

DOCUMENTACIÓN TÉCNICA

Proceso ETL y Modelo de Datos

Proyecto: Laboratorio de Análisis de Datos con Power BI
Análisis de Ventas de E-Commerce Internacional (2022-2024)

Autor: Nicolás Adrián Tomadín
Email: nicolas.tomadin1995@gmail.com
Fecha de Entrega: 04/12/2025
Herramienta: Microsoft Power BI Desktop

PROCESO ETL (EXTRACCIÓN, TRANSFORMACIÓN Y CARGA)

El proceso ETL se implementó íntegramente en Power BI Desktop utilizando el Editor de Power Query para integrar y transformar múltiples fuentes de datos heterogéneas. Se realizó la extracción de cuatro fuentes distintas: tres archivos Excel ([Productos.xlsx](#), [Clientes.xlsx](#), [Ventas.xlsx](#)) y una tabla de objetivos comerciales anuales ingresada manualmente mediante la funcionalidad "Introducir datos". Durante la fase de transformación, se aplicaron procesos exhaustivos de limpieza de datos que incluyeron la eliminación sistemática de valores nulos en campos críticos, la validación y corrección de tipos de datos (fecha, número decimal, texto), el filtrado de todas las órdenes con estado "Cancelado" para enfocarse exclusivamente en transacciones completadas, y la estandarización de valores categóricos como nombres de países, categorías de productos y métodos de pago para garantizar consistencia. Adicionalmente, se enriquecieron las tablas de dimensión mediante la creación de columnas calculadas personalizadas: en Productos se generó un "Rango de Precio" (Bajo, Medio, Alto) basado en umbrales de valor; en Clientes se extrajo el "Dominio de Email" para identificar perfiles B2C vs. B2B y se calculó la "Antigüedad del Cliente" en años para medir lealtad. Estas transformaciones aseguraron la calidad, integridad y coherencia de la información, preparando una base analítica robusta lista para el modelado dimensional y la creación de visualizaciones profesionales.

MODELO DE DATOS Y ESTRUCTURA DIMENSIONAL

Se implementó un modelo de datos bajo esquema estrella (Star Schema), considerado la mejor práctica en Business Intelligence por su rendimiento optimizado y simplicidad para análisis de negocio. La estructura consta de una tabla de hechos central denominada "Ventas" que almacena aproximadamente 300-500 transacciones comerciales con granularidad de línea de venta,

conteniendo métricas cuantitativas (Cantidad, Precio Unitario, Precio Total) y claves foráneas que la vinculan con cuatro tablas de dimensión que proporcionan contexto descriptivo: **Productos** (100 registros con información de catálogo como categoría, marca, proveedor y precio referencia), **Clientes** (~250 registros con datos demográficos incluyendo país, segmento y antigüedad calculada), **Calendario** (1,096 días cubriendo 2022-2024, generada mediante código DAX con atributos derivados como año, mes, trimestre y día de semana), y **Objetivos** (3 registros con metas anuales de ventas para comparación con resultados reales). Las relaciones entre tablas se configuraron meticulosamente con cardinalidad muchos-a-uno (*:1) y dirección de filtro cruzado unidireccional en las relaciones principales (Ventas-Productos, Ventas-Clientes, Ventas-Calendario) para garantizar un flujo de contexto de filtro consistente y eficiente, utilizando filtro bidireccional únicamente en la relación Calendario-Objetivos para permitir análisis comparativos temporales. La tabla Calendario fue marcada explícitamente como "tabla de fechas" en Power BI, habilitando funciones avanzadas de inteligencia temporal (Time Intelligence) como **SAMEPERIODLASTYEAR** para cálculos de crecimiento interanual. Este diseño dimensional garantiza integridad referencial completa, elimina redundancia de datos y proporciona una arquitectura escalable capaz de incorporar nuevas dimensiones o métricas futuras sin modificar la estructura base del modelo.

MEDIDAS DAX Y OPTIMIZACIÓN DEL MODELO

Se desarrollaron ocho medidas DAX explícitas organizadas en una tabla auxiliar denominada "_Medidas" para mejorar la mantenibilidad y profesionalismo del modelo, incluyendo cálculos esenciales como **Ventas Totales** (**SUM(Ventas[Precio Total])**), **Total Órdenes** (**COUNTROWS(Ventas)**), **Ticket Promedio** (**DIVIDE([Ventas Totales], [Total Órdenes], 0)**), **Ventas Año Anterior** (utilizando **CALCULATE** con **SAMEPERIODLASTYEAR**), **Porcentaje** (variación porcentual interanual), **Clientes Únicos** (**DISTINCTCOUNT**), **% Cumplimiento Objetivo** (comparación contra metas anuales usando **FILTER** y **SELECTEDVALUE**), y **Margen Estimado** (simulación de rentabilidad). Estas medidas aprovechan funciones DAX avanzadas que modifican contexto de filtro, manejan divisiones por cero y implementan inteligencia temporal, garantizando cálculos robustos y precisos que alimentan los KPIs visualizados en el dashboard. El modelo está optimizado para rendimiento mediante la compresión automática del motor VertiPaq de Power BI, la eliminación de columnas redundantes que no aportan valor analítico, y el uso eficiente de tipos de datos apropiados. Esta arquitectura proporciona una base técnica sólida y escalable para análisis estratégicos en el sector e-commerce, permitiendo a ejecutivos y analistas tomar decisiones informadas basadas en datos confiables, responder preguntas complejas de negocio mediante segmentación multidimensional, y monitorear el cumplimiento de objetivos comerciales a través de visualizaciones.

interactivas profesionales que se actualizan dinámicamente según los filtros aplicados por el usuario.