



DEFAULT OF CREDIT CARD CLIENTS

SHELLY LEVY AND ORIAN GANOR



Our Team



SHELLY LEVY

data scientist



ALON ORING

proffesional mentor



ORIAN GANOR

data scientist

OUR PROBLEM

- PREDICT WHETHER A CLIENT WILL DEFAULT ON THEIR CREDIT CARD PAYMENT

OUR PROBLEM

- PREDICT WHETHER A CLIENT WILL DEFAULT ON THEIR CREDIT CARD PAYMENT
- MITIGATE POTENTIAL LOSSES

OUR PROBLEM

- PREDICT WHETHER A CLIENT WILL DEFAULT ON THEIR CREDIT CARD PAYMENT
- MITIGATE POTENTIAL LOSSES
- FEATURES OF CREDIT CARD CLIENT

OUR PROBLEM

- PREDICT WHETHER A CLIENT WILL DEFAULT ON THEIR CREDIT CARD PAYMENT
- FEATURES OF CREDIT CARD CLIENT
- MITIGATE POTENTIAL LOSSES
- EVALUATION:
F1, ACCURACY, PRECISION, RECALL, AUC

DATA DESCRIPTION

- 30000 INSTANCES 25 FEATURES

DATA DESCRIPTION

- 30000 INSTANCES 25 FEATURES
- PERSONAL DATA:
AGE,GENDER,EDUCATION, MARITAL
STATUS

DATA DESCRIPTION

- 30000 INSTANCES 25 FEATURES
- PERSONAL DATA:
AGE,GENDER,EDUCATION, MARITAL
STATUS
- FINANCIAL DATA:
HISTORY OF PAST PAYMENT/DEBT
LIMIT BALANCE

DATA DESCRIPTION

	אפריל	מאי	יוני	יולי	אוגוסט	ספטמבר
Bill_Amt	100	200	1000	400	500	1000
Pay_Amt	0	100	200	1000	0	0
Pay_status	-1	-1	-1	1	2	3

DATA DESCRIPTION

	אפריל	מאי	יוני	יולי	אוגוסט	ספטמבר
Bill_Amt	100	200	1000	400	500	1000
Pay_Amt	0	100	200	1000	0	0
Pay_status	-1	-1	-1	1	2	3

- UNVEILING THE ENIGMATIC CATEGORIES

DATA DESCRIPTION

	אפריל	מאי	יוני	יולי	אוגוסט	ספטמבר
Bill_Amt	100	200	1000	400	500	1000
Pay_Amt	0	100	200	1000	0	0
Pay_status	-1	-1	-1	1	2	3

