

Documentación ETL – Dim_Clientes

1. Objetivo

Depurar y consolidar la dimensión **Dim_Customers** para garantizar unicidad de CustomerID, consistencia en atributos descriptivos y trazabilidad de clientes que operan en múltiples ciudades.

2. Pasos realizados

- **Carga inicial:** Se partió de la consulta Dim_customers_ini, con encabezados promovidos y columnas renombradas sin espacios (CustomerID, CustomerName, Segment, Country, City, State, PostalCode, Region).
- **Validaciones de calidad:**
 - **Duplicados:** se detectaron clientes con más de una ciudad asociada.
 - **Nulos críticos:** se identificaron registros con CustomerID, CustomerName, Segment o Country vacíos.
 - **Consistencia de Segment y Country:** se verificó que un mismo cliente no cambie de segmento o país según la ciudad.

3. Tratamiento de duplicados

◆ Opción 1: Consolidación por CustomerID

- Se agrupa por CustomerID.
- Se toma el primer valor de atributos (CustomerName, Segment, City, etc.).
- Resultado: un registro único por cliente.
- Riesgo: se pierde el detalle de ciudades múltiples.
- Uso recomendado: **laboratorios simples** o modelos donde no se analiza geografía a nivel cliente.

◆ Opción 2: Tabla puente CustomerLocations

- Se mantiene **Dim_Customers** con un registro único por cliente.
- Se crea una tabla auxiliar **CustomerLocations** con todas las ciudades asociadas a cada CustomerID.
- Relación en el modelo: Fact_Sales → CustomerLocations → Dim_Customers.
- Resultado: máxima trazabilidad, sin perder detalle geográfico.
- Uso recomendado: **modelos profesionales** donde se requiere analizar clientes por ciudad.

4. Tabla de auditoría (AuditDimCustomer)

Se construyó un resumen con tres métricas clave:

- **Duplicados CustomerID detectados** → cuántos clientes operan en más de una ciudad.
- **Nulos críticos** → cuántos registros tienen problemas en campos clave.

- **CustomerID consolidados automáticamente** → cuántos quedaron en la tabla final limpia (Opción 1).

5. Resultados

- **Dim_Customers** queda con atributos normalizados y trazabilidad clara.
- La tabla AuditDimCustomer documenta duplicados, nulos y consolidación.
- El modelo estrella puede optar por:
 - **Dim_Customers consolidada** (simple).
 - **Dim_Customers + CustomerLocations** (profesional).

Para este laboratorio, se usó **Opción 1 - Consolidación por CustomerID**, se toma el primer valor de atributos (CustomerName, Segment, City, etc.) y se relaciona Fact_Sales → Dim_Customers directamente para no agregar un nivel más al modelo.