Digital Empowerment Pakistan

Python Programming

Submitted By:

Farwa Rubab

Task 03:

Implement a data analysis project using pandas and matplotlib to explore and visualize a dataset of your choice.

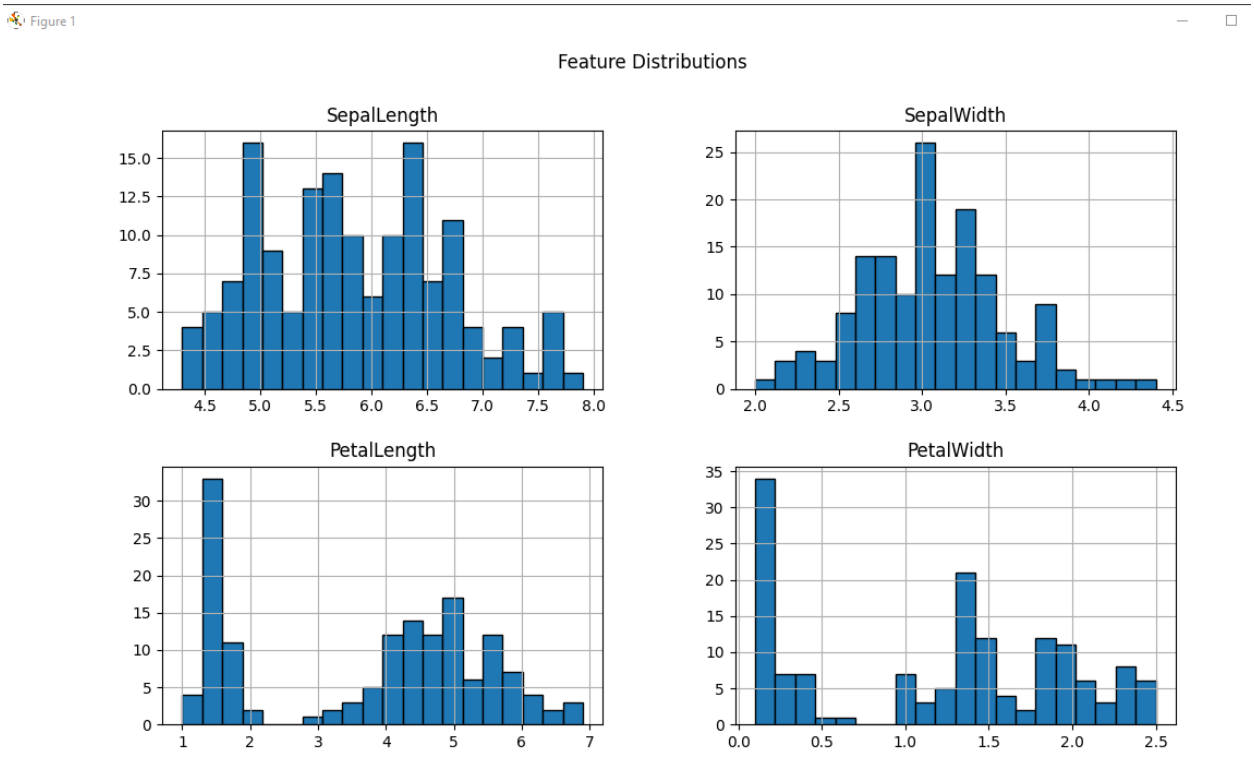
Following is the explanation of the task:

Code:

```
import pandas as pd
import matplotlib.pyplot as plt
from pandas.plotting import scatter_matrix

# Load the dataset
url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'
column_names = ['SepalLength', 'SepalWidth', 'PetalLength', 'PetalWidth',
'Species']
df = pd.read_csv(url, header=None, names=column_names)

# Display the first few rows of the dataframe
print(df.head())

# Check for missing values
print(df.isnull().sum())

# Display the data types
print(df.dtypes)

# Basic statistics
print(df.describe())

# Plot histograms of each feature
df.hist(bins=20, figsize=(12, 8), edgecolor='black')
plt.suptitle('Feature Distributions')
plt.show()

# Scatter plot matrix
scatter_matrix(df, figsize=(12, 8), diagonal='kde')
plt.suptitle('Scatter Matrix')
plt.show()

# Box plots for each feature grouped by species
df.boxplot(by='Species', figsize=(12, 8))
plt.suptitle('Box Plots by Species')
```

```
plt.show()

# Calculate mean values of each feature for each species
mean_values = df.groupby('Species').mean()

# Plot mean values
mean_values.plot(kind='bar', figsize=(12, 6))
plt.title('Mean Values of Features by Species')
plt.ylabel('Mean Value')
plt.show()
```

Output:



	SepalLength	SepalWidth	PetalLength	PetalWidth	Species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

SepalLength 0

SepalWidth 0

PetalLength 0

PetalWidth 0

Species 0

dtype: int64

SepalLength float64

SepalWidth float64

PetalLength float64

PetalWidth float64

Species object

dtype: object

	SepalLength	SepalWidth	PetalLength	PetalWidth
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000