Lending club case study

LOKESHKUMAR RAJENDRAN

BUSINESS OBJECTIVE

The Objective of the case study is to implement EDA techniques for a real world problem and understand the insights to mitigate the credit loss to Lending club

- 1.Identifing the potential candidates who can repay the loan is important, if we reject the person who is an potential barrower and who can repay the loan its an loss to the company
- 2.If we approve a loan to the person who has very less changes of repaying it, then it’s a loss to the leaning club

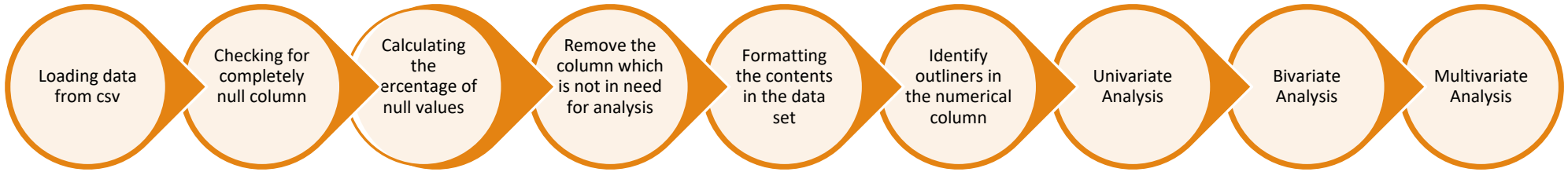
With this case study we will be able to pin point the applications at risk of defaulting the loans, enabling a reduction in credit loss.

By using the EDA techniques we will be able to understand the driving factor behind the loan defaulters i.e the variables which are strong indicator of default.

Data cleaning & pre-processing

Dataset details:

The data given contains information about the past loan applications, weather if they have defaulted or paid fully or the loan is in still progress
The data set also has the lot of columns which is not needed for the analysis, and few columns which are completely blank

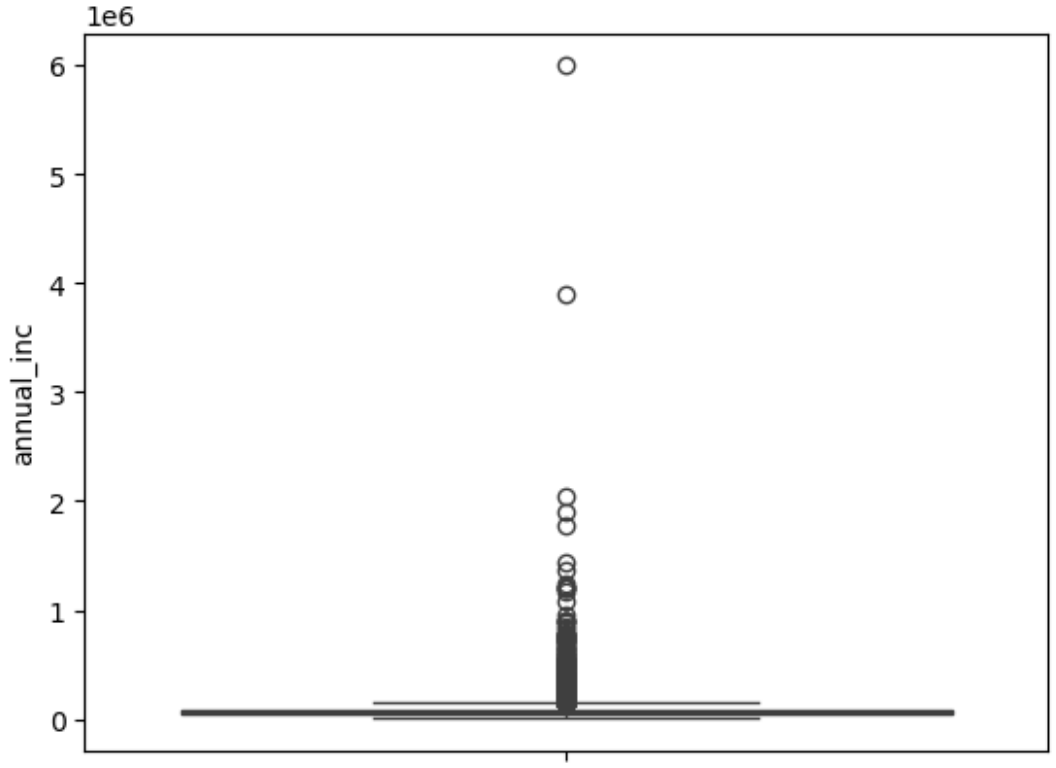


Removing the columns

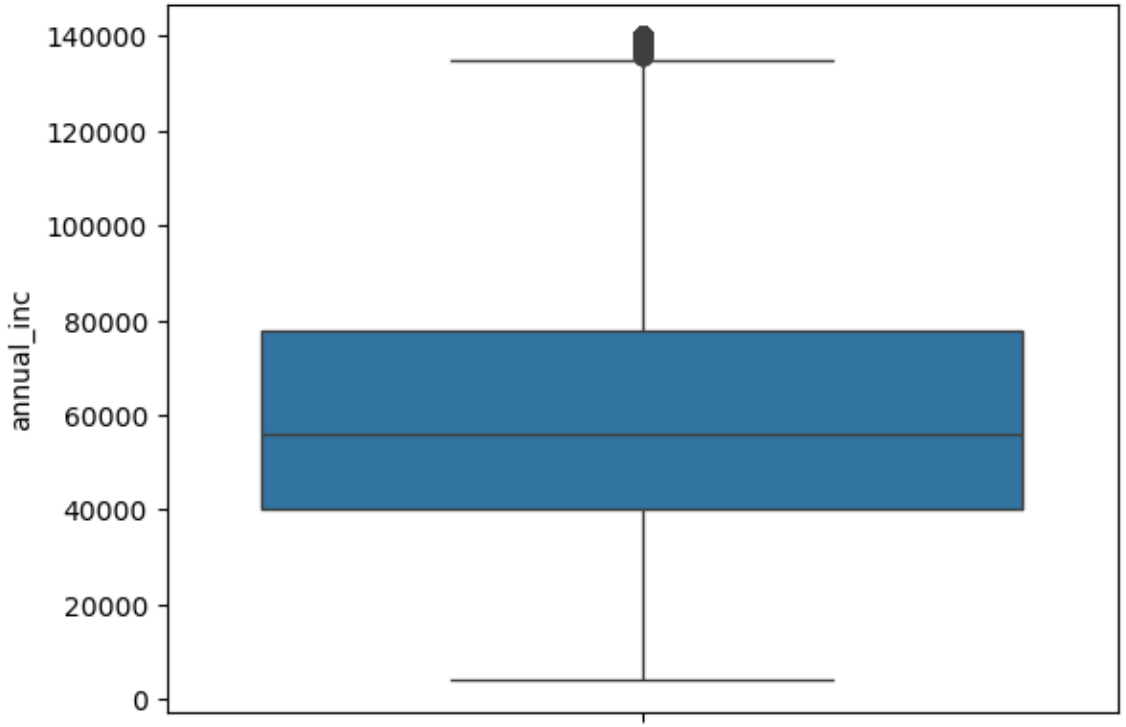
- **Removing the columns which is not needed for the analysis**
  - Zip code only first 3 characters are visible
  - pymnt_plan column as it has sinle value n
  - application_type this column has only one value as INDIVIDUAL
  - policy_code column has single constant value 1
  - delinq_amnt column has single constant value 0
  - tax_liens column has single constant value 0.0 and nan
  - initial_list_status column has single constant value f
  - collections_12_mths_ex_med column has single constant value 0.0 and nan
  - acc_now_delinq' column has single constant value 0
  - chargeoff_within_12_mths column has a single constant value 0.0 and nan
- Identify the columns with the unique values, and remove it as it will not be of any use in analysis

Identifying outliners-Annual Income

Clearly the annual income has the outliners

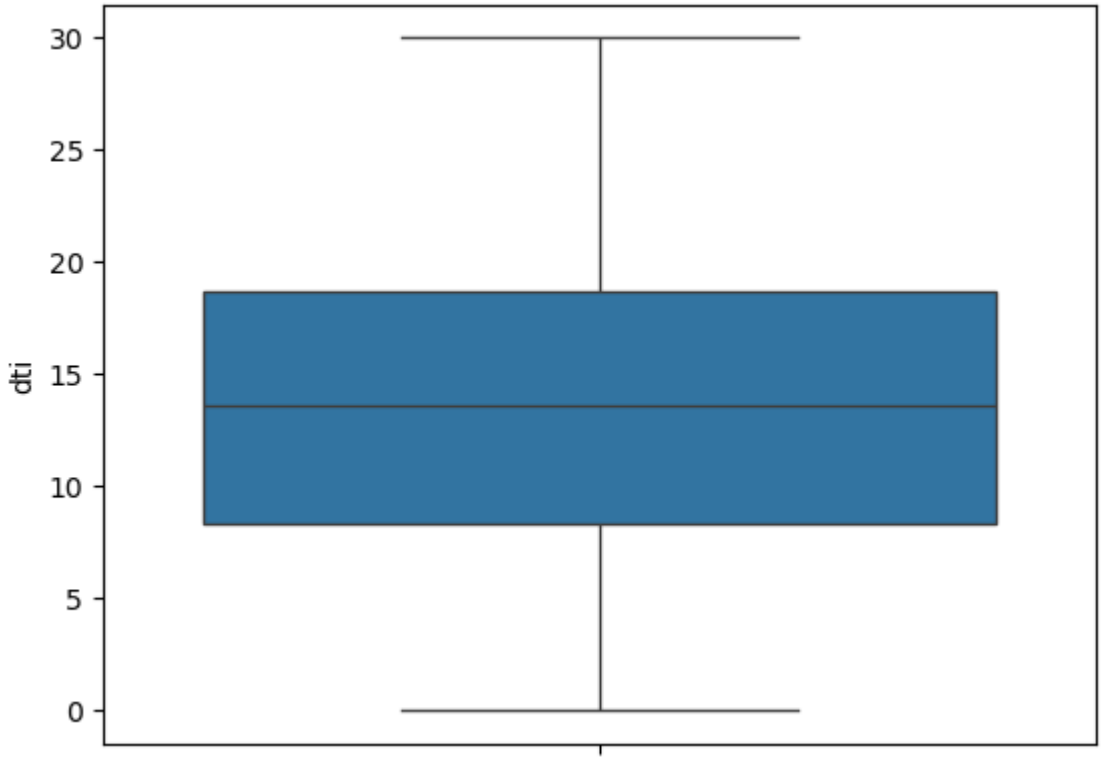


The value after 95% seems disconnected, so removed values after 95%

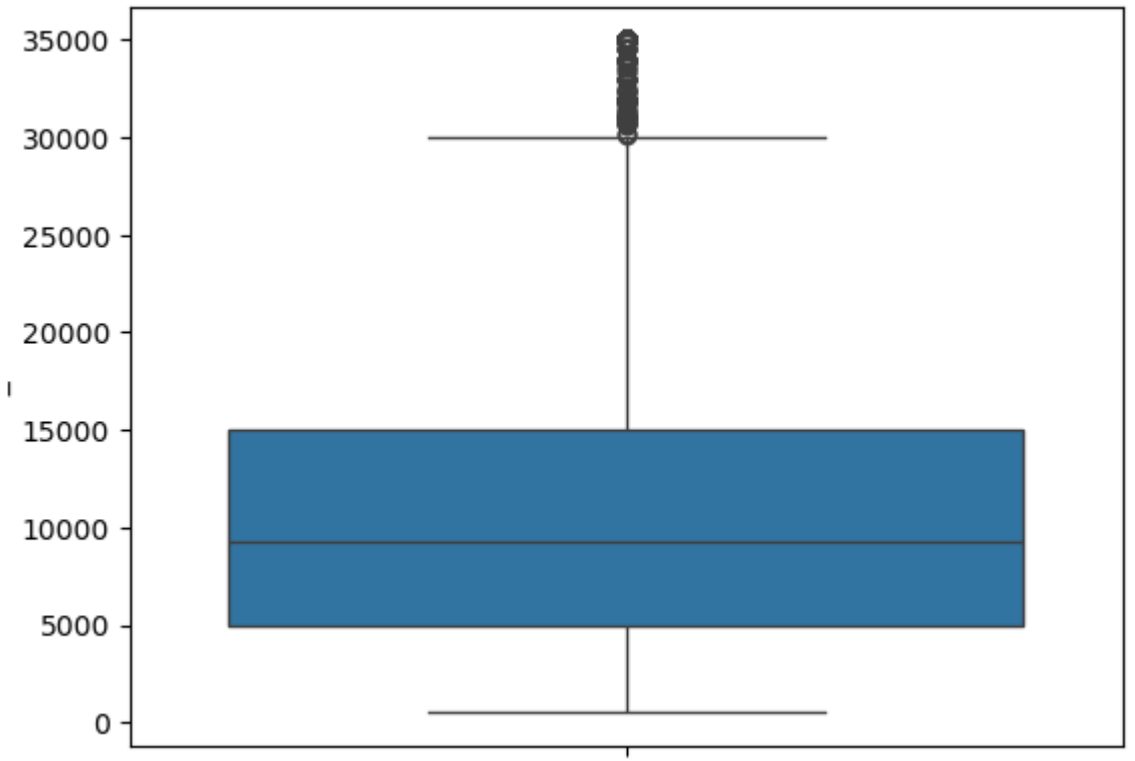


Identifying outliners- Dti, loan amount

DTI column has no outliners

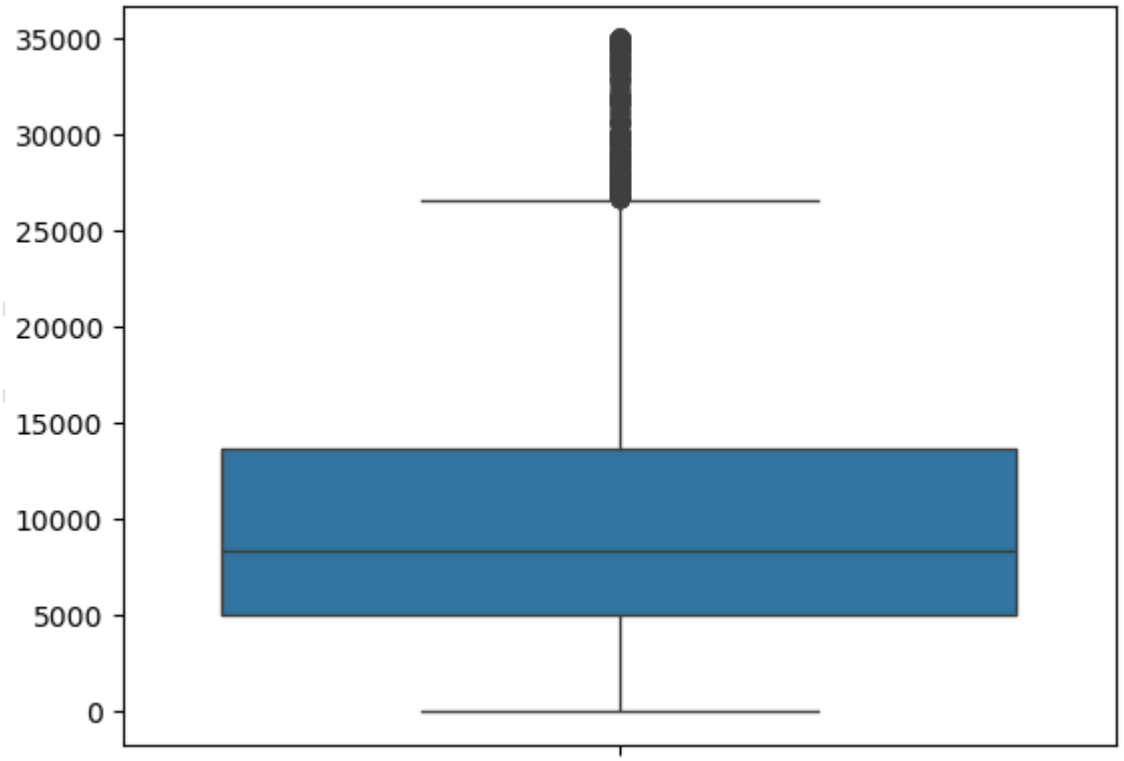


Loan amount column has no outliners

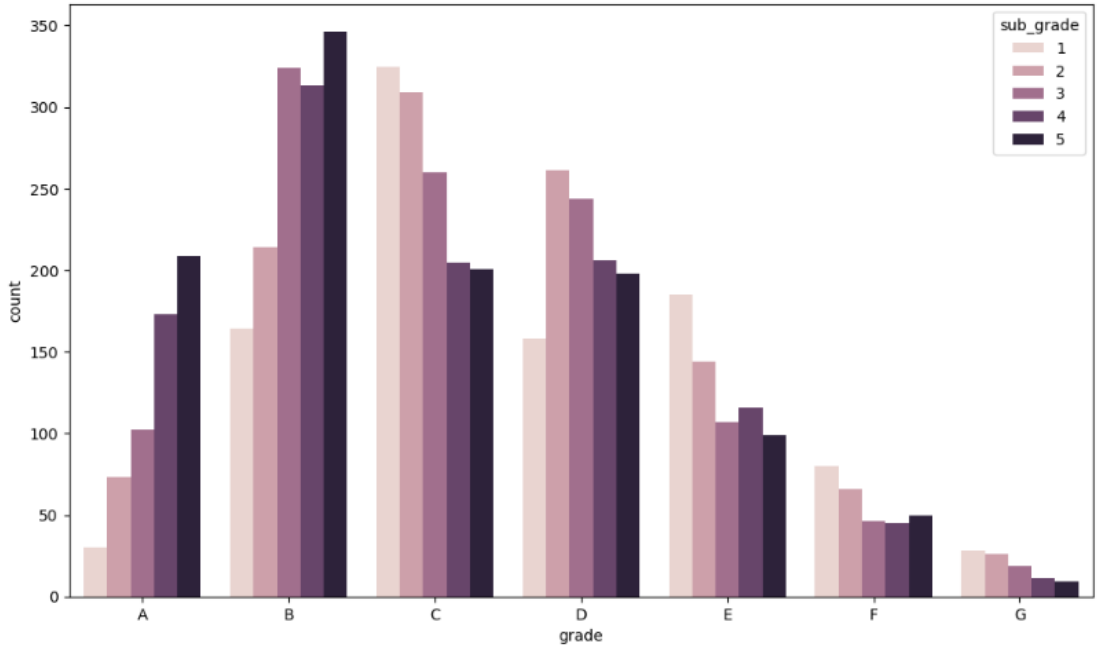


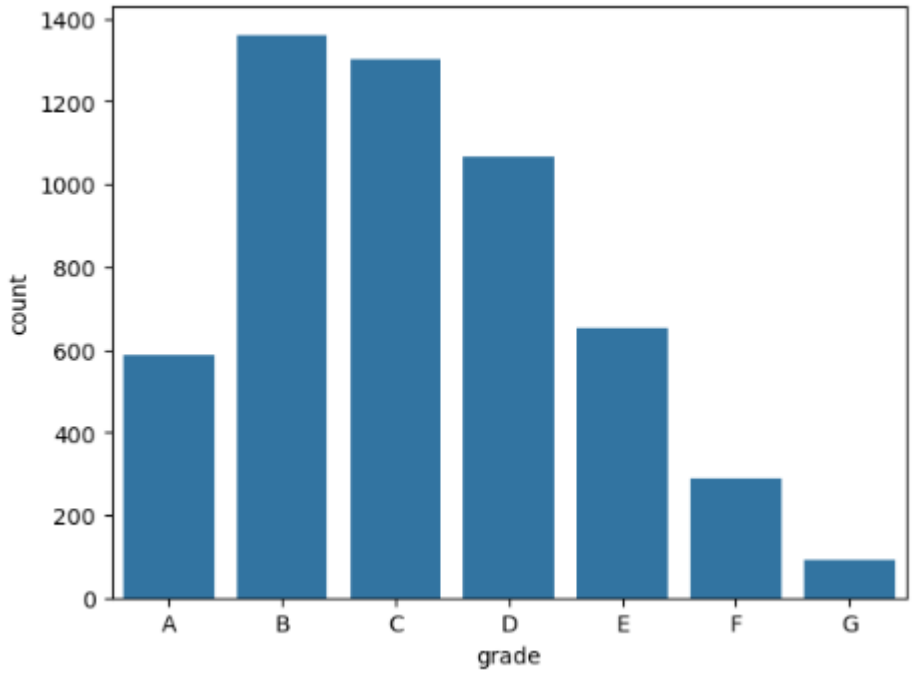
Identifying outliners- Funded amnt inv

The column funded amnt inv has no outliners



Segmented Univariate Analysis-Grade column

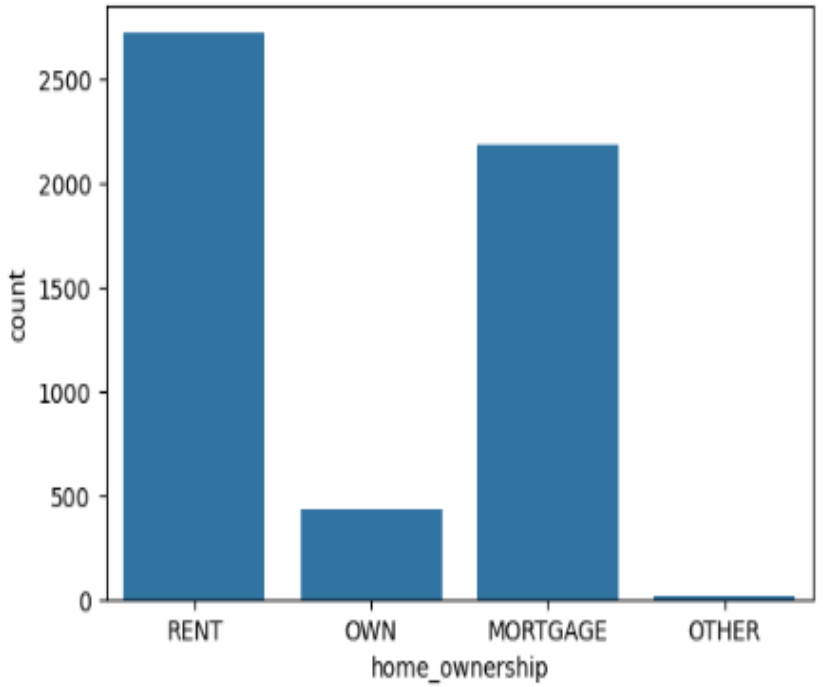


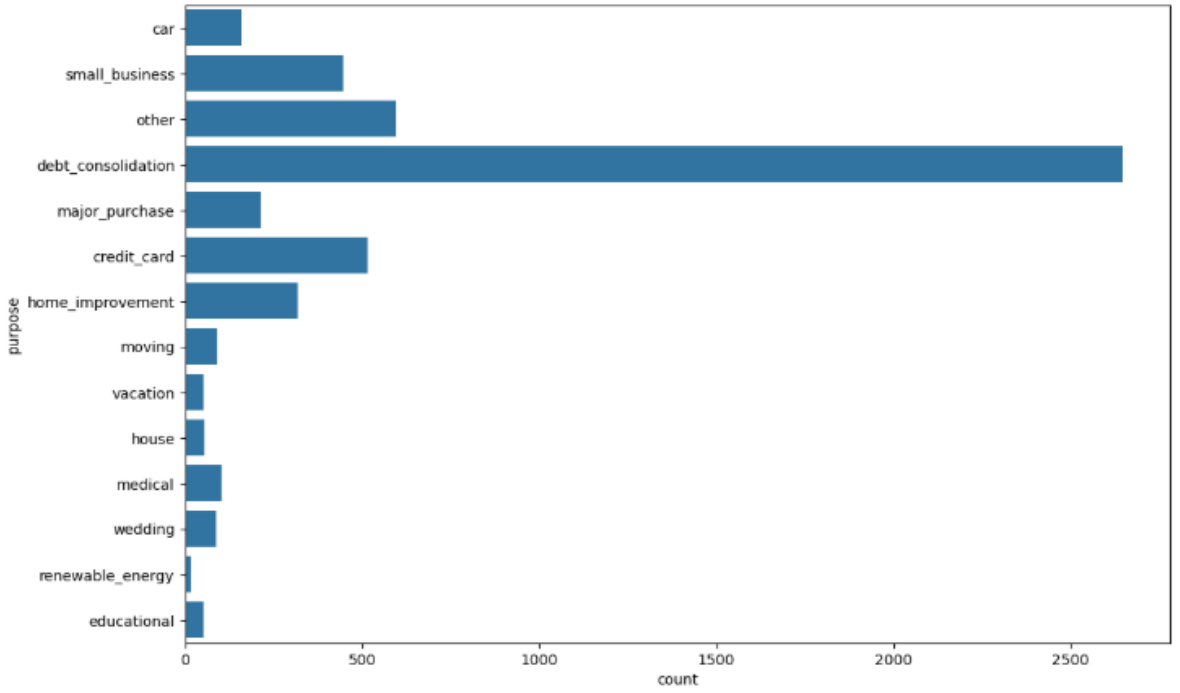


- The analysis done on grade column with subgrade as hue, Indicates that person with grade B5 has the high chance of default

