**Laboratorio 10**

# Sesión # 10 Componente Práctico

**Título del Laboratorio:** Aplicación del uso de la limpieza de los datos en Python.

**Duración:** 2 horas

**Objetivos del Laboratorio:** *Afianzar los conocimientos y manejo básico en Python sobre la Limpieza de los datos con ejercicios prácticos planteados.*

# Materiales Necesarios:

1. *Computador con acceso a internet.*
2. *Colocarlo en el repositorio de Github*
3. *Ampliar el conocimiento con el curso de datos en AWS y Cisco.*
4. *Python en línea: Google colab*.

# Estructura del Laboratorio:

**Parte 1**

Los ejercicios de práctica son escenarios reales que se deberás realizar el código usando las bibliotecas en Python para la manipulación de los datos y limpieza de los datos, pegar la captura de pantalla del resultado.

# Ejercicio de práctica 1.

* 1. **Salud Digital y Análisis de Datos de Pacientes**

Se tiene un conjunto de datos sobre pacientes que incluye nombre, edad, fecha de ingreso y resultados de pruebas. Algunos registros están incompletos o contienen errores.

# Datos:

* + - * Nombre: Carlos, Ana, Luis,Pedro, María, None
      * Edad: 34, 29, None, 45, 25, 33
      * Fecha Ingreso: 2021-01-15, 2020/03/12, 2022-07-01,2021/12/01,2021-05-20, 2020-03-12
      * Resultado Prueba: Positivo,Negativo, Positivo,Positivo, Negativo,Negativo.

# Solución:

# Texto Descripción generada automáticamente

Texto

Descripción generada automáticamente

* 1. **Comercio Electrónico y Recomendaciones Personalizadas**

Se tiene un conjunto de datos sobre transacciones de clientes que incluye el ID del cliente, el producto comprado, la cantidad, el precio y la fecha de la transacción. Algunos registros tienen errores en los precios y fechas.

# Datos:

* + - * **Cliente ID:** 1, 2, 3, 4, 5
      * **Producto:** A, B, C,D, E
      * **Cantidad:** 10, 20, None, 5, 15
      * **Precio:**100, 200$,300, cuatrocientos,500
      * **Fecha Transacción:** 2023-01-15,2023/02/20,2023-03-25,2023/04/30.2023-05-10

# Texto Descripción generada automáticamenteSolucion:

# Texto Descripción generada automáticamente

A. **Escenario:** eres un analista de datos en una pequeña empresa que está recopilando información sobre sus clientes.

El equipo de ventas te entrega un archivo con los siguientes datos:

* Nombre de los clientes
* Edad
* Correo electrónico
* País
* Género
* Número de visitas a la tienda

Sin embargo, cuando abres el archivo, te das cuenta de que hay **problemas** en los datos:

1. **Valores faltantes**: Algunos clientes no tienen información sobre su edad o correo electrónico.
2. **Duplicados**: Algunos clientes han sido registrados más de una vez.
3. **Inconsistencias**: Hay diferentes formas de escribir el género ("M", "Masculino", "F", "Femenino").
4. **Errores tipográficos**: El nombre de algunos países está en minúsculas o mal escrito.
5. **Valores inválidos**: Hay una edad registrada que parece imposible (200 años).

# Datos iniciales (archivo ficticio):

* + - * **Nombre,Edad,Email,Pais,Genero,Visitas**
      * Juan,25,juan@mail.com,colombia,M,5
      * Maria,30,maria@mail.com,Colombia,F,7
      * Pedro,,pedro@mail.com,Mexico,H,6
      * Ana,29,ana@mail.com,Argentina,F,4
      * Luisa,28,luisa@mail.com,Colombia,Femenino,8
      * Carlos,200,carlos@mail.com,Brazil,M,2
      * Juan,25,juan@mail.com,colombia,
    - **Objetivo:** Tu objetivo es **limpiar los datos** para que estén listos para el análisis, aplicando las siguientes técnicas:

# Rellenar valores faltantes.

1. **Eliminar duplicados**.
2. **Estandarizar los valores de las columnas** (como el género y el país).
3. **Corregir valores inválidos** (como la edad de 200 años).
   * + **Solucion:**

Texto

Descripción generada automáticamente

Imagen de la pantalla de una computadora

Descripción generada automáticamente con confianza baja