

Prueba técnica Científico 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
- Estructura y documentación del código Python
- Capacidad de interpretar los resultados del modelo (ejercicio práctico)

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 (10 puntos)

Python (3 puntos)

1. Explica la diferencia entre una lista y un conjunto en Python. ¿En qué situaciones usarías cada uno? (1 puntos)
2. ¿Qué hace el método `apply()` en un DataFrame de pandas? Da un ejemplo de uso. (1 puntos)
3. ¿Cuál es la diferencia entre `deepcopy()` y `copy()` en Python? Explica con un ejemplo. (1 puntos)

SQL (3 puntos)

4. Dado un esquema de base de datos con una tabla `ventas(id, cliente_id, fecha, monto)`, escribe una consulta SQL para obtener el total de ventas por cliente en el último mes. (1 puntos)
5. Explica la diferencia entre un `INNER JOIN` y un `LEFT JOIN`. Da un ejemplo de cada uno. (1 puntos)
6. ¿Qué es la cláusula `HAVING` en SQL y en qué se diferencia de `WHERE`? (1 puntos)

Machine Learning (4 puntos)

7. ¿Cuál es la diferencia entre un modelo supervisado y no supervisado? Da un ejemplo de cada uno. (2 puntos)
8. Menciona tres métricas de evaluación para modelos de clasificación y explica cuándo usar cada una. (2 puntos)

Sección 2:

Prueba práctica SQL (50 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)

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)

Prueba práctica python y machine learning (45 puntos)

Tienes un dataset con información de clientes de un banco. Debes predecir si un cliente hará un default en su crédito.

Dataset: `clientes.csv` (columnas)

- `edad` (int)
- `ingreso_mensual` (float)
- `historial_crediticio` (int, 0-5)
- `default` (int, 0 o 1)

Ejercicio

1. Carga el dataset y realiza una limpieza de datos básica. (5 puntos)
2. Divide los datos en entrenamiento y prueba (80%-20%). (10 puntos)
3. Entrena un modelo de clasificación (indicar cuál modelo se usó y por qué). (15 puntos)
4. Evalúa el modelo con una métrica adecuada y justifica su elección. (15 puntos)