The analysis with grade column shows the B grade employees, have defaulted more

Segmented Univariate Analysis-Home ownership and purpose column

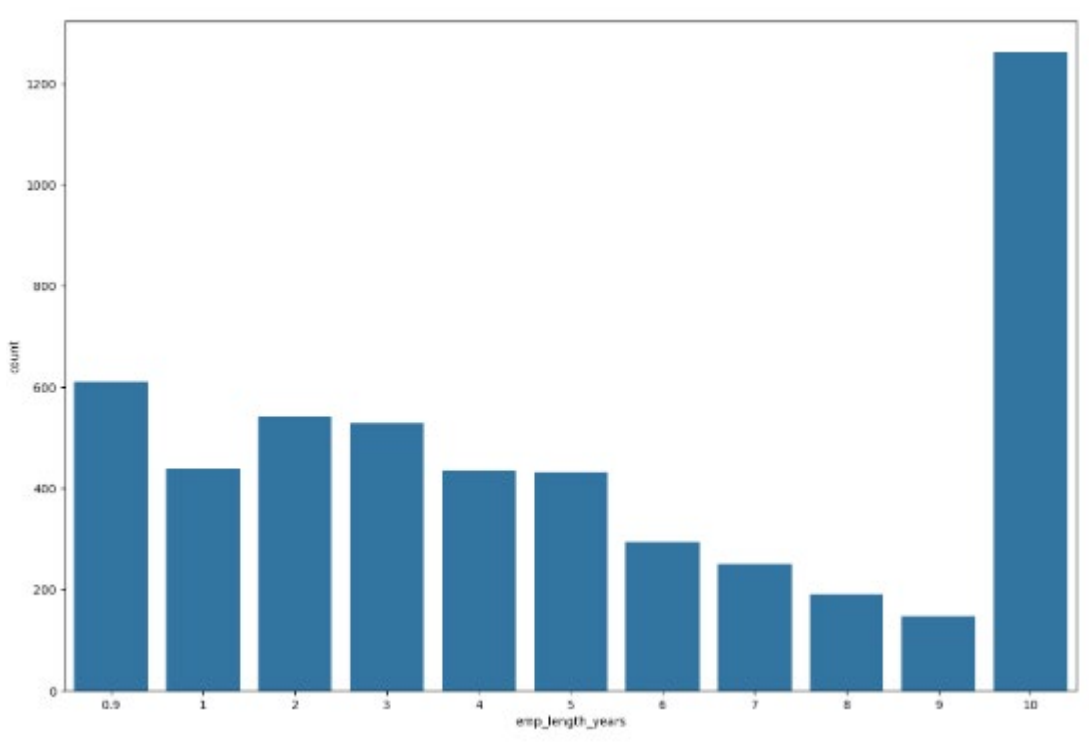




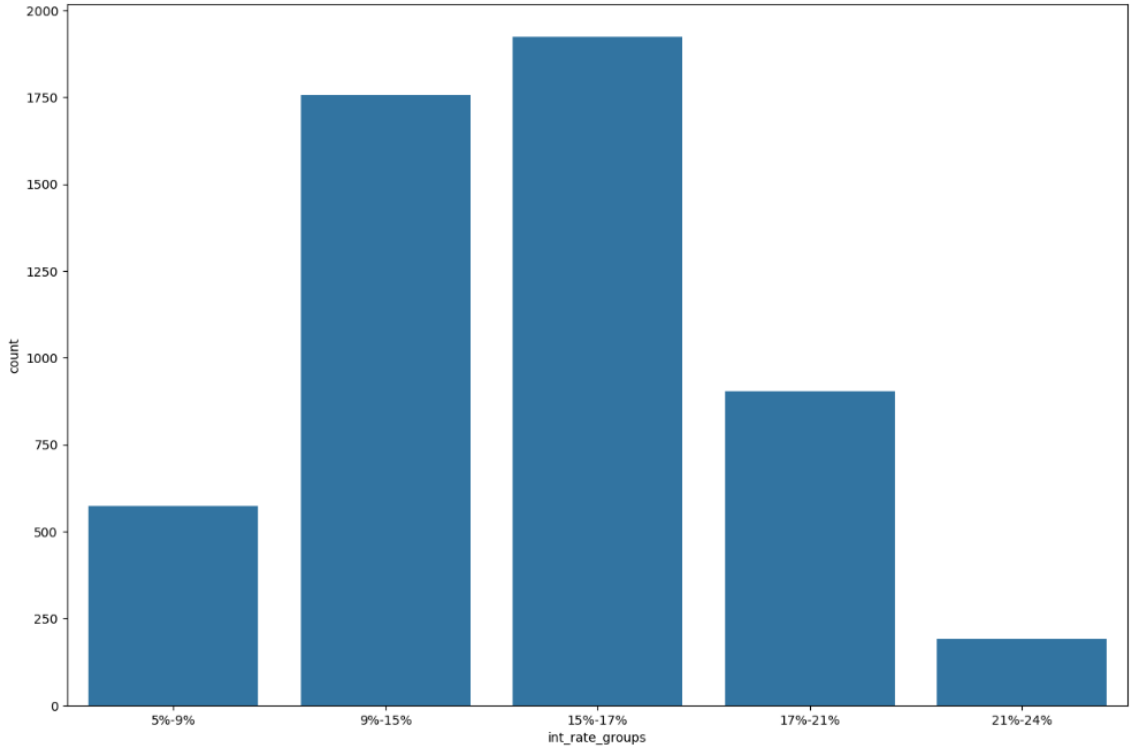
- The person who has the home ownership as Rent has defaulted more when compared to Mortage of home

- The person who bought the loan to pay other loan i.e debt consolidation has very high chances of defaulting when compared to other purposes

Segmented Univariate Analysis-Employee length, interest rate

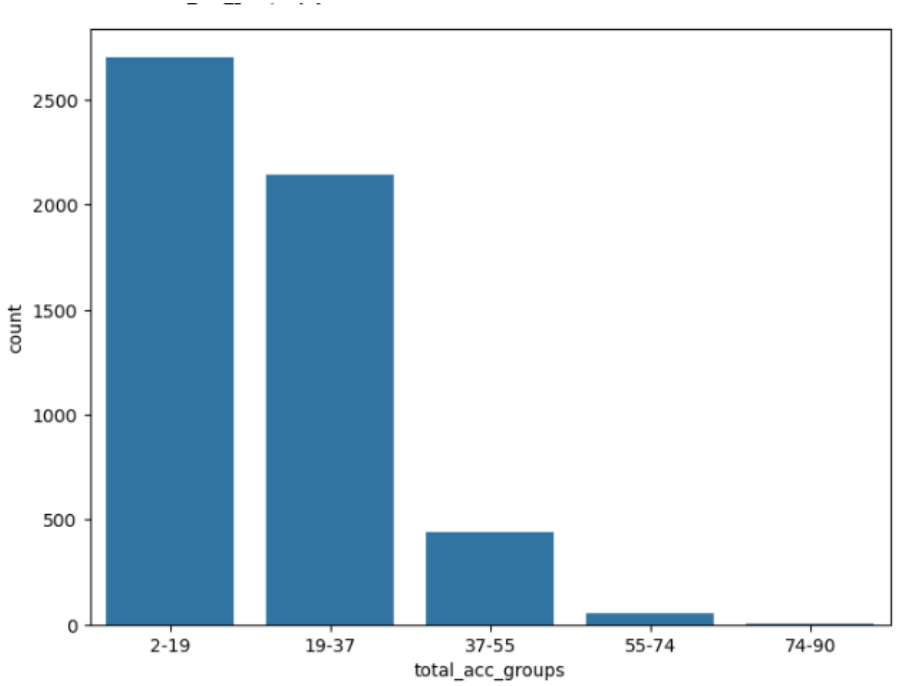


- The Employee who has 10 years or more than 10 years has defaulted more

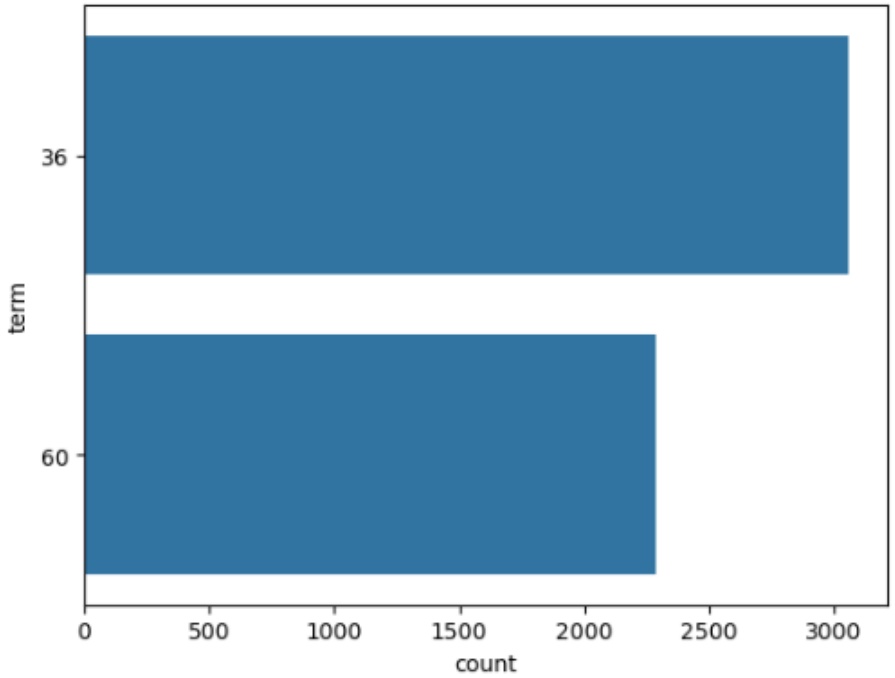


- The loans that was provided with the interest rate of 15-17% has defaulted more

