Lending Club Case Study

Issac Paul

M.Sc AI ML

Problem Statement

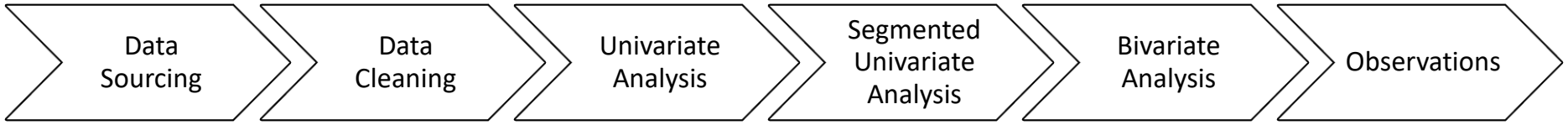
The Lending company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Identification of these risky loan applicants can reduce such loans, thereby cutting down the amount of credit loss.

If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.

Identify Loan applicants that have high probability of loan default using Exploratory Data Analysis

Analysis Approach

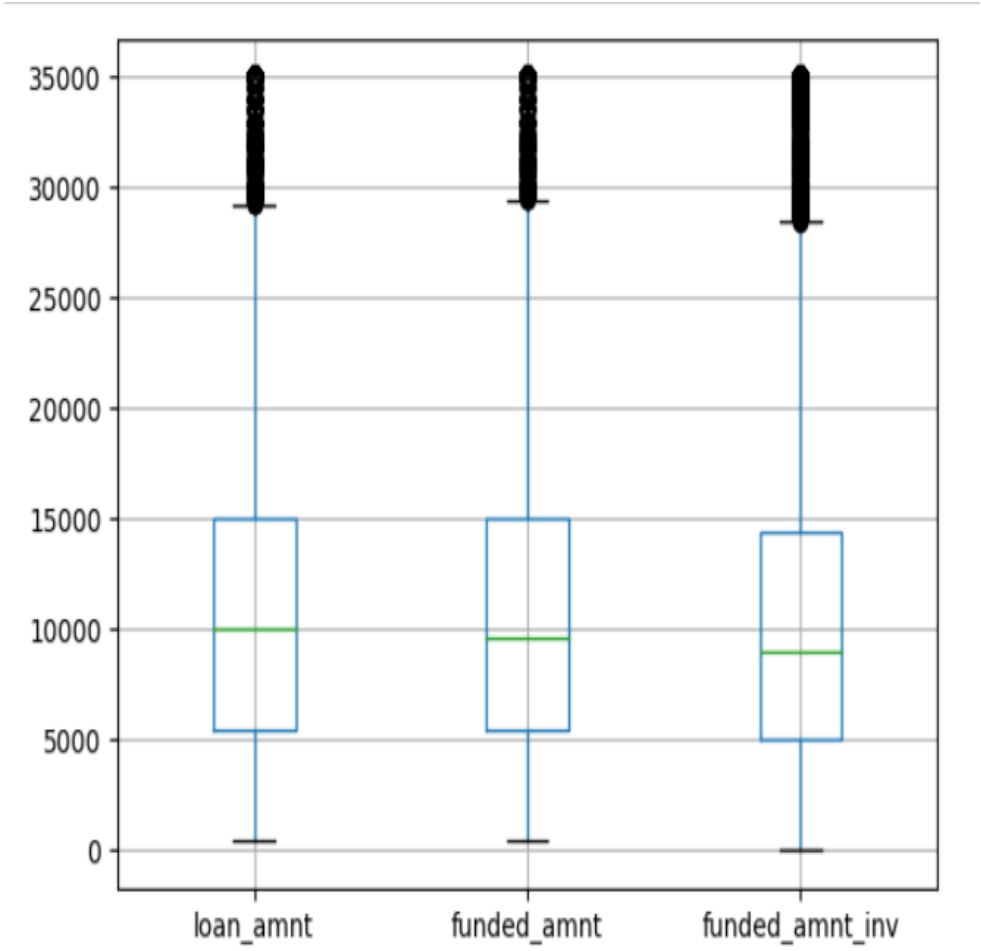


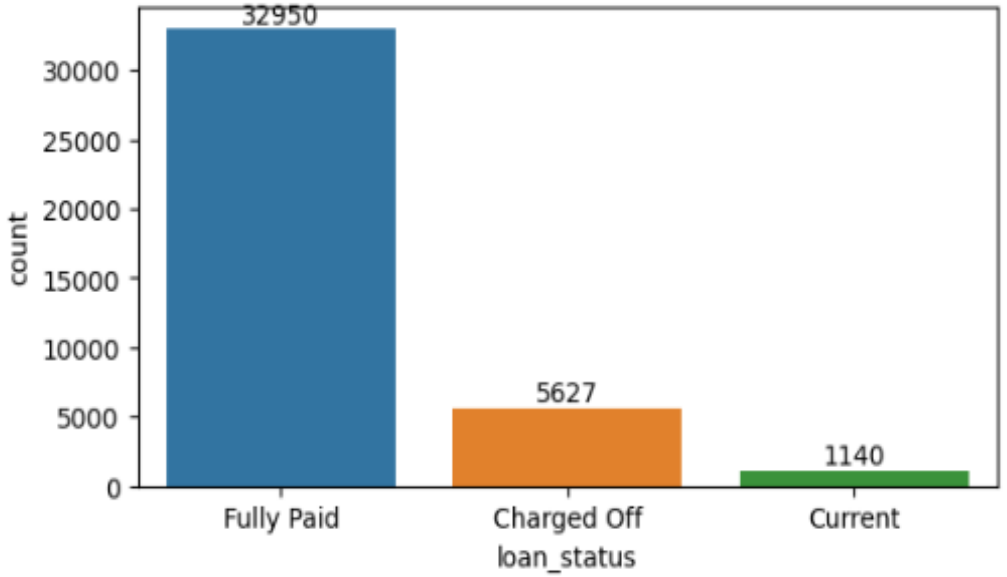
Univariate Analysis on all Loans

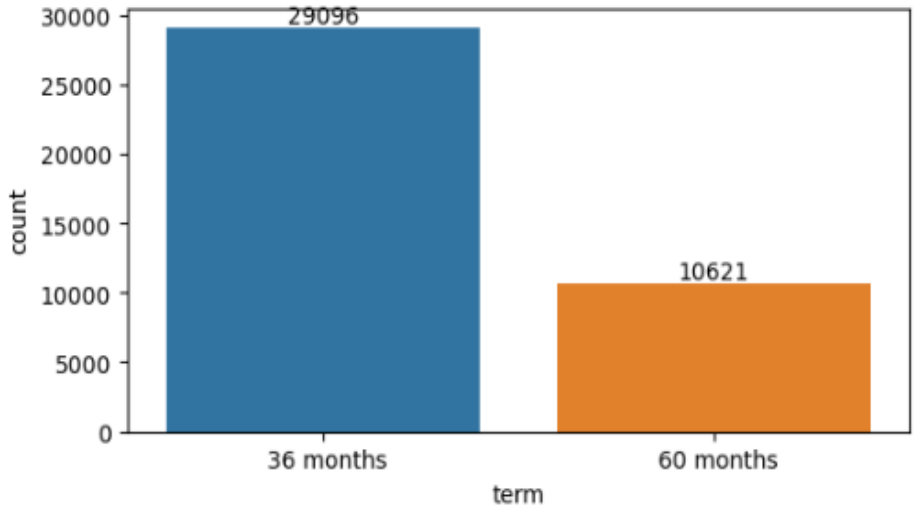
Carrying out univariate analysis on below variables by taking total count to study overall spread of loans in terms of

- Loan Status
- Term of Loan
- Interest Rate range
- Loan Amount Range
- Annual Income Range
- Installment Range
- Home Ownership
- Purpose of Loan

