

The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, creating a modern and dynamic visual effect.

Lending Club Case Study

Exploratory Data Analysis

Vijayasree Patnaik

Viral Vasava

Background

Lending club is the largest online loan marketplace facilitating personal loans, business loans and financing of medical procedures. It provides lucrative interesting rates through online interface

Problem Statement

lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed.

Business Objective

The objective of this assignment is to identify driving factors behind loan default, i.e the variables that are indicators for the default based on the historical loan data. Understand the patterns for the loan defaulters

The company can utilize this analysis for deciding which borrowers potentially has the tendency to default the loan and do the right risk analysis before disbursement of the loan

Data Understanding Summary

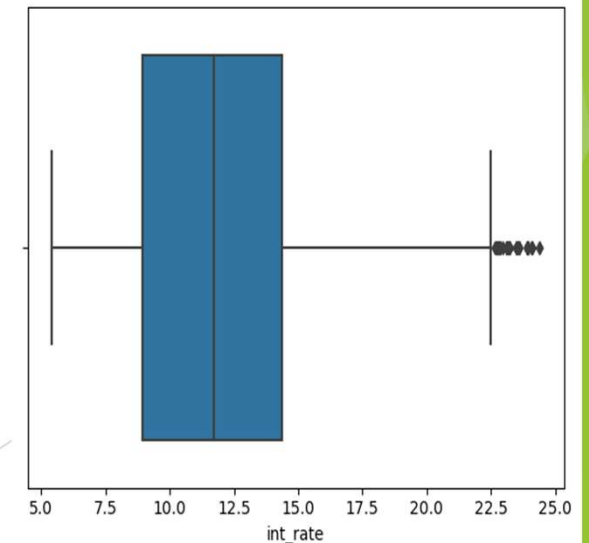
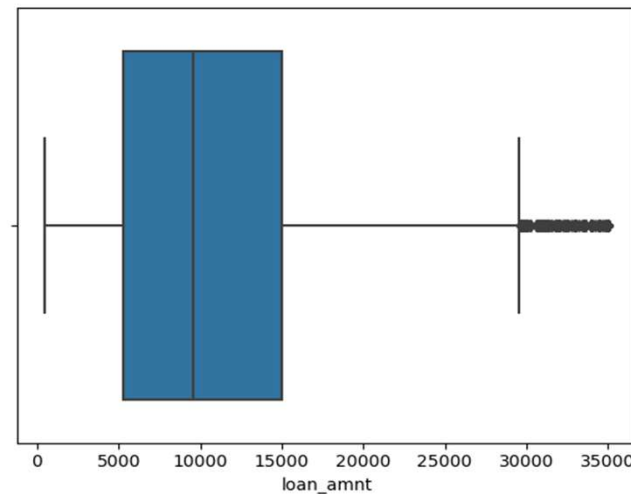
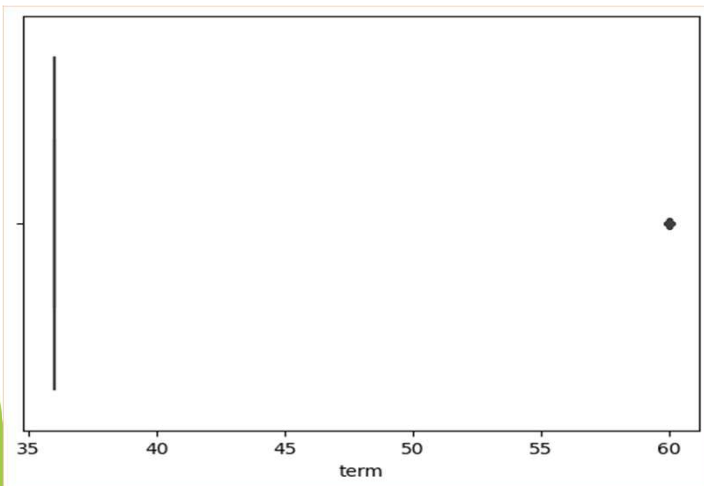
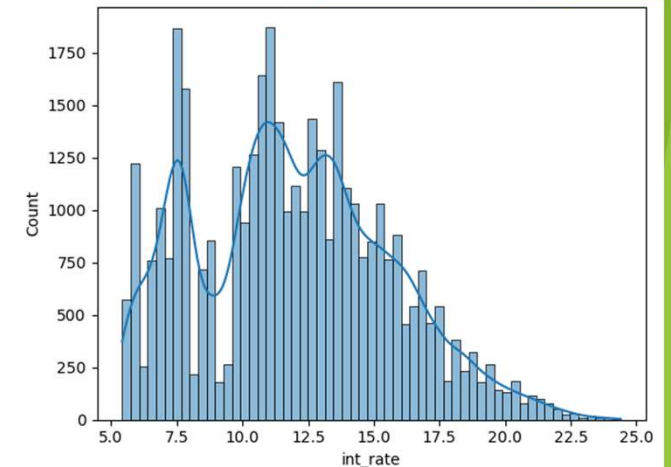
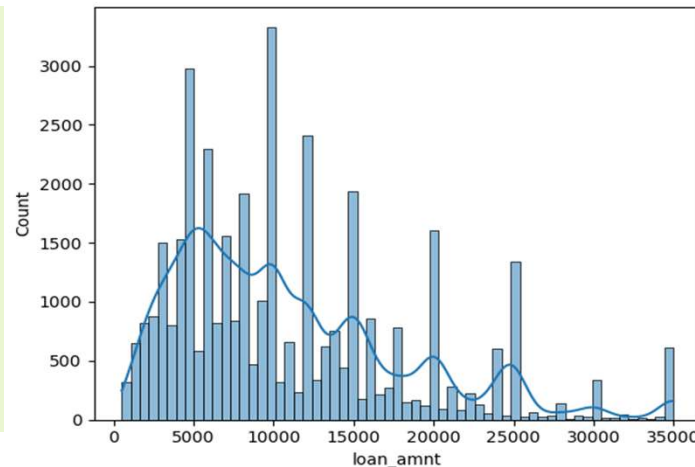
There are primarily 3 categories of data

