Lending Club Case Study

-Ishita Tripathi

Github link: https://github.com/ishitatr/Lending-Club-Case-Study

Github ID:https://github.com/ishitatr

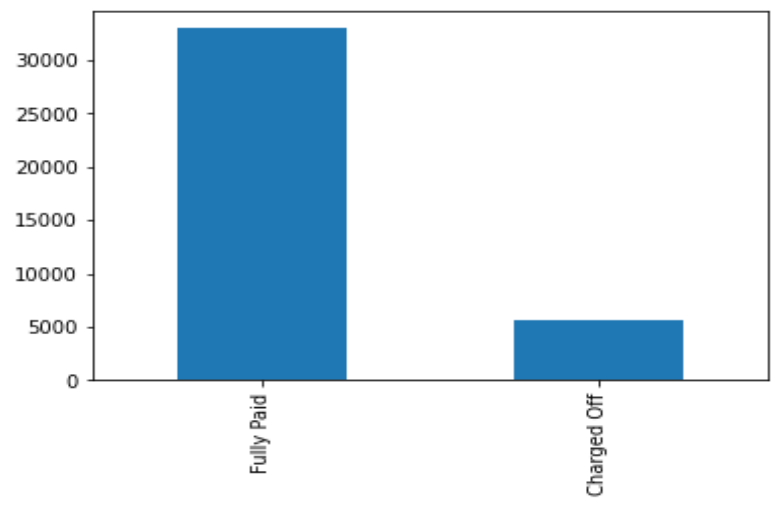
Introduction

- The Lending Club Case Study aims to find driver parameters to estimate default of any borrower
- This is done based on past data available from the lending club from the year 2007 to 2011
- The data is available for people who have either fully paid the loan, their loan was charged off (default) and the people who are currently paying the loan
- Based on the above mentioned past data, precursors which can predict default are determined

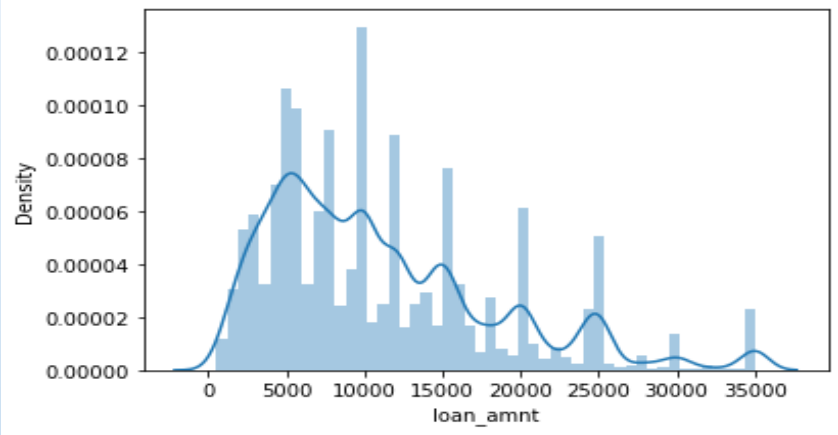
Data Cleaning and Manipulation

- In the source file given by the Lending Club , there are many variables for which data is not available
- There are many columns which are not a precursor to loan default, rather are a result of the loan default e.g. ‘recoveries’ column which is related to recovery of a loan default
- There are some generic columns such as id, member id
- The rows containing NA data are removed
- The int_rate column which is the interest rate of the loan has % symbol
- The loan status is only required for loans which are either fully paid or charged off
- Hot one encoding for loan status is done for easier analysis- 1 for charged off and 0 for fully paid loans

