

Pricing Demo

by Alfonso Cifuentes · a publishing-pricing analytics case study

This project grew out of a technical assessment I completed during a recruitment process for a data analyst role at an international educational publisher. The task: structure a product catalogue, compare genuinely equivalent items and explain which price differences deserved investigation. I rebuilt it with entirely fictional titles, series and prices so the approach can be shared.

362 simulated products, 12 series and one question: are the prices different, or are the products different?

INDEPENDENT RECONSTRUCTION · FICTIONAL DATA · NO EMPLOYER CATALOGUE OR PRICES

A price list is a *list of assumptions.*

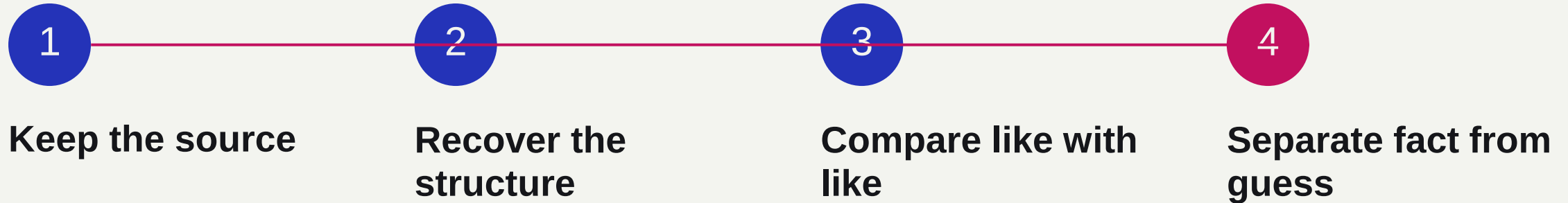
The pooled mean of this fictional catalogue is €44.04, a number that describes no product: workbooks average €24.72, student books €60.51.

362

fictional products,
12 invented series.

5 raw columns → 19 enriched · seeded generator (20261011) · same seed, same bytes · no real publisher, no real prices.

Hold the product still. *Then look at the price.*



A title is *not a schema.*

A rule-based parser that reads titles only matched the reference labels on level, pack, answer key, book part, audio and digital access for all 362 rows. On product type it abstained twice (“TB”, “Gleebook”) instead of guessing. The rules were written for these titles, so this checks the rules, not real-world accuracy. Edition and access duration never appear in a title: they can only be sourced.

6/72

Unknown
is not “no”.

Only 6 of 72 digital workbooks state an access duration (18 months). A blank means unknown, not unlimited.

2 records filed

under the wrong name.

Row	Title	Series field
20	Mosaic English B1 Premium Pack with Digital	Glasshouse English
155	Glasshouse English C1 Study Pack with Digital	Kindred Horizons

Flag and re-file in a reporting view. Never overwrite the source.

Gleebook

is not an eBook.

NAIVE SUBSTRING RULE

"ebook" in title

1 false eBook: digital access granted

WORD-BOUNDARY RULE

`\be-?book\b`

0 matches: printed stays printed

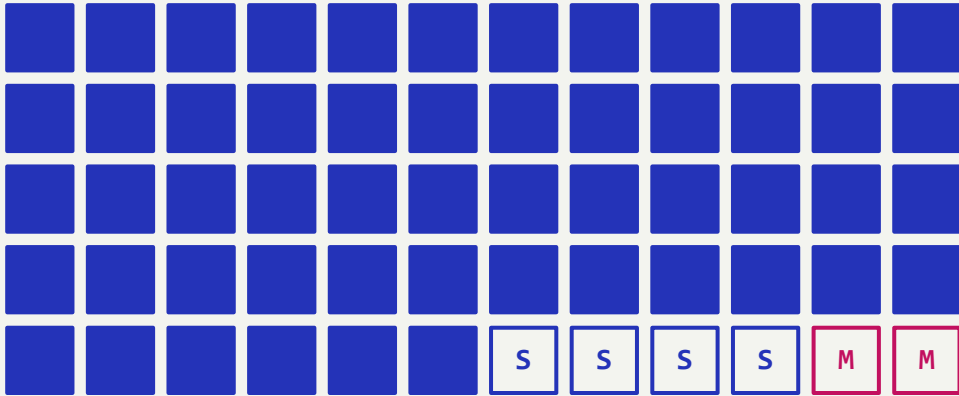
Ambiguity is *operational debt.*

- Ⓜ Edition aliases: “Second Edition” ×3 and “2nd Edition” ×3. Normalise in reporting only.
- Ⓜ Pack wording: “+ Online Access” ×6 vs “with Digital” ×24. Merging is a business decision.
- Ⓜ Taxonomy drift: row 61, a Primary book filed under a Secondary sub-segment.
- Ⓜ “TB” at €108.49: the only price above the upper IQR fence (about €100). Type unresolved.
- Ⓜ Sparse access duration: 66 digital workbooks say nothing.

60

comparable groups
out of 68.

Ten attributes held constant: series, type, role, part, pack, answer key, both digital rights, resources, edition. Only the level changes. 8 groups exist at a single level and are set aside.



54/60

hold one price
at every level.

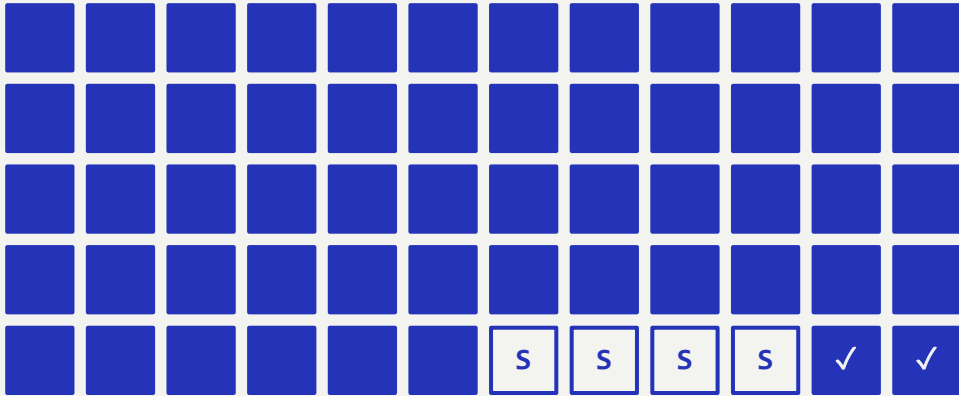
90.0% of comparable groups. Six exceptions remain.

S = Starter priced lower • M = misfiled record

Four Starter prices, *two filing errors.*

Kind	Series (as recorded) · configuration	Odd value	Others	Gap
Starter pricing	Mosaic English · Student Book	Starter €33.49	€41.49	–€8.00
Starter pricing	Mosaic English · Workbook · with Audio	Starter €16.49	€20.49	–€4.00
Starter pricing	Silverline Skills · Student Book	Starter €39.49	€47.49	–€8.00
Starter pricing	Silverline Skills · Workbook · with Audio	Starter €24.49	€28.49	–€4.00
Filing error	Glasshouse English · Student Book · Premium Pack (row 20: Mosaic English)	B1 €69.49	€86.49	–€17.00
Filing error	Kindred Horizons · Student Book · Study Pack (row 155: Glasshouse English)	C1 €67.49	€69.49	–€2.00

Four Starter levels priced €4–€8 below their other levels: plausible strategy, unconfirmed. Two groups contain another series' product, so their “exception” is a filing error.



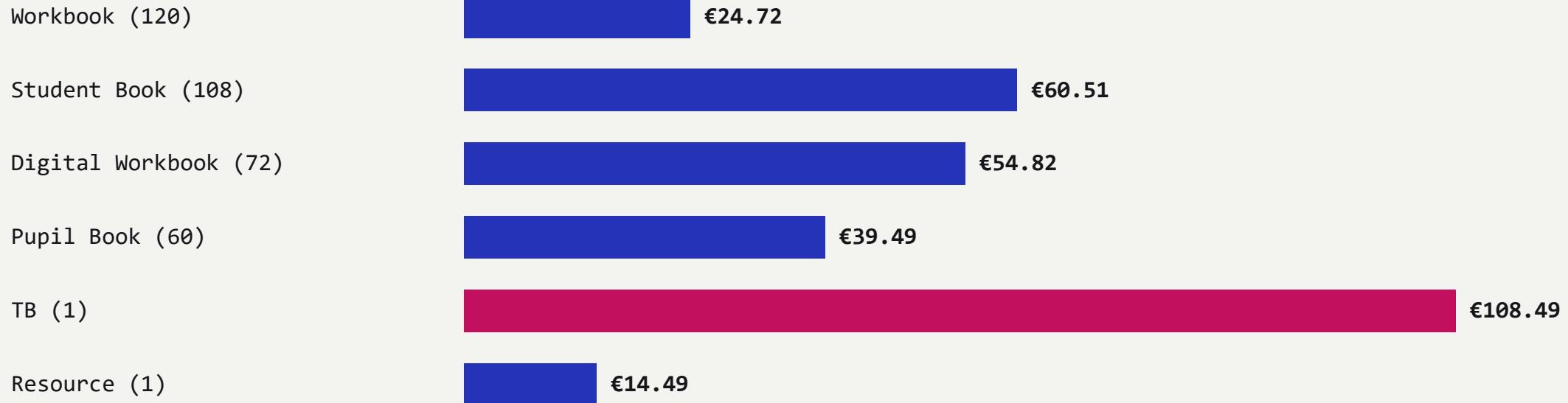
56/60

after re-filing
2 records.

93.3%, with the source file untouched. Two “price exceptions” were data errors.

S = Starter priced lower • M = misfiled record

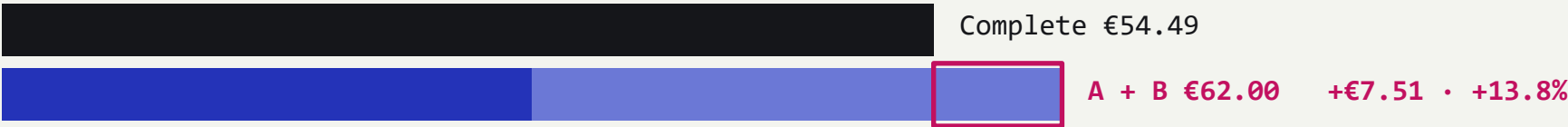
Different products, *different baselines.*



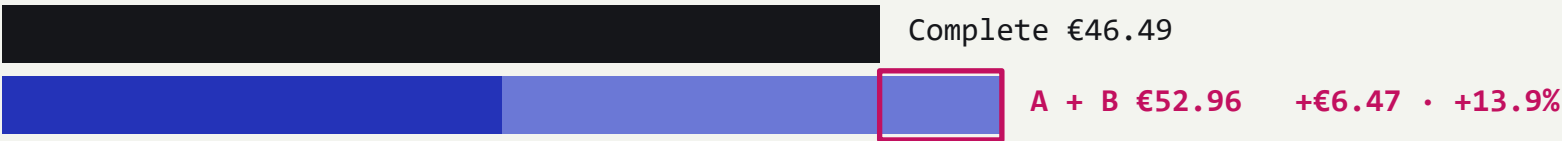
Pooled mean €44.04: context only, never a comparison baseline.

Two parts cost more *than one book.*

LITTLE LANTERNS • SAME LEVEL • STRICT AT 4 OF 6 LEVELS • 2 INDICATIVE (EDITION LABELS DIFFER)



WILDFLOWER WORDS • SAME LEVEL • STRICT AT 6 OF 6 LEVELS



Answer keys: 24 of 24 with-key / without-key workbook pairs share one price when matched on series, level and edition.

What a list price *can't tell you.*

NOT IN THE DATA

~~SALES · COSTS · MARGINS · DISCOUNTS · ELASTICITY~~

Q1

Is the Starter discount in Mosaic English and Silverline Skills deliberate?

Q2

Who owns series metadata, and how were rows 20 and 155 misfiled?

Q3

Is the premium of about 14% for buying a book in two parts intended?

Q4

What is “TB”, and does €108.49 belong in this catalogue?

Pricing Demo

Keep the source. Recover the structure. Compare like with like. Say what you don't know.

Files: synthetic dataset (raw + enriched CSV), analysis JSON, Excel workbook with live formulas, this deck, and four narrated videos (EN/ES, long and short). Generator, analysis and tests are reproducible from seed 20261011.

[PRICINGDEMO-ALFONSOCIFUENTES.VERCEL.APP](#) • [GITHUB.COM/ALFONSOCIFUENTES](#)

Inspired by a recruitment technical assessment; all series, products and prices are invented. Not affiliated with any publisher or employer.