

CREDIT LOAN ASSESSMENT

Rakamin Academy
Project Based Internship
ID/X Partners



READ MORE 



Here's my LinkedIn

About me

ALEISYA ZAHARI SALAM

A Pamulang University student majoring in Mathematics, inspiring to become a data scientist. Having prior experience in various projects such as predicting customer churn and classifying images.



COMPANY BACKGROUND

A lending company is facing challenge streamlining the loan acquisition process to ensure efficiency and swiftness in handling customer applications.

As a Data Science Intern at ID/X Partners, we aim to leverage data analysis and modeling techniques to revolutionize the way we handle loan applications.

By harnessing data, we'll develop predictive models that accurately assess creditworthiness, enabling us to expedite the evaluation process and anticipate potential risks.



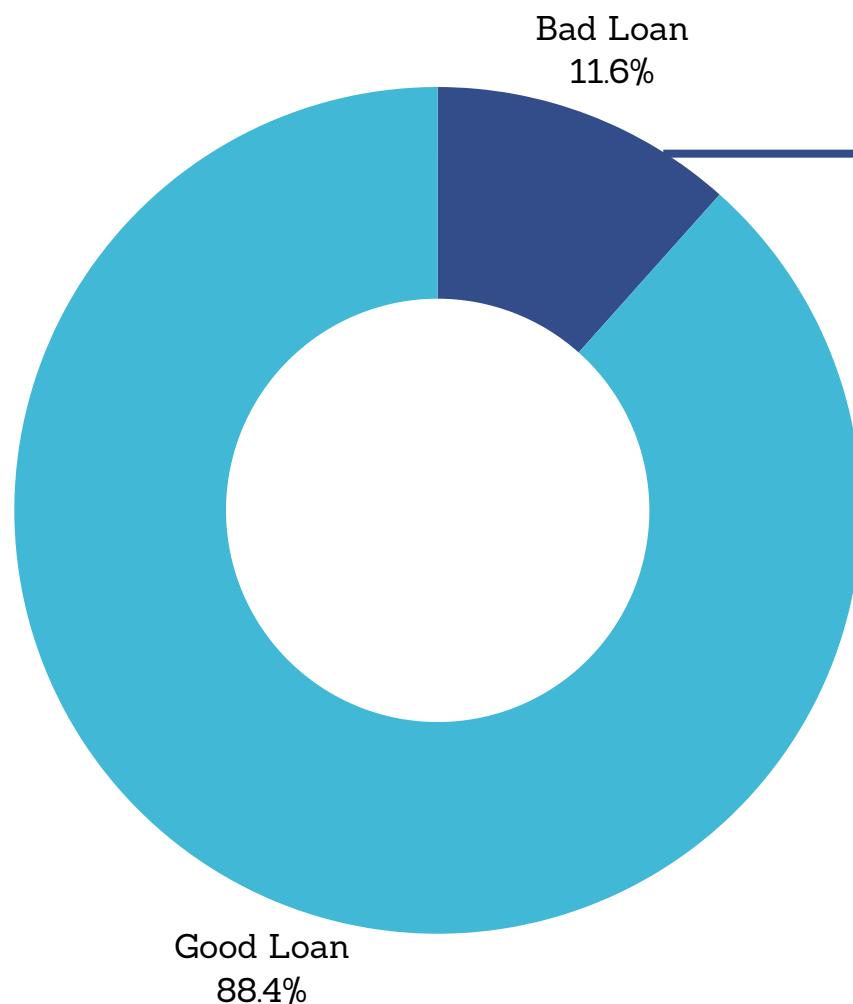
BUSINESS **UNDERSTANDING**

● PROBLEM STATEMENT

● GOALS, OBJECTIVE

Problem Statement

1. Bad Loan

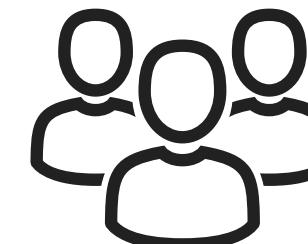
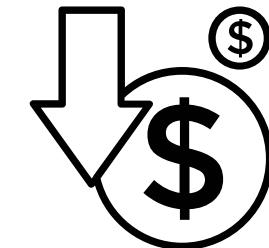


