AltCred AI: Smarter Credit Access

AI-Powered Alternative Credit Scoring for Financial Inclusion
Barclays Hackathon 2025

Developed and validated on two real-world credit datasets:
• Home Credit Default Risk (Primary Production Model)
• Give Me Some Credit (Robustness Benchmark Model)

Algorithms Benchmarked:
Logistic Regression | Random Forest | XGBoost

Core Focus:
Accuracy • Fairness (DI Audit & Mitigation) • SHAP Explainability • Deployment Readiness

Team Covenant Credit Lab
Tanvi Dahale
Rushali Chouhan
Sneha Dahe
Bhumika Desai
Aanandi Bhasme

Problem Statement

Traditional banking mainly depends on detailed credit history (“thick-file” bureau data), which leaves out many deserving people like young professionals, gig workers, and those without past loans. Modern financial institutions aim to increase financial inclusion by using **alternative data (such as income patterns or payment behavior)** while still following strict Basel-III risk rules and maintaining strong credit risk control.

Problem

- **Regulatory Obstacles:** Laws like GDPR (Art. 22) and RBI Fair Practices mandate "Right to Explanation," which complex "Black Box" models cannot provide.
- **Systemic Bias:** Legacy logic often penalizes younger applicants for their lack of history, creating a cycle of exclusion.
- **Data Skew:** Extreme class imbalances (repayment vs. default) often lead to models that overlook high-risk edge cases.

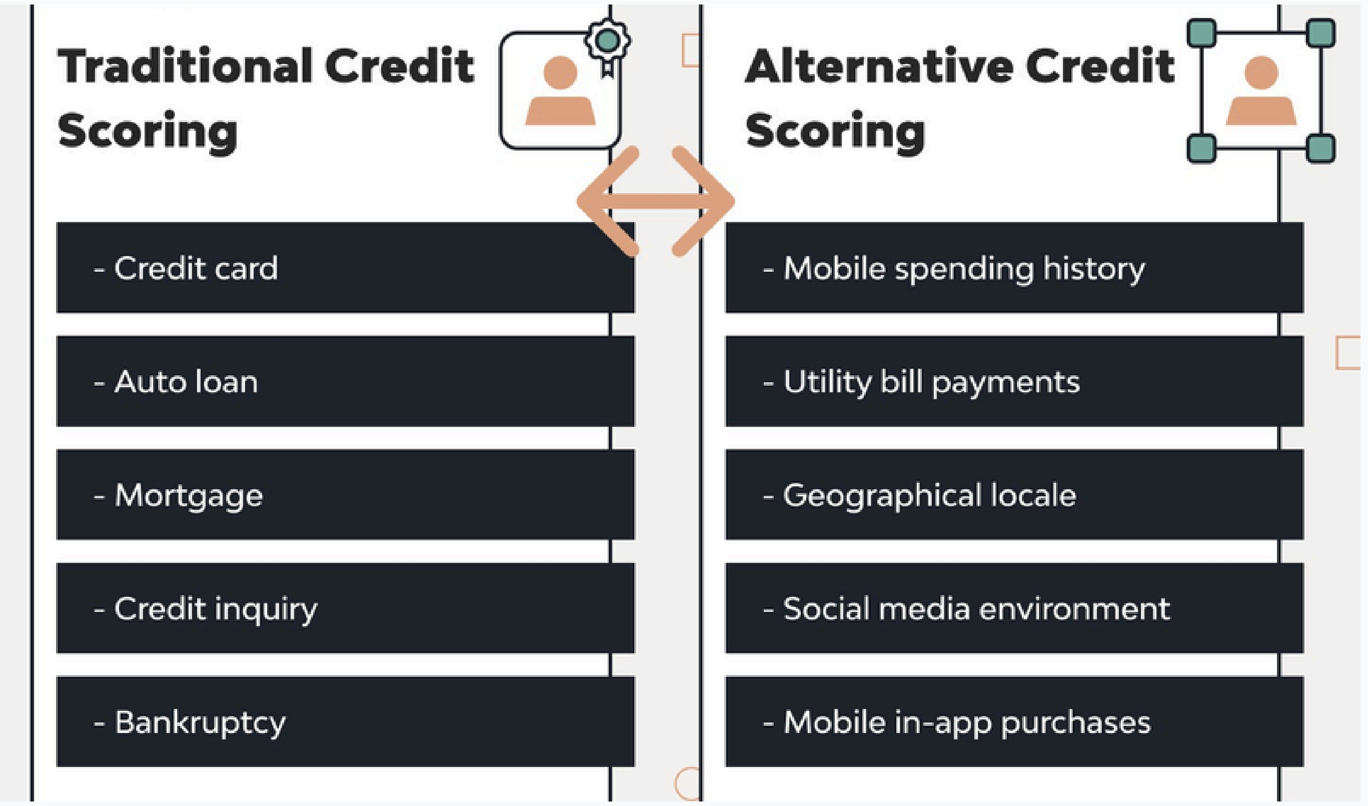
Edge Cases

- **Financial Inclusion:** Uses alternative data like income patterns and job stability to approve capable applicants who lack traditional credit history, while still managing risk.
- **Regulatory Compliance:** Provides clear SHAP-based explanations for every decision, ensuring transparency, auditability, and regulatory compliance.

Problem With Traditional Credit Scoring

Where Current Systems Break Down

Applicant → Bureau Check → No History → Rejected



Limitations of Bureau-Based Models

- Minimum 6 months repayment history required
- First-time borrowers excluded by design
- Built primarily for salaried, urban customers
- Poor generalisation to informal & gig workers

Systemic Gaps & Lack of Transparency

- Rejection with no actionable explanation
- Backward-looking — ignores current income stability
- Static scores, not dynamic risk signals
- No systematic fairness audit across demographics

1.7 Billion adults globally remain unbanked due to lack of formal credit history.
(World Bank Global Findex)

Our Proposed AI Solution

AI-Driven Credit Scoring Pipeline

- End-to-end ML system using **application-level data only**
- **Approach: 2 different models built for 2 datasets to understand development of optimum model w.r.t. different data.**
- Benchmarked LR, Random Forest & XGBoost
- Governance Layer Added: Fairness audit +Post-processing bias mitigation
- Explainability Built-In:SHAP-based explanations
- Selected model based on AUC performance

Alternative signals engineered from:

- Financial stability
- Payment and transaction behavior
- Loan amount & repayment capacity
- Household & financial profile
- External risk indicators (when available)

Category	Technologies & Techniques
Programming Language	Python 3.x
Development Environment	Jupyter Notebook, Anaconda Environment
Data Processing & Manipulation	Pandas, NumPy
Data Visualization & EDA	Matplotlib, Seaborn
Machine Learning Models	Logistic Regression, Random Forest Classifier, XGBoost Classifier (Final Selected Model)
Hyperparameter Optimization	Optuna (Bayesian Optimization), Grid Search (Threshold Tuning)
Imbalance Handling Techniques	SMOTE (Synthetic Minority Oversampling Technique), Class Weighting (scale_pos_weight in XGBoost)
Model Evaluation Metrics	AUC-ROC, Precision, Recall, F1-Score, Accuracy, Youden's J Statistic, Disparate Impact (DI Ratio – Four-Fifths Rule)
Explainability & Interpretability	SHAP (SHapley Additive Explanations)
Fairness & Bias Mitigation	Disparate Impact Audit, Post-processing Threshold Adjustment
Feature Engineering	Behavioral Feature Engineering, Economic Impact Metric (Custom Profit/Loss Calculation), Debt Ratio Cleaning & Transformation, Risk Tier Segmentation

Creditworthiness should reflect repayment capacity — not just credit history.

SYSTEM ARCHITECTURE

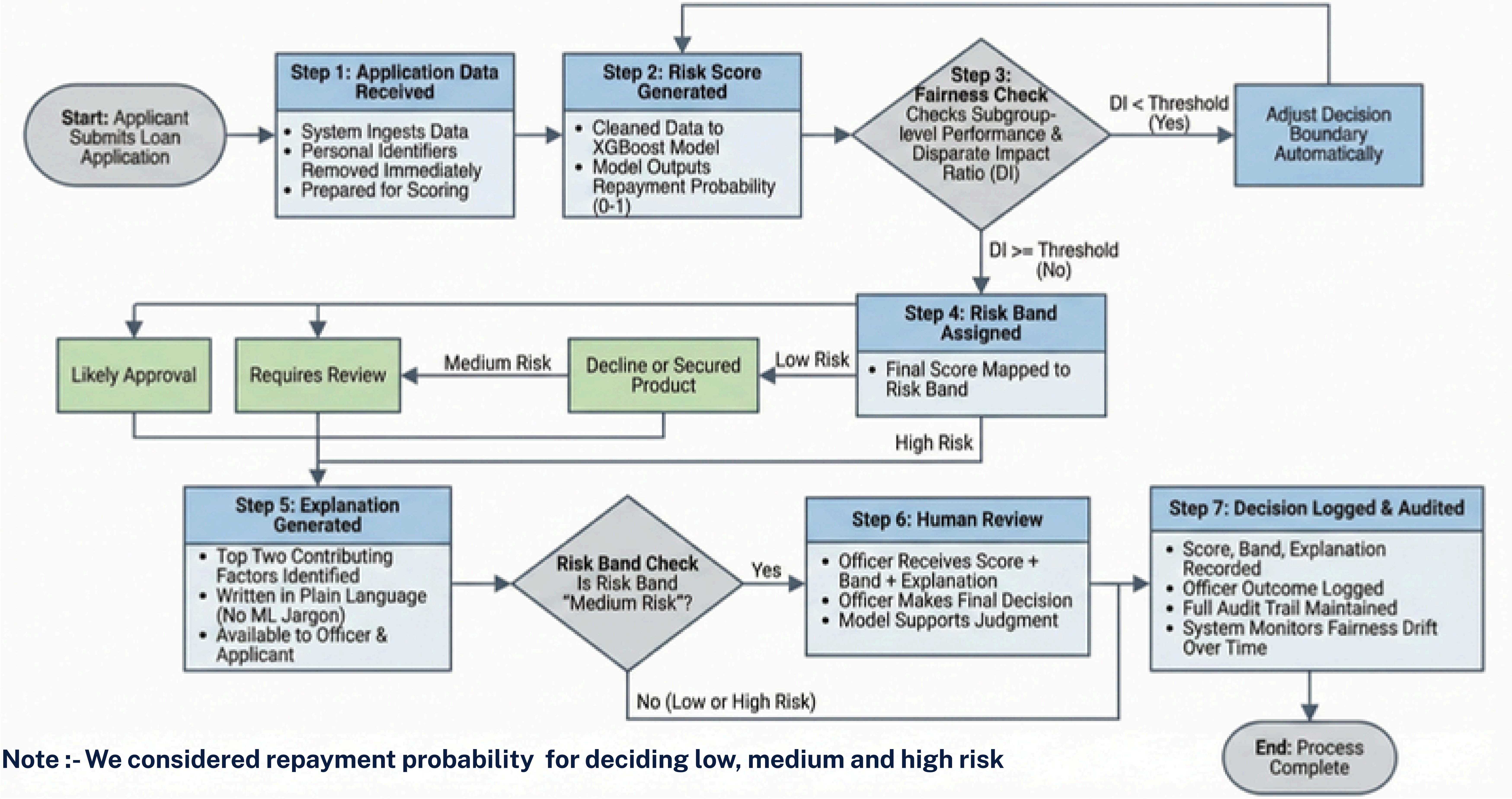
AI-POWERED ALTERNATE CREDIT SCORING · END-TO-END RESPONSIBLE AI PIPELINE

Designed for applicants traditional scoring was never built to evaluate.

Home Credit

WORKFLOW

Home Credit :-
https://www.kaggle.com/datasets/julianocosta/home-credit

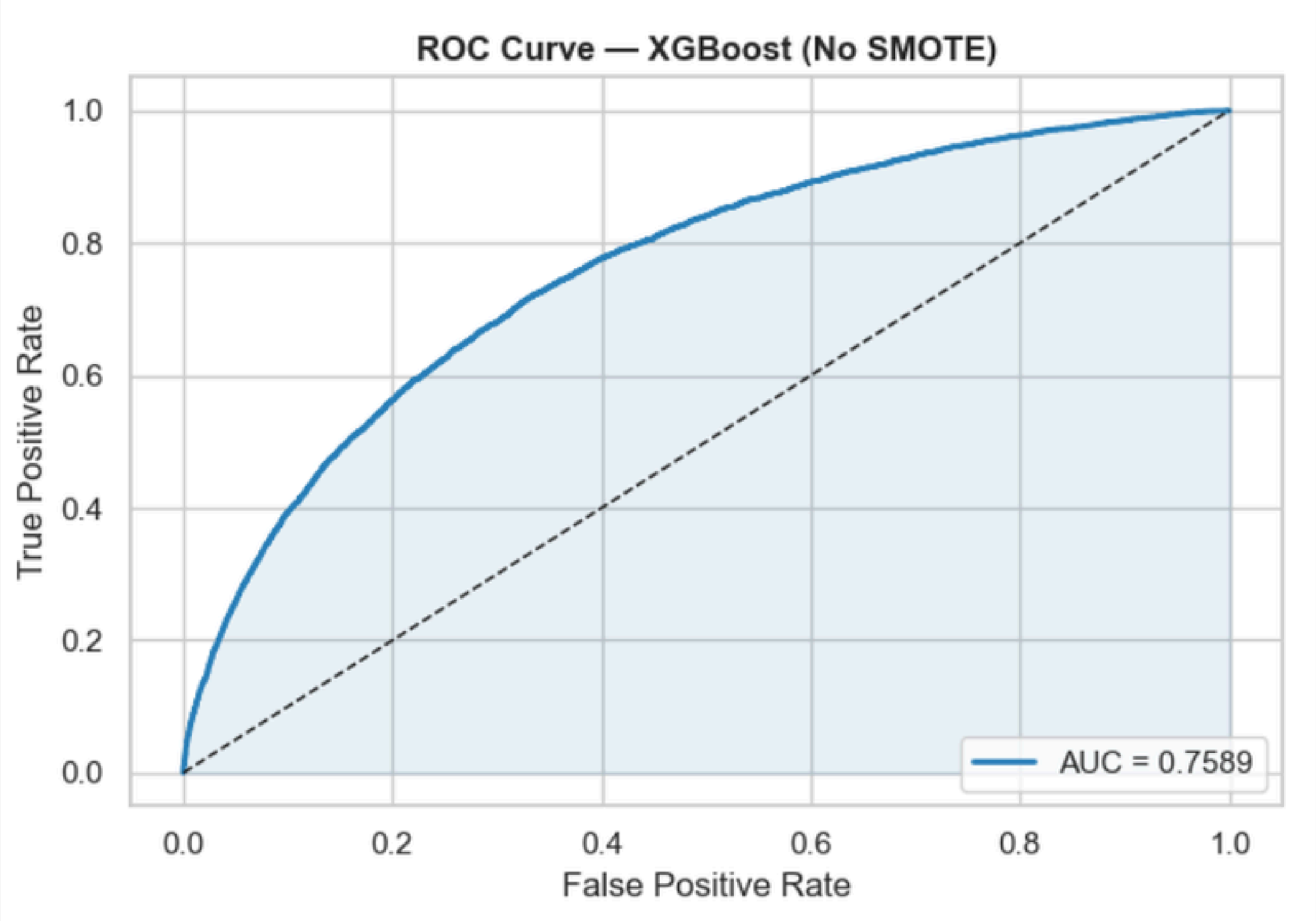


Note :- We considered repayment probability for deciding low, medium and high risk

Data Processing & Model Benchmarking

Home Credit

XGBoost was selected based on highest AUC and stronger minority class(High risk customers) ranking performance.
Accuracy was not used as primary metric due to class imbalance.



Model Selection Rationale

- **Compared 3 models**: Logistic Regression, RandomForest, XGBoost
- Tested both SMOTE and non-SMOTE(Cost-sensitive learning) strategies
- **Primary metric**: AUC-ROC (robust under class imbalance)
- **Key Observations**:

1. SMOTE improved Logistic Regression recall
2. SMOTE degraded XGBoost performance
3. XGBoost (No SMOTE) achieved highest AUC (0.7589)

- **Final Choice**:

XGBoost without SMOTE + scale_pos_weight of 11.4

📊 ALL MODEL PERFORMANCE SUMMARY:

	AUC-ROC	F1-Score	Precision	Recall	Accuracy
LR_Baseline	0.7449	0.2555	0.1575	0.6757	0.6820
LR_SMOTE	0.7443	0.2581	0.1601	0.6647	0.6915
RF_Baseline	0.7448	0.2690	0.1704	0.6375	0.7203
RF_SMOTE	0.6939	0.1062	0.2080	0.0713	0.9031
XGB_Baseline	0.7589	0.2755	0.1748	0.6489	0.7244
XGB_SMOTE	0.7589	0.0452	0.5392	0.0236	0.9195

Empirical validation guided model selection — not assumption.

Performance & Fairness Evaluation **Home Credit**

Model Performance in Context

- **AUC = 0.7589** → The model ranks a defaulter higher than a non-defaulter ~76% of the time.
- **Recall = 0.6489** → ~65% of actual defaulters correctly identified.
- **Precision = 0.1748** → With an 8% base rate, precision is naturally low; flagged applicants default at more than **2× the population rate**.
- **Accuracy (0.72)** is not reliable in imbalanced datasets — approving everyone would still yield ~92% accuracy.

Fairness Audit & Bias Risk

- Evaluated using **Disparate Impact (DI) Ratio** under the **Four-Fifths Rule (≥ 0.80 required).**
- Age split: **≤30 vs >30.**
- **Baseline DI = 0.69 → Fails fairness threshold.**
- Root cause: shorter employment history and lower bureau-linked scores among younger applicants.
- Conclusion:Bias is prsent due to from correlated features

This highlighted the need for a post-processing fairness mitigation layer before deployment.

"A post-processing threshold adjustment was implemented to restore DI ⩾ 0.80 while preserving model AUC."

Fairness Mitigation and Final Results

How We Fixed It — Without Breaking the Model

- We implemented **post-processing threshold adjustment** as our mitigation strategy.
- The model itself was not retrained and no data was modified.
- Only the **decision threshold for younger applicants** was adjusted.
- A grid search was conducted across thresholds (0.05–0.75).

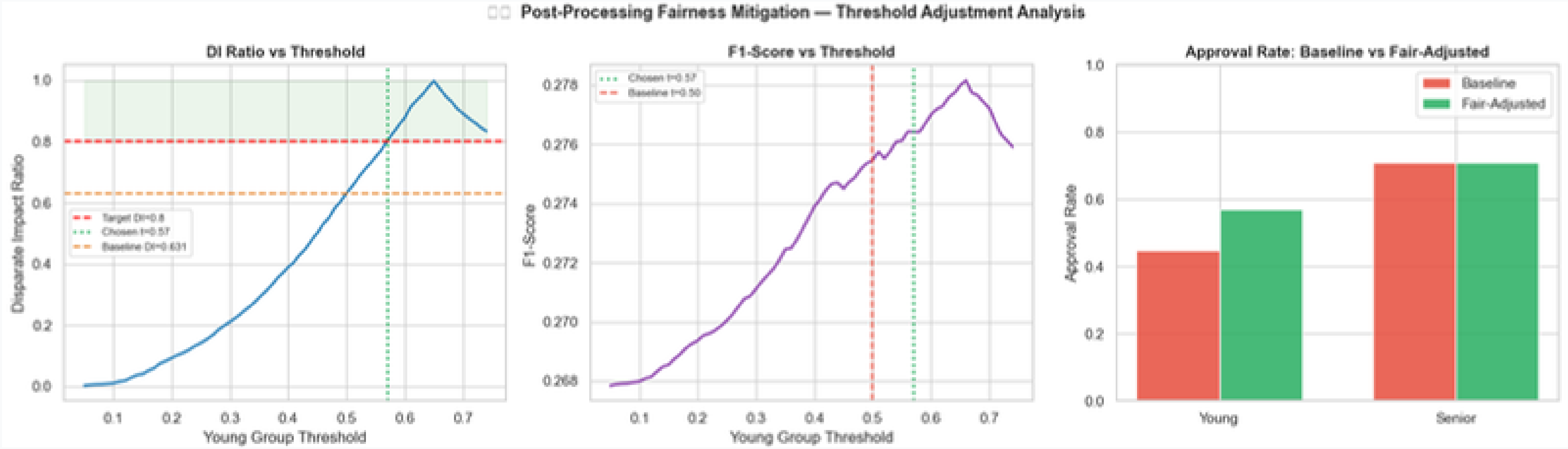
The senior group threshold was fixed at 0.50.

- We selected the **minimum threshold shift** required to raise the **Disparate Impact Ratio to ≥ 0.80**, satisfying the Four-Fifths Rule.

After Mitigation

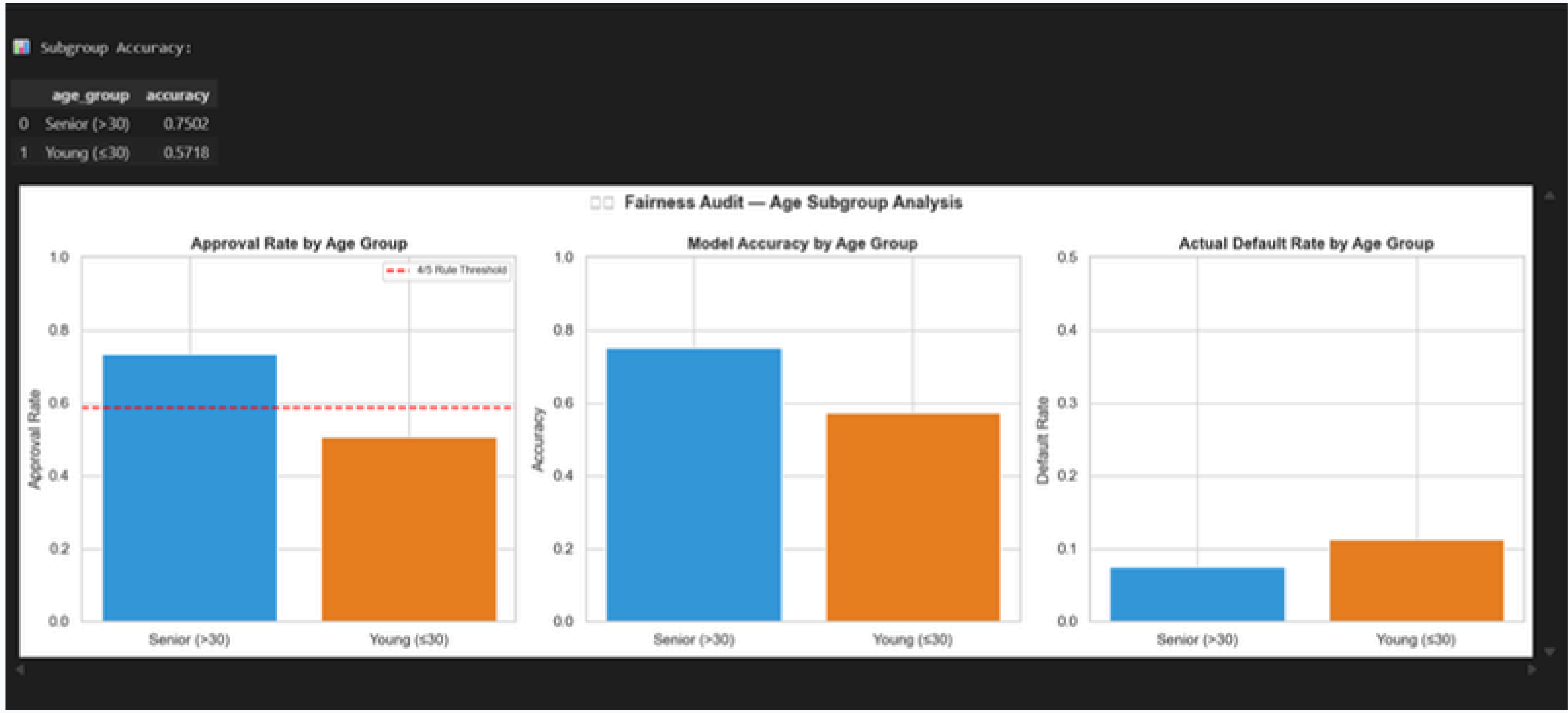
DI Ratio improved from **0.69 → ≥ 0.80 (Fairness Restored)**

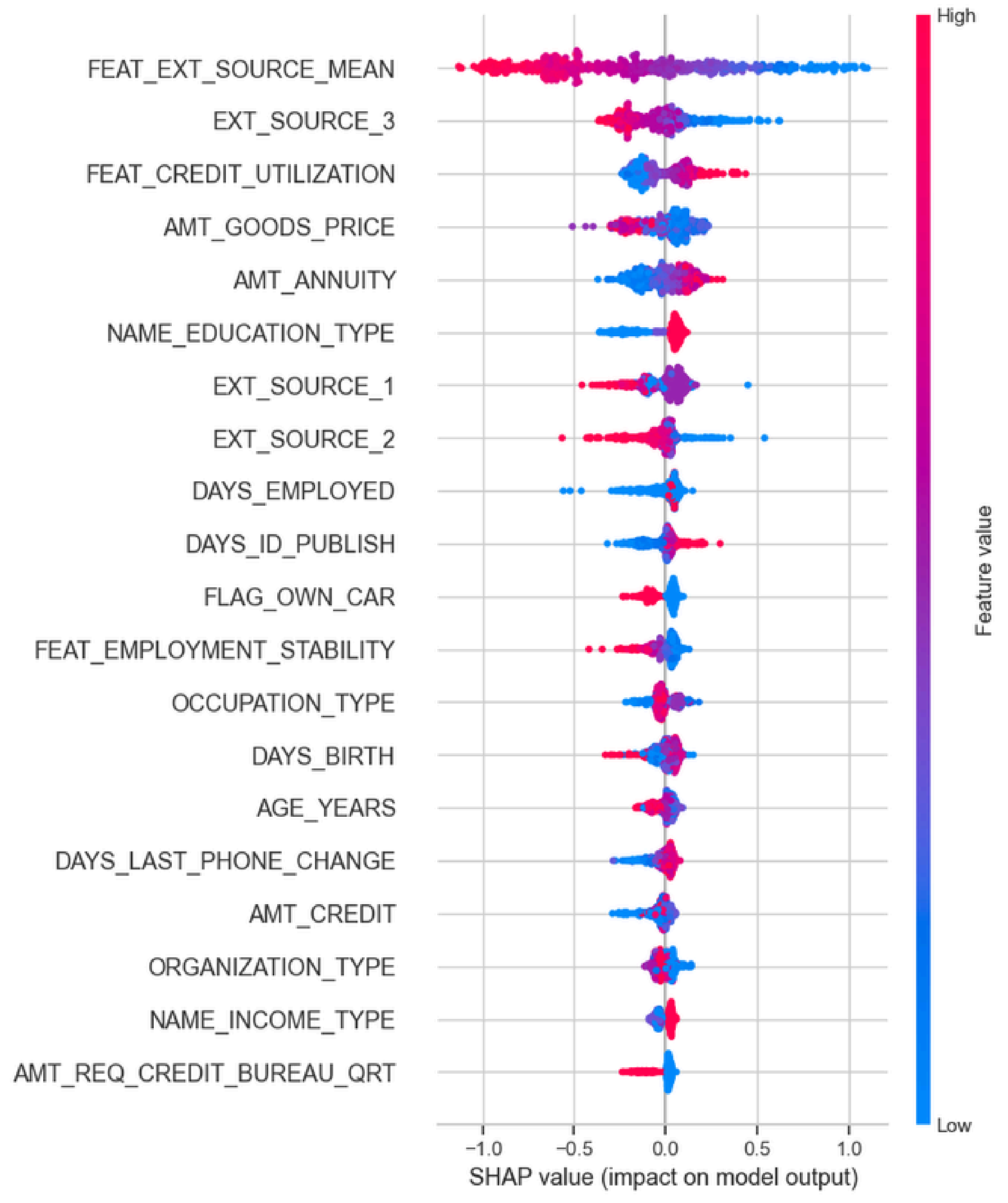
- AUC remained **0.7589** (unchanged, as AUC is threshold-independent)
- Minor changes in F1 and precision were accepted as a deliberate fairness trade-off




```
🏛 AI-POWERED ALTERNATE CREDIT SCORING — FINAL SUMMARY
Barclays Hackathon | Financial Inclusion Track

BEST MODEL      : XGB_Baseline

METRIC              BASELINE     FAIR-ADJUSTED
AUC-ROC             0.7589       0.7589
F1-Score            0.2755       0.2764
Precision           0.1748       0.1760
Recall              0.6489       0.6431
DI Ratio (Age)      0.6932       0.8025
Approval Young      0.4486       0.5702
Approval Senior     0.7105       0.7105
Four-Fifths Rule    ❌ FAIL      ✅ PASS

📊 ALL MODEL PERFORMANCE SUMMARY:
              AUC-ROC  F1-Score  Precision  Recall  Accuracy
LR_Baseline   0.7449   0.2555    0.1575     0.6757  0.6820
...
```

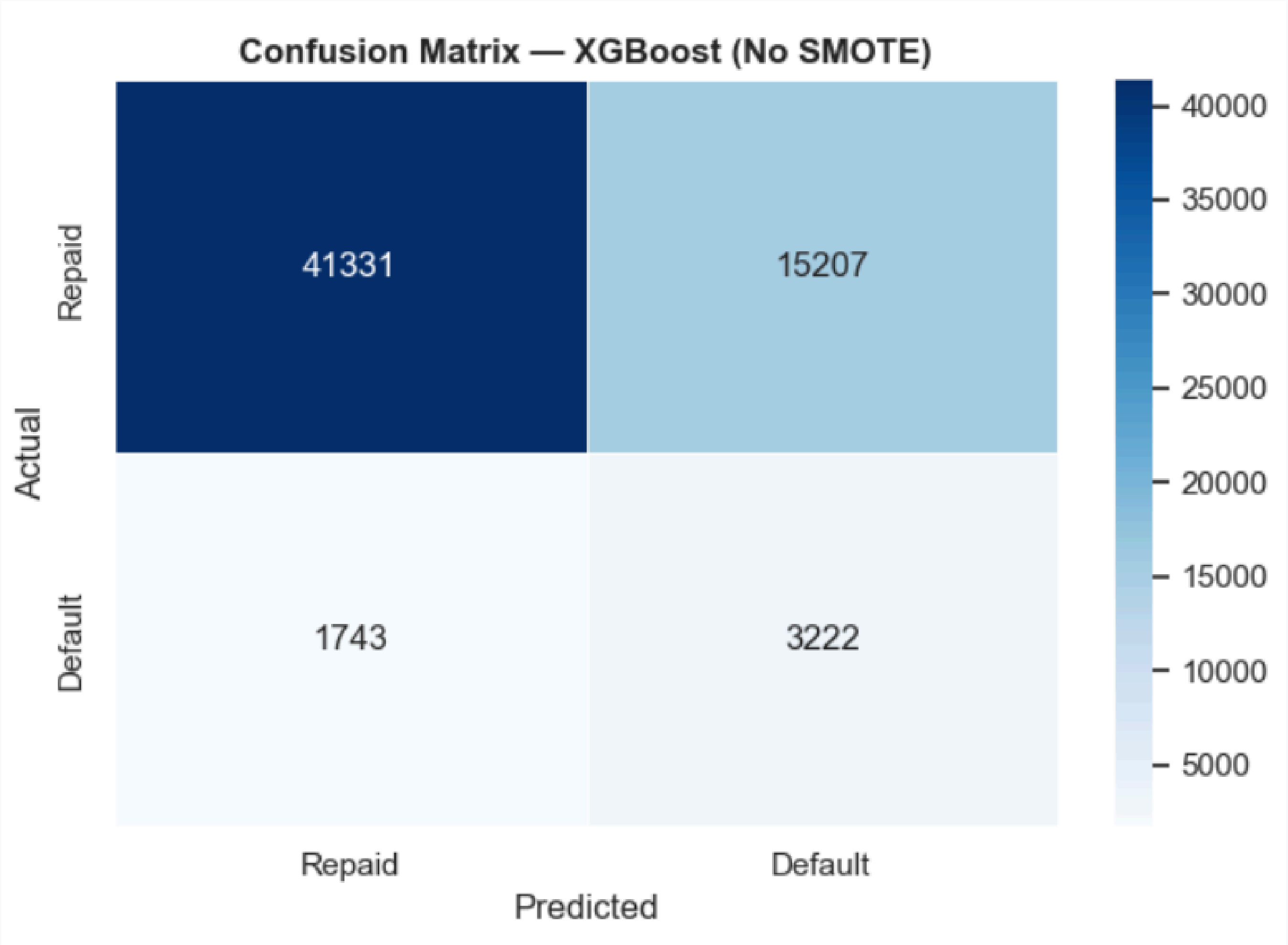
“Bias was corrected without sacrificing ranking power — preserving model integrity while ensuring equitable decision outcomes.”

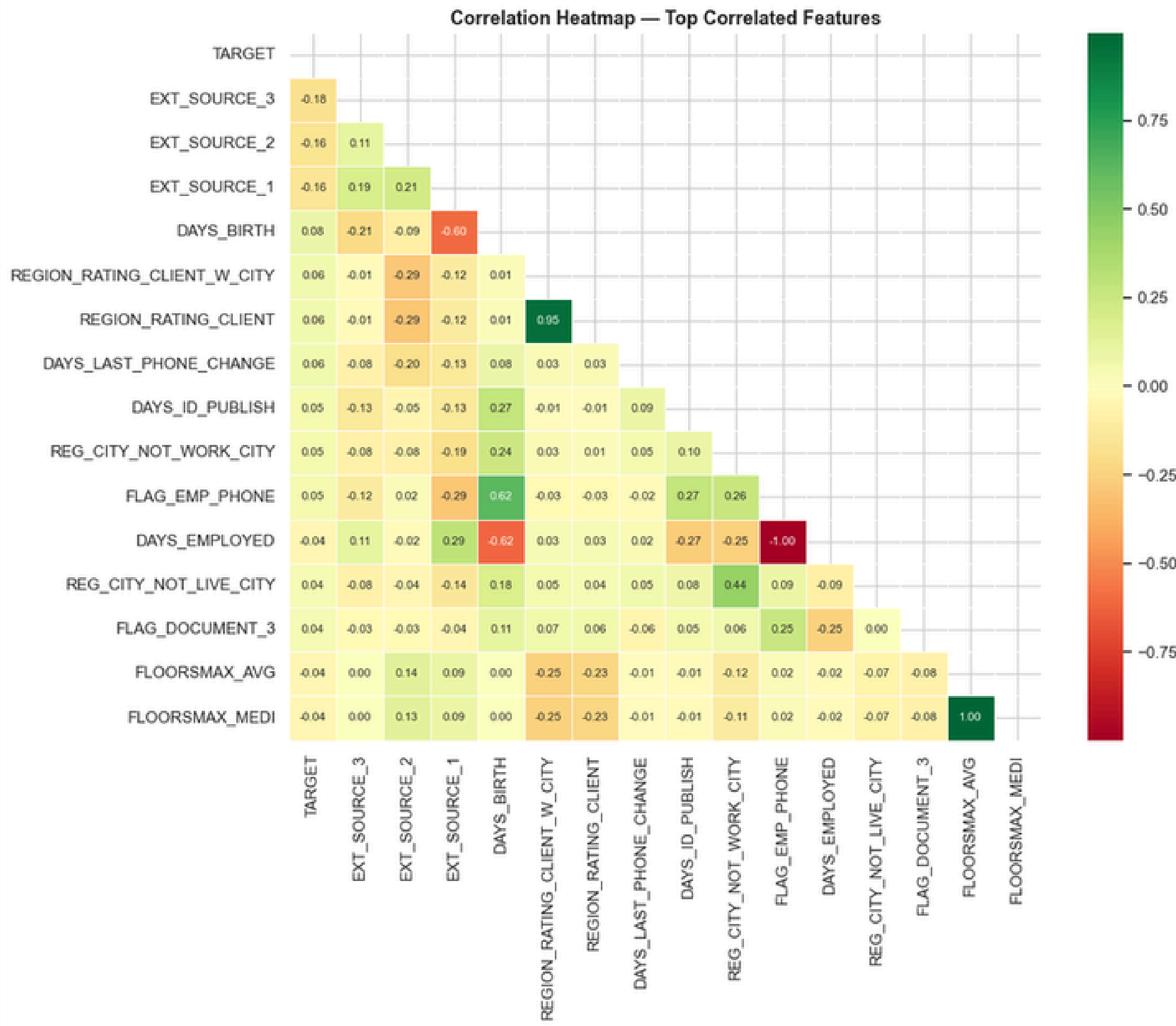




Home Credit

Key Model Outputs & Validation Results

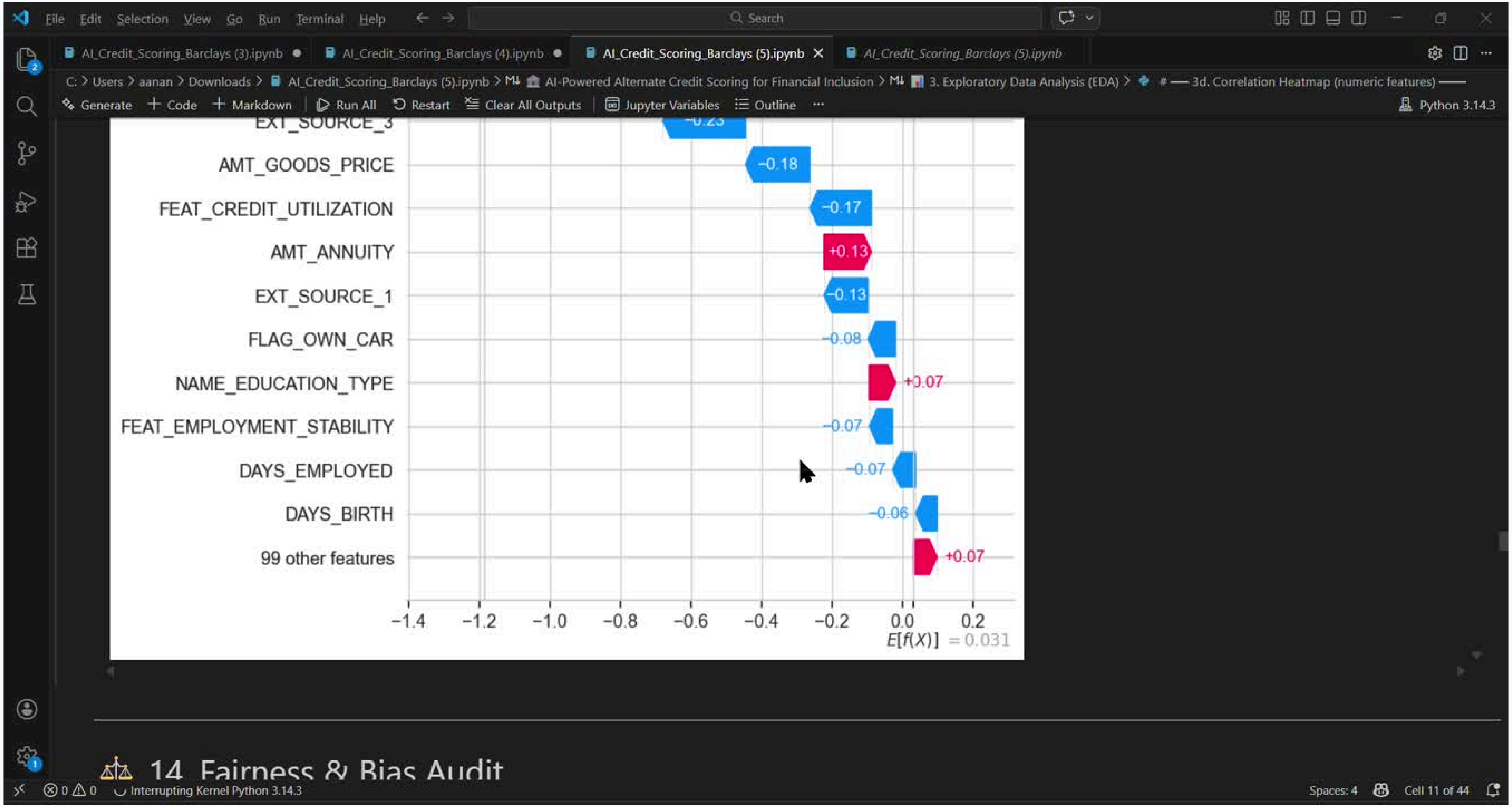




Home Credit

Transparent Credit Decision: Feature-Level Impact

Each bar shows how a specific feature increased or decreased this applicant's default risk. Blue lowers risk, red increases risk.



System Architecure of model based on 2nd Dataset

On Dataset:Give me some credit(link)

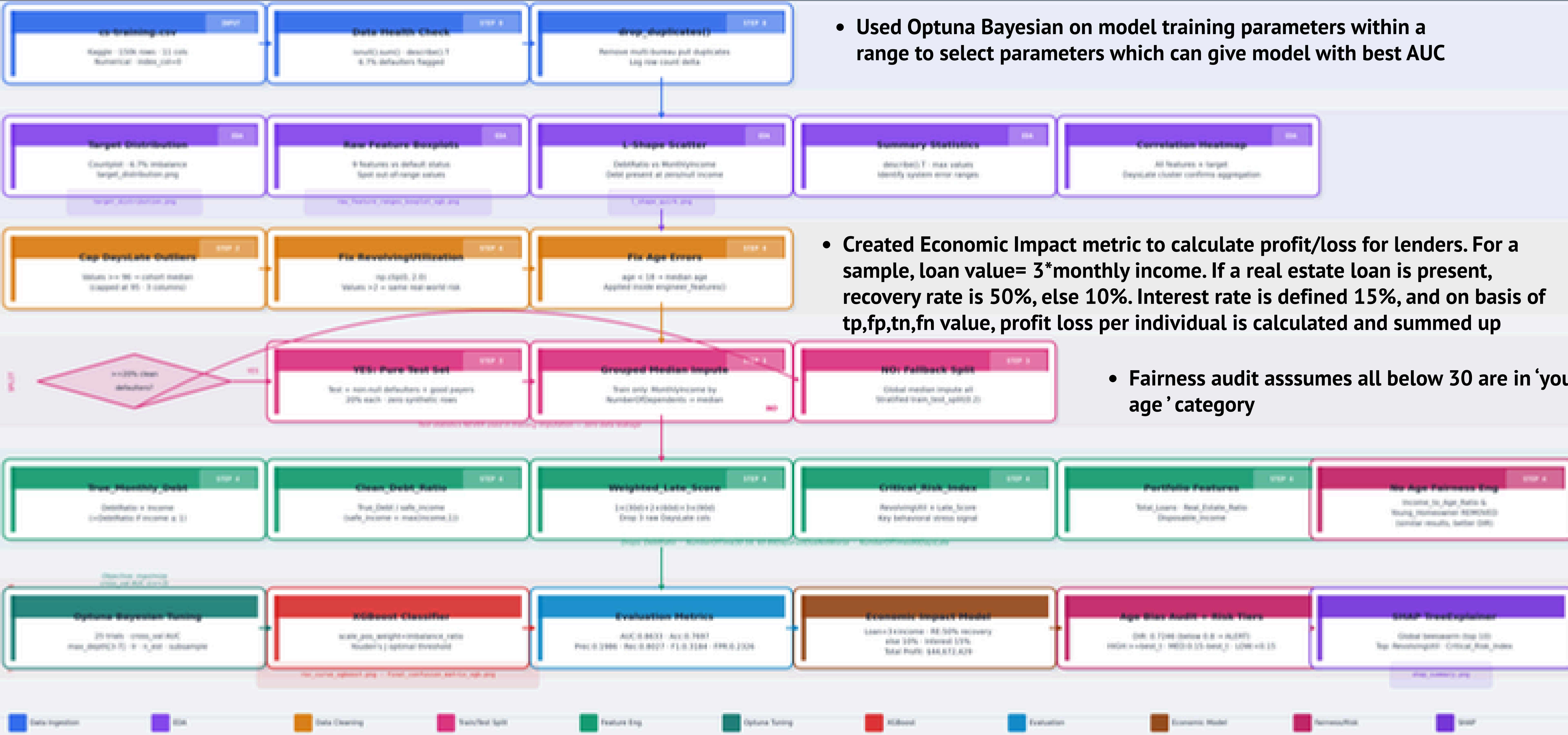
- Creating histogram for defaulter(target=1) and non-defaulter(target=0) count
- EDA + analysing past heatmap to derive new features+ identifying Monthly income v/s DebtRatio relationship
- Fixing outliers ; Payment delay can't exceed 95 days in 2 years; To prevent skewed median, capped delay values and calculated. Utilization capped to 2 as values>2 has same real world impact as utilization of value 2
- Used class weighing to ensure the model is trained mostly non-synthetic data. Hence, since defaulter percentage is around 6.7%, each defaulter will be 93.3/6.7=~14 times more important

Load Data & Deduplicate → EDA (L-Shape Analysis & Outliers) → Clean System Errors → Smart Train/Test Split & Imputation → Feature Engineering & Outlier Capping (Clean Debt) → Standard Scaling → Logistic Regression (Class-Weighted) → Youden's J Threshold Optimization → Evaluation (AUC, Heatmap, Sigmoid) → Age Bias Audit (4/5ths Rule) → Risk Tier Categorization (High/Medium/Low Risk)

System Architecure of model based on 2nd Dataset

CREDIT RISK ML SYSTEM — XGBOOST PIPELINE

On Dataset:[Give me some credit(link)]



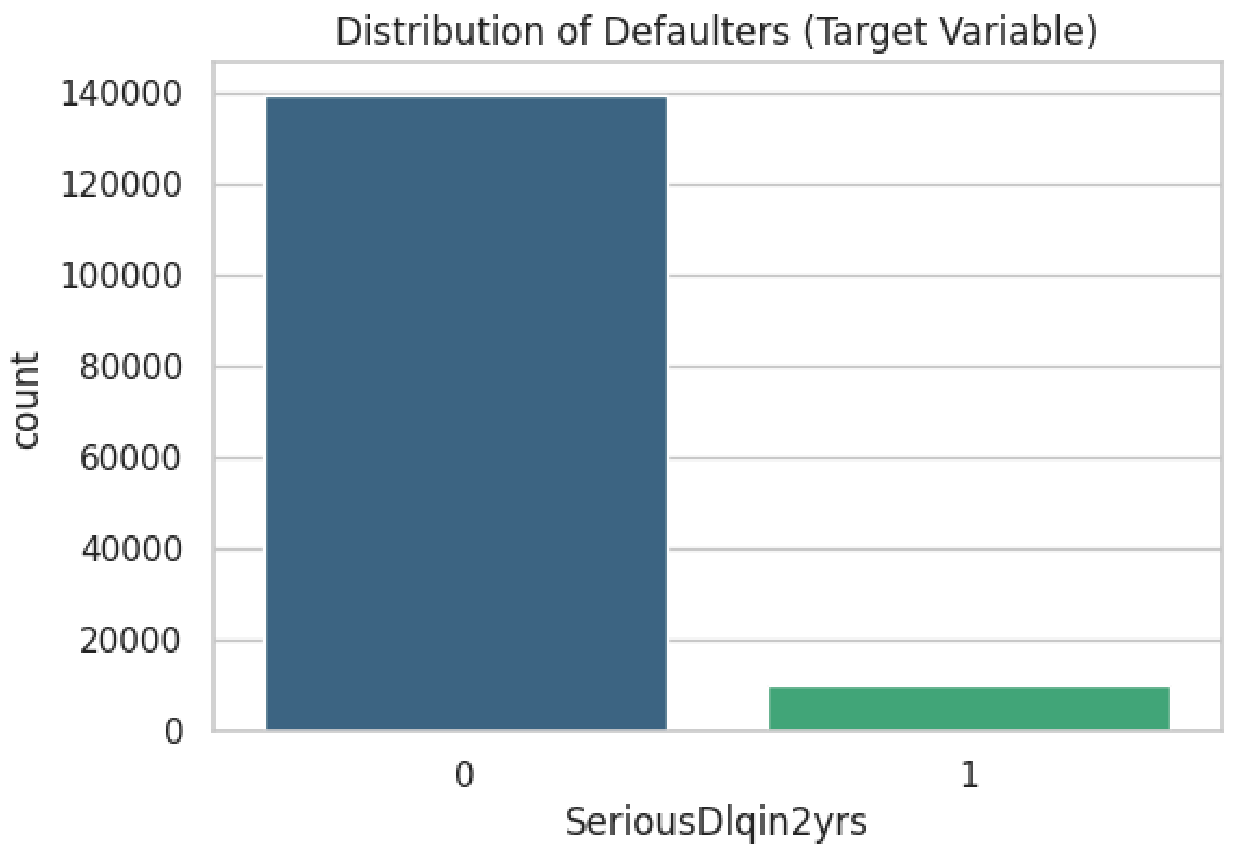
- Used Optuna Bayesian on model training parameters within a range to select parameters which can give model with best AUC

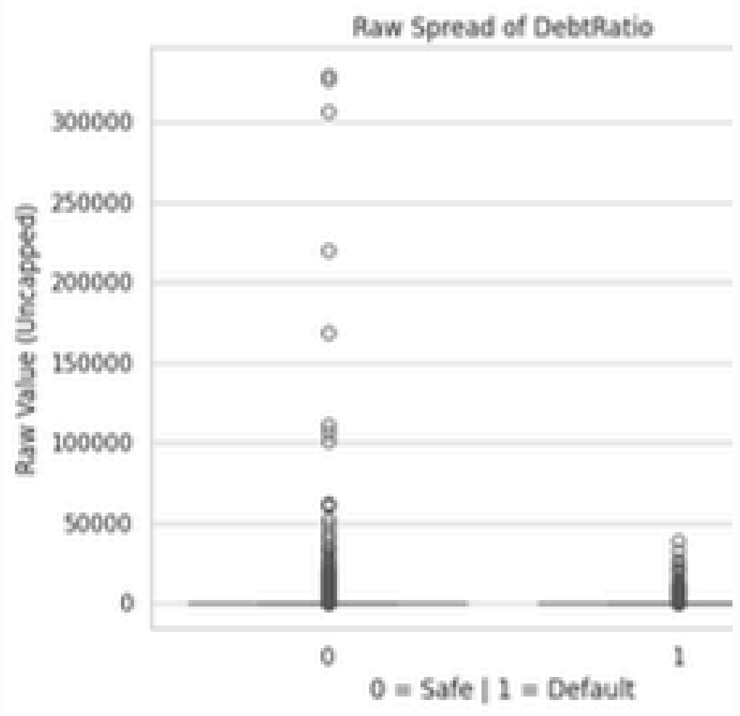
- Created Economic Impact metric to calculate profit/loss for lenders. For a sample, loan value= 3*monthly income. If a real estate loan is present, recovery rate is 50%, else 10%. Interest rate is defined 15%, and on basis of tp,fp,tn,fn value, profit loss per individual is calculated and summed up

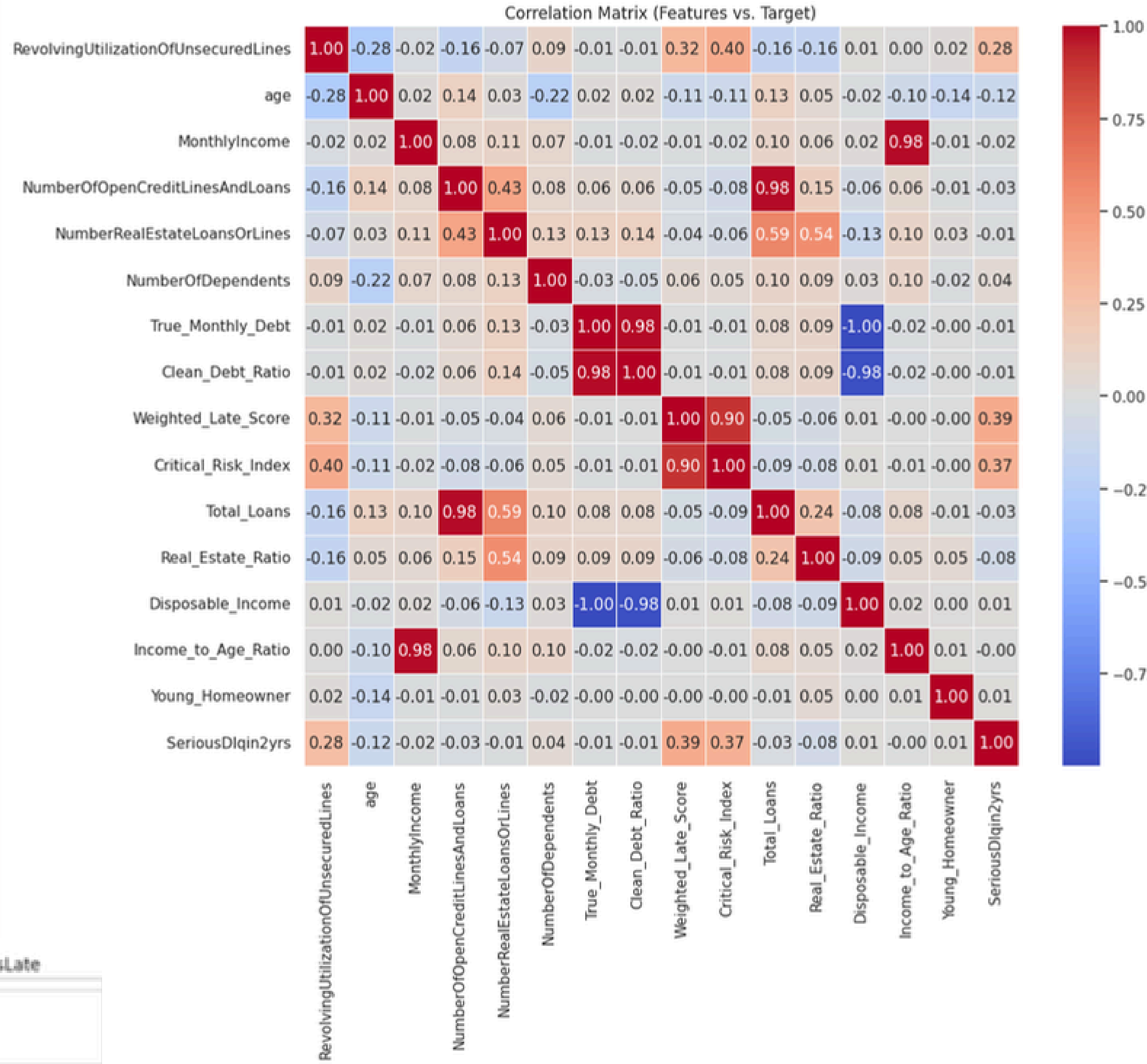
- Fairness audit asssumes all below 30 are in 'you age' category

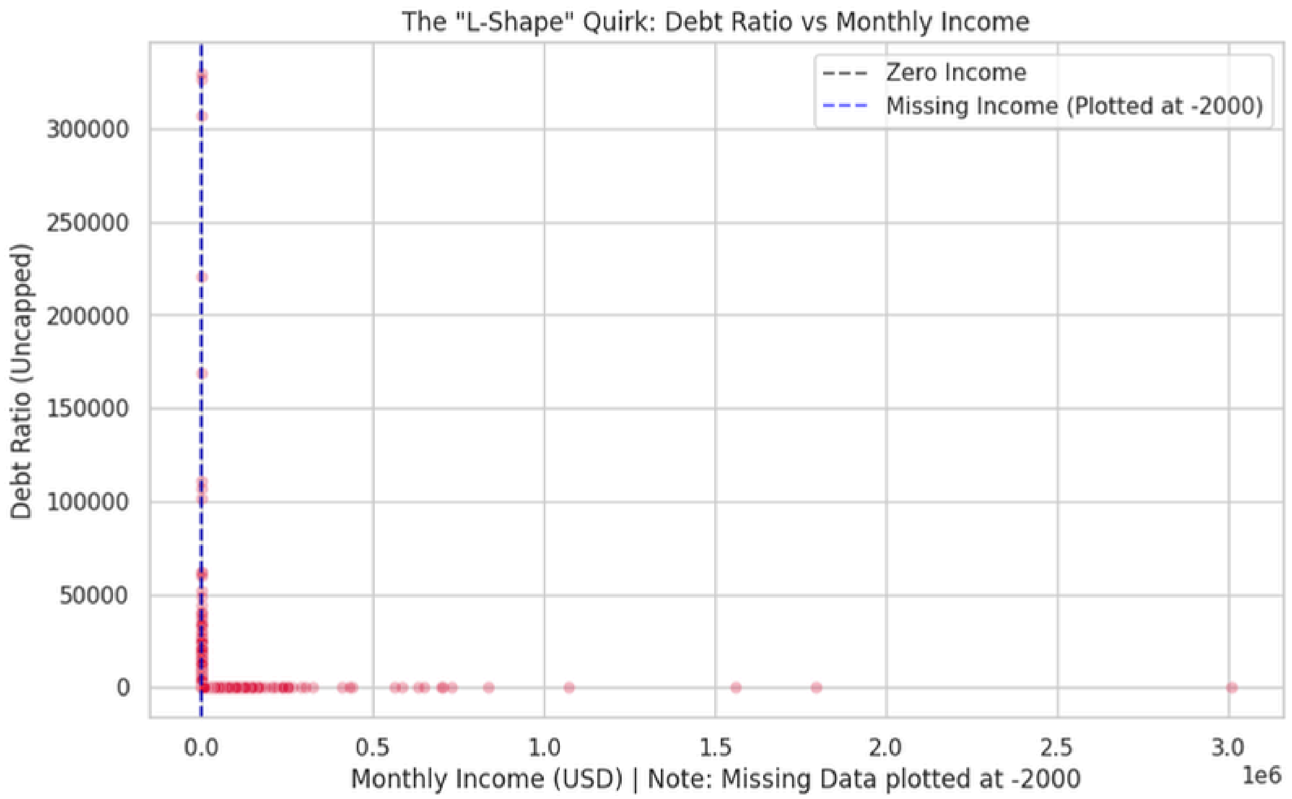
Load Data & Deduplicate → EDA → Clean System Errors → Smart Splitter & Imputation → XGBoost Feature Engineering (No Age Bias Engineering) → Class Weighting (Imbalance Ratio) → Optuna Bayesian Hyperparameter Tuning (25 Trials) → Optimal Model Training (XGBClassifier) → Youden's J Threshold Selection → Evaluation Metrics (AUC, Recall, FPR) → Economic Impact Simulation (Dynamic Recovery Rates) → SHAP Explainability→ Age Bias Audit (4/5ths Rule) →Risk Categorization

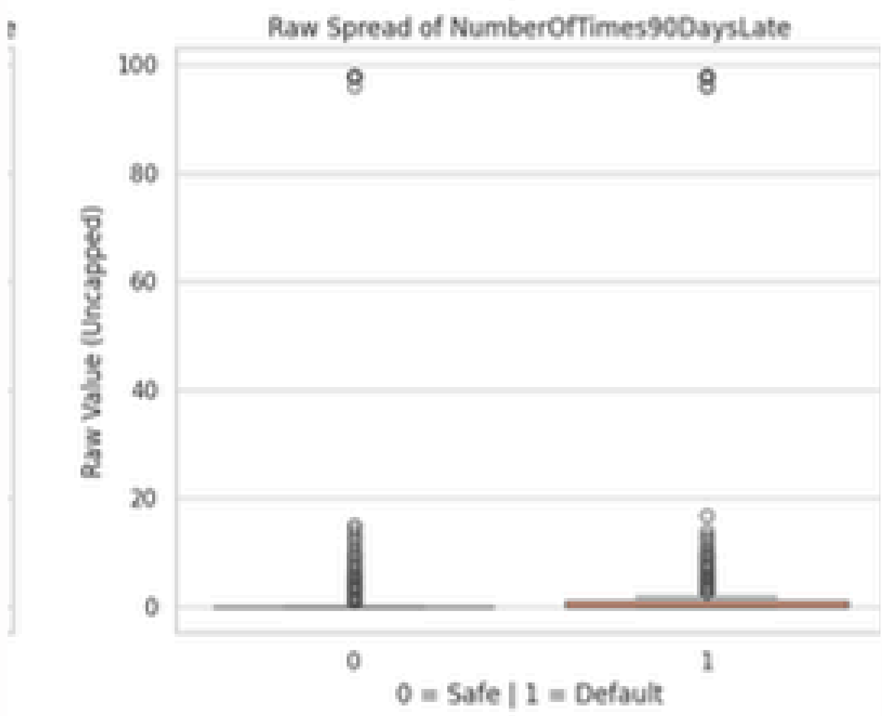
Important EDA data and SHAP outcome:











SHAP Importance Rank	XGBoost Model	Logistic Regression Model
1st Most Important	Revolving Utilization of Unsecured Lines	Weighted Late Score
2nd Most Importan	Critical Risk Index	Revolving Utilization of

Notable steps from workflow

1. **Multiple iterations tried** by varying imputing method, features set, parameters, splitting techniques, data ranges.
2. **Identified DebtRatio is actually Debt for samples with null/0 monthly income** value by plotting a L shaped monthly income v/s debt ratio graph.
3. **Splitting done to ensure non-synthetic test** data through non-null sample selection; imputed data post splitting.
4. **Cleaned DebtRatio feature by first deriving True_Monthly_Debt and replacing it with Clean_Debt_Ratio**;Both new features obtained on the basis monthly data type(0/non zero).
5. Created **behavioral features** like Critical_Risk_Index; **age fairness, transaction flow and stability** based features.
6. Used **class weighing instead** of SMOTE for imbalance data
7. Used **Youden's J Statistic(tpr-fpr) to find optimum threshold**
8. Introduced metric age fairness auditing**(Disparity impact) based on 4/5th ratio**
9. Out of multiple iterations final two were selected, for with and without age fairness features, out of which **one with age fairness features was finalized for logistic regression algorithm** after analyzing evaluation parameters.

Metric	Selected Model (With Age Fairness Features)	Comparison Model (No Age Fairness Features)
AUC Score	0.8557	0.8557
Optimal Threshold	0.4991	0.4794
Accuracy	0.7774	0.7646
Precision	0.1997	0.1925
Recall	0.7722	0.7867
F1-Score	0.3174	0.3093
Young Approval Rate (<30)	53.11%	51.99%
Senior Approval Rate (30+)	75.43%	73.93%
Disparate Impact Ratio	0.7041	0.7032

Notable steps from workflow

1. Observed **non-linear target to feature relationship** by comparing correlationship of feature to target based on heatmap and SHAP list, hence tried XGBoost
2. **Created iterations** where splitting technique was tweaked, null values weren't changed, went ahead with a model based on the same data analysing, data cleaning, data pre-processing and feature engineering steps as Logistic Regression.
3. Used **Optuna Bayesian for tuning XGBoost model training parameters** within an ideal range
4. **Created metric Economic impact** to hypothetically assess profit/loss for lenders
5. Finalised two models, one with age fairness features, one without, **chose the one without age fairness** for more favorable evaluation parameters.

Metric	Model 1(No age fairness features)	Model 2((With age fairness features)
AUC Score	0.8633	0.8644
Accuracy	0.7697	0.7692
Precision	0.1986	0.1983
Recall (TPR)	0.8027	0.8037
F1-Score	0.3184	0.3182
False Alarm Rate (FPR)	0.2326	0.2333
Total Projected Profit	$44,672,429.25	$44,409,549.75
Profit Per Applicant	$1,495.11	$1,486.31
Young Approval Rate (<30)	53.72%	52.32%
Senior Approval Rate (30+)	74.14%	74.16%
Disparate Impact Ratio	0.7246	0.7056

Conclusion for Dataset 2 :

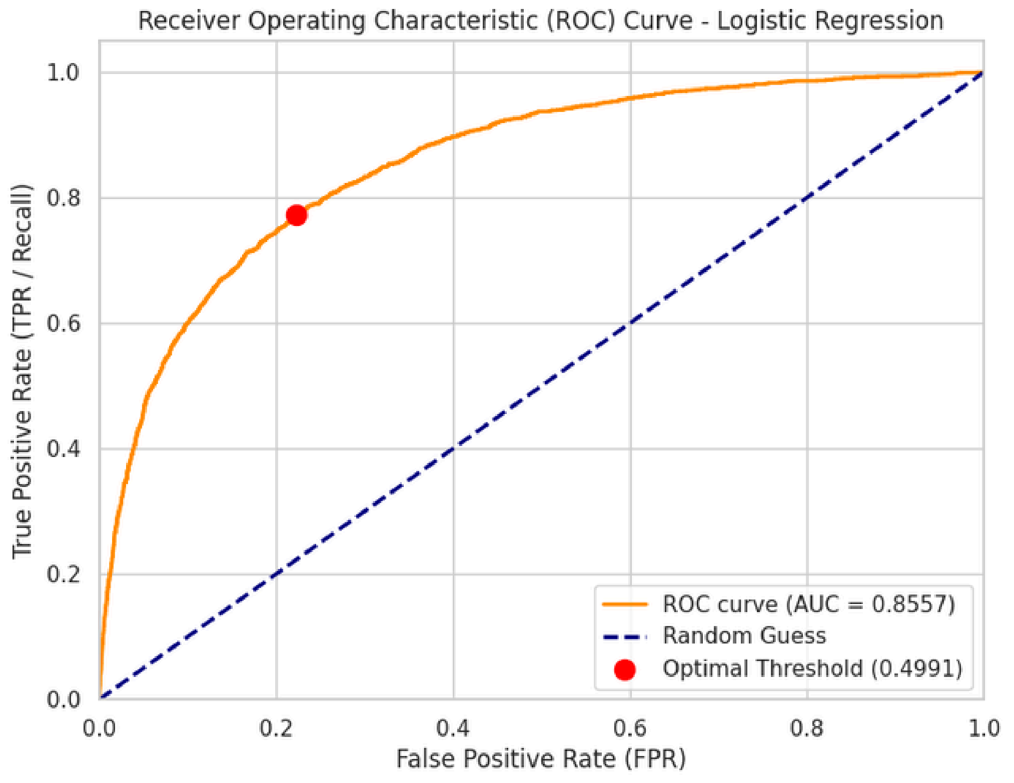
- Problems aimed: Balanced data, optimum feature engineering with behavioral features, fixing age based credit lending disparity, identifying optimum parameters, balancing different evaluation parameters, giving user interpretable output.

- Observations and Conclusion:
- We have countered most problems aimed by obtaining well-acceptable Evaluation Parameters of AUC, recall, through both algorithms.
- Precision and F1 value however is low .
- **We'll prefer the XGBoost algorithm based model with age fairness metrics with due to better evaluation parameter values and its compatibility with our dataset where features can have priority ranks,**
- **However, the age bias problem hasn't been fixed. Hence, we can further optimise this model or we can use a more comprehensive dataset through which more age bias removing features can be engineered.**

- Minor changes possible: Using more optimum data splitting techniques, trying Random Forest algorithm, using KNN for imputing, finding ideal parameters to define 'young age' or 'medium risk threshold', adding synthetic features, finding optimum method to calculate economic impact.

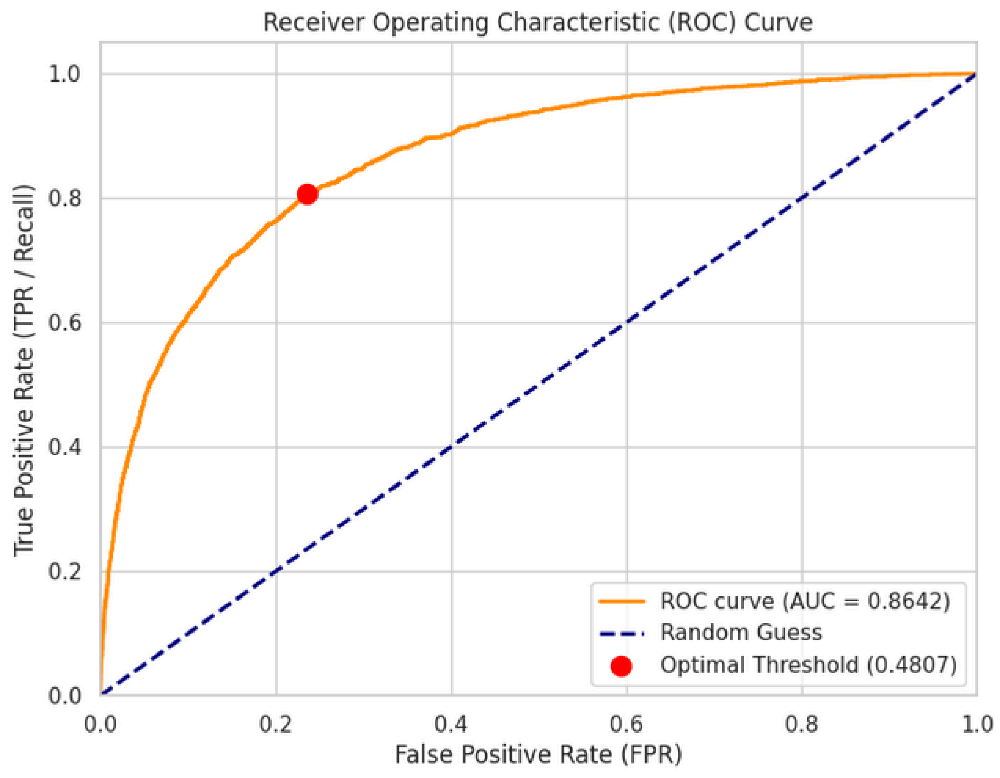
Risk Tier	Total Applicants	Actual Defaulters
High	7,741	1,546
Medium	15,533	422
Low	6,605	34

Logistic regression



Risk Tier	Total Applicants	Actual Defaulters
High	8,200	1,615
Medium	12,448	333
Low	9,231	54

XGBoost



Innovation ,Imapact & Future Roadmap

Novelty

- Fairness is embedded directly into the pipeline, not added later.
- Minimal adjustment applied to satisfy the four-fifths rule while maintaining AUC.
- Prediction, SHAP explainability, and compliance integrated into one system.

Business Impact

- Improves credit access for young and thin-file applicants.
- Loan officers receive probability score, risk band, and SHAP explanation for each case.
- Transparent trade-offs between performance and fairness.

Future Scope

- Add alternative data (utility, wallet, mobile data) to improve AUC.
- Deploy as a real-time scoring API.
- Automate fairness (DI) monitoring.
- Expand to MSME scoring and regulatory dashboard.

We identified a hidden fairness gap, corrected it responsibly, and made the trade-off transparent.
Modern credit systems must not just predict risk – they must prove they are fair and explainable.