Customer Segmentation Analysis with K-Means Clustering

Objective

The goal of this project was to segment retail customers into distinct groups based on their demographic and behavioral patterns, such as age, annual income, and spending score.
These insights can help businesses create targeted marketing strategies and improve customer retention.

Dataset

Mall Customers Dataset

200 customer records

Features:

- CustomerID
- Gender
- Age
- Annual Income (k$)
- Spending Score (1100)

Source: Kaggle (https://www.kaggle.com/datasets/vjchoudhary7/customer-segmentation-tutorial-in-python)

Tools Used

Python

pandas

scikit-learn

matplotlib

seaborn

Methodology

Loaded and explored the dataset to check for missing or inconsistent values.

Selected relevant numerical features: Age, Annual Income, Spending Score.

Scaled the features using StandardScaler for better clustering performance.

Used the Elbow Method to determine the optimal number of clusters.

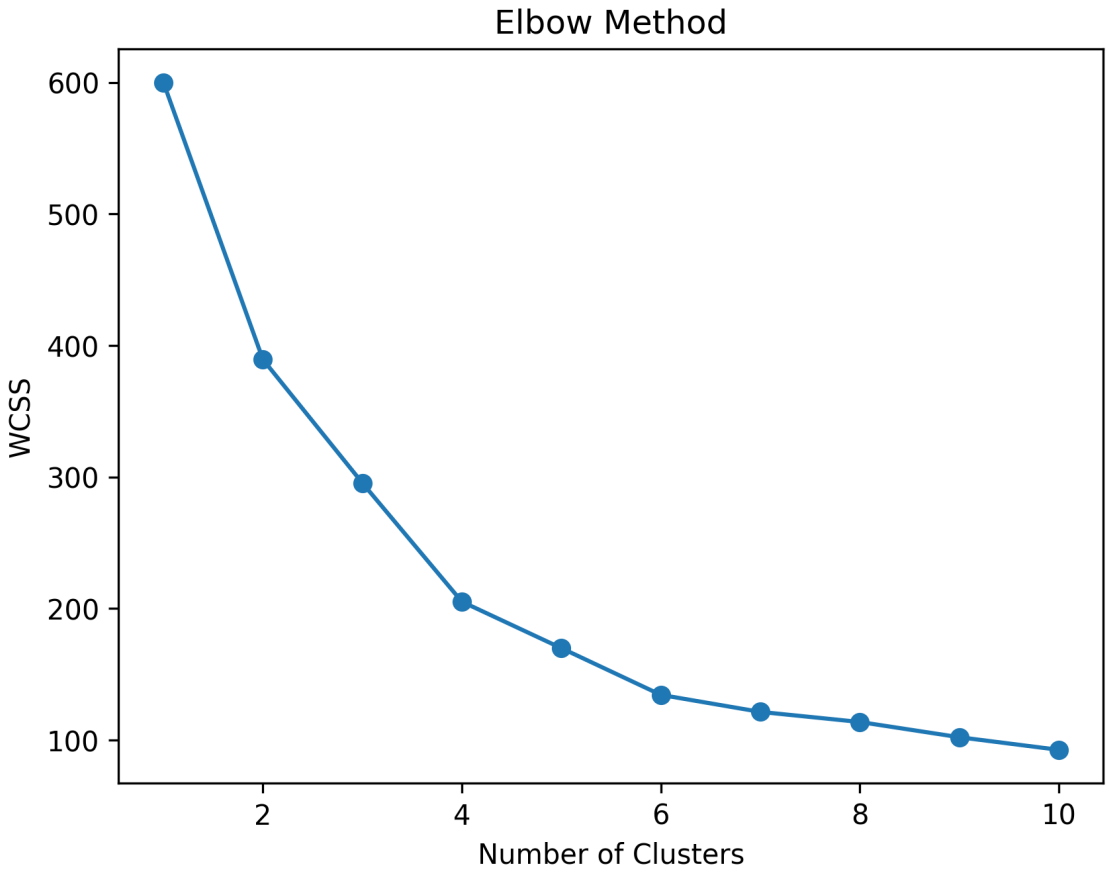
Applied K-Means Clustering with 5 clusters.

Analyzed the resulting segments using group statistics and visualizations.

Results

Optimal Number of Clusters

The Elbow Method indicated that 5 clusters achieved a good balance between WCSS (within-cluster sum of squares) and model complexity.



Cluster Analysis

Below are the average values for each cluster:

(Refer to the notebook or table output for exact values)

Sample insights:

Cluster 0: High-income, high-spending customers - ideal for premium offers.

Cluster 1: Moderate-income, high-spending customers - potential upselling opportunities.

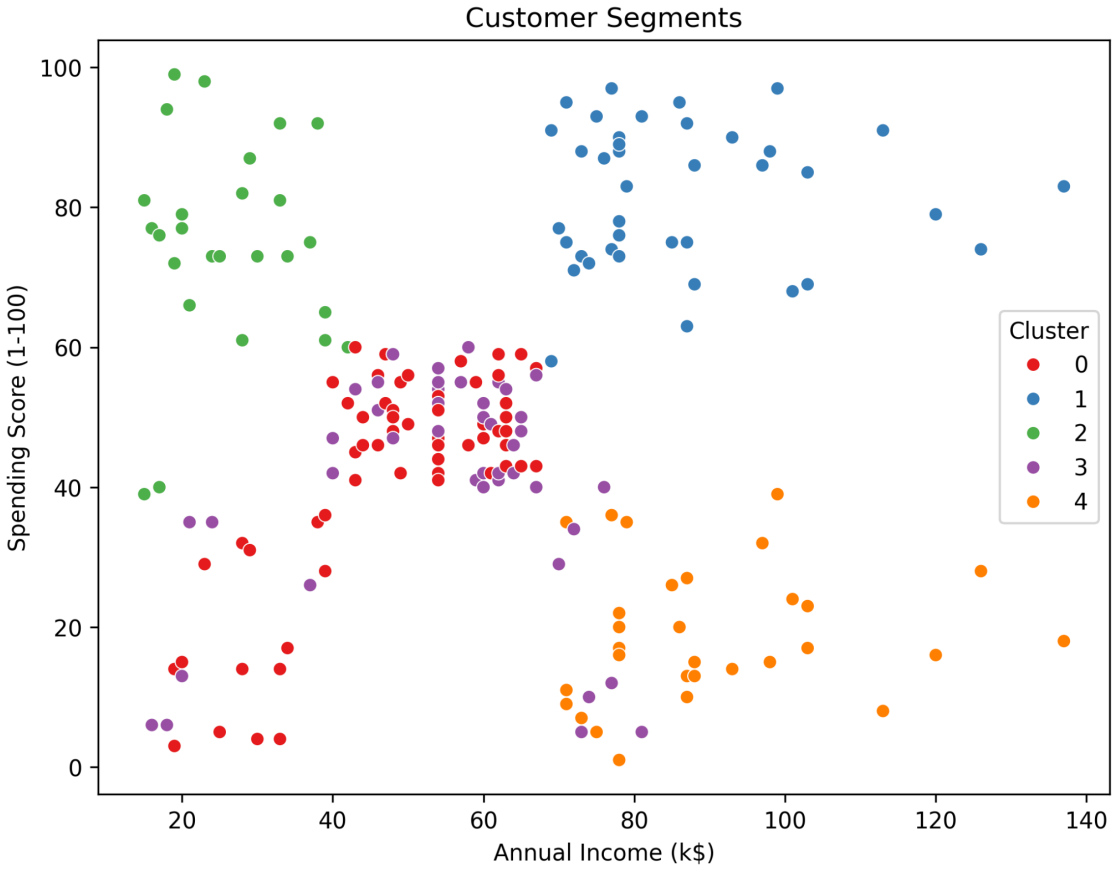
Cluster 2: Low-income, low-spending customers - target with cost-effective solutions.

Cluster 3: Younger, moderate-spending customers - high lifetime value potential.

Cluster 4: Older, frugal customers - maintain loyalty through personalized engagement.

Customer Segments Visualization

The following scatterplot illustrates the segmentation based on Annual Income and Spending Score:



Recommendations

Focus premium marketing efforts on Clusters 0 and 1.

Develop loyalty programs and discounts for Cluster 2.

Upsell and retain Cluster 3 customers with tailored offerings.

Conclusion

The customer segmentation project successfully identified five distinct customer groups with actionable insights for targeted marketing. Businesses can leverage this segmentation to maximize ROI and improve customer satisfaction.

Notes

This analysis is reproducible and available as a Jupyter Notebook. Please contact me if you'd like to view the notebook or discuss further improvements.