**Proyecto Integrador - Módulo 1**

**Análisis y Modelado de Ventas**

Autor: Alejandro N. Herrera Soria

Tecnologías utilizadas: SQL (MySQL), Python, Pandas, Jupyter Notebook (VSC), GitHub

**Descripción General**

Este proyecto aborda la exploración, transformación, automatización y preparación de datos de ventas para su posterior análisis y modelado. Se divide en tres avances técnicos complementarios y progresivos que profundizan desde consultas SQL hasta ingeniería de features en Python.

—

**Avance 1: Consultas SQL Avanzadas**

Objetivo

Explorar el conjunto de datos mediante consultas avanzadas con:

- Funciones de agregación y ventana

- Análisis de concentración de ventas

- Identificación de productos líderes, clientes y vendedores clave

Destacados

- Ranking de productos más vendidos y sus vendedores top

- Distribución de clientes por producto y adopción relativa

- Proporción de ventas dentro de categorías (con funciones ventana)

- Ranking de productos dentro de cada categoría

\*Se entrega script SQL completo + capturas y análisis breves de cada consulta.\*

---

**Avance 2: Automatización y Optimización SQL**

Objetivo

Mejorar la eficiencia operativa y trazabilidad del sistema de ventas.

Componentes clave

- Vistas: creación de vistas reutilizables para las consultas críticas

- Procedimientos almacenados: automatización de operaciones de extracción

- Trigger: activación que registra si un producto supera las 200.000 unidades

- Benchmarking: comparación de tiempos de ejecución antes y después de aplicar índices estratégicos (individuales y compuestos)

\*Se entrega script completo con comparación de tiempos comentados en el propio script\*

---

**Avance 3: Limpieza, Feature Engineering y Preparación para ML**

Objetivo

Transformar los datos en un dataset listo para un proyecto de machine learning.

Procesos realizados

- Carga de datasets en Pandas desde archivos `.csv`

- Limpieza y validación (nulos, descuentos anómalos, precios cero)

- Cálculo de `TotalPriceCalculated` desde `Quantity`, `Price` y `Discount`

- Detección de outliers vía IQR

- Análisis temporal: hora de venta + clasificación "entre semana / fin de semana"

- Feature engineering:

- Edad al momento de contratación (`EdadContratacion`)

- Años de experiencia al momento de venta (`AniosExperiencia`)

- Generación de `dataset\_modelado.csv` para modelos de ML

Transformaciones finales

- Variables categóricas transformadas con one-hot encoding (`TipoDia`)

- Variables numéricas conservadas (con opción a escalar si el modelo lo requiere)

- `TotalPriceCalculated` permanece sin modificar como variable objetivo

---

**Archivos Entregados**

- `.venv`

- `avance1\_consultas.sql`

- `avance2\_automatizacion\_y\_triggers.sql`

- `avance3\_notebook.ipynb`

- `data`

- ‘script\_sql’

- ‘.gitignore’

- `README.md`

- ‘requirements.txt’

---