Financial Economics Research; Vol.1, No.1; 2024
E-ISSN: 3006-3663 P-ISSN: 3006-3655
DOI: https://doi.org/10.70267/t59bcb11
Published by: Zeus Press

E-commerce Platform in Digital Economy: Analysis and Optimization of Amazon's Footwear Store

Bowen Xiong*, Jianwei Sun and Xiangwen Pang

Tianjin University of Finance and Economics, Tianjin 300222, China

**Corresponding author: Bowen Xiong , E-mail:478503648@qq.com*

Abstract

With the continuous deepening of globalization, cross-border e-commerce has become an important part of global trade. Especially driven by digital payment technology, network technology, industrial upgrading and policy support, China's cross-border e-commerce industry has developed rapidly. This research aims to analyze the sales data of men's footwear on the Amazon platform, understand market trends, user behavior patterns, and product characteristics, and then provide data-supported marketing strategies and operational decisions for e-commerce platforms. We have adopted a variety of data analysis techniques, including data cleaning, ARIMA model prediction, price band analysis, repeat purchase rate analysis, user profiling (RFM model), sentiment analysis and cluster analysis. Through comprehensive analysis of multi-dimensional information such as store sales data, product attributes, market hot words, and user evaluations, market dynamics and user preferences are revealed. We have comprehensively applied a variety of data analysis models and techniques, such as using the ARIMA model for sales prediction, the LDA model for sentiment analysis, and the RFM model combined with cluster analysis to construct a refined user profile. Through in-depth analysis of user transaction behaviors and product attributes, accurate marketing strategies can be provided for e-commerce platforms, enhance user experience, increase user stickiness, and thus achieve business growth and improvement of profitability. At the same time, the research also points out the importance of brand building, supply chain management, risk assessment and other aspects, providing strategic suggestions for the long-term stable development of cross-border e-commerce.

Keywords

Amazon cross-border e-commerce, men's footwear industry, data analysis, marketing optimization

1. Introduction

1.1 Current Status of Research at Home and Abroad

1.1.1 Research Review

In China, research on in-depth analysis of e-commerce platform user behavior and demand prediction models is gradually deepening. Researchers use big data technology to collect and analyze user behavior data on e-commerce platforms to reveal user needs and behavior patterns. These studies help e-commerce platforms better understand user needs, optimize products and services, and improve user satisfaction and loyalty. Domestic research also involves the collection, processing, analysis, and mining of user behavior data to understand users' interests, needs, preferences, and other information and apply it to platform

operation and decision-making.

According to "In-Depth Research on Digital Business Empowering the Development of Cross-border E-commerce Platform Enterprises", research on in-depth analysis of user behavior and demand prediction models of e-commerce platforms from the perspective of the digital economy(Xu Limei, 2020) can be integrated into a coherent framework covering aspects such as the digital transformation, supply chain management, market expansion, dispute resolution mechanism, credit system construction, and service trade development of cross-border e-commerce platforms.

This research first focuses on how cross-border e-commerce platforms can optimize user behavior analysis and demand prediction through the application and upgrading of digital technologies. This involves using technologies such as big data and artificial intelligence to deeply understand user behavior patterns, predict market demand, and provide personalized services and product recommendations accordingly.

At the same time, it also emphasizes the importance of supply chain system construction, pointing out that cross-border e-commerce platforms need to build an efficient and transparent supply chain by integrating logistics, payment, customs clearance, and other links to improve user satisfaction and reduce operating costs.

The literature "Research on Logistics Cost Management of Self-operated E-commerce Platforms in the Digital Economy Era - Taking Suning Tesco as an Example" mainly discusses how Suning Tesco, as a self-operated e-commerce platform, optimizes logistics cost management by applying digital information technology in the context of the digital economy. It is pointed out that logistics cost management is a key link in the operation of Suning Tesco. Its optimization not only involves internal value chains such as procurement, warehousing, distribution, and after-sales service but also includes external value chains that interact with suppliers and customers. The literature emphasizes that the advantages of Suning Tesco in logistics cost management include effectively reducing distribution costs, improving the scientific nature of decision-making, and accurately predicting and processing massive data information. To find out the main factors affecting logistics costs, Suning Tesco uses the financial indicator method and analytic hierarchy process (AHP) for analysis and formulates targeted improvement measures accordingly. These improvement measures include optimizing resource allocation, integrating internal and external resources, and reducing unnecessary expense expenditures to create higher benefits for enterprises. Through these measures, Suning Tesco can improve the efficiency and effectiveness of logistics cost management and achieve optimization of enterprise operations.

Foreign research also attaches great importance to the construction of user behavior analysis and demand prediction models. Researchers conduct in-depth analysis of user behavior data through advanced data mining and machine learning technologies to predict users' purchase intentions and needs. These studies emphasize the importance of personalized recommendation systems and how to optimize product design and user experience through user behavior data. In the literature "New marketing strategy model of E-commerce enterprises in the era of digital economy", it first examines the impact of the digital economy on e-commerce enterprises, especially the transformation of marketing models, and analyzes the development trends of e-commerce.

To meet the needs of the main social consumers, e-commerce enterprises need to formulate a comprehensive marketing strategy based on big data and multiple channels and centered on customers. The literature discusses the challenges faced by e-commerce enterprises, such as lack of integrity and homogeneity problems, and proposes corresponding marketing strategies, including product management architecture, customized marketing content, and social media marketing.

According to "SWOT Analysis: Growth of E-commerce Within the Context of Digital Economy(Kumar, Amita;Sharma, Nehaa;Gurna, Kamalpreet Kaurb;Anand, Abhineetc;Patni, Jagdish ChandradCAa; Pinjarkar & Latikae, 2024), it examines the impact of the digital economy on e-commerce enterprises, especially the transformation of marketing models, and analyzes the development trends of e-commerce. To meet the needs of the main social consumers, e-commerce enterprises need to formulate a comprehensive marketing strategy based on big data and multiple channels and centered on customers. The literature discusses the challenges faced by e-commerce enterprises, such as lack of integrity and homogeneity problems, and proposes corresponding marketing strategies, including product management architecture, customized marketing

content, and social media marketing.

In addition, business transformation and technological innovation are also one of the research focuses. The literature "Cognitive Business: A New Frontier of Business Revolution and E-Business Research" discusses how cognitive computing and artificial intelligence promote the evolution and innovation of the digital economy(Ma Shuzhong, Fang Chao & Liang Yinfeng. 2018) ecosystem and proposes "cognitive business" as a new frontier of digital business transformation. Based on an examination of the characteristics of business transformation and the new stage of e-commerce evolution, the concept of "e-commerce 4.0 (eBiz4.0)" is proposed. Cognitive business is regarded as the core issue of the new generation of e-commerce. The literature defines cognitive business from the perspectives of service and system and analyzes its driving forces and scientific, technological, and social foundations. At the same time, it also discusses the challenges faced by cognitive business research and practice.

1.1.2 Literature Review

From the marketing strategy perspective: In the digital economy era, how can e-commerce enterprises attract and retain customers through innovative marketing strategies? This may include developing personalized shopping experiences, using social media platforms for brand promotion, and using data analysis to optimize marketing activities. It emphasizes the key role of digital technology in shaping modern marketing strategies. For example, personalized recommendations are made through machine learning algorithms, market trend analysis is conducted using big data, and social media analysis is used to enhance customer engagement and brand loyalty. (As exemplified in "Research on Logistics Cost Management of Self-operated E-commerce Platforms in the Digital Economy Era - Taking Suning Tesco as an Example" and "New marketing strategy model of E-commerce enterprises in the era of digital economy").

From the economic development perspective: The digital economy can provide impetus for the growth of e-commerce, including exploring new business models, market expansion strategies, and the evolution of consumer behavior. It discusses how e-commerce can promote economic growth through technological innovation, global market access, and consumer-oriented services. At the same time, it may point out the potential of e-commerce in promoting employment, creating new market opportunities, and promoting the development of small businesses. (As seen in "SWOT Analysis: Growth of E-commerce Within the Context of Digital Economy" and "In-Depth Research on Digital Business Empowering the Development of Cross-border E-commerce(Zhai Huafeng, 2020) Platform Enterprises").

From the technology application perspective: Scholars discuss how e-commerce enterprises use digital technologies such as cloud computing, big data analysis, and artificial intelligence to improve service quality, enhance user experience, and increase operational efficiency. It emphasizes that technological innovation is the core driving force for the development of e-commerce. The literature may discuss how to improve data processing capabilities, automate customer service, predict market trends, and develop new business models such as blockchain-based supply chain management. (As mentioned in "Cognitive Business: A New Frontier of Business Revolution and E-Business Research").

In the above research fields, although in-depth insights and strategies are provided, each field may have some weaknesses. For example, marketing strategy research may not fully consider the differences in consumer behavior under different cultural and market backgrounds, resulting in insufficient analysis of specific marketing strategies for cross-cultural or international markets. Economic impact research may focus on the positive impacts of e-commerce, and not explore deeply enough the potential negative impacts on traditional industries, especially traditional retail and the job market.

1.2 Significance of Research

Our research is committed to in-depth analysis of user behavior data, so that e-commerce platforms can more accurately understand user needs and preferences, and then be able to provide more personalized products and services, thereby greatly improving user satisfaction and loyalty. At the same time, the demand prediction model can strongly assist e-commerce platforms to predict market trends and user needs, so as to optimize inventory management and resource allocation, reduce operating costs and improve efficiency. In the highly competitive e-commerce market environment, in-depth and accurate user behavior analysis and demand prediction can enable platforms to have stronger competitive advantages in product recommendation,

marketing strategy and customer service. Moreover, through our research, domestic and foreign small and medium-sized retail e-commerce can have a deeper understanding of the market and can inject energy into traditional retail and promote the development of new-quality productive forces. At the same time, it also strongly promotes e-commerce platforms to carry out digital transformation by using technologies such as big data and artificial intelligence to better adapt to the development requirements of the digital economy era. E-commerce enterprises can attract and retain customers and enhance brand influence by means of innovative marketing strategies such as personalized recommendation and social media marketing.

