

Assignment 10

In [1]:

```
import numpy as np
import pandas as pd
```

In [2]:

```
df = pd.read_csv("iris.csv")
df.columns = ["col1", "col2", "col3", "col4", "col5"]
```

In [3]:

```
df.head()
```

Out[3]:

	col1	col2	col3	col4	col5
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

In [4]:

```
df = df.drop(0)
```

In [5]:

df

Out[5]:

	col1	col2	col3	col4	col5
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
5	5.4	3.9	1.7	0.4	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

149 rows × 5 columns

In [6]:

```
column = len(list(df))
column
```

Out[6]:

5

In [7]:

df.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 149 entries, 1 to 149
Data columns (total 5 columns):
 #   Column  Non-Null Count  Dtype  
---  -
 0   col1    149 non-null    float64
 1   col2    149 non-null    float64
 2   col3    149 non-null    float64
 3   col4    149 non-null    float64
 4   col5    149 non-null    object 
dtypes: float64(4), object(1)
memory usage: 7.0+ KB
```

In [8]:

```
np.unique(df["col5"])
```

Out[8]:

```
array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'], dtype=object)
```

In [9]:

```
df.describe()
```

Out[9]:

	col1	col2	col3	col4
count	149.000000	149.000000	149.000000	149.000000
mean	5.848322	3.051007	3.774497	1.205369
std	0.828594	0.433499	1.759651	0.761292
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.400000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

In [10]:

```
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline
```

In [11]:

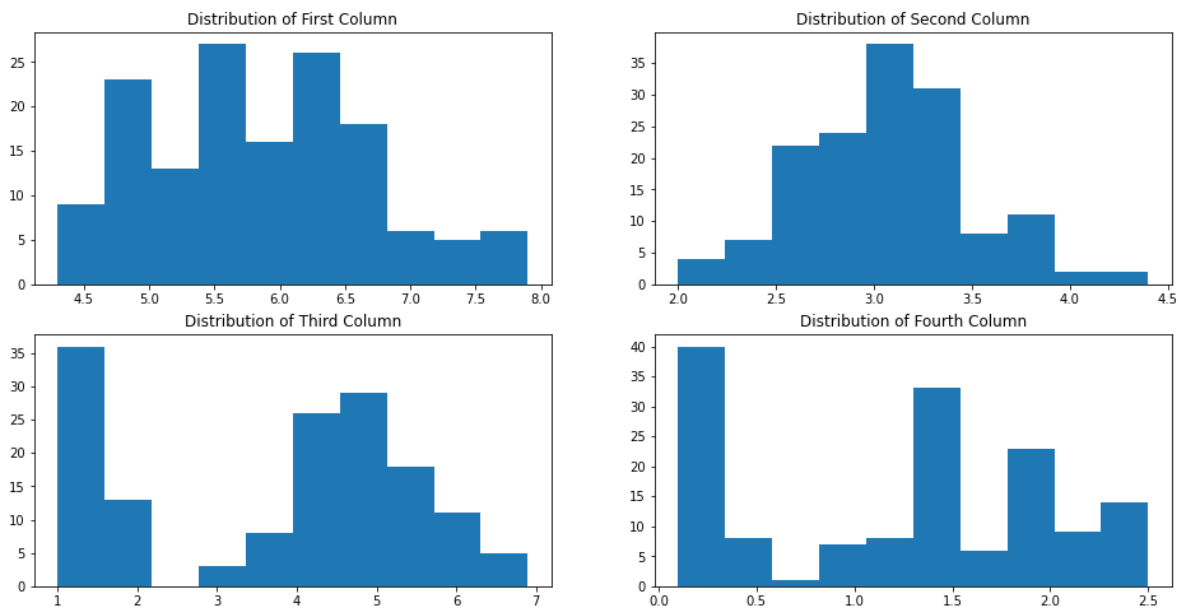
```
fig, axes = plt.subplots(2, 2, figsize=(16, 8))

axes[0,0].set_title("Distribution of First Column")
axes[0,0].hist(df["col1"]);

axes[0,1].set_title("Distribution of Second Column")
axes[0,1].hist(df["col2"]);

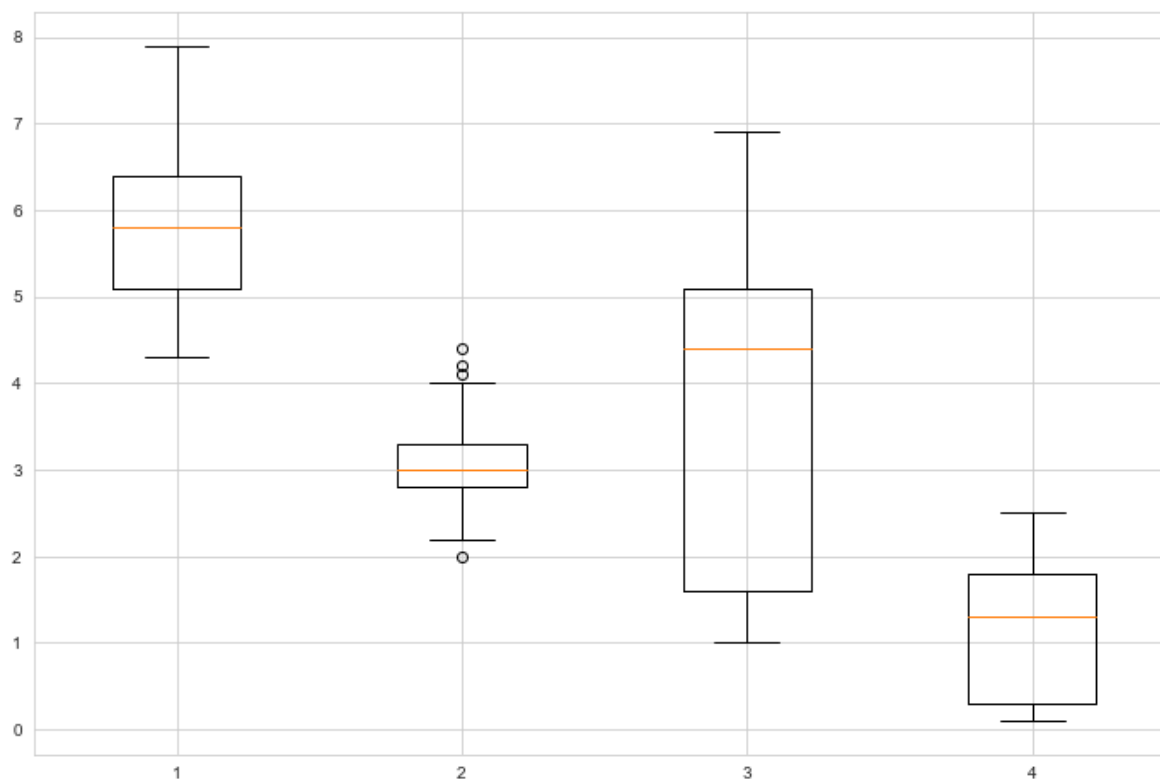
axes[1,0].set_title("Distribution of Third Column")
axes[1,0].hist(df["col3"]);

axes[1,1].set_title("Distribution of Fourth Column")
axes[1,1].hist(df["col4"]);
```



In [12]:

```
data_to_plot = [df["col1"],df["col2"],df["col3"],df["col4"]]  
  
sns.set_style("whitegrid")  
# 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 []: