Import libraries

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Load dataset

In [3]:
```
df=pd.read_csv("Iris.csv")
df
```

Out[3]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

1. list down features and their types

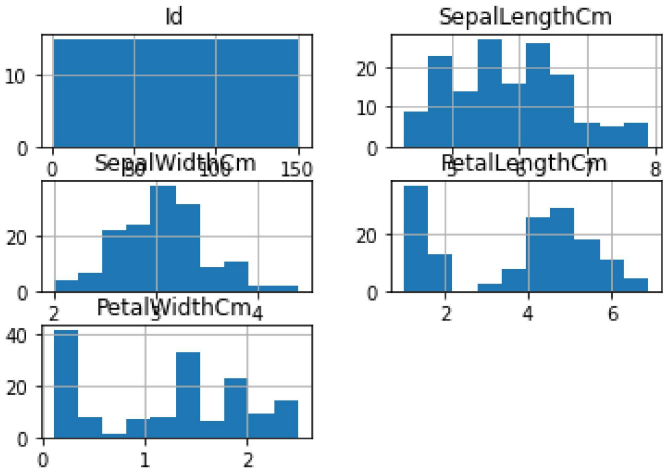
In [4]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   Id             150 non-null    int64  
 1   SepalLengthCm  150 non-null    float64
 2   SepalWidthCm   150 non-null    float64
 3   PetalLengthCm  150 non-null    float64
 4   PetalWidthCm   150 non-null    float64
 5   Species        150 non-null    object 
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
```

2.create a histogram for each feature in dataset to illustrate the feature distribution

In [10]:
```
1 df.hist()
```

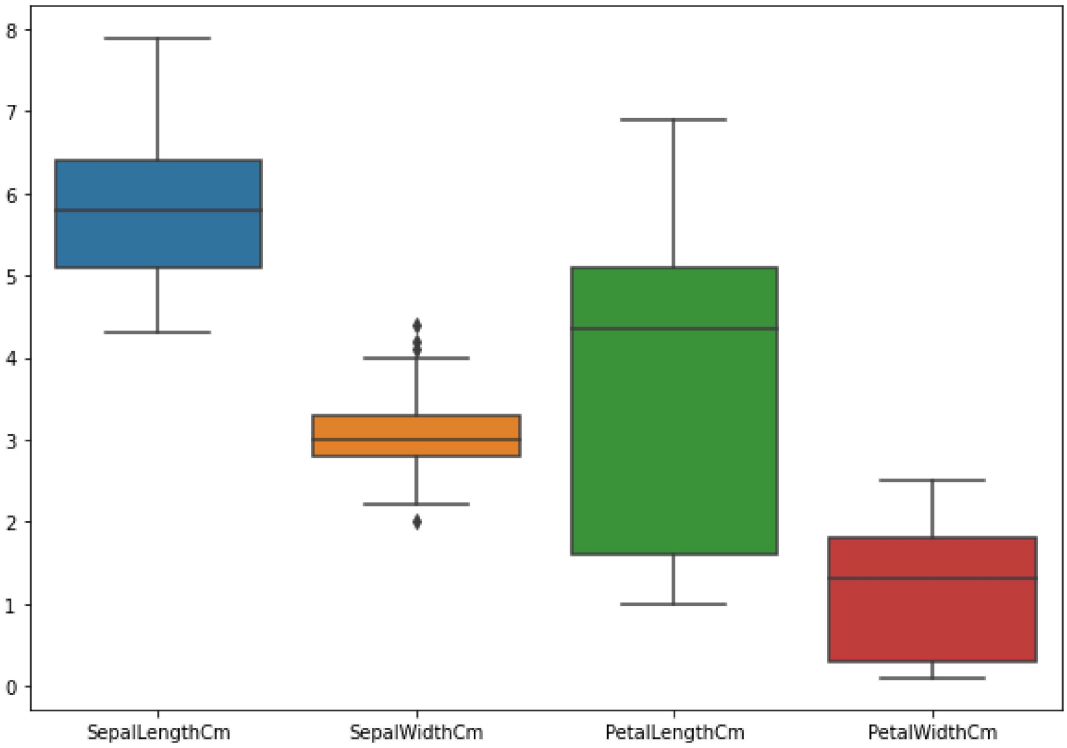
Out[10]:
```
array([[<AxesSubplot:title={'center':'Id'}>,
        <AxesSubplot:title={'center':'SepalLengthCm'}>],
       [<AxesSubplot:title={'center':'SepalWidthCm'}>,
        <AxesSubplot:title={'center':'PetalLengthCm'}>],
       [<AxesSubplot:title={'center':'PetalWidthCm'}>, <AxesSubplot:>]],
      dtype=object)
```



3.create a Boxplot for each feature in the dataset

In [18]:
```
1 plt.figure(figsize=(10,7))
2 sns.boxplot(data=df[['SepalLengthCm','SepalWidthCm','PetalLengthCm','PetalWidthCm']]
```

Out[18]:
```
<AxesSubplot:>
```



4.compare distributions and identify outliers

In [20]:
```
Q1 = df['SepalWidthCm'].quantile(0.25)
Q3 = df['SepalWidthCm'].quantile(0.75)
```

In [21]:
```
iqr = Q3 - Q1 #Interquartile range
minm= Q1 - (1.5*iqr)
maxm = Q3 + (1.5*iqr)
iqr
```

Out[21]: 0.5

In [22]:
```
minm
```

Out[22]: 2.05

In [23]:
```
maxm
```

Out[23]: 4.05

In [24]:
```
df2=df[(df['SepalWidthCm']>minm) & (df['SepalWidthCm']<maxm)]
df2
```

Out[24]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...	...	...	...	...	...	...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

146 rows × 6 columns

In [25]:
```
print("After Shape:",df2.shape)
```

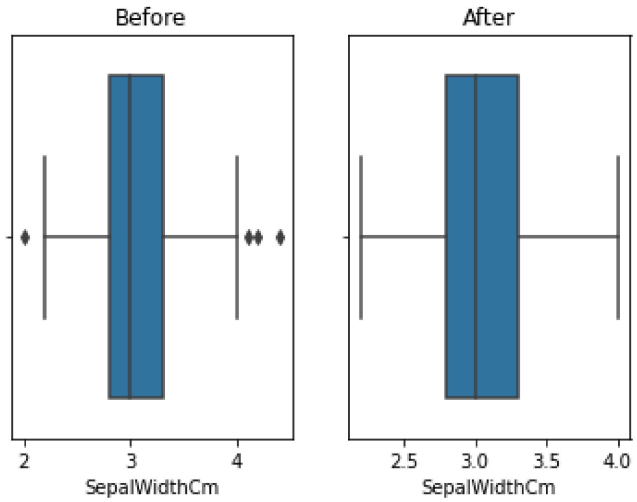
```
After Shape: (146, 6)
```

In [26]:
```
df.shape
```

Out[26]: (150, 6)

In [27]:
```
import warnings
warnings.filterwarnings("ignore")
fig, axes= plt.subplots(1,2)

sns.boxplot(df['SepalWidthCm'],ax=axes[0])
axes[0].title.set_text('Before')
sns.boxplot(df2['SepalWidthCm'],ax=axes[1])
axes[1].title.set_text('After')
plt.show()
```



In [ ]:
```
1
```