Lending Club case EDA

Manish Hemnani

11-05-2022

Lending Club Problem Statement

Business and Problem overview:

Lending Club is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

Borrowers can easily access lower interest rate loans through a fast online interface. lending loans to 'risky' applicants is the largest source of financial loss (called credit loss).

The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

Objective: 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

Analysis and Approach

- ❏ Business Understanding of Features
- ❏ Data Cleaning
- ❏ Data Manipulation
- ❏ Data Analysis
    - ❏ Univariate Analysis
    - ❏ Bivariate Analysis
    - ❏ Multivariate Analysis
- ❏ Recommendations

# Data Understanding and Features Considered

Features were analysed from business perspective and cleaning was performed

- Loan Status: Current status of loan ,there are 3 types of loan
    - Fully Paid: The applicant has fully paid the loan installments
    - Current: Applicant is in process of paying installments , and are not labelled as defaulted
    - Charged Off : Borrower has not paid installments from long time , hence has defaulted on the loan
- Loan Amount: Loan amount

Data Understanding and Features Considered

- ❏ Features taken into consideration for data analysis
  - ❏ Loan Amount,
  - ❏ 'funded_amnt_inv',
  - ❏ 'Term' - Loan Term,
  - ❏ 'Int_rate', - Interest Rate
  - ❏ 'Installment', - Loan Installment
  - ❏ 'Grade', - Loan Grade assigned by Lending Club
  - ❏ 'Sub_grade', - Sub Grade assigned by Lending Club
  - ❏ 'Emp_length', - Employee Length in Years
  - ❏ 'Home_ownership', - Borrower Ownership of Home currently living in
  - ❏ 'Annual_inc', - Annual Income of Borrower
  - ❏ 'Verification_status', - Verification status of borrower
  - ❏ 'Issue_d', - Month in which loan was issues
  - ❏ 'loan_status',- Status of loan - Fully Paid, Current , Charged Off
  - ❏ 'Purpose', - Purpose of Loan taken
  - ❏ 'Addr_state', - The state provided by the borrower in the loan application
  - ❏ 'Dti', - debt and installment ratio
  - ❏ 'Earliest_cr_line', - borrower's earliest reported credit line was opened:
  - ❏ 'Pub_rec', : Number of derogatory public records
  - ❏ 'Revol_util', : Revolving line utilization rate
  - ❏ 'Total_acc',: The total number of credit lines currently in the borrower's credit file
  - ❏ 'Pub_rec_bankrupties': Number of public record bankrupties

Correlation between Features
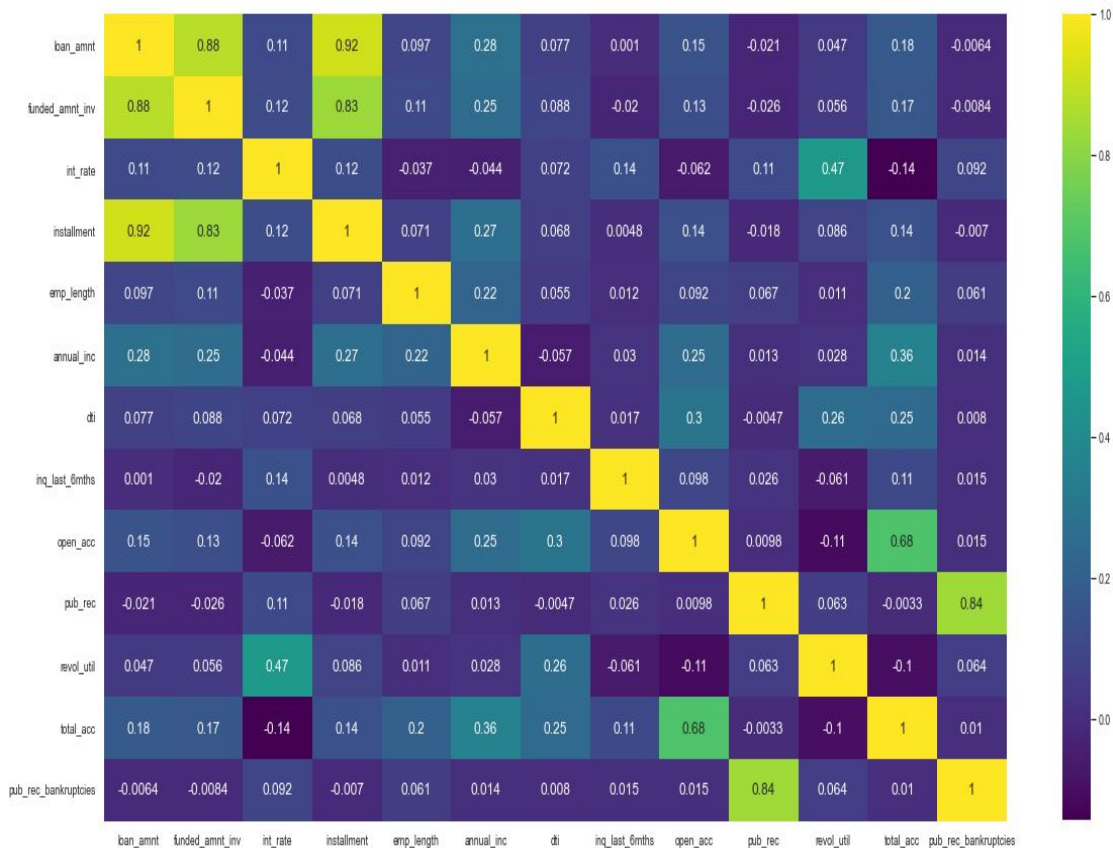
Observation:
- On high level loan_amount , pub_rec_bankruptcies look inversely correlated
- On high level dti and annual_inc looks inversely correlated
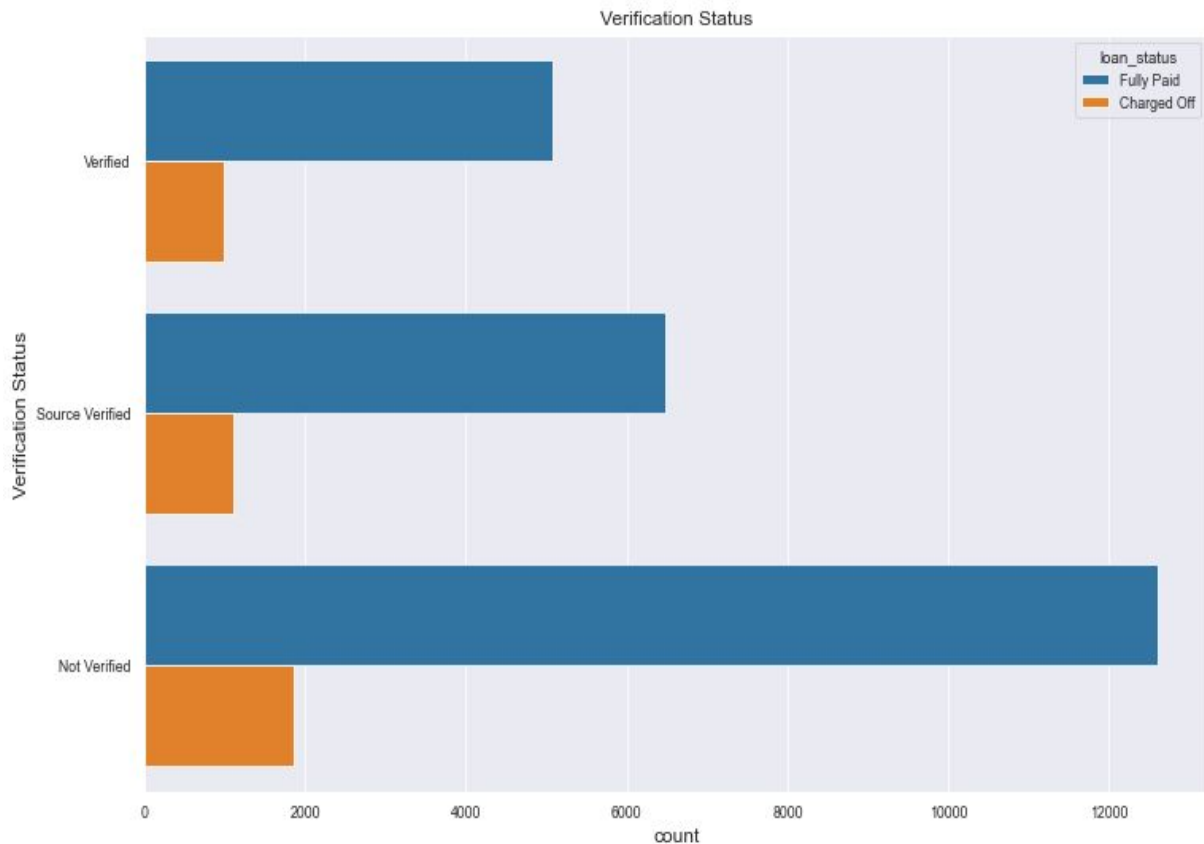- Loan_amount, Funded amount, installment are correlated to each other

These are higher level observation we can go deep down into them to observe in detail

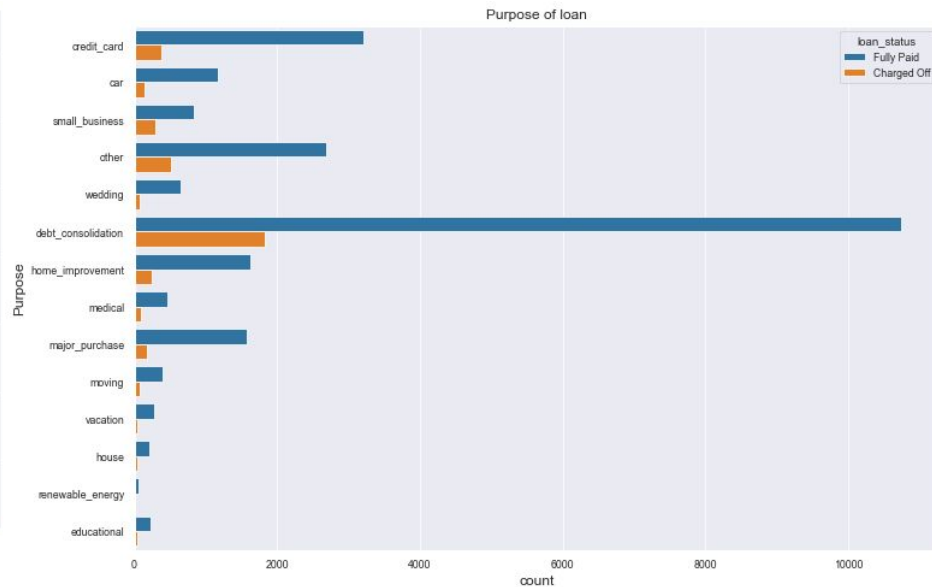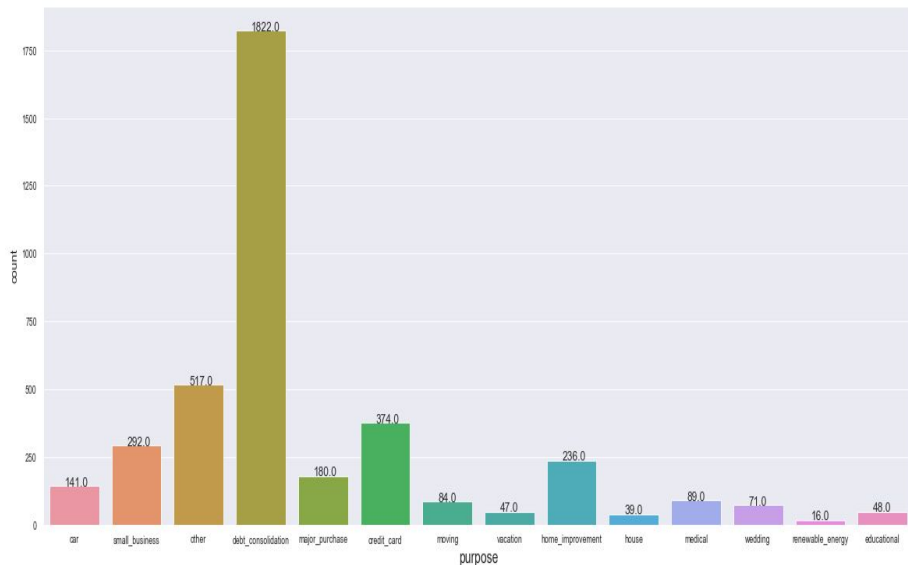Out of total borrowers approx 14% are the defaulters after removing the outliers
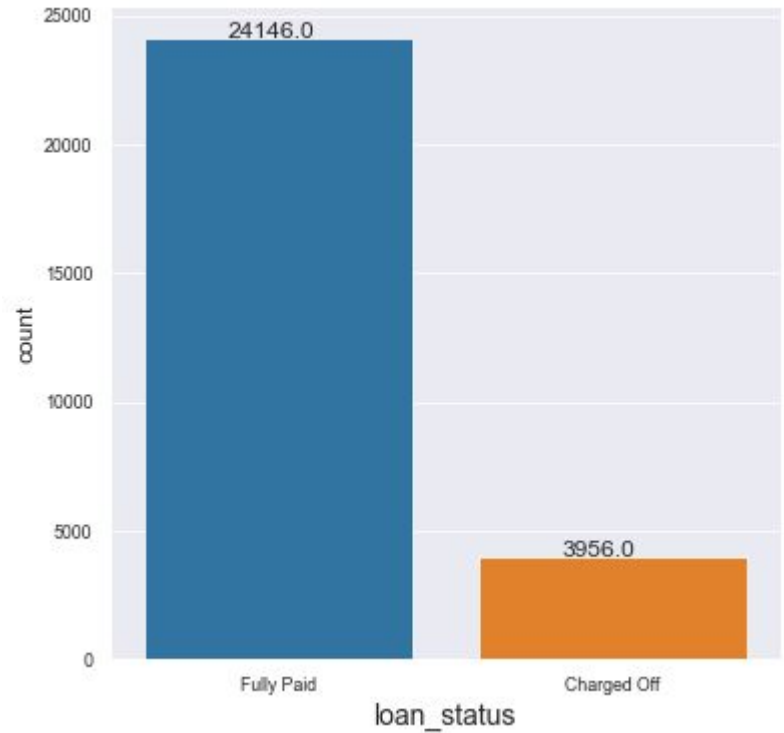
Univariate: Verification status

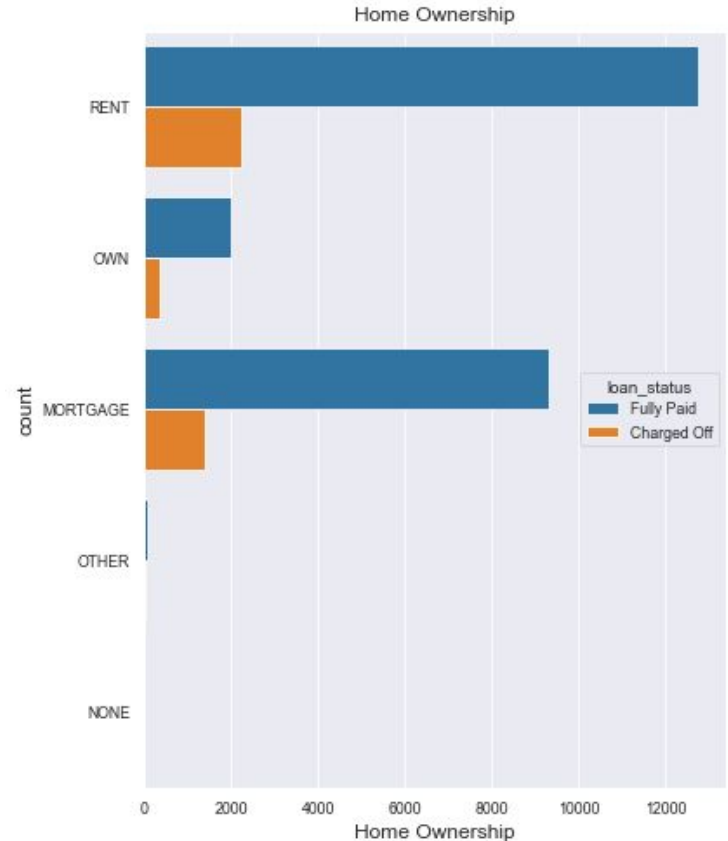| verification_status	loan_status	Charged off %
Verified	Charged Off	16.33
Source Verified	Charged Off	14.51
Not Verified	Charged Off	12.9

There are **13% Non-verified** loans getting **charged off**

Loan taken for purpose of '**small business**' ,'**renewable energy', 'education'** '**debt_consolidation', 'other'** and '**credit card**' are the once which **charged off** the most
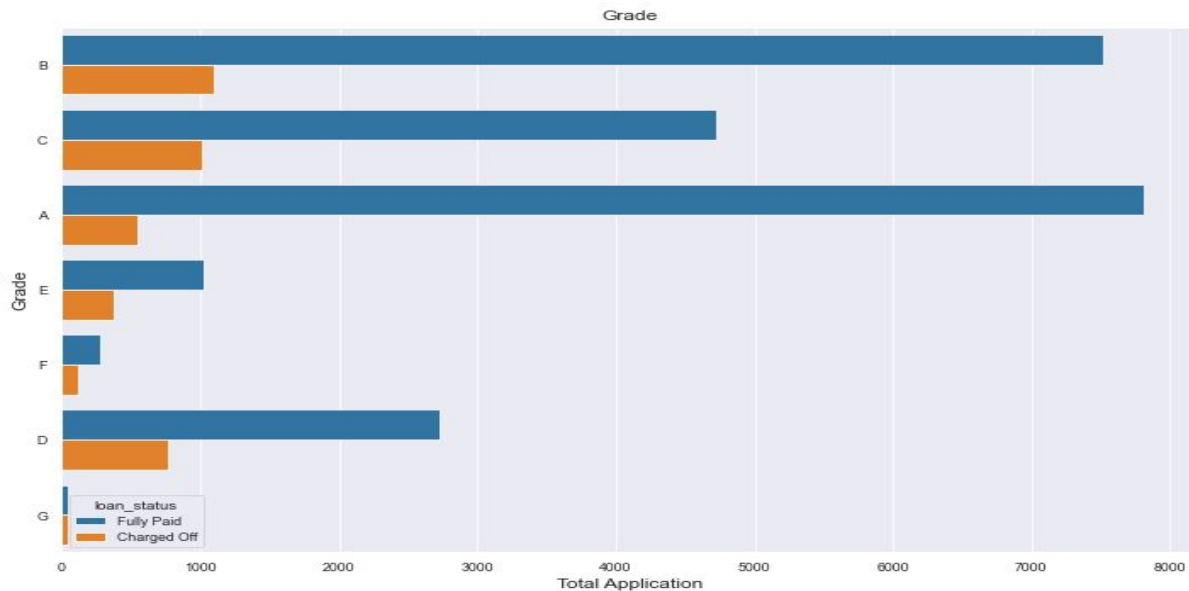
Univariate Analysis : Ownership of House

Large number of people who are living in **rented** property have applied for loan
Or they have **mortgage** on the property .

About 14% of Loan given to Rented home ownership can get charged off

Univariate analysis: Grade

Conclusion: B/C/D have large number of charged off accounts, among which G/F/E have the highest percentage of charged off

Grade	loan_status	Charged off %
G	Charged Off	46.5
F	Charged Off	29.5
E	Charged Off	26.6
D	Charged Off	21.9
C	Charged Off	17.6
B	Charged Off	12.7
A	Charged Off	6.5

Univariate analysis: Loan Issue Month
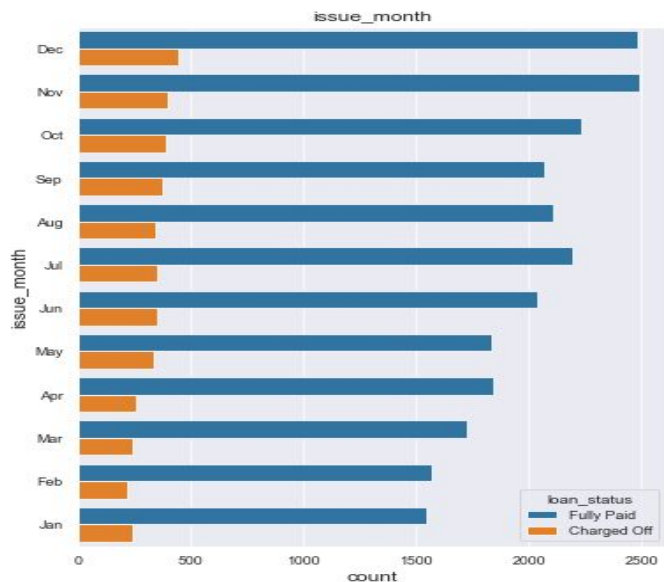
issue_month	loan_status	Charged off %
Sep	Charged Off	15.434871
May	Charged Off	15.345622
Dec	Charged Off	15.145299
Oct	Charged Off	14.998097
Jun	Charged Off	14.834935
Aug	Charged Off	13.965798
Nov	Charged Off	13.836043
Jul	Charged Off	13.802595
Jan	Charged Off	13.598209
Mar	Charged Off	12.322515
Apr	Charged Off	12.214829
Feb	Charged Off	12.10084

Conclusion: Loans issued in Sep / May / Dec have highest percentage of Charged off.
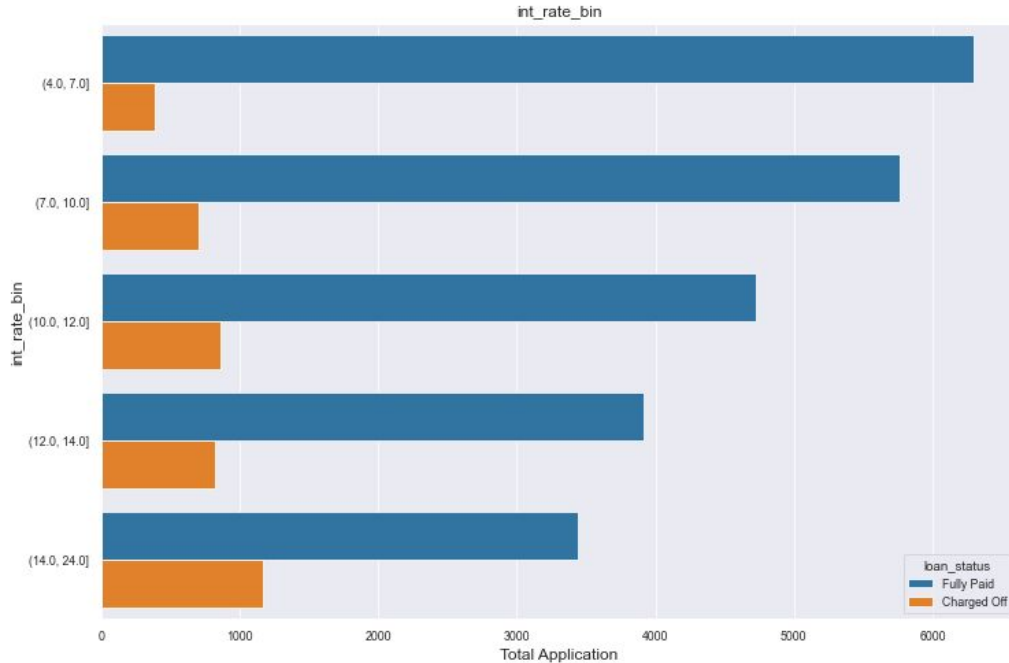Maximum charged off is in December and also has highest number of loans applied

