Lending Club Case Study

By Ranjith Kanhinpuredathil and Pawan Kumar

CONTENT

- Problem Statement
- Solution Statement
- Approaches
- Observations and Results
- Conclusions
- Recommandations

PROBLEM STATEMENT

When a consumer finance company provides a loan based on the applicant's profile. There are two types of risks associated with the bank's decision:

1. Business loss, when the loan application is rejected.
2. Financial loss, when the customer defaults.

We have to solve these two major problems so that banks/finance companies can give more oans to good customers and reduce the rate of defaults.

SOLUTION

- To solve these two major problems, we will use a loan dataset of past applicants and will analyze using EDA.

- By using EDA, we will try to find some significant variables that will help us to find some patterns so that we can reduce the rate of defaults and increase the rate of good loans

APPROACH

We performed the following steps for our EDA analysis

1) Load required libraries and dataset.
2) Data Cleaning.
3) Apply univariate, bivariate, and multivariate approaches to quantitative and categorical variables to get desired results

DATA CLEANING

We performed the following steps for data cleaning

1) Delete unnecessary rows and columns like blank, header, footer total, subtotal, etc.
2) Missing or Null values, If there are significant missing or null values, we should not include that variable in our analysis and drop it. If there are fewer missing or null values, in this case, we can impute values by preferable methods median for numerical columns and mode for categorical columns
3) We standardized values by unit conversion, precision, removed outliers, and removed extra characters
4) We performed type conversion of variables to our desired type
5) We derived some variables from existing ones according to our requirements

OBSERVATIONS AND RESULTS

We did lots of analysis and observations in our Jupiter notebook. In the next few slides, we are including some significant variables with our observations, insights, and visualizations

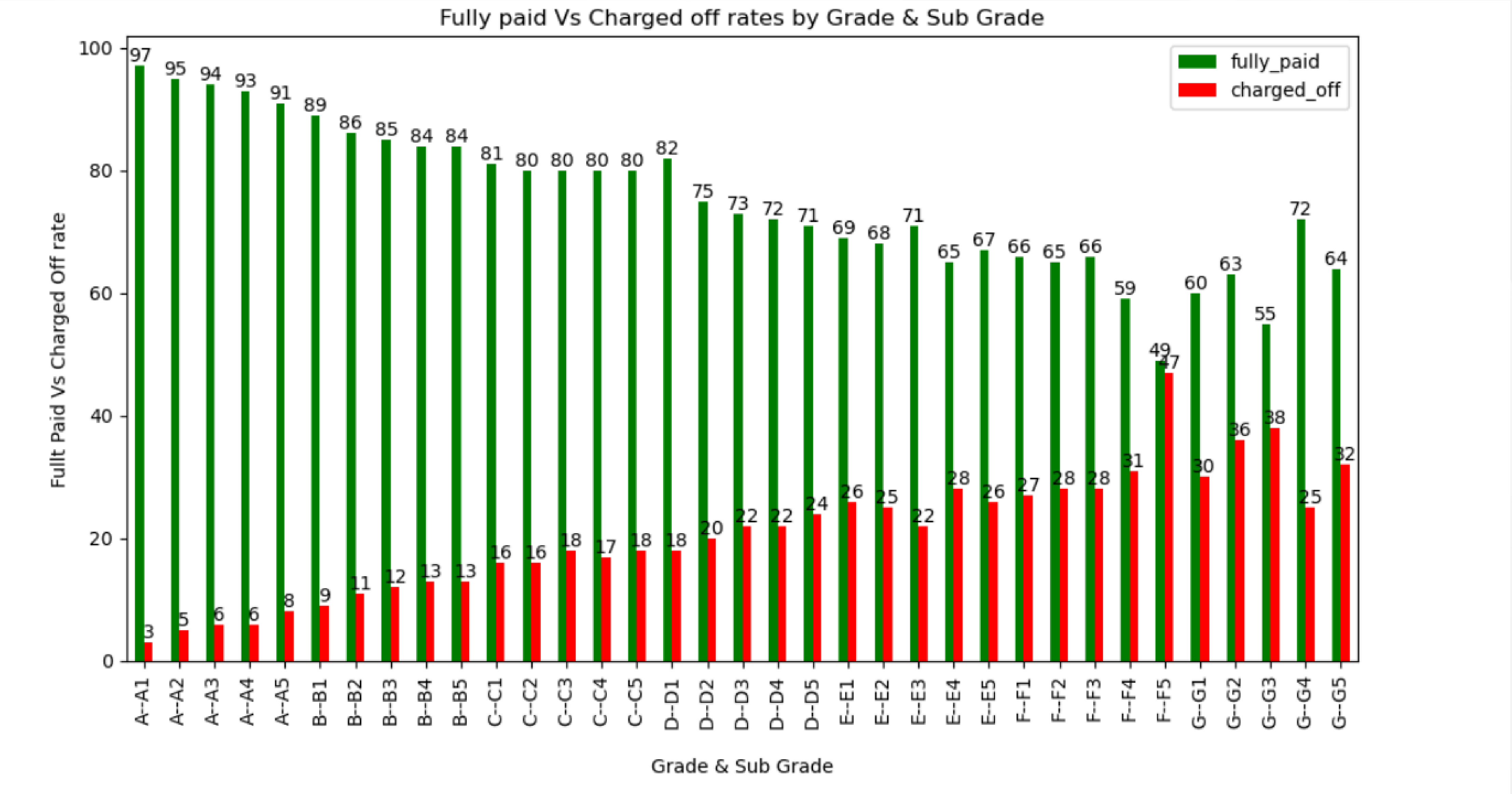
OBSERVATIONS AND RESULTS

Analysis of Fully paid Vs Charged off rates by Grade & Sub Grade

Insights

1. The tendency in general is that the lower the grade-subgrade combination the higher the default rate. For example, A-A1 (higher grade) has a lower default rate than B-B1(lower grade)

