



1
Problem Scope
Definition



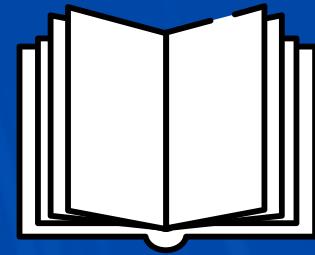
2
Key Customer
Features



3
Model Selection



4
Model Evaluation



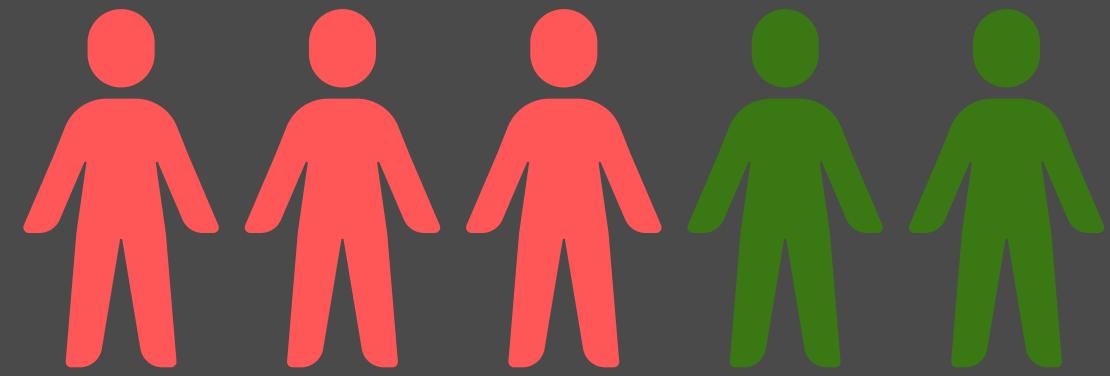
5
Plan Proposal

\$ERA CUSTOMER DEFAULT PREDICTION

By Justin Jimenez

**"HAVING MONEY ISN'T
EVERYTHING, NOT
HAVING IT IS."**

- KANYE WEST

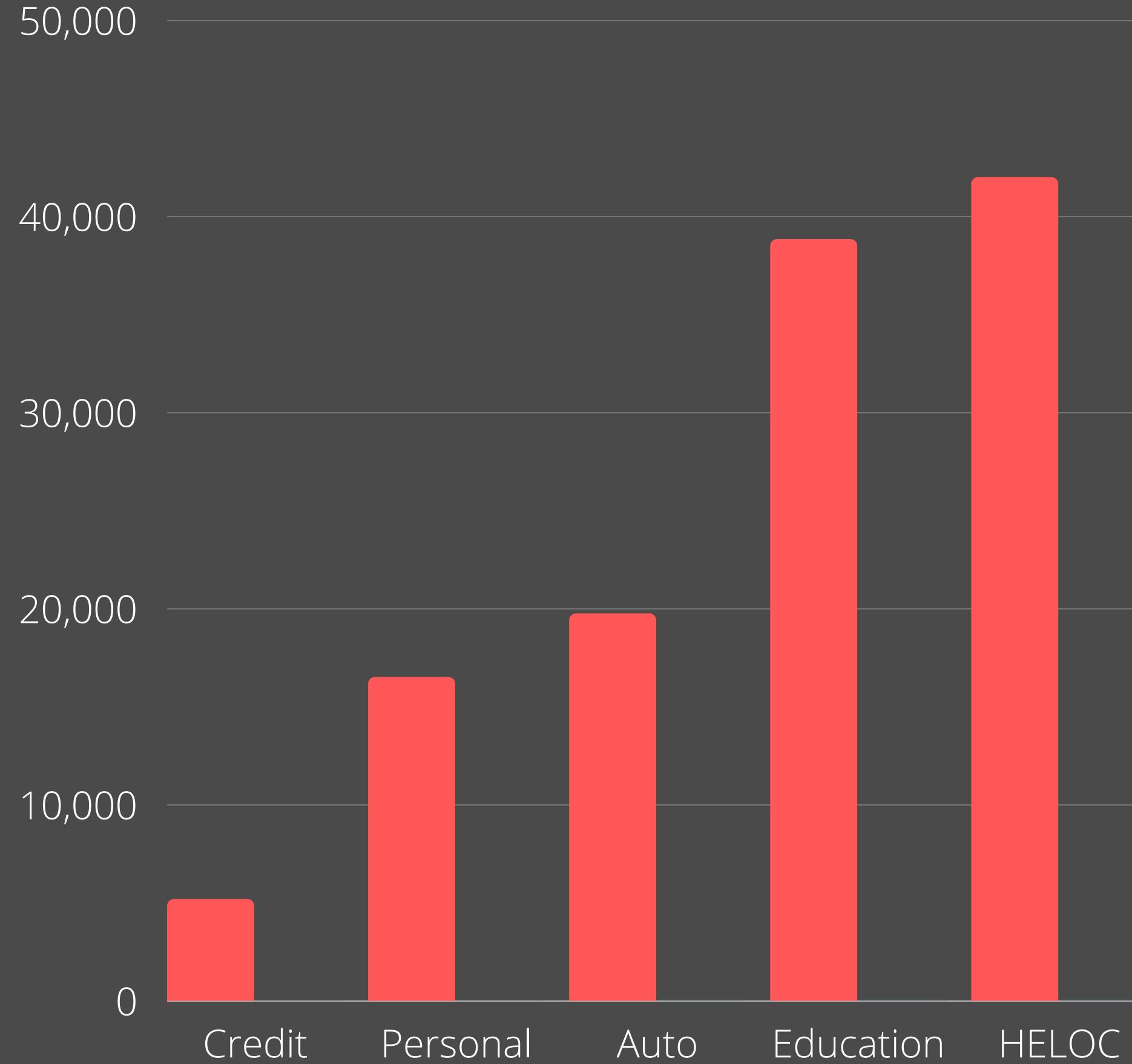


**Approximately 3 in 5
middle class Americans
struggle to pay consumer
debt.**

**\$92,727 Average
Total Balance**



Average Debt by Category



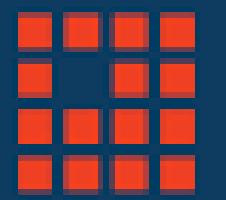
* Data reported from Experian

\$ERA Objective:

**To provide the tools and services
for users to be able to achieve
financial freedom.**

PROBLEM SCOPE DEFINITION

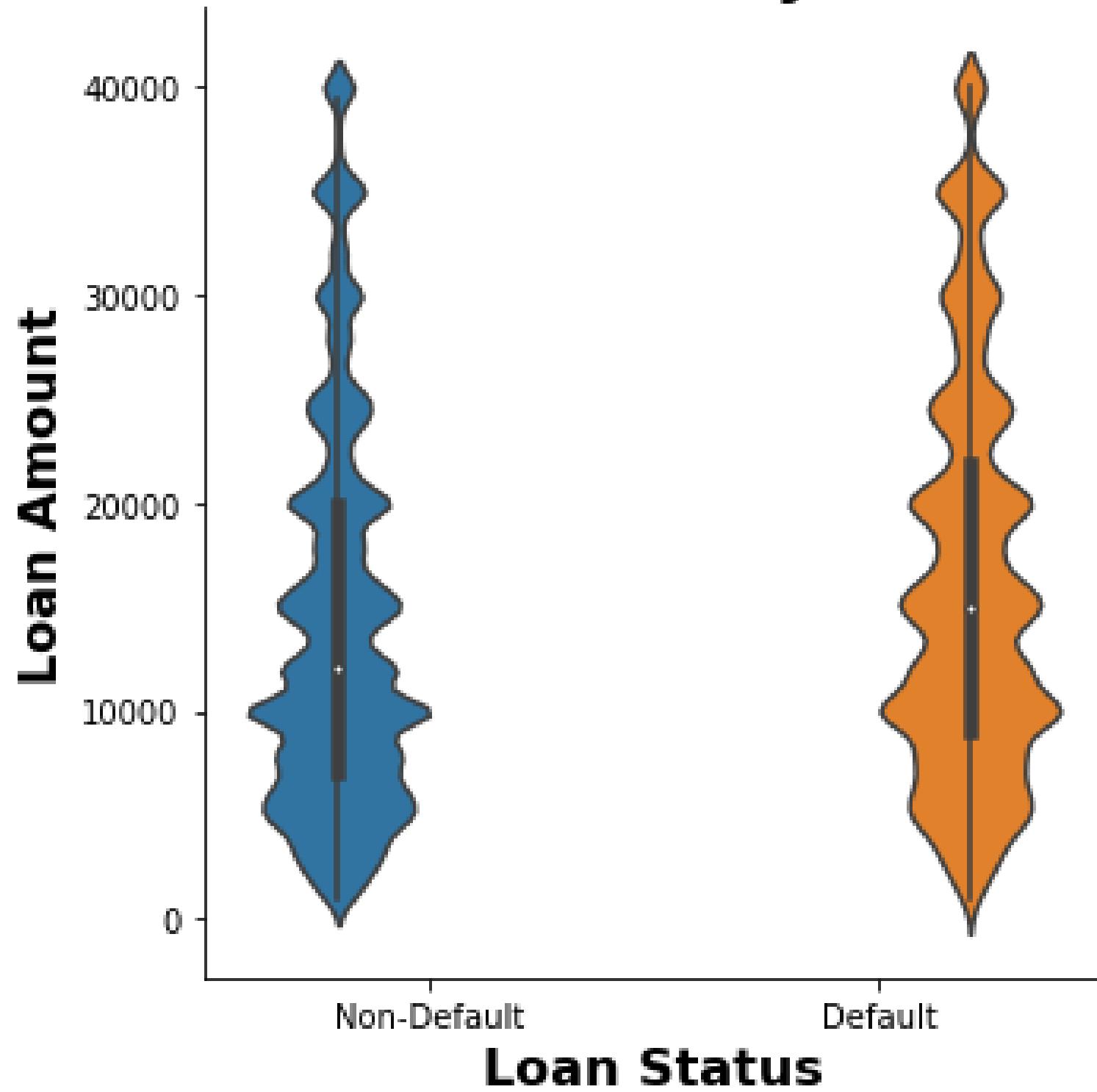
- 1 CLASSIFY USERS LIKELY TO DEFAULT
- 2 EASILY INTERPRETED
- 3 NO FEATURES FROM CREDIT CHECKS
- 4 MAXIMIZE IDENTIFICATION OF BAD BORROWERS
- 5 READY FOR DEPLOYMENT Q2 2022



