WEP Warwick Evans Publishing

International Journal of Global Economics and Management

ISSN: 3005-9690 (Print), ISSN: 3005-8090 (Online) | Volume 4, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijgem.v4n1.02
Journal homepage: https://wepub.org/index.php/IJGEM/index

Data-Driven Market Segmentation: K-means Clustering and STP Analysis in Mainland China's Sportswear Industry

Jiajun Mai *

Stirling College, Chengdu University, Chengdu, Sichuan, 610106, China
*Corresponding Author: J_1064354741@outlook.com

ABSTRACT

Post-Covid-19, mainland China's economy is gradually recovering, leading to increased sportswear sales. This study integrates STP theory (segmentation, targeting, positioning) with the K-means clustering algorithm to analyze the sportswear market in mainland China. Consumer data was collected via a questionnaire survey, pre-processed, and cleaned before applying K-means clustering to segment consumers into four distinct groups. The analysis reveals significant differences in purchasing habits across these segments. This data-driven approach offers sportswear companies insights for targeted marketing strategies, enhancing the precision and effectiveness of their marketing efforts in mainland China.

KEYWORDS

Sportswear Market; Mainland China; Market Segmentation; Target Market; Market Positioning; K-means algorithm

1. INTRODUCTION

As a country with a huge population base and rapid economic growth, China's sportswear market shows strong growth potential. In recent years, the size of the sportswear market has grown rapidly globally with the rise of national fitness awareness and the promotion of healthy lifestyles. According to a market research report by Euromonitor International, the sales of the sportswear market in mainland China have grown gradually since 2009, reaching a peak of CNY 385,801 million in 2023. This represents a 14.2% increase in 2023 compared to 2022, and the sportswear market in mainland China will continue to show growth over the next five years.

The rapid expansion of the sportswear market has introduced new challenges. Consumers' needs are becoming more diverse, necessitating swift adaptation and optimization of marketing strategies by brands to remain competitive. Companies must utilize advanced market analysis tools and methods to accurately identify target markets and develop effective positioning strategies.

This study aims to integrate STP strategies with machine learning methods to analyze the sportswear market in mainland China. By employing questionnaire surveys and market data collection, machine learning techniques will be used for market segmentation, target market identification, and positioning strategy formulation tailored to varying market demands. The primary objective is to achieve precise market segmentation and effective positioning through advanced data analysis techniques, providing practical marketing advice to enterprises to enhance their competitiveness in a highly competitive market.

2. LITERATURE REVIEW

2.1. STP Strategy

STP strategy is divided into segmentation, targeting and positioning, which is an important theoretical and practical tool in the field of marketing. STP strategy focuses more on consumers, which helps enterprises to identify different groups in the market, choose the most attractive target market, and develop differentiated marketing strategies for them [1].

2.1.1. Segmentation

Market segmentation is the initial step in the STP strategy, involving the division of a market into distinct groups of individuals, groups, or organizations with similar buying needs and desires. This consumer-oriented process can be applied to any market. Traditionally, market segmentation variables are categorized into five types: demographic, psychographic, geographic, behavioral, and product-related factors [2, 3]. By analyzing these variables, companies can effectively segment the market and tailor their strategies accordingly.

2.1.2. Targeting

Targeting is the second step in the STP strategy, which is driven by segmentation. In this process, the most profitable markets will be selected. And marketing strategies are specified for the characteristics of the segments. Mark Anthony Camilleri states that there are three types of marketing strategies: undifferentiated marketing, differentiated marketing and focused marketing [4]. These three marketing strategies aim to achieve market coverage.

2.1.3. Positioning

This is the final step in the STP strategy, and more specifically, the process is one of product positioning. In this process, the company will be based on the results of segmentation and targeting, shaping a company's image, targeting the development of a product that meets the requirements of consumers and has a clear and differentiated position in the minds of customers and competing products [1].

2.2. Machine Learning in Marketing Analysis

K-means clustering is a widely used clustering algorithm and is often used as a baseline for comparing other clustering methods [4]. K-means clustering assigns data points to K clusters centers through an iterative optimization process until the cluster to which each data point belongs no longer changes. The method is simple, efficient, and suitable for large-scale datasets. K-means clustering is very effective for market analysis. For example, companies can identify different groups of consumers based on characteristics such as their age, income, frequency of purchases, and brand loyalty by performing cluster analysis of the market.

Cluster analysis can also identify market opportunities and unmet needs. By analyzing consumer purchasing behavior and preferences, companies can pinpoint groups with strong demands for specific product features that are currently underserved. This analysis provides a foundation for developing new products and formulating differentiated marketing strategies.

3. METHODOLOGY

3.1. Research Design

The main objective of this study is to understand the consumer characteristics, purchasing behavior, psychological characteristics, and market potential of the sportswear market in mainland China. Using machine learning methods, the study aims to perform segmentation analysis and subsequently

complete the targeting and positioning sections. This study adopts a quantitative research methodology by designing a structured questionnaire for data collection. The questionnaire is divided into five sections: basic information, purchasing behavior and budget, psychological characteristics and preferences, lifestyle characteristics, and market potential analysis. Each section contains at least five closed-ended questions, covering various aspects such as gender, age, education, income level, purchase frequency, budget, purchase channels, psychological characteristics, lifestyle, and market demand. A Likert scale is used to measure the importance respondents attach to factors such as brand, price, style, functionality, and comfort, as well as their consumption behaviors and preferences, totaling 23 questions.

Convenience sampling was used to distribute the questionnaires through online channels, targeting sportswear consumers in mainland China. The study utilized the Wenjuanxing platform for data collection, gathering 353 questionnaires within two weeks. To ensure the representativeness and reliability of the data, invalid questionnaires (incomplete or apparently random responses) were excluded, resulting in 306 valid questionnaires.

3.2. Data Analysis Process

After cleaning the collected data, missing values, outliers, and duplicates were addressed, and responses using Likert scales were normalized. For categorical variables, such as age group, a one-hot encoding approach was used to convert them into multiple binary variables, allowing efficient data processing by the algorithm. These steps ensured a high-quality database for subsequent cluster analyses, enhancing the reliability and validity of the results.

Before performing the cluster analysis, the optimal number of clusters was determined using the elbow method. This method involved plotting the sum of squared errors (SSE) for different numbers of clusters and identifying the "elbow" point in the plot, which indicates the appropriate number of clusters [5]. The data was then analyzed using the scikit-learn library in Python, employing the K-means clustering algorithm to segment the market of consumers. Cluster analysis divided consumers into different groups based on several dimensions, such as purchasing behavior, and psychological characteristics. The clustering results helped identify consumer groups with similar characteristics, providing a scientific basis for developing market strategies. These results were used to determine the main characteristics and needs of different consumer groups and to formulate corresponding marketing strategies.

The findings are presented through charts and statistical descriptions, detailing the characteristics, purchasing behavior patterns, psychological preferences, and their impact on the market potential of each consumer group. By focusing on the main characteristics and needs of each group, corresponding marketing strategies are proposed to help enterprises better understand consumers and provide data support for market decision-making.

4. RESULT

4.1. Data Visualization

The socio-demographic analysis of the 309 valid questionnaires revealed that 165 respondents (53.4%) were male, and 144 (46.6%) were female. The predominant age group was 26-30 years, comprising 108 individuals (35.3%), followed by the 18-25 age group with 83 respondents (27.2%), and the 31-40 age group with 49 respondents (15.9%). Educationally, 44% held a university degree, and 29.4% had an associate degree. Regarding occupation, private company employees constituted the largest group at 116 (37.9%), followed by 69 government or public agency employees (22.7%), 58 freelancers (18.8%), and 45 students (14.9%).

After performing basic data preprocessing, including the deletion of invalid questionnaires, handling of missing values, unification, and normalization, the elbow method (Fig. 1) was employed. It was observed that the rate of decrease in the total sum of squares of errors (SSE) significantly slows down when the number of clusters reaches 4, indicating that the optimal number of clusters is 4.

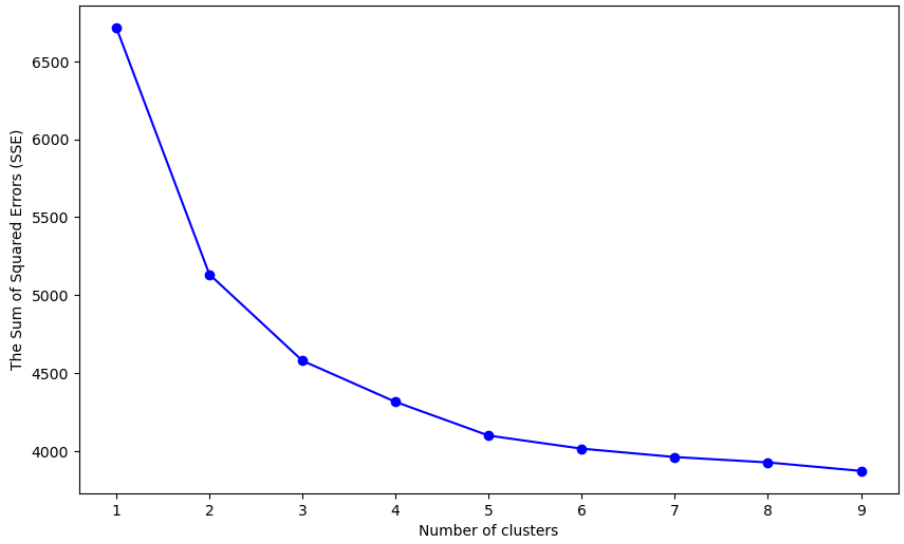

Figure 1. Visualization of the Elbow Method during Data Analysis

The number of k-means clusters was set to 4 for clustering after which PCA dimensionality reduction was performed to reduce the high dimensional data to two dimensions by labelling the clusters and the center of each cluster. Fig. 2 shows that the data was evenly divided into four parts, with clusters 0 and 3 being more dispersed in comparison to clusters 1 and 2.

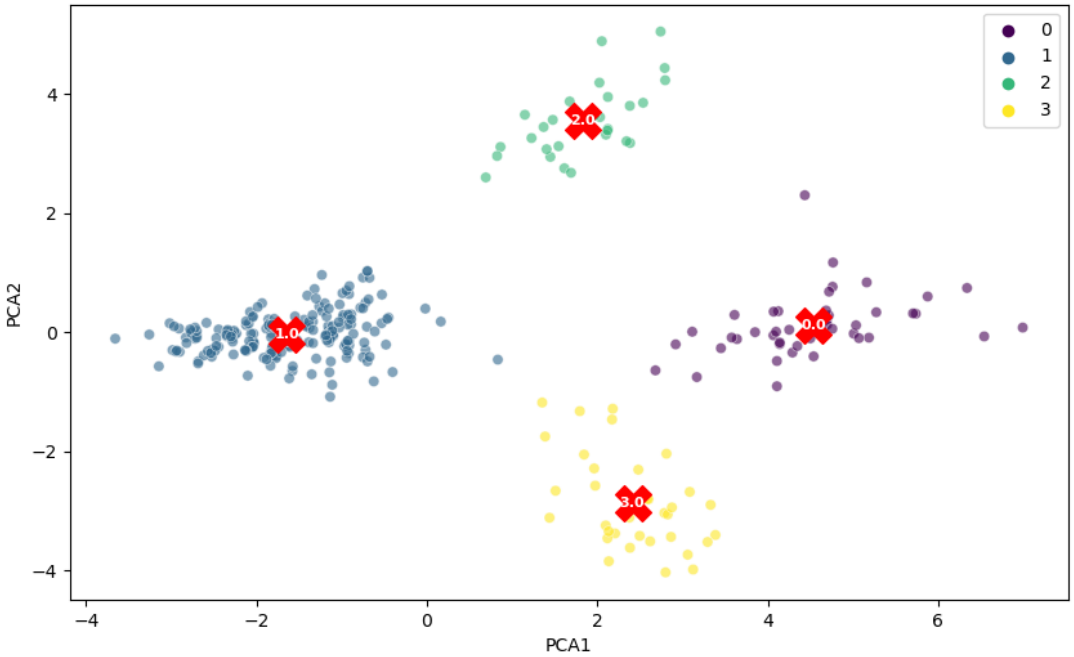

Figure 2. Reduced Dimensional Display of Clustering Results and Cluster Centers

4.2. Marketing Segmentation

An analysis of the cluster centroids was done, and the four clusters were summarized into four mainland China sportswear segments, an analysis of the cluster centers was done, and the following results were summarized:

Table 1. Summary of Clustering Center Points

Cluster	Exercise Habits	Purchase Preferences	Buying Behavior
Price-Sensitive Low-Frequency Exercisers (14.1%, 43 people)	Low frequency, static activities	Price-sensitive, prefers affordable options, high interest in smart sportswear, low consideration for seasonal change and sustainability, low comfort and brand loyalty	Low frequency, low budget, prefers WeChat stores (44%) and online shops (33%)
Brand-Conscious Active Exercisers (66.0%, 202 people)	High frequency, dynamic activities	Values brand and environmental factors, high demand for style and functionality, high consideration for seasonal and environmental factors, high brand loyalty, low interest in smart sportswear	High frequency, balanced budget, prefers physical stores (23%) and online shops (26%)
Professional Exercisers (9.5%, 29 people)	High frequency, gym workouts, professional competitions	Prioritizes functionality and technology, less attention to price and comfort, low consideration for seasonal and environmental factors, high brand loyalty, low interest in smart sportswear	Moderate frequency, high budget, prefers online shops (34%) and physical stores (21%)
Technological Enthusiasts (11.4%, 35 people)	Low frequency, small interest in dynamic activities	High interest in smart sportswear and emerging technologies, low brand loyalty, low consideration for seasonal and environmental factors	Moderate frequency, highest budget, prefers WeChat stores (43%) and online shops (34%)

5. DISCUSSION

5.1. Targeting

Brand-Conscious Active Exercisers: This largest segment, about 66.0%, exhibits diverse purchasing behaviors and psychological characteristics. These consumers are active exercisers who purchase frequently and demonstrate strong brand loyalty. They place significant emphasis on seasonal changes and environmental factors. Given its size and purchasing behavior, this segment should be a primary target market.

Professional Exercisers: Representing about 9.5% of the sample, this group is characterized by high brand loyalty and regular sports participation. They have a moderately high purchasing budget and prioritize professional athletic performance. Many perceive a lack of variety in current market offerings, indicating potential for further market development. Thus, this segment presents an opportunity for targeted marketing strategies.

Technological Enthusiasts: Accounting for approximately 11.4% of the sample, this group has a strong demand for smart sportswear. Despite low brand loyalty and interest in sports, they have moderate purchase frequency and the highest purchasing budgets among all segments. Their receptiveness to emerging products and technologies makes this segment potentially lucrative.

Price-Sensitive Low-Frequency Exercisers: Comprising about 14.1% of the sample, this segment shows high interest in smart sportswear but has a low budget and places less importance on brand loyalty and environmental factors. Their limited acceptance of high-end and mid-range products presents challenges in generating high profit margins. Given the relatively small market share and lower profitability potential, this segment may not be a priority for resource-constrained enterprises.

5.2. Positioning

Brand-Conscious Active Exercisers:

Diversified product lines: offering sportswear in different seasons, price points, and styles.

Brand Image Reinforcement: Promote brand loyalty through advertising and brand campaigns.

Environmental protection and sustainability: Promote environmentally friendly materials and sustainable production processes.

Limited editions and collaborations: Introduce limited editions and collaborations of sportswear to further enhance brand loyalty and market appeal.

Professional Exercisers:

Functionality and technicality: Promote high performance sportswear such as high breathability, waterproofness and abrasion resistance.

Specialized brand image: Promote brand awareness in the professional sports sector.

High-end market positioning: meet its high-end needs through high-quality, high-performance products.

Professional co-operation: To enhance the brand's professional image and market competitiveness through co-operation with professional athletes and sports teams, launching professional-grade sports apparel and promoting it through channels such as professional events and training camps.

Technological Enthusiasts:

Intelligent sports products: develop and promote intelligent sports products with functions such as health monitoring.

Technological innovation: Introduce new and innovative high-tech products to attract this group.

Dynamic activity support: develop sportswear designed for outdoor adventure, running, etc.

Technological experience activities: organize technological experience activities to allow consumers to experience the advantages of smart sports products first-hand, enhancing consumers' willingness to buy and brand identity.

5.3. Limitation and Future Research Directions

The structured questionnaire was insufficient in capturing the complexities of consumer behavior. Future research should incorporate open-ended questions or qualitative methods, such as focus groups and in-depth interviews, to achieve more comprehensive insights. The limited duration of data collection precluded the observation of market dynamics over time. Longitudinal studies are recommended to monitor consumer behavior and market trends. Additionally, this study did not adequately address regional variations. Future research should conduct detailed regional analyses to develop more precisely targeted marketing strategies.

6. CONCLUSION

By combining STP theory and K-means clustering method, this study successfully identifies four major consumer groups in the sportswear market in mainland China: price-sensitive low-frequency exercisers, brand-conscious active exercisers, professional exercisers, and technology enthusiasts. The results of the study show that there are significant differences between the different groups in terms of purchasing behavior, psychological characteristics and market demand. Brand-conscious active athletes are the target group with the most market potential, while professional athletes and

technology enthusiasts also show some market value. Based on these findings, this study provides practical guidance for sportswear companies to formulate differentiated marketing strategies, helping them to enhance their competitiveness in a competitive market. Future research should further optimize sampling methods, conduct longitudinal studies and regional analyses to gain more comprehensive market insights.

REFERENCES

[1] Khandelwal, K., Jakhar, T., Khandelwal, T. (2020) Segmentation, Targeting, and Positioning. International Research Journal of Engineering and Technology, 7(5): 1509-1514.

[2] Camilleri, M.A. (2017). Market segmentation, targeting and positioning. In: M. A. Camilleri, Travel marketing, tourism economics and the airline product, pp. 69-83.

[3] Schlegelmilch, B.B. (2016). Segmenting targeting and positioning in global markets. In Management for Professionals, pp. 63-82.

[4] Murray, P.W., Agard, B., Barajas, M.A. (2017) Market segmentation through data mining: A method to extract behaviors from a noisy data set. Computers & Industrial Engineering, 109: 233-252.

[5] Cui, M. (2020) Introduction to the K-Means Clustering Algorithm Based on the Elbow Method. Geoscience and Remote Sensing, 3: 9-16.