

Análisis Integral de Ventas del Sector Retail mediante Python y Power BI

1. Introducción

El objetivo de este proyecto ha sido desarrollar un análisis completo de un conjunto de datos de ventas del sector Retail, aplicando todo el proceso de trabajo que se espera en un proyecto real de análisis de datos: limpieza, transformación, análisis estadístico y construcción de un Dashboard que permita interpretar los resultados de forma clara.

La idea no ha sido simplemente generar gráficos, sino entender qué está pasando en los datos y extraer conclusiones que tengan sentido desde el punto de vista de negocio.

El proyecto se ha dividido en dos grandes bloques:

- Trabajo en Python (limpieza, transformación y análisis exploratorio).
- Construcción del Dashboard en Power BI.

2. Conjuntos de datos utilizados

Para este proyecto se han utilizado dos conjuntos de datos claramente diferenciados:

2.1 Dataset original (datos en bruto).

El primer conjunto de datos es el Dataset “Online Retail”, que recoge transacciones reales de una empresa de comercio electrónico.

Este Dataset contiene más de 500.000 registros y cada fila representa una línea de factura, es decir, una venta concreta de un producto.

Incluye variables como:

- InvoiceNo (número de factura).
- StockCode (código del producto).
- Description (nombre del producto).
- Quantity (cantidad vendida).
- InvoiceDate (fecha de la transacción).
- UnitPrice (precio unitario).
- CustomerID (identificador del cliente).

- Country (país).

Este conjunto se ha mantenido en el repositorio como Dataset en bruto, sin modificar.

Adjunto las referencias de los datos:

<https://www.kaggle.com/datasets/ulrikthgepedersen/online-retail-dataset>.

<https://www.kaggle.com/datasets/mscdoamit/world-population-by-country-years-1960-2023>.

2.2 Dataset final transformado.

A partir del Dataset original se ha generado un segundo conjunto de datos tras realizar procesos de limpieza y transformación en Python.

En esta fase se realizaron cambios importantes:

- Eliminación de valores nulos en variables clave.
- Revisión y eliminación de registros duplicados.
- Tratamiento de cantidades negativas asociadas a devoluciones.
- Conversión correcta de fechas.
- Creación de nuevas variables temporales.

Además, se creó una nueva variable fundamental para el análisis: • Facturación = Quantity × UnitPrice.

También se añadieron:

- Año.
- Mes.
- Nombre del mes.
- Trimestre.

Este Dataset transformado es el que se utilizó posteriormente en Power BI.

De esta forma se cumplen los requisitos de disponer tanto de los datos originales como del conjunto final preparado para análisis.

3. Limpieza y transformación de datos.

El proceso de limpieza se realizó utilizando Python y la librería Pandas.

Antes de empezar a analizar nada, fue necesario asegurar que los datos eran coherentes.

Para ello:

- Se revisaron valores nulos.
- Se eliminaron registros con CustomerID vacío cuando afectaban al análisis.
- Se eliminaron duplicados.
- Se validaron los tipos de datos.
- Se revisaron posibles valores extremos.

La transformación de fechas fue especialmente importante, ya que permitió crear variables temporales que después fueron clave para el análisis estacional.

La creación de la variable Facturación permitió medir el impacto real de cada transacción.

Esta fase fue esencial para poder trabajar posteriormente con datos fiables.

4. Análisis exploratorio y estadístico.

Una vez preparado el Dataset final, se realizó un análisis exploratorio para entender el comportamiento general del negocio.

Algunos de los puntos analizados fueron:

- Facturación total.
- Distribución de ventas por país.
- Evolución mensual.
- Comportamiento por trimestre.
- Productos con mayor facturación.
- Relación entre volumen vendido y facturación generada.

Al analizar los datos comenzaron a aparecer patrones claros:

- Reino Unido concentra la mayor parte de la facturación.
- El cuarto trimestre muestra un incremento notable en ventas.
- Existen productos con facturación muy elevada pese a no tener el mayor volumen de unidades.
- La estacionalidad influye significativamente en los ingresos.

El análisis no se limitó a obtener cifras, sino que se intentó interpretar cada resultado.

5. Construcción del Dashboard en Power BI.

Una vez finalizada la fase en Python, el Dataset transformado se importó a Power BI.

El Dashboard se estructuró en dos páginas:

Página 1 – Visión General.

En esta página se incluyeron:

- Top 10 países por facturación.
- Evolución mensual de la facturación.
- Indicadores generales de volumen.

El objetivo fue ofrecer una visión clara y rápida del rendimiento global.

Página 2 – Análisis Detallado por Producto.

En esta página se trabajó un nivel más profundo:

- Top 10 productos por facturación.
- Facturación por trimestre.
- Relación entre cantidad vendida y facturación.
- Distribución por país mediante gráfico circular.

Aquí se buscó analizar el comportamiento a nivel más específico.

El diseño se mantuvo coherente utilizando una escala de colores en tonos azules para dar uniformidad visual y profesionalidad.

6. Valor del Dashboard.

El Dashboard permite:

- Identificar mercados clave.
- Detectar productos estratégicos.
- Analizar comportamiento estacional.
- Observar concentración de ingresos.
- Comparar volumen frente a rentabilidad.

No se trata solo de gráficos, sino de una herramienta que permite tomar decisiones con base en datos.

7. Conclusiones.

Después de realizar todo el proceso, se pueden destacar varias conclusiones importantes:

- La facturación está fuertemente concentrada en Reino Unido.
- El cuarto trimestre es el más relevante en términos de ingresos.
- Existe concentración de facturación en productos concretos.
- No siempre el mayor volumen implica mayor facturación.
- La estacionalidad tiene un impacto claro en el negocio.

El proyecto ha permitido aplicar un flujo completo de análisis de datos, desde datos en bruto hasta visualización final.