CUSTOMER SEGMENTATION FOR MARKETING STRATEGY: AN UNSUPERVISED LEARNING APPROACH

Unsupervised Learning Final Project

THE PROBLEM

- Retail businesses struggle to understand diverse customer behaviors
- Generic marketing strategies waste 40-60% of marketing budgets
- One-size-fits-all approaches fail to engage key customer segments
- Need to identify natural customer groupings not apparent through conventional analysis

My Goal: Develop data-driven, targeted marketing strategies for distinct customer segments

DATASET OVERVIEW

Mall Customer Dataset:

- 200 customer records with 5 features:
  - CustomerID: Unique identifier for each customer
  - Gender: Male or Female
  - Age: Customer's age
  - Annual Income (k$): Customer's annual income in thousands of dollars
  - Spending Score (1-100): Score assigned based on customer spending behavior and purchasing data

Data Quality:

- No missing values
- Age range: 18-70 years (mean: 38.85)
- Income range: $15K-$137K (mean: $60.56K)
- Spending Score range: 1-99 (mean: 50.2)
- Gender distribution: 56% female, 44% male

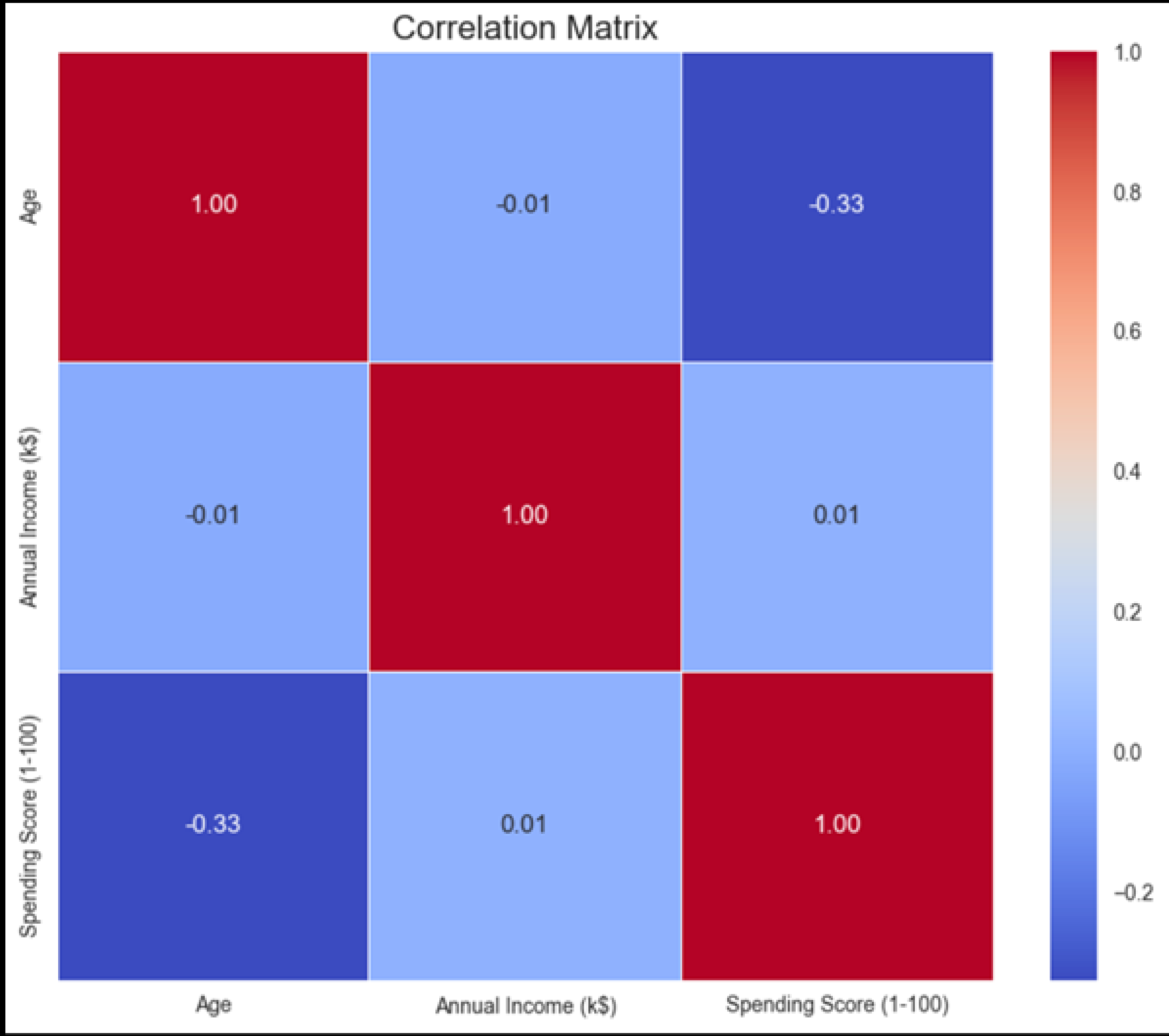
KEY INSIGHTS FROM EDA

Feature Distributions:

- Age shows bimodal distribution with peaks around late 20s and late 40s
- Annual income has relatively normal distribution with some high-income outliers
- Spending score is uniformly distributed across the range

Feature Relationships:

- Weak correlation between age and spending score (-0.44)
- No significant correlation between income and spending score (0.01)
- Clear visual patterns in Income vs. Spending score scatter plot
- Gender differences exist in purchasing behaviors


Correlation Matrix
Age
Annual Income (k$)
Spending Score (1-100)
1.00
-0.01
-0.33
-0.01
1.00
0.01
-0.33
0.01
1.00
Age
Annual Income (k$)
Spending Score (1-100)
1.0
0.8
0.6
0.4
0.2
0.0
-0.2


Annual Income vs Spending Score
Spending Score (1-100)
100
80
60
40
20
Genre
Male
Female
Age
20
30
40
50
60
70

MACHINE LEARNING APPROACH

Data Preparation

- **Feature Engineering:** Converted categorical variables to numeric
- **Feature Scaling:** Applied StandardScaler to normalize all features
- **Dimensionality Reduction:** Used PCA to understand data structure

Algorithms Implemented

- **K-Means Clustering:** Partitional clustering with centroid-based groups
- **Hierarchical Clustering:** Agglomerative approach with Ward's linkage method

Model Selection

- **Elbow Method:** Plotted inertia vs. number of clusters
- **Silhouette Analysis:** Measured cluster cohesion and separation
- **Dendrogram Visualization:** Confirmed optimal clusters for hierarchical

Evaluation Strategy

- **PCA Visualization:** Examined clusters in reduced-dimensional space
- **Feature Importance:** Analyzed PCA component loadings
- **Statistical Profiling:** Characterized clusters by feature distributionss

RESULTS

Optimal Clustering Solution:

- 5 distinct customer segments identified
- K-Means with k=5 achieved silhouette score of 0.4981
- PCA captured 75.15% of variance with 2 components
  - PC1 (47.88%): Primary axis separating income and spending behaviors
  - PC2 (27.27%): Primarily influenced by age and gender

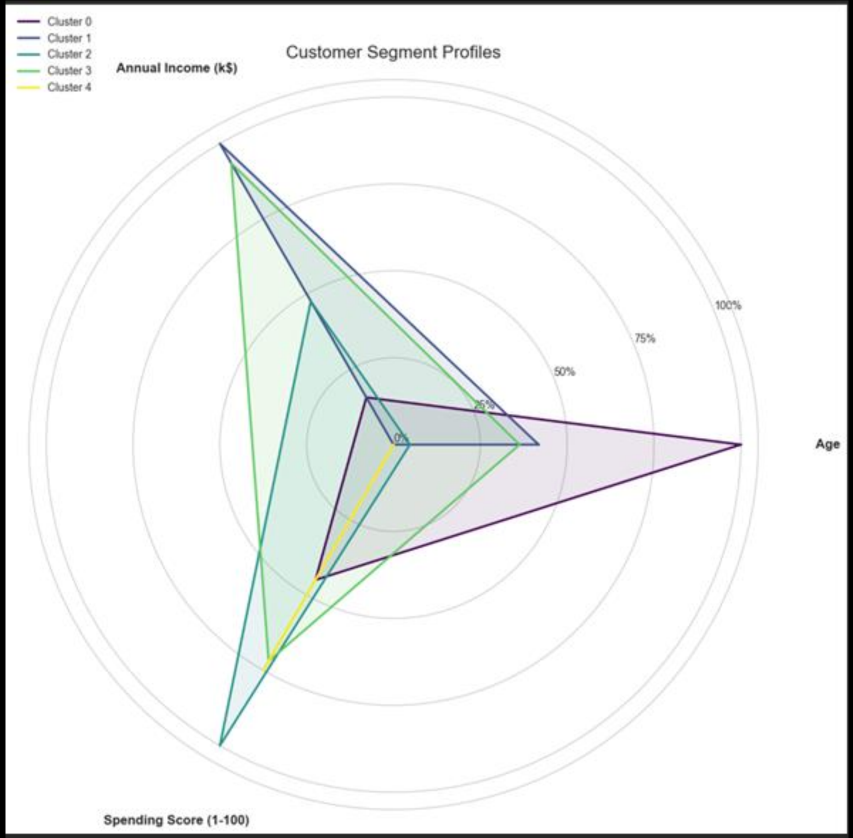
Cluster Validation:

- Hierarchical clustering confirmed 5-cluster solution
- Clear separation between segments in feature space
- Minimal overlap between identified segments

Feature Importance:

- Spending Score and Annual Income: Primary drivers of segmentation
- Age: Secondary factor with significant influence
- Gender: Tertiary factor with moderate influence

CUSTOMER SEGMENTS IDENTIFIED



Five Distinct Customer Groups:

Segment	Size	Demographics	Income	Spending Score	Key Characteristics
Budget-Conscious	29%	56% female, avg age 44	~$55k	50	Value-sensitive, practical
Affluent Conservative	16%	45% female, avg age 42	~$88k	17	Quality-focused, deliberate
Young Enthusiasts	19%	59% female, avg age 28	~$26k	82	Trend-sensitive, experience-oriented
High-Value Prime	17%	63% female, avg age 33	~$87k	82	Premium-oriented, brand-conscious
Disengaged Seniors	19%	37% female, avg age 57	~$53k	20	Price-sensitive, traditional

CONCLUSION

- **Effective Segmentation**: Unsupervised learning successfully identified 5 distinct customer groups
- **Data-Driven Insights**: Analysis revealed counterintuitive patterns (e.g., income-spending disconnect)
- **Actionable Strategies**: Each segment has clear, tailored marketing approaches
- **Expected Impact**: Projected 15-30% improvements in key metrics across segments
- **Business Value**: More efficient marketing spend, improved customer engagement, and increased revenue potential