DSBDA Lab Assignment No. 10

```
Name: Akash Ganesh Padir
Roll No.: TEB04
```

Import Libraries

In [1]:
```
import pandas as pd
import numpy as np
```

Import Dataset

In [4]:
```
csv_url= 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.
```

In [5]:
```
iris= pd.read_csv(csv_url, header=None)
```

In [6]:
```
col_names=['Sepal_Length', 'Sepal_Width', 'Petal_Length', 'Petal_Width', 'Spec
```

In [8]:
```
iris= pd.read_csv(csv_url, names=col_names)
print(iris)
```

```
     Sepal_Length  Sepal_Width  Petal_Length  Petal_Width         Species
0             5.1          3.5           1.4          0.2     Iris-setosa
1             4.9          3.0           1.4          0.2     Iris-setosa
2             4.7          3.2           1.3          0.2     Iris-setosa
3             4.6          3.1           1.5          0.2     Iris-setosa
4             5.0          3.6           1.4          0.2     Iris-setosa
..            ...          ...           ...          ...             ...
145           6.7          3.0           5.2          2.3  Iris-virginica
146           6.3          2.5           5.0          1.9  Iris-virginica
147           6.5          3.0           5.2          2.0  Iris-virginica
148           6.2          3.4           5.4          2.3  Iris-virginica
149           5.9          3.0           5.1          1.8  Iris-virginica

[150 rows x 5 columns]
```

In [9]:
```
iris.head()
```

Out[9]:

	Sepal_Length	Sepal_Width	Petal_Length	Petal_Width	Species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

1. How many features are there and what are their types (e.g., numeric, nominal)?

In [10]:
```
iris.info()
```

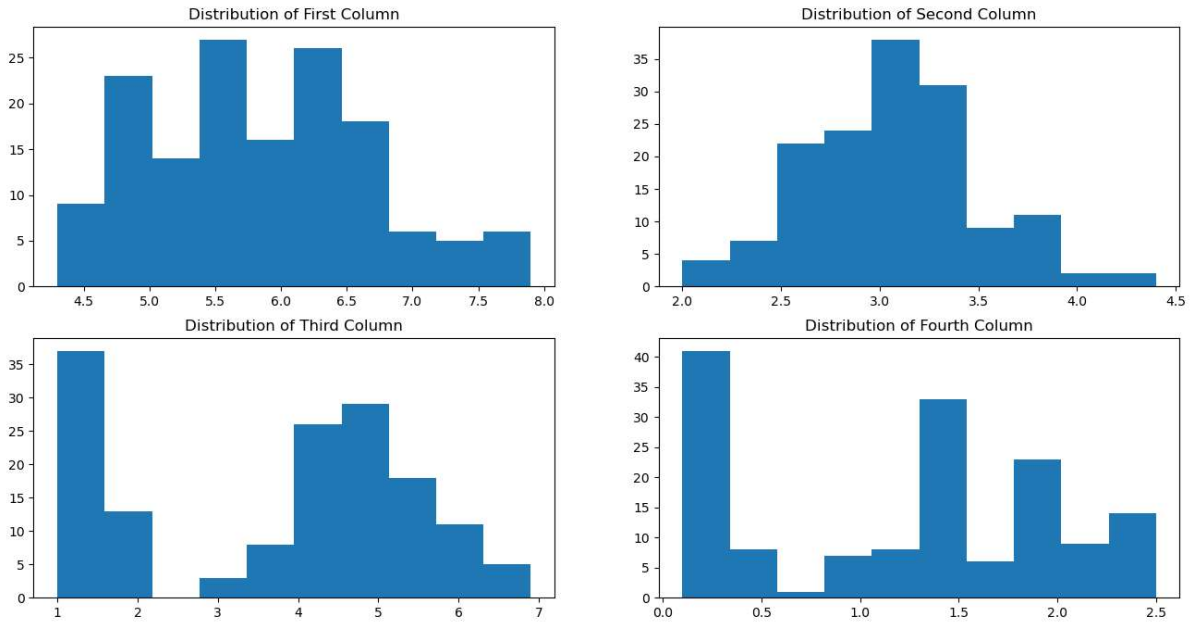
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Sepal_Length  150 non-null    float64
 1   Sepal_Width   150 non-null    float64
 2   Petal_Length  150 non-null    float64
 3   Petal_Width   150 non-null    float64
 4   Species       150 non-null    object 
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

