GRAMENER CASE STUDY

(EDA Case Study)

Members:

1. Ashwini G.M
2. Meenu Jomi
3. Midhun R

• Abstract & Business Objectives

- Lending Club is peer-to-peer lending company, it’s a largest marketplace connecting barrowers with lenders.
- We are trying to analyse the companies are the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.
- Some of their popular products are debt consolidation loans, credit card, house loans etc..
- Lending loans to a ‘risky’ applicants seems to be a largest source of financial loss. Our aim is to identify the risky loan applicants so that those can be reduced.
- Identification of such applicants using EDA is our aim as part this case study.
- Identify key factors on credit loss for Lending Company
- Perform EDA on Lending Company dataset to determine reasons & trends for loan defaults
- Differentiate risky applicants from others
- Provide recommendations to reduce risk & improve profitability

Problem solving methodology (EDA)

- Data Sourcing
- Data Cleaning
- Univariate Analysis
- Segmented Analysis
- Bivariate Analysis
- Derived Metrics

• EDA Approach

- Business problem analysis into following categories
- Overall Loan Status
- Borrower demography
- Loan product profile
- Data cleansing, remove redundancy and pre-process for data consistency
- Perform univariate & bi-variate analysis
- Identify reasons & trends for default through correlation between various variables
- Final recommendations based on EDA take-aways

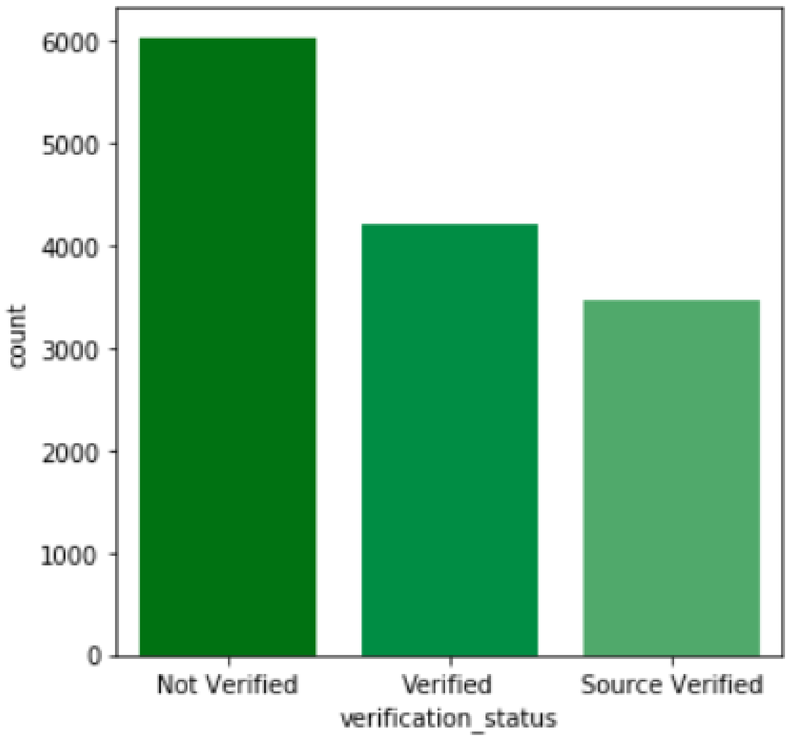
• Data Sourcing / Data Cleaning

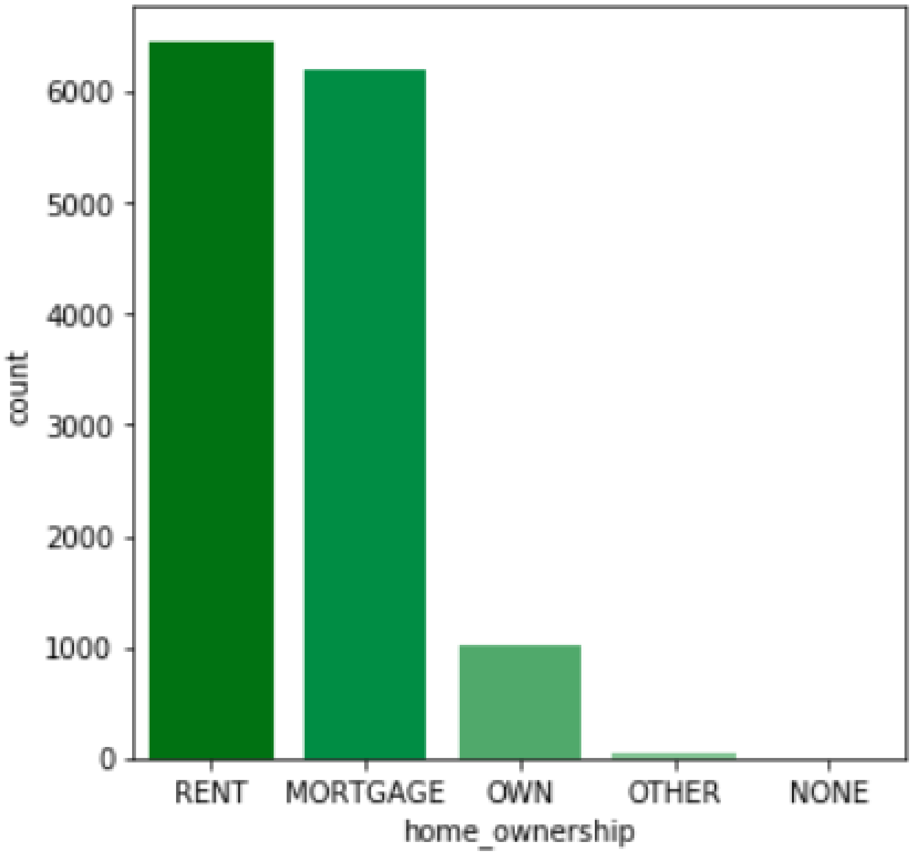
- Data Sourcing:

  The data was provided by Gramener and we have utilized it to our full potential to do the analysis

- Data cleaning:
  - Firstly there were 111 columns. We extracted all the required
  - Then all the missing values in the Rows were dropped
  - We converted interest rate from % to decimal value along with changing it's type

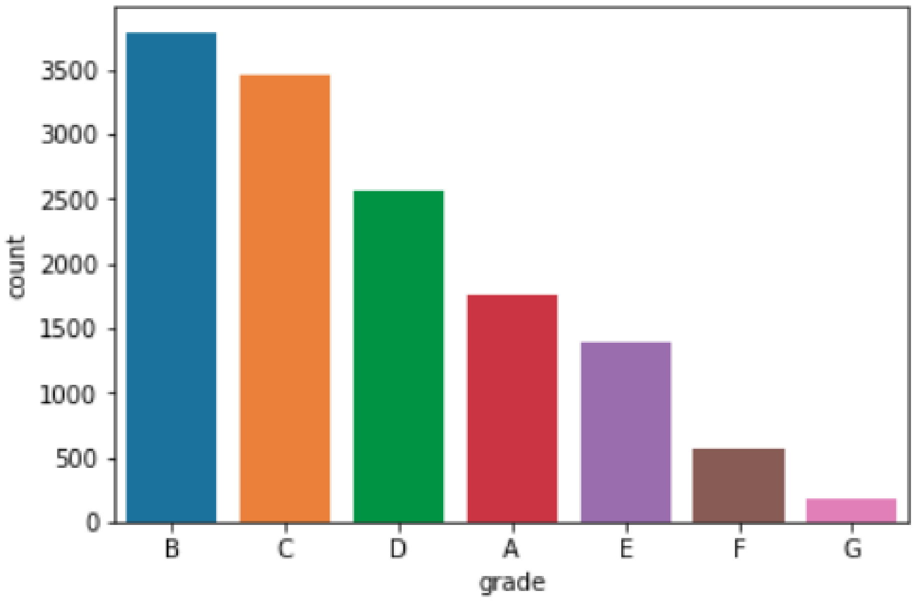
• Plots through EDA

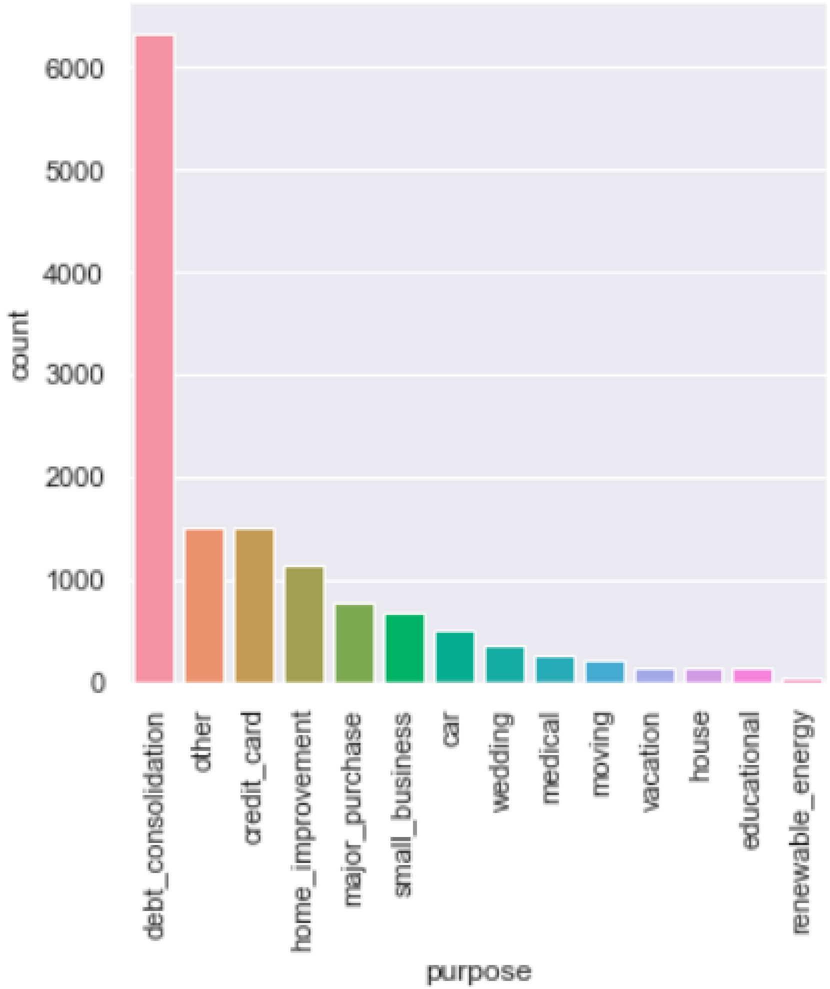




- *Looks like most of the people are on Rent or mortgage*
- *Most of them are Not Verified*

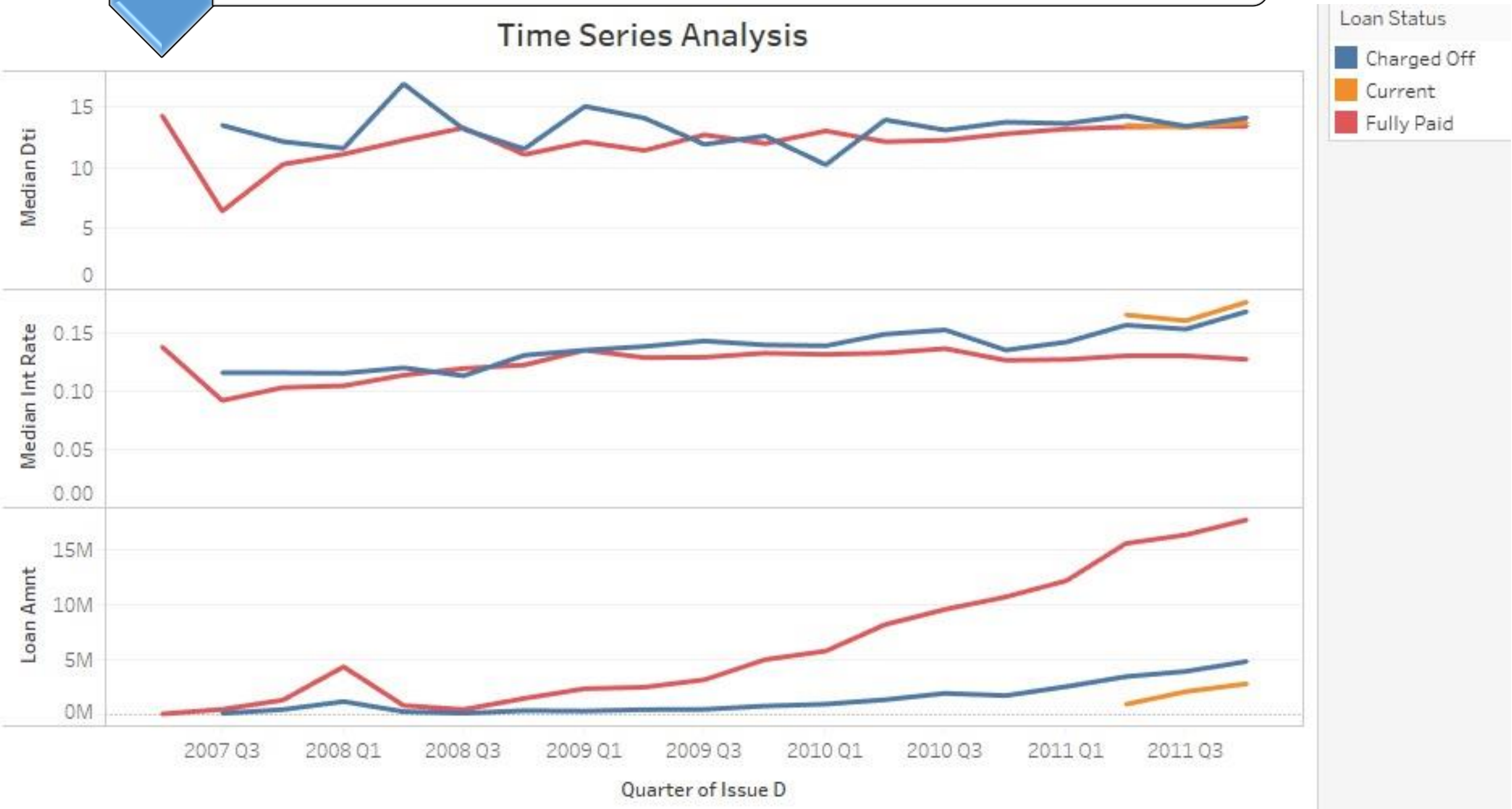
• Plots through EDA





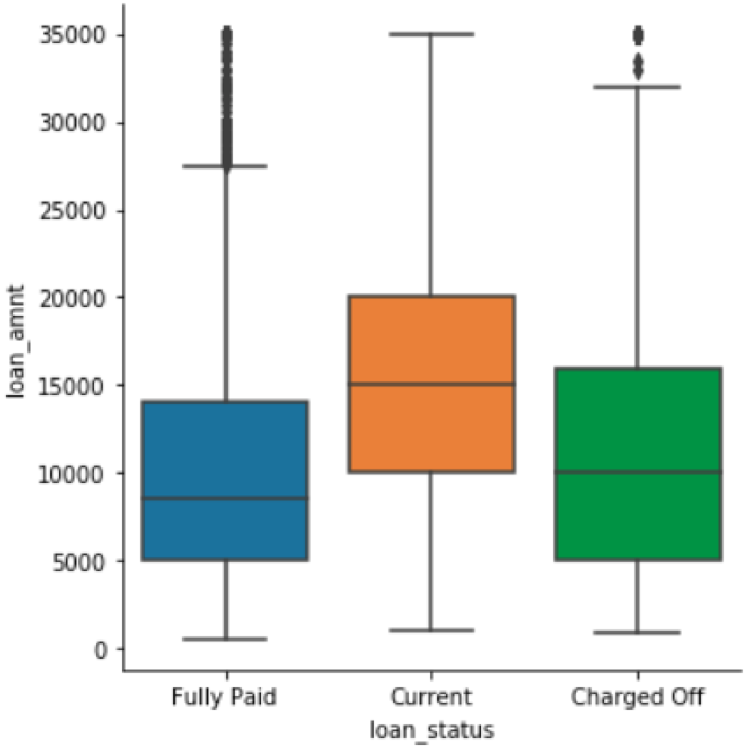
- *B and C grades have more count as compared to G and F*
- *Debt consolidation is the highest in purpose*

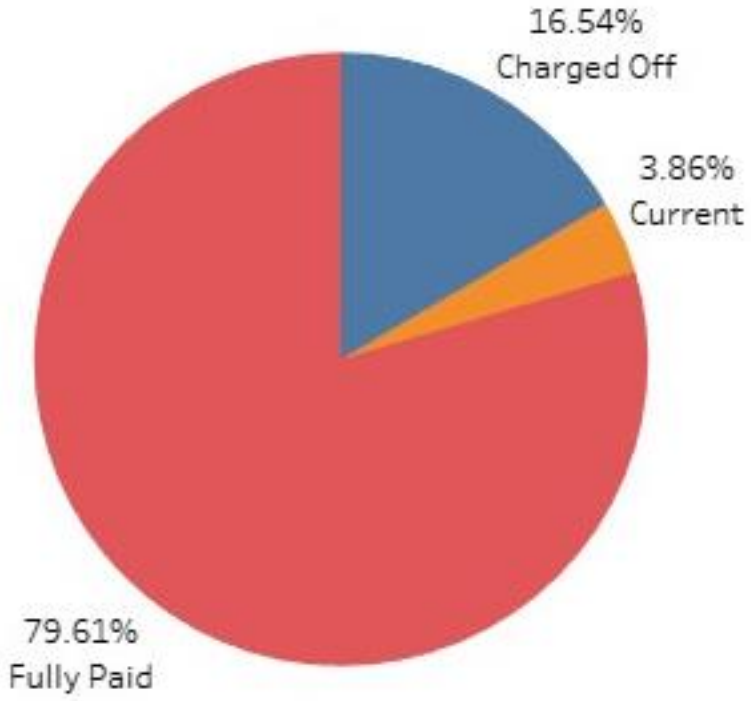
• Plots through EDA



• *The plot shows different types of loan Status against DTI, Interest Rate and Loan Amount*

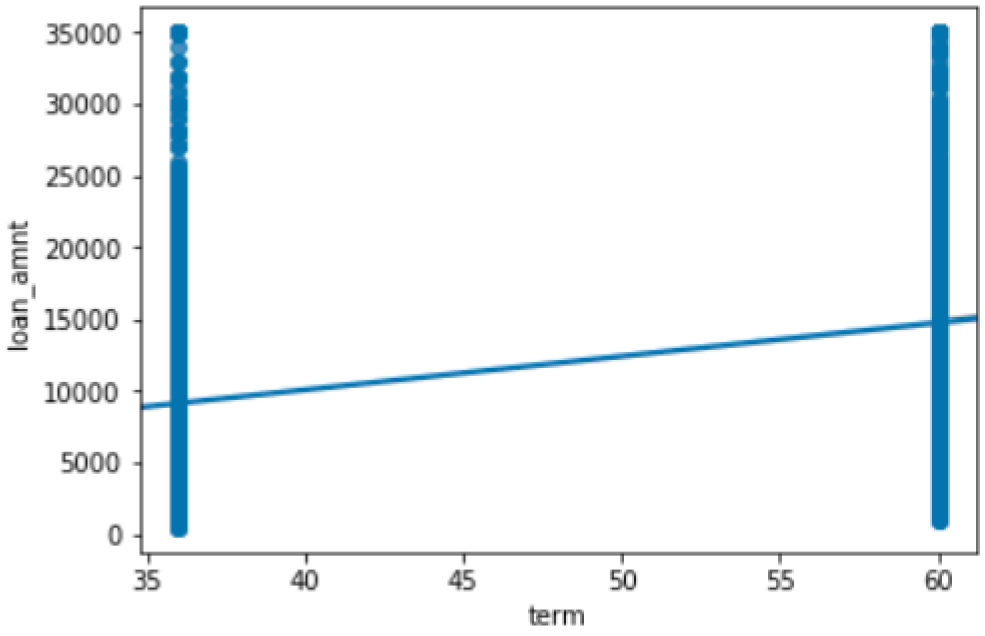
• Plots through EDA

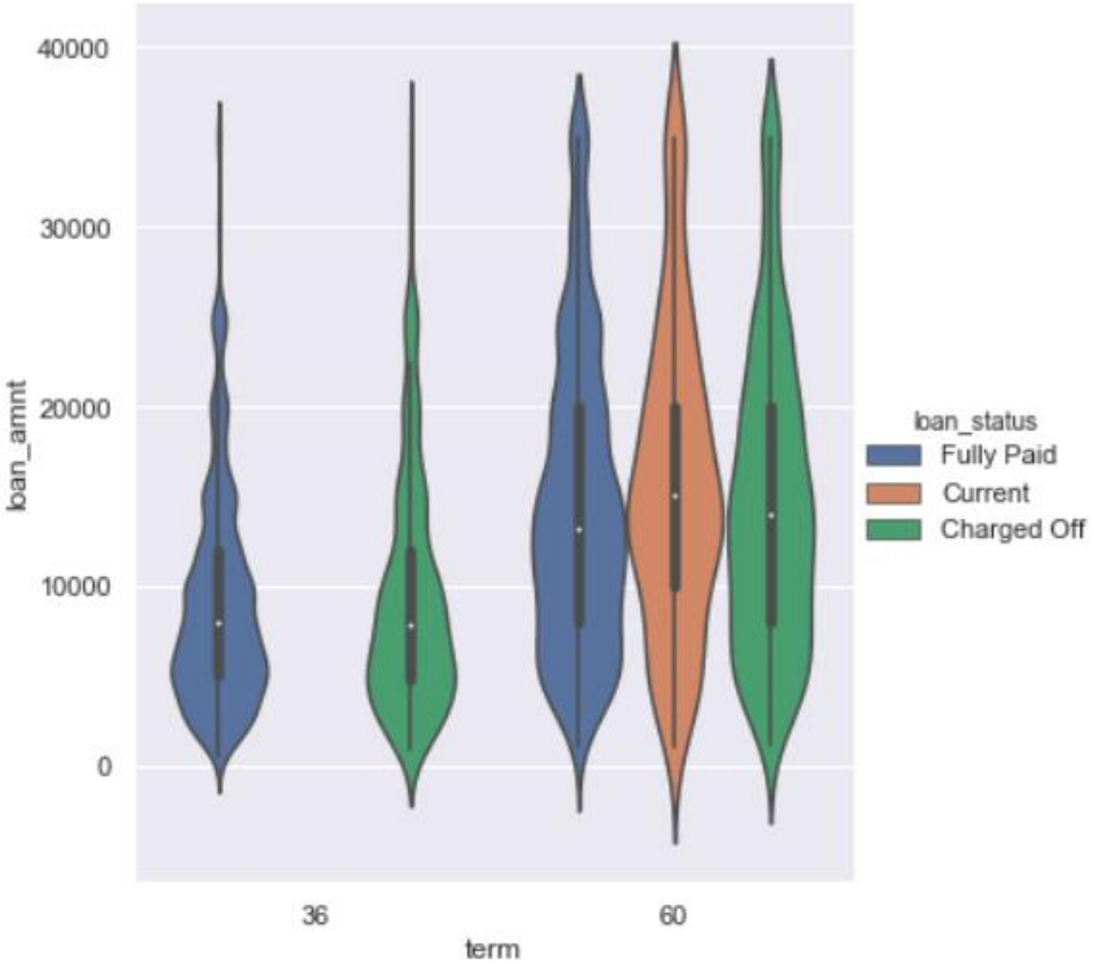




- *Loan amount ranges from 500 to 35000 with mean 10567 and median 9000*
- *Around 80% have fully paid*
- *From the above box plot we can understand that the charged off loan status have higher loan amount compared to the Fully paid loans*
- *The applicants with charged of status can be considered as a defaulter and applicant with a status Fully paid can be considered as a no defaulter.*
- *Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'. Means we have to identify the significant features and the difference in the characteristics of Fully Paid and Charged off users.*

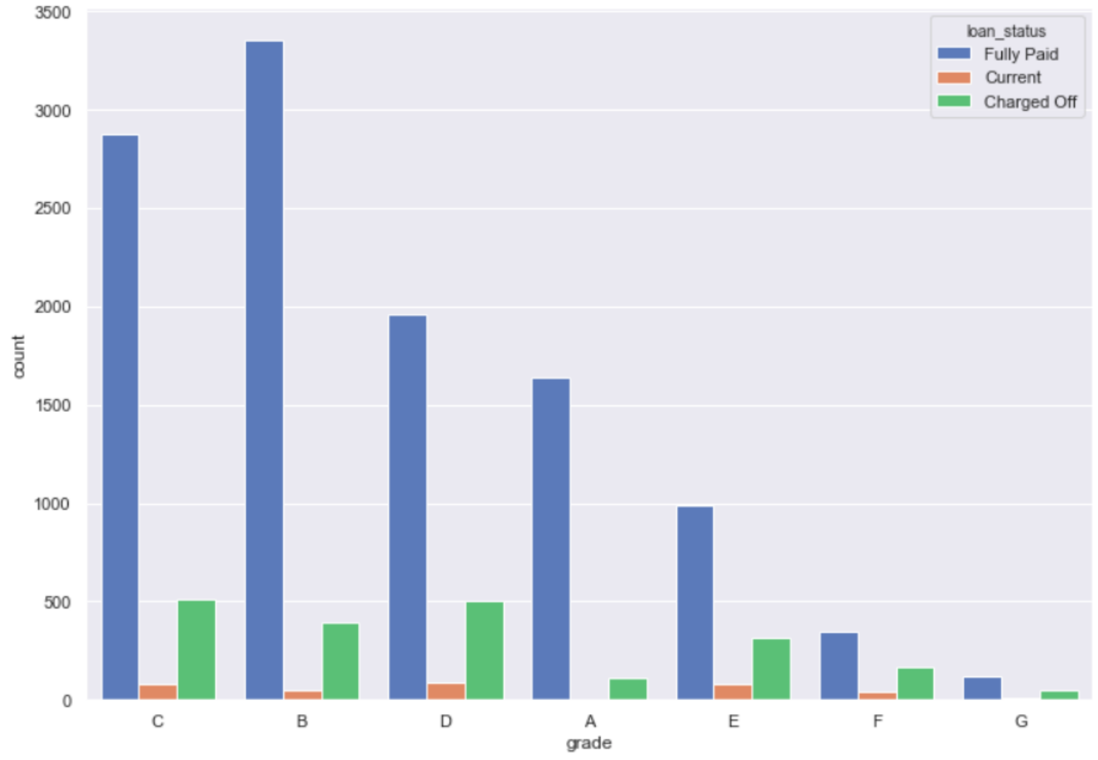
• Plots through EDA

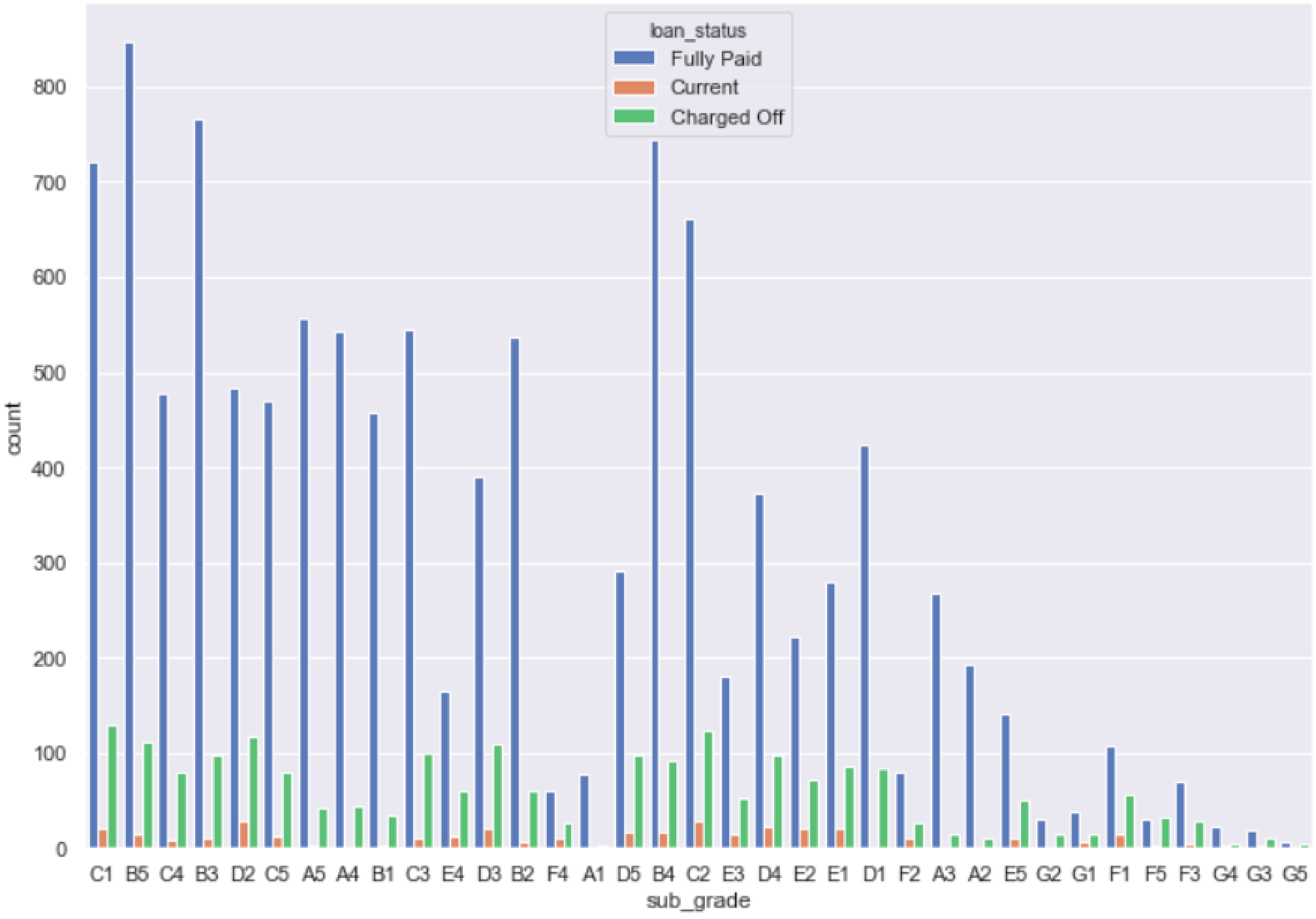




- *Most of the loans are for 36 months and few loans are for 60 months*
- *For 36 months loans 87% is fully paid and only 12% is charged off*
- *For 60 months loans 66% is fully paid and 23% is charged off*
- *Shorter period loans are more likely to pay compared to longer period loans*

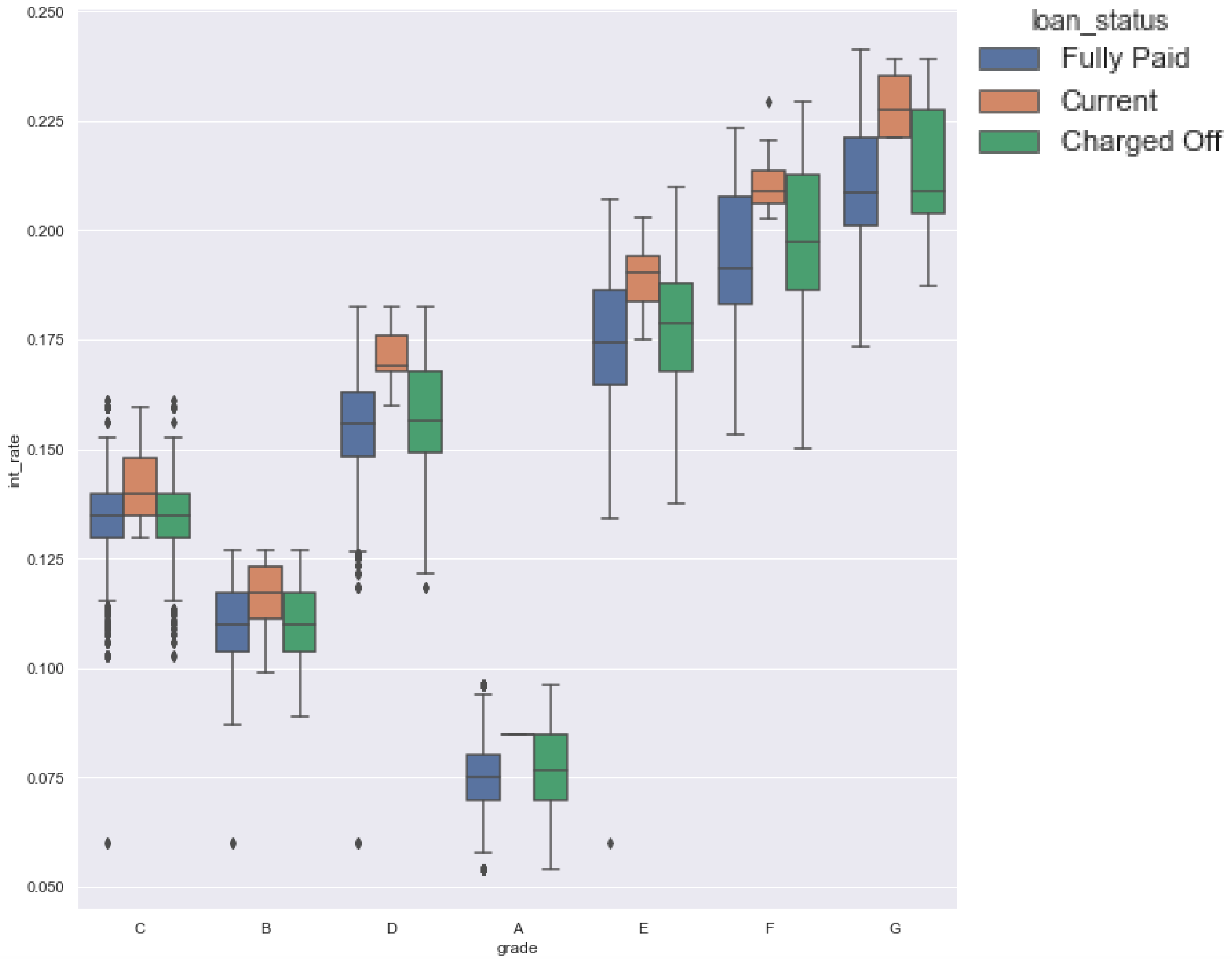
• Plots through EDA





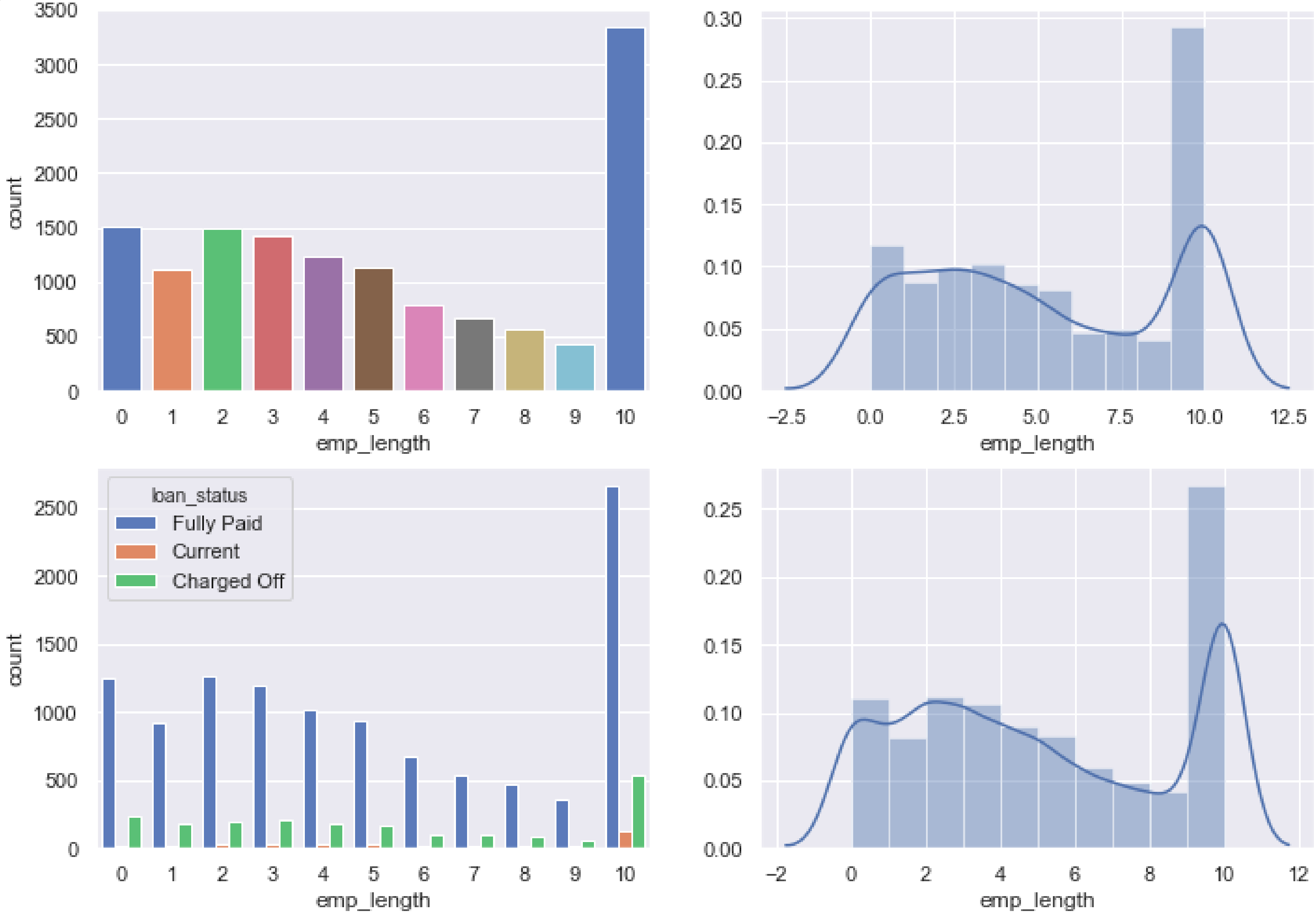
- *From the plots it can be noticied that the number of charged off loans is very less for the grade A and its subgrade compared to other grades.*

• Plots through EDA



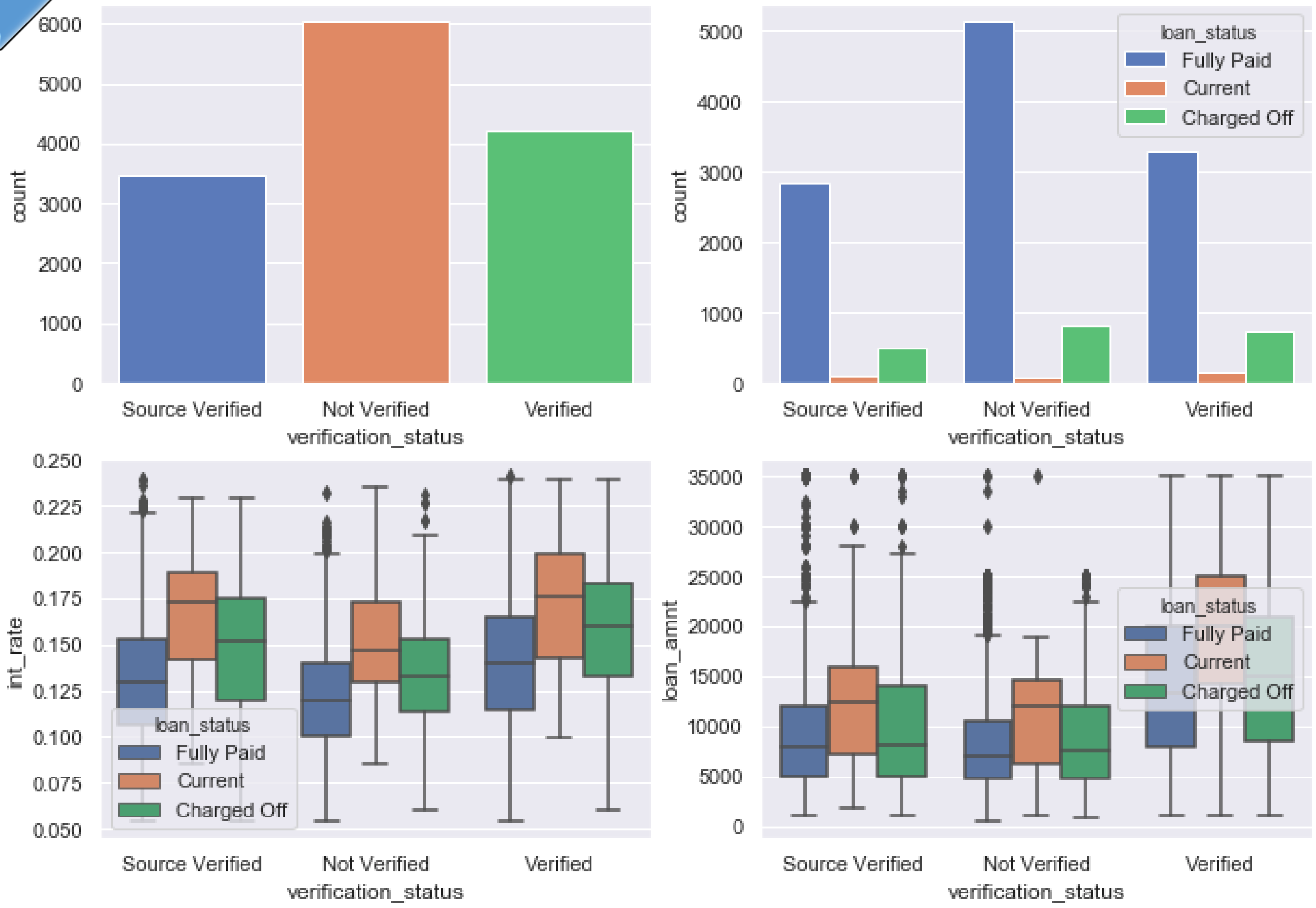
- *From the plots it can be noticed that the number of charged off loans is very less for the grade A and its subgrade compared to other grades. Once the applicant grade decrease from A to G , its most likely to fall in the defaulters list. From the past history the interest rate is increasing along with the applicant grade.*

• Plots through EDA



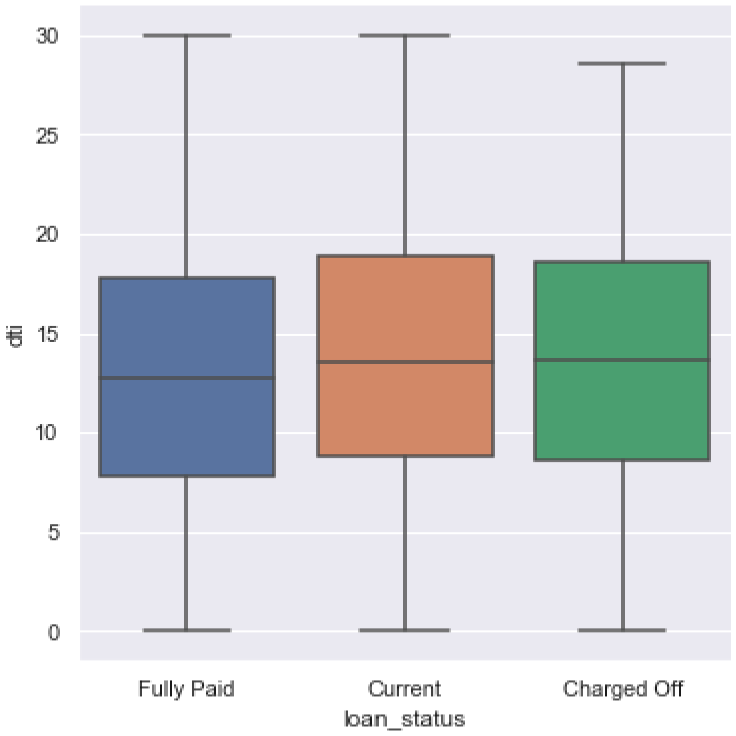
- *The EMP length major number of loans are taken by the employees having an year of experience less than one year to 5 , after that number of loans reducing and we can see a huge increase in the number of loan for 10+ year. The same pattern is followed for the charged off and fully paid loans*

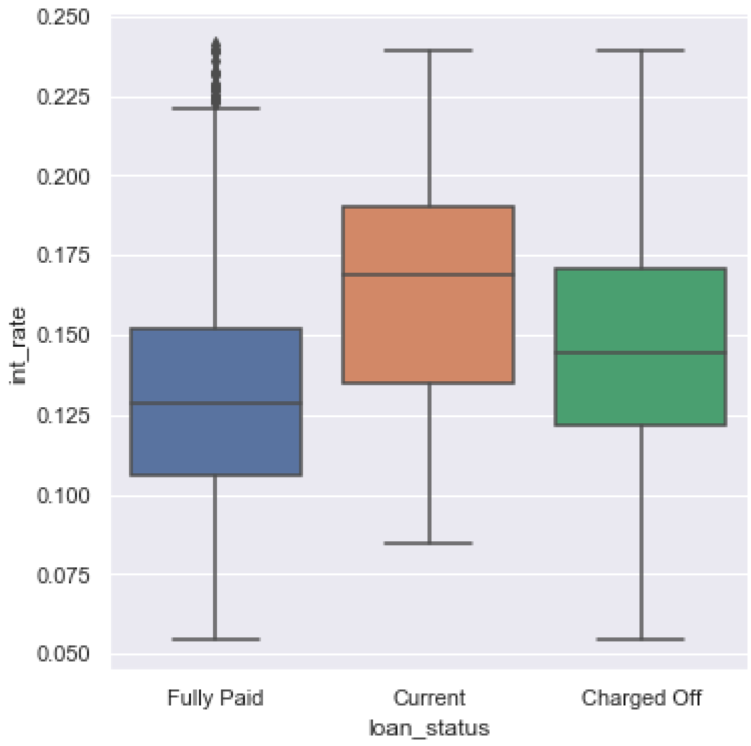
• Plots through EDA



- *From the plots we can make out that the verification status does not play any major role in deciding the defaulter list*

• Plots through EDA





- *DTI- A ratio calculated using the borrower’s total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower’s self-reported monthly income.*
- *Compared to fully paid loans charged off loans are having the higher DTI rate.*
- *Interest rate is more for charged off loans compared to the Fully paid loans*
- *Interest Rate is higher for the charged off loan compared to fully paid loans*

Conclusion

- Looks like most of the people are on Rent or mortgage
- Most of them are Not Verified
- B and C grades have more count as compared to G and F
- Debt consolidation is the highest in purpose
- DTI - Compared to fully paid loans charged off loans are having the higher DTI rate.
- Purpose - 28% of the loan taken for small business has been charged off and only 9% of the loan taken for wedding is charged off. So its shows a clear change in the charged off percentage for different loan purpose
- Term - Shorter period loans are more likely to pay compared to longer period loans
- Grade - The chances of not paying the loan on EMI date (charged off) is increasing while the grade is decreasing from A to G. So higher interest rate should be provided for the lower grade
- Interest rate – Interest Rate is higher for the charged off loan compared to fully paid loans