Name : Amit Kumar

Roll No. : 31

U.Id : 202401100400031

Branch : CSE(AI&ML)

Sec: A

Problem:

Iris Flower Classification – Classify flower species based on petal and sepal dimensions using the Iris dataset.

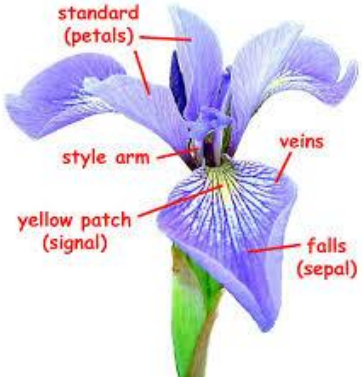
Dataset credits: kaggle

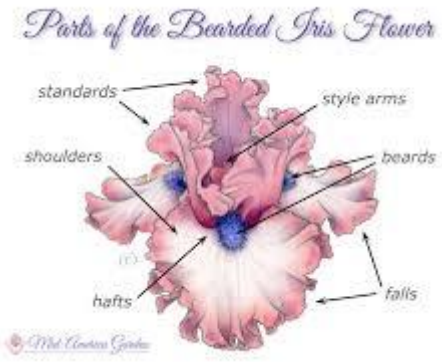
Introduction:

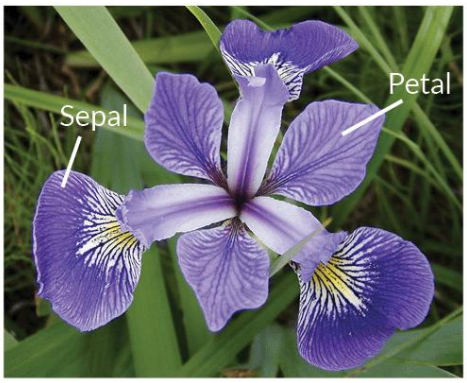
The Iris Flower Classification problem is a well-known machine learning task where the goal is to classify iris flowers into three different species (Setosa, Versicolor, and Virginica) based on the length and width of their petals and sepals. This classification is done using machine learning techniques that analyze patterns in the dataset and make predictions for new data points.

The Iris dataset, originally introduced by Ronald Fisher in 1936, consists of 150 samples with four numerical features:

- Sepal Length
- Sepal Width
- Petal Length
- Petal Width
- Species (target variable – Setosa, Versicolor, or Virginica)







Iris Versicolor

Iris Setosa

Iris Virginica

Approach to Solve the Problem:

1. **Dataset Understanding**
    - Load the **Iris dataset** containing Sepal & Petal dimensions for 3 species.
    - Check for missing values and data types using df.info().
2. **Exploratory Data Analysis (EDA)**
    - Use **scatter plots, pair plots, and bar graphs** to visualize feature relationships.
3. **Data Preprocessing**
    - Convert categorical species labels to numeric values.
    - Split data into **training (80%) and testing (20%)** sets.
4. **Simple Classification Model**
    - Use a **distance-based classification** approach (nearest neighbour).

- Predict species using **Euclidean distance**.

5. **Model Evaluation**
    - Compare **predicted vs actual labels** and calculate **accuracy**.
6. **Additional Feature: Barcode Generation**
    - Generate a barcode using the barcode library for data visualization.

This structured approach efficiently classifies **Iris flowers** using **petal and sepal measurements**, making it a great beginner-friendly **machine learning** project.

CODE:

```
# Importing necessary libraries for data processing and visualization
import numpy as np  # For numerical operations
import pandas as pd  # For data manipulation
import seaborn as sns  # For data visualization
import matplotlib.pyplot as plt  # For plotting graphs



# Load the Iris dataset from a CSV file
file_path = "/IRIS.csv"  # Path to the dataset file
df = pd.read_csv(file_path)  # Read CSV into a Pandas DataFrame
df.head()  # Display the first few rows of the dataset
```

print(df.info()) **# Print dataset information including column names, non-null counts, and data types**

sns.pairplot(df, hue='species')

plt.show()

plt.figure(figsize=(10, 5))

sns.scatterplot(x=df['sepal_length'], y=df['sepal_width'], hue=df['species'], palette='coolwarm')

plt.title("Sepal Length vs Sepal Width")

plt.show()

plt.figure(figsize=(10, 5))

sns.scatterplot(x=df['petal_length'], y=df['petal_width'], hue=df['species'], palette='viridis')

plt.title("Petal Length vs Petal Width")

plt.show()

plt.figure(figsize=(8, 5))

sns.barplot(x=df['species'], y=df['sepal_length'], palette='pastel')

plt.title("Average Sepal Length per Species")

plt.show() **# Bar plot showing the average Sepal Length for each species**

plt.figure(figsize=(8, 5))

sns.barplot(x=df['species'], y=df['petal_length'], palette='husl')

plt.title("Average Petal Length per Species")

plt.show() **# Bar plot showing the average Petal Length for each species**

 # title sepal_length

from matplotlib import pyplot as plt

df['sepal_length'].plot(kind='line', figsize=(8, 4), title='sepal_length')

plt.gca().spines[['top', 'right']].set_visible(False)

 # title sepal_length

from matplotlib import pyplot as plt

df['sepal_length'].plot(kind='hist', bins=20, title='sepal_length')

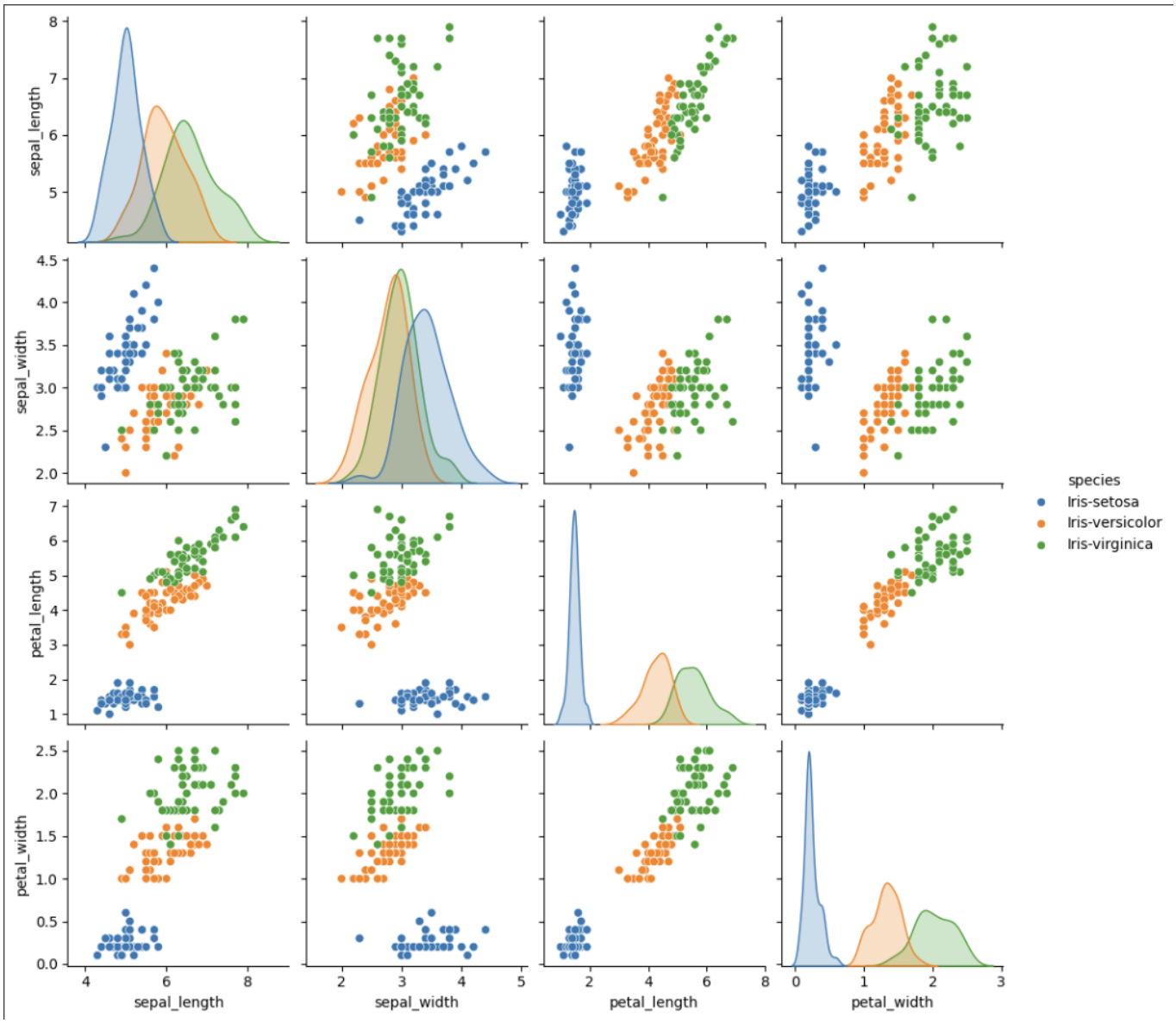
plt.gca().spines[['top', 'right',]].set_visible(False)

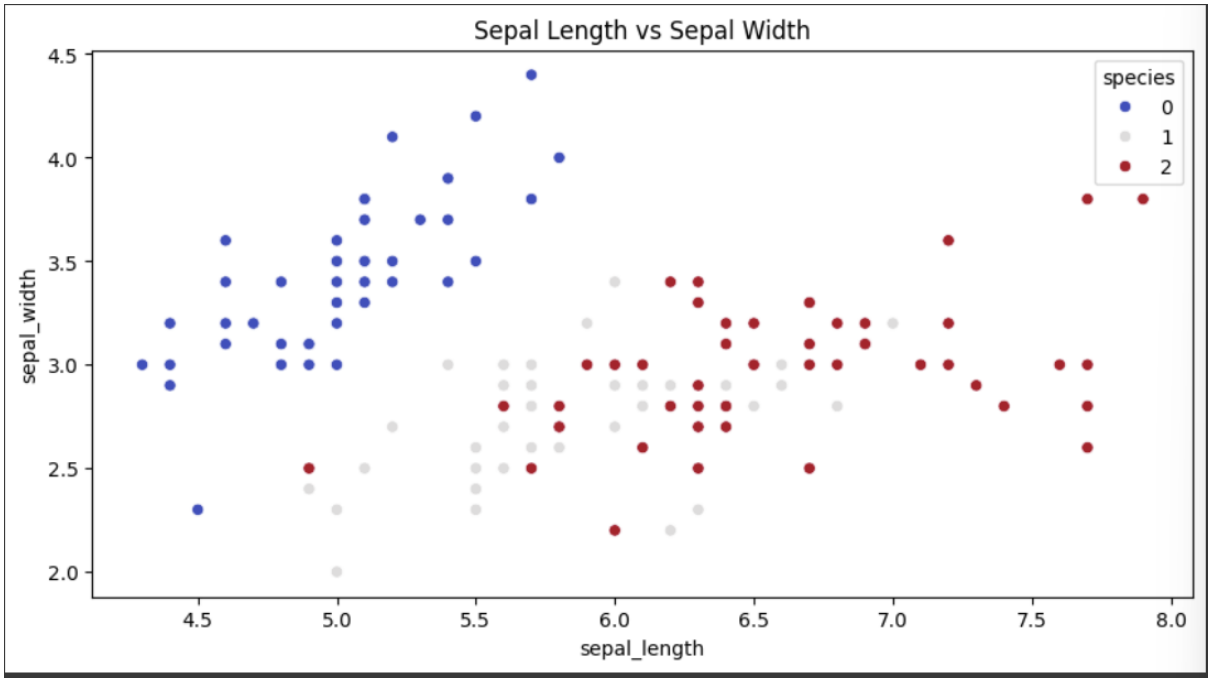
 # Cell 6: Evaluate Model Performance

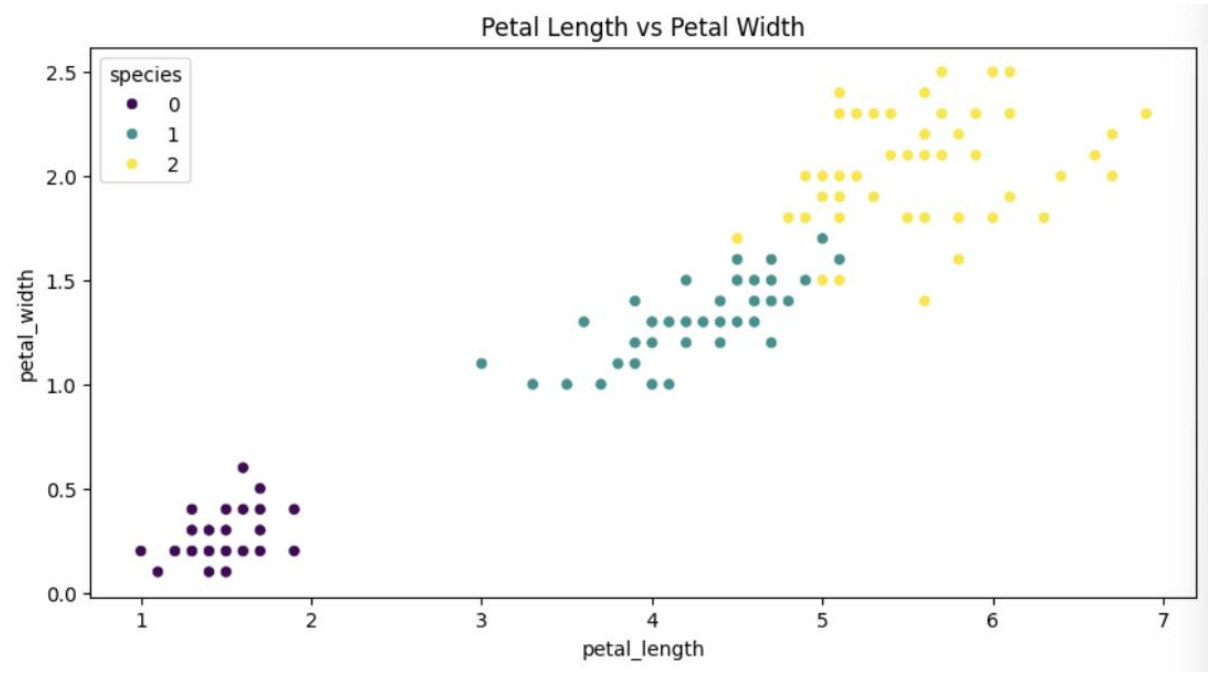
accuracy = np.mean(y_pred == y_test)

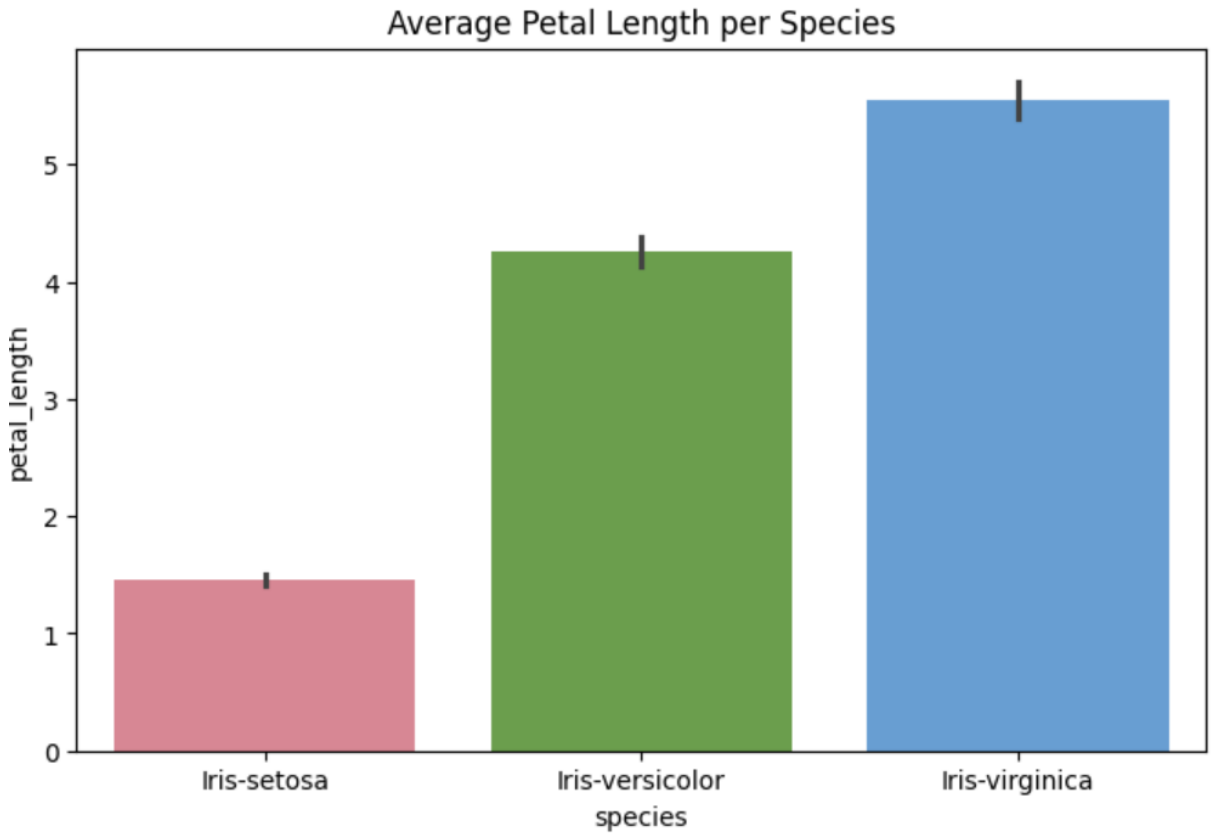
print(f'Accuracy: {accuracy:.2f}')

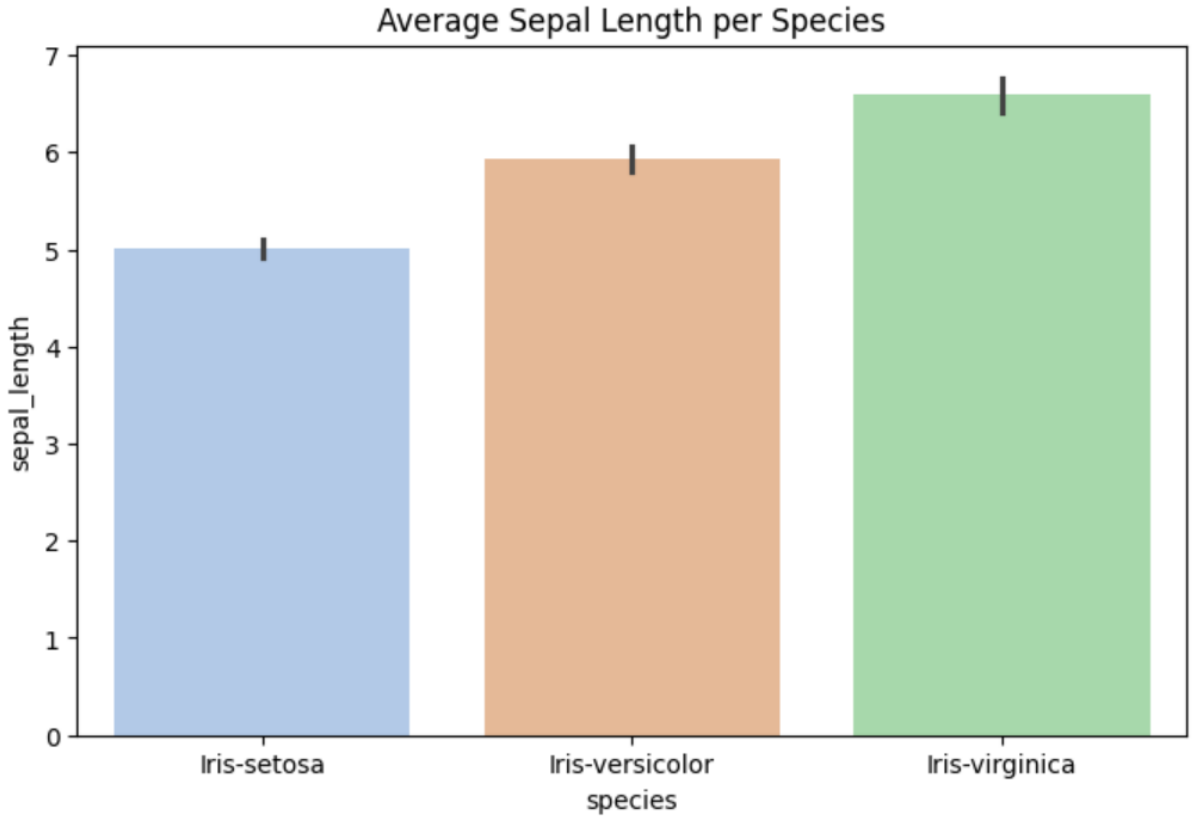
dataset information:

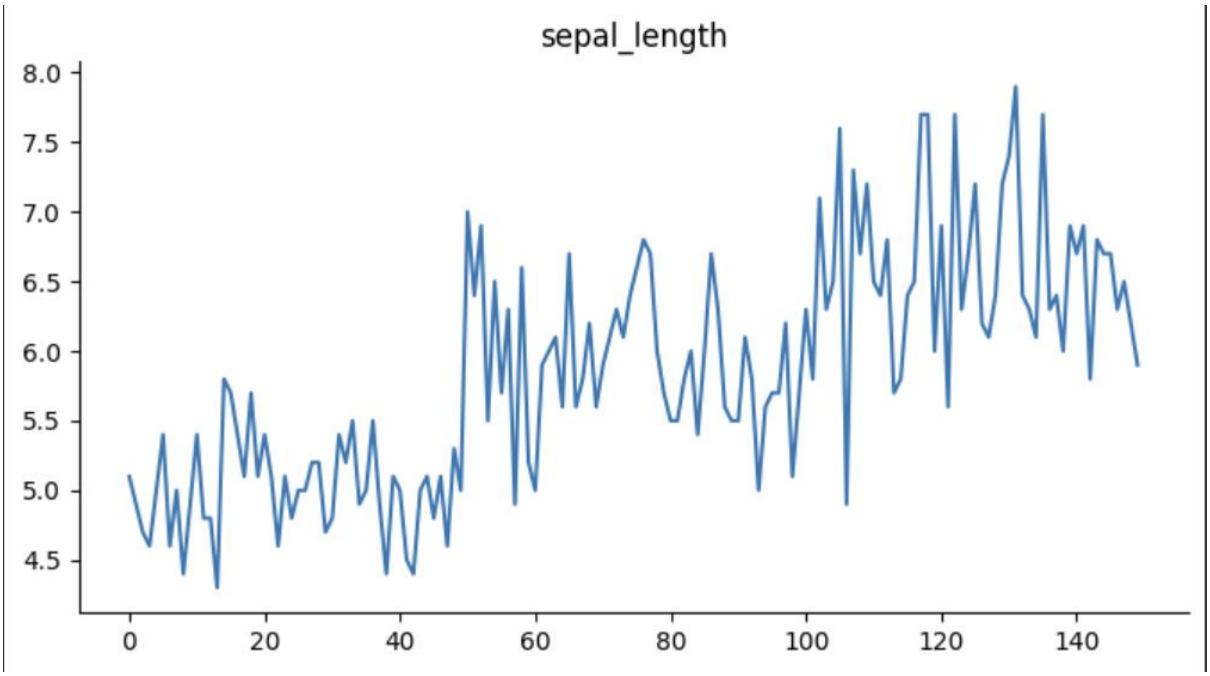


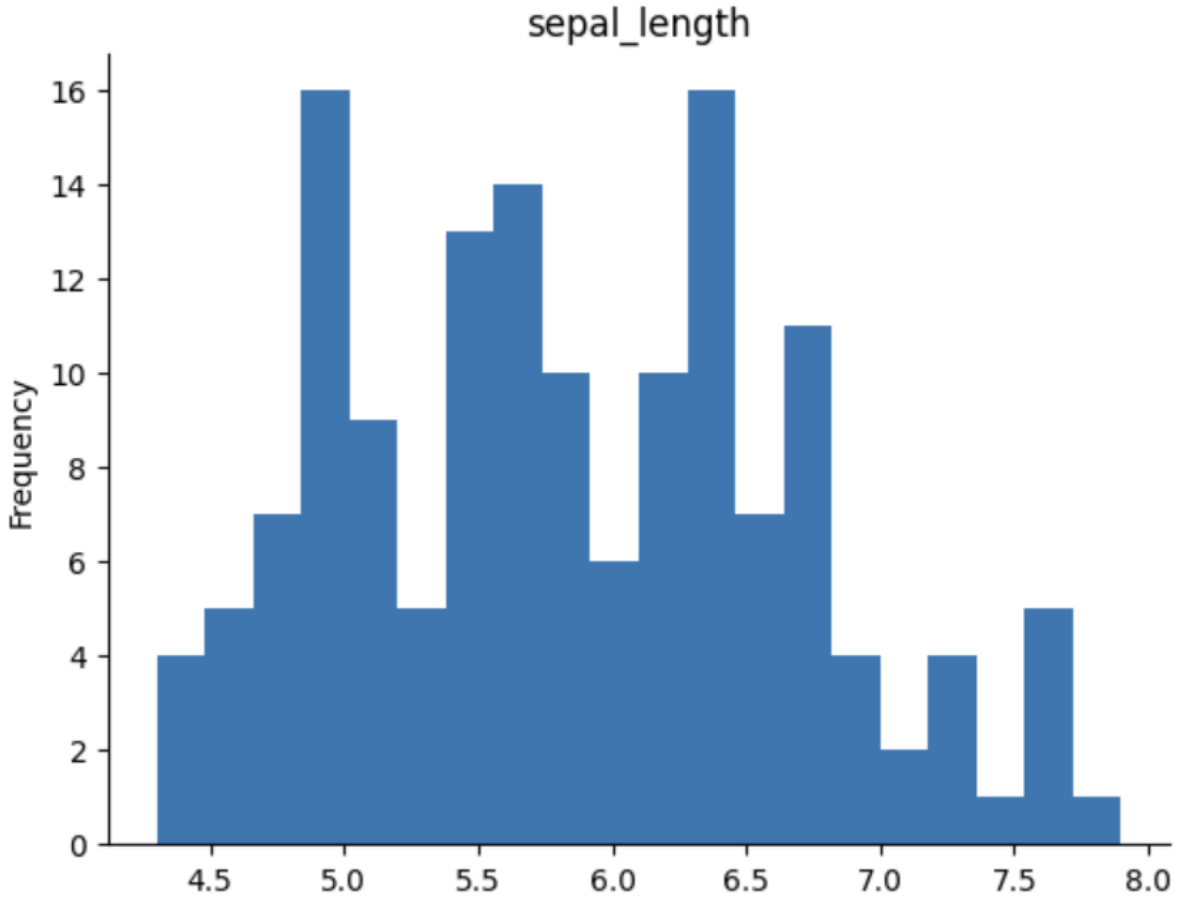

Sepal Length vs Sepal Width
species
0
1
2
sepal_width
sepal_length
4.5
4.0
3.5
3.0
2.5
2.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0


Petal Length vs Petal Width
species
0
1
2
petal_width
petal_length
2.5
2.0
1.5
1.0
0.5
0.0
1
2
3
4
5
6
7


Average Petal Length per Species
petal_length
0
1
2
3
4
5
Iris-setosa
Iris-versicolor
Iris-virginica
species


Average Sepal Length per Species
sepal_length
0
1
2
3
4
5
6
7
Iris-setosa
Iris-versicolor
Iris-virginica
species


sepal_length
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
0
20
40
60
80
100
120
140


sepal_length
Frequency
16
14
12
10
8
6
4
2
0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0