Iris Flower Data

Library **Import**

```
import pandas as pd
# Reading CSV File
df = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/iris_csv.csv")
# printing 5 rows
df.head()    # bref overview of data frame.
```

	sepallength	sepalwidth	petallength	petalwidth	class
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

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

Getting Information from the dataset

```
df.shape    # dataset row, col
```

```
(150, 5)
```

It has 150 rows and 5 coloumns.

```
df.info()   # dataset information.
```

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

```
df.describe()
```

	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

Checking Missing Values

```
df.isnull().sum()
```

```
sepallength    0
sepalwidth     0
petallength    0
petalwidth     0
class          0
dtype: int64
```

Checking Duplicates

```
data = df.drop_duplicates(subset = "class",)
data
```

	sepallength	sepalwidth	petallength	petalwidth	class
0	5.1	3.5	1.4	0.2	Iris-setosa
50	7.0	3.2	4.7	1.4	Iris-versicolor
100	6.3	3.3	6.0	2.5	Iris-virginica

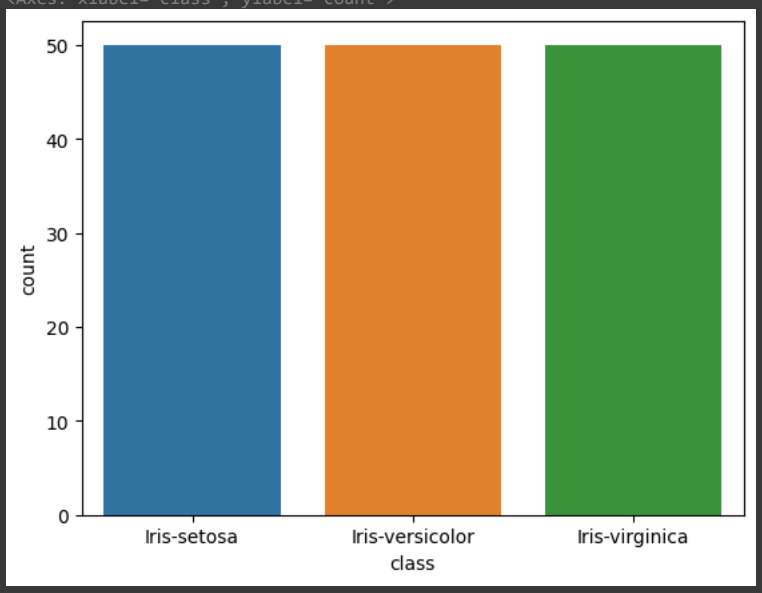
```
#how many value have in this class
df.value_counts("class")
# dataset same number of value is good.
```

```
class
Iris-setosa        50
Iris-versicolor    50
Iris-virginica     50
dtype: int64
```

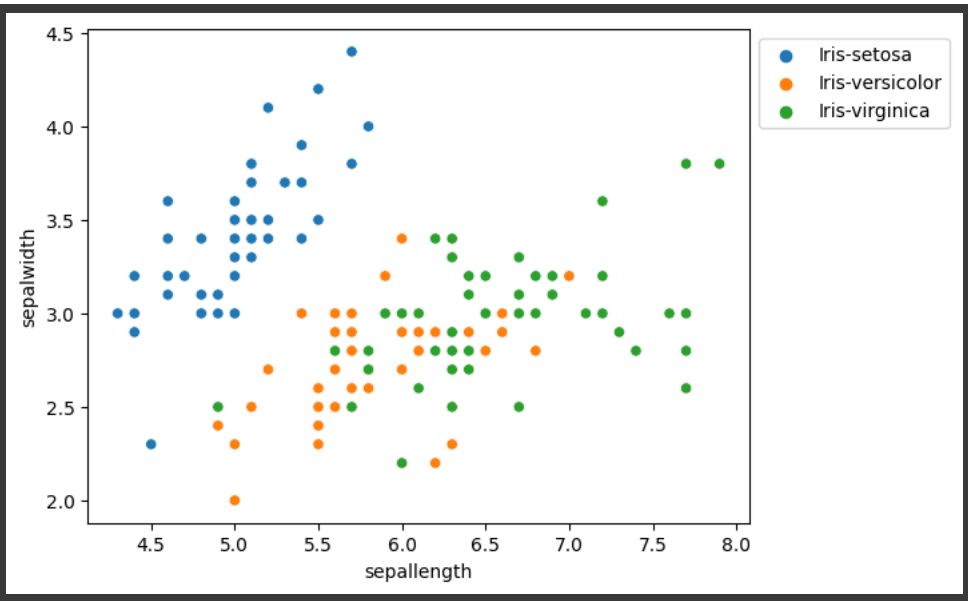
Data Vizualization

```
# importing package
import seaborn as sns     # seaborn used for data analysis.
import matplotlib.pyplot as plt
sns.countplot(x = 'class', data=df)

# ekan e class gulo x okhey. Ar class gulo balanced.
```

```
<Axes: xlabel='class', ylabel='count'>
```

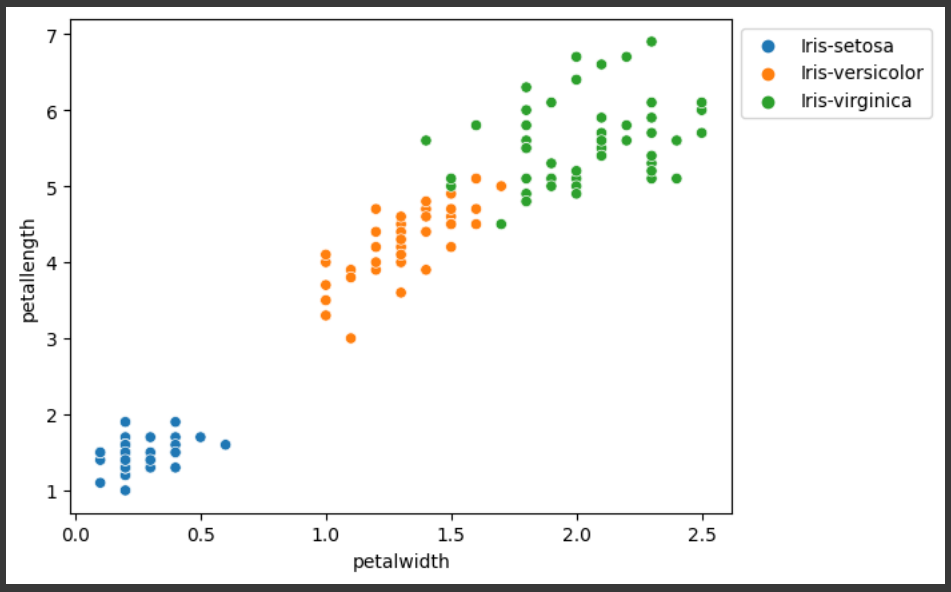



```
#importing packages
import seaborn as sns
import matplotlib.pyplot as plt

sns.scatterplot(x='sepallength', y='sepalwidth',hue='class', data=df, )
#placing legend outside the figure
plt.legend(bbox_to_anchor=(1, 1), loc=2)
plt.show()
```



Double-click (or enter) to edit

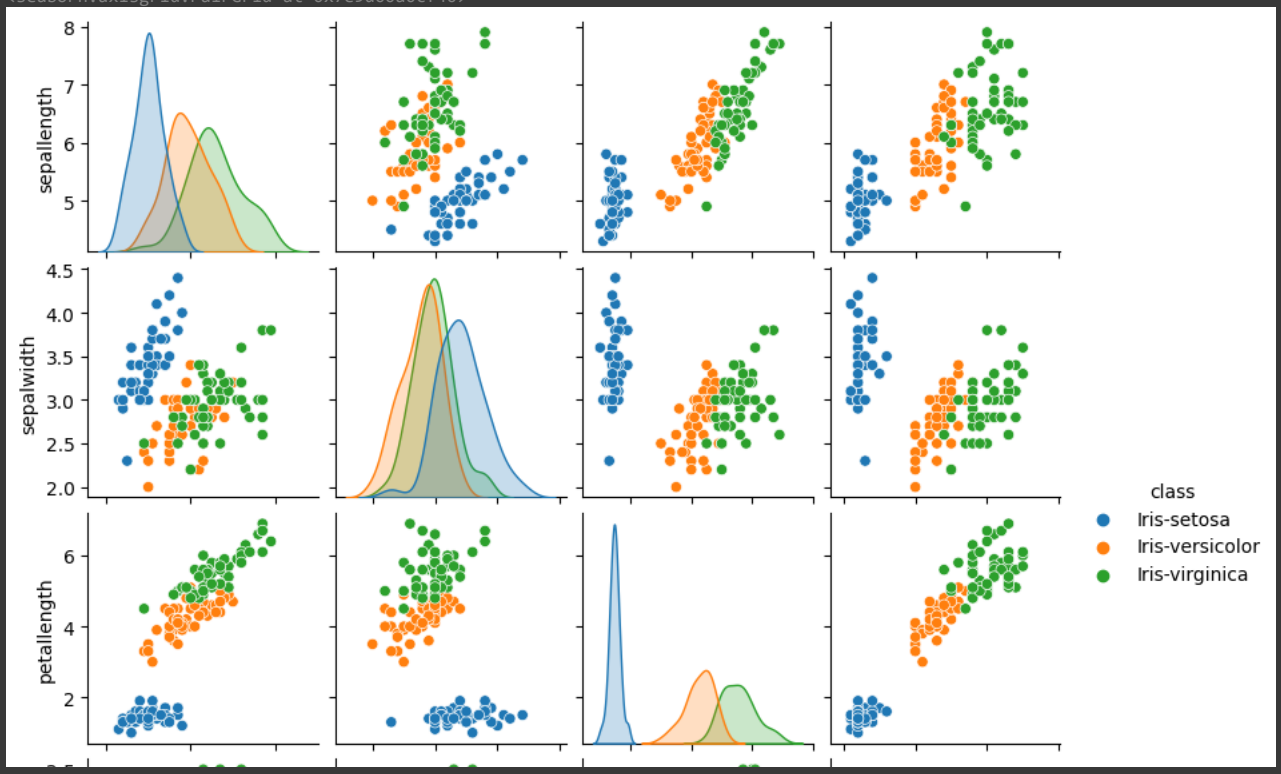
```
#importing packages
import seaborn as sns
import matplotlib.pyplot as plt

sns.scatterplot(x='petalwidth', y='petallength',hue='class', data=df, )
#placing legend outside the figure
plt.legend(bbox_to_anchor=(1, 1), loc=2)
plt.show()
```



setosa have short petalwidth and versicolor and virginica is large petal width.

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

sns.pairplot(df.drop([], axis=1), hue='class', height=2)
```

```
<seaborn.axisgrid.PairGrid at 0x7e9d00a0cf40>
```

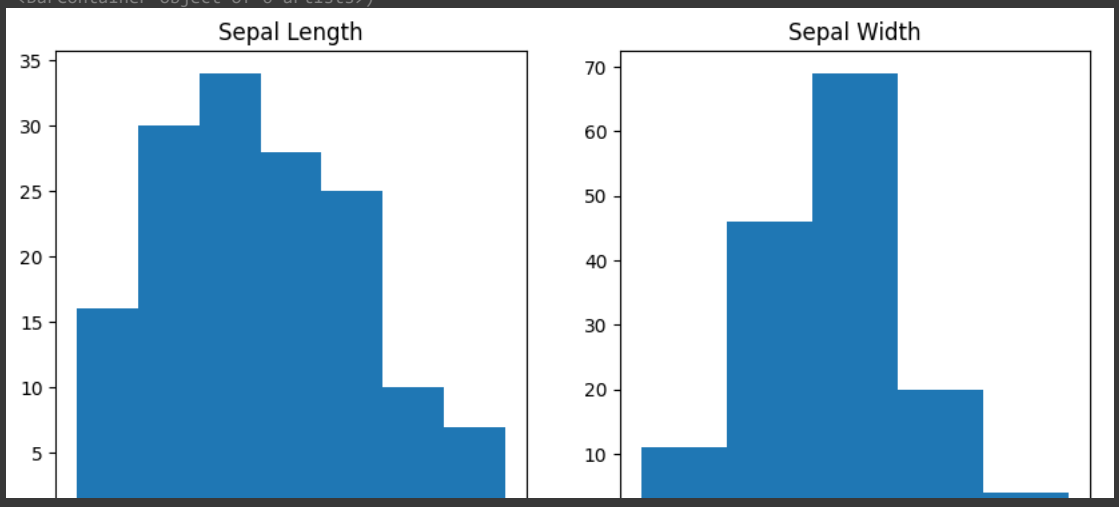


Histograms

axis = 1 (is column), axis = 0 (is row).

```
import seaborn as s
import matplotlib.pyplot as plt

fig, axes = plt.subplots(2, 2, figsize = (10, 10))

axes[0, 0].set_title("Sepal Length")
axes[0, 0].hist(df['sepallength'], bins=7)

axes[0, 1].set_title("Sepal Width")
axes[0, 1].hist(df['sepalwidth'], bins=5)

axes[1, 0].set_title("Petal Length")
axes[1, 0].hist(df['petallength'], bins=6)

axes[1, 1].set_title("Petal Width")
axes[1, 1].hist(df['petalwidth'], bins=6)

```

```
(array([48.,  2., 15., 37., 25., 23.]),
 array([0.1, 0.5, 0.9, 1.3, 1.7, 2.1, 2.5]),
 <BarContainer object of 6 artists>)
```

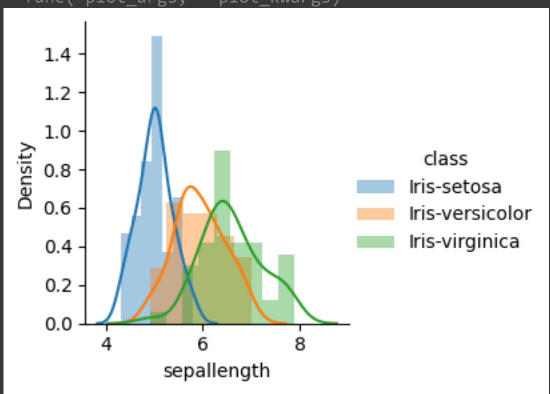


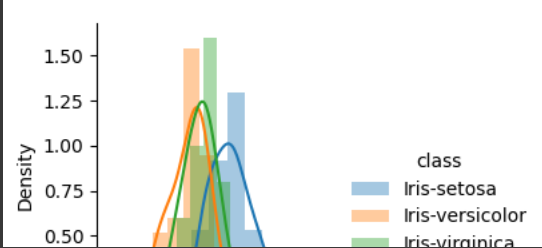
Histograms with Distplot Plot


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

plot = sns.FacetGrid(df, hue="class")
plot.map(sns.distplot, "sepallength").add_legend()

plot = sns.FacetGrid(df, hue="class")
plot.map(sns.distplot, "sepalwidth").add_legend()

plot = sns.FacetGrid(df, hue="class")
plot.map(sns.distplot, "petallength").add_legend()

plot = sns.FacetGrid(df, hue="class")
plot.map(sns.distplot, "petalwidth").add_legend()

plt.show()
```

```
/usr/local/lib/python3.10/dist-packages/seaborn/axisgrid.py:848: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  func(*plot_args, **plot_kwargs)
/usr/local/lib/python3.10/dist-packages/seaborn/axisgrid.py:848: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  func(*plot_args, **plot_kwargs)
/usr/local/lib/python3.10/dist-packages/seaborn/axisgrid.py:848: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  func(*plot_args, **plot_kwargs)
/usr/local/lib/python3.10/dist-packages/seaborn/axisgrid.py:848: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  func(*plot_args, **plot_kwargs)
/usr/local/lib/python3.10/dist-packages/seaborn/axisgrid.py:848: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  func(*plot_args, **plot_kwargs)
/usr/local/lib/python3.10/dist-packages/seaborn/axisgrid.py:848: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  func(*plot_args, **plot_kwargs)
/usr/local/lib/python3.10/dist-packages/seaborn/axisgrid.py:848: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  func(*plot_args, **plot_kwargs)
/usr/local/lib/python3.10/dist-packages/seaborn/axisgrid.py:848: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  func(*plot_args, **plot_kwargs)
/usr/local/lib/python3.10/dist-packages/seaborn/axisgrid.py:848: UserWarning:
```

```
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  func(*plot_args, **plot_kwargs)
/usr/local/lib/python3.10/dist-packages/seaborn/axisgrid.py:848: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  func(*plot_args, **plot_kwargs)
/usr/local/lib/python3.10/dist-packages/seaborn/axisgrid.py:848: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  func(*plot_args, **plot_kwargs)
/usr/local/lib/python3.10/dist-packages/seaborn/axisgrid.py:848: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751

  func(*plot_args, **plot_kwargs)
```





Handling Correlation

```
data.corr(method='pearson')

# proportional (+) and inverse (-) relationship
```

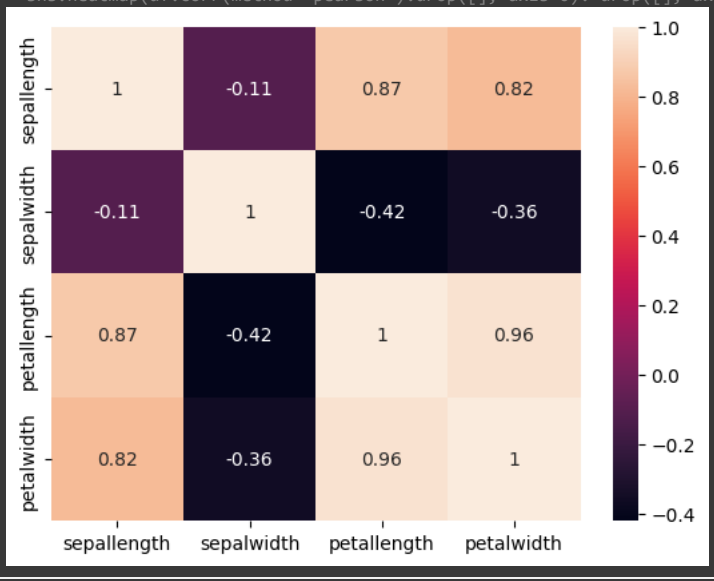
```
<ipython-input-17-2036ded8b4f6>:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future ver
```

Heatmaps

```
sepallength  1.000000  -0.999220  0.795795  0.645817
```

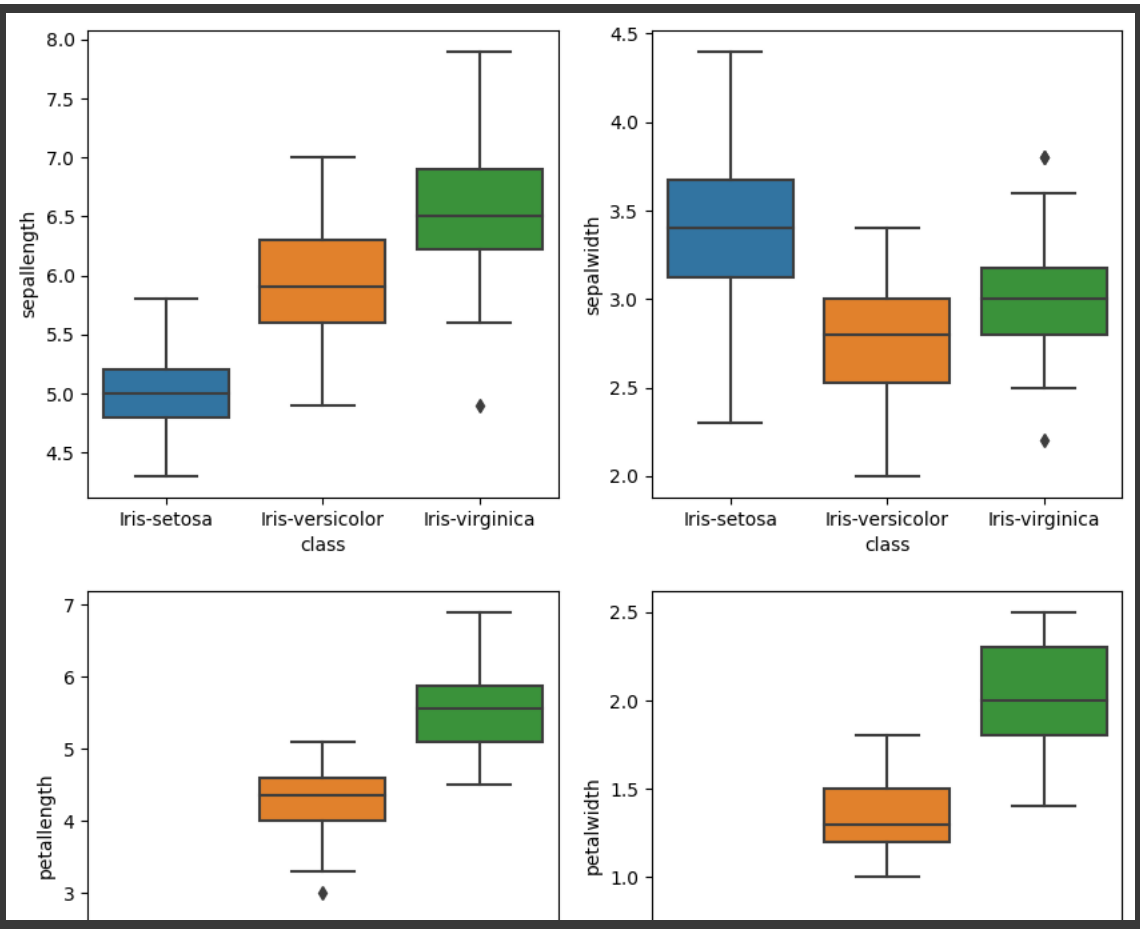
```
import seaborn as sus
import matplotlib.pyplot as pit

sns.heatmap(df.corr(method='pearson').drop([], axis=0). drop([], axis=0), annot=True);
pit.show()

# proportional (+) and inverse (-) relationship
```

```
<ipython-input-18-d2b48a1edcc0>:4: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future ver
  sns.heatmap(df.corr(method='pearson').drop([], axis=0). drop([], axis=0), annot=True);
```



Box Plots

```
# importing packages
import seaborn as sns
import matplotlib.pyplot as plt

def graph(y):
    sns.boxplot(x="class", y=y, data=df)

plt.figure(figsize=(10, 10))

# adding the subplot at the specified
# grid position
plt.subplot(221)
graph('sepallength')

plt.subplot(222)
graph('sepalwidth')

plt.subplot(223)
graph('petallength')

plt.subplot(224)
graph('petalwidth')

plt.show()

```



Handling Outliers

```
1 # importing packages
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4
5 sns.boxplot(x='sepalwidth', data=df)
```

```
<Axes: xlabel='sepalwidth'>
```