

# Ejercicio práctico: Limpieza y transformación de un dataset biosanitario

Tema 3: Aplicación de aspectos avanzados de procesamiento de datos en Python

**Objetivo práctico:** Aplicar técnicas de limpieza y preprocesamiento de datos para preparar un conjunto de datos para su análisis.

## Descripción

En este ejercicio, vais a trabajar con un conjunto de datos sintéticos biosanitarios contenido en un único archivo CSV titulado `pacientes_sintetico.csv`. Este dataset simula registros de pacientes con columnas como edad, presión arterial, niveles de colesterol y frecuencia cardíaca máxima. El archivo ha sido diseñado para incluir imperfecciones comunes: valores nulos, duplicados y columnas numéricas que necesitan normalización.

Vuestra tarea consiste en usar este archivo CSV y realizar las siguientes actividades utilizando Python y bibliotecas como `pandas`, `numpy` y `sklearn`:

1. **Imputar valores nulos:**

Identificad las columnas con valores faltantes (por ejemplo, "colesterol" o "presión arterial") y rellenadlos utilizando el promedio (para datos continuos) o la mediana (para datos con distribuciones no uniformes).

2. **Eliminar duplicados:**

Buscad y eliminad cualquier fila duplicada en el dataset para asegurar que cada paciente esté representado una sola vez.

3. **Normalizar una columna numérica:**

Elegid una columna numérica del CSV (como "edad" o "colesterol") y aplicad la técnica de *min-max scaling* para escalar sus valores al rango `[0, 1]`.

## Instrucciones

- Usad el archivo `pacientes_sintetico.csv` que se os proporcionará (o generadlo con el código de ejemplo que se incluye más abajo).
- Cargad el archivo en Python, limpiadlo y transformadlo.
- Documentad vuestros pasos en el código con comentarios que expliquen qué hacéis y por qué.

## Entregable

- Un script de Python con el código completo que realice las tres tareas a partir del archivo CSV.
- Un resumen breve (puede ser en comentarios en el código) donde indiquéis:
  - Qué columnas tenían valores nulos y cómo las tratasteis.
  - Cuántos duplicados eliminasteis.
  - Qué significa un valor normalizado (por ejemplo, 0 o 1) en la columna que elegisteis.

## Contexto

Imaginad que sois analistas de datos en un hospital preparando un archivo CSV con información clínica simulada para un estudio piloto sobre factores de riesgo cardiovascular. La calidad de los datos es clave para obtener resultados fiables, ¡así que aseguraos de que el dataset esté bien limpio y listo para el análisis!