

```
In [ ]: import pandas as pd
import matplotlib.pyplot as pt
import numpy as np
import requests as r
```

Load Data Set

```
In [ ]: df = pd.read_csv("iris.data")
df
```

```
Out[ ]:
```

	5.1	3.5	1.4	0.2	Iris-setosa
0	4.9	3.0	1.4	0.2	Iris-setosa
1	4.7	3.2	1.3	0.2	Iris-setosa
2	4.6	3.1	1.5	0.2	Iris-setosa
3	5.0	3.6	1.4	0.2	Iris-setosa
4	5.4	3.9	1.7	0.4	Iris-setosa
...
144	6.7	3.0	5.2	2.3	Iris-virginica
145	6.3	2.5	5.0	1.9	Iris-virginica
146	6.5	3.0	5.2	2.0	Iris-virginica
147	6.2	3.4	5.4	2.3	Iris-virginica
148	5.9	3.0	5.1	1.8	Iris-virginica

149 rows × 5 columns

```
In [ ]: df.columns = ["SepalLengthCm", "SepalWidthCm", "PetalLengthCm", "PetalWidthCm", "Species"]
df
```

Out[]:

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	4.9	3.0	1.4	0.2	Iris-setosa
1	4.7	3.2	1.3	0.2	Iris-setosa
2	4.6	3.1	1.5	0.2	Iris-setosa
3	5.0	3.6	1.4	0.2	Iris-setosa
4	5.4	3.9	1.7	0.4	Iris-setosa
...
144	6.7	3.0	5.2	2.3	Iris-virginica
145	6.3	2.5	5.0	1.9	Iris-virginica
146	6.5	3.0	5.2	2.0	Iris-virginica
147	6.2	3.4	5.4	2.3	Iris-virginica
148	5.9	3.0	5.1	1.8	Iris-virginica

149 rows × 5 columns

In []: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 149 entries, 0 to 148
Data columns (total 5 columns):
 #   Column          Non-Null Count  Dtype
---  -
 0   SepalLengthCm   149 non-null   float64
 1   SepalWidthCm    149 non-null   float64
 2   PetalLengthCm   149 non-null   float64
 3   PetalWidthCm    149 non-null   float64
 4   Species         149 non-null   object
dtypes: float64(4), object(1)
memory usage: 5.9+ KB
```

In []: `df.isnull().sum()`

Out[]:

```
SepalLengthCm    0
SepalWidthCm      0
PetalLengthCm     0
PetalWidthCm      0
Species           0
dtype: int64
```

In []: `df.describe()`

Out[]:

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
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 []: `df.drop_duplicates()`

Out[]:

	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	4.9	3.0	1.4	0.2	Iris-setosa
1	4.7	3.2	1.3	0.2	Iris-setosa
2	4.6	3.1	1.5	0.2	Iris-setosa
3	5.0	3.6	1.4	0.2	Iris-setosa
4	5.4	3.9	1.7	0.4	Iris-setosa
...
144	6.7	3.0	5.2	2.3	Iris-virginica
145	6.3	2.5	5.0	1.9	Iris-virginica
146	6.5	3.0	5.2	2.0	Iris-virginica
147	6.2	3.4	5.4	2.3	Iris-virginica
148	5.9	3.0	5.1	1.8	Iris-virginica

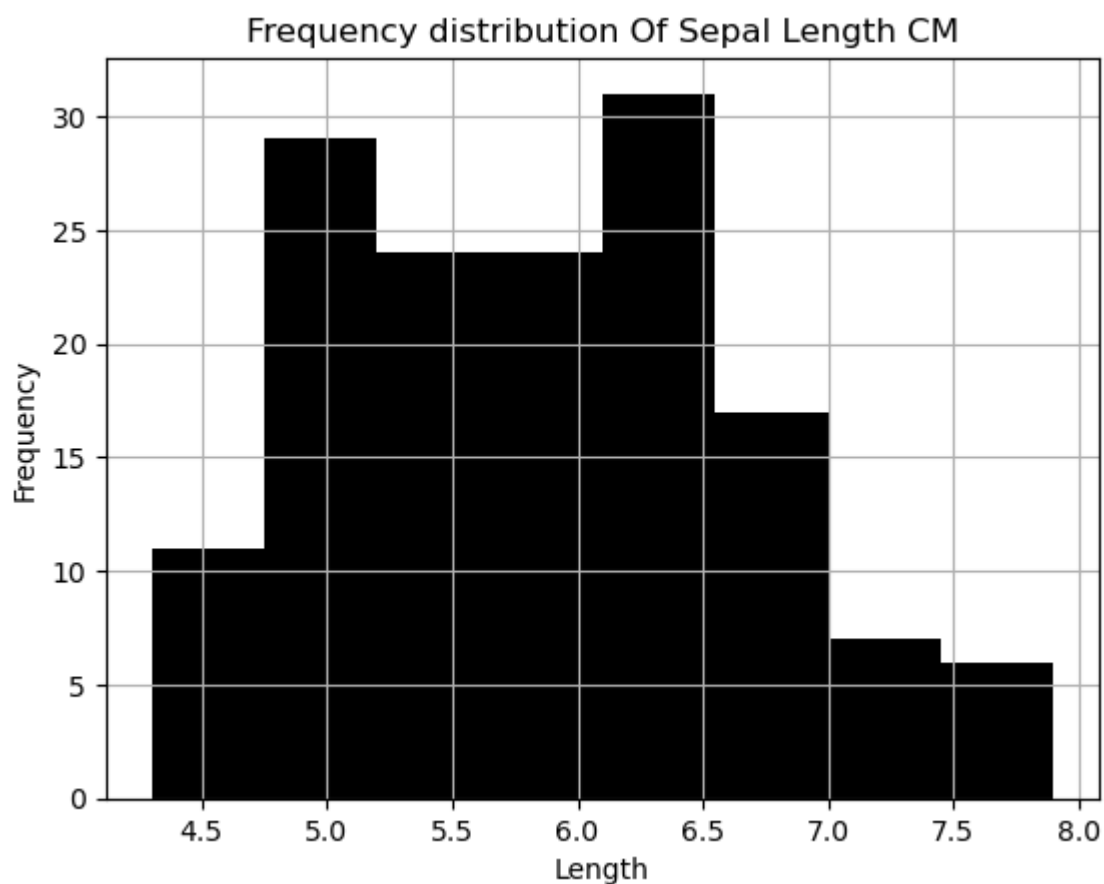
146 rows × 5 columns

Creating Histogram For Each Feature in dataset

1. Sepal Lenght In Centimeter

In []: `pt.title("Frequency distribution Of Sepal Length CM")
pt.xlabel("Length")
pt.ylabel("Frequency")
df["SepalLengthCm"].hist(color="black",bins=8)`

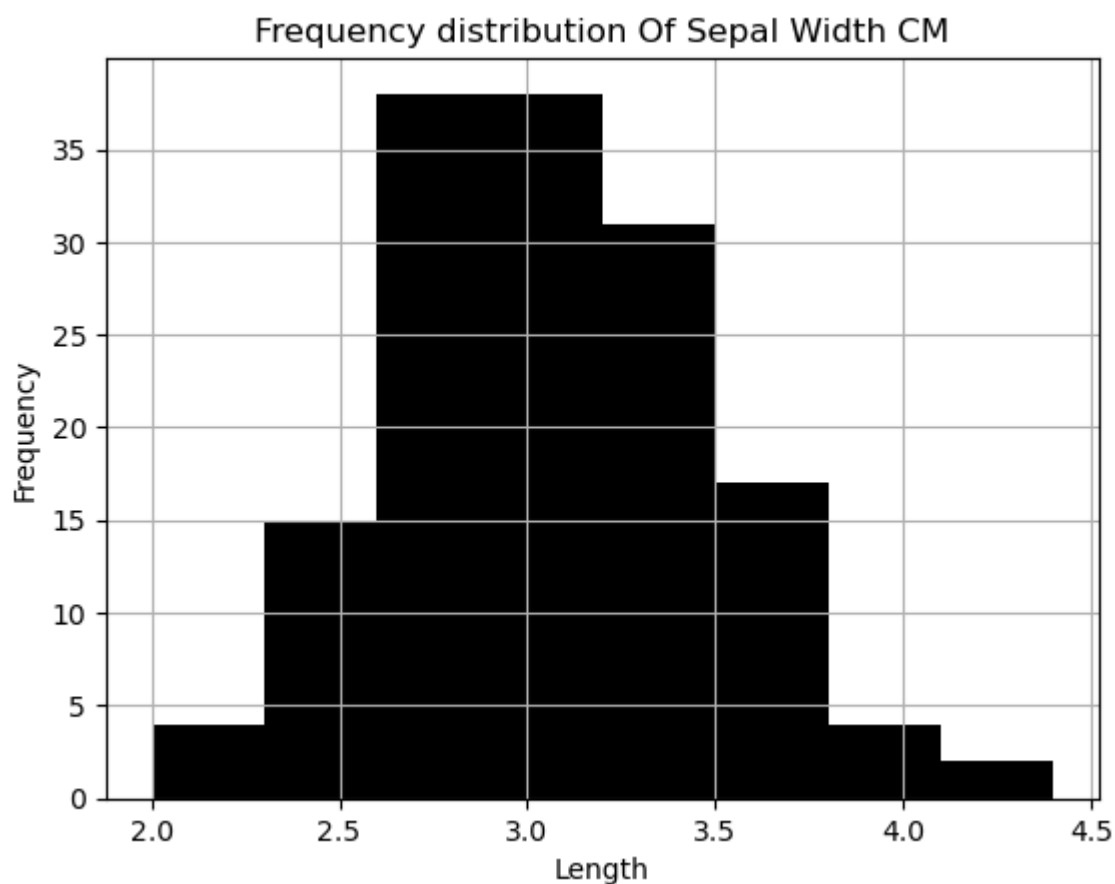
Out[]: `<AxesSubplot:title={'center':'Frequency distribution Of Sepal Length CM'}, xlabel='Length', ylabel='Frequency'>`



1. Sepal Width In Centimeter

```
In [ ]: pt.title("Frequency distribution Of Sepal Width CM")
pt.xlabel("Length")
pt.ylabel("Frequency")
df["SepalWidthCm"].hist(color="black",bins=8)
```

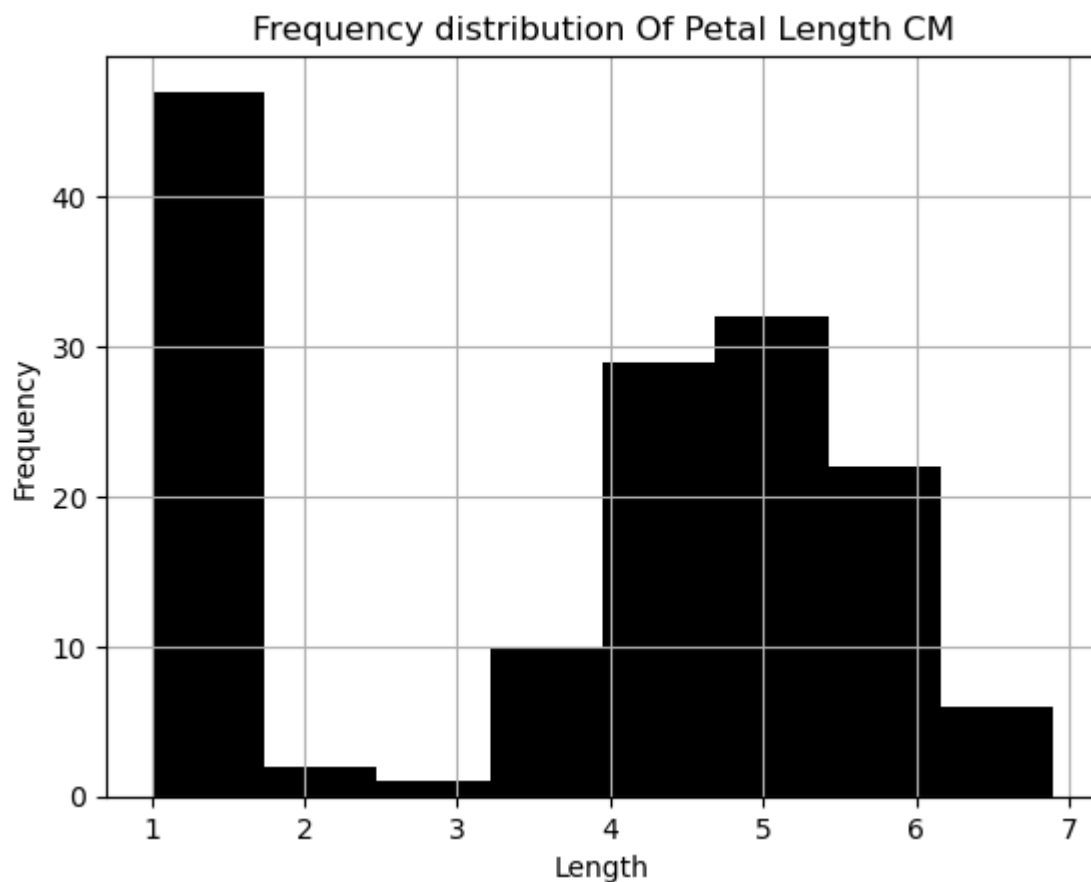
```
Out[ ]: <AxesSubplot:title={'center':'Frequency distribution Of Sepal Width CM'}, xlabel='Length', ylabel='Frequency'>
```



1. Petal Length In Centimeter

```
In [ ]: pt.title("Frequency distribution Of Petal Length CM")
pt.xlabel("Length")
pt.ylabel("Frequency")
df["PetalLengthCm"].hist(color="black",bins=8)
```

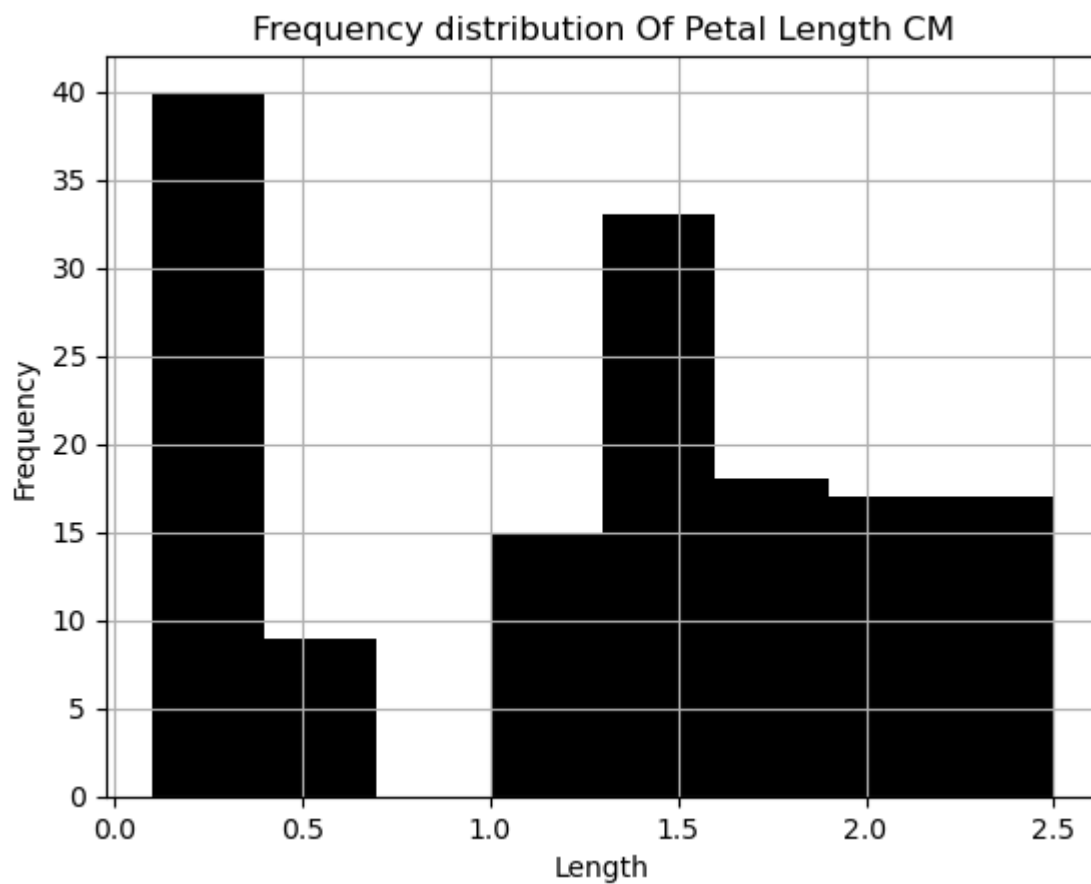
```
Out[ ]: <AxesSubplot:title={'center':'Frequency distribution Of Petal Length CM'}, xlabel='Length', ylabel='Frequency'>
```



1. Petal Width In Centimeter

