EDA Case Study
Risk Analysis of Loan Applications

Group

1. Sudhakar Durgam
2. Bhargav
3. Bishwesh Kumar
4. Antra Banerjee

Overview:

Summary:

- When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Once the application is accepted and the loan is sanctioned then a borrower will re-pay the amount in monthly installments completely or can default leading to credit loss.
- Identification of Loan Applicant traits that tend to 'default' paying back
- Understand the 'Driving Factors' or 'Driver Variables' behind Loan Default phenomena

Objective:

- The largest source of financial loss to the company is through credit loss resulting from lending loans to '**risky applicants**'. If a borrower fails to repay the loan then the loan is termed as '**charged off**' and the pending amount is the credit loss to the company.

Aim:

- The aim is to identify patterns and driving variables which indicate if a loan applicant is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending at a higher interest rate, etc.

Risk Analysis Process:

- Implementing EDA on the input dataset to identify risk analytics in a business environment.
- Perform Uni-variate and BI-variate analysis of variables to identify inconsistencies in the source dataset.
- Determine driving factors which are strong indicators of 'risky applicants'.
- Presentation through visualizations.

Methodology:

Business Knowledge:

- Understand the business context of the problem. Going through online material on risk analytics and loan application approval process

Data Understanding and Preparation:

- Understanding variables in the Source file and refer them through data dictionary. Data cleaning, Treating NA values, Addressing outliers and Identifying relevant columns for analysis

Data Analysis:

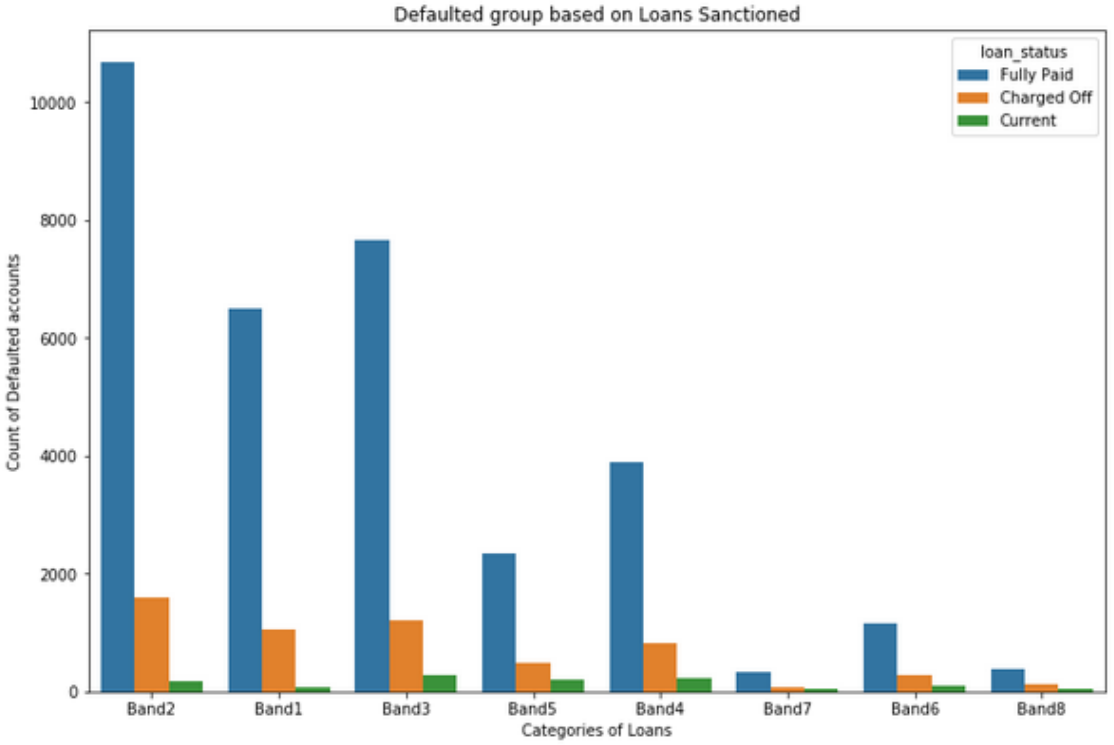
- Find out the trends with categorical and Quantitative variables through EDA and identifying variable relations

Recommendations:

