Analysis on E-Commerce Companies' Business Strategies Based on Shipping Data

Yue Yu[1*], Yingyuan Lin[2]

Corresponding author. Email: yueyu0713@gmail.com[*]

School of Economics, Jilin University, Jilin, 130012, China[1]
Department of Mathematics, University of California, Los Angeles, 90095, United States[2]

ABSTRACT. COVID-19 brings chances accompanied by more fierce competition to the e-commerce business. Therefore, it requires e-commerce business owners to adjust their strategies to succeed. And this essay tries to provide insights into the e-commerce business by focusing on strategy of customer maintenance in the logistics aspect. The research uses multinomial logistic regression to examine what factors have significant impacts on customer satisfaction, and one-way ANOVA to explore strategies to retain customers. For the logistics aspect, the paper applies crosstab and chi-square test, and independent-sample t-test, to examine the significance between specific factors of customer behaviors, product characteristics, and shipping punctuality. The paper further explores the overall significant relationship between customer behaviors, product characteristics, and delivery punctuality by applying logistic regression. The analytical results support that consumer behaviors and product attributes significantly affect e-commerce shipping punctuality. The analysis of the results offers implications for e-commerce shipping strategies to succeed in the industry.

Keywords: e-commerce, customer maintenance strategy, shipping strategy, consumer behaviors.

1 Introduction

The COVID-19 pandemic has impacted various industries, either negatively or positively. Industries such as tourism, catering, and airlines, suffered from severe negative impacts during the pandemic. For instance, the pandemic placed over $100 million in direct tourism jobs at risk, and international travel cratered by 72% in 2020 compared to 2019 accompanied by a decrease in export revenues from 1.7 trillion dollars in 2019 to 651 billion dollars in 2020 [1]. On the other hand, the COVID pandemic has brought huge opportunities to some other sectors including e-commerce, online meeting, biomedical business, and so on. Particularly, the pandemic spurred the development of e-commerce businesses. During the COVID-19 period, companies made more efforts to develop their digital platforms such as adding direct-to-consumer operations and delivery services to satisfy consumers [2]. Also, a recent statistic shows that global e-commerce transactions jumped to $26.7 trillion during the pandemic, and e-commerce experienced a rise in its share of all retail sales, from 16% to 19% in 2020 [3]. A more dramatic data is that the revenue of global e-commerce could potentially grow to 5.545 trillion US dollars, and in 2025 up to 7.835 trillion US dollars, which is 2.2 times compared to 2019 [4]. The above facts and data not only imply a remarkable opportunity for the individual

ICBBEM 2023, May 19-21, Hangzhou, People's Republic of China

DOI 10.4108/eai.19-5-2023.2334253

e-commerce business to develop but creates a more competitive environment within the business. Therefore, COVID-19 requires the owners to adjust their business strategies to adapt to the rapidly changing conditions within the e-commerce sector.

If an individual e-commerce business owner wants to survive under the rapidly changing conditions within e-commerce and even achieve long-term success in this scenario, it is necessary to identify what are the most important factors that contribute to the success of e-commerce. To achieve this, among all factors, customer awareness, employee skills, and organizational infrastructure are some of the most effective factors that contribute to the success of an e-commerce business [5]. Besides these factors, logistics is also often described as the main source of comparative advantage for e-commerce [6]. In the meantime, customer loyalty is widely recognized as a path to long-term success. Finding new customers and doing business with them takes extra time, effort, and money [7]. And one effective way for building and maintaining customer loyalty is to provide outstanding logistic service [8]. This indicates the great role shipping plays in online sales and thus the optimization of the logistic process is meaningful in e-commerce.

The significance of shipping in e-commerce naturally attracts many research interests. Some studies focus on the prediction of the punctuality of the delivery and other research concentrates on the specific strategy's role in shipping. Most of them consider the productivity of workers, transport costs, etc. As the significant factors that impact shipping punctuality. However, there is a lack of attention on how e-commerce companies use strategies for customers with different loyalty and how these strategies work, and how consumers' behaviors and the attributes of the product can influence delivery punctuality. Therefore, this research aims at filling in the potential research gap by examining if companies' logistics strategies correlated to consumer behaviors, and product characteristics, and whether they would significantly affect shipping punctuality. If so, what business implications are behind that, and how do consumer behavior factors and product attributes affect on-time delivery? To resolve the research question, this study utilizes a dataset from an international e-commerce company's shipping data. The research first utilizes logistic regression to predict punctuality based on other factors in the dataset. The result shows that the prediction based on these factors has gratifying predictability. Moving forward, the paper applies multinomial logistic regression to explore whether a certain factor impacts customer rating. Based on the result, the paper uses one-way ANOVA as the main method, combined with Chi-square Tests and T-Tests for contrastive analysis and logistic regression to analyze the inner relationship between consumer behaviors and product attributes to connect it to e-commerce managing strategies and find implications. The paper is organized as follows. Section 2 discusses the data analysis methods that are used in the research. Section 3 describes the dataset used in the paper and provides exploratory data analysis. Section 4 shows the data analysis procedure and offers key insights into the results. Section 5 illustrates further discussion of the results and Section 6 concludes the findings.

