

EDA Case Study

SUBMISSION

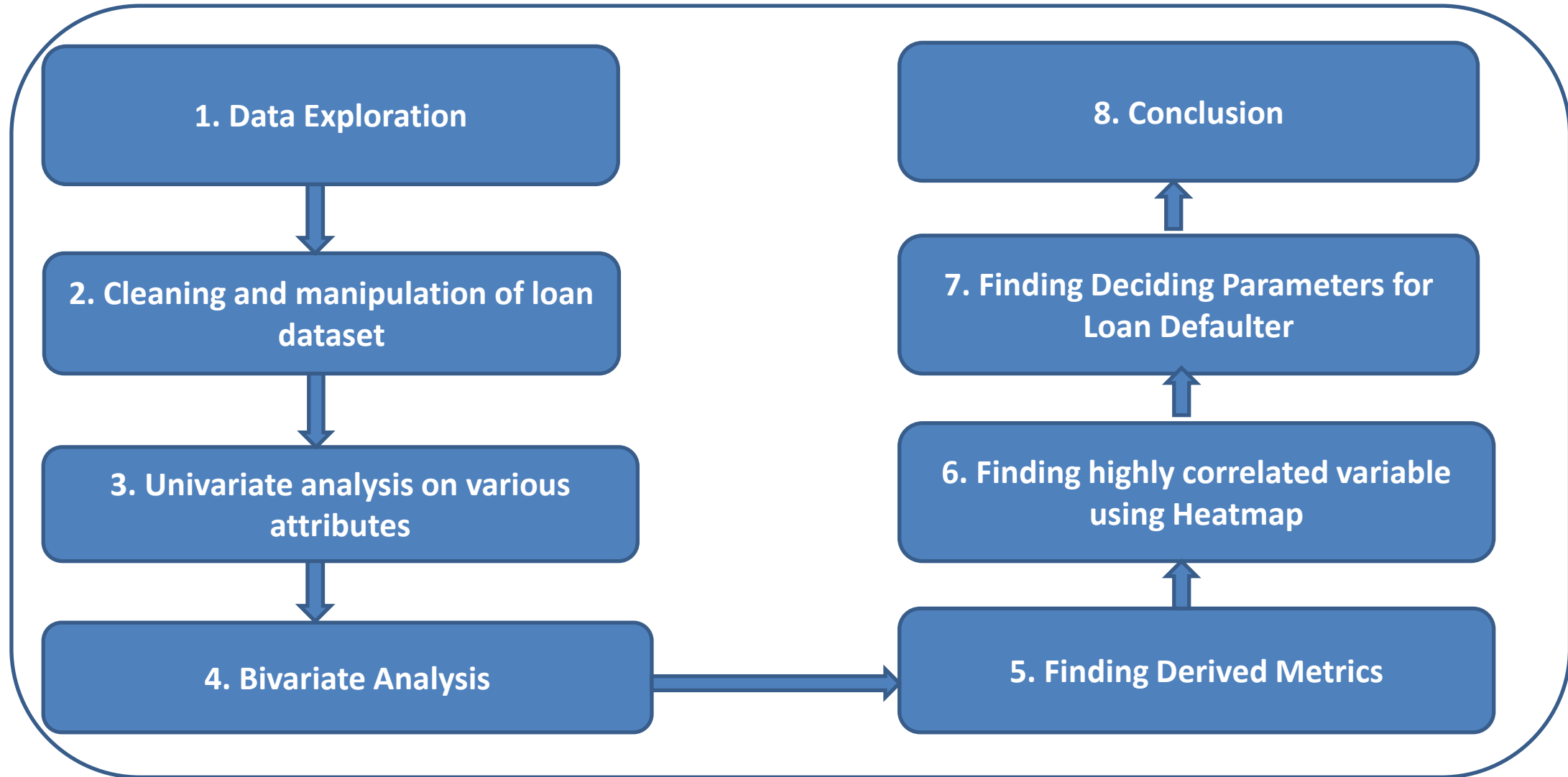
Presented By:

Swami Prem Pranav Kayashyap

Abstract – Loan Risk Analysis

Business Objective:

- To understand how **consumer attributes** and **loan attributes** influence the tendency of default.
- To identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.



Data Exploration

- There are 111 variables in the loan dataset.
- Many attributes contains NA's, 0 or single value
- There are total 39717 records in the dataset.

Data Cleaning and Manipulation

Dealing with Missing values :

- In this dataset 54 variables contain all the observations as NA's so, they are removed.
- 4 columns have all value either 0 or NA's, They are removed
- 4 columns have more than 50% of NA's, so they are removed.
- If some column contain missing values, then we replaced it with the median value.
- 6 columns from remaining has same data in column hence they were removed.
- Date is converted into standard format and % is removed from columns wherever required.
- Removed the variables having too many levels like title.

Univariate Analysis

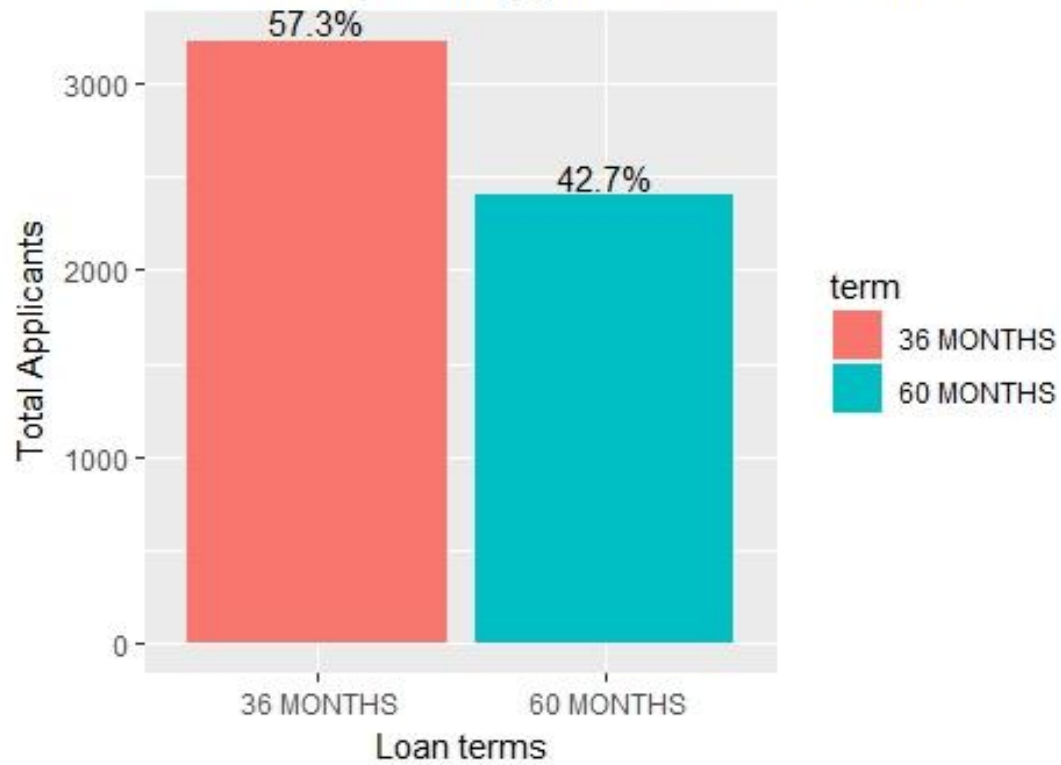
Observations in Univariate Analysis :

- No of Charged off applicants for 36 Months Loan Terms are higher than those for 60 months. Applicants for 36 months=57.3% & 60 months=42.7%.
- No of Applicants for Grade B is the most significant among all with value of 25.3%.
- No of Applicants for Sub-Grade B5 is the highest among all with value of 6.33%.
- The professionals with 10+ years experience secure the highest percentage i.e. 22.6% of the whole population of applicants.
- Nearly half of the population(50.5%) of Charged off applicants are observed to be staying in Rented house.
- Not Verified Charged off applicants are higher in number than verified applicants. Not verified = 38.1% & Verified = 36.4%.
- Nearly half of the Charged off Applicants(49.2%) has the purpose of Debt Consolidation for being charged off.
- The highest value of Applicants with Open account is 557.

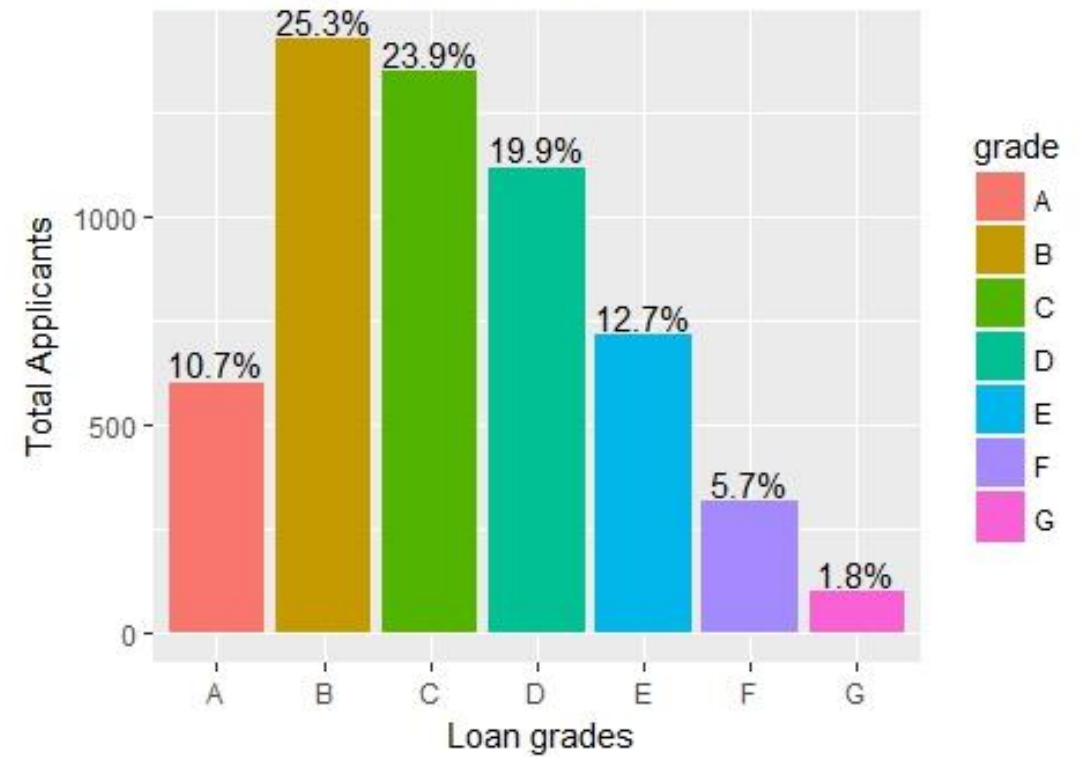
Some graphical analysis are given in following slides for better references.

