

Documentación ETL – Dim_Productos

1. Objetivo

Depurar y consolidar la dimensión **Dim_Productos** para garantizar unicidad de ProductID, consistencia en categorías y subcategorías, y trazabilidad de problemas detectados.

2. Pasos realizados

- **Carga inicial:** Se partió de la consulta Dim_productos_ini, ya con encabezados promovidos y columnas renombradas correctamente (ProductID, ProductName, Category, SubCategory).
- **Validaciones de calidad:**
 - **Duplicados:** se detectaron ProductID con más de un registro asociado.
 - **Nulos críticos:** se identificaron registros con ProductID, Category o SubCategory vacíos.
 - **Consistencia de categorías:** se verificó que los duplicados compartieran la misma categoría y subcategoría.
- **Consolidación automática:**
 - Se agruparon registros por ProductID.
 - Se tomó el primer valor de ProductName como oficial.
 - Se mantuvo la categoría y subcategoría (que eran consistentes).
 - Se normalizaron textos de Category y SubCategory (ej. capitalización).
- **Tabla de auditoría (AuditDimProduct):** Se construyó un resumen con tres métricas:
 - **Duplicados ProductID detectados** → cuántos IDs aparecían más de una vez.
 - **Nulos críticos** → cuántos registros tenían problemas en campos clave.
 - **ProductID consolidados automáticamente** → cuántos quedaron en la tabla final limpia.

3. Resultados

- **Dim_Productos** queda con ProductID único y atributos normalizados.
- La tabla AuditDimProduct documenta problemas detectados y la cantidad de registros consolidados.
- Se mantiene trazabilidad clara y exportable para auditoría técnica.

Para este laboratorio, se usó la **consolidación automática** y se tomó el primer nombre como oficial.

En un proyecto real, lo ideal es exportar la tabla de duplicados y resolverlos con un **mapeo manual** para asegurar consistencia.