LENDING CLUB CASE STUDY

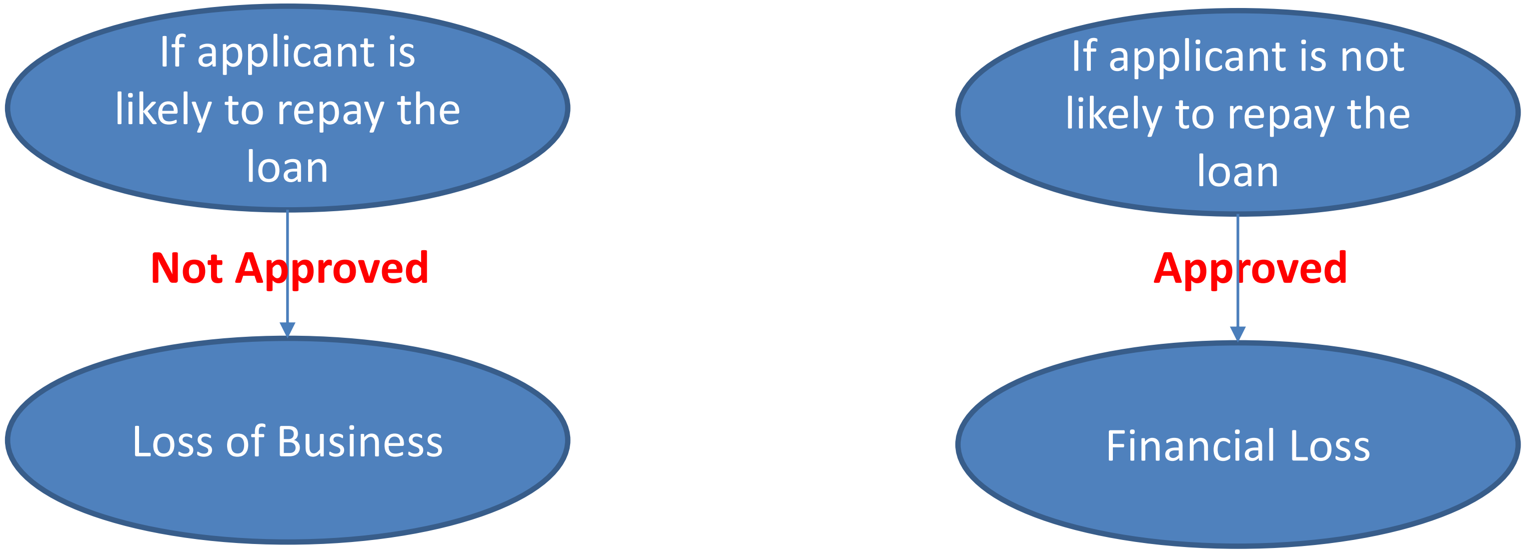
Submission by:

Merrymel George (merrymelgeorge@gmail.com)
Prashant Mittal (meetprashantmittal@gmail.com)

Problem Statement

A **consumer finance company** specializing in lending various types of loans to urban customers. Two **types of risks** are associated with the bank's decision to approve or reject the loan:

RISK 1 **RISK 2**



Objective

Understand and analyse the data to identify the driving factors behind loan default so that the company can utilize this knowledge for its portfolio and risk management

The analysis should help the company recognize a potential defaulter and then avoid/refuse lending loans to that person, hence reducing the risky loan approvals

On a similar line, the company should also be able to identify the potential borrower who pays back the loan fully, so that, the borrower is provided the loan when he applies

Identifying the two categories of borrowers will help the company reduce it's loss of business as well as financial loss

Steps involved in EDA

Data Understanding and Business Understanding

Data Cleansing and derivation based on data understanding and business/domain understanding

Perform Univariate, Segmented Univariate and Bivariate Analysis on the cleansed data

Summarize the inferences and provide the recommendations

Data Understanding

Loans.csv file that contains the various attributes and values for loans for the period of 2007 to 2011 has been provided.

Based on the first hand analysis of the data, the attributes can be broadly classified into 3 categories:

Customer Demographics

- Employment Length
- Employment Title
- Annual Income
- Home Ownership
- Description
- Zip Code
- State Address

Loan Information

- Loan Amount
- Interest Rate
- Funded Loan Amount
- Loan Status
- Grade
- Term
- Issue Data

Customer Behavioural Data

- Application Type
- No :of Delinquent Accounts
- Delinquency 2 years
- Instalment
- Revolving Balance
- Recoveries
- Earliest Credit Line

Data Cleansing and Data Derivation

Columns with more than or equal to 60% NULL (NaN) values were removed

Customer Behaviour columns were removed as it doesn't help for the analysis

Columns with only one unique values were also removed as such attributes doesn't provide any insight

Columns with all values distinct, except the ID(kept to uniquely identify a row), were also removed as these attributes wouldn't add any value to our analysis

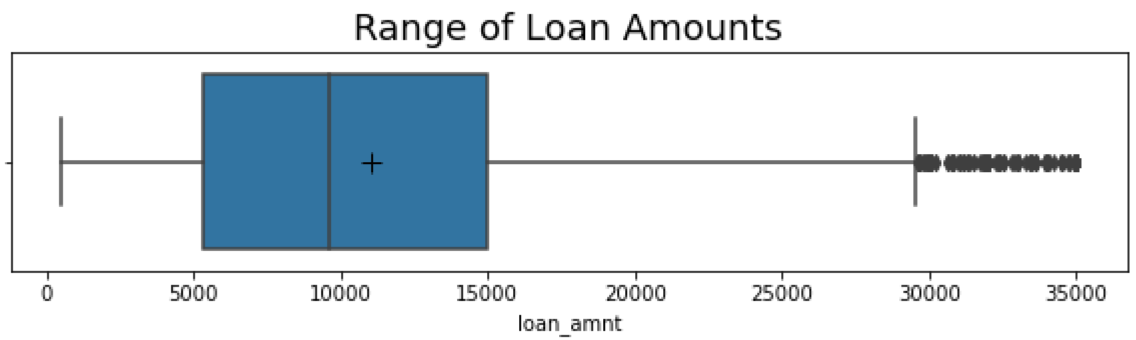
Columns with only two distinct values were checked and on finding that a couple of columns had only one unique value and NULL values, those columns were also removed

Rows that are corresponding to 'Current' loan status were also removed as these rows corresponds to data of progressing repayments and hence doesn't help in the analysis

Outliers in the annual income attribute were handled as and when it was found during the analysis process.

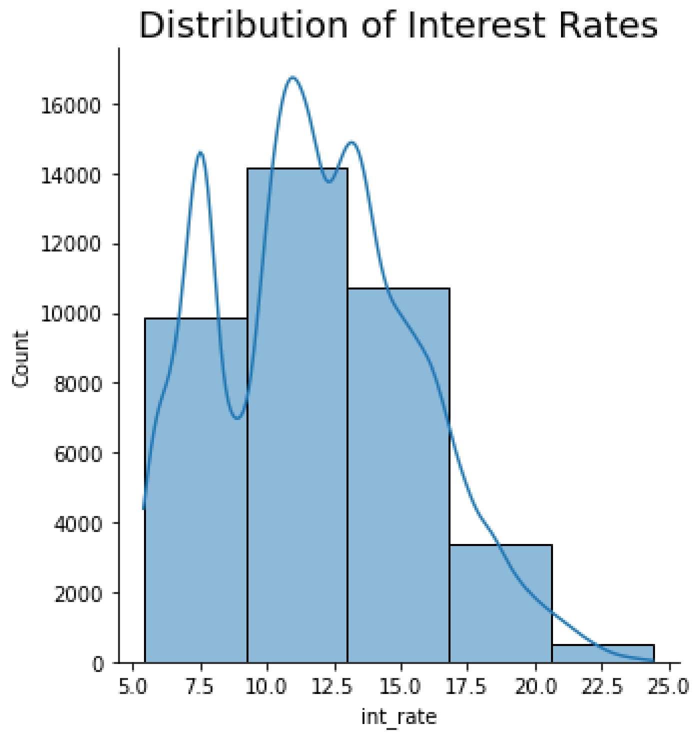
Loan issue month and year were derived from the issue date attribute with a view that these derived attributes may help find relevant insights

Univariate Analysis



Inference

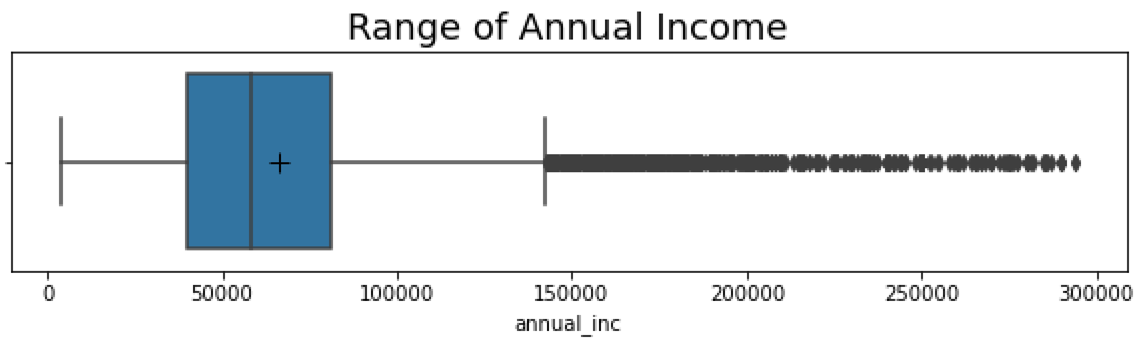
Majority of loan amounts falls in the range of ~$5K and ~$15K with a mean of ~$11K



Inference

Majority of loans falls in the interest rate range of ~9% and 14% with mean at ~12%

Univariate Analysis



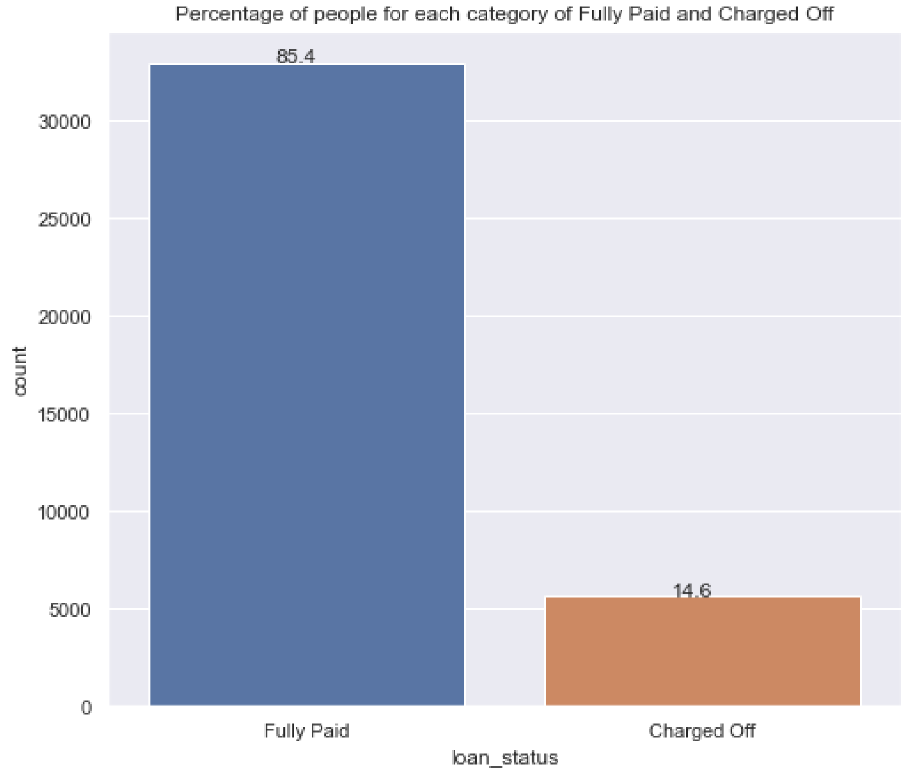
Inference

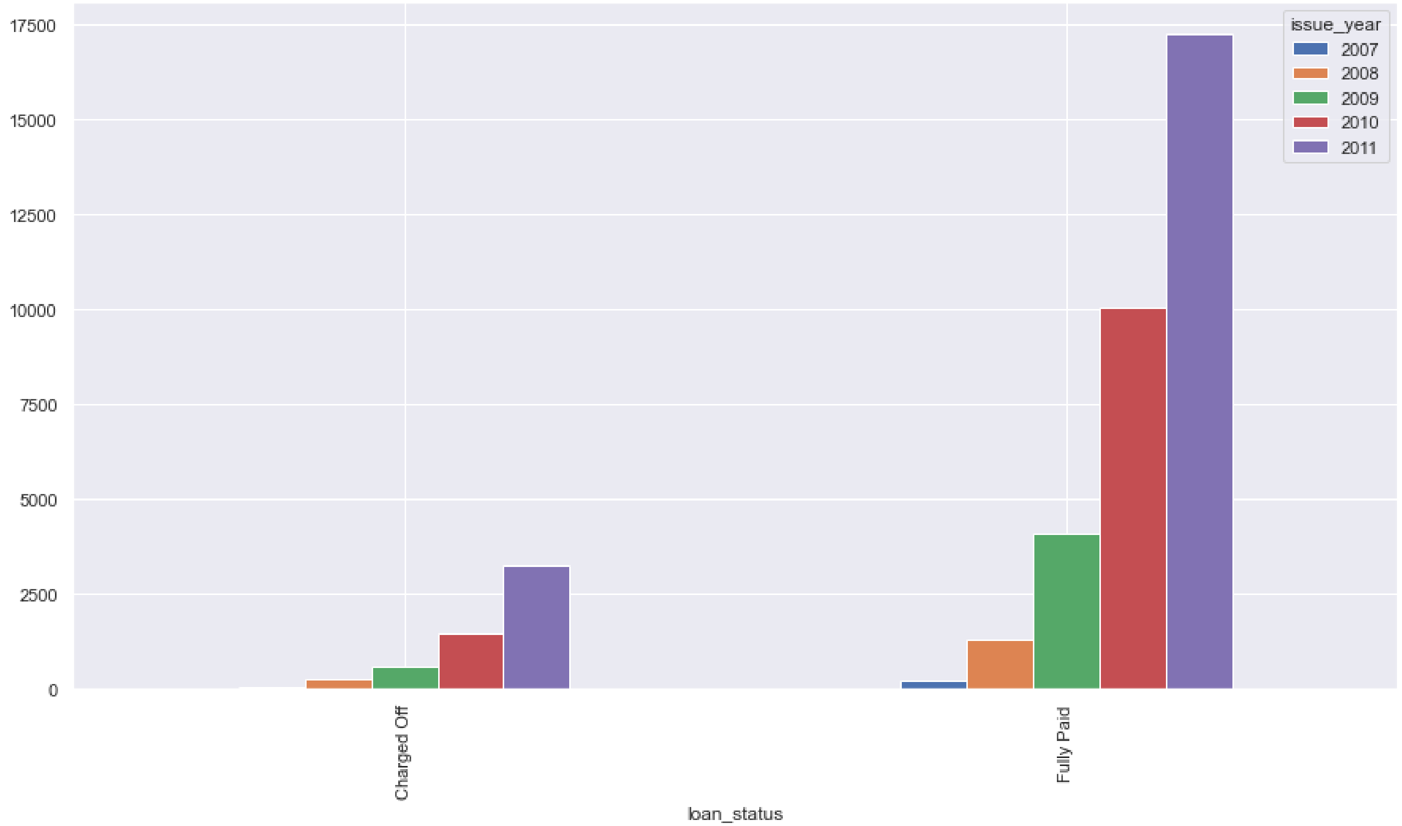
Majority of people who have taken loans falls in annual income range ~$40K and ~$81K with a mean value of $66K

Inference

Majority of people who have taken loans have an employment length ranging from 2 years to 9 years with a mean value of ~5 years

Segmented Univariate Analysis





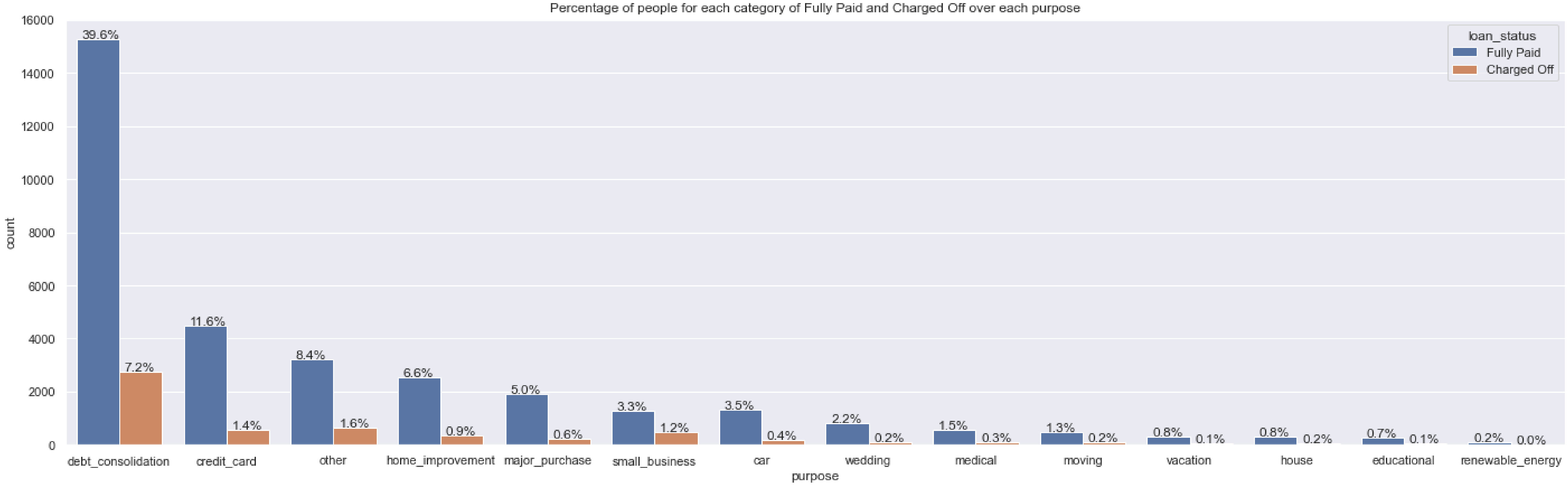
Inference

Only ~15% of loans have not been paid off completely

Inference

Starting from 2007 to 2011, the number of loans that were Charged Off and Fully Paid has kept on increasing

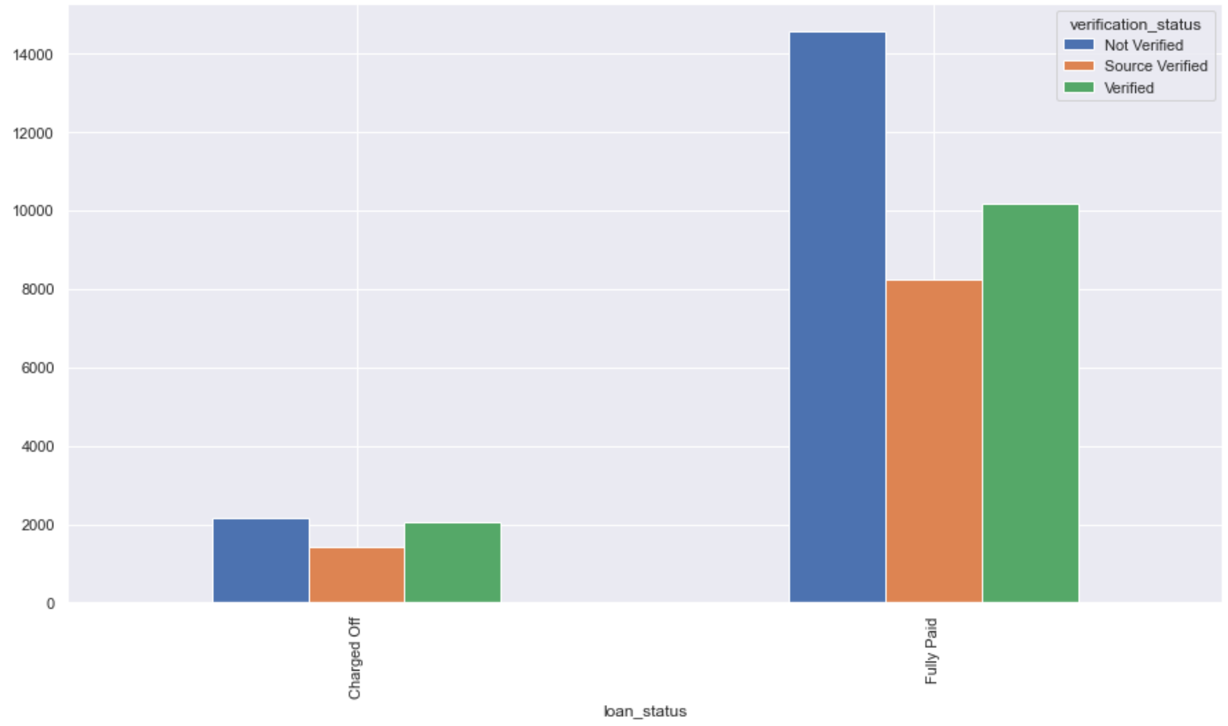
Segmented Univariate Analysis

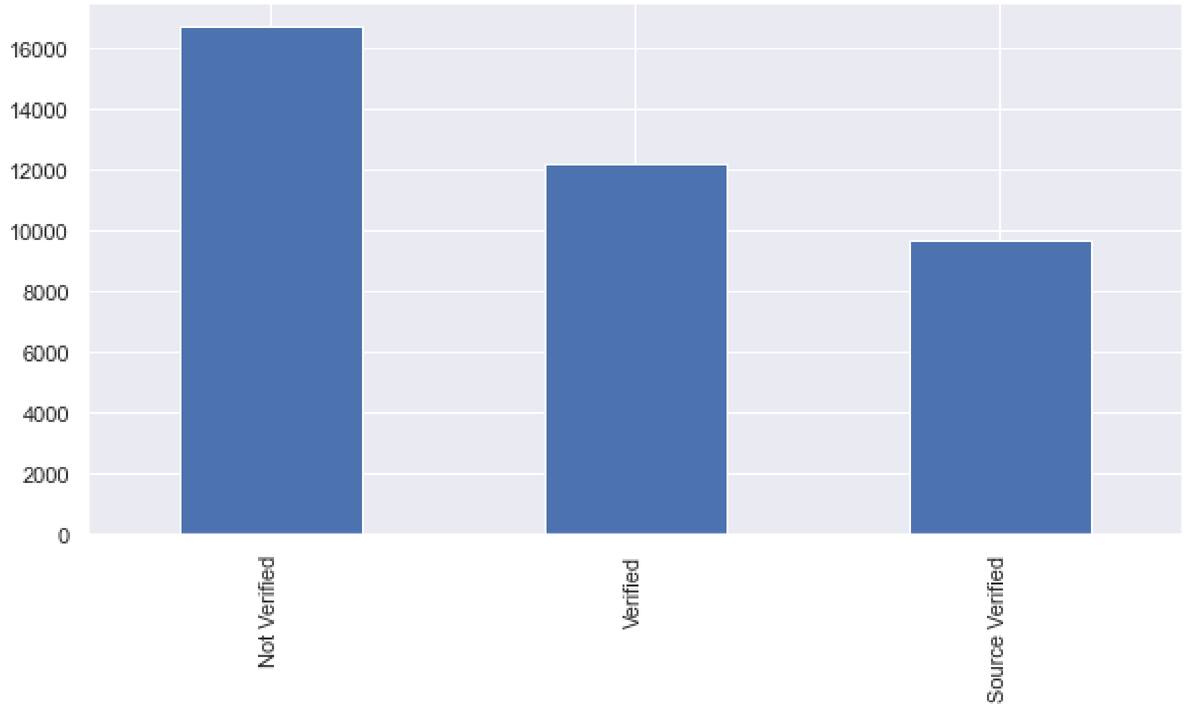


Inference

The major purpose behind people taking loan is Debt consolidation and also, among the loans that are Charged Off, Debt Consolidation comes out to be of high majority

Segmented Univariate Analysis

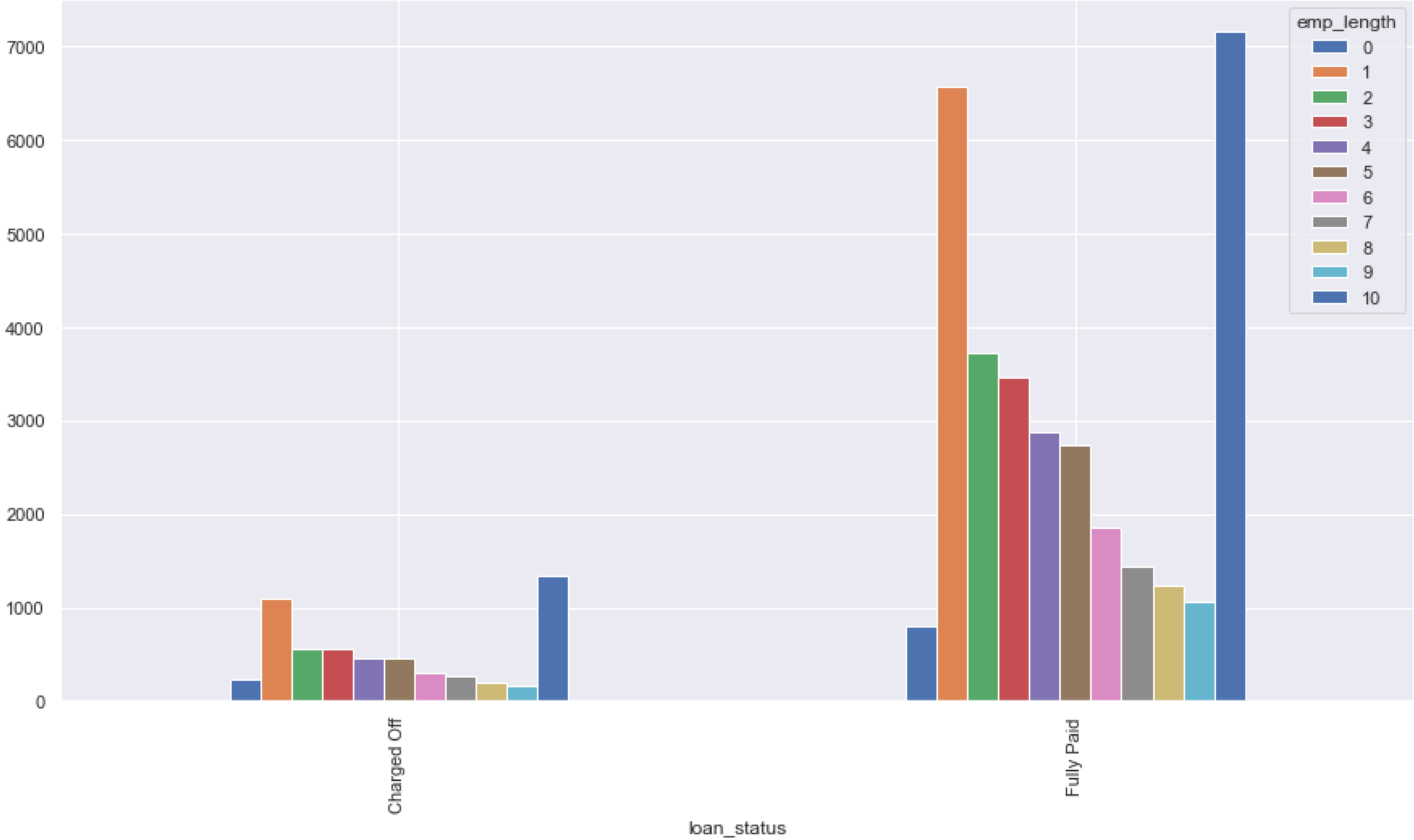




Inference

Be it Charged Off or Fully Paid, the verification status - Not Verified is the highest, meaning verification process has to be improved

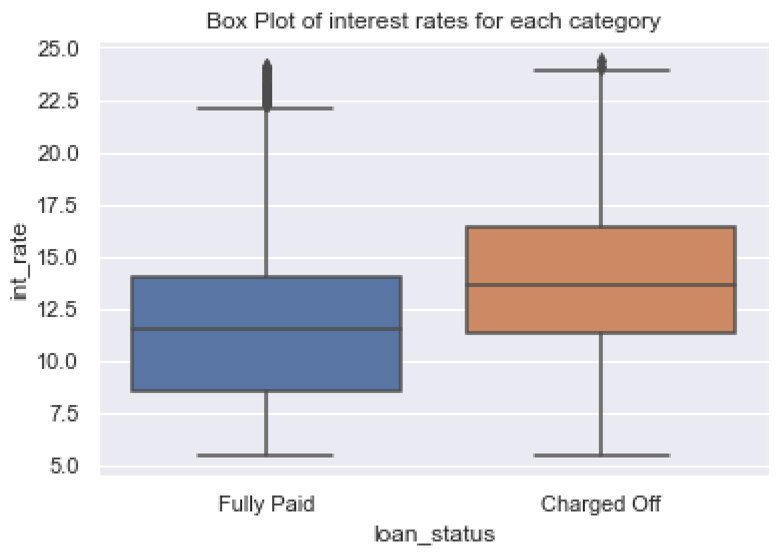
Segmented Univariate Analysis



Inference

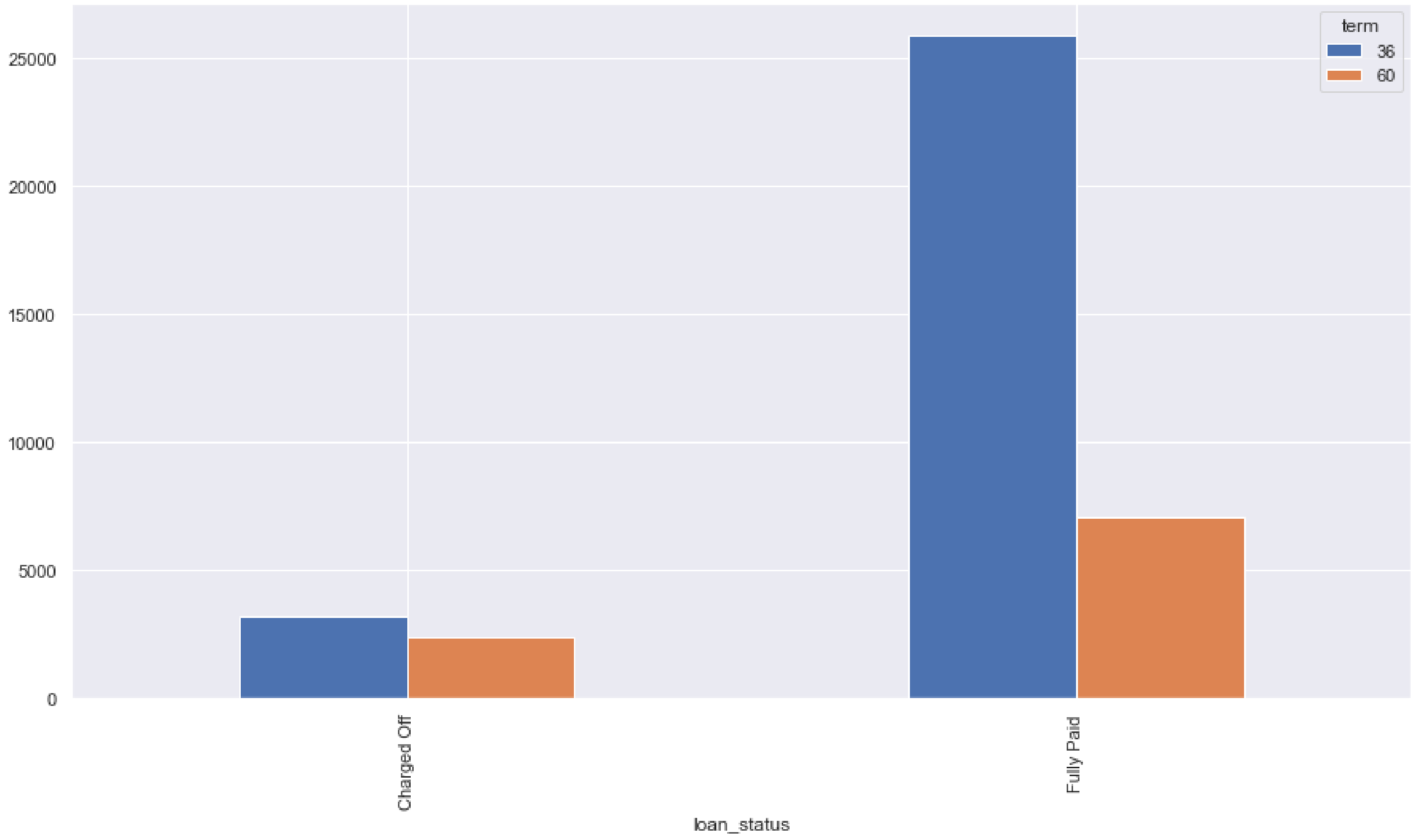
Most of the applicants employment length is either 10+ years and 1 year in both categories of Charged Off and Fully Paid

Segmented Univariate Analysis



Inference

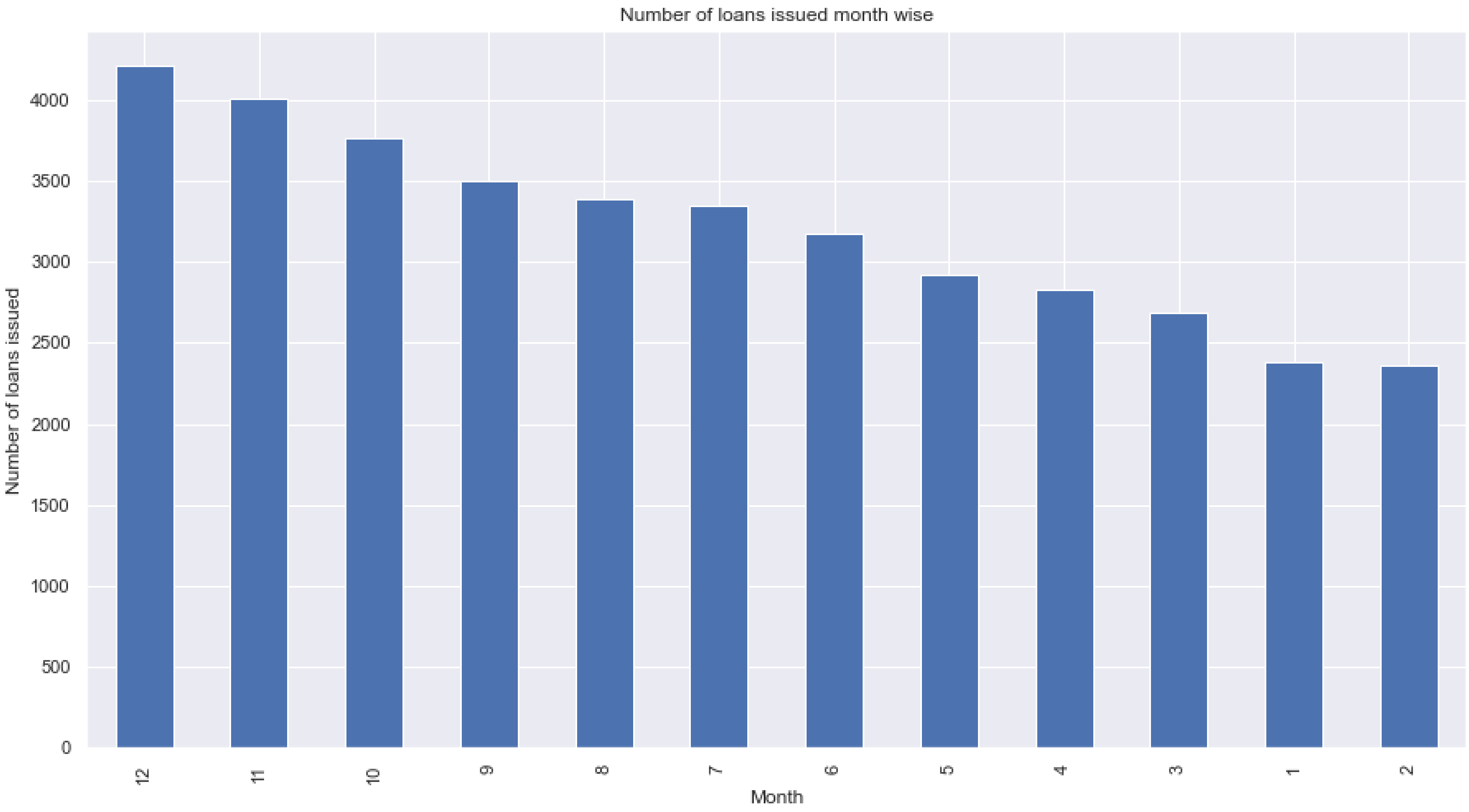
As the rate increases, the chances of Charged Off increases



Inference

36 months term is preferred over 60 months term by people

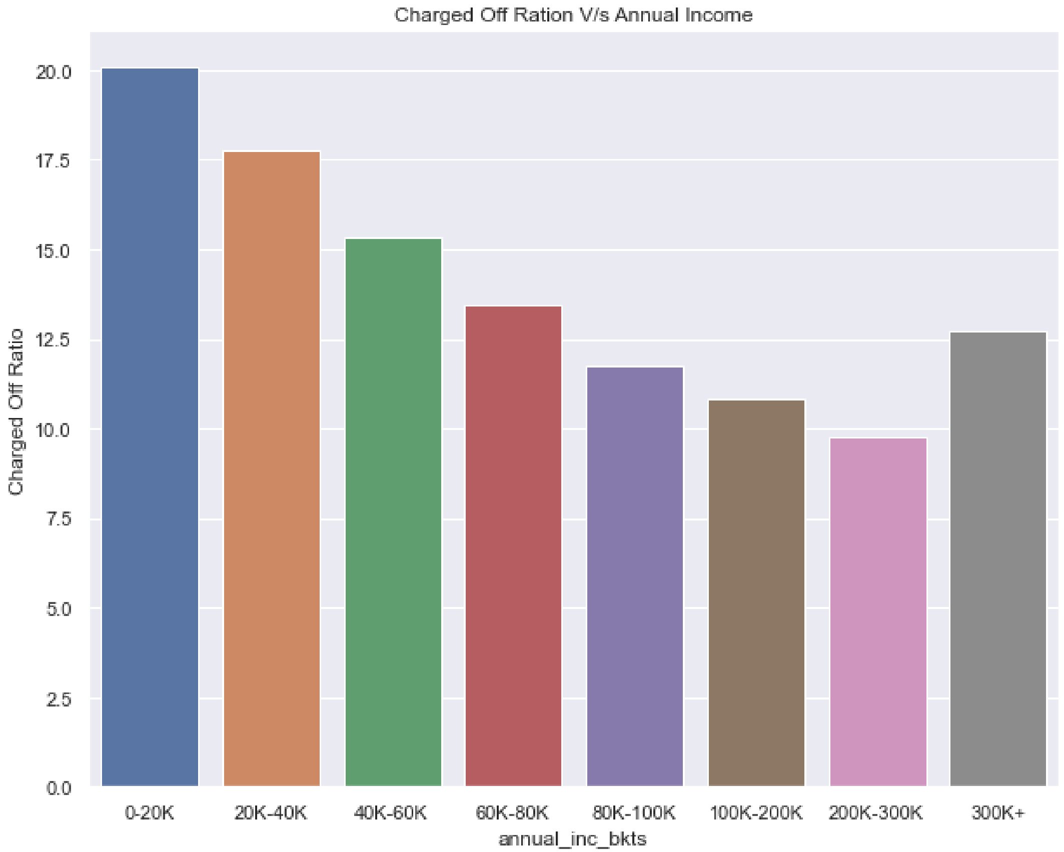
Segmented Univariate Analysis



Inference

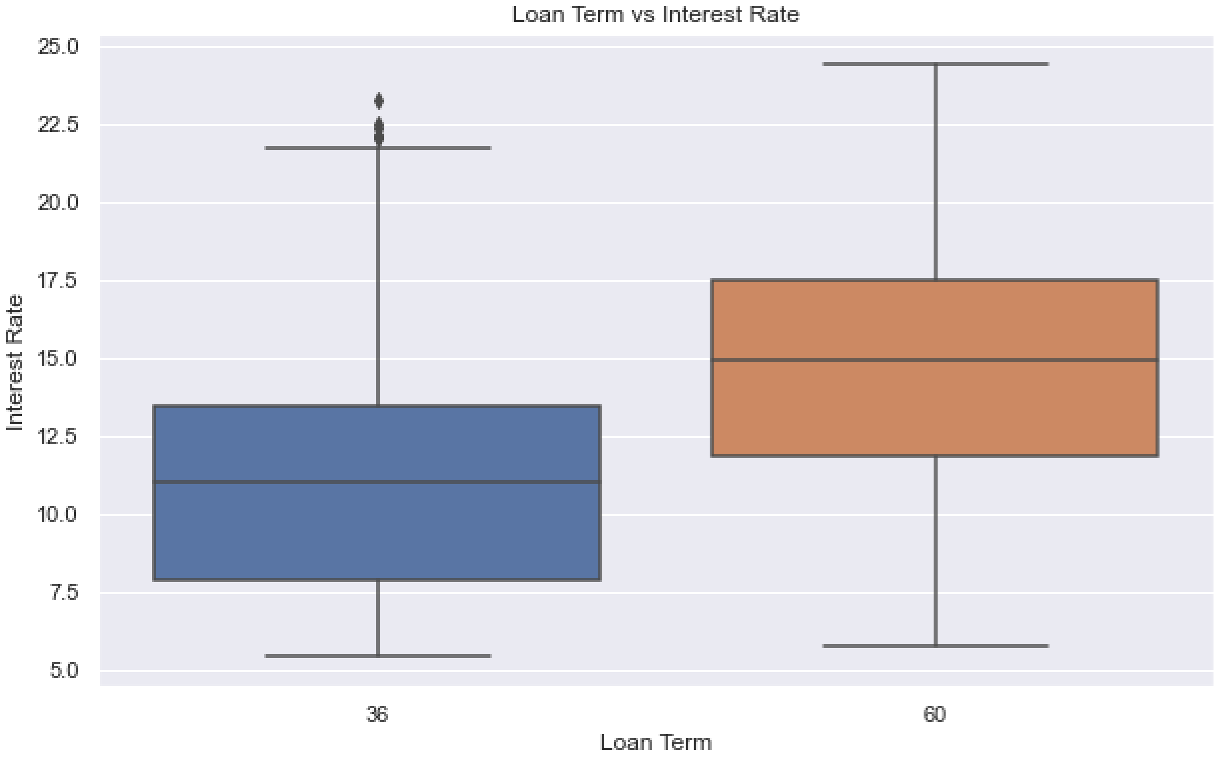
12th month - December month is when the highest number of loans are issued. This could be because of the festive season shopping and holiday season vacations

Bivariate Analysis



Inference

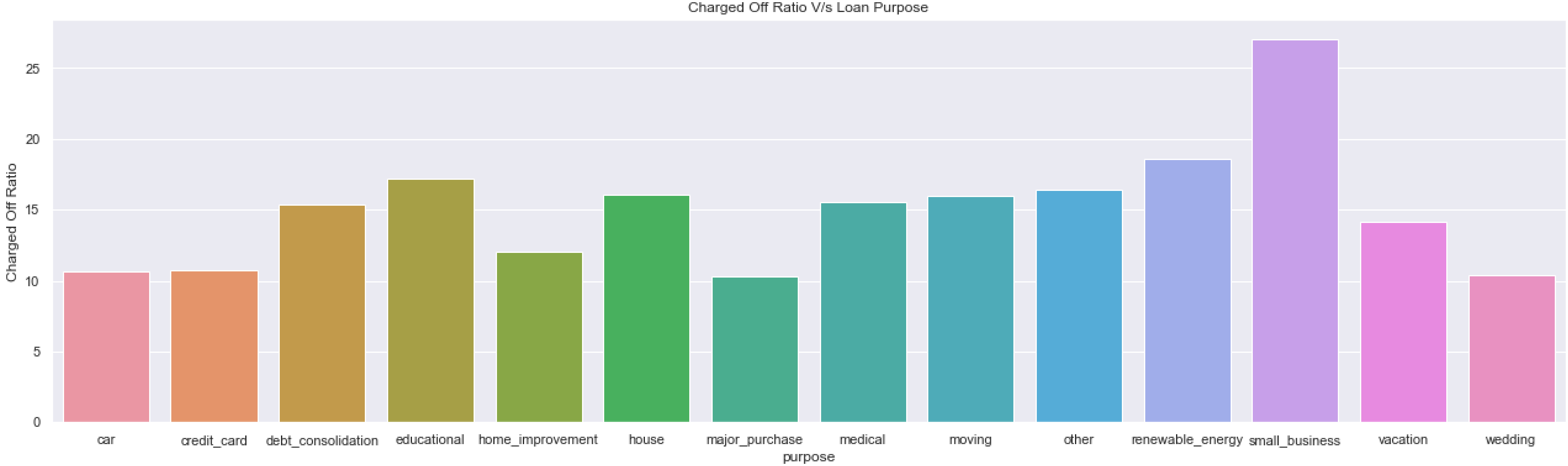
Charged Off ratio generally tends to decrease as the annual income range increases



Inference

Interest rates are generally higher for 60 months term loans than 36 months terms. This also validates the reason for preference for 36 months term loans over 60 months term loans

