

Data Visualization III

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

- 1. How many features are there and what are their types (e.g., numeric, nominal)?
- 2. Create a histogram for each feature in the dataset to illustrate the feature distributions.
- 3. Create a boxplot for each feature in the dataset.
- 4. Compare distributions and identify outliers

By,

Vinayak Jalan

TE B 74

In [1]:

```
import seaborn as sns
iris = sns.load_dataset("iris")
```

In [2]:

```
iris
```

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
...
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

150 rows × 5 columns

In [3]:

```
iris.info
```

Out[3]:

```
<bound method DataFrame.info of
th 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
..      ...      ...      ...      ...      ...
145     6.7      3.0      5.2      2.3  virginica
146     6.3      2.5      5.0      1.9  virginica
147     6.5      3.0      5.2      2.0  virginica
148     6.2      3.4      5.4      2.3  virginica
149     5.9      3.0      5.1      1.8  virginica
```

```
[150 rows x 5 columns]>
```

In [4]:

```
iris.describe()
```

Out[4]:

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

In [5]:

```
type(iris.sepal_length)
```

Out[5]:

```
pandas.core.series.Series
```

In [6]:

```
iris.sepal_length.dtype
```

Out[6]:

```
dtype('float64')
```

In [7]:

```
iris.sepal_width.dtype
```

Out[7]:

```
dtype('float64')
```

In [8]:

```
iris.petal_length.dtype
```

Out[8]:

```
dtype('float64')
```

In [9]:

```
iris.petal_width.dtype
```

Out[9]:

```
dtype('float64')
```

In [10]:

```
iris.species.dtype
```

Out[10]:

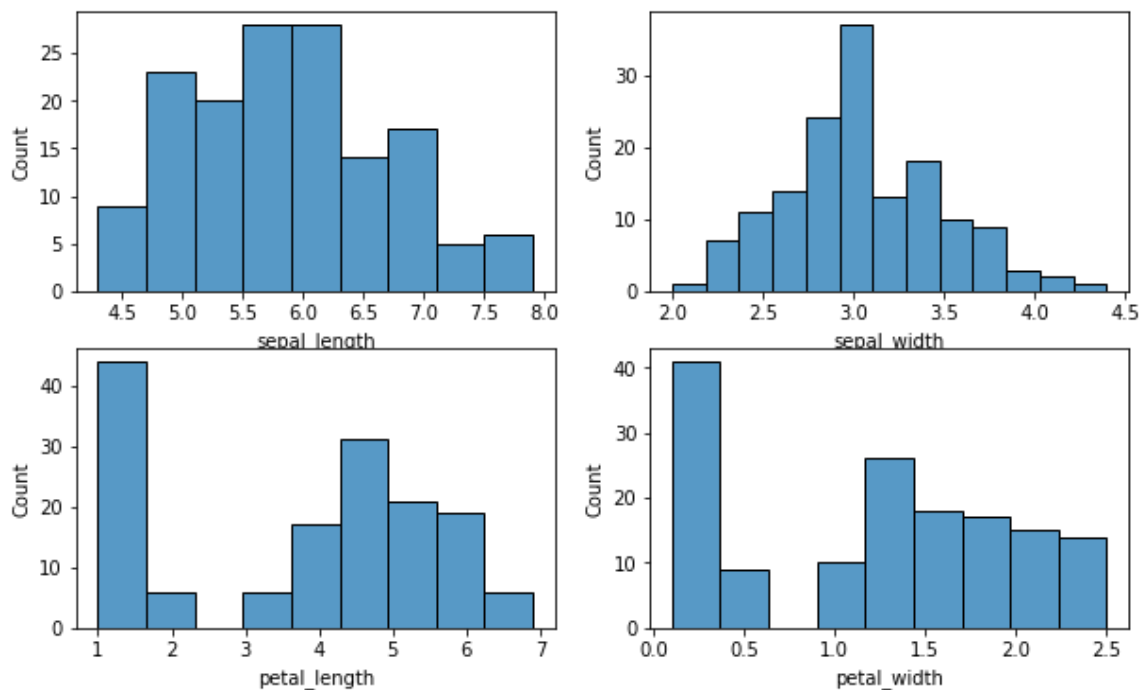
```
dtype('O')
```

In [11]:

```
import matplotlib.pyplot as plt
fig, axes = plt.subplots(2, 2, figsize=(10, 6))
sns.histplot(iris["sepal_length"], ax=axes[0, 0])
sns.histplot(iris["sepal_width"], ax=axes[0, 1])
sns.histplot(iris["petal_length"], ax=axes[1, 0])
sns.histplot(iris["petal_width"], ax=axes[1, 1])
```

Out[11]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f6b7e62b110>



In [12]:

```
#For boxplot
```

```
fig, axes = plt.subplots(2, 2, figsize=(16, 10))
sns.boxplot(x="species", y="sepal_length", data=iris, ax=axes[0, 0])
sns.boxplot(x="species", y="sepal_width", data=iris, ax=axes[0, 1])
sns.boxplot(x="species", y="petal_length", data=iris, ax=axes[1, 0])
sns.boxplot(x="species", y="petal_width", data=iris, ax=axes[1, 1])
```

Out[12]:

<matplotlib.axes._subplots.AxesSubplot at 0x7f6b7dfce110>

