

Proyecto de Análisis de Ventas — SQL · Python · Jupyter · Tableau

Autor: Guille Cases

Tecnologías: PostgreSQL (pgAdmin), Python (Pandas), Jupyter Notebook, Tableau Public

Estado: Finalizado

1. Descripción General

Este proyecto consiste en un análisis completo del ciclo de vida de datos de ventas, utilizando herramientas profesionales de análisis y visualización.

Se replicó un flujo real de analista de datos:

- Modelado y carga de datos en PostgreSQL
- Validaciones SQL para asegurar integridad
- ETL automatizado mediante Python
- Análisis exploratorio en Jupyter
- Dashboard interactivo en Tableau

Todo el proceso es **100% reproducible** mediante el script `run_all.py`.

2. Estructura de Datos y Fuente

Los datos fueron creados manualmente para simular un escenario real con:

- **Clientes (customers)**
- **Productos (products)**
- **Pedidos (orders)**
- **Líneas de pedido (order_items)**

Cada archivo CSV se encuentra en la carpeta `/data`.

Las relaciones entre tablas siguen un modelo relacional **1:N**.

3. Trabajo en SQL (PostgreSQL)

Las tablas fueron creadas en pgAdmin con claves primarias y foráneas.

Se implementaron validaciones esenciales:

- Conteo de filas por tabla
- Verificación de claves foráneas
- Comprobación de cálculos (`qty × unit_price`)
- Identificación de registros huérfanos

La carga de datos se realizó mediante `COPY`.

4. Pipeline Python Reproducible

El archivo `run_all.py` automatiza por completo:

- Carga de CSV a PostgreSQL
- Ejecución de validaciones SQL
- Ejecución de consultas de análisis
- Exportación de resultados a `/outputs/`
- Creación de un Excel resumen (`xlsxwriter`)
- Registro en `run_logs/run_all.log`

Se ejecuta con:

```
python run_all.py
```

5. Análisis Exploratorio (EDA) con Pandas

En Jupyter Notebook se realizó:

- Análisis de ventas por fecha
- Ventas por ciudad
- Ventas por categoría
- Ventas por cliente
- Distribuciones y métricas clave

Se generó un dataset final consolidado:

df_completo_para_tableau.csv

6. Dashboard en Tableau

 [Dashboard público:](#)

Visualizaciones incluidas:

- Serie temporal de ventas diarias
- Ventas por ciudad
- Ventas por categoría
- Top productos
- Top clientes

Métricas analizadas (datos reales):

- SUM(Total)
- Ventas por ciudad (Madrid, Valencia, Barcelona, Sevilla)
- Ventas por categoría

- Ventas por cliente
- Ventas por producto
- Porcentaje por ciudad
- Totales diarios

7. Conclusiones del Análisis

Basado en el dashboard y tus datos reales:

- Madrid representa aproximadamente el 52% de las ventas totales.
- Electrónica es la categoría líder.
- Los clientes top concentran una parte destacada de facturación.
- La serie temporal muestra picos de ventas bien definidos.

8. Valor Profesional del Proyecto

Habilidades:

- SQL para modelado y análisis
- ETL automatizado en Python
- Análisis exploratorio profesional
- Visualización interactiva con Tableau
- Diseño y estructura de proyecto real
- Storytelling y comunicación con datos

9. Estructura del Proyecto

```
 proyecto_sql/
  |
  └── data/
  └── outputs/
  └── queries/
  └── queries_validation/
  └── notebooks/
  └── run_logs/
  └── run_all.py
  └── README.md
```

2. RESUMEN EJECUTIVO

Este proyecto presenta un análisis completo del ciclo de vida de datos de ventas, desarrollado con herramientas profesionales: PostgreSQL, Python, Jupyter Notebook y Tableau Public. El objetivo principal fue simular el flujo de trabajo de un analista de datos en un entorno real, desde la creación de la base de datos hasta la visualización final de resultados.

La base de datos se diseñó con cuatro tablas: clientes, productos, pedidos y líneas de pedido, todas relacionadas mediante claves foráneas. Para garantizar la integridad del modelo, se realizaron validaciones SQL enfocadas en conteos, verificaciones de relaciones y comprobaciones matemáticas de los totales.

Posteriormente, se desarrolló un pipeline reproducible en Python (`run_all.py`) que automatiza la carga de datos, las validaciones, la ejecución de consultas analíticas y la generación de outputs en formato CSV y Excel. Este pipeline permite regenerar todo el proyecto con un solo comando, garantizando trazabilidad y consistencia.

En Jupyter Notebook se llevó a cabo el análisis exploratorio (EDA), generando métricas esenciales como ventas por cliente, categoría, ciudad y evolución diaria. El dataset final consolidado se exportó a Tableau, donde se creó un dashboard interactivo que permite explorar los principales indicadores del negocio.

Los resultados muestran que **Madrid representa aproximadamente el 52% del total de ventas**, la categoría **Electrónica es la más destacada**, y los **clientes top concentran una parte significativa de la facturación**, siguiendo un patrón de concentración típico (Pareto).

Este proyecto demuestra competencias clave como modelado de datos, ETL, análisis exploratorio, visualización ejecutiva y storytelling con datos, convirtiéndose en una pieza sólida del portfolio profesional como Analista de Datos Junior.