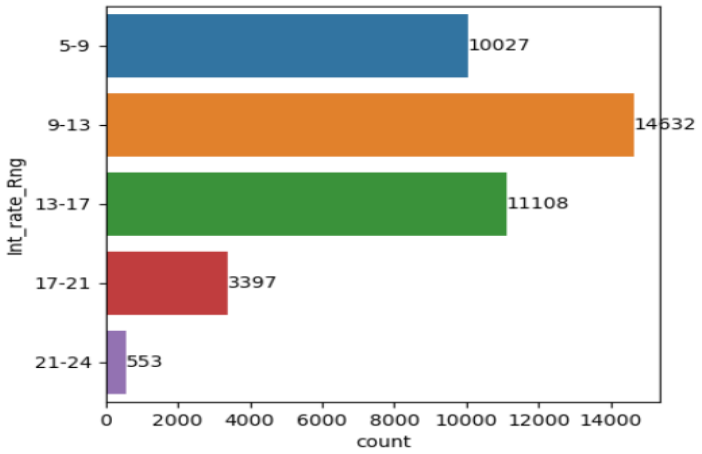
Univariate Analysis on all Loans

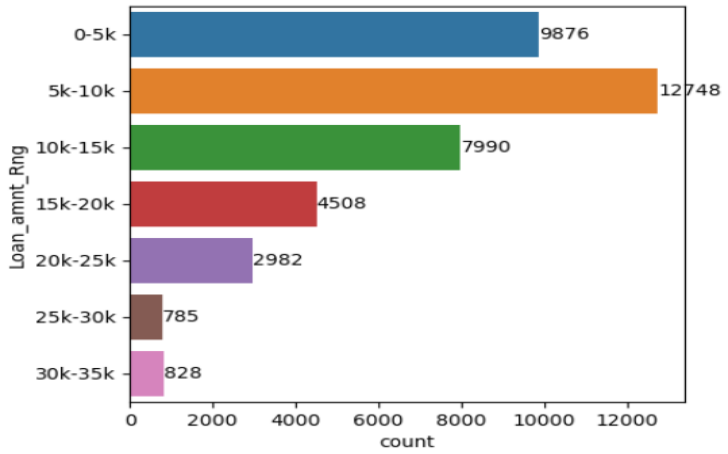


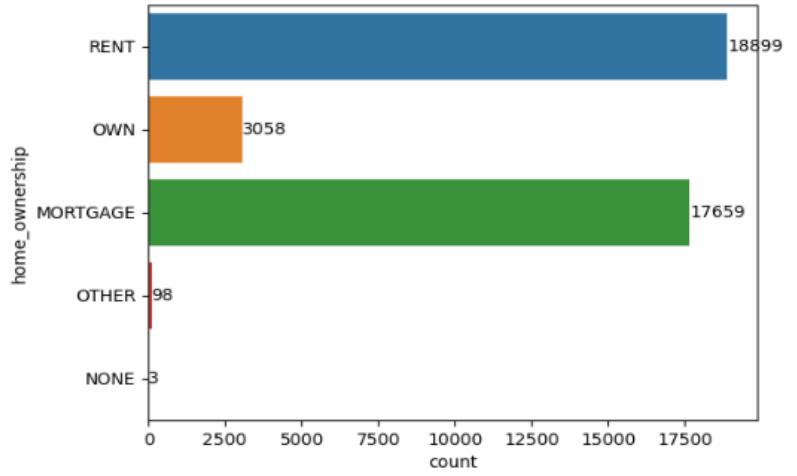


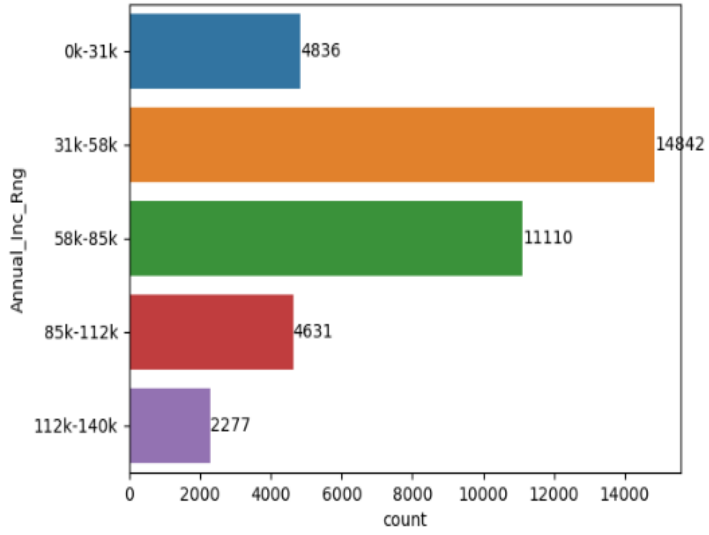


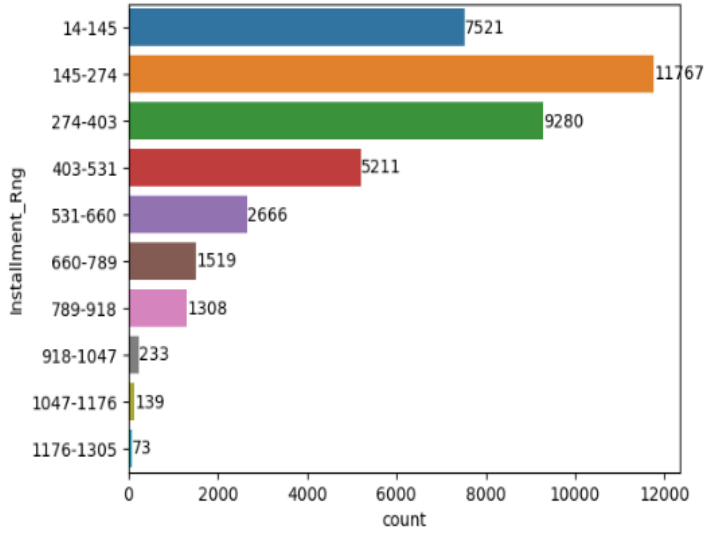
Univariate Analysis on all Loans

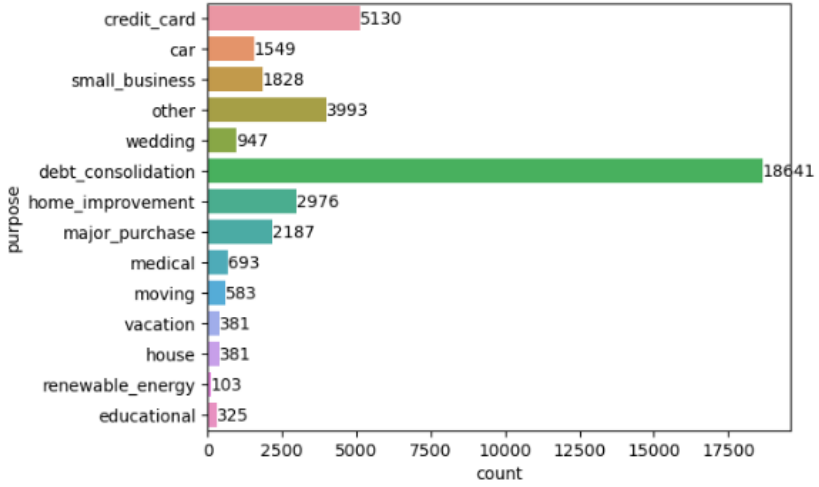












Observations on overall loan characteristics

•Out of **39717** Loans **32950** have been fully paid and **5627** have been charged off. The current **1140** loans will be excluded in the next section where bivariate analysis will be performed as they do not contribute to the analysis since they haven't defaulted yet.

•The highest no of loans are offered in the **36** months tenure

•The highest no of loans are offered in the **(9-13) %** interst rate

•The highest no of loans are disbursed in the range of **5k - 10k**

•The highest no of loans are given to applicants in the annual Income bracket of **31k -58k**

•The highest monthly installments are between **145-274**

•Most of the loans are taken by users who are staying on rent

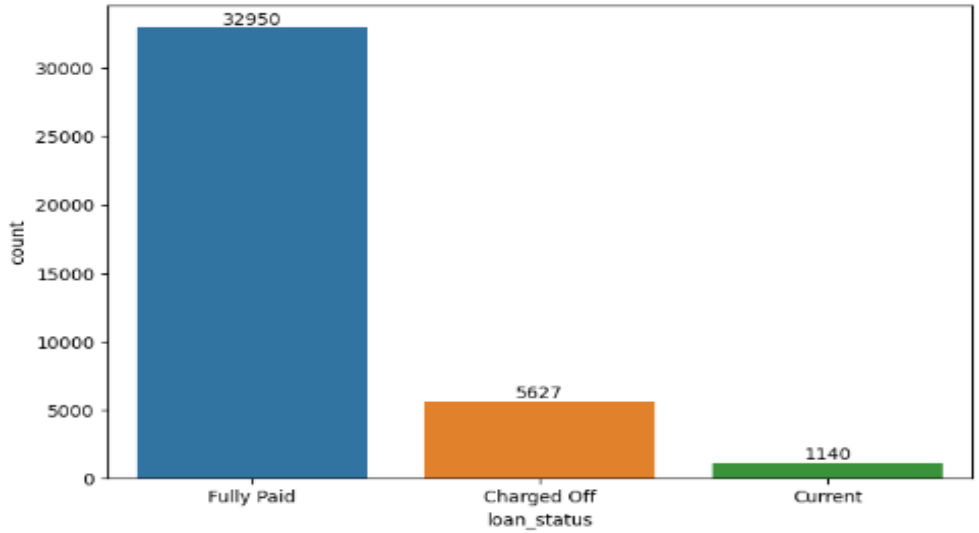
•Most of the loans are taken to repay existing debts

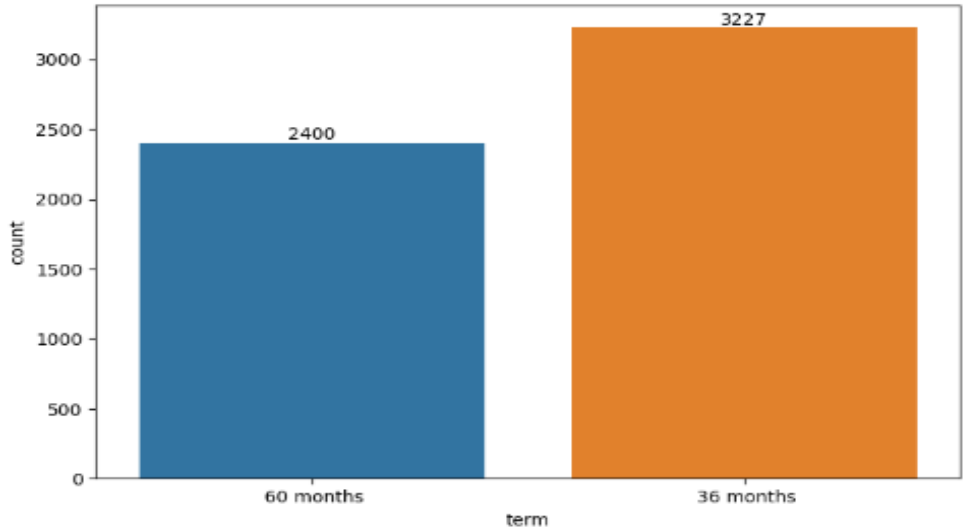
Univariate Analysis on Charged Off loans

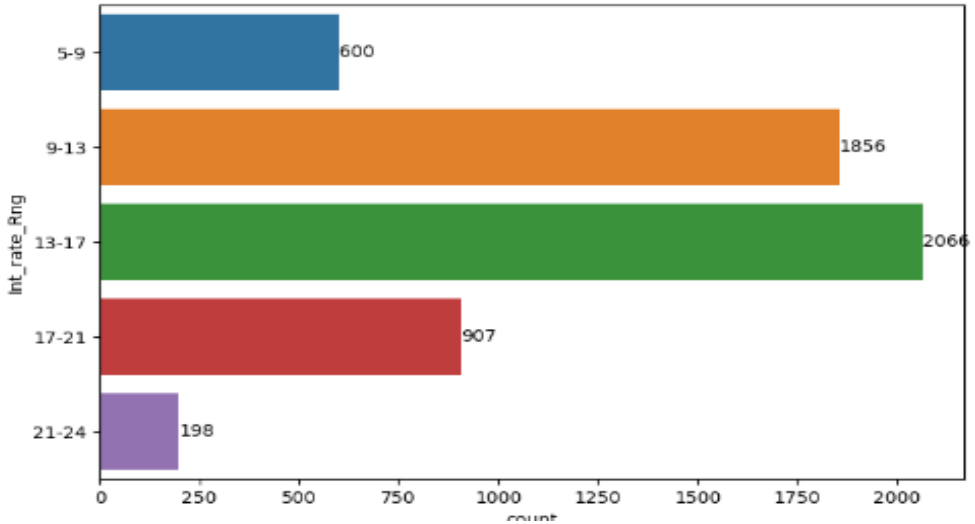
Carrying out univariate analysis on below variables by taking total count to study overall spread of loans in terms of

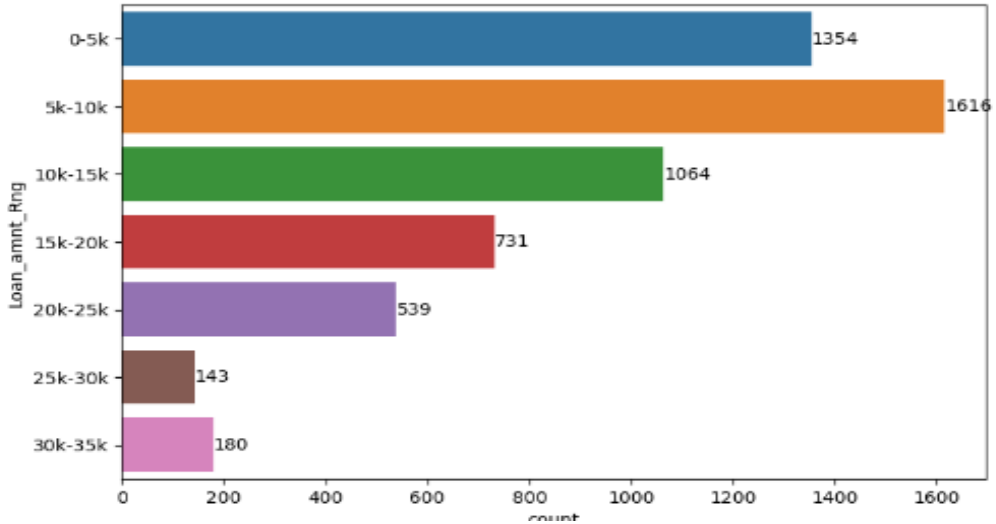
- Loan Status
- Term of Loan
- Interest Rate range
- Loan Amount Range
- Annual Income Range
- Installment Range
- Home Ownership
- Purpose of Loan
- Address State
- Revol_Util
- Inquiries in last 6 months
- Date of issue
- Grade/Subgrade
- Employee Length
- Verification Status
- Open Account
- Total Account

Univariate Analysis on Charged Off loans

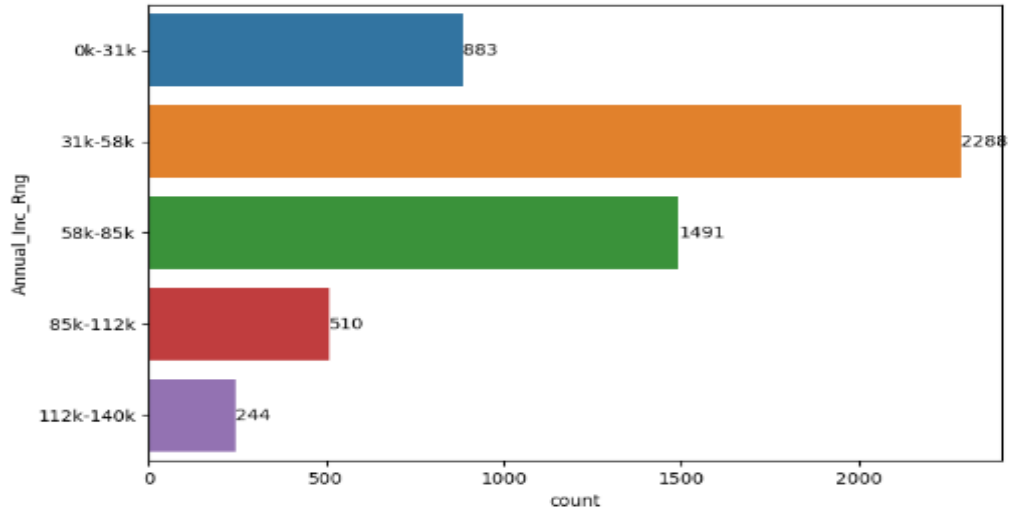


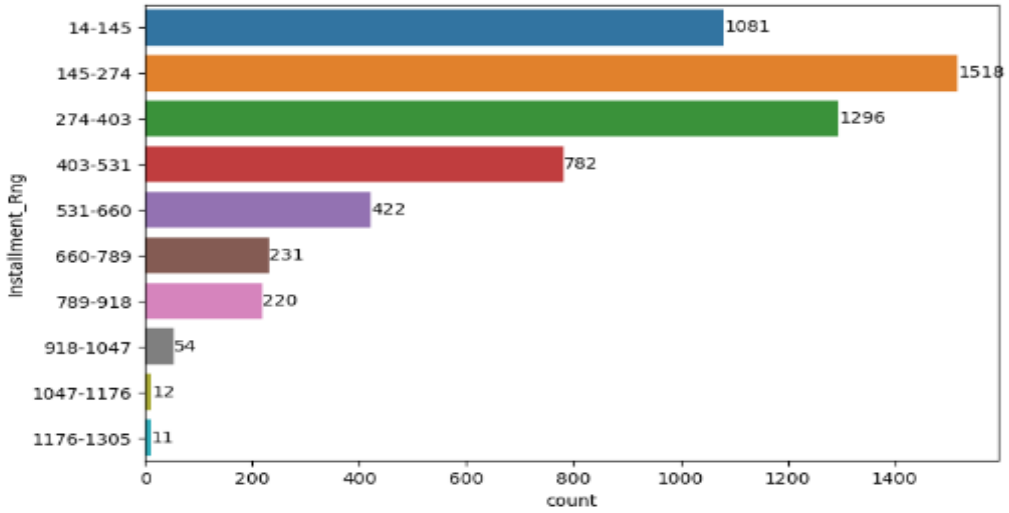


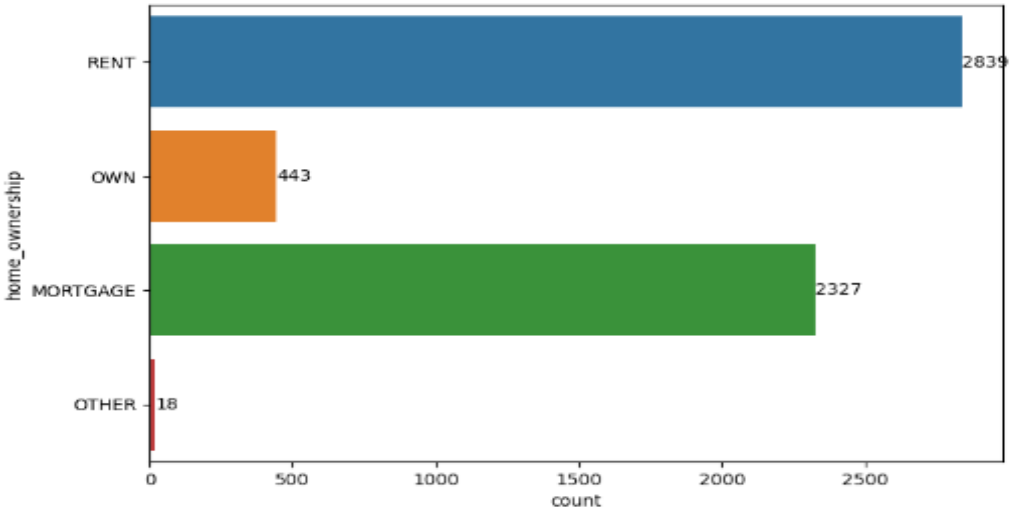


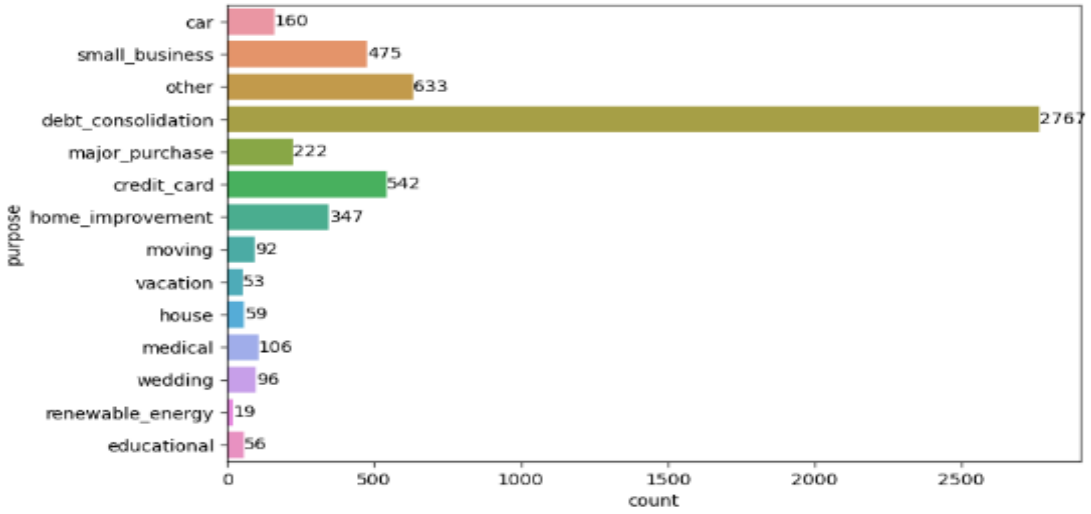


Univariate Analysis on Charged off Loans

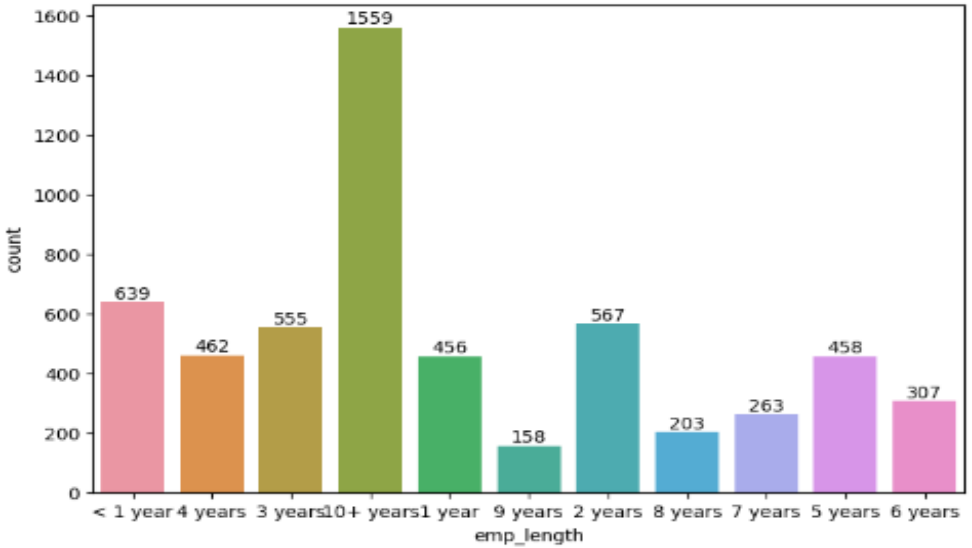


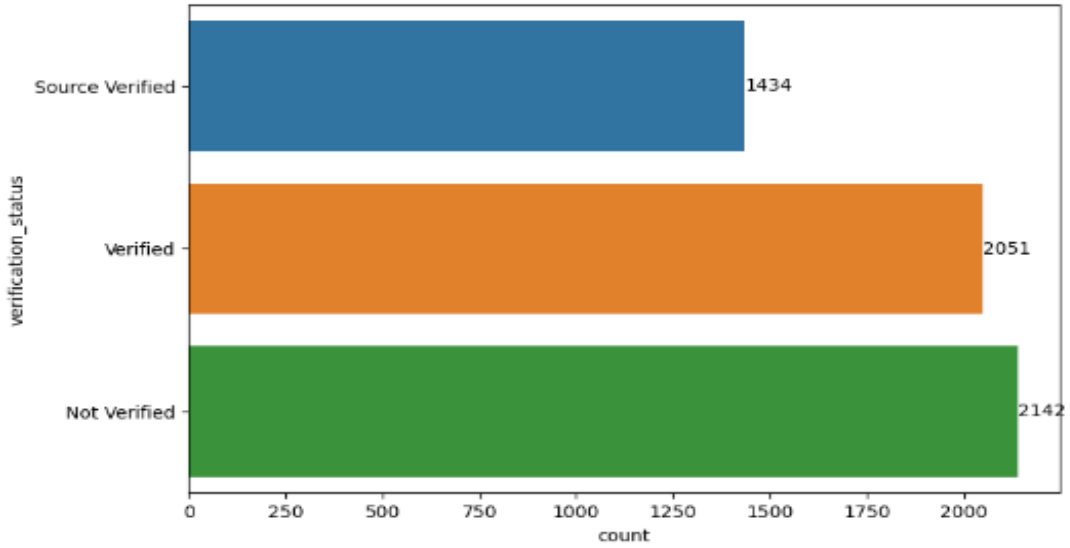


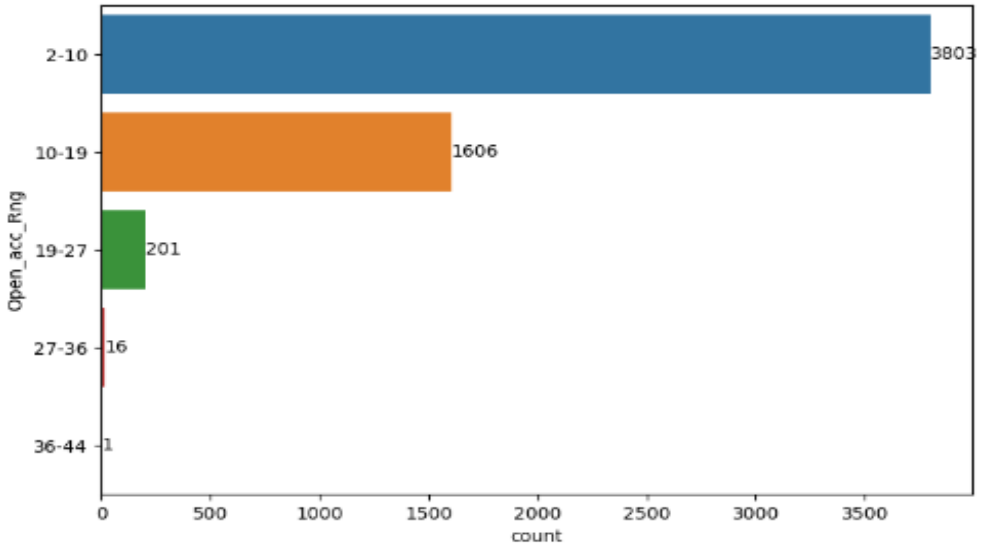


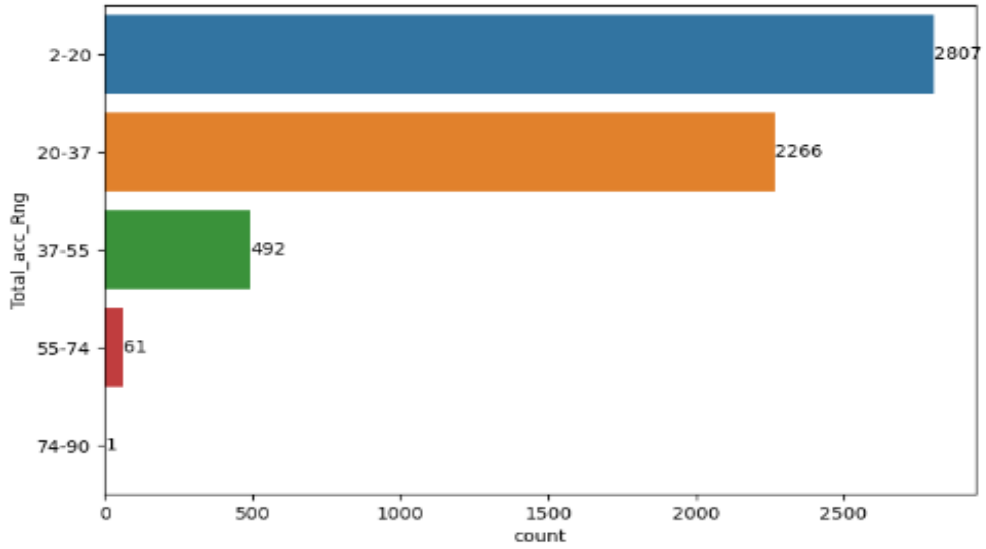


Univariate Analysis on Charged off Loans

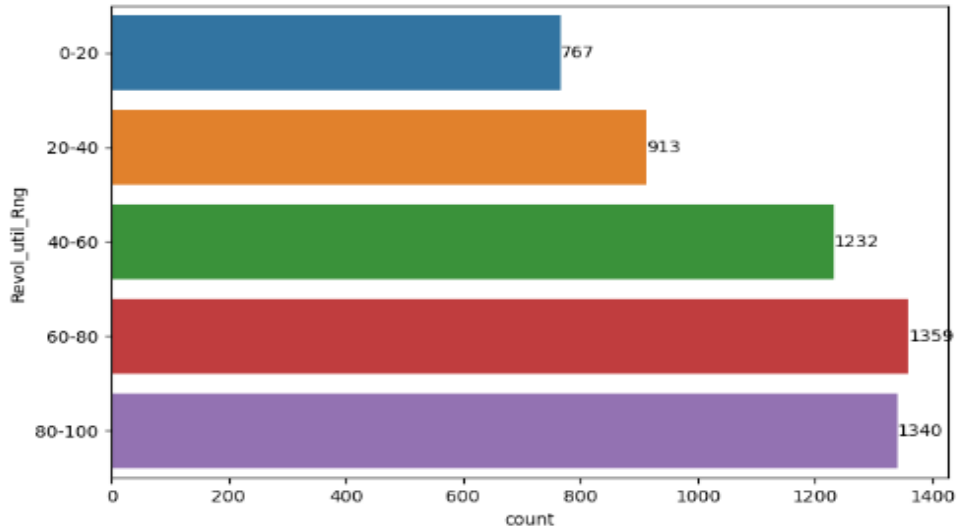


