

Pricing Demo

por Alfonso Cifuentes · caso de estudio de análisis de precios editoriales

Este proyecto nace de una prueba técnica que realicé en un proceso de selección para un puesto de análisis de datos en una editorial educativa internacional. El reto: organizar un catálogo, comparar precios de productos realmente equivalentes y explicar qué diferencias merecían investigarse. Lo reconstruí con títulos, colecciones y precios ficticios para poder compartirlo.

362 productos simulados, 12 colecciones y una pregunta: ¿hay diferencias de precio o estamos comparando cosas distintas?

ADAPTACIÓN INDEPENDIENTE · DATOS FICTICIOS · NINGÚN PRODUCTO NI PRECIO DE LA EMPRESA DE SELECCIÓN

Una lista de precios es *una lista de supuestos.*

La media global de este catálogo ficticio es 44,04 €, una cifra que no describe ningún producto: los cuadernos promedian 24,72 € y los libros del alumno 60,51 €.

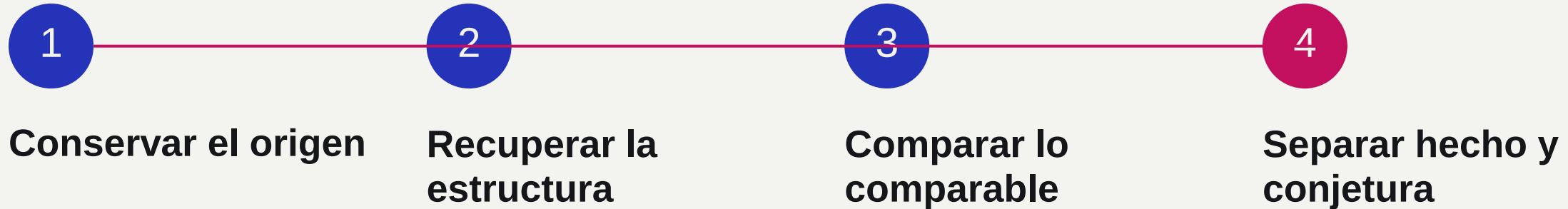
362

productos ficticios,
12 series inventadas.

5 columnas de origen → 19 enriquecidas · generador con semilla (20261011) · misma semilla,
mismos bytes · sin editorial ni precios reales.

Fija el producto.

Después mira el precio.



Un título

no es un esquema.

Un parser de reglas que solo lee títulos coincidió con las etiquetas de referencia en nivel, pack, solucionario, parte del libro, audio y acceso digital en las 362 filas. En tipo de producto se abstuvo dos veces («TB», «Gleebook») en lugar de adivinar. Las reglas se escribieron para estos títulos: esto comprueba las reglas, no la precisión con datos reales. La edición y la duración del acceso nunca aparecen en el título: solo pueden venir de metadatos.

6/72

Desconocido
no es «no».

Solo 6 de 72 cuadernos digitales indican duración de acceso (18 meses). Un vacío significa desconocido, no ilimitado.

2 registros archivados

con el nombre equivocado.

Fila	Título	Campo serie
20	Mosaic English B1 Premium Pack with Digital	Glasshouse English
155	Glasshouse English C1 Study Pack with Digital	Kindred Horizons

Señalar y reubicar en una vista de reporting. Nunca sobrescribir el origen.

Gleebook

no es un eBook.

REGLA INGENUA POR SUBCADENA

"ebook" in title

1 falso eBook: se concede acceso digital

REGLA CON LÍMITES DE PALABRA

\be-?book\b

0 coincidencias: lo impreso sigue impreso

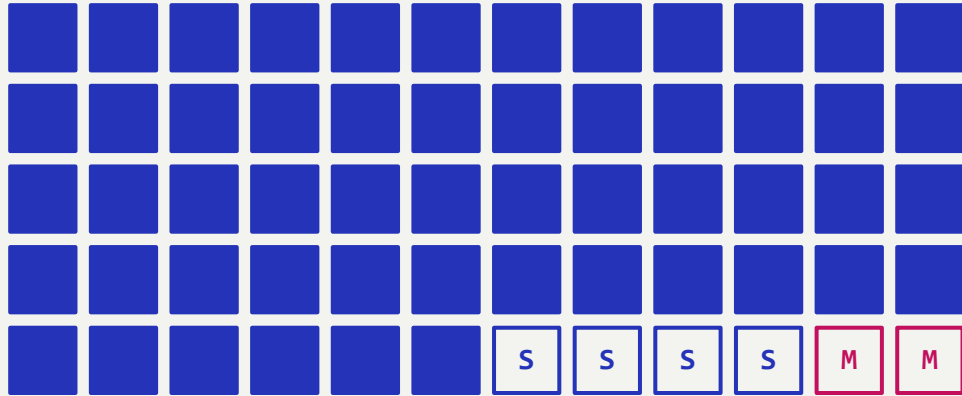
La ambigüedad es *deuda operativa*.

- Ⓜ Alias de edición: «Second Edition» ×3 y «2nd Edition» ×3. Normalizar solo en reporting.
- Ⓜ Nombre del pack: «+ Online Access» ×6 frente a «with Digital» ×24. Fusionarlos es una decisión de negocio.
- Ⓜ Deriva de taxonomía: fila 61, un libro de Primaria en un subsegmento de Secundaria.
- Ⓜ «TB» a 108,49 €: único precio sobre la valla IQR superior (unos 100 €). Tipo sin resolver.
- Ⓜ Duración de acceso escasa: 66 cuadernos digitales no indican nada.

60

grupos comparables
de 68.

Diez atributos fijos: serie, tipo, rol, parte, pack, solucionario, ambos accesos digitales, recursos y edición. Solo cambia el nivel. 8 grupos tienen un único nivel y quedan fuera.



54/60

mantienen un precio
en todos los niveles.

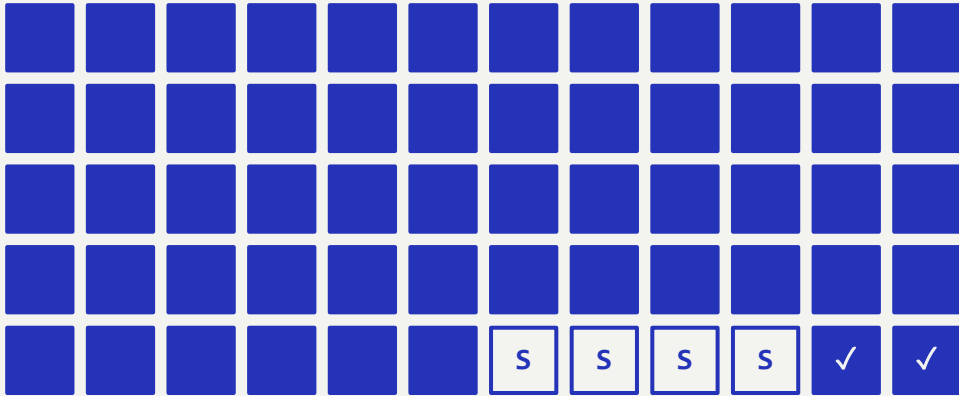
90,0 % de los grupos comparables. Quedan seis excepciones.

S = Starter más barato • M = registro mal archivado

Cuatro precios Starter, *dos errores de archivo.*

Tipo	Serie (registrada) · configuración	Valor atípico	Resto	Diferencia
Precio Starter	Mosaic English · Student Book	Starter 33,49 €	41,49 €	-8,00 €
Precio Starter	Mosaic English · Workbook · with Audio	Starter 16,49 €	20,49 €	-4,00 €
Precio Starter	Silverline Skills · Student Book	Starter 39,49 €	47,49 €	-8,00 €
Precio Starter	Silverline Skills · Workbook · with Audio	Starter 24,49 €	28,49 €	-4,00 €
Error de archivo	Glasshouse English · Student Book · Premium Pack (fila 20: Mosaic English)	B1 69,49 €	86,49 €	-17,00 €
Error de archivo	Kindred Horizons · Student Book · Study Pack (fila 155: Glasshouse English)	C1 67,49 €	69,49 €	-2,00 €

Cuatro niveles Starter entre 4 y 8 € por debajo del resto: estrategia plausible, sin confirmar. Dos grupos contienen un producto de otra serie: su «excepción» es un error de archivo.



