TASK-2 Prediction_using_Unsupervised_ML | Sparks Foundation

Name: Ashishkumar Trada

In [1]:
```
# import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
```
# extracting data
iris_data = pd.read_csv('iris.csv')
```

In [3]:
```
# shape of the data
iris_data.shape
```

Out[3]:
```
(150, 6)
```

In [4]:
```
iris_data.head(10)
```

Out[4]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
9	10	4.9	3.1	1.5	0.1	Iris-setosa

In [5]:
```
# data values
iris_data.Species.value_counts()
```

Out[5]:
```
Iris-versicolor    50
Iris-virginica     50
Iris-setosa        50
Name: Species, dtype: int64
```

In [6]:
```
# getting features from data

X = iris_data.iloc[:, 1:-1].values
```

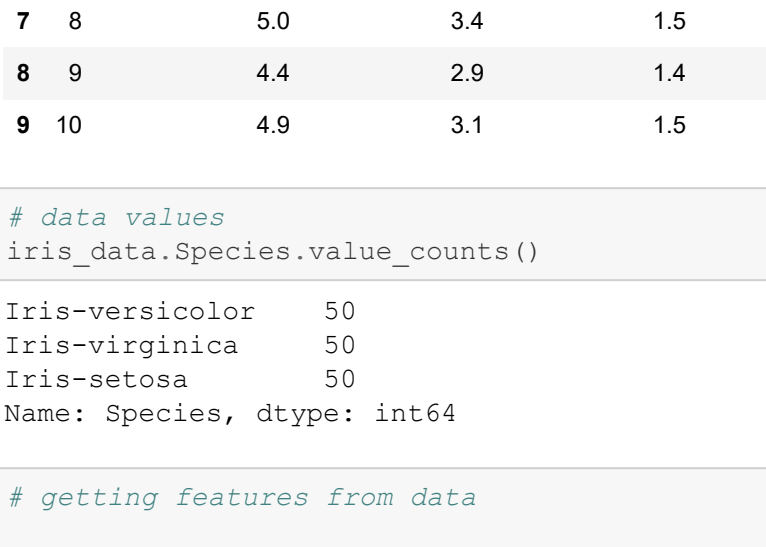
K-Means

In [7]:
```
from sklearn.cluster import KMeans
```

In [8]:
```
# finding the best value of K using Elbow method

wcss = []
for i in range(1,11):
    kmeans = KMeans(init='k-means++', n_clusters=i, random_state=0)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
```

In [9]:
```
plt.plot(range(1,11), wcss)
plt.title('The Elbow Method')
plt.xlabel('clusters')
plt.ylabel('wcss')
plt.show()
```



In [10]:
```
# we have a K(no. of clusters)=3 for our dataset

kmeans = KMeans(n_clusters=3, init='k-means++', random_state=0)
kmeans.fit(X)
y = kmeans.predict(X)
```

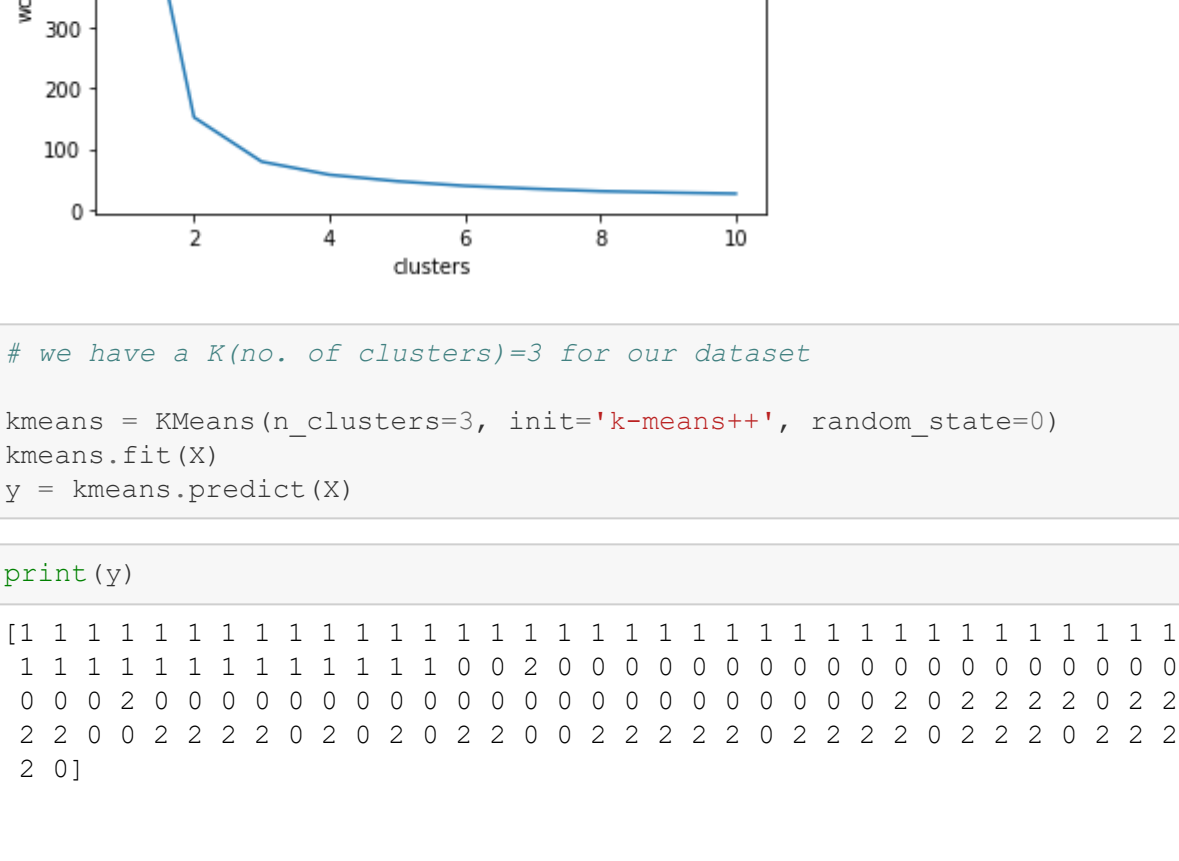
In [11]:
```
print(y)
```

```
[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 1 1 0 0 2 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 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0 2 2 2 2 0 2 2 2 2
 2 2 0 0 2 2 2 2 0 2 0 2 0 2 2 0 0 2 2 2 2 2 0 2 2 2 2 0 2 2 2 0 2 2 2 0 2
 2 0]
```

Clusters

In [12]:
```
# ploating clusters

plt.figure(figsize=(10,7))
plt.scatter(X[y==0, 0], X[y==0, 1], c='red', s=50, label = 'Iris-setosa')
plt.scatter(X[y==1, 0], X[y==1, 1], c='green', s=50, label = 'Iris-versicolor')
plt.scatter(X[y==2, 0], X[y==2, 1], c='blue', s=50, label = 'Iris-virginica')
plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1], c='yellow', s=100, label = 'cen
ters')
plt.title('The flower cluster')
plt.legend()
plt.show()
```



In [ ]:
```
```