Name: Pranav Mehendale

Roll No.: TCOD34

Batch: T11

Data Visualization III:

1. Download the Iris flower dataset or any other dataset into a DataFrame. (e.g., https://archive.ics.uci.edu/ml/datasets/Iris).
2. Scan the dataset and give the inference as:
3. List down the features and their types (e.g., numeric, nominal) available in the dataset.
4. Create a histogram for each feature in the dataset to illustrate the feature distributions.
5. Create a boxplot for each feature in the dataset.
6. Compare distributions and identify outliers.

```
import pandas as pd
import matplotlib as plt
import seaborn as sns
import numpy as np
import warnings

warnings.filterwarnings('ignore')

%matplotlib inline

data = sns.load_dataset('iris')

data.head(5)
```

	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

```
data.sample(5)
```

	sepal_length	sepal_width	petal_length	petal_width	species
100	6.3	3.3	6.0	2.5	virginica
26	5.0	3.4	1.6	0.4	setosa
121	5.6	2.8	4.9	2.0	virginica
9	4.9	3.1	1.5	0.1	setosa
113	5.7	2.5	5.0	2.0	virginica

```
data.describe()
```

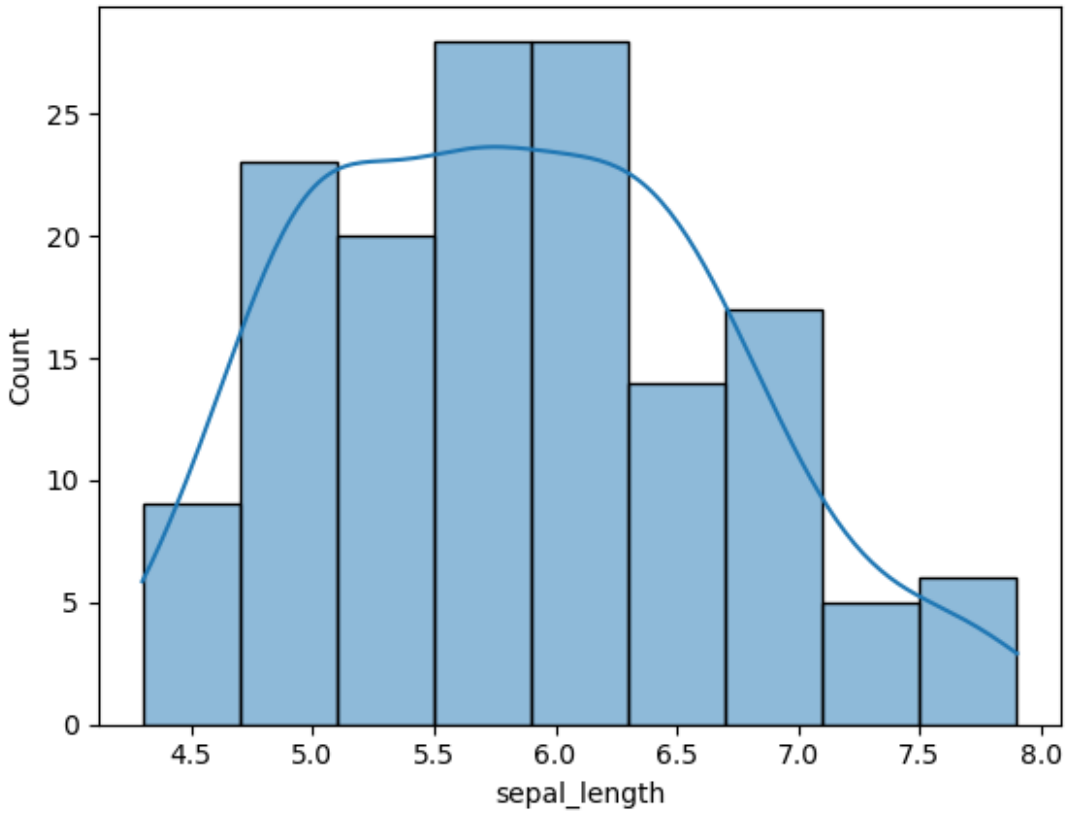
```
       sepal_length  sepal_width  petal_length  petal_width
count    150.000000   150.000000    150.000000   150.000000
mean       5.843333     3.057333      3.758000     1.199333
std        0.828066     0.435866      1.765298     0.762238
min        4.300000     2.000000      1.000000     0.100000
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
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   sepal_length  150 non-null    float64
 1   sepal_width   150 non-null    float64
 2   petal_length  150 non-null    float64
 3   petal_width   150 non-null    float64
 4   species       150 non-null    object 
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

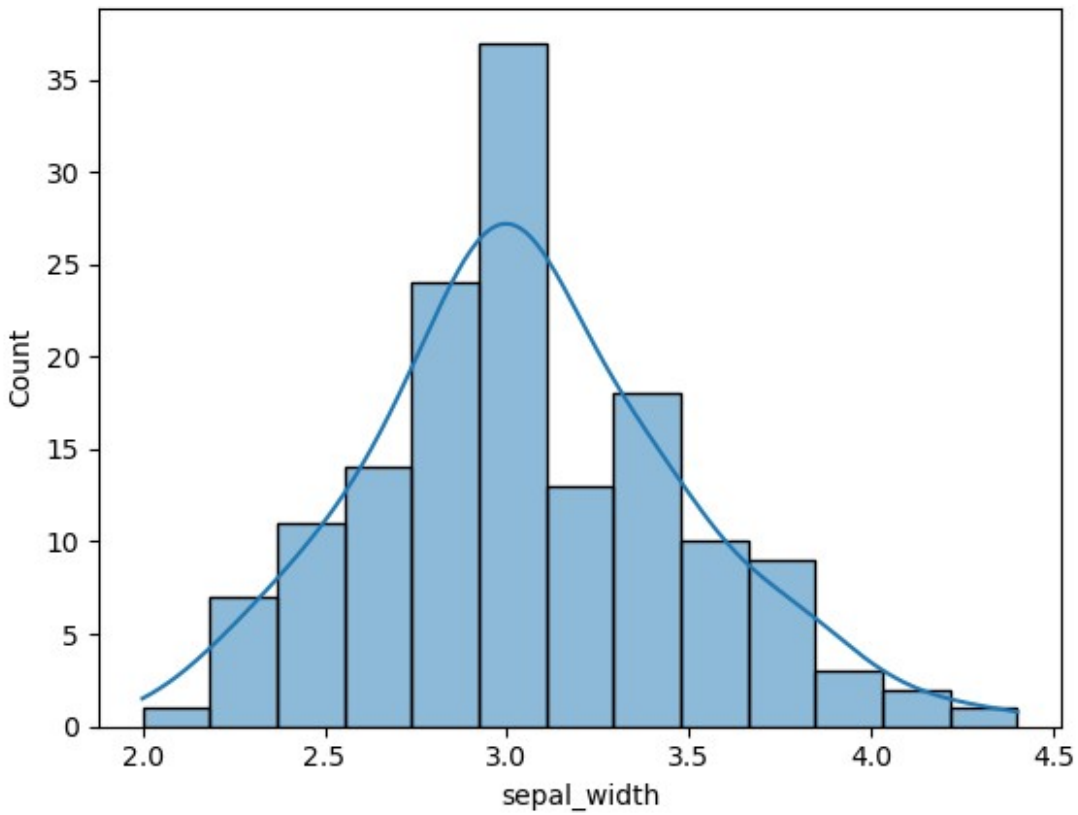
```
sns.histplot(data['sepal_length'],kde=True)
```

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



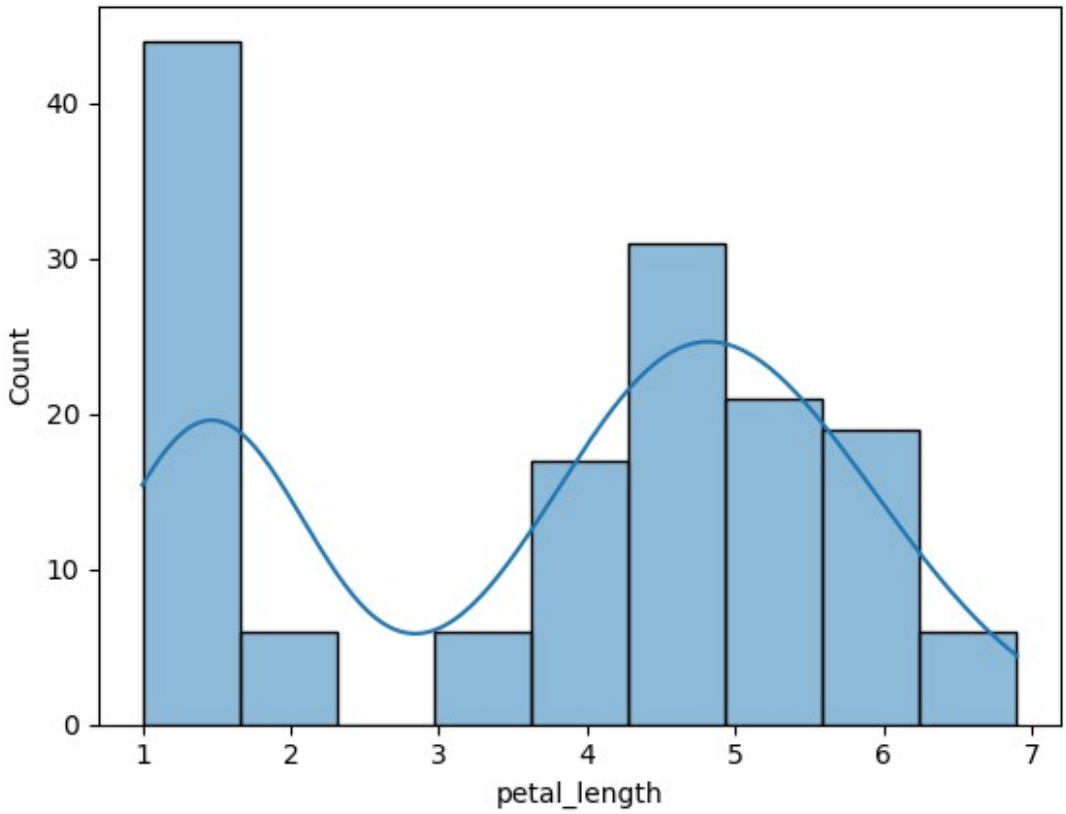

```
sns.histplot(data['sepal_width'],kde=True)
```

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



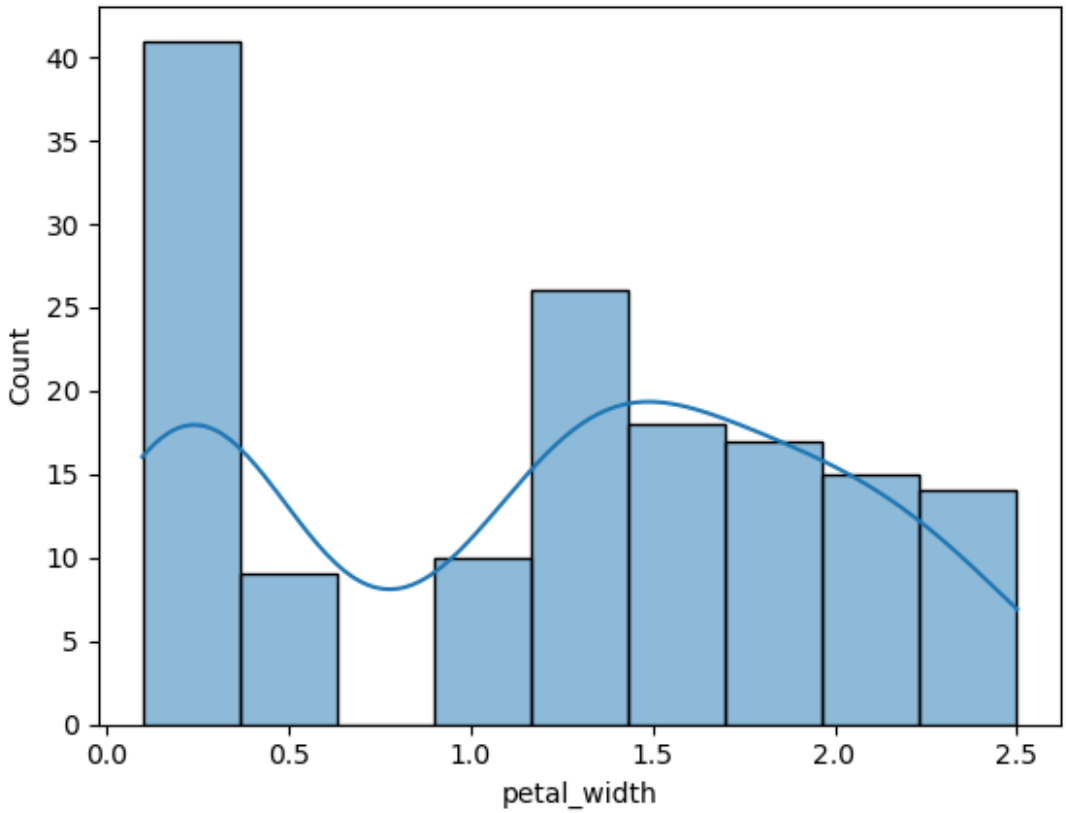