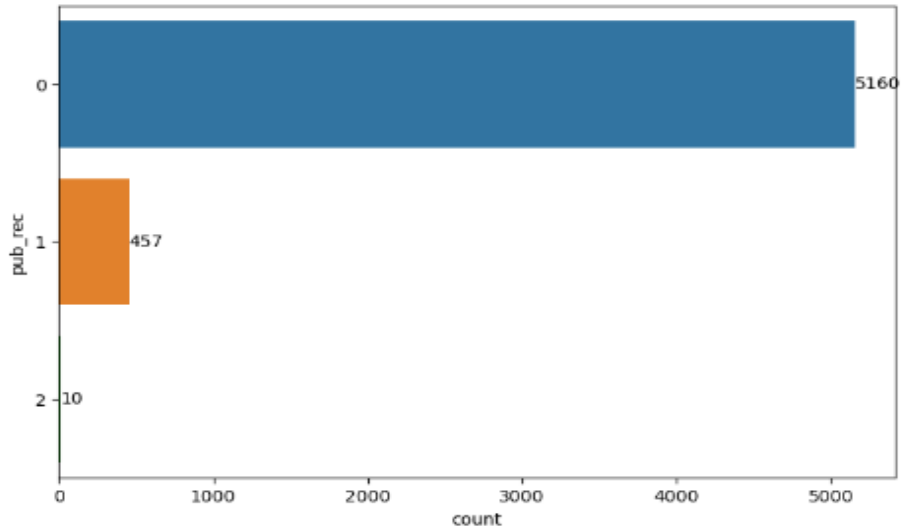


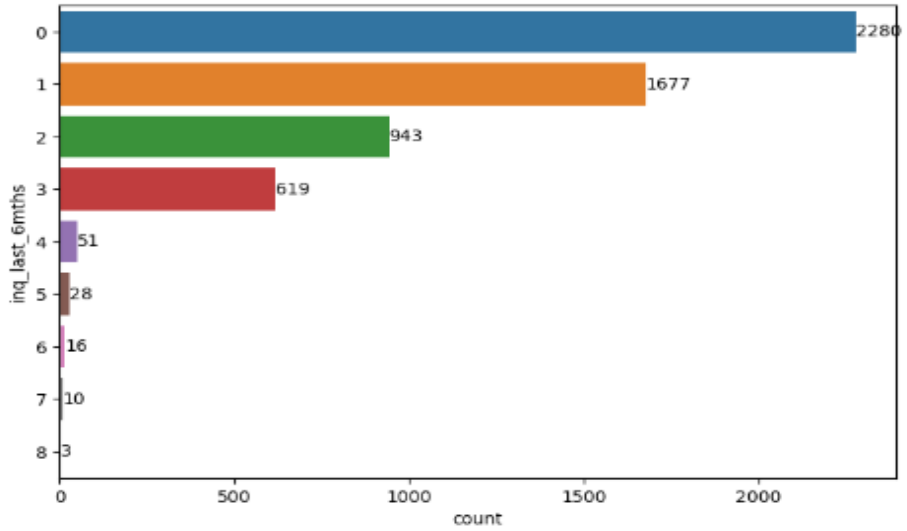


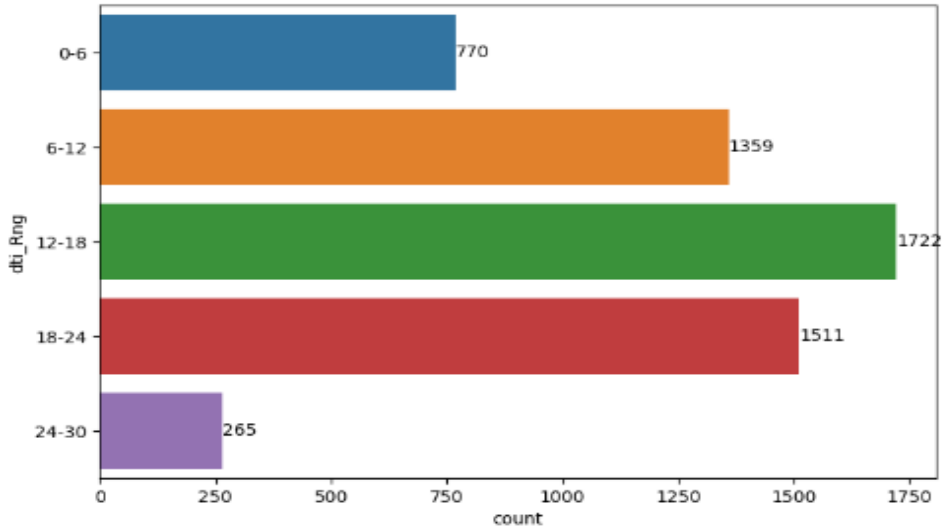


Univariate Analysis on Charged off Loans

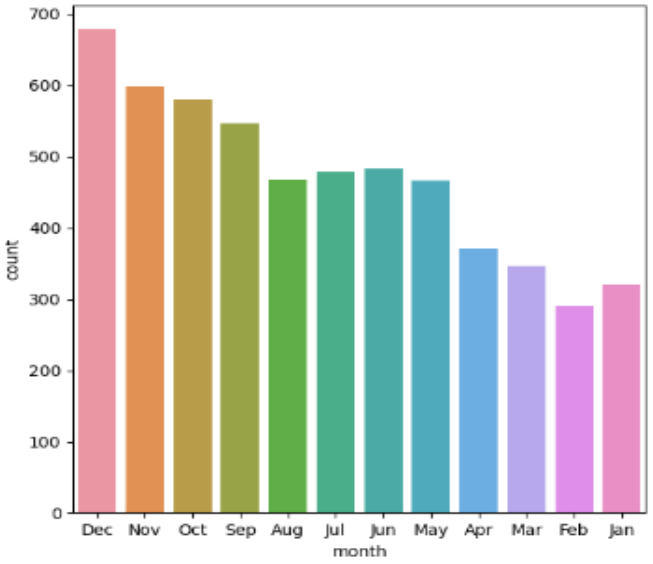


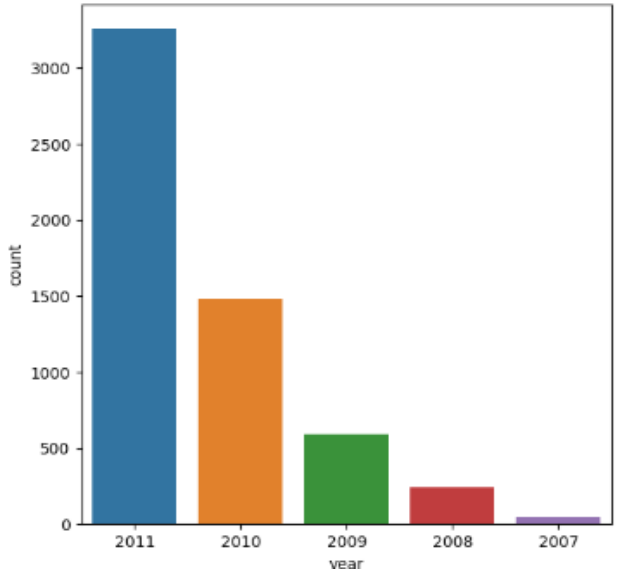


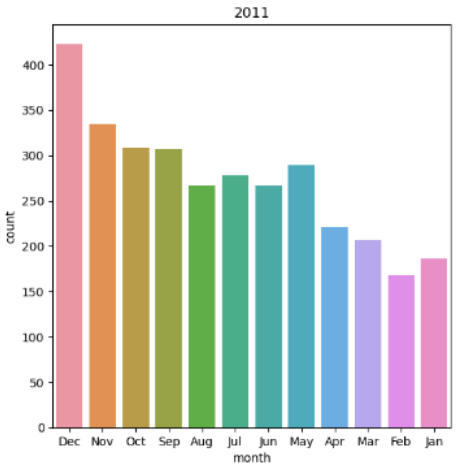


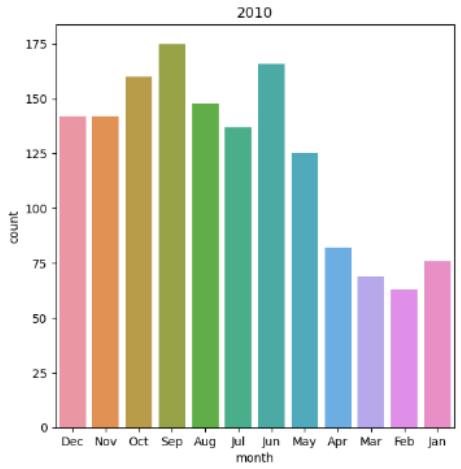


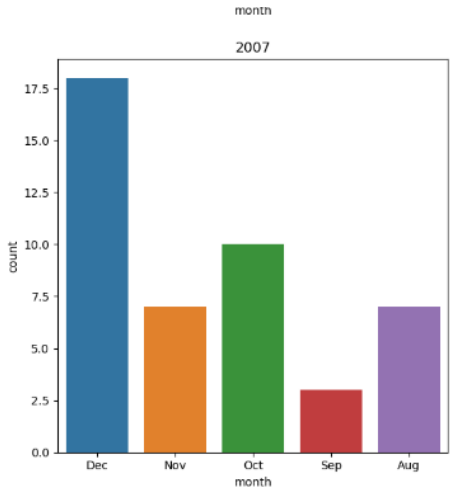
Univariate Analysis on Charged off Loans by Studying Loan Default Month wise for Every year in the range