- UNVEILING THE ENIGMATIC CATEGORIES
- DISTINCTIVE FEATURES

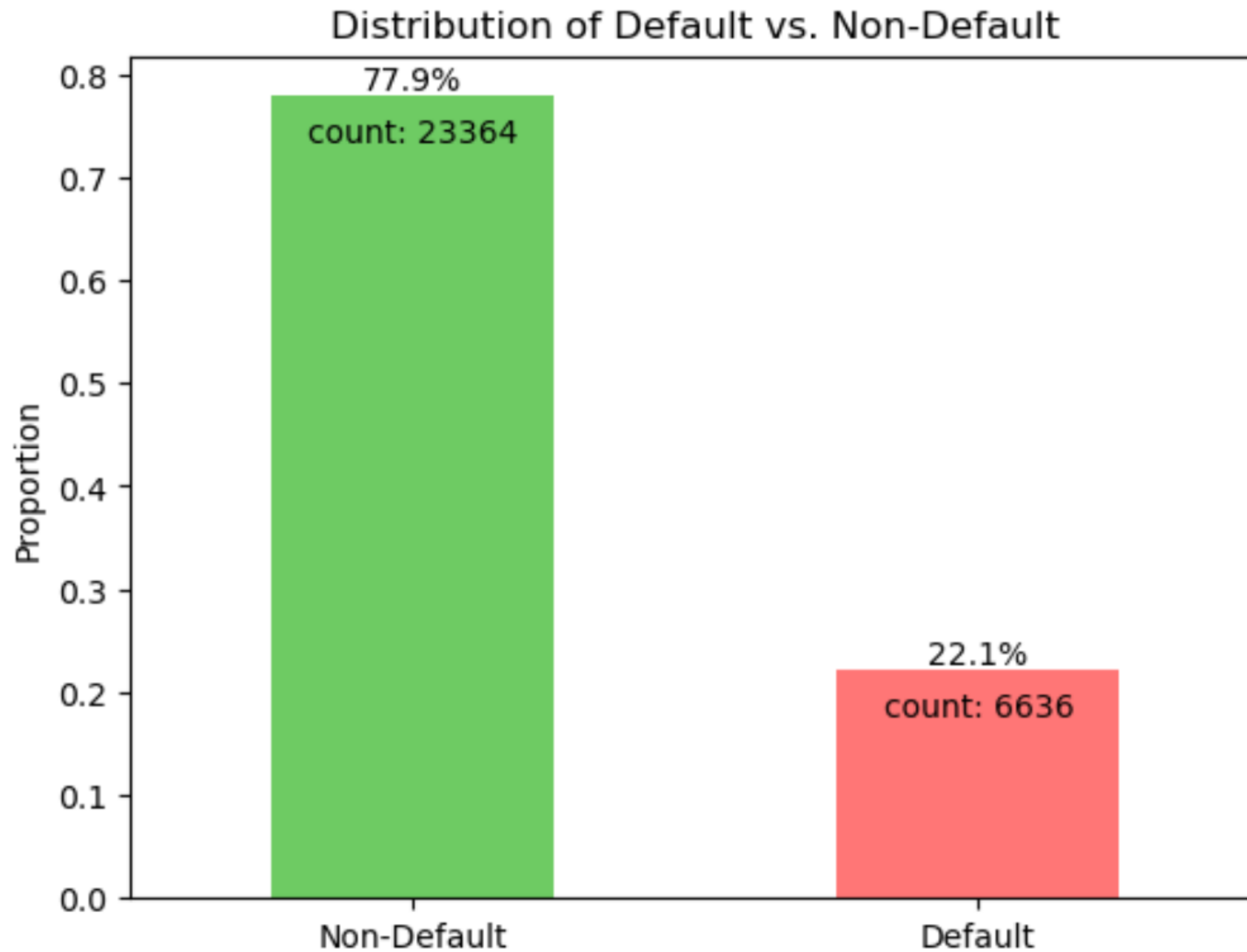
DATA DESCRIPTION

- NA'S: MARRIAGE- 0.18% NA,
