Lending Club Case Study

IIITB ML and AI Program-Feb 24

Poonam Sharma

Selva Kumar

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 39717 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 30 columns from the original dataset post cleanup.

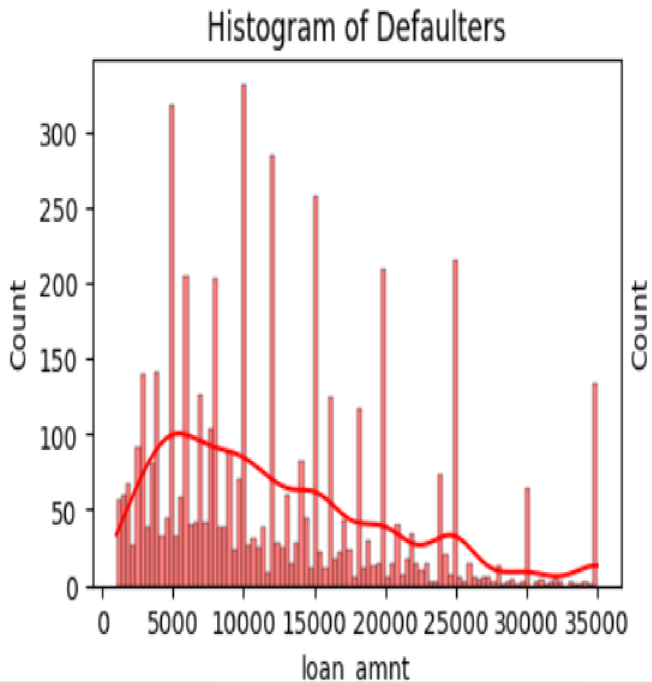
Python Libraries Used In Analysis

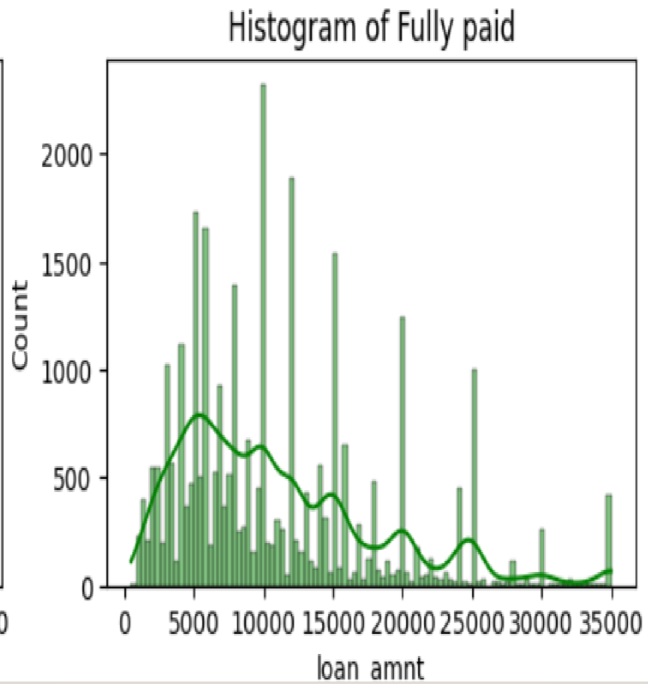
- Numpy
- Pandas
- Matplotlib
- Seaborn

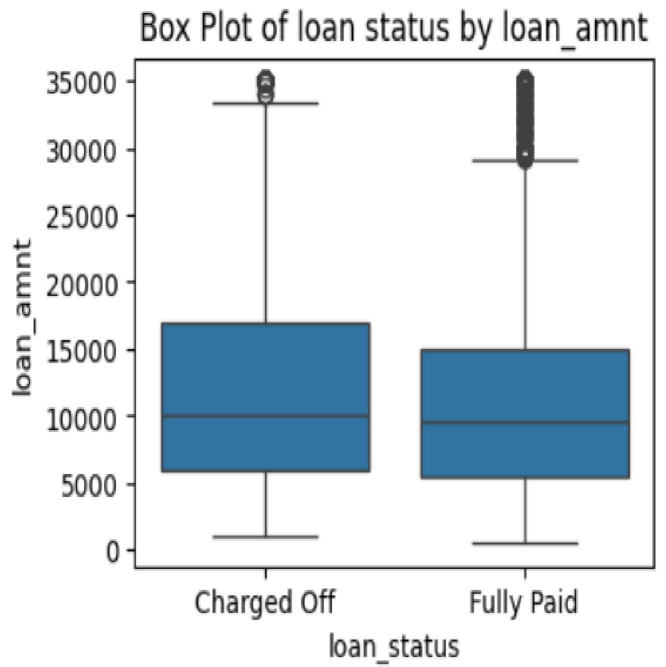
Analysis Approach

- Data Understanding: Load and inspect loan dataset
- Data cleaning: Handle missing values and drop irrelevant columns
- Univariate Analysis
- Bivariate Analysis
- Insights and Recommendations

Loan Amount

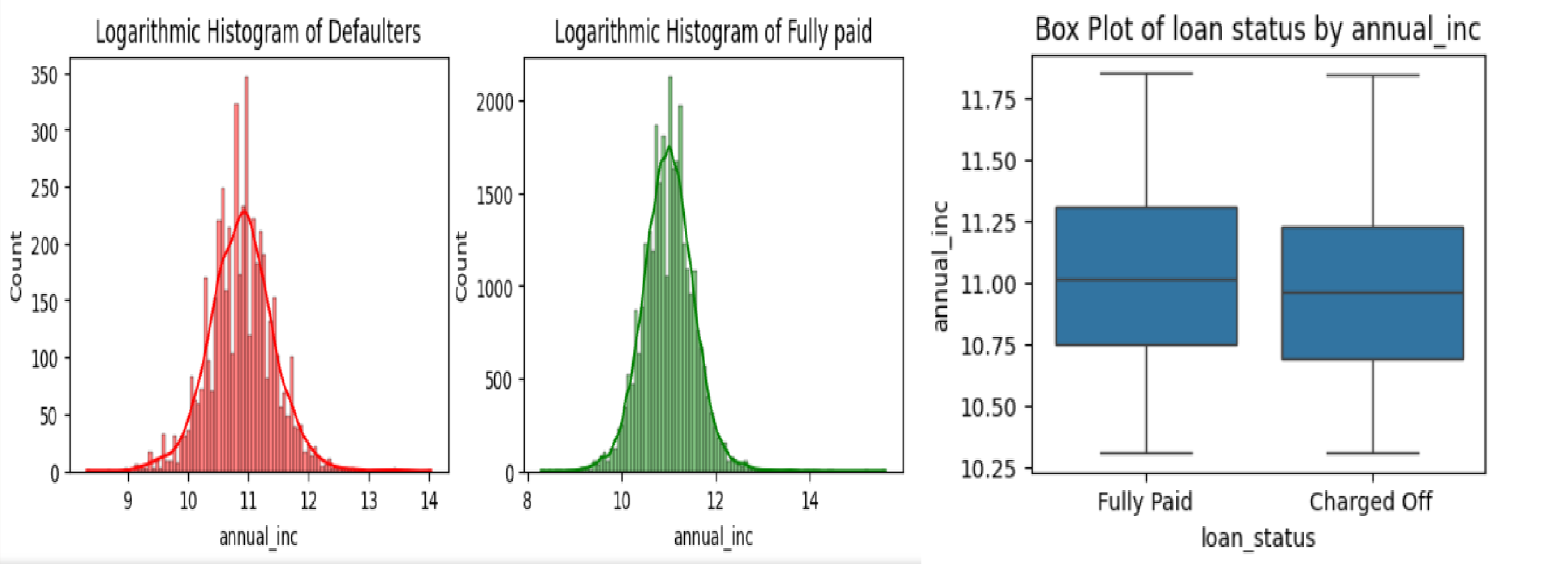

Histogram of Defaulters
Count
300
250
200
150
100
50
0
0 5000 10000 15000 20000 25000 30000 35000
loan_amnt


Histogram of Fully paid
Count
2000
1500
1000
500
0
0 5000 10000 15000 20000 25000 30000 35000
loan_amnt


Box Plot of loan status by loan_amnt
loan_amnt
35000
30000
25000
20000
15000
10000
5000
0
Charged Off
Fully Paid
loan_status

