STRATUM · CASE STUDY

Procure-to-Pay Intelligence

What 1.6 million ERP events reveal about how a company really buys and pays.

Process mining · Compliance analytics · Causal inference · Machine learning

THE BRIEF

A €997M order book and four hard questions

The process owner has compliance questions, and the finance team needs to plan. The answers sit in **1,595,923 ERP events** spread across 76,349 purchase orders from 1,975 vendors.



What really happens?

The official process and the recorded one are rarely the same.



Where do controls fail?

And how much money sits behind each failure?



When does cash go out?

Payment dates and near-term outflow, with honest uncertainty.



What actually causes delay?

Separating real causes from things that merely travel with them.

THE ANSWERS

What the analysis found

€200.2M

CONTROLS

of order value carries at least one control exception (6.7% of lines).

−0.1 days

REWORK

is the adjusted effect of order changes on payment. The naive gap of 5.1 days disappears.

3.6%

CASH

average miss on 30-day cash-out in three backtests, against 5.9% for a naive repeat.

83 days

TIMING

median time to pay when open items count. Looking only at paid items gives 77.

KNOW YOUR DATA

The field that would have broken everything

SUMMED NAIVELY

€5.7B

apparent goods and invoice value

TRUE ORDER BOOK

€997M

total order value

A name that misleads

- “Cumulative net worth” repeats the line's running value on every event, so it isn't the amount of any one document.
- A first version summed it, and the totals didn't reconcile to the order book. The check caught it.
- Value-based matching isn't possible here, so controls test document counts and exposure uses order value.

ENGINEERING

An ERP data model you can trust

Bronze

Streamed, untouched

- 730 MB XES log parsed in about 30 seconds
- Checksum verified on every run
- External XML entities disabled

Staging

dbt on DuckDB

- Typed events and order lines
- Implausible dates flagged, not dropped
- Every event numbered in process order

Marts

Star schema

- Document, line and event grains
- Vendor, user, activity and date dimensions
- Compliance, variants and process graph

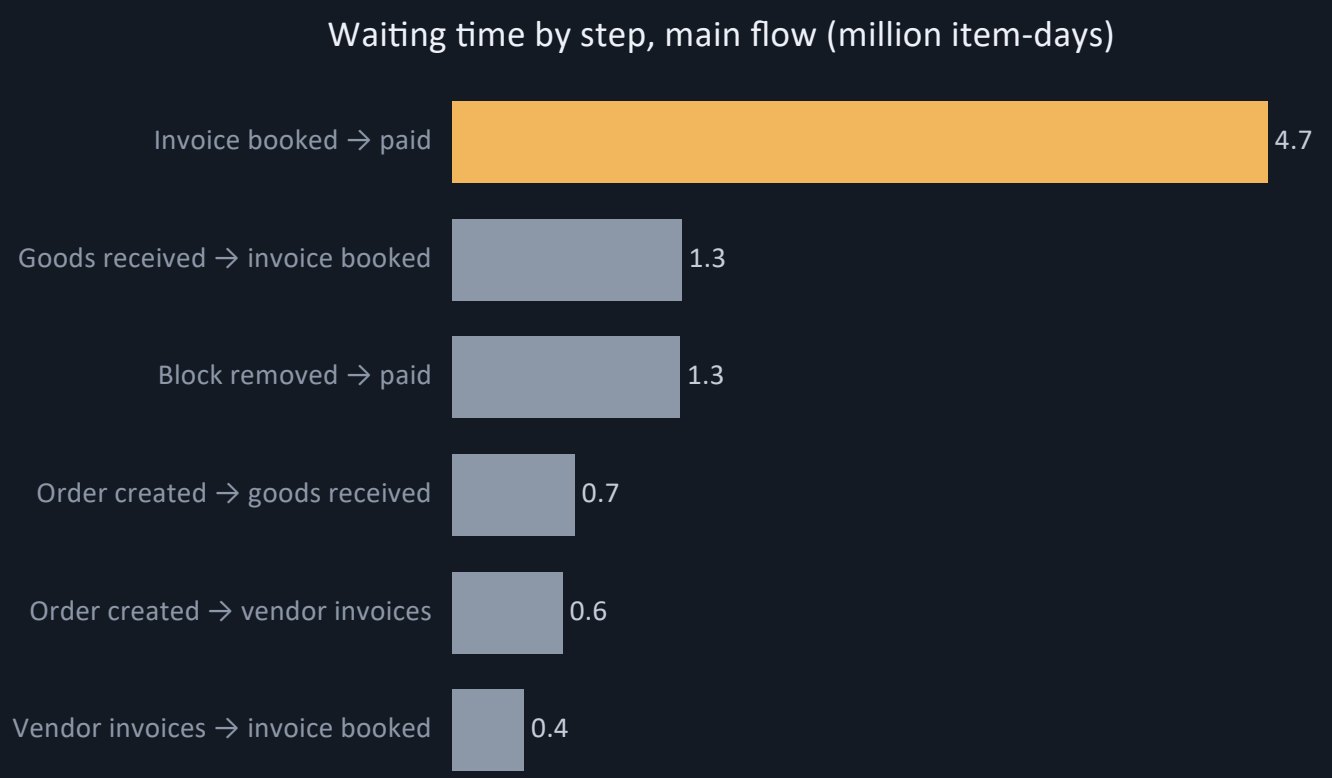
16 dbt models

31 / 31 dbt tests pass, including a rule unit test

1.6M events reconciled to source

PROCESS MINING

Most waiting happens after the invoice is booked



84%

of paid lines in the main flow ran straight through

0.96

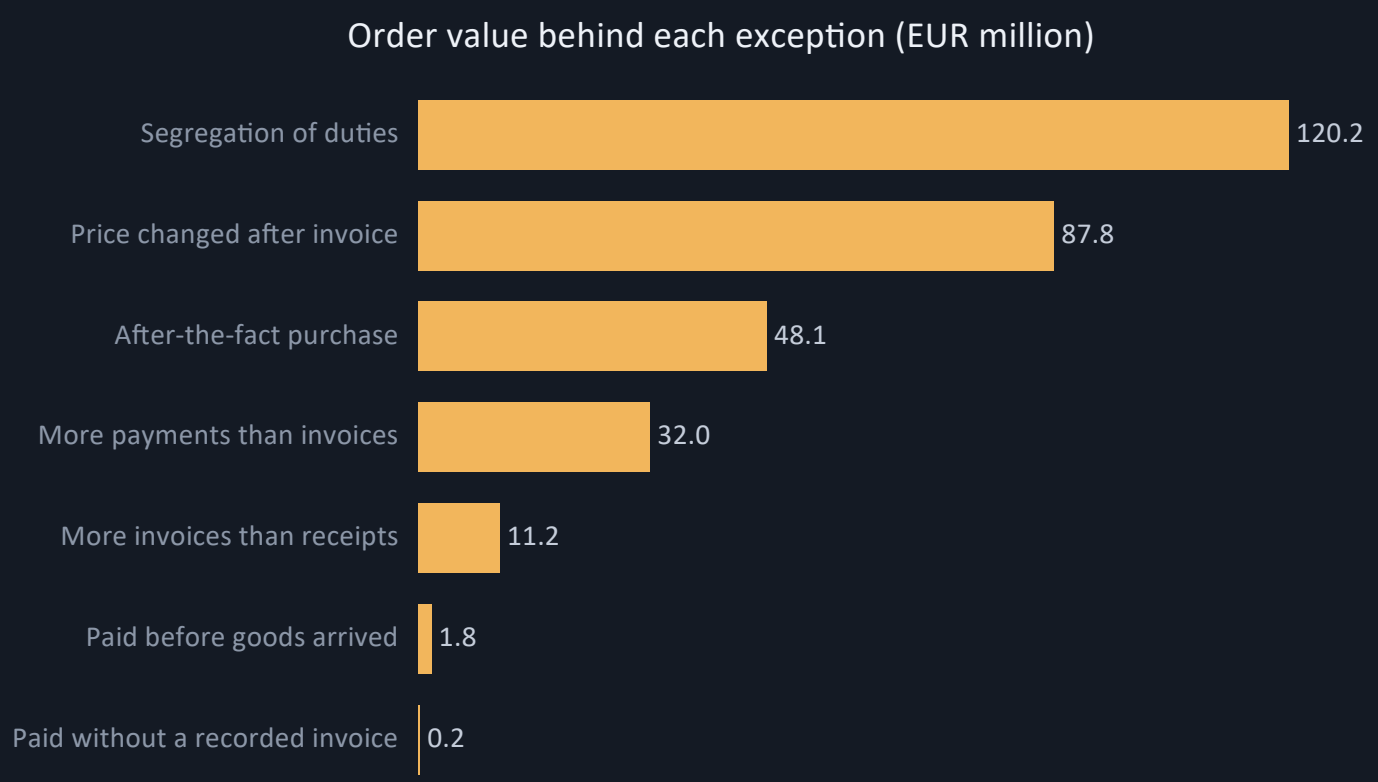
fitness of the discovered model on the main flow

0.68

on the GR-based flow: far less disciplined

COMPLIANCE

€200.2M of orders carry a control exception



7,551

lines with one person on both sides of a control

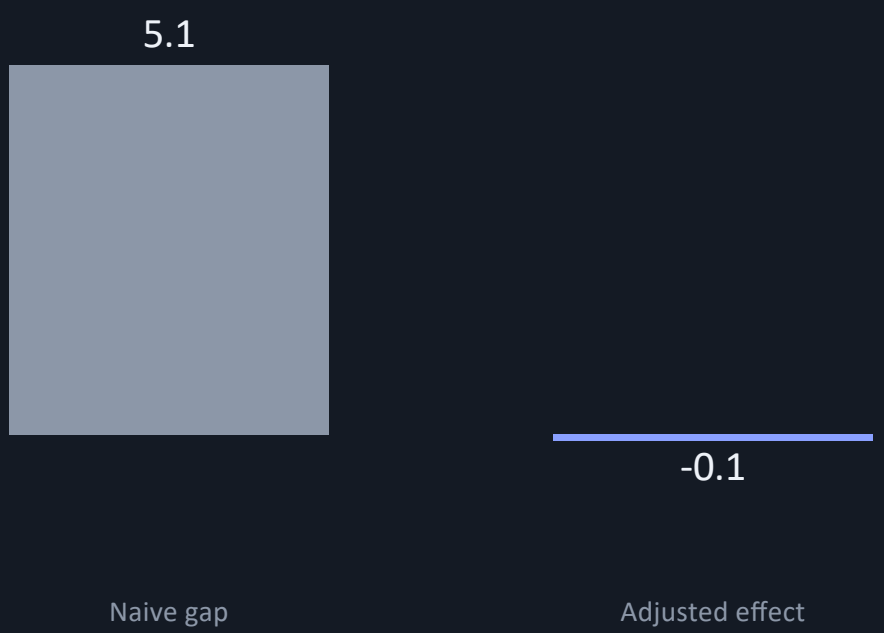
3 of 10

controls recorded no exceptions at all, and goods receivers never pay

An exception is a prompt to review, not proof of wrongdoing. Exposure is the order value of the line.

CAUSAL INFERENCE

Reworked orders look slower. Rework isn't why



Three traps, handled

Confounding: double machine learning adjusts for vendor, value and spend area.