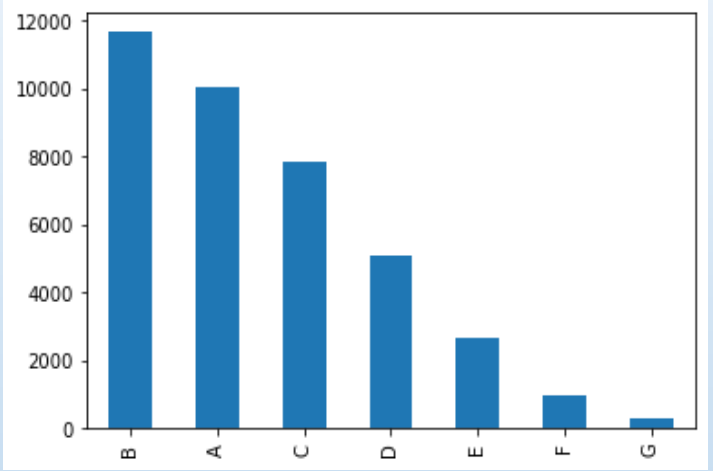
Univariate Analysis



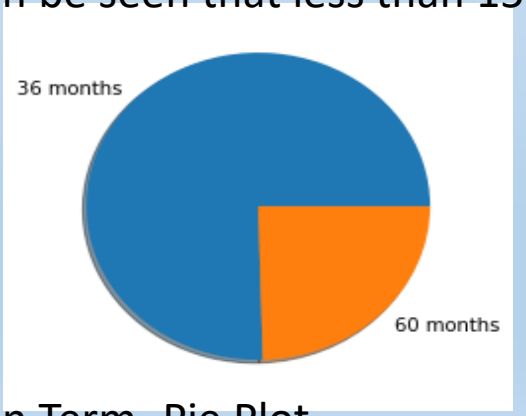
Loan Status- Bar Plot
It can be seen that less than 15%



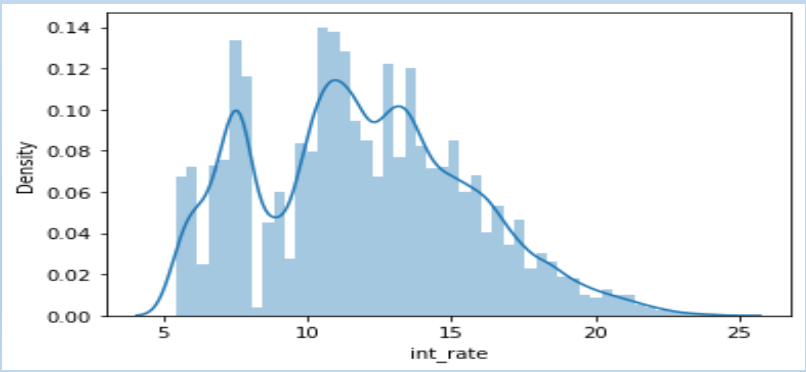
Loan Amount- Distribution Plot
It is right skewed



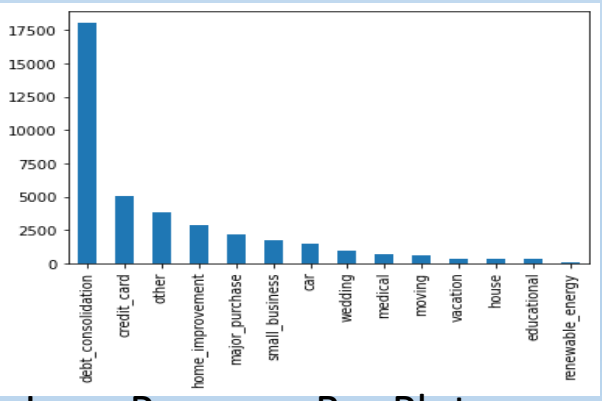
Loan Grade Bar Plot
Most of the loans are A,B,C grade