Univariate Analysis

No of charged off applicants Vs Loan terms

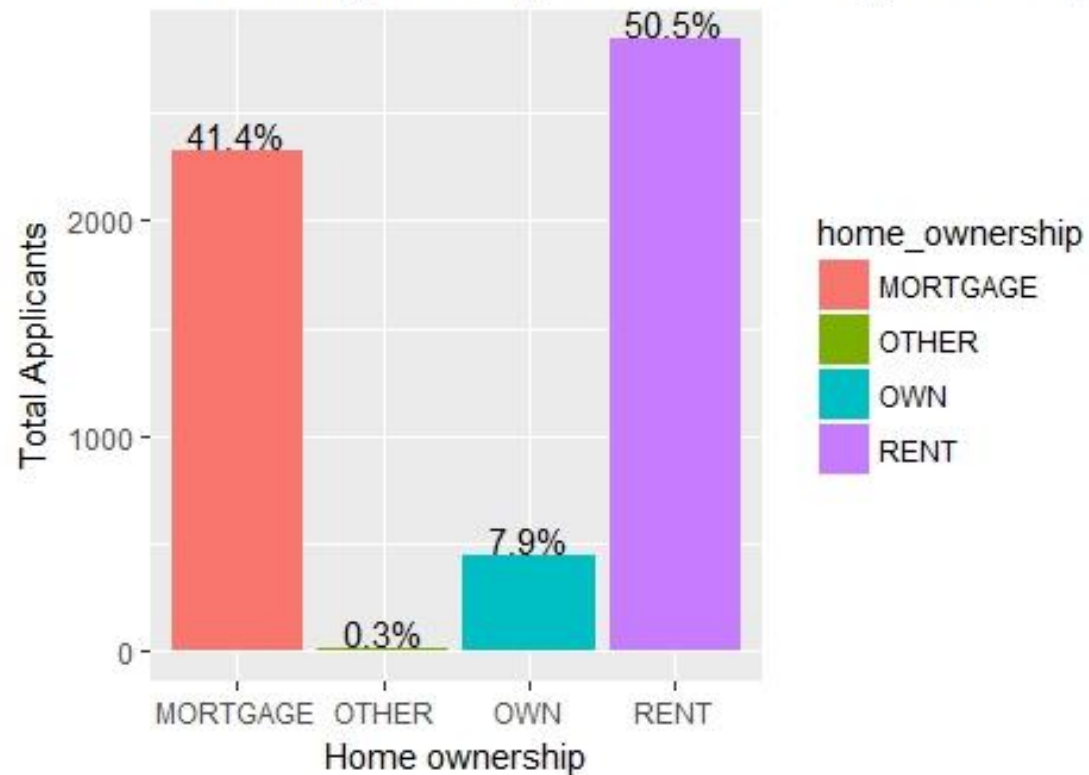


No of applicants Vs Grade

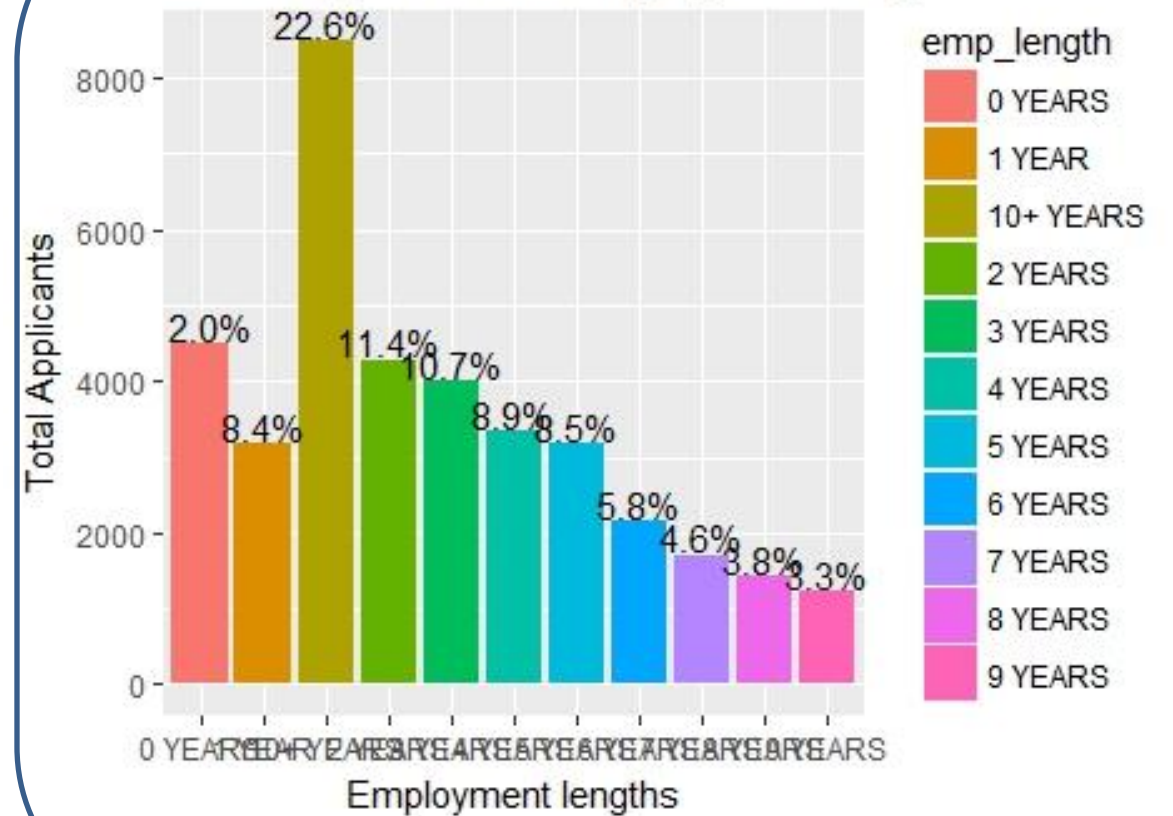


Univariate Analysis

No of charged off applicants Vs Home_ownership

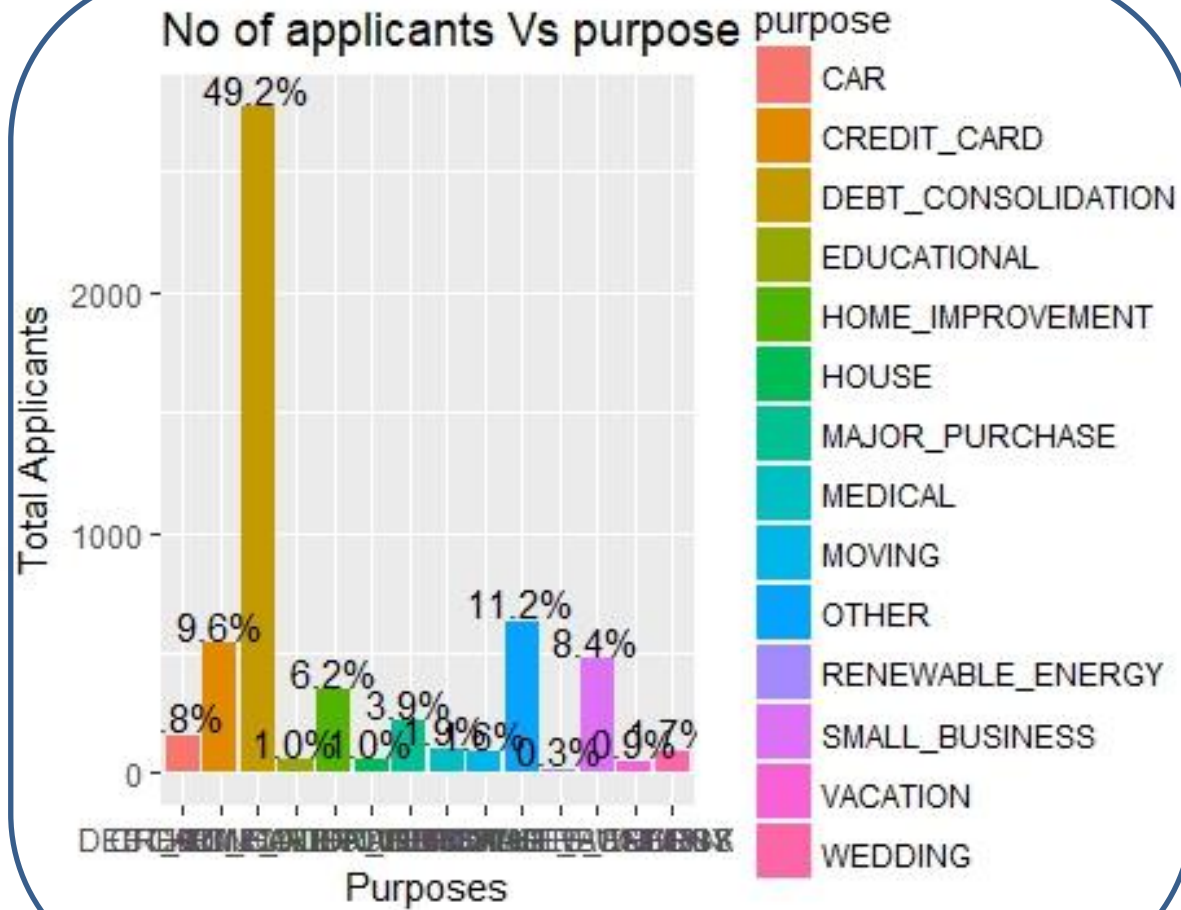


