Practical No-10

Date of Conduction : **Date of Checking:**

Data Visualization III

Download the Iris flower dataset or any other dataset into a DataFrame. (e.g., https://archive.ics.uci.edu/ml/datasets/Iris). Scan the dataset and give the inference as:

1. List down the features and their types (e.g., numeric, nominal) available in the dataset.
2. Create a histogram for each feature in the dataset to illustrate the feature distributions.
3. Create a box plot for each feature in the dataset.
4. Compare distributions and identify outliers.

Python Code:

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Download and load the Iris dataset
url = "https://archive.ics.uci.edu/ml/machine-learning-
databases/iris/iris.data"
column_names = ["sepal_length", "sepal_width", "petal_length",
"petal_width", "class"]
iris = pd.read_csv(url, header=None, names=column_names)

# 1. List down the features and their types
feature_types = iris.dtypes
print("Features and their types:")
print(feature_types)

# 2. Create a histogram for each feature
plt.figure(figsize=(12, 8))
iris.drop("class", axis=1).hist(edgecolor="black", linewidth=1.2, bins=20,
figsize=(12, 8))
plt.suptitle("Histograms for Each Feature", y=0.92)
plt.show()

# 3. Create a box plot for each feature
plt.figure(figsize=(12, 8))
sns.boxplot(data=iris.drop("class", axis=1), palette="Set2")
plt.title("Box Plots for Each Feature")
plt.show()

# 4. Compare distributions and identify outliers
plt.figure(figsize=(12, 8))
sns.boxplot(x="class", y="sepal_length", data=iris, hue="class",
palette="Set2", legend=False)
plt.title("Box Plot for Sepal Length by Class")
plt.show()
```

Explanation:

- The code uses the pandas library to load the Iris dataset from the given URL and names the columns.
- It then prints the types of features in the dataset (numeric).

- A histogram is created for each feature to illustrate their distributions.
- Box plots are generated for each feature to visually represent the distribution, median, and potential outliers.
- A box plot for the Sepal Length is created for each class to compare distributions between classes.

OUTPUT:

"C:\Users\Ram Kumar Solanki\PycharmProjects\pythonProject\venv\Scripts\python.exe"
"C:\Users\Ram Kumar Solanki\PycharmProjects\MBA_BFS\main.py"

Features and their types:

sepal_length float64

sepal_width float64

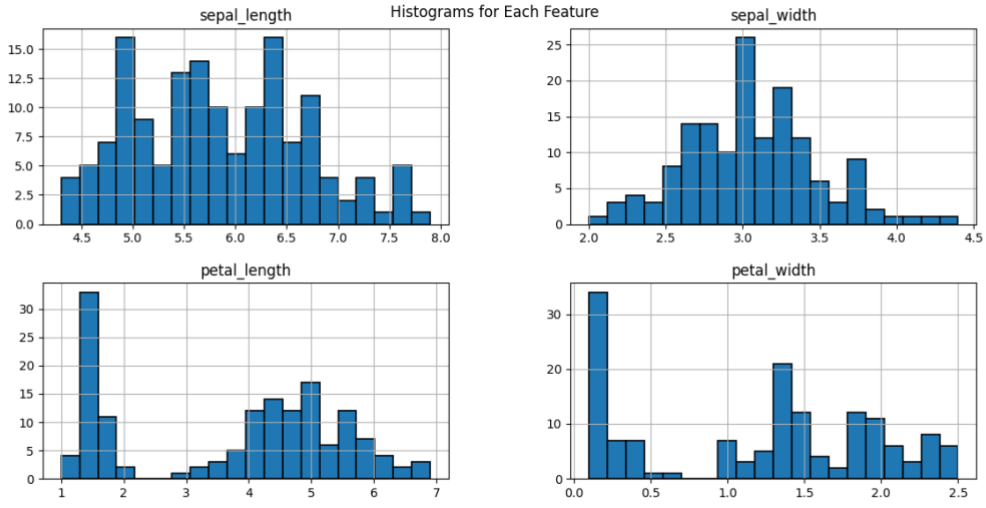
petal_length float64

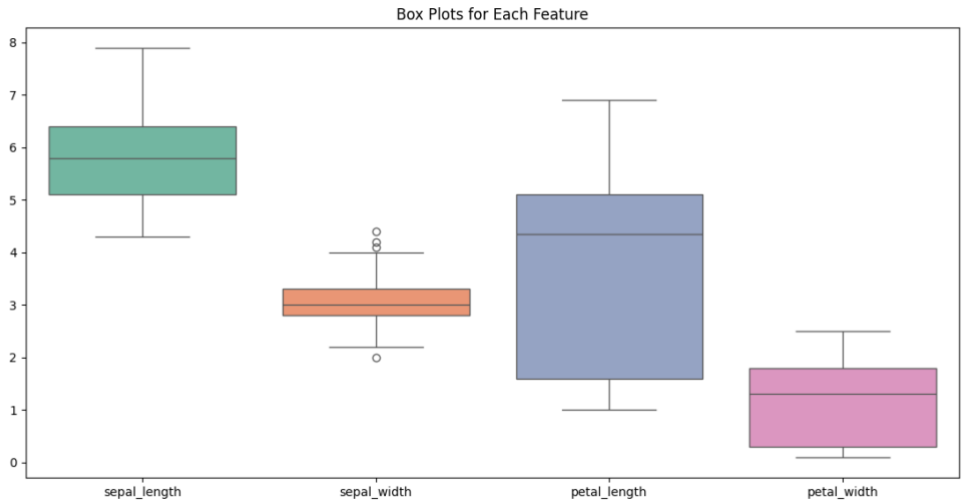
petal_width float64

class object

dtype: object

Process finished with exit code 0




Box Plots for Each Feature
8
7
6
5
4
3
2
1
0
sepal_length
sepal_width
petal_length
petal_width


Box Plot for Sepal Length by Class
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
sepal_length
Iris-setosa
Iris-versicolor
Iris-virginica
class