EDUCATION - 0.05% NA

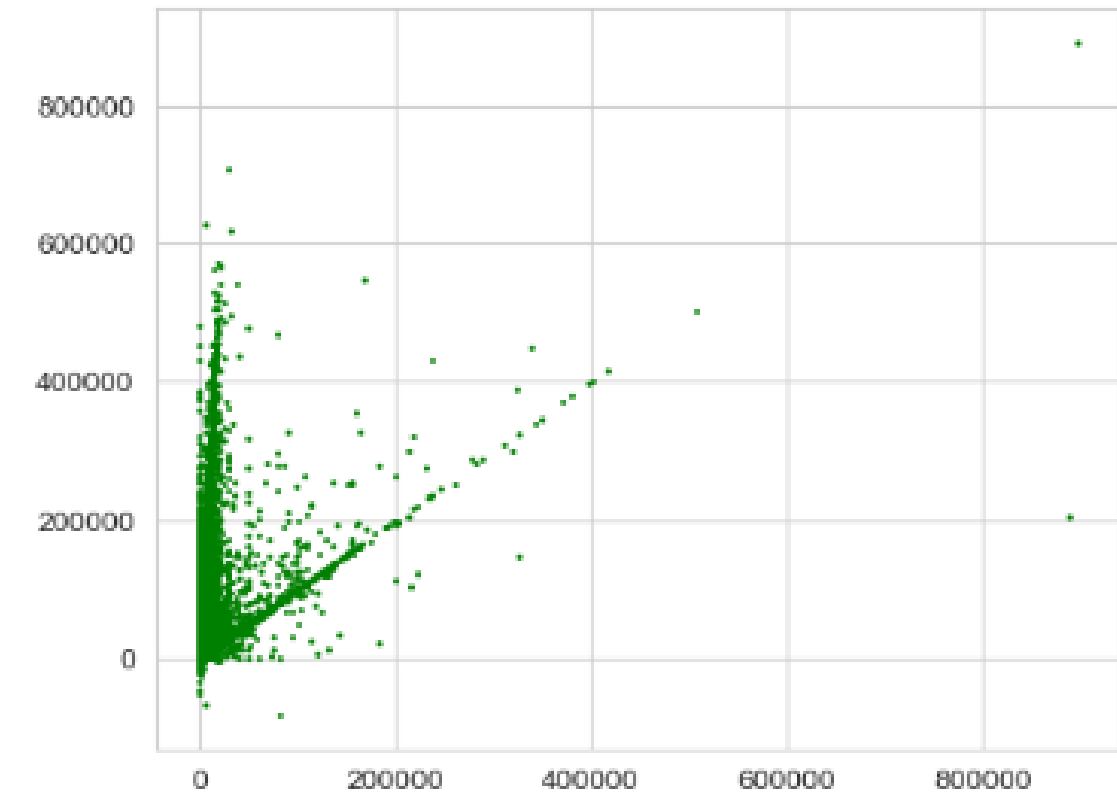
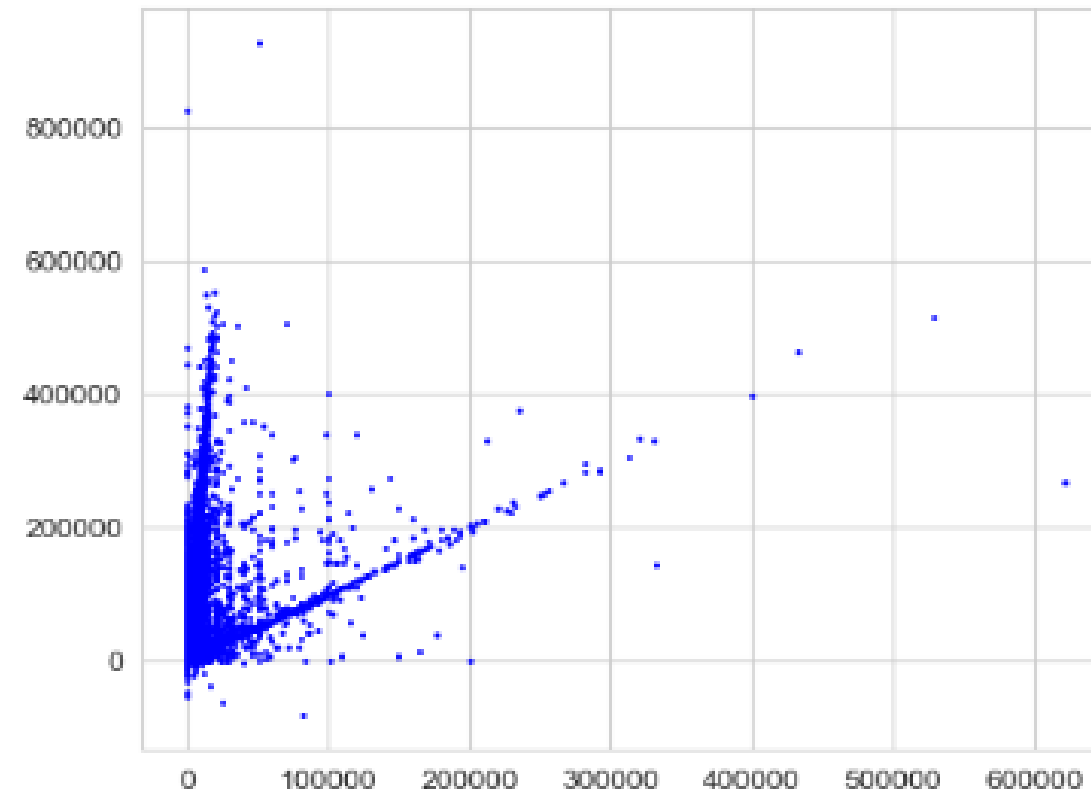
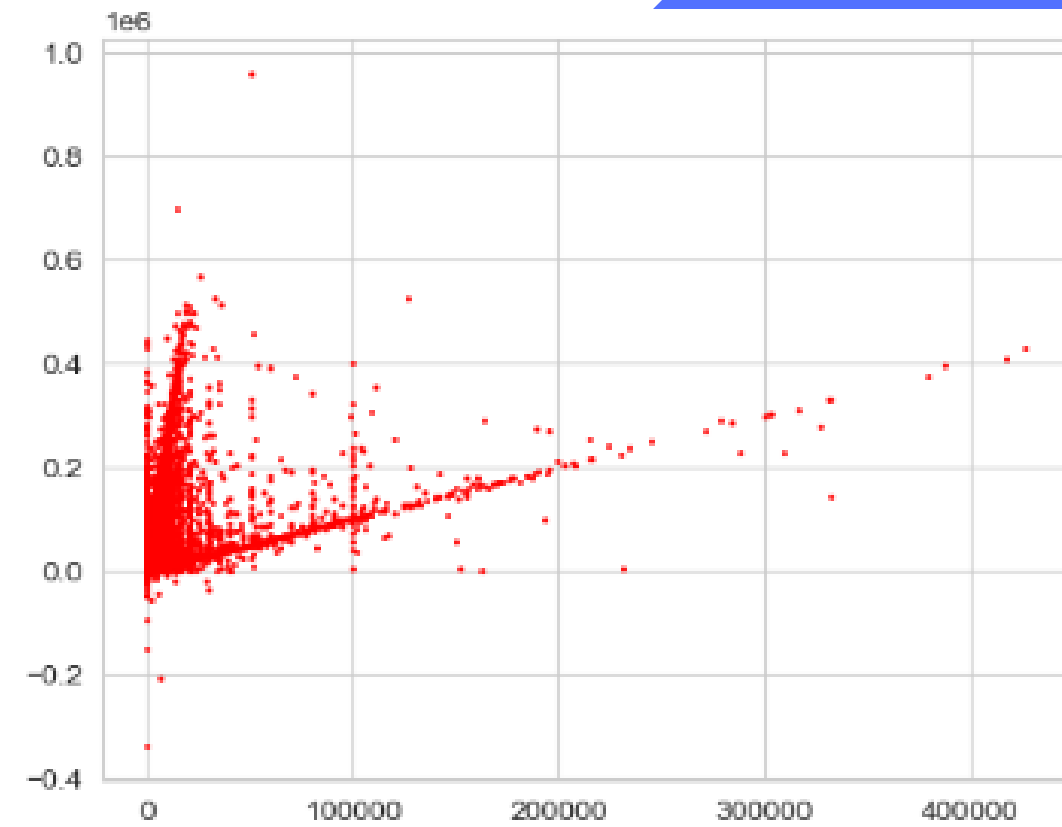
DATA DESCRIPTION

- NA'S: MARRIAGE- 0.18% NA,
EDUCATION - 0.05% NA
- 30% TEST /70% TRAIN

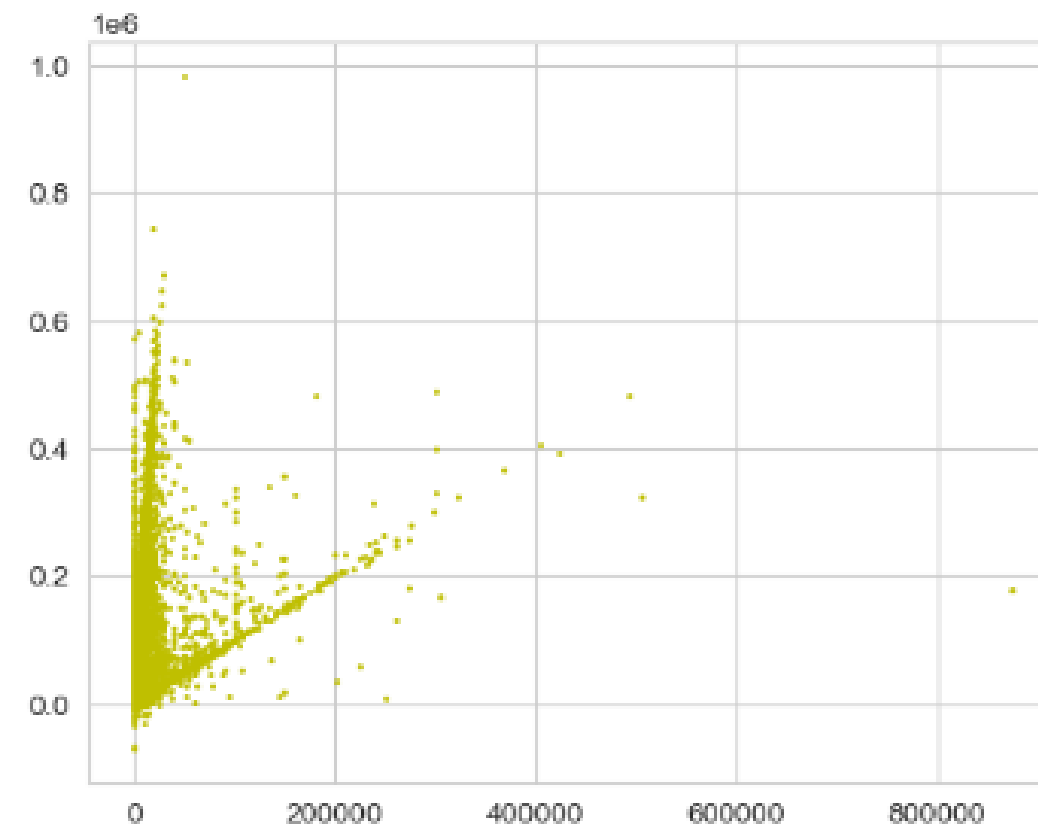
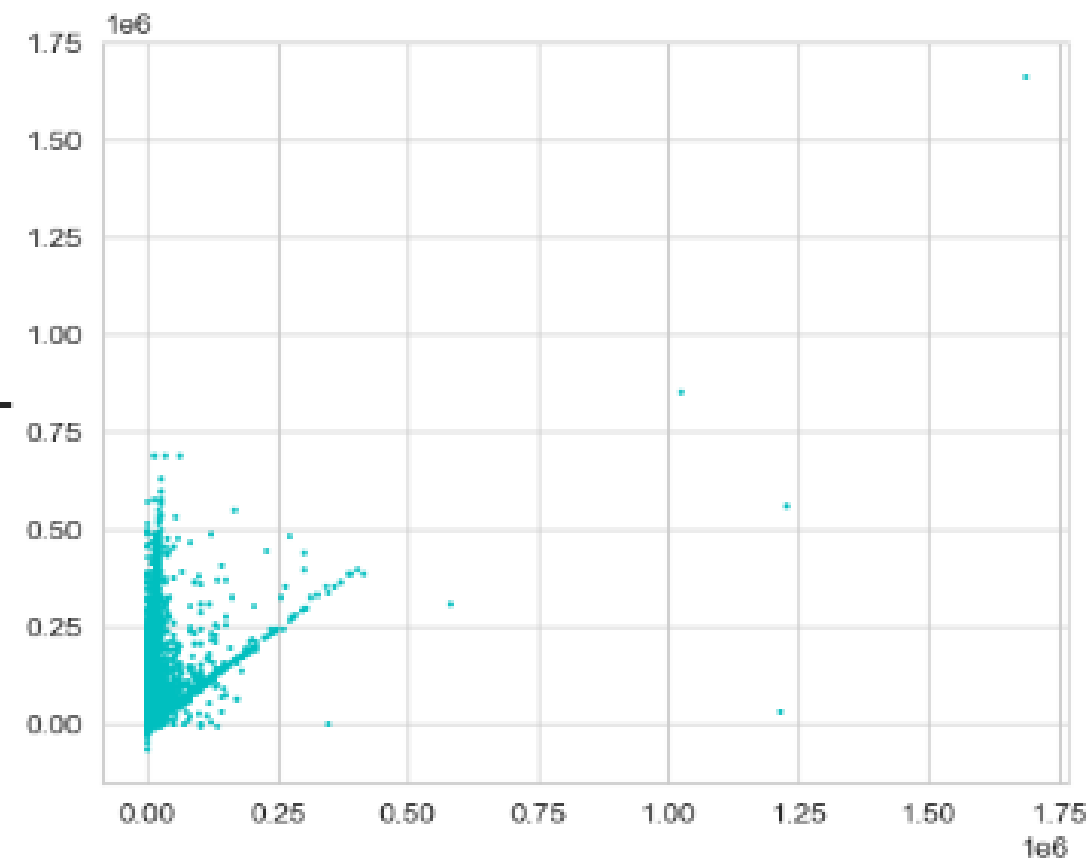
DATA DISTRIBUTION



