# **Customer Segmentation Analysis Report**

## **Executive Summary**

This report details a customer segmentation study conducted on mall customer data using the K-Means clustering algorithm. The objective was to identify distinct customer groups based on Annual Income and Spending Score, allowing for the development of tailored marketing strategies.

## **1. Data Collection and Preprocessing**

### **1.1 Dataset Overview**

* Dataset Used: Mall\_Customers.csv
* Key Features Considered:
  + Annual Income (k$)
  + Spending Score (1–100)

### **1.2 Data Cleaning and Preparation**

* Outliers removed using the IQR method
* Boxplots validated outlier removal effectiveness
* StandardScaler applied for feature normalization

## **2. Exploratory Data Analysis**

### **2.1 Data Distribution**

* Boxplots visualized the distributions and identified skewness
* Post-cleaning data showed improved balance for clustering

## **3. Clustering Analysis**

### **3.1 Determining Optimal Clusters**

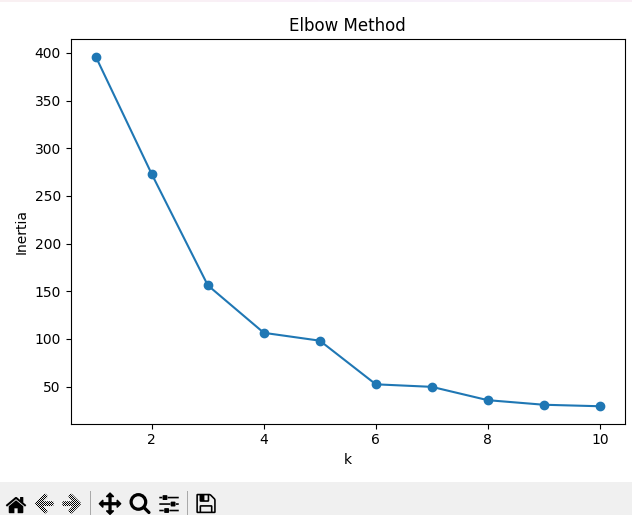
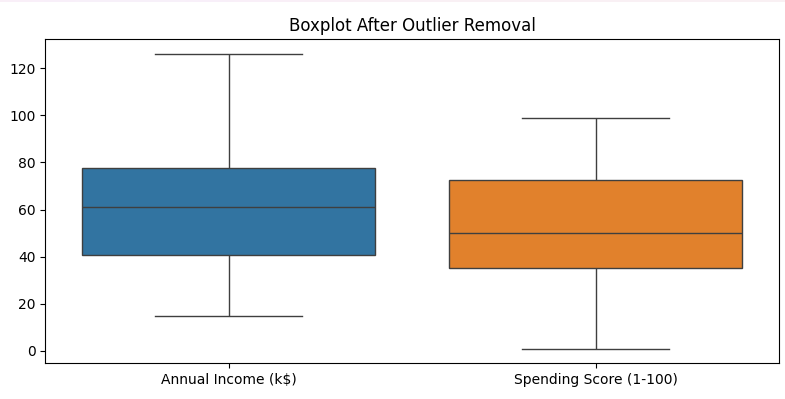
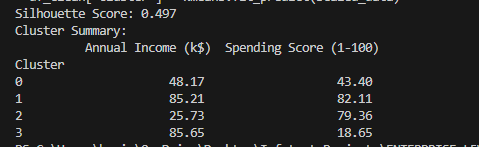
* Elbow Method applied (k = 1 to 10) and Optimal clusters: k = 4
* Silhouette Score achieved:0.497

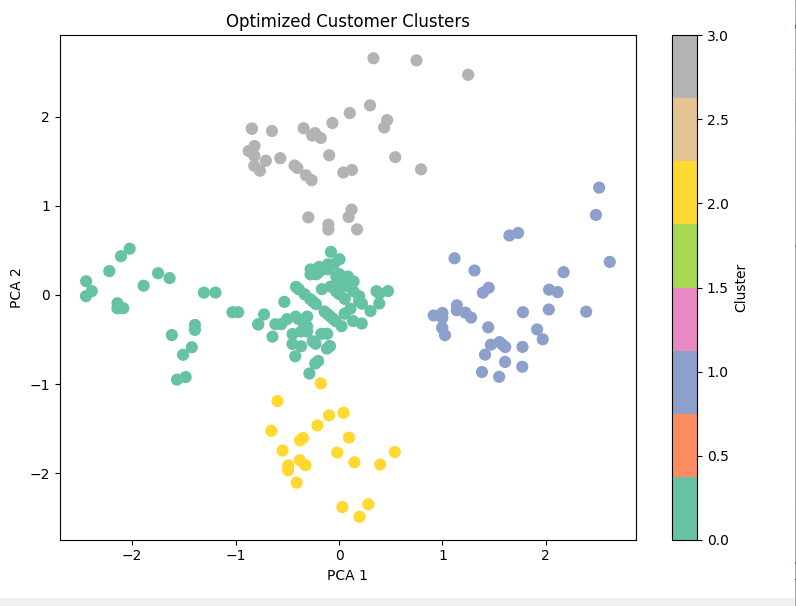
### **3.2 Visualization**

* Applied PCA for dimensionality reduction
* Scatter plot depicted distinct cluster separations
* Color-coded cluster representation ensures easy interpretation

1. **Cluster Profiles**

**4.1 Summary**





### **4.2 Marketing Strategies per Cluster**

#### Cluster 1: High Income – High Spending

* Strategy: VIP programs, premium services, invite-only events

#### Cluster 2: High Income – Low Spending

* Strategy: Personalized product education, loyalty nudges

#### Cluster 3: Low Income – High Spending

* Strategy: Discount bundles, EMI schemes, flash sales

#### Cluster 4: Low Income – Low Spending

* Strategy: Budget-friendly products, clearance deals

## **5. Technical Implementation**

### **5.1 Methodology**

* K-Means Clustering (k=4)
* StandardScaler for feature scaling
* PCA for visualization
* Silhouette Score for validation

### **5.2 Generated Files**

* optimized\_cluster\_summary.csv: Per-cluster stats
* optimized\_clustered\_customers.csv: Customers with assigned clusters

## **6. Conclusion**

The clustering approach effectively segmented customers into four distinct behavioral groups. These insights can drive data-driven marketing, inventory decisions, and strategic planning.

## **7. Future Work**

### **7.1 Data Enhancement**

* Add features like Age, Gender, and Purchase History
* Introduce temporal analysis (e.g., seasonal trends)

### **7.2 Model Advancements**

* Experiment with DBSCAN, Hierarchical Clustering
* Build a real-time dynamic clustering system

### **7.3 Marketing Deployment**

* Launch segment-specific campaigns
* Track KPIs and iteratively refine clusters

Prepared by

Allu Harini