Data Cleaning Steps

Step 1: Load the dataset from the Excel file using pandas.  
 - Use pd.read\_excel('file.xlsx', sheet\_name='SheetName')

Step 2: Inspect the dataset structure and data types.  
 - Use df.info(), df.head() to view the columns and sample data.

No missing value.

No duplicate value.

Order Date formatted to dd-mm-yyyy .

Ship Date formatted to dd-mm-yyyy .

Category text standardized.

Sub-Category text standardized No missing value.

Step 3: Identify numerical columns for outlier detection.  
 - Column: 'Profit'.

Step 4: Remove outliers using the Interquartile Range (IQR) method:  
 - Calculate Q1 (25th percentile) and Q3 (75th percentile).  
 - Compute IQR = Q3 - Q1.  
 - Lower bound = Q1 - 1.5 \* IQR.  
 - Upper bound = Q3 + 1.5 \* IQR.  
 - Remove rows where values fall outside this range for any numeric column.

Step 5: Handle missing values:  
 - Drop rows with NaN values using df.dropna().

Step 6: Convert negative values in 'Profit' column to positive:  
 - df['Profit'] = df['Profit'].abs()

Step 7: Ensure consistent data types for all columns:  
 - Use df.convert\_dtypes() for automatic type conversion.

Step 8: Save the fully cleaned dataset into a new CSV file:  
 - df.to\_csv('cleaned\_data\_final.csv', index=False).