



Analysis of Consumption Behavior of High-value Customers Based on RFM Model and K-means Clustering

Yaxuan Peng¹ , Zirui He¹ ,
Tiancheng Yang¹ and Qingye Huang² *

¹School of Advanced Manufacturing, Guangdong University of Technology, Guangzhou, 510006, China

²Institute of Collaborative Innovation, University of Macau, Macau, 999078, China

3222009113@mail2.gdut.edu.cn (Yaxuan Peng)

3122008758@mail2.gdut.edu.cn (Zirui He)

3124008965@mail2.gdut.edu.cn (Tiancheng Yang)

qingyehuang@163.com (Qingye Huang)

*Correspondence: qingyehuang@163.com

Abstract. This paper proposed a novel model to evaluate the behavioral characteristics of high-value customers using RFM model and the e-commerce consumer behavior data from the Kaggle platform. This model integrates detailed demographic analysis, including age-specific consumption patterns, payment preferences, and device usage to provide a more nuanced understanding of high-value customer segments. The results showed that high-value customers were mainly concentrated in the under-46 age group, and different age groups presented unique consumption patterns. Young people preferred mobile payment and flexible delivery, while the middle-aged and elderly group were more sensitive to discounts and had higher brand loyalty. Finally, this paper provided targeted precision marketing strategies and theoretical and practical guidance for enterprises to achieve refined operation.

Keywords: Consumer behavior analysis, RFM model, K-means, High-value customers, Precision marketing

1 Introduction

Global e-commerce has witnessed a rapid growth in the post-pandemic era [1]. In 2024, global e-commerce sales are expected to exceed 6 trillion US dollars. Consumer behavior is characterized by fragmented decision-making, diversified devices, and personalized demands [2]. Market competition has shifted towards the refined mining of value from existing users. The current research on customer value stratification lacks in-depth analysis of the characteristics of high-value customer groups of different age groups, such as the response of Generation Z to digital promotion and the device preferences

of the elderly, resulting in insufficient matching between marketing strategies and customer group needs and affecting operational efficiency [3].

The remaining parts of this paper are arranged as follows: Section 2 will introduce the data sources and research methods. Section 3 will present the analysis results from dimensions such as age distribution, discount sensitivity, and device preference. Section 4 discusses the practical application value of the research findings. Section 5 summarizes the content of this paper.

2 Methods

In this section, we use the dataset from the Kaggle platform, which can be found at Ecommerce Consumer Behavior Analysis Data [4]. This dataset contains 1,000 consumer behavior records from January to September 2024.

Based on this dataset, the RFM model and K-means clustering method are conducted to screen out high-value user groups [5]. The details are shown as follows.

Step 1: First, according to the RFM model, the three key indicators Recency, Frequency and Monetary are extracted and standardized by Z-score as clustering variables, where μ represents the mean and σ represents the standard deviation.

$$X_{\text{norm}} = \frac{x - \mu}{\sigma} \quad (1)$$

Step 2: Calculate the SSE values using Elbow Method.

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} |x - \mu_i|^2 \quad (2)$$

Step 3: Select k center points randomly and perform iterative operation until the center points are stable or the maximum number of iterations reached.

$$C_i = \{x: |x - \mu_i| \leq |x - \mu_j|, \forall j\} \quad (3)$$

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \quad (4)$$

We set k=5 because customers are divided into 5 categories in the RFM model.

Finally, the center points of the RFM characteristics of the high-value user group are determined as Recency=180, Frequency=5, and Monetary=1000. There are 346 users meeting the characteristics were screened out for subsequent analysis.

3 Results

3.1 Age Distribution

To analyze the distribution characteristics of customers of different age groups, the continuous age variable was divided into four discrete intervals: 18-25 years old, 26-35 years old, 36-45 years old, and 46 years old and above. The distribution of high-value customers in different age groups is shown in Fig. 1.

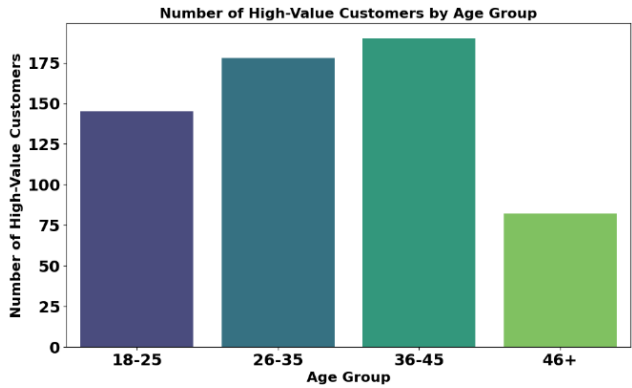


Fig. 1. Number of High-Value Customers by Age Group

3.2 Discount Sensitivity

The responses of high-value customers in different age groups to discounts vary. As shown in Fig. 2, customers aged 18-25 have a relatively neutral attitude towards discounts. Those aged 36-45 are slightly more positive. Customers over 46 are the most interested in discounts. In contrast, customers aged 26-35 are less attracted by discounts and may place more emphasis on product quality or other factors.

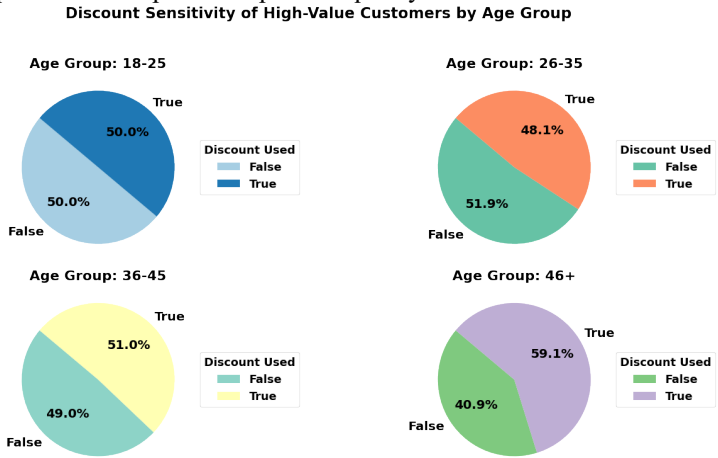


Fig. 2. Discount Sensitivity of High-Value Customers by Age Group

3.3 Device Preference

The differences in shopping device choices among high-value customers of different age groups are shown in Fig. 3. Tablets are the most popular shopping devices among all age groups, demonstrating the preference of high-value customers for portability and multi-functionality. Desktop computers and smart phones account for the second

largest proportion among high-value user groups, while older user groups also have a higher acceptance of mobile shopping.

3.4 Payment Method

From the perspective of the age distribution of payment methods, the consumption habits of high-value customers are shown in Fig. 4. Young customer groups aged 18 to 25 are more inclined to use other payment methods, such as mobile payment. Middle-aged and young customers aged 26-35 and 36-45 prefer to use cash and PayPal. The customer group over 46 years old prefers traditional payment methods such as credit cards or others.

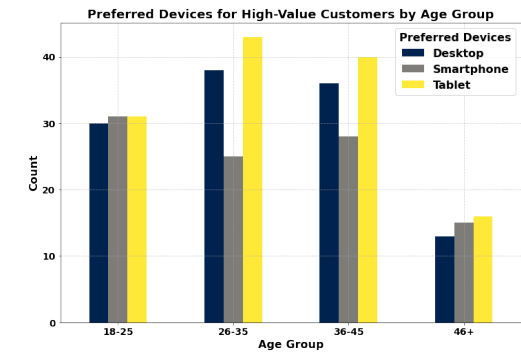


Fig. 3. Preferred Devices for High-Value Customers by Age Group

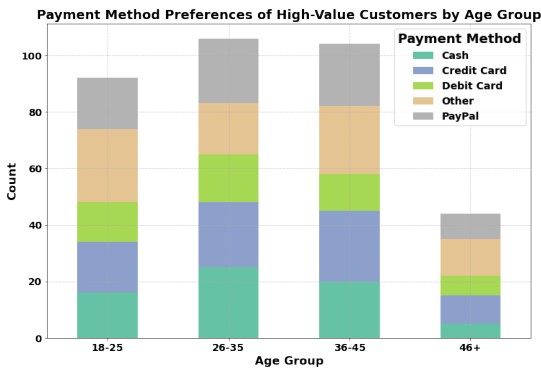


Fig. 4. Payment Method Preferences of High-Value Customers by Age Group

3.5 Delivery Preference

As shown in Fig. 5, it can be seen that high-value customers of different age groups have diversity and differences in their distribution preferences. The distribution of delivery preferences in the age groups of 18-25 years old and 36-45 years old is relatively

uniform, indicating that customers in this age group are more flexible in choosing delivery, while those in the age group of 26-35 years old have no special preferences for delivery methods.

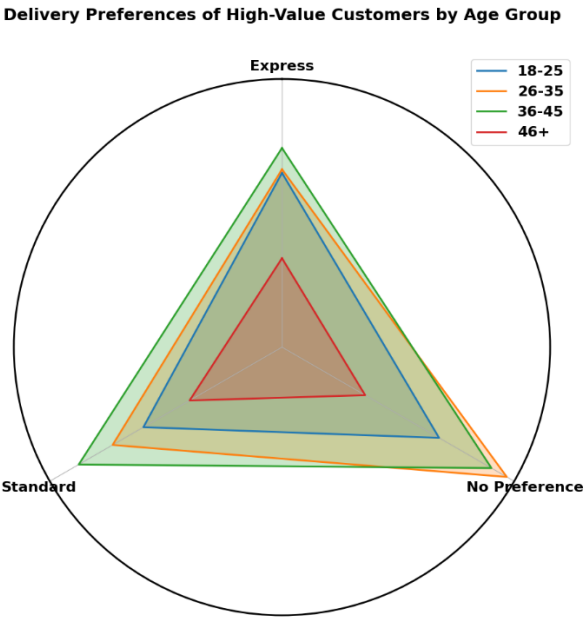


Fig. 5. Delivery Preferences of High-Value Customers by Age Group

3.6 Brand Loyalty

Fig. 6 presents the brand loyalty of high-value customers across different age groups. The brand loyalty of the age groups of 18-25, 26-35 and 36-45 are lower than the age group over 46. This indicates that as customers grow older, their recognition and loyalty to brands increase, and they are more incline to persist in using the brands they trust for a long time.

4 Discussion

This research finds that high-value customers are mainly concentrated in the group under 46 years old. The differences in consumption behavior among different age groups are manifested as follows. For the 18-25 age group, mobile payments and flexible delivery options are the top preferences, while their attitude toward discounts remains relatively neutral. Those aged 26-35 place greater emphasis on product quality and show less sensitivity to promotional discounts and the 36-45 demographic responds more positively to discounts. As for consumers aged 46 and above, they demonstrate

the highest sensitivity to discounts, exhibit strong brand loyalty, and tend to favor traditional payment methods over digital alternatives.

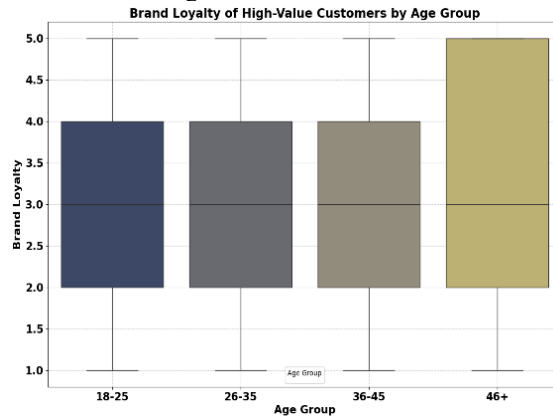


Fig. 6. Brand Loyalty of High-Value Customers by Age Group

5 Conclusion

This study identifies high-value customers who have recent purchase intervals of less than 180 days, purchase frequency exceeding 5 times per year, and spending above 1000, and reveals their varying behaviors across age groups, offering insights for precision marketing. However, this Kaggle data may have limited representativeness and cultural diversity. Future work can expand the data range and use reinforcement learning to dynamically adapt to customer segmentation and enhance marketing strategies.

References

1. Khurjekar S, Kaur R, Puntambekar D, et al. Green growth in e-commerce: Environmental management and customer actions. *Journal of information and optimization sciences*, 2024(6):45.
2. Wang X , Cai B , Salleh K. Research on the application of cluster analysis in cross-border e-commerce customer segmentation and market strategy. *Applied Mathematics and Nonlinear Sciences*, 2024, 9(1).
3. Gumay E C, Bangsawan S, Pandjaitan D R H. The Influence of Social Presence on Impulsive Behavior Buy Gen Z on Live Streaming E-Commerce Shopee with Sor Framework. *Journal of Social Research*, 2024, 3(7).
4. Ecommerce Consumer Behavior Analysis Data, <https://www.kaggle.com/datasets/salahud-dinahmedshuvo/e-commerce-consumer-behavior-analysis-data>, last accessed 2025/05.
5. Wang X, Cai B, Salleh K. Research on the application of cluster analysis in cross-border e-commerce customer segmentation and market strategy. *Applied Mathematics and Nonlinear Sciences*, 2024, 9(1).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

