Presented By : Ashish & Uday
Date:23/07/2024

Overview-Lending club case study

The primary objective of this analysis is to identify loan defaulters and conduct a comprehensive risk analysis to understand the factors contributing to loan defaults.

The analysis utilizes Lending Club loan data from [specific time period, e.g., 2007-2011], focusing on variables such as loan amount, interest rates, borrower's annual income, loan grade, and loan status. Our goals are to distinguish the characteristics of defaulters from non-defaulters, identify high-risk borrowers, and develop a predictive model to forecast potential loan defaults.

Data Description

Data Collection and Initial Overview:

Dataset contains the complete loan data for all loans issued through the time period 2007 t0 2011.

📄 Loan Data Set   📄 Data Dictionary

Data Cleaning and Preprocessing

- Handling Missing Values
- Encoding Categorical Variables.
- Columns which are above 60% of missing values.

Data Understanding

Breaking Attributes as per following types:

1. **Consumer Attributes**: Column provides customer information.

   emp_title, emp_length, home_ownership, annual_inc, verification_status, zip_code, addr_state, dti, delinq_2yrs, earliest_cr_line, inq_last_6mths, open_acc, pub_rec, total_acc, pub_rec_bankruptcies

2. **Loan Attributes**: Column provides loan information.

   loan_amnt, funded_amnt, funded_amnt_inv, term, int_rate, installment, grade, sub_grade, issue_d, loan_status, desc, purpose, title, revol_bal, revol_util, out_prncp, out_prncp_inv, total_pymnt, total_pymnt_inv, total_rec_prncp, total_rec_int, total_rec_late_fee, recoveries, collection_recovery_fee, last_pymnt_d, last_pymnt_amnt, last_credit_pull_d

Data Analysis

Following are strong driver variables:

1. Grade
2. Interest Rate
3. Loan Term
4. Home Ownership
5. Purpose of loan
6. Verification Status
7. pub_rec (Number of derogatory public records)
8. inq_last_6mths (The number of inquiries in past 6 months)
9. total_rec_prncp (Principal received to date.)

Exploratory Data Analysis (EDA)
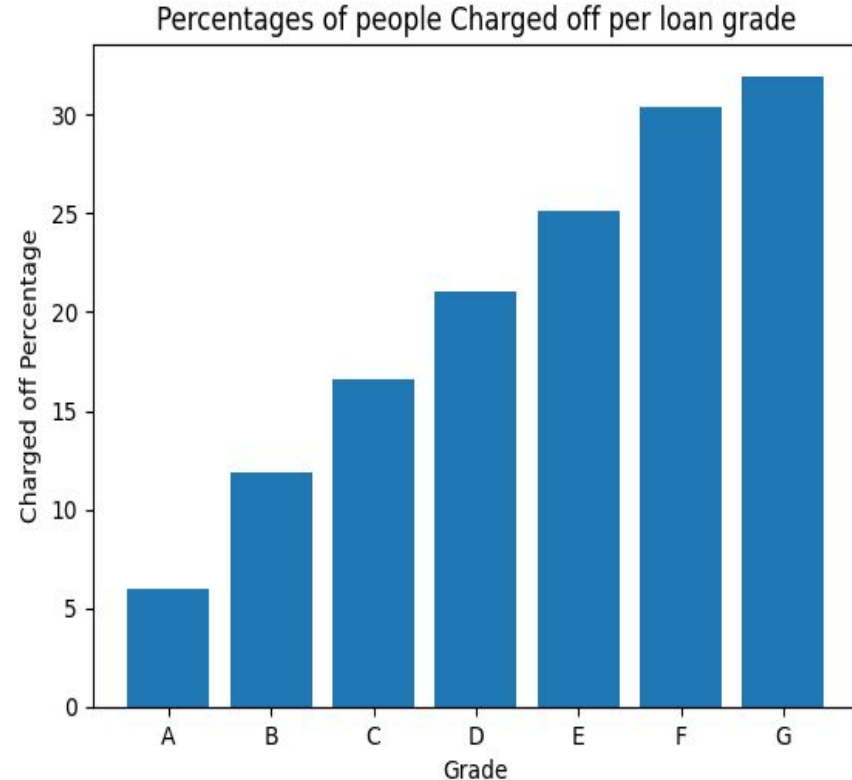
Purpose of analysis:

To analyze the distribution of loan grades, among charged-off loans.

