```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Load data
df = pd.read_csv('Mall_Customers.csv')
print(df.head())

# Key stats
print(df.describe())
```

```
       CustomerID  Gender  Age  Annual Income (k$)  Spending Score (1-100)
0               1    Male   19                  15                      39
1               2    Male   21                  15                      81
2               3  Female   20                  16                       6
3               4  Female   23                  16                      77
4               5  Female   31                  17                      40
           CustomerID         Age  Annual Income (k$)  Spending Score (1-100)
count      200.000000  200.000000          200.000000              200.000000
mean       100.500000   38.850000           60.560000               50.200000
std         57.879185   13.969007           26.264721               25.823522
min          1.000000   18.000000           15.000000                1.000000
25%         50.750000   28.750000           41.500000               34.750000
50%        100.500000   36.000000           61.500000               50.000000
75%        150.250000   49.000000           78.000000               73.000000
max        200.000000   70.000000          137.000000               99.000000
```

```python
from sklearn.preprocessing import StandardScaler

X = df[['Age', 'Annual Income (k$)', 'Spending Score (1-100)']]
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

```python
from sklearn.cluster import KMeans

# Find optimal k
wcss = []
for k in range(1, 11):
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(X_scaled)
    wcss.append(kmeans.inertia_)

# Plot elbow curve
plt.plot(range(1, 11), wcss, marker='o')
plt.xlabel('Number of Clusters (k)')
plt.ylabel('WCSS (Inertia)')
plt.title('Elbow Method')
plt.show()
```
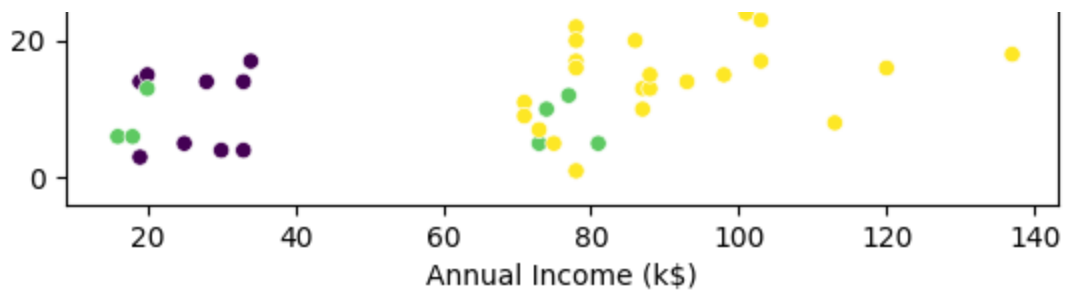
Elbow Method
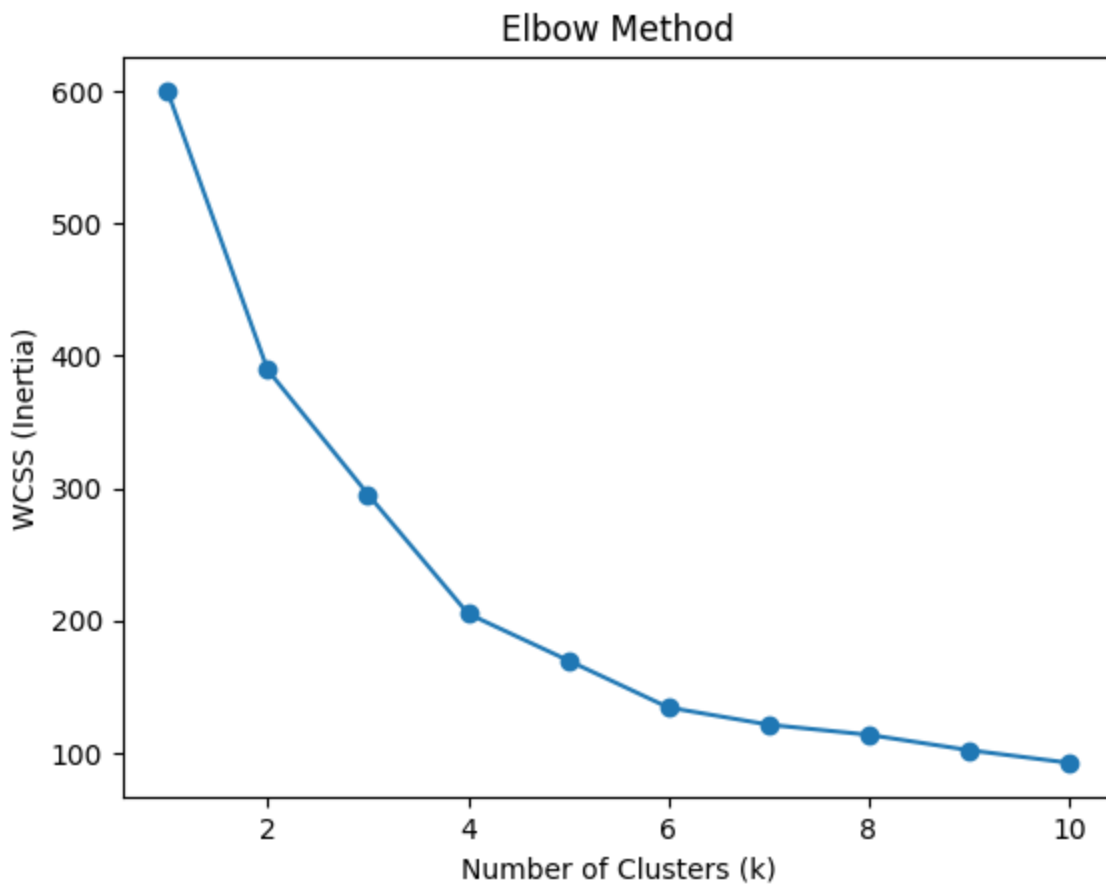

```
kmeans = KMeans(n_clusters=5, random_state=42)
clusters = kmeans.fit_predict(X_scaled)
df['Cluster'] = clusters

# Visualize clusters
sns.scatterplot(data=df, x='Annual Income (k$)', y='Spending Score (1-100)', hue='Cluster
plt.title('Customer Segments')
plt.show()
```

```
plt.figure(figsize=(10, 6))
sns.boxplot(data=df, x='Cluster', y='Spending Score (1-100)')
plt.title('Spending Habits by Cluster')
plt.show()
```