As an important part of the digital economy, e-commerce promotes economic growth through technological innovation and global market access, creates employment opportunities and promotes the development of small businesses. Our research can also further explore how to use emerging technologies such as artificial intelligence and machine learning to improve the accuracy of consumer behavior prediction and the ability to analyze market trends. And with the globalization of the e-commerce market, future research needs to focus on how to implement effective digital marketing strategies in different cultural and regulatory environments, and how to overcome challenges such as logistics, payment and cultural differences faced by cross-border e-commerce(Liqin Wen;Shenglin Ma1CA & Shuping Lyu1, 2024). In addition, in view of the continuous growth of consumers' demand for personalized and customized products, researching how to organically combine customer relationship management systems with personalized marketing strategies to improve customer experience and loyalty is undoubtedly also a field worthy of great attention. Finally, with the rise of emerging social media platforms, future research should also pay attention to the marketing potential of these platforms and how to achieve effective integration of online and offline channels to provide a seamless shopping experience.

1.2.1 Theoretical Significance

Through in-depth mining of user behavior data, we have opened up a new theoretical perspective for understanding consumer behavior, which not only helps enrich the theoretical foundation of disciplines such as marketing, consumer psychology and information systems but also strongly promotes the renewal and development of these fields. This research spans many disciplines such as big data, artificial intelligence, and supply chain management, enhancing interdisciplinary integration and communication, and providing innovative research paths and methodologies for the academic community.

The research results have direct application value in the operational decision-making of e-commerce platforms, can effectively improve the level of enterprise personalized services, realize the optimization of resource allocation, reduce operating costs, and improve overall efficiency, thereby significantly enhancing the practical application effectiveness of academic research. At the same time, the research also promotes the extensive application of emerging technologies such as big data and artificial intelligence in the field of e-commerce, points out the direction for technological innovation, and plays a positive role in promoting the technological progress of the e-commerce industry.

In the context of globalization, we not only focus on the domestic market but also extend to the international market, providing strategic ideas for the international development of small and medium-sized retail e-commerce and helping to improve the international competitiveness of domestic enterprises. Facing the globalization trend of the e-commerce market, the research provides practical strategic guidelines for the challenges faced by cross-border e-commerce in logistics, payment and cultural differences, providing strong support for e-commerce operations in the context of globalization.

At the same time, the research also clearly indicates the key areas of future research, including digital marketing strategies, personalized marketing, and the marketing potential of social media platforms, providing a clear direction for follow-up research and providing valuable references and inspirations for the academic and practical fields. Through these in-depth and comprehensive studies, it has a profound impact on the academic field and also provides strong support for e-commerce practice, making positive contributions to social and economic development and consumer well-being.

1.2.2 Practical Significance

In the context of the digital economy, this project helps deepen the understanding of user behavior patterns and reveal the deep-seated needs and behavior characteristics of users on e-commerce platforms. Through in-depth mining of user behavior data, it can provide a basis for personalized recommendation,

optimize business processes, and improve user satisfaction and corporate revenue.

We improve user shopping experience and satisfaction through personalized recommendation and user behavior analysis. Demand prediction and supply chain management optimize operational efficiency and reduce costs. Through data analysis and prediction, help enterprises better grasp market opportunities and improve competitiveness.

Secondly, this kind of research helps enterprises better understand market needs and consumer behavior, and provides decision-making basis for the formulation of sales strategies and the optimization of products. For example, by analyzing user behavior data, it is possible to discover possible preference conversions and behavior diversions of users during the purchase process, and make targeted adjustments to improve overall efficiency.

In addition, this kind of research also helps promote technological innovation and service optimization of e-commerce platforms. Through the prediction of user behavior, e-commerce enterprises can achieve precise marketing, improve user satisfaction and retention rates, and thus maintain a leading position in the fierce market competition.

To sum up, the research on in-depth analysis of e-commerce platform user behavior and demand prediction models from the perspective of the digital economy(Oh, Jiua & Seong, ByeongchanaCAa. 2024) can not only enrich digital economy theory but also provide scientific guidance for the actual operation of e-commerce platforms, and has important theoretical and practical value.

2. Research Methods

As a professional footwear seller on the Amazon platform(Huang Yi, 2016) , we are acutely aware of the necessity to maintain stable operations and continuously retain a competitive advantage in the rapidly changing market. We have conducted in-depth data analysis specifically on men's footwear. Through comparison with market data, we have analyzed multiple aspects of men's footwear in the past year, including monthly sales trends, sales proportions of different sub-categories, changes in monthly store sales, the structure of men's footwear in the store, price range distribution, product characteristics, and consumer evaluations. These analyses help us understand the sales situation of men's footwear in our store, identify existing problems, and explore potential market opportunities so that we can improve our sales performance.

We have carried out a comprehensive analysis of Amazon's footwear industry stores. The purpose is to identify existing problems and advantages of the store by deeply exploring the trends of store and market sales, category structure and price segments, product conditions, and market hot-selling attributes, and then formulate effective marketing plans.

First, we obtain the sales data of an Amazon store for nearly a year and then start data processing. We properly handle missing values, outliers, and duplicate values in the data to ensure the accuracy of subsequent analyses. In the store sales analysis stage, we focus on comparing the trends of store and market sales to determine whether they are synchronized and whether there is a lag or lead. Next, we deeply study the category structure and price segments of the store. We evaluate their rationality and adaptability based on sales volume indicators. At the same time, we sort out the price distribution and compare it with competitors to determine whether the pricing strategy is appropriate. For store products, we not only need to find out hot-selling styles and analyze consumer public opinions but also pay attention to products with good user satisfaction but poor sales and explore the potential problems. At the same time, we insight into the market's hot-selling attributes and extract key information from search data by means of relevant models and methods. Finally, based on a series of the above analyses, we conduct SWOT analysis of the store, formulate SO, ST, WO, and WT strategies, and plan on-site and off-site promotion plans to enhance the store's competitiveness and improve sales performance. For example, in the analysis of price segments, we may find that the sales volume of products in certain price ranges of the store is scarce and there is a difference from the mainstream price range in the market. In the marketing plan, we may formulate a key on-site promotion plan for a certain type of product according to the store's advantages or an off-site plan to improve the brand image according to the disadvantages.

3. Definition and Processing of Missing Values

According to the data given in the question, this article defines missing values as the following two categories:

(1) The data in the data set is a value of 0.

(2) The data in the data set is a null value.

Perform data preprocessing according to the two missing value situations defined above respectively. The specific steps are as follows:

For the case where the data is 0, we analyze the original data and combine the actual situation to determine whether this value has practical significance. If this value is meaningful, choose to keep it; if this value is meaningless, treat it as a missing value and perform deletion processing. In this question, the values in the discount column are all 0, so this column is directly deleted.

For the case where the data is null, that is, if the data loss situation of this feature is higher than 10%, make a comprehensive judgment combined with the importance of this feature. If the importance of this value is low, it will be directly deleted; if the importance of this value is high, it will be filled according to the upper and lower data combined with the actual situation. In this question, the missing values of postal code data are supplemented according to the same-city order information. The missing values of shipping cost are set as the mean value. The missing values of sales volume are supplemented according to the same type of product and its shipping cost and profit. If it cannot be completely supplemented, it is set as the mode.

Definition and Processing of Outliers and Duplicate Values

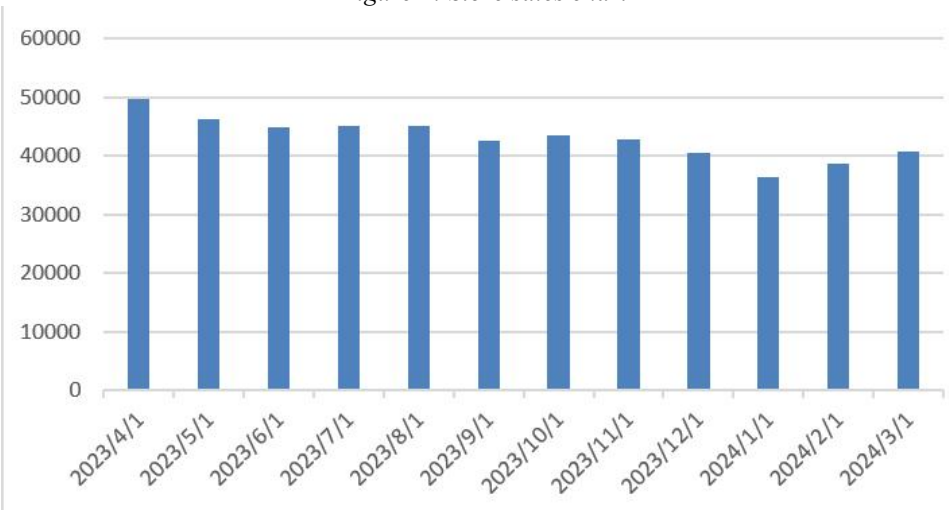
In this article, after completing the processing of missing values, the data set is checked for duplicate values and outliers again. It is found that there are no outliers and duplicate values that need to be deleted. The definition of duplicate values is that all fields used for analysis in the data set are consistent. As defined here, outliers are data in the data set that appear as special symbols or are extremely inconsistent with realistic factors.

After the cleaning process of data preprocessing, the data given in the question can be obtained in a standardized form and can be used normally, providing a rigorous data guarantee for the completion of the following tasks.

4. Sales Analysis of Amazon Footwear Industry Stores

First of all, we need to carry out sales analysis work for the store and present the monthly sales volume of the store intuitively in the form of a bar chart, so as to clearly grasp the sales status of the store.

Figure 1: Store sales chart



It can be seen that the overall sales volume of this store from April 2023 to March 2024 shows a fluctuating situation. The sales volume is the highest in April 2023. The sales volume of this store is relatively evenly distributed and there is no obvious seasonality.

Through data analysis, we compare the monthly sales volume trend of the store in the past year with the monthly sales volume trend of the market to observe whether there is a lag or lead phenomenon. We will perform the following steps:

Load data: Obtain store sales data and market sales data.

Data processing: Convert date data into time series and ensure that the time range of store and market data is consistent.

