Lending club case study

Group Members:

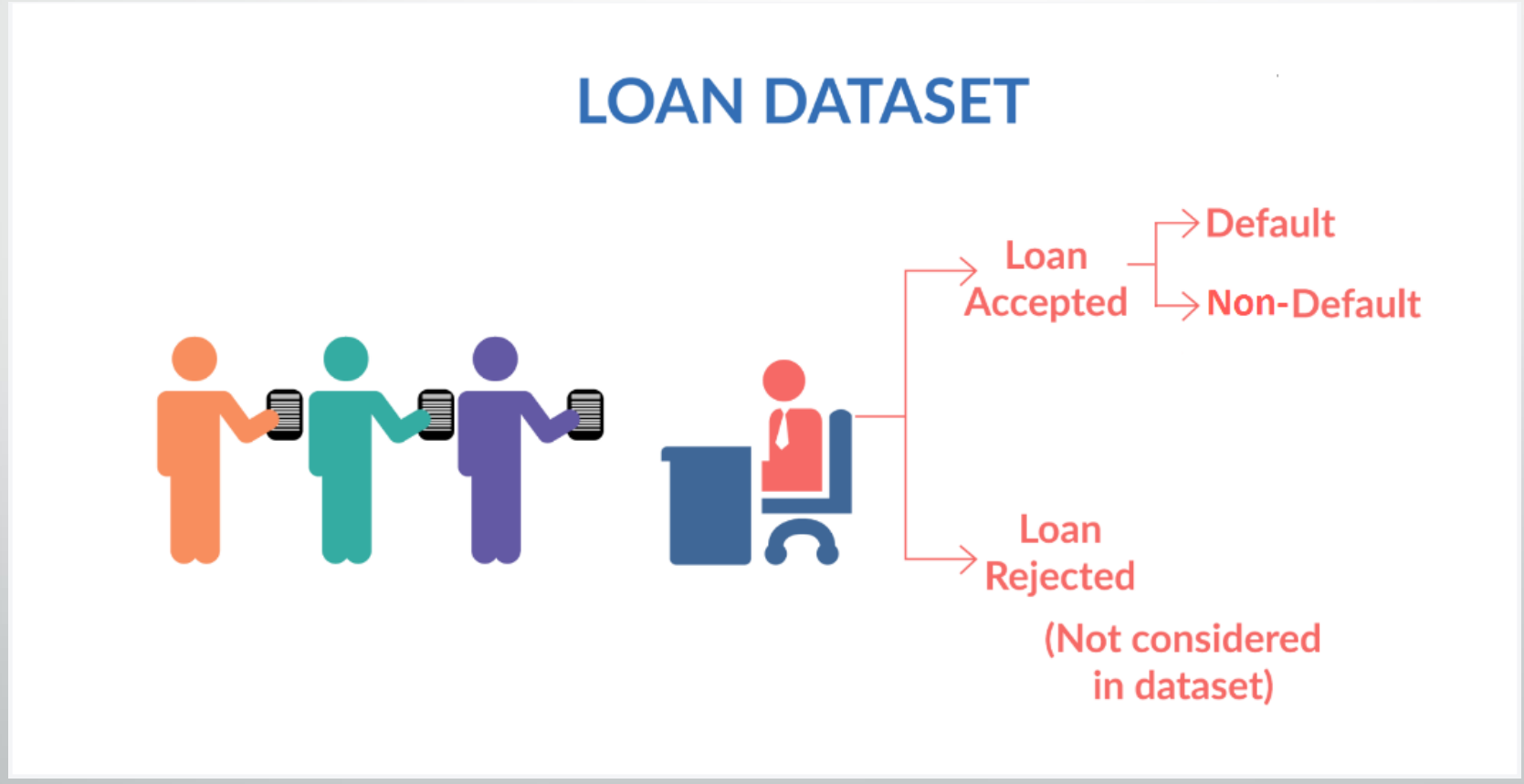
Shamseena VM

Rajesh Kodavalli

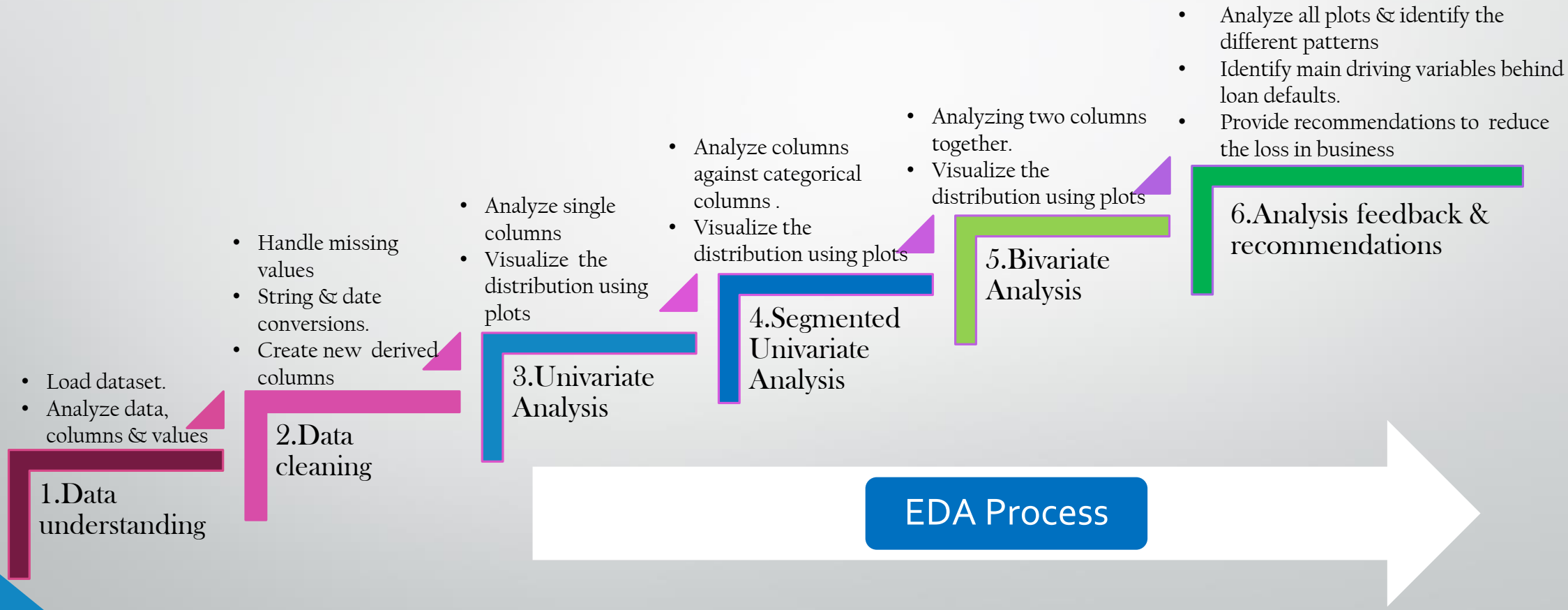
Problem statement

- Lending club case study assignment is to provide an idea about how real business problems are solved using EDA
- It helps in understanding risk analytics in banking and financial services.
- To understand how data is used to minimize the risk of losing money while lending to customers.
- The case study is conducted to analyze the information about past loan customers and understand if they ‘defaulted’ or not.
- To understand the driving variable which causes the customer to default the loan payment

Overview of Loan Applied

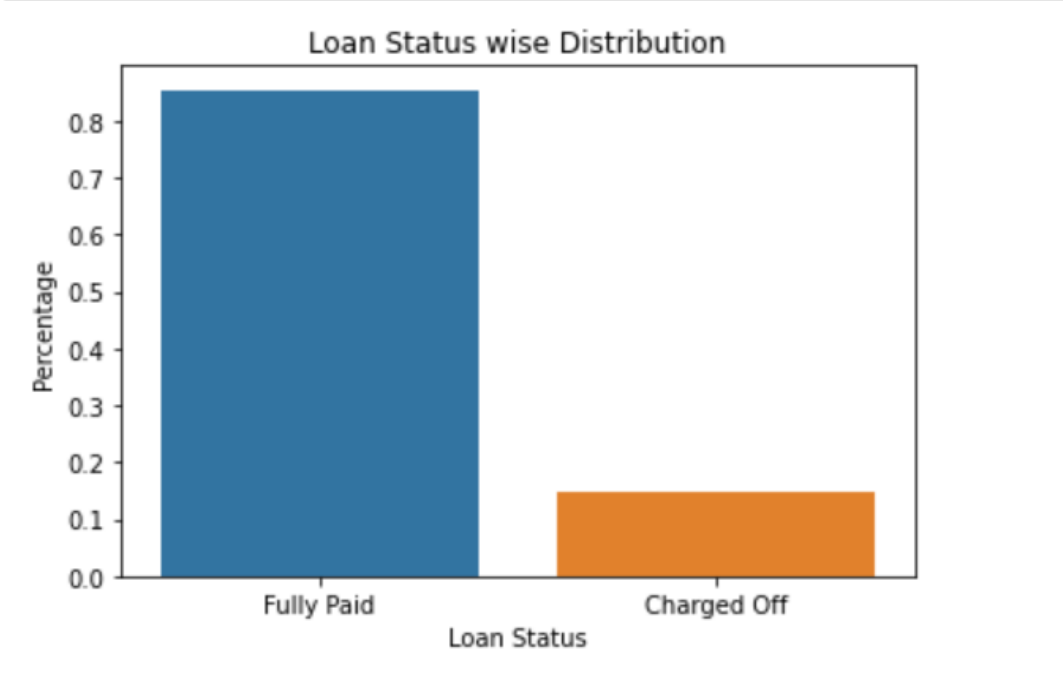


CASE STUDY APPROACH



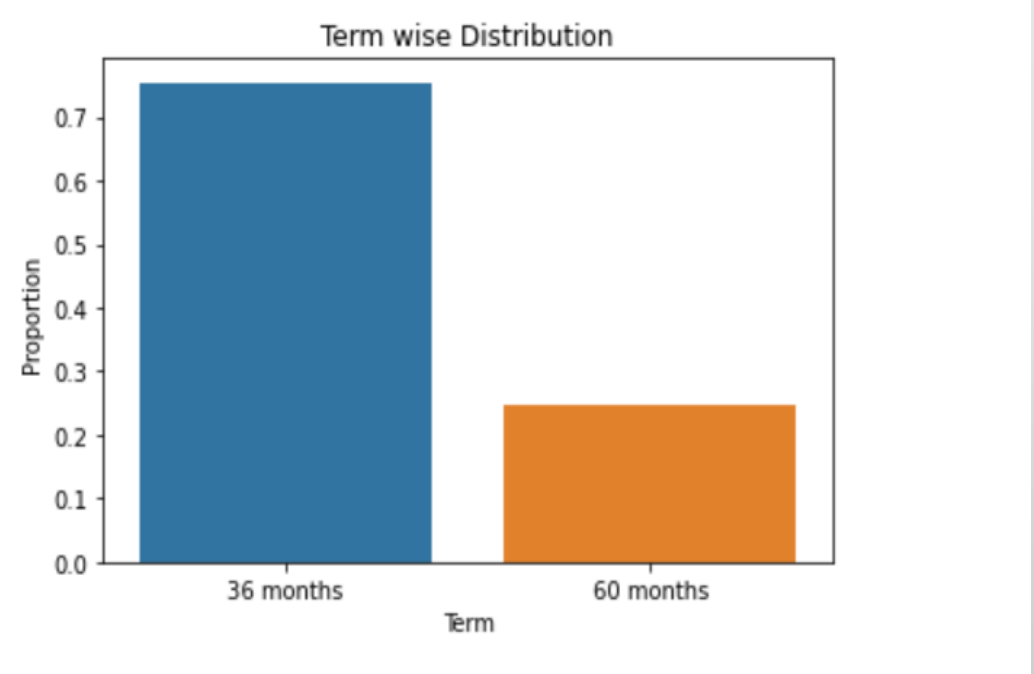
Univariate Analysis

Loan Status column



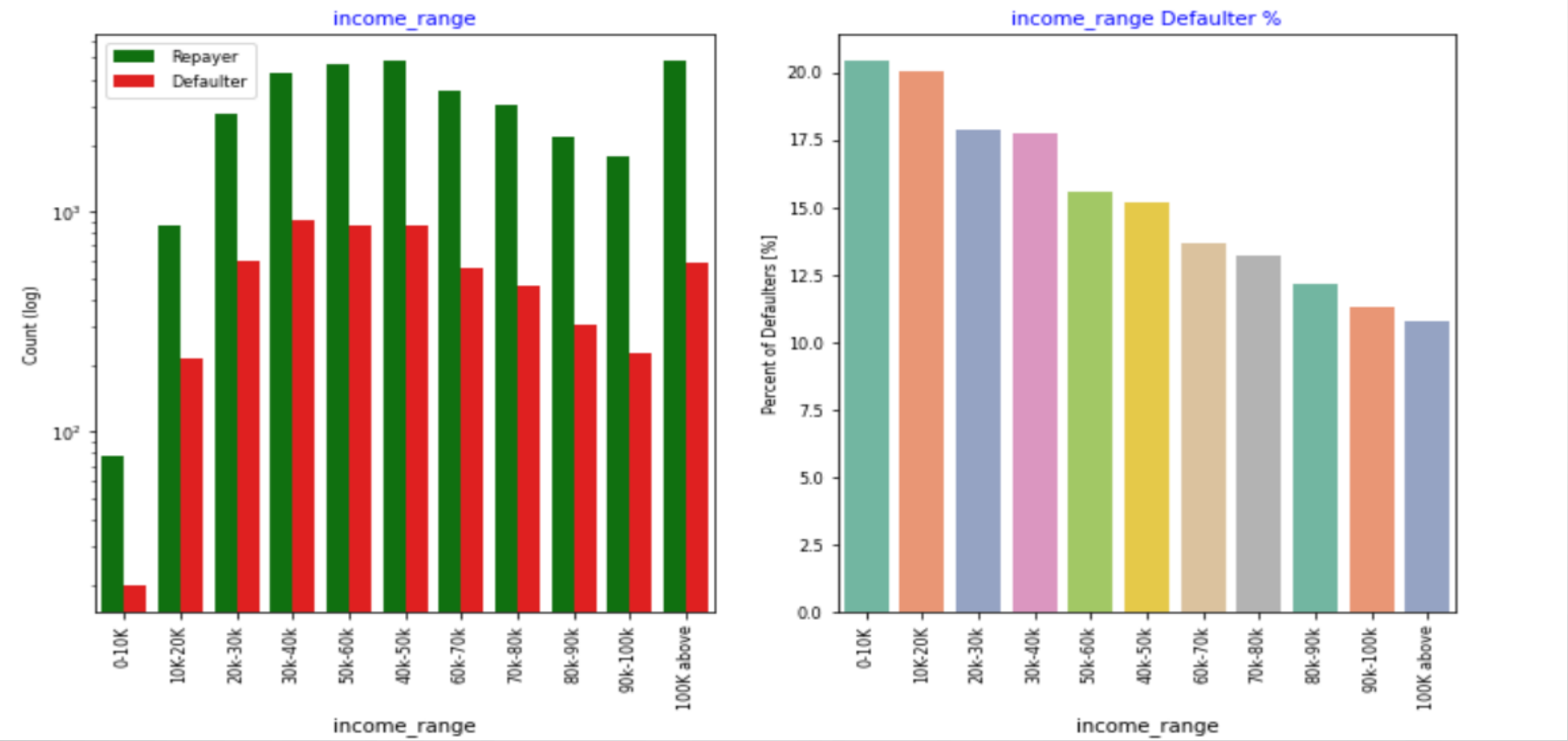
- More then 80 % of customers are fully paid the loan
- Almost 15% of customers charged off/defaulted

Term column



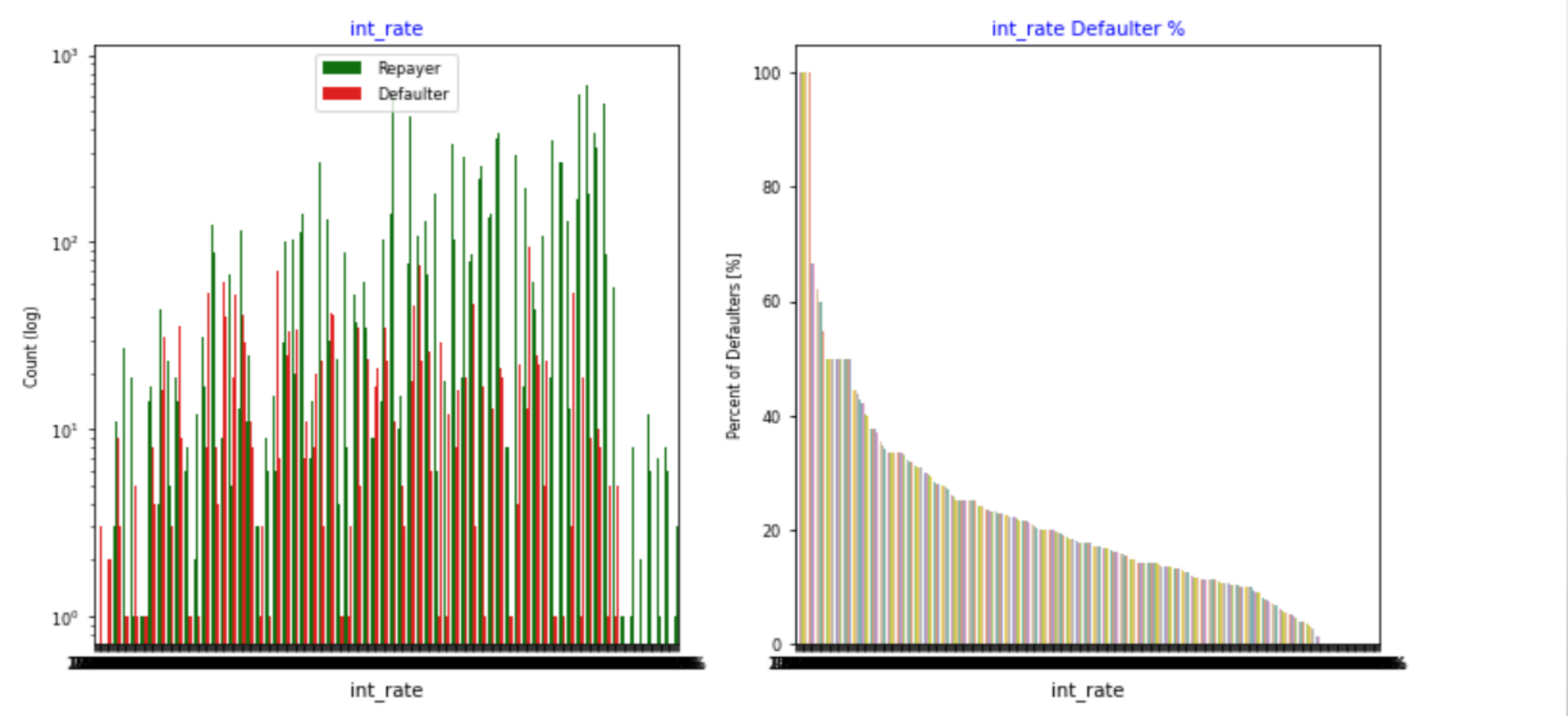
- More then 75 % of customers are fully paid the loan who opted 36 month term for loan payment
- Almost 24% of customers charged off/defaulted who opted for 60 month term for loan payment

Univariate Analysis



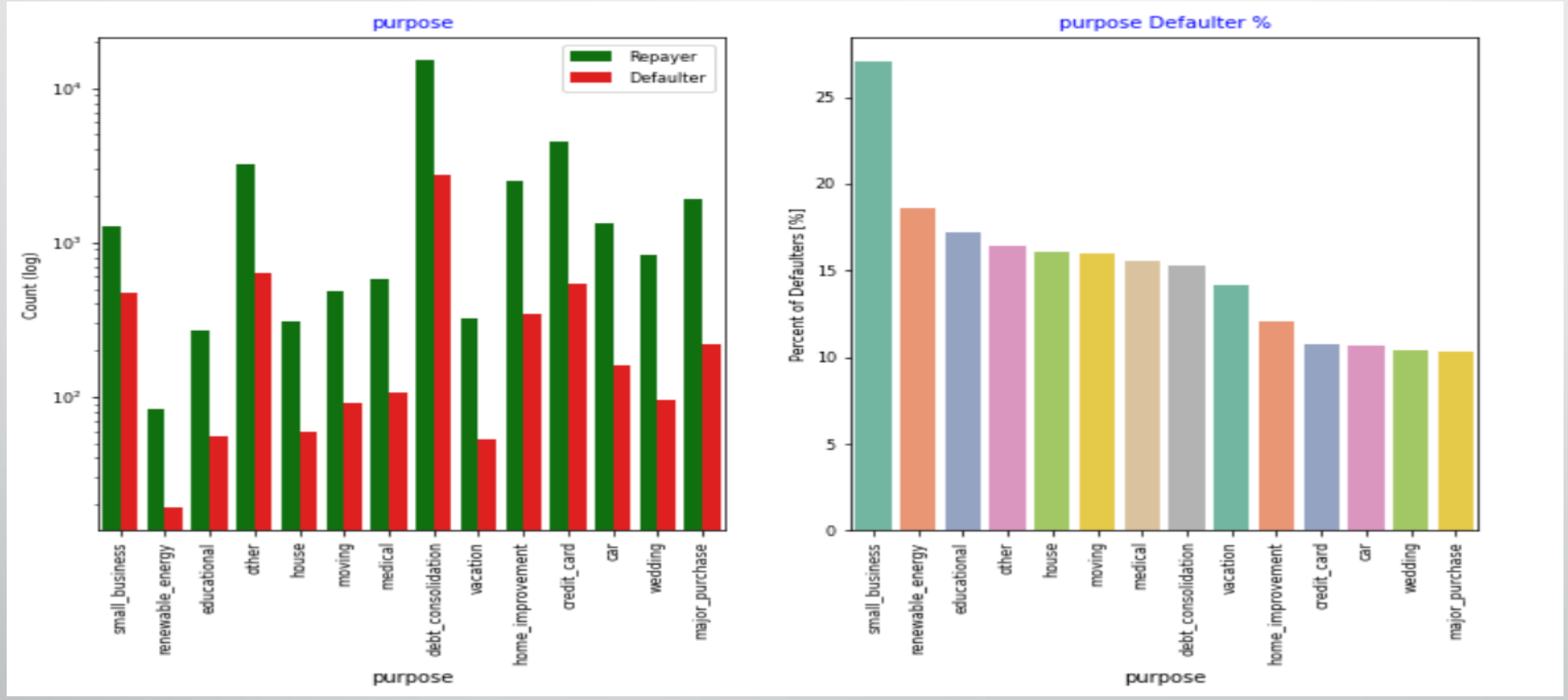
The customers with income below 20k higher default percentage(~21%)

Univariate Analysis



With higher interest rate, higher the default rate

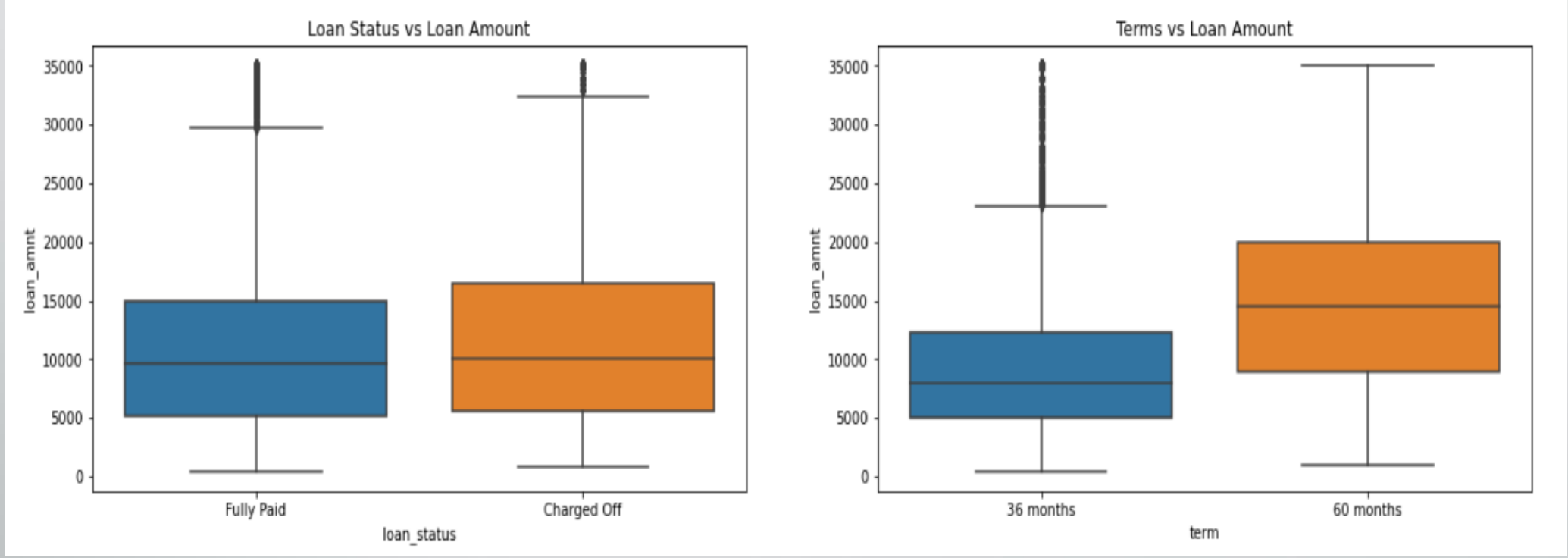
Univariate Analysis



The customers who take loan for samll business have high default percentage(~27%)

