Linear Regression Practice: Gold Price Dataset

This exercise is based on the Gold Price dataset. The goal is to build and evaluate a Linear Regression model to predict gold prices based on historical features. Follow each step carefully to practice data preprocessing, exploration, model building, evaluation, and interpretation.

# Part A: Questions

* 1. Load the dataset and display its shape, column names, and first 5 rows.
* 2. Check for missing values and duplicates. Clean the dataset if necessary.
* 3. Perform exploratory data analysis (EDA): show descriptive statistics and visualize distributions of key features.
* 4. Create a correlation heatmap to identify which features are highly correlated with the target variable (Gold Price).
* 5. Select independent variables (e.g., Open, High, Low, Volume) and target variable (Close/Price).
* 6. Split the dataset into training (80%) and testing (20%) sets.
* 7. Train a Linear Regression model on the training set.
* 8. Evaluate the model using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R² score on the test set.
* 9. Plot Actual vs. Predicted gold prices to visualize model performance.
* 10. Interpret the model coefficients. Which features have the most significant impact on gold price prediction?