assg-on-numpy-pandas

January 24, 2024

1. Read the dataset to python environment.

[2]:
```
import pandas as pd
file_path = '/content/iris.xls'
iris_df = pd.read_excel(file_path)
iris_df.head()
```

[2]:

	SL	SW	PL	PW	Classification
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

2. Display the columns in the dataset.

[3]:
```
iris_columns = iris_df.columns
print("Columns in the Iris dataset:")
print(iris_columns)
```

```
Columns in the Iris dataset:
Index(['SL', 'SW', 'PL', 'PW', 'Classification'], dtype='object')
```

3. Calculate the mean of each column of the dataset.

[4]:
```
column_means = iris_df.mean()
print("Mean of each column :")
print(column_means)
```

```
Mean of each column :
SL    5.843333
SW    3.054000
PL    3.758667
PW    1.198667
dtype: float64

<ipython-input-4-779e1ee7e3ab>:1: FutureWarning: The default value of
numeric_only in DataFrame.mean is deprecated. In a future version, it will
default to False. In addition, specifying 'numeric_only=None' is deprecated.
Select only valid columns or specify the value of numeric_only to silence this
```

```
warning.
  column_means = iris_df.mean()
```

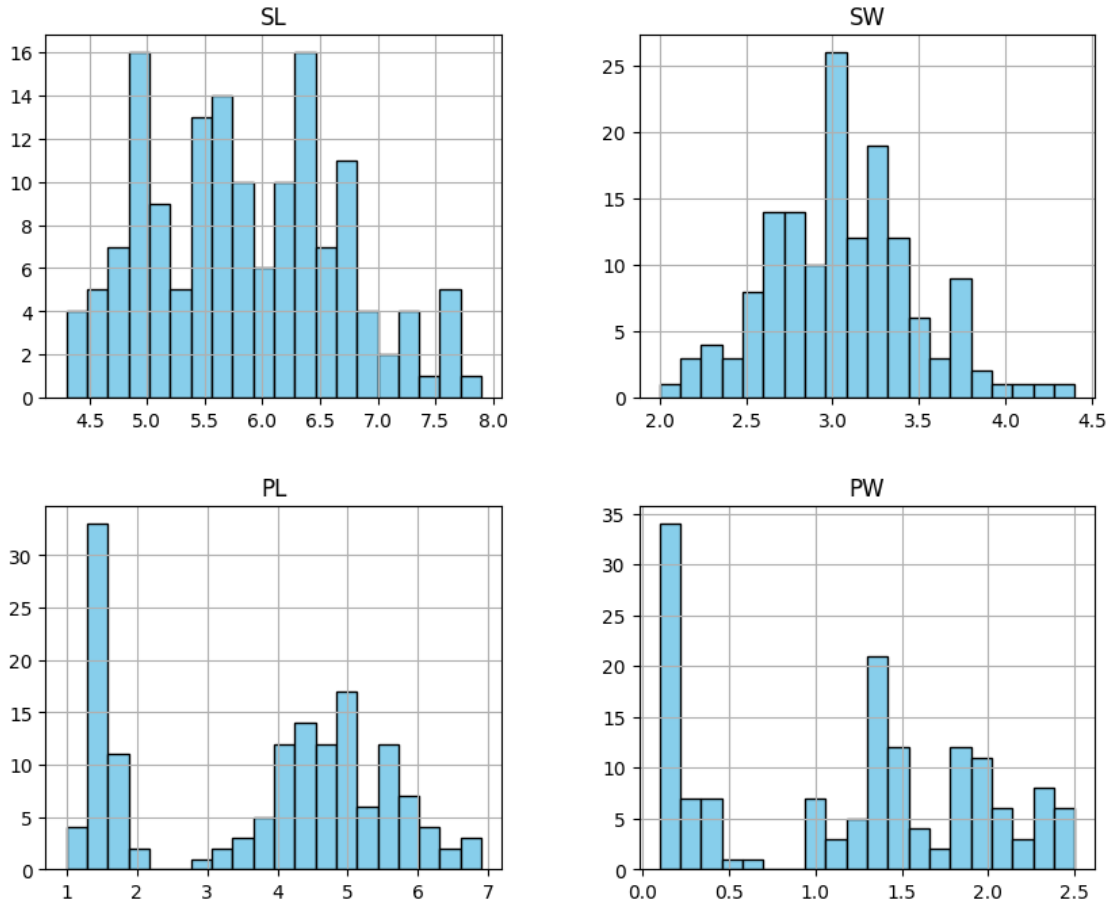
4. Check for the null values present in the dataset.

[5]:
```
null_values = iris_df.isnull().sum()
print("Null values in each column :")
print(null_values)
```

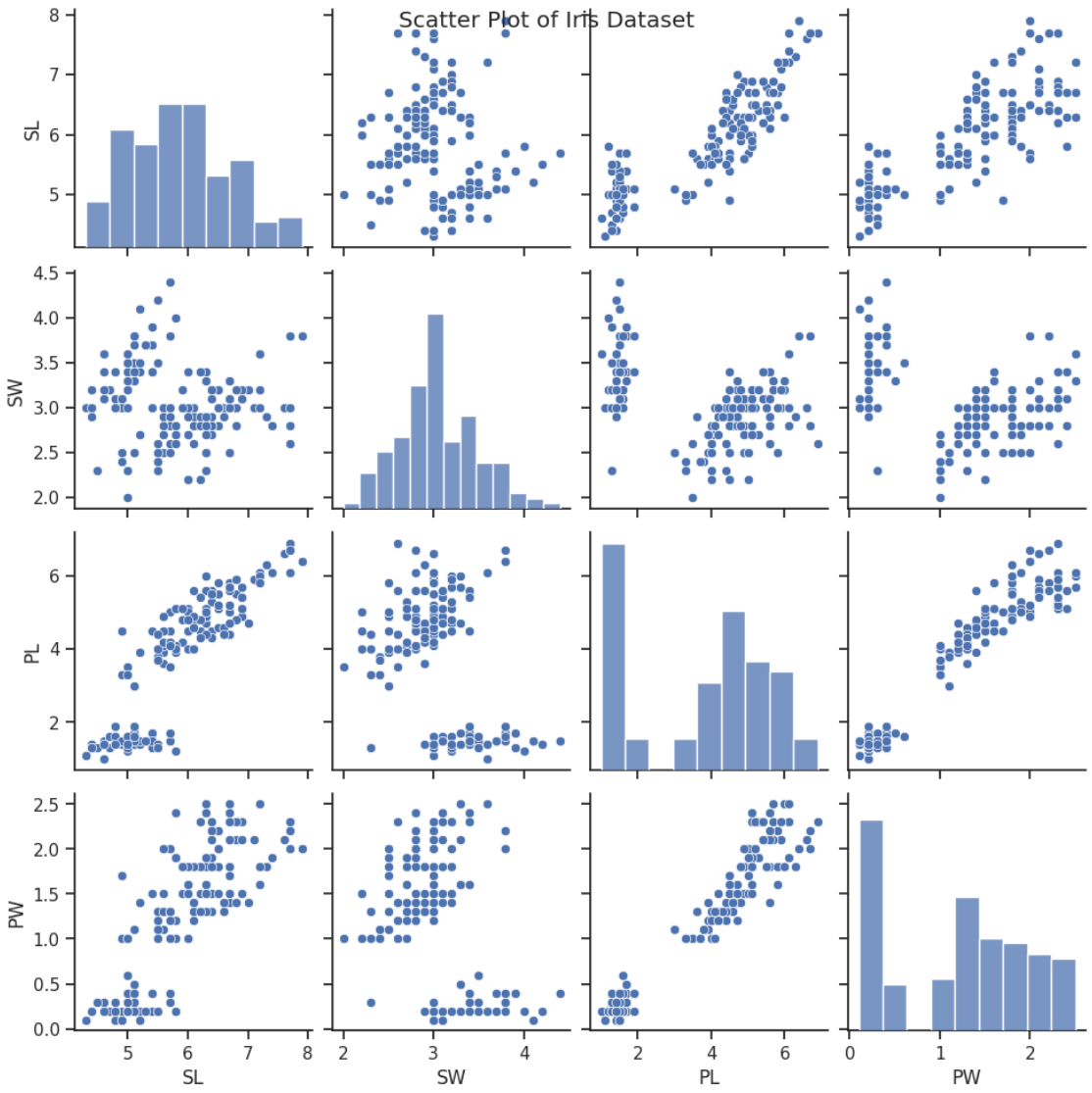
```
Null values in each column :
SL                0
SW                0
PL                0
PW                0
Classification    0
dtype: int64
```

5. Perform meaningful visualizations using the dataset. Bring at least 3 visualizations.

[7]:
```
import seaborn as sns
import matplotlib.pyplot as plt
iris_df.hist(figsize=(10, 8), bins=20, color='skyblue', edgecolor='black')
plt.suptitle('Histograms of Iris Dataset')
plt.show()
```

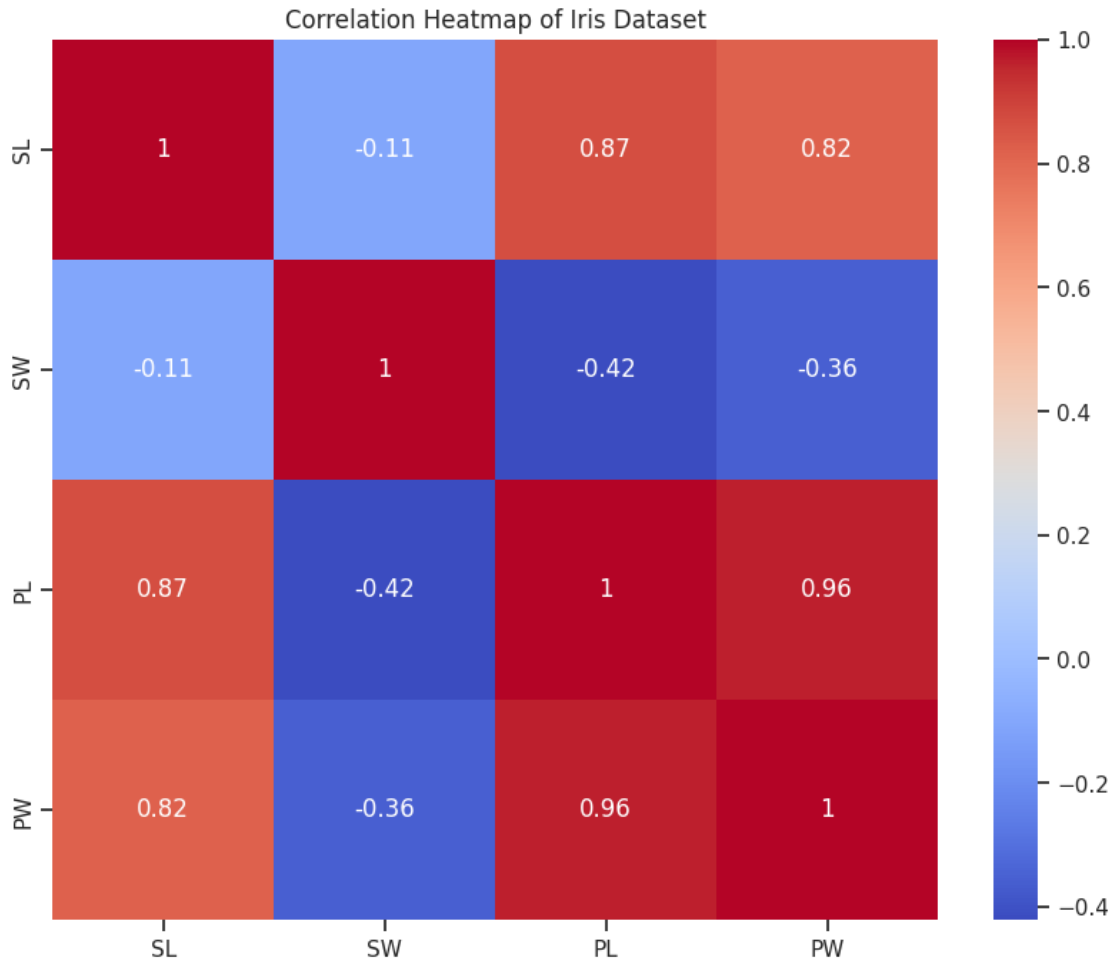


[12]:
```
sns.set_theme(style="ticks")
sns.pairplot(iris_df)
plt.suptitle('Scatter Plot of Iris Dataset')
plt.show()
```



[13]:
```
correlation_matrix = iris_df.corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap of Iris Dataset')
plt.show()
```

```
<ipython-input-13-aeeec30af451>:1: FutureWarning: The default value of
numeric_only in DataFrame.corr is deprecated. In a future version, it will
default to False. Select only valid columns or specify the value of numeric_only
to silence this warning.
  correlation_matrix = iris_df.corr()
```


Correlation Heatmap of Iris Dataset
SL
SW
PL
PW
1
-0.11
0.87
0.82
-0.11
1
-0.42
-0.36
0.87
-0.42
1
0.96
0.82
-0.36
0.96
1
SL
SW
PL
PW
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4