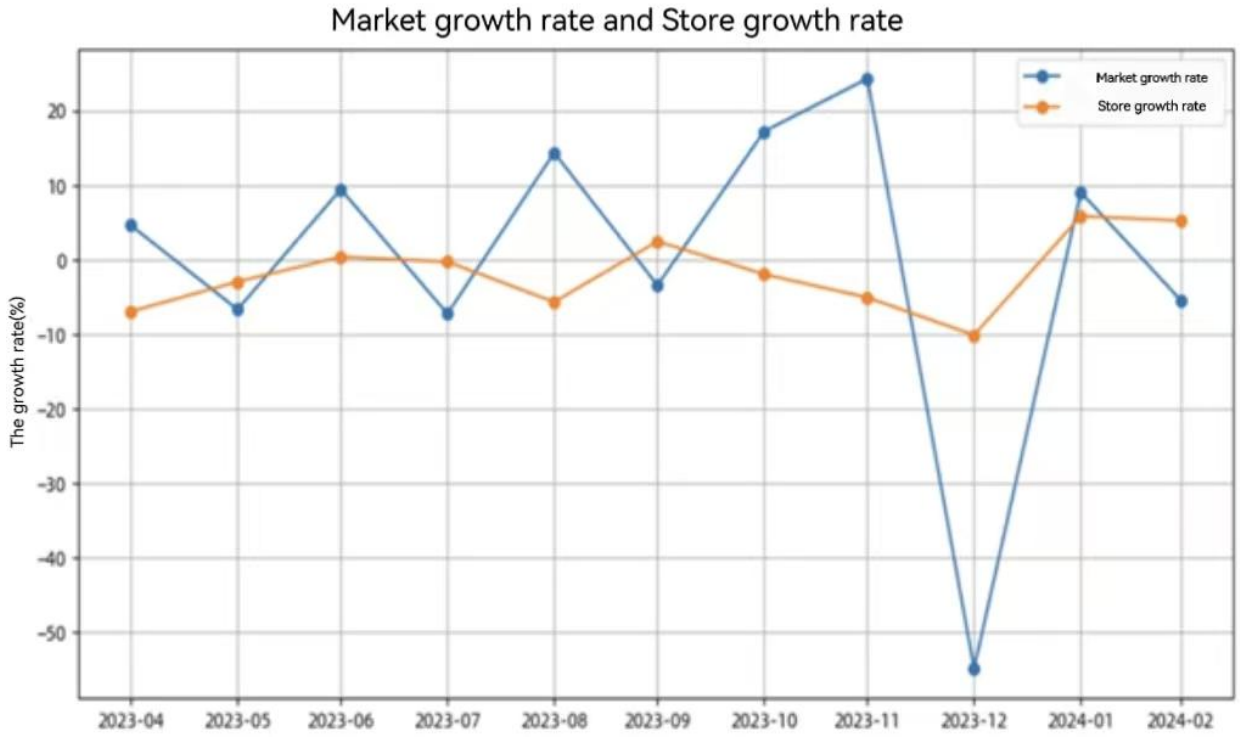
Trend analysis: Draw a time series(Wang Yingwei & Ma Shucai, 2021) trend chart and compare and analyze the changes in store and market sales.

Store sales (in US dollars): Displays the store sales for each month.

Store month-on-month growth rate: Displays the sales change rate relative to the previous month.

Market sales year-on-year growth rate: Displays the change rate of market sales relative to the same month of last year.time

Figure 2: Comparison chart of market sales year-on-year growth rate and store year-on-year growth rate



Comparing the two growth rate lines can help us analyze the relationship between the store sales trend and the overall market.

We can clearly find that the direction of store sales is basically consistent with the market trend. In order to sell more goods, we should always pay attention to market trends so as to increase sales.

5. Constructing an ARIMA Model

5.1 Model Principle

The ARIMA model is an autoregressive integrated moving average model. By adopting past observations and considering the components of differencing, autoregression, and moving average, it separates signals and

noise. The ARIMA model can be described as ARIMA(p, d, q), where:

p represents the order of the autoregressive model, indicating how many historical periods are used to predict the current value.

d represents the order of differencing required for the data. The differencing method can make the data more stationary.

q represents the order of the moving average model. It focuses on the accumulation of error terms in the autoregressive model and can effectively eliminate random fluctuations in predictions.

The basic principle of the ARIMA model is to establish an autoregressive moving average (ARMA) model that describes changes in a time series(Alsuwaylimi & Amjad A.aCAa, 2023) . It can view time series data as an autoregressive moving average model, which consists of two parts: the autoregressive part and the moving average part. First, it fits the time series data into an autoregressive model. The autoregressive model can represent the trend, periodicity, and regularity of time series data. Second, it fits the time series data into a moving average model. The moving average model can extract the random residuals in time series data. By combining the autoregressive model and the moving average model, an ARMA model is constructed, which can simultaneously describe the trend, periodicity, and residuals of time series data(Meertens, Q A 1, Diks, C G H 2, H J van den Herik 3 & Takes, F W 4. 2020) .

The ARIMA prediction model can be written as the following formula(Mingli Song 1, 2CA1; Ruobing Wang1 & Yan Li1, 2024) :

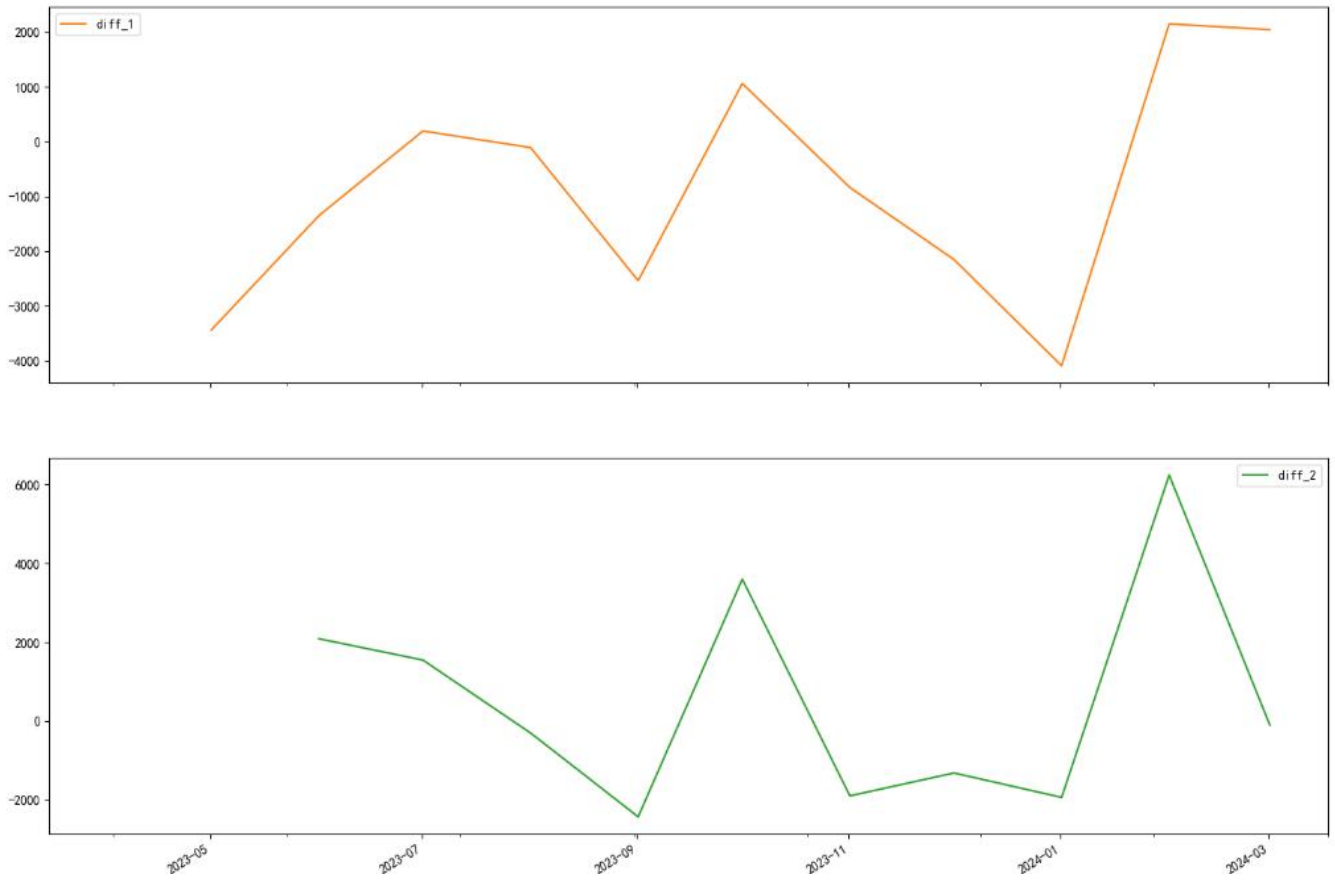
$$\hat{P}^{\{t\}} = P_0 + \sum_{j=1}^{p} \gamma_j p^{\{t-j\}} + \sum_{j=1}^{q} \theta_j \varepsilon^{\{t-j\}} \tag{1}$$

Here, P is the order of the autoregressive model (AR), q is the order of the moving average model (AM).

5.2 ARIMA(p, d, q)

First, by plotting a sequence chart, observe whether the data is stationary.

Figure 3: First-order and second-order difference charts



From the sequence chart, it can be seen that the stationarity of the original time series is relatively poor, while after first-order differencing, the stationarity is better. However, to determine the order of differencing more reasonably, the ADF unit root test is introduced here. The final result of using the unit root for testing is as follows:

Original data: ADF Test Statistic: 0.09840689634523289, p-value:

Table 1: ADF unit root test result table

raw data	ADF Test Statistic: 0.09840689634523289	p-value: 0.9659521570442503
First order difference	ADF Test Statistic: -3.175695395437702	p-value: 0.021435337399735042
Second order difference	ADF Test Statistic: -3.3878424171224113	p-value: 0.011380371043326506

Finally, it can be determined that the parameter d (differencing order) in the ARIMA model is 1. Then, by plotting the autocorrelation function (ACF) and partial autocorrelation function (PACF) charts, determine the specific values of parameter p (autoregressive order) and parameter q (moving average order).

