Practical 3: Descriptive Statistics - Measures of Central Tendency and variability

Descriptive Statistics - Measures of Central Tendency and variability

Perform the following operations on any open source dataset (e.g., data.csv)

1. Provide summary statistics (mean, median, minimum, maximum, standard deviation) for a dataset (age, income etc.) with numeric variables grouped by one of the qualitative (categorical) variable. For example, if your categorical variable is age groups and quantitative variable is income, then provide summary statistics of income grouped by the age groups. Create a list that contains a numeric value for each response to the categorical variable.
2. Write a Python program to display some basic statistical details like percentile, mean, standard deviation etc. of the species of 'Iris-setosa', 'Iris-versicolor' and 'Iris-virginica of iris.csv dataset.

Provide the codes with outputs and explain everything that you do in this step.

Here's how to perform the specified operations using Python with the provided steps:

1. Summary Statistics Grouped by a Categorical Variable

We'll assume the dataset contains the following columns:

- AgeGroup: Categorical variable (e.g., "18-25", "26-35", "36-45")
- Income: Numeric variable (income values for individuals)

```
import pandas as pd
# Example dataset creation
data = {
  "AgeGroup": ["18-25", "18-25", "26-35", "26-35", "36-45", "36-45"],
  "Income": [25000, 27000, 34000, 36000, 45000, 47000],
}
df = pd.DataFrame(data)

df

df.describe()

# Summary statistics grouped by AgeGroup
grouped_stats = df.groupby("AgeGroup").describe()
grouped_stats
```

2. Statistical Details for Iris Dataset

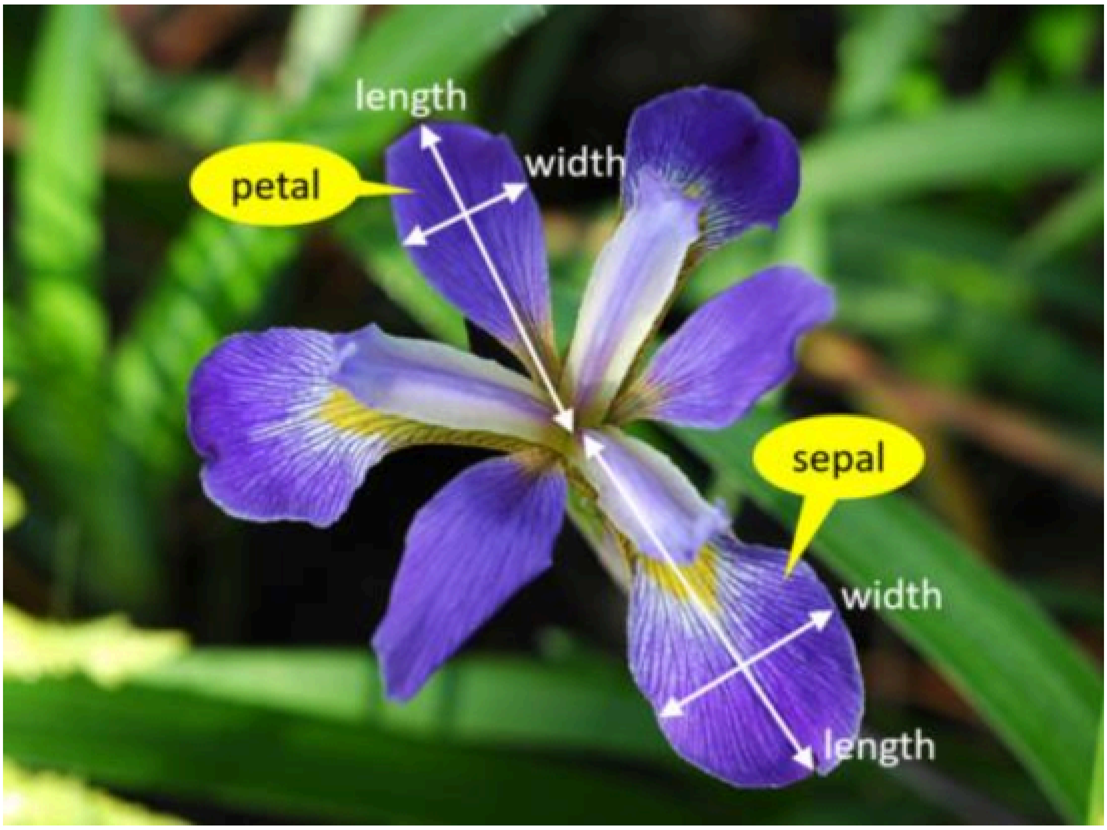
Dataset: Iris Dataset:
https://www.kaggle.com/datasets/vikrishnan/iris-dataset?resource=download