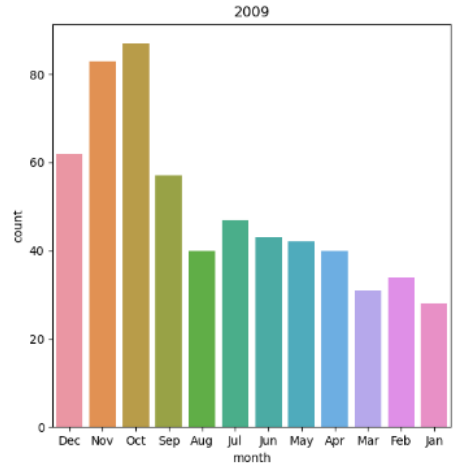


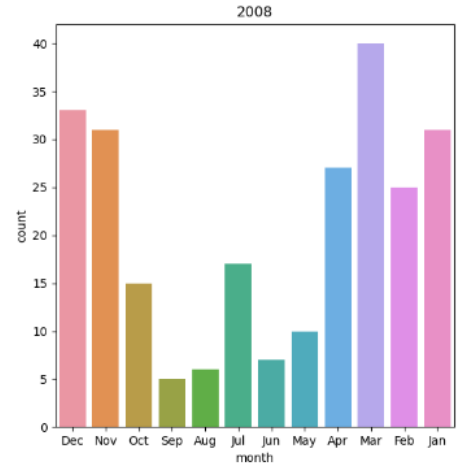




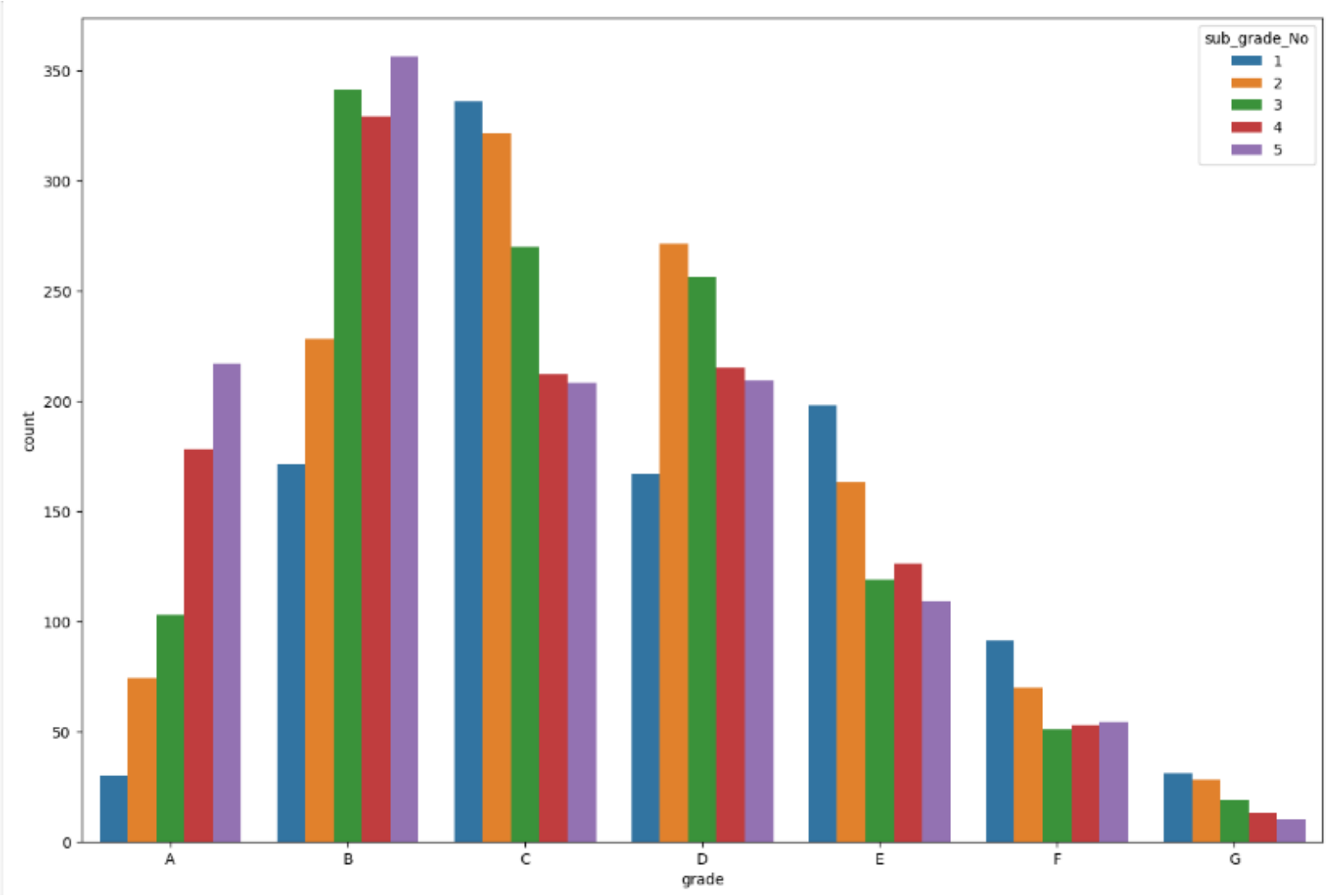




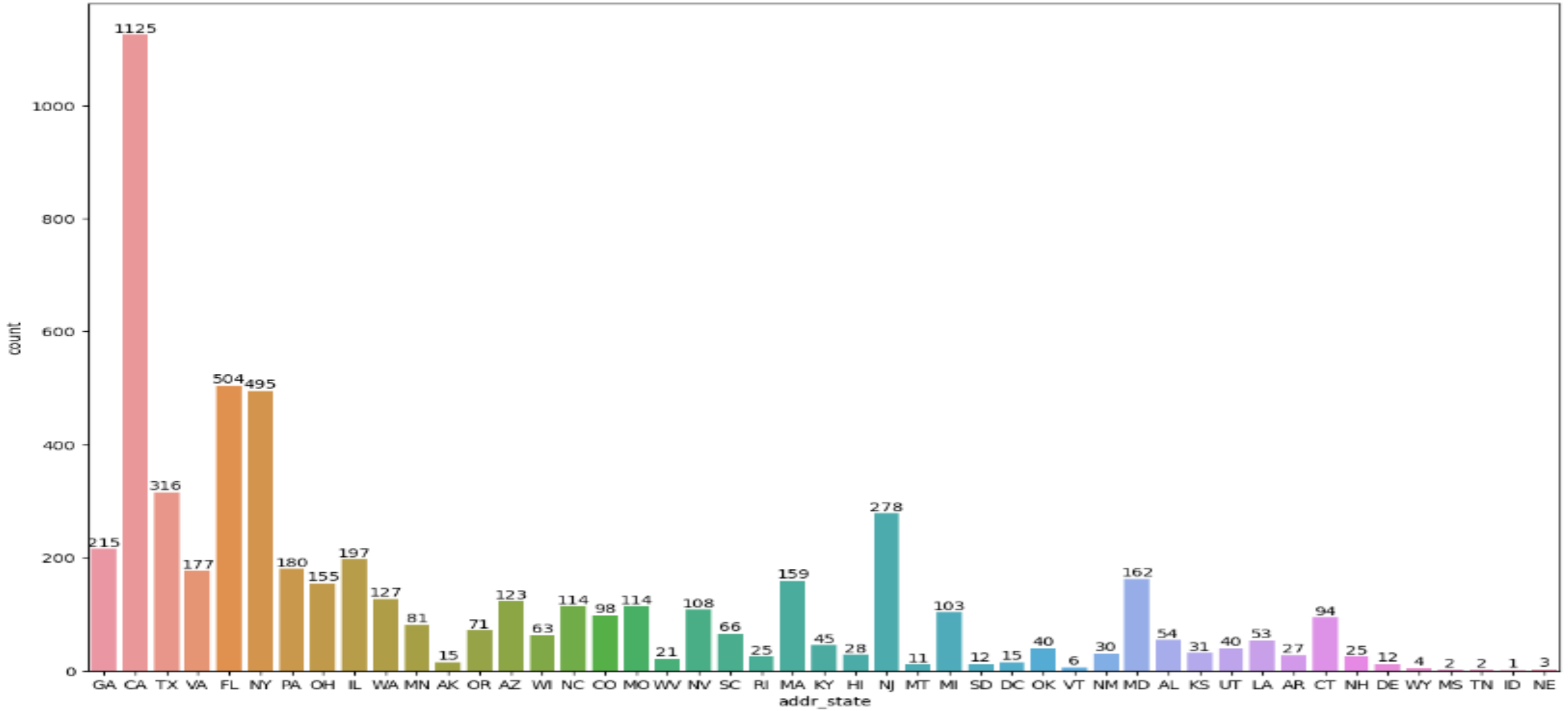




Univariate Analysis on Charged off Loans based on grade and Sub Grade



Univariate Analysis on Charged off Loans based on state



Conclusions on the variables which are strong indicators of default based on bivariate Analysis

By running univariate Analysis on the 5627 Charged off loans. The chances of the Application defaulting is higher if

- The loan is offered to the applicant for **36** months of tenure
- The loans is offered at an interest rate of **(13-17) %** .
- The loan amount disbursed is in the range of **5000 - 10000**
- The loans is offered to an Applicant in the **31000 -58000** Income group
- The monthly installments is between **145-274**
- The loan is given to an applicant who is staying on **rent**
- The loan is taken by applicant to repay existing debts(**debt consolidation**)
- The loan is taken by an Applicant who have **2-10** open credit lines in the borrower's credit file.
- The loan is taken by an Applicant who has **2-20** credit lines(open/closed) in the borrower's credit file.
- The loan is taken by an Applicant who has an employment of **10+** years
- The loan is sanctioned to an Applicant with a Dti ratio is between **12-18**
- The loan is given to an Applicant with a loan status as **Not Verified**
- The loan is given to an Applicant with Grade **'B'**
- The loan is given to an Applicant with s Sub Grade OF '**B5'** .
- The highest default is seen in **CA** state

Observations on variables that had no impact on the default pattern

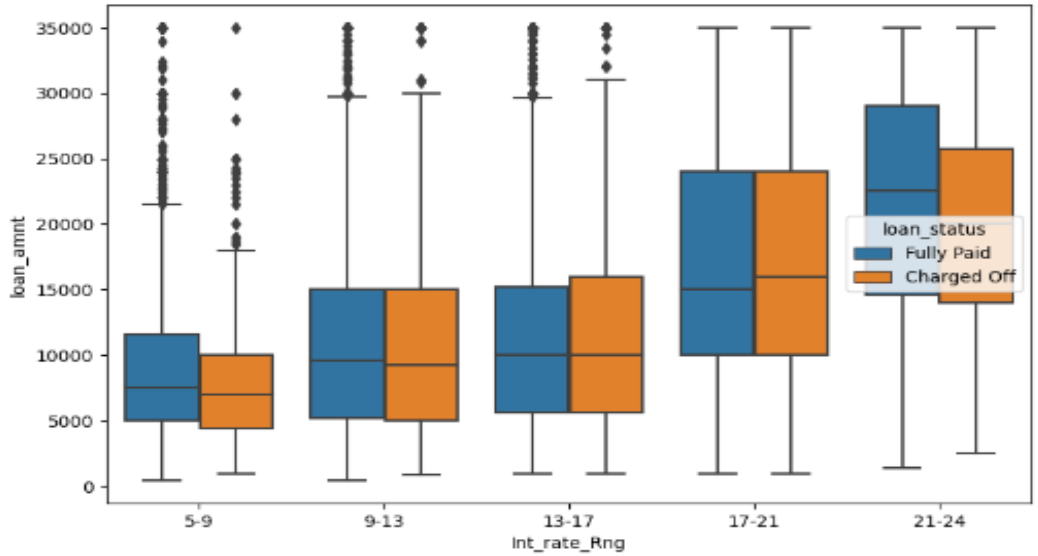
- Even though Applicant who has made 0 enquiries in last 6 months are defaulting on loans this is not a great measure to analyse the defaulting probability
- Even though Applicant who has 0 derogatory public records are defaulting on loans this is not a great measure to analyse the defaulting probability
- No conclusion can be generated from the month in which the loan has been granted because loan disbursements every year has a different default pattern for every month and there is no general trend followed