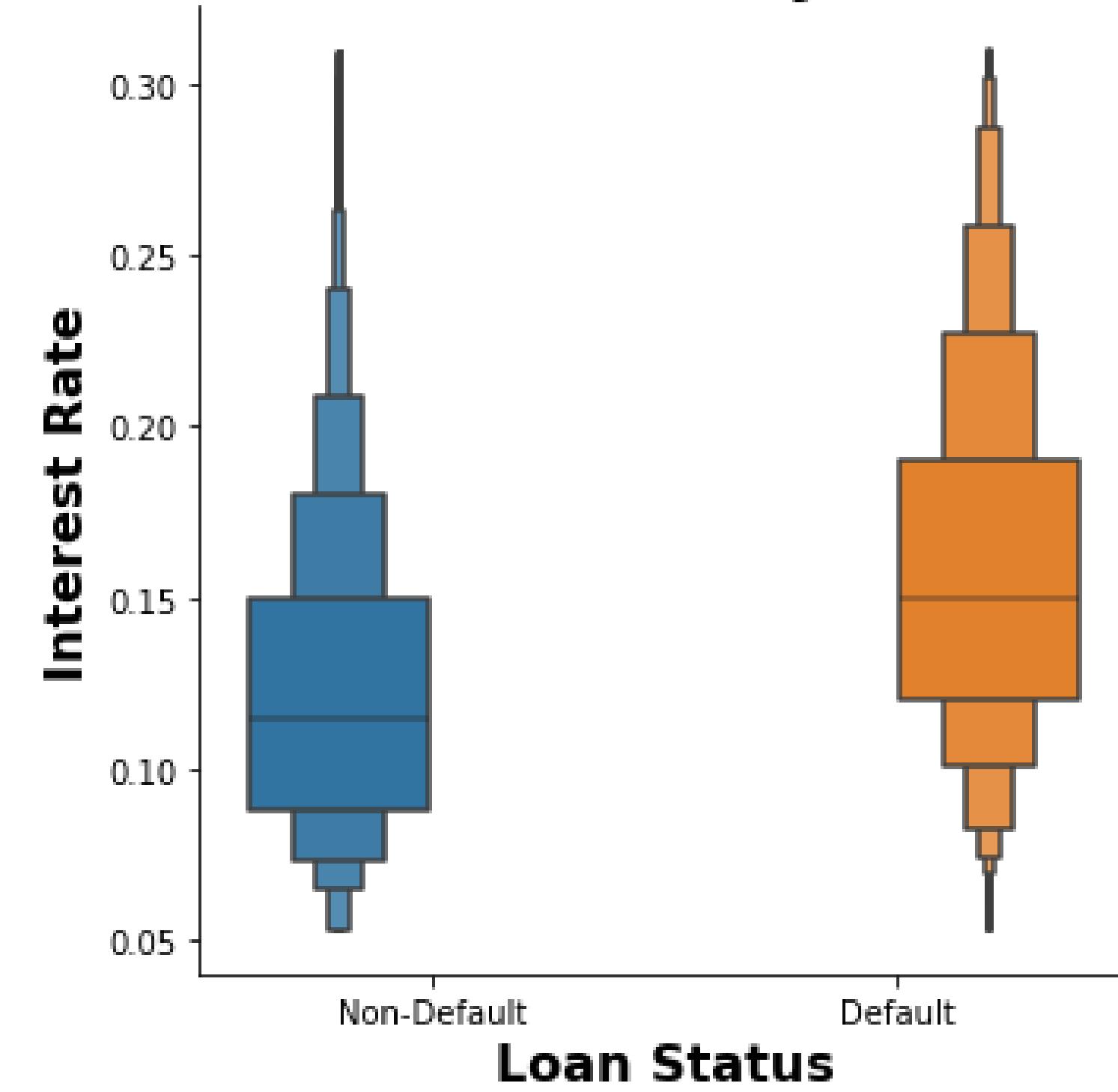
LendingClub

- P2P loan data from 2007 to Q4 2018.
- Original sample of 3.2 million loans.
- Cleaned sample of 800,000 loans.
- 20% Defaulted loans.

