

ML (CSE(AIML) - 5th Semester)

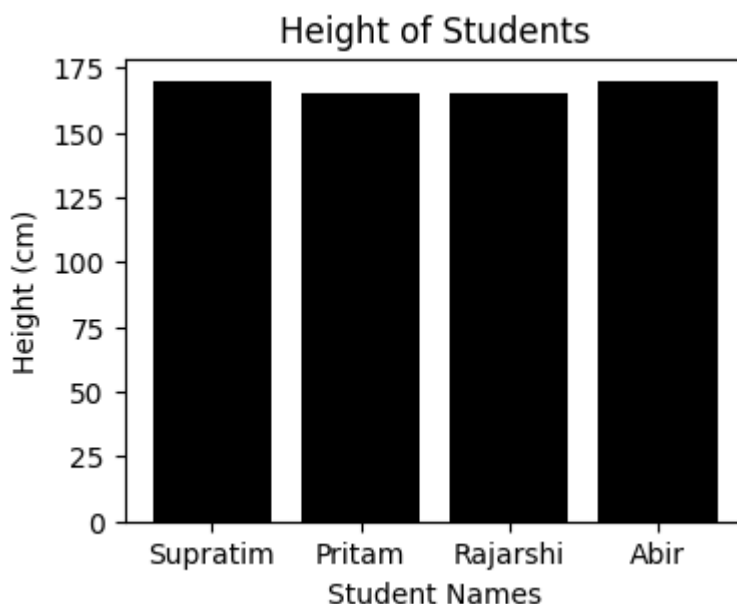
SUPRATIM NAG/AIML/22/057

Assignemnt 1(a)-- PART-1

Q-1: Take the names of four students in an array called X and place their height in an another array called Y. Plot a Bar chart where X axis represents the name of the student and Y axis represents the corresponding height.

```
In [ ]: import matplotlib.pyplot as plt
# Names of students
X = ['Supratim', 'Pritam', 'Rajarshi', 'Abir']
# Heights of students corresponding to the names
Y = [170, 165, 165, 170] # Heights in centimeters

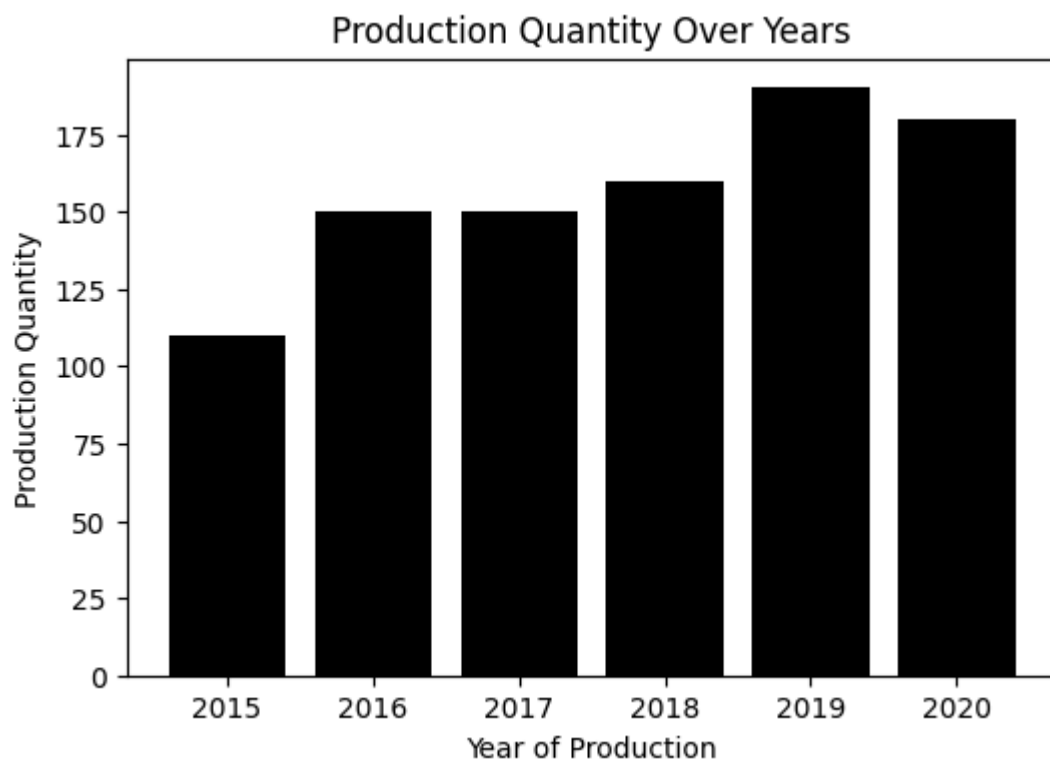
plt.figure(figsize=(4, 3))
plt.bar(X, Y, color='black')
plt.xlabel('Student Names')
plt.ylabel('Height (cm)')
plt.title('Height of Students')
# plt.grid(True)
plt.show()
```



