

# Análisis Integrado de Datos Operativos y Comerciales

---

## Descripción General

Este proyecto realiza un análisis exploratorio y cruzado de datos transaccionales, operativos y de estructura organizacional de una empresa que gestiona ventas, compras, gastos y recursos humanos a través de múltiples sucursales.

Se parte de **10 datasets transformados y normalizados**, con relaciones estructuradas a través de un **modelo entidad-relación (ERD)**.

---

## Transformación y limpieza de datos

Para asegurar la integridad del modelo de datos y permitir los análisis posteriores, se aplicaron los siguientes procesos a cada dataset:

### Clientes\_transformados.csv

- Relleno de nulos (categóricos con valor anterior/posterior; numéricos con mediana)
- Conversión de fechas (*Fecha\_Alta*, *Fecha\_Ultima\_Modificacion*)
- Normalización de textos (minúsculas, espacios eliminados, ; → ,)
- Eliminación de filas con nulos residuales

### Compra\_transformada.csv

- Limpieza de nulos y tipado correcto (*IdProducto*, *IdProveedor*)
- Conversión de fechas
- Normalización de texto y precios

### Empleados\_transformados.csv

- Transformación del campo *Sucursal* para relacionarse con ID
- Añadido *IdSucursal* para permitir joins
- Limpieza y unificación de cargos y salarios

### Gasto\_transformado.csv

- Conversión de fechas
- Preparación para joins con *Sucursales* y *TiposDeGasto*
- Limpieza de textos y nulos

### PRODUCTOS\_transformado.csv

- Eliminación de outliers extremos en precio
- Limpieza de campos *Tipo* y *Concepto*

- Normalización textual

### Proveedores\_transformado.csv

- Revisión y limpieza de nombres y localidades
- Estandarización de provincias/departamentos
- Normalización textual

### Sucursales\_transformado.csv

- Coordenadas geográficas listas para mapa (**Latitud**, **Longitud**)
- Preparado para joins mediante campo **ID** y nombre único

### Venta\_transformado.csv

- Conversión de fechas (**Fecha**, **Fecha\_Entrega**)
- Tipado correcto en claves foráneas
- Revisión de precios y cantidades
- Eliminación de nulos

### TiposDeGasto\_T.csv

- Estandarización de descripciones
- Preparado para join con **Gasto\_transformado.csv**

### CanalDeVenta\_Tranfor.csv

- Limpieza de códigos y descripciones
- Unificación de formato para relación con **Ventas**



## Datasets utilizados

Archivo	Descripción
<b>Clientes_transformados.csv</b>	Información de clientes, edades y localidades
<b>Empleados_transformados.csv</b>	Nómina de empleados, cargos, sucursales y salarios
<b>Sucursales_transformado.csv</b>	Sucursales con datos geográficos
<b>Productos_transformado.csv</b>	Catálogo de productos con tipo y precio
<b>Proveedores_transformado.csv</b>	Datos de proveedores
<b>Compra_transformada.csv</b>	Compras realizadas a proveedores
<b>Venta_transformado.csv</b>	Ventas realizadas a clientes
<b>Gasto_transformado.csv</b>	Gastos operativos por sucursal
<b>TiposDeGasto_T.csv</b>	Clasificación de los tipos de gastos

Archivo	Descripción
CanalDeVenta_Tranfor.csv	Canales a través de los cuales se concretan las ventas

## Modelo de Datos (ERD)

El modelo relacional se construyó con claves primarias y foráneas, y fue visualizado en un **diagrama ERD** que respeta:

- Relaciones entre entidades principales (clientes, productos, empleados, sucursales)
- Tablas transaccionales: ventas, compras, gastos
- Tablas auxiliares: canales, tipos de gasto

Además, se aplicaron transformaciones de tipo y normalización para que las relaciones sean válidas.

## Análisis cruzado realizado

Se desarrollaron 8 análisis cruzados claves con visualizaciones, que permiten generar conocimiento accionable:

1. **Productos más vendidos vs. más comprados**
2. **Sucursales con más ventas vs. más gastos**
3. **Relación entre salario y volumen de ventas generado por empleado**
4. **Preferencia de productos según edad de cliente**
5. **Canal de venta vs. volumen y monto total**
6. **Tipo de gasto predominante por sucursal**
7. **Proveedores con mayor volumen de compras**
8. **Comparativa de precios de compra vs. venta (margen por producto)**

## Aplicaciones posibles

- Optimización de stock y compras
- Gestión de rentabilidad por sucursal
- Evaluación de desempeño comercial
- Segmentación de clientes y canales
- Control presupuestario y gastos
- Análisis de márgenes y estrategia de precios

## Herramientas utilizadas

- **Python:** pandas, seaborn, matplotlib, folium, networkx
- **Power BI (recomendado)** para visualización integrada
- **ERD lógico** generado y visualizado como referencia de modelo