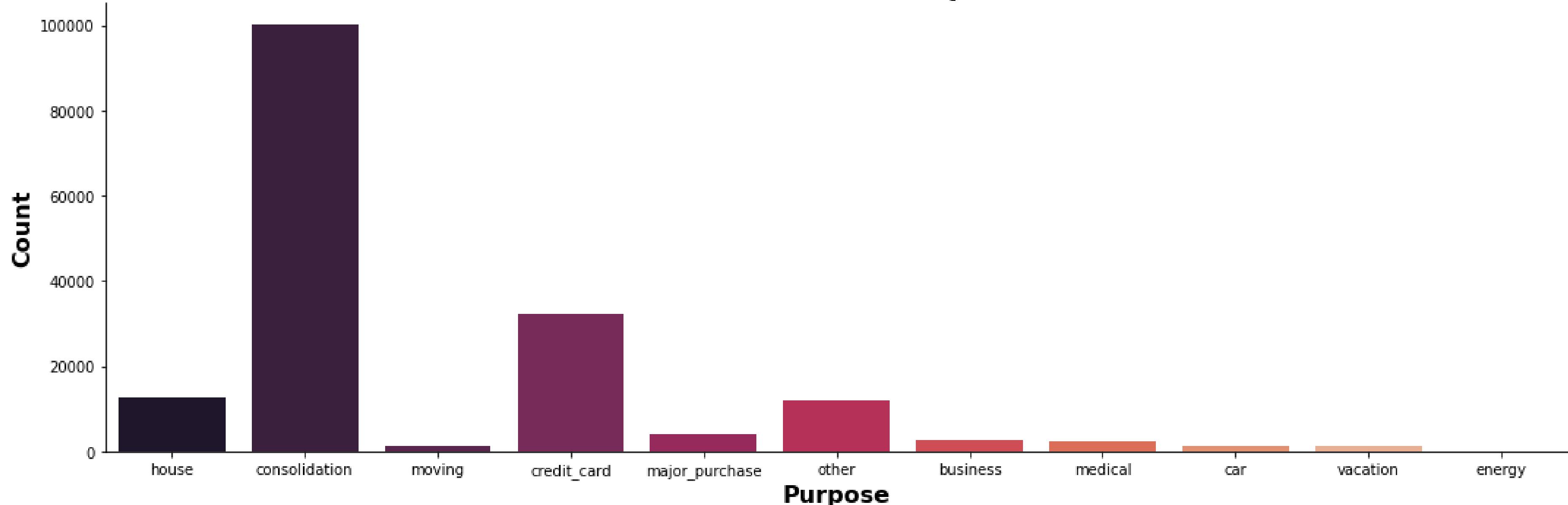
Loan Amount by Status

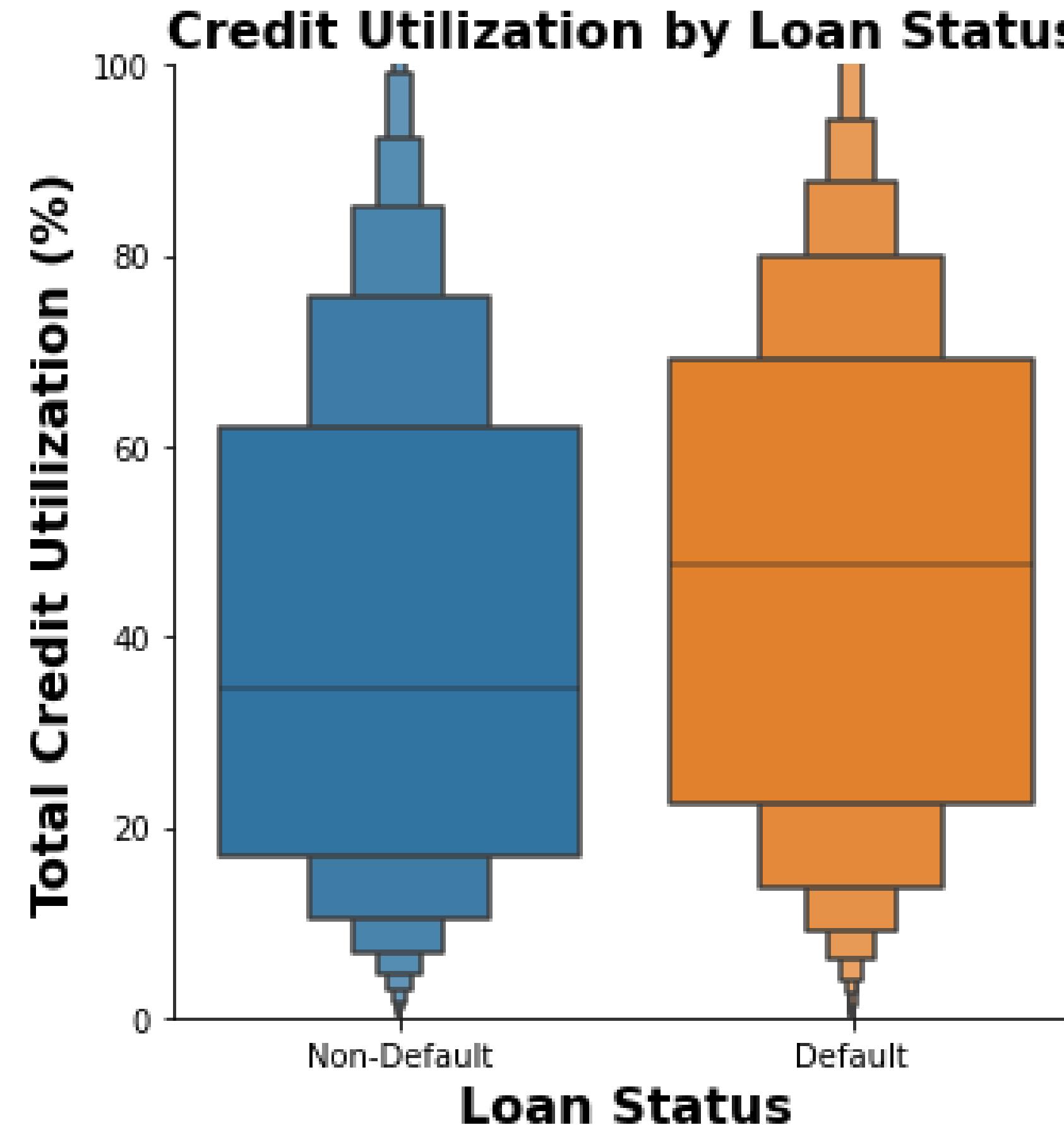


Interest Rate by Status



Defaulted Loan Purposes

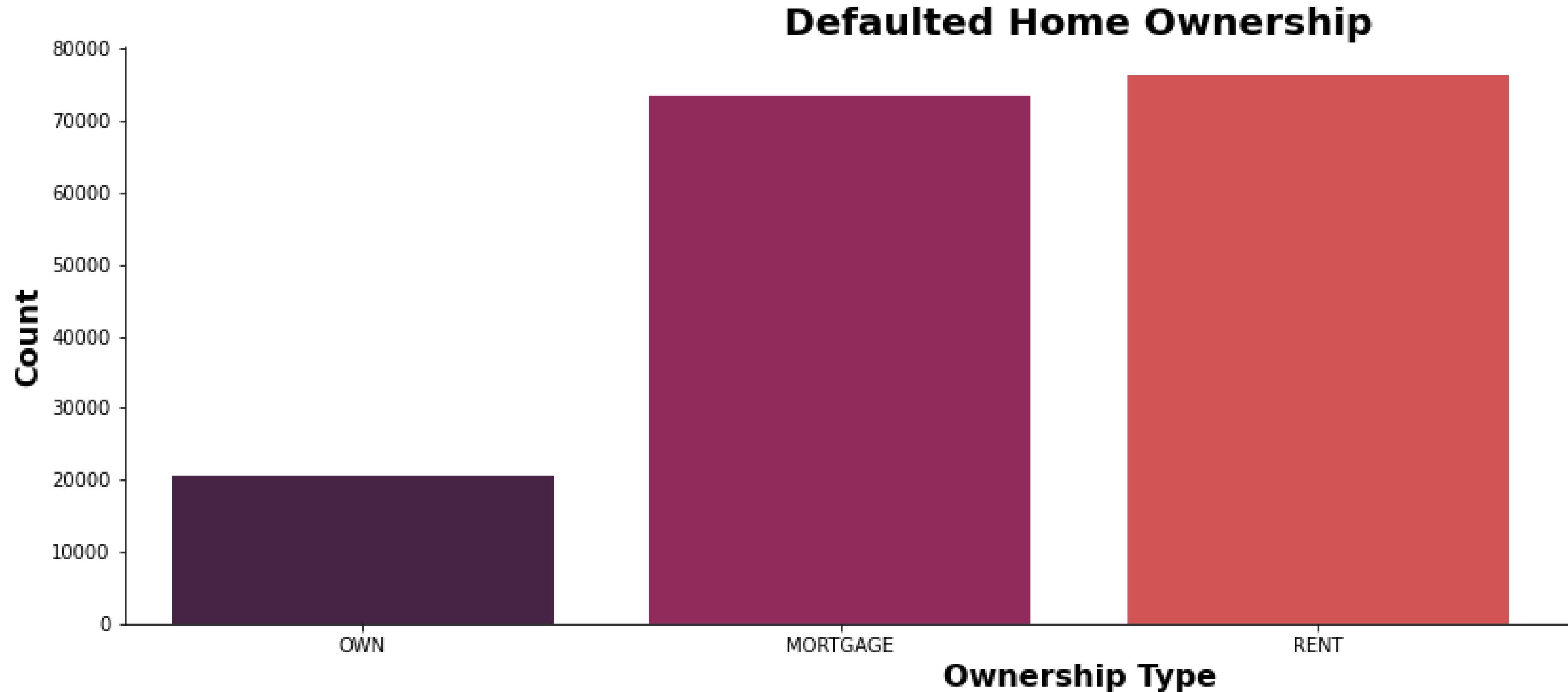




- **48% median utilization among **defaulted** borrowers.**
- **35% median utilization among **non-defaulted** borrowers.**

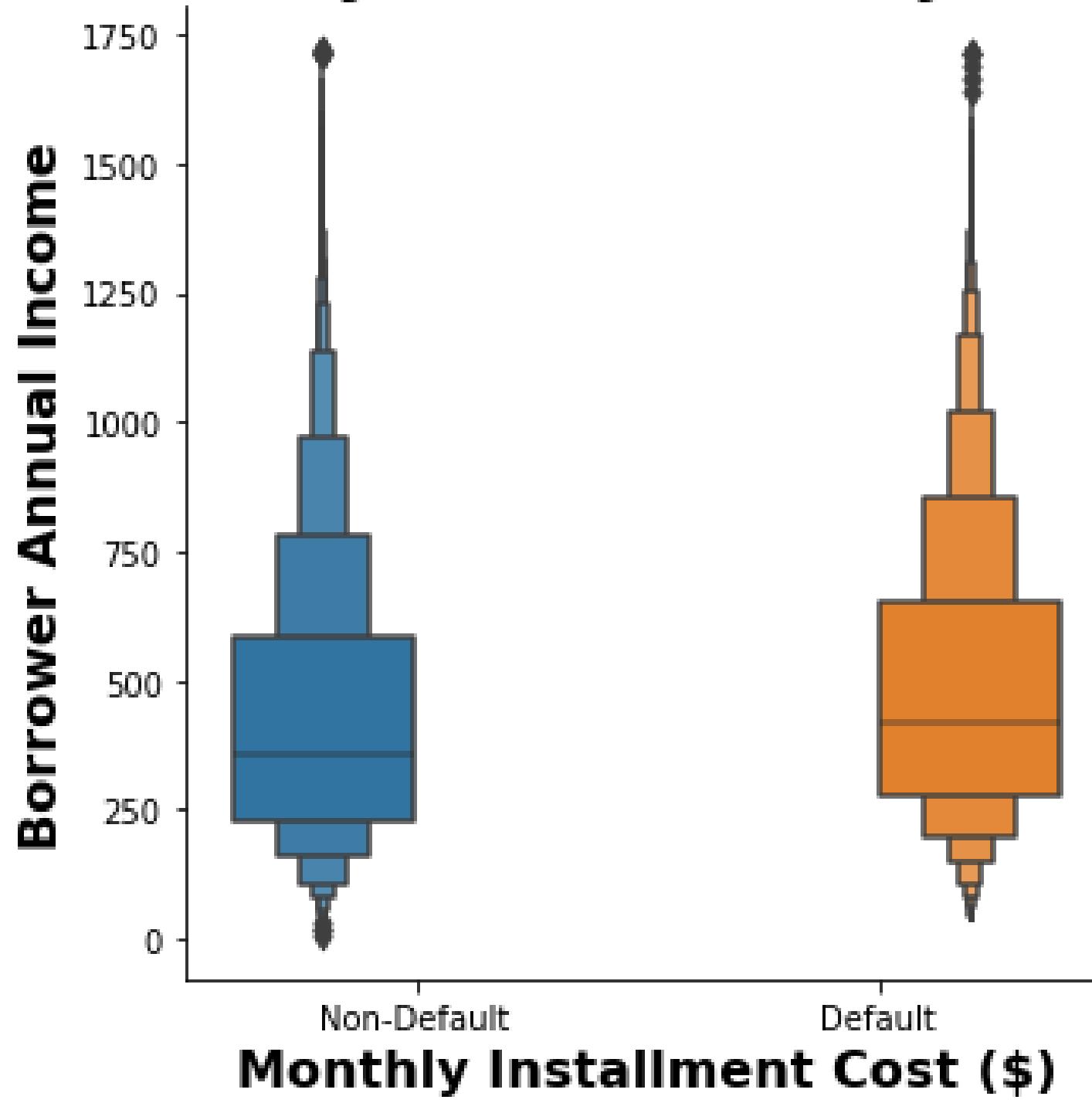
On average, defaulted
borrowers...

Rent rather than own a home.

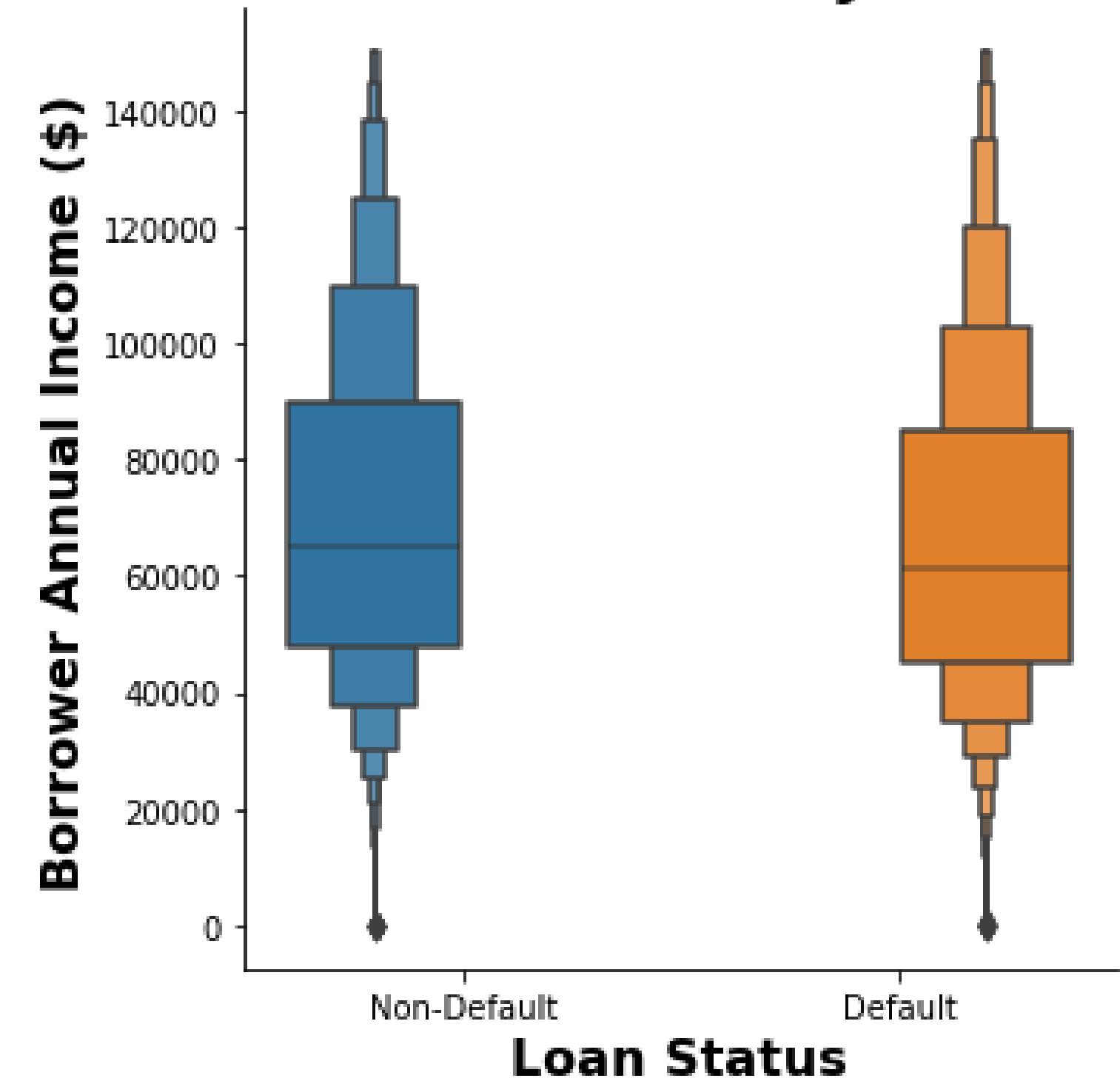


Pay More, Earn Less

Monthly Installments by Status

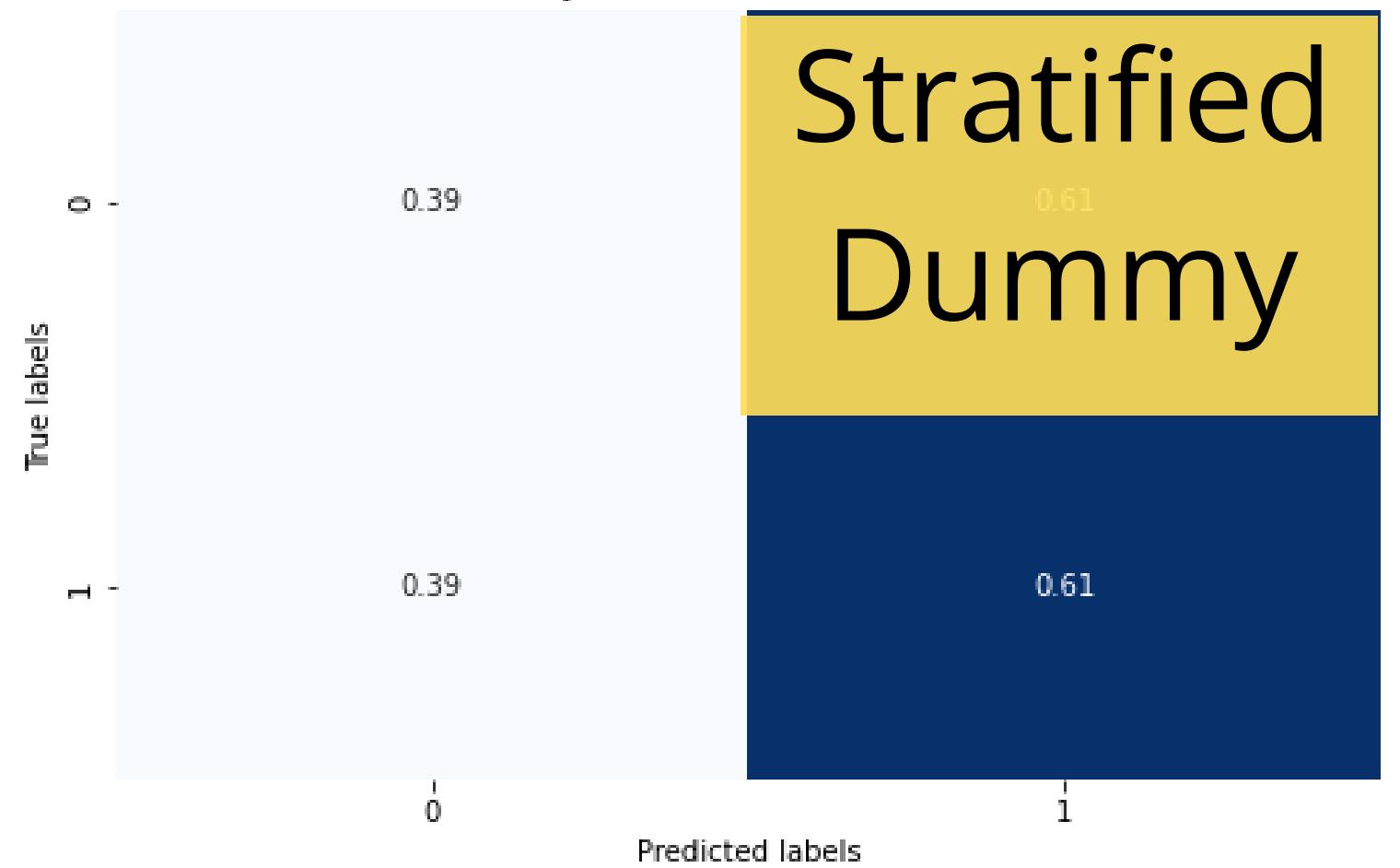


Annual Income by Status

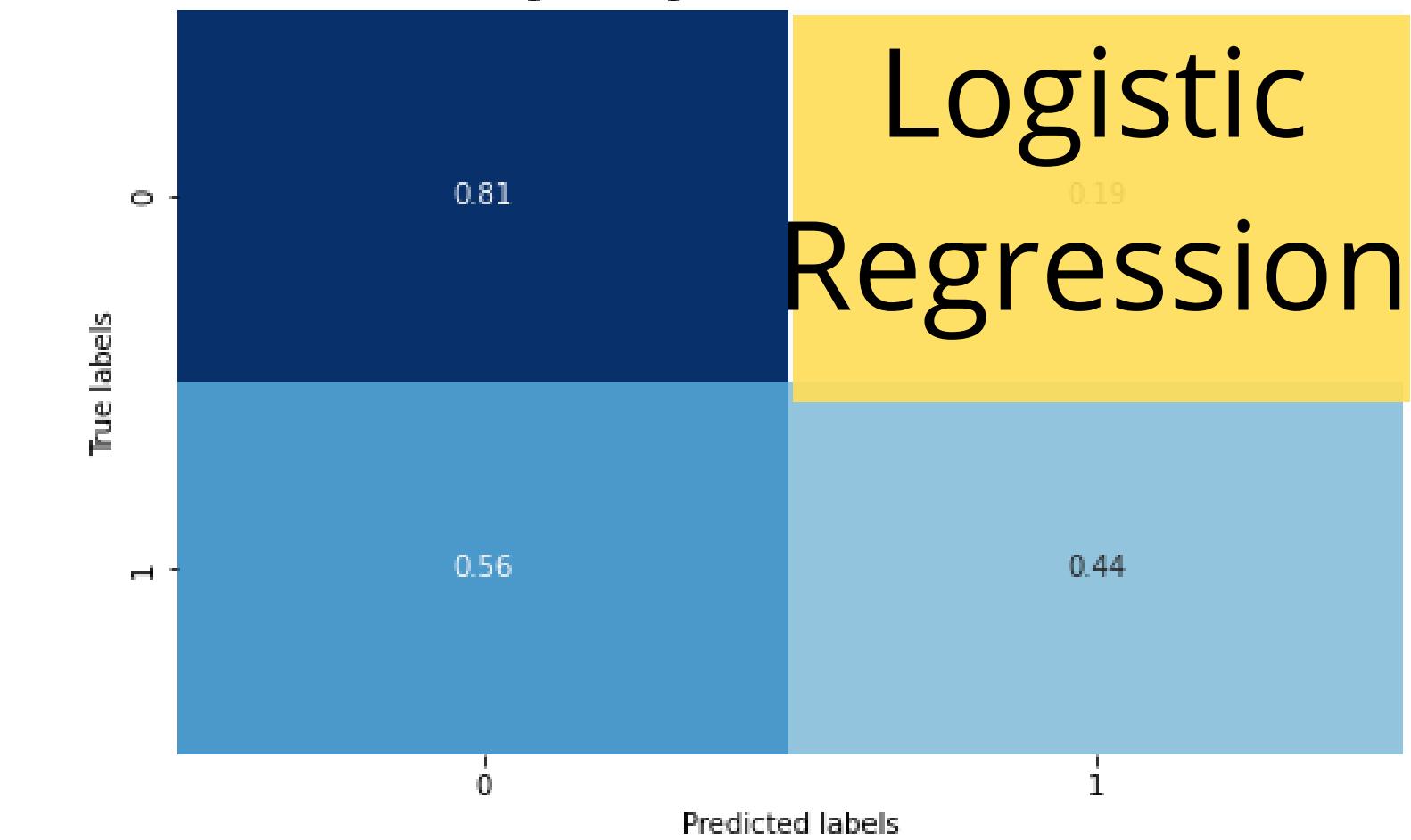


- Lower credit limits
- High utilization rates
- Younger credit
- More recent activity
- More bankruptcies, chargeoffs

