



Taller de Transformación de datos con Power Query : Informe Final

Miguel Ángel Páez García

Programa de Ingeniería de Software, Facultad de Ingeniería

Universidad de Cundinamarca - Extensión Soacha

Grupo INGSOFT701, Arquitectura de Datos CAD65022720

Especialista Jorge Armando Jurado Peralta

Septiembre de 2025

Introducción

Se desarrolló un proceso completo de extracción, limpieza, transformación y consolidación de cinco conjuntos de datos (contactos, productos, órdenes, detalle y ventas mensuales) mediante Power Query y Power BI, con el objetivo de construir un modelo relacional coherente, trazable y analíticamente confiable. El proyecto busca garantizar la consistencia estructural y semántica de la información comercial, así como la identificación de patrones e inconsistencias a través de visualizaciones dinámicas.

Decisiones de Limpieza y Transformación

Los archivos originales en formato CSV presentaban errores de codificación, columnas desalineadas y campos corruptos.

Las principales acciones implementadas fueron:

- **Estandarización de encabezados y tipos de datos**, aplicando nomenclatura uniforme y formatos adecuados (texto, número, fecha, moneda, booleano).
- **Corrección de caracteres no válidos** (acentos, guiones, símbolos) y normalización de países a códigos **ISO 3166-1 alfa-2**.
- **Separación y validación de teléfonos y correos electrónicos**; campos incompletos reemplazados por "Desconocido".
- **Conversión de columnas de meses a filas** (despivotado) y creación de una columna de **fecha unificada** (#date(año, mes, 1)) para análisis temporal.
- **Integración de contactos** (clientes, empleados y proveedores) en una tabla consolidada con columna clasificatoria según el prefijo de ID (C, E, P).
- **Eliminación de duplicados y registros con total = 0**, preservando trazabilidad relacional.

Estas transformaciones respondieron a los desafíos de inconsistencia estructural y semántica, y garantizan integridad referencial entre las tablas principales.

Hallazgos y Análisis

A partir de la visualización del modelo final, se identificaron patrones y anomalías relevantes:

- Solo el **33% de las órdenes** posee cliente vinculado, evidenciando fallas en la trazabilidad comercial.
- Las **ventas 2024 concentran más del 60% del total histórico**, con varianza mensual casi nula, lo que sugiere carga masiva o regularización artificial.
- Predominan las ventas de diferentes tipos en **Chile (66,7%) y Colombia (50%)**, reflejando concentración geográfica y falta de diversificación de mercado.
- Se detectaron **órdenes con valores negativos o idénticos** sin correspondencia en el detalle, posiblemente asociadas a devoluciones o errores de registro.
- El catálogo de productos aumentó de **3 en 2020 a 20 en 2024**, mostrando expansión, aunque sin coherencia proporcional en las ventas por categoría.
- La mayoría de las ventas se registraron en países de habla hispana, destacando **Colombia, México y Chile**, lo que sugiere un enfoque regional del mercado.
- Un número considerable de **transacciones no está vinculado a un cliente identificado**, lo que indica deficiencias en la trazabilidad comercial.
- Las métricas indican **mayor volumen de ventas que de órdenes registradas**, reflejando una posible pérdida de integridad entre fuentes.
- Las categorías de **tecnología y servicios digitales** representaron la mayor proporción de ingresos, mientras que artículos físicos mostraron menor variabilidad.

Conclusión

El proceso ETL permitió obtener un modelo **limpio, normalizado y funcional**, apto para análisis exploratorios y de desempeño. Las decisiones adoptadas (desde la validación de campos hasta la consolidación relacional) aseguran calidad analítica y coherencia semántica del modelo. Sin embargo, las visualizaciones revelan **brechas en la captura y trazabilidad de datos**, lo que subraya la necesidad de fortalecer los mecanismos de registro y control de integridad en las fuentes originales.