Lending Club Case Study (ML C61)

Kumar Ashish
kr.ashish8196@gmail.com

Business Overview

Lending Club is in business of providing loans to Urban Customers. This analysis will try to find factors which can help determine customers who are likely to default on their loans.

The dataset provided had around 40000 rows and 111 columns, detailing the different factors which result in the customer either defaulting or fully paying the loan.

The analysis has been done on around 36000 rows and 41 columns from the original dataset post cleanup.

Loan Amount
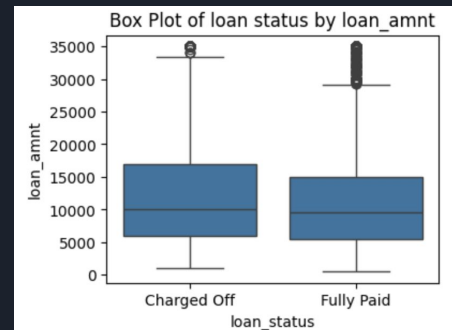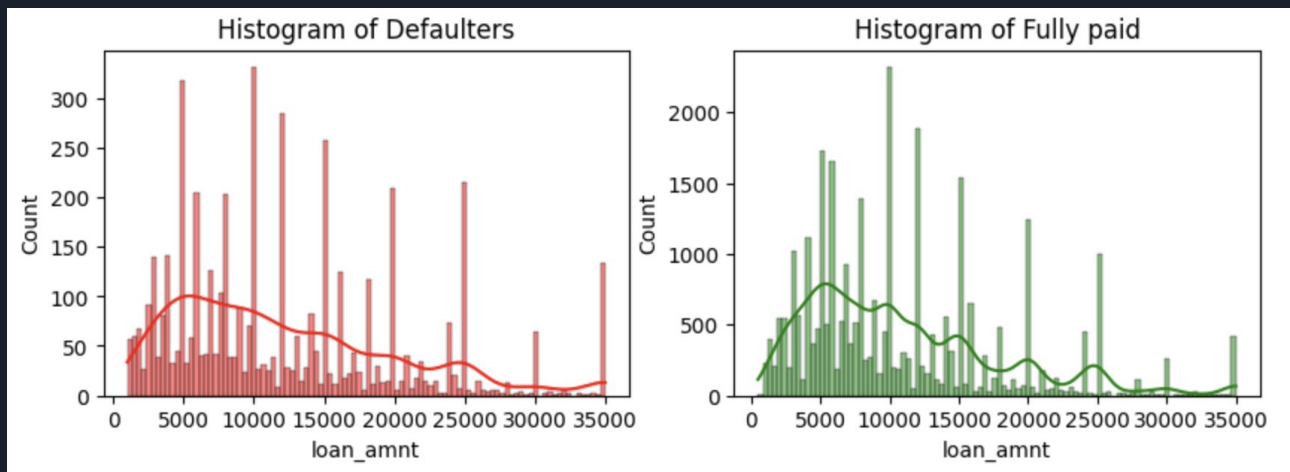
Lending Club gives loan in the range of 1000 to 35000. We can see that both fully paid and defaulters follow a similar pattern of withdrawing loans, albeit there is a huge spike in defaulting for 35000 compared to fully paid. From the box plot also, we can see defaulters usually have higher loan amounts

