

Documentación ETL – Fact_Sales

1. Objetivo

Transformar y depurar la tabla de hechos **Fact_Sales** para que quede lista como tabla oficial del modelo estrella, asegurando trazabilidad, consistencia y métricas confiables.

2. Pasos realizados

- **Carga inicial:** Se importó el CSV Fact_Sales_ini con todas las columnas originales.
- **Deduplicación:**
 - Se eliminaron duplicados exactos (todas las columnas iguales).
 - Se consolidaron registros fragmentados por clave de grano **OrderID + ProductID**, aplicando reglas:
 - OrderDate: mínima fecha.
 - ShipDate: máxima fecha.
 - Sales, Quantity, Profit: suma.
 - Discount: promedio.
 - Devolucion: máximo (para asegurar 1 si hubo devolución).
- **Validaciones de calidad:**
 - **Nulos críticos:** chequeo en claves (OrderID, ProductID, CustomerID, OrderDate).
 - **Rangos:**
 - Discount entre 0 y 1.
 - Quantity > 0.
 - Sales y Profit no negativos.
 - **Codificación de devoluciones:** normalización a valores binarios (0/1).
- **Profit negativo:**
 - Se segmentaron casos en dos categorías:
 - Profit negativo sin devolución.
 - Profit negativo con devolución.
 - Esto permite diferenciar pérdidas por estrategia comercial (descuentos, costos altos) de pérdidas por devoluciones.
- **Tabla de auditoría (AuditSummary):** Se construyó una tabla resumen con tres bloques:
 - **ETL:** conteo de filas en cada etapa (original, sin duplicados, consolidado).
 - **Calidad:** cantidad de registros problemáticos (nulos, rangos, métricas).
 - **Profit negativo:** conteo de casos segmentados por devolución y campañas agresivas a pérdidas para captar clientes.

3. Resultados

- Fact_Sales queda **limpia, consolidada y trazable**.
- La tabla AuditSummary funciona como **dashboard técnico** para monitorear calidad y evolución del ETL.
- Se documentaron las reglas de negocio aplicadas (sumas, promedios, codificación binaria), garantizando transparencia.