Mazars Data Advisory Services
Anti Money Laundering
Fair Lending Compliance

Use Cases

April 2021

mazars

01

BSA AML Model Validation

BSA & AML Model Validation

REGULATORY CONTEXT

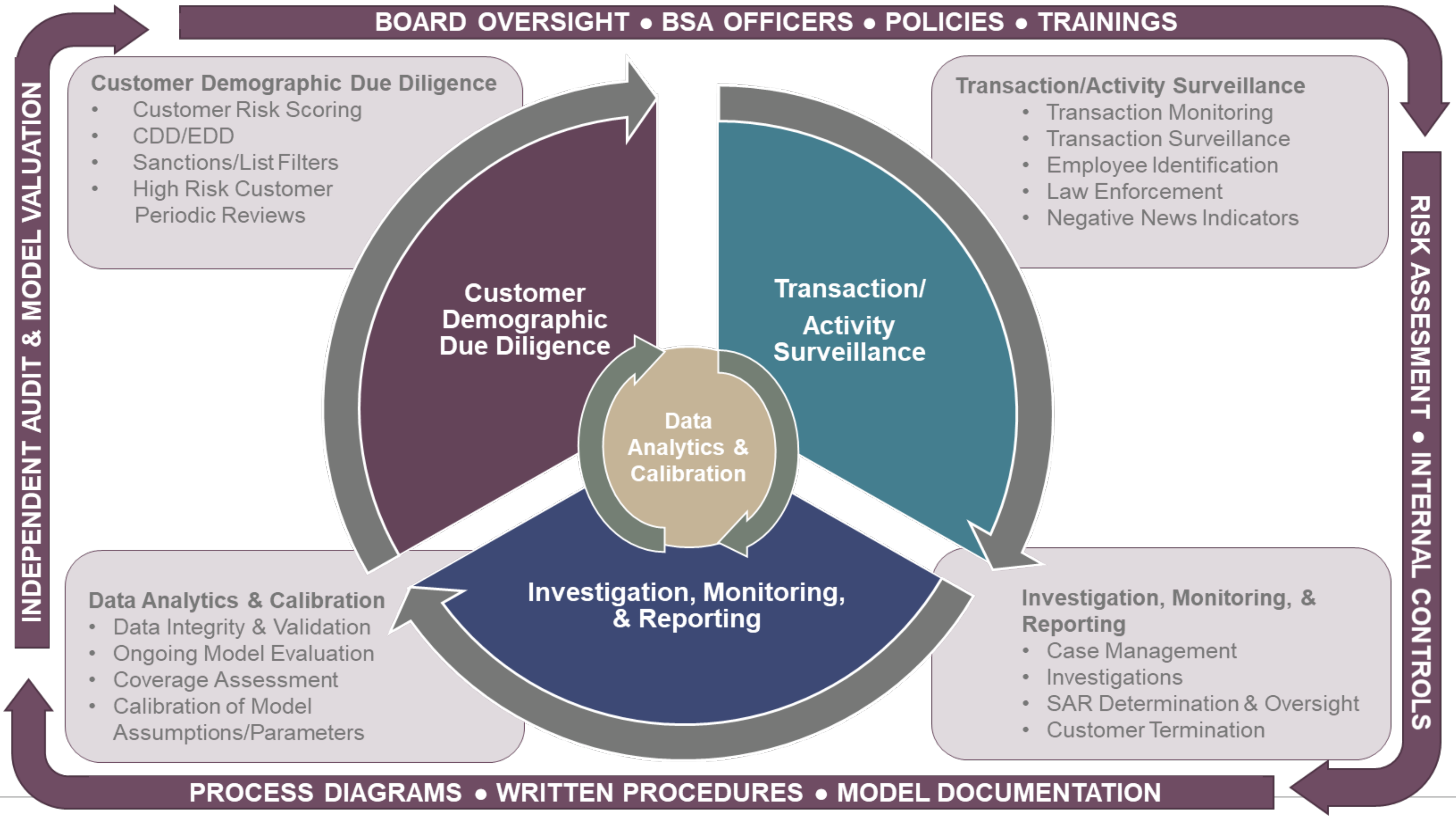
- Current environment mandates periodic independent assessments of financial institutions' BSA-AML monitoring systems
- They require an increased focus on the design, implementation and outputs of transaction monitoring systems in accordance with regulatory expectations, with the objectives of improving decision-making and confidence in the models as well as minimizing exposure to risk while optimizing operational costs
- Main components of a model validation include the validation of the conceptual design, the system, the data and the process

KEY VALIDATION CHALLENGES

- BSA-AML models must perform as expected and are in-line with their design objectives and business uses
- Assumptions must be well documented and supported; outputs are analyzed and presented appropriately
- Data quality and accuracy of data feeds must be validated, including sample data, from source systems to the monitoring database to ensure the integrity of input data and data lineage
- Users must be able to calibrate, optimize and implement scenario thresholds and parameters
- Financial institutions must be able to assure regulators that they have remediated identified alerts and deficiencies, and performed account reviews using lookback methodologies

BSA & AML Model Validation

Illustrative BSA/AML Model Validation Life Cycle



CHOSEN SOLUTION

Our chosen solution consists in running a validation of the bank's model by both **assessing the data inputs relevance and preprocessing method**, as well as **validating the model methodology relevance.** Our approach is tailored to the type of model/system the bank is using and its complexity.

With a **traditional model**, the limited amount of data involved, and complexity level allow us to replicate it and validate its outputs, by:

- Validating data inputs integrity and quality
- Assessing the model conceptual design
- Replicating the model to perform a population-based validation of its performance
- Assessing the model parameters sensitivity

With a **Machine Learning-based model**, the volumes of data and the complexity is too high to reproduce the model itself. To cope with that complexity, our approach is focused on:

- Validating the input data selection
- Assessing the model creation methodology based on the available documentation
- Validating the outputs of the model

BSA & AML Model Validation: ML Illustration

DATA VALIDATION	METHODOLOGY REVIEW	MODEL EXPLAINABILITY
Core Systems	Machine Learning Model	Results validation

Data inputs and processing validation

Data Flow integrity

- Using data from core and model, values from each field are matched to **validate the accuracy of the data flow from the core system**
- **An integrity report is provided** in the workpapers

Training Data selection validation

- Assess training data representativity
- Review logic applied for selecting training data
  - Initial data scope and features (selection bias)
  - Feature creation and selection (Importance level, mutual correlations…)

Methodology and results

Methodology assessment

- Review of **model training and testing procedure**
- Review of **model type selection** (classification, regression, clustering etc.)
- Review of **model selection and hyperparameters tuning methodology** (grid search, random search, Bayesian optimization etc.)

Results validation

- Review **evaluation metrics selection** (accuracy, AUC, f1-score etc.)
- Assess **output robustness testing** (out-of-sample, cross-validation)
- Decision threshold sensitivity testing (ATL/BTL)

Reporting interfaces

Fix black-box effect by going from dataset-level explanation to instance-level explanation

- Build an explanation framework in charge of identifying the variables leading to its result
- Use Local Interpretable Model-Agnostic Explanations (LIME) to detect which variables impacted the prediction

Build a user-friendly interface to visualize the results of the prediction explanation

- Build data visualization to allow for an interpretable representation of the explanation
- The model explainer is deployed as a REST API and integrated in the any related process

Technological stack

Ubuntu / Debian environment

docker GitLab RUNNER Continuous Integration

learn Python (LIME, Scikit-learn)

CSV YML Data Storage

02

Fair Lending Compliance

ECO & FHA Compliance: Defining fairness in AI and ML

- Whereas bias — the systematic favoring of one group over another — can be measured mathematically, fairness is a flexible and subjective concept that must be evaluated in light of the circumstances and goals of the machine learning project

- A fairness definition: “Ensure that algorithmic decisions do not create discriminatory or unjust impacts when comparing across different demographics (e.g., race, sex, etc.…)”

- No one-to-one correspondence between bias and fairness:
  - For example, if an algorithm is more likely to disqualify women applicants from receiving loans to start small businesses, regardless of the applicants’ traits of creditworthiness, that algorithm could be said to be unfair in its treatment of women (or biased against them).
  - However, it is also possible that, in the pursuit of fairness, an algorithm could deliberately introduce a bias as a means of redressing preexisting inequities

Source: Exploring Fairness in Machine Learning for International Development. Spring 2020. Massachusetts Institute of Technology: MIT

ECO & FHA Compliance: Defining fairness in AI and ML

Criterion	Description	Advantages	Disadvantages
Fairness through unawareness	Remove protected attributes from the data set (e.g., race, gender)	• Simple to implement	• Not effective unless some unusual criteria are satisfied (no correlated attributes)
Demographic parity	Require parity of some statistic of the outcome across groups (e.g., rejection rate)	• Conceptually simple • Can have legal standing (disparate treatment)	• Does not address individual-level fairness • May unacceptably compromise prediction accuracy
Equalized opportunity	Force the true positive rates to be the same between the protected groups	• Appeals to a reasonable interpretation of fairness • A good option if the true positive rate is most consequential factor	• Disparate false negative rates may remain between two populations • Requires lots of labeled historical data
Equalized odds	Force both the true positive rates and the false negative rates to be the same between the protected groups	• Appeals to a reasonable interpretation of fairness	• Can be inconsistent with high levels of accuracy

Source: Exploring Fairness in Machine Learning for International Development. Spring 2020. Massachusetts Institute of Technology: MIT

ECO & FHA Compliance

REGULATORY CONTEXT

- The Equal Credit Opportunity Act and the Fair Housing Act are designed to protect consumers from unfair or discriminatory lending practices
- Mazars' Fair Lending compliance professionals help financial institutions pinpoint potential discriminatory practices
- Recent major revamps of HMDA reporting requirements (many more required fields) have created a disruption that generates more complexity in the analysis

CURRENT APPROACH

- Third-party vendors have developed software tools to carry out analyses on loan data, to identify outliers for the Fair Lending auditing work
- BI system with different tabs to address the 5 critical areas of risk (Marketing, Underwriting, Pricing, Steering, Redlining)
- Slice and Dice tabular reports to visualize the data (filters, selectors, etc.)

LIMITATIONS TO BE OVERCOME

- Relatively 'arbitrary' control population (white males) to identify outliers
- Univariate distribution analyses to identify populations above control population threshold level
- Seemingly endless possibilities of slicing and dicing to identify outliers (because they remain "linear")
- Hard to scale to larger datasets (large populations of outliers potentially identified through univariate projections)

mazars

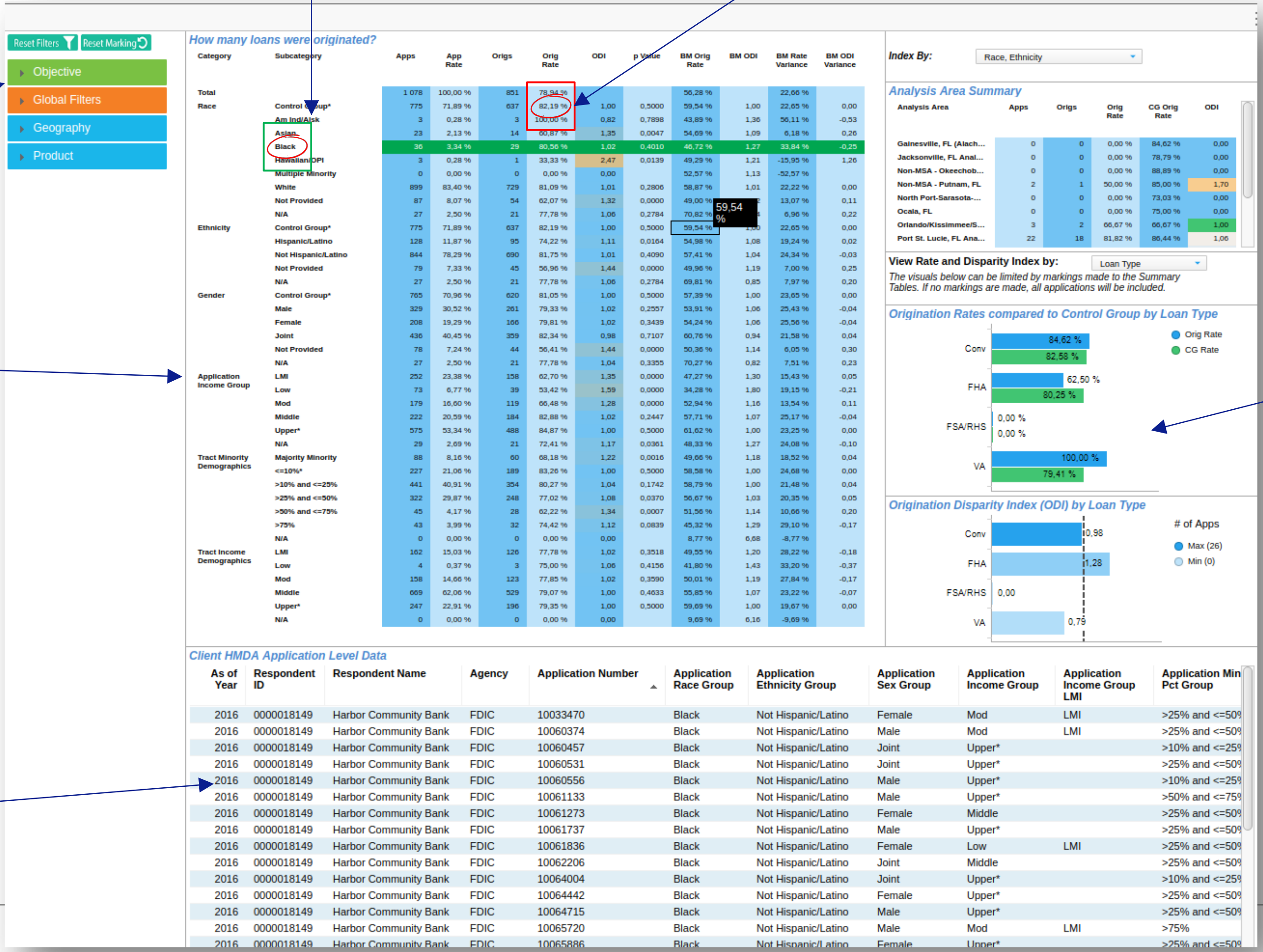
ECO & FHA Compliance

Current approach: A linear, manual, fastidious and incomplete process

Selected subpopulation
Chosen along one of the available axes.
Race = Black
origination rate = 80.56%

Reference Indicator(s)
control group = "white males"
reference origination rate = 82.19%

Filters
allow to focus on a subset of data
- by location
- by date range



Summary statistics
tabular view of given KPIs broken down each time by one category (fixed list): Race, ethnicity, gender, etc.

Other views
They allow to further visualize the data through different angles (dropdown)

Selection records
list of records from selection (highlighted in green above), with lower origination rate compared to control group.
Race=Black: 36 records
 to investigate

ECO & FHA Compliance

New approach: Non-linear, exhaustive, automated (and replicable), fast and more flexible

AUTOMATED

- Ability to process thousands & millions of lines: scalability
- Process can be largely automated (1st order analyses as well as algorithms pipelines)
- No need for cumbersome setup within client’s infrastructure. All of the data processing can be done in the cloud (powerful servers with our toolkit installed) and results are returned to client through web interface

NON LINEAR / EXHAUSTIVE

- Unbiased identification of explanatory variables (no preliminary hypotheses)
- Initial algorithms runs zoom in on specific populations of interest, eliminating the need to analyze the whole client base against axes of interest
- Subsequent algorithms runs reveal the most discriminatory sensitive attributes, eliminating the need to test each and every sensitive attribute
- Algorithms can identify subpopulations defined by a combination of attributes, and certain populations can emerge that wouldn’t have been found through sequential univariate projections

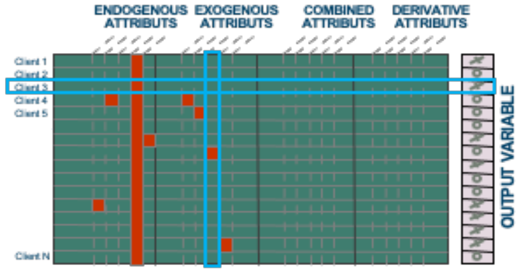
ENHANCED FLEXIBILITY

- Algorithms feature powerful editing capabilities. Adding, removing, modifying variables are split-second operations
- The automation of a large part of the process allows for quicker rerun cycles of the whole data pipeline (from raw data to final population identification). This allows for short implementation times of solutions on slightly different use cases (different type of loan, different target, etc.)
- Due to the technology’s precise subpopulation identification, the auditing effort and time required to study potential outliers can be significantly reduced

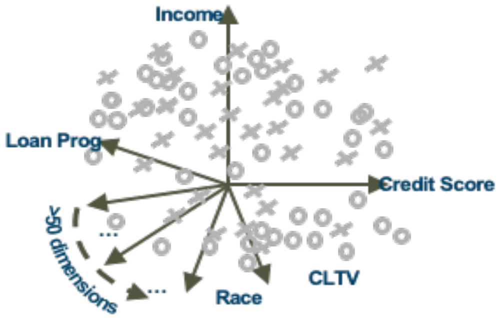
ECO & FHA Compliance

Generating applicant profiles: rules

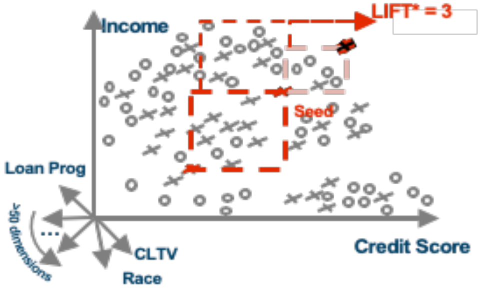
1 Client data is plotted



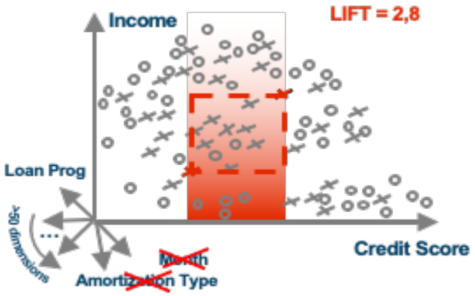
2 Each column represents a dimension in the hyperspace



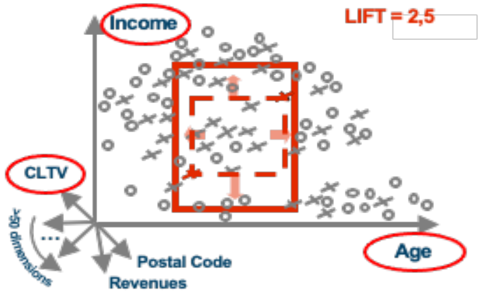
2 Our quantitative approach identifies relevant hypercubes



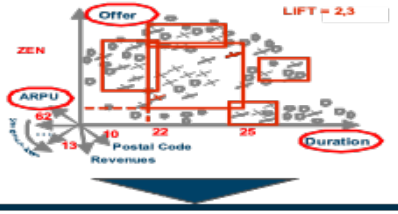
4 Non-explanatory variables are eliminated one by one



5 An optimum is computed that is a trade-off between the hypercubes' lift* and their size



6 Resulting rules capture the largest population with the highest possible lift



Applicants with the following characteristics:

V1 : Debt Ratio	**not in [17, 20]**
V2 : Income	**< 54k**
V3 : CLTV	**not in [65, 72]**

are **3.3 times** more likely to get their application denied

Size = 145 / Lift: 3.3

* Lift: Ratio capturing the "density" of the outcome in a given hypercube vis-à-vis the density in the entire population sample

ECO & FHA Compliance

Advanced analysis: Surfacing potential compliance risk

The refined rules can be explored further under the lens of fair lending compliance. This is done by statistically 'digging' the sensitive attributes space, and uncovering the combinations that best distinguish denied applications from approved ones.

Rules A

P: 44 N: 123 Coverage: 167 Lift: 3.041 Score: 0.365

attribute	p*	n*	lift*	score*	context	domain
DEBT_RATIO_discretized	-2	-19	0.217	0.014	9/10	not{"(17.57, 20.27]"}
COMB_RATIO_discretized	-2	-24	0.290	0.021	9/10	not{"(26.27, 30.3]"}
CLTV_discretized	0	-17	0.281	0.024	8/9	not{"(65.72, 72.09]"}
TINCOME_discretized	-2	-73	0.847	0.074	7/10	not{"(135.0, 192.0]", "(192.0, inf)", "(63.2, 78.0]"}
PNTSFEES_discretized	-15	-258	1.493	0.136	4/7	not{"(1253.84, 1765.62]", "(249.37, 838.85]", "(838.85, 989.0]"}

Rules B

P: 42 N: 105 Coverage: 147 Lift: 3.297 Score: 0.380

attribute	p*	n*	lift*	score*	context	domain
APL_AGE_discretized	-2	-18	0.257	0.015	9/10	not{"(68.0, inf)"}
DEBT_RATIO_discretized	-1	-16	0.271	0.019	9/10	not{"(17.57, 20.27]"}
COMB_RATIO_discretized	-2	-22	0.328	0.021	9/10	not{"(26.27, 30.3]"}
CLTV_discretized	0	-17	0.342	0.027	8/9	not{"(65.72, 72.09]"}
TINCOME_discretized	-2	-67	0.946	0.077	7/10	not{"(135.0, 192.0]", "(192.0, inf)", "(63.2, 78.0]"}
PNTSFEES_discretized	-15	-234	1.636	0.138	4/7	not{"(1253.84, 1765.62]", "(249.37, 838.85]", "(838.85, 989.0]"}

42 denied vs 18 approved
differ only by Applicant Age (within Rule B space)

APL_AGE: not{[68, ∞[} = applicants younger than 68
When looking at **applicants younger or older than 68 'all other things equal'** (Debt Ration, Combined Ration, Income, CLTV, PNTSFEES are in the same ranges), we find a **significant difference in loan origination rates** (71% vs 90%)
→ requires auditor verification to see if the reason lies in other attributes, or if this is a case of unfair lending

P: 42 N: 103 Coverage: 145 Lift: 3.343 Score: 0.384

attribute	p*	n*	lift*	score*	context	domain
DEBT_RATIO_discretized	-1	-16	0.280	0.019	9/10	not{"(17.57, 20.27]"}
CLTV_discretized	0	-16	0.332	0.026	8/9	not{"(65.72, 72.09]"}
LTV_discretized	0	-16	0.332	0.026	8/9	not{"(96.5, inf)"}
TINCOME_discretized	-4	-89	1.112	0.089	6/10	not{"(135.0, 192.0]", "(192.0, inf)", "(54.0, 63.2]", "(63.2, 78.0]"}
PNTSFEES_discretized	-15	-233	1.669	0.139	4/7	not{"(1253.84, 1765.62]", "(249.37, 838.85]", "(838.85, 989.0]"}

P: 38 N: 68 Coverage: 106 Lift: 4.137 Score: 0.422

attribute	p*	n*	lift*	score*	context	domain
MARITALC	-1	-5	0.119	0.003	2/3	not{3.0}
CAPRACE	0	-2	0.077	0.005	4/5	not{6.0}
CO_APL_AGE_discretized	0	-4	0.150	0.009	10/11	not{"(43.2, 48.0]"}
DEBT_RATIO_discretized	0	-11	0.389	0.024	9/10	not{"(17.57, 20.27]"}
LTV_discretized	0	-11	0.389	0.024	8/9	not{"(96.5, inf)"}
CLTV_discretized	0	-13	0.452	0.028	8/9	not{"(65.72, 72.09]"}
APL_AGE_discretized	-2	-22	0.586	0.030	8/10	not{"(63.0, 68.0]", "(68.0, inf)"}
TINCOME_discretized	-4	-63	1.335	0.082	6/10	not{"(135.0, 192.0]", "(192.0, inf)", "(54.0, 63.2]", "(63.2, 78.0]"}
PNTSFEES_discretized	-11	-167	2.146	0.148	4/7	not{"(1253.84, 1765.62]", "(249.37, 838.85]", "(838.85, 989.0]"}

38 denied vs 35 approved
differ only by a combination of Marital status, Age, Race (within Rule B space)

MARITALC: not{3}
CARACE: not{6} = co-applicant was provided or no co-applicant
CO_APL_AGE: not{]43, 48]} = co-applicant younger than 43 or older than 48
APL_AGE: not{[63, ∞[} = applicant younger than 63
→ similarly, there are two populations with different loan origination rates that differ by sensitive attributes. **The rules pinpoints a subpopulation to study more closely.**