

In [25]: *#11. Data Visualization III*  
*#Download the Iris flower dataset or any other dataset into a Dataframe*  
*#1.List down the features and their types (e.g., numeric, nominal) available in the dataset.*  
*#2.Create a histogram for each feature in the dataset to illustrate the feature distributions.*  
*#3.Create a box plot for each feature in the dataset.*  
*#4.Compare distributions and identify outliers.*

*#Iris.csv*

In [26]: `import numpy as np`  
`import pandas as pd`  
`import seaborn as sns`  
`import matplotlib.pyplot as plt`

In [27]: `df1 = pd.read_csv("Iris.csv")`

In [28]: `df1`

Out[28]:

	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

```
In [29]: df1.rename(columns={"SepalLengthCm": "sepal_length",
                             "SepalWidthCm": "sepal_width",
                             "PetalLengthCm": "petal_length",
                             "PetalWidthCm": "petal_width",
                             "Species": "species"}, inplace=True)

df1.head()
```

```
Out[29]:
```

	Id	sepal_length	sepal_width	petal_length	petal_width	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

```
In [30]: df = pd.DataFrame(df1)
df.head()
```

```
Out[30]:
```

	Id	sepal_length	sepal_width	petal_length	petal_width	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

```
In [31]: df.describe()
```

```
Out[31]:
```

	Id	sepal_length	sepal_width	petal_length	petal_width
<b>count</b>	150.000000	150.000000	150.000000	150.000000	150.000000
<b>mean</b>	75.500000	5.843333	3.054000	3.758667	1.198667
<b>std</b>	43.445368	0.828066	0.433594	1.764420	0.763161
<b>min</b>	1.000000	4.300000	2.000000	1.000000	0.100000
<b>25%</b>	38.250000	5.100000	2.800000	1.600000	0.300000
<b>50%</b>	75.500000	5.800000	3.000000	4.350000	1.300000
<b>75%</b>	112.750000	6.400000	3.300000	5.100000	1.800000
<b>max</b>	150.000000	7.900000	4.400000	6.900000	2.500000

```
In [32]: 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   sepal_length    150 non-null   float64   
2   sepal_width     150 non-null   float64   
3   petal_length    150 non-null   float64   
4   petal_width     150 non-null   float64   
5   species         150 non-null   object    
dtypes: float64(4), int64(1), object(1)  
memory usage: 7.2+ KB
```

```
In [33]: df.columns
```

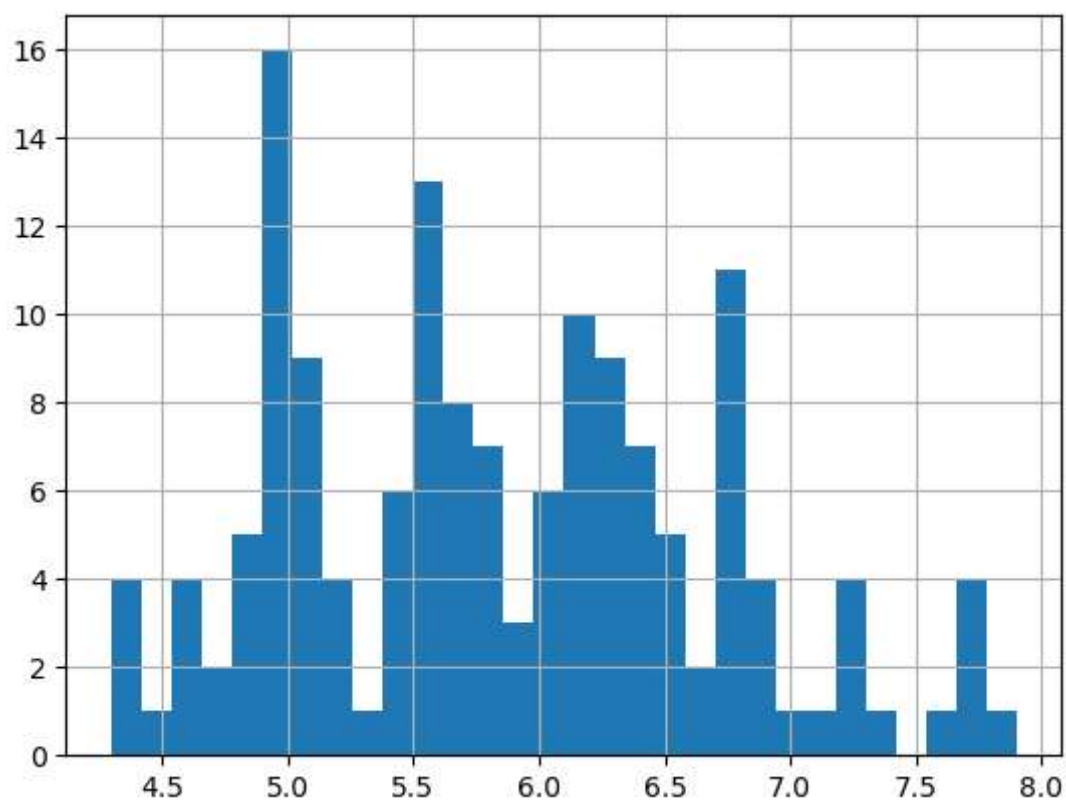
```
Out[33]: Index(['Id', 'sepal_length', 'sepal_width', 'petal_length', 'petal_width',  
              'species'],  
              dtype='object')
```

```
In [13]: df['sepal_length'].max()
```

```
Out[13]: 7.9
```

```
In [34]: df['sepal_length'].hist(bins=30)
```

```
Out[34]: <AxesSubplot:>
```



```
In [36]: df['petal_length'].max()
```

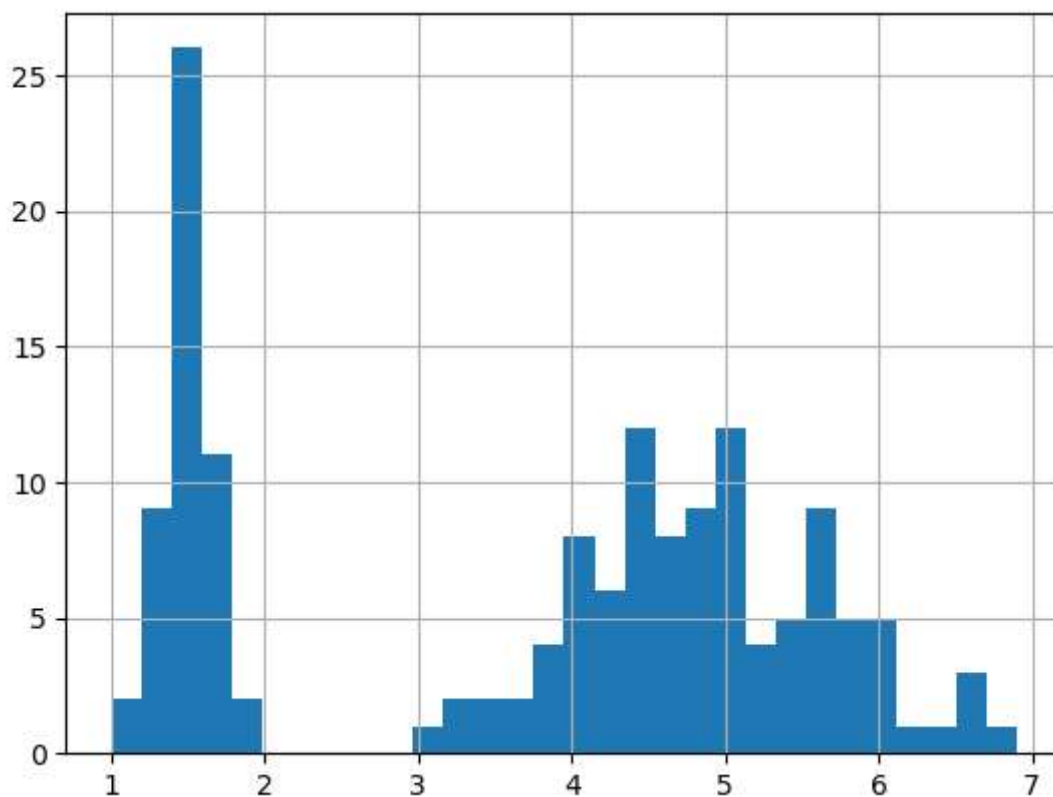
```
Out[36]: 6.9
```

```
In [16]: df['petal_length'].min()
```

```
Out[16]: 1.0
```

```
In [37]: df['petal_length'].hist(bins=30)
```

```
Out[37]: <AxesSubplot:>
```



```
In [39]: df['petal_width'].max()
```

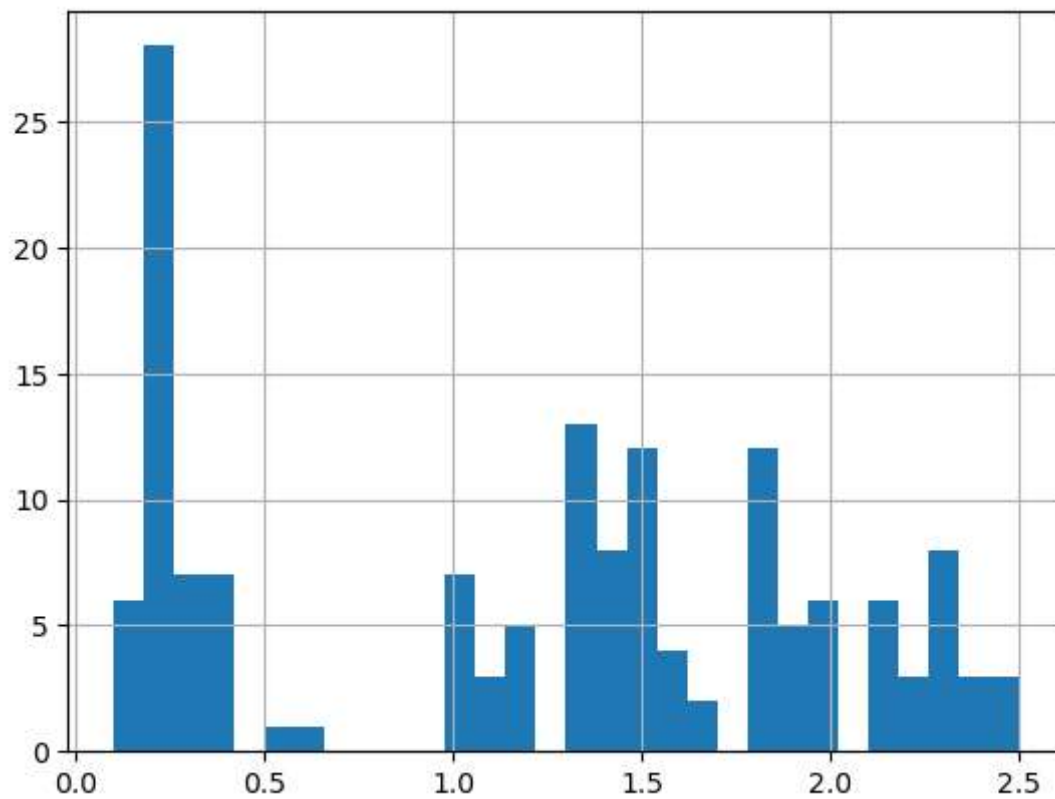
```
Out[39]: 2.5
```

```
In [40]: df['petal_width'].min()
```

```
Out[40]: 0.1
```

```
In [41]: df['petal_width'].hist(bins=30)
```

```
Out[41]: <AxesSubplot:>
```



```
In [42]: df['sepal_width'].max()
```

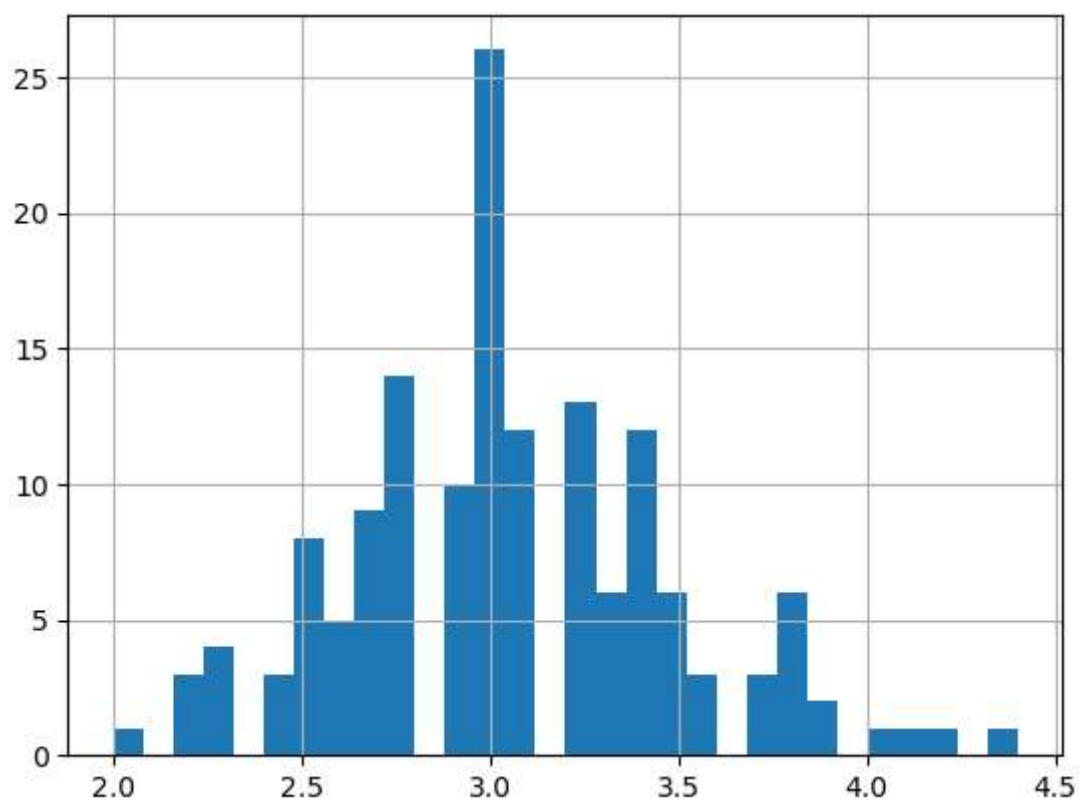
```
Out[42]: 4.4
```

```
In [43]: df['sepal_width'].min()
```

```
Out[43]: 2.0
```

```
In [44]: df['sepal_width'].hist(bins = 30)
```

```
Out[44]: <AxesSubplot:>
```

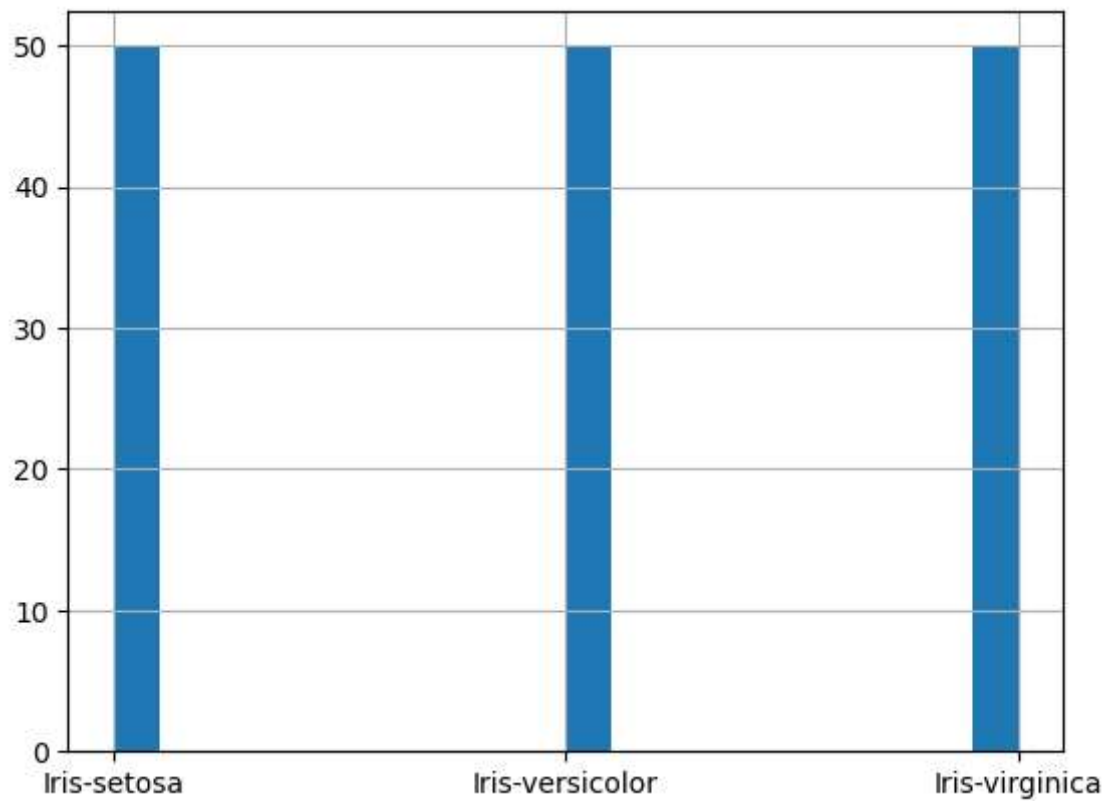


```
In [45]: df['species'].value_counts()
```

```
Out[45]: Iris-setosa      50  
Iris-versicolor    50  
Iris-virginica      50  
Name: species, dtype: int64
```

```
In [46]: df['species'].hist(bins=20)
```

