IMPORT PYTHON LIBRARIES

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

READ THE DATASET TO PYTHON ENVIRONMENT

```
iris = pd.read_excel('/content/iris.xls')
```

DISPLAY THE COLUMNS OF THE DATASET

```
iris.columns
```

```
Index(['SL', 'SW', 'PL', 'PW', 'Classification'], dtype='object')
```

CHECKING THE HEAD OF DATASET

```
iris.head()
```

```
    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
```

CALCULATE THE MEAN VALUES OF EACH COLUMNS OF THE DATASET

```
iris['SL'].mean()
```

```
5.843333333333334
```

```
iris['SW'].mean()
```

```
3.0540000000000003
```

```
iris['PL'].mean()
```

```
3.758666666666666
```

```
iris['PW'].mean()
```

```
1.1986666666666668
```

```
### A good picture of the distribution of data
```

```
iris.describe()
```

```
                SL          SW          PL          PW
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
```

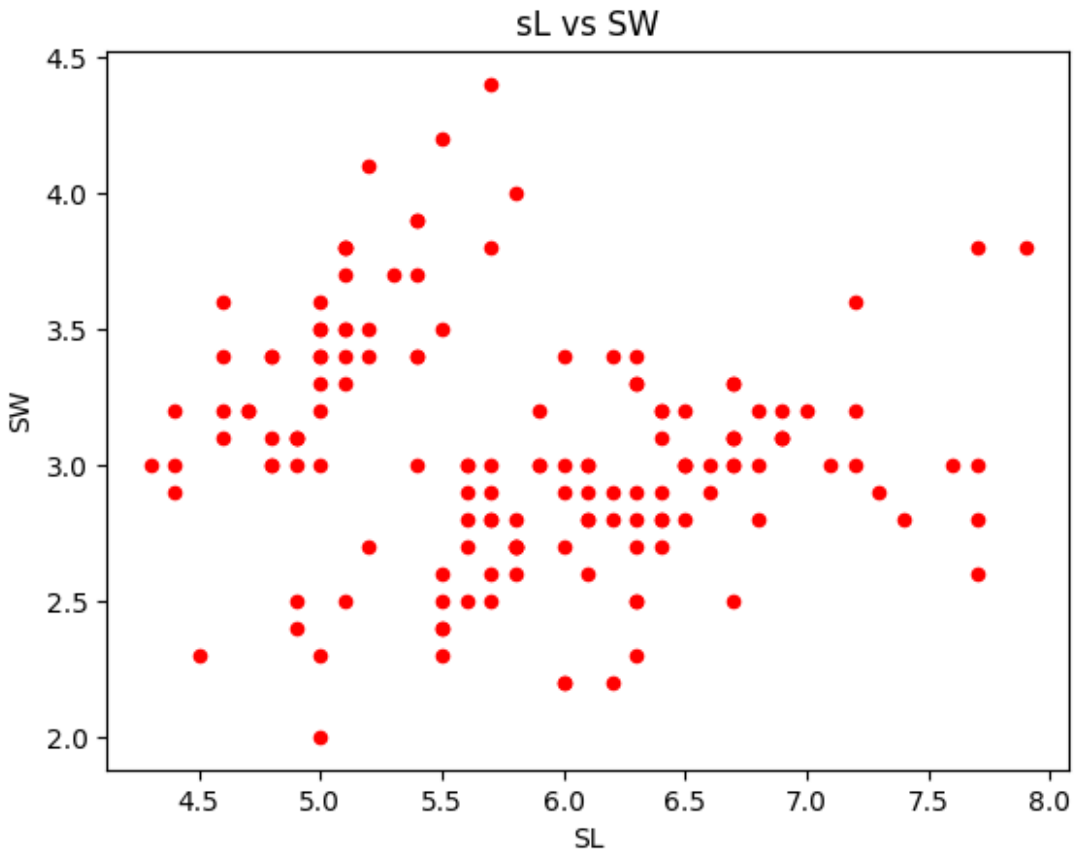
CHECKING THE NULL VALUES PRESENT IN THE DATASET

```
iris.isna()
```

```
        SL     SW     PL     PW  Classification
0    False  False  False  False           False
1    False  False  False  False           False
2    False  False  False  False           False
3    False  False  False  False           False
4    False  False  False  False           False
..     ...    ...    ...    ...             ...
145  False  False  False  False           False
146  False  False  False  False           False
147  False  False  False  False           False
148  False  False  False  False           False
149  False  False  False  False           False

[150 rows x 5 columns]
```