1. Customer information
2. Loan Information
3. Customer Behavior



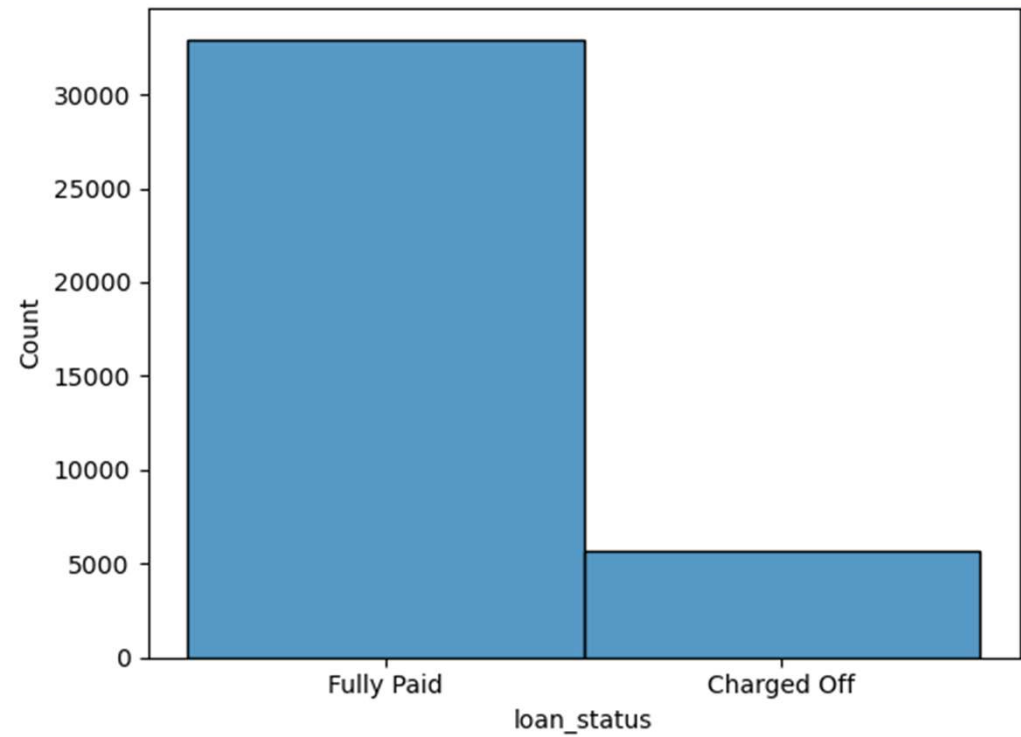
Univariate Analysis - Numeric variables

- Most of the customers have average to low Loan amount and interest rates
- There are few customers that have high interest rates and loan amounts
- Most of the customers are falling into 36 months term



Loan Default

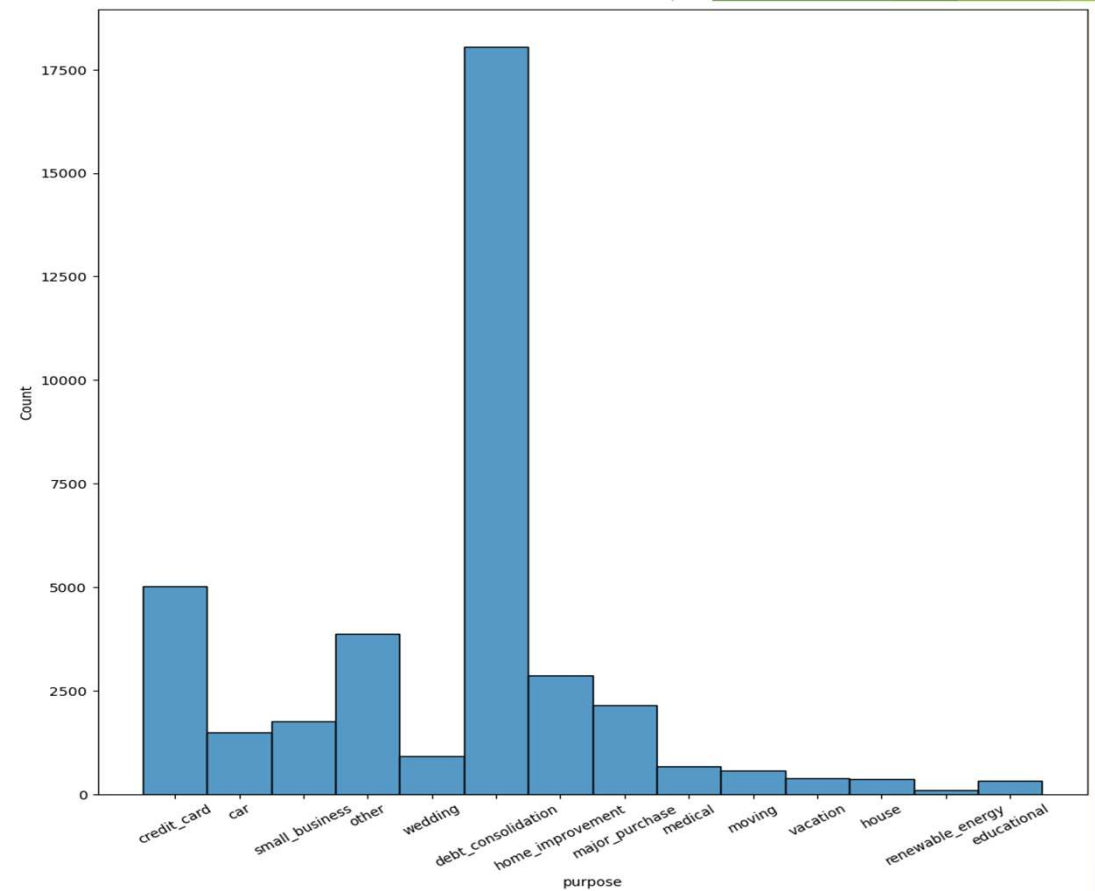
Total 85% of the loans are fully paid
Around 15% loans are defaulted



Loan Purpose

debt_consolidation	46.8%
credit_card	13.0%
other	10.0%
home_improvement	7.5%
major_purchase	5.6%
small_business	4.5%
car	3.9%
wedding	2.4%
medical	1.8%
moving	1.5%
vacation	1.0%
house	1.0%
educational	0.8%
renewable_energy	0.3%

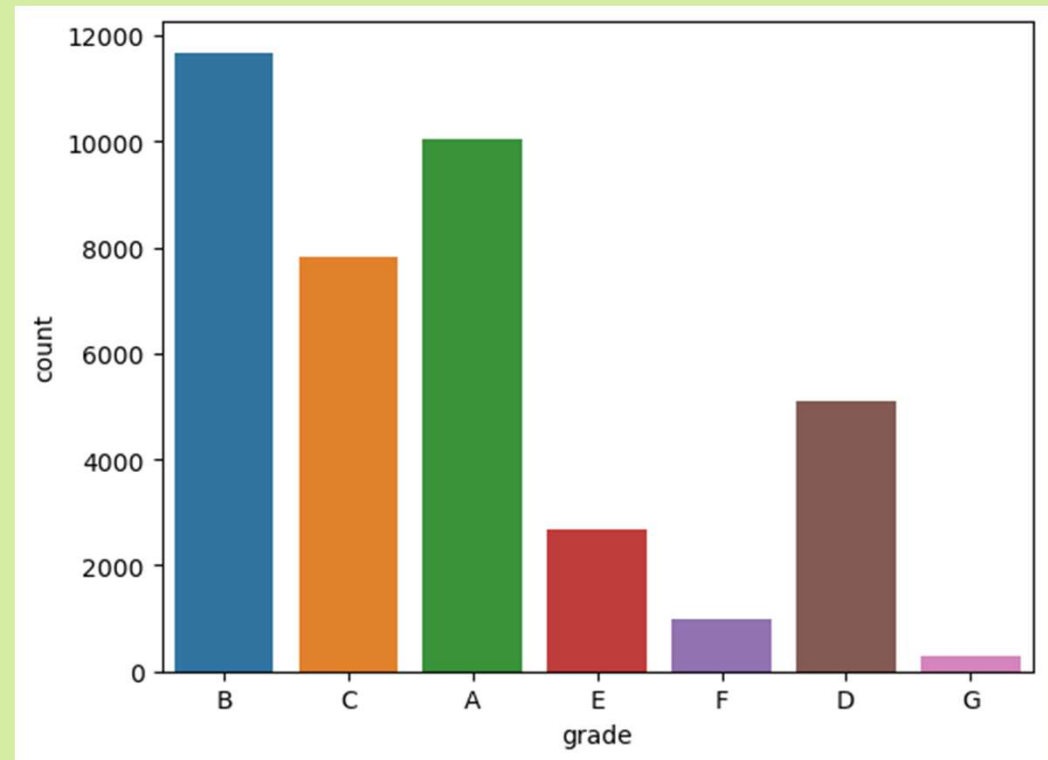
Debt_Consolidation contributes to the major
Percentage of loan data



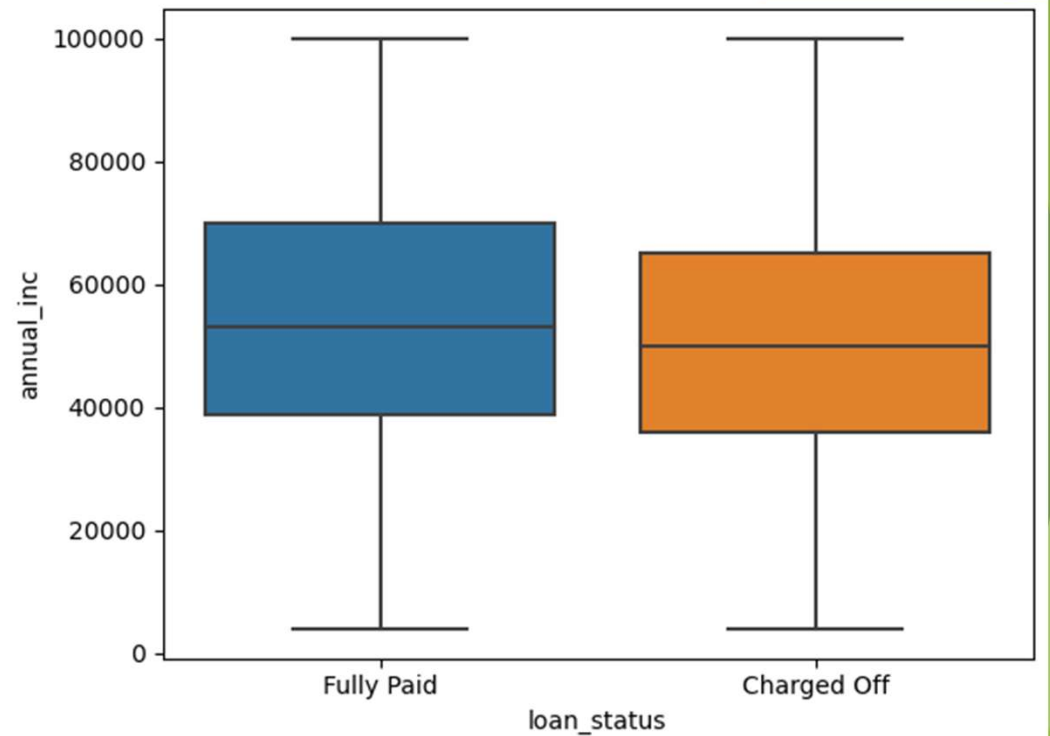
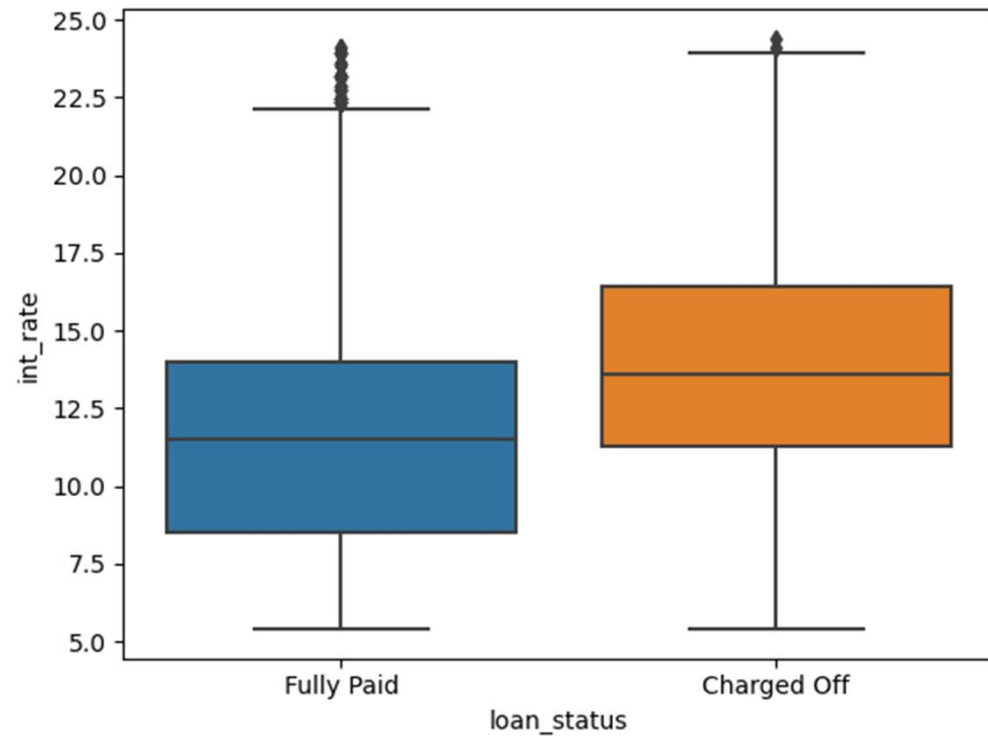
Loan Distribution by Grade

B	30.3%
A	26.0%
C	20.3%
D	13.2%
E	6.9%
F	2.5%
G	0.8%

Loan Distribution for A and B Category is almost equal

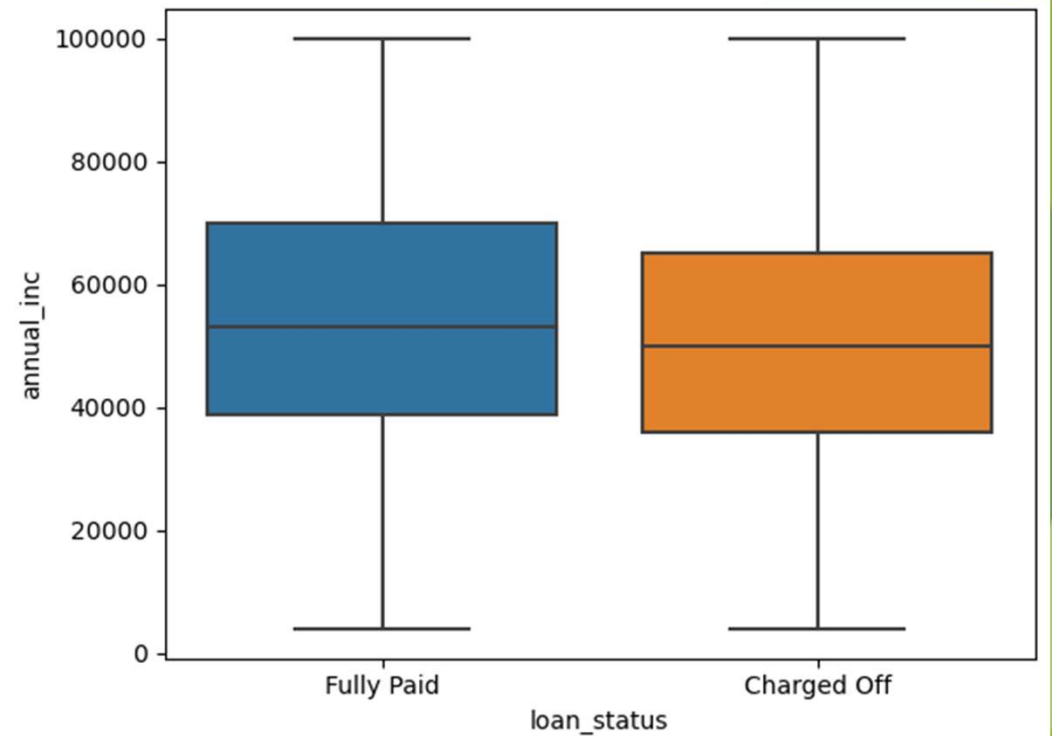
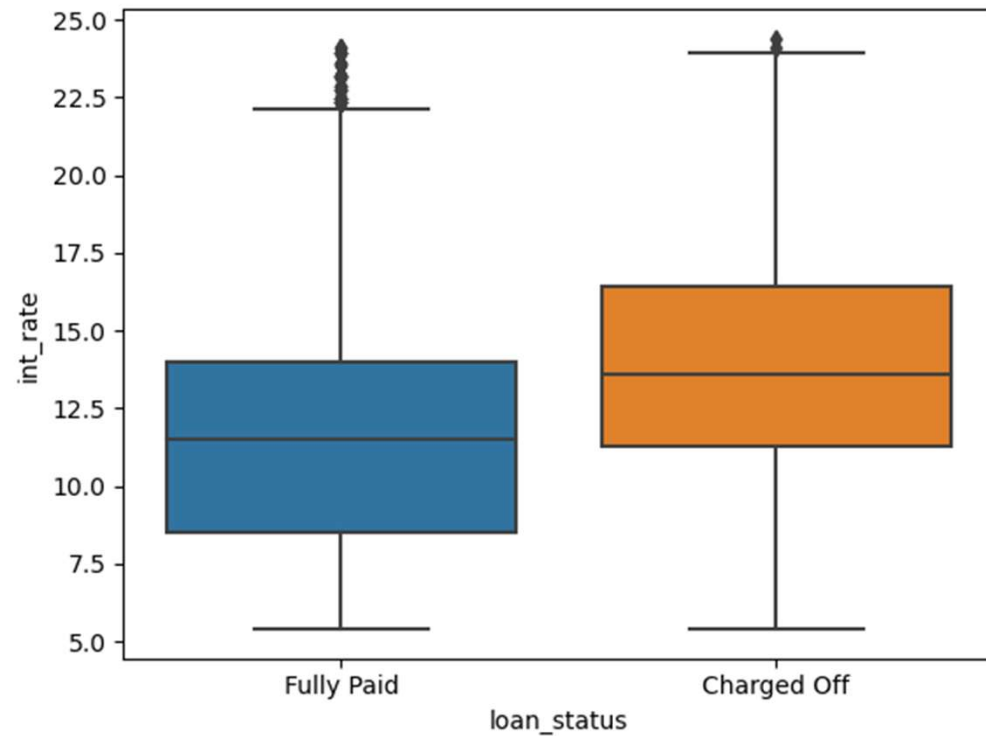


Loan Status



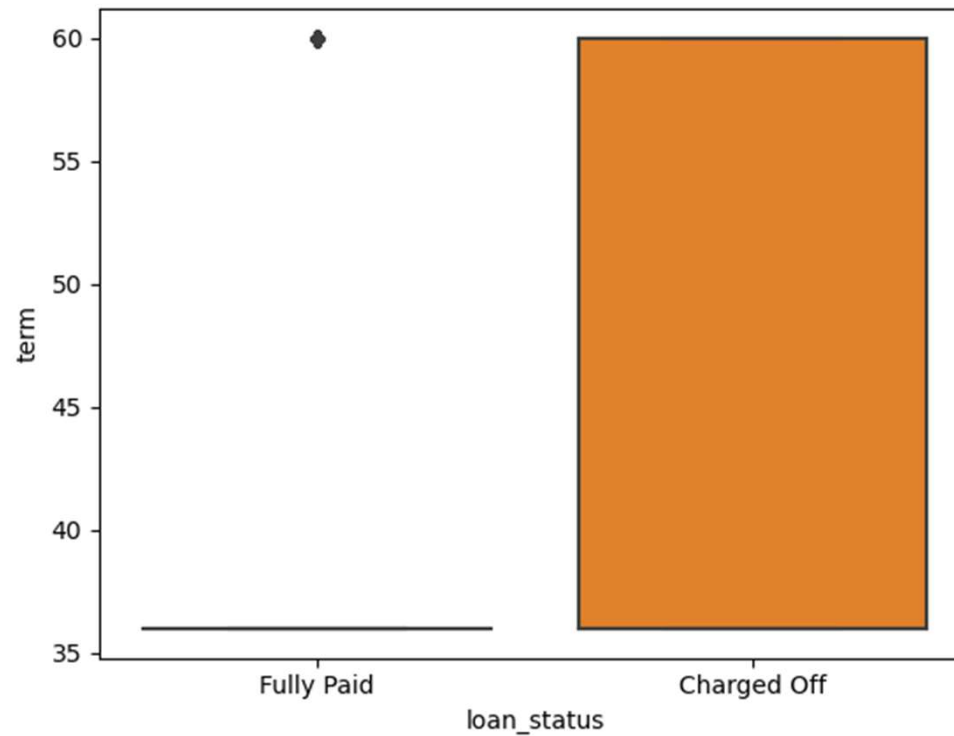
- Customers with high income less default on the loan
- Higher interests has a very strong impact on the loan defaulters

Interest Rate, Annual income rel Loan Status



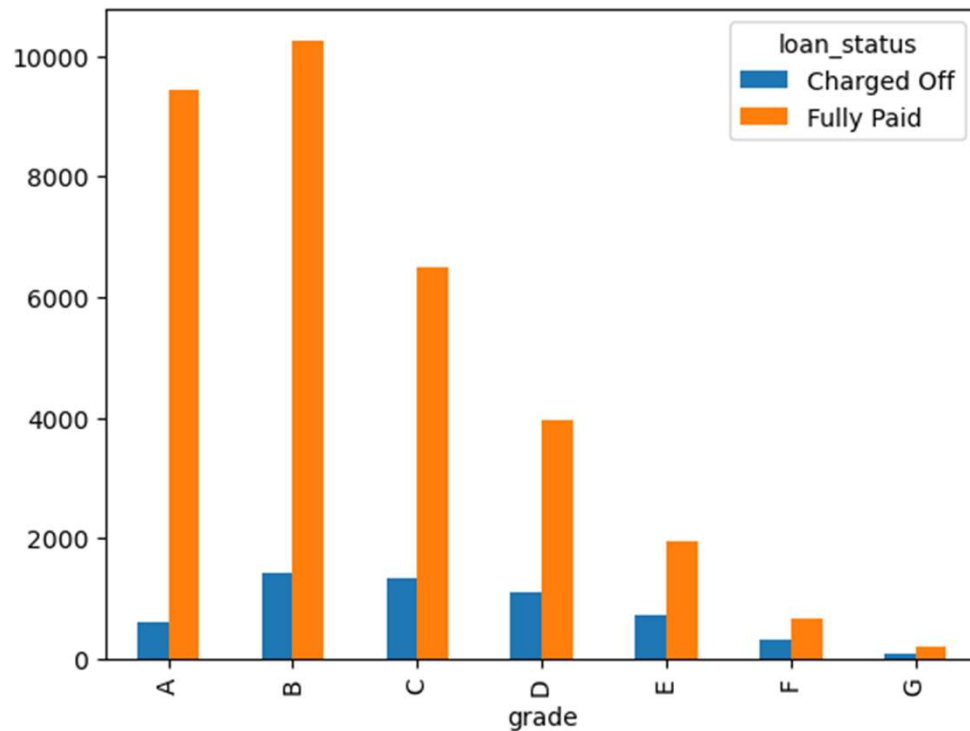
- Customers with high income less default on the loan
- Higher interests has a very strong impact on the loan defaulters

Loan Status Relationship with Term

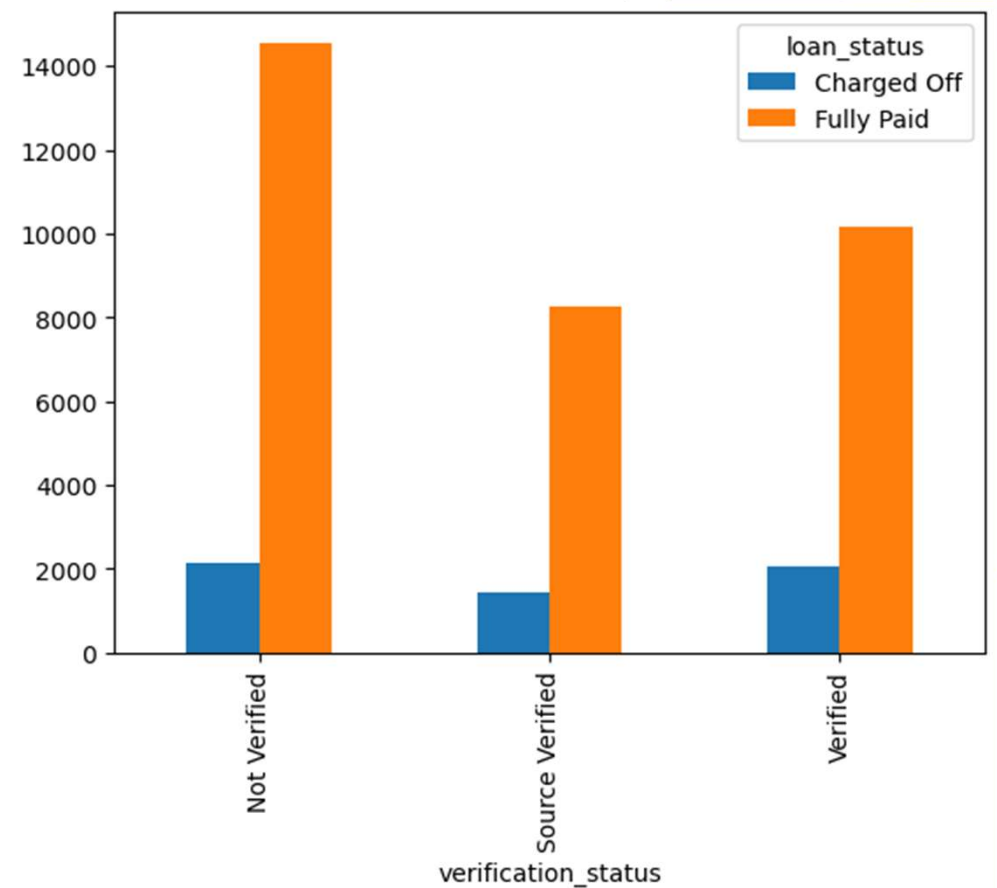


Higher term of loan has more loan defaulters

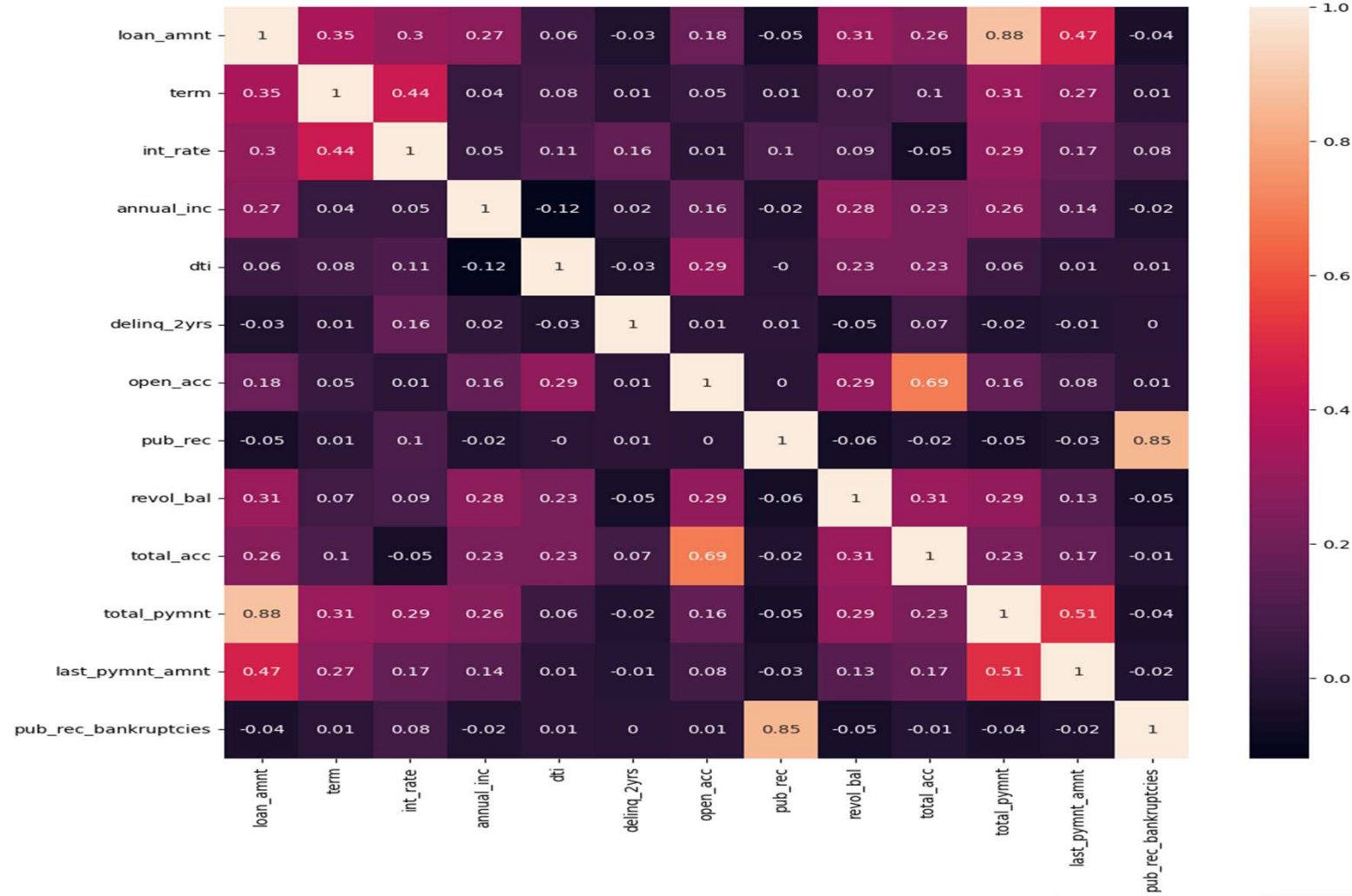
Loan Status, Verification Status, Grade



- More defaulters in Grade B, C and D type of loans. Percentage of defaulters in Grade is higher
- Loan defaulters have not much significance on verification status



Heat Map



Observations from Heatmap

1. funded_amnt, funded_amnt_inv and installment amounts are highly related to loan_amnt data
2. total_pymnt_inv is highly related to total_pymnt data
3. Number of public record bankruptcies data is not much impacting positively or negatively to any of the loan parameters
4. Most of the numeric data values are not much correlated with each other

The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green. These shapes are positioned on the left and right sides of the slide, framing the central text. The overall aesthetic is modern and minimalist.

Thank You!