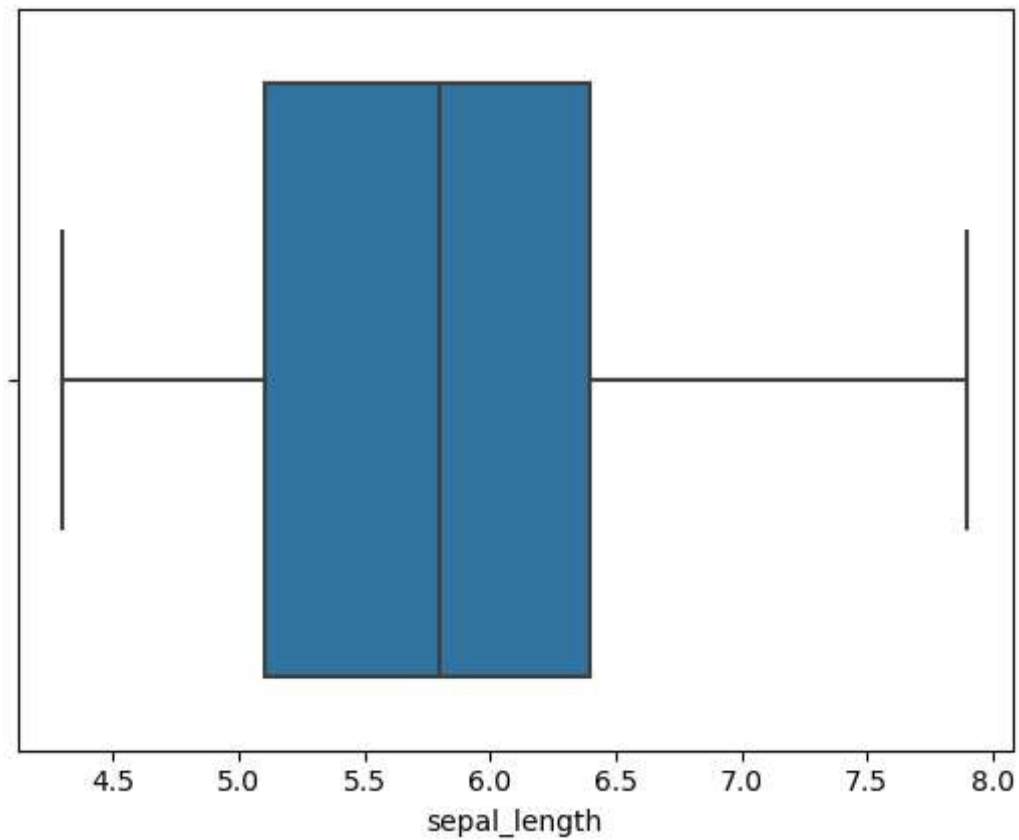
```
Out[46]: <AxesSubplot:>
```





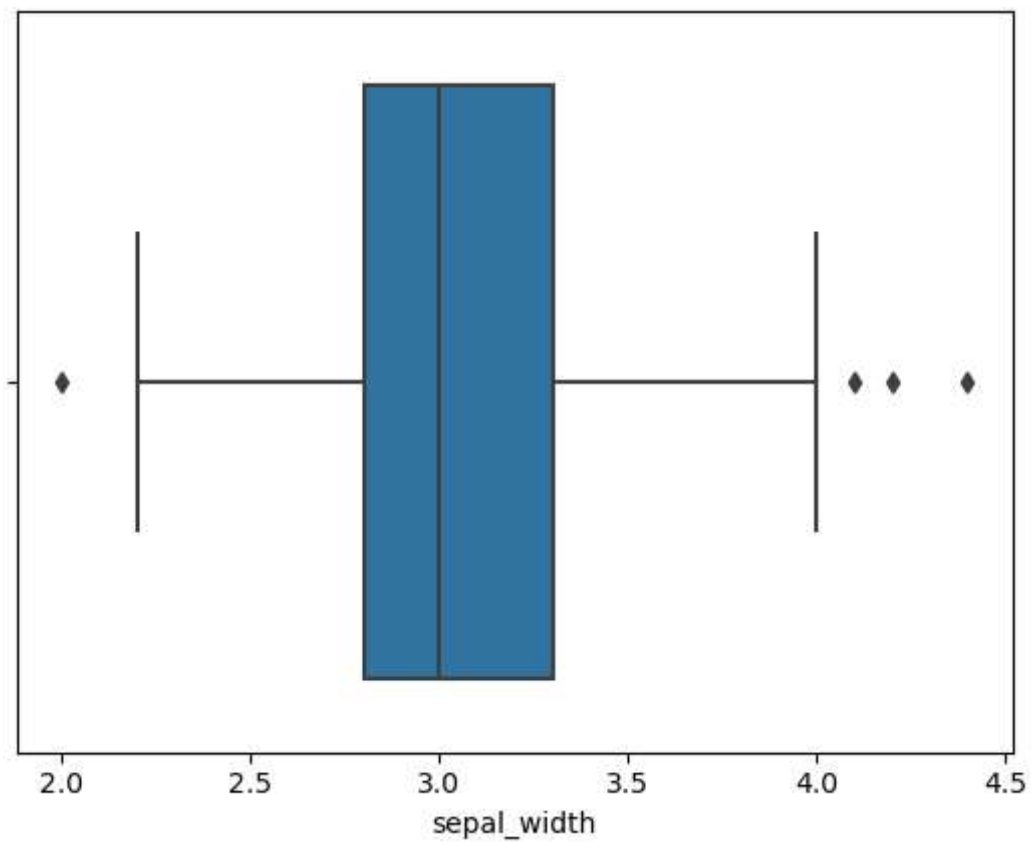
```
In [47]: sns.boxplot(x="sepal_length",data=df)
```

```
