

Proceso ETL y Modelado de Datos

Dataset: Ventas_Boutiques (Nazaret) — Canarias
Relación de transformaciones y ajustes aplicados —

1. Resumen ejecutivo

Se ha depurado el libro origen 00_Dataset_Ventas_Boutiques.xlsx (7 hojas, datos transaccionales de 8 boutiques en Canarias) y se ha construido un modelo de datos en esquema de galaxia con 3 dimensiones y 3 tablas de hechos. El resultado se entrega en 00_Dataset_Ventas_Boutiques_Solucion.xlsx (tablas de Excel listas para Power Pivot) junto con este documento de trazabilidad.

Alcance de este entregable: limpieza y modelado documentado (Fases 1-3 del proceso ETL). La creación del modelo Power Pivot embebido con relaciones y medidas DAX activas (Fase 4) requiere automatización COM sobre Excel de escritorio en Windows, no disponible en el entorno de ejecución de Claude; las medidas quedan documentadas en la hoja Medidas_DAX del libro para su creación manual o vía script.

2. Tablas identificadas

Tabla	Rol	Grano / clave	Filas finales
Dim_Tiendas	Dimensión	Cod_Tienda (PK, única)	8
Dim_Productos	Dimensión	Cod_Producto (PK, única)	32
Dim_Calendario	Dimensión (tabla de fechas)	Fecha (PK, única, continua)	181 (01/01/2026–30/06/2026)
Hechos_Ventas	Hechos	Línea de venta (FK: Cod_Tienda, Cod_Producto, Fecha_Venta)	570
Hechos_Turnos	Hechos	Turno de vendedora (FK: Cod_Tienda, Fecha)	158
Hechos_Objetivo_Mensual	Hechos	Producto x mes (FK: Cod_Producto, Fecha_Mes)	192

Hechos_Ventas se obtiene por anexión (append) de las hojas origen Ventas_Norte y Ventas_Sur, tal y como indica el parámetro Ruta_Fuentes de la hoja Parametros.

3. Transformaciones y ajustes por tabla (Fase 1)

3.1 Ventas_Norte y Ventas_Sur → Hechos_Ventas

Ambas hojas presentaban el mismo patrón de exportación por bloques (probable acumulación de varios volcados). Se aplicó idéntica secuencia de depuración a cada una antes de anexarlas:

Regla aplicada	Norte	Sur	Detalle
Eliminación de cabeceras repetidas intermedias	6	6	Filas de encabezado duplicadas dentro del cuerpo de datos
Eliminación de filas completamente vacías	7	7	Separadores entre bloques de exportación
Eliminación de fila de subtotal	1	1	Fecha_Venta = “TOTAL NORTE/SUR”
Eliminación de duplicados exactos	0	0	No se detectaron

Regla aplicada	Norte	Sur	Detalle
Homogeneización de clave Cod_Producto	44	40	Espacios en blanco al inicio del código
Normalización de Fecha_Venta	300	270	Formatos mixtos dd/mm/aaaa e ISO (aaaa-mm-dd) → tipo fecha único
Normalización numérica Cantidad / Precio_Unitario_Con_IGIC	0 no convertibles	0 no convertibles	Texto → número, sin pérdidas tras excluir cabeceras/vacíos
Normalización de Vendedora	183	169	Mayúsculas/minúsculas inconsistentes → Title Case
Metodo_Pago nulo	57	47	Dato realmente ausente en origen; se marca “No especificado”, no se imputa
Outliers (Cantidad o Precio ≤ 0)	0	0	No se detectaron valores negativos ni ceros anómalos

Tras el anexo (570 filas finales: 314+284 origen – 28 filas eliminadas por hoja), se verificó integridad referencial: 0 filas huérfanas de Cod_Producto contra Dim_Productos y 0 de Cod_Tienda contra Dim_Tiendas.

Medidas económicas calculadas por línea (a partir de Precio_Unitario_Con_IGIC, que incluye el IGIC repercutido):

- $\text{Importe_Total_Linea} = \text{Cantidad} \times \text{Precio_Unitario_Con_IGIC}$
- $\text{Base_Imponible_Linea} = \text{Importe_Total_Linea} / (1 + \text{tipo IGIC del producto})$
- $\text{IGIC_Linea} = \text{Importe_Total_Linea} - \text{Base_Imponible_Linea}$
- $\text{Margen_Linea} = \text{Base_Imponible_Linea} - (\text{Costo_Unitario} \times \text{Cantidad})$

El tipo de IGIC aplicado depende de Dim_Productos[Tipo_IGIC]: General = 7% (prendas y calzado) o Incrementado = 13,5% (joyería fina), según los parámetros de la hoja Parametros.

3.2 Dim_Tiendas

- Sin nulos ni duplicados; clave Cod_Tienda única sobre 8 filas.
- Normalización de tipo de dato: Fecha_Apertura (texto dd/mm/aaaa) → fecha, 8 filas afectadas.

3.3 Dim_Productos

- Sin depuración necesaria: 0 nulos, 0 duplicados, clave Cod_Producto única sobre 32 filas.

3.4 Turnos_Vendedoras → Hechos_Turnos

Regla aplicada	Filas afectadas	Detalle
Relleno hacia adelante (forward-fill) de Nombre / Apellidos / Supervisora	134	Cada bloque de turnos solo llevaba el nombre en la primera fila; verificado que todo cambio de tienda coincide con una fila con nombre (0 discrepancias)
Normalización de tipo Fecha	158	Texto dd/mm/aaaa → fecha
Descomposición de Comisiones_Venta	158	Campo de texto con lista separada por ',' → Nº_Ventas_Turno (recuento) y Comisiones_Total (suma)
Duplicados exactos	0	No se detectaron

Nota de diseño: Hechos_Ventas[Vendedora] (nombre + primer apellido) no se relaciona con el censo de Turnos_Vendedoras (nombre + apellidos completos): la clave no es fiable 1 a 1 por distinta granularidad de captura, y 2 vendedoras de Ventas (David Rodríguez, Elena Marrero) no tienen correspondencia verificable en Turnos. Se mantiene como atributo de texto en Hechos_Ventas, no como relación modelada, para evitar cruces incorrectos.

3.5 Ventas_Mensuales_Objetivo → Hechos_Objetivo_Mensual

- Sin nulos ni duplicados en origen (32 filas × 6 meses).
- Transformación estructural: unpivot de formato ancho (columnas Ene_2026...Jun_2026) a formato largo (Cod_Producto, Fecha_Mes, Objetivo_Unidades) → 192 filas (32 productos × 6 meses), compatible con el modelo de estrella y la relación con Dim_Calendario.

4. Modelo de datos (Fase 2)

Esquema en galaxia: tres tablas de hechos comparten dimensiones comunes (Dim_Tiendas, Dim_Productos, Dim_Calendario), sin dimensión compartida única para las tres.

Tabla origen (1)	Tabla destino (N)	Columna de enlace
Dim_Tiendas	Hechos_Ventas	Cod_Tienda
Dim_Tiendas	Hechos_Turnos	Cod_Tienda
Dim_Productos	Hechos_Ventas	Cod_Producto
Dim_Productos	Hechos_Objetivo_Mensual	Cod_Producto
Dim_Calendario	Hechos_Ventas	Fecha ↔ Fecha_Venta
Dim_Calendario	Hechos_Turnos	Fecha ↔ Fecha
Dim_Calendario	Hechos_Objetivo_Mensual	Fecha ↔ Fecha_Mes

Cada tabla de hechos aporta una única columna de fecha activa contra Dim_Calendario: no hay ambigüedad de relaciones activas/inactivas que resolver con USERELATIONSHIP.

