Iris Dataset Analysis using Python

Description:

This Python script performs Exploratory Data Analysis (EDA) on the Iris dataset. The dataset is loaded using pandas from a CSV file and includes columns for sepal length, sepal width, petal length, petal width, and species. Basic dataset information is displayed using .shape, .head(), .info(), and .describe(). Visualizations are created using seaborn and matplotlib: a pairplot shows feature relationships and species separation, histograms show value distributions, and box plots help identify outliers in the numeric columns.

Python Code:

```
# Import necessary libraries
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Load the Iris dataset from a CSV file
iris = pd.read_csv('IRIS.csv')  # Ensure IRIS.csv is in the same directory

# Print basic information
print("Shape of dataset:", iris.shape)
print("\nColumn names:", iris.columns.tolist())
print("\nFirst five rows:")
print(iris.head())

# Summary statistics
print("\nDataset info:")
print(iris.info())

print("\nDescriptive statistics:")
print(iris.describe())

# ------------------------------
# Visualization
# ------------------------------

# Scatter plot to show relationships
sns.pairplot(iris, hue='species')
plt.suptitle("Scatter Plot Matrix", y=1.02)
plt.show()

# Histograms for value distributions
iris.hist(edgecolor='black', figsize=(10, 8))
plt.suptitle("Histograms of Iris Features")
plt.tight_layout()
plt.show()

# Box plots to identify outliers
plt.figure(figsize=(12, 6))
for i, column in enumerate(iris.columns[:-1]):  # Skip 'species'
    plt.subplot(1, 4, i+1)
    sns.boxplot(y=iris[column])
    plt.title(f'Boxplot of {column}')
plt.tight_layout()
plt.show()
```