

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.datasets import load_iris
from sklearn.decomposition import PCA
from sklearn.datasets import make_blobs
```

Number Cluster define

```
File "<ipython-input-3-e3fef81eb25>", line 1
    Number Cluster define
      ^
SyntaxError: invalid syntax
```

Number cluster define

k=4

Generate sample data

```
x,y=make_blobs(n_samples=300,centers=4,cluster_std=0.60,random_state=0)
```

```
kmeans=KMeans(n_clusters=k)
```

```
kmeans.fit(x)
centroids=kmeans.cluster_centers_
```

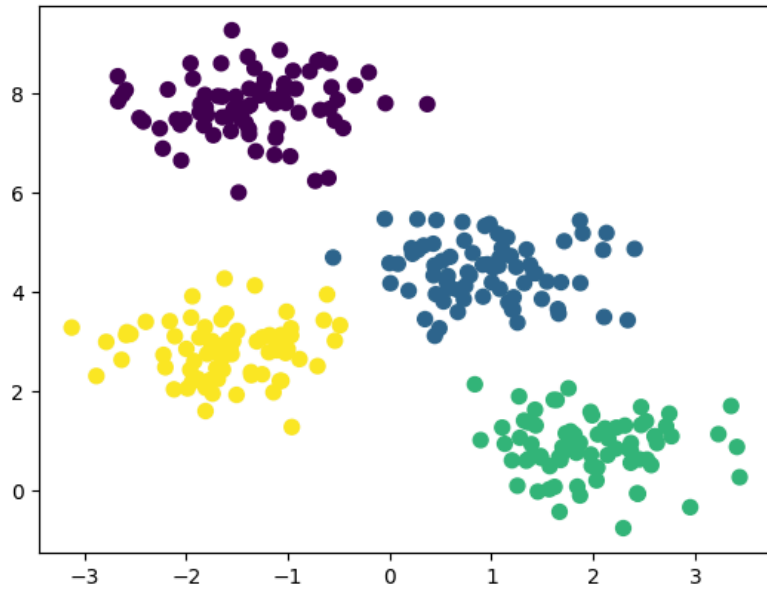
```
/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will be changed from 1 to 10 in version 1.4. To suppress this warning, please use `n_init='auto'` or explicitly set `n_init` to the desired value.
  warnings.warn(
```

```
labels=kmeans.labels_
```

plot the data point

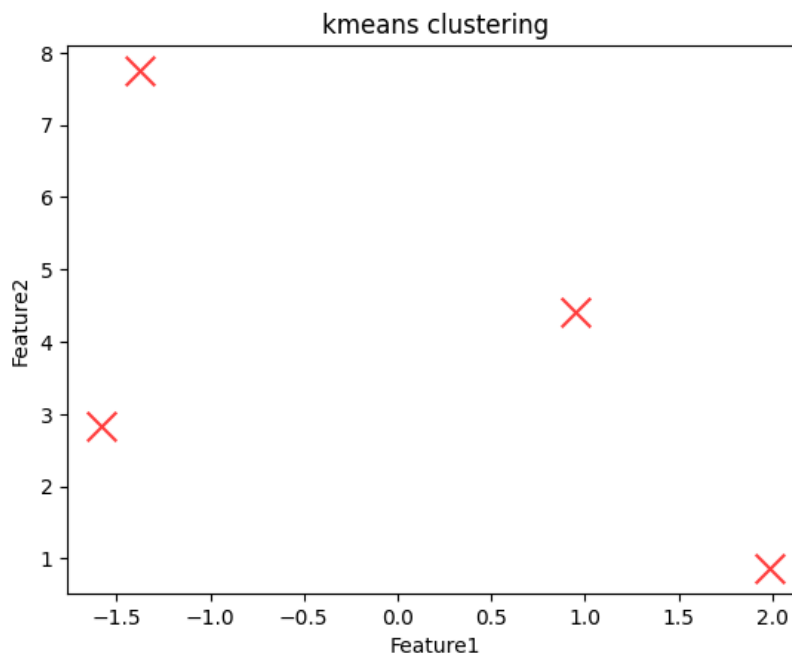
```
plt.scatter(x[:,0],x[:,1],c=labels, s=50,cmap='viridis')
```

 <matplotlib.collections.PathCollection at 0x7e05be6e72b0>

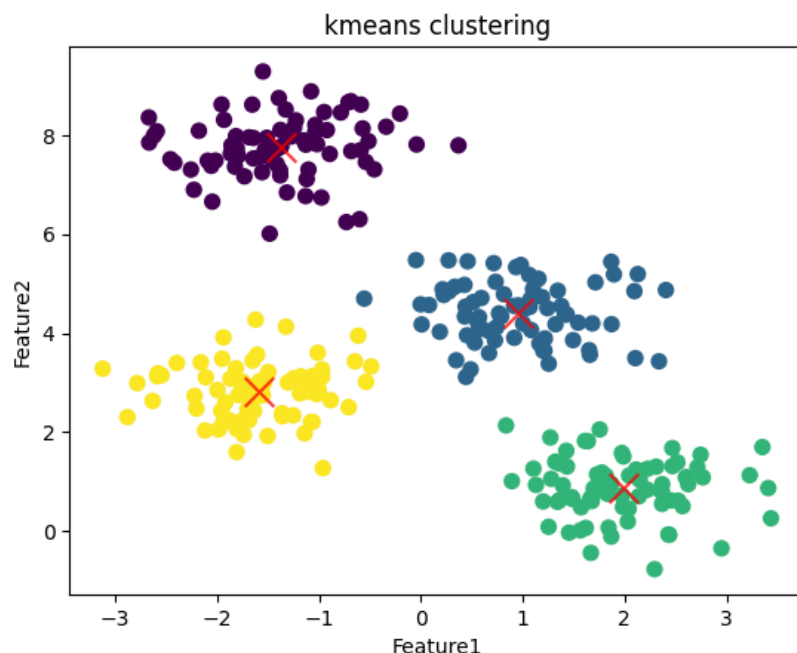


```
plt.scatter(centroids[:,0],centroids[:,1],c='red',s=200,alpha=0.75,marker='x')
plt.title("kmeans clustering")
plt.xlabel("Feature1")
plt.ylabel("Feature2")
plt.show()
```





```
plt.scatter(x[:,0],x[:,1],c=labels,s=50,cmap='viridis')
plt.scatter(centroids[:,0],centroids[:,1],c='red',s=200,alpha=0.75,marker='x')
plt.title("kmeans clustering")
plt.xlabel("Feature1")
plt.ylabel("Feature2")
plt.show()
```



```
import pandas as pd
```

```
path_of_data="/content/drive/MyDrive/Iris.csv"
df=pd.read_csv(path_of_data)
```

```
x=df[['SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm']]
```

```
k=3
kmeans=KMeans(n_clusters=k)
kmeans.fit(x)
centroids=kmeans.cluster_centers_
labels=kmeans.labels_
```

/usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 1 in the future. This will affect the results of the fit method. warnings.warn(

```
pca=PCA(n_components=2)
x_pca=pca.fit_transform(x)
centroids_pca=pca.transform(centroids)
```

/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does not have valid feature names, but PCA will assume all columns in X contain numerical features. warnings.warn(

```
plt.scatter(x_pca[:,0],x_pca[:,1],c=labels,s=50,cmap='viridis')
```

 <matplotlib.collections.PathCollection at 0x7e05be6acbb0>

