Lending Club Case Study

March , 2022

- Alok Kumar
- Pradip Tiwari

Executive PG Program In Machine Learning and AI

Introduction

Lending Club is a 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.**

When a person applies for a loan, there are two types of decisions that could be taken by the company:

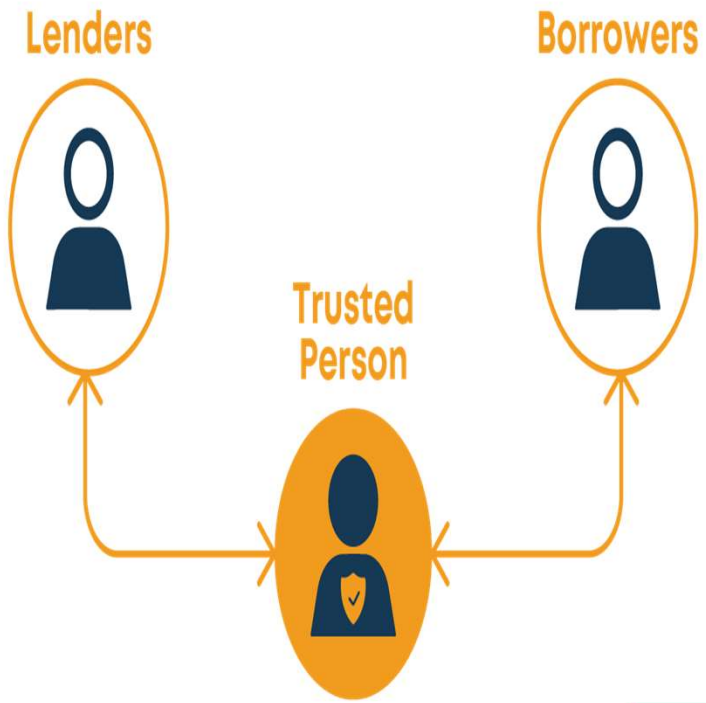
1. **Loan accepted:** If the company approves the loan, there are 3 possible scenarios described below:

    - **Fully paid:** Applicant has fully paid the loan (the principal and the interest rate).

    - **Current**: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted„ but certainty is not clear.

    - **Charged-off**: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

2. **Loan rejected:** The company had rejected the loan (because the candidate does not meet their requirements).

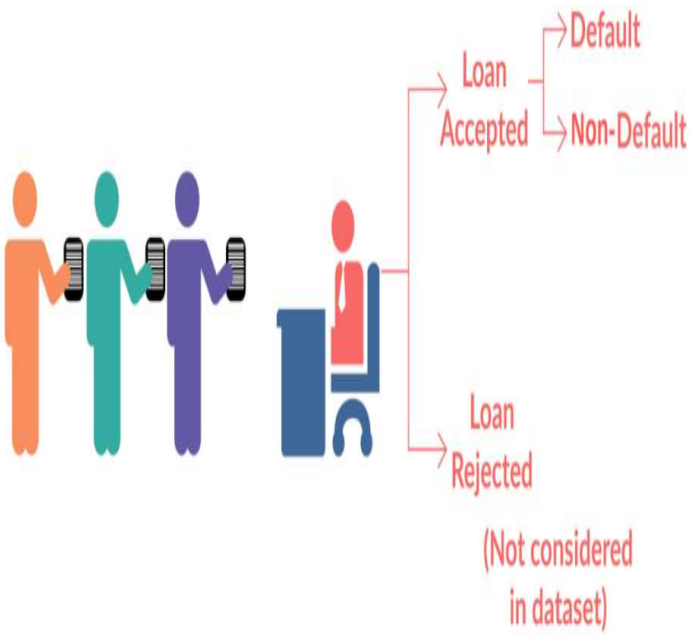


Case Study Objective

- Lending loans to Risky applicants is the largest source of financial loss in any Financial institution and is “called Credit Loss”. The credit loss is the amount of money lost by the lender when the borrower refuses or incapable to pay the money owed.

- Objective is to identify the risky loan applicants at the time of loan application so that such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA and different libraries of Python is the main aim of this case study.

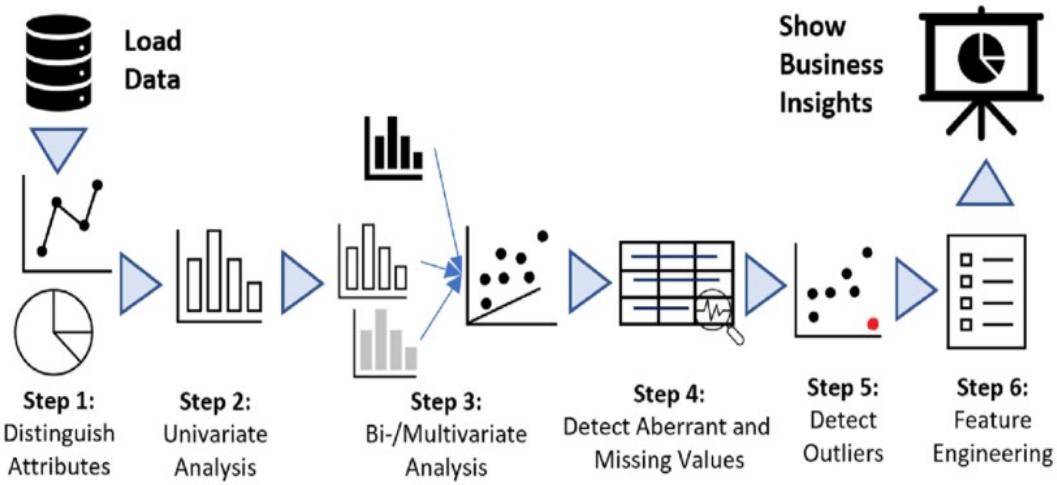
- In other words, to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its analysis and risk assessment. And thus minimize the risk of losing money while lending to customers.



Data Analysis (EDA methodology)

International Institute of Information Technology Bangalore

- The entire EDA process is based on the simple steps which are to be taken to reach the conclusion.
- These steps define and signifies the key components to analyses the risk profile.
- The necessary steps are Data sourcing , Data cleaning , Univariate analysis ,Bivariate analysis Derived metrics



upGrad

Data Assumption

- In this case we have been provided with the Private Data of Lending Club for all loans issued through the time period 2007 to 2011.
- We also have a data dictionary which describes the meaning of these variables.
- The loan request if graded by LC and then listed for loan. Now it is up to the investor who decides which loan listing to invest in looking at the loan detail and borrowers information.
- Only those variables be chosen for analysis that will be available to the investor at the time of deciding whether to invest in a loan request or not.

Data Cleaning

The Data Cleaning process concluded are as follows:

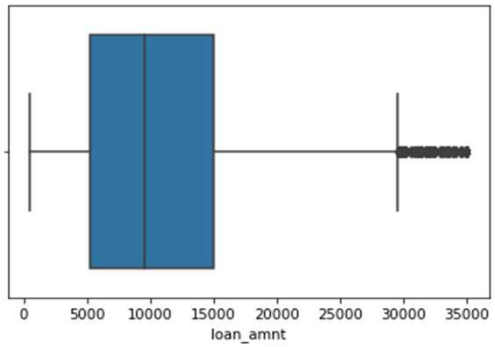
- *Fixing Rows & Columns:* In this process we have removed the columns having null values, single valued and irrelevant columns which are not affecting the Loan defaulting criteria.
- *Fixing Missing Values:* *Two* of the columns had some missing values which were dropped out of analysis.
- *Standardising Values:* We standardise some of the columns through Lambda function.
- *Fixing Invalid Values:* We fixed invalid values in columns.
- *Filtered Data:* Finally we achieved the filtered data after entire process of Data Cleaning.

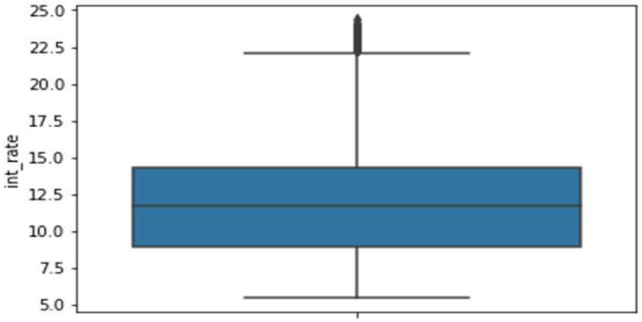
Note: Data cleaning is a very vital and most basic process of Exploratory Data Analysis and any fault in cleaning to disastrous results.

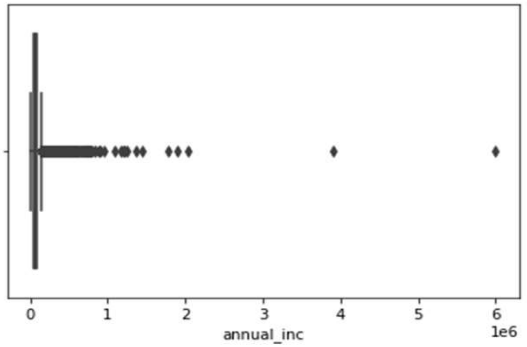
Statistics using Univariate Analysis

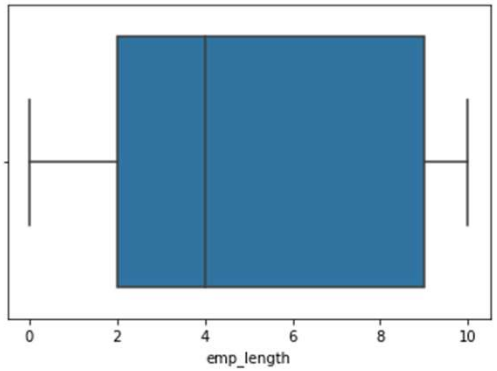
International Institute of Information Technology Bangalore

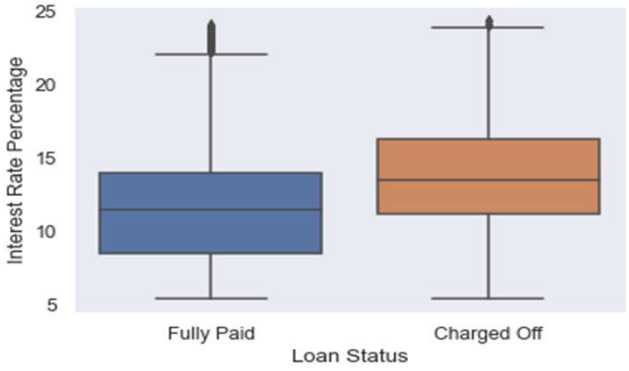
- Statistics of annual_inc shows presence of outliers which was removed in next step.
- Statistics of emp_length shows average of emp_length is 4 years, between 50%-75% percentile emp_length is 5-10 years.
- Statistics of int_rate shows average int. is 12%, after 75% percentile interest rate increased to 25% from 15%.
- Interest rate and Loan Status describe as less intrest rates are fully paid compared to high interest

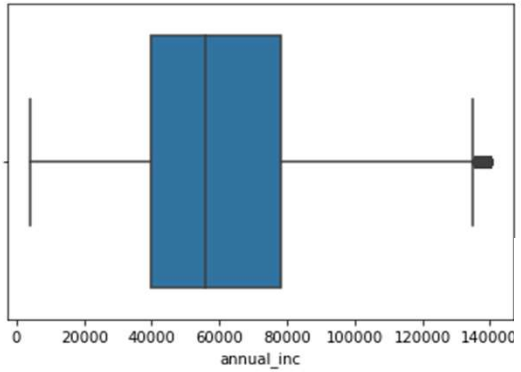










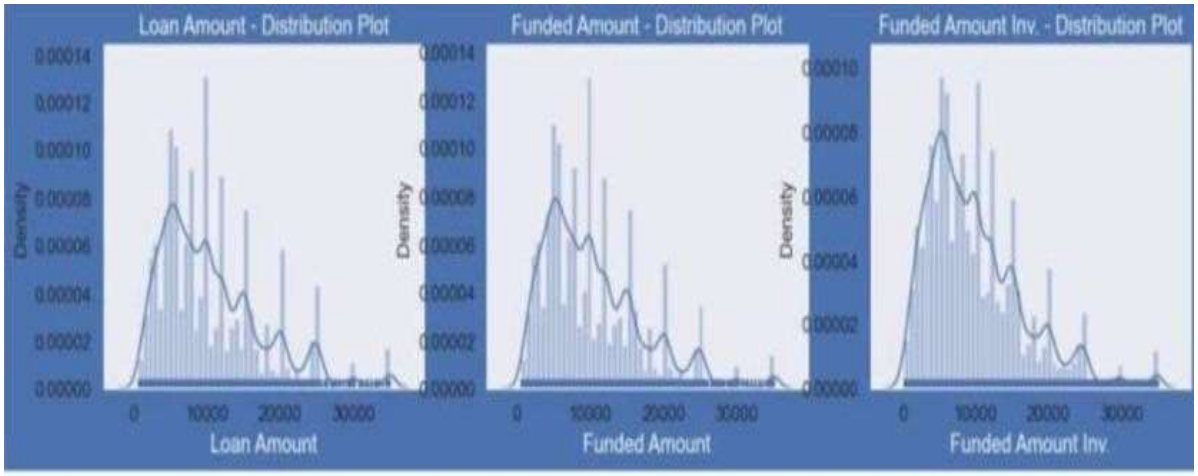


upGrad

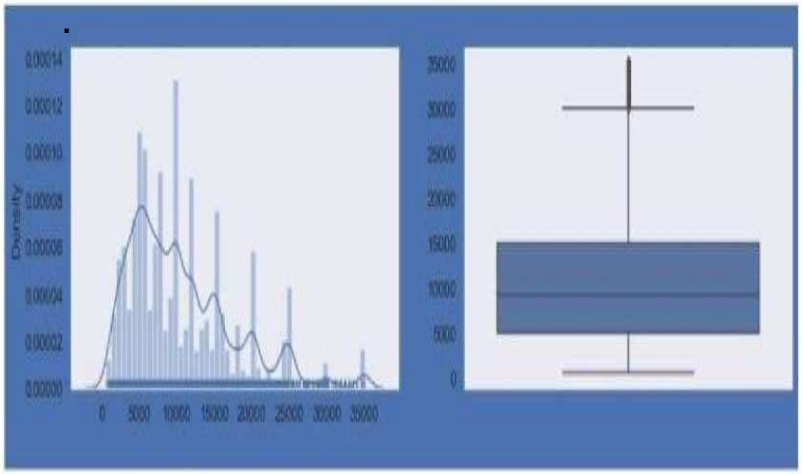
Univariate Analysis

International Institute of Information Technology Bangalore

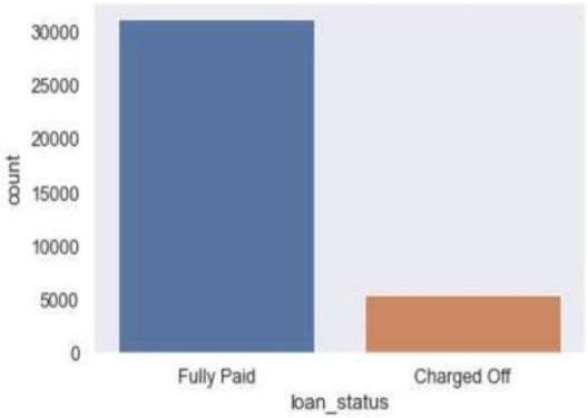
Distribution of Three Loan Amount Fields using Distribution Plot.



Distribution and Box Plot



Loan Status using Count Plot



Observation:

- *Plot for 3 variables looks very much similar.*
- *Majority of Borrower's Annual Income is in the range of 40000-80000*
- *Most of the Loan Amounts are in the range 5000-15000*

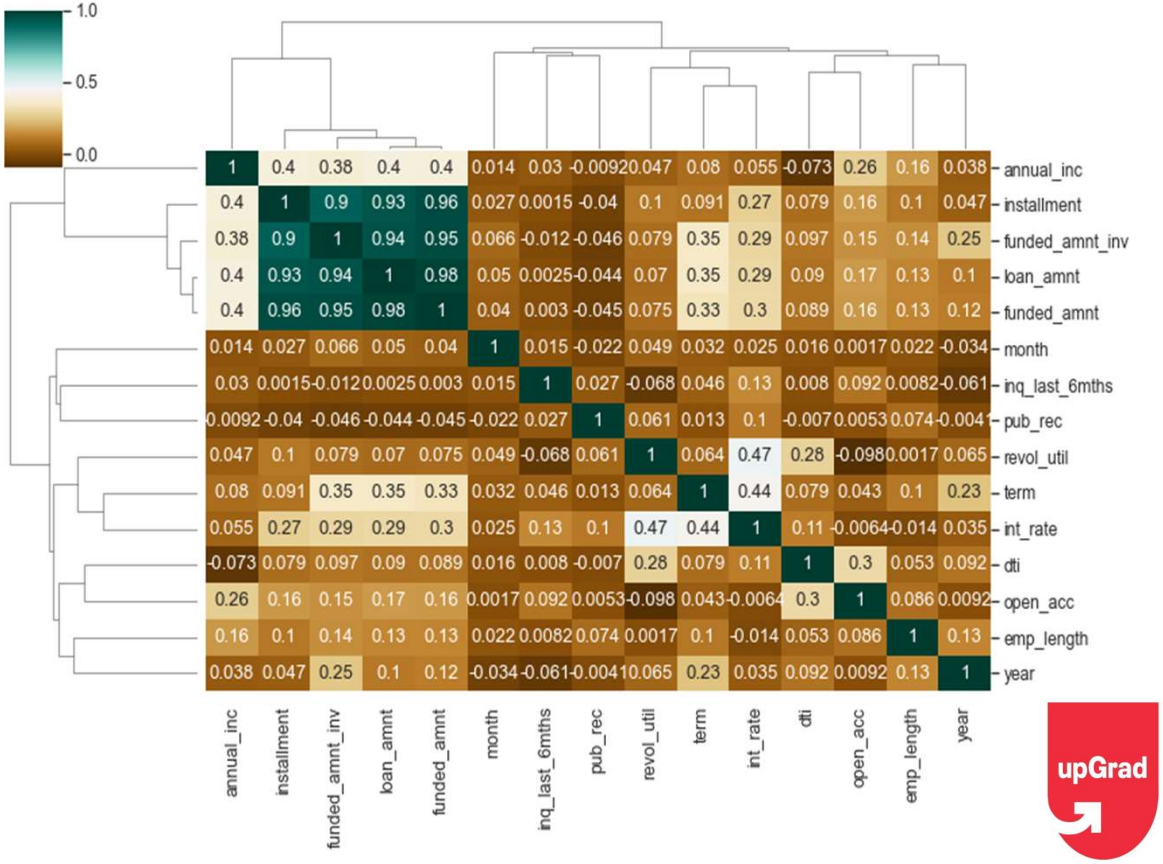
upGrad

Statistics using Bivariate Analysis

International Institute of Information Technology Bangalore

(Bivariate Analysis –Correlation Matrix –Quantitative Variables)

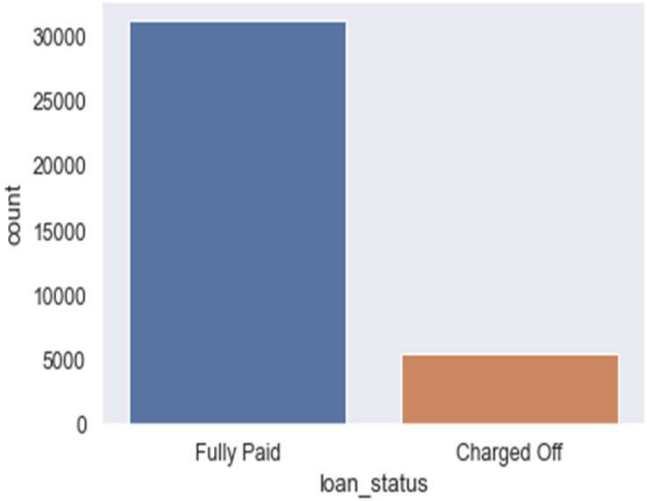
HEATMAP WITH DENDOGRAM (CLUSTERMAPT TO SHOW CLOSENESS AMONG NUMERICAL VARIABLES

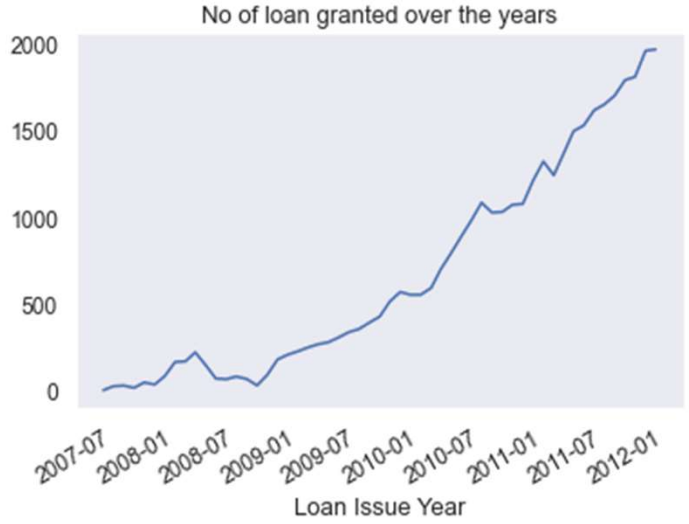


Observations:

- Loan Amount, Investor Amount and Funding Amount is strongly correlated.
- Annual Income with DTI is negatively correlated.
- Annual Income and Employment Years is positively correlated.

Analysis of Loan Status





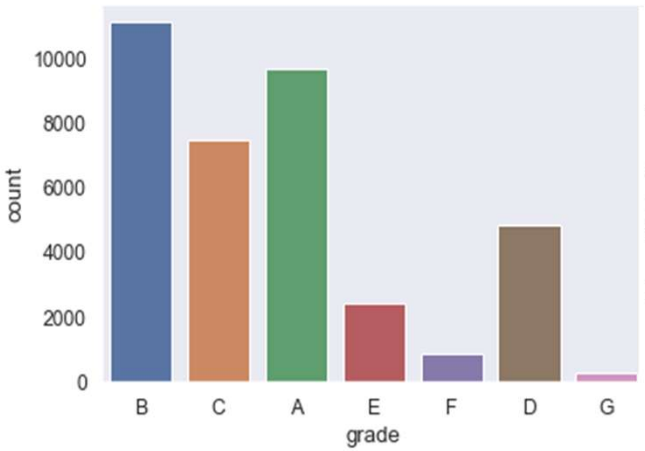
Observation:

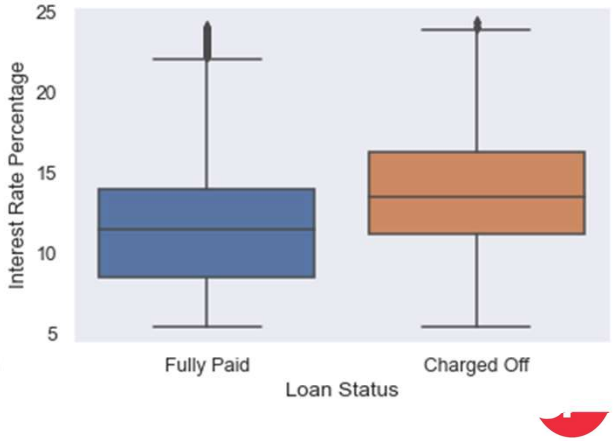
- Out of Total, 14% loans were defaulted.
- 78% Loans were fully paid and 8% are Current Loans.
- Loan Applications were increasing exponentially during the period.

Analysis of Loan Status

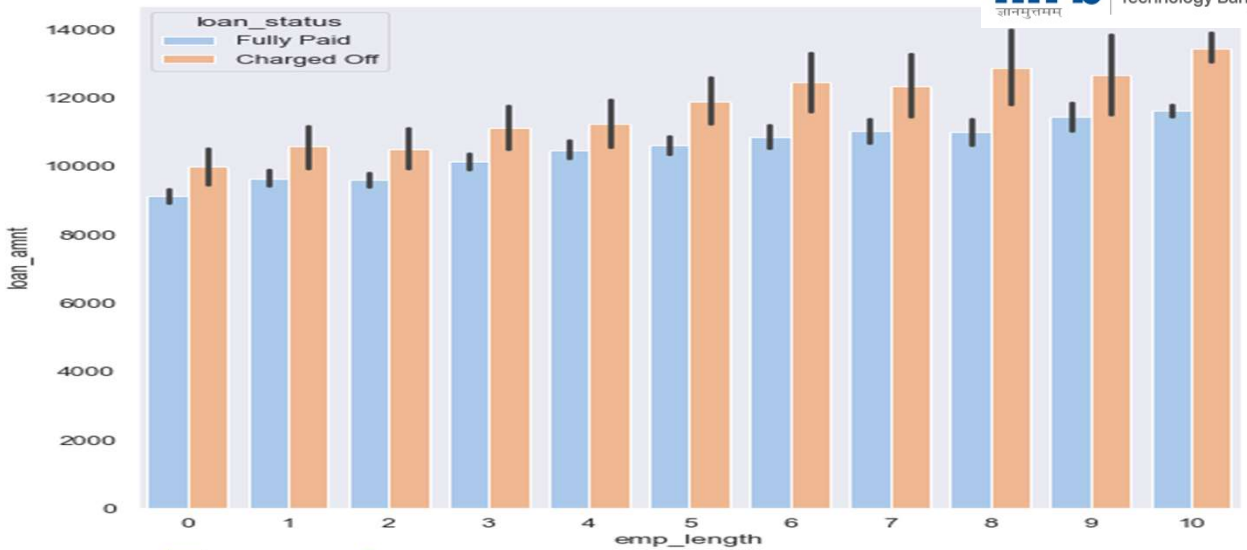
Observation:

- Most of the Loans are of High Grade.
- Higher the loan lower the Interest Rate.
- Higher the Interest Rate higher chances of Loan Default.





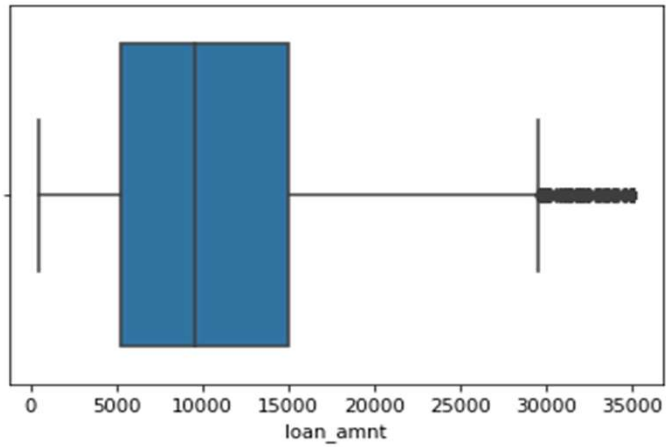
Loan Amount vs Term

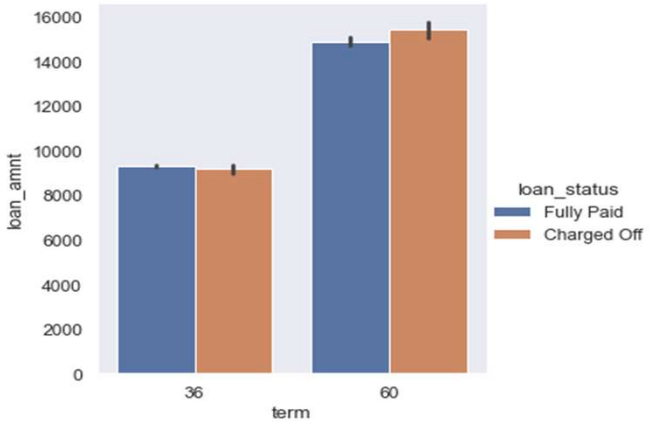


Observation:

