Supervised Learning - Linear Regression With Error Metrics:

The **`microsoft_stock.csv`** file contains historical stock data for Microsoft Corporation (MSFT). This data is often used in financial and machine learning projects to analyze and predict stock prices.

Here's a brief explanation of the typical columns you might find in this file:

1. Date → The date on which the trading took place. The format is usually `YYYY-MM-DD`.
2. Open → The opening price of the stock on that particular day.
3. High → The highest price of the stock during the trading day.
4. Low → The lowest price of the stock during the trading day.
5. Close → The closing price of the stock on that particular day. This is often the target variable for prediction.
6. Volume → The number of shares traded during the day.

Exercise instructions:
Use the **'Microsof_stock.csv'** file provided to you and execute the following instructions:

Data Exploration:

1. Load the provided `microsoft_stock.csv` file using pandas.
1. Display the first few rows of the dataframe to understand its structure.
2. Display basic information about the dataframe using appropriate pandas methods.
3. Check for any missing values in the dataset.
4. Print summary statistics of the dataframe to understand the distribution of the data.
5. Check the correlation between every column in the data frame and any other column
6. Find what columns has positive / negative and zero correlation

Data Preparation:

1. Convert the `Date` column to a datetime format and set the `Date` as the dataframe index.
2. From the entire dataframe select the following columns as the features:
    a. `Open`
    b. `High`
    c. `Low`
    d. `Volume`
3. Select the `Close` column as the label variable.

Linear Regression Machine Learning:

1. Perform Linear Regression machine learning on the features selected
2. Use train / test split to split the data (select 30% columns for the test set and seed of 42).
3. Apply a simple linear regression model with default values.
4. Evaluate your model results using MAE, MSA, RMSE error metrics.
5. Check what is the error percentage of your MAE and RMSE metric results.
6. Decide if your error percentage is acceptable or not.
7. Plot the y_test vs residuals scatter plot and determine if your data set was a good fit for linear regression modeling.

Model Deployment:

1. Once you happy with your model results train your Linear Regression model on the entire dataset
2. Print the final beta coefficient the model found for each feature
3. Export your final model into a joblib file
4. Import your final model from the joblib file and load it back to your working area
5. Use the import model to predict the ‘close’ value of the following unknown data points:
    a. **Open:** 250.00, **High:** 255.00, **Low:** 249.50, **Volume:** 23,000,000
    b. **Open:** 260.00, **High:** 265.00, **Low:** 259.50, **Volume:** 21,500,000
    c. **Open:** 245.00, **High:** 250.00, **Low:** 244.50, **Volume:** 22,000,000