Loan Term- Pie Plot
Most of the loans are 36 months loans



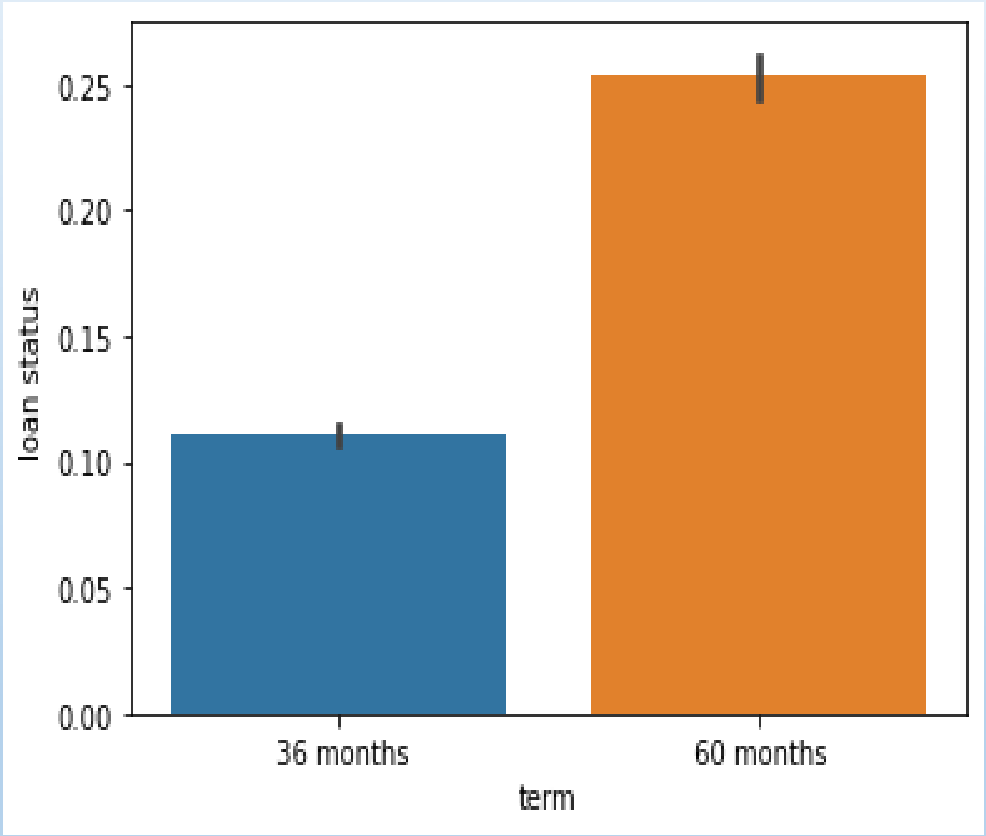
Loan Interest Rate-Distribution Plot
There are two types of loans-less than 9% and greater than 9%



Loan Purpose- Bar Plot
Most of the loans are taken for debt consolidation

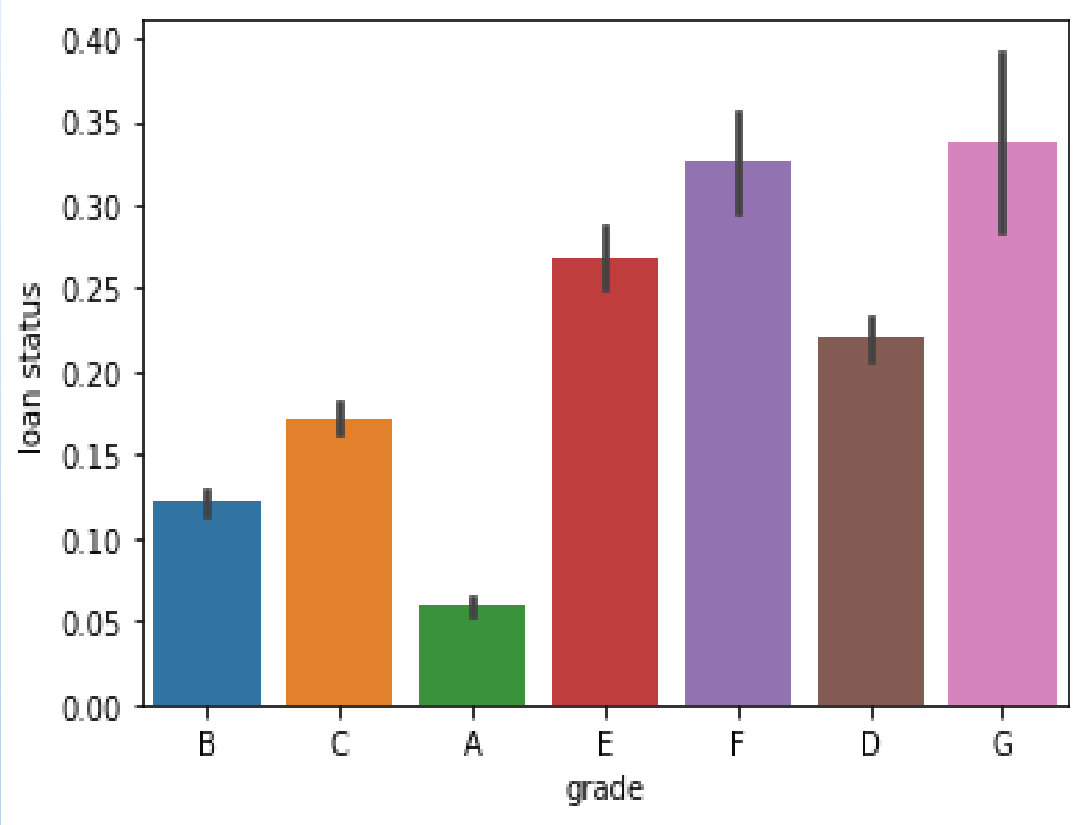
Bivariate Analysis-Loan Term and Loan Status

- The loans can be of two terms- 36 months or 60 months
- According to the barplot analysis of the loan term against loan status, most of the loan defaults occur for 60 months loan
- Almost 25% of the 60 months loans end up in default as compared to almost 11% of the 36 months loans
- The average interest earned per loan on 36 months loan is Rs 3254 whereas it is Rs 11835 for 60 months loan



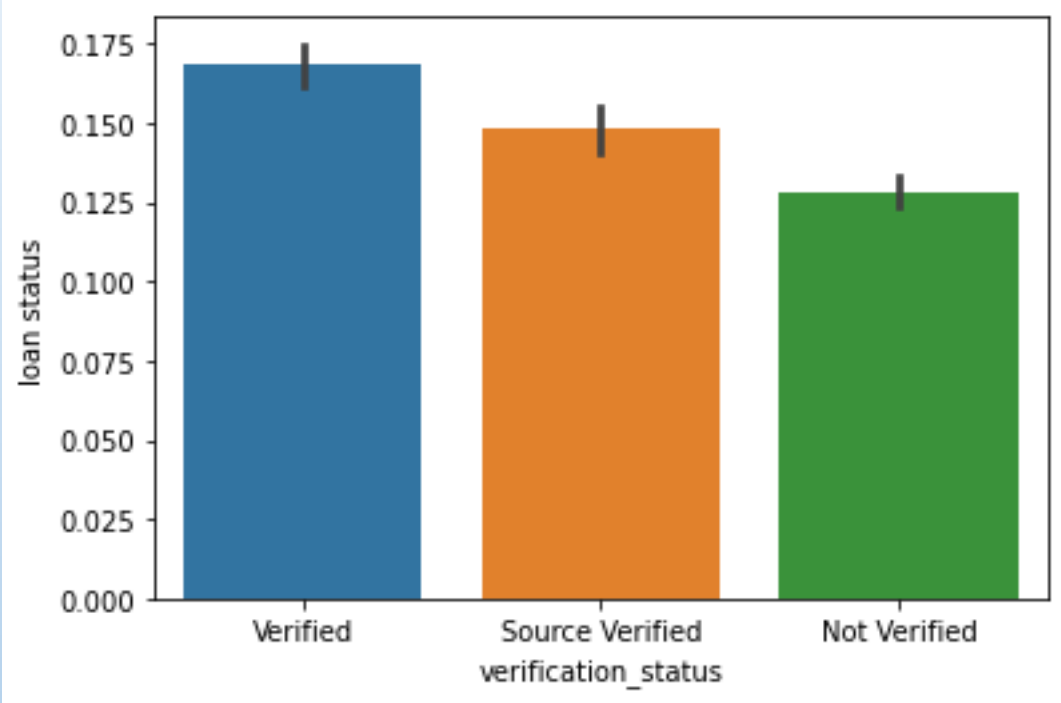
Bivariate Analysis-Loan Grade and Loan Status

- There are 7 loan grades- A to G, in decreasing order of loan safety
- i.e. loans with grade A are rated well as compared to grade G
- This can also be seen the bar plot where loans with grade A default the least and loans with grade G default the highest



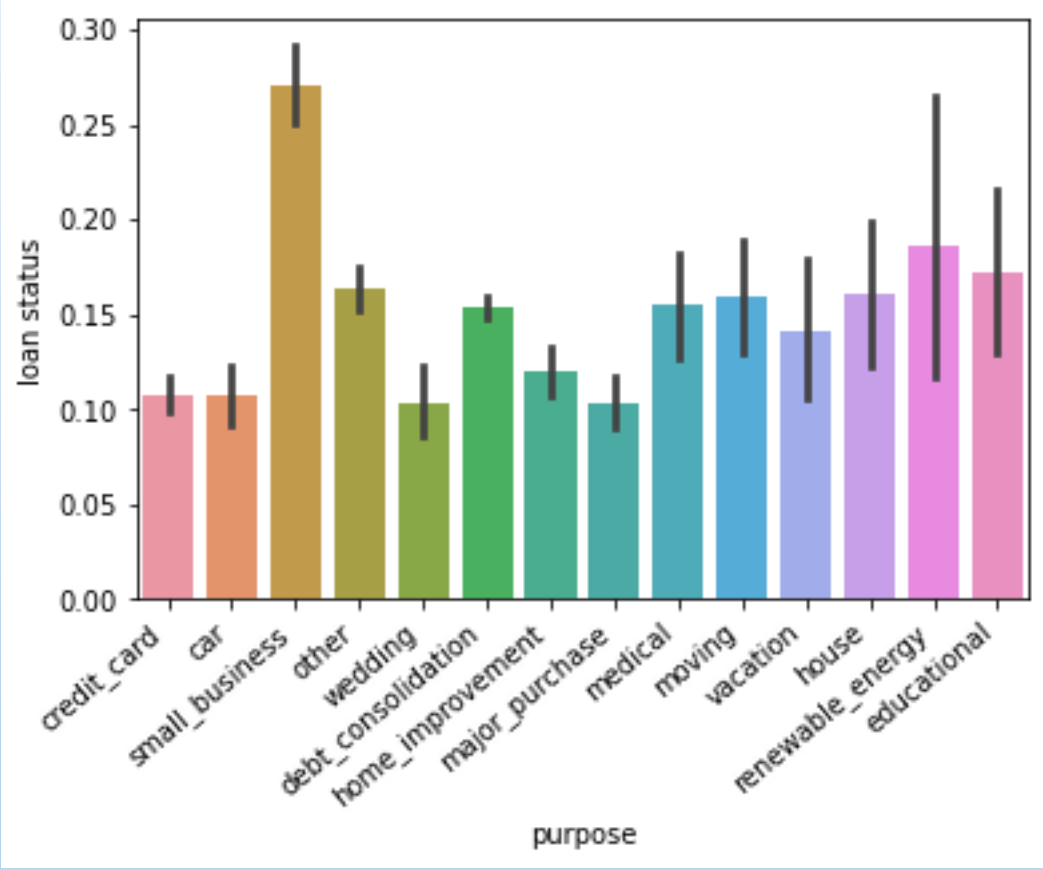
Bivariate Analysis-Verification Status and Loan Status

- The verification status can be either verified, source verified or not verified
- It was observed that verified and source verified loans default the highest