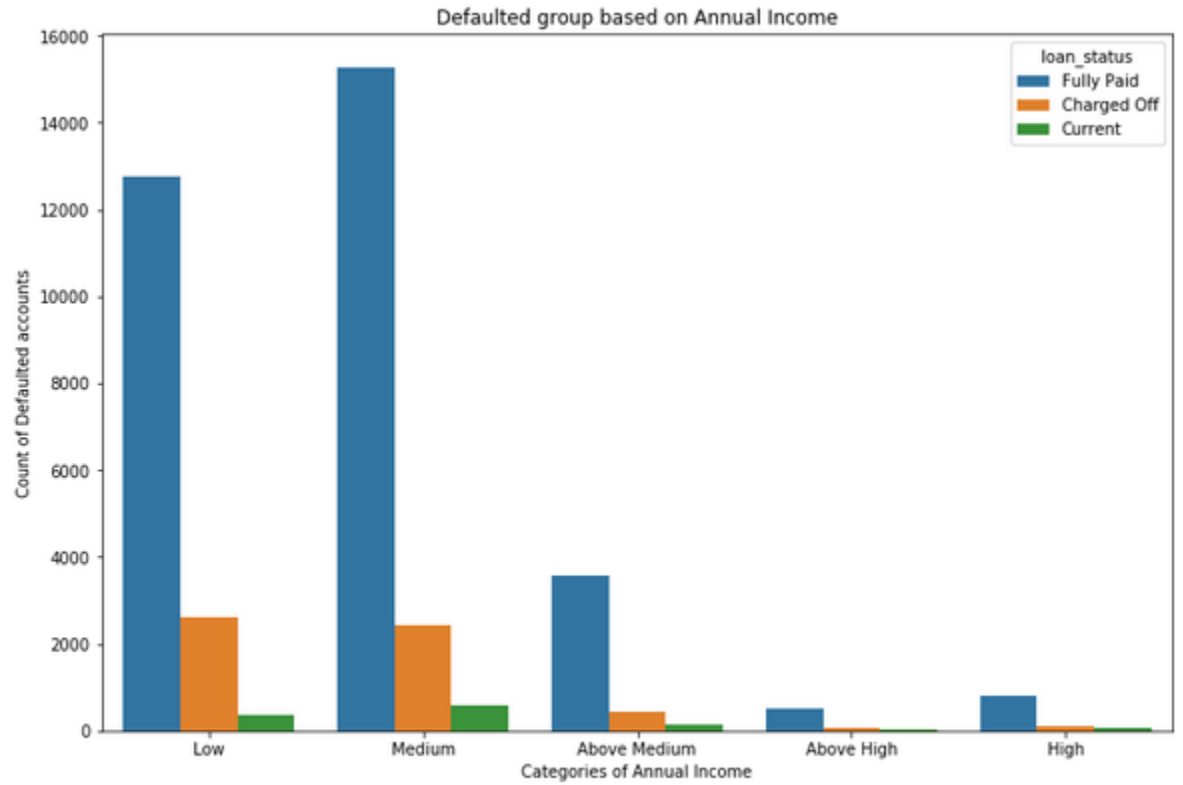
- Present Visualized insights from analysis and provide recommended factors for loan defaulting

Conclusions:

- Defaulted group on loan sanctioned and annual income

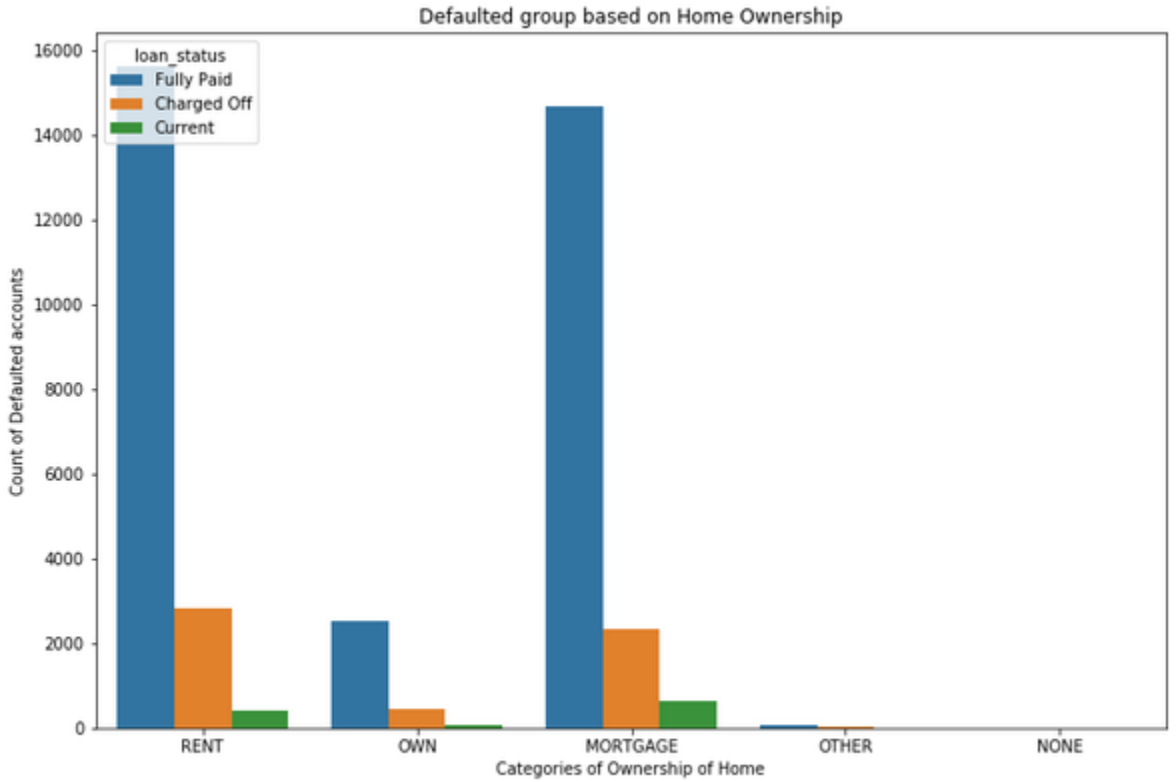


From the above, though there is less visibility about the percentages of 'Charged off' accounts.
From the calculation we could see Band 8 has high number of 'charged off'
accounts which is close to 33%.
Also from bank perspective, we need to target high loan accounts to reduce the maximum loss

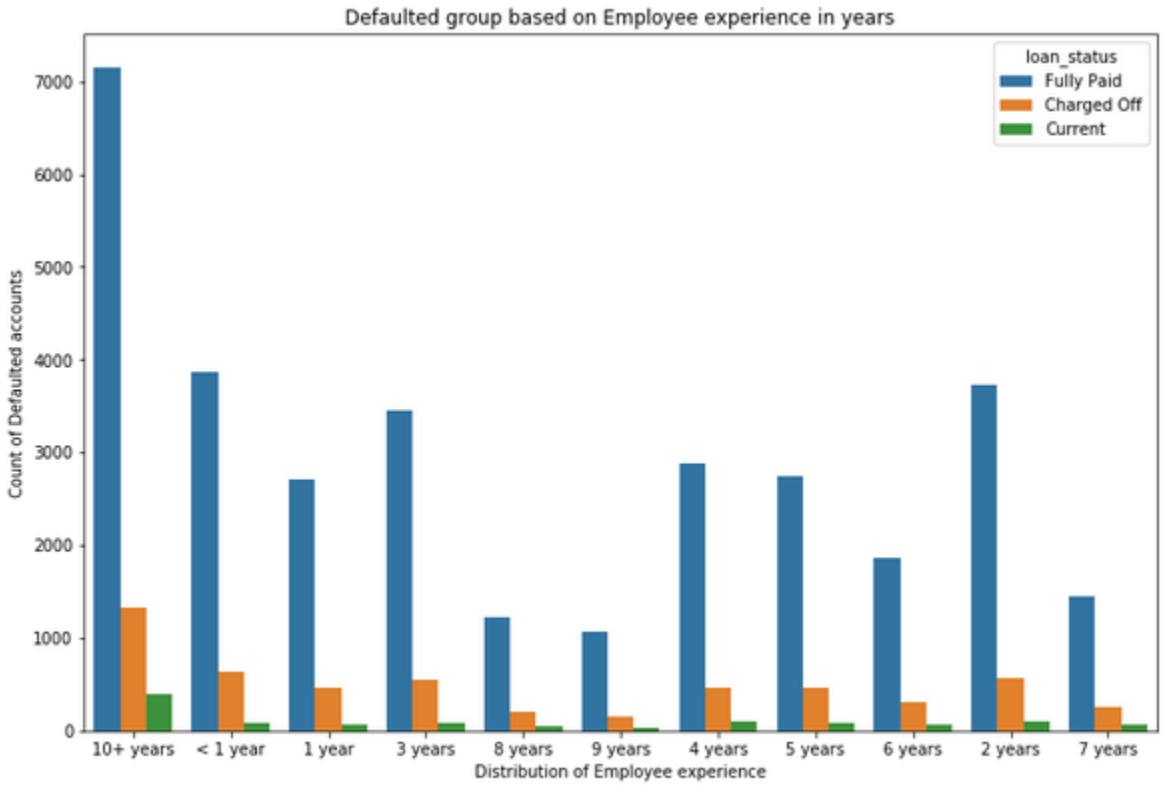


From the above we can calculate, Low income category have high percentage
of default loan accounts around 16.6%. Other next is Medium close to 13%

- Defaulted group on Home ownership and Years of experience

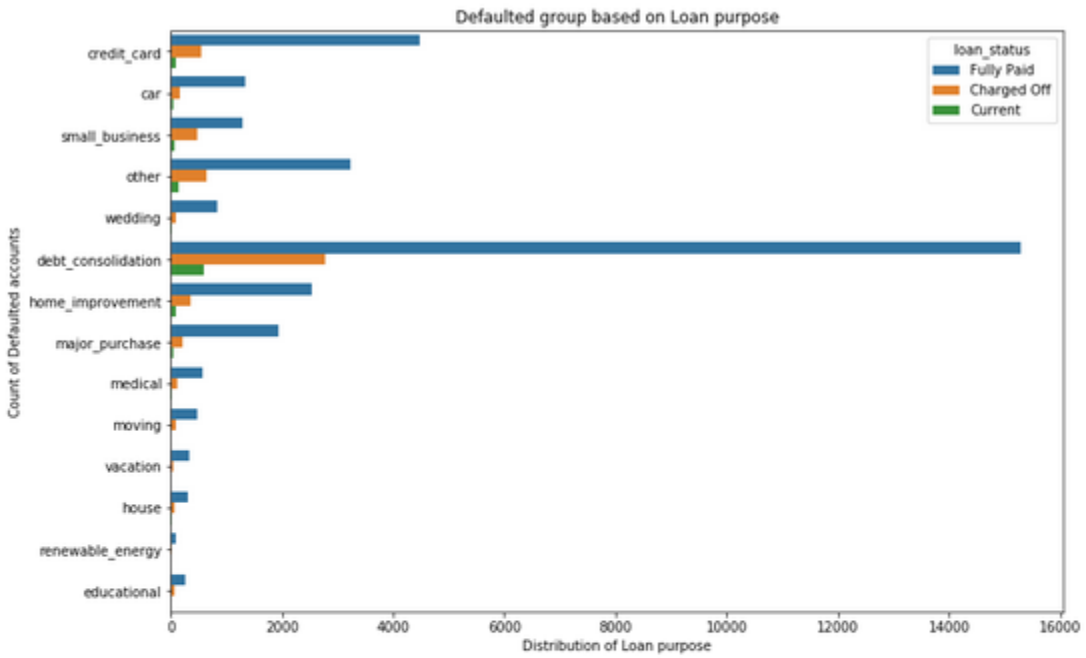


From the above we could see 'RENT' with home ownership has the highest default accounts around 18%