11.6%
466k
are classified as
bad loan

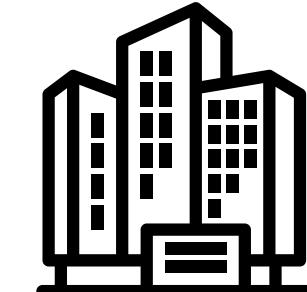
Out of
customer
classified as
bad loan

Impact?

Financial loses



Increase risk
exposure

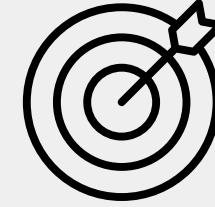


Reputation Damage

Goals & Objective

STRATEGIES ARE NEEDED TO PREVENT
BAD LOANS

Goals



Identifies factors
that influence bad
loans

Objective



1. Deep analytics (EDA)
2. Build model machine learning classification





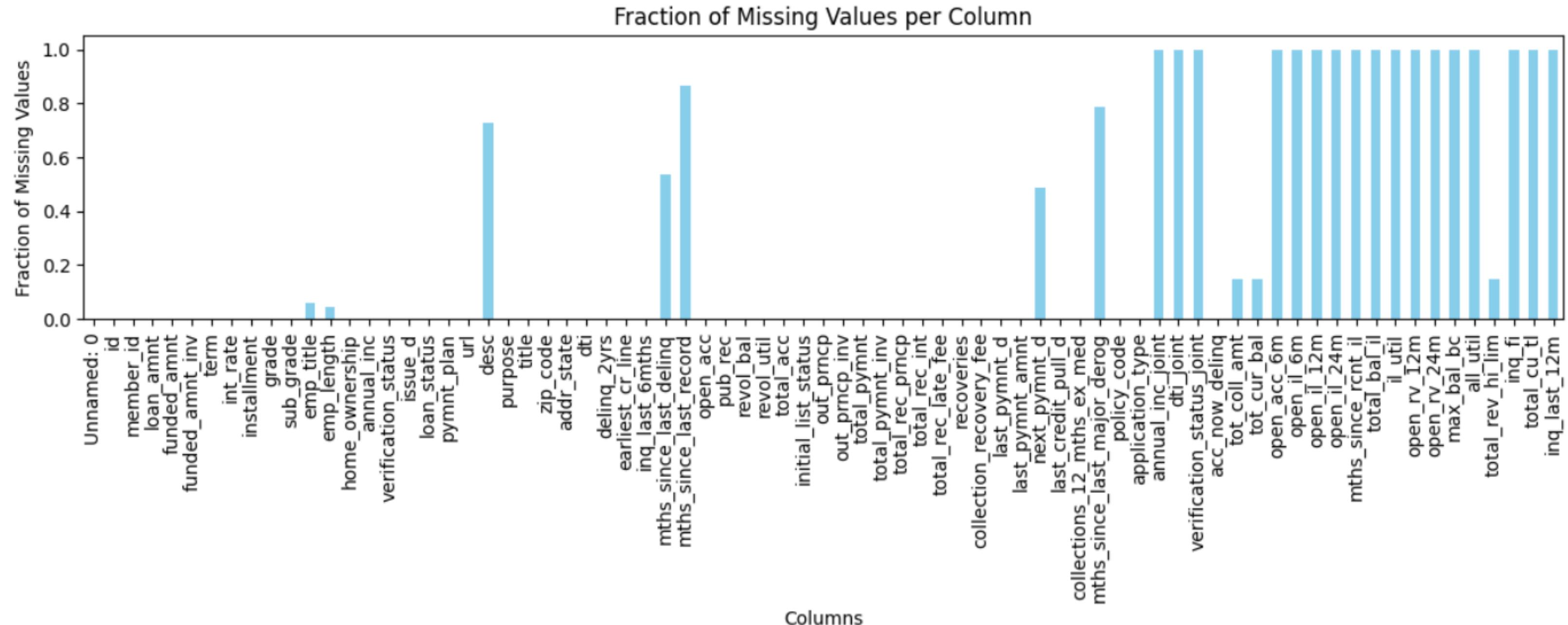
DATA **PREPROCESSING**



DATA CLEANSING

Data Cleansing

missing value



Remove columns that have missing values exceeding 80%, except for the last_pymnt_d column, which will be retained for feature engineering purposes.

Data Cleansing

1. Check Duplicate
2. Remove unnecessary column
3. Handle feature that have correlation between independent features
4. handle feature that didn't have correlation to target





FEATURE **ENGINEERING**

FEATURE ENGINEERING

- 1 Change data type for date
- 2 Adding new feature based from existing features

Column pymnt_time = the number of month between next_pymnt_d and last_pymnt_d

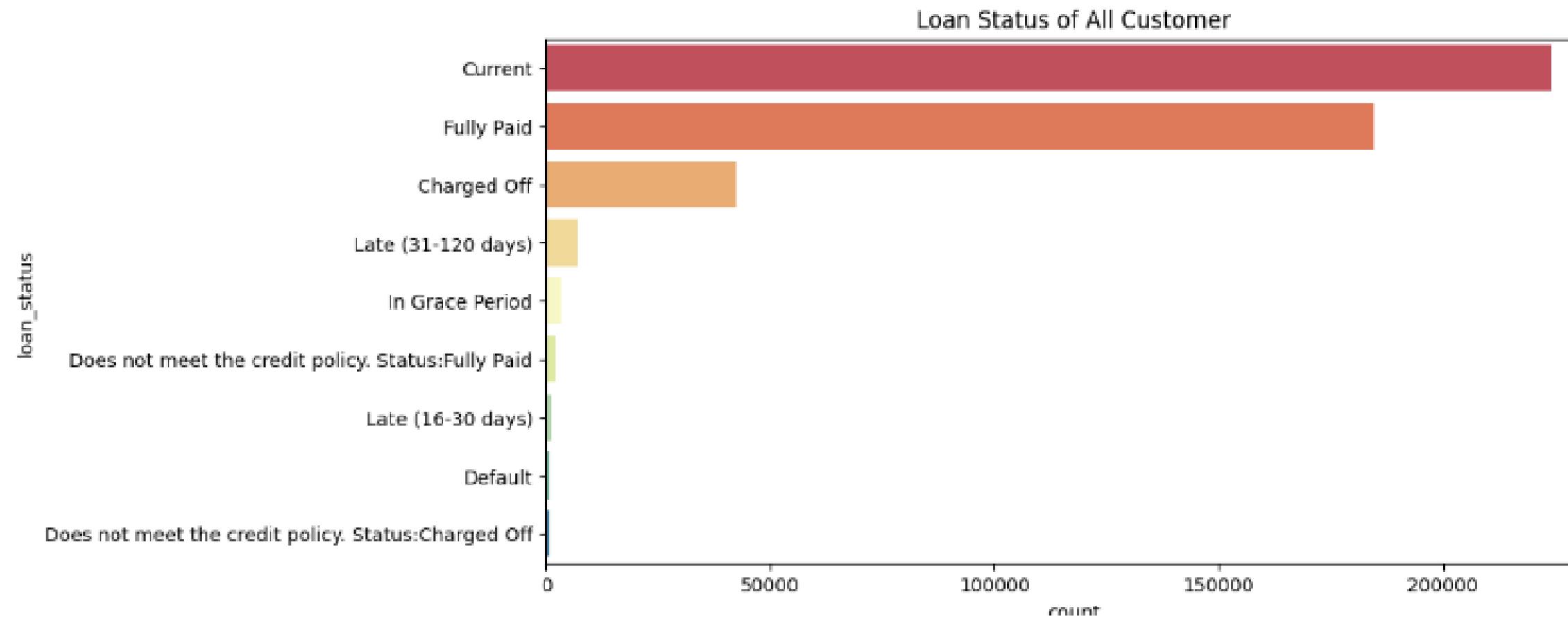
Column credit_duration_year = the number of year between earliest_cr_line and last_credit_pull_d

- 3 Change feature term
In further analysis, the term column consisting of '36 months' and '60 months' will be transformed into '36' and '60' only.
- 4 Feature selection using correlation, biserial correlation, weight of evidence and information value

FEATURE ENGINEERING

4

Check Target value (loan_status)



- Between 9 unique value in `loan_status`, `Current`, `Fully Paid`, and `In Grace Period` are defined as good loan.
- Meanwhile, bad loans will be defined as other than the mentioned good loans.
- For the model, the feature will be devided into two group namely '`good loan`' that will be represents 1, and '`bad loan`' with 0

FEATURE ENGINEERING

6

Feature encoding for the model with one hot encoding and label encoding

Feature that we will be encode are categorical ordinal and nominal





MODELING AND EVALUATION

Modeling and Evaluation: Machine Learning Techniques



Decision Tree

1. Does not require that the data set follow a normal distribution.
2. Robust to outliers

```
▶ dt = DecisionTreeClassifier()  
    dt.fit(x_train, y_train)
```

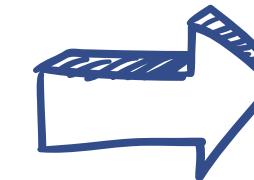
```
⇒ ▾ DecisionTreeClassifier  
    DecisionTreeClassifier()
```



Modeling and Evaluation: Model Evaluation Parameters



ROC-AUC

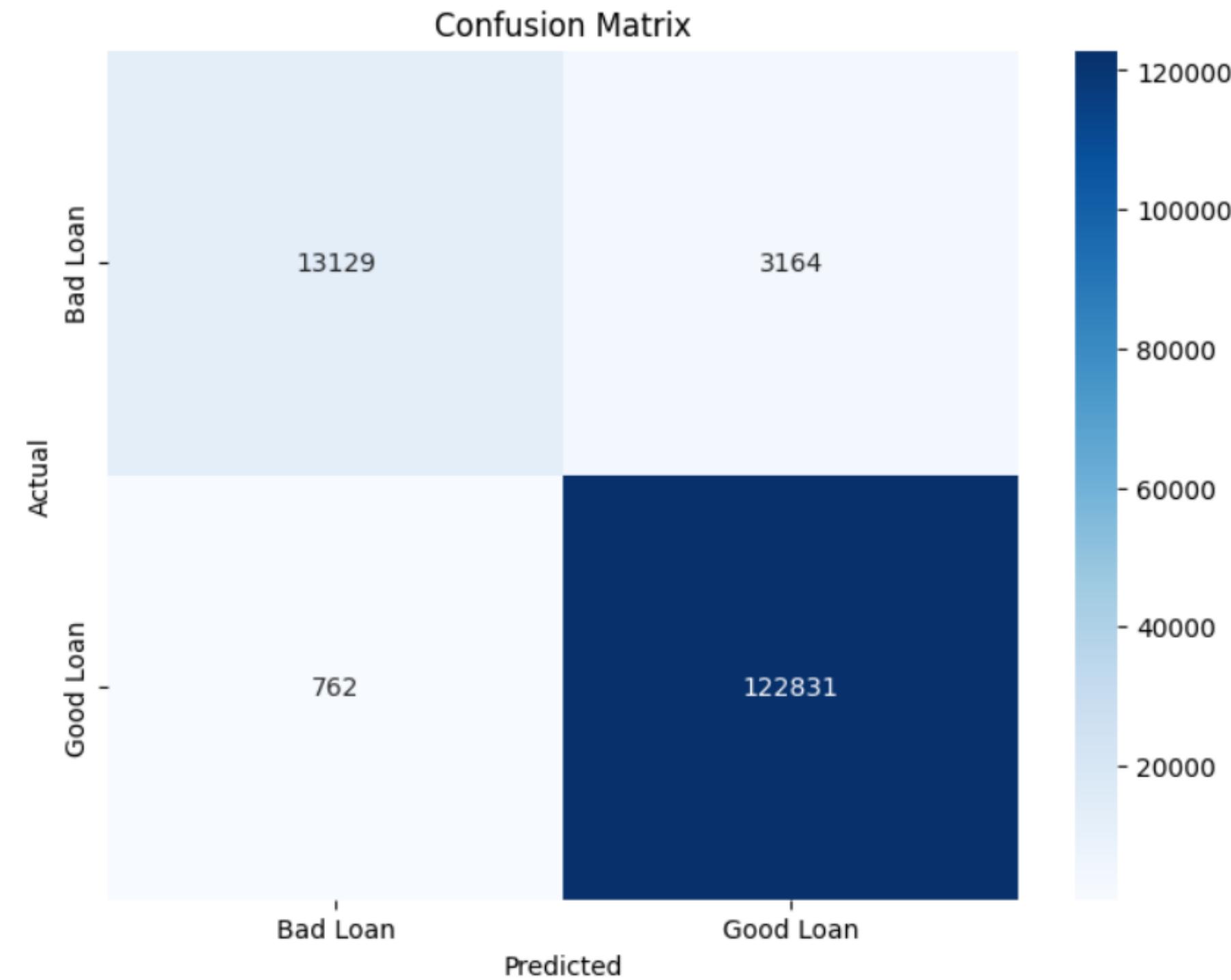


measures how well a model can distinguish between positive and negative classes across various thresholds.

To ensure accurate assessment



Modeling and Evaluation: Model Evaluation Parameters



Modeling and Evaluation: Model Evaluation Parameters

▼ Cross Validation

```
[193] scores = cross_val_score(dt_model, x_train, y_train, cv=5, scoring='roc_auc')

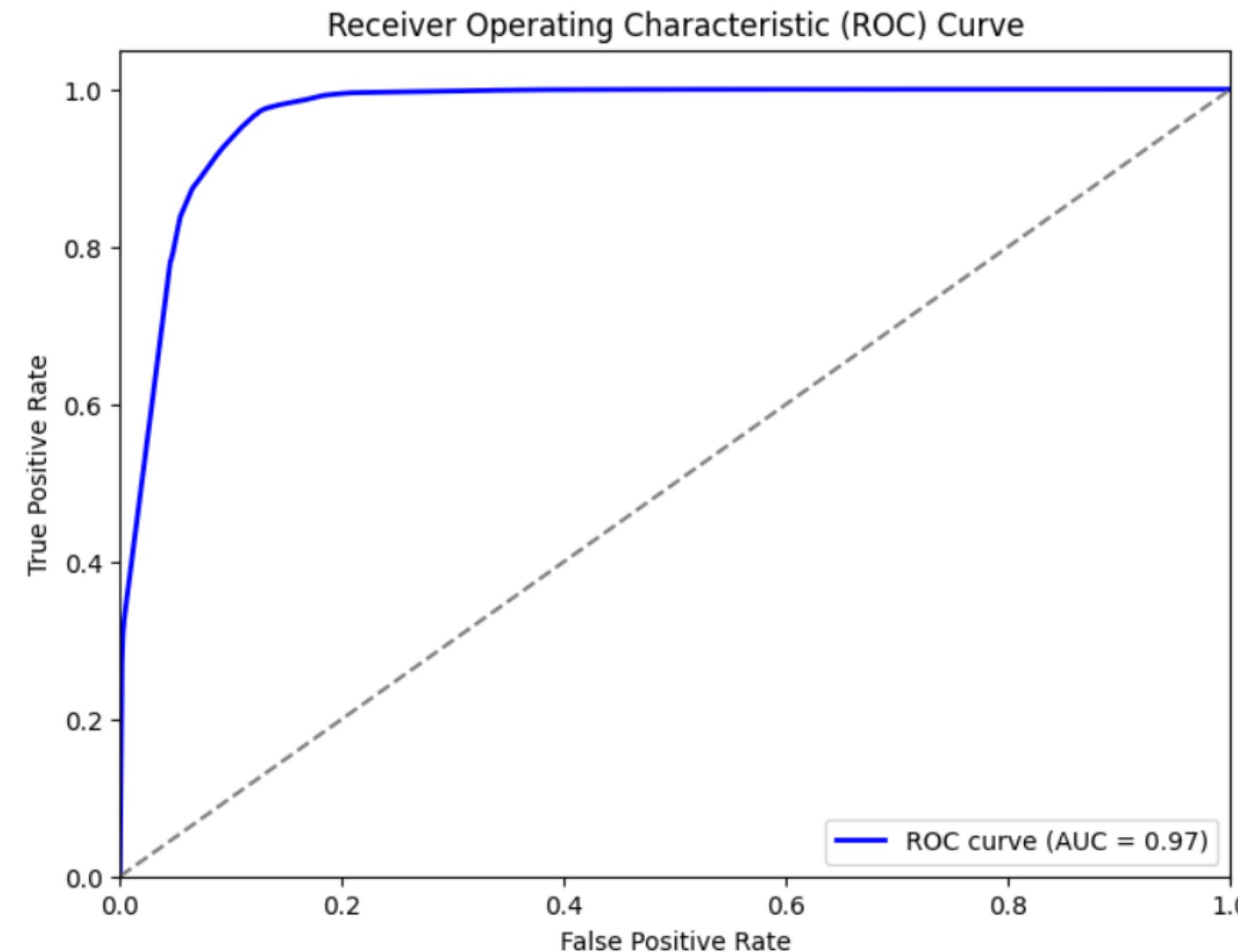
# Print the cross-validation scores
print("Cross-validation scores:", scores)

# Calculate and print the mean score
print("Mean cross-validation score:", scores.mean())

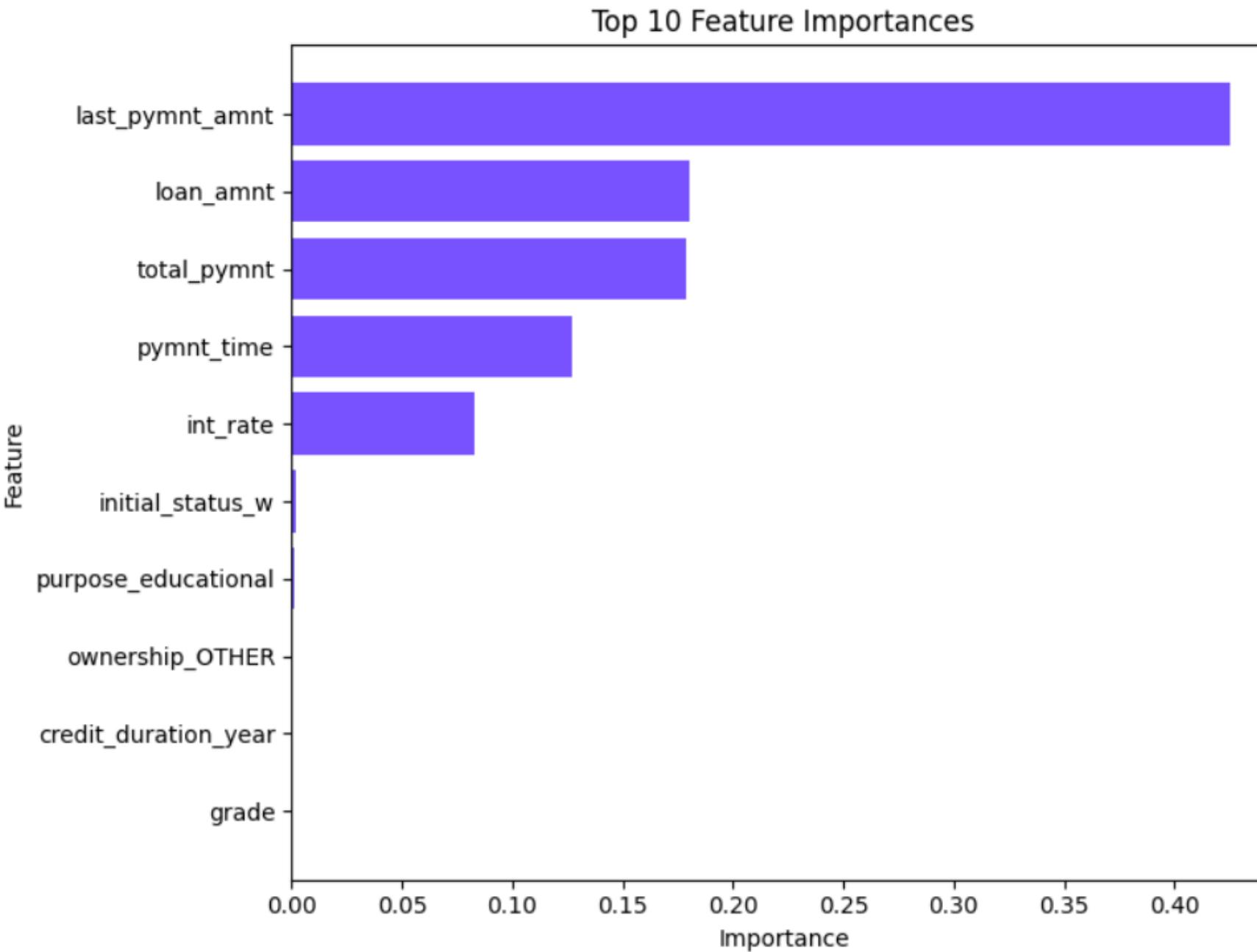
Cross-validation scores: [0.96671877 0.96837874 0.967064    0.96815854 0.96736511]
Mean cross-validation score: 0.9675370296583965
```