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

In [11]:
```
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline
```

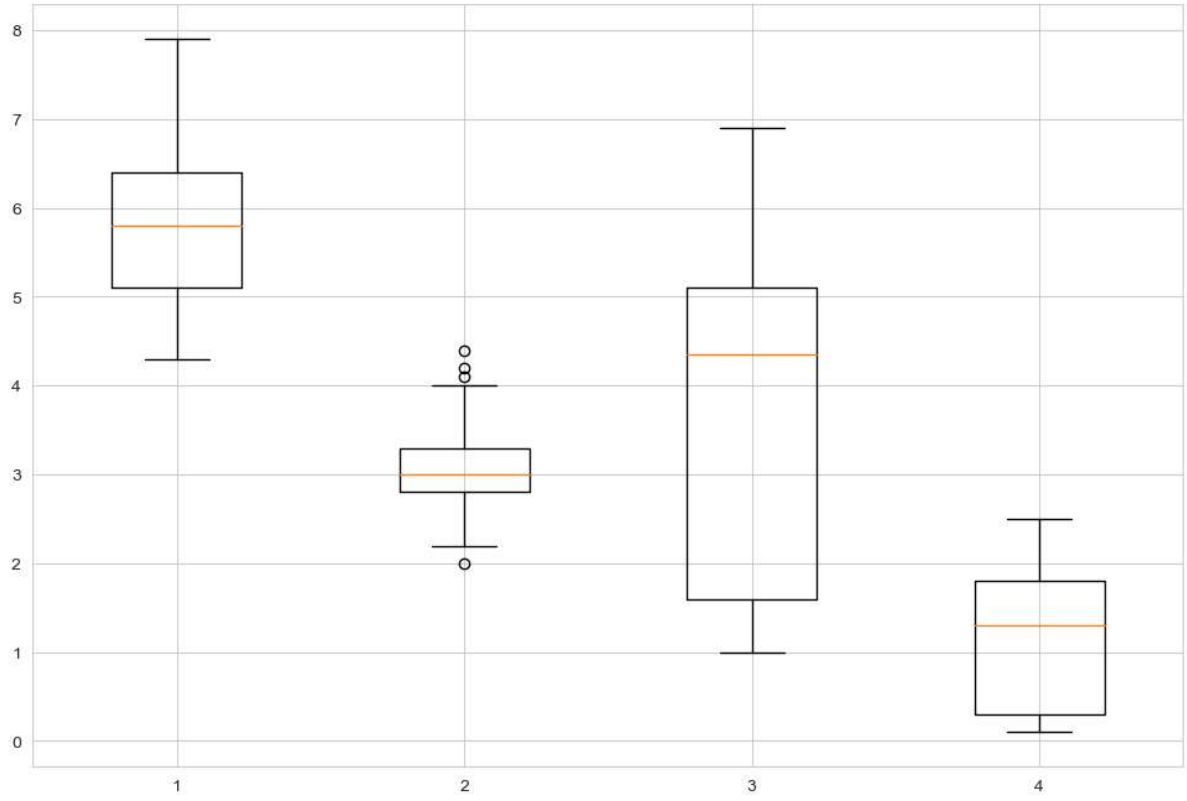
In [13]:
```
fig, axes = plt.subplots(2,2, figsize=(16,8))

axes[0,0].set_title("Distribution of First Column")
axes[0,0].hist(iris["Sepal_Length"])

axes[0,1].set_title("Distribution of Second Column")
axes[0,1].hist(iris["Sepal_Width"])

axes[1,0].set_title("Distribution of Third Column")
axes[1,0].hist(iris["Petal_Length"])

axes[1,1].set_title("Distribution of Fourth Column")
axes[1,1].hist(iris["Petal_Width"])
```

Out[13]:
```
(array([41.,  8.,  1.,  7.,  8., 33.,  6., 23.,  9., 14.]),
 array([0.1 , 0.34, 0.58, 0.82, 1.06, 1.3 , 1.54, 1.78, 2.02, 2.26, 2.5 ]),
 <BarContainer object of 10 artists>)
```



3. Create a boxplot for each feature in the dataset.Compare distributions and identify outliers

In [14]:
```
data_to_plot = [iris["Sepal_Length"],iris["Sepal_Width"],iris["Petal_Length"],

sns.set_style("whitegrid")
#sns.set_style("dark")

#Creating a figure instance
fig=plt.figure(1, figsize=(12,8))

#Creating an axes instance
ax=fig.add_subplot(111)

#Creating the boxplot
bp=ax.boxplot(data_to_plot);
```



In [ ]:
```
```