No of applicants Vs Employment length

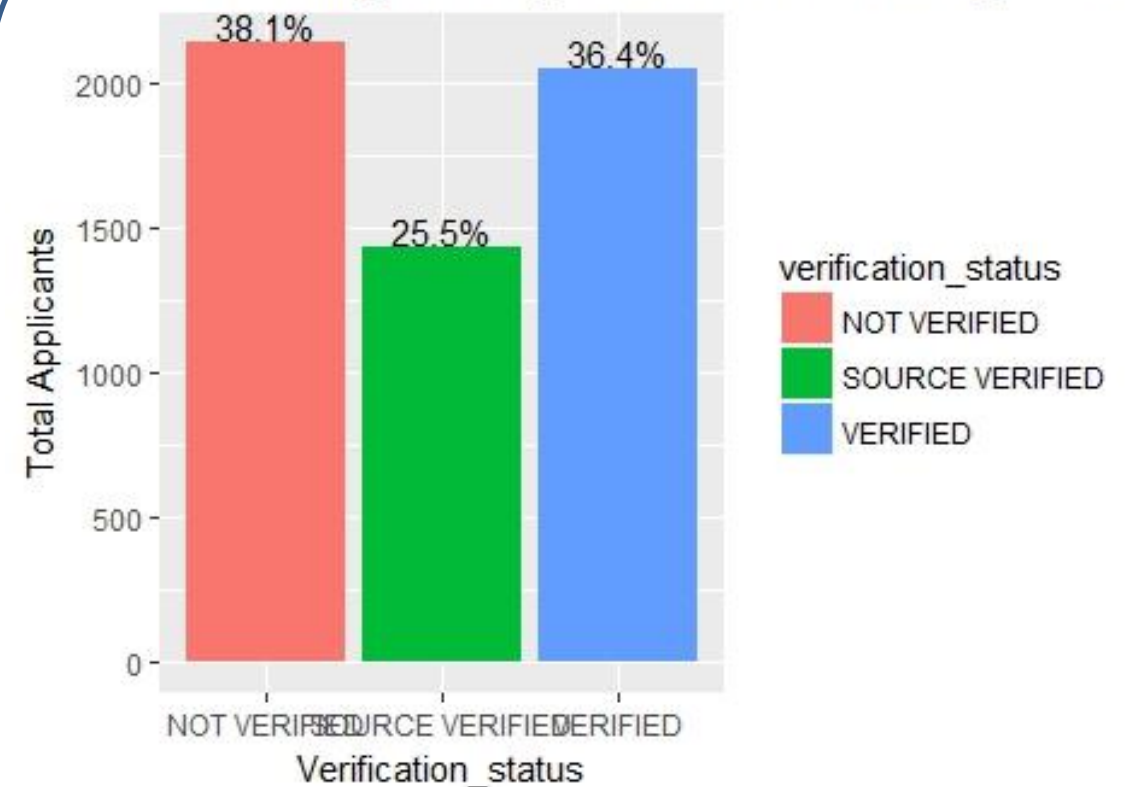


Univariate Analysis

No of applicants Vs purpose



No of charged off applicants Vs verification_status



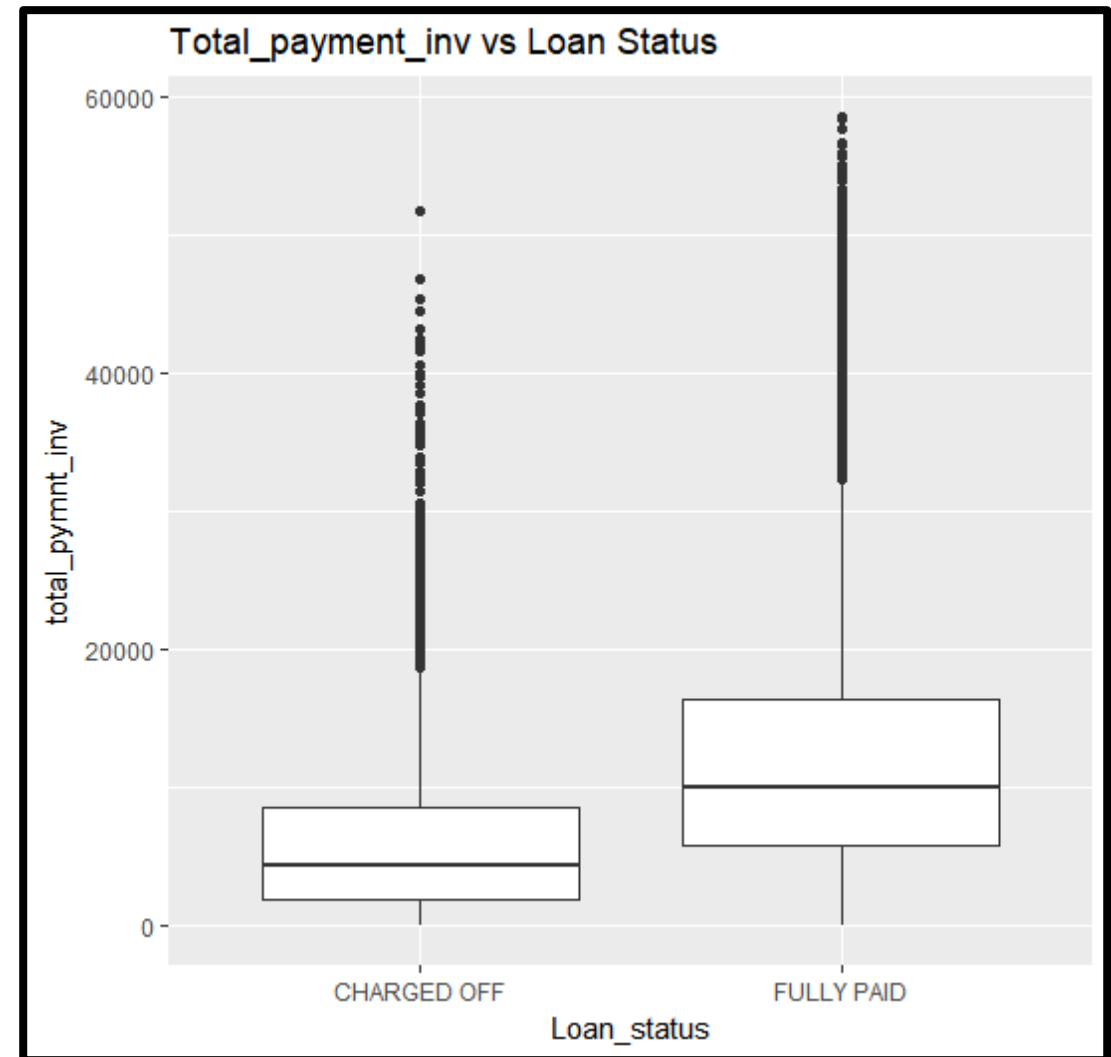
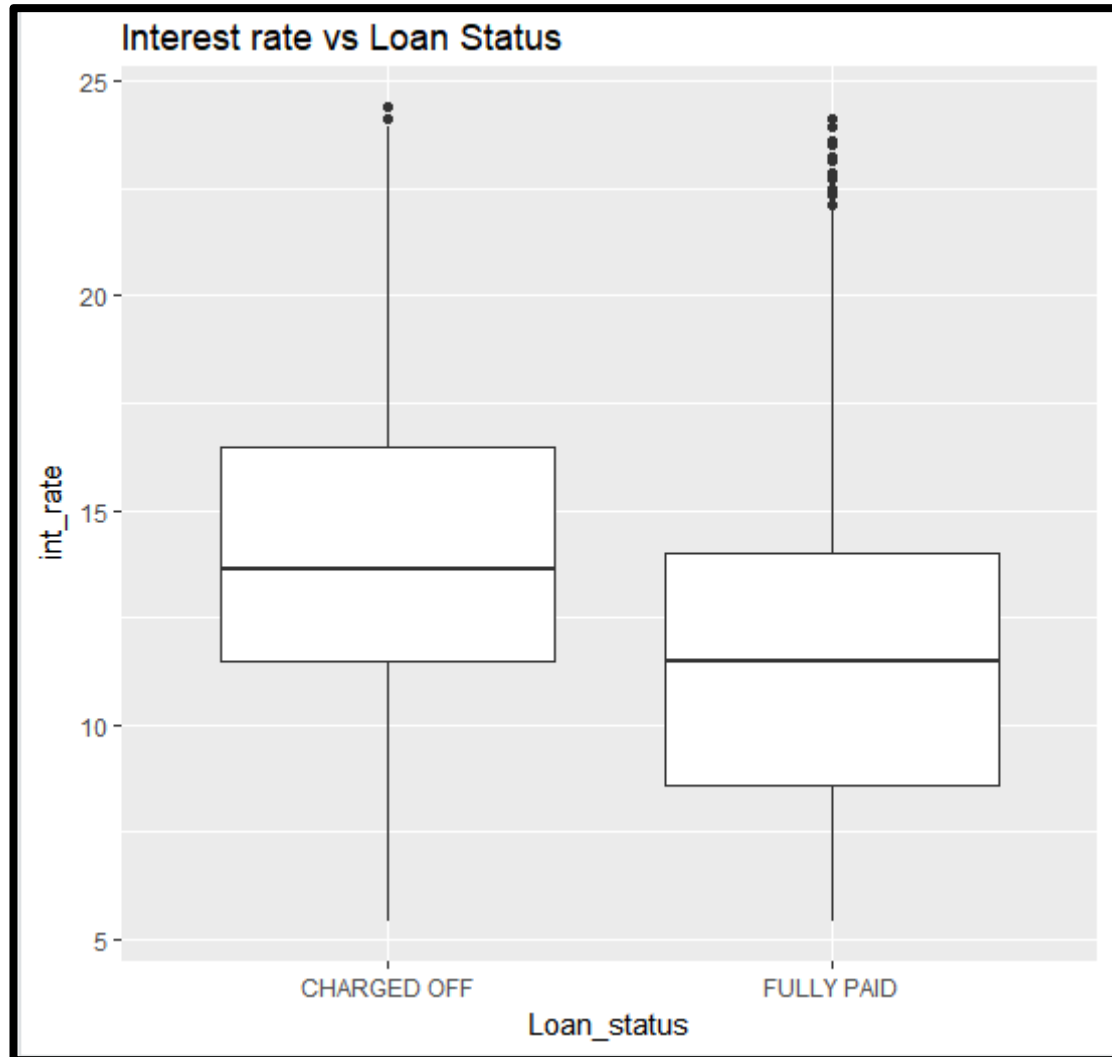
Bivariate analysis

