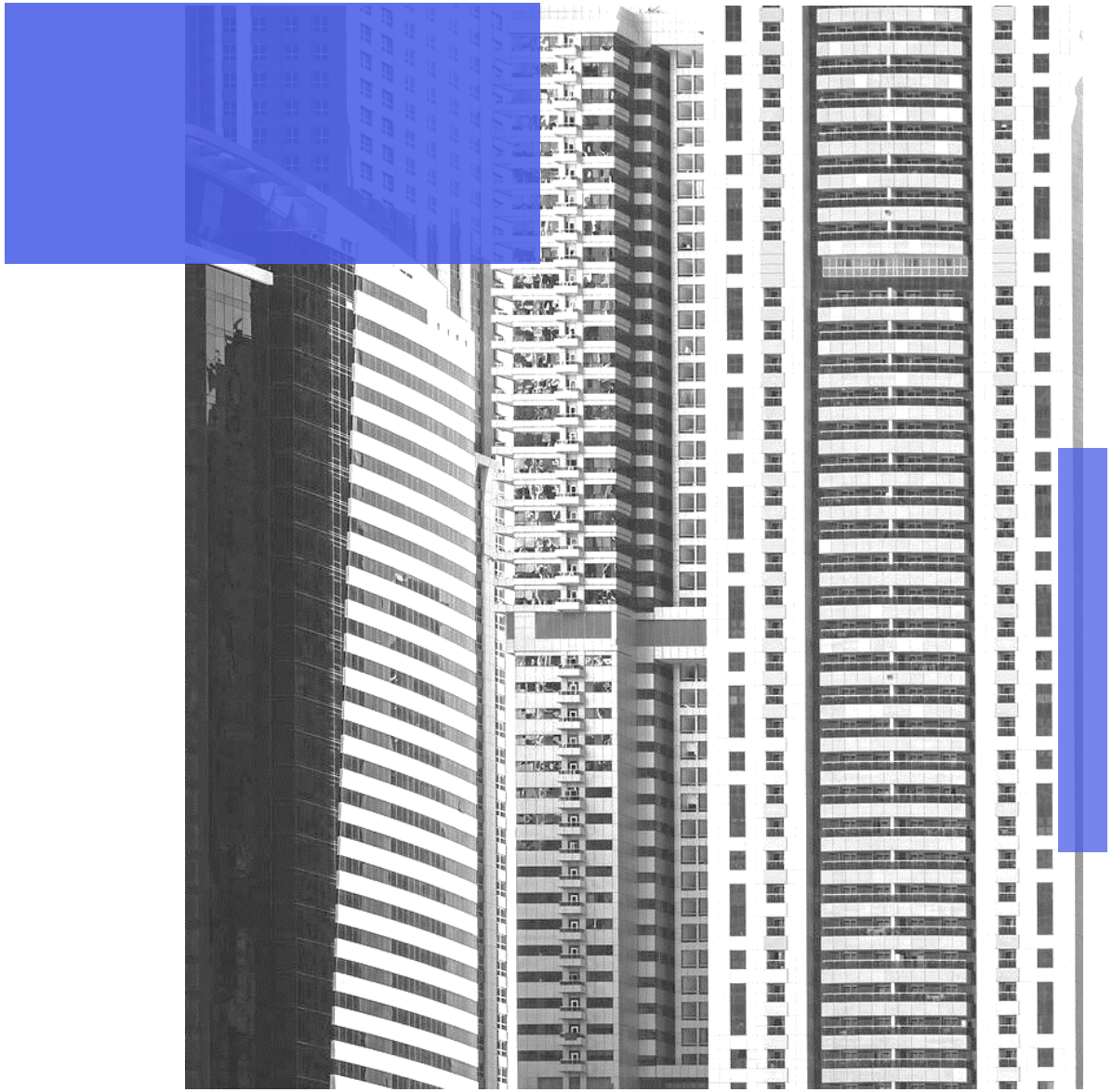
CREDIT CARD SEGMENTATION PROFITABILITY ANALYSIS



AGENDA

PROBLEM DESCRIPTION

DATA ANALYSIS

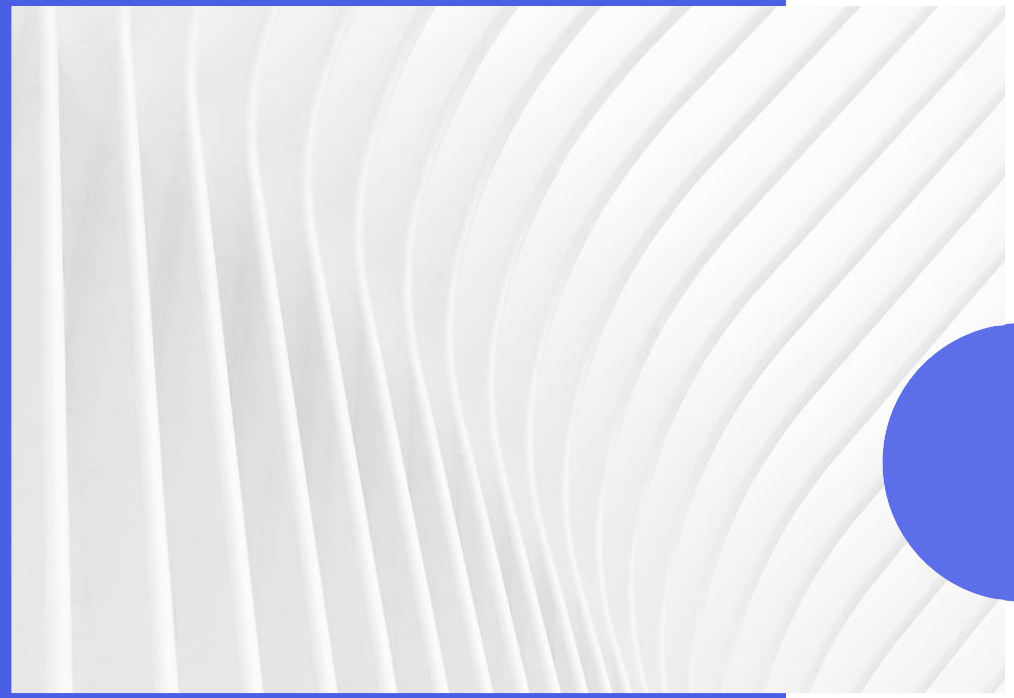
ASSUMPTIONS

METHODOLOGY

EVALUATION

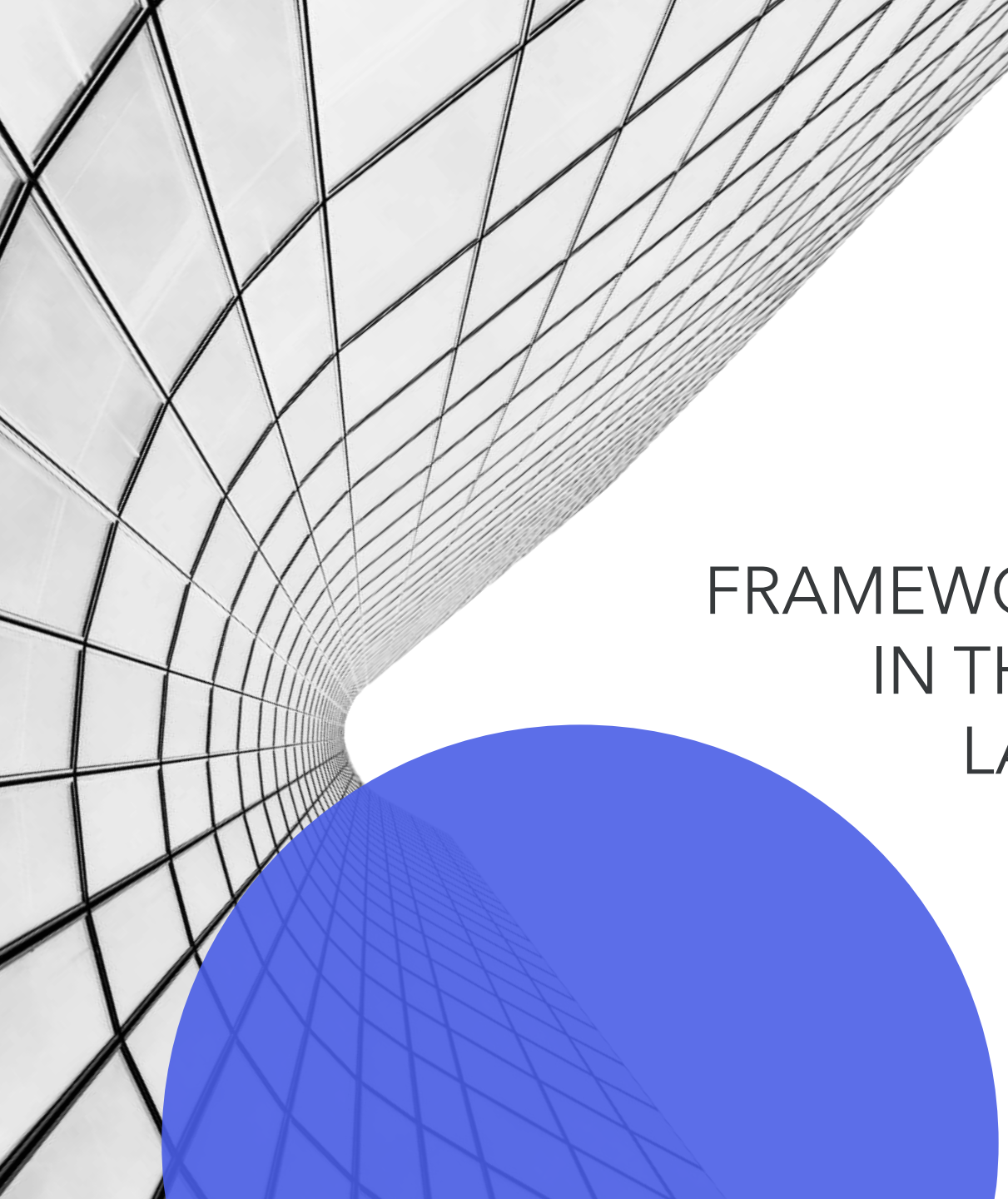
CONCLUSION

PROBLEM DEFINITION



The primary driver of commercial bank failures in recessionary high interest rate environments are losses and weak loan demand.

Consumer credit card debt amounted to $841B in 2022. Credit loans remain to be lucrative to banks given their comparatively high interest rates.



*the **objective***

PROVIDE A **RISK DRIVEN** FRAMEWORK FOR ISSUING CREDIT IN THE **CURRENT** ECONOMIC LANDSCAPE TO **MINIMIZE LOSSES**

*the **originality***

BRING **TRANSPARENCY** TO **BLACK BOX**
APPLICANT APPROVAL SIZING MODELS

*the **contribution***

MODERNIZE OUTDATED METHODS
CURRENTLY IN INDUSTRY

DATASET

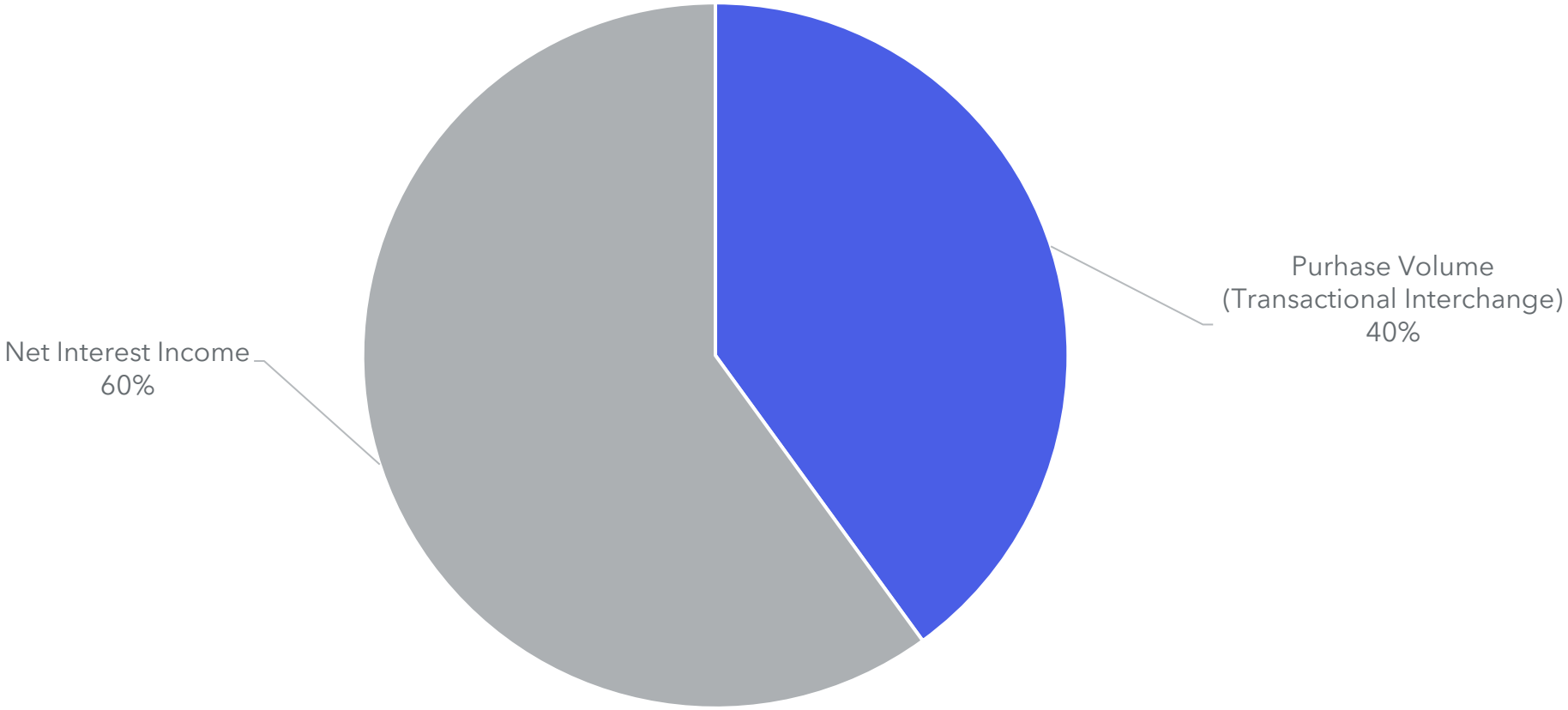
application_record.csv

ID	Client Number
CODE_GENDER	Gender
FLAG_OWN_CAR	Owns car?
FLAG_OWN_REALITY	Owns property?
CNT_CHILDREN	Number of children
AMT_INCOME_TOTAL	Annual income
NAME_INCOME_TYPE	Income category/source
NAME_EDUCATION_TYPE	Education level
NAME_FAMILY_STATUS	Marital status
NAME_HOUSING_TYPE	Way of living
DAYS_BIRTH	Age
DAYS_EMPLOYED	Days Employed
FLAG_MOBIL	Mobile phone?
FLAG_WORK_PHONE	Work phone?
FLAG_PHONE	Phone?
FLAG_EMAIL	Email?
OCCUPATION TYPE	Occupation category
CNT_FAM_MEMBERS	Family Size

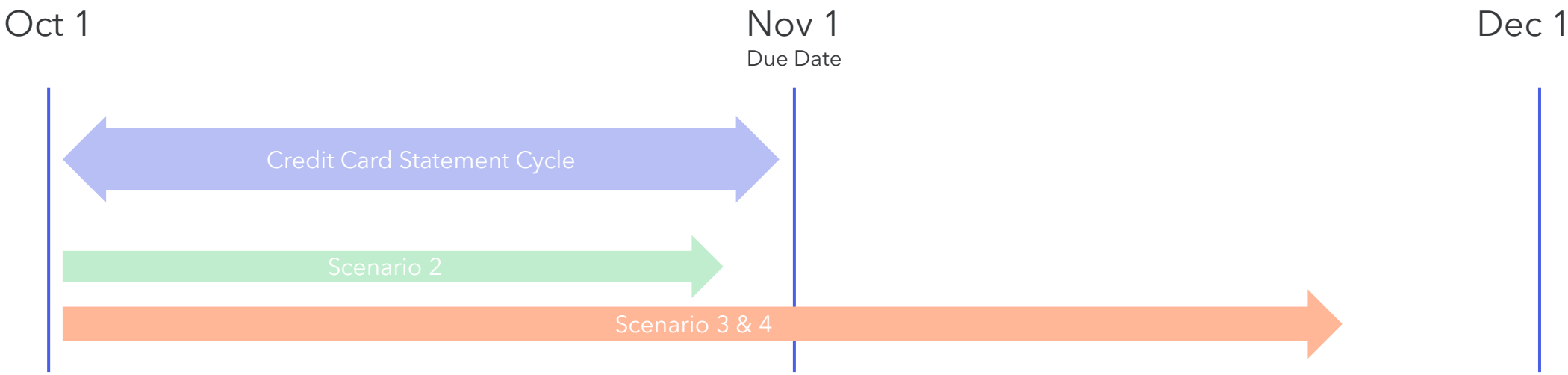
credit_record.csv

ID	Client Number
MONTHS_BALANCE	Number of months on book
STATUS	No loan / loan paid off / overdue

HOW CREDIT CARDS MAKE MONEY



NET INTEREST INCOME (60%)



Scenario 1: No transactions made during the statement period. No interest income.
Scenario 2: Full balance paid by due date, hence no interest charged. No interest income.
Scenario 3: Interest assessed on overdue balance and compounds for subsequent cycles until paid off.
Scenario 4: Write off clients as bad debts after many cycles, resulting in a net loss.

CLIENT PROFILING

Scenario 1

No Purchases Made

$0 Revenue

Scenario 2

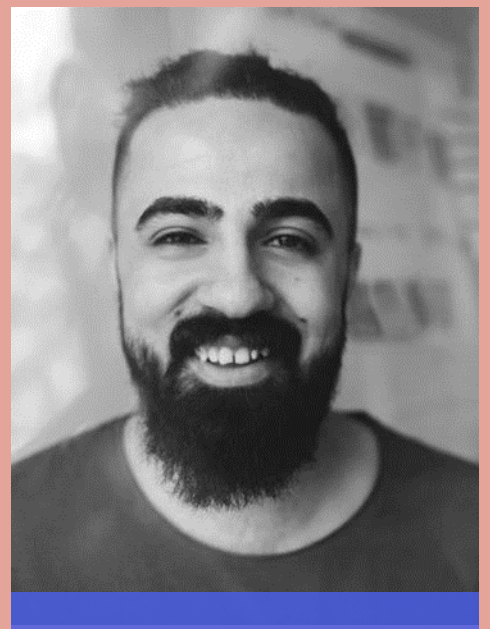
Balance Fully Paid Off

Interchange

Scenario 3

Late Payment (Overdue)

Interchange + Net Interest



Scenario 4

No Payment (Write off)

Net loss

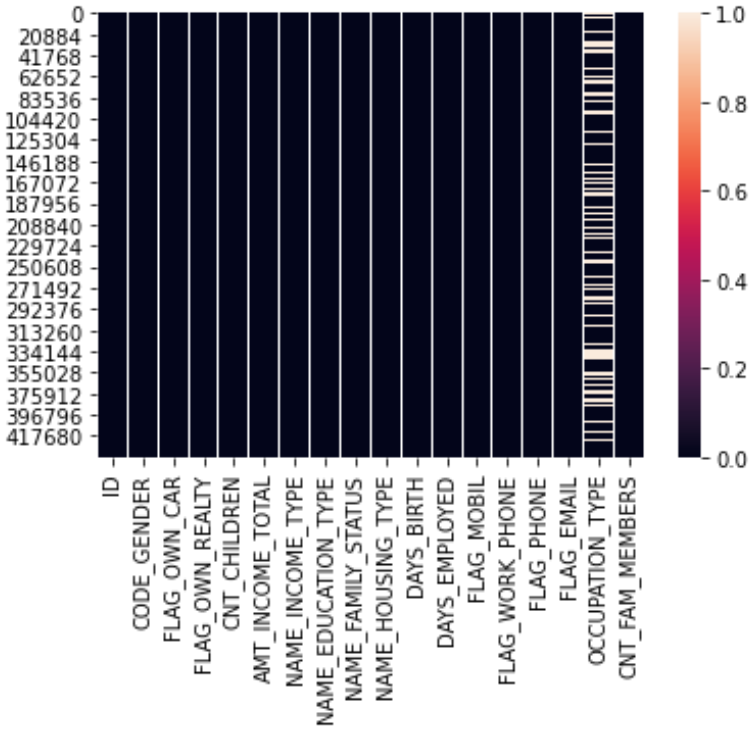
ASSUMPTIONS

ADDITIONAL REVENUE	COSTS	SPEND CONTROL	LOSSES
Finance Charges	Cost of Funds	Balance Utilization (Interest Income)	Unit Write Off Rate
Earnings on Capital	**Management Fees**	**Limit Increases**	**Balance Control Ratio**
Balance Transfers	Marginal Operating Expenses	Purchase Utilization (Interchange Income)	Loss Assumption Factors
Transaction Fees	**Cost of Rewards**		
Over-limit Fees	Insurance		
Annual Fees			

Modelled as zero/constant.

EXPLORATORY ANALYSIS

UNIVARIATE



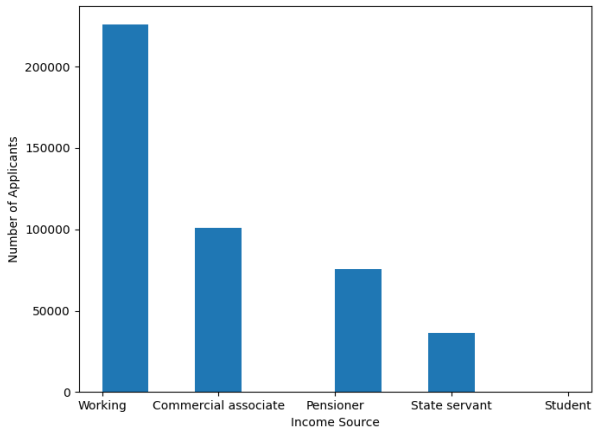
Applicant Record NULL CHECK

Credit Record NULL CHECK

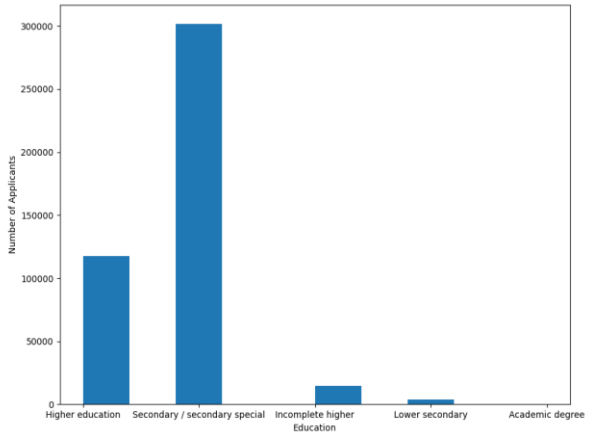
Only feature missing values is Occupation Type.

