# Lending Club Case Study

# Problem statement:

You work for a consumer finance company which specialises 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.

Find and understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. And utilise this knowledge for portfolio and risk assessment.

# Data Understanding:

- Remove the columns with percentage of Null values more than 60-70.

- loan_amnt : The listed amount of loan applied for by the borrower.

- loan_amnt : The listed amount of loan applied for by the borrower.

- funded_amnt : The total amount committed to that loan at that point of time.

- funded_amnt_inv: The total amount committed by investors for tat loan at that point in time.

- grade: LC assigned loan grade

- loan_status: cuurent status of the loan( fully paid, charged off, current)

- verification_status: Indicates whether income was verified by LC or not.

- Remove all customer behavioural columns which are not available during the submission of loan

# Data Cleaning:

- Fill Null values with appropriate data using methods like mean, median,mode or some constant
- Remove rows or columns which have significant number of nulls
- Loan status as Current means it is neither fully paid nor charged off, hence removing rows with loan status as current.

# Univariate Analysis:

By doing Univariate analysis on term, int_rate,dti with loan_status as condition, we were able to decide where the maximum distribution is present and decide outliers as well.
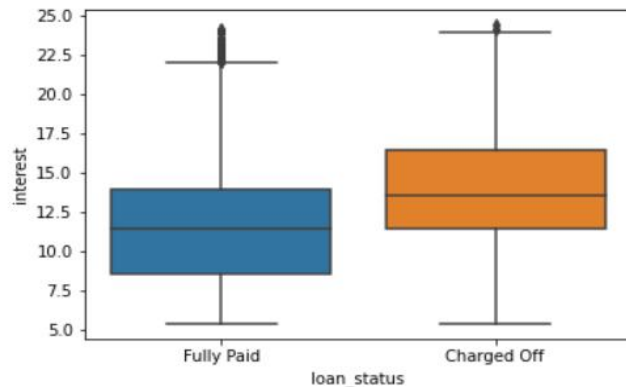
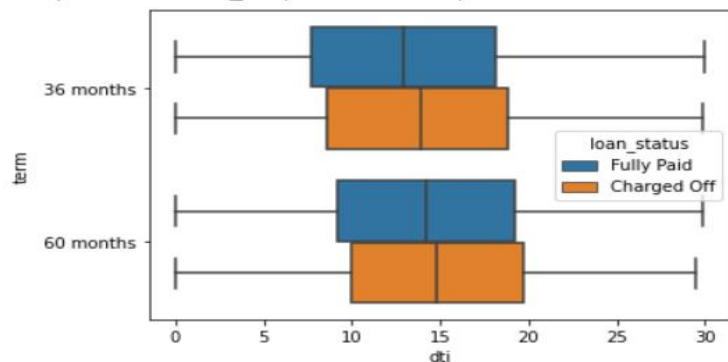Below are the dstribution plots.
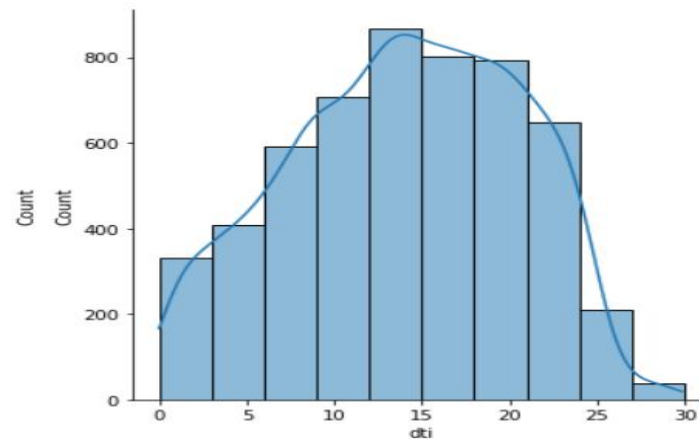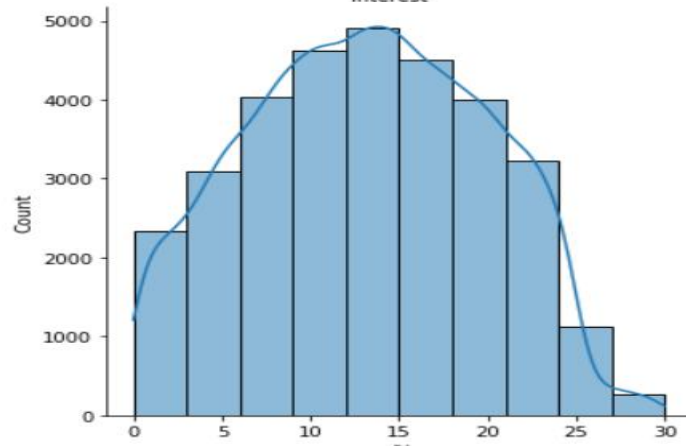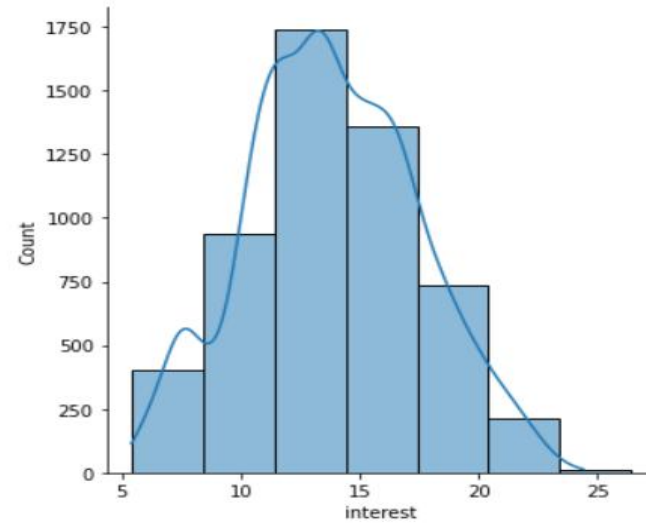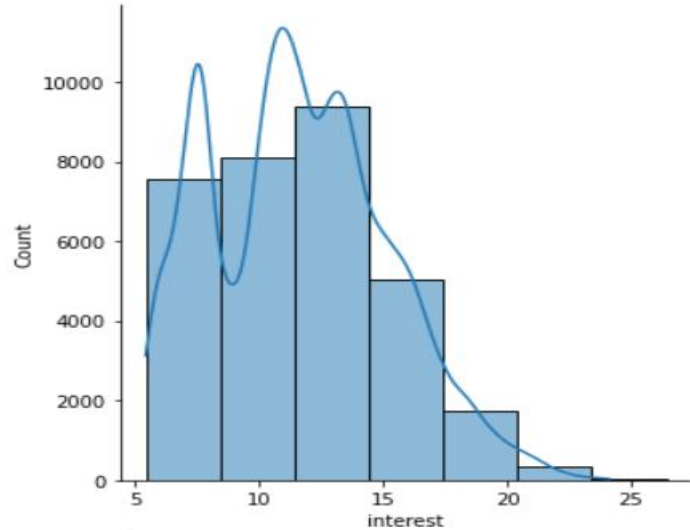
# Bivariate Analysis:

By doing Bivariate analysis like below, we were able to decide threshold values based on the relation ships or dependencies by which the "loan_status" can go to "Charged Off".
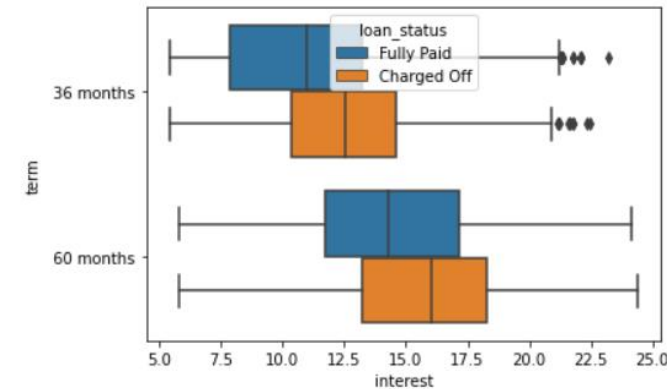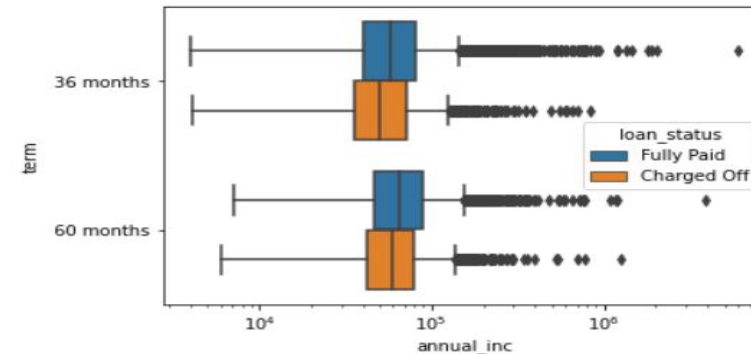
"int_rate" and "loan_status"


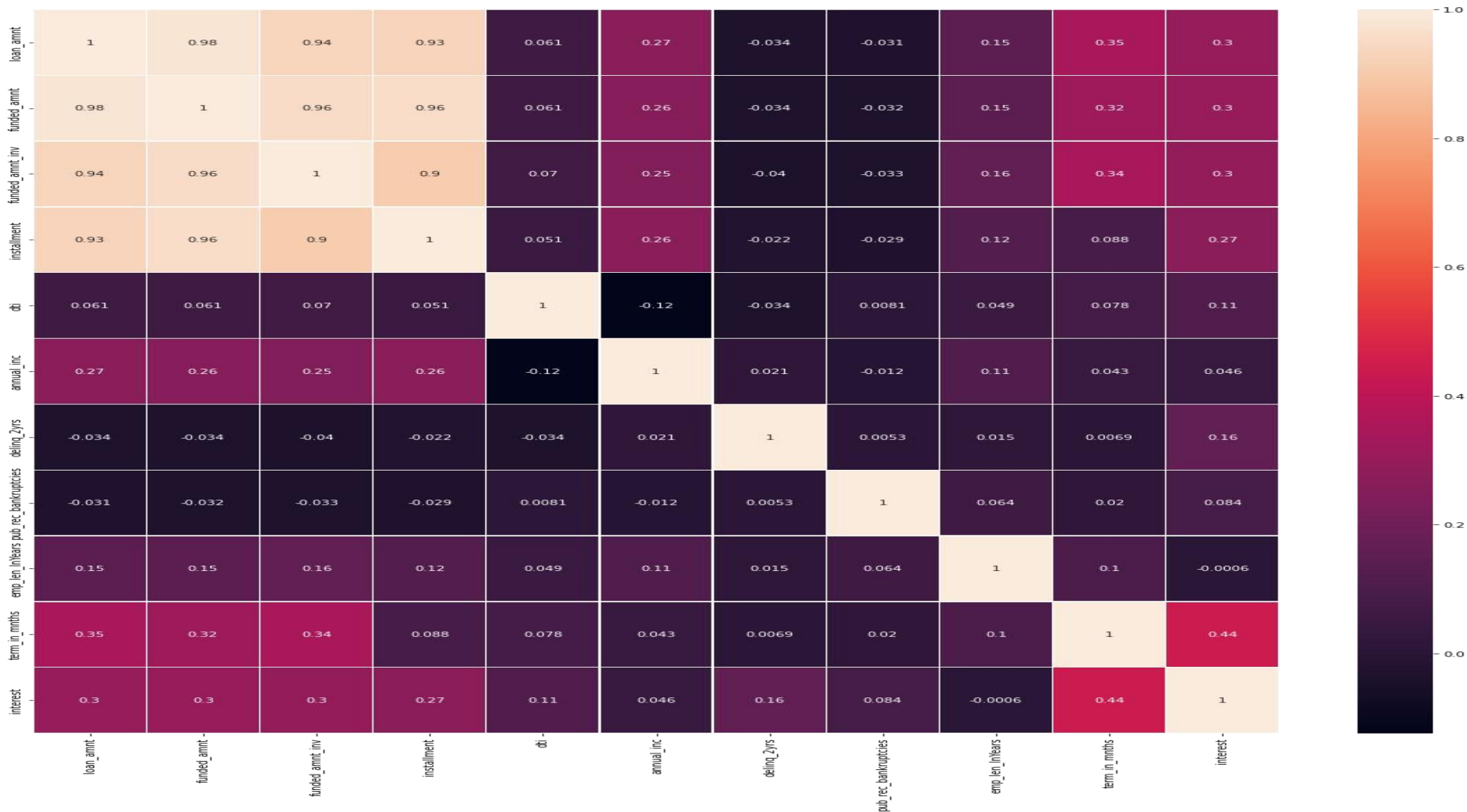
"int_rate" and "term" with "loan_status" as condition



"dti" and "term" with "loan_status" as condition



"annual_inc" and "term" with "loan_status" as condition

# Heatmap plot tells the correlation between each of the variables

# Conclusion:

- [int_rate, annual_inc,dti,term ] are the variables or driving factors which mainly tells about whether a borrower can go "Default" or "not".

- If "int_rate" is between 10% to 18%, mostly the borrower can go to "Default".

- For 36 months term, "int_rate" should be less than 13%.

- For 60 months term, "int_rate" should be less than 16%.

- and "dti" should be between 10 and 15.(ideally lesser the dti the better)

- Good "annual_inc" greater than 50000.

# Thank You