Project ID - #CC69855 Project Title - Exploratory Data Analysis (EDA) on Iris Dataset Internship Domain - Data Science Intern Project Level - Entry Level Assigned By- CodeClause Internship

Conduct exploratory data analysis on the famous Iris dataset to understand its characteristics and relationships between features.

In [64]:
```
import pandas as pd
```

In [65]:
```
import seaborn as sns
```

In [66]:
```
import matplotlib.pyplot as plt
```

In [67]:
```
df = pd.read_csv("iris_csv.csv")
```

Print top 5 rows in the dataset

In [68]:
```
df.head()
```

Out[68]:

	sepallength	sepalwidth	petallength	petalwidth	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

In [69]:
```
df.shape
```

Out[69]:
```
(150, 5)
```

Summary of the dataset

In [70]:
```
df.describe()
```

Out[70]:

	sepallength	sepalwidth	petallength	petalwidth
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

In [71]:
```
df.info()
```

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

Checking Missing Values

In [72]:
```
df.isnull().sum()
```

Out[72]:
```
sepallength    0
sepalwidth     0
petallength    0
petalwidth     0
class          0
dtype: int64
```

Checking Duplicates

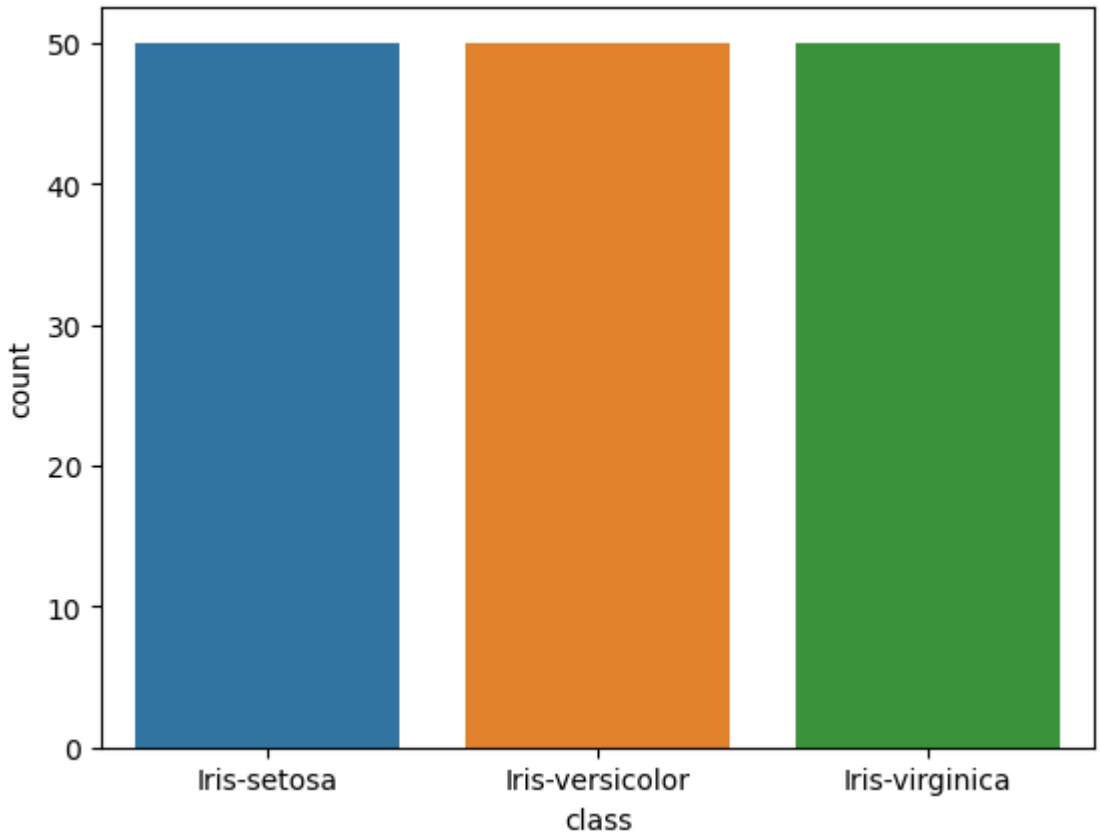
In [73]:
```
data = df.drop_duplicates(subset ="class",)
data
```

Out[73]:

	sepallength	sepalwidth	petallength	petalwidth	class
0	5.1	3.5	1.4	0.2	Iris-setosa
50	7.0	3.2	4.7	1.4	Iris-versicolor
100	6.3	3.3	6.0	2.5	Iris-virginica

