IMPORTING THE LIBRARIES

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

from sklearn import datasets
```

Loading the data set and understanding the data

```
iris = datasets.load_iris()

df = pd.DataFrame(iris.data, columns = iris.feature_names)

df
```

```
     sepal length (cm)  sepal width (cm)  petal length (cm)  petal
width (cm)
0                  5.1               3.5                1.4
0.2
1                  4.9               3.0                1.4
0.2
2                  4.7               3.2                1.3
0.2
3                  4.6               3.1                1.5
0.2
4                  5.0               3.6                1.4
0.2
..                 ...               ...                ...
...
145                6.7               3.0                5.2
2.3
146                6.3               2.5                5.0
1.9
147                6.5               3.0                5.2
2.0
148                6.2               3.4                5.4
2.3
149                5.9               3.0                5.1
1.8

[150 rows x 4 columns]
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 4 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   sepal length (cm)  150 non-null    float64
 1   sepal width (cm)   150 non-null    float64
```

```
 2   petal length (cm)  150 non-null    float64
 3   petal width (cm)   150 non-null    float64
dtypes: float64(4)
memory usage: 4.8 KB
```

Checking for null values

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

```
sepal length (cm)    0
sepal width (cm)     0
petal length (cm)    0
petal width (cm)     0
dtype: int64
```

No null values

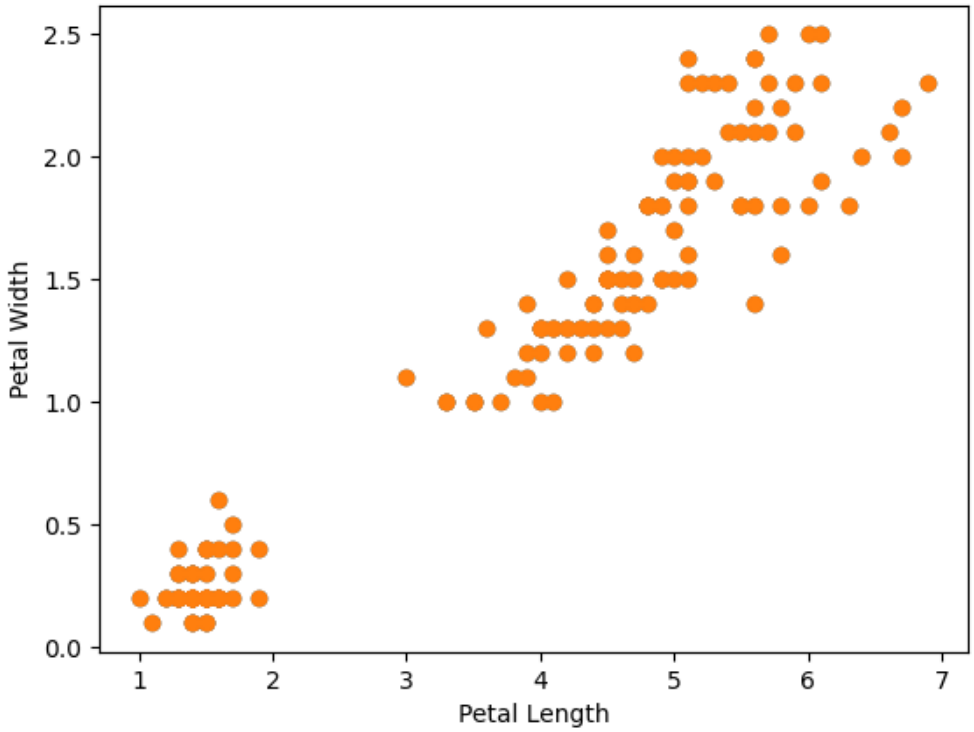
Selecting our features: We choose to form clusters based on two features.

1. Petal length
2. Petal width

```
X = df.iloc[:,2:4]
X
```

```
     petal length (cm)  petal width (cm)
0                  1.4               0.2
1                  1.4               0.2
2                  1.3               0.2
3                  1.5               0.2
4                  1.4               0.2
..                 ...               ...
145                5.2               2.3
146                5.0               1.9
147                5.2               2.0
148                5.4               2.3
149                5.1               1.8

[150 rows x 2 columns]
```

```
plt.scatter(X['petal length (cm)'],X['petal width (cm)'])
plt.xlabel('Petal Length')
plt.ylabel('Petal Width')
plt.show()
```

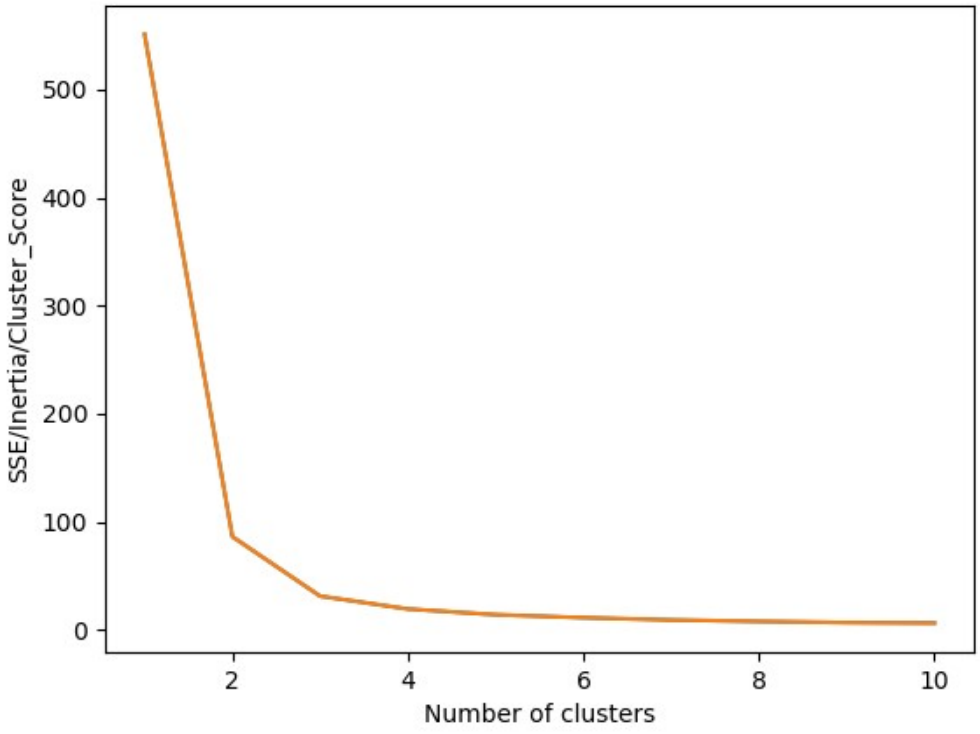


FINDING NUMBER OF CLUSTERS: ELBOW METHOD

```
from sklearn.cluster import KMeans

sse = []

for i in range(1,11):
    kms = KMeans(n_clusters = i, init = 'random', random_state = 43)
    kms.fit(X)
    sse.append(kms.inertia_)

sse1 = pd.DataFrame(sse, range(1,11))

sse1
```

```
              0
1   550.895333
2    86.390220
3    31.371359
4    19.483001
5    14.171086
6    11.447138
7     9.355972
8     8.012587
9     6.916963
10    6.483080
```

```
plt.plot(range(1,11),sse)
plt.ylabel('SSE/Inertia/Cluster_Score')
plt.xlabel('Number of clusters')
plt.show()
```



Creating clusters through sklearn.KMeans

```
Kmeans = KMeans(n_clusters =3, init = 'random', random_state = 43)
Kmeans.fit(X)
pred = Kmeans.predict(X)

pred
```

```
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
1,
       1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 2, 0, 0, 0,
0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 2, 2, 2, 2, 2, 0, 2, 2,
2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 2, 2, 2, 2, 2, 2, 0, 2, 2, 2, 2,
```

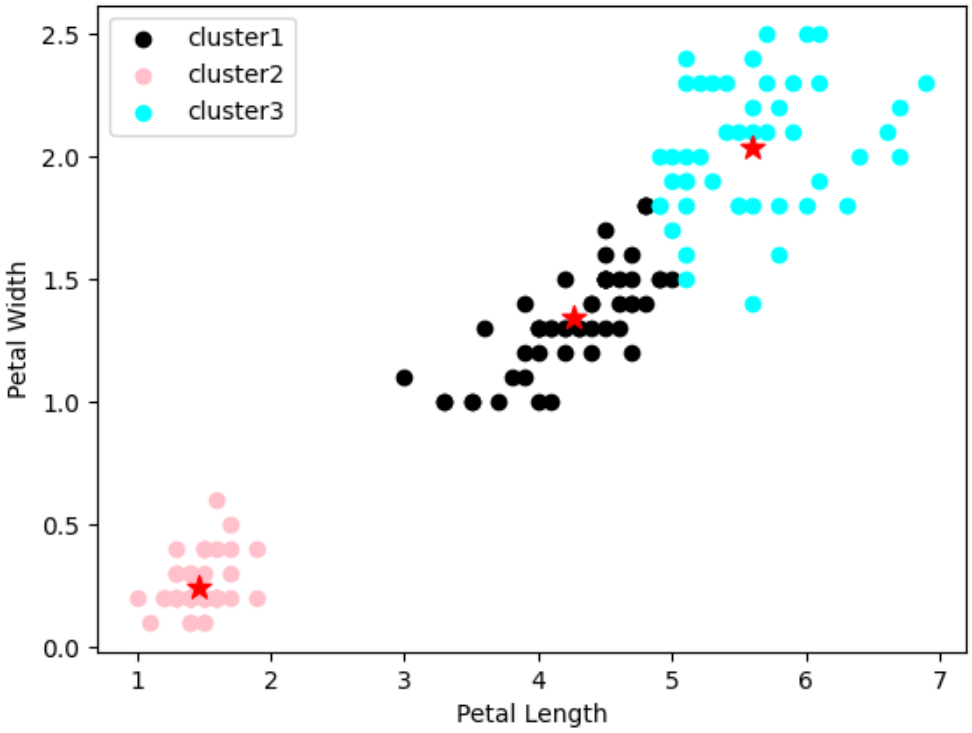
```
2,
       2, 2, 2, 2, 2, 2, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2])
```

```
CC = Kmeans.cluster_centers_
CC
```

```
array([[4.26923077, 1.34230769],
       [1.462     , 0.246     ],
       [5.59583333, 2.0375    ]])
```

VISUALIZING THE CLUSTERS

```
pred1 = X[pred == 0]
pred2 = X[pred == 1]
pred3 = X[pred == 2]

plt.scatter(pred1.iloc[:,0],pred1.iloc[:,1],c = 'black', label =
'cluster1')
plt.scatter(pred2.iloc[:,0],pred2.iloc[:,1],c = 'pink', label =
'cluster2')
plt.scatter(pred3.iloc[:,0],pred3.iloc[:,1],c = 'cyan', label =
'cluster3')
plt.xlabel('Petal Length')
plt.ylabel('Petal Width')
plt.legend()

plt.scatter(CC[:,0],CC[:,1],c='red',s = 100, marker = '*')
plt.show()
```


cluster1
cluster2
cluster3
Petal Width
Petal Length
0.0
0.5
1.0
1.5
2.0
2.5
1
2
3
4
5
6
7