**Customer Segmentation Analysis Report**

**Objective**

The goal of our group was to conduct an analysis which would segment the customers from the Women’s Clothing E-Commerce Reviews dataset using various clustering techniques. The focus was primarily to understand the customer behaviors based on age, rating, and positive feedback count and then combined it with categorical features such as division, department, and class. Thus, based on the identified clusters, we proposed specific targeted marketing strategies for each customer group to benefit the store.

**DataPreprocessing:** We started by cleaning the data, dropping rows with missing values in important columns such as "Title," "Review Text," "Division Name," "Department Name," and "Class Name." We also removed irrelevant columns.

**Feature Engineering**:

Numerical Features: Age, Rating, and Positive Feedback Count had been standardized the usage of StandardScaler.

Categorical Features: OneHotEncoder became used to at least one-warm encode the Division, Department, and Class names.

After that, a single characteristic matrix changed into created through combining those capabilities for Clustering**.**

**Clustering Methods**:

**K-Means Clustering**: After the usage of the Elbow Method to decide the best range of clusters, we selected four.

**DBSCAN**: To realise the density-based totally clustering sample of the information, we employed DBSCAN (Density-Based Spatial Clustering of Applications with Noise) as a supplementary technique.

**PCA**: The clusters have been shown through the application of Principal Component Analysis (PCA) for dimensionality discount.

**Cluster Characteristics**

**K-Means Clustering Results**

| **Cluster** | **Age** | **Rating** | **Positive Feedback Count** | **Division Name** | **Department Name** | **Class Name** |
| --- | --- | --- | --- | --- | --- | --- |
| 0 | 57.06 | 4.64 | 1.92 | General | Tops | Knits |
| 1 | 40.89 | 2.35 | 2.29 | General | Tops | Dresses |
| 2 | 35.52 | 4.71 | 1.61 | General | Tops | Dresses |
| 3 | 45.05 | 4.08 | 26.07 | General | Tops | Dresses |

**Cluster Interpretation**

Cluster 0: This group consists of older customers who have acquired short, tremendous remarks counts and high ratings. They mainly obtain knits.

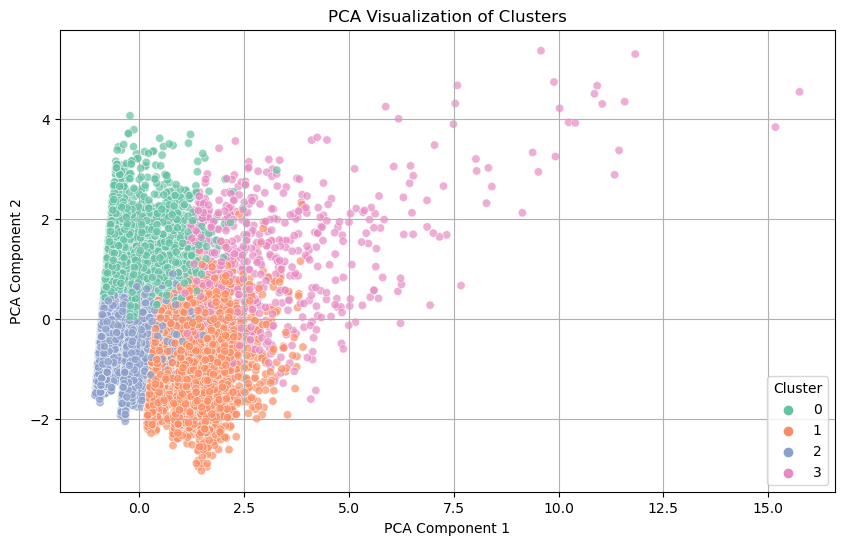
Cluster 1: This cluster consists of younger customers who have received poor evaluations. Dresses are what they buy most customarily.

Cluster 2: These clients have desirable ratings and are a tad younger. They also have a choice of dresses.

Cluster 3: Customers of their middle age who acquire common rankings but an excessive quantity of favorable comments. They have a desire for dresses.

**Visualization Results**

1. **PCA Visualization of Clusters**



The PCA scatterplot helps visualize the K-Means clusters in two dimensions:

* **Cluster Separation**: Clusters 0, 1, and 3 show clear separation, indicating distinct customer segments. However, there is some overlap between clusters 1, 2, and 3, indicating that these groups may share some similar characteristics.
* **Cluster Characteristics**: Cluster 0, composed of older customers with high ratings, stands out as quite distinct. Clusters 1, 2, and 3, mostly younger customers, tend to overlap, especially in the preferences for dresses.

**2. DBSCAN Clustering ResultsA graph of a number of data

Description automatically generated with medium confidence**

The DBSCAN clustering results show dense clusters with horizontal lines indicating segmentation based primarily on customer ratings:

**Noise Points:** Several factors have been recognized with the aid of DBSCAN as noise, representing customers whose styles are wonderful and do not agree to the primary clusters.

**Rating-Driven Segmentation:** The DBSCAN clusters are broadly speaking formed based totally on scores as opposed to age, suggesting that client pleasure tiers are the dominant aspect in forming segments.

**3. Age vs Rating Visualization by K-Means Clusters**

The scatter plot of Age vs Rating for the K-Means clusters shows:

• Cluster 0: This cluster is made of older clients with remarkable ratings.

• Cluster 1: Younger clients who regularly offer low reviews make up this cluster.

• Cluster 2: Generally talking, these are more youthful customers with erratic better rankings.

• Cluster 3: Customers inside the middle age institution who've above-common scores and a huge range of fine comments counts.

**4. Elbow Method for Optimal K:**

A graph of a number of clusters

Description automatically generated

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Age | Rating | Recommended IND | Positive Feedback Count |
| 0 | 35.309 | 4.60137 | 0.99966 | 1.67761 |
| 1 | 56.101 | 4.61788 | 0.99867 | 3.41174 |
| 2 | 42.3681 | 2.29044 | 0.0036 | 3.39372 |

**Customer segmentation:**

****

**Analysis and Marketing Strategies**

**Cluster 0: More Experienced, Content Clients**

• **Characteristics:** Older customers who're commonly glad about their purchases, with excessive rankings and low fantastic remarks counts. Their favourite clothing is knit.

**• Approach to Marketing:**

Loyalty applications: To keep those clients who've a high degree of delight, offer unique reductions or loyalty points.

Product Suggestions: Pay interest to knit-based recommendations that correspond with their buying patterns.

Encourage more feedback from these clients, as their range of properly evaluations is low despite their excessive ratings**.**

**Cluster 1: Younger, Dissatisfied Customers**

**• Characteristics**: Younger clients with low ratings, who typically buy dresses.   
**• Marketing Plan:** Product Enhancement: Examine the reasons behind this group's lower ratings. To allay their worries, raise the standard of the dresses or customer service.   
Targeted Promotions: To re-engage this demographic, provide discounts or promotions on dresses.   
Customer Service: Respond to complaints and restore confidence by offering individualized customer service.   
  
**Cluster 2: Younger, Highly Satisfied Customers**

**• Features:** Younger customers who price products surprisingly and purchase garments on a everyday basis.

**• Marketing Strategy:** Cross-promote and Upsell: To increase basket size, propose related merchandise like shoes or accessories.

It's feasible that those customers are more energetic on social media. To make stronger logo loyalty, use influencer relationships and social media advertising.

To sustain their interest, offer this group special sneak peeks at subsequent get dressed collections.

**Cluster 3: Middle-Aged, Variable Feedback Customers**

**• Features:** Customers of their centre age range with mediocre rankings and a large quantity of top critiques. They buy dresses as well.

**• Marketing Strategy:** Engagement Campaigns: Use their excessive response charge in your gain via asking them to participate in awareness corporations or client surveys whilst growing new products.

Loyalty Program: Provide greater perks, such early get admission to deals or special discounts, in exchange for everyday feedback.

Customize product pointers primarily based on previous remarks to improve the shopping for experience.

**Conclusion**

This clustering analysis has provided valuable insights into the different customer segments in the Women’s Clothing E-Commerce dataset. Each segment has unique characteristics, and the proposed marketing strategies are aimed at increasing engagement, satisfaction, and revenue for each group. By leveraging these strategies, the company can target specific customer behaviours and preferences more effectively.

**References:**

Barnes, L., & Lea‐Greenwood, G. (2006). Fast fashioning the supply chain: Shaping the research agenda. *Journal of Fashion Marketing and Management: An International Journal*, 10(3), 259-271. <https://doi.org/10.1108/13612020610679259>

Beverland, M. B. (2004). Uncovering “theories‐in‐use”: Building luxury wine brands. *European Journal of Marketing*, 38(3/4), 446-466. https://doi.org/10.1108/03090560410518637