EXPLORATORY ANALYSIS

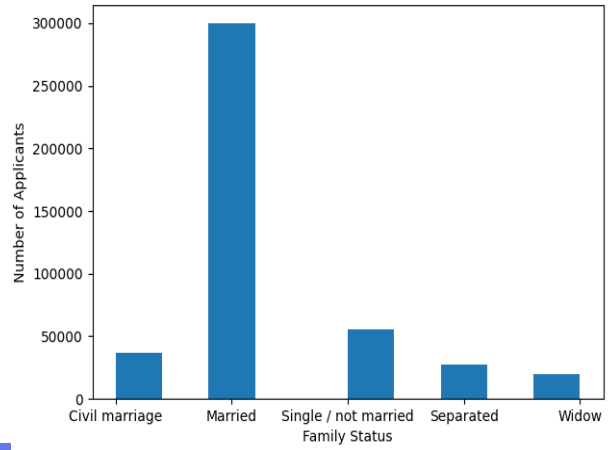
UNIVARIATE



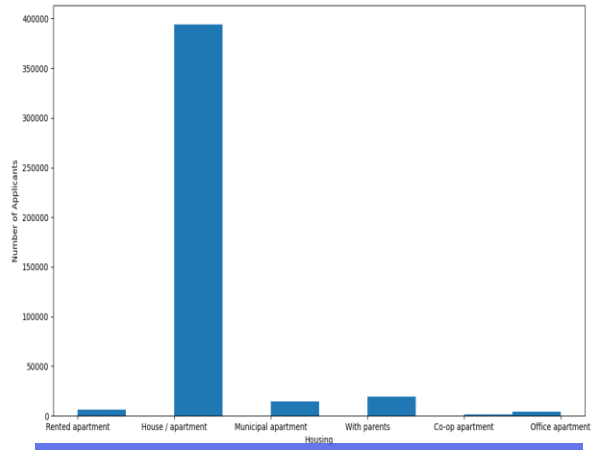
Income Source



Education Type



Family Status

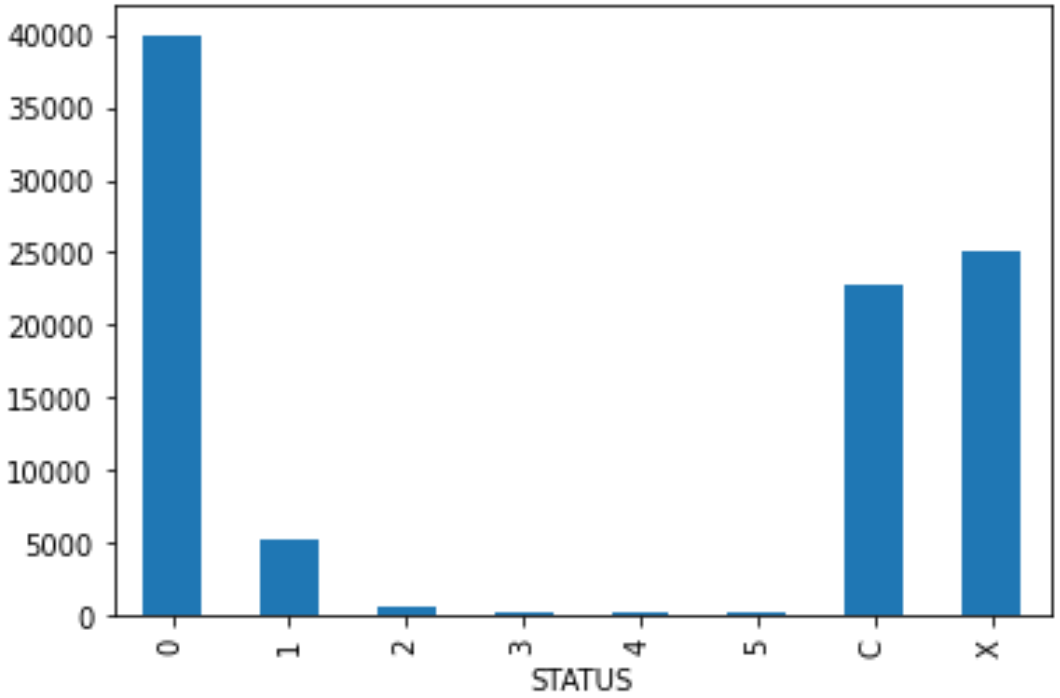


Housing Type

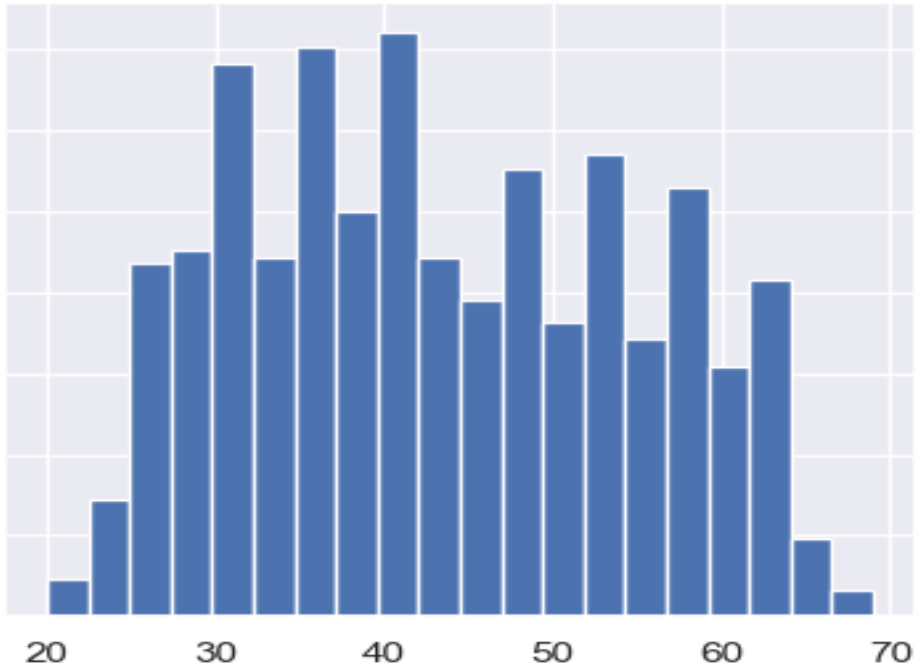
Most applicants are working class, secondary education, married and own a house/apartment.

EXPLORATORY ANALYSIS

UNIVARIATE



Payback Window Distribution



Age Distribution

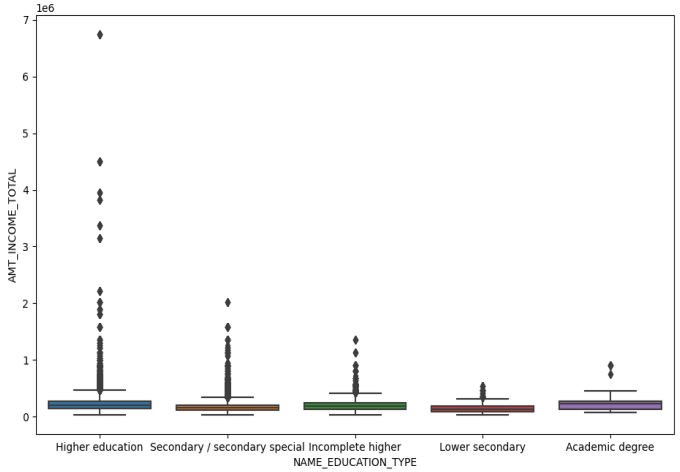
Equal number of clients make no purchases vs. those who pay off in full (affects interchange revenue).

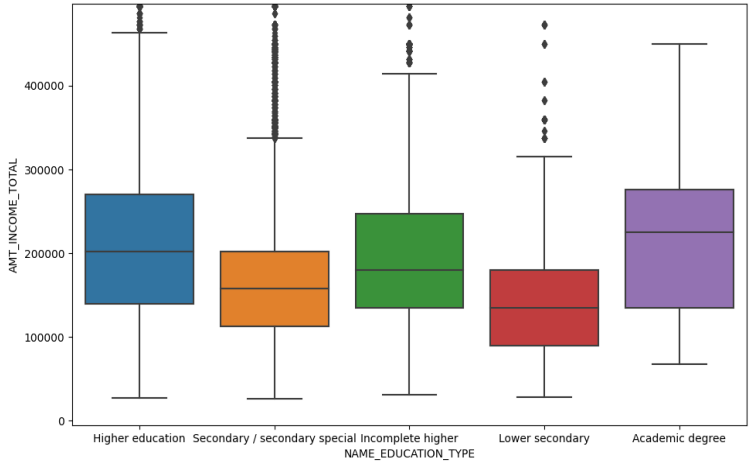
Most overdue loans are paid off within 30 days after the statement cycle.

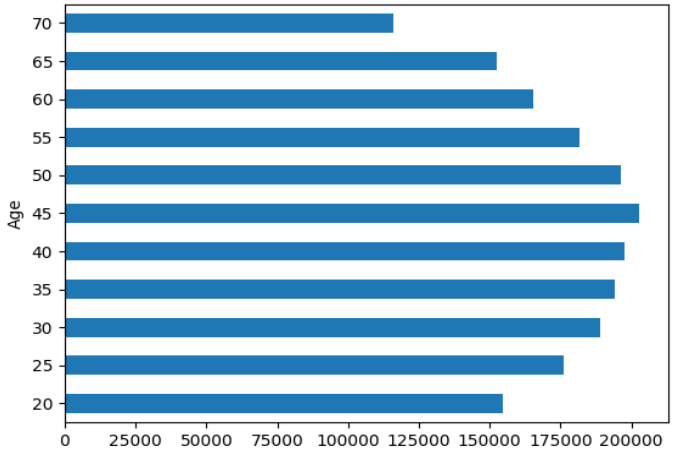
People seldom wait up to 150+ days to repay their loan.

EXPLORATORY ANALYSIS

BIVARIATE







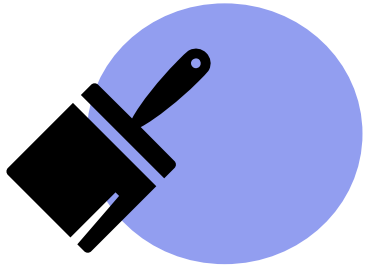
Income Distribution by Education

Income Distribution by Age

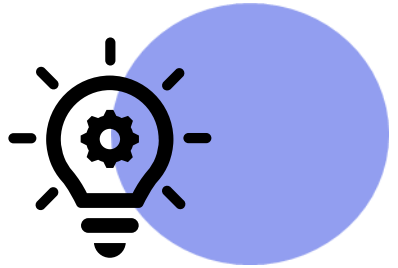
Some outliers earn 14x higher than IQR range [$500,000 - $7,000,000].