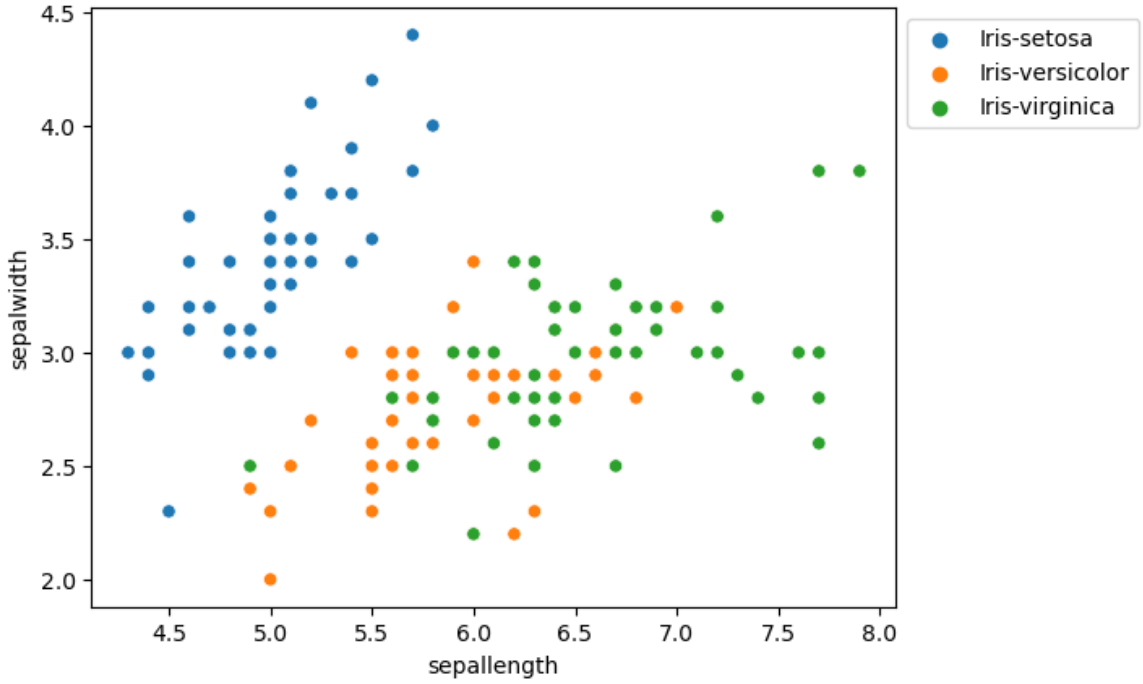
Data Visualization

In [74]:
```
sns.countplot(x='class', data=df, )
plt.show()
```



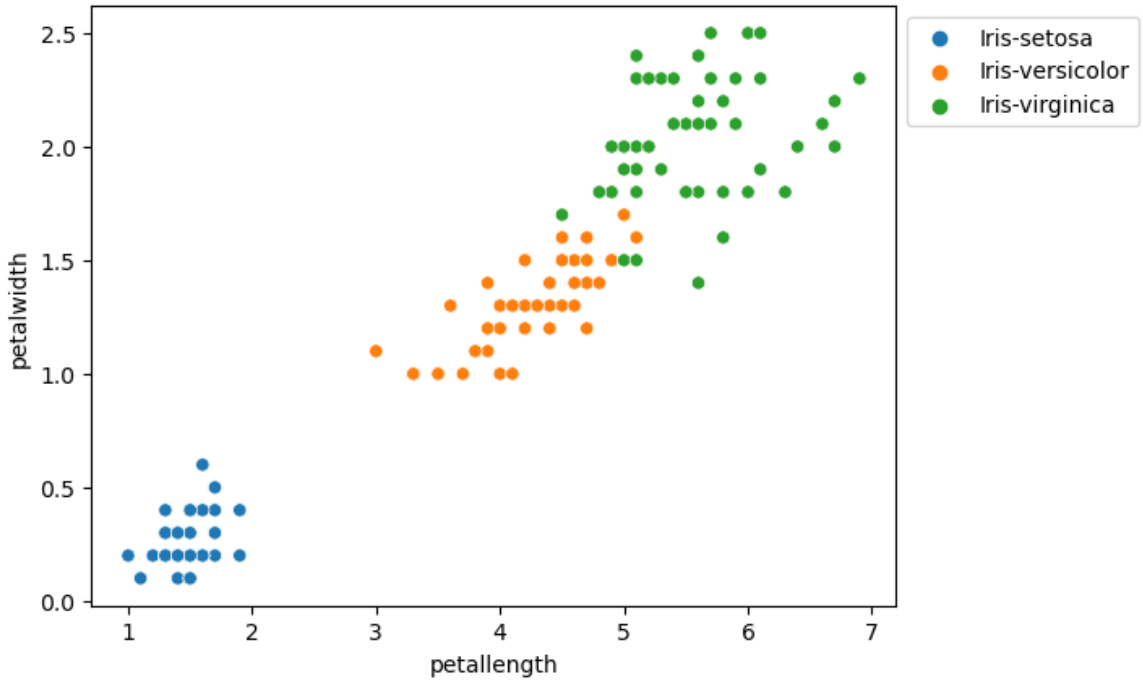
Comparing Sepal Length and Sepal Width

In [75]:
```
sns.scatterplot(x='sepallength', y='sepalwidth',hue='class', data=df,
plt.legend(bbox_to_anchor=(1, 1), loc=2)
plt.show()
```



Comparing Petal Length and Petal Width

In [76]:
```
sns.scatterplot(x='petallength', y='petalwidth',hue='class', data=df,
plt.legend(bbox_to_anchor=(1, 1), loc=2)
plt.show()
```



Histograms

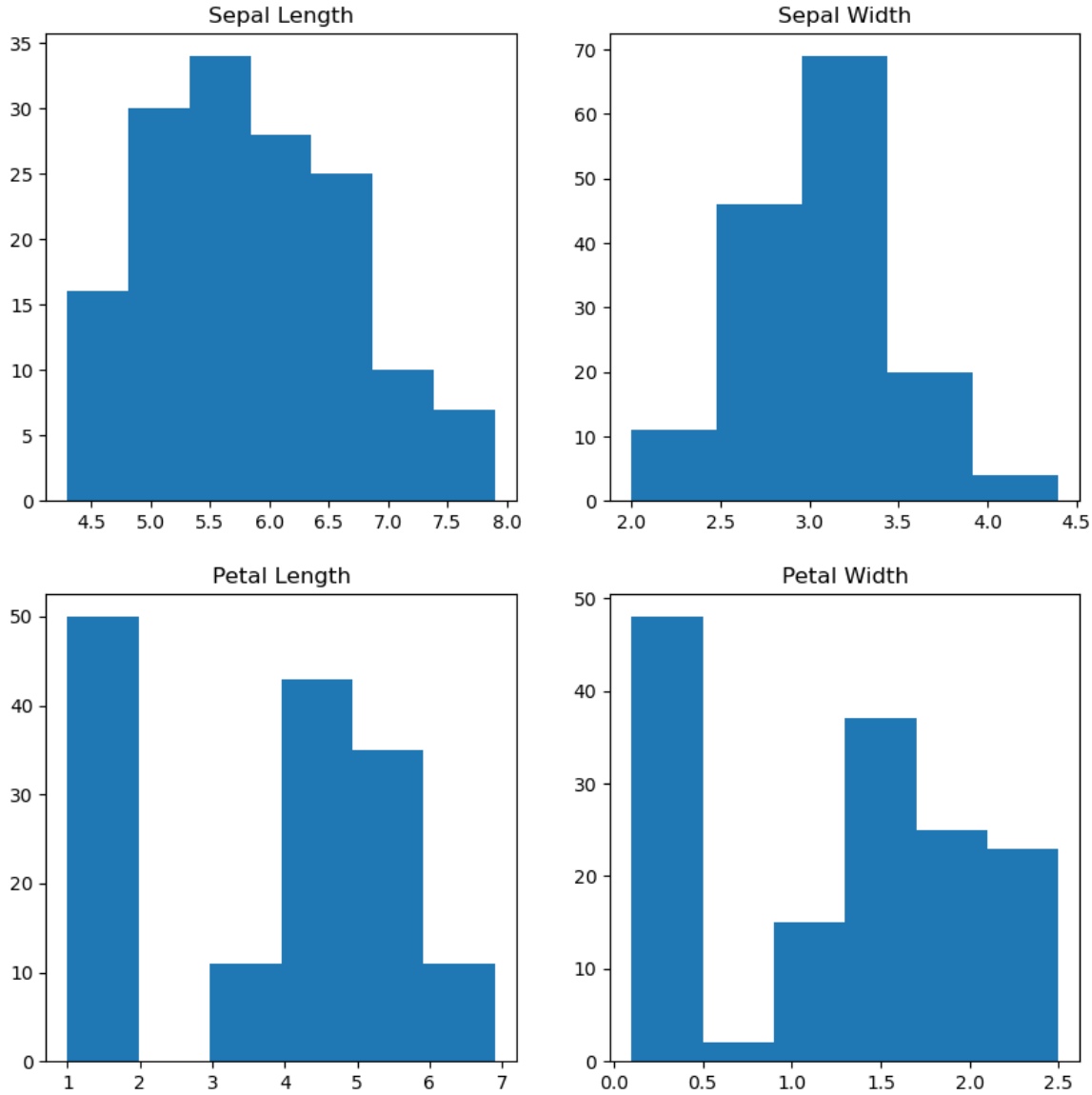
In [77]:
```
fig, axes = plt.subplots(2, 2, figsize=(10,10))

axes[0,0].set_title("Sepal Length")
axes[0,0].hist(df['sepallength'], bins=7)

axes[0,1].set_title("Sepal Width")
axes[0,1].hist(df['sepalwidth'], bins=5);

axes[1,0].set_title("Petal Length")
axes[1,0].hist(df['petallength'], bins=6);

axes[1,1].set_title("Petal Width")
axes[1,1].hist(df['petalwidth'], bins=6);
```



Correlation

In [78]:
```
data.corr(method='pearson', numeric_only=True)
```

Out[78]:

	sepallength	sepalwidth	petallength	petalwidth
sepallength	1.000000	-0.999226	0.795795	0.643817
sepalwidth	-0.999226	1.000000	-0.818999	-0.673417
petallength	0.795795	-0.818999	1.000000	0.975713
petalwidth	0.643817	-0.673417	0.975713	1.000000

Handling Outliers

In [79]:
```
sns.boxplot(x='sepalwidth', data=df)
```

Out[79]:
```
<Axes: xlabel='sepalwidth'>
```