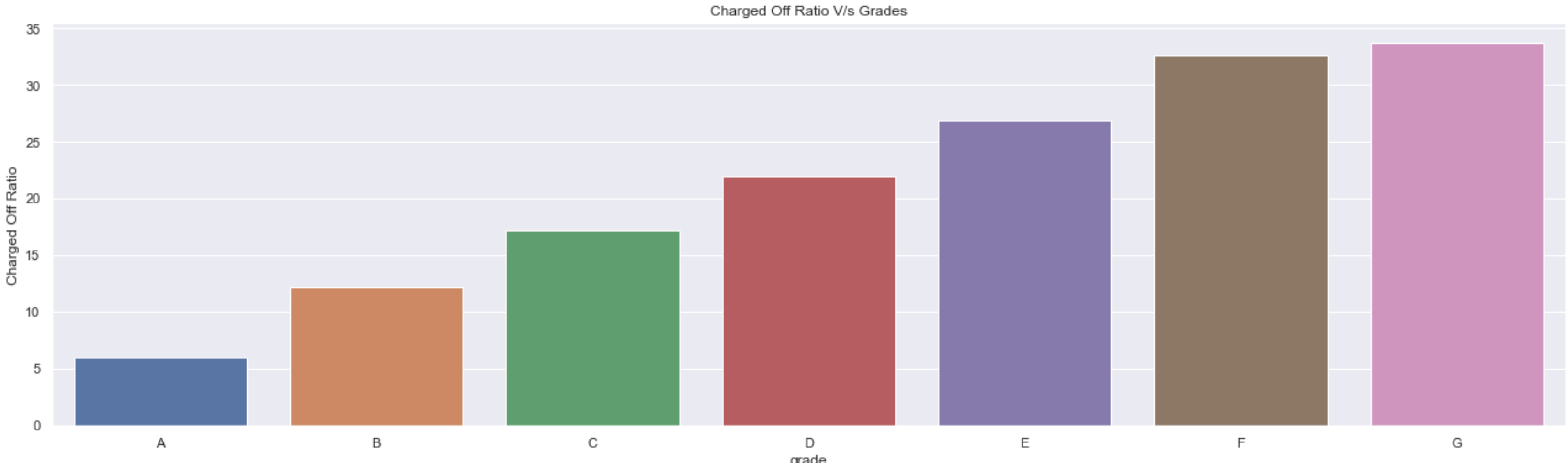
Bivariate Analysis

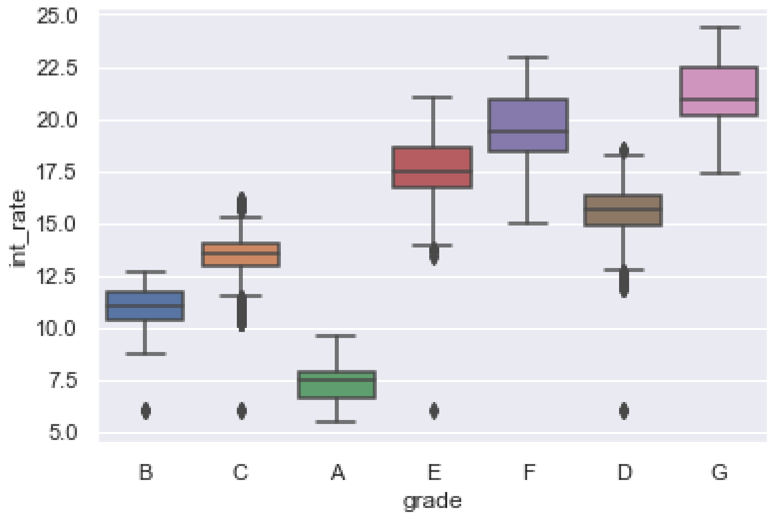


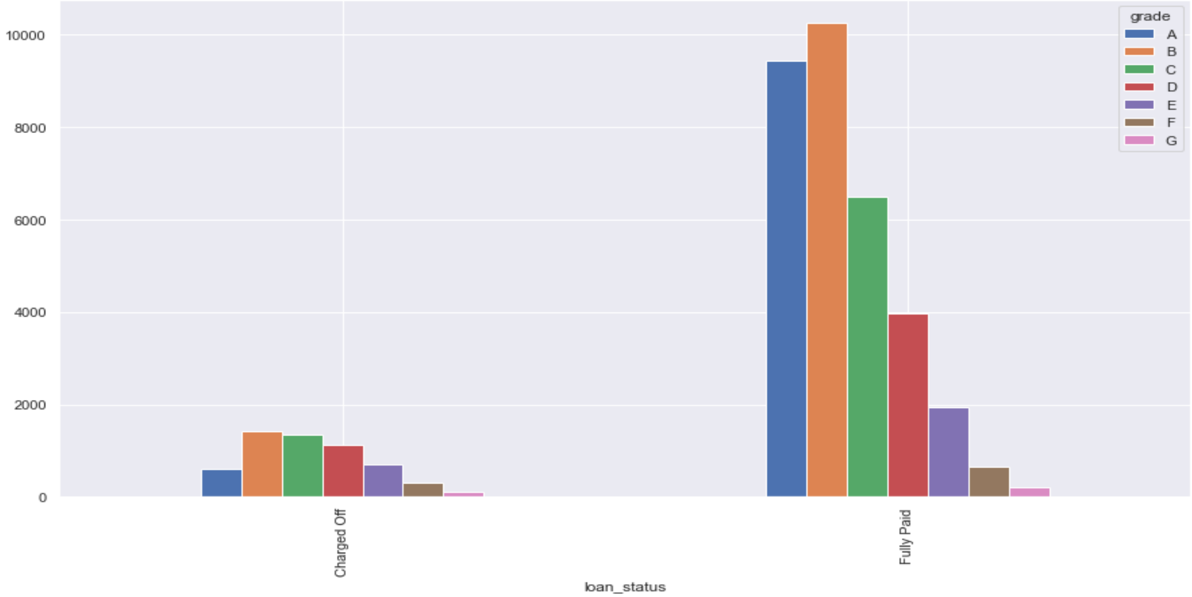
Inference

Charged Off ratio is highest for Small business loan purpose

Bivariate Analysis



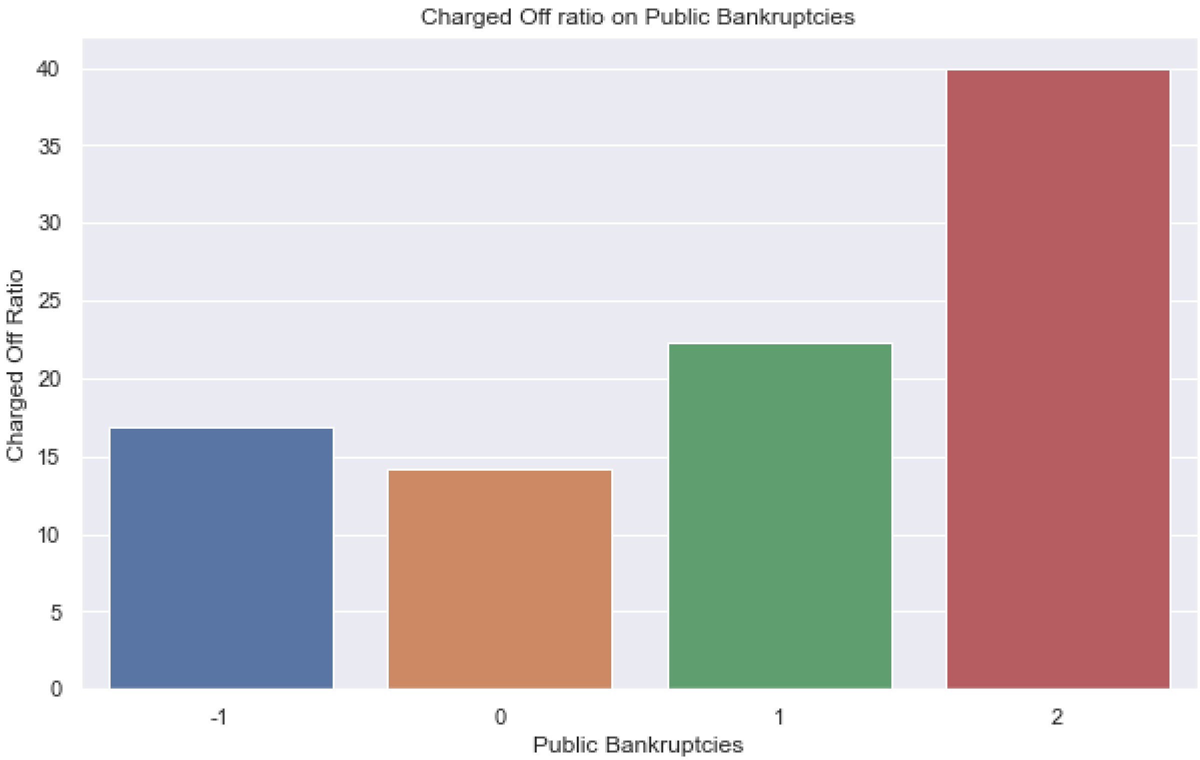


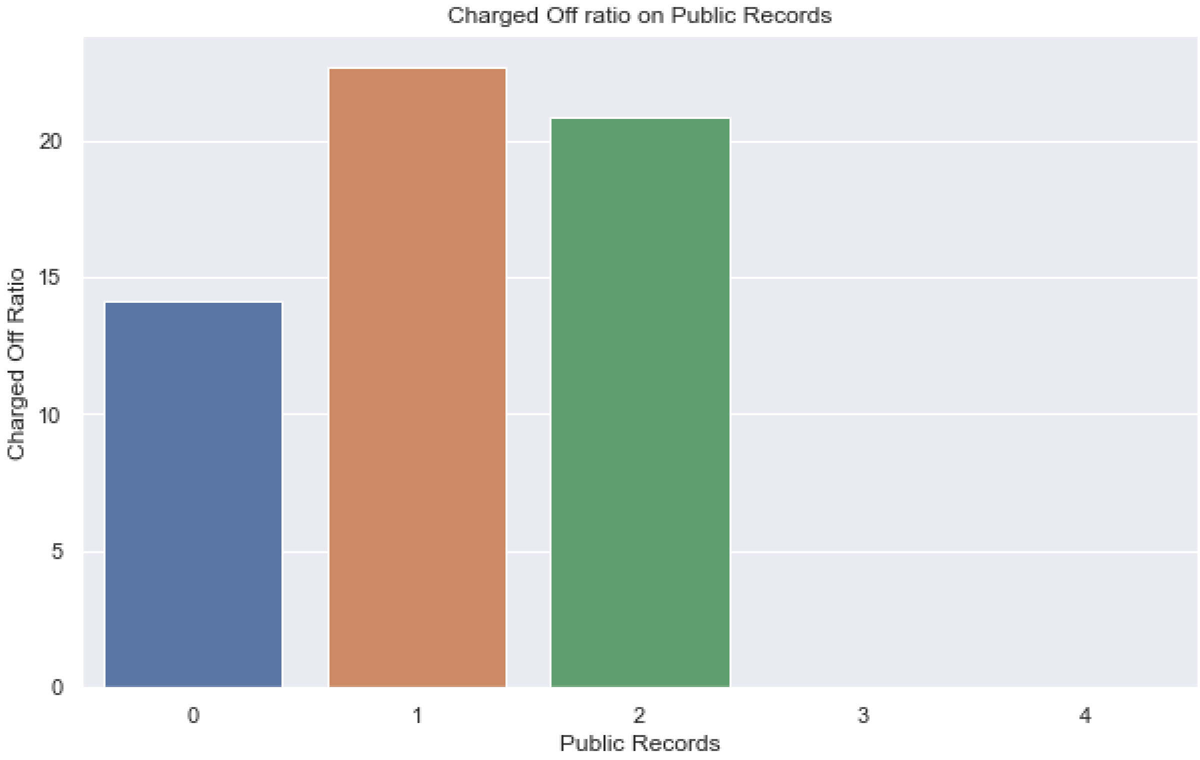


Inference

As the grades moves from A to G, the interest rates increases, and the charged off ratio also increases

Bivariate Analysis

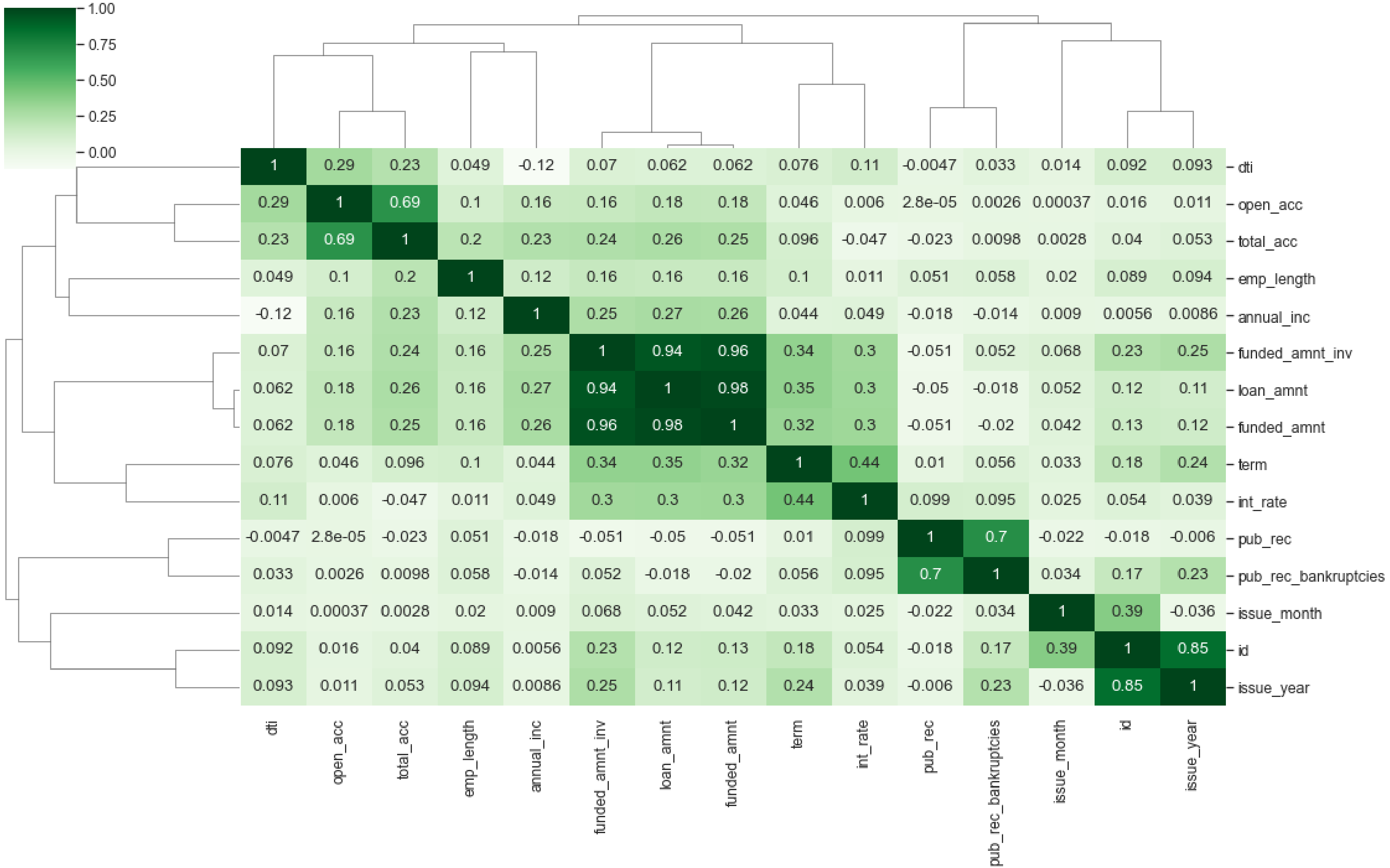




Inference

Charged Off ratio increases with increase in number of bankruptcies and Derogatory public records

Bivariate Analysis



Inference

Negative correlation between dti and annual income. This means if the annual income is more, the dti is less

Recommendations

Driving Variables	Recommendation
Annual Income	It is highly risky to lend loans to applicants in the 0-20K annual income range
Interest Rates	As the interest rates increases, the chances of charge off increases, hence the company should formulate more loan plans with lesser interest rates
Purpose for loan	Small Business Purpose loan lenders show high tendency for defaulting and hence it should be considered as risky category
Loan Term	Short term loans should be encouraged as such loans have lesser interest rates which ensures the loan is Fully Paid
Grade	Grade A loans should be encouraged as such loans have lesser interest rates which ensures the loan is Fully Paid. The order of preference should be Grade A, B, C, D, E, F and G
Public Bankruptcies and Public Derogatory	It's highly risky to lend loans to applicants who has previous bankruptcy record history or public derogatory records. The more the number of records, the higher the risk