2 Methodology

This paper applies an international e-commerce company's shipping data. First, exploratory data analysis results are shown to visualize the distribution of the variables in the dataset. To analyze if certain variables notably affect customer satisfaction and companies' strategies for different customers, the study applies multinomial logistic regression, the one-way ANOVA. Multinomial logistic regression is used to find out which factors significantly influence customer satisfaction. To see if other variables in the dataset contribute to the distinction between the proposed variable and the customers with various prior purchases, one-way ANOVA is used to assess their inner correlation. It can allow further analysis to explain the strategies companies use to promote customer loyalty. Moreover, combining crosstab and the chi-square test to study how discrete variables can potentially affect shipping punctuality is useful since the chi-square test first shows whether there is a significant difference among different groups in the discrete variable. And the crosstab method further shows what the difference between groups specifically is like by giving the percentage among groups. For continuous variables in the dataset, the study mainly applies an independent-sample t-test since it is useful for showing whether there is a significant distinction when one variable is binary and the other is continuous. These methods contribute to research on exploring what other potential factors affect order punctuality. For the prediction of delivery punctuality to show the overall significance between consumer behaviors, product characteristics, and shipping punctuality, the paper applies logistic regression. It is commonly used in the field of e-commerce research for prediction. For example, logistic regression is used as a basis to build a hybrid prediction model for customer churn in the realm of e-commerce [9]. This article applies logistic regression based on the forward conditional method which shows the prediction accuracy for whether the delivery is on time separately and gives the overall correctness of the prediction. Moving forward, the paper also tries to use factor analysis to group the variables with similar properties together. This skill reduces the original variables' dimensions that are used in the logistic regression. Using the new factors after grouping in the logistics can achieve better predictability of the regression model which makes the result more convincing.

3 Data description

The paper aims at discovering key insights from customer behaviors and their impact on the logistic aspects of e-commerce, through an international e-commerce company database with 10999 observations of 12 variables. After data filtering, four selected logistic criteria including Customer care calls, Cost of the product (in dollars), Discount offered, and Prior purchase are adopted for the analysis to explore their significant impacts on Reach on time_Y.N. Reached on time_Y. N is a binary variable where 0 represents the product has reached on time, and 1 indicates the unpunctual deliveries. The variable Customer care calls, which indicates the number of calls made from inquiries for shipment, is a discrete variable, with 6 groups including 2,3,4,5,6,7. The variable Cost of the product (in dollars) is a continuous variable, representing the price of the product purchased by the customers. The paper first determines the relationship between these three factors, which will be discussed in section 4.1.

Discount offered indicates a discount on the specific product, which varies from 1% to 65%. For the variable Prior Purchases, it has 8 groups, varying from 2 to 10. Before the detailed analysis of the data, some exploratory data analysis results will be shown below. This preliminary analysis shows the distribution of Reached on Time_Y.N's distribution among several variables in the dataset.

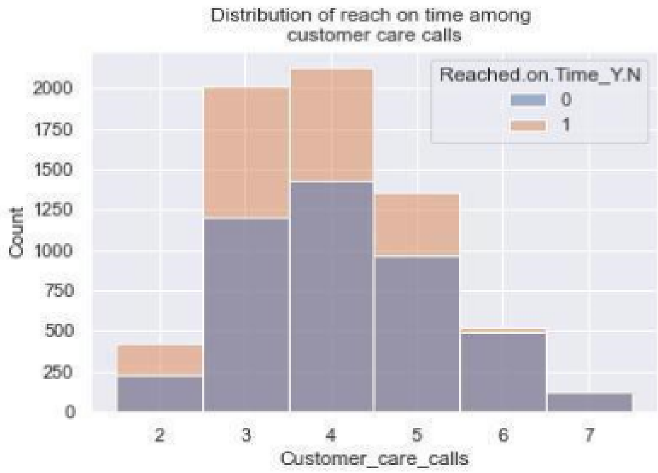

Fig. 1. Distribution of reach on time among customer care calls. (Owner-draw)

Figure 1 shows Reached on Time_Y.N's distribution among customer care calls. The result illustrates that the percentage of on-time delivery is higher when the customer calls are over three. This result spurs further analysis on exploring if customer care calls substantially impact shipping punctuality. Additionally, the chart below provides the primitive intention of doing a detailed analysis on whether prior purchases significantly affect Reached on Time_Y.N.

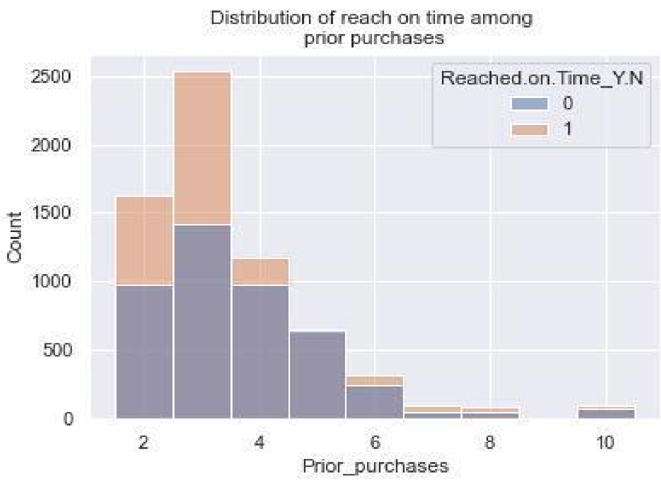

Fig. 2. Distribution of reach on time among prior purchases. (Owner-draw)

According to Figure 2, when the prior purchases are either 2 or 3, the percentage of on-time deliveries is around 60%. Also, the percentage of on-time deliveries is notably higher when the customer has prior purchases from 4 to 6. And when prior purchases range from 7 to 10, the percentage of on-time deliveries is also around 60%.

4 Data Analysis and Results

4.1 Related to the company's customer maintenance strategy

Under the fierce competition during the epidemic period, e-commerce faces the buyer's market, in which customers' preference decides whether they shop from this company or change to another one. So it is necessary for e-commerce companies to maintain customer relationships. Bowersox, Mentzer, and Speh found that improving a company's logistics capability can build a closer relationship with customers. Different customers may expect different levels of logistics services [10], so logistics services should be tailored to the needs of different customer groups [8]. Therefore, we use Prior purchases as a measure of customer loyalty and explore what logistics strategies e-commerce companies use to maintain customer loyalty. We take the Prior purchases as a Factor and other variables as a dependent list to do one-way ANOVA, the results indicating that customers having 2 and 3 prior purchase experiences show a similar pattern with customers having bought 7 or 8, or 10 times before. These two populations are named New and Old Customers. People with prior purchase times of 4,5,6 are grouped and named Intermediate Customers, showing different characteristics from New-and-old customers.

4.2 Factors that influence Customer satisfaction

4.2.1 Discount: the most influential factor in customer satisfaction

To find out what kind of factors impact customer satisfaction, the paper uses the method of multinomial logistics regression, with Customer Rating as the independent variable and Prior purchases, Reached on time, Discount offered, Mode of shipment, and Product importance as the factors, the results are shown below:

Table 1. Factors impact on customer satisfaction [Owner-draw]

Likelihood Ratio Tests				
Effect	Model Fitting Criteria -2 Log Likelihood of Reduced Model	Likelihood Ratio Tests		
		Chi-Square	df	Sig.
Intercept	14362.492[a]	000	0	
Prior purchases	14388.140	25.647	28	.592
Reached.on.Time Y.N	14368.538	6.046	4	.196
Discount offered	14661.393	298.901	256	.034
mode of shipment	14369.519	7.026	8	.534
Product importance	14369.545	7.053	8	.531
Note: a.This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.				

The Likelihood Ratio Tests show that only the Discount offered passed the significance test (significance <0.05), so the product discount is the most concerning factor in customer satisfaction (see Table 1).

4.2.2 Discount strategy: to satisfy new and old customers

Then the paper researched whether the company uses discount as a strategy by the One-way ANOVA method, using the Discount offered as the dependent list and prior purchase as the Factor, the results in Table 2 show that significant differences exist among the Prior purchase groups (significance<0.001). Table 3 illustrates that the mean of Intermediate customers is lower than the total mean (13.37), while the mean of the New and Old Customers is higher than 13.37, so the discount of the middle-distributed customers is less than the new and old customers, which can be considered as a marketing strategy for e-commerce companies to attract new customers by discounts to improve their satisfaction and retain the old customers by discounts for gratitude and rewards.

Table 2. ANOVA of each group in discount_offered [Owner-draw]

			Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)		60928.191	7	8704.027	33.836	<.001
	Linear Term	Weighted	19786.938	1	19786.938	76.919	<.001
		Deviation	41141.253	6	6856.875	26.655	<.001
Within Groups			2827356.758	10991	257.243		
Total			2888284.949	10998			

Note:
df : abbreviation of degree of freedom
Sig. : abbreviation of significance

Table 3. Descriptive features of each group in discount_offered [Owner-draw]

					95% Confidence interval for Mean			
	N	Mean	Std.Deviation	Std.Error	Lower Bound	Upper Bound	Minimum	Maximum
2	2599	15.32	17.422	.342	14.65	15.99	1	65
3	3955	14.98	17.318	.275	14.44	15.52	1	65
4	2155	10.85	14.040	.302	10.25	11.44	1	65
5	1287	9.27	11.980	.334	8.61	9.92	1	65
6	561	11.06	14.800	.625	9.84	12.29	1	65
7	136	17.53	18.954	1.625	14.32	20.74	1	65
8	128	13.32	15.702	1.388	10.57	16.07	1	62
10	178	13.60	15.277	1.145	11.34	15.86	1	64
Total	10999	13.37	16.206	.155	13.07	13.68	1	65

Note:
Std. : abbreviation of standard

4.3 Companies' strategy to maintain intermediate customers

Reflecting the current state of products and the arrival date of products are features of Customer service management [11], and logistics capabilities like customer responsiveness and competing on time provide core competencies for corporate strategy [12]. So the paper served Customer care calls and Reach on time as factors representing the company's service level, researching its customer strategies toward different customers.

