**Executive Summary**

Acme Inc. faces challenges in targeting customers effectively, leading to **low conversion rates and inefficient marketing spending**. To address this, we conducted a **customer segmentation analysis** using clustering techniques. This report details our approach, findings, and recommendations for **personalized marketing strategies** based on customer behavior. To address the challenge, we:

* Conducted an in-depth data analysis to extract insights.
* Evaluated multiple models and conducted market research to determine the optimal segmentation approach.
* Tested different clustering models, including:
* Fine-tuned the GMM model to produce the best result.
* Developed a prototype recommender system based on the segmentation.

**Business Statement**

Acme Inc.'s current **marketing approach lacks precision**, resulting in **wasted resources and low ROI**. Our goal is to identify distinct **customer segments** to optimize targeted marketing efforts.

**Data Overview and EDA**

**Data Glossary**

There are 3900 entries and 18 columns.

**Customer ID**: A unique identifier assigned to each individual customer, facilitating tracking and analysis of their shopping behavior over time.  
**Age**: The age of the customer, providing demographic information for segmentation and targeted marketing strategies.  
**Gender**: The gender identification of the customer, a key demographic variable influencing product preferences and purchasing patterns.  
Item Purchased: The specific product or item selected by the customer during the transaction.  
**Category**: The broad classification or group to which the purchased item belongs (e.g., clothing, electronics, groceries).  
**Purchase Amount (USD)**: The monetary value of the transaction, denoted in United States Dollars (USD), indicates the cost of the purchased item(s).  
Location: The geographical location where the purchase was made, offering insights into regional preferences and market trends.  
**Size**: The size specification (if applicable) of the purchased item, relevant for apparel, footwear, and certain consumer goods.  
**Color**: The color variant or choice associated with the purchased item, influencing customer preferences and product availability.  
**Season**: The seasonal relevance of the purchased item (e.g., spring, summer, fall, winter), impacting inventory management and marketing strategies.  
**Review Rating**: A numerical or qualitative assessment provided by the customer regarding their satisfaction with the purchased item.  
**Subscription Status**: Indicates whether the customer has opted for a subscription service, offering insights into their level of loyalty and potential for recurring revenue.  
**Shipping Type**: Specifies the method used to deliver the purchased item (e.g., standard shipping, express delivery), influencing delivery times and costs.  
**Discount Applied**: Indicates if any promotional discounts were applied to the purchase, shedding light on price sensitivity and promotion effectiveness.  
**Promo Code Used**: Notes whether a promotional code or coupon was utilized during the transaction, aiding in the evaluation of marketing campaign success.  
**Previous Purchases**: Days passed from the last purchase  
**Payment Method**: Specifies the mode of payment employed by the customer (e.g., credit card, cash), offering insights into preferred payment options.  
**Frequency of Purchases**: Indicates how frequently a customer purchases items.

**Numerical Variables Analysis**

|  | **count** | **mean** | **std** | **min** | **25%** | **50%** | **75%** | **max** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Customer ID** | 3900 | 1950.50 | 1125.98 | 1.0 | 975.75 | 1950.5 | 2925.25 | 3900.0 |
| **Age** | 3900 | 44.07 | 15.21 | 18.0 | 31.00 | 44.0 | 57.00 | 70.0 |
| **Purchase Amount (USD)** | 3900 | 59.76 | 23.69 | 20.0 | 39.00 | 60.0 | 81.00 | 100.0 |
| **Review Rating** | 3900 | 3.75 | 0.72 | 2.5 | 3.10 | 3.7 | 4.40 | 5.0 |
| **Previous Purchases** | 3900 | 25.35 | 14.45 | 1.0 | 13.00 | 25.0 | 38.00 | 50.0 |

The distribution of **Age, Previous Purchases, and Purchase Amount (USD) appears uniform,** suggesting a balanced dataset. A graph with different colored squares

AI-generated content may be incorrect.

Figure 1The distribution of Age, Purchase Amount (USD), Previous Purchases

**Categorical Variables**

* **Gender**: Disproportionately male. It's surprising because the top items purchased are associated with women.
* **Category**: The most popular category is clothing.
* **Size**: Size 'M' is the most popular.
* **Season**: Evenly distributed.
* **Subscription Status**: Most customers don't have subscriptions.
* **Discount Applied and Promo Code Used**: They are identical. We don't need one or another.
* **Payment Method**: There isn't any pattern.
* **Frequency of Purchases**: Every three Months seems to be the most frequent. But this is misleading because every three Months and Quarterly are essentially the same. The same goes for Bi-weekly and Fortnightly. With that consideration, Quarterly is the most frequent value, followed by Bi-weekly.

A chart of different colored squares

AI-generated content may be incorrect.

Figure 2Distribution of categorical variables

* **Item Purchased**: Blouses, Jewelry, and Pants are the most popular items.
* **Location**: Montana is the most frequent value, which is surprising because Montana is a relatively small state. The data might have been sampled with strata.

A graph of different colored bars

AI-generated content may be incorrect.

Figure 3Distribution of Item Purchased, and Location

**Feature Engineering**

1. **Frequency of Purchase**: This indicates how frequently a customer shops on the website. We turned this into a numerical scale per year. For instance, Weekly is 52, as there are 52 weeks in a year. In the same manner, the Monthly would be 12.
2. **Loyalty\_Score**: A combination of metrics. Frequency + Recency + Subscription\_Status \* 10.

**Models Tested**

**Model Selected**

A graph of different colored bars

AI-generated content may be incorrect.

Figure 4Silhouette Score Results

Based on the silhouette scores, we concluded that the clusters created with GMM clustering present the best results. The following are the characteristics of each cluster.A diagram of a group of boxes

AI-generated content may be incorrect.

A chart of a cluster of data

AI-generated content may be incorrect.

A diagram of a distribution across clusters

AI-generated content may be incorrect.

A chart of different colored squares

AI-generated content may be incorrect.

**Cluster Summary**

**Cluster 0: Premium High-Margin Shoppers**

Average Spending: $82 per transaction (82% of maximum possible spend)  
Purchase pattern: Buy expensive items approximately once every 25-26 days  
Engagement: Exceptionally high (4.3/5.0) - actively interacting with brand touchpoints  
Business value: Likely responsible for 30-40% of total profit despite lower purchase frequency  
Real-world example: Luxury shoppers who carefully select high-end items rather than making frequent purchases

**Cluster 1: Steady Mid-Tier Customers**

Average Spending: $57 per transaction  
Purchase pattern: Shop approximately every 2-3 weeks (15 times per period)  
Recency: Last purchased about 30 days ago  
Engagement: Good but not exceptional (3.7/5.0)  
Real-world example: Regular customers who reliably purchase mid-priced items on a predictable schedule

**Cluster 2: Enthusiast Power Users**

Average Spending: $68 per transaction  
Purchase pattern: Most frequent shoppers at 38 purchases (approaching weekly visits)  
Recency: Longest time since last purchase (39 days) - potential risk of churn  
Engagement: Very high (4.2/5.0) - brand enthusiasts  
Real-world example: Passionate hobbyists or product category enthusiasts who frequently buy related items

**Cluster 3: Recent New/Reactivated Customers**

Average Spending: $51 per transaction  
Purchase pattern: Infrequent (8-9 purchases total)  
Recency: Most recent shoppers (purchased within the last 13 days)  
Engagement: Moderate (3.4/5.0)  
Real-world example: Newly acquired customers or recently reactivated dormant customers who haven't established a regular pattern yet

**Cluster 4: Budget-Conscious Frequent Shoppers**

Average Spending: $40 per transaction (lowest spending group)  
Purchase pattern: Very frequent (35 purchases) - about every 10 days  
Recency: Moderately recent (29 days since last purchase)  
Engagement: Lowest of all groups (3.2/5.0)  
Real-world example: Price-sensitive frequent shoppers who make smaller purchases, likely deal-hunters or necessity-based shoppers

**Conclusion / Recommendation**

During the exercise, we evaluated multiple customer segmentation approaches. Our model produced five distinctly different segments: Premium High-Margin Shoppers, Steady Mid-Tier Customers, Enthusiastic Power Users, recent new/Reactivated Customers, and Budget-Conscious Frequent Shoppers.

|  |  |  |
| --- | --- | --- |
| Segment | Description | Actionable Recommendations |
| Premium High-Margin Shoppers | High-spending, frequent, highly selective shopper | Exclusive product offers and VIP Perks |
| Enthusiast Power Users | Most frequent shopper with medium spending, potential risk of churn | More personalized marketing through recommender system |
| Steady Mid-Tier Customers | Frequent shopping with lower spending | Loyalty incentives |
| Recent New/Reactivated Customers | Recently joined, low engagement | Automated welcome sequences with vouchers |
| Budget-Conscious Frequent Shoppers | Low engagement / bang per buck | Discount driven promotions |

With the new market segmentation, we can implement a highly personalized recommender system, which will improve campaign efficiency and marketing ROI.

A screenshot of a product recommendation service

AI-generated content may be incorrect.

Figure 5 Recommender System Prototype