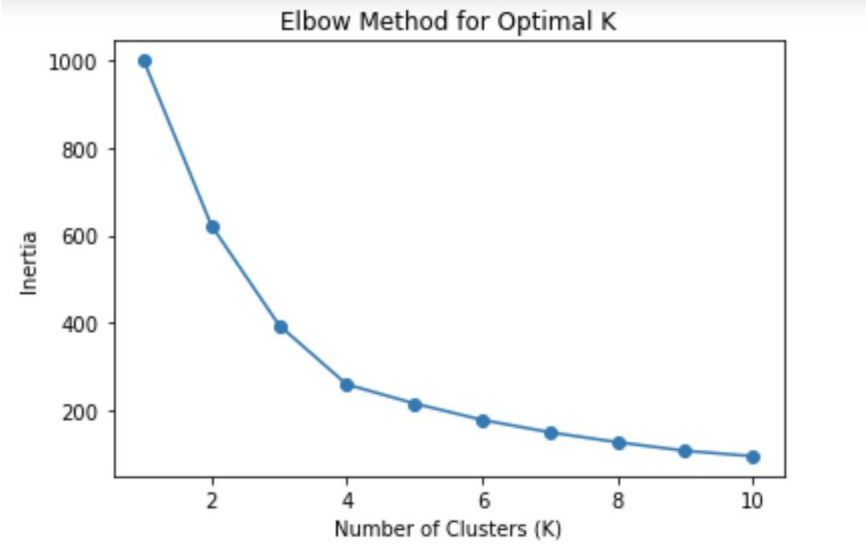
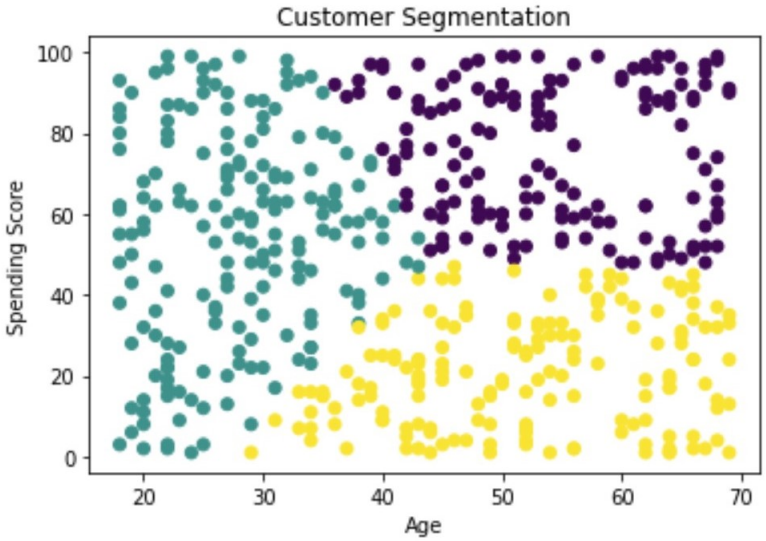
Project 2: Customer Segmentation Script

```
# Import necessary libraries
import pandas as pd
import numpy as np
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

# Generate random customer data for demonstration
np.random.seed(42)

# Number of customers
num_customers = 500

# Features: Spending Score and Age
spending_score = np.random.randint(1, 100, num_customers)
age = np.random.randint(18, 70, num_customers)

# Create a DataFrame
data = pd.DataFrame({'SpendingScore': spending_score, 'Age': age})

# Standardize the data
scaler = StandardScaler()
scaled_data = scaler.fit_transform(data)

# Determine the optimal number of clusters (K) using the Elbow method
inertia = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(scaled_data)
    inertia.append(kmeans.inertia_)

# Plot the Elbow method results
plt.plot(range(1, 11), inertia, marker='o')
plt.title('Elbow Method for Optimal K')
plt.xlabel('Number of Clusters (K)')
plt.ylabel('Inertia')
plt.show()

# Based on the Elbow method, let's choose K=3
optimal_k = 3

# Apply K-means clustering with the optimal number of clusters
kmeans = KMeans(n_clusters=optimal_k, random_state=42)
data['Cluster'] = kmeans.fit_predict(scaled_data)
```

```
# Visualize the clusters
plt.scatter(data['Age'], data['SpendingScore'], c=data['Cluster'], cmap='viridis')
plt.title('Customer Segmentation')
plt.xlabel('Age')
plt.ylabel('Spending Score')
plt.show()

# Save the segmented customer data to a CSV file
data.to_csv('customer_segmentation_data.csv', index=False)

# Print a summary message
print(f"Customer segmentation completed with {optimal_k} clusters.")
```

```
C:\Users\jilal\Anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:1036: UserWarning: KMe
  warnings.warn(
```






```
Customer segmentation completed with 3 clusters.
```

This script generates random customer data, performs K-means clustering, and visualizes the clusters. The resulting segmented customer data is saved to a CSV file named customer_segmentation_data.csv. You can replace the random data with your actual customer data for a real-world scenario.

The code performs customer segmentation using the K-means clustering algorithm and provides insights into the optimal number of clusters (K) using the Elbow method. Here's an interpretation of the key components:

1. Elbow Method:

- The first plot, known as the Elbow method, helps determine the optimal number of clusters (K). It shows the inertia (within-cluster sum of squares) for different values of K.
- The elbow point, where the rate of decrease in inertia slows down, indicates an optimal number of clusters.

2. Customer Segmentation:

- The second plot visualizes the clusters formed based on the chosen optimal number (K=3 in this case).
- Each point represents a customer, positioned in the space defined by 'Age' and 'Spending Score.' The color indicates the assigned cluster.

3. Summary Message:

- The code concludes by saving the segmented customer data to a CSV file and printing a summary message, indicating the completion of customer segmentation with the chosen number of clusters (3 in this case).

4. Insights:

- Analyzing the scatter plot can provide insights into distinct customer segments based on age and spending behavior.
- The Elbow method aids in choosing a reasonable number of clusters for meaningful segmentation.

Overall, the code provides a visual representation of customer segmentation and the optimal number of clusters for further analysis or targeted marketing strategies.

The output message "Customer segmentation completed with 3 clusters." indicates that the K-means clustering algorithm was applied, and the data was successfully segmented into three distinct clusters based on the chosen features (in this case, 'Age' and 'Spending Score'). The number of clusters (K=3) was determined as optimal using the Elbow method.

Additionally, the visualization graph accompanying this output likely displays a scatter plot of the customer data points in the 'Age' and 'Spending Score' space, with different colors representing the three clusters. This visual representation helps to observe the separation of customers into distinct groups, allowing for further analysis and targeted strategies for each segment.

Conclusion:

The completion of customer segmentation using the K-means clustering algorithm with three clusters suggests that distinct patterns exist in the dataset, grouping customers based on their age and spending score. This segmentation enables a more targeted approach in understanding and addressing the preferences or behaviors of different customer segments, fostering personalized strategies for improved customer satisfaction and business outcomes.