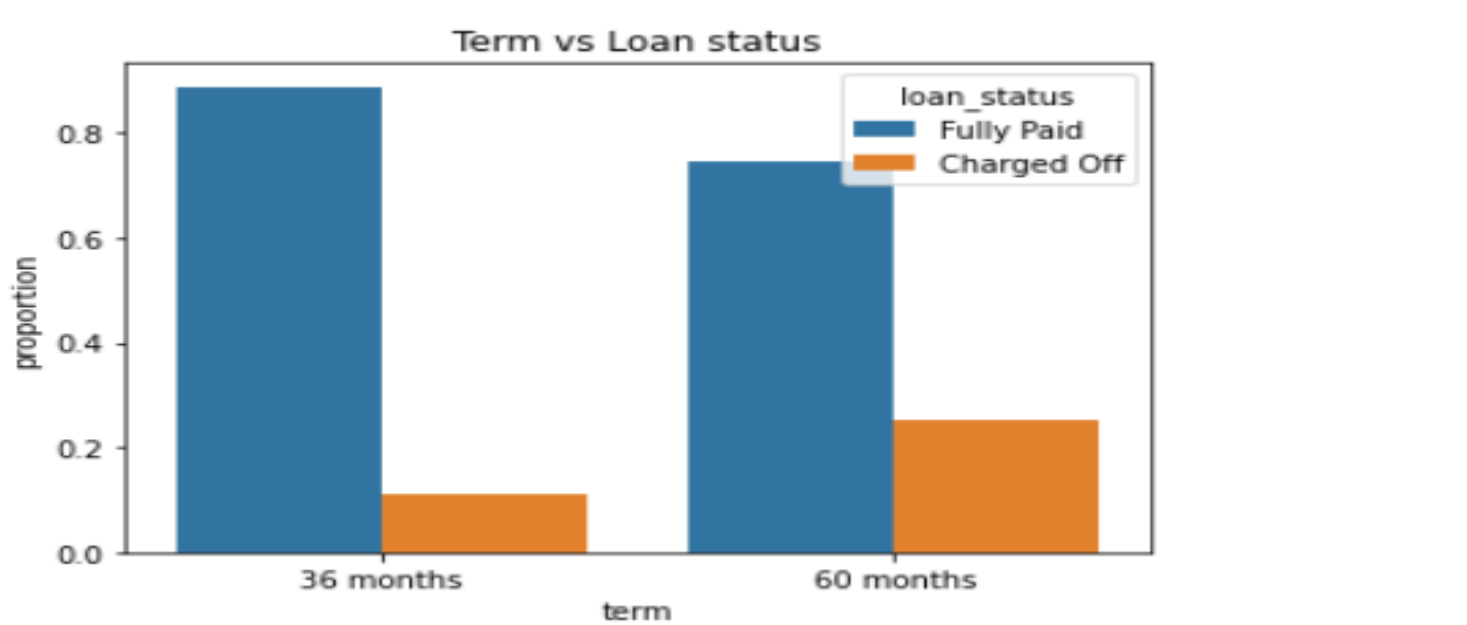
Segmented Univariate Analysis

loan_status vs loan_amount & term vs loan_amount column



Bivariante Analysis

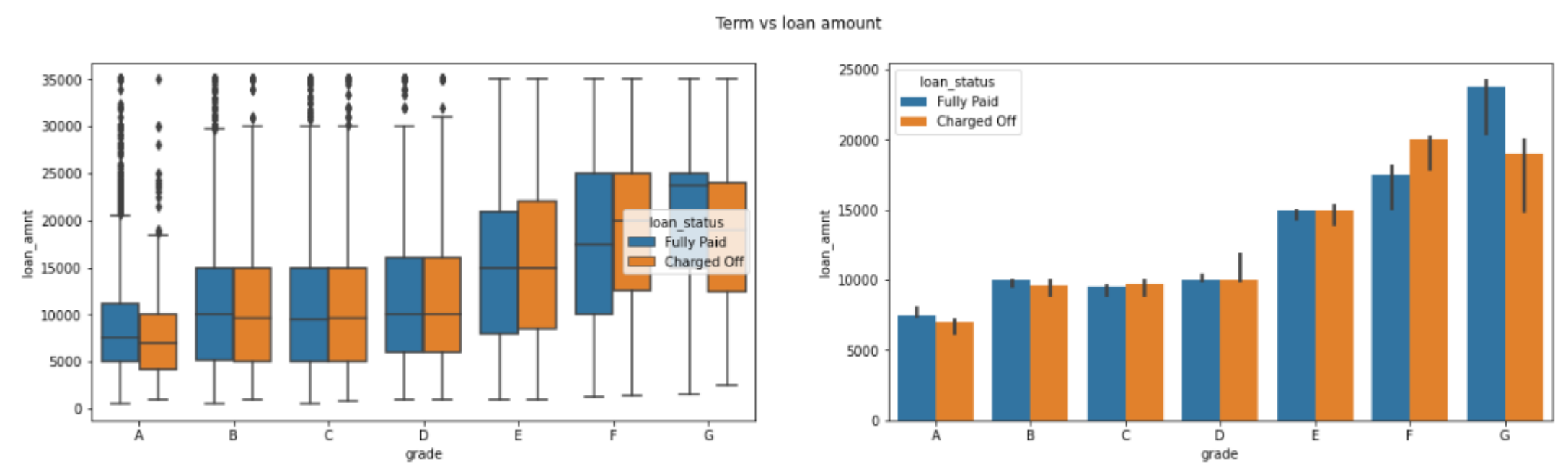
Term vs loan_status



The default rate is high in 6o months tenure because most people took high loan amount with high interest rate in it and they faced difficulties in returning the sum to bank

