

## Tarea: Proceso de Limpieza de Datos

### Objetivo

Aplicar correctamente las técnicas de limpieza y preparación de datos vistas en clase, asegurando la comprensión del flujo completo: detección, corrección y validación de inconsistencias en una base de datos real.

---

### Instrucciones Generales

1. Selecciona una base de datos real (puede ser del tema que prefieras, pero debe tener al menos 200 registros y 8 columnas).
  2. Realiza todo el proceso de limpieza en **Python (pandas)**, documentando tu trabajo con **comentarios** y **celdas Markdown** descriptivas.
  3. No se permite el uso de `dropna()` como solución principal.
    - Se penalizará su uso, ya que el objetivo es limpiar y recuperar información, no eliminarla.
  4. Guarda tu archivo `.ipynb` con el código completo y súbelo a un repositorio en **GitHub** junto con:
    - La base de datos original (sucia)
    - La base de datos limpia (final)
- 

### Estructura del Reporte (en formato PDF)

Tu reporte debe incluir lo siguiente:

#### 1. Portada

- Nombre completo
  - Materia: *Introducción a la Ciencia de Datos*
  - Nombre del profesor
  - Fecha de entrega
  - Link al repositorio de GitHub
-

## 2. Descripción inicial de la base de datos

Incluye un bloque Markdown donde expliques:

- Fuente o contexto de la base de datos.
  - Descripción general del contenido.
  - Significado de cada columna.  
(Puedes usar una tabla para mayor claridad).
- 

## 3. Proceso de Limpieza (con evidencias)

Incluye capturas de pantalla y fragmentos de código donde muestres los pasos realizados:

- Revisión de datos faltantes (isnull(), info(), etc.)
- Detección y manejo de duplicados
- Corrección de valores atípicos o inconsistentes
- Detección y corrección de palabras extrañas o mal escritas
- Traducción de textos al español (cuando aplique)
- Cambio de nombres de columnas para mayor claridad o consistencia
- Conversión de tipos de datos (numéricos, fechas, etc.)
- Validación final mostrando que la base quedó limpia y coherente

Cada paso debe estar claramente documentado con explicación, código y resultado.

---

## 4. Conclusiones

Redacta brevemente:

- Qué problemas principales presentaba la base.
  - Qué técnicas aplicaste para solucionarlos.
  - Qué aprendiste del proceso.
-

## Entrega

Debes entregar:

1. Un archivo **PDF** con el reporte completo (máximo 10 páginas).
2. El **link al repositorio en GitHub**, que debe incluir:
  - La base de datos original
  - La base de datos limpia
  - El archivo .ipynb con el código

---

## Penalizaciones

- Uso del comando dropna() → -2 puntos
- No incluir capturas de pantalla → -1 punto
- No entregar el link de GitHub o repositorio incompleto → -2 puntos
- No documentar el significado de las columnas → -1 punto

---

## Criterios de Evaluación

Criterio	Ponderación
Claridad del reporte y redacción	15%
Documentación del proceso (Markdown y comentarios)	25%
Corrección y limpieza efectiva de datos	35%
Entrega completa (bases, código y PDF)	15%
Presentación y orden general	10%