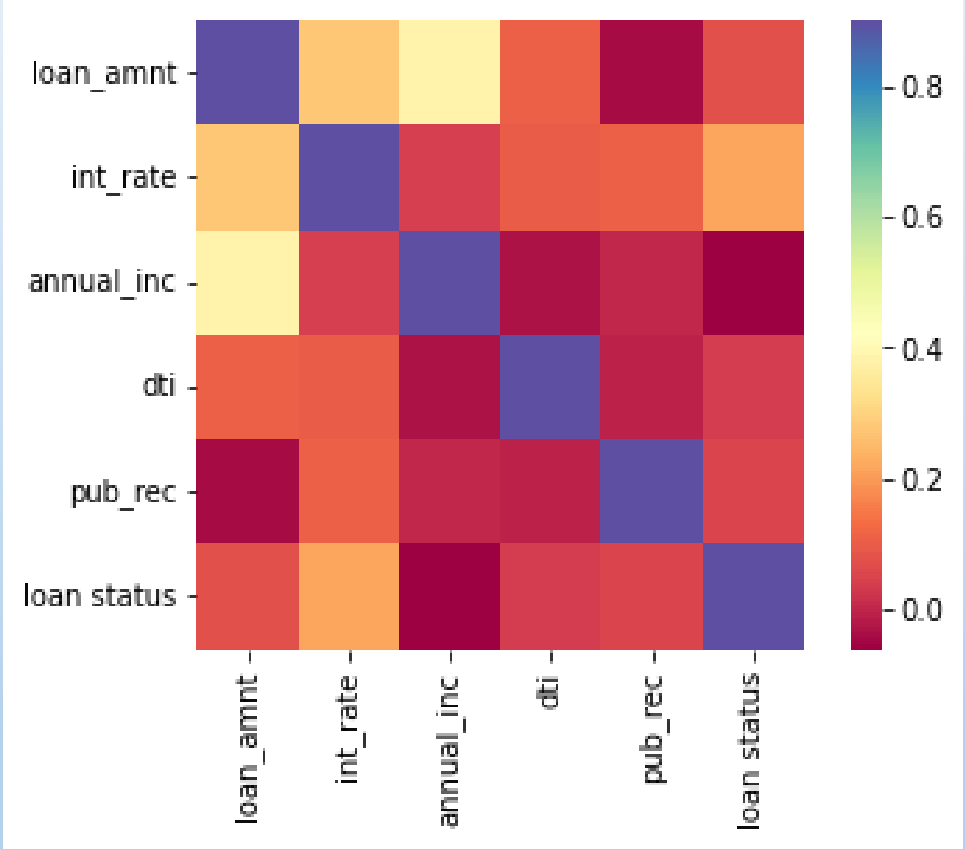
Bivariate Analysis-Loan Purpose and Loan Status

- It can be seen that loans taken for the purpose of small businesses have a chance greater than 25% of default
- Loans for credit card, car, wedding and major purchases have the least probability of default (approx. 10%)