EMp_length having more than 10+ years of experience - Around 15%

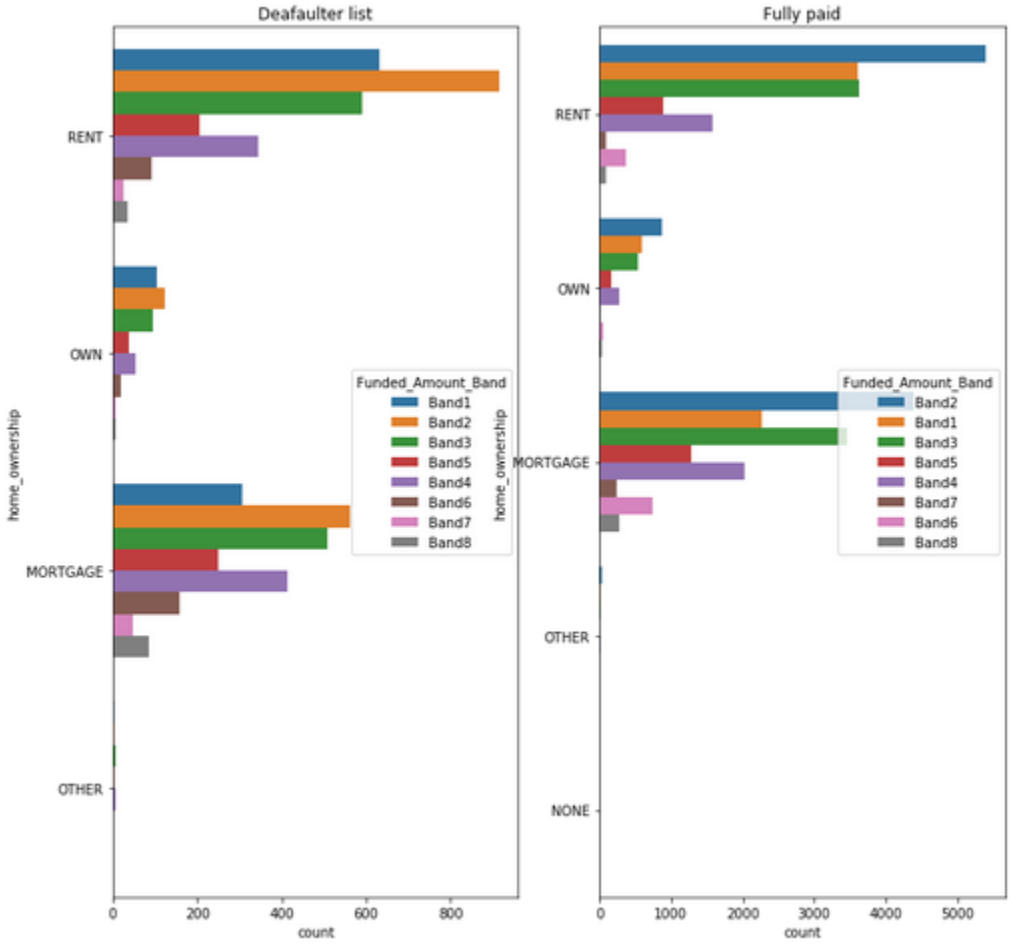
- Defaulted group on Loan purpose and Grades



Purpose with 'Debt consolidation' has the highest defaulters

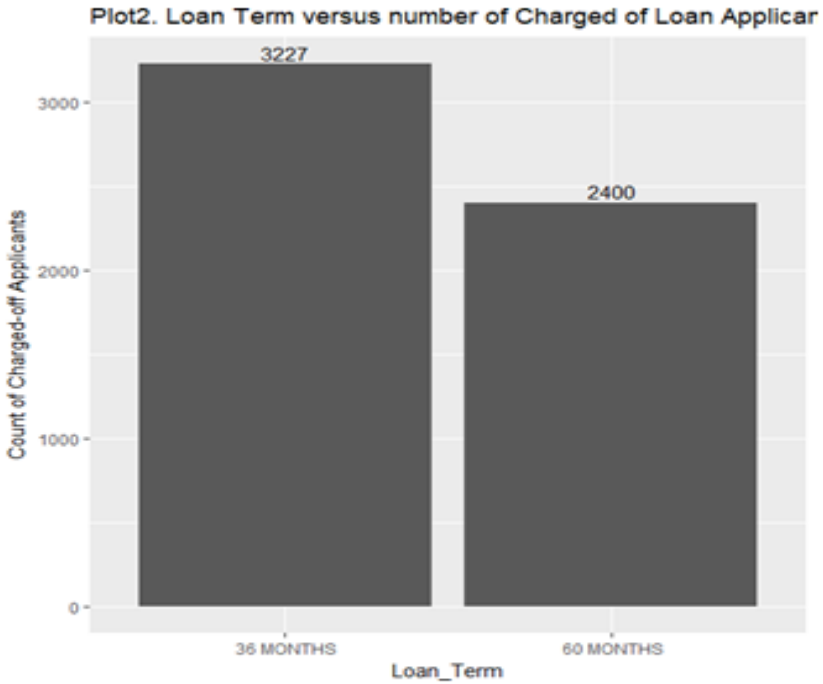
UNIVARIATE ANALYSIS PARAMETERS FOR LOAN DEAFULT

1) Band 8 is the problem (loan amount > 30000). This is also big loss to company as the loan amount is huge -33%
2) Low income category (annual income <50000) have high percentage of default loan accounts around 16.6%
3) Home ownership as 'RENT' is having the highest defaulted account - 18.5%
4) Interest rates more than 17% interest have more defaulters. This increase is seen if the interest increases. - MOre than 25%
5) DTI between 18-25% have more defaulted accounts - 16-17%
6) EMp_length having more than 10+ years of experience - Around 15%
7) Purpose for requesting the loan 'Debt consolidation'



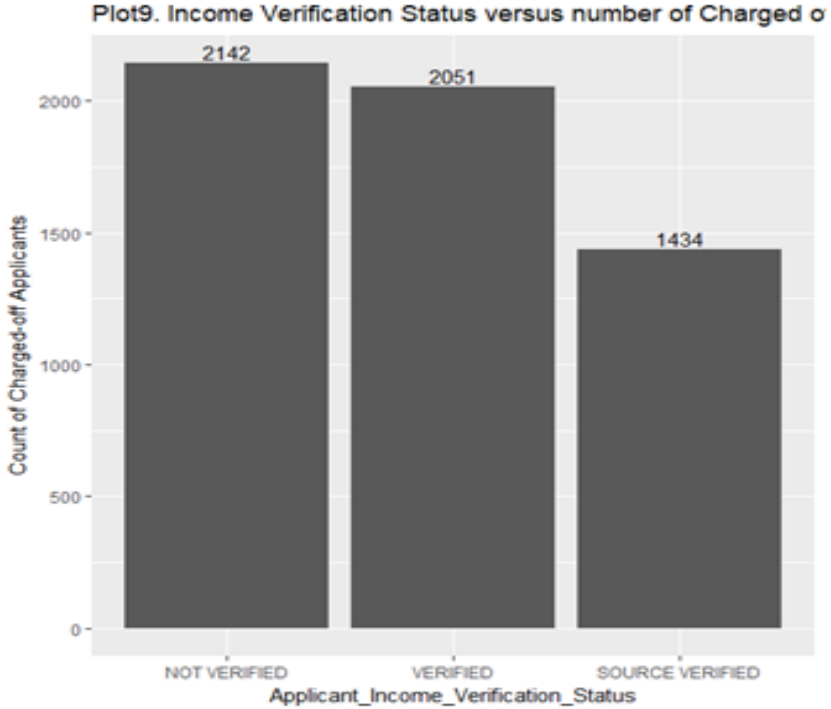
Loans below 10,000 & with the applicants having 'RENT' as home_ownership is another indicator for default account

Analysis of Loan Term versus Count of Charged-Off Loans



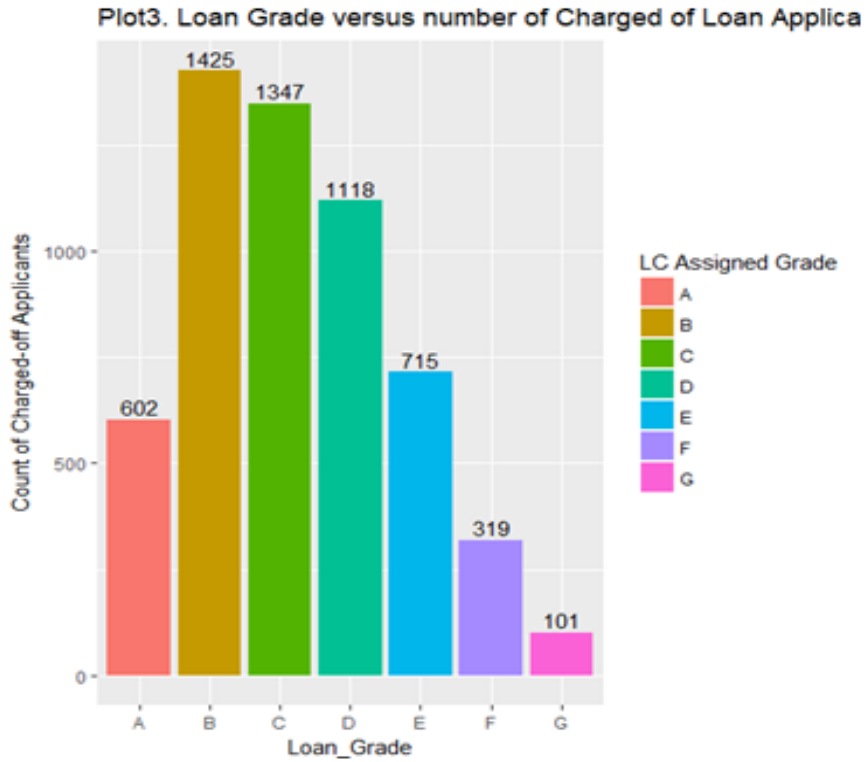
- Observation [A]1. The number of charged-off loans for 3 years is higher than that of 5 years

