

PHAN HOAI THU (KATHERINE)

Commercial Data Analyst

0817 375 930 · phanhoaitu114@gmail.com · Ho Chi Minh City
kathuatwork.com · linkedin.com/in/phanhoaitu114 · github.com/kath-pahotu

I build the reporting layer commercial teams run on — plan versus actual, price-volume-mix, margin, and channel efficiency — and the controls that prove it reconciles. Three and a half years in e-commerce, marketplace and agency analytics. I would rather hand over a number with its limits stated than a cleaner one without them.

Evidence — *TheLook*: 35 checks across three gates, a reconciliation statement, and a forecast rebuilt after its own audit found two defects · *Dunnhumby*: 147 category elasticities, and a campaign result that reversed once the comparison groups were matched · *BE Group*: real operating data, and a scope labelled "All Vietnam" that was two cities

Commercial — price elasticity · promotion effectiveness and trade spend · CAC, CIR and payback · cohort retention

Engineering — SQL (T-SQL, DuckDB, BigQuery, PostgreSQL) · Python (pandas, scikit-learn, statsmodels) · star-schema modelling · automated data contracts

Measurement — A/B testing and randomisation audits · difference-in-differences · fixed-effects panel models · matching · rolling-origin backtesting

Reporting — Power BI (DAX, Power Query) · Excel (Power Pivot) · GA4

EXPERIENCE

Independent Data & Commercial Analyst (*Freelance*) · Jan 2026 – Present · Ho Chi Minh City

- Analytics engagements for e-commerce and retail clients: channel efficiency, funnel diagnosis and reporting consolidation.

E-commerce Analyst · Laptop SGN (*Freelance*) · Jul 2025 – Dec 2025

- Built the definition layer the reporting ran on — **five documented tables** across sessions, users, funnel events, transactions and segments, every column typed and every derived rule written, so conversion and revenue carried one definition in every report.
- Analysed 100,000+ session, user and transaction records to find where revenue was leaking: **21.5% of product-page visitors reached the cart, against 63.5% of checkout visitors completing** — mid-funnel, not checkout.
- Built the customer view: recency, frequency and order-value segments with a named action each, cuts by channel, device, age and product category, and a derived cross-category browsing measure separating focused buyers from comparison shoppers — which set the repeat-purchase baseline at **36.4% second purchase, 17.8% third**.

Digital Marketing Analyst · Golden Owl Solutions · Apr 2024 – Sep 2025

- Built the lead-quality reporting across **15 client businesses serving 250,000+ monthly users**: roughly 3,100 leads a year by source, country, owner and status, of which **close to a quarter were spam or unqualified** — every per-lead metric had been running on an inflated denominator.
- Joined GA4 behavioural data to HubSpot sales records to trace a qualified lead from first visit to contact, and identified the step where that path broke for everyone else.
- Delivered **70+ analyses** across GA4, Looker Studio and Search Console, including a rate-versus-mix decomposition showing one account's falling click-through rate came from impressions widening into lower-intent queries, not from declining performance.

Digital Marketing Executive · Canawan Global · Sep 2022 – Apr 2024

- Owned the data loop for a **1,000+ SKU** Etsy and Amazon catalogue: weekly product, acquisition and conversion reporting in GA4 and Looker Studio showing which products carried the revenue, and automated listing publication with n8n and Google Sheets.
-

SELECTED PROJECTS

TheLook — **commercial reporting layer** · DuckDB, SQL, Python, Power BI

github.com/kath-pahotu/the-look-ecommerce-data-analysis

- Reproducible ELT from 7 source files and 2.42M events; rebuilt 680,862 sessions, modelled a conformed star schema, and reconciled a second source — regional plans in five currencies — onto one country-month grain. Rebuilds from source with a single command.
- **35 automated checks across three gates**, with a stated rule for what blocks a build and what quarantines, plus a reconciliation statement a reviewer can audit: 292 plan rows submitted, 5 quarantined, 287 accepted, money tying at every stage.
- Audited the forecasting module before presenting it — seven checks specified in advance, of which **two found a defect in my own code**: models were scored one month ahead while the forecast ran eight months ahead. Rebuilt to 192 rolling-origin folds; 8.0% MAPE.

Dunnhumby — price elasticity and promotion effectiveness · Python, DuckDB, statsmodels

github.com/kath-pahotu/dunnhumby-pricing-promotion

- Own-price elasticity for **147 grocery categories** across 2.6M transactions and 36.8M promotion records, using two-way fixed effects with robust errors. Price, display and mailer modelled in one equation so each is purged of the others; 134 of 147 significant.
- Found stockpiled goods are price-elastic while fresh meat is inelastic on price but moves +161% on the mailer — two facts implying opposite promotional strategies.
- Evaluated 17 coupon campaigns by difference-in-differences. The naive estimate said coupons *reduced* spend and **the parallel-trends test endorsed it on 15 of 17**. Diagnosed the cause — the comparison group spent 7× less before the campaign — and rebuilt with matching on pre-period spend. Every sign flipped: 13 of 17 significantly positive, median +23.7%.

BE Group — marketing channel economics · Python, DuckDB · real operating data, indexed

github.com/kath-pahotu/be-group-channel-economics

- Rebuilt three operating workbooks into a reproducible pipeline: 54 weeks × 8 acquisition channels with spend, CAC, CPA and cost-to-income ratio, brand-level take rate, and a promotion margin model. Every published figure indexed; no real value disclosed.
- Reconciliation found a scope labelled "All Vietnam" is **exactly two cities, to 100.000% in every month** — a label that would have overstated coverage wherever it appeared. Also showed a commission increase created ~5 points of margin headroom, of which promotion consumed 3.4 within four days.

Rokt — holdout experiment audit · Python, statistics

github.com/kath-pahotu/au_rokt

- Audited 852,197 records, removed 36,893 duplicates and found 26,426 users in both arms before estimating any effect. Measured +13.9% uplift (95% CI +9.1% to +19.0%); a user-cohort sensitivity analysis returned +5.5% (p = 0.016).
- Recommended a **conditional** rollout pending confirmation of the randomisation unit, rather than treating statistical significance as sufficient evidence.

ZaloPay — promotion budget concentration · T-SQL, Python, Power BI

github.com/kath-pahotu/zalopay-promo-abuse-detection

- T-SQL -> pandas -> Power BI workflow over 254K transactions and 90.6K users; six behavioural signals, with the SQL and Python outputs reconciled independently against each other.
- Found 4.26% of scored users associated with 26.94% of credited promotion cost. Scoped the recommendation to a shadow-mode review queue rather than auto-blocking: the data carries no fraud labels, so it cannot support a fraud claim.

EDUCATION

Bachelor of Business Administration · University of Economics Ho Chi Minh City · 2018 – 2022

TheHau Data Analytics Mentoring Programme · selected 15-person cohort · Mar – Jul 2025

IELTS Overall Band 7.5 · 2022