Data Cleaning, Transformation, and Aggregation on the Iris Dataset

Scenario Overview:

You are working with the Iris dataset, which contains various flower attributes, including sepal and petal measurements, and the species of the flowers. Your goal is to clean and transform the dataset to prepare it for further analysis, including handling missing values, reshaping the data, and performing basic aggregations.

Tasks:

Task 1: Data Inspection and Missing Value Handling

- **Inspect the Dataset:** Examine the dataset and identify which columns contain missing values. Report how many missing values are present in each column.
- **Handle Missing Values in Numeric Columns:** Replace any missing values in the numeric columns (`sepal_length`, `sepal_width`, `petal_length`, and `petal_width`) with the average (mean) value of the respective column.
- **Handle Missing Values in Categorical Column:** Impute any missing values in the **species** column by replacing them with the most frequent value in that column.

Task 2: Data Cleaning and Transformation

- **Remove Duplicate Entries:** Check if there are any duplicate rows in the dataset and remove them if found, ensuring the dataset only contains unique entries.
- **Create a New Column by Modifying Existing Ones:** Create a new column that calculates the **petal area** by multiplying the `petal_length` and `petal_width` columns. Add this column to the dataset.
- **Drop Rows with Any Remaining Missing Values:** After handling missing data in the previous step, drop any rows that still contain missing values.

Task 3: Aggregation and Transformation

- **Convert Categorical Data to Numeric:** Convert the **species** column (which is categorical) into numeric values by assigning each unique species a distinct integer value.
- **Aggregation:** Calculate the mean of each numeric column (`sepal_length`, `sepal_width`, `petal_length`, `petal_width`) grouped by the species of the flowers. This will give you insights into the average measurements for each species.

Task 4: Advanced Reshaping

- **Reshape the Data:** Reshape the dataset from a wide format to a long format. The goal is to create a new version of the dataset where each row corresponds to a single measurement (sepal length, sepal width, petal length, or petal width) for each flower. You should also create a column that identifies the type of measurement.