4.3.1 Intermediate customers receive more calls

The research found that Intermediate customers received more care calls than new and old customers. Use Customer care calls as the dependent list and Prior purchases as the Factor for One-way ANOVA, the results in Table 4 demonstrates that significant differences exist among groups in the Customer care calls.

Table 4. ANOVA of each group in Customer care calls [Owner-draw]

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2188.902	7	312.700	283.070	<.001
Within Groups	12141.477	10991	1.105		
Total	14330.379	10998			

Table 5. Descriptive features of each group in Customer care calls [Owner-draw]

					95% Confidence interval for Mean			
	N	Mean	Std.Deviation	Std.Error	Lower Bound	Upper Bound	Minimum	Maximum
2	2599	3.71	0.884	0.017	3.68	3.74	2	6
3	3955	3.73	0.925	0.015	3.70	3.76	2	7
4	2155	4.76	1.418	0.031	4.70	4.82	2	7
5	1287	4.57	1.035	0.029	4.51	4.62	2	7
6	561	4.27	1.042	0.044	4.19	4.36	2	6
7	136	4.01	0.996	0.085	3.85	4.18	2	6
8	128	3.83	1.028	0.091	3.65	4.01	2	6
10	178	3.67	0.912	0.068	3.54	3.81	2	6
Total	10999	4.05	1.141	0.011	4.03	4.08	2	7

Table 5 shows that the mean of new and old customers is lower than the total mean (4.05), while the mean of the intermediate customers is higher than 4.05, so the calls that respond to new and old customers are fewer than the intermediate customers. However, Intermediate customers are more punctual and get more care calls than New and old customers, which may be because the e-commerce company pays more attention to these customers' demands, so provide more care calls and punctual orders to cultivate their intermediate customers to older customers.

4.3.2 Intermediate customers get more punctual orders

The Chi-square Tests and Crosstab results illustrate the relationship between punctual products and customer types. From Table 6, it can be seen that there are significant differences between the binary groups of Reach on time(significance<0.001), and the intermediate customers received more punctual goods than the other customers.

Table 6. Chi-Square Tests of Prior purchase and Reached on time [Owner-draw]

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	125.922[a]	7	<.001
Likelihood Ratio	125.226	7	<.001
Linear-by-Linear Association	33.895	1	<.001
N of Valid Cases	10999		
Note: a.0 cells (0.0%) have expected count less than 5. The minimum expected count is 51.62.			

Table 7. Reached.on.Time_Y.N * Prior_purchases Crosstabulation [Owner-draw]

			Prior Purchase								Total
			2	3	4	5	6	7	8	10	
Reached.on.Time	0	Count	974	1421	984	645	247	44	45	76	4436
		% within Reached.on.Time	22.0%	32.0%	22.2%	14.5%	5.6%	1.0%	1.0%	1.7%	100.0%
	1	Count	1625	2534	1171	642	314	92	83	102	6563
		% within Reached.on.Time	24.8%	38.6%	17.8%	9.8%	4.8%	1.4%	1.3%	1.6%	100.0%
Total		Count	2599	3955	2155	1287	561	136	128	178	10999
		% within Reached.on.Time	23.6%	36.0%	19.6%	11.7%	5.1%	1.2%	1.2%	1.6%	100.0%

From Table 7, it can be noticed that in 22% of the punctual orders, customers have bought two times before, while 24.8% of customers with unpunctual orders have 2 purchases. The comparison of these two numbers illustrates that among the group with 2 prior purchases, the probability of the order being unpunctual will be higher than being on time. By comparing the other groups, it seems more possible for the new and old ones to get an unpunctual order, while the intermediate customers have a higher rate of reaching on time. Because close relations are resource intensive, only some customers are worth caring for, and this is probably the company's strategy that cares more about the intermediate customers than others[13].

4.4 Further exploration for punctuality

Section 4.3 explains the company’s strategies to the kind of customers well, but we further research the possible reason why intermediate customers get more punctual products.

4.4.1 Intermediate customers’ behaviors impact punctuality

4.4.1.1 Customers’ calls urge the company to reach on time

Using the method of crosstab and chi-square, with Reach on time as the row and Customer care calls as the column, the results are shown as Table 8.

Table 8. Chi-Square Tests of Customer care calls and Reach on time [Owner-draw]

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	54.274[a]	5	<.001
Likelihood Ratio	53.936	5	<.001
Linear-by-Linear Association	49.556	1	<.001
N of Valid Cases	10999		
Note：a.0 cells (0.0%) have expected count less than 5. The minimum expected count is 99.21.			

The chi-square tests result shows that the Asymptotic significance is <0.001, indicating that there are significant differences among groups.