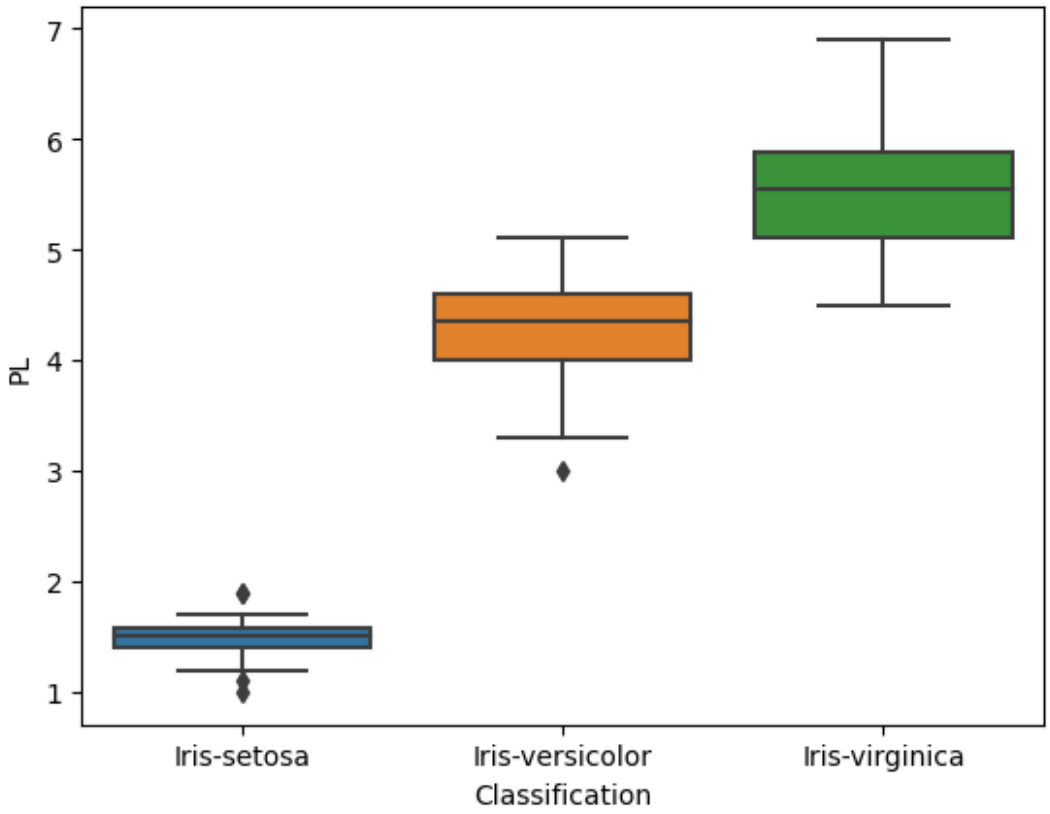
DATA VISUALIZATION

```
# scatter plot==>SL vs SW

iris.plot(kind="scatter",x="SL",y="SW",color='red')
plt.title('sL vs SW')

Text(0.5, 1.0, 'sL vs SW')
```



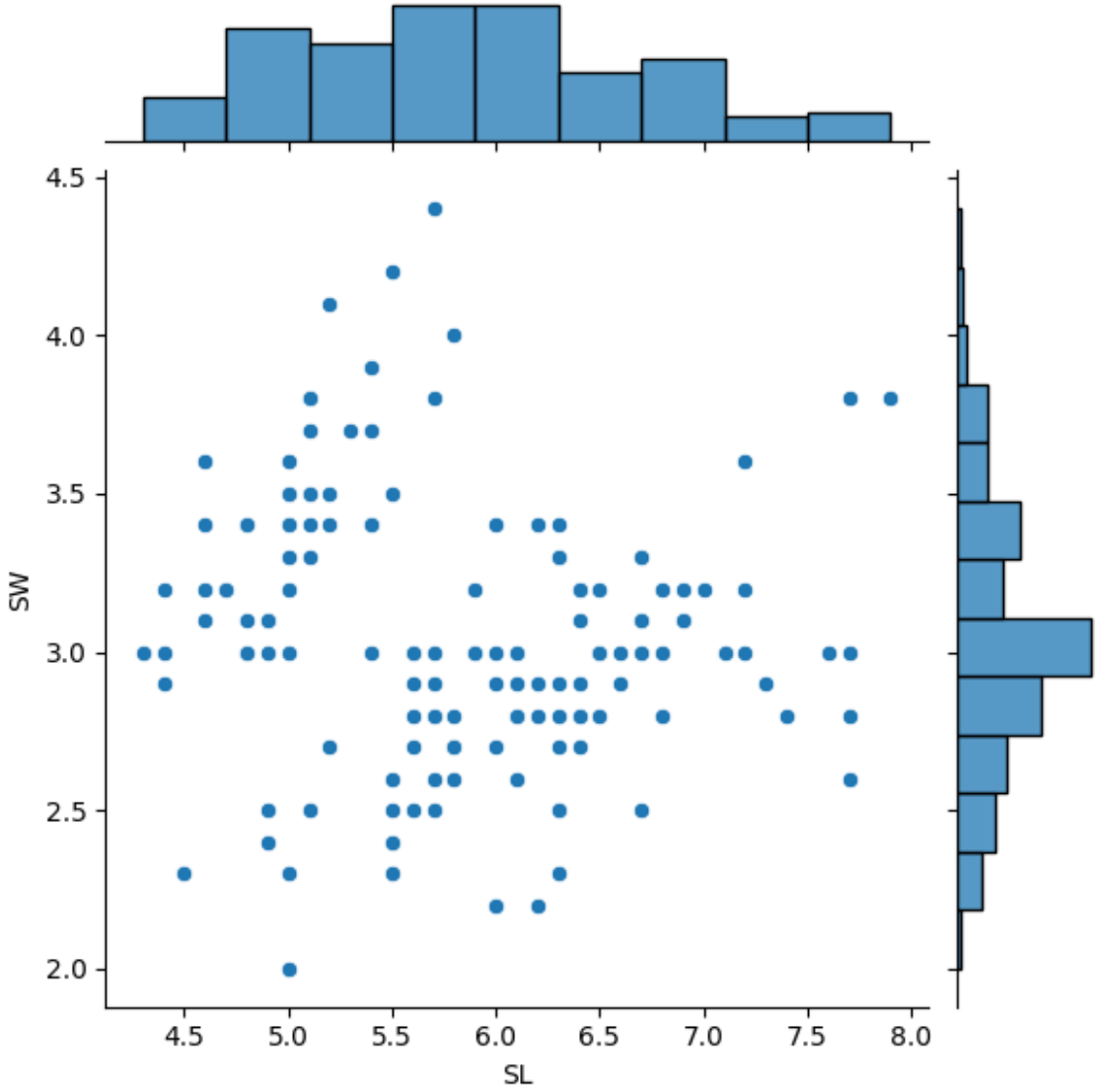

```
#boxplot ==>Classification vs PL

sns.boxplot(x='Classification',y='PL',data=iris)
```

```
<Axes: xlabel='Classification', ylabel='PL'>
```



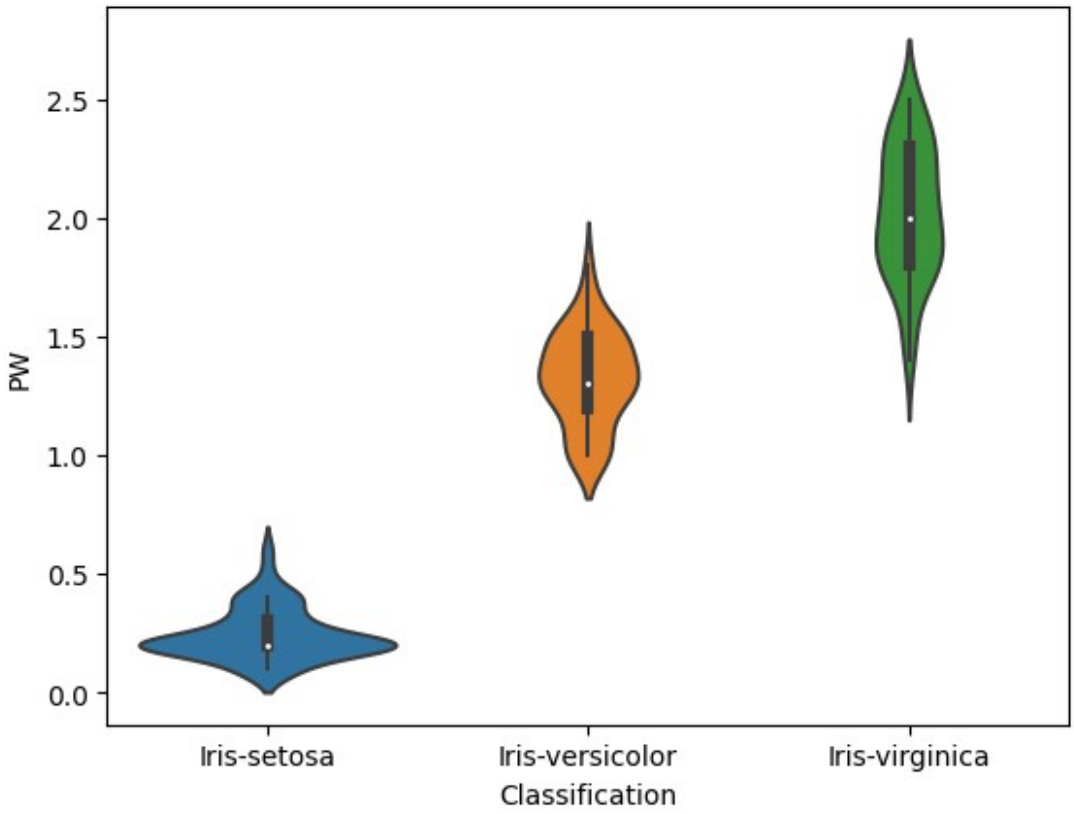

```
#joint plot of PL vs PW

sns.jointplot(x='SL',y='SW',data=iris)
```

```
<seaborn.axisgrid.JointGrid at 0x7899b11a28c0>
```




```
# violin plot

sns.violinplot(x='Classification',y='PW',data=iris,size=8)
```

```
<Axes: xlabel='Classification', ylabel='PW'>
```

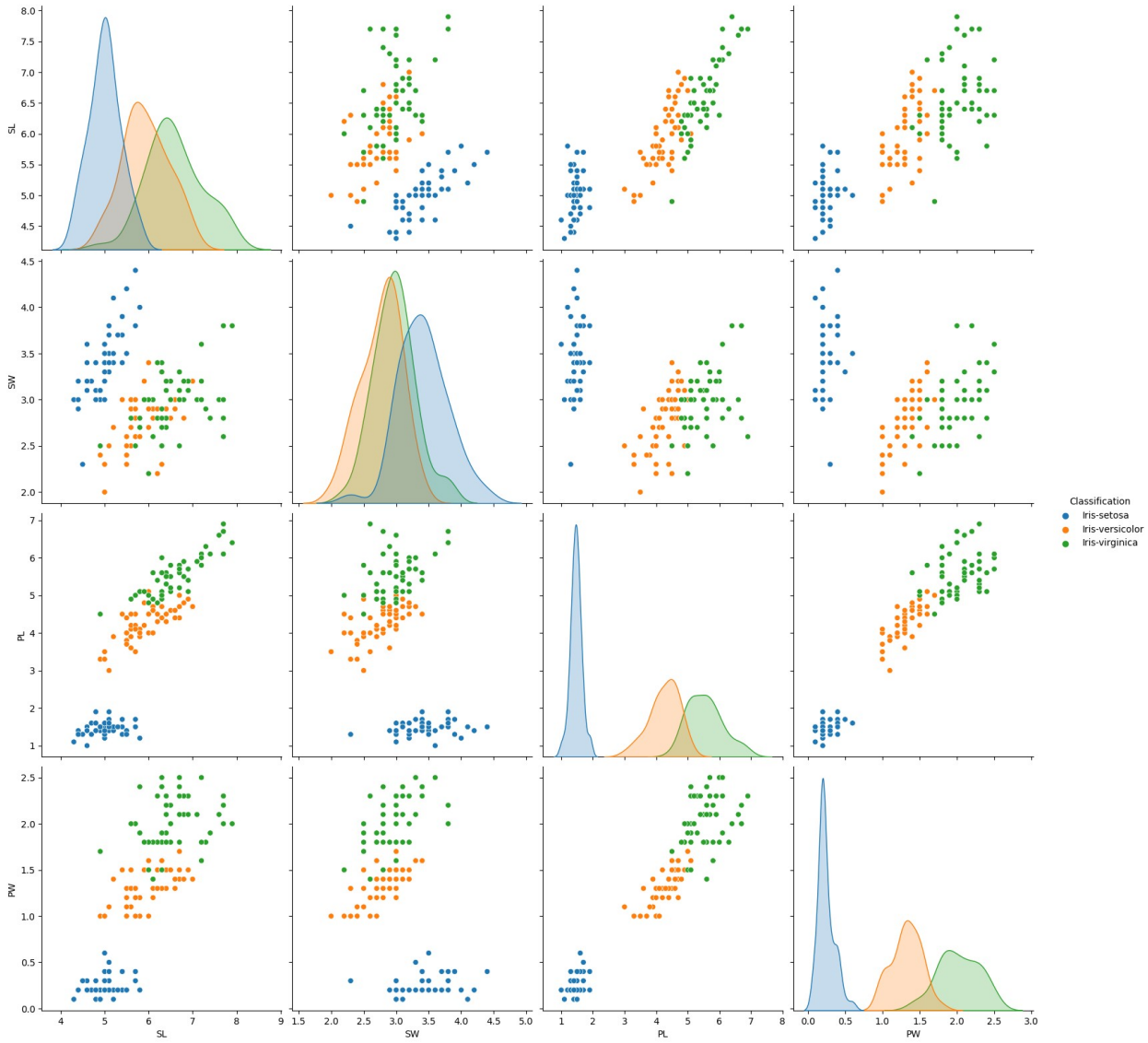



```
iris.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   SL              150 non-null    float64
 1   SW              150 non-null    float64
 2   PL              150 non-null    float64
 3   PW              150 non-null    float64
 4   Classification  150 non-null    object 
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

```
# pairplot ==> relationship between each pair of features

sns.pairplot(iris,hue='Classification',size =4)
```

```
/usr/local/lib/python3.10/dist-packages/seaborn/axisgrid.py:2095:
UserWarning: The `size` parameter has been renamed to `height`; please
update your code.
  warnings.warn(msg, UserWarning)

<seaborn.axisgrid.PairGrid at 0x7899abe98b20>
```


SL
SW
PL
PW
Classification
Iris-setosa
Iris-versicolor
Iris-virginica