Analysis of Applicants Income Verification Status vs. Count of Charged-Off Loans.



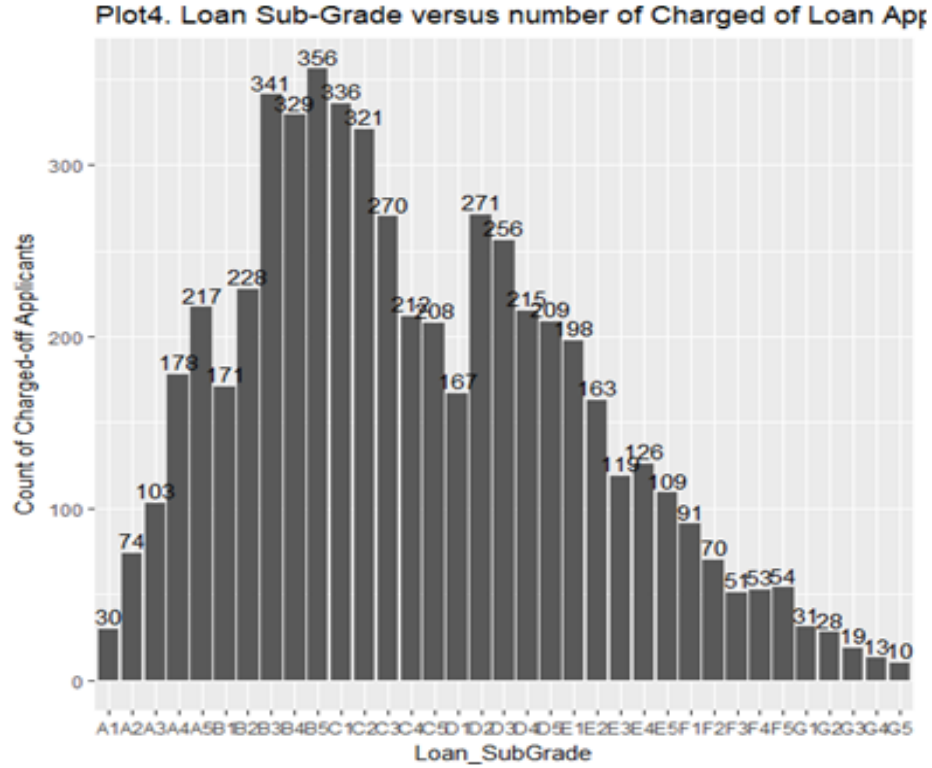
- The number of charged-off loans for applicants with not verified income is higher than that of verified and source verified.

Analysis of LC Assigned grade versus Count of Charged-Off Loans



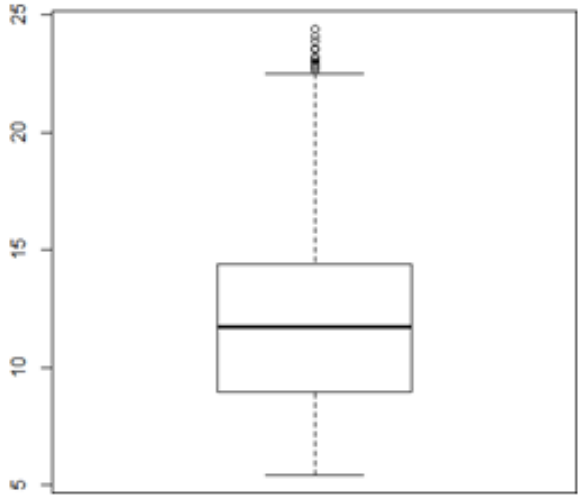
- LC Grades B,C and D contribute to the majority number of charged-off loans for

Analysis of LC Assigned sub-grade versus Count of Charged-Off Loans



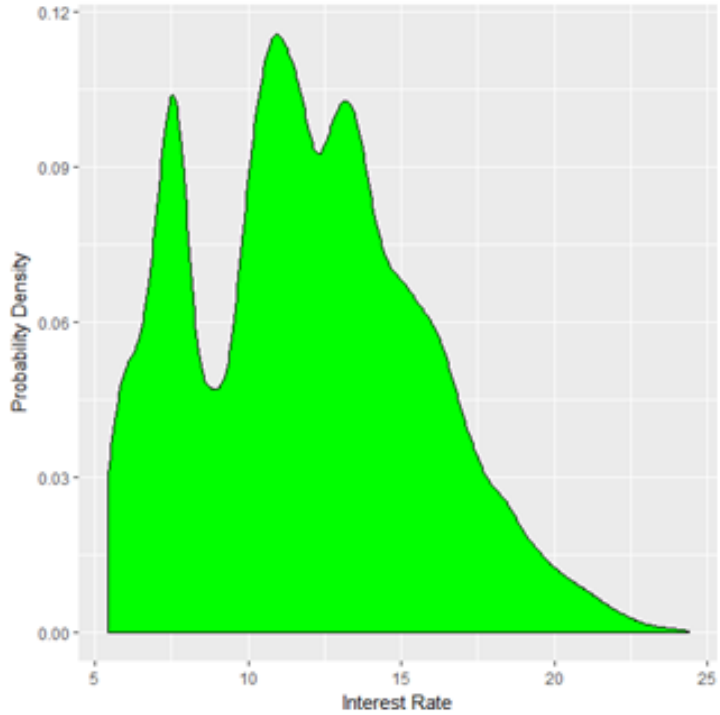
The highest number of charged-off loans are in B3 ~ C3 and also D2~E1 sub grades.

Interest Rate boxplot and summary and Analysis

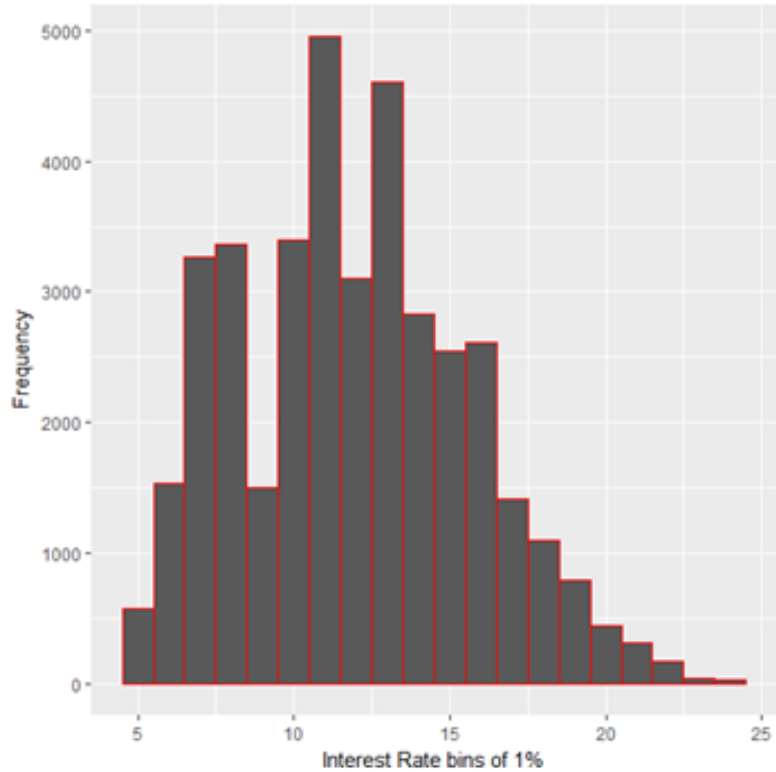



```
> summary(loan_dataset$int_rate)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   5.42    8.94   11.71   11.93   14.38   24.40
```

Plot15. Density plot of Interest Rate



Plot16. Frequency Plot of Binned Interest Rate



- From the box plot & the distribution, we can see that the count of loans with interest rate spikes between 7-8% and again between 11-13%.