

```
In [5]: import pandas as pd
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

# Load the dataset
file_path = "E:\\serious\\archive (2)\\avocado.csv"
customer_data = pd.read_csv(file_path)

# Explore the data (optional)
print(customer_data.head())

# Assuming the data contains columns related to purchase behavior, you can filter/s
# For this example, we'll assume numerical features are all you need for clustering
X = customer_data.select_dtypes(include=['float64', 'int64'])

# Data scaling (standardize the data)
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Find the optimal number of clusters using the 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_)

# Plot the elbow graph
plt.plot(range(1, 11), wcss)
plt.title('Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()

# Apply K-Means with the optimal number of clusters (e.g., 3)
kmeans = KMeans(n_clusters=3, init='k-means++', random_state=42)
y_kmeans = kmeans.fit_predict(X_scaled)

# Add the cluster labels to the original dataset
customer_data['Cluster'] = y_kmeans

# Save the clustered data
customer_data.to_csv('customer_segmented.csv', index=False)

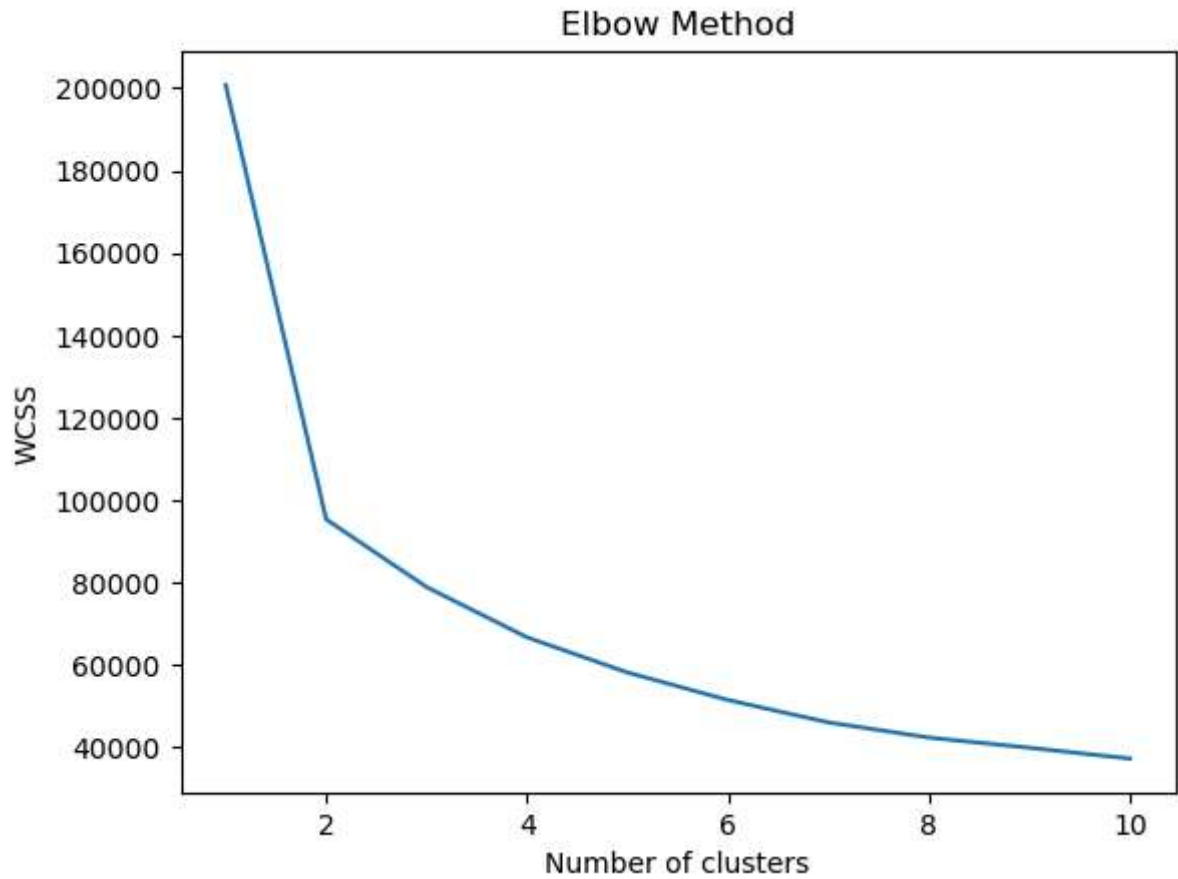
print("Clustering complete. Segmented data saved as 'customer_segmented.csv'.")
```

	Unnamed: 0	Date	AveragePrice	Total Volume	4046	4225	\
0	0	2015-12-27	1.33	64236.62	1036.74	54454.85	
1	1	2015-12-20	1.35	54876.98	674.28	44638.81	
2	2	2015-12-13	0.93	118220.22	794.70	109149.67	
3	3	2015-12-06	1.08	78992.15	1132.00	71976.41	
4	4	2015-11-29	1.28	51039.60	941.48	43838.39	

	4770	Total Bags	Small Bags	Large Bags	XLarge Bags	type	\
0	48.16	8696.87	8603.62	93.25	0.0	conventional	
1	58.33	9505.56	9408.07	97.49	0.0	conventional	
2	130.50	8145.35	8042.21	103.14	0.0	conventional	
3	72.58	5811.16	5677.40	133.76	0.0	conventional	
4	75.78	6183.95	5986.26	197.69	0.0	conventional	

	year	region
0	2015	Albany
1	2015	Albany
2	2015	Albany
3	2015	Albany
4	2015	Albany

[illegible]



```
C:\Users\samee\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 'auto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
```

```
warnings.warn(
```

```
Clustering complete. Segmented data saved as 'customer_segmented.csv'.
```