

Lending Club Case Study

Submitted By : Sundar Balakrishnan

Sandeep Patel

Problem Statement

This case study is about identifying factor which impacts the loan default from the customers.

Analysis has been performed using python and excel spread sheet for impact assessment of various key attributes like loan term, interest rate, classification of customers etc.

The business ask is to identify the factor impacting the potential default customers (charged off) vis a vis fully paid customers.

So that the Consumer Finance Company can make adjustment/amendment in its policies to avoid potential loss due to loan default .

Approach

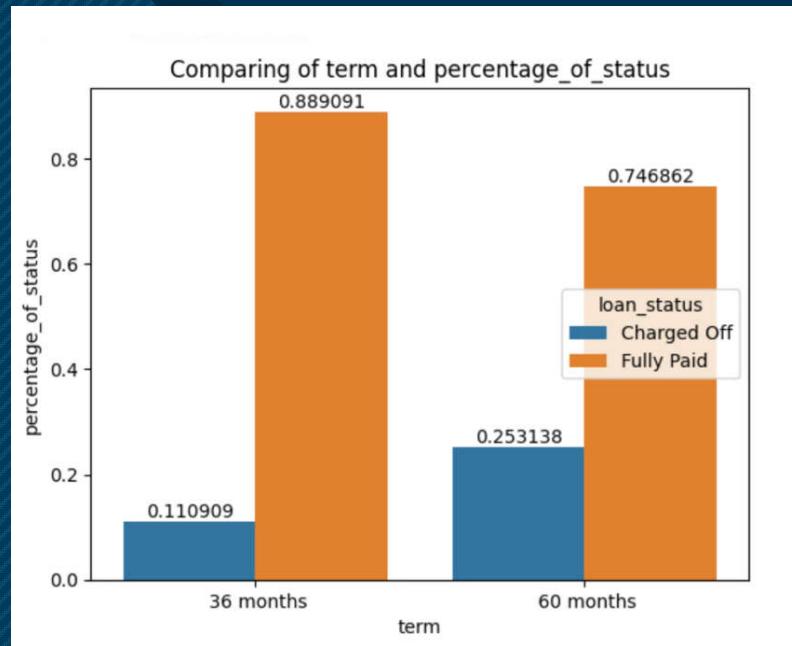
1. We have identified the data fields in the given data sheet. We performed the sanity check around the data, the data types, max values, mean values, Inter quartile range, outliers for each variable.
2. We then checked the fields for null values.
3. Removed fields/columns where all rows are either NA or null.
4. This reduced the analysis fields to around 50 from 100+.
5. We then performed univariate analysis on fields to check the distribution of the fields.
6. Then we checked the relationship between different variable (Bivariate Analysis).
7. Then we shared our finding along with the plots (using python)

Findings

Univariate Analysis

1. Defaults had a strong linking with the “Term” of the loan.

60 Months loan has greater default rate (25%) than 36 months loan (11%)

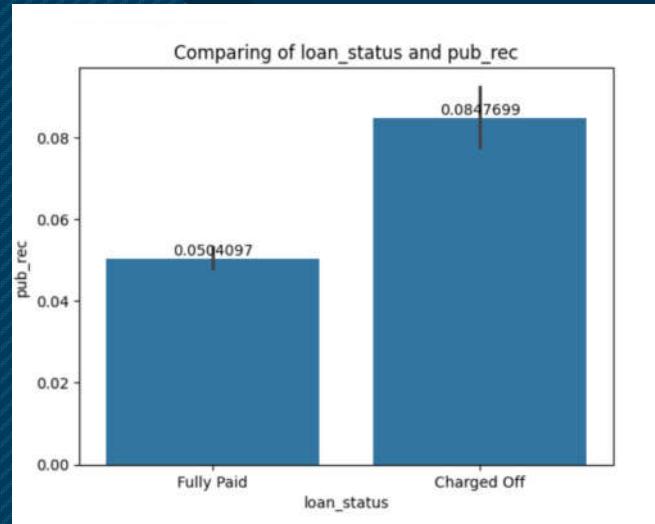


Findings

Univariate Analysis

2."Number of derogatory public records" also had a relation with the default.

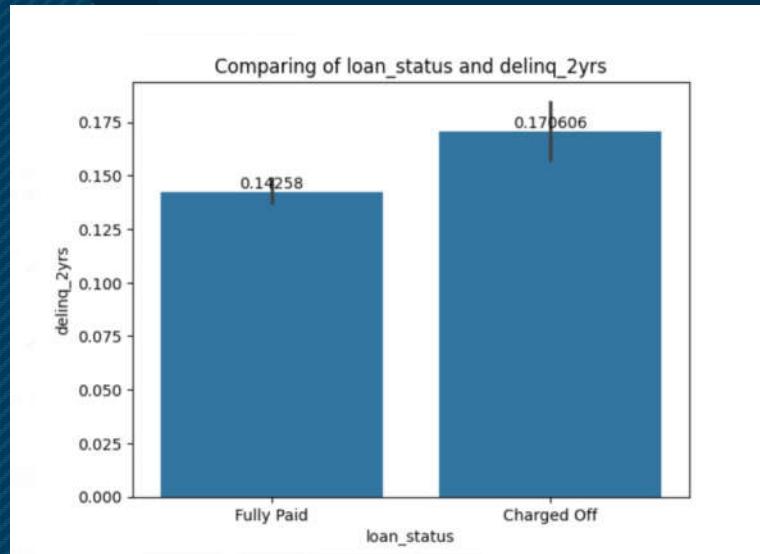
8% of the defaulters were having at least one derogatory public record against them compared with 5% for Fully paid customer and 4 % for current customer.



Findings

Univariate Analysis

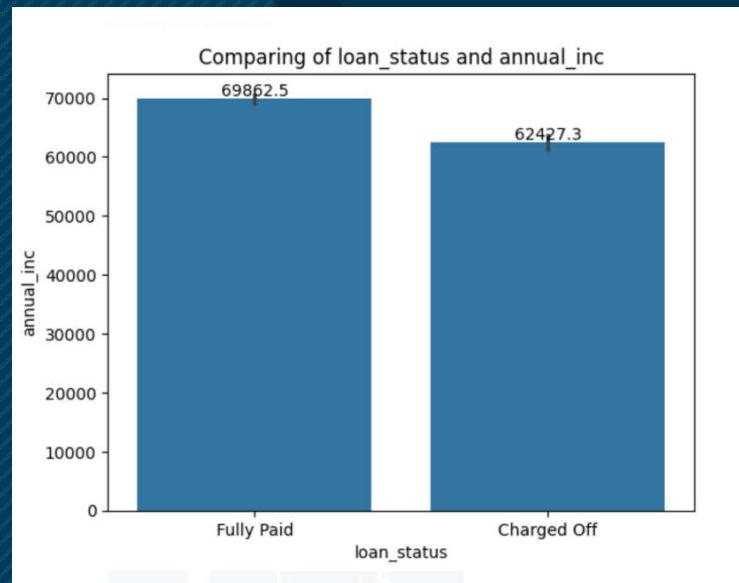
3. The number of "30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years" were more for defaulting customers than fully paid ones



Findings

Univariate Analysis

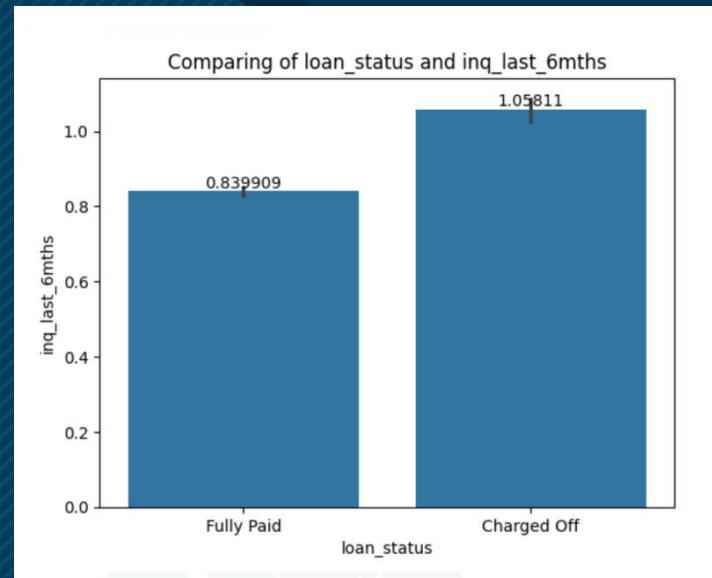
4. Average “Self reported annual income during registration” was low for the default customer than Fully Paid customer (60 K vs. 70 K)



Findings

Univariate Analysis

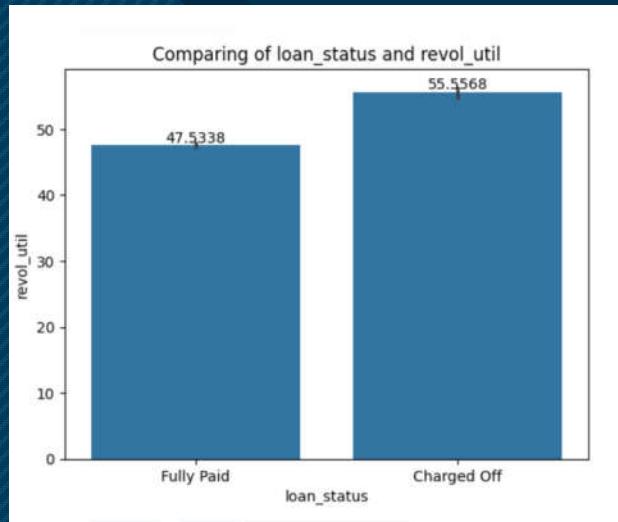
5. “**Inquiry in last 6 month**” are more for defaulted customer compared to Fully paid customers.



Findings

Univariate Analysis

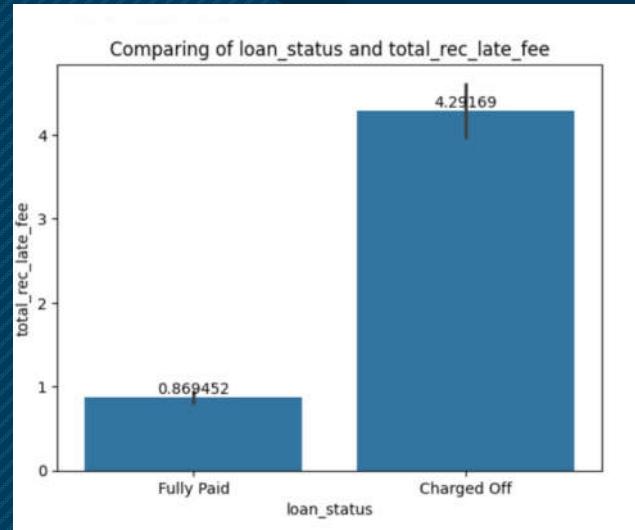
6. “Revolving Credit utilization rate” are more for defaulted customer compared to Fully paid customers.



Findings

Univariate Analysis

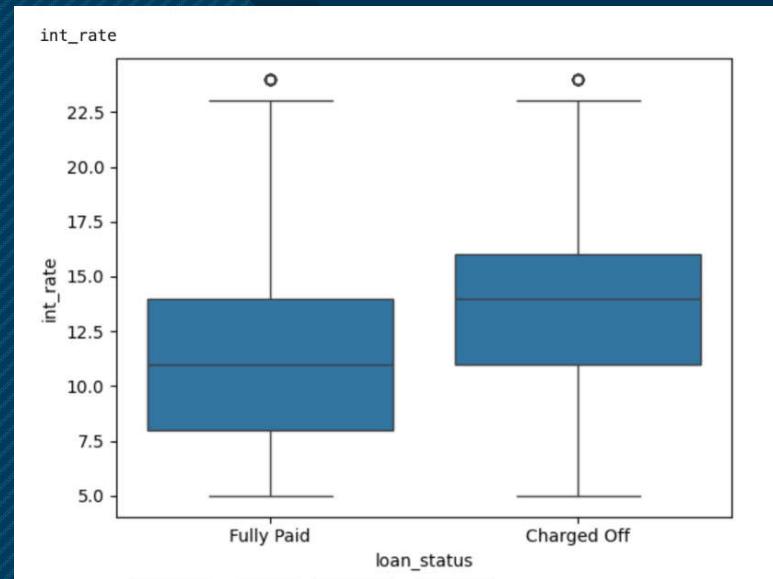
7. “Late fees received to date” are more for defaulted customer compared to Fully paid customers.



Findings

Bivariate Analysis

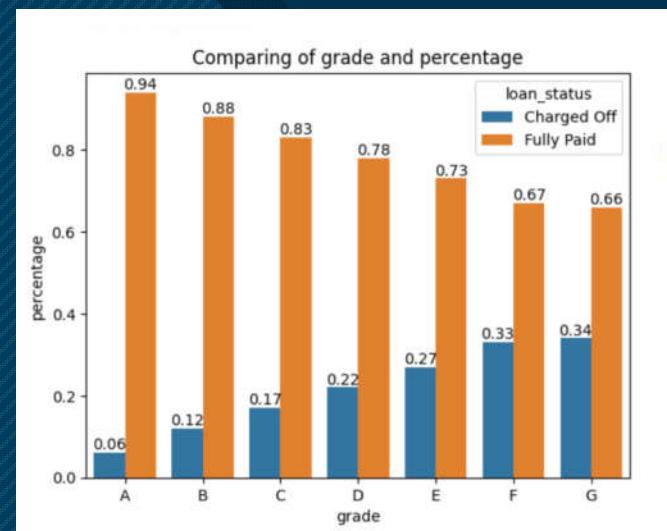
1. Defaulted Customer on an average paid more (14%) than Fully Paid Customer (12%) vs Current Customer (15%)



Findings

Bivariate Analysis

2. Comparing of grade with percentage



Findings

Bivariate Analysis

3. The bank on an average charged greater Interest for the Verified Customer(13%) Vs. Source Verified Vs. (12%)the Non Verified (11%)

4. 62 % of the defaulted customers were verified/source verified vs. 38 % of Non verified.



Thank You!!