Segmented Univariate Analysis-Total account and term column

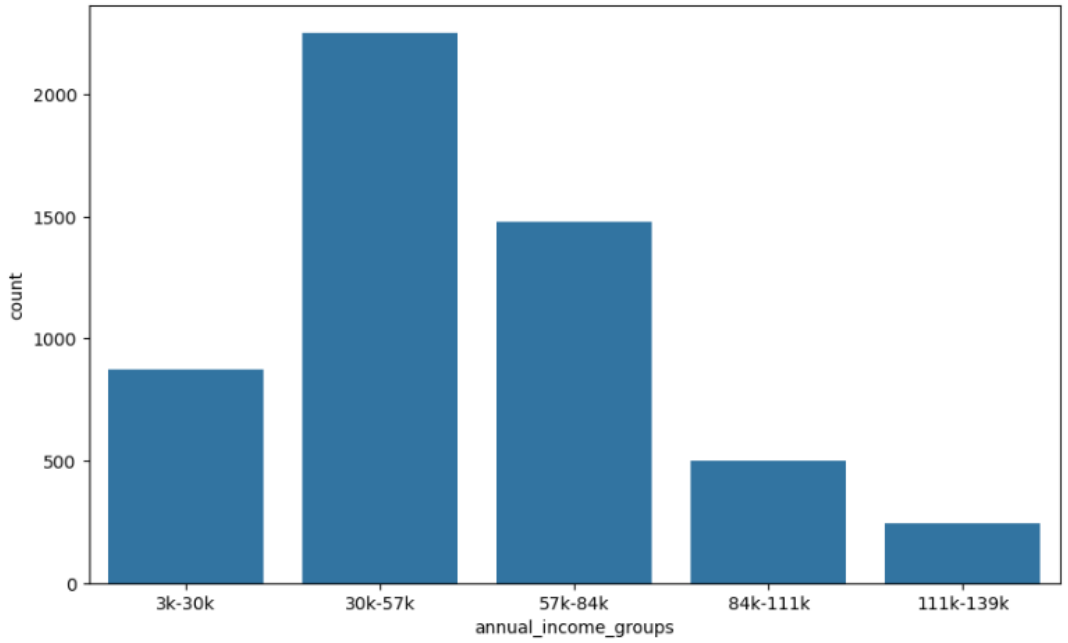


- The person who has total account between 2-19 has defaulted more

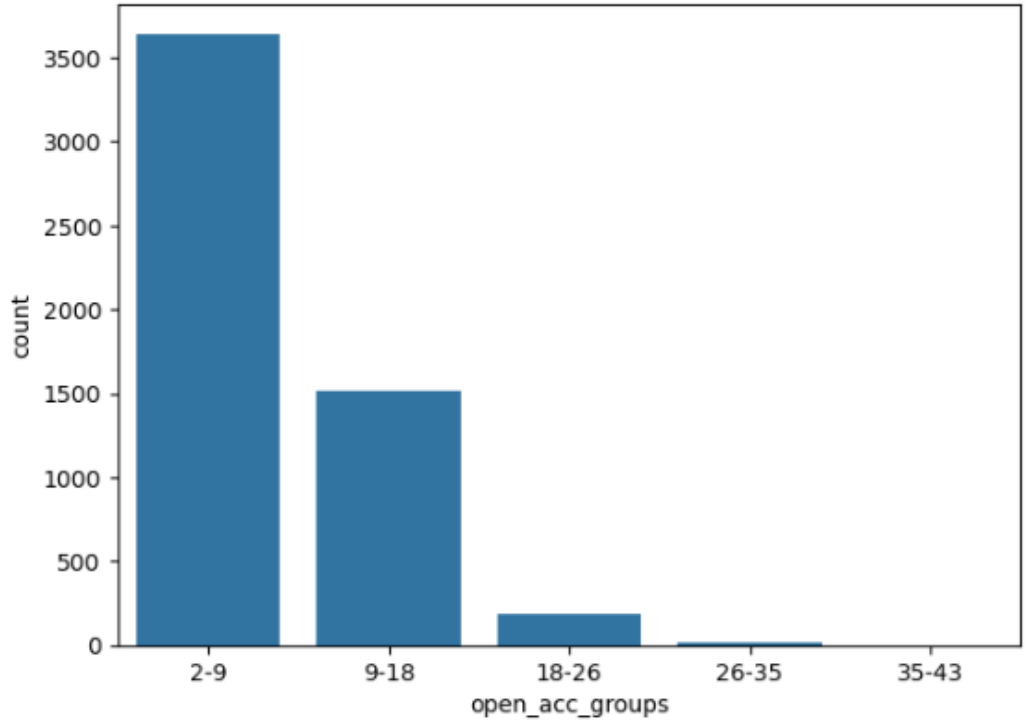


- The loan which has the 36 months term has defaulted more when compared to 60 months

Segmented Univariate Analysis-Annual Income and open account

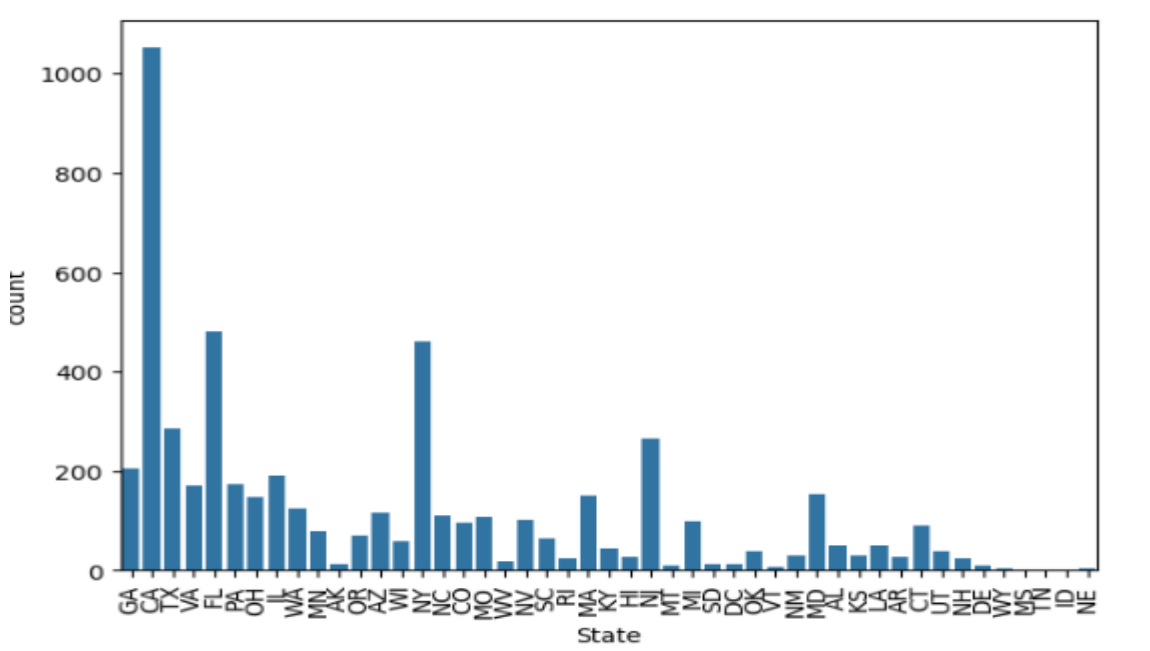


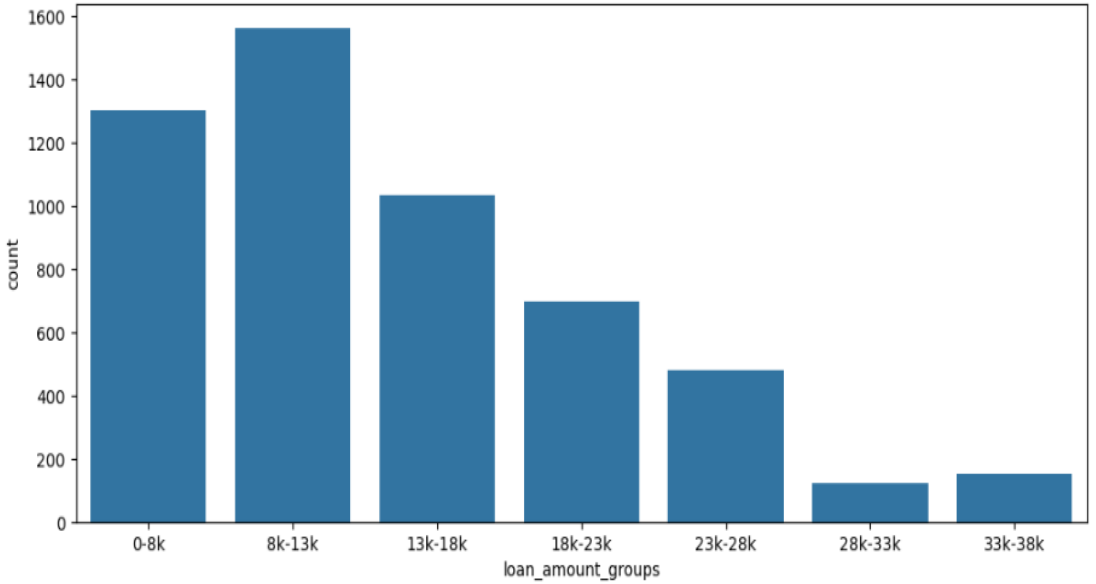
- The person with the income range of 30K to 57K has defaulted more, an extra caution has to be taken for the loans with salary of this range



- People with the open account of 2-9 has charged off more when compared to others

Segmented Univariate Analysis-State and loan amount

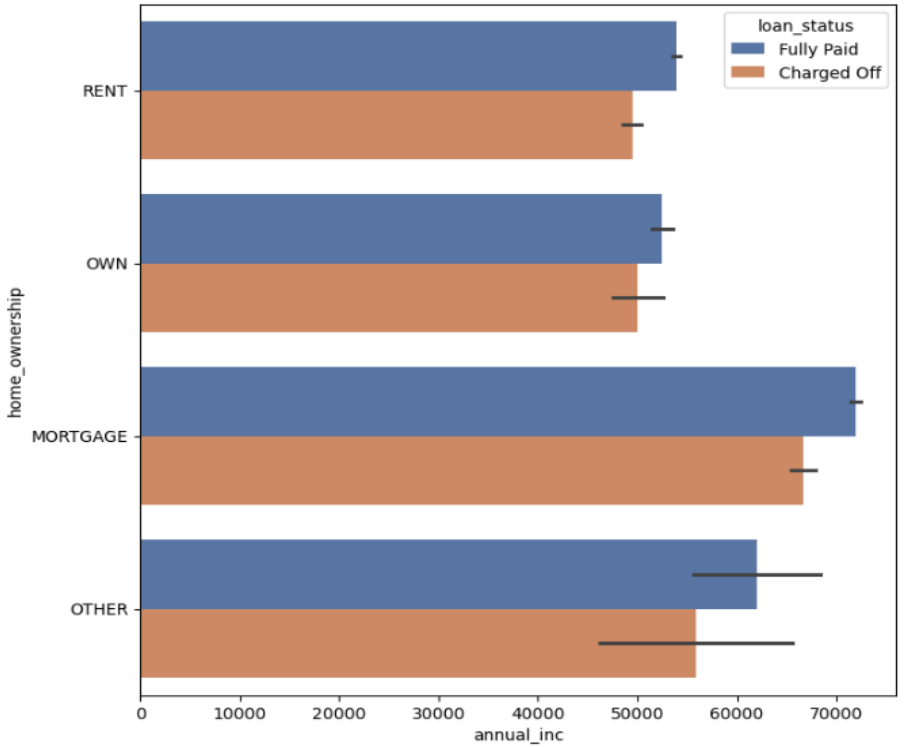




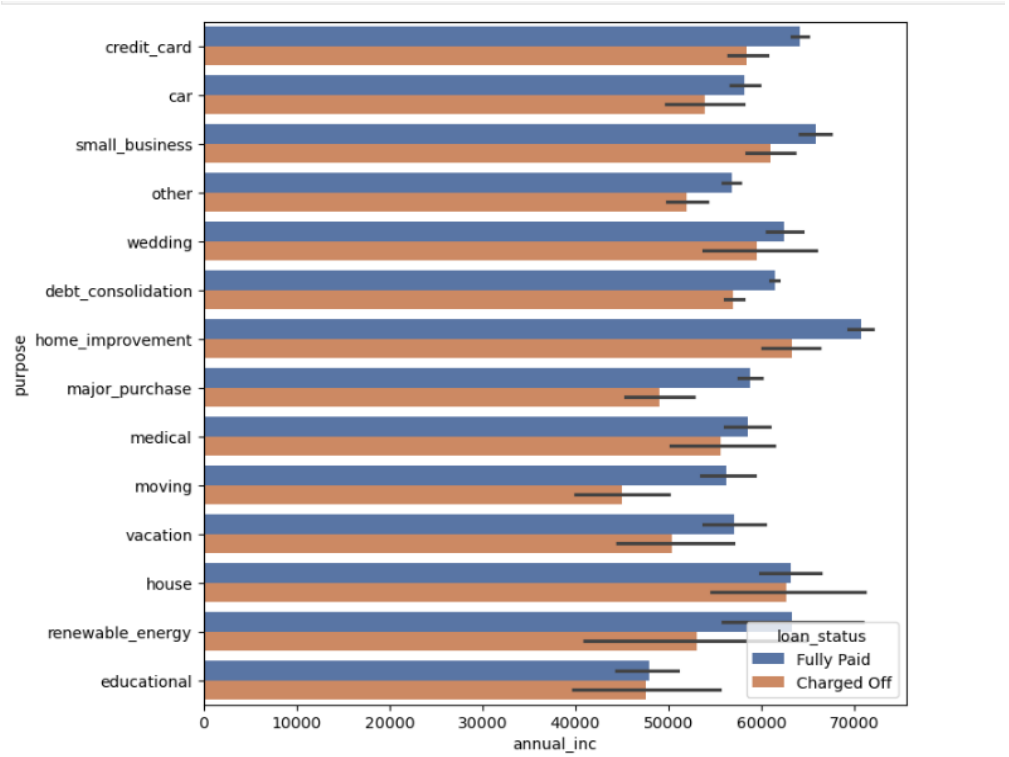
- People from the CA state has defaulted more when compared to others. Lending club should have extra due diligence on the CA state

- Loans provided between 8K to 13K has defaulted more, more the loan amount there is less changes of defaulting

Bivariate Analysis-Annual income vs home ownership and Annual income vs loan purpose

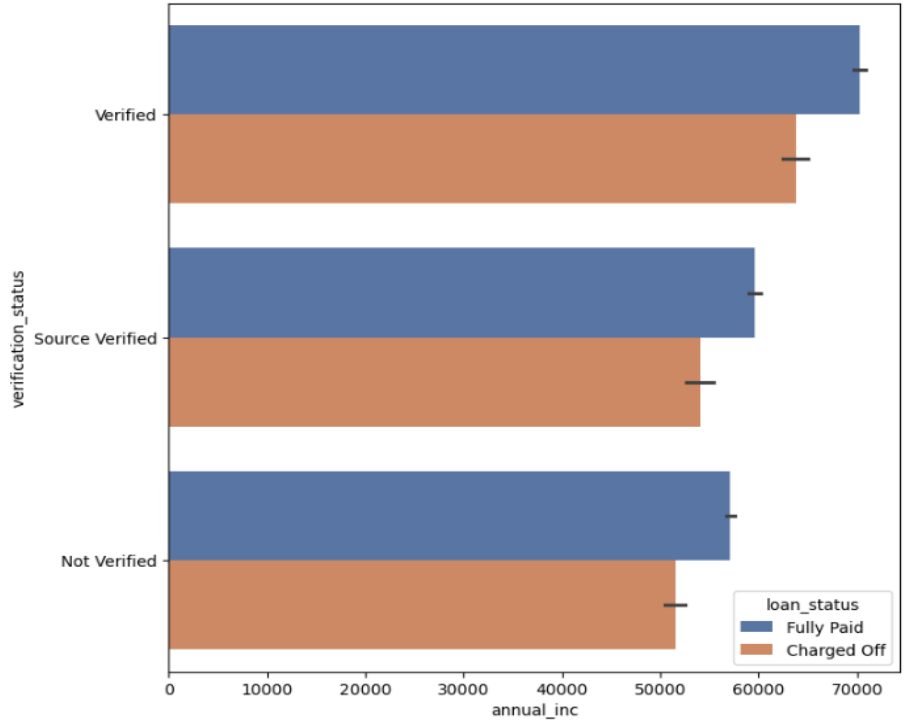


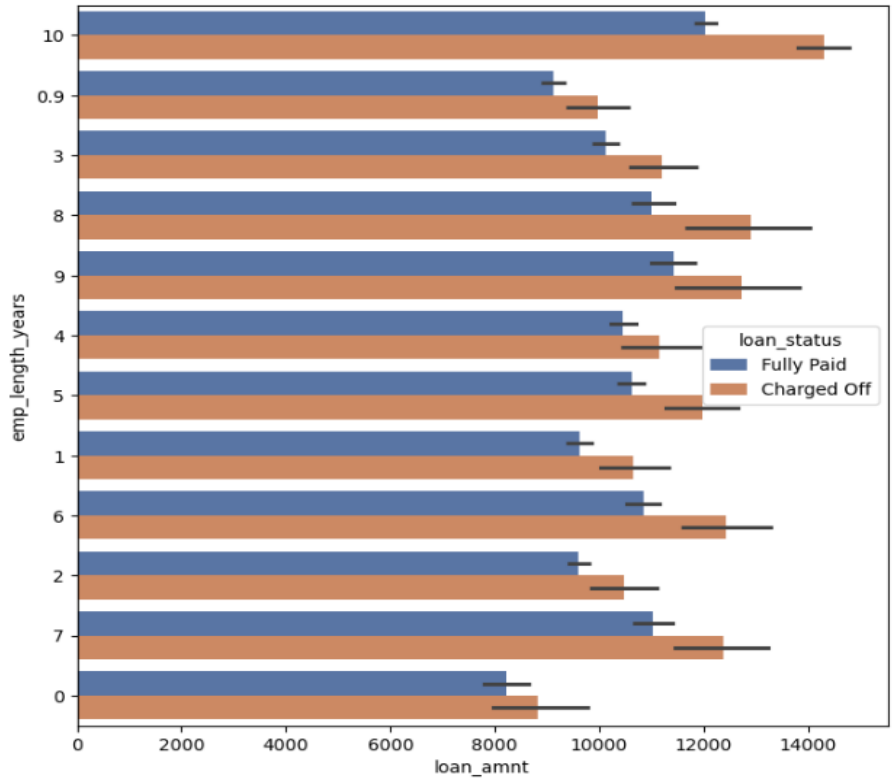
- Persons with mortage having the annual income of 60k to 70k has more possibility of defaulting



- person who is taking loan for home improvement and if he has salary of 60k to 70k then he is likely to default

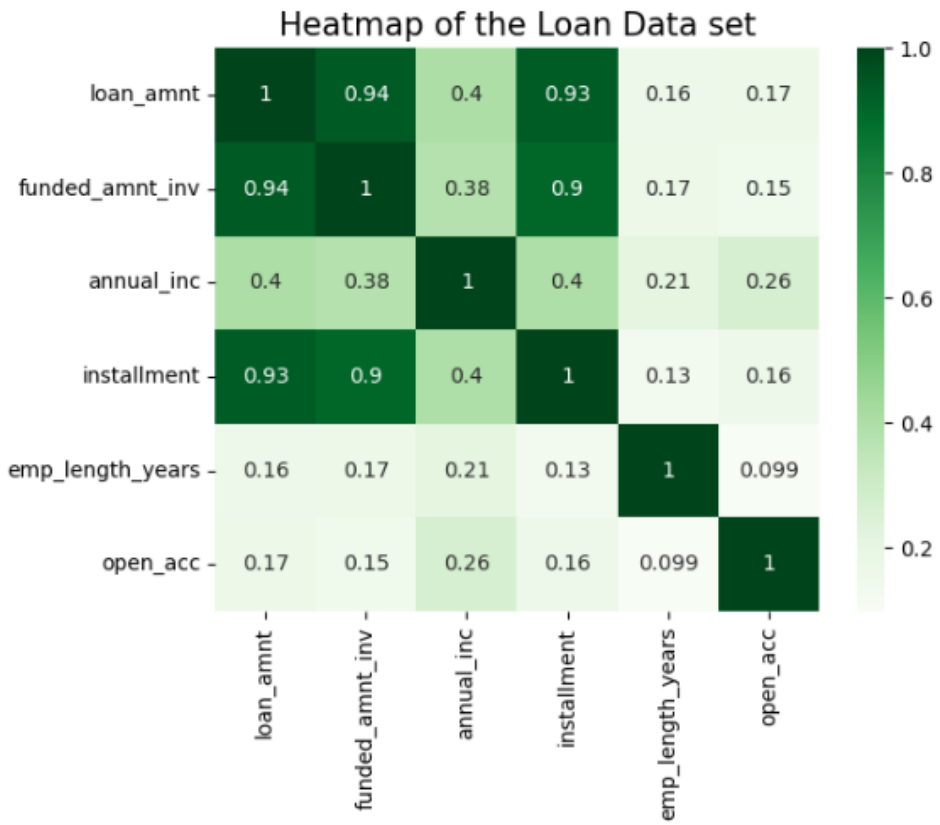
Bivariate Analysis-Annual income vs verification status and Loan amount vs employee length





- person with verified status who is having the salary range from 65K to 70K is likely to default
- Employee with equal to or more than 10 years of experience has high amount of loan approved and they have charged off more

Multivariate Analysis



The multi variate analysis is carried on the below columns

- Loan amount
- Funded amount inv
- Annual inc
- Installment
- Emp_length
- Open_acc

The dark green colour indicates the positive correlation between the columns

Positive corelation between the loan amount, funded amount and the installment is seen

Weak correlation is found on open account and employee length

Suggestions

The major driving factors which is used to predict the possibility of defaulting is below

- Grades
- Verification status
- Annual income
- Loan amount
- State
- Purpose of the loan
- Employee experience

✓ **Implement Stricter check for Grade B,C,D employee :** we should implement a stricter check or should implement an extra verification for applicants falling under B,C and D category to default minimal risk.

Suggestions

- ✓**Focus more on subgrades B3,B4,B5,C1 and C2:** Analysis shows the employee with this subgrade has more possibility of defaulting , lending club should consider additional risk mitigation measures
- ✓**Interest Rate decrease for 60 months loan:** lending club should consider decreasing the interest rate for the long term loans
- ✓**Careful Evaluation for Debt Consolidation Loans**: Carefully evaluate applicants seeking loan for the purpose of debt consolidation, considering potential interest rate adjustments or offering financial counseling services to manage the associated risks.
- ✓**Consider home ownership:** Home ownership should be considered as an potential factor for repaying the loan, the person who has own home has no need of rent, so the possibility of repayment of loan is high

Suggestions

Strict verification on city: A strict verification has to be conducted on the loan from city's like CA,NY and FL cities

Thorough Assessment for High Loan Amounts: Conduct more thorough assessments for loan amounts between 8K to 13K. Consider capping loan amounts for higher-risk applicants to mitigate potential defaults

GitHub Repository Link: