LENDING CLUB ASSIGNMENT

SUBMISSION

Name:

Kiran Gupta

Parama Bhattacharya

Lending Club Analysis

Introduction: Consumer Finance Company which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.
Two **types of risks** are associated with the bank's decision:

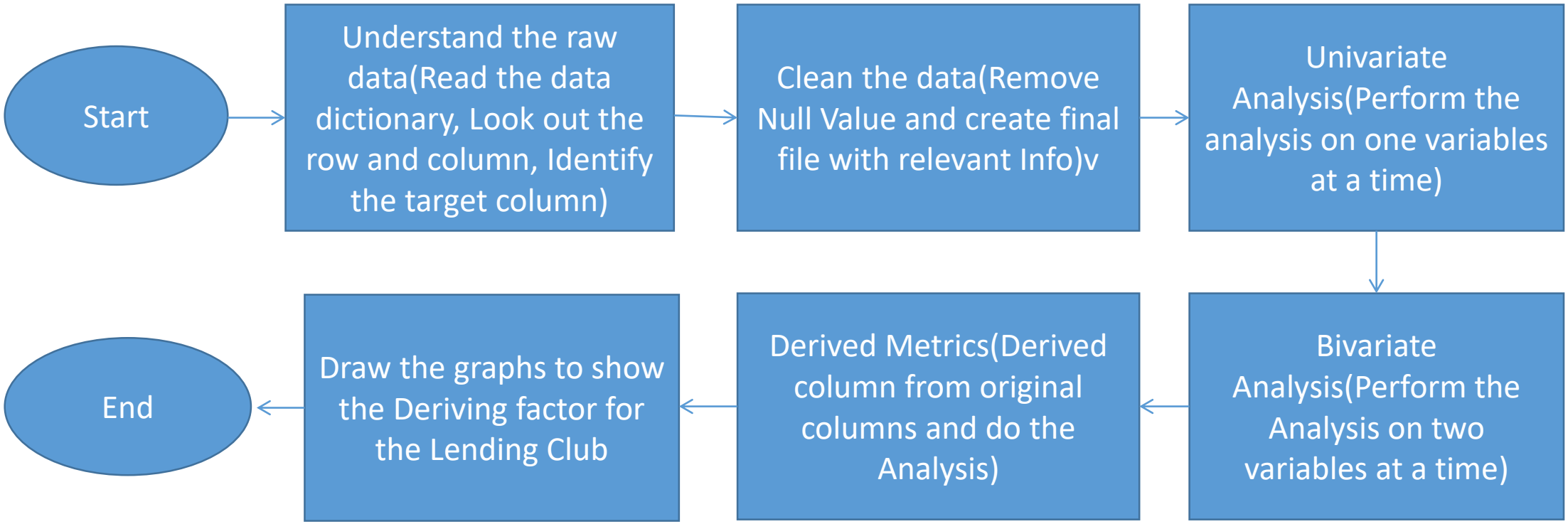
- If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company
- If the applicant is **not likely to repay the loan,** i.e. he/she is likely to default, then approving the loan may lead to a **financial loss** for the company

  Sample data of loan contains the information about past loan applicants and whether they 'defaulted' or not.

Objective:

The aim is 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.

Flowchart


Start
Understand the raw data(Read the data dictionary, Look out the row and column, Identify the target column)
Clean the data(Remove Null Value and create final file with relevant Info)v
Univariate Analysis(Perform the analysis on one variables at a time)
Bivariate Analysis(Perform the Analysis on two variables at a time)
Derived Metrics(Derived column from original columns and do the Analysis)
Draw the graphs to show the Deriving factor for the Lending Club
End

Analysis

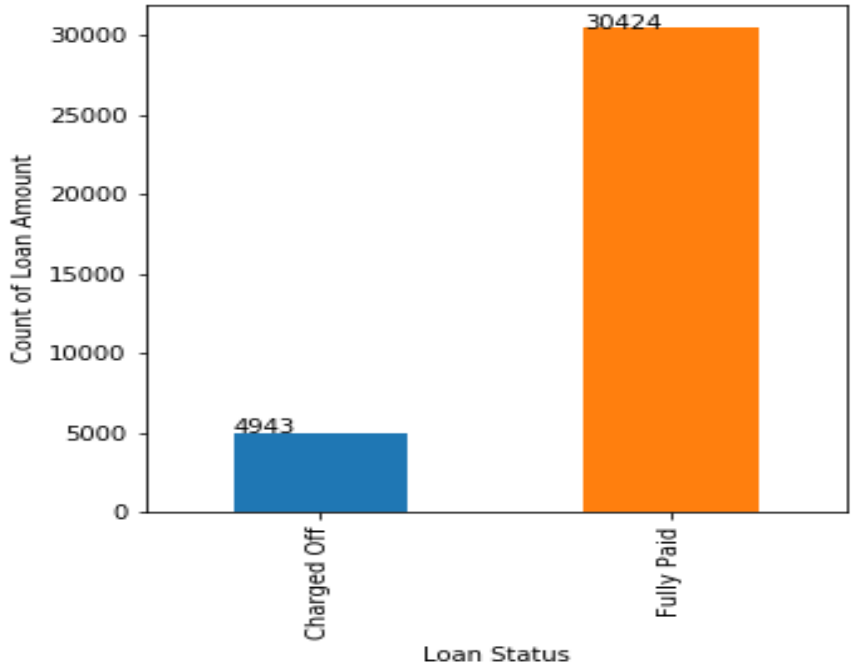
- There are one loan file and one data dictionary provided in the file.
- Remove the null value and do relevant data cleaning
- Identify the major columns which indicate the default on the loan.

