Boston University
BA 305 Final Project

RISK OR REWARD?

Predicting Loan Outcomes

Ania Shaheed

CONTENT

01 CONTEXT


application
LendingClub
capital
BORROWERS
INVESTORS

BAD BORROWERS

Late Payments
Missed deadlines

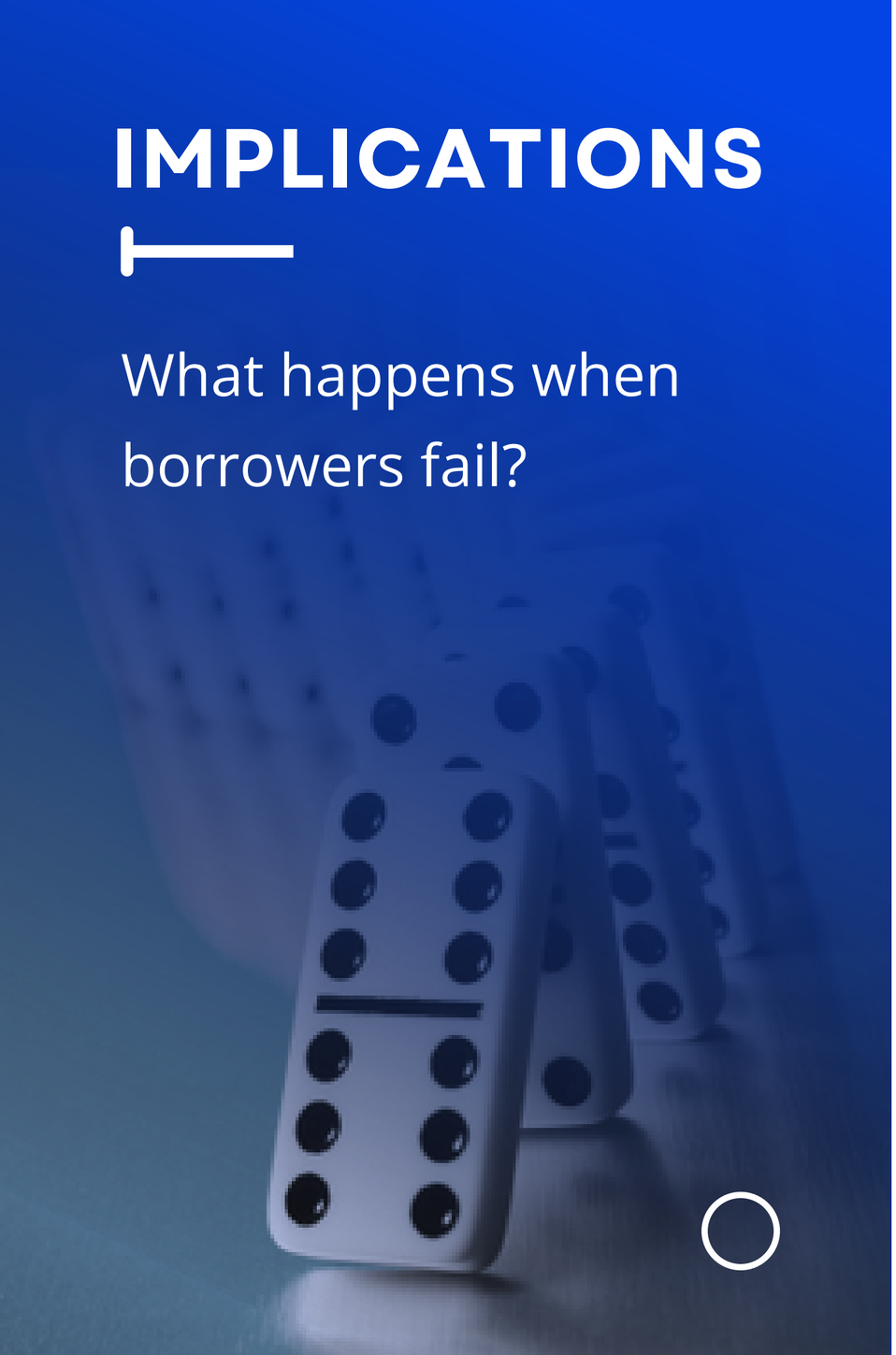
Charged Off
Lender gives up

Default
No repayment

High Risk
Bad credit history

IMPLICATIONS

What happens when borrowers fail?



01 Investors lose money

They don't get repaid and lose their principal and interest

02 LendingClub loses trust

If too many loans fail, new investors won’t trust LendingClub

03 Higher rates for everyone

Good borrowers pay higher rates because lenders demand more protection

04 Borrowers’ credit drop

Defaulting crushes their credit score, hurting future borrowing

02 THE DATA

OVERVIEW

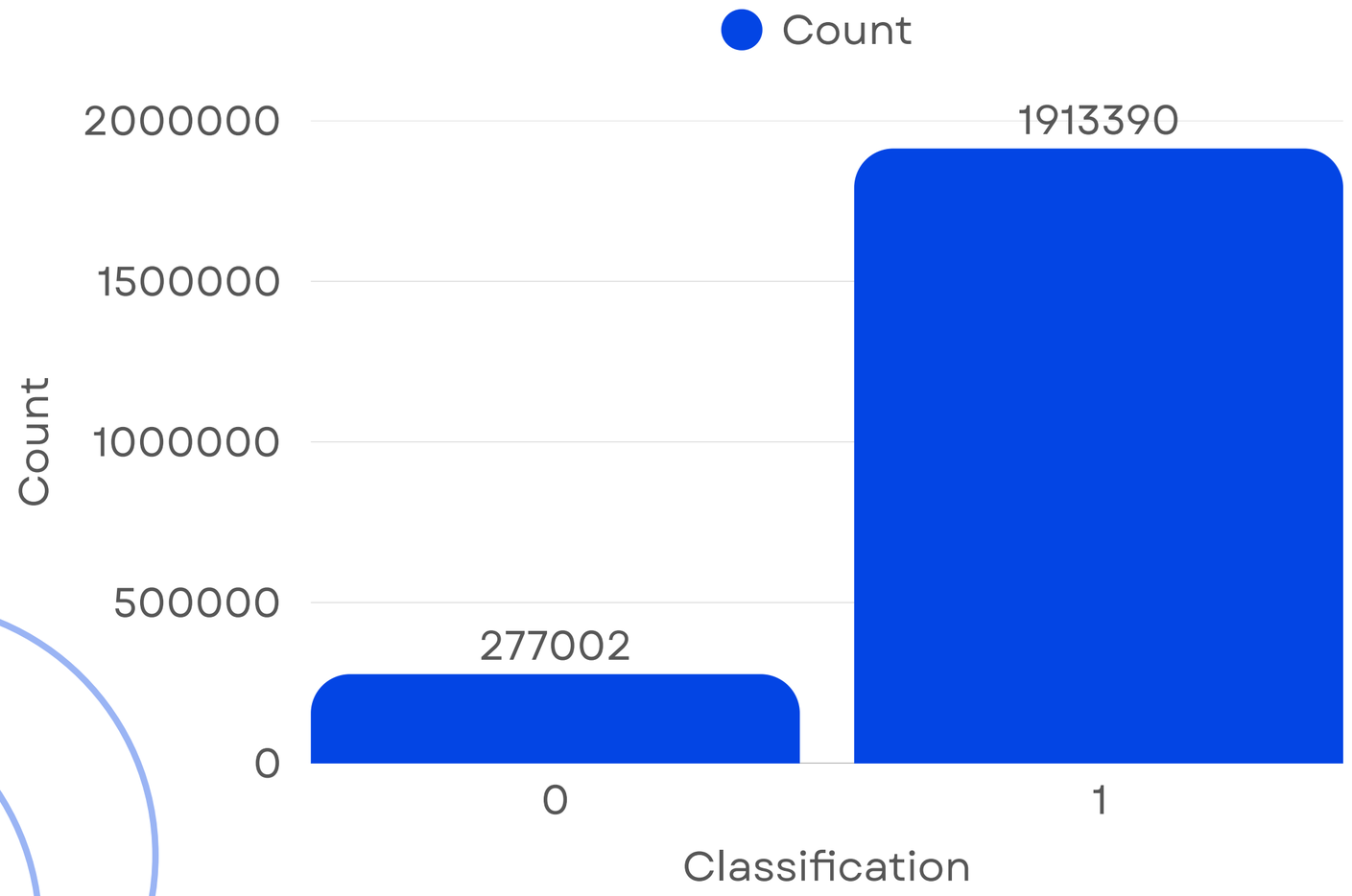
2,260,668
ROWS

145
FEATURES

LOANS ISSUED
2012 - 2019

FEB. 2019
SNAPSHOT

TARGET VARIABLE | LOAN STATUS



Classification	Evaluation
Fully Paid	
Current	
In Grace Period	
Late (16-30 days)	
Late (31-120 days)	
Charged Off	
Default	
Does not meet the credit policy - Fully Paid	
Does not meet the credit policy - Charged Off	

MEET THE AVERAGE BORROWER

“I want to break free from credit card debt”

Career

Professional Chemist

- **6 years** of experience
- **$78K** annual income
- **19%** debt-income ratio

LAUREN SMITH

Demographics

31 years old
NYC | Renting

Loan Terms

$15K
36 months

Risk Profile

13% interest
C1 subgrade

03

OUR PROJECT

OBJECTIVES

PREDICT DEFAULT RISK

Use machine learning to **estimate the likelihood of a borrower defaulting** after a loan is issued. Sort given loans into "**high risk**" or "**low risk**" for early warning and intervention.

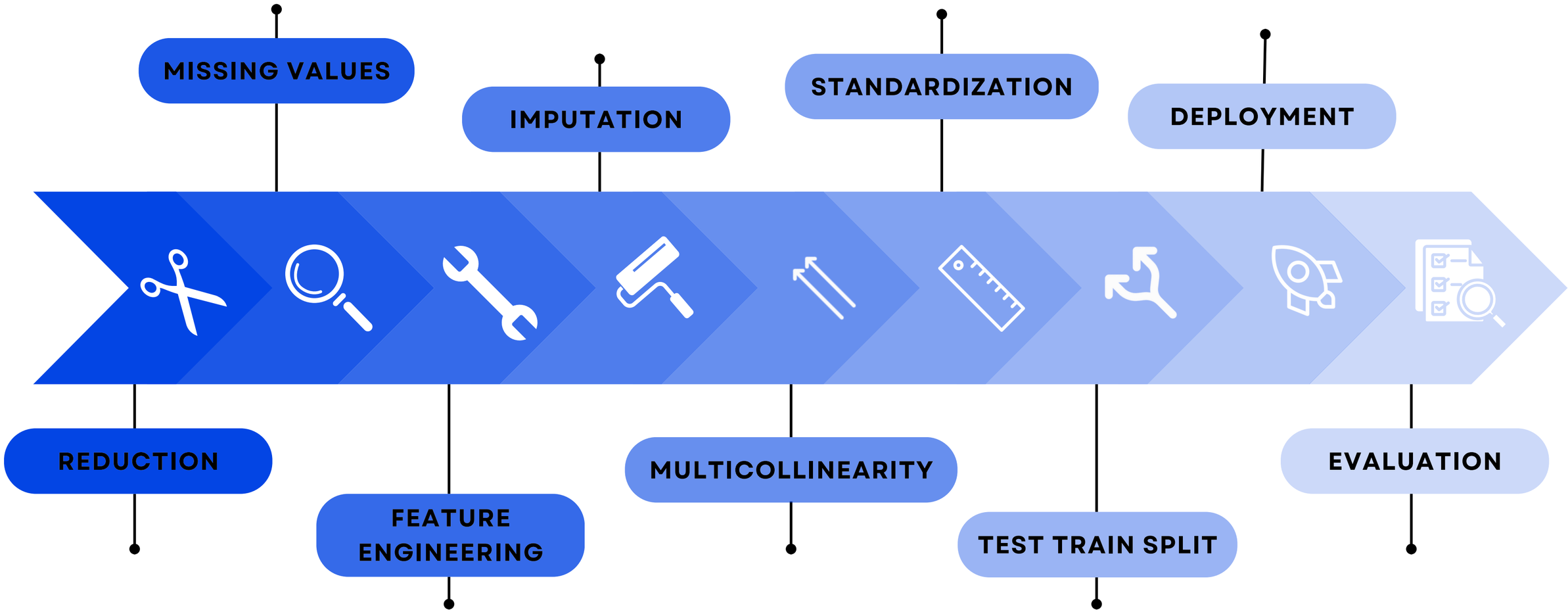
BUSINESS IMPACT

Protect lenders by spotting risky borrowers sooner. Enable smarter lending decisions and loss reduction. Use both **pre-loan** and **post-loan** data to capture real-time borrower actions and **assess risk throughout repayment**.

UNCOVER KEY FEATURES

Which borrower and loan characteristics are most strongly linked to future delinquency? **Strongest signals of risk** can guide decisions for investors and the LendingClub platform.

METHODOLOGY



04

DATA PRE-PROCESSING

Dimensionality Reduction

Missing Values

Remove metadata

Eliminating data leakage

Filtering out joint applications

Missing chunks of info

High nulls

Redundancies

145 features

↓

89 features

↓

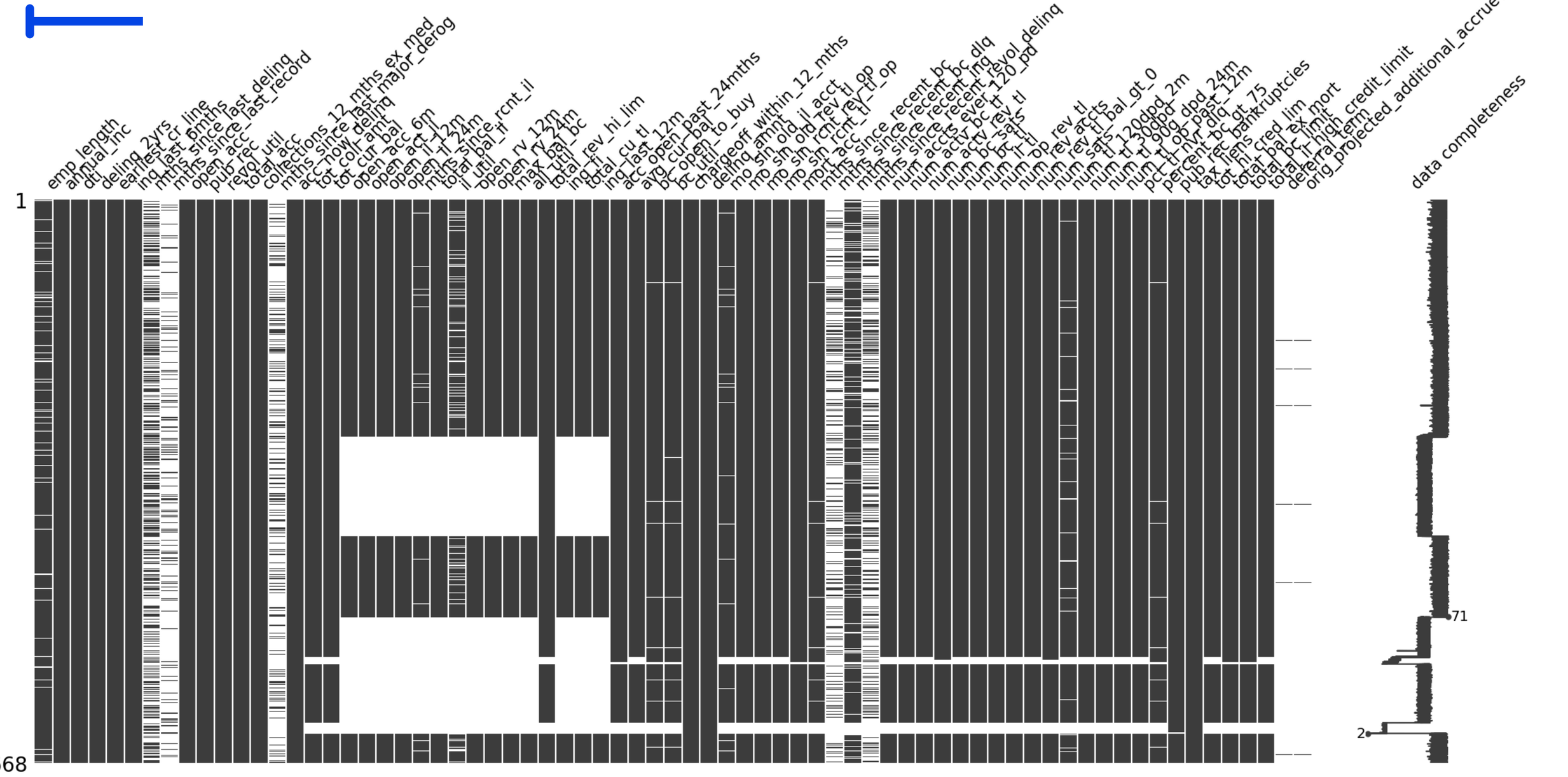
73 features

↓

46 features

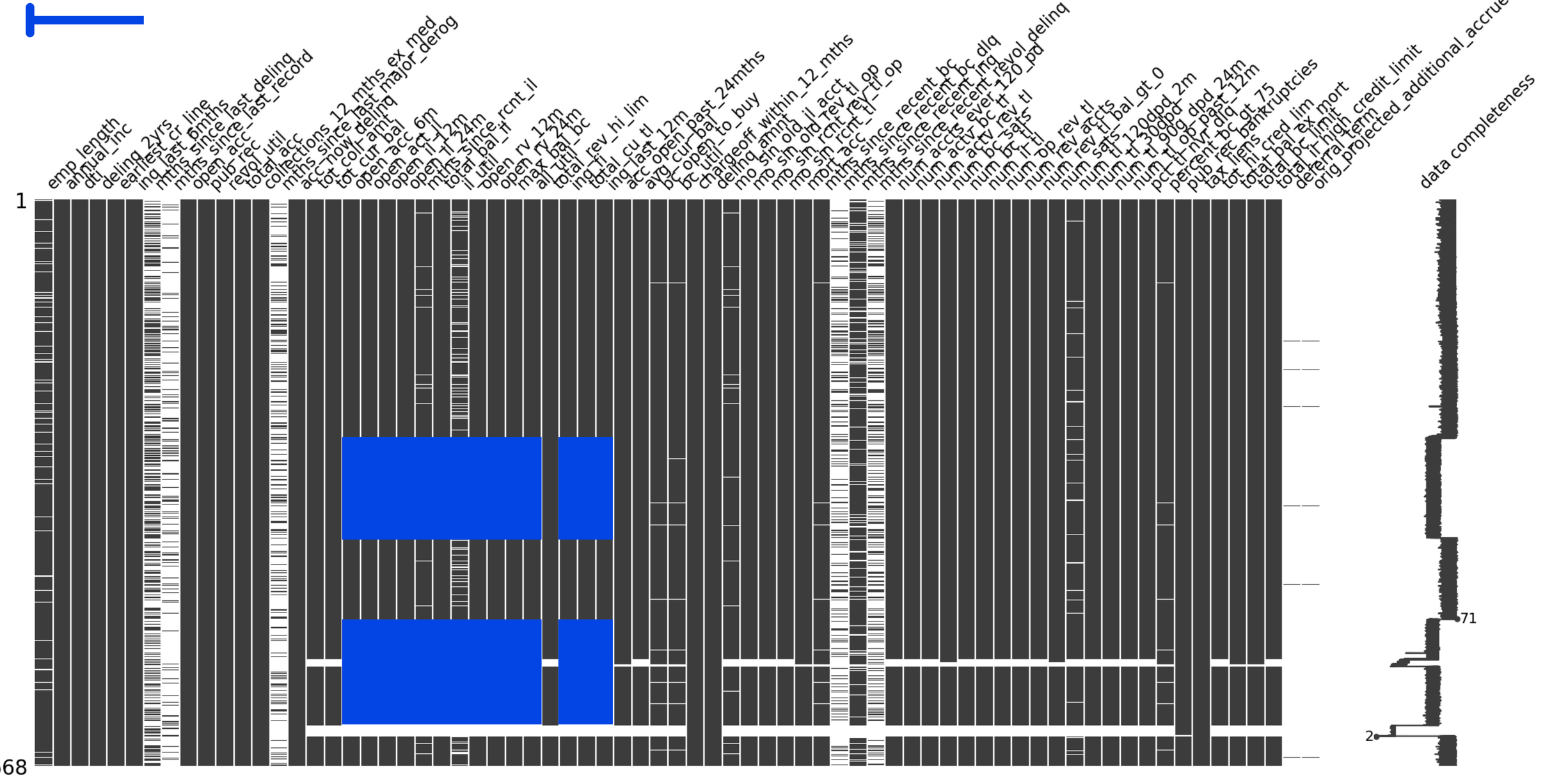
+ 20 dummies

MISSING VALUES MATRIX



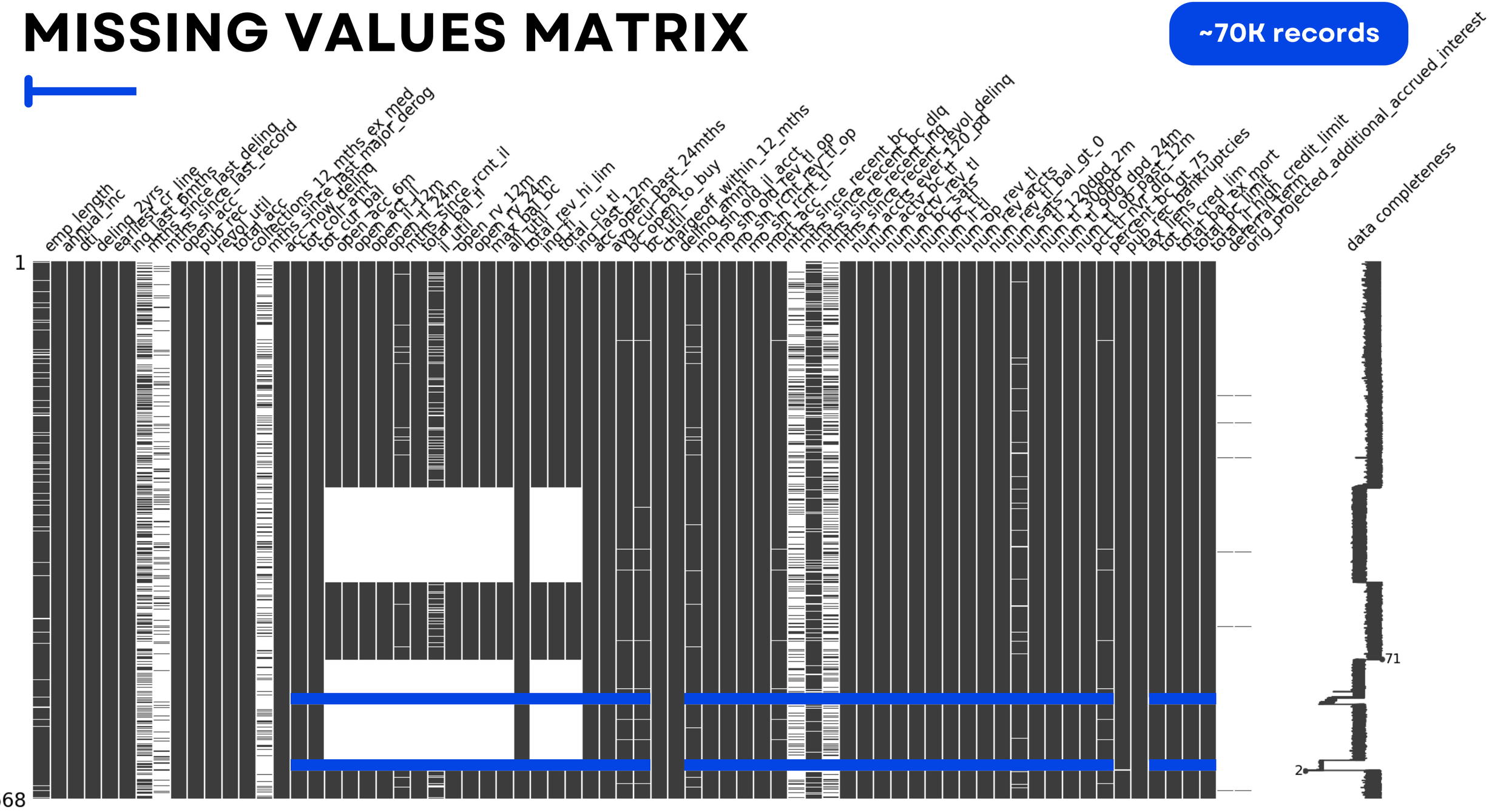
MISSING VALUES MATRIX

14 features

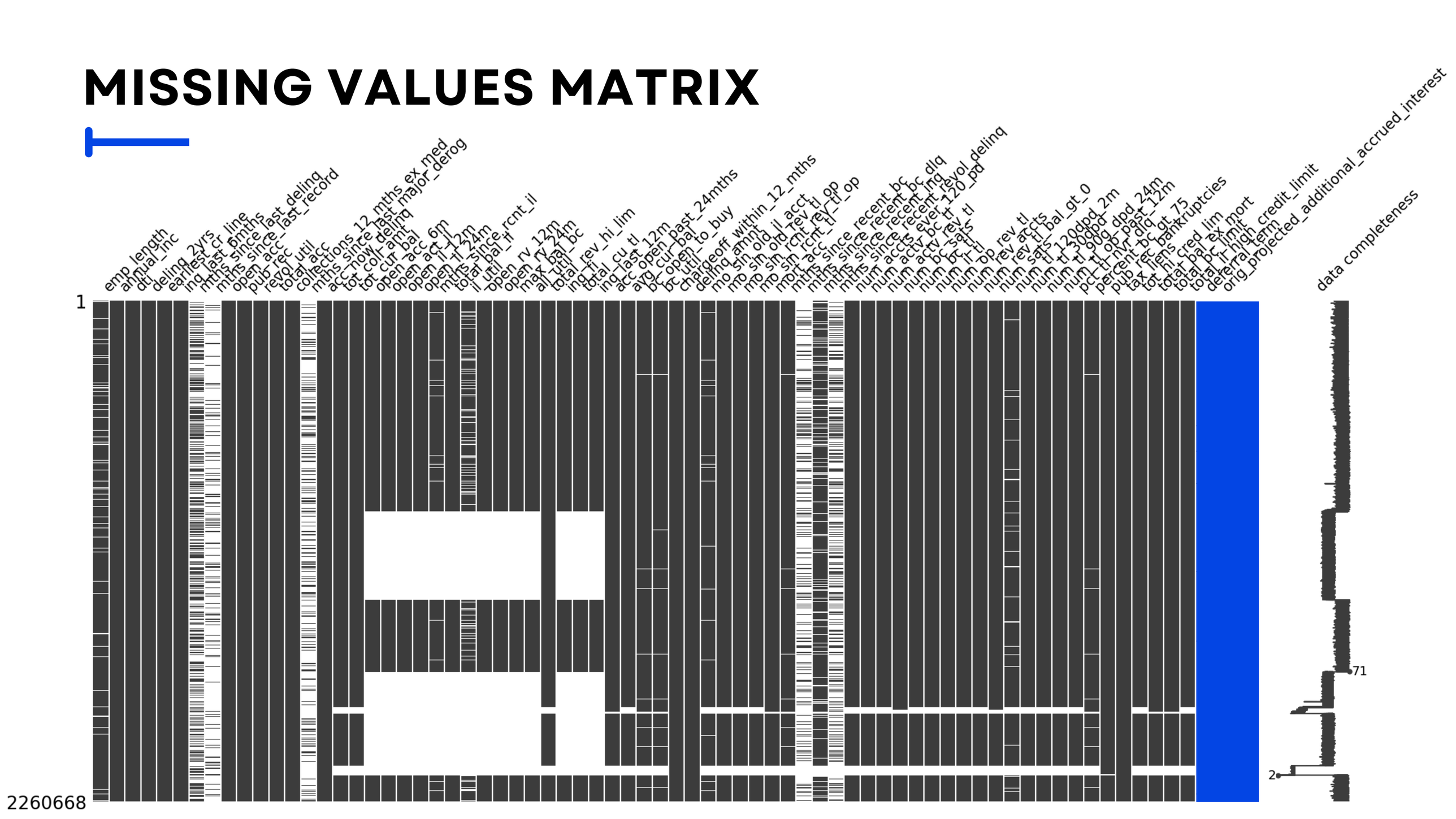


MISSING VALUES MATRIX

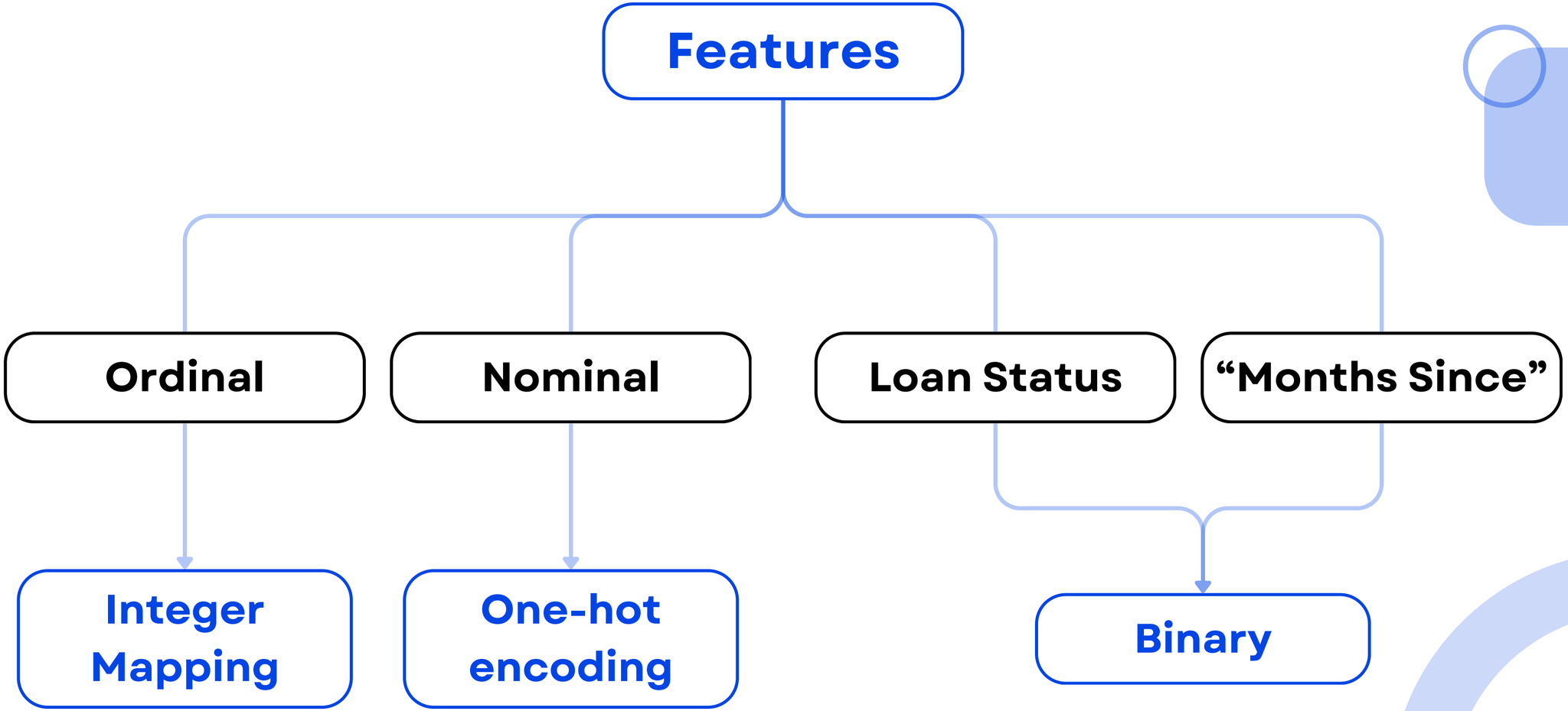
~70K records



MISSING VALUES MATRIX



FEATURE ENGINEERING SUMMARY



05

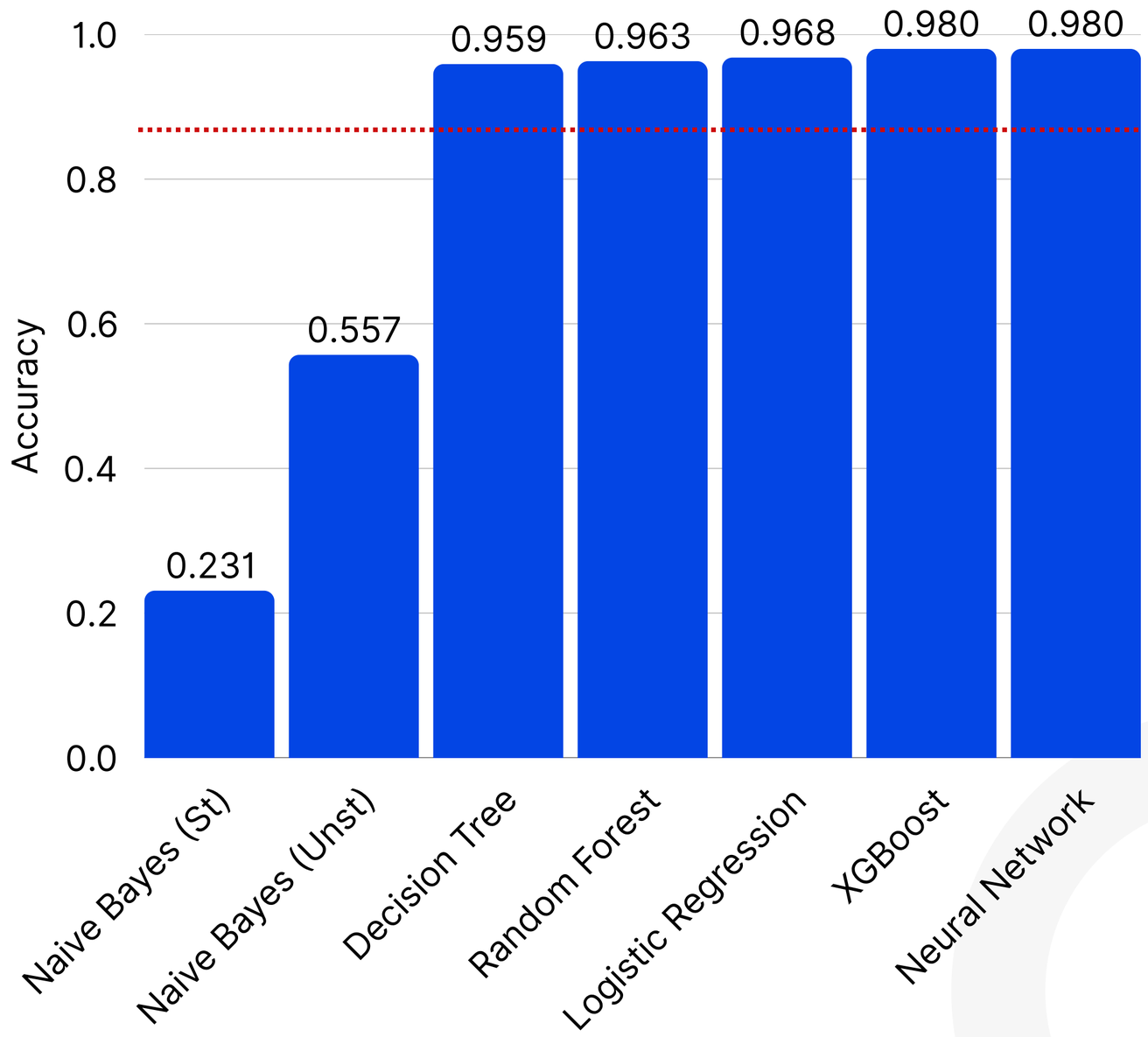
PREDICTIVE MODELS

ACCURACY

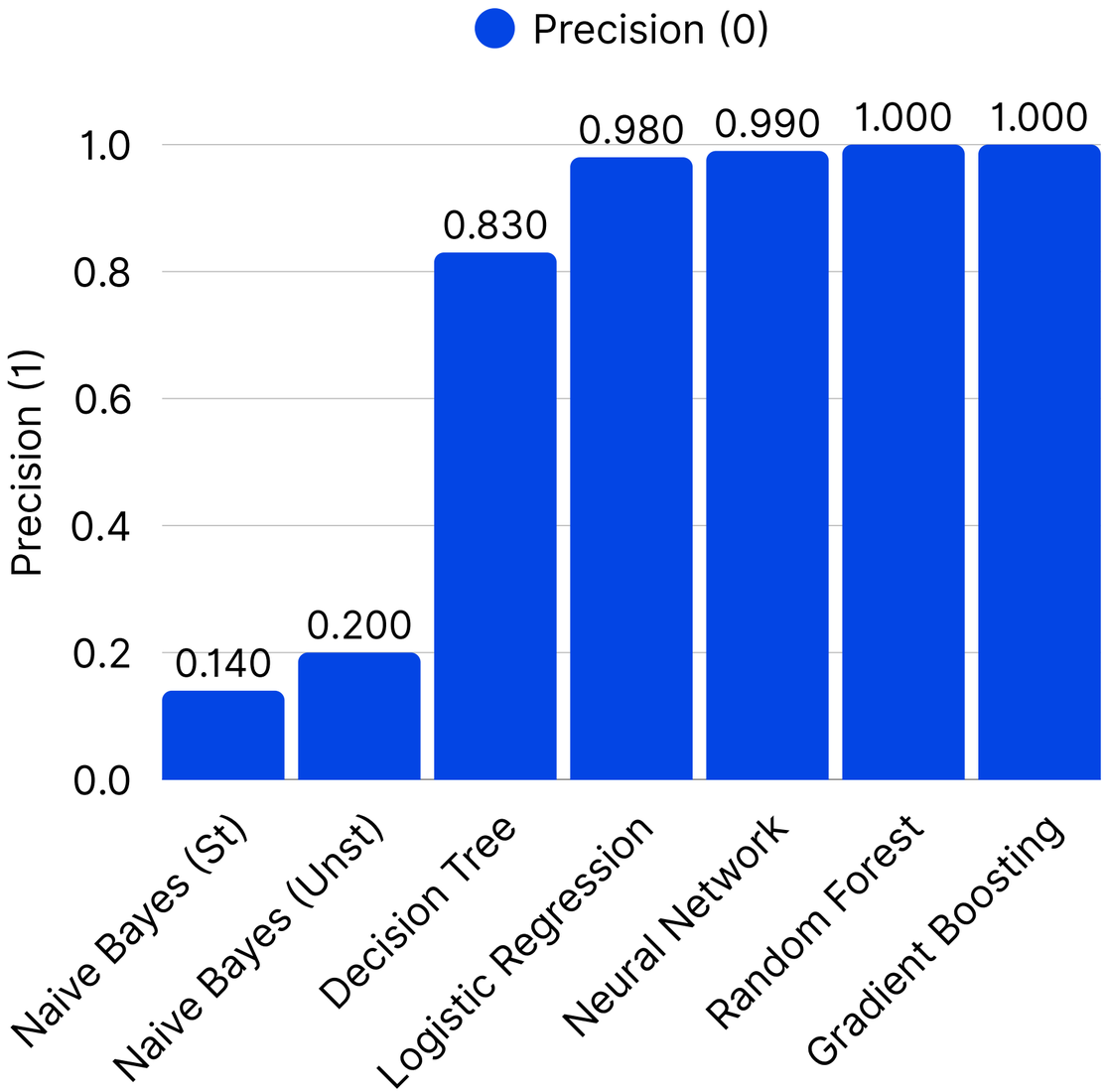
WINNERS: XGBoost and Neural Network

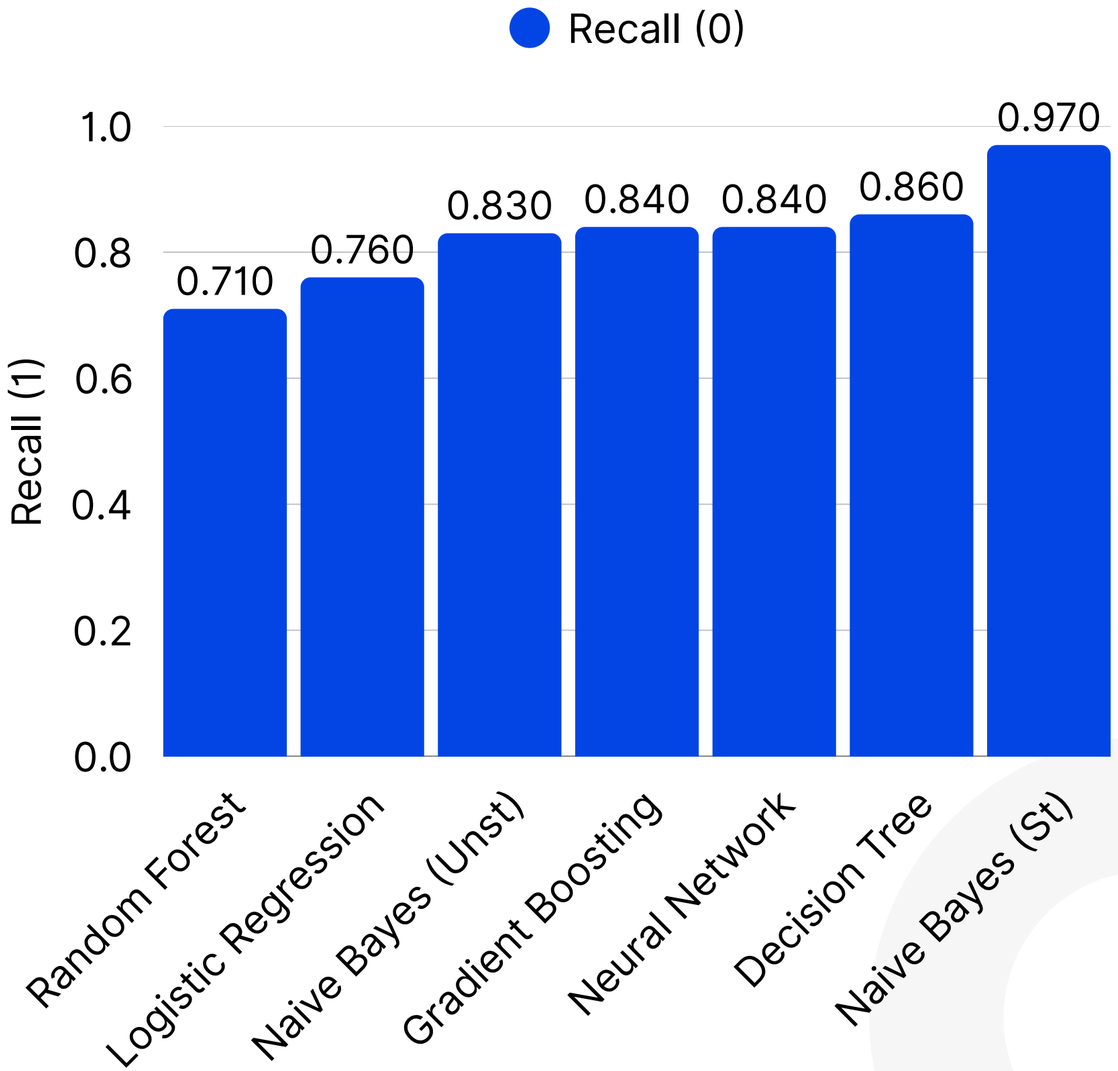
LOSER: Naive Bayes

Other models performed well above baseline, comparable to winners



PRECISION AND RECALL - BAD BORROWERS (0)





Model	Accuracy	Precision (0)	Recall (0)	Precision (1)	Recall (1)	F1 Score
Logistic Regression	96.8%	0.98	0.76	0.97	1.00	0.97
Decision Tree	95.9%	0.83	0.86 ★	0.98 ★	0.97	0.96
Random Forest	96.3%	1.00 ★	0.71	0.96	1.00	0.96
Naive Bayes (Raw)	55.7%	0.20	0.83	0.95	0.52	0.63
Naive Bayes (Standardized)	23.1%	0.14	0.97 ★	0.97	0.12	0.22
Gradient Boosting	98.0% ★	1.00 ★	0.84	0.98 ★	1.00	0.98 ★
Neural Network (NN)	98.0% ★	0.99	0.84	0.98 ★	1.00	0.98 ★

06 CONCLUSIONS

OUR MODEL: XGBOOST

The gradient boosting classifier achieved **98% ACCURACY**, a significant improvement over the 87% baseline. It is a lightweight model that **CATCHES 84% OF RISKY BORROWERS**. Given the top two features, **LOAN AMOUNT** and **TOTAL PAYMENT**, the model can spot underperforming loans early and make smarter **CHARGE-OFF** and recovery decisions.

CHALLENGES

Balancing performance with real world constraints

01 Big data

High dimensionality and slow run times limiting models

02 Imbalanced classes

Not enough delinquent cases

03 Deployment concerns

How useful is a post-issue model that informs investors when to charge off? Preventative models are more ideal, but less accurate

04 Interpretability

Neural networks are complex to a layman

Boston University
BA 305 Final Project

THANK YOU

FOR YOUR ATTENTION

Group B2

Ania Shaheed, Sydney Kennedy,
Amira Zuniga, Sloane Payse