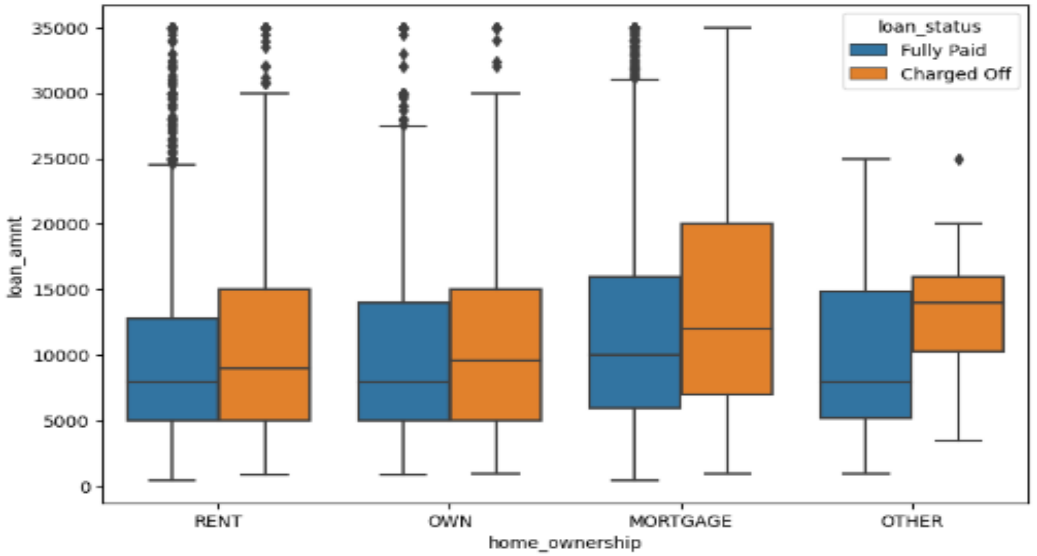
Bivariate Analysis on Charged Off loans(loan Amount)

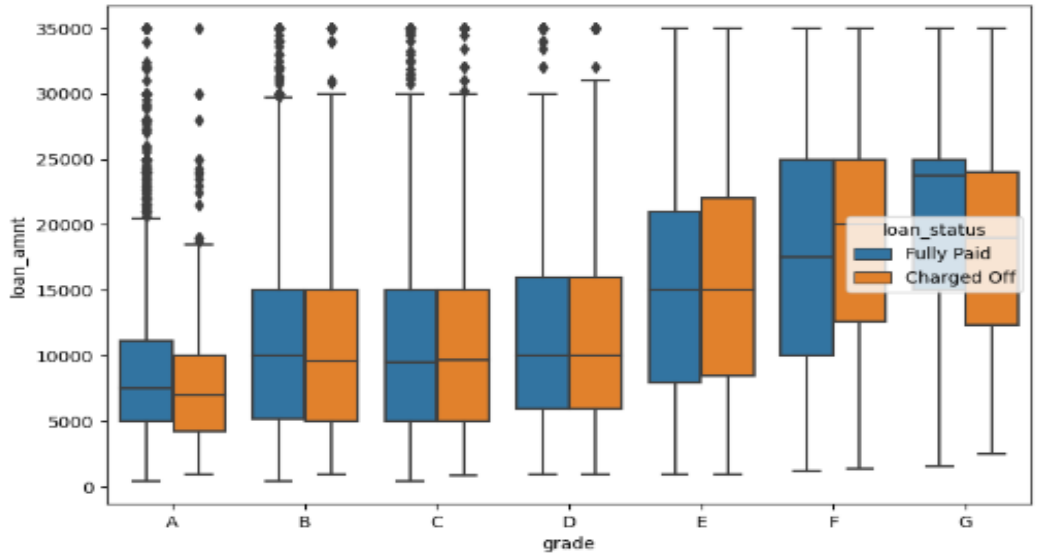
Bivariate analysis is performed for loan Amount against

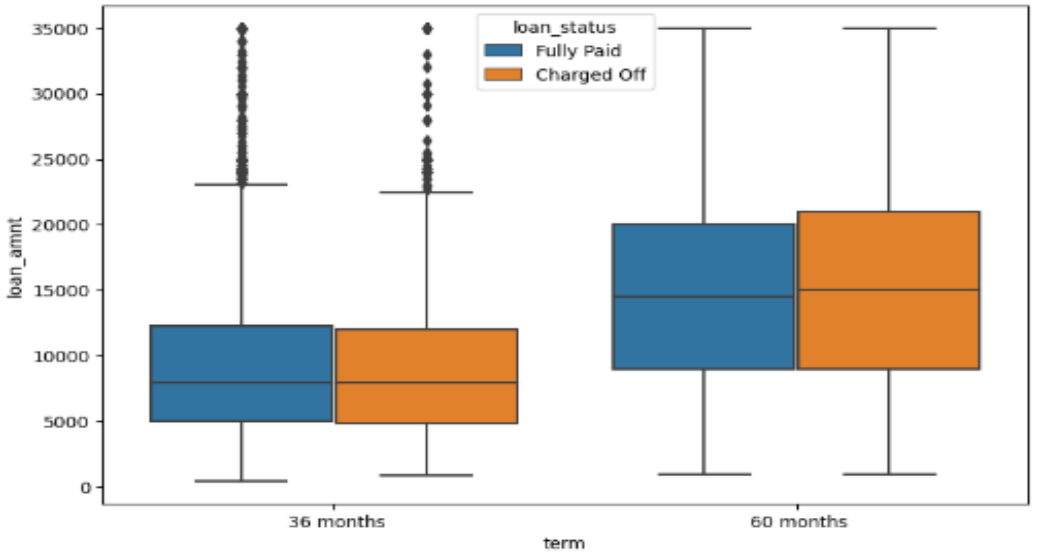
- Interest Rate
- Home Ownership
- Grade
- Term
- Emp Length
- Verification Status
- Tenure
- Annual Income
- Purpose

Bivariate Analysis on Charged Off loans(loan Amount)

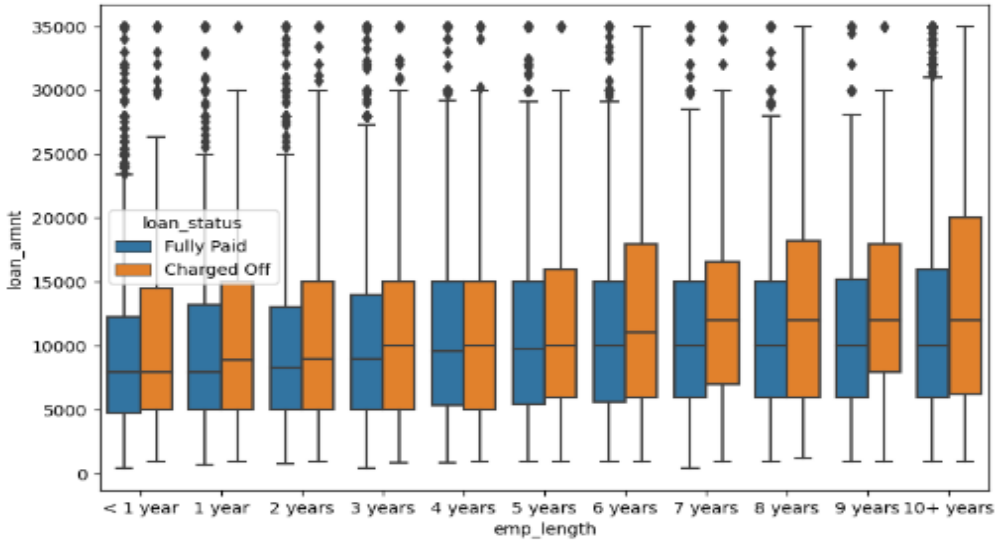


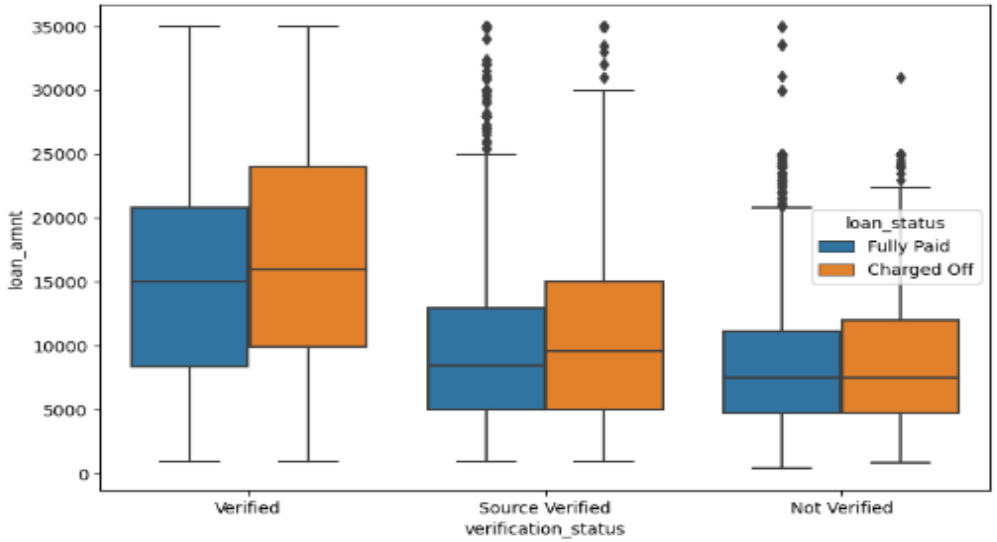


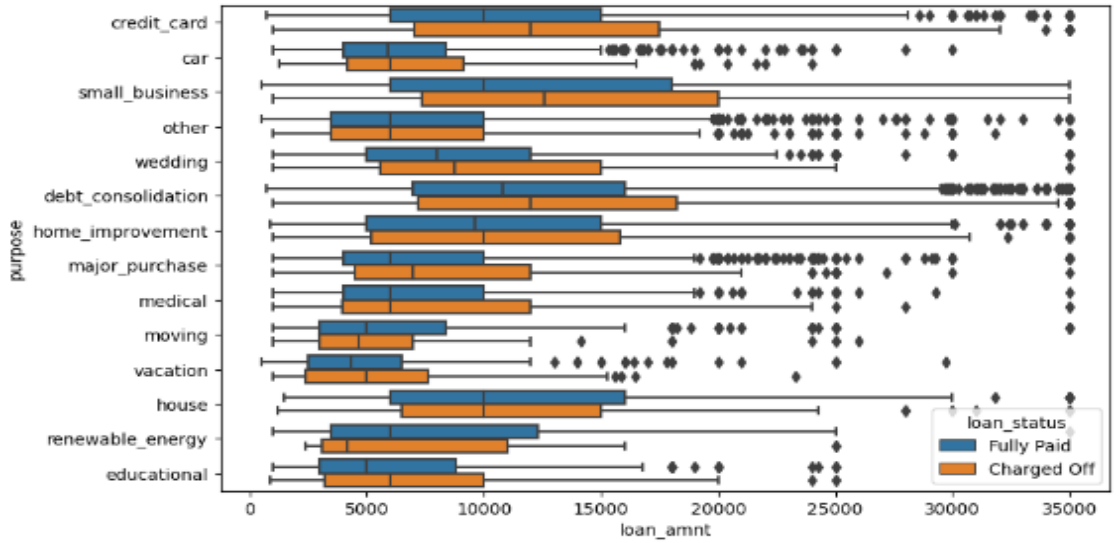


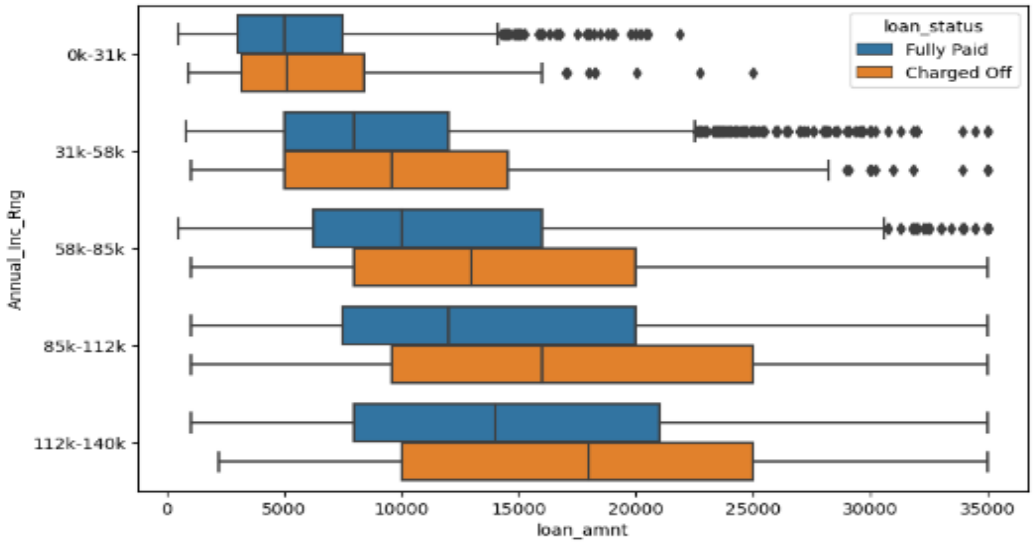


Bivariate Analysis on Charged Off loans(loan Amount)