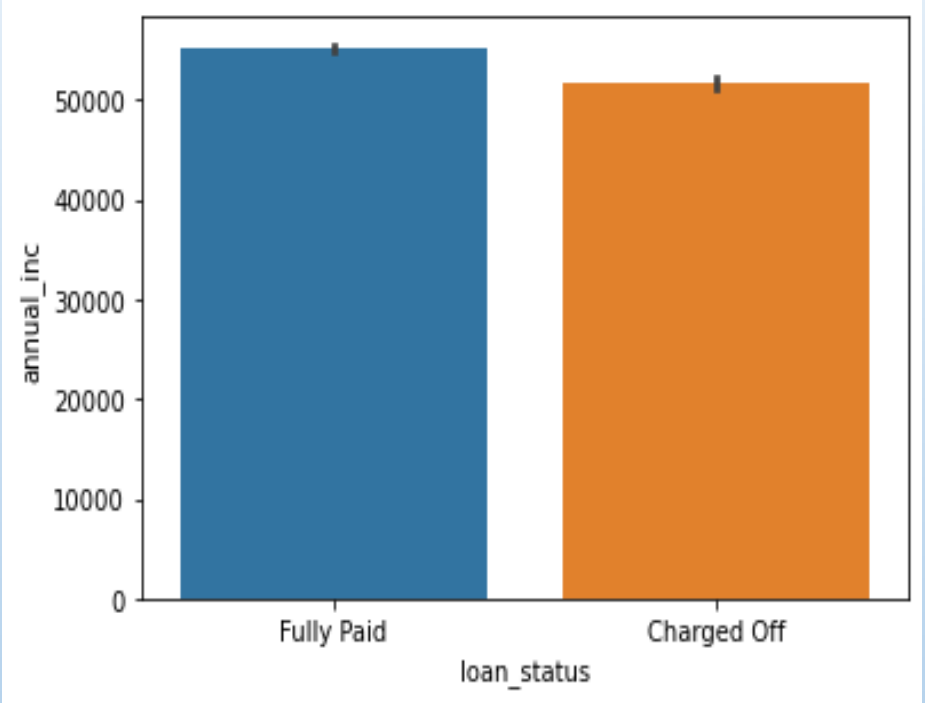
Correlation Matrix

- Correlation Matrix is constructed to find out the correlation between various numerical variables and loan status
- This can be used to further analyze the relationship between the high correlated variables with loan status
- It can be seen that loan status is correlated with interest rate of the loan, loan amount, annual income, public record and DTI in descending order of correlation



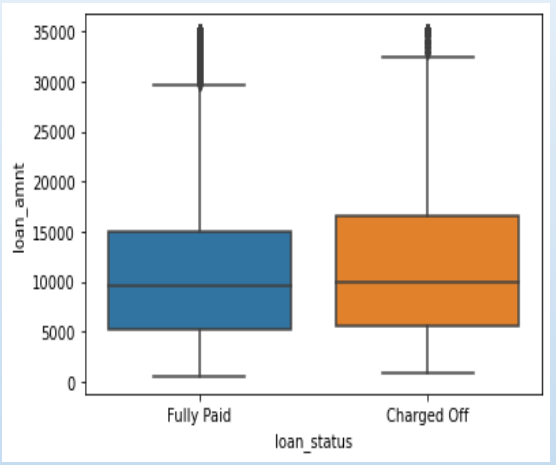
Bivariate Analysis- Annual Income with Loan Status

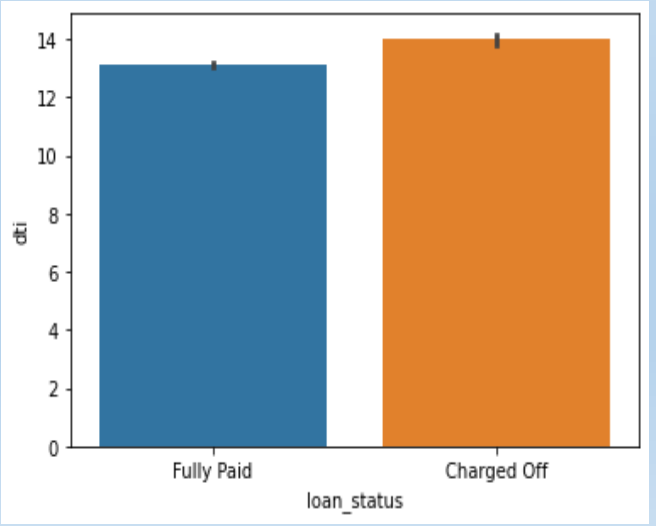
- The average annual income of people who default on loans is around Rs 52,000 whereas for people who fully pay the loan, the average annual income is Rs 50,000
- People with an annual income greater than Rs 1,00,000 all have paid the loan