Q-2: Take a dataset containing the information of Production Quantity and Year of Production in a data frame and plot the Bar Chart where X axis represents the Year of Production and Y axis represents the corresponding Production Quantity.

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt
# Sample dataset
data = {
    'Year of Production': [2015, 2016, 2017, 2018, 2019, 2020],
    'Production Quantity': [110, 150, 150, 160, 190, 180]
}
df = pd.DataFrame(data)

plt.figure(figsize=(6, 4))
plt.bar(df['Year of Production'], df['Production Quantity'], color='black')
plt.xlabel('Year of Production')
plt.ylabel('Production Quantity')
plt.title('Production Quantity Over Years')
plt.show()
```

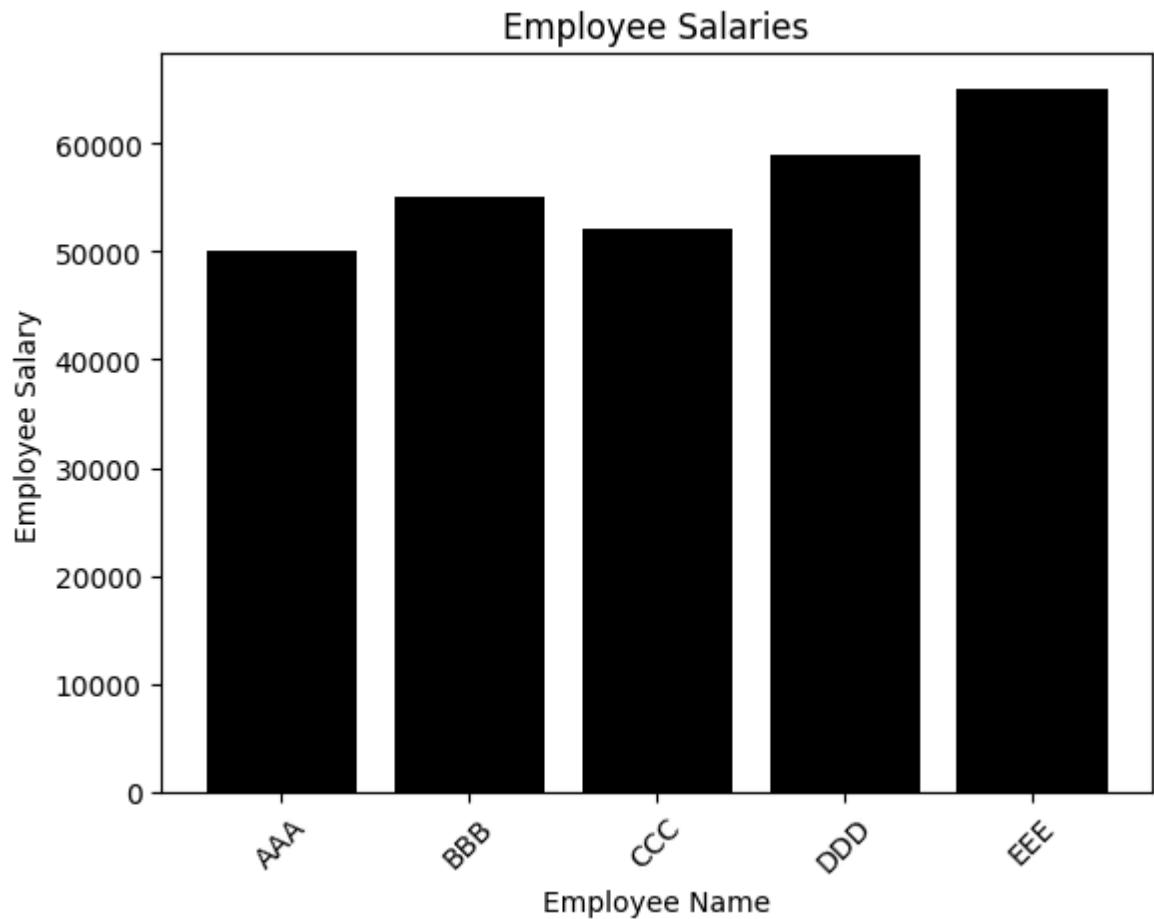


Q-3: Take a csv file with name, employee_id, and salary. Plot a Bar Chart with employee name on the X-axis and employee salary on the Y-axis.

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt

file_path='DSD Datasets/employee.csv'
df = pd.read_csv(file_path)

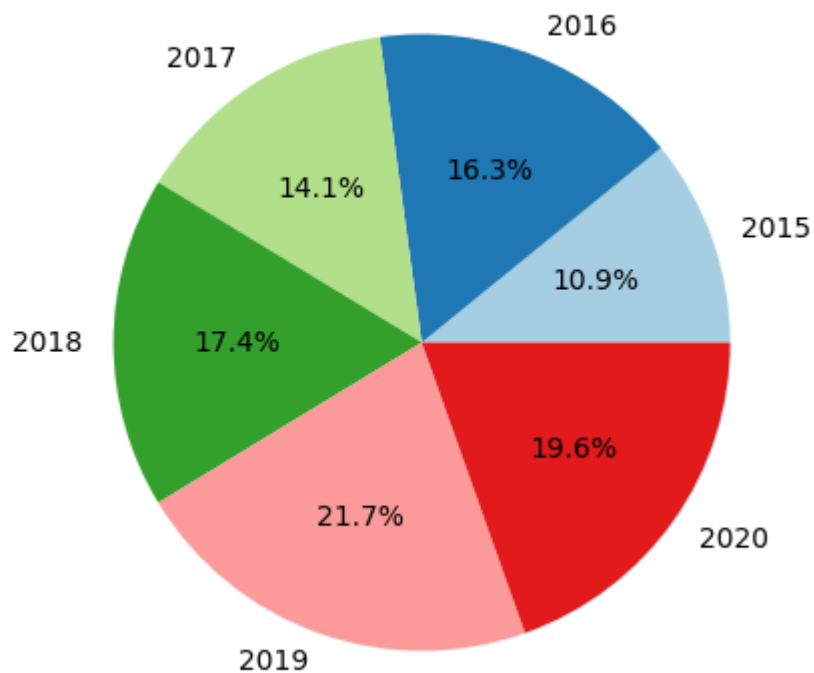
plt.bar(df['name'], df['salary'], color='black')
plt.xlabel('Employee Name')
plt.ylabel('Employee Salary')
plt.title('Employee Salaries')
plt.xticks(rotation=45)
plt.show()
```



Q-4: Consider the dataset created in Q-2 and plot the corresponding pie chart.

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt
# Sample dataset (same as Q-2)
data = {
    'Year of Production': [2015, 2016, 2017, 2018, 2019, 2020],
    'Production Quantity': [100, 150, 130, 160, 200, 180]
}
# Creating a DataFrame
df = pd.DataFrame(data)
# Plotting the pie chart
plt.pie(df['Production Quantity'], labels=df['Year of Production'], autopct='%1.1f%%')
plt.title('Production Quantity Distribution by Year')
plt.show()
```

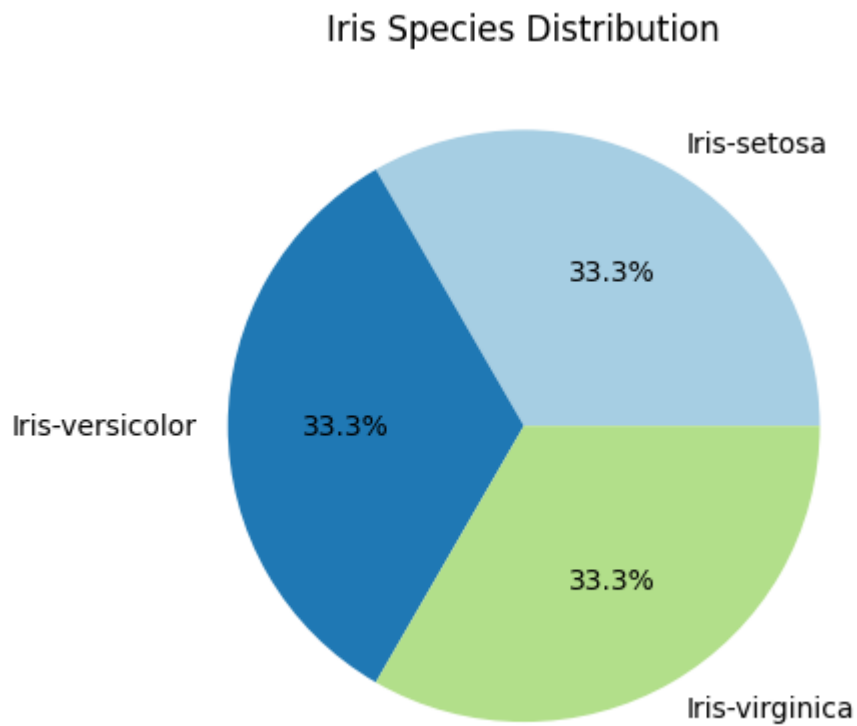
Production Quantity Distribution by Year



Q-5: Consider the iris dataset and plot a pie chart describing Iris Species Distribution.

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt

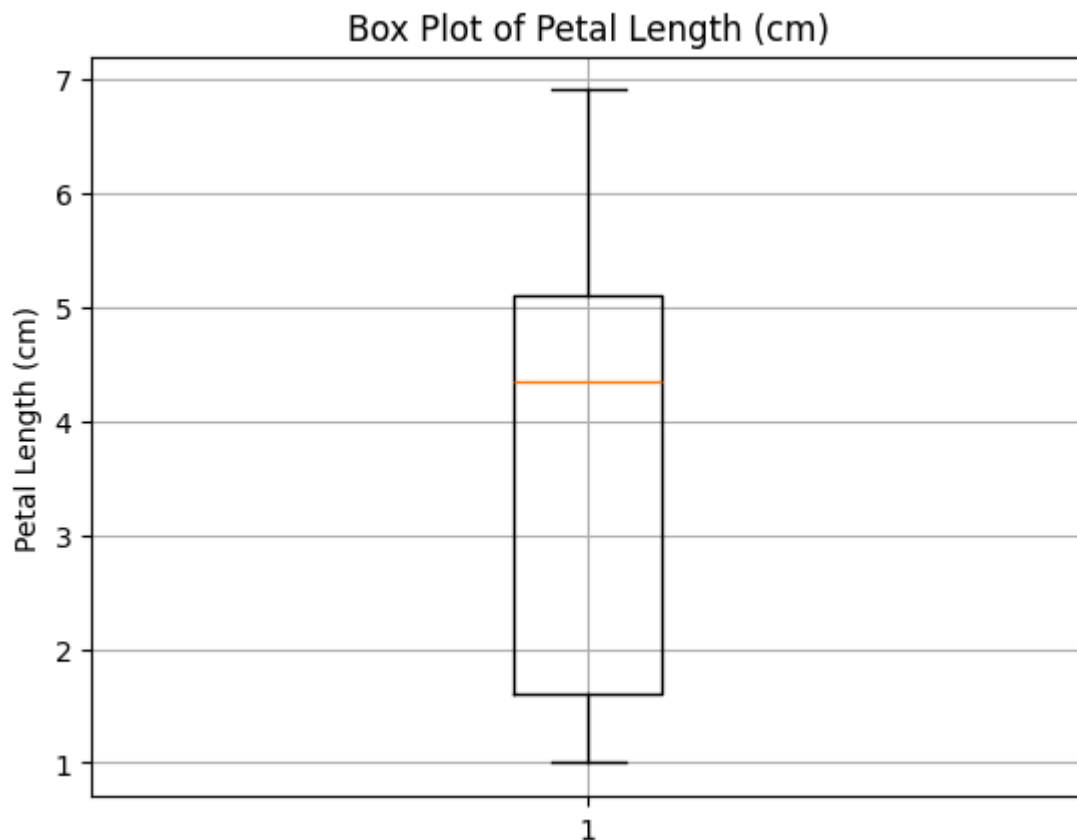
file_path='DSD Datasets/Iris.csv'
df = pd.read_csv(file_path)
# Count the number of instances for each species
species_counts = df['Species'].value_counts()
# Plotting the pie chart
plt.pie(species_counts, labels=species_counts.index, autopct='%1.1f%%', colors=p
plt.title('Iris Species Distribution')
plt.show()
```



Q-6: Take the iris dataset as input and plot the box-plot for the column " PetalLengthCm".

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt

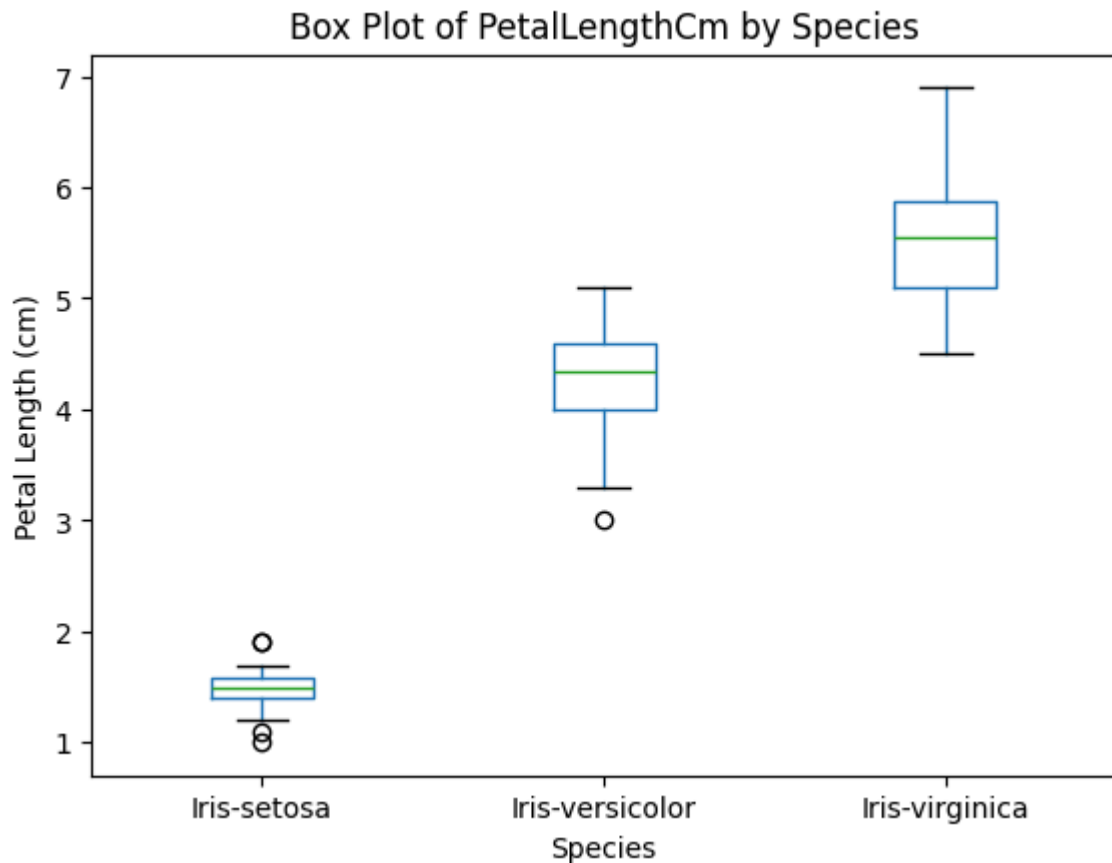
file_path='DSD Datasets/Iris.csv'
df = pd.read_csv(file_path)
# Plotting the box plot for the "PetalLengthCm" column
plt.boxplot(df['PetalLengthCm'])
plt.title('Box Plot of Petal Length (cm)')
plt.ylabel('Petal Length (cm)')
plt.grid(True)
plt.show()
```



Q-7: Take the iris dataset as input and plot the box-plot for the column " PetalLengthCm" grouped by Species.

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt

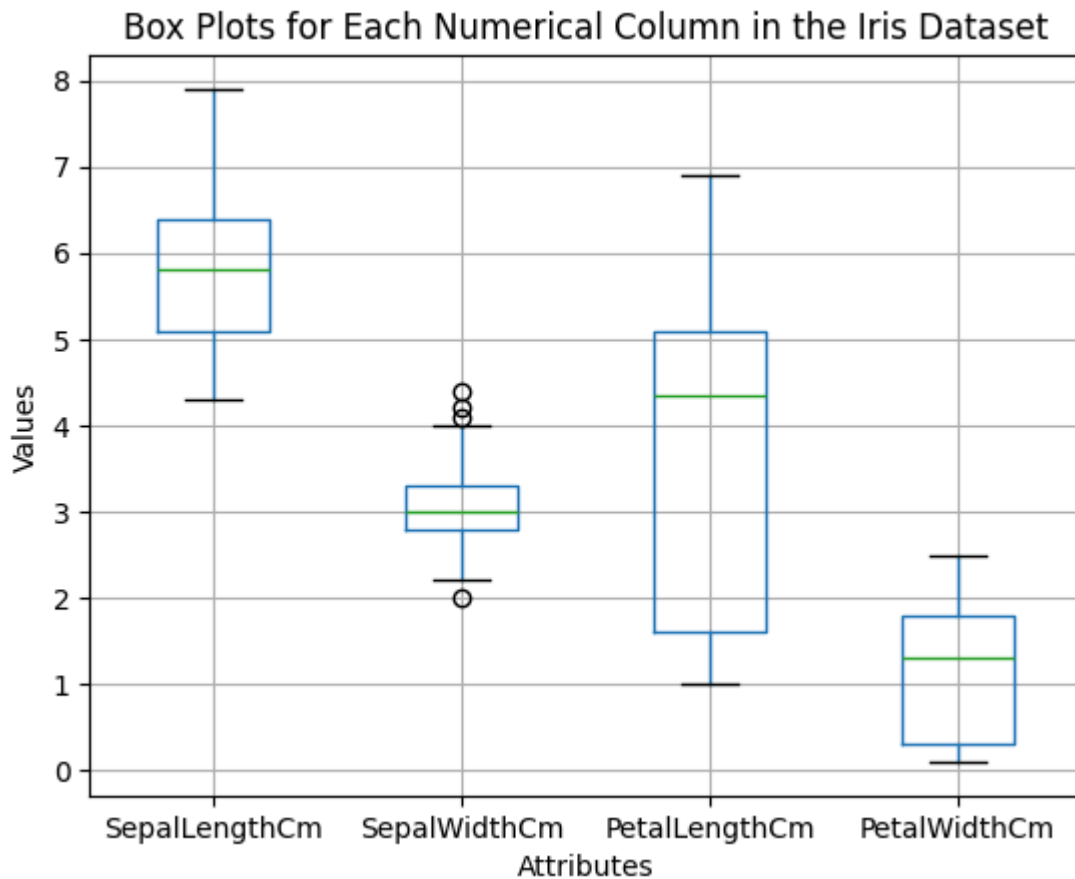
file_path='DSD Datasets/Iris.csv'
df = pd.read_csv(file_path)
# Plotting the box plot for the "PetalLengthCm" column grouped by "Species"
df.boxplot(column='PetalLengthCm', by='Species', grid=False)
plt.title('Box Plot of PetalLengthCm by Species')
plt.suptitle('') # Suppress the automatic title
plt.xlabel('Species')
plt.ylabel('Petal Length (cm)')
plt.show()
```