Condición de Power Pivot verificada: la clave en el lado “1” de cada relación es única y sin duplicados en las tres dimensiones.

5. Medidas DAX propuestas (Fase 3)

Definidas exclusivamente a partir de columnas reales presentes en las tablas de hechos depuradas; no se han asumido supuestos no soportados por los datos.

Tabla	Medida	Fórmula DAX
Hechos_Ventas	Importe Total	SUM(Hechos_Ventas[Importe_Total_Linea])
Hechos_Ventas	Transacciones	COUNTROWS(Hechos_Ventas)
Hechos_Ventas	Unidades Vendidas	SUM(Hechos_Ventas[Cantidad])
Hechos_Ventas	Ticket Medio	DIVIDE([Importe Total],[Transacciones])
Hechos_Ventas	Base Imponible	SUM(Hechos_Ventas[Base_Imponible_Linea])
Hechos_Ventas	IGIC Repercutido	SUM(Hechos_Ventas[IGIC_Linea])
Hechos_Ventas	Margen Bruto	SUM(Hechos_Ventas[Margen_Linea])
Hechos_Ventas	Pct Margen sobre Base	DIVIDE([Margen Bruto],[Base Imponible])
Hechos_Ventas	Importe Mes Actual	CALCULATE([Importe Total], DATESMTD(Dim_Calendario[Fecha]))
Hechos_Ventas	Importe Mes Anterior	CALCULATE([Importe Total], DATEADD(Dim_Calendario[Fecha],-1,MONTH))

Tabla	Medida	Fórmula DAX
Hechos_Ventas	Variación Pct Intermensual	DIVIDE([Importe Mes Actual]-[Importe Mes Anterior],[Importe Mes Anterior])
Hechos_Objetivo_Mensual	Objetivo Unidades	SUM(Hechos_Objetivo_Mensual[Objetivo_Unidades])
Hechos_Ventas	Pct Cumplimiento Objetivo	DIVIDE([Unidades Vendidas],[Objetivo Unidades])
Hechos_Turnos	Horas Trabajadas	SUM(Hechos_Turnos[Horas_Trabajadas])
Hechos_Turnos	Comisiones Totales	SUM(Hechos_Turnos[Comisiones_Total])
Hechos_Turnos	Ventas en Turno	SUM(Hechos_Turnos[Nº_Ventas_Turno])
Hechos_Turnos	Comisión Media por Venta	DIVIDE([Comisiones Totales],[Ventas en Turno])

6. Limitación técnica — Fase 4

openpyxl (usado para generar el libro de salida) puede escribir tablas de Excel pero no el modelo tabular embebido (Power Pivot / Data Model) ni medidas DAX nativas: esa capa reside en el motor de Analysis Services integrado en Excel, con formato interno no accesible desde librerías Python puras ni desde el entorno sandbox de Claude (Linux, sin Excel instalado).

Por decisión del usuario, este entregable se limita al libro depurado más esta documentación. Las tablas ya están dadas de alta como Tablas de Excel (ListObject), listas para “Power Pivot > Agregar al Modelo de datos”; las relaciones (sección 4) y medidas (sección 5) deben crearse manualmente en Power Pivot, o mediante un script de automatización COM (xlwings) ejecutado en local con Excel de escritorio — como el ya disponible en 05_Construir_Modelo_PowerPivot.py de esta misma carpeta, adaptable a esta versión del libro.

7. Entregables

- 00_Dataset_Ventas_Boutiques_Solucion.xlsx — tablas depuradas (Dim_Tiendas, Dim_Productos, Dim_Calendario, Hechos_Ventas, Hechos_Turnos, Hechos_Objetivo_Mensual, Parametros) + hojas Log_Depuracion, Modelo y Medidas_DAX.
- 04_Relacion_Transformaciones_Ventas_Boutiques.docx (este documento) — trazabilidad de transformaciones y ajustes.