Loan Status: Current status of the loan. There are three categories:

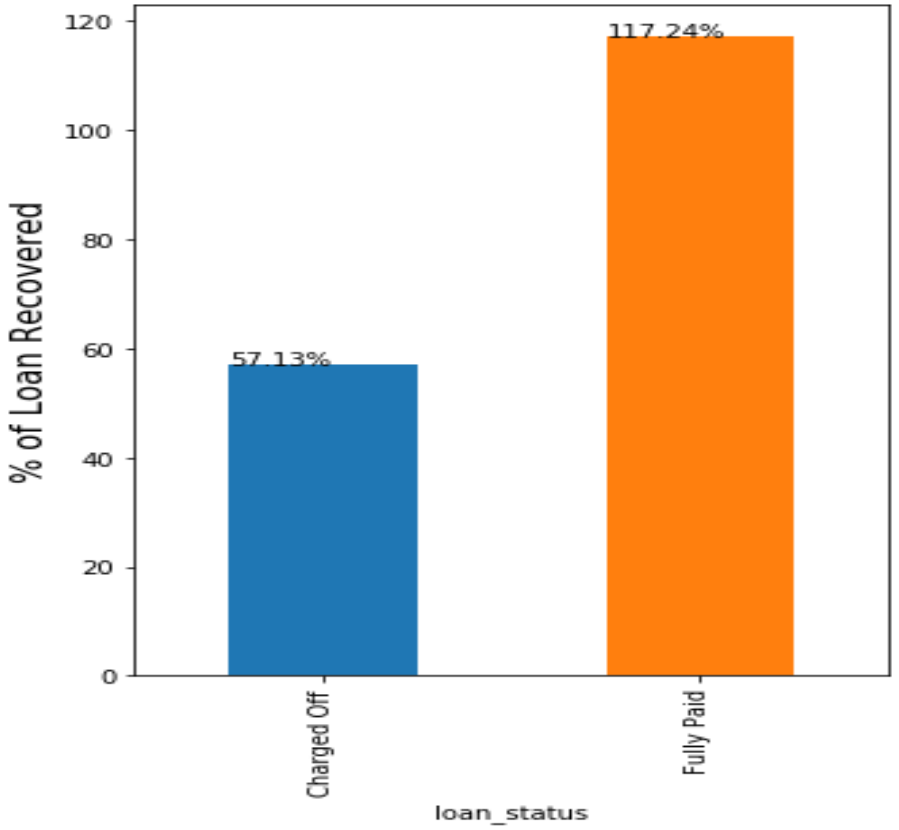
- **Fully paid:** Applicant has fully paid the loan (the principal and the interest rate)
- **Current:** Applicant is in the process of paying the installments, i.e. the tenure of the loan is not yet completed. These candidates are not labeled as 'defaulted'.
- **Charged-off:** Applicant has not paid the installment's in due time for a long period of time, i.e. he/she has defaulted on the loan

Univariate Analysis

- Below graph shows relation Between Loan Amount Count vs Loan Status
- It is evident that around 5K Charged off Loan and 30K Fully Paid.
- Only around 57% of money is recovered from Charged off Loans.
- From Fully Paid loan, Lending club earning around 17% Profit.



