

# Research on the Influencing Factors of Online Clothing Sales Based on Binary Logit Regression

Yichen Gu<sup>1</sup>, Zejian Wang<sup>1</sup> and Yunfeng Xia<sup>2</sup>

<sup>1</sup>*School of Economics and Management, Xidian University, Xian, 710000, China*

<sup>2</sup>*Jurong Country Garden School, Jurong, 212446, China*

**Keywords:** Data Analysis, Customer Satisfaction, Online Clothing Sales, Payment, Influence Factor.

**Abstract:** The main purpose of this study is to use the method of binary logit regression to conduct a comprehensive discussion on "factors affecting online clothing sales". In the era of more and more developed online sales, online clothing sales occupy a large part of online commodity sales, but at the same time, there are still many problems in online clothing sales that need to be improved. The overall goal is to find the basic determinants of customer satisfaction in online clothing sales, so as to give the factors for enterprises to increase online clothing sales. In this study, some online questionnaires were first obtained to obtain the data of people choosing online or offline shopping. Then, empirical analysis was conducted using the data set obtained from Kaggle to identify the correlation of online fields that significantly affects product sales, and binary logit regression analysis was carried out after processing these data. Through calculation, it is determined that the payment method and the discount application has a significant positive impact on customer satisfaction, which shows that it is essential for relevant managers to optimize these two aspects to promote the virtuous cycle of online clothing sales.

## 1 INTRODUCTION

In 2022, the new driving force index of China's economic development (100 in 2014) was 766.8, an increase of 28.4 percent over the previous year, and all sub-indicators have improved over the previous year. Among them, the Internet economy index has grown the fastest (Zheng, 2018). The internet economy refers to the broad spectrum of economic activities that rely on internet technology as a platform, with the network serving as the medium and application technology innovation at its core. The present state of the Internet economy primarily encompasses five main types: e-commerce, Internet finance (ITFIN), instant messaging, search engines, and online games. As a new economic form, the Internet economy is highly innovative, which not only shows the obvious technological sacrifice of The Times, but also integrates the essence of the traditional economy. The tangible economy encompasses economic endeavors involving the manufacturing and distribution of physical goods, intangible products, and services. It comprises both the material production and service sectors across various industries, as well as the creation and

provision of intangible products, serving as a crucial foundation for human sustenance and progress (Zheng, 2018).

In recent years, China's online economy has grown faster and faster, while the offline economy has gradually lost its temperature. According to the survey, by 2022, the national online shopping replacement rate (the replacement rate of online consumption to offline consumption) will reach 80.7% (Li & Shi, 2023). It can be seen that online shopping has become the main way of residents' consumption. With the gradual change of residents' consumption patterns, an increasing number of brick-and-mortar retail enterprises have started to undergo transformation and enhancement. In this context, the main sales channels at present are online sales on the Internet platform and offline sales in traditional stores (Yu, 2018). In addition, the domestic online game industry was only in a stable growth trend in the first half of 2017. As of June 2017, domestic online game users have exceeded 420 million, 4.6 million more than in 2016, accounting for 56.2% of the total amount of domestic netizens (Gong, 2019).

In contrast, although China's real economy has entered the 21st century, with the enhancement of

people's material living standards, China's domestic physical retail industry has embarked on a phase of swift advancement, experiencing an average growth rate of around 10%, which showcases the increasing purchasing power and consumer demand in the market. Nevertheless, given the current circumstances, the domestic physical retail industry still struggles when compared to international physical retail conglomerates in relation to overall sales and operational scale. (Lei, 2016). In this case, the real economy is not strong enough, and the real economy has ushered in the impact of the Internet economy. The rapid rise of the Internet economy has made the traditional economy face unprecedented challenges in terms of business philosophy, the business model and the service model, which is embodied in the loss of customer resources, the decrease of deposit ratio and the decrease of intermediate business volume (Yu, 2015).