Out[47]: <AxesSubplot:xlabel='sepal_length'>
```



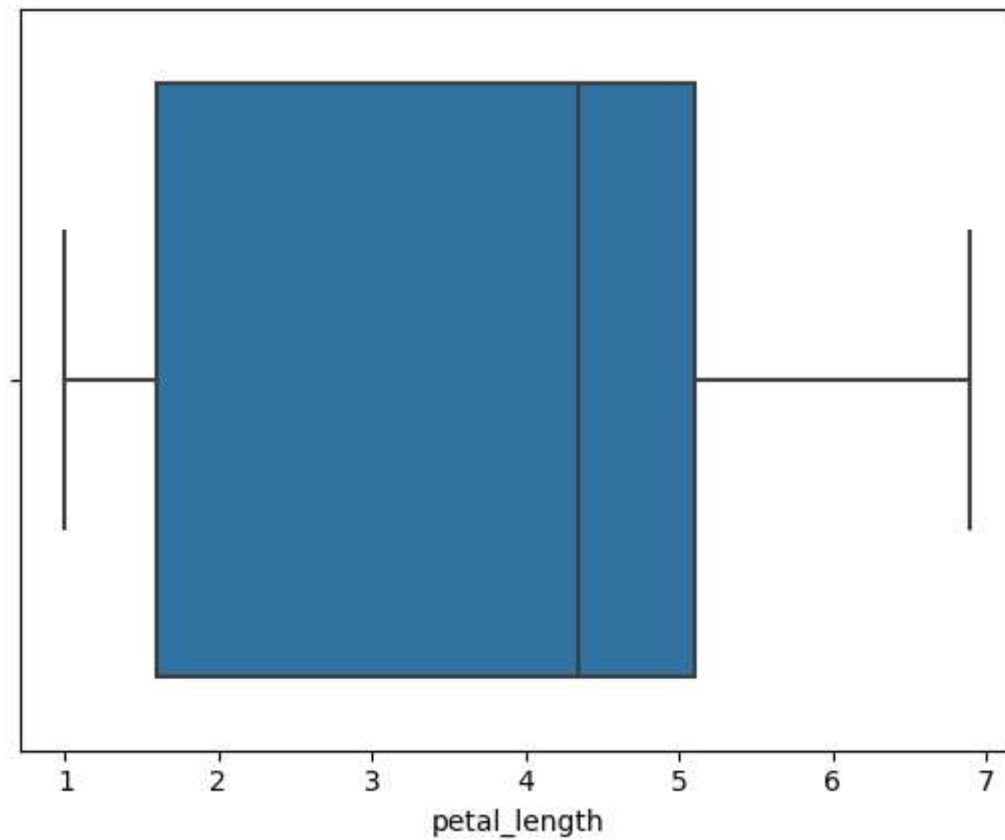
```
In [48]: sns.boxplot(x="sepal_width",data=df)
```

```
Out[48]: <AxesSubplot:xlabel='sepal_width'>
```



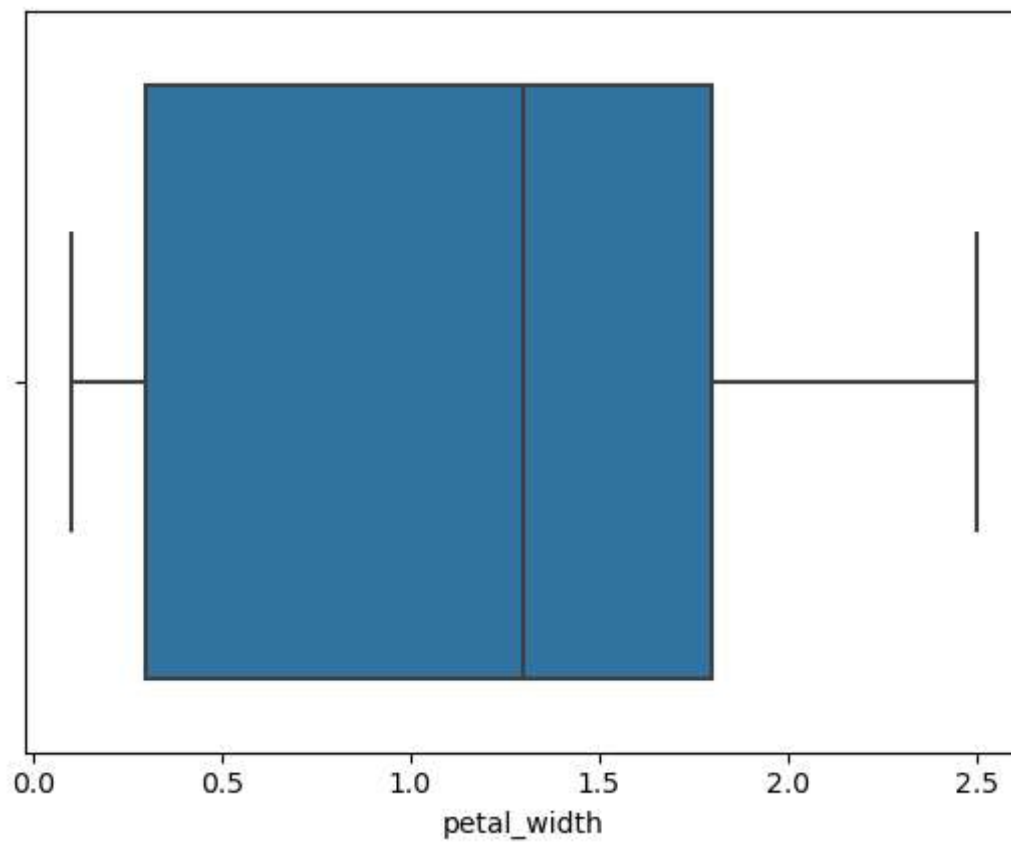
```
In [49]: sns.boxplot(x='petal_length',data=df)
```

```
Out[49]: <AxesSubplot:xlabel='petal_length'>
```



```
In [50]: sns.boxplot(x='petal_width',data=df)
```

```
Out[50]: <AxesSubplot:xlabel='petal_width'>
```



```
In [ ]:
```