Key Variables:

- Grade(loan attributes): LC assigned loan grade.
- loan_status(loan attributes): Current status of the loan.

Conclusion:

As grade increase chances of "Charged off" increases.Grade and loan status has strong relation.

Percentages of people Charged off per loan grade

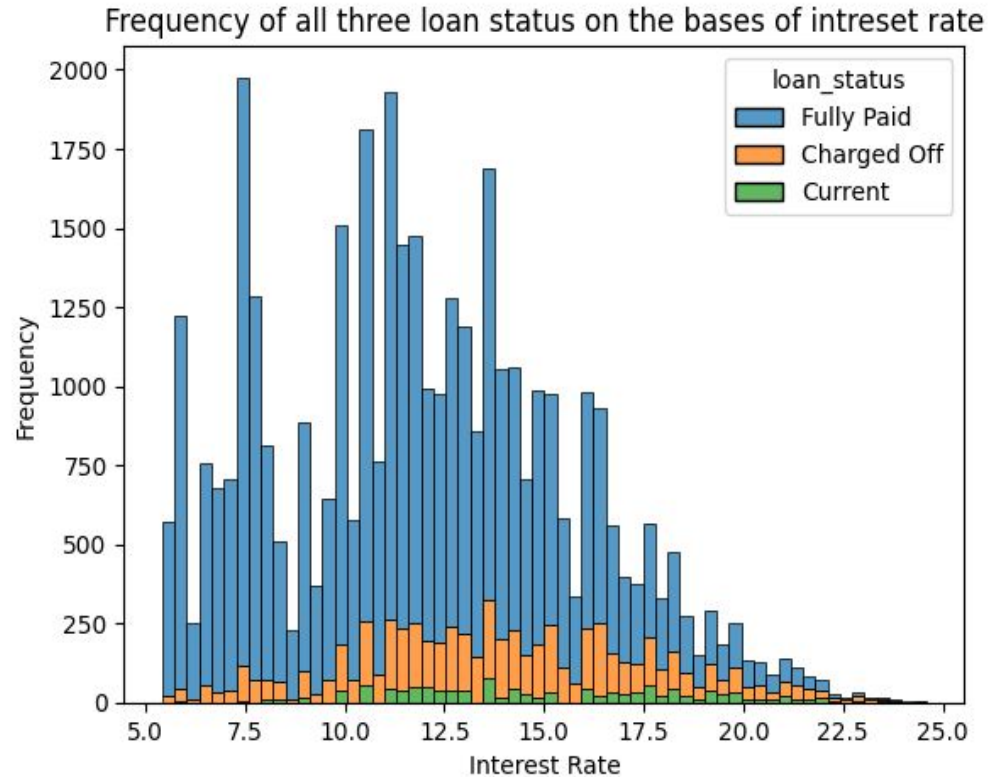# Exploratory Data Analysis (EDA)

Purpose of analysis:

To analyze the distribution of loan grades, among charged-off.

Key Variables:

- int_rate(loan attributes): Interest Rate on the loan
- loan_status(loan attributes): Current status of the loan.

Conclusion:

Probability of "Charged off" increases as interest rate increases.

Frequency of all three loan status on the bases of intreset rate

Exploratory Data Analysis (EDA)

Purpose of analysis:

Analysing effect of loan term on charged off loans. Also the graph shows

we do not have any "current" in 36 month period.
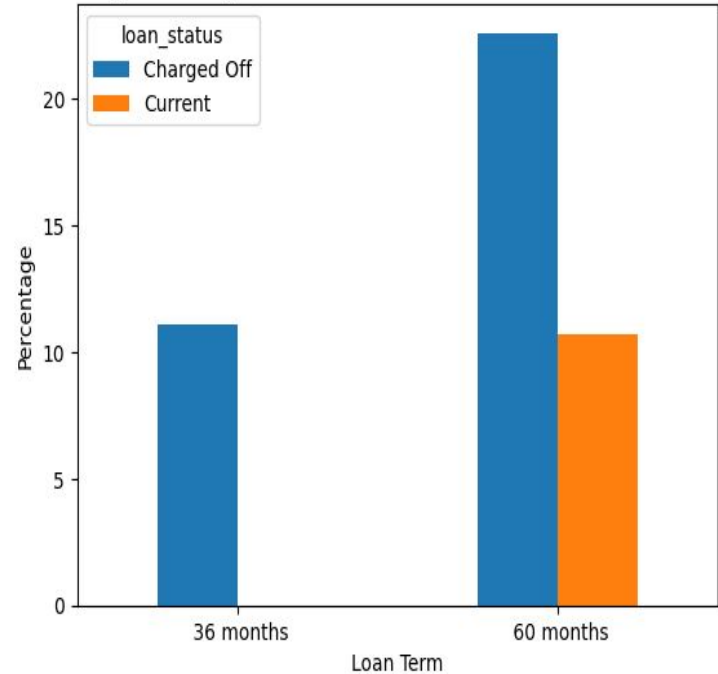
Key Variables:

- term(loan attributes): The number of payments on the loan.

  Values are in months and can be either 36 or 60.

- loan_status(loan attributes): Current status of the loan

Conclusion:

Chances of Charged off is high for 60 month loan period.

Also there is no current loan in 36 month term.

Percentage of Charged off and Current loans on the bases of loan term

Exploratory Data Analysis (EDA)

Purpose of analysis:

Analysing effect of home ownership on charged off loans.

Key Variables:

- Home_ownership(consumer attributes):

  The home ownership status provided by the borrower

  during registration.

  Our values are: RENT, OWN, MORTGAGE, OTHER.

- loan_status(loan attributes): Current status of the loan

Conclusion:

Other' home ownership type has more change of get "Charged off".

Percentage of Charged off and Current loans on the bases of Home ownership
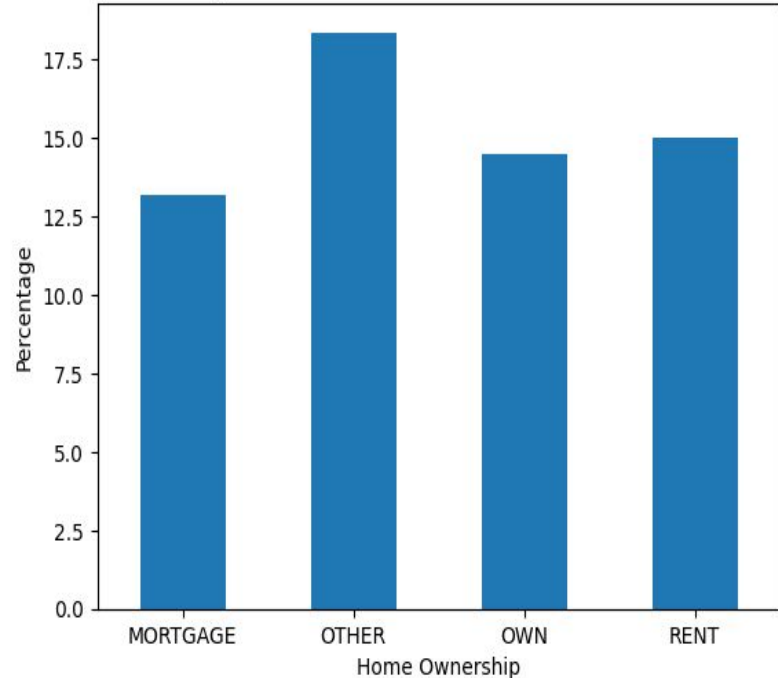
Exploratory Data Analysis (EDA)