- Loan Amount continuously increasing with Employee Experience

Loan Amount Vs Employee Experience

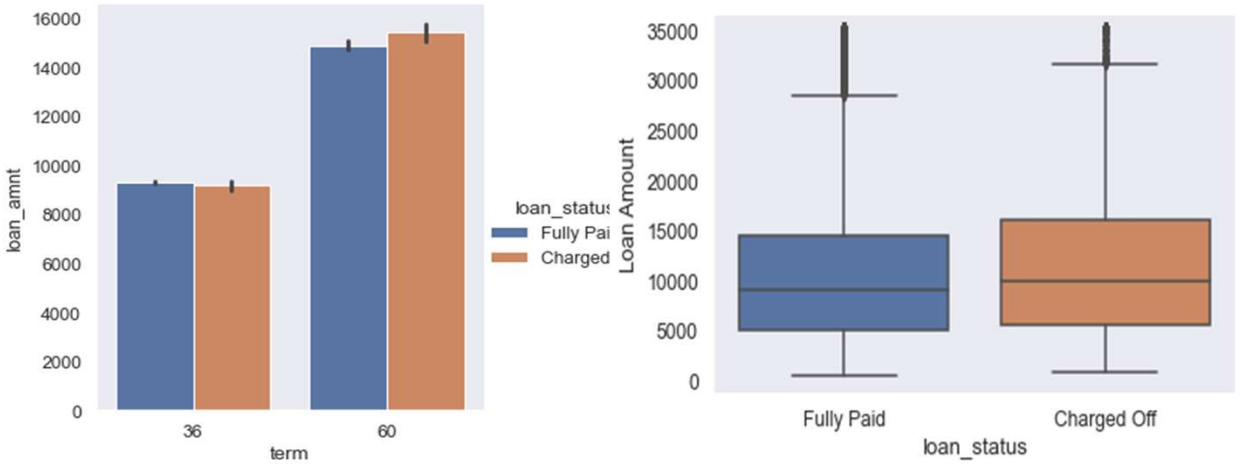




Observation:

- Most of the loans were in the range of 5000-15000.
- Higher the loan higher chance of Loan Default.

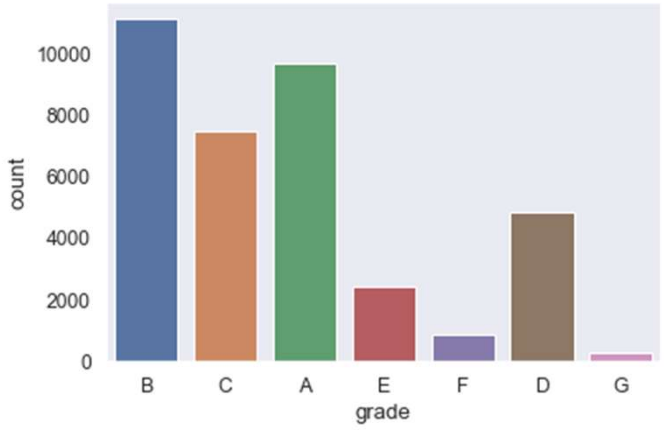
Loan Amount, Loan Status, Tenure

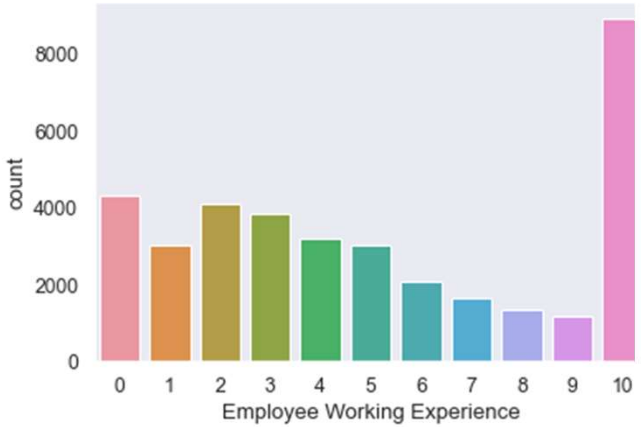


Observation:

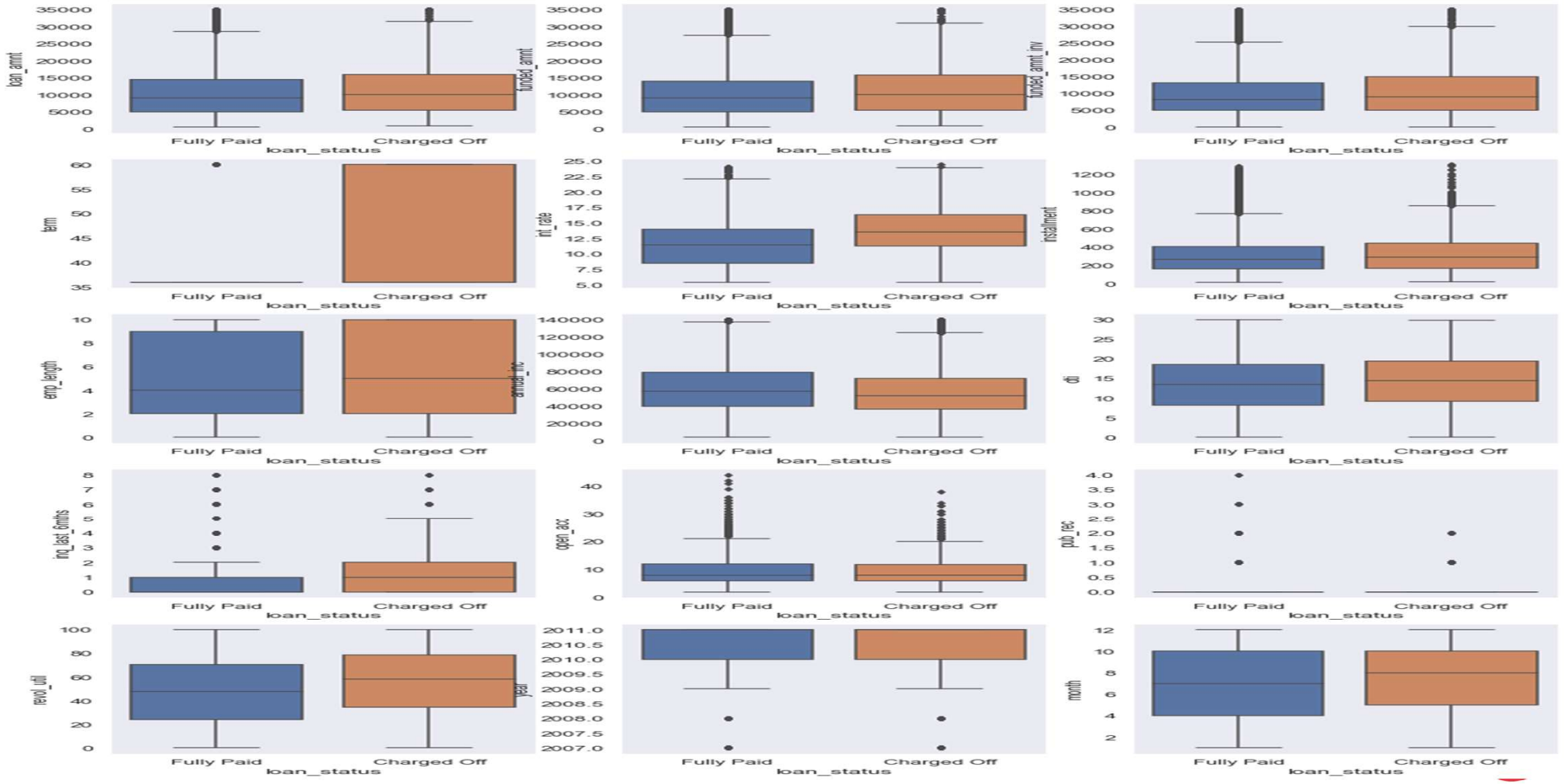
- Higher the Loan term tends to higher Loan Amount and Defaults.
- Higher the Loan Amount higher the Default.

