



Ejercicio Formativo 3 Capítulo 6

Aspectos generales

- **Objetivos:** aplicar los contenidos de bases de datos de manera completa.
- **Entrega:** lunes 03 de noviembre a las 17:30 hrs. en el repositorio privado y respondiendo el ticket de salida.
- **Formato de entrega:** archivo Python Notebook (**E3.ipynb**) con el avance logrado para el ejercicio. El archivo debe estar ubicado en la carpeta **C6**. Utilice múltiples celdas de texto y código para facilitar el trabajo del cuerpo docente.
- **ULTRA IMPORTANTE:** todas las celdas utilizadas deben estar ejecutadas al momento de entregar el ejercicio, de modo que las salidas generadas sean visibles. En caso de no cumplir con esto, su entrega no será considerada como validación del ticket de salida.

Introducción

Este ejercicio consiste en construir una base de datos relacional a partir de archivos CSV, para luego realizar consultas SQL que permitan extraer información relevante. Para este propósito, trabajaremos con un esquema que incluye datos relacionados con productos, ventas, clientes y tiendas.

Antes de comenzar, se recomienda revisar el esquema conceptual y el contenido de los archivos CSV disponibles. Además, puede utilizar herramientas como DB Browser for SQLite para inspeccionar la base una vez construida.

Descripción de los datos

Considere una base de datos para una cadena de tiendas, la cual recopila información sobre productos, categorías, marcas, clientes, tiendas, empleados y transacciones.

El modelo de datos incluye relaciones claras entre tablas (por ejemplo: cada producto tiene una categoría y una marca; cada orden incluye ítems y está asociada a un cliente, empleado y tienda). Existen llaves primarias, llaves foráneas y restricciones de dominio evidentes que deben ser respetadas durante la creación del esquema.

Misiones

1. **Exploración inicial de los datos:** cargue los archivos CSV entregados y realice un análisis preliminar del contenido. Para cada archivo:

- Indique la cantidad de registros y columnas.
- Muestre las primeras filas y comente brevemente qué representa cada columna.
- Identifique posibles claves primarias o relaciones implícitas entre tablas.

La idea de esta pregunta es familiarizarse con el dominio de los datos y detectar patrones que guíen el modelado posterior.

2. **Propuesta de estructura relacional:** con base en la exploración anterior, proponga una estructura relacional razonable para organizar la información. Para cada tabla, indique:

- Nombre y propósito de la tabla.
- Lista de atributos con su tipo de dato estimado.
- Restricciones que deberían cumplirse: claves primarias, claves foráneas, unicidad o dominio (por ejemplo: campos que solo aceptan ciertos valores, campos no nulos).

No se requiere diagrama ni notación formal, pero la propuesta debe ser coherente y reflejar una buena interpretación de los datos.

3. **Implementación del esquema en SQLite:** cree una base de datos en SQLite (por ejemplo, `store.db`) e implemente las tablas y relaciones definidas en la pregunta anterior. Para esto, debe utilizar comandos `CREATE TABLE` que incluyan las restricciones mencionadas.

4. **Carga de datos y verificación:** cargue la información desde los archivos CSV a la base de datos.

Se espera que:

- Las tablas independientes se carguen primero.
- Se respeten todas las restricciones definidas.
- Se muestre la cantidad de registros cargados por tabla.

Si se producen errores al cargar los datos, explique brevemente qué ocurrió y cómo lo resolvió.

5. **Consultas:** para cada una de las siguientes preguntas, escriba una única consulta que entregue la respuesta, utilizando agentes basados en IA de la manera que prefiera (incluso para generar la consulta completa de manera automática). Una vez generada la consulta, escriba código en Python para validar los resultados de esta, de modo que los resultados coincidan siempre entre ambas versiones.

Las preguntas son las siguientes:

- a) Liste todos los productos junto con su nombre de categoría y marca.
- b) Muestre cuántos clientes hay registrados por ciudad.
- c) Calcule el precio promedio de los productos vendidos.
- d) Obtenga las 5 tiendas con más órdenes realizadas.
- e) Liste los empleados que han vendido productos de al menos 5 categorías distintas.
- f) Calcule el total vendido por tienda y mes.
- g) Encuentre los productos que han generado más ingresos que el promedio de su categoría.
- h) Liste los 10 clientes que más gastaron en productos que nunca se han vuelto a vender después.
- i) Encuentre los productos con más unidades en stock que ventas totales acumuladas.
- j) Calcule la variación mensual del total vendido por tienda, expresada como porcentaje respecto al mes anterior.
- k) Liste los productos cuyo precio promedio de venta ha bajado mes a mes durante al menos tres meses seguidos.
- l) Encuentre todos los clientes que han comprado productos de todas las categorías existentes.
- m) Calcule la secuencia más larga de días consecutivos con ventas en cada tienda.
- n) Determine los productos que fueron top 1 en ingresos mensuales durante al menos 6 meses no consecutivos.
- o) Encuentre el subconjunto mínimo de productos que representa al menos el 80% del total de ingresos.