TASK 2

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

```
In [4]: import pandas as pd

# Load the data
data = pd.read_csv('/Users/manojt/Downloads/Mall_Customers (1).csv')

# Display the first few rows of the data
print(data.head())
```

```
   CustomerID  Gender  Age  Annual Income (k$)  Spending Score (1-10
0)
0           1    Male   19                  15
39
1           2    Male   21                  15
81
2           3  Female   20                  16
6
3           4  Female   23                  16
77
4           5  Female   31                  17
40
```

```
In [5]: # Drop non-numeric columns if necessary (e.g., 'CustomerID' and 'Gende
data_processed = data.drop(['CustomerID', 'Gender'], axis=1)

# Display the processed data
print(data_processed.head())
```

```
   Age  Annual Income (k$)  Spending Score (1-100)
0   19                  15                      39
1   21                  15                      81
2   20                  16                       6
3   23                  16                      77
4   31                  17                      40
```

```
In [6]: # The features for clustering will be 'Age', 'Annual Income (k$)', and
features = data_processed[['Age', 'Annual Income (k$)', 'Spending Scor

# Display the selected features
print(features.head())
```

```
   Age  Annual Income (k$)  Spending Score (1-100)
0   19                  15                      39
1   21                  15                      81
2   20                  16                       6
3   23                  16                      77
4   31                  17                      40
```

```python
In [7]: from sklearn.cluster import KMeans
        import matplotlib.pyplot as plt

        # Determine the optimal number of clusters using the Elbow Method
        wcss = []
        for i in range(1, 11):
            kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300, n_ir
            kmeans.fit(features)
            wcss.append(kmeans.inertia_)

        # Plot the Elbow Method graph
        plt.plot(range(1, 11), wcss)
        plt.title('Elbow Method')
        plt.xlabel('Number of clusters')
        plt.ylabel('WCSS')
        plt.show()
```
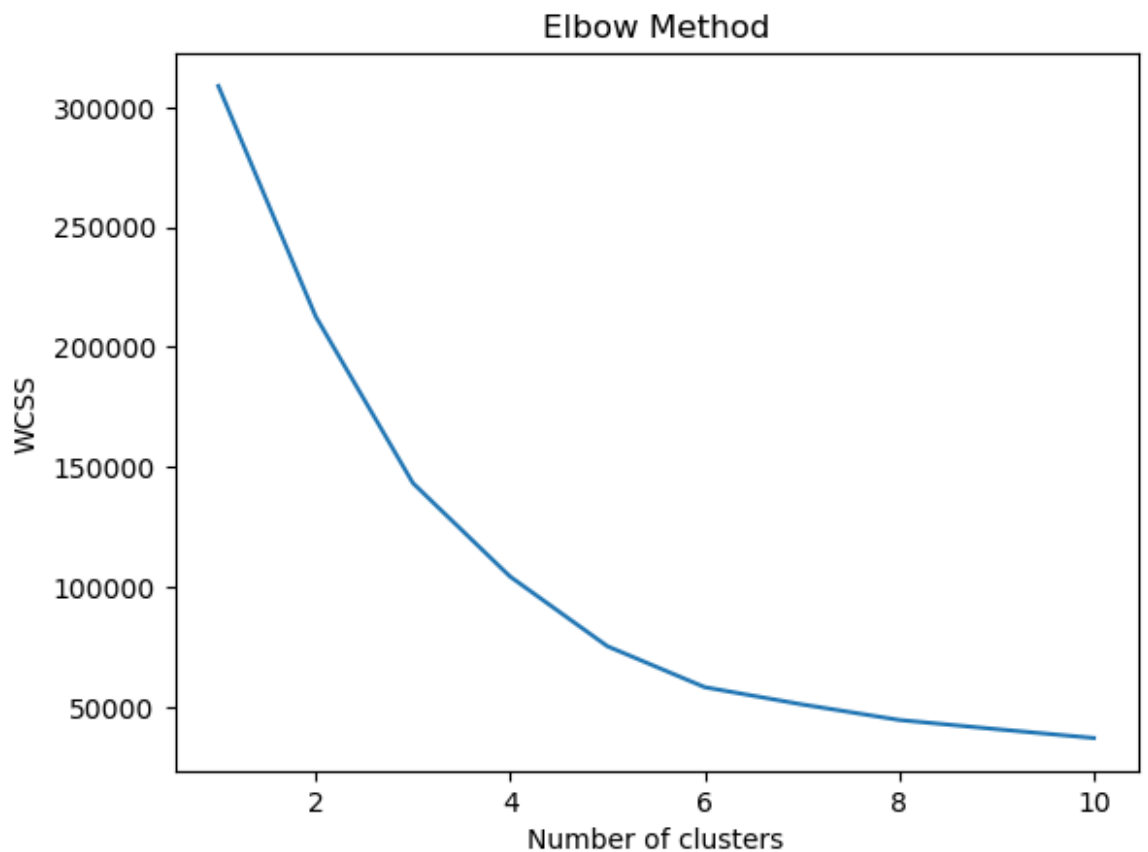
In [8]:
```python
# Fit the K-means model
optimal_clusters = 5
kmeans = KMeans(n_clusters=optimal_clusters, init='k-means++', max_ite
clusters = kmeans.fit_predict(features)

# Add the cluster labels to the original data
data['Cluster'] = clusters

# Display the data with cluster labels
print(data.head())
```

```
   CustomerID  Gender  Age  Annual Income (k$)  Spending Score (1-10
0)  \
0           1    Male   19                  15
39
1           2    Male   21                  15
81
2           3  Female   20                  16
6
3           4  Female   23                  16
77
4           5  Female   31                  17
40

   Cluster
0        0
1        4
2        0
3        4
4        0
```
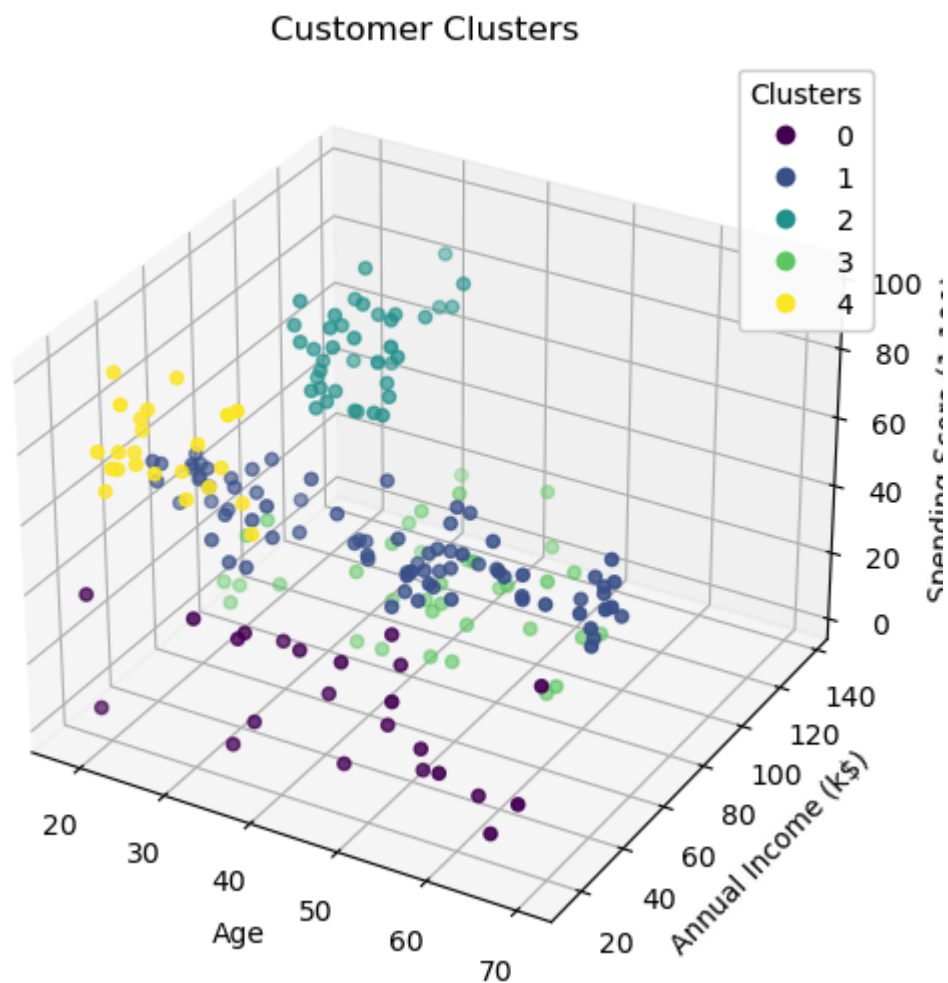
In [9]:
```python
from mpl_toolkits.mplot3d import Axes3D

# Create a 3D plot
fig = plt.figure(figsize=(8, 6))
ax = fig.add_subplot(111, projection='3d')
scatter = ax.scatter(features['Age'], features['Annual Income (k$)'],

# Set plot labels
ax.set_xlabel('Age')
ax.set_ylabel('Annual Income (k$)')
ax.set_zlabel('Spending Score (1-100)')
ax.set_title('Customer Clusters')

# Add legend
legend = ax.legend(*scatter.legend_elements(), title="Clusters")
ax.add_artist(legend)

plt.show()
```


In [ ]: