Assignment 9 - Data Visualization III (Data Science)

Atharva Taras (TE A - 73)

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

In [1]:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:

```
data = pd.read_csv('Iris.csv')
data.head()
```

Out[2]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa

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

In [3]:

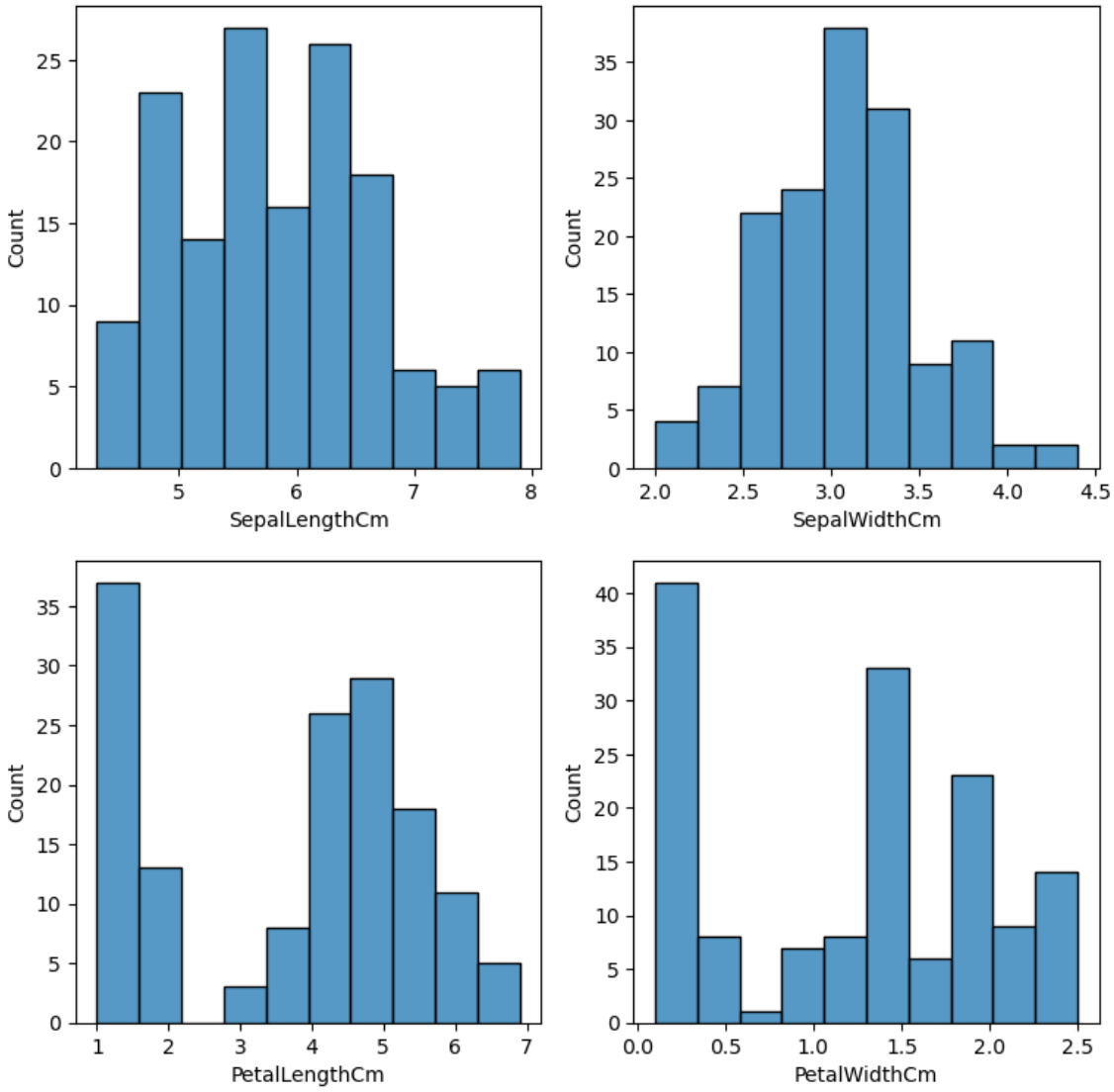
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Id             150 non-null    int64
 1   SepalLengthCm  150 non-null    float64
 2   SepalWidthCm   150 non-null    float64
 3   PetalLengthCm  150 non-null    float64
 4   PetalWidthCm   150 non-null    float64
 5   Species        150 non-null    object
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
```

2. Create a histogram for each feature in the dataset to illustrate the feature distributions.

In [4]:

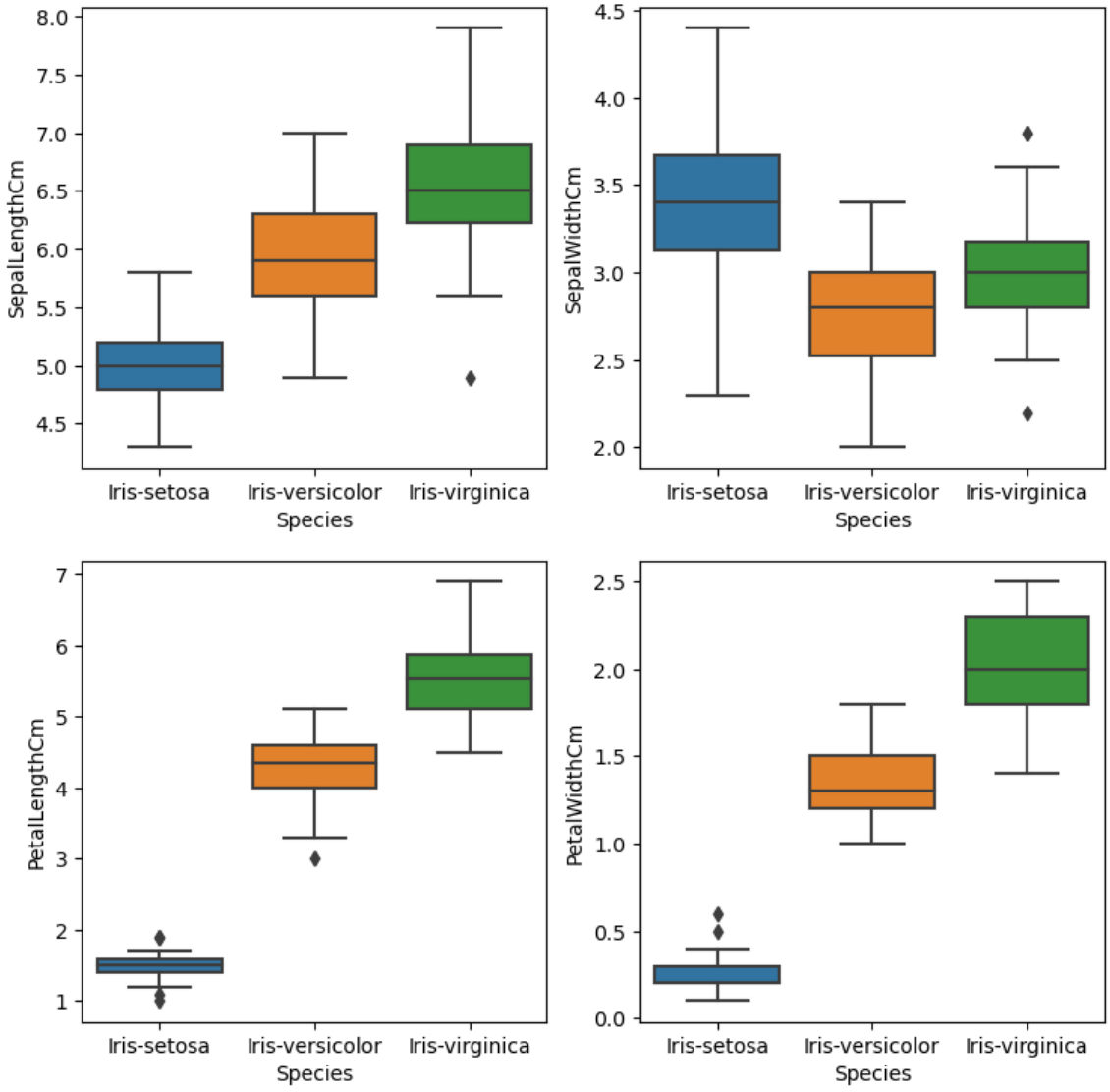
```
fig, axes = plt.subplots(nrows=2, ncols=2, figsize=(9, 9))
axes = axes.flatten()
ax = sns.histplot(x="SepalLengthCm", data=data, bins=10,ax=axes[0])
ax = sns.histplot(x="SepalWidthCm", data=data, bins=10,ax=axes[1])
ax = sns.histplot(x="PetalLengthCm", data=data, bins=10,ax=axes[2])
ax = sns.histplot(x="PetalWidthCm", data=data, bins=10, ax=axes[3])
```



3. Create a boxplot for each feature in the dataset.

In [5]:

```
fig, axes = plt.subplots(nrows=2, ncols=2, figsize=(9, 9))
axes = axes.flatten()
ax = sns.boxplot(x="Species", y="SepalLengthCm", data=data, orient='v', ax=axes[0])
ax = sns.boxplot(x="Species", y="SepalWidthCm", data=data, orient='v', ax=axes[1])
ax = sns.boxplot(x="Species", y="PetalLengthCm", data=data, orient='v', ax=axes[2])
ax = sns.boxplot(x="Species", y="PetalWidthCm", data=data, orient='v', ax=axes[3])
```



4. Compare distributions and identify outliers.

Outliers found in

1 - Iris Setosa (Petal Length, Petal Width)

2 - Iris Versicolor (Petal Lenght)

3 - Iris Viginica (Sepal Lenght, Sepal Width)