Observations in Bivariate Analysis :

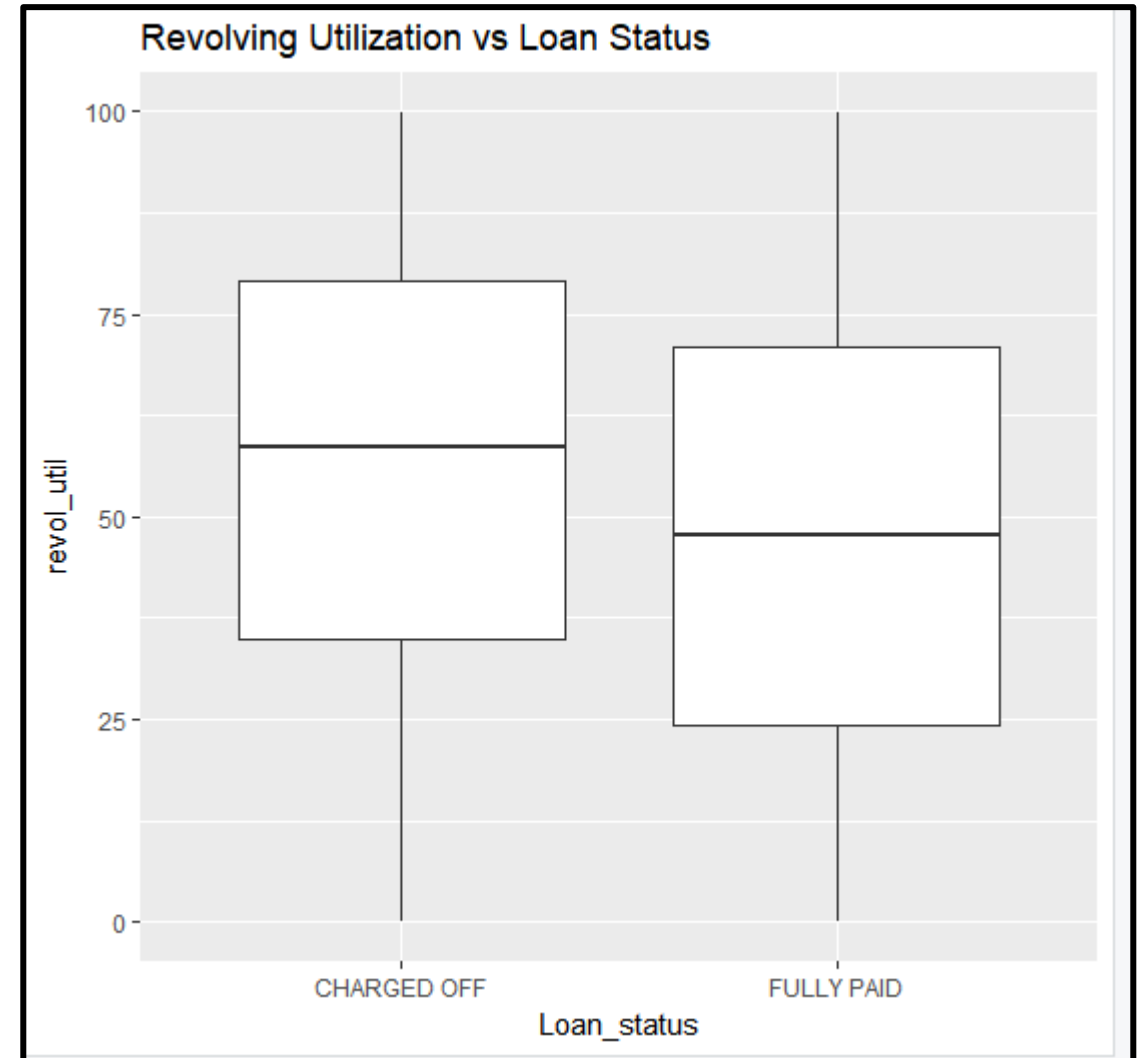
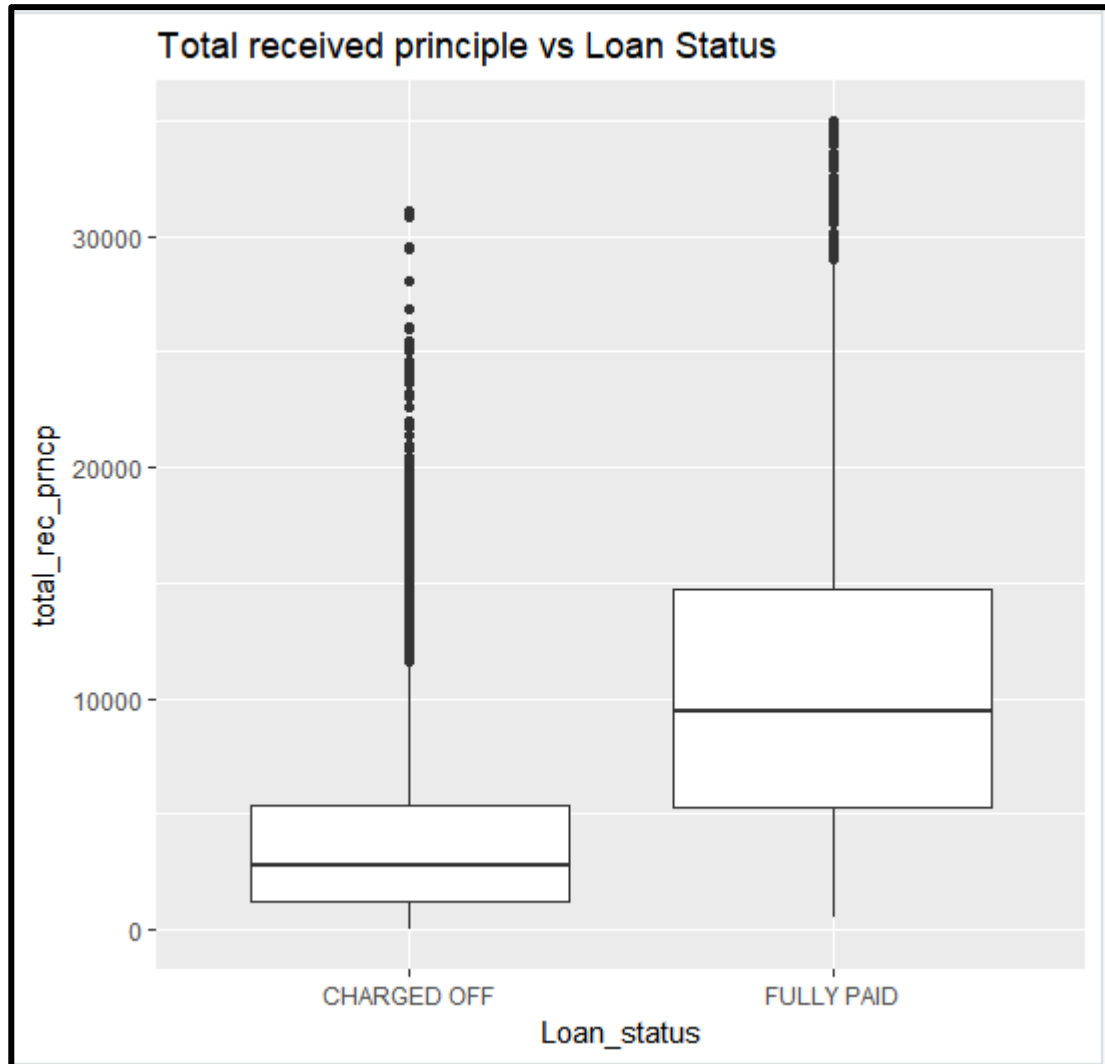
- Based on interest rate and loan status - We can see that `int_rate` might be significant factor in deciding defaulters.
- Based on loan status and `total_payment_inv` - we can see that `total_pymnt_inv` might be significant factor in deciding defaulters
- Based on loan status and `total_rec_prncp` - we can see that `total_rec_prncp` can be a significant factor in deciding the defaulters
- Based on boxplot of `loan_status` and `revol_util` - From boxplot it seems that `revol_util` is significant factor to decide defaulters.

Some graphical analysis are given in following slides for better references.

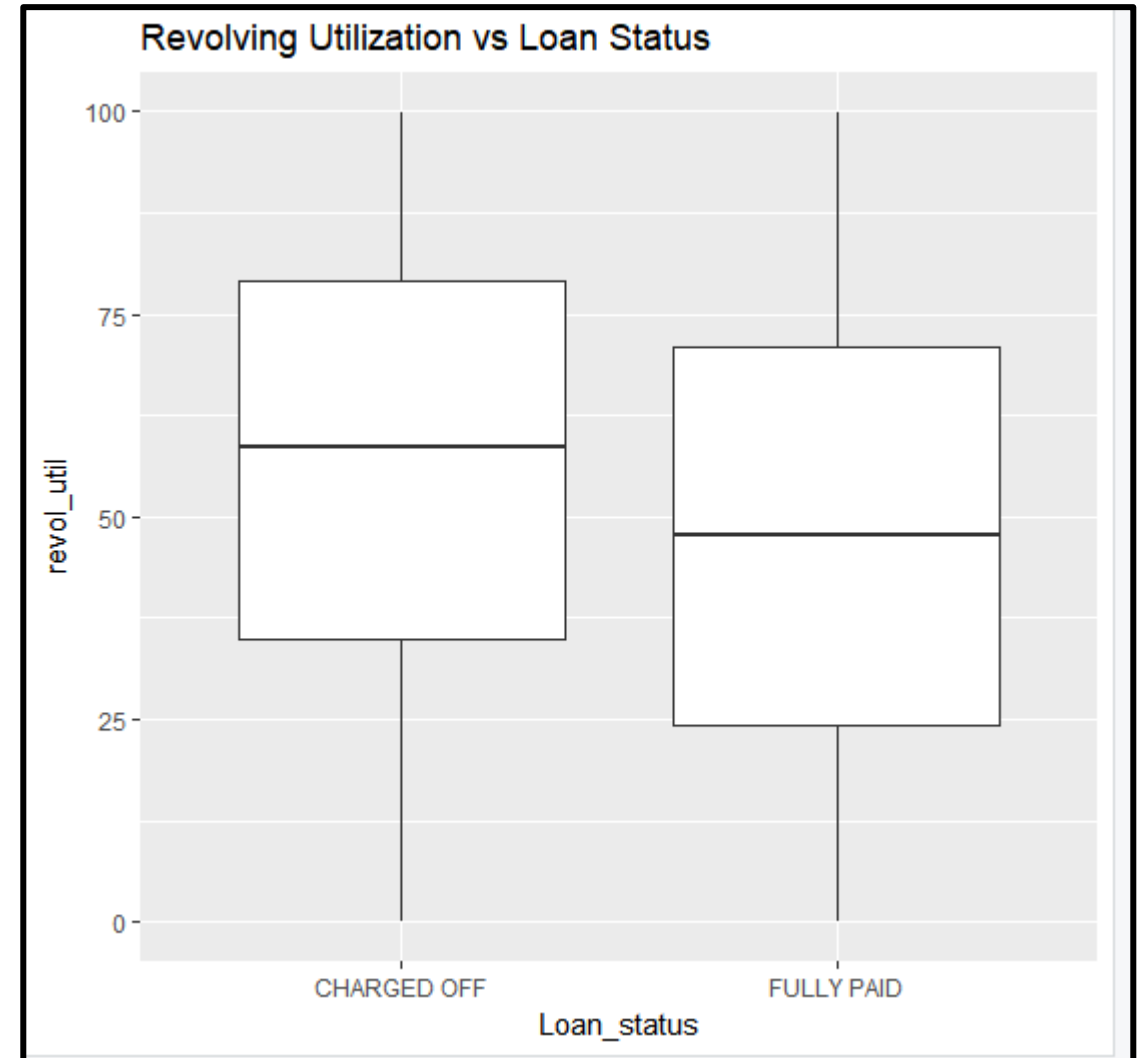
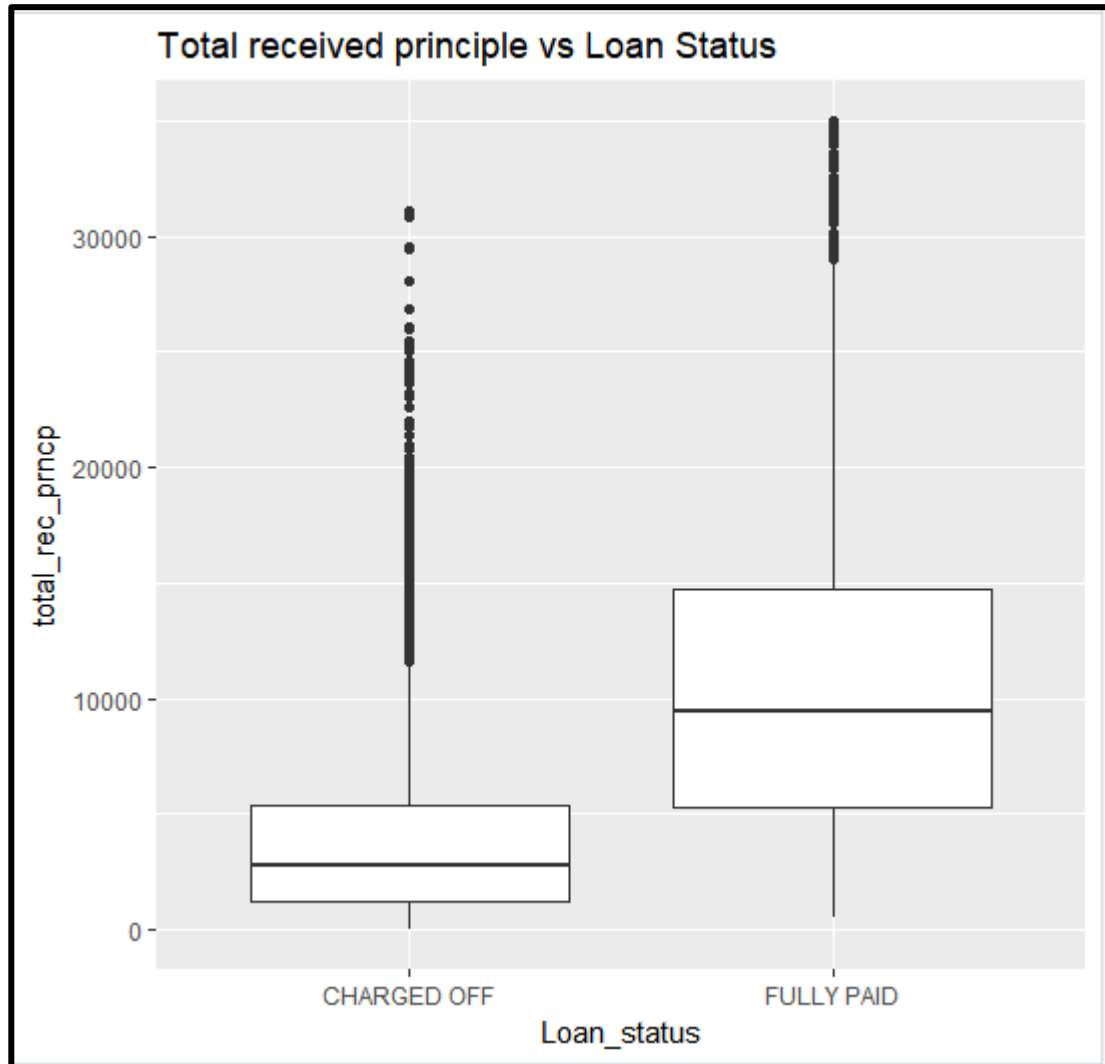
Bivariate analysis



Bivariate analysis



Bivariate analysis

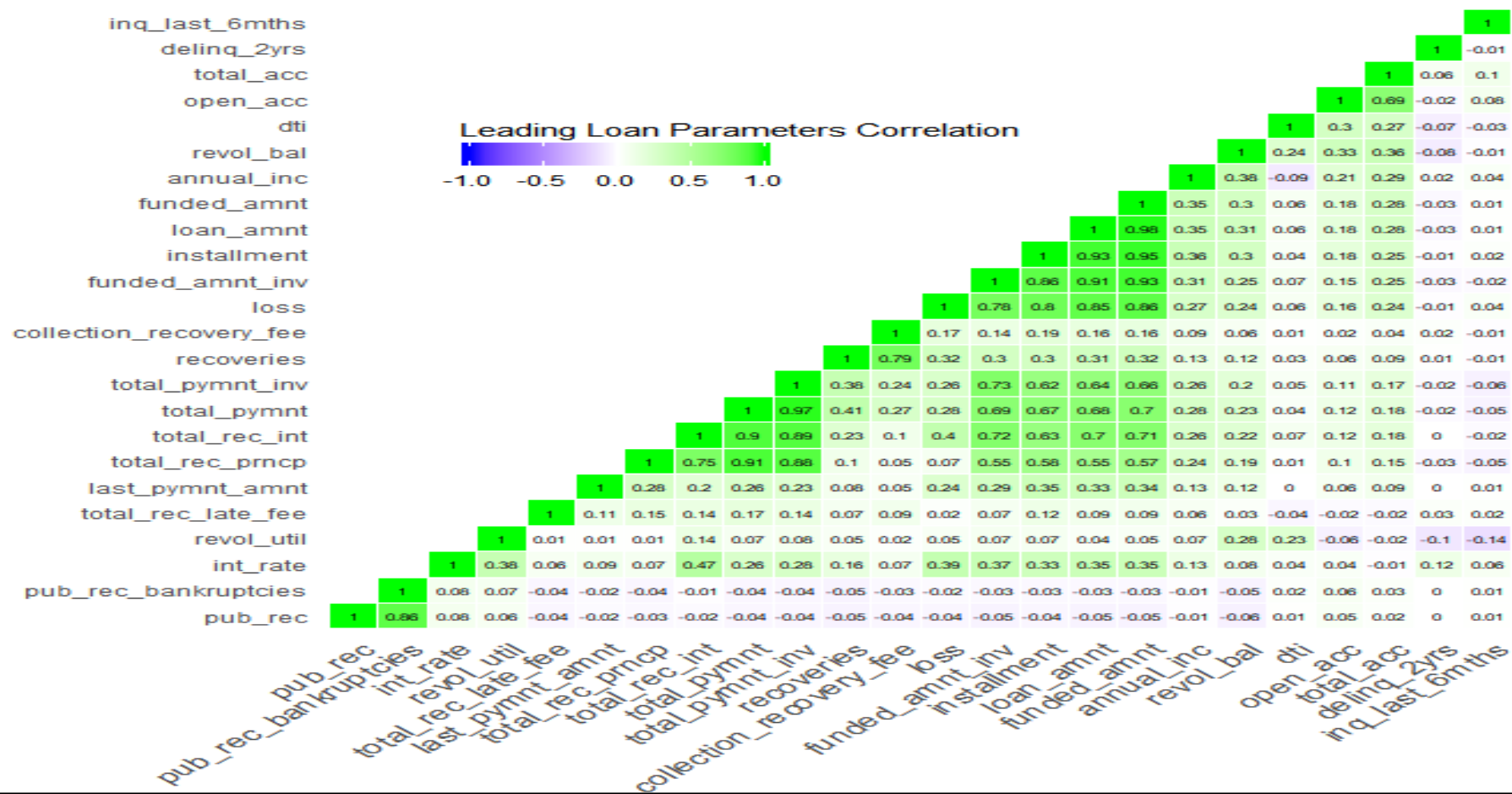


Derived Metrics

As per the requirement we have created 3 derived metrics as below :

- Loan Amount levels (Low, moderate, high, extreme high etc .)
- Interest rate levels (Low, moderate, high, extreme high etc.)
- Loss