2. However the highest default rate is among F-F5 close to 45 percent default rate which is very high.

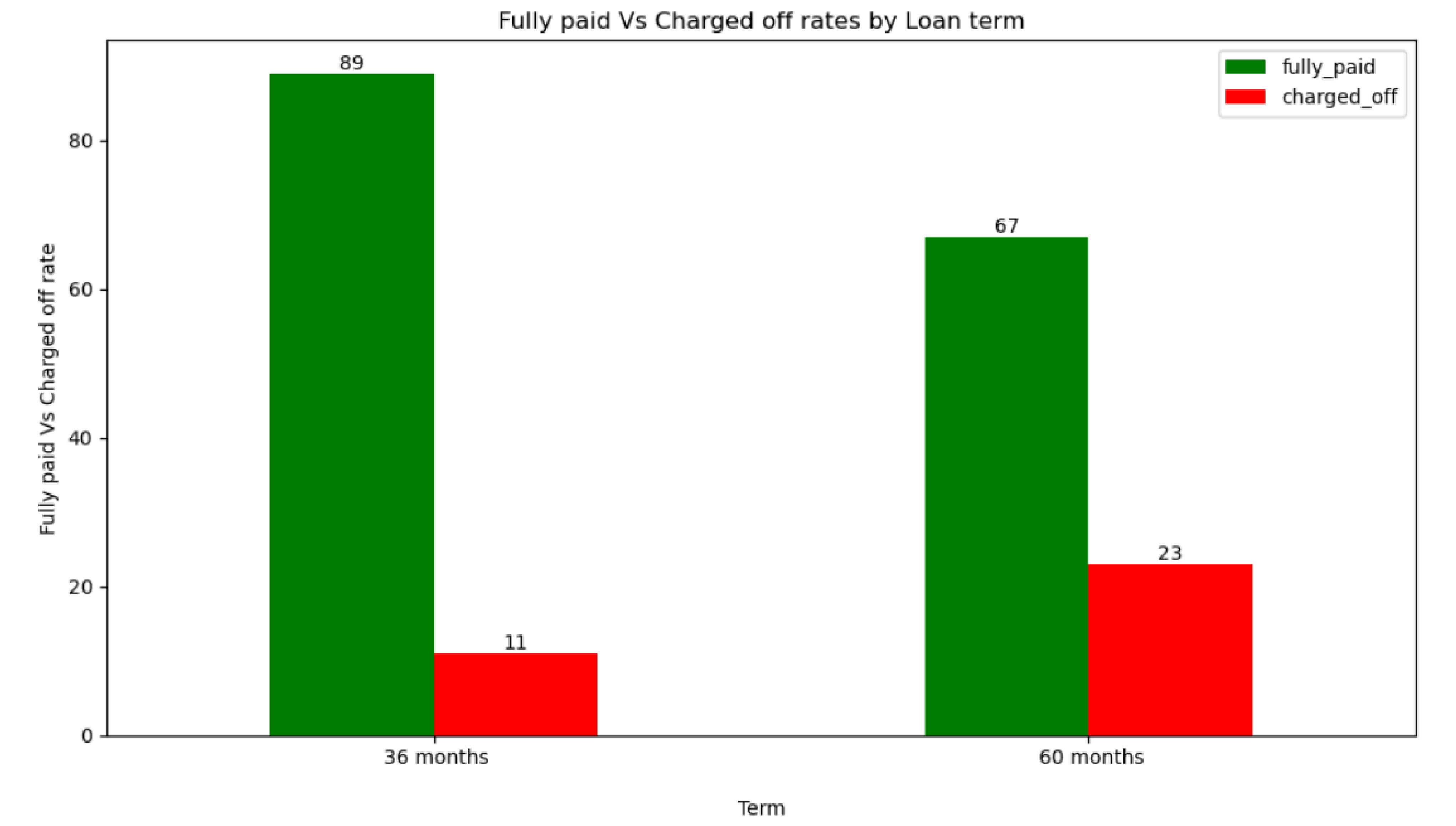


OBSERVATIONS AND RESULTS

Analysis of Fully paid Vs Charged off rates by Loan term

Insights

Loans with a 60-month term has a double default rate than that of a 36-month loan term. This could be due to the amount borrowed being high. The impact of the combination of term and amount can be checked further. For this, we may have to create a loan bucket to limit the number of loan amounts.

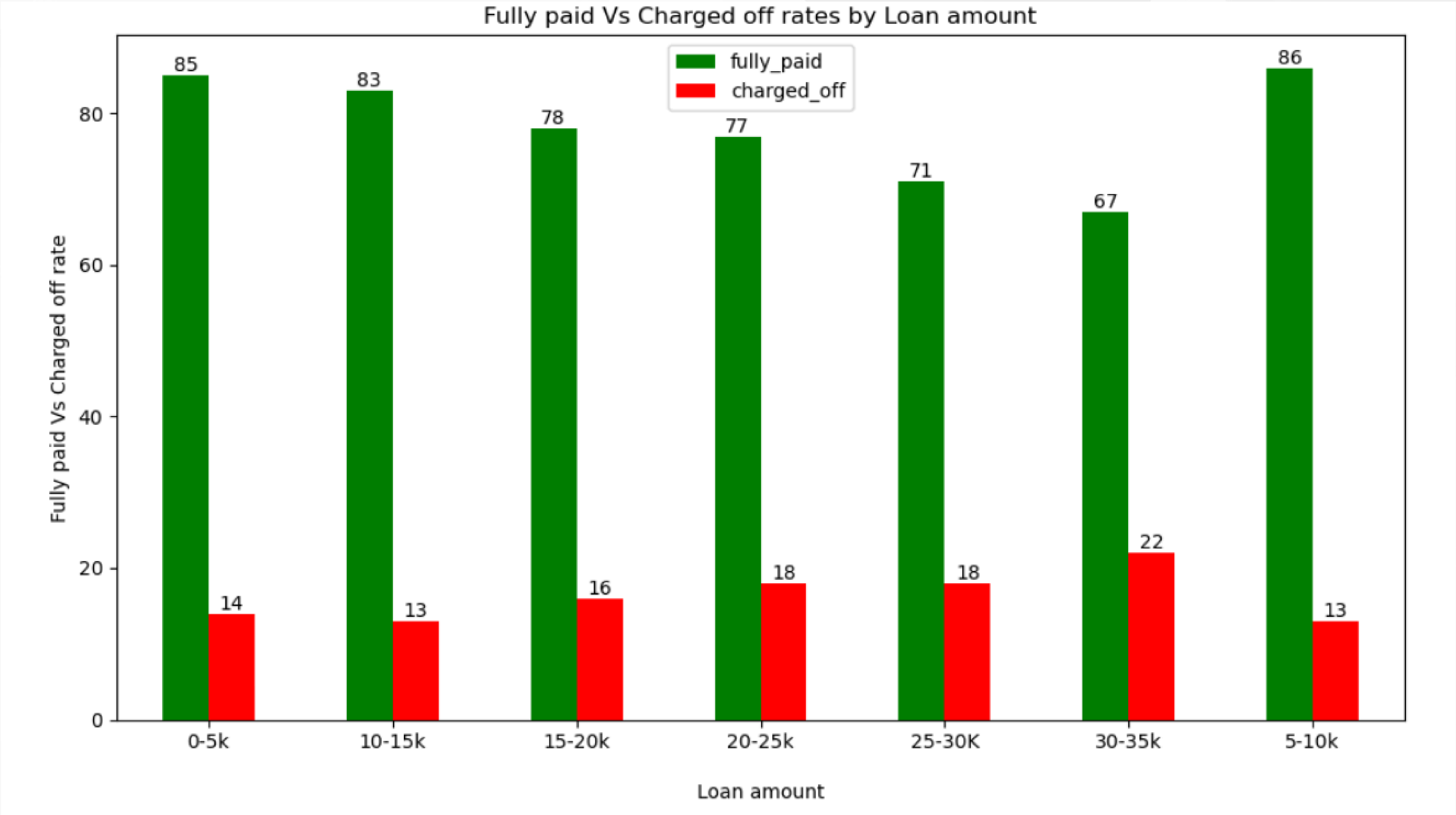


OBSERVATIONS AND RESULTS

Analysis ofFully paid Vs Charged off rates by Loan amount

Insights

1. The loan amount between 5k and 10k has the lowest default rate
2. As the loan amount increases, default rates also increase.

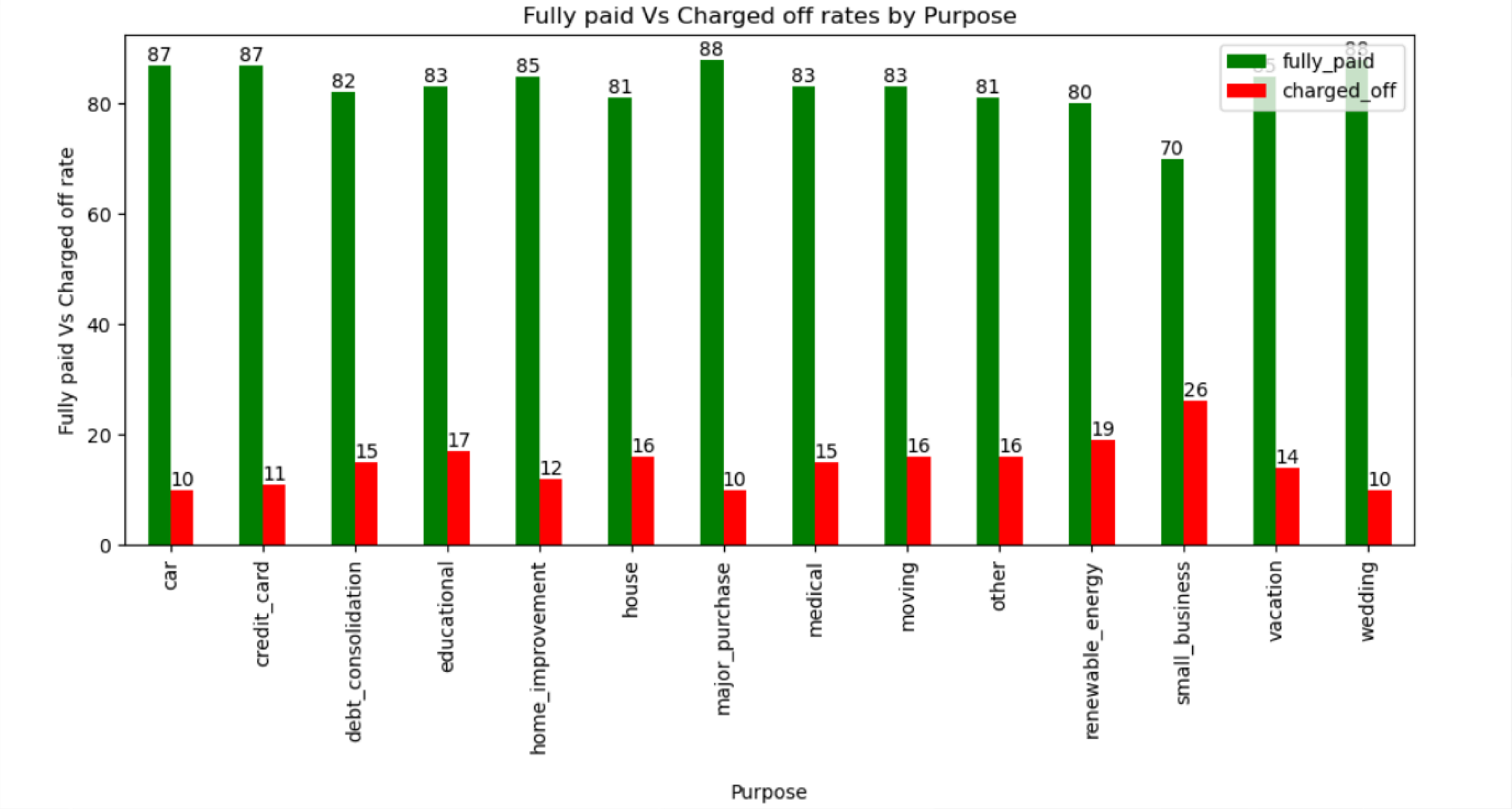


OBSERVATIONS AND RESULTS

Analysis of Fully paid Vs Charged off rates by Purpose

Insights

1. The highest default rate of about 25% is for loans acquired for small businesses.
2. Mostly everything else stays within 15% or lower.
3. Loans availed for educational purposes and renewable energy also go slightly above 15%

