**Name:** Aravind Kumar Kaspe **Banner ID:** 001291145

**ASSIGNMENT 2**

**CSCI 5930 – Homework 2: Charts**

1. **Make the pie, histogram, and line chart in Python for iris features for sepal and petal lengths.  
   a) Copy the code and graph.  
   b) Extract information from the chart.**

**Copy the code and graph:**

1. Import all the required packages for the visualization.

import matplotlib.pyplot as plt

import pandas as pd

from sklearn.datasets import load\_iris

1. Transform the iris data into a data frame.

iris = load\_iris()

data = pd.DataFrame(data=iris.data, columns=iris.feature\_names)

data.head()

1. Separate sepal and petal features for better understanding

sepal\_features = ['sepal length (cm)', 'sepal width (cm)']

petal\_features = ['petal length (cm)', 'petal width (cm)']

1. Pie chart for Sepal Length and Petal Length mean

sepal\_lengths = data[sepal\_features[0]].mean(), data[petal\_features[0]].mean()

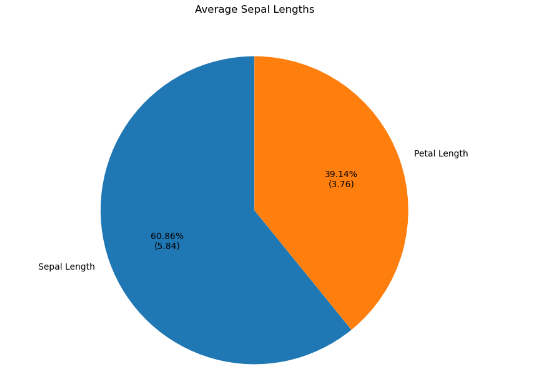
labels = ['Sepal Length', 'Petal Length']

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

plt.pie(sepal\_lengths, labels=labels, autopct=lambda p: '{:.2f}%\n({:.2f})'.format(p, p \* sum(sepal\_lengths)/100), startangle=90)

plt.title('Average Sepal Lengths')

plt.show()



1. Histogram for Sepal Length

plt.figure(figsize=(10, 6))

plt.hist(data[sepal\_features[0]], bins=50, label='Sepal Length')

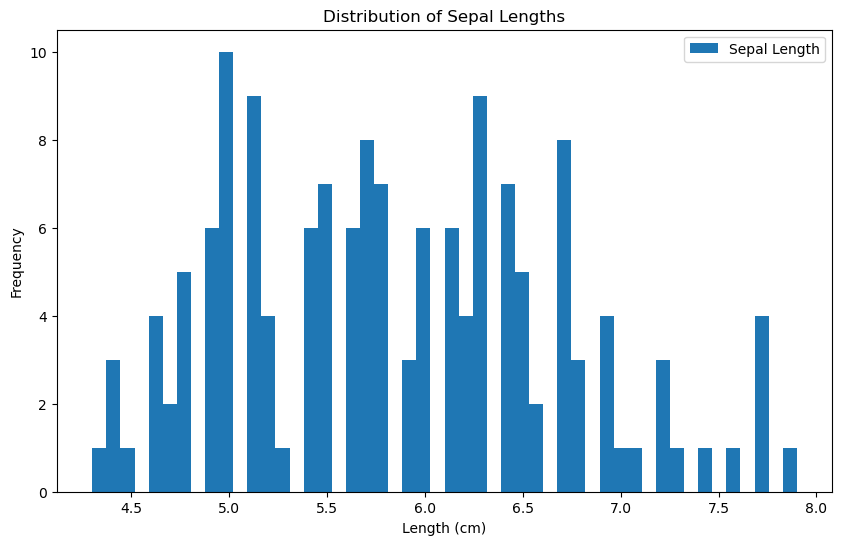
plt.title('Distribution of Sepal Lengths')

plt.xlabel('Length (cm)')

plt.ylabel('Frequency')

plt.legend()

plt.show()



1. Line graph for Petal lengths

plt.figure(figsize=(10, 6))

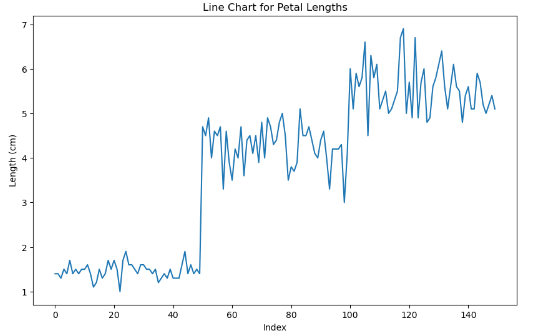
plt.plot(data[petal\_features[0]])

plt.title('Line Chart for Petal Lengths')

plt.xlabel('Index')

plt.ylabel('Length (cm)')

plt.show()



**Extract information from the chart.**

1. The pie chart above shows the average sepal lengths v/s average petal lengths for Iris. The blue circle represents the average sepal length for Iris, the orange circle represents the average petal length for Iris. Average Iris sepal length of 5.84 which is higher than the Iris petal length of 3.76.
2. The Histogram above represents the distribution of numerical data. It shows the frequency of different sepal length values for the Iris flower dataset.

The x-axis represents the sepal length in centimeters, and the y-axis represents the frequency of each sepal length value. The bars in the histogram show how many sepals have a particular length. For example, the bar at 5.5 cm is the tallest, meaning there are more sepals with a length of 5.5 cm than any other length.

The data visualization also shows a normal distribution curve plotted over the histogram. The normal distribution is a bell-shaped curve that represents the theoretical distribution of a continuous variable. The closer the data aligns with the normal distribution curve, the more likely the data is normally distributed. In this case, the sepal lengths are approximately normally distributed, with most of the sepals having lengths between 5 cm and 6.5 cm.

The average sepal length is about 5.8 cm. The sepal lengths range from about 4.3 cm to 8 cm. There is a small group of sepals that are much shorter or much longer than the average.

1. The line graph shows the average iris petal length over time. The y-axis shows the petal length in centimetres, and the x-axis shows the number of flowers with that petal length. The graph shows that the petal lengths of Iris are normally distributed, with a mean of about 5 cm and a standard deviation of about 1 cm. The range of petal lengths is from about 3 cm to about 7 cm. There are a few outliers, which are data points that fall outside of the normal distribution.