0qftqlmgp

April 19, 2025

Import essential Python libraries for data analysis and visualization. `seaborn` is used for advanced statistical visualizations with built-in datasets and themes. `pandas` is used for loading, manipulating, and analyzing data in DataFrame format. `matplotlib.pyplot` is used for creating static, animated, and interactive plots and acts as the base library behind many other visualization tools.

```
[1]: import seaborn as sns
     import pandas as pd
     import matplotlib.pyplot as plt
```

Load the classic Iris dataset using Seaborn's built-in `load_dataset()` function. The Iris dataset contains measurements for 150 iris flowers from three different species, with features like sepal length, sepal width, petal length, and petal width.

```
[3]: # Load the Iris dataset
     df = sns.load_dataset('iris')
```

Display the first 5 rows of the Iris dataset using the `.head()` method. This gives a quick preview of the structure, column names, and sample values in the DataFrame, helping to understand the kind of data you're working with.

```
[4]: df.head()
```

```
[4]:    sepal_length  sepal_width  petal_length  petal_width species
     0           5.1          3.5           1.4          0.2  setosa
     1           4.9          3.0           1.4          0.2  setosa
     2           4.7          3.2           1.3          0.2  setosa
     3           4.6          3.1           1.5          0.2  setosa
     4           5.0          3.6           1.4          0.2  setosa
```

Create histograms for each numeric feature in the Iris dataset using the `.hist()` method. The figure size is set to (10, 8) for clarity, with 15 bins per feature for granularity. The bars are colored sky blue with black edges for better visibility. `plt.suptitle()` adds a descriptive title to the entire figure. `plt.tight_layout()` ensures that subplots do not overlap, and `plt.show()` renders the final plot.

```
[5]: # Create histograms for each numeric feature
     df.hist(figsize=(10, 8), bins=15, color='skyblue', edgecolor='black')
     plt.suptitle("Feature Distributions (Histograms)- Iris Dataset", fontsize=14)
     plt.tight_layout()
```
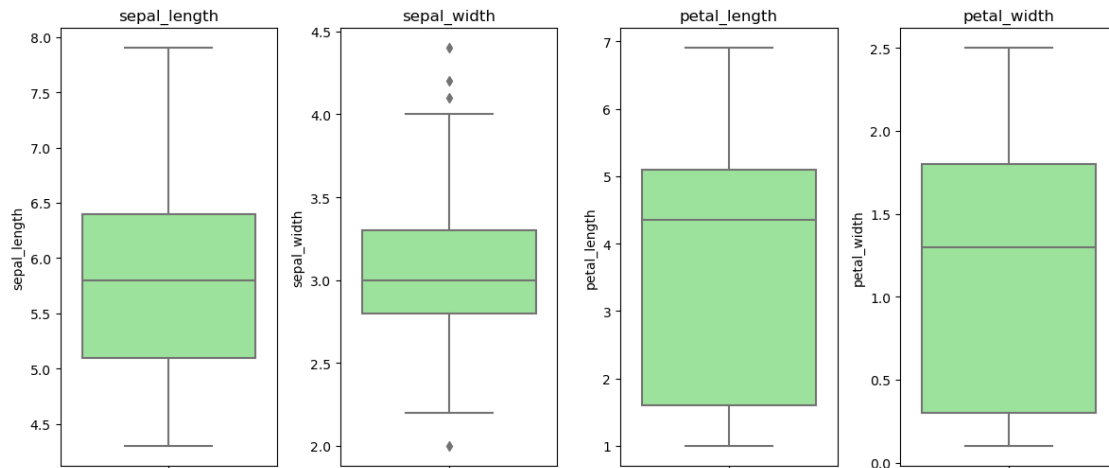
```
plt.show()
```

Feature Distributions (Histograms)- Iris Dataset
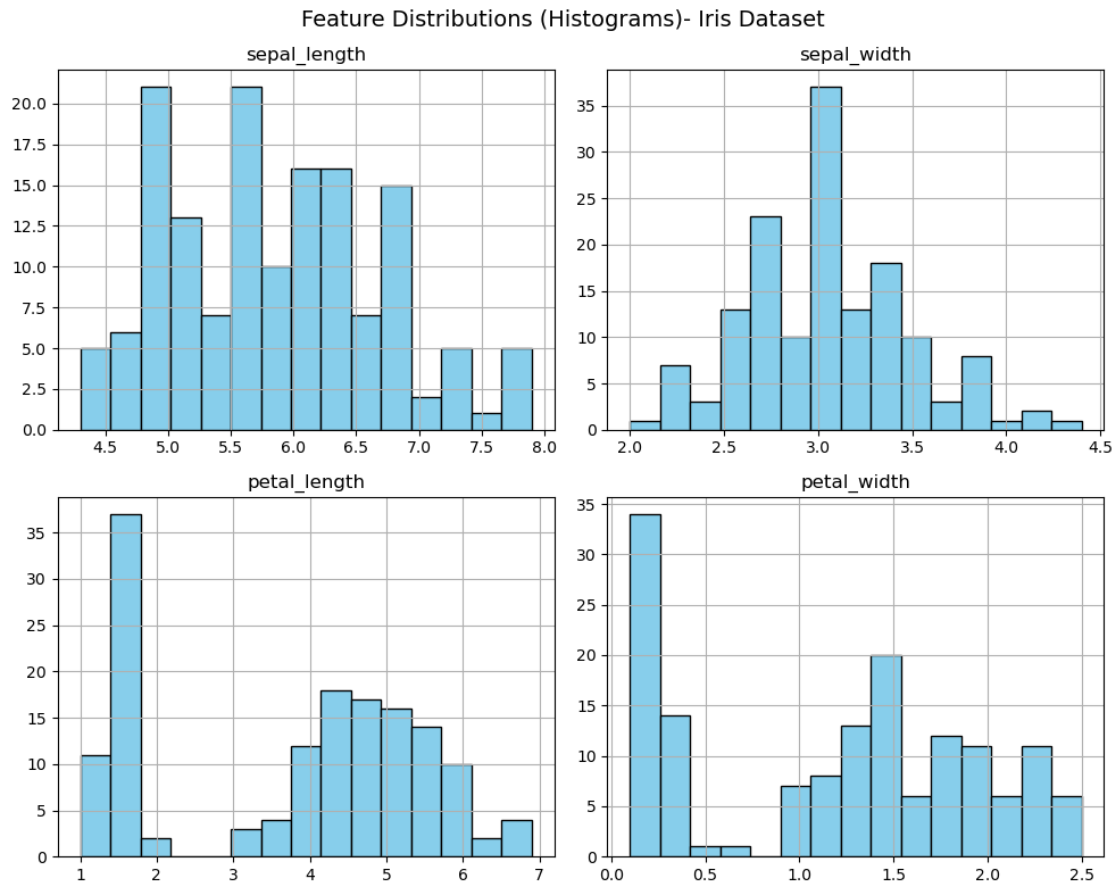
Create box plots for each numeric feature in the Iris dataset. A figure is initialized with a width of 12 and height of 6 for proper spacing. The `for` loop iterates through all numeric columns in the DataFrame using `select_dtypes()`. Inside the loop, each subplot is positioned side by side using `plt.subplot(1, 4, i + 1)` for a 1-row, 4-column layout. A box plot is generated for each numeric feature using Seaborn's `sns.boxplot()`, with a light green color for better visual appeal. The title for each subplot is set to the column name. A main title is added to the__

```
[7]: # Create box plots for each numeric feature
     plt.figure(figsize=(12, 6))
     numeric_columns = df.select_dtypes(include='number').columns
     for i, column in enumerate(numeric_columns):
         plt.subplot(1, len(numeric_columns), i + 1)
         sns.boxplot(y=df[column], color='lightgreen')
         plt.title(column)
     plt.suptitle("Box Plots of Numeric Features - Iris Dataset", fontsize=14)
     plt.tight_layout(rect=[0, 0.03, 1, 0.95])  # to make space for suptitle
     plt.show()
```

Box Plots of Numeric Features - Iris Dataset

[ ]: