

Ejercicio de postulación | Ingeniero/a de datos junior

Introducción

El presente ejercicio tiene como objetivo evaluar tus competencias técnicas en el ámbito de la ingeniería de datos. Se plantea un desafío ficticio, sin ningún fin comercial ni estratégico.

Desafío

El dataset adjunto ([Sample Analytics Dataset](#)), contiene información sobre cuentas, clientes y transacciones de una aplicación de servicios financieros. Tu misión es analizar estos datos para responder las siguientes preguntas:

1. ¿Cuál es el promedio, mínimo, máximo y desviación estándar del límite de las cuentas de usuarios?
2. ¿Cuántos clientes poseen más de una cuenta?
3. ¿Cuál es el monto promedio y el número de transacciones del mes de junio?
4. ¿Cuál es el id de cuenta con la mayor diferencia entre su transacción más alta y más baja?
5. ¿Cuántas cuentas tienen exactamente 3 productos y, además, uno de esos productos es "Commodity"?
6. ¿Cuál es el nombre del cliente que, en total entre todas sus cuentas, ha realizado la mayor cantidad de transacciones de tipo sell?
7. ¿Cuál es el usuario del cliente cuya cuenta tiene entre 10 y 20 transacciones de tipo "buy", y que presenta el promedio de inversión más alto por operación de este tipo?
8. ¿Cuál es el promedio de transacciones de compra y de venta por acción (campo "symbol")?
9. ¿Cuáles son los diferentes beneficios que tienen los clientes del tier "Gold"?
10. Obtener la cantidad de clientes por rangos etarios ([10–19], [20–29], etc.), que hayan realizado al menos una compra de acciones de "amzn". La edad debe calcularse como la diferencia entre la fecha de corte 2025-05-16 y el campo "birthdate".

Solución requerida

1. **Diseño del Data Warehouse:** Debes crear el diagrama de un data warehouse (DW) basado en el dataset proporcionado. Puedes utilizar el enfoque que consideres más apropiado según los requerimientos, entre ellos:
 - Modelado dimensional (Kimball)
 - One Big Table (OBT)
 - Data Vault
 - Modelado normalizadoSegún el diseño elegido, entregar una breve justificación del enfoque escogido.
2. **Implementación en SQLite:** Después de diseñar el modelo del DW, debes:
 - Crear las tablas, relaciones (si existen) y restricciones en SQLite
 - Desarrollar un script en el lenguaje de programación de tu elección que realice el proceso ETL:
 - **Extracción:** Lectura de los archivos JSON
 - **Transformación:** Ajuste de los datos al modelo DW creado
 - **Carga:** Inserción de los datos en las tablas SQLite

3. **Consultas SQL:** Finalmente, debes desarrollar las consultas SQL que respondan a cada uno de los 10 requerimientos planteados.

Consideraciones importantes

- Solo es necesario almacenar los campos del dataset que sean utilizados para responder a los requerimientos.
- No redondees los campos almacenados en la base de datos.
- Queda a tu criterio convertir campos de texto a valores numéricos si lo consideras apropiado.
- Todas las consultas SQL deben realizar el redondeo de valores numéricos a 2 decimales.
- Es obligatorio utilizar control de versiones (Git y GitHub).
- Para dudas adicionales, puedes consultar el [documento de preguntas frecuentes](#).

Entregables

- Archivo con el diagrama realizado junto a la justificación del diseño escogido en formato PDF.
- Archivo .db con la base de datos SQLite junto a los datos ya cargados.
- Archivo .sql con las consultas que respondan los requerimientos.
- Código fuente en repositorio **privado** de GitHub, y acceso al usuario **postulaciones-bsale**
- Enviar mediante [este formulario](#) la siguiente información:
 - ☐ CV en formato PDF.
 - ☐ Diagrama del DW en formato PDF
 - ☐ Archivo con la base de datos SQLite
 - ☐ Archivo con las consultas SQL
 - ☐ URL del repositorio de GitHub.
 - ☐ Respuestas a los requerimientos.

¡Mucho éxito! 🤖