But in fact, the fundamental goals of the development of the Internet economy and the real economy are the same. Whether it is the network economy or real economy, the ultimate goal of its development is to develop productivity (Cao, 2017). Therefore, it is necessary to maintain the balance between the Internet economy and the real economy. From 2011 to 2018, China experienced a significant surge in online retail sales, escalating from 0.78 trillion yuan to 9.01 trillion yuan, reflecting an average annual growth rate of 43.6 percent. By 2018, online retail sales in China constituted 10 percent of the GDP. Meanwhile, offline retail sales saw a notably lower average annual growth rate of only 4.2%, significantly beneath the average annual growth rate of the GDP (Yu & Zou, 2020). However, online shopping on the Internet and offline shopping in the physical industry each has advantages and disadvantages. Online shopping realizes cross-domain transactions, so that consumers can buy the goods they want without leaving their homes. On shopping websites, consumers can easily understand product information and make purchasing decisions. Conversely, traditional offline shopping offers consumers the chance to physically interact with goods. Shoppers can meticulously examine products and even engage in hands-on experiences with the products themselves (Bai, 2019). As online-to-offline integration becomes a new choice in the field of e-commerce, the industry generally believes that the "entry into the market" of online retailers will harm the interests of offline retailers (Ding, 2019).

This study aims at analyzing the factors influencing online sales in the fashion industry,

specifically identifying which types of clothing have the greatest potential for online sales.

## 2 METHODS

### 2.1 Data Source

The study used data from China's National Bureau of Statistics and the website Kaggle. The first dataset collected people's choices for online and offline shopping in 2022. The other dataset includes 14 factors that may affect people's online shopping experience and related data, with a total sample size of 3,900, which can fully illustrate the research question.

### 2.2 Indicators and Analysis

The analysis has carefully chosen specific indicators to deepen the understanding of the relationship between online and offline shopping. These indicators include product categories, prices, purchase quantities, and the male-female ratio of consumer groups. The analysis ensures that these indicators will serve as effective tools for analyzing and elucidating the complex dynamics of online and offline shopping. In addition, targeted surveys have been conducted to gain a deeper understanding of consumers' attitudes towards online and offline consumption. The survey aims at providing a detailed understanding of consumer perspectives for research purposes, enabling a more comprehensive analysis of the factors influencing consumers' choices to purchase clothing online in both e-commerce and traditional retail contexts (Table 1).

By leveraging, we sought to delve into the complexities surrounding online and offline shopping dynamics. While acknowledging the strengths of these datasets, it is important to remain cognizant of their limitations, particularly regarding temporal aspects and the challenges associated with assessing popularity. These considerations are essential for maintaining the integrity and validity of the analyses. The meticulous selection of indicators, combined with the targeted consumer survey, forms the cornerstone of the approach to unraveling the multifaceted relationship between online and offline shopping. With this comprehensive approach, the goal is to provide valuable contributions to the current knowledge base in this field, illuminating the complex interplay of elements influencing consumer behaviors within the spheres of e-commerce and traditional retail.

Table 1: Delving into Metrics.

Indicator	Mean ± standard deviation	Variance	Median	Standard error
Age	44.068±15.208	231.271	44.000	0.244
Gender	1.680±0.467	0.218	2.000	0.007
Item Purchased	13.035±7.199	51.828	13.000	0.115
Category	2.002±0.897	0.804	2.000	0.014
Size	2.120±0.930	0.866	2.000	0.015
Color	13.109±7.222	52.151	13.000	0.116
Season	2.493±1.117	1.248	2.000	0.018
Review Rating	3.750±0.716	0.513	3.700	0.011
Subscription Status	1.270±0.444	0.197	1.000	0.007
Shipping Type	3.514±1.698	2.882	4.000	0.027
Discount Applied	1.430±0.495	0.245	1.000	0.008
Promo Code Used	1.430±0.495	0.245	1.000	0.008
Previous Purchases	25.352±14.447	208.719	25.000	0.231
Payment Method	3.512±1.691	2.858	3.000	0.027

### 2.3 Method Introduction

The study first conducted data screening, selecting variables that may be related and analyzing the data using binary logit regression. The Review Rating was divided into two parts: 0 (dissatisfied) for ratings between 2.5-3.5, and 1 (satisfied) for ratings between 3.5-5, which were used as the dependent variable *y*, referred to as Customer Satisfaction. The study selected Age, Gender, Category, Location, Purchase Amount (USD), Size, Color, Previous Purchases, Payment Method, Frequency of Purchases, Subscription Status and Discount Applied as independent variables. Descriptive analysis and frequency analysis were performed on these variables to highlight their characteristics and facilitate the final binary logit regression analysis of Customer Satisfaction.

Table 2: Overview of Binary Logit Regression Analysis.

Name	Options	Frequency	Percentage
Customer Satisfaction	0	2106	54.00%
	1	1794	46.00%
	Total	3900	100.0%
Summary	valid	3900	100.00%
	Hiatus	0	0.00%
	Total	3900	100.0%

The model's fitting quality is assessed based on the accuracy of its predictions. Table 3 indicates that the overall predictive accuracy of the research model stands at 82.74%, demonstrating an acceptable level of model fitting. The prediction accuracy is 84.14% and when the true value is 1 is 81.10%.

## 3 RESULTS AND DISCUSSION

### 3.1 Basic Information

Using Age, Gender, Category, Location, Purchase Amount (USD), Size, Color, Previous Purchases, Payment Method, Frequency of Purchases, Subscription Status and Discount Applied as independent variables and Customer Satisfaction as the dependent variable for binary logistic regression analysis. The table above indicates the participation of 3900 samples in the analysis, revealing the absence of any missing data. (Table 2).

### 3.2 Model Results

The following table lists the data obtained by binary logit regression analysis and the relevant results obtained by these data (Table 4).

As can be seen from the above table, Age, Gender, Category, Location, Purchase Amount (USD), Size, Color, Previous Purchases, Payment Method, Frequency of Purchases, Subscription Status and Discount Applied are independent variables, Customer Satisfaction is considered the dependent variable for binary Logit regression analysis.

Table 3: Summary of prediction accuracy with binary Logit regression.

		Predicted Value		Forecast Accuracy	Predicting Error Rate
		0	1		
true value	0	1772	334	84.14%	15.86%
	1	339	1455	81.10%	18.90%
gather				82.74%	17.26%

Table 4: Summary of the results of the binary Logit regression analysis.

Sum	regression coefficient	standard error	z	Wald $\chi^2$	p	OR price 95% CI
Age	0.002	0.003	0.530	0.281	0.596	0.995 ~ 1.008
Gender	0.042	0.129	0.323	0.104	0.747	0.810 ~ 1.342
Category	0.024	0.054	0.456	0.208	0.649	0.922 ~ 1.139
Location	0.006	0.003	1.642	2.697	0.101	0.999 ~ 1.012
Purchase Amount (USD)	-0.002	0.002	-0.879	0.773	0.379	0.994 ~ 1.002
Size	-0.060	0.053	-1.138	1.294	0.255	0.849 ~ 1.045
Color	0.009	0.007	1.267	1.605	0.205	0.995 ~ 1.022
Previous Purchases	0.004	0.003	1.313	1.725	0.189	0.998 ~ 1.011
Payment Method	0.061	0.029	2.121	4.497	0.034	1.005 ~ 1.123
Frequency of Purchases	0.035	0.024	1.439	2.071	0.150	0.987 ~ 1.086
Subscription Status	-0.218	0.154	-1.421	2.021	0.155	0.595 ~ 1.086
Discount Applied	0.372	0.155	2.399	5.757	0.016	1.071 ~ 1.966
intercept	-14.877	0.575	-25.862	668.849	0.000	0.000 ~ 0.000

As can be seen from the above table 4, that means Age, Gender, Category, Location, Purchase Amount (USD), Size, Color, Previous Purchases, Payment Method, Frequency of Purchases, Subscription Status, Discount Applied can illustrate the 0.49 variation in Customer Satisfaction. It can be seen from the above table that the formula of the model is:

$$\ln\left(\frac{p}{1-p}\right) = -14.877 + 0.002 \times \text{Age} + 0.042 \times \text{Gender} + \dots + 0.372 \times \text{Discount Applied} \tag{1}$$

The regression coefficient value of the Payment Method is 0.061, and presents the significance level of 0.05 ( $z = 2.121, p = 0.034 < 0.05$ ), which means that the Payment Method will have a significant benefit on Customer Satisfaction. And the OR value is 1.062, which means that when the Payment Method is increased by one unit, the change (increase) of Customer Satisfaction is 1.062 times.

The regression coefficient value of Discount Applied is 0.372, and presents the significance of 0.05 level ( $z = 2.399, p = 0.016 < 0.05$ ). It means that Discount Applied will have a significant positive

impact on Customer Satisfaction. And the OR value is 1.451, meaning that when the Discount Applied is increased by one unit, the increase in Customer Satisfaction is 1.451 times.

### 3.3 Discussion

The summary analysis shows that Payment Method and Discount Applied will significantly benefit Customer Satisfaction. However, Age, Gender, Category, Location, Purchase Amount (USD), Size, Color, Previous Purchases, Frequency of Purchases, Subscription Status does not affect Customer Satisfaction. In a business environment, payment methods and discounts are crucial factors that impact customer satisfaction. Many retailers enhance customer satisfaction by offering a wide range of payment options and appealing discount policies. It is also believed that customer satisfaction can be influenced by various other factors, including age, gender, category, location, purchase amount, size, color, previous purchase history, purchase frequency,

and subscription status. Firstly, providing diverse payment methods such as credit cards, debit cards, Alipay or WeChat Pay can cater to customers' different payment habits and needs while enhancing shopping convenience and flexibility. Allowing customers to select their preferred payment method during the purchasing process will contribute to their comfort and overall satisfaction. Additionally, discount policies serve as effective promotional tools that entice customers to make purchases thereby increasing their desire to buy and overall satisfaction.

#### 4 CONCLUSION

This study uses a binary logit regression model that takes satisfaction as the dependent variable, using age, gender, category, location, purchase amount (USD), size, color, previous purchase, payment method, purchase frequency, subscription status, and discount application as the independent variables. At the same time, this paper also takes into account some control variables, such as purchase frequency, the number of items in the shopping basket, etc. To ensure the accuracy of the study results. The relationship between many factors of online clothing sales satisfaction is deeply discussed. By analyzing a large number of sales data, this paper gets a series of statistical results and calculates the extent to which all independent variables affect consumer satisfaction. When other potential variables are taken into account, the study finds that Payment Method and Discount Applied have a significant impact on consumers' satisfaction with online clothing sales. Based on the regression model of the research, some suggestions for merchants' future sales can be made. When formulating sales strategies, merchants should seriously consider these two factors and improve consumers' shopping experience through flexible use of discount activities and targeted payment promotion, so as to enhance sales performance. At the same time, future research can further explore other factors that may influence satisfaction to more fully understand the market dynamics of online clothing sales.

#### AUTHORS CONTRIBUTION

All the authors contributed equally and their names were listed in alphabetical order.

#### REFERENCES

- Y. Zheng, *Rel. Inv. Entr.* 8, 2 (2018).  
 B. Li and Y. X. Shi, *Economic daily news*, (2023).  
 H. Y. Yu, *Journal of marketing research* 1, 3 (2018).  
 Z. Y. Gong, *Global Market* 8, 17 (2019).  
 C. L. Lei, *Business research* 5, 3 (2016).  
 Y. M. Yu, *Harbin Institute of Technology*, (2015).  
 J. P. Cao, *Management* 1, 231 (2017).  
 W. T. Yu and M. Zou, *Journal of UESTC* 22(5), 10 (2020).  
 L. Bai, *Journal of commercial age*, (2019).  
 J. W. Ding, *Business Economics Research*, (2019).