# **Experiment 1: Data Loading, Cleaning, and Preprocessing**

## **Aim:**

The goal of this experiment is to understand how to load a dataset, clean it, and perform basic preprocessing steps to prepare it for analysis

### **1. Importing Required Libraries**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

**Explanation:**We import numpy (np) is used for numerical operations, pandas (pd) for data manipulation, matplotlib.pyplot (plt) for plotting, and seaborn (sns) for statistical data visualization.

### **2. Loading the Dataset**

df=pd.read\_csv('/content/gold.csv')

**Explanation:**We use pd.read\_csv() to load a CSV file containing Gold data.

### **3. Understanding the Data**

df.info()

**Explanation:**df.info() provides a summary of the DataFrame, including the number of non-null values, column data types, and memory usage. It helps in understanding the dataset's structure and identifying missing values.

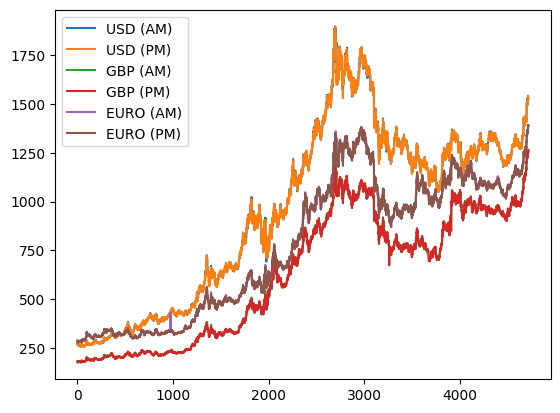
### **4. Displaying the First Few Rows**

df.head()

**Explanation:**df.head() shows the first five rows of the dataset, giving us an overview of the available columns and their values.

### **5. Visualizing Outliers**

df.plot()



### **7. Adding new column to the dataset**

df['Return'] = df['USD (PM)'].pct\_change()\*100

df['Lagged\_Return'] = df.Return.shift()

df = df.dropna()

### **8. Visualizing the Distribution**

df.plot()

