A8: Apply Principal Component Analysis (PCA) using sci-kit-learn to reduce the dimensionality of a dataset and visualize the reduced features.

Load Iris Dataset

In [1]:
```
from sklearn.datasets import load_iris
from sklearn.preprocessing import StandardScaler
```

In [2]:
```
# Step 1: Load a sample dataset (Iris dataset)
iris = load_iris()
X = iris.data        # features
y = iris.target      # target labels
```

In [3]:
```
X
```

Out[3]:
```
array([[5.1, 3.5, 1.4, 0.2],
       [4.9, 3. , 1.4, 0.2],
       [4.7, 3.2, 1.3, 0.2],
       [4.6, 3.1, 1.5, 0.2],
       [5. , 3.6, 1.4, 0.2],
       [5.4, 3.9, 1.7, 0.4],
       [4.6, 3.4, 1.4, 0.3],
       [5. , 3.4, 1.5, 0.2],
       [4.4, 2.9, 1.4, 0.2],
       [4.9, 3.1, 1.5, 0.1],
       [5.4, 3.7, 1.5, 0.2],
       [4.8, 3.4, 1.6, 0.2],
       [4.8, 3. , 1.4, 0.1],
       [4.3, 3. , 1.1, 0.1],
       [5.8, 4. , 1.2, 0.2],
       [5.7, 4.4, 1.5, 0.4],
       [5.4, 3.9, 1.3, 0.4],
       [5.1, 3.5, 1.4, 0.3],
       [5.7, 3.8, 1.7, 0.3],
       [5.1, 3.8, 1.5, 0.3],
       [5.4, 3.4, 1.7, 0.2],
       [5.1, 3.7, 1.5, 0.4],
       [4.6, 3.6, 1. , 0.2],
       [5.1, 3.3, 1.7, 0.5],
       [4.8, 3.4, 1.9, 0.2],
       [5. , 3. , 1.6, 0.2],
       [5. , 3.4, 1.6, 0.4],
       [5.2, 3.5, 1.5, 0.2],
       [5.2, 3.4, 1.4, 0.2],
       [4.7, 3.2, 1.6, 0.2],
       [4.8, 3.1, 1.6, 0.2],
       [5.4, 3.4, 1.5, 0.4],
       [5.2, 4.1, 1.5, 0.1],
       [5.5, 4.2, 1.4, 0.2],
       [4.9, 3.1, 1.5, 0.2],
       [5. , 3.2, 1.2, 0.2],
       [5.5, 3.5, 1.3, 0.2],
       [4.9, 3.6, 1.4, 0.1],
       [4.4, 3. , 1.3, 0.2],
       [5.1, 3.4, 1.5, 0.2],
       [5. , 3.5, 1.3, 0.3],
       [4.5, 2.3, 1.3, 0.3],
```

```
       [4.4, 3.2, 1.3, 0.2],
       [5. , 3.5, 1.6, 0.6],
       [5.1, 3.8, 1.9, 0.4],
       [4.8, 3. , 1.4, 0.3],
       [5.1, 3.8, 1.6, 0.2],
       [4.6, 3.2, 1.4, 0.2],
       [5.3, 3.7, 1.5, 0.2],
       [5. , 3.3, 1.4, 0.2],
       [7. , 3.2, 4.7, 1.4],
       [6.4, 3.2, 4.5, 1.5],
       [6.9, 3.1, 4.9, 1.5],
       [5.5, 2.3, 4. , 1.3],
       [6.5, 2.8, 4.6, 1.5],
       [5.7, 2.8, 4.5, 1.3],
       [6.3, 3.3, 4.7, 1.6],
       [4.9, 2.4, 3.3, 1. ],
       [6.6, 2.9, 4.6, 1.3],
       [5.2, 2.7, 3.9, 1.4],
       [5. , 2. , 3.5, 1. ],
       [5.9, 3. , 4.2, 1.5],
       [6. , 2.2, 4. , 1. ],
       [6.1, 2.9, 4.7, 1.4],
       [5.6, 2.9, 3.6, 1.3],
       [6.7, 3.1, 4.4, 1.4],
       [5.6, 3. , 4.5, 1.5],
       [5.8, 2.7, 4.1, 1. ],
       [6.2, 2.2, 4.5, 1.5],
       [5.6, 2.5, 3.9, 1.1],
       [5.9, 3.2, 4.8, 1.8],
       [6.1, 2.8, 4. , 1.3],
       [6.3, 2.5, 4.9, 1.5],
       [6.1, 2.8, 4.7, 1.2],
       [6.4, 2.9, 4.3, 1.3],
       [6.6, 3. , 4.4, 1.4],
       [6.8, 2.8, 4.8, 1.4],
       [6.7, 3. , 5. , 1.7],
       [6. , 2.9, 4.5, 1.5],
       [5.7, 2.6, 3.5, 1. ],
       [5.5, 2.4, 3.8, 1.1],
       [5.5, 2.4, 3.7, 1. ],
       [5.8, 2.7, 3.9, 1.2],
       [6. , 2.7, 5.1, 1.6],
```

```
       [5.4, 3. , 4.5, 1.5],
       [6. , 3.4, 4.5, 1.6],
       [6.7, 3.1, 4.7, 1.5],
       [6.3, 2.3, 4.4, 1.3],
       [5.6, 3. , 4.1, 1.3],
       [5.5, 2.5, 4. , 1.3],
       [5.5, 2.6, 4.4, 1.2],
       [6.1, 3. , 4.6, 1.4],
       [5.8, 2.6, 4. , 1.2],
       [5. , 2.3, 3.3, 1. ],
       [5.6, 2.7, 4.2, 1.3],
       [5.7, 3. , 4.2, 1.2],
       [5.7, 2.9, 4.2, 1.3],
       [6.2, 2.9, 4.3, 1.3],
       [5.1, 2.5, 3. , 1.1],
       [5.7, 2.8, 4.1, 1.3],
       [6.3, 3.3, 6. , 2.5],
       [5.8, 2.7, 5.1, 1.9],
       [7.1, 3. , 5.9, 2.1],
       [6.3, 2.9, 5.6, 1.8],
       [6.5, 3. , 5.8, 2.2],
       [7.6, 3. , 6.6, 2.1],
       [4.9, 2.5, 4.5, 1.7],
       [7.3, 2.9, 6.3, 1.8],
       [6.7, 2.5, 5.8, 1.8],
       [7.2, 3.6, 6.1, 2.5],
       [6.5, 3.2, 5.1, 2. ],
       [6.4, 2.7, 5.3, 1.9],
       [6.8, 3. , 5.5, 2.1],
       [5.7, 2.5, 5. , 2. ],
       [5.8, 2.8, 5.1, 2.4],
       [6.4, 3.2, 5.3, 2.3],
       [6.5, 3. , 5.5, 1.8],
       [7.7, 3.8, 6.7, 2.2],
       [7.7, 2.6, 6.9, 2.3],
       [6. , 2.2, 5. , 1.5],
       [6.9, 3.2, 5.7, 2.3],
       [5.6, 2.8, 4.9, 2. ],
       [7.7, 2.8, 6.7, 2. ],
       [6.3, 2.7, 4.9, 1.8],
       [6.7, 3.3, 5.7, 2.1],
       [7.2, 3.2, 6. , 1.8],
```

```
       [6.2, 2.8, 4.8, 1.8],
       [6.1, 3. , 4.9, 1.8],
       [6.4, 2.8, 5.6, 2.1],
       [7.2, 3. , 5.8, 1.6],
       [7.4, 2.8, 6.1, 1.9],
       [7.9, 3.8, 6.4, 2. ],
       [6.4, 2.8, 5.6, 2.2],
       [6.3, 2.8, 5.1, 1.5],
       [6.1, 2.6, 5.6, 1.4],
       [7.7, 3. , 6.1, 2.3],
       [6.3, 3.4, 5.6, 2.4],
       [6.4, 3.1, 5.5, 1.8],
       [6. , 3. , 4.8, 1.8],
       [6.9, 3.1, 5.4, 2.1],
       [6.7, 3.1, 5.6, 2.4],
       [6.9, 3.1, 5.1, 2.3],
       [5.8, 2.7, 5.1, 1.9],
       [6.8, 3.2, 5.9, 2.3],
       [6.7, 3.3, 5.7, 2.5],
       [6.7, 3. , 5.2, 2.3],
       [6.3, 2.5, 5. , 1.9],
       [6.5, 3. , 5.2, 2. ],
       [6.2, 3.4, 5.4, 2.3],
       [5.9, 3. , 5.1, 1.8]])
```

Standardize the Data

PCA is effected by scale so you need to scale the features in your data before applying PCA. Use StandardScaler to help you standardize the dataset's features onto unit scale (mean = 0 and variance = 1) which is a requirement for the optimal performance of many machine learning algorithms.

In [4]:
```
X = StandardScaler().fit_transform(X)
```

In [5]:
```
X
```

Out[5]:
```
array([[-9.00681170e-01,  1.01900435e+00, -1.34022653e+00,
        -1.31544430e+00],
       [-1.14301691e+00, -1.31979479e-01, -1.34022653e+00,
        -1.31544430e+00],
       [-1.38535265e+00,  3.28414053e-01, -1.39706395e+00,
        -1.31544430e+00],
       [-1.50652052e+00,  9.82172869e-02, -1.28338910e+00,
        -1.31544430e+00],
       [-1.02184904e+00,  1.24920112e+00, -1.34022653e+00,
        -1.31544430e+00],
       [-5.37177559e-01,  1.93979142e+00, -1.16971425e+00,
        -1.05217993e+00],
       [-1.50652052e+00,  7.88807586e-01, -1.34022653e+00,
        -1.18381211e+00],
       [-1.02184904e+00,  7.88807586e-01, -1.28338910e+00,
        -1.31544430e+00],
       [-1.74885626e+00, -3.62176246e-01, -1.34022653e+00,
        -1.31544430e+00],
       [-1.14301691e+00,  9.82172869e-02, -1.28338910e+00,
        -1.44707648e+00],
       [-5.37177559e-01,  1.47939788e+00, -1.28338910e+00,
        -1.31544430e+00],
       [-1.26418478e+00,  7.88807586e-01, -1.22655167e+00,
        -1.31544430e+00],
       [-1.26418478e+00, -1.31979479e-01, -1.34022653e+00,
        -1.44707648e+00],
       [-1.87002413e+00, -1.31979479e-01, -1.51073881e+00,
        -1.44707648e+00],
       [-5.25060772e-02,  2.16998818e+00, -1.45390138e+00,
        -1.31544430e+00],
       [-1.73673948e-01,  3.09077525e+00, -1.28338910e+00,
        -1.05217993e+00],
       [-5.37177559e-01,  1.93979142e+00, -1.39706395e+00,
        -1.05217993e+00],
       [-9.00681170e-01,  1.01900435e+00, -1.34022653e+00,
        -1.18381211e+00],
       [-1.73673948e-01,  1.70959465e+00, -1.16971425e+00,
        -1.18381211e+00],
       [-9.00681170e-01,  1.70959465e+00, -1.28338910e+00,
        -1.18381211e+00],
       [-5.37177559e-01,  7.88807586e-01, -1.16971425e+00,
        -1.31544430e+00],
```

```
       [-9.00681170e-01,  1.47939788e+00, -1.28338910e+00,
        -1.05217993e+00],
       [-1.50652052e+00,  1.24920112e+00, -1.56757623e+00,
        -1.31544430e+00],
       [-9.00681170e-01,  5.58610819e-01, -1.16971425e+00,
        -9.20547742e-01],
       [-1.26418478e+00,  7.88807586e-01, -1.05603939e+00,
        -1.31544430e+00],
       [-1.02184904e+00, -1.31979479e-01, -1.22655167e+00,
        -1.31544430e+00],
       [-1.02184904e+00,  7.88807586e-01, -1.22655167e+00,
        -1.05217993e+00],
       [-7.79513300e-01,  1.01900435e+00, -1.28338910e+00,
        -1.31544430e+00],
       [-7.79513300e-01,  7.88807586e-01, -1.34022653e+00,
        -1.31544430e+00],
       [-1.38535265e+00,  3.28414053e-01, -1.22655167e+00,
        -1.31544430e+00],
       [-1.26418478e+00,  9.82172869e-02, -1.22655167e+00,
        -1.31544430e+00],
       [-5.37177559e-01,  7.88807586e-01, -1.28338910e+00,
        -1.05217993e+00],
       [-7.79513300e-01,  2.40018495e+00, -1.28338910e+00,
        -1.44707648e+00],
       [-4.16009689e-01,  2.63038172e+00, -1.34022653e+00,
        -1.31544430e+00],
       [-1.14301691e+00,  9.82172869e-02, -1.28338910e+00,
        -1.31544430e+00],
       [-1.02184904e+00,  3.28414053e-01, -1.45390138e+00,
        -1.31544430e+00],
       [-4.16009689e-01,  1.01900435e+00, -1.39706395e+00,
        -1.31544430e+00],
       [-1.14301691e+00,  1.24920112e+00, -1.34022653e+00,
        -1.44707648e+00],
       [-1.74885626e+00, -1.31979479e-01, -1.39706395e+00,
        -1.31544430e+00],
       [-9.00681170e-01,  7.88807586e-01, -1.28338910e+00,
        -1.31544430e+00],
       [-1.02184904e+00,  1.01900435e+00, -1.39706395e+00,
        -1.18381211e+00],
       [-1.62768839e+00, -1.74335684e+00, -1.39706395e+00,
        -1.18381211e+00],
```

```
[-1.74885626e+00,  3.28414053e-01, -1.39706395e+00,
 -1.31544430e+00],
[-1.02184904e+00,  1.01900435e+00, -1.22655167e+00,
 -7.88915558e-01],
[-9.00681170e-01,  1.70959465e+00, -1.05603939e+00,
 -1.05217993e+00],
[-1.26418478e+00, -1.31979479e-01, -1.34022653e+00,
 -1.18381211e+00],
[-9.00681170e-01,  1.70959465e+00, -1.22655167e+00,
 -1.31544430e+00],
[-1.50652052e+00,  3.28414053e-01, -1.34022653e+00,
 -1.31544430e+00],
[-6.58345429e-01,  1.47939788e+00, -1.28338910e+00,
 -1.31544430e+00],
[-1.02184904e+00,  5.58610819e-01, -1.34022653e+00,
 -1.31544430e+00],
[ 1.40150837e+00,  3.28414053e-01,  5.35408562e-01,
  2.64141916e-01],
[ 6.74501145e-01,  3.28414053e-01,  4.21733708e-01,
  3.95774101e-01],
[ 1.28034050e+00,  9.82172869e-02,  6.49083415e-01,
  3.95774101e-01],
[-4.16009689e-01, -1.74335684e+00,  1.37546573e-01,
  1.32509732e-01],
[ 7.95669016e-01, -5.92373012e-01,  4.78571135e-01,
  3.95774101e-01],
[-1.73673948e-01, -5.92373012e-01,  4.21733708e-01,
  1.32509732e-01],
[ 5.53333275e-01,  5.58610819e-01,  5.35408562e-01,
  5.27406285e-01],
[-1.14301691e+00, -1.51316008e+00, -2.60315415e-01,
 -2.62386821e-01],
[ 9.16836886e-01, -3.62176246e-01,  4.78571135e-01,
  1.32509732e-01],
[-7.79513300e-01, -8.22569778e-01,  8.07091462e-02,
  2.64141916e-01],
[-1.02184904e+00, -2.43394714e+00, -1.46640561e-01,
 -2.62386821e-01],
[ 6.86617933e-02, -1.31979479e-01,  2.51221427e-01,
  3.95774101e-01],
[ 1.89829664e-01, -1.97355361e+00,  1.37546573e-01,
 -2.62386821e-01],
```

```
[ 3.10997534e-01, -3.62176246e-01,  5.35408562e-01,
  2.64141916e-01],
[-2.94841818e-01, -3.62176246e-01, -8.98031345e-02,
  1.32509732e-01],
[ 1.03800476e+00,  9.82172869e-02,  3.64896281e-01,
  2.64141916e-01],
[-2.94841818e-01, -1.31979479e-01,  4.21733708e-01,
  3.95774101e-01],
[-5.25060772e-02, -8.22569778e-01,  1.94384000e-01,
 -2.62386821e-01],
[ 4.32165405e-01, -1.97355361e+00,  4.21733708e-01,
  3.95774101e-01],
[-2.94841818e-01, -1.28296331e+00,  8.07091462e-02,
 -1.30754636e-01],
[ 6.86617933e-02,  3.28414053e-01,  5.92245988e-01,
  7.90670654e-01],
[ 3.10997534e-01, -5.92373012e-01,  1.37546573e-01,
  1.32509732e-01],
[ 5.53333275e-01, -1.28296331e+00,  6.49083415e-01,
  3.95774101e-01],
[ 3.10997534e-01, -5.92373012e-01,  5.35408562e-01,
  8.77547895e-04],
[ 6.74501145e-01, -3.62176246e-01,  3.08058854e-01,
  1.32509732e-01],
[ 9.16836886e-01, -1.31979479e-01,  3.64896281e-01,
  2.64141916e-01],
[ 1.15917263e+00, -5.92373012e-01,  5.92245988e-01,
  2.64141916e-01],
[ 1.03800476e+00, -1.31979479e-01,  7.05920842e-01,
  6.59038469e-01],
[ 1.89829664e-01, -3.62176246e-01,  4.21733708e-01,
  3.95774101e-01],
[-1.73673948e-01, -1.05276654e+00, -1.46640561e-01,
 -2.62386821e-01],
[-4.16009689e-01, -1.51316008e+00,  2.38717193e-02,
 -1.30754636e-01],
[-4.16009689e-01, -1.51316008e+00, -3.29657076e-02,
 -2.62386821e-01],
[-5.25060772e-02, -8.22569778e-01,  8.07091462e-02,
  8.77547895e-04],
[ 1.89829664e-01, -8.22569778e-01,  7.62758269e-01,
  5.27406285e-01],
```

```
       [-5.37177559e-01, -1.31979479e-01,  4.21733708e-01,
         3.95774101e-01],
       [ 1.89829664e-01,  7.88807586e-01,  4.21733708e-01,
         5.27406285e-01],
       [ 1.03800476e+00,  9.82172869e-02,  5.35408562e-01,
         3.95774101e-01],
       [ 5.53333275e-01, -1.74335684e+00,  3.64896281e-01,
         1.32509732e-01],
       [-2.94841818e-01, -1.31979479e-01,  1.94384000e-01,
         1.32509732e-01],
       [-4.16009689e-01, -1.28296331e+00,  1.37546573e-01,
         1.32509732e-01],
       [-4.16009689e-01, -1.05276654e+00,  3.64896281e-01,
         8.77547895e-04],
       [ 3.10997534e-01, -1.31979479e-01,  4.78571135e-01,
         2.64141916e-01],
       [-5.25060772e-02, -1.05276654e+00,  1.37546573e-01,
         8.77547895e-04],
       [-1.02184904e+00, -1.74335684e+00, -2.60315415e-01,
        -2.62386821e-01],
       [-2.94841818e-01, -8.22569778e-01,  2.51221427e-01,
         1.32509732e-01],
       [-1.73673948e-01, -1.31979479e-01,  2.51221427e-01,
         8.77547895e-04],
       [-1.73673948e-01, -3.62176246e-01,  2.51221427e-01,
         1.32509732e-01],
       [ 4.32165405e-01, -3.62176246e-01,  3.08058854e-01,
         1.32509732e-01],
       [-9.00681170e-01, -1.28296331e+00, -4.30827696e-01,
        -1.30754636e-01],
       [-1.73673948e-01, -5.92373012e-01,  1.94384000e-01,
         1.32509732e-01],
       [ 5.53333275e-01,  5.58610819e-01,  1.27429511e+00,
         1.71209594e+00],
       [-5.25060772e-02, -8.22569778e-01,  7.62758269e-01,
         9.22302838e-01],
       [ 1.52267624e+00, -1.31979479e-01,  1.21745768e+00,
         1.18556721e+00],
       [ 5.53333275e-01, -3.62176246e-01,  1.04694540e+00,
         7.90670654e-01],
       [ 7.95669016e-01, -1.31979479e-01,  1.16062026e+00,
         1.31719939e+00],
```

```
       [ 2.12851559e+00, -1.31979479e-01,  1.61531967e+00,
         1.18556721e+00],
       [-1.14301691e+00, -1.28296331e+00,  4.21733708e-01,
         6.59038469e-01],
       [ 1.76501198e+00, -3.62176246e-01,  1.44480739e+00,
         7.90670654e-01],
       [ 1.03800476e+00, -1.28296331e+00,  1.16062026e+00,
         7.90670654e-01],
       [ 1.64384411e+00,  1.24920112e+00,  1.33113254e+00,
         1.71209594e+00],
       [ 7.95669016e-01,  3.28414053e-01,  7.62758269e-01,
         1.05393502e+00],
       [ 6.74501145e-01, -8.22569778e-01,  8.76433123e-01,
         9.22302838e-01],
       [ 1.15917263e+00, -1.31979479e-01,  9.90107977e-01,
         1.18556721e+00],
       [-1.73673948e-01, -1.28296331e+00,  7.05920842e-01,
         1.05393502e+00],
       [-5.25060772e-02, -5.92373012e-01,  7.62758269e-01,
         1.58046376e+00],
       [ 6.74501145e-01,  3.28414053e-01,  8.76433123e-01,
         1.44883158e+00],
       [ 7.95669016e-01, -1.31979479e-01,  9.90107977e-01,
         7.90670654e-01],
       [ 2.24968346e+00,  1.70959465e+00,  1.67215710e+00,
         1.31719939e+00],
       [ 2.24968346e+00, -1.05276654e+00,  1.78583195e+00,
         1.44883158e+00],
       [ 1.89829664e-01, -1.97355361e+00,  7.05920842e-01,
         3.95774101e-01],
       [ 1.28034050e+00,  3.28414053e-01,  1.10378283e+00,
         1.44883158e+00],
       [-2.94841818e-01, -5.92373012e-01,  6.49083415e-01,
         1.05393502e+00],
       [ 2.24968346e+00, -5.92373012e-01,  1.67215710e+00,
         1.05393502e+00],
       [ 5.53333275e-01, -8.22569778e-01,  6.49083415e-01,
         7.90670654e-01],
       [ 1.03800476e+00,  5.58610819e-01,  1.10378283e+00,
         1.18556721e+00],
       [ 1.64384411e+00,  3.28414053e-01,  1.27429511e+00,
         7.90670654e-01],
```

```
[ 4.32165405e-01, -5.92373012e-01,  5.92245988e-01,
  7.90670654e-01],
[ 3.10997534e-01, -1.31979479e-01,  6.49083415e-01,
  7.90670654e-01],
[ 6.74501145e-01, -5.92373012e-01,  1.04694540e+00,
  1.18556721e+00],
[ 1.64384411e+00, -1.31979479e-01,  1.16062026e+00,
  5.27406285e-01],
[ 1.88617985e+00, -5.92373012e-01,  1.33113254e+00,
  9.22302838e-01],
[ 2.49201920e+00,  1.70959465e+00,  1.50164482e+00,
  1.05393502e+00],
[ 6.74501145e-01, -5.92373012e-01,  1.04694540e+00,
  1.31719939e+00],
[ 5.53333275e-01, -5.92373012e-01,  7.62758269e-01,
  3.95774101e-01],
[ 3.10997534e-01, -1.05276654e+00,  1.04694540e+00,
  2.64141916e-01],
[ 2.24968346e+00, -1.31979479e-01,  1.33113254e+00,
  1.44883158e+00],
[ 5.53333275e-01,  7.88807586e-01,  1.04694540e+00,
  1.58046376e+00],
[ 6.74501145e-01,  9.82172869e-02,  9.90107977e-01,
  7.90670654e-01],
[ 1.89829664e-01, -1.31979479e-01,  5.92245988e-01,
  7.90670654e-01],
[ 1.28034050e+00,  9.82172869e-02,  9.33270550e-01,
  1.18556721e+00],
[ 1.03800476e+00,  9.82172869e-02,  1.04694540e+00,
  1.58046376e+00],
[ 1.28034050e+00,  9.82172869e-02,  7.62758269e-01,
  1.44883158e+00],
[-5.25060772e-02, -8.22569778e-01,  7.62758269e-01,
  9.22302838e-01],
[ 1.15917263e+00,  3.28414053e-01,  1.21745768e+00,
  1.44883158e+00],
[ 1.03800476e+00,  5.58610819e-01,  1.10378283e+00,
  1.71209594e+00],
[ 1.03800476e+00, -1.31979479e-01,  8.19595696e-01,
  1.44883158e+00],
[ 5.53333275e-01, -1.28296331e+00,  7.05920842e-01,
  9.22302838e-01],
```

```
       [ 7.95669016e-01, -1.31979479e-01,  8.19595696e-01,
         1.05393502e+00],
       [ 4.32165405e-01,  7.88807586e-01,  9.33270550e-01,
         1.44883158e+00],
       [ 6.86617933e-02, -1.31979479e-01,  7.62758269e-01,
         7.90670654e-01]])
```

In [6]:
```
y
```

Out[6]:
```
array([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, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 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, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2])
```

	sepal length	sepal width	petal length	petal width
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

Standardization →

	sepal length	sepal width	petal length	petal width
0	-0.900681	1.032057	-1.341272	-1.312977
1	-1.143017	-0.124958	-1.341272	-1.312977
2	-1.385353	0.337848	-1.398138	-1.312977
3	-1.506521	0.106445	-1.284407	-1.312977
4	-1.021849	1.263460	-1.341272	-1.312977

The array x (visualized by a pandas dataframe) before and after standardization

PCA Projection to 2D

The original data has 4 columns (sepal length, sepal width, petal length, and petal width). In this section, the code projects the original data which is 4 dimensional into 2 dimensions

In [7]:
```
from sklearn.decomposition import PCA
import pandas as pd
pca = PCA(n_components=2) # creating an instance of PCA
```

```python
principalComponents = pca.fit_transform(X) #fitting the pca instance to the dataset
principalDf = pd.DataFrame(data = principalComponents
             , columns = ['principal component 1', 'principal component 2'])
```

In [8]:
```python
principalDf
```

Out[8]:

	principal component 1	principal component 2
0	-2.264703	0.480027
1	-2.080961	-0.674134
2	-2.364229	-0.341908
3	-2.299384	-0.597395
4	-2.389842	0.646835
...	...	...
145	1.870503	0.386966
146	1.564580	-0.896687
147	1.521170	0.269069
148	1.372788	1.011254
149	0.960656	-0.024332

150 rows × 2 columns

	sepal length	sepal width	petal length	petal width
0	-0.900681	1.032057	-1.341272	-1.312977
1	-1.143017	-0.124958	-1.341272	-1.312977
2	-1.385353	0.337848	-1.398138	-1.312977
3	-1.506521	0.106445	-1.284407	-1.312977
4	-1.021849	1.263460	-1.341272	-1.312977

PCA
(2 components)

	principal component 1	princial component 2
0	-2.264542	0.505704
1	-2.086426	-0.655405
2	-2.367950	-0.318477
3	-2.304197	-0.575368
4	-2.388777	0.674767

PCA and Keeping the Top 2 Principal Components

In [13]:
```
# Combine PCA results with target column
finalDf = pd.concat([principalDf, pd.Series(y, name='target')], axis=1)
finalDf
```

Out[13]:

	principal component 1	principal component 2	target
0	-2.264703	0.480027	0
1	-2.080961	-0.674134	0
2	-2.364229	-0.341908	0
3	-2.299384	-0.597395	0
4	-2.389842	0.646835	0
...	...	...	...
145	1.870503	0.386966	2
146	1.564580	-0.896687	2
147	1.521170	0.269069	2
148	1.372788	1.011254	2
149	0.960656	-0.024332	2

150 rows × 3 columns

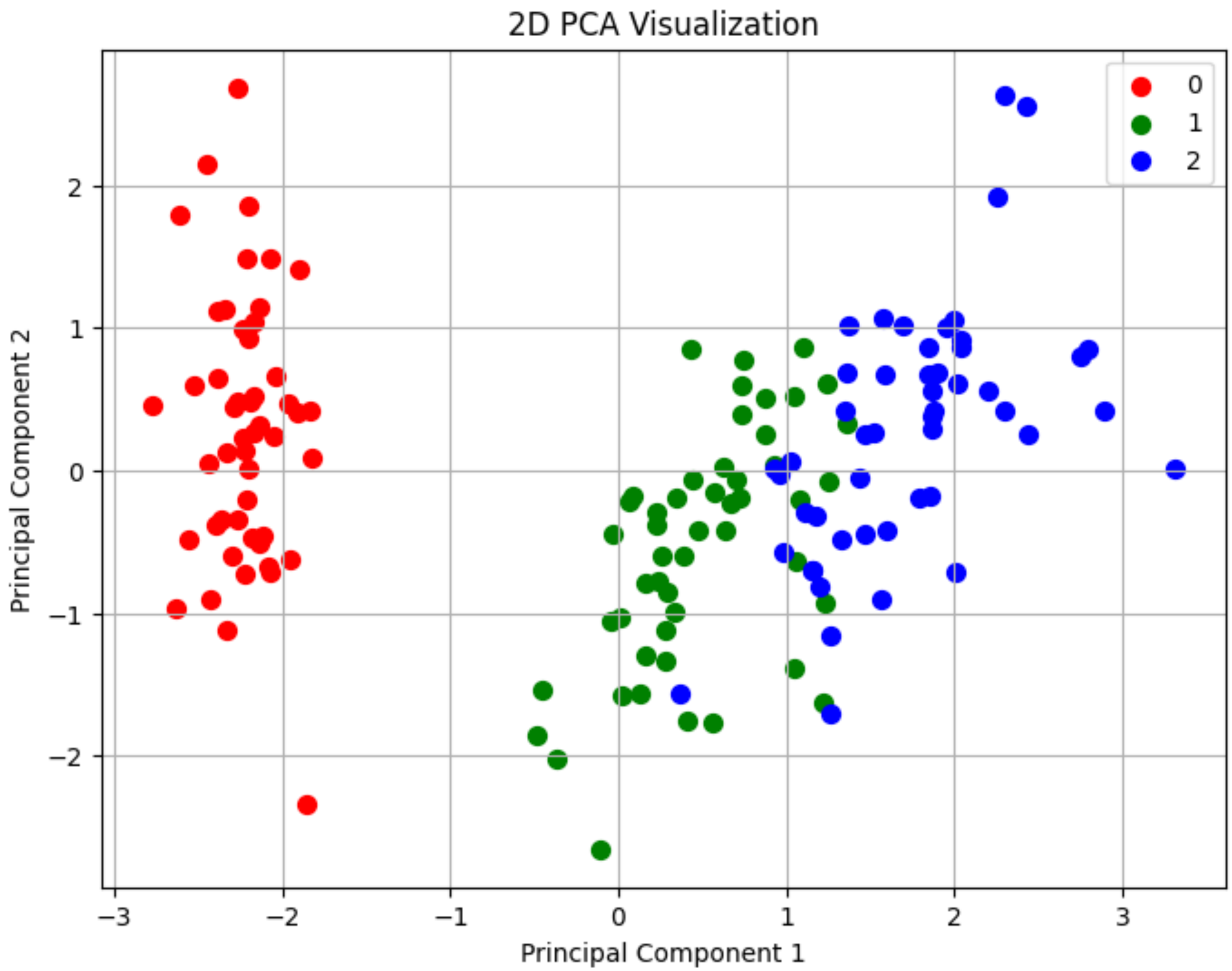
In [16]:
```
# Plot the 2D PCA projection
import matplotlib.pyplot as plt

plt.figure(figsize=(8,6))
targets = list(set(y))  # unique target classes
colors = ['r', 'g', 'b', 'c', 'm', 'y', 'k']  # add more if needed

for target, color in zip(targets, colors):
    indicesToKeep = finalDf['target'] == target
    plt.scatter(finalDf.loc[indicesToKeep, 'principal component 1'],
```

```
                finalDf.loc[indicesToKeep, 'principal component 2'],
                c=color,
                s=50,
                label=target)

plt.xlabel('Principal Component 1')
plt.ylabel('Principal Component 2')
plt.title('2D PCA Visualization')
plt.legend()
plt.grid()
plt.show()
```


2D PCA Visualization
Principal Component 1
Principal Component 2
0
1
2
−3
−2
−1
0
1
2
3
−2
−1
0
1
2