```
sns.histplot(data['petal_length'],kde=True)
```

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



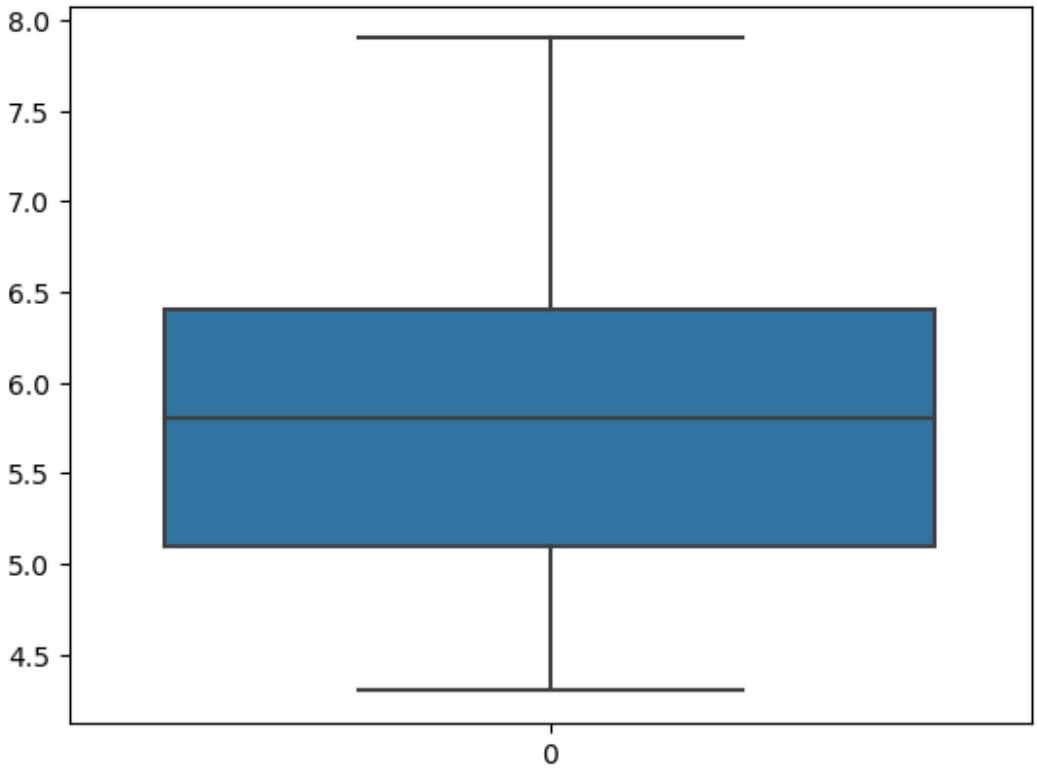

