Cedar Kestrel

SYNTHETIC EXAMPLE / NOT A CLIENT RESULT

Independent model review

A concise example of the questions, checks and limits that make an analysis useful. Every observation in this example was generated by code.

800	0.771	0.192
SIMULATED OBSERVATIONS	AUC	BRIER SCORE

1. The question

Does the prediction output separate the two simulated outcomes, and how well do its probabilities match the observed event frequency?

2. What was checked

800 rows; binary outcomes; probabilities between 0 and 1; discrimination; probability error; threshold counts; and an explicitly reproducible data export. The seed is 2026.

3. Findings, with interpretation

The AUC is 0.771. This describes how often a randomly chosen event receives a higher score than a randomly chosen non-event; it is not a probability that an individual prediction is correct.

The mean squared probability error (Brier score) is 0.192. The mean predicted probability is 0.459, compared with an event frequency of 0.370. This gap raises a calibration question; the two averages alone do not establish full calibration.

At the illustrative threshold of 0.50: 198 true positives, 145 false positives, 359 true negatives and 98 false negatives. No threshold is recommended for real-world decisions.

Cedar Kestrel

SYNTHETIC EXAMPLE / NOT A CLIENT RESULT

Limits are part of the result

4. Threshold counts

Observed outcome	Predicted event	Predicted non-event
Event	198	98
Non-event	145	359

5. What this example cannot establish

The dataset is artificial. It cannot establish clinical utility, fairness in a real population, external validity, regulatory compliance, safety, or the quality of a customer model. No patient, customer or proprietary research data was used.

A real review would also depend on the study question, lawful access to the relevant data, participant-level splitting, the training history, missing-data handling, calibration by subgroup, and a fit-for-purpose validation design.

6. Reproduce the numbers

Source: assets/demo/synthetic_validation.csv. Metrics: assets/demo/metrics.json. Generator: tools/synthetic_demo.py. The included quality test independently recalculates AUC, Brier score and the confusion counts from the exported CSV.

The reference Brier score (0.234) uses a constant predicted probability of 0.34 for every observation. It is a descriptive synthetic reference, not an out-of-sample benchmark.

7. A real engagement

Scope, fee, rights, data handling and deliverables are agreed before work begins. This sample illustrates reporting style; it does not promise a result or constitute a commissioned assessment.