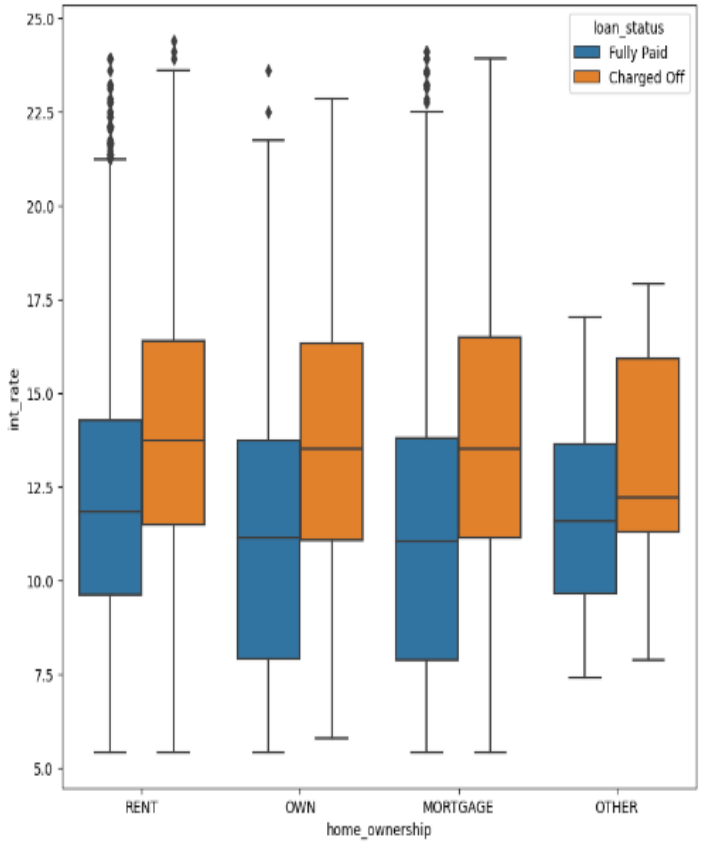


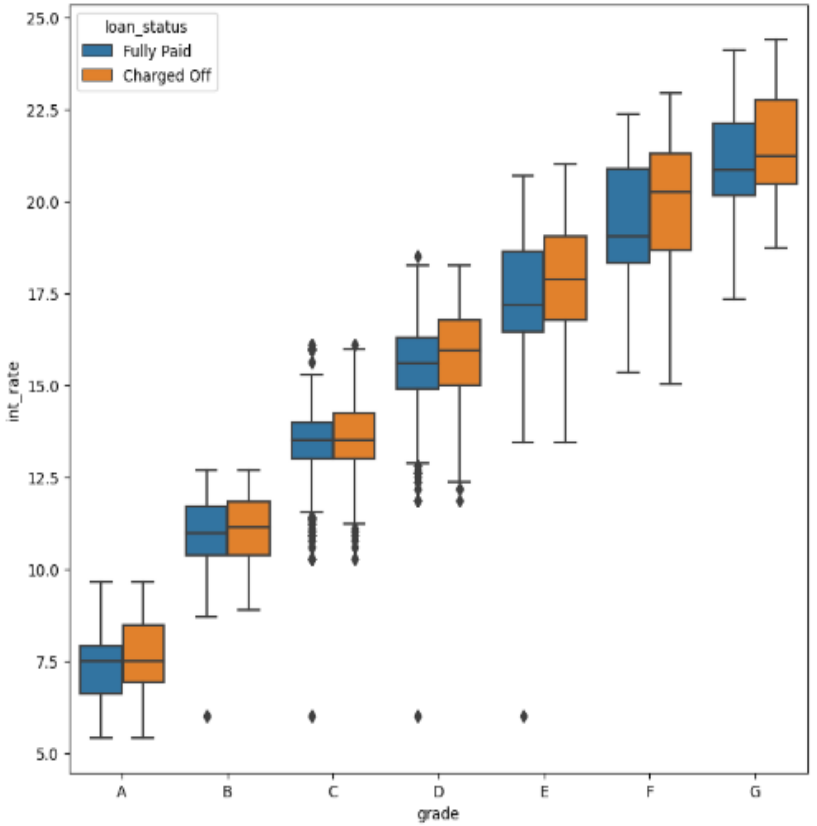
Bivariate Analysis on Charged Off loans

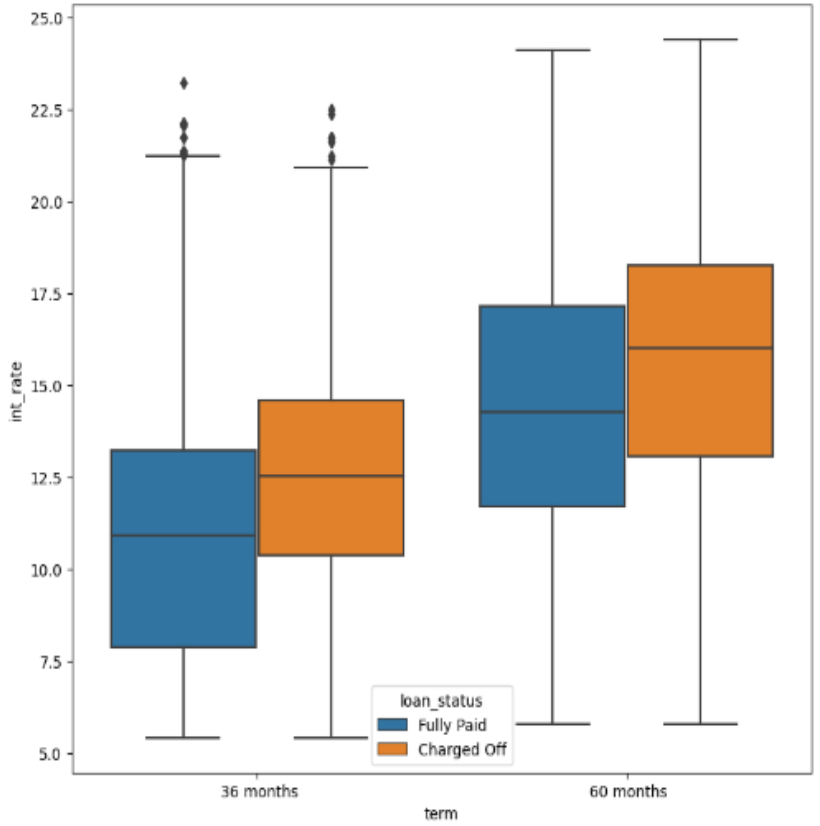
Bivariate analysis is performed for Interest Amount against

- Loan Amount
- Home Ownership
- Grade
- Term
- Emp Length
- Verification Status
- Tenure
- Annual Income
- Purpose

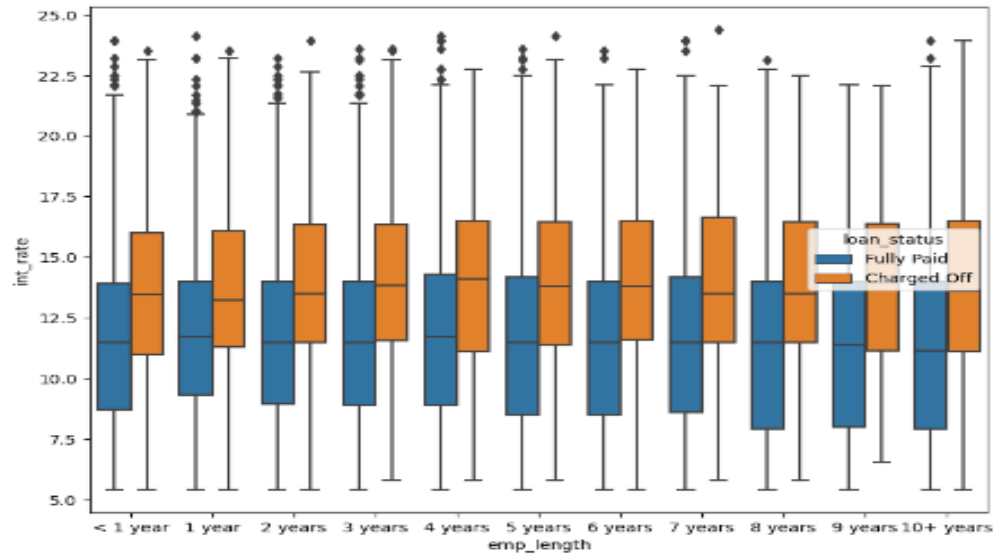
Bivariate Analysis on Charged Off loans(Interest Amount)