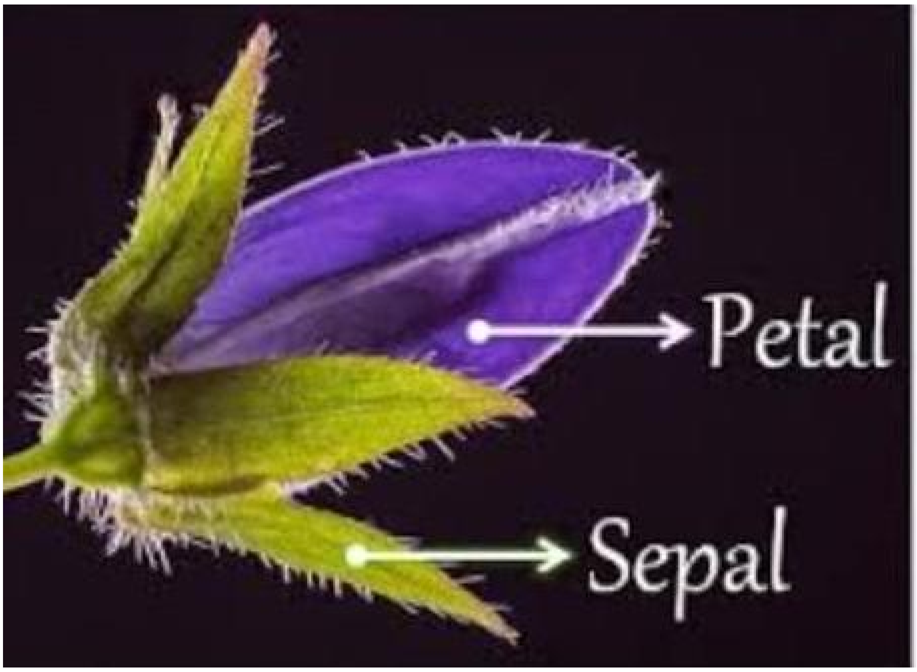
The dataset contains:

- sepal_length, sepal_width, petal_length, petal_width: Numeric variables
- species: Categorical variable with three categories: 'Iris-setosa', 'Iris-versicolor', 'Iris-virginica'

```
# Importing the Iris dataset

# Load the dataset
file_path = "C:/Users/Talha Ahmed/Desktop/My Practicals/Practical 3/iris.data.csv"
# Update the path as per your file location
data = pd.read_csv(file_path)
data

data = pd.read_csv(file_path, header=None) # remove header
data

# Assign meaningful column names based on Iris dataset
data.columns = ["sepal_length", "sepal_width", "petal_length", "petal_width", "species"]
data

#group by species
grouped_species = data.groupby("species").describe()
grouped_species

# Specify the output file path
output_file_path = "C:/Users/Talha Ahmed/Desktop/My Practicals/Practical 3/grouped_species.csv"

# Export the grouped data to a CSV file
grouped_species.to_csv(output_file_path)
```

```
# Filter the data for Iris-setosa
iris_setosa = data[data["species"] == "Iris-setosa"]
iris_setosa

# Display basic statistics for Iris-setosa
iris_setosa_statistics = iris_setosa.describe()
iris_setosa_statistics


# Filter the data for Iris-versicolor
iris_versicolor = data[data["species"] == "Iris-versicolor"]
iris_versicolor_statistics = iris_versicolor.describe()

iris_versicolor_statistics


# Filter the data for Iris-virginica
iris_virginica = data[data["species"] == "Iris-virginica"]
iris_virginica_statistics = iris_virginica.describe()
iris_virginica_statistics
```




Petal
Sepal