Figure 4: PACF chart

Figure 5:ACF chart

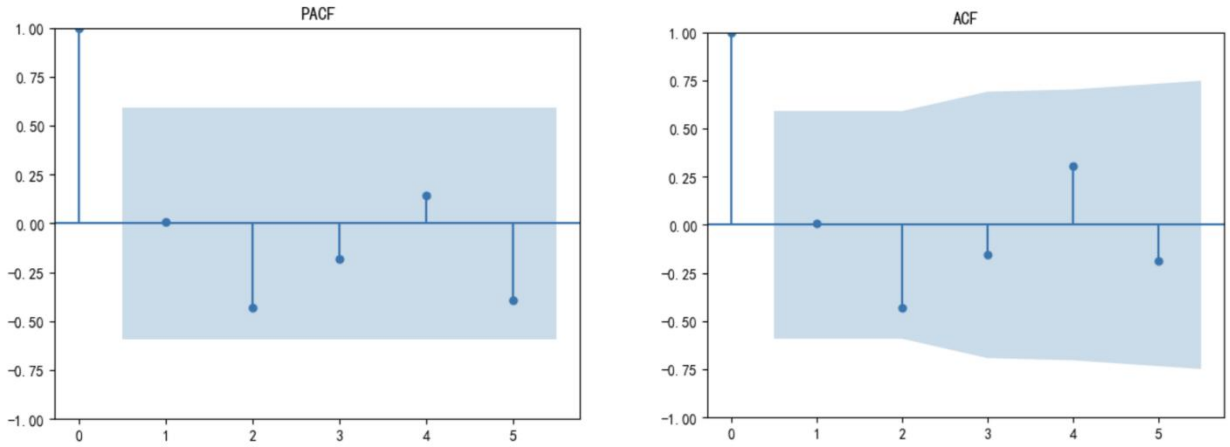
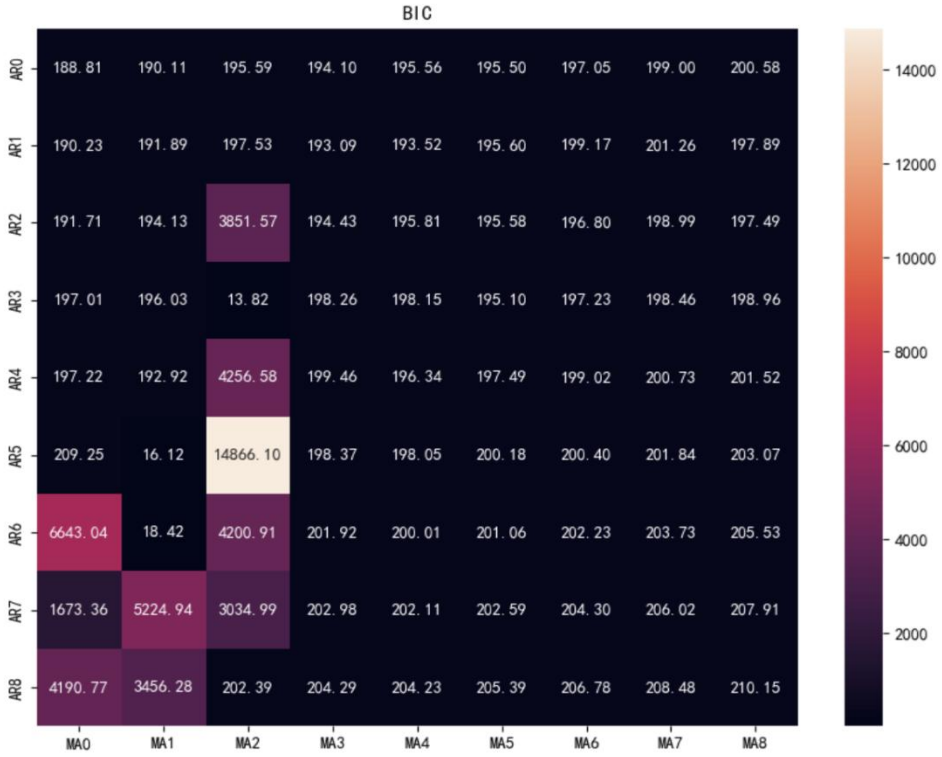


Figure 6: BIC test chart

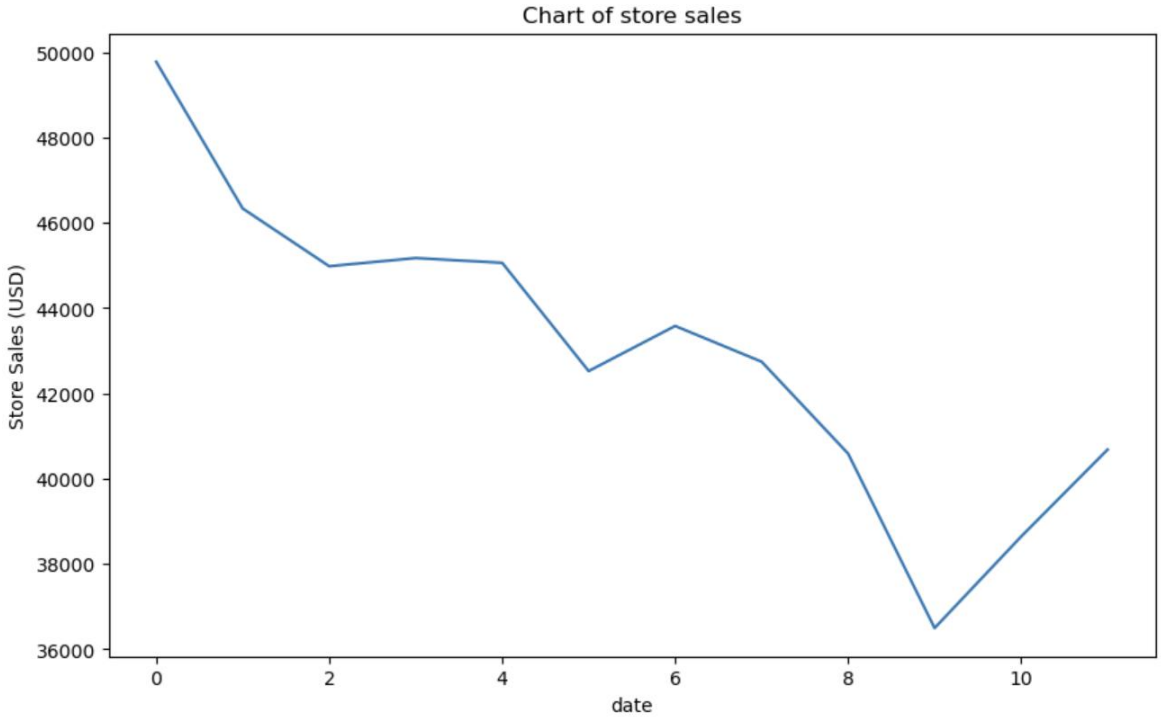


After analysis, it is known that p = 6 and q = 2. Therefore, an ARIMA(6, 1, 2) model is established for prediction(KENNEY, MARTIN1, ZYSMAN & JOHN2, 2016)

5.3 Establishment of ARIMA(6, 2, 2) Model

The ARIMA(6, 2, 2) model is solved through Python, and the store sales volume is predicted. The results are as follows:

Figure 7: ARIMA model prediction result chart



6. Price Band Analysis

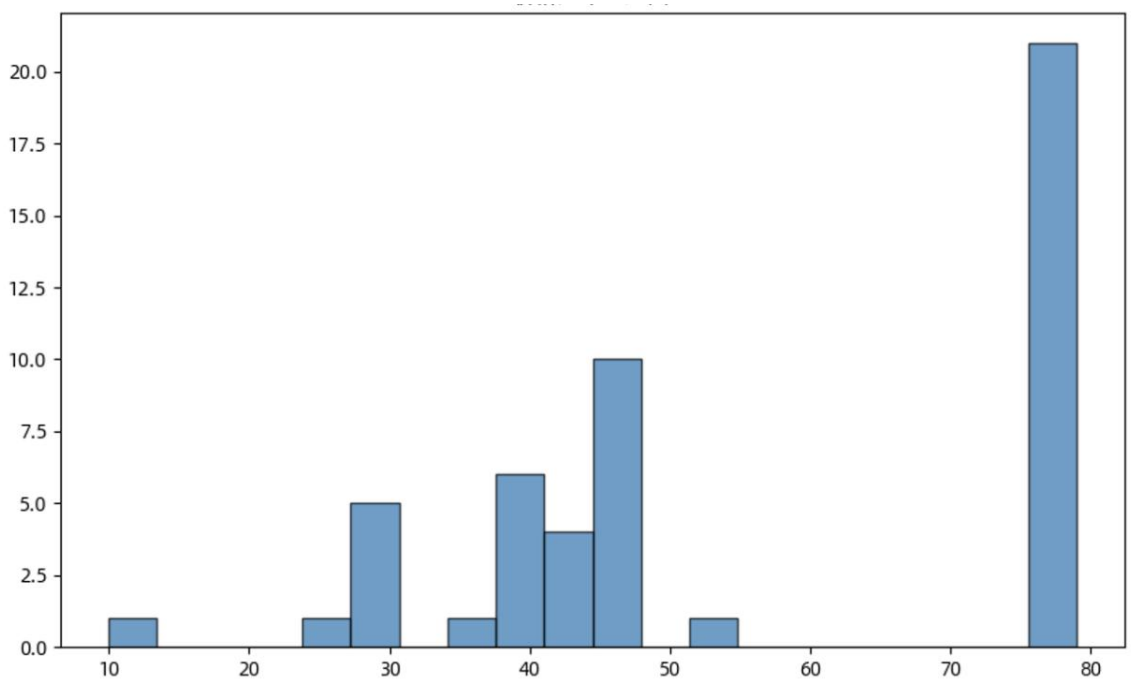
We conduct an analysis on the store's products, starting with the aspect of product prices. The following information has been obtained:

- The mean price of products is 55.12 yuan.
- The standard deviation of prices is 19.96 yuan.
- The price range is from 9.99 yuan to 78.99 yuan.
- The median price is 45.99 yuan.

We can comprehensively analyze the price distribution of store products. First of all, the mean price of products is 55.12 yuan, while the median is 45.99 yuan. The median is lower than the mean, which indicates that there may be skewness in the price distribution of products. That is, there are a few products with relatively high prices. These high-priced products have a greater impact on the mean and raise the overall average price. At the same time, the standard deviation of product prices is 19.96 yuan, which is a relatively large value. This further confirms the large volatility of prices, indicating that product prices are widely dispersed above and below the mean. In addition, the price range of products is from 9.99 yuan to 78.99 yuan. This large price range indicates that the store's product prices span a wide range and can cover multiple price points from low-end to high-end, thus meeting the purchasing power and needs of different consumers. This pricing strategy may help attract a wider customer base, but at the same time, the store needs to make corresponding strategic adjustments in inventory management and cost control.

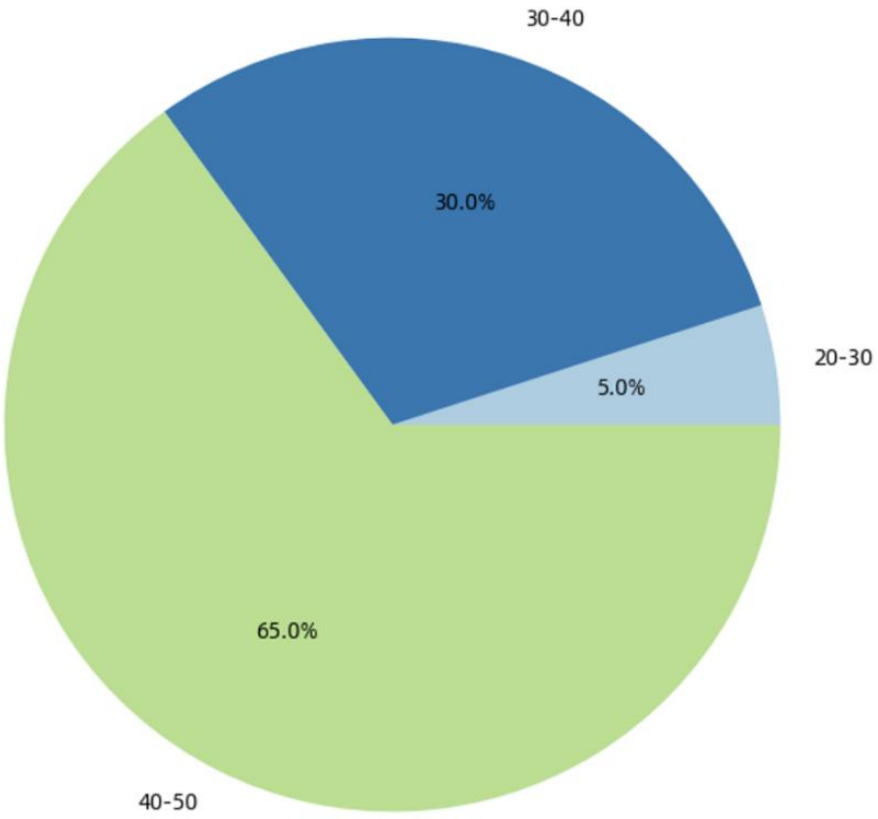
Next, we will draw a price distribution histogram to more intuitively view the price distribution.

Figure 8: Price distribution histogram



Then we conduct price band analysis. Price band analysis is a method of evaluating market demand, consumer preferences, and product positioning by studying the distribution range and intervals of product prices. Price band analysis helps enterprises more accurately grasp market dynamics, meet consumer needs, improve operating efficiency and market competitiveness.

Figure 9: Sales proportion chart of detailed price bands



We find that the price band of 40 - 50 yuan accounts for as high as 65% of the sales proportion. This

indicates that products in this price range are extremely popular and competitive in the market. It is very likely that products in this price range better meet the needs and expectations of consumers in terms of quality, function, and cost performance.

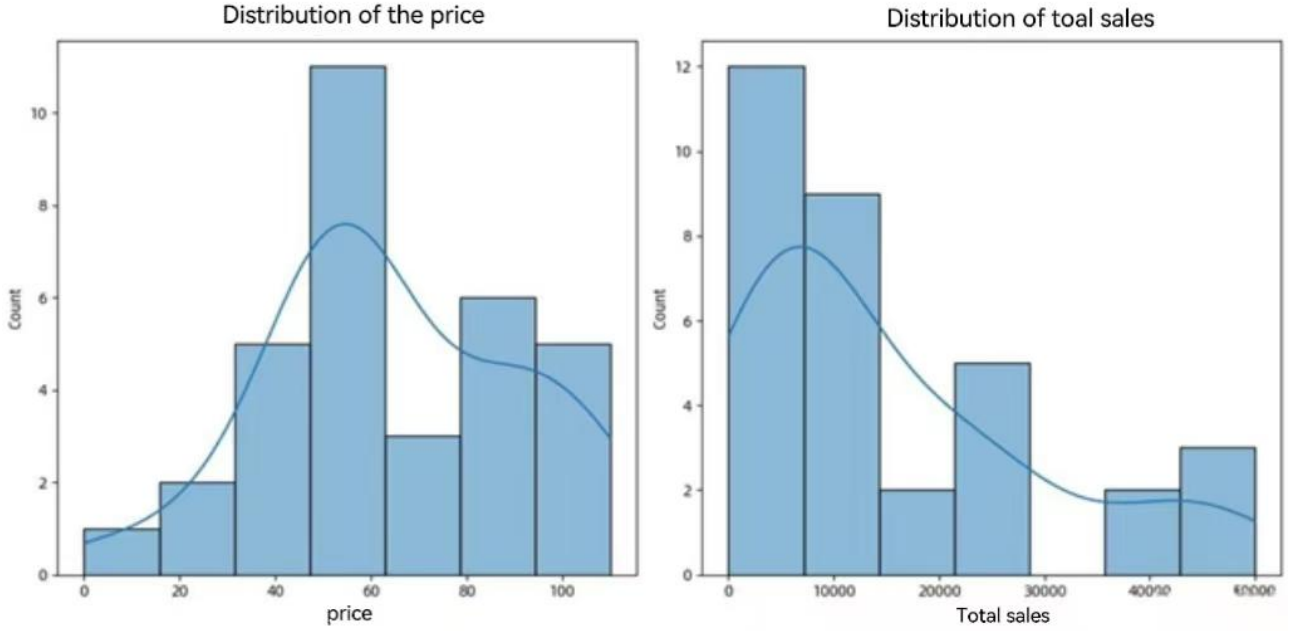
In contrast, the price band of 30 - 40 yuan only accounts for 30% of the sales proportion. This may mean that products in this price range fail to fully attract consumers in some aspects or are at a relative disadvantage in competition with products in the 40 - 50 yuan price band.

From the perspective of the overall sales structure, the store's sales performance largely depends on products in the 40 - 50 yuan price band. This provides an important basis for the store's business decisions. For example, the store can consider adding more products with different styles, functions, or qualities in the 40 - 50 yuan price band to further meet the diverse needs of consumers and consolidate the sales performance of this advantageous price band.

At the same time, for the price band of 20 - 30 yuan, it is necessary to deeply analyze the reasons for its relatively low sales. It may be necessary to improve the design and quality of products, or adjust marketing strategies, such as increasing promotional efforts and optimizing product display, to improve the attractiveness and market share of products in this price band.

We can now continue to analyze the relationship between "total sales" and "price". First, draw a scatter plot to intuitively show the relationship between the two. Subsequently, we can use linear regression analysis to determine the most appropriate price range.

Figure 10: Histogram of price distribution and total sales distribution



Through observing the statistical information and distribution charts of prices and total sales, the following conclusions can be drawn:

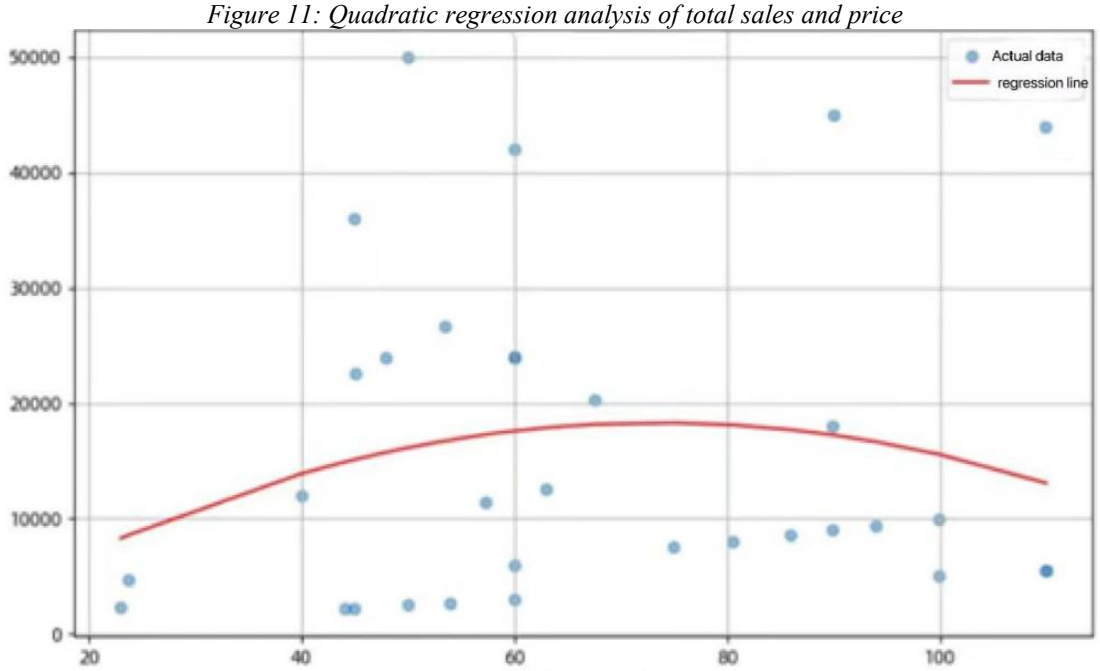
- The average price is about 65 yuan.
- The average total sales is about 15,354.
- The median price is 59.99 yuan, which may be an important reference point.

Next, we will conduct regression analysis to find the best price range. We will use a linear regression model to analyze the relationship between "total sales" and "price".

We further optimize the pricing strategy by excluding abnormal data and using an appropriate regression model to estimate a reasonable price range. Considering the distribution of prices and total sales, we can assume that data with too low prices are outliers and filter out these data, and then conduct regression analysis again.

We conduct quadratic regression analysis:

Figure 11: Quadratic regression analysis of total sales and price



According to the results of quadratic regression analysis, we obtain the optimal pricing of 73.49 yuan. This means that based on data analysis, this price may bring the largest total sales. Of course, in actual situations, other external factors such as market competition and brand positioning may also need to be considered.

7. Commodity Attribute Analysis

We have conducted Pearson correlation analysis on the commodity information on the search page, aiming to explore the impact of various variables on sales volume. The main intentions of correlation analysis cover the following aspects:

First, it can determine whether there is a correlation between different variables and the strength and direction of this correlation. For example, through correlation analysis, we can know whether there is a positive correlation (higher price, higher sales volume), a negative correlation (higher price, lower sales volume), or no obvious correlation between product price and sales volume.

Second, it helps predict and explain changes in variables.

Third, it can be applied to optimize decision-making.

Fourth, it can identify potential causal relationships. Although correlation does not necessarily mean causality, a strong correlation can provide clues and directions for further exploring causal relationships.

Subsequently, we have drawn a correlation heat map:

Figure 12: Correlation analysis chart

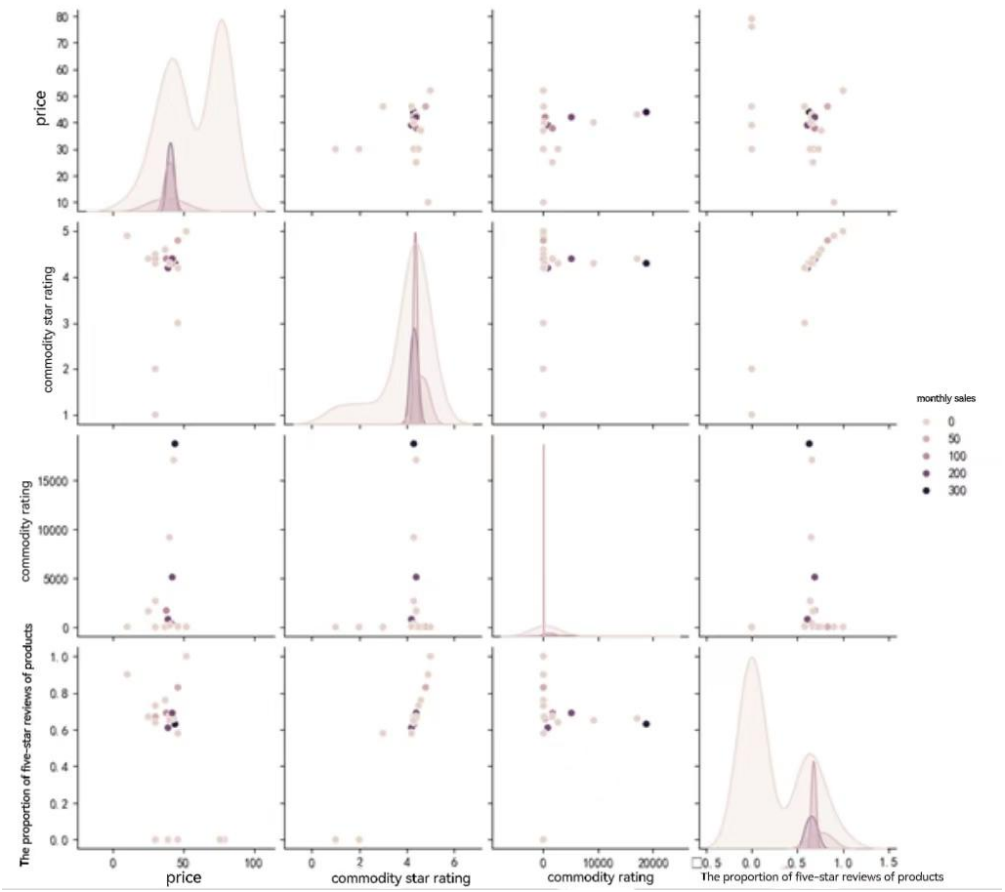
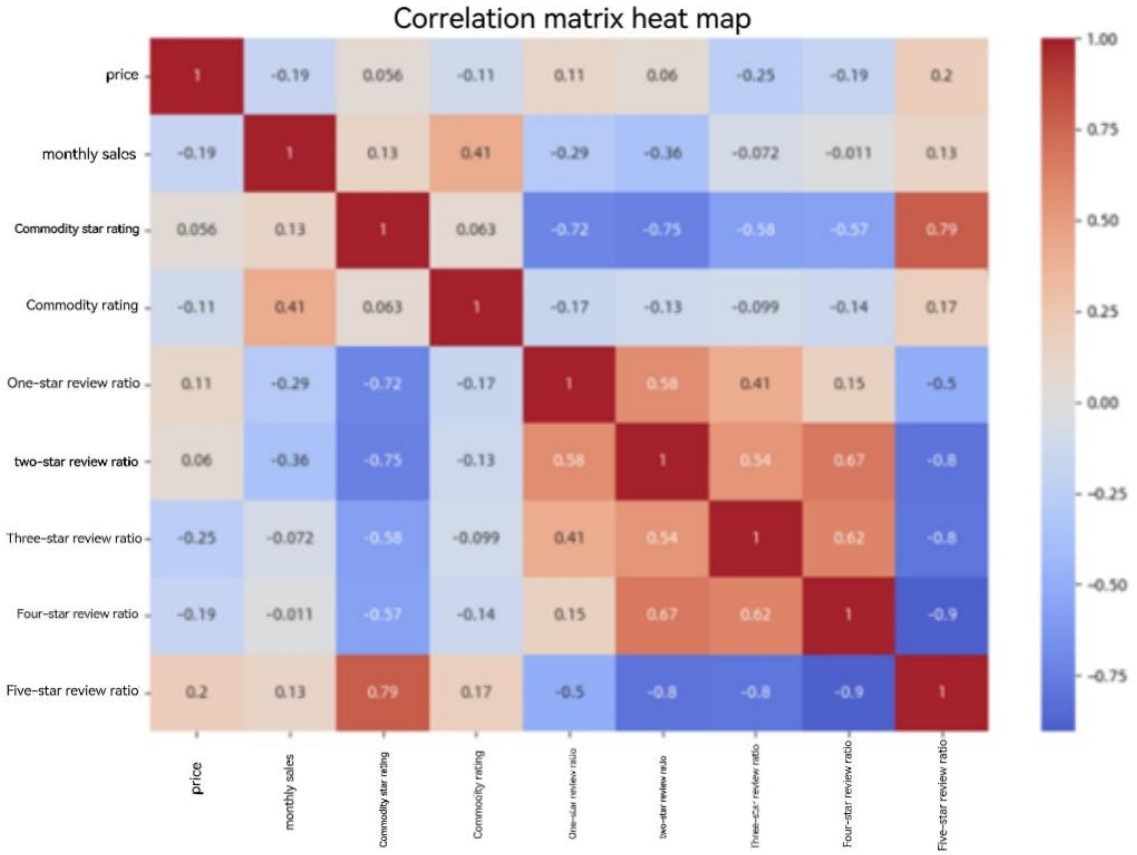


Figure 13: Correlation matrix heat map



The correlation coefficient between price and monthly sales volume is -0.185367, which shows that there is a weak negative correlation between the two. That is to say, generally, the higher the price, the lower the sales volume is often.

The correlation coefficient between the number of commodity reviews and monthly sales volume is 0.413369, which indicates that the more the number of commodity reviews, the higher the sales volume. There is a moderate positive correlation between the two.

The correlation coefficient between the proportion of five-star reviews of commodities and monthly sales volume is 0.128030. Although the correlation is not very strong, it can be considered that five-star reviews have a certain positive impact on sales volume.

For other indicators, such as commodity star ratings and the proportion of reviews at all levels and monthly sales volume, the correlations are relatively weak (the absolute values are all less than 0.5), which means that these factors have relatively little impact on sales volume.

Subsequently, combined with the three indicators of monthly commodity sales volume, the number of commodity reviews, and the proportion of five-star reviews of commodities, we have systematically sorted out men's shoes products in the entire store, focusing on finding those products that although have a relatively high number of commodity reviews and a large proportion of five-star reviews, that is, relatively good user satisfaction, but do not show corresponding market competitiveness in monthly sales volume.

We will view the descriptive statistical information of the data, and then identify those products with a relatively high number of commodity reviews and a large proportion of five-star reviews but low monthly sales volume by setting thresholds. We can define the following criteria:

The number of commodity reviews is higher than the average.

The proportion of five-star reviews of commodities is higher than the average.

Monthly sales volume is lower than the average.

Screen out a list of products that meet the conditions according to these criteria.

We have found products that meet the conditions. Although these products have a relatively high number of reviews and a large proportion of five-star reviews, they do not show corresponding market competitiveness in monthly sales volume.

Table 2:

Product ID	brand name	Monthly Sales volume	product review score	percentage of 5-star reviews for the product
B01CUME4UW	HEYDUDESHOES	50	19208	0.84
B09RRS54CW	Brooks	200	14472	0.81
B07CYXDDZ7	Salomon	50	12957	0.82

8. Commodity Reviews: Word Segmentation, Part-of-Speech Tagging, and Removal of Stop Words

8.1 Word Segmentation and Removal of Stop Words

8.1.1 Word Segmentation of Review Data

Word segmentation is a fundamental step in text information processing and involves splitting a sequence of words into individual words. Accurate word segmentation can significantly enhance the computer's ability to recognize and understand text information. Conversely, inaccurate word segmentation will generate a large amount of noise, seriously interfering with the computer's recognition and understanding ability and having a greater impact on the subsequent processing of this information.

The most commonly used package for word segmentation is the jieba word segmentation package. Jieba

word segmentation is an open-source library written in Python specifically for word segmentation. It has three basic principles, which are the implemented technologies.

① It achieves efficient word graph scanning based on the Trie tree structure and generates a directed acyclic graph (DAG) composed of all possible word formation situations of Chinese characters in a sentence. Jieba word segmentation comes with a dictionary called dict.txt, which contains more than 20,000 words, including the frequency of occurrence of entries (this frequency is trained by the author based on resources such as People's Daily corpus) and part of speech. The Trie tree is a well-known prefix tree. If the first few characters of a word are the same, indicating that the word has the same prefix, it can be stored using the Trie tree. The Trie tree storage method has the advantage of fast search speed. The latter sentence "generating a directed acyclic graph composed of all possible word formation situations of Chinese characters in a sentence" means that given a sentence to be segmented, a directed acyclic graph as shown is generated.

② It uses dynamic programming to find the maximum probability path and identify the maximum segmentation combination based on word frequency. First, it looks for segmented words in the sentence to be segmented, then finds the frequency of occurrence of the word. According to the method of dynamic programming to find the maximum probability path, it calculates the maximum probability in reverse from right to left for the sentence (the reverse is because the center of gravity of Chinese sentences often falls on the right. Calculating from right to left has a higher accuracy rate than calculating from left to right, which is similar to reverse maximum matching). Finally, the maximum probability segmentation combination is obtained. Directed acyclic graph generated by segmentation of "There are differences of opinion".

