Madhur Jaripatke

Roll No. 48

TE A Computer

RMDSSOE, Warje, Pune

Data Visualization III

Download the Iris flower dataset or any other dataset into a DataFrame. (e.g., https://archive.ics.uci.edu/ml/datasets/Iris). Scan the dataset and give the inference as:

1. List down the features and their types (e.g., numeric, nominal) available in the dataset.
2. Create a histogram for each feature in the dataset to illustrate the feature distributions.
3. Create a box plot for each feature in the dataset.
4. Compare distributions and identify outliers.

Importing the Libraries

```python
In [1]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
```

Load the Iris dataset

```python
In [2]: iris = sns.load_dataset('iris')
        iris.head()
```

Out[2]:

	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

List down the features and their types (e.g., numeric, nominal) available in the dataset.

```
In [3]: fig, axes = plt.subplots(2, 2, figsize=(10, 10))
        sns.histplot(iris['sepal_length'], kde=True, ax=axes[0, 0])
        sns.histplot(iris['sepal_width'], kde=True, ax=axes[0, 1])
        sns.histplot(iris['petal_length'], kde=True, ax=axes[1, 0])
        sns.histplot(iris['petal_width'], kde=True, ax=axes[1, 1])
```

Out[3]: <Axes: xlabel='petal_width', ylabel='Count'>
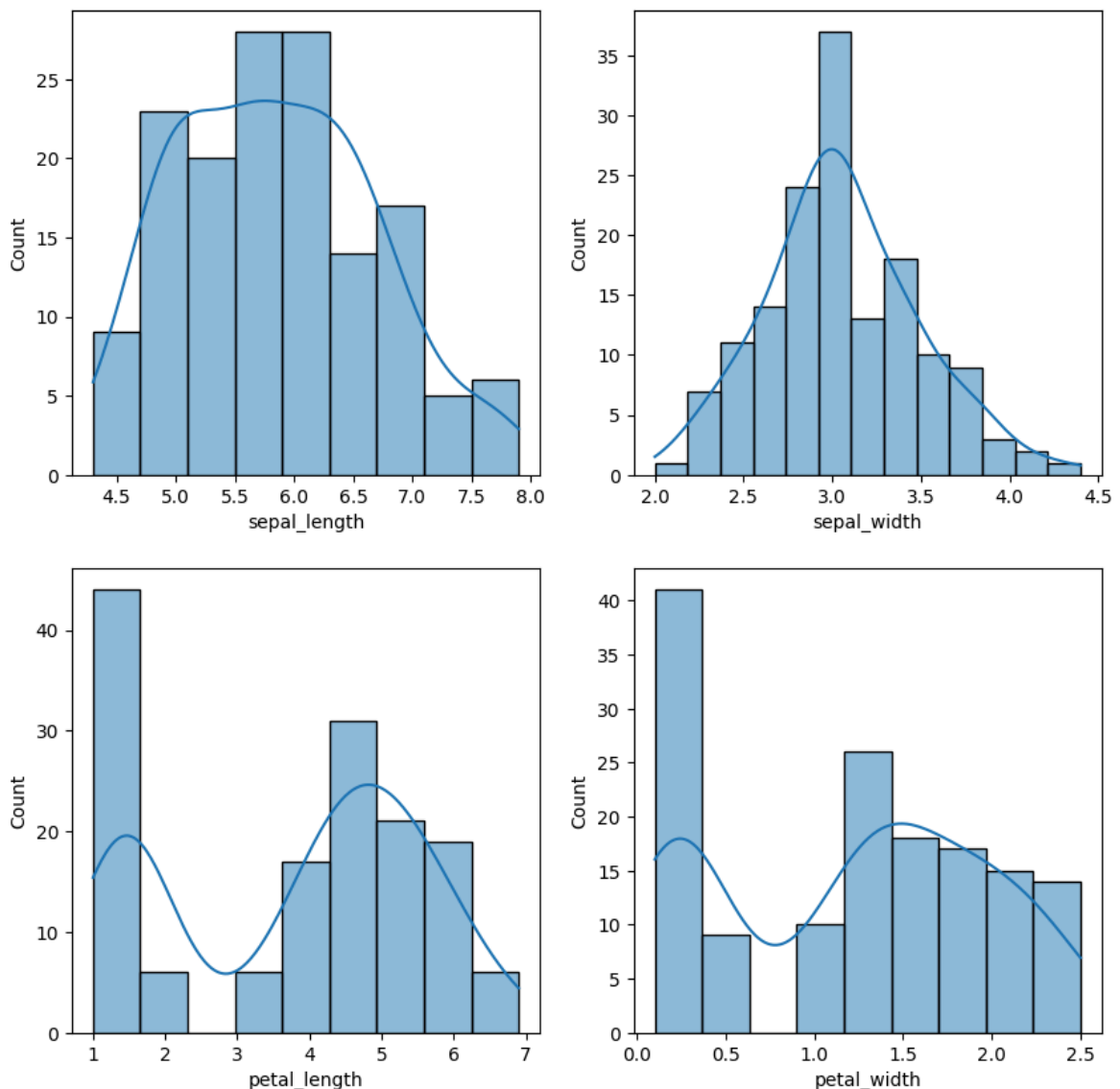
Create a box plot for each feature in the dataset.

```
In [4]: fig, axes = plt.subplots(2, 2, figsize=(10, 10))
        sns.boxplot(y='petal_length', x = 'species', data = iris, ax=axes[0, 0])
        sns.boxplot(y='petal_width', x = 'species', data = iris, ax=axes[0, 1])
```

```
sns.boxplot(y='sepal_length', x = 'species', data = iris, ax=axes[1, 0])
sns.boxplot(y='sepal_width', x = 'species', data = iris, ax=axes[1, 1])
```

Out[4]:   <Axes: xlabel='species', ylabel='sepal_width'>