**Laboratorio 9**

# Sesión # 9 Componente Práctico

**Título del Laboratorio:** aplicación del uso de la calidad de los datos en Python.

**Duración:** 2 horas

**Objetivos del Laboratorio:** *Afianzar los conocimientos y manejo básico en Python sobre la calidad 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**

El primer ejercicio de práctica deberás realizar el código usando las bibliotecas en Python para la manipulación de los datos y aplicar la calidad de los datos, pegar la captura de pantalla del resultado y copiar el código según el caso.

# Ejercicio de práctica 1.

**Aseguramiento de la Calidad de los Datos**

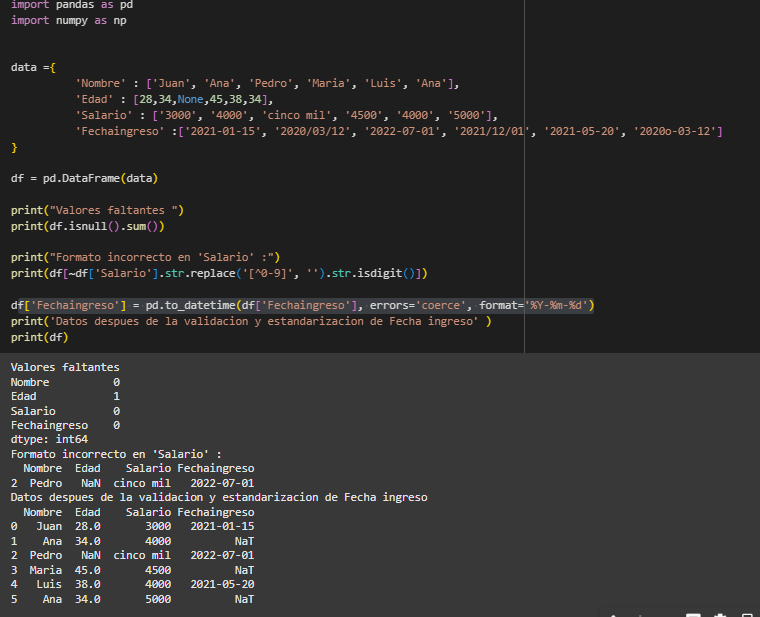
* 1. Se tiene un conjunto de datos sobre empleados que incluye su nombre, edad, salario y fecha de ingreso. Se requiere implementar algunos pasos de aseguramiento de la calidad de los datos, como la verificación de valores faltantes y la validación de formatos de datos.

# Datos

* + **Nombre:** Juan,Ana, Pedro,María,Luis,Ana
  + **Edad:**28, 34, None, 45, 38, 34.
  + **Salario:**3000, 4000, cinco mil, 4500, 4000, 5000
  + **Fecha Ingreso:**2021-01-15, 2020/03/12, 2022-07-01,2021/12/01,2021-05-20, 2020-03-12

# Se requiere:

* + 1. Identificar y contar los valores faltantes en cada columna.
    2. Verificar el formato de la columna 'Salario' para asegurar que todos los valores sean numéricos.
    3. Validar y estandarizar el formato de la columna 'Fecha Ingreso' a YYYY-MM-DD.

**SOLUCION:**

* 1. Se tiene un conjunto de datos sobre empleados que incluye su nombre, edad, salario, puesto de trabajo y fecha de ingreso. Se requiere implementar aseguramiento de la calidad de los datos, verificando la completitud de los datos, la validez de los salarios y la consistencia de las fechas de ingreso.

# Datos:

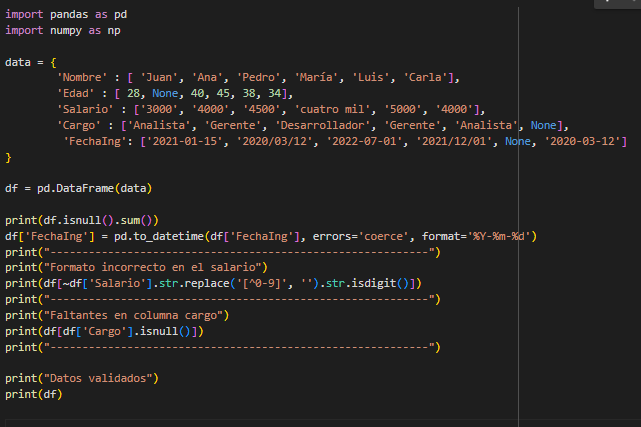
* + **Nombre:** Juan, Ana, Pedro, María, Luis, Carla
  + **Edad:** 28, None, 40, 45, 38, 34
  + **Salario:** 3000, 4000, 4500, cuatro mil, 5000, 4000
  + **Cargo:** Analista, Gerente, Desarrollador, Gerente, Analista, None
  + **Fecha Ingreso:** 2021-01-15, 2020/03/12, 2022-07-01, 2021/12/01, None, 2020-

03-12

# Se requiere:

* + 1. Identificar y contar los valores faltantes en cada columna.
    2. Verificar el formato de la columna "Salario" para asegurarse de que todos los valores sean numéricos.
    3. Validar y estandarizar el formato de la columna "FechaIngreso" a YYYY-MM-DD.
    4. Verificar que no haya valores faltantes en la columna "Cargo".

# Solución:

****

**Texto

Descripción generada automáticamente**

**Parte 2**

El segundo ejercicio de práctica deberás realizar el código usando las bibliotecas en Python para la manipulación de los datos y aplicar la calidad de los datos, pegar la captura de pantalla del resultado y copi ar el código según el escenario.

# Ejercicio de práctica 2.

* 1. **Se requiere validar los datos:** Precios de Productos, se tiene una lista de productos con los respectivos precios y se necesita validar que ningún precio sea negativo o nulo.

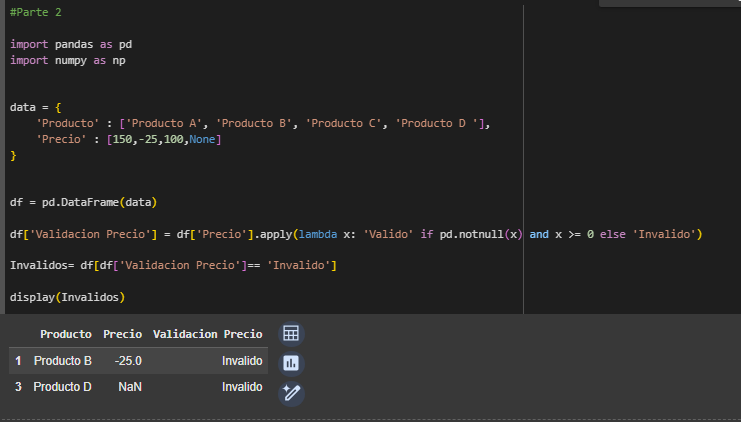
# Se requiere:

* + 1. Crear un DataFrame con productos y precios.
    2. **Producto:** Producto A, Producto B,Producto C,Producto D.

**3. Precio:**150, -25, 100, None.

1. Validar que todos los precios sean mayores o iguales a cero y no nulos.
2. Mostrar los productos con precios inválidos.

* **Resultado esperado:** Una tabla que muestra qué productos tienen precios incorrectos.



* 1. **Se requiere validar los datos duplicados:** Clientes en una Base de Datos, se tiene una lista de clientes donde algunos registros tienen el mismo ID y necesitas identificarlos.

# Se requiere:

* + 1. Crear un DataFrame con información de clientes.

**2. ID Cliente:**1, 2, 3, 2, 4.

1. **Nombre:** Ana, Luis, María, Luis, Carlos
2. Verificar si hay duplicados en la columna de ID del cliente.
3. Mostrar los registros duplicados.

* **Resultado esperado:** Una tabla que contiene solo los clientes que tienen el mismo ID, facilitando su corrección.

# Solución:

**Texto

Descripción generada automáticamentew**

* 1. **Se requiere tener seguridad en los datos:** Encriptación y Desencriptación de Datos Sensibles, se tiene un número de tarjeta de crédito que deseas encriptar para protegerlo y luego desencriptarlo para el uso.

# Se requiere:

* + 1. Crear una cadena con datos sensibles (un número de tarjeta 1234-5678-9101- 1121).
    2. Encriptar esos datos utilizando una clave generada.
    3. Desencriptar los datos y mostrar ambos (encriptado y desencriptado).
* **Resultado esperado:** Ver el número de tarjeta encriptado y luego desencriptado, demostrando cómo proteger la información sensible.

# Solución:

# 