Bivariate Analysis

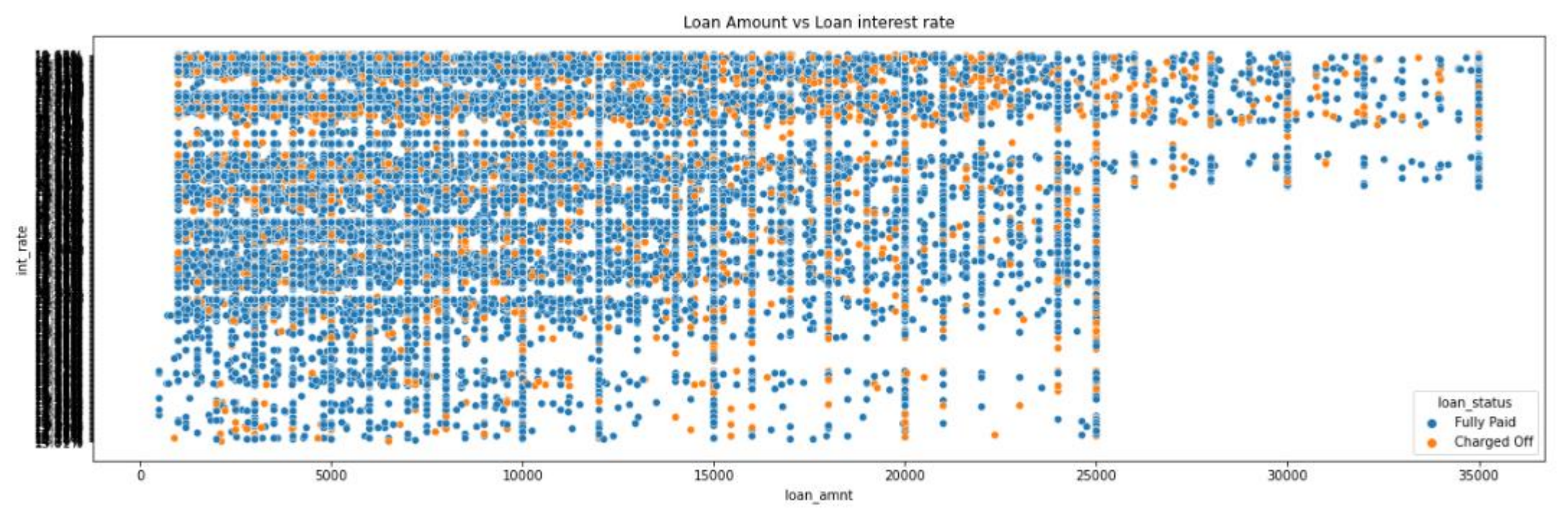
Term vs loan_amount



- For lower grades 'F' and 'G' there are more difference between charged-off and fully paid.
- The lower grade people has taken higher amount of loans and also they are more prone to default the loan.

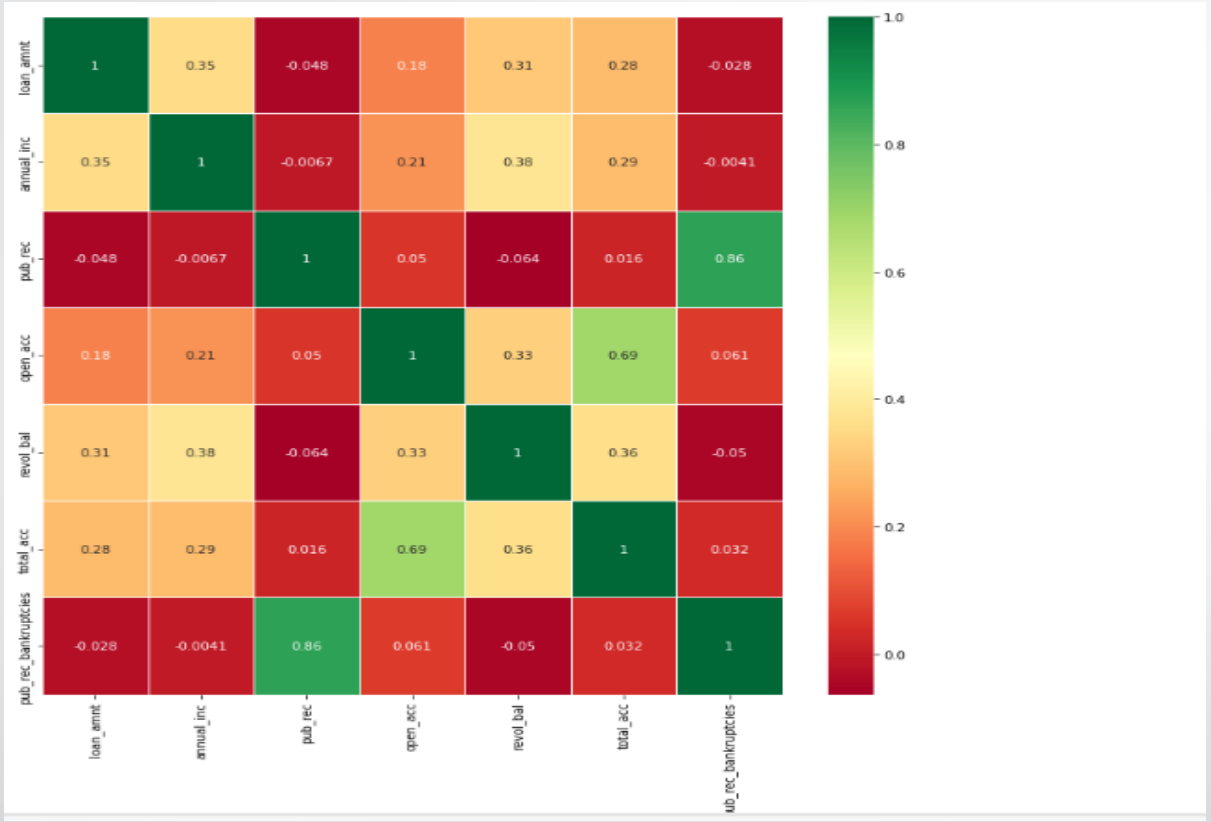
Bivariate Analysis

Loan amount vs interest rate



- For lower grades 'F' and 'G' there are more difference between charged-off and fully paid.
- The lower grade people has taken higher amount of loans and also they are more prone to default the loan.

Correlation



- The above figure showsthe correlation between of defaulted customer

Main driving variables for customer default

- Term
- Loan amount
- Purpose
- Interest rate
- Income

Conclusion

- We can work on the Data with fully paid and Defaulted customers base.
- We can create model on the above data to figure out the driver variables of default customers.
- The model from above steps can be used to figure out the probability of an existing customer/ New customer to default