

# ***Clustering Analysis of Online Shopping Behavior on the TikTok Platform: Revealing Different User Characteristics***

**Hanjing Chen<sup>1,a,\*</sup>**

<sup>1</sup>*School of International Studies, Communication University of China, Beijing, 100024, China*  
*a. hj041105@163.com*

*\*corresponding author*

**Abstract:** With the integration of social media and e-commerce platforms, TikTok has become an important online shopping platform. This study aims to explore the online shopping behavior patterns of different user groups on the TikTok platform through cluster analysis methods. By designing and implementing a questionnaire survey, data on users' personal characteristics, usage habits, and purchasing behaviors were collected. Applying cluster analysis techniques, users were divided into different groups, and the characteristics of each group were deeply analyzed. This paper mainly focuses on identifying whether users' characteristics are closely related to their shopping behaviors. This study finds that engagement with live streaming shopping, diversity of products purchased, and frequency of shopping on TikTok are key factors influencing their online shopping behaviors. This research not only provides targeted marketing strategy suggestions for the TikTok platform but also offers a new perspective for understanding e-commerce behavior on social media platforms.

**Keywords:** Online Shopping Behavior, Cluster Analysis, User Characteristics, Social Media E-Commerce

## **1. Introduction**

In recent years, social media platforms have become more than just venues for information exchange; they have turned into important e-commerce platforms. As a standout, the online shopping behavior on the TikTok platform has attracted the attention of many researchers and marketers. This study focuses on revealing the online shopping behavior patterns of different users on the TikTok platform through cluster analysis. By designing a questionnaire survey, the study collects data on users' personal characteristics, usage habits, and purchasing behaviors, and applies cluster analysis techniques to process and analyze the data. Through comparative analysis, this paper mainly focuses on identifying whether users' characteristics are closely related to their shopping behaviors., thereby providing a basis for marketing strategy and user experience optimization on the TikTok platform. At the same time, the results of the study are also of significant importance for understanding the trends of e-commerce behavior on social media platforms today.

## 1.1. The Trend of Integration between Social Media and E-commerce

The "China Internet Development Status Statistical Report" by the China Internet Network Information Center (CNNIC) indicates a significant growth in China's online shopping user base, reaching 710 million by March 2020, with mobile shoppers comprising 78.9% of this figure, highlighting the dominance of mobile shopping [1]. The evolution of online shopping from PC-centric to mobile app-driven platforms, and now to a more diversified approach including live streaming and mini-programs, reflects changes in technology, user purchasing behaviors, and consumption habits. This trend underscores the transition towards shopping within varied life scenarios, facilitated by the integration of social, entertainment, and shopping experiences.

In the Web 2.0 era, internet applications have become more people-oriented, with end-users generating and leading content and actions. This shift has given rise to social e-commerce, a new model of e-commerce that leverages interpersonal networks and social media for the promotion and sale of products through social interactions and user-generated content. Distinguished from traditional e-commerce by its low cost of traffic acquisition and high conversion rates, social commerce benefits from precise marketing, enhanced social effects, and strong user loyalty, offering unique advantages like discovery shopping and rich scenarios.

Social commerce, or social + e-commerce, integrates and mutually embeds social media and e-commerce, manifesting in various forms from embedding social features into e-commerce platforms to incorporating shopping channels into social media platforms. The rapid growth of social commerce, as evidenced by its continued expansion since 2014, presents new opportunities and challenges for business model innovation and sustainable development [2]. This growth also sparks significant interest in academic research, aiming to understand and innovate upon the social commerce trend and its implications for future e-commerce and social media integration.

## 1.2. The Position and Role of the TikTok Platform in Social Commerce

In recent developments, short videos and live streaming have emerged as innovative channels in the publishing industry, driving online transformation among book publishers. TikTok E-commerce released the "2023 TikTok E-commerce Book Consumption Data Report", highlighting that over 400 million books were sold through the platform in the past year, averaging more than 2 million books per day [3]. This surge in book sales is attributed to the platform's ability to stimulate reading interest among viewers, with book-related live streams and short video clips garnering over 11.3 billion and 101.3 billion views respectively. The platform saw a 17% year-over-year increase in the number of book merchants generating revenue, with nearly 20% of them experiencing over 100% growth in transaction volume [4].

The report, analyzed by "China News Publishing & Broadcasting Journal", explores why publishers are gravitating towards TikTok E-commerce and its impact on the industry. It highlights the dual benefits of sales growth and profit, with live streaming and short videos offering a dynamic way to showcase books, thereby igniting reader interest and driving sales. The platform's ability to reflect user interest in trending topics also plays a significant role in boosting book sales, as seen with the surge in sales related to the animated film "Chang'an Twenty Thousand Li" last summer.

TikTok E-commerce has also expanded the influence and reach of publishers, becoming a premier platform for new book promotions. Notable authors and literary figures have taken to the platform to recommend books, enhancing cultural consumption and offering a more accessible way for readers to discover new titles. The platform's support for big data has significantly contributed to promoting reading among younger audiences, with a notable increase in book purchases among post-80s and post-90s generations [5].

Furthermore, the report emphasizes the importance of a healthy ecosystem for sustained platform growth, detailing efforts to protect copyright and promote legitimate publishing through various initiatives and collaborations. These efforts aim to support the healthy development of the publishing industry while ensuring a safe environment for children's reading and fostering a culture of legal book consumption. TikTok E-commerce's proactive measures in intellectual property protection and its "Good Books for Everyone" campaign underline its commitment to fostering a robust and positive ecosystem for book sales and consumption.

### **1.3. The current research status of e-commerce data mining**

E-commerce data mining, a crucial aspect of leveraging the "big data" wave that swept across the internet landscape in 2012, has seen varied applications and challenges across domestic and international markets. In China, despite the booming interest in big data around 2013, small and medium-sized enterprises (SMEs) have struggled to harness their potential fully [6]. The optimism surrounding e-commerce and big data did not translate into significant growth for many SMEs, primarily due to short-term investment mindsets, lack of strategic planning, and inadequate follow-through on e-commerce initiatives. The key value of big data lies not in the abundance of data itself but in the insights and cost-saving operational efficiencies it can uncover. Moving forward, the era demands refined management practices over broad-stroke marketing strategies, emphasizing precision in business operations through data analysis.

Contrastingly, in international contexts, collaboration and shared data resources among companies have marked a progressive trend towards understanding consumer behaviors and tailoring product offerings accordingly. A notable example is in Japan, where enterprises like Yahoo Japan and ASKUL have partnered with major manufacturers like P&G and Ajinomoto to analyze consumer purchasing habits and search records. This collaborative effort aims to break down competitive barriers and foster the development of products that cater to specific customer preferences and scenarios. By pooling resources, including sales data and search trends, these companies can accelerate product development and introduce innovations that meet emerging consumer demands more effectively.

This approach of leveraging shared big data for product development and marketing is increasingly becoming a norm in e-commerce. For instance, Amazon Japan's collaboration with Kagome to sell tomato juice and Rakuten's potential future initiatives highlight a growing trend towards data-driven product innovation. As consumer preferences diversify, making it challenging to predict bestsellers, the synergy between food, daily necessities manufacturers, and online enterprises is expected to expand, demonstrating the transformative power of big data in e-commerce across different market dynamics.

In summary, while domestic enterprises, particularly SMEs in China, face challenges in adopting and benefiting from e-commerce data mining due to strategic and operational limitations, international examples showcase the potential of collaborative and data-driven approaches in enhancing product development and marketing strategies. The contrast between domestic struggles and international successes underscores the need for a strategic shift towards more integrated and analytical approaches to e-commerce, emphasizing the critical role of big data in shaping the future of the industry.

## **2. Methodology**

Big data serves as a foundational resource and advantage for the future development of e-commerce platforms, offering informational support and fostering a "snowball effect" through the continuous collection of effective data and data circulation. This process is not limited to a single product

component or business area within e-commerce platforms. Despite past limitations due to technology, methods, and industry characteristics, which confined massive data collection primarily online, offering e-commerce businesses a short-term processing window, it's crucial to recognize that the largest source of market data lies offline. The essence of e-commerce is to serve traditional commerce processes electronically, enhancing efficiency and reducing costs. Establishing direct offline data collection interfaces, rather than relying solely on online data, is a key step. Merging online and offline data collection into data centers enhances analysis, feeding back into online platforms for more precise utility. This research collected data through questionnaires distributed via online communication tools for data mining analysis, focusing on the K-Means clustering analysis model.

## 2.1. Principles and implementation steps of questionnaire design

The design of questionnaires hinges on the goal to collect either exploratory (qualitative) or confirmatory (quantitative) information. Exploratory surveys, suited for qualitative data, may not require a formal questionnaire but rather a guide with open-ended questions to facilitate in-depth discussions. Conversely, confirmatory surveys necessitate a formal, standardized questionnaire with prescribed questions and response formats to ensure consistency and facilitate statistical analysis.

Effective questionnaire design is crucial and should:

- (1) Align with the research objectives, avoiding omissions due to inadequate preparation or understanding.
- (2) Capture complete and accurate information, designed to be easily understood by respondents, minimizing refusal or biased responses.
- (3) Facilitate easy response from participants, ensuring data is straightforward to analyze and interpret.
- (4) Remain concise and engaging to keep respondents interested throughout the survey process.

Conducting a survey involves the following nine basic steps:

- (1) Determine the information required.
- (2) Define the target respondents.
- (3) Choose the method to achieve the objective.
- (4) Determine the content of the questions.
- (5) Consider the wording of the questions.
- (6) Arrange the questions in a meaningful order and format.
- (7) Review the length of the questionnaire.
- (8) Test the questionnaire before the formal survey.
- (9) Determine the final form of the survey.

## 2.2. Data collection and sample characteristics description

The data collection for this study was conducted through an online questionnaire distributed via social media platforms, with a specific focus on TikTok users. The questionnaire was designed to capture a comprehensive range of responses concerning the users' demographics, TikTok usage habits, and their online shopping behaviors. The survey received a total of 132 valid responses over a period of two weeks, allowing for a diverse and representative sample of the TikTok user base. Below is a detailed breakdown of the sample characteristics based on the analysis of the collected data:

### **Gender Distribution:**

Female: 84.09% ; Male: 9.09% ; Prefer not to say: 6.82%

The sample predominantly consists of female respondents, highlighting the platform's appeal among women.

**Age Distribution:**

Ages 18-24: 92.42%; Below 18: 5.30%; Over 45: 2.27%

The majority of the respondents fall within the 18-24 age group, indicating that TikTok's user base skews younger.

**Occupation Distribution:**

Students: 96.97%; Other: 2.27%; Teachers: 0.76%

A significant portion of the sample identifies as students, underscoring the platform's popularity among the youth and academic communities.

**Education Level Distribution:**

Bachelor's Degree: 66.67%; High School or Below: 19.70%; Associate Degree: 7.58%

Master's Degree or Above: 6.06%

Most participants possess at least a bachelor's degree, reflecting a relatively well-educated user demographic.

**Daily Time Spent on TikTok:**

More than 2 hours: 33.33%; Less than 30 minutes: 25.00%; 1-2 hours: 24.24%

30 minutes to 1 hour: 17.42%

A third of the users spend more than 2 hours daily on TikTok, indicating high engagement levels with the platform.

**Discovery of New Products:**

The primary mode of discovering new products is through TikTok's recommended videos (26.52%), followed by a mix of recommended videos, live streaming, search, and friend's shares.

**Shopping Frequency on TikTok:**

1-2 times per month: 51.52%; 1-2 times per week: 21.97%; Never: 21.21%

Almost daily: 5.30%

Most users shop on TikTok 1-2 times per month, highlighting periodic engagement with the platform's e-commerce features.

**Main Factors Influencing Purchase Decisions:**

Price (15.91%) and a combination of price, product reviews, and seller reputation (11.36%) are significant factors, showcasing the importance of cost and trust in purchasing decisions.

**Types of Products Purchased:**

Apparel and beauty products are among the most commonly purchased items, indicating these categories' popularity on the platform.

**Concerns When Shopping on TikTok:**

Product quality is the primary concern for 80.30% of users, emphasizing the importance of trust in product authenticity and quality assurance.

**Purchasing Through Live Streaming:**

62.12% of respondents occasionally purchase products through live streaming, demonstrating the effectiveness of this feature in driving sales.

### **Product Review Habits:**

Nearly half of the users (48.48%) occasionally leave reviews, indicating a balanced approach to feedback sharing after purchase.

This detailed sample characteristic description provides insights into the demographics, usage habits, and shopping behaviors of TikTok users. These findings underscore the platform's significance in the digital commerce ecosystem and its influence on modern consumer habits.

## **2.3. Theoretical Foundations and Implementation Process of Cluster Analysis**

### **2.3.1. Purchasing Behavior**

The purchasing behavior of consumers in e-commerce is influenced by a variety of factors, both internal and external. Internal factors refer to the user's own conditions or perceptions towards products and purchasing actions, which are the main issues affecting purchasing behavior. External factors are limited to the shopping environment created by e-commerce platforms, which have been found to have a significant impact in empirical research.

In summary, the factors can be categorized as follows:

(1) **Social Factors:** These have a significant influence on user purchasing behavior, including social class and environment. **Class:** It's undeniable that differences exist among users in terms of job positions, income, education level, and titles. Users within similar categories can be classified into the same social class, sharing common values, social views, purchasing tendencies, tastes, and methods. Hence, e-commerce sales strategies need to be tailored to different classes. **Environment:** Individuals are inevitably influenced by colleagues, friends, and family, which over time can form habits. This also applies to purchasing behavior, where people in the same environment tend to have similar purchasing desires and methods.

(2) **Cultural Factors:** The individual's cultural background. An individual's living environment helps establish fixed behavior patterns and value orientations, which also translate into product needs in purchasing behavior. These cultural differences include race, ethnicity, religion, and geography.

(3) **Family Factors:** The purchasing power of a family cannot be ignored. Families can be diverse, including harmonious and conflicted, newlywed and unmarried, single-parent, affluent, and poor. These factors have an obvious impact on purchasing behavior.

(4) **Personal Factors:** These refer to individual characteristics, such as gender, age, personality, preferences, and tastes. Cultural, social, and family factors mentioned also encompass personal aspects like cultural orientation, economic income, and social status, which determine an individual's purchasing power and willingness.

### **2.3.2. K-Means Algorithm**

The K-Means algorithm is a classic method within cluster analysis, also known as a centroid-based algorithm technique. Its process is as follows: Initially,  $k$  objects are randomly selected from the dataset  $D$ , each representing the initial center of a cluster. For every other object, its Euclidean distance to each cluster center is calculated, and the object is assigned to the closest cluster. Subsequently, the K-Means algorithm iteratively improves the variation within clusters. For each cluster, it uses the objects that were iteratively classified to that cluster to calculate a new mean or center. Then, the updated mean is used as the new cluster center, and all objects are reclassified. The iteration continues until the assignment stabilizes, and the final updated cluster is the same as in the previous round. This paper adopts the Euclidean distance, with the sum of squared errors between all sample points in the cluster and the cluster center, defined as:

$$E = \sum_{i=1}^k \sum_{p \in C_j} \text{dist}(p, C_j)^2 \quad (1)$$

Here,  $E$  is the sum of squared errors for all objects in the dataset;  $P$  represents a point in space, denoting a given data object; and  $C_j$  is the centroid of cluster  $C_j$  (both  $P$  and  $C_j$  are multidimensional). In other words, for each object in a cluster, the square of the distance between the object and its cluster center is calculated, and then these values are summed. This objective function aims to make the resulting clusters as compact and distinct as possible.

### 3. Cluster Analysis Results and Discussion

Through the application of K-Means clustering on the dataset derived from the survey responses of TikTok users, I have successfully segmented the users into three distinct groups based on their online shopping behavior. This section delves into the characteristics that define each cluster, providing insights into their preferences and behaviors on the TikTok platform.

#### 3.1. Description of Characteristics of Different User Groups

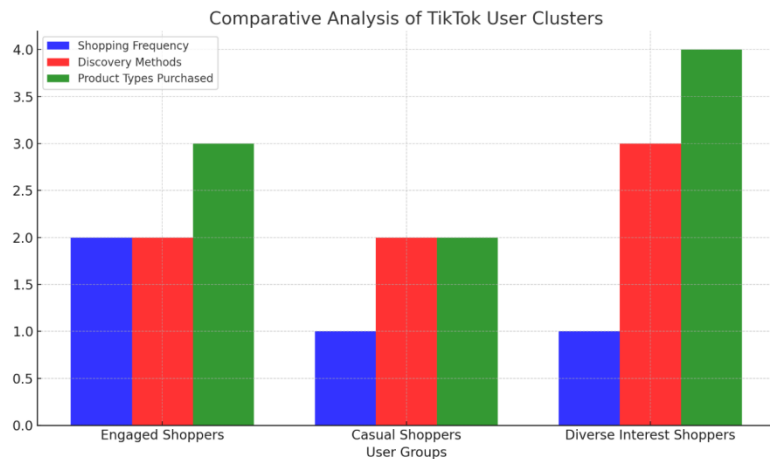


Figure 1: Different User Groups

The analysis revealed three clusters, each representing a unique user group with distinct shopping patterns (Figure 1).

**Cluster 0 - Engaged Shoppers:** This group consists predominantly of female users aged 18-24, with diverse occupations but largely holding a bachelor's degree. They spend 30 minutes to 1 hour daily on TikTok, frequently shopping on the platform 1-2 times per week. They are influenced by a variety of factors including product price, reviews, seller reputation, the recommendation algorithm, friends' suggestions, and livestream shopping. They mainly purchase apparel, beauty products, and electronics, with product quality being their primary concern. Their engagement with live streaming shopping is occasional, and they rarely leave reviews after making a purchase.

**Cluster 1 - Casual Shoppers:** Characterized mainly by female students aged 18-24 with bachelor's degrees, these users also spend 30 minutes to 1 hour daily on TikTok but shop less frequently, typically 1-2 times per month. The discovery of new products is through recommended videos and friends' shares. Their purchasing decisions are influenced by product price, reviews, seller reputation, recommendation algorithms, and friends' suggestions. Beauty products and household items are their primary purchases, with product quality being a major concern. Similar to Cluster 0, they occasionally engage in livestream shopping but tend not to leave reviews.

Cluster 2 - Diverse Interest Shoppers: This cluster is similar in demographic composition to the others, with a majority being female users aged 18-24 from various occupations. They spend 30 minutes to 1 hour on TikTok daily and shop 1-2 times per month. They discover new products through a combination of recommended videos, live streaming, and search functions. Their shopping is influenced by reviews, seller reputation, friends' recommendations, and live streaming. This group buys a wide range of products, including apparel, household items, food, and books, with product quality as their top concern. They occasionally purchase through live streaming and rarely leave reviews.

### 3.2. Correlation Analysis between Users' Characteristics and Shopping Behaviors

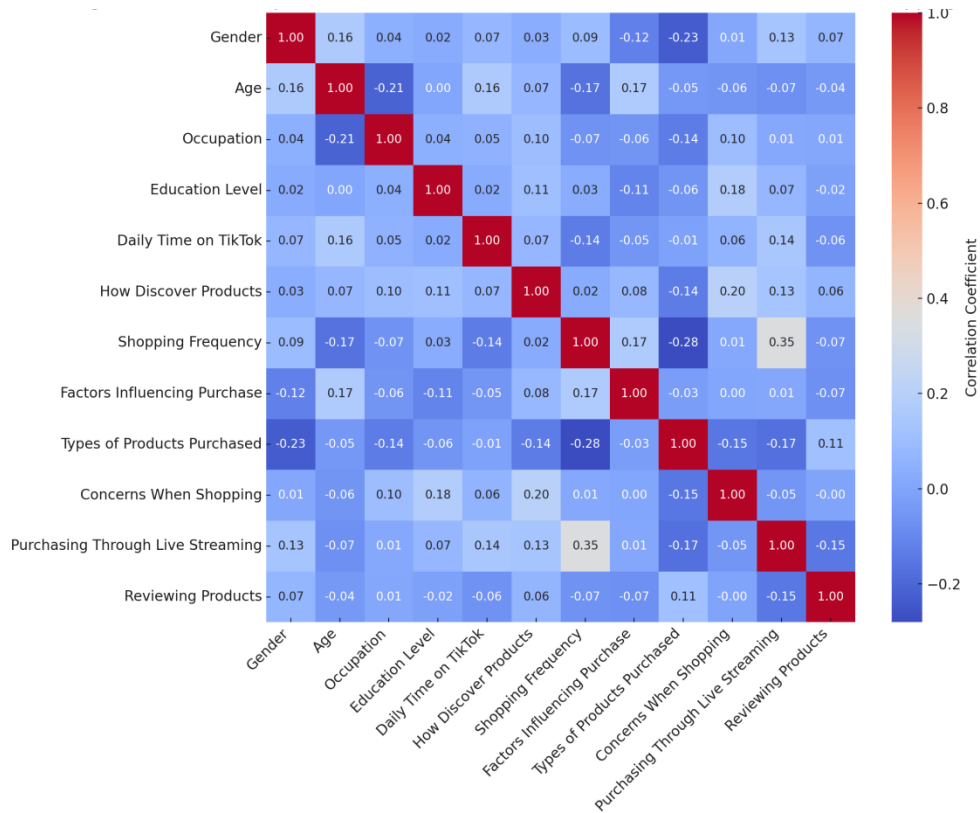


Figure 2: Correlation coefficient



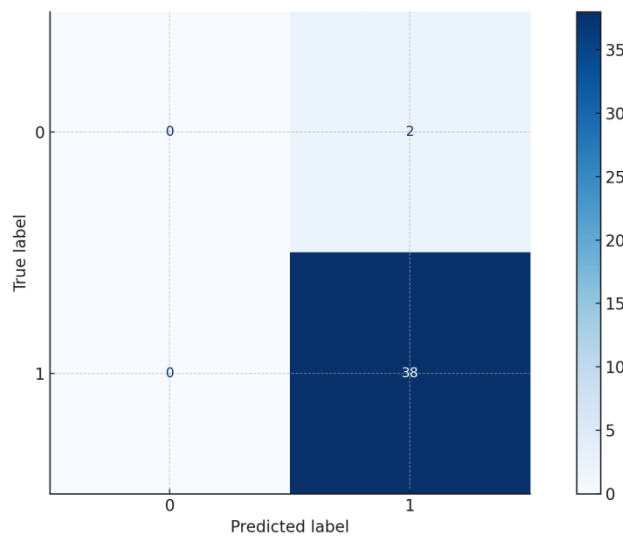


Figure 3: True-predicted label

This section of the study delves into the relationship between user characteristics and their shopping behavior on the TikTok platform. The analysis employs statistical methods to quantify the strength and direction of associations between various user demographics (such as age, gender, and education level) and their purchasing patterns (Figure 2). The investigation draws upon a dataset comprising responses from 132 users, meticulously collected through an online survey.

The correlation analysis was conducted using Python's pandas, seaborn, and matplotlib libraries for data manipulation and visualization. Initially, categorical data were encoded into numerical formats suitable for correlation analysis. Key user characteristics analyzed included gender, age, occupation, education level, and daily time spent on TikTok (Figure 3). These were correlated with shopping behavior indicators such as shopping frequency, main factors influencing purchasing decisions, types of products purchased, and engagement with live streaming for shopping.

The correlation matrix revealed several noteworthy relationships: (1) Shopping Frequency and Live Streaming: A positive correlation (0.35) was identified between the frequency of shopping on TikTok and the likelihood of purchasing through live streaming. This suggests that users who shop more frequently are also more inclined to engage with livestream shopping experiences. (2) Types of Products and Shopping Frequency: A negative correlation (-0.28) between the diversity of products purchased and shopping frequency indicates that users who shop more frequently tend to purchase a broader range of products. (3) Age and Shopping Frequency: The analysis uncovered a negative correlation (-0.17) between user age and shopping frequency, suggesting younger users are more active shoppers on TikTok compared to older users.

The logistic regression model aimed to predict the likelihood of users making a purchase based on their characteristics. The model, trained on gender, age, education level, and daily TikTok usage, achieved an accuracy of 95%. However, it predominantly classified users as frequent shoppers (at least once a month), highlighting a potential bias towards the majority class in the dataset.

## 4. Commercial Applications of Analysis Results and Strategic Recommendations

### 4.1. Tailored Marketing Campaigns

Segment-Specific Strategies: Develop marketing strategies that are specifically tailored to the distinct user groups identified in the cluster analysis. For Value Seekers, emphasize affordability and value

for money. For Engaged Shoppers, focus on showcasing product quality and brand reputation. And for Social Influencers, leverage user-generated content and social proof to drive engagement.

**Personalized Advertising:** Utilize TikTok's robust data analytics to deliver personalized ads to users based on their past engagement and shopping behaviors. This could involve showing users ads for products they've viewed but haven't purchased, or products similar to those they have shown interest in.

#### **4.2. Enhancing Product Discovery**

- **AI-Driven Recommendations:** Leverage artificial intelligence to refine product recommendation algorithms, providing users with personalized product suggestions based on their browsing and shopping history. This approach could significantly enhance user experience and increase sales conversion rates.
- **Interactive Content:** Employ interactive content such as polls, quizzes, and challenges to engage users and subtly guide them towards discovering new products that match their preferences and behaviors.

#### **4.3. Content Strategy Optimization**

- **Diverse Content Creation:** Create a diverse range of content tailored to the preferences of different user clusters. For example, product tutorials and reviews for Engaged Shoppers, and entertaining, shareable content for Social Influencers.
- **Influencer Collaborations:** Partner with influencers who resonate with your target segments to amplify your message and reach. Influencers can help lend credibility to your brand and products, making your offerings more appealing to potential buyers.

#### **4.4. Strategic Business Alliances**

- **Cross-Promotion Partnerships:** Forge partnerships with other brands and influencers to tap into new audiences and cross-promote products. This strategy can help brands leverage each other's strengths and user bases for mutual benefit.
- **Technology Integration:** Explore integrations with emerging technologies such as augmented reality (AR) for virtual try-ons or immersive product experiences, which can significantly enhance the online shopping experience on TikTok.

#### **4.5. Leveraging Big Data for Decision Making**

- **Analytics for Insight Generation:** Utilize TikTok's comprehensive analytics tools to continuously monitor user engagement and shopping behaviors. Insights gathered can inform product development, marketing strategies, and content creation to better meet the needs and preferences of your target audience.

### **5. Conclusion**

This study presents a comprehensive analysis of TikTok's impact on e-commerce, offering insights into user behavior and platform strategies. However, it acknowledges several limitations: the sample size and diversity are limited, potentially not capturing the full spectrum of TikTok's user base. The research focuses on a specific geographic area, which may overlook the global intricacies of TikTok shopping habits. Additionally, the rapid evolution of social media and e-commerce could render the findings less relevant over time.

To address these limitations and advance understanding, future research should aim to expand the geographic scope to include a more diverse array of cultures and regions, providing a global perspective on TikTok user behavior. Longitudinal studies are recommended to track how user behaviors and platform features evolve, offering insights into shifting trends in the social media e-commerce landscape. Exploring the impact of technological advancements like augmented reality (AR), virtual reality (VR), and artificial intelligence (AI) could provide a forward-looking view of online shopping experiences. Moreover, comparing shopping behaviors across different social media platforms would help identify unique challenges and opportunities, refining e-commerce strategies.

By tackling these avenues, future research can deepen the understanding of social media's role in e-commerce. This study lays a foundational path for continued exploration and innovation at the intersection of social media and online shopping, aiming to optimize engagement strategies and drive sales more effectively on platforms like TikTok. The conclusion and outlook underscore the need for a broader, more adaptive approach to researching TikTok's e-commerce potential, suggesting a multifaceted strategy for scholars and practitioners to enhance user engagement and commercial success in the rapidly evolving digital marketplace.

## References

- [1] Juncheng Zhang. "Douyin E-commerce Data Report Provides Answers" [N]. *China News Publishing and Broadcasting News*, 2024-02-05(007). DOI: 10.28907/n.cnki.nxwcb.2024.000267.
- [2] Meiling Liang. "Challenges and Optimization Strategies of Social Media Marketing in Cross-border E-commerce" [J]. *China Business Review*, 2023, (23): 53-56. DOI: 10.19699/j.cnki.issn2096-0298.2023.23.053.
- [3] Xu Li. "Prospects for Research on Socialized E-commerce User Decision Behavior" [J]. *New Economy*, 2022, (01): 46-49.
- [4] Yiwun Sun, Ronglei Luo. "Evaluation and Clustering Analysis of Live Broadcasting Competitiveness of Women's Clothing Brands—Based on Douyin Platform" [J]. *Journal of Zhejiang Textile and Garment Vocational Technical College*, 2023, 22(01): 32-36.
- [5] Sui Yi, Zongge Du. "Development of Natural Gas Customer Segmentation Model" [J]. *Oil & Gas and New Energy*, 2021, 33(02): 67-71.
- [6] Puxiong Zhou, Chufan Gong. "Research on Recommendation Model Based on Dynamic User Interest Tags" [J]. *New Century Library*, 2021, (09): 65-69. DOI: 10.16810/j.cnki.1672-514X.2021.09.010.