**Assignement**

Complete the following tasks in a Jupyter Notebook or Python script, ensuring thorough documentation with markdown explanations and well-commented code.

**1. Data Preprocessing and Exploratory Analysis**

* **Load** the dataset and perform an initial exploration.
* Conduct **Exploratory Data Analysis (EDA)**:
  + Generate summary statistics and describe the dataset.
  + Visualize the distribution of each variable.
  + Identify correlations between variables.
  + Create new features if necessary (e.g., extracting the hour from InvoiceDate).

**2. Data Cleaning & Handling Missing Values**

* **Identify** and handle missing values appropriately:
  + Discuss potential methods to address missing data (e.g., imputation strategies).
  + Implement the chosen methods and justify your decisions.
* **Remove** duplicate records if any.
* **Filter** out anomalies or outliers in the data (e.g., negative quantities or prices).

**3. Feature Engineering**

* **Create** new features to enhance model performance:
  + Compute TotalSales as Quantity × UnitPrice.
  + Extract temporal features from InvoiceDate (e.g., day of the week, month).
  + Encode categorical variables such as Country using appropriate encoding techniques.

**4. Training the Model**

* **Split** the dataset into training and testing subsets (e.g., 70%-30% split).
* **Train** a Linear Regression model and at least one other regression model of your choice (e.g., RandomForestRegressor).

**5. Hyperparameter Tuning Using GridSearchCV**

* Perform **GridSearchCV** to optimize hyperparameters for your chosen model.
* Document:
  + The parameters considered.
  + The best parameters identified.
  + The performance of the model with these parameters.

**6. Performance Metrics and Accuracy**

* Evaluate and compare the models using:
  + Mean Absolute Error (MAE)
  + Mean Squared Error (MSE)
  + Root Mean Squared Error (RMSE)
  + R-squared (R² score)
* Summarize the performance metrics in a clear table.

**7. Interpretation of Results**

* **Analyze** the results:
  + Determine which features significantly influence sales predictions.
  + Discuss the impact of hyperparameter tuning on model performance.
  + Suggest potential improvements or alternative approaches.