🛒 Customer Segmentation for E-Commerce

Author: Kavya

📧 maherarakavya@gmail.com

🌐 GitHub Profile

🎯 Objective

The purpose of this analysis is to segment e-commerce customers based on their shopping behavior, demographics, and satisfaction levels.

By identifying groups of customers who share similar characteristics, the goal is to help the business create **targeted marketing strategies**, **personalized offers**, and **improve customer retention**.

Customer segmentation is essential for every online business.

Not all customers behave the same — some are frequent buyers, others are price-sensitive, and some may purchase only during sales.

This study uses **K-Means Clustering** to uncover these hidden customer patterns.

🧩 Dataset Overview

The dataset includes customer information such as:

- **Age**
- **Purchase Amount (USD)**
- **Previous Purchases**
- **Review Rating**
- Plus encoded categorical features such as **Gender**, **Category**, **Subscription Status**, **Shipping Type**, and **Payment Method**.

These features together provide a holistic view of how customers behave — from how much they spend to how satisfied they feel about their purchases.

⚙️ Methodology

1. Data Preparation

- The dataset was cleaned and encoded to ensure all features were numerical.
- Missing values were handled appropriately.
- All numerical features were standardized using **StandardScaler** to bring them to a common scale.

2. K-Means Clustering

- The **Elbow Method** was applied to determine the optimal number of clusters by observing where the Within-Cluster Sum of Squares (WCSS) started to flatten.

- The result indicated that **4 clusters** provided the best segmentation — giving the right balance between interpretability and accuracy.

3. Dimensionality Reduction (PCA)

- To visualize high-dimensional data, **Principal Component Analysis (PCA)** was used.
- The data was reduced to two components (PC1 and PC2), which made it possible to visualize the clusters clearly in a 2D plot.

4. Visualization

- **Elbow Plot** → To select the optimal number of clusters.
- **PCA Scatter Plot** → To visualize how customers are grouped.
- **Heatmap of Cluster Centers** → To understand what makes each cluster unique.

📊 Analysis & Insights

1. Determining the Optimal Number of Clusters (Elbow Method)

The Elbow Method revealed that **four clusters** offer the best structure for segmentation.

The curve dropped sharply until four clusters, after which it began to flatten.

This means creating more clusters would only make the model complex without providing additional meaningful insight.

Each cluster represents a **distinct customer group**, such as loyal spenders, price-sensitive buyers, first-time customers, and inactive users.

This forms the foundation for a behavior-based marketing strategy.

---

2. Visualizing Customer Groups (PCA Plot)

The PCA scatter plot visually confirmed that the customer data naturally forms **four well-separated clusters**.

Each colored region in the plot represents customers with similar behavioral patterns.

The clusters revealed the following behavioral types:

- 🏆 **Cluster 0 — Loyal High Spenders:**

  Customers who frequently buy high-value items and give good reviews.

  They are consistent and show strong brand loyalty.

- 💸 **Cluster 1 — Price-Sensitive Buyers:**

  Customers who make moderate purchases and are highly responsive to discounts or offers.

- 🌱 **Cluster 2 — New or Deal-Driven Customers:**

  Young users making high-value purchases for the first time, often influenced by promotional campaigns.

- 💤 **Cluster 3 — Low-Engagement Customers:**

  Older or low-spending customers with fewer purchases and lower satisfaction.

This separation validates that the chosen features — **Age, Purchase Amount, Previous Purchases, and Review Rating** — effectively captured distinct customer behaviors.

---

3. Understanding Behavioral Differences (Cluster Centers Heatmap)

The heatmap reveals how these four groups differ numerically:

Cluster	Age	Purchase Amount (USD)	Previous Purchases	Review Rating	Key Behavior
0	52	78	29	4.3	Loyal high spenders

1	35	39	22	4.2	Moderate buyers
2	32	77	20	3.3	New, deal-driven shoppers
3	55	47	29	3.2	Older, low-engagement users

Key Patterns Identified:

- **Cluster 0:** Customers who trust the brand and buy regularly — strong candidates for loyalty programs.
- **Cluster 1:** Balanced customers who are consistent but may need small incentives to spend more.
- **Cluster 2:** New or impulsive buyers — opportunities for re-targeting to encourage future purchases.
- **Cluster 3:** Older, disengaged users — likely to churn without reactivation efforts.

This analysis highlights how spending behavior, satisfaction, and customer age vary across segments.

💼 Business Implications

Understanding these four customer segments gives the e-commerce company a strategic edge:

1. **Personalized Marketing:**
    - Send targeted offers to each group instead of generic campaigns.
    - Example: Discounts for price-sensitive buyers; premium rewards for loyal customers.

2. **Retention Strategy:**
    - Use loyalty programs to keep Cluster 0 engaged.
    - Offer reactivation emails or discounts to Cluster 3.
3. **Customer Lifetime Value (CLV):**
    - High-value clusters (like 0 and 2) should be nurtured to increase long-term profitability.
4. **Product & Pricing Strategy:**
    - Identify what kind of products attract each cluster and adjust pricing or bundling accordingly.

🧭 Conclusion

This analysis successfully segmented customers into **four meaningful groups**, each with unique behaviors and preferences.

By applying **unsupervised learning (K-Means)** and **dimensionality reduction (PCA)**, the project uncovers patterns that traditional reports might overlook.

The findings provide a strong base for **data-driven decision-making** in e-commerce — enabling:

- Smarter marketing campaigns
- Improved retention
- Higher customer satisfaction
- And ultimately, better revenue growth

🧰 Tools & Libraries

Python · Pandas · NumPy · Scikit-learn · Seaborn · Matplotlib

---