

✓ DSBDAL Assignment 06 - Data Visualization 3

Download the Iris flower dataset or any other dataset into a DataFrame. (e.g., <https://archive.ics.uci.edu/ml/datasets/Iris>). Scan the dataset and give the inference as:




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. Compare distributions and identify outliers.

```
from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

```
import numpy as np
import seaborn as sns
import pandas as pd
```

```
ds = pd.read_csv('/content/drive/My Drive/DSBDL/Assignment6/iris.csv')
ds
```

	sepal_length	sepal_width	petal_length	petal_width	species	
0	5.1	3.5	1.4	0.2	setosa	
1	4.9	3.0	1.4	0.2	setosa	
2	4.7	3.2	1.3	0.2	setosa	
3	4.6	3.1	1.5	0.2	setosa	
4	5.0	3.6	1.4	0.2	setosa	
...	
145	6.7	3.0	5.2	2.3	virginica	
146	6.3	2.5	5.0	1.9	virginica	
147	6.5	3.0	5.2	2.0	virginica	
148	6.2	3.4	5.4	2.3	virginica	
149	5.9	3.0	5.1	1.8	virginica	

150 rows × 5 columns

Next steps:

[Generate code with ds](#)



[View recommended plots](#)

✓ Feature Information



Numeric features: sepal_width, sepal_length, petal_length and petal_width

Nominal features: species

```
ds.dtypes
```

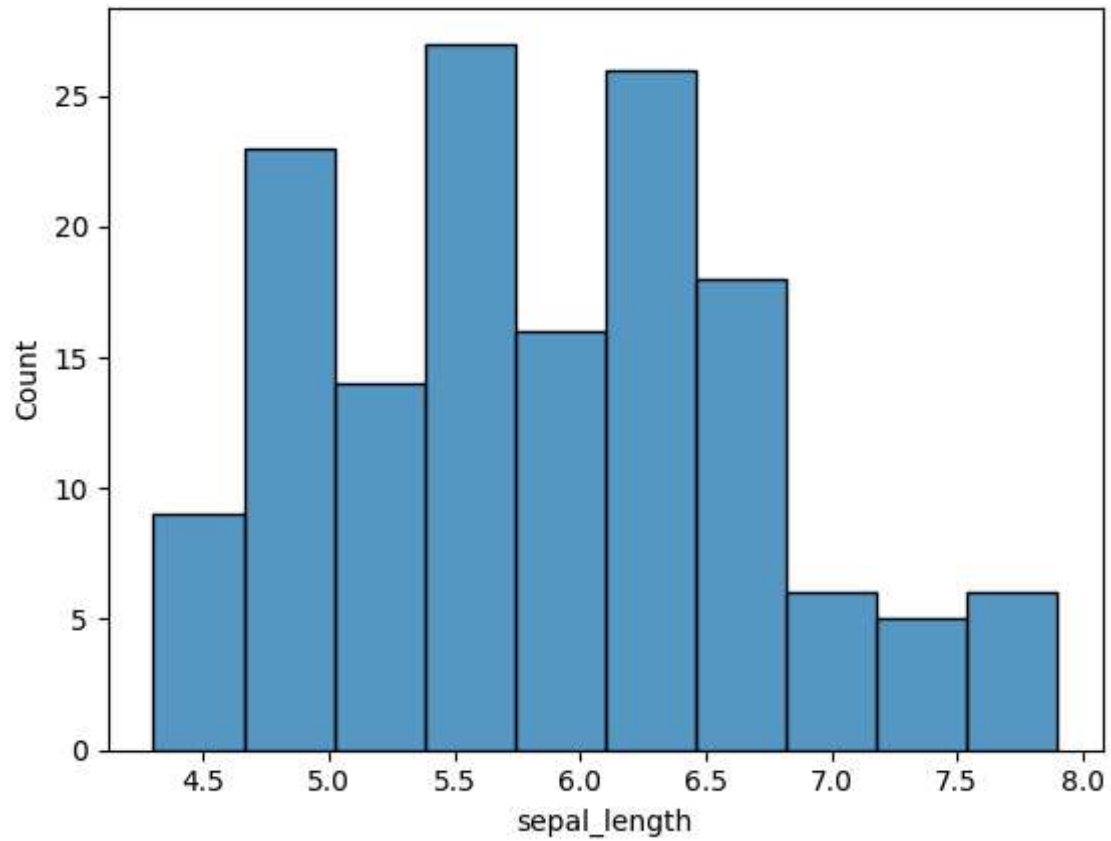
```
sepal_length    float64
sepal_width     float64
petal_length     float64
petal_width     float64
species         object
dtype: object
```

```
ds.describe()
```

	sepal_length	sepal_width	petal_length	petal_width	
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	

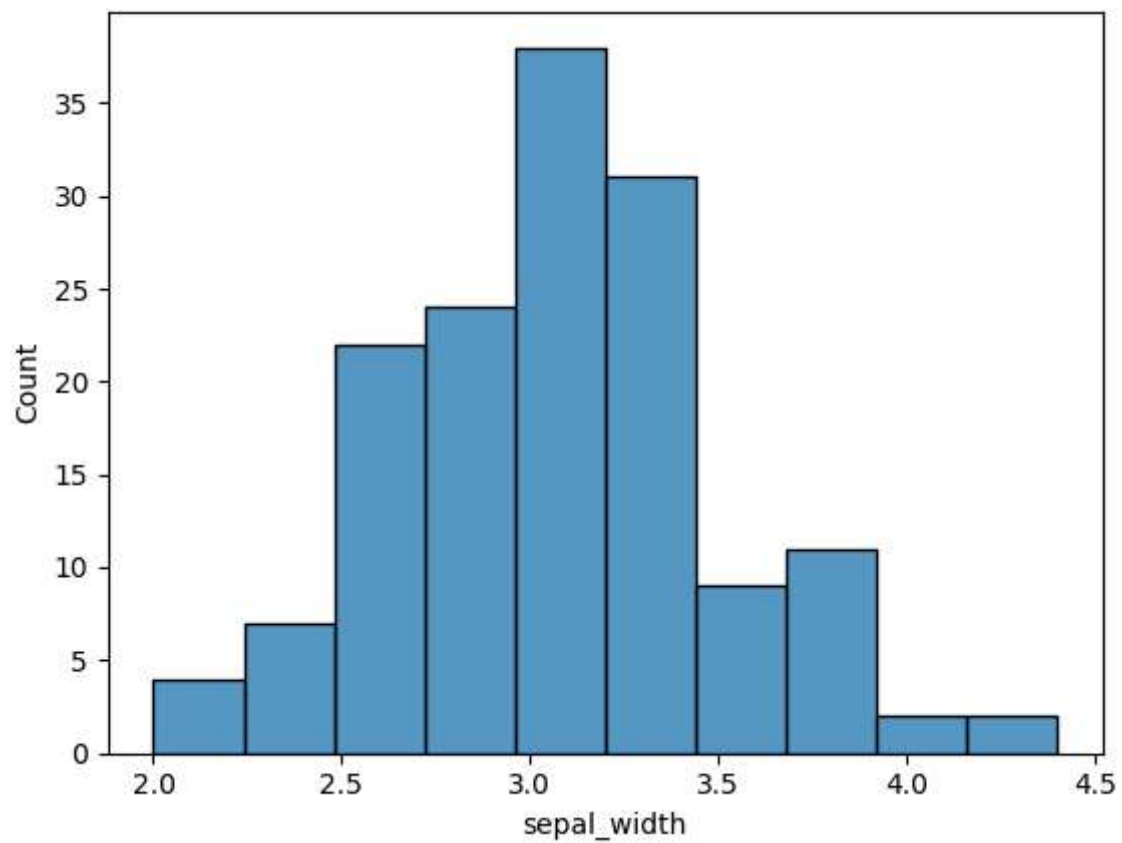
```
sns.histplot( data=ds , x="sepal_length" , bins=10 )
```

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



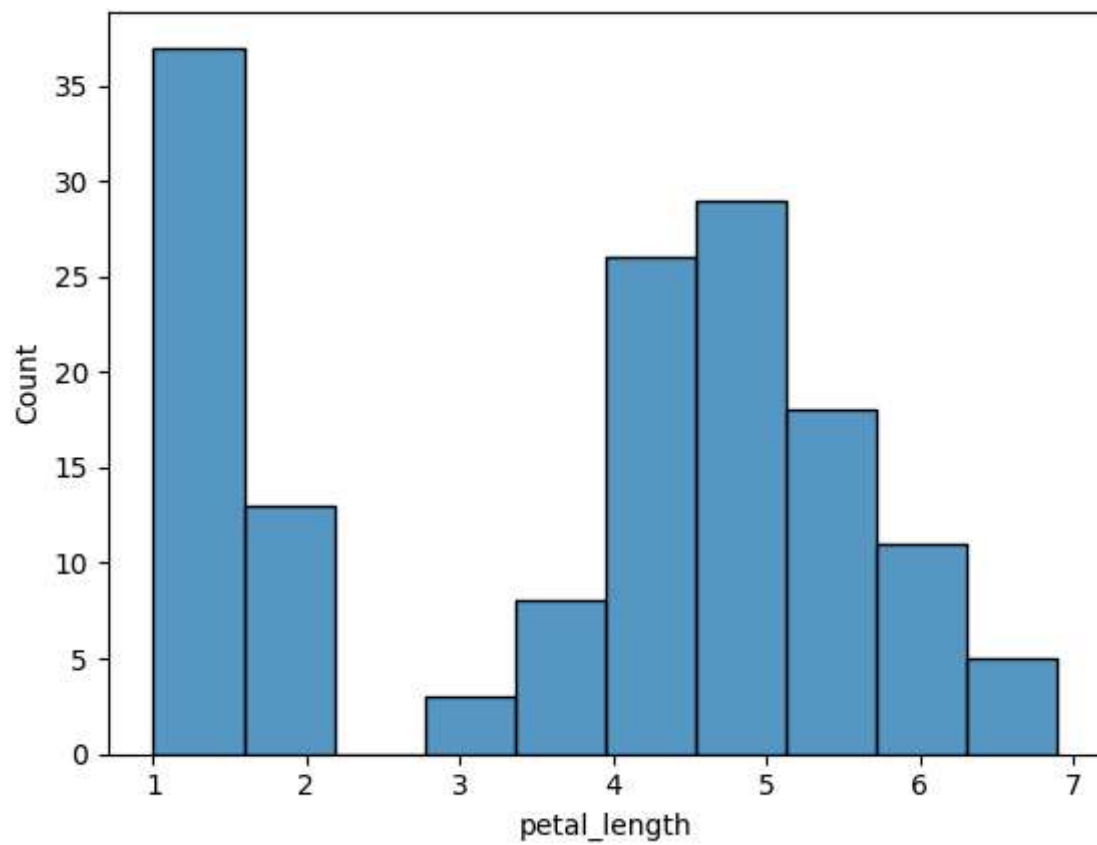
```
sns.histplot( data=ds , x="sepal_width" , bins=10 )
```

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



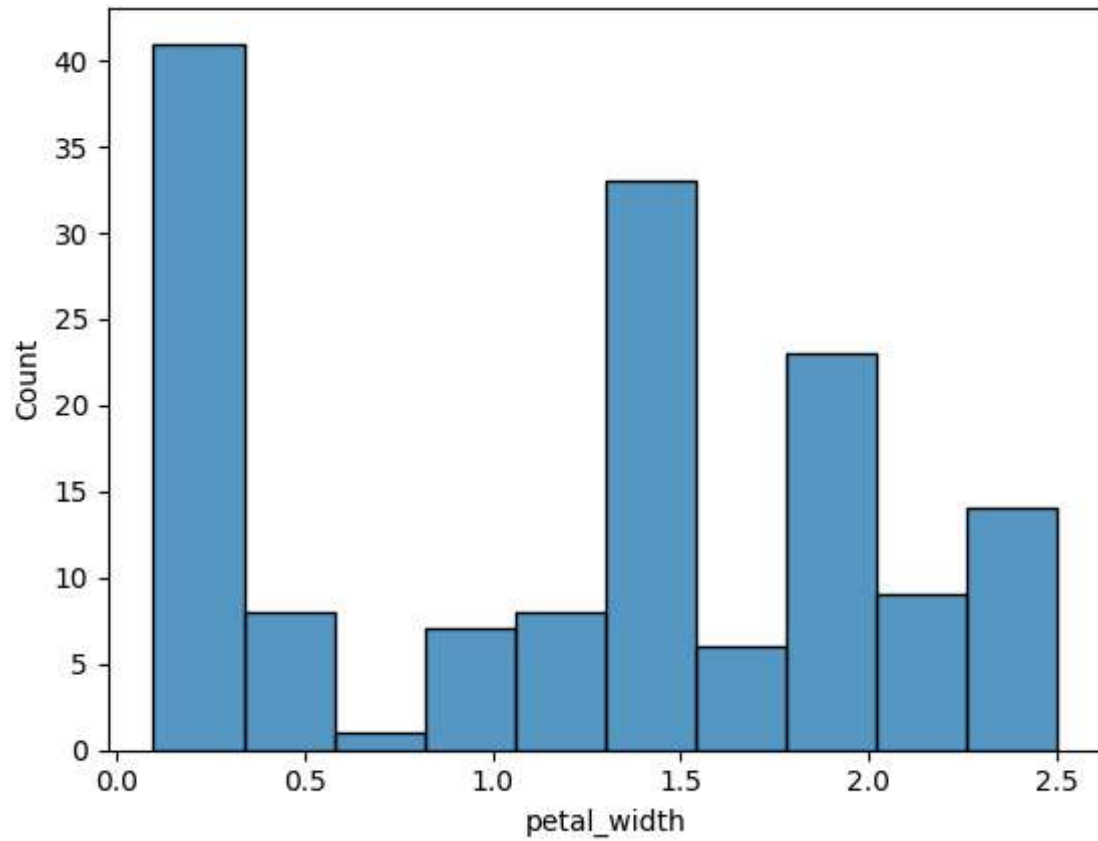
```
sns.histplot( data=ds , x="petal_length" , bins=10 )
```

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



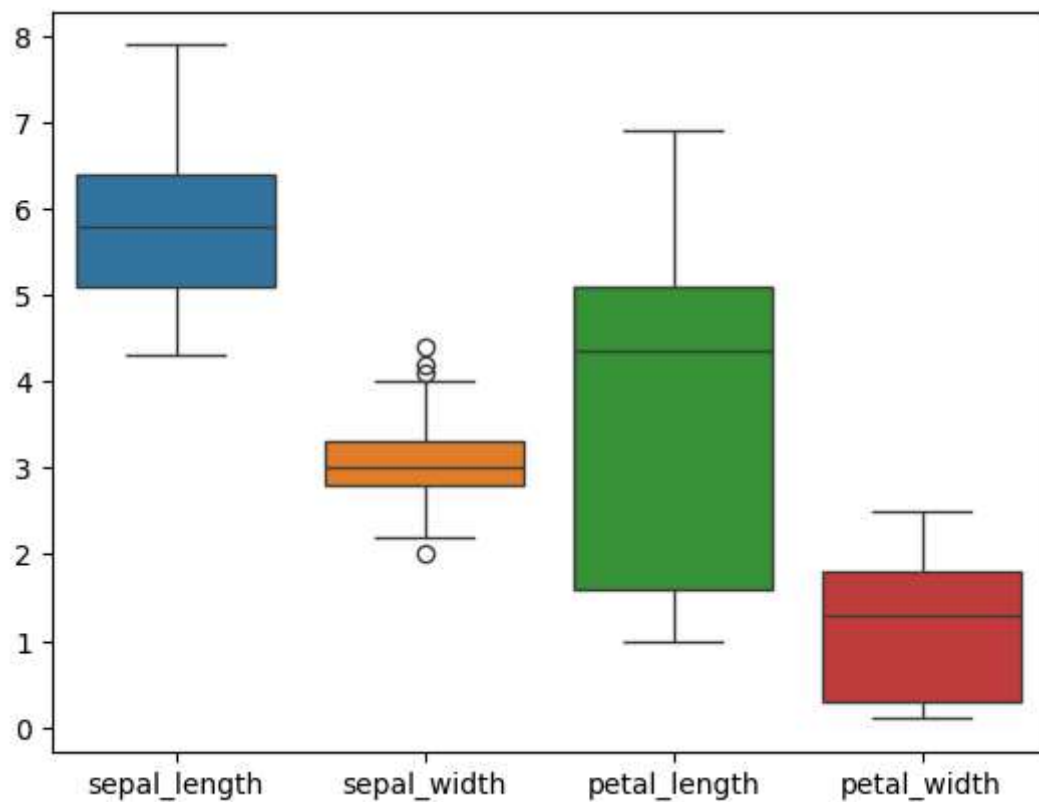
```
sns.histplot( data=ds , x="petal_width" , bins=10 )
```

<Axes: xlabel='petal_width', ylabel='Count'>



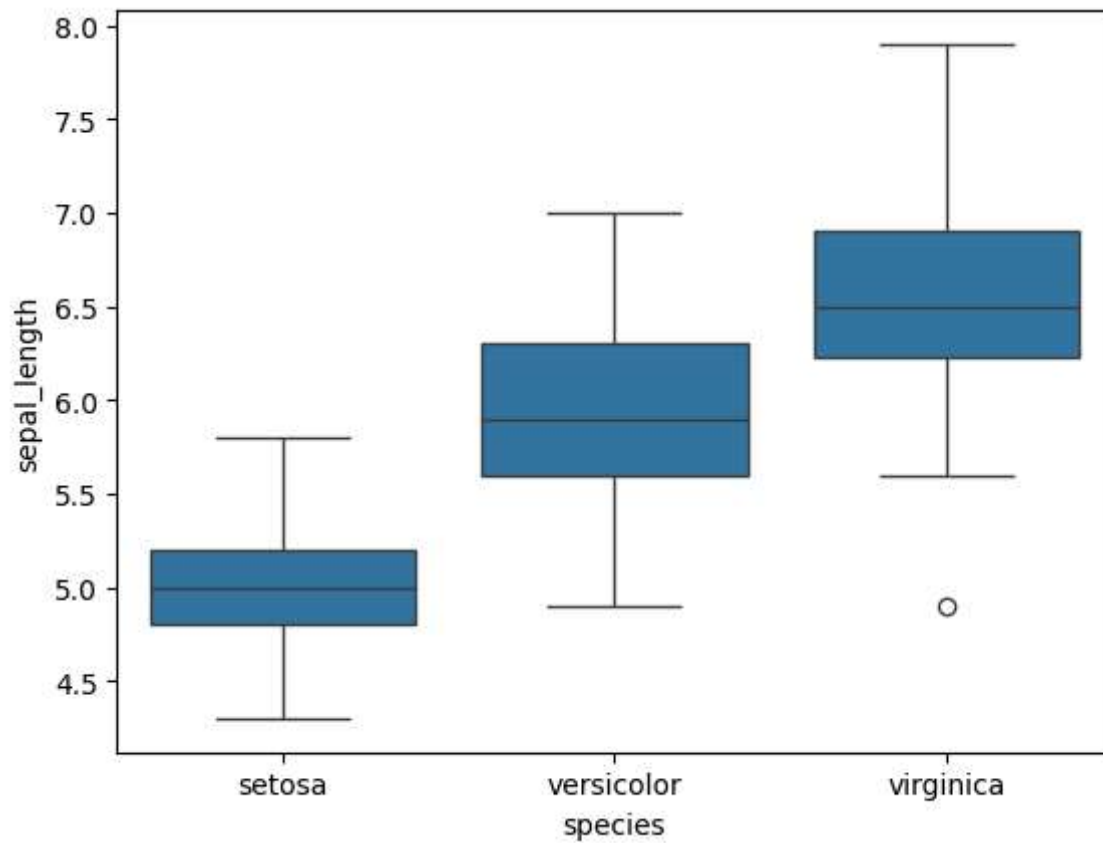
```
sns.boxplot( data=ds.drop( [ "species" ] , axis=1 ) )
```

<Axes: >



```
sns.boxplot( data=ds , x="species" , y="sepal_length" )
```

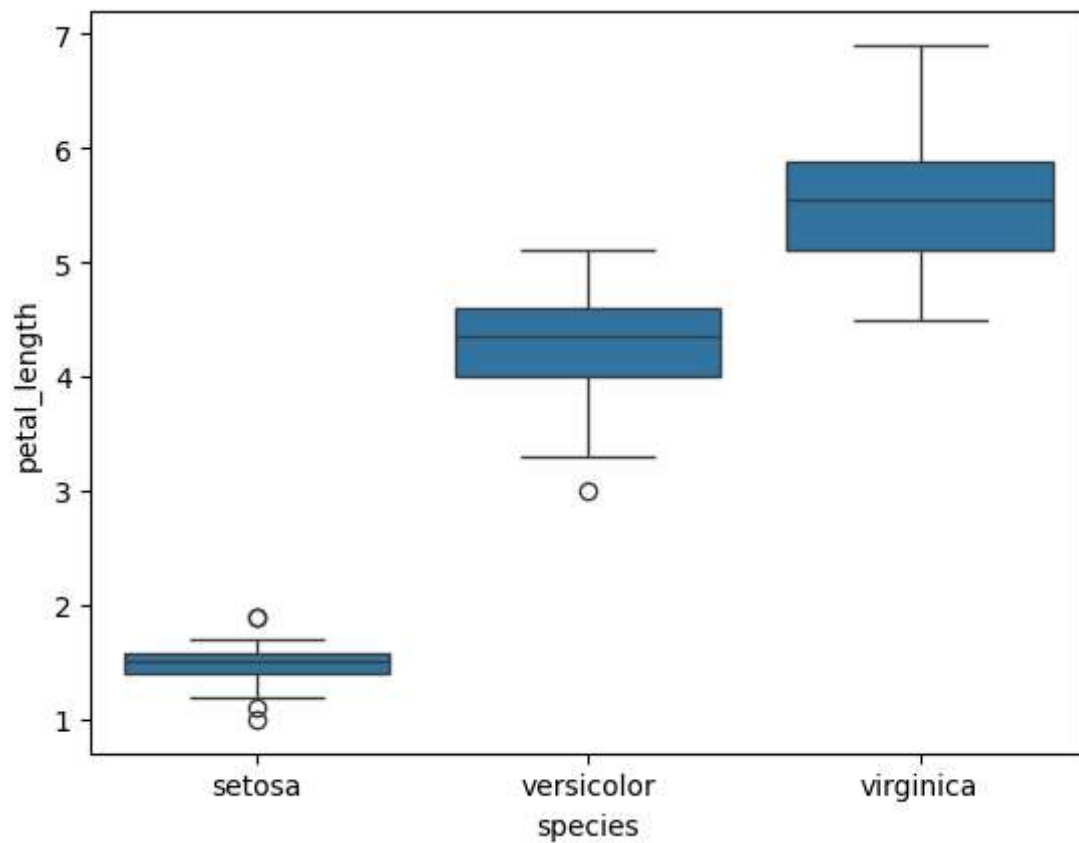
<Axes: xlabel='species', ylabel='sepal_length'>



```
sns.boxplot( data=ds , x="species" , y="sepal_width" )
```

```
<Axes: xlabel='species', ylabel='sepal width'>  
sns.boxplot( data=ds , x="species" , y="petal_length" )
```

```
<Axes: xlabel='species', ylabel='petal_length'>
```



```
sns.boxplot( data=ds , x="species" , y="petal_width" )
```