Caso de Estudio – Data Engineer (Validación Alertas Smart)

# Contexto

La empresa realiza auditorías de campo en tiendas de cadenas de supermercados. Como parte del proceso, se registran fotografías por cada categoría auditada, junto con dos fotos obligatorias de inicio y fin de la auditoría. El área de operaciones ha venido desarrollando un sistema de alertas inteligentes para validar la calidad y completitud de esta información.

# Objetivo del Caso

Diseñar un flujo de validación automatizado utilizando Python, que verifique si las auditorías cumplen con el número mínimo de fotos esperadas según el cliente auditado y las categorías asociadas. El resultado debe mostrarse en un dashboard y además generarse un archivo Excel con las alertas encontradas y enviarse por correo automáticamente desde el notebook.

# Modelo de Datos y Generación Simulada

Deberá construir su propia base de datos simulada, utilizando un LLM (como ChatGPT o Copilot) u otra técnica generativa. Los datos deben ser ficticios pero representativos del contexto planteado. El modelo mínimo debe incluir:

* • Clientes: ID y nombre del cliente auditado.
* • Categorias: ID y nombre de las categorías posibles.
* • Cliente\_Categoria: tabla que indique qué categorías audita cada cliente.
* • Auditorias: ID, fecha, tienda, auditor responsable y cliente al que pertenece.
* • FotosCargadas: registro de fotos subidas por auditoría y categoría.

Por ejemplo:  
• El cliente 'Cruz Verde' audita 3 categorías → por lo tanto, se esperan mínimo 5 fotos (3 categorías + 2 de control).  
• El cliente 'Embonor' audita 5 categorías → se esperan mínimo 7 fotos.  
Una auditoría con menos fotos que el mínimo esperado según el cliente se considera un caso de alerta.

# Herramientas permitidas

Se puede utilizar cualquier entorno o herramienta a su alcance, incluyendo Google Colab, Jupyter Notebooks, Visual Studio Code u otros entornos de desarrollo compatibles con Python. Se valorará la capacidad de estructurar una solución funcional y bien presentada, más allá del entorno específico.

# Requerimientos técnicos

1. Crear un modelo de datos realista y generar datos simulados de acuerdo con el contexto.

2. Leer y procesar los datos usando Python.

3. Calcular la cantidad mínima de fotos requeridas por auditoría según el cliente y sus categorías.

4. Identificar auditorías con número de fotos insuficiente.

5. Generar una tabla de alertas con auditoría, cliente y número de fotos faltantes.

6. Mostrar el resultado en un dashboard o visualización de datos.

7. Enviar por correo un archivo Excel con las alertas desde el mismo notebook.

# Entregables

• Notebook en Python con todo el flujo implementado.

• Captura o link del dashboard generado.

• Archivo Excel generado automáticamente con las alertas.

• Breve explicación del enfoque utilizado.

# Criterios de Resultado

• Calidad del modelo de datos y realismo de los datos generados.

• Correcta lectura y manipulación de datos con Python.

• Lógica de validación implementada correctamente.

• Automatización del envío por correo.

• Claridad y visualización del dashboard.

• Presentación oral y capacidad de comunicar el flujo propuesto.

# Presentación

La presentación del caso se realizará en una sesión individual de aproximadamente 30 minutos por candidato, donde se espera una explicación del flujo, código utilizado, visualizaciones y resultados obtenidos.