③For unknown words, the HMM model is used, and the Viterbi algorithm is employed to mark Chinese words according to four states of BEMS. Among them, B represents begin, indicating the start position; E represents end, indicating the end position; M represents middle, indicating the middle position; S represents single, indicating the position where a single word is formed. The HMM model uses these four states (B, E, M, S) to mark words. For example, Beijing can be marked as BE, that is, north/B Beijing/E, indicating that north is the start position and Beijing is the end position. The Chinese nation can be marked as BMME, which means start, middle, middle, and end.

8.1.2 Removing Stop Words

Stop words (Stop Words), as translated in dictionaries, are "function words and non-indexed words in computer retrieval". In SEO search engines, to save storage space and improve search efficiency, search engines will automatically ignore certain words or phrases when indexing pages or processing search requests. These words or phrases are called stop words.

To a certain extent, stop words are equivalent to filter words. The difference is that the scope of filter words is larger. Keywords containing pornographic, political, and other sensitive information will be treated as filter words for processing. Stop words themselves do not have this restriction. Generally speaking, stop words can be roughly divided into the following two categories:

One category consists of words that are very widely used or even overly frequent. For example, in English, "i", "is", "what", and in Chinese, "I", "just", etc. These words appear in almost every document. Querying such words cannot guarantee that the search engine can provide truly relevant search results. Therefore, it is impossible to improve the accuracy of search results by narrowing the search scope, and at the same time, it will reduce the search efficiency. Therefore, when searching, search engines such as Google and Baidu will ignore specific common words. If too many stop words are used, it is possible that accurate results cannot be obtained, and even a large number of completely unrelated search results may be obtained.

The other category is words that appear very frequently in texts but have little practical significance. This category mainly includes modal particles, adverbs, prepositions, conjunctions, etc. These words usually have no clear meaning on their own and only have a certain role when placed in a complete sentence. Common ones are "of", "in", "and", "then", etc. For example, in the sentence "I am your best friend", "is" and "of" are two stop words.

After word segmentation, reviews change from a string form to a form of multiple strings composed of words or phrases. At this point, it can be determined whether the words in the review are stop words. According to the definition of stop words above, a stop word library is compiled, and stop words in reviews

are removed according to this library. The key code is as follows:

Since the goal of this project is to analyze the advantages and disadvantages of product features, reviews like "Good, very good product" and "Very good, continue to support" although expressing an emotional tendency towards the product, in fact, it is impossible to extract which product features are satisfactory to users based on these reviews. Only when there are clear nouns, such as organization names and other proper nouns, in the review does it have meaning. Therefore, part-of-speech tagging needs to be performed on the words after word segmentation. Then, reviews containing nouns are extracted according to the part of speech. Regarding dictionary part-of-speech tagging, jieba adopts the tagging method of ICTCLAS.

According to the obtained part of speech, extract reviews with part of speech containing "n". The key code is as follows:

```
ind = result[['n' in x for x in result['nature']]]['index_content'].unique()
result = result[[x in ind for x in result['index_content']]]
```

8.2 Drawing Word Clouds

After data preprocessing, word clouds can be drawn to view the word segmentation effect. Word clouds will visually highlight "keywords" with high frequencies in the text. First, word frequency statistics need to be performed on words, and the word frequencies are sorted in descending order, exported in Excel format, and imported into Tableau.

Emotional tendency is also called emotional polarity. In a commodity review, it can be understood as whether the attitude of the user in expressing their own views on the commodity is supportive, opposing, or neutral, that is, the positive emotion, negative emotion, and neutral emotion usually referred to. Since this project is mainly to analyze the advantages and disadvantages of products, only the direction of emotional tendency in user review information needs to be analyzed, and the emotional degree of each review does not need to be analyzed.

Figure 14: User evaluation word cloud diagram

To analyze the emotional tendency of reviews, emotional words need to be matched first. The main method used is dictionary matching. The emotional word list used in this project is the "Word Set for Sentiment Analysis (beta version)" released by HowNet. Mainly use word lists such as "English positive evaluation", "English negative evaluation", "English positive emotion", and "English negative emotion". Merge the two word lists of "English positive evaluation" and "English positive emotion", and assign an initial weight of 1 to each word as the positive review emotional word list of this project. Merge the two word lists of "English negative evaluation" and "English negative emotion", and assign an initial weight of -1 to each word as the negative review emotional word list of this project.

Read in the positive and negative review emotional word lists, assign an initial weight of 1 to positive words, and an initial weight of -1 to negative words. Use the merge function to match the word emotional word list with the word segmentation result. The code is as follows:

```
positive = set(pos_comment.iloc[:, 0]) | set(pos_emotion.iloc[:, 0])
negative = set(neg_comment.iloc[:, 0]) | set(neg_emotion.iloc[:, 0])
intersection = positive & negative  # 正负面情感词表中相同的词语
positive = list(positive - intersection)
negative = list(negative - intersection)
positive = pd.DataFrame({"word": positive,
                         "weight": [1] * len(positive)})
negative = pd.DataFrame({"word": negative,
                         "weight": [-1] * len(negative)})
```

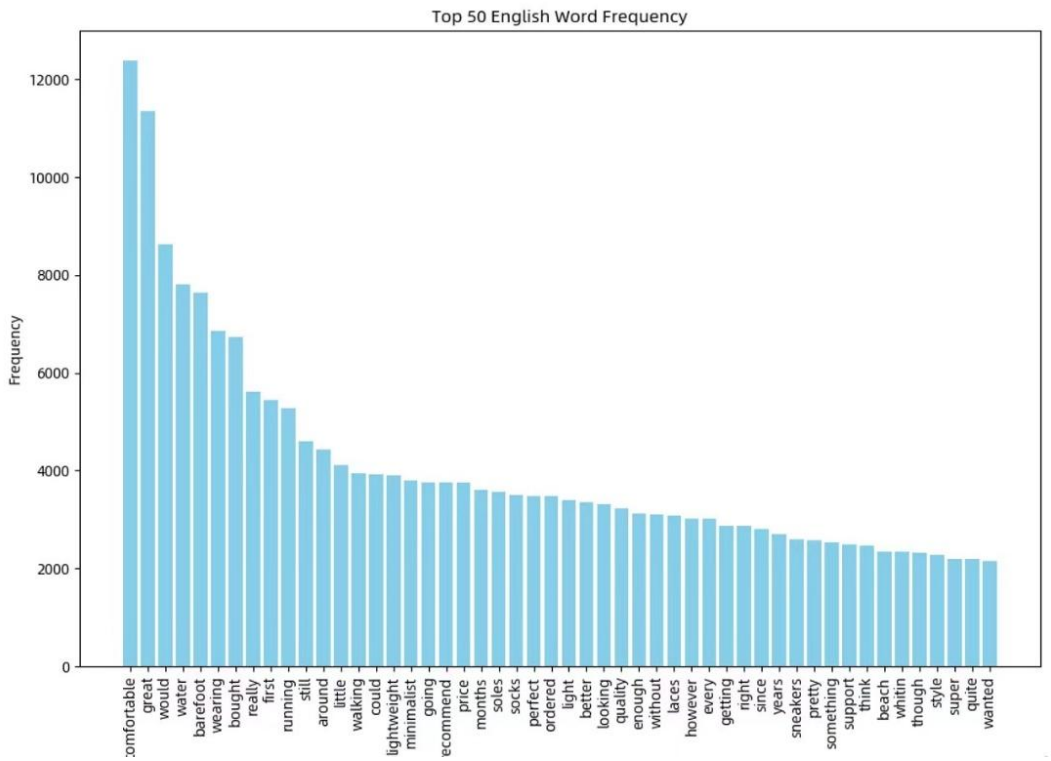
8.2.1 Correcting Emotional Tendency

Emotional tendency correction mainly judges whether the emotional value is correct based on whether there are negative words in the two positions in front of the emotional word. Since there is a phenomenon of multiple negations in Chinese, that is, when a negative word appears an odd number of times, it means negation; when a negative word appears an even number of times, it means affirmation. According to Chinese language habits, search for the first two words before each emotional word. If there are an odd number of negative words, adjust to the opposite emotional polarity.

8.2.2 Viewing the Effect of Emotional Analysis

Use the WordCloud function under the wordcloud package to count the word frequencies of positive and negative reviews respectively to view the effect of emotional analysis. Finally, we obtain the top 50 positive emotional words. According to the research results, we can further optimize our products and increase product sales.

Figure 15: Top 50 positive emotional words



9. Marketing Plan

9.1 Store SWOT Analysis

Strengths:

Market position: The store performs well on the Amazon platform. In 2022, its sales reached 201 billion U.S. dollars, showing strong market competitiveness.

Product diversity: The store offers a wide price range from 9.99 yuan to 78.99 yuan, which can meet the needs of different consumers.

Consumer base: It has a certain proportion of repeat customers, indicating high customer loyalty.

Technology application: Utilize technical means such as data analysis and ARIMA models for market trend prediction and inventory management.

Weaknesses:

Price volatility: The large standard deviation of commodity prices indicates that there may be instability in price management.

Inventory management: There may be insufficient supplier management, resulting in increased difficulty in warehouse inventory counting and delivery delays.

Market dependence: The store may be highly dependent on the Amazon platform. Consider multi-channel sales strategies to diversify risks.

Product homogeneity: In some price ranges, products may face fierce homogeneous competition.

Opportunities:

Market expansion: With the growth of global cross-border e-commerce, the store has the opportunity to further expand the international market.

Belt and Road Initiative policy: Policy support provides Chinese enterprises with a broader market and more convenient trade channels.

Technological innovation: The application of automated production lines and 3D printing technology can improve production efficiency and product quality.

Changes in consumer behavior: After the epidemic, consumers are more inclined to shop online, providing new growth points for e-commerce.

Threats:

Market competition: There are many competitors in the market, and price wars may lead to a decline in profits.

Macroeconomic fluctuations: The uncertainty of the global economic environment may affect consumer purchasing power.

Policy changes: Changes in trade policies and cross-border e-commerce policies may affect store operations.

Technological changes: The emergence of new e-commerce platforms and sales models may pose a threat to existing business models.

9.2 Store SO, ST, WO, WT Strategies

SO (Leverage strengths to seize opportunities) strategy:

Expand the global market: Utilize the store's strong sales performance and brand influence on the Amazon platform to further explore the international market, especially in regions supported by the "Belt and Road Initiative" policy.

Product diversification: Leverage the advantage of a wide price range of products to develop more new

products that meet market demands to cover a wider consumer group.

ST (Leverage strengths to reduce threats) strategy:

Strengthen brand positioning: Through strengthening brand building and marketing, enhance consumers' awareness of the store's brand to resist the threat of market competition.

Optimize supply chain management: Utilize the store's technological advantages to improve supply chain management and reduce the impact of macroeconomic fluctuations and policy changes on inventory and supply.

WO (Improve weaknesses and leverage opportunities) strategy:

Improve price management: In response to the problem of large price volatility of products, formulate a more scientific pricing strategy through market research and data analysis to take advantage of the growth opportunity of cross-border e-commerce.

Enhance customer loyalty: In response to the problem of room for improvement in repeat purchase rates, increase customer stickiness by providing personalized services and loyalty reward programs.

WT (Reduce weaknesses and threats) strategy:

Diversify sales channels: Given the risks that may arise from dependence on the Amazon platform, develop other e-commerce platforms or establish own sales channels to disperse the threat of market concentration.

Improve product quality and innovation: In the face of homogeneous product competition, provide products with unique selling points through research and development and innovation to reduce market competition pressure.

By implementing these strategies, your store can better leverage its own advantages, seize market opportunities, and at the same time address and mitigate weaknesses and external threats. This will help the store achieve long-term stable development.

9.3 On-site and Off-site Promotion Plans

On-site promotion plan

Optimize product listings:

Ensure that all product titles, descriptions, and images are attractive and comply with SEO best practices.

Use relevant keywords to improve search rankings.

Utilize Amazon advertising:

Run Sponsored Products and Sponsored Brands ads to increase product visibility.

Participate in Amazon promotional activities:

Use promotional activities such as Lightning Deals and Today's Deals to attract buyers.

Optimize price strategy:

Adjust prices according to price band analysis to attract more consumers.

User review management:

Encourage satisfied customers to leave positive reviews to improve store reputation.

Off-site promotion plan

Social media marketing:

Create brand accounts on platforms such as Facebook, Instagram, and Twitter, and post attractive content and promotional information.

Influencer cooperation:

Cooperate with opinion leaders or influencers in the industry to attract potential customers through their recommendations.

Content marketing:

Create blog posts, videos, and other valuable content to provide information and usage tips about products.

Search engine marketing:

Use tools such as Google AdWords for paid search advertising to attract targeted traffic.

Email marketing:

Develop an email subscription list and send regular newsletters and exclusive offers.

Affiliate marketing:

Join Amazon or other platforms' affiliate marketing programs and let other websites help promote products.

Local community engagement:

Participate in local community activities and markets to increase brand awareness in the local area.

Analysis and optimization:

Use the data feedback from off-site promotions to continuously optimize promotion strategies and content.

10. Conclusion

10.1 Summary

In the current wave of globalization, cross-border e-commerce has become an important part of global trade. Through in-depth analysis of men's footwear sales data on the Amazon platform, we have reached the following key conclusions:

First, it is necessary to deeply analyze the target market, especially countries along the "Belt and Road Initiative", understand their needs, preferences, cultural and consumption habits, and then carry out market segmentation and adjust the product line accordingly to meet the diverse needs of different consumer groups.

In the field of product strategy, achieve product differentiation based on the results of data analysis, launch products with unique selling points and attributes that attract buyers, so as to enhance market competitiveness. At the same time, combine market research and cost analysis to formulate a scientific and reasonable pricing strategy.

The importance of user behavior analysis is self-evident. Further refine user portraits based on the results of RFM model and cluster analysis, accurately identify high-value user groups, and then provide personalized product recommendations and promotional activities according to users' behaviors and preferences.

For sales analysis and prediction, regularly analyze store sales data, monitor sales trends, and adjust marketing strategies in a timely manner. At the same time, use prediction tools such as the ARIMA model to predict future sales trends and optimize inventory management.

In terms of commodity attribute analysis, identify the common attributes of hot-selling products such as style, color, and material through data analysis, and then optimize the product portfolio to increase overall sales.

In terms of enhancing user loyalty, based on the analysis results of repeat purchase rates, explore the factors that affect repeat purchases, such as star ratings, reviews, prices, etc., and design loyalty reward programs such as point systems and member privileges to increase user stickiness.

In the field of supply chain management, optimize inventory management according to sales forecasts and

market demands, reduce the risks of overstocking and out-of-stock, and at the same time strengthen communication and cooperation with suppliers to improve the response speed and flexibility of the supply chain.

In the aspect of brand building and marketing, create a good brand image with high-quality products and excellent services, improve brand awareness, and use multiple channels such as social media, advertising, and cooperation with KOLs to enhance brand exposure and influence.

In terms of data analysis and decision support, ensure that all decisions are based on data analysis to improve the scientific nature and effectiveness of decisions, and continuously optimize operating strategies and improve operating efficiency according to the results of data analysis.

Finally, risk management must be carried out. Regularly assess market risks such as exchange rate fluctuations and policy changes, and formulate response strategies and emergency plans to ensure that rapid responses can be made in the event of emergencies and reduce losses.

10.2 Outlook

Looking to the future, we will continue to promote development and optimization in many dimensions.

First, in terms of market expansion, with the booming development of global cross-border e-commerce, stores have a valuable opportunity to further expand the international market, especially in regions strongly supported by the "Belt and Road Initiative" policy. These regions often contain huge consumption potential and undeveloped market space, providing an ideal stage for the store's international layout.

Second, technological innovation has become a key force driving industry progress. The application of data technology can not only significantly improve production efficiency but also greatly improve product quality. This undoubtedly injects strong innovation power into the footwear industry and strongly leads the trend of technological change.

Furthermore, after the epidemic, consumers' preference for online shopping has become more obvious. This trend has created new growth opportunities for the e-commerce industry. Stores must continuously pay close attention to and actively adapt to this significant change in consumer behavior in order to better meet consumer needs.

The importance of brand building cannot be underestimated. By continuously improving brand awareness and reputation and enhancing customer loyalty, the store's competitiveness in the market can be effectively enhanced.

In the field of supply chain management, optimizing inventory and supplier management is crucial. Efforts should be made to reduce the difficulty of warehouse inventory counting and reduce the occurrence of delivery delays to ensure timely shipments. This can not only improve operational efficiency but also enhance customer satisfaction.

Customer experience is an important link that cannot be ignored. By continuously improving the quality and level of customer service and accurately providing personalized recommendations, customer satisfaction and repeat purchase rates can be significantly improved, thereby strengthening the long-term cooperation relationship between the store and customers.

The role of multi-channel marketing is also very important. In addition to effective promotion activities within the Amazon platform, diversified off-site promotion channels should also be actively considered, such as leveraging the powerful communication power of social media and the precise positioning function of search engine marketing to further expand the market coverage and enhance brand influence.

Product innovation is one of the core elements to maintain competitiveness. For products with high ratings but relatively low sales, the reasons should be deeply analyzed and effective improvement measures should be taken. For example, optimize product characteristics to make them more in line with market needs; adjust marketing strategies to enhance the market promotion effect of products.

In addition, it is necessary to always pay attention to the technological innovation trends and changes in consumer preferences in the footwear industry so that the product line can be adjusted in time to perfectly

match market demands and always maintain product attractiveness and competitiveness.

Finally, user portrait analysis is an important means to achieve refined operations. Fully utilizing advanced tools such as the RFM model to deeply mine and analyze user behavior characteristics can provide accurate data support for store operation decisions and achieve more efficient and accurate operation strategies.

At the same time, risk management is also an indispensable important link. Formulate comprehensive response strategies to deal with potential threats such as macroeconomic fluctuations and policy changes, so as to ensure that the business can maintain stability and achieve sustainable development in a complex and changing market environment.

By fully implementing the above strategies, stores can give full play to their own advantages, keenly capture market opportunities.

References

Alsuwaylimi & Amjad A.aCAa. (2023).Comparison of ARIMA, ANN and Hybrid ARIMA-ANN Models for Time Series Forecasting. Information Sciences Letters, Vol. 12(2), pp. 1003-1016.

Huang Yi. (2016). Digital economy dividends: Rewards for pioneers. Big Data Era (3), 6-9.

KENNEY, MARTIN1, ZYSMAN & JOHN2. (2016). The Rise of the Platform Economy. ISSUES IN SCIENCE AND TECHNOLOGY, Vol. 32(3), pp. 61-69.

Kumar, Amita;Sharma, Nehaa;Gurna, Kamalpreet Kaurb;Anand, Abhineetc;Patni, Jagdish ChandradCAa; Pinjarkar & Latikae. (2024). Forecasting Stellar XLM Prices: Insights from ARIMA Analysis. International Journal of Religion, Vol. 5(6), pp. 273-288.

Liqin Wen;Shenglin Ma1CA & Shuping Lyu1. (2024). The influence of internet celebrity anchors' reputation on consumers' purchase intention in the context of digital economy: from the perspective of consumers' initial trust. Applied Economics, pp. 1-22.

Ma Shuzhong, Fang Chao & Liang Yinfeng. (2018). Digital Trade and Its Values and Research Prospects. International Trade Issues (10),16-30.

Meertens, Q A 1, Diks, C G H 2, H J van den Herik 3 & Takes, F W 4. (2020). A data-driven supply-side approach for estimating cross-border Internet purchases within the European Union. Journal of the Royal Statistical Society. Series A, Statistics in Society, Vol. 183(1), pp. 61-90.

Mingli Song 1, 2CA1; Ruobing Wang1 & Yan Li1. (2024). Hybrid time series interval prediction by granular neural network and ARIMA. Granular Computing, Vol. 9(1).

Oh, Jiua & Seong, ByeongchanaCAa. (2024). Forecasting with a combined model of ETS and ARIMA. Communications for Statistical Applications and Methods, Vol. 31(1), pp. 143-154.

Wang Yingwei & Ma Shucai. (2021). Time series prediction based on ARIMA and LSTM hybrid model. Computer Applications and Software (02), 291-298.

Xu Limei. (2020). Research on the competitiveness evaluation of global digital economy enterprises. Shanghai Economy (3), 77-90.

Zhai Huafeng. (2020). Research on the challenges and countermeasures faced by China's cross-border e-commerce under the background of digital economy. Management and Technology of SMEs (34), 30-31.

Funding

This research received no external funding

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

Not Applicable.