Students and 20-year-olds have higher median incomes than is seen in the real world.

METHODOLOGY
APPROACHES & CHALLENGES

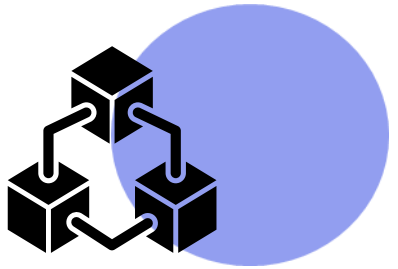


DATA CLEANING



FEATURE ENGINEERING

DATA PARTITIONING



MODEL DESIGN

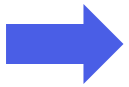
DATA CLEANING

ATTRIBUTE	ACTION	REASONING
CNT_CHILDREN	**DROP FEATURE**	High Correlation To **CNT_FAMILY_MEMBERS** (0.88)
OCCUPATION_TYPE	**GROUPED CATEGORICAL REPLACEMENT**	Replace N/A With Most Common Occupation Per Income Amount (Socio-Economic Grouping)
AMT_INCOME_TOTAL	**NONE**	Outliers Have Little Effect On Ensembles

FEATURE ENGINEERING

BINARY CATEGORICAL ENCODING

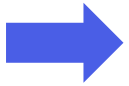
ATTRIBUTE	VALUE RANGE
CODE_GENDER	[M, F]
FLAG_OWN_CAR	[Y, N]
FLAG_OWN_REALITY	[Y, N]



VALUE RANGE
[1, 0]
[1, 0]
[1, 0]

NUMERIC EQUAL WIDTH BINNING

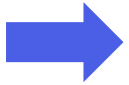
ATTRIBUTE	VALUE RANGE
DAYS_BIRTH	[-7705, -25201]
AMT_INCOME_TOTAL	[26100, 6750000]
DAYS_EMPLOYED	[-17531, 365243]



VALUE RANGE	NOTES
[21,69]	Rounded To Nearest Integer
[25000, 7000000]	Rounded To Nearest 10,000
[0,51]	Currently Unemployed Range = [0, (Age-18)]

MULTICLASS CATEGORICAL ENCODING

ATTRIBUTE	VALUE RANGE
NAME_EDUCATION_TYPE	Secondary, Higher Ed., etc.
OCCUPATION_TYPE	Laborer, Manager, etc.
NAME_FAMILY_STATUS	Single, Married, Widow, etc.
NAME_INCOME_TYPE	Working, Student, Pension, etc.
NAME_HOUSING_TYPE	House, Rented Apartment, etc.



VALUE RANGE	NOTES
[1,5]	Ordinal Encoding To Maintain Hierarchy
[1,0]	One-Hot Dummy Encoding
[1,0]	One-Hot Dummy Encoding
[1,0]	One-Hot Dummy Encoding
[1,0]	One-Hot Dummy Encoding

FEATURE ENGINEERING

$$\textbf{MONTH SPEND} = \frac{\text{AMT INCOME TOTAL} * \text{EXPENDITURE (50\%)}}{\text{12 MONTHS}}$$

$$\text{INTERCHANGE} = \textbf{MONTH SPEND} * \text{INTERCHANGE. RATE (2\%)}$$

$$\text{INTEREST} = \textbf{MONTH SPEND}\left[1 + \frac{\text{APR (19.99\%)}}{\text{12 MONTHS}}\right]^{\text{\# MONTHS OVERDUE}} - \textbf{MONTH SPEND}$$

Clients modelled as spending 50% of their gross income, to account for net earnings after tax and savings.

DATA PARTITIONING
CLASSIFICATION - APPROACH #1

OBJECTIVE: MINIMIZE LOSSES (MULTICLASS)

Scenario 1
No Purchases Made

$0 Revenue

Scenario 2
Balance Fully Paid Off

Interchange

Scenario 3
Late Payment (Overdue)

Interchange + Net Interest

MAXIMIZE

Scenario 4
No Payment (Write off)

Net loss

MINIMIZE

PROBLEMS

High Income Scenario 2 yield more Interchange Revenue than Low Income Scenario 3 Interest Revenue.
Under-Sampling Issues.
Poor Model Performance.

DATA PARTITIONING
CLASSIFICATION - APPROACH #2

OBJECTIVE: MINIMIZE LOSSES (MULTICLASS)

Scenario 1
No Purchases Made

$0 Revenue

Scenario 2 + 3
Balance Fully Paid Off or Overdue

Interchange or Interchange + Net Interest

25%	25%	25%	25%

MAXIMIZE IQR

Scenario 4
No Payment (Write off)

Net loss

MINIMIZE

PROBLEMS

Revenue Generation is a Function of Time. Not all Clients have equal Credit History.
Under-Sampling Issues.
Poor Model Performance.

DATA PARTITIONING
CLASSIFICATION - APPROACH #3

OBJECTIVE: MINIMIZE LOSSES (BINARY)

Scenario 1 + 2 + 3

No Purchases Made, Balance Fully Paid Off or Overdue

$0 Revenue, Interchange or Interchange + Net Interest

Scenario 4

No Payment (Write off)

Net loss

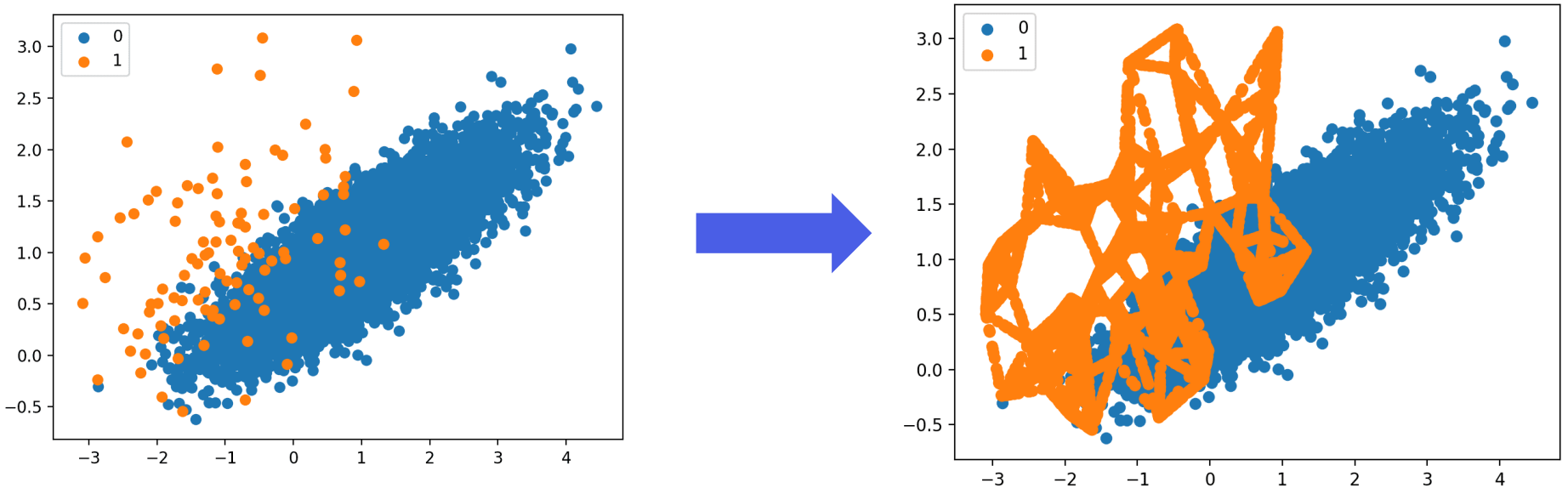
MINIMIZE

PROBLEM

Under-Sampling.

DATA PARTITIONING
SAMPLING - APPROACH #1

TECHNIQUE: SMOTE
METHOD: KNN TO GENERATE NEW DATA

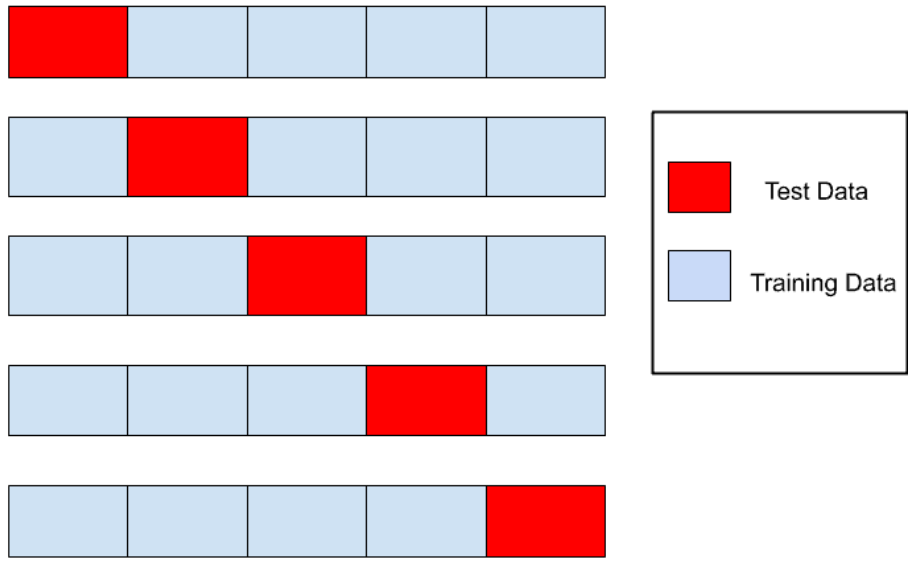


PROBLEMS

Fake Data with Little Variance.
Majority : Minority Ratio too Big, Over-Sampling = Over Correction.
Under-Fitting.

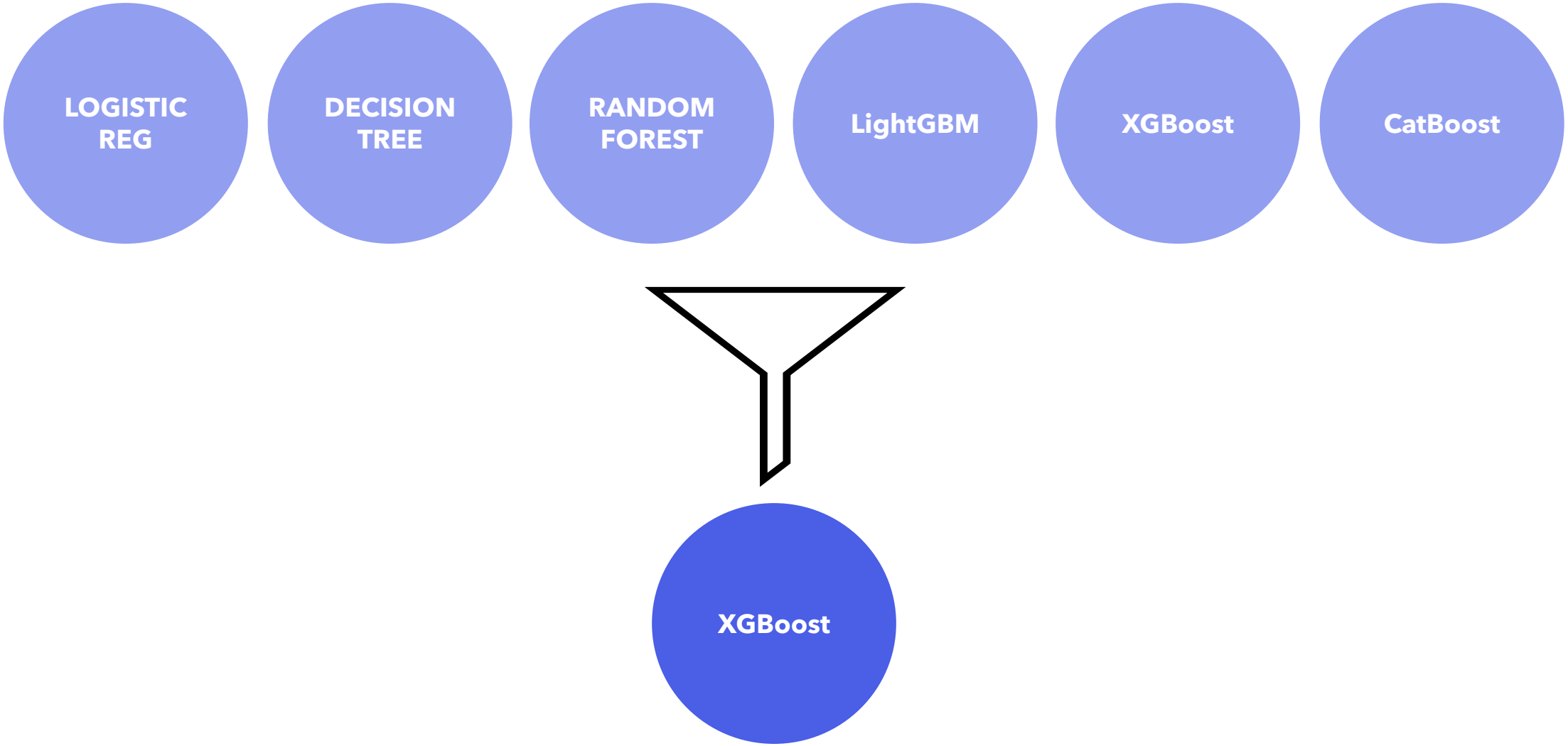
DATA PARTITIONING
SAMPLING - APPROACH #2

TECHNIQUE: K-FOLD CROSS VALIDATION
METHOD: SPLIT DATA INTO K GROUP, TRAIN ON K-1 FOLDS



MODEL DESIGN

APPROACH: TRY NUMEROUS MODELS, FINE TUNE BEST ONE



MODEL DESIGN

FINE TUNING: BAYSIAN SEARCH BACK-PROPOGATION

XGBoost

OPTIMAL PARAMETERS

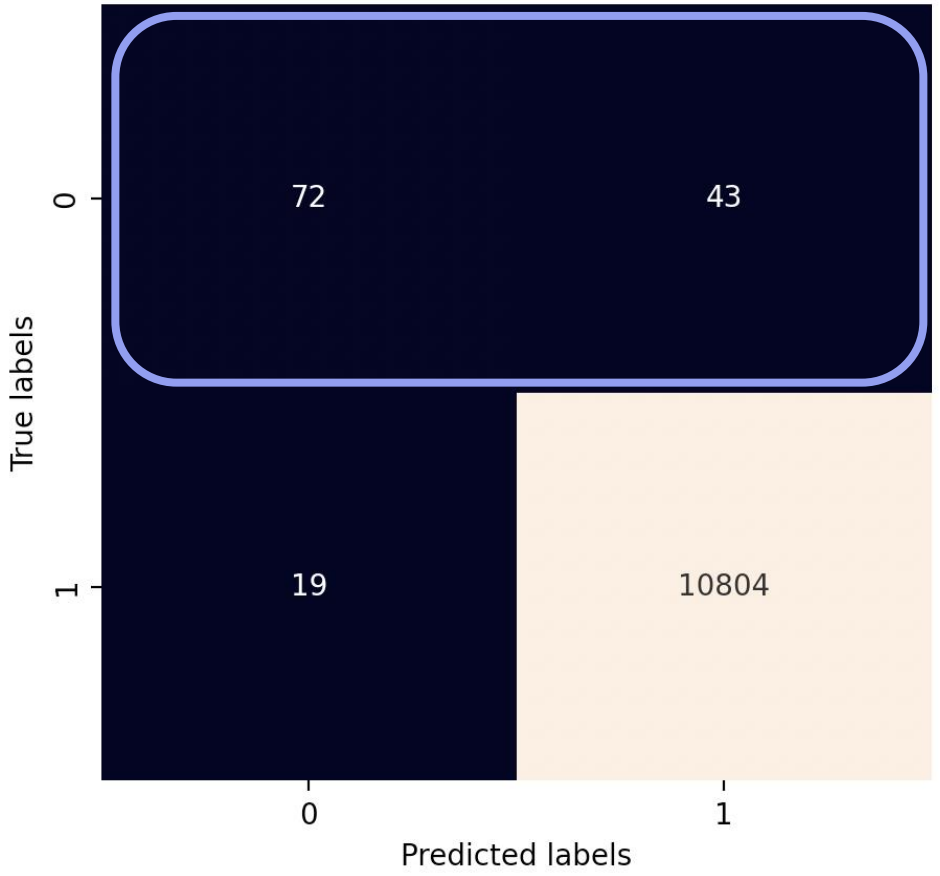
TREE DEPTH
11

 # TREES
285

SAMPLE RATIO
1 : 1

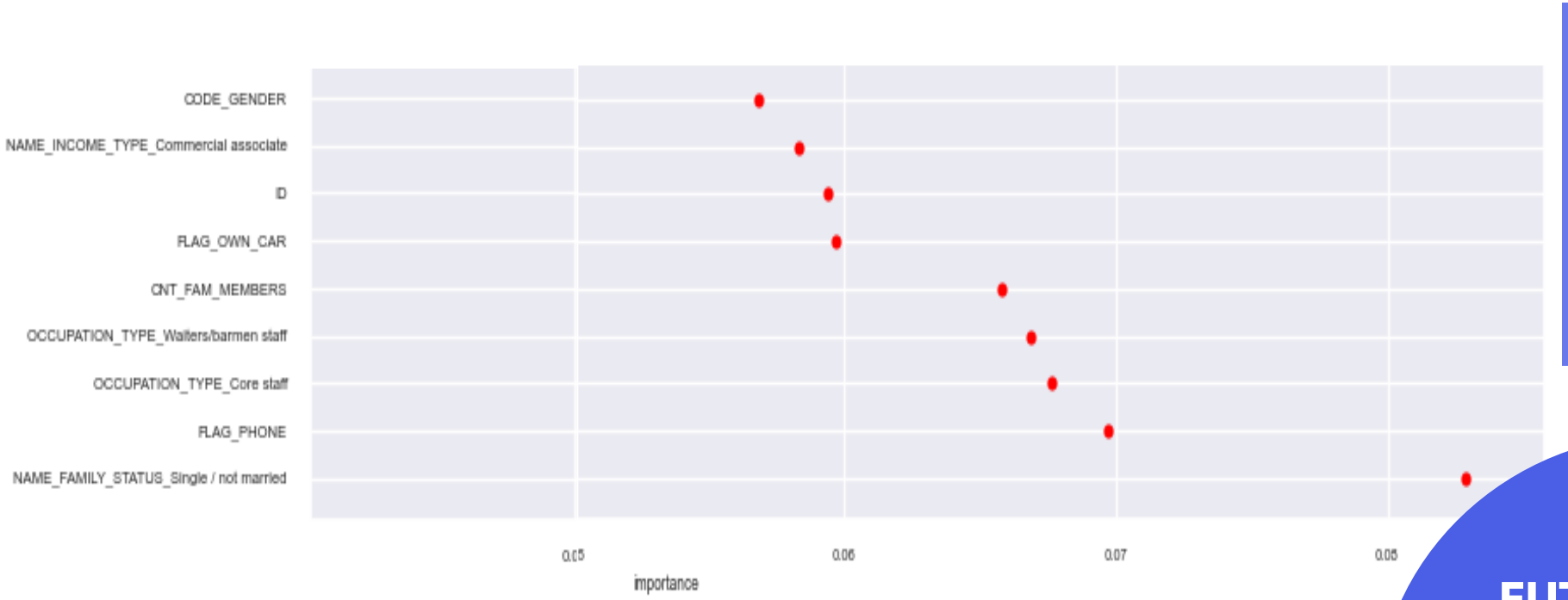
RESULTS

OPTIMIZE WRITE-OFF CLASSIFICATION TO MINIMIZE LOSSES



RECALL	**59.8%**
ACCURACY	99.6%
PRECISION	79.1%

CONCLUSIONS



LOW SOCIO-ECONOMIC FACTORS YIELD HIGHEST IMPORTANCE

SINGLE PROLETARIANS WITH **NO DEPENDANTS/ASSETS** MOST LIKELY TO BECOME **WRITE-OFFS**

FUTURE WORK

Added Applicant Features
Added Quantitative Features
Explore Multiclass Approach
Real Credit Bureau Data
Add Assumptions Back


THANK YOU