Table 9. Reached.on.Time_Y.N * Customer_care_calls Crosstabulation [Owner-draw]

			Customer_care_calls						Total
			2	3	4	5	6	7	
Reached.on.Time	00	Count	222	1206	1431	968	490	119	4436
		% within Reached.on.Time	5.0%	27.2%	32.3%	21.8%	11.0%	2.7%	100.0%
	11	Count	416	2011	2126	1360	523	127	6563
		% within Reached.on.Time	6.3%	30.6%	32.4%	20.7%	8.0%	1.9%	100.0%
Total		Count	638	3217	3557	2328	1013	246	10999
		% within Reached.on.Time	5.8%	29.2%	32.3%	21.2%	9.2%	2.2%	100.0%

The percentage of customer care calls within Reached on Time_Y.N. indicates the proportion of various times of care calls in punctual or unpunctual orders. The data shows that in 5% of the punctual orders, customers have made two phone calls, while 6.3% of unpunctual order has 2 customer care calls **(see Table 9).** The comparison of these two numbers shows that among the group with 2 customer care calls, the probability of the order being unpunctual will be higher than being on time. By comparing the other groups, it is more possible for customers who called 2 / 3 / 4 calls to get an unpunctual order, while customers who made 5 / 6 / 7 calls have a higher rate of reaching on time. So the customers with more calls are more likely to get on-time deliveries, which may be because customers' urging makes the company pay more attention.

4.4.1.2 Intermediate customers buy more expensive products and insurance incentives for more punctuality

Intermediate customers show another feature that they buy more expensive products than the other customers. The one-way ANOVA and descriptive results are as below (see Table10&Table 11), indicating that different groups of product price show significant differences, and the average product prices cost by intermediate customers are all higher than the total mean ($210.20), while the cost by new and old customers are all lower than $210.20.

Table 10. ANOVA of each group in Cost of the Product [Owner-draw]

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1781655.519	7	254522.217	118.413	.000
Within Groups	23624583.33	10991	2149.448		
Total	254.6238.85	10998			

Table 11. Descriptive features of each group in Cost of the Product [Owner-draw]

					95% Confidence interval for Mean			
	N	Mean	Std.Deviation	Std.Error	Lower Bound	Upper Bound	Minimum	Maximum
2	2599	201.36	43.562	854	199.69	203.04	96	296
3	3955	200.15	43.953	699	198.78	201.52	96	310
4	2155	228.99	53.251	1.147	226.74	231.24	96	310
5	1287	226.88	46.495	1.296	224.34	229.42	98	310
6	561	217.67	47.038	1.986	213.77	221.57	97	293
7	136	204.96	45.726	3.921	197.20	212.71	104	286
8	128	199.92	45.831	4.051	191.91	207.94	127	279
10	178	202.17	46.320	3.472	195.32	209.02	108	284
Total	10999	210.20	48.063	458	209.30	211.10	96	310

The T-Test results in Table 12 concerning Reach on time and Cost of products shows that significant differences exist among groups in the variable Reach on time and Table 13 demonstrates that the average price of punctual products ($214.5) is higher than the unpunctual ones ($209.29).

Table 12. Independent Samples Test of cost of the product and reach on time. [Owner-draw]

		t-test for Equality of Means				
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Cost_of_the_Product	Equal variances assumed	7.738	10997	<0.001	7.209	0.932
	Equal variances not assumed	7.747	9557.164	<0.001	7.209	0.931
Note: 1.df refers to degree of freedom; 2.Sig. refers to significance level						

Table 13. Group statistics of price divided by punctuality [Owner-draw]

	Reached.on.Time Y.N	N	Mean	Std.Deviation	Std.Error Mean
Cost_of_the_Product	0	4436	214.50	47.757	.717
	1	6563	207.29	48.055	.593

The possible reason may be that most e-commerce firms would purchase cargo transport insurance for substantial monetary compensation, and the insurance amount increases as the monetary amount of losses incurred rises [14]. Therefore, the company will be more careful to arrive on time to avoid the high premium. On the one hand, intermediate customers' high expenditure seems to provocate punctuality through insurance, but on the other hand, there may be other substantial policies that lead to their preferences of delivering more expensive products, which could not be confirmed due to limited columns of the dataset.

4.4.2 Logistic regression for predicting shipping punctuality

To show the overall significant relationship between consumer behaviors, product attributes, and shipping punctuality, this paper applies logistic regression to predict delivery time based on other variables in the dataset. Logistic regression is a commonly used model in various research fields, including prediction related to e-commerce. For example, the linear logistic model is applied to try to optimize online shopping behavior prediction [15]. Here, if the model shows a satisfying predicting power, it means that consumer behaviors and product characteristics significantly impact shipping punctuality. Using the forward conditional method in the logistic regression based on a 95% confidence interval, Table 14 shows the significant level of variables in the equation.

Table 14. Variables in the logistic equation before Factor Analysis.

