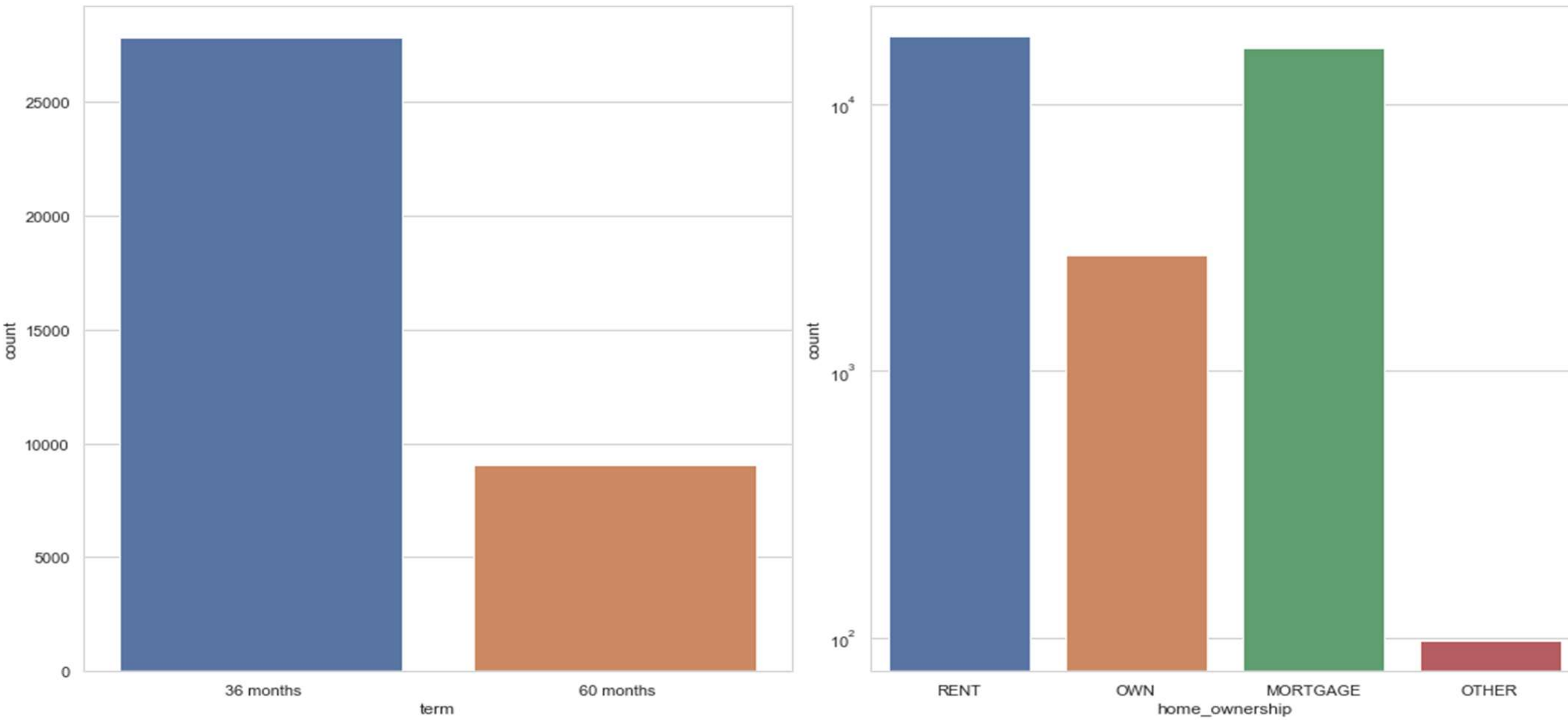
For Standard loan amount of 5,10,15,20,25.. Demand amount is inversely proportional to charged off money



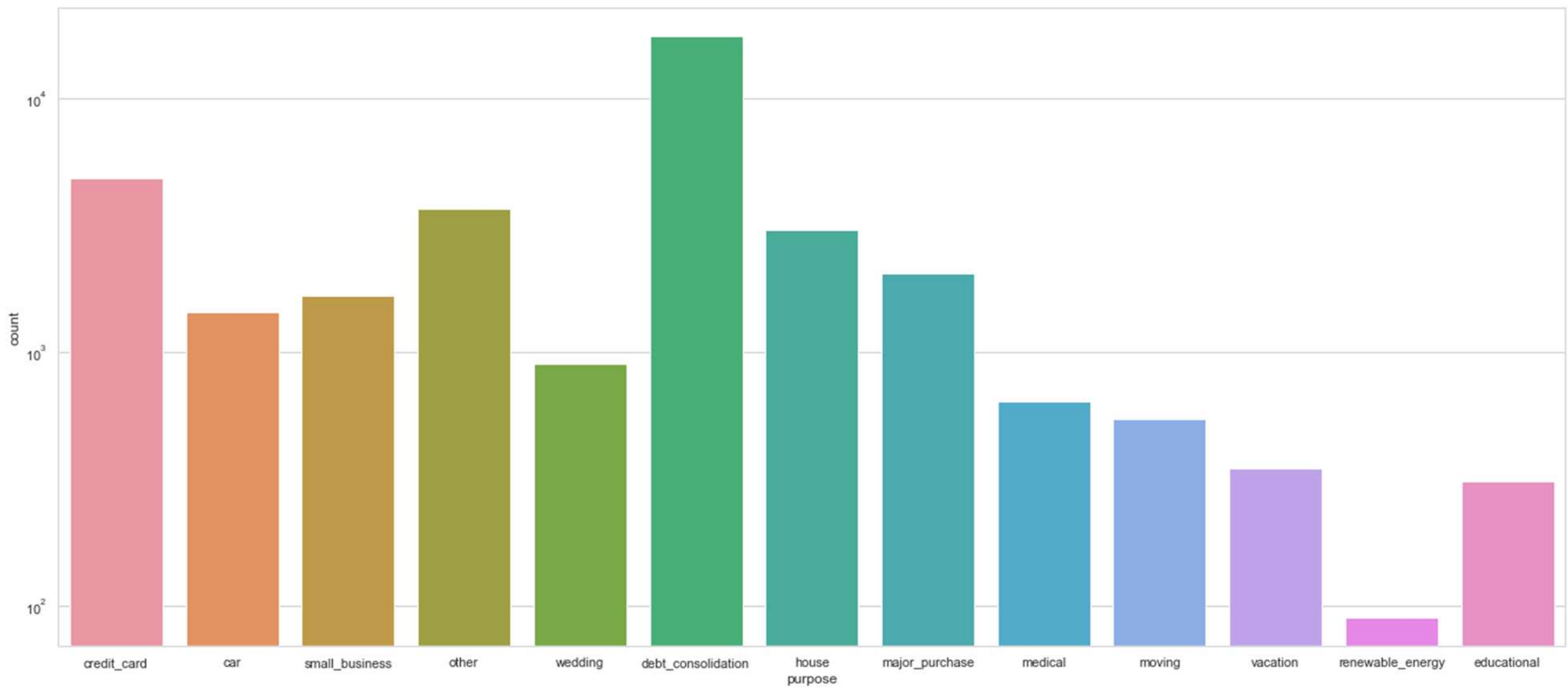
More Demand for lower loan amounts



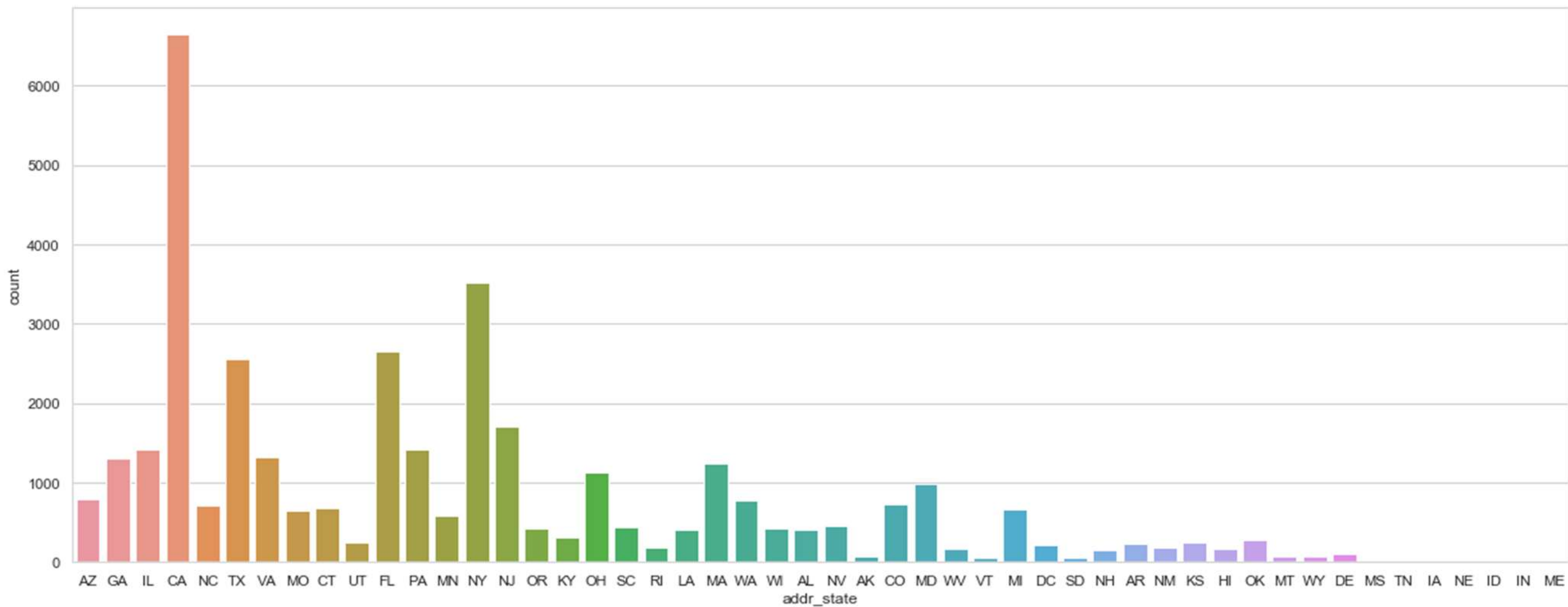
More Demand for 36 month slab by mainly for borrower living on rent and mortgaged homes



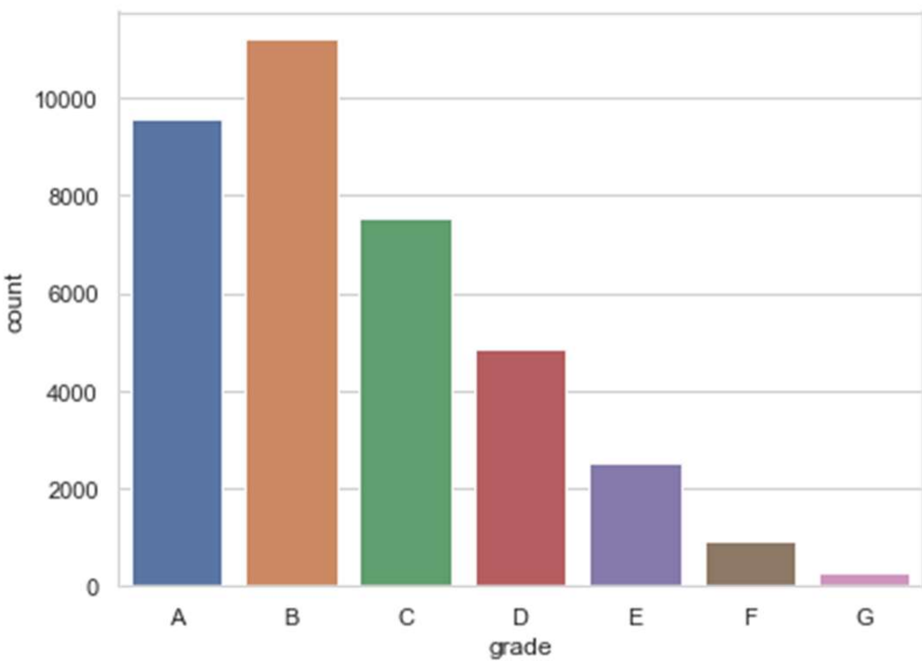
Loan Purpose

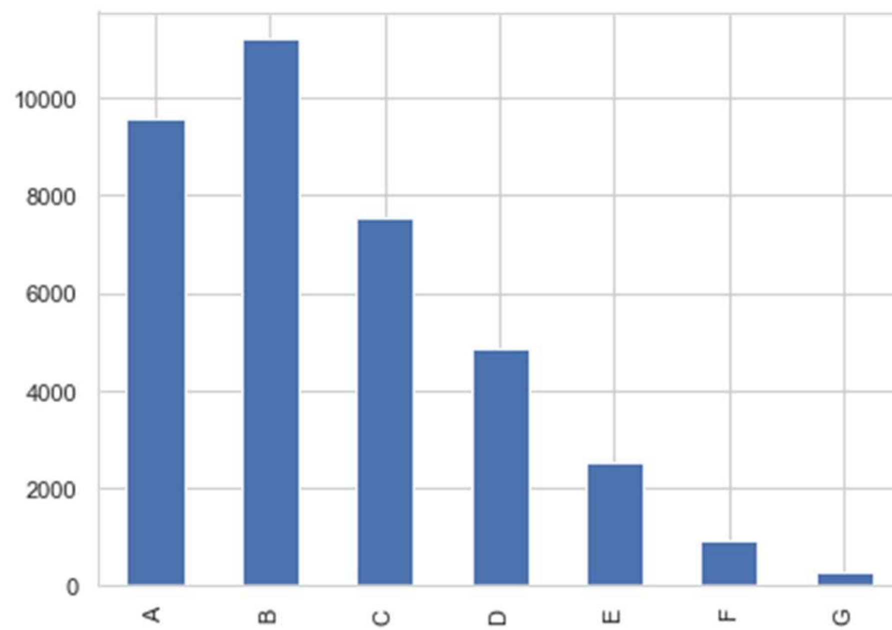


State wise - CA>NY>FL>TX

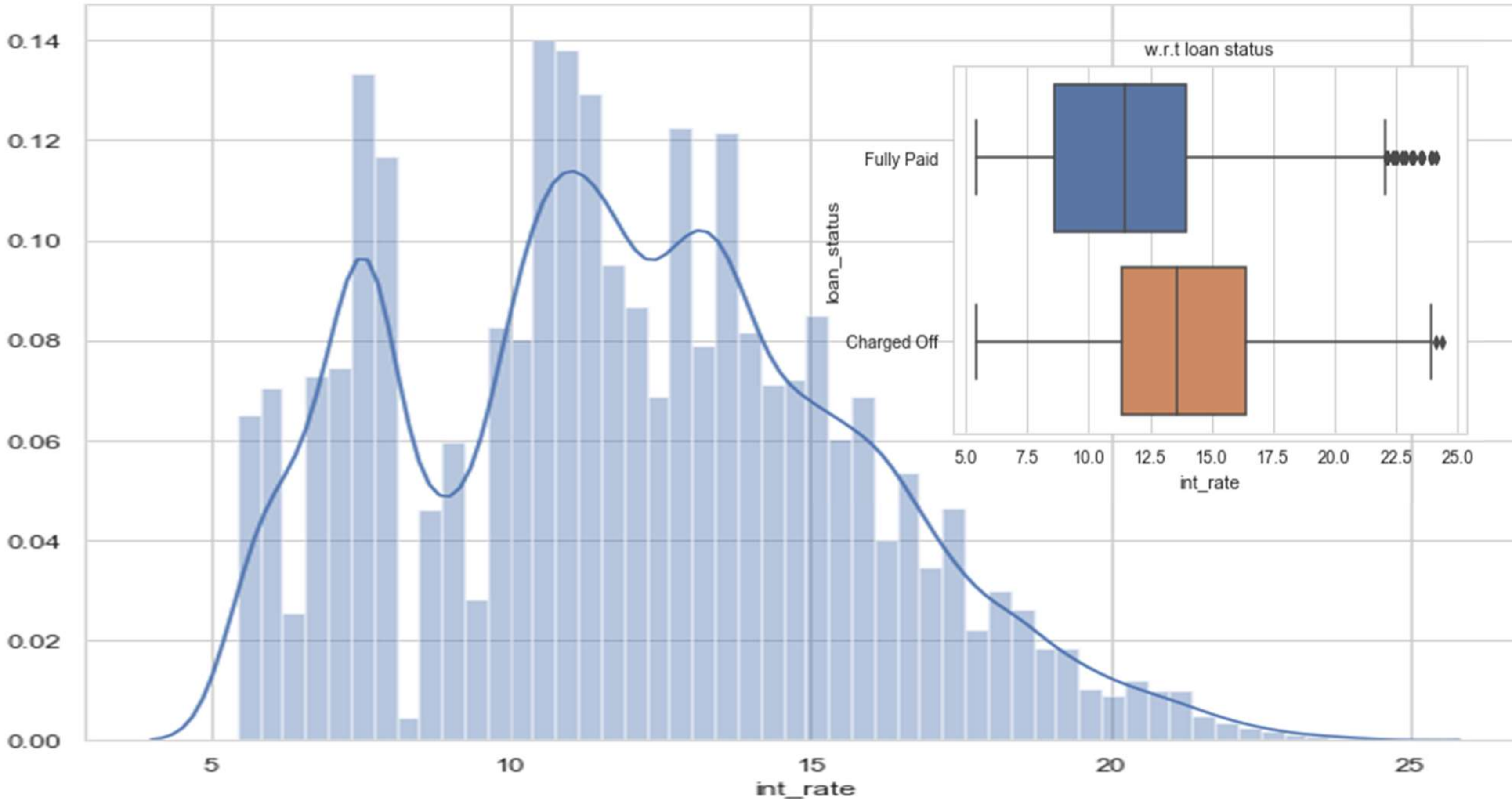


Grade wise loan – A and B are more safe

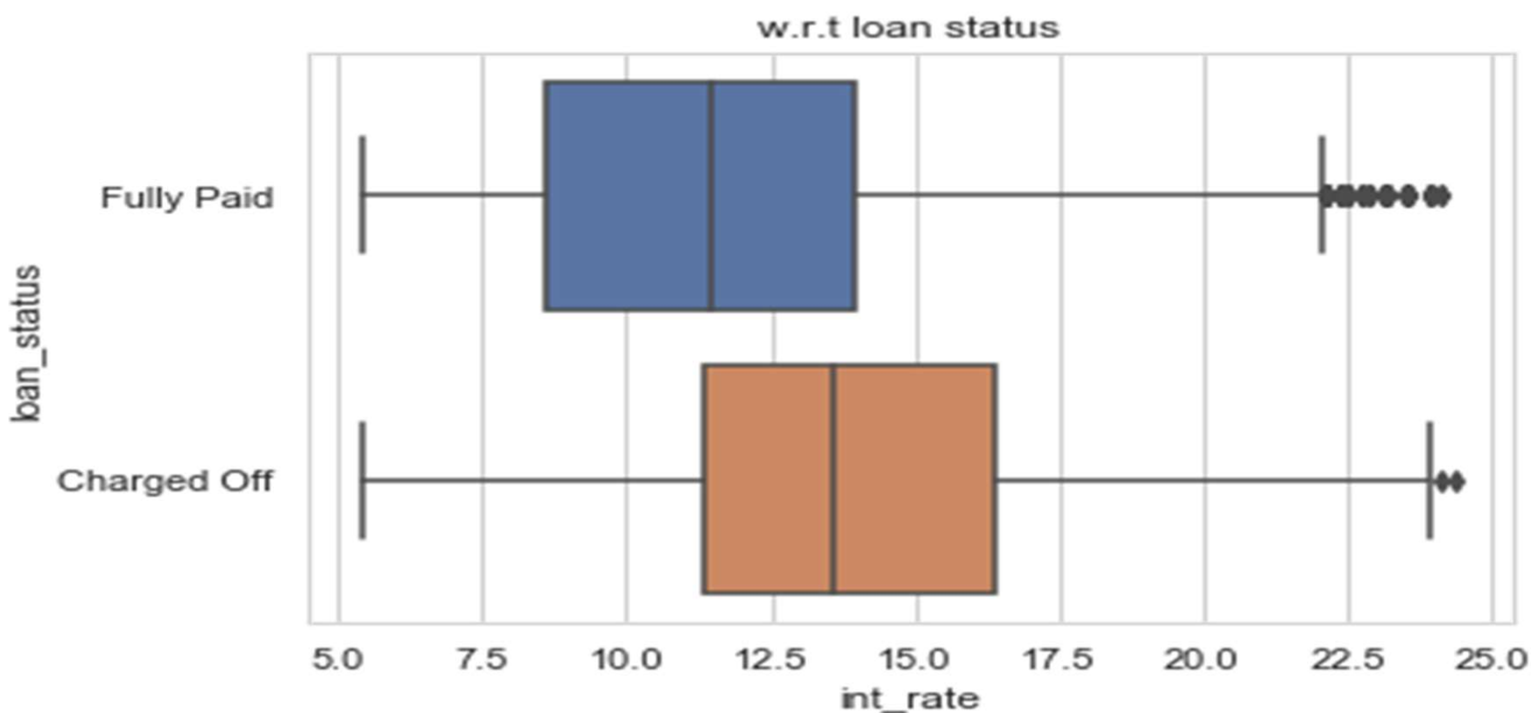




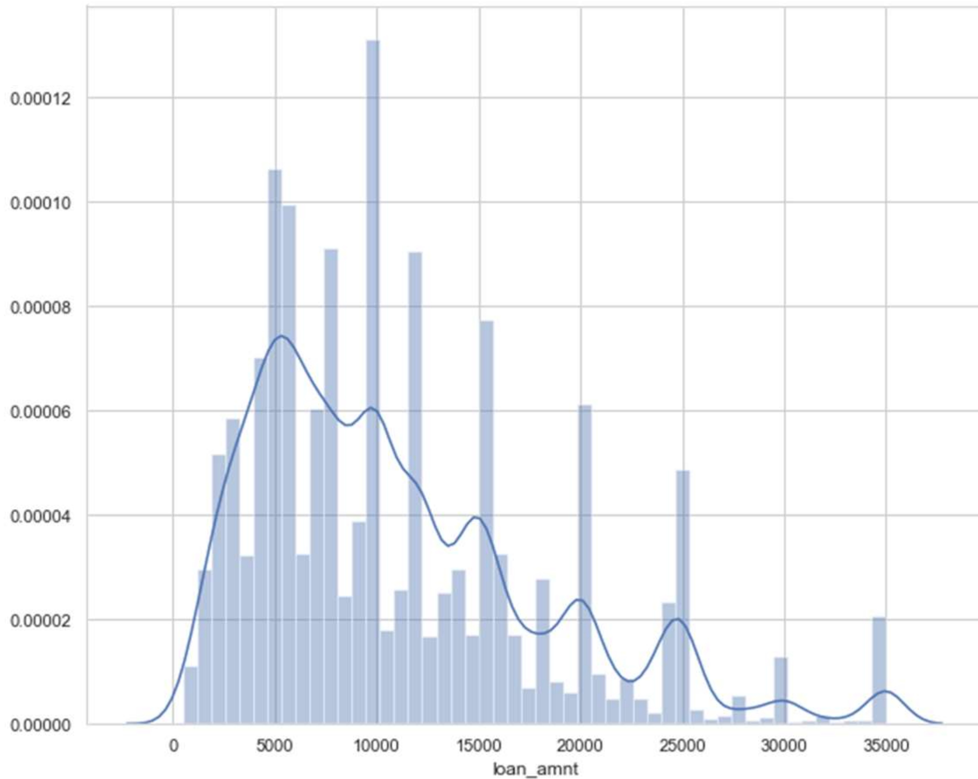
frequent interest rate are from 7 to 14 percent

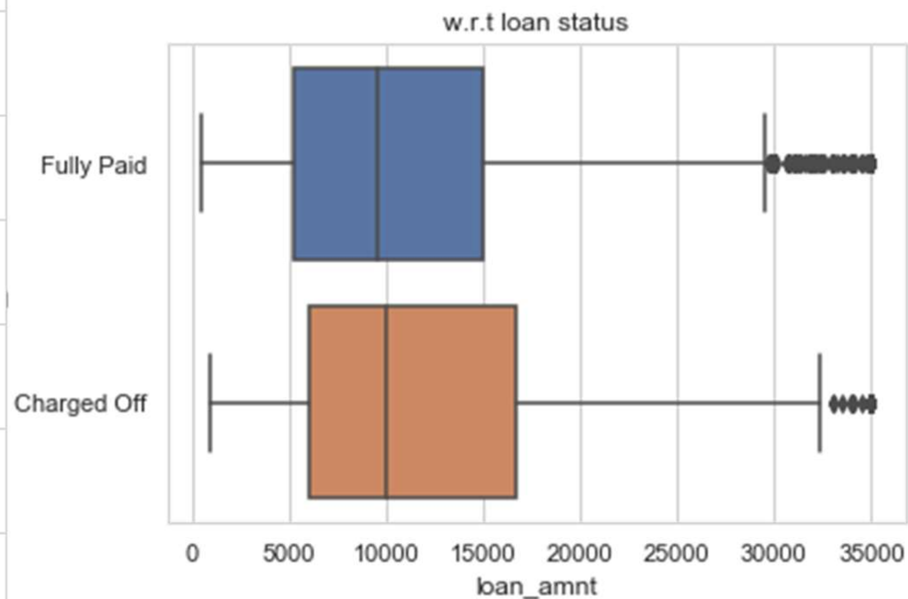


the more interest rate applied has more tendency to default.

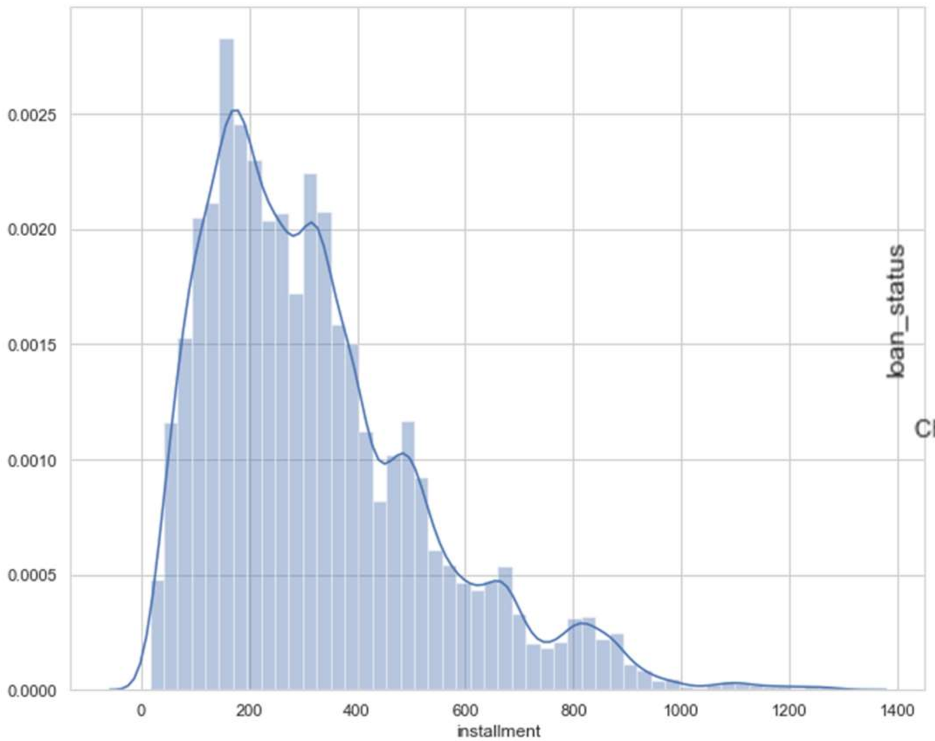


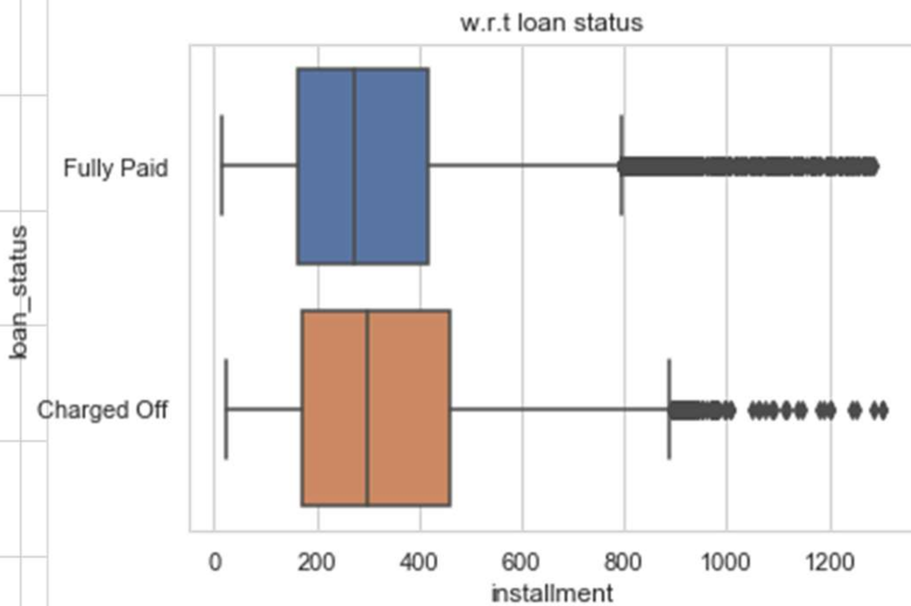
Bigger Loan amount has more tendency to default





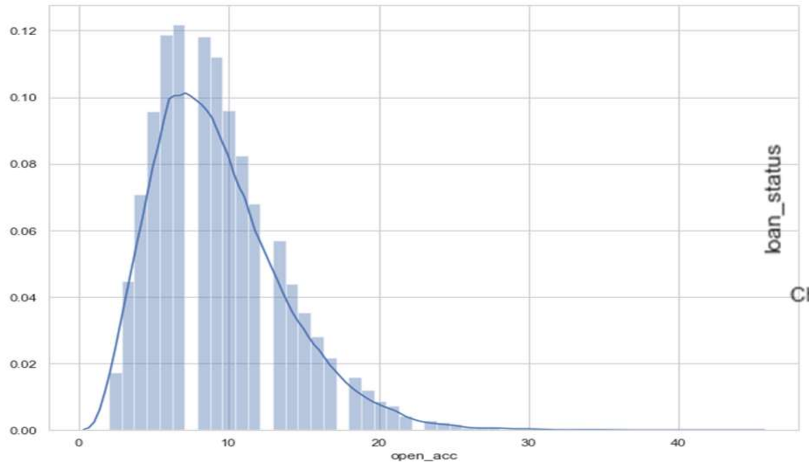
Installment amount

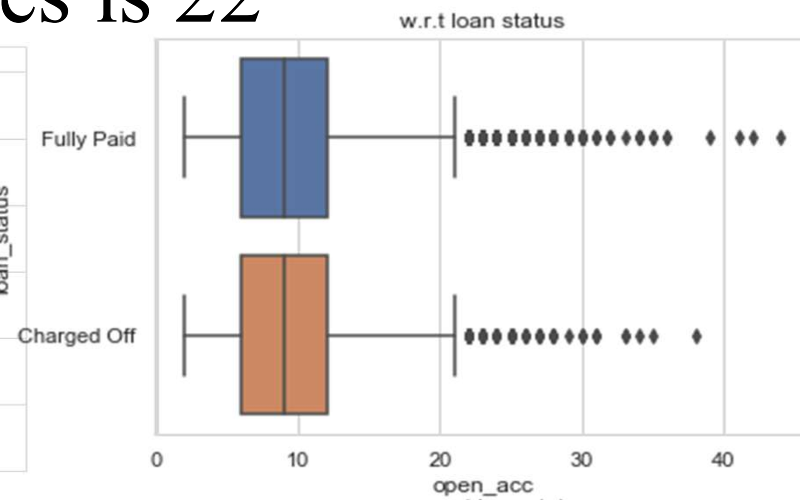




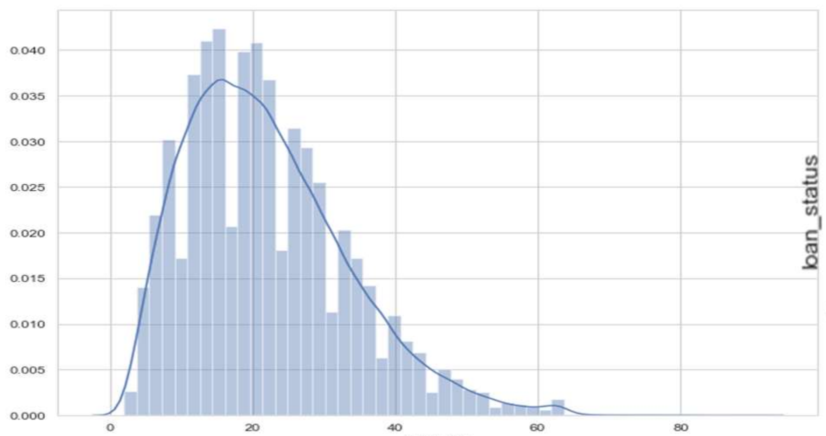
Avg credit lines is 22

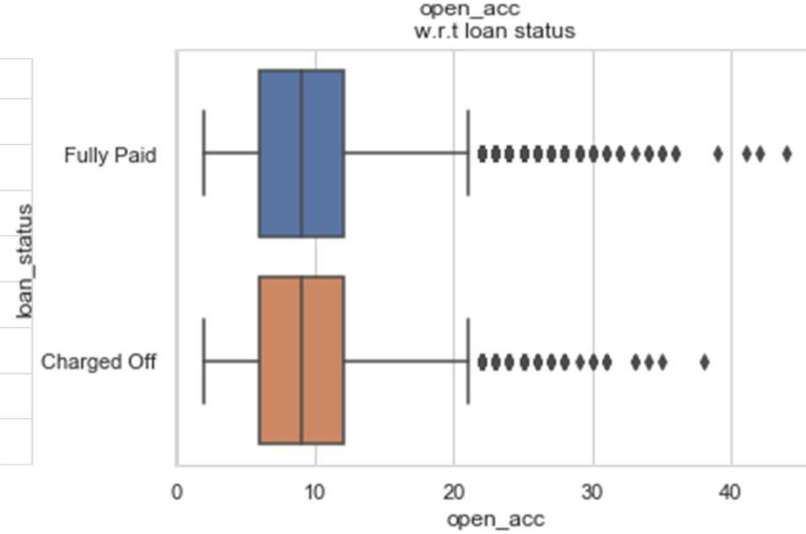
open



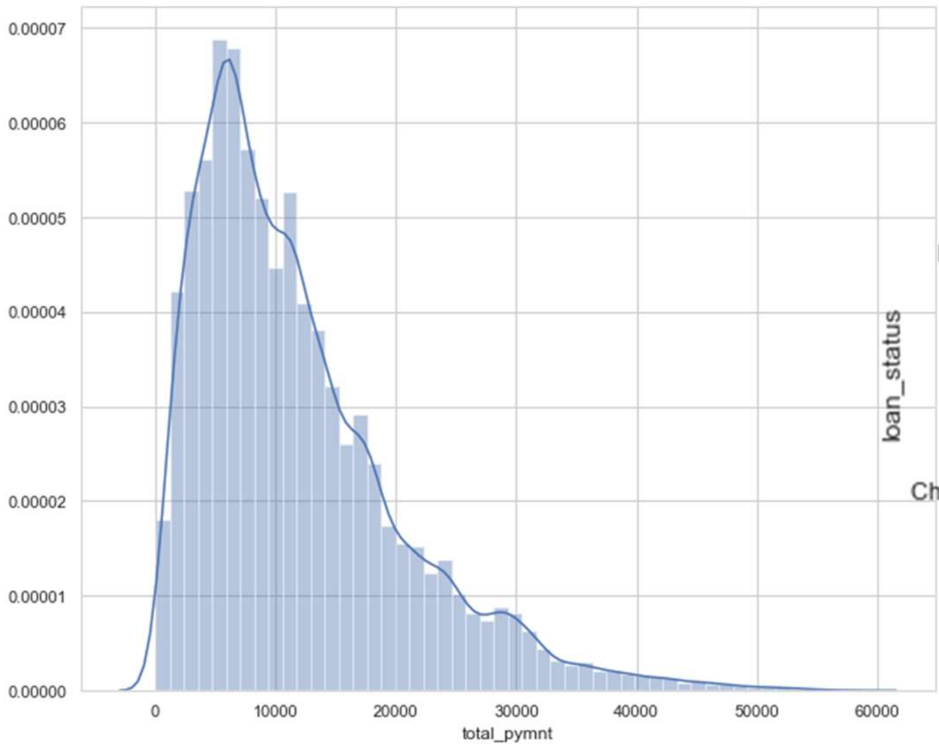


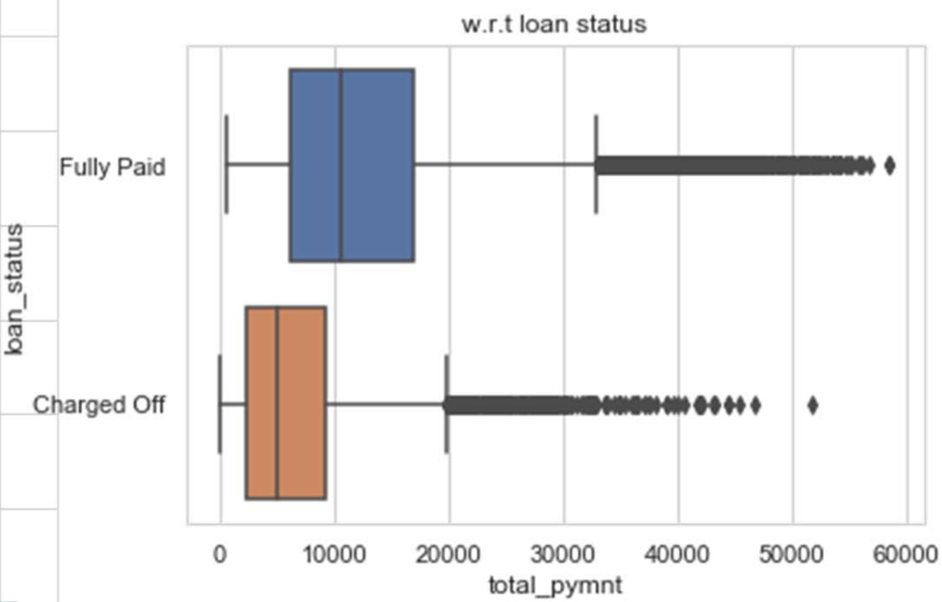
total



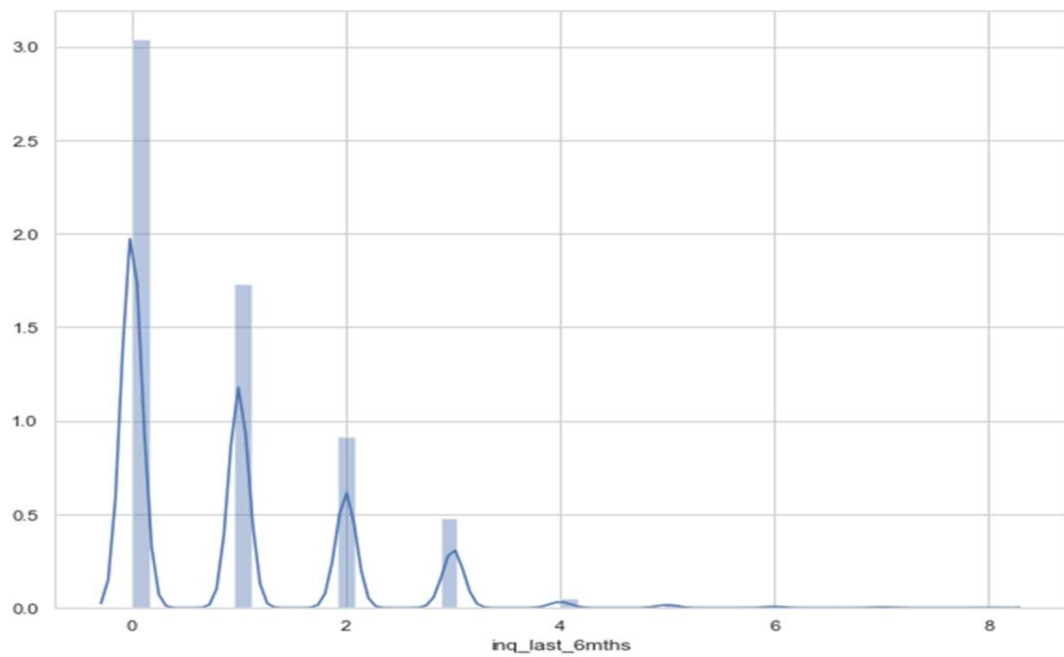


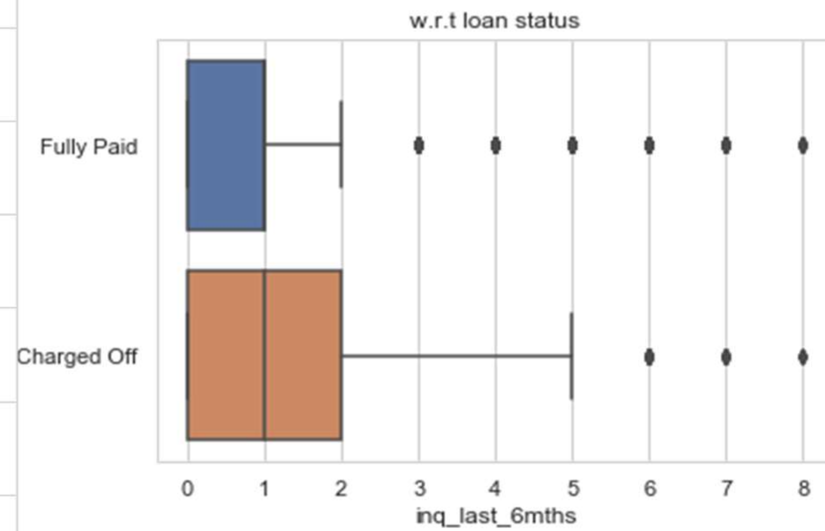
avg total payment for charged off loans is approximately half of the that of fully paid loans.





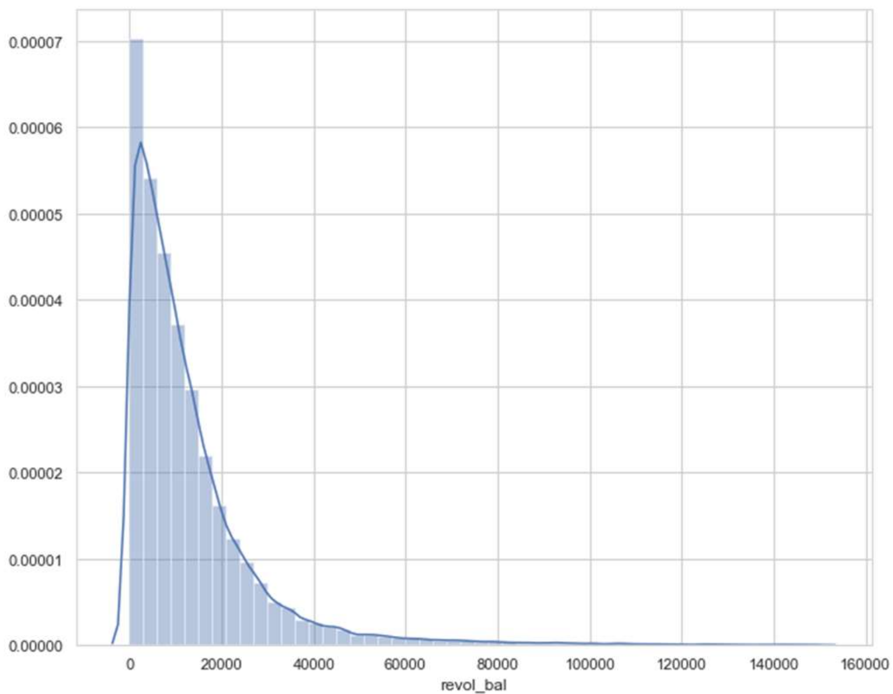
Enquiries in last 6 month

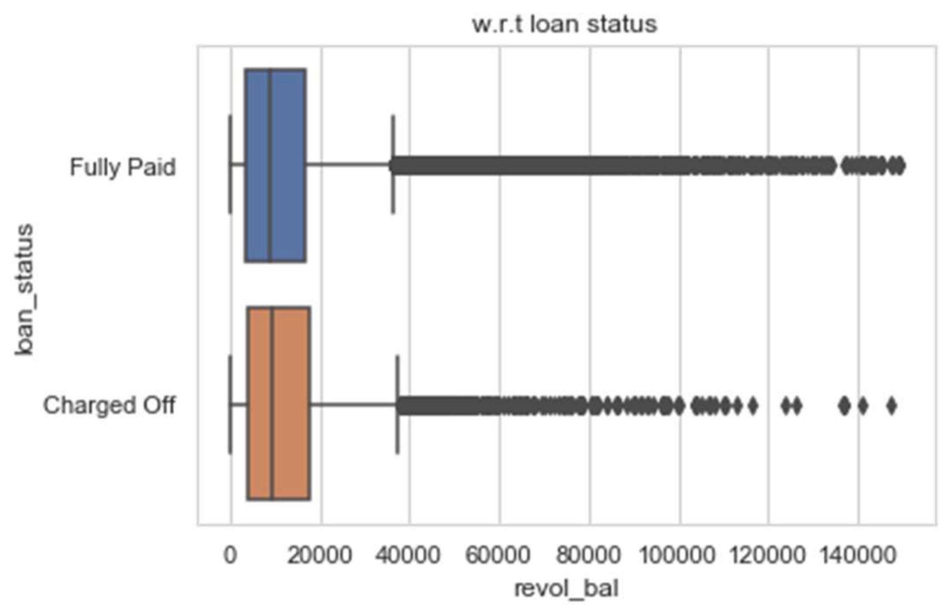




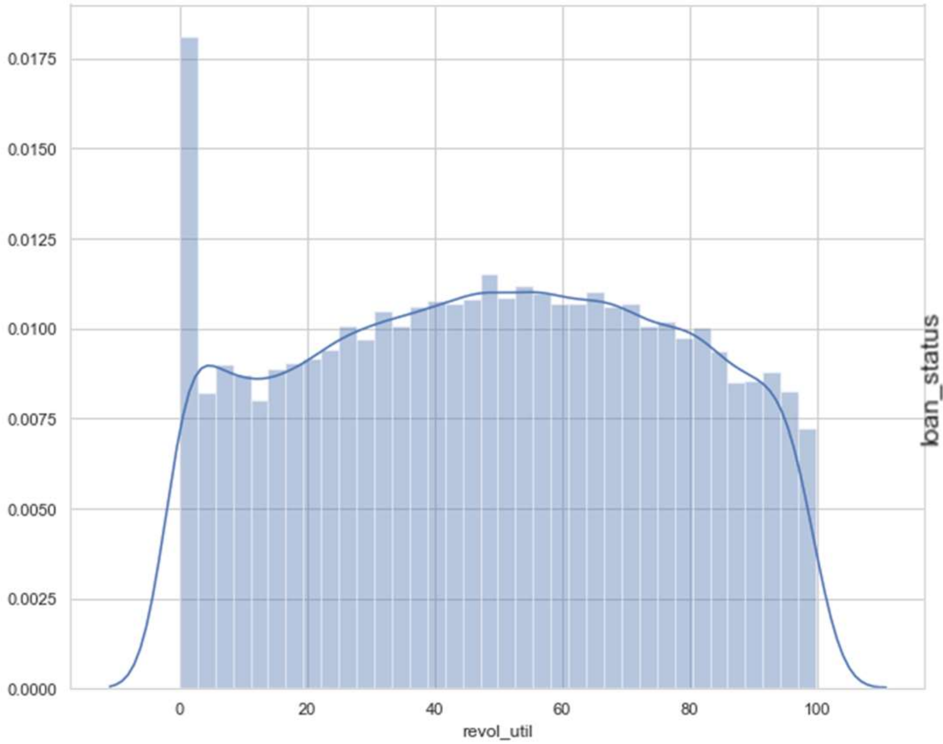
Clearly defaulter enquired more than paying

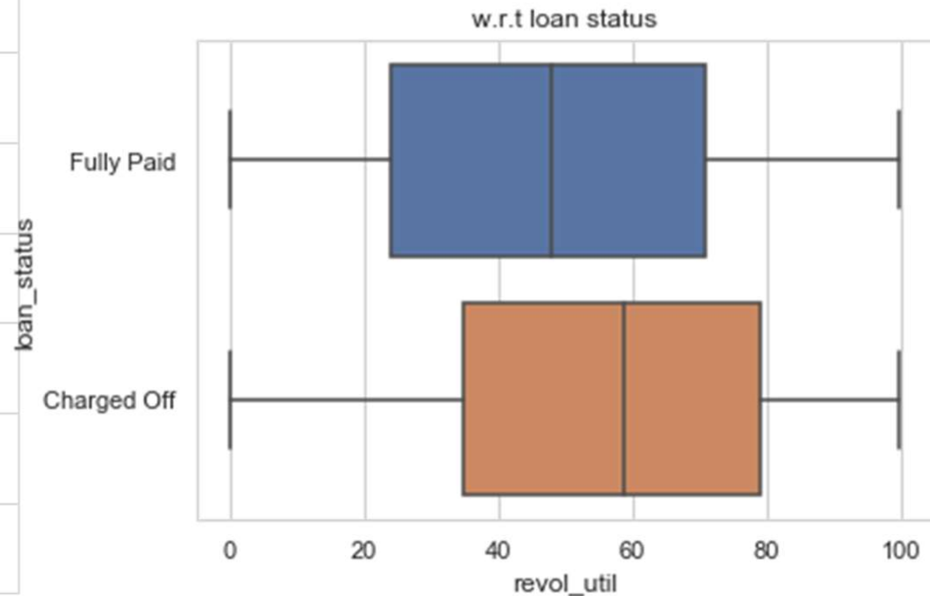
credit revolving balance



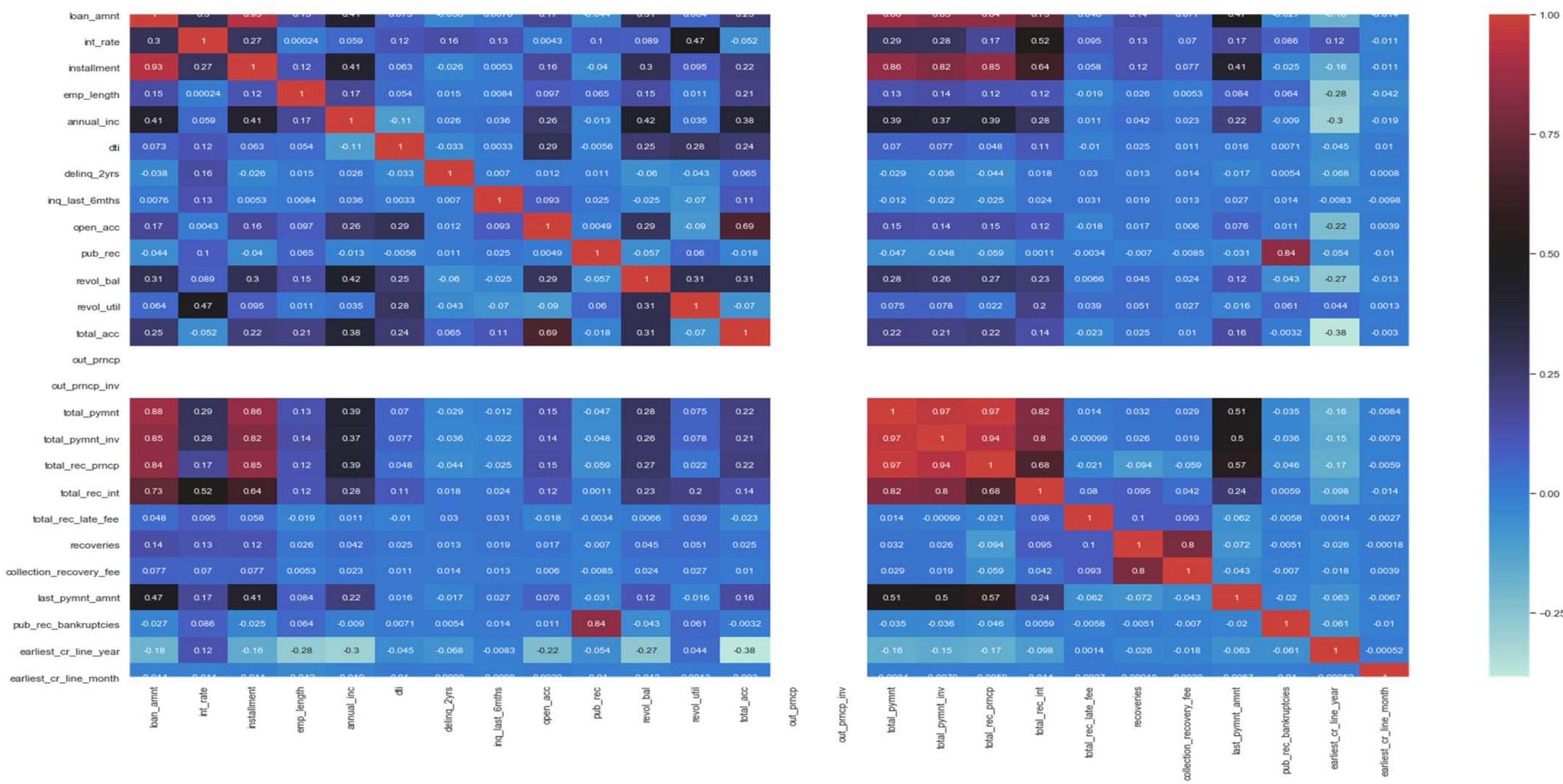


the amount of credit the borrower is using relative to all available revolving credit

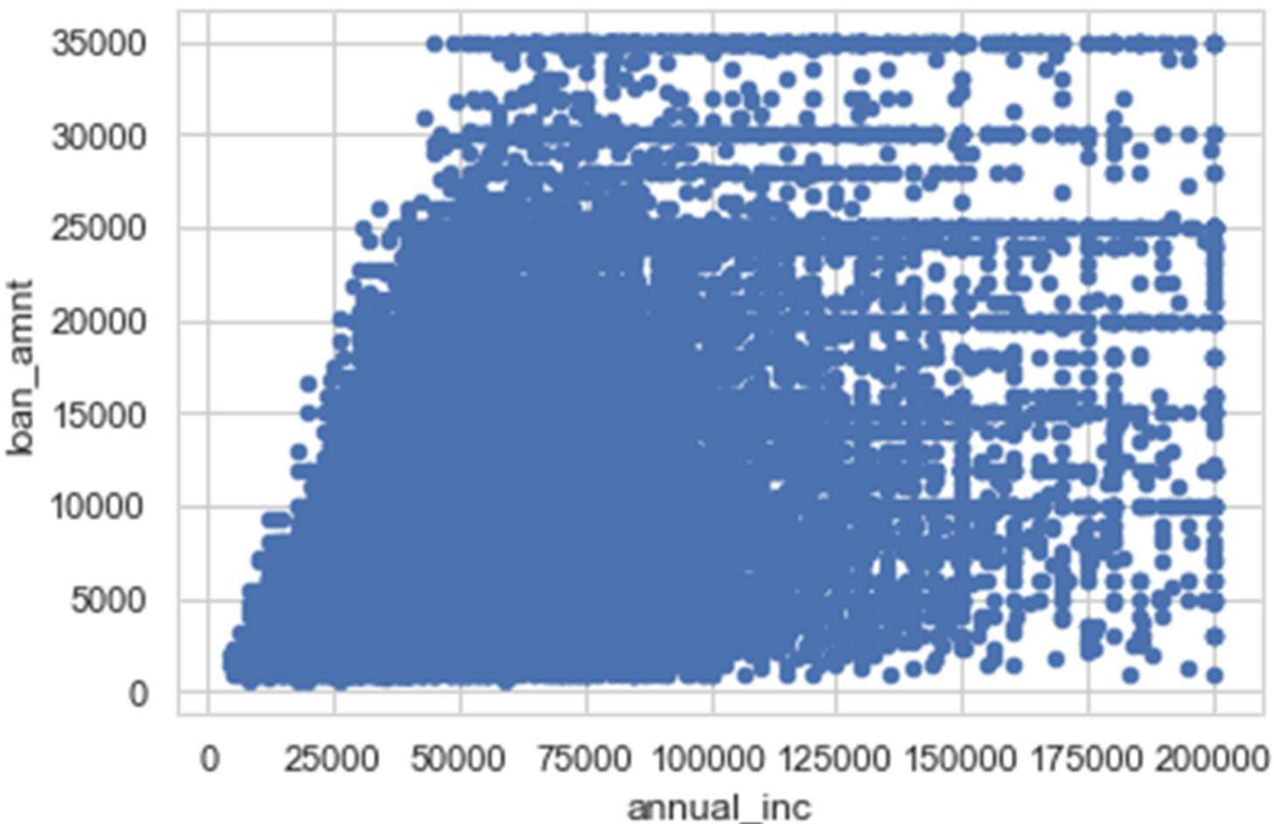




Correlation plot with all features



Annual income vs loan amount



mostly people asking for loan is lower income group for lower value loan

Inferences from EDA

1. People looking for loan from LC are usually looking for small loan amount
2. Majority of borrower are residing in rented or mortgaged home
3. More interest has more tendency to default. As for higher value loans more interest is charged by LC
4. It was observed that approximately 14.2% of people were charged off with 8.9% of total loaned amount
5. Grade A and Grade B are relatively lower in risk to provide loan
6. Defaulter enquired more than once in last 6 month
7. 10k loan amount requested more it seems
8. 5k is most given loan amount
9. Defaulter had slightly more installment amount
10. Maximum number of loans were distributed in theses states in order CA>NY>FL>TX
11. 36 month term is most wanted loan term slab
12. Interest rates varies with maximum observation between 7 to 14 percent