Graph1: Loan Status vs Loan Amount

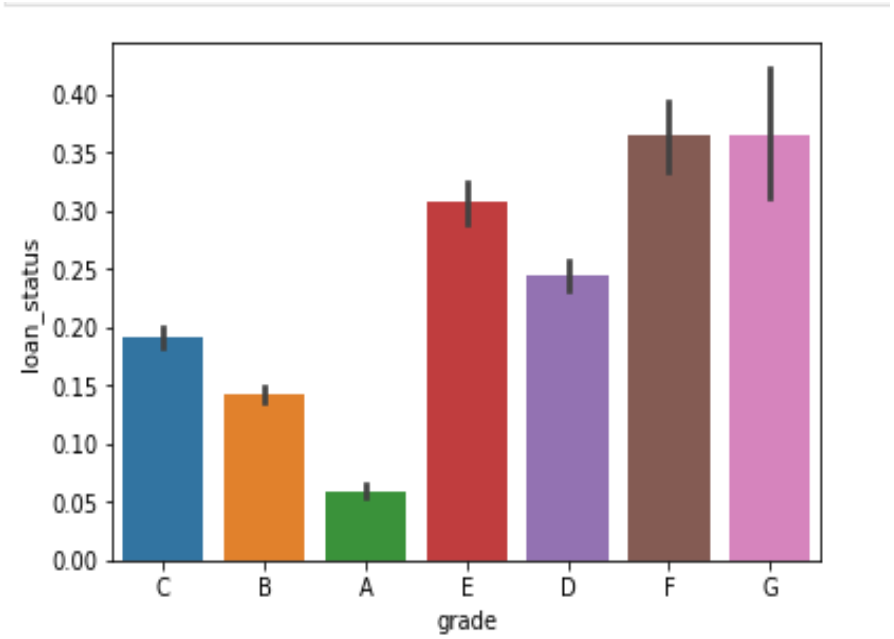


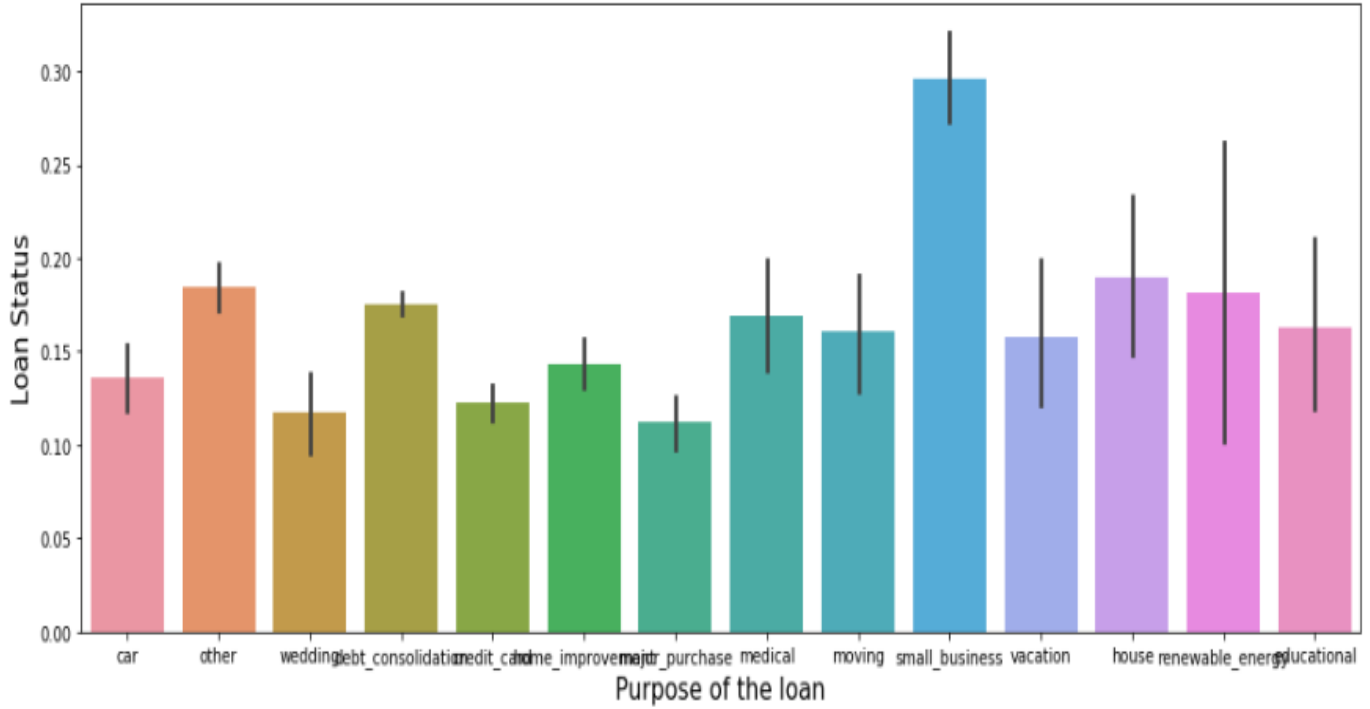
Graph2: Loan Status vs Loan Recovered

Univariate Analysis

- It is evident from the plot that as the grade of loan goes from A to G, the default rate increases. This is expected because the grade is decided by Lending Club based on the risk nature of the loan.
- From the above graph, we can see that the following "purpose" are the ones which is strongest contender of being defaulted.

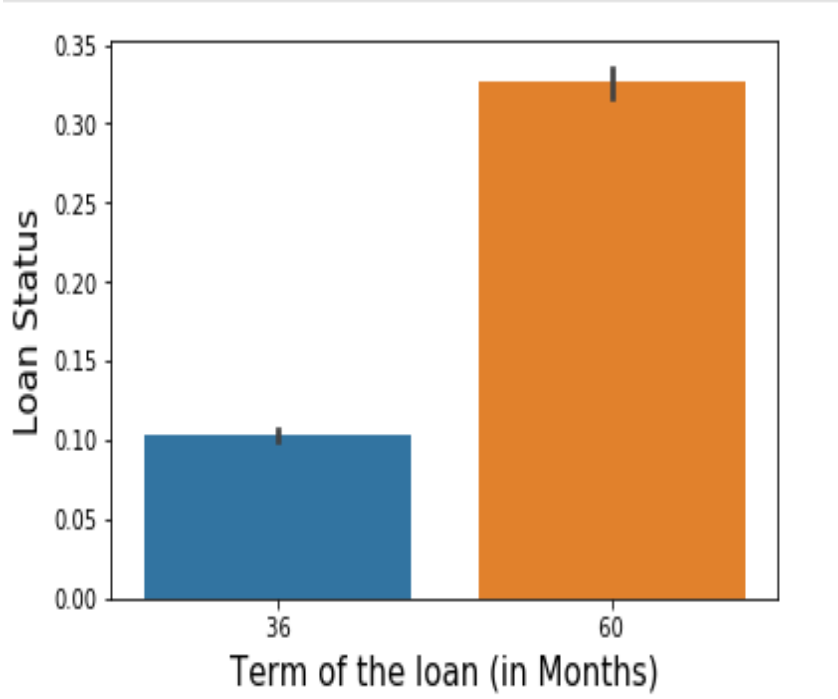
→Small business , Renewable Energy and educational

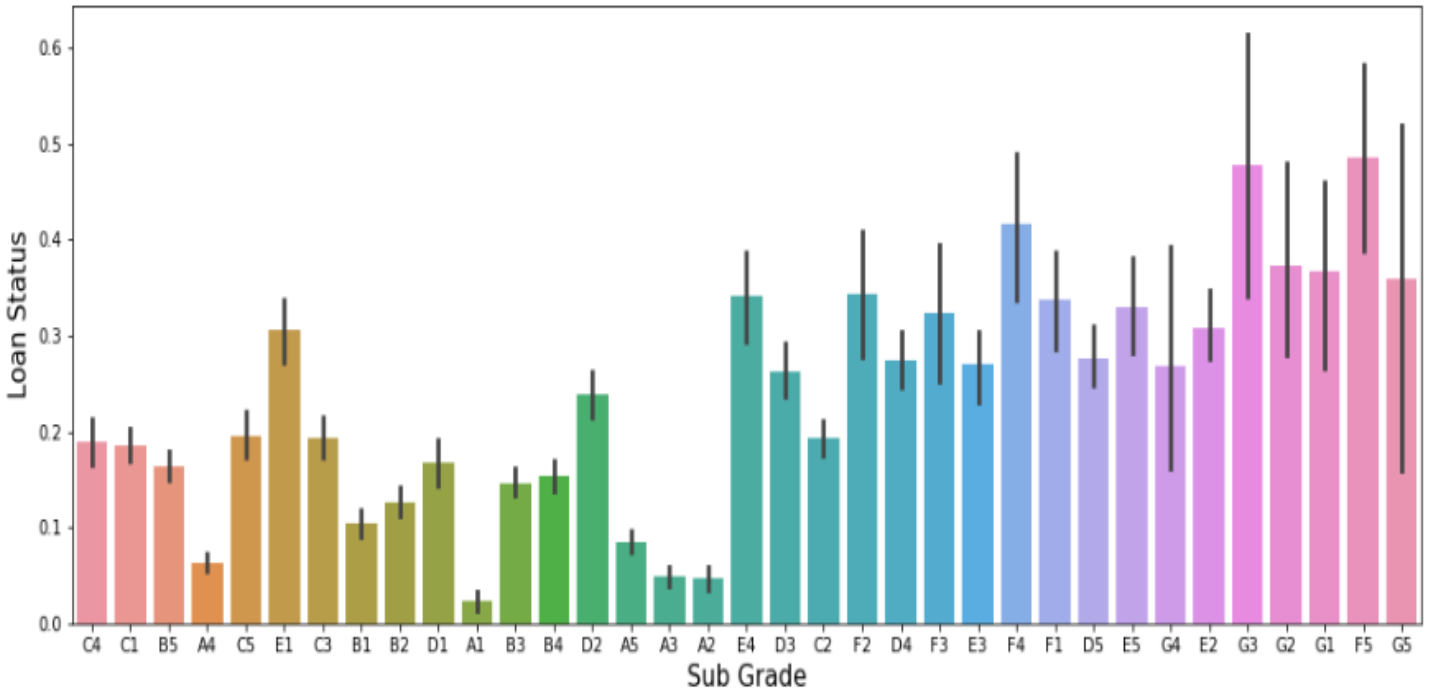




Univariate Analysis

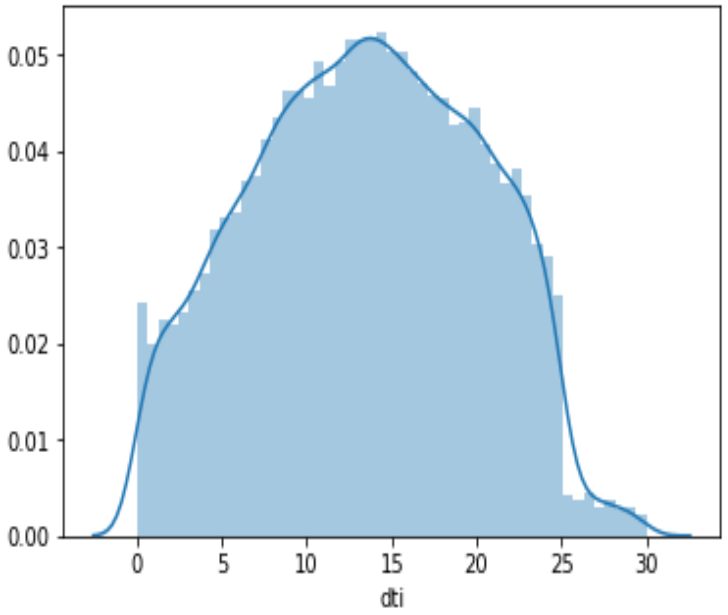
- It clearly shows that loan taken for a longer term which is 60 months is a stronger candidate for default.
- It is evident from the plot that as the sub-grade of loan goes from A1 to A2, A2 to A3, A3 to A4 and A4 to A5, the default rate increases. This is expected because the sub-grade is also decided by Lending Club based on the riskiness of the loan.

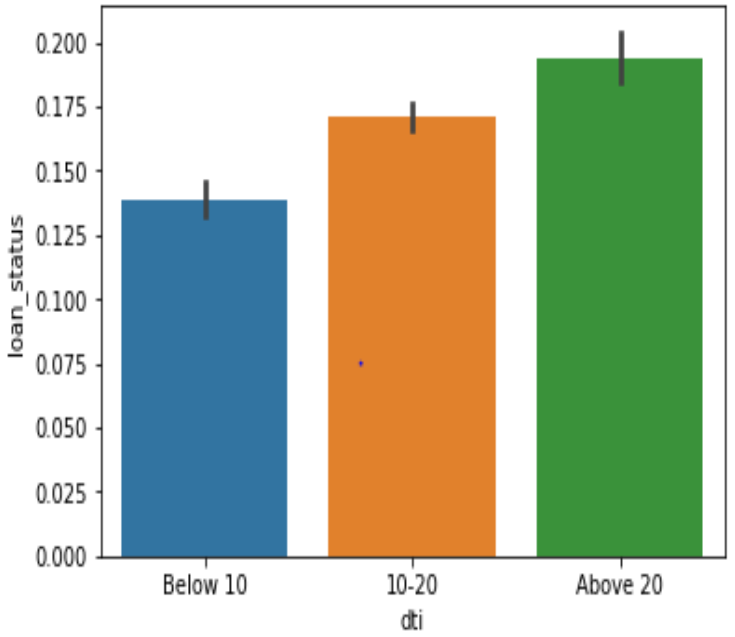


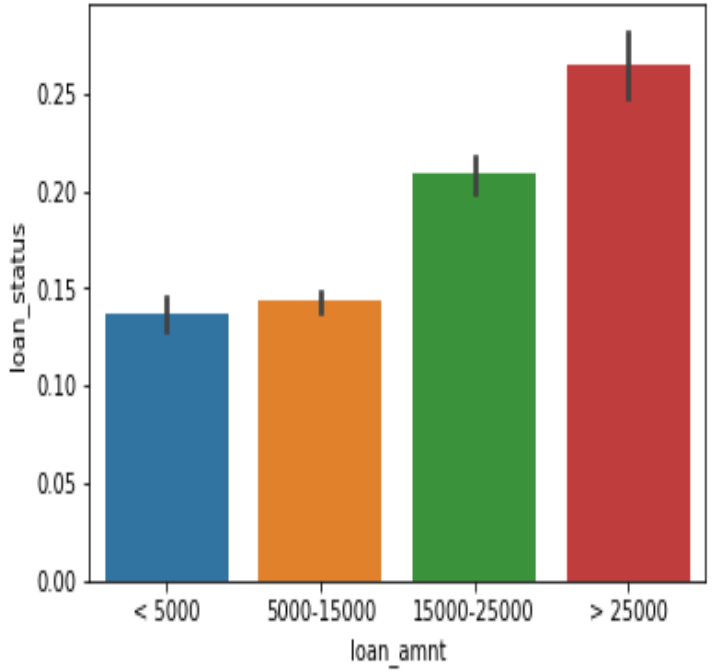


Univariate Analysis

- It clearly shows that higher risk of default where DTI Relation is greater than 15.
- It is clearly visible that higher debt to ratio values have higher risk of defaulting.
- It is clear from the plot that higher the value of "loan Amount", greater is the risk of defaulting.

