



Lending Club Case Study

SUBMISSION

Name: Vedika Pawale and Jagruti Kadam

Business Understanding :

There is one 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:

- 1) If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- 2) 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

Business Objectives :

The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Problem Solving Methodology :

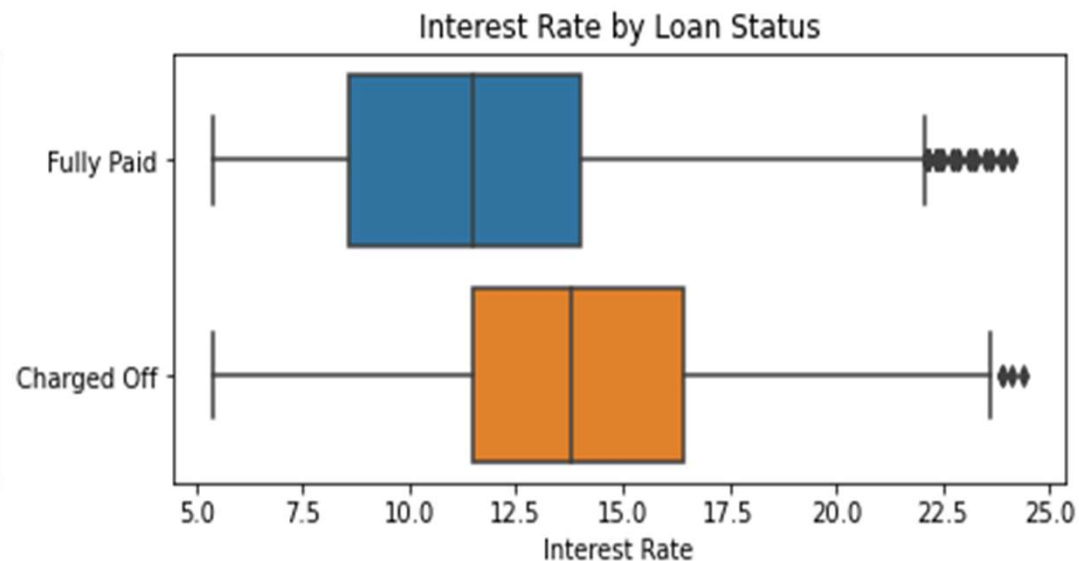
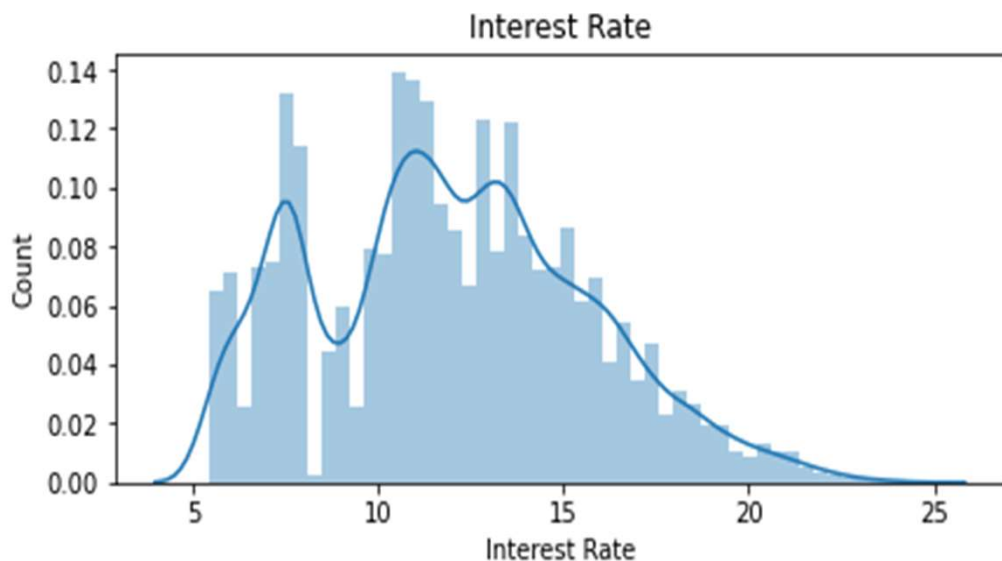
- ☐ Identify target variable as a loan status. Considered data related to only charge-off and fully paid and dropped data for current type.
- ☐ Grouped variables into 3 categories:
 1. Customer Demographic.
 2. Loan Characteristics.
 3. Customer Behaviour.

Dropped data related to customer behaviour as they are post loan approval variables.

- ☐ Data Cleaning : Handle null values for rows and columns.
- ☐ Derived new variables from existing ones(Extract month and year from datetime using pandas) like ex-issue_date, earliest cr line etc
- ☐ Data-type correction
- ☐ Default rate Vs variable analysis
- ☐ Plot variables spread and rate of charge-off against given variable.
- ☐ Bivariate analysis using heatmap and barplot.

Analysis : Interest rate

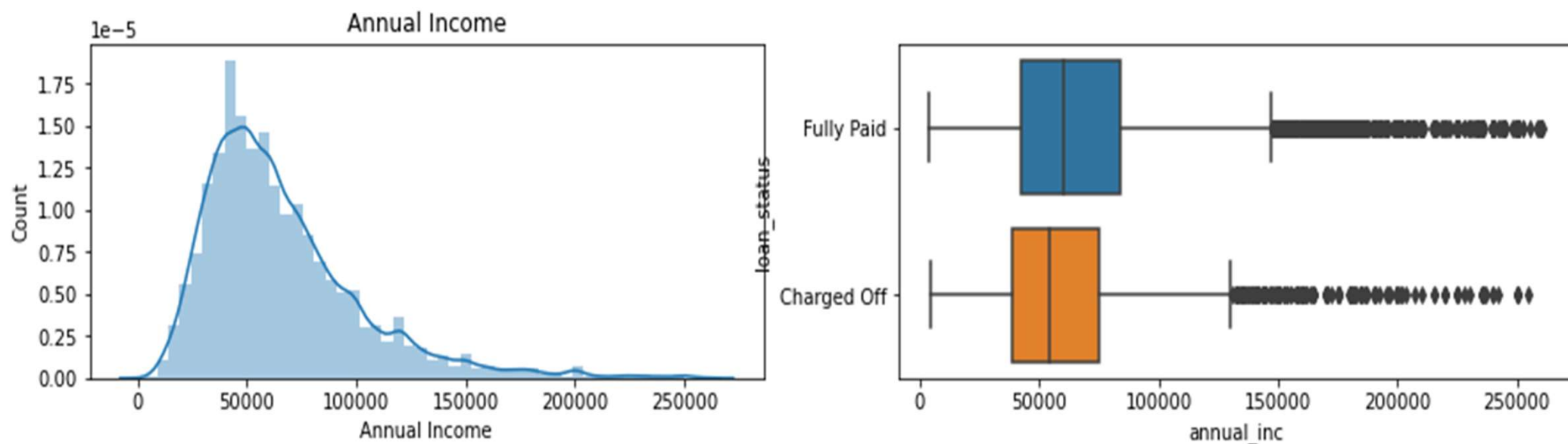
Data Dictionary : Interest Rate on the loan



Observation : Interest rate is directly proportional to Loan Charged Off. We observe that median and percentile values are higher for Charged Off.

Analysis : Annual Income

Data Dictionary : The self-reported annual income provided by the borrower during registration.

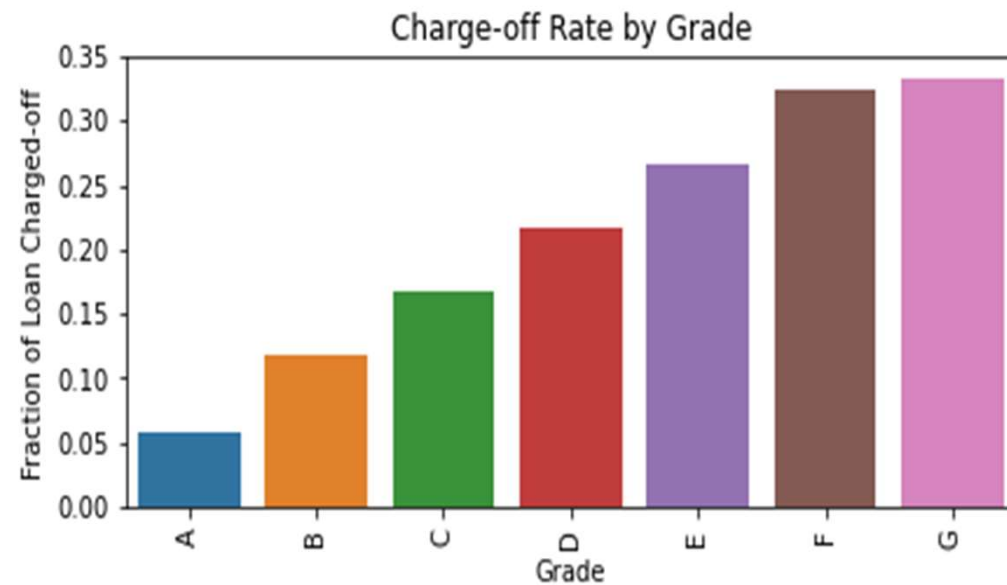
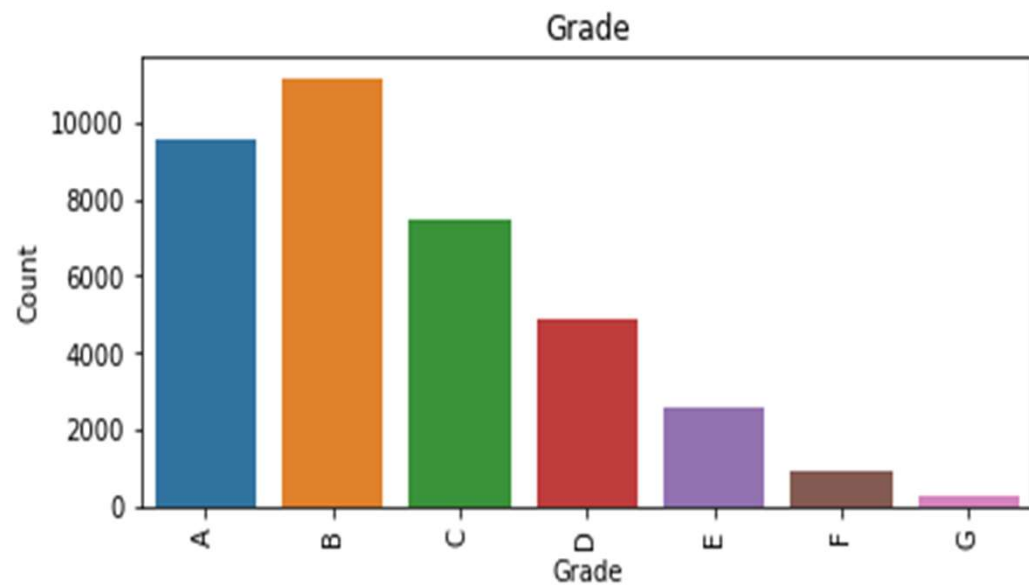


Observation : Annual Income is inversely proportional to Loan status. We observe that median and percentile values are higher for fully-paid compared charged off.



Analysis : Grade

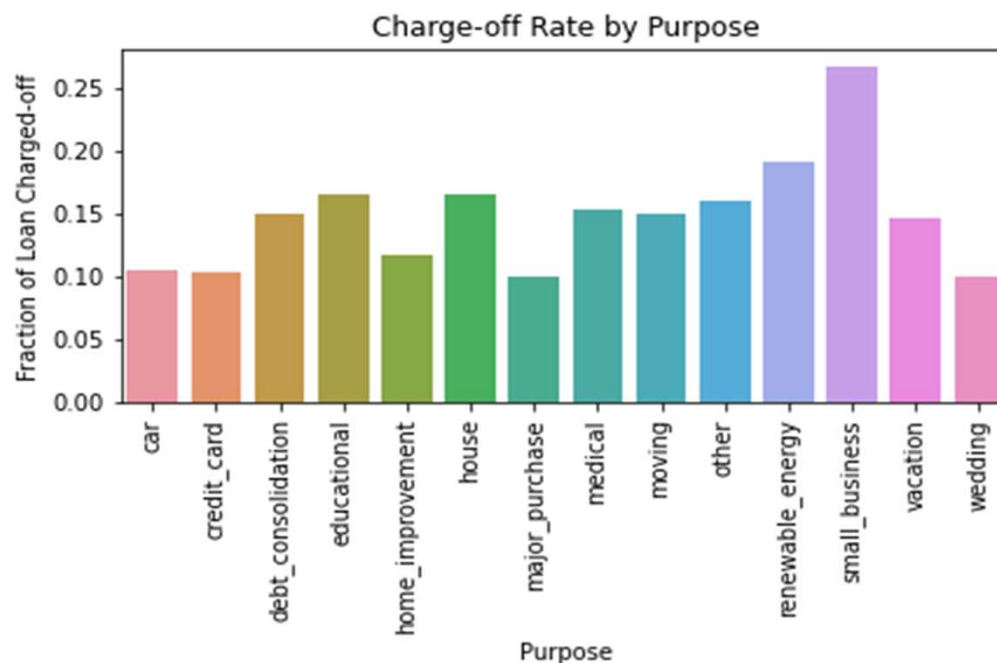
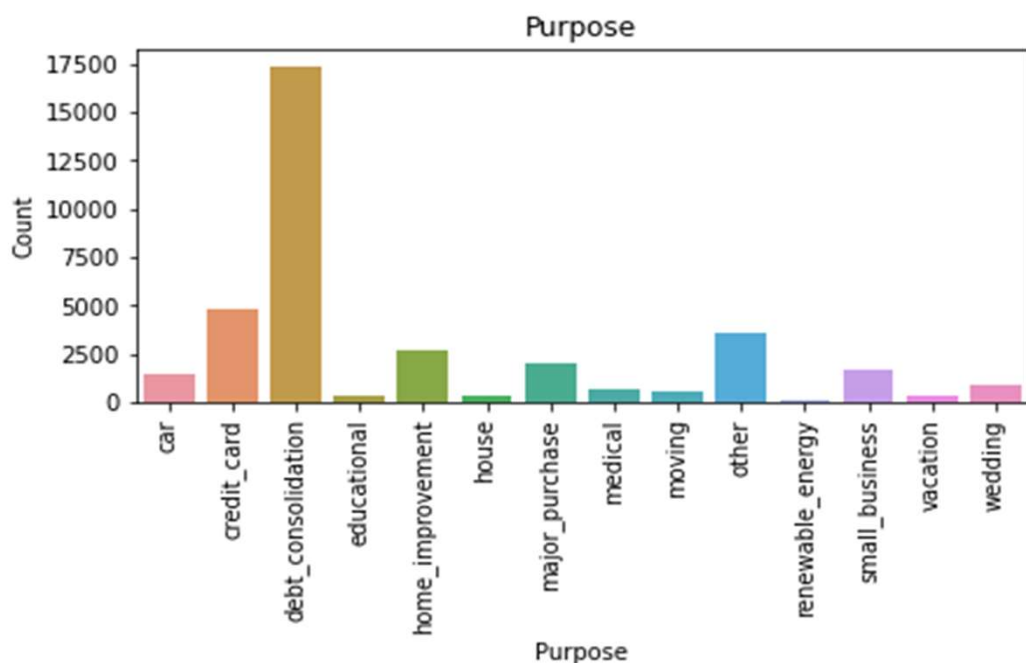
Data Dictionary : LC assigned loan grade



Observation : Grade is directly proportional to Loan Charged Off

Analysis : Purpose of loan

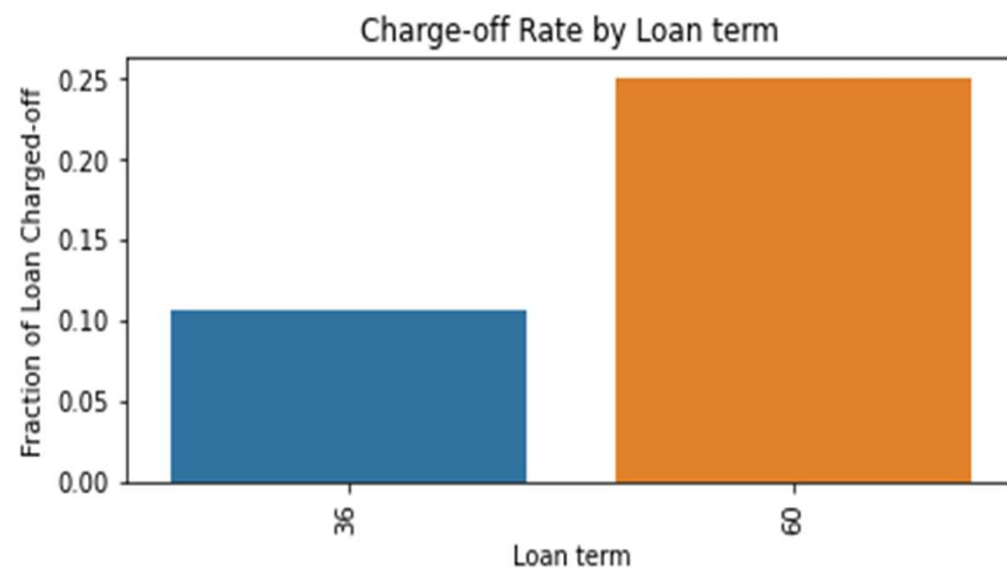
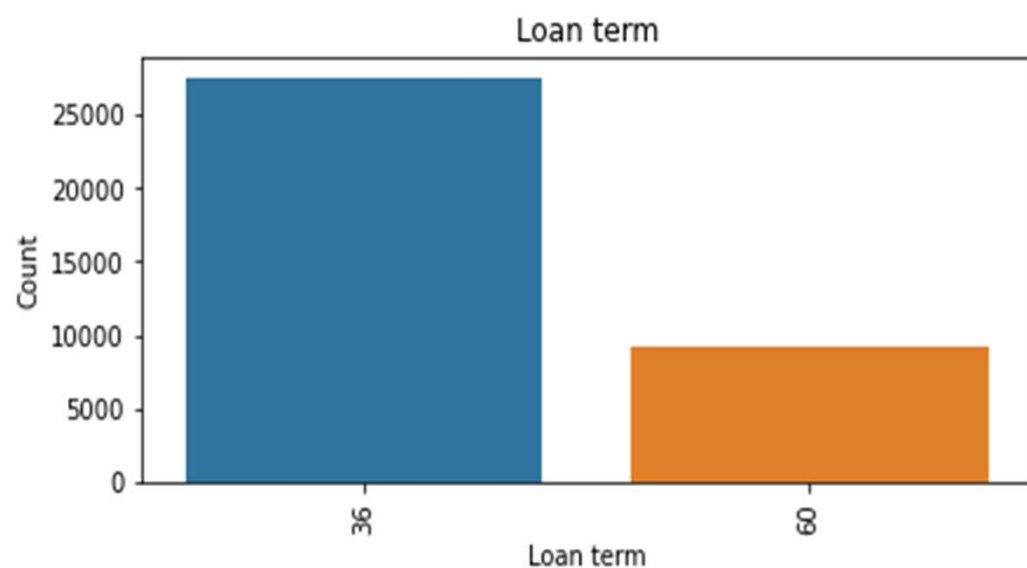
Data Dictionary : A category provided by the borrower for the loan request.



Observation : Charge-off rate varies with purpose of loan, applicants from "small business" category has the highest probability of Charged Off.