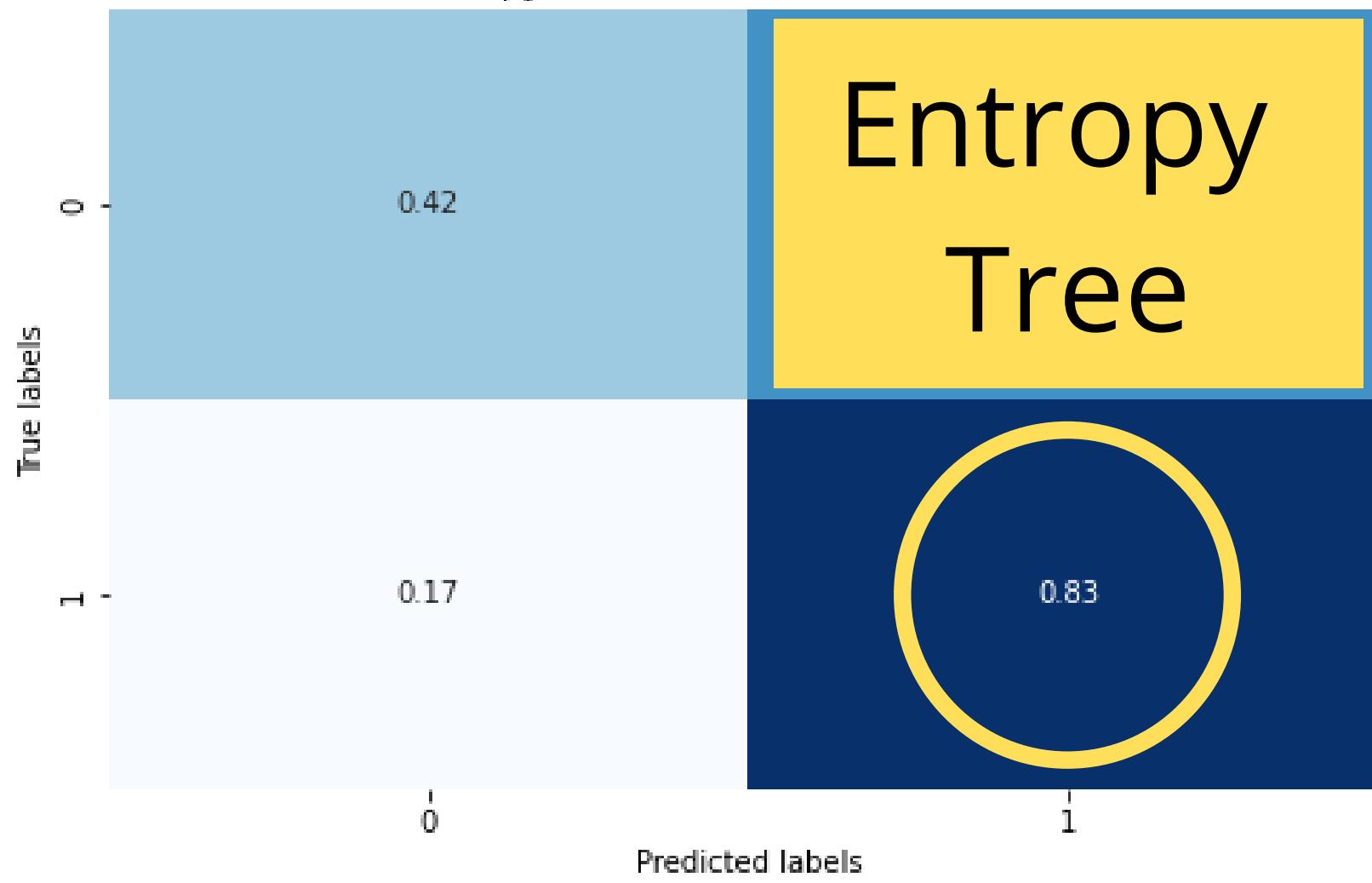
Dummy Classifier Confusion Matrix



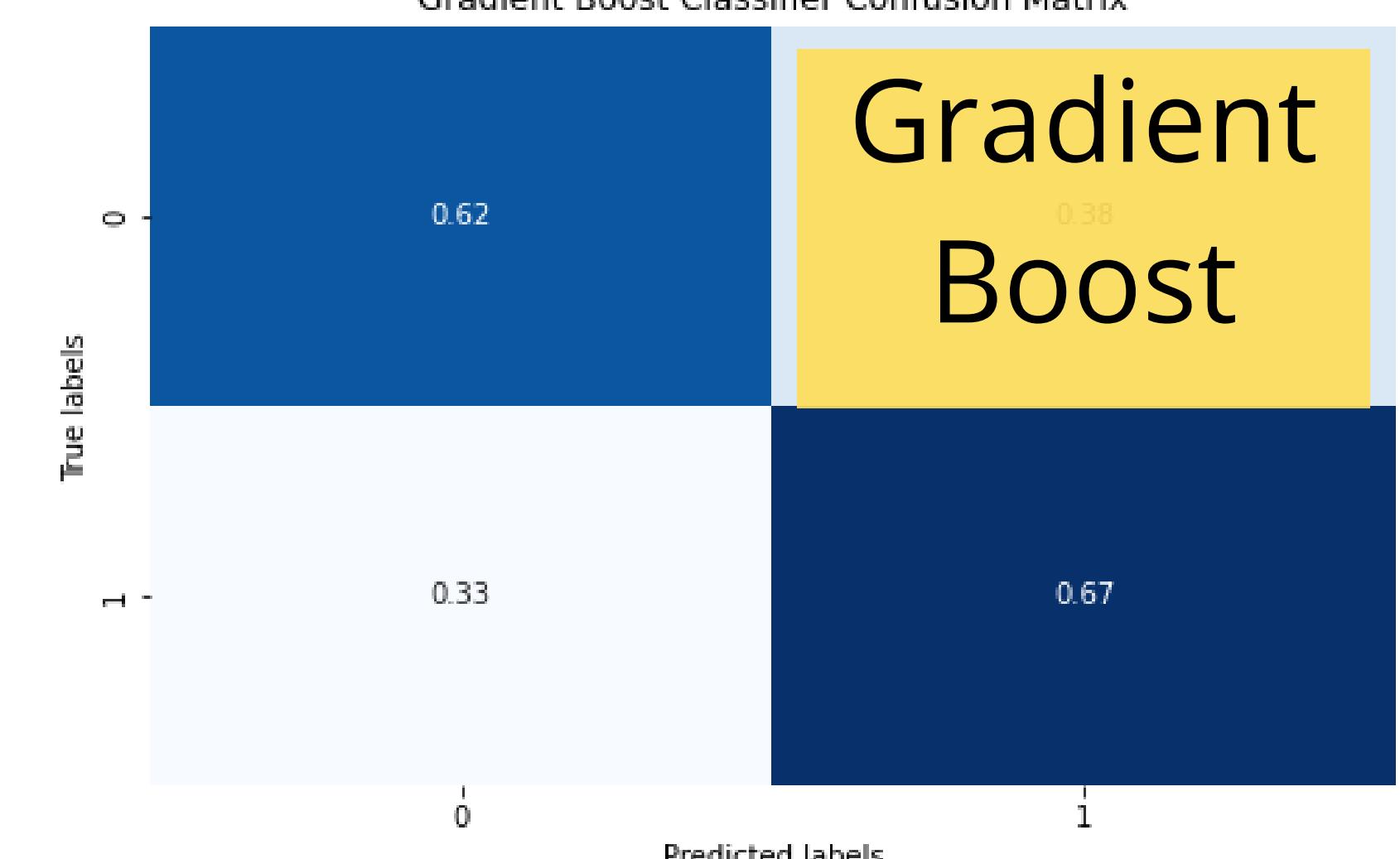
Logistic Regression Confusion Matrix



Entropy Decision Tree Confusion Matrix



Gradient Boost Classifier Confusion Matrix



Entropy Decision Tree

Optimized for recall performance

80/20 Train-Test Split

5 fold Cross Validation

Hyperparameter tuning with RandomizedSearchCV

