

Entregable Intermedio — TS2

Tema: Analítica de Datos y Business Intelligence aplicado a datos farmacéuticos

Periodo: 1 sep – 28 nov 2025 — Entregable intermedio (10/oct/2025)

Autor: Luis Mario Ayala Castellanos ID: 174902

1. Resumen

Este proyecto aplica técnicas de Analítica de Datos y Business Intelligence (BI) a un conjunto de ventas farmacéuticas con el objetivo de generar KPIs accionables y un pipeline reproducible para la toma de decisiones en inventarios y marketing. En este entregable se presenta el marco teórico, la definición del caso, objetivos SMART, dataset y diccionario de datos, propuesta de KPIs, y un avance de ETL con visualizaciones preliminares basadas en una muestra (50 registros) como prueba de concepto.

2. Marco teórico y estado del arte

Analítica descriptiva, diagnóstica y predictiva: La analítica descriptiva se enfoca en resumir qué ocurrió en el pasado, la diagnóstica busca explicar causas, y la predictiva intenta anticipar comportamientos futuros. La arquitectura BI integra datos mediante procesos ETL en un data warehouse, seguido por análisis y dashboards. Se incluyen KPIs como ventas netas, ticket promedio, rotación de inventario y ventas por categoría.

3. Definición del problema y objetivos SMART

Caso: Optimizar la gestión comercial de una farmacia mediante analítica de ventas históricas.

Problema: Falta de métricas estandarizadas que permitan priorizar surtido/inventario y detectar productos con baja rotación.

Objetivos:

1. Definir y calcular KPIs
2. Implementar un pipeline ETL reproducible.
3. KPIs propuestos
 1. Ventas Totales = $\Sigma (\text{Cantidad} \times \text{PrecioUnitario})$.
 2. Número de Transacciones = conteo de filas.
 3. Ticket Promedio = Ventas Totales / Número de Transacciones.
 4. Clientes Únicos = `COUNT_DISTINCT(ClienteID)`.

5. Ventas por Categoría = Σ TotalVenta por CategoríaProducto.
6. Top N Productos por ventas = orden descendente por TotalVenta.
7. Porcentaje de ventas por Ciudad = $(\text{Ventas Ciudad} / \text{Ventas Totales}) \times 100$.

Visualizacion de los KPIs en el notebook

```
KPIs preliminares
Ventas Totales: $14,154.74
Número de Transacciones: 49
Ticket Promedio: $288.87
Clientes Únicos: 49

Ventas por Categoría:
CategoríaProducto
Antihistamínico    3252.26
Antidiabético     2809.95
Antiácido          2587.15
Analgésico         2077.86
Suplemento        2044.32
Antibiótico       1383.20
Name: TotalVenta, dtype: float64

Top 5 Productos más vendidos:
NombreProducto
Glibenclamida     1642.66
Diclofenaco       1238.54
Desloratadina     1165.66
Ranitidina        1113.23
Fexofenadina      1079.10
Name: TotalVenta, dtype: float64
```

Fig 1. KPIs preliminares

Porcentaje de ventas por Ciudad:	
Ciudad	
San Octavio los altos	6.843291
San Javier los altos	6.303542
San Elisa los altos	5.281058
Nueva Bahrein	5.208432
San Flavio los altos	4.705138
San Timoteo los bajos	4.612589
Vieja Bolivia	4.103926
Nueva Guyana	3.659198
Nueva Jordania	3.622532
Nueva Nueva Zelandia	3.381200
Vieja Malta	3.147144
San Nicolás los bajos	2.944597
San Marcela los bajos	2.930467
Nueva Australia	2.624704
Nueva Colombia	2.573979
Nueva Paraguay	2.486446
San José Eduardo de la Montaña	2.431341
Nueva Israel	2.345221
Vieja Bhután	2.328831
San Aida los altos	2.149245
San Luis Manuel los bajos	2.145854
San Renato de la Montaña	1.903673
Vieja Angola	1.886647

Fig 2. KPIs preliminares

5. Dataset y data dictionary

Fuente: Dataset propio de la farmacia (original 30,000 registros). Para este entregable se utilizó una muestra de 50 registros.

Columnas incluidas: ClienteID, NombreCliente, Edad, Ciudad, ProductoID, NombreProducto, CategoríaProducto, FechaCompra, Cantidad, PrecioUnitario, TotalVenta.

Ejemplo de data dictionary:

ClienteID (string) — ID cliente. Ej: C001.

NombreCliente (string) — Nombre completo (PII, se recomienda anonimizar).

Edad (int) — Edad en años.

Ciudad (string) — Ciudad de residencia.

ProductoID (string) — ID producto.

NombreProducto (string) — Descripción producto.

CategoríaProducto (string) — Categoría.
FechaCompra (datetime) — Fecha de compra.
Cantidad (int) — Unidades vendidas.
PrecioUnitario (float) — Precio por unidad.
TotalVenta (float) — $\text{Cantidad} \times \text{PrecioUnitario}$.

6. Avance ETL / limpieza Operaciones realizadas:

- Carga del archivo Excel con pandas.
- Limpieza de nombres de columnas.
- Conversión de fechas a datetime.
- Conversión de variables numéricas (Cantidad, PrecioUnitario).
- Creación del campo TotalVenta.
- Chequeo de nulos y exportación de la muestra procesada.

```
# ETL preliminar
import pandas as pd

def load_data(path="C:\\Users\\luism\\OneDrive\\Documentos\\A. 9no Semestre\\Temas 2\\Reprocesamiento de datos por acentos\\demoDatosRaw.xlsx"):
    df = pd.read_excel(path)
    df.columns = df.columns.str.strip()
    df["FechaCompra"] = pd.to_datetime(df.get("FechaCompra"), errors="coerce")
    df["Cantidad"] = pd.to_numeric(df.get("Cantidad"), errors="coerce")
    df["PrecioUnitario"] = pd.to_numeric(df.get("PrecioUnitario"), errors="coerce")
    df["TotalVenta"] = df["Cantidad"] * df["PrecioUnitario"]
    return df

if __name__ == "__main__":
    df = load_data("C:\\Users\\luism\\OneDrive\\Documentos\\A. 9no Semestre\\Temas 2\\Reprocesamiento de datos por acentos\\demoDatosRaw.xlsx")
    # Guardar
    output_path = "C:\\Users\\luism\\OneDrive\\Documentos\\A. 9no Semestre\\Temas 2\\Reprocesamiento de datos por acentos\\demoDatosProcesados.csv"
    df.to_csv(output_path, index=False, encoding="utf-8")
    print("Archivo procesado guardado en:", output_path)
```

Fig 3. ETL preliminar

7. Visualizaciones preliminares

Se generaron gráficos de ventas por ciudad, ventas por categoría y ventas por fecha. En la muestra se observó concentración de ventas en algunas ciudades y categorías.