Annual Income
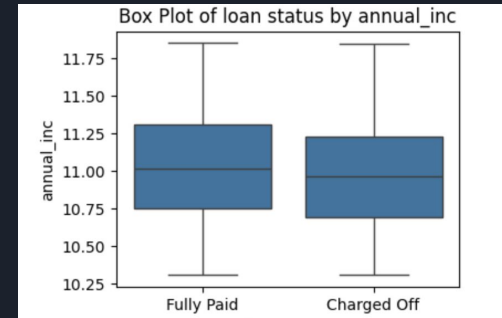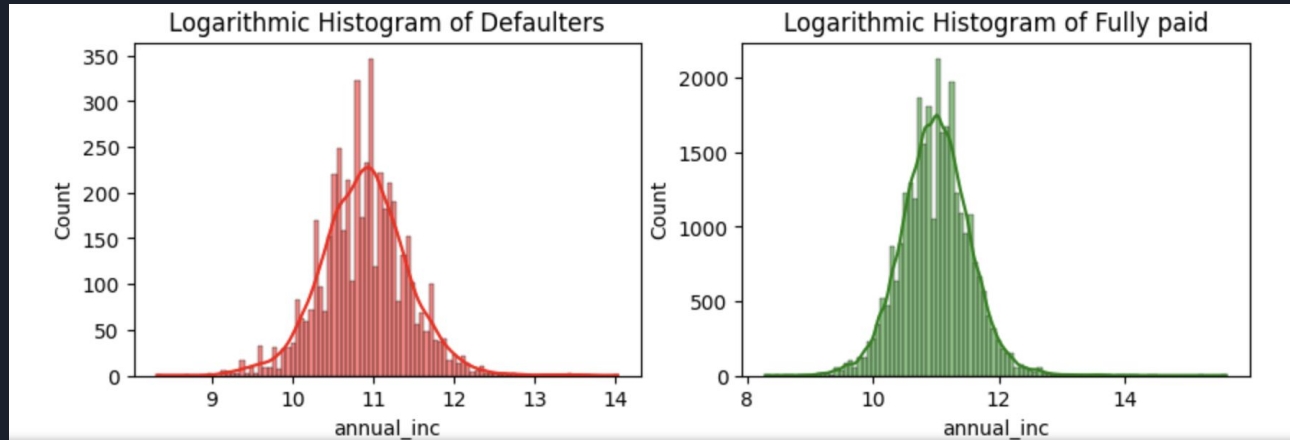
Annual Income is following a normal distribution for both defaulters and fully paid, with mean somewhere around 11. From box plot we can see that defaulters generally tend to have lower annual income

Interest Rate
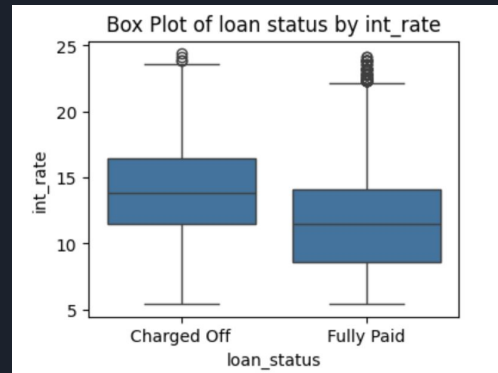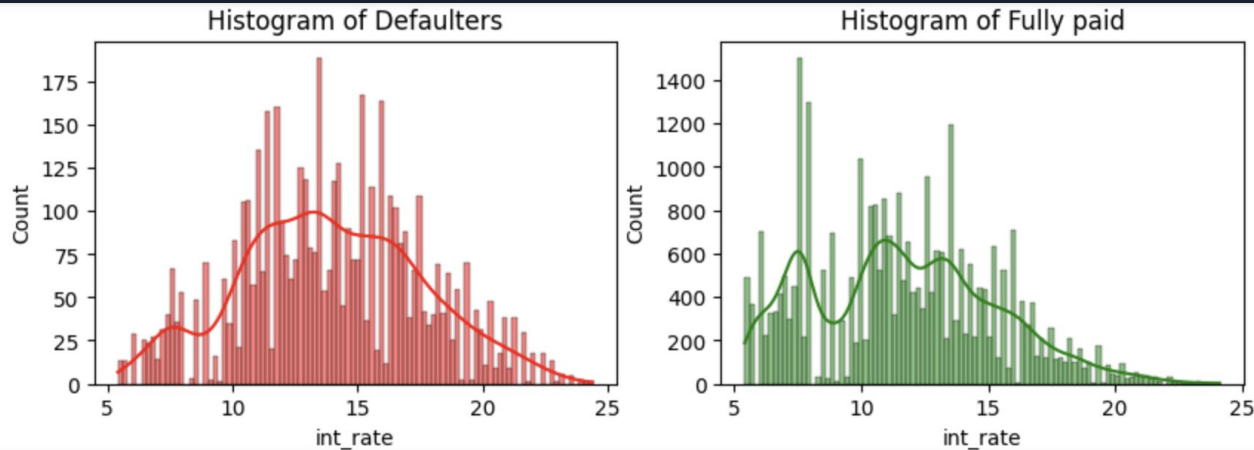
From the histogram, we can see that fully paid users have a generally lower interest rate. From box plot, it is obvious that higher interest rates are definitely causing users to default
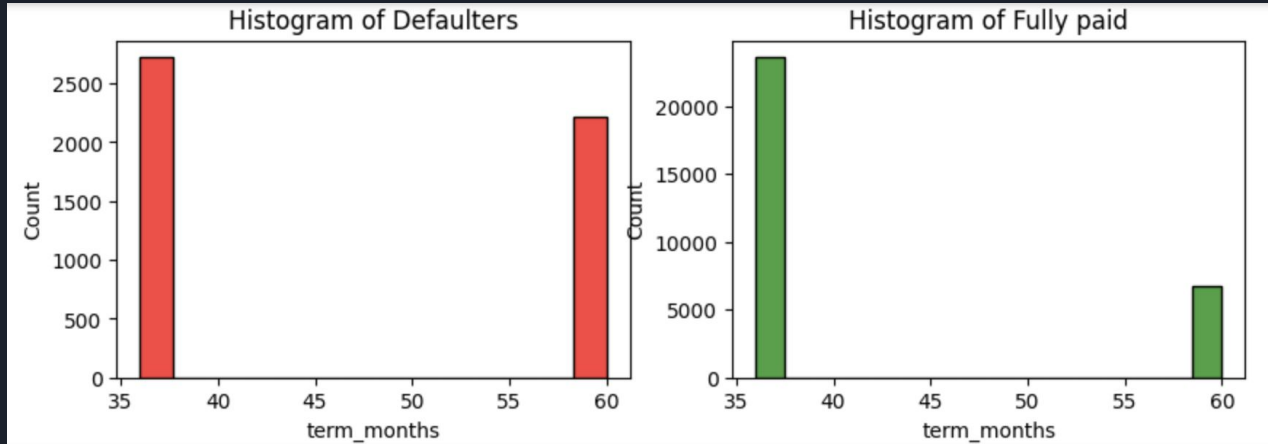
Grade Value
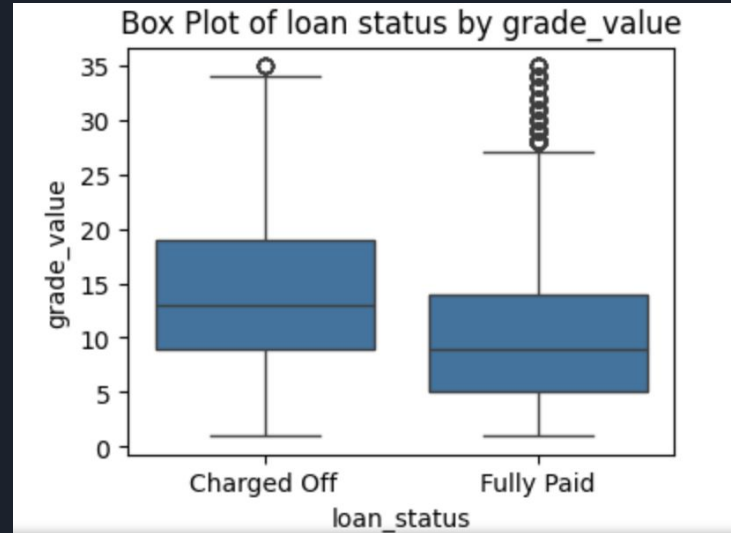
A1	1
A2	2
A3	3
B1	6
B2	7
C1	16

With grades taking lower values as follows, i.e. A1 = 1, A2 = 2... C1 = 16 etc. , we observed that when this value is higher, chance of defaulting is high. So C1 is more likely to default compared to A1.

Term Months

Loan duration can be of either 36 months or 60 months. It was observed that people taking loan of 60 months are more likely to default compared to those taking loans of 36 months

Recommendations

1. Higher loan amounts are riskier than lower loan amounts, need to do more due diligence when processing the same
2. People with lower annual income tend to default more, hence we should try to prioritise people with more income
3. Higher interest rates cause more defaults, Lending club should try to provide more manageable interest rates
4. Grade values like E, F, G are much riskier compared to A, B, C. We should try to avoid them
5. Higher loan terms tend to default more, we should try to process these very carefully