Analysis : Term

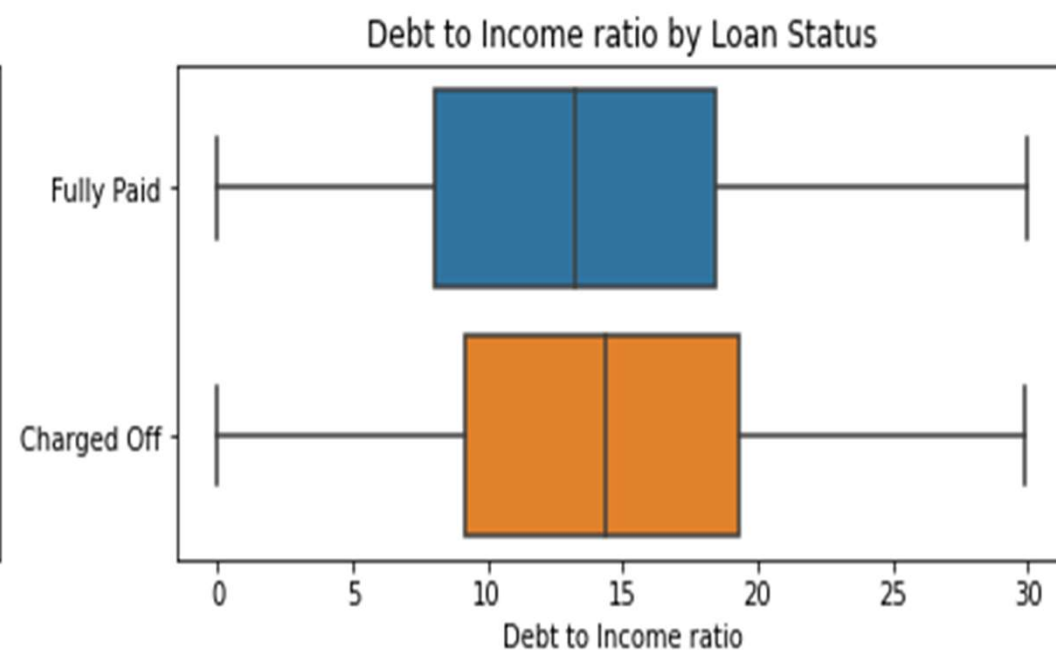
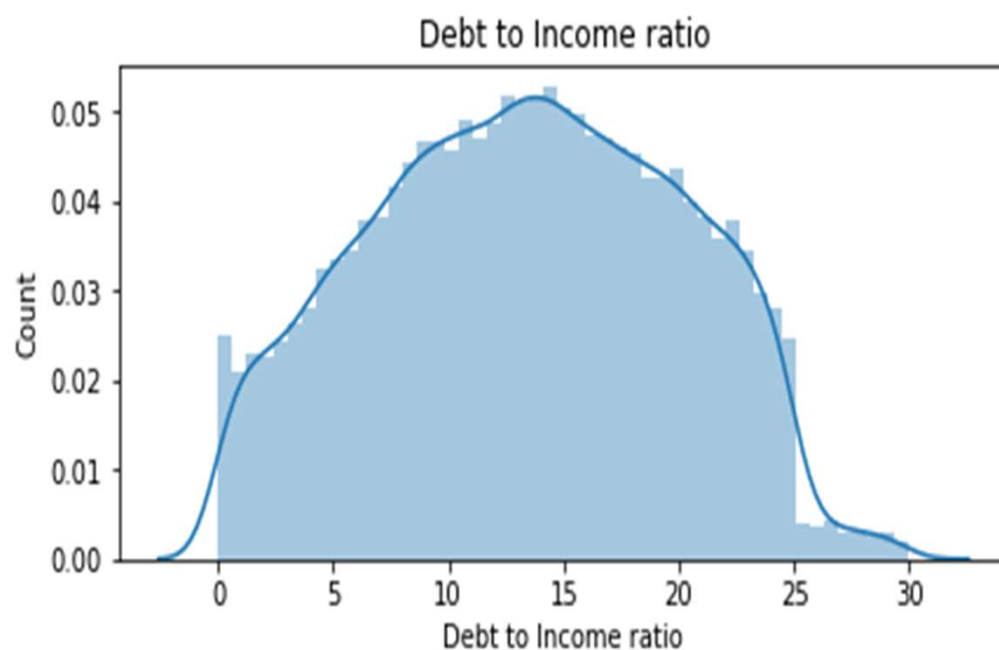
Data Dictionary : The number of payments on the loan. Values are in months and can be either 36 or 60.



Observation : Loan Term is directly proportional to Loan Charged Off .We observed that more applicants applied loan for 36 months term period compared for 60 Months in which fraction of loan charged off rate is higher in 60 months.

Analysis : Debt to Income Ratio

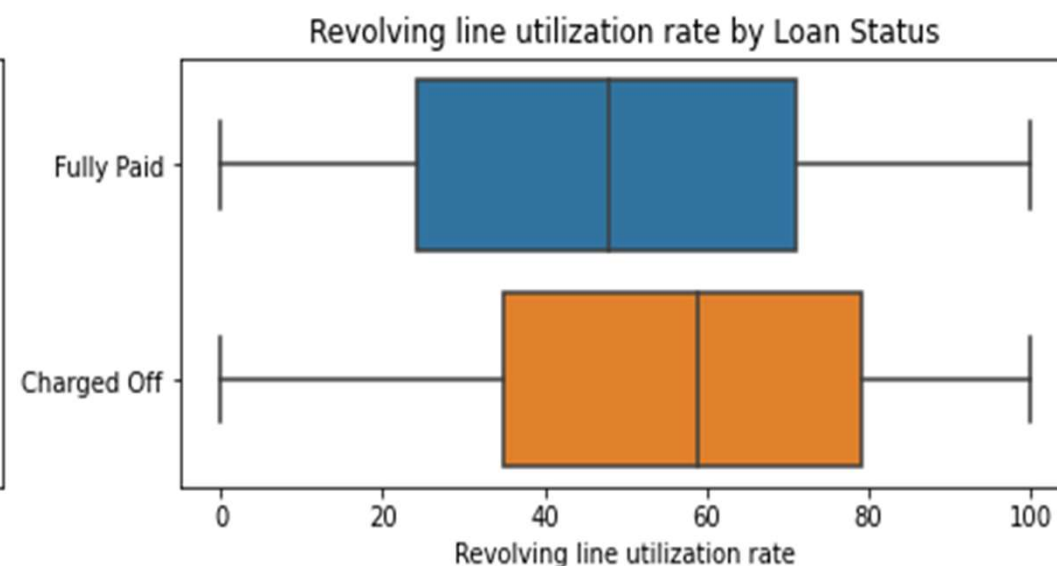
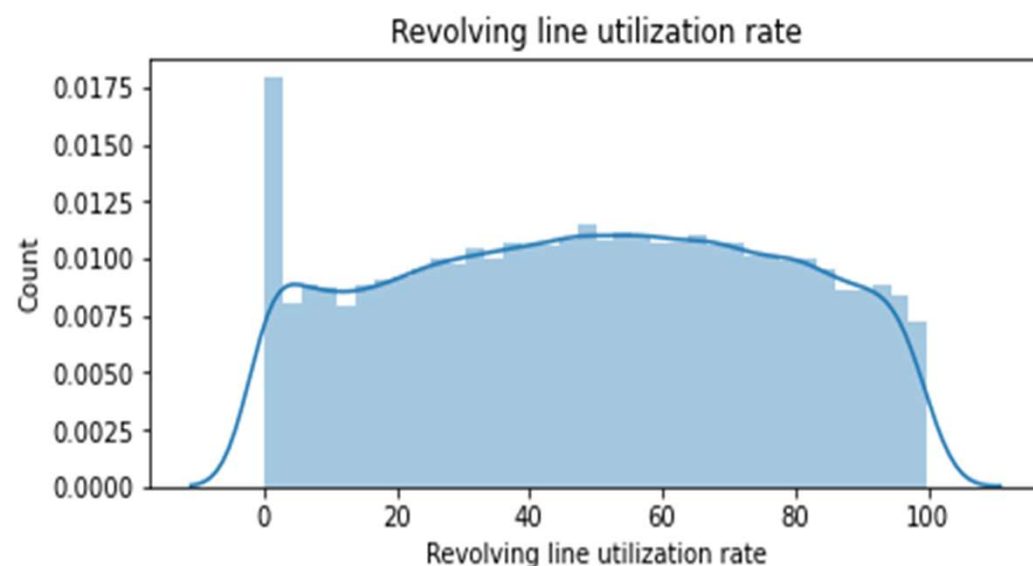
Data Dictionary : A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's



Observation : dti ratio is directly proportional to charged-off rate. We observe that median and percentile values are slightly higher for charged off compared to fully paid .

Analysis : Revolving line utilization rate

Data Dictionary : Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.



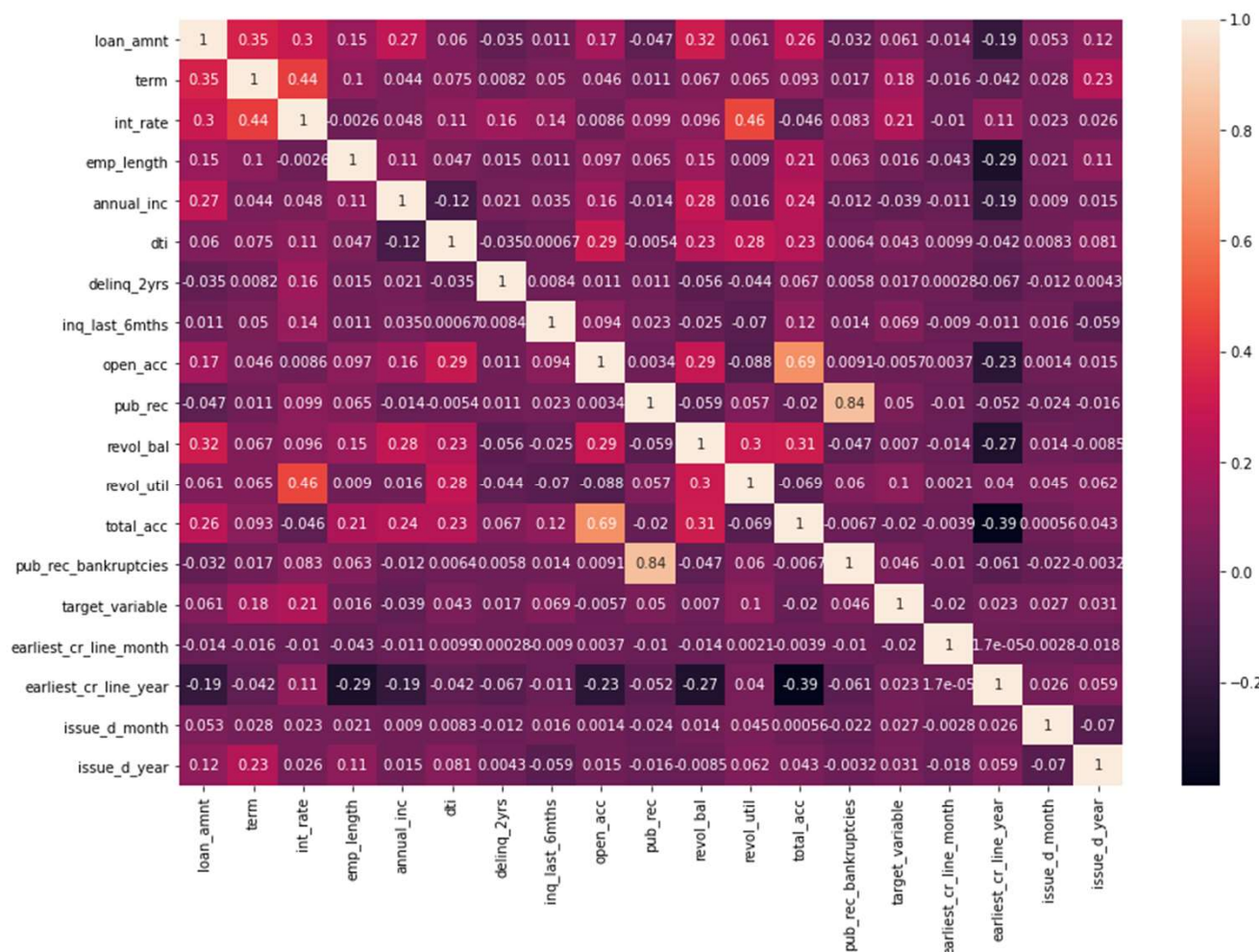
Observation : Revolving line utilization rate is directly proportional to Loan status. We observe that median and percentile values are higher for fully-paid compared charged off.

Analysis :Heatmap for loan dataset

Observation:

1.Target Variable(charge off rate) has positive correlation with following variables:-loan amount, term, interest rate, employment length, dti , delinquency records, inquiries in last 6 months, pub derogatory records, revolving balance ,credit utilization rate, public bankruptcy records

2.Target Variable(charged off rate) has negative correlation with following variables:-annual income, total accounts, open acc, earliest_cr_line_month



Conclusion :

From our observation we conclude that,

- Loan Status is a Target Variable.

Below is a list of Top 7 variables responsible to predict the Loan Status :

- Interest rate
- Annual Income
- Grade
- Purpose of loan
- Term
- Debt to Income Ratio
- Revolving line utilization rate