

Ratu Aubrey Khairani  
24060120130124  
B1

## ▼ Import library pandas dari python

```
import pandas
```

## ▼ Memuat Datasets

memuat datasets iris dengan memasukan link dataset yang diambil dari repository, kemudian di paste pada variable url yang telah dibuat

```
url = "http://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"  
names = ['sepal-length', 'sepal-width', 'petal-length', 'petal-width', 'class']  
dataset = pandas.read_csv(url, names=names)
```

variable dataset berfungsi untuk memasukkan library dan datasets yang telah diimport

## ▼ Menentukan Dimensi Dataset

disini saya menentukan dimensi dataset dengan menggunakan dataset.shape

```
print(dataset.shape)  
  
(150, 5)
```

## ▼ Melihat Isi dari Dataset

Kode dibawah merupakan kode untuk melihat 10 baris pertama dalam dataset

```
print(dataset.head(10))
```

	sepal-length	sepal-width	petal-length	petal-width	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa

B

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
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
9	4.9	3.1	1.5	0.1	Iris-setosa

## ▼ Distribusi Kelas Data

Mengelompokkan data berdasarkan kelas dan ukuran data dalam dataset

```
print(dataset.groupby('class').size())
```

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

## ▼ Ringkasan Statistik

Mendeskripsikan data yang ada pada dataset yang meliputi count, mean, dan lainnya.

```
print(dataset.describe())
```

	sepal-length	sepal-width	petal-length	petal-width
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

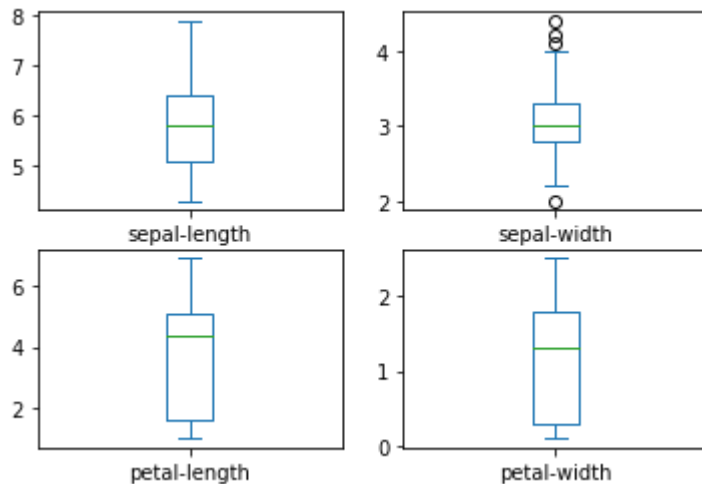
## ▼ Visualisasi dengan Plot Univariat

Plot Univariat merupakan plot dari setiap variable individu. Berikut digunakan plot berjenis plotbox dengan input numerik.

```
import matplotlib.pyplot as plt
dataset.plot(kind='box', subplots=True,
```

B

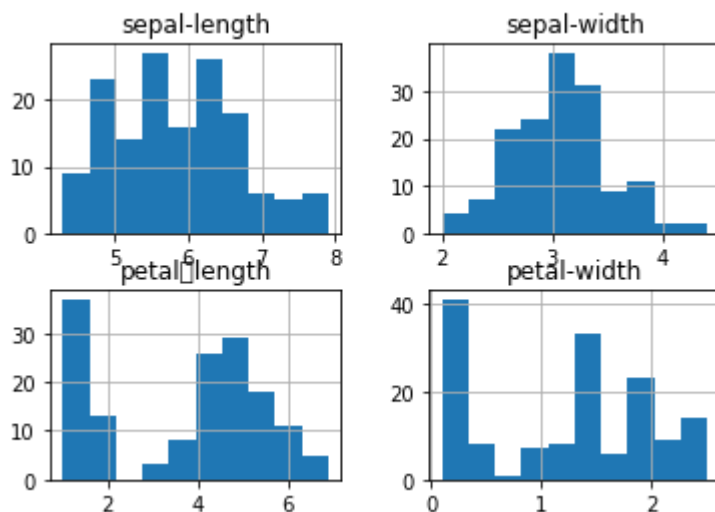
```
layout=(2,2), sharex=False, sharey=False)
plt.show()
```



## ▼ Histogram

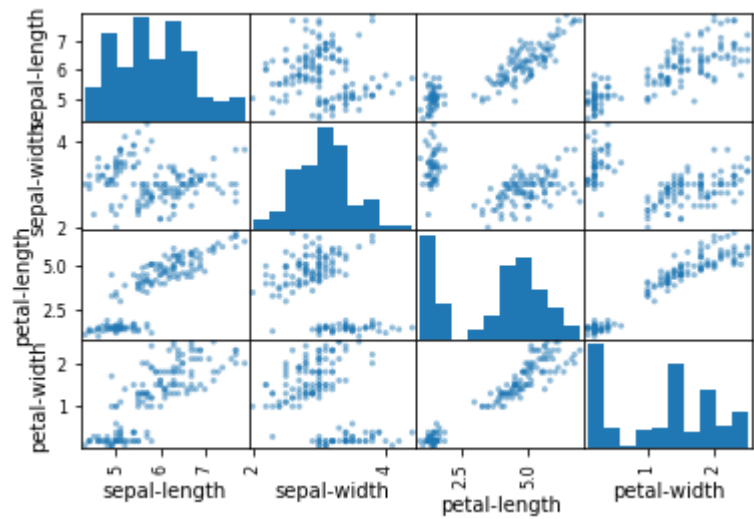
Berikut merupakan kode untuk membuat histogram untuk masing-masing variable input

```
dataset.hist()
plt.show()
```



## ▼ Visualisasi Data Menggunakan Plot Multivariat

```
from pandas.plotting import scatter_matrix
scatter_matrix(dataset)
plt.show()
```



[Colab paid products](#) - [Cancel contracts here](#)

✓ 0s completed at 10:31 PM

● ✕