

Tarea 1

Objetivo

Escribe un script de Python (.py) que limpie el conjunto de datos proporcionado (dataset_uncleaned.csv) para que coincida con la versión limpia (dataset_v1.csv) que mostré en clase.

No debes enviar el archivo limpio, sino solo el script de Python (ej., clean_data.py) que, al ejecutarse, genera la versión limpia.

Instrucciones:

1. Descarga y examina el archivo dataset_uncleaned.csv
2. Escribe un script en python que:
 - a. Limpie los datos paso a paso
 - b. Produzca un conjunto de datos limpio dataset_clean.csv que coincida con la estructura y los valores del archivo de referencia que usamos en la clase.
3. El script debe:
 - a. Ser completamente reproducible (lo ejecutaré yo misma)
 - b. Incluya comentarios explicando tu lógica
 - c. Use funciones donde sea apropiado
 - d. Guarde el conjunto de datos limpio como dataset_clean.csv (misma estructura y datos que el archivo dataset_v1.csv)

Pasos de limpieza a tener en cuenta (consejos):

Necesitarás:

- Tratar valores faltantes, caracteres no válidos y patrones de texto extraños
- Convertir columnas a tipos de datos correctos (p. ej., fechas, números, conversión de hexadecimal a entero)
- Normalizar entradas inconsistentes (p. ej., mayúsculas y minúsculas, ruido en las palabras, formatos de fecha incorrectos)
- Limpiar e imputar valores categóricos mediante lógica de grupos
- Detectar y corregir valores numéricos no válidos mediante grupos de clientes
- Reconstruir columnas como Credit_History_Age y gestionar columnas codificadas como SSN
- Reemplazar valores de marcador de posición como !@9#%8, nan, etc.