BILL vs. PAY

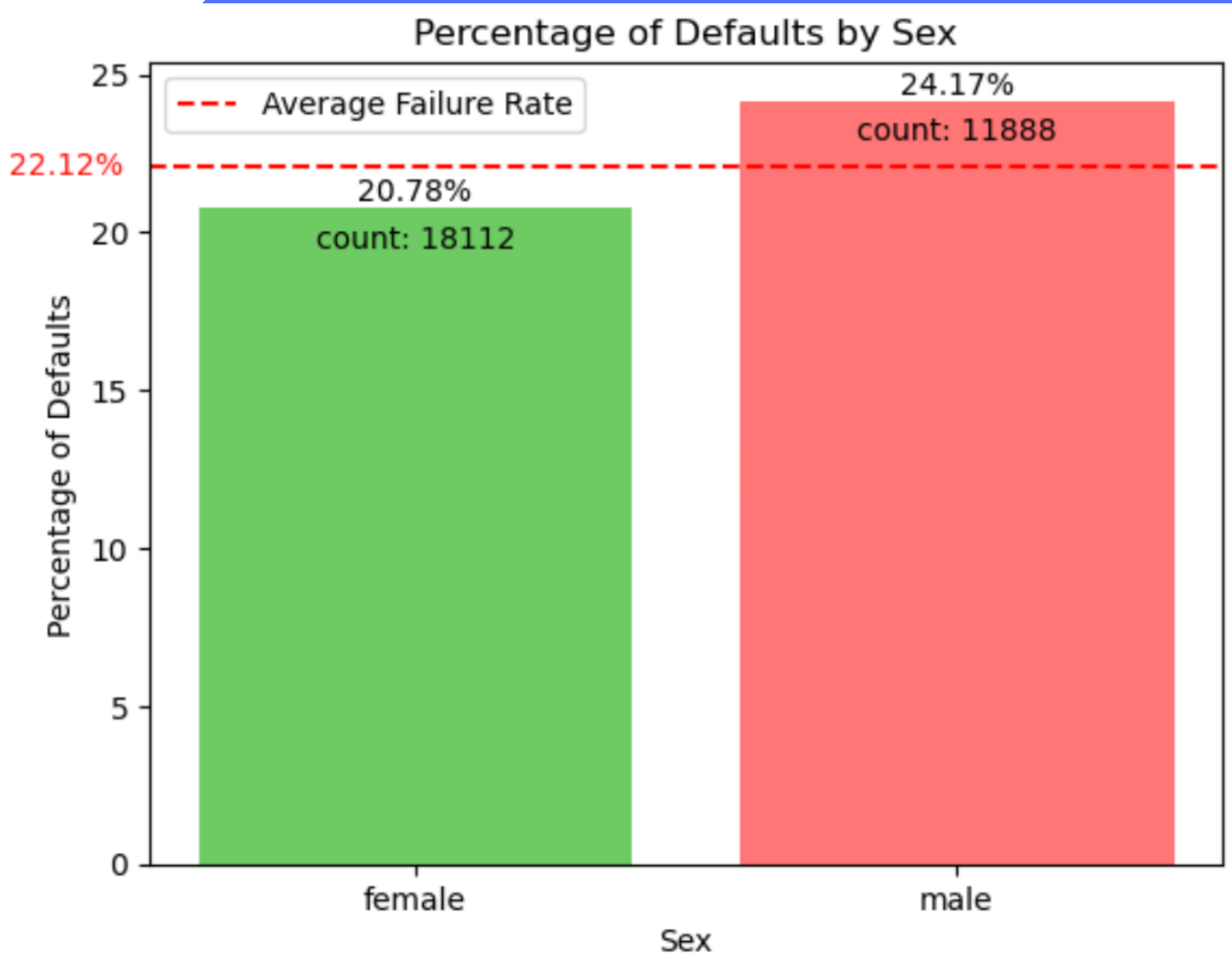


Bill Amount in past 6 months

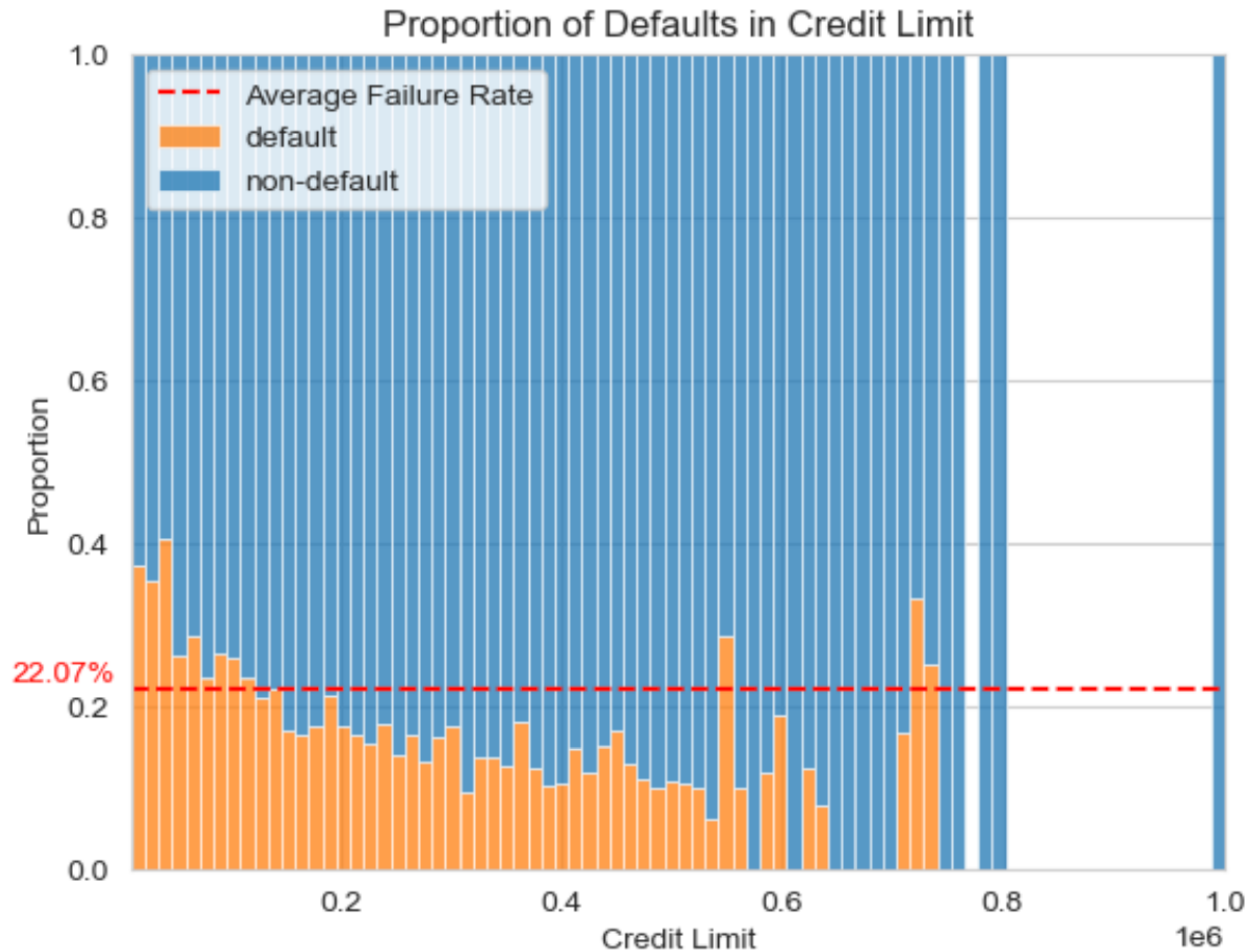


Payment in past 6 months

DEFAULTERS BY GENDER



CREDIT LIMIT DEFAULTERS



DATA ENGINEERING

MISSING VALUES:

- NA's: Marriage- 0.18% NA,
Education - 0.05% NA

DATA ENGINEERING

MISSING VALUES:

- NA's: Marriage- 0.18% NA,
Education - 0.05% NA
- 3 Different Metrics: Remove, Knn, Binning

DATA ENGINEERING

