TASK 02

Create a K-means clustering algorithm to group customers of a retail store based on their purchase history.

```python
# TASK 02 - KMeans Clustering (Dataset Created Inside Code)

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

# Create Mall Customer Dataset (200 customers)
np.random.seed(42)

data = pd.DataFrame({
    "CustomerID": range(1, 201),
    "Gender": np.random.choice(["Male", "Female"], 200),
    "Age": np.random.randint(18, 70, 200),
    "Annual Income (k$)": np.random.randint(15, 140, 200),
    "Spending Score (1-100)": np.random.randint(1, 100, 200)
})

# Select Features
X = data[['Annual Income (k$)', 'Spending Score (1-100)']]

# Standardize Data
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Elbow Method
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', random_state=42)
    kmeans.fit(X_scaled)
    wcss.append(kmeans.inertia_)

plt.figure()
plt.plot(range(1, 11), wcss)
plt.title("Elbow Method")
plt.xlabel("Number of Clusters")
```

```python
plt.ylabel("WCSS")
plt.show()

# Apply KMeans (K=5)
kmeans = KMeans(n_clusters=5, init='k-means++', random_state=42)
y_kmeans = kmeans.fit_predict(X_scaled)

# Add Cluster Column
data['Cluster'] = y_kmeans

# Print Output
print("\nClustered Data (First 10 Customers):")
print(data.head(10))

# Plot Clusters
plt.figure()
plt.scatter(X_scaled[y_kmeans == 0, 0], X_scaled[y_kmeans == 0, 1])
plt.scatter(X_scaled[y_kmeans == 1, 0], X_scaled[y_kmeans == 1, 1])
plt.scatter(X_scaled[y_kmeans == 2, 0], X_scaled[y_kmeans == 2, 1])
plt.scatter(X_scaled[y_kmeans == 3, 0], X_scaled[y_kmeans == 3, 1])
plt.scatter(X_scaled[y_kmeans == 4, 0], X_scaled[y_kmeans == 4, 1])
plt.title("Customer Segments")
plt.xlabel("Annual Income (Scaled)")
plt.ylabel("Spending Score (Scaled)")
plt.show()

# Print Cluster Centers
print("\nCluster Centers (Scaled Values):")
print(kmeans.cluster_centers_)
```
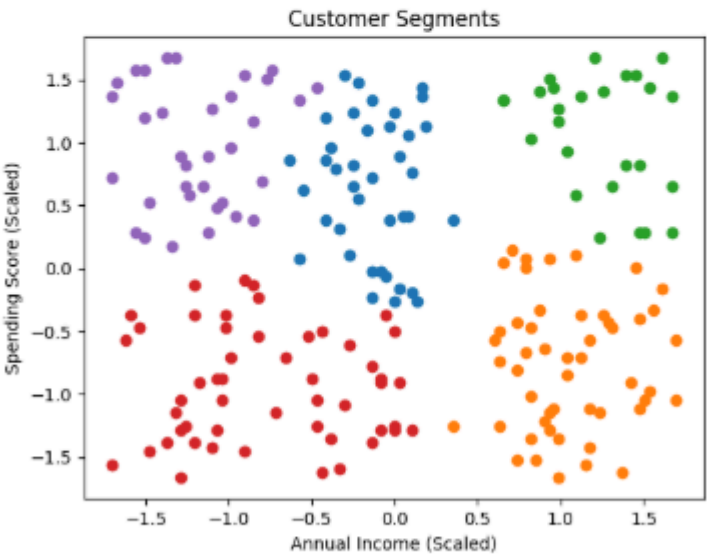
Output

```
Clustered Data (First 10 Customers):
   CustomerID  Gender  Age  Annual Income (k$)  Spending Score (1-100)  \
0           1    Male   49                  68                      86
1           2  Female   56                  22                      57
2           3    Male   66                  41                      29
3           4    Male   69                  41                      78
4           5    Male   49                 112                      92
5           6  Female   21                  35                      69
6           7    Male   47                  44                      47
7           8    Male   54                 111                      94
8           9    Male   40                  42                      62
9          10  Female   56                 125                      69

   Cluster
0        0
1        4
2        3
3        4
4        2
5        4
6        3
7        2
8        4
9        2
```


Customer Segments

```
Cluster Centers (Scaled Values):
[[-0.13455481  0.63951769]
 [ 1.0515019  -0.76715941]
 [ 1.23198176  1.07761523]
 [-0.76770428 -0.93550817]
 [-1.18253133  0.97381476]]
```