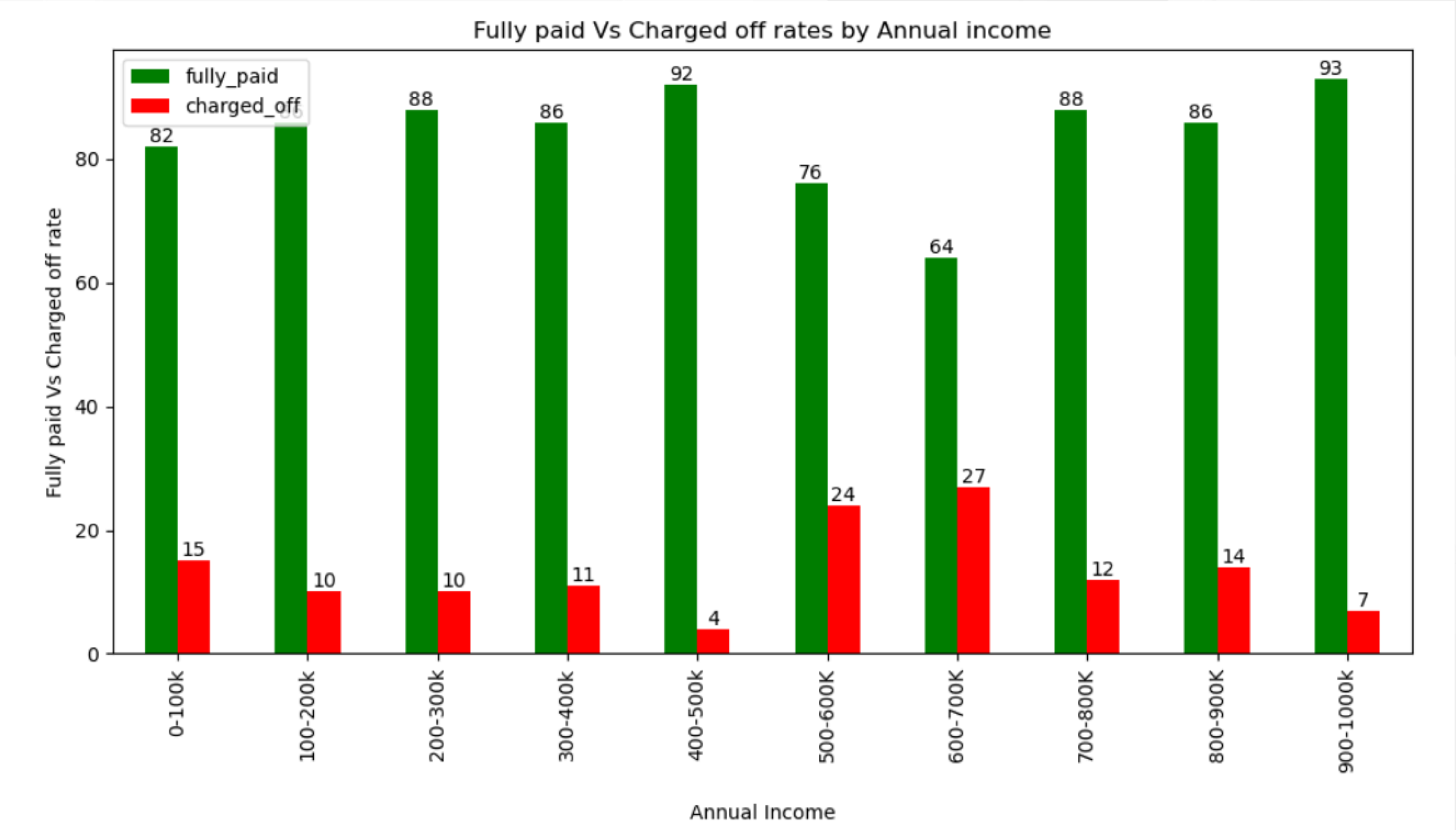


OBSERVATIONS AND RESULTS

Analysis of Fully paid Vs Charged off rates by Annual income

Insights

1. The highest default rate is for people with income levels between 500-600k and 600-700k at close to 25%.
2. The lowest default rate is for people with income levels between 400-500k.

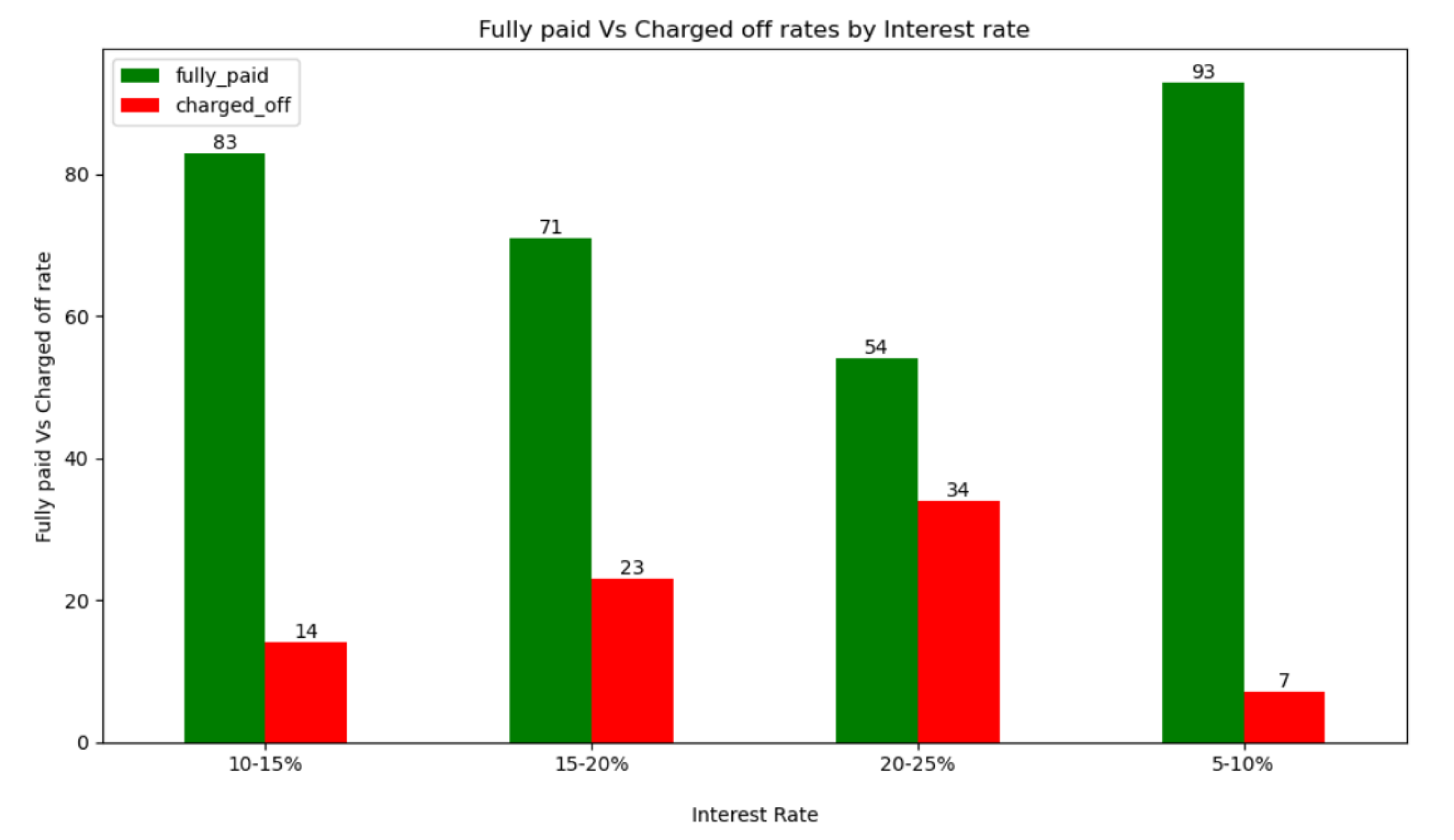


OBSERVATIONS AND RESULTS

Analysis of Fully paid Vs Charged off rates by Interest rate

Insights

Straight forward insight. The higher the interest rate, the higher the loan default rate. A higher interest rate of 20-25% has a very high risk. The loans given at this rate have more than 30% default rate. The loans given at 15 to 20 percent interest rate which has more than 20% default rate.



CONCLUSIONS

After completing our analysis successfully, we are able to find some significant variables that can help a financial company make the right decision when analyzing a loan applicant's application, reduce its default rate, and increase the rate of good loans

We mentioned the significant variables in the next slide

RECOMMANDATIONS

After doing lots of analysis, we found some significant variables that can be used by the risk management team for making further decisions and models

We found mentioned below variables significant

1) Grade and Sub Grade (grade and sub_grade)
2) Loan Term (term)
3) Address State (addr_state)
4) Purpose (purpose)
5) Annual Income (annual_inc)
6) Interest Rate (int_rate)
7) Loan Amount (loan_amnt)
8) Public Record bankruptcies (pub_rec_bankruptcies)
9) Public Records (pub_rec)
10) Inquiries in past 6 months (inq_last_6mths)
11) Revolving line utilization rate (revol_uti)