

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

```
In [2]: df= pd.read_csv("Iris.csv",header=None)
df.columns=["col1","col2","col3","col4","col5","col6"]
df.head()
```

```
Out[2]:
```

	col1	col2	col3	col4	col5	col6
0	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
1	1	5.1	3.5	1.4	0.2	Iris-setosa
2	2	4.9	3.0	1.4	0.2	Iris-setosa
3	3	4.7	3.2	1.3	0.2	Iris-setosa
4	4	4.6	3.1	1.5	0.2	Iris-setosa

```
In [3]: column=len(list(df))
column
```

```
Out[3]: 6
```

```
In [4]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 151 entries, 0 to 150
Data columns (total 6 columns):
 #   Column  Non-Null Count  Dtype  
---  -
 0   col1    151 non-null    object  
 1   col2    151 non-null    object  
 2   col3    151 non-null    object  
 3   col4    151 non-null    object  
 4   col5    151 non-null    object  
 5   col6    151 non-null    object  
dtypes: object(6)
memory usage: 7.2+ KB
```

```
In [5]: np.unique(df["col6"])
```

```
Out[5]: array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica', 'Species'],
              dtype=object)
```

```
In [7]: import matplotlib.pyplot as plt
import matplotlib
import seaborn as sns
%matplotlib inline
```

```
C:\Users\avcoe\anaconda3\lib\site-packages\scipy\__init__.py:155: UserWarning: A NumPy
y version >=1.18.5 and <1.25.0 is required for this version of SciPy (detected versio
n 1.26.4
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}")
```

```
In [8]: import pandas as pd
import numpy as np
df= pd.read_csv("Iris.csv",header=None)
```

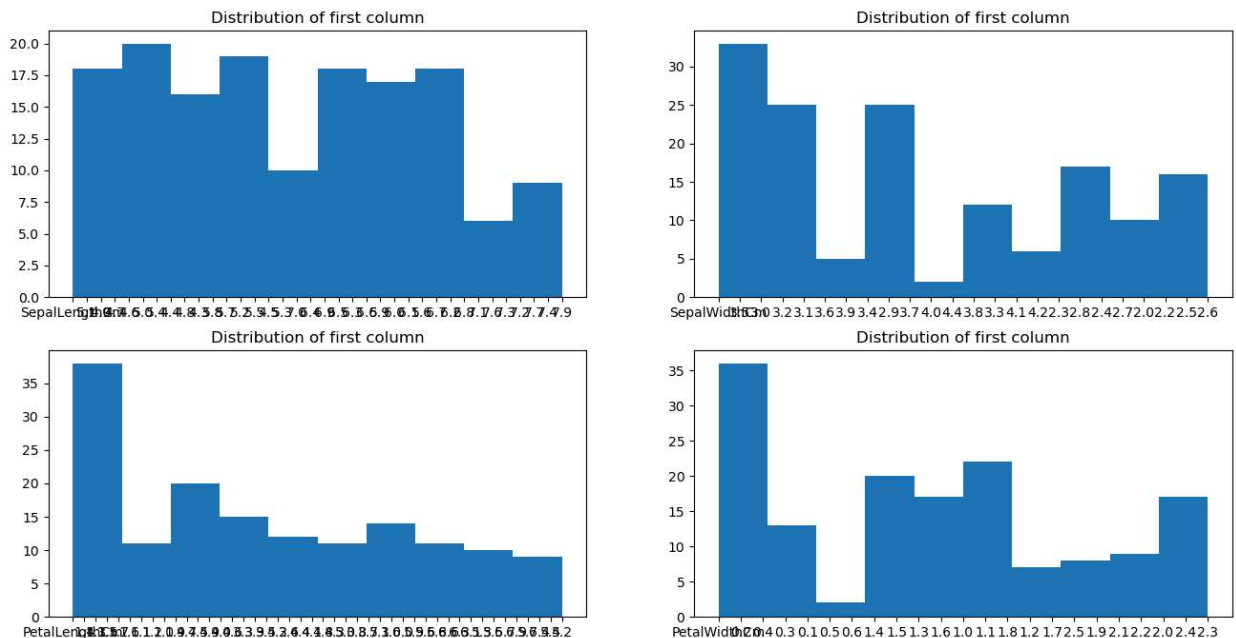
```
df.columns=["col1","col2","col3","col4","col5","col6"]
fig, axes=plt.subplots(2,2,figsize=(16,8))

axes[0,0].set_title("Distribution of first column")
axes[0,0].hist(df["col2"]);

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

axes[1,0].set_title("Distribution of first column")
axes[1,0].hist(df["col4"]);

axes[1,1].set_title("Distribution of first column")
axes[1,1].hist(df["col5"]);
```



```
In [9]: df["col2"]=pd.to_numeric(df["col2"],errors='coerce')
df["col3"]=pd.to_numeric(df["col3"],errors='coerce')
df["col4"]=pd.to_numeric(df["col4"],errors='coerce')
df["col5"]=pd.to_numeric(df["col5"],errors='coerce')

df.dropna(inplace=True)

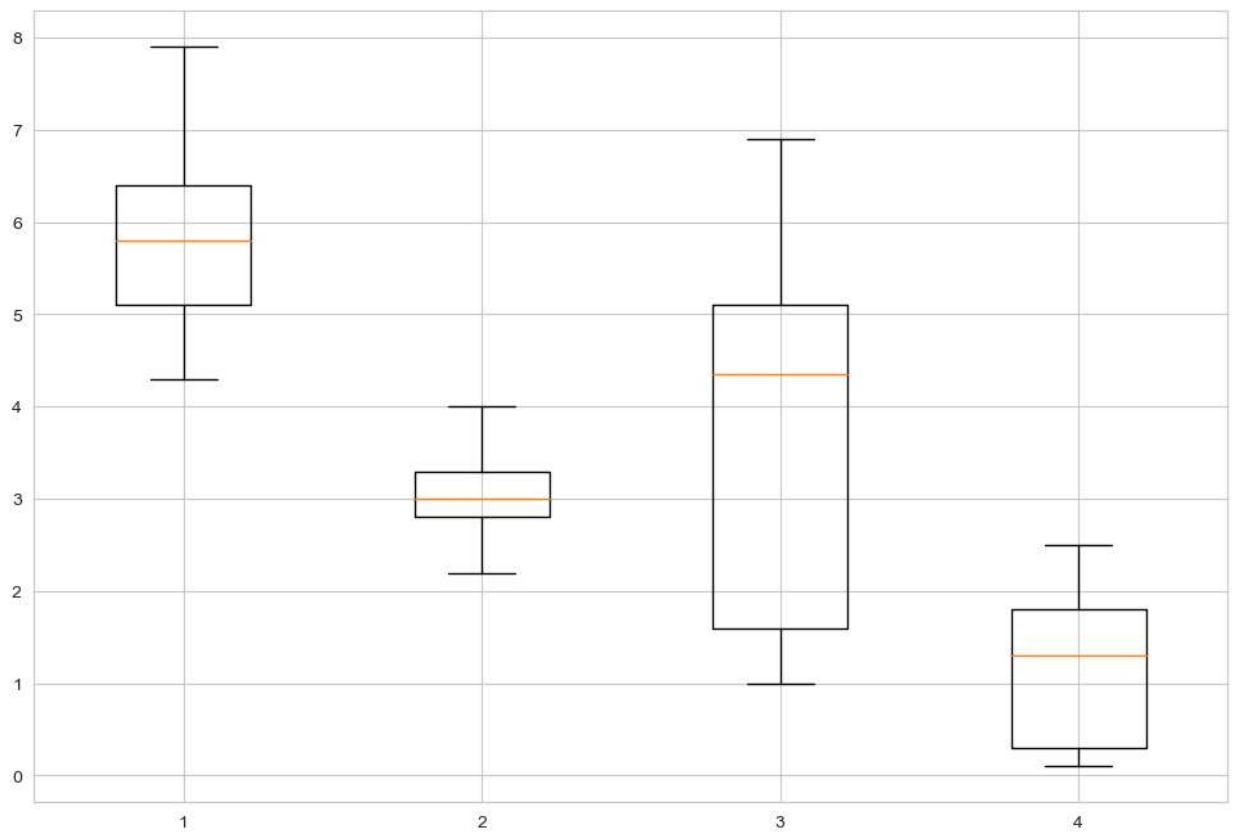
data_to_plot =[df["col2"],df["col3"],df["col4"],df["col5"]]

sns.set_style("whitegrid")

fig =plt.figure(1,figsize=(12,8))

ax=fig.add_subplot(111)

bp =ax.boxplot(data_to_plot,showfliers=False)
plt.show()
```



In [10]: `df.describe()`

Out[10]:

	col2	col3	col4	col5
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
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 []: