

Prueba técnica analista de datos - Puntored

Duración: 24 horas a partir de la fecha de envío

Criterios de Evaluación

- Precisión de las respuestas teóricas.
- Claridad y eficiencia del código SQL.
- Precisión, eficiencia y claridad en la presentación de la visualización de datos.

Modo de entrega

Cargue los archivos pertinentes al desarrollo de la prueba a un repositorio público de GitHub y comparta la URL. (15 puntos)

Sección 1: Preguntas Teóricas (11 puntos)

SQL (4 puntos)

1. ¿Cuál es la diferencia entre un JOIN y un UNION en SQL? Proporcione un ejemplo de cada uno. (1 punto)
2. Explique qué es una consulta CTE (Common Table Expression) y mencione un caso de uso. (1 punto)
3. ¿Qué es la cláusula HAVING en SQL y en qué se diferencia de WHERE? (1 punto)
4. Dado el siguiente conjunto de datos en una tabla llamada **ventas**, escriba una consulta SQL para obtener el total de ventas por cliente, incluyendo sólo aquellos con un monto total superior a 250. (1 punto):

id_venta	fecha	cliente_id		monto
1	2025-03-01	101	200	
2	2025-03-02	102	150	
3	2025-03-02	101	300	

Python (3 puntos)

1. ¿Cuál es la diferencia entre una lista y un diccionario en Python? Dé un ejemplo de cada uno. (1 punto)
2. ¿Qué librerías en Python se utilizan comúnmente para análisis de datos? Mencione al menos tres. (1 punto)
3. ¿Qué salida generará el siguiente código en Python? (1 punto)

Python

```
data = [5, 3, 9, 1]
print(sorted(data)[-2])
```

Visualización de datos - Tableau (4 puntos)

1. ¿Cuál es la diferencia entre una **medida** y una **dimensión** en Tableau? (1 punto)
2. ¿Qué tipo de gráfico es más adecuado para comparar porcentajes de una variable categórica? (1 punto)
3. ¿Cómo se puede crear un **filtro dinámico** en Tableau para que el usuario pueda seleccionar un rango de fechas? (1 punto)
4. ¿Que es un archivo **hyper** en tableau, y en que caso de uso se puede aplicar? (1 punto)

Sección 2: Prueba práctica SQL (55 puntos)

Dada la siguiente tabla **empleados**:

id	nombre	departamento	salario	fecha_contratacion
1	Ana	Ventas	5000	2020-01-15
2	Luis	IT	6000	2019-08-21
3	María	Ventas	5500	2021-05-10
4	Juan	IT	7000	2018-11-03

Escriba una consulta SQL para obtener el promedio de salario por departamento, excluyendo empleados contratados después del 1 de enero de 2020. (5 puntos).

Tienes las siguientes tablas en una base de datos:

clientes

id (INT, PRIMARY KEY)

nombre (VARCHAR) apellido

(VARCHAR) **ventas** id

(INT, PRIMARY KEY)

```
cliente_id (INT, FOREIGN  
KEY a clientes.id)  
producto (VARCHAR) fecha  
(DATE) monto (DECIMAL)
```

Escribe una consulta para obtener los 5 clientes con mayor monto total de ventas en los últimos 6 meses. (5 puntos)

Escribe una consulta para calcular el ticket promedio de ventas por cliente en el último año. (5 puntos)

Escribe una consulta para obtener el nombre completo de los clientes y su monto total de ventas. (10 puntos)

Escribe una consulta para obtener el ingreso promedio de ventas por mes. (10 puntos)

Escribe una consulta para calcular el ranking de clientes por ventas en el último año. (10 puntos)

Escribe una consulta para calcular el total de ventas por cliente y luego selecciona solo los clientes cuyo total de ventas es superior al promedio general. (10 puntos)

Sección 3: Python (30 puntos)

Dado el archivo CSV llamado `datos_ventas.csv` realiza las siguientes acciones:

1. Carga y exploración básica

- Carga el dataset en un DataFrame.
- Muestra los primeros 5 registros.
- Describe brevemente las columnas y sus tipos.
- ¿Hay valores nulos? ¿Cómo los manejarías?

2. Limpieza de datos

- Corrige los tipos de datos si es necesario (por ejemplo, fechas).
- Elimina columnas irrelevantes si aplica.
- Crea una columna nueva calculada, por ejemplo: $Ganancia = Ingreso - Costo$.

3. Análisis exploratorio

- ¿Cuáles son los 5 productos más vendidos por unidades?
- ¿Cuál es la categoría que más ingresos genera?
- ¿Hay diferencias significativas de ventas por región o canal?

4. Visualización

- Crea una visualización que muestre la evolución de las ventas mensuales.

- Muestra un gráfico comparativo entre ingresos y costos por categoría.
- Usa un gráfico de dispersión (scatter) para mostrar la relación entre el número de ventas y las ganancias.

5. Insight y comunicación

- En máximo 5 líneas, escribe dos hallazgos interesantes del análisis.
- Propón una acción que el negocio podría tomar basada en esos hallazgos.

Sección 4: Visualización de datos - Tableau (30 puntos)

Usando un dataset con ventas por ciudad (*ciudad*, *fecha*, *monto_venta*), cree un dashboard en Tableau que incluya:

- Un gráfico de líneas mostrando la evolución de las ventas en el tiempo.
- Un mapa geográfico con las ventas por ciudad.
- Un filtro interactivo para seleccionar un rango de fechas.

Entrega: Comparta una captura de pantalla del dashboard y un archivo *.twbx* con su visualización.