Lending Club Case Study

Analysis on Loan defaulters

Group Members:
Ruchika Janefalkar & Ruchika Agarwal

Business Objectives

- 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.

- If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.

- The Analysis needs to perform on the past loan applicants and find defaulters

Methodology

Exploratory Data Analysis

Technology used
Python - Pandas ,numpy

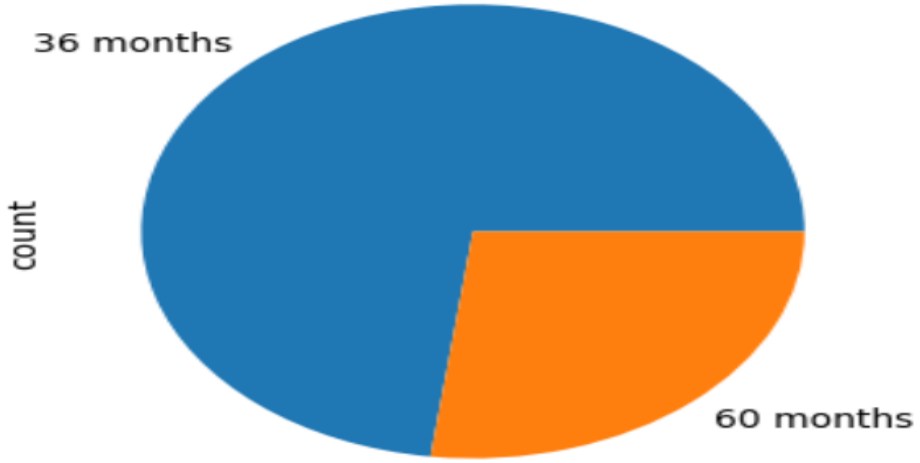
Industry

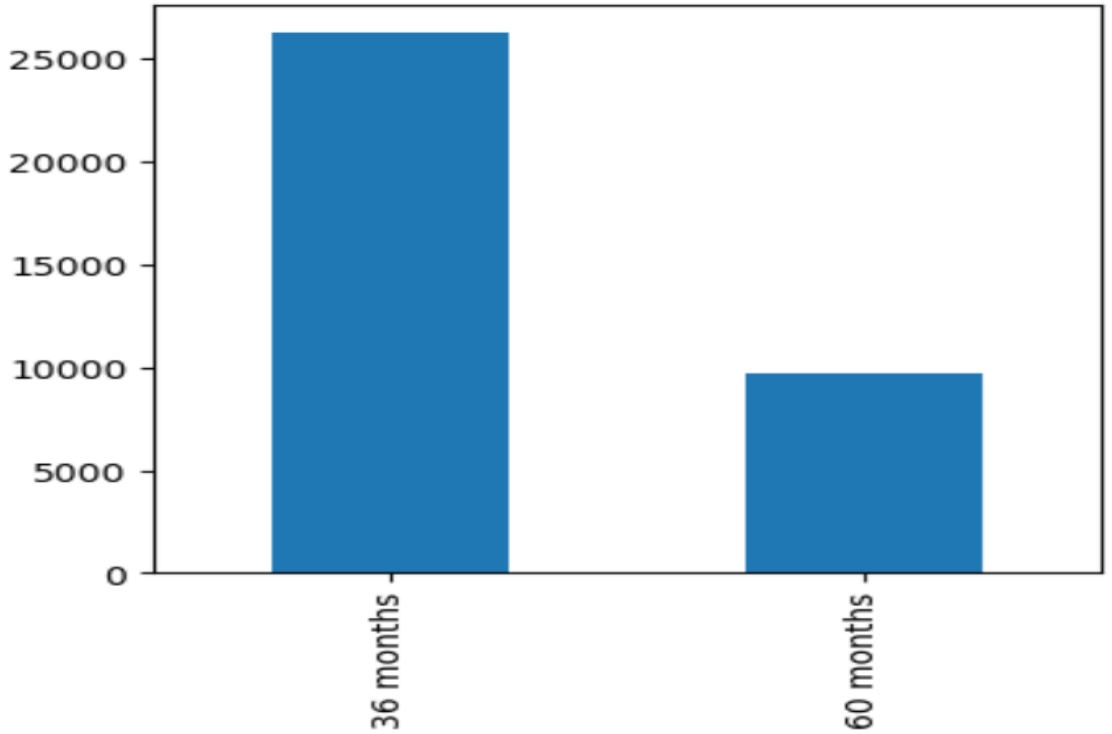
- Banking - online loan marketplace

Analyse on:

- Past applicants who were defaulter.
- Identify risky loan applicants

	EDA	
Data Cleaning & Sanity check	Data Analysis	Observations
1.Removing the null value columns. 2. Removing columns with single value. 2.Removing null values rows. 3.Make the datatype changes for analysis. 4.Check the +, ,,% signs so analysis becomes easier. 5. Categorize - numerical, categorical, extra columns for analysis purpose.	Univariate Analysis – Analyse each column and plotting charts to see distribution of each column. Bivariate Analysis – Analyse relationship between two columns. Multivariate Analysis – Analyse relationship between 3 or more columns by using different graph types.	Share insight based on all earlier analysis.





Term Column analysis

This has two terms only 36 Months and 60 Months.

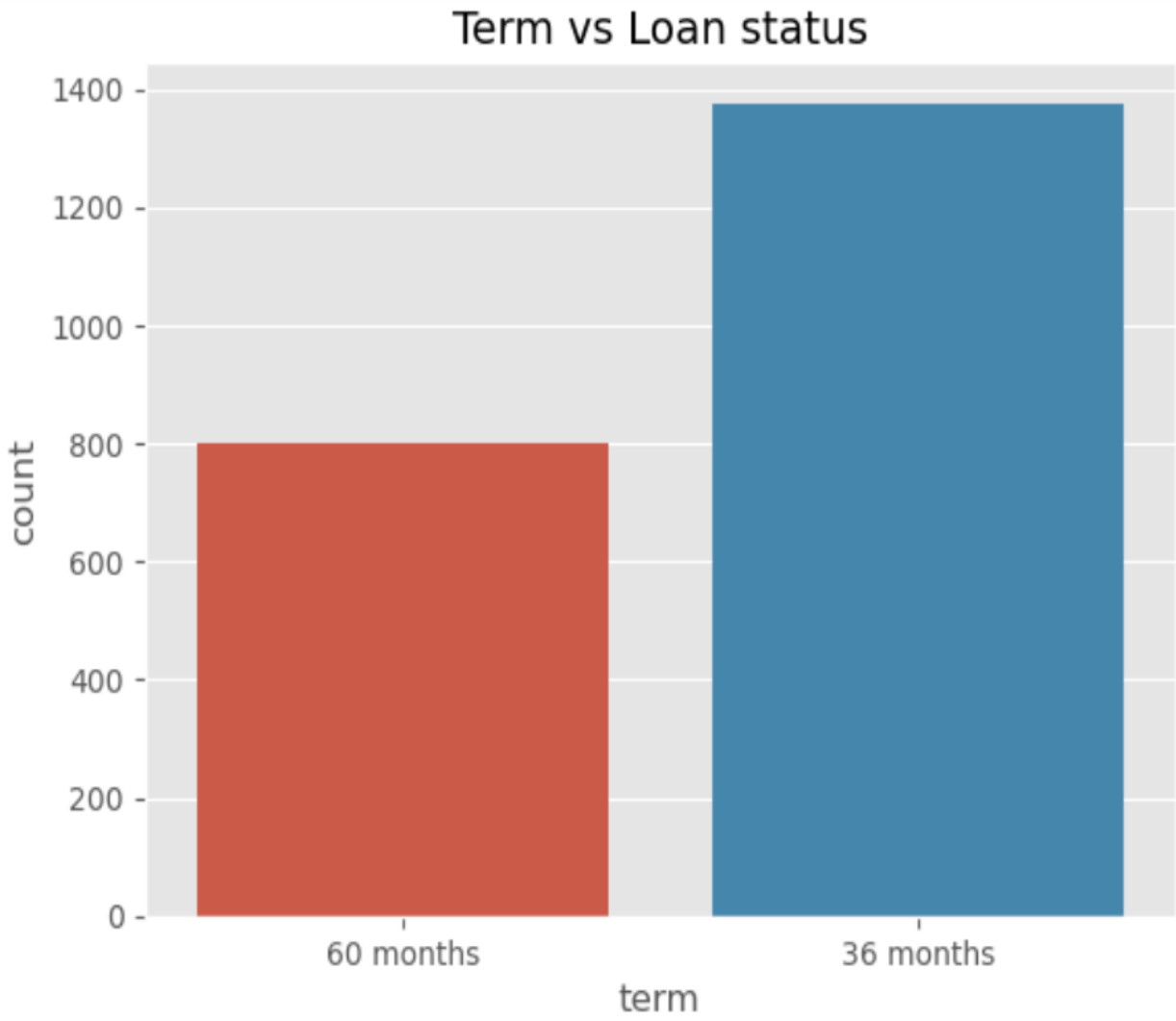
This indicates that borrowers have taken 36 months tenure more than 60 months

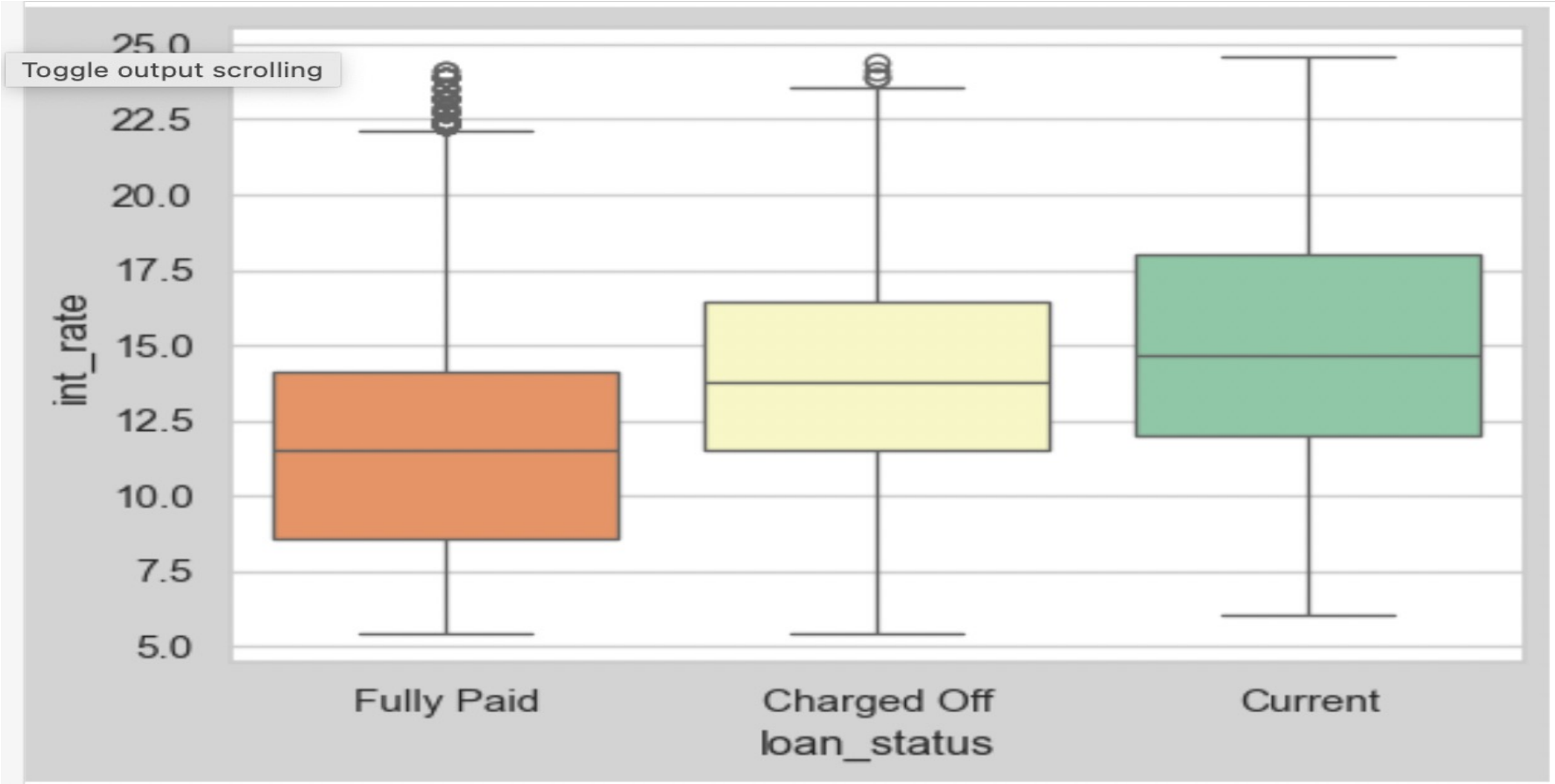
Term vs Loan status

1. 10.5% loans with 36 months defaulted.

2. 22.3% loans with 60 months defaulted.

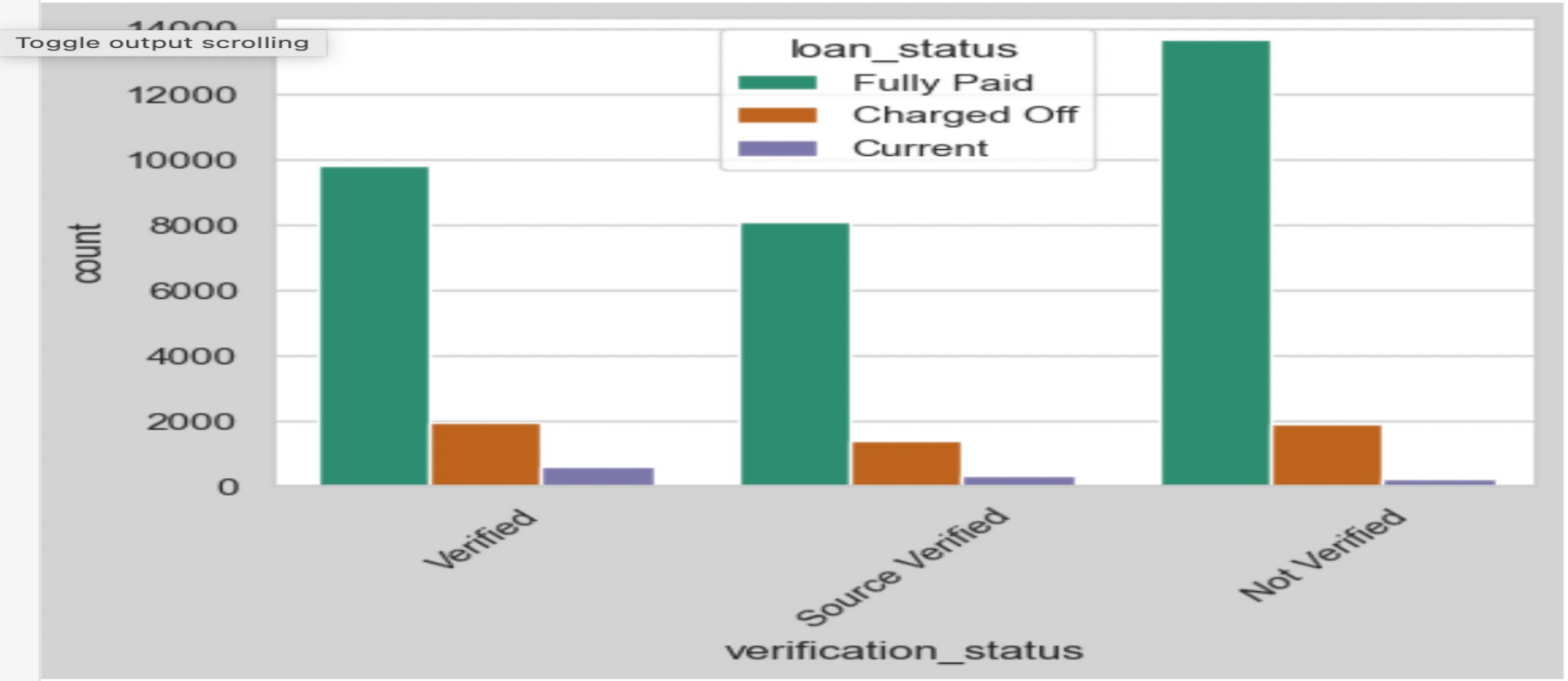
3. The defaulters are 13.7% and fully paid is 83.3% in the data.





Loan status vs interest rate

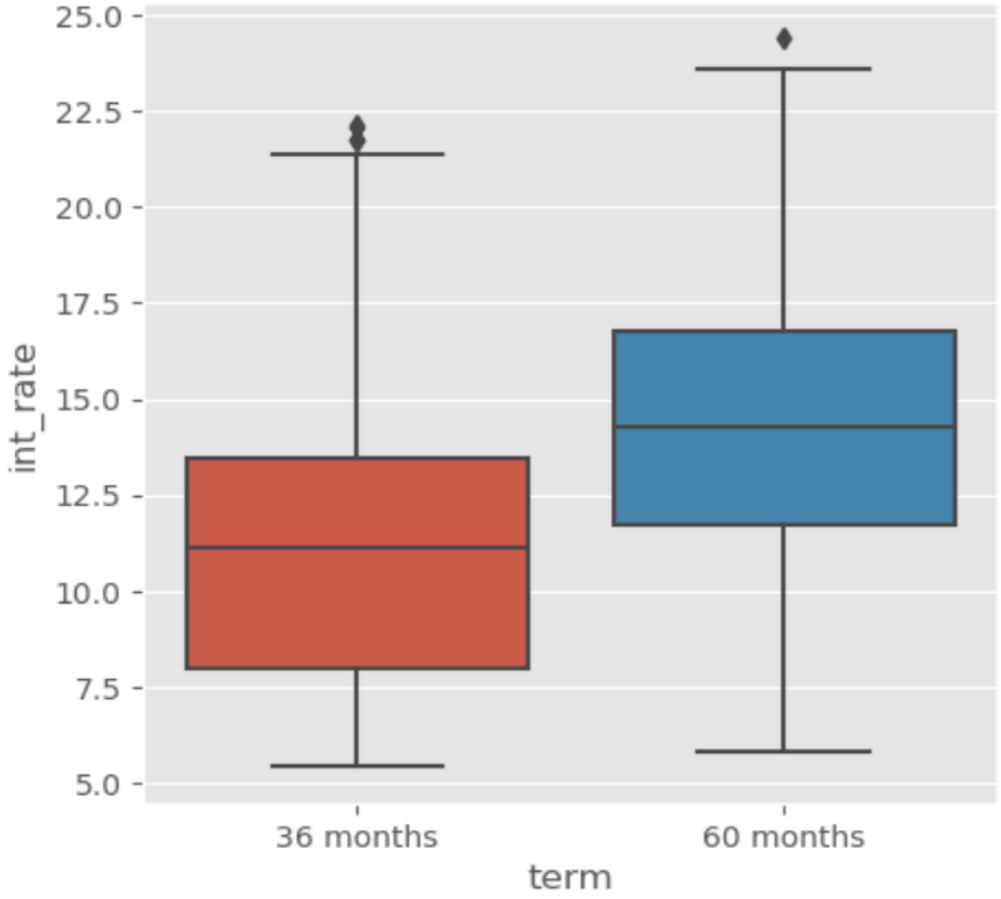
This indicates that customer defaults loan when interest rate is on higher side between 11%-16%

