

# Taller: Verificación y Limpieza de DataFrames

Este taller está diseñado para practicarse en Google Colab utilizando pandas, numpy y matplotlib o seaborn. El objetivo es que los estudiantes aprendan a detectar, eliminar y rellenar valores faltantes en un DataFrame real. Tiempo estimado: 40 minutos.

Dataset: 5 columnas (ID, Edad, Ingreso\_Mensual, Ciudad, Satisfaccion), 70 filas con datos faltantes en las columnas numéricas. Archivo proporcionado: dataset\_ejercicios\_limpieza.csv

## Ejercicios:

- 1 Ejercicio 1 – Inspección de valores faltantes: 1. Cargar el dataset. 2. Usar `df.info()` y `df.isnull().sum()` para detectar valores nulos. 3. Graficar un barplot con la cantidad de nulos por columna.
- 2 Ejercicio 2 – Eliminación de filas incompletas: 1. Crear una copia del DataFrame. 2. Eliminar filas con cualquier valor nulo usando `dropna()`. 3. Comparar el número de filas antes y después. 4. Graficar un histograma de la columna Edad después de la limpieza.
- 3 Ejercicio 3 – Relleno con la media o mediana: 1. Rellenar los nulos de Edad con la media. 2. Rellenar los nulos de Ingreso\_Mensual con la mediana. 3. Verificar que ya no queden nulos. 4. Graficar un boxplot de Ingreso\_Mensual antes y después.
- 4 Ejercicio 4 – Relleno condicional por grupo: 1. Agrupar por Ciudad. 2. Rellenar los nulos de Satisfaccion con la media por ciudad usando `transform` y `fillna`. 3. Verificar los nulos restantes y graficar un countplot de Satisfaccion por ciudad.
- 5 Ejercicio 5 – Documentación de resultados: 1. Crear un resumen en markdown describiendo los métodos usados. 2. Incluir número de filas/columnas antes y después. 3. Guardar el DataFrame limpio en `dataset_limpio.csv`.