**Materia**: Introducción al Análisis de Datos  
**Clase**: 2 – Tipos de datos, ETL, Datawarehouse y Data Lake

**1️⃣ Parte A – Tipos de datos**

1. **Datos estructurados** → ID Pedido, Producto, Cantidad, Precio Unitario, Fecha.
2. **Semi estructurados** → Si exportamos el archivo a JSON (ejemplo: {"ID Pedido": 1, "Producto": "Maleta A", "Cantidad": 2, ...}).
3. **No estructurados** → Comentarios de clientes en texto libre, fotos de productos, audios de reclamos.

**2️⃣ Parte B – Mini-ETL en papel**

* **Problemas detectados:**
  1. Nombres inconsistentes (“MaletaA” → debería ser “Maleta A”).
  2. Mayúsculas desordenadas (“BOLSA” → “Bolsa”).
  3. Columnas mal definidas (ejemplo: Cantidad debería ser número entero, Precio Unitario número decimal).
* **Correcciones propuestas:**
  1. Estandarizar nombres de productos.
  2. Homogeneizar uso de mayúsculas/minúsculas.
  3. Ajustar tipo de datos en cada columna.
* **Nueva columna calculada:**  
  Monto Total = Cantidad × Precio Unitario

**3️⃣ Parte C – DWH vs Data Lake**

1. Si cargamos bd\_ventas **sin limpiar** se parece a un **Data Lake**, porque guarda los datos crudos tal cual llegan.
2. Si lo **limpiamos y normalizamos** se parece a un **Data Warehouse**, porque los datos quedan confiables, organizados y listos para análisis.

**4️⃣ Cierre reflexivo**

* Si hiciéramos un dashboard con bd\_ventas sin limpiar:  
  Los gráficos serían poco confiables (productos duplicados, sumas mal calculadas).
* Las empresas invierten en ETL y Data Warehouses porque:  
  → Necesitan **información de calidad, unificada y consistente** para tomar decisiones estratégicas.