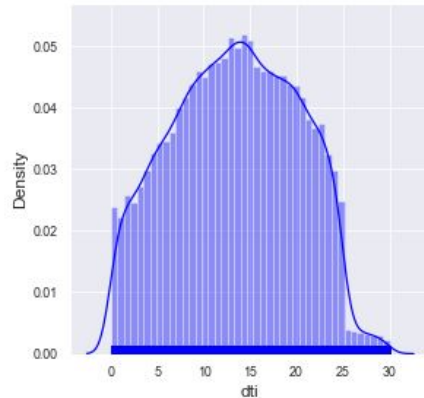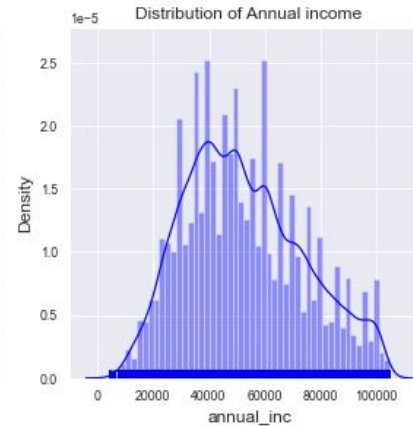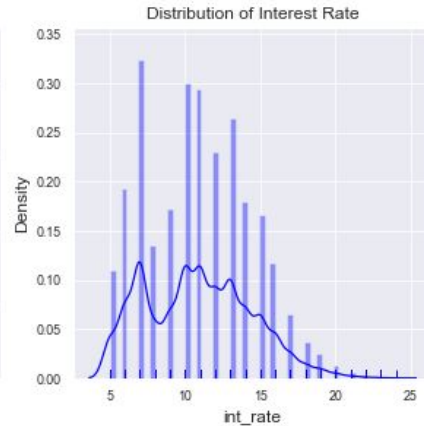# Segmented Univariate Analysis : Debt to Income Ratio
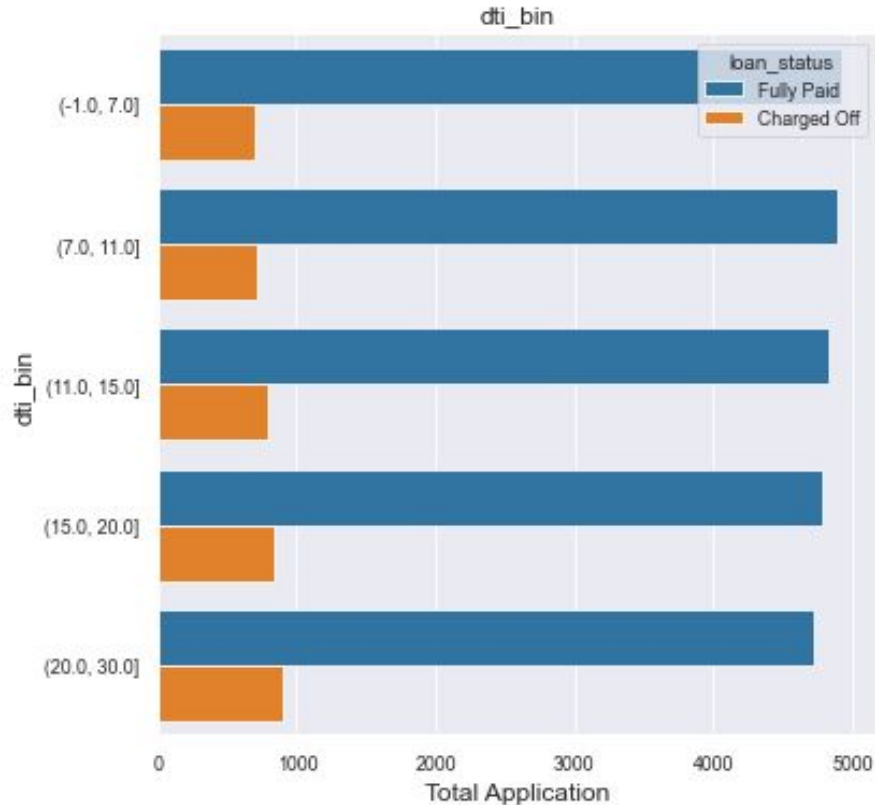
Observation: Higher the DTI value higher charge off is possible for the borrower

Segmented Univariate Analysis : Interest Rate

Observation: Higher the rate of interest higher the chances of loan getting Charged off

# Distribution Plots: Analysing features

- ❏ Loan Amount and int_rate have similar distribution can be analysed in correlation
- ❏ Dti can be analysed individually

Bivariate Analysis

Below Features can be Analysed for Bivariate Analysis
- Int_rate vs Term
- Interest rate vs Loan Amount
- Loan amount vs Grade
- Dti vs annual Income
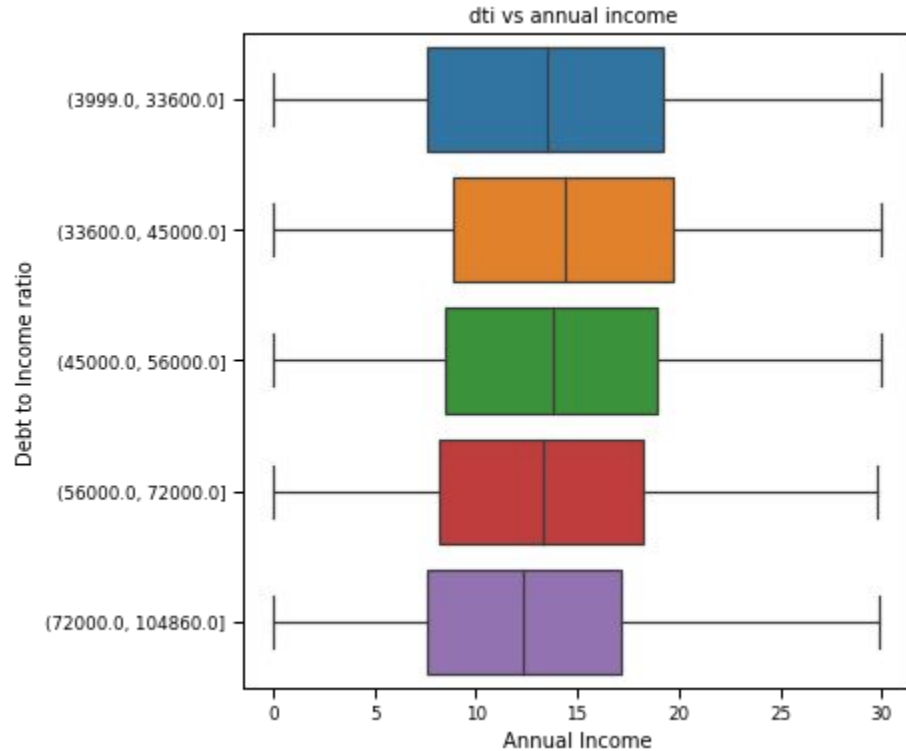- Loan Amount Vs Grade
- Grade vs Interest rate

Bivariate Analysis : Interest Rate vs Term and interest rate vs Loan Amount

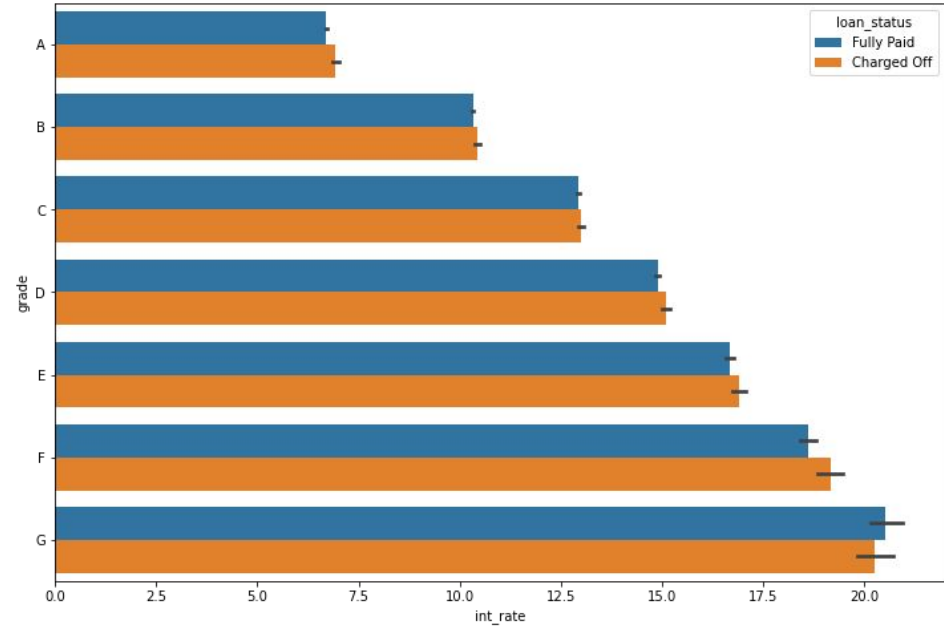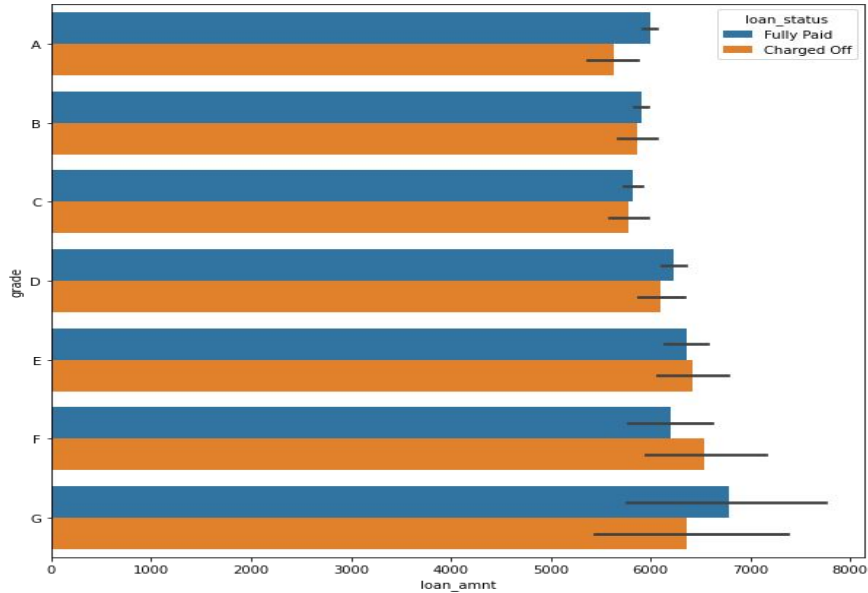Term vs Interest Rate

Loan Amount vs Interest Rate
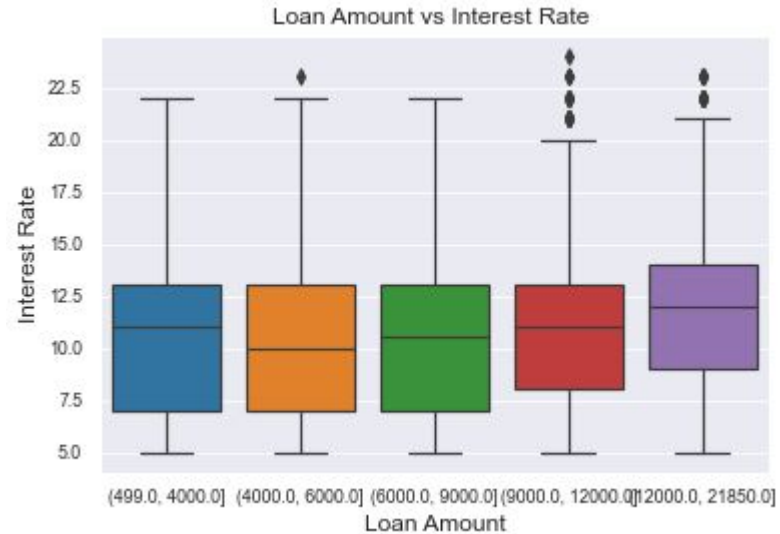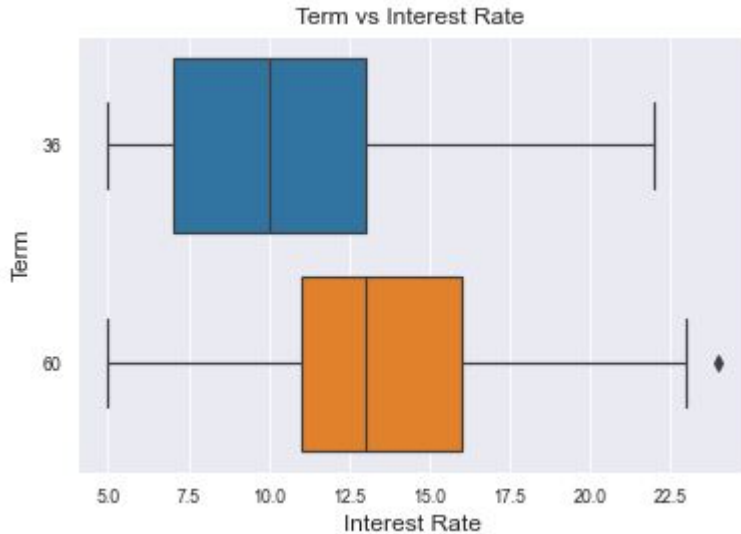
Observation:
❏ Interest amount is higher for 60 month loan term
❏ Interest rate increase with the amount of loan

Bivariate analysis: DTI vs Annual Income

DTI is good for borrower with higher annual income hence could be less chances of loan getting charged off

Observation: Higher the loan amount and interest rate for group E/F/G higher the chances of loan Charge off

Conclusion

Conclusion

- Higher the rate of interest and Longer Team Increase are likely to be charged off.
- Loan Issued in month of December/November/October are more likely to get charged Off.
- Organisation should avoid giving loans to unverified borrowers and change the verification process.
- Applicants with E/F/G Grade if given loan with higher rate of interest are likely to be charged off
- Observation: Higher the loan amount and interest rate for group E/F/G higher the chances of loan Charge off
- higher Interest Rate and Longer Team Increase the Probability of charged off.
- DTI is good for borrower with higher annual income hence could be less chances of loan getting charged off