Name:- Muhammad Huzaifa Waseem (2303-KHI-DEG-021)
Pair Partner 1:- Muhammad Faizan Rafique (2303.005.KHI.DEG)
Pair Partner 2:- Syed Muhammad Hammad Irshad(2303.KHI.DEG.032)

UNIT 3.3:

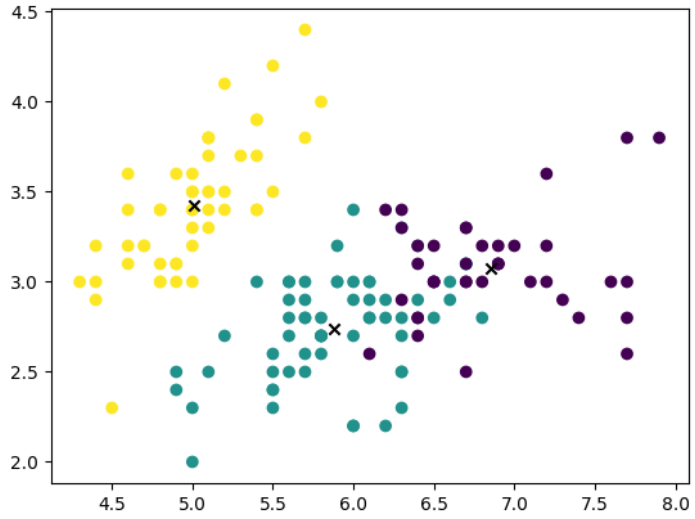
Assignment

K-Mean Clusterization without PCA:-

[5]:
```
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
import numpy as np
from sklearn import datasets
iris = datasets.load_iris()
x = iris.data
y = iris.target
x = iris.data
y = iris.target
```

[6]:
```
model = KMeans(n_clusters=3, n_init=1, max_iter=100)
model.fit(x)

all_predictions = model.predict(x)
centroids = model.cluster_centers_
plt.scatter(x[:,0], x[:,1], c=all_predictions)
plt.scatter(centroids[:,0], centroids[:,1], marker='x', color="black")
plt.show
```

[6]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



K-Mean Clusterization with PCA:-

[7]:
```
from sklearn.decomposition import PCA
x.shape
```

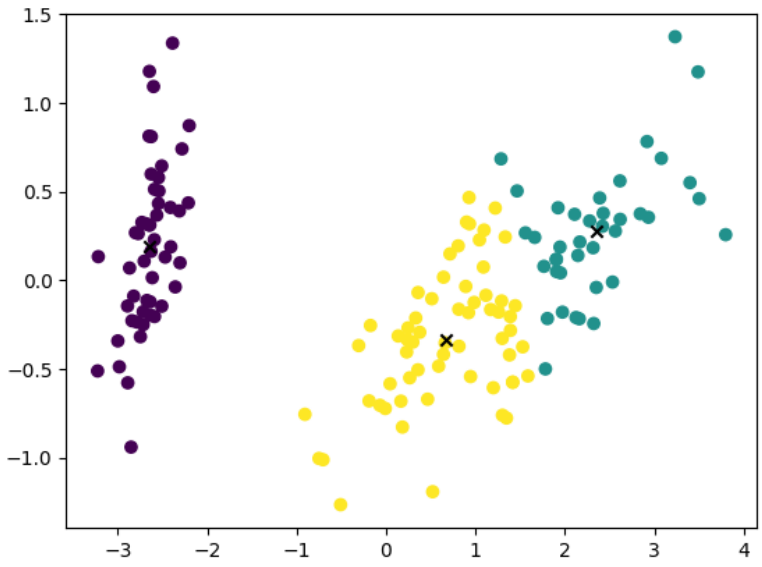
[7]:
```
(150, 4)
```

[8]:
```
pca = PCA(n_components=2)
x_reduced = pca.fit_transform(x)

x_reduced.shape
```

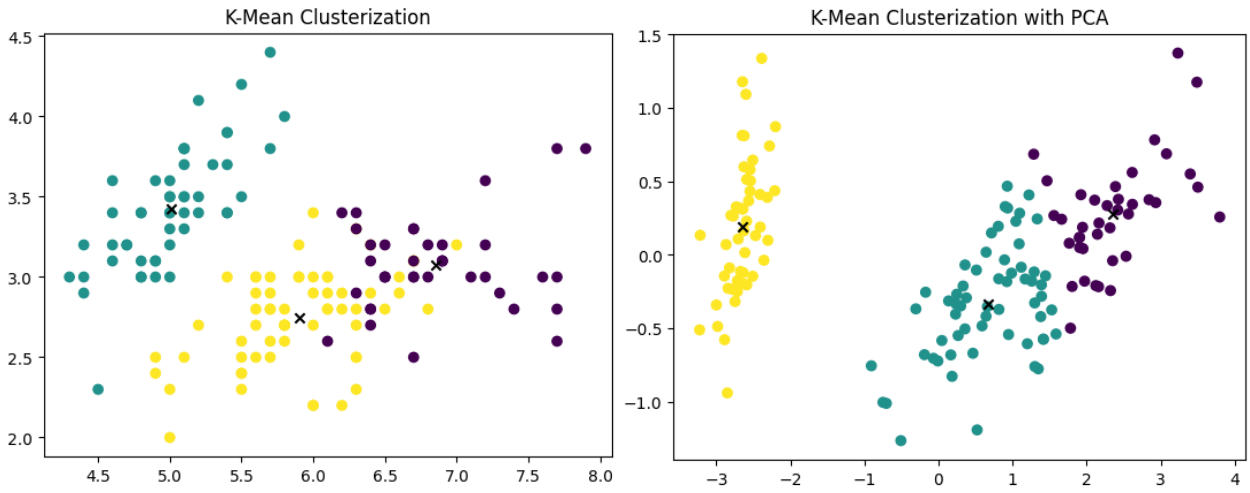
[8]:
```
(150, 2)
```

[10]:
```
model = KMeans(n_clusters=3, n_init=1, max_iter=100)
model.fit(x_reduced)

all_predictions = model.predict(x_reduced)
centroids = model.cluster_centers_
plt.scatter(x_reduced[:,0], x_reduced[:,1], c=all_predictions)
plt.scatter(centroids[:,0], centroids[:,1], marker='x', color="black")
plt.show
```

[10]:
```
<function matplotlib.pyplot.show(close=None, block=None)>
```



Comparison



Explanation

Before applying PCA, the data points were scattered all over the graph and it was difficult to separate them into distinct groups. However, after applying PCA, the data points were transformed so that they were closer to the centroid, making it easier to read and analyze the data. This is because PCA helps to reduce the dimensionality of the data, highlighting the most important features and patterns in the data.