```
In [ ]: pt.title("Frequency distribution Of Petal Length CM")
pt.xlabel("Length")
pt.ylabel("Frequency")
df["PetalWidthCm"].hist(color="black",bins=8)
```

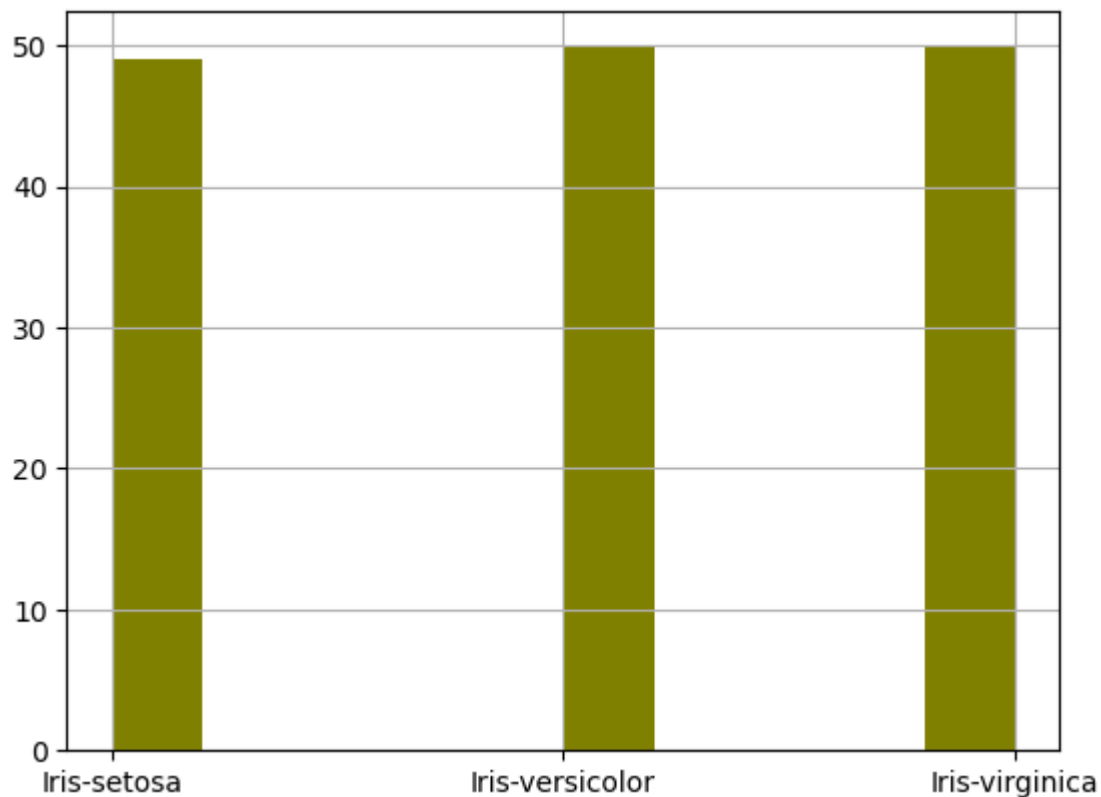
```
Out[ ]: <AxesSubplot:title={'center':'Frequency distribution Of Petal Length CM'}, xlabel='Length', ylabel='Frequency'>
```



5.Species

```
In [ ]: df["Species"].hist(color="olive")
```

```
Out[ ]: <AxesSubplot:>
```

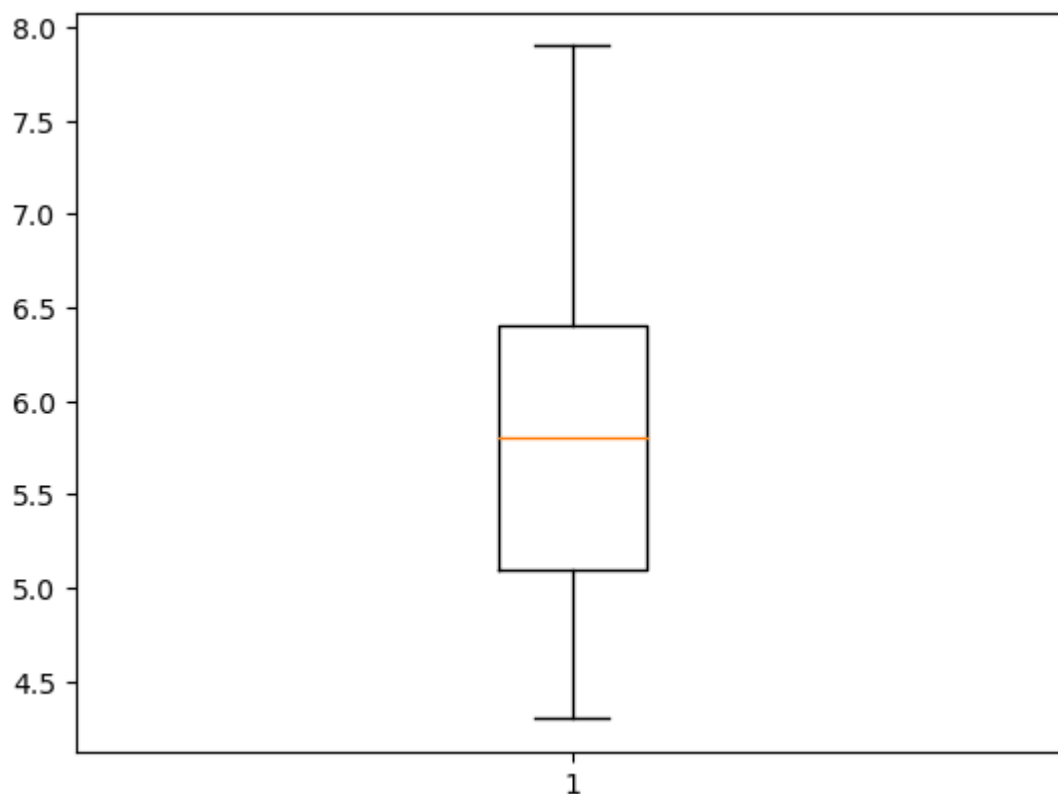


Creating a Box Plot For Each Feature In The Dataset

1.Sepal Lenght

```
In [ ]: pt.boxplot(df["SepalLengthCm"])
```

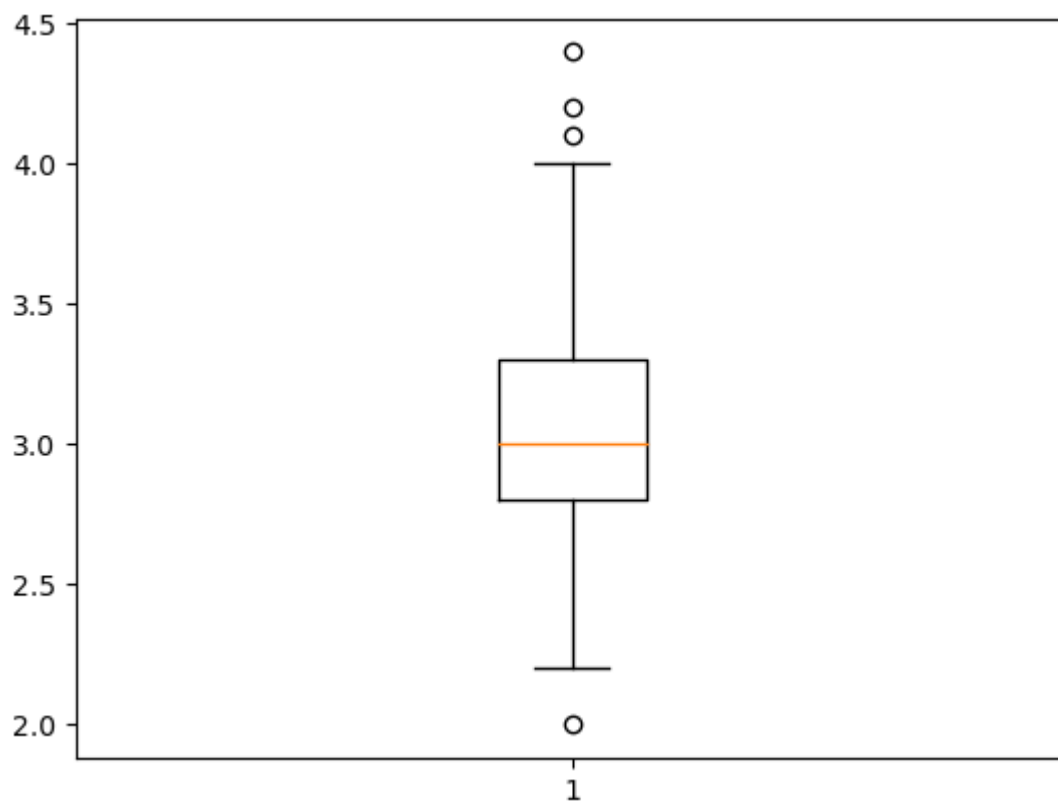
```
Out[ ]: {'whiskers': [<matplotlib.lines.Line2D at 0x21318723250>,  
  <matplotlib.lines.Line2D at 0x21318723520>],  
  'caps': [<matplotlib.lines.Line2D at 0x213187237f0>,  
  <matplotlib.lines.Line2D at 0x21318723ac0>],  
  'boxes': [<matplotlib.lines.Line2D at 0x21318716fa0>],  
  'medians': [<matplotlib.lines.Line2D at 0x21318723d90>],  
  'fliers': [<matplotlib.lines.Line2D at 0x213187300a0>],  
  'means': []}
```

2.Sepal Width

```
In [ ]: pt.boxplot(df["SepalWidthCm"])
```

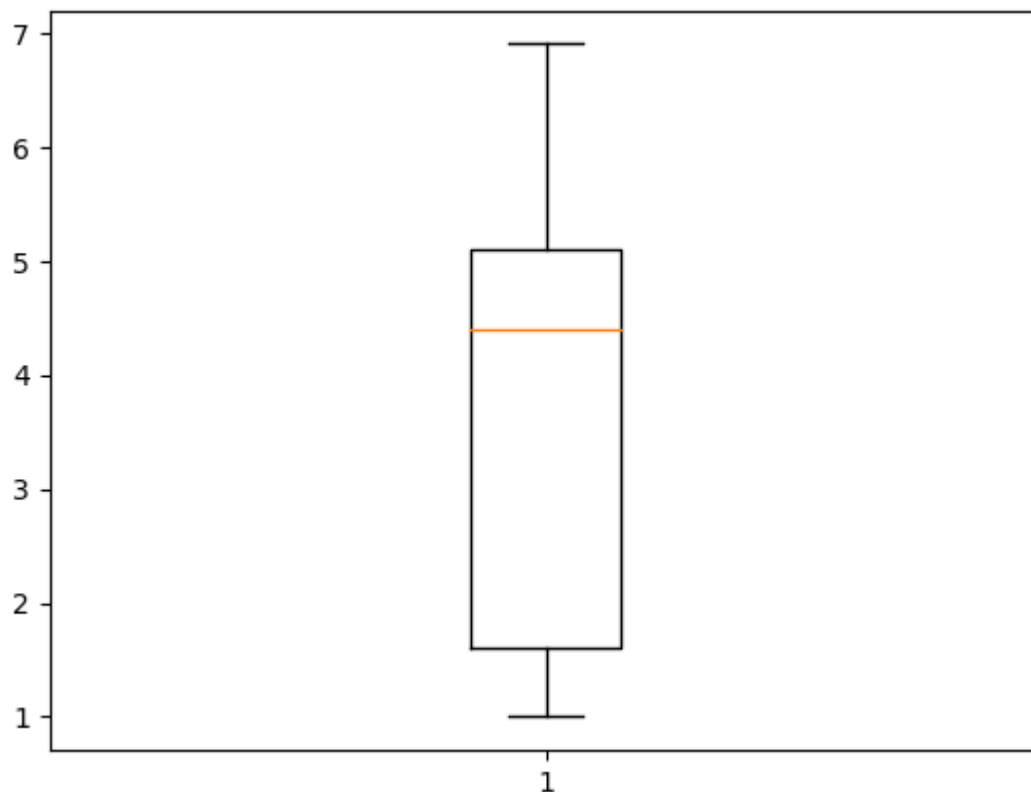
```
Out[ ]: {'whiskers': [<matplotlib.lines.Line2D at 0x21318742e80>,  
                    <matplotlib.lines.Line2D at 0x21318755190>],  
         'caps': [<matplotlib.lines.Line2D at 0x21318755460>,  
                  <matplotlib.lines.Line2D at 0x21318755730>],  
         'boxes': [<matplotlib.lines.Line2D at 0x21318742cd0>],  
         'medians': [<matplotlib.lines.Line2D at 0x21318755a00>],  
         'fliers': [<matplotlib.lines.Line2D at 0x21318755cd0>],  
         'means': []}
```



3. Petal Length in Centimeter

```
In [ ]: pt.boxplot(df["PetalLengthCm"])
```

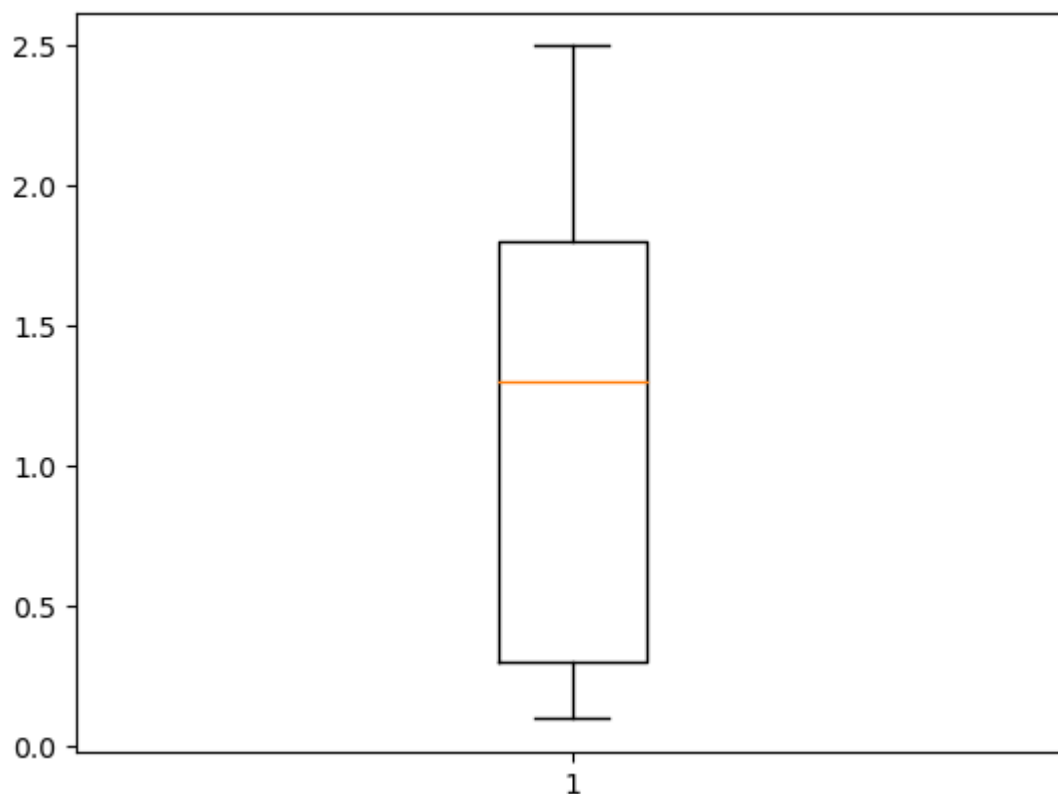
```
Out[ ]: {'whiskers': [<matplotlib.lines.Line2D at 0x213187b61f0>,
<matplotlib.lines.Line2D at 0x213187b64c0>],
'caps': [<matplotlib.lines.Line2D at 0x213187b6790>,
<matplotlib.lines.Line2D at 0x213187b6a60>],
'boxes': [<matplotlib.lines.Line2D at 0x213187a6ee0>],
'medians': [<matplotlib.lines.Line2D at 0x213187b6d30>],
'fliers': [<matplotlib.lines.Line2D at 0x213187c3040>],
'means': []}
```



4.Petal Width

```
In [ ]: pt.boxplot(df["PetalWidthCm"])
```

```
Out[ ]: {'whiskers': [<matplotlib.lines.Line2D at 0x21318815790>,  
  <matplotlib.lines.Line2D at 0x21318815a60>],  
  'caps': [<matplotlib.lines.Line2D at 0x21318815d30>,  
  <matplotlib.lines.Line2D at 0x21318823040>],  
  'boxes': [<matplotlib.lines.Line2D at 0x21318815490>],  
  'medians': [<matplotlib.lines.Line2D at 0x21318823310>],  
  'fliers': [<matplotlib.lines.Line2D at 0x213188235e0>],  
  'means': []}
```



Comparing All Boxplots

```
In [ ]: df[{"SepalLengthCm", "SepalWidthCm", "PetalLengthCm", "PetalWidthCm"}].boxplot()  
pt.title("Comparing all boxplots")
```

C:\Users\student\AppData\Local\Temp\ipykernel_5380\2625316568.py:1: FutureWarning: Passing a set as an indexer is deprecated and will raise in a future version. Use a list instead.

```
df[{"SepalLengthCm", "SepalWidthCm", "PetalLengthCm", "PetalWidthCm"}].boxplot()  
Out[ ]: Text(0.5, 1.0, 'Comparing all boxplots')
```

