

## Reto técnico – Data Engineer

### Objetivo del reto

Evaluar tu capacidad para:

- Tomar datos crudos
- Limpiarlos y transformarlos
- Modelarlos y cargarlos en una base de datos
- Exponer valor a través de SQL
- Resolver el problema de forma independiente

### Contexto

Eres Data Engineer en una plataforma de e-commerce. Recibes datos operativos en archivos planos que deben ser preparados para análisis.

### Datos de entrada

Revisar archivos JSON anexados.

### Lo que debes hacer

1. Transformación de datos  
Implementar un script (Python u otro lenguaje) que:
  - Limpie y normalice los datos
  - Convierta fechas a un formato estándar
  - Convierta montos a valores numéricos
  - Normalice emails (lowercase, trim)
  - Relacione órdenes con clientes usando email
  - Identifique y descarte registros inválidos, documentando el criterio
2. Modelo de datos  
Diseñar un modelo simple en PostgreSQL con dos tablas:
  - customers
  - orders

El modelo debe incluir:

- Primary keys
- Foreign key correcta entre órdenes y clientes
- Tipos de datos adecuados

### 3. Carga de datos

- Crear las tablas (DDL)
- Insertar los datos transformados
- El proceso debe poder ejecutarse más de una vez sin duplicar datos

### 4. SQL de valor

Entregar tres consultas SQL:

- Total de ventas por cliente
- Número de órdenes por país
- Ticket promedio

### Entregables

No es obligatorio usar GitHub ni generar diagramas complejos.

Se espera recibir:

- Script ETL (un archivo)
- Script SQL con DDL y queries
- README corto explicando:
  - Cómo ejecutar la solución
  - Decisiones técnicas relevantes
  - Supuestos realizados
- Logging básico
- Validaciones simples (por ejemplo, órdenes sin cliente)
- Uso de librerías como pandas o SQLAlchemy

En caso de apoyarte del ChatGPT o cualquier modelo de AI, por favor, compartenos el prompt 😊

¡Buena suerte!