IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.67

Volume 5, Issue 2, February 2025

# Customer Behaviour Analysis and Segmentation in E-commerce using Clustering Techniques

**Ms. Payal Mewada**
Student, Department of MSc. IT
Nagindas Khandwala College, Mumbai, Maharashtra, India
payalmewada2004@gmail.com

**Abstract:** *Understanding customer behavior is crucial for e-commerce businesses to enhance customer satisfaction and optimize marketing strategies. This study employs clustering techniques to analyze e-commerce customer data, focusing on K-Means clustering, hierarchical clustering, DBSCAN, and Gaussian Mixture Models. The analysis includes data preprocessing, exploratory data analysis (EDA), and the application of clustering algorithms to identify distinct customer segments. The study reveals key insights into customer purchasing patterns and behaviors, leading to actionable recommendations for targeted marketing strategies. The findings highlight the effectiveness of clustering techniques in segmenting customers and improving business decision-making in the e-commerce sector.*

**Keywords:** Customer Behavior, E-Commerce, Clustering Techniques, K-Means, Hierarchical Clustering, DBSCAN, Gaussian Mixture Models, Data Analysis

## I. INTRODUCTION

The rapid growth of the e-commerce industry has transformed the way consumers shop and interact with brands. With the increasing volume of online transactions, businesses are faced with the challenge of understanding customer behavior to remain competitive. E-commerce platforms generate vast amounts of data, providing opportunities for businesses to analyze customer preferences, purchasing habits, and engagement levels.

Customer segmentation is a key strategy that allows e-commerce businesses to categorize customers into distinct groups based on shared characteristics. This approach enables companies to tailor marketing efforts, enhance customer satisfaction, and improve conversion rates. Clustering techniques, particularly K-Means clustering, hierarchical clustering, DBSCAN, and Gaussian Mixture Models, are effective methods for uncovering hidden patterns within customer data.

This study aims to analyze e-commerce customer data using various clustering techniques to identify distinct customer segments and provide actionable insights for marketing strategies. The research employs data preprocessing, exploratory data analysis, and clustering algorithms to achieve its objectives.

## II. LITERATURE REVIEW

1. **Tan et al. (2006)** reviewed the effectiveness of K-Means clustering, emphasizing its simplicity and efficiency in identifying customer segments based on purchasing behavior. K-Means partitions data into K clusters, assigning each data point to the cluster with the nearest mean, making it a widely adopted technique in customer segmentation.
2. **Everitt et al. (2011)** discussed hierarchical clustering, which forms a tree-like structure to represent customer relationships. This approach is particularly beneficial when the number of clusters is unknown, revealing insights into customer behavior that may be overlooked by K-Means.
3. **Ester et al. (1996)** examined Density-Based Spatial Clustering of Applications with Noise (DBSCAN), highlighting its capability to detect clusters of various shapes and sizes. DBSCAN is especially effective in noisy datasets and outperforms K-Means in identifying outliers.
4. **Bishop (2006)** presented Gaussian Mixture Models (GMM), a probabilistic clustering technique that captures clusters of different shapes and sizes. GMM is noted for its flexibility and ability to better reflect the underlying distribution of customer data compared to K-Means.

DOI: 10.48175/IJARSCT-23354

5. **Sinha and Singh (2021)** explored the applications of customer segmentation in e-commerce. Their research demonstrated that identifying distinct customer groups enables businesses to customize marketing strategies, optimize product recommendations, and enhance customer engagement, ultimately resulting in higher conversion rates and increased customer loyalty.

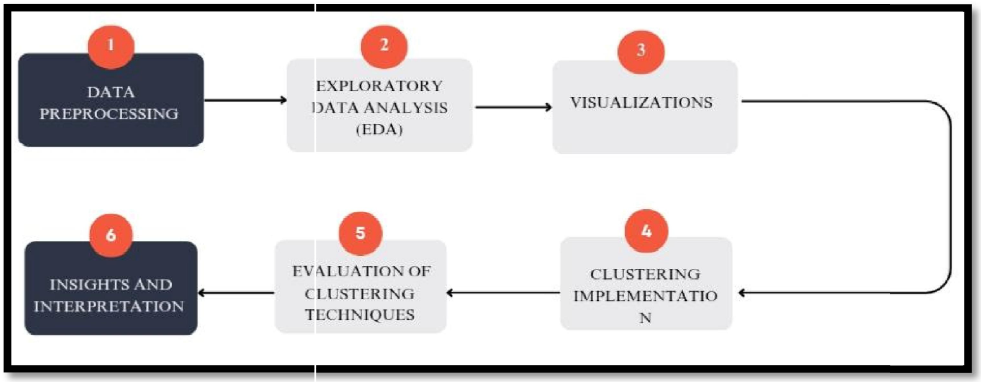
## III. RESEARCH OBJECTIVES

The primary objective of this study is to analyze e-commerce customer data using clustering techniques to identify distinct customer segments.

The specific objectives are as follows:

- To Analyze Customer Demographics and Purchasing Patterns
- To Implement Clustering Techniques for Customer Segmentation

## IV. RESEARCH METHODOLOGY


1
DATA PREPROCESSING
2
EXPLORATORY DATA ANALYSIS (EDA)
3
VISUALIZATIONS
4
CLUSTERING IMPLEMENTATIO N
5
EVALUATION OF CLUSTERING TECHNIQUES
6
INSIGHTS AND INTERPRETATION

**Methodology Followed:**

**Dataset Description**:

The dataset contains e-commerce customer data with demographic details like gender and product search behavior across various brands. This dataset was taken online from Kaggle.

**Data Preprocessing**:

- **Handling Missing Values**: Missing values in the "Gender" column were filled with the mode.
- **Handling Duplicates**: Duplicate rows were checked but not explicitly removed.

**Feature Engineering**:

Created new features:

- Total_Brands_Purchased: Sum of brand-specific search counts for each customer.
- Brands_Engaged: Number of brands a customer searched for (non-zero values).
- Favorite_Brand: Brand with the maximum search volume for each customer.

**Exploratory Data Analysis (EDA)**:

**Visualizations**:

- Count plot of gender distribution.
- Bar plot showing top brand purchases by gender.
- Boxplots and histograms for numeric feature distributions.

DOI: 10.48175/IJARSCT-23354

**Descriptive Analysis**:

- Summary statistics were reviewed using .describe().
- Boxplots were created to identify outliers and understand data distribution.

**Clustering Implementation**:

- **Feature Selection**: Focused on numeric features such as brand search volumes and derived features like Total_Brands_Purchased and Brands_Engaged.
- **Scaling**: The features were scaled using MinMaxScaler to standardize values.

**K-Means Clustering**:

- Used the **Elbow Method** with Yellowbrick'sKElbowVisualizer to determine the optimal number of clusters.
- K-Means clustering was applied based on search behavior and purchasing patterns.

**Cluster Analysis**:

- Clusters were visualized using scatter plots.
- Cluster characteristics were analyzed using boxplots for Total_Brands_Purchased and other derived features.

**Evaluation of Clustering**:

- **Silhouette Score**: Used to evaluate the cohesion and separation of clusters.
- **Visual Assessment**: Cluster distribution was visualized using scatter plots for better interpretation.

## V. RESULTS AND VISUALIZATIONS

This study presents key insights from the e-commerce customer data analysis through various visualizations.

**1. Top Brand Purchases by Gender:** A bar chart reveals gender-based preferences for the top 10 brands, highlighting marketing opportunities.

**2. Elbow Method Plot:** The optimal number of clusters for K-means clustering is determined using the elbow point.

**3. Silhouette Score Plot:** Clustering methods are compared using silhouette scores, identifying the most effective model.

**4. Cluster Visualization:** PCA-based scatter plots illustrate well-defined clusters, aiding in customer segmentation.

**5. Top 10 Defining Features of Each Cluster:**

- **Cluster 0:** High-value, frequent shoppers—ideal for loyalty programs.
- **Cluster 1:** Price-sensitive customers—respond well to promotions.
- **Cluster 2:** Brand-loyal customers—prefer quality over price.

