```
# Import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load the Iris dataset
iris_df = pd.read_csv('C:/Users/KSK/Downloads/iris/iris.csv')

# Display the first few rows of the dataset
iris_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
```

```
iris_df.isnull().sum()
```

```
Id               0
SepalLengthCm    0
SepalWidthCm     0
PetalLengthCm    0
PetalWidthCm     0
Species          0
dtype: int64
```

```
# Summary statistics
iris_df.describe()
```

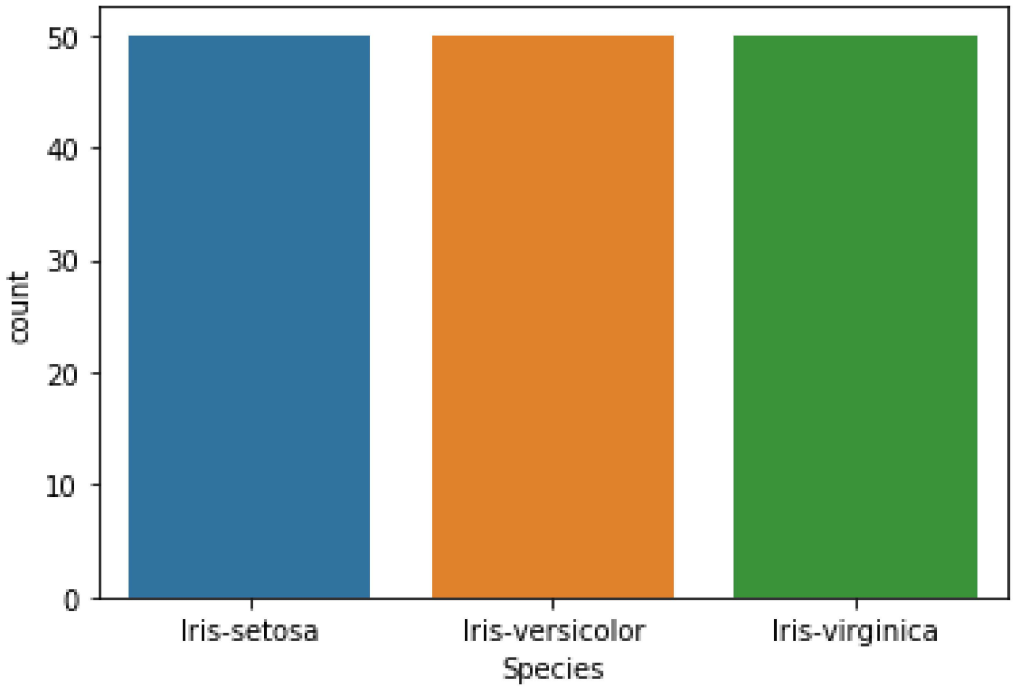
```
                Id  SepalLengthCm  SepalWidthCm  PetalLengthCm
PetalWidthCm
count  150.000000     150.000000    150.000000     150.000000
150.000000
mean    75.500000       5.843333      3.054000       3.758667
1.198667
std     43.445368       0.828066      0.433594       1.764420
0.763161
min      1.000000       4.300000      2.000000       1.000000
0.100000
25%     38.250000       5.100000      2.800000       1.600000
0.300000
50%     75.500000       5.800000      3.000000       4.350000
1.300000
```

```
75%    112.750000        6.400000      3.300000       5.100000
1.800000
max    150.000000        7.900000      4.400000       6.900000
2.500000
```

```
iris_df['Species'].value_counts()
```

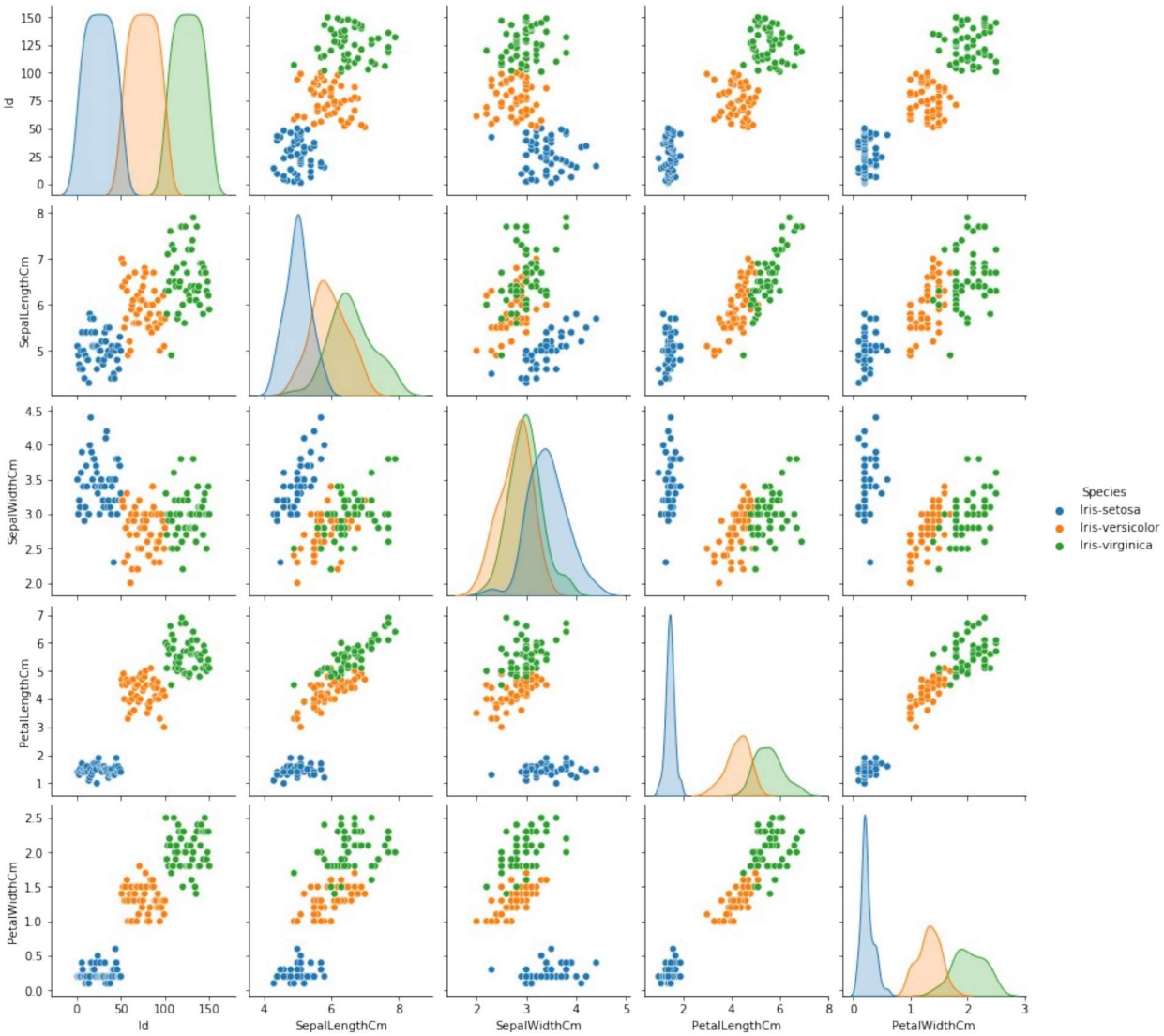
```
Iris-setosa        50
Iris-versicolor    50
Iris-virginica     50
Name: Species, dtype: int64
```

```
# visualizing target variable
sns.countplot(iris_df['Species'])
plt.show()
```

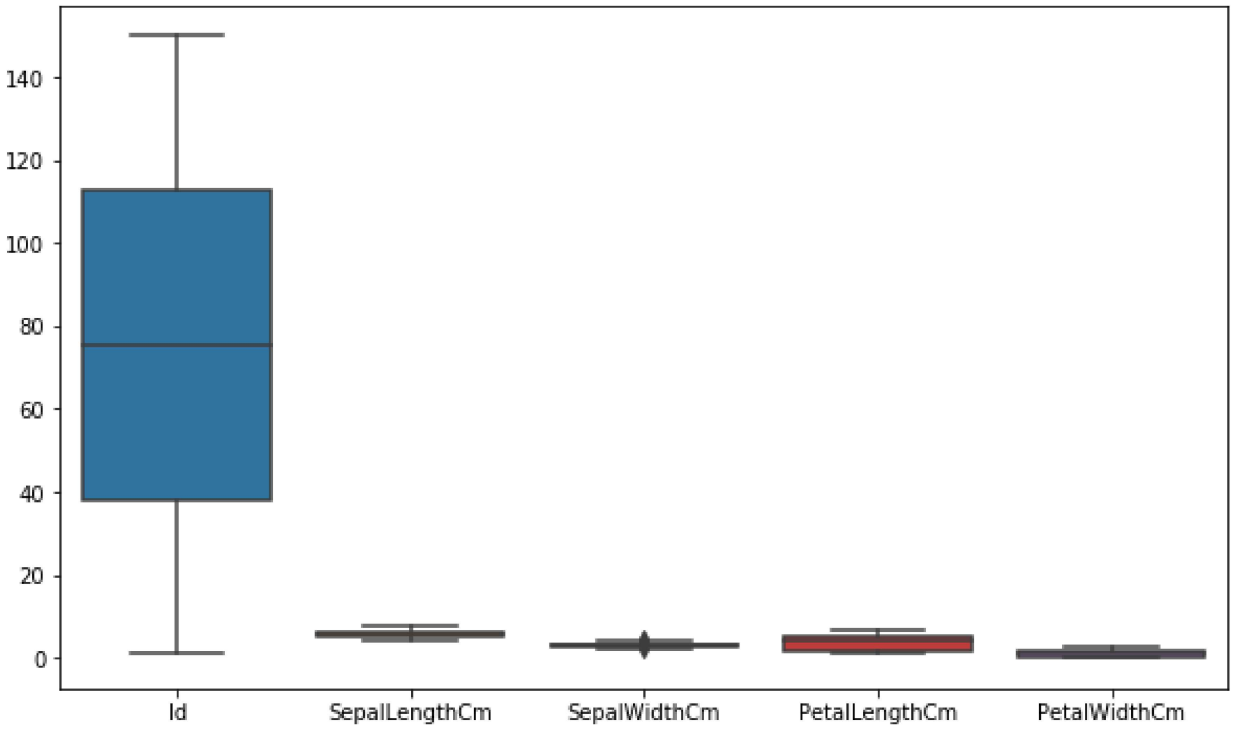


ALl the class labels are equal in number

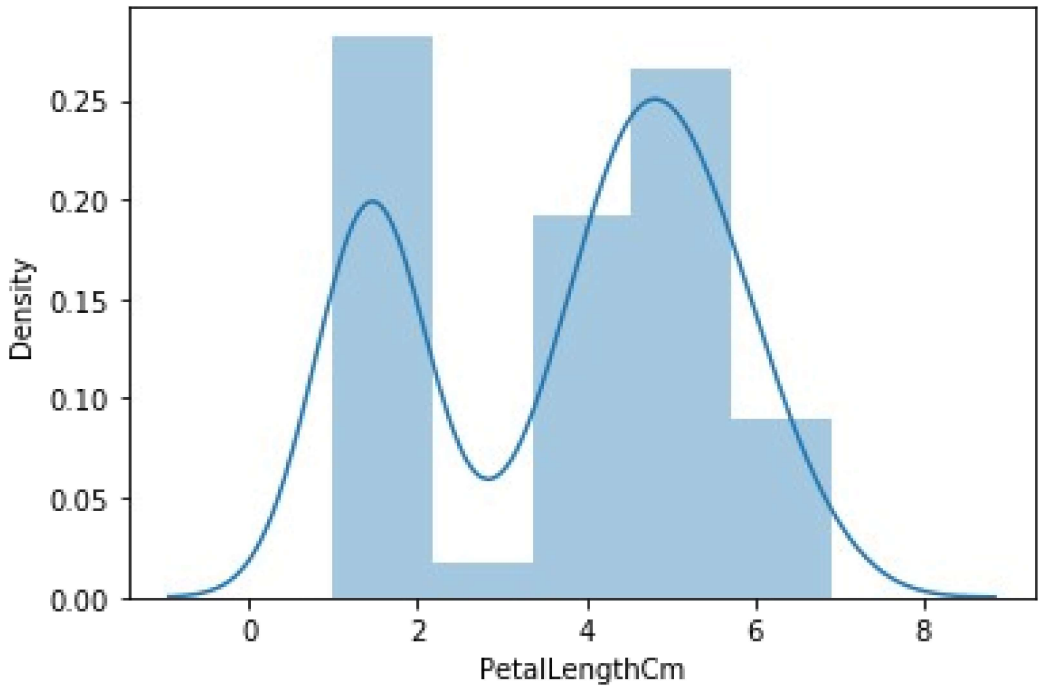
```
# Pairplot to visualize relationships between variables
sns.pairplot(iris_df, hue='Species')
plt.show()
```



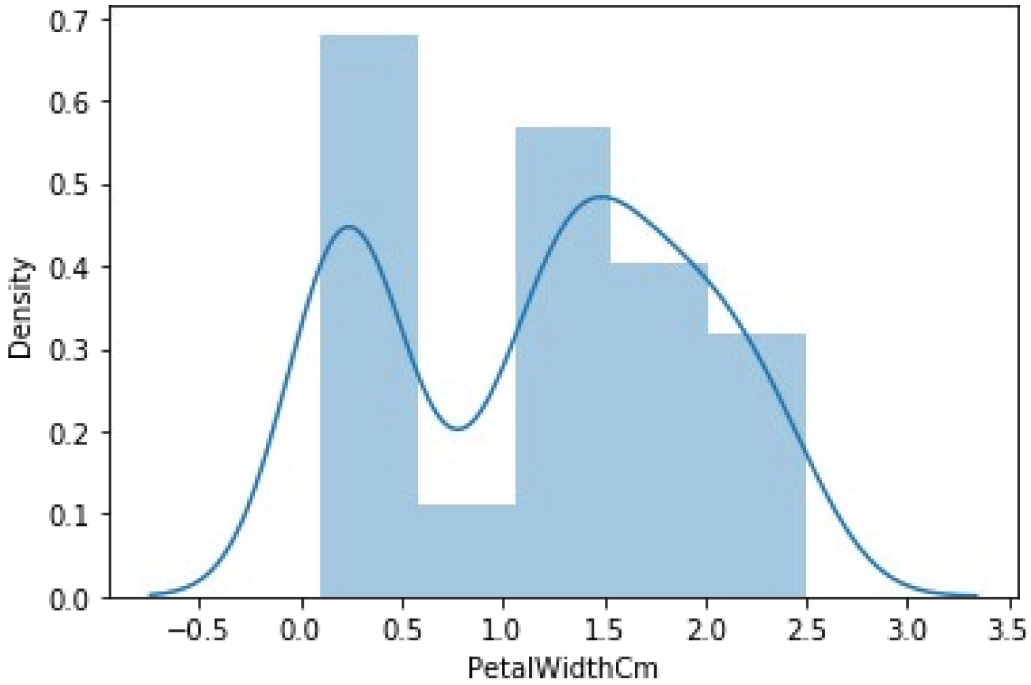

```
# Boxplot to visualize distributions of numerical variables
plt.figure(figsize=(10, 6))
sns.boxplot(data=iris_df.drop('Species', axis=1))
plt.show()
```



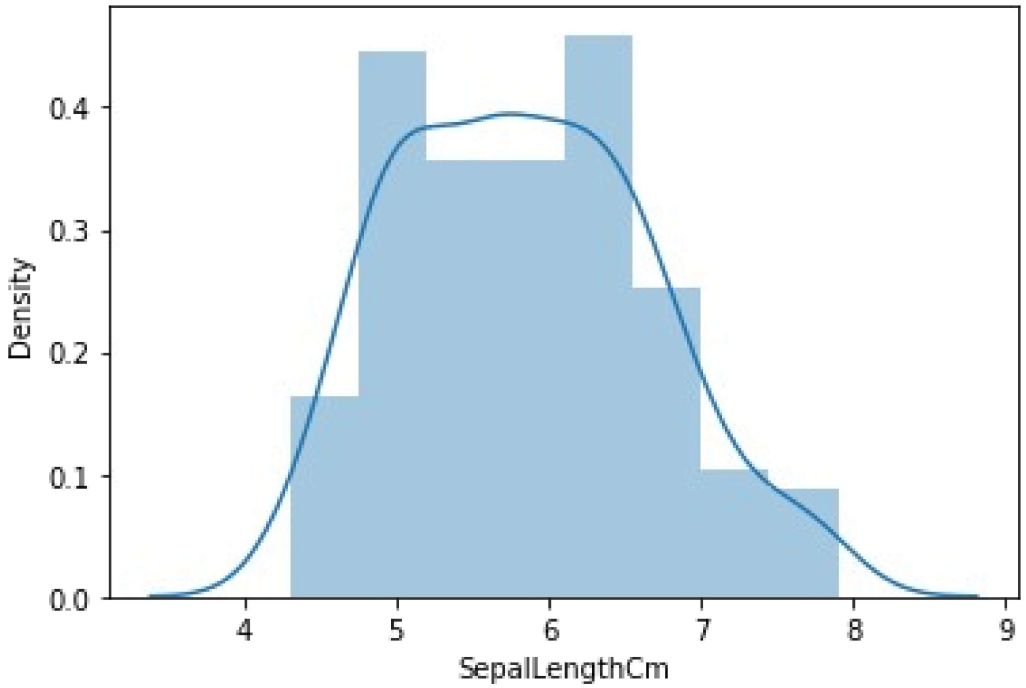

```
# Histograms for each numerical variable
import warnings
warnings.filterwarnings('ignore')
sns.distplot(iris_df['PetalLengthCm'])
plt.show()
```



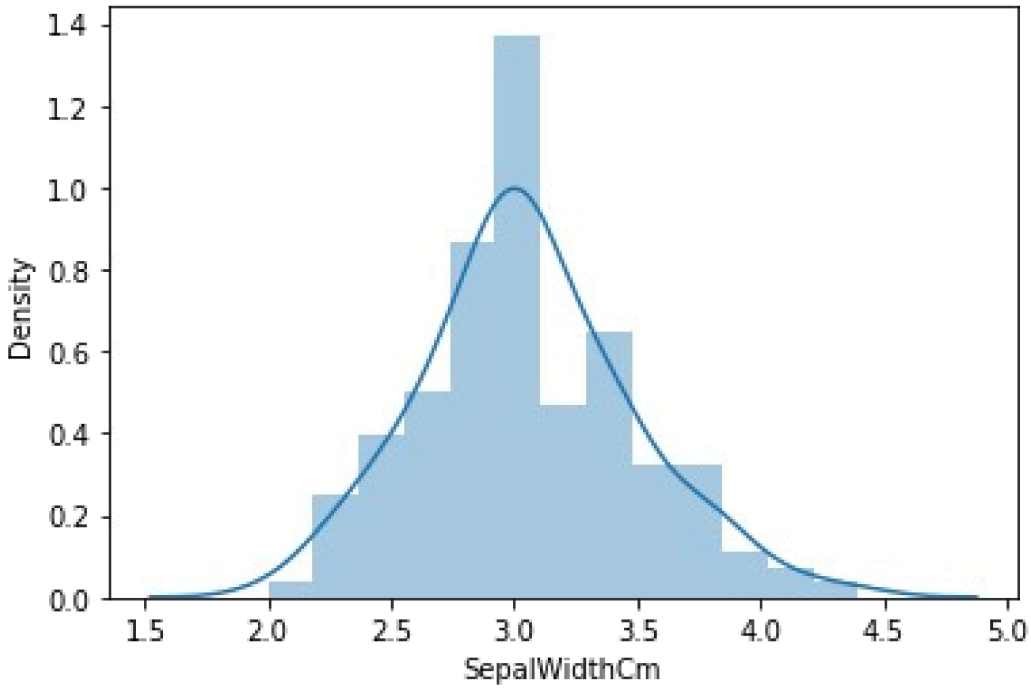
```
# Histograms for each numerical variable
import warnings
warnings.filterwarnings('ignore')
sns.distplot(iris_df['PetalWidthCm'])
plt.show()
```




```
# Histograms for each numerical variable
import warnings
warnings.filterwarnings('ignore')
sns.distplot(iris_df['SepalLengthCm'])
plt.show()
```




```
# Histograms for each numerical variable
import warnings
warnings.filterwarnings('ignore')
sns.distplot(iris_df['SepalWidthCm'])
plt.show()
```

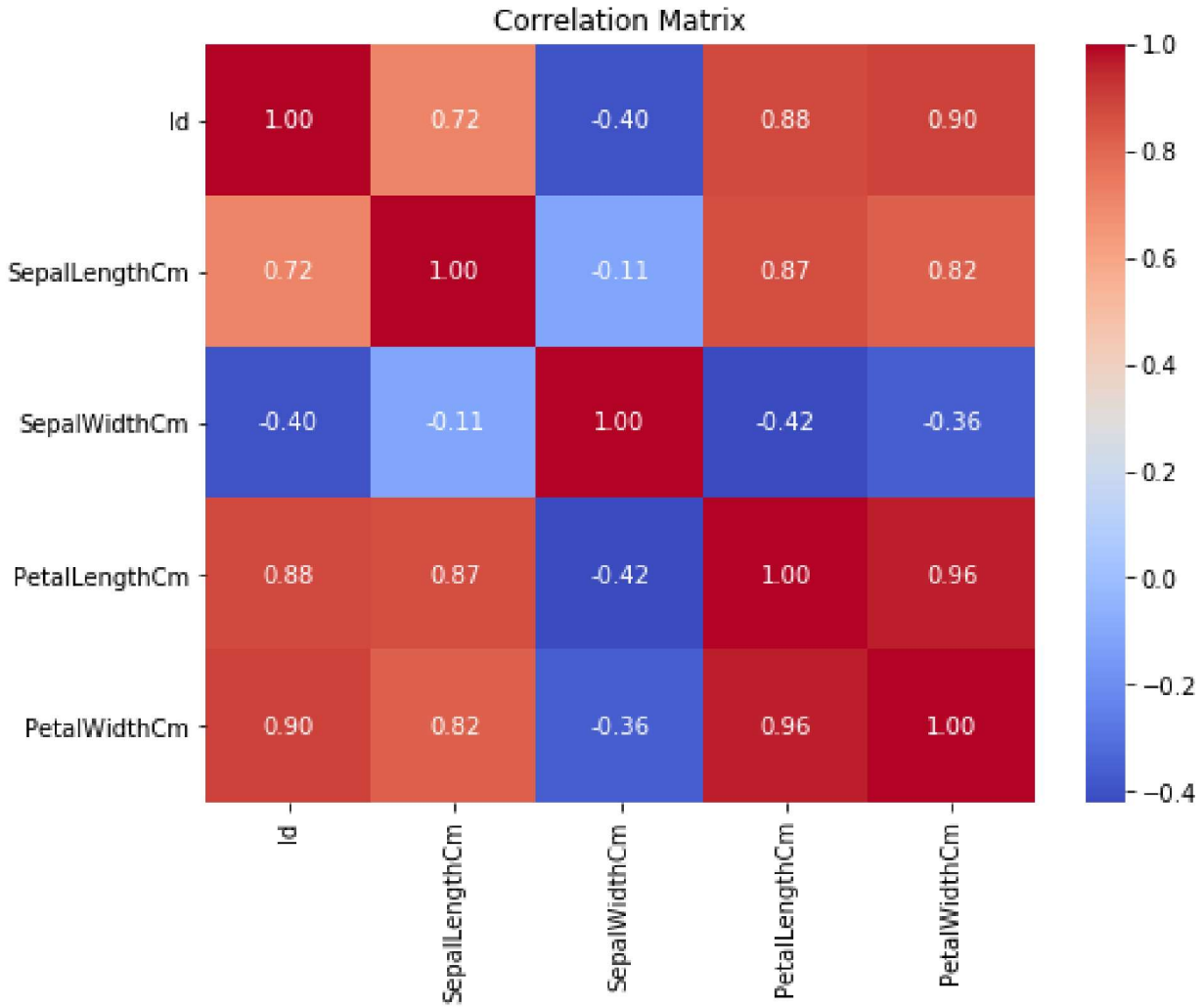


```
iris_df.corr()
```

```
                    Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  \
Id            1.000000       0.716676     -0.397729       0.882747   
SepalLengthCm 0.716676       1.000000     -0.109369       0.871754   
SepalWidthCm -0.397729      -0.109369      1.000000      -0.420516   
PetalLengthCm 0.882747       0.871754     -0.420516       1.000000   
PetalWidthCm  0.899759       0.817954     -0.356544       0.962757   

               PetalWidthCm  
Id                 0.899759  
SepalLengthCm      0.817954  
SepalWidthCm      -0.356544  
PetalLengthCm      0.962757  
PetalWidthCm       1.000000  
```

```
# Correlation matrix
plt.figure(figsize=(8, 6))
sns.heatmap(iris_df.corr(), annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Matrix')
plt.show()
```



There is a strong positive correlation between petal length and petal width