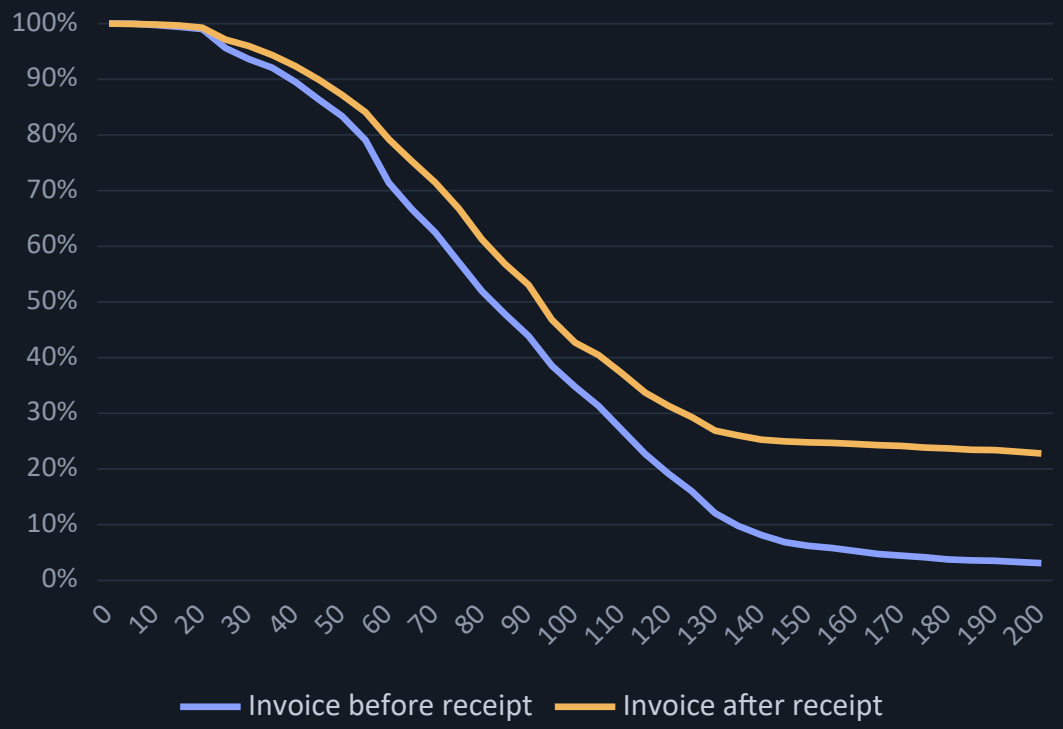
Immortal time: only orders still unpaid at day 14 count, with changes inside those 14 days.

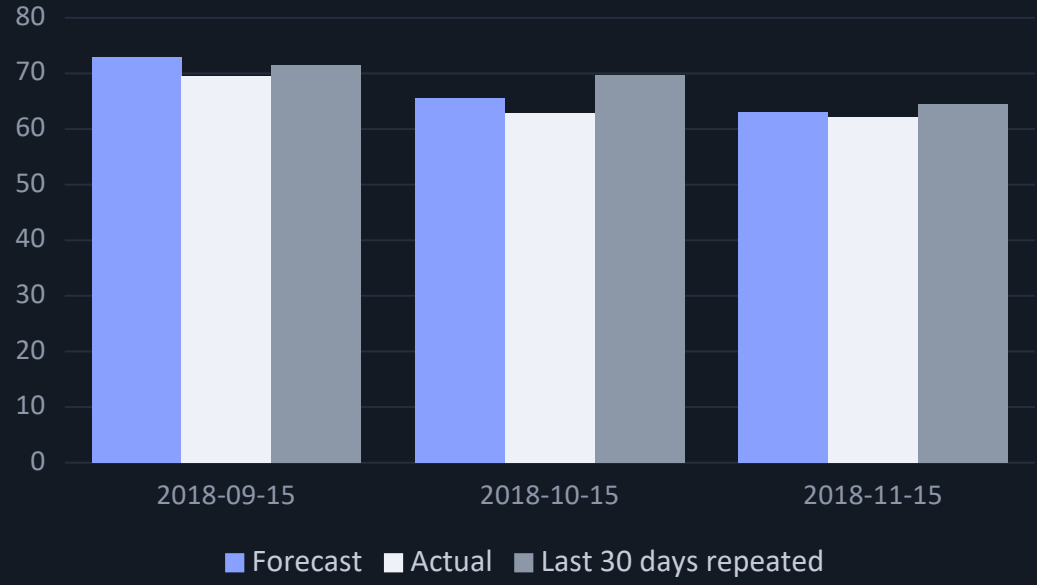
Censoring: the outcome is unpaid days in a fixed 180-day window.

Result: 95% CI -1.6 to +1.4 days on 108,383 lines. Placebo: +0.5.

SURVIVAL ANALYSIS

Forecasting cash from the open book





Next 30 days: about €67.7M expected to go out. Average backtest miss 3.6% vs 5.9% naive.

HIERARCHICAL BAYES

Finding vendors that are truly a problem

1

median lines per vendor in the raw top 20: league tables crown one-order flukes

55

median lines in the Bayesian top 20: real volume, worth a conversation

1 of 20

vendors appear on both lists, so the choice of method changes who you call

Predicting each vendor's second half	Log score (higher is better)	Weighted error
Hierarchical Bayes	**-0.224**	**0.061**
Raw vendor rate	-0.294	0.064
Overall rate	-0.240	0.077

UNCERTAINTY AND LIVE PREDICTION

Predictions you can rely on

Conformal prediction

Does a “90% range” really hold 90% of the time?

79%

coverage from the raw quantile model

92%

after conformal calibration

Coverage on the smaller flow reaches 88%, and ranges are 105 days wide: honest rather than impressive.

Deep sequence model

A PyTorch GRU reads events as they happen.

19.8 d

error on time remaining, vs 24.2 days for the baseline

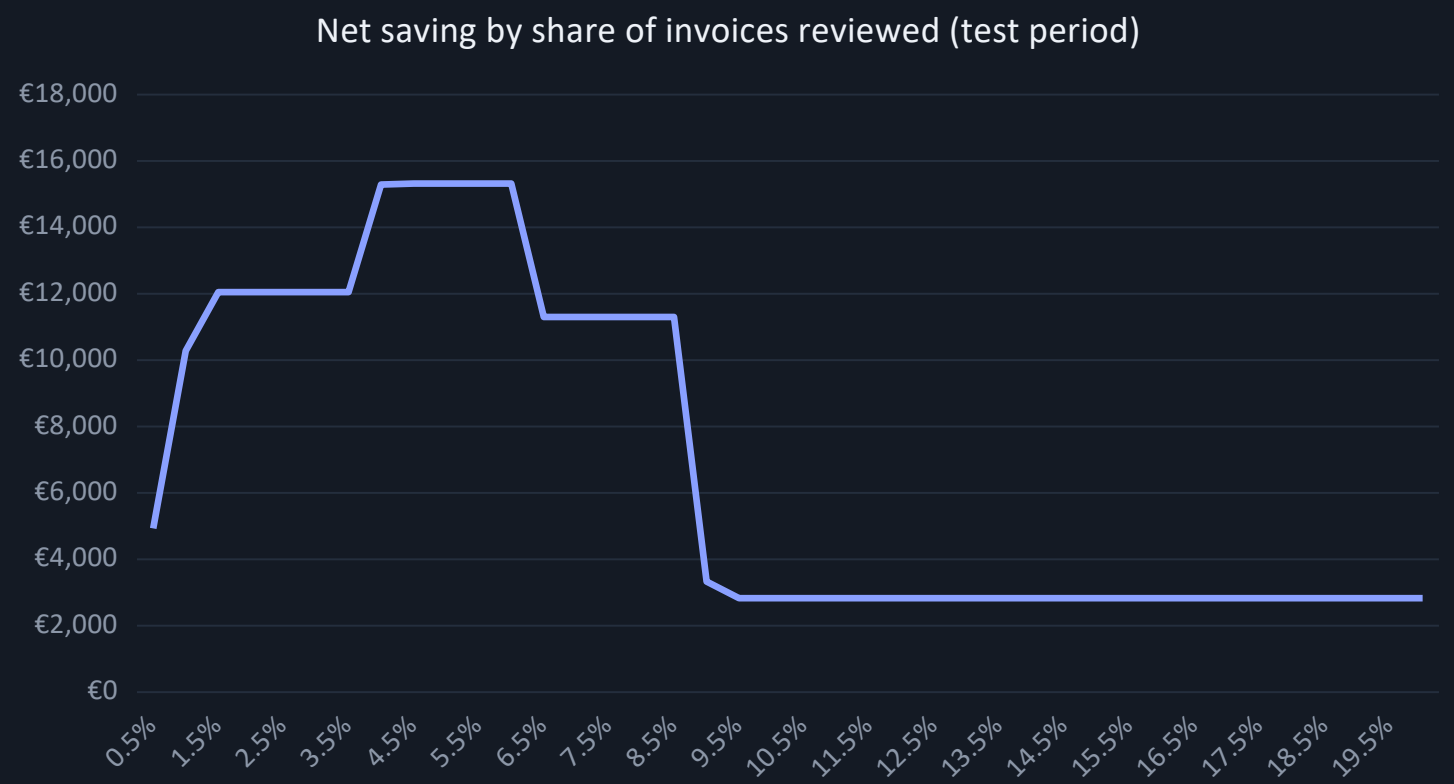
79%

next step correct, vs 76% for a Markov chain

Tested on orders from a later period, with no future information in any prediction (unit-tested).

FROM SCORE TO DECISION

A good model isn't automatically a business case



4.2%

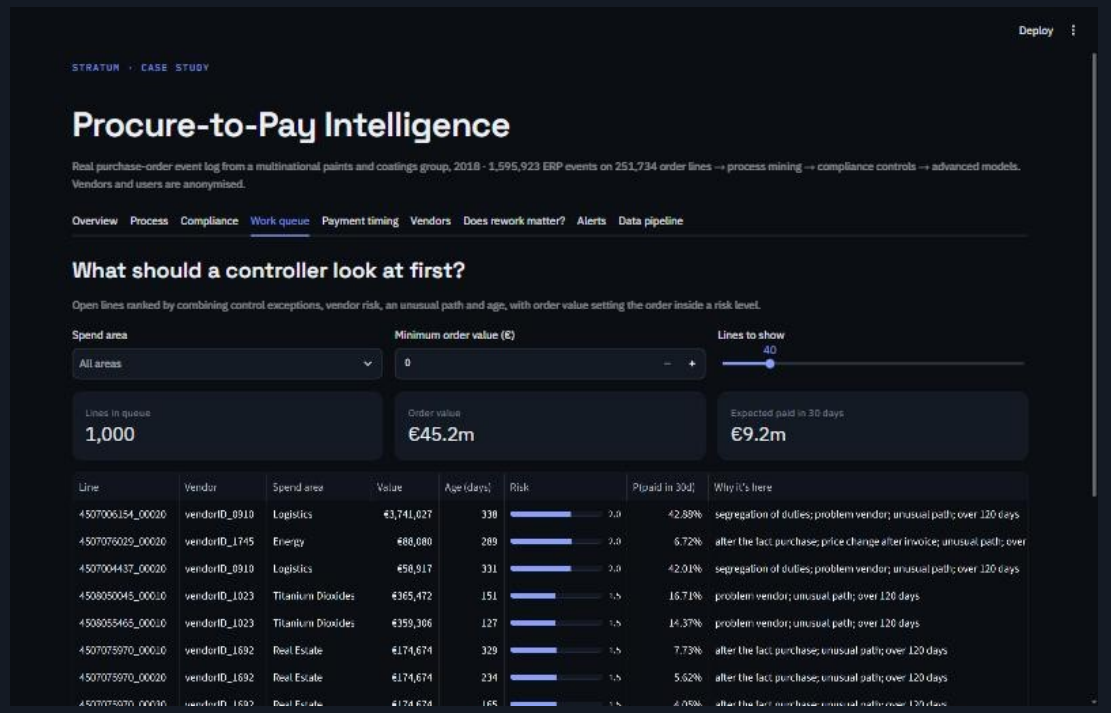
of invoices is the best share to review, not all of them

34%

of future corrections sit in that small slice

ROC-AUC 0.75. Assumes €10 a review, €180 a correction and half prevented, which are inputs rather than findings. At €25 a review the saving falls to about €5k.

A work queue, not just a report



Work queue: open lines ranked, with the reason each one is there



Causal view: naive gap against the adjusted effect

Nine tabs: Overview · Process · Compliance · Work queue · Payment timing · Vendors · Rework · Alerts · Data pipeline

CONCLUSIONS

What I'd tell the CFO

Review the leaking controls

7,551 lines had one person on both sides of a control. Three other controls held fully: find out why, and copy it.

Look at the wait after booking

That step holds the most waiting time. Check it against payment terms before changing anything.

Don't chase order changes

Reducing rework is unlikely to speed payment. Spend that effort elsewhere, with Packaging the one area to look at.

Run cash from the open book

A forecast built on line-level timing gives finance a 30-day view that's better than repeating last month.

HOW IT'S BUILT

What makes this work trustworthy

Tested forward in time

Every model is scored on later data, against something simple

Reported what didn't work

Limits and weak spots are stated, not hidden

Reproducible

One command from public download to dashboard, with identical results

Tested

31 dbt tests plus 12 Python tests, including recovering known effects

TOOLKIT

Python · SQL · dbt · DuckDB · pm4py · lifelines · NumPyro · PyTorch · scikit-learn · SciPy · Plotly · Streamlit · pytest · GitHub Actions

STRATUM

What's your ERP not telling you?

I turn the transaction history your systems already hold into clear answers on process, control and cash.

stratumdata.org

stratumfreelance@gmail.com