# Experiment 3.3

**Student Name: Sumant Rao** **UID:21BCS7561**
**Branch: CSE** **Section/Group: 911 "B"**
**Semester: 5th** **Date of Performance: 18/10/23**
**Subject Name: AIML** **Subject Code:21CSH-316**

## 1. Aim:

implement Exploratory Data Analysis on any data set.

## 2. Objective:

**To Learn about Meta-data**

## 3. Algorithm:

Step 1: Separate By Class.

Step 2: Summarize Dataset.

Step 3: Summarize Data By Class.

Step 4: Gaussian Probability Density Function.

Step 5: Class Probabilities.

## 4. Code:

```
[1]  import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
     from sklearn.datasets import load_iris
```

```
[2]  iris = load_iris()
     df = pd.DataFrame(data=iris.data, columns=iris.feature_names)
     df['target'] = iris.target
     df['species'] = df['target'].map({0: 'setosa', 1: 'versicolor', 2: 'virginica'})
```

```
[3]  # Display the first few rows of the dataset
     print(df.head())

     # Get summary statistics of the dataset
     print(df.describe())
```

```
[4]  sns.pairplot(df, hue='species')
     plt.show()
```

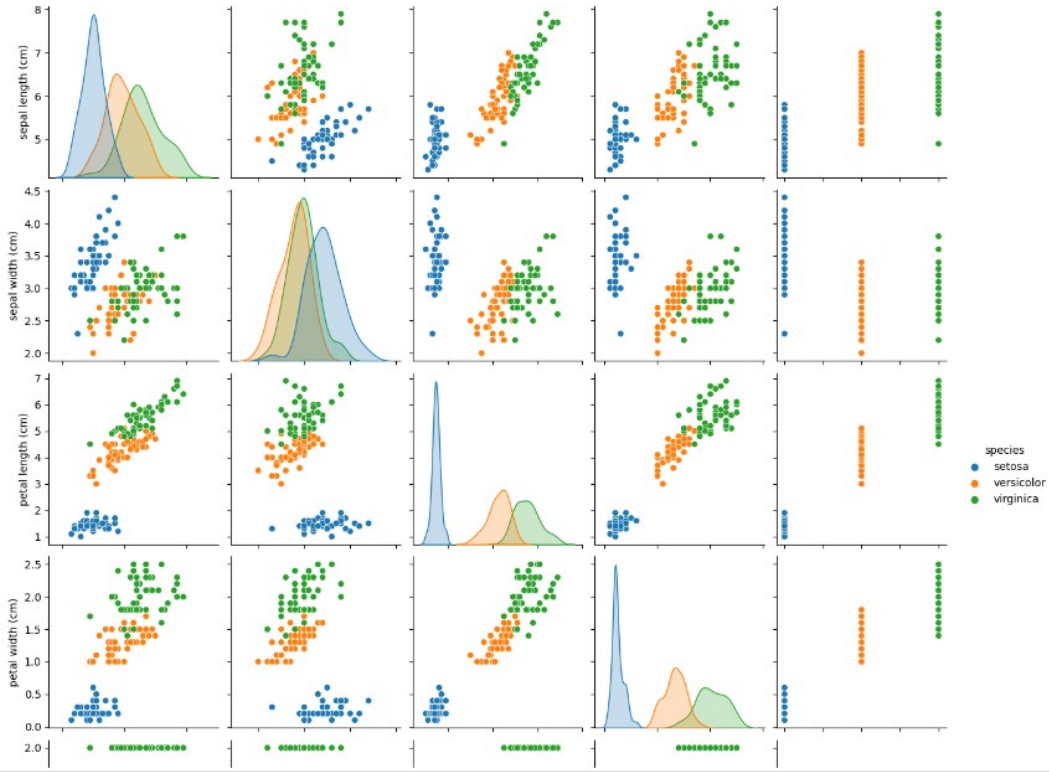
```
0s   plt.figure(figsize=(10, 6))
     sns.boxplot(x='species', y='sepal length (cm)', data=df)
     plt.xlabel('Species')
     plt.ylabel('Sepal Length (cm)')
     plt.show()
```

```
plt.figure(figsize=(10, 6))
for species in df['species'].unique():
    sns.distplot(df[df['species'] == species]['petal width (cm)'], label=species)
plt.xlabel('Petal Width (cm)')
plt.legend()
plt.show()
```

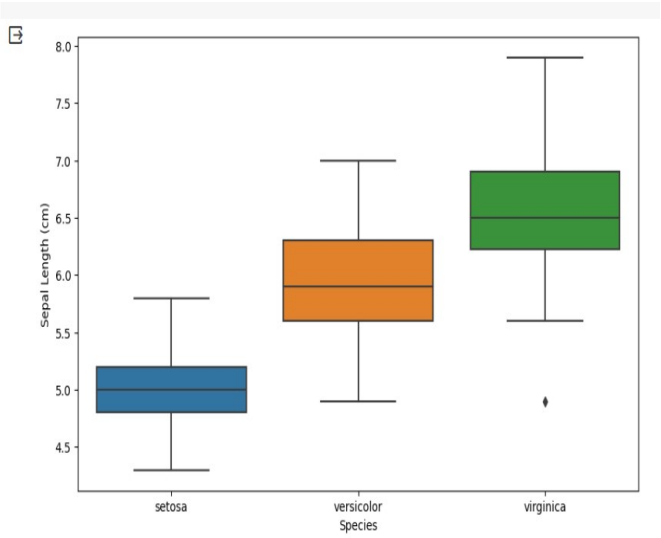
```
correlation_matrix = df.corr()
plt.figure(figsize=(10, 6))
sns.heatmap(correlation_matrix, annot=True, cmap="YlGnBu")
plt.show()
```

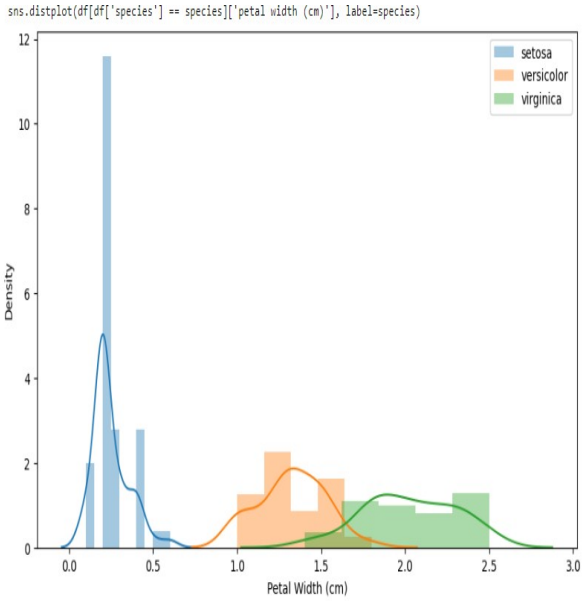
## 5. Output

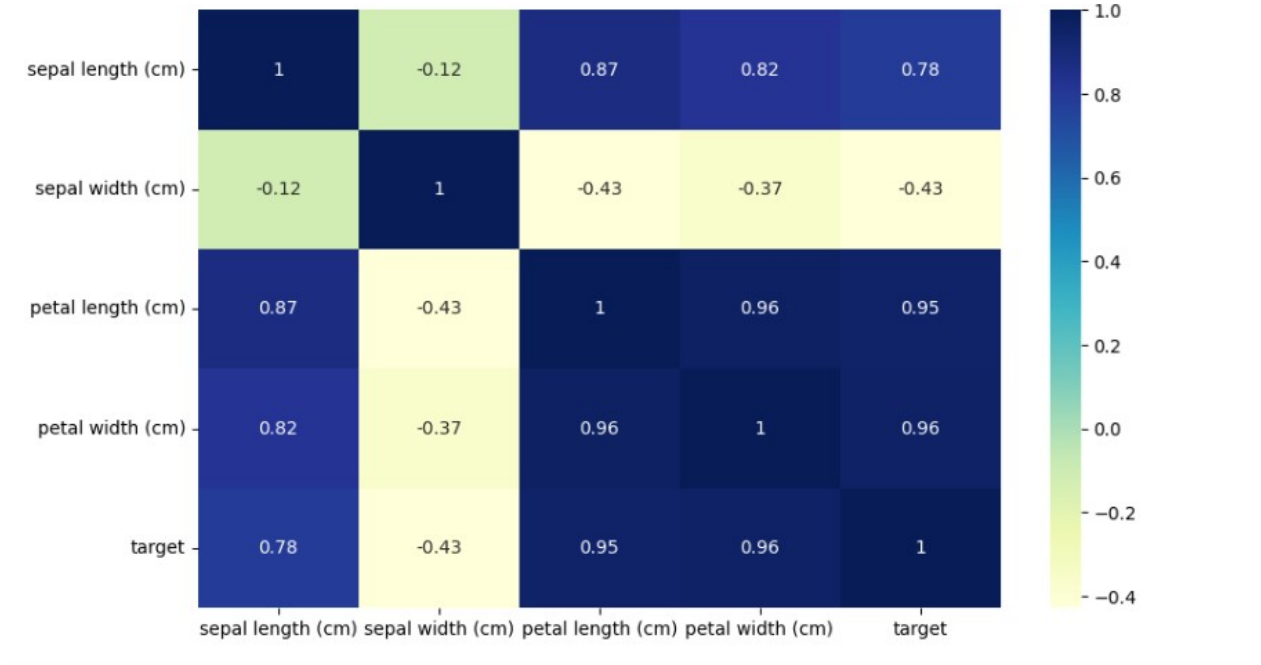
```
   sepal length (cm)  sepal width (cm)  petal length (cm)  petal width (cm)  \
0                5.1               3.5                1.4               0.2
1                4.9               3.0                1.4               0.2
2                4.7               3.2                1.3               0.2
3                4.6               3.1                1.5               0.2
4                5.0               3.6                1.4               0.2

   target species
0       0  setosa
1       0  setosa
2       0  setosa
3       0  setosa
4       0  setosa
       sepal length (cm)  sepal width (cm)  petal length (cm)  \
count         150.000000        150.000000         150.000000
mean            5.843333          3.057333           3.758000
std             0.828066          0.435866           1.765298
min             4.300000          2.000000           1.000000
25%             5.100000          2.800000           1.600000
50%             5.800000          3.000000           4.350000
75%             6.400000          3.300000           5.100000
max             7.900000          4.400000           6.900000

       petal width (cm)      target
count        150.000000  150.000000
mean           1.199333    1.000000
std            0.762238    0.819232
min            0.100000    0.000000
25%            0.300000    0.000000
50%            1.300000    1.000000
75%            1.800000    2.000000
max            2.500000    2.000000
```




+ Code + Text


Sepal Length (cm)
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
setosa
versicolor
virginica
Species


sns.distplot(df[df['species'] == species]['petal width (cm)'], label=species)
setosa
versicolor
virginica
Density
12
10
8
6
4
2
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Petal Width (cm)


sepal length (cm)
sepal width (cm)
petal length (cm)
petal width (cm)
target
1
-0.12
0.87
0.82
0.78
-0.12
1
-0.43
-0.37
-0.43
0.87
-0.43
1
0.96
0.95
0.82
-0.37
0.96
1
0.96
0.78
-0.43
0.95
0.96
1
sepal length (cm)
sepal width (cm)
petal length (cm)
petal width (cm)
target
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4