	B	S.E.	Wald	df	Sig.	Exp(B)
step 1 Discount offered	.126	.004	845.438	1	<.001	1.134
Constant	-.737	.035	437.298	1	<.001	479
Step 2 Discount offered	.118	.004	716.529		<.001	1.125
Weight in gms	.000	.000	139.648	1	<.001	1.000
Constant	-.052	.068	593	1	.441	949
Step3 Customer care calls	-.137	.021	43.765	1	<.001	872
Djscount offered	.114	.004	665.690	1	<.001	1.121
Weight in gms	.000	.000	180.229	1	<.001	1.000
Constant	.702	.133	27.986	1	<.001	2.018
Step 4 Customer care calls	-.129	.021	38.463	1	<.001	879
Prior purchases	-.075	.015	24.596	1	<.001	928
Discount offered	.113	.004	649.675	1	<.001	1.120
Weight in gms	.000	.000	199.551	1	<.001	1.000
Constant	1.013	.147	47.325	1	<.001	2.753
Step 5 Customer care calls	-.110	.021	26.056	1	<.001	896
Cost of the Product	-.002	.000	15.452	1	<.001	998
Prior purchases	-.073	.015	23.395	1	<.001	930
Discount offered	.112	.004	634.360	1	<.001	1.118
Weight in gms	.000	.000	212.317	1	<.001	1.000

Constant	1.406	.179	61.852	1	<.001	4.080
Step 6 Customer care calls	-.107	.021	24.970	1	<.001	898
Cost of the Product	-.002	.001	15.265	1	<.001	998
Prior purchases	-.076	.015	25.241	1	<.001	927
Product importance	.101	.035	8.522	1	.004	1.107
Discount offered	.112	.004	630525	1	<.001	1.118
Weight in gms	.000	.000	218.731	1	<.001	1.000
Constant	1.266	.185	46.765	1	<.001	3.545

Note:
a.Variable(s)entered on step 1: Discount offered.
b. Variable(s) entered on step 2: Weight in gms.
c.Variable(s) entered on step 3:Customer care calls.
d.Variable(s)entered on step 4: Prior purchases.
e.Variable(s)entered on step 5:Cost of the Product.
f.Variable(s) entered on step 6:Product importance.
g.S.E. refers to Standard Error
h.df refers to degree of freedom
i.Sig. refers to significance level

From Table 14, it can be seen that the model includes one more variable in the equation for each step. In the end, the model includes customer care calls, product cost, prior purchase, product importance, discount offered, and product weight into the model and excludes gender, customer rating, and mode of shipment. Among the variables excluded, customer rating is a process after the product is delivered and customers rate the service based on punctuality and other criteria. Therefore, it is legitimate to exclude it. Due to the e-commerce company giving different predicted delivery times for different methods, the model precludes the shipping method. Thus, it is unrealistic to predict shipping punctuality based on this. More importantly, the significance column illustrates that all the variables in this model have a significant level of less than 0.05 which means that this logistic regression model is statistically significant. Therefore, it is fair to assume that the model will show an overall satisfying predictability.

Table 15. Logistic regression predicting accuracy.

				Predicted
Observed		Reached on Time Y.N.		Percentage Correct
		0	1	
Step 1 Reached.on.Time Y.N	0	2186	2250	49.3
	1	1971	4592	70.0
Overall Percentage				61.6
Step 2 Reached.on.Time Y.N	0	2482	1954	56.0
	1	2102	4461	68.0
Overall Percentage				63.1
Step 3 Reached.on.Time Y.N	0	2517	1919	56.7
	1	2085	4478	68.2
Overall Percentage				63.6
Step 4 Reached.on.Time Y.N	0	2531	1905	57.1
	1	2077	4486	68.4
Overall Percentage				63.8
Step 5 Reached.on.Time Y.N	0	2551	1885	57.5
	1	2089	4474	68.2
Overall Percentage				63.9
Step 6 Reached.on.Time Y.N	0	2565	1871	57.8
	1	2102	4461	68.0
Overall Percentage				63.9
Note: a. The cut value is .500				

The classification table 15 demonstrates that the overall percentage correct in predicting shipping punctuality is 63.9% based on the variables chosen from the data. It implies that most of the variables of customer behaviors and product attributes are significantly affecting shipping punctuality.

4.4.3 Factor analysis

To further enhance this argument, the study uses factor analysis. This approach is commonly used in prediction in various fields. For example, factor analysis is used to establish a more effective model in financial prediction combined with logistic regression [16]. Here, the paper includes all the variables except reaching on time. First, the KMO and Bartlett's Test in Table 16 shows that the model has a significant level that is less than 0.01 and thus it is statistically significant, which makes further analysis convincing.

Table 16. KMO and Bartlett's Test result for Factor Analysis

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.490
Bartlett's Test of Sphericity Approx.Chi-Square	Approx.Chi-Square	5744.015
	df	36
	Sig.	<.001
Note: a.df refers to degree of freedom b.Sig refers to significance level		

Table 17. Total variance explained by factors

Total Variance Explained									
	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	Sums of Square % of Variance	Loadings Cumulative %	Rotation Total	Sums of Square % of Variance	Loadings Cumulative %
1	1.632	18.130	18.130	1.632	18.130	18.130	1.438	15.977	15.977
2	1.355	15.055	33.185	1.355	15.055	33.185	1.379	15.326	31.303
3	1.026	11.398	44.584	1.026	11.398	44.584	1.047	11.630	42.933

4	1.011	11.238	55.822	1.011	11.238	55.822	1.012	11.241	54.174
5	1.000	11.109	66.931	1.000	11.109	66.931	1.001	11.125	65.299
6	.984	10.929	77.860	.984	10.929	77.860	1.000	11.115	76.414
7	.870	9.665	87.525	.870	9.665	87.525	1.000	11.111	87.525
8	.675	7.501	95.026						
9	.448	4.974	100.000						
Note: Extraction Method: Principal Component Analysis									

Furthermore, Table 17 shows that the first 7 factors can explain 87.525% total variance, which is a significant amount. Based on this, the factor analysis forms new seven factors by grouping the original 9 factors to reduce the dimension. And the rotated component matrix shows the result after grouping.

Table 18. Factors after regrouping using rotated component matrix

Rotated Component Matrix							
Component							
	1	2	3	4	5	6	7
Customer care calls	.776						
Cost of the Product	.820						
Prior purchases			.974				
Product importance				.994			
Customer rating							1.000
Discount offered		.854					
Weight in gms		-.799					
Mode of shipment					.999		
gender						1.000	
Note: Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization a. Rotation converged in 5 iterations.							

The rotated component matrix in Table 18 indicates that the level of importance is decreasing from column 1 to column 7. Thus, the model groups customer care calls, and the cost of the product together to form the most important factor. Similarly, discounted offered and weight forms the second most important factor.

4.4.4 Logistic Regression After Factor Analysis

The significance of the factor analysis model, the total variance explained by the 7 factors, and the decreasing importance of the seven factors after grouping allow the paper to do another logistic regression based on the result. The new logistic regression includes the first four of the seven grouped factors in the model since the previous result indicates that the first four grouped factors are the most important ones that have significant impacts on shipping punctuality. And the classification table shows the prediction results based on these four factors.

Table 19. Logistic regression predicting accuracy after factor analysis

Classification Table				
Predicted				
Observed		Reached.on.Time Y.N		Percentage Correct
		0	1	
Step 1 Reached.on.Time Y.N	0	2922	1514	65.9
	1	2234	4329	66.0
Overall Percentage				65.9
Step 2 Reached.on.Time Y.N:	0	2471	1965	55.7
	1	1944	4619	70.4
Overall Percentage				64.5
Step 3 Reached.on.Time Y.N:	0	2385	2051	53.8
	1	1875	4688	71.4
Overall Percentage				64.3
Step 4 Reached.on.Time Y.N:	0	2407	2029	54.3
	1	1874	4689	71.4
Overall Percentage				64.5
Note: The cutvalue is .500				

Table 19 shows that the model overall predicts 64.5% of the reach on time correctly. Compared to the previous result which is 63.9%, this is only a fraction increase in overall accuracy. However, this model performs better in predicting the product is not reached on time by increasing the accuracy from 68% to 71.4%. In this sense, the new model based on the result of factor analysis is more powerful for predicting the product is not on time compared to the original intention of improving the overall performance. Overall, the logistic regression before and after the factor analysis shows that variables except for sex, shipping mode, and customer rating have an eminent predicting power in predicting shipping punctuality. Among the precluded variables, sex is not related to either customer behaviors or product attributes. For customer rating, although it is part of customer behaviors, the rating is always based on the experience after the consumers receive the product. Therefore, it is reasonable not to incorporate it as a predictor variable. For shipping mode, although it can be considered as a part of product attributes and predicted delivery time is based on different delivery methods, it can not affect whether a product arrives on time or not. Therefore, the satisfying model predicting performance support that consumer behaviors and product attributes significantly affect shipping punctuality.

5 Conclusion

In today's dynamic and competitive environment, it is difficult for e-commerce companies to compete on products alone as the market provides more opportunities for customers to choose similar products and product features. Instead, of its ability to create barriers, supply chain relationships can be a stable source of competitive advantage [13]. Therefore, it is necessary to explore how e-commerce companies maintain customers in their delivery aspect. From the above analysis, we identify that consumer behaviors and product attributes have significantly affected shipping punctuality overall. Going further, we find that in order to maintain customers, e-commerce companies have different logistics service strategies for new and old customers and intermediate customers. E-commerce companies can retain new and old customers by focusing on discounts, while they can retain intermediate customers through care calls and on time. So for new and old customers, e-commerce companies can carefully choose discount methods to spur them to consume. For example, they can apply the most common restrictive discount promotion strategy the minimum purchase limit and the maximum purchase requirement to attract both new and old consumers. Note that they should carefully choose their use of them since they both generate a stronger purchase intention and a sense of trouble when their purchases are restricted [17]. For intermediate users, e-commerce companies can provide customer care calls that effectively deal with customer concerns and endeavor to make more on-time deliveries to maintain them. Besides, the middle customer seems to deliver more expensive products, we guess this may also be one of the reasons for their goods reaching more on time, as more compensation would be paid by insurance companies for more expensive products reaching over-time, and it will eventually reflect on company's increasing insurance premium. However, whether some substantial policies would lead to the customers' high expenditure still need to explore. The paper interprets the data analysis result and provides insights into how e-commerce can maintain new, intermediate, and old customers by differentiating their shipping and other strategies. Based on this, the article explores the impact of shipping punctuality by providing justification regarding factors in consumer behaviors and product attributes that can significantly.

REFERENCE

[1] UNWTO, July 2022, “Impact Assessment of the COVID-19 Outbreak On International Tourism,” https://www.unwto.org/impact-assessment-of-the-covid-19-outbreak-on-international-tourism

[2] Lin, Xiao, Xusen Cheng and Jian Mou, Journal of Global Information Technology Management, 2022, “Understanding Global E-Commerce Development During the COVID-19 Pandemic: Technology - organization - environment Perspective,” https://www.tandfonline.com/doi/abs/10.1080/1097198X.2022.2026658

[3] UN News, May 3, 2021, “Global e-commerce jumps to $26.7 trillion, fuelled by COVID-19,” https://news.un.org/en/story/2021/05/1091182

[4] Zenkina, E. V., BENEFICIUM научное периодическое сетевое издание 1 (2022): 68-73, 2022 "ABOUT CURRENT TRENDS IN GLOBAL E-COMMERCE," https://beneficium.pro/index.php/beneficium/article/view/BENEFICIUM.2022.1%2842%29.68-73/186

[5] Mahdi Choshin, Ali Ghaffari, Computers in Human Behavior, Volume 66, 2017, Pg 67-74, “An investigation of the impact of effective factors on the success of e-commerce in small- and medium-sized companies”, https://www.sciencedirect.com/science/article/abs/pii/S0747563216306537#preview-section-cited-by

[6] Gajewska, T., i D. Zimon. Archives of Transport, t. 45, nr 1s. 25–34., Polska Akademia Nauk. Czytelnia Czasopism PAN, 2018, “Study of the Logistics Factors That Influence the Development of E-Commerce Services in the Customer’s Opinion,” https://bibliotekanauki.pl/articles/223556

[7] Mittal, B., and Lassar, W.M., Journal of Services Marketing, Vol. 12 No. 3, pp. 177-194., 1998, "Why do customers switch? The dynamics of satisfaction versus loyalty," https://doi.org/10.1108/08876049810219502

[8] Davis, Elizabeth R., 2006, “The role of logistics service quality in creating customer loyalty,” https://www.proquest.com/docview/304980562?fromopenview=true&pq-origsite=gscholar

[9] Li, Xueling, Zhen Li, EBSCO, October 2019, “A Hybrid Prediction Model for E-Commerce Customer Churn Based on Logistic Regression and Extreme Gradient Boosting Algorithm,”https://web.p.ebscohost.com/abstract?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=16331311&AN=141446653&h=O6atOxGl1QXhU0gRqxq4869JMIYUAQgIFdjfroUQfMjVJfrx1ysnyOcQf8dTWts0cVh%2fytHRWrxhHxiji5R6Kg%3d%3d&crl=c&resultNs=AdminWebAuth&resultLocal=ErrCrlNotAuth&crlhashurl=login.aspx%3fdirect%3dtrue%26profile%3dehost%26scope%3dsite%26authtype%3dcrawler%26jrnl%3d16331311%26AN%3d141446653

[10] Bowersox, Donald J., John. T. Mentzer, and Thomas W. Speh, Journal of Business Strategy, Vol. 12, No. 1, pp. 36-49, 1995, “Logistics Leverage,” https://jbs-ojs-shsu.tdl.org/jbs/article/view/260

[11] Ying, Wang, and Sang Dayong, Expert Systems with Applications 29.2, 2005, "Multi-agent framework for third party logistics in E-commerce," https://www.sciencedirect.com/science/article/abs/pii/S0957417405000916

[12] Morash, Edward A., Cornelia LM Droge, and Shawnee K. Vickery, Journal of business logistics, 1996, "Strategic logistics capabilities for competitive advantage and firm success," https://www.econbiz.de/Record/strategic-logistics-capabilities-for-competitive-advantage-and-firm-success-morash-edward/10007001249

[13] Day, George S., Journal of the academy of marketing science, 2000, "Managing market relationships," https://link.springer.com/article/10.1177/0092070300281003

[14] Wu, Pei-Ju, Mu-Chen Chen, and Chih-Kai Tsau, International Journal of Physical Distribution & Logistics Management, 2017, "The data-driven analytics for investigating cargo loss in logistics

systems," https://scholar.nycu.edu.tw/en/publications/the-data-driven-analytics-for-investigating-cargo-loss-in-logisti
[15] Liu C-J, Huang T-S, Ho P-T, Huang J-C, Hsieh C-T, 2020, “Machine learning-based e-commerce platform repurchase customer prediction model,” https://doi.org/10.1371/journal.pone.0243105
[16] D. Han, L. Ma, and C. Yu, 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing, 2008, "Financial Prediction: Application of Logistic Regression with Factor Analysis," https://ieeexplore.ieee.org/abstract/document/4680497
[17] J. Tong and J. Shen, 2021 2nd International Conference on E-Commerce and Internet Technology (ECIT), 2021, "Research on Customer Satisfaction under Two Restrictive Discount Promotion Strategies Based on SPSS analysis software," https://ieeexplore.ieee.org/abstract/document/9406926