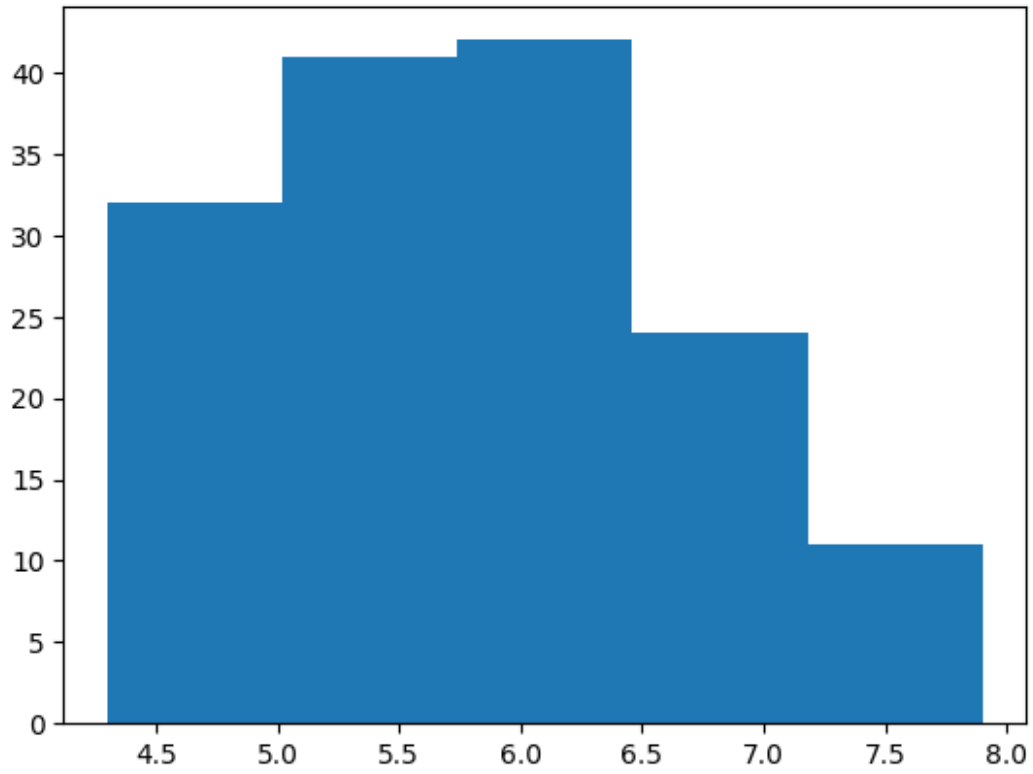
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv('iris.csv')

df.head()
```

```
   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
```

SEPALLENGTH ANALYSIS

```
plt.hist(df['SepalLengthCm'],bins=5)
```

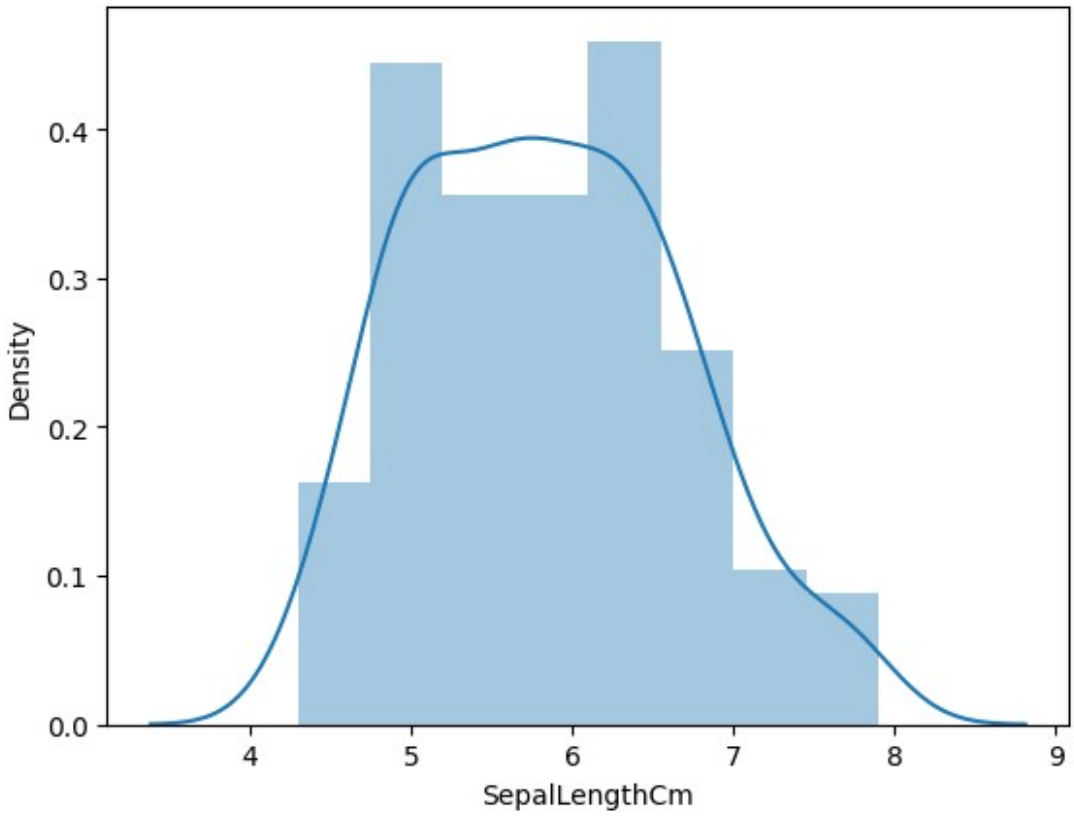
```
(array([32., 41., 42., 24., 11.]),
 array([4.3 , 5.02, 5.74, 6.46, 7.18, 7.9 ]),
 <BarContainer object of 5 artists>)
```




```
sns.distplot(df['SepalLengthCm'])
```

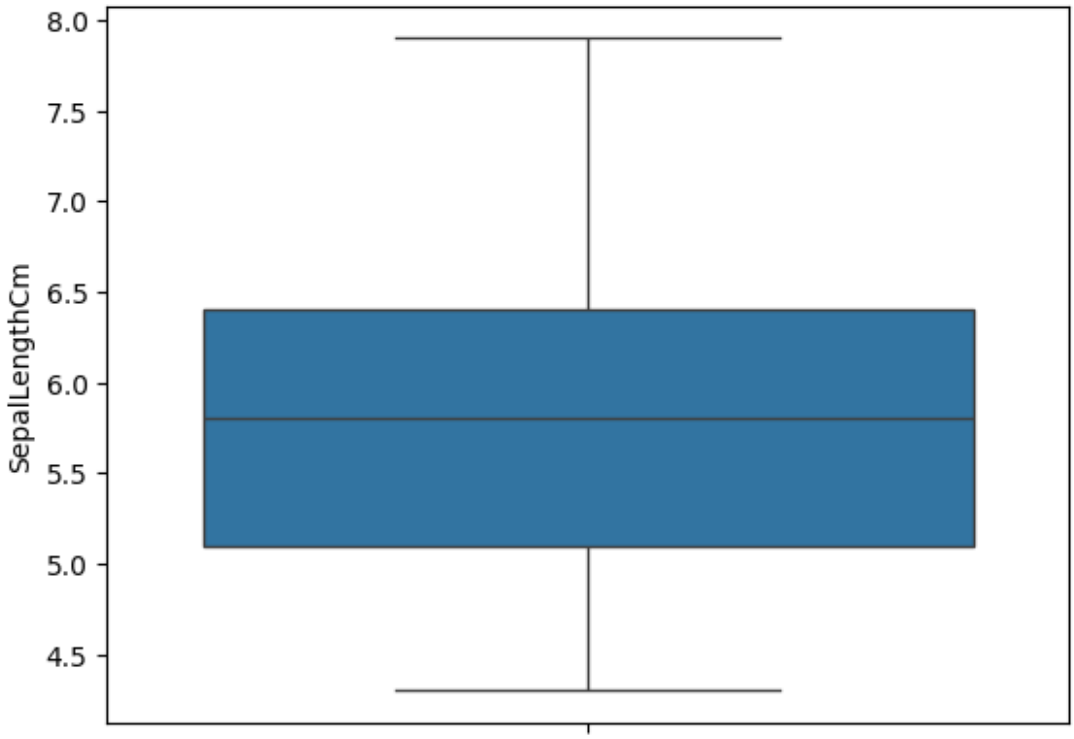
```
C:\Users\KIIT\AppData\Local\Temp\ipykernel_30844\789205228.py:1:
UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['SepalLengthCm'])
```

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




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

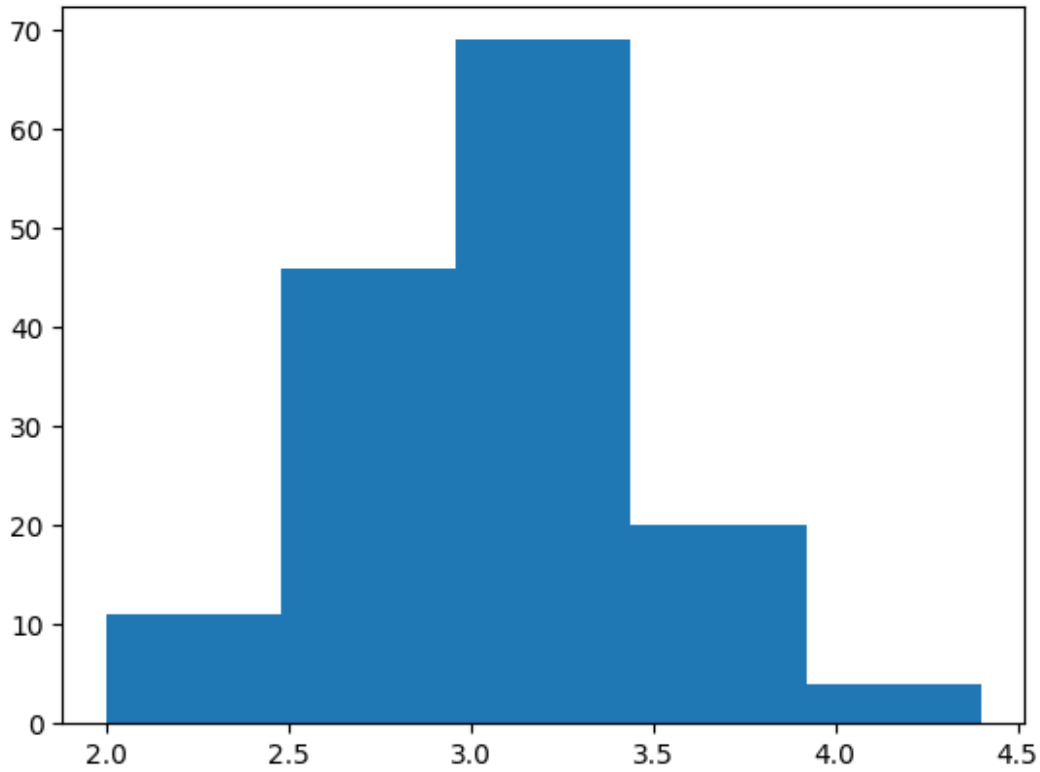
```
<Axes: ylabel='SepalLengthCm'>
```



SEPALWIDTHCM

```
plt.hist(df['SepalWidthCm'],bins=5)
```

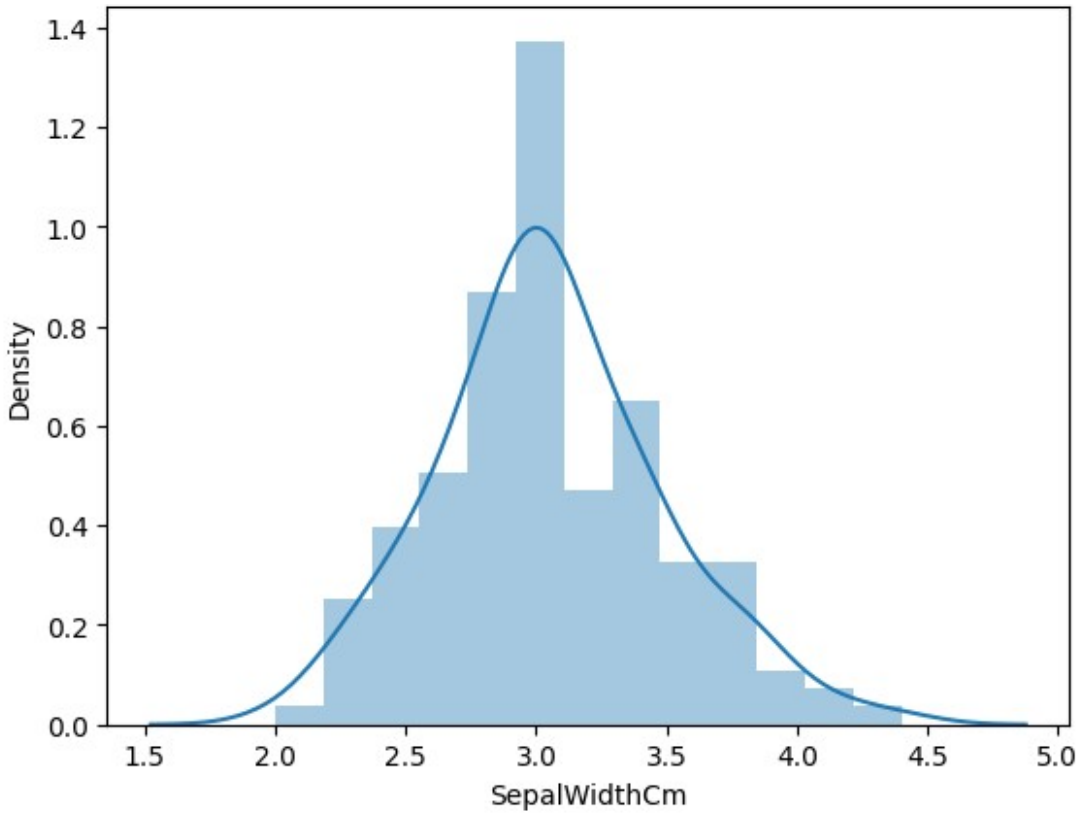
```
(array([11., 46., 69., 20.,  4.]),
 array([2.  , 2.48, 2.96, 3.44, 3.92, 4.4 ]),
 <BarContainer object of 5 artists>)
```




```
sns.distplot(df['SepalWidthCm'])
```

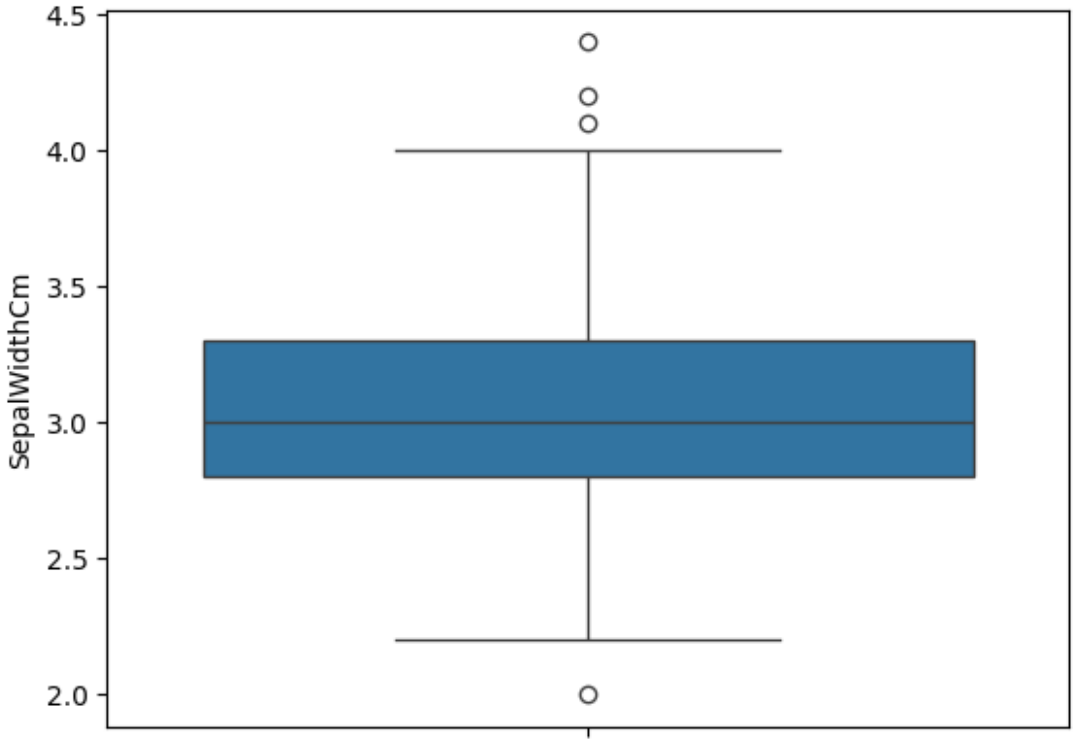
```
C:\Users\KIIT\AppData\Local\Temp\ipykernel_30844\3402425195.py:1:
UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['SepalWidthCm'])
```

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




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

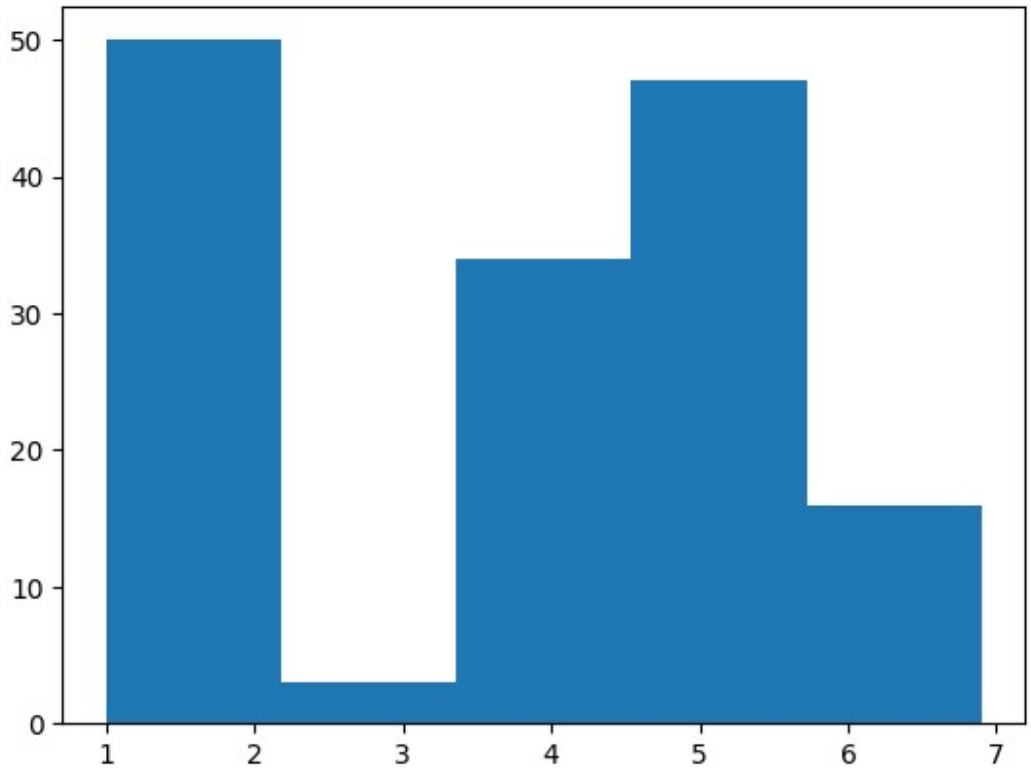
```
<Axes: ylabel='SepalWidthCm'>
```



PETALLENGTHCM