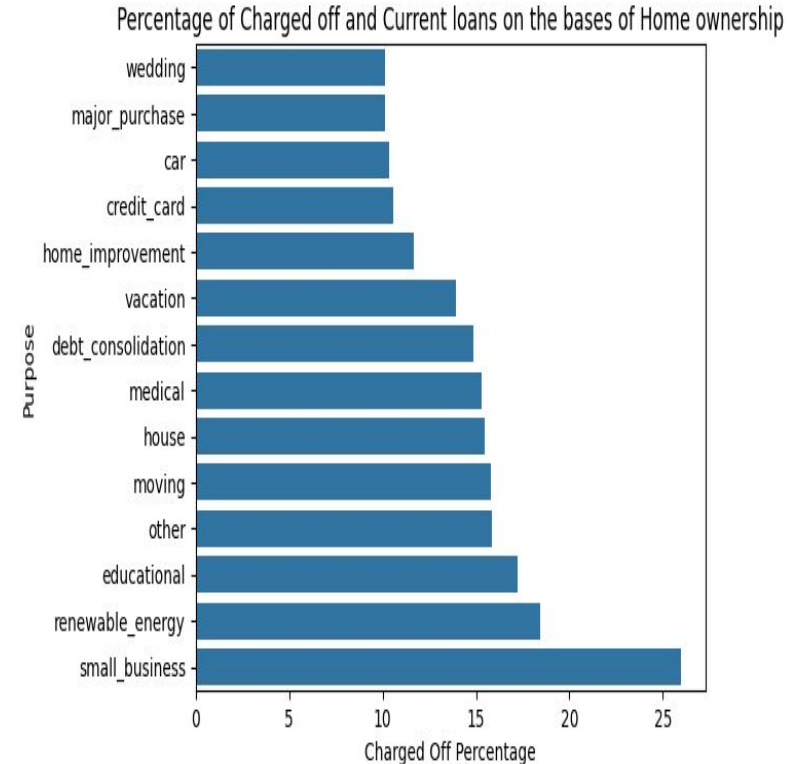**Purpose of analysis:**

Analysing effect of purpose of loan mentioned on

charged off loans.

Key Variables:

- purpose(consumer attributes): A category provided

  by the borrower for the loan request.

- loan_status(loan attributes): Current status of the loan.

Conclusion:

Small Business are more prone to "Charged Off".

Percentage of Charged off and Current loans on the bases of Home ownership

Exploratory Data Analysis (EDA)

Purpose of analysis:

Analysing effect of verification status and derogatory.

public records on charged off loans.

Key Variables:

- verification_status(consumer attributes):
- pub_rec(consumer attributes): Number of derogatory public records.
- loan_status(loan attributes): Current status of the loan.

Conclusion:

Source verified has more number of derogatory records.   Which leads to "Charged off" loans.

Percentage of Charged off loans on the bases of Verification Status and Derogatory Records
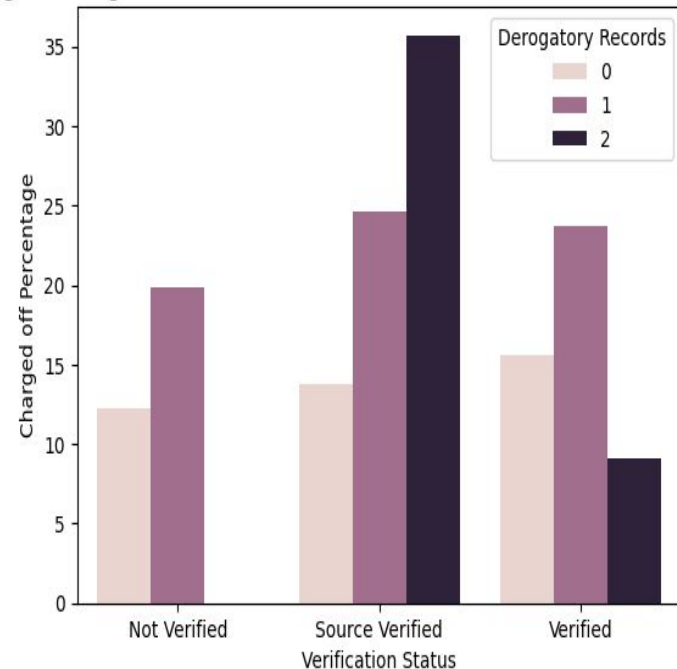
Exploratory Data Analysis (EDA)

Purpose of analysis:

Analysing effect of loan inquiries in past 6 months (excluding auto and mortgage inquiries) on charged off loans.
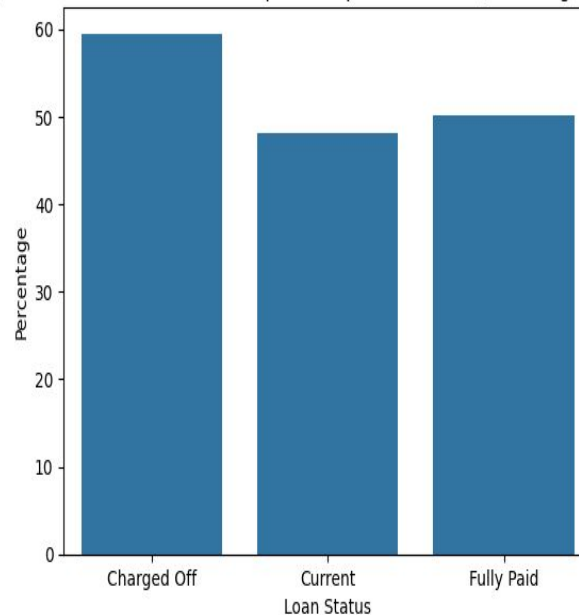
Key Variables:

- inq_last_6mths(consumer attributes): The number of inquiries in past 6 months (excluding auto and mortgage inquiries)
- loan_status(loan attributes): Current status of the loan.

Conclusion:

People having "Charged Off" loan tend to do more loan inquiries in last 6 months.

Percentage of Charged off on the bases of loan inquiries in past 6 months (excluding auto and mortgage inquiries)

Exploratory Data Analysis (EDA)
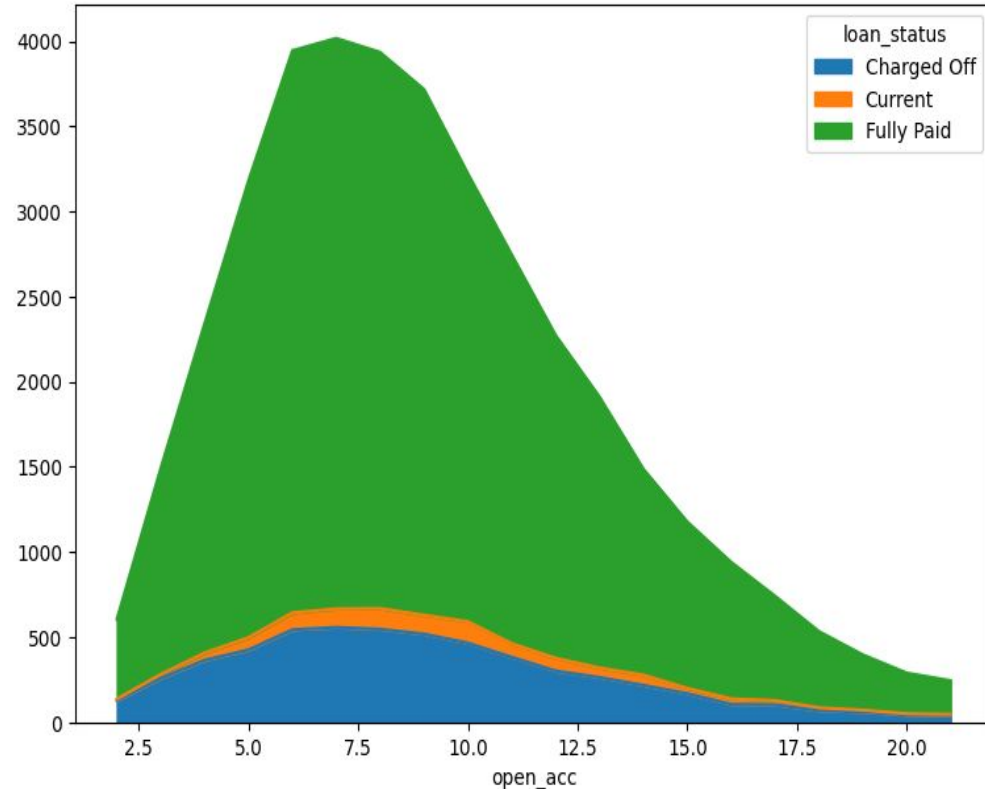
Purpose of analysis:

Analysing ratio of monthly installments and income on charged off loans

Key Variables:

- open_acc(consumer attributes): The number of open credit lines in the borrower's credit file.
- loan_status(loan attributes): Current status of the loan.

Conclusion:

Open Credit line has no correlation with charged off loans

Exploratory Data Analysis (EDA)

Purpose of analysis:

Analysing borrower's earliest reported credit line month on charged off loans.

Key Variables:

- earliest_cr_line(consumer attributes): The month the borrower's earliest reported credit line was opened.
- loan_status(loan attributes): Current status of the loan.

Conclusion:

As per graph charged off loan increases with volume of the loans in each year. The borrower's earliest reported credit line open year do not have direct correlation with loan status.

Infomations about loan status on the bases of the borrower's earliest reported credit line was opened year
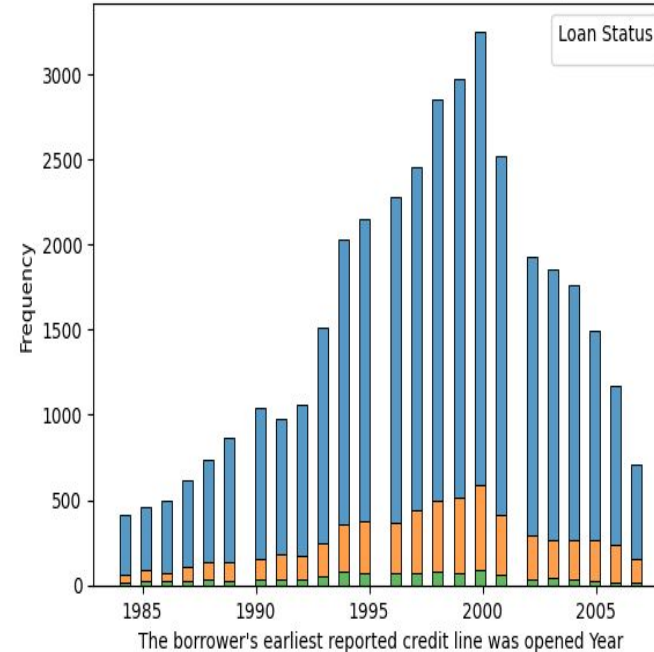
Exploratory Data Analysis (EDA)

Purpose of analysis:

Analysing at principal outstanding charged off loan installments get stopped.

Key Variables:

- Total_rec_prncp(loan attributes): Principal received to date.
- funded_amnt(loan attributes): The total amount committed to that loan at that point in time.

Conclusion:

Mosty charged off loan installments stops at initial tenure of the loans.

Infomations about loan status on the bases of percentage of principle out standing

Frequency vs Percentage of principle out standing
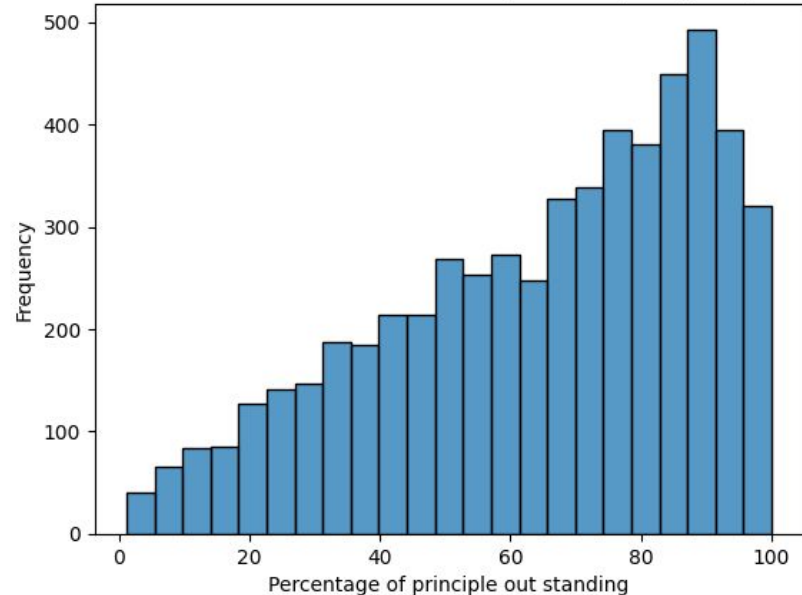
Conclusion

Summary of Key Findings:

Example: "We observed that loan grades have a strong correlation with the likelihood of a loan being 'Charged Off'. Higher grades (A, B) have lower charge-off rates, while lower grades (F, G) exhibit higher charge-off rates."

Insights and Interpretations:

- Example: "The analysis suggests that loan grade is a critical factor in assessing the risk of loan defaults. This insight can be leveraged to improve loan approval processes and risk management strategies."

Implications:

- Example: "By understanding the relationship between loan grades and charge-off rates, lenders can make more informed decisions, potentially reducing the overall default rate and improving portfolio performance."

EDA has provided valuable insights into the factors influencing loan defaults. By continuously exploring and analyzing data, we can make data-driven decisions that enhance our understanding and management of loan risks."