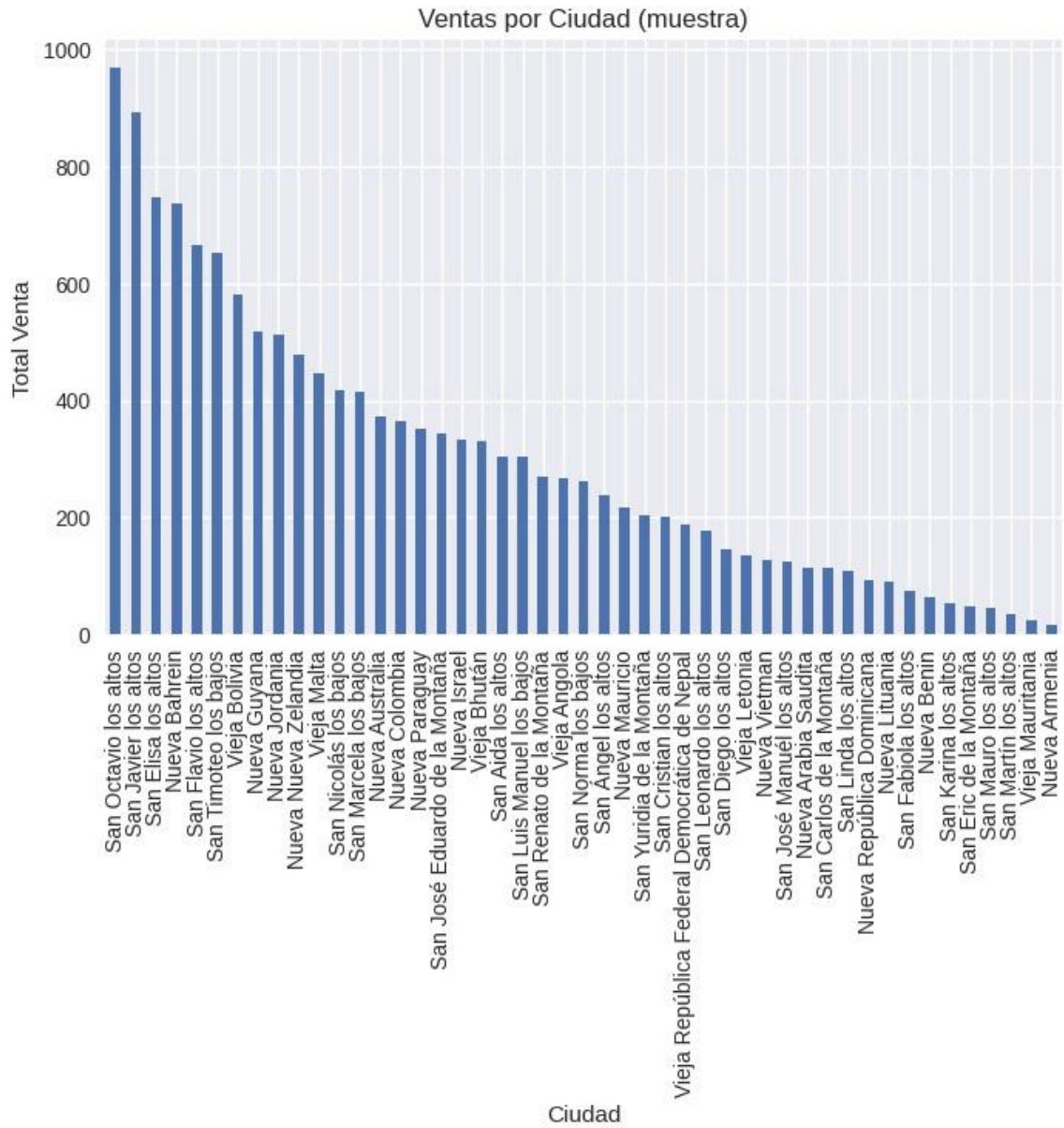


Fig 4. Ventas por Ciudad

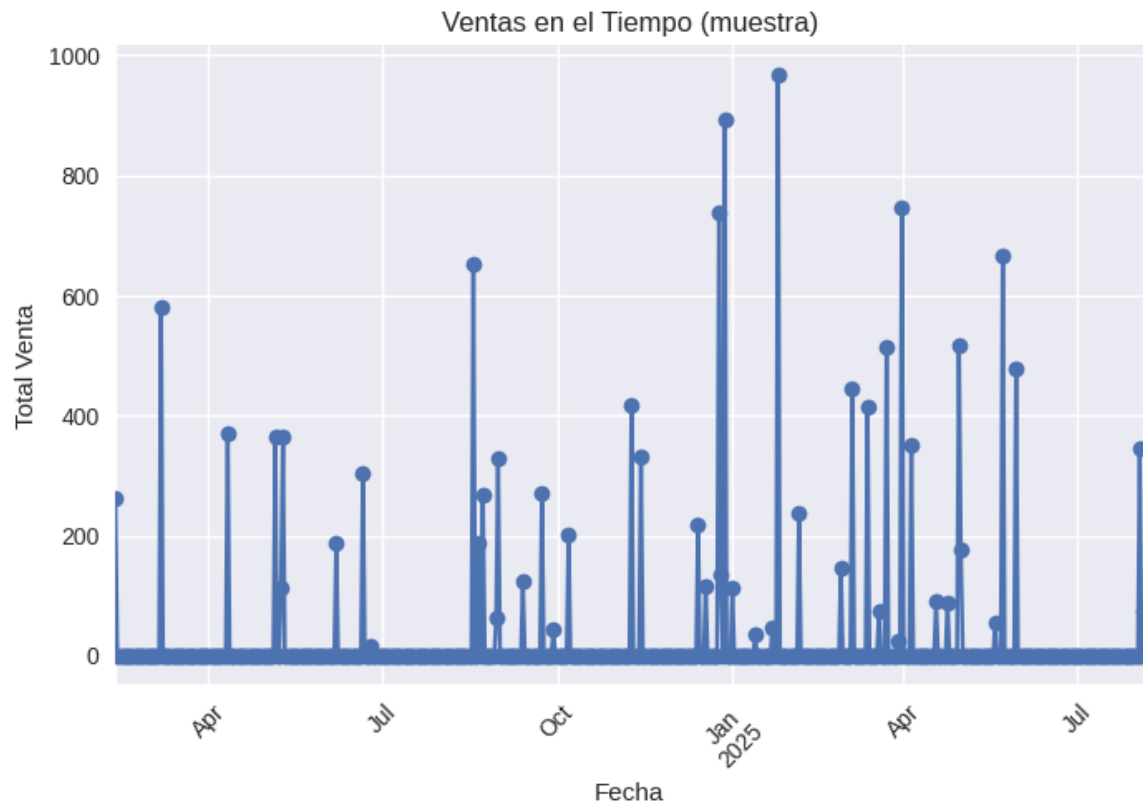


Fig 5. Ventas por fechas

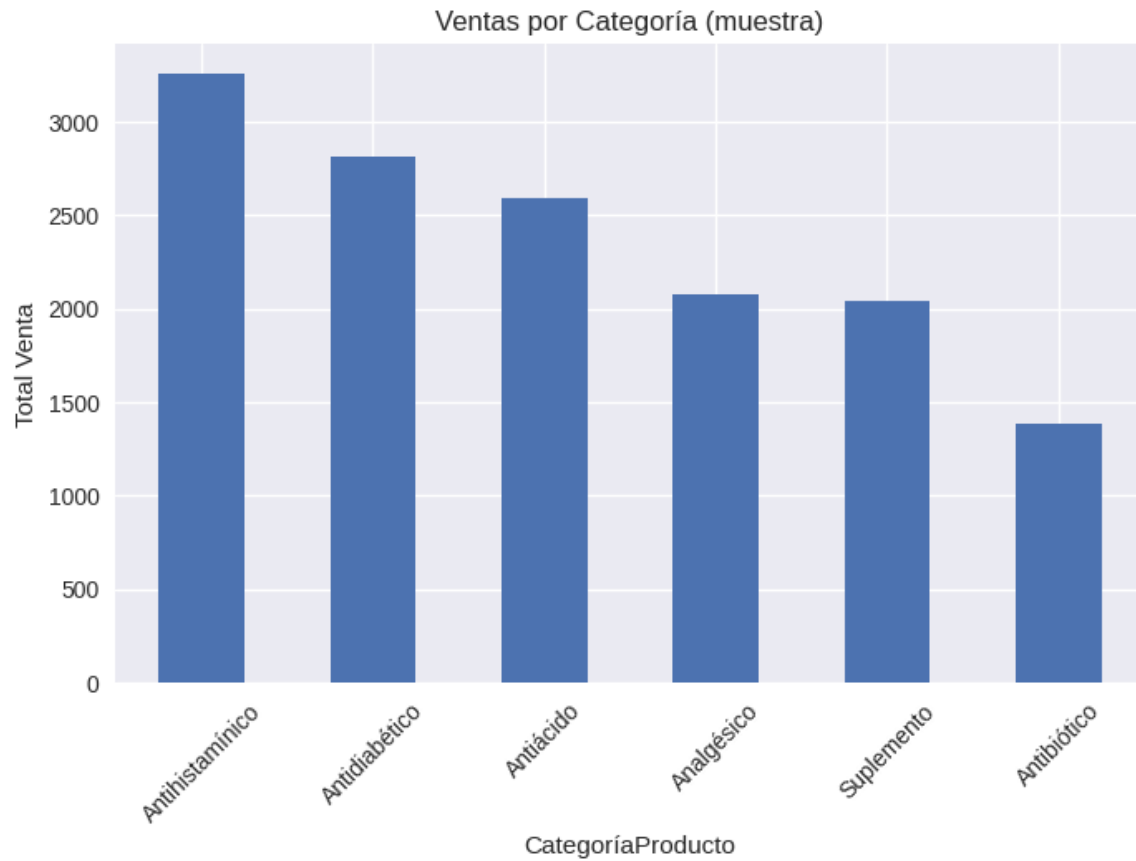


Fig 6. Ventas por Categoría

Referencias

- Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice (2nd ed.). OTexts.
- Few, S. (2012). Show Me the Numbers: Designing Tables and Graphs to Enlighten. Analytics Press.
- Provost, F., & Fawcett, T. (2013). Data Science for Business. O'Reilly.