Employee Experience, Grade

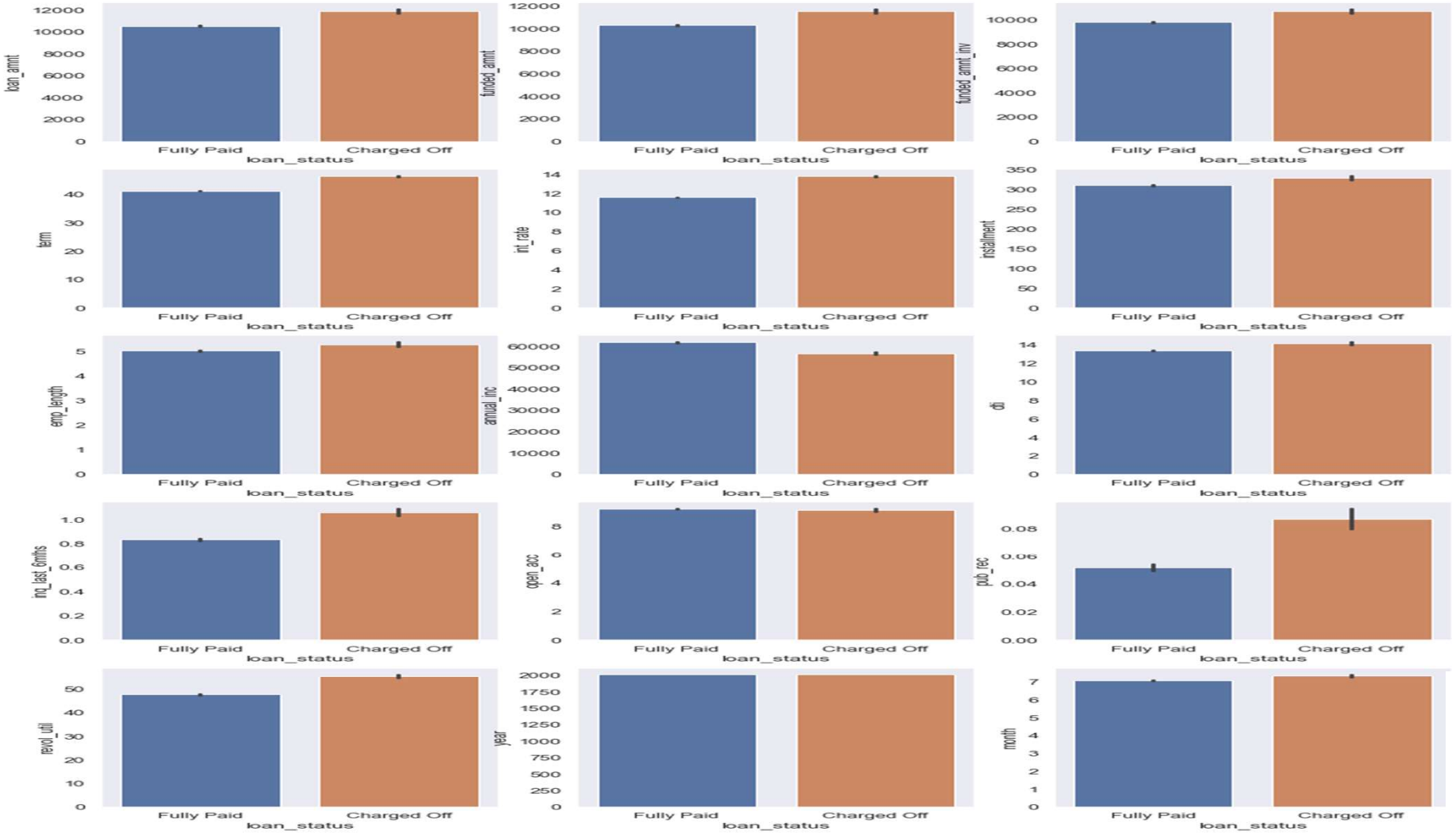




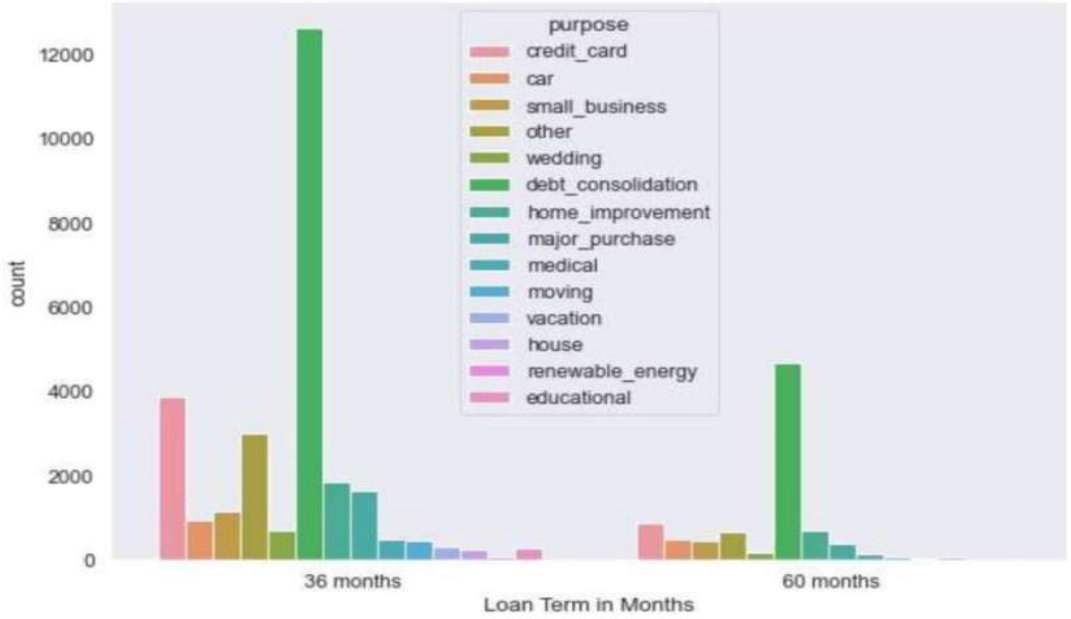
Various Loan parameter against Loan status analysis using box plot

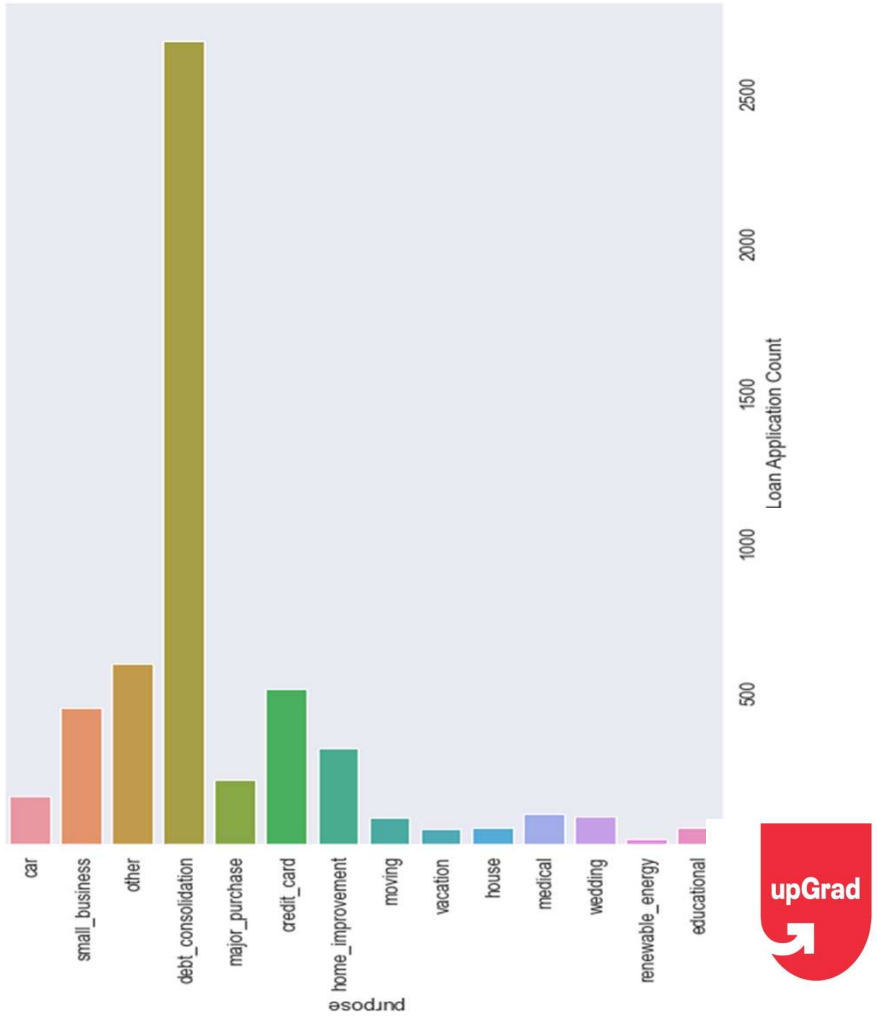


Various Loan parameter against Loan status analysis using bar plot



Loan Term, Loan Count, Purpose





Observations:

- Majority of the employee apply loas for debt consolidation purpose.
- Loan Term of 36 months have high chances to be Defaulters

Conclusion & Observations:

International Institute of Information Technology Bangalore

- **Interest rate is increasing with loan amount increase.**
- **longer the loan term more the interest rate.**
- **Verified loan applications tend to have higher loan amount. Which might indicate that the firms are first verifying the loans with higher values**
- **Most of the loans granted were of 36 months tenure**
- **Loans granted for 36 months have comparitively high tendency to default.**
- **Debt Consolidation is the most common purpose and most loans are granted for 36 and 60 months.**
- **Loan having higher rate of interest have more defaulters.**
- **Majority of employees applying for loans have equal to or more than 10 years of experience.**
- **Tendancy to default the loan with 10 years experience is also high.**
- **Applicants with higher salary mostly applied loans for "home_improvment", "house", "renewable_energy" and "small_businesses"**
- **The high number of loan defaults in 2011 could be due to the financial crisis in USA (Assuming the data is of US origin)**
- **Across all the income groups, the loan_amount is higher for people who defaulted**

upGrad


THANK YOU FOR
YOUR ATTENTION