```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('Downloads/Iris.csv')

df
```

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

            Species
0       Iris-setosa
1       Iris-setosa
2       Iris-setosa
3       Iris-setosa
4       Iris-setosa
..              ...
145  Iris-virginica
146  Iris-virginica
147  Iris-virginica
148  Iris-virginica
149  Iris-virginica

[150 rows x 6 columns]
```

```python
df.head(5)
```

```
   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
```

```
df.describe()
```

```
                Id  SepalLengthCm  SepalWidthCm  PetalLengthCm
PetalWidthCm
count  150.000000     150.000000    150.000000     150.000000
150.000000
mean    75.500000       5.843333      3.054000       3.758667
1.198667
std     43.445368       0.828066      0.433594       1.764420
0.763161
min      1.000000       4.300000      2.000000       1.000000
0.100000
25%     38.250000       5.100000      2.800000       1.600000
0.300000
50%     75.500000       5.800000      3.000000       4.350000
1.300000
75%    112.750000       6.400000      3.300000       5.100000
1.800000
max    150.000000       7.900000      4.400000       6.900000
2.500000
```

```
df.dtypes
```

```
Id                int64
SepalLengthCm    float64
SepalWidthCm     float64
PetalLengthCm    float64
PetalWidthCm     float64
Species           object
dtype: object
```

```
df.isnull().sum()
```

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

```
df['Species'].unique()
```

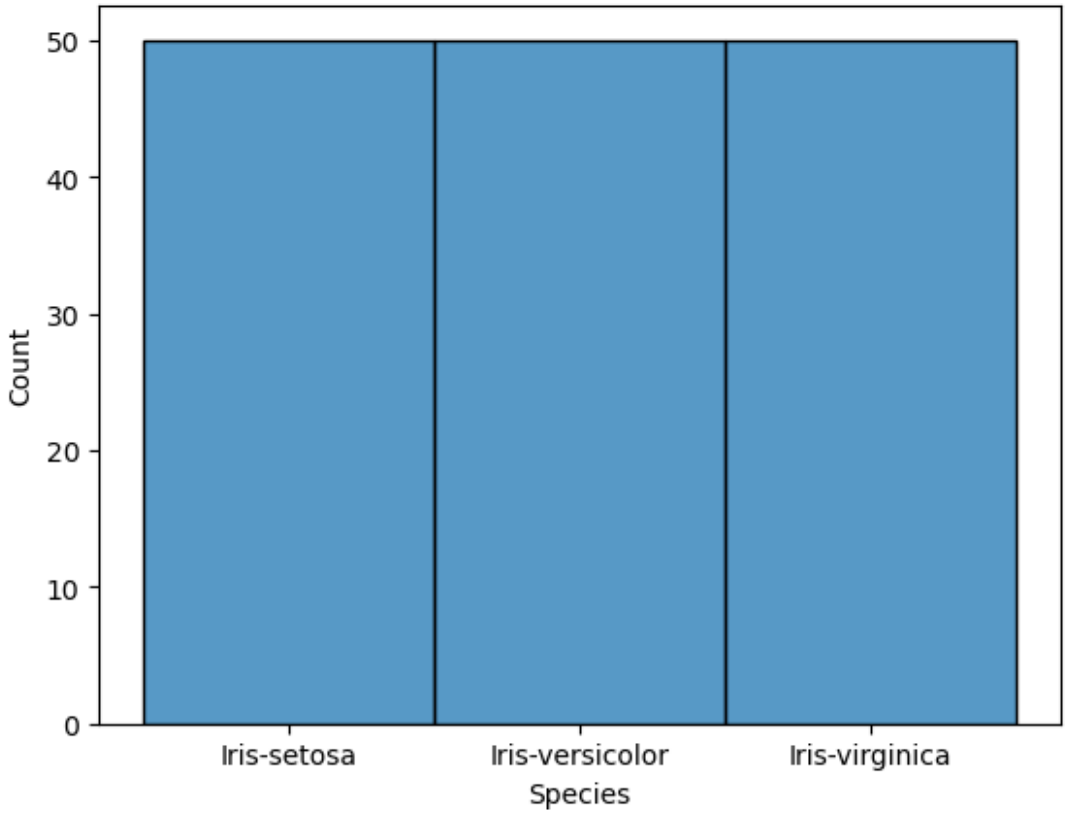
```
array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'],
dtype=object)
```

Histogram

Histogram - Species

```
sns.histplot(data=df, x='Species')
```

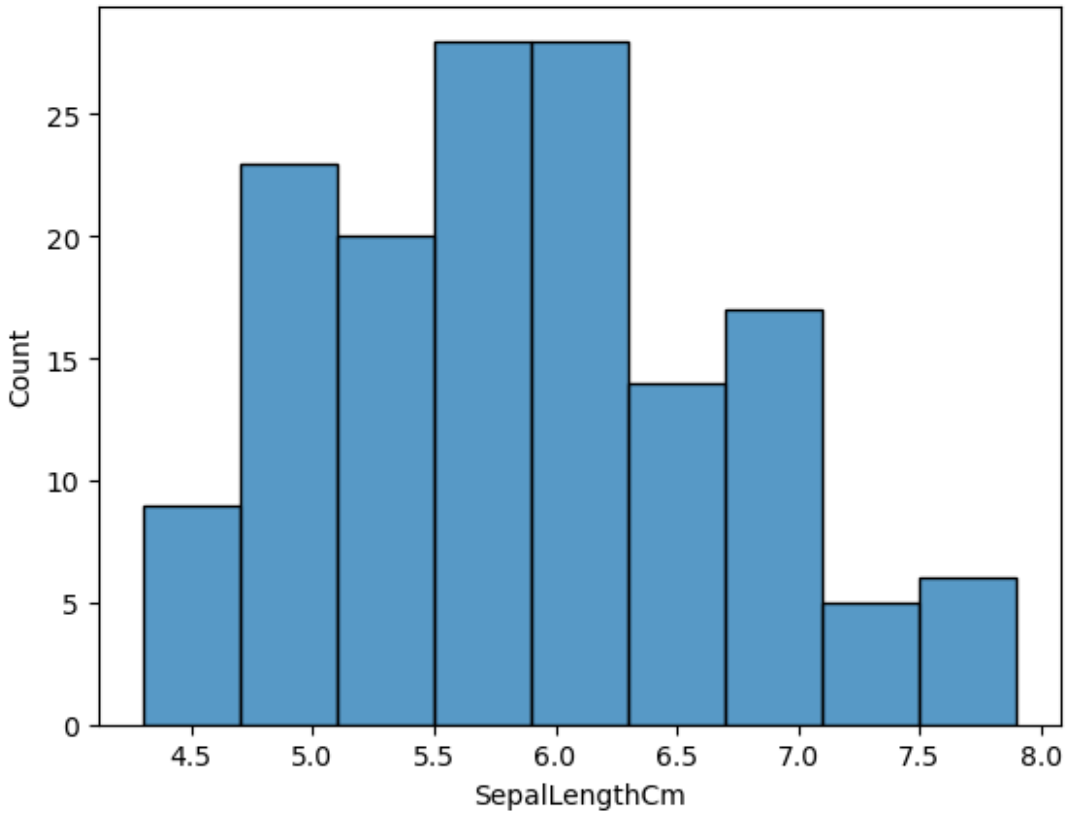
```
<Axes: xlabel='Species', ylabel='Count'>
```



Histogram - SepalLengthCM

```
sns.histplot(data=df, x="SepalLengthCm")
```

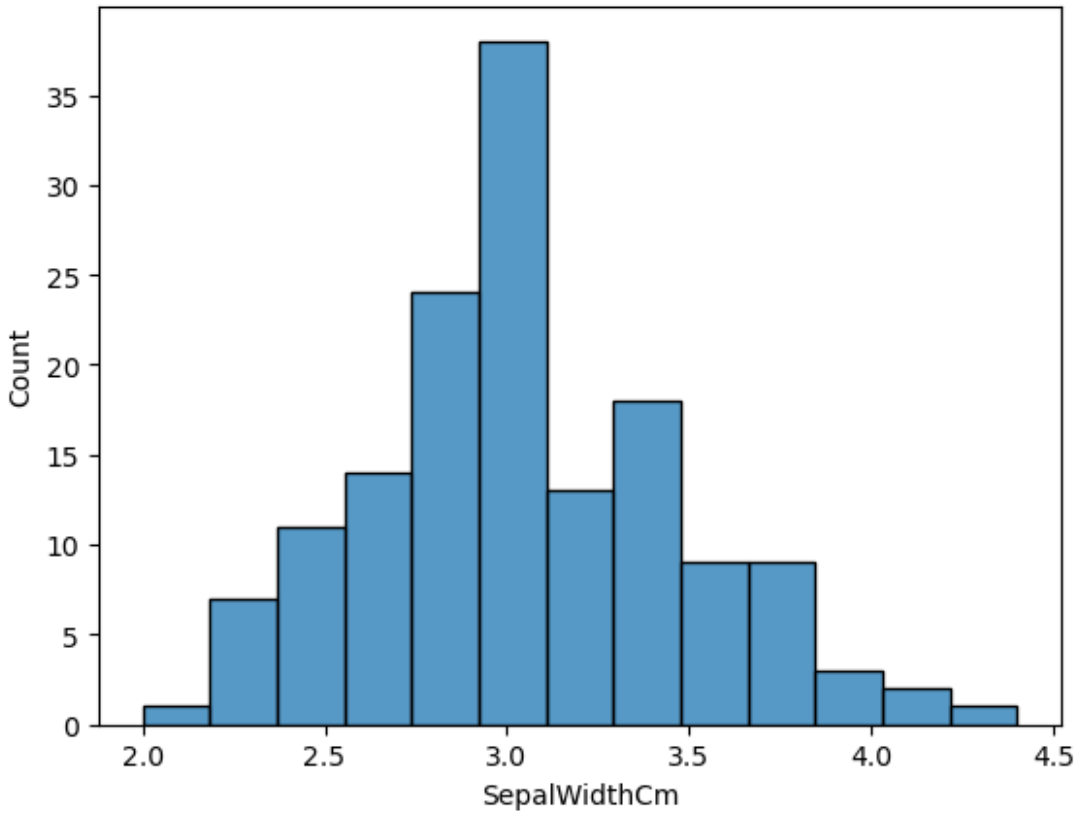
```
<Axes: xlabel='SepalLengthCm', ylabel='Count'>
```



Histogram - SepalWidthCm

```
sns.histplot(data=df, x='SepalWidthCm')
```

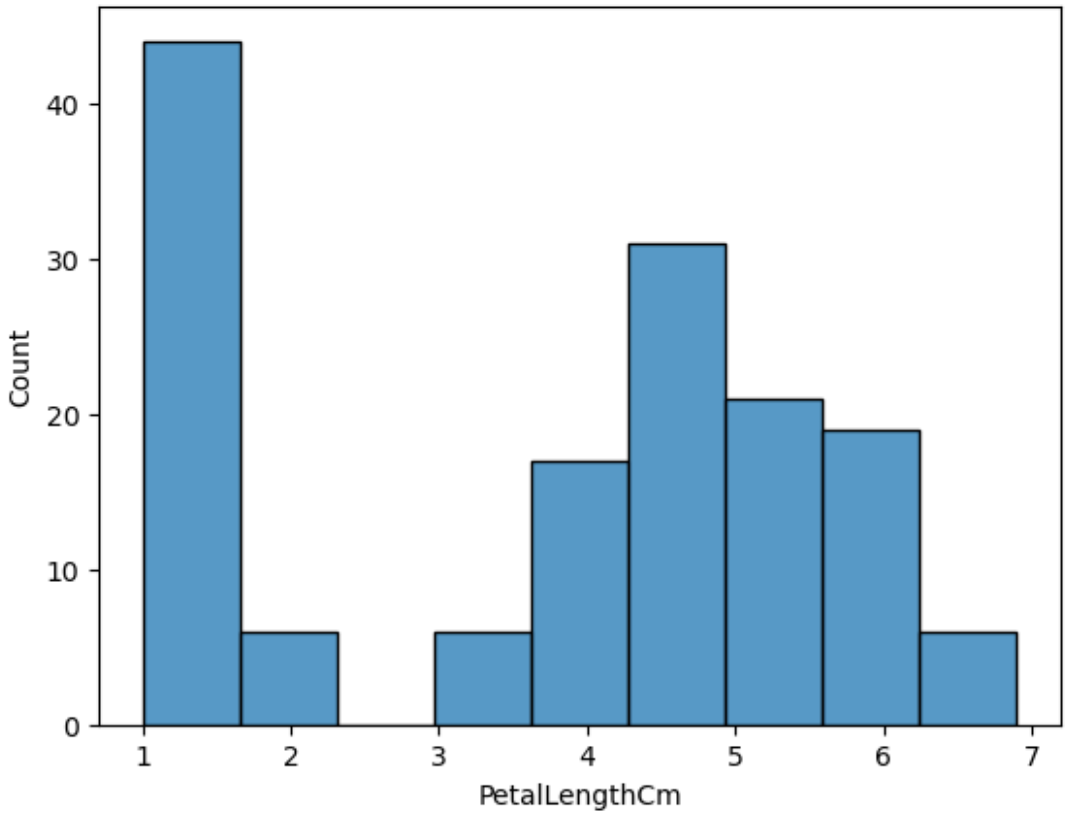
```
<Axes: xlabel='SepalWidthCm', ylabel='Count'>
```



Histogram - PetalLengthCM

```
sns.histplot(data=df, x='PetalLengthCm')
```

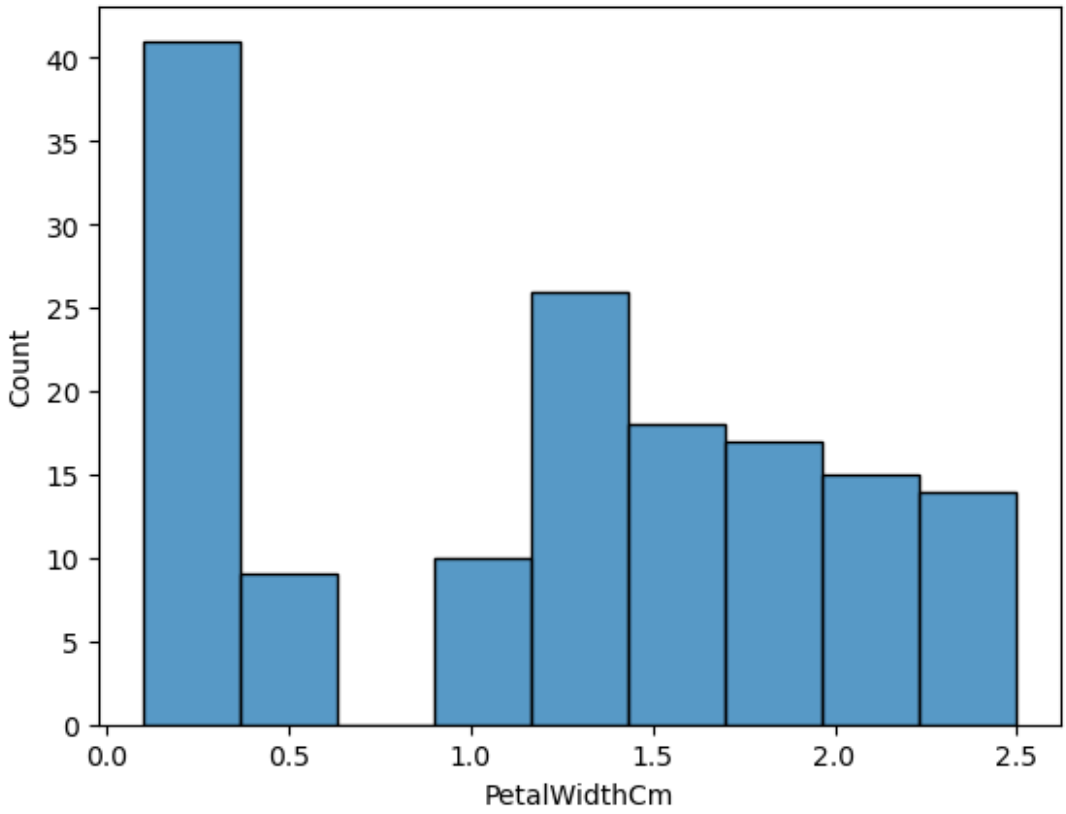
```
<Axes: xlabel='PetalLengthCm', ylabel='Count'>
```



Histogram - PetalWidthCM

```
sns.histplot(data=df, x='PetalWidthCm')
```

```
<Axes: xlabel='PetalWidthCm', ylabel='Count'>
```

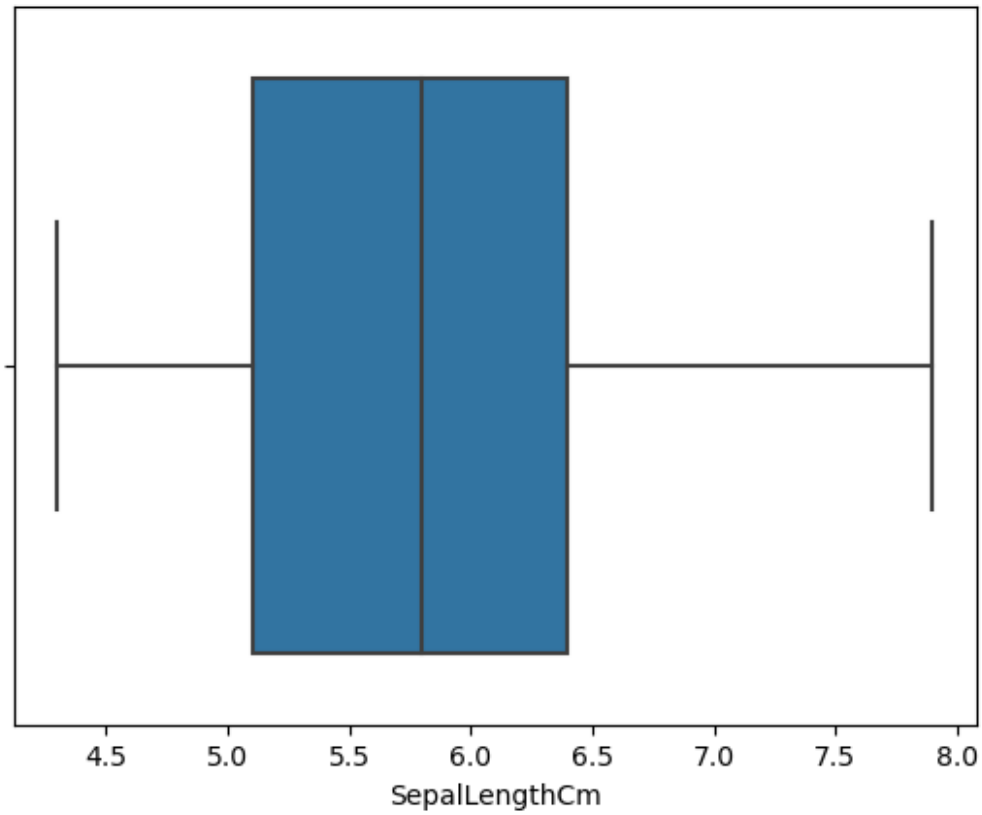


BoxPlot

SepalLengthCm

```
sns.boxplot(x=df['SepalLengthCm'])
```

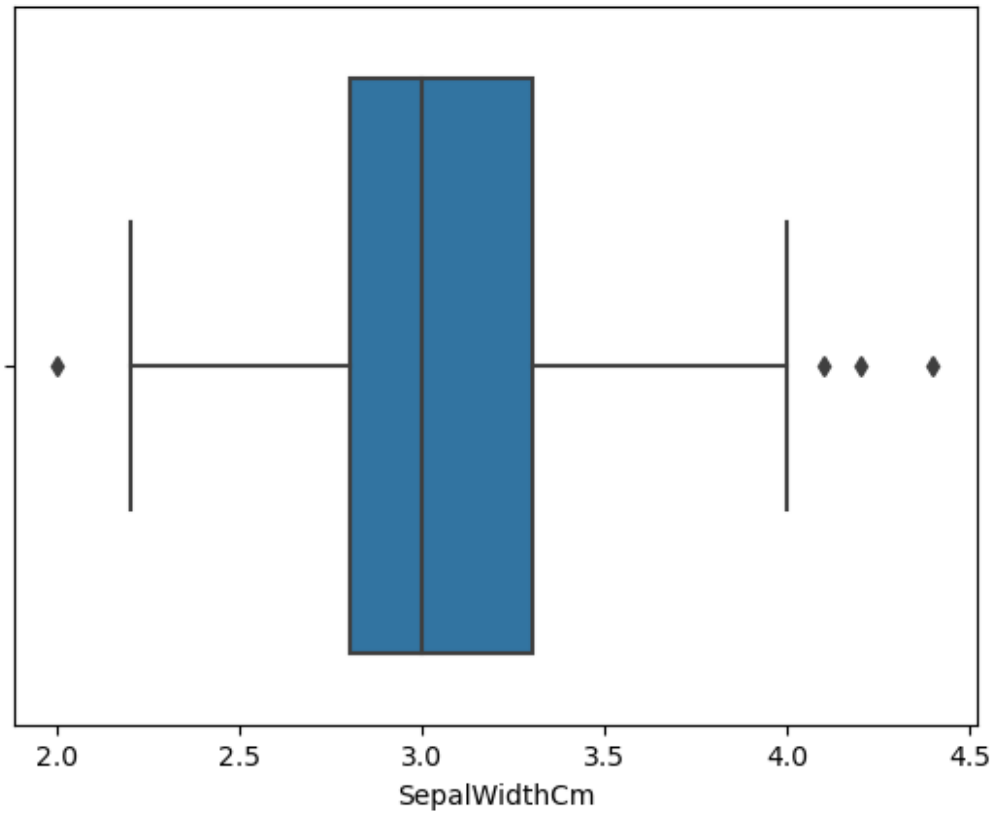
```
<Axes: xlabel='SepalLengthCm'>
```



SepalWidthCm (Outliers)

```
sns.boxplot(x=df['SepalWidthCm'])
```

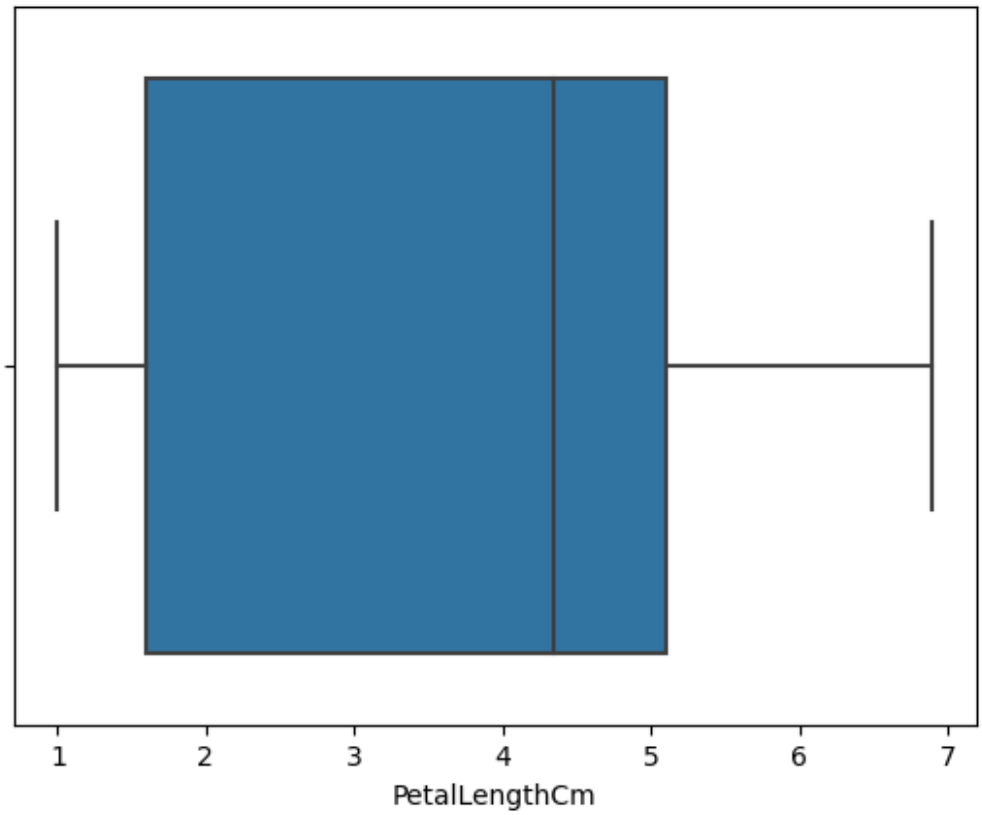
```
<Axes: xlabel='SepalWidthCm'>
```



PetalLengthCm

```
sns.boxplot(x=df['PetalLengthCm'])
```

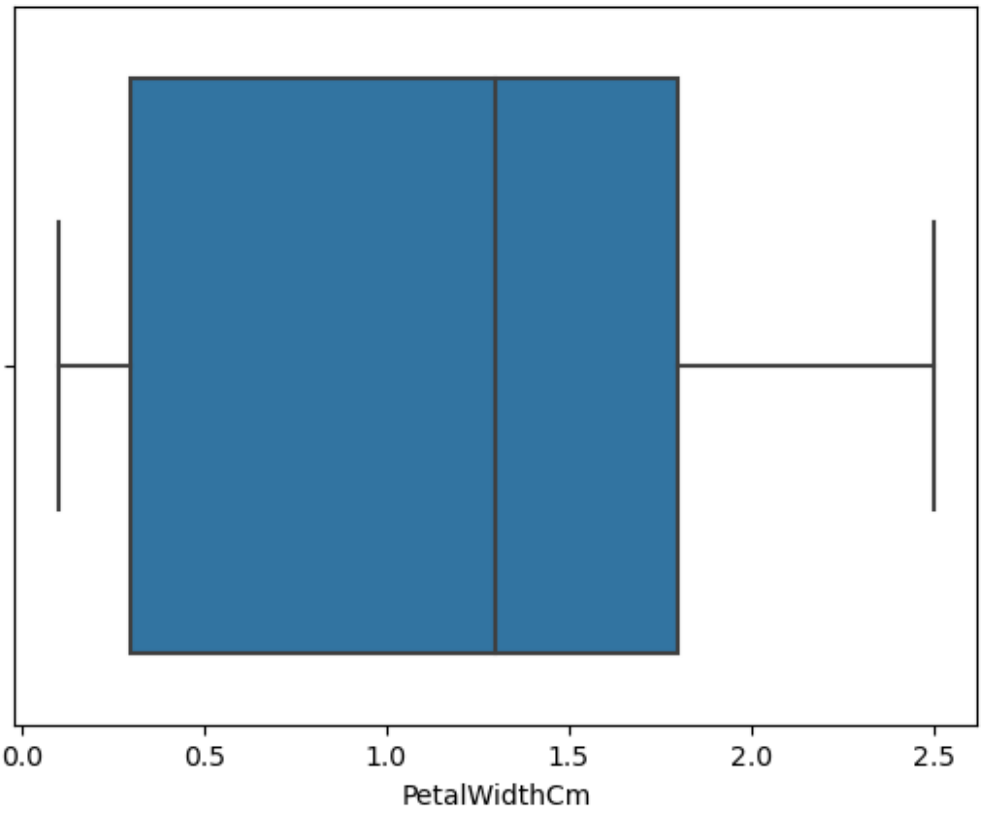
```
<Axes: xlabel='PetalLengthCm'>
```



PetalWidthCm

```
sns.boxplot(x=df['PetalWidthCm'])
```

```
<Axes: xlabel='PetalWidthCm'>
```

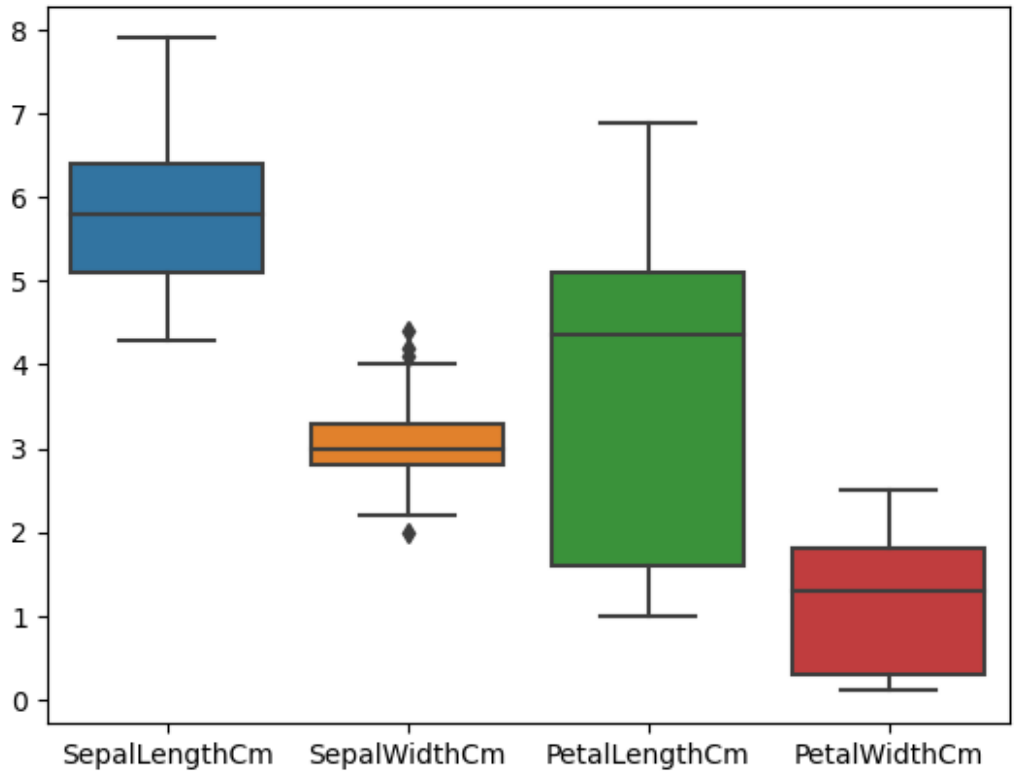



```
data2 = df.iloc[:, 1:]
```

```
figure = plt.figure(figsize = (12,8))
```

```
<Figure size 1200x800 with 0 Axes>
```

```
sns.boxplot(data= data2)
plt.show()
```



Identifying the outliers

```
from matplotlib.cbook import boxplot_stats
stats = boxplot_stats(df['SepalWidthCm'])
stats
```

```
[{'mean': 3.0540000000000003,
  'iqr': 0.5,
  'cilo': 2.9359050183971735,
  'cihi': 3.0640949816028265,
  'whishi': 4.0,
  'whislo': 2.2,
  'fliers': array([2. , 4.4, 4.1, 4.2]),
  'q1': 2.8,
  'med': 3.0,
  'q3': 3.3}]
```

```
outliers = stats[0].get('fliers')
```

```
print(outliers)
```

```
[2.  4.4 4.1 4.2]
```

Removing the outliers

```
Q1=df['SepalWidthCm'].quantile(0.25)
Q3=df['SepalWidthCm'].quantile(0.75)
Q1,Q3
```

```
(2.8, 3.3)
```

```
IQR = Q3-Q1
```

```
lower_limit=Q1-1.5*IQR
upper_limit=Q3+1.5*IQR
```

```
df[(df['SepalWidthCm']<lower_limit)|(df['SepalWidthCm']>upper_limit)]
```

```
    Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  \
15  16            5.7           4.4            1.5           0.4
32  33            5.2           4.1            1.5           0.1
33  34            5.5           4.2            1.4           0.2
60  61            5.0           2.0            3.5           1.0

           Species
15     Iris-setosa
32     Iris-setosa
33     Iris-setosa
60  Iris-versicolor
```

```
df_without_outliers=df[(df['SepalWidthCm']>lower_limit)&(df['SepalWidt
hCm']<upper_limit)]
df_without_outliers
```

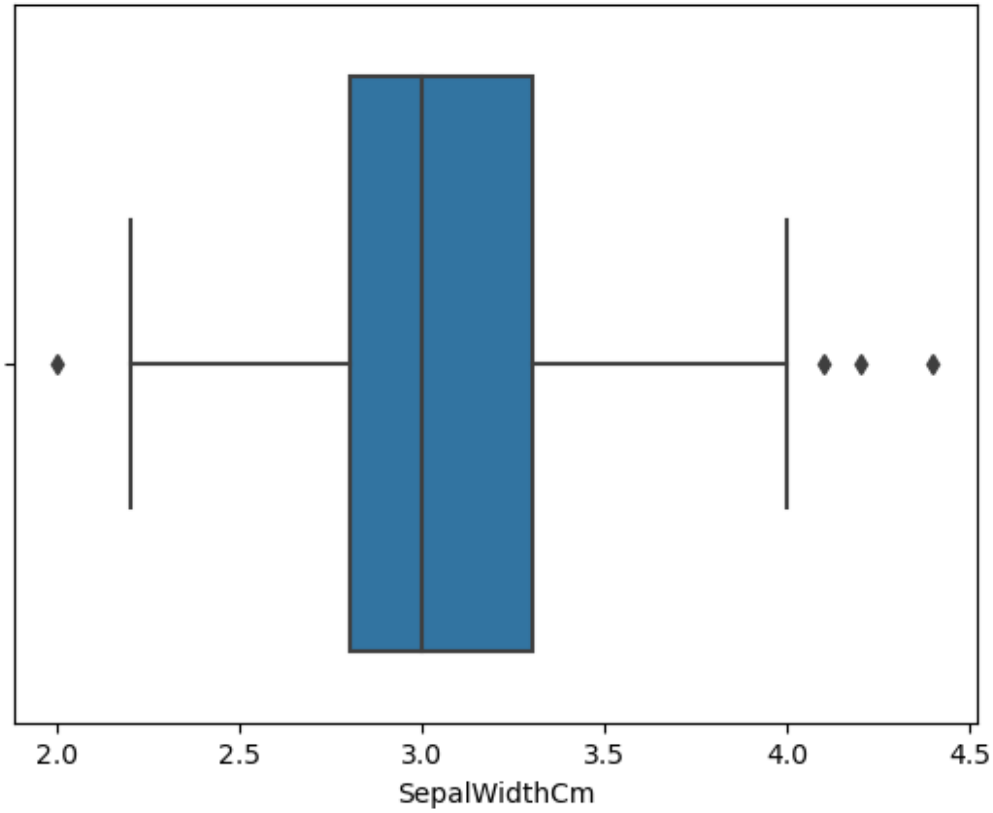
```
      Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  \
0      1            5.1           3.5            1.4           0.2
1      2            4.9           3.0            1.4           0.2
2      3            4.7           3.2            1.3           0.2
3      4            4.6  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

           Species
0      Iris-setosa
1      Iris-setosa
2      Iris-setosa
3      Iris-setosa
4      Iris-setosa
..             ...
145  Iris-virginica
146  Iris-virginica
147  Iris-virginica
148  Iris-virginica
149  Iris-virginica
```

[146 rows x 6 columns]

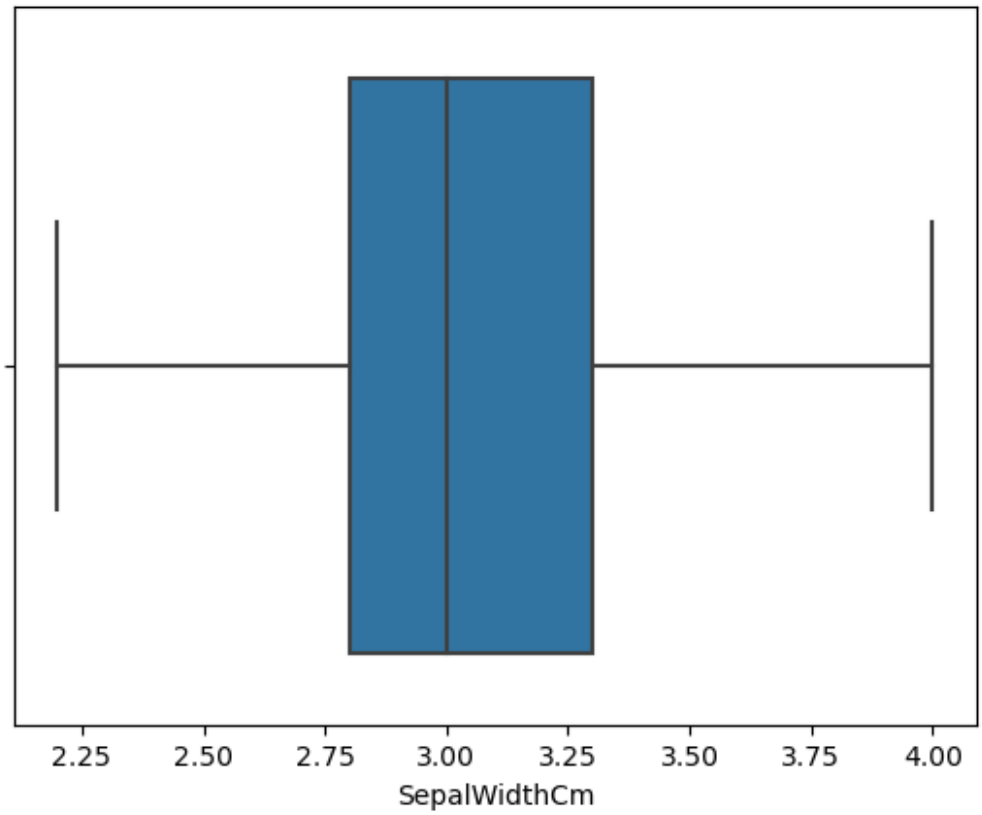
```
sns.boxplot(x=df['SepalWidthCm'])
```

```
<Axes: xlabel='SepalWidthCm'>
```




```
sns.boxplot(x=df_without_outliers['SepalWidthCm'])
```

```
<Axes: xlabel='SepalWidthCm'>
```


2.25
2.50
2.75
3.00
3.25
3.50
3.75
4.00
SepalWidthCm