Lending Club Case Study

Ankit Kumar

Praneta M

Abstract

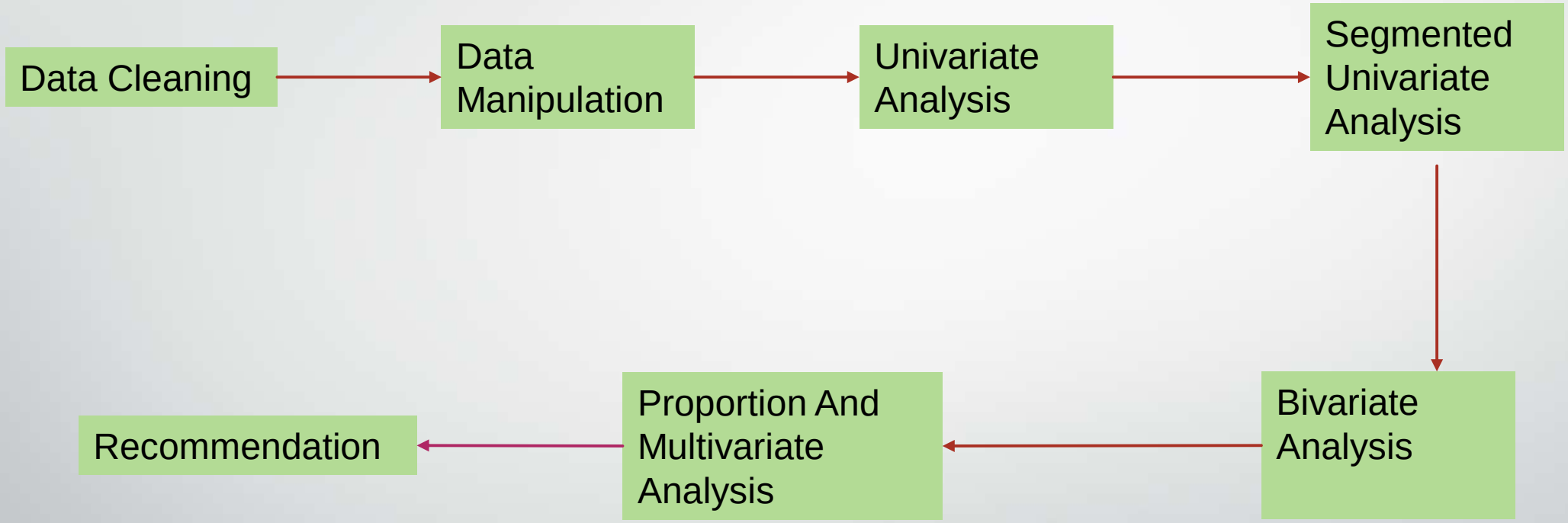
Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.

When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's 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

The 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.
Hence we have done EDA for the same which identifies those factors recommending company on whom to give loans which may not go as defaulters.

EDA Process


Data Cleaning
Data Manipulation
Univariate Analysis
Segmented Univariate Analysis
Bivariate Analysis
Proportion And Multivariate Analysis
Recommendation

Data Understanding

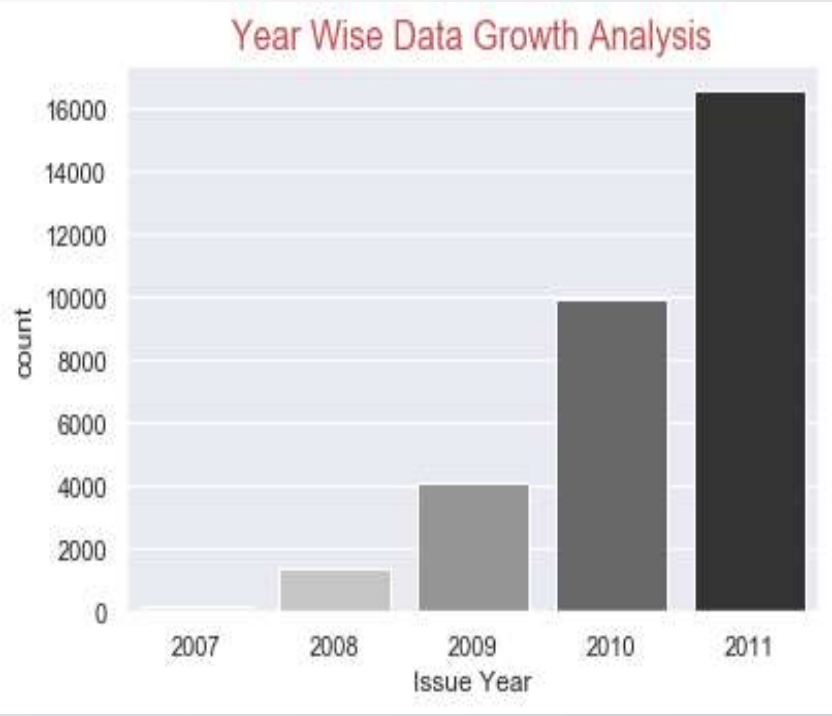
Once we have read the problem statement, going through the data to get clear understanding of the data.

Data is given in the form of csv file with 111 columns around 39k rows.

There are broadly three types of variables –

- Those which are related to the applicant (demographic variables such as age, occupation, employment details etc.),
- Loan characteristics (amount of loan, interest rate, purpose of loan etc.) and
- Customer behavior variables (those which are generated after the loan is approved such as delinquent 2 years, revolving balance, next payment date etc.).

The data file is read by pandas and stored in a dataframe.



With each year the loan count is growing.

Data Cleaning

Checking the null variables of the data and dropping the columns with all null values.

Checking unique variable count is each column so that single unique value columns can also be deleted as we will not use it further.

Demographic, ID and behavioural attribute columns are also dropped as it will not be deterministic variables in identifying the defaulters.

We are left with 16 rows for Data Analysis.

In those rows two columns (employee length and bankruptcies has missing value which is handled with Mode imputation.

Data Manipulation and Standardization

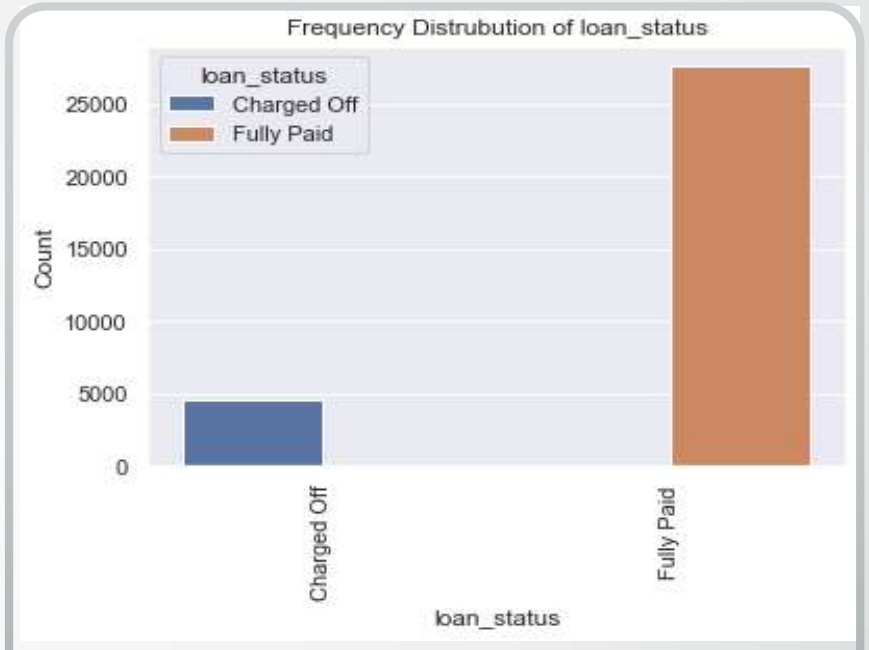
Deriving attributes from existing columns and modifying columns is done to derive month and year from date and converting object columns to float, string or category.

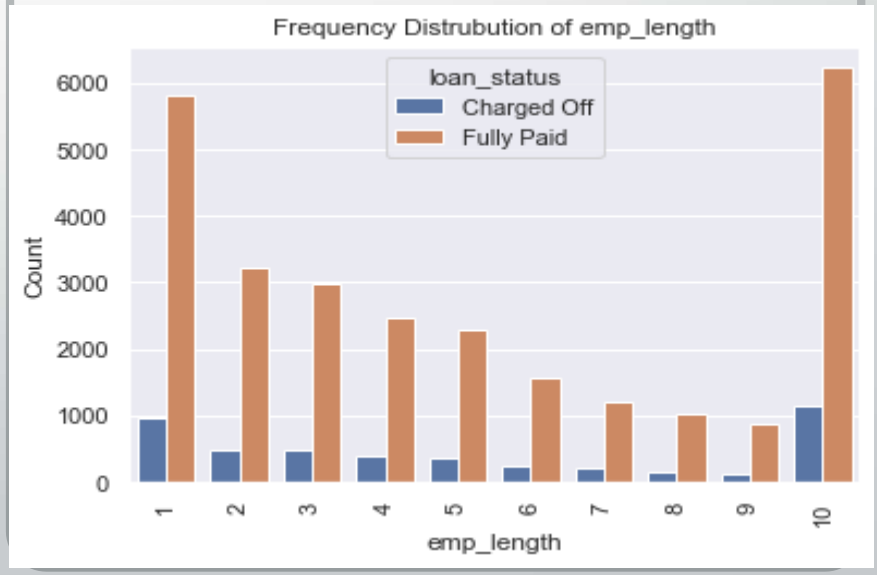
The data is divided in **Categorical** and **Continuous** columns for our analysis.

Data Analysis
Univariate

Count plot is plotted for categorical columns.

- 36 months term that is shorter term has more defaulters
- 'B' grade has high count for both fully paid and charged off
- 10 years and more has high defaulters
- Rented and Mortgage owners both have high defaulters count
- Around 5000 records have been defaulters in our dataset

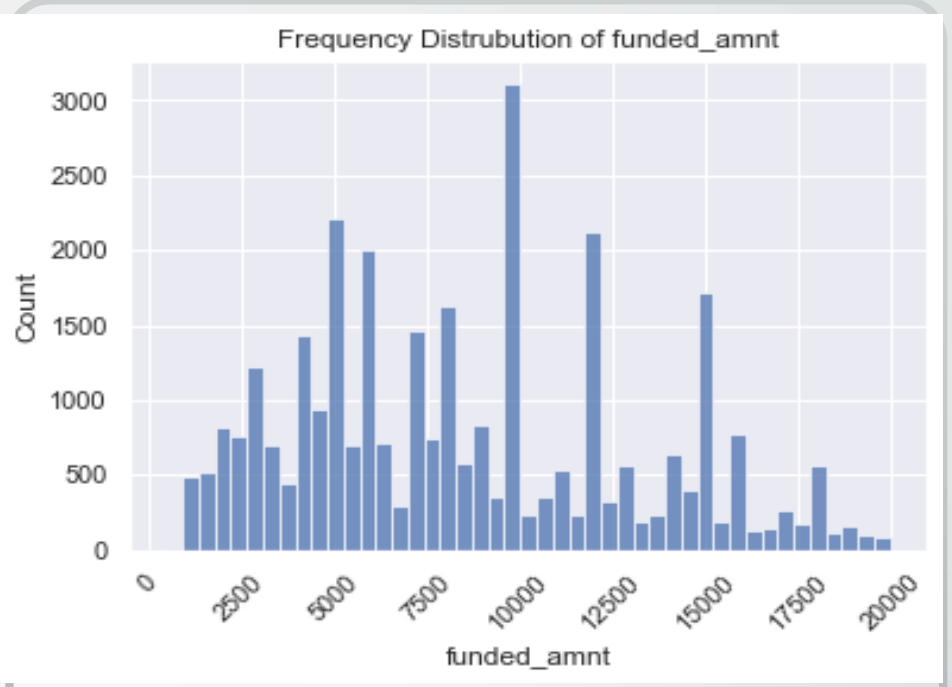


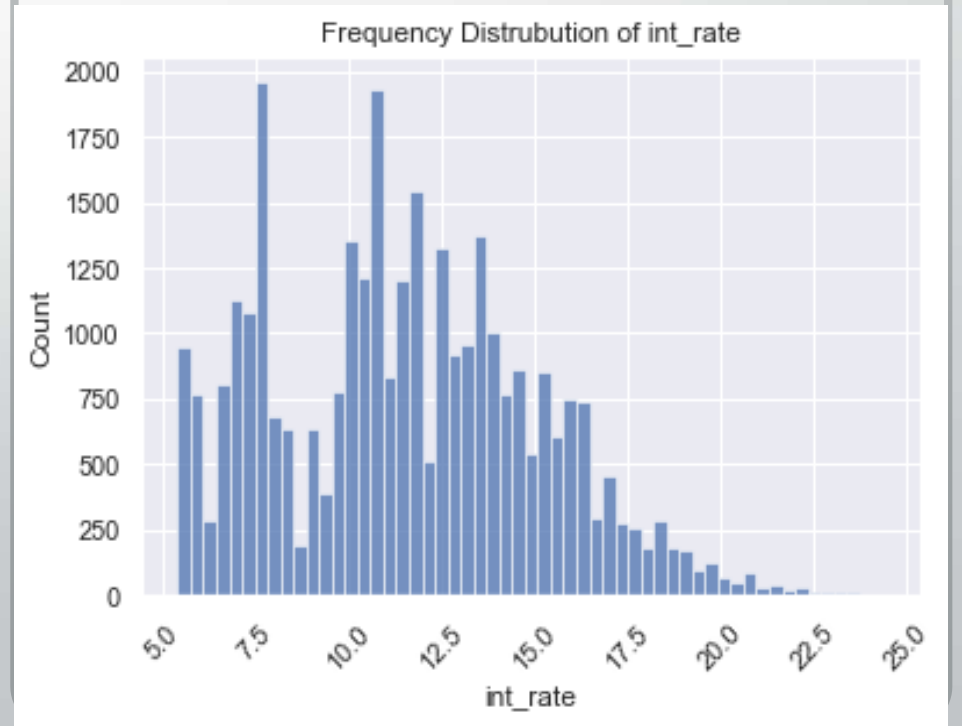


Histogram for Continuous Columns

Observations from frequency distribution of categorical columns

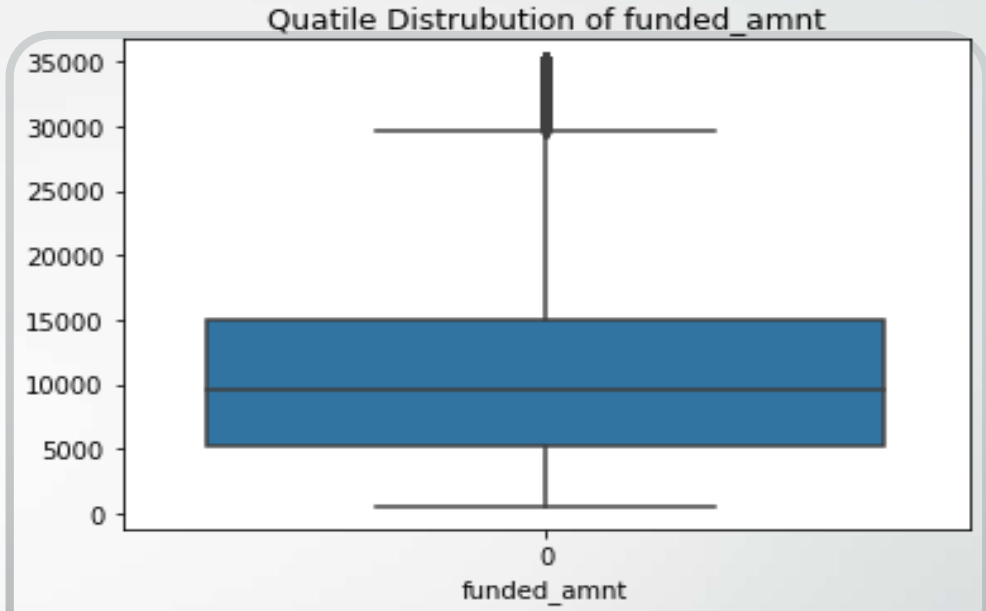
- Around 10000 is most funded amount
- 11% is the interest rate given for most of loans.
- 180 installments are mostly needed to complete the loan

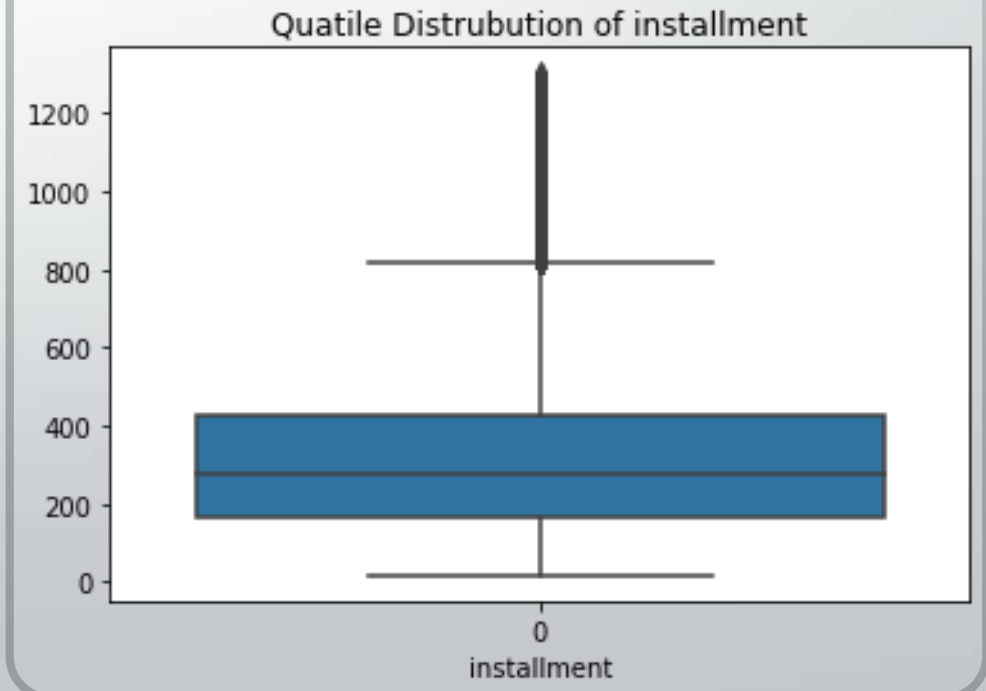




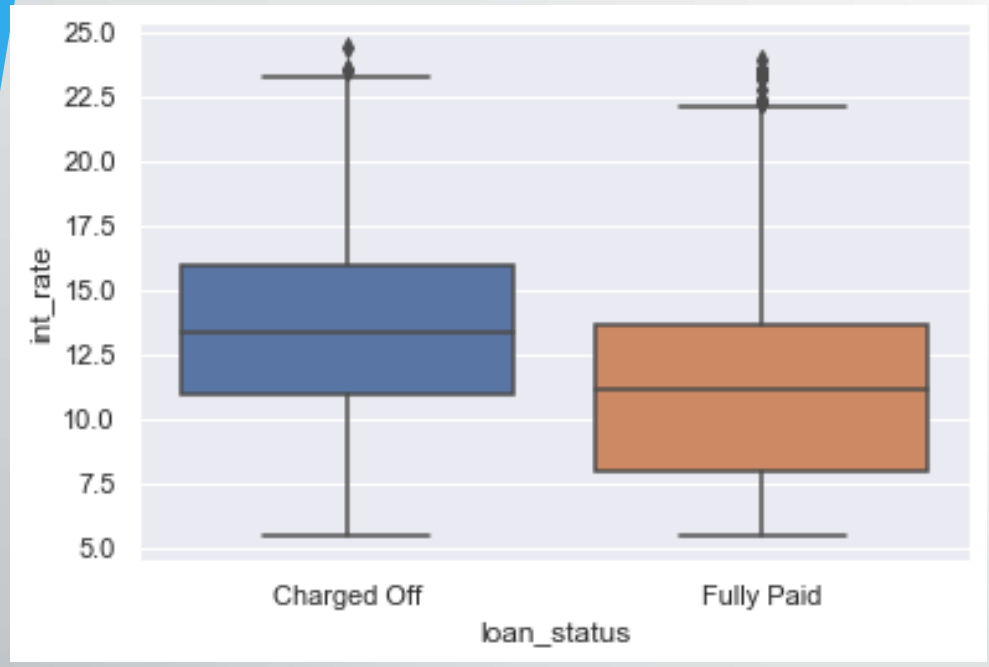
Box Plot- Continuous Variables Outliers Removal

- Most of the loan amount is in between 5400 to 30000
- There are outliers and having loan amount value more than 30000





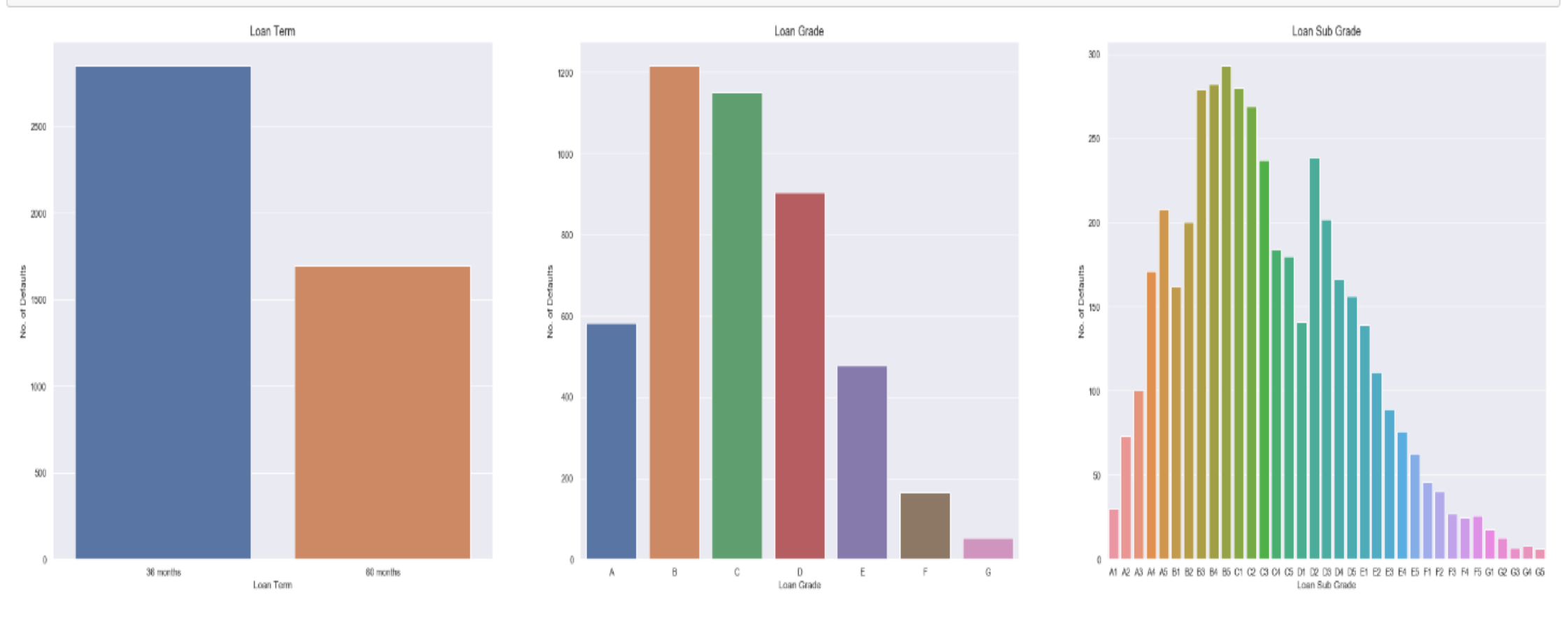
Bivariate Analysis



With high interest rate Defaulters are more.

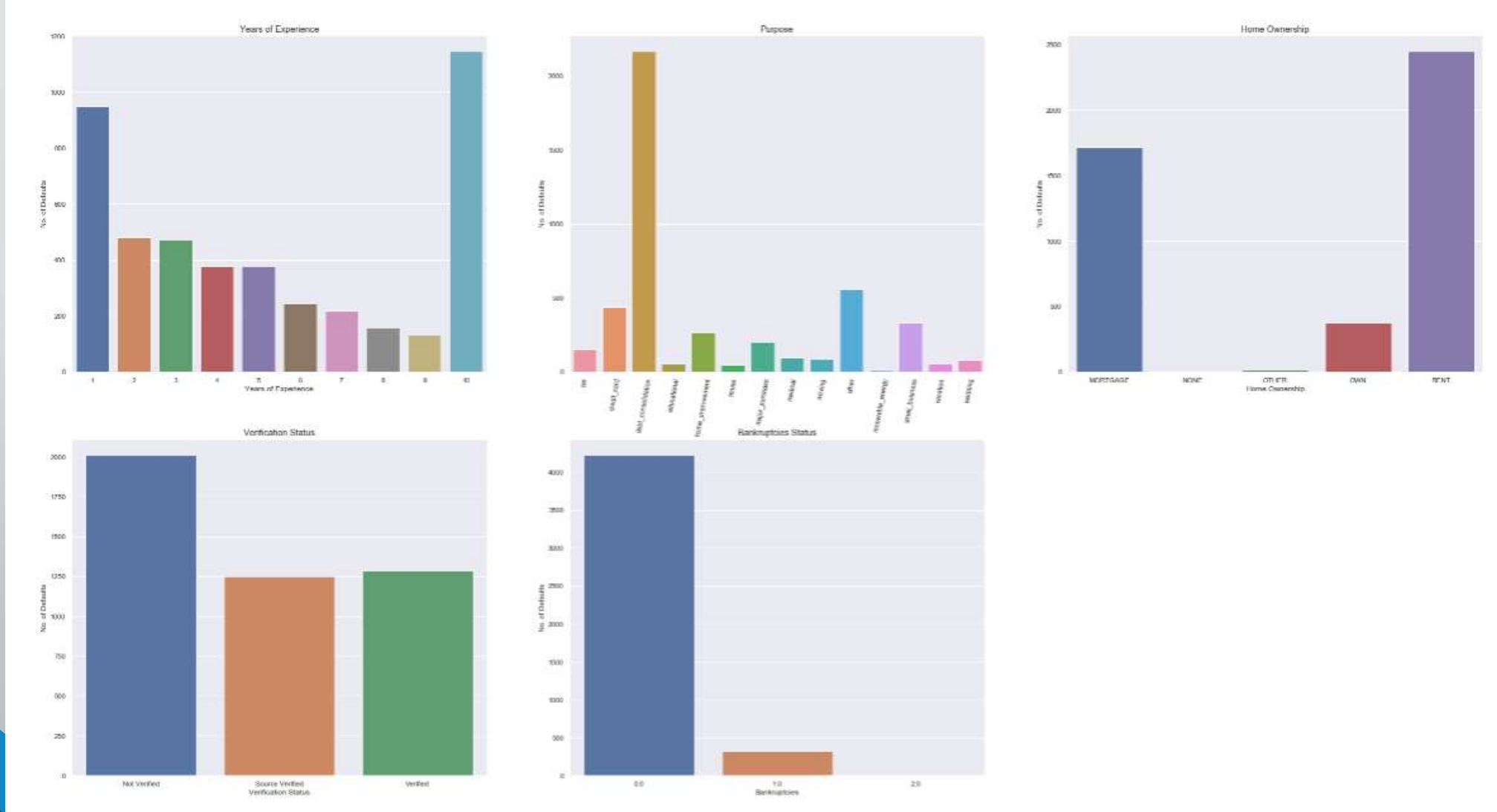
Defaulters are slightly more with debt ratio.

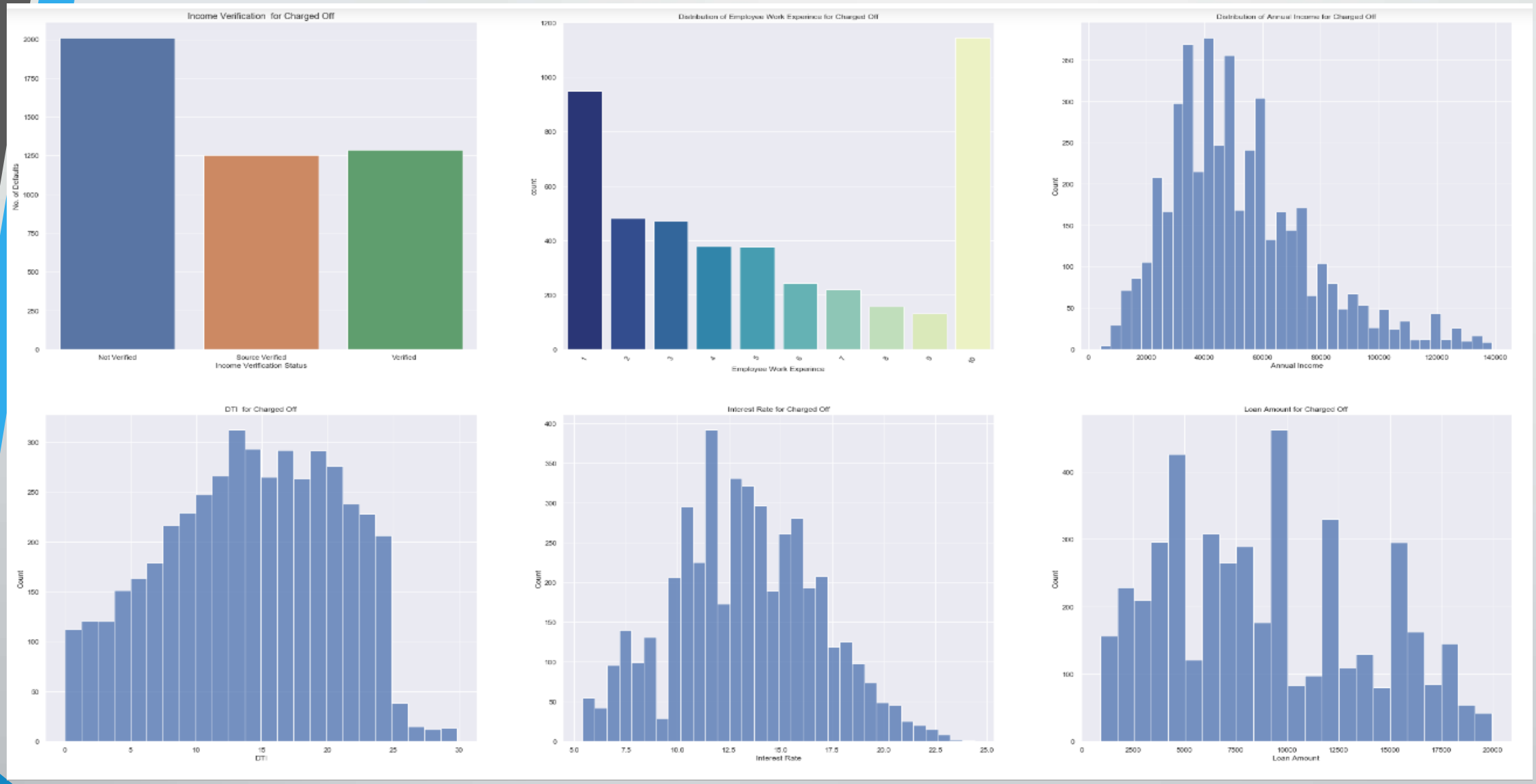
Univariate Analysis –Defaulters Data



Defaulters are more with lesser term
B and C Grade has high defaulters
B3 to C4 sub grades has more defaulters

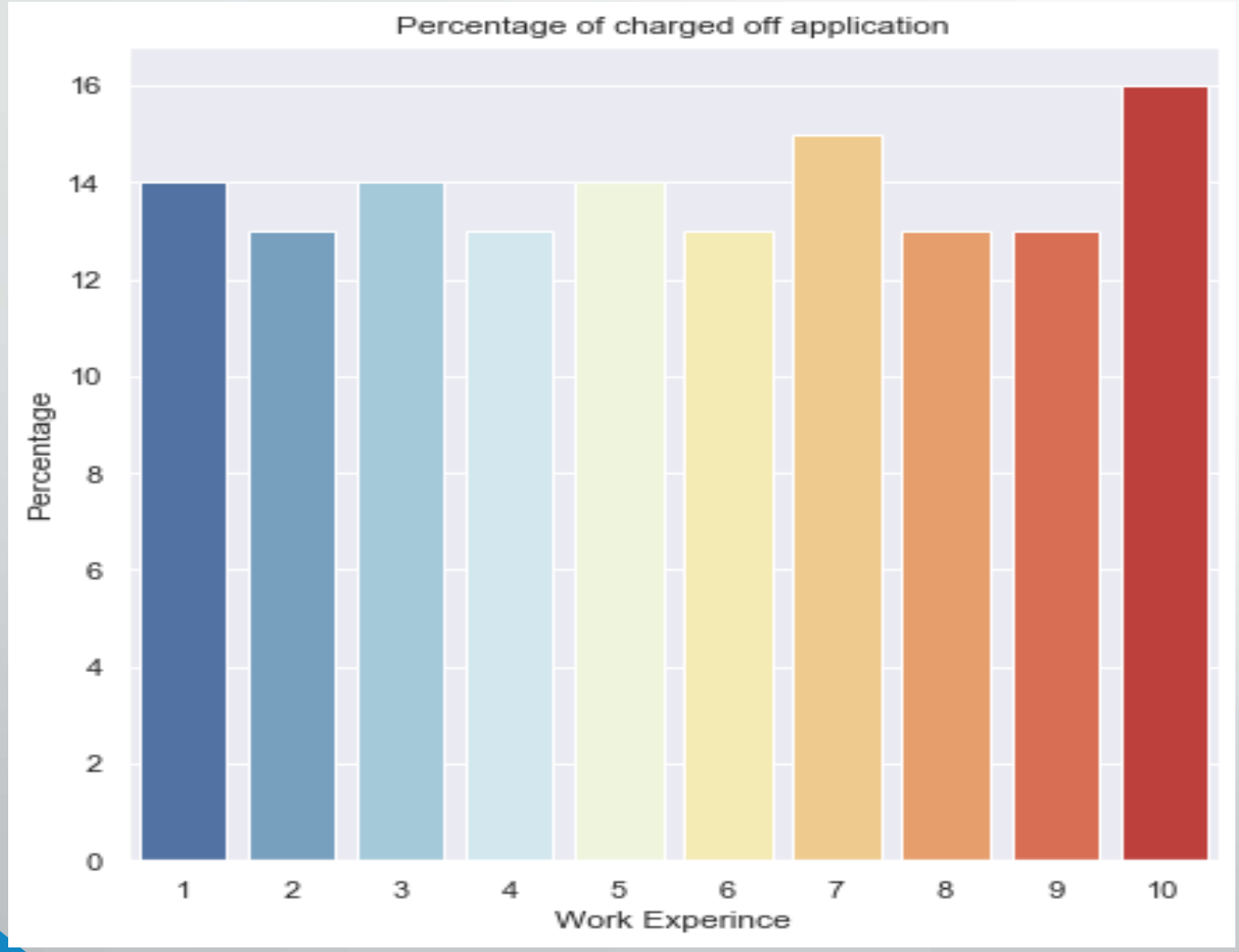
Years of experience 10 and above has more charged off frequency count.
High defaulters have purpose of debt-consolidation.
Rented ownership has more chance of charge off.
Applications which are not verified are more of defaulters.



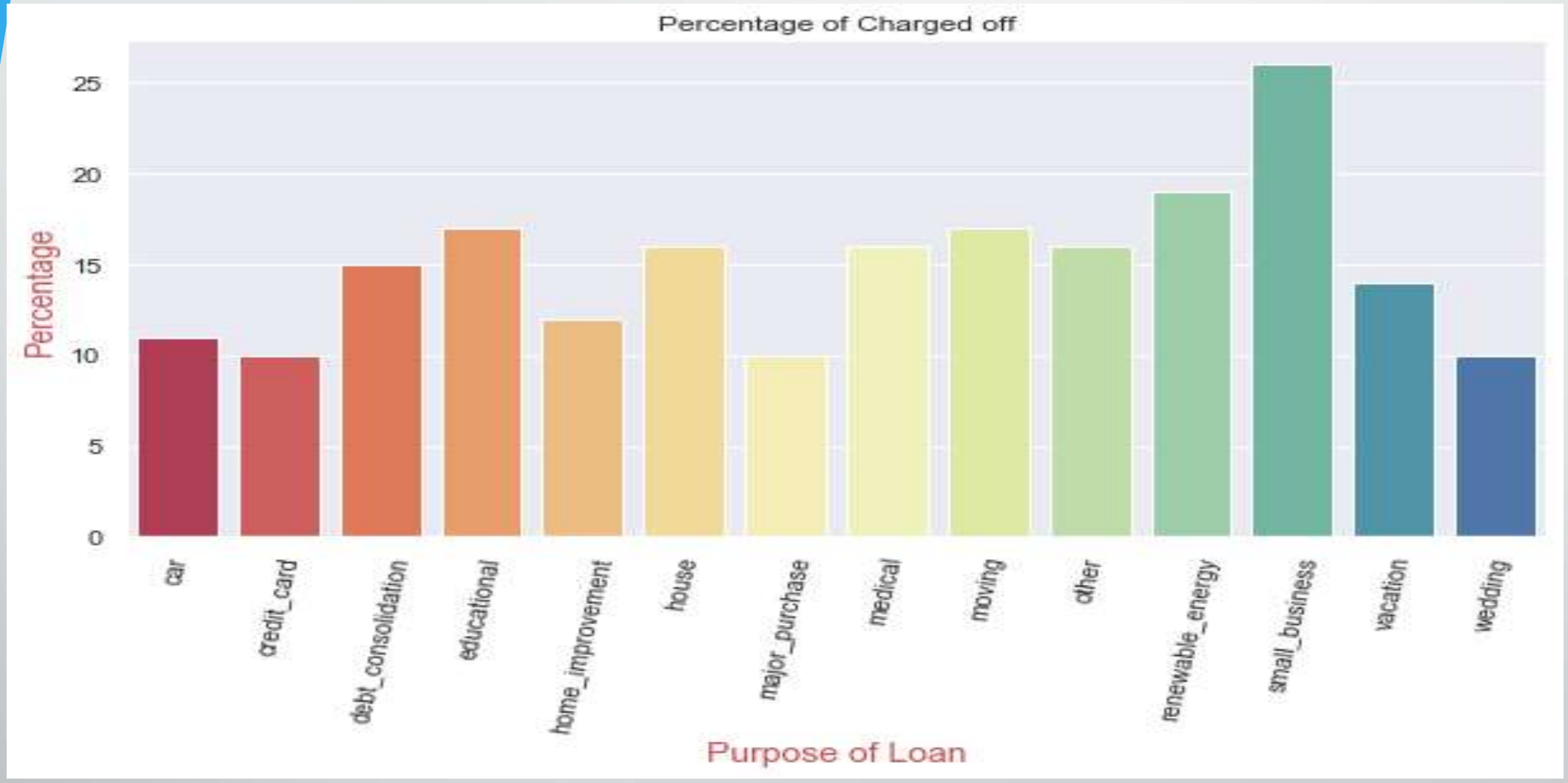


- Annual Income of charged off is in between 40000 to 70000
- DTI for charges is in between 11 to 21
- Maximum charged off is for interest rate in between 12 - 16
- 5000 - 15000 loan amount are getting maximum defaults

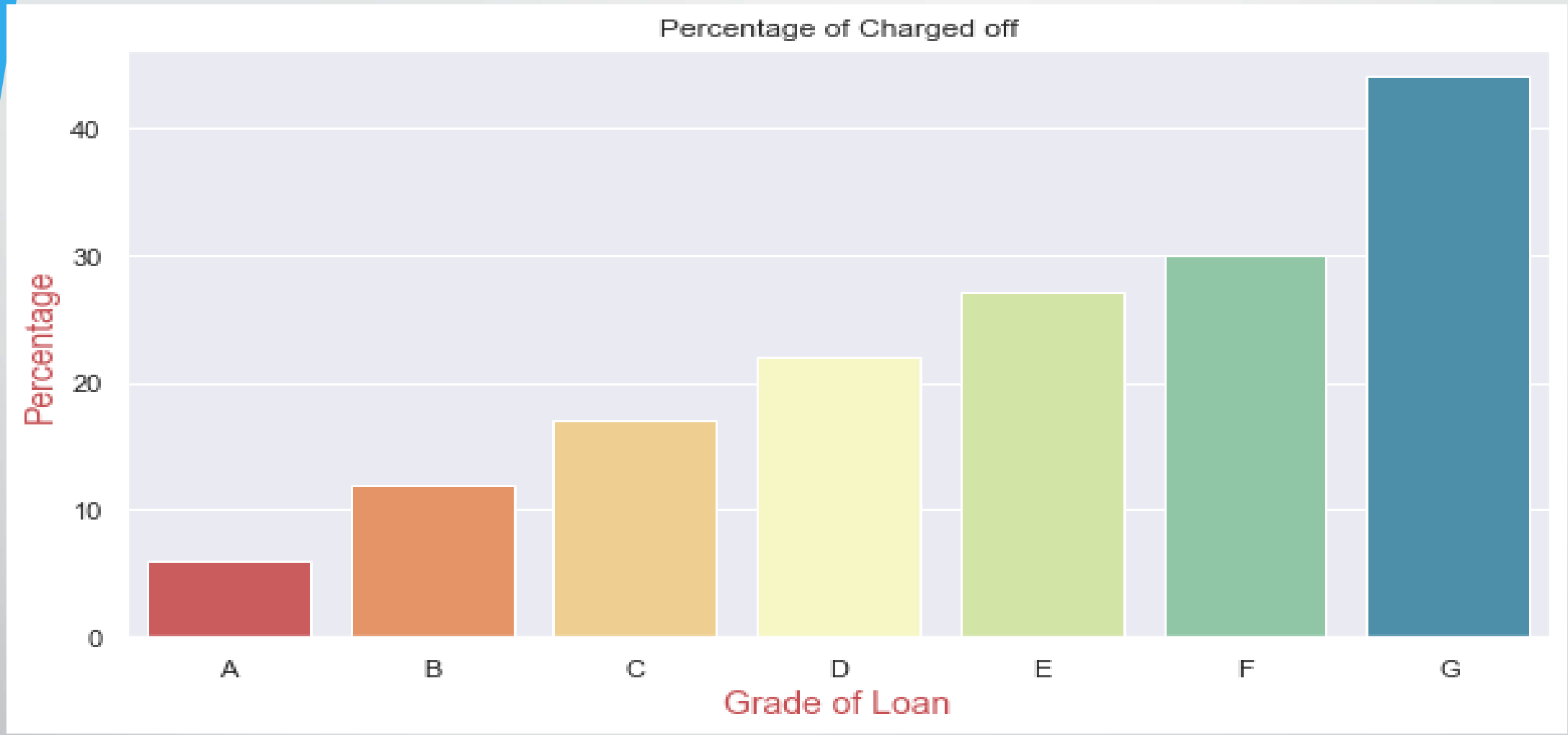
Proportion Analysis



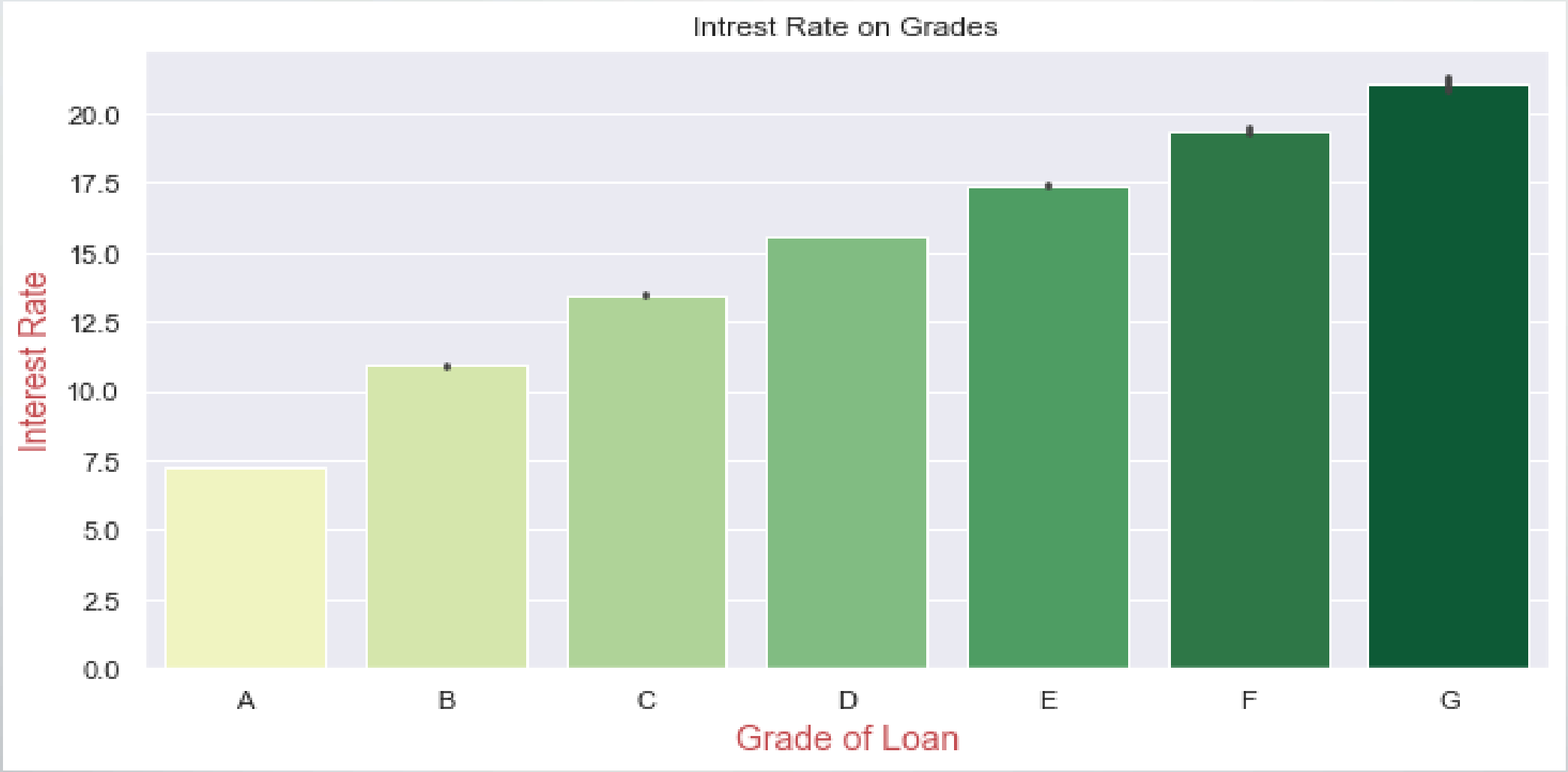
Work experience with 10 years and above, 7 years have 15-16% defaulters.



- Small business, renewable energy, education and debt consolidation has 15 to 25% defaulters.

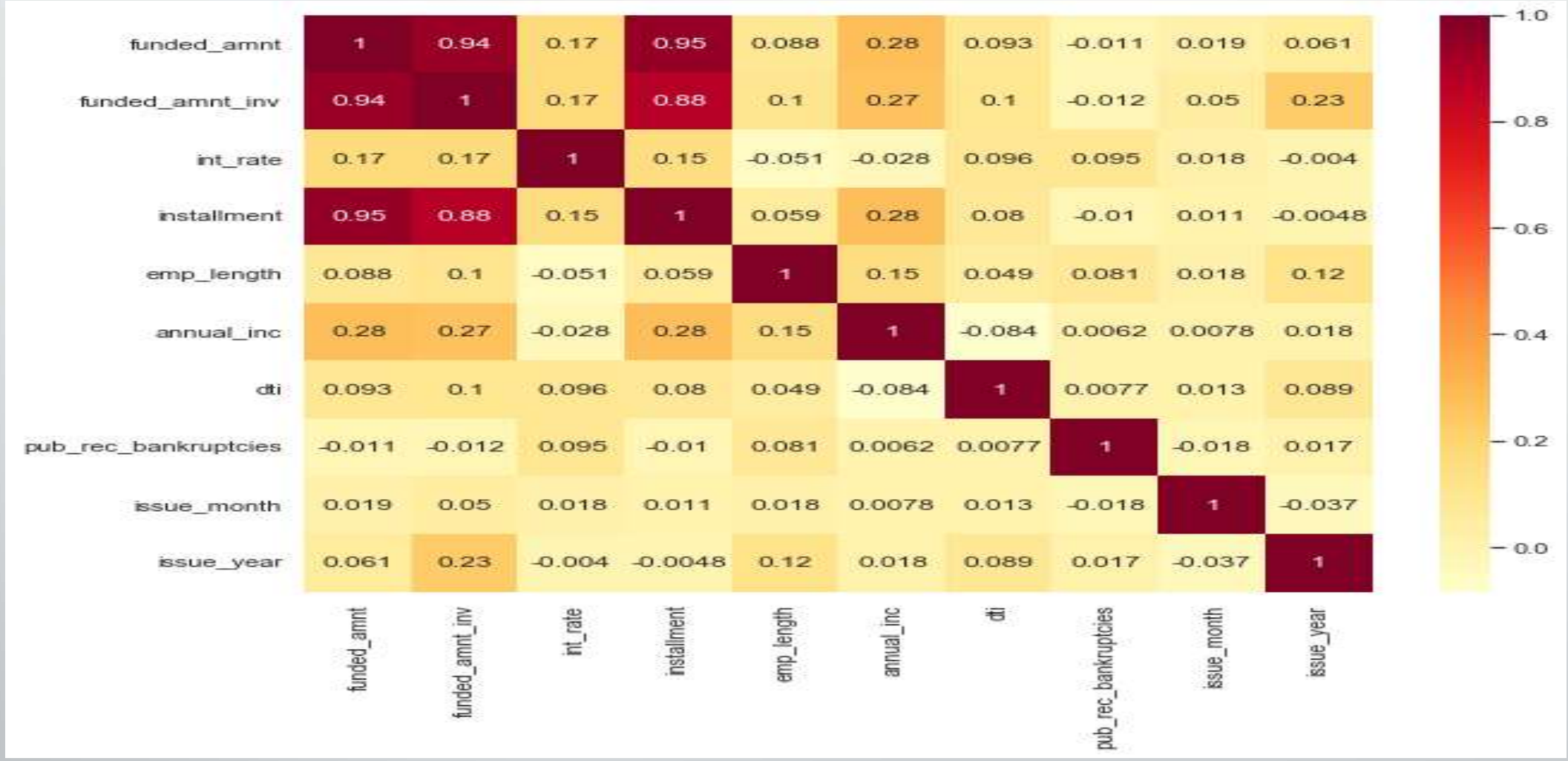


- Grades F and G has defaulters with 30 to 40% ratio.



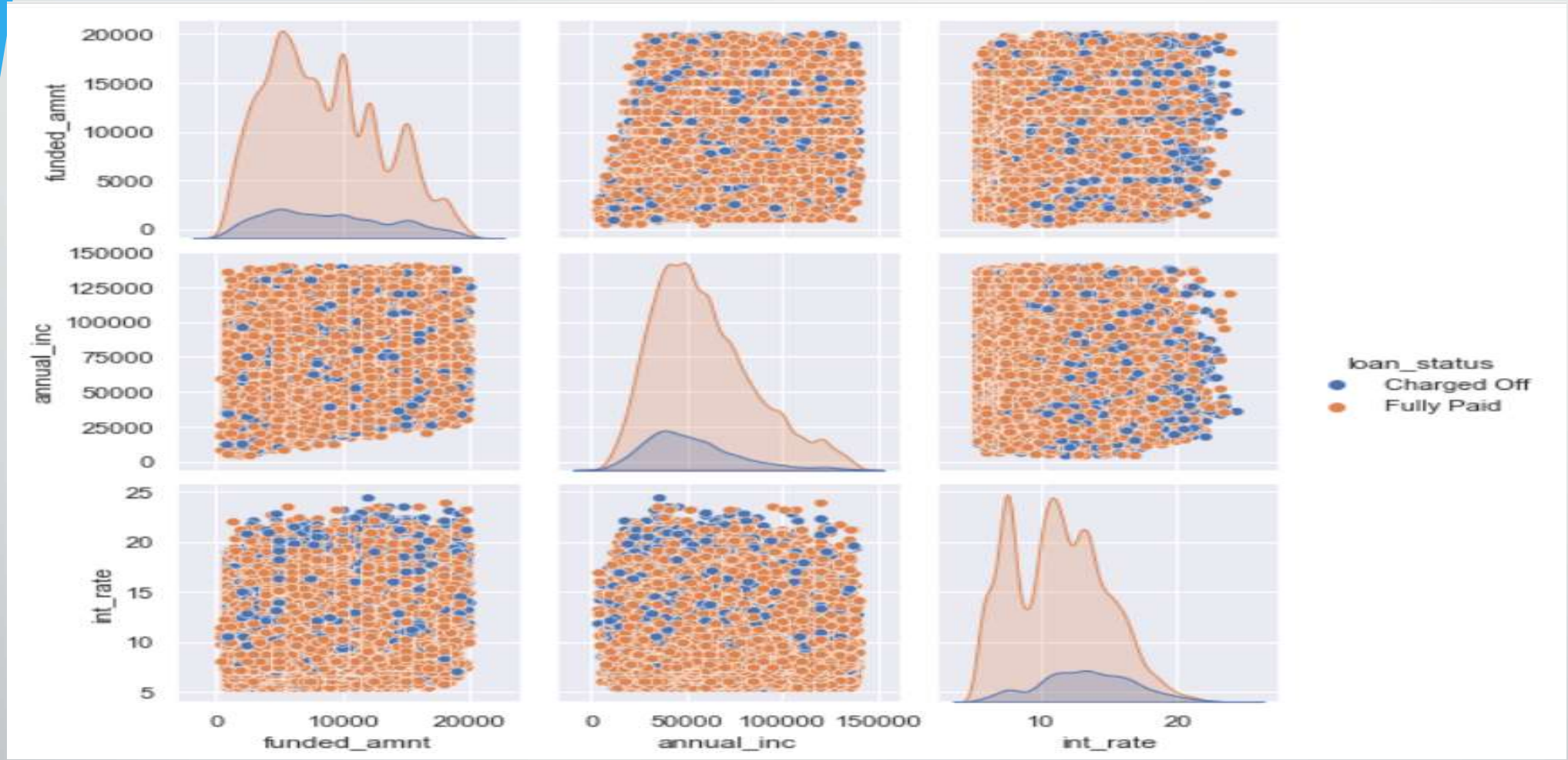
- As the Grade moves from A to G interest rate also increase.

Correlation matrix



- Funded amount and instalment has high correlation.

  Interest Rate, experience of employee and annual income has negative correlation



- Applications having higher Interest rate and Higher Loan Amount having more chances of being default.
- Application having higher Interest rate and lower Annual Income having more chances of being default.

Insights

- Shorter term of loan has more defaulters, but higher term has more chances.
- Rented and mortgaged house owners have high default rate
- Annual Income of charged off is in between 40000 to 70000
- Debt to income ratio(DTI) for charges is in between 11 to 21
- Maximum charged off is for interest rate in between 12 - 16
- 5000 - 15000 loan amount are getting maximum defaults
- Purpose of Debt consolidation has maximum defaults and loan for small business has more chances to be default.
- As the Grade move from A to G the number of defaulters also increase, subgrade also have similar behavior
- Applications having higher Interest rate and Higher Loan Amount having more chances of being default.
- Application having higher Interest rate and lower Annual Income having more chances of being default.

Recommendation

- Increase term of loan to reduce the amount of defaulters with lesser funding amount.
- Prefer own house owners to give loan to rather than rented or mortgaged
- Consider lower loan amount and more term for the applicants in grade and subgrade in F and G
- Rented and Mortgaged home owners loan for debt consolidation or home repair may not be preferable.
- Small businesses may go to default.