```
plt.hist(df['PetalLengthCm'],bins=5)
```

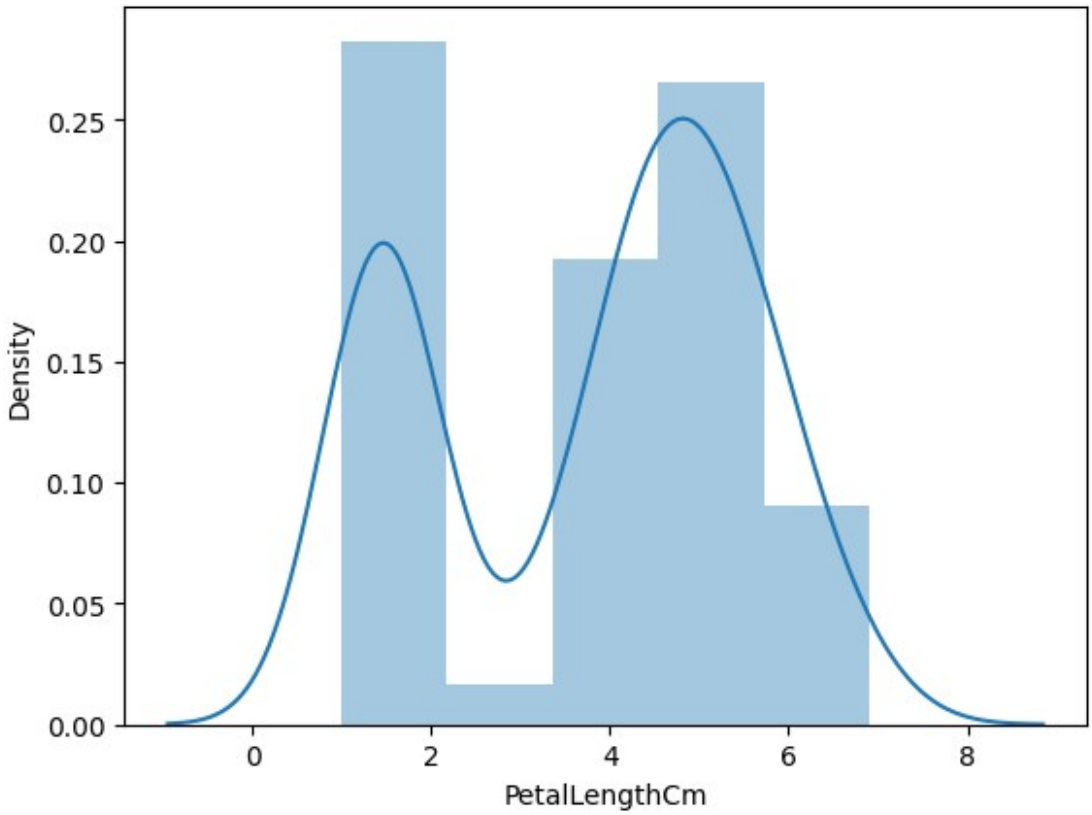
```
(array([50.,  3., 34., 47., 16.]),
 array([1.  , 2.18, 3.36, 4.54, 5.72, 6.9 ]),
 <BarContainer object of 5 artists>)
```




```
sns.distplot(df['PetalLengthCm'])
```

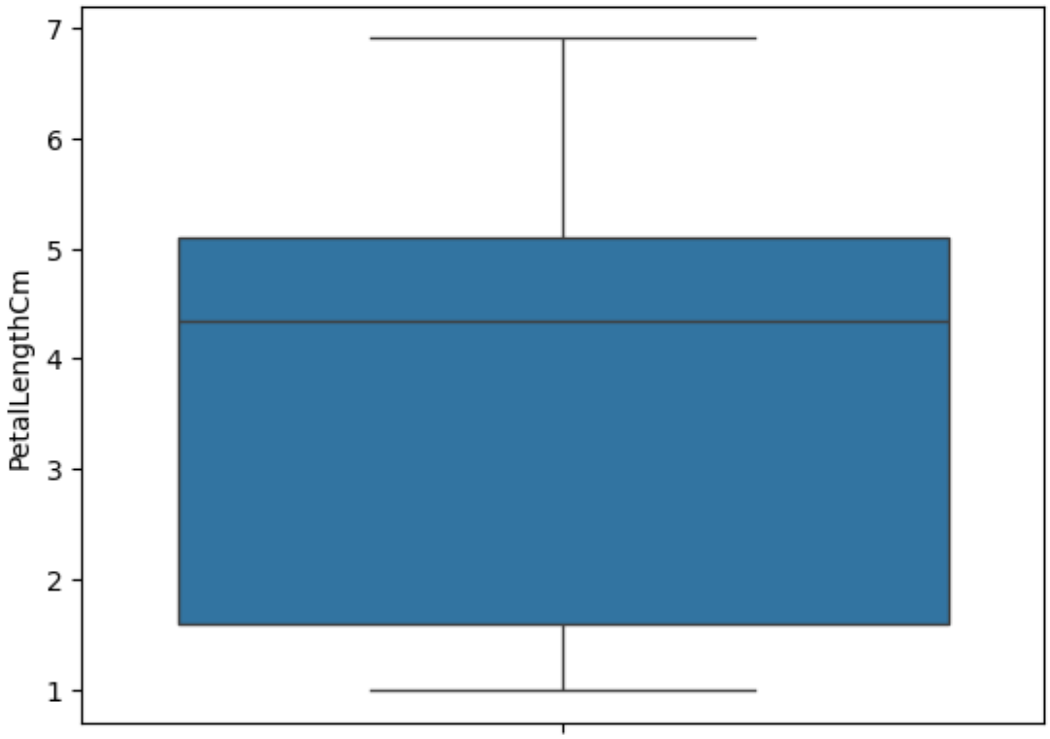
```
C:\Users\KIIT\AppData\Local\Temp\ipykernel_30844\2781944064.py:1:
UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['PetalLengthCm'])
```

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




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

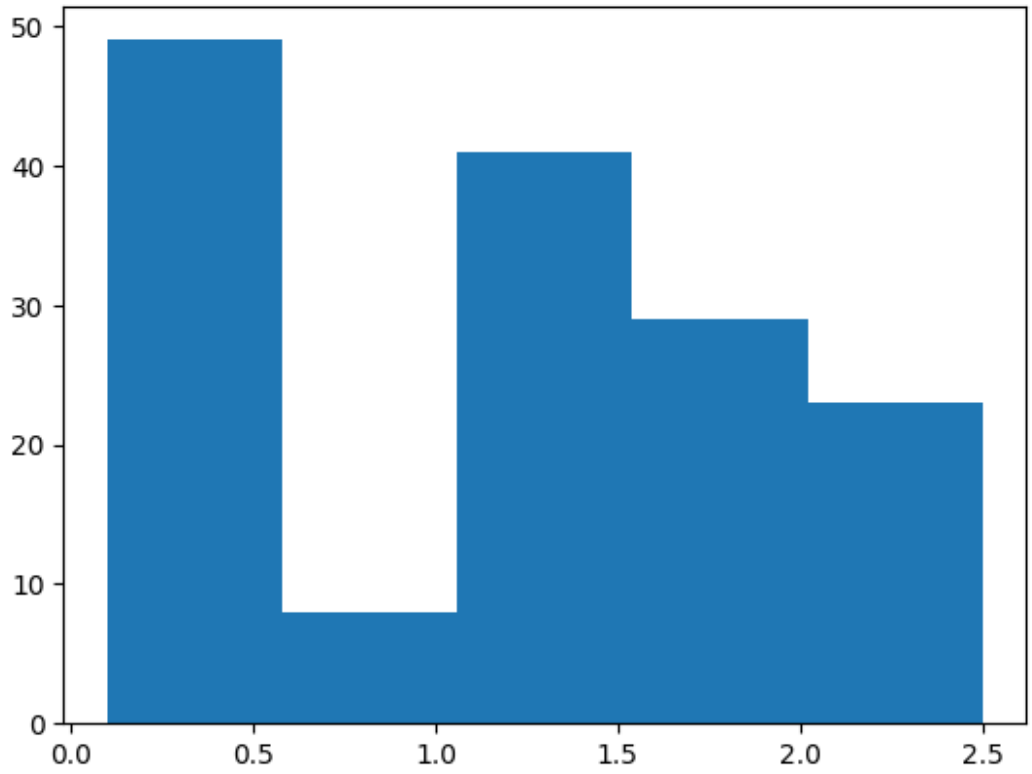
```
<Axes: ylabel='PetalLengthCm'>
```



PETALWIDTHCM

```
plt.hist(df['PetalWidthCm'],bins=5)
```

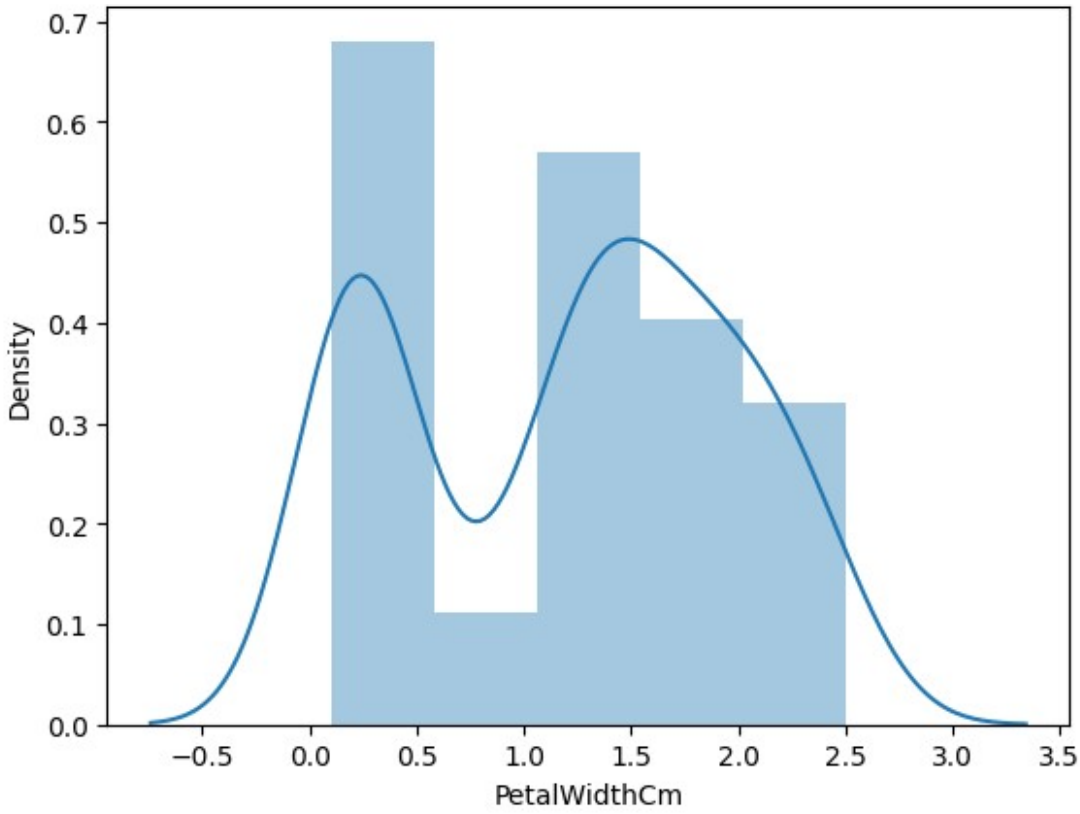
```
(array([49.,  8., 41., 29., 23.]),
 array([0.1 , 0.58, 1.06, 1.54, 2.02, 2.5 ]),
 <BarContainer object of 5 artists>)
```




```
sns.distplot(df['PetalWidthCm'])
```

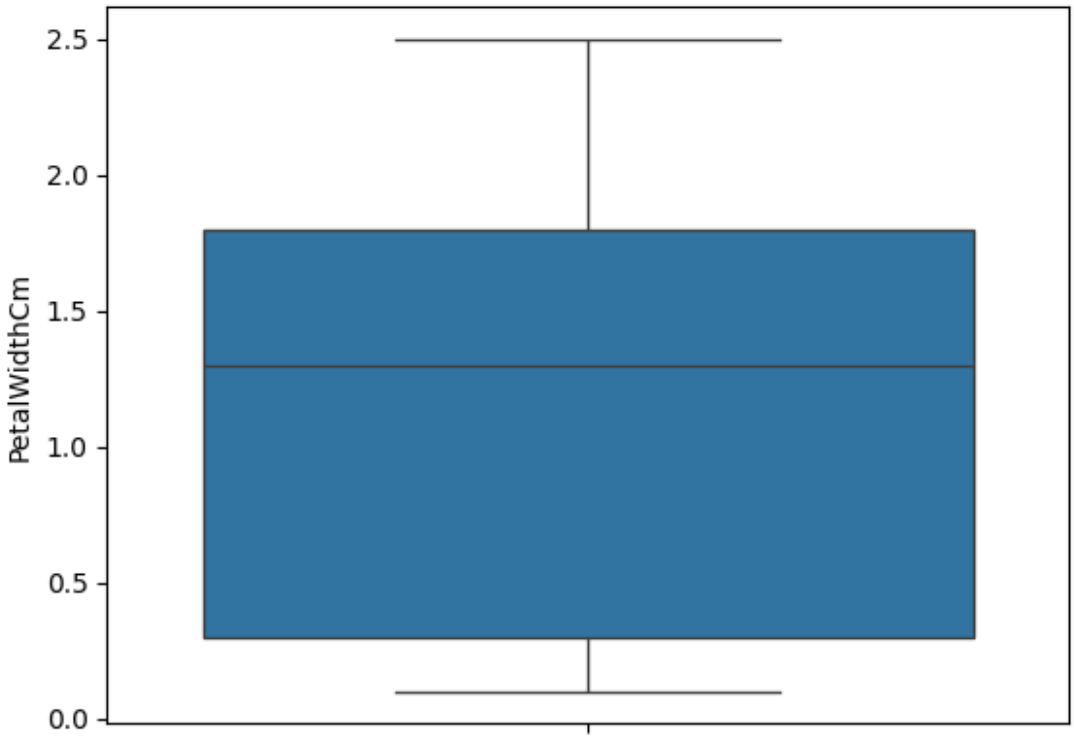
```
C:\Users\KIIT\AppData\Local\Temp\ipykernel_30844\3666008087.py:1:
UserWarning:

`distplot` is a deprecated function and will be removed in seaborn
v0.14.0.

Please adapt your code to use either `displot` (a figure-level
function with
similar flexibility) or `histplot` (an axes-level function for
histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  sns.distplot(df['PetalWidthCm'])
```

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




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

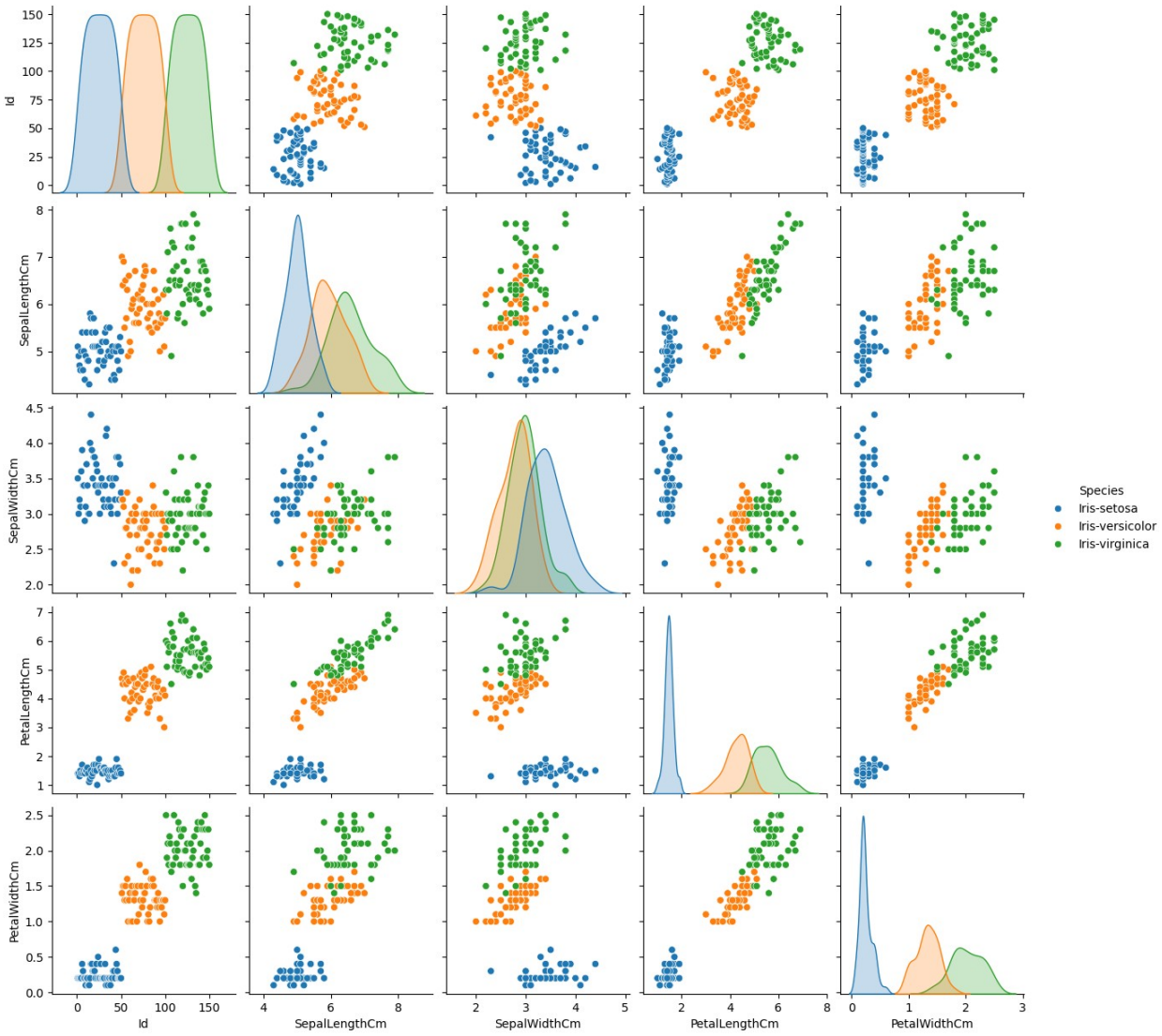
```
<Axes: ylabel='PetalWidthCm'>
```



SPECIES

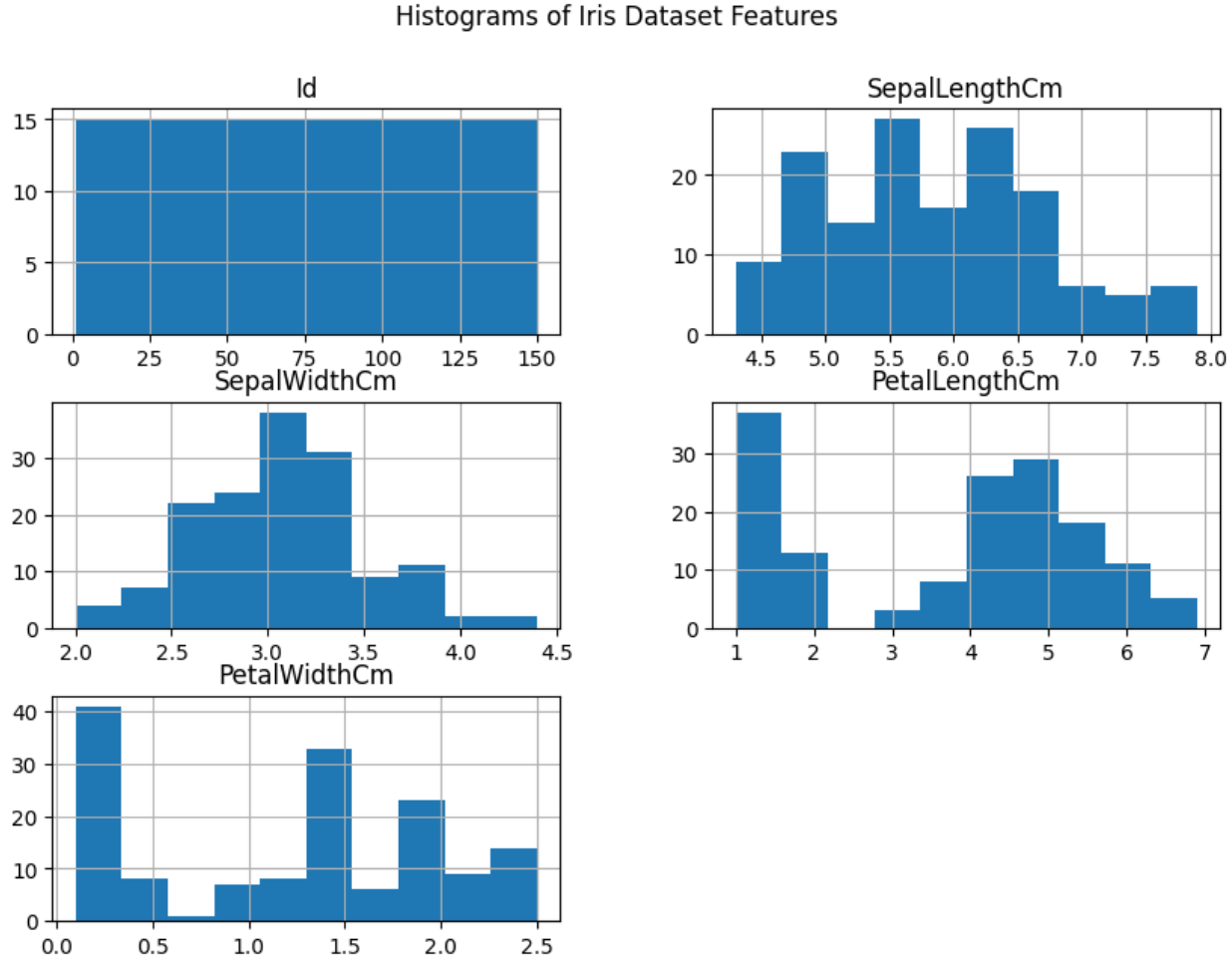
```
sns.pairplot(df,hue='Species')
```

```
<seaborn.axisgrid.PairGrid at 0x1ce883cc710>
```

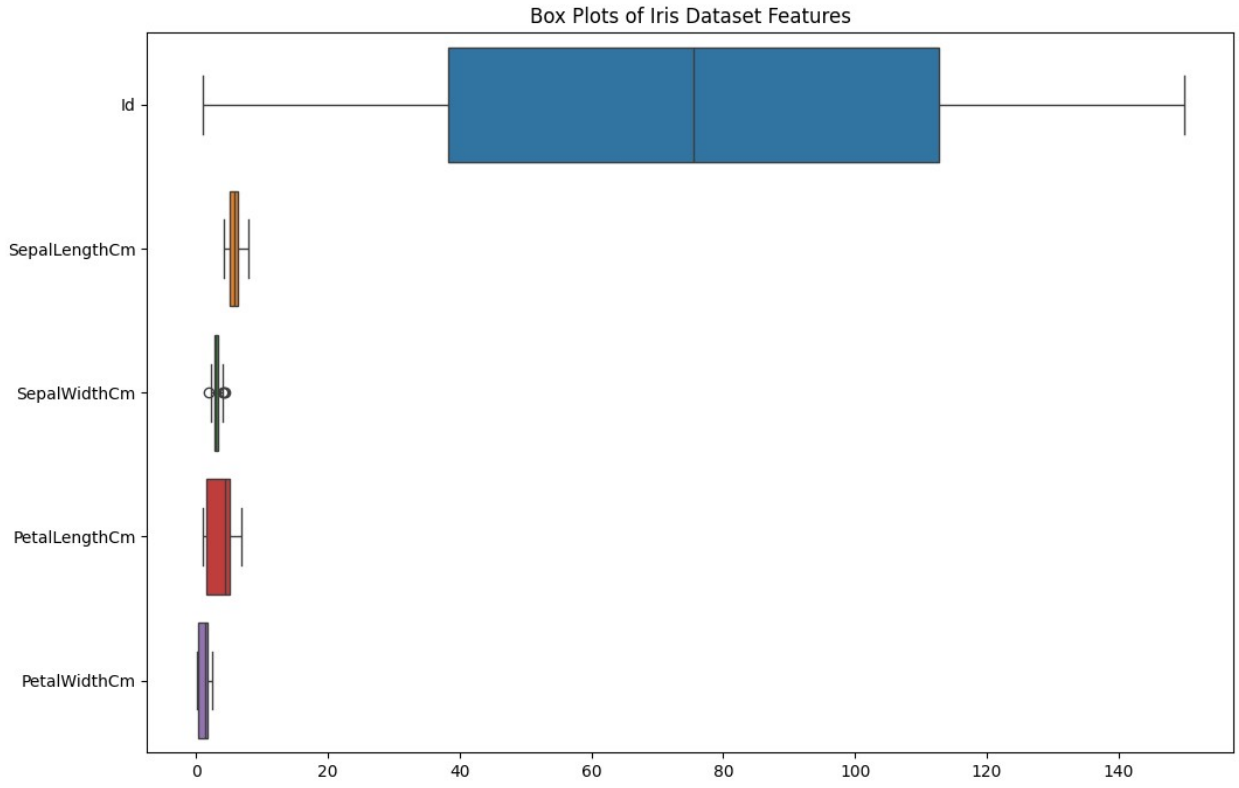


Univariate Analysis