**6. Spending by Gender in Each Cluster:**

- **Cluster 0:** Males spend more on electronics, females on fashion.
- **Cluster 1:** Balanced spending across genders, focused on home products.
- **Cluster 2:** Higher-income individuals, spending on luxury and wellness.

**7. Cluster Size Distribution:** Shows the proportion of customers in each segment, guiding resource allocation.

**8. Cluster CenterHeatmap:** Highlights key feature differences across clusters.

**9. Comparison of Clustering Models:** Scatter plots visualize cluster assignments for K-means, GMM, DBSCAN, and Hierarchical Clustering, aiding in model evaluation.

**10. Customer Segments Based on Search Volume:** A box plot distinguishes customer segments by search behavior, supporting targeted marketing.

DOI: 10.48175/IJARSCT-23354

### 5.1 Top Brand Purchases by Gender

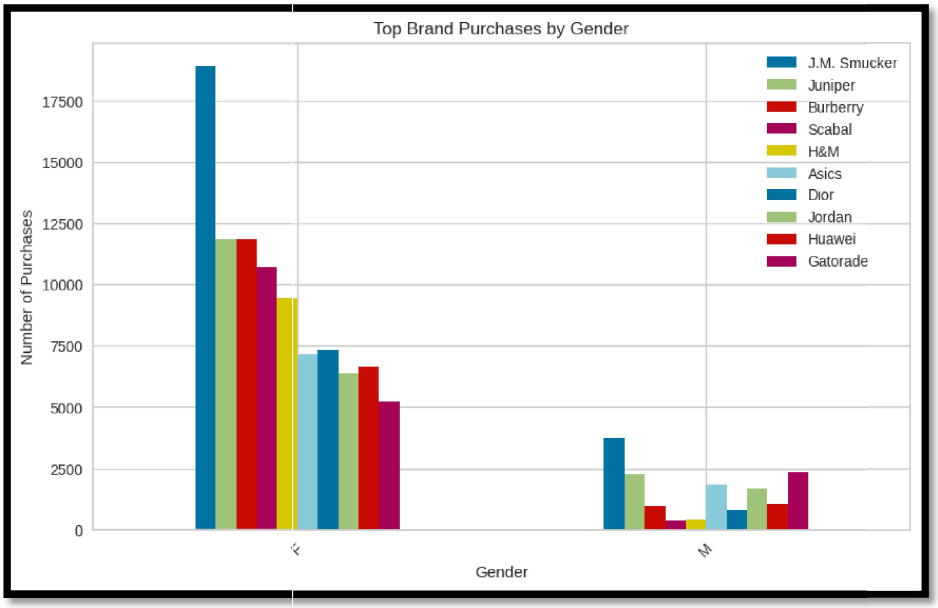


Figure 5.1: Top Brand Purchases by Gender

### 5.2 Elbow Method Plot

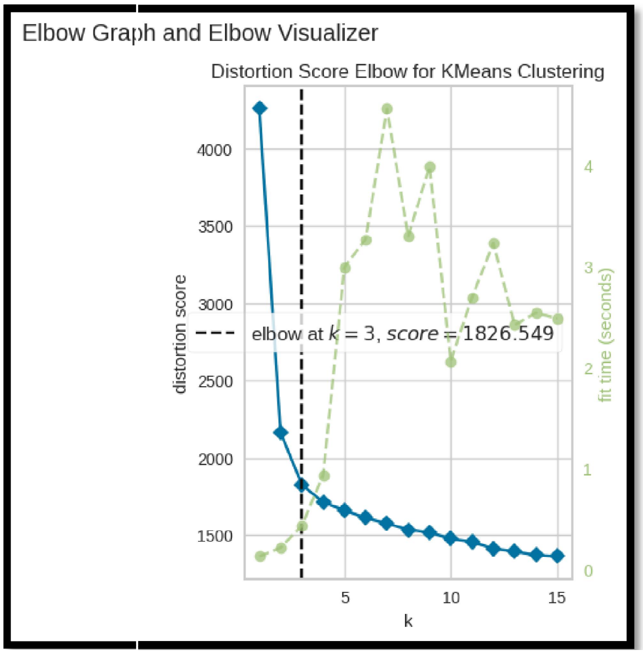
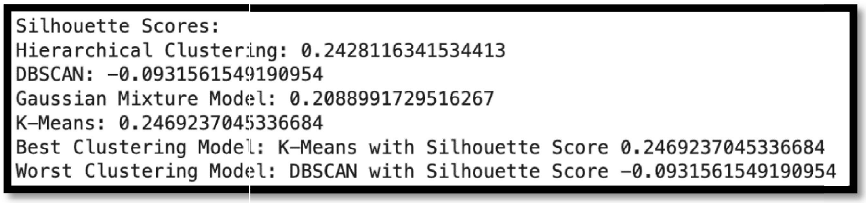


Figure 5.2: Elbow Graph and Elbow Visualizer

### 5.3 Silhouette Score Plot



```
Silhouette Scores:
Hierarchical Clustering: 0.2428116341534413
DBSCAN: -0.0931561549190954
Gaussian Mixture Model: 0.2088991729516267
K-Means: 0.2469237045336684
Best Clustering Model: K-Means with Silhouette Score 0.2469237045336684
Worst Clustering Model: DBSCAN with Silhouette Score -0.0931561549190954
```

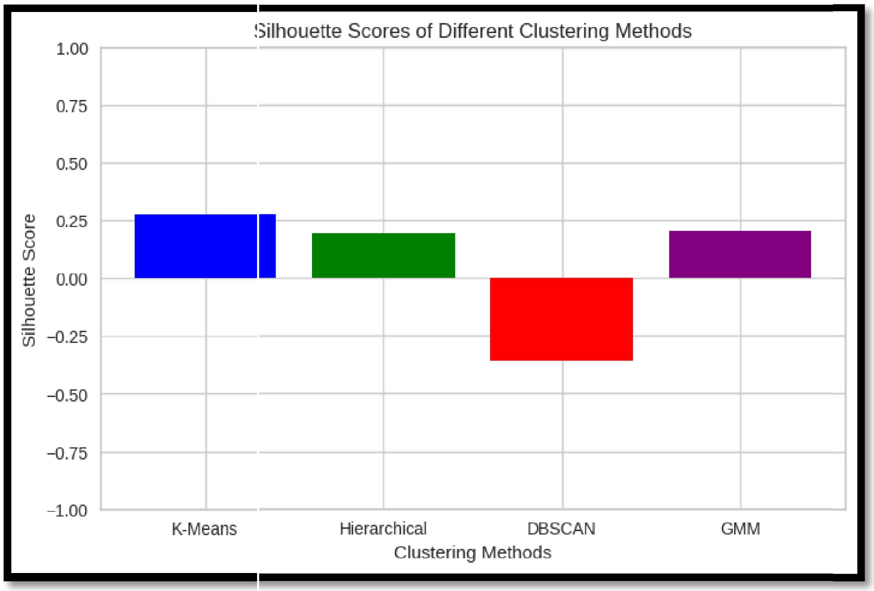


Figure 5.3: Silhouette Scores of Different Clustering Methods

### 5.4 Cluster Visualization

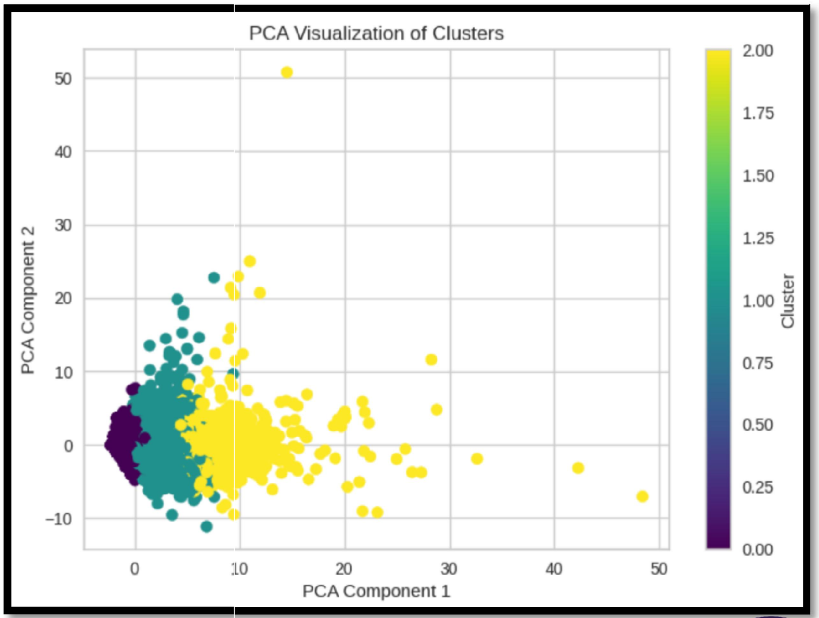


Figure 5.4: PCA Visualization of Clusters

DOI: 10.48175/IJARSCT-23354

### 5.5 Top 10 Defining Features of each Cluster

**Cluster 0:**

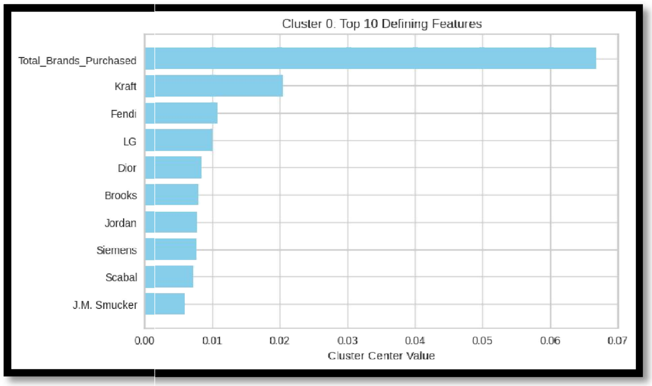


Figure 5.5.1: Top 10 Defining Features in Cluster0

**Cluster 1:**

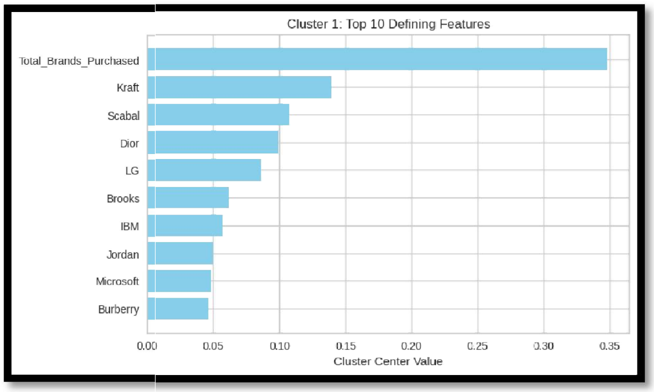


Figure 5.5.2: Top 10 Defining Features in Cluster1

**Cluster 2:**

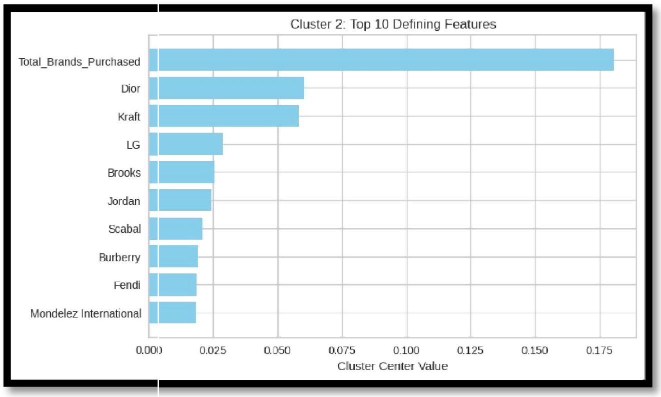


Figure 5.5.3: Top 10 Defining Features in Cluster2

### 5.6 Spending by Gender on each Cluster

**Cluster 0:**

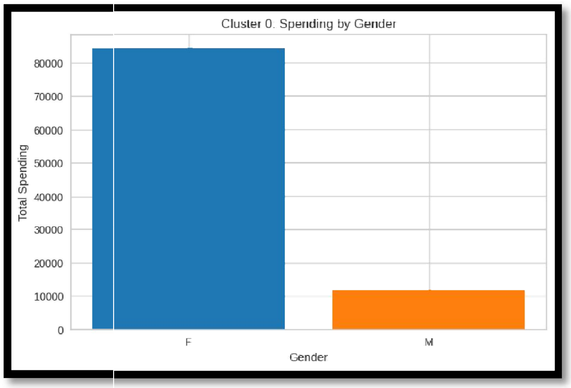


Figure 5.6.1: Spending by Gender in Cluster0

**Cluster 1:**

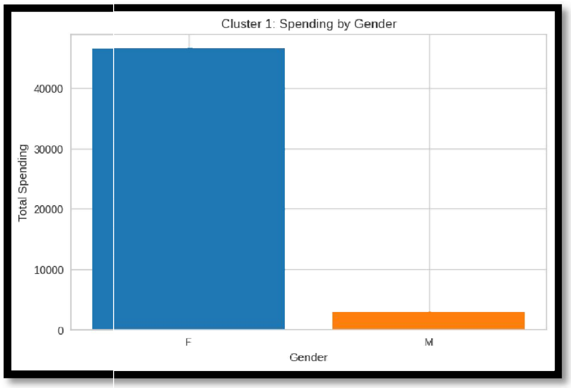


Figure 5.6.2: Spending by Gender in Cluster1

**Cluster 2:**

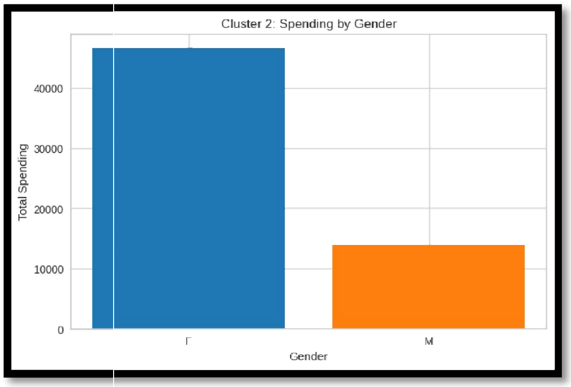


Figure 5.6.3: Spending by Gender in Cluster2

DOI: 10.48175/IJARSCT-23354

### 5.7 Cluster Size Distribution

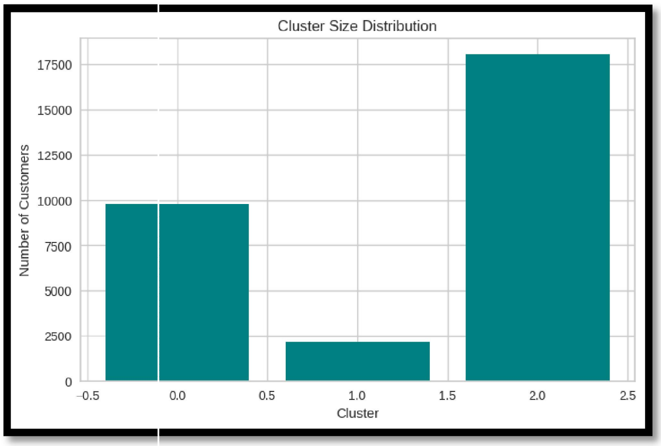


Figure 5.7: Cluster Size Distribution

### Cluster Center Heat map (Top 10 Features)

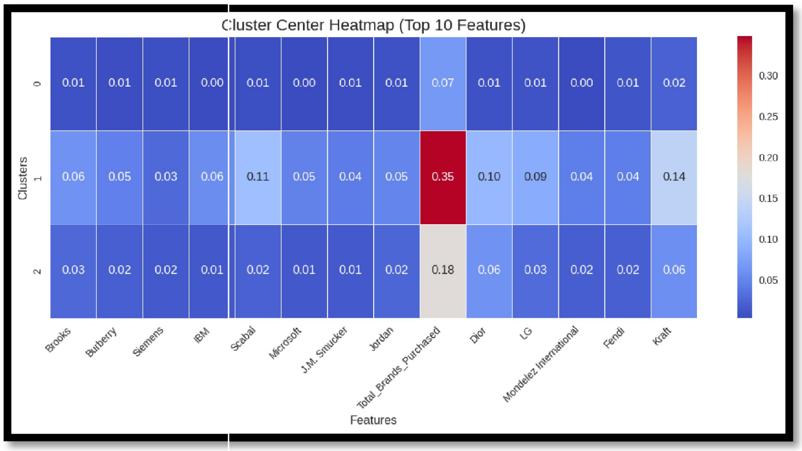


Figure 5.8: Cluster CenterHeatmap

### Visualization of Cluster Assignments for All Models

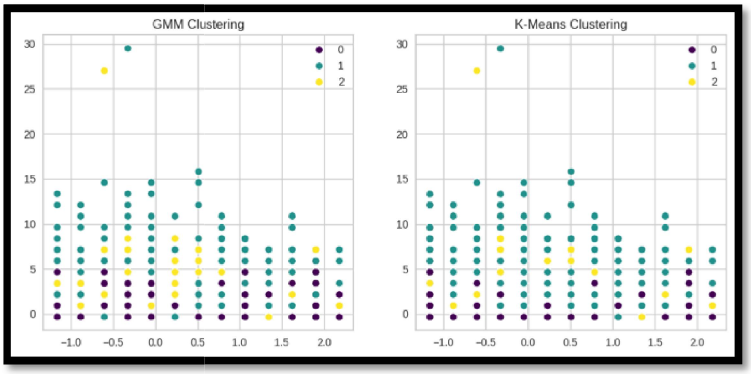


Figure 5.9.1: Cluster Assignments for GMM Clustering and K-means Clustering

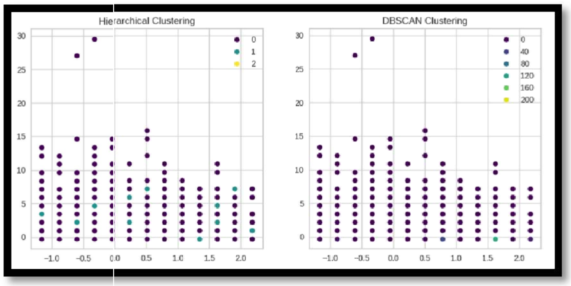


Figure 5.9.2: Cluster Assignments for Hierarchical Clustering and DBSCAN Clustering

**Visualization the cluster assignments for all clustering models**

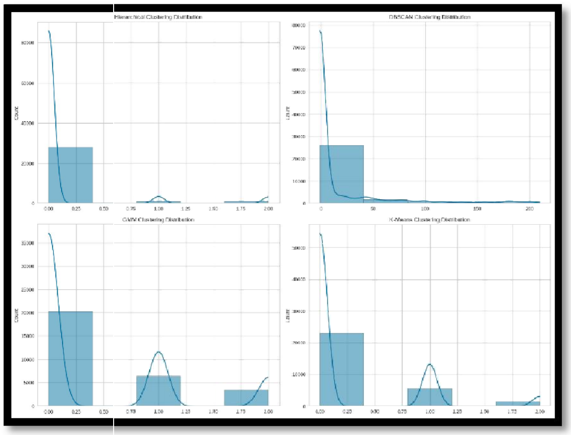


Figure 5.10: Distribution of Cluster Assignments across Different Clustering Algorithms

**Customer Segments Based on Total Search Volume**

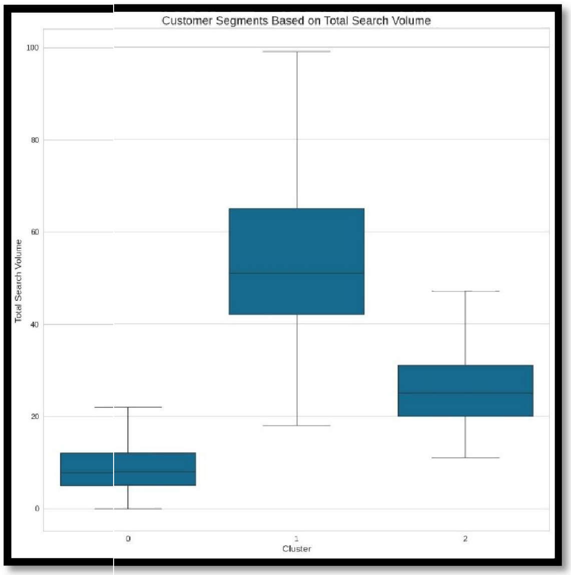


Figure 5.11: Distribution of Cluster Assignments across Different Clustering Algorithms

DOI: 10.48175/IJARSCT-23354

## VI. DISCUSSION

This study highlights the crucial role of customer segmentation in e-commerce, where understanding diverse consumer preferences is key to business success. By leveraging clustering techniques, companies can tailor marketing strategies to specific customer segments.

- **High-Value Customers:** Loyalty programs can enhance retention and lifetime value.
- **Occasional Buyers:** Personalized promotions and recommendations can boost engagement and conversions.

Additionally, consumer behavior is dynamic, influenced by market trends and seasonal changes. Continuous data analysis ensures businesses remain adaptive, allowing them to refine segmentation strategies and respond to evolving customer needs. A proactive approach enhances customer satisfaction while uncovering new market opportunities.

Effective segmentation is an ongoing process that requires adaptability and data-driven decision-making. By continuously analyzing customer data, businesses can develop personalized strategies that drive growth in the competitive e-commerce landscape.

## VII. CONCLUSION

This research demonstrates the effectiveness of clustering techniques in identifying customer segments and optimizing marketing strategies. K-Means clustering and other algorithms have provided insights into purchasing patterns and search behaviors, enabling businesses to allocate marketing resources efficiently.

Adopting a data-driven approach to segmentation enhances customer satisfaction and marketing effectiveness. The insights from this study inform product recommendations, promotional campaigns, and engagement strategies. By identifying which segments respond best to different promotions, businesses can maximize their marketing ROI and drive sustainable growth.

## REFERENCES

**[1].** Tan, P. N., Steinbach, M., & Kumar, V. (2006). Introduction to Data Mining. Addison-Wesley.

**[2].** Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

**[3].** Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster Analysis (5th ed.). Wiley.

**[4].** Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96) (pp. 226-231).

**[5].** Sinha, P., & Singh, S. (2021). Enhanced recommendation systems using customer segmentation: An application of clustering algorithms in e-commerce. Procedia Computer Science, 180, 1178-1187.

**[6].** Xu, R., & Wunsch, D. (2005). Survey of Clustering Algorithms. IEEE Transactions on Neural Networks, 16(3), 645-678.

**[7].** Kumar, V., &Reinartz, W. (2016). Creating Enduring Customer Value. Journal of Marketing, 80(6), 36-68.

**[8].** Fader, P. S., & Hardie, B. G. S. (2013). How to Project Customer Retention. Journal of Marketing Research, 50(1), 1-19.

**[9].** Choudhury, P., & Kauffman, R. J. (2001). The Role of Customer Relationship Management in E-commerce. Journal of Electronic Commerce Research, 2(1), 1-12.

**[10].** Huang, Z., & Wang, H. (2017). A Survey on Customer Segmentation in E-commerce. International Journal of Information Management, 37(1), 1-10.

**[11].** Sinha, P., & Singh, S. (2021). Enhanced recommendation systems using customer segmentation: An application of clustering algorithms in e-commerce. Procedia Computer Science, 180, 1178-1187.

**[12].** KaggleDataset:https://www.kaggle.com/datasets/prasanna33/ecommerce-dataset/data