MISSING VALUES:

- NA's: Marriage- 0.18% NA, Education - 0.05% NA
- 3 Different Metrics: Remove, Knn, Binning

FEATURE ENGINEERING:

- 8 new features

DATA ENGINEERING

MISSING VALUES:

- NA's: Marriage- 0.18% NA, Education - 0.05% NA
- 3 Different Metrics: Remove, Knn, Binning

FEATURE ENGINEERING:

- 8 new features
- Late payment (yes/no, count)

DATA ENGINEERING

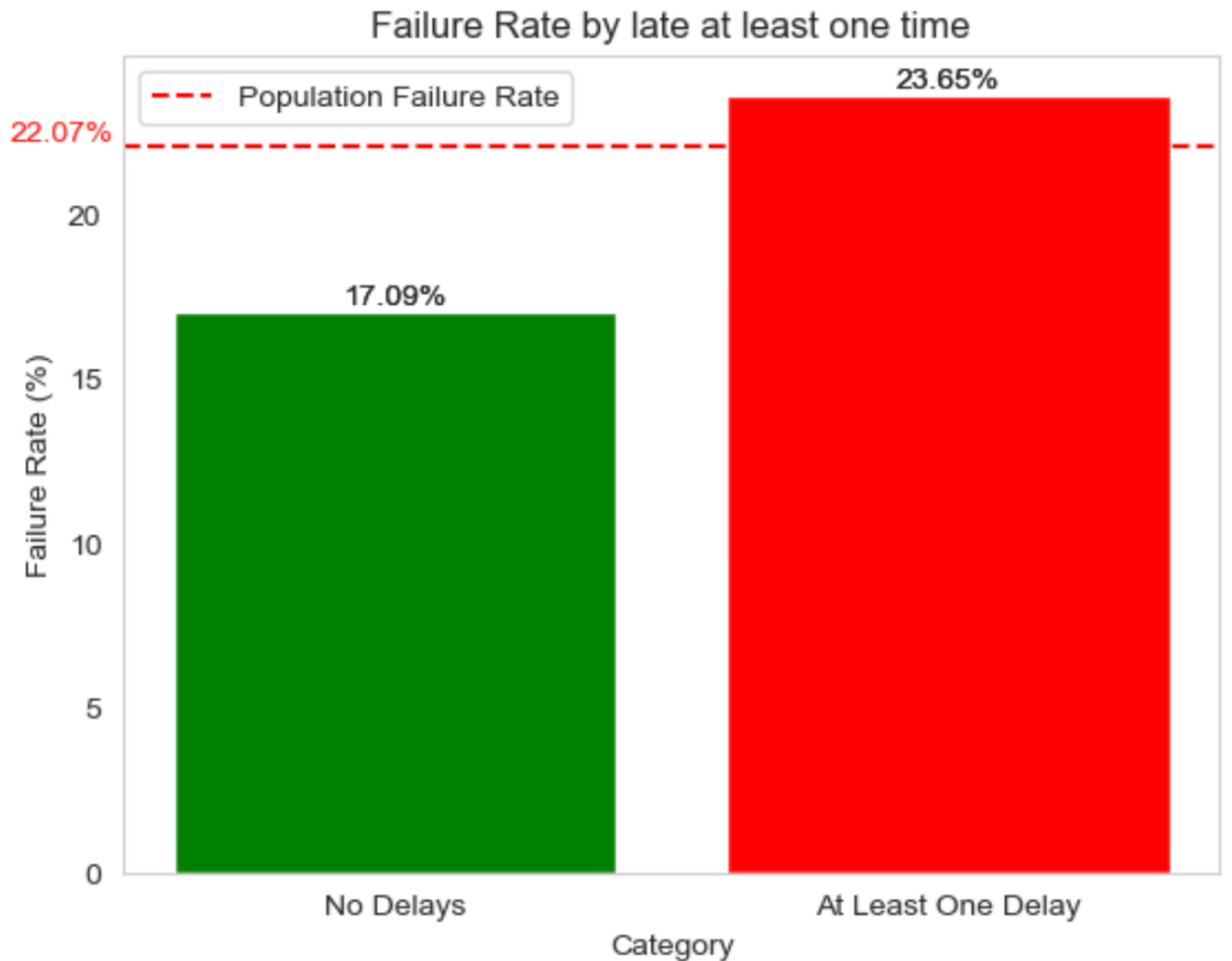
MISSING VALUES:

- NA's: Marriage- 0.18% NA, Education - 0.05% NA
- 3 Different Metrics: Remove, Knn, Binning

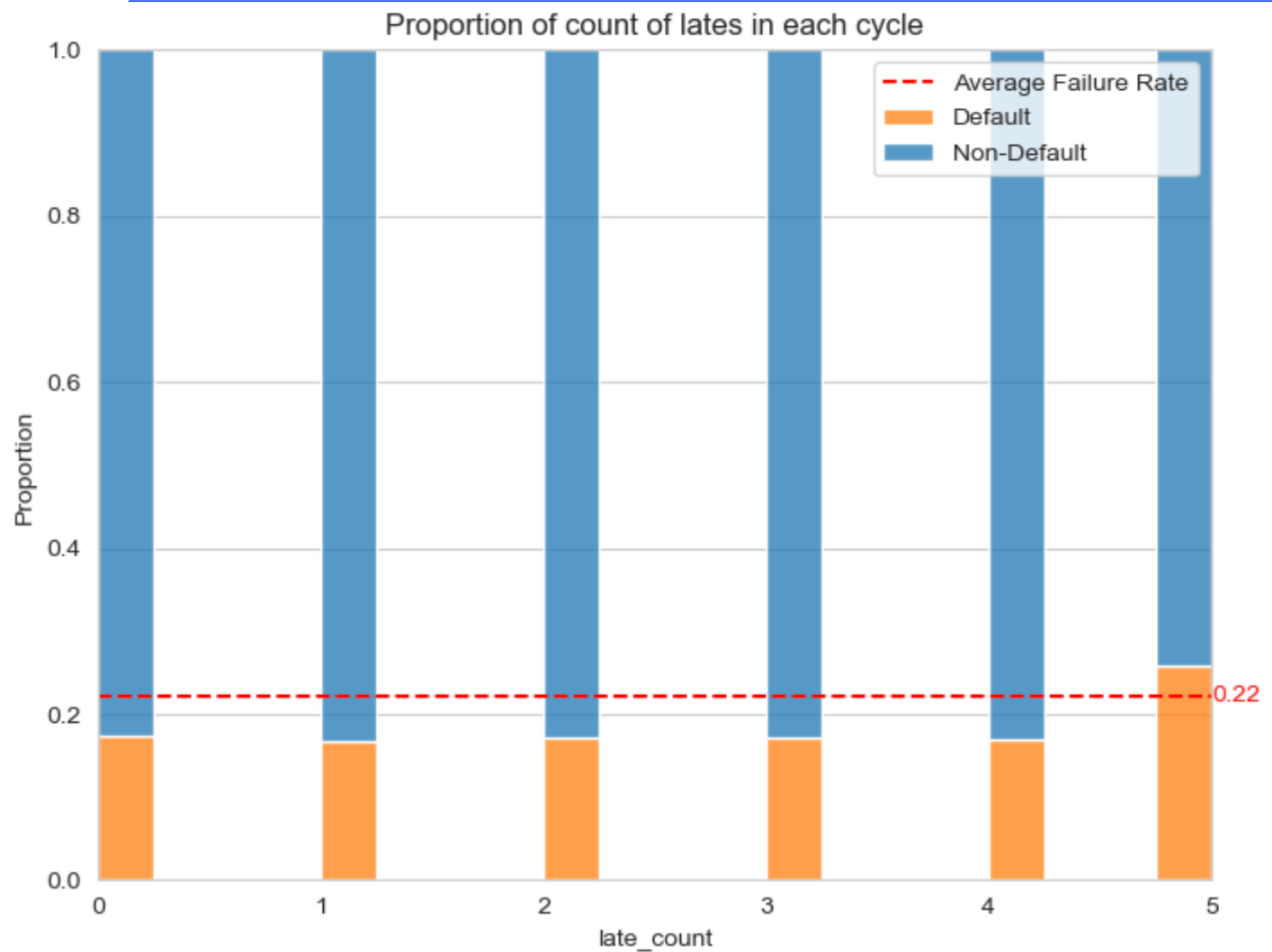
FEATURE ENGINEERING:

- 8 new features
- Late payment (yes/no, count)
- Std of bill/payment

DEFAULT RATE FOR LATE CUSTOMERS



PROPORTION OF LATE COUNT



BASELINE MODEL

- BASELINE MODEL
- MAJORITY OF NON DEFAULT INSTANCES:
 - Training Accuracy: 0.77
 - Test Accuracy: 0.78

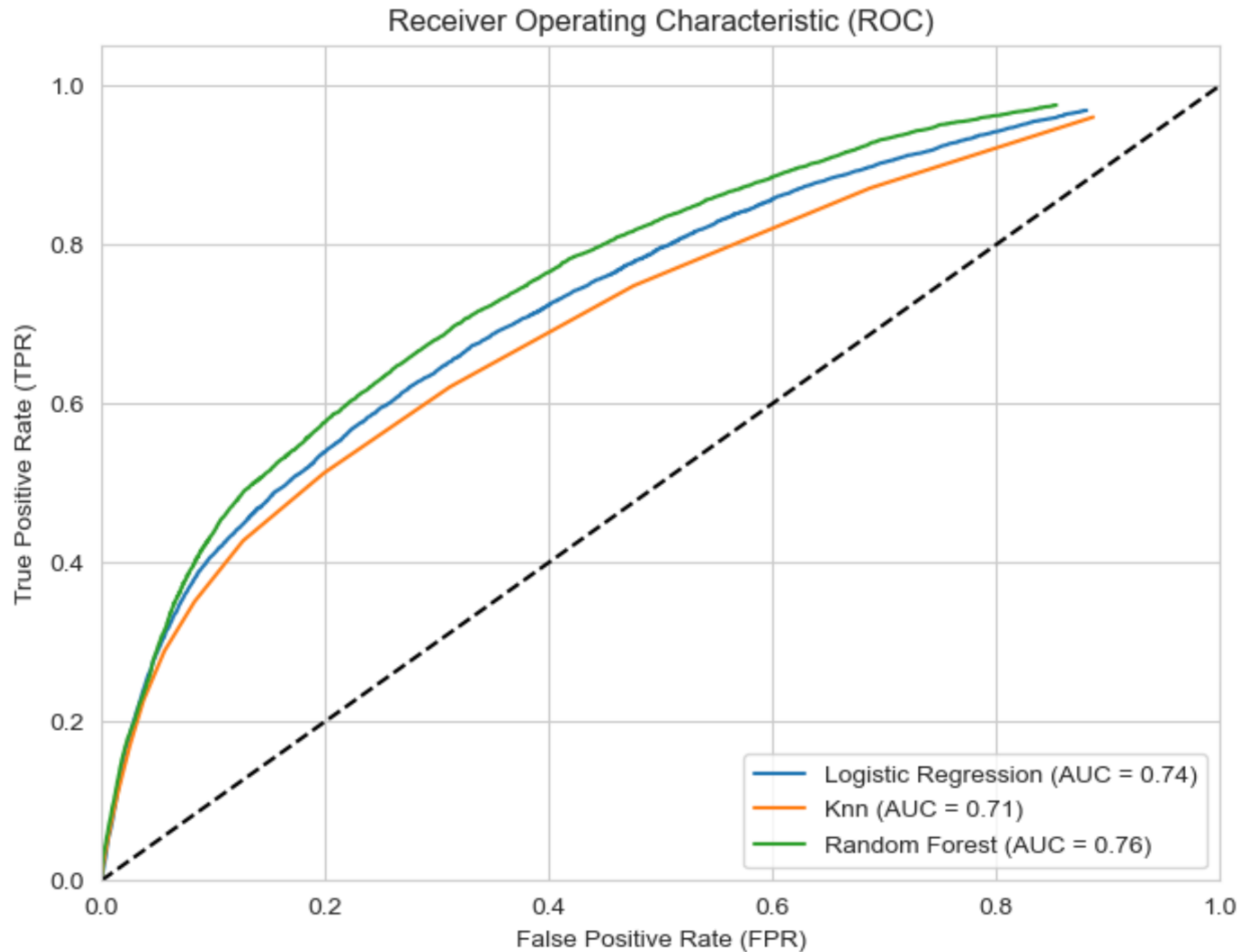
MODELING

- Knn, Logistic Regression, Random Forest, Ada-Boost

MODELING

- Knn, Logistic Regression, Random Forest, Ada-Boost
- **Hyper Parameter Tuning**
- **Feature Selection:** Forward, Backward, Bootstrapping, Random Selection
- **Scalling Metrics:**
Min-Max, Robust, Standard
- **Evaluation:** K-folds, F1 socre, AUC
- **Dealing With Imbalance:** boost strap, class weight
- Ada-Boost with the best model

ROC CURVE



TIME IMPACT ANALYSIS

4 Months

Metric	Value
F1	0.783383
recall	0.808244
precision	0.785415
accuracy	0.808244

6 Months

Metric	Value
F1	0.776789
recall	0.802819
precision	0.779792
accuracy	0.802819

OUR NEXT STEPS

- HYPER PARAMETER TUNING
- EVALUATE DIFFERENT TRAIN/TEST RATIO
- PCA
- EVALUATE MODEL ON DIFFERENT TIME SNAPSHOTS



QUESTIONS?