Modeling and Evaluation: Model Evaluation Parameters



Modeling and Evaluation: Model Evaluation Parameters





BUSINESS **RECOMMENDATION**



RECOMMENDATIONS

Business Recommendations



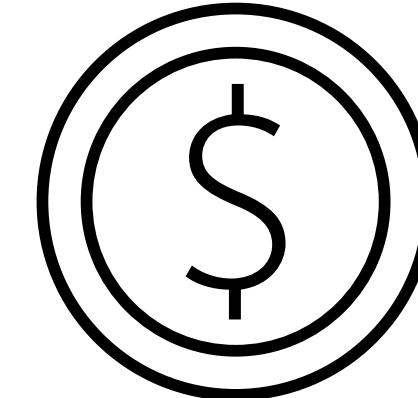
Last payment amount

Increasing the payment amount taken from customers can reduce the chances of them becoming high-risk loans



Loan amount

The loan amount granted is linked to the associated interest rate to be paid. As the loan size increases, the corresponding interest rate also increases.

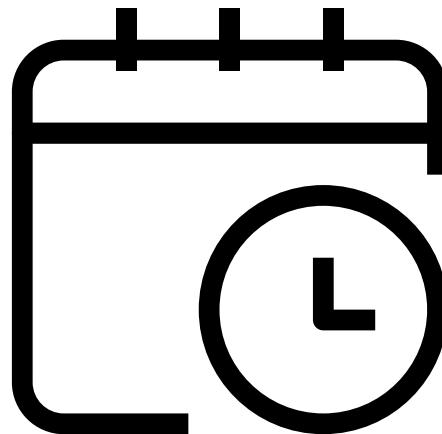


Total Payment

If payments fall within the range of 0 to 5000, it's probable that they would be deemed as problematic loans. This range often signals an increased likelihood of loan repayment issues or default.



Business Recommendations



Payment Time

The greater the duration customers have for repayment, the increased chance of them becoming high-risk loans.



Interest Rate

Higher interest rates taken by customers correlate with a rise in the rate of bad loans. Consider to maintain low interest rate.





Thank You