```
sns.histplot(data['petal_width'],kde=True)
```

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



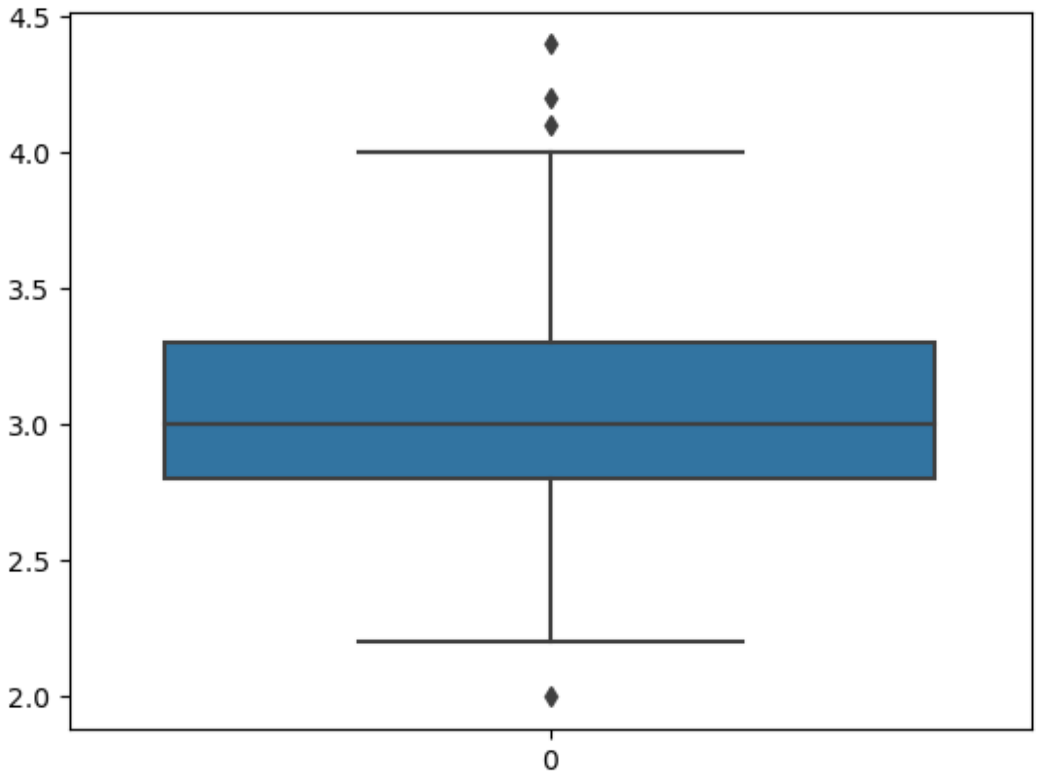

```
sns.boxplot(data['sepal_length'])
```

```
<Axes: >
```



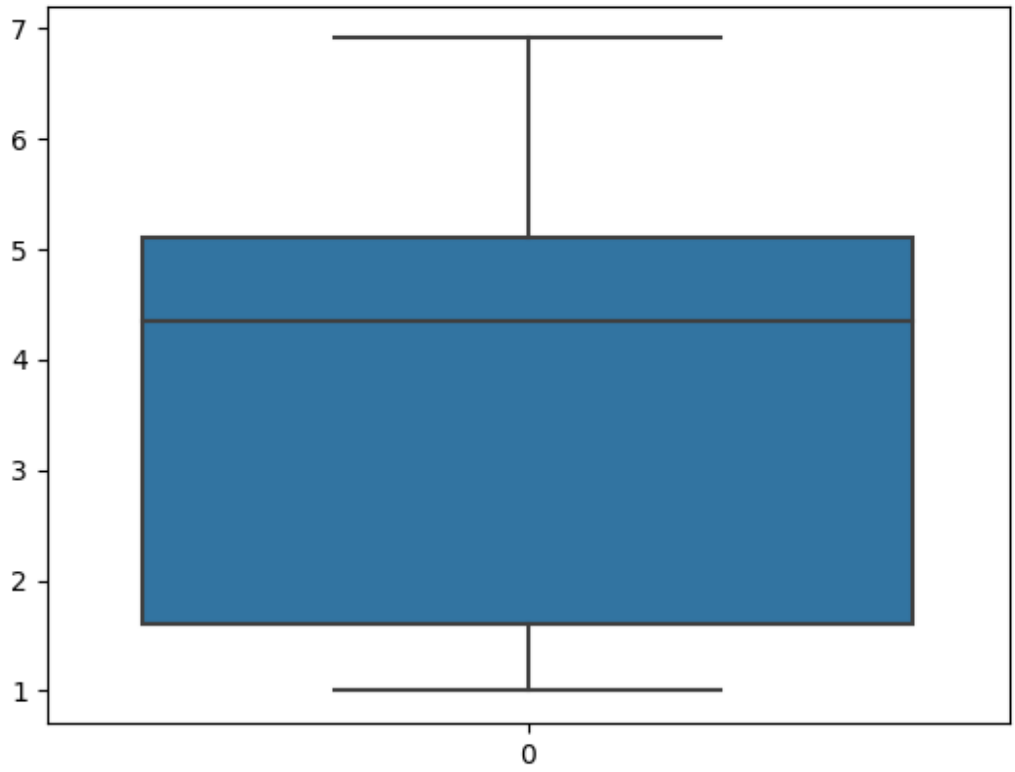

```
sns.boxplot(data['sepal_width'])
```

```
<Axes: >
```



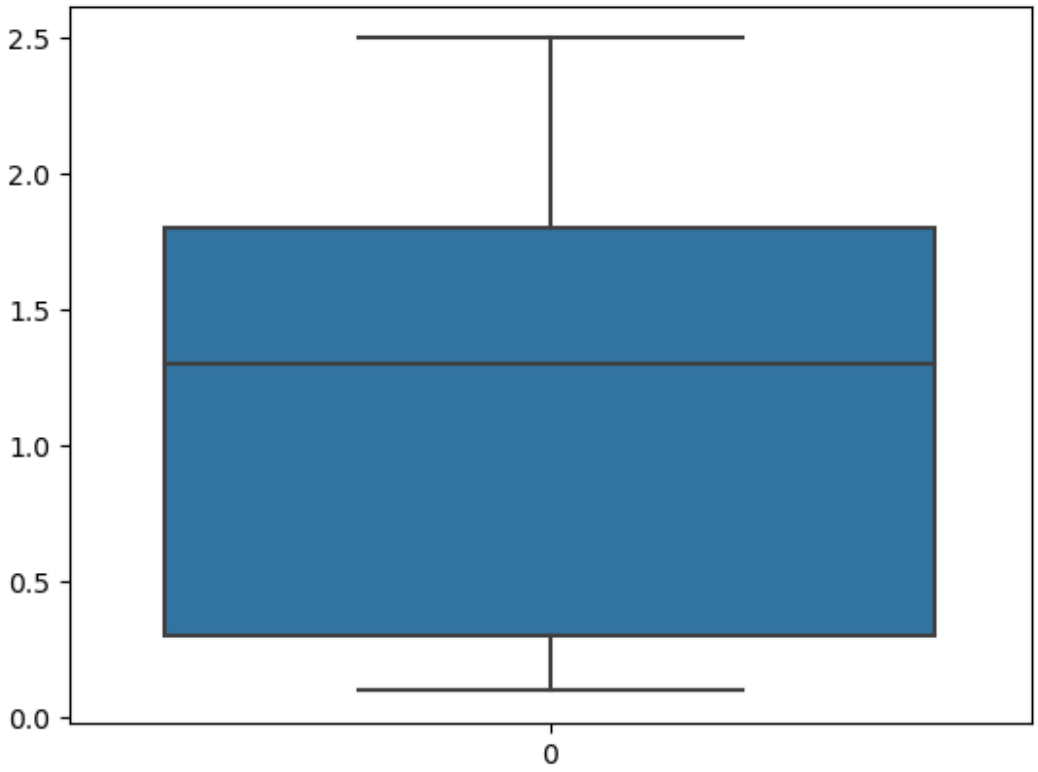

```
sns.boxplot(data['petal_length'])