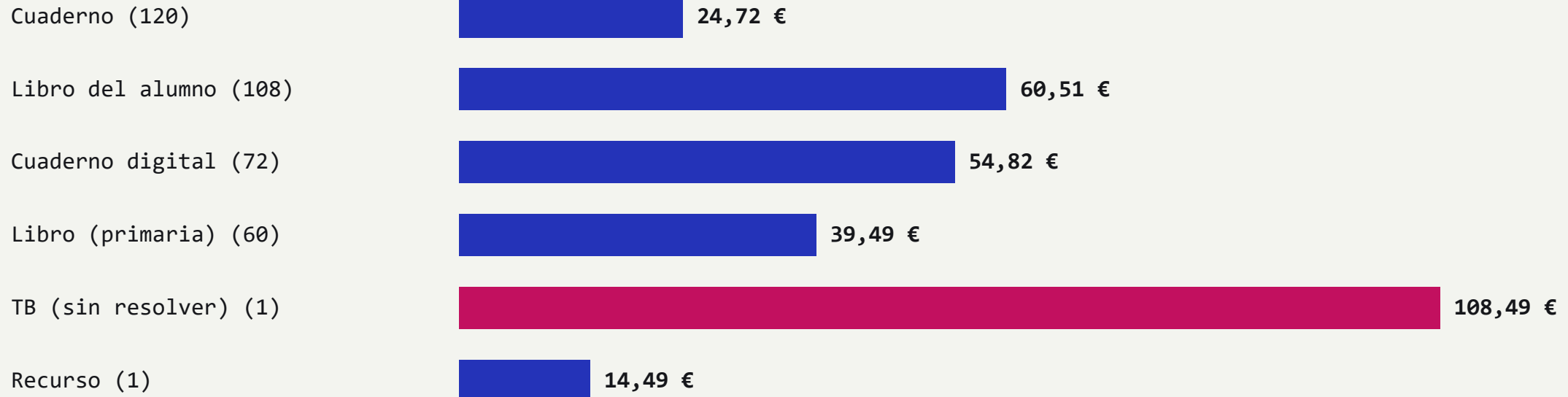
56/60

tras reubicar
2 registros.

93,3 %, sin tocar el fichero de origen. Dos «excepciones de precio» eran errores de datos.

S = Starter más barato • M = registro mal archivado

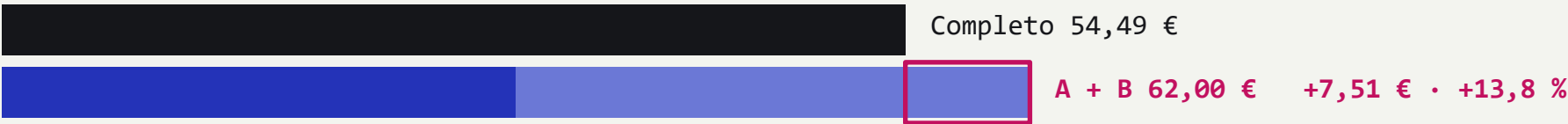
Productos distintos, *referencias distintas.*



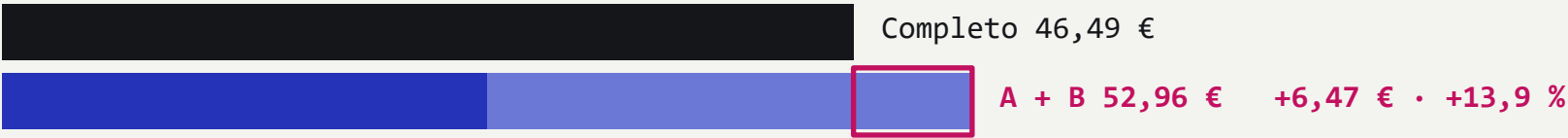
Media global 44,04 €: solo contexto, nunca referencia de comparación.

Dos partes cuestan más *que un libro.*

LITTLE LANTERNS • MISMO NIVEL • ESTRUCTA EN 4 DE 6 NIVELES • 2 ORIENTATIVAS (EDICIÓN DISTINTA)



WILDFLOWER WORDS • MISMO NIVEL • ESTRUCTA EN 6 DE 6 NIVELES



Solucionarios: 24 de 24 parejas de cuaderno con y sin soluciones comparten precio al emparejar por serie, nivel y edición.

Lo que un precio de catálogo *no puede decir.*

NO ESTÁ EN LOS DATOS

~~VENTAS · COSTES · MÁRGENES · DESCUENTOS · ELASTICIDAD~~

Q1

¿Es deliberado el descuento Starter en Mosaic English y Silverline Skills?

Q2

¿Quién gestiona los metadatos de serie y cómo se archivaron mal las filas 20 y 155?

Q3

¿Es intencionado el sobrecoste de en torno al 14 % por comprar un libro en dos partes?

Q4

¿Qué es «TB» y encaja 108,49 € en este catálogo?

Pricing Demo

*Conserva el origen. Recupera la estructura. Compara lo comparable.
Di lo que no sabes.*

Ficheros: dataset sintético (CSV de origen y enriquecido), JSON de análisis, Excel con fórmulas vivas, esta presentación y cuatro vídeos narrados (ES/EN, largo y corto). Generador, análisis y tests reproducibles con la semilla 20261011.

[PRICINGDEMO-ALFONSOCIFUENTES.VERCEL.APP](https://pricingdemo-alfonsocifuentes.vercel.app) · [GITHUB.COM/ALFONSOCIFUENTES](https://github.com/alfonsocifuentes)

Inspirado en una prueba técnica de selección; todas las series, productos y precios son inventados. Sin relación con ninguna editorial ni empleador.