✅ Task 1 – Data Cleaning and Preprocessing

🔍 Objective:

To clean and preprocess the raw **Iris dataset** by handling missing values, duplicates, and inconsistent data formats to make it analysis-ready for further tasks.

🧾 Dataset Used:

- **Name:** `iris.csv`
- **Description:** A classic dataset containing measurements (sepal length, sepal width, petal length, petal width) of three Iris flower species – *setosa*, *versicolor*, and *virginica*.

🔧 Steps Performed:

1. **Data Loading:**
    - Used `pandas` to load the CSV file into a DataFrame.
2. **Missing Values Handling:**
    - Checked for null/missing values using `.isnull().sum()`.
    - Although the dataset was mostly clean, numerical columns were set to use column mean if any missing values were found.
3. **Duplicate Removal:**
    - Identified duplicates using `.duplicated().sum()`.
    - Removed all duplicate rows using `.drop_duplicates()` to avoid redundancy in analysis.
4. **Standardization of Categorical Data:**
    - Converted text columns like `species` to lowercase and removed leading/trailing spaces using `.str.lower().str.strip()`.
5. **Export Cleaned Data:**
    - The final cleaned dataset was saved as `cleaned_iris.csv` for further analysis and visualization.

🛠️ Tools & Libraries Used:

- **Language:** Python
- **Libraries:** pandas