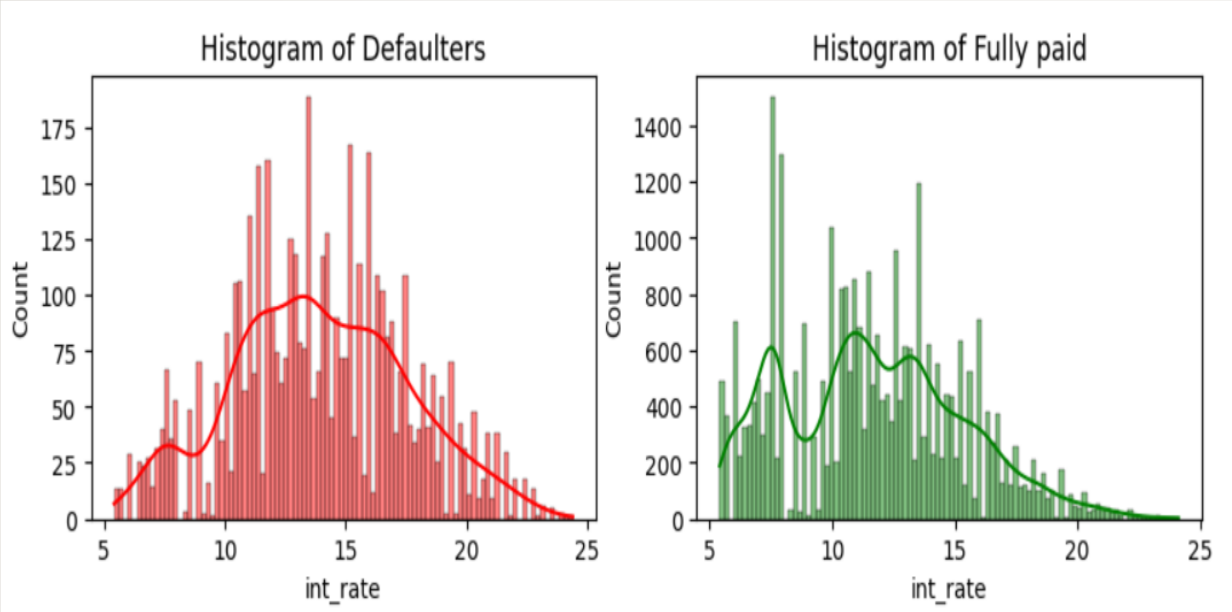
- Lending Club gives loan in the range of 1000 to 35000. We can see that both fully paid and defaulters follow same pattern of withdrawing loans, albeit there is a huge spike in defaulting for 35000 compared to fully paid. From the box plot also , we can see that defaulters usually have higher loan amounts.

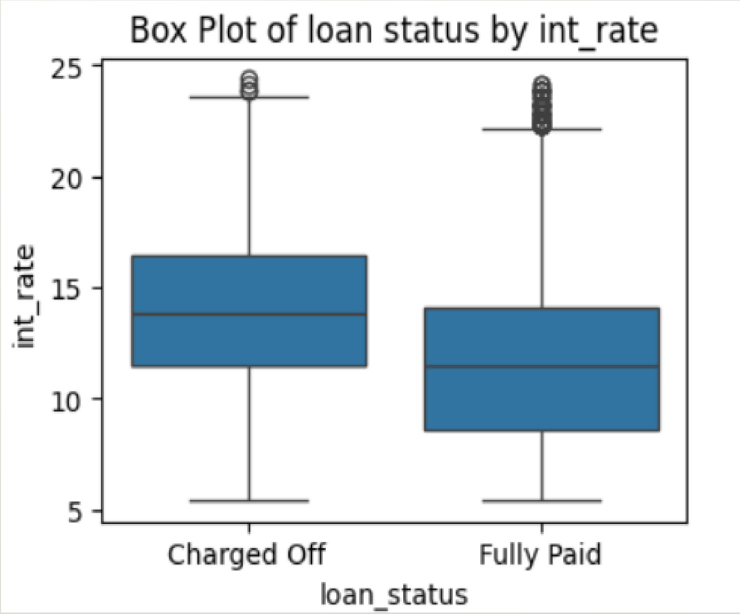
Annual Income



- Annual Income is following a normal distribution for both defaulters and fully paid, with mean somewhere around 11. Defaulters generally tend to have lower median annual income

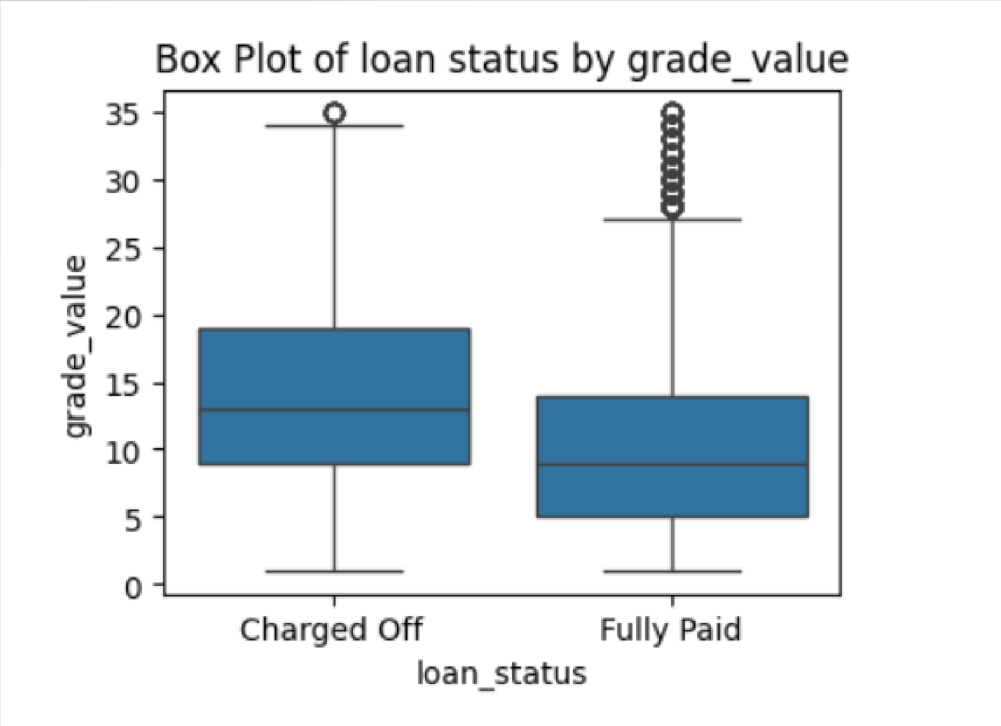
Interest Rate





- 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.

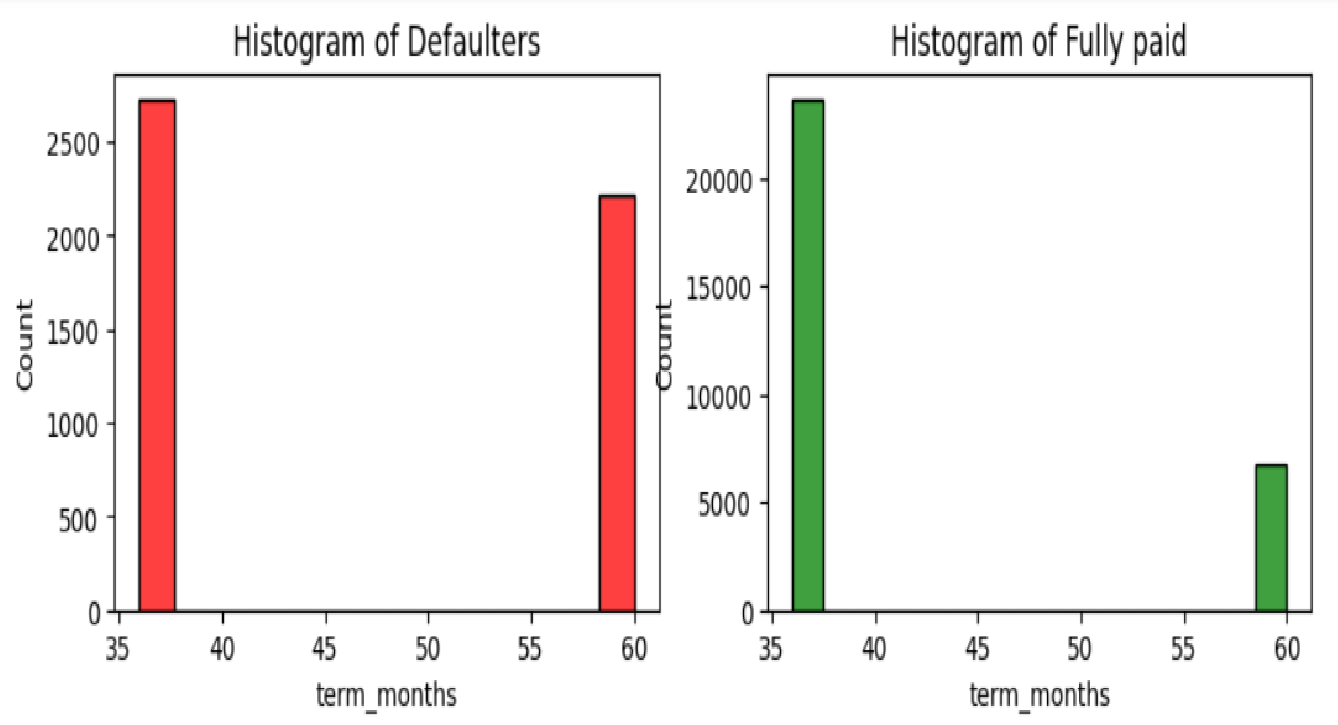
Grade Value



A1	1
A2	2
A3	3
B1	6
B2	7
C1	16

- Loans with lower grade values as follows, i.e. A1=1,A2=2 …C1=16 etc., we observed that when this value is higher, chances of defaulting is high. So C1 is more likely to default compared to A1.

Term Months


Histogram of Defaulters
Count
2500
2000
1500
1000
500
0
35
40
45
50
55
60
term_months
Histogram of Fully paid
Count
20000
15000
10000
5000
0
35
40
45
50
55
60
term_months

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

Recommendations

- Higher loan amounts are riskier than lower loan amounts, need to do more due diligence when processing the same
- People with lower annual income tend to default more.
- Grade values like E,F,G are much riskier compared to A,B,C. These lower grades have higher interest rates, which eventually leads to default.
- Higher loan amounts tend to default more, we should try to process these more carefully.
- Borrower's taking loan for the term 60 months has more chances of default.