

AssignmentNo_A10

Assignment no.10

Title:Data Visualization III

Download the Iris flower dataset or any other dataset into a DataFrame. 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 boxplot for each feature in the dataset. 4. Compare distributions and identify outliers.

```
[ ]: import numpy as np
import pandas as pd
```

```
[ ]: df = pd.read_csv("Iris.csv")
df
```

```
[ ]:      Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  \
0      1          5.1          3.5          1.4          0.2
1      2          4.9          3.0          1.4          0.2
2      3          4.7          3.2          1.3          0.2
3      4          4.6          3.1          1.5          0.2
4      5          5.0          3.6          1.4          0.2
..  ...
145  146          6.7          3.0          5.2          2.3
146  147          6.3          2.5          5.0          1.9
147  148          6.5          3.0          5.2          2.0
148  149          6.2          3.4          5.4          2.3
149  150          5.9          3.0          5.1          1.8
```

```
      Species
0      Iris-setosa
1      Iris-setosa
2      Iris-setosa
3      Iris-setosa
4      Iris-setosa
..  ...
145  Iris-virginica
146  Iris-virginica
147  Iris-virginica
148  Iris-virginica
```

149 Iris-virginica

[150 rows x 6 columns]

```
[ ]: df.head()
```

```
[ ]:      Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  Species
0     1         5.1         3.5         1.4         0.2  Iris-setosa
1     2         4.9         3.0         1.4         0.2  Iris-setosa
2     3         4.7         3.2         1.3         0.2  Iris-setosa
3     4         4.6         3.1         1.5         0.2  Iris-setosa
4     5         5.0         3.6         1.4         0.2  Iris-setosa
```

```
[ ]: df.tail()
```

```
[ ]:      Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm  \
145  146         6.7         3.0         5.2         2.3
146  147         6.3         2.5         5.0         1.9
147  148         6.5         3.0         5.2         2.0
148  149         6.2         3.4         5.4         2.3
149  150         5.9         3.0         5.1         1.8
```

```
      Species
145  Iris-virginica
146  Iris-virginica
147  Iris-virginica
148  Iris-virginica
149  Iris-virginica
```

```
[ ]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Id              150 non-null   int64
1   SepalLengthCm   150 non-null   float64
2   SepalWidthCm    150 non-null   float64
3   PetalLengthCm   150 non-null   float64
4   PetalWidthCm    150 non-null   float64
5   Species         150 non-null   object
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
```

```
[ ]: df.describe()
```

```
[ ]:      Id  SepalLengthCm  SepalWidthCm  PetalLengthCm  PetalWidthCm
count  150.000000      150.000000      150.000000      150.000000      150.000000
mean    75.500000        5.843333        3.054000        3.758667        1.198667
std     43.445368        0.828066        0.433594        1.764420        0.763161
min      1.000000        4.300000        2.000000        1.000000        0.100000
25%     38.250000        5.100000        2.800000        1.600000        0.300000
50%     75.500000        5.800000        3.000000        4.350000        1.300000
75%    112.750000        6.400000        3.300000        5.100000        1.800000
max    150.000000        7.900000        4.400000        6.900000        2.500000
```

Mean

```
[ ]: print("Sepal length",df.SepalLengthCm.mean())
print("Sepal Width",df.SepalWidthCm.mean())
print("Petal length",df.PetalLengthCm.mean())
print("Petal Width",df.PetalWidthCm.mean())
```

```
Sepal length 5.8433333333333334
Sepal Width 3.0540000000000003
Petal length 3.7586666666666666
Petal Width 1.1986666666666668
```

Median

```
[ ]: print("Sepal length",df.SepalLengthCm.median())
print("Sepal Width",df.SepalWidthCm.median())
print("Petal length",df.PetalLengthCm.median())
print("Petal Width",df.PetalWidthCm.median())
```

```
Sepal length 5.8
Sepal Width 3.0
Petal length 4.35
Petal Width 1.3
```

Standard Deviation

```
[ ]: print("Sepal length",df.SepalLengthCm.std())
print("Sepal Width",df.SepalWidthCm.std())
print("Petal length",df.PetalLengthCm.std())
print("Petal Width",df.PetalWidthCm.std())
```

```
Sepal length 0.828066127977863
Sepal Width 0.4335943113621737
Petal length 1.7644204199522626
Petal Width 0.7631607417008411
```

Mode

```
[ ]: print("Sepal length",df.SepalLengthCm.mode())
print("Sepal Width",df.SepalWidthCm.mode())
```

```
print("Petal length",df.PetalLengthCm.mode())
print("Petal Width",df.PetalWidthCm.mode())
```

```
Sepal length 0    5.0
Name: SepalLengthCm, dtype: float64
Sepal Width 0    3.0
Name: SepalWidthCm, dtype: float64
Petal length 0    1.5
Name: PetalLengthCm, dtype: float64
Petal Width 0    0.2
Name: PetalWidthCm, dtype: float64
```

Max

```
[ ]: print("Sepal length",df.SepalLengthCm.max())
      print("Sepal Width",df.SepalWidthCm.max())
      print("Petal length",df.PetalLengthCm.max())
      print("Petal Width",df.PetalWidthCm.max())
```

```
Sepal length 7.9
Sepal Width 4.4
Petal length 6.9
Petal Width 2.5
```

Min

```
[ ]: print("Sepal length",df.SepalLengthCm.min())
      print("Sepal Width",df.SepalWidthCm.min())
      print("Petal length",df.PetalLengthCm.min())
      print("Petal Width",df.PetalWidthCm.min())
```

```
Sepal length 4.3
Sepal Width 2.0
Petal length 1.0
Petal Width 0.1
```

```
[ ]: column = len(list(df))
      column
      np.unique(df["Species"])
```

```
[ ]: array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'], dtype=object)
```

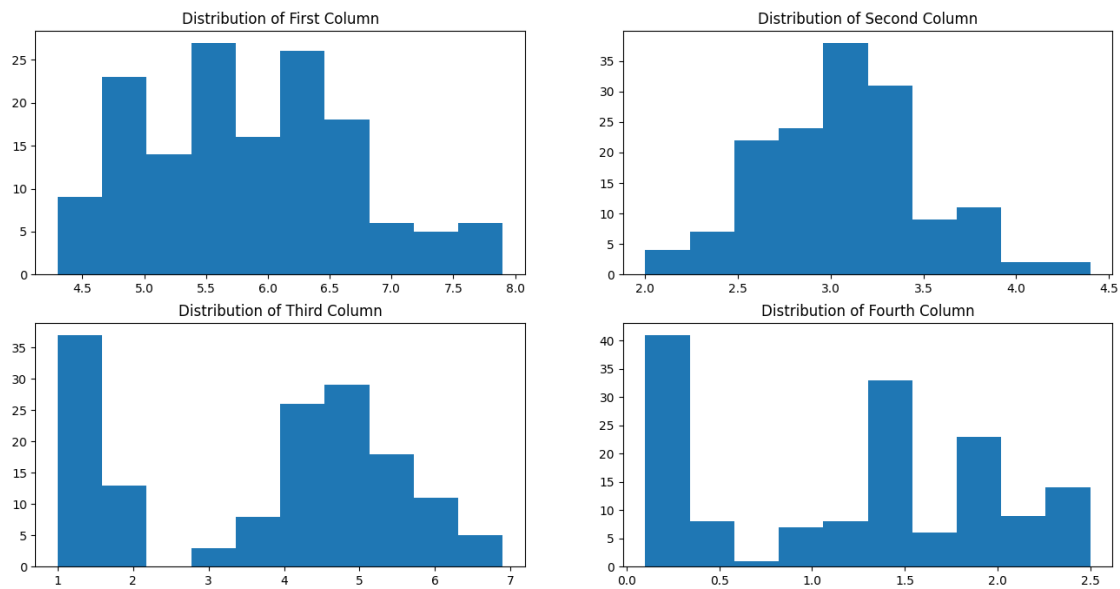
```
[ ]: import seaborn as sns
      import matplotlib
      import matplotlib.pyplot as plt

      %matplotlib inline
      fig, axes = plt.subplots(2, 2, figsize=(16, 8))
      axes[0,0].set_title("Distribution of First Column")
```

```

axes[0,0].hist(df["SepalLengthCm"]);
axes[0,1].set_title("Distribution of Second Column")
axes[0,1].hist(df["SepalWidthCm"]);
axes[1,0].set_title("Distribution of Third Column")
axes[1,0].hist(df["PetalLengthCm"]);
axes[1,1].set_title("Distribution of Fourth Column")
axes[1,1].hist(df["PetalWidthCm"]);

```



```

[ ]: data_to_plot = [
    ↪ [df["SepalLengthCm"], df["SepalWidthCm"], df["PetalLengthCm"], df["PetalWidthCm"]]
sns.set_style("whitegrid") # Creating a figure instance
fig = plt.figure(1, figsize=(12,8))
# Creating an axes instance
ax = fig.add_subplot(111) # Creating the boxplot
bp = ax.boxplot(data_to_plot);

```