```

```
<Axes: >
```



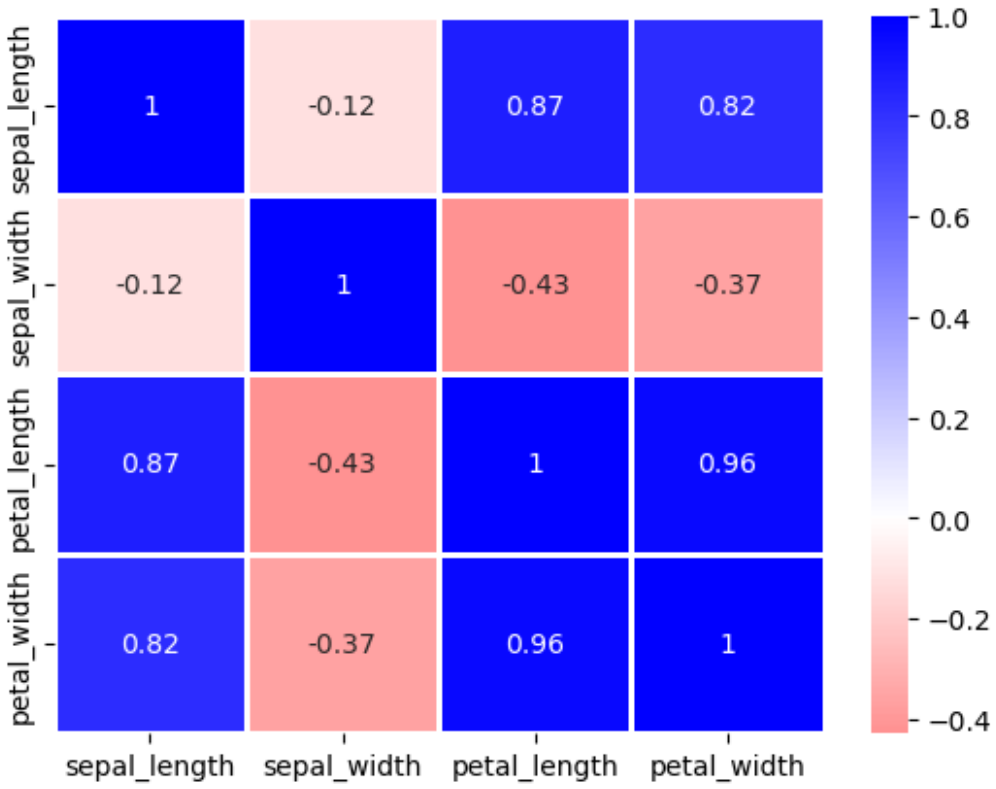

```
sns.boxplot(data['petal_width'])
```

```
<Axes: >
```

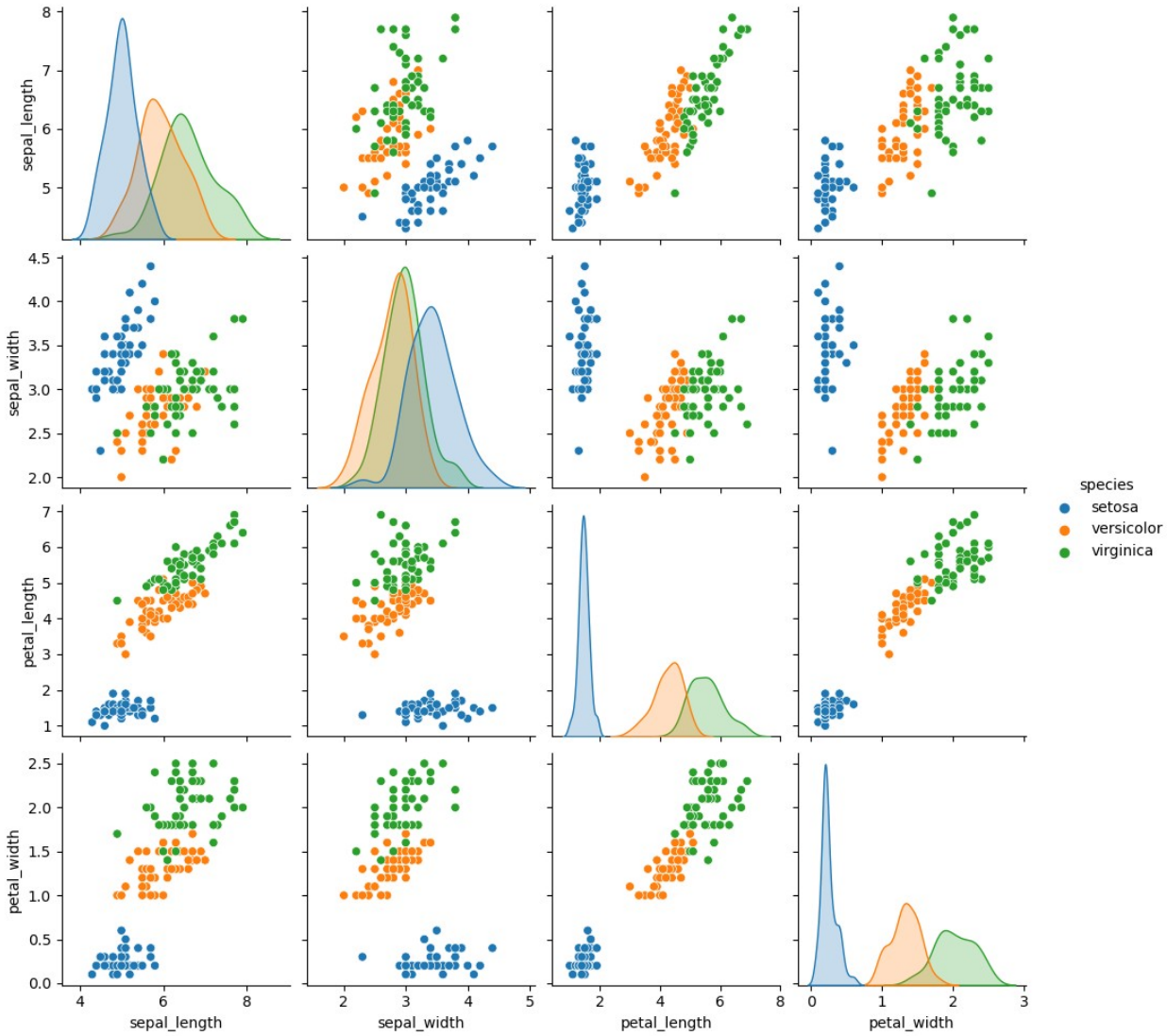



```
numerical_data = data.select_dtypes(include=[np.float64])
sns.heatmap(numerical_data.corr(), cmap='bwr_r',linecolor='white',
linewidths=1,annot=True, center=0)

<Axes: >
```




```
g = sns.pairplot(data,hue="species")
```


sepal_length
sepal_width
petal_length
petal_width
species
setosa
versicolor
virginica