Bivariate Analysis- Loan Amount, Interest Rate and DTI with Loan Status

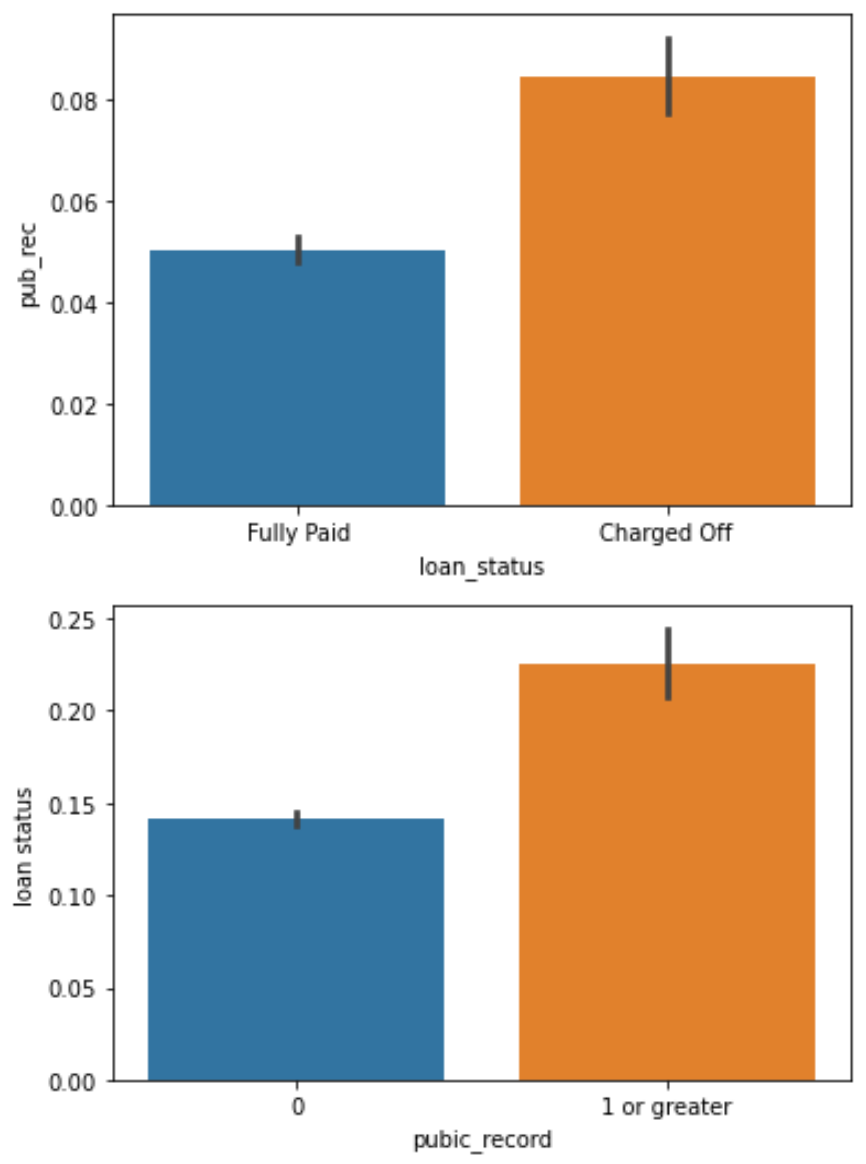
- The average loan amount is found to be higher for charged off loans
- The average interest rate is higher for charged off loans
- The average monthly debt to income ratio (DTI) is around 14% for fully paid loans as compared to 13% for charged off loans





Bivariate Analysis- Public Record and Loan Status

- Almost 8 people in 100 people with derogatory public record defaulted on their loan as compared to 5 people in 100 people
- Almost 15% people defaulted on their loan with no public records, as compared to 22% of the people who have public records and defaulted on their loan



Conclusion

- There are two types of attributes which can predict the default of the loan – loan attributes and person attributes
- Loan attributes which can predict default- Loan Term , Loan Grade, Verification Status, Loan Amount, Interest Rates
- Person Attributes- monthly debt to income ratio, public records and annual income of the person

Thank You!!