Ironhack

FINAL PROJECT: CREDIT RISK PD MODEL :

María Martínez Gil

PROJECT FRAMING

- **Objective**: build a calibrated Probability-of-Default (PD) model to rank applicants by risk and set a clear approval threshold that aligns to a target loss appetite / volume goal.

- **Business use**: underwriting cutoffs, expected-loss (EL) sizing, capital planning, and pricing/line management.

EXECUTIVE SUMMARY ...

- Objective: Increase approvals while holding loss rate within guardrails.
- Primary KPI: KS (ranking power for cut-offs).
- Secondary: AUC, Brier/RMSE for probability quality.
- Decision unit: Approve/decline at PD threshold; support line assignment later.

Policy (Recent Backtest): Threshold ~0.010 → ~20% approvals, ~0.49% approved bad rate on recent cohort; overall bad-rate ≈ 2.28%.

Performance: KS ≈ 0.415, AUC ≈ 0.754, Brier (after cal.) ≈ 0.0110, RMSE ≈ 0.105.

Model: Gradient-boosted trees (LightGBM), 24-month default label.

DATA CONTEXT & LABELING

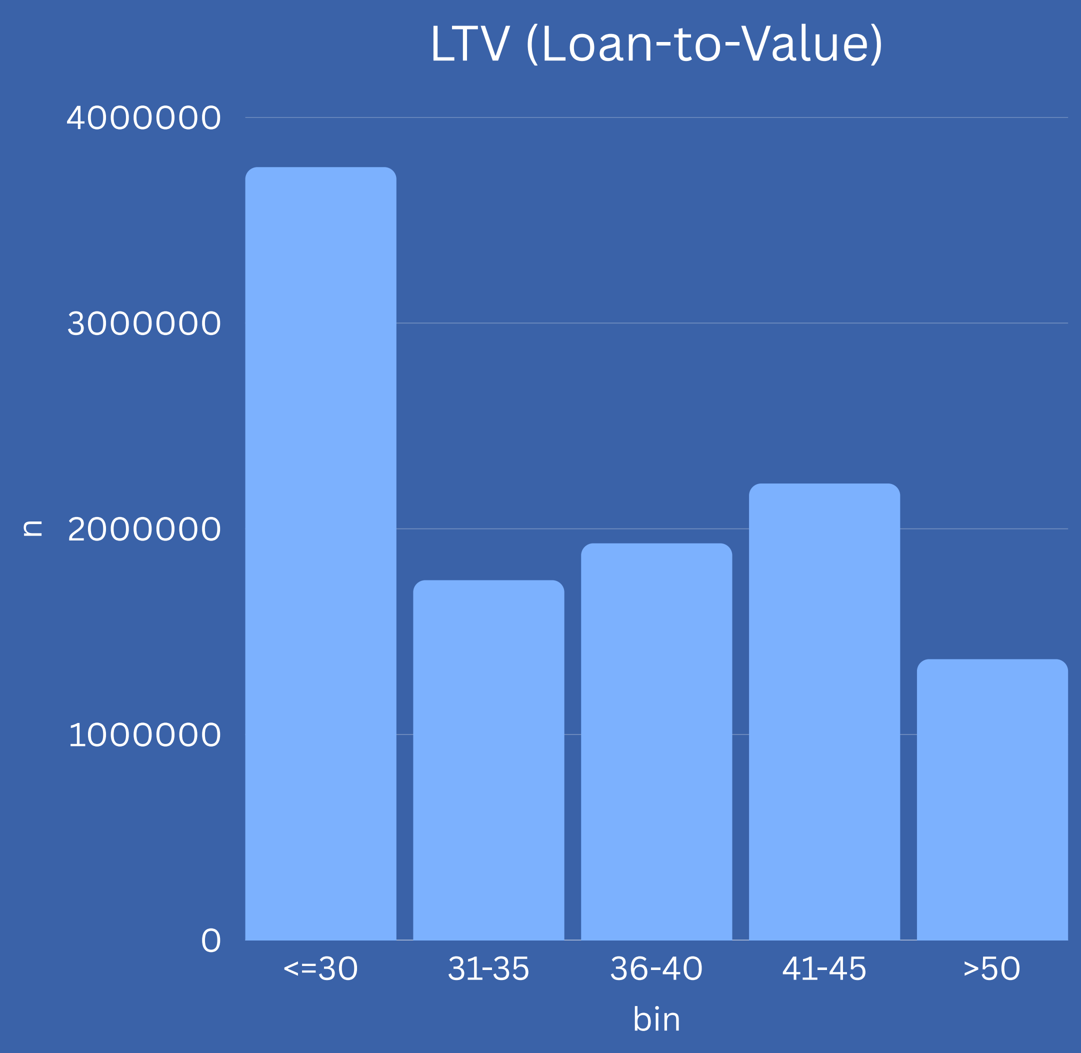
- **Source & scope:** Public Freddie Mac Single-Family Loan-Level dataset.
- **Time window used:** Originations 2017-2023.
- **Size (the build)**: Train 2,663,021 loans post-filters & Test: 5,751,411 loans from reports/metrics. Plus a recent 2023 hold-out of 163,210 loans.

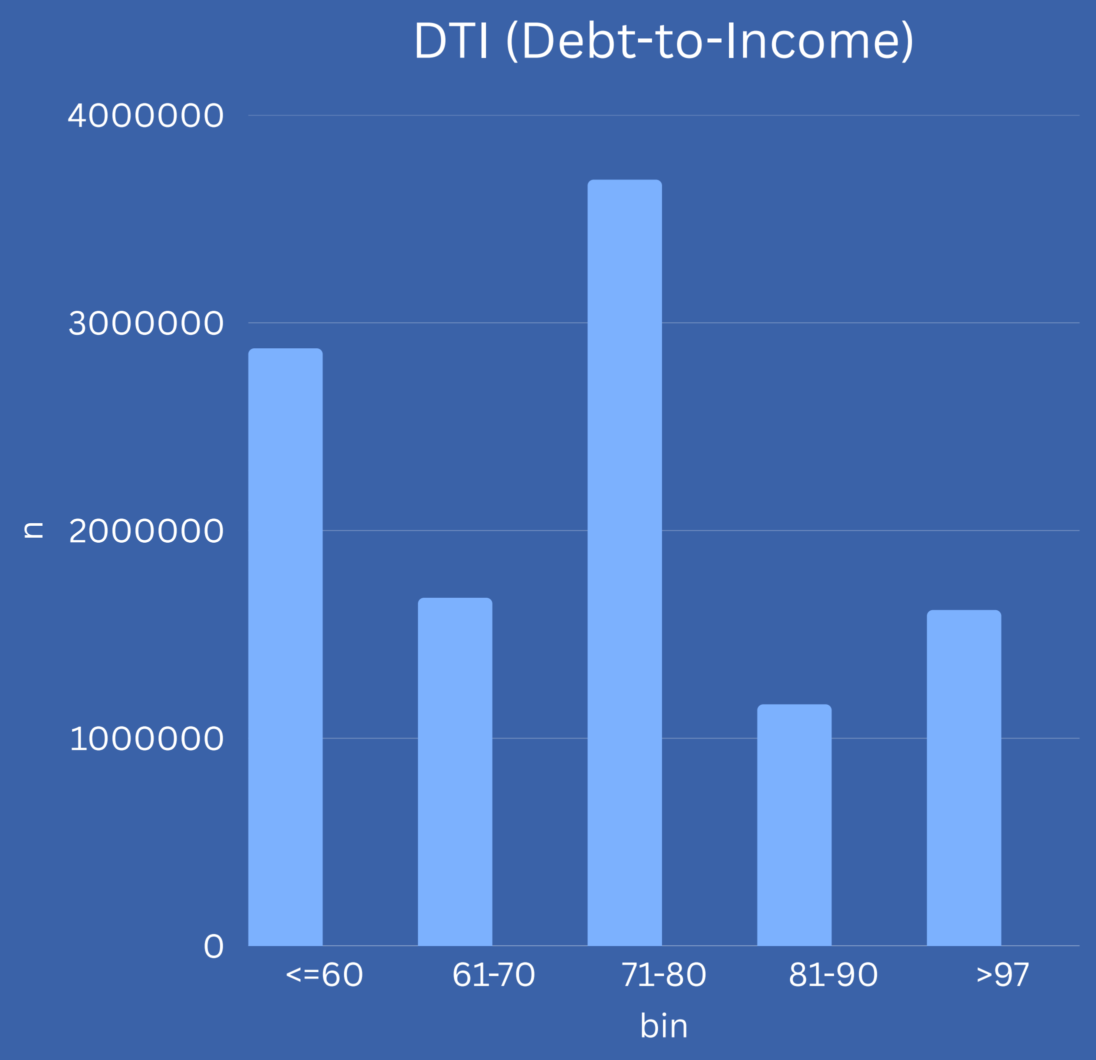
Split	Vintages	N loans	Pos rate
Train	2016-2019	2,663,021	1% (varies)
Valid	2020	(subset)	1% (varies)
Test	2021-2022	5,751,411	1,12%
Recent	2023	163,120	(used for back-test)

Freddie Mac

EDA OVERVIEW - KEY RISK DRIVERS :

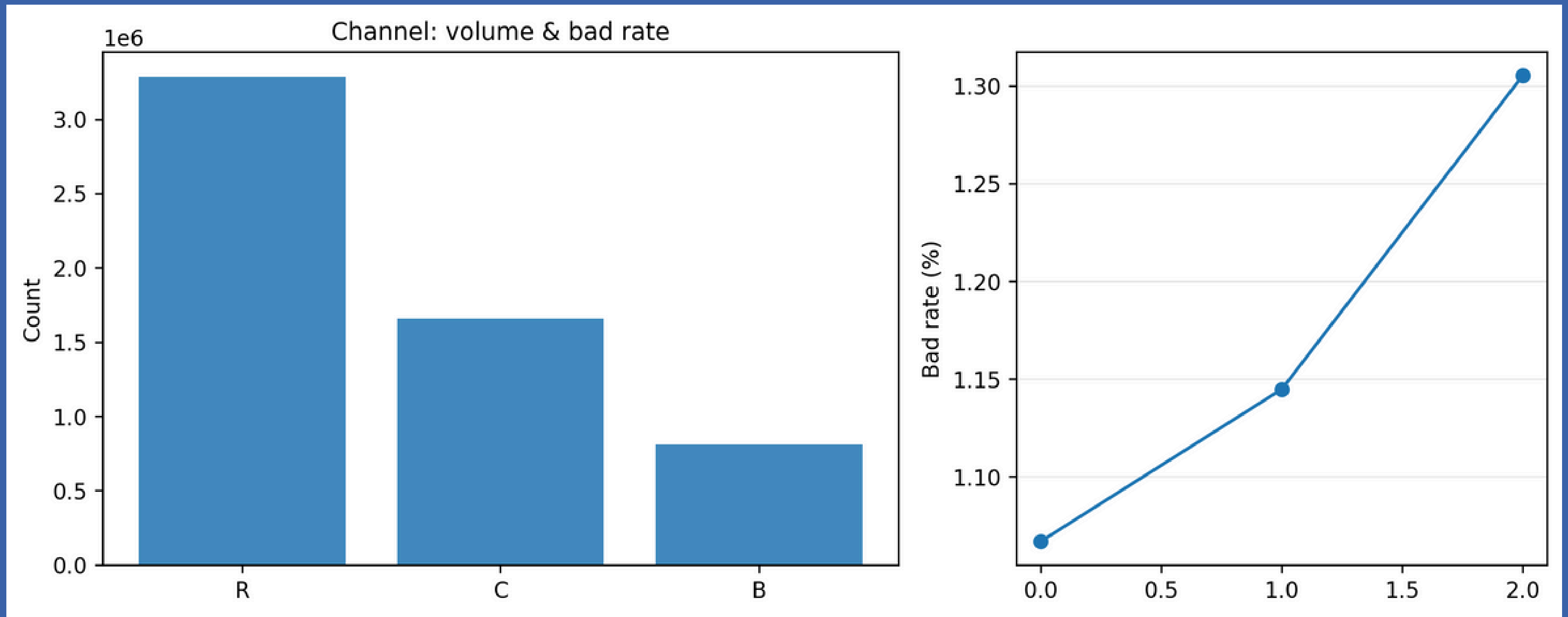
EDA covered univariate distributions (FICO/DTI/LTV), risk monotonicity checks, segment heterogeneity (channel/purpose/state), temporal drift (vintage), missingness and outliers. We binned key drivers and validated economic shape constraints prior to model fitting; results informed feature engineering and guardrails.

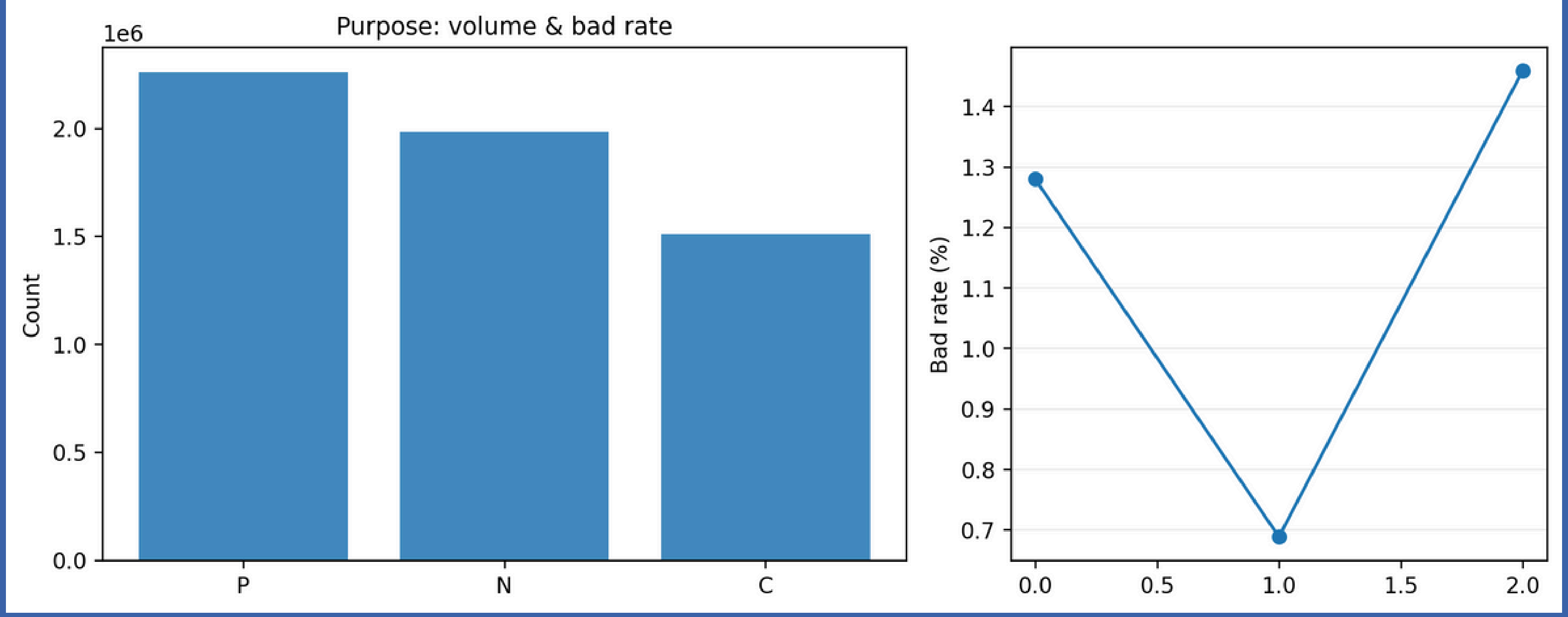




PORTFOLIO MIX: CHANNEL & PURPOSE

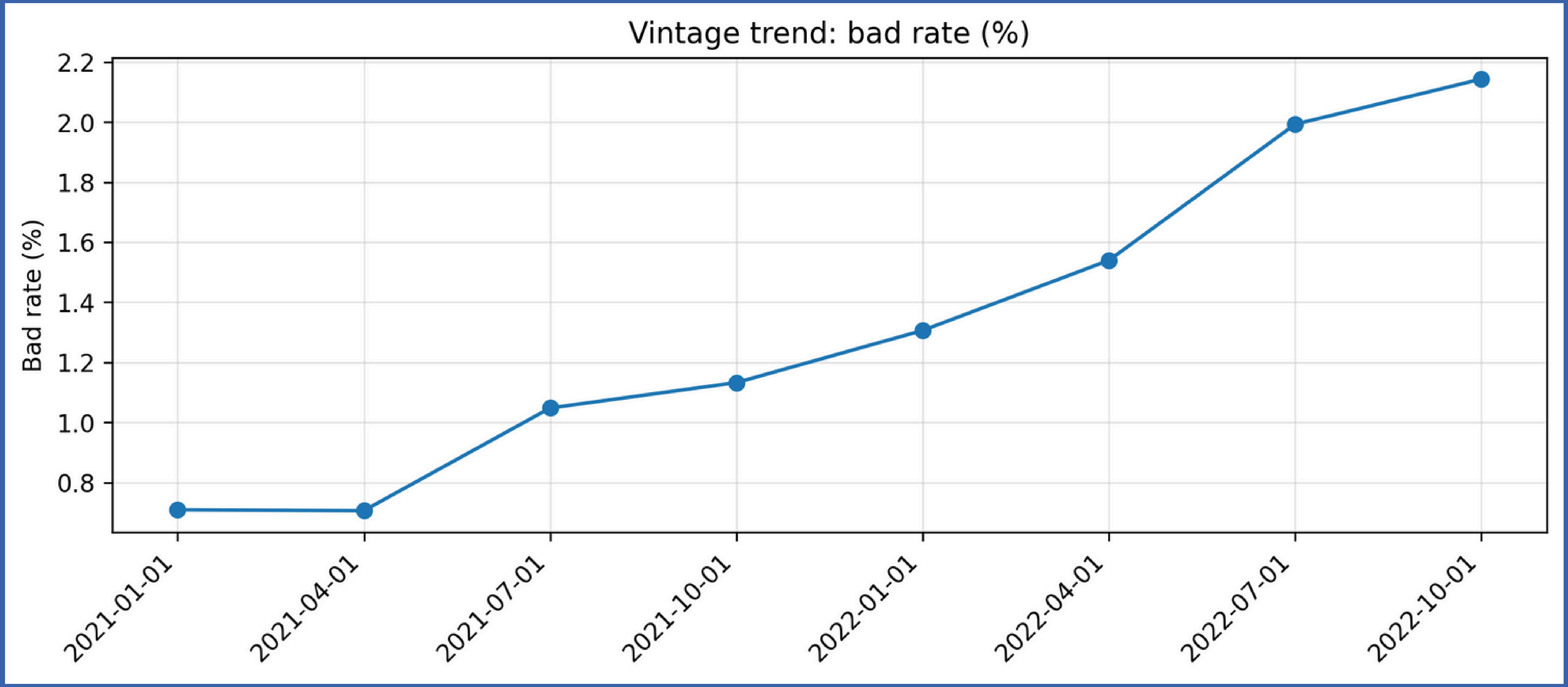
- Channel: Retail dominates; Broker slightly higher bad rate.
- Purpose: Cash-out refi = highest risk; No-cash refi = lowest; Purchase mid.



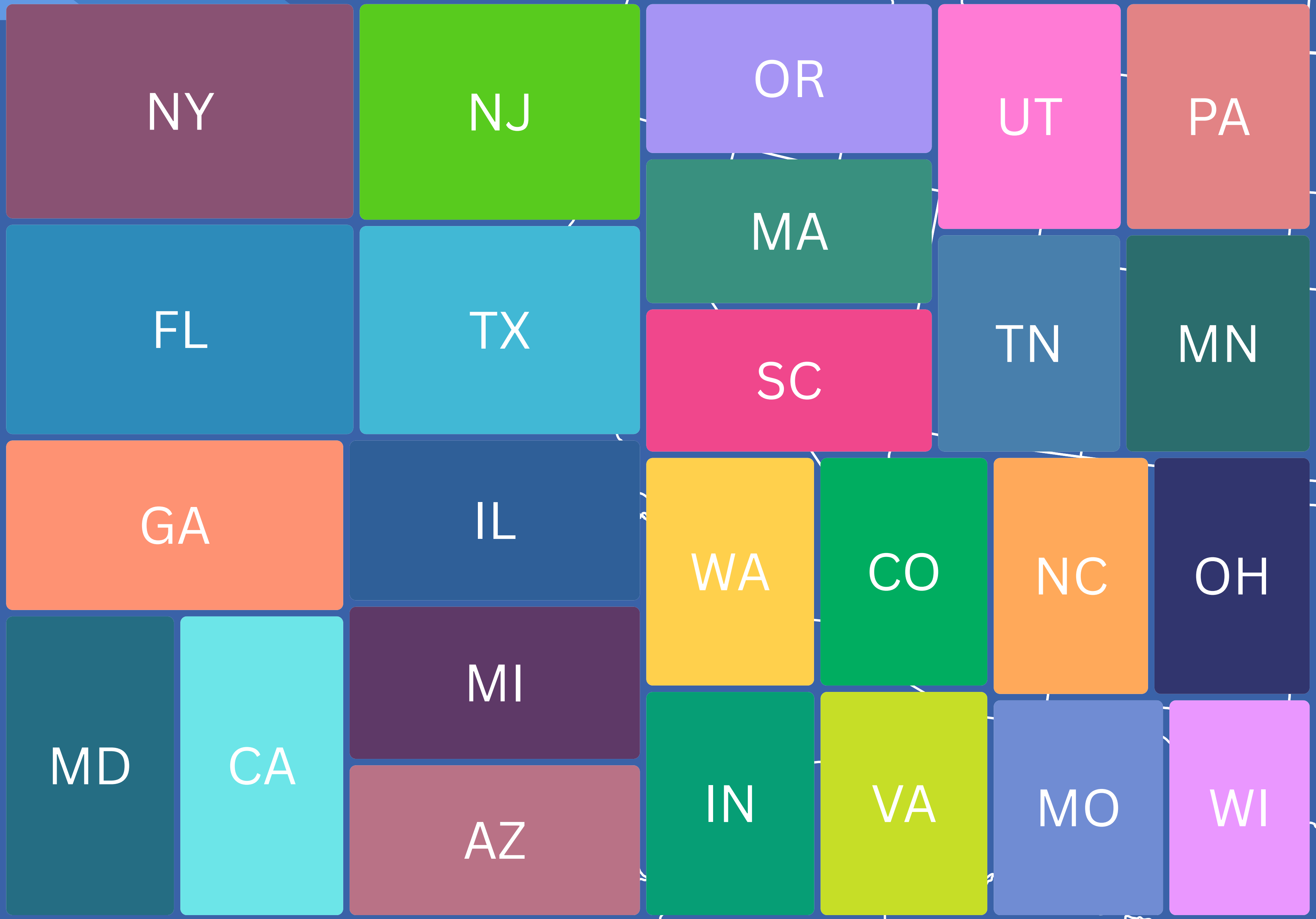


VINTAGE TREND (DRIFT)

Bad rate increased ~3× from early-2021 to late-2022; validates calibration and conservative thresholds.



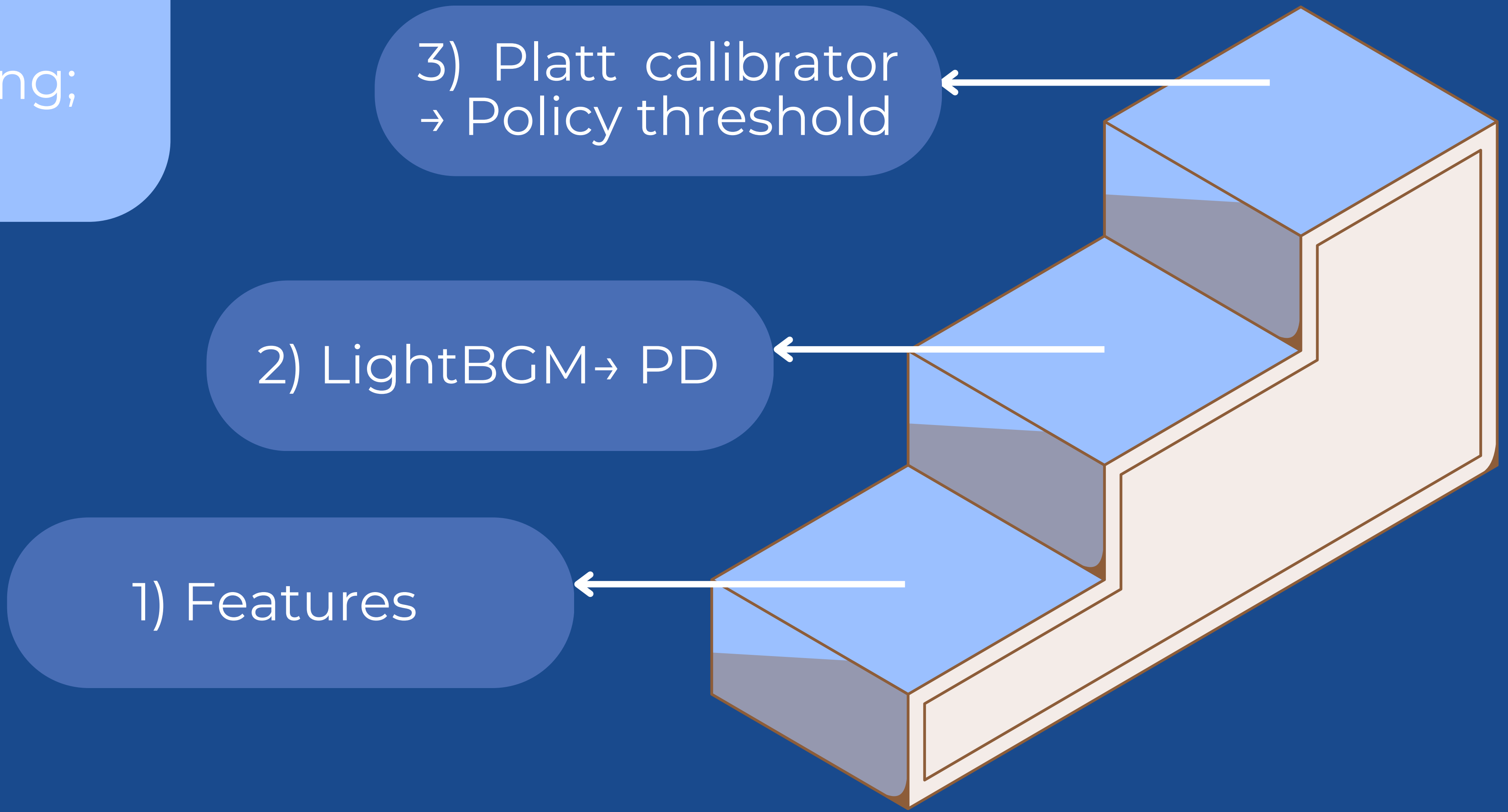
GEOGRAPHY – TOP STATES



- Volume is concentrated in a few large states; for example New York, New Jersey, Texas and Florida make up a big share.
- The rest is a long tail of smaller states.
- The top 10–12 states cover most of the book, so focus monitoring and pricing there.
- When you add risk coloring, tighten cut-offs where high volume meets above-average bad rate, and consider modest loosening where high volume pairs with below-average risk.

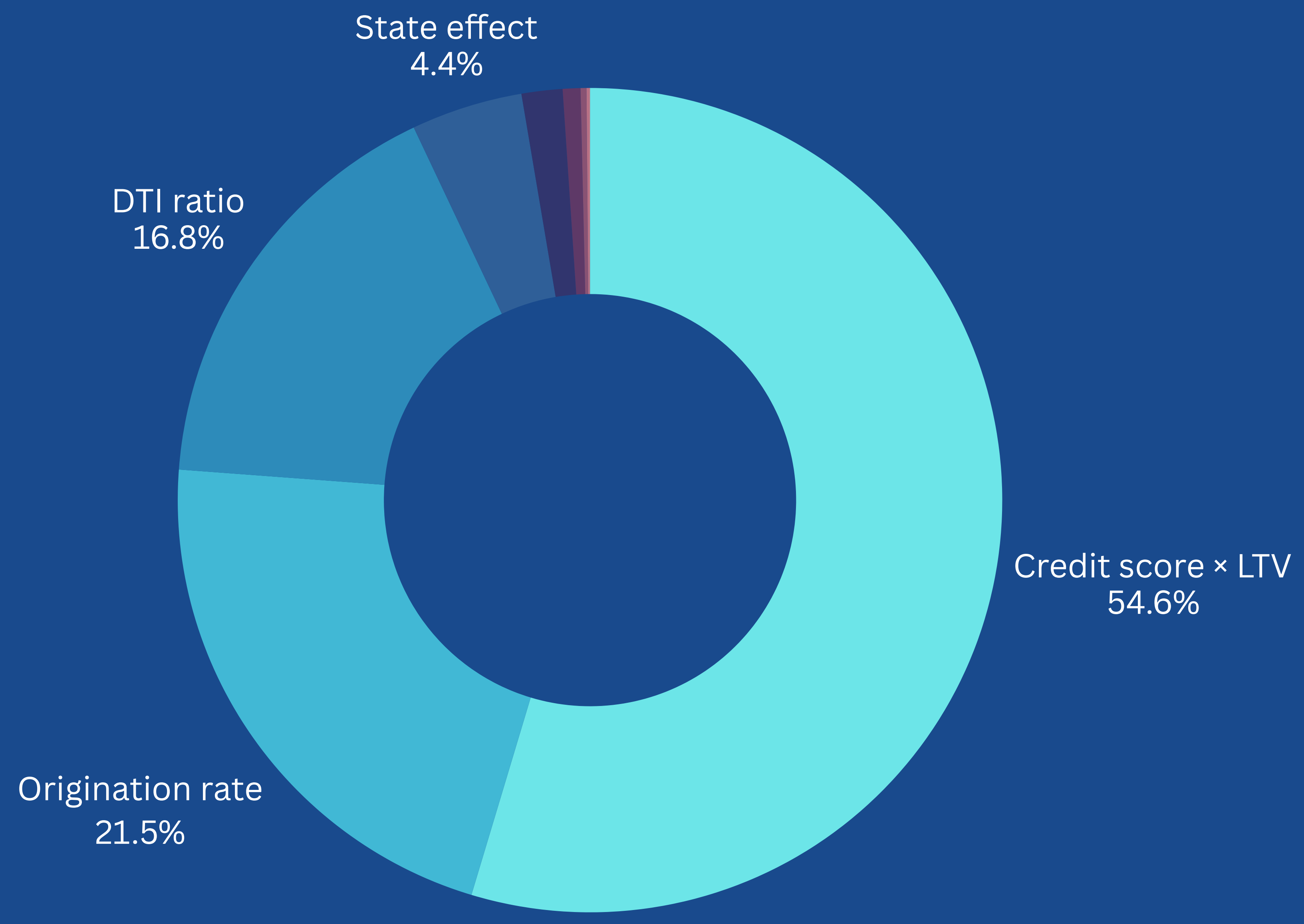
MODEL ARCHITECTURE & TRAINING SETUP

- Gradient-boosted trees (LightGBM), origination-time features only; 24-month default label.
- Time-split training: Train 2017–2019, Validate 2020 (early stopping), Test 2021–2022; 2023 held out for business backtest.
- Class imbalance handled via sampling/weighting; conservative regularization to prevent overfit.



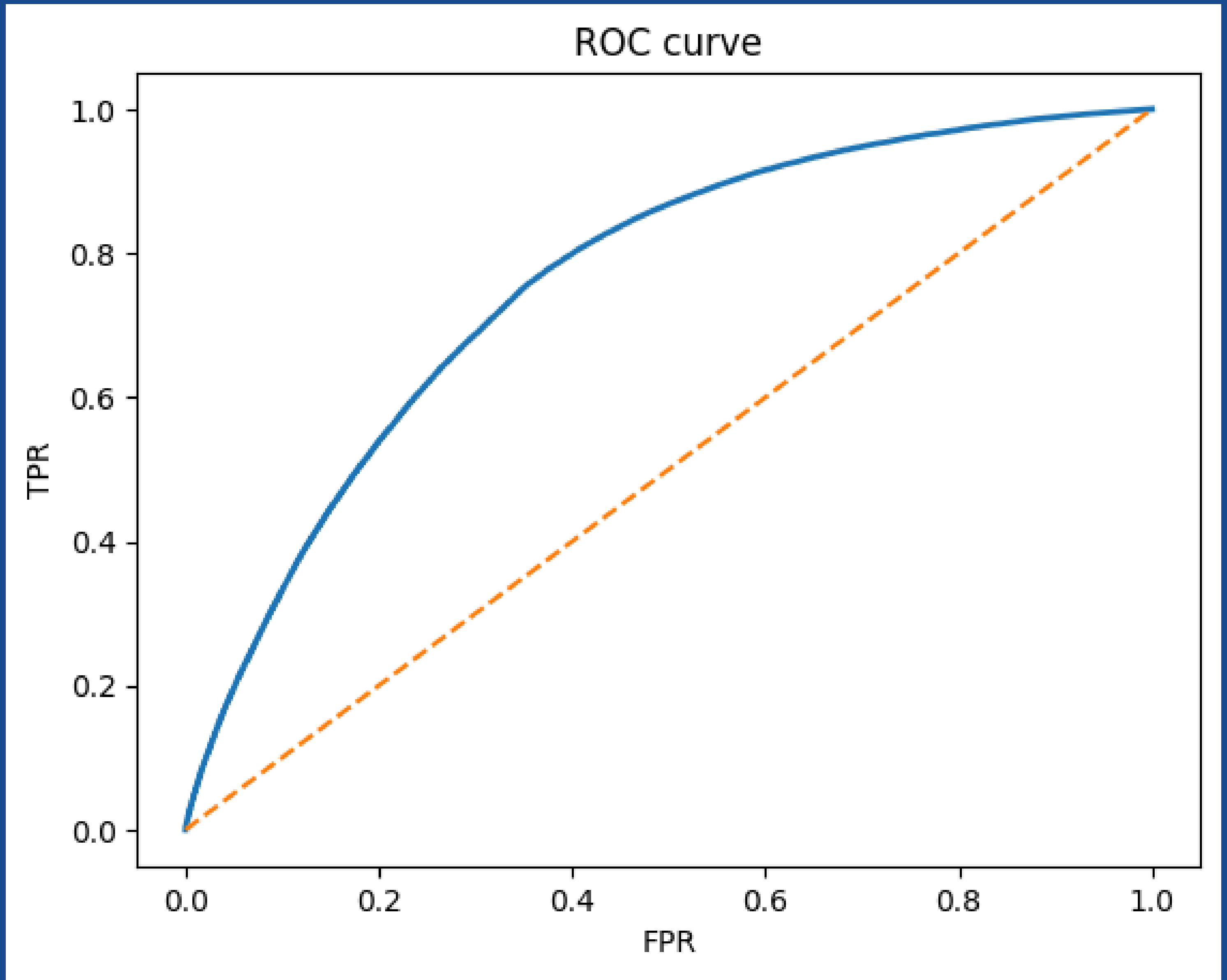
FEATURE SET & GOVERNANCE

- **Inputs at decision time**: credit score, DTI, LTV/CLTV, rate, purpose, channel, occupancy, property type, state, vintage.
- **Light engineering:** caps/bins and a few safe interactions. No macro variables in v1.
- **Risk intuition + calibration:** PD falls with higher score and rises with LTV/DTI; calibrated to reliable PDs.
- **Controls:** fixed time splits, versioned code/data, threshold saved in config, monthly monitoring (drift, calibration, slices).

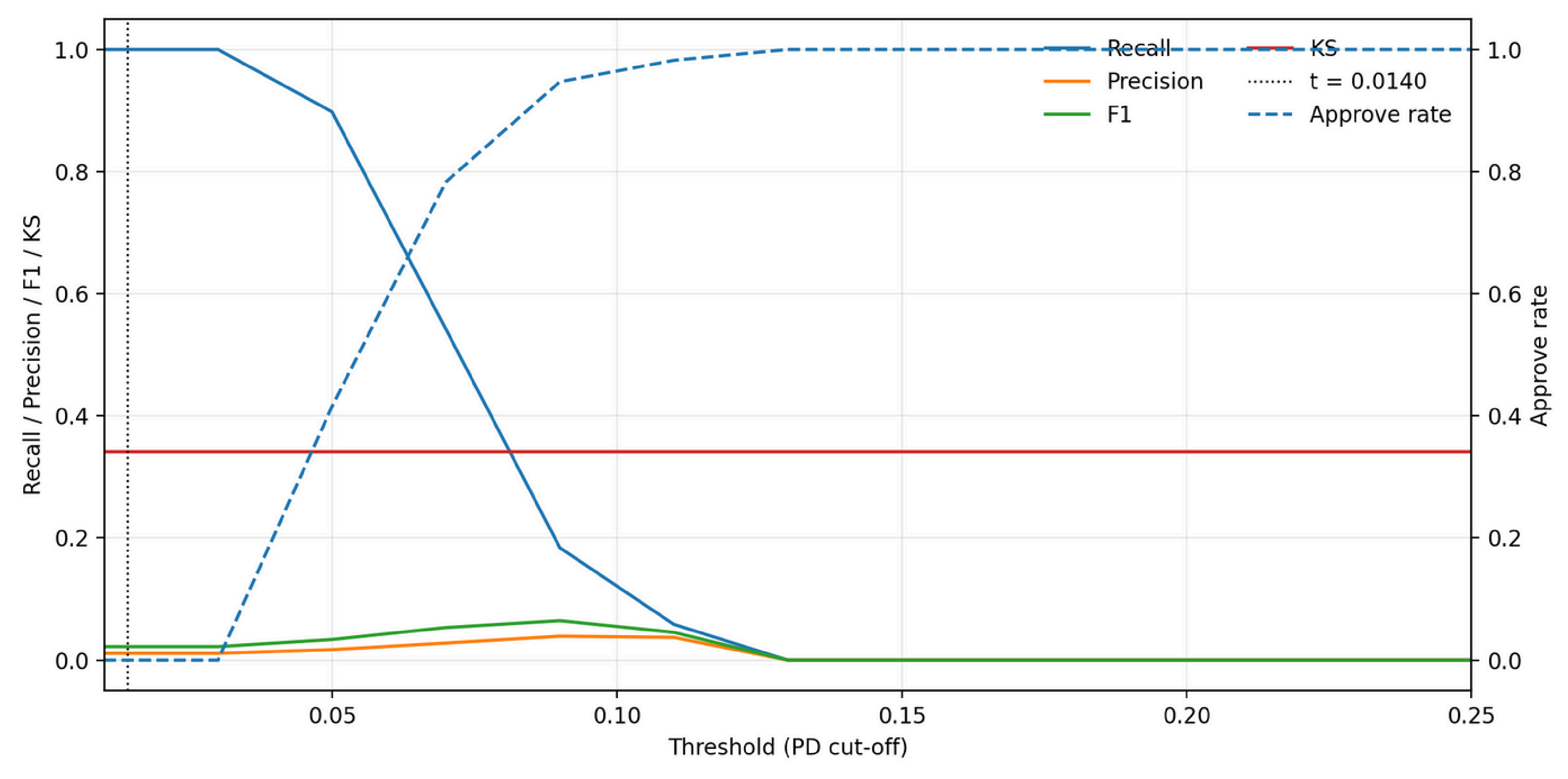


DISCRIMINATION (HOW WELL WE RANK)

KS ≈ 0.415; AUC ≈ 0.754 (stable across test).
Strong separation for underwriting cut-offs and pricing tiers.



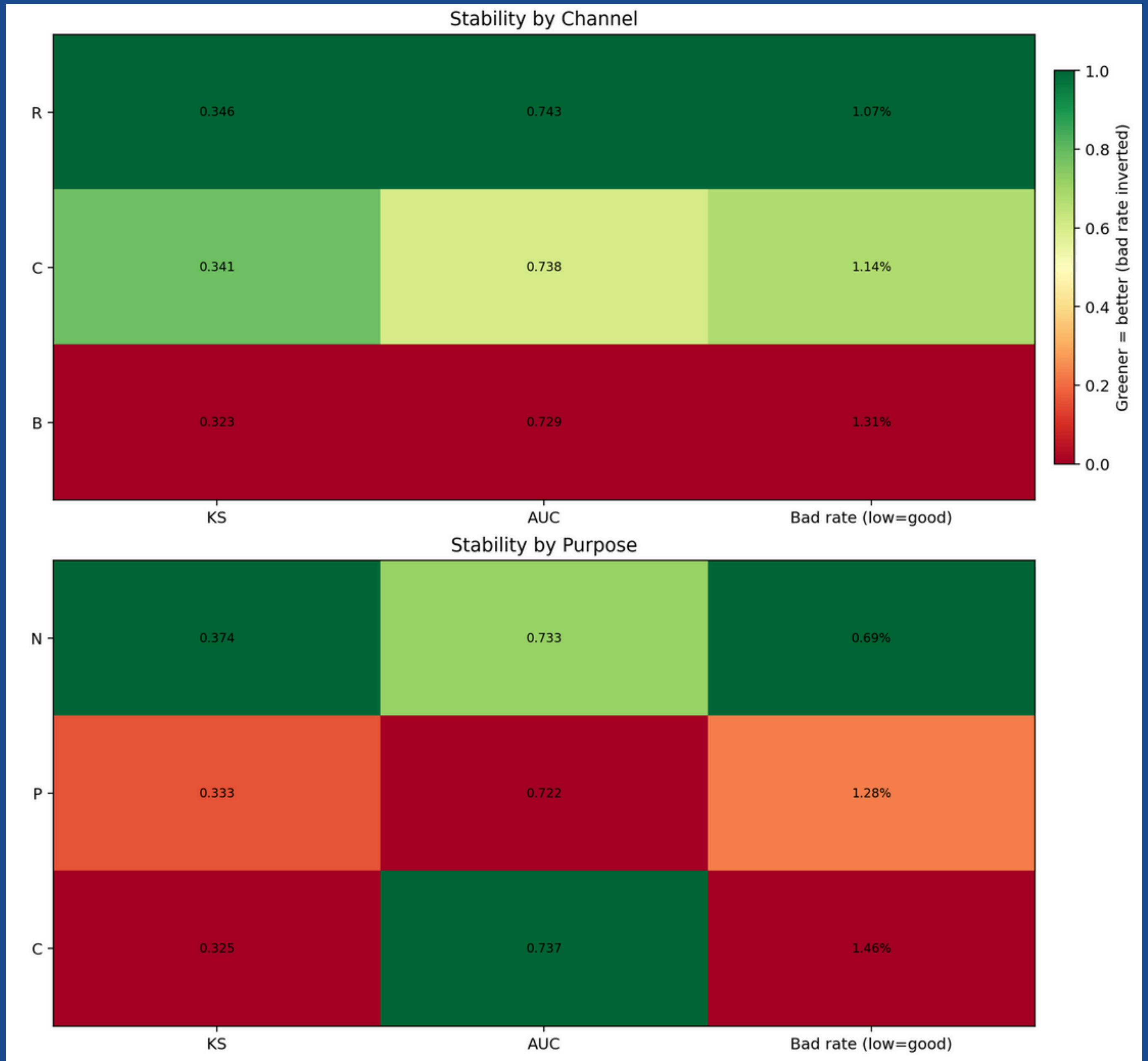
POLICY CURVE & CHOSEN THRESHOLD



- Swept thresholds across KS, precision, recall and approval rate.
- Chosen balanced policy at PD ≈ 0.010 to target ~20% approvals.
- Threshold stored in policy_choice.json for reproducibility.

SLICE STABILITY (STATE, CHANNEL, PURPOSE, VINTAGE)

1. **Discrimination:** is steady across slices. KS ≈ 0.32–0.37, AUC ≈ 0.72–0.74 for all groups → the model ranks well regardless of channel or purpose.
2. **Channel (risk levels):**
   - **2.1 Retail** ≈ 1.07% bad-rate (best).
   - **2.2 Correspondent** ≈ 1.14% (slightly worse).
   - **2.3 Broker** ≈ 1.31% (highest).
3. **Purpose (risk levels):**
   - **3.1 No-cash-out refi** ≈ 0.69% (safest).
   - **3.2 Purchase** ≈ 1.26% (mid).
   - **3.3 Cash-out refi** ≈ 1.46% (riskiest).



MONITORING & MRM

- **AUC (test):** Target ≥ 0.74, Alert if ↓ ≥ 0.02 m/m → 0.759 ✅
- **KS @ t:** Target ≥ 0.38, Alert if ↓ ≥ 0.03 m/m → 0.415 ✅
- **Brier (calibrated)**: Target ≤ 0.0125, Alert if ↑ ≥ 10% m/m → 0.0110 ✅
- **Bad rate (recent):** Guardrail ≤ 2.30%, Alert if ↑ ≥ 40 bps m/m → 1.99% ✅
- **Approval rate (recent):** Show the current value and a guardrail band (e.g., 18–24%).

DEMO ON STREAMLIT

Network URL: http://192.168.1.154:8501
Local Network: http://localhost:8501

CHALLENGES & LIMITATIONS

Problems encountered

- **Pathing/nesting:** streamlit_app\streamlit_app\... caused “Could not open” until we moved the model and switched to pathlib.
- **Format & features:** App expected a LightGBM text booster; we exported it and saved feature_order.json to lock inference order.
- **Heavy data wrangling:** Multi-million-row Freddie files → slow/oom until we standardized schemas, downcasted dtypes, and used chunked reads.

Project limitations

- **Data scope/bias:** Trained on Freddie-Mac 2017–2024 with approval-only outcomes → limited generalization + selection bias.
- **Ops maturity:** Batch scoring only; no live API, automated monitoring/retraining, or formal fairness/MRM governance yet.

Ironhack

THANK YOU FOR LISTENING!

María Martínez Gil