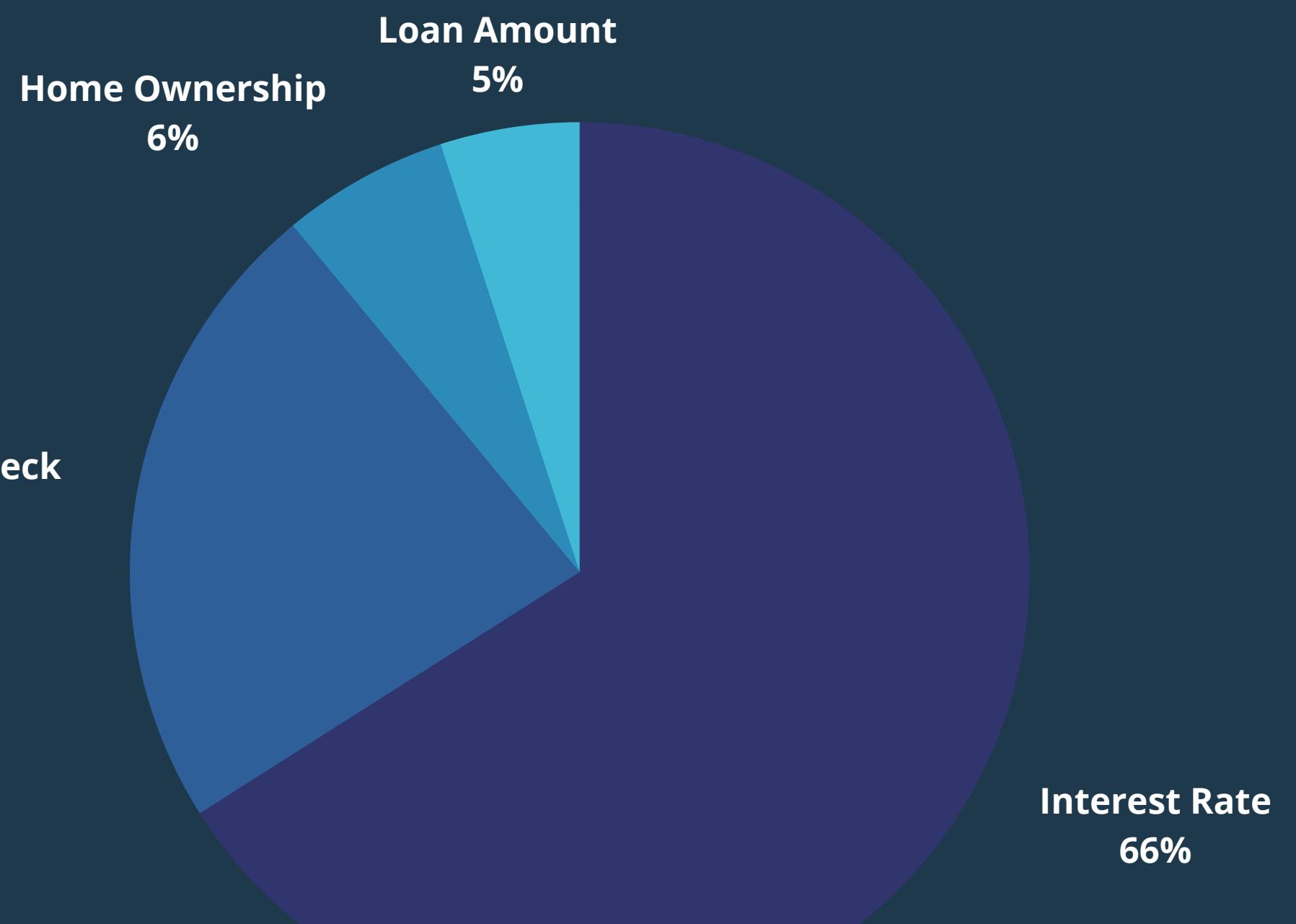
Max Depth = 3

Class Weight Ratio 0.76:1 (Non-Default : Default)

Min Samples Split = 46, Min Samples Leaf = 2

TOP 5 PREDICTION FEATURES

Coefficients of Key Features



INTEREST RATE

7 - 30%



LAST CREDIT PULL

More Recent Pulls Are More Likely



HOME OWNERSHIP

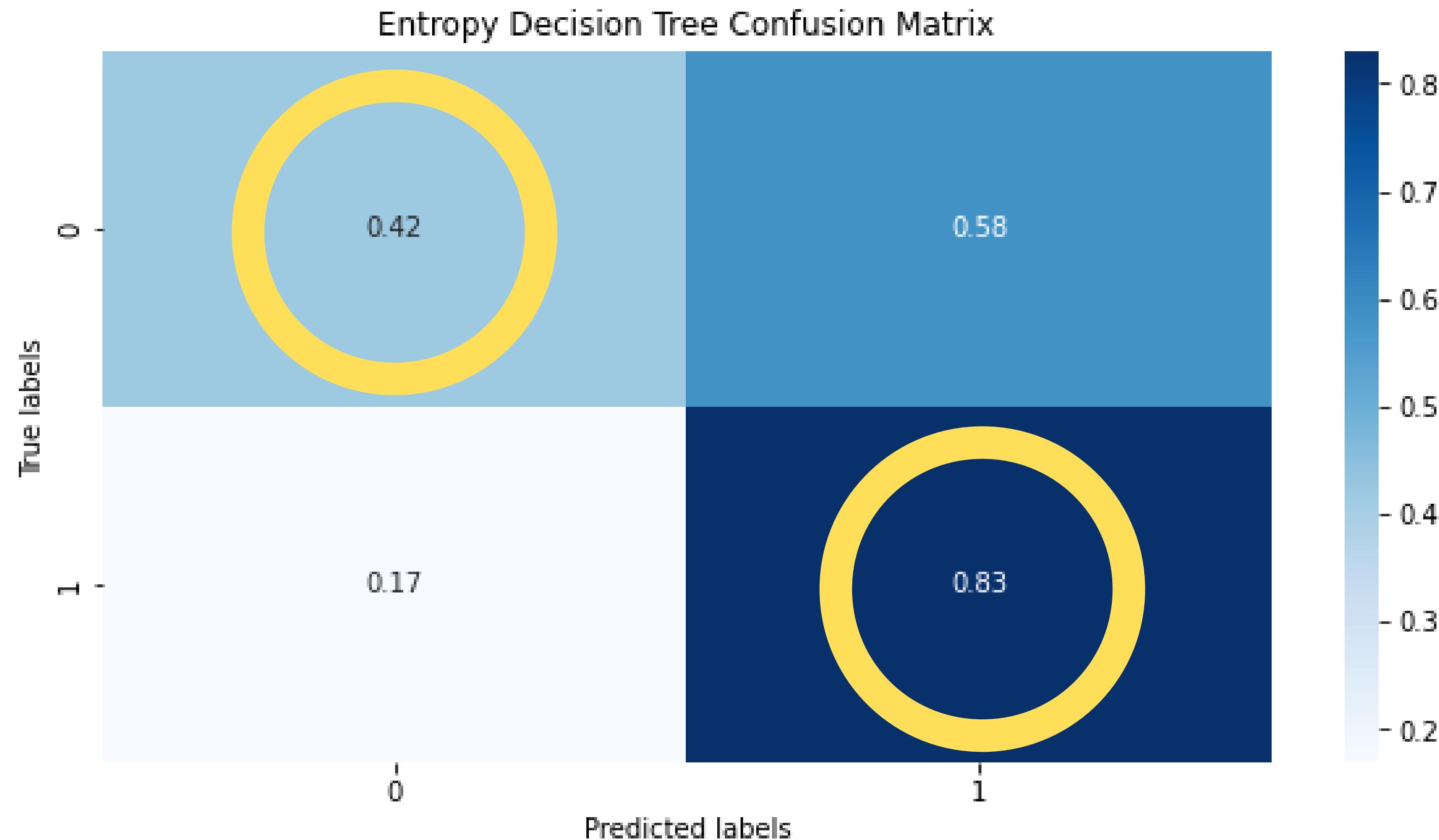
Mortgage or No Mortgage?

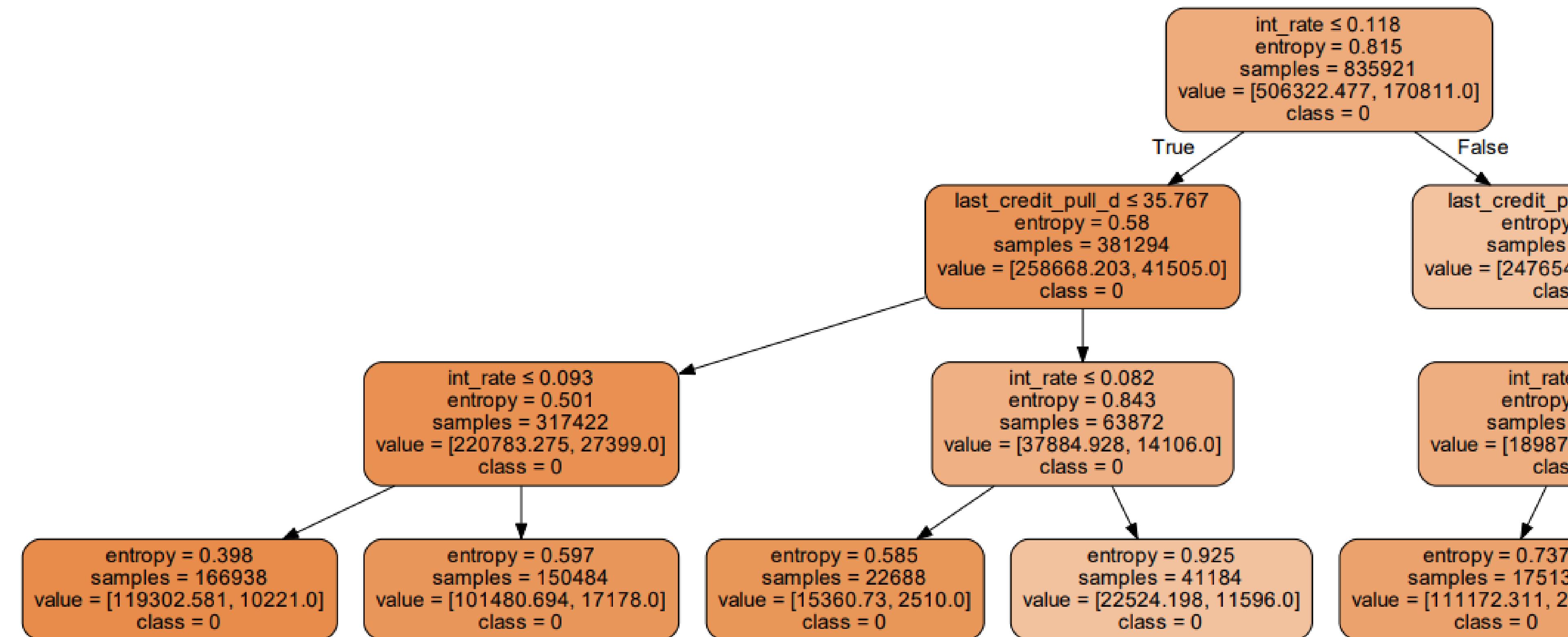


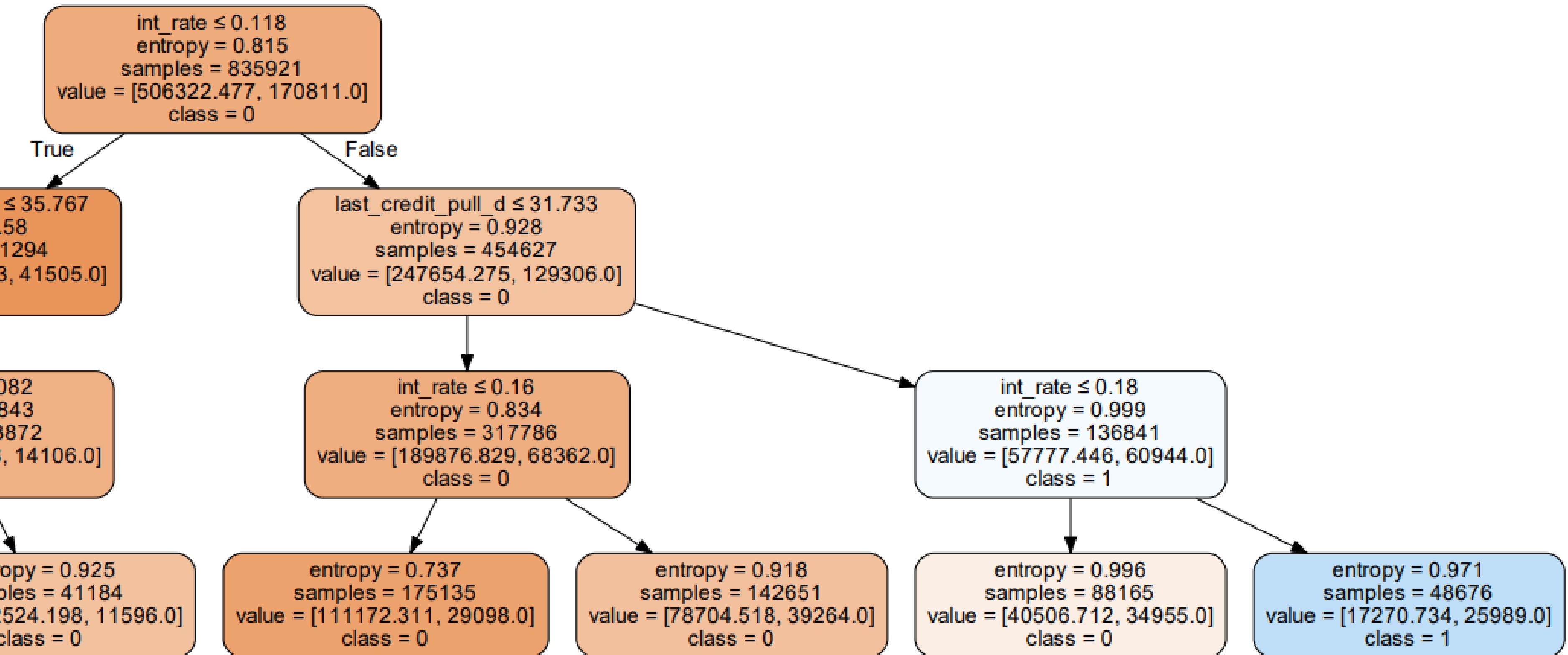
LOAN AMOUNT

\$1,000 - 40,000

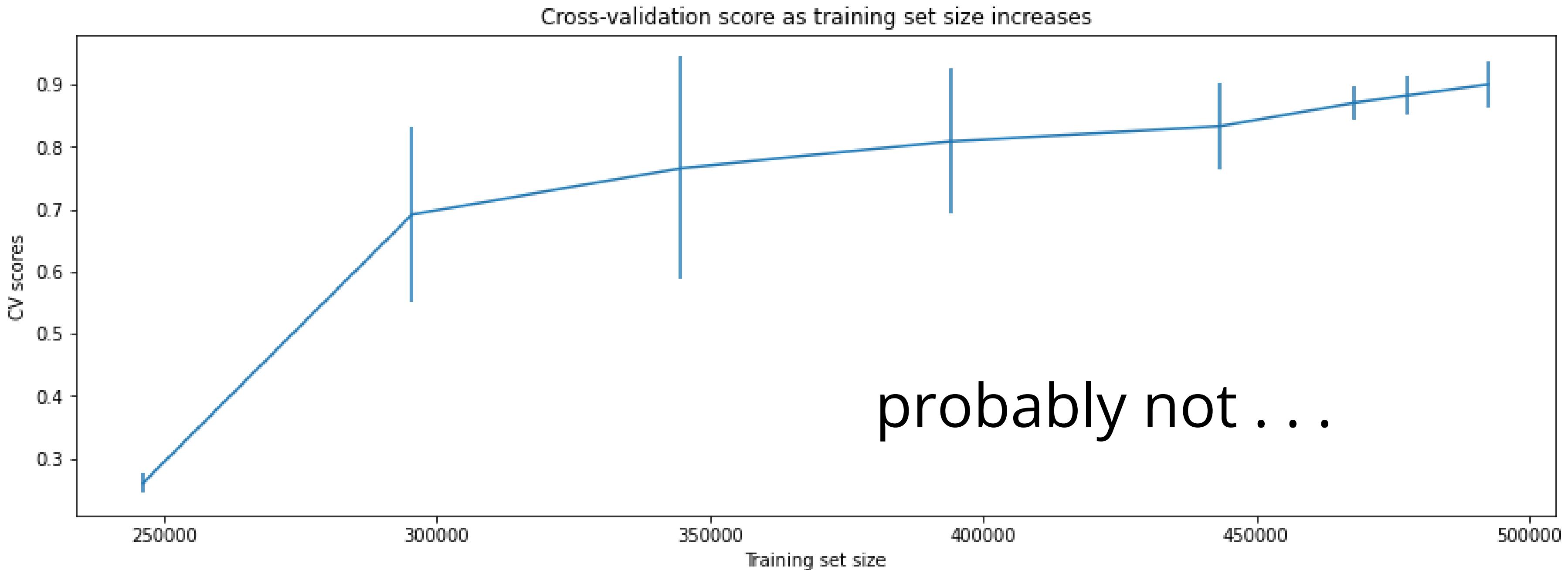
Model Performance

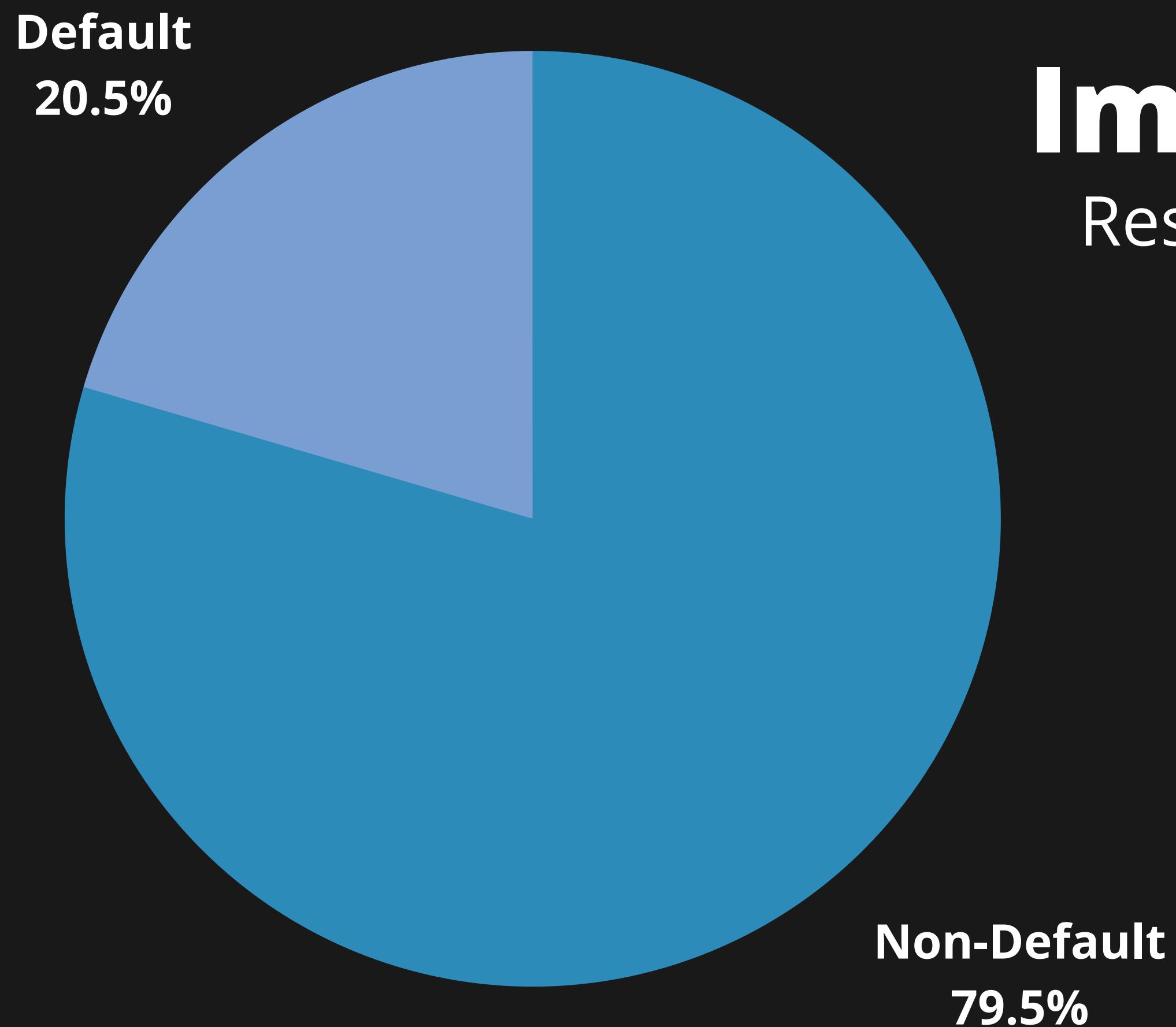






Need more data?





Imbalanced Data

Resampled using SMOTEENN

- One Hot Encoded and Standardized
- Only used features that users could provide, no credit report data

All models are wrong,
some models are useful.

Is this one?

LOSS Prevention

418,000 loans in test set:
86,000 (26%) defaulted.