Q-8: Take the iris dataset as input and plot the multiple box-plots for each column.

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt

file_path='DSD Datasets/Iris.csv'
df = pd.read_csv(file_path)
# Selecting only the numerical columns (excluding 'Id' and 'Species')
numeric_columns = df.drop(columns=['Id', 'Species'])
# Plotting the box plots for each numerical column
numeric_columns.boxplot()
plt.title('Box Plots for Each Numerical Column in the Iris Dataset')
plt.xlabel('Attributes')
plt.ylabel('Values')
plt.show()
```



Q-9: Take the iris dataset as input and plot each column with Histogram.

```
In [ ]: import pandas as pd
import matplotlib.pyplot as plt

file_path='DSD Datasets/Iris.csv'
df = pd.read_csv(file_path)
# Selecting only the numerical columns (excluding 'Id' and 'Species')
numeric_columns = df.drop(columns=['Id', 'Species'])
# Plotting histograms for each numerical column
for i, column in enumerate(numeric_columns.columns, 1):
    plt.subplot(2, 2, i)
    plt.hist(numeric_columns[column], bins=20, edgecolor='black', alpha=0.7)
    plt.title(f'Histogram of {column}')
    plt.xlabel(column)
    plt.ylabel('Frequency')

plt.tight_layout()
plt.show()
```