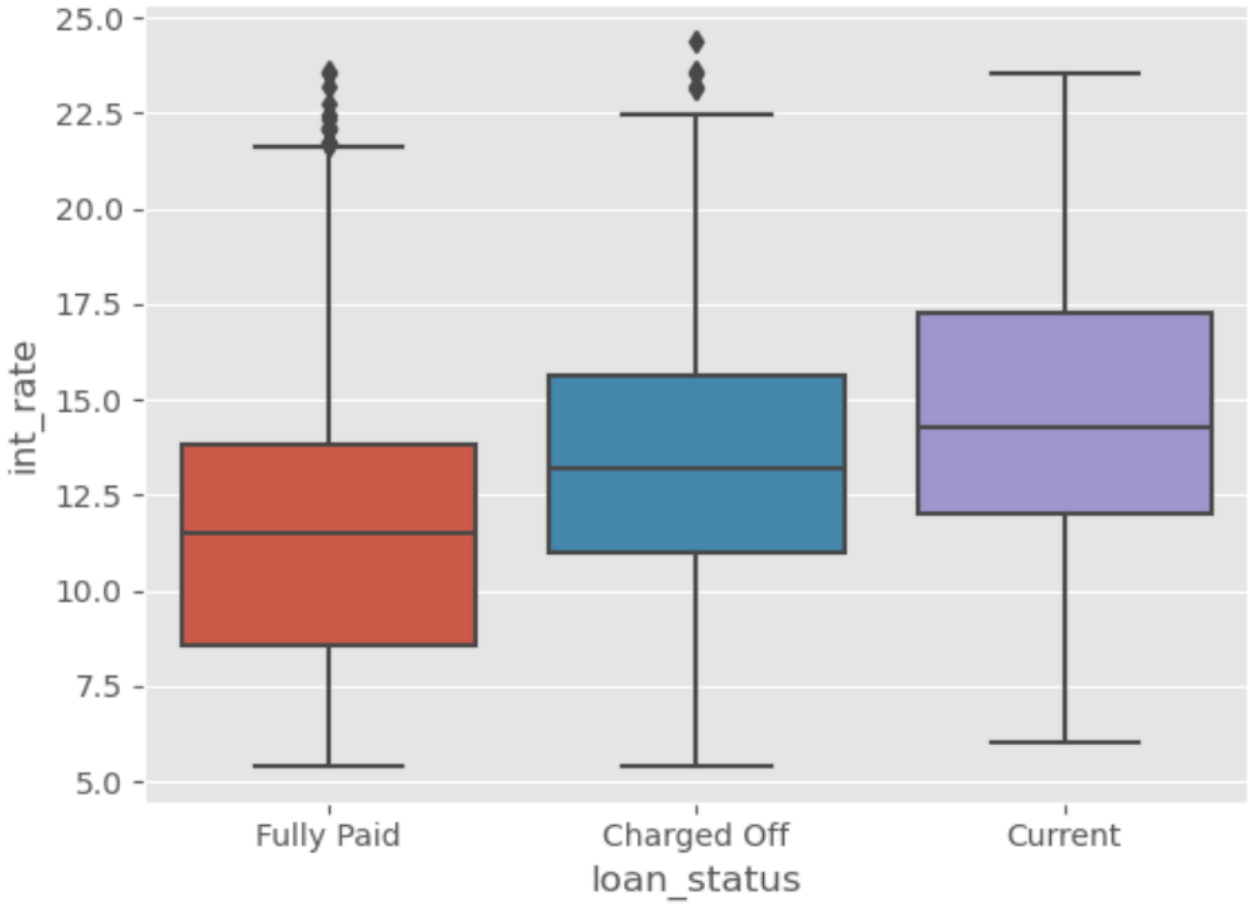


Loan Status vs Verification status

Majority of loans are 'Not Verified'. Probability of verified loans to default are more than source verified. Not verified loans are also close behind to default.

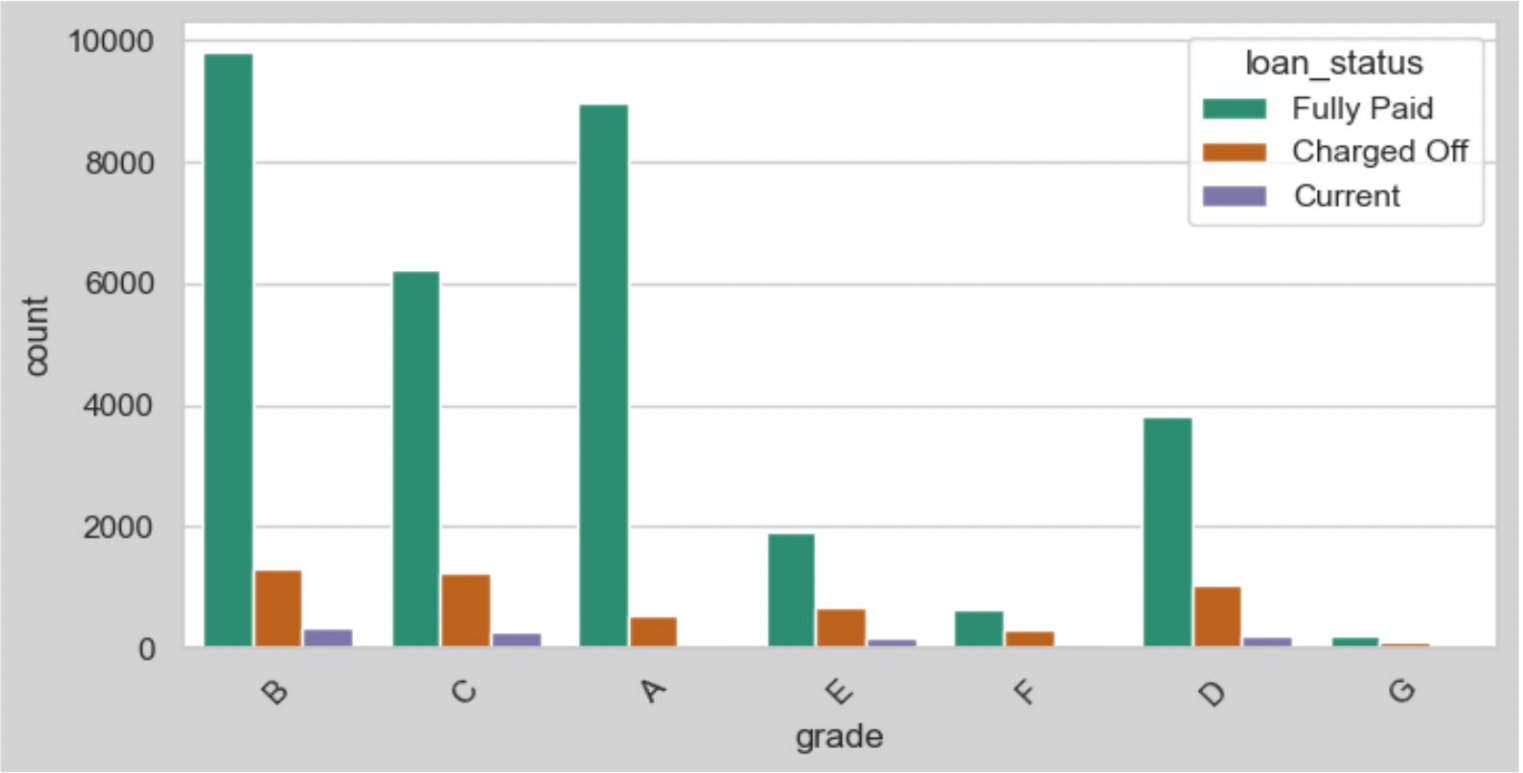
It means LC is not verifying most of the loans and the ones that are verified are not done properly and there probability to default is high.





Analyze Term, Interest rate, Loan status=Charged off

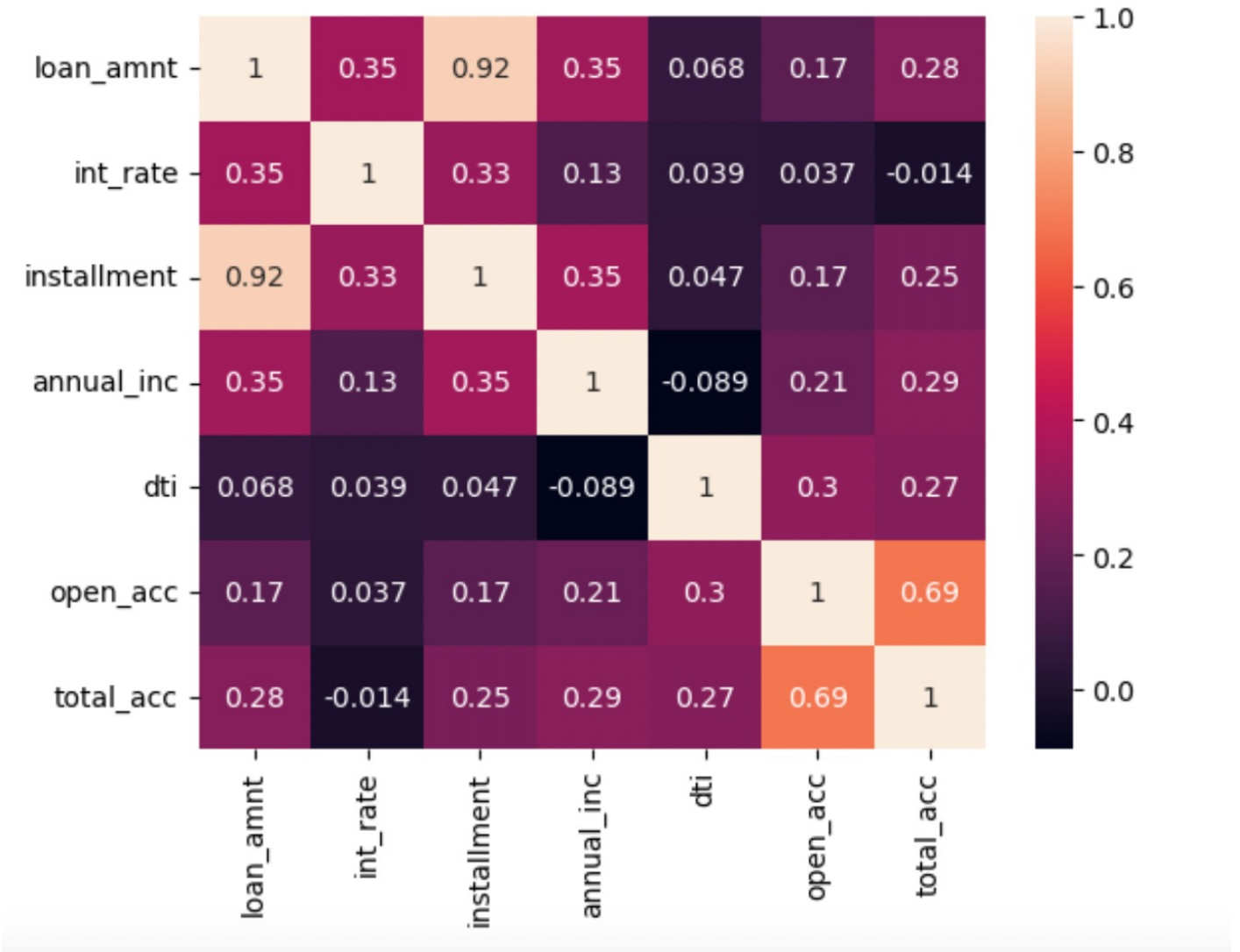
The probability of increase in defaulters when interest rate is between 11.5%-16.5% and 60 months term.



Grade vs loan status vs loan amount

There is high probability of defaulting for customer having grade G,F,E,D,C as their percentage of charged off is more as compared to the total loans of these grades.

Multivariate Analysis for charged off loans



Inferences towards the loan to default:

1.Majority of loans are ’Not Verified’. Probability of verified loans to default are more than source verified. Not verified loans are also close behind to default. It means LC is not verifying most of the loans and the ones that are verified are not done properly and their probability to default is high.

2.On higher interest rates the average defaulters are more. Also, the loans with status - 'current' are also of higher int rates so probability of current loans to default are high.

3.There is no other prominent variable for a loan to default apart from interest rate. The occurrences with highest 'Charged Off's in any graph is majorly because of high amount of loan sanctioned for that value.