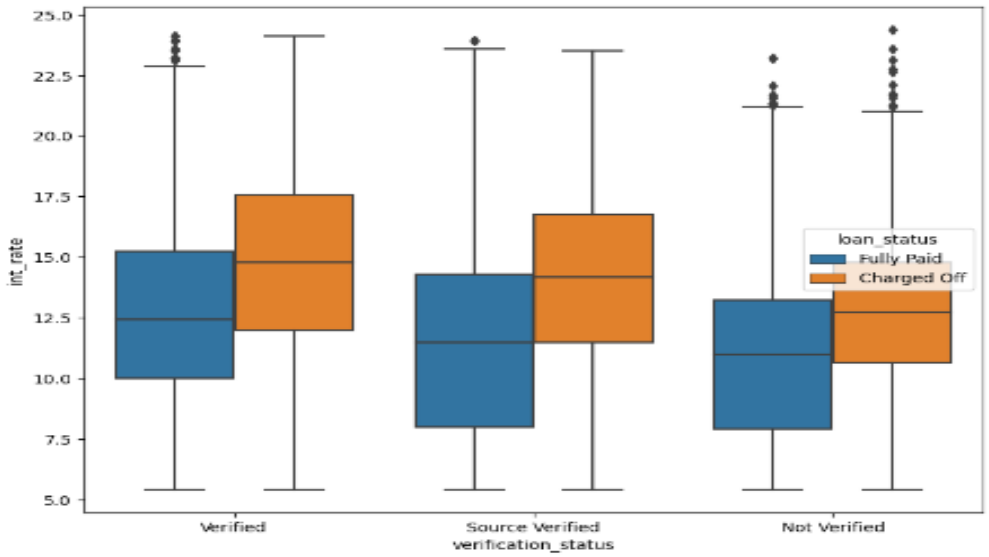


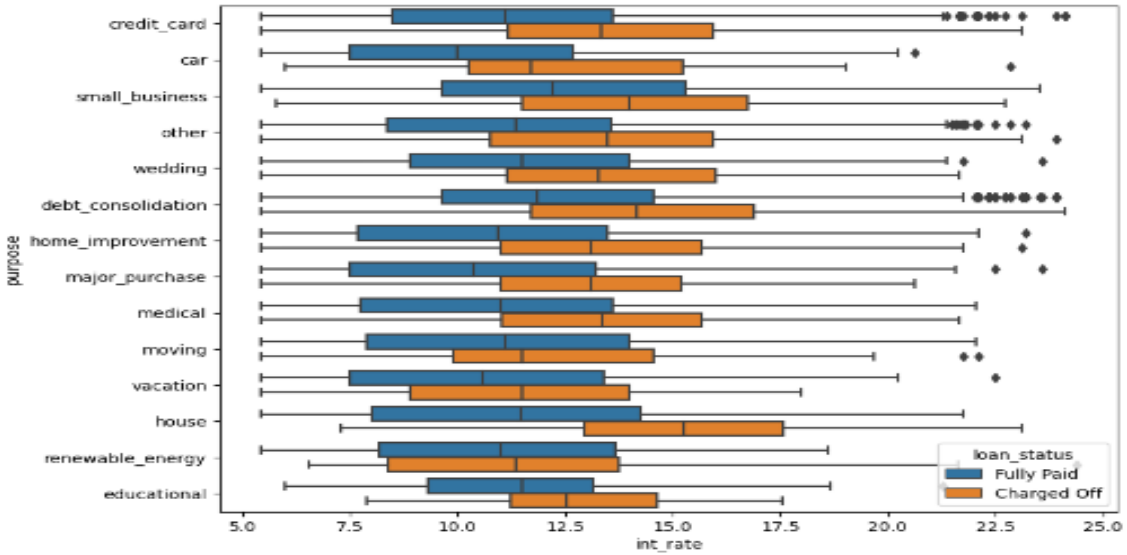


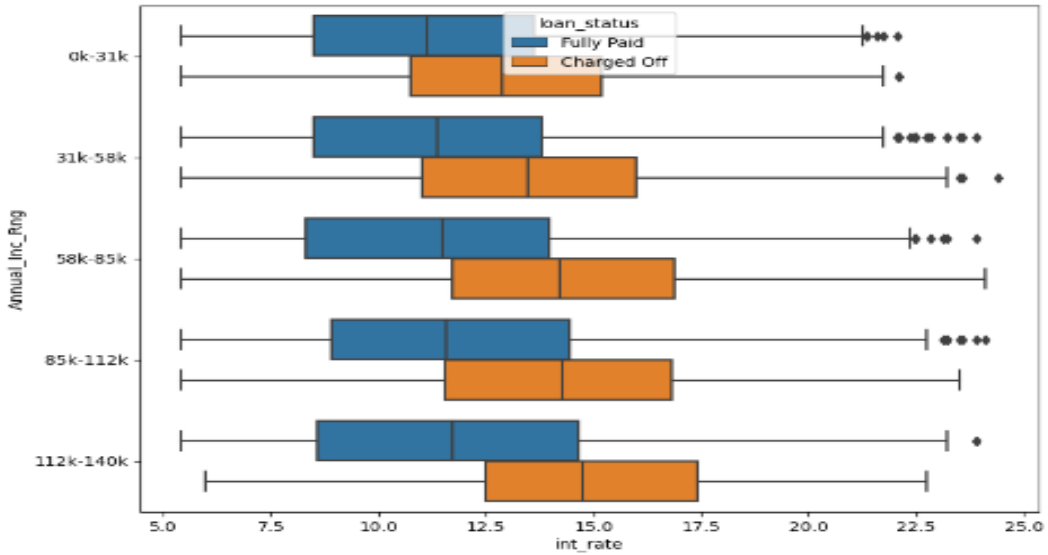


Bivariate Analysis on Charged Off loans(Interest Amount)







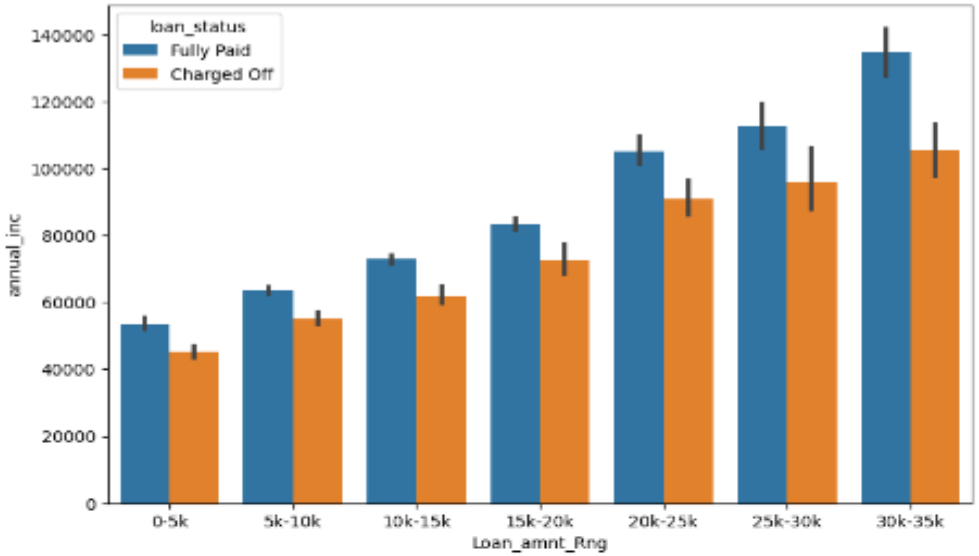


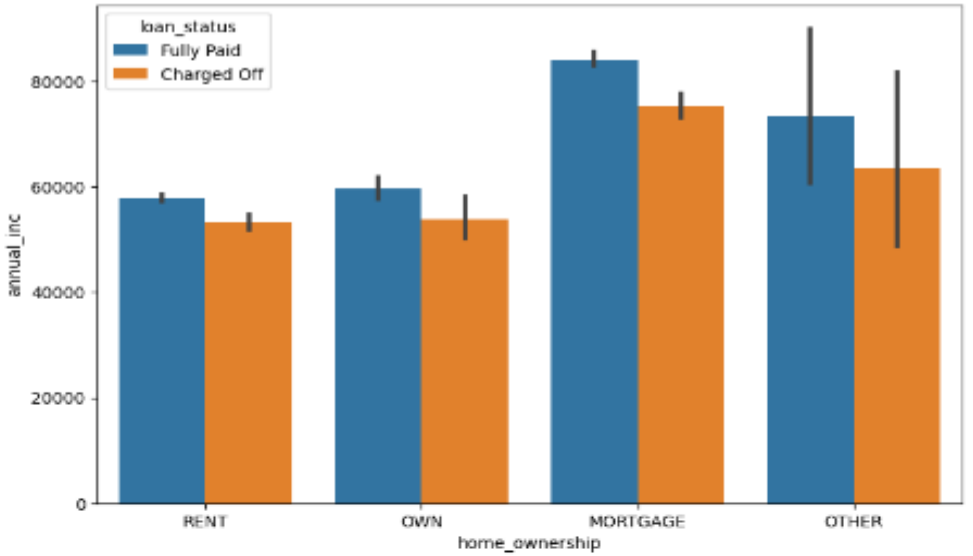
Bivariate Analysis on Charged Off loans

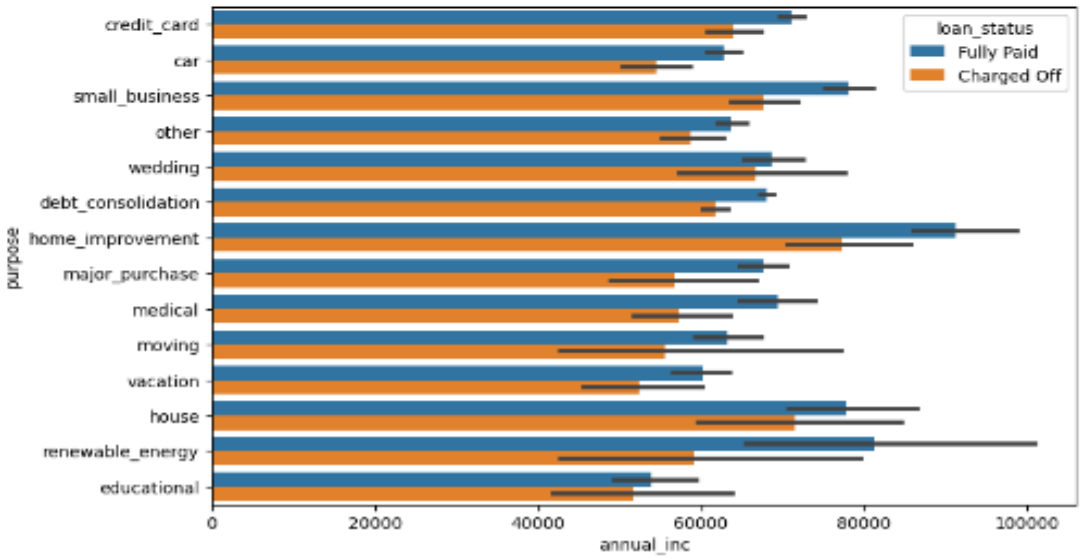
Bivariate analysis is performed for Annual Income against

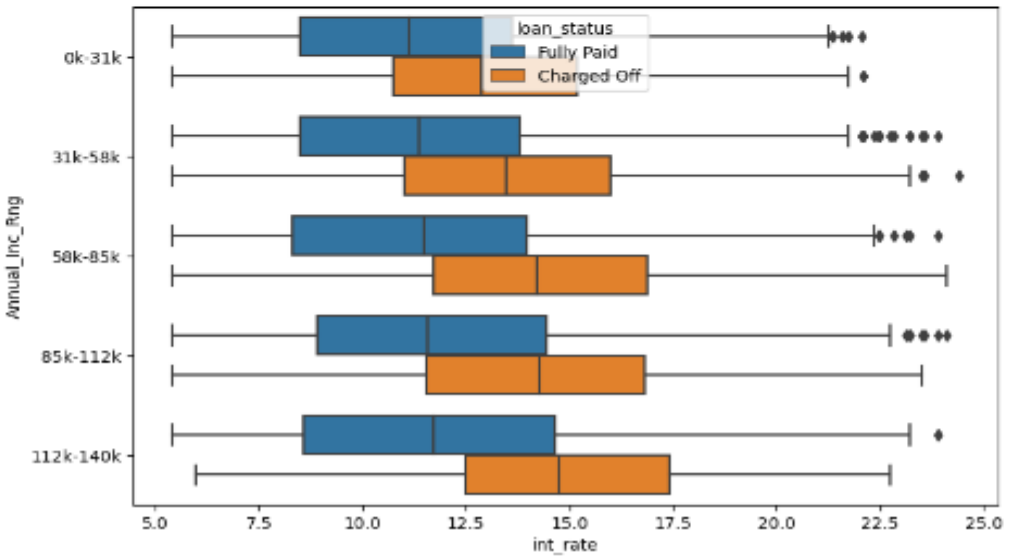
- Interest Rate
- Home Ownership
- Loan Amount
- Purpose

Bivariate Analysis on Charged Off loans(Annual Income)









Conclusions on the variables which are strong indicators of default based on bivariate Analysis

By running bivariate Analysis on the 5627 Charged off loans. The chances of the Application defaulting is higher if

- The large loan (10-25 k) offered to the applicant at a high rate of interest 17-24 has a higher chance of default
- Applicants on mortage and granted higher loan amount are bound to default.
- Grade F loans on high amounts/High Interest Rates tend to default
- Higher tenure 60 mnths loans for high amounts and high interest rates again have a tendency to default
- Applicants with 10+ years of employement and granted higher loan amnt on higher interests also tend to default
- Small business owner given Higher loan amountss tend to default
- Loans granted for debt consolidation on high interst tend to default
- Applicants with an annual income of 50k-85k and granted high interst loans have a high default rate