Heat map

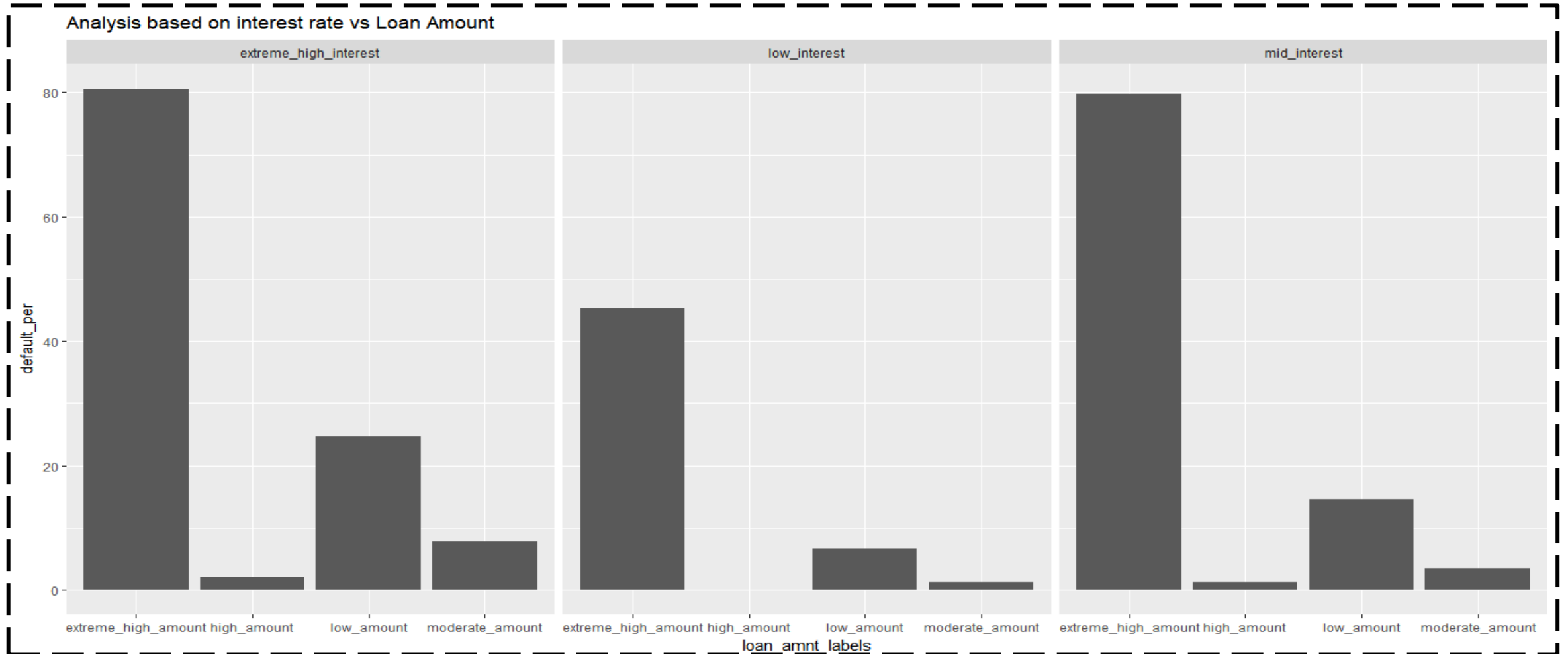


Top 7 Significant Parameters for Loan Defaulter

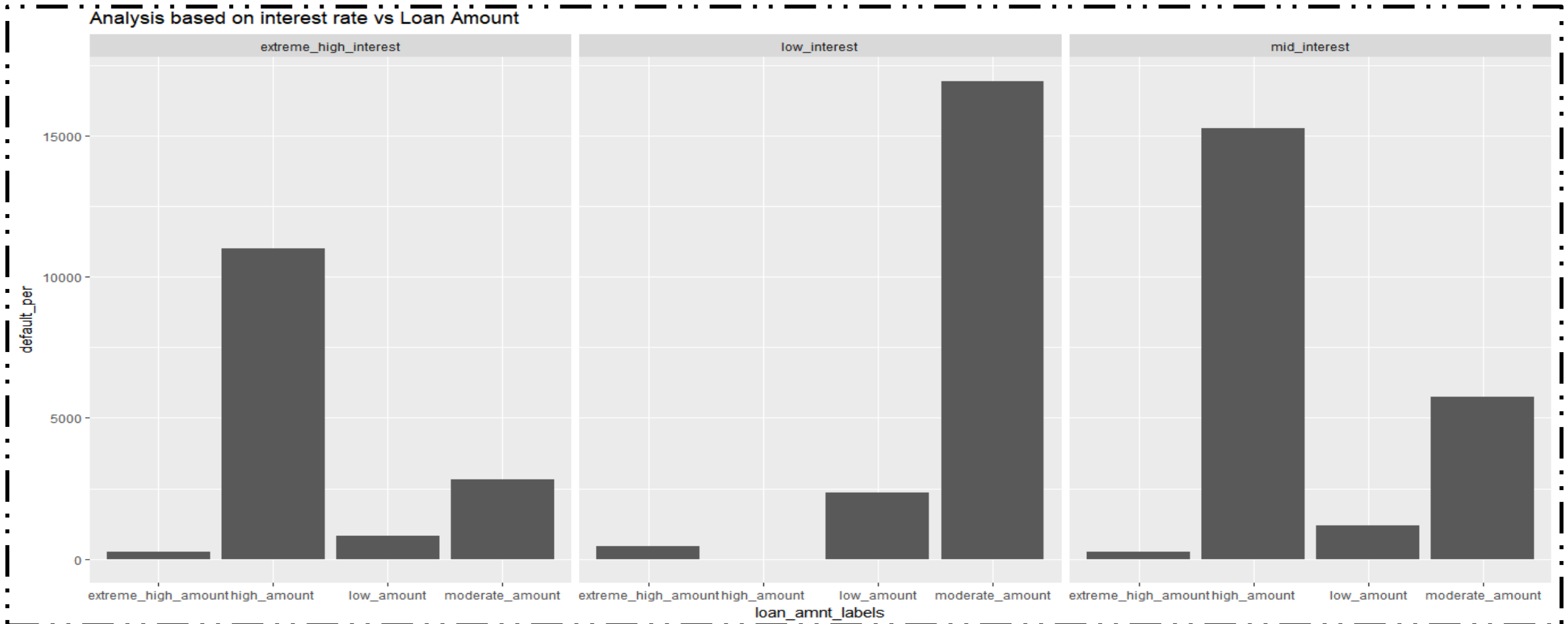
Based on the various analysis done we have come up with the following significant parameter

- Loan Amount
- Interest rate
- Grade
- Purpose of Loan
- Loan terms
- State
- Home Ownership

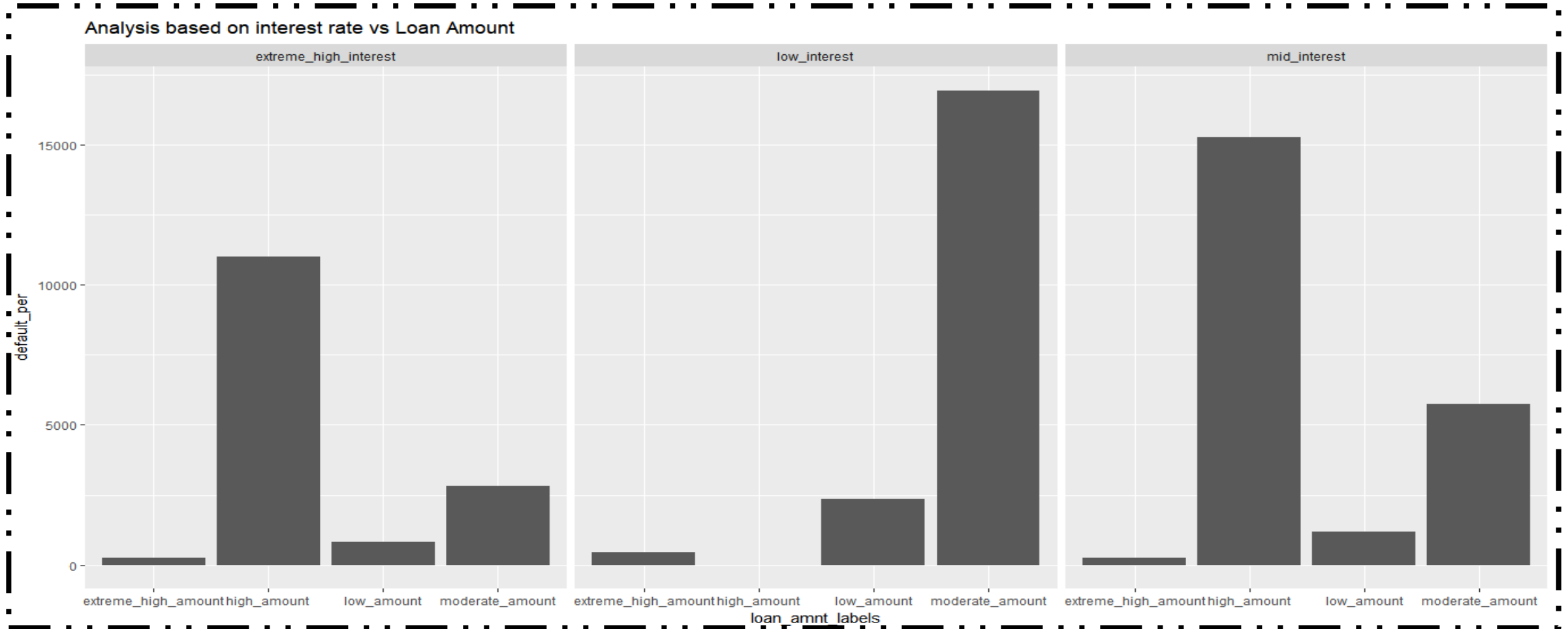
Interest rate Vs Loan Amount (Loan Term)



Interest rate Vs Loan Amount (Loan Purpose)



Interest rate Vs Loan Amount (Home ownership)



Conclusion

Based on various analysis done through Box plot, Bar Chart and HeatMap, we have come up with the following conclusion.

- we found few deciding factors which ends up in determining the loan defaulter applicant.
- Bank can consider the deciding factors before sanctioning the loan to avoid the credit loss.