

## PRIMER TALLER - MINERIA DE DATOS 2026-1

### Universidad Distrital Francisco José De Caldas



Seleccionar un data set o crear uno, mínimo de 1000 registros (Excel, CSV, etc) (por ejemplo de la pagina Kaggle), este debe contener inconsistencias comunes como: valores nulos, duplicados, tipos incorrectos, outliers, etc.

#### Desarrollo:

1. Describir los atributos y cuál es el objetivo del análisis (problema a resolver)
2. Cargar el dataset en Python
3. Realiza las siguientes tareas:
  - a) Identificar y eliminar las filas duplicadas
  - b) Analizar columnas con valores faltantes, proponer una solución
  - c) Corregir cualquier error de tipo de dato (por ejemplo, fechas o números que estén como texto, etc).
  - d) Detectar y manejar posibles outliers en una variable numérica.
  - e) Investigar y aplicar una técnica de normalización o estandarización (al menos en una columna numérica)

#### CONDICIONES DE ENTREGA:

1. Código Python completo.
2. Incluir análisis gráfico.
3. Informe con explicación de cada paso realizado, justificación de técnicas utilizadas y breve análisis final de cómo quedó el dataset después del preprocesamiento.
4. Exposición en clase con la síntesis del desarrollo del taller (máximo 10 min). Deben participar todos los integrantes del grupo.

**Correo:** [mtelloc@udistrital.edu.co](mailto:mtelloc@udistrital.edu.co) con:

**Asunto:** Taller 1 – Minería de datos

**Detalle:** Nombres y códigos

**Fecha de entrega y sustentación:** sábado 28 de febrero de 2026

**Entrega:** Máximo 2 estudiantes

Esta calificación corresponde al 10% del primer corte (2026-1)