```
# Histograms for each feature
df.hist(figsize=(10, 7))
plt.suptitle('Histograms of Iris Dataset Features')
plt.show()
```

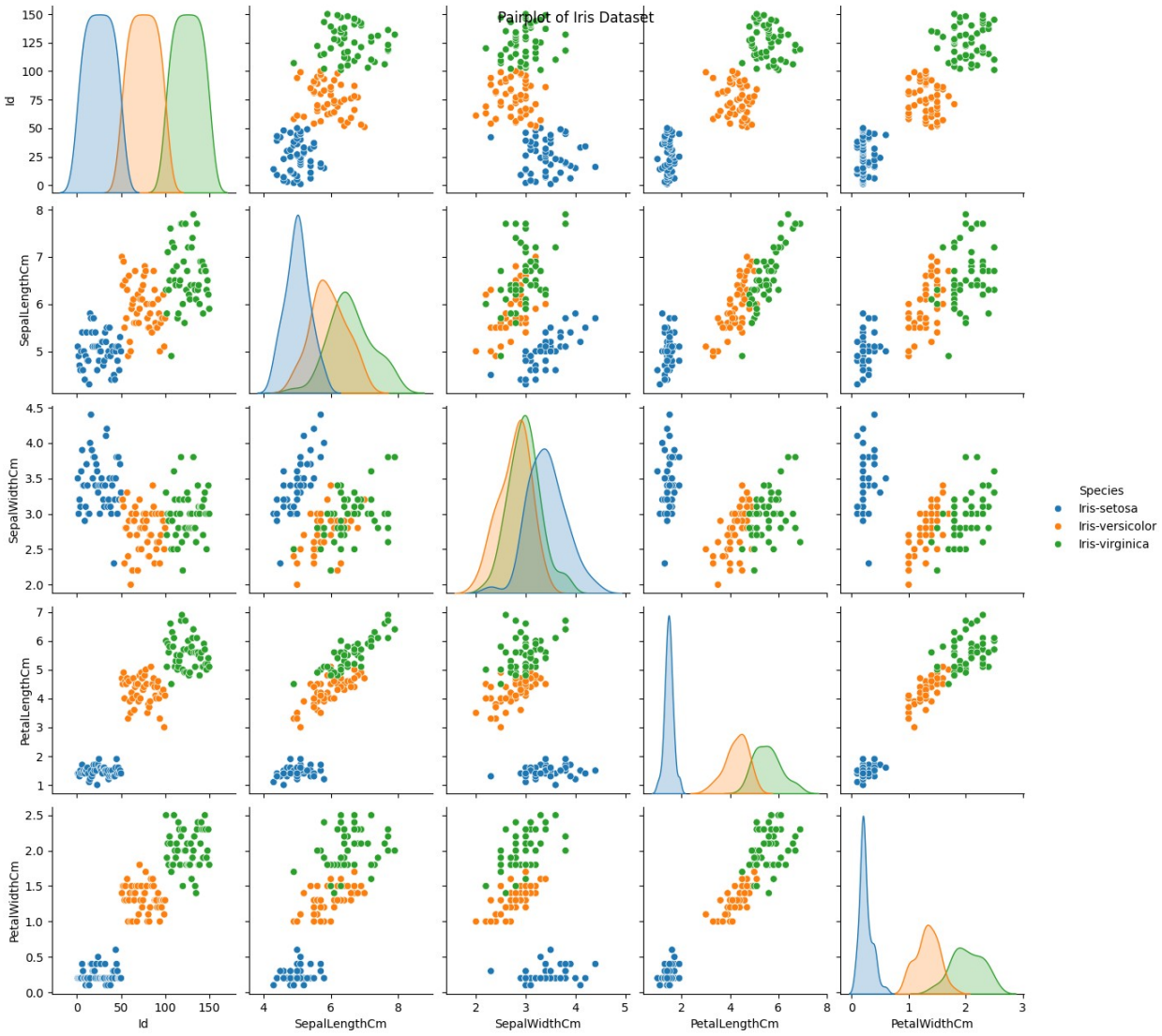



```
# Box plots for each feature
plt.figure(figsize=(12, 8))
sns.boxplot(data=df, orient='h')
plt.title('Box Plots of Iris Dataset Features')
plt.show()
```

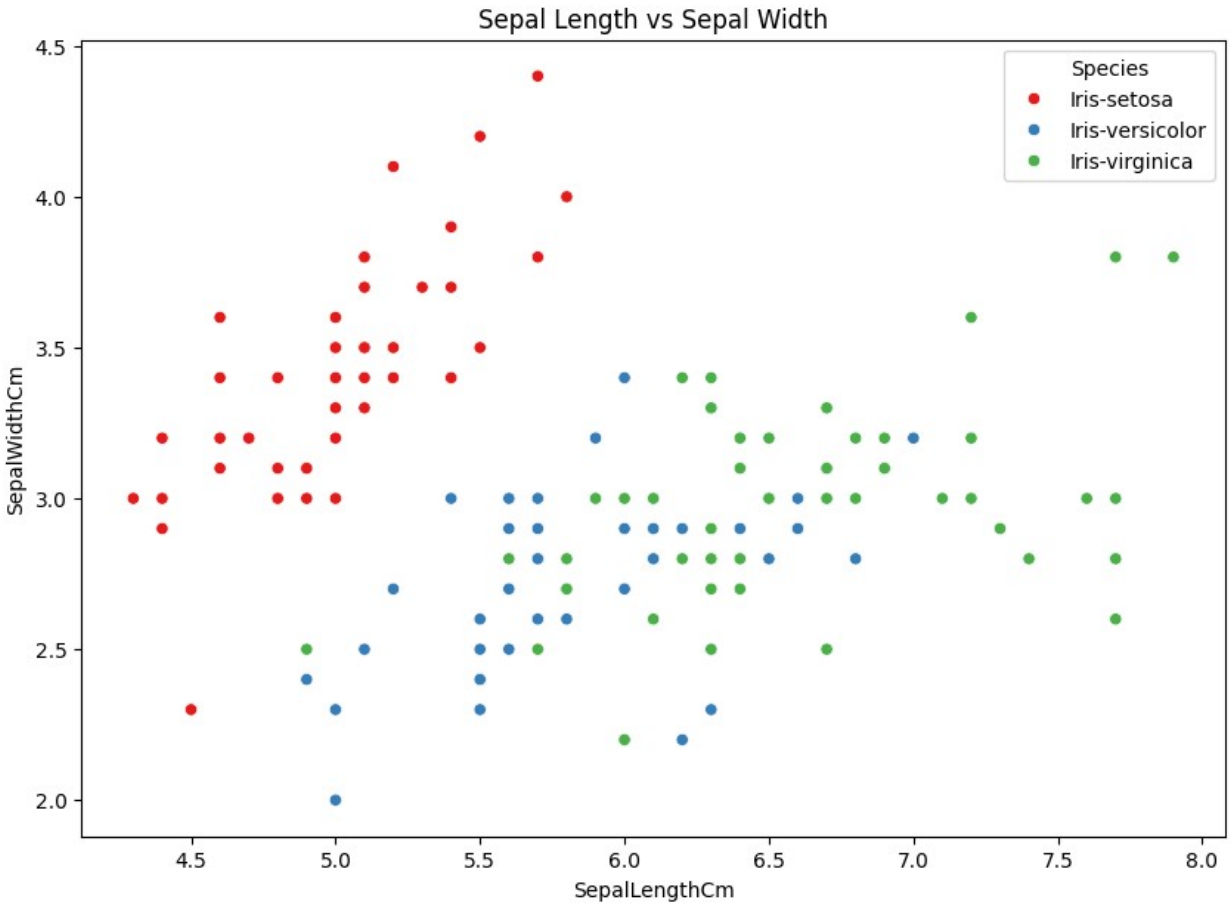


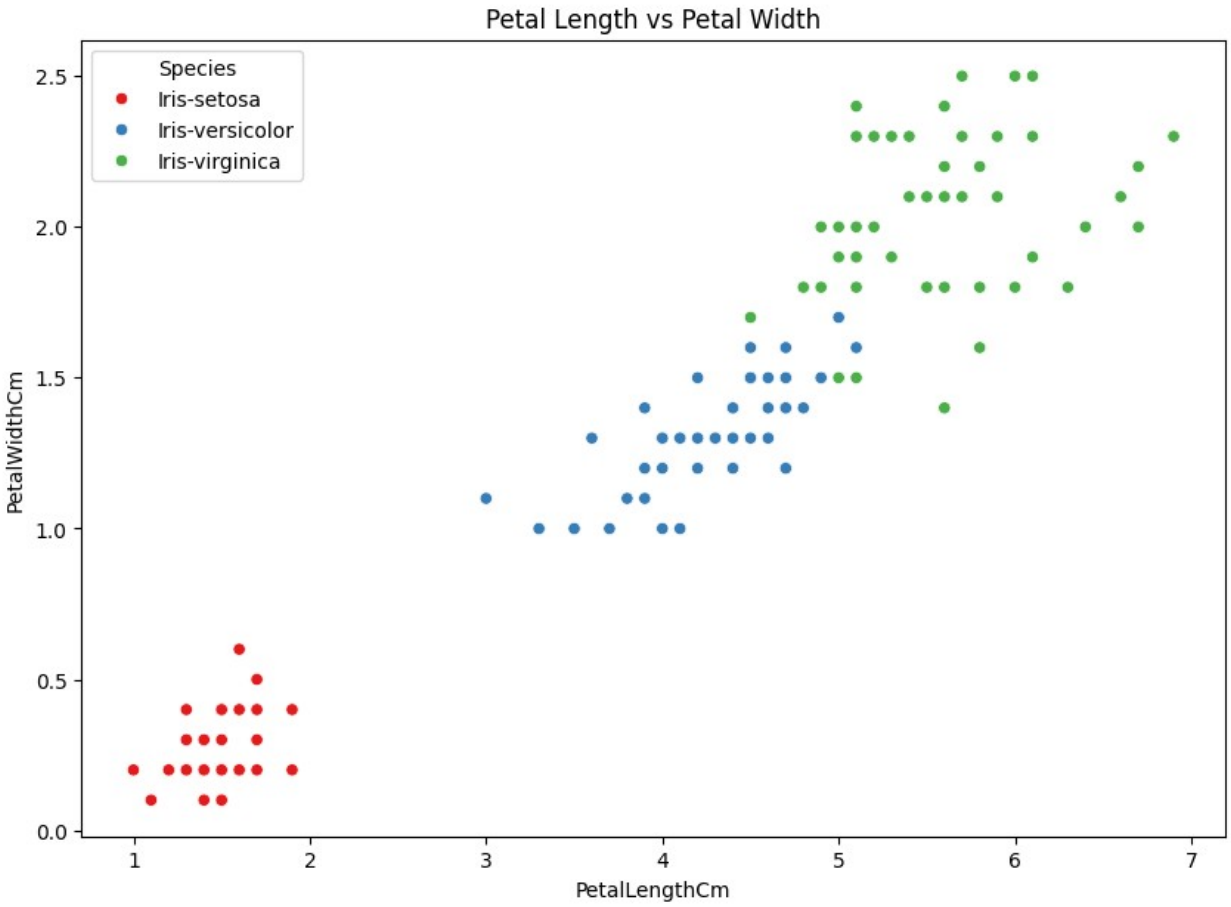
Bivariate Analysis

```
# Pairplot to show relationships between all features
sns.pairplot(df, hue='Species')
plt.suptitle('Pairplot of Iris Dataset')
plt.show()
```



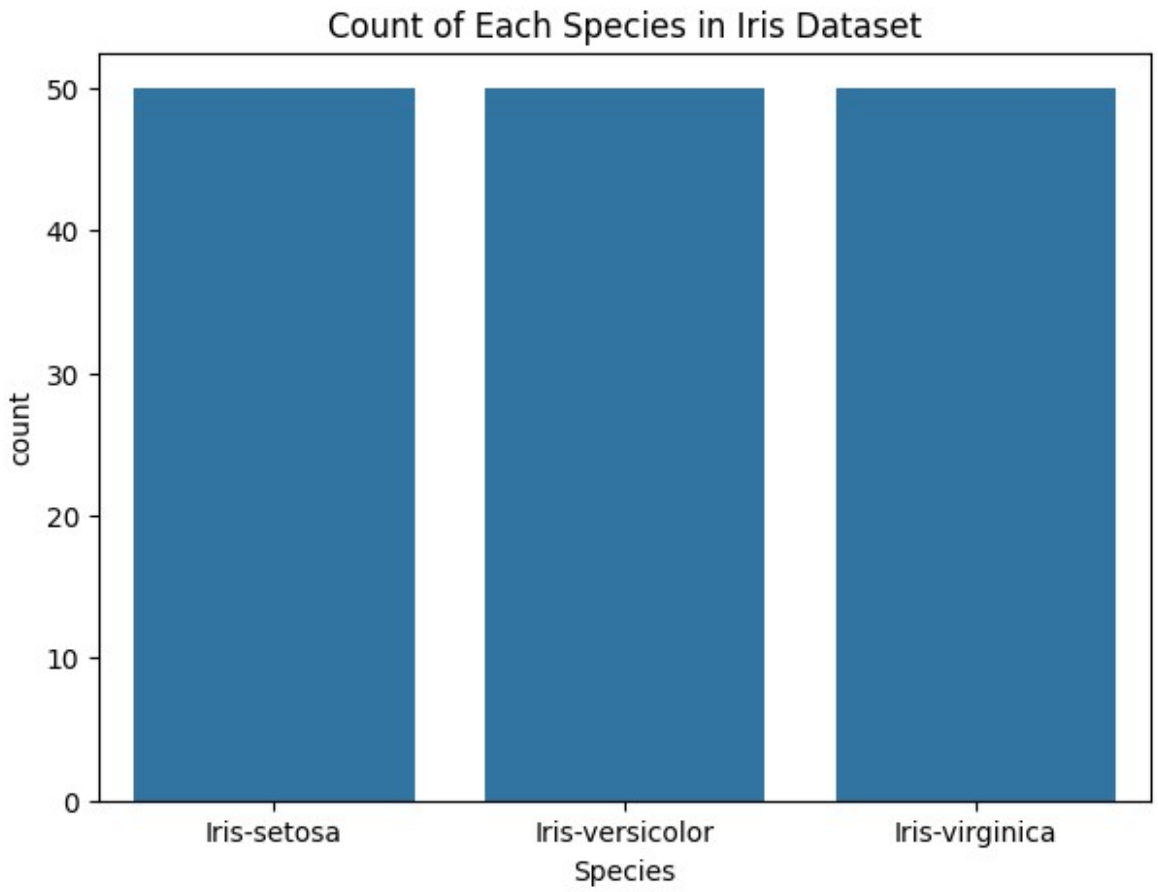

```
# Scatter plots for feature relationships
plt.figure(figsize=(10, 7))
sns.scatterplot(data=df, x='SepalLengthCm', y='SepalWidthCm',
hue='Species', palette='Set1')
plt.title('Sepal Length vs Sepal Width')
plt.show()

plt.figure(figsize=(10, 7))
sns.scatterplot(data=df, x='PetalLengthCm', y='PetalWidthCm',
hue='Species', palette='Set1')
plt.title('Petal Length vs Petal Width')
plt.show()
```


Sepal Length vs Sepal Width
Species
Iris-setosa
Iris-versicolor
Iris-virginica
SepalWidthCm
SepalLengthCm
4.5
4.0
3.5
3.0
2.5
2.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0



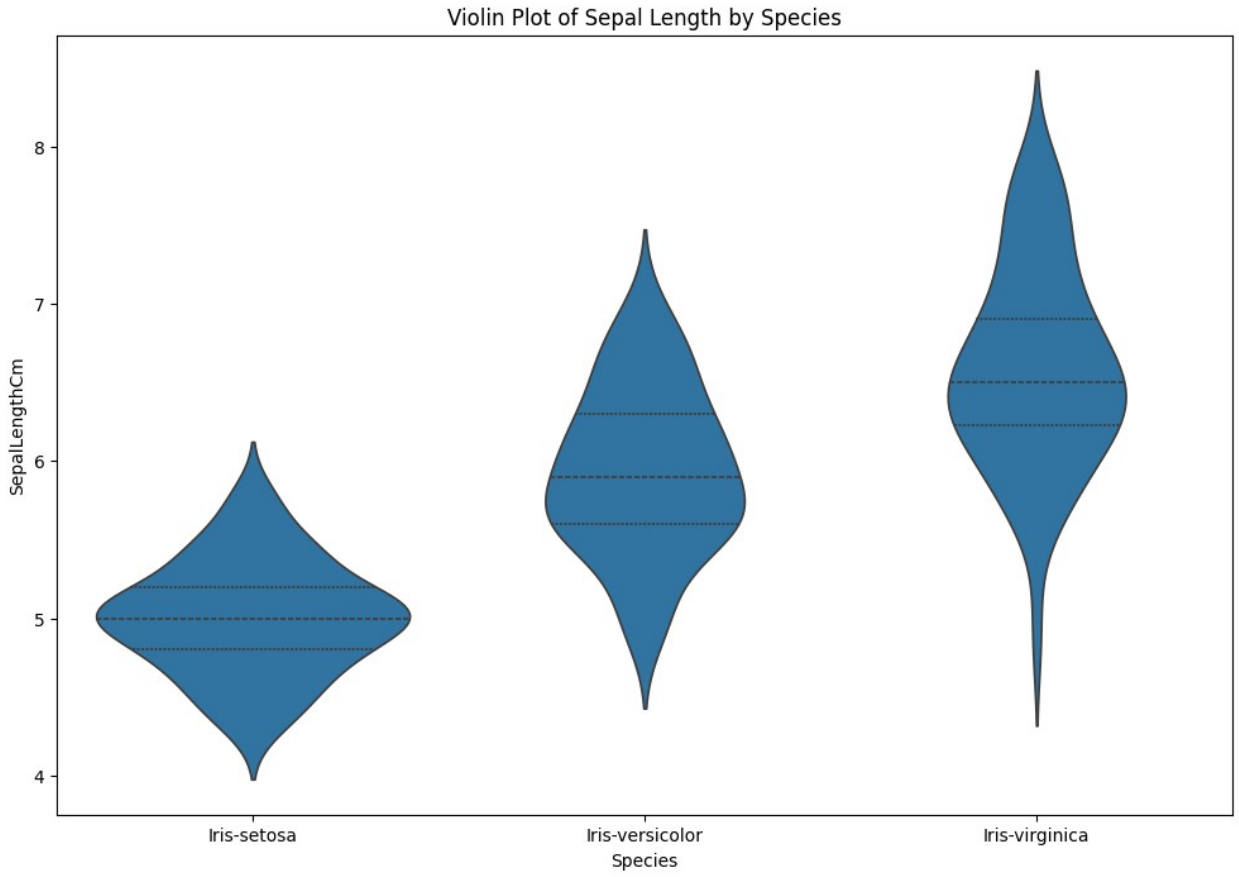
Distribution of Target Variable

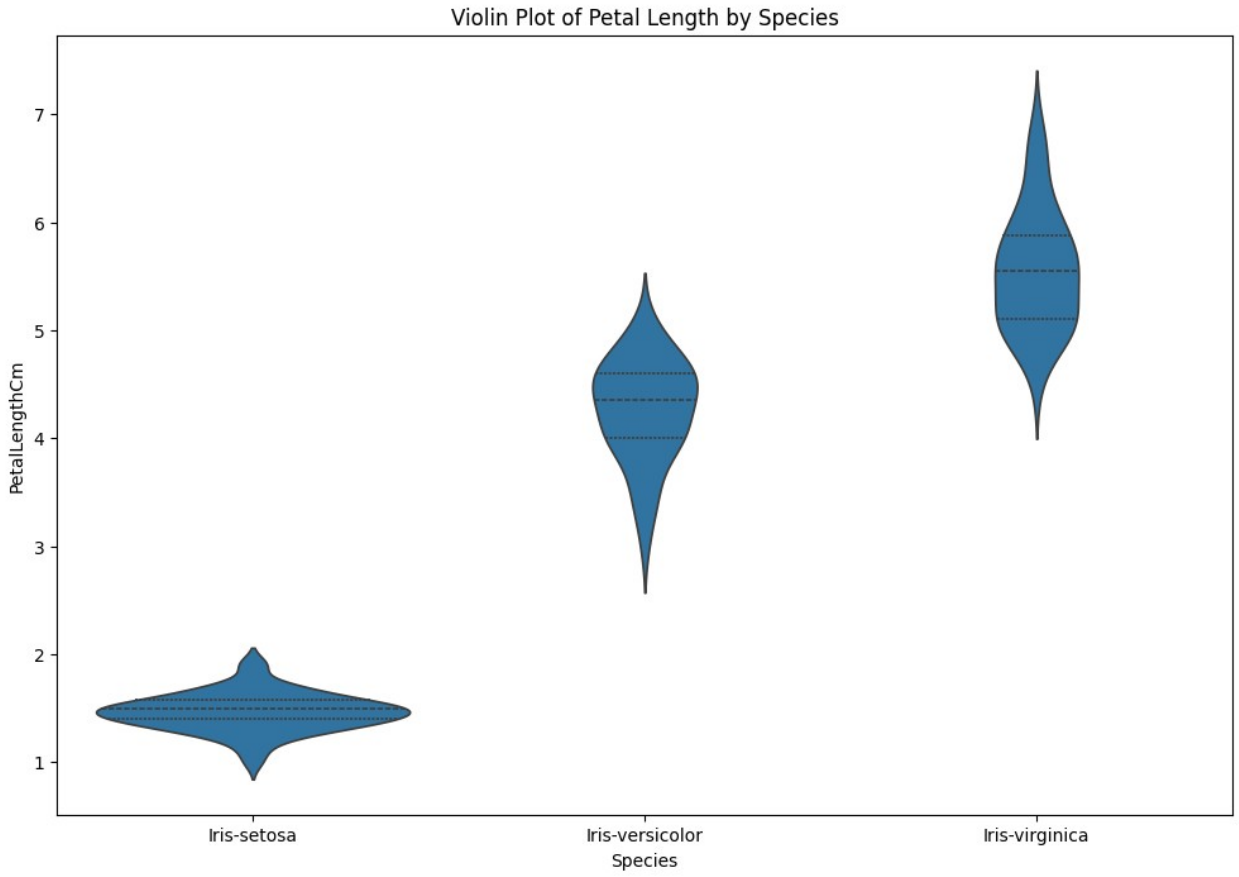
```
# Count plot of the species
plt.figure(figsize=(7, 5))
sns.countplot(data=df, x='Species')
plt.title('Count of Each Species in Iris Dataset')
plt.show()
```



Advanced Visualization

```
# Violin plots for feature distributions by species
plt.figure(figsize=(12, 8))
sns.violinplot(data=df, x='Species', y='SepalLengthCm',
inner='quartile')
plt.title('Violin Plot of Sepal Length by Species')
plt.show()

plt.figure(figsize=(12, 8))
sns.violinplot(data=df, x='Species', y='PetalLengthCm',
inner='quartile')
plt.title('Violin Plot of Petal Length by Species')
plt.show()
```


Violin Plot of Sepal Length by Species
SepalLengthCm
8
7
6
5
4
Iris-setosa
Iris-versicolor
Iris-virginica
Species



1. Histograms:

Analysis:

Distribution of Features: Histograms reveal the distribution of sepal length, sepal width, petal length, and petal width. Skewness: Check for skewness in feature distributions. Observations:

Sepal Length: Values range from approximately 4.3 to 7.9 cm, with most values clustering between 5.0 and 6.0 cm. Petal Length: Shows a clear separation, with Iris-setosa clustering around 1.0-1.5 cm, Iris-versicolor around 4.0-5.0 cm, and Iris-virginica around 4.5-6.9 cm.

1. Box Plots:

Analysis:

Outliers: Identifies any outliers in features using box plots. Spread and Central Tendency: Compares the spread and median of features across species. Observations:

Petal Length: Iris-setosa: Median ~1.5 cm, IQR from 1.4 to 1.6 cm. Iris-versicolor: Median ~4.5 cm, IQR from 4.0 to 5.0 cm. Iris-virginica: Median ~5.5 cm, IQR from 4.5 to 6.0 cm. Outliers: Iris-virginica may show outliers with petal lengths exceeding 6.5 cm.

1. Pair Plot:

Analysis:

Relationships Between Features: Visualizes scatterplots of feature pairs. Species Separation: Checks for clear separation between species based on feature combinations. Observations:

Petal Length vs. Petal Width: Shows clear clustering of different species, indicating effective separation. Sepal Length vs. Sepal Width: Iris-setosa has a smaller range for sepal width (~3.0-4.0 cm) and moderate range for sepal length (~4.3-5.8 cm), while Iris-virginica shows a larger range.

1. Scatter Plots:

Analysis:

Feature Relationships: Examines relationships between specific feature pairs. Species Distinction: Determines how well scatter plots distinguish between species. Observations:

Petal Length vs. Petal Width: Iris-setosa: Petal Length ~1.0-1.5 cm, Petal Width ~0.1-0.3 cm. Iris-versicolor: Petal Length ~3.0-5.0 cm, Petal Width ~1.0-1.5 cm. Iris-virginica: Petal Length ~4.5-6.9 cm, Petal Width ~1.5-2.5 cm. Clustering: Clear separation is observed among species, particularly with petal features.

1. Correlation Matrix:

Analysis:

Correlation Coefficients: Analyzes correlations between numeric features. Feature Relationships: Identifies strong and weak correlations. Observations:

Petal Length and Petal Width: High positive correlation (~0.96), indicating that as one increases, the other also tends to increase. Sepal Length and Sepal Width: Weak correlation (~-0.12), suggesting minimal relationship between these features.

1. Count Plot of Species:

Analysis:

Class Distribution: Determines the distribution of different species in the dataset. Balance: Checks for balance among species. Observations:

Iris-setosa: ~50 instances. Iris-versicolor: ~50 instances. Iris-virginica: ~50 instances. Balance: The dataset is evenly balanced among the three species.

1. Violin Plots:

Analysis:

Feature Distribution by Species: Examines how feature distributions vary by species. Spread and Density: Understands density and spread of features. Observations:

Sepal Length for Iris-setosa: Median ~5.0 cm, with a narrow spread. Petal Length for Iris-virginica: Median ~5.5 cm, with a broader spread compared to other species. Summary of Key Insights:

Feature Distribution: Sepal Length and Petal Length show significant variability, with Petal Length particularly distinguishing between species. Species Separation: Petal Length and Petal

Width provide excellent separation between species, especially when compared to Sepal features. Correlation: Strong correlation between Petal Length and Petal Width highlights their interdependence. Outliers: Certain species, like Iris-virginica, exhibit outliers in petal length. Class Balance: The dataset is well-balanced, with an equal number of instances for each species. This comprehensive analysis provides a detailed understanding of the Iris dataset and highlights the relationships between its features, as well as the distinctions among species.