

Transformación de Datos con Power Query

Nicolas David Gómez

Sergio Amaya

Fernanda Carantonio

Universidad de Cundinamarca

Ingeniería de software

Jorge Jurado Peralta

Soacha, Cundinamarca

30 de septiembre de 2025

Introducción

Con el fin de disponer de un modelo de datos limpio, estructurado y listo para análisis en Power BI, se aplicaron una serie de pasos de normalización, depuración y consolidación sobre los archivos fuente.

1. Carga Inicial de Datos

Se descargaron e importaron los archivos CSV, configurando como delimitador el carácter “;” para que cada archivo se estructurara en una sola columna y fuera posible su manipulación inicial.

2. Normalización de Tablas

Durante la preparación se identificaron inconsistencias y se aplicaron las siguientes transformaciones:

- Eliminación de dobles comas y caracteres innecesarios en direcciones de correo electrónico.
- Compleción de datos faltantes, asegurando uniformidad en la cantidad de columnas por fila.
- División de columnas a partir del delimitador “;” para garantizar la correcta separación de atributos y lograr una estructura tabular normalizada.

3. Una tabla de Ventas normalizadas lista para análisis temporal y categórico.

Se trabajó con las tablas `ventas_detalle` y `ventas_ordenes` para conformar una sola tabla de análisis.

- Se eliminaron filas vacías y se corrigieron valores inconsistentes.
- Se filtraron registros con cantidades negativas o nulas cuando no aportaban al análisis, y en los casos necesarios se marcaron como devoluciones para conservar trazabilidad.

Cálculo de ventas: En `ventas_detalle` se creó una columna calculada $Venta = Cantidad * PrecioUnitario$, garantizando valores numéricos en ambas columnas.

Unión con órdenes: Se aplicó un merge con `ventas_ordenes` usando `OrderID`.

Esto permitió incorporar campos adicionales como `CustomerID`, `FechaOrden` y `Estado`.

Se expandieron las columnas resultantes y se eliminaron registros duplicados o incompletos.

Normalización de fechas y claves: Se ajustaron las columnas de fecha a tipo `Date`.

Se validó la unicidad de `OrderID + ProductoID` para evitar inconsistencias.

El resultado fue una tabla de Ventas lista para análisis temporal y categórico, manteniendo detalle de transacciones y vinculación con clientes y órdenes.

4. Una tabla consolidada de Contactos que reúna información de clientes, proveedores y empleados.

Con el objetivo de unificar la información dispersa en diversas fuentes, se estandarizaron los nombres de columnas en las tablas de:

- contactos_clientes
- contactos_proveedores
- contactos_empleados

Una vez normalizadas, se empleó la herramienta “Anexar consultas” para integrar los tres orígenes en una sola tabla consolidada. El resultado fue denominado “Contactos”, lo que permite contar con una visión centralizada y coherente de todos los registros de clientes, proveedores y empleados.

5. Relaciones establecidas entre ventas, productos y contactos.

En la vista Modelo de Power BI se establecieron las siguientes relaciones:

Ventas → Productos Ventas[ProductoID] → productos_erp[ProductoID].

Relación muchos a uno, unidireccional de Productos a Ventas.

Ventas → Contactos Ventas[CustomerID] → Contactos[ContactoID].

Relación muchos a uno, unidireccional de Contactos a Ventas.

Estas relaciones garantizan un modelo analítico robusto y permiten el cruce de información temporal, categórica y de entidades.

Conclusión

El proceso aplicado permitió depurar errores, unificar criterios de columnas y consolidar datos en tablas maestras como Contactos. Estas decisiones de transformación garantizan calidad, consistencia y escalabilidad en el análisis posterior, resolviendo los desafíos de duplicidad, formatos heterogéneos y datos incompletos.