This model would have caught 71,380 loans.

\$15,000 Median Loan Amount * 71,380 loans =
Approximately \$1.07 billion.

\$284 million more than the baseline dummy classifier.

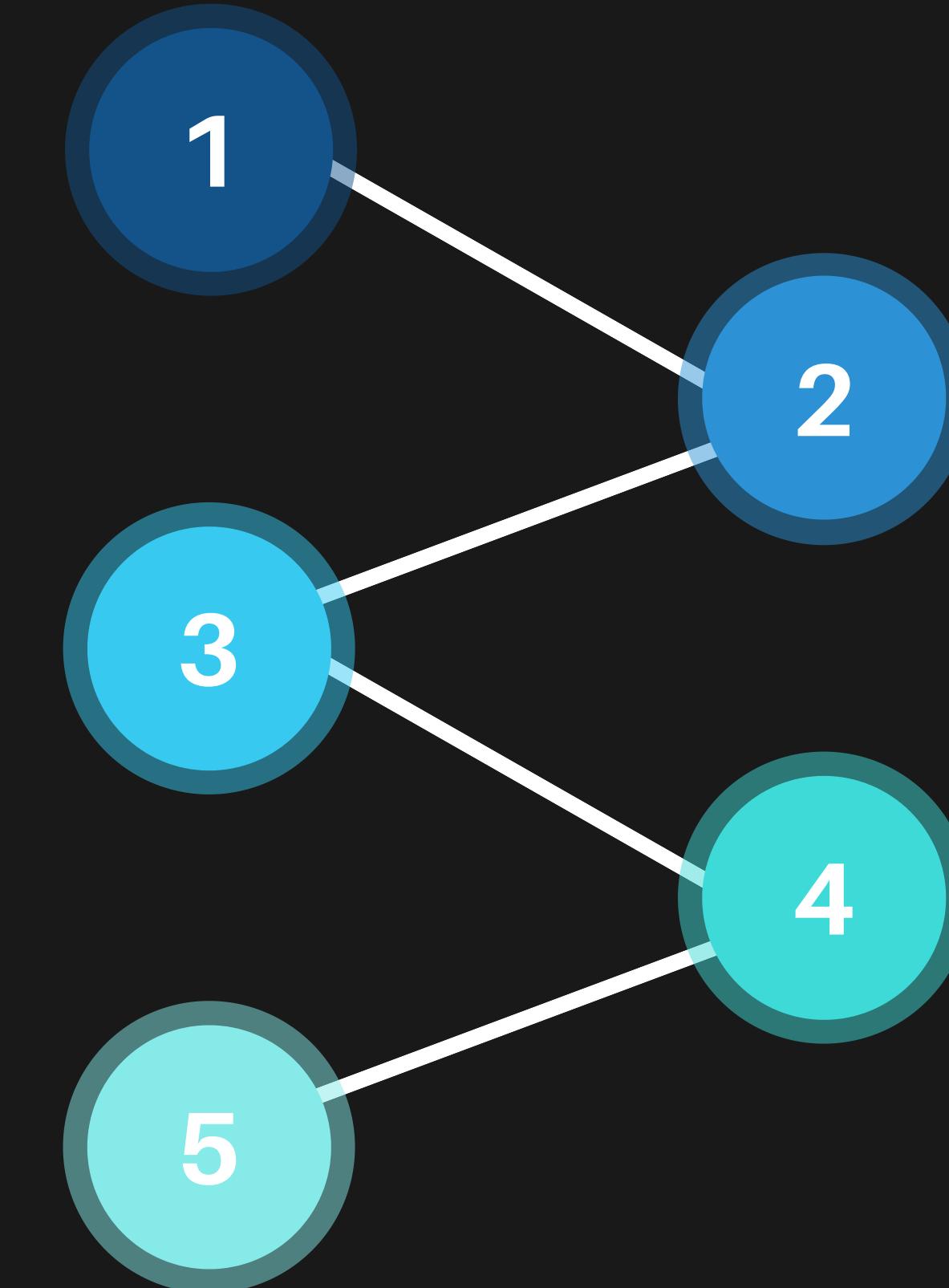
IMPLEMENTATION PLAN PROPOSAL

5-Step Implementation Plan

Develop calculator web app to collect data and leads

Create log and dashboard to track model performance and degradation

Develop related models to predict interest rates and other features.



Integrate model into web app for deployment

Update the model with batch learning process