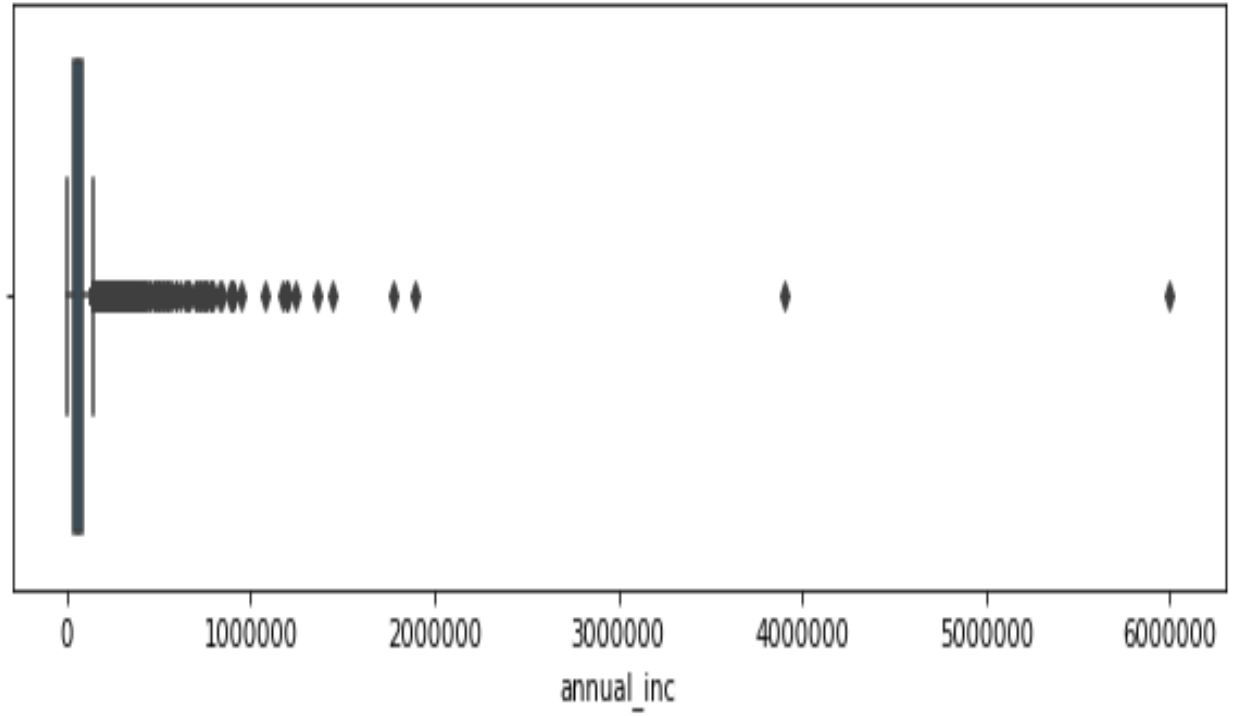


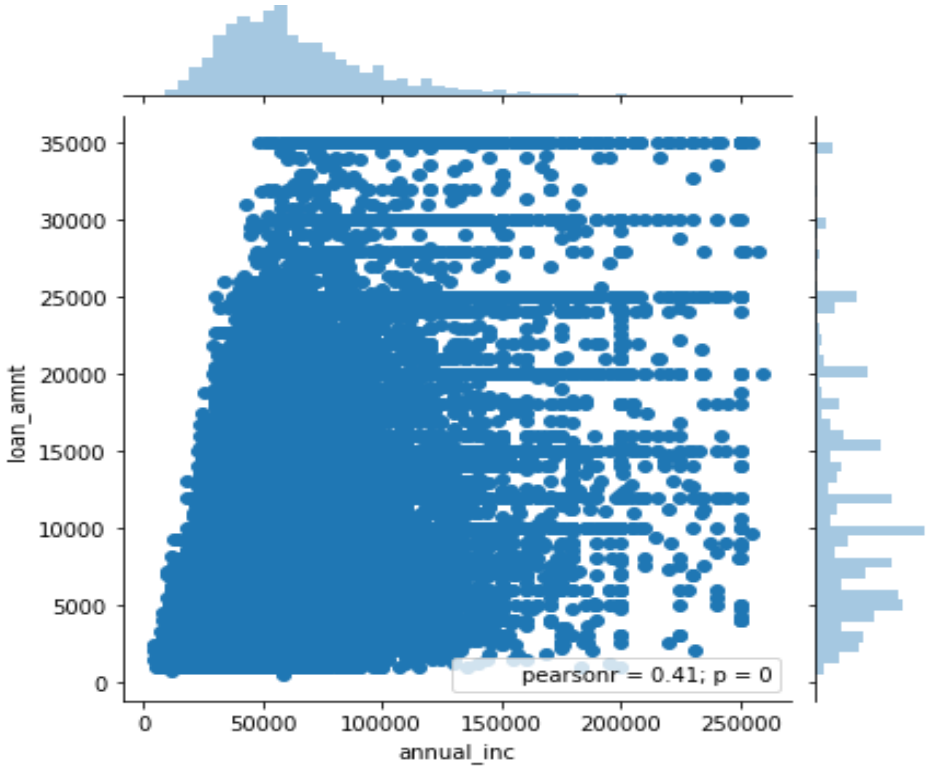




Bivariate Analysis

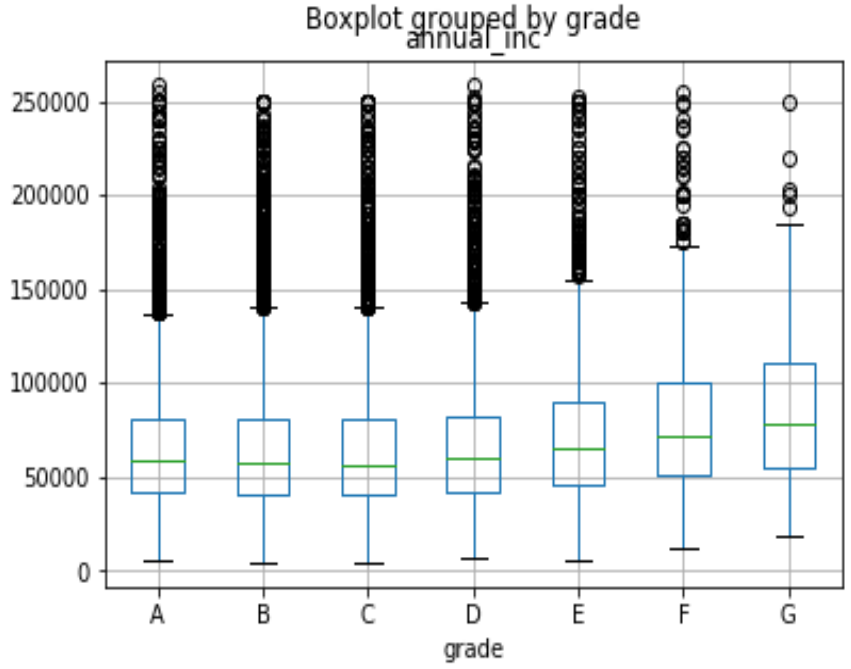
- From the Box Plot it clears that Annual Income varies from 4000 to 6000000
- It is evident that people having lesser salary, less than 50000 and who have taken loan of 25000 and higher has a higher chance of defaulting.

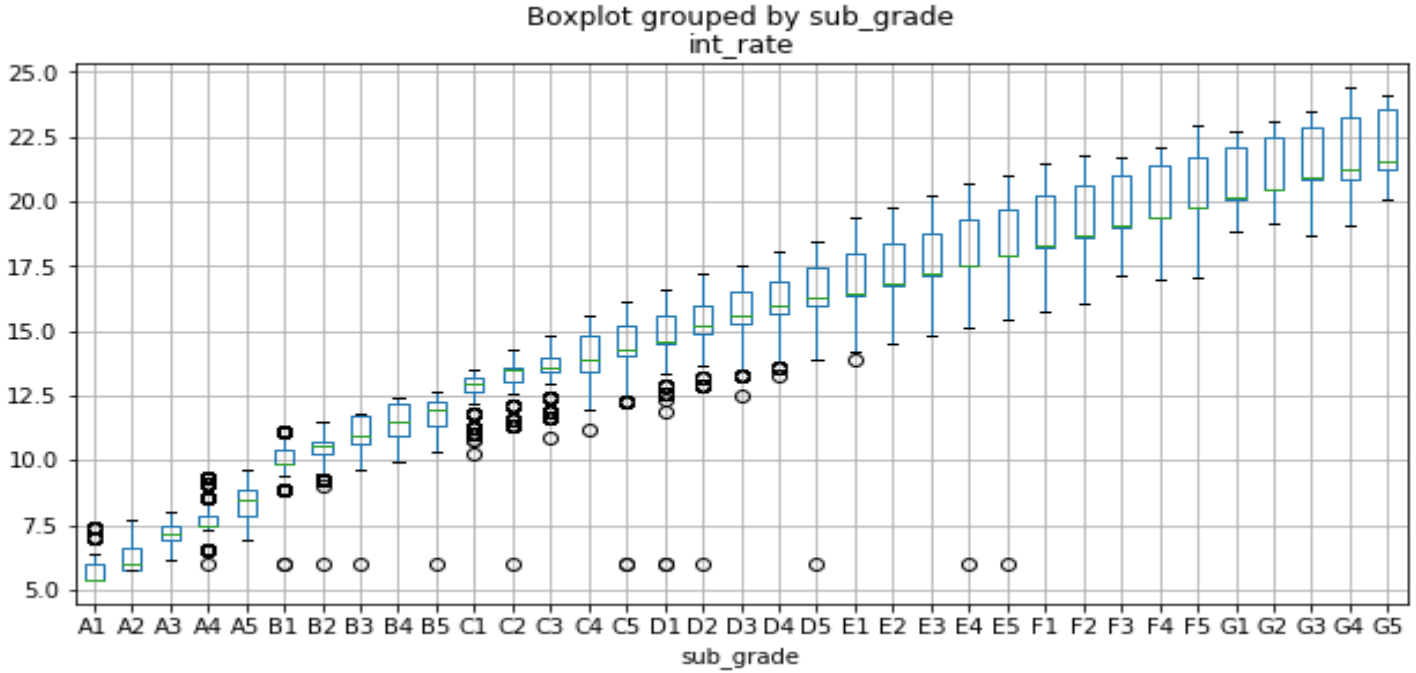




Bivariate Analysis

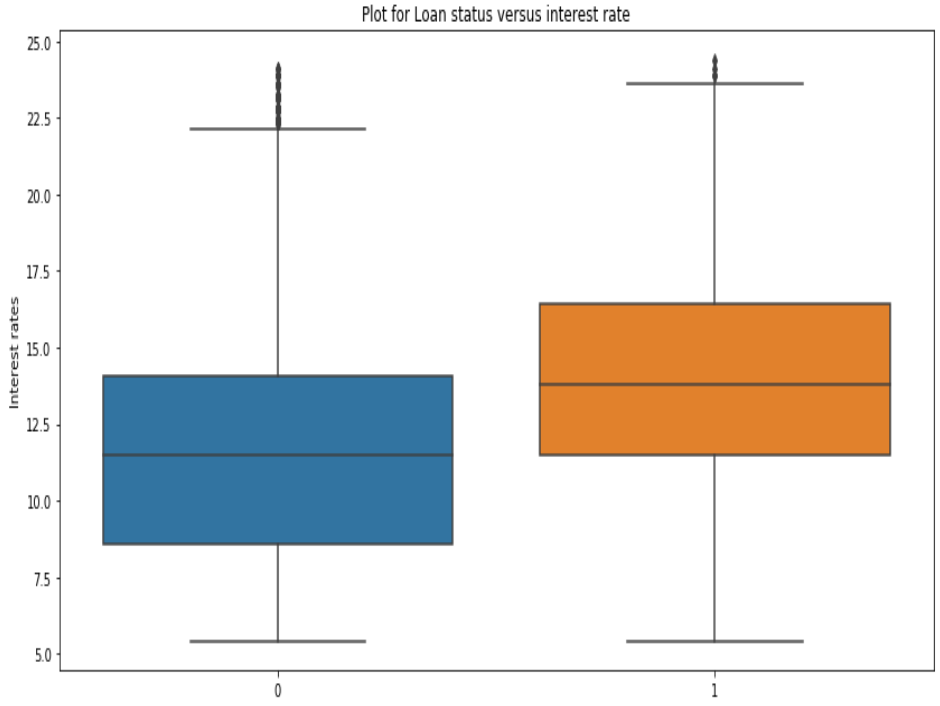
- From the graph, it shows that grade actually decreases with the increase in the loan amount
- It is evident from the plot that as the sub-grade of loan goes from A1 to A2, A2 to A3, A3 to A4 and A4 to A5, the default rate increases. This is expected because the sub-grade is also decided by Lending Club based on the riskiness of the loan.

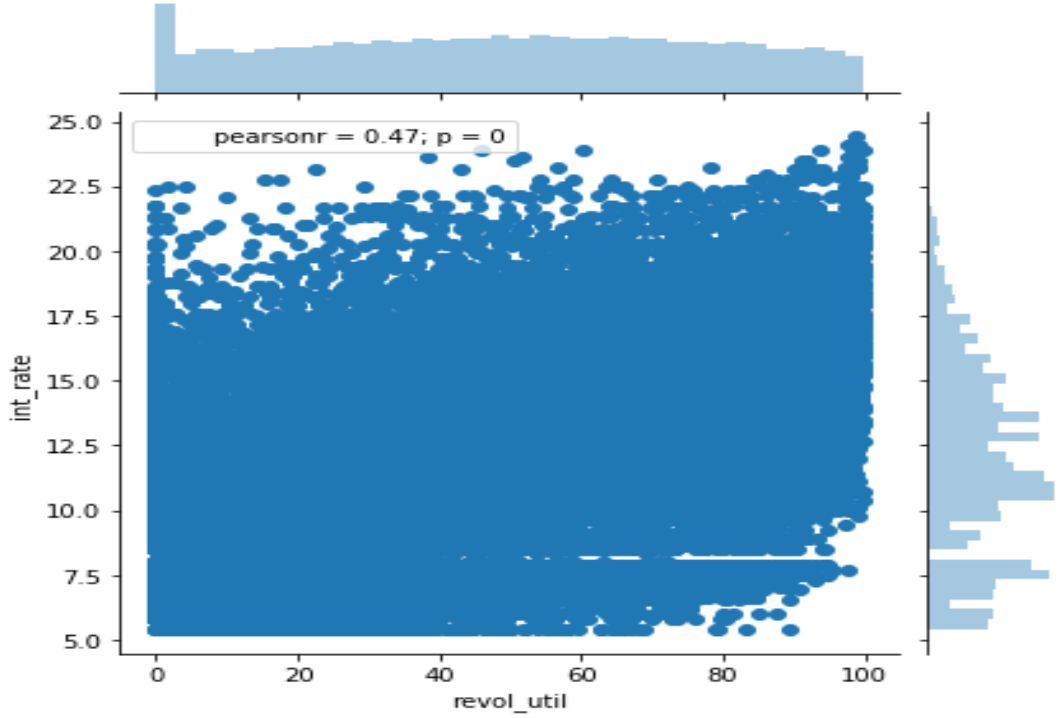




Bivariate Analysis

- From the graph, it shows that **Higher interest rates are likely to be charged off.**
- It is evident from the People having higher utilization revolving line of credit having higher interest rate which drives potential defaults.





Conclusion

- Stop the approval of loan where purpose is Small business.
- Stop approval of loan with higher amount.
- Stop approval of loan where annual income is low.
- Reduce the loan amount where DTI ratio is higher.
- Approve the loan where loan amount is lower.